Trees classification based on Fourier coefficients of the sapflow density flux[*]

Dmitry Efrosinin[ab]**, Irina Kochetkova**[bc]**, Natalia Stepanova**[d]**,
Alexey Yarovslavtsev**[be]**, Konstantin Samouylov**[bc]**,
Riccardo Valentini**[fb]

[a]Johannes Kepler University Linz, Austria
`dmitry.efrosinin@jku.at`

[b]Peoples' Friendship University of Russia (RUDN University), Russia
`gudkova-ia@rudn.ru, yaroslavtsev-am@rudn.ru, samuylov-ke@rudn.ru`

[c]Institute of Informatics Problems, Federal Research Center
"Computer Science and Control" of RAS, Russia

[d]V.A. Trapeznikov Institute of Control Sciences of RAS, Russia
`natalia0410@rambler.ru`

[e]LAMP, Russian Timiryazev State Agrarian University, Russia,

[f]Tuscia University, Viterbo, Italy
`rik@unitus.it`

Abstract

In this paper we study the possibility to use the artificial neural networks for trees classification based on real and approximated values of the sap flow density flux describing water transport in trees. The data sets were generated by means of a new tree monitoring system TreeTalker[©]. The Fourier series-based model is used for fitting the data sets with periodic patterns. The multivariate regression model defines the functional dependencies between sap flow density and temperature time series. The paper shows that Fourier

coefficients can be successfully used as elements of the feature vectors required to solve different classification problems. Here we train multilayer neural networks to classify the trees according to different types of classes. The quality of the developed model for prediction and classification is verified by numerous numerical examples.

Keywords: TreeTalker monitoring system, Fourier coefficients, neural network, classification of trees

AMS Subject Classification: 65C60, 62M10

1. Introduction

In the last decade, monitoring systems, which can be treated as a part of smart technologies and are used for generating large amounts of data sets from a network of sensors, have evolved rapidly. This work is a continuation of a previous survey related to a new sensor tree monitoring system TreeTalker$^{©}$ (TT) [12]. The TT is a system used for real-time ecological forest plantation monitoring made with the concept of the Internet of Things (IoT). It is responsible for collecting data from all sensors, which are fixed to trees, and transforming the analog signal to meaningful variables. With this system a database was created which is expected to be published shortly. It includes, among other things like temperature, humidity of the air and wood, spectral characteristics of the canopy, radial growth of the trunk, and data from accelerometer about 3D position of trunk, a large amount of information on the sap density flux describing water transport process in different tree species that also differ in age, health status, metric characteristics, etc.

We report in this paper our first experiments carried out on data sets extracted by the TT monitoring system as well as on the estimated values of the density flux and dedicated to trees classification. Classification is a very common use case of a machine learning. Artificial neural networks [6, 9] (NN) are parts of a supervised machine learning which are most popular in different problems of data classification, pattern recognition, regression, clustering, time series forecasting. Some of applications of the NN to the real data sets can be found, e.g. in [1, 4]. We study the possibility to use NN for classifying the trees of different species within the same age group and visual-tree-assessment (VTA) score, the same species but with different age groups and/or VTA scores as well as with trunk diameters. As classification features we use Fourier coefficients obtained by fitting the truncated Fourier series to the sapflow density flux data sets. This paper also shows that in addition to the Fourier coefficients obtained directly for the sap density function, coefficients estimated using multivariate regression on air temperature data can also be successfully used to classify trees. Thus, it is possible to simulate the sap flow process of a particular tree species. In the long term, this approach, which incorporates data generated by the TT with the proposed Fourier coefficient estimation method, can be used to trace ecological anomalies in the health status of an individual tree or entire forest area.

The rest of the paper is organized as follows. In Section 2 we describe the data

sets and methods. In Sections 3 we study the problem of trees classification which is based on Fourier coefficients estimated in observable time period. Final remarks are given in Section 4.

2. Data sets and methods

In most cases, data is collected hourly, but in Russian conditions, 1.5 hours are recommended. The performance of TreeTalkers is being tested at several sites all over Eurasia continent from Spain to China, with a wide variety of tree species, climate, topography, and land use with multiple tests of device reliability in terms of sensors operational limits of the sensors, data transmission, and battery effectiveness. In May 2019, 60 TT sensors were installed on different species of trees growing in different areas, belonging to different age groups with varying VTA scores summarized in Table 1. Data from all devices was collected till the end of November 2019 and stored on a remote web server. All data for basic variables was 3 sigma filtered, runaways were eliminated and gap filled with linear 4D interpolation for gaps smaller than 4 measurements. Sapwood area data was combined with TT data, and individual tree sap flux was calculated utilizing R software. While there are several papers where modern modeling techniques implemented for predicting evapotranspiration of planted areas with environmental data [7, 11, 14], there are very limited amount of papers about individual tree sap flow modeling [10]. Taking into account growing trend of IoT devices used in environmental monitoring [3, 13] modeling of one of the main physiological tree characteristics can be of great interest. This metric is in many cases a very contrasting reflection of the degree to which individual qualitative factors influence the state of the tree. As an example, consider Figure 1, which shows the sap density functions for two tree species: *Salix alba* and *Acer platanoides*. In first case we have three classes (III,2), (IV,2) and (IV,3), where trees differ in age group and VTA factor. In second case trees belong to the same age group but differ in VTA value according to four classes, (VI,1), (VI,2), (VI,3) and (VI,4). As can be seen, the functions $y_{\mathrm{flux},t}$ have a significant difference, depending on the respective class. Thus, it can be expected that this characteristic can be successfully applied for classification.

The data sets for the the air temperature (*tair*) and the sapflow density flux (*flux*) are represented respectively in form of time series $\vec{y}_{\mathrm{tair}} = (y_{\mathrm{tair},1}, y_{\mathrm{tair},2}, \dots)$ and $\vec{y}_{\mathrm{flux}} = (y_{\mathrm{flux},1}, y_{\mathrm{flux},2}, \dots)$ and with time-ordered sequence of observations. These time series are characterized by fluctuations which exhibit a periodic nature with a cycle length T. Each cycle includes mostly $N = 16$ measurements that are made at equally spaced time intervals $\Delta t = 1.5h$. The total time within a cycle is then $T = N\Delta t = 24h$. Since the fluctuations may have different amplitudes and shapes within each period, the data sets can not be treated as pure periodic ones, i.e. in general case $y_{\mathrm{tair},t_0} \neq y_{\mathrm{tair},t_0+kT}$ and $y_{\mathrm{flux},t_0} \neq y_{\mathrm{flux},t_0+kT}$ for all $k \in \mathbb{N}$. The data preprocessing step includes the denoising by locally data smoothing. We use for that a low pass filter which passes signals with a frequency lower than a selected cutoff frequency $\omega_c = 0.9$. Smaller values of ω_c result in greater smoothing.

(a) (b)
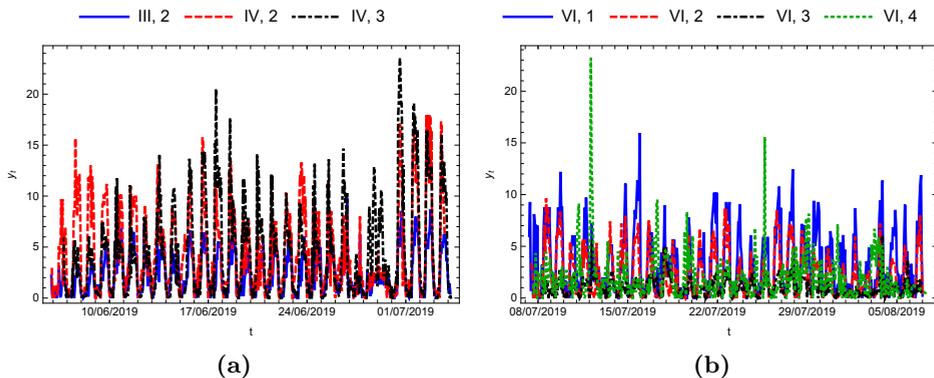
Figure 1. Sapflow density flux y_t *Salix alba* (a)
and *Acer platanoides* (b).

Table 1. Key elements of the database.

Sorts	Area	Age	VTA score
Acer fraxinifolium	MS-7	IV,IV	2,3
Acer platanoides	OL-1,OL-2,OL-3,MS-6	III,IV,VI	1,2,3,4
Aesculus flava	OL-3	VI	3
Betula pendula	OL-1,MS-4,MS-6	IV,VI	1,2,3
Carpinus betulus	OL-4	IV	2
Fraxinus excelsior	OL-3	VI	3
Fraxinus ornus	OL-3	VI	2
Juglans mandshurica	OL-3	VI	3
Junglas cinerea	OL-3	VI	2
Larix decidua	MS-6	IV,VI	2
Larix sibirica	OL-2,OL-3,MS-6	V,VI	2,3,4
Malus domestica	OL-3	IV	3
Picea abies	OL-1,MS-4	IV,V,VI	1,2,3
Pinus sylvestris	OL-1	III,IV,VI	2,3
Populus nigra	OL-3,MS-7	VI	2,3
Populus tremula	OL-1	III,VI	1,2
Prunus avium	OL-3	VI	2
Pyrus commutis	OL-3	VI	2
Quercus Rrobur	OL-2,OL-3,MS-4	VI	2,3,4
Robinia pseudoacacia	OL-3	IV	2
Salix alba	OL-3,MS-5	III,IV,VI	2,3
Tilia cordata	OL-1,OL-2,OL-3,MS-4,MS-6	III,IV,VI	1,2,3,4

A truncated Fourier series can be used to find approximations for periodic functions
of the air temperature $f_{\text{tair}}(t)$ and the sapflow density $f_{\text{flux}}(t)$ with a fundamental

period T that passes through all of the points,

$$f_{\text{tair}}(t) \approx a_0 + \sum_{n=1}^{m} \left[a_n \cos\left(\frac{2\pi nt}{T}\right) + b_n \sin\left(\frac{2\pi nt}{T}\right) \right],$$

$$f_{\text{flux}}(t) \approx \alpha_0 + \sum_{n=1}^{m} \left[\alpha_n \cos\left(\frac{2\pi nt}{T}\right) + \beta_n \sin\left(\frac{2\pi nt}{T}\right) \right]. \tag{2.1}$$

The coefficients a_n, b_n and α_n, β_n can not be explicitly derived since the functions $f_{\text{tair}}(t)$ and $f_{\text{flux}}(t)$ are not available in explicit form and hence they must be estimated. We have only data $\vec{y}_{\text{tair}} = (y_{\text{tair},1}, y_{\text{tair},2}, \ldots, y_{\text{tair},n_s})'$ and $\vec{y}_{\text{flux}} = (y_{\text{flux},1}, y_{\text{flux},2}, \ldots, y_{\text{flux},n_s})'$ generated by the sensors. The known periodic patterns of the approximated functions $f_{\text{tair}}(t)$ and $f_{\text{flux}}(t)$ are expressed through vectors of parameters $\vec{a} = (a_0, a_1, \ldots, a_m, b_1, \ldots, b_m)'$ and $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m)'$. These parameters are estimated using the method of the linear least squares

$$\sum_{t=(i-1)T}^{iT} (y_{\text{tair},t} - f_{\text{tair}}(t))^2 \Rightarrow \min_{\vec{a}}, \quad \sum_{t=(i-1)T}^{iT} (y_{\text{flux},t} - f_{\text{flux}}(t))^2 \Rightarrow \min_{\vec{\alpha}}, \ 1 \leq i \leq n_p,$$

where $n_p = \frac{n_s}{N}$ is a number of cycles of length T within the observations with a total sample size n_s.

For trees classification the feature vectors $(\alpha_{i,0}, \alpha_{i,1}, \ldots, \alpha_{i,m}, \beta_{i,1}, \ldots, \beta_{i,m})$ of the Fourier coefficients (2.1) in an observable period $1 \leq i \leq n_p$ are used. Recall that in the previous paper [5] we presented a method for predicting the density flux during the day based on data on air temperature during the observed cycle. For this purpose, Fourier series and a multivariate regression model were used, establishing the functional relationship between the respective Fourier coefficients for temperature data sets and density flux values,

$$\alpha_{i,n} = \theta_{0,n} + \theta_{1,n} a_{i,0} + \sum_{j=1}^{m} [\theta_{j+1,n} a_{i,j} + \theta_{m+j+1,n} b_{i,j}], \ 0 \leq n \leq m, \tag{2.2}$$

$$\beta_{i,n} = \theta_{0,n+m} + \theta_{1,n+m} a_{i,0} + \sum_{j=1}^{m} [\theta_{j+1,n+m} a_{i,j} + \theta_{m+j+1,n+m} b_{i,j}], \ 1 \leq n \leq m,$$

where $1 \leq i \leq n_p$, $\vec{\theta}_k = (\theta_{0,k}, \theta_{1,k}, \ldots, \theta_{2m+1,k})', 0 \leq k \leq 2m$, denotes the vector of parameters of the multidimensional regression model. We discuss here the results of experiments carried out on data sets extracted by the TT monitoring system as well as on the estimated values of the density flux and dedicated to trees classification. We study the possibility to use artificial neural networks to classify the trees of the same species but with different age groups and visual-tree-assessment (VTA) scores. As classification features we use a predicted Fourier coefficients of the sap flow density flux approximation function. In the long term, this approach which incorporates data generated by the TT with the proposed Fourier coefficient estimation method can be used to determine the anomalous state of a tree or generally monitor forest ecology.

As was mentioned above, as features for trees classification we use the sets of vectors S, consisting of eleven original coefficients of the truncated Fourier series fitted to the density flux function $y_{\mathrm{flux},t}$. Moreover, the classifier will be applied also to the sets $\hat{S}$ for the predicted coefficients of the function $\hat{y}_{\mathrm{flux},t}$ by using the multiple regression (2.2) for the Fourier coefficients of the air temperature data $y_{\mathrm{tair},t}$. The data sets for the classification problem were prepared in form of the set of mappings,

$$S = \{(\alpha_{i,0}, \alpha_{i,1}, \ldots, \alpha_{i,m}, \beta_{i,1}, \ldots, \beta_{i,m}) \to \text{Class } N : 1 \leq i \leq n_p\},$$
$$\hat{S} = \{(\hat{\alpha}_{i,0}, \hat{\alpha}_{i,1}, \ldots, \hat{\alpha}_{i,m}, \hat{\beta}_{i,1}, \ldots, \hat{\beta}_{i,m}) \to \text{Class } N : 1 \leq i \leq n_p\},$$

where $m = 5$ and n_p is a number of observable periods. 70% of samples S and $\hat{S}$ is referred to as training data and the rest – as validation data. The data were chosen so that the sample in each class was more or less balanced, i.e. the sample size in each class did not differ significantly. The multilayer neural network is used for the data classification. It can be formally defined as a function $f : \vec{\alpha} \to \vec{y}$, which maps an input vector $\vec{\alpha}$ of dimension $2m + 1$ to an estimate output $\vec{y} \in \mathbb{R}^{N_c}$ of the class number $N = 1, \ldots, N_c$. The network is decomposed into 6 layers as illustrated in Figure 2, each of which represents a different function mapping vectors to vectors. The successive layers are: a linear layer with an output vector of size k, a nonlinear elementwise activation layer, other three linear layers with output vectors of size k, and a nonlinear normalization layer.
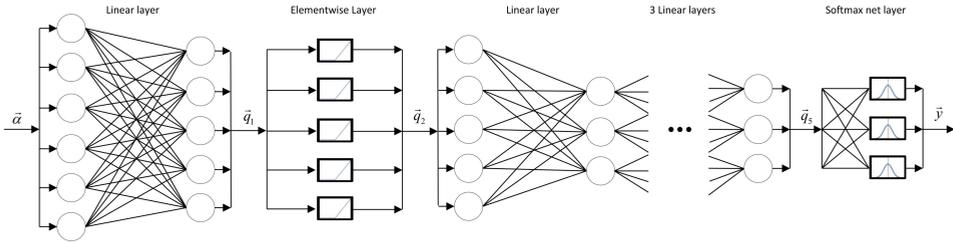
Figure 2. Architecture of the neural network.

The first layer is an affine transformation

$$\vec{q}_1 = W_1\vec{\alpha} + \vec{b}_1,$$

where $\vec{q}_1 = \mathbb{R}^{2m+1}$ is the output vector, $W \in \mathbb{R}^{2m+1 \times k=30}$ is the weight matrix, $\vec{b}_1 \in \mathbb{R}^{2m+1}$ is the bias vector. The rows in W_1 are interpreted as features that are relevant for differentiating between corresponding classes. Consequently, $W_1\vec{\alpha}$ is a projection of the input $\vec{\alpha}$ onto these features. The second layer is an elementwise activation layer which is defined by the nonlinear function $\vec{q}_2 = \max(0, \vec{q}_1)$ setting negative entries of q_1 to zero and uses only positive entries. The next three layers layers are another affine transformations,

$$\vec{q}_i = W_i\vec{q}_{i-1} + \vec{b}_i,$$

where $\vec{q}_i \in \mathbb{R}^k, W_i \in \mathbb{R}^{k \times k}$, and $b_i \in \mathbb{R}^k, i = 3, 4, 5$. The last layer is the normalization layer $\vec{y} = \text{softmax}(\vec{q}_5)$, which componentwise is of the form

$$y_N = \frac{e^{q_{5N}}}{\sum_N e^{q_{5N}}}, \; N = 1, \ldots, N_c.$$

The last layer normalizes the output vector $\vec{y}$ with the aim to get the values between 0 and 1. The output $\vec{y}$ can be treated as a probability distribution vector, where the Nth element y_N represents the likelihood that $\vec{\alpha}$ belongs to class N.

The neural networks in our experiments are trained by the ADAM (adaptive moment estimation method) [8] which is a modification of stochastic gradient descent (SGD). The neural network toolbox in *Mathematica*$^©$ of the Wolfram Research is used. We verify the classifier which should be accurate enough to be used to predict new output from verification data. The algorithm was ran many times on samples and networks with different sizes. In all cases the results were quite positive and indicate the potential of machine learning methodology for trees classification problem based on the estimated Fourier coefficients.

To follow the classification progress, we summarize the results in form of a confusion matrix evaluated on the trained and verification data. with these matrices it is possible to observe the relations between the classifier outputs and the true ones. Each row of these matrices represents the instances in a predicted value while each column represents the instances in an actual value. Different statistical measures of the performance of a binary classification, such as the overall accuracy (ACC), sensitivity (true positive rate – TPR), specificity (true negative rate – TNR) as well as F-1 Scores which is the harmonic mean of precision and sensitivity. For more details about these measures, refer to [2]. Note that in multi-class classification problem we calculate the F-1 Score per class in a one-vs-rest manner, i.e. we estimate successful occurrence of the class as if there are individual classifiers for each class.

3. Experiments

Five main examples are discussed in this section.

Example 3.1. In this example, we test the feasibility of using the data to classify tree varieties within the same age group and VTA score. Data for four tree varieties such as *Acer platanoides*, *Betula pendula*, *Salix albe* and *Tilla cordata* in the age group IV and with the VTA score 2 were selected for 4-class classification problem as is shown in Table 2.

The Figure 3 shows the confusion matrices evaluated using data set S and $\hat{S}$ for real and estimated Fourier coefficients of the data flux density function $y_{\text{flux},t}$. Obviously, the matrices are diagonally dominant and the frequencies of correctly recognized classes are almost identical. Four statistical quantities described above, which are used to represent some aspect of a classification quality, are summarized in Table 3. The quality of the classification is quite high, the overall accuracy is over

than 85%. The quality metrics per class take also the high values. Moreover, the classification of trees according to the estimated Fourier coefficients exhibits slightly reduced values for quality parameters, but the difference is quite insignificant.

Table 2. Classes within the same age group IV and VTA score 2.

Class N	Sort	Age group	VTA score
1	Acer platanoides	IV	2
2	Betula pendula	IV	2
3	Salix albe	IV	2
4	Tilia cordata	IV	2

(a)   (b)
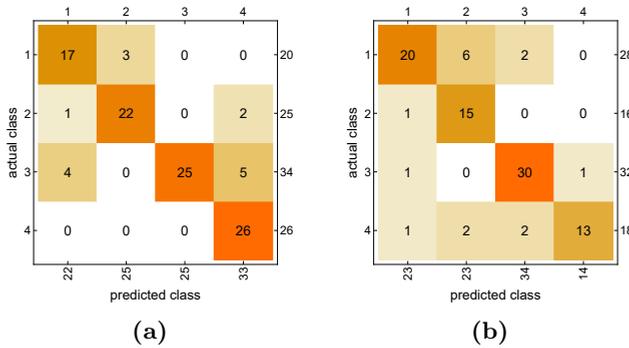
Figure 3. Confusion matrices for sorts classification based on S (a) and $\hat{S}$ (b) data sets.

Table 3. Classification performance.

Data \ Metric	ACC	TPR	TNR	F-1 Scores
S	0.8571	$1 \to 0.8500$ $2 \to 0.8800$ $3 \to 0.7353$ $4 \to 1.000$	$1 \to 0.9412$ $2 \to 0.9625$ $3 \to 1.000$ $4 \to 0.9114$	$1 \to 0.8095$ $2 \to 0.8800$ $3 \to 0.8474$ $4 \to 0.8814$
$\hat{S}$	0.8298	$1 \to 0.7143$ $2 \to 0.9375$ $3 \to 0.9375$ $4 \to 0.7222$	$1 \to 0.9955$ $2 \to 0.8974$ $3 \to 0.9355$ $4 \to 0.9864$	$1 \to 0.7843$ $2 \to 0.7692$ $3 \to 0.9091$ $4 \to 0.8125$

Based on the available samples, it can thus be stated that the sap flow process varies considerably among the different tree varieties. We believe that by obtaining an appropriate trained neural network for each tree variety of a certain age group and

VTA score, it is feasible to recognize anomalies in the growth process of a particular tree, which in turn will make it possible to produce an environmental health map of forest plantations in urban parks or large forest areas outside of cities.

The experiments carried out comparing different tree varieties in terms of sap density values show the possibility to use not only the real values obtained directly by the TT monitoring system, but also their estimates obtained by multivariate linear regression as a functional relationship between air temperature and sap density values. This allows considerable savings in the purchase and installation of a large number of sensors, as a sensor network of a limited number of devices installed on different types of trees will be sufficient to cover large areas.

Example 3.2. Consider data sets with n_p observable periods for *Salix albe*. We divide the data set into three subgroups according to Table 4.

Table 4. Classes of *Salix albe*.

Class N	Age group	VTA score
1	IV	2
2	IV	3
3	III	2
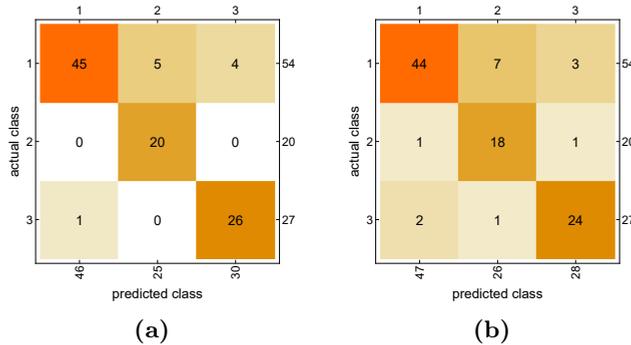
Figure 4. Confusion matrices for classification of *Salix alba* based on $y_{\text{flux},t}$ (a) and $\hat{y}_{\text{flux},t}$ (b).

The trees of this species can belong to different age groups and have different VTA scores. Figure 4 illustrates two confusion matrices which are obviously diagonally dominant. As we can see, factors such as age group and VTA have a significant influence on the values of the density flux function. As can be seen in Table 5, the classification accuracy reaches more than 90% and the overall accuracy is almost indistinguishable from the qualitative characteristics for each class. The results of experiments were quite positive and indicate the potential of machine learning methodology for trees classification problem based on the Fourier coefficients for the fitted density flux data.

Table 5. Classification performance.

Data \ Metric	ACC	TPR	TNR	F-1 Scores
Salix albe, S	0.9009	$1 \to 0.8333$ $2 \to 1.000$ $3 \to 0.9629$	$1 \to 0.9787$ $2 \to 0.9382$ $3 \to 0.9459$	$1 \to 0.9000$ $2 \to 0.8889$ $3 \to 0.9123$
Salix albe, $\hat{S}$	0.8514	$1 \to 0.8148$ $2 \to 0.9000$ $3 \to 0.9231$	$1 \to 0.9348$ $2 \to 0.9125$ $3 \to 0.9459$	$1 \to 0.8713$ $2 \to 0.8000$ $3 \to 0.8889$

Example 3.3. Next we study the possibility to classify the trees of the same species according to different age groups but with the equal VTA scores. The presented experiment includes the gathered data for *Tilia cordata* and the task is to provide a classification according to the classes in Table 6.

Table 6. Classes of *Tilia cordata*.

Class N	Age group	VTA score
1	III	3
2	IV	3
3	VI	3

(a)   (b)
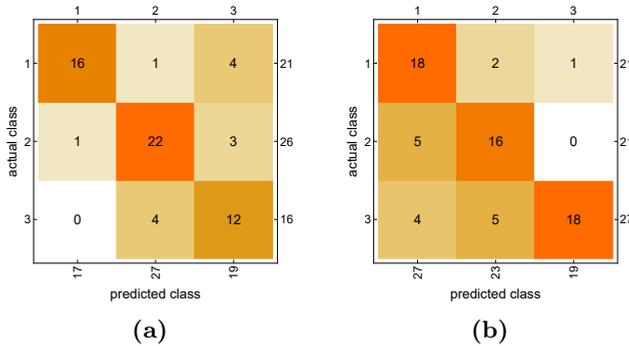
Figure 5. Confusion matrices for classification of *Tilia cordata* based on real $y_{\text{flux},t}$ (a) and estimated $\hat{y}_{\text{flux},t}$ (b) density flux function.

The Figure 5 shows two confusion matrices in a 3-class classification problem. As we see here, the matrices are also diagonally dominated but nevertheless there are non-zero false positive and false negative elements. The overall accuracy together with other quality characteristics per class are summarized in Table 7. In this example we obtained over 79% accuracy for trees classification. The age classification of other tree species yielded fairly similar results. Hence a the sapflow

density can be treated as a characteristic for determining the age of a tree. This result can also be considered encouraging given the high noise content of the raw data, erroneous measurements, and missing values within individual classes. The use of Fourier coefficients derived from air temperature for classification can also be considered acceptable, although of course the quality is slightly degraded and in average is near 74%. During the experiments, we also noticed that when the VTA score is increased, the trees are more accurately classified according to the age group. Finally, it can be noticed that the higher the age of the trees, the more closely the sap density function takes on values. Classification then becomes in this case a more difficult task. This can be seen from the low values of the classification quality characteristics for classes 2 and 3 in Table 7.

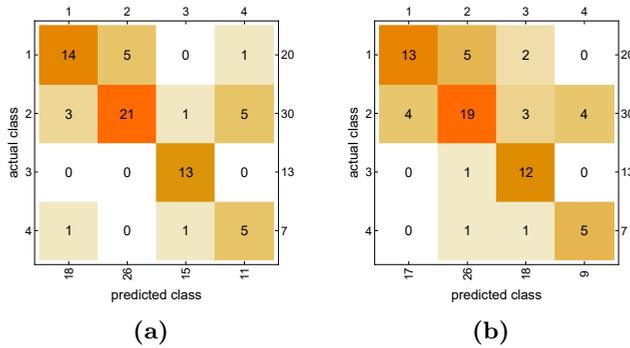**Table 7.** Classification performance.

Data \ Metric	ACC	TPR	TNR	F-1 Scores
$y_{\text{flux},t}$	0.7937	$1 \to 0.8421$ $2 \to 0.8302$ $3 \to 0.6857$	$1 \to 0.7619$ $2 \to 0.8461$ $3 \to 0.7500$	$1 \to 0.9762$ $2 \to 0.8649$ $3 \to 0.8511$
$\hat{y}_{\text{flux},t}$	0.7430	$1 \to 0.8571$ $2 \to 0.7619$ $3 \to 0.6429$	$1 \to 0.8163$ $2 \to 0.8367$ $3 \to 0.9762$	$1 \to 0.7500$ $2 \to 0.7111$ $3 \to 0.7659$

Example 3.4. Now we will fix the age group and try to classify the trees by VTA scores only. Consider data sets for *Acer platanoides*. The data were divided into four subgroups according to Table 8.

Table 8. Classes of *Acer platanoides*.

Class N	Age group	VTA score
1	VI	1
2	VI	2
3	VI	3
4	VI	4

The confusion matrices in Figure 6, although diagonally dominant, contain many non-zero elements outside the main diagonal. The reason is, that there was not much variation in the density flux data in each group. Therefore, we consider the classification accuracy of more than 75% as a very good result, taking into account that VTA is still a somewhat subjective characteristic. Table 9 shows that some classes are better recognized than others. This is not surprising, as it is obvious that in addition to age group and VTA there are other factors that influence the value of juice density, such as trunk diameter. In the following example we investigate the task of classification according to this characteristic.

Figure 6. Confusion matrices for classification of *Acer platanoides* based on data set S (a) and $\hat{S}$ (b).

Table 9. Classification performance.

Metric / Data	ACC	TPR	TNR	F-1 Scores
Acer platanoides, S	0.7571	$1 \rightarrow 0.7000$ $2 \rightarrow 0.7000$ $3 \rightarrow 1.000$ $4 \rightarrow 0.7143$	$1 \rightarrow 0.9200$ $2 \rightarrow 0.8750$ $3 \rightarrow 0.9649$ $4 \rightarrow 0.9048$	$1 \rightarrow 0.7368$ $2 \rightarrow 0.7500$ $3 \rightarrow 0.9286$ $4 \rightarrow 0.5556$
Acer platanoides, $\hat{S}$	0.7000	$1 \rightarrow 0.6500$ $2 \rightarrow 0.6333$ $3 \rightarrow 0.9230$ $4 \rightarrow 0.7143$	$1 \rightarrow 0.9200$ $2 \rightarrow 0.8333$ $3 \rightarrow 0.8947$ $4 \rightarrow 0.9365$	$1 \rightarrow 0.7027$ $2 \rightarrow 0.6786$ $3 \rightarrow 0.7742$ $4 \rightarrow 0.5039$

Example 3.5. In this example we try to classify the trees by trunk diameter based on the density flux information with fixed factors of the age group and the VTA score. Two tree species are selected for the illustration: *Betula pendula* and *Tilia cordata*.

Table 10. Classes of *Betula pendula* (a) and *Tilia cordata* (b).

Class N	Diam	Age	VTA
1	25.46	VI	1
2	26.73	VI	1
3	30.24	VI	1

(a)

Class N	Diam	Age	VTA
1	37.87	VI	2
2	48.76	VI	2
3	51.24	VI	2

(b)

Three different classes for each species are enumerated respectively in Table 10. Here we see that the quality of classification by trunk diameter is slightly higher than by age group, although these two factors have a large positive correlation for

almost all the tree species under consideration. As features we use here only the Fourier coefficients of data sets of type S obtained by fitting the truncated Fourier series to the density flux data sets. But we expect that the results will be similar to the case of the estimated Fourier coefficients. The results of classification are illustrated as usual in form of the confusion matrices in Figure 7 and in Table 11 of performance measures.
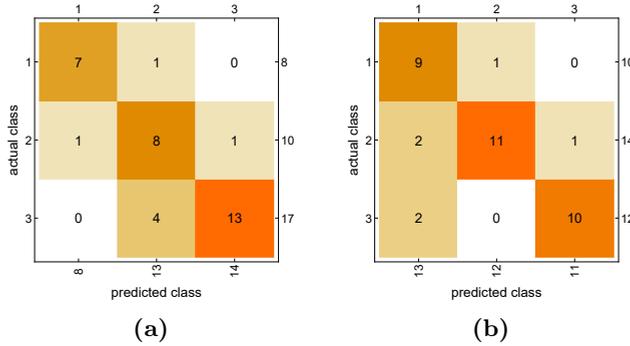
Figure 7. Confusion matrices for classification of *Betula pendula* (a) and *Tilia cordata* based on real $y_{\text{flux},t}$ density flux function.

Table 11. Classification performance.

Metric / Data	ACC	TPR	TNR	F-1 Scores
Betula pendula, S	0.8000	$1 \rightarrow 0.8750$ $2 \rightarrow 0.8000$ $3 \rightarrow 0.7647$	$1 \rightarrow 0.9629$ $2 \rightarrow 0.8000$ $3 \rightarrow 0.9444$	$1 \rightarrow 0.8750$ $2 \rightarrow 0.6956$ $3 \rightarrow 0.8387$
Tilia cordata, S	0.8333	$1 \rightarrow 0.9000$ $2 \rightarrow 0.7857$ $3 \rightarrow 0.8333$	$1 \rightarrow 0.8462$ $2 \rightarrow 0.9545$ $3 \rightarrow 0.9583$	$1 \rightarrow 0.7816$ $2 \rightarrow 0.8461$ $3 \rightarrow 0.8677$

Here we see that the quality of classification by trunk diameter is more than 80% which is slightly higher than the classification by the age group, although these two factors have a large positive correlation for almost all the tree species under consideration.

4. Conclusion

On the basis of the proposed experiments, it can be noticed that the temperature observations can be mapped to the values of the sap flow density flux through the corresponding Fourier coefficients which is resulting in high quality predictions. Moreover, the estimated coefficients for the function approximating the sap

flow density have a good potential to be used as feature vector in trees classification tasks even within the same species. From this we can draw a conclusion about the perspective to use the TreeTalker equipment together with the proposed mathematical approach for solving problems of trees monitoring and anomaly state recognition. Moreover, if a tree's sapflow density pattern does not match what a healthy tree with similar characteristics should have, this can be seen as an indirect sign of problems with soil, groundwater or the general environment. As new data become available, we plan to continue our research on tree classification based on the monitoring system. We will also take into account the reviewer's suggestion related to the use of alternative classifiers and a comparative analysis of classification quality.

References

[1] L. Ahrens, J. Ahrens, H. Schotten: *A machine-learning phase classification scheme for anomaly detection in signals with periodic characteristics*, Journal of Advances in Signal 27 (2019), p. 23,
DOI: https://doi.org/10.1186/s13634-019-0619-3.

[2] D. G. Altman, J. M. Bland: *Statistics Notes: Diagnostic tests 1: sensitivity and specificity*, BMJ 308.6943 (1994), p. 1552,
DOI: https://doi.org/10.1136/bmj.308.6943.1552.

[3] A. Boursianis, M. Diamantoulakis, A. Liopa-Tsakalidi, P. Barouchas, G. Salahas, S. Goudos: *Internet of Things (IoT) and Agricultural UnmannedAerial Vehicles (UAVs) in Smart Farming: A Comprehensive Review*, Internet of Things 100187 (2020).

[4] B. Colvert, E. Kanso, E. Alsalman: *Classifying vortex wakes using neural networks*, Bioinspiration and Biomimetics 13.2 (2017), pp. 1–11,
DOI: https://doi.org/10.1088/1748-3190/aaa787.

[5] D. Efrosinin, I. Kochetkova, N. Stepanova, A. Yarovslavtsev, K. Samouylov, R. Valentini: *The Fourier Series Model for Predicting Sapflow Density Flux based on TreeTalker Monitoring System*, in: LNCS, NEW2AN 2020 (to be published), St. Petersburg, Russia: Springer, 2020.

[6] C. Gershenson: *Artificial Neural Networks for Beginners*, 2003, arXiv: cs/0308031 [cs.NE].

[7] F. Junliang, W. Yue, F. Zhang, H. Cai, X. Wang, X.-A. Lu, Y. Xiang: *Evaluation of SVM, ELM and four tree-based ensemble modelsfor predicting daily reference evapotranspiration using limited meteorological data in differentclimates of China*, Agricultural and Forest Meteorology 263 (2018).

[8] D. P. Kingma, J. Ba: *Adam: A Method for Stochastic Optimization*, 2014, arXiv: 1412.6980 [cs.LG].

[9] S. Russell, P. Norvig: *Artificial Intelligence: A Modern Approach*, 3rd, USA: Prentice Hall Press, 2009, ISBN: 0136042597.

[10] J. Siqueira, T. Pac, J. Silvestre, F. Santos, A. Falcao, L. Pereira: *Generating fuzzy rules by learning from olive tree transpiration measurement – An algorithm to au-tomatize Granier sap flow data analysis*, Computers and Electronics in Agriculture 101 (2014).

[11] D. Tang, Y. Feng, W. Hao, N. Cui: *Evaluation of artificial intelligence models for actual crop evapotranspiration modeling in mulched and non-mulchedmaize croplands*, Computers and Electronics in Agriculture 152 (2018).

[12] R. Valentini, L. Marchesini, D. Gianelle, G. Sala, A. Yarovslavtsev, V. Vasenev, S. Castaldi: *New Tree Monitoring Systems: From Industry 4.0 to Nature 4.0.* Annals of Silvicultural Research 43.2 (2019), pp. 84–88, DOI: `http://dx.doi.org/10.12899/asr-1847`.

[13] G. Xu, Y. Shi, X. Sun, W. Shen: *Internet of things in marine environment moni-toring: A review*, Sensors 19.7 (2019).

[14] S. Yamac, M. Todorovic: *Estimation of daily potato crop evapotranspirationusing three different machine learning algorithms and four scenarios of available meteorolog-ical data*, Agricultural Water Management 228 (2020).