This is an ACCEPED MANUSCRIPT of an article published by Taylor & Francis in Enterprise Information Systems on 16 July 2020, available online: http://www.tandfonline.com/10.1080/17517575.2020.1793391

RealForAll: Real-time System for Automatic Detection of Airborne Pollen

Danijela Tešendić[a], Danijela Boberić Krstićev[a], Predrag Matavulj[b], Sanja Brdar[b], Marko Panić[b], Vladan Minić[b] and Branko Šikoparija[b]

[a]Faculty of Sciences, University of Novi Sad, Serbia; [b]BioSense Institute - Research Institute for Information Technologies in Biosystems, University of Novi Sad, Novi Sad, Serbia

The aim of this paper is to describe a solution suitable for the automation of standard pollen information service (EN 16868:2019). We are describing the RealForAll integrated information system developed for automatic airborne pollen detection and real-time data delivery to end-users. This solution is based on the measurements from the Rapid-E airborne particle monitor. The system incorporates an AI-enabled subsystem based on a convolutional neural network that continuously retrieves raw data from Rapid-E and performs the classification of airborne pollen. The main advantages of this system reflect in real-time data delivery and independence of aerobiology experts during the pollen season.

Keywords: process automation; integrated information system; neural networks; classification; health care;

1. Introduction

Pollen is one of the most common triggers of seasonal allergies and around 30% of the world population suffers from some form of allergic disease (Akdis, Hellings, and Agache 2015). In most European countries, national organizations of various kinds provide information about pollen concentration in the air, publish pollen forecasts and issue warnings. Method for sampling and analysis of airborne pollen is standardized by EN 16868:2019 (Ambient air - Sampling and analysis of airborne pollen grains and fungal spores for networks related to allergy - Volumetric Hirst method). That standard prescribes the use of Hirst-type volumetric devices (Hirst 1952) for sampling airborne particles. The device sucks in 10 l of air with suspended aerosols per minute, which then impacts on the adhesive coated, transparent plastic tape that moves past the inlet at 2 mm per hour to give a time-related sample. The collected samples are analysed manually using a light microscope (Galán et al. 2014; Buters et al. 2018). This is a very tedious, labour-intensive and time consuming method and measurement data are always delayed from a few days up to a few weeks.

Real-time measurements of airborne pollen concentrations can improve the quality of life in a pollen sensitive population. Timely information can help people to prevent allergy symptoms and to better manage their allergic diseases (Bousquet et al. 2019). If patients had access to information about immediate exposure levels, they could take appropriate medication and plan their activities. Also, pollen forecasts play an integral role in the management of pollen allergies. Real-time measurements can be used for improvement of short term forecasts, particularly by enabling assimilation of measurement data and models (Sofiev et al. 2017). In addition, up-to-date information on the concentration of harmful airborne particles in the air (e.g. fungi and spores) can be very effective in agriculture and forestry (Garzia-Mozo, 2011). For example, application of protective agents at the right moment can prevent crop damage and increase yields.

At present, a number of web portals and mobile applications provide outdated information about airborne pollen concentration or estimated current values resulting from coupling the previous day’s observations with pollen forecasts (Pasyfo 2019; Polleninfo 2019; Norkko 2019). In order to improve the pollen information service and provide real-time measurements to end users it is necessary to automate the whole process of pollen detection. Automation of this process requires a measurement device (particle monitor) capable of producing data in digital format which can be further processed automatically. Such a device should sample and characterize single airborne particles with sufficient detail to enable their identification. Application of advanced technologies made real-time pollen monitoring possible only recently. Currently, two types of technologies seem to be the most suitable for detecting airborne pollen: automatic multi stack image recording and laser-induced fluorescence (Huffman et. al, 2019). Another challenge in automatic pollen detection is to develop an information system which will harvest raw data from particle monitors, process it and disseminate classification results to end users. Despite suitable particle monitors being commercially available, there is no out-of-the-box system that integrates detection and classification particles of interest. Such automatic integrated systems have to be developed considering the needs of stakeholders.

An interactive map (https://oteros.shinyapps.io/pollen_map/) visualises distribution of pollen monitoring stations throughout the world (Buters et al. 2018). Search for only automatic stations on the map, results in 4 types of operational solutions for pollen detection. However, they have some limitations regarding time necessary to produce information or the number of pollen types which can be identified. Japan is the pioneer in automatic detection of airborne pollen using the KH-3000 particle monitor. Their solution is limited to monitoring of only concentrations of Japanese cedar

(Hanakosan 2019; Kawashima et al. 2017). BAA500 particle monitors are used across Germany. A solution integrating those monitors delivers pollen concentrations for various pollen types but with an average delay of 3-6 h (Oteros et al. 2015; Hund 2019; Pollenflug 2019). The remaining two solutions are PolenSense (https://pollensense.com/) used in North America and Swisens Poleno (https://swisens.ch/) used in Switzerland, but contrary to information from their web sites, there are no available online resources showing operational full season real-time measurements.

The aim of this paper is to describe a solution to automate pollen information service, which overcomes limitations associated with previously mentioned solutions. This is achieved by developing the RealForAll integrated information system for automatic detection of various pollen types and real-time data delivery to end-users (RealForAll 2019). This is the first solution which integrates Rapid-E particle monitor, a state-of-the-art technology in the field of automatic pollen monitoring (Plair 2019), and artificial intelligence (AI) techniques in order to classify full spectra of allergy relevant airborne pollen types. Classified data are transferred to a subsystem responsible for storing and delivering airborne pollen concentration to end-users through web and mobile applications in real-time.

The Rapid-E particle monitor is based on laser-induced fluorescence technique and this is the first study which evaluates its performance in automatic real-time pollen monitoring comparable with the requirements given in EN 16868:2019. So far only proof of concept study was conducted (Sauliene et al. 2019) which set the basis for this full season performance evaluation on larger spectra of pollen types in an operational environment.

Since automatisation is a prerequisite for fostering the mobile health concept in allergy care (Matricardi et al. 2019) the results of this work are expected to support an ongoing rapid change of pollen monitoring (Buters, Schmidt-Weber, and Oteros 2018).

2. Related work

Automation of a process related to airborne pollen detection may be positioned in the domain of environmental informatics. Environmental informatics applies computer science disciplines to environmental information processing (Hilty et al. 1995). Environmental data has a complex nature and its processing requires the application of advanced information technologies like machine learning, deep learning, data analysis and data mining. There are a lot of recent examples of AI techniques application in solving environmental problems (Wang et al. 2019; McGovern et al. 2017; Manogaran and Lopez 2018), which serves developing integrated information systems for the monitoring and management of environmental data (Fang et al. 2014; Fang et al. 2015).

There is a notable interest in a real-time detection of bioaerosols which is extensively overviewed by Huffman et. al (2019). The authors gave an overview of major techniques and devices for real-time airborne particle detection. They emphasized laser-induced fluorescence as the most promising technology for automatic detection of bioaerosols of interest including different allergenic pollen types. This technique uses monochromatic light to trigger scattering and fluorescence which are then detected to analyse chemical composition, size and morphology of individual particles. Compared to simple particle counters, the laser-induced fluorescence approach is more suitable for real-time pollen monitoring where identification of diverse pollen species is required since they provide diversity of data needed for precise classification of bioaerosols. Rapid-E pollen monitor integrated in the RealForAll system records both scattering and fluorescence characteristics for each sampled airborne particle (Kiselev, Bonacina, and

Wolf 2013). Result of that recording is a complex dataset that requires usage of advanced machine learning tools for identification differences between different pollen types.

The resulting measurements from bioaerosol monitors are further analysed for the purpose of discriminating between bacteria, fungal spores and pollen and different AI based strategies are tested (Crawford et al., 2015; Pan, Huang, and Chang, 2012 ; Robinson et al., 2013; Ruske et al., 2017; Ruske et al., 2018; Swanson and Huffman, 2020). It was shown that for the task of airborne particle classification, clustering in general performs slightly worse than the supervised learning methods (Ruske et al. 2017). Same authors also noted that use of neural networks may improve accuracy of classification. Šauliene et al. (2019) used three different architectures of convolutional neural networks to analyse scattering and the fluorescence properties for each particle reaching the Rapid-E device. That was the first analysis of the pollen monitoring capabilities of the Rapid-E pollen monitor. They found that the Rapid-E has the potential to identify pollen types in real time but it is necessary to improve classification algorithms to include more pollen types. Recently, Sauvageat et al. (2020) conducted research on utilizing convolutional neural networks to classify data from Swisens Poleno device monitor. They applied digital holography technique on fluorescence data to reconstruct images of airborne particles. These images are further processed by the neural network and they succeeded to identify up to ten different pollen species.

However, a good classification is just one step toward the automation of the whole process of airborne pollen monitoring. None of the results from previously mentioned studies haven't been yet implemented in an operational environment for real-time pollen monitoring. In order to develop a fully operational solution, it is necessary to incorporate all activities of that process in an integrated information system, starting from raw data preprocessing to presenting real-time measurements in a user-friendly manner.

3. RealForAll system

RealForAll system is an integrated system for real-time monitoring of airborne allergens and dissemination of information about their concentration.

The system has been developing since 2018 and has already monitored one full pollen season. The system currently provides pollen measurements from two locations: Novi Sad, Serbia and Osijek, Croatia. By implementation of this system, the whole process of pollen detection is successfully automated.

The software architecture of this system is presented in figure 1. The system consists of several subsystems where each subsystem has a particular role. Role of the Rapid-E device, which is represented as a component in figure 1, is to collect airborne particles and generate raw optical data. These data are further processed by the AI-enabled subsystem for classification in order to detect different types of particles (component *Data classification*). Classified data are sent to the *RealForAllHub* subsystem whose role is to store classified data and to transform it in an appropriate format for the end-user applications. Detailed description of these subsystems is given below.

3.1 Rapid-E

The Rapid-E device is an airborne particle monitoring station. It is designed for automatic and real-time analysis of single particles suspended in air. The device aspirates ambient air with suspended particles that interact with the laser light sources (Šauliene et al. 2019) resulting in scattered light and fluorescence that are combined for characterizing each particle (Kiselev, Bonacina, and Wolf 2013). The scattered photons are captured from different angles by 24 time-resolving detectors (Kiselev, 2019), resulting in an image which size depends on the particle's morphology (i.e. size and shape). Chemical characteristics of detected particles are represented by their emission spectrum and fluorescence lifetime. After excitation by the deep-UV laser (337 nm) emitted fluorescence is recorded at 32 measuring channels within a spectral range of 350–800 nm and eight sequential acquisitions/bands with 500 ns retention. In addition, the rate of decrease of the fluorescence intensity (fluorescence lifetime) after double excitation by a laser beam is recorded at four spectral bands (350-400, 420-460, 511-572, 672-800) and 2 ns temporal resolution (Kiselev and Kiseleva, 2019).

Rapid-E device provides a JSON file containing scattered light, fluorescence spectrum and lifetime properties of each particle sampled in a minute. The device has a LAN connector and provides a secure shell that can be used to access the data in real-time.

The RealForAll system currently incorporates two Rapid-E devices. One is installed in Novi Sad, Serbia and the other is in Osijek, Croatia. The devices are connected to a local network of institutions hosting those devices which provides a stable connection between devices and the subsystem for classification. Both devices are operational and generated data in real-time during pollen season, from February to October 2019.

3.2 Subsystem for classification

The AI-enabled subsystem for classification continuously retrieves raw data from Rapid-E and performs the classification of pollen particles. The subsystem detects and counts particles larger than 8 microns and identifies several different pollen types. Classification is based on artificial neural networks implemented in Python using PyTorch for neural network implementation (details are given in Chapter 4).

The subsystem classifies minute measurements in real-time. Classification is performed with a latency of a few minutes to ensure that raw data have been retrieved from the device. As an output of the classification, the subsystem generates JSON documents for each measurement device and time-related sample. The document contains a device's identifier, time of measurement, and measured values for each classified pollen type. JSON documents are sent to the *RealForAllHub* subsystem to be stored and further processed.

3.3 RealForAllHub subsystem

This subsystem is designed to store and maintain classified data and it is implemented using Java EE technologies. Classified data are stored in PostgreSQL relational database (PostgreSQL 2019). The subsystem provides REST service for importing data into the database (*ImportService* component in figure 1). This service is used by the subsystem for classification but any other institution with real-time pollen measurements can be joined easily. The only restriction imposed by technology is to provide a continuous flow of classified data in the format described by this REST service. Those measurements will be accessible through our end-user applications.

The end-user applications show hourly pollen concentrations, but the *RealForAllHub* subsystem receives minute measurements. This requires aggregation of measurement data and it is done on each hour but postponed by several minutes to ensure that all data for a given hour are received (*Aggregation* component in figure 1). Aggregation calculates the average hourly value from minute values within the last hour. Those aggregated data are also stored in the database. There is a configuration in the system regarding how many minute measurements are expected to be received during an hour. A notification email is sent to the system admin in the case that some measurements are missing (*Notification* component in figure 1).

This subsystem provides REST services to end-user applications. Mobile and web applications use the *AppService* component (figure 1) to obtain and visualize data about pollen concentrations. Also, there is the *AdminService* component (figure 1) used by the web application for system administration.

3.4 End-user applications

The main aim of the RealForAll system is to disseminate information about pollen concentration. Appropriate Android and iOS mobile applications, as well as the web application, have been developed for that purpose (Android app 2019; iOS app 2019; Web app 2019).

Mobile applications show real-time pollen measurements from available Rapid-E devices as well as hourly averages for a selected device (figure 2). Presented measurements can be filtered by pollen types and compared to measurements from other devices. The applications also provide a forecast of pollen distribution over Europe generated by SILAM (SILAM 2019). In addition, the applications may be used to keep personal allergy symptoms diary in order to find a correlation between recorded symptoms and airborne pollen measurements. Information from this diary may be useful in the evaluation of prescribed treatments and for better management of allergic diseases. The web application for end-users has fewer features than mobile applications and it only provides measured hourly average concentrations and the forecast.

The RealForAll system also has a web application for system administration. It is not exposed publicly and only authorised users have access. The application allows the management of the RealForAll system (i.e. adding new pollen types and devices as well as configuration of some system properties). It also allows the export of minute and hourly classifications for a selected period.

4. Classification

The output of the Rapid-E device is a JSON file containing scattered light, fluorescence spectrum and lifetime of fluorescence signals for every particle sampled in a minute. The detailed description of the output files structure is given in the earlier pilot study (Šauliene et al. 2019). The character of the measurements (i.e involves a temporal component for scattering light and fluorescence and multiple wavelength bands for lifetime of fluorescence) allows transferring light intensity signals into two dimensional image format suitable for analysis using Convolutional Neural Network. This section provides details regarding classification methodology used in the RealForAll system.

4.1 Data collection

Labeled data for training the classifier are obtained in calibration events, where the domain expert is exposing the Rapid-E device with collected aerosol samples in a controlled environment. Each calibration resulted in JSON files, belonging to the same aerosol class. Calibration was performed for 24 the most common pollen classes (Acer, Alnus, Ambrosia, Artemisia, Betula, Broussonetia, Carpinus, Corylus, Fraxinus excelsior, Fraxinus ornus, Juglans, Morus, Other pollen, Pinaceae, Plantago, Platanus, Poaceae, Populus, Quercus, Salix, Taxaceae, Tilia, Ulmus, Urticaceae). In addition, real-time measurements at a time when there was no pollen in the air are labeled as “other” and “starch” and used in training the classifier in order to prevent mixing other bioaerosols (i.e. fungal spores) and starch with pollen.

4.2 Data preprocessing

Laser-induced data tend to be noisy and using them in their raw form can result in poor generalization. To avoid this, scattered light images are centred with respect to the time axis around the mean of indices with maximum values over 24 angle pixels and then cut or padded with zeros to fit the size of 20x120 (4 boundary angle pixels are removed due to device dependence). Fluorescence spectrum and lifetime signals are normalized into 0-1 range. The signals of scattered light and fluorescence spectrum are smoothed with the Savitzky–Golay filter (Savitzky and Golay 1964) to additionally reduce the noise. Fluorescence spectrum signals were converted into a 4x32 pixels image by stacking second to fourth acquisitions/bends. Similarly, fluorescence lifetime signals were converted into a 4x24 pixels image. Particle size approximation calculated from the scattered light image and lifetime weights calculated from the fluorescence lifetime signal were also added as features to the classification model.

To ensure only high-quality records are analyzed, detected particles with the scattering image width larger than 450 pixels, at least one of four maximal spectral peaks lower than 408 nm or larger than 495 nm, maximum spectral intensity less than 2500 and lifetime maximum peak not detected between 20 ns and 88 ns are filtered out.

The processed data contains 103593 pollen samples and 6285 samples from real-time measurements. The data is split into train and test datasets, where the train set contains 90% of samples from each class while the remaining events were used for testing.

4.3 Neural network architecture

The classification algorithm is based on convolutional neural networks (CNN), which have so far shown better performance on similar problems in image processing and object classification compared to other machine learning classifiers (Krizhevsky, Sutskever, and Hinton 2012). CNN allows automatic feature extraction, which is crucial when dealing with complex homogeneous data, as well as combining multiple inputs from the Rapid-E device to perform classification. The network is processing each data-type individually using the combination of 2-D convolutions, ReLU activations, batch normalization, max pooling and dropout, considered as a convolutional block (figure 3). By doing so, it learns the most important features and reduces the dimensionality of data provided by Rapid-E. The features are first equalized by passing each of them to one fully connected layer of the same size, so that each input has the same contribution to the feature vector, and then concatenated, together with the additional features of size and lifetime weights and are passed to the fully connected layer consisting of 26 nodes since there are 26 aerosol classes for identification, after which the classification is performed with the log-softmax activation function. In this way the network architecture allows the gradient to flow through the whole network, updating the weights for each distinct source, based on the joint decision. The cost function used for training the network is negative log-likelihood loss and the updater is the stochastic gradient descent with a learning rate of 0.001 and a momentum of 0.9. We created batches used for training the model in such a way that each batch contains the same number of samples from each class and thus resolved the unbalanced dataset problem. The detailed description of the network is given in Šauliene et al. (2019).

4.4 Classification results

On the test dataset, the model yields an accuracy of 65.3% on 26 classes (Figure 4). The precision, recall and F1 score of the model are 59%, 69% and 61%, respectively. The number of classes involved in the test is rather high making the task unrealistic for real life monitoring. Therefore the real performance is evaluated by comparison to standard monitoring of airborne pollen (EN 16868:2019) performed in Novi Sad in 2019. In order to neutralize losses from data preprocessing, RealForAll data were multiplied by a scaling factor (SF) corresponding to the relationship between quantity measured by standard method EN 16868:2019 and quantity obtained by RealForAll system. The performance was tested by analyzing Pearson correlation coefficients (R) between average daily pollen concentrations measured by two systems while focusing on the periods when standard method detects pollen of interest. Good performance (R > 0.7) was confirmed for 11 pollen types (Figure 5) while the rest classifications underperformed (Figure 6). Pearson's correlation coefficients (R) and scaling factor (SF) are given on both figures. For both good and underperforming classifications there is a notable amount of false positive detections that are eliminated by manual limitation of the pollen season. Apart from the further improvement of the classification model by introducing the shortcut connections between network layers (He et al. 2016) and increasing the width of these networks (Zagoruyko and Komodakis 2016) for better feature extraction, the next developments will strive to automate the limitation of the season by introducing confidence thresholds for the classifications under which the model will not deliver data to the *RealForAllHub* subsystem. Future development of the classification should involve additional separation systems for the mixing classes since some classes are very well separated while some are not (Figure 4). It should be noted that for nearly all underperformed classifications, signals are characterized by low intensity. This is also characteristic for the standard method EN 16868:2019, but this is even more augmented in automatic classification by strict filtering which ensures that only high quality detections are analyzed but it decreases the detection limit of the method.

5. Discussion

By implementing the RealForAll system in a production environment, we succeeded in the automation of the pollen detection process and the dissemination of real-time measurements. Currently, our mobile applications are installed on more than 1700 mobile devices. The main advantages of this system in comparison to the standard method EN 16868:2019 reflect in its extensibility, real-time data delivery and independence of aerobiology experts during the pollen season.

RealForAll system can, with relative ease, be adapted to a wider user base. Namely, introducing new Rapid-E devices in the RealForAll system would not have a significant impact on overall system performance. In the case of the standard method EN 16868:2019, adding new Hirst devices require additional manual effort directly influencing operational time and cost. Also, it enables interoperability with other systems for pollen detection. They can easily send and store their measurements to the RealForAll system with the advantage that their data will be efficiently disseminated through RealForAll end-user applications.

The irreplaceable step in the standard method EN 16868:2019 is the manual classification of data performed by aerobiology experts. Operation of the RealForAll system doesn't require this kind of expertise because the process of classification is automated by the AI module.

Finally, the significant difference between those two approaches is the time needed to provide relevant pollen measurements. Table 1 shows the approximate duration of activities carried out in order to get hourly pollen measurements using the standard method and the RealForAll system, respectively. The Rapid-E device is sampling in real-time so the sample characteristics are measured for every minute and the hourly sample is available already after 60 minutes of measurements. In the case of the Hirst device, it is not economical to process samples every hour and because of that sample is usually available after either 24 hours sampling or more common after one week. A 24-hour sample obtained from the Hirst device requires a minimum 2 hours for preprocessing and classification while in the case of the RealForAll system that activity is done in 15 minutes. Taking everything into consideration, we can conclude that the

RealForAll system can disseminate pollen information at least 20 times faster than the standard method.

Although the RealForAll system shows great results in performance and operability, there is still some room for improvement. Analyzing values from table 1, it can be seen that the RealForAll system has latency in data provision but it is not due to the long-lasting classification process as it may seem. This delay is a consequence of batch data processing. The subsystem for classification downloads data from the Rapid-E device at periodic intervals. The classification process is not aware of whether the download is complete and because of that, it is postponed for 10 minutes to provide enough time for finishing the download process. However, this does not guarantee that all data will be processed. Size of Rapid-E minute recordings can vary from 10 to 100Mb and those files may not be transferred in 10 minutes in the case of poor Internet connection. This problem can be solved by implementing streaming data processing. In that manner, we would have a continuous flow of raw data from Rapid-E devices and data will be immediately classified as they arrive without any delay. However, the latency of several minutes is still inevitable to ensure that the RealForAllHub subsystem has received most of the classified data before it performs aggregation.

6. Conclusion

In this paper, we introduce an integrated system for real-time monitoring of airborne allergens and the dissemination of information about their concentration. This system automatises the standard method EN 16868:2019 for pollen detection. The system provides hourly measurements with a latency of 15 minutes which is a significant improvement in comparison with the standard method. Also, the system provides easy integration of new devices as well as pollen measurements from other systems, which brings an advantage to application users who get a single point of access to real-time measurements from different locations.

Biological contaminants pose severe threats to the manufacturing processes of numerous industrial, food and pharmaceutical products. Additionally, microorganisms such as fungi, bacteria, and viruses can cause significant damage to workers' health and plant health in agriculture. The introduction of automatic bioaerosols monitoring in industrial enterprises is expected to minimize negative occupational health effects and maximize profit. The discrimination of bioaerosols is often a prerequisite for successful implementation of mitigation measures. For example, for successful allergy management, it is not sufficient to know total pollen concentrations but the quantity of each allergen in the atmosphere so sensitive individuals could be selectively warned. Similarly, the presence of only specific fungal spores should be a trigger for fungicide spraying in glass houses indicating that discrimination of bioaerosols has more value than information on their bulk quantity in production enterprises.

The RealForAll system proved that AI enables automation for monitoring of airborne allergens which is part of routine environmental monitoring in about 700 stations worldwide (Map 2019). This opens possibilities for the application of aerobiology in a variety of industries in particular relation to human, animal and plant health. Despite the fact that further improvement of classification models is needed to enable identification full spectra of bioaerosols suspended in the atmosphere, the RealForAll system is an example of how automation of tedious manual process that requires a notable amount of domain expertise (i.e. identification of pollen) is supported by advanced laser-induced fluorescence measurements and AI. Systems for real-time pollen identification are still in their ongoing phase of development and a lot of effort should be made especially regarding classification accuracy. Opening raw data from pollen monitors worldwide as well as making classification models publicly available to other researchers would be very beneficial in order to get scientific feedback and speed up further research in this field.

To sum up, comparing to the Hirst method, the main drawback of implementing RealForAll system reflects in its initial investment cost but on the other side, it provides real-time pollen measurements which help allergic people to better manage their allergic disease and are essential for the improvement of forecasting models (Sofiev 2019).

Acknowledgement: RealForAll project (2017HR-RS151) co-financed by the Interreg IPA Cross-border Cooperation programme Croatia – Serbia 2014-2020 and Provincial secretariat for Science, Autonomous Province Vojvodina, Republic of Serbia (contract no. 102-401-337/2017-02-4-35-8) and the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grant No. 451-03-68/2020-14/ 200358).

References

Akdis, Cezmi A., Peter W. Hellings, and Ioana Agache, eds. *Global atlas of allergic rhinitis and chronic rhinosinusitis*. European Academy of Allergy and Clinical Immunology, 2015.

Android app. 2019. “RealForAll.“ Accessed November 2019. https://play.google.com/store/apps/details?id=com.realforall

Bousquet, Jean, Peter W. Hellings, Ioana Agache, Flore Amat, Isabella Annesi-Maesano, Ignacio J. Ansotegui, Josep M. Anto et al. "Allergic Rhinitis and its Impact on Asthma (ARIA) Phase 4 (2018): Change management in allergic rhinitis and asthma multimorbidity using mobile technology." *Journal of Allergy and Clinical Immunology* 143, no. 3 (2019): 864-879. https://doi.org/10.1016/j.jaci.2018.08.049

Buters, Jeroen TM, C. Antunes, Ana Galveias, Karl C. Bergmann, Michel Thibaudon, Carmen Galán, Carsten Schmidt-Weber, and Jose Oteros. "Pollen and spore monitoring in the world." *Clinical and translational allergy* 8, no. 1 (2018): 9. https://doi.org/10.1186/s13601-018-0197-8

Buters, Jeroen, Carsten Schmidt-Weber, and Jose Oteros. "Next-generation pollen monitoring and dissemination." *Allergy* 73, no. 10 (2018): 1944-1945. https://doi.org/10.1111/all.13585

Crawford, I., S. Ruske, D. O. Topping, and M. W. Gallagher. "Evaluation of hierarchical agglomerative cluster analysis methods for discrimination of primary biological aerosol." *Atmospheric Measurement Techniques* 8, no. 11 (2015): 4979. https://doi.org/10.5194/amt-8-4979-2015

Crouzy, Benoît, Michelle Stella, Thomas Konzelmann, Bertrand Calpini, and Bernard Clot. "All-optical automatic pollen identification: towards an operational system." *Atmospheric environment* 140 (2016): 202-212. https://doi.org/10.1016/j.atmosenv.2016.05.062

Fang, Shifeng, Li Da Xu, Yunqiang Zhu, Jiaerheng Ahati, Huan Pei, Jianwu Yan, and Zhihui Liu. "An integrated system for regional environmental monitoring and management based on internet of things." *IEEE Transactions on Industrial Informatics* 10, no. 2 (2014): 1596-1605. https://doi.org/10.1109/TII.2014.2302638

Fang, Shifeng, Lida Xu, Yunqiang Zhu, Yongqiang Liu, Zhihui Liu, Huan Pei, Jianwu Yan, and Huifang Zhang. "An integrated information system for snowmelt flood early-warning based on internet of things." *Information Systems Frontiers* 17, no. 2 (2015): 321-335. https://doi.org/10.1007/s10796-013-9466-1

Galán, C., Matt Smith, M. Thibaudon, G. Frenguelli, J. Oteros, R. Gehrig, U. Berger, B. Clot, R. Brandao, and EAS QC working group. "Pollen monitoring: minimum requirements and reproducibility of analysis." *Aerobiologia* 30, no. 4 (2014): 385-395. https://doi.org/10.1007/s10453-014-9335-5

Garzia-Mozo, H. "The use of aerobiological data on agronomical studies." *Annals of Agricultural and Environmental Medicine* 18, no. 1 (2011).

Hanakosan. 2019. "Hanakosan". Accessed November 2019. http://kafun.taiki.go.jp/

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.

Hilty, L. M., Bernd Page, F. J. Radermacher, and W-F. Riekert. "Environmental informatics as a new discipline of applied computer science." In *Environmental Informatics*, pp. 1-11. Springer, Dordrecht, 1995. https://doi.org/10.1007/978-94-017-1443-3_1

Hirst, J.M. "An automatic volumetric spore trap." *Annals of applied Biology* 39, no. 2 (1952): 257-265. https://doi.org/10.1111/j.1744-7348.1952.tb00904.x

Huffman, J. Alex, Anne E. Perring, Nicole J. Savage, Bernard Clot, Benoît Crouzy, Fiona Tummon, Ofir Shoshanim et al. "Real-time sensing of bioaerosols: Review and current perspectives." *Aerosol Science and Technology* (2019): 1-31. https://doi.org/10.1080/02786826.2019.1664724

Hund. 2019. "Pollen monitor." Accessed November 2019. https://www.hund.de/en/service/pollen-monitor.html

iOS app. 2019. "RealForAll." Accessed November 2019. https://apps.apple.com/us/app/realforall/id1441311324

Kawashima, S., Thibaudon, M., Matsuda, S., Fujita, T., Lemonis, N., Clot, B. and Oliver, G., 2017. Automated pollen monitoring system using laser optics for observing seasonal changes in the concentration of total airborne pollen. *Aerobiologia*, *33*(3), pp.351-362. https://doi.org/10.1007/s10453-017-9474-6

Kiselev, Denis. "Method and device for detection and/or morphologic analysis of individual fluid-borne particles." U.S. Patent Application 16/072,548, filed January 31, 2019. http://www.freepatentsonline.com/y2019/0033191.html

Kiselev, Denis, and Svetlana Kiseleva. "Device and method for detecting and/or characterizing fluid-borne particles." U.S. Patent Application 16/310,275, filed October 31, 2019. http://www.freepatentsonline.com/y2019/0331601.html

Kiselev, Denis, Luigi Bonacina, and Jean-Pierre Wolf. "A flash-lamp based device for fluorescence detection and identification of individual pollen grains." *Review of Scientific Instruments* 84, no. 3 (2013): 033302. https://doi.org/10.1063/1.4793792

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In *Advances in neural information processing systems*, pp. 1097-1105. 2012.

Manogaran, Gunasekaran, and Daphne Lopez. "Disease surveillance system for big climate data processing and dengue transmission." In *Climate Change and Environmental Concerns: Breakthroughs in Research and Practice*, pp. 427-446. IGI Global, 2018. https://doi.org/10.4018/978-1-5225-5487-5.ch022

Map. 2019. "Worldwide Map of Pollen Monitoring Stations." Accessed November 2019. https://www.eaaci.org/19-activities/task-forces/4342-pollen-monitoring-stations-of-the-world.html

Matricardi, Paolo Maria, Stephanie Dramburg, Alberto Alvarez-Perea, Darío Antolín-Amérigo, Christian Apfelbacher, Marina Atanaskovic-Markovic, Uwe Berger et al. "The Role of Mobile Health Technologies in Allergy Care: an EAACI Position Paper." *Allergy* (2019). https://doi.org/10.1111/all.13953

McGovern, Amy, Kimberly L. Elmore, David John Gagne, Sue Ellen Haupt, Christopher D. Karstens, Ryan Lagerquist, Travis Smith, and John K. Williams. "Using artificial intelligence to improve real-time decision-making for high-impact weather." *Bulletin of the American Meteorological Society* 98, no. 10 (2017): 2073-2090. https://doi.org/10.1175/BAMS-D-16-0123.1

Norkko. 2019. "Situation." Accessed November 2019. http://www.norkko.fi/?lang=en

Oteros, Jose, Gudrun Pusch, Ingrid Weichenmeier, Ulrich Heimann, Rouven Möller, Stefani Röseler, Claudia Traidl-Hoffmann, Carsten Schmidt-Weber, and Jeroen TM Buters. "Automatic and online pollen monitoring." *International archives of allergy and immunology* 167, no. 3 (2015): 158-166. https://doi.org/10.1159/000436968

Pan, Yong-Le, Hermes Huang, and Richard K. Chang. "Clustered and integrated fluorescence spectra from single atmospheric aerosol particles excited by a 263- and 351-nm laser at New Haven, CT, and Adelphi, MD." *Journal of Quantitative Spectroscopy and Radiative Transfer* 113, no. 17 (2012): 2213-2221. https://doi.org/10.1016/j.jqsrt.2012.07.028

Pasyfo. 2019. "Personal Allergy Symptom Forecasting System." Accessed November 2019. http://pasyfo.lt/en/

Plair. 2019. "Instruments for Precise Environmental Monitoring." Accessed November 2019. http://www.plair.ch/

Poleninfo. 2019. "Pollen load map of Europe." Accessed November 2019. https://www.polleninfo.org/

Pollenflug. 2019. "Pollenflug aktuell." Accessed November 2019. https://epin.lgl.bayern.de/pollenflug-aktuell

PostgreSQL. 2019. "PostgreSQL: The World's Most Advanced Open Source Relational Database." Accessed November 2019. https://www.postgresql.org/

RealForAll. 2019. "What do we breathe?." Accessed November 2019. https://www.realforall.com/language/en/welcome/

Robinson, Niall H., J. D. Allan, J. A. Huffman, Paul H. Kaye, V. E. Foot, and M. W. Gallagher. "Cluster analysis of WIBS single-particle bioaerosol data."

Atmospheric Measurement Techniques (2013). https://doi.org/10.5194/amt-6-337-2013

Ruske, Simon, David O. Topping, Virginia E. Foot, Paul Kaye, Warren Stanley, I. P. Crawford, Andrew Morse, and Martin W. Gallagher. "Evaluation of machine learning algorithms for classification of primary biological aerosol using a new UV-LIF spectrometer." *Atmospheric Measurement Techniques* (2017). https://doi.org/10.5194/amt-10-695-2017

Ruske, Simon, David O. Topping, Virginia E. Foot, Andrew P. Morse, and Martin W. Gallagher. "Machine learning for improved data analysis of biological aerosol using the WIBS." *Atmospheric Measurement Techniques* 11, no. 11 (2018): 6203-6230. https://doi.org/10.5194/amt-11-6203-2018

Šaulienė, Ingrida, Laura Šukienė, Gintautas Daunys, Gediminas Valiulis, Lukas Vaitkevičius, Predrag Matavulj, Sanja Brdar et al. "Automatic pollen recognition with the Rapid-E particle counter: the first-level procedure, experience and next steps." *Atmospheric Measurement Techniques* 12, no. 6 (2019): 3435-3452. https://doi.org/10.5194/amt-12-3435-2019

Sauvageat, Eric, Yanick Zeder, Kevin Auderset, Bertrand Calpini, Bernard Clot, Benoît Crouzy, Thomas Konzelmann, Gian Lieberherr, Fiona Tummon, and Konstantina Vasilatou. "Real-time pollen monitoring using digital holography." *Atmospheric Measurement Techniques* (2020). https://doi.org/10.5194/amt-2019-427

Savitzky, Abraham, and Marcel JE Golay. "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry* 36, no. 8 (1964): 1627-1639. https://doi.org/10.1021/ac60214a047

SILAM. 2019. “System for Integrated modeLling of Atmospheric coMposition.” Accessed November 2019. http://silam.fmi.fi/index.html

Sofiev, Mikhail, Olga Ritenberga, Roberto Albertini, Joaquim Arteta, Jordina Belmonte, Carmi Geller Bernstein, Maira Bonini et al. "Multi-model ensemble simulations of olive pollen distribution in Europe in 2014: current status and outlook." *Atmospheric Chemistry and Physics* 17, no. 20 (2017): 12341-12360. https://doi.org/10.5194/acp-17-12341-2017

Sofiev, Mikhail. "On possibilities of assimilation of near-real-time pollen data by atmospheric composition models." *Aerobiologia* 35, no. 3 (2019): 523-531. https://doi.org/10.1007/s10453-019-09583-1

Swanson, Benjamin E., and J. Alex Huffman. "Pollen clustering strategies using a newly developed single-particle fluorescence spectrometer." *Aerosol Science and Technology* (2020): 1-33. https://doi.org/10.1080/02786826.2019.1711357

Wang, Puze, Jiping Yao, Guoqiang Wang, Fanghua Hao, Sangam Shrestha, Baolin Xue, Gang Xie, and Yanbo Peng. "Exploring the application of artificial intelligence technology for identification of water pollution characteristics and tracing the source of water quality pollutants." *Science of The Total Environment* 693 (2019): 133440. https://doi.org/10.1016/j.scitotenv.2019.07.246

Web app. 2019. "Measurements." Accessed November 2019. https://www.realforall.com/language/en/measurements/

Zagoruyko, Sergey, and Nikos Komodakis. "Wide residual networks." *arXiv preprint arXiv:1605.07146* (2016).

Table 1. Comparison of activity's duration for standard method EN 16868:2019 and RealForAll system when delivering concentrations of airborne pollen at 1h resolution

Analyzed features		Standard method EN 16868:2019	RealForAll system
Sample is available		24-168h	60 minutes
Identification bioaerosols	Sample preprocessing	1 h	/
	Classification	1 h	15 minutes
Overall time		**26-170h**	**75 minutes**

Figure 1. System architecture

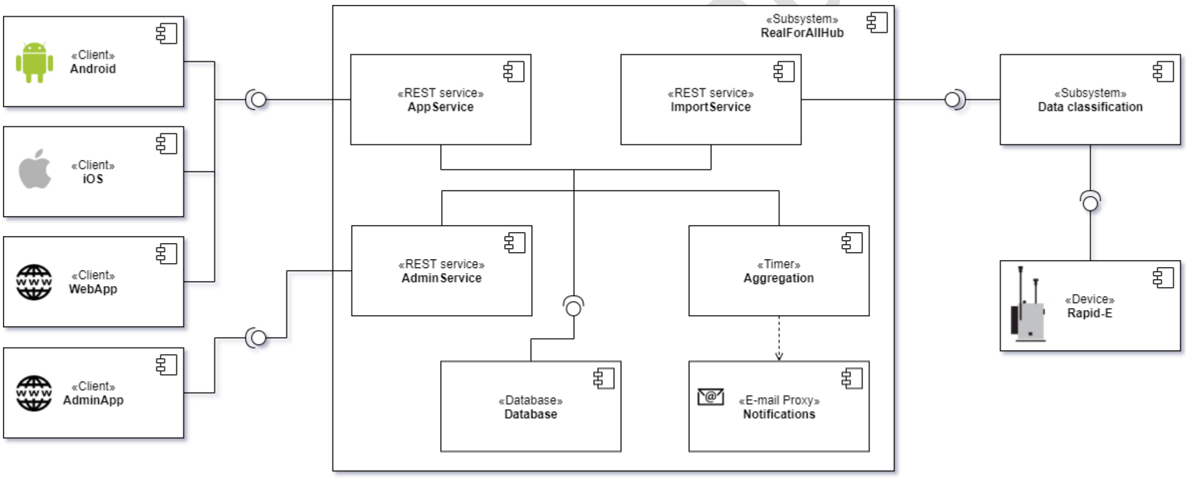


Figure 2. Pages of the RealForAll android application that disseminate real-time measurements

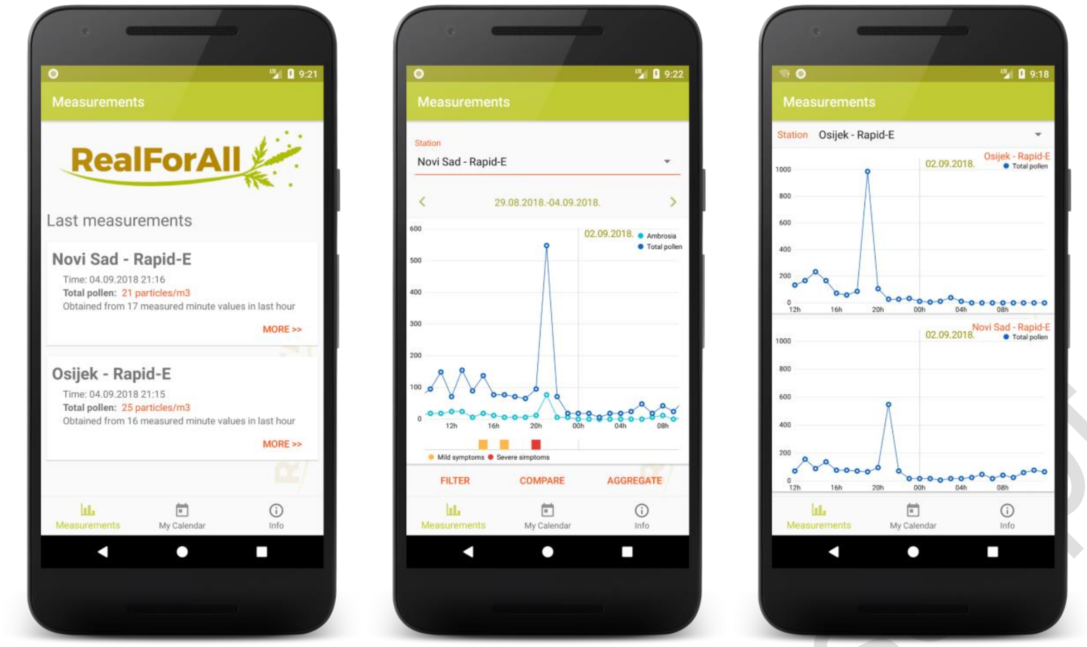


Figure 3. Architecture of neural network used for operational classification of pollen in 2019

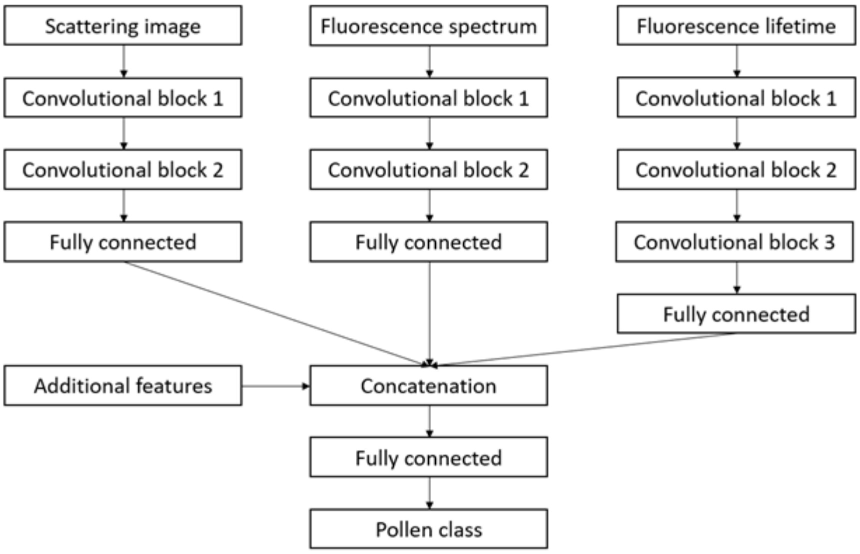


Figure 4. Confusion matrix for the classification model

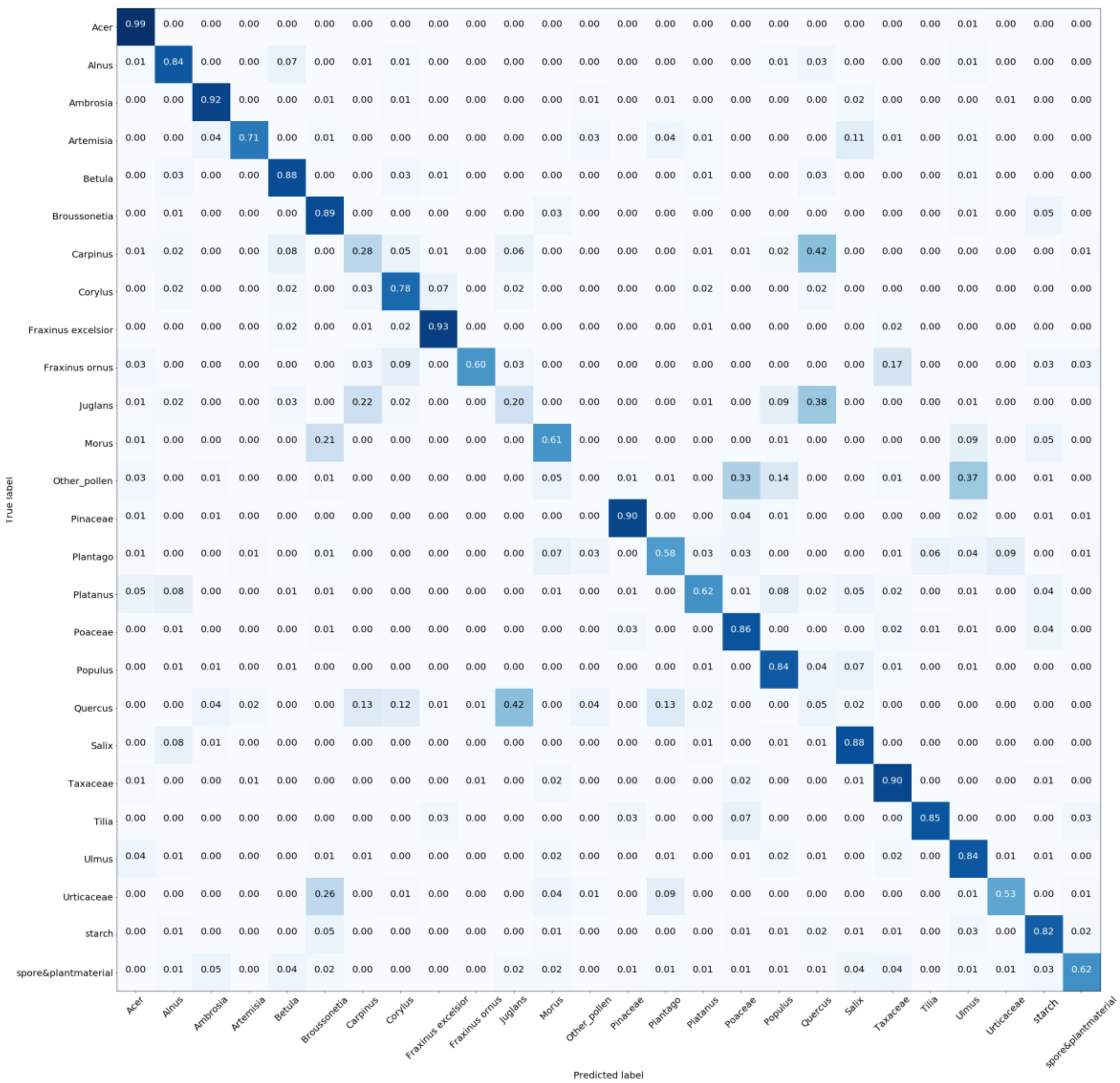


Figure 5. Good performance of the RealForAll system in comparison to standard method EN 16868:2019 measurements in Novi Sad during 2019 (*Fraxinus* represents the sum of *Fraxinus ornus* and *Fraxinus excelsior* that were classified separately)

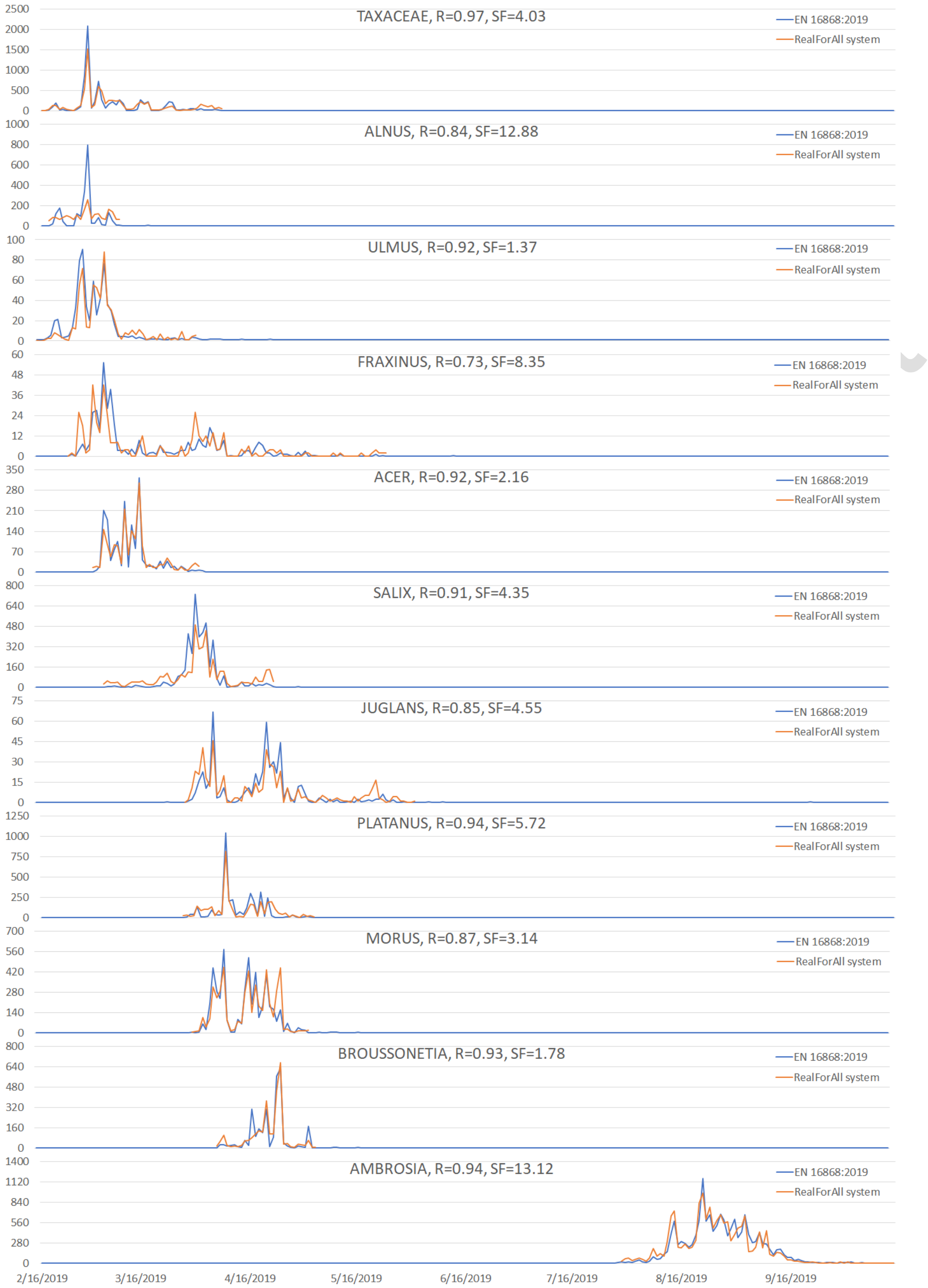
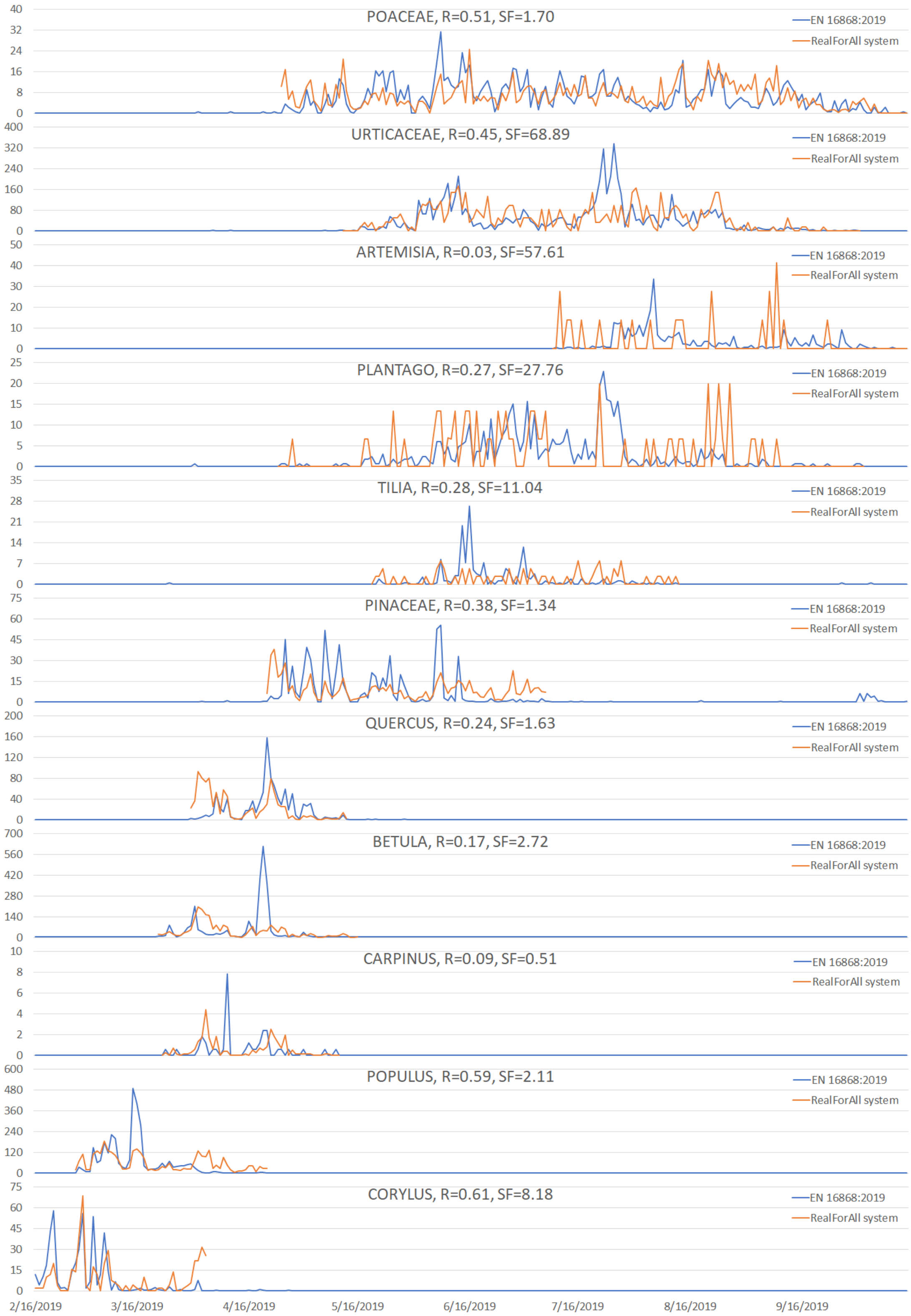


Figure 6. Underperformance of the RealForAll system in comparison to standard method EN 16868:2019 measurements in Novi Sad during 2019 (*Urticaceae* includes *Parietaria*)



POACEAE, R=0.51, SF=1.70
URTICACEAE, R=0.45, SF=68.89
ARTEMISIA, R=0.03, SF=57.61
PLANTAGO, R=0.27, SF=27.76
TILIA, R=0.28, SF=11.04
PINACEAE, R=0.38, SF=1.34
QUERCUS, R=0.24, SF=1.63
BETULA, R=0.17, SF=2.72
CARPINUS, R=0.09, SF=0.51
POPULUS, R=0.59, SF=2.11
CORYLUS, R=0.61, SF=8.18
EN 16868:2019
RealForAll system
2/16/2019
3/16/2019
4/16/2019
5/16/2019
6/16/2019
7/16/2019
8/16/2019
9/16/2019