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] and Rodolfo Souza [10]

Keywords– remote sensing; surface energy balance; deep learning.

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)]. Despite its small magnitude, G plays an important role in the determination of available energy for hydrological processes, such as evapotranspiration (ET) [Purdy *et al.* (2016)]. Many models integrated remote sensing data with meteorological monitoring to calculate SEB fluxes, of which G was determined empirically [Jackson *et al.* (1987); Kalma & Jupp (1990); Menenti & Choudhury (1993); Norman *et al.* (1995); Anderson *et al.* (1997); Bastiaanssen (1995); Roerink *et al.* (2000); Su (2002); Allen *et al.* (2007); Anderson *et al.* (2007)]. However, the wide spatio-temporal variability of G and the developed empirical models limited sampling usually implicates that their application to conditions other than the ones for which they were developed can lead to regional biases of G predictions, which may result in large uncertainties in SEB closure [Purdy *et al.* (2016)]. Therefore, a reduction in the biases in G estimation may lead to more accurate available energy prediction and improvements in ET estimates.

To overcome the issue of regional biases when developing a model for G prediction, 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

1) Universidade Federal do Rio Grande do Sul, 91501-970, Porto Alegre, Brazil; cominideandrade@gmail.com; olavo.pedrollo@ufrgs.br; anderson.ruhoff@ufrgs.br; adriana.moreira@ufrgs.br; leonardo.laipelt@ufrgs.br; rafael.kayser@ufrgs.br;
2) Universidade Federal do Mato Grosso, 78060-900, Cuiabá, Brazil; marcelo@fisica.ufmt.br
3) Universidade Federal de Campina Grande, 58429-900, Campina Grande, Brazil; carlos.santos@ufcg.edu.br
4) 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) Universidade de Cuiabá, 78005-300, Cuiabá, Brazil; higo.dalmagro@kroton.com.br
7) Universidade Federal de Pernambuco, 50740-540, Recife, Brazil; anto-nio.antonino@ufpe.br
8) Universidade Federal do Agreste de Pernambuco, 55292-278, Garanhuns, Brazil; romual-do.lima@ufape.edu.br
9) Universidade Federal Rural de Pernambuco, 56900-000, Serra Talhada, Brazil; eduardo.ssouza@ufrpe.br
10) Texas A&M University, College Station, TX 77843, USA; rodolfo.souza@tamu.edu

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 [Ruhoff *et al.* (2012); Laipelt *et al.* (2020), Danelichen *et al.* (2014)].

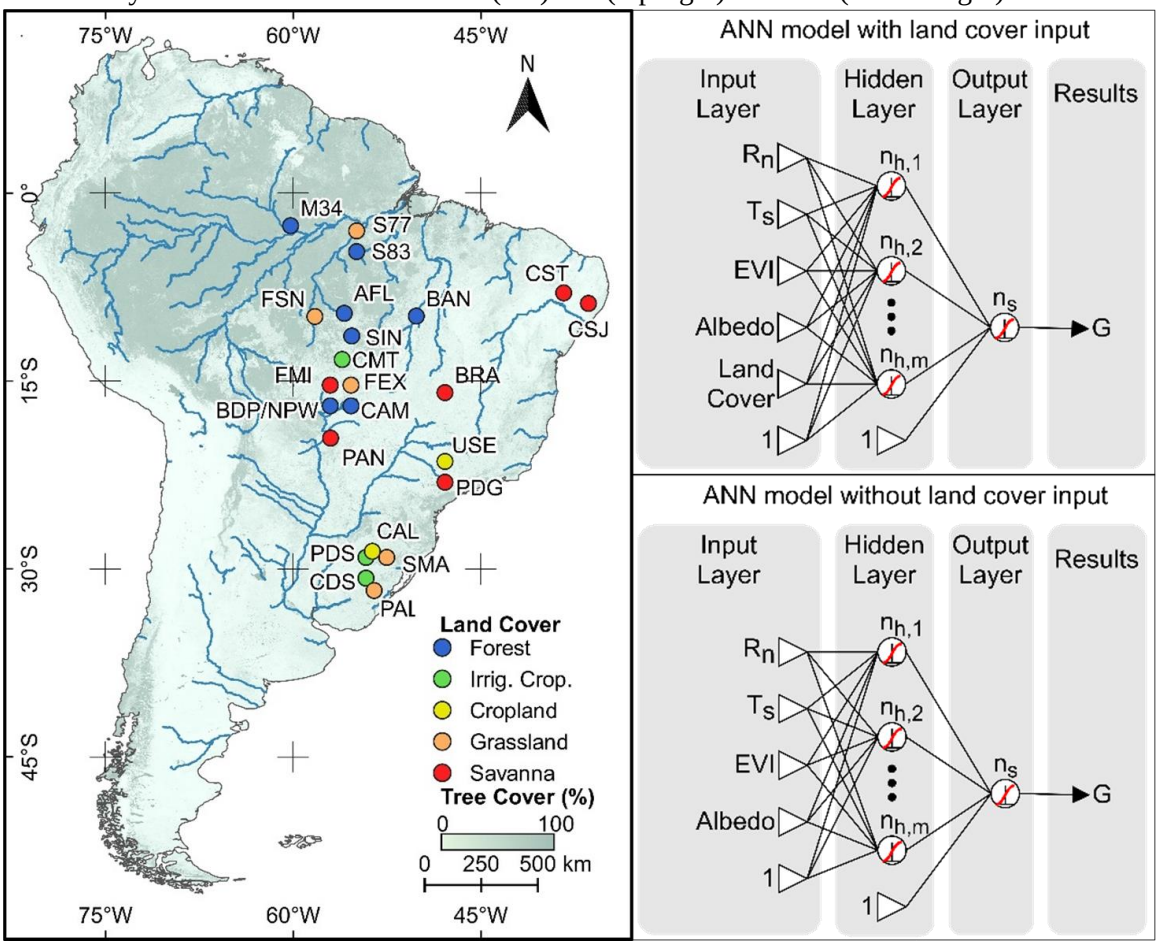
An alternative to the existing methods of modeling G is the use of artificial neural networks (ANNs), which are universal approximators [Hornik *et al.* (1989)] and have shown better performance than existing conceptual and empirical models [Silverman & Dracup (2000); Zanetti *et al.* (2007); Tabari *et al.* (2010); Canelón & Chávez (2011); Jimeno-Sáez *et al.* (2018); Käfer *et al.* (2020)]. 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 *et al.* (1989)].

Therefore, given the diversity of G estimation methods and their limited application conditions, this study focused 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 ANNs' performance to that of commonly used G models. With this study, we hope to assist future SEB closure studies.

MATERIAL AND METHODS

23 flux towers located in South America were used to calibrate and validate the ANNs. Figure 1 displays the location and land cover of the study sites, and the scheme of the ANNs that were built, with (Gc) and without (Gnc) land cover data. Surface temperature (Ts), albedo (α), and enhanced vegetation index (EVI), obtained from the moderate resolution imaging spectroradiometer (MODIS), and net radiation (Rn) from the global land data assimilation system 2.1 (GLDAS 2.1) product, were used as inputs, both available at the Google Earth Engine platform (https://earthengine.google.com/ [Accessed date: June 30, 2020]). The land cover information was acquired from flux towers data files and, for simplification, were grouped into one of five types and replaced by a number, as follows: (1) cropland; (2) irrigated cropland; (3) forest; (4) grassland; (5) savanna.

Figure 1 – Study sites location and land cover (left). Gc (top right) and Gnc (bottom right) ANNs' structure.



ANN training was performed on the randomized series, partitioned into calibration and validation sets. Priorly, one third of the data was reserved for verification and comparison with the Jackson *et al.* (1987) and Bastiaanssen (1995) equations. The existing models were adjusted for minimization of the quadratic error in the verification set.

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 1. These values also represent the valid range for applications of the trained ANNs.

Table 1 – Performance metrics of the Gc and Gnc ANNs, and the Jackson *et al.* (1987) and Bastiaanssen (1995) models, grouped by land cover.

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.

RESULTS

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 security reasons, 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, 1,745 samples for the validation series, and 1,285 samples for the verification series, for both the Gc and the Gnc ANNs.

Table 2 displays the Nash-Sutcliffe coefficient (NS), the mean absolute error (MAE) and the correlation coefficient (r) of the existing models (after adjustment) and of the developed ANNs, grouped by land cover.

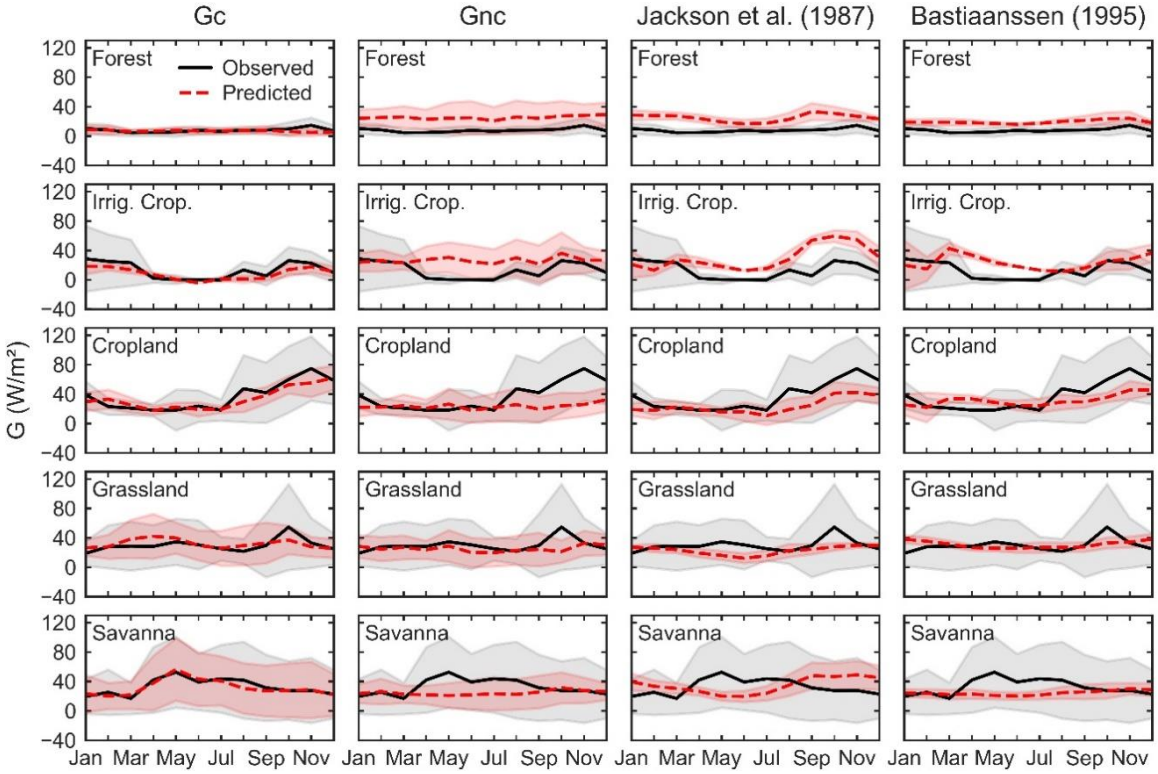
Table 2 – Performance metrics of the Gc and Gnc ANNs, and the Jackson *et al.* (1987) and Bastiaanssen (1995) models (denoted as Jack. And Bast., respectively), grouped by land cover, and summarized for all 23 flux towers (general).

Land Cover	NS				MAE (W/m²)				r			
	Gc	Gnc	Jack.	Bast.	Gc	Gnc	Jack.	Bast.	Gc	Gnc	Jack.	Bast.
Forest	-0.16	0.70	-8.32	-4.79	4.60	11.43	14.77	11.90	0.22	0.57	0.15	-0.13
Irrigated Cropland	0.08	-0.58	-0.77	-0.89	19.90	28.70	33.41	30.87	0.36	0.68	-0.14	-0.38
Cropland	0.22	0.11	-0.05	0.05	23.03	24.80	24.61	25.46	0.48	0.51	0.30	0.29
Grassland	0.19	0.41	-0.05	-0.06	19.08	20.29	21.41	23.97	0.49	0.51	0.16	-0.02
Savanna	0.72	-0.10	-0.33	-0.01	15.12	18.09	42.12	31.75	0.86	0.48	-0.28	0.38
General	**0.53**	**0.27**	**-0.17**	**0.02**	**14.04**	**18.06**	**24.43**	**21.78**	**0.73**	**0.53**	**0.02**	**0.21**

Based on Table 2, both ANNs and models yielded low NS values, with exceptions for Gc over savanna (NS = 0.72) and Gnc over forest (NS = 0.70). The ANNs yielded lower MAE values and higher r values consistently over the different land covers. Overall, The ANNs performed better than the existing models, with great improvement in NS and r values, and between 17.1% and 42.5% reduction of MAE.

Figure 2 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 2 – 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.



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 the mapping of 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.

REFERENCES

ALLEN, R.G.; TASUMI, M.; TREZZA, R. (2007). Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)—Model. J. Irrig. Drain. Eng., 133, 380–394, doi:10.1061/(asce)0733-9437(2007)133:4(380).

ANDERSON, M.C.; NORMAN, J.M.; DIAK, G.R.; KUSTAS, W.P.; MECIKALSKI, J.R. (1997). A two-source time-integrated model for esti-mating surface fluxes using thermal infrared remote sensing. Remote Sens. Environ., 60, 195–216.

ANDERSON, M.C.; KUSTAS, W.P.; NORMAN, J.M.; HAIN, C.R.; MECIKALSKI, J.R.; SCHULTZ, L.; GONZÁLEZ-DUGO, M.P.; CAMMALLERI, C.; D'URSO, G.; PIMSTEIN, A.; *et al.* (2007). Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. Hydrol. Earth Syst. Sci. 2011, 15, 223–239, doi:10.5194/hess-15-223-2011.

BASTIAANSSEN, W.G.M. (1995). Regionalization of Surface Flux Densities and Moisture Indicators in Composite Terrain. Ph. D. Thesis, Wageningem Agricultural University, Wageningen, The Netherlands, November; p. 273.

BIUDES, M.S.; VOURLITIS, G.L.; MACHADO, N.G.; DE ARRUDA, P.H.Z.; NEVES, G.A.R.; DE ALMEIDA LOBO, F.; NEALE, C.M.U.; DE SOUZA NOGUEIRA, J. (2015). Patterns of energy exchange for tropical ecosystems across a climate gradient in Mato Grosso, Brazil. Agr. For. Meteorol., 202, 112–124, doi:10.1016/j.agrformet.2014.12.8.

BORGES, C.K.; DOS SANTOS, C.A.C.; CARNEIRO, R.G.; DA SILVA, L.L.; DE OLIVEIRA, G.; MARIANO, D.; SILVA, M.T.; DA SILVA, B.B.; BEZERRA, B.G.; PEREZ-MARIN, A.M.; *et al.* (2020). Seasonal variation of surface radiation and energy balances over two contrasting areas of the seasonally dry tropical forest (Caatinga) in the Brazilian semi-arid. Env. Monit. Assess. 2020, 192, 524, doi:10.1007/s10661-020-08484-y.

CANELÓN, D.J.; CHÁVEZ, J.L. (2011). Soil Heat Flux Modeling Using Artificial Neural Networks and Multispectral Airborne Remote Sensing Imagery. Remote Sens., 3, 1627–1643, doi:10.3390/rs3081627.

DA ROCHA, H.R.; MANZI, A.O.; CABRAL, O.M.; MILLER, S.D.; GOULDEN, M.L.; SALESKA, S.R.; COUPE, N.; WOFSY, S.C.; BORMA, L.S.; ARTAXO, P.; *et al.* (2009). Patterns of water and heat flux across a biome gradient from tropical forest to savanna in Brazil. J. Geophys. Res., 114, doi:10.1029/2007JG000640.

DANELICHEN, V.H.M.; BIUDES, M.S.; SOUZA, M.C.; MACHADO, N.G.; SILVA, B.B.; NOGUEIRA, J.S. (2014). Estimation of soil heat flux in a neotropical Wetland region using remote sensing techniques. Rev. Bras. Meteorol., 29, 469-482, doi:10.1590/0102-778620120568.

DAVIDSON, E.A.; ARTAXO, P. (2004). Globally significant changes in biological processes of the Amazon Basin: Results of the Large-scale Biosphere–Atmosphere Experiment. Glob. Chang. Biol., 10, 519–529, doi:10.1111/j.1529-8817.2003.00779.x.

EVA, H.D.; BELWARD, A.S.; DE MIRANDA, E.E.; DI BELLA, C.M.; GOND, V.; HUBER, O.; JONES, S.; SGRENZAROLI, M.; FRITZ, S. (2004). A land cover map of South America. Glob. Chang. Biol., 10, 731–744, doi:10.1111/j.1529-8817.2003.00774.x.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. (1989). Multilayer feedforward networks are universal approximators. Neural Netw., 2, 359–368.

JACKSON, R.D.; MORAN, M.S.; GAY, L.W.; RAYMOND, L.H. (1987). Evaluating evaporation from field crops using airborne radiometry and ground-based meteorological data. Irrig. Sci., 8, 81–90.

JIMENO-SÁEZ, P.; SENENT-APARICIO, J.; PÉREZ-SÁNCHEZ, J.; PULIDO-VELAZQUEZ, D. (2018). A Comparison of SWAT and ANN Models for Daily Runoff Simulation in Different Climatic Zones of Peninsular Spain. Water, 10, 192, doi:10.3390/w10020192.

KÄFER, P.; SOUZA DA ROCHA, N.; RIBEIRO DIAZ, L.; KAISER, E.; SANTOS, D.; VEECK, G.; ROBÉRTI, D.; ROLIM, S.; OLIVEIRA, G. (2020). Artificial neural networks model based on remote sensing to retrieve evapotranspiration over the Brazilian Pampa. J. Appl. Remote Sens., 14, 038504, doi:10.1117/1.JRS.14.038504].

KALMA, J.D.; JUPP, D.L.B. (1990). Estimating evaporation from pasture using infrared thermometry: Evaluation of a one-layer resistance model. Agr. For. Meteorol., 51, 223–246.

KUSTAS, W.P.; NORMAN, J.M. (1999). Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover. Agr. For. Meteorol., 94, 13–29, doi:10.1016/s0168-1923(99)00005-2.

LAIPELT, L.; RUHOFF, A.L.; FLEISCHMANN, A.S.; KAYSER, R.H.B.; DE MELLO KICH , E.; DA ROCHA, H.R.; NEALE, C.M.U. (2020). Assessment of an Automated Calibration of the SEBAL Algorithm to Estimate Dry-Season Surface-Energy Partitioning in a Forest–Savanna Transition in Brazil. Remote Sens., 12, 1108, doi:10.3390/rs12071108.

MENENTI, M.; CHOUDHURY, B.J. (1993). Parameterization of land surface evaporation by means of location dependent potential evaporation and surface temperature range. In Exchange Processes at the Land Surface for a Range of Space and Time Scales. In Proceedings of the Yokohama Symposium, Yokohama, Japan; pp. 561–568.

NORMAN, J.M.; KUSTAS, W.P.; HUMES, K.S. (1995). Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. Agr. For. Meteorol, 77, 263–293, doi:10.1016/0168-1923(95)02265-y.

PURDY, A.J.; FISHER, J.B.; GOULDEN, M.L.; FAMIGLIETTI, J.S. (2016). Ground heat flux: An analytical review of 6 models evaluated at 88 sites and globally. J. Geophys. Res. Biogeosciences, 121, 3045–3059, doi:10.1002/2016jg003591.

ROERINK, G.J.; SU, Z.; MENENTI, M. (2000). S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance. Phys. Chem. Earth Part B Hydrol., 25, 147–157, doi:10.1016/s1464-1909(99)00128-8.

RUHOFF, A.L.; PAZ, A.R.; COLLISCHONN, W.; ARAGAO, L.E.O.C.; DA ROCHA, H.R.; MALHI, Y.S. (2012). A MODIS-Based Energy Balance to Estimate Evapotranspiration for Clear-Sky Days in Brazilian Tropical Savannas. Remote Sens., 4, 703–725, doi:10.3390/rs4030703.

SALESKA, S.R.; DA ROCHA, H.R.; HUETE, A.R.; NOBRE, A.D.; ARTAXO, P.E.; SHIMABUKURO, Y.E. (2013). LBA-ECO CD-32 Flux Tower Network Data Compilation, Brazilian Amazon: 1999–2006. ORNL Distributed Active Archive Center. Available online: http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1174 (accessed on June 30, 2020)

SILVERMAN, D.; DRACUP, J.A. (2000). Artificial Neural Networks and Long-Range Precipitation Prediction in California. J. Appl. Meteorol., 39, 57–66, doi:10.1175/1520-0450(2000)039<0057:annalr>2.0.co;2.

SU, Z. (2002). The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci., 6, 85–100, doi:10.5194/hess-6-85-2002.

TABARELLI, M.; DA ROCHA, C.F.D.; ROMANOWSKI, H.P.; ROCHA, O.; DE LACERDA, L.D. (2013). PELD—CNPq: Dez Anos do Programa de Pesquisas Ecológicas de Longa Duração do Brasil: Achados, Lições e Perspectivas. Universitária da UFPE: Recife, Brazil, p. 446, ISBN: 978-85-415-0329-7.

TABARI, H.; SABZIPARVAR, A.-A.; AHMADI, M. (2010). Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region. Meteorol. Atmos. Phys., 110, 135–142, doi:10.1007/s00703-010-0110-z.

VILLARREAL, S.; VARGAS, R. (2021). Representativeness of FLUXNET sites across Latin America. J. Geophys. Res. Biogeosci., 126, e2020JG006090, doi: 10.1029/2020JG006090.

ZANETTI, S.S.; SOUSA, E.F.; OLIVEIRA, V.P.; ALMEIDA, F.T.; BERNARDO, S. (2007). Estimating Evapotranspiration Using Artificial Neural Network and Minimum Climatological Data. J. Irrig. Drain. Eng., 133, 83–89, doi:10.1061/(asce)0733-9437(2007)133:2(83).