Generalising Tree–Level Sap Flow Across the European Continent

Ralf Loritz[1], Chen Huan Wu[1], Daniel Klotz[2], Martin Gauch[3], Frederik Kratzert[4] and Maoya Bassiouni[5]

[1] Institute of Water and River Basin Management, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2] Helmholtz Centre for Environmental Research – UFZ, Department Computational of Hydrosystems, Leipzig, Germany

[3] Google Research, Zurich, Switzerland

[4] Google Research, Vienna, Austria

[5] Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA

Correspondence to: Ralf Loritz (Ralf.Loritz@kit.edu)

Key Points:

- LSTMs demonstrate their ability in predicting hourly sap flow for diverse trees and climates across Europe.
- Training on extensive datasets enhances LSTM's ability to simulate sap flow accurately in both seen and unseen environments.
- The study underscores the value of large-scale deep learning models and advocates for the expansion of datasets like SAPFLUXNET.

Abstract.

Sap flow observations provide a basis for estimating transpiration and understanding forest water use dynamics and plant-climate interactions. This study developed a continental modeling approach using Long Short-Term Memory networks (LSTMs) to predict hourly tree-level sap flow across Europe based on the SAPFLUXNET database. We developed models with varying levels of training sets to evaluate performance in unseen conditions. The average Kling-Gupta Efficiency was 0.77 for gauged models trained on 50 % of time series across all forest stands and was 0.52 for ungauged models trained on 50 % of the forest stands. Continental models matched or exceeded performance of specialized and baseline models for all genera and forest stands, demonstrating the potential of LSTMs to generalize hourly sap flow across tree, climate, and forest types. This work highlights hence the potential of deep learning models to generalize sap flow for enhancing tree to continental ecohydrological investigations.

Plain language summary.

Transpiration is the dominant flux of water from the land surface to the atmosphere, yet it remains challenging to measure and estimate especially given different climates and tree types. This study shows how large-scale deep learning models can simulate sap flow, the movement of water within trees, with high precision, even in forests and conditions not previously studied. We show that these models excel when they learn from large and diverse datasets. The flexible design of our model training means that every new sap flow measurement can potentially be used to further optimize our networks. Our findings indicate that this approach of continuously updating the model with new information greatly improves its performance to simulate and predict tree-level sap flow. This work thus highlights the potential of deep learning models to generalize sap flow, thereby enhancing ecohydrological investigations from the tree to the continental scale.

1 Introduction

Accurate quantification of plant transpiration is a critical component in hydrological research, accounting for approximately 65 % of global terrestrial evapotranspiration (e.g. Good et al., 2015). Plants play thereby a pivotal role in controlling the exchange of water between the atmosphere and the land surface. Yet, capturing complex plant water use responses to environmental conditions and estimating transpiration across spatio-temporal scale, remains challenging. Among the limited number of available measurements for plant transpiration, in-situ sap flow sensors are the most widespread technique due to their (relative) low cost and ease of use (Dugas et al., 1993). Sap flow has long been recognized, especially in ecology and plant physiology fields, as a fundamental measurement for deciphering vegetation functionality and transpiration dynamics in both forested (Granier & Loustau, 1994) and agricultural ecosystems (Dugas et al., 1994).

While the analysis of sap flow has been less prevalent in catchment hydrology, recent studies highlight how sap flow measurements can enhance understanding of intricate relationships among vegetation characteristics, hydrometeorological factors, and catchment properties. For instance, Hassler et al. (2018) conducted an extensive study to determine the relative influence of tree-, stand-, and site-specific characteristics on sap velocity patterns, using data from 61 beech and oak trees across 24 sites in Luxembourg. Their findings suggest that transpiration estimates at the catchment scale could be significantly improved by taking into account not just hydro-meteorological drivers, but also the spatial patterns of the composition of forests in a catchment. Renner et al. (2016) showed that variability in sap flow driven by topography and aspect could be balanced out by the forest stand composition, resulting in equivalent transpiration rates across south and north facing hillslopes. This exemplifies how vegetation dynamics adapt to environmental conditions to effectively use available resources. Hoek van Dijke et al. (2019) used sap flow measurements to explore the link between normalized difference vegetation index (NDVI) and transpiration. They showed that NDVI is not always reliable for modeling transpiration, especially during drought periods, as the correlation between NDVI and sap flow can vary positively or negatively, influenced by seasonal changes, moisture availability, and hydrogeological factors. Integrating sap flow into catchment studies and understanding its spatio-temporal variability is thus key for improving our ability to estimate transpiration at landscape scales

Sap flow measurement campaigns have typically been designed to examine plant-soil interaction at plot or plant scales (e.g. Jacksich et al., 2020, Seeger et al., 2022). Nevertheless, recent modeling demonstrates avenues for incorporating individual sap flow measurements to advance our understanding of large-scale ecohydrological dynamics, for instance, by validating large scale environmental models and / or by improving catchment wide transpiration estimates (Loritz et al., 2022). Further, the SAPFLUXNET initiative was founded to combine and harmonize numerous, individual small-scale field campaigns in a global open-source sap flow database and hence overcome the spatial and temporal scarcity of sap flow data sets (Poyatos et al., 2021). The SAPFLUXNET database therefore presents unique opportunities to learn generalizable relationships across different plant genera and different climates, critical for estimating sap flow in ungauged regions at the tree level. Such relationships have yet to be encoded into data-driven models to predict sap flow regionally based on tree type, stand characteristics and climate.

Deep learning offers powerful avenues to find generalizations in large and diverse datasets (e.g. LeCun et al., 2015). For example, Koppa et al. (2022) employed a feed-forward neural network to analyze daily, global datasets from eddy covariance stations, satellites and sap flow measurements from SAPFLUXNET. Their aim was to create an accurate global vegetation stress model that improves simulations of the reduction of evaporation from its theoretical maximum, for instance during periods of water limitations. By integrating the feed-forward neural network

into an existing process-based model, they improved the ability of the process-based model to estimate global evaporation rates. Further, Loritz et al. (2022) showed that gated recurrent units (Cho et al., 2014), a form of recurrent neural network that drops old information as it ingest new information, could simulate vegetation dynamics in form of catchment-averaged sap velocities with low residuals even in areas where the model has not been trained. Similar to Koppa et al. 2022, Loritz et al. (2022) showed that the deep-learning-based sap velocity simulations can be coupled with a process-based, hydrological model, in this case with the objective to replace the semi-empirical Steward-Jarvis equation, resulting in more accurate transpiration estimates and ultimately better soil moisture simulations particularly during a drought year. In addition, Li et al. (2022), highlighted the ability of recurrent neural networks to estimate the vegetation dynamics of a specific tree species in New Zealand from standard meteorological variables. Both Loritz et al. (2022) and Li et al. (2022) found that particular recurrent neural networks like Long Short-Term Memory (LSTM; Hochreiter and Schmidhuber, 1997) models are suitable architectures to simulate and predict sap flow when they compared different networks. However, both studies trained models on relatively small and local datasets, representing only the dynamics of a forest stand or a small catchment without deciphering different tree species behavior. The extent to which recurrent neural networks can detect consistent relationships between tree-level sap flow and meteorological drivers across different species, measurement methods, climates and forest stands remains an open question.

This study aims to investigate the potential of LSTMs, a well-suited architecture for simulating environmental time series (e.g. Shen et al., 2017, Kratzert et al., 2018, 2019, Besnard et al., 2019), for modeling tree-level sap flow at an hourly time scale across the European continent. We developed continental tree-level sap flow models leveraging the SAPFLUXNET database to extract generalized relations between sap flow from different tree genera, dynamic atmospheric drivers and forest stand characteristics. We developed different experimental training setups with the aim of evaluating the performance of continental deep learning sap flow models in-sample and out-of-sample against simple baseline models and more specialized, local models. Such a deep learning approach for continental tree-level sap flow modeling offers avenues to overcome limitations of transpiration models when tree–level parameterizations for stomatal conductance are locally unavailable, could be used to assess different forest structures and their implications for regional transpiration rates, and ultimately provide robust sap flow based transpiration estimates across scales.

2 Material and Methods

2.1 The SAPFLUXNET database

The SAPFLUXNET database (Poyatos et al., 2021) represents a comprehensive global repository of tree-level sap flow measurements and their ancillary data including tree and forest characteristics and meteorological observations. We used version 0.1.5 of the database

encompassing 202 datasets providing information on 2714 individual trees. The sap flow datasets span 174 species (141 angiosperms and 33 gymnosperms), 95 genera, and 45 families and covers the temporal period 1995-2018. The measurement periods for individual trees within the database extends from a minimum of 3 months up to a maximum of 16 years.

2.1.1 A European subset of the SAPFLUXNET database

We selected a European subset of the SAPFLUXNET dataset, comprising 64 forest stands out of 202 stands in the global datasets, encompassing 738 individual trees. We specifically focused on the European subset of SAPFLUXNET due to the high density of measurements and strong overlap of tree genera found within this region. In total, we included six tree genera with > 20 individual tree measurements: 282 plants with Pinus, 159 plants with Picea, 144 plants with Quercus, 94 plants with Fagus, 30 plants with Larix, and 29 plants with Pseudotsuga. Selected datasets represented six forest types according to the International Geosphere-Biosphere Programme (IGBP) classification: 34 evergreen needle-leaf forest (ENF), 11 mixed forest (MF), 8 deciduous broadleaf forest (DBF), 5 deciduous needle-leaf forest (DNF), 4 evergreen broadleaf forest (EBF), and 2 savannas (SAV). We treated each individual tree seasonal sap flow time series separately resulting in a total of 2279 years of sap flow ($cm^3\ h^{-1}$) observations, each corresponding to about one vegetation season (April to September, 3-6 months depending if the sensor was installed for a shorter period). The division into individual seasonal time series is important and allows us to include all sensors in our study, even if they cover only a few months at any possible point in time. The winter period (October to March) was excluded as transpiration is low or zero at most stands.

2.1.2 Selected model features

We considered six dynamic features comprising meteorological variables at an hourly resolution available in the SAPFLUXNET database: air temperature (°C), relative humidity (%), vapor pressure deficit (kPa), shortwave incoming radiation ($W\ m^{-2}$), precipitation (mm), and wind speed ($m\ s^{-1}$). We considered six static features, comprising four forest stand characteristics: mean elevation (m), long-term mean annual temperature (°C), long-term mean annual precipitation (mm), and forest type (DBF, DNF, EBF, ENF, MF, SAV), and two individual tree-level characteristics: diameter at breast height (DBH; cm) and tree genera (Fagus, Larix, Picea, Pinus, Pseudotsuga, and Quercus). We implemented one-hot-encoding for each of the 6 genera and 6 forest types in the dataset. We note that we included the DBH as a scaling variable in the model features to adequately estimate tree-level sap flow instead of standardizing and upscaling the sap flow measurements into transpiration rates per unit area. As such, one can apply the resulting models in unseen locations, potentially test novel stand structure scenarios and have the flexibility to estimate forest-level transpiration given different tree genera, sizes, and density.

2.2 Deep learning model - Long Short-Term Memory

LSTMs are recurrent neural networks that are specifically engineered to circumvent the vanishing and exploding gradient problem encountered in regular recurrent neural networks (Hochreiter and Schmidhuber, 1997; Hochreiter 1998). This is achieved through the introduction of a cell state, which provides the network with the capability to learn long-term dependencies that are typically important in environmental, sequential data. The memory cell works in conjunction with so-called "gates", mechanisms that evolve the memory and output over time, while allowing the error to propagate consistently through the network, thereby facilitating the learning process. LSTMs have showcased their efficiency and aptitude in hydrological modeling and have emerged as one of the top-performing models for simulating various state-dependant ecohydrological phenomena, such as streamflow, soil moisture and ecosystem water and carbon fluxes (e.g. Kratzert et al., 2018; 2019, Besnard et al., 2019; Bartolomeis et al., 2023).

2.2.1 LSTM hyperparameters

We explored ranges of LSTM hyperparameter settings based on previous hydrological studies (Mai et al., 2022) and saw that the LSTM performance was relatively stable with a wide range of hyperparameters. Thus, for the sake of simplicity, we only show the results of a single set of parameters here. We chose the settings from Mai et al. 2022, with modifications made to (1) a reduced sequence length (which represents the number of time steps the network looks back to process and learn from sequential inputs) and (2) a reduced number of epochs. Both changes only minimally influenced the model performance while greatly decreasing training times of the LSTMs. The hyperparameter setting used for all model variants are: number of hidden layers = 1; hidden layer neurons = 256; learning rate = 0.0005; dropout rate = 0.4; batch size = 64; sequence length = 24 hours; epoch 20; iterative optimization algorithm = ADAM.

2.2.2 Model setups and evaluation

We developed several experimental setups that differ in the amounts of data that were used to train the models with the goal of testing the model performance in different conditions. Our objective is to compare simple baseline, specialized and continental model setups 1) to assess the value of training deep learning models on larger and more diverse datasets instead of building models for each site or genera and 2) quantify model performance in predicting sap flow for unseen periods and for unseen locations.

For all model setups, the data were divided such that a single vegetation growing season of tree-level sap flow data is treated as an individual time series and entirely part of either the training, validation or test data (in total 2279 individual time series). We split the 2279 individual time series ten times into 1140 years (50%) for training, 912 years (40%) for testing and set aside the remaining 227 years (10 %) for validation. This so-called Monte Carlo Cross Validation scheme allows us to assess the robustness of our models with respect to the information content of the

training data by training LSTMs on ten different random subsamples that are drawn without repetitions (Maier et al., 2023). We assessed each model's performance using the Kling-Gupta efficiency (KGE) and its three components: Pearson correlation (r), bias ratio (α), and variability ratio (β), the Nash-Sutcliff efficiency (NSE), and the mean absolute error (MAE). For definitions of these performance metrics please see Gupta et al. (2009). All numeric input features were standardized by subtracting the mean and by dividing them by the standard deviation of the training data.

We developed two types of specialized models that have been trained on smaller subsets of the data described in section 2.1.1: For the single forest stand models, we trained LSTMs on a single forest stand (location) at a time and across several tree genera; if present at the forest stand. For the single tree genera models, we trained LSTMs on a single genera at a time (e.g. Fagus, Pinus) and across several forest stands where the genera is present. Further we trained gauged-continental models across all 64 forest stands. With the gauged-continental models, we assess the ability of the LSTMs to generalize in new time periods but in seen forest stands. Finally, we compare gauged-continental models to gauged-baseline models, representing the monthly averaged hourly diurnal cycle of sap flow for each stand and for each genera across the European continent. These baseline models are built ten times on the same randomly selected training data as the gauged-continental models.

We developed ungauged-continental models to fully examine the LSTM networks' ability to generalize at unseen forest stands and trees. We divided the data into random subsets of 50 % training (33 stands) and 50 % testing (31 stands). We stratified the splits to maintain the fractions of forest type (IGBP) within each split (Kang et al., 2023). For the forest type evergreen broad-leaved forest (EBF), which had three stands, two were used for training and one for testing. The total number of stands for training was 33 versus 31 stands for testing. We compare ungauged-continental models to ungauged-baseline models, representing the monthly averaged hourly diurnal cycle of sap flow for each genera across the european continent. These ungauged-baseline models are built ten times on the same training data as the ungauged-continental models.

3 Results

3.1 Performance of gauged-continental models

The gauged-continental models, representing the upper bound in performance, were capable of simulating sap flow with an average KGE of 0.77 ± 0.04 (Figure. 1) in comparison to the gauged-baseline models of 0.64 ± 0.05. The comparison is, however, in favor of the gauged-baseline models as not all stands have several trees with the same genera and as we use a random subsampling to generate our training data. The latter entails that the genera-specific monthly averaging works only at sites with several sensors or long observation time series where we

typically also observe higher performances of the gauged-continental models. The performance differences between the ten random training splits were small (± 0.04 KGE). This indicates that the randomly selected training data each hold a similar amount of information about the relationship between input features and sap flow and that the model is capable of generalizing. Looking closely at model performance across tree genera, there is a consistent pattern of high KGE values for Quercus, Fagus, and Pseudotsuga trees. Notably, even Pseudotsuga, the tree genera with one of the smallest dataset (80 years), achieves a KGE of 0.76 ± 0.07 in contrast to a KGE of 0.34 ± 0.07 for the gauged-baseline model. On the other hand, sap flow simulations of the Picea trees showed weaker performances with KGEs of 0.55 ± 0.06, despite being frequently found in the data (gauged-baseline models = 0.28 ± 0.03). The amount of data for a tree genera does not directly correlate with the model performance.

Figure. 2 illustrates three sequences of hourly sap flow simulations and observations of a Quercus tree, a Pseudotsuga tree and a Picea tree for five consecutive days. While there is some uncertainty, most models of the ensemble agree on the diurnal sap flow pattern, which is underpinned by the fact that Pearson correlations between different members of the model ensemble are all higher than 0.8. In agreement with the overall findings, the gauged-continental models capture the sap flow dynamics from the shown Quercus tree well (Pearson correlation = 0.9) and also match the absolute values ($\beta = 1.03$). The model performance for the shown Picea tree was lower and matches the dynamics to a certain extent (Pearson correlation = 0.77) but misses the absolute values. For the Pseudotsuga, the patterns and absolute values are well matched and even the drop of sap flow during midday for two days is matched reasonably well by the gauged-continental model. Model uncertainty across random subsampling is low for both the Quercus and Pseudotsuga tree compared to the Picea tree.

3.2 Performance of specialized models

Overall, models trained on a single genera did not outperform the gauged-continental models. At best they achieved an equivalent performance to the gauged-continental models. However, single genera models were found to be more sensitive to the amount of data, particularly if trained only on a few sites, and can exhibit large performance differences, even leading to negative KGEs for some tree genera. In contrast, the gauged-continental models remain relatively stable and less affected if the data representing a genera is reduced or if the number of stands is varied. Here, no simulation, not even removing a tree genera completely, resulted in a negative KGE. This opens the avenue to extend the training dataset by new tree types even if they have only been measured at a few locations for a short period.

We also compared the outcomes of the single-stand and gauged-continental models, focusing on the specific tree genera measured at each site as not all tree types are present in each forest stand. At certain forest stands, for example in some locations in central Europe, the performance of the single stand models closely mirrors the gauged-continental model for the dominant genera at

these sites (Fagus). On the other hand, in some locations in South France, performance was lower than the gauged-continental model for the dominant genera (Quercus trees) with KGEs around 0.5 or 0.6 versus KGEs of 0.79. At no forest stand did the single-stand models outperform the gauged-continental model.

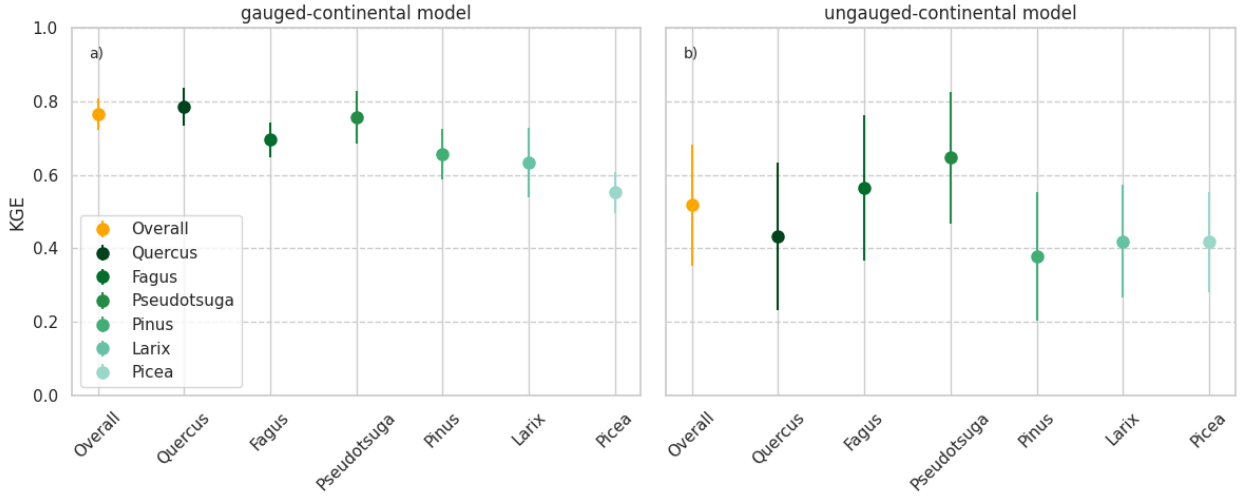

Figure. 1: Performance and standard deviation (±) of the ten a) gauged-continental models and b) the ten ungauged-continental models measured by the KGE for all testing data (overall; orange) and for each of the six tree genera (genera name; green).

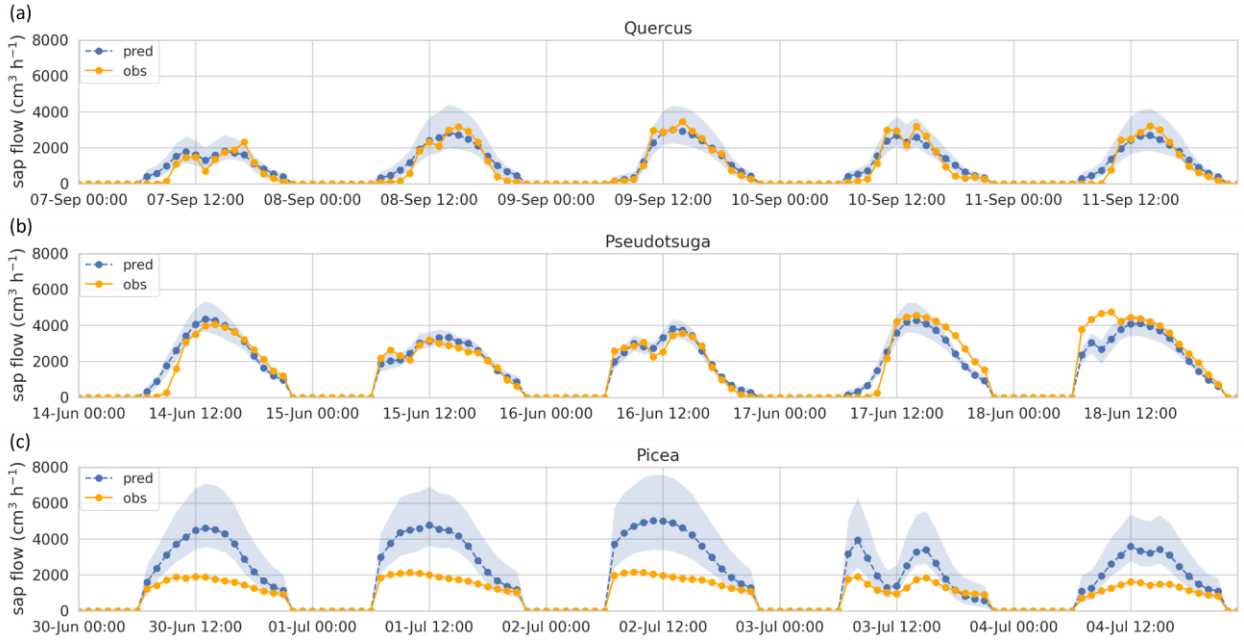

Figure. 2: Observed and simulated hourly sap flow using the gauged-continental model for five consecutive days in the testing data for a) a Quercus tree in the year 2011 (France), b) a Pseudotsuga tree in the year 2012 (Germany) and c) for a Pinus tree in the year 2001 (Sweden). Blue bands visualize the uncertainty of the gauged-continental model given the five random training subsampling.

3.3 Performance of ungauged-continental models

As expected, the performance of the ungauged-continental model is on average lower than that of the gauged-continental model yet proved reasonable with an average KGE of 0.52 ± 0.16 in comparison to a KGE of -0.11 ± 0.15 of the ungauged-baseline model. The difference for the

genera-specific performance is also in this range with the highest difference of 0.79 (KGE) for Quecrus trees (ungauged-baseline models = 0.43 and ungauged-continental models = -0.36). Model performance of the ungauged-continental models increased rather quickly if, for instance, 70 % instead of 50% of the data was used for training (KGE ~= 0.65). Diving into the individual models, we found that the performance was particularly affected by the frequency of forest types, despite stratifying random subsampling of the training data by IGBP. In other words, forest types with lower frequencies had a larger effect on the overall test scores than the amount of samples of a tree genera. The observed performance variance of the ungauged-continental models (standard deviation of 0.16) highlights the variability of the information content in the training data. Again this value reduces rather quickly if the models are trained on more data.

4 Discussion

4.1 Gauged-continental models provide robust sap flow estimates across the European continent compared to specialized and baseline models

Gathering sap flow data can result in strong variations even at the same tree, making it challenging to achieve consistent and accurate readings (Steppe et al., 2010). For instance, sap flow measurements taken just a few meters apart in trees of similar genera, size, and height can show significant differences (Vandegehuchte and Steppe, 2013). Particularly absolute values vary, while the overall dynamics, akin to what is frequently found with respect to in-situ soil moisture observations, are typically well-matched if similar sensors are installed in different trees located in the same forest stand (e.g. Zehe et al., 2005; Loritz et al., 2017; Hassler et al., 2018). Differences in absolute values might thereby arise from various small-scale structural characteristics, such as properties of the sap wood or heterogeneous flow paths inside the stem. These factors are typically unknown and not provided to our models as input. The model has, therefore, no way to learn such small-scale differences that might explain why at two similar trees in close proximity different absolute sap flows have been measured. This might be one reason why all models in this study generally learn to represent the dynamics of sap flow well, but can exhibit a high bias for certain trees. Nevertheless, given that SAPFLUXNET uses various sensors and measurement techniques across the globe (see the discussion of the SAPFLUXNET publication during the review process in Earth System Science Data; Poyatos et al., 2021) it is striking that LSTMs can generalize tree-level sap flow across diverse climate zones and genera with a performance akin to the local models described by Li et al. (2022), Loritz et al. (2022) and the specialized models trained in this study (e.g. in terms of sensor type, installing method, tree type, climate zone). Our findings reinforce thereby previous research suggesting that deep learning models, trained on large and diverse datasets, often outperform those trained on more localized, less diverse data (e.g. Kratzert et al., 2019, Sunwook and Steinschneider, 2022). While many single genera or single stand models (specialized models) in our study performed comparably to the gauged-continental models, there were notable instances where the latter was

superior. The gauged-continental models displayed reduced variance during the random subsampling of the training data and are able to simulate a tree genera even if it was entirely omitted with a KGE ranging from 0.2 to 0.4 depending on the tree genera. This result indicates that forest type (IGBP) provides more valuable insights for the models than information about which specific tree genera it should simulate. The reason being, variations within some of the forest types are less even among different tree genera, than those across forest types.

4.2 Ungauged-continental LSTMs provide reasonable sap flow estimates at ungauged forest stands, and new data have the potential to further enhance this capability

We quantified LSTMs performance for predicting hourly sap flow at forest stands that were unseen during training. The results show that the overall ungauged-continental model's performance, although lower than that of the gauged-continental model, was still reasonable with an average KGE of around 0.52. We found that the best-performing forest stands were often in the most frequent forest types (e.g. ENF). While the model was capable of predicting sap flow also in boreal and mediterranean forests, where measurements are more scarce, these predictions became less reliable the further they deviated from the training data showing clear limitations of the ungauged-continental model and the chosen training data set. The random subsampling of the training data and experiments with increasing the training data highlight that each new set of sap flow data can make the LSTM more robust, particularly in vegetation types that are less frequently monitored. Given that there are many tax funded large sap flow datasets available in Europe (and likely in other parts of the world) that are yet not openly available and not included in SAPFLUXNET, we argue that our study hints towards the currently unused potential that lies ahead when these data sets are shared in a consistent manner or if new measurement campaigns would be designed specifically to close the spatial and ecological gaps in the SAPFLUXNET dataset.

The ungauged-continental models can make reasonable hourly predictions of sap flow in unseen forest stands for a majority of the European continent. This entails that with prescribed forest type and weather data (e.g. from climate models), it is possible to create hourly sap flow maps for Europe with one, continental deep learning model. Surely these dynamic maps have clear limitations but there are limited options to gather hourly information about plant water use at the tree level at ungauged sites. Furthermore, as we simulate tree-level sap flow based on different forest stand characteristics (DBH, genera) this model could be used to assess different forest structures and their implications for regional transpiration rates and how they potentially change under different forest management strategies. Such dynamic sap flow maps could also be used to evaluate or replace transpiration models that are frequently found in land surface or hydrological models as shown in Loritz et al. (2022). Our study hence demonstrates an avenue towards developing ensembles of continental sap flow models to predict sap flow and evaluate vegetation dynamics around the globe.

5. Conclusion

This study evaluated the current potential and limits of deep learning to generalize tree-level sap flow dynamics across the European continent. Key technical criteria for developing robust LSTMs include the random subsampling strategy based on the vegetation seasons and the requirement to train the networks on large and diverse data. If trained properly we demonstrate that LSTMs can achieve a reasonable level of performance in predicting hourly sap flow for different tree genera, climate zones and forest types. Training deep learning models on large and diverse datasets and on several tree genera at the same time proved beneficial compared to specialized models and supported that LSTMs are capable generalizing vegetation dynamics beyond the individual tree-level sap flow measurement. This research paves the way for producing hourly tree-level sap flow maps across Europe, harnessing the combination of forest characteristics and dynamic meteorological drivers. Our study hints that as more sap flow datasets become openly available, the accuracy and coverage of such models are expected to improve significantly particularly for forest types that are less frequently found in the SAPFLUXNET dataset. This holds the promise of increasing our ability to simulate the vegetation water use dynamics that are encoded in sap flow and could serve as a valuable benchmark for different land surface and hydrological models.

Acknowledgments

This research contributes to the ViTamins (Invigorating Hydrological Science and Teaching: merging key Legacies with new Concepts and Paradigms) Project funded by the Volkswagen Foundation.

Open Research

SAPFLUXNET is openly available at https://zenodo.org/records/3971689. All python codes to run the models are openly available at: 10.5281/zenodo.10118262.

References

de Bartolomeis, P., Meterez, A., Shu, Z., & Stocker, B. D. (n.d.). An effective machine learning approach for predicting ecosystem CO 2 assimilation across space and time. doi: 10.5194/egusphere-2023-1826

Besnard, S., Carvalhais, N., Altaf Arain, M., Black, A., Brede, B., Buchmann, N., Chen, J., Clevers, J. G. P. W., Dutrieux, L. P., Gans, F., Herold, M., Jung, M., Kosugi, Y., Knohl, A., Law, B. E., Paul-Limoges, E., Lohila, A., Merbold, L., Roupsard, O., … Reichstein, M. (2019). Memory effects of climate and vegetation affecting net ecosystem CO2 fluxes in global forests. PLoS ONE, 14(2). doi: 10.1371/journal.pone.0211510

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. http://arxiv.org/abs/1406.1078

Dugas, W. A., Wallace, J. S., Allen, S. J., & Roberts, J. M. (1993). Heat balance, porometer, and deuterium estimates of transpiration from potted trees. *Agricultural and Forest Meteorology*, *64*(1–2), 47–62. doi: 10.1016/0168-1923(93)90093-W

Dugas, W. A., Heuer, M. L., Hunsaker, D., Kimball, B. A., Lewin, K. F., Nagy, J., & Johnson, M. (1994). Sap flow measurements of transpiration from cotton grown under ambient and enriched CO2 concentrations. Agricultural and Forest Meteorology, 70(1–4), 231–245. doi: 10.1016/0168-1923(94)90060-4

Good, S. P., Noone, D., & Bowen, G. (2015). Hydrologic connectivity constrains partitioning of global terrestrial water fluxes. Science, 349(6244), 175–177. doi: 10.1126/science.aaa5931

Granier, A., & Loustau, D. (1994). Measuring and modelling the transpiration of a maritime pine canopy from sap-flow data. Agricultural and Forest Meteorology, 71(1–2), 61–81. doi: 10.1016/0168-1923(94)90100-7

Gupta, H. v., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. doi: 10.1016/j.jhydrol.2009.08.003

Hassler, S. K., Weiler, M., & Blume, T. (2018). Tree-, stand- and site-specific controls on landscape-scale patterns of transpiration. Hydrology and Earth System Sciences, 22(1), 13–30. doi: 10.5194/hess-22-13-2018

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Maier, H. R., Zheng, F., Gupta, H., Chen, J., Mai, J., Savic, D., Loritz, R., Wu, W., Guo, D., Bennett, A., Jakeman, A., Razavi, S., & Zhao, J. (2023). On how data are partitioned in model development and evaluation: Confronting the elephant in the room to enhance model generalization. Environmental Modelling & Software, 167, 105779. doi: 10.1016/j.envsoft.2023.105779

Hoek van Dijke, A. J., Mallick, K., Teuling, A. J., Schlerf, M., Machwitz, M., Hassler, S. K., Blume, T., & Herold, M. (2019). Does the Normalized Difference Vegetation Index explain spatial and temporal variability in sap velocity in temperate forest ecosystems? *Hydrology and Earth System Sciences*, *23*(4), 2077–2091. doi: 10.5194/hess-23-2077-2019

Jackisch, C., Knoblauch, S., Blume, T., Zehe, E., & Hassler, S. K. (2020). Estimates of tree root water uptake from soil moisture profile dynamics. Biogeosciences, 17(22), 5787–5808. doi: 10.5194/bg-17-5787-2020

462 Kang, Y., Gaber, M., Bassiouni, M., Lu, X., and Keenan, T.: CEDAR-GPP: spatiotemporally
463 upscaled estimates of gross primary productivity incorporating CO2 fertilization, Earth Syst. Sci.
464 Data Discuss. [preprint], doi: 10.5194/essd-2023-337, in review, 2023.
465
466 Kling, H., & Gupta, H. (2009). On the development of regionalization relationships for lumped
467 watershed models: The impact of ignoring sub-basin scale variability. Journal of Hydrology,
468 373(3–4), 337–351. doi: 10.1016/j.jhydrol.2009.04.031
469
470 Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff
471 modelling using Long Short-Term Memory (LSTM) networks. Hydrology and Earth System
472 Sciences, 22(11), 6005–6022. doi: 10.5194/hess-22-6005-2018
473 Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019).
474 Towards learning universal, regional, and local hydrological behaviors via machine learning
475 applied to large-sample datasets. Hydrology and Earth System Sciences. doi: 10.5194/hess-23-
476 5089-2019
477
478 Koppa, A., Rains, D., Hulsman, P., Poyatos, R., & Miralles, D. G. (2022). A deep learning-based
479 hybrid model of global terrestrial evaporation. Nature Communications, 13(1), 1912. doi:
480 10.1038/s41467-022-29543-7
481
482 Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In Nature (Vol. 521, Issue 7553, pp.
483 436–444). Nature Publishing Group. doi: 10.1038/nature14539
484
485 Li, Y., Ye, J., Xu, D., Zhou, G., & Feng, H. (2022). Prediction of sap flow with historical
486 environmental factors based on deep learning technology. Computers and Electronics in
487 Agriculture, 202. doi: 10.1016/j.compag.2022.107400
488
489 Loritz, R., Bassiouni, M., Hildebrandt, A., Hassler, S. K., & Zehe, E. (2022). Leveraging sap
490 flow data in a catchment-scale hybrid model to improve soil moisture and transpiration
491 estimates. Hydrology and Earth System Sciences, 26(18), 4757–4771. doi: 10.5194/hess-26-
492 4757-2022
493
494 Mai, J., Shen, H., Tolson, B. A., Gaborit, É., Arsenault, R., Craig, J. R., Fortin, V., Fry, L. M.,
495 Gauch, M., Klotz, D., Kratzert, F., O'Brien, N., Princz, D. G., Rasiya Koya, S., Roy, T.,
496 Seglenieks, F., Shrestha, N. K., Temgoua, A. G. T., Vionnet, V., & Waddell, J. W. (2022). The
497 Great Lakes Runoff Intercomparison Project Phase 4: the Great Lakes (GRIP-GL). Hydrology
498 and Earth System Sciences, 26(13), 3537–3572. doi: 10.5194/hess-26-3537-2022
499
500 Poyatos, R., Granda, V., Flo, V., Adams, M. A., Adorján, B., Aguadé, D., Aidar, M. P. M.,
501 Allen, S., Alvarado-Barrientos, M. S., Anderson-Teixeira, K. J., Aparecido, L. M., Altaf Arain,
502 M., Aranda, I., Asbjornsen, H., Baxter, R., Beamesderfer, E., Berry, Z. C., Berveiller, D.,
503 Blakely, B., … Martínez-Vilalta, J. (2021). Global transpiration data from sap flow
504 measurements: The SAPFLUXNET database. In Earth System Science Data (Vol. 13, Issue 6).
505 doi: 10.5194/essd-13-2607-2021
506

Renner, M., Hassler, S. K., Blume, T., Weiler, M., Hildebrandt, A., Guderle, M., Schymanski, S. J., & Kleidon, A. (2016). Dominant controls of transpiration along a hillslope transect inferred from ecohydrological measurements and thermodynamic limits. Hydrology and Earth System Sciences, 20(5), 2063–2083. doi: 10.5194/hess-20-2063-2016

Seeger, S., & Weiler, M. (2021). Temporal dynamics of tree xylem water isotopes: In situ monitoring and modeling. Biogeosciences, 18(15), 4603–4627. doi: 10.5194/bg-18-4603-2021

Steppe, K., de Pauw, D. J. W., Doody, T. M., & Teskey, R. O. (2010). A comparison of sap flux density using thermal dissipation, heat pulse velocity and heat field deformation methods. Agricultural and Forest Meteorology, 150(7–8), 1046–1056. doi: 10.1016/j.agrformet.2010.04.004

Wi, S., & Steinschneider, S. (2022). Assessing the Physical Realism of Deep Learning Hydrologic Model Projections Under Climate Change. Water Resources Research, 58(9). doi: 10.1029/2022WR032123

Vandegehuchte, M.W. and Steppe, K. 2013. Corrigendum to: Sap-flux density measurement methods: working principles and applicability. Functional Plant Biology 40(10), 1088-1088. doi: 10.1071/FP12233_CO

Zehe, E., Graeff, T., Morgner, M., Bauer, A., & Bronstert, A. (2010). Plot and field scale soil moisture dynamics and subsurface wetness control on runoff generation in a headwater in the Ore Mountains. Hydrology and Earth System Sciences, 14(6), 873–889. doi: 10.5194/hess-14-873-2010