UNIVERSADADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE PESQUISAS HIDRÁULICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS HÍDRICOS E SANEAMENTO AMBIENTAL

BRUNO CÉSAR COMINI DE ANDRADE

MULTI-SCALE ACTUAL EVAPOTRANSPIRATION MAPPING IN SOUTH AMERICA WITH REMOTE SENSING DATA AND THE GEESEBAL MODEL

PORTO ALEGRE

2023

BRUNO CÉSAR COMINI DE ANDRADE

MULTI-SCALE ACTUAL EVAPOTRANSPIRATION MAPPING IN SOUTH AMERICA WITH REMOTE SENSING DATA AND THE GEESEBAL MODEL

Tese apresentada ao Programa de Pós-graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do grau de doutor.

Orientador: Anderson Ruhoff

PORTO ALEGRE

2023

Comini de Andrade, Bruno César
MULTI-SCALE ACTUAL EVAPOTRANSPIRATION MAPPING IN SOUTH AMERICA WITH REMOTE SENSING DATA AND THE GEESEBAL MODEL / Bruno César Comini de Andrade. -- 2023.
191 f.
Orientador: Anderson Ruhoff.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas, Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental, Porto Alegre, BR-RS, 2023.

1. Balanço de Energia da Superfície. 2. Mudança Climática. 3. Desmatamento. 4. Evapotranspiração real. 5. Hidrologia. I. Ruhoff, Anderson, orient. II. Título.

BRUNO CÉSAR COMINI DE ANDRADE

MULTI-SCALE ACTUAL EVAPOTRANSPIRATION MAPPING IN SOUTH AMERICA WITH REMOTE SENSING DATA AND THE GEESEBAL MODEL

Tese apresentada ao Programa de Pós-graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do grau de doutor.

Aprovado em: Porto Alegre, 28 de abril de 2023.

_____

Prof. Dr. Anderson Ruhoff – Universidade Federal do Rio Grande do Sul
Orientador

_____

Prof. Dr. Walter Collischonn – Universidade Federal do Rio Grande do Sul
Examinador

_____

Dr. Gabriel Senay – United States Geological Survey
Examinador

_____

Dr. Trent Wade Biggs – San Diego State University
Examinador

AGRADECIMENTOS

Agradeço as seguintes instituições pelo suporte financeiro: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Agência Nacional de Águas e Saneamento Básico (ANA), Fulbright, Universidade Federal do Rio Grande do Sul (UFRGS) e Instituto de Pesquisas Hidráulicas (IPH).

Sou grato aos meus professores, pelo aprendizado e amizade, e aos meus colegas no IPH, em especial ao Leonardo, Rafael e Ayan, cuja colaboração foi crucial para o desenvolvimento desta pesquisa.

Agradeço aos colegas pesquisadores que colaboraram comigo nos trabalhos que executei durante o doutorado. Sou grato aos membros da banca de qualificação, por seus comentários e sugestões, que contribuíram para a qualidade da minha tese e me inspiraram a avançar nas investigações que iniciei.

Sou grato aos pesquisadores do Desert Research Institute (DRI), Charles Morton, Chris Pearson, Chris Dunkerly, Christine Albano, e John Volk, por toda a ajuda que me deram durante meu doutorado sanduíche. Agradeço especialmente ao meu supervisor de intercâmbio, Justin Huntington, por me acolher e tornar minha experiência longe de casa muito mais agradável.

Sou profundamente grato ao meu orientador, Anderson Ruhoff, por sua dedicação e perseverança que viabilizaram meu sonho de concluir o doutorado. Sua mentoria, profissionalismo e amizade são um exemplo que espero seguir como pesquisador e professor.

Agradeço aos meus pais e à minha família que sempre me amaram e me apoiaram.

Finalmente, agradeço à Carolina, minha companheira, por seu amor, apoio, muita paciência, e pelas diversas dificuldades e alegrias que passamos juntos.

RESUMO

O monitoramento preciso da evapotranspiração (ET) é crucial para gerenciar os recursos hídricos, garantir a segurança alimentar e avaliar os impactos das mudanças climáticas. Modelos de Balanço de Energia da Superfície (SEB) que usam dados de sensoriamento remoto são os mais confiáveis para estimar a ET, mas muitas vezes são difíceis de aplicar em grande escala devido ao longo tempo de processamento, necessidade de calibração local, entre outros obstáculos. Esta tese tem como foco a melhoria do geeSEBAL, uma implementação do modelo *Surface Energy Balance Algorithm for Land* (SEBAL) na plataforma *Google Earth Engine* (GEE), adaptando-o para modelagem em escala continental, usando imagens do *Moderate Resolution Imaging Spectroradiometer* (MODIS). O novo modelo, chamado geeSEBAL-MODIS, foi usado para gerar uma série temporal de ET a cada 8 dias para a América do Sul com pixels de 500 m. Estudos de validação mostram que o geeSEBAL-MODIS é mais preciso do que outros produtos globais de ET, com uma redução do erro de 13% na escala de campo e 30% na escala de bacia hidrográfica. O conjunto de dados está disponível para o público e pode ser usado para monitorar tanto mudanças climáticas em grande escala quanto as variações locais de ET relacionadas às atividades humanas. A análise de tendências mostra um aumento de 8,4% na ET sobre a América do Sul, associado ao aumento da demanda atmosférica, e à redução da precipitação e disponibilidade de água. Esses resultados destacam a importância de informações precisas sobre os processos do ciclo hidrológico para auxiliar no planejamento e gerenciamento dos recursos hídricos em um cenário de maior escassez. Nesse contexto, projetos como o OpenET, que fornece dados confiáveis e de alta resolução espacial de ET nos Estados Unidos, são cruciais para monitorar o consumo de água e auxiliar no desenvolvimento sustentável. Este trabalho também apresenta uma reprodução parcial do processo do OpenET para a intercomparação de modelos de sensoriamento remoto com dados de torres de fluxo, usando torres micrometeorológicas na América do Sul. Os resultados são promissores e abrem caminho para a expansão do OpenET além dos Estados Unidos e em direção a uma aplicação global.

Palavras-chave: Balanço de energia da superfície. Mudança climática. Desmatamento. MODIS. Landsat. SEBAL. OpenET.

ABSTRACT

Accurately monitoring evapotranspiration (ET) is crucial for managing water resources, ensuring food security, and assessing the impacts of climate change. Surface Energy Balance (SEB) models that use remote sensing data are the most reliable for estimating ET, but they are often challenging to apply on a large scale due to long processing times, and local calibration requirements, among other obstacles. This dissertation focuses on improving geeSEBAL, an implementation of the Surface Energy Balance Algorithm for Land (SEBAL) model on the Google Earth Engine (GEE) platform, by adapting it for continental-scale modeling using Moderate Resolution Imaging Spectroradiometer (MODIS) images. The new model, called geeSEBAL-MODIS, was used to generate a temporal series of ET every 8 days for South America with pixels of 500 m. Validation studies show that geeSEBAL-MODIS is more accurate than other global ET products, with a reduction in error of 13% at the field scale and 30% at the basin scale. The dataset is publicly available and can be used to monitor both large-scale climate change and local ET variations related to human activities. Trend analysis shows an 8.4% increase in ET over South America, associated with increased atmospheric demand, and reductions in precipitation and water availability. These findings underscore the importance of accurate information on hydrological cycle processes to assist in planning and managing water resources in a scenario of greater scarcity. In this context, projects like OpenET, which provides reliable and high spatial-resolution ET data in the United States, are crucial for monitoring water consumption and aiding in sustainable development. This work also presents a partial reproduction of the OpenET process for intercomparing remote sensing models with flux tower data, using micrometeorological towers in South America. The results are promising and pave the way for expanding OpenET beyond the United States and toward global application.

Keywords: Surface energy balance. Climate change. Deforestation. MODIS. Landsat. SEBAL. OpenET.

ACRONYMS AND ABBREVIATIONS

A_0	Linear regression slope coefficient
ALEXI	Atmosphere–Land Exchange Inverse
ANN	Artificial Neural Network
AOI	Area of Interest
ASTER	Advanced Spaceborne Thermal Emission and Reflectance Radiometer
AVHRR	Advanced Very High-Resolution Radiometer
BESS	Breathing Earth System Simulator
BR	Bowen Ratio
CONUS	Continental USA
DisALEXI	Atmosphere–Land Exchange Inverse flux disaggregation approach
DRI	Desert Research Institute
dS/dt	Storage variation in time
dT	Temperature gradient
e_a	Atmospheric actual vapor pressure
EC	Eddy Covariance
ECOSTRESS	ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station
EDF	Environmental Defense Fund
ENSO	*El Niño* Southern Oscillation
e_s	Atmospheric saturation vapor pressure
ET	Actual evapotranspiration
ET_0	Evapotranspiration for a grass reference surface
ET_p	Potential evapotranspiration
ET_r	Evapotranspiration for an alfalfa reference surface
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization
G	Soil heat flux
Gc	ANN model with land cover input
GEE	Google Earth Engine
geeSEBAL	SEBAL implementation on the GEE platform
geeSEBAL-MODIS	MODIS adaptation of the geeSEBAL model; Dataset derived from this model version
GLDAS	Global Land Data Assimilation System
GLEAM	Global Land Evaporation Amsterdam Model
Gnc	ANN model without land cover input

GOES	Geostationary Operational Environmental Satellite
GRACE	Gravity Recovery and Climate Experiment
H	Sensible heat flux
ITCZ	Inter Tropical Convergence Zone
k_c	Crop coefficient
k_s	Soil water stress coefficient
LAI	Leaf Area Index
LBA-ECO	Large-scale Biosphere-Atmosphere Experiment in the Amazon
LE	Latent heat flux
LST	Land Surface Temperature
MAE	Mean Absolute Error
METRIC	Mapping Evapotranspiration with Internalized Calibration
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared radiation
ONDACBC	National Observatory of Water and Carbon Dynamics in the Caatinga Biome
P	Precipitation
PELD	Long-term ecological research in Pantanal
PM	Penman Monteith
PML	Penman-Monteith-Leuning
PT	Priestley-Taylor
PT-JPL	Priestley–Taylor Jet Propulsion Laboratory
Q	Discharge
r	Correlation coefficient
RMSE	Root Mean Square Error
R_n	Surface net Radiation
RS	Remote Sensing
R_s	Downwards solar radiation
SA	South America
SACZ	South Atlantic Convergence Zone
SEB	Surface Energy Balance
SEBAL	Surface Energy Balance Algorithm for Land
SEBI	Surface Energy Balance Index
SEBS	Surface Energy Balance System
S-SEBI	Simplified Surface Energy Balance Index

SSEBop	Operational Simplified Surface Energy Balance
SULFLUX	South Brazilian Network of Surface Fluxes and Climate Change
T_{air}	Air temperature
T_{dew}	Dew point temperature
TIR	Thermal-infrared radiation
T_s	Surface temperature
TWS	Total Water Storage
UFMT	Federal University of Mato Grosso
UFRGS	Federal University of Rio Grande do Sul - Brazil
u_n	Air wind speed at a height of n meters above the surface
USA	United States of America
USP	University of Sao Paulo
UV	Ultra-violet radiation
VI	Vegetation Index
VIIRS	Visible Infrared Imaging Radiometer Suite
VPD	Vapor Pressure Deficit
WaPOR	Water Productivity through Open access of Remotely sensed derived data
WB	Water Balance
γ	Psychrometric constant
Δ	Slope of the saturation vapor pressure-temperature curve
ΔS	land's water storage variation
λ	Specific latent heat of vaporization

SUMMARY

Chapter 1: General Introduction

1.1 Motivation for this study

Actual evapotranspiration (ET) is the combination of two separate processes of water transference from land to the atmosphere: evaporation from wet surfaces; and transpiration from vegetation. ET links the water, energy, and carbon cycles (ZHANG; KIMBALL; RUNNING, 2016), deeming it essential for water availability, agricultural productivity, and climate change assessments. Despite the inherent difficulty to measure ET, especially at large spatial scales, it is an essential climate variable, and accurate estimates of the spatiotemporal trends and variability in ET are necessary from local to continental scales.

ET represents the major water withdrawal from land, making it a major driver of a region's hydrological and climatological conditions. ET monitoring is used for climatological classification, as well as water availability drought monitoring, which allows for the implementation of strategic adaptation strategies in face of a changing environment, with more frequent and severe water scarcity conditions. ET is also key for the determination of agricultural water use, leading to a more conscient and sustainable food production in a world with increasing demand (FISHER *et al.*, 2017).

Over the past three decades, the focus of ET determination went from local measurements to estimations assisted by satellite remote sensing, due to the need to better assess the spatial variability of ET rather than accurately measure it in a controlled environment. As consequence, various spatially distributed ET datasets are available at global coverage. However, most models rely on empirical and simplified relationships, with coarse spatial resolution, leading to high uncertainties and lack of representativeness over spatially heterogeneous landscapes (JIANG; LIU, 2021; LAIPELT *et al.*, 2021; SINGH *et al.*, 2013). Meanwhile, more physically sound models have been limited to field and regional scales due to computational costs and time constraints, providing incomplete understanding over the land-atmosphere water-energy-carbon exchange over larger ecosystems.

Advancements in cloud computing over the past years have changed this scenario, allowing faster computation and utilization of more complex mathematical models over larger regions, with finer spatial resolution. As a result, regional projects have been launched, providing more reliable ET estimations to the public. Among them, the OpenET project (MELTON *et al.*, 2021) provides high resolution ET estimates over the western United States of America (USA), based on various widely tested models. Initiatives like the OpenET benefit from the connection of researchers from different countries, strengthening scientific research

based on the exchange of experience, of measured data for model validation, and of the best coding practices.

The Surface Energy Balance Algorithm for Land, developed into the Google Earth Engine (geeSEBAL) (LAIPELT *et al.*, 2021) is one of the models used in OpenET project. Because of its physically based formulation, geeSEBAL shows potential for ET representation at heterogenous landscapes, with high accuracy. Given its ingestion into the Google Earth Engine (GEE) platform (GORELICK *et al.*, 2017), it is now possible to use it to generate ET data over wide areas and long time series. In this context, the geeSEBAL is a great candidate for continental and global ET monitoring at moderate resolution and high accuracy. However, geeSEBAL has some limitations, such as sensitivity to the scene context (LONG; SINGH; LI, 2011) and errors in the soil heat flux (G) estimation (LAIPELT *et al.*, 2020), which is a critical component of the surface energy balance equation. Therefore, this study aims to explore the implementation of geeSEBAL at the continental level, with both moderate and high spatial resolutions, and to propose improvements to some of the model's limitations. The results of this study will contribute to the development of a more robust and accurate ET monitoring tool, with potential applications in agriculture, water resources management, and climate modeling.

1.2 Objectives

The objective of this research is to provide ET data at both moderate (500 m pixels) and high (30 m pixels) spatial resolutions for South America, helping understand how this process relates to land cover and climate changes. The scientific papers presented in chapter 3 and 4 and the investigations in chapters 5 and 6 present the following specific objectives:

- Assessment of soil heat flux estimation accuracy using artificial neural networks, focusing on improvements of surface energy balance models;
- Adapt the geeSEBAL model to the large scale, with continental coverage for South America (SA), and assess its accuracy compared to existing datasets. Using this new model version, named geeSEBAL-MODIS, provide an ET time series at moderate spatial resolution for the entire SA;
- Use geeSEBAL-MODIS ET dataset to investigate anomalies and long-term trends of ET in South America, as well as other 10 meteorological and remote sensing variables relevant to the hydrological cycle;
- Adapt the OpenET intercomparison approach for global applications, and apply it to South America, using the geeSEBAL and SSEBop models, as well as data from 36 flux towers.

1.3 Dissertation Structure

This dissertation is comprised of seven chapters. The introductory chapter provides an overview of the topic, while the second one deepens into it, with a review of literature on ET description and its relevance, the history of ET monitoring, and a description of South America, which is the selected study area. Chapter 3 presents the findings of a study published in Remote Sensing in 2021 (https://doi.org/10.3390/rs13122337), which proposes the replacement of empirical soil heat flux equations by models based on artificial neural networks. Chapter 4 focuses on the adaptation of the geeSEBAL model to continental scale estimation and the development of publicly available moderate resolution ET dataset covering SA. The results presented in this chapter composed a second paper that is undergoing peer review at ISPRS Journal of Photogrammetry and Remote Sensing. Chapter 5 covers an investigation of ET and other variables trends in South America. In chapter 6, the focus shifts to the reproduction of the OpenET project approach for the intercomparison of remote sensing and flux tower ET data, in South America. Finally, chapter 7 presents the dissertation's concluding remarks.

Chapter 2: Background

2.1 What is evapotranspiration?

2.1.1 The process of evapotranspiration and its participation in the water, energy and carbon cycles

Actual evapotranspiration (ET) (also known as evapotranspiration), is process of water withdrawal from the land to the atmosphere. It combines the vaporization of water in wet soil, water intercepted by plants and water bodies, with the transpiration from vegetation (Figure 2.1). ET is one of the most important components of the water cycle and, given the large water and energy fluxes is accounts for, it is a driver of regional hydrological and climatological conditions, as well as meteorological conditions at multiple spatial scales (RUHOFF *et al.*, 2012).

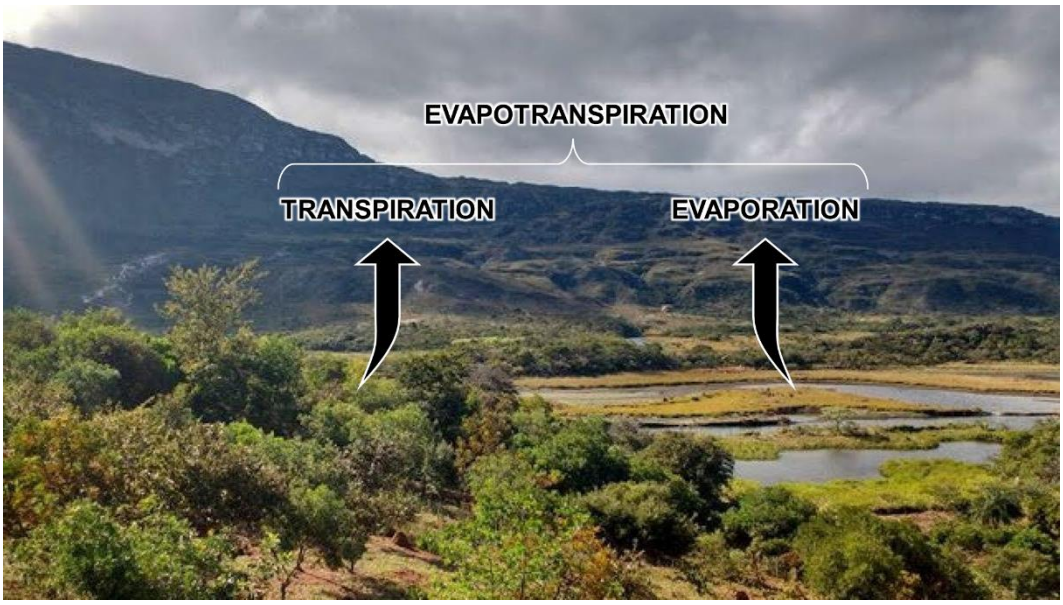


Figure 2.1 – Representation of evapotranspiration as a sum of the evaporation and the transpiration. Source: the author.

ET is a complex process, that connects numerous biogeochemical and biogeophysical processes, such as the water, energy, carbon and nutrients cycles (ALFIERI; KUSTAS; ANDERSON, 2018).

Energy and water availability are the main forcing drivers that affect ET. Direct and diffuse radiation heat the Earth's surface, providing the energy for water in the soil and canopy to vaporize. Given its primary driver position, the surface energy balance (SEB) equation, Equation (1), is the primary boundary condition to consider when calculating ET rates (KUSTAS; NORMAN, 1996):

$$\lambda \times ET = LE = R_n - G - H \quad (1)$$

where λ is the specific latent heat of vaporization; LE is the latent heat flux; R_n is the net radiation; G is the soil heat flux; and H is the sensible heat flux. Figure 2.2 shows a schematic representation of the Earth's energy cycle, in which absorbed shortwave radiation, such as ultra-violet (UV) and visible, is emitted as longwave radiation, both near-infrared (NIR) and thermal-infrared (TIR). LE represents the energy used by water on the surface of the Earth to vaporize into the atmosphere. In the absence of water, all energy taken by the surface is transformed into G and H, with LE = 0. On the other hand, the abundance of water leads to LE = R_n – G, with H = 0 (ALLEN et al., 1998). Since H is the energy used for heating the near surface atmosphere, evapotranspiration plays an important role in regulating the air temperature, so that high ET promotes an "evaporative cooling" effect.

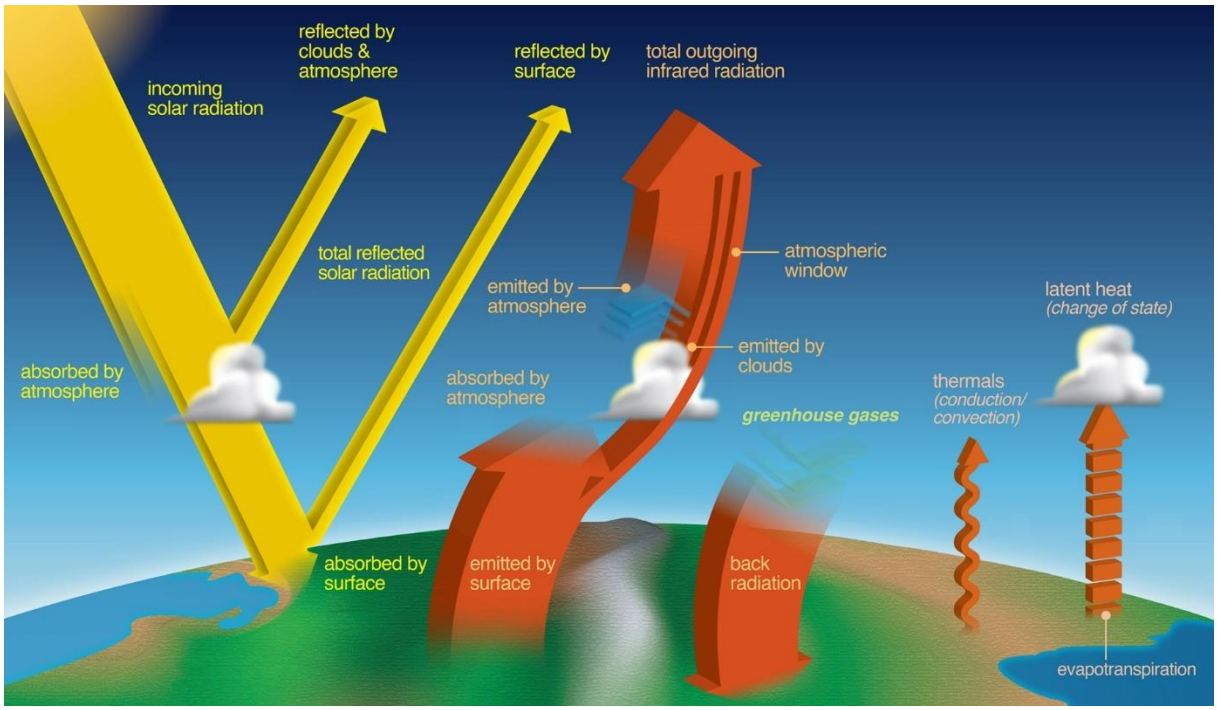


Figure 2.2 – Schematic drawing of the major components of the Earth's energy cycle, which is open due to the exterior energy source and losses to space. Yellow arrows represent shortwave radiation, while orange arrows represent longwave radiation. Source: adapted from Atkinson (2017).

Figure 2.3 presents the Earth's water cycle. Over the Earth's land, ET accounts for between 50% and 80% of precipitation, making the second largest flux of the water cycle (VARGAS GODOY *et al.*, 2021). On land, the water balance can be expressed as the sum of the fluxes into (positive) and from (negative) the land:

$$\Delta S = P - Q - ET \quad (2)$$

where ΔS is the land's water storage variation; P in the precipitation over land; and Q is the discharge.

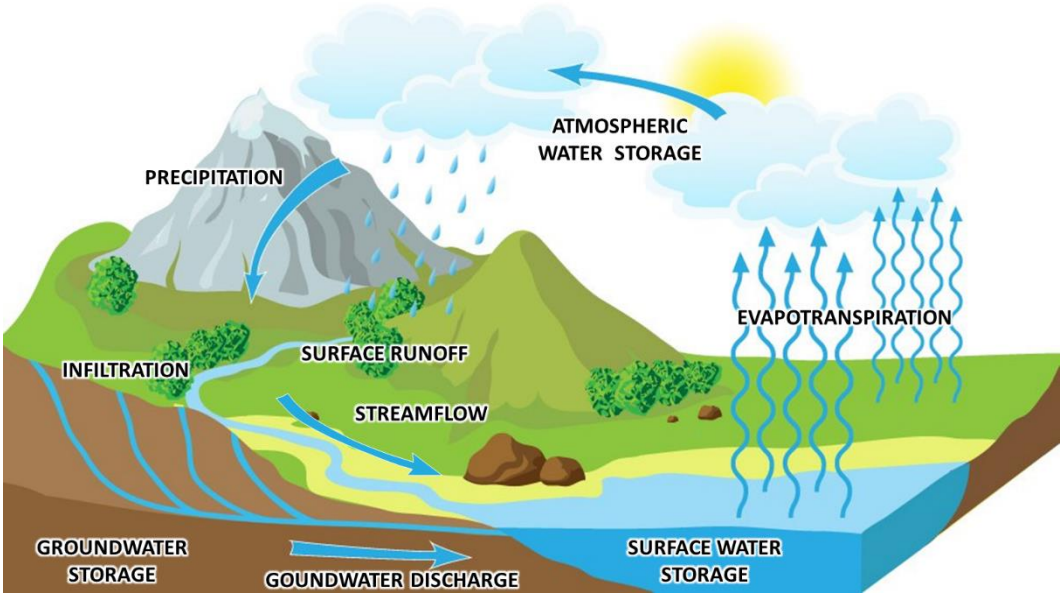


Figure 2.3 – Schematic drawing of the major components of the water cycle. Source: the author.

Another important factor that controls ET is the atmospheric capacity for water uptake, also known as atmospheric demand, or potential evapotranspiration (ET_p). Air temperature, wind speed and air moisture influence the atmosphere power to take up water from the surface. Penman (1948) developed a mathematical model to estimate evaporation rates of water bodies, considering the SEB equation, in association with the mass transport method, that takes into account the aforementioned atmospheric drivers of ET.

For surfaces other than water bodies, vegetation type, age, health and distribution influence the ET rates. On one hand, plant shading limits evaporation, since less energy is available at the soil level (ALLEN et al., 1998). On the other, plant transpiration, controlled by stomatal aperture, can increase ET losses. Monteith (1965) extended Penman's equation by adding surface and aerodynamic resistances, making it adequate for ET estimation over vegetated surfaces (MONTEITH; UNSWORTH, 2013).

Figure 2.4 shows the components of the terrestrial carbon cycle. In plants, carbon is taken in by the leaves and, along with water taken up by the roots, is used in the photosynthetic process to produce energy and promote plant growth. Excess carbon is released by plants via the respiration process. Water availability plays a critical role on the carbon cycle, since soil

moisture controls plant photosynthesis and respiration (MU; ZHAO; RUNNING, 2011a). Plants transpiration accounts for almost all water they consume, with very little remaining in their bodies (ALLEN et al., 1998; HUNSAKER et al., 2003; LAZZARA; RANA, 2010; TASUMI et al., 2005). Therefore, ET monitoring is key for assessing vegetation biomass production and the surface carbon budget, which, in turn, support the study of climate variations and trends.

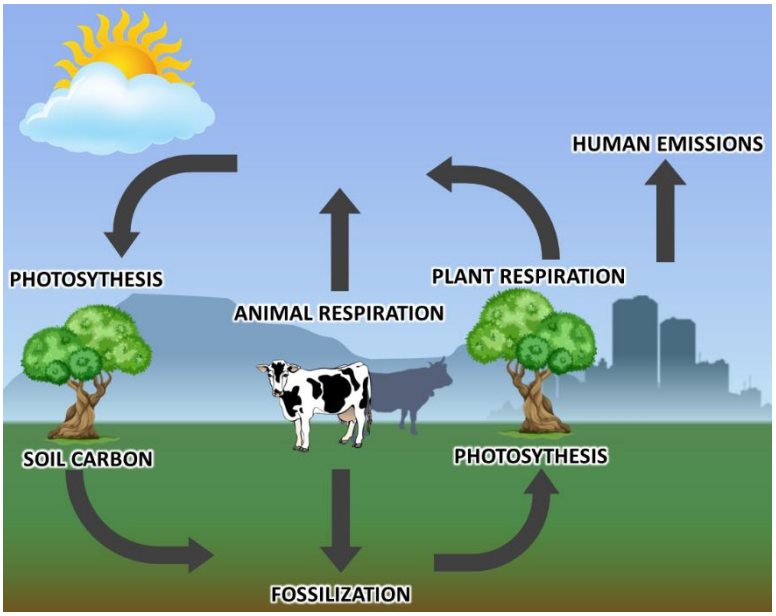


Figure 2.4 – Simplified schematic drawing of the fast carbon cycle's land components. Source: the author.

2.2 Evolution of evapotranspiration monitoring

2.2.1 From local measurements to regional estimations

Historically, ET monitoring has been performed at local scale. The two most used instruments to measure ET are the weighing lysimeter and the micrometeorological tower (also known as flux towers). In the lysimeter, ET is obtained as a residual of the water balance in the soil block. Lysimeters utilization usually limited to short crops, as plant size influences the size of the structure. This way, the weighing lysimeter is not indicated for ET monitoring in forest formations (TEIXEIRA *et al.*, 2009). It is not appropriate for natural vegetation as well, since its area tends to be small and does not represent plant diversity (RANA; KATERJI; DE LORENZI, 2005).

Micrometeorological approaches to ET measuring take two forms, the Bowen Ratio (BR) (BOWEN, 1926) and the Eddy Covariance (EC) (BURBA, 2013; MONTEITH; UNSWORTH, 2013). These methods do not suffer from the same limitations as weighing lysimeters, and flux

towers are used over a variety of land covers and biomes to measures energy, water and carbon fluxes. BR stations uses simultaneous measurements of air temperature and humidity at different heights to calculate vertical gradients, and from those, the BR, which is a relationship between H and LE. Together with R_n and G measurements, ET is calculated at the station. EC stations, on the other hand, makes vector (wind direction and speed) and scalar (pressure, temperature, concentration of water, carbon dioxide and other gases) measurements to determine H and LE separately. Although EC data has been extensively used to validate remote ET estimates over the last decades (FISHER *et al.*, 2017; LAIPELT *et al.*, 2020; MELTON *et al.*, 2021; MORTON *et al.*, 2013; RUHOFF *et al.*, 2013; SENAY *et al.*, 2022a), it often lacks surface energy balance closure. Therefore, an additional verification step should be taken to assess data quality and correct it for SEB closure (TWINE *et al.*, 2000).

Despite the accuracy of such methods, they provide values for specific sites and fail to represent the regional heterogeneity of ET. Other methods try to simplify ET estimation, based on land surface and atmospheric conditions. The FAO Penman-Monteith (PM) method (ALLEN et al., 1998; ASCE-EWRI, 2005) employs the surface energy balance and mass transport approaches to estimate ET for a reference crop surface (ET_r) and combines it with crop (k_c) and soil (k_s) coefficients to estimate actual evapotranspiration. Equations (3) and (4) describe daily ET estimation via the FAO PM method:

$$ET = k_s \times k_c \times ET_r \tag{3}$$

$$ET_r = \frac{0.408\,\Delta(R_{n\,24h} - G_{24h}) + \frac{1600}{T_{air\,24h} + 273.15} \cdot \gamma \cdot u_{2\,24h}(e_{s\,24h} - e_{a\,24h})}{\Delta + \gamma(1 + 0.38\,u_{2\,24h})} \tag{4}$$

where k_s and k_c are water stress and crop coefficients; $R_{n\,24h}$ is the daily net radiation (MJ day^{-1} m^{-2}); G_{24h} is the daily soil heat flux (MJ day^{-1} m^{-2}); $T_{air\,24h}$ is the daily mean air temperature (°C); γ is the psychrometric constant (kPa °C^{-1}); $u_{2\,24h}$ is the daily mean wind speed (m/s) at 2 m height; $e_{s\,24h}$ and $e_{a\,24h}$ are the saturation and actual vapor pressure for the daily time step (kPa); and Δ is the slope of the saturation vapor pressure-temperature curve (kPa °C^{-1}).

ET_r values are generally obtained from atmospheric conditions gathered at meteorological stations. the k_s and k_c coefficients are obtained from information about soil moisture, crop type, and crop age. Given the lower spatial variability of atmospheric variables when compared to ET, estimated ET_r has larger footprint and, with information on crop and soil conditions, it is possible to map ET over cropland areas. However, this method is highly sensitive to k_s and k_c

accuracy and upscaling it to regional can raise uncertainties, due to heterogeneity of climate, terrain, land cover, soil and soil moisture (TUCCI, 2004; DE ANDRADE, 2018).

River basin (or catchment) water balance is another method for ET estimation, that is appropriate for regional applications. Eq. (5) shows a rewritten and simplified form of Eq. (2), in which ET is computed from the water balance of a river basin:

$$\mathrm{ET} = \mathrm{P} - \mathrm{Q} - \Delta\mathrm{S} \tag{5}$$

where P in the precipitation over the catchment area; Q is the discharge; and ΔS is the water storage variation in the basin.

At daily and monthly time-scales, multi-parameter hydrological models are used for water fluxes prediction, with a need for previous calibration (MISHRA; SINGH, 2004; COLLISCHONN et al., 2007; SIQUEIRA et al., 2018). At annual and multi-annual time intervals, ΔS is usually assumed as equal to zero, if no trends or significant anomalies in annual Q are observed (CAVALCANTE *et al.*, 2019), and ET can simply be obtained as the difference between measured values of P and Q. Optionally, ΔS can be retrieved from the Gravity Recovery and Climate Experiment (GRACE) (SWENSON, 2012; SWENSON; WAHR, 2002, 2006a; WEISE, 2015) mission's total water storage (TWS) data (REAGER; FAMIGLIETTI, 2013), with restrictions to the river basin's area (LANDERER; SWENSON, 2012; RODELL *et al.*, 2011). Catchment water balance is a widely used method for ET estimation, but it provides average values for the catchment, or sub-catchments in the case of distributed approaches, and still does not represent the true spatial variability of ET (ZHANG; KIMBALL; RUNNING, 2016). In addition, global improvements in ET measurements are not sufficient for capturing the ET patterns in regions with limited ground data, which are, unfortunately, the ones that suffer the most with climate change and drought (GOSWAMI *et al.*, 2006; IM; PAL; ELTAHIR, 2017; XU; ZHOU; SHIMIZU, 2010).

Despite the limitations of the aforementioned methods for ET monitoring, they are well-established and are still useful, especially for calibration and validation of the new methods being developed (SENAY et al., 2013; SENAY; KAGONE; VELPURI, 2020; BHATTARAI et al., 2019; DE ANDRADE et al., 2021). Over the past decades, growing access to satellite imagery initiated a new phase in ET estimation. As a result, several approaches based on remote sensing emerged, addressing the issue of spatial representativeness, but not without their own problems.

2.2.2 Remote sensing of evapotranspiration

Although satellite remote sensing data has been available for over four decades, it took a while before it started being widely used in hydrology, because of limitations like the unavailability of capable hardware and software for big data processing, lack of knowledge and trust in the application of remote sensing techniques, and reluctance to move away from conventional and well-established methods (SCHULTZ; ENGMAN, 2000). However, constant computational advancements and the multiplication of spaceborne remote sensors have made remote sensing a more attractive approach to modeling hydrological variables, with the advantage of the spatial representativeness of land features at multiple spatial scales, in a globally consistent and cost-effective manner (KUSTAS; NORMAN, 1996; SCHMUGGE *et al.*, 2002).

2.2.2.1 ET models based on remote sensing data

The most commonly used remote sensing (RS) based ET models can be divided into two main approaches: one with the land surface temperature (LST) as the main driver, and another based on vegetation indexes (VI). Table 2.1 lists some of the most used models in each approach. The first group uses LST retrieved from satellite images to estimate SEB fluxes, so they are also known as SEB models. Most of them rely on the selection of boundary conditions, or pixels, on which ET is either null or potential (the same as atmospheric demand), and a fraction of evapotranspiration is computed for each image pixel, based on its LST value in relation to the extremes. These models tend to be limited in regard to large scale application and operationalization, since the selection methods for the boundary conditions can be subjective, making extrapolations for larger areas uncertain (LIOU; KAR, 2014a; SENAY *et al.*, 2013).

VI based models generally use vegetation indexes, such as the leaf area index (LAI) or the normalized difference vegetation index (NDVI), and other surface characteristics, like albedo and land cover classification, to compute the resistance terms of the Penman-Monteith (PM) (MONTEITH, 1965) or of the Priestley-Taylor (PT) (PRIESTLEY; TAYLOR, 1972) equations. The PM equation presents a robust solution to both the SEB and of the mass transport models, while the PT equation represents a simplification of PM, in which the mass transport parcel is replaced by an empirical value, that is a function of the area of interest climate.

Table 2.1 – Brief description of the most commonly used ET models, separated between SEB and physical approaches.

Model	Source	Characteristics
LST approach		
Surface Energy Balance Index (SEBI)	(MENENTI, M; CHOUDHURY, 1993)	FE is based on the difference between LST and the temperature of the atmospheric boundary layer, and two boundary conditions for extreme surface moisture conditions, derived from the SEB equation.
Atmosphere–Land Exchange Inverse (ALEXI)	(ANDERSON *et al.*, 1996)	Utilizes geostationary data to correlate LST sub-daily changes to the evaporative cooling effect. Estimates both plant transpiration and soil evaporation, at a coarse spatial resolution.
Surface Energy Balance Algorithm for Land (SEBAL)	(BASTIAANSSEN *et al.*, 1998a)	Estimates a temperature gradient to estimates SEB fluxes, via an internal calibration method, and with selection of pixels under extreme surface moisture conditions.
Simplified Surface Energy Balance Index (S-SEBI)	(ROERINK; SU; MENENTI, 2000)	Graphical approach to the SEBI model, with FE computation based on the relationship between LST and albedo.
Surface Energy Balance System (SEBS)	(SU, 2002)	FE is based on the solution of the SEB equation at extreme surface moisture conditions for sensible heat flux retrieval, and comparison to local H value.
Atmosphere–Land Exchange Inverse flux disaggregation approach (DisALEXI)	(NORMAN *et al.*, 2003)	Vegetation index and LST data from fine remote sensing data are used to upscale ALEXI's ET estimations.
Mapping Evapotranspiration with Internalized Calibration (METRIC)	(ALLEN; TASUMI; TREZZA, 2007)	Employs the SEBAL methodology for SEB fluxes computation, with correction of FE based on instant ET_r values, and additional correction related to surface topography.
Operational Simplified Surface Energy Balance (SSEBop)	(SENAY *et al.*, 2013)	FE is based on LST and the LST of extreme moisture boundary conditions, derived from vegetation index and the SEB equation.
VI approach		
MOD16	(MU; ZHAO; RUNNING, 2011a)	Uses vegetation index and surface properties, associated to meteorological conditions to estimate ET, based on the Penman-Monteith equation.
Priestley–Taylor Jet Propulsion Laboratory (PT-JPL)	(FISHER; TU; BALDOCCHI, 2008)	Uses vegetation indexes to estimate transpiration and soil evaporation, based on the Priestley-Taylor equation.
Penman-Monteith-Leuning (PML)	(LEUNING *et al.*, 2008)	Uses leaf area index to derive surface conductance, and the Penman-Monteith equation to estimate ET.
Global Land Evaporation Amsterdam Model (GLEAM)	(MIRALLES *et al.*, 2011)	Employs the Priestley-Taylor equation to estimate ET, based on meteorological reanalysis and remote sensing surface data.
Breathing Earth System Simulator (BESS)	(RYU *et al.*, 2011)	Uses the Penman-Monteith equation to estimate soil and canopy ET, as well as gross primary productivity.

Although VI based methods are widely used for regional and global applications, not using LST information makes them inaccurate over areas with high soil evaporation, which is quite common over agricultural areas and natural wetlands (BIGGS; MARSHALL; MESSINA, 2016). Over such areas, SEB models are indicated, for LST is highly sensitive to effects of

evaporative cooling and surface moisture (ANDERSON *et al.*, 2011; MELTON *et al.*, 2021), making them ideal for the assessment of ET variations and anomalies related to water stress (BHATTARAI *et al.*, 2019).

Table 2.2 lists some of the most relevant remote sensing instruments for ET estimation by SEB models. Visible and infrared numeric data provide information about the surface at multiple spatiotemporal scales. Usually, due to energy output differences, the data from a single instrument are distributed with a finer spatial resolution for visible and near infrared bands, and a coarser one for thermal infrared. Among the different instruments, spatial and temporal resolution have opposing patterns, where finer spatial resolution leads to longer revisit times. Conversely, satellites at a higher orbit, and coarser spatial resolution, take less time to reimage the same area.

Table 2.2 – Spaceborne remote sensing instruments commonly used in ET estimation by LST-driven models (SEB models).

Satellite / Sensor	**Spatial Resolution**	**Temporal resolution**	**Activity**
GOES (Geostationary Operational Environmental Satellite)	1, 4 and 8 km	8 times per hour	1978 - now
Landsat 5 TM	30 and 120 m	16 days	1984 - 2013
AVHRR/3 (Advanced Very High Resolution Radiometer)	1090 m	twice a day	1998 - now
Landsat 7 ETM+	15, 30 and 100 m	16 days	1999 - now
MODIS (Moderate Resolution Imaging Spectroradiometer) Terra	250, 500 and 1000 m	Twice a day	1999 - now
ASTER (Advanced Spaceborne Thermal Emission and Reflectance Radiometer)	15, 30 and 90 m	daily	1999 - now
MODIS Aqua	250, 500 and 1000 m	Twice a day	2002 - now
VIIRS (Visible Infrared Imaging Radiometer Suite)	375 and 750 m	daily	2012 - now
Landsat 8	15, 30 and 100 m	16 days	2013 - now
Sentinel-3A	1000 m	daily	2016 - now
ECOSTRESS (ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station)	70 m	1 to 5 days	2018 - now
Landsat 9	15, 30 and 100 m	16 days	2022 - now

2.2.2.2 The SEBAL model

The SEBAL model was developed by Bastiaanssen et al. (1998) and has been applied over a wide range of surfaces and climates. The original model uses physical (R_n, H) and empirical (G) relationships to estimate the components of the SEB equation, driven exclusively

by remote sensing data. One of the model's innovations is the calculation of a near surface temperature gradient (dT), that relies on LST data alone. This removes the need for LST calibration and the dependance on spatially accurate air temperature data (ALLEN; TASUMI; TREZZA, 2007). Overall, the model's sound theoretical foundation and relative simplicity inspired researchers to apply and adapt SEBAL to their needs over the past two decades. It is also the selected model for this research, and a detailed description of its formulation is presented in Chapter 3.

At the time of its creation, SEBAL's independence from meteorological data was ideal for areas with poor monitoring of atmospheric variables, since the satellite images have global coverage. However, the base model does not solve the mass transport and other atmospheric factors that affect ET, which could lead to errors in some scenarios. Allen et al. (2007) and Allen, Tasumi, Trezza (2007) proposed an improvement to SEBAL by including ET_r data into the formulation, thus creating the METRIC model. Like SEBAL, the METRIC model did not propose a solution for the subjectivity of endmembers (hot and cold pixels) selection, that hinders large scale operationalization for both models. (SENAY *et al.*, 2013) proposed a new method for operational ET estimation inspired by SEBAL, the SSEBop model. In SSEBop, the hot and cold pixels selection is replaced by boundary conditions, diminishing the contextual nature of the model and facilitating continental application. Despite the creation of new models after it, SEBAL is still used worldwide, and ranks among the best performing ET models (LAIPELT *et al.*, 2020; LIOU; KAR, 2014a; MELTON *et al.*, 2021; WAGLE *et al.*, 2017).

The development of the GEE platform provided access to satellite data and powerful cloud computing, enabling the incorporation of SEBAL framework within the GEE, renamed to geeSEBAL. The geeSEBAL model is able to provide long ET time series for Landsat imagery, spanning over 30 years. It has been applied to areas in Brazil (LAIPELT *et al.*, 2021) and the USA (MELTON *et al.*, 2021), and has shown its potential as tool for monitoring agricultural water use and effects of other human activities on ET.

The main issue that inhibits SEBAL, and METRIC, large scale application is the contextual nature of the endmember selection process, which is dependent on the area of interest (AOI) size and location, and the imagery spatial resolution (LONG; SINGH; LI, 2011). Bhattarai, Mallick and Jain (2019) describe that this issue leads to uncertainties when scaling the application to larger areas, such as a country or continent. Figure 2.5 illustrates this effect over the India's territory. While the "all India" AOI yields a seamless map of ET, it computes overestimations on colder areas. On the other hand, AOI splitting led to contrasting LE values

over the borders of the agro-ecological regions or grid cells. Additionally, finer grids resulted in errors in the model's computation due to insufficient valid endmembers. Therefore, the large-scale implementation of SEBAL still remains a challenge.

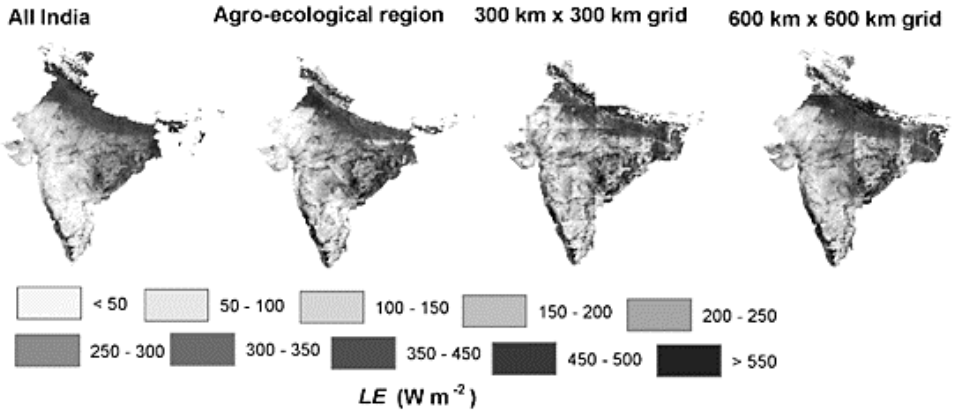


Figure 2.5 – Instantaneous LE maps (March 12, 2005) from the SEBAL model using the four spatial domains. Source: adapted from Bhattarai, Mallick and Jain (2019).

A more in-depth description of the SEBAL model formulation is presented in Chapter 3, along with the solution proposed in this work for large scale implementation.

2.2.2.3 Multi model and multi scale approaches to ET monitoring

Despite the advancements in data availability and computational processing power, mapping ET at high spatiotemporal resolution is still challenging for large areas, mainly for three reasons: the incompatibility between spatial resolution and satellite revisit time; the processing time for SEB models to compute ET, which makes on demand ET mapping unfeasible; and the computer memory consumption of ET maps at continental scales. Therefore, coupling multi-scale ET mapping is a fitting alternative (ANDERSON *et al.*, 2011). In this context, the FAO's (Food and Agriculture Organization) Water Productivity through Open access of Remotely sensed derived data (WaPOR) project (FAO, 2020) provides ET data in Africa at three spatial resolutions, with a coarser one (250 m pixels) covering the whole continent, and a finer one (30 m pixels) covering irrigation areas. The WaPOR data helps monitoring ET at different levels and contexts, from small field irrigation to large river basins storage assessment.

Another challenge to ET monitoring is the large range of ET models, that often yield diverging results, without a consensus on which is better or more suitable for specific locations. One solution to this problem has been the merging of the results from multiple models, generating an ensemble (ANDAM-AKORFUL *et al.*, 2015; BHATTARAI *et al.*, 2019;

RUHOFF *et al.*, 2022; VINUKOLLU *et al.*, 2011). The ensemble approach tends to mitigate potential inaccuracies from individual models, yielding better performance in validation studies. One of the largest efforts to provide an ensemble from different models is the OpenET project (MELTON *et al.*, 2021), that uses six models to provide an ensemble ET time series at high spatial resolution, covering the western USA. The project includes many stakeholders and collaborators from the USA, led by the Desert Research Institute (DRI), the Environmental Defense Fund (EDF), the National Aeronautics and Space Administration (NASA), and Google. In Brazil, the Federal University of Rio Grande do Sul (UFRGS) is also involved in the project, implementing the geeSEBAL model. The easily accessible ET data from the OpenET project helps monitor water use and, as a consequence develop better water management practices.

Projects like the WaPOR and OpenET indicate a new phase in ET modeling, a phase that benefits from big geodata availability, increasing computational power and expanding collaboration between research teams around the globe.

2.2.3 Insights on the future of ET monitoring

The incorporation of remote sensing data into ET modeling has improved our understanding of this and other hydrological process from field to continental scale, at multiple temporal resolutions. SEB-based models have the advantage of representing the surface characteristics more accurately, because of the close relationship between LST and surface moisture. However, their dependence on cloud free images restricts their usage in some areas and time periods, which is a major drawback, and can lead to inconsistences in ET seasonality of larger areas, when compared to ET estimated from river basin water balance (FASSONI-ANDRADE *et al.*, 2021; RUHOFF *et al.*, 2012). SEB-based models also present divergent results, making the selection of a single one a difficult task. Finally, the contextual nature and expensive computational demand of some models hinder their application at an operational level (BHATTARAI *et al.*, 2017; BHATTARAI; MALLICK; JAIN, 2019; LONG; SINGH; LI, 2011; WAGLE *et al.*, 2017). Nonetheless, SEB-based models have been used extensively over the past years and show the potential to solve several of the main outstanding knowledge gaps for the ET-based science. To do so, improvements in ET estimation must aim for higher accuracy, higher spatiotemporal resolution, larger spatial coverage, lower latency, and long-term monitoring (FISHER *et al.*, 2017).

Since LST plays a major role in ET prediction, acting as a surrogate for surface moisture (ANDERSON *et al.*, 2012), recent missions have consistently incorporated thermal infrared

sensors aboard their satellites. That is the case for the Visible Infrared Imaging Radiometer Suite (VIIRS), Landsat 9, and ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) missions, which provide thermal infrared images with spatiotemporal resolutions of 1,000 m/1 day, 100 m/16 days, and 70 m/1-5 days, respectively (CAO *et al.*, 2013; FISHER *et al.*, 2020a; MASEK *et al.*, 2020). These missions enable the improvement and continuity of LST (and, as consequence, also ET) monitoring for the years to come.

The ever-growing range of satellite imagery for ET modeling creates a new challenge, which is accessing and using petabytes of data. Cloud storage and computing has been presented as a solution for managing "big data". Platforms such as the GEE democratize powerful processing worldwide, making both large amounts of data and powerful computer processing accessible to anyone. In this context, the use and innovation of ET models is booming. Novel techniques are also starting to be widely used, such as machine-learning models (CHEN; ADAMS, 2006; DE ANDRADE *et al.*, 2021a; KÄFER *et al.*, 2020; SARI; DOS REIS CASTRO; PEDROLLO, 2017), that require little to no prior knowledge of the physical processes underlying the observed variables, enabling the discovery of unforeseen relationships as new observations are made (LARY *et al.*, 2016; MCCABE *et al.*, 2017).

2.3 South America

South America (SA) represents 12% of the Earth's land area, and contributes with around 30% of total runoff into the oceans (CLARK *et al.*, 2015), earning the nickname of "Continent of Water" (STEVAUX *et al.*, 2009). Distributed across a wide latitudinal range (55º S to 12º N), SA presents a variety of climates, seasonally influenced by the Inter Tropical Convergence Zone (ITCZ), near the Equator Line; and the South Atlantic Convergence Zone (SACZ), formed from the Amazon towards the South of Brazil and the Atlantic Ocean (MARENGO *et al.*, 2012; VERA *et al.*, 2006). While the ITCZ is a worldwide phenom derived from the Earth's rotation and incident solar radiation difference, the SACZ is a particular effect to SA, emanating from both the ITCZ during the southern hemisphere summer and the presence of the Andes Mountain Range, that acts as an obstacle to clouds and moisture formed over the Atlantic Ocean and the Amazon Forest, which get redirected towards the center of SA (GARREAUD *et al.*, 2009). At the annual time scale, climate in SA is recurrently influenced by the *El Niño* Southern Oscillation (ENSO), in which the variations in the Pacific and Atlantic oceans' surface

temperature propagate to changes in atmospheric conditions, often leading to diverging extreme events across SA (CAI *et al.*, 2020; MARENGO *et al.*, 2016).

These factors are strong drivers of the SA climates and biomes, presented in Figure 2.6. In the northern land portion of the continent, humid winds of the ITCZ are responsible for intense rainfall with gauges reaching up to 13,000 mm year^{-1}, in western Colombia (POVEDA; MESA, 2000). In this region is located the Amazon Forest, the largest biome in SA and the largest tropical forest in the world. Rainfall in the Amazon decreases gradually from Northwest to the South range on average from 3,000 to 1,500 mm year^{-1} (ESPINOZA VILLAR *et al.*, 2009). Southeast of the Amazon Forest, is the Brazilian Savanna, formed by the Cerrado and Caatinga biomes. This region receives rainfall between 600 and 2,000 mm year^{-1}, presenting a well-defined seasonality with dry winters and wet summers, typically influenced by the SAMS and the SACZ (MARENGO *et al.*, 2012, 2022). Southwest of the Cerrado, the Pantanal and Chaco regions are home to seasonally flooded and dry forests, respectively, despite the relatively lower annual precipitation (GARCÍA *et al.*, 2017; REBOITA *et al.*, 2010). In the southern portion of the continent, the semi-arid climate of Patagonia contrasts with the wetter regions like the Pampa. The Andes Mountain Range acts as a barrier for moisture from the Pacific Ocean to reach the eastern portions of the continent (GARREAUD; ACEITUNO, 2001), presenting a gradient of over 4,000 mm year^{-1} in parts of the west coast, to under 200 mm year^{-1}, in northern Chile and western Argentina (PARUELO; JOBBÁGY, 2007).

Over the past decades, great efforts have been employed on monitoring water cycle processes throughout the South American continent. For ET, stations based on BR and EC methods have been installed on several landcover types and climates, as a result of cooperation of several institutions from Brazil and other countries. Figure 2.7 (A) shows some of the micrometeorological towers installed in South America over the past two decades. Flux towers were installed over croplands (CABRAL *et al.*, 2003; CARNEIRO *et al.*, 2013; DIAZ *et al.*, 2019; MOREIRA *et al.*, 2013; SAKAI *et al.*, 2004; SOUZA *et al.*, 2019; UCKER TIMM *et al.*, 2009; ZIMMER *et al.*, 2016), grasslands (BIUDES *et al.*, 2015; HASLER; AVISSAR, 2007; OLIVEIRA *et al.*, 2006; RUBERT *et al.*, 2016; VON RANDOW *et al.*, 2012), forests (ARAÚJO, 2002; BIUDES *et al.*, 2015; BORMA *et al.*, 2009; DALMAGRO *et al.*, 2019; GOULDEN *et al.*, 2004; SALESKA *et al.*, 2003; ULKE; GATTINONI; POSSE, 2015; VIHERMAA *et al.*, 2016; VON RANDOW *et al.*, 2012), wetlands (GRIFFIS *et al.*, 2020; HOLL *et al.*, 2019), and savanna formations (BIUDES *et al.*, 2015; BORGES *et al.*, 2020; DA ROCHA *et al.*, 2009; SANTOS *et al.*, 2003). However, given the cost and technical difficulties

of sustaining a flux tower, most sites present a series of under four years, with several gaps and measurement errors, that must go through a treatment phase before being used for ET monitoring.

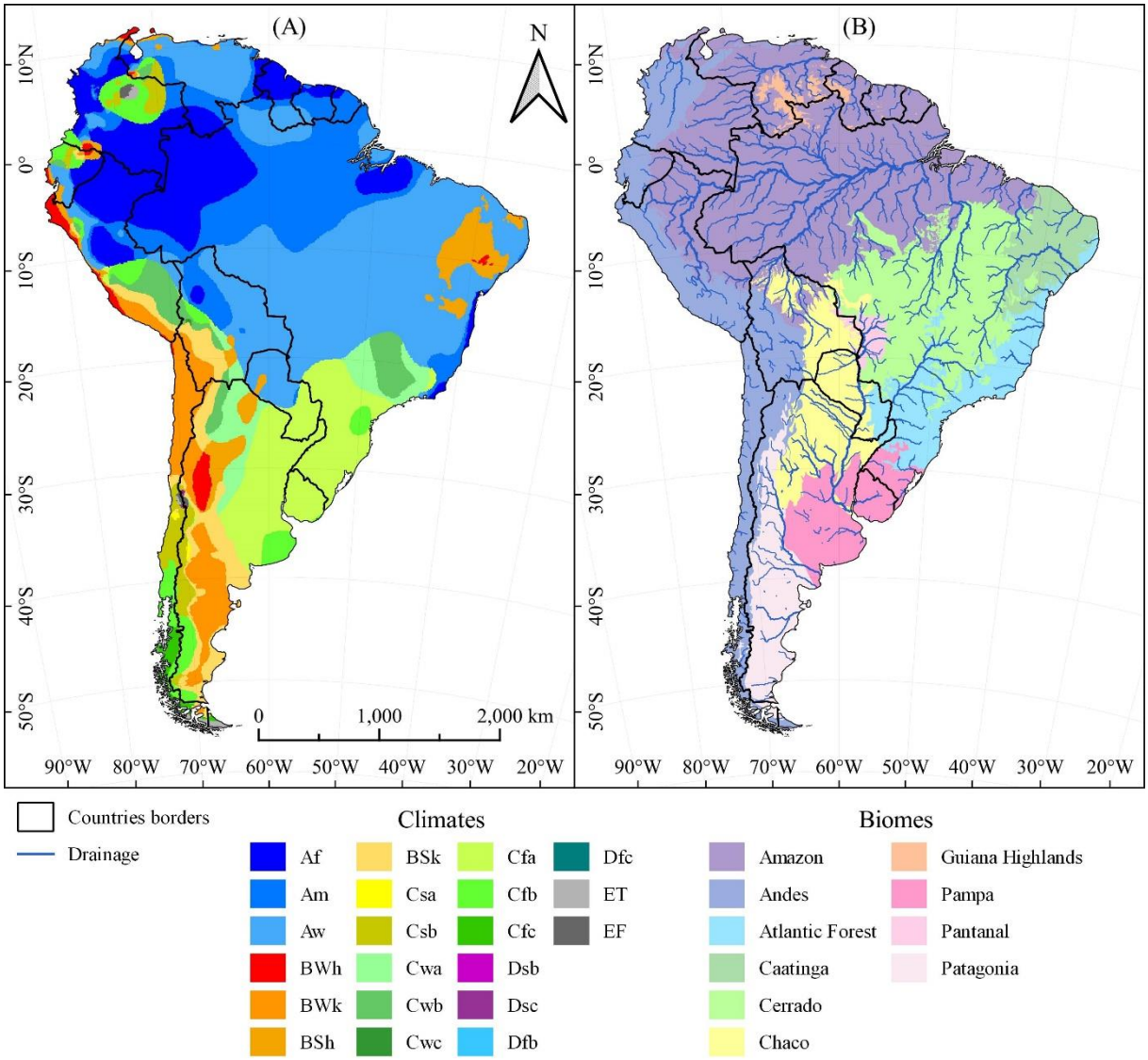


Figure 2.6 – (A) Climates distribution in South America (PEEL; FINLAYSON; MCMAHON, 2007), denoted as follows: tropical rainforest (Af), tropical monsoon (Am), tropical wet savanna (Aw), arid hot (BWh), arid cold (BWk), semi-arid hot (BSh), semi-arid cold (BSk), temperate dry hot summer (Csa), temperate dry and warm summer (Csb), temperate dry and hot summer (Cwa), temperate dry winter and warm summer (Cwb), temperate dry winter and cold summer (Cwc), temperate without dry season and hot summer (Cfa), temperate without dry season and warm summer (Cfb), temperate without dry season and cold summer (Cfc), cold climate with dry and warm summer (Dsb), cold dry summer (Dsc), cold without dry season and warm summer (Dfb), cold without dry season and cold summer (Dfc), polar tundra (ET), and polar frost (EF). (B) South American biomes from the Terrestrial Ecoregions of the World (OLSON *et al.*, 2001; TURCHETTO-ZOLET *et al.*, 2013).

At a larger scale, ET has been consistently estimated from river basin water balance for several decades, based on rainfall and river discharge data, connected through a wide array of

mathematical models that predict intermediary processes, such as ET and water storage variations. One of the most laborious steps of applying rainfall-discharge models is to collect the input data, particularly the discharge one, that may have several sources and inconsistencies. To overcome this obstacle, Ruhoff et al., (2022) prepared and distributed streamflow measurements for approximately fifty points in SA, corresponding to several of the largest river basins in the continent, at a monthly time scale. The data is available at https://metadados.snirh.gov.br/geonetwork/srv/eng/catalog.search#/metadata/ebcdf49c-706e-421f-aa70-bf866a0ab94e . Figure 2.7(B) displays the stations location and the respective drainage area of each one, corresponding to seven major river basins in SA. Despite the efforts on collecting and sharing measured data for ET monitoring in South America, strategic regions are still lacking attention, such as the Andes Mountains, the Argentinian Patagonia, and the Orinoco region, Northwest of the Amazon Forest.

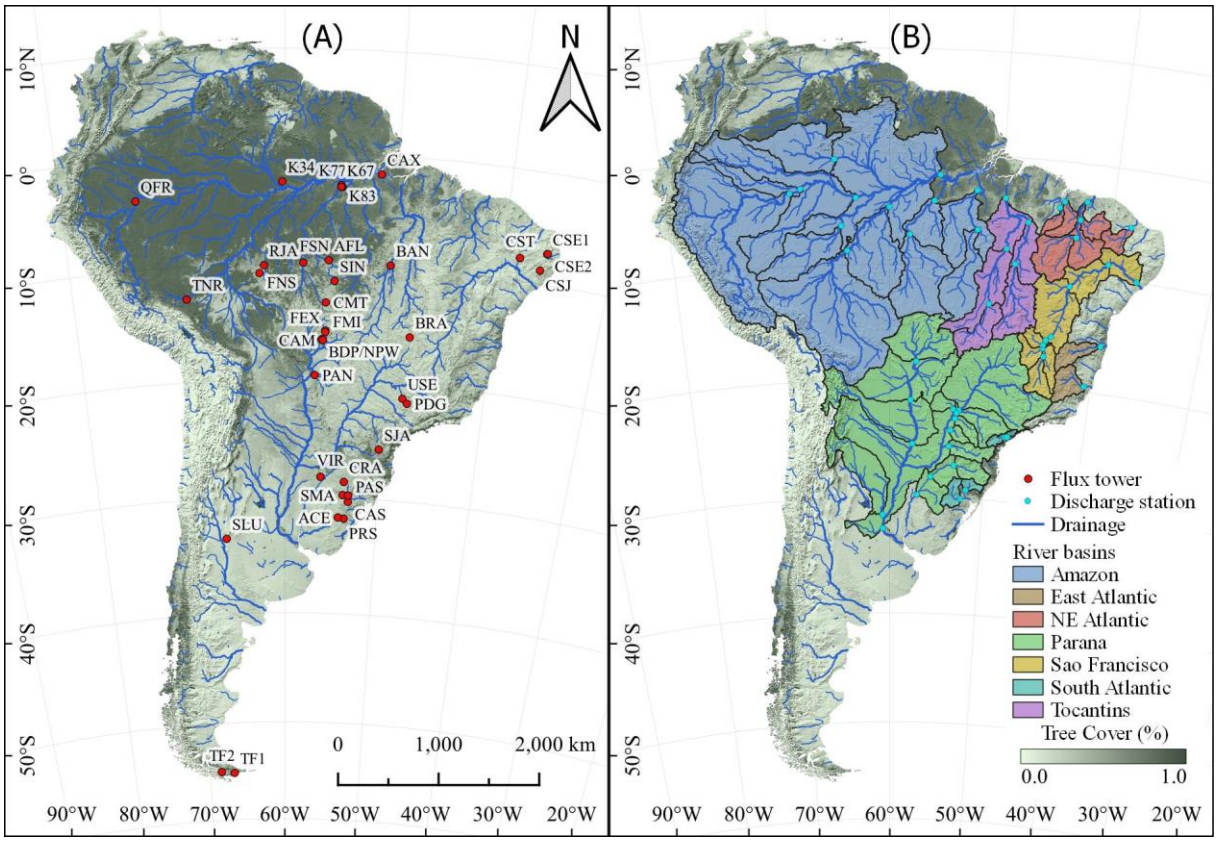


Figure 2.7 – (A) Flux towers distribution in South America. (B) river discharge stations and their respective drainage areas on seven major river basins in South America. Source: the author.

Chapter 3: Artificial Neural Network Model of Soil Heat Flux over Multiple Land Covers in South America

Bruno César Comini de Andrade [1, *]

Olavo Correa Pedrollo [1]

Anderson Ruhoff [1]

Adriana Aparecida Moreira [1]

Leonardo Laipelt [1]

Rafael Bloedow Kayser [1]

Marcelo Sacardi Biudes [2]

Carlos Antonio Costa dos Santos [3]

Debora Regina Roberti [4]

Nadja Gomes Machado [5]

Higo Jose Dalmagro [6]

Antonio Celso Dantas Antonino [7]

José Romualdo de Sousa Lima [8]

Eduardo Soares de Souza [9]

Rodolfo Souza [10]

[1] Institute of Hydraulic Research, Universidade Federal do Rio Grande do Sul, 91501-970, Porto Alegre, Brazil; olavo.pedrollo@ufrgs.br (O.C.P.); anderson.ruhoff@ufrgs.br (A.R.); adriana.moreira@ufrgs.br (A.A.M.); leonardo.laipelt@ufrgs.br (L.L.); rafael.kayser@ufrgs.br (R.B.K.);
[2] Physics Institute, Universidade Federal do Mato Grosso, 78060-900, Cuiabá, Brazil; marcelo@fisica.ufmt.br
[3] Department of Atmospheric Sciences, Universidade Federal de Campina Grande, 58429-900, Campina Grande, Brazil; carlos.santos@ufcg.edu.br
[4] Physics Institute, Universidade Federal de Santa Maria, 97105-900, Santa Maria, Brazil; debora@ufsm.br
[5] Instituto Federal de Mato Grosso, 78060-900, Cuiabá, Brazil; nadja.machado@blv.ifmt.edu.br
[6] Postgraduate Program in Environmental Science, Universidade de Cuiabá, 78005-300, Cuiabá, Brazil; higo.dalmagro@kroton.com.br
[7] Department of Nuclear Energy, Universidade Federal de Pernambuco, 50740-540, Recife, Brazil; antonio.antonino@ufpe.br
[8] Universidade Federal do Agreste de Pernambuco, 55292-278, Garanhuns, Brazil; romualdo.lima@ufape.edu.br
[9] Academic Unit of Serra Talhada, Universidade Federal Rural de Pernambuco, 56900-000, Serra Talhada, Brazil; eduardo.ssouza@ufrpe.br
[10] Department of Biological & Agricultural Engineering, Texas A&M University, College Station, TX 77843, USA; Rodolfo.souza@tamu.edu
***** Correspondence: cominideandrade@gmail.com

Abstract

Soil heat flux (G) is an important component for the closure of the surface energy balance (SEB) and the estimation of evapotranspiration (ET) by remote sensing algorithms. Over the

last decades, efforts have been focused on parameterizing empirical models for G prediction, based on biophysical parameters estimated by remote sensing. However, due to the existing models' empirical nature and the restricted conditions in which they were developed, using these models in large-scale applications may lead to significant errors. Thus, the objective of this study was to assess the ability of the artificial neural network (ANN) to predict mid-morning G using extensive remote sensing and meteorological reanalysis data over a broad range of climates and land covers in South America. Surface temperature (T_s), albedo (α), and enhanced vegetation index (EVI), obtained from a moderate resolution imaging spectroradiometer (MODIS), and net radiation (R_n) from the global land data assimilation system 2.1 (GLDAS 2.1) product, were used as inputs. The ANN's predictions were validated against measurements obtained by 23 flux towers over multiple land cover types in South America, and their performance was compared to that of existing and commonly used models. The Jackson *et al.* (1987) and Bastiaanssen (1995) G prediction models were calibrated using the flux tower data for quadratic errors minimization. The ANN outperformed existing models, with mean absolute error (MAE) reductions of 43% and 36%, respectively. Additionally, the inclusion of land cover information as an input in the ANN reduced MAE by 22%. This study indicates that the ANN's structure is more suited for large-scale G prediction than existing models, which can potentially refine SEB fluxes and ET estimates in South America.
Keywords: remote sensing; surface energy balance; deep learning; MODIS; GLDAS 2.1; flux towers

3.1 Introduction

Soil heat flux (G) is one of the main components of the surface energy balance (SEB) and accounts for the energy transferred to and from the land surface and deeper layers of the ground (KUSTAS; NORMAN, 1999). The SEB determines the energy and mass exchanges of the soil-plant-atmosphere continuum, surface temperature (Ts), the amount of energy available for evaporation and, ultimately, SEB defines local climate conditions and water availability (TRENBERTH; FASULLO; KIEHL, 2009). Most energy that is transferred into the soil during sunlight hours is transferred back at night. Thus, G in vegetated surfaces is generally the smallest of the SEB fluxes and has been neglected in daily considerations of the SEB with some success (SENAY *et al.*, 2013, 2017). However, significant energy transfers occur both during the day and night, and G's exclusion from instantaneous SEB estimations can lead to substantial

errors, e.g., mid-morning G (usually between 9 am and 12 pm) can be as large as sensible heat flux (H) for some land covers (HEUSINKVELD *et al.*, 2004; SAUER; HORTON, 2015).

Several remote sensing products provide valuable knowledge at different spatiotemporal scales to assess land surface and meteorological conditions. Motivated by this availability, many models were developed to calculate SEB fluxes from the integration of remote sensing data, mainly focused on actual evapotranspiration (ET) estimation (ALLEN; TASUMI; TREZZA, 2007; ANDERSON *et al.*, 1996, 2011; BASTIAANSSEN *et al.*, 1998a, 1998b; KALMA; JUPP, 1990; MENENTI, M.; CHOUDHURY, 1993; NORMAN; KUSTAS; HUMES, 1995; ROERINK; SU; MENENTI, 2000; SENAY *et al.*, 2007; SENAY; KAGONE; VELPURI, 2020; SU, 2002). Among them, the SEBAL (surface energy balance algorithm for land) is one of the most used models around the world (BASTIAANSSEN *et al.*, 2010, 1998a, 1998b, 2005; RUHOFF *et al.*, 2012; SINGH *et al.*, 2008; ZHANG *et al.*, 2010), with relatively accurate predictions of ET (BHATTARAI *et al.*, 2016; LIOU; KAR, 2014b; WAGLE *et al.*, 2017).

However, the empirical nature of the developed G prediction models usually implicates that their application to conditions other than the ones for which they were developed (e.g., land cover, climate) may result in large uncertainties (PURDY *et al.*, 2016), suggesting a need for local calibration for accurate estimations (A. IRMAK *et al.*, 2011; CAMMALLERI; LA LOGGIA; MALTESE, 2009). Moreover, errors in G can profoundly decrease the accuracy of SEB models (A. IRMAK *et al.*, 2011; DHUNGEL; BARBER, 2018; KUSTAS; NORMAN, 1999; LAIPELT *et al.*, 2020; RUHOFF *et al.*, 2012; RUSSELL *et al.*, 2015). To address this limitation, a local calibration can be applied (DANELICHEN *et al.*, 2014; SINGH *et al.*, 2008), but this may be impractical for operational purposes, such as SEB mapping over large areas. Comparative studies revealed a wide range in models' performance, especially across different land covers (DANELICHEN *et al.*, 2015; PURDY *et al.*, 2016), and seasonal and regional biases have been identified, indicating potential issues with model comparisons at the local scale (PURDY *et al.*, 2016).

To overcome the issue of regional biases when performing a local scale model calibration/validation, we built a network of flux towers from regional initiatives to represent the wide-ranging climate and ecosystem diversity of South America (VILLARREAL; VARGAS, 2021). In addition to the most used flux tower network data compilation (SALESKA *et al.*, 2013) from the large-scale biosphere-atmosphere experiment in the Amazon (LBA-ECO) (DAVIDSON; ARTAXO, 2004), we also included data from the SULFLUX (South Brazilian

network of surface fluxes and climate change), ONDACBC (National Observatory of Water and Carbon Dynamics in the Caatinga Biome) (BORGES *et al.*, 2020), and the PELD Pantanal (long-term ecological research in Pantanal) (TABARELLI *et al.*, 2013), in conjunction with flux measurements supported by Brazilian universities, including the UFMT (Federal University of Mato Grosso) (BIUDES *et al.*, 2015) and the USP (University of Sao Paulo) (DA ROCHA *et al.*, 2009), funded by national and regional research agencies. This network encompasses most land covers in South America, corresponding to over 80% of its area (except for urban, barren, and high altitude forest land covers) (EVA *et al.*, 2004), and has been used as validation data for SEB flux estimations via remote sensing (DANELICHEN *et al.*, 2014; LAIPELT *et al.*, 2020; RUHOFF *et al.*, 2012).

An alternative to the existing methods of modeling G is the use of artificial neural networks (ANNs), which are universal approximators (HORNIK; STINCHCOMBE; WHITE, 1989) and have been successfully tested (used or applied) in several hydrological processes and modeling attempts and have shown better performance than existing conceptual and empirical models (CANELÓN; CHÁVEZ, 2011; HORNIK; STINCHCOMBE; WHITE, 1989; JIMENO-SÁEZ *et al.*, 2018; KÄFER *et al.*, 2020; SILVERMAN; DRACUP, 2000; TABARI; SABZIPARVAR; AHMADI, 2011; ZANETTI *et al.*, 2007). ANNs are computational models analogous to the brain's biological behavior, simulating its capabilities of learning and memorizing. The ANN relates input data to specified outputs through a series of intertwined layers of neurons. The neurons connect and transform input variables via activation functions. An ANN can be trained by pairing historical data and calibrating the connections' synaptic weights to obtain the best relationship between inputs and outputs (HORNIK; STINCHCOMBE; WHITE, 1989).

One of the main disadvantages of ANNs is that their structure does not describe the underlying processes that explain the target variable, thus being labeled as black-box models (ADADI; BERRADA, 2018). On the other hand, the empirical origin of the existing models for instantaneous G prediction also presents the same limitation. With this in mind, ANNs become an attractive methodology for G prediction. In fact, previous applications of ANNs for G prediction at the local scale (CANELÓN; CHÁVEZ, 2011; JIMENO-SÁEZ *et al.*, 2018) have yielded promising results. Considering the availability of the long time series of measured G, representative of the South American landscape (EVA *et al.*, 2004; VILLARREAL; VARGAS, 2021), and the readily available remote sensing data, the usage of the ANN approach for large-scale applications is encouraged as a next step to assess its efficiency in predicting G.

Therefore, given the diversity of G estimation methods and their limited application conditions, this paper focuses on assessing the ability of ANNs to predict G over a wide range of ecosystems in South America, using long-term remote sensing and meteorological time series, and comparing the ANN's performance to that of commonly used G models. With this study, we hope to assist future SEB closure studies and ET estimates with more accurate estimates of G and consequently H and LE.

3.2 Materials and Methods

3.2.1 Study Sites

In this study, data from 23 flux towers were used to calibrate and validate the ANNs, located in South America. The area surrounding each flux tower was considered one study site. Table 3.1 lists the flux towers surrounding land covers and the data acquisition periods, while Figure 3.1 displays the location and land cover of the study sites. For simplification and repeatability, the land covers were grouped into one of five types: forest, irrigated cropland, cropland, grassland, and savanna.

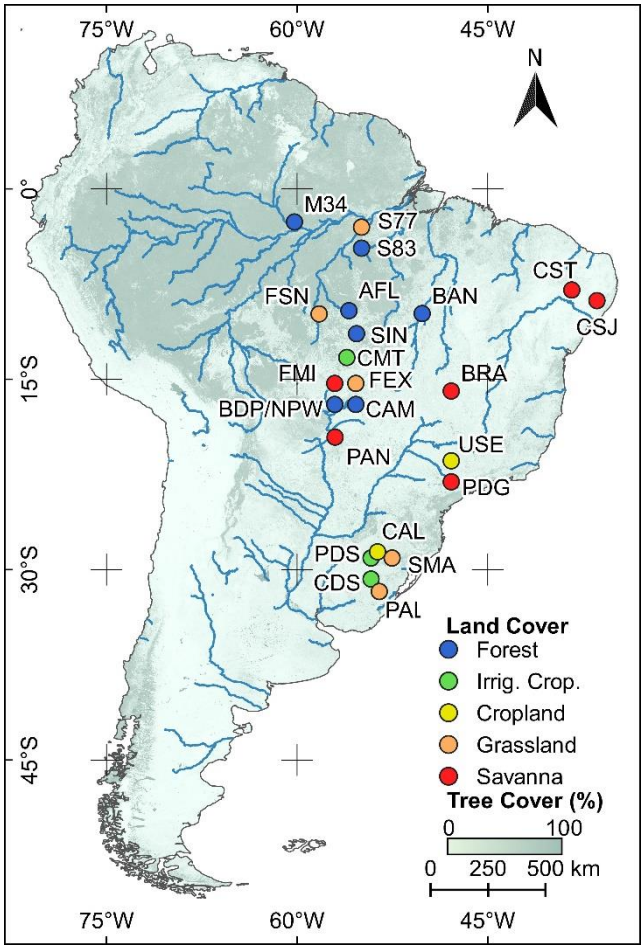


Figure 3.1 – Flux tower locations and main land cover types used in this research.

Table 3.1 – Description of the flux towers used in this study.

Name	Land Cover	Main Land Cover	Available Data Period	Source
AFL	Rainforest	Forest	February 2003 – March 2004	(BIUDES *et al.*, 2015)
BAN	Seasonally flooded forest	Forest	October 2003 – December 2006	(BORMA *et al.*, 2009)
BDP	Seasonally flooded forest	Forest	June 2011 – August 2015	(BIUDES *et al.*, 2015)
BRA	Cerrado	Savanna	January 2011 – December 2011	(SANTOS *et al.*, 2003)
CAL	Cropland (soy)	Cropland	January 2009 – September 2014	(AGUIAR, 2011; MOREIRA, 2012)
CAM	Seasonally flooded forest	Forest	January 2007 – January 2009	(TABARELLI *et al.*, 2013)
CDS	Cropland (flooded rice)	Irrigated cropland	January 2011 – December 2014	(CARNEIRO, 2012; SOUZA *et al.*, 2016)
CMT	Irrigated cropland	Irrigated cropland	September 2015 – February 2017	(LATHUILLIÈRE *et al.*, 2018)
CSJ	Caatinga	Savanna	April 2019 – December 2019	(BRITO *et al.*, 2020)
CST	Caatinga	Savanna	February 2014 – March 2020	(SILVA *et al.*, 2017; SOUZA *et al.*, 2016)
FEX	Pasture	Grassland	September 2006 – March 2010	(BIUDES *et al.*, 2015)
FMI	Cerrado	Savanna	April 2009 – May2013	(BIUDES *et al.*, 2015)
FSN	Pasture	Grassland	March 2002 – July 2003	(HASLER; AVISSAR, 2007)
M34	Tropical forest	Forest	January 2000 – September 2006	(ARAÚJO, 2002)
NPW	Seasonally flooded forest	Forest	January 2015 – July 2017	(DALMAGRO *et al.*, 2019)
PAL	Grassland	Grassland	September 2013 – September 2016	(RUBERT *et al.*, 2016)
PAN	Cerrado	Savanna	January 2001 – December 2006	(OLIVEIRA *et al.*, 2006)
PDG	Cerrado	Savanna	January 2001 – December 2006	(ROCHA *et al.*, 2002)
PDS	Cropland (flooded rice)	Irrigated cropland	October 2003 – March 2004	(TIMM *et al.*, 2014; ZIMMER *et al.*, 2016)
SIN	Transitional forest (Cerradão)	Forest	May2005 – October 2008	(BIUDES *et al.*, 2015)
S77	Pasture / agriculture	Grassland	August 2000 – November 2005	(SAKAI *et al.*, 2004)
S83	Tropical forest	Forest	June 2000 – March 2004	(GOULDEN *et al.*, 2004)
SMA	Grassland	Grassland	January 2013 – December 2016	(RUBERT *et al.*, 2016)
USE	Cropland (sugarcane)	Cropland	January 2001 – December 2002	(CABRAL *et al.*, 2003)

3.2.2 Input Data

Table 3.2 shows the input data used in the development of the ANN. Surface temperature (T_s), surface albedo (α), and enhanced vegetation index (EVI) data were obtained from a moderate resolution imaging spectroradiometer (MODIS) product at daily time intervals. Given that the training process of ANN requires a large amount of data, the MODIS products were selected due to their high temporal resolution. MODIS's image acquisition (on-board of the

Terra satellite) occurs at around 10:30 AM (local time). The net radiation (R_n) dataset was acquired from the global land data assimilation system version 2 (GLDAS 2.1) reanalysis product at a 3 hour time intervals, averaged for the 9 AM to 12 PM window. Both satellite and reanalysis products were available from the Google Earth Engine platform (https://earthengine.google.com/ [Accessed date: June 30, 2020]).

Table 3.2 – Input data description.

Data	Product	Spatiotemporal Resolution	Source
Net radiation (R_n)	GLDAS 2.1	0.25°/3 h	(RODELL *et al.*, 2004b)
Land surface temperature (T_s)	MOD11A2.V6	1000 m/1 day	(WAN, 2014)
Albedo (α)	MCD43A3.V6	500 m/1 day	(SCHAAF; WANG, 2015)
Enhanced vegetation index (EVI)	MOD13Q1.V6	250 m/16 days	(DIDAN, 2015a)

Flux tower data were obtained at hourly and sub-hourly time intervals and were also resampled and averaged to the 9 AM to 12 PM local time window. A search was executed on each flux tower dataset to remove incomplete (gap-filling) data. Rn values 10% lower than the five days moving average were discarded to remove cloudy sky data. Finally, G values beyond the range of -50% to 50% of Rn were also removed from the analysis (IDSO; AASE; JACKSON, 1975; KUSTAS; DAUGHTRY, 1990; KUSTAS; DAUGHTRY; VAN OEVELEN, 1993). After this filtering procedure, a total of 3,778 samples remained available.

Figure 3.2 displays the correlation plot between the observed G from the flux towers and the R_n, T_s, EVI, and α datasets. The correlation plots and r values are distinguished by land cover. Correlation among the input datasets is generally weak, with $|r| < 0.50$ for most land covers, where |r| is the absolute value of r. Exceptions ($|r| > 0.50$) are mainly found for T_s and R_n on most land covers. The correlation of observed G to input datasets is also low in general. The correlation plots are useful to identify potentially redundant inputs of the ANN. Due the low correlation among the variables, it can be assumed that they may contribute independently to the ANN predictions.

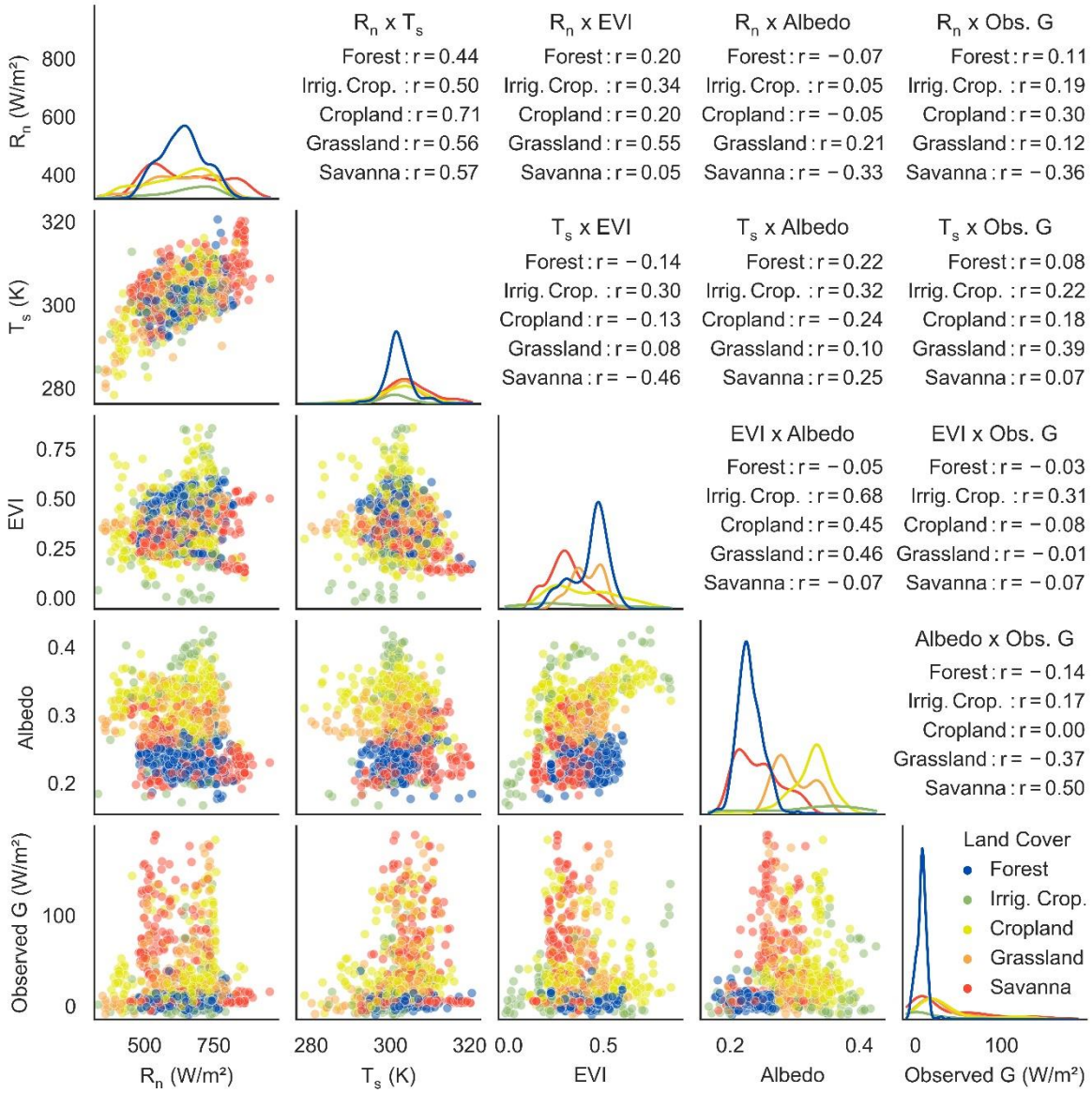


Figure 3.2 – Correlation plot and r values of the observed G and the input datasets for all flux towers, distinguished by land cover.

3.2.3 ANN's Structure

Figure 3.3 displays a generic scheme of the ANNs that were built. The input data are combined into a series of processing units in the hidden layer, called neurons. These "hidden" neurons are then connected to an output neuron that converts the hidden layer's information into the desired output. In order to transform the data, each neuron passes the received information through an activation function, such as the bipolar and unipolar sigmoid functions. The connections between neurons are established by synaptic weights (or simply weights), and these, together with the activation functions, determine the relationship between the input data and the corresponding outputs. The weight values are defined during the training process. The

number of neurons in the hidden layer (m) determines the ANN complexity and is assessed based on a search process for a network with the least complexity that still obtains the same performance, with a validation sample, as an initial network, that is purposefully over-sized (LUCCHESE; DE OLIVEIRA; PEDROLLO, 2020). The ANNs were developed with the most common inputs used in models for G computation (CANELÓN; CHÁVEZ, 2011; JIMENO-SÁEZ *et al.*, 2018; SILVERMAN; DRACUP, 2000; TABARI; SABZIPARVAR; AHMADI, 2011; ZANETTI *et al.*, 2007). This allows for the assessment of the benefits of using ANNs compared to preexisting models.

The study sites were grouped into different land cover types to consider characteristics that are difficult to capture from remote sensing data, such as subcanopy vegetation density, vegetation height, and soil skin temperature. Land cover data were inputted as numerical data in the ANNs, as follows: 1 for cropland; 2 for irrigated cropland; 3 for forest; 4 for grassland; and 5 for savanna. The impact of the inclusion of land cover data as an input was assessed using a comparison of the ANNs with and without the land cover information. Thus, two networks were generated. From this point on, the two networks will be referred to as: (i) Gc: an ANN that considers land cover as an input; and (ii) Gnc: an ANN that does not consider land cover as an input.

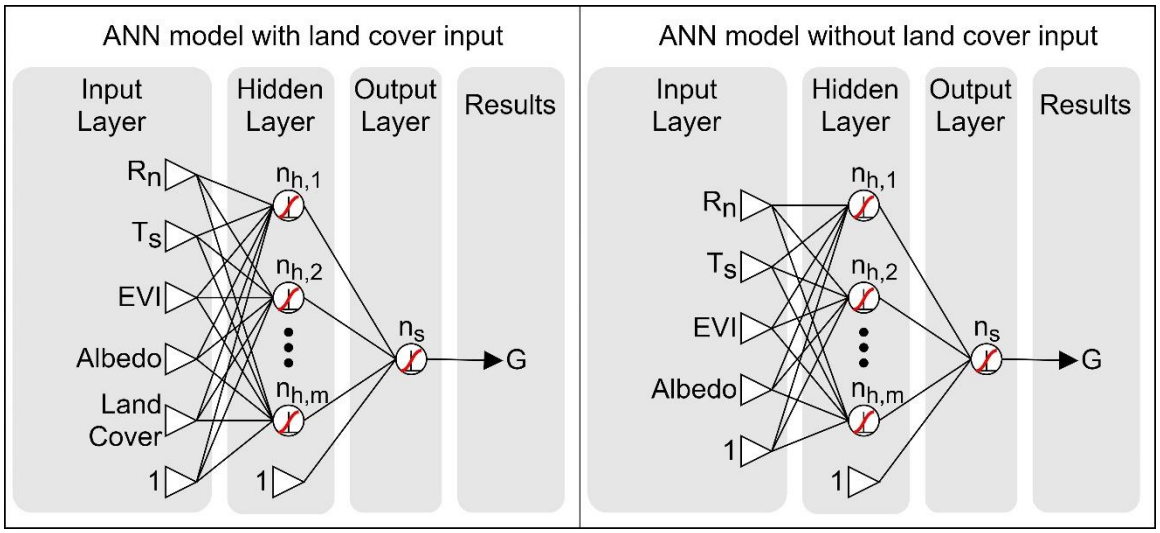


Figure 3.3 – Structure of the two ANNs used for this study. One includes the site's land cover as an input (Gc) and another is without land cover information (Gnc).

In order to avoid preferential treatment of inputs with different magnitudes, input and output data were scaled using a normalization procedure, shown by Equation (1):

$$X' = \frac{(X - X_{bottom})}{X_{top} - X_{bottom}} \quad (1)$$

where X' is the scaled variable; X is the raw variable; and X_{bottom} and X_{top} are the data bottom and top limits, respectively. The top and bottom limits were chosen based on observations of the 23 study sites' available series and are shown in Table 3.3. These values also represent the valid range for applications of the trained ANNs.

Table 3.3 – Data limits used for the scaling procedures presented in Equation (1).

Data	Bottom Limit	Top Limit
Soil heat flux (G)	-20.0 W/m^2	200.0 W/m^2
Net radiation (R_n)	150.0 W/m^2	1000.0 W/m^2
Land surface temperature (T_s)	275.0 K	325.0 K
Albedo (α)	0.1	0.5
Enhanced vegetation index (EVI)	-0.1	1.0
Land cover	0.0	6.0

The developed ANNs are generically represented by Equation (2):

$$O' = f_o\left\{\sum_{i=1}^{m} w_{o,i} f_h\left[\sum_{j=1}^{n}\left(w_{h,i,j} X'_j\right) + b_{h,i}\right] + b_o\right\} \quad (2)$$

where O' is the ANN's output, equivalent to G in its scaled form; n is the number of input datasets; m is the number of neurons in the hidden layer; wh, bh, fh, wo, bo, and fo are the synaptic weights (w), biases (b), and activation functions (f) of the hidden (h) and output (o) layers, respectively.

The unipolar sigmoid function was used as the activation function for both the hidden and output layers.

3.2.4 ANN's Training Process

The ANN's training consisted of adjusting of the synaptic weights and biases by minimization of the output error using the backpropagation algorithm (RUMELHART; HINTON; WILLIAMS, 1986). w and b values were updated using the method presented by WIDROW and HOFF (1989) and the learning rate was calculated according to the training acceleration techniques proposed by Vogl *et al.* (1988).

The cross-validation method was used to control overfitting. This method consists of dividing the datasets into three series of similar sizes: a training series, a validation series, and a verification series. Weight and bias updates occur to reduce the total quadratic error in the training dataset (S_t) with each iteration, but the quadratic error in the validation series (S_v) is

also observed. The final values of w and b are the ones that provide the minimal Sv value. The ANN performance is then assessed for the verification series.

The number of iterations (or cycles) must be sufficient to identify the best iteration (minimal error). If the best cycle is too close to the end of the network training, it is possible, and probable, that a better solution could be found if there were more iterations. Thus, a method for increasing the total number of cycles was applied to the ANN training, as follows:

- A starting minimum of 200,000 cycles was defined;
- If the best iteration occurred before the 100,000th cycle, the training was considered complete. If it happened after, the total number of cycles was doubled, up to 400,000 cycles.

This second step was repeated either until the best cycle was found in the first half of the training process or until a limit of five duplications was reached, reaching up to 6.4 million iterations. This limit was applied to reduce computational running time and to avoid infinite loops.

In order to assess the ANN's complexity, the training process was repeated with the progressive reduction in the number of neurons in the hidden layer, from a maximum of 12 neurons, until a decrease in its generalization capacity was observed, measured by the quadratic error in the validation series. The smallest number of neurons that did not yield a decrease in the network's generalization capacity was then chosen.

After assessing the ANN's complexity, the data series division was evaluated, as recommended by (CROWTHER; COX, 2005). One-third of the total data was reserved for the verification series. The remaining data were split into calibration and validation series. The tested split ratio range was larger than that recommended by (CROWTHER; COX, 2005), because it was noticed that better predictions of G happened for smaller training datasets. Therefore, a broader range of split ratios needed to be tested. A sequence of tests between 10/90 and 90/10 was executed to determine the optimal calibration/validation split ratio. The quadratic error present in the validation series was used to determine the optimal data division.

Due to the ANN's empirical nature and the randomness of the initial conditions, training runs could prematurely stop at a local minimum on the objective function surface. To avoid this, after each randomization and partitioning of the data series, the cross-validation process was repeated ten times. Since data randomization was performed 40 times, ANN training was repeated in parallel 400 times for each of the two networks created. The trained ANNs with the

best performance in the validation series were then used to predict G in the verification data series.

3.2.5 Performance Assessment and Input Data Contribution Analysis

The ANN's predictions were assessed against data measured in the flux towers via a set of statistical indexes, calculated for the validation data series of each of the 40 data randomizations. The Nash–Sutcliffe coefficient (NS) (NASH; SUTCLIFFE, 1970) value in the validation dataset was used as the criterion for selecting the optimal ANN.

For comparison purposes, G was also estimated by the original and adjusted Jackson *et al.* (1987) (JACKSON *et al.*, 1987) and Bastiaanssen (1995) (BASTIAANSSEN, 1995) equations. Reparameterization was performed using the minimum quadratic error as the goal function. Table 3.4 presents the originals and adjusted equations (for results, see Section 3.2., Models' performance evaluation), where Ts is the surface temperature (°C); α is the surface albedo; and NDVI is the normalized difference vegetation index.

Table 3.4 – Original Jackson *et al.* (1987) and Bastiaanssen (1995) equations and adjusted equations calibrated for minimum quadratic error.

Original Equations	**Adjusted Equations**
Jackson *et al.* (1987):	Jackson *et al.* (1987):
$G = 0.583 R_n \exp(-2.13 \cdot NDVI)$	$G = 0.150 R_n \exp(-0.89 \cdot NDVI) - 34.3$
Bastiaanssen (1995):	Bastiaanssen (1995):
$G = T_s R_n (0.0038 + 0.0074\alpha)(1 - 0.98 NDVI^4)$	$G = T_s R_n (-0.0019 + 0.0106\alpha)(1 - 1.10 NDVI^4) + 10.4$

The performances of the Gc and Gnc ANNs, and of the Jackson *et al.* (1987) and Bastiaanssen (1995) models, were evaluated via the comparison of NS, mean absolute error (MAE), root-mean-square error (RMSE) (WILLMOTT; MATSUURA, 2005), bias (BOYLAN; RUSSELL, 2006), the slope coefficient (A_0) (SEN, 1968), and the linear correlation coefficient (r) (STATISTICAL METHODS IN THE ATMOSPHERIC SCIENCES, 2019), calculated for the verification series. The MAE, RMSE, and bias metrics were calculated as absolute values and as a percentage of the mean of the observed verification data series.

The input data sets contribution to the ANNs was calculated by the connection weights method proposed by (GARSON, 1991) and modified by (GOH, 1995). In this method the relative importance of the input variable is entirely derived from the synaptic weights of the ANN post-training.

3.3 Results

3.3.1 ANN Training Process

The complexity analysis of the ANNs indicated that its performance improves initially with higher complexity but is stable for complexities greater than four neurons. Therefore, for prudency, seven neurons for Gc and five neurons for Gnc were chosen.

After assessing the ANN's complexity, the optimal dataset training/validation split ratio was verified, resulting in 748 samples for the training series, 1745 samples for the validation series, and 1285 samples for the verification series, for both the Gc and the Gnc ANNs. The synaptic weights of the trained ANNs, as well as the relative importance of each input variable to each ANN, are available as Supplementary Information (Tables SI.1-SI.3).

3.3.2 Model Performance Evaluation

Performance metrics of the Gc and Gnc ANNs, and the Jackson *et al.* (1987) and Bastiaanssen (1995) original and adjusted models, are displayed in Table 3.5. The adjusted versions of the Jackson *et al.* (1987) and Bastiaanssen (1995) models yielded significantly better performance metrics in relation to the original versions. Given this difference in performance, the original versions were not further assessed. The ANNs performed better than the adjusted models, with lower values of MAE and RMSE, and higher NS, A_0, and r values. The Gc ANN yielded the best NS, MAE, RMSE, A_0, and r values.

Table 3.5 – Performance metrics of the Gc and Gnc ANNs, and of the Jackson *et al.* (1987) and Bastiaanssen (1995) original and adjusted models.

Metric	Gc	Gnc	Jackson *et al.* (1987)		Bastiaanssen (1995)	
			Original	Adjusted	Original	Adjusted
NS	0.53	0.27	−23.11	−0.17	−7.16	0.02
MAE (W/m^2)	14.04	18.06	145.66	24.43	84.69	21.78
MAE (%)	56%	70%	603%	101%	351%	90%
RMSE (W/m^2)	23.95	28.90	159.15	35.00	92.61	32.00
RMSE (%)	95%	113%	659%	145%	384%	133%
Bias (W/m^2)	−1.87	−0.82	145.36	1.07	81.73	0.99
Bias (%)	−7%	−3%	602%	4%	339%	4%
A_0	0.52	0.30	0.08	0.01	0.12	0.08
r	0.73	0.53	0.04	0.02	0.12	0.21

3.3.3 Model Performance by Land Cover

Figure 3.4 compares the seasonal G in each land cover, given by the flux tower observations and by the Gc and Gnc ANNs, and the Jackson *et al.* (1987) and Bastiaanssen

(1995) adjusted models. While forest land cover displays a relatively constant observed G throughout the year, other land covers present a broader seasonal range and larger monthly standard deviations (given by shaded areas). Overall, the Gc ANN yielded the best adherence to the observed G series, showing similar seasonal patterns and standard deviations, especially in the savanna land cover. On the other hand, the Gnc ANN and the adjusted models yielded a smaller seasonal range and standard deviation, not adhering to the observed G series as well as the Gc ANN.

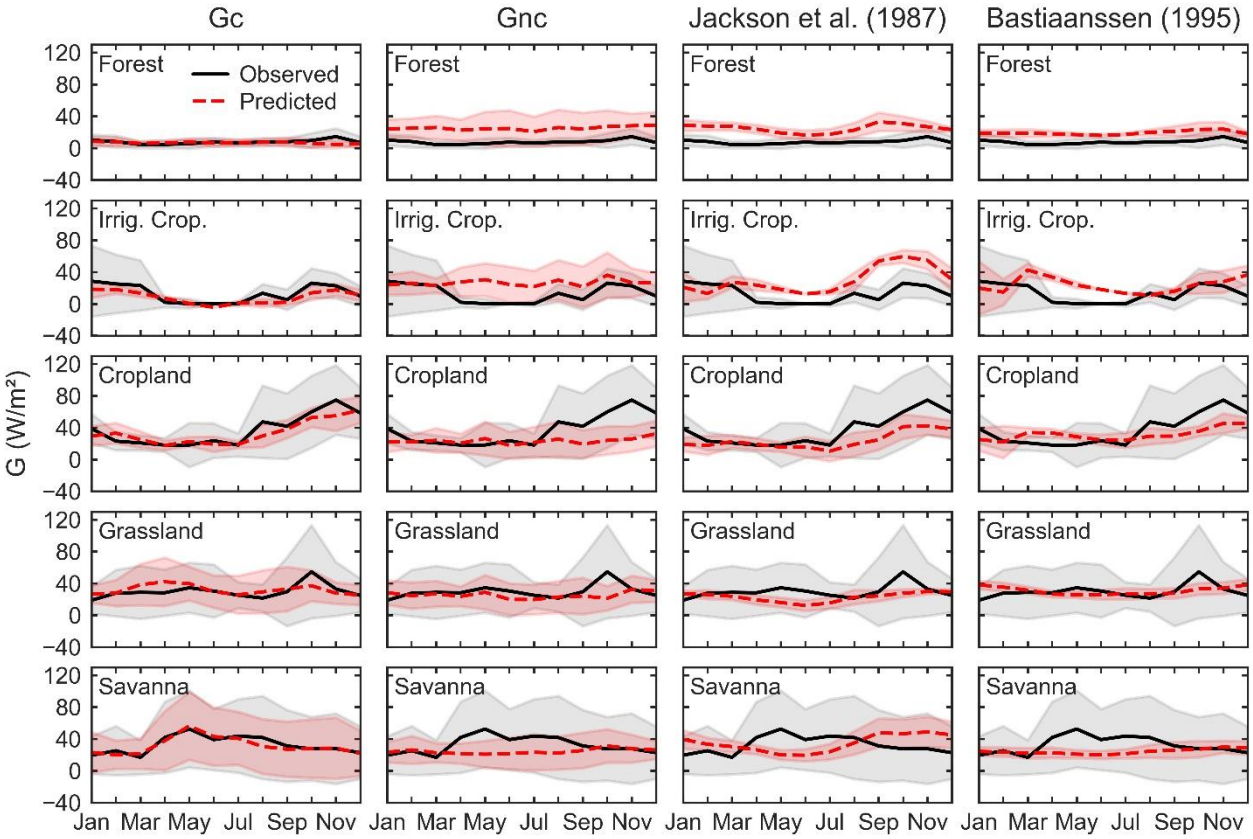


Figure 3.4 – Observed (black, continuous) and predicted (red, dashed) seasonal monthly values of G, for each land cover. Shaded areas represent each series monthly standard deviations, while the lines represent the monthly averages.

Figure 3.5 presents the correlation plots of observed G in the flux towers to the predicted G by the Gc and Gnc ANNs, and by the Jackson *et al.* (1987) and Bastiaanssen (1995) adjusted models, distinguished by land cover. The Gc ANN presents the closest adherence to the 1:1 relationship line (black line in each plot), as the other methods underestimate higher values of G. H. A higher correlation was observed for the Gc ANN over savanna (r = 0.86), followed by grassland (r = 0.49). In contrast, the lowest correlation values were found over forest (r = 0.22) and irrigated cropland (r = 0.36). The Gnc model yielded a similar correlation for all land covers, ranging from r = 0.48 over savanna to r = 0.68 over irrigated cropland. The adjusted

Jackson *et al.* (1987) best and worst performances were observed for grassland (r = 0.30) and savanna (r = −0.28), respectively. The Bastiaanssen (1995) adjusted model performed better over savanna (r = 0.38) and cropland (r = 0.29), and worse over irrigated cropland (r = −0.38) and forest (r = −0.13).

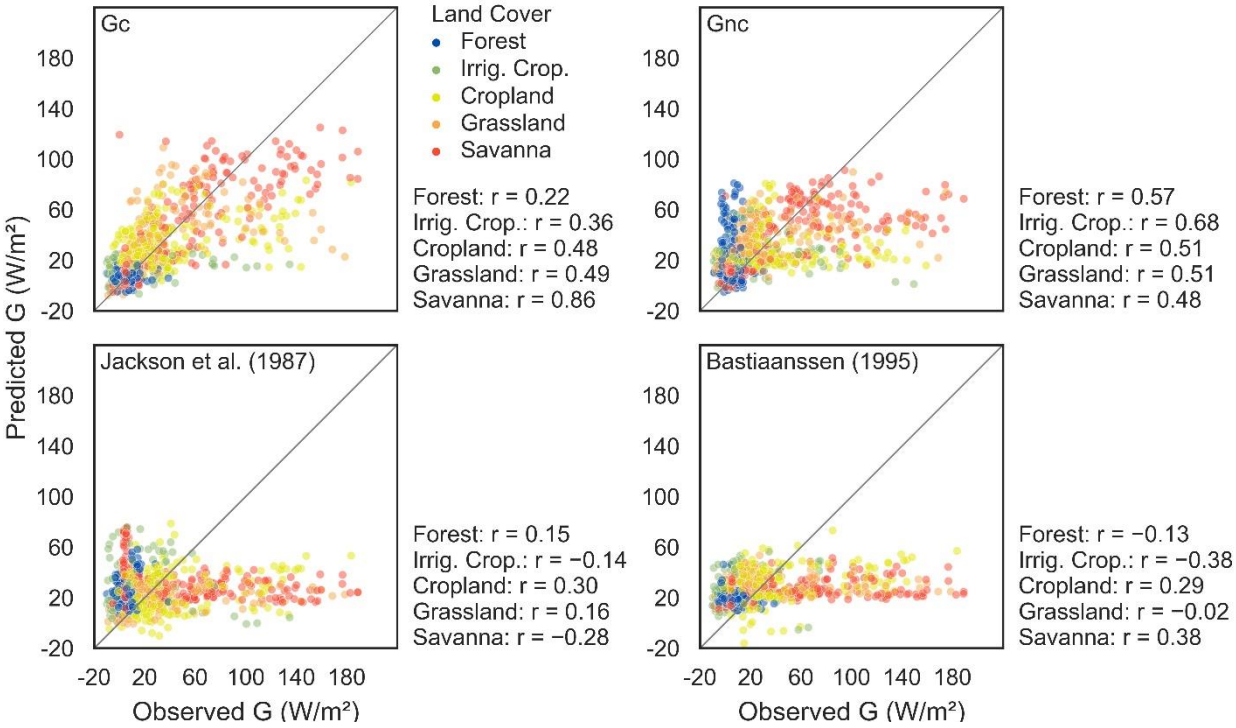


Figure 3.5 – Correlation plots and linear correlation coefficient (r) values of observed G (horizontal axis) to predictions (vertical axis) by Gc and Gnc ANNs, as well as adjusted Jackson *et al.* (1987) and adjusted Bastiaanssen (1995) models, distinguished by color for each land cover. The diagonal black line represents the perfect 1:1 correlation between observed and predicted G.

Figure 3.6 displays other performance metrics of the neural networks and adjusted models for each land cover. G predictions over forest yielded generally low NS and A_0 values, except for Gnc (NS = 0.70). On the other hand, errors (MAE, RMSE, and bias) were also lower for forest in comparison to other land covers. Except Gc bias (−0.19 W/m^2), the positive bias yielded by Gnc and the adjusted models indicates an overestimation of G for forests by models that do not consider land cover as an input. The Gnc ANN yielded the lowest errors of all models. The models and ANNs performed similarly over irrigated cropland, cropland, and grassland, being slightly better for grassland and worse for irrigated cropland. NS and A_0 values were low, ranging between −0.89 and 0.41 and between −0.23 and 0.35, respectively. The MAE ranged between 19.08 and 33.41 W/m^2; RMSE ranged from 27.67 to 42.47 W/m^2; and bias between −13.60 and 14.10 W/m^2. Over savanna, the performance of the ANNs and adjusted models differ the most. Gc predictions yielded NS = 0.72, while the others ranged between

−0.33 and 0.01. Errors (MAE and RMSE) of the Gc and Gnc ANNs were nearly 50% lower than errors from the Jackson *et al.* (1987) and Bastiaanssen (1995) adjusted models. A_0 values were better for the ANNs as well, being 0.66 for Gc and 0.41 for Gnc, while the adjusted models ranged between −0.10 and 0.06.

Some contradictions can be identified between Figure 3.5 and Figures 3.4 and 3.6. Although higher correlation values are yielded by Gnc in comparison to the Gc ANN, Gc presents systematic biases of seasonal G values, as well as higher errors (MAE and RMSE) and lower values of NS and A_0. This discrepancy indicates that the correlation coefficient alone can be misleading as an accuracy metric.

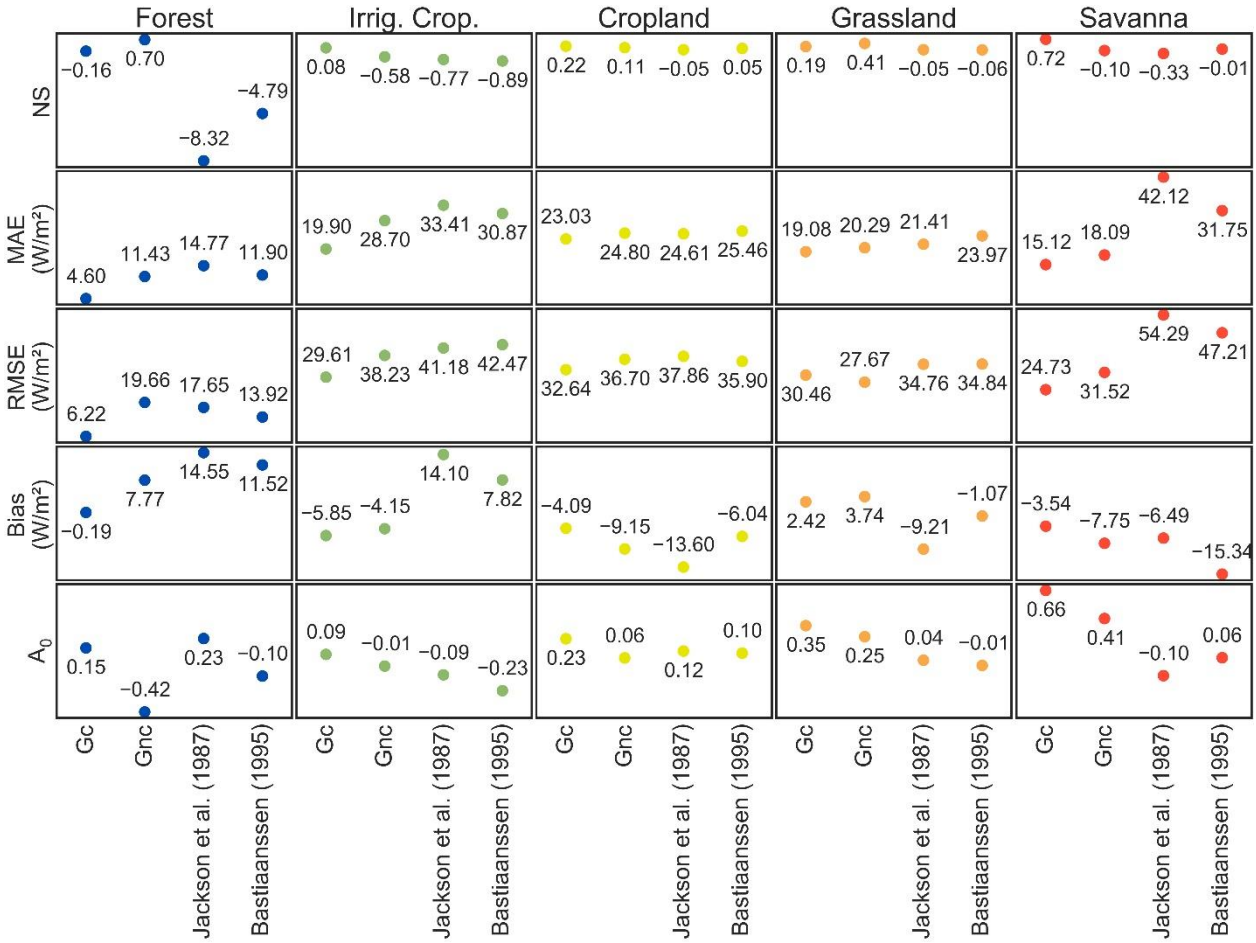


Figure 3.6 – Other performance metrics (NS, MAE, RMSE, bias, and A_0) of the Gc, Gnc, adjusted Jackson *et al.* (1987), and adjusted Bastiaanssen (1995), distinguished by land cover.

3.4 Discussion

Uncertainties in the estimation of SEB fluxes from remote sensing data have motivated the use of ANNs as an alternative to existing models (FENG *et al.*, 2020; HAMED ALEMOHAMMAD *et al.*, 2017; WALLS *et al.*, 2020), but reference material regarding the

validation of G estimates is still scarce, compared to other SEB fluxes. Table 3.6 presents the performance metrics of the G predictions via remote sensing presented by some recent studies. These works used either airborne (ANDERSON *et al.*, 2008; CANELÓN; CHÁVEZ, 2011; SU, 2002) or orbiting (CAMMALLERI; LA LOGGIA; MALTESE, 2009; DANELICHEN *et al.*, 2014; KÄFER *et al.*, 2020; LAIPELT *et al.*, 2020; RUHOFF *et al.*, 2012; TEIXEIRA *et al.*, 2009) platforms associated with meteorological data to estimate G. Validation occurred via comparison to flux tower data in one or more land covers. In most studies, the number of samples was low, ranging between 12 and 698. An exception is the work performed by Purdy *et al.* (2016), with over 100,000 samples. The lowest MAE and RMSE values were found by Su (2002) in cotton crops, while the highest values were obtained by Laipelt *et al.* (2020) over forest. Generally, the highest correlation and lowest error values were obtained with airborne remote sensing data, which have higher spatial resolutions and are used over more homogenous land surfaces. This indicates that using remote sensing data with a finer spatial resolution may better correlate to G measurements.

Table 3.6 – Compilation of other validation studies of G estimation from remote sensing data.

Source	Land Cover	Number of Samples	MAE (W/m^2)	RMSE (W/m^2)	r
(SU, 2002)	Cropland, shrubland, and grassland	620	4.5-41.5	5.4-46.3	0.96-0.97
(RUHOFF *et al.*, 2012)	Savanna and cropland	56	35.0-41.0	-	0.66-0.75
(PURDY *et al.*, 2016)	Various	100,234	-	26.9-42.1	0.32-0.46
(LAIPELT *et al.*, 2020)	Forest	12	108.3-108.9	109.3-110.0	-
(DANELICHEN *et al.*, 2014)	Croplands and forest	131	-	29.2-45.3	0.06-0.76
(CANELÓN; CHÁVEZ, 2011)	Croplands	50	-	-	0.73-0.96
(KÄFER *et al.*, 2020)	Grasslands	67	40.6	43.0	0.70
(ANDERSON *et al.*, 2008)	Grasslands	698	-	23.0-28.0	0.57-0.77
(TEIXEIRA *et al.*, 2009)	Irrigated cropland and savanna	18	-	13.3	0.90

Compared to the metrics presented in Table 3.6, the Gc and Gnc ANNs and the Jackson *et al.* (1987) and Bastiaanssen (1995) models performed within expectations. However, unlike other studies, the models' adjustments and the ANN's training were performed for various land covers and a larger data series. This way, correlation values may be lower, however the resulting models and ANNs provide a higher generalization potential. Overall, the MAE and RMSE values of the ANNs and adjusted models in Table 3.5 reside in the lower portion of the ranges presented in Table 3.6.

As in our study, G predictions of the studies from Table 3.6 showed a lower correlation in densely forested areas. One possible reason for this is the low variance of observed G in

forests. Another possible reason is that G is dependent on the soil skin temperature, while remote sensing products provide canopy surface temperature, which can be uncorrelated in dense vegetation. In fact, Kim *et al.* (2016) verified that, despite the generally high correlation between soil and canopy temperatures, this relationship is weaker in cloud-free images, when radiative heating leads to fast changes in canopy temperature while soil temperature changes slowly. This limitation hinders the improvement of G prediction with orbiting remote sensors on forests. However, the Gc ANN yielded visibly lower errors (MAE, RMSE, and bias), indicating the superiority of this method.

The Bastiaanssen (1995) equation provided better G predictions than the Jackson *et al.* (1987) model, following observations of Danelichen *et al.* (2014) and the assumption by Purdy *et al.* (2016) that models that consider T_s outperform those that do not. When compared to the existing models, both ANNs yielded better performance metrics, with Gc yielding the best predictions and closest seasonal patterns to those observed in the flux towers. Therefore, the introduction of land cover as a numerical input has also proved to be effective.

In this study, the optimal training/validation split ratio contrasts with that recommended by Crowther and Cox (2005) (30/70 against 50/50 to 70/30). It is supposed that this difference is a result of the dataset size. As observed by Crowther and Cox (2005), ANN training with large data sets is less sensitive to partitioning, since there is enough data to support pattern recognition. Thus, a smaller part of the data can be reserved for the training set.

The Gnc ANN yielded better performance than that of the Bastiaanssen (1995) model, despite both models using the same input data. This indicates that this ANN's structure is more suited for large-scale G estimation. In addition, ANNs can incorporate many more inputs, potentially achieving better performance. However, including too many inputs requires expanding the complexity of the network, which causes longer computational times. The low correlation among the considered input variables (R_n, T_s, EVI, and α) provides a low redundancy level to the developed ANNs, and each one is essential to the accurate prediction of G, as also indicated by Canelón *et al.* (2011). The addition of land cover information complemented the input data series and improved performance without significantly impacting computational times. In Figure 3.4, systematic biases were identified for Gnc, Jackson *et al.* (1987), and Bastiaanssen (1995). This indicates that existing models' datasets do not fully comprehend the dynamics of G, further demonstrating the importance of land cover as a key factor in G prediction. Thus, additional datasets should be considered in the future development of models for G prediction. Figure 3.4 shows that the average value and seasonality ranges of

G vary significantly between the different land covers. It also shows that, even though this flux is of small magnitude compared to others of the SEB, it is not negligible. To the best of our knowledge, this work is one of the first large-scale model developments for G prediction in South America, comprehending multiple land covers and climates representative of most South American ecosystems. This work can potentially improve ET estimates in South America and similar regions worldwide.

3.5 Conclusions

In this study, ANNs were developed to calculate G via the integration of satellite remote sensing and meteorological reanalysis data. Compared to that of existing models, the performance of the generated ANNs was better overall, yielding lower errors and higher correlation values. This indicates that the ANN's structure can better approximate the behavior of G over various land covers and climate conditions.

The inclusion of land cover information into the ANNs as an input improved the accuracy of the G predictions. The superior performance of the ANNs with land cover as an input indicates that the commonly used remote sensing data may be insufficient to fully capture the differences among the surfaces and appropriately predict G. However, it is recommended to include additional remote sensing datasets in such models, especially those used in image land cover classification, instead of land cover data. This would remedy possible issues with the lack of standardization of land cover classification systems. On the other hand, the addition of input data sets increases the complexity of the ANNs and may even reduce accuracy. Thus, parsimony is recommended when selecting predictor data sets.

These findings demonstrate that the developed ANNs can predict G spatiotemporal variability more accurately than existing models. Despite the limitation of the distribution of the available flux towers, the wide variety of land covers considered, encompassing most of South America, and the length of the time series used in the ANN's training mean that the developed ANNs also yielded a higher generalization ability than the existing models. However, the ANN's accuracy over high altitude and meridional land covers should also be assessed for greater reliability.

For future studies, we suggest mapping G over the whole South America using the ANNs and a comparison of this to existing global products. The investigation of the effects of ANN-based G estimates on error reduction of surface energy balance fluxes and evapotranspiration modelling is also recommended.

3.6 Supplementary Information – ANN's synaptic weights and input variable relative importance

Tables SI.1 and SI.2 display the connection weights of the trained Gc and Gnc ANN's, respectively, which were used in Eq. (2) to calculate G. The weights are discriminated by layer: first, the connection weights of the inputs and input layer bias ($b_{h,i}$) to the hidden neurons ($w_{h,i,j}$); then, the weights between the output neuron ($w_{o,i}$), and the hidden neurons and hidden layer bias (b_o).

Table SI.1 – Synaptic weights of the trained Gc ANN.

Hidden Layer								
j	**Input Layer**	**$w_{h,1,j}$**	**$w_{h,2,j}$**	**$w_{h,3,j}$**	**$w_{h,4,j}$**	**$w_{h,5,j}$**	**$w_{h,6,j}$**	**$w_{h,7,j}$**
1	R_n	-18.4696	12.7293	1.2707	-8.7263	-8.2317	-10.7398	-12.4299
2	T_s	-10.4746	-11.5024	-1.7972	-4.4871	7.5452	-0.0310	3.0522
3	α	6.0675	3.6202	-1.5794	-5.8049	-2.9073	-2.6185	4.1044
4	EVI	10.1455	-25.1095	-4.7671	5.7186	2.5864	23.4429	8.4487
5	Land Cover	9.6609	3.7326	2.0444	21.5455	14.5421	-6.2494	6.4693
6	$b_{h,i}$	-0.4436	-1.7756	1.0148	-10.9672	-6.8910	2.7639	-1.5949
Output Layer								
Hidden Layer		**$w_{h,1,j}$**	**$w_{h,2,j}$**	**$w_{h,3,j}$**	**$w_{h,4,j}$**	**$w_{h,5,j}$**	**$w_{h,6,j}$**	**$w_{h,7,j}$**	**b_o**
$w_{o,i}$		7.4148	-9.1672	-6.2833	-6.2346	8.6429	-10.0445	-11.2160	12.8041

Table SI.2 – Synaptic weights of the trained Gnc ANN.

Hidden Layer						
j	**Input Layer**	**$w_{h,1,j}$**	**$w_{h,2,j}$**	**$w_{h,3,j}$**	**$w_{h,4,j}$**	**$w_{h,5,j}$**
1	R_n	8.4093	17.3811	-8.2284	-5.7787	0.0505
2	T_s	-2.0826	-0.3879	-9.8736	-1.2015	3.4544
3	α	1.7308	-5.5842	-2.5098	2.7142	-14.6466
4	EVI	-29.1761	-14.6965	40.1691	-6.4490	7.5938
5	$b_{h,i}$	6.0127	-0.1840	-9.0590	-1.7264	-2.6003
Output Layer						
Hidden Layer		**$w_{h,1,j}$**	**$w_{h,2,j}$**	**$w_{h,3,j}$**	**$w_{h,4,j}$**	**$w_{h,5,j}$**	**b_o**
$w_{o,i}$		-10.8514	1.7316	-8.6113	-0.2803	-6.6283	7.4246

Table SI.3 presents the input variables relative importance to the Gc and Gnc ANN's based on the synaptic connections' weights. R_n and T_s presented similar importance to both ANN's. However, lower importance was yielded by α and EVI in the Gc ANN than in the Gnc one. Land cover yielded similar relative importance to that of R_n.

Table SI.3 – Input variables relative importance to the Gc and Gnc ANN's.

Input Variable	**Gc**	**Gnc**
R_n	24.3%	23.1%
T_s	13.5%	8.6%
α	10.0%	19.3%
EVI	29.0%	48.9%
Land Cover	23.2%	-

Chapter 4: geeSEBAL-MODIS: Continental scale evapotranspiration based on the surface energy balance for South America

Bruno Comini de Andrade [1]

Anderson Ruhoff [1]

Leonardo Laipelt [1]

Ayan Fleischmann [2]

Justin Huntington [3]

Charles Morton [3]

Forrest Melton [4, 5]

Tyler Erickson [6]

Debora R. Roberti [7]

Vanessa de Arruda Souza [1]

Marcelo Biudes [8]

Nadja Gomes Machado [9]

Carlos Antonio Costa dos Santos [10]

Eric G. Cosio [11]

Corresponding Author: Bruno Comini de Andrade (bruno.comini@ufrgs.br)

[1] Federal University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

[2] Mamirauá Institute for Sustainable Development, Tefé, Amazonas, Brazil

[3] Desert Research Institute, Reno, Nevada, United States of America

[4] California State University Monterey Bay, Seaside, California, United States of America

[5] NASA Ames Research Center, Moffett Field, California, United States of America

[6] Google, Mountain View, California, United States of America

[7] Federal University of Santa Maria, Santa Maria, Rio Grande do Sul, Brazil

[8] Physics Institute, Federal University of Mato Grosso, Cuiabá, Mato Grosso, Brazil

[9] Federal Institute of Education Science and Technology of Mato Grosso, Cuiabá, Mato Grosso, Brazil

[10] Dep. of Atmospheric Sciences, Federal University of Campina Grande, Campina Grande, Paraíba, Brazil

[11] Pontifical Catholic University of Peru, Lima, Peru

Highlights

- The geeSEBAL algorithm was adapted to run with MODIS data, providing 8-day evapotranspiration data at 500 m, hereon named geeSEBAL-MODIS;
- Development of a 20-yr (2002-2021) actual evapotranspiration dataset with complete spatial coverage for South America;
- The dataset was compared to eddy covariance measurements and water balance data across the continent;
- geeSEBAL-MODIS yielded higher accuracy and performance than existing global ET datasets at both field and basin scales.

Abstract

Monitoring actual evapotranspiration (ET) is critical for the accurate assessment of water availability and water resources management, especially in areas with dry climates and frequent droughts. The Surface Energy Balance Algorithm for Land (SEBAL) has been used over several land and climate conditions, and is able to estimate ET at field scale with high accuracy. However, model complexity and subjective parameterization have hindered its operationalization, until the recent development of the geeSEBAL model, which implements the SEBAL model on the Google Earth Engine platform. Here, we present a unique methodology for a continental-scale application of SEBAL, called geeSEBAL-MODIS, that employs novel land surface temperature normalization techniques, enabling the application of contextual ET models to very large scales. We introduce a dynamic ET dataset for the entire South American continent, between 2002 and 2021, at 500 m spatial and 8 days temporal resolution. The satellite-based data were compared against daily ET measured at 27 flux towers as well as water balance-based annual ET from 29 large river basins. geeSEBAL-MODIS data were also compared to eight state-of-the-art global ET datasets. At local scale, geeSEBAL-MODIS demonstrated a satisfactory performance ($r = 0.65$, $KGE = 0.64$, $MAE = 0.83$ mm day^{-1} (24.7%) and $RMSE = 1.07$ mm day^{-1} (31.8%)), with negligible bias. At basin scale, geeSEBAL-MODIS generally underestimated ET (bias = -85 mm year^{-1}, $r = 0.65$, $KGE = 0.47$, $MAE = 107$ mm year^{-1} (10.1%) and $RMSE = 137$ mm year^{-1} (12.9%)). Compared to other global datasets, geeSEBAL-MODIS demonstrated better performance over multiple South American biomes, climates and land cover types. The developed dataset caused an error reduction of 13% at local scale, and 30% at the basin scale, when compared to the average performance of the global datasets. The analyses demonstrate that geeSEBAL-MODIS can be

used as a tool for monitoring climate change and human-related impacts on ET. The geeSEBAL-MODIS model opens the path for high accuracy global ET monitoring at moderate to high resolution, supporting advances in water resources management around the globe.

Keywords: eddy covariance, SEBAL; MODIS; South America; water balance.

4.1.1 Introduction

Evapotranspiration (ET) is the sum of water evaporation from the Earth's surface and transpiration from plants. ET is a driver of the terrestrial climate system, recycling precipitation and controlling temperature at local and regional scales (MARENGO *et al.*, 2018). For example, ET from the Amazon rainforest is responsible for a considerable portion of rainfall in southeastern South America, through atmospheric low-level currents (VAN DER ENT *et al.*, 2010), which makes accurate calculation of ET at regional scales necessary to understand the hydrological cycle and to improve water resources management (BALDOCCHI *et al.*, 2019; FENG *et al.*, 2016).

ET measuring instruments, such as weighing lysimeters and eddy covariance (EC) flux towers, provide accurate measures of ET at local scale (approximately 10 to 1,000 m). However, these sensors present high costs of installation and maintenance, being limited to representing ET rates only at the specific sites where they are installed. Thus, other methods have been developed to provide readily available data at field, regional (e.g., an entire river basin) and very large scales (continental and global) based on satellite-driven models. At the very large scale, ET information is essential for the understanding of the Earth's water, energy and carbon cycles (MA; SZILAGYI; ZHANG, 2021). Therefore, ET modeling is key to the understanding of the spatiotemporal variations and dynamics of climate variables. Finally, knowing ET responses to limited water availability is necessary for the development of effective adaptation strategies (IPCC, 2022).

Over the past decades, efforts to perform global scale mapping of ET led to the development of multiple global datasets (CLEUGH *et al.*, 2007; FISHER; TU; BALDOCCHI, 2008; MU; ZHAO; RUNNING, 2011b; RUNNING *et al.*, 2017; ZHANG *et al.*, 2009). However, these approaches are based on models that do not solve the surface energy balance (SEB) equation and yield significant biases as a result, particularly, the underestimation of ET over irrigated croplands and seasonally flooded areas (AGUILAR *et al.*, 2018; BHATTARAI *et al.*, 2019; SOUZA *et al.*, 2019). Additionally, SEB-based models that are mainly driven by

land surface temperature (LST) have shown better performance at both local and regional scales (YIN *et al.*, 2020; ZHANG; KIMBALL; RUNNING, 2016), especially over irrigated agricultural lands. These models use satellite-derived land surface temperature (LST) to calculate latent heat flux (LE), either as a residual of the SEB equation or through partitioning of the available energy. This is possible due to the high sensitivity of LST to the evaporative cooling effect (ANDERSON *et al.*, 2012), which makes SEB models ideal for assessing variations in ET related to limitations on water availability at the land surface. The most widely used SEB models include the Surface Energy Balance Algorithm for Land (SEBAL) (BASTIAANSSEN *et al.*, 1998a, 1998b), Mapping ET with Internalized Calibration (METRIC) (ALLEN *et al.*, 2007; ALLEN; TASUMI; TREZZA, 2007), Disaggregated Atmosphere-Land Exchange Inverse (DisALEXI) (ANDERSON; KUSTAS; NORMAN, 2007), Simplified Surface Energy Balance Index (S-SEBI) (ROERINK; SU; MENENTI, 2000), Surface Energy Balance System (SEBS) (SU, 2002), Two-Source Energy Balance (TSEB) (KUSTAS; NORMAN, 1999), operational Simplified Surface Energy Balance (SSEBop) (SENAY *et al.*, 2013), and Hybrid Single-source Energy Balance (HSEB) (JAAFAR; MOURAD; SCHULL, 2022).

SEB models have been extensively validated and compared at field (BHATTARAI *et al.*, 2017; MELTON *et al.*, 2021; SINGH; SENAY, 2016; WAGLE *et al.*, 2017) and regional scales (MOREIRA *et al.*, 2019; RUHOFF *et al.*, 2022; VELPURI *et al.*, 2013; WEERASINGHE *et al.*, 2020) and, despite showing great potential for ET mapping, the inability to provide results under cloudy skies (BHATTARAI *et al.*, 2019), the time consuming computational processing and the lack of automation have, until recently, presented challenges to their application at larger scales and at operational levels. Moreover, there is no established consensus on the models' performance ranking due to the uncertainties related to input data, as well as assumptions made in the models' parameterization. Nonetheless, SEB models are still the most reliable remote sensing-based method for monitoring ET, since LST serves as an effective quantitative indicator for soil moisture and surface water stress (ANDERSON *et al.*, 2011).

Recently, multi-model approaches have been implemented to improve the performance of SEB models over heterogeneous climatic conditions, with ensembles yielding more accurate ET estimates and more stable performance than individual models (BHATTARAI *et al.*, 2019; ELNASHAR *et al.*, 2021; MELTON *et al.*, 2021; RUHOFF *et al.*, 2022). Among others, the OpenET (MELTON *et al.*, 2021) and the WaPOR (FAO, 2020) initiatives stand out by merging ET from multiple SEB models to aid water resources management, especially to achieve food

security via sustainable agricultural water use in the context of a changing climate and intensified droughts. OpenET combines several SEB models, providing an ensemble dataset at high spatial resolution for the western United States of America (USA), while WaPOR provides ET at multiple spatial resolutions over strategic areas of Africa.

Due to the computationally intensive algorithms used in SEB models, their application at continental and global scales is still limited. Currently, only SSEBop presents a freely available global ET dataset (SENAY; KAGONE; VELPURI, 2020). The SEBAL model has been widely validated over different surface and climate conditions, yielding consistent and accurate calculations of ET (BANDARA, 2003; BASTIAANSSEN *et al.*, 1998b; BASTIAANSSEN; ALI, 2003; BHATTARAI *et al.*, 2017; KIPTALA *et al.*, 2013; LAIPELT *et al.*, 2020; RUHOFF *et al.*, 2012). However, global applications of SEBAL are unknown, and country scale applications are scarce (BHATTARAI; MALLICK; JAIN, 2019; CHENG *et al.*, 2021). Therefore, the SEBAL limitations related to the model's parameterization over heterogeneous climates and complex terrains have hindered its operationalization. Automation attempts for the model's parameter calibration, such as the internal Calibration using Inverse Modeling at Extreme Conditions (CIMEC) (ALLEN, R. *et al.*, 2011; ALLEN *et al.*, 2007, 2013; ALLEN; TASUMI; TREZZA, 2007; KILIC *et al.*, 2020), are sensitive to surface terrain, climatic conditions and the size of the area of interest (AOI), which requires extremes surface conditions, one dry and another wet, often leading to subjective results, especially over humid areas where dry surfaces are hard to identify (BHATTARAI; MALLICK; JAIN, 2019; KHAND *et al.*, 2017; LONG; SINGH; LI, 2011). While other automation attempts were validated at regional scales (BHATTARAI *et al.*, 2017; SABOORI *et al.*, 2021), they still require adaptations for global application.

In the context of the lack of ET datasets based on the SEB equation at continental and global scales, and building upon the development of the geeSEBAL framework (LAIPELT *et al.*, 2021) inside the Google Earth Engine (GEE) cloud computing platform (GORELICK *et al.*, 2017), here we present a new version of geeSEBAL to continental-scale ET mapping, with an example presented for South America. The model innovations include a novel method for LST normalization and a new endmember selection procedure optimized for Moderate Resolution Imaging Spectroradiometer (MODIS) imagery at the continental scale, thus named geeSEBAL-MODIS. The new methodology is validated with multiple in situ ET measurements from a large flux tower network across South America, and with ET derived from water balance

from multiple river basins, and is further compared to eight previously developed global ET datasets.

4.2 Methodology

4.2.1 Input data

The remote sensing and climate reanalysis data utilized to implement the geeSEBAL-MODIS model in this study are described in Table 4.1. Daily datasets were aggregated to the 8-day time window of MODIS AQUA. For datasets with coarser temporal resolution, the image with closest date to each 8-day period was retrieved. All the data were resampled with the bicubic method and reprojected to a 500 m spatial resolution prior to model computation.

Table 4.1 – Datasets used in the implementation of the geeSEBAL-MODIS.

Dataset	Variable	Symbol	Resolution		Source
			Temporal	Spatial	
SRTM	Elevation	Z	-	90 m	(JARVIS *et al.*, 2008)
MYD11A2 V6	Land Surface Temperature	LST	8 days	1,000 m	(WAN, 2014)
MYD11A2 V6	Band 31 emissivity	ε_0	8 days	1,000 m	(WAN, 2014)
MYD13A1 V6	Normalized Difference Vegetation Index	NDVI	16 days	500 m	(DIDAN, 2015b)
MYD13A1 V6	Enhanced Vegetation Index	EVI	16 days	500 m	(DIDAN, 2015b)
MCD43A4 V6	Surface nadir reflectance	SNR	16 days	500 m	(SCHAAF; WANG, 2015)
MCD12Q1 V6	Land Cover Type 1	LC	1 year	500 m	(FRIEDL; SULLA-MENASHE, 2015)
ERA5-Land Hourly	Air temperature at 2 m height	T_a	1 hour	11,132 m	(MUÑOZ SABATER, 2019)
ERA5-Land Hourly	u component of wind speed at 10 m height	u_u	1 hour	11,132 m	(MUÑOZ SABATER, 2019)
ERA5-Land Hourly	v component of wind speed at 10 m height	u_v	1 hour	11,132 m	(MUÑOZ SABATER, 2019)
ERA5-Land Hourly	Hourly downwards surface solar radiation	$R_{S\downarrow}$	1 hour	11,132 m	(MUÑOZ SABATER, 2019)
Soil Texture Class	Soil texture class (USDA system) at 0 cm depth	STC	-	250 m	(HENGL, 2018)

4.2.2 The geeSEBAL-MODIS algorithm

In this section, the general formulation of the geeSEBAL-MODIS model is described, followed by the adaptations to continental scale application, namely, the LST normalization and the automated endmembers selection.

4.2.2.1 Model description

The geeSEBAL-MODIS algorithm (Figure 4.1) uses multiple remote sensing data and complementary meteorological data to estimate instantaneous SEB fluxes (usually between 1:30 and 2:00 PM local time, for MODIS AQUA imagery). Instantaneous ET computation is based on the residual of the Energy Balance Equation (1):

$$LE = R_n - G - H \quad (1)$$

where LE is the instantaneous latent heat flux (W m^{-2}); R_n is the instantaneous net radiation (W m^{-2}); G is the instantaneous soil heat flux (W m^{-2}); and H is the instantaneous sensible heat flux (W m^{-2}). LE represents the energy used on the land surface to vaporize water into the atmosphere. In SEBAL, it is computed as the residual of the SEB closure equation, while the other energy fluxes are calculated from remote sensing and meteorological data.

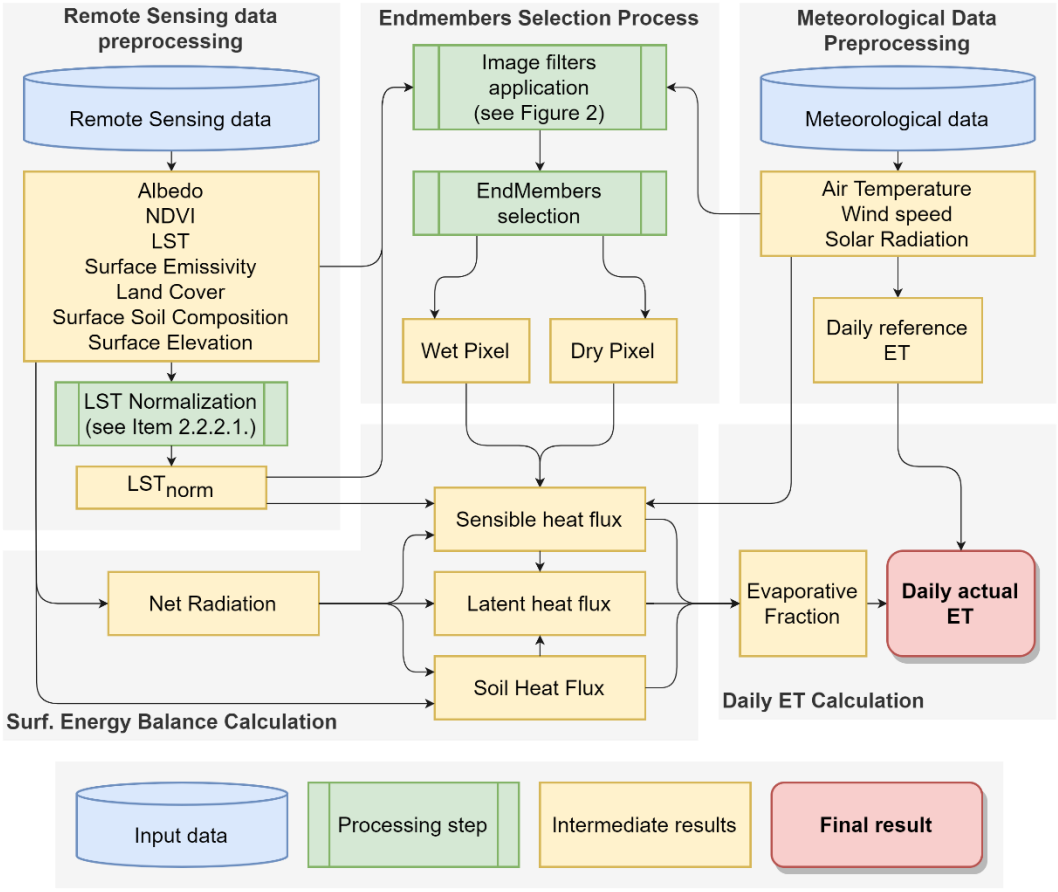


Figure 4.1 – Data processing flowchart of the geeSEBAL-MODIS algorithm.

R_n is the computed through the radiation budget as the difference between incoming and outgoing radiation at the surface, as presented by Eq. (2):

$$R_n = (1 - \alpha)R_{S\downarrow} - R_{L\uparrow} + \varepsilon_0 \cdot R_{L\downarrow} \quad (2)$$

where α is the surface albedo (dimensionless), calculated with a commonly used formulation (TASUMI; ALLEN; TREZZA, 2008); $R_{S\downarrow}$ is the incoming shortwave radiation (W m^{-2}); $R_{L\uparrow}$ is the outgoing longwave radiation (W m^{-2}); $R_{L\downarrow}$ is the incoming longwave radiation (W m^{-2}); and ε_0 is broadband surface thermal emissivity (dimensionless). The variables in Eq. (2) are derived for clear sky conditions (ALLEN, R. *et al.*, 2011), which a plausible assumption given that the instantaneous energy fluxes are computed for visible pixels only.

G is computed from Eq. (3) (BASTIAANSSEN *et al.*, 1998a):

$$\frac{G}{R_n} = (LST - 273.15)(0.0038 + 0.0074\,\alpha)(1 - 0.98\,NDVI^4) \quad (3)$$

where LST is the surface temperature (K); and NDVI is the normalized difference vegetation index (dimensionless). Even though there are other empirical equations for the relationship between G and R_n found in the literature (DANELICHEN *et al.*, 2014; PURDY *et al.*, 2016), with improved results observed in South America machine learning models (DE ANDRADE *et al.*, 2021a), we opted for the original equation, since all methods present limitations when applied to a broad set of climate and surface cover conditions. For water bodies, G was arbitrarily defined as 20% of R_n, to maintain spatial continuity. Although the G/R_n rate changes significantly over time, this ratio falls in the observed range shown in the literature (LIU; ZHANG; DOWLER, 2012; ZHAO; LIU, 2018) and is considered a reasonable first estimate.

H is determined by Eq. (4):

$$H = \frac{\rho_{air} \cdot Cp \cdot dT}{r_{ah}} \quad (4)$$

where ρ_{air} is the air density (kg m^{-3}); C_p is the specific heat capacity (J kg^{-1} K^{-1}); dT is the near surface temperature gradient (K); and r_{ah} is the aerodynamic resistance (s m^{-1}) between two heights z_1 and z_2.

One of the main characteristics of SEBAL is the use of an estimated dT value instead of the difference between LST and air temperature (T_a), which allows it to overcome the T_a uncertainties (ALLEN, R. *et al.*, 2011). dT is obtained via the assumption of a linear relationship to LST, according to Eq. (5):

$$dT = a + b \cdot LST \quad (5)$$

where a and b are empirically determined for each satellite image. To define the linear relationship, two extreme conditions are required, a hot and dry one, where LE is assumed as zero; and another that is colder and wet, where the H value is assumed to be equal to zero. The "cold" and "hot" pixels (endmembers) selection methods are described in Item 4.2.2.2.

The value of H is determined iteratively, via corrections to the aerodynamic resistance and to the friction velocity based on the Monin-Obukhov similarity (BASTIAANSSEN *et al.*, 1998a). In the first step, the value of r_{ah} is estimated by Eq. (6):

$$r_{ah} = \frac{\ln \frac{z_1}{z_2}}{u^* \cdot k} \tag{6}$$

where z_1 and z_2 are heights above the zero-plane displacement height (0.1 and 2.0 m are commonly used values); u^* is the friction velocity (m s^{-1}); and k is the von Karman constant (dimensionless), equal to 0.41. The friction velocity in the first step is computed for each pixel, according to Eq. (7):

$$u^* = \frac{k \cdot u_{200}}{\ln\left(\frac{200}{z_{om}}\right)} \tag{7}$$

where u_{200} is the wind speed (m s^{-1}) at 200 m above the surface; and z_{om} is the momentum roughness length (m) (WATERS *et al.*, 2002).

After the iterative computation of H, instantaneous LE can be determined via Eq. (1). To compute daily ET values, equations (8) and (9) are used, in a similar approach used by geeSEBAL in the OpenET ET ensemble (MELTON *et al.*, 2021):

$$FE = \frac{LE}{R_n - G} \tag{8}$$

$$ET = FE \cdot ET_r \tag{9}$$

where FE is the evapotranspiration fraction (dimensionless); ET is the daily evapotranspiration (mm day^{-1}); and ET_r is the daily reference evapotranspiration (mm day^{-1}). ET_r values are calculated for a rough surface crop, such as alfalfa (ASCE-EWRI, 2005), Eq. (10):

$$ET_r = \frac{0.408\,\Delta(R_{n\,24h} - G_{24h}) + \frac{1600}{T_{air\,24h} + 273.15} \cdot \gamma \cdot u_{2\,24h}(e_{s\,24h} - e_{a\,24h})}{\Delta + \gamma(1 + 0.38\,u_{2\,24h})} \tag{10}$$

where $R_{n\ 24h}$ is the daily net radiation (MJ day^{-1} m^{-2}); G_{24h} is the daily soil heat flux (MJ day^{-1} m^{-2}); $T_{air\ 24h}$ is the daily mean air temperature (°C); γ is the psychrometric constant (kPa °C^{-1}); $u_{2\ 24h}$ is the daily mean wind speed (m s^{-1}) at 2 m height; $e_{s\ 24h}$ and $e_{a\ 24h}$ are the saturation and actual vapor pressure for the daily time step (kPa); and Δ is the slope of the saturation vapor pressure-temperature curve (kPa °C^{-1}). G_{24h} is assumed equal to zero. $T_{air\ 24h}$ and $u_{2\ 24h}$ are acquired from meteorological reanalysis data, and $e_{s\ 24h}$, $e_{a\ 24h}$ and Δ are calculated with commonly used equations (ALLEN, R. *et al.*, 2011; ASCE-EWRI, 2005; BASTIAANSSEN *et al.*, 1998a; WATERS *et al.*, 2002) and meteorological reanalysis data. $R_{n\ 24h}$ is computed by the radiation budget (ALLEN *et al.*, 1998-) with $R_{S\downarrow}$ obtained from reanalysis to account for cloud cover during the 8-day period.

4.2.2.2 geeSEBAL model adaptations for continental-scale application

Land surface temperature normalization

One of the main assumptions of SEBAL is the direct relationship between LST and H, correlating the available water for ET to LST. However, other factors also influence LST such as elevation and solar radiation (CHEN *et al.*, 2009; MALBÉTEAU *et al.*, 2017; RIGON; BERTOLDI; OVER, 2006; ZHAO *et al.*, 2019a, 2019b). As such, LST is generally lower at higher altitudes and latitudes (CHEN *et al.*, 2009). Until recently, these phenomena have limited the application of the SEBAL model to flat terrains and to small AOI (ALLEN *et al.*, 2013; BHATTARAI; MALLICK; JAIN, 2019; LONG; SINGH; LI, 2011).

The partitioning of the AOI into several cells that form a grid, with the separate computation of ET for each grid cell, mitigates this issue. However, vegetation distribution and SEBAL's parametrization have an influence on the final ET estimate, which often lead to inconsistent estimates along the edges of adjacent cells (BHATTARAI; MALLICK; JAIN, 2019).

To overcome this major limitation of SEBAL, we present a new and simple method for LST normalization as a function of surface elevation (z) and solar zenith angle (θ). Equations (11-13) show the steps in the novel LST normalization method:

$$LST_{dem} = LST + 0.0065 \cdot z \quad (11)$$

$$LST_{reg} = lr_{cos\theta} \cdot \cos\theta + LST_{cos\theta=0} \quad (12)$$

$$LST_{norm} = LST_{dem} - LST_{reg\,\theta} + 300 \quad (13)$$

where LST_{dem} (K) is the LST adjusted to a common elevation datum (sea level); LST_{reg} is the linear regression (K) between LST_{dem} and the cosine of θ (cos θ); $lr_{\cos\theta}$ and $LST_{\cos\theta = 0}$ are the slope and intercept of LST_{reg}(cos θ); and LST_{norm} (K) is the LST_{dem} subtracted by LST_{reg} and offset by 300 K. Eq. (11) is the widely used LST correction for the adiabatic rate, which removes the surface elevation effect in LST, while Eq. (12-13), combined with the steps described in the next paragraph, represent a novel method to remove the Earth's curvature effect on LST, based on the fact that the radiation incidence angle also affects surface heating (MALBÉTEAU *et al.*, 2017).

To calculate the LST_{reg} coefficients, only densely vegetated pixels on flat terrain were selected (NDVI ≥ 0.70, α < 0.23, terrain slope < 15°). Thus, a mask was applied prior to Eq. (12) computation. To consider the spatial variability of $lr_{\cos\theta}$ and $LST_{\cos\theta = 0}$, South America was divided into numerous 1,200 x 1,200 km cells (see examples in Figure 4.2), randomly distributed across the available densely vegetated pixels of each image, with a minimum distance between the centers of the cells around 600 km, aiming for maximum coverage. For every LST image, a different grid had to be generated, as available densely vegetated pixels vary according to surface conditions as well as cloud cover. $lr_{\cos\theta}$ and $LST_{\cos\theta = 0}$ were calculated for each cell, and images of the entire study area for each coefficient were generated through spatial interpolation using the inverse-distance weighted method. After the interpolation process, Eq. (12) is applied with the gridded $lr_{\cos\theta}$ and $LST_{\cos\theta = 0}$ to compute LST_{reg}.

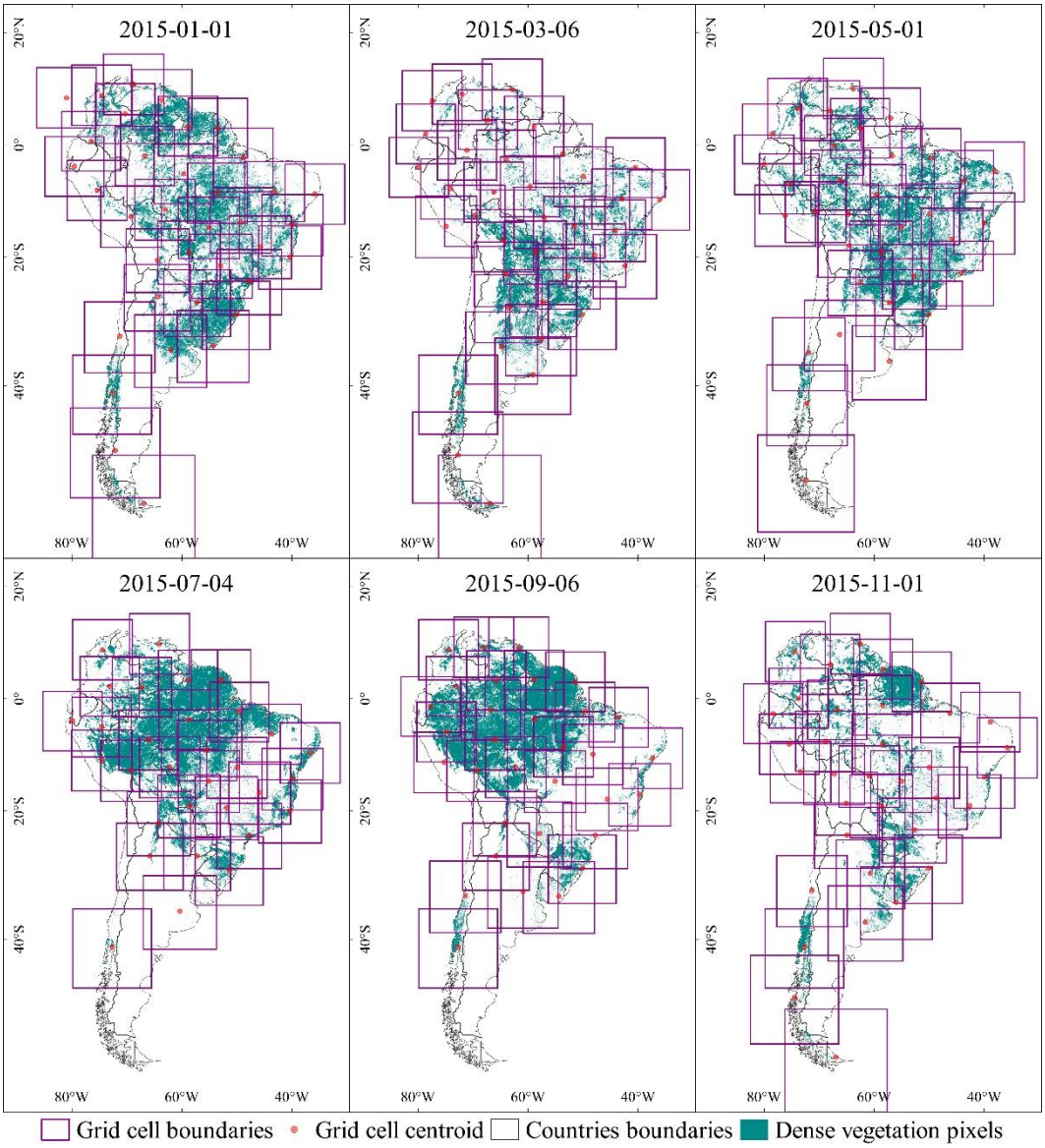


Figure 4.2 – Grids used for the computation of LST_{norm} for six dates of 2015. The grid configuration is dependent on densely vegetated pixels availability and varies with each LST image used.

Automated endmembers selection procedure

The bottleneck for most SEBAL applications lies on the proper selection of the locations with extreme conditions, also known as endmembers. Sensitivity analysis results (FENG *et al.*, 2016; KAYSER *et al.*, 2022; LONG; SINGH; LI, 2011) indicate that the model is most sensitive to hot and cold endmembers' temperatures. Numerous methodologies have been proposed to optimally and operationally locate the endmembers (ALLEN *et al.*, 2013; BHATTARAI *et al.*, 2017; MORTON *et al.*, 2013; SABOORI *et al.*, 2021). The automation of the hot and cold endmember selection enables the operational computation of ET image datasets. However, other issues remain regarding complex topographic landscapes and very large study areas.

In this study, the selection of the endmember candidates followed the steps described in Figure 4.3. The procedure is based on theoretical and empirical knowledge regarding SEB fluxes and LST patterns (ALLEN *et al.*, 2013; BHATTARAI *et al.*, 2017; LAIPELT *et al.*, 2020; SABOORI *et al.*, 2021). As a result, one cold and one hot endmember pixels are selected for each image of South America.

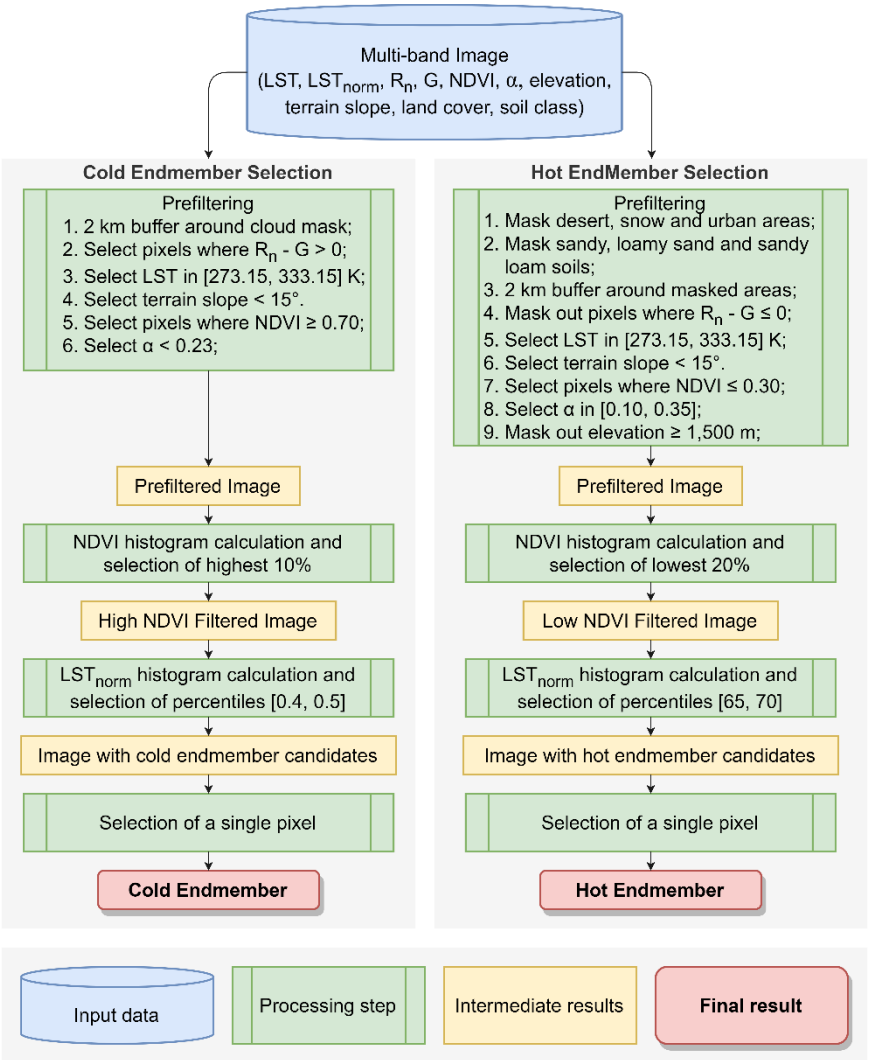


Figure 4.3 – Flowchart of the selection procedure of the cold and hot endmembers for automated calibration.

Both groups of candidates were filtered for positive available energy (R_n – G). This removes regions where there is not enough energy for ET to occur. Additionally, pixels with LST ≤ 273.15 K (0°C) and LST ≥ 333.15 K (60°C) were masked out to ensure the linearity of the relationship between surface temperature and dT. Finally, areas of strong terrain slope were also masked out to avoid extreme illumination conditions. Based on visual inspection, only pixels with slope ≤ 15° were selected as acceptable endmember candidates.

The cloud masking algorithm of MODIS accurately identifies clouds and cloud shadow pixels. However, it commonly leaves residual pixels at the border of such features. These pixels are very detrimental to the cold endmember automatic selection procedure, as these areas tend to be colder than their surroundings. Thus, a buffer is applied to the MODIS cloud masked areas. Through observation of several images, a buffer of 2 km was deemed sufficient to avoid the cloud and cloud shadow residuals. The cold endmember candidates are considered to be well vegetated, so NDVI ≥ 0.70 was selected as the initial criterion. In addition, the dense vegetation generally displays lower reflectivity values, so pixels where $\alpha \leq 0.23$ were selected. From the remaining pixels, the 10% highest NDVI were selected. Then, the cold candidates were defined between the LST_{norm} percentiles of 0.4% and 0.5%, ensuring that the cold endmember is in a very well vegetated area and probably with the highest soil moisture content.

For the hot endmember, pixels identified as desert, snow or urban areas were masked out, as well as pixels classified as sand, loamy sand and sandy loam soils. In such areas, LST reaches higher values (or very low values, in snow), which leads to an overestimation of the hot endmember LST. The masked area received a 2 km buffer to mitigate land cover and soil classification potential issues. The hot endmember candidates are considered to be poorly vegetated or bare soil, so NDVI ≤ 0.30 was selected as the initial criterion. To avoid high emissivity areas such as dark lava rocks and high albedo desert areas, a mask was applied so only pixels with $0.10 \leq \alpha \leq 0.35$ were selected. An additional mask was applied to exclude pixels with elevation over 1,500 m to avoid possible inconsistencies in the LST topographical normalization process, which could strongly influence the hot endmember selection. From the remaining pixels, the 20% lowest NDVI were selected. Then, the hot candidates were defined between the LST_{norm} percentiles of 65% and 70%, ensuring that the hot endmember is in a bare soil area with high probability of negligible soil moisture content. The upper limit of 70% is to avoid pixels that are excessively hot, which would be detrimental to ET modeling, based on sensitivity tests over the South American continent. Given the vegetation distribution and geography of South America, the steps described above ensure that a large number of endmembers are available for each MODIS image, from which one cold pixel and one hot pixel are selected.

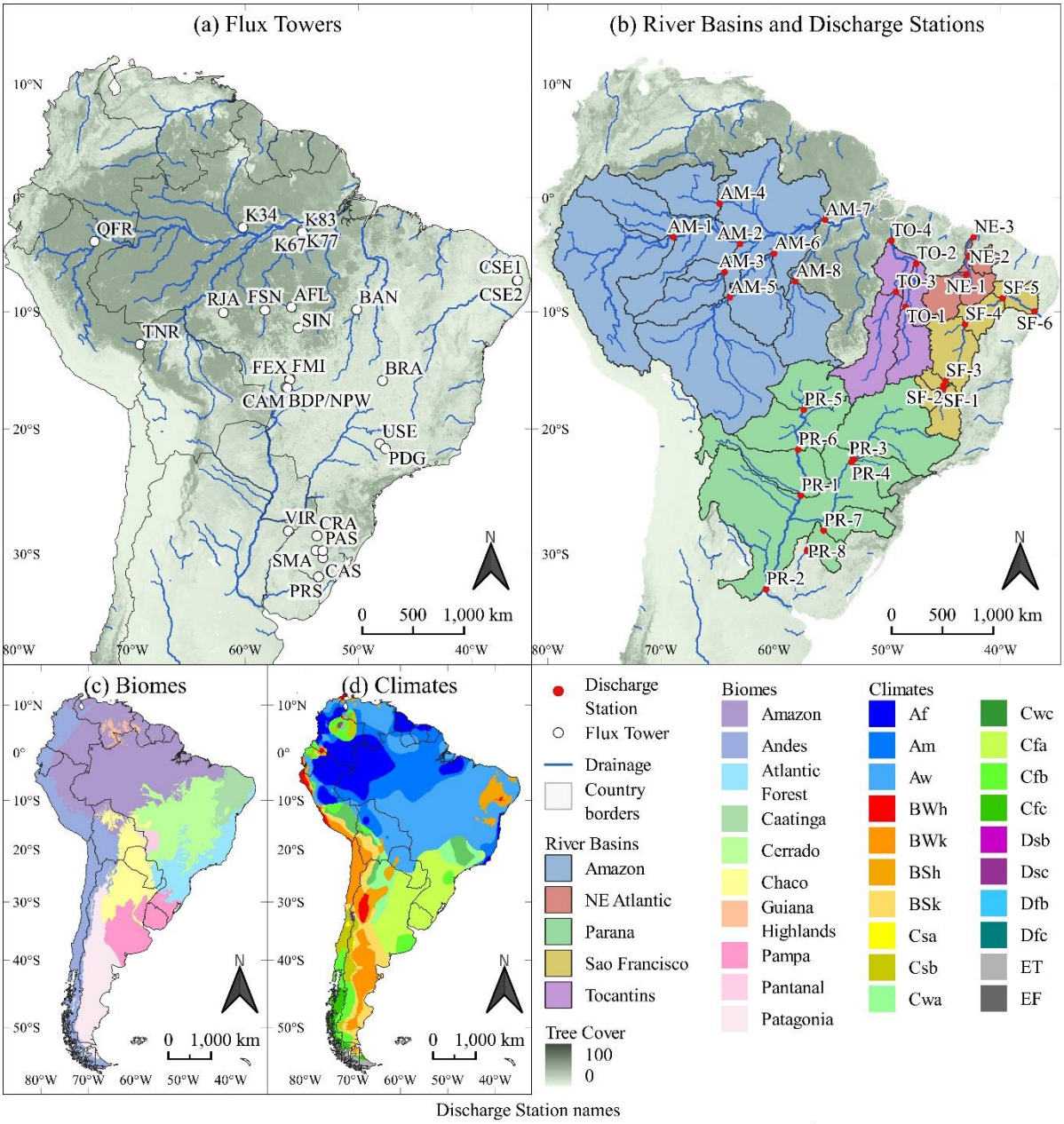
4.2.3 Validation of ET estimates

In this study, we used the correlation coefficient (r), the linear regression slope, the Kling-Gupta efficiency (KGE) (GUPTA *et al.*, 2009), the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the bias as accuracy assessment metrics for the validation of

geeSEBAL-MODIS and other models against ET data from flux towers and calculated from water balance.

4.2.3.1 Validation at local scale

To assess the geeSEBAL-MODIS performance, the satellite-based ET was compared against ET data computed using the EC approach and calculated using observations from micrometeorological instrumentation installed at flux towers sites across South America over the last two decades. This network of EC towers is the culmination of many researchers' efforts from multiple South American institutions, including the large-scale biosphere-atmosphere experiment in the Amazon (LBA-ECO), the South Brazilian network of surface fluxes and climate change (SULFLUX), the National Observatory of Water and Carbon Dynamics in the Caatinga Biome (ONDACBC), the long-term ecological research in Pantanal (PELD Pantanal), and networks from the Federal University of Mato Grosso (UFMT), and the University of Sao Paulo (USP). The location of the flux towers is presented in Figure 4.4, while Table 4.2 shows additional details about each flux tower site.



Discharge Station names

Amazon: São Paulo de Olivença (1), Itapéua (2), Canutama (3), Serrinha (4), Porto Velho (5), Fazenda Vista Alegre (6), Óbidos (7), Barra do Sao Manoel (8); NE Atlantic: Barao de Grajaú (1), Teresina (2), Luzilândia (3); Paraná: Puerto Pilcomayo (1), Timbúes (2), UHE Porto Primavera (3), Porto São Jose (4), Sao Francisco (5), Porto Murtinho (6), Garruchos (7), Uruguaiana (8); São Francisco: Cachoeira da Manteiga (1), São Romão (2), São Francisco (3), Barra (4), Santa Maria da Boa Vista (5), Traipu (6); Tocantins: Miracema dos Tocantins (1), Descarreto (2), Conceicão do Araguaia (3), Tucuruí (4).

Figure 4.4 – Flux towers (a) and discharge stations with each respective river basin (b) used for ET validation, as well as biomes (c) and climates (d) in South America.

Table 4.2 – Flux towers used for ET validation, with information about the surrounding biome and land cover type, measurement technique (Eddy Covariance or Bowen Ratio), ET data availability and source.

Tower	Biome	Land cover	Measurement Technique	Data Availability	Source
AFL	Amazon	Evergreen broadleaf forest	EC	Feb/2003-Apr/2004	(BIUDES *et al.*, 2015)
BAN	Cerrado	Seasonally flooded forest	EC	Oct/2003-Dec/2006	(BORMA *et al.*, 2009)
BDP	Pantanal	Seasonally flooded woodland	EC	Jun/2011-Aug/2015	(BIUDES *et al.*, 2015)
BRA	Cerrado	Savanna	EC	Dec/2010-Jan/2012	(SANTOS *et al.*, 2003)
CAM	Pantanal	Seasonally flooded forest	BR	Dec/2006-Oct/2010	(BIUDES *et al.*, 2015)
CAS	Pampa	Croplands (irrigated)	EC	Oct/2009-Mar/2015	(SOUZA *et al.*, 2019)
CRA	Pampa	Croplands (rainfed)	EC	Dec/2008-May/2013	(TEICHRIEB *et al.*, 2013)
CSE1	Caatinga	Closed shrubland	EC	Jan/2013-Sep/2017	(BORGES *et al.*, 2020)
CSE2	Caatinga	Open shrubland	EC	Mar/2013-Feb/2017	(BORGES *et al.*, 2020)
FEX	Cerrado	Pasture	BR	Aug/2006-Mar/2010	(BIUDES *et al.*, 2015)
FMI	Cerrado	Savanna grassland	BR	Apr/2009-Jun/2013	(BIUDES *et al.*, 2015)
FSN	Amazon	Pasture	EC	Mar/2002-Jul/2003	(PRIANTE-FILHO *et al.*, 2004)
K34	Amazon	Tropical rainforest	EC	Jul/2000-Sep/2006	(ARAÚJO, 2002)
K67	Amazon	Primary tropical moist forest	EC	Feb/2002-Jan/2006	(SALESKA, 2016)
K77	Amazon	Pasture then agriculture after Dec/2001	EC	Dec/2000-Oct/2005	(SAKAI *et al.*, 2004)
K83	Amazon	Primary tropical moist forest	EC	Jul/2000-Mar/2004	(GOULDEN *et al.*, 2004)
NPW	Pantanal	Woodland savanna (Seasonally flooded)	EC	Nov/2013-Jul/2017	(DALMAGRO *et al.*, 2019)
PAS	Pampa	Natural grassland	EC	Aug/2014-Sep/2016	(RUBERT *et al.*, 2019)
PDG	Cerrado	Woodland savanna	EC	Jan/2002-Jan/2004	(DA ROCHA *et al.*, 2009)
PRS	Pampa	Croplands (irrigated)	EC	Jun/2003-Jul/2004	(TIMM *et al.*, 2014)
QFR	Amazon	Permanent wetlands	EC	Feb/2018-Nov/2019	(GRIFFIS *et al.*, 2020)
RJA	Amazon	Tropical dry forest	EC	Jul/2000-Nov/2002	(VON RANDOW *et al.*, 2004)
SIN	Amazon	Tropical dry forest	EC	Apr/2005-Jun/2008	(BIUDES *et al.*, 2015)
SMA	Pampa	Natural grassland	EC	Nov/2013-Jun/2016	(RUBERT *et al.*, 2022)
TNR	Amazon	Primary evergreen forest	EC	Dec/2016-Jul/2018	(VIHERMAA *et al.*, 2016)
USE	Cerrado	Cropland (rainfed)	EC	Feb/2001-Jan/2003	(CABRAL *et al.*, 2003)
VIR	Chaco	Evergreen needleleaf forests	EC	Dec/2009-May/2012	(ULKE; GATTINONI; POSSE, 2015)

4.2.3.2 Validation at basin scale

Due to the high spatial variability of ET, the ability of the 27 flux towers to fully represent the heterogeneous South American territory is limited to the land cover over which they were installed and to their relatively small footprint, which is often in the order of a few hundred meters (VOLK *et al.*, 2023). Thus, to evaluate large-scale ET performance, a regional validation of remotely sensed ET data is also recommended. Many studies evaluated the performance of

remotely sensed ET data using water balance and hydrological models (DE ANDRADE, 2018; RUHOFF *et al.*, 2022, 2012; SENAY *et al.*, 2017), which is recommended especially for regions with scarce monitoring data (RUHOFF *et al.*, 2022). Therefore, we coupled precipitation and water storage with in situ streamflow data to perform water balance calculations for 29 river basins in South America (MOREIRA *et al.*, 2019; RUHOFF *et al.*, 2022). As presented in Figure 4.4, they cover five major South American rivers, including the Amazon, Tocantins, São Francisco, Paraná and Parnaíba. Water balance-based ET was calculated annually, between 2003 and 2014, according to Eq. (14):

$$ET = P - Q - \frac{dS}{dt} \quad (14)$$

where ET is the annual evapotranspiration on the basin's surface (mm year^{-1}), P is the annual precipitation over the basin (mm year^{-1}), Q is the annual streamflow measured at the discharge station (mm year^{-1}), and dS/dt is the basin's total water storage variation during a hydrological year (mm year^{-1}). Precipitation was calculated as an ensemble of five datasets: the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) (FUNK *et al.*, 2015); ERA5 Land (MUÑOZ SABATER, 2019); Global Land Data Assimilation System (GLDAS) version 2.1 (RODELL *et al.*, 2004b); Global Precipitation Measurement (GPM) (HUFFMAN *et al.*, 2019); and Terra Climate (ABATZOGLOU *et al.*, 2018). Streamflow data is based on the measurements provided by the Brazilian Water and Sanitation Agency (ANA) (available at https://www.snirh.gov.br, last accessed on 04 November 2022) and the Argentinian Hydrological Database System (SNIH) (available at https://snih.hidricosargentina.gob.ar/, last accessed on 04 November 2022). Storage variation was calculated using the Gravity Recovery and Climate Experiment (GRACE) dataset of total water storage (TWS) (LANDERER; SWENSON, 2012; SWENSON, 2012; SWENSON; WAHR, 2006b). Annual TWS variation (dS/dt) was calculated as the difference of TWS between two time steps, corresponding to the last day of two consecutive hydrological years. Hydrological years were defined with their endings falling on the last day of the month with the lowest streamflow monthly mean value. Due to the temporal availability of all the datasets, the water balance approach was applied from 2003 to 2014. Based on recent findings (VISHWAKARMA; DEVARAJU; SNEEUW, 2018), only basins with areas over 100,000 km² were selected. No gap filing methods were employed. Thus, only years with full data availability were used.

Uncertainty in the ET based on water balance was estimated by Eq. (15) (LONG; LONGUEVERGNE; SCANLON, 2014; RODELL *et al.*, 2004a, 2011; SENAY; BUDDE; VERDIN, 2011; SHEFFIELD *et al.*, 2009):

$$\upsilon_{\overline{ET}} = \frac{\sqrt{\upsilon_{\overline{P}}^2 \overline{P}^2 + \upsilon_{\overline{Q}}^2 \overline{Q}^2 + \upsilon_{\frac{\overline{dS}}{dt}}^2 \frac{\overline{dS}}{dt}^2}}{\overline{P} - \overline{Q} - \frac{\overline{dS}}{dt}} \quad (15)$$

where υ is the relative uncertainty of each water balance component. The absolute uncertainties of dS/dt and P were assumed to be the average of the standard deviation among the different datasets' estimations. Q uncertainty was assumed as equal to 5% of average annual streamflow.

4.2.3.3 Comparison with other global ET datasets

To assess geeSEBAL-MODIS performance relative to previously available ET datasets, we selected eight state-of-the-art global ET datasets (Table 4.3) and compared the validation results of each one at both local and river basin scales, between 2003 and 2014. Three of these models are based on vegetation phenology methods, the Global Land Evaporation Amsterdam Model (GLEAM), MOD16 and Penman-Monteith-Leuning (PML), while the SSEBop is a SEB based model. ERA5 and GLDAS are based on land surface models, the Breathing Earth System Simulator (BESS) is a Biophysical process-based model, and Terra Climate utilizes a simplified water balance approach to estimate ET.

Table 4.3 – Description of the eight ET datasets used for comparison with geeSEBAL-MODIS.

Model	Spatial and temporal resolution	ET method	Advantages	Limitations	References
BESS	5 km, 8 days	Biophysical process-based model	Linkage between carbon and water fluxes	Complex terrain and heterogeneity of land surface are not considered	(JIANG; RYU, 2016; RYU *et al.*, 2011)
ERA5 (v. Land)	0.1 degree, Hourly	Land surface model	High temporal resolution	Complex terrain and heterogeneity of land surface are not considered	(HERSBACH *et al.*, 2020)
GLDAS (v. 2.1)	0.25 degree, 3 hours	Land surface model	High temporal resolution	Complex terrain and heterogeneity of land surface are not considered	(RODELL *et al.*, 2004b)
GLEAM (v. 3.3b)	0.25 degree, daily	Remote sensing (Priestley-Taylor equation)	Moderate meteorological inputs requirements	Limitations in areas with high soil and water evaporation	(MARTENS *et al.*, 2017; MIRALLES *et al.*, 2011)
MOD16 (v. 6)	500 meters, 8 days	Remote sensing (Penman-Monteith equation)	Low complexity for implementation	Requires measured data for model calibration Limitations in areas with high soil and water evaporation	(MU *et al.*, 2007; MU; ZHAO; RUNNING, 2011b)
PML (v. 2)	500 meters, 8 days	Remote sensing (Penman-Monteith equation)	Physically sound	High meteorological data requirements	(ZHANG *et al.*, 2019)
SSEBop (v. 4)[1]	1 km, monthly	Remote Sensing (Simplified surface energy balance)	Low complexity for implementation	Requires clear-sky conditions	(SENAY *et al.*, 2013; SENAY; KAGONE; VELPURI, 2020)
Terra Climate	2.5 arcmin, monthly	One-dimensional water balance (modified Thornthwaite-Mather equation)	Simple application	Sensitive to precipitation error	(ABATZOGLOU *et al.*, 2018)

[1] – SSEBop versions 4 and 5 were compared with both flux towers and water balance ET. Accuracy assessment indicated that SSEBop version 4 yielded better performance than version 5 in South America. Thus, we chose to include SSEBop version 4 in this intercomparison.

4.2.4 ET application cases

To demonstrate breakthrough research opportunities with the new geeSEBAL-MODIS, we provide two applications to foster our comprehension of large and continental-scale ET dynamics.

4.2.4.1 Deforestation and cropland expansion impacts on ET

To evaluate the impacts of land cover change on ET, we investigate ET patterns over two regions in Brazil, the Amazon deforestation arc (CAVALCANTE *et al.*, 2019; GATTI *et al.*, 2021; SPERA *et al.*, 2016; SPRACKLEN; GARCIA-CARRERAS, 2015) and the Urucuia Aquifer System (DE ANDRADE *et al.*, 2021b; GASPAR; CAMPOS; DE MORAES, 2012; GONÇALVES *et al.*, 2020), that are undergoing intensive anthropic alterations, based on land cover data from the MapBiomas project (SOUZA *et al.*, 2020) between 2003 and 2020. Deforestation and cropland expansion areas were identified, and local ET time series were extracted as the pixel value in each area for every processed image. The time series were then resampled to monthly time intervals. Seasonal ET means were also computed for the periods before and after the land cover change.

4.2.4.2 Net evaporation of artificial lakes

Net evaporation (E_n) of an artificial lake is defined as the difference of evaporation from the lake and the ET of the lake if there was no lake, for which the ET of the surrounding area can be used as a surrogate (ANA, 2021b). The E_n constitutes a water use that can heavily impact water resources availability. In this study, we considered the surrounding area as the land within a 5 km buffer around the lake margins. The lake's limits were obtained from MapBiomas (SOUZA *et al.*, 2020) between 2003 and 2020. Seasonal values of lake evaporation, surrounding area ET and the lake E_n were computed, as well as annual trends, for two huge artificial lakes in Brazil (Porto Primavera and Sobradinho reservoirs).

4.3 Results and Discussion

4.3.1 Multi-scale model validation

Figure 4.5 shows the ET time series calculated using geeSEBAL-MODIS and measured at the 27 flux towers in South America. Measured values were averaged for each 8-day time window of the geeSEBAL-MODIS dataset. geeSEBAL-MODIS data agree with the flux tower ET, generally falling within the shaded area that captures the daily variability in ET over each 8-day period. Also, geeSEBAL-MODIS is able to explain local ET seasonality and is sensitive to interannual variability. Scatterplots for all towers with additional performance metrics are available in Supplementary Material (Figure S1). geeSEBAL-MODIS performs best in areas of high seasonal variation. geeSEBAL-MODIS tends to overestimate ET over central Brazil

(BDP, BRA, NPW and PDG) during September, which represents the end of the dry season. This phenomenon is still not fully understood, and additional investigations are needed to understand why the model predictions deviate from measured values in this context. However, a previous study (ALLEN *et al.*, 2021) indicates that the FAO Penman-Monteith method for calculating ET_r tends to be overestimated when reanalysis data are used over hot and arid climates. This artificial increase in atmospheric demand overcompensates for the reduction in available surface moisture observed in the dry season, generating higher values of ET. The results indicate that errors of geeSEBAL-MODIS are within the range found by other validation studies of SEBAL with MODIS data, which are around 20% of average ET from local measurements (RUHOFF *et al.*, 2012; ZHENG; WANG; LI, 2016).

Figure 4.6 shows the annual ET anomalies time series estimated by geeSEBAL-MODIS and the water balance approach for 29 river basins. Overall, geeSEBAL-MODIS underestimates ET compared to ET computed from water balance, with an average bias of -86 mm $year^{-1}$ (around 8%). Higher uncertainty in water balance ET is observed over Amazon basins, which is mainly caused by a larger distribution among the precipitation datasets used. The lowest deviations occur in the São Francisco, Paraná and Northeast Atlantic basins. Meanwhile, ET is consistently underestimated in the Tocantins and Amazon basins. geeSEBAL-MODIS interannual variations mostly occur inside the water balance ET range of uncertainty and agrees with the overall long-term trends. However, ET anomalies for some years present opposite signals. Further investigation is needed to identify the cause of these diverging variations. On one hand, the relatively higher ET anomalies from geeSEBAL-MODIS could be caused by the overestimation in ET_r in dry climates (ALLEN *et al.*, 2021) or by a possible insensitivity of the water balance approach to plant responses and soil hydraulic conductivity effects on positive ET anomalies during short dry periods, due to low stomatal control, deep root zone and antecedent increased vegetation cover caused by a prior wet period (GOULDEN; BALES, 2019; ZHAO *et al.*, 2022). On the other hand, the negative correlation between moisture (precipitation) and atmospheric demand (ET_r) (HUNTINGTON *et al.*, 2011a) can lead to lower geeSEBAL-MODIS ET anomalies than those derived from water balance, especially in areas of humid climate. Scatterplots for all basins with additional performance metrics are available in Supplementary Material (Figure S2).

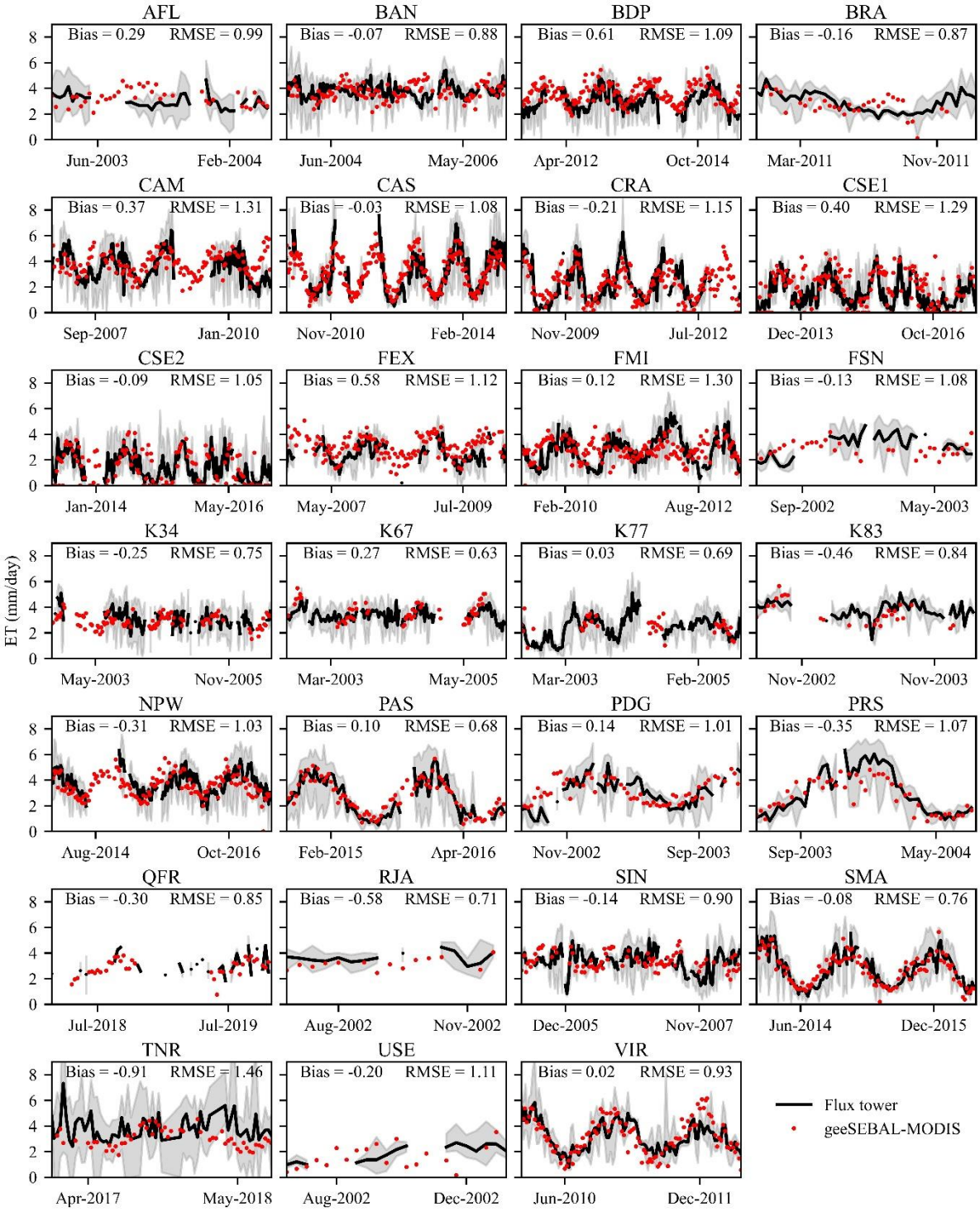


Figure 4.5 – Time series of geeSEBAL-MODIS ET and ET measured at flux tower sites, averaged for every 8-day interval, at 27 sites in South America, with RMSE and Bias in mm day^{-1}. The grey shaded areas represent flux tower ET variation (maximum and minimum values) during each 8-day time window.

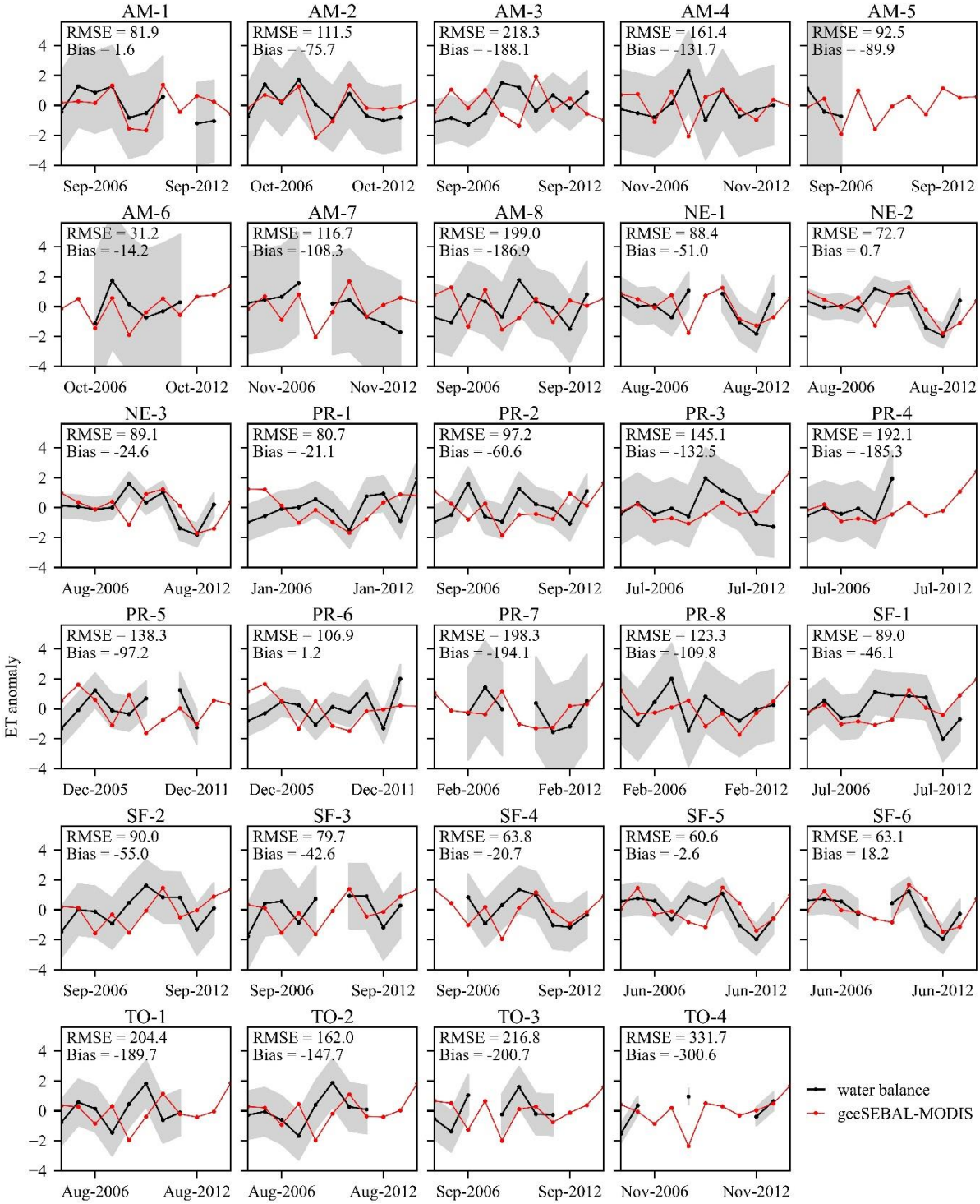


Figure 4.6 – Annual time series of anomalies of geeSEBAL-MODIS and water balance ET of 29 river basins in South America, with RMSE and Bias in mm year^{-1}. The grey shaded areas represent the uncertainty of the ET from water balance.

4.3.2 Intercomparison with global ET datasets

Figure 4.7 shows the distribution of the comparison statistics between the nine datasets (i.e., including geeSEBAL-MODIS) and ET measured at each individual flux tower. All models generally yielded an average slope higher than 0.5. The KGE, r, bias, MAE and RMSE metrics

indicate a superior performance of geeSEBAL-MODIS at the local scale, when compared to the eight global datasets. Scatterplots for all models with aggregated performance metrics are available in Supplementary Material (Figure S3). Given that geeSEBAL-MODIS uses ERA5-Land data for computing ET_r, it was expected to yield similar results in areas of higher FE, where ET values are close to the atmospheric demand. However, ERA5 overestimated ET by around 1 mm day^{-1} for towers located in the Amazon rainforest (K34, K67, K77 and K83), whereas geeSEBAL-MODIS bias in these towers ranged between -0.56 and 0.13 mm day^{-1}. On the other hand, both models yielded similar results over rice paddy croplands (CAS and PRS).

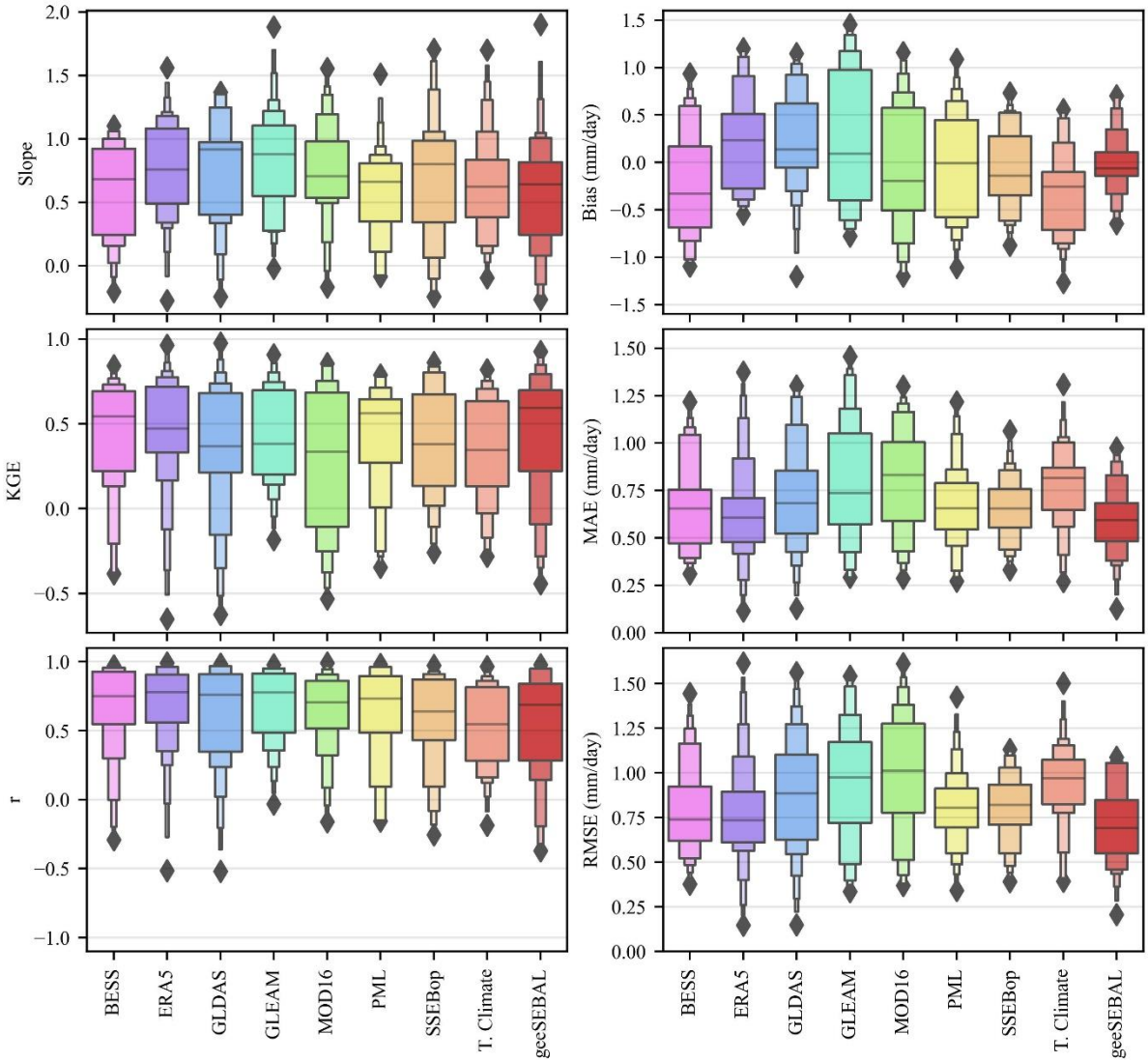


Figure 4.7 – Boxenplots of performance metrics between ET estimated by nine models (BESS, ERA5, GLDAS, GLEAM, MOD16, PML, SSEBop, Terra Climate and geeSEBAL-MODIS) and ET measured at 21 flux towers in South America. Slope, bias (mm day^{-1}), KGE, MAE (mm day^{-1}), correlation coefficient (r), and RMSE (mm day^{-1}) are displayed.

Figure 4.8 presents the comparison of annual ET obtained from water balance and annual ET from geeSEBAL-MODIS and other ET datasets, for 29 river basins in South America, between 2003 and 2014. In contrast to the comparison at local scale, the datasets showed distinct patterns at the basin scale. Wider ranges are identified for slope, r and KGE. ERA5, GLDAS, GLEAM, MOD16 and PML overestimate ET in the Amazon, while underestimating it in other basins. Most models showed lower sensitivity to ET interannual variations in the Amazon when compared to water balance ET. SSEBop yielded consistently lower bias, but also the lowest slope values. Despite the considerable bias of geeSEBAL-MODIS, it yielded the lowest overall errors (MAE and RMSE) of all datasets, and the errors are 30% lower than the average of the other eight datasets. geeSEBAL-MODIS also yielded the highest r-value when compared to the water balance ET.

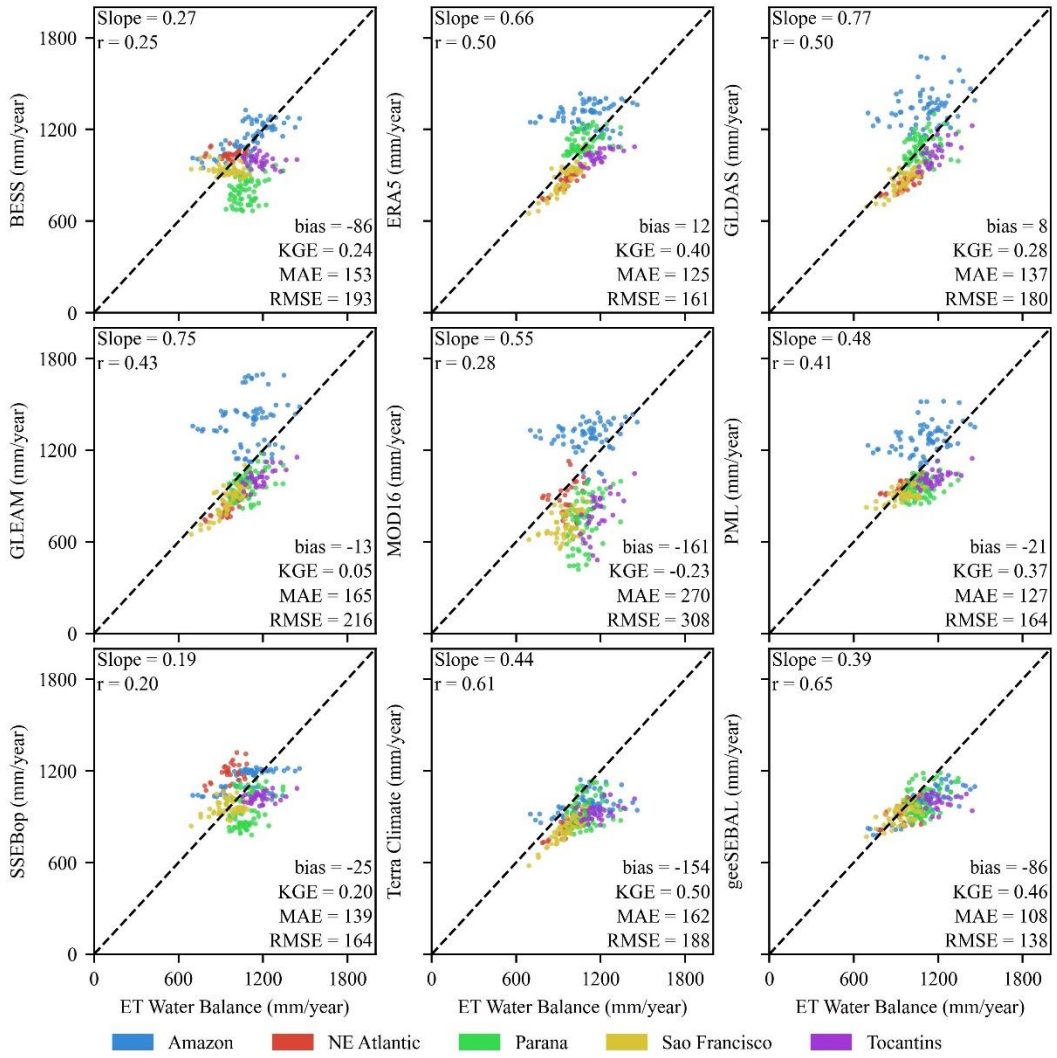


Figure 4.8 – Comparison between annual ET estimated by 9 models (BESS, ERA5, GLDAS, GLEAM, MOD16, PML, SSEBop, Terra Climate and geeSEBAL-MODIS) and aggregated ET calculated from water balance of 29 river basins in South America. The black dashed line represents the 1:1 relationship, and point color represents the major river basin. Slope, correlation coefficient (r), bias (mm year^{-1}), KGE, MAE (mm year^{-1}) and RMSE (mm year^{-1}) are also displayed.

Out of the eight ET datasets used in the geeSEBAL-MODIS cross-validation, we selected Terra Climate, MOD16, PML and SSEBop to perform a spatial analysis of ET patterns. Despite their performance, BESS, ERA5, GLDAS and GLEAM were excluded because of their coarser spatial resolution. Figure 4.9 presents the dataset comparisons over five areas in South America: (a) the reservoir of the Sobradinho Dam, in northeastern Brazil; (b) a portion of the Amazon deforestation arc, in northwestern Brazil; (c) the Pantanal biome; (d) a portion of the Andes Mountains and vegetated plains with areas of cropland, in northern Argentina; and (e) the mouths of the Uruguay and Paraná Rivers into the Atlantic Ocean, near the border between Argentina and Uruguay. Different periods from 2013 were selected for each region, in order to

maximize the ET contrast among the various land covers. For reference, EVI images for each location are also shown.

Terra Climate's coarse spatial resolution is insufficient to represent the spatial features and ET spatial variations in all locations. The other datasets better represent the ET spatial distribution. MOD16 and PML display higher ET over dense forests and lower ET over sparser vegetation, when compared to SSEBop and geeSEBAL-MODIS. Also, both MOD16 and PML do not estimate ET rates over water bodies, as seen at Sobradinho Dam's reservoir and Pantanal wetlands, yielding either masked or extreme values for the analyzed time period (especially PML over the Pantanal). SSEBop and geeSEBAL-MODIS showed similar spatial patterns and are more sensitive to ET differences between natural vegetation and croplands. However, SSEBop yields lower ET rates in deep water pixels, contrasting with the ET rates observed at water pixels near the water bodies margins. All datasets yielded lower ET for the Andean region. For the PML dataset, a gridded pattern is observed, probably due to a coarse resolution of one of its input datasets.

Among the assessed continental to global-scale ET datasets, geeSEBAL-MODIS has one of the highest spatial (500 m) and temporal (8 days) resolutions. Because of this and due to its SEB-based structure, it is able to represent several natural and human-affected surface features, such as land cover change, including deforestation and cropland expansion; irrigated cropland water use; surface water evaporation and net evaporation from artificial lakes; vegetation greening/wilting and water stress conditions. In addition, through coupling of trend analysis and climate change studies, the dataset can also be used for prediction of future conditions, which can improve decision making at both regional and continental scales.

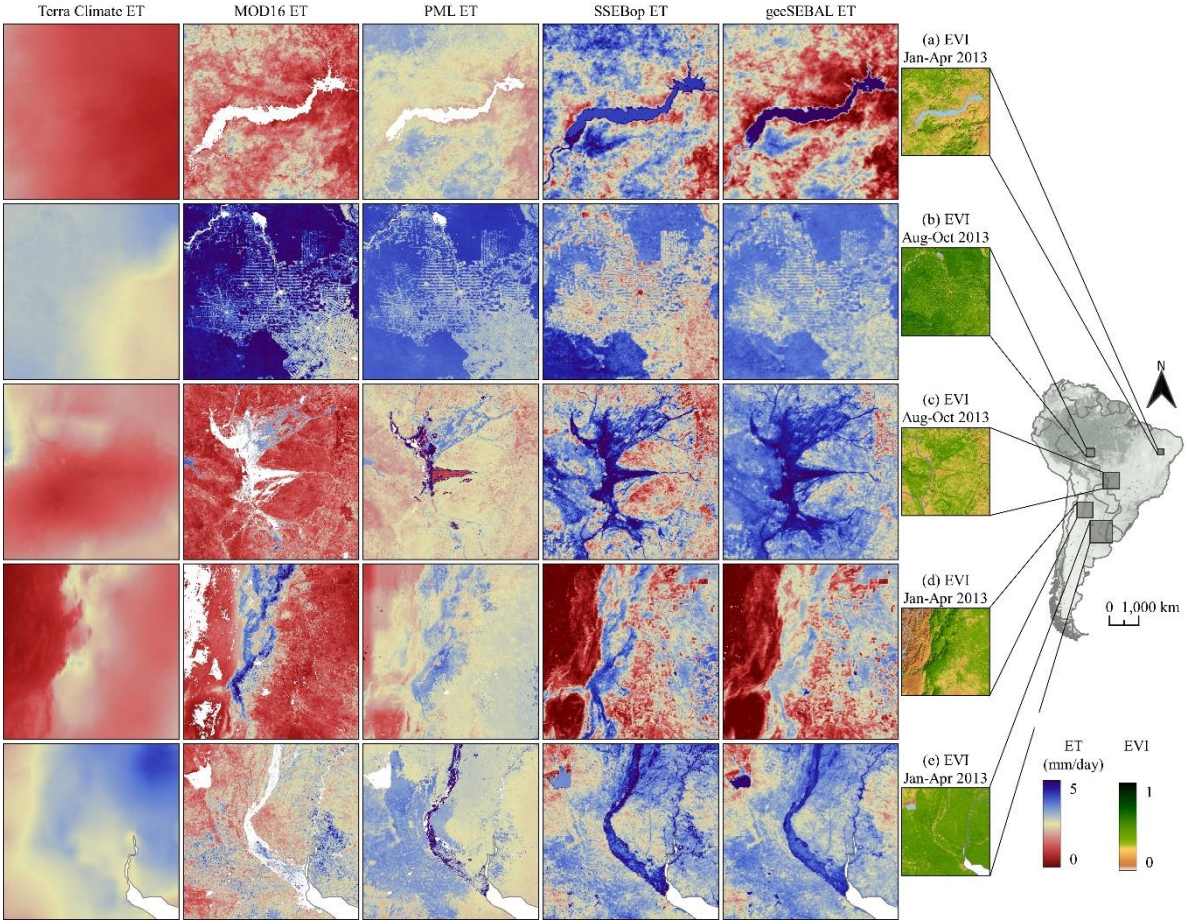


Figure 4.9 – Mean daily ET (Terra Climate, MOD16, PML, SSEBop and geeSEBAL-MODIS) over the reservoir of the Sobradinho Dam (a); a portion of the Amazon deforestation arc (b); the Pantanal biome (c); a portion of the Andes Mountains and vegetated plains with areas of cropland (d); and the lower portions of the Uruguay and Paraná rivers into the Atlantic Ocean (e). Details with EVI values for each location are presented as well. White areas represent pixels with no data.

Incorporation of features such as lower latency (from three months to near-real time) and global spatial coverage should provide geeSEBAL-MODIS the conditions to be used as a tool for land ET monitoring worldwide. Additionally, the combination of multi-sensor moderate resolution LST data, such as MODIS AQUA, MODIS TERRA and, with the decommissioning of these two in the near future, the transition to Visible Infrared Imaging Radiometer Suite (VIIRS) are potential next steps for optimizing geeSEBAL-MODIS accuracy and continuation of global ET long-term monitoring over the next decades. Multi-source satellite images also help overcome the limitation of temporal resolution of the geeSEBAL-MODIS dataset, which was set to 8 days due to the MODIS satellite orbit and to cloud cover conditions. Finally, regional biases identified for some regions in South America need to be investigated and may result in improvements to the geeSEBAL-MODIS model formulation and parameterization.

4.3.3 LST normalization procedure

Figure 4.10 displays bimonthly means of LST (Figure 4.10a), LST_{norm} (Figure 4.10b) and NDVI (Figure 4.10c), derived between 2003 and 2021, as well as scatterplots of LST and LST_{norm} against latitude (Figure 4.10d). Higher slopes of LST versus latitude occur between April and September, due to a higher radiation differential during winter. In the summer, higher temperatures are found in the mid latitudes and in drier areas, such as the Brazilian Northeast and the Patagonia region. LST over the Andes Mountains follows is lower because of the higher altitudes, despite the lack of vegetation. The scatterplots evidence that the procedure successfully removes the geographical position effects on LST, as areas with similar NDVI presented similar LST_{norm} consistently throughout the year. Differences in LST_{norm} for pixels in the same range of NDVI values can, therefore, be associated with the effects of evaporative cooling. Although LST_{norm} on low NDVI areas showed a trend to decrease with higher latitudes, this effect is not detrimental to the geeSEBAL-MODIS algorithm, as the lower dT values in these latitudes can be associated to lower available energy, resulting in lower R_n and H rates.

The LST normalization procedure adjusted LST values based on pixel illumination. The resulting LST_{norm} yielded a higher correlation with vegetation cover and land cover types. This procedure provided the grounds for geeSEBAL-MODIS continental-scale application and the development of the geeSEBAL-MODIS dataset, which overcame the limitations presented in previous very large scale applications of SEBAL (BHATTARAI; MALLICK; JAIN, 2019; CHENG *et al.*, 2021), and maintained the model's independence from reanalysis T_a data during surface energy fluxes computation. The model validation at local and basin scale demonstrated the model's high accuracy and its potential for a future global implementation.

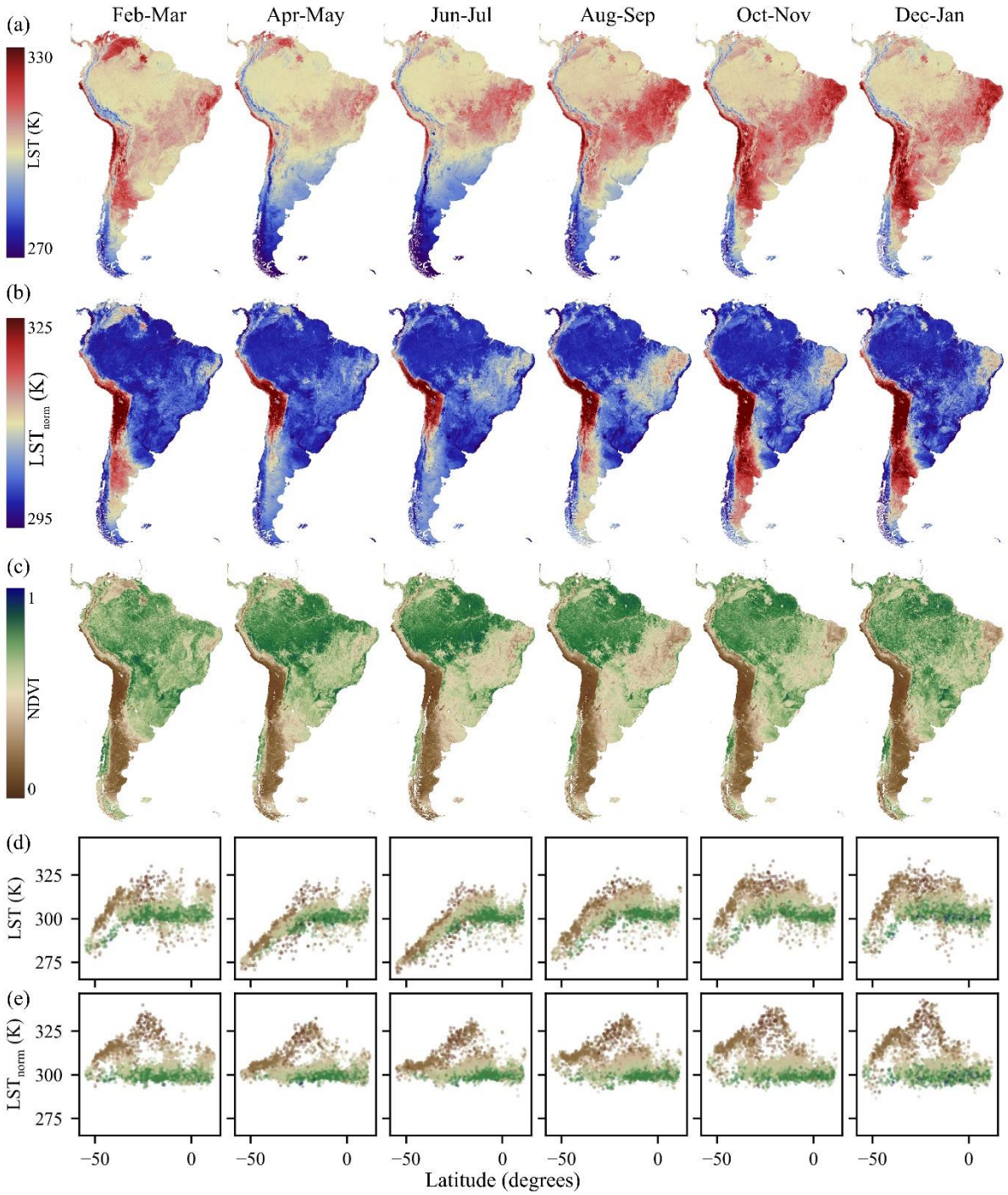


Figure 4.10 – Effects of the LST normalization procedure for the period of Feb-2003 to Nov-2021. Bimestrial average maps of LST (a), LST_{norm} (b), and NDVI (c). Scatterplots for each bimester of LST (d) and LST_{norm} against Latitude (e). The colors in the scatterplots correspond to the NDVI value.

4.3.4 The South American ET regime

4.3.4.1 Long-term spatiotemporal patterns of ET

Figure 4.11 shows long-term (2003-2021) spatiotemporal patterns of ETr, FE and ET in South America. Seasonal variation is observed for ET_r, with lower values during winter and higher ones during summer, especially for higher latitudes, which is related to the seasonality of radiation and T_a. Higher FE values are also observed during summer, in parts of central and northeastern Brazil, but this seasonality is associated with the rainfall regime in the Cerrado and Caatinga biomes, with most rainfall concentrated between November and March. Throughout the year, the highest values of ET_r (and one of the lowest ET) occur in the Brazilian Northeast, where there is a combination of high T_a, wind speed, and radiation, with low atmospheric water content. FE reflects the surface water availability for ET (DE ANDRADE *et al.*, 2021b). Generally, high FE values (between 0.8 and 1.0) are associated with densely forested areas, wetlands, irrigated crops and open water surfaces, as well as along the eastern shoreline. The highest FE are mainly observed in water bodies. Lower FE occurs in the Andes mountains, as well as in Patagonia and parts of Brazil's Northeast, which can be related to the sparser vegetation and are subjected to lower precipitation rates. Average monthly values (2003-2021) of ETr, FE and ET are available in Supplementary Material (Figure S4).

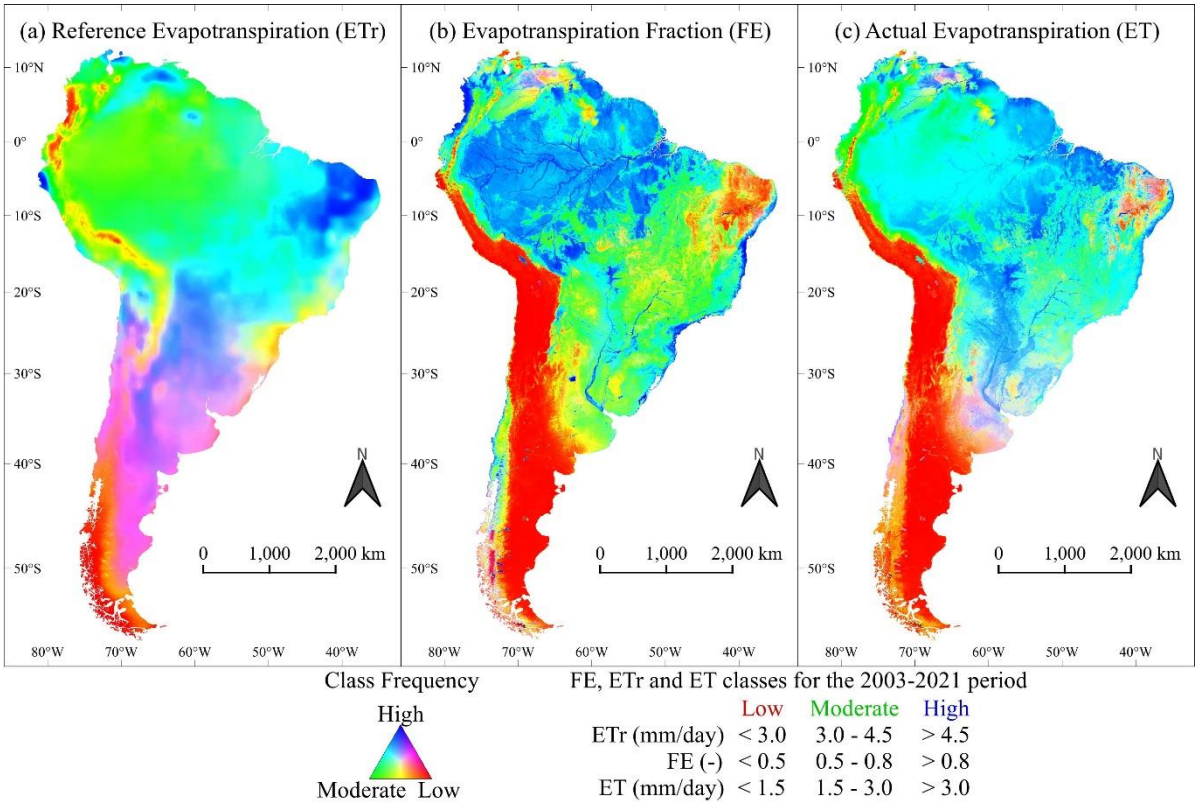


Figure 4.11 – Frequency of (a) ET_r, (b) FE and (c) ET (2003-2021). Colors indicate the frequency at which a pixel yields high (blue), moderate (green) or low (red) ET_r, FE or ET.

4.3.4.2 The drivers of ET in South America

Figure 4.12 presents the atmospheric and surface drivers of geeSEBAL-MODIS ET, derived from the absolute value the correlation coefficient between ET (a) atmospheric (Wind, VPD and R_s) and (b) surface (Albedo, EVI and LST). In areas where ET is classified as low, weak correlations are identified to both atmospheric and surface variables, since seasonal variations of ET are almost absent and rare anomalies can be associated with numerous factors. Over the humid biomes of the Amazon, Chaco, Pampa, Atlantic Forest and south of the Andes, the most dominant drivers of ET are VPD, R_s and LST. On the Brazilian Cerrado and Caatinga biomes, wind speed and EVI play a stronger influence on ET variation. Albedo and LST were found to be the dominant forces over wetlands, riparian forests, and higher elevation areas. These findings agree with results similar studies (PENG *et al.*, 2019), which indicate that EVI and VPD are opposite drivers of ET. EVI is associated with stomatal activity, the parts of plants responsible for transpiration, and correlates positively to ET. Meanwhile, VPD, which is associated to the water content, controls stomatal functions, closing the stomata when VPD is high, thus correlating negatively to ET. r-values between ET and each driver is available in the Supplementary Material (Figure S5).

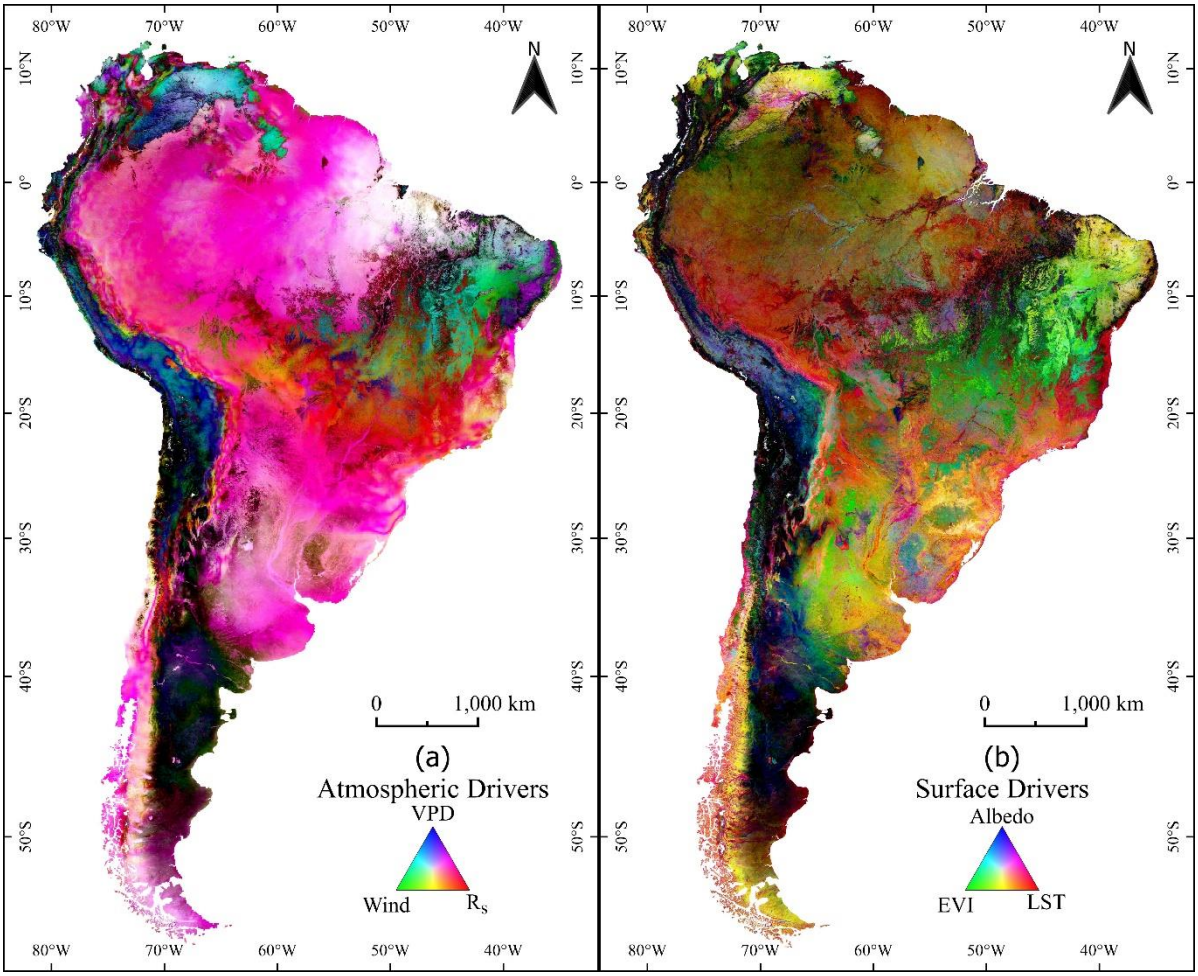


Figure 4.12 – (a) Atmospheric (VPD, Wind and R_s), and (b) surface (Albedo, EVI and LST) drivers of ET in South America, based on the period 2003-2021. Dark areas indicate low correlation between ET and the drivers.

4.3.5 ET application cases

4.3.5.1 Deforestation and cropland expansion impacts on ET

Figure 4.13 displays deforestation and cropland expansion for two areas in Brazil between 2003 and 2020: one in the Brazilian Amazon Arc of Deforestation, and another in the Urucuia Aquifer System, in the Northeast of Brazil. In the Amazon, the black arrow points to an area where deforestation occurred in 2010, becoming a pastureland. In the time series of anomalies, consistent positive and negative anomalies are distinguished before and after the deforestation event, which is expected, given the ET rates of arboreous vegetation compared to grasslands. The chart of monthly means for the area shows lower ET values throughout the year after the deforestation event, especially during the dry season, with a decrease of up to 30 mm month^{-1} in August. In the Urucuia region, the black arrow indicates an area where savanna was replaced by irrigated soy crop in 2009. Higher ET rates occur after the conversion of the area. Also, the

cropland shows less sensitivity to variations in climate conditions, as evidenced by the lower ET rates in the 2008 drought event that are not perceived for other drought events (2012, 2015 and 2020). The higher ET rates follow the crop cycles, occurring twice a year with harvests in March and September/October. These results agree with findings by (LAIPELT *et al.*, 2021), that evaluated high resolution (30 m) ET over human-affected areas in Brazil. Despite the coarse spatial resolution of the MODIS LST datasets, geeSEBAL-MODIS is sensitive to changes in land cover smaller than the 1 km pixel of MODIS Aqua. This example demonstrates that geeSEBAL-MODIS is a valuable tool for quantifying the hydrological impacts of deforestation and cropland expansion and monitoring the impacts of land conversion on evapotranspiration and water consumption.

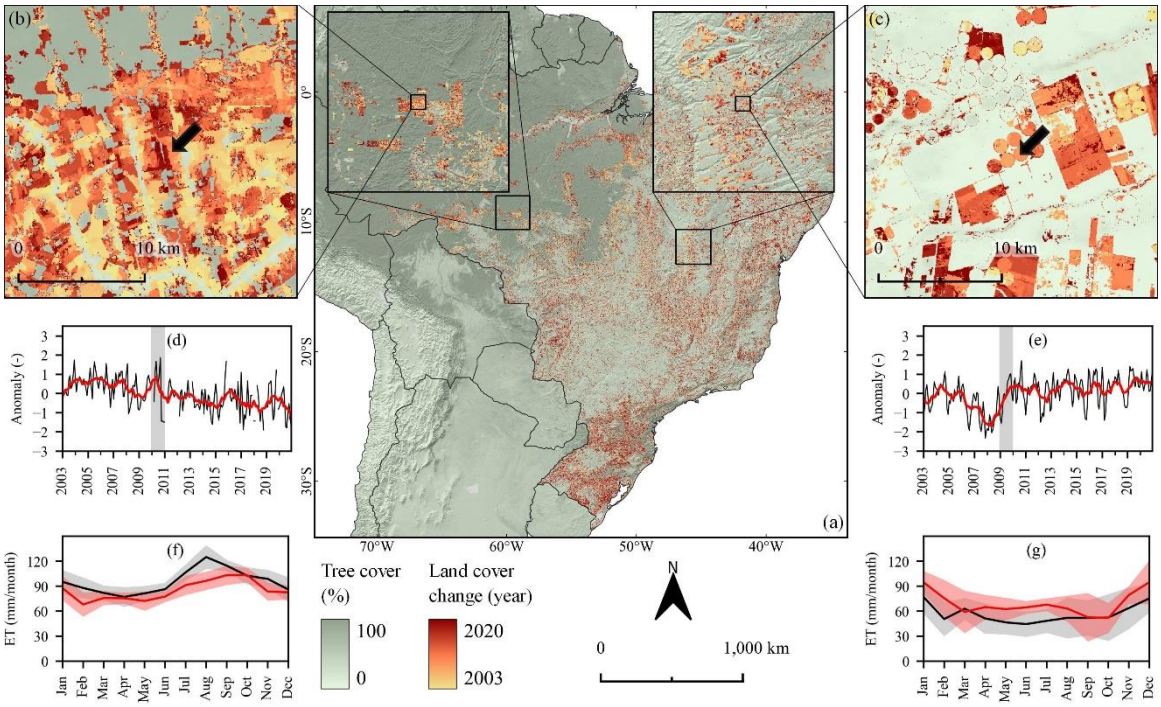


Figure 4.13 – (a) Land cover change corresponding to deforestation and cropland expansion between 2003 and 2020. Details for areas of (b) deforestation occurred in the Amazon, in 2010, and (c) irrigated soy cropland that started in 2009, in the Brazilian Cerrado. ET anomalies time series with three month (black lines and two years (red lines) rolling averages for both areas (d and e), where shaded areas represent the year of land cover change for each location. Mean monthly ET values for both areas before (black lines) and after (red lines) the land cover change (f and g) are also presented, where the shaded areas represent the monthly standard deviation.

4.3.5.2 Net evaporation of artificial lakes

Figure 4.14 displays the E_n analysis for the Porto Primavera and Sobradinho reservoirs, between 2003 and 2020. During this period, Sobradinho yielded a high interannual storage variation than Porto Primavera, with a lower water cover frequency along the borders and the

upstream portion of the reservoir. The same occurs for the ET variation, with higher monthly evaporation standard deviation over Sobradinho. Such differences could be attributed either to climatic variation or to dam operation rules. E_n over Sobradinho averages 1,280 mm $year^{-1}$, which represents a 207% increase in water lost via evaporation because of the open water surface. In Porto Primavera, average E_n is 459 mm $year^{-1}$, representing an increase of 42% of water loss over the artificial lake. The contrasting results for both reservoirs are associated with the different atmospheric demand (ET_r), which is higher over the semi-arid climate around Sobradinho, as well as to the higher vegetation cover around the Porto Primavera reservoir, that yields higher ET rates. Along with the higher water loss in Sobradinho, a positive trend (18.6 mm $year^{-2}$) in E_n is observed between 2003 and 2020, resulting from the decrease of ET in the reservoir surroundings. Porto Primavera, on the other hand, shows decreasing E_n trend (-1.2 mm $year^{-2}$). Compared to findings by (ANA, 2021b), our results of E_n are similar for Porto Primavera, but much lower for Sobradinho. The seasonality, however, is similar for both locations. The magnitude differences might be related to the approach for ET_r calculation, which can cause overestimations of ET, especially over arid and semi-arid regions (Allen *et al.*, 2021).

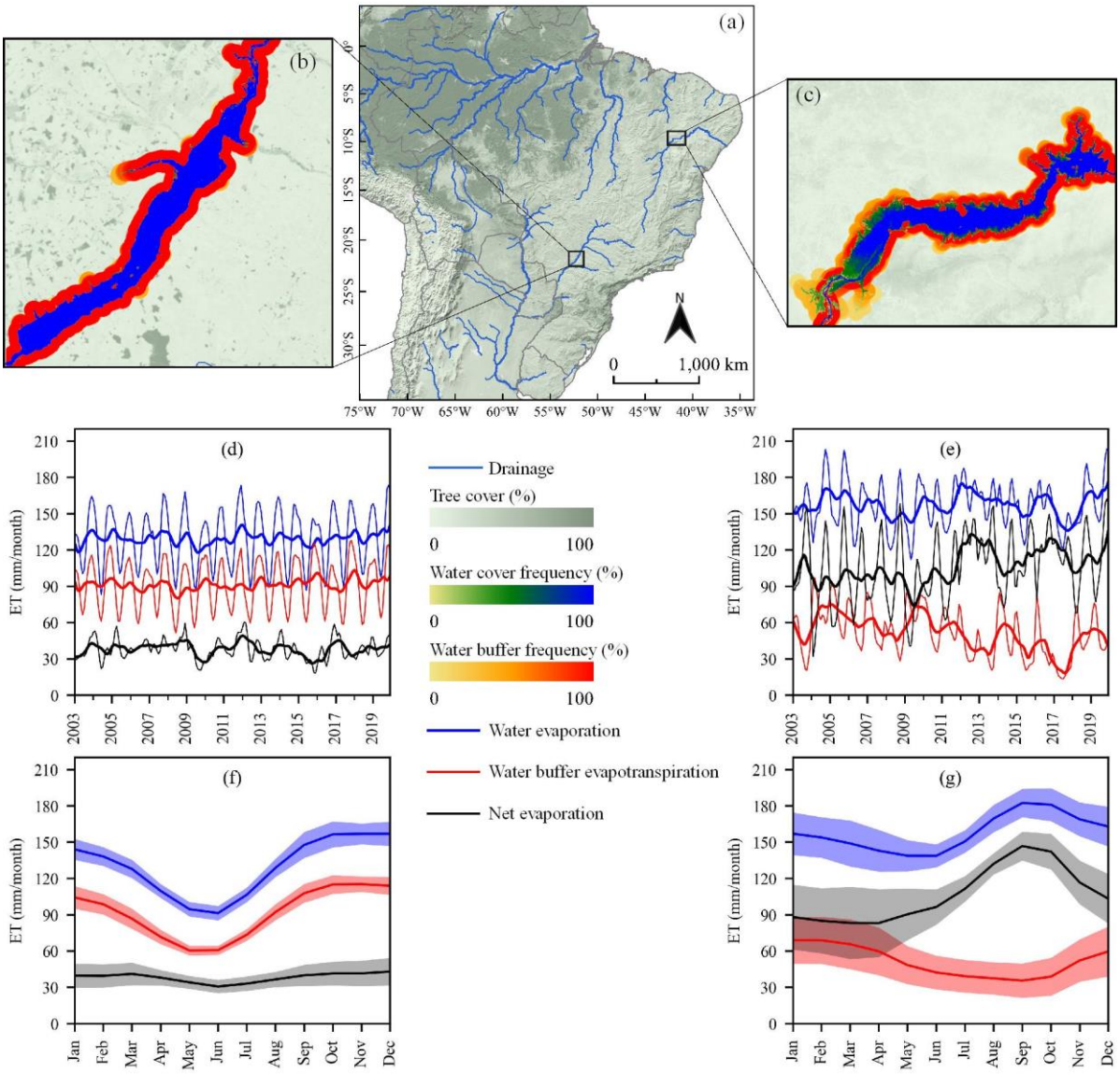


Figure 4.14 – Net evaporation of artificial lakes using ET from the geeSEBAL-MODIS dataset. Tre cover and drainage map of Brazil (a); Porto Primavera and Sobradinho reservoirs with water cover frequency (as in years with water cover over the total of years in the series) and frequency of the 5 km water buffer (b and c); time series of water evaporation, buffer area evapotranspiration and reservoir net evaporation for both artificial lakes (d and e), where the stronger lines represent a rolling average of 12 months; and monthly means of water evaporation, buffer area evapotranspiration and reservoir net evaporation for both artificial lakes (f and g), where the shaded areas represent the monthly standard deviation.

4.3.6 Advances in making continental scale ET datasets locally relevant

geeSEBAL-MODIS ET estimates were generally more accurate than other ET model products, with error reduction of 13% at the local and 30% at the regional scale, when compared to the average error of the other global datasets. geeSEBAL-MODIS also displayed a more consistent performance over the different surface and climatic conditions found in South America, with higher KGE and correlation values, revealing its robustness for continental-scale

ET modeling. We also demonstrate the model's ability to represent local processes, such as deforestation and cropland area expansion. This represents a new era for SEB-based ET modeling, in which the spatial coverage obstacle is no longer a limitation to models such as SEBAL and METRIC.

SEB models are known to better represent the surface-atmosphere interactions than vegetation-based and low spatial resolution models (ANDERSON *et al.*, 2012; BHATTARAI *et al.*, 2019; LU *et al.*, 2019; YIN *et al.*, 2020), especially over irrigated areas. Furthermore, the general trend for improvements in big data computation, quality of meteorological datasets and spatiotemporal resolution of thermal imagery suggest a great potential to increase the accuracy of SEB models beyond what is possible with other available methods (ANDERSON *et al.*, 2012; FISHER *et al.*, 2017). These conditions pave the way for the implementation of novel, very large-scale SEB models, such as the geeSEBAL-MODIS and SSEBop (SENAY; KAGONE; VELPURI, 2020), and expansion of projects like OpenET and WaPOR, helping improve their accuracy and spatial coverage and, consequently, aiding water resources management.

Although recent missions like Landsat 9 (MCCORKEL *et al.*, 2018) and ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) (FISHER *et al.*, 2020a) support a trend towards high-resolution modelling of *ET* (FISHER *et al.*, 2017; MASEK *et al.*, 2020), the generation of datasets at high resolution for very large areas is still computationally expensive with additional demands for storage of hundreds of terabytes to petabytes of data. Therefore, moderate spatial resolution models, like the ones based on MODIS and VIIRS data, provide a near-term pathway to continental and global *ET* modeling, with the advantage of a higher temporal resolution, which can help to address cloud cover limitations in many regions around the world. Given the near-term plans for decommissioning the MODIS instruments and de-orbiting the Terra and Aqua satellites, the VIIRS sensor is a candidate for the continuity of the global scale *ET* model development (TANG *et al.*, 2020)., with the benefit of a higher *LST* spatial resolution (750 m against 1 km from MODIS) while preserving the global daily coverage.

4.4 Conclusions

In this study, we presented the first continental-scale application of the geeSEBAL model at moderate spatial resolution, resulting in the geeSEBAL-MODIS dataset. The ET dataset was developed with 500 m / 8-day spatiotemporal resolution, between July-2002 and December-

2021, covering the entire South America. Accuracy assessment demonstrated that geeSEBAL-MODIS improves ET estimations at field scale by 13% and at basin scale by 30% when compared to the average performance existing global ET datasets. geeSEBAL-MODIS also yielded with a more stable performance, without the overestimations made by vegetation phenology and land surface models in densely vegetated areas. The higher spatiotemporal resolution also suggests that geeSEBAL-MODIS is more appropriate for land monitoring, helping to identify local impacts of human activities on ET, indicating a decrease in deforested areas of the Amazon rainforest and an increase for irrigated cropland expansion in the Brazilian Cerrado biome.

Based on the geeSEBAL-MODIS results, the South American territory was classified regarding ET spatiotemporal patterns. Applications of the model to surface monitoring were suggested at multiple scales, indicating the ability of geeSEBAL-MODIS model to represent both local and regional processes. The geeSEBAL-MODIS development represents a new era in ET modeling, with the ability to combine powerful cloud computing with complex mathematical models based on the SEB approach. This opens possibilities for the development and expansion of SEB-based global datasets, helping to improve water resources management worldwide.

4.5 Data availability statement

The geeSEBAL-MODIS dataset with evapotranspiration estimates for South America is freely available at https://code.earthengine.google.com/?asset=projects/et-brasil/assets/geesebal/myd11a2/sa/v0-02.

4.6 Supplementary material

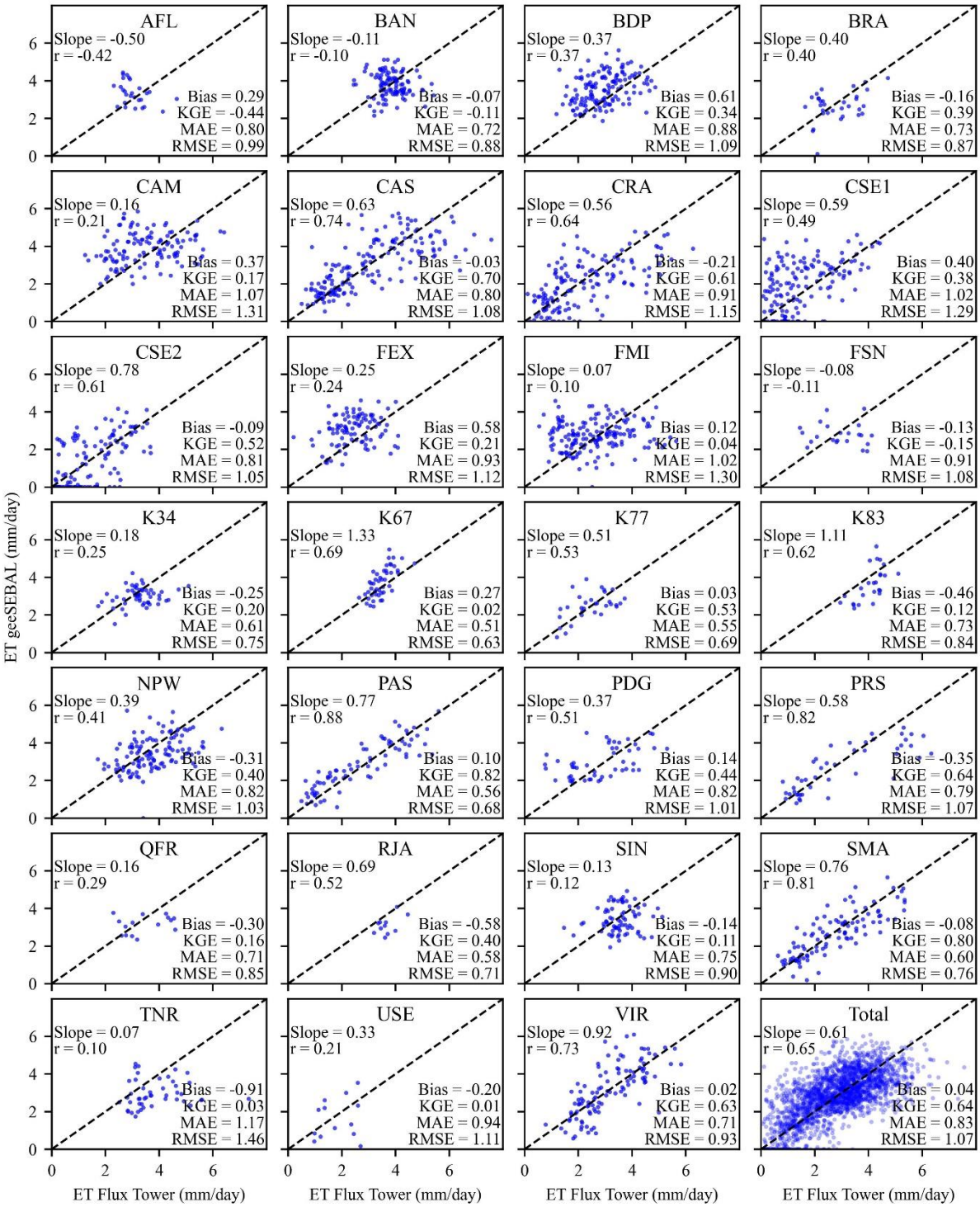


Figure S1. Comparison between ET estimated by geeSEBAL and measured at the 27 flux towers in South America, averaged for every 8-day interval from the geeSEBAL product. The black dashed line represents the 1:1 relationship. Correlation coefficient (r), bias (mm day^{-1}), MAE (mm day^{-1}) and RMSE (mm day^{-1}) are also displayed.

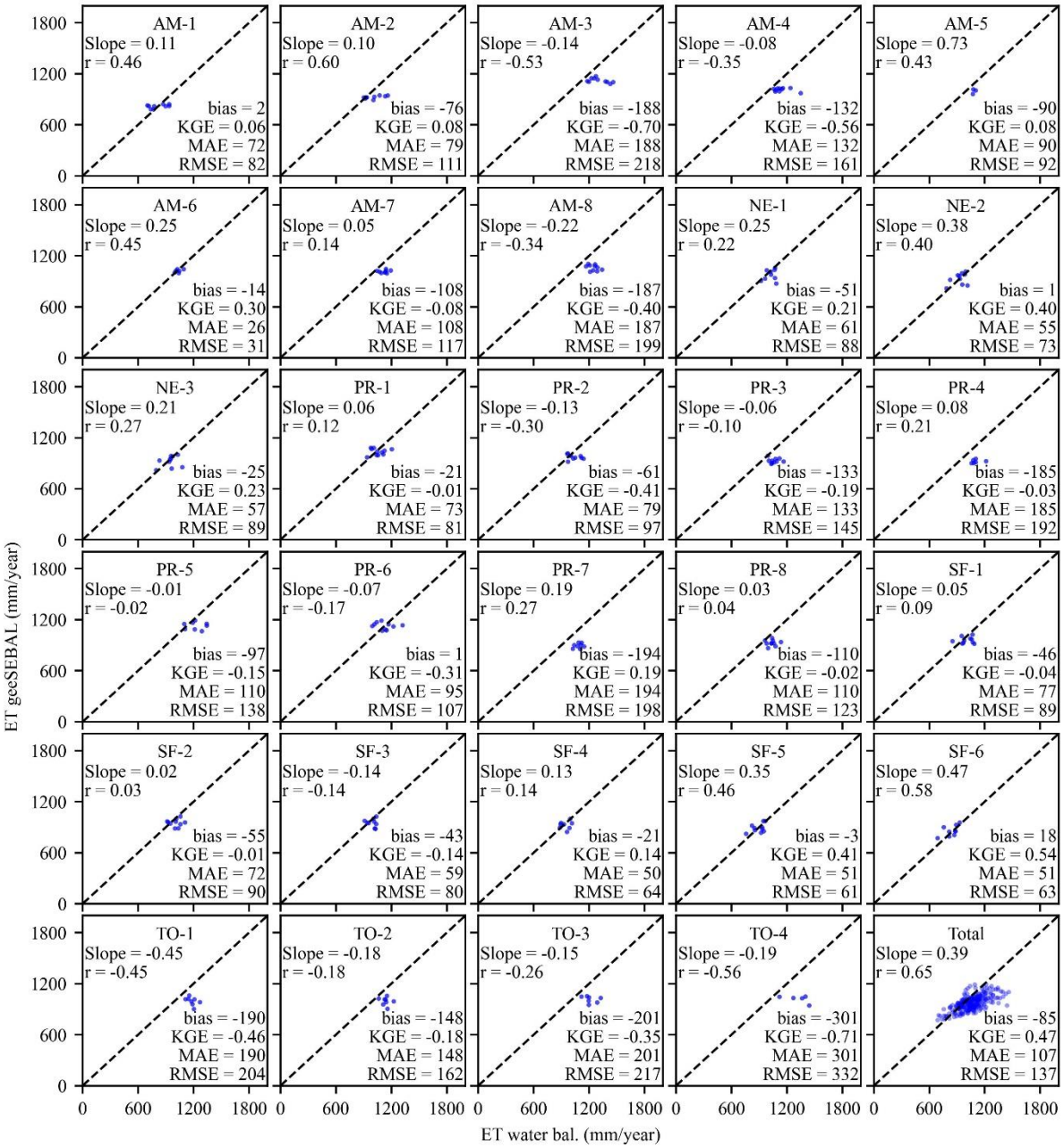


Figure S2. Comparison between annual ET estimated by geeSEBAL and calculated from water balance of 29 river basins in South America. The black dashed line represents the 1:1 relationship. Correlation coefficient (r), bias (mm year^{-1}), MAE (mm year^{-1}) and RMSE (mm year^{-1}) are also displayed.

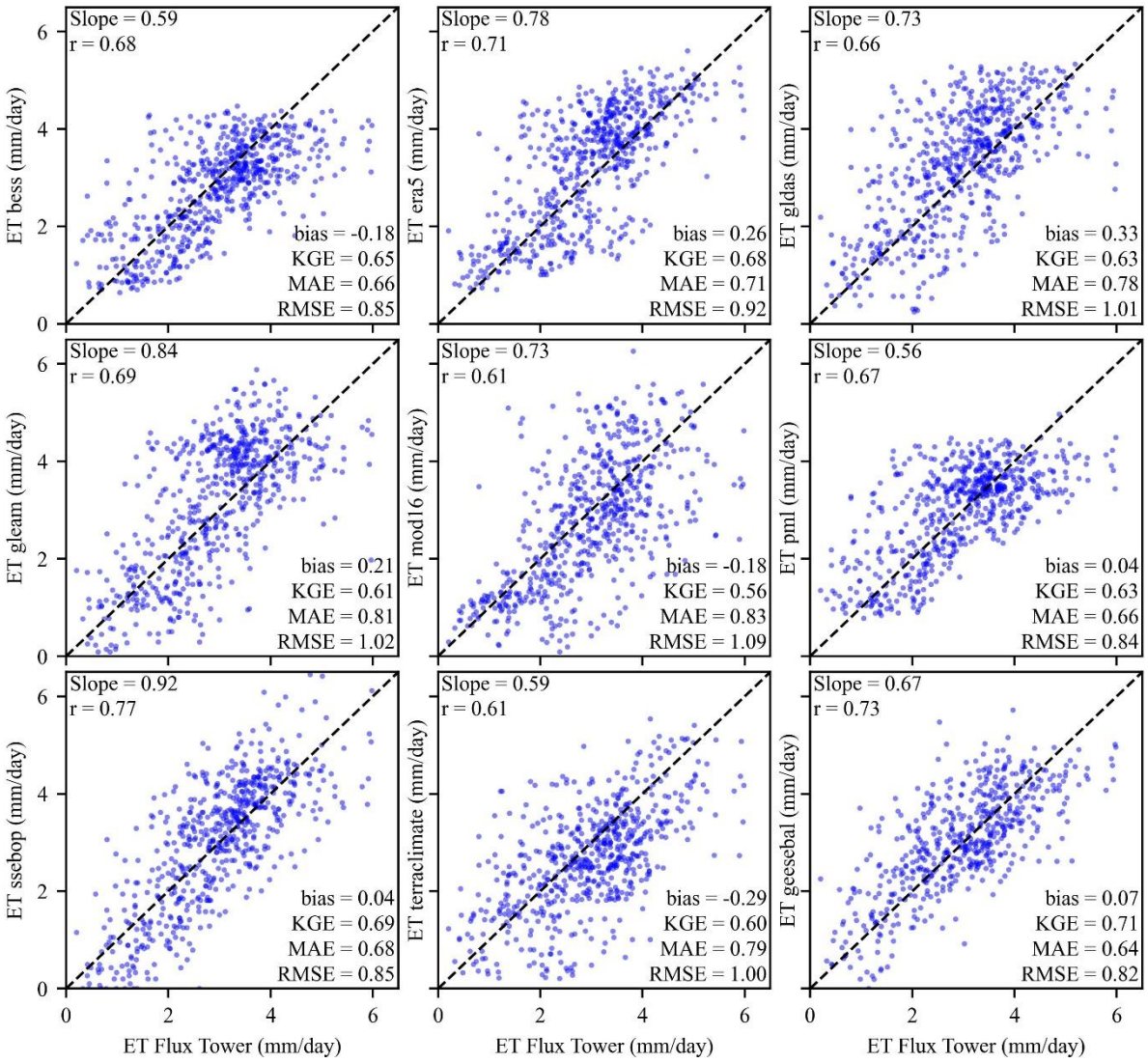


Figure S3. Comparison between ET estimated by 9 models (BESS, ERA5, GLDAS, GLEAM, MOD16, PML, SSEBop, Terra Climate and geeSEBAL-MODIS) and aggregated ET measured at 21 flux towers in South America, at a monthly time interval. The black dashed line represents the 1:1 relationship. Slope, correlation coefficient (r), bias (mm day^{-1}), KGE, MAE (mm day^{-1}) and RMSE (mm day^{-1}) are also displayed.

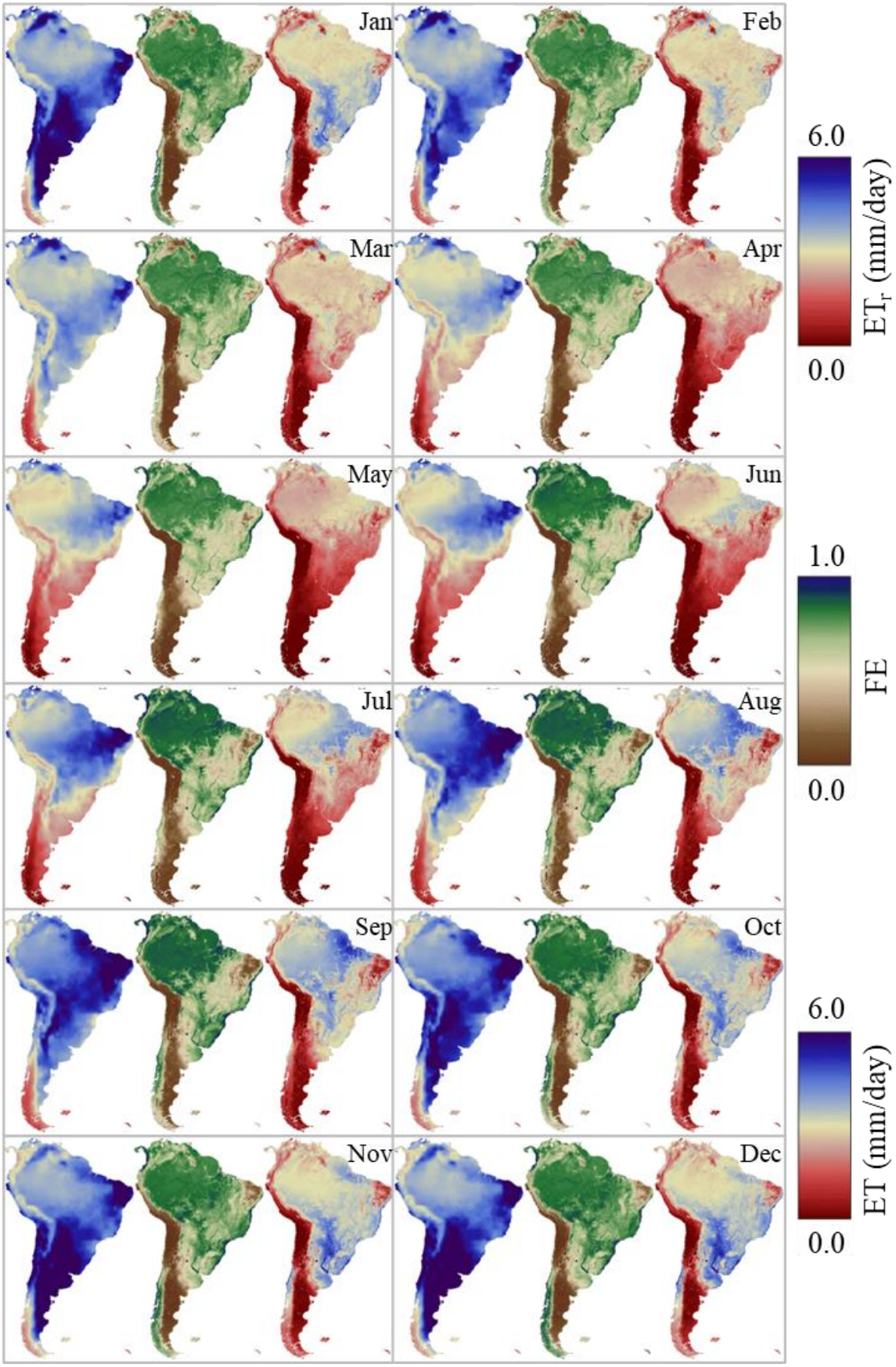


Figure S4. Average monthly values (2003-2021) of ETr, FE and ET, respectively, in South America.

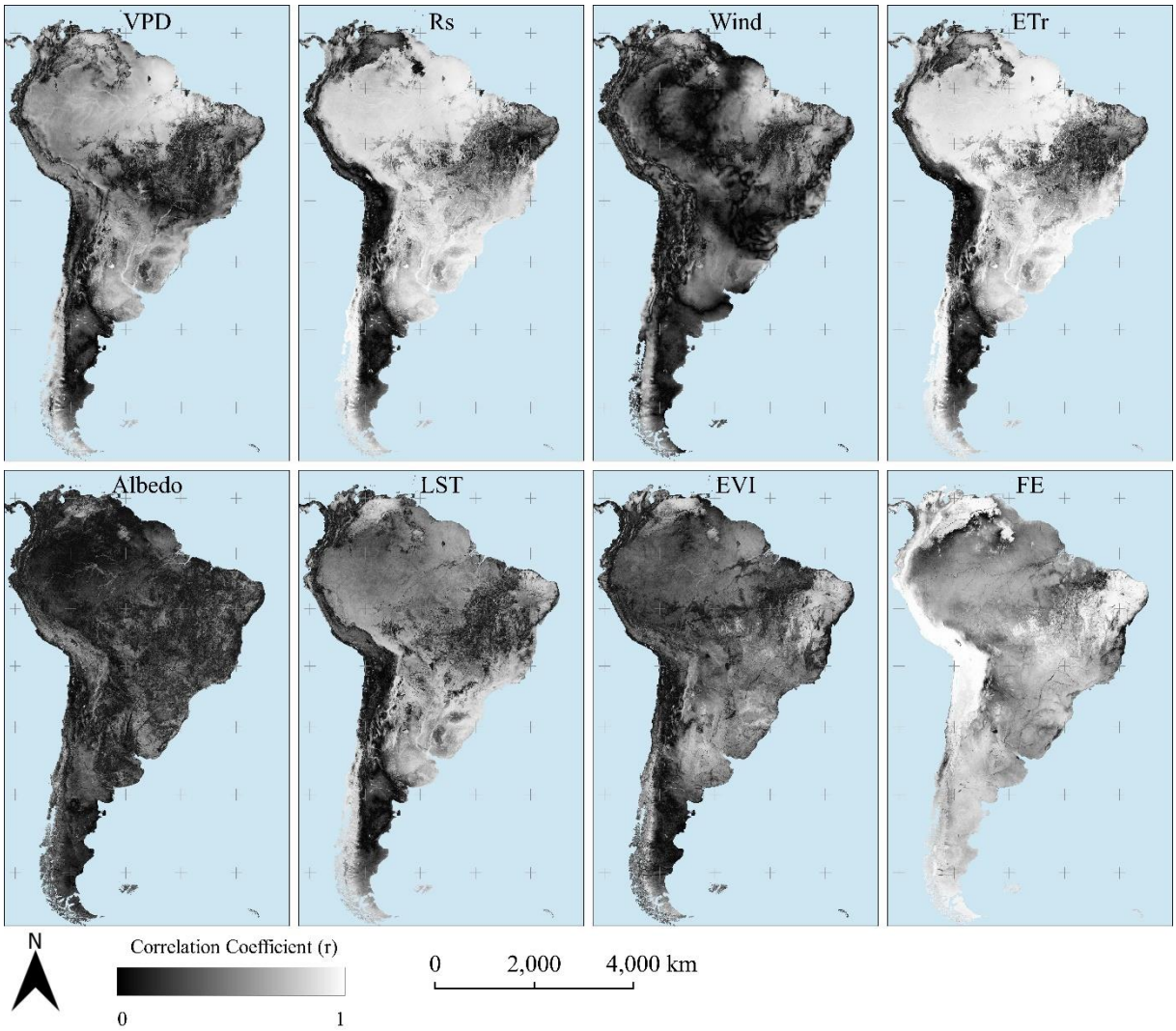


Figure S5. Maps of correlation coefficient (r-value) between ET and various meteorological and remote sensing variables, including FE and ET_r used in geeSEBAL-MODIS, in South America.

Chapter 5: Spatiotemporal Patterns of Evapotranspiration and other Hydrological Processes in Twenty First Century South America

5.1 Introduction

Climate change is a global phenomenon that poses a great challenge on humanity, with particular effects on water management, as changing patterns of precipitation (P) and evapotranspiration (ET) alter the availability and quality of freshwater resources. The effects of climate change on the hydrological cycle are complex, far-reaching, and often divergent ([MASSON-DELMOTTE *et al.*, 2021). As atmospheric temperature rises, more water evaporates from oceans, lakes, and rivers, and transpires from vegetation, leading to increased atmospheric moisture content. This, in turn, results in more intense and frequent precipitation events. However, increased ET also means that some regions are experiencing more frequent and severe droughts, as water is lost from the land surface and vegetation at a faster rate than it can be replenished by precipitation.

In South America, studies indicate that climate change leads to more extreme events of precipitation and droughts, which compromise food security, ecosystems survival, and water supply (CABALLERO; RUHOFF; BIGGS, 2022; CARVALHO, 2020). In order to assess water availability, it is essential to monitor hydrological variables such as P, ET, discharge (Q), and total water storage (TWS). And given the evidence of climatological non-stationarity, it is also necessary to calculate trends and patterns of the water cycle processes changes, which aid in the prediction of future scenarios and development effective adaptation strategies.

This chapter provides an investigation of the closure of river basin water balance in major South American basins, using the novel geeSEBAL-MODIS dataset as the evapotranspiration source, in association with well-established data sources of P Q and TWS. The chapter also describes the trends of ET and correlates it with ten other remote sensing and meteorological variables, investigating their relation and potential causes, such as global warming ([MASSON-DELMOTTE *et al.*, 2021).

5.2 Material and Methods

5.2.1 Input Data

Figure 5.1 presents the 29 river basins located in South America that were used to compare remote sensing and water balance (WB) variables, as well as the climates and biomes of the region. For this study, I used the datasets shown in Table 5.1 to compute WB based

hydrological processes and analyze the trend of meteorological and remote sensing variables. All computations were performed on the Google Earth Engine (GEE) platform (Gorelick et al., 2017). I gathered the datasets' time series between 2000 and 2022 and aggregated them monthly. To ensure pixel area independence from latitude, I performed spatial aggregations for each month using the MODIS sinusoidal projection (SR-ORG:6974).

To calculate precipitation data, I took the ensemble mean of six datasets and assumed that the standard deviation of these datasets represented the uncertainty, as follows: the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) (FUNK *et al.*, 2015); ERA5 Land (MUÑOZ SABATER, 2019); Global Land Data Assimilation System (GLDAS) version 2.1 (RODELL *et al.*, 2004b); Global Precipitation Measurement (GPM) (HUFFMAN *et al.*, 2019); Terra Climate (ABATZOGLOU *et al.*, 2018); and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks–Climate Data Record (PERSIANN-CDR) (ASHOURI *et al.*, 2015). Monthly streamflow data were obtained from measurements provided by the Brazilian Water and Sanitation Agency (ANA) (available at https://www.snirh.gov.br, last accessed on 07 March 2023) and the Argentinian Hydrological Database System (SNIH) (available at https://snih.hidricosargentina.gob.ar/, last accessed on 07 March 2023). Monthly TWS variation (dS/dt) was calculated as the difference of TWS between two time steps, corresponding to the last day of two consecutive hydrological months. Based on recent findings (VISHWAKARMA; DEVARAJU; SNEEUW, 2018), only basins with areas over 100,000 km² were selected. No gap filing methods were employed.

Vapor Pressure Deficit (VPD) and wind speed at 2 m height (u_2) data were obtained from Eq. (16-18):

$$\mathrm{VPD} = 0.6108\left(e^{\frac{17.27\,T_{air}}{T_{air}+237.3}} - e^{\frac{17.27\,T_{dew}}{T_{dew}+237.3}}\right) \quad (16)$$

$$u_{10} = \sqrt{{u_x}^2 + {u_y}^2} \quad (17)$$

$$u_2 = u_{10}\frac{4.87}{log(678-5.42)} \quad (18)$$

where u_{10} is wind speed at 10 m height (m/s).

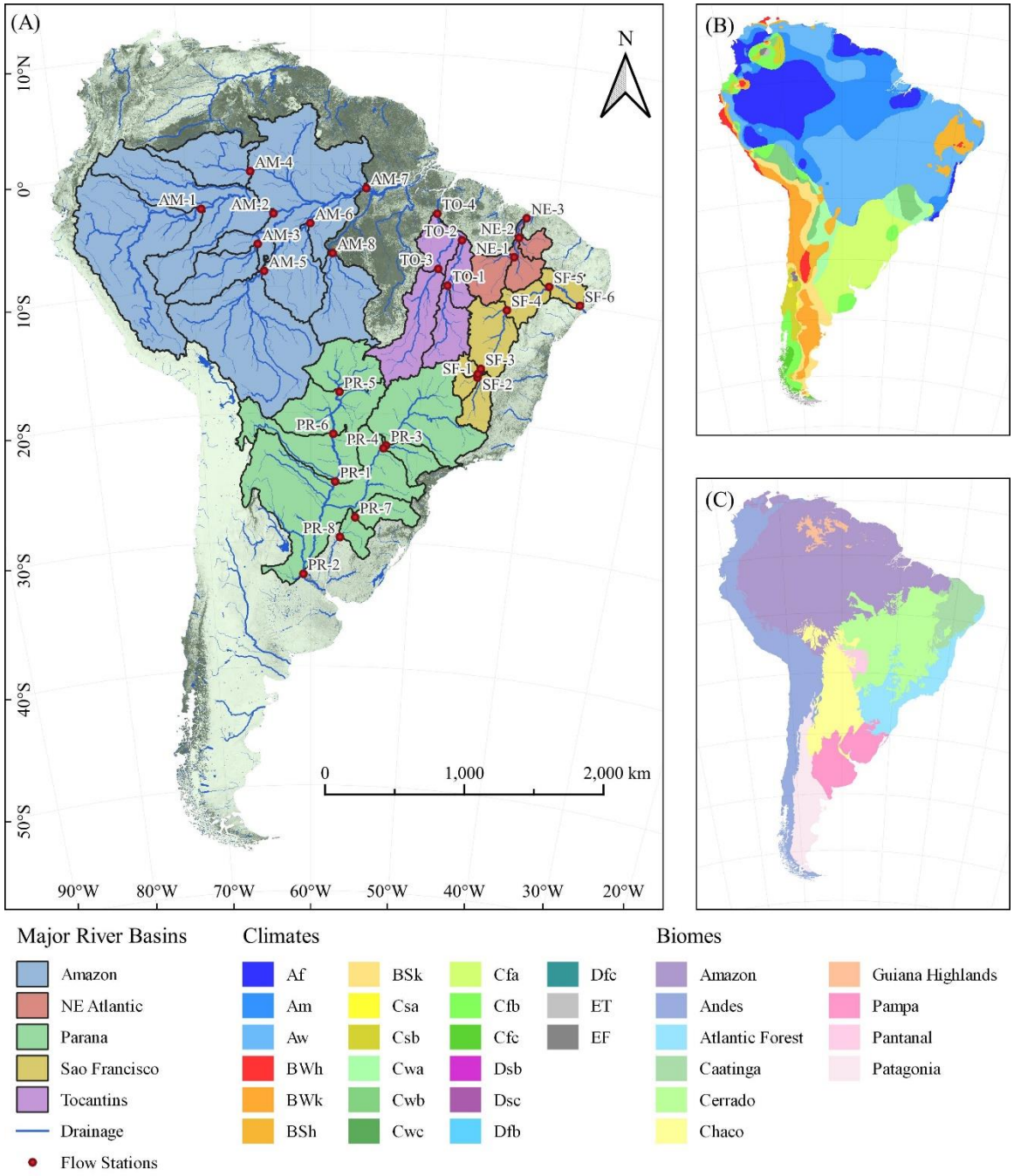


Figure 5.1 – 29 flow stations with their respective drainage areas, grouped into five major river basins (A). The climates (PEEL; FINLAYSON; MCMAHON, 2007) in South America (B). The South American biomes (OLSON *et al.*, 2001) (C).

Table 5.1 – Remote sensing and meteorological variables used in this study.

Dataset	Variable	Symbol	Resolution		Source
			Temporal	Spatial	
GLDAS 2.1	Precipitation	P	3 hours	27,830 m	(RODELL *et al.*, 2004b)
ERA5-Land	Precipitation	P	Monthly	11,132 m	(MUÑOZ SABATER, 2019)
Terra Climate	Precipitation	P	Monthly	4,638 m	(ABATZOGLOU *et al.*, 2018)
CHIRPS	Precipitation	P	Daily	5,566 m	(FUNK *et al.*, 2015)
GPM	Precipitation	P	Monthly	11,132 m	(HUFFMAN *et al.*, 2019)
PERSIANN-CDR	Precipitation	P	Daily	27,830 m	(ASHOURI *et al.*, 2015)
GRACE – Global Mascons	Total Water Storage	TWS	Monthly	55,660 m	(WATKINS *et al.*, 2015)
geeSEBAL-MODIS	Evapotranspiration	ET	8 days	500 m	(COMINI DE ANDRADE *et al.*, 2023)
geeSEBAL-MODIS	Reference Evapotranspiration	ET_r	8 days	500 m	(COMINI DE ANDRADE *et al.*, 2023)
MYD11A2 V6.1	Land Surface Temperature	LST	8 days	1,000 m	(WAN, 2014)
MYD13A1 V6.1	Normalized Difference Vegetation Index	NDVI	16 days	500 m	(DIDAN, 2015b)
MYD13A1 V6.1	Enhanced Vegetation Index	EVI	16 days	500 m	(DIDAN, 2015b)
ERA5-Land Monthly	Air temperature at 2 m height	T_{air}	Monthly	11,132 m	(MUÑOZ SABATER, 2019)
ERA5-Land Monthly	Dew point temperature at 2 m height	T_{dew}	Monthly	11,132 m	(MUÑOZ SABATER, 2019)
ERA5-Land Monthly	u component of wind speed at 10 m height	u_u	Monthly	11,132 m	(MUÑOZ SABATER, 2019)
ERA5-Land Monthly	v component of wind speed at 10 m height	u_v	Monthly	11,132 m	(MUÑOZ SABATER, 2019)
ERA5-Land Hourly	Hourly downwards surface solar radiation	R_S	1 hour	11,132 m	(MUÑOZ SABATER, 2019)

5.2.2 River basin water balance calculation

5.2.2.1 Evapotranspiration from water balance

Many studies evaluated the performance of remotely sensed ET data using WB and hydrological models (DE ANDRADE, 2018; RUHOFF *et al.*, 2022, 2012; SENAY *et al.*, 2017), which is recommended especially for regions with scarce monitoring data (RUHOFF *et al.*, 2022). Therefore, I coupled precipitation and water storage with in situ streamflow data to perform WB calculations for 29 river basins in South America (MOREIRA *et al.*, 2019; RUHOFF *et al.*, 2022). As presented in Figure 5.1, they compose five major South American rivers, including the Amazon, Tocantins, São Francisco, Paraná and Parnaíba. WB-based ET was calculated monthly, between 2000 and 2022 (23 years), according to Eq. (16):

$$ET = P - Q - \frac{dS}{dt} \quad (16)$$

where ET is the monthly evapotranspiration on the basin's surface (mm month^{-1}), P is the monthly precipitation over the basin (mm month^{-1}), Q is the monthly average streamflow measured at the discharge station (mm month^{-1}), and dS/dt is the basin's total water storage variation during a month (mm month^{-1}).

Uncertainty in the ET from WB was computed with Eq. (17) (PASCOLINI-CAMPBELL; REAGER; FISHER, 2020), which is less conservative than the one often used in literature (LONG; LONGUEVERGNE; SCANLON, 2014; RODELL *et al.*, 2004a, 2011; SENAY; BUDDE; VERDIN, 2011; SHEFFIELD *et al.*, 2009):

$$\upsilon_{\overline{ET}} = \sqrt{\upsilon_{\overline{P}}^2 + \upsilon_{\overline{Q}}^2 + \upsilon_{\frac{\overline{dS}}{dt}}^2} \quad (17)$$

where υ is the relative uncertainty of each WB component. The absolute uncertainties of dS/dt and P were assumed to be the average of the standard deviation among the different datasets' estimations. Q uncertainty was assumed as equal to 5% of average monthly streamflow.

Similar to ET, dS/dt was also computed from WB, based on Eq. (18), a rewritten form of Eq. (16):

$$\frac{dS}{dt} = P - Q - ET \quad (18)$$

where ET is the actual evapotranspiration obtained from geeSEBAL-MODIS (mm month^{-1}). Uncertainty of dS/dt estimates were computed according to Eq. (18):

$$\upsilon_{\frac{dS}{dt}_i} = \sqrt{\upsilon_{TWS_{i+1}}^2 + \upsilon_{TWS_i}^2} \quad (18)$$

where $\upsilon_{\frac{ds}{dt}_i}$ is the uncertainty of TWS variation on the i-th month(mm); υ_{TWS_i} and $\upsilon_{TWS_{i+1}}$ are the uncertainties in TWS data of the i-th month and the next one (mm), which are provided with the TWS dataset.

Monthly anomalies were also calculated, based on Eq. (19):

$$x_{anom,i} = x_i - \mu_x \quad (19)$$

where $x_{anom,\,i}$ is the monthly anomaly of the x variable (ET or dS/dt) on the i-th month; x_i is the variable value for the same month; and μ_x is the average value of x for each month of the year.

5.2.3 Long-term trends calculation

Long-term trends were calculated for 11 remote sensing and meteorological variables at pixel scale, for the entire South American continent between July-2002 and July-2022 (20 years). Trend calculations were performed using the Theil-Sen's slope estimator (SEN, 1968) and trend significance analysis was done using the Mann-Kendall trend test at 95% confidence level (KENDALL, 1975; MANN, 1945).

Monthly anomalies and monthly normalized anomalies were also calculated, based on Eq. (20-21), respectively:

$$x_{anom,i} = x_i - \mu_x \tag{20}$$

$$\hat{x}_{anom,i} = \frac{x_i - \mu_x}{\sigma_x} \tag{21}$$

where $x_{anom,\,i}$ is the monthly anomaly of the x variable on the i-th month; $\hat{x}_{anom,\,i}$ is normalized anomaly; x_i is the variable value for the same month; μ_x is the average value of x for each month of the year; and σ_x is the standard deviation of x for each month of the year.

5.3 Results and Discussion

5.3.1 Comparison of dS/dt and ET from Remote Sensing and Water Balance

5.3.1.1 Long-term statistical comparison of dS/dt

Figure 5.2 presents a Budyko framework for the 29 South American basins. The relationship between the dryness and the evaporative indexes follows an asymptotic curve (called Budyko curve) that approaches both the demand and supply limits. The basins' indexes correlate to their location climates, where basins located at humid climates (Af, Am, Cf), such as the Amazon ones, PR-7, and PR-8, yield both lower evaporative and dryness indexes, while ones located at dry (Bw and Bs) climates, such as the NE Atlantic, SF-5, and SF-6, show the highest indexes. The Tocantins, upper São Francisco, and upper Paraná basins lie in between, with $ET_r/P > 1$, as they are located at a tropical climate with a dry winter. Differences between ET from geeSEBAL-MODIS and water balance are small, and both approaches lie within the behavioral limits of the Budyko framework (below the energy and water limits), indicating they are both physically sound methods.

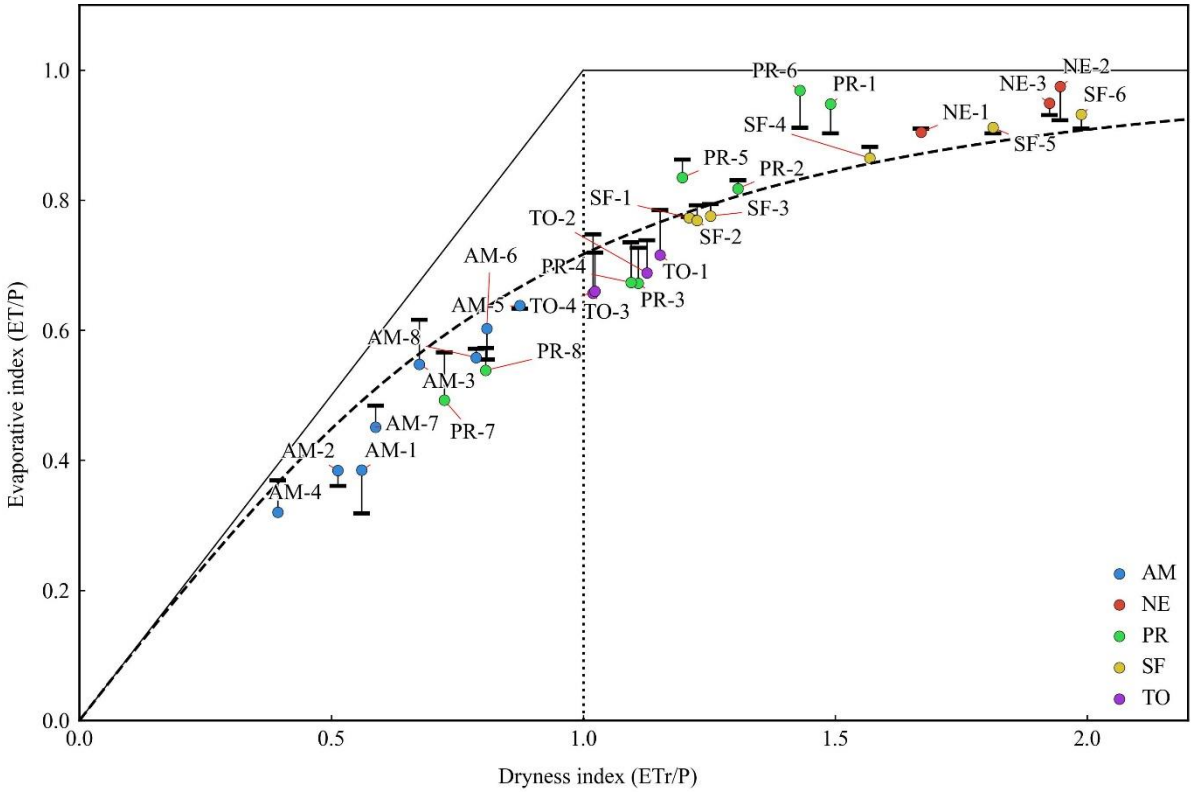


Figure 5.2 – Budyko framework relating long-term (2002-2022) evaporative index (ET/P) with the dryness index (ET_r/P), where $ET_r/P > 1$ refers to water-limited basins, and $ET_r/P < 1$ to energy-limited ones. The circles represent ET obtained using the geeSEBAL-MODIS data, while the horizontal bars represent ET estimated from the river basins water balance. The diagonal line represents the demand/energy limit, where $ET = ET_r$. The horizontal line ET/P = 1 represents the supply/water limit. The dashed curve represents the Budyko curve (BUDYKO; MILLER, 1974).

Figure 5.3 presents the comparison statistics of dS/dt and ET obtained from remote sensing and from river basin WB. Generally, dS/dt data yielded greater agreement than ET, with the values of the slope of linear regression (slope) and determination coefficient (R^2) closer to one. However, the root mean square error (RMSE) and mean absolute error (MAE) values were slightly higher for dS/dt. Bias values indicate that WB dS/dt was overestimated in comparison to remote sensing, and the opposite pattern is observed for the ET comparison. The highest errors for both variables occur at the Tocantins River basin. These results agree with findings from other researches (COMINI DE ANDRADE *et al.*, 2023; RUHOFF *et al.*, 2022). Further investigation is needed to comprehend the reasons for such discrepancies in the Tocantins basin. The larger range of the comparison statistics over the Amazon River basin can be related to the higher uncertainty of the precipitation data as well as to higher magnitude of streamflow values, which also lead to higher uncertainty. Scatter plots with the comparison statistics for each river basin are available at the Supplementary Material (Figures S1 and S2), as well as time series of dS/dt and ET (Figures S3 and S4).

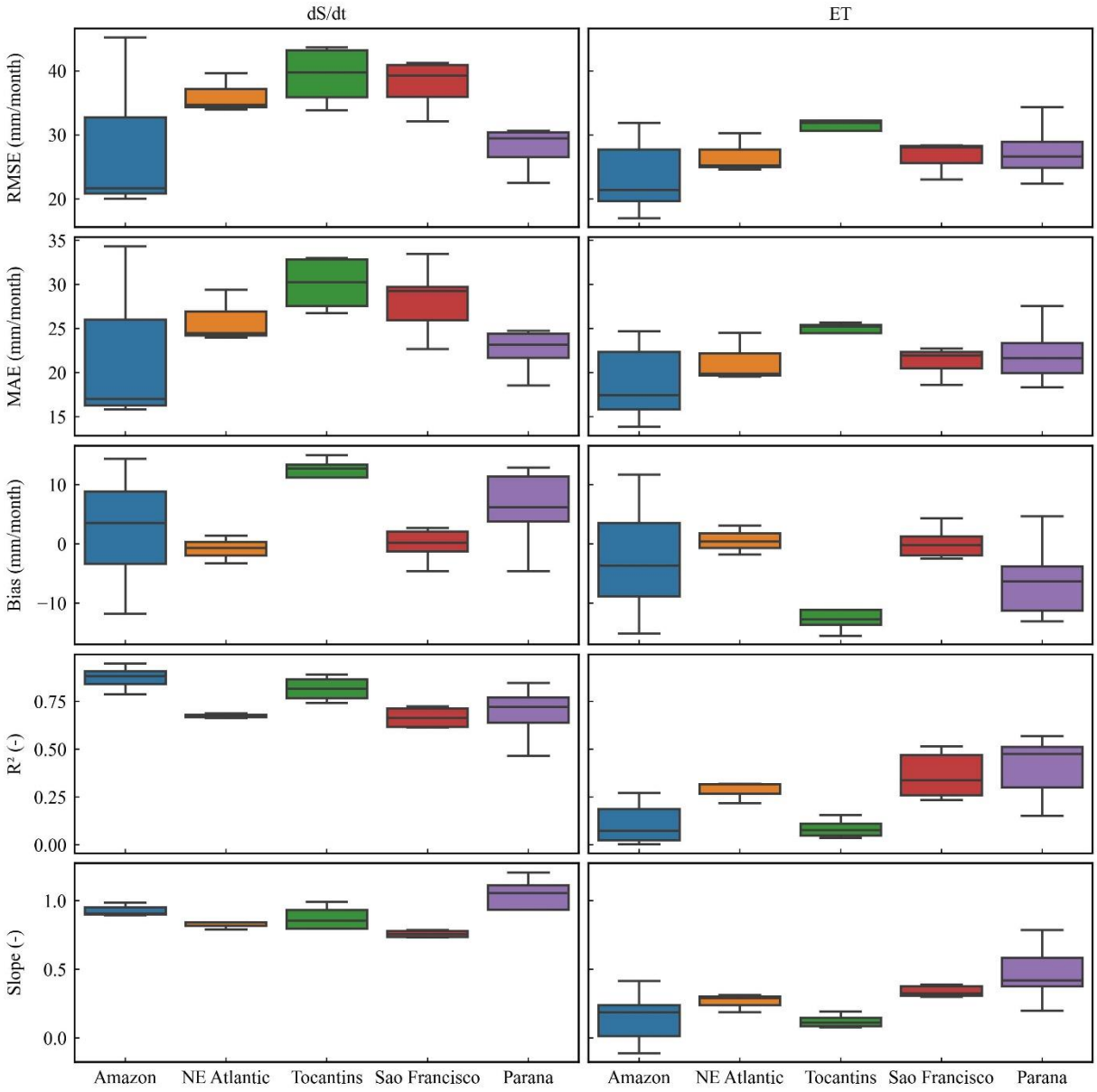


Figure 5.3 – Comparison statistics between water balance and remote sensing dS/dt (left column) and between remote sensing and water balance ET (right column) at 29 river basins in South America, discriminated by major river basin. Remote sensing dS/dt and water balance ET are considered the observed data in the comparisons, while water balance dS/dt and remote sensing ET are considered the estimated data.

5.3.1.2 Seasonal patterns of dS/dt and ET

Figure 5.4 presents the seasonal values of precipitation and discharge, as well as dS/dt obtained from GRACE and from WB, while Figure 5.5 presents seasonal ET_r and ET from geeSEBAL-MODIS and from WB. GRACE and WB present similar seasonal patterns, with a strong correlation to precipitation. The major differences lie in the first trimester (Jan-Feb-Mar), when WB yields higher dS/dt, and in the third trimester (Jul-Aug-Sep), with lower WB estimates. In contrast, WB and geeSEBAL-MODIS present different seasonal patterns,

although the remote sensing ET lies mostly within the WB ET uncertainty range. WB ET shows larger seasonal differences and stronger correlation to water availability (precipitation), with minimum values mostly occurring in September, while geeSEBAL-MODIS correlates more strongly to energy availability (reference ET).

Potential issues about ET estimates from geeSEBAL-MODIS were raised, regarding its sensitivity to the ET_r input, as well as the influence of fire events, which could lead to overestimated ET overall, especially during August and September, due to wrong selection of anchor pixels in the model internal calibration (COMINI DE ANDRADE *et al.*, 2023). On the other hand, higher ET rates on August and September can also be attributed to the presence of riparian vegetation, deep-rooted forests, and irrigated croplands, all of which have the ability to supply high atmospheric demand due to their access to subsurface and groundwater, despite the preceding months of low precipitation. A study in the São Francisco basin used the SSEBop ET model and indicated that forested areas are able to keep ET rates around 100 mm $month^{-1}$ in the driest months, while other surfaces (grassland, savanna and croplands) yield much lower ET rates (DE ANDRADE *et al.*, 2021b). In the same study, a comparison was made between the ET estimates of SSEBop and a rainfall-discharge lumped model. The results showed that both methods produced more similar results to geeSEBAL-MODIS than to WB.

ET_r values can be viewed as the upper limit for ET rates, representing the energy limit beyond which the rate of evapotranspiration cannot exceed. However, ET estimates from WB are sometimes higher than ET_r, specifically during the first trimester. During these months, when there are higher precipitation values, the changes in dS/dt may not correspond with these values, resulting in unrealistic estimates of ET. The sources of errors from WB ET can be attributed to various sources, including errors in the ensemble precipitation estimates, inaccuracies in streamflow gauged measurements, and to errors in GRACE TWS measurements (LEHMANN; VISHWAKARMA; BAMBER, 2022). Assessing the accuracy of streamflow discharge measurements can be challenging. However, ensemble precipitation estimates can be validated and calibrated using in situ measurements. Ultimately, the major limitation of ET estimation based on WB lies in the errors of dS/dt from GRACE, that can be caused by leakage, noise, measurements gaps, and by the coarse spatial resolution of the TWS dataset (WIESE; LANDERER; WATKINS, 2016).

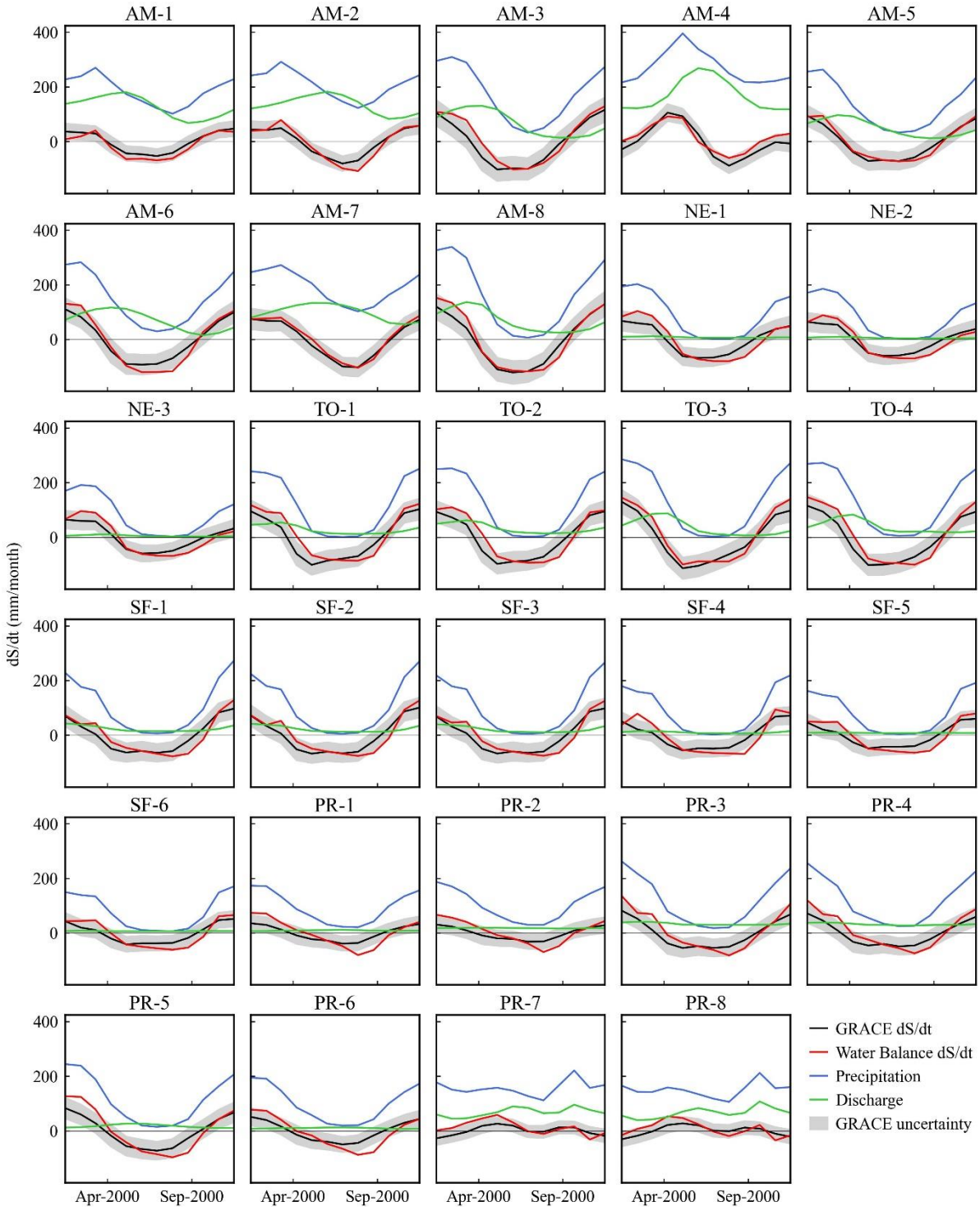


Figure 5.4 – Monthly averages of dS/dt from GRACE and from water balance, as well as precipitation and discharge for the 29 basins. The shaded areas indicate the uncertainty of the GRACE dS/dt.

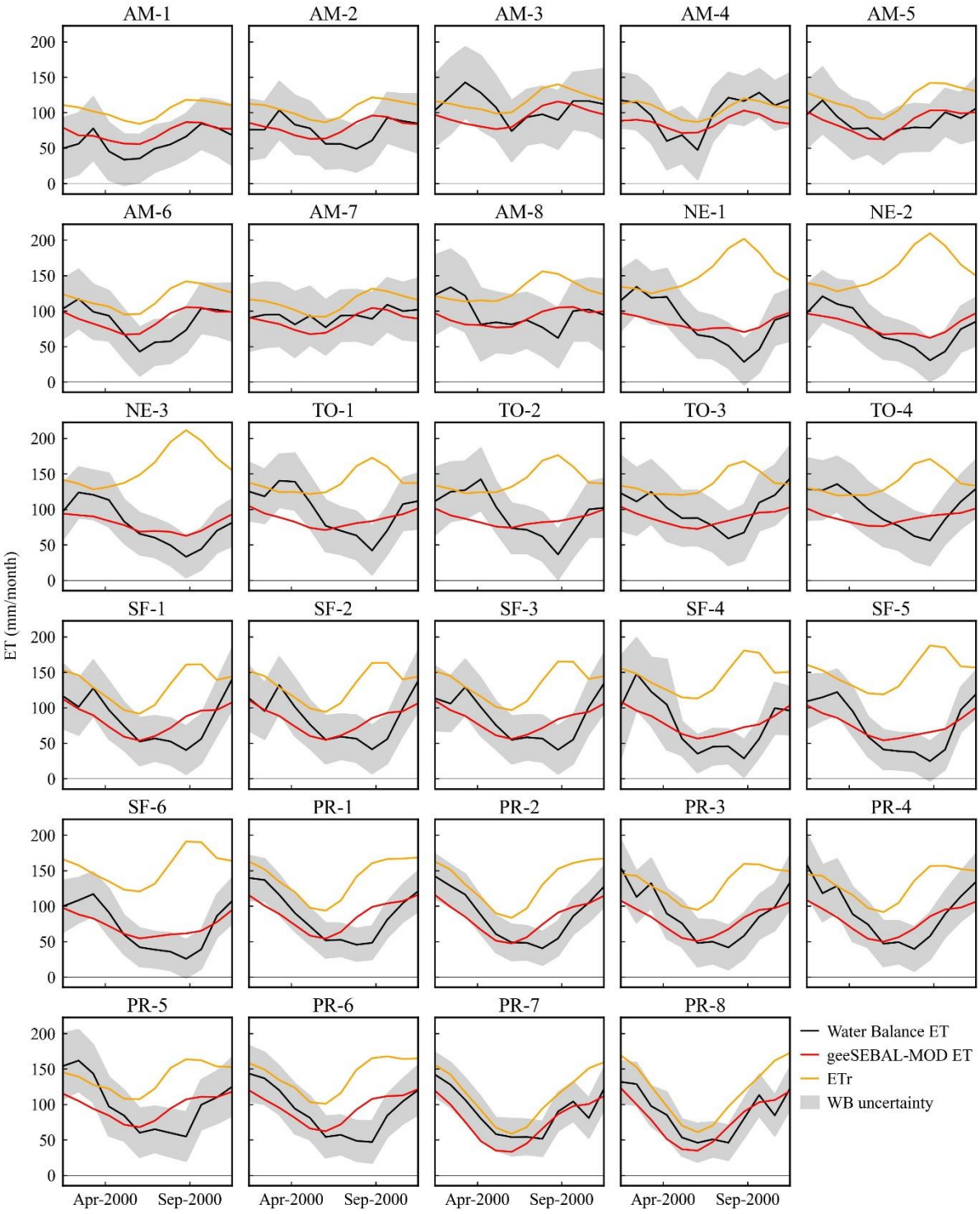


Figure 5.5 – Monthly averages of ET from water balance and from geeSEBAL-MODIS, as well as precipitation and discharge for the 29 basins. The shaded areas indicate the uncertainty of the water balance ET.

Figure 5.6 presents the mean seasonal fluxes, as well as the imbalance of the water balance (imbalance = P – dS/dt – ET – Q). The imbalance values vary from -60 to 60 mm month^{-1}, with most values being sufficiently close to zero that the uncertainty range includes a balance of zero. The positive imbalance values in the first semester are concurrent with the

overestimation of ET by the WB approach, forced by the high values of P, indicating the underestimation of either dS/dt or ET values. The negative values usually happen in the end of the dry season, around September, and indicate the overestimation of either ET, forced by the higher values of ET_r, or dS/dt. The imbalance results agree with previous studies, and reveal that, despite the availability of several sources of data for P, Q, ET, and dS/dt, closure of the water balance remains challenging (LEHMANN; VISHWAKARMA; BAMBER, 2022; MOREIRA *et al.*, 2019; PASCOLINI-CAMPBELL; REAGER; FISHER, 2020; RUHOFF *et al.*, 2022).

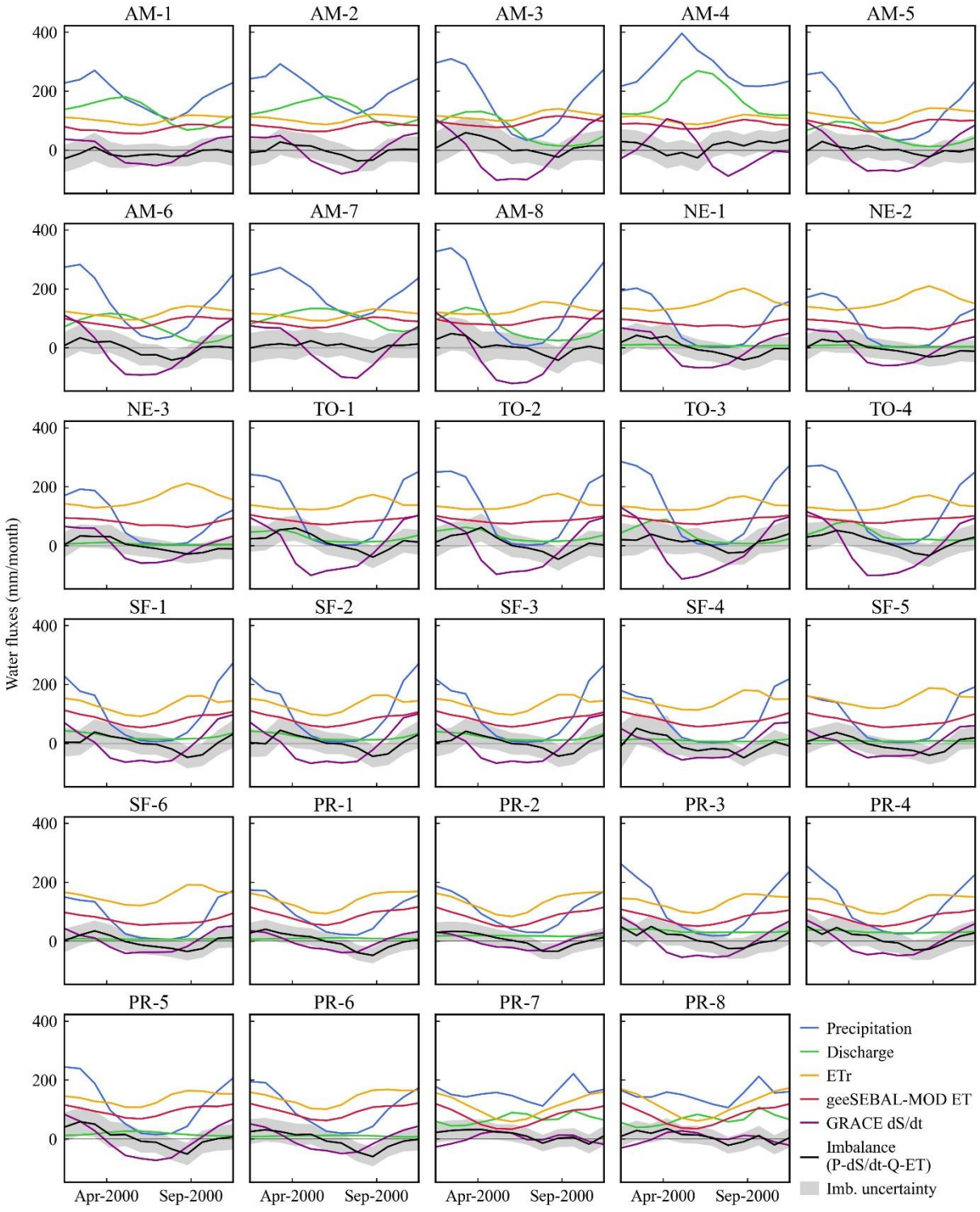


Figure 5.6 – Seasonal water fluxes for South America basins. Shaded areas represent the uncertainty in the imbalance, which was considered the same as the uncertainty in the ET derived from water balance.

5.3.1.3 Anomalies and trends of dS/dt and ET

Figure 5.7 and Figure 5.8 present the monthly anomalies of dS/dt and ET, respectively, from water balance and remote sensing, over the 29 South American river basins. To enhance visualization and clarity, a 12-month double moving average filter was utilized to smooth the series. The temporal phase and signal of the dS/dt anomalies are similar for both sources, which correlate to precipitation anomalies. However, anomalies tend to be smaller than that of WB dS/dt. This phenomenon is more prevalent in Amazon basin, but is also identified in all others, to lesser degree, and indicates that GRACE data has a low sensitivity to precipitation and discharge anomalies, which could be associated to several factors, such as sensor sensitivity, inter basin leakage, or the overcorrection of atmospheric effects (VELICOGNA; WAHR, 2013; WATKINS *et al.*, 2015).

The low amplitude of GRACE dS/dt is compensated by ET in the WB approach, which is generally larger than geeSEBAL-MODIS ET anomalies, as shown in Figure 5.7. Identification of the main drivers and their weight into ET anomalies is a complex task, as they often show opposing patterns, such as energy and water availability, due to the complementary feedback between ET and atmospheric demand (BRUTSAERT; STRICKER, 1979; HUNTINGTON *et al.*, 2011b). In addition, local, regional, and global phenomena can affect the rate of ET. For example, in the Parana basins, precipitation presents a behavior that correlates positively to the Oceanic Niño Index (ONI), which is an index to classify the Pacific Ocean surface temperatures into La Niña (cool) and El Niño (warm) events (NOAA CLIMATE PREDICTION CENTER, 2014), while ET_r correlation to ONI is mostly negative (Figure 5.9). The opposite can be said for NE Atlantic and Tocantins basins. ET from both WB and geeSEBAL-MODIS generally correlates positively with ONI, but with values close to zero. Overall, geeSEBAL-MODIS ET correlates more strongly to ONI over basins of central Brazil, which are subject to Tropical climate with dry winter (Aw climate), and present a savanna landscape.

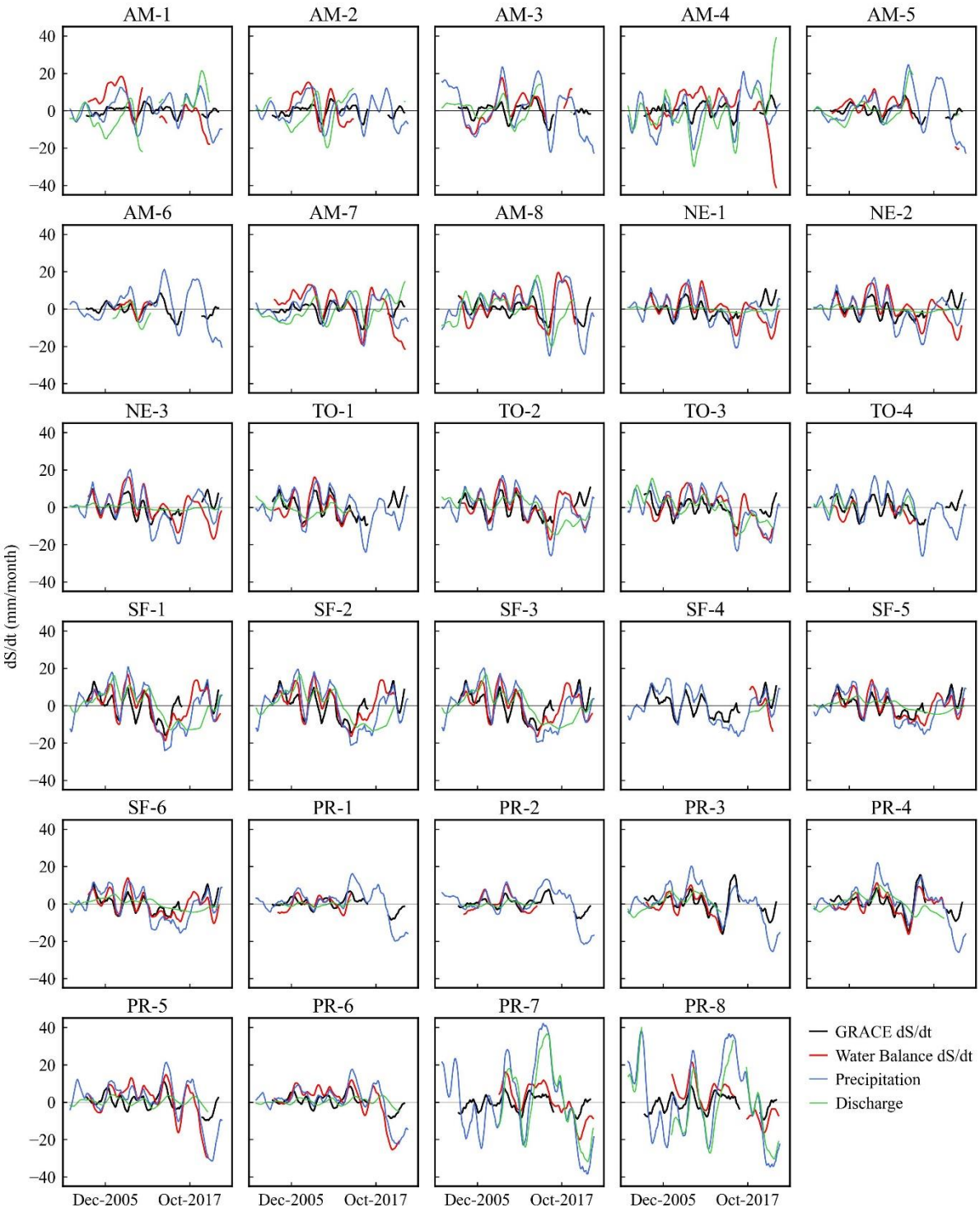


Figure 5.7 – Monthly dS/dt anomalies time series derived from GRACE and from water balance data at 29 river basins. All the lines represent 12-month double moving averages.

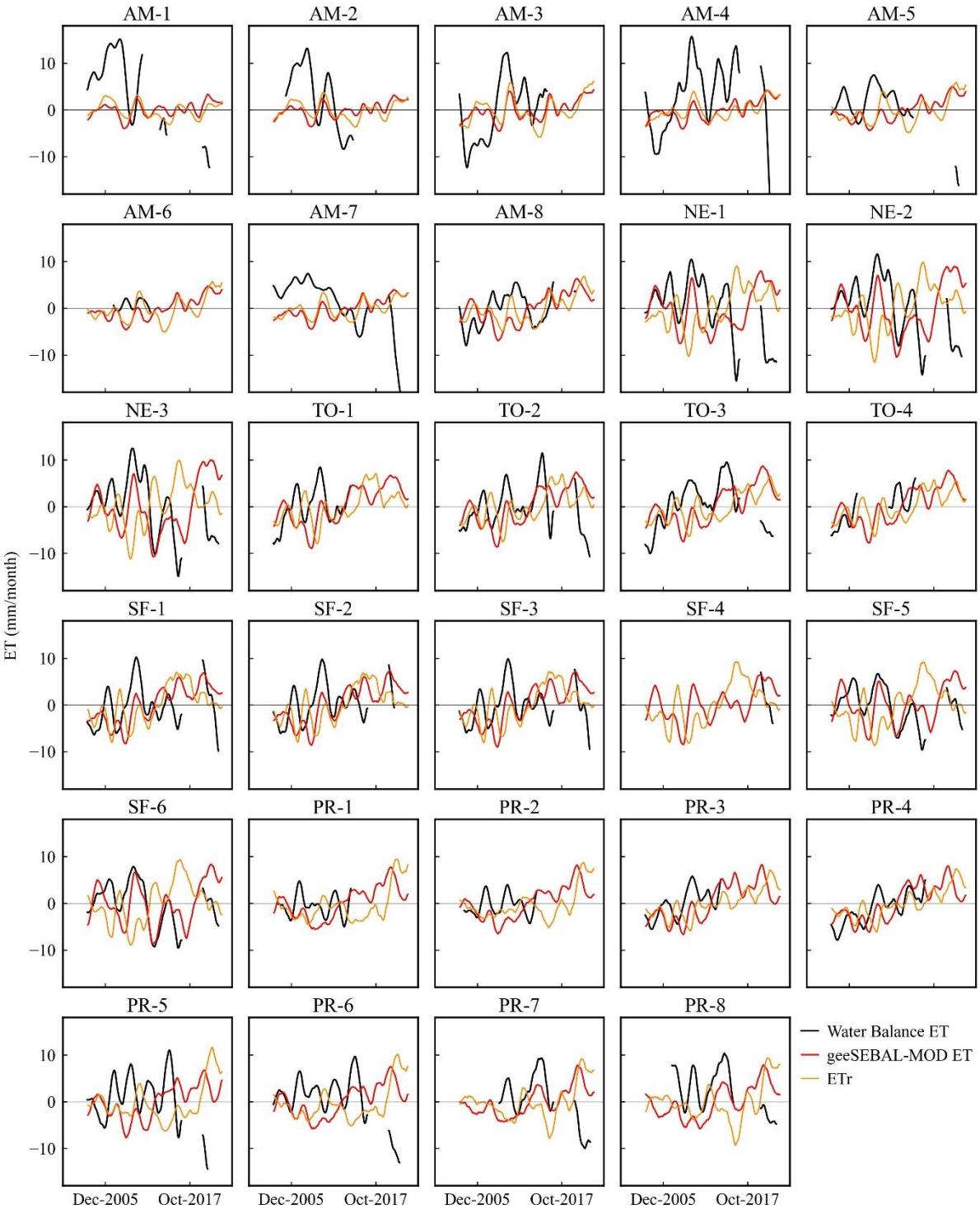


Figure 5.8 – Monthly ET anomalies time series derived from geeSEBAL-MODIS and from water balance data at 29 river basins. All the lines represent 12-month double moving averages.

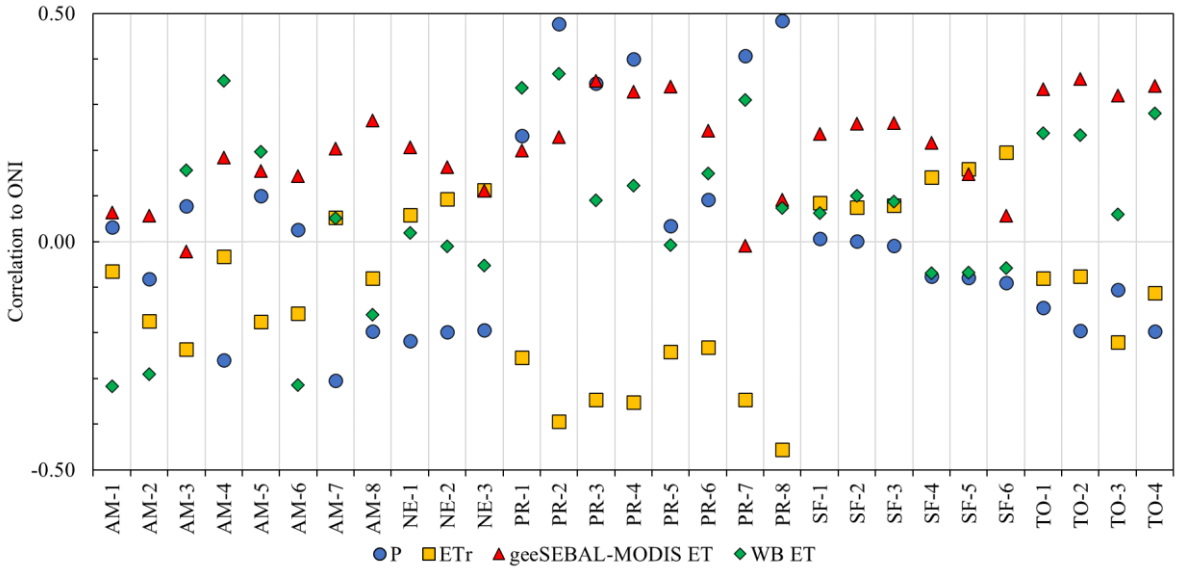


Figure 5.9 – Correlation between the Oceanic Niño Index (ONI) and the monthly anomalies (12-month double moving averages) of precipitation (P), reference ET (ET_r), and actual evapotranspiration (ET) from geeSEBAL-MODIS and WB. Values are calculated for the 2002-2022 period.

Table 5.2 presents the long-term annual trends of dS/dt (from GRACE data and from WB estimates), ET (from WB estimates and from geeSEBAL data), P and Q, derived from monthly data. Significant trends were identified only for ET and Q, in some basins. Most basins yielded decreasing trends of dS/dt and P, with increasing trend of ET_r. At several basins, Q values are determined by the operation of hydropower plants distributed along the developed rivers of each basin, and no correlation was identified between P and Q trends. Nonetheless, negative Q trends are observed for several basins in this study.

Table 5.2 also shows the agreement between the different sources of dS/dt and ET. The agreement was verified if there was no significance to the trends of both, or if both were significant and with the same sign. Otherwise, it was decided that there is no agreement. Out of the 29 river basins, 7 presented disagreements with dS/dt trends from GRACE and from water balance, while the ET trends from geeSEBAL-MODIS and from WB disagreed on all basins but PR-4.

On energy-limited basins, such as the Amazon ones and PR7 and PR-8, since precipitation is usually higher than reference evapotranspiration, changes in ET more associated to anomalies of ET_r, rather than P. Therefore, the increase in ET_r verified in these basins would lead to an increase in ET and, consequently, a decrease in dS/dt. This leads to believe that GRACE data is not sensitive enough to measure water storage variations in humid basins. Furthermore, uncertainties in precipitation, dS/dt from GRACE, and Q add to the total uncertainty in ET

estimates from water balance, that range from 30% to 65%, depending on the basin. Remote sensed ET, is able to provide global land evapotranspiration at good enough spatiotemporal resolution, that can be validated locally using micrometeorological data (BHATTARAI *et al.*, 2019; CHEN *et al.*, 2016; LAIPELT *et al.*, 2021; MELTON *et al.*, 2021; SENAY *et al.*, 2013).

Table 5.2 – The long-term trends of dS/dt, ET, P, ET_r and Q, with significant trends (95% confidence) highlighted in bold. Additionally, the time window with common data is also presented, as well as the trend agreement between the different sources of dS/dt and ET, with disagreeing trends highlighted in bold and colored according to their sign (red for negative and blue for positive trends). Agreement is discriminated by sign and by the combination of sign and significance.

		Trend							**Agreement**			
		dS/dt		**ET**					**Sign**		**Sign & Significance**	
Basin	**Common data time window (years)**	**GRACE**	**WB**	**WB**	**geeSEBAL-MODIS**	**P**	**ET_r**	**Q**	**dS/dt**	**ET**	**dS/dt**	**ET**
AM-1	16	-1.7	**-16.3**	**-15.0**	**2.0**	-4.4	0.2	**8.4**	Yes	**No**	**No**	**No**
AM-2	16	-0.4	**-10.0**	**-9.4**	**2.5**	-3.6	1.1	**4.3**	Yes	**No**	**No**	**No**
AM-3	15	-1.3	-1.7	6.4	**3.3**	**-6.6**	**2.9**	**-2.4**	Yes	Yes	Yes	**No**
AM-4	19	-0.5	**-13.0**	-8.0	**3.8**	3.3	**3.2**	**10.9**	Yes	**No**	**No**	**No**
AM-5	17	-1.5	**-8.2**	-5.1	**3.4**	-0.8	**2.9**	1.8	Yes	**No**	**No**	**No**
AM-6	8	-1.0	-3.3	-9.1	**3.5**	-1.9	**3.1**	**14.6**	Yes	**No**	Yes	**No**
AM-7	19	1.0	**-12.7**	**-10.8**	**3.6**	-1.5	**2.4**	**8.0**	**No**	**No**	**No**	**No**
AM-8	16	-4.7	-3.6	0.2	**4.6**	0.0	**3.9**	**2.5**	Yes	Yes	Yes	**No**
NE-1	18	6.0	-1.2	-7.2	**3.8**	-0.6	**4.7**	**-1.2**	**No**	**No**	Yes	**No**
NE-2	18	6.2	-2.3	-7.6	**4.6**	-0.2	**3.9**	**-0.8**	**No**	**No**	Yes	**No**
NE-3	19	5.7	-1.5	-5.8	**5.5**	0.0	**3.3**	**-0.8**	**No**	**No**	Yes	**No**
TO-1	12	-2.4	-2.2	14.0	**6.1**	-1.4	**3.3**	**-3.8**	Yes	Yes	Yes	**No**
TO-2	19	-1.2	-3.7	1.1	**6.1**	-1.8	**3.4**	**-6.3**	Yes	Yes	Yes	**No**
TO-3	18	-1.9	-1.2	4.1	**6.1**	-3.9	**4.4**	**-6.5**	Yes	Yes	Yes	**No**
TO-4	11	-1.8	-6.5	8.4	**5.9**	-3.4	**3.9**	-1.6	Yes	Yes	Yes	**No**
SF-1	19	-0.7	-3.7	-0.4	**6.2**	-3.9	**3.8**	**-5.7**	Yes	**No**	Yes	**No**
SF-2	18	-0.6	-5.0	-0.2	**6.3**	-3.7	**3.7**	**-5.4**	Yes	**No**	Yes	**No**
SF-3	18	-0.4	-4.0	-0.9	**6.2**	-3.5	**3.7**	**-5.3**	Yes	**No**	Yes	**No**
SF-4	5	-0.3	9.9	5.7	**3.9**	-1.7	**4.0**	**10.1**	**No**	Yes	Yes	**No**
SF-5	18	0.3	-0.4	-2.2	**3.4**	-1.5	**3.4**	**-3.3**	**No**	**No**	Yes	**No**
SF-6	19	0.6	-0.1	-1.9	**3.8**	-1.4	2.6	**-3.0**	**No**	**No**	Yes	**No**
PR-1	12	-1.7	-0.8	1.4	**4.7**	-5.0	**3.1**	1.1	Yes	Yes	Yes	**No**
PR-2	12	-0.6	0.6	1.8	**4.6**	**-6.7**	**3.4**	**2.0**	**No**	Yes	Yes	**No**
PR-3	12	-2.0	-3.5	10.8	**6.1**	**-7.8**	**4.5**	1.4	Yes	Yes	Yes	**No**
PR-4	16	-2.0	-0.8	**10.0**	**6.1**	**-7.3**	**4.6**	**-3.4**	Yes	Yes	Yes	Yes
PR-5	17	-4.1	**-11.1**	-3.5	**5.0**	**-7.9**	**3.3**	-0.5	Yes	**No**	**No**	**No**
PR-6	18	-2.6	**-10.0**	-4.6	**4.5**	**-6.9**	**3.5**	-0.6	Yes	**No**	**No**	**No**
PR-7	13	1.9	-18.0	-7.6	**3.9**	-9.7	**2.8**	-12.8	**No**	**No**	Yes	**No**
PR-8	15	2.8	-7.6	-3.6	**3.7**	-11.3	**2.8**	-7.0	**No**	**No**	Yes	**No**

5.3.2 Spatiotemporal patterns of the trends of various meteorological and remote sensing variables

Table 5.3 presents the long-term annual trends of several variables in South America, calculated between 2002 and 2022. Values are divided between the whole South America continent and areas with statistically significant trends. Over the whole South American continent, increasing trends were identified for ET, ET_r, EVI, LST, R_s, T_{air}, and VPD; TWS presented a decreasing trend; and P and wind yielded no significant trends (at the 95% confidence level) between 2002 and 2022. Areas with significant trends ranged from 20.5% to 83.5% and their spatial means generally yielded stronger trends than the whole continent aggregation.

Table 5.3 – Means, long-term annual trends, and aggregated changes over 20 years (2002-2022) of 11 meteorological and remote sensing variables in South America. Percentage of South America with statistically significant and their value is also presented for each variable.

	Whole South America				**Area with significant trends**	
Variable	**Spatial mean**	**Mean annual trend**	**Change over 20 years (% of mean)**		**Percentage of total area**	**Mean annual trend**
ET (mm year^{-1})	896.63	3.78	75.53	(8.42%)	70.0%	5.07
ET_r (mm year^{-1})	1474.96	2.33	46.54	(3.16%)	65.1%	3.09
EVI (-)	0.37	0.00031	0.00616	(1.68%)	49.2%	0.00051
LST (K)	303.48	0.03	0.54	(0.18%)	67.3%	0.04
TWS (mm) [1]	-6.63	-1.17	-23.31	(-)	83.5%	-1.43
NDVI (-)	0.60	0.00003	0.00055	(0.09%)	46.0%	0.00006
P (mm year-^{1})	1643.64	-2.22	-44.37	(-2.70%)	20.5%	-5.15
R_s (W m^{-2})	205.55	0.17	3.43	(1.67%)	39.0%	0.26
T_{air} (K)	294.80	0.02	0.44	(0.15%)	69.1%	0.03
VPD (kPa)	0.74	0.0032	0.0649	(8.75%)	62.4%	0.0045
Wind (m s^{-1})	0.86	-0.0006	-0.0119	(-1.38%)	24.7%	-0.0005

[1] It is not possible to calculate the percentage of the mean value of TWS as this variable is not an absolute value, but rather an anomaly.

As shown in Table 5.3, the trends in South America vary in space, and there are both areas with significant and insignificant trends. While aggregated trends provide a summary of information, they fail to reveal crucial spatiotemporal details necessary for a comprehensive understanding of the dynamics behind the changes in the variables that drive the regional hydrological cycle. Figure 5.10 expands on this by presenting the long-term trends of the 11 variables over the 10 South American biomes. Figures 5.11 to 5.21 present the annual average, and trends spatial distribution of all 11 variables, along with the anomalies' series (three-month and two-year moving averages) for each biome in South America.

The continent presents an overall increasing trend of ET (Figure 5.11), with a spatial average of 3.78 mm/year, with an aggregated growth of 8.42% of the mean annual ET. Increasing ET is also identified for all South American biomes, with higher magnitude on central and eastern South America. However, non-monotonous patterns are observed in all biomes, with opposing trends between the first and second decades. Results agree with the literature (ANABALÓN; SHARMA, 2017; JAVADIAN *et al.*, 2020; JUNG *et al.*, 2010; LIU *et al.*, 2021), and highlight a growing concern regarding South America water availability, due to decreasing precipitation and increasing atmospheric demand (Figure 5.12), associated with growing demands for water use (BRÊDA *et al.*, 2020; COOK *et al.*, 2020). TWS trends (Figure 5.15) corroborate this concern, with significant TWS decrease over eastern South America, which puts hundreds of millions of people at risk of water scarcity.

Localized ET trends are associated with land cover changes, such as deforestation, that lead to a decrease in ET, and cropland development, which may lead to opposing ET trends depending on the crop, prior vegetation cover, and presence of irrigation. These anomalies are also observed for NDVI (Figure 5.16), EVI (Figure 5.13), and are opposite for LST (Figure 5.14).

Changes in land cover are fairly easy to identify and generally lead to large changes in LST and NDVI values, leading to abrupt variations in ET. However, over natural landscapes there are several factors that may influence changes in ET. In the southern Chaco and Pampa biomes, positive changes in ET are strongly correlated ($r \approx 0.85$) to EVI and NDVI, as well as LST. This phenomenon is also observed to a lesser degree in natural areas of the Cerrado biome, and shows low correlation to precipitation ($r \approx 0.30$), suggesting that the natural vegetation in these areas is going through a greening effect caused by increases in available carbon dioxide in the atmosphere (PIAO *et al.*, 2020; ZHU *et al.*, 2016), although the mechanisms behind CO_2 controls on stomatal are complex and can vary based on vegetation species and drought conditions (HAMIM, 2005; LEAKEY, 2009; TAYLOR *et al.*, 2018). On the other hand, the increase in ET over the primary forests of the Amazon deforestation arc are not related to greening, but rather to higher R_s (Figure 5.18), T_{air} (Figure 5.19) and VPD (Figure 5.20), which lead to an increase in ET_r. In contrast, decreases in ET over the Caatinga biome natural landscapes are related to negative trends of P ($r \approx 0.65$) and NDVI ($r \approx 0.73$).

The ensemble precipitation time series yielded negative trends for most biomes, except for the Amazon and the Guiana Highlands. However, only in the Andes, Pantanal and Patagonia it was considered significant. Figure 5.17 shows that significant P trends are not widespread

across the biomes, which leads to the insignificant trend results for the whole regions. Nonetheless, strong negative signals are observed in southern South America, as well as in parts of Atlantic Forest, Cerrado, Chaco, Amazon, and most of Pantanal. In contrast, a strong positive trend is observed in northeastern Amazon, near to the Atlantic Ocean. These findings are somewhat different from other investigations (DARAMOLA; XU, 2022; MU; JONES, 2022; SHIMIZU; ANOCHI; KAYANO, 2022), however various factor play a role in the trend calculation results. For example, the limits of the studied time window may influence results, especially if there are extreme anomalies close to the start or to the end of the time series. The limits of spatial aggregation are also important to consider. In the Amazon, both negative and positive trends are identified. However, these signals cancel each other, and the trend for the whole Amazon biome is insignificant. Finally, the precipitation sources show considerable differences, which can lead to diverging trends.

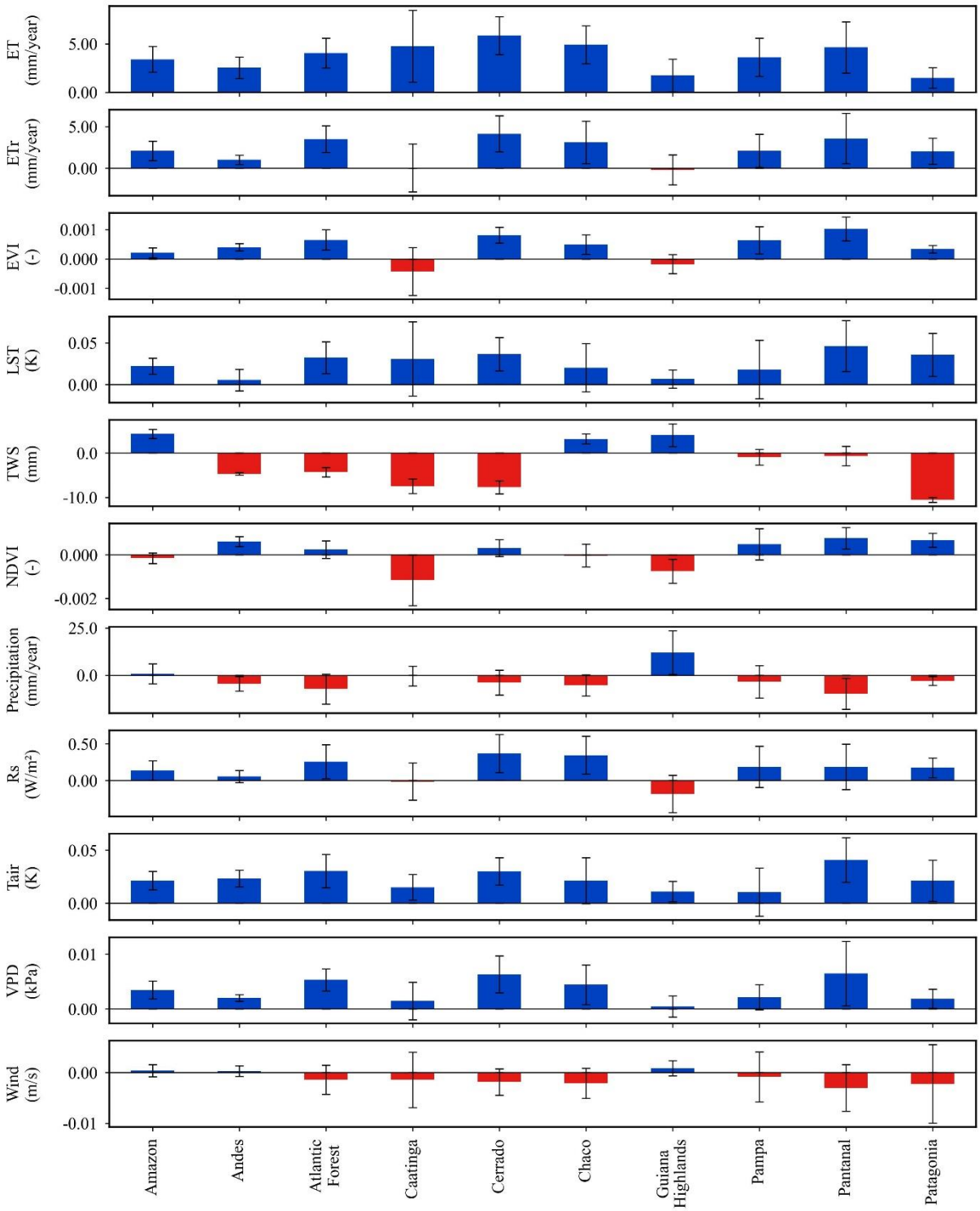


Figure 5.10 – Long-term annual trends (2002-2022) of 11 remote sensing and meteorological variables, for each of the 10 major biomes in South America. Error bars indicate the 95% confidence interval, while bar colors indicate positive (blue) or negative (red) trends. Error bars that cross the null trend line indicate that the trend is not statistically significant at the 95% confidence level.

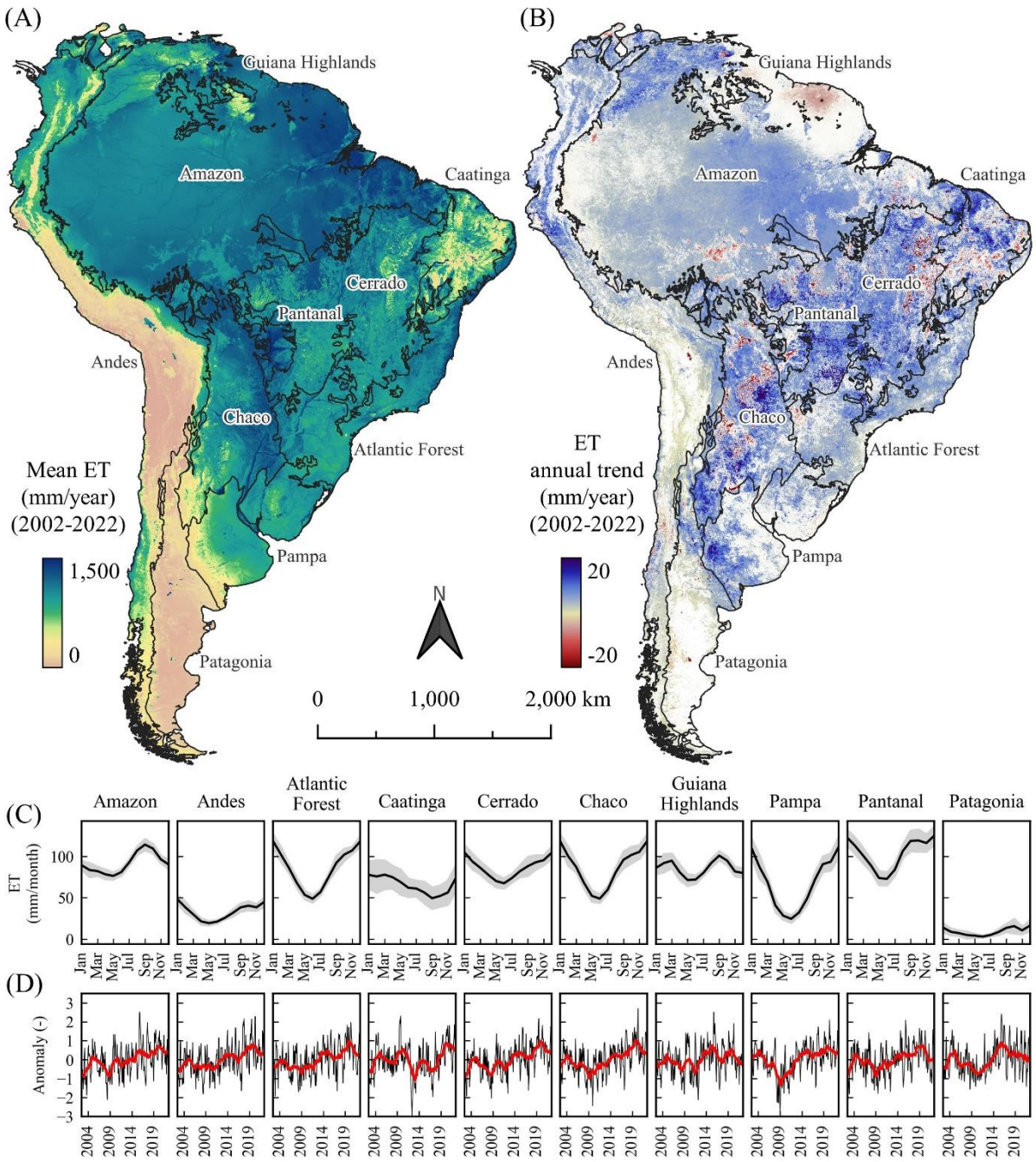


Figure 5.11 – Long-term mean of ET in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

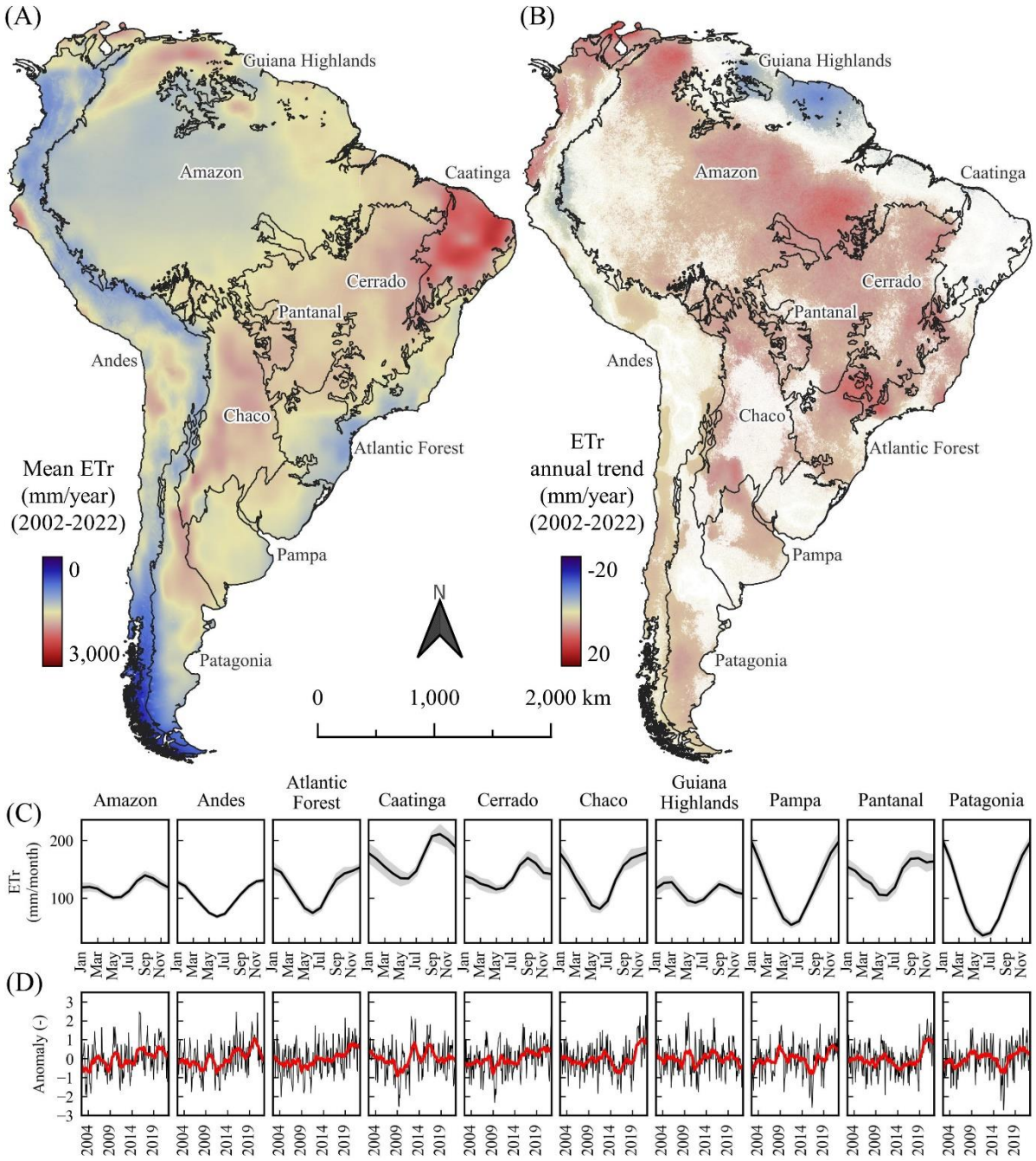


Figure 5.12 – Long-term trend of ET_r in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

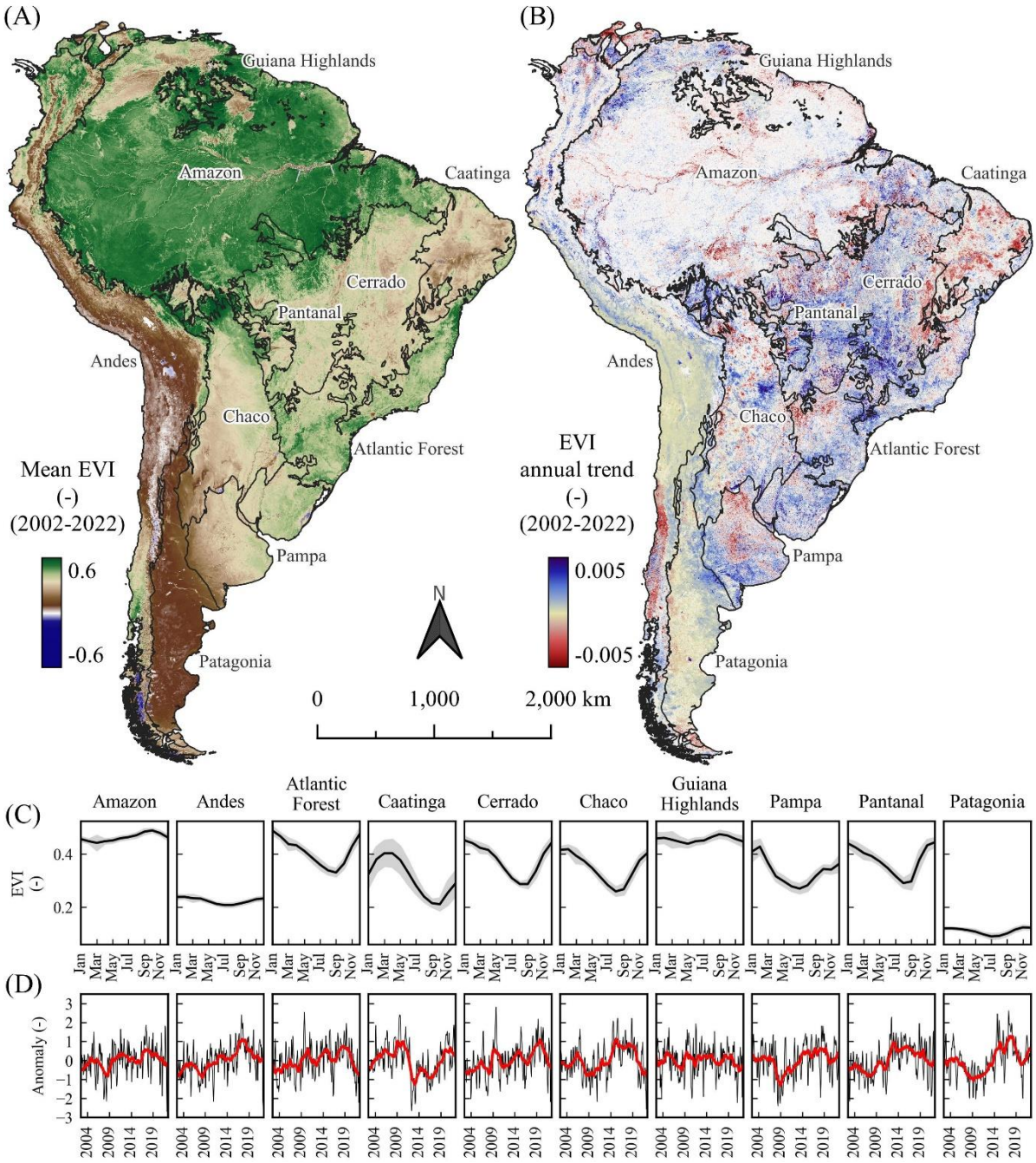


Figure 5.13 – Long-term trend of EVI in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

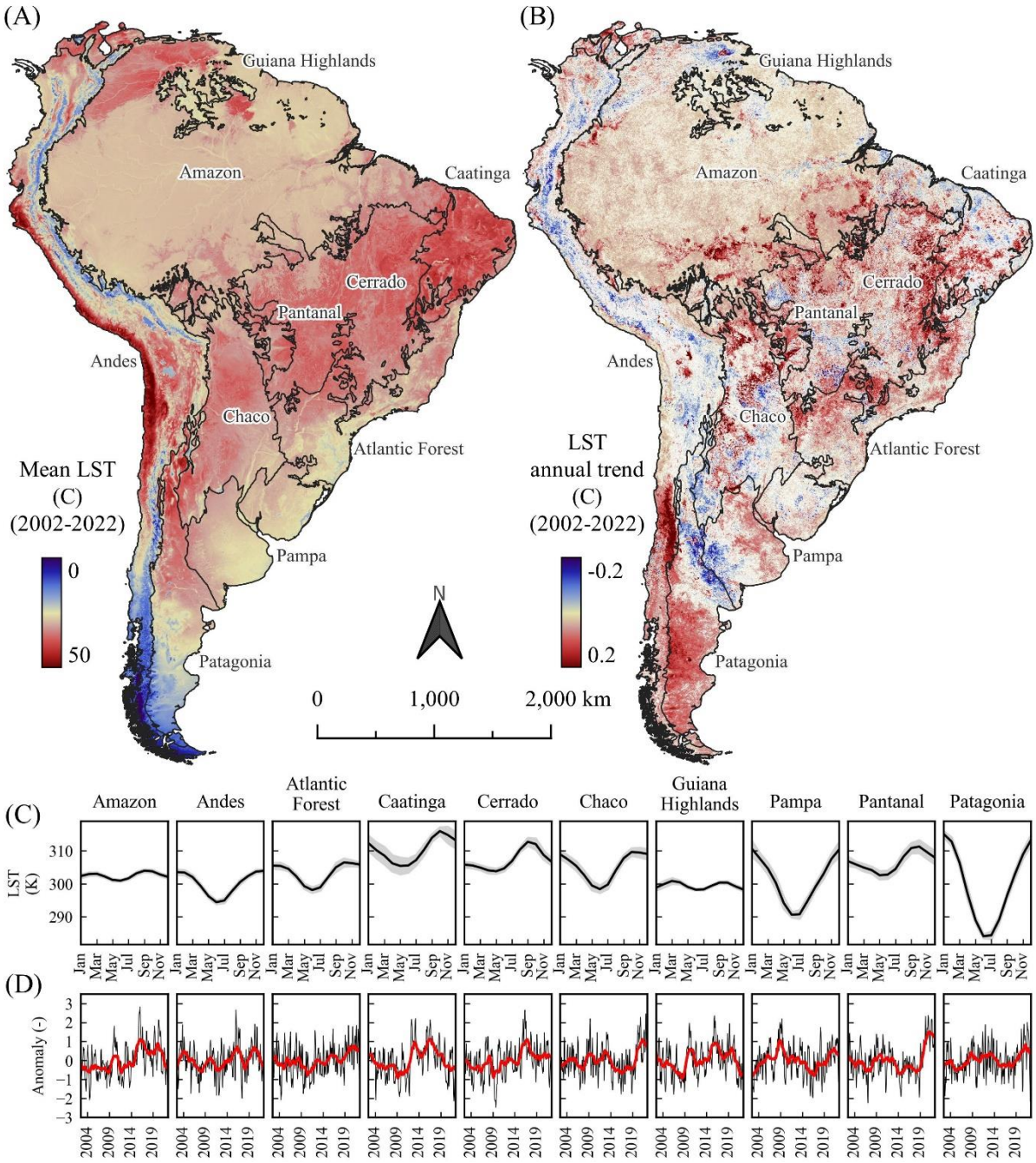


Figure 5.14 – Long-term trend of LST in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

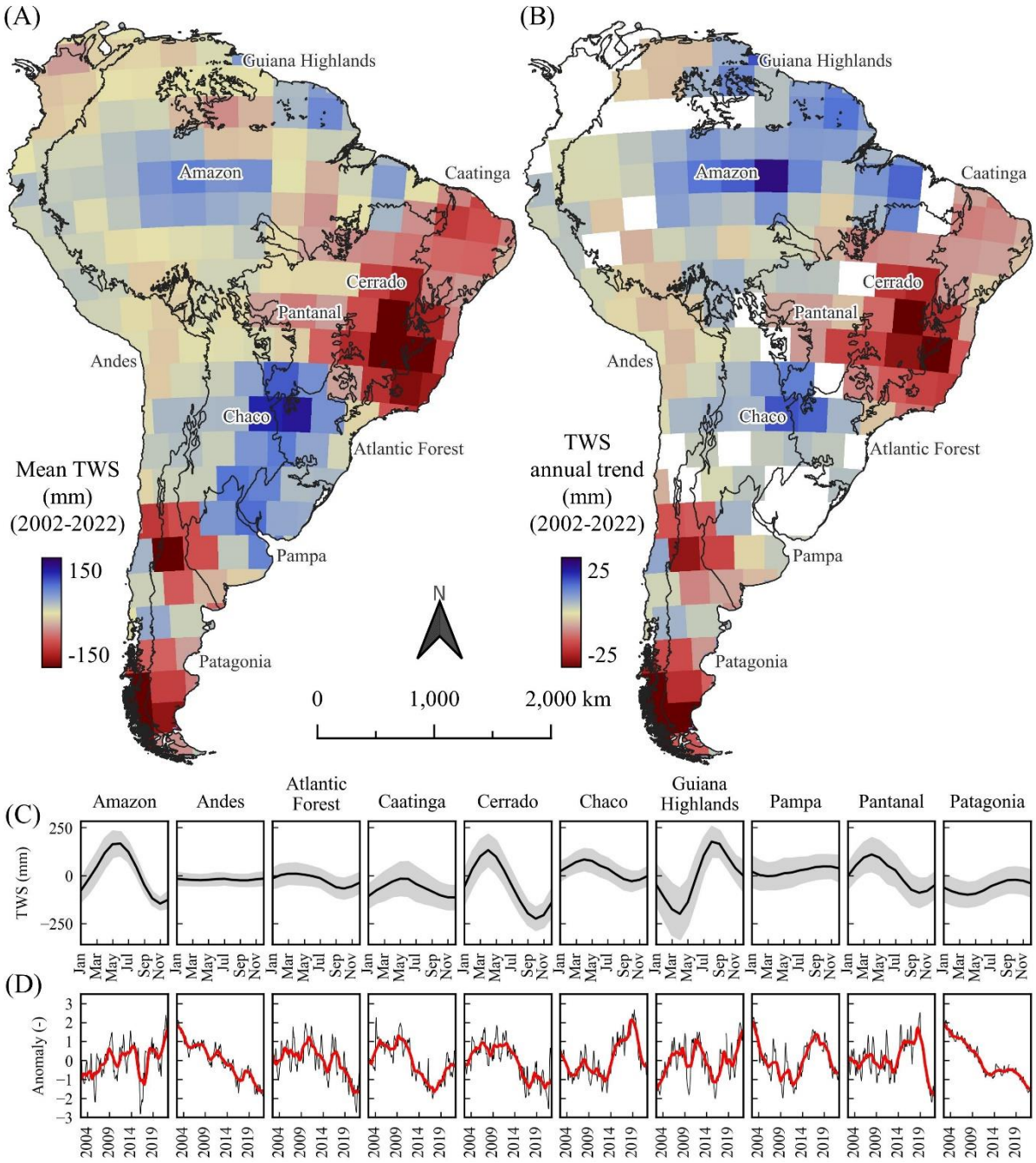


Figure 5.15 – Long-term trend of TWS in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

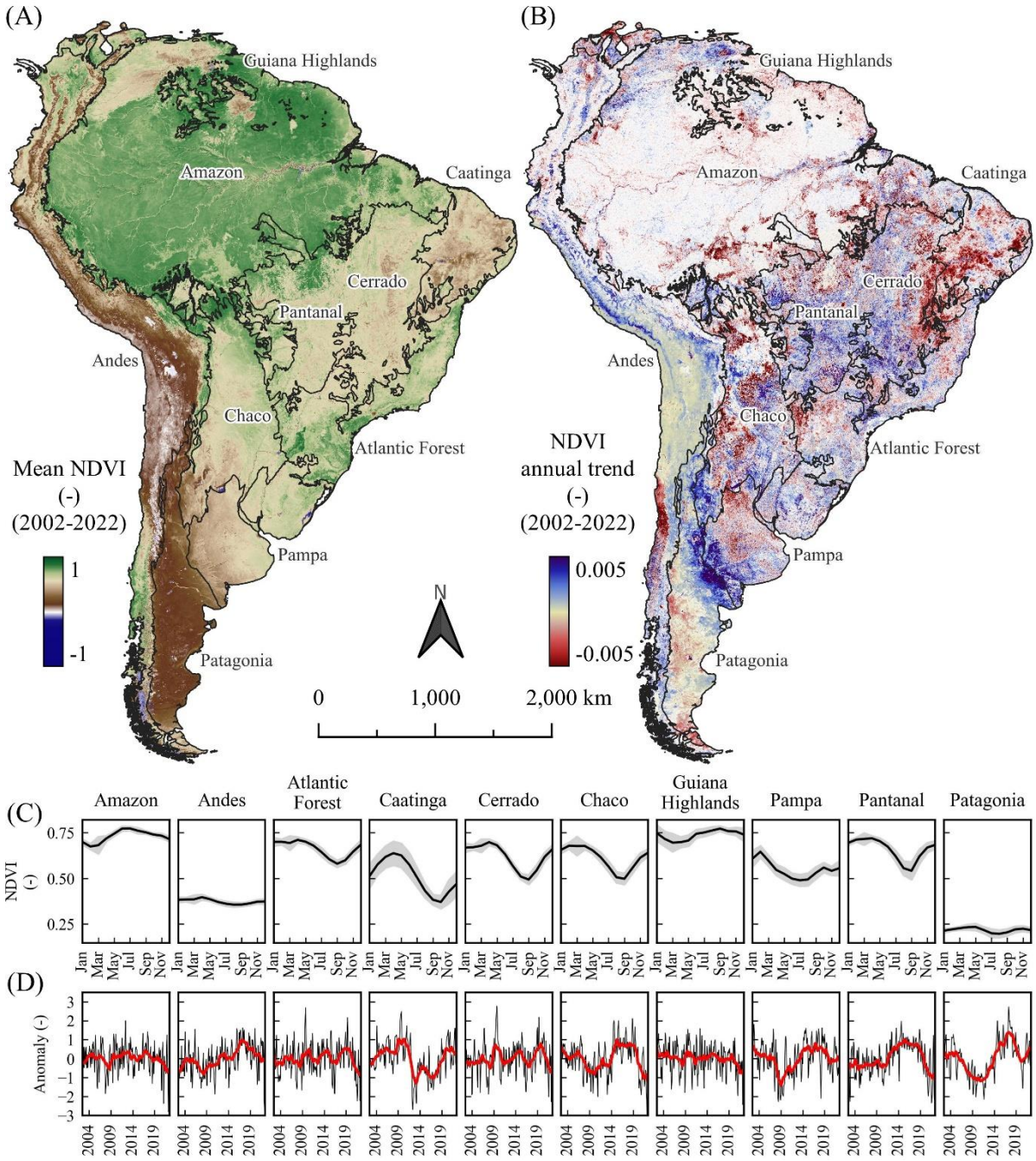


Figure 5.16 – Long-term trend of NDVI in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

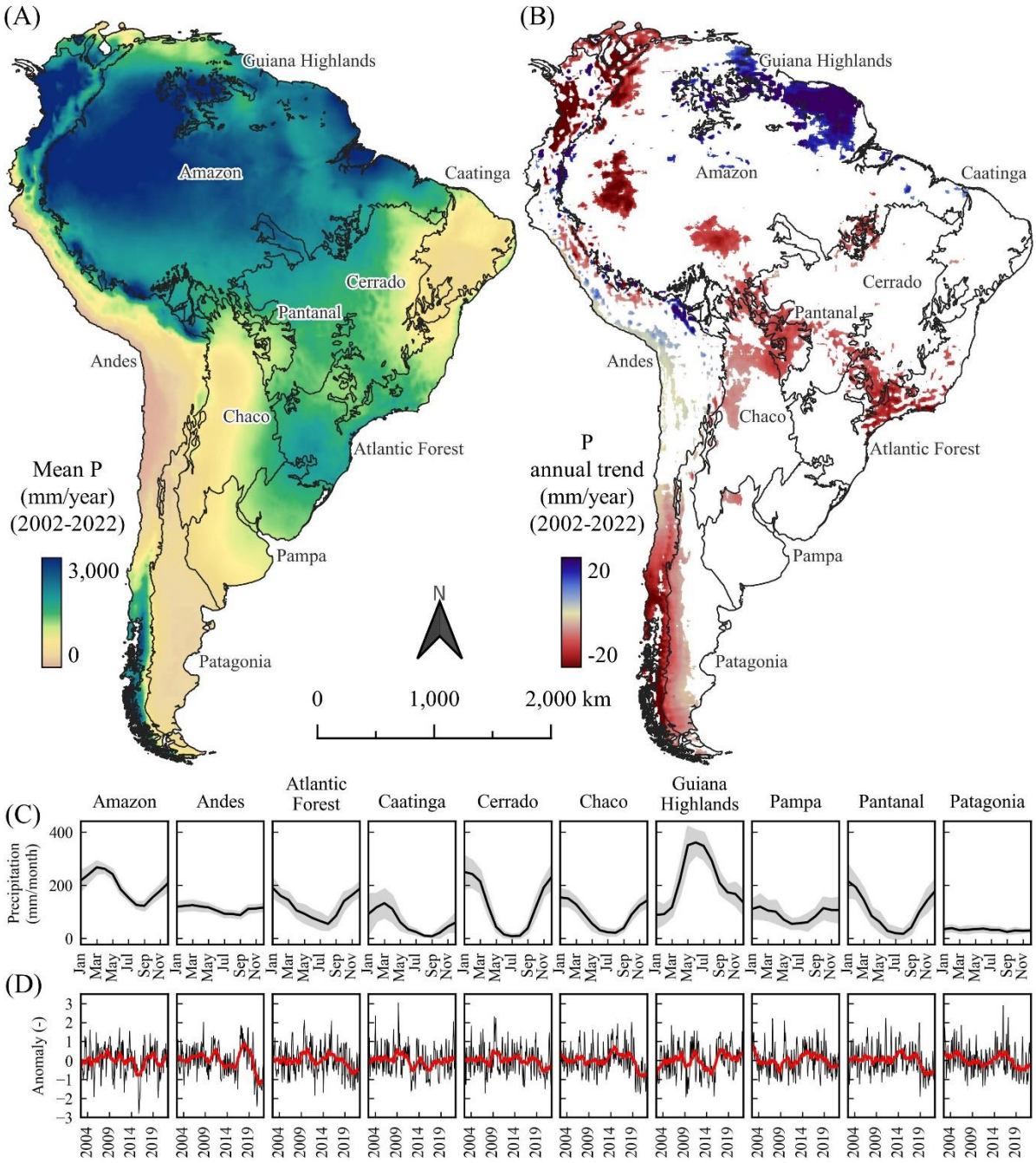


Figure 5.17 – Long-term trend of precipitation in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

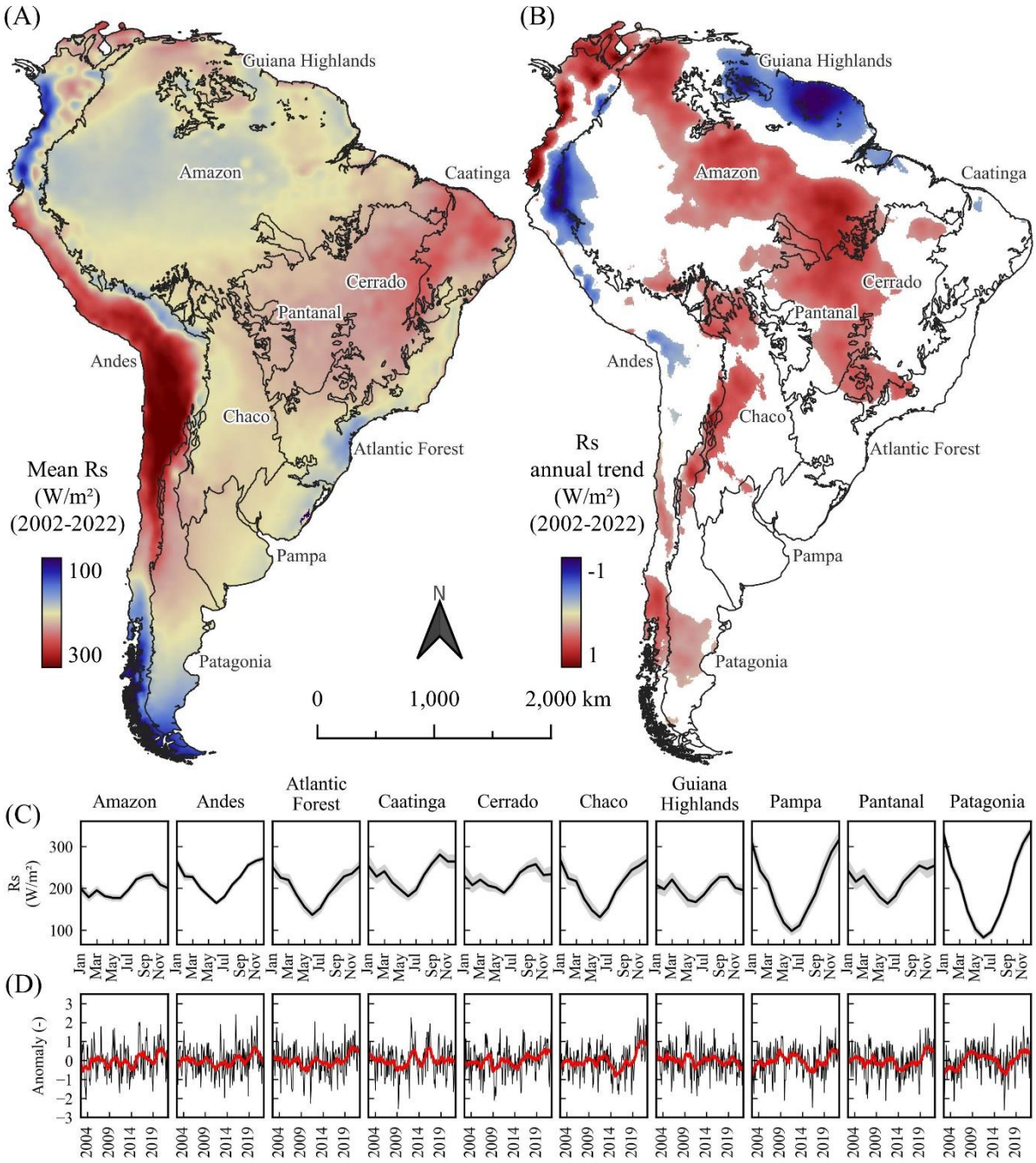


Figure 5.18 – Long-term trend of R_s in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

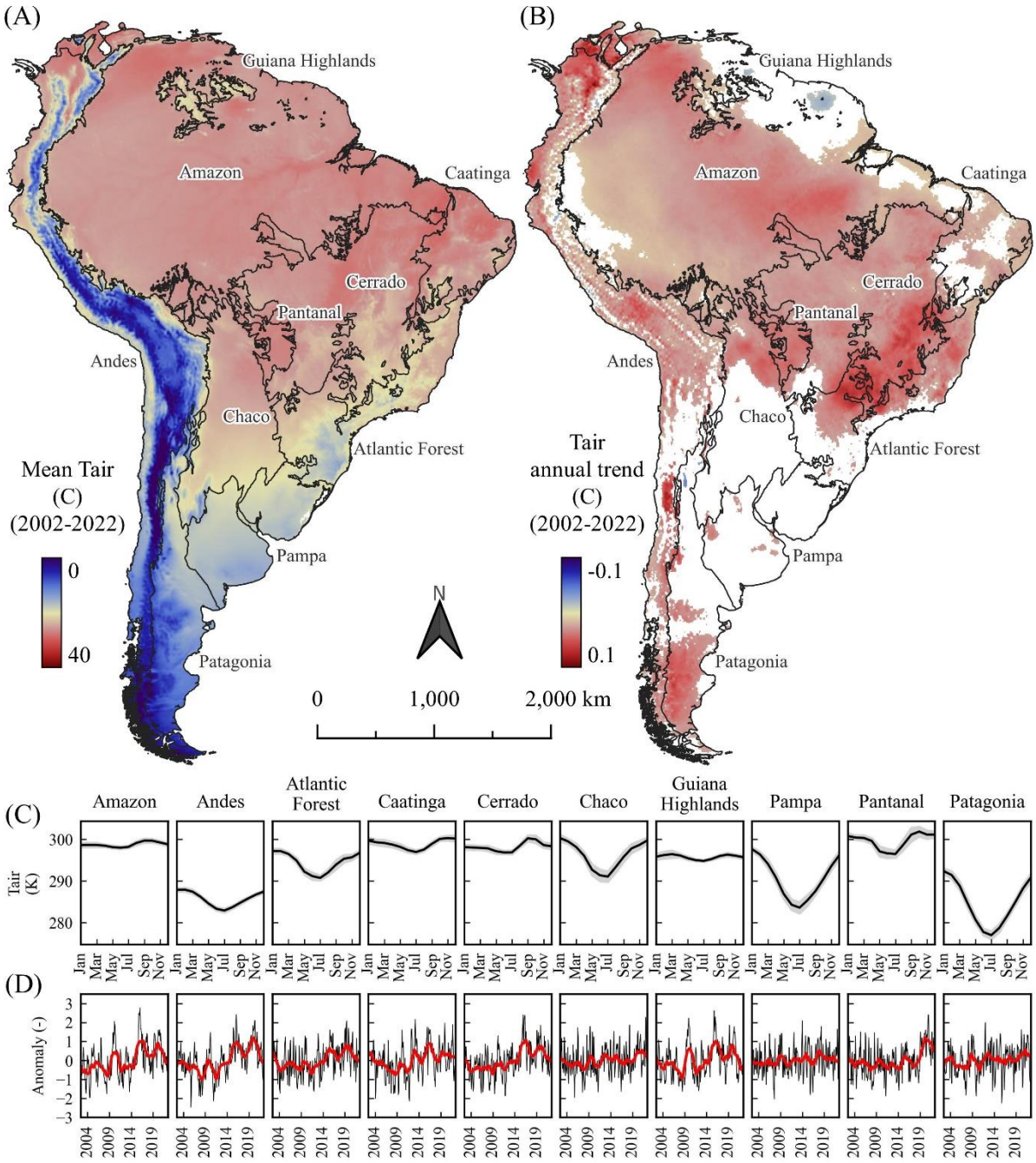


Figure 5.19 – Long-term trend of T_{air} in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

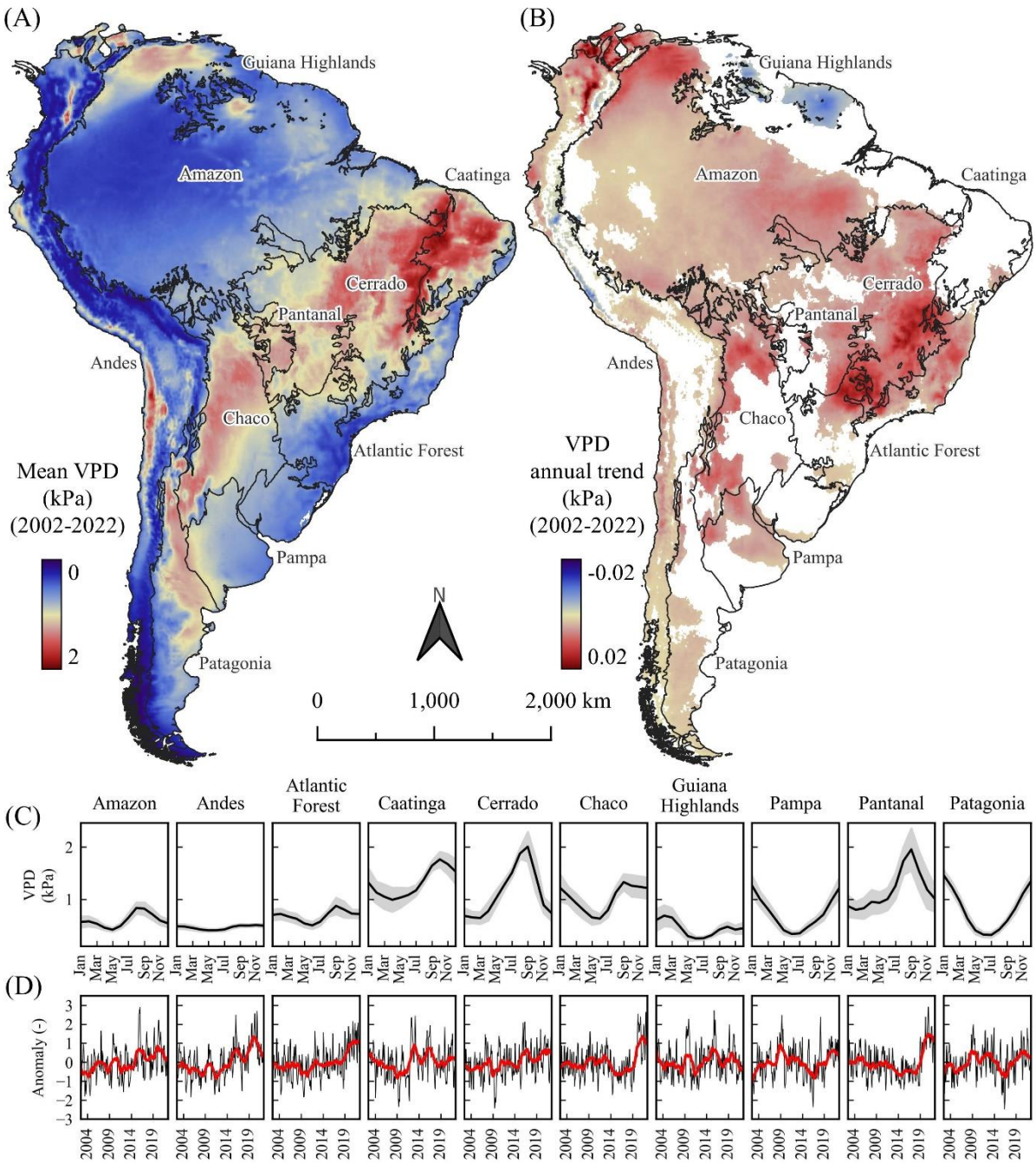


Figure 5.20 – Long-term trend of VPD in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

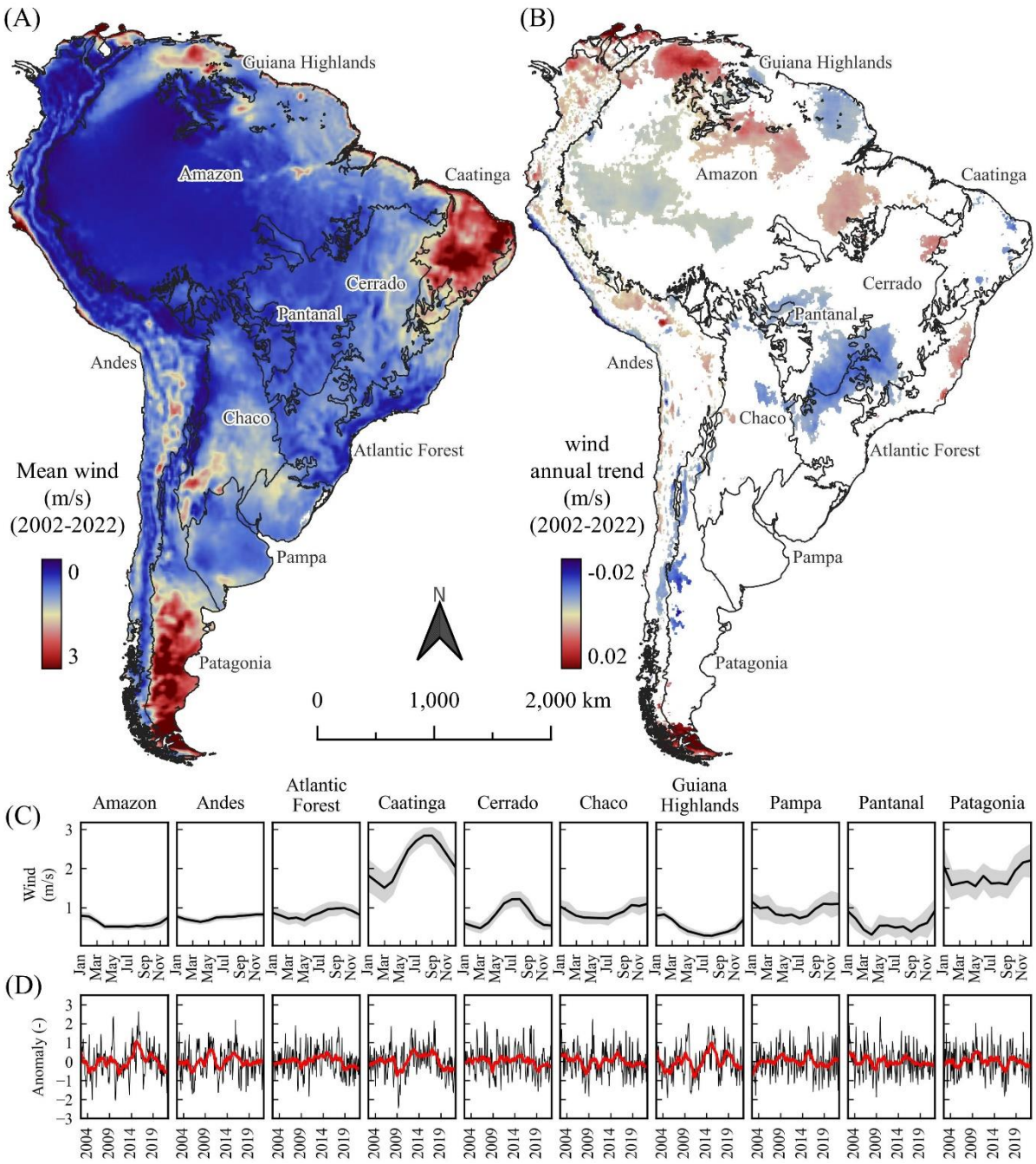


Figure 5.21 – Long-term trend of wind in South America (A); statistically significant long-term trend (B); seasonal average (black lines) and standard deviation (shaded areas) for ten major biomes (C); and time series of monthly anomalies for the same biomes (D), with a rolling average of both three months (black lines) and two years (red lines).

5.4 Conclusions

River basin water balance is a widely used method to estimate the availability of water. However, it has some limitations due to the lack of monitoring of ground water storage (or total water storage) and evapotranspiration, which leads to significant gaps in this approach. This chapter proposes an assessment of remote sensing datasets of TWS (Total Water Storage) and

ET (Evapotranspiration) to close the river basin water budget. The results indicate that while some imbalances occur, they were generally within the uncertainty range of the datasets. Minor systematic biases were observed for TWS during the rainy season and for ET during the dry season. Despite these limitations, TWS and ET remote sensing datasets are considered valuable tools for assessing water availability in river basins.

Another focus of this chapter was the assessment of trends of ET and 10 other remote sensing and meteorological variables in South America. The results indicate an overall increase in ET, driven by atmospheric warming and drying, in association with increased energy availability and, possibly, to carbon dioxide fertilization. On the other hand, precipitation presented no significant trends over the whole continent, while being negative over some areas. These findings point to the straining of water resources availability in a major part of South America, particularly in central and southeastern Brazil, an area of strategic importance for food production.

5.5 Supplementary Material

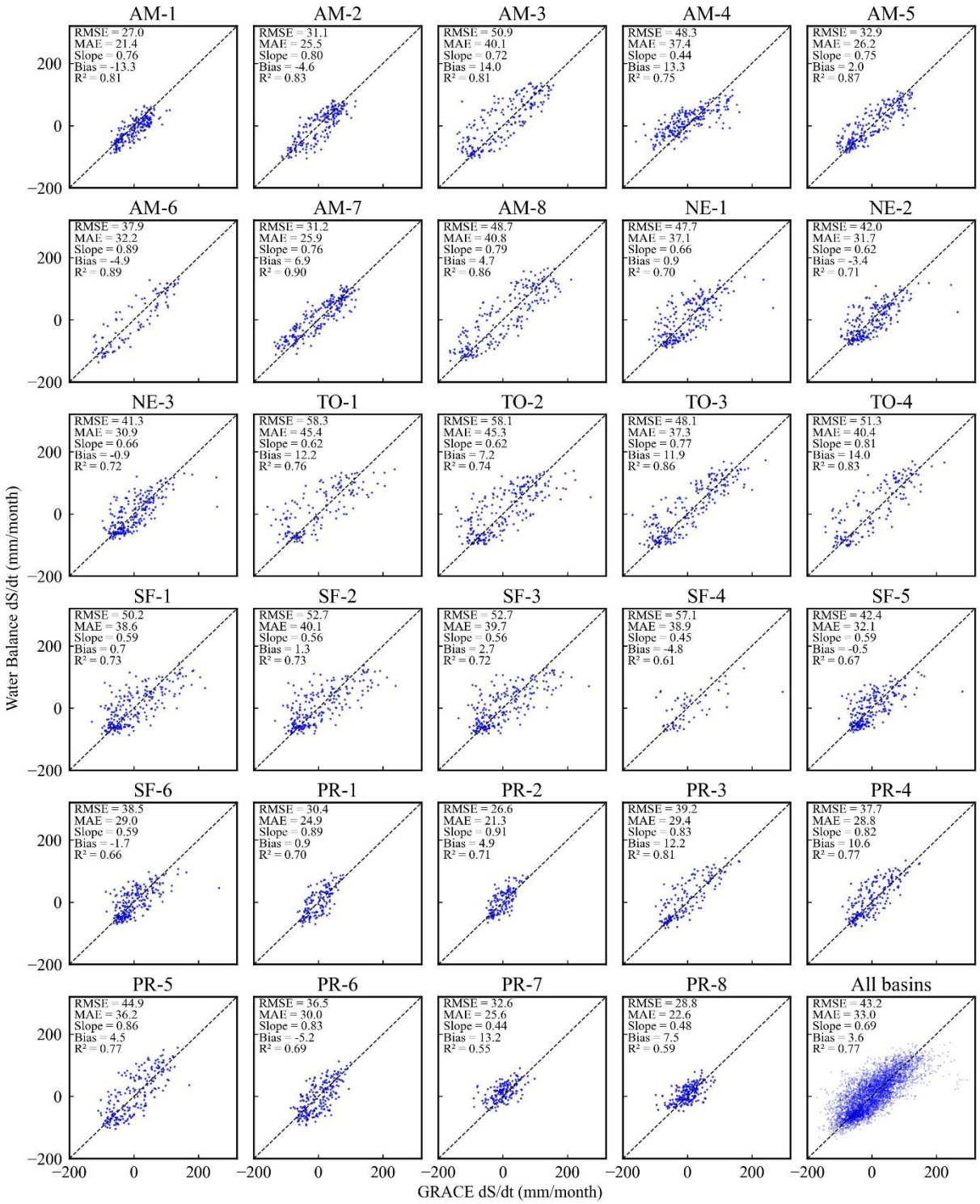


Figure S 1 – Scatterplots comparing dS/dt estimates from GRACE and from water balance at 29 river basins.

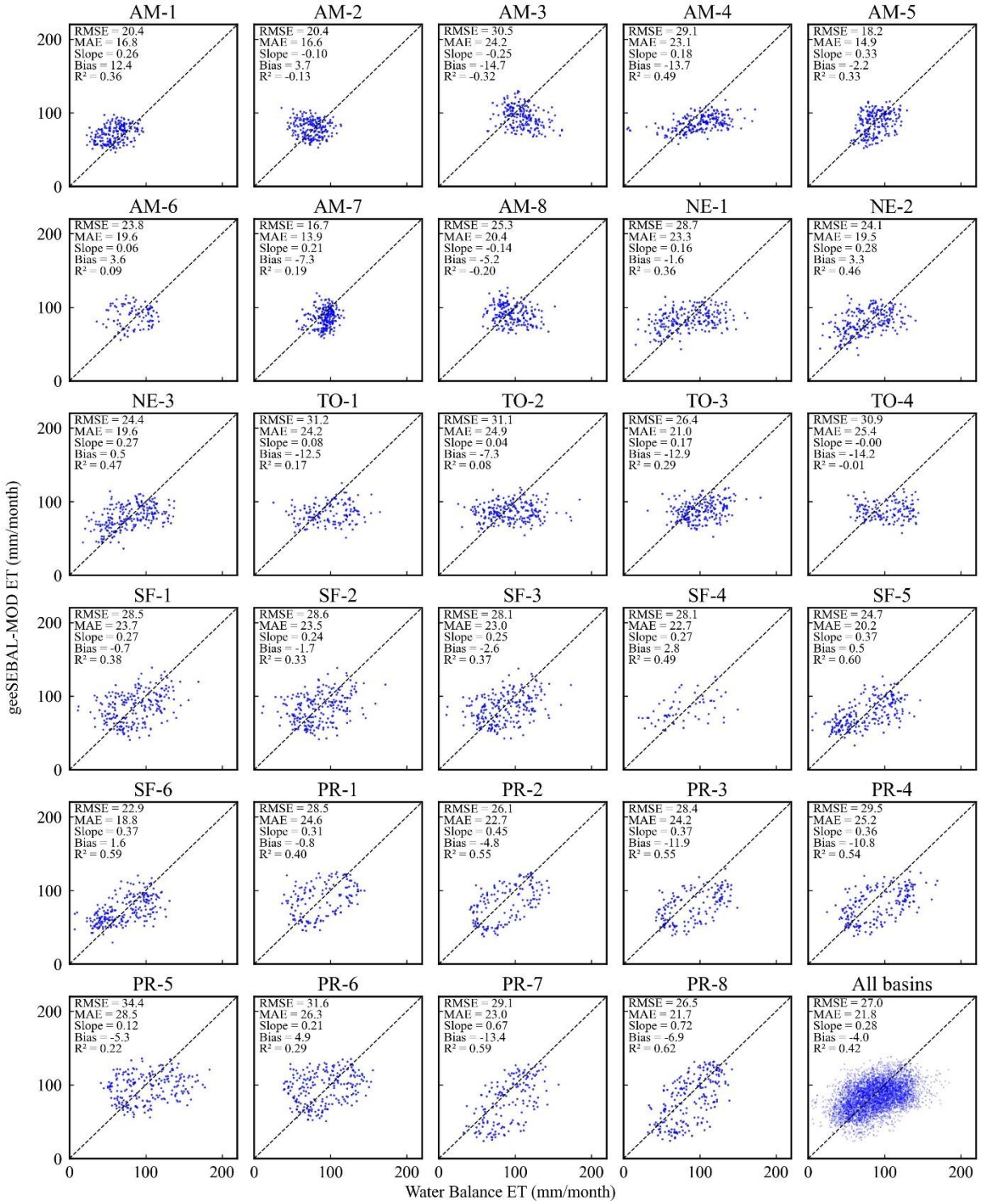


Figure S 2 – Scatterplots comparing ET estimates from geeSEBAL-MODIS and from water balance at 29 river basins.

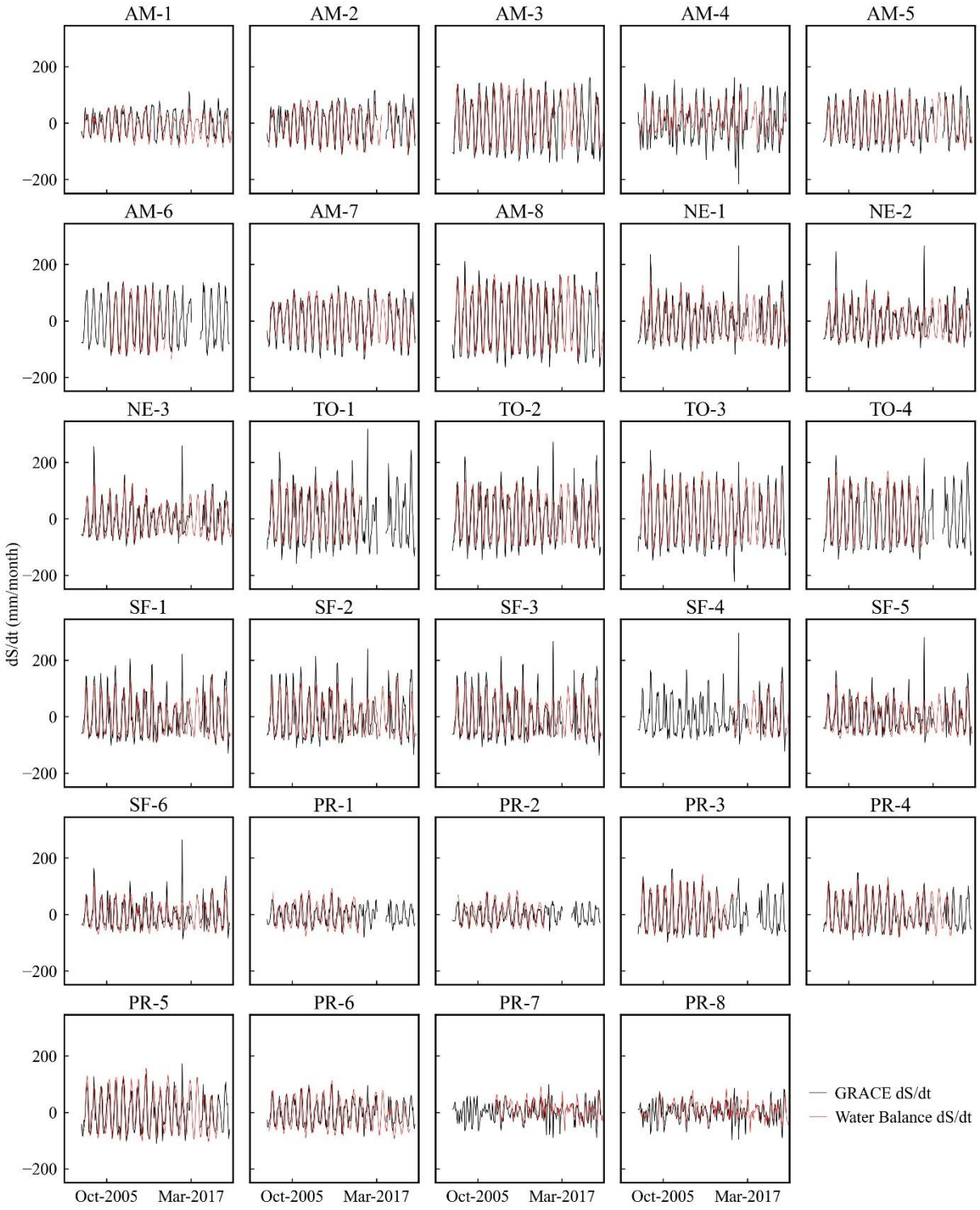


Figure S 3 – Time series of dS/dt estimates from GRACE and from water balance at 29 river basins.

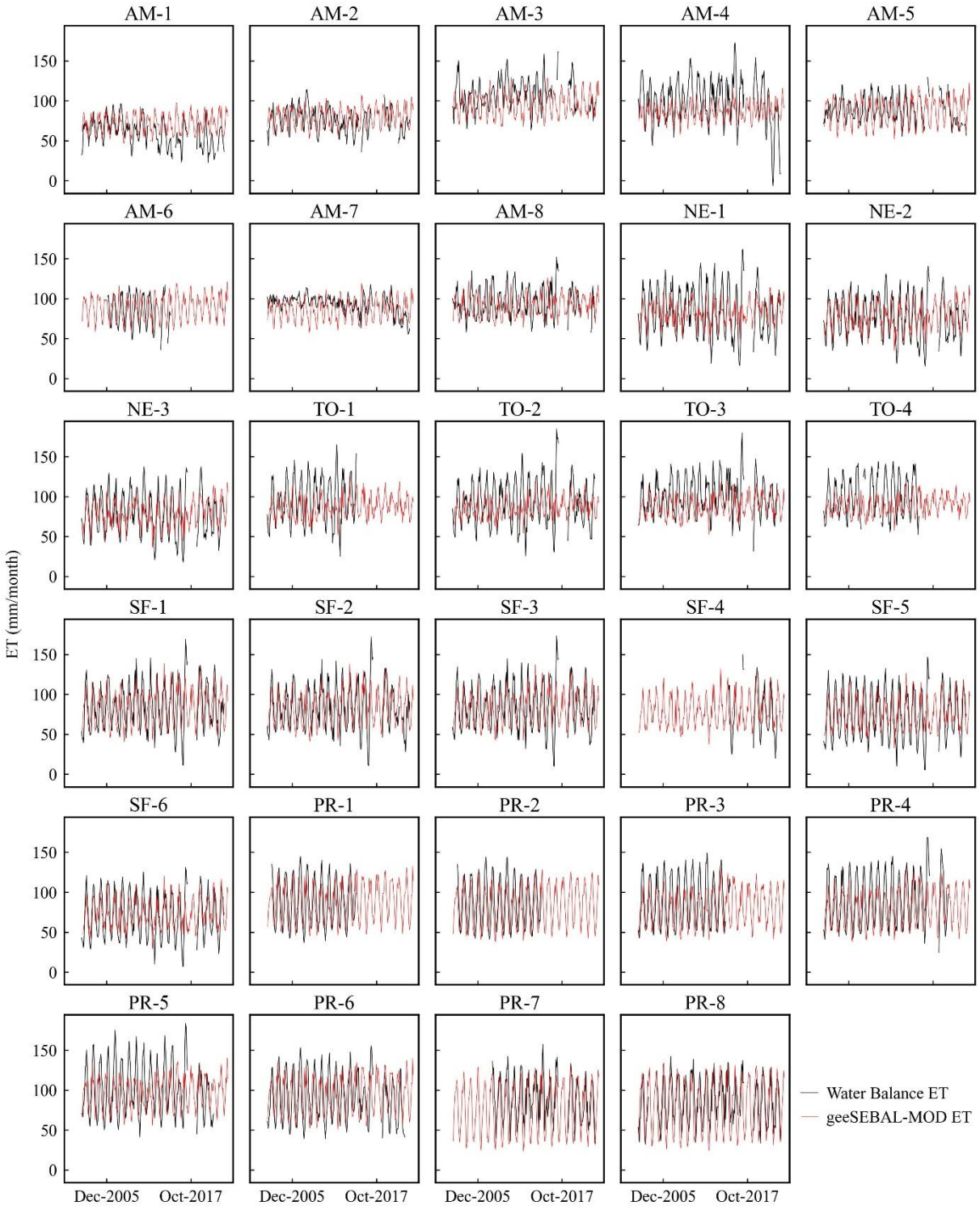


Figure S 4 – Time series of ET estimates from geeSEBAL-MODIS and from water balance at 29 river basins.

Chapter 6: Continuity and Expansion of The OpenET Project

6.1 Introduction

This chapter presents the OpenET framework (MELTON *et al.*, 2021) for intercomparison of remote sensing evapotranspiration (ET) data with in situ flux tower ET data, as well as its application in South America. The current version of OpenET uses six models to produce the ensemble ET data, namely Atmosphere-Land Exchange Inverse/Disaggregation of the Atmosphere-Land Exchange Inverse (ALEXI/DisALEXI) (ANDERSON *et al.*, 2018; ANDERSON; KUSTAS; NORMAN, 2007), Google Earth Engine implementation of the Mapping Evapotranspiration at High Resolution with Internalized Calibration (eeMETRIC) (ALLEN *et al.*, 2005; ALLEN, R. G. *et al.*, 2011; ALLEN; TASUMI; TREZZA, 2007), Surface Energy Balance Algorithm for Land using Google Earth Engine (geeSEBAL) (BASTIAANSSEN *et al.*, 1998a; LAIPELT *et al.*, 2021), Priestley-Taylor Jet Propulsion Laboratory (PT-JPL) (FISHER; TU; BALDOCCHI, 2008), Satellite Irrigation Management Support (SIMS) (MELTON *et al.*, 2012; PEREIRA *et al.*, 2020), and Operational Simplified Surface Energy Balance (SSEBop) (SENAY, 2018; SENAY *et al.*, 2013, 2023). From the six models' ET estimates, OpenET calculates an ensemble mean and uses it as the official evapotranspiration data. OpenET provides ET data for western and central United States from 2017 to 2022 at a monthly timescale and 30 meter spatial resolution (https://openetdata.org/).

The OpenET project provides accurate, consistent, reliable, and easily accessible ET information, which helps water users and managers make better decisions towards sustainable development. The single value ensemble approach addresses questions of the user community regard the multiple ET models options, with a lack of an established ranking. In the future, flux tower data validation results will be used to calculate a new ensemble, which will select the worst performing models over particular regions, crops, or land cover type will be filtered out or receive a lower weighting.

The need for reliable ET data is not unique to the United States, as irrigation practices, climate variability, and water use conflicts are world spread. Countries from South America, Brazil in particular, can benefit from the local replication of the OpenET project to address these challenges, ensuring food security and water availability for all. The timing for initiating this endeavor is good, as a collaborative environment was built between American and Brazilian researchers, in situ ET data from several locations across South America is available for the intercomparison procedure, and there is growing interest from local water managers to monitor irrigation water use (ANA, 2020, 2021a).

The study in this chapter focuses on intercomparing ET data from SSEBop and geeSEBAL, as well as their mean, to flux tower data from 36 sites in South America, following the OpenET framework. This work represents the first step towards the development of a reliable, multi-model ET data source in South America.

6.2 Material and Methods

6.2.1 Flux Station Data

6.2.1.1 Flux Station Sites

The first stage of expansion of the OpenET project collects ET data from 36 micrometeorological stations located across South America. Figure 6.1 presents the sites location, while Table 6.1 lists the stations with details on site land cover, climate, LE acquisition method, time range of data availability, and missing SEB fluxes data. The sites use either eddy covariance (EC) or Bowen ratio (BR) measurements to provide LE and ET data at sub-daily intervals. The EC method is considered the most direct and generally accurate method for measuring latent and sensible heat fluxes, and it is available at scale over many locations, vegetation, and climate conditions, allowing for robust intercomparisons in time and space (BALDOCCHI, 2020).

The sites used in this study occur in a variety of climates and land covers. Although they are not well distributed over South America, they represent the majority (in terms of spatial distribution) of climatological and surface conditions of this continent. Land cover classifications were obtained using the International Geosphere-Biosphere Programme (IGBP) scheme, based on metadata and aerial imagery inspection.

This flux tower catalog results from the collaboration of researcher from several institutions, namely from the large-scale biosphere-atmosphere experiment in the Amazon (LBA-ECO), the South Brazilian network of surface fluxes and climate change (SULFLUX), the National Observatory of Water and Carbon Dynamics in the Caatinga Biome (ONDACBC), the long-term ecological research in Pantanal (PELD Pantanal), the Pontifical Catholic University of Peru (PUC-PE), the Federal University of Mato Grosso (UFMT), the University of Sao Paulo (USP), and the Federal university of Rio Grande do Sul (UFRGS), as well as funding from national and regional research agencies.

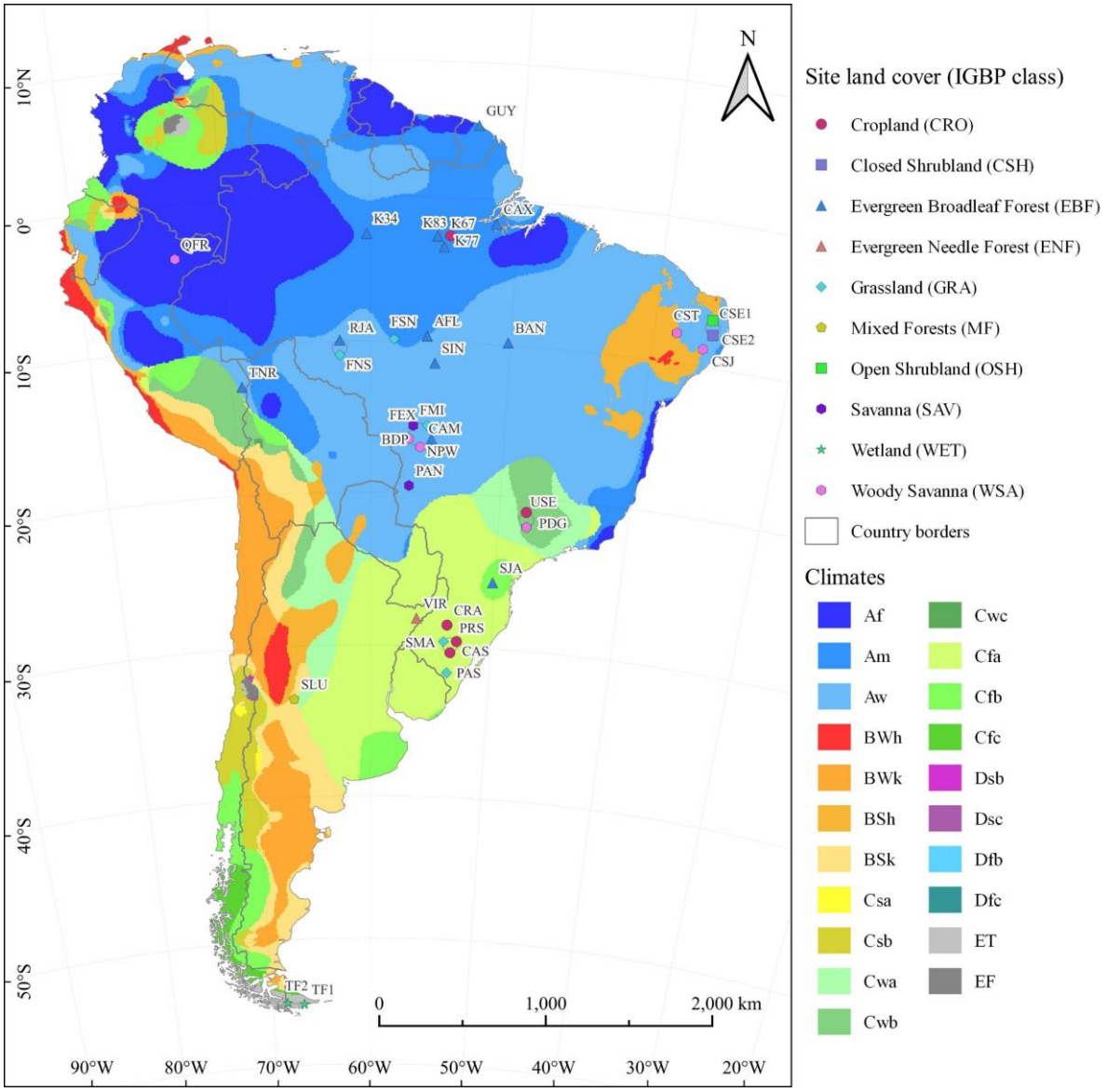


Figure 6.1 – Locations of ground-based stations used in the first stage of expansion of the OpenET benchmark dataset.

Table 6.1 – List of micrometeorological stations used for the intercomparison.

Site ID	Climate	IGBP land cover	Method	Time series (yyyy-mm-dd)		Missing SEB flux data
				Start	End	
AFL	Aw	EBF	EC	2003-02-27	2004-04-17	G
BAN	Aw	EBF	EC	2003-10-24	2006-12-09	
BDP	Aw	WSA	EC	2011-06-15	2015-08-12	
CAM	Aw	EBF	BR	2006-08-27	2010-09-28	
CAS	Cfa	CRO	EC	2009-10-19	2015-03-31	
CAX	Am	EBF	EC	1999-01-01	2003-07-31	G
CRA	Cfa	CRO	EC	2009-01-01	2014-09-07	
CSE1	Aw	CSH	EC	2013-01-09	2017-09-04	
CSE2	Aw	OSH	EC	2013-03-19	2017-02-08	
CSJ	Aw	WSA	EC	2019-04-01	2019-12-24	
CST	BSh	WSA	EC	2014-01-17	2020-03-14	
FEX	Aw	GRA	BR	2006-09-01	2010-03-11	
FMI	Aw	SAV	BR	2009-04-10	2013-06-01	
FNS	Aw	GRA	EC	1999-02-04	2002-11-05	G
FSN	Am	GRA	EC	2002-03-26	2003-10-03	
GUY	Am	EBF	EC	2004-01-01	2015-01-01	G
K34	Am	EBF	EC	1999-06-14	2006-10-01	
K67	Am	EBF	EC	2002-01-01	2012-01-01	G
K77	Am	CRO	EC	2000-08-08	2005-11-02	
K83	Am	EBF	EC	2000-06-28	2004-03-12	
NPW	Aw	WSA	EC	2013-01-01	2017-06-30	G
PAN	Aw	SAV	EC	2000-09-01	2002-06-01	
PAS	Cfa	GRA	EC	2013-09-26	2016-09-07	
PDG	Cwb	WSA	EC	2001-01-01	2004-01-01	G
PRS	Cfa	CRO	EC	2003-07-02	2004-07-22	
QFR	Af	WSA	EC	2018-01-01	2020-01-01	G
RJA	Aw	EBF	EC	1999-03-23	2002-11-15	G
SIN	Aw	EBF	EC	2005-05-01	2008-10-18	
SJA	Cfb	EBF	EC	2009-10-31	2012-04-20	Rn, G
SLU	BSk	MF	EC	2009-01-01	2012-01-01	
SMA	Cfa	GRA	EC	2013-01-01	2016-06-15	
TF1	ET	WET	EC	2016-01-01	2018-05-17	Rn, G
TF2	EF	WET	EC	2016-01-01	2018-04-17	Rn, G
TNR	Aw	EBF	EC	2017-01-01	2019-05-06	G
USE	Cwb	CRO	EC	2001-01-01	2003-01-05	G
VIR	Cfa	ENF	EC	2009-01-01	2013-01-01	G

To assess and build confidence in the ground data, I conducted energy balance closure analyses at the daily timestep, using measurements of SEB components included: latent energy (LE), sensible heat flux (H), net radiation (R_n), and soil heat flux (G), all of which are required for energy balance closure analysis and correction. On sites where G data was not available, a fraction of R_n was used, corresponding to 0.05 on forested sites, and 0.10 for all other surfaces.

6.2.1.2 Flux Data Processing and Closure Corrections

The processing, closure corrections and quality assessment for flux data are outlined by the following steps:

1) standardization of the multi-source data to single format. This step was necessary due to the multiple sources of flux data, with several formats, which hinder the application of the closure corrections;

2) gap-filling of missing or faulty sub-daily energy balance components. Gaps up to 4 hours long during the night (defined as periods with $R_n < 0$) and 2 hours during the day (defined as $R_n >= 0$) were linearly interpolated. Days with larger gaps were flagged for later inspection;

3) daily aggregation;

4) energy balance closure correction;

5) visual inspection and screening of post-processed data.

Steps 2-4 were done using the software developed using the QA/QC algorithm developed by the OpenET team (VOLK *et al.*, 2021), which available via open-source Python package in a GitHub repository. The whole flux data processing and correction procedure is described in detail by Volk et al. (2023).

SEB fluxes data is subject to SEB closure imbalances at daily time scale. To enforce the SEB closure, several methods have been developed, of which the most common are the Bowen ratio closure, where LE and H are both adjusted while preserving the observed daily Bowen ratio (H/LE) (TWINE *et al.*, 2000); and the energy balance ratio (EBR) closure, where EBR = (LE + H)/(Rn – G), and in which LE and H are both adjusted such that the EBR averages to 1 over some timescale larger than a day (PASTORELLO *et al.*, 2020). In previous studies comparing remote sensed ET to flux data in South America, the BR method was used (DE ANDRADE, 2018; KAYSER *et al.*, 2022; LAIPELT *et al.*, 2020, 2021; RUHOFF *et al.*, 2012). However, the EBR method is used in this study, to conform to the OpenET benchmark dataset (MELTON *et al.*, 2021; VOLK *et al.*, 2023). The primary benefit of the EBR is that it does not require the SEB to be closed at daily time scale, instead using a time window (in this study, 15 days) for closure. On one hand, if the EBR extremes in the time window are accurate, the correction will be overly conservative; on the other, if the extremes result from measurement errors, the EBR method will smoothen the anomalies and approximate to more realistic values.

6.2.1.3 Static Footprint Calculation

In this study, I followed a simplified methodology for estimating the upwind footprint of the SEB fluxes around the micrometeorological towers (Figure 6.2). An initial polygon is centered on the flux tower site. Rectangles of three sizes were produced, with 90 m, 150 m, and 210 m sides. Based on the flux site wind speed and direction daytime data (between 06:00 and

18:00 local time), the rectangles were shifted in order to capture a more representative area. The maximum shift was capped in order to include the tower location. Finally, an individualized post-processing was performed to remove non-representative features, such roads, constructions, open water, and different land covers. For the intercomparison with remote sensing data, only the 210 m polygon was used, as it has been demonstrated that it is more representative of a homogenous area and provides less remote sensing data gaps (FISHER *et al.*, 2020b; MELTON *et al.*, 2021; VOLK *et al.*, 2023).

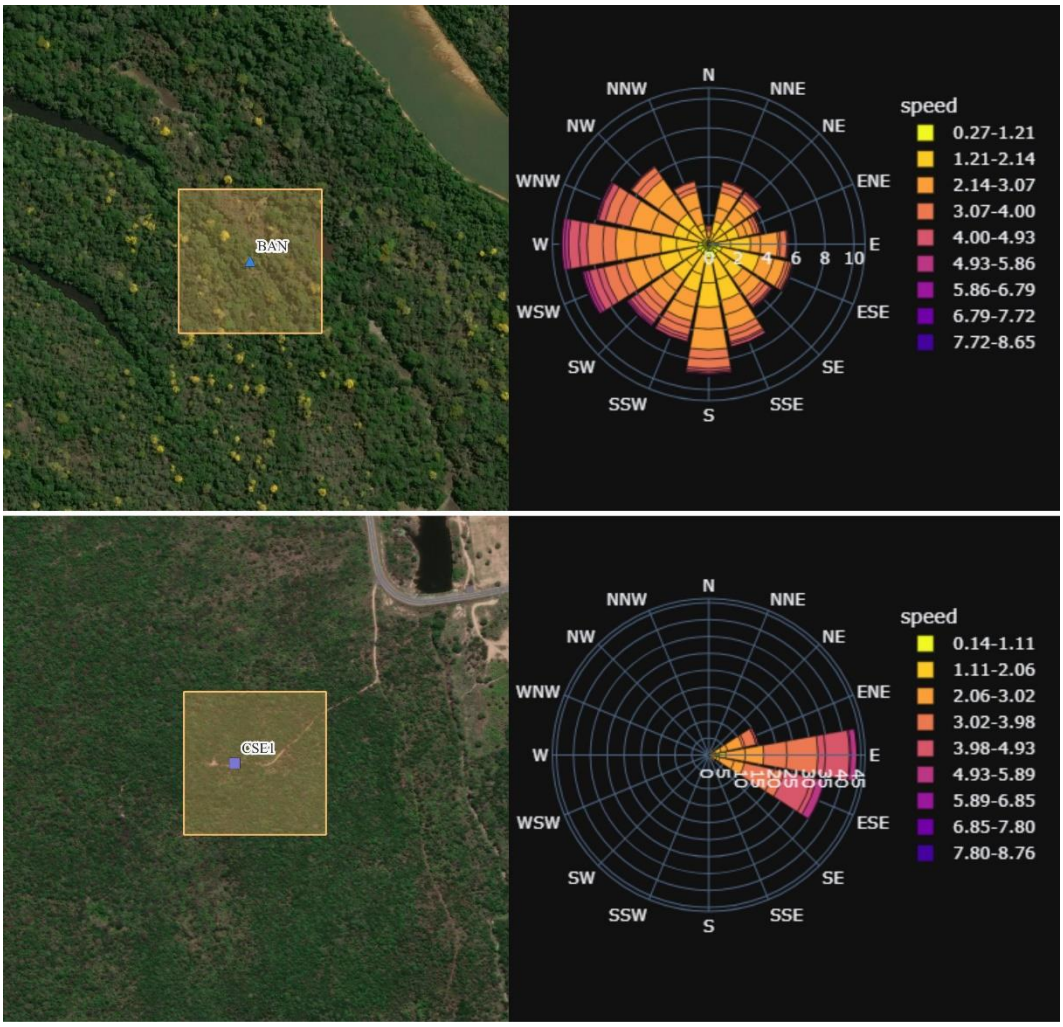


Figure 6.2 – Example of the footprint calculation algorithm for two flux towers: BAN (top), and CSE1 (bottom). The wind rose with wind speed and direction frequency for each tower is also shown.

6.2.2 Remote Sensing Models

6.2.2.1 Input Data

Table 6.2 presents the open-source data used for the application of the SSEBop and geeSEBAL models. The major difference from the original OpenET is that global

meteorological and land cover datasets were used here. A land cover dataset was generated by merging data from Copernicus Global Land Cover (CGLS) – Collection 2 (BUCHHORN *et al.*, 2020) from 2015 to 2019, and European Space Agency (ESA) WorldCover 10m (ZANAGA *et al.*, 2021) for 2020-2021. For periods before 2015, the first image of CGLS was used, while for periods after 2021, the last image of ESA was used. Other datasets that are model-specific are mentioned in the models' description that follows.

Table 6.2 – Input data for the remote sensing ET models.

			Resolution	
Variable	**Datasets**	**Symbols**	**Temporal**	**Spatial**
Elevation	USGS SRTM	z	-	30 m
Land surface temperature	LANDSAT 5 TM LANDSAT 7 ETM+ LANDSAT 8 OLI/TIRS	LST	16 days	100 m
Surface reflectance	LANDSAT 5 TM LANDSAT 7 ETM+ LANDSAT 8 OLI/TIRS	S	16 days	30 m
Dew point temperature	ERA5-Land Hourly	T_d	1 hour	11,132 m
Air temperature	ERA5-Land Hourly	T_a	1 hour	11,132 m
Wind	ERA5-Land Hourly	u	1 hour	11,132 m
Solar radiation	ERA5-Land Hourly	R_s	1 hour	11,132 m
Land cover	CGLS-LC100 ESA WorldCover 10m v100 ESA WorldCover 10m v200	LC	1 year	100 m 10 m 10 m

6.2.2.2 SSEBop Model

Model description

The Operational Simplified Surface Energy Balance (SSEBop) model is a widely used tool for global ET mapping, which computes actual ET as a product of ET fraction (ET_f) and maximum ET under water-unlimited environmental conditions, often considered equivalent to alfalfa reference ET (ET_r), as shown in Eq. 22:

$$ET = ET_f \cdot ET_r \tag{22}$$

To estimate ET_f (Eq. 23), SSEBop uses the concept of satellite psychrometry that employs two model parameters to define the model boundary conditions for minimum and maximum ET, in association with observed satellite land surface temperature (LST).

$$ET_f = 1 - \gamma_s \cdot (LST - T_c) \tag{23}$$

The first parameter, γ_s, is the surface psychrometric constant over a dry-bare surface, determined based on energy balance principles and may be assumed constant for a given

location and day of year, and can be interpreted as the inverse of the dT parameter (SENAY *et al.*, 2013); and the second is the wet-bulb reference surface temperature (T_c) limit.

γ_s pre-calculated for each day-of-year, using Eq. (24):

$$\gamma_s = \frac{\rho \cdot C_p}{R_n \cdot r_{ah}} \tag{24}$$

Where ρ is the density of air (kg m^{-3}), calculated as a function of elevation (ALLEN *et al.*, 1998-; SENAY *et al.*, 2013); C_p is specific heat of air at constant pressure, 1.013×10^{-3} (MJ kg^{-1} $°C^{-1}$); R_n is the daily average net radiation (MJ m^{-2} d^{-1}); and r_{ah} (s m^{-1}) is the aerodynamic resistance over dry-bare soil, 180 s m^{-1}. γ_s is derived using R_n is derived from ERA5 downwards shortwave radiation, along with other weather datasets, T_a from Daymet and Chelsea, to create a daily median (SENAY et al., 2022).

Recently, a new approach was developed to improve the accuracy of the T_c parameter estimation and expand the spatiotemporal coverage of the SSEBop model, addressing the limitations of previous methods, such as the lack of high NDVI pixels in arid and semi-arid regions, non-uniform distribution of high NDVI calibration landscapes, and widely varying T_c results due to the NDVI threshold for greenest pixel selection (SENAY *et al.*, 2023). The new method, named forcing and normalizing operation (FANO) for the determination of T_c is given by Eq. (25-27):

$$T_c^* = LST^* - \frac{f}{\gamma_s^*}(0.9 - NDVI^*) \tag{25}$$

$$c = \frac{T_c^*}{T_{air}^*} \tag{26}$$

$$T_c = c \cdot T_{air} \tag{27}$$

where * denotes the average value for a large area; f is a proportionality "FANO" constant, globally assumed as 1.25, that relates the ratio of LST- T_c to 0.9 – NDVI. Implementation of this set of equations is dependent on NDVI mean value and water pixels for each coarse pixel. Initially, a 5 km pixel is used, and Eq. (25) is applied if $NDVI^*$ is between 0.0 and 0.9; if mean $NDVI^*$ is higher than 0.9 or lower than 0.0, T_c^* is considered equal to LST^*; if $NDVI^*$ is between 0.0 and 0.9, but too many water pixels are identified (10%), the large area is set to 100 km and Eq. (25) is reapplied. T_c^* is resampled to the T_{air} dataset resolution using Eq. (26-27), where T_{air} is the spatially averaged daily max air temperature.

6.2.2.3 geeSEBAL Model

A detailed overview of geeSEBAL formulation can be found in Laipelt et al. (2021), which is based on the original model developed by Bastiaanssen et al. (1998a). Differently from geeSEBAL-MODIS, described in chapter 4 of this dissertation, the OpenET version of geeSEBAL uses Landsat images for retrieval of LST data, and evapotranspiration fraction is converted to daily ET using daily average net radiation instead of daily average ET_r.

The latest version of geeSEBAL incorporates several changes based on the accuracy assessment and intercomparison results from OpenET, as follows:

1) the simplified version of CIMEC was improved by using additional filters to select the endmembers, including the use of land cover data and filters for NDVI, LST and albedo;
2) corrections to LST for endmembers selection based on antecedent precipitation;
3) definition of wind speed thresholds to reduce model instability;
4) improvements to estimate daily net radiation, using FAO-56 as reference (ALLEN et al., 1998).

6.3 Results

6.3.1 Flux Data Closure Corrections

The closure correction method results are presented in Figure 6.3. This technique rarely results in the perfect closure of the energy balance on any given date, but the SEB fluxes are adjusted so that closure converges to 1 over the sliding window periods (15 days). Figure 6.4 presents the comparison of ET data from the flux towers, before and after the closure correction algorithm application. For the towers in South America, the EBR closure correction generally results in higher daily ET rates. The corrected value are used for the intercomparison with the remote sensing models.

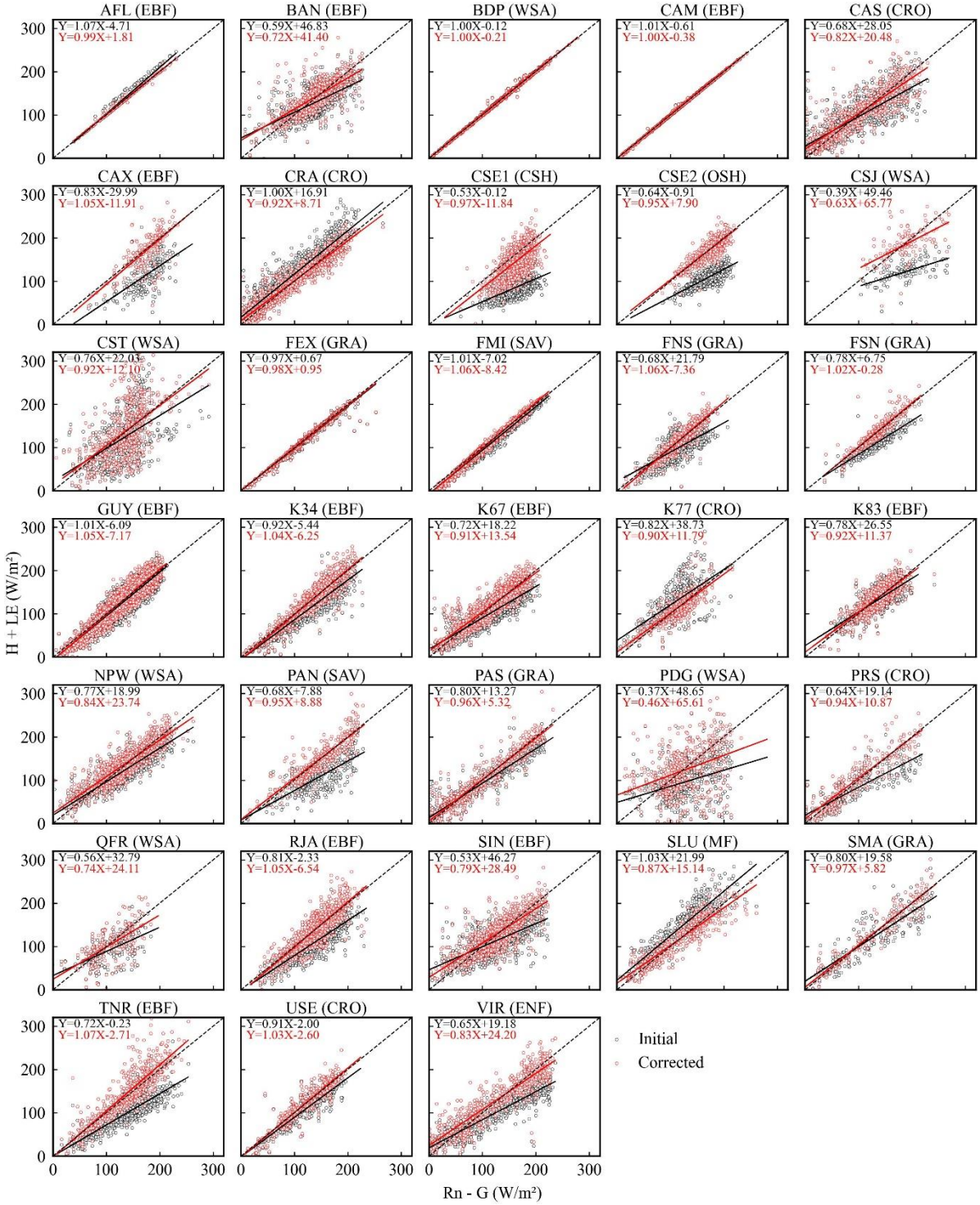


Figure 6.3 – Daily average of turbulent fluxes and closure correction results. Regression lines and equation are presented for SEB fluxes before (black) and after (red) EBR closure corrections.

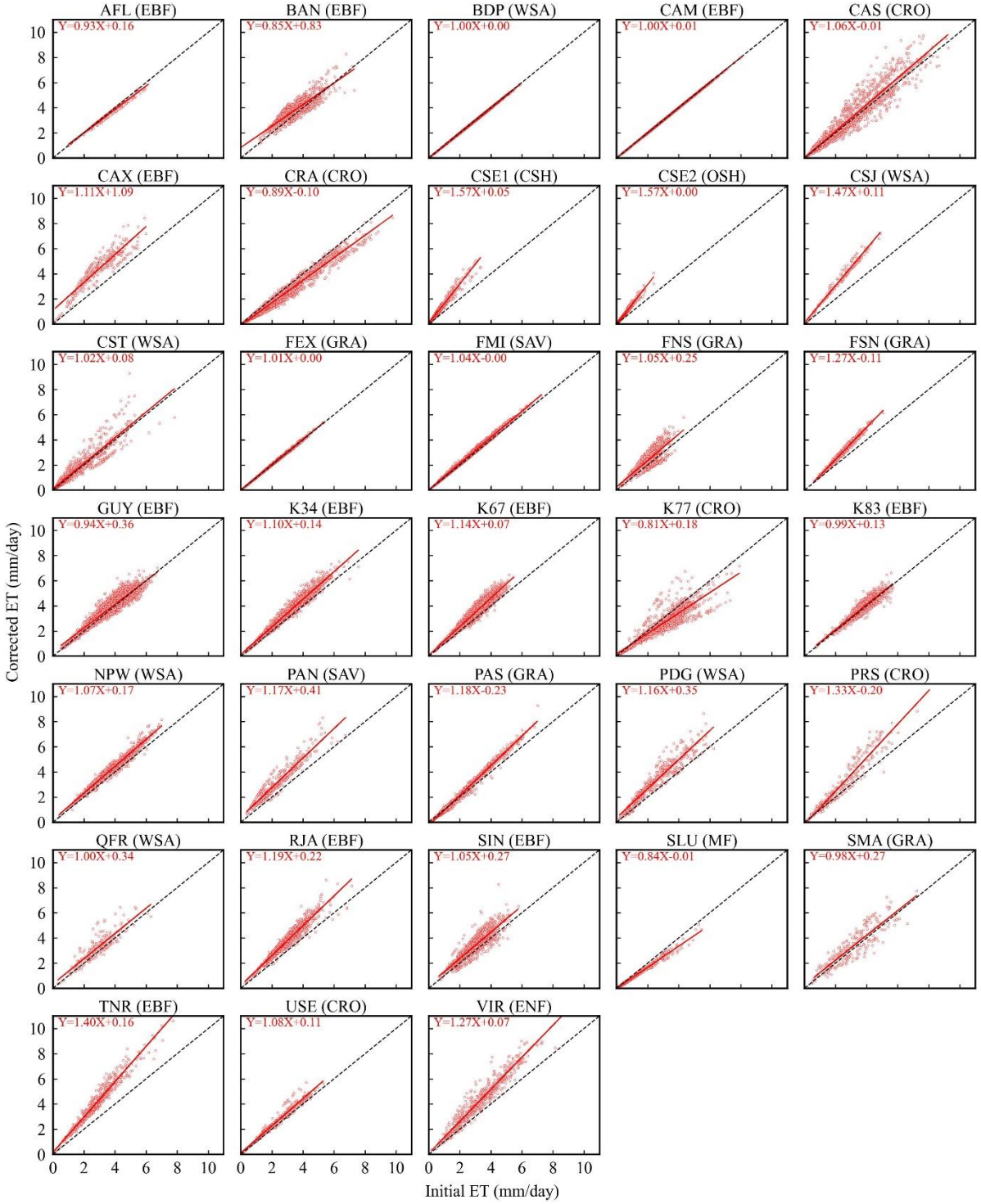


Figure 6.4 – Comparison of flux tower ET data before and after closure corrections.

6.3.2 Intercomparison

Figure 6.5 shows ET time series by day of (DOY) for the 36 flux sites of South America, used in the first stage of the OpenET expansion project outside of the United States. The remote sensing models (SSEBop, geeSEBAL, and ensemble, which is a simple mean of the two) are capable of representing the seasonality of flux tower ET data. However, a general

overestimation is observed for remote sensing ET. This overestimation occurs more frequently on drier climates (Figure 6.6), such as Aw (Tropical, dry winter), Bs (Dry, semi-arid), and Cw (Temperate, dry winter); and drier landscape (Figure 6.7), such as shrublands (OSH and CSH) and mixed forests (MF). On woody savannas (WSA), lower ET values from flux towers are overestimated by the remote sensing models, while the max values yield stronger agreement. Statistical comparison metrics for each climate (Figure 6.8) and for each land cover (Figure 6.9) reveal that the best performance of OpenET models occurs in temperate (Cfa) and in tropical monsoon (Am) climates, and over croplands (CRO) and grasslands (GRA). The intercomparison statistical metrics for each flux tower are available in Table S 1.

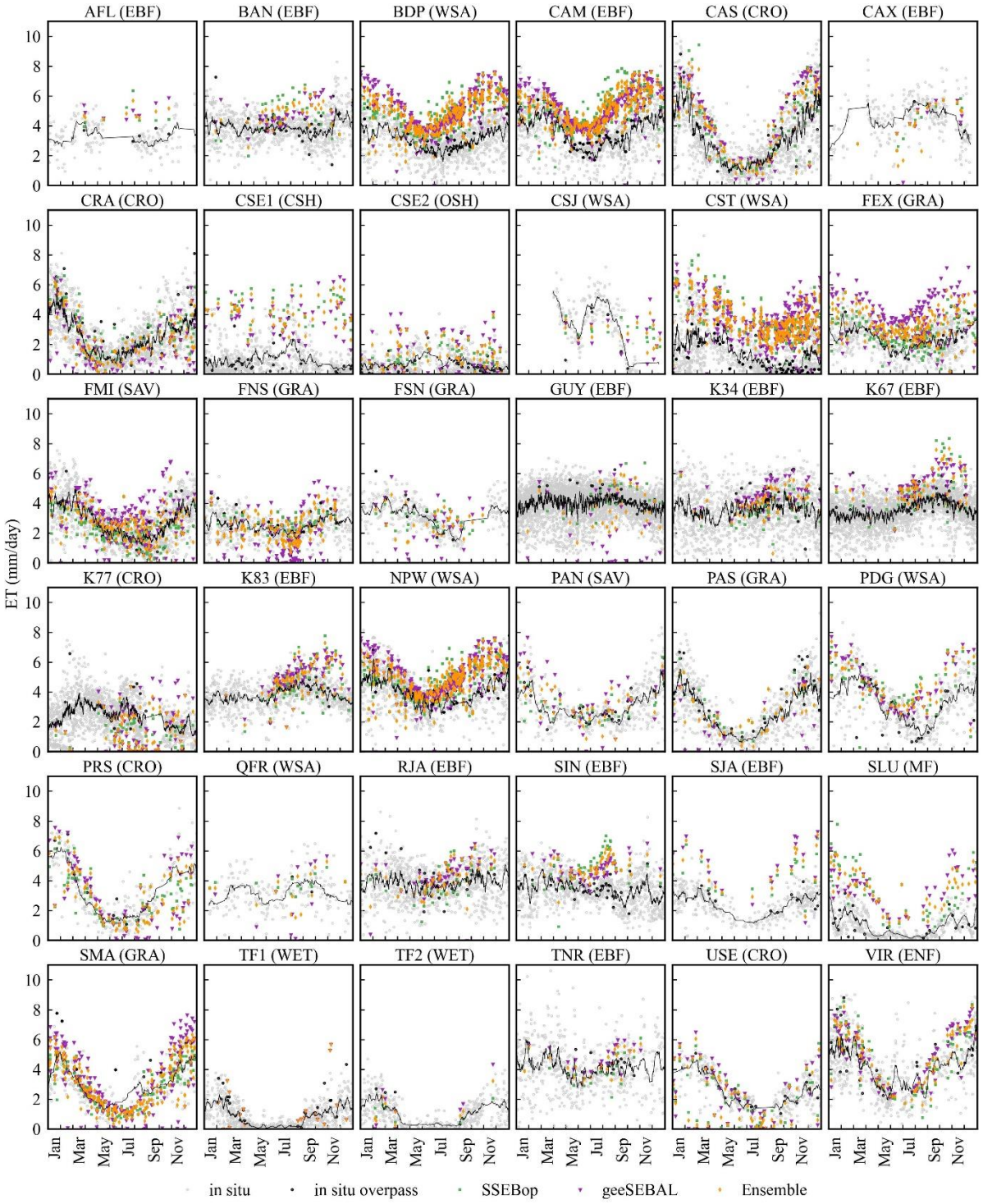


Figure 6.5 – Time series of ET from the towers (light grays is all tower data and black is data on days of satellite overpass, which provide remote sensing data as well), SSEBop, geeSEBAL, and the ensemble mean. Values are displayed by day of year (DOY), the black line represents the tower data 15 DOY moving average.

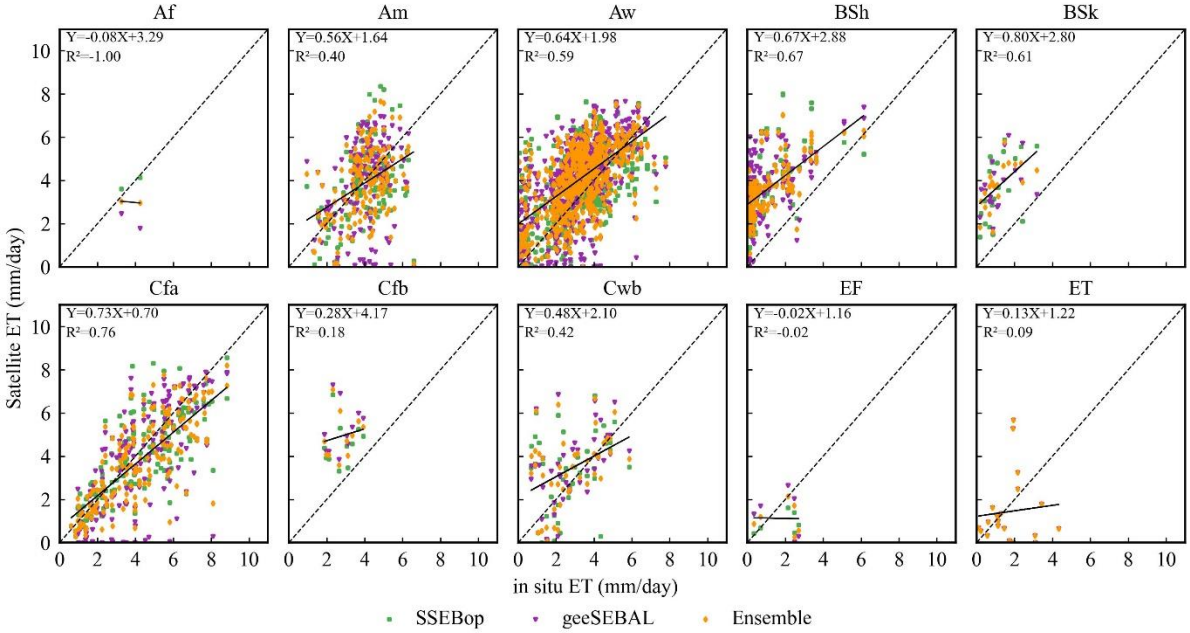


Figure 6.6 – Comparison scatterplots between daily ET data from the towers and from the remote sensing models, for each climate classification. The linear regression equation and determination coefficient (R^2) for the comparison between in situ (X) and Ensemble (Y) ET is also shown

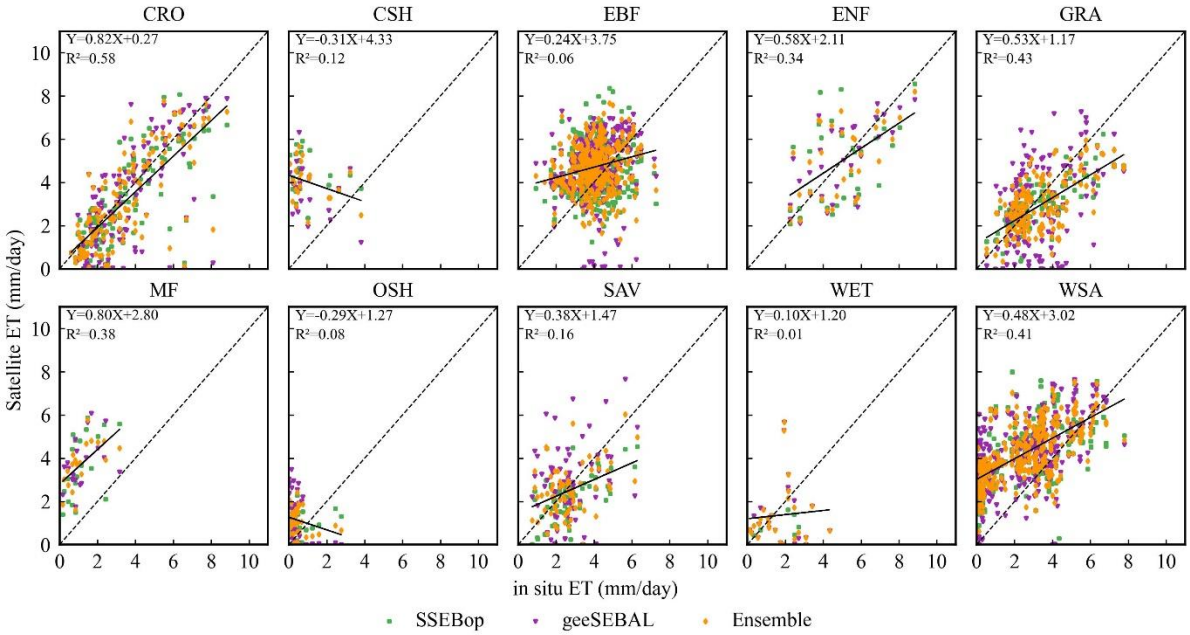


Figure 6.7 – Comparison scatterplots between daily ET data from the towers and from the remote sensing models, for each land cover classification. The linear regression equation and determination coefficient (R^2) for the comparison between in situ (X) and Ensemble (Y) ET is also shown.

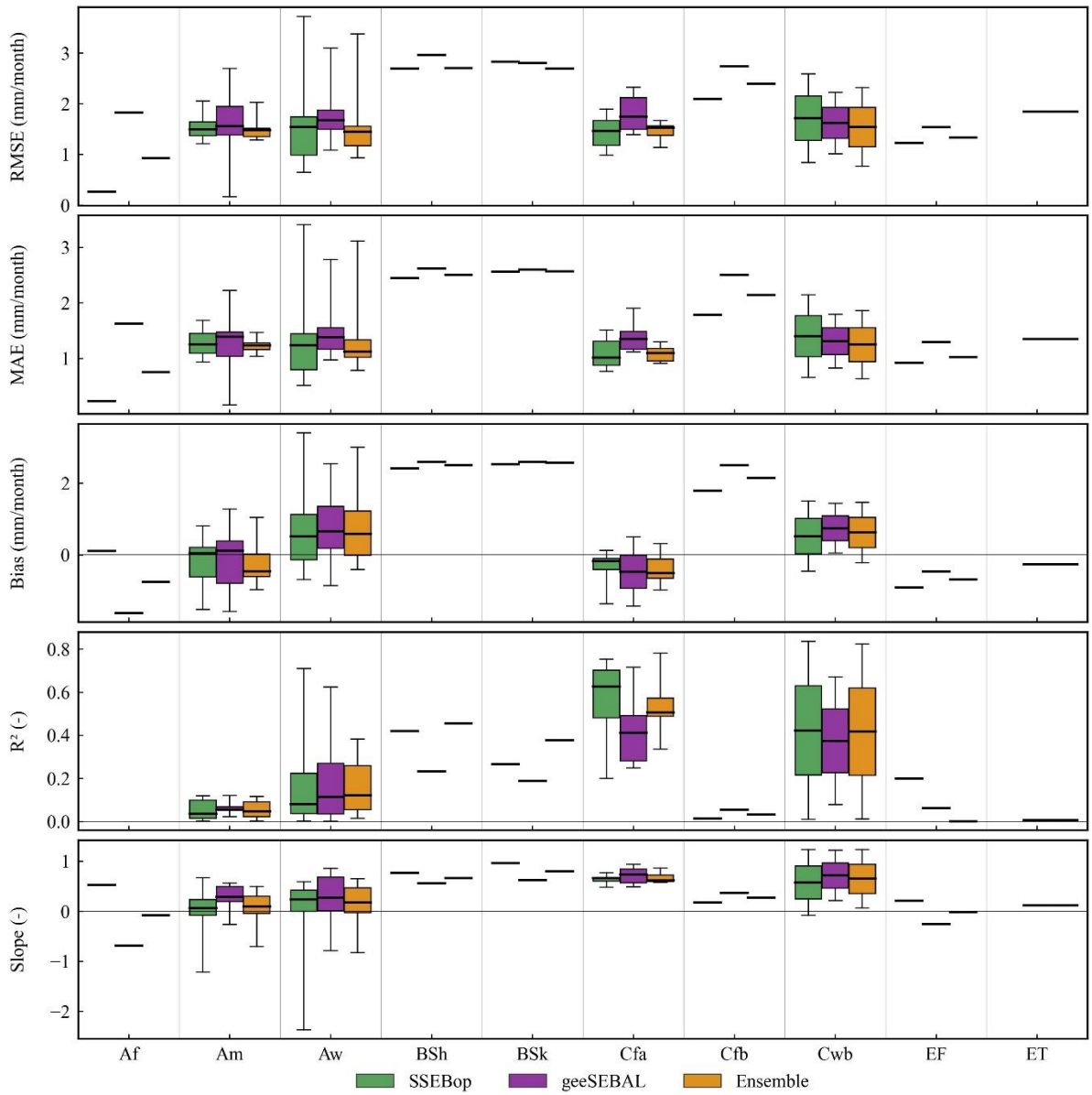


Figure 6.8 – Intercomparison statistics boxplots between flux tower ET and the remote sensing models for each climate.

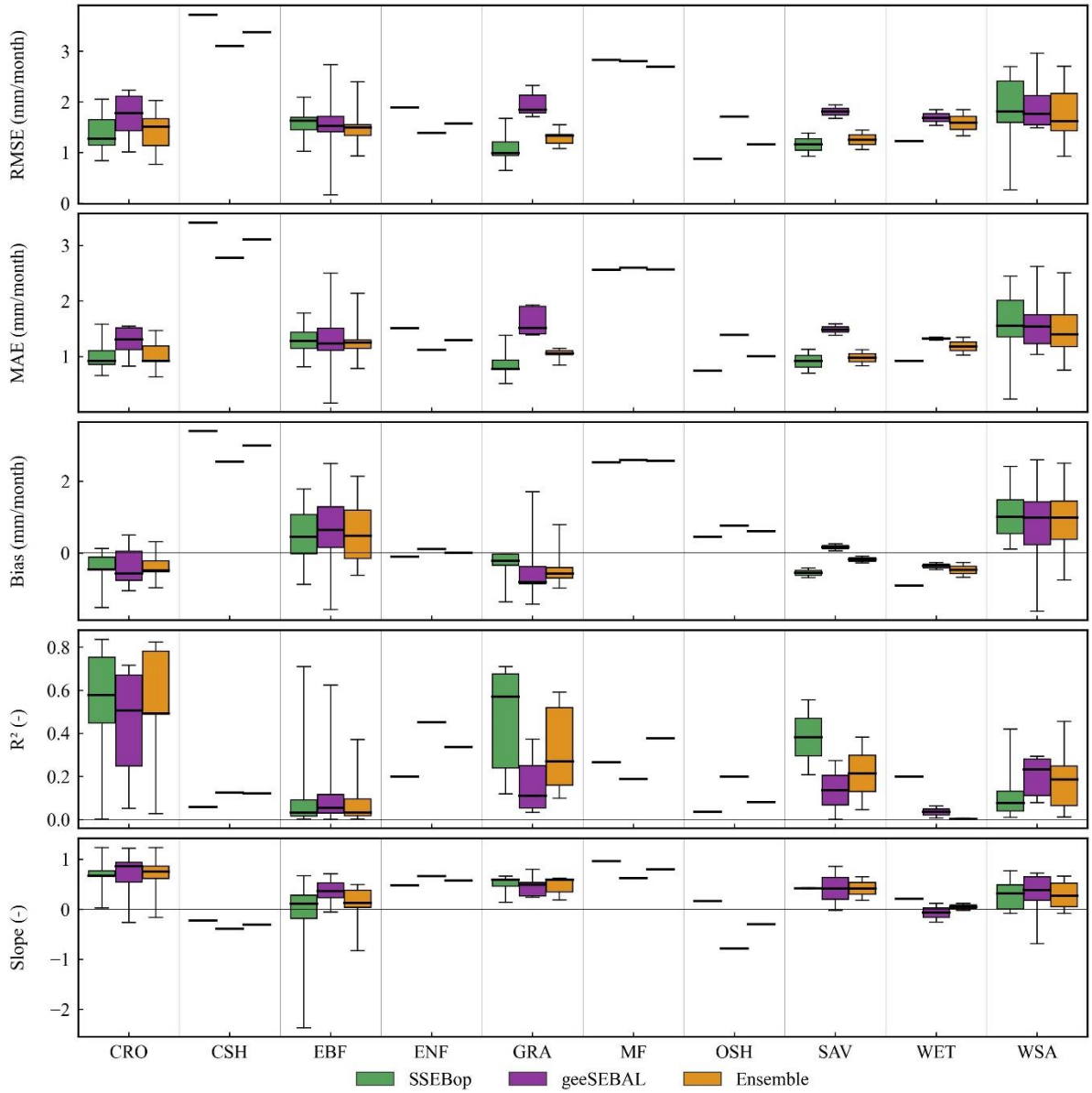


Figure 6.9 – Intercomparison statistics boxplots between flux tower ET and the remote sensing models for each land cover.

6.4 Discussion

6.4.1 Limitations of Flux Station Data

Flux towers are recognized among the best alternatives for measuring actual evapotranspiration, with applications over various surface and climate conditions (BALDOCCHI, 2020; JIANG; LIU, 2021; KOLARS *et al.*, 2013), and that is traditionally used as the basis for comparison with remote sensing methods (BHATTARAI et al., 2019; DE ANDRADE, 2018; LI et al., 2018; MELTON et al., 2021; SENAY et al., 2022). A necessary assumption for the ET measurement from micrometeorological stations is that there is sufficient "fetch", which refers to the upwind distance with uniform features, to ensure that the

measurements accurately represent the underlying surface and are not affected by flux from neighboring surfaces (DREXLER *et al.*, 2004). One of the towers with available data in South America was removed from the analysis due to its location between irrigated and non-irrigated croplands (LATHUILLIÈRE *et al.*, 2018).

In situ data gaps due to instrument malfunction, power outages, or other issues, can significantly impact the accuracy of analysis. First, gaps reduce data availability for the accuracy assessment. Second, they can influence the data processing, correction, and quality control, introducing biases and inaccuracies. Data gaps during extreme SEB events, the EBR moving average value may be underestimated or overestimated, leading to inaccurate corrections.

The lack of G measurement in many flux towers is a major issue, as it decreases trust in other fluxes measurement, due to the inability to verify the SEB closure. In this study, G was estimated as a fraction of R_n at 12 sites. This simplified strategy allowed the closure correction algorithm application, but does not increase confidence in data. In the original OpenET benchmark dataset development (VOLK *et al.*, 2023), only towers with the four SEB flux (R_n, H, LE, and G) measurements were selected. Nonetheless, the dataset still included over 300 towers. Applying this criterion to our study would mean that more one-third (12 without G, and 2 without Rn and G) of stations with available data would be eliminated. For this reason, the principal investigator decided for the maintenance of all towers. In future studies, it is recommended that more robust methods for estimating G are used (DE ANDRADE *et al.*, 2021a). Even then, data from such towers will always present high uncertainty associated.

Energy balance closure correction methods exist because of the uncertainty in H and LE measurements, compared to R_n and G. Although they adjust H and LE to enforce energy balance closure, these methods can introduce additional uncertainties and biases in the data, depending on the method used and the assumptions made. The EBR method used in this study tends to apply overly conservative corrections, especially for data which presents high variability in the ratio. On the other hand, since it does not force EBR = 1.0 for all days, it reduces the effects of unrealistic values in measurements errors.

6.4.2 OpenET models intercomparison

Compared to flux tower ET data, SSEBop, geeSEBAL, and the ensemble performed similarly, with overall RMSE equal to 1.57, 1.81 and 1.58 mm day^{-1}, respectively. The worst performance was identified in closed shrublands, followed by mixed forests, which occur in regions of dry climate (Bsh, Bsk). In these areas, both models overestimate ET, possibly due to

the parameterization in geeSEBAL endmember selection procedure, and for SSEBop, due to the ET_r overestimation from ERA5-Land data (ALLEN *et al.*, 2021). The r_{ah} parameter in SSEBop might also be overestimated, although ET is less sensitive to it in the new FANO formulation (SENAY *et al.*, 2023). Generally, the ensemble model yielded errors within the other two or slightly lower. The same pattern is observed for other metrics, with the ensemble performance between the two models or very close to the best one. Overall, the results of this study reveal that even the simple average of two models provides more robust estimates of ET.

The original OpenET has a distinct advantage over this study, which is the utilization of higher spatial resolution and bias-corrected meteorological data. This approach results in a significant reduction in errors within the gridded atmospheric dataset, which in turn improves the accuracy of the ET estimates generated by remote sensing models. In future studies, accuracy assessment and bias correction of ERA5 Land (or another global meteorological dataset) is recommended.

In comparison to previous studies conducted in South America utilizing geeSEBAL and SSEBop (DE ANDRADE, 2018; LAIPELT *et al.*, 2020, 2021; RUHOFF *et al.*, 2012), the validation results in this study are relatively poorer. This decrease in performance is anticipated, as earlier studies were conducted on a smaller scale, with a limited number of study sites, and the models underwent a local calibration process to improve their accuracy. In the present study, the models were used in their default formulation, which provides average performance with occasional misestimations.

Other potential sources of error in the remote sensing ET data are:

1) the land cover classification error: For this study, global land cover datasets were selected due to the coverage and spatial resolution. However, the time range of data coverage was limited to 2015-2021. Since a big portion of flux tower data was obtained prior to 2015, erroneous land cover was assumed in geeSEBAL endmember selection, potentially leading to large errors between 1999 and 2014;

2) the errors associated with Landsat collection 2 LST data: collection 2 provides LST data calculated from the sensor thermal bands and from the Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Emissivity Dataset (ASTER GED) Version 3 (GEDv3) emissivity data. The combination of the different spatial resolutions from Landsat and ASTER, in addition to the resampling method, results in artifacts that make LST images look blocky (BLOCKINESS ARTIFACT IN LANDSAT COLLECTION 2 SURFACE TEMPERATURE PRODUCTS | U.S. GEOLOGICAL SURVEY, [*s. d.*]). Despite these issues,

no significant improvement is observed overall in the intercomparison with flux tower data when using Landsat collection 1 imagery. Examples of this effect are exemplified in Figure 6.10 for two flux tower sites, and a summary of the statistical metrics for the intercomparison with flux tower data is available in the supplementary material (Table S 2). Furthermore, data gaps in ASTER are replicated in the Landsat LST band (LANDSAT COLLECTION 2 SURFACE TEMPERATURE DATA GAPS DUE TO MISSING ASTER GED | U.S. GEOLOGICAL SURVEY, [*s. d.*]). These issues can be detrimental to both the SSEBop and the geeSEBAL models, as it may lead to an incorrect relationship between vegetation indexes and LST. Given the discontinuation of Landsat collection 1, replacement of Landsat LST band with user-computed LST is recommended for the calculation of ET.

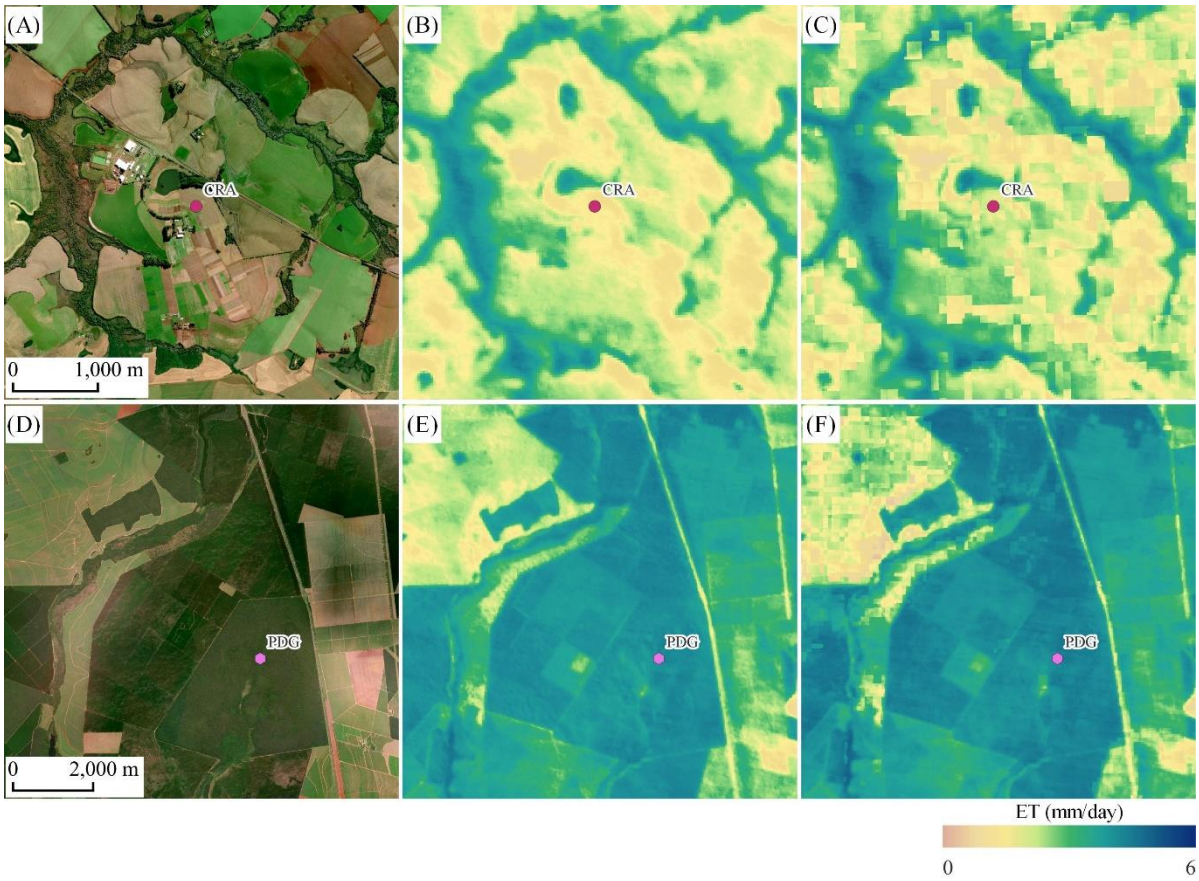


Figure 6.10 – True color images (A, D) and ensemble average ET using Landsat collection 1 (B, E) and collection 2 (C, F), over CRA and PDG flux tower sites.

6.4.3 Challenges in the OpenET implementation in South America

The main drawback of remote sensing ET modeling is the requirement of clear-sky conditions (BHATTARAI *et al.*, 2017). Cloud cover produces data gaps and hinders the construction of a time series and the comparison with ground data. For example, in the QFR site, despite a two-year window of available data, the comparison was possible for only two

dates, because of data gaps in both ground and satellite data. Cloud cover poses a challenge to the application of projects such as the OpenET, but investments in the increase of simultaneous orbiting satellites help reduce revisit time in mitigate this limitation.

This study demonstrated the reproducibility of OpenET outside the United States, using South America as the starting point. Continuation of this research will require the inclusion of the other OpenET models and maybe even others, increasing remote sensing ET robustness and accuracy. Moreover, collaboration with regional and national water management agencies in South America, such as the Brazilian Water Agency (ANA), could provide valuable insights for the definition of the main interest areas for ET monitoring, such as irrigation districts and basins experiencing water use conflicts. Ultimately, this collaboration could contribute to improving water and food security in key regions, thereby advancing sustainable development efforts.

6.5 Conclusions

This chapter describes the implementation of the OpenET framework for intercomparison of remote sensing evapotranspiration (ET) models in South America. Ground truth data were obtained from flux data collected at 36 micrometeorological stations and corrected using the same state-of-the-art methodology used in the original project. Two OpenET models, SSEBop and geeSEBAL, were adapted for global application, and an ensemble mean was calculated. The results showed reasonable agreement between remote sensing models and flux tower data, although overestimations were observed in semi-arid climates. To increase the accuracy of the models, potential improvements could come from bias correction of meteorological data, following the example set by the original project. This work also demonstrated the reproducibility of the OpenET approach and its potential for expansion beyond the United States borders. Future efforts should focus on applying more models to increase the robustness and accuracy of the ensemble model.

The implementation of the OpenET intercomparison framework in South America is a testament to the importance of a collaborative community in advancing scientific research. Given the challenging prospect of growing constraints and increasing interannual variability in water supplies, continued collaboration between researchers, water users, and water managers will be crucial in developing more accurate, accessible and reliable evapotranspiration data. Such information will aid in better management of water resources and agricultural practices,

addressing the challenges towards sustainable water resource management in South America and beyond.

6.6 Supplementary Material

Table S 1 – Summary of intercomparison statistics of daily evapotranspiration from flux tower and remote sensing data, including the SSEBop, geeSEBAL and the Ensemble models, derived with Landsat Collection 2 images. RMSE, MAE and Bias are in mm d^{-1}; R^2 and the slope of the linear regression are nondimensional.

Flux Site	Climate	IGBP Class	SSEBop					geeSEBAL					Ensemble				
			RMSE	MAE	Bias	R^2	Slope	RMSE	MAE	Bias	R^2	Slope	RMSE	MAE	Bias	R^2	Slope
AFL	Aw	EBF	1.67	1.24	1.24	0.71	-2.36	1.90	1.86	1.86	0.27	0.71	1.68	1.55	1.55	0.37	-0.82
BAN	Aw	EBF	1.68	1.46	0.51	0.02	-0.17	1.48	1.16	0.66	0.00	-0.05	1.47	1.19	0.57	0.01	-0.10
BDP	Aw	WSA	1.87	1.48	1.43	0.04	0.24	1.71	1.46	1.39	0.29	0.73	1.71	1.44	1.40	0.19	0.47
CAM	Aw	EBF	1.76	1.38	1.02	0.08	0.26	1.53	1.36	1.32	0.62	0.69	1.57	1.28	1.16	0.33	0.48
CAS	Cfa	CRO	1.15	0.93	0.13	0.75	0.78	1.43	1.13	0.51	0.72	0.95	1.14	0.92	0.31	0.78	0.87
CAX	Am	EBF	1.69	1.69	0.05	-	-1.21	0.17	0.17	0.17	-	0.52	1.34	1.30	-0.35	-	-0.70
CRA	Cfa	CRO	1.28	0.86	-0.12	0.58	0.67	2.23	1.52	-1.05	0.25	0.55	1.51	0.93	-0.49	0.49	0.62
CSE1	Aw	CSH	3.73	3.41	3.40	0.06	-0.22	3.10	2.78	2.55	0.13	-0.39	3.38	3.11	3.00	0.12	-0.31
CSE2	Aw	OSH	0.87	0.74	0.44	0.03	0.17	1.71	1.39	0.76	0.20	-0.78	1.16	1.01	0.60	0.08	-0.29
CSJ	Aw	WSA	1.76	1.62	0.52	0.08	-0.06	1.49	1.16	0.11	0.11	0.18	1.53	1.37	0.31	0.07	0.06
CST	BSh	WSA	2.70	2.45	2.41	0.42	0.77	2.97	2.62	2.60	0.23	0.57	2.71	2.51	2.50	0.46	0.67
FEX	Aw	GRA	0.94	0.78	-0.03	0.24	0.60	2.13	1.93	1.71	0.11	0.54	1.19	1.04	0.79	0.27	0.60
FMI	Aw	SAV	1.39	1.13	-0.68	0.21	0.41	1.94	1.59	0.07	0.00	-0.02	1.45	1.13	-0.28	0.05	0.18
FNS	Aw	GRA	0.65	0.52	-0.03	0.57	0.46	1.85	1.52	-0.86	0.03	0.27	1.08	0.85	-0.40	0.16	0.35
FSN	Am	GRA	1.22	0.94	-0.34	0.12	0.14	1.78	1.41	-0.81	0.05	0.25	1.36	1.11	-0.58	0.10	0.19
GUY	Am	EBF	1.25	1.01	0.05	0.02	-0.19	2.70	2.23	-1.58	0.02	0.48	1.53	1.23	-0.62	0.00	0.11
K34	Am	EBF	1.50	1.26	-0.88	0.01	0.07	1.23	0.88	0.12	0.05	0.16	1.28	1.04	-0.45	0.02	0.09
K67	Am	EBF	1.60	1.32	0.81	0.11	0.67	1.56	1.39	1.28	0.07	0.29	1.49	1.27	1.04	0.12	0.50
K77	Am	CRO	2.06	1.59	-1.52	0.00	0.03	2.11	1.55	-0.76	0.05	-0.26	2.03	1.47	-0.97	0.03	-0.16
K83	Am	EBF	1.50	1.19	0.38	0.05	0.34	1.54	1.21	0.62	0.12	0.57	1.51	1.21	0.40	0.07	0.42
NPW	Aw	WSA	1.54	1.32	0.58	0.13	0.39	1.50	1.04	0.59	0.28	0.68	1.40	1.12	0.58	0.25	0.54
PAN	Aw	SAV	0.93	0.71	-0.42	0.56	0.44	1.68	1.39	0.26	0.27	0.86	1.06	0.84	-0.09	0.38	0.66
PAS	Cfa	GRA	0.99	0.77	-0.22	0.71	0.67	2.33	1.90	-1.42	0.25	0.49	1.33	1.06	-0.69	0.59	0.59
PDG	Cwb	WSA	2.59	2.14	1.50	0.01	-0.08	2.23	1.80	1.44	0.08	0.22	2.32	1.86	1.47	0.01	0.07
PRS	Cfa	CRO	1.65	1.11	-0.47	0.45	0.66	1.78	1.31	-0.56	0.51	0.86	1.67	1.19	-0.51	0.49	0.76
QFR	Af	WSA	0.26	0.24	0.11	-	0.52	1.83	1.63	-1.63	-	-0.68	0.93	0.76	-0.76	-	-0.08
RJA	Aw	EBF	1.33	1.03	-0.24	0.03	0.16	1.65	1.18	0.09	0.00	0.04	1.34	1.04	-0.09	0.02	0.11
SIN	Aw	EBF	1.72	1.44	1.31	0.00	0.07	1.48	1.27	1.25	0.11	0.35	1.55	1.31	1.29	0.03	0.15
SJA	Cfb	EBF	2.09	1.79	1.79	0.01	0.18	2.74	2.50	2.50	0.05	0.37	2.40	2.14	2.14	0.03	0.28
SLU	BSk	MF	2.83	2.57	2.53	0.27	0.97	2.81	2.60	2.60	0.19	0.63	2.70	2.57	2.57	0.38	0.80
SMA	Cfa	GRA	1.68	1.39	-1.36	0.68	0.59	1.71	1.39	-0.38	0.37	0.80	1.55	1.15	-0.97	0.52	0.63
TF1	ET	WET	-	-	-	-	-	1.85	1.35	-0.27	0.01	0.13	1.85	1.35	-0.27	0.01	0.13
TF2	EF	WET	1.23	0.93	-0.90	0.20	0.22	1.54	1.30	-0.46	0.06	-0.25	1.33	1.03	-0.68	0.00	-0.02
TNR	Aw	EBF	1.03	0.82	-0.37	0.10	0.48	1.09	0.97	0.48	0.04	0.26	0.93	0.79	0.05	0.07	0.37
USE	Cwb	CRO	0.84	0.66	-0.46	0.84	1.24	1.02	0.83	0.05	0.67	1.22	0.77	0.64	-0.22	0.82	1.24
VIR	Cfa	ENF	1.90	1.51	-0.09	0.20	0.49	1.39	1.12	0.10	0.45	0.67	1.57	1.30	0.01	0.34	0.58

Table S 2 – Summary of intercomparison statistics of daily evapotranspiration from flux tower and remote sensing data, including the SSEBop, geeSEBAL and the Ensemble models, derived with Landsat Collection 1 images. RMSE, MAE and Bias are in mm d^{-1}; R^2 and the slope of the linear regression are nondimensional.

Flux Site	Climate	IGBP Class	SSEBop					geeSEBAL					Ensemble				
			RMSE	MAE	Bias	R^2	Slope	RMSE	MAE	Bias	R^2	Slope	RMSE	MAE	Bias	R^2	Slope
AFL	Aw	EBF	1.60	1.07	1.06	0.65	-2.49	1.67	1.66	1.66	0.72	0.82	1.51	1.36	1.36	0.38	-0.84
BAN	Aw	EBF	1.63	1.36	0.31	0.02	-0.15	1.46	1.14	0.53	0.00	0.03	1.42	1.12	0.40	0.00	-0.05
BDP	Aw	WSA	1.65	1.30	1.22	0.06	0.27	1.75	1.52	1.46	0.38	0.90	1.60	1.35	1.31	0.25	0.55
CAM	Aw	EBF	1.57	1.25	0.78	0.09	0.25	1.49	1.33	1.26	0.61	0.72	1.43	1.17	0.99	0.36	0.48
CAS	Cfa	CRO	1.15	0.86	-0.24	0.77	0.68	1.44	1.16	0.35	0.67	0.87	1.10	0.85	0.05	0.77	0.78
CAX	Am	EBF	1.89	1.79	-0.59	-	-1.35	1.40	1.40	-0.05	-	-0.83	1.63	1.60	-0.32	-	-1.09
CRA	Cfa	CRO	1.25	0.86	-0.32	0.61	0.67	2.48	1.80	-1.53	0.24	0.52	1.57	1.05	-0.71	0.49	0.58
CSE1	Aw	CSH	3.27	2.99	2.93	0.02	-0.13	2.97	2.70	2.41	0.06	-0.29	3.14	2.90	2.75	0.05	-0.23
CSE2	Aw	OSH	0.95	0.73	0.50	0.01	0.13	1.67	1.28	1.14	0.05	-1.22	1.08	0.85	0.67	0.00	-0.06
CSJ	Aw	WSA	1.53	1.34	-0.11	0.01	0.03	1.43	1.20	-0.38	0.16	0.21	1.41	1.27	-0.24	0.13	0.12
CST	BSh	WSA	2.85	2.69	2.66	0.41	0.74	2.56	2.14	2.12	0.23	0.70	2.63	2.44	2.42	0.39	0.69
FEX	Aw	GRA	0.92	0.74	0.14	0.24	0.56	2.42	2.29	2.03	0.15	0.69	1.31	1.20	1.03	0.28	0.54
FMI	Aw	SAV	1.05	0.81	-0.10	0.30	0.47	2.11	1.81	1.75	0.23	0.46	1.27	1.04	0.77	0.31	0.44
FNS	Aw	GRA	0.62	0.50	0.01	0.64	0.45	2.15	1.90	-1.39	0.03	0.30	0.98	0.77	-0.37	0.21	0.37
FSN	Am	GRA	1.18	1.00	-0.04	0.16	0.16	1.81	1.50	-1.20	0.16	0.40	1.28	1.09	-0.54	0.19	0.25
GUY	Am	EBF	1.33	1.10	-0.08	0.02	-0.22	2.46	1.87	-1.67	0.31	1.58	1.43	1.20	-0.77	0.14	0.68
K34	Am	EBF	1.58	1.31	-1.05	0.05	0.17	1.32	1.02	-0.26	0.10	0.29	1.42	1.18	-0.76	0.07	0.24
K67	Am	EBF	1.37	1.11	0.48	0.13	0.67	1.39	1.16	0.57	0.11	0.62	1.28	1.10	0.50	0.16	0.69
K77	Am	CRO	1.15	0.78	-0.47	0.23	0.31	1.64	1.31	-0.75	0.00	0.06	1.29	0.90	-0.63	0.16	0.28
K83	Am	EBF	1.26	1.03	0.22	0.07	0.35	1.72	1.41	0.00	0.23	1.05	1.39	1.18	0.07	0.15	0.67
NPW	Aw	WSA	1.42	1.23	0.37	0.15	0.42	1.47	1.08	0.50	0.29	0.70	1.30	1.06	0.39	0.26	0.56
PAN	Aw	SAV	0.82	0.65	-0.19	0.60	0.49	1.70	1.34	0.88	0.23	0.63	0.99	0.75	0.36	0.46	0.57
PAS	Cfa	GRA	0.98	0.74	-0.29	0.71	0.68	3.17	2.73	-2.61	0.16	0.32	1.49	1.14	-0.95	0.56	0.52
PDG	Cwb	WSA	2.79	2.17	1.15	0.12	-0.24	2.39	1.95	1.11	0.00	0.00	2.82	2.18	0.84	0.10	-0.26
PRS	Cfa	CRO	1.51	1.00	-0.71	0.56	0.66	1.91	1.24	-0.43	0.26	0.57	1.57	0.98	-0.56	0.50	0.68
QFR	Af	WSA	0.26	0.25	0.05	-	0.50	2.86	2.86	-2.86	-	1.35	1.40	1.40	-1.40	-	0.93
RJA	Aw	EBF	1.42	1.08	-0.50	0.02	0.13	1.54	1.10	-0.22	0.01	0.13	1.40	1.06	-0.44	0.03	0.15
SIN	Aw	EBF	1.39	1.12	0.90	0.06	0.34	1.44	1.23	1.20	0.13	0.38	1.28	1.09	1.02	0.15	0.38
SJA	Cfb	EBF	1.85	1.50	1.47	0.01	0.14	2.65	2.39	2.39	0.05	0.35	2.22	1.93	1.93	0.03	0.24
SLU	BSk	MF	2.52	2.24	2.21	0.28	1.00	2.51	2.29	2.03	0.02	0.26	2.34	2.11	2.11	0.21	0.66
SMA	Cfa	GRA	1.67	1.41	-1.39	0.71	0.62	2.55	1.73	-0.68	0.08	0.37	1.87	1.26	-1.11	0.32	0.51
TF1	ET	WET	-	-	-	-	-	1.38	1.01	-0.71	0.14	0.30	1.38	1.01	-0.71	0.14	0.30
TF2	EF	WET	1.36	1.03	-1.02	0.10	0.15	1.69	1.52	-0.89	0.41	-0.44	1.50	1.28	-0.96	0.09	-0.15
TNR	Aw	EBF	1.17	0.97	-0.61	0.25	0.47	1.06	0.89	0.27	0.06	0.36	1.07	0.88	-0.28	0.22	0.47
USE	Cwb	CRO	0.77	0.58	-0.32	0.78	1.12	1.11	0.77	-0.36	0.48	0.99	0.77	0.57	-0.34	0.76	1.07
VIR	Cfa	ENF	1.86	1.55	-0.39	0.21	0.46	1.48	1.18	0.01	0.41	0.63	1.60	1.32	-0.23	0.32	0.53

Chapter 7: Final Conclusions

This dissertation is centered on the estimation of actual evapotranspiration (ET) at multiple spatial resolutions using remote sensing data, and provides information on a major knowledge gap in water management. An adaptation of the Surface Energy Balance Algorithm for Land (SEBAL) model was developed in the Google Earth Engine platform for continental actual ET estimation. The resulting dataset, named geeSEBAL-MODIS, covers South America between 2002 and 2022, with pixels of 500 m and images every 8 days. The study shows the potential of cloud computing and provides a pathway for global ET estimation using complex and physically sound models.

The geeSEBAL-MODIS dataset was validated through ground measurements, regional estimates based on river basin water balance, and global ET datasets. Results indicate that the dataset is sensitive to ET seasonality, temporal anomalies, and spatial heterogeneity, making it useful for land and climate monitoring. Furthermore, it yielded lower errors than existing global ET datasets at both field and basin scales, improving trust in remotely sensed ET data.

An improvement for the soil heat flux (G) estimation was also developed, using artificial neural networks (ANN). The assessment of G from ANN shows potential for further improvements in surface energy balance fluxes modeling, leading to better ET estimations. Future comparison studies between geeSEBAL-MODIS estimates and local measured data are recommended for G and other surface energy balance fluxes, as well as ET fraction and reference evapotranspiration.

The positive results of the continental version of geeSEBAL motivate the expansion of the model's coverage towards global application. However, several steps are required before this goal can be accomplished, including model validation in climatic conditions found outside of South America, further verification and improvements to the LST normalization method, endmember selection, and estimation of SEB fluxes in complex surface and climate conditions, as well as resource acquisition regarding computational power and memory.

Moreover, the LST normalization procedure presented in this work enabled the geeSEBAL model computation for the very large scale, paving the way for continental (or global) application of other contextual ET models. Given the forecasted discontinuation of MODIS image acquisition, an implementation of geeSEBAL with VIIRS imagery is necessary to ensure the construction of longer monitoring series.

The novel geeSEBAL-MODIS dataset was utilized to explore trends in ET and other hydrologically significant variables. The findings indicate a general rise in ET due to factors

such as global warming, atmospheric drying, increased solar radiation at the Earth's surface, and potentially, atmospheric carbon dioxide fertilization. As precipitation dwindles across the continent, projections for the future suggest that water resources availability will be strained in many parts of South America, particularly in central and southeastern Brazil. This finding highlights the need for proactive measures to address water scarcity in these regions.

In the final chapter of this dissertation, a reproduction of the OpenET approach for intercomparison of remote sensing and in situ ET measurements is presented. Despite implementing only two models, as opposed to the original project's six models, this study successfully demonstrated the feasibility of replicating OpenET beyond the borders of the United States. Continuation of this research is highly recommended, by increasing the number of models tested, the number of flux tower sites, as well as creating a feedback framework to improve remote sensing estimations. Establishing the OpenET ensemble as reliable source for ET data would be a significant contribution to the field of water management in South America, as much as has been in the United States, aiding in the development of effective water management policies and strategies.

This dissertation approached ET modeling using various remote sensing sources and methods, providing multi-scale ET data, thereby providing water managers with essential information at the local and regional levels. By employing this multifaceted approach to ET modeling, water managers can make better informed decisions about water allocation, usage, and conservation, thereby contributing to the sustainable management of this vital resource.

References

A. IRMAK *et al.* Comparison and Analysis of Empirical Equations for Soil Heat Flux for Different Cropping Systems and Irrigation Methods. **Transactions of the ASABE**, [*s. l.*], v. 54, n. 1, 2011.

ABATZOGLOU, J. T. *et al.* TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. **Scientific Data**, [*s. l.*], v. 5, 2018.

ADADI, A.; BERRADA, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). **IEEE Access**, [*s. l.*], v. 6, 2018.

AGUIAR, L. J. G. **Fluxos de massa e energia para a cultura de milho (Zea mays L.) no Rio Grande do Sul**. 2011. - Universidade Federal de Viçosa, Viçosa, 2011.

AGUILAR, A. L. *et al.* Performance assessment of MOD16 in evapotranspiration evaluation in Northwestern Mexico. **Water (Switzerland)**, [*s. l.*], v. 10, n. 7, 2018.

ALFIERI, J. G.; KUSTAS, W. P.; ANDERSON, M. C. A Brief Overview of Approaches for Measuring Evapotranspiration. **Agronomy Monographs**, [*s. l.*], p. 109–127, 2018.

ALLEN, R. G. *et al.* A landsat-based energy balance and evapotranspiration model in Western US water rights regulation and planning. **Irrigation and Drainage Systems**, [*s. l.*], v. 19, n. 3–4, p. 251–268, 2005. Disponível em: https://link.springer.com/article/10.1007/s10795-005-5187-z. Acesso em: 9 abr. 2023.

ALLEN, R. G. *et al.* Conditioning point and gridded weather data under aridity conditions for calculation of reference evapotranspiration. **Agricultural Water Management**, [*s. l.*], v. 245, 2021.

ALLEN, R. G. *et al.* Evapotranspiration information reporting: I. Factors governing measurement accuracy. **Agricultural Water Management**, [*s. l.*], v. 98, n. 6, p. 899–920, 2011. Disponível em: Acesso em: 9 abr. 2023.

ALLEN, R. G. *et al.* **FAO Irrigation and Drainage Paper No. 56 - Crop Evapotranspiration (guidelines for computing crop water requirements)**. Rome: Food and Agriculture Organization of the United Nations, 1998-. ISSN 02545284.

ALLEN, R. G. *et al.* Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)—Applications. **Journal of Irrigation and Drainage Engineering**, [*s. l.*], v. 133, n. 4, 2007.

ALLEN, R. *et al.* Satellite-based ET estimation in agriculture using SEBAL and METRIC. **Hydrological Processes**, [*s. l.*], v. 25, n. 26, p. 4011–4027, 2011.

ALLEN, R. G. *et al.* Sensitivity of Landsat-Scale Energy Balance to Aerodynamic Variability in Mountains and Complex Terrain. **Journal of the American Water Resources Association**, [*s. l.*], v. 49, n. 3, p. 592–604, 2013.

ALLEN, R. G.; TASUMI, M.; TREZZA, R. Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)—Model. **Journal of Irrigation and Drainage Engineering**, [*s. l.*], v. 133, n. 4, p. 380–394, 2007. Disponível em: https://ascelibrary.org/doi/10.1061/%28ASCE%290733-9437%282007%29133%3A4%28380%29. Acesso em: 10 ago. 2022.

ANA. **Atlas irrigação: uso da água na agricultura irrigadaAgência Nacional de Águas e Saneamento Básico**. [*S. l.: s. n.*], 2021a.

ANA. **Estimativas de evapotranspiração real por sensoriamento remoto no Brasil**. [*S. l.: s. n.*], 2020. Disponível em: www.ana.gov.br. .

ANA. **Evaporação líquida de reservatórios artificiais no Brasil**. [*S. l.: s. n.*], 2021b. *E-book*. Disponível em: www.ana.gov.br.

ANABALÓN, A.; SHARMA, A. On the divergence of potential and actual evapotranspiration trends: An assessment across alternate global datasets. **Earth's Future**, [*s. l.*], v. 5, n. 9, 2017.

ANDAM-AKORFUL, S. A. *et al.* Multi-model and multi-sensor estimations of evapotranspiration over the Volta Basin, West Africa. **International Journal of Climatology**, [*s. l.*], v. 35, n. 10, 2015.

ANDERSON, M. C. *et al.* A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales. **Remote Sensing of Environment**, [*s. l.*], v. 112, n. 12, p. 4227–4241, 2008.

ANDERSON, M C *et al.* **A Two-Source Time-Integrated Model for Estimating Surface Fluxes Using Thermal Infrared Remote SensingREMOTE SENS. ENVIRON**. [*S. l.*]: ©Elsevier Science Inc, 1996.

ANDERSON, M. *et al.* Field-scale assessment of land and water use change over the California delta using remote sensing. **Remote Sensing**, [*s. l.*], v. 10, n. 6, 2018.

ANDERSON, M. C. *et al.* Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. **Hydrology and Earth System Sciences**, [*s. l.*], v. 15, n. 1, p. 223–239, 2011.

ANDERSON, M. C. *et al.* Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. **Remote Sensing of Environment**, [*s. l.*], v. 122, 2012.

ANDERSON, M. C.; KUSTAS, W. P.; NORMAN, J. M. Upscaling flux observations from local to continental scales using thermal remote sensing. *Em*: , 2007. **Agronomy Journal**. [*S. l.*: *s. n.*], 2007. p. 240–254.

ARAÚJO, A. C. Comparative measurements of carbon dioxide fluxes from two nearby towers in a central Amazonian rainforest: The Manaus LBA site. **Journal of Geophysical Research**, [*s. l.*], v. 107, n. D20, 2002.

ASCE-EWRI. The ASCE standardized reference evapotranspiration equation: ASCE-EWRI Standardization of Reference Evapotranspiration Task Committe Report. **American Society of Civil Engineers**, [*s. l.*], 2005.

ASHOURI, H. *et al.* PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. **Bulletin of the American Meteorological Society**, [*s. l.*], v. 96, n. 1, p. 69–83, 2015. Disponível em: https://journals.ametsoc.org/view/journals/bams/96/1/bams-d-13-00068.1.xml. Acesso em: 13 mar. 2023.

BALDOCCHI, D. D. **How eddy covariance flux measurements have contributed to our understanding of Global Change Biology**. [*S. l.*: *s. n.*], 2020.

BALDOCCHI, D. *et al.* How Much Water Is Evaporated Across California? A Multiyear Assessment Using a Biophysical Model Forced With Satellite Remote Sensing Data. **Water Resources Research**, [*s. l.*], v. 55, n. 4, 2019.

BANDARA, K. M. P. S. Monitoring irrigation performance in Sri Lanka with high-frequency satellite measurements during the dry season. **Agricultural Water Management**, [*s. l.*], v. 58, n. 2, 2003.

BASTIAANSSEN, W. G. M. *et al.* A remote sensing surface energy balance algorithm for land (SEBAL): 1. Formulation. **Journal of Hydrology**, [*s. l.*], v. 212–213, n. 1–4, 1998a.

BASTIAANSSEN, W. G. M. *et al.* A remote sensing surface energy balance algorithm for land (SEBAL): 2. Validation. **Journal of Hydrology**, [*s. l.*], v. 212–213, n. 1–4, 1998b.

BASTIAANSSEN, W. *et al.* Discussion of "Application of SEBAL Model for Mapping Evapotranspiration and Estimating Surface Energy Fluxes in South-Central Nebraska" by Ramesh K. Singh, Ayse Irmak, Suat Irmak, and Derrel L. Martin. **Journal of Irrigation and Drainage Engineering**, [*s. l.*], v. 136, n. 4, 2010.

BASTIAANSSEN, W. G. M. **Regionalization of surface flux densities and moisture indicators in composite terrain: a remote sensing approach under clear skies in Mediterranean climates**. [*S. l.: s. n.*], 1995.

BASTIAANSSEN, W. G. M. *et al.* SEBAL Model with Remotely Sensed Data to Improve Water-Resources Management under Actual Field Conditions. **Journal of Irrigation and Drainage Engineering**, [*s. l.*], v. 131, n. 1, 2005.

BASTIAANSSEN, W. G. M.; ALI, S. A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. **Agriculture, Ecosystems and Environment**, [*s. l.*], v. 94, n. 3, 2003.

BHATTARAI, N. *et al.* A new optimized algorithm for automating endmember pixel selection in the SEBAL and METRIC models. **Remote Sensing of Environment**, [*s. l.*], v. 196, p. 178–192, 2017.

BHATTARAI, N. *et al.* An automated multi-model evapotranspiration mapping framework using remotely sensed and reanalysis data. **Remote Sensing of Environment**, [*s. l.*], v. 229, p. 69–92, 2019.

BHATTARAI, N. *et al.* Evaluating five remote sensing based single-source surface energy balance models for estimating daily evapotranspiration in a humid subtropical climate. **International Journal of Applied Earth Observation and Geoinformation**, [*s. l.*], v. 49, 2016.

BHATTARAI, N.; MALLICK, K.; JAIN, M. Sensitivity of four contextual remote sensing based surface energy balance models to spatial domain. *Em*: , 2019. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives**. [*S. l.: s. n.*], 2019.

BIGGS, T. W.; MARSHALL, M.; MESSINA, A. Mapping daily and seasonal evapotranspiration from irrigated crops using global climate grids and satellite imagery: Automation and methods comparison. **Water Resources Research**, [*s. l.*], v. 52, n. 9, 2016.

BIUDES, M. S. *et al.* Patterns of energy exchange for tropical ecosystems across a climate gradient in Mato Grosso, Brazil. **Agricultural and Forest Meteorology**, [*s. l.*], v. 202, 2015.

BLOCKINESS ARTIFACT IN LANDSAT COLLECTION 2 SURFACE TEMPERATURE PRODUCTS | U.S. GEOLOGICAL SURVEY. [*S. l.*], [*s. d.*]. Disponível em: https://www.usgs.gov/landsat-missions/blockiness-artifact-landsat-collection-2-surface-temperature-products. Acesso em: 6 abr. 2023.

BORGES, C. K. *et al.* **Seasonal variation of surface radiation and energy balances over two contrasting areas of the seasonally dry tropical forest (Caatinga) in the Brazilian semi-arid**. [*S. l.: s. n.*], 2020.

BORMA, L. S. *et al.* Atmosphere and hydrological controls of the evapotranspiration over a floodplain forest in the Bananal Island region, Amazonia. **Journal of Geophysical Research: Biogeosciences**, [*s. l.*], v. 114, n. 1, 2009.

BOWEN, I. S. The ratio of heat losses by conduction and by evaporation from any water surface. **Physical Review**, [*s. l.*], v. 27, n. 6, 1926.

BOYLAN, J. W.; RUSSELL, A. G. PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. **Atmospheric Environment**, [*s. l.*], v. 40, n. 26, 2006.

BRÊDA, J. P. L. F. *et al.* Climate change impacts on South American water balance from a continental-scale hydrological model driven by CMIP5 projections. **Climatic Change**, [*s. l.*], v. 159, n. 4, 2020.

BRITO, T. R. do C. *et al.* Mudanças no Uso da Terra e Efeito nos Componentes do Balanço Hídrico no Agreste Pernambucano. **Revista Brasileira de Geografia Física**, [*s. l.*], v. 13, n. 2, 2020.

BRUTSAERT, W.; STRICKER, H. An advection-aridity approach to estimate actual regional evapotranspiration. **Water Resources Research**, [*s. l.*], v. 15, n. 2, p. 443–450, 1979. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1029/WR015i002p00443. Acesso em: 1 abr. 2023.

BUCHHORN, M. *et al.* Copernicus Global Land Cover Layers—Collection 2. **Remote Sensing 2020, Vol. 12, Page 1044**, [*s. l.*], v. 12, n. 6, p. 1044, 2020. Disponível em: https://www.mdpi.com/2072-4292/12/6/1044/htm. Acesso em: 4 abr. 2023.

BUDYKO, M. I. (Mikhail I.; MILLER, D. H. Climate and life. [*s. l.*], p. 508, 1974. Disponível em: Acesso em: 31 mar. 2023.

BURBA, G. **Eddy Covariance Method-for Scientific, Industrial, Agricultural, and Regulatory Applications**. [*S. l.: s. n.*], 2013.

CABALLERO, C. B.; RUHOFF, A.; BIGGS, T. Land use and land cover changes and their impacts on surface-atmosphere interactions in Brazil: A systematic review. **Science of The Total Environment**, [*s. l.*], v. 808, p. 152134, 2022. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0048969721072107. Acesso em: 2 abr. 2023.

CABRAL, O. M. R. *et al.* Fluxos turbulentos de calor sens'ivel, vapor d'água e CO2 sobre plantação de cana-de-açúcar (Saccharum sp.) em Sertãozinho-SP. **Revista Brasileira de Meteorologia**, [*s. l.*], v. 18, p. 61–70, 2003.

CAI, W. *et al.* **Climate impacts of the El Niño–Southern Oscillation on South America**. [*S. l.: s. n.*], 2020.

CAMMALLERI, C.; LA LOGGIA, G.; MALTESE, A. Critical analysis of empirical ground heat flux equations on a cereal field using micrometeorological data. *Em*: , 2009. **Remote Sensing for Agriculture, Ecosystems, and Hydrology XI**. [*S. l.: s. n.*], 2009.

CANELÓN, D. J.; CHÁVEZ, J. L. Soil heat flux modeling using artificial neural networks and multispectral airborne remote sensing imagery. **Remote Sensing**, [*s. l.*], v. 3, n. 8, 2011.

CAO, C. *et al.* Suomi NPP VIIRS sensor data record verification, validation, and long-term performance monitoring. **Journal of Geophysical Research Atmospheres**, [*s. l.*], v. 118, n. 20, 2013.

CARNEIRO, J. V. **Estimativas das trocas líquidas de carbono em duas áreas de cultivo de arroz irrigado na região central do RS**. 2012. - Universidade Federal de Santa Maria, Santa Maria, 2012.

CARNEIRO, J. *et al.* NEE SOBRE UMA CULTURA DE ARROZ IRRIGADO NO SUL DO BRASIL. **Ciência e Natura**, [*s. l.*], v. 0, n. 0, 2013.

CARVALHO, L. M. V. **Assessing precipitation trends in the Americas with historical data: A review**. [*S. l.: s. n.*], 2020.

CAVALCANTE, R. B. L. *et al.* Opposite Effects of Climate and Land Use Changes on the Annual Water Balance in the Amazon Arc of Deforestation. **Water Resources Research**, [*s. l.*], v. 55, n. 4, 2019.

CHEN, S. Y. *et al.* A study on the normal annual variation field of land surface temperature in China. **Acta Geophysica Sinica**, [*s. l.*], v. 52, n. 9, 2009.

CHEN, M. *et al.* Uncertainty analysis of the Operational Simplified Surface Energy Balance (SSEBop) model at multiple flux tower sites. **Journal of Hydrology**, [*s. l.*], v. 536, p. 384–399, 2016.

CHEN, J.; ADAMS, B. J. Integration of artificial neural networks with conceptual models in rainfall-runoff modeling. **Journal of Hydrology**, [*s. l.*], v. 318, n. 1–4, p. 232–249, 2006.

CHENG, M. *et al.* Long time series of daily evapotranspiration in China based on the SEBAL model and multisource images and validation. **Earth System Science Data**, [*s. l.*], v. 13, n. 8, 2021.

CLARK, E. A. *et al.* Continental runoff into the oceans (1950-2008). **Journal of Hydrometeorology**, [*s. l.*], v. 16, n. 4, 2015.

CLEUGH, H. A. *et al.* Regional evaporation estimates from flux tower and MODIS satellite data. **Remote Sensing of Environment**, [*s. l.*], v. 106, n. 3, 2007.

COLLISCHONN, W. *et al.* The MGB-IPH model for large-scale rainfall-runoff modelling. **Hydrological Sciences Journal**, [*s. l.*], v. 52, n. 5, 2007.

COMINI DE ANDRADE, B. *et al.* geeSEBAL-MODIS: Continental scale evapotranspiration based on the surface energy balance for South America. [*s. l.*], 2023.

COOK, B. I. *et al.* Twenty-First Century Drought Projections in the CMIP6 Forcing Scenarios. **Earth's Future**, [*s. l.*], v. 8, n. 6, 2020.

CROWTHER, P. S.; COX, R. J. A method for optimal division of data sets for use in neural networks. *Em*: , 2005. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [*S. l.: s. n.*], 2005.

DA ROCHA, H. R. *et al.* Patterns of water and heat flux across a biome gradient from tropical forest to savanna in brazil. **Journal of Geophysical Research: Biogeosciences**, [*s. l.*], v. 114, n. 1, 2009.

DALMAGRO, H. J. *et al.* Radiative forcing of methane fluxes offsets net carbon dioxide uptake for a tropical flooded forest. **Global Change Biology**, [*s. l.*], v. 25, n. 6, 2019.

DANELICHEN, V. H. M. *et al.* Estimating of gross primary production in an Amazon-cerrado transitional forest using MODIS and landsat imagery. **Anais da Academia Brasileira de Ciencias**, [*s. l.*], v. 87, n. 3, 2015.

DANELICHEN, V. H. D. M. *et al.* Estimation of soil heat flux in a neotropical wetland region using remote sensing techniques. **Revista Brasileira de Meteorologia**, [*s. l.*], v. 29, n. 4, 2014.

DARAMOLA, M. T.; XU, M. Recent changes in global dryland temperature and precipitation. **International Journal of Climatology**, [*s. l.*], v. 42, n. 2, 2022.

DAVIDSON, E. A.; ARTAXO, P. Globally significant changes in biological processes of the Amazon Basin: Results of the large-scale Biosphere-Atmosphere Experiment. **Global Change Biology**, [*s. l.*], v. 10, n. 5, 2004.

DE ANDRADE, B. C. C. *et al.* Artificial Neural Network Model of Soil Heat Flux over Multiple Land Covers in South America. **Remote Sensing**, [*s. l.*], v. 13, n. 12, 2021a.

DE ANDRADE, B. C. C. **ESTIMATIVA DA EVAPOTRANSPIRAÇÃO REAL VIA SENSORIAMENTO REMOTO**. 2018. [*s. l.*], 2018.

DE ANDRADE, B. C. C. *et al.* Remote sensing-based actual evapotranspiration assessment in a data-scarce area of Brazil: A case study of the Urucuia Aquifer System. **International Journal of Applied Earth Observation and Geoinformation**, [*s. l.*], v. 98, 2021b.

DHUNGEL, S.; BARBER, M. E. Estimating calibration variability in evapotranspiration derived from a satellite-based energy balance model. **Remote Sensing**, [*s. l.*], v. 10, n. 11, 2018.

DIAZ, M. B. *et al.* Dynamics of the superficial fluxes over a flooded rice paddy in southern Brazil. **Agricultural and Forest Meteorology**, [*s. l.*], v. 276–277, 2019.

DIDAN, K. **MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data Set]**. [*S. l.*], 2015a.

DIDAN, K. MYD13A2 MODIS/Aqua Vegetation Indices 16-Day L3 Global 1km SIN Grid V006 [Data set]. **NASA EOSDIS Land Processes DAAC**, [*s. l.*], 2015b.

DREXLER, J. Z. *et al.* A review of models and micrometeorological methods used to estimate wetland evapotranspiration. **Hydrological Processes**, [*s. l.*], v. 18, n. 11, p. 2071–2101, 2004. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1002/hyp.1462. Acesso em: 6 abr. 2023.

ELNASHAR, A. *et al.* Synthesis of global actual evapotranspiration from 1982 to 2019. **Earth System Science Data**, [*s. l.*], v. 13, n. 2, 2021.

ESPINOZA VILLAR, J. C. *et al.* Contrasting regional discharge evolutions in the Amazon basin (1974-2004). **Journal of Hydrology**, [*s. l.*], v. 375, n. 3–4, 2009.

EVA, H. D. *et al.* A land cover map of South America. **Global Change Biology**, [*s. l.*], v. 10, n. 5, 2004.

FAO. **WaPOR V2 quality assessment – Technical Report on the Data Quality of the WaPOR FAO Database version 2**. [*S. l.*]: FAO, 2020.

FASSONI-ANDRADE, A. C. *et al.* **Amazon Hydrology From Space: Scientific Advances and Future Challenges**. [*S. l.: s. n.*], 2021.

FENG, J. *et al.* Estimating surface heat and water vapor fluxes by combining two-source energy balance model and back-propagation neural network. **Science of the Total Environment**, [*s. l.*], v. 729, 2020.

FENG, X. *et al.* Revegetation in China's Loess Plateau is approaching sustainable water resource limits. **Nature Climate Change**, [*s. l.*], v. 6, n. 11, 2016.

FISHER, J. B. *et al.* ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. **Water Resources Research**, [*s. l.*], v. 56, n. 4, 2020a.

FISHER, J. B. *et al.* ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. **Water Resources Research**, [*s. l.*], v. 56, n. 4, 2020b.

FISHER, J. B. *et al.* The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. **Water Resources Research**, [*s. l.*], v. 53, n. 4, p. 2618–2626, 2017. Disponível em: Acesso em: 25 ago. 2022.

FISHER, J. B.; TU, K. P.; BALDOCCHI, D. D. 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**, [*s. l.*], v. 112, n. 3, 2008.

FRIEDL, M.; SULLA-MENASHE, D. MCD12Q1 MODIS/Terra+ aqua land cover type yearly L3 global 500m SIN grid V006 [data set]. **NASA EOSDIS Land Processes DAAC**, [*s. l.*], v. 10, 2015.

FUNK, C. *et al.* The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. **Scientific Data**, [*s. l.*], v. 2, 2015.

GARCÍA, A. G. *et al.* Patterns and controls of carbon dioxide and water vapor fluxes in a dry forest of central Argentina. **Agricultural and Forest Meteorology**, [*s. l.*], v. 247, 2017.

GARREAUD, R. D. *et al.* Present-day South American climate. **Palaeogeography, Palaeoclimatology, Palaeoecology**, [*s. l.*], v. 281, n. 3–4, 2009.

GARREAUD, R. D.; ACEITUNO, P. Interannual rainfall variability over the South American Altiplano. **Journal of Climate**, [*s. l.*], v. 14, n. 12, 2001.

GARSON, D. Interpreting neural-network connection weights. **AI Expert**, [*s. l.*], 1991. Disponível em: https://dl.acm.org/doi/10.5555/129449.129452. Acesso em: 4 dez. 2022.

GASPAR, M. T. P.; CAMPOS, J. E. G.; DE MORAES, R. A. V. Determinação das espessuras do Sistema Aquífero Urucuia a partir de estudo geofísico. **Revista Brasileira de Geociencias**, [*s. l.*], v. 42, n. SUPPL.1, 2012.

GATTI, L. v. *et al.* Amazonia as a carbon source linked to deforestation and climate change. **Nature**, [*s. l.*], v. 595, n. 7867, p. 388–393, 2021.

GOH, A. T. C. Back-propagation neural networks for modeling complex systems. **Artificial Intelligence in Engineering**, [*s. l.*], v. 9, n. 3, 1995.

GONÇALVES, R. D. *et al.* Using GRACE to quantify the depletion of terrestrial water storage in Northeastern Brazil: The Urucuia Aquifer System. **Science of the Total Environment**, [*s. l.*], v. 705, 2020.

GORELICK, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, [*s. l.*], v. 202, p. 18–27, 2017.

GOSWAMI, B. N. *et al.* Increasing trend of extreme rain events over India in a warming environment. **Science**, [*s. l.*], v. 314, n. 5804, 2006.

GOULDEN, M. L. *et al.* Diel and seasonal patterns of tropical forest CO2 exchange. **Ecological Applications**, [*s. l.*], v. 14, n. 4 SUPPL., 2004.

GOULDEN, M. L.; BALES, R. C. California forest die-off linked to multi-year deep soil drying in 2012–2015 drought. **Nature Geoscience**, [*s. l.*], v. 12, n. 8, 2019.

GRIFFIS, T. J. *et al.* Hydrometeorological sensitivities of net ecosystem carbon dioxide and methane exchange of an Amazonian palm swamp peatland. **Agricultural and Forest Meteorology**, [*s. l.*], v. 295, 2020.

GUPTA, H. v. *et al.* Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. **Journal of Hydrology**, [*s. l.*], v. 377, n. 1–2, p. 80–91, 2009. Disponível em: Acesso em: 5 dez. 2022.

H. L. PENMAN. Natural Evaporation from Open Water, Bare Soil and Grass. **Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences**, [*s. l.*], v. 193, n. 1032, 1948.

HAMED ALEMOHAMMAD, S. *et al.* Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence. **Biogeosciences**, [*s. l.*], v. 14, n. 18, 2017.

HAMIM. Photosynthesis of C3 and C4 Species in Response to Increased CO2 Concentration and Drought Stress. **HAYATI Journal of Biosciences**, [*s. l.*], v. 12, n. 4, p. 131–138, 2005. Disponível em: Acesso em: 3 abr. 2023.

HASLER, N.; AVISSAR, R. What controls evapotranspiration in the Amazon basin?. **Journal of Hydrometeorology**, [*s. l.*], v. 8, n. 3, 2007.

HENGL, T. Soil texture classes (USDA system) for 6 soil depths (0, 10, 30, 60, 100 and 200 cm) at 250 m. [*s. l.*], 2018. Disponível em: https://zenodo.org/record/2525817. Acesso em: 12 set. 2022.

HERSBACH, H. *et al.* The ERA5 global reanalysis. **Quarterly Journal of the Royal Meteorological Society**, [*s. l.*], v. 146, n. 730, 2020.

HEUSINKVELD, B. G. *et al.* Surface energy balance closure in an arid region: Role of soil heat flux. **Agricultural and Forest Meteorology**, [*s. l.*], v. 122, n. 1–2, 2004.

HOLL, D. *et al.* Cushion bogs are stronger carbon dioxide net sinks than moss-dominated bogs as revealed by eddy covariance measurements on Tierra del Fuego, Argentina. **Biogeosciences**, [*s. l.*], v. 16, n. 17, 2019.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. **Neural Networks**, [*s. l.*], v. 2, n. 5, 1989.

HUFFMAN, G. J. *et al.* Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). Algorithm Theoretical Basis Document (ATBD) Version 06. **Global Precipitation Measurement (GPM) National Aeronautics and Space Administration (NASA) Prepared**, [*s. l.*], n. March, 2019.

HUNSAKER, D. J. *et al.* Estimating cotton evapotranspiration crop coefficients with a multispectral vegetation index. **Irrigation Science**, [*s. l.*], v. 22, n. 2, p. 95–104, 2003.

HUNTINGTON, J. L. *et al.* Evaluating the complementary relationship for estimating evapotranspiration from arid shrublands. **Water Resources Research**, [*s. l.*], v. 47, n. 5, 2011a.

HUNTINGTON, J. L. *et al.* Evaluating the complementary relationship for estimating evapotranspiration from arid shrublands. **Water Resources Research**, [*s. l.*], v. 47, n. 5, p.

5533, 2011b. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1029/2010WR009874. Acesso em: 1 abr. 2023.

IDSO, S. B.; AASE, J. K.; JACKSON, R. D. Net radiation - soil heat flux relations as influenced by soil water content variations. **Boundary-Layer Meteorology**, [*s. l.*], v. 9, n. 1, 1975.

IM, E. S.; PAL, J. S.; ELTAHIR, E. A. B. Deadly heat waves projected in the densely populated agricultural regions of South Asia. **Science Advances**, [*s. l.*], v. 3, n. 8, 2017.

IPCC. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Em*: PÖRTNER, H. O. *et al.* (org.). **Climate Change 2022: Impacts, Adaptation and Vulnerability**. Cambridge, UK and New York, NY, USA: Cambridge University Press, 2022.

JAAFAR, H.; MOURAD, R.; SCHULL, M. A global 30-m ET model (HSEB) using harmonized Landsat and Sentinel-2, MODIS and VIIRS: Comparison to ECOSTRESS ET and LST. **Remote Sensing of Environment**, [*s. l.*], v. 274, p. 112995, 2022. Disponível em: Acesso em: 5 dez. 2022.

JACKSON, R. D. *et al.* Evaluating evaporation from field crops using airborne radiometry and ground-based meteorological data. **Irrigation Science**, [*s. l.*], v. 8, n. 2, 1987.

JARVIS, A. *et al.* Hole-filled SRTM for the globe Version 4. **available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org)**, [*s. l.*], v. 15, 2008.

JAVADIAN, M. *et al.* Global trends in evapotranspiration dominated by increases across large cropland regions. **Remote Sensing**, [*s. l.*], v. 12, n. 7, 2020.

JIANG, Y.; LIU, Z. remote sensing Evaluations of Remote Sensing-Based Global Evapotranspiration Datasets at Catchment Scale in Mountain Regions. [*s. l.*], 2021. Disponível em: https://doi.org/10.3390/rs13245096.

JIANG, C.; RYU, Y. Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS). **Remote Sensing of Environment**, [*s. l.*], v. 186, 2016.

JIMENO-SÁEZ, P. *et al.* A comparison of SWAT and ANN models for daily runoff simulation in different climatic zones of peninsular Spain. **Water (Switzerland)**, [*s. l.*], v. 10, n. 2, 2018.

JUNG, M. *et al.* Recent decline in the global land evapotranspiration trend due to limited moisture supply. **Nature**, [*s. l.*], v. 467, n. 7318, 2010.

KÄFER, P. S. *et al.* Artificial neural networks model based on remote sensing to retrieve evapotranspiration over the Brazilian Pampa. **Journal of Applied Remote Sensing**, [*s. l.*], v. 14, n. 03, 2020.

KALMA, J. D.; JUPP, D. L. B. Estimating evaporation from pasture using infrared thermometry: evaluation of a one-layer resistance model. **Agricultural and Forest Meteorology**, [*s. l.*], v. 51, n. 3–4, 1990.

KAYSER, R. H. *et al.* Assessing geeSEBAL automated calibration and meteorological reanalysis uncertainties to estimate evapotranspiration in subtropical humid climates. **Agricultural and Forest Meteorology**, [*s. l.*], v. 314, 2022.

KENDALL, M. G. Rank Correlation Methods (4th edn.) Charles Griffin. **San Francisco, CA**, [*s. l.*], 1975.

KHAND, K. *et al.* Dry season evapotranspiration dynamics over human-impacted landscapes in the southern Amazon using the landsat-based METRIC model. **Remote Sensing**, [*s. l.*], v. 9, n. 7, 2017.

KILIC, A. *et al.* Global production and free access to Landsat-scale Evapotranspiration with EEFlux and eeMETRIC. *Em*: , 2020. **6th Decennial National Irrigation Symposium**. [*S. l.: s. n.*], 2020.

KIM, Y. *et al.* Canopy skin temperature variations in relation to climate, soil temperature, and carbon flux at a ponderosa pine forest in central Oregon. **Agricultural and Forest Meteorology**, [*s. l.*], v. 226–227, 2016.

KIPTALA, J. K. *et al.* Mapping evapotranspiration trends using MODIS and SEBAL model in a data scarce and heterogeneous landscape in Eastern Africa. **Water Resources Research**, [*s. l.*], v. 49, n. 12, 2013.

KOLARS, K. *et al.* Using Eddy Covariance, Soil Water Balance, and Photosynthetically Active Radiation Methods for Corn Evapotranspiration Measurements in the Red River Valley. **American Society of Agricultural and Biological Engineers Annual International Meeting 2013, ASABE 2013**, [*s. l.*], v. 1, p. 781–791, 2013. Disponível em: Acesso em: 6 abr. 2023.

KUSTAS, W. P.; DAUGHTRY, C. S. T. Estimation of the soil heat flux/net radiation ratio from spectral data. **Agricultural and Forest Meteorology**, [*s. l.*], v. 49, n. 3, 1990.

KUSTAS, W. P.; DAUGHTRY, C. S. T.; VAN OEVELEN, P. J. Analytical treatment of the relationships between soil heat flux/net radiation ratio and vegetation indices. **Remote Sensing of Environment**, [*s. l.*], v. 46, n. 3, 1993.

KUSTAS, W. P.; NORMAN, J. M. 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**, [*s. l.*], v. 94, n. 1, 1999.

KUSTAS, W. P.; NORMAN, J. M. Utilisation de la télédétection pour le suivi de l'évapotranspiration sur les terres. **Hydrological Sciences Journal**, [*s. l.*], v. 41, n. 4, p. 495–516, 1996.

LAIPELT, L. *et al.* Assessment of an automated calibration of the SEBAL Algorithm to estimate dry-season surface-energy partitioning in a Forest-Savanna Transition in Brazil. **Remote Sensing**, [*s. l.*], v. 12, n. 7, 2020.

LAIPELT, L. *et al.* Long-term monitoring of evapotranspiration using the SEBAL algorithm and Google Earth Engine cloud computing. **ISPRS Journal of Photogrammetry and Remote Sensing**, [*s. l.*], v. 178, p. 81–96, 2021.

LANDERER, F. W.; SWENSON, S. C. Accuracy of scaled GRACE terrestrial water storage estimates. **Water Resources Research**, [*s. l.*], v. 48, n. 4, 2012.

LANDSAT COLLECTION 2 SURFACE TEMPERATURE DATA GAPS DUE TO MISSING ASTER GED | U.S. GEOLOGICAL SURVEY. [*S. l.*], [*s. d.*]. Disponível em: https://www.usgs.gov/landsat-missions/landsat-collection-2-surface-temperature-data-gaps-due-missing-aster-ged. Acesso em: 6 abr. 2023.

LARY, D. J. *et al.* Machine learning in geosciences and remote sensing. **Geoscience Frontiers**, [*s. l.*], v. 7, n. 1, 2016.

LATHUILLIÈRE, M. J. *et al.* Rain-fed and irrigated cropland-atmosphere water fluxes and their implications for agricultural production in Southern Amazonia. **Agricultural and Forest Meteorology**, [*s. l.*], v. 256–257, 2018.

LAZZARA, P.; RANA, G. The crop coefficient (K c) values of the major crops grown under Mediterranean climate. **Italian Journal of Agrometeorology**, [*s. l.*], v. 15, n. 2, 2010.

LEAKEY, A. D. B. Rising atmospheric carbon dioxide concentration and the future of C4 crops for food and fuel. **Proceedings of the Royal Society B: Biological Sciences**, [*s. l.*], v. 276, n. 1666, p. 2333, 2009. Disponível em: /pmc/articles/PMC2690454/. Acesso em: 3 abr. 2023.

LEHMANN, F.; VISHWAKARMA, B. D.; BAMBER, J. How well are we able to close the water budget at the global scale?. **Hydrology and Earth System Sciences**, [*s. l.*], v. 26, n. 1, p. 35–54, 2022. Disponível em: Acesso em: 1 abr. 2023.

LEUNING, R. *et al.* A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation. **Water Resources Research**, [*s. l.*], v. 44, n. 10, 2008.

LI, S. *et al.* Assessment of Multi-Source Evapotranspiration Products over China Using Eddy Covariance Observations. **Remote. Sens.**, [*s. l.*], v. 10, n. 11, 2018. Disponível em: Acesso em: 6 abr. 2023.

LIOU, Y. A.; KAR, S. K. **Evapotranspiration estimation with remote sensing and various surface energy balance algorithms-a review**. [*S. l.*]: MDPI AG, 2014a.

LIOU, Y. A.; KAR, S. K. **Evapotranspiration estimation with remote sensing and various surface energy balance algorithms-a review**. [*S. l.*]: MDPI AG, 2014b.

LIU, Y. *et al.* A framework for actual evapotranspiration assessment and projection based on meteorological, vegetation and hydrological remote sensing products. **Remote Sensing**, [*s. l.*], v. 13, n. 18, 2021.

LIU, H.; ZHANG, Q.; DOWLER, G. Environmental controls on the surface energy budget over a large Southern Inland water in the united states: An analysis of one-year eddy covariance flux data. **Journal of Hydrometeorology**, [*s. l.*], v. 13, n. 6, 2012.

LONG, D.; LONGUEVERGNE, L.; SCANLON, B. R. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. **Water Resources Research**, [*s. l.*], v. 50, n. 2, p. 1131–1151, 2014.

LONG, D.; SINGH, V. P.; LI, Z. L. How sensitive is SEBAL to changes in input variables, domain size and satellite sensor?. **Journal of Geophysical Research Atmospheres**, [*s. l.*], v. 116, n. 21, 2011.

LU, Z. *et al.* Differences among evapotranspiration products affect water resources and ecosystem management in an Australian catchment. **Remote Sensing**, [*s. l.*], v. 11, n. 8, 2019.

LUCCHESE, L. V.; DE OLIVEIRA, G. G.; PEDROLLO, O. C. Attribute selection using correlations and principal components for artificial neural networks employment for landslide susceptibility assessment. **Environmental Monitoring and Assessment**, [*s. l.*], v. 192, n. 2, 2020.

MA, N.; SZILAGYI, J.; ZHANG, Y. Calibration-Free Complementary Relationship Estimates Terrestrial Evapotranspiration Globally. **Water Resources Research**, [*s. l.*], v. 57, n. 9, 2021.

MALBÉTEAU, Y. *et al.* Normalizing land surface temperature data for elevation and illumination effects in mountainous areas: A case study using ASTER data over a steep-sided valley in Morocco. **Remote Sensing of Environment**, [*s. l.*], v. 189, 2017.

MANN, H. B. Non-Parametric Test Against Trend. **Econometrica**, [*s. l.*], v. 13, n. 3, 1945.

MARENGO, J. A. *et al.* **Changes in Climate and Land Use Over the Amazon Region: Current and Future Variability and Trends**. [*S. l.: s. n.*], 2018.

MARENGO, J. A. *et al.* Extreme Seasonal Climate Variations in the Amazon Basin: Droughts and Floods. *Em*: [*S. l.: s. n.*], 2016.

MARENGO, J. A. *et al.* Increased climate pressure on the agricultural frontier in the Eastern Amazonia–Cerrado transition zone. **Scientific Reports**, [*s. l.*], v. 12, n. 1, 2022.

MARENGO, J. A. *et al.* **Recent developments on the South American monsoon system**. [*S. l.: s. n.*], 2012.

MARTENS, B. *et al.* GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. **Geoscientific Model Development**, [*s. l.*], v. 10, n. 5, 2017.

MASEK, J. G. *et al.* Landsat 9: Empowering open science and applications through continuity. **Remote Sensing of Environment**, [*s. l.*], v. 248, 2020.

[MASSON-DELMOTTE, V. *et al.* **IPCC, 2021: Climate Change 2021: The Physical Science BasisCambridge University Press. In Press.** [*S. l.: s. n.*], 2021.

MCCABE, M. F. *et al.* The future of Earth observation in hydrology. **Hydrology and Earth System Sciences**, [*s. l.*], v. 21, n. 7, 2017.

MCCORKEL, J. *et al.* Landsat 9 thermal infrared sensor 2 characterization plan overview. *Em*: , 2018. **International Geoscience and Remote Sensing Symposium (IGARSS)**. [*S. l.: s. n.*], 2018.

MELTON, F. S. *et al.* **OpenET: Filling a Critical Data Gap in Water Management for the Western United StatesJournal of the American Water Resources Association**. [*S. l.: s. n.*], 2021. Disponível em: https://doi.org/10. .

MELTON, F. S. *et al.* Satellite irrigation management support with the terrestrial observation and prediction system: A framework for integration of satellite and surface observations to support improvements in agricultural water resource management. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, [*s. l.*], v. 5, n. 6, p. 1709–1721, 2012. Disponível em: Acesso em: 9 abr. 2023.

MENENTI, M; CHOUDHURY, B. J. **Exchange Processes at the Land Surface for a Range of Space and Time Scales**. [*S. l.*]: IAHS Publ, 1993.

MENENTI, M.; CHOUDHURY, B. J. Parameterization of land surface evaporation by means of location dependent potential evaporation and surface temperature range. **Exchange processes at the land surface for a range of space and time scales. Proc. international symposium, Yokohama, 1993**, [*s. l.*], 1993.

MIRALLES, D. G. *et al.* Magnitude and variability of land evaporation and its components at the global scale. **Hydrology and Earth System Sciences**, [*s. l.*], v. 15, n. 3, 2011.

MISHRA, S. K.; SINGH, V. P. Long-term hydrological simulation based on the Soil Conservation Service curve number. **Hydrological Processes**, [*s. l.*], v. 18, n. 7, 2004.

MONTEITH, J. L. **Evaporation and environment.** [*S. l.: s. n.*], 1965.

MONTEITH, J.; UNSWORTH, M. **Principles of Environmental Physics: Plants, Animals, and the Atmosphere: Fourth Edition**. [*S. l.: s. n.*], 2013.

MOREIRA, V. *et al.* ANÁLISE DA EVAPOTRANSPIRAÇÃO E ÍNDICE DE ÁREA FOLIAR NUMA CULTURA DE SOJA SOB DOIS SISTEMAS DE MANEJO. **Ciência e Natura**, [*s. l.*], v. 0, n. 0, 2013.

MOREIRA, A. A. *et al.* Assessment of terrestrial water balance using remote sensing data in South America. **Journal of Hydrology**, [*s. l.*], v. 575, p. 131–147, 2019.

MOREIRA, V. S. **Balanço de água no ciclo da cultura de soja: representação no modelo de vegetação dinâmica Agro-IBIS**. 2012. - Universidade Federal de Santa Maria, Santa Maria, 2012.

MORTON, C. G. *et al.* Assessing Calibration Uncertainty and Automation for Estimating Evapotranspiration from Agricultural Areas Using METRIC. **Journal of the American Water Resources Association**, [*s. l.*], v. 49, n. 3, p. 549–562, 2013.

MU, Q. *et al.* Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. **Remote Sensing of Environment**, [*s. l.*], v. 111, n. 4, 2007.

MU, Y.; JONES, C. An observational analysis of precipitation and deforestation age in the Brazilian Legal Amazon. **Atmospheric Research**, [*s. l.*], v. 271, p. 106122, 2022. Disponível em: Acesso em: 17 abr. 2023.

MU, Q.; ZHAO, M.; RUNNING, S. W. Evolution of hydrological and carbon cycles under a changing climate. **Hydrological Processes**, [*s. l.*], v. 25, n. 26, 2011a.

MU, Q.; ZHAO, M.; RUNNING, S. W. Improvements to a MODIS global terrestrial evapotranspiration algorithm. **Remote Sensing of Environment**, [*s. l.*], v. 115, n. 8, 2011b.

MUÑOZ SABATER, J. **ERA5-Land hourly data from 1981 to present**. [*S. l.: s. n.*], 2019.

NASH, J. E.; SUTCLIFFE, J. v. River flow forecasting through conceptual models part I - A discussion of principles. **Journal of Hydrology**, [*s. l.*], v. 10, n. 3, 1970.

NOAA CLIMATE PREDICTION CENTER. **Cold and warm episodes by seasons**. [*S. l.*], 2014.

NORMAN, J. M. *et al.* Remote sensing of surface energy fluxes at 101-m pixel resolutions. **Water Resources Research**, [*s. l.*], v. 39, n. 8, 2003.

NORMAN, J. M.; KUSTAS, W. P.; HUMES, K. S. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. **Agricultural and Forest Meteorology**, [*s. l.*], v. 77, n. 3–4, 1995.

OLIVEIRA, M. B. L. *et al.* Fluxos turbulentos de energia sobre o Pantanal Sul Mato-Grossense. **Revista Brasileira de Meteorologia**, [*s. l.*], v. 21, 2006.

OLSON, D. M. *et al.* Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. **BioScience**, [*s. l.*], v. 51, n. 11, 2001.

PARUELO, J. M.; JOBBÁGY, E. G. The Grasslands and Steppes of Patagonia and the Río de la Plata Plains. *Em*: THE PHYSICAL GEOGRAPHY OF SOUTH AMERICA. [*S. l.: s. n.*], 2007.

PASCOLINI-CAMPBELL, M. A.; REAGER, J. T.; FISHER, J. B. GRACE-based Mass Conservation as a Validation Target for Basin-Scale Evapotranspiration in the Contiguous United States. **Water Resources Research**, [*s. l.*], v. 56, n. 2, p. e2019WR026594, 2020. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1029/2019WR026594. Acesso em: 2 abr. 2023.

PASTORELLO, G. *et al.* The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. **Scientific data**, [*s. l.*], v. 7, n. 1, 2020.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, [*s. l.*], v. 11, n. 5, 2007.

PENG, L. *et al.* Determinants of the ratio of actual to potential evapotranspiration. **Global Change Biology**, [*s. l.*], v. 25, n. 4, 2019.

PEREIRA, L. S. *et al.* Prediction of crop coefficients from fraction of ground cover and height. Background and validation using ground and remote sensing data. **Agricultural Water Management**, [*s. l.*], v. 241, p. 106197, 2020. Disponível em: Acesso em: 9 abr. 2023.

PIAO, S. *et al.* **Characteristics, drivers and feedbacks of global greening**. [*S. l.: s. n.*], 2020.

POVEDA, G.; MESA, O. J. On the existence of Lloró (the rainiest locality on earth): Enhanced ocean-land-atmosphere interaction by a low-level jet. **Geophysical Research Letters**, [*s. l.*], v. 27, n. 11, 2000.

PRIANTE-FILHO, N. *et al.* Comparison of the mass and energy exchange of a pasture and a mature transitional tropical forest of the southern Amazon Basin during a seasonal transition. **Global Change Biology**, [*s. l.*], v. 10, n. 5, 2004.

PRIESTLEY, C. H. B.; TAYLOR, R. J. On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. **Monthly Weather Review**, [*s. l.*], v. 100, n. 2, 1972.

PURDY, A. J. *et al.* Ground heat flux: An analytical review of 6 models evaluated at 88 sites and globally. **Journal of Geophysical Research: Biogeosciences**, [*s. l.*], v. 121, n. 12, 2016.

RANA, G.; KATERJI, N.; DE LORENZI, F. Measurement and modelling of evapotranspiration of irrigated citrus orchard under Mediterranean conditions. **Agricultural and Forest Meteorology**, [*s. l.*], v. 128, n. 3–4, 2005.

REAGER, J. T.; FAMIGLIETTI, J. S. Characteristic mega-basin water storage behavior using GRACE. **Water Resources Research**, [*s. l.*], v. 49, n. 6, 2013.

REBOITA, M. S. *et al.* Climatological features of cutoff low systems in the Southern Hemisphere. **Journal of Geophysical Research Atmospheres**, [*s. l.*], v. 115, n. 17, 2010.

RIGON, R.; BERTOLDI, G.; OVER, T. M. GEOtop: A distributed hydrological model with coupled water and energy budgets. **Journal of Hydrometeorology**, [*s. l.*], v. 7, n. 3, 2006.

ROCHA, H. R. da *et al.* Measurements of CO2 exchange over a woodland savanna (Cerrado Sensu stricto) in southeast Brasil. **Biota Neotropica**, [*s. l.*], v. 2, n. 1, 2002.

RODELL, M. *et al.* Basin scale estimates of evapotranspiration using GRACE and other observations. **Geophysical Research Letters**, [*s. l.*], v. 31, n. 20, 2004a.

RODELL, M. *et al.* Estimating evapotranspiration using an observation based terrestrial water budget. **Hydrological Processes**, [*s. l.*], v. 25, n. 26, 2011.

RODELL, M. *et al.* The Global Land Data Assimilation System. **Bulletin of the American Meteorological Society**, [*s. l.*], v. 85, n. 3, 2004b.

ROERINK, G. J.; SU, Z.; MENENTI, M. 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**, [*s. l.*], v. 25, n. 2, 2000.

RUBERT, G. C. *et al.* **Correction: Evapotranspiration of the Brazilian pampa biome: Seasonality and influential factors [Water, 10, 1864 (2018)] DOI: 10.3390/w10121864**. [*S. l.: s. n.*], 2019.

RUBERT, G. C. D. *et al.* ESTIMATIVA DA EVAPOTRANSPIRAÇÃO EM ÁREA DE PASTAGEM EM SANTA MARIA – RS. **Ciência e Natura**, [*s. l.*], v. 38, 2016.

RUBERT, G. C. D. *et al.* Patterns and controls of the latent and sensible heat fluxes in the Brazilian Pampa Biome. **Atmosphere**, [*s. l.*], v. 13, n. 1, 2022.

RUHOFF, A. L. *et al.* A MODIS-based energy balance to estimate Evapotranspiration for clear-sky days in Brazilian tropical savannas. **Remote Sensing**, [*s. l.*], v. 4, n. 3, p. 703–725, 2012.

RUHOFF, A. L. *et al.* Evaluation de l'algorithme MODIS d'estimation de l'évapotranspiration globale utilisant des mesures de covariance de la turbulence et la modélisation hydrologique dans le bassin du Rio Grande. **Hydrological Sciences Journal**, [*s. l.*], v. 58, n. 8, p. 1658–1676, 2013.

RUHOFF, A. *et al.* Global Evapotranspiration Datasets Assessment Using Water Balance in South America. **Remote Sensing**, [*s. l.*], v. 14, n. 11, 2022. Disponível em: https://www.mdpi.com/2072-4292/14/11/2526.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, [*s. l.*], v. 323, n. 6088, 1986.

RUNNING, S. *et al.* MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V006 [Data set]. **NASA EOSDIS Land Processes DAAC**, [*s. l.*], v. 1.5, 2017.

RUSSELL, E. S. *et al.* Impacts of soil heat flux calculation methods on the surface energy balance closure. **Agricultural and Forest Meteorology**, [*s. l.*], v. 214–215, 2015.

RYU, Y. *et al.* Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales. **Global Biogeochemical Cycles**, [*s. l.*], v. 25, n. 4, 2011.

SABOORI, M. *et al.* Automatically selecting hot and cold pixels for satellite actual evapotranspiration estimation under different topographic and climatic conditions. **Agricultural Water Management**, [*s. l.*], v. 248, 2021.

SAKAI, R. K. *et al.* Land-use change effects on local energy, water, and carbon balances in an Amazonian agricultural field. **Global Change Biology**, [*s. l.*], v. 10, n. 5, 2004.

SALESKA, S. AmeriFlux AmeriFlux BR-Sa1 Santarem-Km67-Primary Forest. [*s. l.*], 2016.

SALESKA, S. R. *et al.* Carbon in Amazon Forests: Unexpected Seasonal Fluxes and Disturbance-Induced Losses. **Science**, [*s. l.*], v. 302, n. 5650, 2003.

SALESKA, S. R. *et al.* LBA-ECO CD-32 Flux Tower Network Data Compilation, Brazilian Amazon: 1999-2006. *Em*: ORNL DAAC. [*S. l.: s. n.*], 2013. p. 1999–2006. *E-book*. Disponível em: http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1174.

SANTOS, A. J. B. *et al.* Effects of fire on surface carbon, energy and water vapour fluxes over campo sujo savanna in central Brazil. **Functional Ecology**, [*s. l.*], v. 17, n. 6, 2003.

SARI, V.; DOS REIS CASTRO, N. M.; PEDROLLO, O. C. Estimate of Suspended Sediment Concentration from Monitored Data of Turbidity and Water Level Using Artificial Neural Networks. **Water Resources Management**, [*s. l.*], v. 31, n. 15, p. 4909–4923, 2017.

SAUER, T. J.; HORTON, R. Soil Heat Flux. **Micrometeorology in Agricultural Systems**, [*s. l.*], p. 131–154, 2015. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.2134/agronmonogr47.c7. Acesso em: 1 dez. 2022.

SCHAAF, C.; WANG, Z. **MCD43A4 MODIS/Terra+Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global - 500m V006**. [*S. l.*], 2015.

SCHMUGGE, T. J. *et al.* **Remote sensing in hydrology**. [*S. l.: s. n.*], 2002. Disponível em: www.elsevier.com/locate/advwatres. .

SCHULTZ, G. A.; ENGMAN, E. T. **Remote Sensing in Hydrology and Water Management**. 2011. ed. [*S. l.: s. n.*], 2000.

SEN, P. K. Estimates of the Regression Coefficient Based on Kendall's Tau. **Journal of the American Statistical Association**, [*s. l.*], v. 63, n. 324, 1968.

SENAY, G. B. *et al.* A coupled remote sensing and simplified surface energy balance approach to estimate actual evapotranspiration from irrigated fields. **Sensors**, [*s. l.*], v. 7, n. 6, p. 979–1000, 2007.

SENAY, G. B. *et al.* Improving the Operational Simplified Surface Energy Balance Evapotranspiration Model Using the Forcing and Normalizing Operation. **Remote Sensing**, [*s. l.*], v. 15, n. 1, 2023.

SENAY, G. B. *et al.* Mapping actual evapotranspiration using Landsat for the conterminous United States: Google Earth Engine implementation and assessment of the SSEBop model. **Remote Sensing of Environment**, [*s. l.*], v. 275, p. 113011, 2022.

SENAY, G. B. *et al.* Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach. **Journal of the American Water Resources Association**, [*s. l.*], v. 49, n. 3, p. 577–591, 2013.

SENAY, G. B. Satellite psychrometric formulation of the operational simplified surface energy balance (ssebop) model for. **Applied Engineering in Agriculture**, [*s. l.*], v. 34, n. 3, p. 555–566, 2018.

SENAY, G. B. *et al.* Satellite-based water use dynamics using historical Landsat data (1984–2014) in the southwestern United States. **Remote Sensing of Environment**, [*s. l.*], v. 202, p. 98–112, 2017.

SENAY, G. B.; BUDDE, M. E.; VERDIN, J. P. Enhancing the Simplified Surface Energy Balance (SSEB) approach for estimating landscape ET: Validation with the METRIC model. **Agricultural Water Management**, [*s. l.*], v. 98, n. 4, p. 606–618, 2011.

SENAY, G. B.; KAGONE, S.; VELPURI, N. M. Operational global actual evapotranspiration: Development, evaluation, and dissemination. **Sensors (Switzerland)**, [*s. l.*], v. 20, n. 7, 2020.

SHEFFIELD, J. *et al.* Closing the terrestrial water budget from satellite remote sensing. **Geophysical Research Letters**, [*s. l.*], v. 36, n. 7, 2009.

SHIMIZU, M. H.; ANOCHI, J. A.; KAYANO, M. T. Precipitation patterns over northern Brazil basins: climatology, trends, and associated mechanisms. **Theoretical and Applied Climatology**, [*s. l.*], v. 147, n. 1–2, p. 767–783, 2022. Disponível em: https://link.springer.com/article/10.1007/s00704-021-03841-4. Acesso em: 17 abr. 2023.

SILVA, P. F. da *et al.* Seasonal patterns of carbon dioxide, water and energy fluxes over the Caatinga and grassland in the semi-arid region of Brazil. **Journal of Arid Environments**, [*s. l.*], v. 147, 2017.

SILVERMAN, D.; DRACUP, J. A. Artificial neural networks and long-range precipitation prediction in California. **Journal of Applied Meteorology**, [*s. l.*], v. 39, n. 1, 2000.

SINGH, R. K. *et al.* Actual evapotranspiration (water use) assessment of the colorado river basin at the landsat resolution using the operational simplified surface energy balance model. **Remote Sensing**, [*s. l.*], v. 6, n. 1, p. 233–256, 2013.

SINGH, R. K. *et al.* Application of SEBAL Model for Mapping Evapotranspiration and Estimating Surface Energy Fluxes in South-Central Nebraska. **Journal of Irrigation and Drainage Engineering**, [*s. l.*], v. 134, n. 3, 2008.

SINGH, R. K.; SENAY, G. B. Comparison of four different energy balance models for estimating EvapoTranspiration in the Midwestern United States. **Water (Switzerland)**, [*s. l.*], v. 8, n. 1, 2016.

SIQUEIRA, V. A. *et al.* Toward continental hydrologic-hydrodynamic modeling in South America. **Hydrology and Earth System Sciences**, [*s. l.*], v. 22, n. 9, p. 4815–4842, 2018.

SOUZA, V. de A. *et al.* Evaluation of MOD16 algorithm over irrigated rice paddy using flux tower measurements in Southern Brazil. **Water (Switzerland)**, [*s. l.*], v. 11, n. 9, 2019.

SOUZA, C. M. *et al.* Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. **Remote Sensing**, [*s. l.*], v. 12, n. 17, 2020.

SOUZA, R. *et al.* Vegetation response to rainfall seasonality and interannual variability in tropical dry forests. **Hydrological Processes**, [*s. l.*], v. 30, n. 20, 2016.

SPERA, S. A. *et al.* Land-use change affects water recycling in Brazil's last agricultural frontier. **Global change biology**, [*s. l.*], v. 22, n. 10, 2016.

SPRACKLEN, D. v.; GARCIA-CARRERAS, L. The impact of Amazonian deforestation on Amazon basin rainfall. **Geophysical Research Letters**, [*s. l.*], v. 42, n. 21, 2015.

STATISTICAL METHODS IN THE ATMOSPHERIC SCIENCES. [*S. l.: s. n.*], 2019.

STEVAUX, J. C. *et al.* Floods in Urban Areas of Brazil. **Developments in Earth Surface Processes**, [*s. l.*], v. 13, n. C, p. 245–266, 2009. Disponível em: Acesso em: 6 set. 2022.

SU, Z. **The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxesHydrology and Earth System Sciences**. [*S. l.: s. n.*], 2002.

SWENSON, S. C. GRACE MONTHLY LAND WATER MASS GRIDS NETCDF RELEASE 5.0. Ver. 5.0. **PO.DAAC, CA, USA**, [*s. l.*], 2012.

SWENSON, S.; WAHR, J. **Estimating Large-Scale Precipitation Minus Evapotranspiration from GRACE Satellite Gravity Measurements**. [*S. l.: s. n.*], 2006a.

SWENSON, S.; WAHR, J. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. **Journal of Geophysical Research: Solid Earth**, [*s. l.*], v. 107, n. B9, p. ETG 3-1-ETG 3-13, 2002.

SWENSON, S.; WAHR, J. Post-processing removal of correlated errors in GRACE data. **Geophysical Research Letters**, [*s. l.*], v. 33, n. 8, 2006b.

TABARELLI, M. *et al.* **PELD - CNPq dez anos do Programa de Pesquisas Ecológicas de Longa Duração do Brasil : achados, lições e perspectivas**. [*S. l.*]: Ed. Universidade da UFPE, 2013.

TABARI, H.; SABZIPARVAR, A. A.; AHMADI, M. Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region. **Meteorology and Atmospheric Physics**, [*s. l.*], v. 110, n. 3, 2011.

TANG, X. *et al.* Can VIIRS continue the legacy of MODIS for near real-time monitoring of tropical forest disturbance?. **Remote Sensing of Environment**, [*s. l.*], v. 249, 2020.

TASUMI, M. *et al.* **Irrigation and Drainage Systems (2005) 19: 355-376 C Operational aspects of satellite-based energy balance models for irrigated crops in the semi-arid U.S**. [*S. l.: s. n.*], 2005.

TASUMI, M.; ALLEN, R. G.; TREZZA, R. At-Surface Reflectance and Albedo from Satellite for Operational Calculation of Land Surface Energy Balance. **Journal of Hydrologic Engineering**, [*s. l.*], v. 13, n. 2, 2008.

TAYLOR, S. H. *et al.* CO2 availability influences hydraulic function of C3 and C4 grass leaves. **Journal of Experimental Botany**, [*s. l.*], v. 69, n. 10, p. 2731, 2018. Disponível em: /pmc/articles/PMC5920307/. Acesso em: 3 abr. 2023.

TEICHRIEB, C. A. *et al.* Characterizing the relative role of low-frequency and turbulent processes in the nocturnal boundary layer through the analysis of two-point correlations of the wind components. **Physica A: Statistical Mechanics and its Applications**, [*s. l.*], v. 392, n. 6, 2013.

TEIXEIRA, A. H. de C. *et al.* Reviewing SEBAL input parameters for assessing evapotranspiration and water productivity for the Low-Middle São Francisco River basin, Brazil. Part A: Calibration and validation. **Agricultural and Forest Meteorology**, [*s. l.*], v. 149, n. 3–4, p. 462–476, 2009.

TIMM, A. U. *et al.* Energy partitioning and evapotranspiration over a rice paddy in Southern Brazil. **Journal of Hydrometeorology**, [*s. l.*], v. 15, n. 5, 2014.

TRENBERTH, K. E.; FASULLO, J. T.; KIEHL, J. Earth's global energy budget. **Bulletin of the American Meteorological Society**, [*s. l.*], v. 90, n. 3, 2009.

TUCCI, C. E. M. **Hidrologia: Ciência e Aplicação**. [*S. l.: s. n.*], 2004.

TURCHETTO-ZOLET, A. C. *et al.* **Phylogeographical patterns shed light on evolutionary process in South America**. [*S. l.: s. n.*], 2013.

TWINE, T. E. *et al.* Correcting eddy-covariance flux underestimates over a grassland. **Agricultural and Forest Meteorology**, [*s. l.*], v. 103, n. 3, 2000.

UCKER TIMM, A. *et al.* Employing turbulent and meandering time scales to modeling the contaminants enhanced horizontal dispersion. **Atmospheric Research**, [*s. l.*], v. 93, n. 4, 2009.

ULKE, A. G.; GATTINONI, N. N.; POSSE, G. Analysis and modelling of turbulent fluxes in two different ecosystems in Argentina. *Em*: , 2015. **International Journal of Environment and Pollution**. [*S. l.: s. n.*], 2015.

VAN DER ENT, R. J. *et al.* Origin and fate of atmospheric moisture over continents. **Water Resources Research**, [*s. l.*], v. 46, n. 9, 2010.

VARGAS GODOY, M. R. *et al.* **The Global Water Cycle Budget: A Chronological Review**. [*S. l.: s. n.*], 2021.

VELICOGNA, I.; WAHR, J. Time-variable gravity observations of ice sheet mass balance: Precision and limitations of the GRACE satellite data. **Geophysical Research Letters**, [*s. l.*], v. 40, n. 12, p. 3055–3063, 2013. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1002/grl.50527. Acesso em: 1 abr. 2023.

VELPURI, N. M. *et al.* A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. **Remote Sensing of Environment**, [*s. l.*], v. 139, p. 35–49, 2013.

VERA, C. *et al.* **Toward a unified view of the American monsoon systems**. [*S. l.: s. n.*], 2006.

VIHERMAA, L. E. *et al.* Fluvial carbon export from a lowland Amazonian rainforest in relation to atmospheric fluxes. **Journal of Geophysical Research: Biogeosciences**, [*s. l.*], v. 121, n. 12, 2016.

VILLARREAL, S.; VARGAS, R. Representativeness of FLUXNET Sites Across Latin America. **Journal of Geophysical Research: Biogeosciences**, [*s. l.*], v. 126, n. 3, 2021.

VINUKOLLU, R. K. *et al.* Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends. **Hydrological Processes**, [*s. l.*], v. 25, n. 26, 2011.

VISHWAKARMA, B. D.; DEVARAJU, B.; SNEEUW, N. What is the spatial resolution of GRACE satellite products for hydrology?. **Remote Sensing**, [*s. l.*], v. 10, n. 6, 2018.

VOGL, T. P. *et al.* Accelerating the convergence of the back-propagation method. **Biological Cybernetics 1988 59:4**, [*s. l.*], v. 59, n. 4, p. 257–263, 1988. Disponível em: https://link.springer.com/article/10.1007/BF00332914. Acesso em: 4 dez. 2022.

VOLK, J. M. *et al.* Development of a Benchmark Eddy Flux Evapotranspiration Dataset for Evaluation of Satellite-Driven Evapotranspiration Models Over the CONUS. **Agricultural and Forest Meteorology**, [*s. l.*], v. 331, p. 109307, 2023. Disponível em: Acesso em: 26 fev. 2023.

VOLK, J. *et al.* flux-data-qaqc: A Python Package for Energy Balance Closure and Post-Processing of Eddy Flux Data. **Journal of Open Source Software**, [*s. l.*], v. 6, n. 66, 2021.

VON RANDOW, C. *et al.* Comparative measurements and seasonal variations in energy and carbon exchange over forest and pasture in South West Amazonia. **Theoretical and Applied Climatology**, [*s. l.*], v. 78, n. 1–3, 2004.

VON RANDOW, R. C. S. *et al.* Evapotranspiration of deforested areas in central and southwestern Amazonia. **Theoretical and Applied Climatology**, [*s. l.*], v. 109, n. 1–2, 2012.

WAGLE, P. *et al.* Performance of five surface energy balance models for estimating daily evapotranspiration in high biomass sorghum. **ISPRS Journal of Photogrammetry and Remote Sensing**, [*s. l.*], v. 128, p. 192–203, 2017.

WALLS, S. *et al.* Prediction of actual evapotranspiration by artificial neural network models using data from a Bowen ratio energy balance station. **Neural Computing and Applications**, [*s. l.*], v. 32, n. 17, 2020.

WAN, Z. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. **Remote Sensing of Environment**, [*s. l.*], v. 140, 2014.

WATERS, R. *et al.* **S E B A L Surface Energy Balance Algorithms for Land Idaho Implementation Advanced Training and Users Manual List of Symbols**. [*S. l.: s. n.*], 2002.

WATKINS, M. M. *et al.* Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. **Journal of Geophysical Research: Solid Earth**, [*s. l.*], v. 120, n. 4, p. 2648–2671, 2015.

WEERASINGHE, I. *et al.* Can we trust remote sensing evapotranspiration products over Africa. **Hydrology and Earth System Sciences**, [*s. l.*], v. 24, n. 3, 2020.

WEISE, D. GRACE monthly global water mass grids NETCDF RELEASE 5.0. Ver. 5.0. **Po.Daac, Ca, Usa.**, [*s. l.*], 2015.

WIDROW, B.; HOFF, M. E. Adaptive switching circuits. *Em*: , 1989. **Wescon Conference Record**. [*S. l.: s. n.*], 1989.

WIESE, D. N.; LANDERER, F. W.; WATKINS, M. M. Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. **Water Resources Research**, [*s. l.*], v.

52, n. 9, p. 7490–7502, 2016. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1002/2016WR019344. Acesso em: 1 abr. 2023.

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate Research**, [*s. l.*], v. 30, n. 1, 2005.

XU, Z.; ZHOU, G.; SHIMIZU, H. **Plant responses to drought and rewatering**. [*S. l.: s. n.*], 2010.

YIN, L. *et al.* A comparison of SSEBop-model-based evapotranspiration with eight evapotranspiration products in the Yellow River Basin, China. **Remote Sensing**, [*s. l.*], v. 12, n. 16, 2020.

ZANAGA, D. *et al.* ESA WorldCover 10 m 2020 v100. [*s. l.*], 2021. Disponível em: https://zenodo.org/record/5571936. Acesso em: 4 abr. 2023.

ZANETTI, S. S. *et al.* Estimating Evapotranspiration Using Artificial Neural Network and Minimum Climatological Data. **Journal of Irrigation and Drainage Engineering**, [*s. l.*], v. 133, n. 2, 2007.

ZHANG, Y. *et al.* Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. **Remote Sensing of Environment**, [*s. l.*], v. 222, 2019.

ZHANG, K. *et al.* Satellite based analysis of northern ET trends and associated changes in the regional water balance from 1983 to 2005. **Journal of Hydrology**, [*s. l.*], v. 379, n. 1–2, 2009.

ZHANG, S. *et al.* Temporal-spatial variation in crop evapotranspiration in Hebei Plain, China. **Journal of Food, Agriculture and Environment**, [*s. l.*], v. 8, n. 2, 2010.

ZHANG, K.; KIMBALL, J. S.; RUNNING, S. W. **A review of remote sensing based actual evapotranspiration estimation**. [*S. l.*]: John Wiley and Sons Inc, 2016.

ZHAO, W. *et al.* A practical method for reducing terrain effect on land surface temperature using random forest regression. **Remote Sensing of Environment**, [*s. l.*], v. 221, 2019a.

ZHAO, M. *et al.* Evapotranspiration frequently increases during droughts. **Nature Climate Change 2022 12:11**, [*s. l.*], v. 12, n. 11, p. 1024–1030, 2022. Disponível em: https://www.nature.com/articles/s41558-022-01505-3. Acesso em: 14 nov. 2022.

ZHAO, W. *et al.* Normalization of the temporal effect on the MODIS land surface temperature product using random forest regression. **ISPRS Journal of Photogrammetry and Remote Sensing**, [*s. l.*], v. 152, 2019b.

ZHAO, X.; LIU, Y. Variability of Surface Heat Fluxes and Its Driving Forces at Different Time Scales Over a Large Ephemeral Lake in China. **Journal of Geophysical Research: Atmospheres**, [*s. l.*], v. 123, n. 10, 2018.

ZHENG, C.; WANG, Q.; LI, P. Coupling SEBAL with a new radiation module and MODIS products for better estimation of evapotranspiration. **Hydrological Sciences Journal**, [*s. l.*], v. 61, n. 8, p. 1535–1547, 2016. Disponível em: Acesso em: 6 dez. 2022.

ZHU, Z. *et al.* Greening of the Earth and its drivers. **Nature Climate Change**, [*s. l.*], v. 6, n. 8, 2016.

ZIMMER, T. *et al.* Estimating the Soil Thermal Conductivity Using Experimental Soil Heat Flux in a Rice Paddy Area. **American Journal of Environmental Engineering**, [*s. l.*], v. 6, n. 4A, 2016.