Estimation of Forest Leaf Area Index Using Meteorological Data: Assessment of Heuristic Models

S. Karimi[1] [*], A. H. Nazemi[1], A. A. Sadraddini[1], T. R. Xu[2] [*], S. M. Bateni[3], and A. F. Fard[1]

[1] Water Engineering Department, Faculty of Agriculture, University of Tabriz, Tabriz 5166616471, Iran
[2] State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science,
Beijing Normal University, Beijing 100875, China
[3] Department of Civil and Environmental Engineering and Water Resources Research Center,
University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA

ABSTRACT. Leaf Area Index (LAI) is an important structural feature of our ecosystem as it affects energy, carbon, and water exchanges between the land surface and overlying atmosphere. Global scale LAI datasets have been obtained by regression, heuristic data driven, and radiative transfer models using remotely sensed land surface reflectance data. However, the estimation of LAI from remotely sensed data is limited only to clear sky conditions. Also, it is problematic to estimate LAI in forests by using conventional remote sensing image analysis of multi-spectral data. Due to the above-mentioned shortcomings of estimating LAI from remotely sensed data, this study obtained LAI from meteorological data using the Gene Expression Programming (GEP) technique. The new approach was tested in different forest sites with broad-leaf and needle-leaf trees in USA. The results showed that the GEP technique can accurately estimate LAI from meteorological data in different forest sites.

Keywords: meteorological data, leaf area index, gene expression programming

1. Introduction

Leaf area index (LAI) is a dimensionless variable, which is defined as the total one-sided area of photosynthetic tissues per unit ground surface area (Watson, 1947). LAI is an important structural characteristic of our ecosystem because it influences the exchanges of water, energy, and carbon between the land surface and atmosphere (Sellers et al., 1988; Wulder et al., 1998; Sonnentag et al., 2007). It is the main variable for modeling canopy photosynthesis and transpiretion, and highly affects the plant-atmosphere interaction. Therefore, it plays a key role in the energy and water exchanges between the canopy and the atmosphere (Weiss et al., 2004; Bateni et al., 2013a, b; Xu et al., 2014). Information on LAI is also critical to describe the canopy microclimate including wind, temperature, and moisture profiles within and above the canopy (Welles, 1990). Furthermore, it is an important parameter in the rainfall-runoff simulations (Li et al., 2009), drought stress analyses (Cáceres et al., 2015), and soil water content predictions (Coopersmith et al., 2014; Chen et al., 2015).

LAI may vary from day to day, and its evolution depends on various factors such as species composition, stage of development, site conditions, and the management practices (Welles, 1990; Jonckheere et al., 2004). Direct and indirect methods have been developed to obtain LAI (Jonckheere et al., 2004). The direct measurement of LAI is costly, labor intensive and time consuming (Colaizzi et al., 2017). Hence, indirect methods have been developed that relate LAI to the radiation intensity below the canopy via a radiative transfer model (Norman et al., 1989; Chason et al., 1991; Olivas et al., 2013).

Indirect methods are generally faster, and can be easily applied over large-scale domains. They can be categorized into three major groups, namely ground-based, space-borne, and crop growth methods (Bonan, 1993; Li et al., 2011). Ground-based approaches determine LAI at the ground via contact and/or non-contact methods (Anser et al., 2003; Qu et al., 2014). The ground-based methods can readily determine LAI at point or relatively small scales, but their application is too difficult at regional and global scales (Biudes et al., 2014). Thus, space-borne methods have been built based on differences in the spectral reflection of vegetation and surrounding area (Ripple et al., 1991). There are a variety of methods for retrieving LAI from remotely sensed data such as radiative transfer, regression, and heuristic models.

The radiative transfer models are built based on the physical relationship between the vegetation characteristics and spectral reflectance, and can estimate LAI accurately at regional scales (Meroni et al., 2004; He et al., 2013). However, these mod-
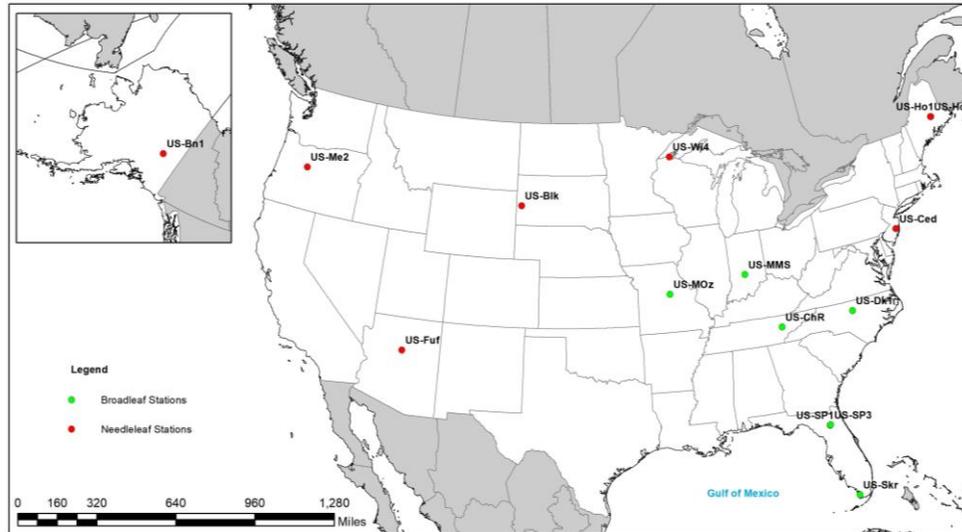
Figure 1. Geographical position of the studied locations.

els are usually intricate, and their application is time consuming (Jacquemoud et al., 2000). Furthermore, as the structural presentation of different vegetation is variable, different radiative transfer models will be needed for various types of vegetation/ land covers (Yin et al., 2015).

The regression-based methods usually relate LAI to vegetation indices such as simple ratio (SR), normalized difference vegetation index (NDVI), optimized soil adjusted vegetation index (OSAVI), and wide dynamic range vegetation index (WDRVI) (Makki and Homma, 2014). The empirical methods are site-specific, and need to be calibrated when applying at different regions. The crop growth models are complicated and need a large number of input parameters that limit their applications in large-scale domains (Zhou et al., 2013).

Hence, heuristic data driven approaches have been used to estimate LAI at larger scales (Feng and Liang, 2003; Walthal et al., 2004; Dunea and Moise, 2008; Xiao et al., 2014; Liang et al., 2014). For example, the Global LAnd Surface Satellite (GLASS)-LAI product is produced by using reflectance data in the visible and infrared bands in heuristic data driven methods (Liang et al., 2014; Xiao et al., 2014). As mentioned above, the current LAI retrieval methods are highly dependent on remotely sensed data. The estimation of LAI from remotely sensed data is limited only to clear sky conditions. Moreover, due to the impact of the aerosols, snow cover, and sensor failure, many satellite-based LAI products have high noise, low accuracy, and large errors in the time series and spatio-temporal distributions (Heinsch et al., 2006; Mao et al., 2017). Particularly, it is problematic to estimate LAI in forests by using conventional remote sensing image analysis of multi-spectral data (Nemani et al., 1993). For example, the reflectance in the near-infrared band is different for broadleaf and needle-leaf forests, and therefore they have different NDVI-LAI relationships (Franklin et al., 1997).

Due to the above-mentioned drawbacks of estimating LAI from space-borne data and crop growth models, this study obtains LAI in forests from meteorological data via employing the GEP technique. In the recent years, GEP has been applied to many hydrological problems such as estimation of solar radiation (Landeras et al., 2012), air temperature (Kisi et al., 2013a; Kisi and Shiri, 2014; Shiri et al., 2014a), pan evaporation (Kisi et al., 2012; Pour Alibaba et al., 2013; Kisi, 2015; Kim et al., 2015), forest carbon fluxes (Dou and Yang, 2018), crop evapotranspiration (Torres et al., 2011; Shiri et al., 2014b, c; Shrestha and Shukla, 2015; Karimi et al., 2016), soil parameters (Shiri et al., 2017a, b), river flows (Liu et al., 2014; Karimi et al., 2017) and river qualitative parameters (Kisi et al., 2013b). Recently, Karimi et al. (2018) utilized GEP and random forest to estimate LAI in croplands and grasslands from in-situ meteorological data.

Genetic programming (GP), a generalization of Genetic Algorithm (GA) (Goldberg, 1989), was proposed by Koza (1992). It employs a "parse tree" structure to search for solutions. GEP is comparable to GP. The chromosomes are composed of multiple genes that each gene encodes a smaller subprogram. The structural and functional organization of the linear chromosomes allows the unconstrained operation of important genetic operators such as mutation, transposition, and recombination (Ferreira, 2006). GEP can be applied to datasets where (1) the interrelationships among the relevant variables are poorly understood, (2) the ultimate solution is hard to find, (3) conventional mathematiccal analyses cannot provide analytical solutions, (4) there is a large amount of data that require examination, classification, and integration, and (5) there is a need to generate an explicit relationship between the input and output parameters (Banzhaf et al., 1998).

In this study, the GEP approach is used to estimate LAI from meteorological data. The new approach does not suffer from the above-mentioned shortcomings of the indirect approaches.

2. Materials and Methods

2.1. Data

Half-hourly meteorological records (air temperature, vapor

Table 1. Geographical Positions of the Studied Stations

Station ID	Station	Latitude (°N)	Longitude (°W)	Altitude (m)	Land cover	Available data	Data sampling interval (min)	Climatic conditions (Köppen)
				Broad-leaf stations				
US-SP1	Austin Cary	29.738	-82.218	50	Deciduous broadleaf trees	2003-2015	30	Warm temperate fully humid with hot summer
US-ChR	Chestnut Ridge	35.931	-84.332	286	Deciduous broadleaf trees	2006-2015	30	Warm temperate fully humid with hot summer
US-SP3	Donaldson	29.754	-82.163	50	Evergreen broadleaf trees	1998-2015	30	Warm temperate fully humid with hot summer
US-Dk1	Duke Forest	35.978	-79.094	169	Evergreen broadleaf trees	1998-2011	30	Warm temperate fully humid with hot summer
US-Skr	Florida Everglades	25.364	-81.077	1	Evergreen broadleaf trees	2003-2015	30	Equatorial savannah with dry winter
US-MMS	Morgan Monroe State Forest	39.323	-86.413	275	Deciduous broadleaf trees	1998-2015	30	Warm temperate fully humid with hot summer
US-MOz	Missouri Ozark	38.744	-92.200	212	Deciduous broadleaf trees	2004-2015	30	Warm temperate fully humid with hot summer
				Needle-leaf stations				
US-Blk	Black-Hills	44.158	-103.65	1718	Evergreen needleleaf forest	2004-2008	30	Warm summer continental: significant precipitation in all seasons
US-Bn1	Delta Junction 1920 Control Site	63.920	-145.378	518	Evergreen needleleaf forest	2002-2011	30	Snow fully humid cool summer
US-Ced	Cedar Bridge	39.837	-74.379	58	Evergreen needleleaf trees	2004-2015	30	Warm temperate fully humid with hot summer
US-Fuf	Flagstaff-Unmanaged Forest	35.089	-111.762	2180	Evergreen needleleaf trees	2005-2011	30	Warm temperate with dry, warm summer
US-Ho1	Howland Forest (Main)	45.204	-68.740	60	Evergreen needleleaf trees	1995-2015	30	Snow fully humid warm summer
US-Ho2	Howland Forest (Wezzst Tower)	45.209	-68.747	91	Evergreen needleleaf trees	1999-2015	30	Snow fully humid warm summer
US-Wi4	Mature Red Pine	46.739	-91.166	352	Evergreen needleleaf trees	2002-2005	30	Snow fully humid warm summer
US-Me2	Metolius Mature Ponderosa Pine	44.452	-121.557	1237	Evergreen needleleaf trees	2003-2015	30	Warm temperate with dry, warm summer

pressure deficit, and net radiation data) from 15 sites in the United States of America were used in this study. The broad-leaf and needle-leaf forests were the main land covers in the 15 sites. Figure 1 shows the location of the studied stations. Table 1 summarizes the characterizations of these sites. The meteorological data in these sites are available on the AmeriFlux archive (www.ameriflux.ornl.gov). The corresponding remotely sensed GLASS-LAI for each site during the study period was obtained from the Beijing Normal University website (http://glass-product.bnu.edu.cn/). The temporal smoothness of the GLASS-LAI product is the best among the existing LAI products (Liang et al., 2014). The spatial resolution of GLASS-LAI

products is 1 ~ 5 km and its temporal resolution is 8 days (Liang et al., 2014). In each station, the cross-correlation technique was employed to identify the meteorological variables that had the most significant influence on LAI. Based on the cross-correlation results, air temperature (T_A), net radiation (R_n), and vapor pressure deficit (VPD) had the highest correlation with LAI. This is in agreement with the outcomes reported by Savoy and Mackay (2015). Hence, it was hypothesized that the dynamics of LAI could be related to these variables.

2.2. Study Flowchart and Input Selection

This study aimed at developing generalized GEP-based models (GGEP) to simulate LAI. The meteorological data (T_A, R_n, and VPD) from 15 AmeriFlux sites with broad-leaf and needle-leaf trees were used to train and test the GGEP models. Seven input configurations were built to assess the effect of the

utilized meteorological variables (i.e., air temperature, net radiation, and vapor pressure deficit) on LAI (Table 2). Also, different data management scenarios were considered.

In the first case scenario, meteorological data from four broad-leaf sites (namely, Austin Cary, Chestnut Ridge, Donaldson, and Duke Forest; total 1,732 patterns) were pooled and used to train the GGEP models. Then, the developed broad-leaf GGEP models were tested in all the needle-leaf, and broad-leaf sites that were not used in the training step. In the second case scenario, meteorological data from four needle-leaf sites (namely, Cedar Bridge, Delta Junction, Howland Forest, and Mature Red Pine; total 1,569 patterns) were used to train the GGEP models. The trained needle-leaf GGEP models were tested in all the broad-leaf, and needle-leaf stations that were not used in the training process. Finally, training data from the eight broad-leaf and needle-leaf stations were pooled and a

Table 2. Different Input Configurations for Feeding the GGEP Model

	GGEP1	GGEP2	GGEP3	GGEP4	GGEP5	GGEP6	GGEP7
R_n	■			■	■		■
T_A		■		■		■	■
VPD			■		■	■	■

Table 3a. Preliminary Selection of GEP Fitness Function using *RRSE* Index

Fitness function based on the absolute error	*RRSE*	Fitness function based on the relative error	*RRSE*
Absolute error with selection range	0.466	Relative error with selection range	0.523
Absolute/hits	0.911	Relative/hits	0.998
Mean square error (*MSE*)	0.469	r-*MSE*	0.541
Root mean square error (*RMSE*)	0.466	r-*RMSE*	0.474
Mean absolute error (*MAE*)	**0.451**	r-*MAE*	0.566
Relative square error (*RSE*)	0.586	r-*RSE*	0.480
Root relative square error (*RRSE*)	0.565	r-*RRSE*	0.522
Relative absolute error (*RAE*)	0.490	r-*RAE*	0.634

Table 3b. Preliminary Selection of GEP Function Set using RRSE Index

Function set		RRSE
F1	$\{+, -, \times, \div\}$	0.559
F2	$\{+, -, \times, \div, \ln, e^x\}$	**0.471**
F3	$\{+, -, \times, \div, \sqrt[3]{}, \sqrt{}, x^3, x^2\}$	0.510
F4	$\{+, -, \times, \div, \sqrt[3]{}, \sqrt{}, x^3, x^2, \min(x, y)\}$	0.474
F5	$\{+, -, \times, \div, \sqrt[3]{}, \sqrt{}, x^3, x^2, \min(x, y), \max(x, y)\}$	0.483
F6	$\{+, -, \times, \div, \min(x, y), \min(x, y, z), \max(x, y), \max(x, y, z), \sin x\}$	0.489
F7	$\{+, -, \times, \div, \sqrt[3]{}, \sqrt{}, \ln, e^x, x^2, x^3\}$	0.557
F8	$\{+, -, \times, \div, \sqrt[3]{}, \sqrt{}, \ln, e^x, x^2, x^3, \sin x, \cos x, arctgx\}$	0.535
Linking functions		
1	Addition	**0.471**
2	Multiplication	0.579
3	Subtraction	0.522
4	Division	0.452

Table 4. Statistical Indices of the Broad-Leaf GGEP Models in Test Stations

Model	Input parameter	R	SI	MAE	R	SI	MAE
		Broad-leaf stations			Needle-leaf stations		
		Florida			Black Hills		
GGEP1	R_n	0.874	0.084	0.219	0.831	0.217	0.341
GGEP2	T	0.835	0.163	0.476	0.823	0.216	0.308
GGEP3	VPD	0.708	0.129	0.357	0.692	0.822	0.984
GGEP4	R_n, T	0.721	0.123	0.314	0.828	0.209	0.292
GGEP5	R_n, VPD	0.706	0.119	0.307	0.780	0.235	0.332
GGEP6	T, VPD	0.889	0.158	0.428	0.783	0.541	0.837
GGEP7	R_n, T, VPD	0.906	0.116	0.312	0.830	0.209	0.293
		Missouri Ozark			Flagstaff unmanaged forest		
GGEP1	R_n	0.985	0.139	0.316	0.672	0.455	0.311
GGEP2	T	0.964	0.214	0.472	0.896	0.265	0.181
GGEP3	VPD	0.943	0.287	0.654	0.869	0.270	0.185
GGEP4	R_n, T	0.977	0.165	0.358	0.893	0.263	0.177
GGEP5	R_n, VPD	0.901	0.300	0.400	0.685	0.440	0.299
GGEP6	T, VPD	0.936	0.292	0.626	0.881	0.294	0.200
GGEP7	R_n, T, VPD	0.981	0.163	0.363	0.894	0.265	0.181
		Morgan Monroe			Metolius		
GGEP1	R_n	0.983	0.151	0.332	0.885	0.288	0.477
GGEP2	T	0.961	0.209	0.607	0.718	0.251	0.391
GGEP3	VPD	0.938	0.300	0.809	0.638	0.480	0.456
GGEP4	R_n, T	0.973	0.207	0.474	0.735	0.241	0.385
GGEP5	R_n, VPD	0.889	0.312	0.823	0.700	0.318	0.480
GGEP6	T, VPD	0.900	0.314	0.428	0.593	0.671	0.900
GGEP7	R_n, T, VPD	0.978	0.172	0.415	0.729	0.242	0.382

Table 5. Correlation between the Meteorological Variables and the Errors of LAI Estimates from Different Models for Training Data Set

Broad-leaf					Needle-leaf				
	R_n	T_A	VPD	LAI		R_n	T_A	VPD	LAI
R_n	1.000				R_n	1.000			
T_A	0.216	1.000			T_A	0.776	1.000		
VPD	0.016	0.064	1.000		VPD	0.645	0.654	1.000	
LAI	0.424	0.184	-0.030	1.000	LAI	0.407	0.638	0.397	1.000
Err. GGEP1	-0.247	-0.115	0.001	-0.236	Err. GGEP1	0.240	0.213	0.124	0.161
Err. GGEP2	-0.279	-0.111	0.002	-0.065	Err. GGEP2	0.112	0.171	0.092	0.081
Err. GGEP3	-0.287	-0.108	0.044	-0.217	Err. GGEP3	0.192	0.227	0.097	0.147
Err. GGEP4	-0.270	-0.234	-0.003	-0.109	Err. GGEP4	0.069	0.124	0.051	0.020
Err. GGEP5	-0.250	-0.107	0.036	-0.191	Err. GGEP5	0.200	0.223	0.113	0.119
Err. GGEP6	-0.262	-0.106	0.017	-0.084	Err. GGEP6	0.099	0.146	0.074	0.054
Err. GGEP7	-0.260	-0.109	0.005	-0.076	Err. GGEP7	0.076	0.151	0.068	0.082

Note: Err. denotes the error (residual) values of each model.

global GGEP model was developed and tested in the remaining broad-leaf and needle-leaf stations.

Alternatively, the GEP models could be tested at the same sites in which they were trained. However, the major downside of such GEP models is their limited applicability to the study sites where they were trained and tested (Marti et al., 2013; 2015). Therefore, in this study, the cross-station approach was utilized to generalize the applicability of GEP models in various sites. A similar strategy was used in other studies (i.e., Kumar et al., 2009; Marti and Gasque, 2010; Beriro et al., 2013; Shiri et al., 2013, 2015; Wang et al., 2014).

2.3. Gene Expression Programming (GEP)

GEP evolves computer programs of different sizes and shapes, encoded in linear chromosomes of fixed lengths. The chromosomes are composed of multiple genes, each gene encoding a smaller subprogram. Furthermore, the structural and functional organization of the linear chromosomes allows the unconstrained operation of important genetic operators such as mutation, transposition, and recombination. The advantages of GEP are (Ferreira, 2006): (i) the chromosomes are simple entities: linear, compact, relatively small, and easy to manipulate genetically (replicate, mutate, recombine, etc.), (ii) the expres-

Table 6. Statistical Indices of the Needle-Leaf GGEP Models in Test Stations

Model	Input parameter	R	SI	MAE	R	SI	MAE
		Broad-leaf stations			Needle-leaf stations		
		Florida			Black Hills		
GGEP1	R_n	0.889	0.095	0.259	0.920	0.229	0.314
GGEP2	T_A	0.839	0.189	0.555	0.925	0.215	0.314
GGEP3	VPD	0.840	0.201	0.589	0.779	0.886	0.990
GGEP4	R_n, T	0.803	0.157	0.444	0.908	0.197	0.288
GGEP5	R_n, VPD	0.806	0.299	0.600	0.815	0.286	0.399
GGEP6	T_A, VPD	0.779	0.300	0.600	0.878	0.342	0.504
GGEP7	R_n, T, VPD	0.773	0.101	0.272	0.944	0.207	0.306
		Missouri Ozark			Flagstaff unmanaged forest		
GGEP1	R_n	0.978	0.222	0.529	0.919	0.294	0.197
GGEP2	T_A	0.972	0.233	0.346	0.864	0.264	0.178
GGEP3	VPD	0.900	0.300	0.408	0.775	0.389	0.354
GGEP4	R_n, T_A	0.836	0.230	0.900	0.894	0.257	0.174
GGEP5	R_n, VPD	0.711	0.595	1.102	0.781	0.367	0.235
GGEP6	T_A, VPD	0.881	0.527	1.348	0.882	0.275	0.189
GGEP7	R_n, T_A, VPD	0.967	0.215	0.474	0.897	0.262	0.175
		Morgan Monroe			Metolius		
GGEP1	R_n	0.984	0.169	0.391	0.708	0.243	0.398
GGEP2	T_A	0.928	0.311	0.691	0.759	0.226	0.350
GGEP3	VPD	0.884	0.309	0.685	0.700	0.316	0.473
GGEP4	R_n, T_A	0.969	0.223	0.517	0.804	0.224	0.348
GGEP5	R_n, VPD	0.856	0.388	0.821	0.702	0.245	0.389
GGEP6	T_A, VPD	0.870	0.498	0.984	0.764	0.453	0.753
GGEP7	R_n, T_A, VPD	0.975	0.176	0.400	0.760	0.225	0.348

sion trees are exclusively the expression of their respective chromosomes; they are entities upon which selection acts, and according to fitness, they are selected to reproduce with modification. However, there are also some problems regarding the GEP application. For instance, in some cases, the depth of parse tree starts growing which leads to producing nested functions (i.e., the Bloat Phenomena) (Shiri et al., 2014b). In such cases, penalization of complex models by the parsimony pressure procedure should be established for producing parsimonious relations (Poli and McPhee, 2008).

2.4. Performance Evaluation Parameters

Four statistical metrics, namely correlation coefficient (*r*), scatter index (*SI*), mean absolute error (*MAE*), and root relative square error (*RRSE*) were used to assess the performance of models:

$$r = \frac{\sum_{i=1}^{n}\left(LAI_{io} - \overline{LAI_o}\right)\left(LAI_{iM} - \overline{LAI_M}\right)}{\sqrt{\sum_{i=1}^{n}\left(LAI_{io} - \overline{LAI_o}\right)^2 \sum_{i=1}^{n}\left(LAI_{iM} - \overline{LAI_M}\right)^2}} \qquad (1)$$

$$SI = \frac{RMSE}{\overline{LAI_o}} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(LAI_{i0} - LAI_{iM}\right)^2}}{\overline{LAI_o}} \qquad (2)$$

$$MAE = \frac{1}{n}\left|\sum_{i=1}^{n}\left(LAI_{i0} - LAI_{iM}\right)\right| \qquad (3)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}\left(LAI_{i0} - LAI_{iM}\right)^2}{\sum_{i=1}^{n}\left(LAI_{io} - \overline{LAI_o}\right)^2}} \qquad (4)$$

where LAI_{i0} and LAI_{iM} are respectively the target (observed) and simulated LAI values at the *i*th time step, *n* denotes the number of time steps, $\overline{LAI_o}$ is the mean of the observed LAI values, and $\overline{LAI_M}$ represents the mean value of the LAI simulations. *RRSE* and *SI* metrics show the relative error of the estimations, while *MAE* denotes the linear error values. *RRSE* gives information on the variance of the utilized data as well as the obtained errors. *SI* presents the weighted *RMSE* of LAI estimations; thereby it can provide information on the performance of different models. A combined use of these statistical metrics allows evaluating various models.

3. Results and Discussion

3.1. Model Implementation

There are neither predefined functions in GEP (GEP produces functions and selects the one that best fits the experimental outputs) nor restrictions in the complexity and structure of the

formed functions (Guven et al., 2008). Compared to other heuristic models such as neural networks, GEP is less sensitive to the number of inputs. In GEP, the structure and coefficients for a solution are evolved simultaneously. This increases GEP's degree of freedom over other function fitting methods such as regression and other techniques that utilize a prescribed mathematical structure (Deschaine, 2014; Shiri, 2017).

Prior to evaluating the applied input configurations, a preliminary test was carried out on the parse tree of the GEP model by introducing all the input variables (i.e., T_A, VPD, and R_n) to GEP (GGEP7 model). The first step with the GEP approach is to select a fitness function. In this study, different fitness functions were evaluated (Table 3), and finally the best function was picked. As shown in Table 3, the MAE gives the lowest $RRSE$ and thus was chosen as the best fitness function. The MAE fitness function of an individual program (i) can be defined as (Ferreira, 2006):

$$MAE_i = \frac{1}{n}\sum_{j=1}^{n}\left|P_{ij} - T_j\right| \tag{5}$$

where P_{ij} denotes the simulated value of the individual program i for fitness case j (out of sample case), and T_j is the target value for fitness case j. For a perfect fit, $P_{ij} = T_j$ and $MAE_i = 0$. To evaluate the fitness f_i of an individual program i, the following equation is applied:

$$f_i = 1000\frac{1}{1+E_i} \tag{6}$$

f_i ranges from 0 to 1,000, with 1,000 corresponding to the ideal case (Ferreira, 2006). MAE may be implemented both with and without parsimony pressure. The version with parsimony pressure puts a little pressure on the size of the evolving solutions, allowing the discovery of more compact models. The parsimony pressure uses the fitness measure as raw fitness, its raw maximum fitness (rf_{max}) is 1,000, and its overall fitness (f_{ppi}) is calculated by:

$$f_{ppi} = rf_i \cdot [1 + \frac{1}{5000} \cdot \frac{S_{max} - S_i}{S_{max} - S_{min}}] \tag{7}$$

where S_i is the program size, and S_{max} and S_{min} are respectively the maximum and minimum program sizes, which are given by:

$$S_{max} = G(h+t) \tag{8}$$

$$S_{min} = G \tag{9}$$

where G is the number of genes, and h and t are the head and tail sizes, respectively. Here, maximum overall fitness (f_{ppmax}) is evaluated by the following formula:

Table 7. Statistical Indices of Global GGEP Models in Test Stations

Model	Input parameter	R	SI	MAE	R	SI	MAE
		Broadleaf stations			Needleleaf stations		
		Florida			Black Hills		
GGEP1	R_n	0.795	0.087	0.221	0.800	0.228	0.341
GGEP2	T_A	0.790	0.182	0.545	0.823	**0.215**	0.306
GGEP3	VPD	0.799	0.168	0.458	0.799	0.236	0.332
GGEP4	R_n, T_A	0.707	0.245	0.774	0.871	**0.184**	0.253
GGEP5	R_n, VPD	0.777	0.164	0.443	0.781	0.239	0.351
GGEP6	T_A, VPD	0.880	0.233	0.701	0.800	0.255	0.370
GGEP7	R_n, T_A, VPD	0.749	0.268	0.845	0.862	**0.194**	0.275
		Missouri Ozark			Flagstaff unmanaged forest		
GGEP1	R_n	0.956	0.225	0.482	0.896	0.268	0.176
GGEP2	T_A	0.966	0.259	0.618	0.906	**0.253**	0.173
GGEP3	VPD	0.703	0.607	0.995	0.770	0.381	0.251
GGEP4	R_n, T_A	0.926	0.280	0.591	0.910	**0.245**	0.165
GGEP5	R_n, VPD	0.734	0.610	0.998	0.801	0.356	0.235
GGEP6	T_A, VPD	0.898	0.325	0.700	0.889	0.276	0.189
GGEP7	R_n, T_A, VPD	0.921	0.288	0.614	0.905	**0.252**	0.173
		Morgan Monroe			Metolius		
GGEP1	R_n	0.930	0.281	0.583	0.851	0.242	0.392
GGEP2	T_A	0.928	0.285	0.627	0.768	**0.222**	0.349
GGEP3	VPD	0.705	0.453	0.970	0.689	0.250	0.393
GGEP4	R_n, T_A	0.930	0.277	0.593	0.781	**0.215**	0.335
GGEP5	R_n, VPD	0.725	0.460	0.981	0.702	0.245	0.389
GGEP6	T_A, VPD	0.877	0.347	0.780	0.800	0.288	0.475
GGEP7	R_n, T_A, VPD	0.855	0.506	1.323	0.800	**0.216**	0.339

Table 8. Mathematical Expressions of the Optimum Global GGEP Models

GGEP1	$LAI = Arctg\left[42.804\left(Exp(R_n + 6.118255)\right)^{1.5}\right] + Arctg\left[0.00013R_n^2\right] + \left[Arctg\left(Arctg(0.241R_n^{0.66})\right)\right]^2$
GGEP2	$LAI = Arctg\left[\left(-0.0152(T_A - 8.5276)\right)^3 + 7.670105\right] + Arctg\left[\left\{(12.415 - T_A)(4.0432T_A)\right\}^2\right]$ $+ Arctg\left[\dfrac{Arctg\left(-9.8597T_A\right) + T_A - 9.8597}{6.2527}\right]$
GEP4	$LAI = \cos\left(-6.023Arctg\left[\dfrac{T_A}{R_n}\right]\right) + \left[Exp(T_A) - T_A - 4.9714\right]^{0.22} + 0.656\left[Arctg(0.123T_A)\right]^3$
GEP7	$LAI = Exp\left[\sin\left(-4.110 - arctg(R_n - 14.329)\right)\right] + Arctg\left[R_n - VPD - 0.8429(T_A + 1)\right]$ $+ Exp\left[\sin^2\left(-0.4867VPD\right)\right]$

Table 9. Summary of the Performance of the Applied Models in Test Stations

Model	Input parameter	Broad-leaf test stations			Needle-leaf test stations		
		R	SI	MAE	R	SI	MAE
		Training with broad-leaf stations					
GGEP1	R_n	0.947	0.125	0.289	0.796	0.320	0.376
GGEP2	T_A	0.920	0.195	0.518	0.812	0.244	0.293
GGEP3	VPD	0.863	0.239	0.607	0.733	0.524	0.542
GGEP4	R_n, T_A	0.890	0.165	0.382	0.819	0.238	0.285
GGEP5	R_n, VPD	0.832	0.244	0.510	0.722	0.331	0.370
GGEP6	T_A, VPD	0.908	0.255	0.494	0.752	0.502	0.646
GGEP7	R_n, T_A, VPD	0.955	0.150	0.363	0.818	0.239	0.285
		Training with needle-leaf stations					
GGEP1	R_n	0.950	0.162	0.393	0.849	0.255	0.303
GGEP2	T_A	0.913	0.244	0.531	0.849	0.235	0.281
GGEP3	VPD	0.875	0.270	0.561	0.751	0.530	0.606
GGEP4	R_n, T_A	0.869	0.203	0.620	0.869	0.226	0.270
GGEP5	R_n, VPD	0.791	0.427	0.841	0.766	0.299	0.341
GGEP6	T_A, VPD	0.843	0.442	0.977	0.841	0.357	0.482
GGEP7	R_n, T_A, VPD	0.905	0.164	0.382	0.867	0.231	0.276
		Global (pooled) training					
GGEP1	R_n	0.894	0.198	0.429	0.849	0.246	0.303
GGEP2	T_A	0.895	0.242	0.597	0.832	0.230	0.277
GGEP3	VPD	0.736	0.409	0.808	0.753	0.289	0.325
GGEP4	R_n, T_A	0.854	0.267	0.653	0.854	0.215	0.251
GGEP5	R_n, VPD	0.745	0.411	0.807	0.761	0.280	0.325
GGEP6	T_A, VPD	0.885	0.302	0.727	0.830	0.273	0.345
GGEP7	R_n, T_A, VPD	0.842	0.354	0.927	0.856	0.221	0.262

$$f_{pp\max} = 1.0002rf_{\max} \tag{10}$$

The second step consists of choosing the set of terminals and functions to create the chromosomes. In this study, the terminal set included *TA*, R_n, and *VPD*. The selection of an appropriate function set was carried out by examining different function sets (Table 3). The third step involves choosing the chromosomal architecture. Following Ferreira (2001; 2006), length of the head is set to 8 ($h = 8$), and three genes per chromosome is used. In the fourth step, the linking function is chosen. The linking function for algebraic sub-trees must be "addition"

or "multiplication" (Ferreira, 2001a). Results showed that the addition linking function produced better results (Table 3). The final step is to choose the genetic operators. Following the literature (i.e., Roushangar et al., 2014a, b; Karimi et al., 2016), the default values of GeneXpro were used in this study.

3.2. Models Evaluation and Discussion

3.2.1. Broad-Leaf GGEP Models

Table 4 summarizes the testing statistical indices of the GGEP models trained with data in broad-leaf stations. As shown, GGEP1 (using R_n as the unique input variable) performs better

than the other models in all broad-leaf test stations, followed by the GGEP7 and GGEP4. However, in the needle-leaf stations, GGEP4 produces the most accurate results followed by GGEP7 and GGEP2. Analyzing the GEP models per station using a local GEP model (not presented here) confirms these findings. Different performances of the models in broad-leaf and needle-leaf stations might be linked to the land covers of the stations where the net radiation (R_n) would be affected by the land cover and the percentage of the land surface that has been shaded by the vegetation. Nonetheless, such performance discrepancy might be assessed through the correlation between the input variables (i.e., T_A, VPD, and R_n) and the error values in Table 5. As indicated in Table 5, the highest correlation in broad-leaf land cover belonged to R_n-*LAI*, followed by T_A-*LAI* (positive correlation) and VPD-*LAI* (negative correlation). Similarly, the highest correlation of the models error corresponded to R_n. For the GGEP2 (with T_A as the only input), GGEP3 (with VPD as the only input), and GGEP6 (with T_A and VPD as inputs) models, the error values showed the highest correlation with R_n although R_n was not used in these models. Thus, it can be concluded that R_n strongly affects the LAI simulations in broad-leaf forests. For the needle-leaf stations, the highest correlation was found between T_A and LAI. Moreover, a high influence of T_A on LAI simulations can be observed in needle-leaf stations.

3.2.2. Needle-Leaf GGEP Models

Needle-leaf GGEP models (trained with data in needle-leaf stations) were tested in broadleaf and needle-leaf stations. Testing statistics of the needle-leaf GGEP models are presented in Table 6. In the broad-leaf test stations, GGEP1 produced the most accurate results, followed by GGEP7 and GGEP4. Also, in the needle-leaf stations, GGEP4 was the best model and GGEP7 and GGEP2 models were ranked as the second and third accurate models. Similar to the broad-leaf stations, the highest correlation values for needle-leaf stations were observed in R_n-*LAI*, followed by T_A-*LAI* (positive correlation) and VPD-*LAI* (negative correlation).

Both the broad-leaf and needle-leaf GGEP models show a high correlation in R_n-*LAI*, and to a lesser extent in T_A-*LAI*. These findings indicate that variations in LAI are highly affected by changes in net radiation and air temperature. This is in agreement with the results reported by Zhu et al. (2013). This might be linked to the higher influence of temperature and solar radiation (light) on photosynthesis as discussed by Hew et al. (1969) and Gomez et al. (1998).

3.2.3. Global GGEP Models

The Global GGEP models were calibrated with data from the broad-leaf and needle-leaf training stations. Then, they were tested in the broad-leaf and needle-leaf testing stations. Table 7 gives the testing statistics of the global GGEP models. As shown, GGEP1, GGEP7 and GGEP4 models were the most accurate models for the broad-leaf test stations, while GGEP4, GGEP7 and GGEP2 were the best models for needle-leaf test stations. Comparing Tables 4, 6 and 7 showed that the global GGEP models gave more accurate results than the needle-leaf GGEP mod-

els in needle-leaf testing stations, while the broad-leaf GGEP models outperformed the global GGEP models in broad-leaf testing stations. Figure 2 shows the *SI* reduction (decrease in *SI* values of the models) for the global GGEP models application in the needle-leaf testing stations, which was computed as follows:

$$SIreduction = 1 - \frac{SIglobal}{SIneedleleaf} \qquad (11)$$

It is evident that the global GGEP model (which includes the broad-leaf and needle-leaf training stations) has improved the accuracy of the LAI predictions because it used all the data. Among the temperature-based GGEP models, the maximum *SI* reduction was seen for Flagstaff station, while the maximum decrease of *SI* for the GGEP4 and GGEP7 models were for to Black Hills testing station. Figure 3 shows the observed versus simulated LAI values of the best global GGEP models in the needle-
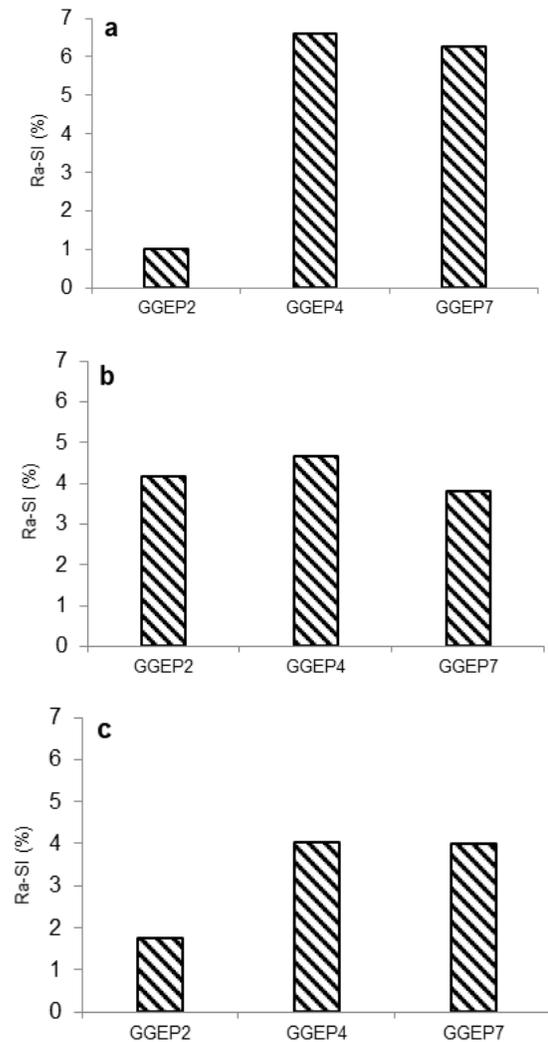
Figure 2. *Ra-SI* values for global GGEP models in needle-leaf test stations: a) Black Hills, b) Flagstaff, and c) Metolius stations.
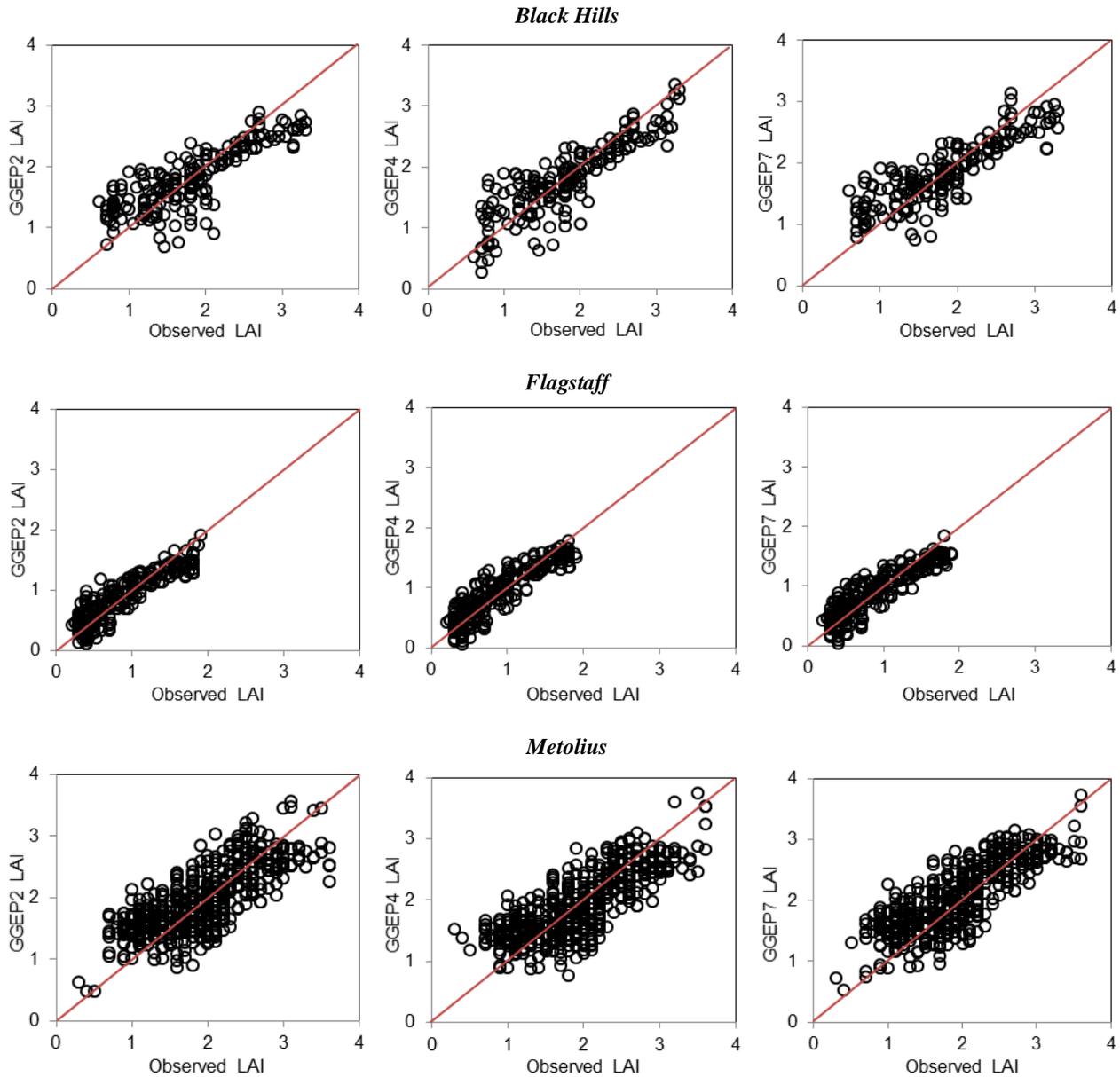
Figure 3. *Ra-SI* values for global GGEP models in needle-leaf test stations: a) Black Hills, b) Flagstaff, and c) Metolius stations.

leaf testing stations. As indicated, most of the points fall around the 1:1 line, and GGEP7 outperforms GGEP2 and GGEP4. All of these results indicate that the global GGEP7 model can predict LAI accurately over the needle-leaf stations. Table 8 presents the mathematical expressions for the optimum global GGEP models. Compared to the other GEP models, the single-input GEP2 model generated the most complex expression. This is because of difficulties in relating LAI to only air temperature. However, it should be noted that the parsimony pressure tool is used to compress the model tree. Francone and Deschaine (2004) and Deschaine (2014) stated that there are different viewpoints on the parsimony tool because it is a binding or forcing function on an otherwise freely search strategy (i.e., uncon-

strained with maximum degree of freedom) when the program is allowed to grow unrestricted. Therefore, giving different weights to input variables of different GEP models may alter the complexity of the final expression.

Table 9 presents the performance indicators of the GGEP models. As indicated, the GGEP models trained in broad-leaf (needle-leaf) sites generate less accurate results in needle-leaf (broad-leaf) testing sites. Also, the global GEP models can produce accurate results in both training and testing stations. Therefore, the models trained in broad-leaf (needle-leaf) sites can only be used to estimate LAI in broad-leaf (needle-leaf) sites, while the global GEP models can be employed for predicting LAI in both broad-leaf and needle-leaf sites.

4. Conclusions

In this study, the GEP method was employed to characterize the relationships between meteorological variables and leaf area index (LAI). The meteorological variables (i.e., T_A, VPD, and R_n) and GLASS-LAI data were used to develop and generalize the GEP models. The models were trained using meteorological data from several AmeriFlux stations with broad-leaf and needle-leaf trees and then tested in a number of independent stations. Finally, a GGEP model was developed by taking into account meteorological data in all the broad-leaf and needle-leaf training stations. The results showed that:

1. GEP is a promising tool for characterizing nonlinear relationships between the meteorological variables (i.e., T_A, VPD, and R_n) and LAI.

2. The broad-leaf GGEP model that used only R_n generated the most accurate results in broad-leaf stations. The broad-leaf GGEP model with T_A and R_n as inputs yielded the best results in the needle-leaf stations.

3. The needle-leaf GGEP model using T_A and R_n led to the best results in the needle-leaf stations, while the GGEP model that relied on R_n was ranked as the best model for the broad-leaf stations.

4. The needle-leaf GGEP models performed better than the broad-leaf GGEP models in the needle-leaf stations, and the broad-leaf GGEP model led to better results in the broad-leaf stations.

5. The global GGEP models could accurately estimate LAI in both the needle-leaf and broad-leaf stations.

The results revealed that the GGEP models can provide accurate LAI values in the testing sites. This is one of the significant outcomes of this study because the developed models are not site-specific. Future studies should focus on estimating LAI in different sites with various land covers and adopting more data management scenarios.

References

Asner, G.P., Scurlock, J.M.O., and Hicke, J.F. (2003). Global synthesis of leaf area index observation: implications for ecological and remote sensing studies, *Glob. Ecol. Biogeogr.*, 12, 191-205. https://doi.org/10.1046/j.1466-822X.2003.00026.x.

Banzhaf, W., Nordin, P., Keller, P.E., and Francone, F.D. (1998). *Genetic Programming. Morgan Kaufmann*, San Francisco, CA, pp. 512. https://doi.org/10.1007/BFb0055923.

Bateni, S.M., Entekhabi, D., and Castelli, F. (2013). Mapping evaporation and estimation of surface control of evaporation using remotely sensed land surface temperature from a constellation of satellites, *Water Resour. Res.*, 49(2), 950-968. https://doi.org/10.1002/wrcr.20 071.

Bateni, S.M., Entekhabi, D., and Jeng, D.S. (2013). Variational assimilation of land surface temperature and the estimation of surface energy balance components, *J. Hydrol.*, 481, 143-156. https://doi.org/10.1016/j.jhydrol.2012.12.039.

Beriro, D.J., Abrahart, R.J., Nathanail C.P., Moreno, J., and Bawasir, A.S. (2013). A typology of different development and testing options for symbolic regression modelling of measured and calculated datasets, *Environ. Model. Soft.*, 47, 29-41. https://doi.org/10.1016/j.envsoft. 2013.03.020.

Biudes, M.S., Machado, N.G., Danelichen, V.H., Souza, M.C., Vourlitis, G.L., and Nogueira Jde S. (2014). Ground and remote sensing-based measurements of leaf area index in a transitional forest and seasonal flooded forest in Brazil, *Int. J. Biometeorol.*, 58(6), 1181-1193. https://doi.org/10.1007/s00484-013-0713-4.

Bonan, G. (1993). Importance of LAI and forest type when estimating photosynthesis in boreal forests, *Remote Sens. Environ.*, 43, 303-314. https://doi.org/10.1016/0034-4257(93)90072-6.

Cáceres, M.D., Martínez-Vilalta, J., Coll, L., Llorens, P., Casals, P., Poyatos, R., Pausas, J.G., and Brotons, L. (2015). Coupling a water balance model with forest inventory data to predict drought stress: the role of forest structural changes vs. climate changes, *Agric. Forest Meteor.*, 213, 77-90. https://doi.org/10.1016/j.agrformet.20 15.06.012.

Chason, J.W., Baldocchi, D.D., and Huston, M.A. (1991). Comparison of direct and indirect methods for estimating forest canopy leaf-area, *Agric. Forest Meteor.*, 57, 107-128. https://doi.org/10. 1016/0168-1923(91)90081-Z.

Chen, M., Willgoose, G.R., and Saco, P.M. (2015). Investigating the impact of leaf area index temporal variability on soil moisture predictions using remote sensing vegetation data, *J. Hydrol.*, 522, 274-284. https://doi.org/10.1016/j.jhydrol.2014.12.027.

Colaizzi, P., Evett, S.R., Brauer, D.K., Howell, T.R., Tolk, J.R., and Copeland, K.S. (2017). Allometric method to estimate leaf area index for row crops, *Agron. J.*, 109(3), 883-894. https://doi.org/10. 2134/agronj2016.11.0665.

Coopersmith, E.J., Cosh, M.H., and Daughtry, C.S.T. (2014). Field-scale moisture estimates using COSMOS sensors: A validation study with temporary networks and leaf-area-indices, *J. Hydrol.*, 519(Part A), 637-643.

Deschaine, L.M. (2014). *Decision support for complex planning challenges: Combining expert systems, engineering-oriented modelling, machine learning, information theory, and optimization technology*, Chalmers University of Technology, Sweden, pp. 233.

Dou, X. and Yang, Y. (2018). Estimating forest carbon fluxes using four different data-driven techniques based on long-term eddy covariance measurements: Model comparison and evaluation, *Sci. Total. Environ.*, 627, 78-94. https://doi.org/10.1016/j.scitotenv.201 8.01.202.

Dunea, D. and Moise, V. (2008). Artificial neural networks as a support for leaf are index modeling in crop canopies, *12th WSEAS International Conference on COMPUTERS*, Heraklion, Greece, 2008, pp: 440-445.

Everingham, Y.L., Smyth, C.W., and Inman-Bamber, N.G. (2009). Ensemble data mining approaches to forecast regional sugarcane crop production, *Agric. Forest Meteor.*, 149(3-4), 689-696. https://doi.org/10.1016/j.agrformet.2008.10.018.

Ferreira, C. (2001). Gene expression programming: a new adaptive algorithm for solving problems, *Compl. Sys.*, 13(2), 87-129.

Ferreira, C. (2006). *Gene expression programming: Mathematical Modeling by an artificial intelligence*, Springer, Berling, Heidelberg New York, pp.478. https://doi.org/10.1007/3-540- 32849-1_2.

Francone, F.D. and Deschaine., L.M. (2004). Getting it right at the very start-building project models where data is expensive by combining human expertise, machine learning and information theory, *2004 Business and Industry Symposium*, Washington, DC, 2004.

Franklin, S.E., Lavigne, M.B., Deuling, M.J., Wulder, M.A., and Hunt, E.R. (1997). Estimation of forest leaf area index using remote sensing and GIS data for modelling net primary production, *Int. J.*

Remote Sens., 18(16), 3459-3471. https://doi.org/10.1080/0143116 97216973.

Goldberg, D.E. (1989). *Genetic algorithms in search, optimization, and machine learning*, Addison -Wesley, Boston, pp. 432.

Gomez, I., Perez-Rodriguez, E., Vinegla, B., Figureoa, F.L., and Karsten, U. (1998). Effects of solar radiation on photosynthesis, UV-absorbing compounds and enzyme activities of the green alga Dasycladus vermicularis from southern Spain, *J. Photochem. Photobiol. B: Biology.*, 47(1), 46-57. https://doi.org/10.1016/S1011 -1344(98)00199-7.

Guven, A., Aytek, A., Yuce, M.I., and Aksoy, H. (2008). Genetic programming-based empirical model for daily reference evapotranspiration estimation, *Clean Soil, Air, Water.*, 36(10-11), 905-912. https://doi.org/10.1002/clen.200800009.

He, B.B., Quan, X.W., and Xing, M.F. (2013). Retrieval of leaf area index in alpine wetlands using a two-layer canopy reflectance model, *Int. J. Appl. Earth. Obser. Geoinformation.*, 21, 78-91. https://doi.org/10.1016/j.jag.2012.08.014.

Hew, C.S., Krotkov, G., and Canvin, D.T. (1969). Effects of temperature on photosynthesis and CO2 evolution in light and darkness by green leaves, *Plant. Physiol.*, 44(5), 671-677. https://doi.org/10.11 04/pp.44.5.671.

Heinsch, F.A., Zhao, M.S., Running, S.W., Kimball, J.S., Nemani, R.R., Davis, K.J., Bolstad, P.V., Cook, B.D., Desai, A.R., and Ricciuto, D.M. (2006). Evaluation of remote sensing based terrestrial productivity from modis using regional tower eddy flux network observations, *IEEE Trans. Geosci. Remote Sens.*, 44, 1908-1925. https://doi./ 10.1109/TGRS.2005.853936.

Jacquemoud, S., acour, C., Poilve, H., and Frangi, J.P. (2000). Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode, *Remote Sens. Environ.*, 74(3), 471-481. https://doi.org/10.1016/S0034-4257(00)00139-5.

Jonckheere, I., Fleck, S., Nackaerts, K., Muysa, B., Coppin, P., Weiss, M., and Baret, F. (2004). Review of methods for in situ leaf area index determination Part I. Theories, sensors and hemispherical photography, *Agric. Forest Meteor.*, 121, 19-35. https://doi.org/10.1016/j.agrformet.2003.08.027.

Karimi, S., Kisi, O., Kim, S., Nazemi, A.H., and Shiri, J. (2016). Modelling daily reference evapotranspiration in humid locations of South Korea using local and cross-station data management scenarios, *Int. J. Climatol.*, 33, 3238-3246.

Karimi, S., Shiri, J., Kisi, O., and Xu, T. (2017). Forecasting daily streamflow values: assessing heuristic models, *Hydrol. Res.*, 49 (3), 658-669. https://doi.org/10.2166/nh.2017.111.

Karimi, S., Sadraddini, A.A., Nazemi, A.H., Xu, T., and Fakherifard, A. (2018). Generalizability of gene expression programming and random forest methodologies in estimating cropland and grassland leaf area index, *Comput. Electron. Agric.*, 144, 232-240. https://doi.org/10.1016/j.compag.2017.12.007.

Kim, S., Shiri, J., Singh, V.P., Kisi, O., and Landeras, G. (2015). Predicting daily pan evaporation by soft computing models with limited climatic data, *Hydrol. Sci. J.*, 60(6), 1120-1136. https://doi.org/10.1080/02626667.2014.945937.

Kisi, O., Kim, S., and Shiri, J. (2013a). Estimation of dew point temperature using neuro-fuzzy and neural network techniques, *Theor. Appl. Climatol.*, 114, 365-373. https://doi.org/10.1007/s00704-013 -0845-9.

Kisi, O., Akbari, N., Sanatipour, M., Hashemi, A., Teimourzadeh, K., and Shiri, J. (2013b). Modeling of dissolved oxygen in river water using artificial intelligence techniques, *J. Environ. Inform.*, 22(2), 92-101. https://doi.org/10.3808/jei.201300248.

Kisi, O. and Shiri, J. (2014). Prediction of long-term monthly air temperature using geographical inputs, *Int. J. Climatol.*, 34, 179- 186. https://doi.org/10.1002/joc.3676.

Kisi, O., Pour Ali Baba, A., and Shiri, J. (2012). Generalized neuro-fuzzy models for estimating daily pan evaporation values from weather data, *J. Irrig. Drain. Eng.*, 138(4), 1-14. https://doi.org/10. 1061/(ASCE)IR.1943-4774.0000403.

Kisi, O. (2015). Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree, *J. Hydrol.*, 528, 312-320. https://doi.org/10.1016/j.jhyd rol.2015.06.052.

Koza, J.R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, The MIT Press, Cambridge, MA, pp.840.

Kumar, K., Raghuwanshi, N.S., and Singh, R. (2009). Development and validation of GANN model for evapotranspiration estimation, *J. Irrig. Drain. Eng.*, 14(2), 131-140. https://doi.org/10.1061/(AS CE)1084-0699(2009)14:2(131).

Landeras, G., Lopez, J.J., Kisi, O., and Shiri, J. (2012). Comparison of Gene Expression Programming with neuro-fuzzy and neural network computing techniques in estimating daily incoming solar radiation in the Basque Country (Northern Spain), *Energy Conver. Manag.*, 62, 1-13. https://doi.org/10.1016/j.enconman.2012.03.025.

Li, H., Zhang, Y., Chiew, F.S.H., and Xu, S. (2009). Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index, *J. Hydrol.*, 370(1-4), 155-162. https://doi.org/10. 1016/j.jhydrol.2009.03.003.

Li, R., Li, C.J., Dong, Y.Y., Liu, F., Wang, J.H., Yang, X.T., and Pan, Y.C. (2011). Assimilation of Remote Sensing and Crop Model for LAI Estimation Based on Ensemble Kaiman Filter, *Agric. Sci. China*, 10(10), 1595-1602. https://doi.org/10.1016/S1671-2927(11) 60156-9.

Liang, S., Zhang, X., Xiao, Z., Cheng, J., Liu, Q., and Zhao, X. (2014). Leaf Area Index. *Global LAnd Surface Satellite (GLASS) Products, Springer Briefs in Earth Sciences*, pp.3-31. https://doi. org/10.100 7/978-3-319-02588-9_2.

Liu, S., Xie, Z., and Zeng, Y. (2016). Estimation of streamflow in ungauged basins using a combined model of black-box model and semi-distributed model: case study in the Yingluoxia watershed, *J. Beijing Normal University*, 52 (3), 393-401. (Published in Chinese with English abstract).

Lovell, J., Jupp, D.L., Culvenor, D., and Coops, N. (2003). Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests, *Can. J. Remote Sens*., 29:5, 607-622. https://d oi.org/10.5589/m03-026.

Makki, M. and Homma, K. (2014). Empirical regression models for estimating multiyear leaf area index of rice from several vegetation indices at the field scale, *Remote Sens.*, 6(6), 4764-4779. https:// doi.org/10.3390/rs6064764.

Mao, F., Li, X., Du, H., Zhou, G., Han, N., Xu, X., Liu, Y., Chen, L., and Cui, L. (2017). Comparison of two data assimilation methods for improving MODIS LAI time series for Bamboo forests, *Remote Sens.*, 9(401), 1-17. https://doi.org/10.3390/rs9050401.

Marti, P. and Gasque, M. (2010). Ancillary data supply strategies for improvement of temperature-based ETo ANN models, *Agric. Water. Manage.*, 97, 939-955. https://doi.org/10.1016/j.agwat.2010.02.002.

Marti, P., Shiri, J., Duran-Ros, M., Arbat, G., Cartagena, F.R., and Puig-Bargues, J. (2013). Artificial neural networks vs. gene expressions programming for estimating outlet dissolved oxygen in micro irrigation sand filters fed with effluents, *Comput. Electron. Agric.*, 99, 176-185. https://doi.org/10.1016/j.compag.2013.08.016.

Marti, P., González-Altozano, P., López-Urrea, R., Mancha, L.A., and Shiri, J. (2015). Modeling reference evapotranspiration with calculated targets: Assessment and implications, *Agric. Water. Manage.*, 149, 81-90.https://doi.org/10.1016/j.agwat.2014.10.028.

Meroni, M., Colombo, R., and Panigada, C. (2004). Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations, *Rem. Sens. Environ.*, 92(2), 195-206. https://doi.org/10.1016/j.rse.2004.06.005.

Nemani, R., Pierce, L., Running, S., and Band, L. (1993). Forest ecosystem processes at the watershed scale: sensitivity to remotely-

sensed leaf area index estimates, *Int. J. Remote Sens.*, 14, 2519-2534. https://doi.org/10.1080/01431169308904290.

Norman, J.M. and Campbell, G.S. (1989). *Canopy structure*. Pearcy, R.W., Ehlringer, J., Mooney, H.A., Rundel, P.W. (Eds.), Plant Ecology: Field Methods and Instrumentation, Chapman & Hall, London, pp. 301-325. https://doi.org/10.1007/978-94-009-2221-1_14.

Olivas, P.C., Oberbauer, S.F., Clark, D.B., Clark, D.A., Ryan, M.G., Brien, J.J., and Ordonez, H. (2013). Comparison of direct and indirect methods for assessing leaf area index across a tropical rain forest landscape, *Agric. Forest Meteor.*, 177, 110-116. https://doi.org/10.1016/j.agrformet.2013.04.010.

Pierce, L.L. and Running, S.W. (1988). Rapid estimation of coniferous forest leaf area index using a portable integrating radiometer, *Ecology*, 69, 1762-1767. https://doi.org/10.2307/1941154.

Ploi, R. and McPhee, N.F. (2008). *Covariant parsimony pressure for genetic programming*, Technical report CES-480, ISSN: 1744-8050.

Pour Ali Baba, A., Shiri, J., Kisi, O., FakheriFard, A., Kim, S., and Amini, A. (2013). Estimating daily reference evapotranspiration using available and estimated climatic data by adaptive neuro-fuzzy inference system (ANFIS) and Artificial neural network (ANN), *Hydrol. Res.*, 44(1), 131-146. https://doi.org/10.2166/nh.2012.074.

Qu, Y., Han, W., Fu, L., and Li, C. (2014). LAINet - a wireless sensor network for coniferous forest leaf area index measurement: Design, algorithm and validation, *Comput. Electron. Agric.*, 108, 200-208. https://doi.org/10.1016/j.compag.2014.08.003.

Ripple, W.J., Wang, S., Isaacson, D.L., and Paine, D.P. (1991). A preliminary comparison of Landsat Thematic Mapper and SPOT-1-HRV multispectral data for estimating coniferous forest volume, *Int. J. Remote Sens.*, 12, 1971-1977. https://doi.org/10.1080/0143 1169108955230.

Roushangar, K., Vojoudi, F., and Shiri, J. (2014a). Modeling river total bed material load discharge using artificial intelligence approaches (based on conceptual inputs), *J. Hydrol.*, 514, 144-122. https://doi.org/10.1016/j.jhydrol.2014.03.065.

Roushangar, K., Mouaze, D., and Shiri, J. (2014b). Evaluation of genetic programming-based models for simulating friction factor in alluvial channels, *J. Hydrol.*, 517, 1154-1161. https://doi.org/10.1016/j.jhydrol.2014.06.047.

Sellers, P.J., Hall, F.G, Asrar, G. Strebel, D.E., and Murphy, R.E. (1988). The first ISLSCP field experiment (FIFE), *Bull. American. Meteor. Soc.*, 69, 22-27. https://doi.org/10.1175/1520-0477(1988) 069<0022:TFIFE>2.0.CO;2.

Shiri, J., Nazemi, A.H., Sadraddini, A.A., Landeras, G., Kisi, O., and Marti, P. (2013). Global cross-station assessment of neuro-fuzzy models for estimating daily reference evapotranspiration, *J. Hydrol.*, 480, 46-57. https://doi.org/10.1016/j.jhydrol.2012.12.006.

Shiri, J., Kim, S., and Kisi, O. (2014a). Estimation of daily dew point temperature using genetic programming and neural networks approaches, *Hydrol. Res.*, 45(2), 165-181. https://doi.org/10.2166/nh.2013.229.

Shiri, J., Nazemi, A.H., Sadraddini, A.A., Landeras, G., Kisi, O., Fakheri Fard, A., and Marti, P. (2014b). Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran, *Comput. Electron. Agric.*, 108, 230-241. https://doi.org/10.1016/j.compag.2014.08.007.

Shiri, J., Sadraddini, A.A., Nazemi, A.H., Kisi, O., Landeras, G., Fakheri Fard, A., and Marti, P. (2014c). Generalizability of gene expression programming-based approaches for estimating daily reference evapotranspiration in coastal stations of Iran, *J. Hydrol.*, 508, 1-11. https://doi.org/10.1016/j.jhydrol.2013.10.034.

Shiri, J., Sadraddini, A.A., Nazemi, A.H., Marti, P., Fakheri Fard, A., Kisi, O., and Landeras, G. (2015). Independent testing for assessing the calibration of the Hargreaves-Samani equation: New heuristic alternatives for Iran, *Comput. Electron. Agric.*, 117, 70-80. https://doi.org/10.1016/j.compag.2015.07.010

Shiri, J. (2017). Evaluation of FAO56-PM, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of Iran, *Agric. Water. Manage.*, 188, 101-114. https://doi.org/10.1016/j.agwat.201 7.04.009

Shiri, J., Keshavarzi, A., Kisi, O., Iturraran-Viveros, U., Bagherzadeh, A., Mousavi, R., and Karimi, S. (2017a). Modeling soil cation exchange capacity using soil parameters: Assessing the heuristic models, *Comput. Electron. Agric.*, 135, 242-251. https://doi.org/10.1016/j.compag.2017.02.016

Shiri, J., Keshavarzi, A., Kisi, O., Karimi, S., and Iturraran-Viveros, U. (2017b). Modeling soil bulk density through a complete data scanning procedure: Heuristic alternatives, *J. Hydrol.*, 549, 592-602. https://doi.org/10.1016/j.jhydrol.2017.04.035

Shrestha, N.K., and Shukla, S. (2015). Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment, *Agric. Forest Meteor.*, 200, 172-184. https://doi.org/10.1016/j.agrformet.2014.09.025

Sonnentag, O., Talbot, J., Chen, J.M., and Roulet, N.T. (2007). Using direct and indirect measurements of leaf area index to characterize the shrub canopy in an ombrot rophic peatland, *Agric. Forest Meteor.*, 144, 200-212. https://doi.org/10.1016/j.agrformet.2007.03.001

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

Walthall, C., Dulaney, W., Anderson, M., Norman, J., Fang, H., and Liang, S. (2004). A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery, *Remote Sens. Environ.*, 92, 465-474. https://doi.org/10.1016/j.rse.2004.06.003

Wang, Z., Wu, P., Zhao, X., Cao, X., and Gao, Y. (2014). GANN models for reference evapotranspiration estimation developed with weather data from different climatic regions, *Theor. Appl. Climatol.*, 116, 481-489. https://doi.org/10.1007/s00704-013-0967-0

Watson, D.J. (1947). Comparative physiological studies in the growth of field crops. I: Variation in net assimilation rate and leaf area between species and varieties, and within and between years, *Ann. Bot.*, 11, 41-76. https://doi.org/10.1093/oxfordjournals.aob.a083148

Weiss, M., Baret, F., Smith, G.J., Jonckheere, I., and Coppin, P. (2004). Review of methods for in situ leaf area index (LAI) determination. Part II. Estimation of LAI, errors and sampling, *Agric. Forest Meteor.*, 121, 37-53. https://doi.org/10.1016/j.agrformet.2003.08.001

Welles, J.M. (1990). Some indirect methods of estimating canopy structure, *Rem. Sens. Rev.*, 5, 31-43. https://doi.org/10.1080/027572590 09532120.

Wulder, M.A., LeDrew, E.F., Franklin, S.E., and Lavigne, M.B. (1998). Aerial Image Texture Information in the Estimation of Northern Deciduous and Mixed Wood Forest Leaf Area Index (LAI), *Remote Sens. Environ.*, 64(1), 64-76. https://doi.org/10.1016/S0034-4257(97)00169-7.

Xiao, Z. Q., Liang, S., Wang, J. D., Chen, P., Yin, X.J., Zhang, L.Q., and Song, J.L. (2014). Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance, *IEEE Trans. Geosci. Remote Sens.*, 52, 209-223. https://doi.org/10.1109/TGRS.2013.2237780.

Xu, T., Bateni, S.M., Liang, S., Entekhabi, D., and Mao, K. (2014). Estimation of surface turbulent heat fluxes via variational assimilation of sequences of land surface temperatures from Geostationary Operational Environmental Satellites, *J. Geoph. Res. Atm.*, 119(10), 780-10,798.

Yin, G., Li, J., Liu, Q., Fan, W., Xu, B., Zeng, Y., and Zhao, J., (2015). Regional leaf area index retrieval based on remote sensing: the role of radiative transfer model selection, *Remote Sens.*, 7(4), 4604-4625. https://doi.org/10.3390/rs70404604.

Zhou, H., Chen, P., and Wang, J. (2013). A data-based mechanistic

assimilation method to estimate time series LAI, *IGARSS 2013 - 2013 IEEE International Geoscience and Remote Sensing Symposium*. https://doi.org/10.1109/IGARSS.2013.6723235.

Zhu, X., Zhao, Y., and Feng, X. (2013). A methodology for estimating leaf area index by assimilating remote sensing data into crop model based on temporal and spatial knowledge, *China Geogra.*

Sci., 23(5), 550-561. https://doi.org/10.1007/s11769-013- 0621-x.

Zhao, K., Garcia, M., Liu, S., Guo, Q., Chen, G., Zhang, X., Zhou, Y., Meng, X. (2015). Terrestrial lidar remote sensing of forests: Maximum likelihood estimates of canopy profile, leaf area index, and leaf angle distribution, *Agric. Forest. Meteor.*, 100-113, 209-210. https://doi.org/10.1016/j.agrformet.2015.03.008.