Simultaneous retrieval of sugarcane variables from Sentinel-2 data using Bayesian regularized neural network

Mohammad Hajeb[a], Saeid Hamzeh[a,*], Seyed Kazem Alavipanah[a], Lamya Neissi[b], Jochem Verrelst[c]

[a]Department of Remote Sensing and GIS, Faculty of Geography, University of Tehran, Tehran, Iran

[b]Sugarcane Research and Training Institute and By-products Development of Khuzestan, Khuzestan, Iran

[c]Image Processing Laboratory (IPL), Parc Científic, Universitat de Val encia, València, Spain

Abstract

Quantifying biophysical and biochemical vegetation variables is of great importance in precision agriculture. Here, the ability of artificial neural networks (ANNs) to generate multiple outputs is exploited to simultaneously retrieve Leaf area index (LAI), leaf sheath moisture (LSM), leaf chlorophyll content (LCC), and leaf nitrogen concentration (LNC) of sugarcane from Sentinel-2 spectra. We apply a type of ANNs, Bayesian Regularized ANN (BRANN), which incorporates the Bayes' theorem into a regularization scheme to tackle the overfitting problem of ANN and improve its generalizability. Quantitatively assessing the result accuracy indicated RMSE values of 0.48 (m^2/m^2) for LAI, 2.36 (% wb) for LSM, 5.85 ($\mu g/cm^2$) for LCC, and 0.23 (%) for LNC, applying simultaneous retrieval. It was demonstrated that simultaneous retrievals of the variables outperformed the individual retrievals. The superiority of the proposed BRANN over a conventional ANN trained with the Levenberg-Marquardt algorithm was confirmed through statistical comparison of their results. The model was applied over the entire Sentinel-2 images to map the considered variables. The maps were probed to qualitatively evaluate the model performance. The results indicated that the retrievals reasonably represent spatial and temporal variations of the variables. Generally, this study demonstrated that the BRANN simultaneous retrieval model can provide faster and more accurate retrievals than those obtained from conventional ANNs and individual retrievals.

Keywords

Vegetation parameter retrieval; Sentinel-2; Sugarcane; Multi-output ANN; Bayesian regularization

[*]Corresponding author at: Faculty of Geography of University of Tehran, No. 1, Zarrinkoub Alley, Vesal Shirazi St. Enqelab St. Post Box: 14155-6465, Tehran 1417853933, Iran. saeid.hamzeh@ut.ac.ir (S. Hamzeh).

1 Introduction

Vegetation covers are explained by their structural, biophysical, and biochemical variables. These variables play an important role in climate, hydrology, and ecology models, as well as in understanding the agricultural ecosystem processes (Baret et al., 2007; Sellers et al., 1997). In precision farming, monitoring these variables has been targeted to evaluate nutrition status and plant growth, leading to food security at national and regional levels (Weiss et al., 2020). Near-real time mapping the vegetation variables can therefore help better management of water and soil resources which preserves productivity at lower environmental costs (Moran et al., 1997), leading to sustainable development.

Sugarcane, a tall perennial grass in the genus Saccharum, is one of the economic products in tropics and subtropics regions (Som-Ard et al., 2021). Due to contribution of sugarcane in sugar and ethanol bio-fuel production (Moraes et al., 2015), its global demand is rapidly growing, which requires its cultivation to be as efficient as possible. Quantifying sugarcane biophysical and biochemical variables such as leaf area index (LAI), chlorophyll content, nitrogen content, and water content which are involved in important physical and physiological processes is of great significance in cultivating it with more productivity. LAI plays a critical role in the energy, water, and carbon exchanges between the continents and the atmosphere (Sellers et al., 1997). LAI is a good indicator of sugarcane growth and yield, so its variations during the crop cycle are used in sugarcane growth models (Teruel et al., 1997). Chlorophyll contributes to verifying vegetation health, and physiological and nutritional status (Delegido et al., 2010). Like in all plants, chlorophyll in sugarcane is correlated to vegetation stress, photosynthetic capacity, and productivity (Oliveros et al., 2021). Nitrogen is a macro-nutrient that once absorbed by plants becomes a biochemical variable participating in constitution of organics such as protein, chlorophyll, and nucleic acid (Féret et al., 2021). In the case of sugarcane, plant nitrogen affects leaf and stalk growth (Miphokasap et al., 2012) and consequently crop yield and sugar production (Wiedenfeld, 1995). Monitoring the response of sugarcane growth to varying nitrogen application rates demonstrated the significant benefit of increased nitrogen availability (Sofonia et al., 2019). Water content is one of the main controlling factors of photosynthesis and respiration in plant leaves (Zhang et al., 2019). Leaf sheath moisture (LSM) affects sugarcane growth and reflects the balance between the plant nutrients (Keshavaiah et al., 2013). Several studies were conducted to investigate or retrieve sugarcane LAI (Abebe et al., 2022; Lin et al., 2009; Yang et al., 2017), chlorophyll content (Oliveros et al., 2021), nitrogen (Abdel-Rahman et al., 2010; Abdel-Rahman et al., 2013; Miphokasap et al., 2012; Miphokasap and Wannasiri, 2018; Shendryk et al., 2020), LSM (Keshavaiah et al., 2013), and salinity stress (Hamzeh et al., 2013; Hamzeh et al., 2016). Som-Ard et al. (2021) reviewed remote sensing applications in sugarcane cultivation.

Remote sensing has provided an unprecedented opportunity to acquire information about vegetation traits at both local and global scales (Sellers et al., 1997; Weiss et al., 2020). Based on remotely sensed surface radiance, vegetation variables can be retrieved by either statistical-empirical methods or physically-based approaches (Baret and Buis, 2008). Although physically-based approaches have the advantages of allowing more system insight and easily transferring to different data acquisition conditions and crop types, they are

difficult to use and inherently ill-posed for inversion (Baret and Buis, 2008; Combal et al., 2003). Additionally, they are only capable to retrieve variables considered as the inputs into the Radiative Transfer Models (RTM) that have to be inverted during retrieval process (Verrelst et al., 2015a). On the other hand, statistical-empirical methods are easy to use; once calibrated, their implementation is instantaneous; and they are considered as benchmark models to evaluate the performance of physically-based RTMs (Kimes et al., 1998).

Artificial Neural Networks (ANNs) are among the non-linear non-parametric regression models frequently applied to retrieve surface biophysical/biochemical properties such as biomass (see review in (Ali et al., 2015)), fractional vegetation cover (fCover) (Bacour et al., 2006), LAI (see review in (Fang et al., 2019)), chlorophyll content (Wang et al., 2022), nitrogen content (see review in (Berger et al., 2020b)), and water content (Neinavaz et al., 2017; Mirzaie et al., 2014; Trombetti et al., 2008). Verrelst et al. (2019) reviewed successful studies applying ANNs to retrieve different vegetation properties. ANNs, however, suffer from the overfitting problem (Kimes et al., 1998). Overfitting causes a network cannot generalize well to unseen data outside the training set (Beale et al., 2010). Regularization, however, is the practice done to tackle overfitting and improve the generalizability of ANNs. Several regularization methods have been introduced, the most well-known of which are early stopping (Yao et al., 2007) and Bayesian regularization (MacKay, 1992). Bayesian regularization offers more flexible generalization performance than early stopping, especially when dataset is small (Beale et al., 2010). By performing Bayesian regularization, almost all the disadvantages of ANNs can be mitigated while their benefits are preserved (Burden and Winkler, 2008).

Based on Bayes' rule, the Bayesian procedure incorporates *a priori* information about the solution (Qu et al., 2008). Using the *a priori* information can restrict the variable space to a smaller subspace (Combal et al., 2003), resulting in more numerical stability of the model inversion. Bayesian procedures have already been used in the field of vegetation variable retrieval for regulating physically-based RTM inversion methods (Laurent et al., 2014; Xu et al., 2019), assimilating data (Lewis et al., 2012), and estimating uncertainty for the retrievals (Shiklomanov et al., 2016). Bayesian approach has also been utilized to regulate ANNs; the Bayesian regularized ANN (BRANN) has been applied in applications such as environmental studies (Liu et al., 2022; Ye et al., 2021), economy (Sariev and Germano, 2020), and social studies (Kayri, 2016). In the field of vegetation studies, BRANN was applied to map sub-pixel land cover distribution with application in estimating patterns of deforestation and recovery (Braswell et al., 2003), model water status of grapevine (Pôças et al., 2017), and predict cotton yield (Xu et al., 2021). The studies attributed the good performance of BRANN to its ability to generate a more robust network, resulting in lower overfitting.

Simultaneous multi-variable retrieval using statistical-empirical methods is accomplished using a multi-output regression that can be solved by either training independent models or training a single model which directly generates multiple outputs. The former is a shortcut and easily performed, but does not take into account the internal relationships between the target variables. The single model approach, however, considers cross-relationships between multiple target variables (Zhu and Gao, 2018; Verrelst et al., 2015a). Exploiting

co-varying relationships between multiple dependent output variables, the single model approach can improve the prediction precision (Bacour et al., 2006). Especially in the case of vegetation properties, considering the cross-relations among biophysical/biochemical parameters is of great importance in the retrieval process. For instance, since the LAI and the fCover are indicators of the plant density and thereby correlated, a retrieval model should consider not only the underlying relations between the input spectral bands and the parameters to be predicted but also the internal relationships between the output parameters (Tuia et al., 2011). A prime benefit of the single model approach is to speed up the processing because it is implemented just once in the prediction phase (Verrelst et al., 2015a). Nevertheless, inclusion of unrelated variables to this scheme can make the training more complex and raise the risk of ending in local minima (Verrelst et al., 2015a), which can reduce the model performance. For example, Baret et al. (2007) reported more robust results applying a single-output model than multiple output models. Rivera et al. (2013) stated that simultaneously retrieving using a single inversion strategy was not the best choice for LAI and leaf chlorophyll content retrieval, because of the non-linear correlation between these variables. So depending on their cross-relations across the spectral domain, some variables can be more successfully simultaneously retrieved than others (Mousivand et al., 2014).

With the motivation of simultaneously retrieving vegetation bio-physical/chemical variables and overcoming the overfitting problem of ANNs, this paper presents a multi-output BRANN technique to retrieve LAI, LSM, leaf chlorophyll content (LCC), and leaf nitrogen concentration (LNC) of sugarcane from Sentinel-2 spectra. The predictions are both quantitatively and qualitatively assessed. To compare their performance in estimating the sugarcane variables, a comparison between BRANN and a conventional ANN trained with the Levenberg-Marquardt (LM_ANN) algorithm is conducted. For both BRANN and LM_ANN, the retrieval is achieved both simultaneously (all variables at the same time using a single model) and individually (each variable using its own separate independent model) to compare their results.

2 Materials and methods

2.1 Case study

The study has been conducted in Amir Kabir Sugarcane Agro-Industrial zone, one of the seven units in Khuzestan province of Iran, located from 48° 12′ 19″ E to 48° 21′ 23″ E latitude and 30° 58′ 21″ N to 31° 5′ 37″ N longitude. Fig. 1 shows the location of the study area. The region is morphologically flat and its total area is 14,000 ha, of which about 10,000 ha were under cultivation in 2020. The farms are almost homogenous and most of them are 25 ha (1000 (m) × 250 (m)). The area is climatologically semi-arid with about 266 mm annual precipitation and 2788 mm/yr annual evaporation from open pans.

2.2 Data in use

2.2.1 Ground measurements—The ground measurements of the target variables were performed during a field campaign carried out in eight dates from 15th May to 23rd August 2020, concurrently to Sentinel-2A image acquisition. The dates were chosen according to the distinct stages of sugarcane phenology during its growing season to characterize

adequately the global variations of the variables. The sampling strategy was based on measuring the target variables on several samples within an elementary sampling unit (ESU) with an area of 3 (m) × 1.83 (m), and averaging their values. A total of 136 ESUs were taken in the field campaign (see Fig. 1(c)). The ESUs were distributed in fields with 5 different sugarcane varieties (see Fig. 1(d)), and 7 different ratoons (see Fig. 1(e)). These are good representative of all fields throughout the study area.

LAI was derived through a destructive manner so that, in each ESU, three plant samples were harvested and scanned to determine their one-sided leaf area. The total area of leaves in the ESU was then calculated by multiplying the leaf area of samples by the number of plants in the ESU which was counted during the fieldwork, and dividing the result by 3 (the number of samples). LAI (m^2/m^2) was finally calculated by dividing the total area of leaves by the ESU area, i.e. 5.49 (m^2).

To measure the LSM, the leaf sheaths were disconnected and their fresh weigh (FW) was immediately measured. Next, the leaf sheaths were dried in an oven at 80 °C for 24 h and their dry weight (DW) was measured (Fig. 2(a and b)). The percentage of LSM in wet basis (% *wb*) was then obtained using Eq. (1) as:

$$LSM(\% \, wb) = \frac{FW - DW}{FW} \times 100 \, . \tag{1}$$

In order to measure LCC a Minolta SPAD-502 was used, *in situ*, to take chlorophyll from 10 samples per each ESU. For each sample the SPAD reading was repeated 3 times and the average of these 30 values was used. The sampling was not performed on the veins. A widely used exponential equation (Eq. (2)) proposed by Markwell et al. (1995) was applied to convert the unit-less values of the SPAD readings (*SPAD* in Eq. (2)) into chlorophyll concentration (*Chl* in Eq. (2)), as:

$$Chl(\mu mol/m^2) = 10^{SPAD^{0.265}}, \, \left(with \, r^2 = 0.94\right). \tag{2}$$

Finally, the unit of the leaf chlorophyll concentration was converted to $\mu g/cm^2$ regarding the molar mass of chlorophyll, 893.51 (*g/mol*), and the new value was considered as LCC ($\mu g/cm^2$) of the ESU.

After drying and grinding the samples, the LNC (%) was determined through the titration method using the Kjeldahl device by calculating the consumed acid in the laboratory (Fig. 2(c and d)). The widest part of the leaf lamina of leaves 3 to 6 from the top of the straw was considered, and the midrib and veins were discarded since their presence reduces nitrogen concentration (Miphokasap et al., 2012).

Fig. 3 presents some basic statistics of the measurements, and depicts their histogram and density plot.

2.2.2 Satellite data—Sentinel-2 Level-2A images acquired concurrent with the ground measurements were used to retrieve the target variables. Sentinel-2A, launched on 23rd June 2015, is equipped with multispectral optical sensors capable of acquiring 13 spectral bands in the range 400–2500 nm with spatial resolutions of 10 (m), 20 (m), and 60 (m). The 60-meter spatial resolution bands were discarded because of their low resolution, and the remaining ten spectral bands (bands 2–8, 8a and 11–12) were only considered. To match the spatial resolution, the 10-meter spatial resolution images were down-sampled into 20 (m) using nearest neighbor interpolation.

Besides the Sentinel-2A images, Sentinel-2B images were used to map the variables on a pixel-by-pixel full scene basis.

Preparing the Sentinel-2 images was done in the Sentinel Application Platform (SNAP).

2.3 Methodology

2.3.1 Bayesian regularized artificial neural network—BRANN incorporates the Bayes' theorem into a regularization scheme to deal with the overfitting problem of ANNs and improve their generalizability. While conventional training aims to reduce only the sum squared error (E_D) as performance function, a regularized method also considers the model weights into a weight attenuation term (E_W) which penalizes the large weights. The regularized objective function becomes a linear combination of E_D and E_W as:

$$F = \beta E_D + \alpha E_W, \tag{3}$$

where E_W is the sum of squares of network weights. α and β are regularization hyper-parameters. The ratio α/β controls the trade-off between goodness-of-fit and model complexity. The larger the ratio, the more emphasis on weight decay, resulting in a smoother network response. If the ratio becomes smaller, the training algorithm drives the errors smaller (Dan Foresee and Hagan, 1997). Finding the optimum values for the regularization hyper-parameters is therefore the main problem with implementing regularization. MacKay (1992) proposed a probability-based iterative manner to automatically optimize the hyper-parameters. This manner starts with a broad prior distribution for the model parameters, before the data are seen. After the data are taken, our knowledge is updated by calculating a posterior distribution, which is narrower than the prior distribution, using Bayes' rule (*Posterior = Likelihooh × Prior/Evidance*). The goal is choosing the weights that maximize the posterior distribution. In each iteration, the posterior distribution is updated according to the Bayes' rule while large weights are penalized. The steps to determine the optimum regularization hyper-parameters by BRANN are summarized as follows:

i. Set an initial value of α, β and weights. The values are used, after the first training step, to recover the regularization hyper-parameters.

ii. Take one step of the Levenberg-Marquardt algorithm to minimize the objective function Eq. (3).

iii. Compute the effective number of parameters, γ, using the Gauss-Newton approximation to the Hessian matrix (H) in the Levenberg-Marquardt training algorithm as:

$$\gamma = m - \alpha Trace H^{-1}, \tag{4}$$

where m is the number of network weights. γ expresses how many network parameters are effectively used in reducing the objective function (Dan Foresee and Hagan, 1997).

iv. Compute new estimates for α and β using Eq. (5) and Eq. (6), respectively.

$$\alpha = \frac{\gamma}{2E_w(w)}. \tag{5}$$

$$\beta = \frac{N - \gamma}{2E_D(w)}, \tag{6}$$

where N is the number of training samples.

v. Iterate steps 2 through 4 until convergence.

2.3.2 BRANN design for simultaneous multi-variable retrieval—A key factor that affects the performance of (BR)ANNs is the network architecture. For this purpose, the optimum number of hidden layers as well as their neurons should be appropriately determined (Bacour et al., 2006). ANNs with a complicated structure may follow the noise in used data resulting in poor generalization. Conversely, a network with a low number of neurons will not be capable to capture nonlinear relationships between inputs and output(s) (Göçken et al., 2016). In this work, to find the best network architecture, one and two-hidden layer networks were examined, and the number of neurons in the hidden layer(s) was optimized by trial-and-error. According to the dimension of the considered input and target variables, the networks have 10 neurons in the input layer based on the 10 selected Sentinel-2A wavebands, and 4 neurons in the output layer in case of simultaneous retrieval of the four variables and 1 neuron in retrieving the variables individually.

Tangent sigmoid was used as the transfer function in the hidden layer because of its ability to capture the inputs-output(s) nonlinear relationships. For the output layer, however, a linear transfer function was utilized since it is not restricted to produce output values in a specified range. A network with this combination of transfer functions can approximate any continuous function well (Beale et al., 2010).

Since the target variables are of a different dynamic range, their values were scaled into the range of [0, 1], in order to prevent the scaling factor problem and enhance the convergence performance (Bacour et al., 2006). Because the input spectra are as reflectance, they are intrinsically of the same scale, [0, 1], so no further normalization was required over them. An inverse process was needed to invert the scaled predicted values of the target variables into their actual dynamic range.

The same procedure was followed for both LM_ANN and BRANN models.

The retrieval process using the BRANN and LM_ANN models was implemented in Matlab R2022a.

2.3.3   Accuracy assessment procedure—To assess the accuracy and precision of the models used, the randomized bootstrapping procedure was conducted by randomly dividing the dataset into 2 subsets, 70 % for model calibration (95 samples out of 136) and the remaining 30 % for independent validation (41 samples). The procedure was repeated 201 times to create the bootstrap replicate datasets. An odd number was chosen so that the median value (of the statistical indicators) is produced by one of the models, participating in bootstrapping, itself. The median of the statistical indicators was considered as a measure of the model performance.

Four statistical indicators, root mean square error (RMSE), mean bias error (MBE), coefficient of determination (R^2), and relative RMSE (*RRMSE = RMSE/Mean of measurments*) were used for model validation. Each of these statistical indicators was calculated for each of the target variables in each iteration. For example, $RMSE_i^{LAI}$ represents the RMSE of the LAI predictions in the ith bootstrap out of the 201 repetitions. For each of the target variables, the median and standard deviation (Std.) of these 201 values of the statistical indicators was considered as a representative of the performance/accuracy and robustness/precision of the models in retrieving the variable, respectively. For example, in the case of LAI, we considered $RMSE_{\text{Median}}^{LAI} = Median\left(RMSE_i^{LAI}, i = 1:201\right)$ as the final value of RMSE representing the accuracy of LAI retrievals and $RMSE_{Std}^{LAI} = Std\left(RMSE_i^{LAI}, i = 1:201\right)$ as a measure of the precision of LAI retrievals.

In order to evaluate the overall performance of the simultaneous retrieval models in predicting all the target variables, a measure that considers the average of the RMSE values of all variables was defined. This measure, here after called averaged RMSE $\left(RMSE_i^{AVG}\right)$, is computed as:

$$RMSE_i^{AVG} = \frac{RMSE_i^{LAI} + RMSE_i^{LSM} + RMSE_i^{LCC} + RMSE_i^{LNC}}{4}, i = 1 : 201 \, . \tag{7}$$

Note that the measure has no physical meaning, and has been used only to simplify the comparison of the overall performance of the retrieval models in retrieving all the target variables altogether.

3   Experimental results and discussion

The proposed Bayesian regularized ANN was implemented on a dataset consisting of Sentinel-2 spectra as the independent variables and ground measurements of the considered sugarcane variables as the target variables. The obtained results were both quantitatively and qualitatively evaluated, as presented in subsections 3–2 and 3–3, respectively. The

sugarcane variable retrieval was done both simultaneously and individually. In order to provide a comparison, besides the proposed BRANN, the conventional LM_ANN model was also applied. The same procedure was followed in the implementation of both models. For both BRANN and LM_ANN, the results of simultaneous and individual retrievals were compared.

To enable comparing the performance of the different models used for the variable retrieval, a statistical hypothesis test was conducted. For each model, there were 201 RMSEs derived from the bootstrap replicates. These RMSEs were considered as a group. So, in each pairwise comparison, we have two RMSE groups which have to be compared. The goal is to estimate the significance of difference between these two groups. For this purpose, first, the Shapiro–Wilk normality test was utilized to examine the normality of the RMSEs of each group. Based on the results of the normality test, if the RMSEs of both groups followed a normal distribution, then the parametric paired sample t-test was applied to give the significance of difference between these two groups. Otherwise, the nonparametric Wilcoxon Signed-Rank test was utilized to give the significance of difference between these two non-normal groups. The results of the statistical comparisons are presented in subsections 3–2-1 and 3–2-2.

According to the comparison result, the best model was implemented on Sentinel-2A and B images to map the target variables. The prediction maps were interpreted to qualitatively evaluate the retrieval quality.

3.1 Optimizing the network architecture of the models used

In this section the results of the trial and error process performed for optimizing the network architecture of the model used are presented. During the trial and error process, each of the used models (LM_ANN and BRANN) was constructed with networks consisting of one and two hidden layers. In the one-layer networks, the number of neurons from one to 15 was examined. In the 2-layer networks, different combinations of odd numbers from one to 15 were considered as the number of neurons for each of the hidden layers (i.e. a total of 64 different network architectures for each of the models used). Fig. 4 compares the performance of the LM_ANN and BRANN models with different architectures in both simultaneous and individual retrievals in terms of their median of averaged RMSEs (i.e. the median of the 201 $RMSE_i^{AVG}$ calculated from the bootstrap replicates, $RMSE_{Median}^{AVG}$). According to the results presented in this figure, in our experiments, it was demonstrated that the networks with two hidden layers did not perform better than those with one hidden layer. Among all the architectures, the best result was obtained from the one-hidden layer network with 4 neurons for the BRANN model, and with 2 neurons for the LM_ANN model. The $RMSE_{Median}^{AVG}$ values of 2.23 (in detail, 0.48 (m^2/m^2), 2.36 (% wb), 5.85 (μg/cm^2), and 0.23 (%) for LAI, LSM, LCC, and LNC, respectively) and 2.43 (in detail, 0.50 (m^2/m^2), 2.66 (% wb), 6.33 (μg/cm^2), and 0.24 (%) for LAI, LSM, LCC, and LNC, respectively) were achieved by the most efficient BRANN and LM_ANN models, respectively (see the bold labels in Fig. 4(a)).

From the charts of Fig. 4, it can be seen that LM_ANN with simpler networks provided less error. As the number of neurons increases, the retrieval error raises, in both individual and simultaneous retrievals. In fact, increasing model parameters without increasing the size of dataset (136 samples in our study) and without regulating the model leads to overfitting, as discussed later in this subsection. Also, generally by LM_ANN with simpler networks (in particular, the one-hidden layer network (Fig. 4(a)) and the two-hidden layer networks with 3 and 5 neurons in the first hidden layer (Fig. 4(b and c))), the simultaneous variable retrieval led to superior results to individual retrievals.

As seen in Fig. 4, the BRANN model outperformed the LM model in almost all considered architectures. BRANN was also less sensitive to network architecture, since its errors have low fluctuation in different architectures. This is in agreement with (Demuth & Beale, 2004) in which it is stated that BRANN can reduce the difficulty of determining the optimum network architecture.

Since high training accuracy but low testing accuracy is considered as the evidence of overfitting (Skidmore et al., 1997), to investigate the overfitting problem in the used models, their testing and training accuracies were analyzed in different architectures. Figs. 5 and 6 compare the testing and training accuracies for the LM_ANN and BRANN models with different architectures, respectively.

As seen in Fig. 5, as the complexity of the network increases with increase of the number of neurons/layers, the error in the training data decreases, but the measured error in the independent test data increases, which it is a sign of overfitting of the LM_ANN when a complex network is used. This effect can be seen in both individual and simultaneous retrievals, but it is more observable in the individual retrievals. In Fig. 5 (a), it can be seen that in the number of neurons more than 8, the model was perfectly fitted to the training data and the training error is almost zero, while the testing error is high. In the case of simultaneous retrievals, the difference between testing and training errors was smaller than that of individual retrievals. This shows that the simultaneous retrieval using the LM_ANN has alleviated the overfitting problem in it. This is consistent with the findings of Atzberger (2004), which used a simultaneous retrieval to reduce overfitting.

As it can be observed in Fig. 6, by applying BRANN, the difference between testing and training errors is generally small, in both simultaneous and individual retrievals. This indicates the ability of BRANN to overcome the overfitting problem and its more generalizability than LM_ANN. By applying BRANN, even in more complex models, although the performance of the model was reduced, there was a good balance between testing and training errors. This is due to the fact that the BRANN model, by penalizing large network weights, uses only an effective number of network parameters instead of all available parameters.

3.2 Quantitative assessment

This section involves to quantitatively evaluate the performance of the models applied. For this purpose, an accuracy assessment was performed based on the comparison between the model results and the ground measurements in terms of RMSE, RRMSE, R^2, and MBE as

statistical indicators. Table 1 shows the results of the quantitative assessment of the BRANN model by presenting a summary statistics of the four statistical indicators including their mean, median, and Standard deviation. The statistics were calculated based on the 201 bootstrap replicates. As seen in Table 1, low values of MBE have been achieved for all variables, indicating that, generally, their retrievals were on average neither underestimated nor overestimated. It can be explained by the fact that ANN is trained globally to provide unbiased predictions of variables of interest (Bacour et al., 2006). Nevertheless, some underestimations in high values and somewhat overestimations in low values of the variables are observable in the scatterplots of Fig. 7. Achieving more accurate retrievals in medium values of the considered variables than those in low and high values was expected since the low and high values of the variables and their corresponding spectral properties were less represented in the training data than the medium ones (see the frequency distribution of the measured variables in Fig. 3). In the case of LAI, some underestimations for LAI higher than 4 applying ANN were also reported in several previous studies (Bacour et al., 2006; Verrelst et al., 2015b; Xie et al., 2021). The LAI underestimations can be explained by saturation of radiometric signals in dense vegetation (typically for LAI > 5), which causes that small variations in the spectral reflectance cannot be correctly related to the actual canopy LAI (Bacour et al., 2006). In these conditions, ANN overestimates low LAIs to compensate for the underestimation in high LAI values to give predictions without bias, as achieved in our experiments.

Fig. 7 shows the scatterplots of measured values of the sugarcane variables versus their corresponding BRANN estimations.

3.2.1 BRANN vs LM_ANN comparison

—Comparison between BRANN and LM_ANN results indicated that, although there is accordance between their results, BRANN retrieved all target variables more successfully than LM_ANN in both simultaneous and individual retrievals. Fig. 8 compares the accuracy of BRANN and LM_ANN predictions. As seen in Fig. 8, compared to LM_ANN, BRANN reduced $RMSE_{Median}^{LAI}$ from 0.67 to 0.51 (m^2/m^2), $RMSE_{Median}^{LSM}$ from 2.69 to 2.38 (% wb), $RMSE_{Median}^{LCC}$ from 6.89 to 6.03 (µg/cm^2), and $RMSE_{Median}^{LNC}$ from 0.296 to 0.247 (%) applying individual retrieval. In the case of simultaneous retrieval, also, applying BRANN instead of LM_ANN resulted in reduction of $RMSE_{Median}^{LAI}$ from 0.50 to 0.48 (m^2/m^2), $RMSE_{Median}^{LSM}$ from 2.66 to 2.37 (% wb), $RMSE_{Median}^{LCC}$ from 6.33 to 5.85 (µg/cm^2), and $RMSE_{Median}^{LNC}$ from 0.242 to 0.232 (%).

The statistical hypothesis test comparing the BRANN and LM_ANN results showed that the median of RMSEs of the BRANN retrievals is statistically significantly lower than that of LM_ANN retrievals, for all sugarcane variables in both simultaneous and individual retrievals. The values of z-statistic of Wilcoxon Signed-Ranks test, and p-value and significance level of the test are given in Table 2.

LM_ANN provided very poor results in a few bootstrap repetitions. In these replicates, the model presented a small error in the training data but did not in the testing data, which is an evidence of overfitting.

Comparing the optimal network architecture of these models explains the superiority of BRANN over LM_ANN. The optimal network of LM_ANN includes 2 neurons in its hidden layer, which is a simpler network than that of BRANN with 4 neurons. This means that this LM_ANN model is less capable to capture nonlinear relationships between inputs and output(s). Thanks to penalizing large weights, BRANN, however, can explore a more complex architecture without overfitting. For more clarity, in our experiments in the simultaneous retrieval, the optimum BRANN network (the network consisting of 10 input, 4 hidden and 4 output neurons) used only 38 effective parameters out of the 64 available network weights and biases. Such a parsimonious BRANN model can more properly capture nonlinearities without overfitting, resulting in higher generality and lower error than LM_ANN that uses all network parameters. The finding is in accordance with some studies comparing BRANN with non-regularized ANNs (e.g. Gianola et al., 2011; Kayri, 2016).

From the precision viewpoint, also, BRANN showed to be more robust than LM_ANN since it provided lower Std. in all statistical indicators. This superiority is especially significant in the individual retrievals.

3.2.2 Simultaneous vs Individual retrieval comparison—Statistical comparison between the results of simultaneous and individual retrievals showed slight improvement in the retrieval of the sugarcane variables when estimated together. The general improvement in all statistical indicators, RMSE, RRMSE, R^2, and MBE, was achieved by both BRANN and LM_ANN. For example, in the case of RMSE, by applying simultaneous retrieval, the median value of RMSEs of LAI, LSM, LCC, and LNC had decreased, respectively, from 0.51 (m^2/m^2), 2.38 (% wb), 6.03 ($\mu g/cm^2$) and 0.247 (%) to 0.48 (m^2/m^2), 2.37 (% wb), 5.85 ($\mu g/cm^2$), and 0.232 (%) by BRANN, and from 0.67 (m^2/m^2), 2.69 (% wb), 6.89 ($\mu g/cm^2$) and 0.296 (%) to 0.50 (m^2/m^2), 2.66 (% wb), 6.33 ($\mu g/cm^2$), and 0.242 (%) by LM_ANN. Fig. 8 also compares the simultaneous and individual retrieval accuracies. The values of t-score of t-test and z-statistic of Wilcoxon Signed-Ranks test, as well as p-value and significance level of these tests comparing simultaneous retrievals over individual ones are given in Table 3.

The general retrieval improvement can be attributed to the fact that in the simultaneous retrievals, in addition to underlying relations between the input spectra and the target variables, the cross-relations between the target variables themselves linking them together are also considered. So the multi-output simultaneous retrieval model can provide a more realistic representation of the variable retrieval problem than the single-output individual retrieval model that discards the correlations between the target variables. This was confirmed in some previous studies which emphasized the superiority of multi-output regression methods over single-output ones, in the field of vegetation variable retrieval (Bacour et al., 2006; Tuia et al., 2011), and other fields (Zhu and Gao, 2018).

To investigate the effect of inter-relationships of the variables on simultaneously retrieving them, their cross-relations were examined by performing a correlation analysis. The highest correlations were found for LSM_LAI, LCC_LNC, and LSM_LNC with correlation coefficient of 0.49, 0.35, and 0.33, respectively. The relatively high positive correlation

between LSM and LAI can somehow be explained by the finding of Peñuelas et al. (1994), in which it was expressed under water stress, plants decrease leaf expansion and leaves shrink, resulting in LAI reduction. The positive correlation between LCC and LNC was expected since nitrogen is essentially involved in constitution of chlorophyll, as already stated in many studies, (e.g. Bacour et al., 2006; Berger et al., 2020a). Regarding the correlation between LSM and LNC, Ding et al. (2018) found out plant nitrogen level can raise up its root water uptake. On the other hand, a strong correlation between water mass flow and soil nitrogen mobility was found by Cramer et al. (2009). They figured out increasing water flow can reach more nitrate to plant root. These close relationships between the nitrogen and water content of plants were reflected in the moderate positive correlation between LSM and LNC in our experiments. Comparing simultaneous retrievals with individual retrievals in each of the 201 bootstrap repetitions turned out that in most of the replicates in which the simultaneous retrieval of a variable was superior to its individual retrieval, its correlated variable was also retrieved simultaneously superior to individually. For example, LCC was retrieved simultaneously in 148 (out of 201) bootstrap replicates more accurate than individually; of these 148 replicates, in 119 ones (about 80 % of them), the LNC simultaneous retrieval was superior to individual retrieval, as well. It can be attributed to the inherent correlation between LCC and LNC. The number of bootstrap replicates in which LSM simultaneous retrieval was more accurately than its individual retrievals is 115; in 100, out of these 115, replicates (almost 87 % of them), the simultaneously retrieving model could also predict LNC more successfully than the individually-one. This indicates that retrieving two correlated variables simultaneously can lead to superior results to retrieving each of them individually, as already confirmed in some studies (e.g. Tuia et al., 2011).

From the precision viewpoint, by applying BRANN, the simultaneous retrievals were a little more stable than individual ones. As seen in Table1, the BRANN simultaneous retrievals provided $RMSE_{\text{Std.}}^{LAI}$ and $RMSE_{\text{Std.}}^{LSM}$ less than, and $RMSE_{\text{Std.}}^{LCC}$ and $RMSE_{\text{Std.}}^{LNC}$ equal to the BRANN individual ones. By applying LM_ANN, it was turned out that the overfitting problem of LM_ANN occurred much less in simultaneous retrievals than in individual ones. Hence, the Std. of RMSEs calculated from the LM_ANN simultaneous retrievals was significantly less than that calculated from the LM_ANN individual retrievals, indicating that the LM_ANN simultaneous retrieval model is more robust that the individual one. This emphasizes the advantage of simultaneous retrieval, especially when standard non-regularized ANNs is used.

An interesting advantage of the proposed simultaneous retrieval is a remarkable reduction in computational time. The total time elapsed to retrieve all four variables was reduced from 1766 to 403 (s), and from 1055 to 369 (s) performing simultaneous retrieval using BRANN and LM_ANN, respectively. This can be attributed to the fact that in simultaneous retrieval of the four variables, a single model should be calibrated instead of four separate models each of which has to be calibrated separately, as already stated in some previous studies (Tuia et al., 2011; Verrelst et al., 2015b). Tuia et al. (2011) reported that computational load was almost divided by the number of output variables, which was somehow confirmed in our study.

The computational times given here are based on using a core i5-4460 3.20 GHz personal computer with 8 GB installed memory (RAM).

3.2.3 Comparing the retrievability of the target variables by the BRANN simultaneous retrieval model

—In this subsection we discuss how well each of the considered sugarcane variables is estimated by the BRANN simultaneous retrieval model (as the most efficient model used in our experiments). For this purpose, the retrievabilities of the variables are compared in terms RRMSE as the most widely-used dimensionless index (Richter et al., 2012).

Considering the median of the 201 RRMSEs ($RRMSE_{Median}$), LSM was retrieved excellently providing the lowest $RRMSE_{Median}$ (<3 (%)). LNC and LCC with $RRMSE_{Median}$ values of 15.46 (%) and 16.40 (%) respectively can be considered as good retrievals. LAI, however, with 37.83 (%) $RRMSE_{Median}$ was estimated less accurately. Important to note here is that LAI retrievals with RMSE<0.5 (m^2/m^2) are considered excellent in the literature (Richter et al., 2012), and our LAI retrieval has $RMSE_{Median}$ of 0.48 (m^2/m^2).

Our investigations revealed that the quality of retrievals of a model is governed by the variability level of the variable to be retrieved; the less the variability, the higher the performance of the results. To be more explicit, when the variability of a variable is low, the model encounters less difficulty in retrieving that variable. The variability of a variable can be expressed by coefficient of variation ($CV = Std./Mean \times 100(\%)$) of the variable measured values. For example, LAI of sugarcane has high rate of changes during its growing season (as discussed in the last paragraph of subsection 3–3); thereby, the variation range of the LAI values is comparatively large relative to its average value (see Fig. 3(a)). LAI has the highest CV, 51.54 (%), among the considered variables, which can explain why it has provided the highest $RRMSE_{Median}$. LSM, however, has relatively low variations compared to its average value (see Fig. 3(b)), which can be explained by the fact that water requirement of sugarcane is supplied sufficiently by performing regular irrigation. Among the considered variables LSM has the lowest CV, 3.84 (%), explaining the lowest $RRMSE_{Median}$ for LSM retrieval. Although LCC and LNC have quite different average values, they have somewhat partly similar CVs, 19.44 (%) and 16.66 (%) respectively (see Fig. 3(c and d)), which can give a reason for their almost similar $RRMSE_{Median}$ values, 16.40 (%) and 15.46 (%) respectively. Generally, the CV values of 51.54, 19.44, 16.66, and 3.84 (%) for measured values of LAI, LCC, LNC, and LSM, respectively, seem to be meaningfully related to $RRMSE_{Median}$ values of about 38, 16, 15, and 3 (%) provided by these variables, respectively.

In a further investigation, we computed the CV of the target variables of the 41 randomly selected testing samples for each of the 201 testing set participating in bootstrapping, and then calculated Pearson's correlation coefficient between the 201 CVs and their corresponding RRMSEs. Interestingly, a strong positive correlation between the CVs and the RRMSEs was found. Fig. 9 depicts the scatter plot of CV versus RRMSE. As observed in this figure, the correlation coefficient was 0.73, 0.37, 0.50, and 0.49 for LAI, LSM, LCC, and LNC, respectively. This shows that, even for a single variable, a set with lower CV value of the variable can provide a lower RRMSE and vice versa.

3.3 Qualitative assessment

In this subsection the performance of the proposed model is qualitatively assessed by visually interpreting the variable maps generated by applying the BRANN simultaneous retrieval model over the entire Sentinel-2 images. The Sentinel-2 images acquired on 25 different dates throughout the sugarcane growing season were used to produce a time series of the considered variables. The generated variable maps are probed to find out how much the predictions in different places and times are reasonable.

Fig. 10(b1-b4) depicts the prediction maps of the four sugarcane variables derived from the 14th July 2020 Sentinel-2 image, as an instance. As seen in the figures, the variable maps closely resemble the spatial pattern of fields. Within-field variations, caused by farm management practices such as irrigation, fertilization, or harvesting, are observed in the variable maps regarding characteristics of the variable. The fallow lands and man-made areas are clearly differentiated. There are some relatively low density vegetated areas, in the easternmost part of the study area, which their pattern in the prediction maps well mimics the spatial pattern seen in the images. The predicted values of each of the variables seem to be reasonable in these vegetated areas. The bare soils covering the northwestern part of the images are marked by close to zero and even negative values, in all variable maps. The western part of the image is contaminated by a thin cloud, and a dense cloud covered the northwestern part (see Fig. 10(a)); these clouds have affected the predictions of the variables in these areas.

Considering the 201 predictions of each variable obtained by the models participating in bootstrap replicates, the associated Std. maps of the variables were generated. The Std. maps can be treated as a measure of uncertainty of the retrievals (Rivera et al., 2013). Higher values of Std. indicate that the model had more difficulty to retrieve a variable, resulting in higher variation in retrievals. Fig. 10(c1-c4) shows the Std. maps of the sugarcane variables on 14th July 2020. As seen in the figure, the sugarcane fields have provided the lowest Std. values. This is due to the fact that the retrieval models have basically calibrated over the measurements carried out only in these fields. Vegetated areas, even low density ones, show to have low Std. values, indicating that the model was more stable in these areas. Higher values of Std. are observed in non-vegetated areas such as fallow lands, bare soils and man-made features. The highest Std. values, however, are seen in the pixels contaminated by thick clouds having the most spectral distance with the vegetation spectrum. Generally, it can be inferred from the Std. maps that retrievals of non-vegetation areas, where the spectral properties were unrepresented in the training data, were low confidant.

To draw the temporal variations of the variables, time series were prepared for each of the sugarcane variables (see Fig. 11), at three spatial scales. The time series were made of the mean of the predicted values over all sugarcane fields in the first spatial scale, the fields of a certain variety in the second one, and an individual sugarcane field in the third one. In the second spatial scale, three sugarcane varieties having the most number of fields in the study area, i.e. CP69-1062, IRC9902, and CP73-21 (see Fig. 1(d)), were considered. For each of these varieties, one field of the ratoon PC and another of the ratoon R4 were selected as the fields to be investigated separately at the third spatial scale. The selected fields are marked in Fig. 10(a) as colored rectangles. As seen in Fig. 11(a), generally, the known LAI evolution

stages, an initial slow increase, a rapid increase, another slow increase, and finally a decline phase (Teruel et al., 1997), were shown in the LAI time series. The relative decrease in LAI on 18th and 28th August is likely related to heat stress. Under hot wind conditions that occur in summer in the study area, sugarcane shrinks its leaves, resulting in reduction in LAI. As observed in Fig. 11(b), like in LAI, the overall trend in LSM time series is upward, but with a lower slope. The generally similar shape of the time series of LAI and LSM reflects the relatively high correlation found in their ground measurements. In other words, as the ground measurements of LAI and LSM are correlated, their prediction values are correlated, as well, resulting in the rather analogous shape of their time series. The LNC time series (Fig. 11(d)) shows an initial slow increment until the middle of July, followed by a rather steady trend. This is consistent with (Wiedenfeld, 1995), which expressed most sugarcane nitrogen uptake occurs during early growth phase up to canopy closure. Comparing Fig. 11(c) with Fig. 11(d) shows that the time series of LNC and LCC follow approximately the same trend, reflecting their relatively high correlation. As seen in Fig. 11(a and b), the highest values of LAI and LSM were predicted for the IRC9902 variety; the laboratory and field researches conducted in the study area have also indicated that this variety has more and larger leaves than the others. The considerable decrease in the predicted values of the variables, observed in mid to late August for the selected fields of all three varieties of the ratoon PC, and after 17th October for the selected field of the variety CP73-21_ratoon R4, is attributed to harvesting these fields. Due to rapid growth of the ratoon PC, its harvesting begins earlier than the other ratoons.

In general, quantitative assessment of the results indicates that the applied BRANN simultaneous retrieval model can provide predictions of the considered variables that reasonably represent the spatial and temporal variations of them. Considering different sugarcane varieties and ratoons shows that the predicted maps can well represent the differences between the phenology of the considered sugarcane varieties and ratoons.

4 Conclusion

This paper reports the outcomes of our efforts for retrieving LAI, LSM, LCC, and LNC of sugarcane from Sentinel-2 data. The multi-output BRANN was used to simultaneously retrieve the variables. The performance of BRANN was compared against that of the most conventional neural network, LM_ANN. To evaluate the performance of the simultaneous retrievals, individually retrieving each of the variables was also performed to compare their results. The statistical test comparing the performance of BRANN and LM_ANN indicated that BRANN outperformed LM_ANN in both simultaneous and individual retrievals. It is due to the capability of the model to penalize large network weights resulting in a more robust model with lower overfitting and higher generality. Statistical comparison between simultaneous and individual retrievals showed a marginal gain in accuracy and precision of the results applying simultaneously retrieval. Also, the simultaneous retrieval significantly reduced the runtime of the retrievals.

Investigating the retrievability of the variables by the BRANN model turned out that the variability level of the variable rules the quality of retrievals. According to this finding, LSM

having the lowest CV has been retrieved more accurately than the other variables; LNC, LCC, and LAI were ranked as the next most accurate retrievals, respectively.

Qualitative assessment of the results of the BRANN simultaneous retrieval model indicated that the retrievals reasonably represent the spatial and temporal variations of the variables. Generally, the variables were retrieved more confidently in vegetated areas than in non-vegetated areas since the model was calibrated exclusively using training data collected from sugarcane fields.

This study confirms the usability of the BRANN simultaneous retrieval model providing more accurate and precise, and much faster retrievals of sugarcane variables from Sentinel-2 images than those of conventional LM_ANN and individual retrievals. Our predictions of the sugarcane variables, especially the LSM and LNC predictions, are being used practically for irrigation and fertilizer management in the Amir Kabir Sugarcane Agro-Industrial zone, in the time of writing this paper.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data availability

Data will be made available on request.

References

Abdel-Rahman EM, Ahmed FB, Van den Berg M. Estimation of sugarcane leaf nitrogen concentration using in situ spectroscopy. Int J Appl Earth Obs Geoinf. 2010; 12: S52–S57.

Abdel-Rahman EM, Ahmed FB, Ismail R. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. Int J Remote Sens. 2013; 34: 712–728.

Abebe G, Tadesse T, Gessesse B. Assimilation of leaf Area Index from multisource earth observation data into the WOFOST model for sugarcane yield estimation. Int J Remote Sens. 2022; 43: 698–720.

Ali I, Greifeneder F, Stamenkovic J, Neumann M, Notarnicola C. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. Remote Sens. 2015; 7: 16398–16421.

Atzberger C. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. Remote Sens Environ. 2004; 93 (1–2) 53–67.

Bacour C, Baret F, Béal D, Weiss M, Pavageau K. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: principles and validation. Remote Sens Environ. 2006; 105: 313–325.

Baret F, Buis S. Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems. Adv L Remote Sens Syst Model Invers Appl. 2008. 173–201.

Baret F, Hagolle O, Geiger B, Bicheron P, Miras B, Huc M, Berthelot B, Nino F, Weiss M, Samain O, Louis J, et al. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION Part 1 : Principles of the algorithm. 2007; 110: 275–286.

Beale, MH, Hagan, MT, Demuth, HB. Neural network toolbox. Vol. 2. Natick: User's Guide, MathWorks, Inc; 2010.

Berger K, Verrelst J, Féret J-B, Hank T, Wocher M, Mauser W, Camps-Valls G. Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. Int J Appl Earth Obs Geoinf. 2020a; 92 102174 [PubMed: 36090128]

Berger K, Verrelst J, Feret J-B, Wang Z, Wocher M, Strathmann M, Danner M, Mauser W, Hank T. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. Remote Sens Environ. 2020b; 242 111758 [PubMed: 36082364]

Braswell BH, Hagen SC, Frolking SE, Salas WA. A multivariable approach for mapping sub-pixel land cover distributions using MISR and MODIS: application in the Brazilian Amazon region. Remote Sens Environ. 2003; 87: 243–256.

Burden F, Winkler D. Bayesian regularization of neural networks. Methods Mol Biol. 2008; 458: 25–44. [PubMed: 19065804]

Combal B, Baret F, Weiss M, Trubuil A, Macé D, Pragnère A, Myneni R, Knyazikhin Y, Wang L. Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. Remote Sens Environ. 2003; 84: 1–15.

Cramer MD, Hawkins H-J, Verboom GA. The importance of nutritional regulation of plant water flux. Oecologia. 2009; 161: 15–24. [PubMed: 19449035]

Dan Foresee, F; Hagan, MT. Gauss-Newton approximation to bayesian learning; IEEE Int Conf Neural Networks – Conf Proc; 1997. 1930–1935.

Delegido J, Alonso L, Gonzalez G, Moreno J. Estimating chlorophyll content of crops from hyperspectral data using a normalized area over reflectance curve (NAOC). Int J Appl Earth Obs Geoinf. 2010; 12: 165–174.

Demuth, H, Beale, M. Neural Network Toolbox for use with MATLAB. User's Guide, MathWorks; 2004. 2004

Ding L, Lu Z, Gao L, Guo S, Shen Q. Is nitrogen a key determinant of water transport and photosynthesis in higher plants upon drought stress? Front Plant Sci. 2018; 9

Fang H, Baret F, Plummer S, Schaepman-Strub G. An overview of global Leaf Area Index (LAI): methods, products, validation, and applications. Rev Geophys. 2019; 57: 739–799.

Féret J-B, Berger K, de Boissieu F, Malenovský Z. PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. Remote Sens Environ. 2021; 252 112173

Gianola D, Okut H, Weigel KA, Rosa GJ. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. BMC Genet. 2011; 12: 87. [PubMed: 21981731]

Göçken M, Ozçalici M, Boru A, Dosdogru AT. Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction. Expert Syst Appl. 2016; 44: 320–331.

Hamzeh S, Naseri AA, Alavipanah SK, Mojaradi B, Bartholomeus HM, Clevers JG, Behzad M. Estimating salinity stress in sugarcane fields with spaceborne hyperspectral vegetation indices. Int J Appl Earth Obs Geoinf. 2013; 21: 282–290.

Hamzeh S, Naseri AA, AlaviPanah SK, Bartholomeus H, Herold M. Assessing the accuracy of hyperspectral and multispectral satellite imagery for categorical and quantitative mapping of salinity stress in sugarcane fields. Int J Appl Earth Obs Geoinf. 2016; 52: 412–421.

Kayri M. Predictive abilities of Bayesian regularization and levenberg-marquardt algorithms in artificial neural networks: A comparative empirical study on social data. Math Comput Appl. 2016; 21: 20.

Keshavaiah KV, Palled YB, Shankaraiah C, Nandihalli BS. Effect of sheath moisture and relation of SPAD on yield of sugarcane. Adv Res J Crop Improv. 2013; 4: 98–102.

Kimes DS, Nelson RF, Manry MT, Fung AK. Review article: Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. Int J Remote Sens. 1998; 19: 2639–2663.

Laurent VCE, Schaepman ME, Verhoef W, Weyermann J, Chávez RO. Remote Sensing of Environment Bayesian object-based estimation of LAI and chlorophyll from a simulated Sentinel-2 top-of-atmosphere radiance image. Remote Sens Environ. 2014; 140: 318–329.

Lewis P, Gómez-Dans J, Kaminski T, Settle J, Quaife T, Gobron N, Styles J, Berger M. An Earth Observation Land Data Assimilation System (EO-LDAS). Remote Sens Environ. 2012; 120: 219–235.

Lin H, Chen J, Pei Z, Zhang S, Hu X. Monitoring sugarcane growth using ENVISAT ASAR data. IEEE Trans Geosci Remote Sens. 2009; 47: 2572–2580.

Liu H, He B, Zhou Y, Kutser T, Toming K, Feng Q, Yang X, Fu C, Yang F, Li W, Peng F. Trophic state assessment of optically diverse lakes using Sentinel-3-derived trophic level index. Int J Appl Earth Obs Geoinf. 2022; 114 103026

MacKay DJC. Bayesian Interpolation. Neural Comput. 1992; 4: 415–447.

Markwell J, Osterman JC, Mitchell JL. Calibration of the Minolta SPAD-502 leaf chlorophyll meter. Photosynth Res. 1995; 46: 467–472. [PubMed: 24301641]

Miphokasap P, Wannasiri W. Estimations of nitrogen concentration in sugarcane using hyperspectral imagery. Sustainability. 2018; 10: 1266.

Miphokasap P, Honda K, Vaiphasa C, Souris M, Nagai M. Estimating canopy nitrogen concentration in sugarcane using field imaging spectroscopy. Remote Sens. 2012; 4: 1651–1670.

Mirzaie M, Darvishzadeh R, Shakiba A, Matkan AA, Atzberger C, Skidmore A. Comparative analysis of different uni-and multi-variate methods for estimation of vegetation water content using hyper-spectral measurements. Int J Appl Earth Obs Geoinf. 2014; 26: 1–11.

Moraes MAFD, Oliveira FCR, Diaz-Chavez RA. Socio-economic impacts of Brazilian sugarcane industry. Environ Dev. 2015; 16: 31–43.

Moran MS, Inoue Y, Barnes EM. Opportunities and limitations for imagebased remote sensing in precision crop management. Remote Sens Environ. 1997; 61: 319–346.

Mousivand A, Menenti M, Gorte B, Verhoef W. Remote Sensing of Environment Global sensitivity analysis of the spectral radiance of a soil – vegetation system. Remote Sens Environ. 2014; 145: 131–144.

Neinavaz E, Skidmore AK, Darvishzadeh R, Groen TA. Retrieving vegetation canopy water content from hyperspectral thermal measurements. Agric For Meteorol. 2017; 247: 365–375.

Oliveros N, Tinini R, dos Costa D, Ramos R, Wetterich C, Teruel B. Predictive models of chlorophyll content in sugarcane seedlings using spectral images. Eng Agric. 2021; 41: 475–484.

Peñuelas J, Gamon JA, Fredeen AL, Merino J, Field CB. Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. Remote Sens Environ. 1994; 48: 135–146.

Pôças I, Gonçalves J, Costa PM, Gonçalves I, Pereira LS, Cunha M. Hyperspectral-based predictive modelling of grapevine water status in the Portuguese Douro wine region. Int J Appl Earth Obs Geoinf. 2017; 58: 177–190.

Qu Y, Wang J, Wan H, Li X, Zhou G. A Bayesian network algorithm for retrieving the characterization of land surface vegetation. Remote Sens Environ. 2008; 112: 613–622.

Richter K, Atzberger C, Hank TB, Mauser W. Derivation of biophysical variables from Earth observation data: validation and statistical measures. J Appl Remote Sens. 2012; 6 063557-63561

Rivera JP, Verrelst J, Leonenko G, Moreno J. Multiple cost functions and regularization options for improved retrieval of leaf chlorophyll content and LAI through inversion of the PROSAIL model. Remote Sens. 2013; 5: 3280–3304.

Sariev E, Germano G. Bayesian regularized artificial neural networks for the estimation of the probability of default. Quant Financ. 2020; 20: 311–328.

Sellers PJ, Dickinson RE, Randall DA, Betts AK, Hall FG, Berry JA, Collatz GJ, Denning AS, Mooney HA, Nobre CA, Sato N, et al. Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. Science 80-). 1997; 275: 502–509.

Shendryk Y, Sofonia J, Garrard R, Rist Y, Skocaj D, Thorburn P. Fine-scale prediction of biomass and leaf nitrogen content in sugarcane using UAV LiDAR and multispectral imaging. Int J Appl Earth Obs Geoinf. 2020; 92 102177

Shiklomanov AN, Dietze MC, Viskari T, Townsend PA, Serbin SP. Quantifying the influences of spectral resolution on uncertainty in leaf trait estimates through a Bayesian approach to RTM inversion. Remote Sens Environ. 2016; 183: 226–238.

Skidmore AK, Turner BJ, Brinkhof W, Knowles E. Performance of a neural network: mapping forests using GIS and remotely sensed data. Photogrammetric Eng Remote Sens. 1997; 63 (5) 501–514.

Sofonia J, Shendryk Y, Phinn S, Roelfsema C, Kendoul F, Skocaj D. Monitoring sugarcane growth response to varying nitrogen application rates: a comparison of UAV SLAM LiDAR and photogrammetry. Int J Appl Earth Obs Geoinf. 2019; 2 101878

Som-ard J, Atzberger C, Izquierdo-Verdiguier E, Vuolo F, Immitzer M. Remote sensing applications in sugarcane cultivation: a review. Remote Sens. 2021; 13 (20) 4040.

Teruel DA, Barbieri V, Ferraro LA Jr. Sugarcane leaf area index modeling under different soil water conditions. Sci Agric. 1997; 54: 39–44.

Trombetti M, Riano D, Rubio MA, Cheng YB, Ustin SL. Multi-temporal vegetation canopy water content retrieval and interpretation using artificial neural networks for the continental USA. Remote Sens Environ. 2008; 112 (1) 203–215.

Tuia D, Verrelst J, Alonso L, Perez-Cruz F, Camps-Valls G. Multioutput support vector regression for remote sensing biophysical parameter estimation. IEEE Geosci Remote Sens Lett. 2011; 8: 804–808.

Verrelst J, Camps-valls G, Munoz-marí J, Pablo J, Veroustraete F, Clevers JGPW, Moreno J. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. ISPRS J Photogramm Remote Sens. 2015a; 108: 273–290.

Verrelst J, Rivera JP, Veroustraete F, Munoz-Marí J, Clevers JGPW, Camps-Valls G, Moreno J. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison. ISPRS J Photogramm Remote Sens. 2015b; 108: 260–272.

Verrelst J, Malenovský Z, Van der Tol C, Camps-Valls G, Gastellu-Etchegorry J-P, Lewis P, North P, Moreno J. Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. Surv Geophys. 2019; 40: 589–629. [PubMed: 36081834]

Wang T, Gao M, Cao C, You J, Zhang X, Shen L. Winter wheat chlorophyll content retrieval based on machine learning using in situ hyperspectral data. Comput Electron Agric. 2022; 193 106728

Weiss M, Jacob F, Duveiller G. Remote sensing for agricultural applications: a meta-review. Remote Sens Environ. 2020; 236 111402

Wiedenfeld RP. Effects of irrigation and N fertilizer application on sugarcane yield and quality. Field Crops Research. 1995; 43 (2) 101–108. DOI: 10.1016/0378-4290(95)00043-P

Xie R, Darvishzadeh R, Skidmore AK, Heurich M, Holzwarth S, Gara TW, Reusen I. Mapping leaf area index in a mixed temperate forest using Fenix airborne hyperspectral data and Gaussian processes regression. Int J Appl Earth Obs Geoinf. 2021; 95 102242

Xu W, Chen P, Zhan Y, Chen S, Zhang L, Lan Y. Cotton yield estimation model based on machine learning using time series UAV remote sensing data. Int J Appl Earth Obs Geoinf. 2021; 104 102511

Xu XQ, Lu JS, Zhang N, Yang TC, He JY, Yao X, Cheng T, Zhu Y, Cao WX, Tian YC. Inversion of rice canopy chlorophyll content and leaf area index based on coupling of radiative transfer and Bayesian network models. ISPRS J Photogramm Remote Sens. 2019; 150: 185–196.

Yang Q, Ye H, Huang K, Zha Y, Shi L. Estimation of leaf area index of sugarcane using crop surface model based on UAV image. Nongye Gongcheng Xuebao/Transactions Chinese Soc Agric Eng. 2017; 33: 104–111.

Yao Y, Rosasco L, Caponnetto A. On early stopping in gradient descent learning. Constr Approx. 2007; 26: 289–315.

Ye L, Jabbar SF, Abdul Zahra MM, Tan ML. Bayesian regularized neural network model development for predicting daily rainfall from sea level pressure data: investigation on solving complex hydrology problem. Complexity. 2021; 2021 6631564

Zhang Z, Tang BH, Li ZL. Retrieval of leaf water content from remotely sensed data using a vegetation index model constructed with shortwave infrared reflectances. Int J Remote Sens. 2019; 40: 2313–2323.

Zhu X, Gao Z. An efficient gradient-based model selection algorithm for multioutput least-squares support vector regression machines. Pattern Recognit Lett. 2018; 111: 16–22.
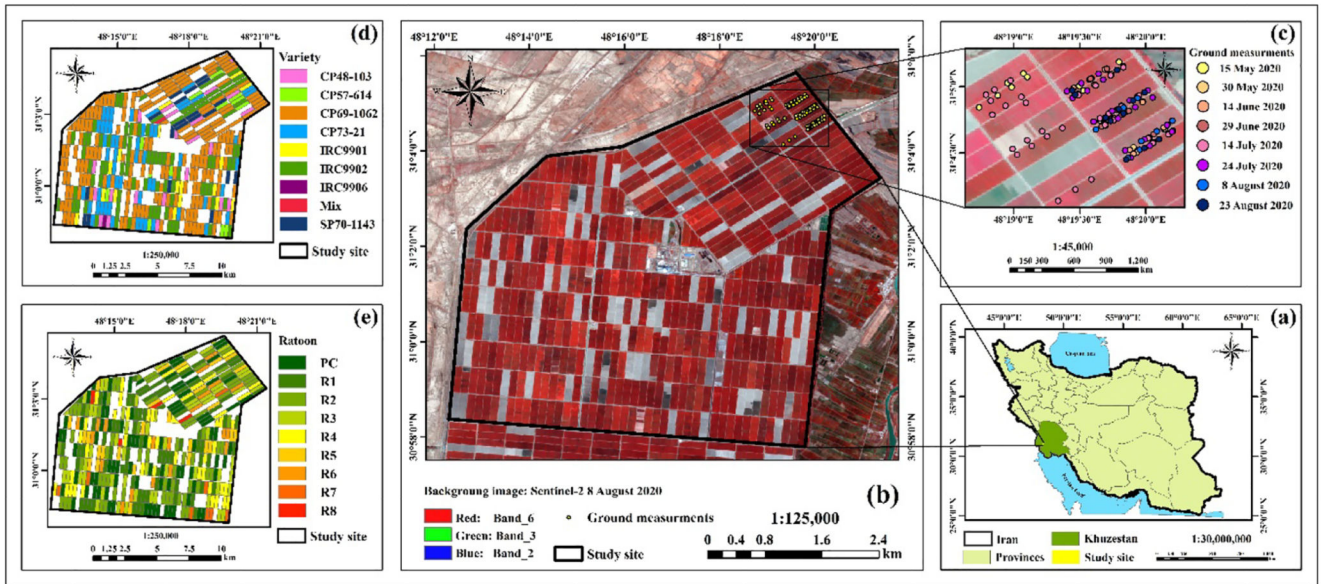
Fig. 1. Study area.

(a) The location of the case study; (b) The boundary of Amir Kabir Sugarcane Agro-Industrial zone superimposed on Sentinel-2 image; (c) Distribution of ground measurement samples; (d) Map of sugarcane varieties; and (e) Map of sugarcane ratoons. Plant Cultivation (PC) represents newly cultivated sugarcanes. R1-R8 represent sugarcanes of ratoon1-8.

Fig. 2.
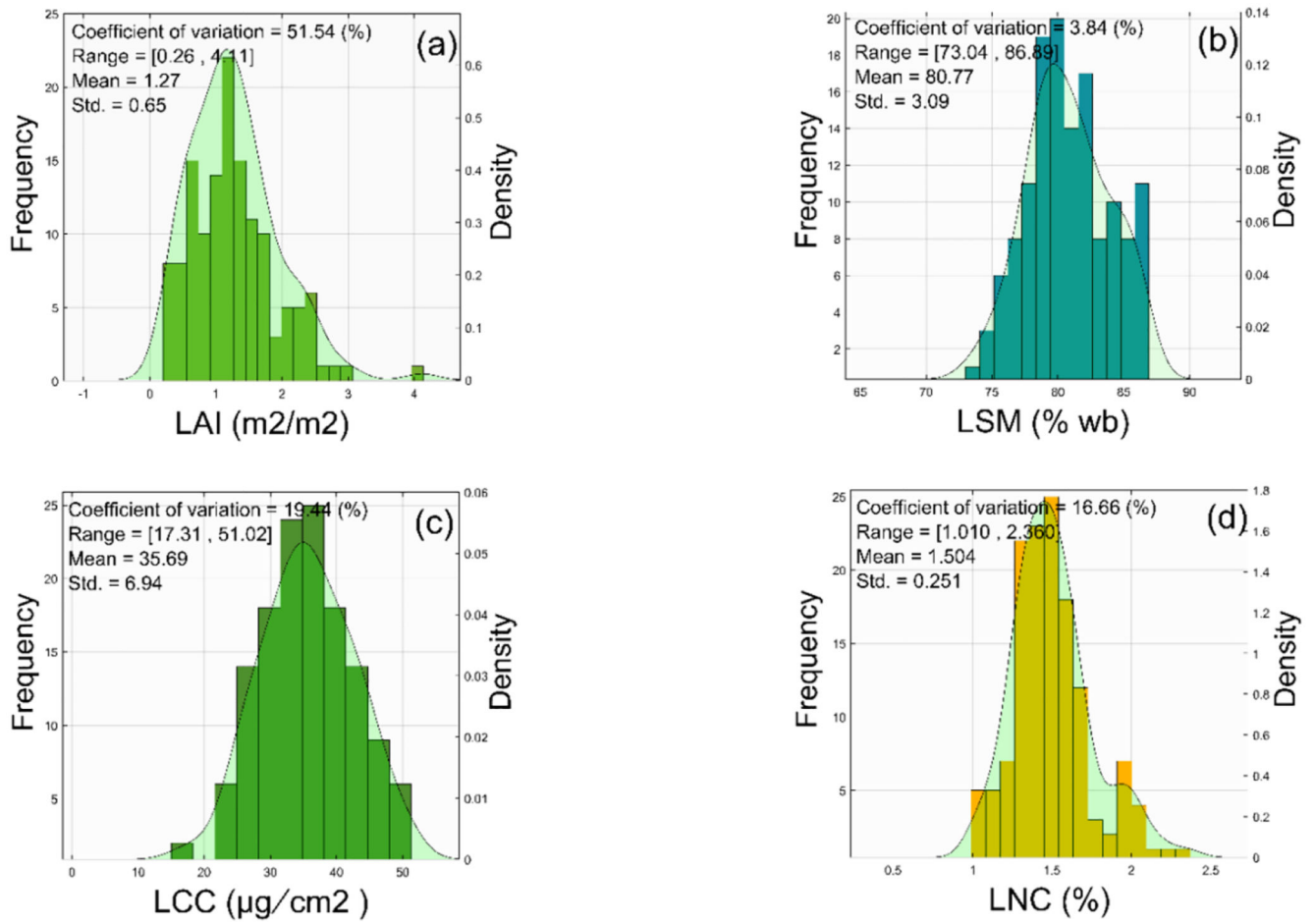Laboratory proceedings for measuring (a and b) LSM; and (c and d) LNC.

Fig. 3. Histogram and density plot of the measured variables
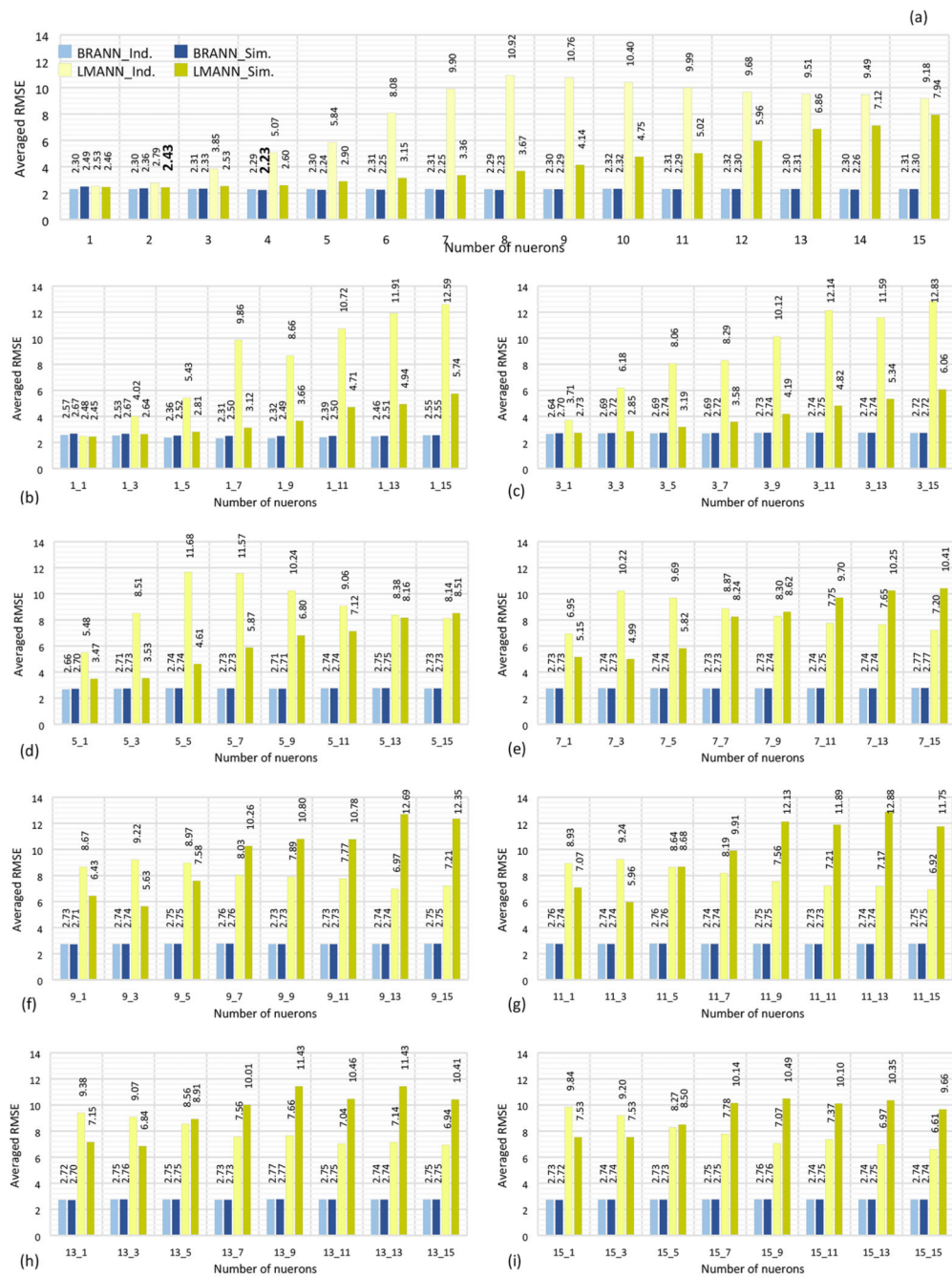(a) LAI; (b) LSM; (c) LCC; and (d) LNC.

Fig. 4.
Comparing the performance of BRANN and LM_ANN in terms of averaged RMSE calculated for testing dataset for networks with (a) One-hidden layer; and (bi) two-hidden layer. The X-axis represents the number of neurons in hidden layers as: the number of neurons in the first hidden layer_ the number of neurons in the second hidden layer. The averaged RMSEs of the optimum network for BRANN and LM_ANN are presented as bold label.
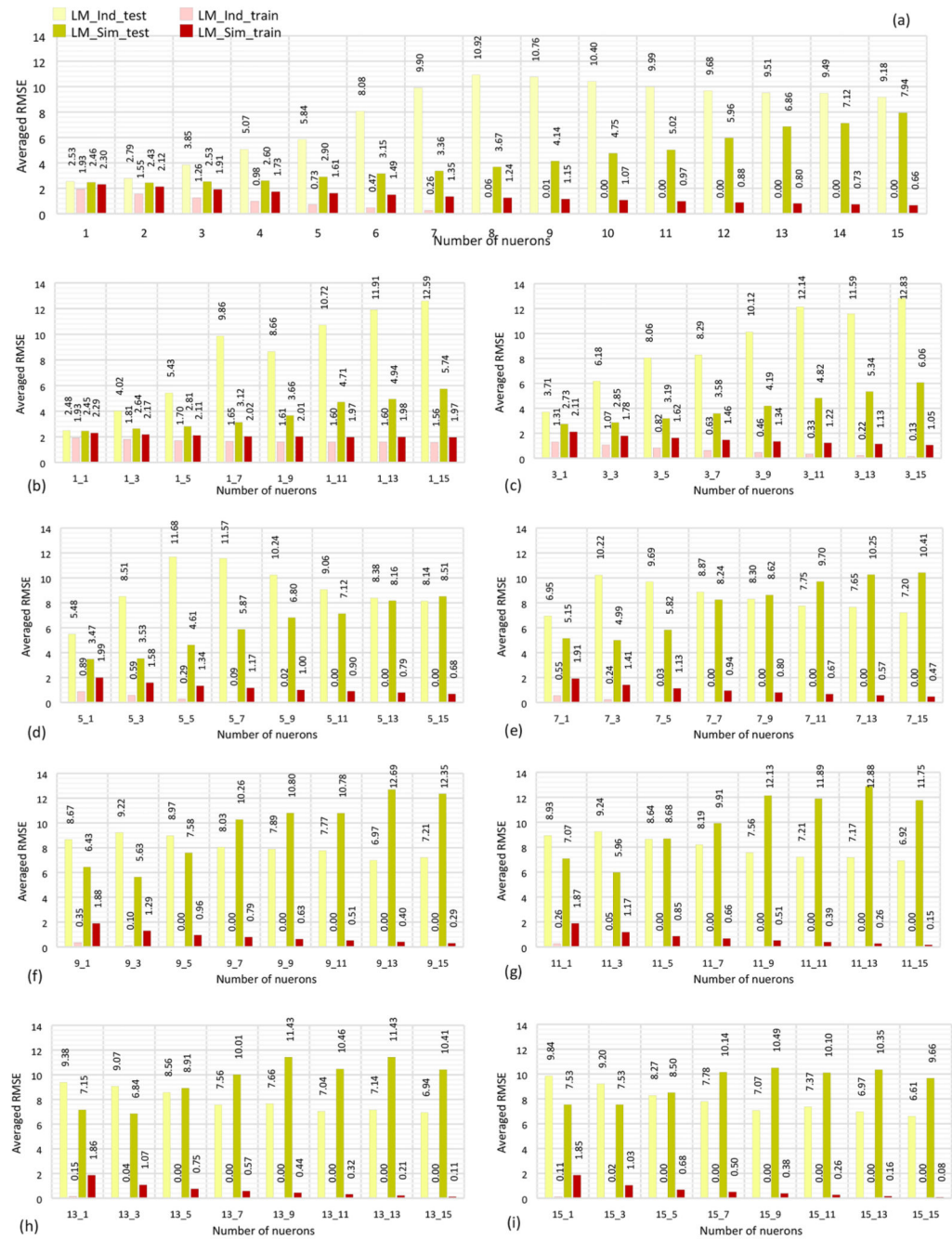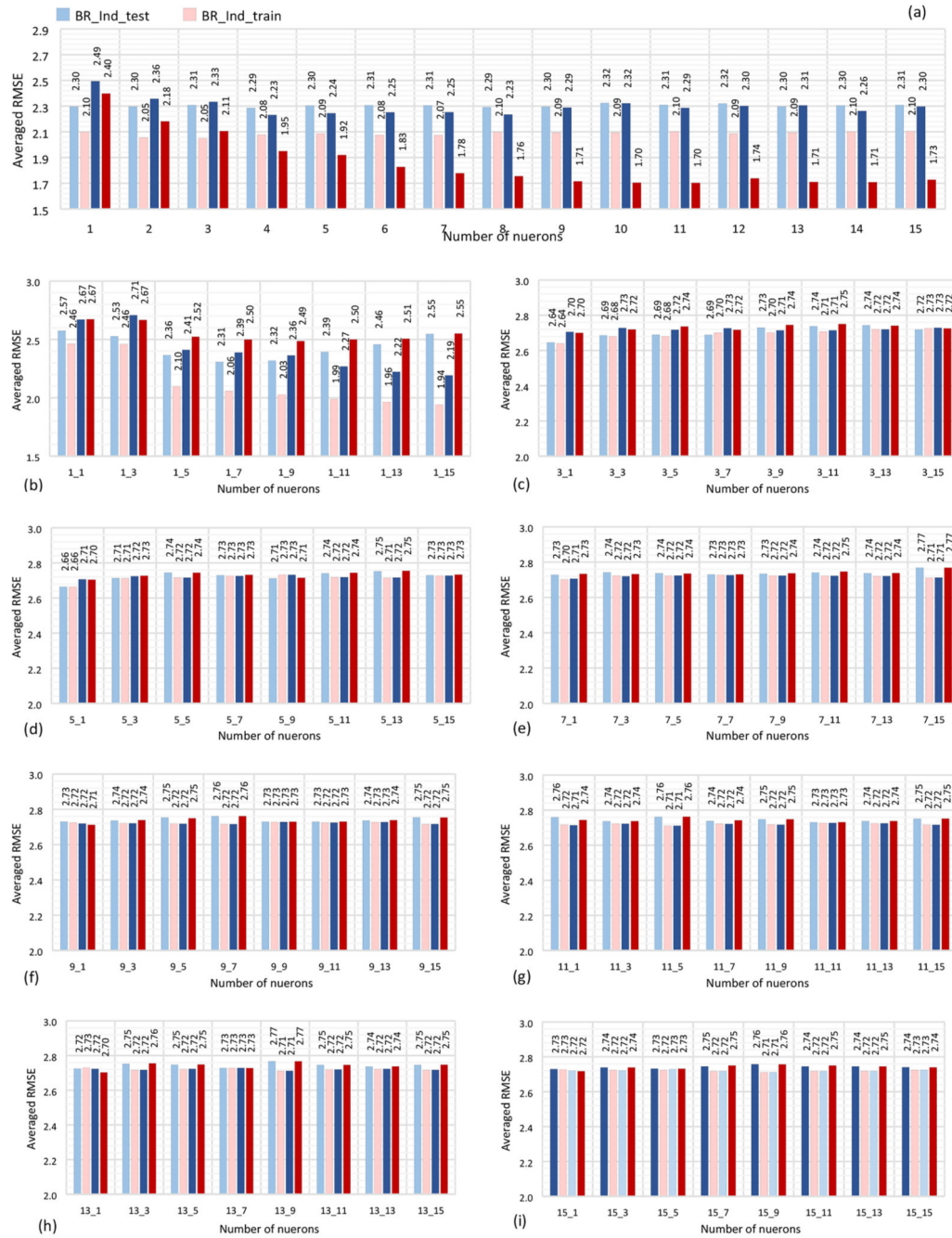
Fig. 5.
Comparing the testing and training errors of LM_ANN in terms of averaged RMSE for networks with (a) One-hidden layer; and (b-i) two-hidden layer. The X-axis represents the number of neurons in hidden layers as: the number of neurons in the first hidden layer_ the number of neurons in the second hidden layer.

Fig. 6.

Comparing the testing and training errors of BRANN in terms of averaged RMSE for networks with (a) One-hidden layer; and (b-i) two-hidden layer. The X-axis represents the number of neurons in hidden layers as: the number of neurons in the first hidden layer_ the number of neurons in the second hidden layer.
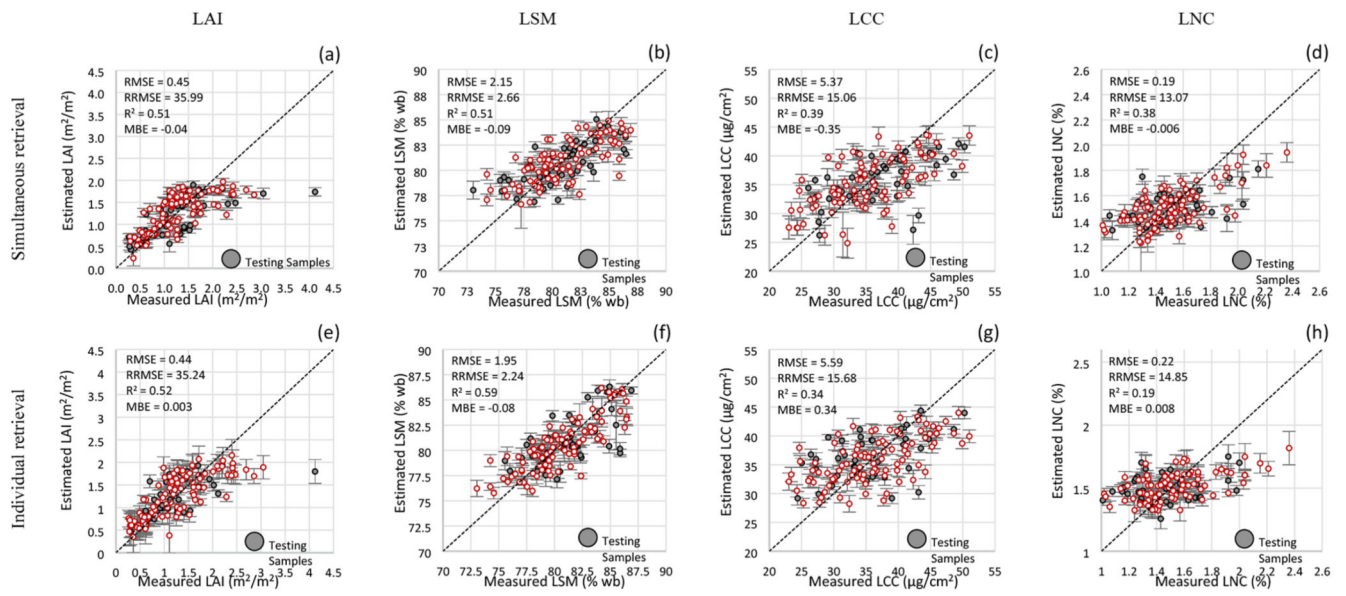
Fig. 7.
Scatterplot of the measurements versus the BRANN estimations of (a) LAI; (b) LSM; (c) LCC; and (d) LNC, for the simultaneous retrieval, and (e) LAI; (f) LSM; (g) LCC; and (h) LNC, for the individual retrieval. Vertical error bar represents the standard deviation of the 201 estimations. RMSE, RRMSE, R^2, and MBE were calculated considering all samples.
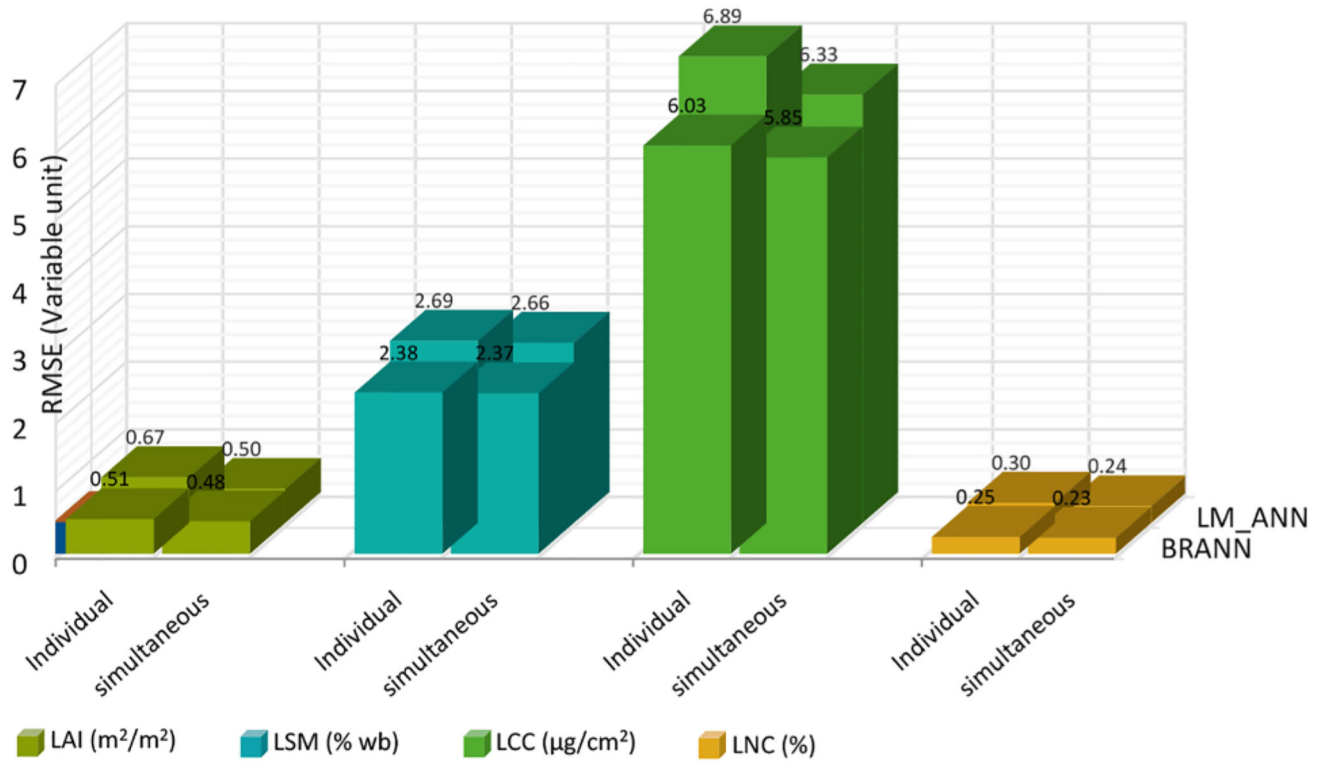
Fig. 8. RMSE of retrievals of BRANN and LM_ANN, applying individual and simultaneous retrievals.
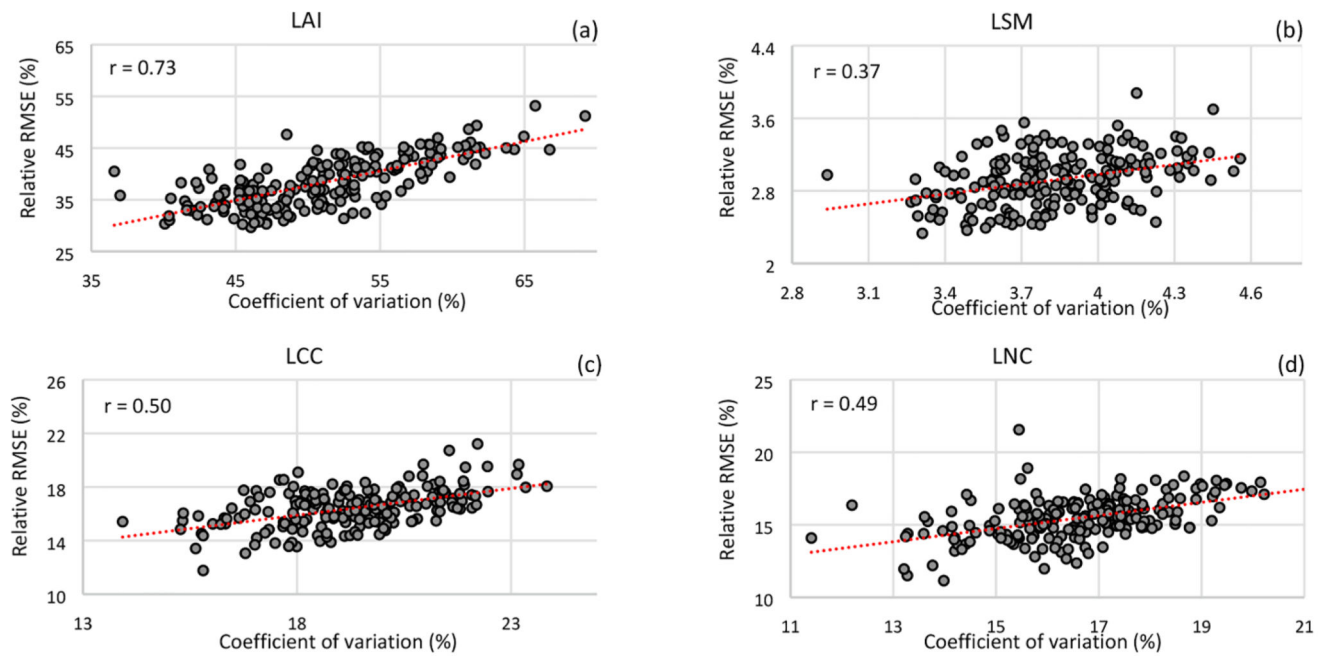
Fig. 9.

Scatter plot of CV versus RRMSE for (a) LAI; (b) LSM; (c) LCC; and (d) LNC. The dotted red line represents the linear trend line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
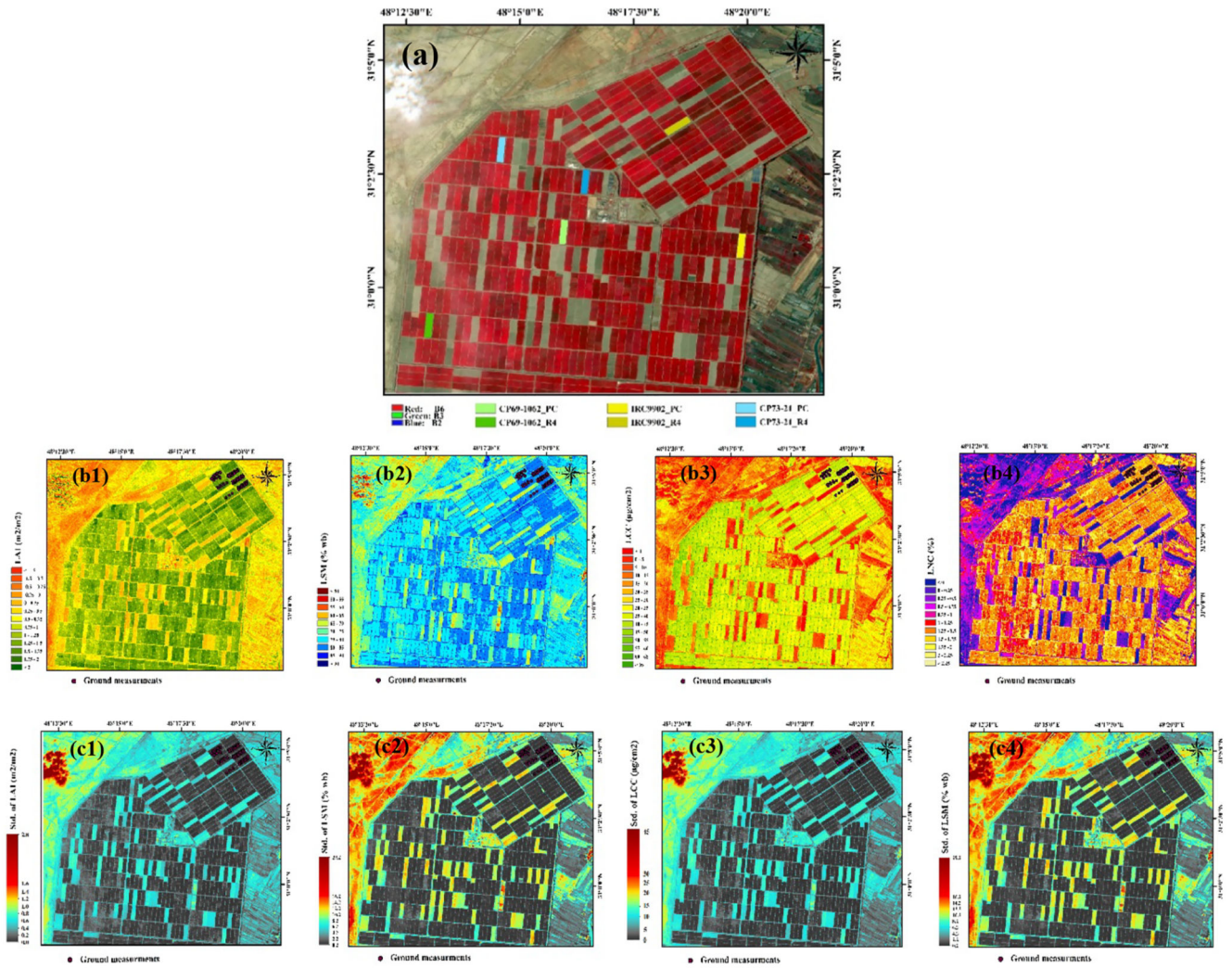
Fig. 10.
(a) Sentinel-2 image of 14th July 2020. The selected fields for the time series investigation at the third spatial scale are marked as colored rectangles. The map of (b1) LAI, (b2) LSM, (b3) LCC, and (b4) LNC predicted from the image; The Std. map of (c1) LAI, (c2) LSM, (c3) LCC, and (c4) LNC of the same date.
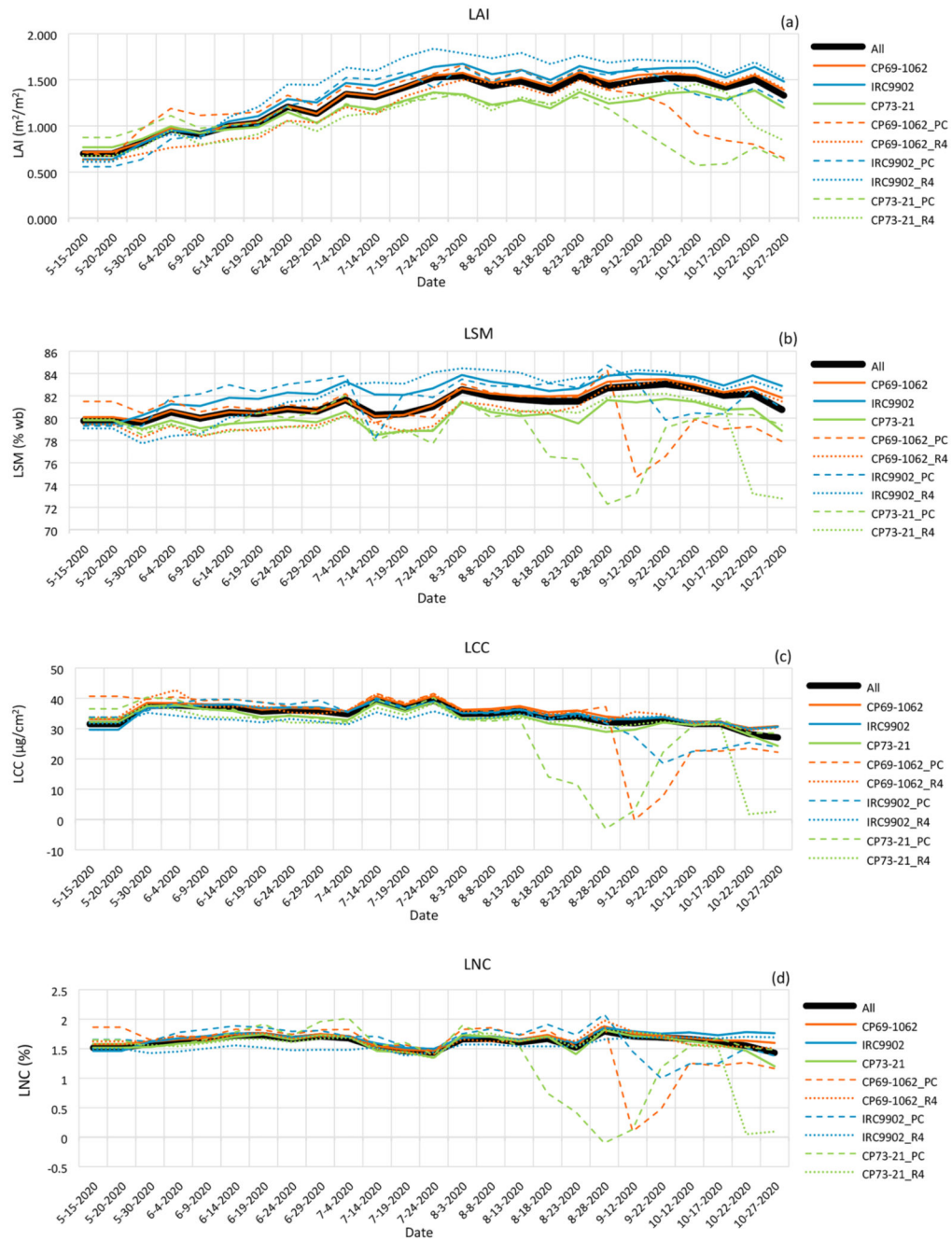
Fig. 11.
Time series of the predictions of (a) LAI; (b) LSM; (c) LCC; and (d) LNC. The time series made of the mean of the predicted values over: all sugarcane fields is shown as a bold black line; the fields of each considered variety is drawn as thin colored lines; and the selected field of the ratoons PC and R4 is presented as dashed and dotted colored lines, respectively.

Table 1

Summary statistic of the statistical indicators obtained from BRANN results. The statistics include mean, median, and standard deviation (Std.) calculated based on the 201 bootstrap replicates. Retrieving time (in second) is given in the last two rows of the table.

		Individual retrieval				Simultaneous retrieval			
		LAI (m²/m²)	LSM (% wb)	LCC (µg/cm²)	LNC (%)	LAI (m²/m²)	LSM (% wb)	LCC (µg/cm²)	LNC (%)
RMSE (Variable unit)	Mean	0.52	2.40	6.00	0.247	0.49	2.36	5.85	0.232
	Median	0.51	2.38	6.03	0.247	0.48	2.37	5.85	0.232
	Std.	0.09	0.26	0.49	0.022	0.07	0.23	0.49	0.022
RRMSE (%)	Mean	40.48	2.96	16.81	16.41	38.32	2.92	16.40	15.46
	Median	40.21	2.95	16.73	16.45	37.83	2.93	16.41	15.46
	Std.	5.92	0.33	1.48	1.41	4.78	0.28	1.43	1.47
R^2 (-)	Mean	0.34	0.37	0.21	−0.01	0.41	0.40	0.24	0.09
	Median	0.39	0.42	0.24	0.03	0.42	0.42	0.27	0.12
	Std.	0.19	0.20	0.13	0.13	0.12	0.12	0.15	0.20
MBE (Variable unit)	Mean	−0.01	−0.04	−0.06	0.00	0.00	−0.06	−0.05	−0.01
	Median	0.00	−0.03	0.03	0.00	0.00	−0.05	0.05	−0.01
	Std.	0.10	0.45	1.11	0.05	0.10	0.43	1.08	0.04
Retrieving time(Second)		416	481	518	351	–	–	–	–
Total time(Second)		1766				403			

Table 2

Statistical hypothesis test for comparing BRANN vs LM_ANN.

	BRANN vs LM_ANN	z-statistics	p-value	Median of RMSE (BRANN: LM_ANN)
Individual retrieval	LAI	−10.48	1.05e-25	(0.51: 0.67) ****
	LSM	− 8.64	5.47e-18	(2.38: 2.69) ****
	LCC	−10.19	2.02e-24	(6.03: 6.89) ****
	LNC	− 9.62	6.18e-22	(0.25: 0.30) ****
Simultaneous retrieval	LAI	− 2.39	0.017	(0.48: 0.50) *
	LSM	−10.96	5.57e-28	(2.37: 2.66) ****
	LCC	− 7.49	6.51e-14	(5.85: 6.33) ****
	LNC	− 2.66	0.008	(0.23: 0.24) *

*
significant at p < 0.05

**
significant at p < 0.005

***
significant at p < 0.001

****
significant at p < 0.0001.

Table 3

Statistical hypothesis test for comparing simultaneous (Sim.) vs individual (Ind.) retrieval.

	Sim. vs Ind.	z-statistics	*t*-score	p-value	Median of RMSE (Sim.: Ind.)
BRANN	LAI	− 3.48		0.0004	(0.48: 0.51) **
	LSM	− 2.33		0.02	(2.37: 2.38) *
	LCC		7.61 [a]	1.03e-12	(5.85: 6.03) ****
	LNC		11.04 [a]	1.84e-22	(0.23: 0.25) ****
LM_ANN	LAI	−11.46		2.03e-30	(0.50: 0.67) ****
	LSM	−1.37		0.169	(2.66: 2.69) [b]
	LCC	− 7.66		1.72e-14	(6.33: 6.89) ****
	LNC	−10.50		8.56e-26	(0.24: 0.30) ****

[a]RMSEs of LCC and LNC, applying BRANN in both simultaneous and individual retrievals, passed the Shapiro–Wilk normality test, so the statistical comparison of them was performed using the parametric paired sample *t*-test.

*significant at p < 0.05

**significant at p < 0.005

***significant at p < 0.001

****significant at p < 0.0001.

[b]The alternative hypothesis is rejected.