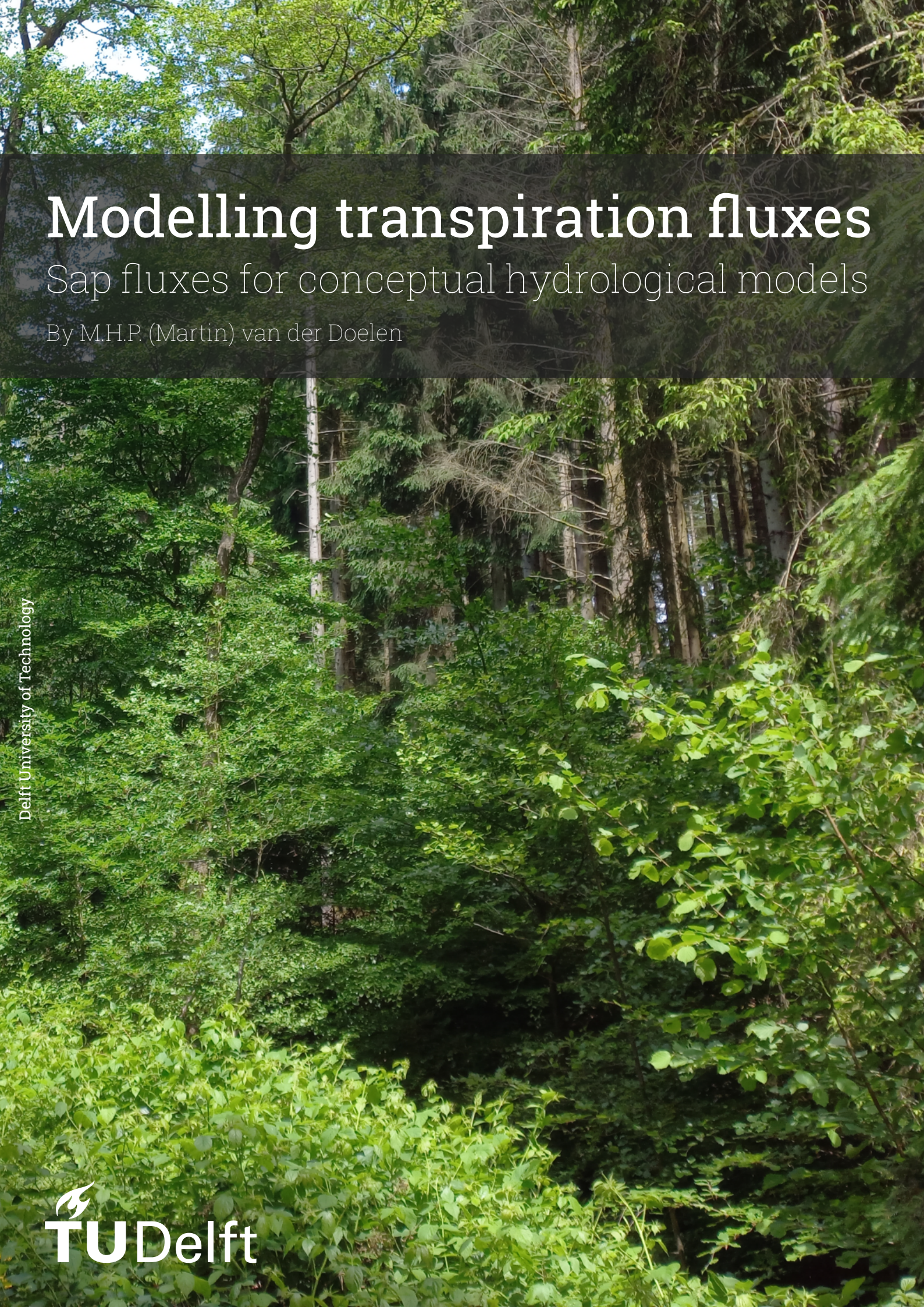


Modelling transpiration fluxes

Sap fluxes for conceptual hydrological models

By M.H.P. (Martin) van der Doelen

Delft University of Technology

TU Delft

Modelling transpiration fluxes

Sap fluxes for conceptual hydrological models

by

M.H.P. (Martin) van der Doelen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on 30 September, 2025.

Student number:	4606647
Project duration:	19 February, 2025 – 30 September, 2025
Thesis committee:	Dr. habil. M. Hrachowitz, TU Delft, supervisor, chair
	Dr. Ir. A.M.J. Coenders, TU Delft, supervisor

Cover: Weierbach Experimental Catchment in Luxembourg.
Picture taken in June, 2025.
Source: Photograph by the author.

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft Delft University of Technology

Preface

Before you lies the master thesis *Modelling transpiration fluxes: Sap fluxes for conceptual hydrological models*, which is the result of my graduation project to obtain the degree Master of Science in Environmental Engineering at Delft University of Technology. This project has been an incredible journey through the world of vegetation and hydrological modelling. While I gained valuable knowledge, each new insight also revealed further fascinating questions, inspiring me to keep expanding my understanding. This leaves me with mixed feelings: I am proud to present this thesis, yet a little sad that such an inspiring journey has come to an end.

I would like to thank my supervisors, Markus Hrachowitz and Miriam Coenders, for their enthusiasm, expertise, and constructive feedback. Their encouragement and inspiring guidance were invaluable throughout the entire process.

I am also grateful to Christophe Hissler for providing this research topic and sharing his knowledge about the study area. I would like to thank Jean François Iffly for providing additional datasets that were essential for this research.

Finally, I would like to thank my girlfriend and family, in particular my parents and sister, for their ongoing support. I am also grateful to my friends for their encouragement and invaluable companionship during this process. I especially want to thank Loïs, who shared many hours working alongside me on our theses, making this journey both productive and enjoyable.

I hope you will enjoy reading this thesis as much as I enjoyed working on it.

M.H.P. (Martin) van der Doelen
Delft, September 2025

Abstract

Vegetation activity plays a key role in the water cycle, with transpiration accounting for 50–80% of total terrestrial evaporation globally. Yet many conceptual hydrological models neglect phenology, estimating transpiration as a function of relative soil moisture and assuming constant vegetation activity. This simplification leads to systematic errors, particularly in winter, when conventional models often overestimate transpiration.

Because transpiration is strongly correlated with sap flow, this study developed three methods to incorporate tree phenology into a semi-distributed conceptual hydrological model. The model structure distinguished between coniferous and deciduous trees. Sap flow dynamics were included either directly, using sap flow data, or indirectly, using temperature as a proxy for seasonal variation.

To address the limited availability of sap flow data, a Generalized Additive Model (GAM) was developed to predict normalised sap flow. Using temperature, relative humidity, incoming shortwave radiation, volumetric soil water content, and normalised accumulated growing degree-day as predictors, the model reliably reproduced sap flow dynamics for both tree types. In addition, the GAM framework enabled separate analysis of each predictor's relationship with sap flow, providing clearer insights into the underlying processes and the relative influence of individual variables.

The results show that including tree phenology improves the ability of conceptual hydrological models to represent transpiration seasonality and vegetation dynamics. Although discharge simulations were not substantially improved, the added internal realism reduces equifinality and makes the model more robust under changing conditions.

Among the three developed methods, the direct inclusion of sap flow dynamics resulted in the highest model performance, particularly in transpiration simulations. This highlights the added value of the sap flow prediction model itself, which provided a robust link between environmental drivers and vegetation dynamics, thereby strengthening the integration of phenology into conceptual hydrological modelling.

Contents

List of Figures

List of Tables

1 Introduction

Water is a fundamental component of the Earth system. Understanding its distribution and movement through the hydrological cycle is essential for managing natural resources such as forests, predicting flood and drought risks, and sustaining ecosystems. Hydrological processes govern the availability of freshwater and are highly sensitive to climate variability, land-use changes, and vegetation dynamics. In light of the increasing impacts of climate change, hydrology plays a critical role in both environmental science and water resource management.

1.1. Background and motivation

In hydrology, hydrological models are developed to quantify and simulate hydrological processes. They are used to predict discharge, groundwater recharge, floods, and droughts, among other things. Additionally, they can help improve the physical understanding of the system's behaviour. However, due to the complexity of the hydrological cycle and its interacting processes, these models are necessarily simplified representations of reality. As such, they inevitably include uncertainties and limitations.

1.1.1. Hydrological models

Hydrological models can be classified into three categories. The first category is black-box models, which use statistical relations between the input and output without explicit knowledge of the underlying processes. Therefore, they do not aid in the physical understanding of the system's behaviour (Liu et al., 2019). The second and third categories are both process-based models, namely the conceptual and physically based models. Physically-based models describe the system from a micro-scale perspective, explicitly representing as many processes as possible (Hrachowitz and Clark, 2017). The CATFLOW hydrological model, for example, uses the Darcy-Richards and Saint-Venant equations to represent soil water and surface runoff, respectively (Loritz et al., 2022). These micro-scale processes are then integrated to catchment scale using a spatial grid, which comes at a computational cost and requires detailed input data. Conceptual models are in between the black-box and physically based models. They provide a macro-scale description of the hydrological system, using a more abstract and more simplified representation of the processes involved by combining micro-scale processes into a macro-scale manifestation (Hrachowitz and Clark, 2017). These models can be spatially lumped or (semi-) distributed.

Because conceptual models are more simplified, they have lower computational costs and lower data requirements compared to physically based models. At the same time they still provide a better physical understanding of the system's behaviour than black-box models.

It could be argued that physically-based models should always produce an accurate result based on their detailed representation of reality using micro-scale physics. Although this might be true on point-scale, these models encounter multiple issues at larger spatial scales (Beven, 2001). This suggests that micro-scale properties and processes do not necessarily produce accurate results on a large spatial scale, such as a catchment. This is also related to heterogeneity, as some properties can be highly variable over a spatial domain while this data is not always available at the required scale. Additionally, these complex models are essentially 'unverifiable' at catchment scale (Blöschl, 2001). Another issue is equifinality, the concept that the same outcome can be achieved in different ways, e.g., different combinations of (calibrated) parameters. This is often linked with over-parametrisation of a model, which is common for complex physically-based models. As a result, a model might produce promising results for the wrong reasons (Beven, 2001; Blöschl, 2001).

These problems are not only limited to physically-based models, as conceptual models are also affected—albeit to a lesser degree. Where physically-based models might have hundreds or thousands of parameter values to be specified, conceptual models require far fewer parameters, reducing the issue of equifinality. Furthermore, by focusing on macro-scale effects, such as the hydrological patterns shaped by the catchment's self-organising capacity, conceptual models can effectively capture the dominant system dynamics through calibration. This allows them to implicitly include the combined effects of complex, micro-scale processes and heterogeneities without the need to explicitly model them. This makes conceptual models more robust.

1.1.2. Transpiration and phenology

Evaporation (E) is the transport of water from the Earth's surface into the atmosphere and consists of two major processes. The first, evaporation, is the conversion of liquid water from a variety of surfaces to water vapour. This includes evaporation from open water (E_o), soil (E_s), and interception (E_i). The second process is transpiration (E_t), a biological process where water contained in plant tissues is vaporized and removed to the atmosphere (Allen et al., 1998).

Transpiration plays a crucial role in plant physiological processes such as photosynthesis (Shao et al., 2023). Photosynthesis takes place in the leaves of trees. Using the chlorophyll in its leaves, it absorbs sunlight and combines CO_2 with water to produce glucose, producing oxygen as a by-product. The needed CO_2 is obtained by opening the stomata in the leaves, which also leads to water loss through transpiration. To compensate for this loss, water (with dissolved minerals) is transported via the xylem from the roots to the rest of the tree. This flow through the xylem is called the sap flow. Only a tiny fraction of the water taken up is used within the tree, with the remaining water being used for transpiration (Allen et al., 1998). Therefore, sap flow is strongly correlated with tree transpiration and its water use, and can be used as an indicator of plant water status (Gimenez et al., 2005; Köstner et al., 1992).

There are two main categories of trees: deciduous and coniferous, each with its own phenology—the seasonal development and activity. Deciduous trees shed their leaves annually while coniferous trees typically have needles and retain them year-round, resulting in different transpiration behaviour throughout the year. In the early spring, deciduous trees still have to develop their leaves. After budburst, when new leaves emerge, the transpiration rates gradually increase. In autumn, they undergo abscission, the process where deciduous trees shed their leaves to conserve water and energy during the winter. Because coniferous trees do not lose their leaves, they have a higher transpiration rate at the beginning and end of, as well as outside, the growing season. Deciduous trees, however, usually have a higher transpiration rate during the growing season. Furthermore, the timing of these phenological events is highly sensitive to temperature (Polgar and Primack, 2011).

1.1.3. Impact of vegetation

Tree activity is a key component of the water cycle, with transpiration accounting for 50–80% of total terrestrial evaporation globally (Coenders-Gerrits et al., 2014; Good et al., 2015; Wei et al., 2017). A change in vegetation can significantly impact the water balance and stream flow in a forest ecosystem (Breil et al., 2023; Juice et al., 2016). Higher transpiration rates reduce runoff, resulting in lower discharge. This change in vegetation can either be intentional or unintentional.

Intentional tree species conversion is a common technique in forest management, which can strongly influence catchment discharge (Rahmat et al., 2018). A change from coniferous to deciduous, or vice versa, can increase or decrease the annual runoff. Additionally, the distribution of runoff over a year can also change due to differences in phenology between coniferous and deciduous trees.

Tree species conversion can also happen unintentionally. One example occurred in central Massachusetts (United States) and was investigated by Daley et al. (2007). Here, eastern hemlocks (coniferous) were infected by *hemlock woolly adelgid*, which can be fatal for this species. Black birch (deciduous) has been found to be the dominant replacement in this region. During the peak growing season, the daily rates of transpiration were 1.6 times higher in black birch. This shift in tree species is expected to significantly alter the region's water balance, particularly during the peak growing season, potentially reducing stream flow in late summer due to increased water uptake by black birch. As a result, stream flow in previously low- or moderate-flow streams may become unsustainable during this period.

Furthermore, climate-driven forest die-off from drought and heat stress is expected to increase with climate change, impacting forests worldwide (Anderegg et al., 2013). Changes in temperature and precipitation strongly influence the relationships between trees and insects, such as bark beetles, causing widespread forest mortality (Jaime et al., 2024). For example, stressors such as heat and drought activate stress hormones and proteins in Norway spruce, which attract bark beetles (*Ips typographus*). These bark beetles cause extensive tree mortality in European Norway spruce forests (Netherer et al., 2021). Bearup et al. (2014) found that changes in forest transpiration due to *mountain pine bark beetle* infestations in the Rocky Mountains of North America strongly influenced the streamflow. The fractional late-summer groundwater contributions to stream flow increased by 30 $\pm$ 15% after infestation.

1.1.4. Actual evaporation in hydrological models

Determining the actual amount of evaporation (E_a) is difficult as measuring it is neither easy nor cheap. It is easier to determine the maximum possible evaporation based on available energy and assuming water is not limiting. This is called the potential evaporation (E_p). By definition, actual evaporation can never exceed the potential evaporation ($E_a \leq E_p$).

In conceptual hydrological models, E_a is often estimated by scaling E_p with a reduction factor that usually depends on the soil moisture content (Bouaziz et al., 2021; Jayathilake and Smith, 2020; Zhao et al., 2013). This approach neglects dynamic vegetation processes such as phenology, which significantly influence transpiration rates. It treats E_a not as an important process but as a static quantity to estimate, whereas in reality, vegetation activity is a major part of the total evaporation. This highlights a key limitation of such models.

As E consists of several different processes, dynamic partitioning of E into E_t and other evaporation components is a common practice in more complex models to increase their accuracy (e.g., Shuttleworth-Wallace, ENWATBAL, Cupid-DPEVAP, SWEAT, TSEB, FAO dual Kc model, HYDRUS-1D (Kool et al., 2014)). However, the implementation of these dynamic partitioning methods in conceptual models has not been sufficiently investigated.

Previous research by Mert (2021) and Pierik (2022) showed promising results when combining conceptual models with a dynamic partitioning based on phenology. Mert (2021) developed three methods to partition E in conceptual models. The first method uses a part of the Jarvis model developed by Jarvis and McNaughton (1986), using stomatal responses to temperature to partition evaporation. The second method was based on the crop evaporation method developed by Allen et al. (1998). Here, the leaf area index (LAI) was used to identify the growing stages of vegetation, which were used to partition the evaporation. The third method was a combination of the first two methods. These methods were implemented in lumped FLEX and GR4J models for two catchments. The modified models were able to provide discharge simulations as accurate as those of models that do not include plant phenology, while their internal transpiration was more accurate. Furthermore, discharge simulations from the modified FLEX models were slightly closer to observations during spring months. Pierik (2022) tested the suggested method based on Allen et al. (1998) developed by Mert (2021) on 28 catchments located in the United States and found similar results.

1.2. Problem statement

Conceptual hydrological models are highly simplified representations of hydrological systems. Because they are process-based, they provide a better physical understanding of the system's behaviour compared to black-box models. They have advantages over physically-based models due to their lower data requirements and computational costs. However, assumptions are necessary to implement these simplifications. One of these assumptions is the fixed partitioning of evaporation, neglecting the phenology of vegetation. As a result, even though the model might provide accurate discharge simulations, the internal transpiration flux might be unrealistic, while this is a key hydrological process. Research by Mert (2021) and Pierik (2022) demonstrated the potential to improve the accuracy of the model's transpiration estimates by incorporating phenological information. Both studies focused on homogeneous land cover.

Although the results were promising, Pierik (2022) emphasised the need for further research on alternative methods to incorporate phenological information into conceptual hydrological models. They

also recommended extending the analysis from catchments with relatively homogeneous land cover to those with more heterogeneous conditions.

1.3. Research objectives and questions

Transpiration and sap flow are strongly correlated. Therefore, sap flow is a possible candidate to include tree phenology in a conceptual hydrological model. However, sap flow data are usually temporally limited, whereas conventional hydrological models require several years of data.

The objective of this study is to incorporate tree phenology into a conceptual hydrological model for a heterogeneous forest catchment by using environmental variables to represent sap flow dynamics.

Hypothesis

It is hypothesised that incorporating tree phenology into a conceptual hydrological model through environmental variables representing sap flow dynamics will improve the model's performance in simulating both discharge and transpiration, compared to models without vegetation dynamics.

Research questions

1. How can environmental variables representing sap flow dynamics be used to include tree phenology in a conceptual hydrological model for a catchment with heterogeneous tree coverage?
2. How can sap flow be predicted based on environmental variables?
3. How do conceptual hydrological models that include phenology information perform on discharge and transpiration simulations compared to the conventional conceptual hydrological model?

1.4. Thesis outline

Chapter 2 describes the methodology of this study. This chapter first describes the study area and how the required data were obtained and prepared for use in the developed models. Next, the structure of the conventional lumped model is described, along with how it was modified to a semi-distributed structure. Furthermore, it explains how phenology was incorporated into this model using three different methods. Subsequently, it is described how the models were calibrated and evaluated. Finally, the model developed to predict normalised sap flow from environmental variables is presented.

Chapter 3 includes the results obtained from the methodology. To start, the sap flow prediction model was analysed and its outcomes evaluated. Next, the methods for incorporating phenology are described. Finally, the conventional and modified hydrological models are evaluated. This includes a sensitivity analysis and assessment of the discharge and transpiration simulations.

Chapter 4 discusses the results of each part of the study, interprets the findings, and highlights their implications and limitations.

Chapter 5 presents the conclusions of this study.

2 Methodology

This chapter outlines the methods used to investigate the incorporation of tree phenology into a conceptual hydrological model for a heterogeneous forest catchment. It starts with a description of the study area (Section 2.1). This is followed by a description of the datasets obtained and the procedures used to prepare continuous time series for model inputs (Section 2.2). Next, a conventional conceptual hydrological model was chosen and modified to fit this study's requirements (Section 2.3). Three methods were then developed to modify the conventional model, incorporating sap flow dynamics directly or indirectly (Section 2.4). After that, Section 2.5 describes how the models were calibrated and validated. Finally, the development of a sap flow prediction model is presented (Section 2.6), which allowed the extension of limited sap flow observations to the longer period required for the modified hydrological models.

2.1. Study area

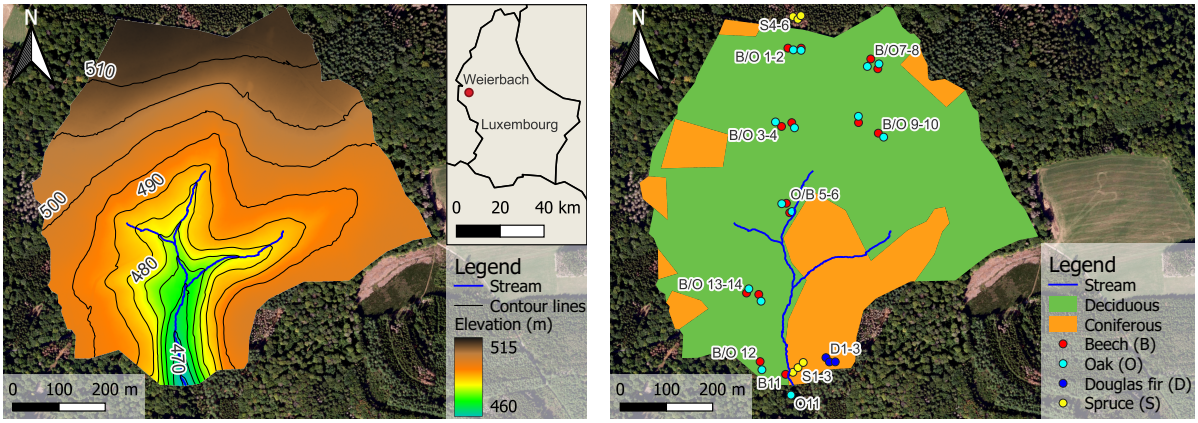
The Weierbach Experimental Catchment (WEC) in Luxembourg (49°49'38"N, 5°47'44"E) is a fully forested catchment covering 0.45 km^2 and drains south into the Attert River, a tributary of the Alzette River. It lies within the Ardennes massif and is characterised by deep V-shaped valleys. The slopes range from 0.6° to 31° with a mean of 5.86° (Schoppach et al., 2021). Elevation in the catchment ranges from 458 to 514 m a.s.l. with a mean of 498 m a.s.l. (Martínez-Carreras et al., 2016; Pfister et al., 2017), as illustrated in Figure 2.1a.

The WEC has a semi-marine climate with a mean annual temperature of 8.9 °C. The mean monthly temperatures range between 0.8 °C in January to 17 °C in July (2007–2019) (Schoppach et al., 2021). Between 2006 and 2014, the mean annual precipitation was 953 mm $year^{-1}$, the mean annual potential evaporation was 593 mm $year^{-1}$, and the average annual stream discharge was 478 mm $year^{-1}$ (Pfister et al., 2017). Mean monthly precipitation between 2007 and 2019 ranged from 39 mm $month^{-1}$ in April to 103 mm $month^{-1}$ in December (Schoppach et al., 2021). Higher evaporation with lower precipitation rates during the summer result in discharge seasonality, with the lowest streamflow between July and October.

The vast majority of the catchment is covered by deciduous forest, composed of a mix of European beech (*Fagus sylvatica*) and pedunculate and sessile oak (*Quercus robur* and *Quercus petraea*). Douglas fir (*Pseudotsuga menziesii*) and Norway spruce (*Picea abies*), both coniferous, cover the remaining part of the catchment (Fabiani, 2022). Here, the Douglas fir is located in the south-east of the catchment while the spruce trees grow in smaller stands across the catchment. In the remainder of this report, these species will be referred to by their common names: beech, oak, Douglas fir, and spruce.

Estimates of tree species coverage within the catchment vary slightly across studies. Hissler et al. (2021) reports 70% coverage for the beech-oak forest, 15% for Douglas fir, and 15% for spruce. In contrast, Schoppach et al. (2021) estimates 89% for beech-oak, with only 6% and 5% for Douglas fir and spruce, respectively. Jiménez-Rodríguez et al. (2024) estimated 77.6% coverage for the beech-oak forest and 22.4% combined coverage for Douglas fir and spruce. Martínez-Carreras et al. (2016) found similar coverages, with 76% for beech-oak and 24% for coniferous stands.

In this study, the combined estimates of 76% for beech-oak and 24% for Douglas fir and spruce are adopted, in line with Martínez-Carreras et al. (2016). These values closely match those derived from digitizing the vegetation cover figure in Schoppach et al. (2021) using QGIS, based on aerial imagery



(a) Location and elevation of the catchment.

(b) Tree distribution and numbering.

Figure 2.1: (a) Elevation of the Weierbach Experimental Catchment (WEC) and its location within Luxembourg. (b) Distribution of deciduous and coniferous forest within the catchment, including the number of trees monitored for sap flow in 2019 and 2020.

from Esri (2022). Notably, the spatial distribution visible in the figure appears inconsistent with the percentages reported in the text, supporting the use of an independently derived estimate.

The WEC is the most instrumented and studied catchment in Luxembourg, with high-frequency hydro-climatological information recorded since 2009 (Hissler et al., 2021). Additionally, tree physiological functioning, including sap flow, was monitored during the 2019 and 2020 growing seasons. An overview of the measurement locations is shown in Figure 2.1b.

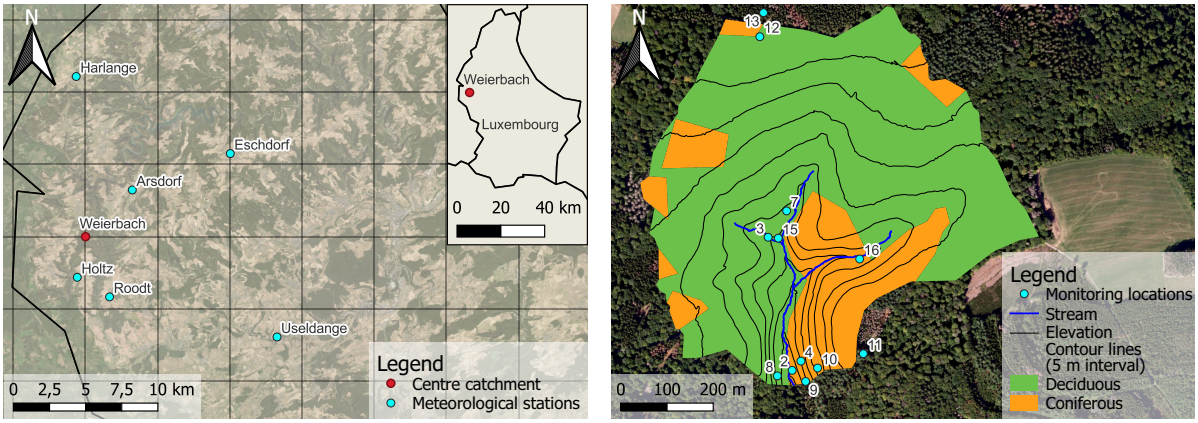
2.2. Data availability & preparation

Several different time series were needed for the sap flow prediction model and the rainfall-runoff model (see Sections 2.3, 2.4 and 2.6). For the variables listed below, multiple datasets were obtained.

1. Temperature
2. Relative Humidity
3. Incoming Shortwave Radiation
4. Soil Volumetric Water Content
5. Discharge
6. Precipitation
7. Sap flow velocity
8. Transpiration

Some of the data were retrieved from the Administration des Services Techniques de l'Agriculture (ASTA, 2025), which operates several automatic meteorological stations in the area (Figure 2.2a) with measured parameters varying between stations. The locations of monitoring sites within the WEC are shown in Figure 2.2b.

This section describes how this data were obtained and prepared to be used as inputs for the models in study, as these require continuous time series. First, hydro-meteorological and soil data were prepared, followed by data related to vegetation water usage.



(a) Nearby meteorological stations.

(b) Monitoring sites within the WEC.

Figure 2.2: (a) Locations of nearby meteorological stations used in this study. (b) Monitoring sites within the Weierbach Experimental Catchment, including hydrological, soil, and vegetation measurement points.

2.2.1. Hydro-meteorological and soil data

This subsection describes the data obtained for climatic and soil-related variables. The data includes temperature, relative humidity, incoming shortwave radiation, soil moisture, discharge and precipitation. Additionally, potential evaporation was derived from the temperature dataset.

Temperature

Temperature data from within the WEC were retrieved from Hissler et al. (2021). Here, four HMP 45 C (Vaisala Oyj, Vantaa, Finland) probes were installed (sites 8–11 in Figure 2.2b) that measured temperature and relative humidity at a resolution of 15 minutes from 2012–2019. The distance from one probe to the next is approximately 60 m on average. Therefore, it is expected that the probes show similar patterns. It was observed that data from location 11 were shifted by 12 days between March 2016 and December 2019. This shift was corrected for by shifting the data back, resulting in missing values for location 11 at the end of December 2019.

In addition to the publicly available dataset covering the period 2012–2019, supplementary data for the years 2020–2022 was obtained directly from the Luxembourg Institute of Science and Technology (LIST) upon request. This data come from the same probes and locations but with a higher temporal resolution of 5 minutes. Both datasets were combined into one merged dataset ranging from 2012–2022. Afterwards, the resulting dataset was resampled to an hourly frequency by taking the hourly mean values. Finally, a single time series was created by taking the mean values of all four locations.

Additionally, temperature data were retrieved from ASTA (2025) for nearby meteorological stations located in Holtz, Roodt and Arsdorf. These datasets have an hourly resolution and measurements were taken at 2 m above the surface. Roodt and Arsdorf data were obtained from 2012 to 2022. Temperature data at Holtz were available from 15 October 2015 onwards and were obtained from 2015 to 2022.

Lastly, temperature data from the ERA5 reanalysis product was retrieved from Hersbach et al. (2023) via the Copernicus Climate Change Service (C3S) Climate Data Store for the WEC area. The temperature was calculated in kelvin for an elevation of 2 m by interpolating between the lowest model level and the Earth's surface, where atmospheric conditions are also taken into account (Copernicus Climate Change Service, Climate Data Store, 2023). This dataset has an hourly resolution and was obtained for 2012 to 2022.

The resulting time series were compared. It was noticed that the temperatures measured in Arsdorf often had much lower minimum daily temperatures than the other time series, while the daily maximum temperatures were similar. Furthermore, it was noticed that daily maximum temperatures measured in the WEC were a few degrees lower during the summer compared to the other time series. This is possibly caused by the thermal insulating function of forests (Renaud et al., 2011).

A selection was made from these temperature datasets. Data from Arsdorf was not used because of its deviation compared to the other data. Temperature data from the WEC were also excluded due to the suspected thermal insulating effect of the forest. Because photosynthesis occurs in the canopy and temperatures in the WEC were measured below the canopy, these measurements might be an underestimation of the temperature at the canopy level.

A continuous temperature time series was constructed ranging from 2012 to 2022 using a hierarchical gap-filling approach based on three data sources, prioritizing direct observations over modelled data. Data from the Holtz station was chosen as primary dataset, which served as the main reference temperature due to its proximity to the WEC. Data gaps in the Holtz time series were initially filled using observations from the Roodt station. Any remaining gaps were subsequently filled using the ERA5 reanalysis dataset, ensuring full temporal coverage.

Relative humidity

Relative humidity data from within the WEC were retrieved from Hissler et al. (2021) for 2012–2019. These data were measured at the same locations and by the same four HMP 45 C probes (Vaisala Oyj, Vantaa, Finland) as the temperature with a resolution of 15 minutes. Data from 2020–2022 were obtained directly from LIST upon request and had an increased resolution of 5 minutes. As also done for the temperature data, both datasets were combined into one merged dataset ranging from 2012–2022. Afterwards, the resulting dataset was resampled to an hourly frequency by taking the hourly mean values. Finally, a single time series was created by taking the mean values of all four locations.

Additionally, hourly relative humidity data were retrieved from ASTA (2025) from two nearby meteorological stations (Roodt and Arsdorf) for 2012–2022. These are the same stations used for the temperature data. However, data were not available for the Holtz station. Outliers in the data from the Roodt station were removed.

Lastly, the relative humidity was calculated using the hourly 2 m dew point temperature and the already obtained 2 m temperature, both from the same ERA5 reanalysis product. First, the saturated vapour pressure e_s was calculated in kPa using the Clausius-Clapeyron equation (Koutsoyiannis, 2012), shown below as Equation 2.1. Here, e_s is a function of temperature (T) in °C. Afterwards, the actual vapour pressure e_a was calculated using the same equation but using the dew point temperature. Finally, the relative humidity was calculated using Equation 2.2, where h is the relative humidity as a percentage.

$$e_\mathsf{s}(T) = 0.61 \exp\left(\frac{19.9T}{T+273}\right) \tag{2.1}$$

$$h = \frac{e_\mathsf{a}(T)}{e_\mathsf{s}(T)} \cdot 100\% \tag{2.2}$$

The resulting time series were compared. Local measurements in the WEC showed a high number of hourly values where the relative humidity was 99% or higher (33%). For Arsdorf this was 19% and for Roodt 12%. The ERA5 time series had the lowest percentage with 3% of the values being above 99% relative humidity. Furthermore, the relative humidity measured in Arsdorf often had a much higher daily maximum value compared to the other time series, while the daily minimums were similar. It was also noticed that the relative humidity measured in the WEC itself was often higher than that in the other datasets, even when excluding values above 99%. This is possibly caused by the same thermal insulating function of forests that affected the local temperature measurements (Renaud et al., 2011).

The Roodt and ERA5 datasets were selected to construct a continuous time series. The Arsdorf dataset was excluded because of its deviation compared to the other data. Due to its high number of values above 99% relative humidity and the suspected thermal insulating effect of the forest, data from the WEC was also excluded.

The continuous relative humidity time series for 2012–2022 was constructed using the same hierarchical gap-filling approach as used for the temperature. Again, direct observations (Roodt dataset) were prioritized over modelled data (ERA5). Therefore, data from the Roodt station was chosen as the primary dataset where gaps were filled using the ERA5 data.

Incoming shortwave radiation

Incoming shortwave radiation (R_s) is the total solar radiation arriving at the Earth's surface from the sun and the sky. This includes both direct and diffuse shortwave radiation. Diffuse shortwave radiation is sunlight that was scattered in the atmosphere before reaching the surface, while direct shortwave radiation is sunlight that directly reaches the surface without scattering.

R_s can be expressed in two ways, namely irradiance and irradiation. Irradiance is the short-wave radiation flux (radiant energy flow per unit time, e.g., in $W\,m^{-2}$) and irradiation is the integral of irradiance over a time period (typically an hour) and can be expressed in the unit $Wh\,m^{-2}$ (Page, 2012). In this study, R_s is expressed as irradiance in $W\,m^{-2}$.

The first dataset was obtained directly from LIST upon request and had a resolution of 5 minutes. It contained the incoming shortwave irradiance in $W\,m^{-2}$ for Roodt from 2018–2021. This data were resampled to an hourly resolution by taking the hourly mean. In this dataset, only the last two years (2020–2021) were actual measurements, measured using a CMP3 pyranometer (Kipp & Zonen B.V., Delft, The Netherlands). The first two years (2018–2019) were estimations based on sunshine duration and showed an overestimation for this period compared to the datasets described below.

The second dataset contains hourly global shortwave irradiation (reported as *Globalstrahlung*) in $Wh\,m^{-2}$ for 2012–2022. It was retrieved from ASTA (2025) for meteorological stations in Eschdorf, Harlange and Useldange, which are the closest stations where this data were available. Measurements at Harlange were available from early 2015 onwards. The data were converted to an hourly mean irradiance in $W\,m^{-2}$. Data from Useldange and Eschdorf showed minimal seasonality between 2012 and early 2017, while showing strong seasonality afterwards. Data from Harlange showed the expected seasonality.

The third dataset was the *Surface solar radiation downwards* dataset from the ERA5 reanalysis product and was retrieved from Hersbach et al. (2023) via the Copernicus Climate Change Service (C3S) Climate Data Store for the WEC area. The dataset spans 2012–2022 and provides hourly values of solar radiation accumulated over one hour, expressed in $J\,m^{-2}$. This data were converted to $W\,m^{-2}$.

From these datasets, the Roodt dataset was excluded from further usage because it contained two different methods and the exact moment where the method was changed is unknown. Furthermore, data from Eschdorf and Useldange between 2012 and February 2017 were excluded due to its lack of seasonality.

Harlange, Eschdorf, and Useldange are located around the WEC at similar distances: 11 km North, 11 km North-East, and 14 km South-East, respectively. To approximate R_s at the WEC, the inverse distance weighting method was applied (Equation 2.3). In this approach, the value at the target location (X_o) is estimated as a weighted average of m surrounding observations (X_i), where the weights are inversely proportional to the distance (r_i) raised to the power n. A power of $n = 1$ corresponds to linear inverse-distance weighting, which was used here. In cases of missing data from one or more stations, only the available stations were included in the calculation.

$$X_o = \frac{\sum_{i=1}^{m} \frac{X_i}{r_i^n}}{\sum_{i=1}^{m} \frac{1}{r_i^n}} \tag{2.3}$$

The continuous time series for R_s was created using the remaining data. In contrast to previous time series, no hierarchical gap-filling approach was used due to the differences in temporal data availability. Here, ERA5 data were used between 2012 and February 2015. From March 2015 until February 2017 data from Harlange were used. The remaining time span (March 2017 until December 2022) was filled with data resulting from the inverse distance approach described above. Remaining gaps were filled with ERA5 data.

Soil volumetric water content

Soil volumetric water content (θ) in $m^3\,m^{-3}$ was retrieved from Hissler et al. (2021). This dataset contained θ measurements from 7 locations (sites 7–13 in Figure 2.2b) at multiple depths using CS650 reflectometers (Campbell Scientific Ltd, Logan, UT, USA), with data recorded at 30-minute intervals. At site 7, θ was measured at a depth of 10 cm from 2011 to 2019. At sites 8–13, θ was measured at depths of 10, 20, 40 and 60 cm (θ_{10}, θ_{20}, θ_{40}, θ_{60}, respectively) from 2012 to 2019. A continuation of this data for 2020–2022 was obtained directly from LIST upon request. Both datasets were combined and resampled to an hourly resolution by taking the hourly mean, and outliers were removed. Because θ reacts relatively slowly, especially at greater depths, and data gaps generally had a short time span (order of days), gaps were filled using linear interpolation.

Per location, θ showed very similar temporal variation, with differences in baseline values between depths. Generally, θ was higher at shallower depths and fluctuated more. At greater depths, θ was more stable. The values were highest from autumn to spring and lowest during summer, when θ_{10} was sometimes lower than at the other depths.

In general, the different locations also showed very similar temporal variation, again with differences in the baselines. A potential explanation for this is the terrain, as the sites located at a lower altitude (e.g., site 9) were generally wetter than ones located further uphill (e.g., site 11). However, site 12 showed the highest θ values while being located uphill, contradicting this hypothesis.

Because the absolute θ values varied considerably between locations, averaging across sites for a given depth would neglect spatial variability and introduce bias. Instead, θ was spatially interpolated for each time step and depth using squared inverse distance weighting to approximate catchment-wide values (Equation 2.3 with n=2). Due to the spatial variability in slopes of the terrain, squared inverse distance was chosen over its linear version to put more weight on the distance. This results in continuous hourly time series for depths of 10, 20, 40, and 60 cm across the catchment.

Discharge

Discharge data were obtained in $m^3\,s^{-1}$ at a 15-minute interval from Hissler et al. (2021). They were measured using ISCO 4120 flow loggers (Teledyne ISCO, Lincoln, NE, USA) at multiple locations within the WEC (sites 2, 3, 15 & 16, see Figure 2.2b). Recordings at site 2 ranged from September 2009 until December 2019. For sites 3, 15, and 16 recordings ranged from September 2015 until December 2019. A continuation of this data for 2020–2022 was obtained directly from LIST upon request. Both datasets were combined and resampled to a daily resolution in $m^3\,d^{-1}$ using daily means.

The CAMELS-LUX dataset (Catchment Attributes and MEteorology for Large-sample Studies – LUXembourg) (Nijzink et al., 2025) was used for additional discharge data, ranging from November 2004 until October 2021. This dataset was mostly based on the same observation points as the data provided by Hissler et al. (2021) and contained interpolations for periods of missing data. For details on the interpolation methods, see Nijzink et al. (2025). The dataset had a 15-minute temporal resolution and was available in $m^3\,s^{-1}$. It was resampled to a daily resolution in $m^3\,d^{-1}$ by taking the daily mean.

It was noticed that discharge had a higher base flow and peaks during November–April, with May–October showing a low base flow with occasional higher peak flows, depending on the year. This was in line with the seasonality described previously in the study area description (Section 2.1).

The two discharge datasets were combined into a single continuous time series. Observations from Hissler et al. (2021) and LIST were used as the primary data, with gaps filled using information from Nijzink et al. (2025). The resulting continuous time series was converted from $m^3\,d^{-1}$ to specific discharge in $mm\,d^{-1}$ using the catchment area (0.45 km^2).

Precipitation

Precipitation data were obtained from Hissler et al. (2021) which were measured in Holtz between 2009 and 2019 in mm using an OTT Pluvio^2 L weighing precipitation gauge (OTT HydroMet GmbH, Kempten, Germany). The temporal resolution between 2009–2015 was 15 minutes, which was increased to 10 minutes between 2016 and 2019. Data from 2020–2022 were obtained directly from LIST upon request and had a reduced resolution of 15 minutes. Outliers were removed and both sets were resampled to a daily resolution by taking daily sums.

Additional precipitation data for Holtz were retrieved from ASTA (2025) which was available from 2015 onwards. The hourly data were resampled to daily total precipitation by taking the daily sum.

Both datasets showed similar precipitation patterns. Monthly precipitation was generally higher during winter than during summer, although the seasonal contrast was less pronounced than for discharge, as precipitation occurred throughout the year.

These datasets were combined into one continuous time series, where data obtained from Hissler et al. (2021) and LIST were used as the primary dataset. Gaps were filled using data for Holtz retrieved from ASTA (2025), resulting in a continuous daily time series from 2012–2022 in mm day^{-1}.

Potential evaporation

Numerous methods exist to estimate potential evaporation, and several studies have compared their performance across contexts (Oudin et al., 2005; Tanguy et al., 2018). A widely used method is the Penman-Monteith equation (Monteith, 1965) because of its physically-based mechanism. However, it has a high information requirement and is sensitive to errors in the input data. Furthermore, Zhao et al. (2013) compared different studies that analysed the sensitivity of several conceptual hydrological models and found that temperature-based methods often improved the models' performances compared to the Penman-Monteith method. For this reason, an alternative method was chosen.

Hamon (1961) developed a method based on mean air temperature and hours of daylight for a given day to calculate the potential evaporation. Two years later, this method was slightly updated by Hamon (1963), see Equation 2.4. The hours of daylight can be calculated, making the mean daily temperature its only variable. This method was chosen because Oudin et al. (2005) showed that it performs well for different conceptual models. Additionally, it was also used by Fenicia et al. (2008) to estimate the potential evaporation as input for their conceptual model of the Alzette river basin, of which the WEC is a sub-catchment, further supporting this choice.

Although a function of T, the calculation of E_p using Hamon involves several steps. The main equation is shown below as Equation 2.4. Here, the hours of daylight (N) and the saturated water vapour density (ρ_vs) have to be calculated.

The daily total potential evaporation was calculated in mm day^{-1} using the daily mean temperature from the continuous temperature time series.

$$E_\mathrm{p} = cN\rho_\mathrm{vs} \tag{2.4}$$

where:

E_p	Potential evaporation	[mm d^{-1}]
c	Constant (= 0.1651)	[−]
N	Hours of daylight	[h]
ρ_vs	Saturated water vapour density	[g m^{-3}]

The hours of daylight (N) were calculated using Equations 2.5–2.7. Here, Equations 2.5 and 2.6 were according to Allen et al. (1998), and Equation 2.7 according to Hodges and Ritchie (1991).

$$N = \frac{24}{\pi}\omega_\mathrm{s} \tag{2.5}$$

$$\omega_s = \arccos\left(-\tan\left(\phi\right)\tan\left(\delta\right)\right) \tag{2.6}$$

$$\delta = 0.409\sin\left(\frac{2\pi}{365}J - 1.39\right) \tag{2.7}$$

where:

N	Hours of daylight	[h]
ω_s	Sunset hour angle	[rad]
ϕ	Latitude (= 49.8 for WEC)	[rad]
δ	Solar decimation	[rad]
J	Day of the year	[−]

The saturated water vapour density ρ_{vs} was calculated using Equation 2.8 (Wiederhold, 2012). Here, the saturation vapour pressure e_s was calculated using Equation 2.1.

$$\rho_{vs} = \frac{216.7 e_s}{T + 273.16} \tag{2.8}$$

where:

ρ_{vs}	Saturation vapour density	[g m^{-3}]
e_s	Saturation vapour pressure	[kPa]
T	Temperature	[°C]

2.2.2. Vegetation data

In addition to hydro-meteorological and soil data, data on vegetation water use were also obtained. These variables provide insights into the role of trees in catchment-scale water fluxes and are essential for linking phenology with hydrological processes. They include sap flow velocity and transpiration.

Sap flow

The dataset used for sap flow velocity was retrieved from Klaus et al. (2024). This dataset contained the dasometric data of monitored trees, hourly sap flow velocity (cm h^{-1}), and tree surveys carried out in the WEC. In total, 36 trees were monitored during the 2019 and 2020 growing seasons. This included 3 Douglas Fir, 6 spruce, 13 oak, and 14 beech (36 trees in total). The tree locations are shown in Figure 2.1b.

The sap velocities were determined based on the heat ratio method (HRM) using SFM1 sap flow sensors (ICT International Pty Ltd, Armidale, NSW, Australia). The sensors consist of a central heating needle and two needles with thermistors to record the temperature upstream and downstream of the heater. The needles were positioned in a straight line vertically above each other with 0.5 cm spacing. The sensors were installed at breast-height (1.3 m above the ground). Further technical details of the sensors and the sap velocity calculations can be found in Fabiani et al. (2022) and Schoppach et al. (2021).

The sap flow velocity time series had an hourly resolution, and outliers were removed. Most series started in April–May 2019 and continued until December 2019, while in 2020, they generally ran from May to November, with some series missing for certain trees. Many time series also contained gaps ranging from days to months. A more detailed overview of start and end dates, as well as gaps, is provided in Appendix A, Table A.1.

Noticeable were the data gaps in most time series, especially during 2019. Furthermore, significant differences in sap flow velocity were noticed between 2019 and 2020 for a few trees. While some trees showed a strong increase in sap flow velocity (e.g. beech tree B2, see Figure 2.3), others showed a strong decrease. This was likely due to recalibration or replacement of measurement probes.

It was also noticed that the beech trees generally had a higher sap flow velocity than the other tree species. Furthermore, deciduous trees showed lower values early in the growing season, increasing over time with their maximum in July. Afterwards, the sap flow velocity started to decrease again. Some coniferous trees showed a similar temporal pattern but already had a higher rate at the start of the growing season.

The sap flow velocity (v_{sf}) of trees is influenced by tree-specific characteristics such as sapwood area. Larger trees often have a higher v_{sf} than smaller ones, and rates may also vary between tree species. These differences between trees complicated direct comparison. As described above, inspection of the data revealed that v_{sf} values for the same tree could differ substantially between years, likely due to recalibration or replacement of measurement probes.

To account for these inconsistencies, and to facilitate the analysis of v_{sf} responses to environmental conditions—as well as comparisons between individual trees and species—the v_{sf} time series were normalised per year using min-max scaling to a range between 0 and 1 (Equation 2.9). This approach has also been employed in related studies (Fabiani et al., 2022; Wang et al., 2020).

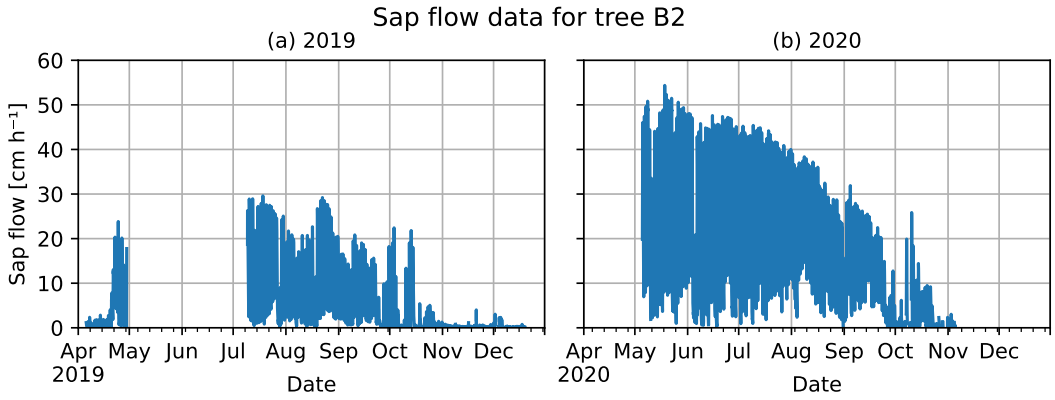


Figure 2.3: Sap flow velocity example plot: the sap flow velocity for tree B2 (beech) for 2019 and 2020.

$$X_{\mathrm{n},y} = \frac{X_y - X_{\mathrm{min},y}}{X_{\mathrm{max},y} - X_{\mathrm{min},y}} \tag{2.9}$$

where:

$X_{\mathrm{n},y}$	Normalised values in year y
X_y	Original values in year y
$X_{\mathrm{min},y}$	Minimum value in year y
$X_{\mathrm{max},y}$	Maximum value in year y

This normalisation approach assumed that both the minimum and maximum possible sap flow velocities occurred within each year's selection of data for each tree. For the minimum value, this was a reasonable assumption as the lowest value would be 0 cm h^{-1}, which was commonly observed in the dataset. The maximum, however, was less likely to be captured consistently within each year, especially as the data contains gaps for some trees. The real physiological maximum v_{sf} only occurs under ideal conditions, which were not likely to occur every year.

This first assumption was justified based on practical constraints and the goal of enabling comparisons of v_{sf} dynamics between different trees. Although the maximum possible rate may not have been observed each year, the scaled sap flow ($v_{\mathrm{sf,n}}$) still captured the shape and timing, which were considered to be the most important properties for further analysis.

A second assumption was that each tree's v_{sf} response remained comparable under similar environmental conditions across both years. This assumption may not always hold, as trees grow over time, increasing their sapwood area and potential maximum sap flow. Other factors such as disease or physical damage can also alter their v_{sf} capacity.

The second assumption was primarily justified based on practical constraints. Moreover, since the v_{sf} dataset only spans 2 years, no significant changes in tree sizes were expected. Additionally, scaling was applied separately for each year, which helped to minimize the influence of gradual physiological changes. By normalising per year, the focus was on relative dynamics within each year, rather than on the absolute difference across years.

Transpiration

Transpiration data were obtained from the NASA Global Land Data Assimilation System (GLDAS, (Li et al., 2020)) for the WEC area between 2012 and 2022. GLDAS used satellite and ground-based observational data (data assimilation) and land surface models to simulate processes like water and energy exchange between land and the atmosphere. It had global coverage at a high resolution of 0.25°.

GLDAS-2 has three sub-versions. Here, version 2.0 uses only meteorological fording data and covers 1948–2014. Version 2.1 combines model and observational data, starting in 2000. Version 2.2 utilizes data assimilation: the total terrestrial water anomaly observation from Gravity Recovery and Climate

Experiment (GRACE) was assimilated (Li et al., 2019). This study used version 2.2 because of the inclusion of data assimilation. For more detailed information about the working of GLDAS-2.2, see Li et al. (2020).

The obtained transpiration data had an hourly temporal resolution in kg m^{-2} s^{-1}. This was converted to mm d^{-1}, assuming the density of water to be 1000 kg m^3.

The data showed strong seasonality, with E_t being near zero between November and March. Most years, E_t started to increase between the middle and end of March and went to zero between October and November. Furthermore, it showed a strong fluctuation between April and October with sharp peaks.

The transpiration data had already been converted to mm d^{-1}. Furthermore, the dataset contained no gaps and had a daily resolution. Therefore, no additional processing was required for future usage.

2.3. Conventional FLEX model

There are many different conceptual hydrological models with various levels of complexity. For this research, the catchment was modelled using the base structure of the FLEX$_{\text{I}}$ (Flux Exchange) model described by Fenicia et al. (2008), which was based on Fenicia et al. (2006). The base structure is explained below in Section 2.3.1.

This research made a distinction between deciduous and coniferous trees in the catchment based on their different phenology (see 2.4.1 Tree categorisation). For this reason, the lumped model structure was reconfigured into a semi-distributed framework. This is explained in Section 2.3.2.

This semi-distributed model served as a reference model, showing how a conventional model structure behaves when phenology is not included. This made it possible to evaluate the influence of phenology inclusion in the modified models (see Section 2.4). It should be mentioned that this model was influenced more by vegetation coverage than the original (lumped) model because the semi-distributed structure was based on vegetation cover. However, phenology was still not accounted for, as the vegetation parameters were constant over time. Thus, the model was considered to serve sufficiently well as a reference model.

It was chosen to not use a lumped model of the catchment as a reference because the process of modifying it from lumped to semi-distributed may also influence the outcomes, limiting the analysis of the phenology integration influence itself.

The model was run at daily time steps. This was preferred over hourly steps, as it reduces the total number of steps by a factor of 24, resulting in a much faster runtime. Reducing runtime is particularly important for the calibration method used (see Section 2.5 Model calibration).

The model took precipitation (P) and potential evaporation (E_p) in mm d^{-1} as input and simulated the discharge (Q) in mm d^{-1}.

2.3.1. Base structure

The FLEX$_{\text{I}}$ (Flux Exchange) model described by Fenicia et al. (2008) is a lumped conceptual model that is composed of four interconnected reservoirs which indicate various water storages and their scales. These reservoirs are the interception reservoir (IR), the unsaturated soil reservoir (UR), the fast reacting reservoir (FR), and the slow reacting reservoir (SR).

Interception reservoir

The interception reservoir IR takes interception processes into account. Precipitation first runs through this reservoir. The storage of IR is referred to as S_i with a maximum capacity of I_{max} in mm, which value needs to be determined through calibration. After the precipitation P is added to S_i and S_i reaches I_{max}, excess precipitation results in throughfall P_{tf} (Equation 2.10). During periods of no precipitation, water evaporates from S_i at a rate that is either limited by E_p or S_i (Equation 2.11).

$$P_{tf} = \begin{cases} 0 & \text{if } S_i \leq I_{max} \\ S_i - I_{max} & \text{if } S_i > I_{max} \end{cases} \tag{2.10}$$

$$E_i = \begin{cases} 0 & \text{if } P > 0 \\ \min\left[S_i, E_p\right] & \text{if } P = 0 \end{cases} \tag{2.11}$$

Unsaturated reservoir

To determine the flow into UR, throughfall is partitioned into various components using an effective runoff coefficient C_r. This coefficient is expressed as a S-shaped function (Equation 2.12) that depends on the ratio between current storage of UR (S_u) and its maximum storage ($S_{u,max}$), and where β is a shape factor (Figure 2.4). Values of both $S_{u,max}$ and β are determined through calibration. The part of P_{tf} that infiltrates into UR (recharge of UR: R_{ur}) is determined by Equation 2.13. When UR is filled to capacity, the excess water is routed to FR.

$$C_r = \frac{1}{1 + \exp\left(\frac{-S_u/S_{u,max} + 0.5}{\beta}\right)} \tag{2.12}$$

$$R_u = (1 - C_r) \cdot P_{tf} \tag{2.13}$$

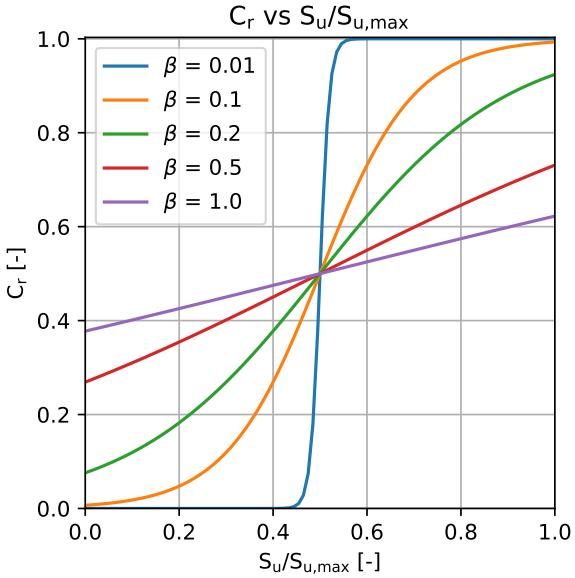


Figure 2.4: The relationship between the effective runoff coefficient (C_r) and relative soil moisture ($S_u/S_{u,max}$) for different values of shape parameter β.

There are two outflows for UR. The first one is percolation P_s, which is modelled as a linear function of the relative soil moisture (Equation 2.14) and is routed to SR. The maximum percolation rate ($P_{s,max}$) is found through calibration.

$$P_s = \left(\frac{S_u}{S_{u,max}}\right) \cdot P_{s,max} \tag{2.14}$$

The second outflow is transpiration (E_t). In Fenicia et al. (2008), E_t was modelled using Equation 2.15, regarding E_t to be independent of E_i. Here, C_e represents the limit of S_u relative to $S_{u,max}$ below which vegetation experiences water stress. When S_u is above $S_{u,max} \cdot C_e$, soil moisture is not limiting and transpiration can occur at the rate of potential evaporation. The value of C_e is determined by calibration.

$$E_t = \min\left[1, \frac{S_u}{S_{u,max} \cdot C_e}\right] \cdot E_p \tag{2.15}$$

Fast and slow reacting reservoir

The remaining part of P_{tf} that did not infiltrate is partitioned into preferential recharge R_{sr} to SR and runoff R_{fr} routed to FR. This is done using a runoff partitioning coefficient D, resulting in Equations 2.16 and 2.17.

$$R_{sr} = (P_{tf} - R_{ur}) \cdot D \tag{2.16}$$

$$R_{fr} = (P_{tf} - R_{ur}) \cdot (1 - D) \tag{2.17}$$

Originally, the FLEX$_I$ model by Fenicia et al. (2008) included two lag functions, one for each reservoir. This introduced two additional parameters that required calibration, increasing the model's equifinality issue. Considering the daily time steps of the model and the small size of the WEC (0.45 km^2), it was therefore decided not to include time lags.

Both SR and FR are modelled as linear reservoirs, meaning their discharge is linearly related to their storage (S_s and S_f, respectively). This is done using Equations 2.18 and 2.19 where K_s and K_f are the timescales of the slow and fast reservoirs respectively. A summation of the slow discharge Q_s and fast discharge Q_f results in the total modelled (or simulated) discharge Q_m (Equation 2.20).

$$Q_s = \frac{S_s}{K_s} \tag{2.18}$$

$$Q_f = \frac{S_f}{K_f} \tag{2.19}$$

$$Q_m = Q_s + Q_f \tag{2.20}$$

2.3.2. Model coupling

The original model described by Fenicia et al. (2008) is a lumped model. As described in Section 2.1 (Study area), the WEC consists of a forest subdivided into deciduous and coniferous parts. This distinction is important because the study focuses on integrating phenology, and these tree categories exhibit different phenological behaviour. For this reason, the original model was modified into a semi-distributed model. The multiple coniferous stands across the catchment were combined into a single lumped coniferous component (see Section 2.4 for details).

The semi-distributed model had a parallel structure based on the FLEX^T model by Gao et al. (2014). Two lumped models were run in parallel, except for the slow reservoir (SR), which was considered to be shared. Consequently, both parts of the model shared the same structure and equations but with differences in parameter values to allow for vegetation specific differences. The resulting model structure is shown in Figure 2.5.

$$Q_m = Q_s + Q_f \tag{2.21}$$

The lumped model has eight parameters that need to be determined by calibration. Using the structure suggested above for the semi-distributed model, this would have resulted in a total of fifteen parameters to calibrate (K_s was shared). Increasing the number of calibration parameters increases the issue of equifinality and calibration complexity.

Most of these parameters were not (strongly) dependent on vegetation type but were mainly soil or terrain characteristics. Because of the small size of the catchment and because coniferous stands were located across the catchment, these parameters were considered to be catchment-wide parameters which did not depend on the vegetation category. Therefore, only I_{max} and C_e were defined as specific to vegetation. This reduced the total of parameters to ten.

I_{max} was considered to be vegetation specific because of the difference in leaves and canopy structure between deciduous and coniferous trees. C_e influenced when and how much E_t was limited by soil moisture, which depended on the vegetation and could have been influenced by multiple factors (e.g., root depth, maximum transpiration rate).

Hereafter, the coupled semi-distributed model described above is referred to as the *conventional model*.

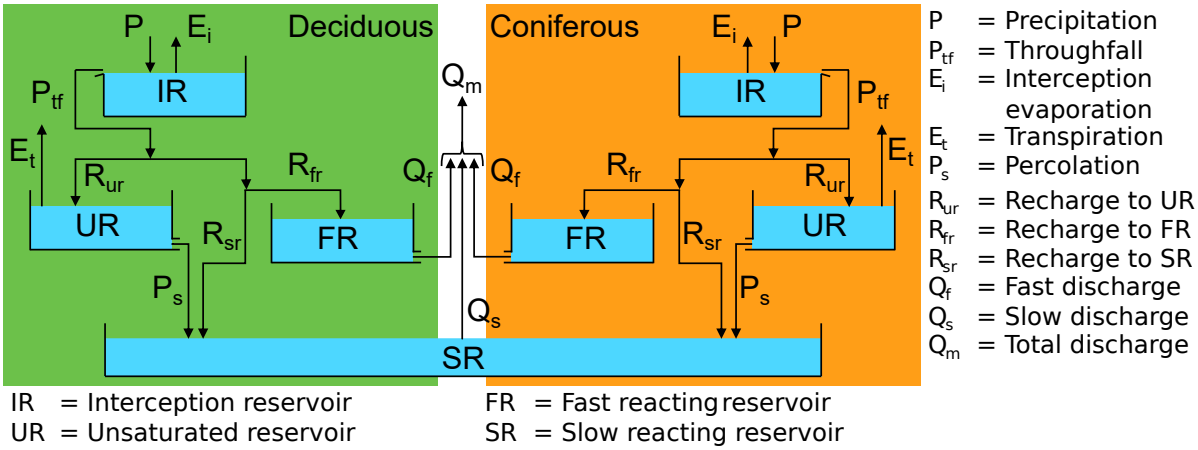


Figure 2.5: Semi-distributed structure of the conventional model. Deciduous and coniferous sub-catchments each use the same lumped model structure with partially differing parameters, sharing a single slow reservoir. Total discharge is calculated as the sum of the two fast discharges and the shared slow discharge.

2.4. Phenology inclusion

The conventional model described in Section 2.3 did not include phenological dynamics. It had a semi-distributed structure based on tree category (deciduous/coniferous). The reasoning behind this is first described in Section 2.4.1.

To include phenology, three methods were designed. The first one involved using normalised sap flow data, replacing the relative soil moisture factor used in the transpiration equation (Equation 2.15). This is explained in Section 2.4.2. The second method used the growing degree-day as an indication of the annual development stages of deciduous trees, which was used to dynamically partition potential evaporation E_p in interception evaporation E_i and transpiration E_t. This is explained in Section 2.4.3. Finally, the third method combined the sap flow and dynamic partitioning methods and is explained in Section 2.4.4.

2.4.1. Tree categorisation

As described in Section 2.1 Study area, there are four main tree species (beech, oak, Douglas fir and spruce). Although both deciduous trees, beech and oak trees each have their own water use strategy. Fabiani et al. (2022) found that the beech trees in the WEC exploit a shallower and seasonally less persistent water source than oak trees. This was attributed to the shallower root system of beech trees compared to the oaks. However, because the beech and oak are completely mixed, no clear distinction could be made in terms of modelling. Therefore, the mixed beech and oak forest was lumped into one deciduous forest.

Douglas fir and spruce trees were lumped into a single conifer category due to the limited number of trees with sap flow measurements (3 Douglas fir and 6 spruce), compared to the deciduous trees (14 beech and 13 oak). Although Douglas fir and spruce may differ in phenology, root structure, and water-use strategies, their physiological contrast with deciduous trees is likely to be more influential, as suggested by Schoppach et al. (2021). Furthermore, coniferous trees accounted for 24% of the catchment area. Dividing them into separate categories would likely result in minimal influence of each species on the catchment's total discharge.

As described above, the model was divided into two vegetation categories (deciduous and coniferous) to capture key phenological differences, while avoiding an overly detailed representation that would have been impractical given the available data and likely minimal influence on catchment-scale results.

2.4.2. Sap flow method

As described in the introduction (Section 1.1.2), transpiration is closely linked to photosynthesis. Photosynthesis requires the stomata to open for CO_2-uptake, during which water is lost via transpiration. To compensate for this loss, water is transported via the xylem from the roots to the canopy, resulting in a sap flow. In this study, sap flow velocity—the speed of sap movement in the xylem—is hereafter referred to simply as sap flow (SF or v_{sf}) for brevity, except where the full term is needed for clarity. This means that transpiration and sap flow are positively correlated: when transpiration is high, the sap flow will also be high (Köstner et al., 1992). For this reason, transpiration and sap flow have the same general pattern.

Under the same atmospheric conditions, different trees can have different sap flow velocities. For example, the diameter at breast height, landscape characteristics have been identified as drivers for tree-to-tree variability (Fabiani et al., 2022; Schoppach et al., 2021). Schoppach et al. (2021) also found indications that larger oaks accessed deeper water sources than the smaller ones. This complicates the upscaling of sap-flow data from individual trees to stand- or even catchment-scale.

The conventional hydrological model used relative soil moisture and potential evaporation to calculate transpiration (Equation 2.15). In this formulation, E_t equalled E_p under non-water-stressed conditions, since E_p was multiplied by one (i.e. stomatal resistance is neglected). This meant that E_t occurred at its maximum rate. When no soil moisture was available for vegetation, E_p was multiplied with a (near-)zero value, resulting in E_t that was also (near-)zero. In this first method of phenology inclusion, it was hypothesised that the normalised sap flow velocity ($v_{sf,n}$) could be used as a replacement of the factor that is multiplied with E_p.

The sap flow was scaled between 0–1 using Equation 2.9, resulting in $v_{sf,n}$. So, the maximum sap flow value was converted to 1. Normalisation was applied separately for each year, assuming that the true maximum v_{sf} occurred within that year. While this assumption may not always hold, it helped account for physiological changes in trees. For example, the maximum v_{sf} increase when trees grow larger. If one would normalise over a time series spanning multiple years, the maximum rate would have never been reached in the early years because the tree itself was smaller. Another scenario could be that a tree was damaged, reducing its maximum v_{sf}. By applying normalisation per year, these changes affected the other years in the time series to a lesser extent, as the scaling adapted to the new situation (i.e., new maximum rate).

Next to physiological changes to the trees themselves, inconsistencies were observed in the measurement of v_{sf}. Dix and Aubrey (2021) found that most tree sap flow publications did not apply or use calibration, reducing the accuracy of these estimates. Therefore, the actual values of these measurements may be inaccurate, but by normalising v_{sf}, the general response dynamics to a changing environment were still usable.

Another benefit of using $v_{sf,n}$ over v_{sf} was the possibility to compare and use data from different trees simultaneously. Because v_{sf} is strongly influenced by the characteristics of the trees themselves, such as its size, it is difficult to upscale. Taking the mean v_{sf} of multiple trees per time step does not necessarily reflect the mean v_{sf} over the catchment. However, it is likely that v_{sf} of multiple trees shows the same relative response, meaning it increases and decreases at the same moments when better or worse conditions present themselves. In other words, they show the same response but not in absolute values. By normalising v_{sf} to $v_{sf,n}$, the response in v_{sf} of different trees to environmental changes could be compared in order to upscale to catchment size.

Coniferous and deciduous trees showed different transpiration rates during a year. As explained in the Introduction (Section 1.1.2), and also observed in Section 2.2.2, coniferous trees had a higher E_t at the start of the growing season, as deciduous trees were still developing their leaves during this time. Therefore, both categories each had their own $v_{sf,n}$ input.

To implement this inclusion of $v_{sf,n}$ in the conventional model, only the transpiration equation (Equation 2.15) was modified. The term related to the relative soil moisture was replaced by $v_{sf,n}$. Here, $v_{sf,n}$ was the $v_{sf,n}$ of the corresponding part of the model (deciduous/coniferous).

$$E_t = v_{sf,n} \cdot E_p \tag{2.22}$$

Due to the modification, this model required two v_{sf} time series as forcing data, one deciduous and one coniferous, in addition to the standard precipitation and potential evaporation forcing. Additionally, the vegetation specific calibration parameter C_e was removed, reducing the total number of calibration parameters from ten to eight.

Hereafter, this method is referred to as the *SF method*.

2.4.3. Evaporation partitioning method

Deciduous trees have a strong seasonality. They start the year with no leaves. During the spring, when temperatures rise, they come out of dormancy and begin developing their leaves. There is a substantial variation in timing for budburst and leaf-out within and between tree species, but the leaf development of most temperate trees is highly sensitive to temperature (Polgar and Primack, 2011). These events generally occur earlier during warmer years than during cooler years, making accumulated temperature a useful indicator to predict the timing of these events.

Accumulated growing degree-days

The growing degree-days (T_{GDD}) method is frequently used to estimate and describe the timing of biological processes such as leaf development. It is more accurate than counting calendar days due to the strong influence of temperature. T_{GDD} is usually calculated using Equation 2.23. Here, T_{max} is the daily maximum air temperature, T_{min} the daily minimum air temperature and T_{base} the threshold temperature below which the biological processes do not progress (McMaster and Wilhelm, 1997), in this case leaf development. The value of T_{base} is species specific. For beech and oak, a value of 5 °C is often used (Fu et al., 2019).

$$T_{GDD,avg} = \left(\frac{T_{max} + T_{min}}{2} \right) - T_{base} \tag{2.23}$$

Equation 2.23 uses the daily minimum and maximum temperatures to assess whether the temperature exceeded the threshold, and thus whether leaf development occurred. If the minimum temperature is low enough, no development would be expected, while it is possible that temperatures during the afternoon exceeded T_{base}. Therefore, in this study a slightly modified version of Equation 2.23 was used.

Here, T_{GDD} was calculated using an hourly integration approach, comparing the hourly temperature measurement T_i with T_{base}. For each hour, the temperature difference $\Delta T = T_i - T_{base}$ was computed. Negative values were excluded, as temperatures below the threshold do not contribute to leaf development. The daily T_{GDD} value was then obtained by summing all positive ΔT values and dividing by 24, converting the units from [°C h] to [°C d]. This procedure is equivalent to a Riemann sum approximation of the continuous-time integral. The resulting equation is shown below as Equation 2.24.

$$T_{GDD} = \frac{1}{24} \sum_{i=1}^{24} \max\left[0, T_i - T_{base}\right] \tag{2.24}$$

Using an hourly time scale to calculate T_{GDD} results in different outcomes compared to the daily scale from the original method. This is especially true when the minimum temperature is below T_{base}, as shown in Figure 2.6. When the daily minimum temperature is above T_{base}, both methods show similar results.

Accumulated T_{GDD} (T_{CGDD}) is the summation of T_{GDD} over time, starting from 1 January. It can serve as a proxy for the timing of certain stages in plant development, such as the budburst and leaf-out of deciduous trees (Crimmins and Crimmins, 2019). There are no exact T_{CGDD} values at which these events occur, as they are also influenced by other factors (e.g., the chilling requirement, where a warmer winter can delay the start of development (Dantec et al., 2014)). Furthermore, T_{CGDD} thresholds are species-specific.

For the reasons mentioned above, this study did not define specific T_{CGDD} thresholds for the start of different stages in tree development. Instead, the general S-shaped pattern of T_{CGDD} was used. At the start of the year, T_{GDD} accumulation was low due to cold temperatures. As temperatures rose, T_{GDD} increased more rapidly, producing a steeper slope of T_{CGDD} (higher first derivative). The rate of increase

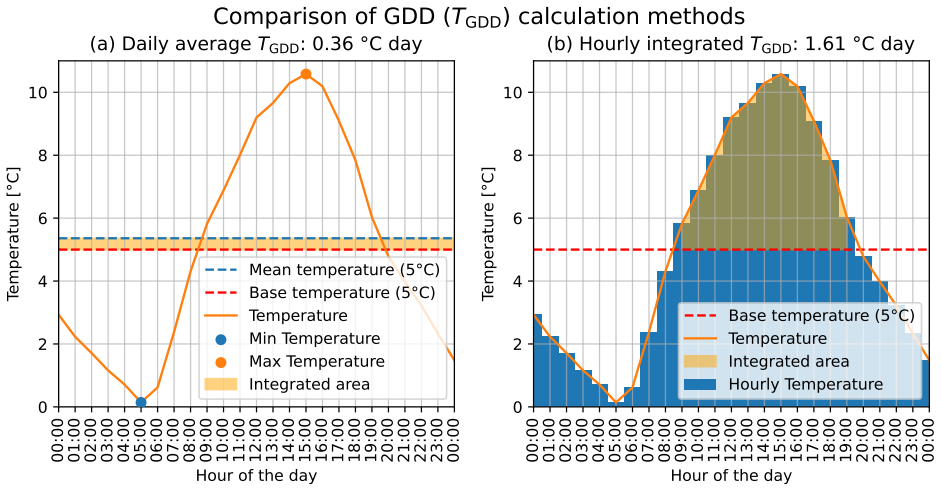


Figure 2.6: Visualisation of two different methods to determine Growing Degree-Days (GDD). (a) The original method using an daily average. (b) The Hourly GDD method used in this study.

then decreased (second derivative decreased), which was assumed to correspond approximately to full leaf development. Midway through the summer, temperatures began to decline, reducing T_{GDD} and flattening the slope of the T_{CGDD} curve. The peak of the first derivative was considered the tipping point, and the subsequent minimum of the second derivative marked the approximate end of the growing season, after which vegetation activity, such as transpiration, slowed.

The T_{CGDD} over the year has a shape similar to a logistic curve, a common sigmoid function (Equation 2.25) with an analytical solution. Therefore, observed temperature can be used to calculate T_{GDD}, which in turn can be used to compute T_{CGDD}, onto which a scaled sigmoid function can be fitted. As a result, the first and second derivatives can be determined analytically (Equations 2.26 and 2.27), as shown in Figure 2.8. This allows an explicit definition of the tipping points described above based on the local minima and maxima of the first and/or second derivative.

$$f(t) = \frac{L}{1 + e^{-k(t-t_0)}} \tag{2.25}$$

$$\frac{\mathrm{d}f}{\mathrm{d}t}(t) = \frac{Lke^{-k(t-t_0)}}{\left(1 + e^{-k(t-t_0)}\right)^2} \tag{2.26}$$

$$\frac{\mathrm{d}^2 f}{\mathrm{d}t^2}(t) = -\frac{Lk^2 e^{-k(t-t_0)}}{\left(1 + e^{-k(t-t_0)}\right)^2} + \frac{2Lk^2 e^{-2k(t-t_0)}}{\left(1 + e^{-k(t-t_0)}\right)^3} \tag{2.27}$$

where:

$f(t)$	Logistic function (T_{CGDD} curve)	[°C d]
L	Maximum value (scaling parameter)	[°C d]
k	Steepness of the curve	[d^{-1}]
T	Time	[d]
t_0	T-value of the function's midpoint	[d]

During a warm year, the fitted function has a higher value for L, a lower value for k, and a smaller t_0, resulting in an earlier start and a longer growing season. In cooler years, the opposite occurs.

Partitioning factor K_v

The Food and Agriculture Organization (FAO) has presented guidelines to calculate the crop evaporation ($E_{t,c}$) where a reference crop evaporation ($E_{t,0}$) is multiplied with a crop coefficient K_c, shown in Equation 2.28 (Allen et al., 1998). Here, $E_{t,0}$ is the atmospheric water demand for a reference crop and therefore similar to E_p. The FAO refers to this method as the *single crop coefficient approach*.

$$E_{t,c} = K_c \cdot E_{t,0} \tag{2.28}$$

By using a K_c that depends on vegetation cover, $E_{t,0}$ can be used to determine the transpiration of a different crop, as K_c is crop specific. The value of K_c also changes over time, depending on the growing stage of a crop. Early in the season, K_c will be low as the crop is still small and other evaporation terms will be dominant. Mid-season, transpiration is higher as the crops are larger, increasing its relative importance compared to other evaporation terms. This is shown in Figure 2.7.

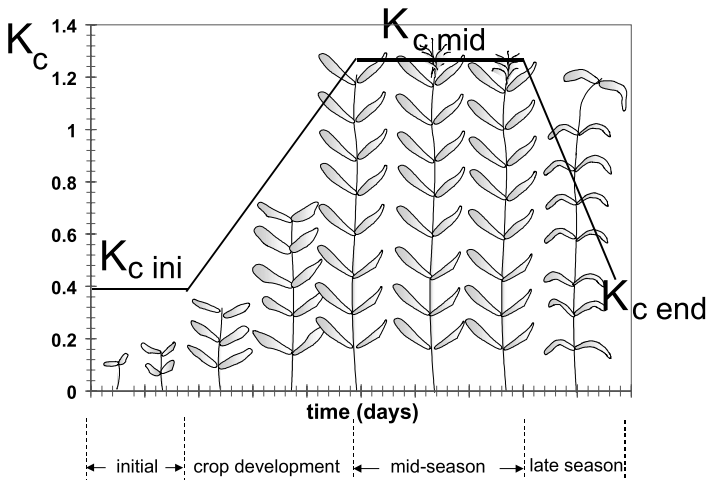


Figure 2.7: The FAO generalised crop coefficient (K_c) curve for the single crop coefficient approach (Source: Allen et al. (1998)).

Mert (2021) and Pierik (2022) modified this approach by using the leaf area index (LAI) as an indication for the growing stages of vegetation in a catchment. The values of K_c for different growing stages were determined through model calibration. Next, K_c was used to partition E_p between E_i and E_t. To do so, E_p was replaced by $[(1 - K_c(t)) \cdot E_p]$ in Equation 2.11. In Equation 2.15, E_p was replaced by $[K_c(t) \cdot E_p]$.

One downside of using LAI is the limited spatial and temporal resolution, as it is usually determined using remote sensing. Because the WEC is small and the coniferous stands are located within the deciduous forest, it is not possible to reliably obtain LAI data for the tree categories separately.

Therefore, instead of LAI, the T_{CGDD} curve was used to approximate the growing stages. The corresponding values of each growing stage, previously defined as K_c, was replaced by K_v for clarity, indicating vegetation instead of crops. The shape of the first derivative of T_{CGDD} was used for the shape of the K_v curve.

Instead of using a constant value during the initial stage and a linear interpolation during crop development, the initial and crop development stages were combined into a single developing stage. The first derivative of T_{CGDD} defined the shape of this stage. Although it technically started at 1 January with zero, the curve remained near zero until temperatures began to rise, at which point T_{GDD} and consequently the T_{CGDD} slope increased. The mid-season stage began when the second derivative of T_{CGDD} reached its maximum, represented by a plateau. This was assumed to correspond to full leaf development in deciduous trees, resulting in the maximum K_v. After this point, the mid-season ended, and the K_v curve decreased following the shape of the first derivative of T_{CGDD}, approaching near zero.

Next, the K_v curve was normalised for each year, scaling it between 0 and 1 using Equation 2.9. The resulting curve is the normalised K_v shape function, which is shown in Figure 2.8.

The actual K_v values ($K_{v,a}$) were determined using Equation 2.29. Here, $K_{v,max}$ is the maximum K_v value that was reached during the mid-season and was obtained through calibration. This allowed K_v to be transformed into $K_{v,a}$ without changing the seasonality shape.

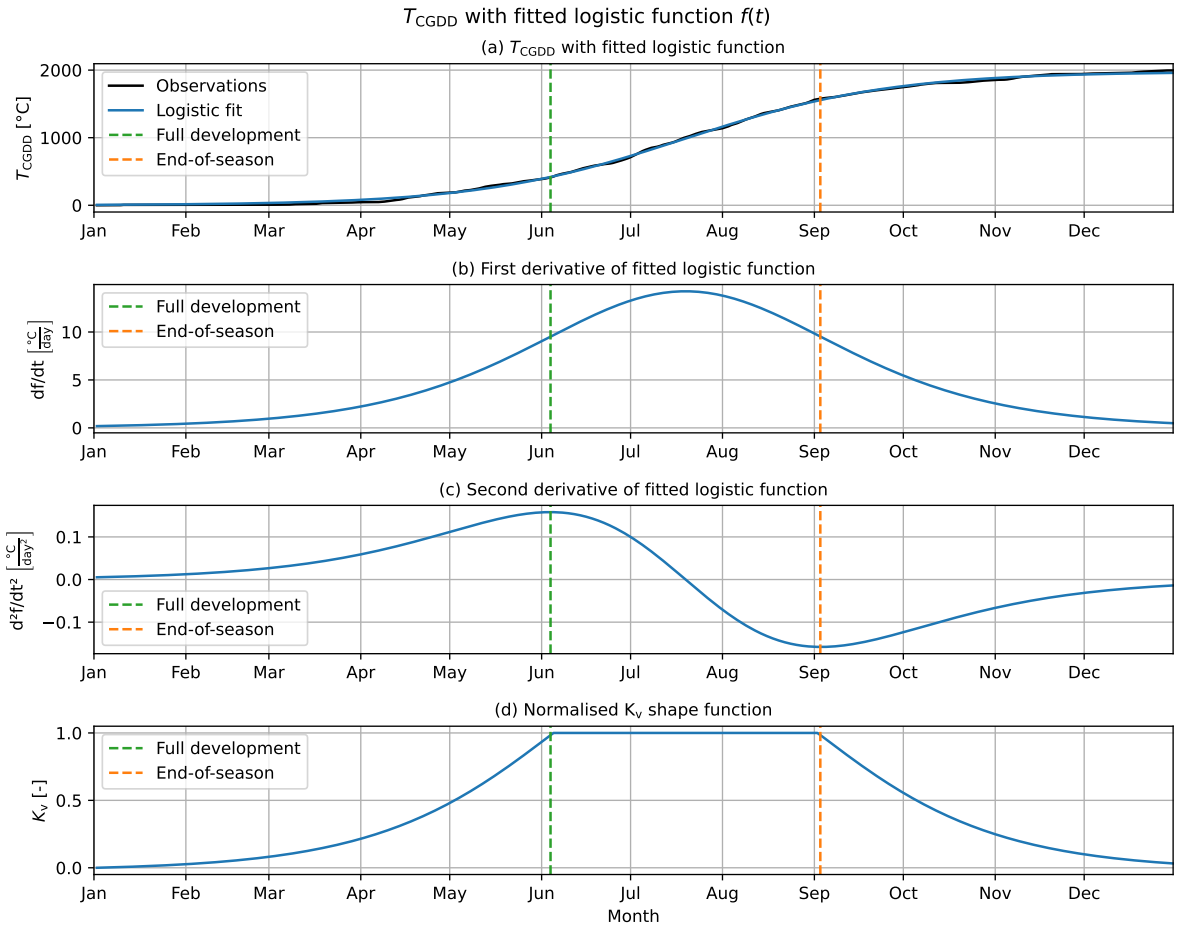


Figure 2.8: Overview of the functions used to derive the normalised K_v shape function. (a) Fitted logistic function on the T_{CGDD} curve. Its first derivative is shown in (b) First derivative of the logistic function. (c) Second derivative of the logistic function. (d) Normalised K_v shape function.

The method described above was only applied for deciduous trees as they have a strong seasonality due to their annual phenological cycle of growing and shedding leaves. Coniferous trees also have an annual phenological cycle, but its influence on E_t is weaker because they do not shed their needles or enter full dormancy like deciduous species. For this reason, this method strongly simplified the coniferous seasonality by assigning a constant K_v during the year. $K_{v,a}$ was determined in the same way as for deciduous trees, using Equation 2.29 with its own $K_{v,max}$ and constant K_v.

$$K_{v,a} = K_v \cdot K_{v,max} \tag{2.29}$$

Finally, $K_{v,a}$ was implemented similarly to the approaches of Mert (2021) and Pierik (2022). Equations 2.30 and 2.31 show the modified functions for E_i and E_t, which were incorporated into the conventional model. This method required one additional forcing time series (K_v) and two extra calibration parameters ($K_{v,max}$), one for each tree category, resulting in a total of 12 calibration parameters.

Hereafter, this method is referred to as the *K_v method*.

$$E_i = \begin{cases} 0 & \text{if } P > 0 \\ \min\left[S_i, (1 - K_{v,a}) \cdot E_p\right] & \text{if } P = 0 \end{cases} \tag{2.30}$$

$$E_t = \min\left[1, \frac{S_u}{S_{u,max} \cdot C_e}\right] \cdot (K_{v,a} \cdot E_p) \tag{2.31}$$

2.4.4. Combined method

The third method to include phenology into the conceptual model was a combination of the SF method and K_v method. In summary, the SF method used $v_{sf,n}$ to replace the relative soil moisture factor in the original E_t equation (Equation 2.15) in order to represent the transpiration more accurately, directly using sap flow dynamics. In the K_v method, $K_v(t)$ was used to partition the availability of E_p between E_i and E_t, indirectly using sap flow dynamics. It used T_{CGDD} to approximate the growing stages and corresponding vegetation coefficient K_v. At the start of the year, K_v was low, resulting in a small fraction of E_p being allocated to E_t, as E_i was dominant. During mid-season, E_t was more dominant, requiring a larger fraction of E_p, resulting in a lower E_i.

These methods worked at different parts of the equations for a different reasoning. The third method, referred to as the *Combined method*, integrates both approaches. Both $v_{sf,n}$ and K_v represent seasonality by multiplying E_p by a factor between 0 and 1. Applying both directly would overly restrict E_p, as multiplying two values between 0 and 1 produces an even smaller result. To avoid this, the mean value of K_v and $v_{sf,n}$ was used. The modified equations are shown in Equations 2.32 and 2.33.

$$E_i = \begin{cases} 0 & \text{if } P > 0 \\ \min\left[S_i, \left(1 - \frac{v_{sf,n} + K_{v,a}}{2}\right) \cdot E_p\right] & \text{if } P = 0 \end{cases} \tag{2.32}$$

$$E_t = \left(\frac{v_{sf,n} + K_{v,a}}{2}\right) \cdot E_p \tag{2.33}$$

The combined method had ten calibration parameters, the same number as the conventional model. The original vegetation specific calibration parameter C_e was replaced by $K_{v,max}$, where both were defined per tree category (deciduous/coniferous).

2.5. Model calibration

Conceptual models usually contain parameters whose values need to be obtained during calibration. In this case, calibration is the process of adjusting the calibration parameters in order to align the model outcomes more closely with observed outcomes or real-world data. This is often done by using objective functions to assess the performance of the model, where the outcome of the objective function must be maximised or minimised, depending on its formulation. Instead of using one objective function, it is also possible to use multiple functions, called multi-objective calibration (also called multi-objective optimization).

2.5.1. Objective function

A widely used method to assess model performance in hydrology is the Nash-Sutcliffe efficiency metric (NSE, or N_{NSE}), introduced by Nash and Sutcliffe (1970). It quantifies the degree to which the simulated discharge from the model (Q_m) matches the observed discharge (Q_o) using the mean observed discharges ($\overline{Q}_o$) as shown in Equation 2.34. The NSE ranges from $-\infty$ to 1, with 1 indicating a perfect match, 0 indicating that the model performs no better than using the mean observed discharge, and values below 0 indicating performance worse than the mean. According to Ritter and Muñoz-Carpena (2013), values between 0.65 and 0.8 indicate acceptable performance, 0.8–0.9 good performance, and 0.9–1.0 very good performance.

$$N_{NSE} = 1 - \frac{\sum_{i=1}^{n} \left(Q_{o,i} - Q_{m,i}\right)^2}{\sum_{i=1}^{n} \left(Q_{o,i} - \overline{Q}_o\right)^2} \tag{2.34}$$

The NSE is highly sensitive to large deviations and consequently strongly influenced by high peaks. Therefore, using only NSE during calibration results in a model that is mainly fitted to peak flows, possibly resulting in a model that has problems predicting low flows. The logarithmic version of N_{NSE} (logNSE, or N_{logNSE}) applies a natural logarithm transformation on both observed and simulated values prior to calculating the model performance. This reduces the influence of peak flows, increasing the model's accuracy for low flow. Due to the usage of a logarithm, input values should be positive.

Discharge observations showed that the WEC had a low flow during the summer with occasional peak flows. Discharge was higher during winter with more peak flows. Due to the occurrence of both high

and low flows, both N_{NSE} and N_{logNSE} were used in a multi-objective function during calibration.

A third metric was included in the objective function to assess how well the model partitioned precipitation in runoff and evaporation. The monthly runoff coefficient (C_{mr}) expressed the fraction of precipitation that results in discharge (Equation 2.35). The remaining part of the precipitation either evaporated or resulted in storage change. The NSE was computed using the observed and simulated monthly C_{mr} values, providing a measure of the model's ability to reproduce the temporal variability in precipitation-runoff partitioning. High $N_{\text{NSE,C}_{\text{mr}}}$ values indicated that the model accurately represented the seasonal distribution of precipitation into discharge versus evaporation and storage changes, independent of the absolute magnitude of precipitation.

$$C_{\text{mr}} = \frac{\overline{Q}}{\overline{P}} \tag{2.35}$$

The multi-objective function was calculated using the Euclidean distance in a three-dimensional performance space, using $N_{\text{NSE,Q}}$, $N_{\text{logNSE,Q}}$ and $N_{\text{NSE,C}_{\text{mr}}}$. Using these three performance indicators assesses the model on peak flow, low flow and seasonal precipitation-runoff partitioning performance simultaneously. In order to minimise the multi-objective function, each individual metric was subtracted from 1, resulting in the multi-objective function F_{obj} shown in Equation 2.36.

$$F_{\text{obj}} = \sqrt{\left(1 - N_{\text{NSE,Q}}\right)^2 + \left(1 - N_{\text{logNSE,Q}}\right)^2 + \left(1 - N_{\text{NSE,C}_{\text{mr}}}\right)^2} \tag{2.36}$$

2.5.2. Calibration procedure

Model calibration was performed using a Monte Carlo simulation approach. For each calibration parameter, a realistic range of values was defined. During each Monte Carlo iteration, a random value for each parameter was drawn from a uniform distribution within its predefined range using a fixed random seed to ensure reproducibility. The range is shown per parameter and per model in Table 2.1.

The model was run for 2012–2022, using the years 2012–2014 as a warm-up period to minimise the influence of initial storage conditions. Model performance was assessed for the calibration period (2015–2019) using the multi-objective function F_{obj} (Equation 2.36), which combined the Nash-Sutcliffe Efficiency (NSE) for daily discharge, the NSE for log-transformed daily discharge, and the NSE for monthly runoff coefficients.

A total of 90 000 parameter sets were evaluated for each model. Parameter sets with an objective function value below a predefined threshold during the calibration period were subsequently evaluated for the validation period (2020–2022). For these selected runs, an overall performance score was computed as the mean of the calibration and validation objective values. The parameter set with the lowest overall score was selected as the final calibrated model.

2.5.3. Model validation and comparison

This study used one conventional model and three models in which phenology was integrated, each using a different method. This resulted in a total of four models. In order to validate and compare each model, multiple methods were used.

First, the model parameter sensitivity for each model was analysed using scatterplots, showing all combinations tried during calibration which resulted in an overall performance score below 0.6. This provided insights into the model's sensitivity to slight changes in the corresponding parameters value.

Next, simulated discharges were evaluated using the $N_{\text{NSE,Q}}$ and $N_{\text{logNSE,Q}}$ metrics. Additionally, the Root Mean Squared Error (N_{RMSE}) was calculated to provide an absolute measure of model error. Besides the annual performance (January–December), the year was divided in two subsets based on the hydrological year which starts 1 October. The first subset covered the winter semester (October–March) and the second covered the summer semester (April–September). Splitting the dataset in seasons allowed for an analysis of the seasonal performance using the same set of metrics. Furthermore, the discharge time series were visually compared.

Finally, the daily simulated transpiration was be evaluated using the N_{NSE}, N_{logNSE} and N_{RMSE} metrics. To do so, transpiration data obtained from the NASA Global Land Data Assimilation System (GLDAS,

Table 2.1: Overview of model parameters estimated during calibration for each model, including units and realistic value ranges.

Parameter	Unit	Value range per model			
		Conventional	***SF***	***K_v***	***Combined***
Catchment					
$S_{u,max}$	[mm]	(100, 600)	(100, 600)	(100, 600)	(100, 600)
β	[−]	(0.01, 0.1)	(0.01, 0.1)	(0.01, 0.1)	(0.01, 0.1)
$P_{s,max}$	[mm d^{-1}]	(0.001, 1.2)	(0.001, 1.2)	(0.001, 1.2)	(0.001, 1.2)
K_f	[d]	(0.0, 7.0)	(0.0, 7.0)	(0.0, 7.0)	(0.0, 7.0)
K_s	[d]	(0.0, 50.0)	(0.0, 50.0)	(0.0, 50.0)	(0.0, 50.0)
D	[−]	(0.0, 1.0)	(0.0, 1.0)	(0.0, 1.0)	(0.0, 1.0)
Deciduous					
I_{max}	[mm]	(1.0, 5.0)	(1.0, 5.0)	(1.0, 5.0)	(1.0, 5.0)
C_e	[−]	(0.2, 1.0)	N/A	(0.2, 1.0)	N/A
$K_{v,max}$	[−]	N/A	N/A	(0.0, 1.0)	(0.0, 1.0)
Coniferous					
I_{max}	[mm]	(1.0, 5.0)	(1.0, 5.0)	(1.0, 5.0)	(1.0, 5.0)
C_e	[−]	(0.2, 1.0)	N/A	(0.2, 1.0)	N/A
$K_{v,max}$	[−]	N/A	N/A	(0.0, 1.0)	(0.0, 1.0)

(Li et al., 2020)) were used as observational data. Following the same approach as for discharge, the metrics were used to assess the annual and seasonal performances. Furthermore, the transpiration time series were visually compared using a 7-day rolling mean. Additionally, the contributions of the coniferous and deciduous parts of the models were compared with each other.

2.6. Sap flow prediction model

Conceptual hydrological rainfall–runoff models typically rely on internal fluxes and storages (e.g., soil moisture), which respond dynamically to meteorological inputs. To simulate these processes, the models require initial conditions for each storage component; however, these values are often unknown or imprecise. To address this limitation, models are generally run through a warm-up period, allowing internal states to adjust from arbitrary starting values toward more realistic conditions. The warm-up period typically ranges from one to several years, depending on the accuracy of the initial conditions (Kim et al., 2018). Consequently, several years of input data are usually needed—not only to ensure an adequate warm-up, but also to support robust model calibration and validation. In this study, a three-year warm-up period (2012–2014) was used, followed by a five-year calibration period (2015–2019) and a three-year validation period (2020–2022).

Two of the designed models required normalised sap flow $v_{sf,n}$ as input (see Section 2.4). Direct sap flow measurements were only available for the 2019 and 2020 growing seasons, which was insufficient for running the designed models. Other inputs required for the model were available for multiple years, making sap flow data the limiting factor.

Limited temporal coverage is common in sap flow datasets. The SAPFLUXNET database, the first global compilation of whole-plant transpiration data from sap flow measurements, contains sap flow time series for 2714 plants (mostly trees) of which 50% is shorter than three years (Poyatos et al., 2021). Moreover, these datasets are often restricted to the growing season only.

To address this challenge, a model to predict normalised sap flow was designed. This predictive model enabled the extension of the sap flow time series, allowing for its incorporation into the hydrological models.

A wide range of modelling approaches exist. Multiplicative models such as the popular Jarvis model multiply stress functions that depend on environmental factors, acting as reduction factors (Jarvis, 1976). A downside of the multiplicative nature is that a (wrong) near-zero value of any of the stress functions will cause a near-zero outcome.

Machine learning models are another option. A major downside of these models is their 'black-box' structure. The model internals are not interpretable and they can produce physically inconsistent predictions (Daw et al., 2020).

In this study, an additive model was chosen to predict $v_{\text{sf,n}}$. Specifically, a Generalized Additive Model (GAM) was chosen as these allow for the use of smooth functions (see Section 2.6.1). GAMs add different functions that can depend on environmental variables together to predict the outcome, which is visualised in Figure 2.9. The additive structure allows separate analysis of each predictor's relationship with sap flow. This gives a better understanding of the internal processes of the model and the influence of each variable. Additionally, adjustments to the model can be made informed by this information (e.g., introducing constraints). GAMs have been used in previous research but mostly to analyse the relation of sap flow with environmental variables (e.g., Luo et al. (2022), Tian et al. (2018), Flo et al. (2022) and Ohana-Levi et al. (2020)). In this study, it is hypothesised that if the influence of environmental variables on sap flow can be analysed using GAMs, the reverse should also be possible: the fitted model could predict sap flow based on the derived relationships and the same environmental variables.

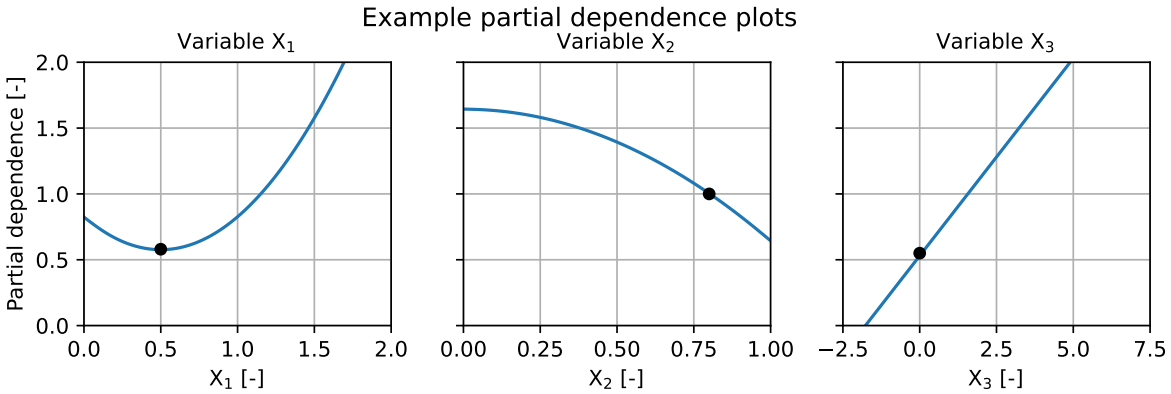


Figure 2.9: An example of partial dependence plots. If the conditions (x_1, x_2, x_3) = (0.5, 0.8, 0) occur, the model predicts a value by taking the sum of the partial dependency functions: $0.58+1.0+0.55 = 2.13$.

2.6.1. Generalized additive models

The relationship between a response variable Y and one or more predictors X can be modelled using linear regression. With only one predictor, the model is a simple linear regression. A flexible generalisation of linear models was formulated by Nelder and Wedderburn (1972) as Generalized Linear Models (GLMs), unifying various other statistical models such as linear regression, logistic regression and Poisson regression.

A GLM has three main components. First, the linear predictor η combines the input variables (predictors X) to a linear equation, using coefficients β as weights (Eq. 2.37). Here, the parameters are estimated by maximum likelihood. Second, the distribution of the response variable Y which can follow any distribution of the exponential family. Third, the link function $g(\cdot)$ that links the linear predictor η to the expected value $\mu = \mathbb{E}[Y]$ (Eq. 2.38). As η has a domain of $(-\infty, +\infty)$, the link function maps it to the domain of the response variable. For example, a logarithmic link function is often used in combination with Gamma and Inverse Gaussian distributions, to ensure $\mu > 0$ (Dunn and Smyth, 2018).

The linear predictor for a set of predictors $X_1, X_2, \ldots, X_n$ can be expressed by:

$$\eta_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_n x_{n,i} \quad \rightarrow \quad \eta = \beta_0 + \sum_{i=1}^{n} \beta_i X_i \tag{2.37}$$

and the GLM itself can be expressed by:

$$g(\mathbb{E}[Y]) = \eta \tag{2.38}$$

Generalized Additive Models (GAMs), introduced by Hastie and Tibshirani (1986), extend the traditional Generalized Linear Models (GLMs) by replacing linear predictors (Eq. 2.37) with a set of smooth functions (Eq. 2.39) where $f(X)$ are smooth functions. This allows greater flexibility in capturing complex, non-linear relationships between predictors X and the response variable Y that may be difficult to represent with purely parametric functions. The distribution of the response variable and the link function are not changed compared to GLMs.

$$\eta = \beta_0 + \sum_{i=1}^{n} f_i(X_i) \tag{2.39}$$

A widely adopted approach for estimating the smooth functions are penalized B-splines (basis splines), or P-splines (Eilers and Marx, 2002). Introduced by Eilers and Marx (1996), P-splines combine the flexibility of B-spline basic functions with a penalty on the spline coefficient to control the smoothness. This penalty helps to avoid overfitting by limiting the wiggliness while still capturing the complex (non-linear) relations. Therefore, it balances the fit and complexity of the model.

B-splines are constructed from polynomial pieces, joined at certain values of x, called the knots. Each B-spline is only non-zero for a limited domain, giving it local control. Using these B-splines, a smooth curve $s(x)$ can be constructed by taking a weighted sum of multiple B-splines as shown in Equation 2.40 below.

$$s(x) = \sum_{j=1}^{n} \beta_j B_j(x) \tag{2.40}$$

where $B_j(x)$ is B-spline basis function j and β_j its corresponding scaling coefficient. For a set with m datapoints (x_i, y_i) on a set of n B-splines, the least square objective function (F_{bs}) to minimize becomes Equation 2.41.

$$F_{\text{bs}} = \sum_{i=1}^{m} \left[y_i - \sum_{j=1}^{n} \beta_j B_j(x_i) \right]^2 \tag{2.41}$$

Using too many knots leads to overfitting while too few knots lead to underfitting. To make the results less flexible, Eilers and Marx (1996) proposed to introduce a penalty based on finite differences of the coefficients of adjacent B-splines (hence, P-splines). The least square objective function (F_{ps}) to minimize becomes Equation 2.42.

$$F_{\text{ps}} = \sum_{i=1}^{m} \left[y_i - \sum_{j=1}^{n} \beta_j B_j(x_i) \right]^2 + \lambda \sum_{j=k+1}^{n} (\Delta^k \beta_j)^2 \tag{2.42}$$

where Δ^k is the k-th order derivative of the coefficients and the parameter λ for control over the smoothness of the fit. As the smoothness penalty increases, the differences between neighbouring coefficients will decrease, reducing the overall flexibility of the model. This is indicated by the effective degrees of freedom (EDoF), which quantifies the model's flexibility in fitting the data. A lower EDoF indicates a smoother, less complex fit. When there is no penalty introduced ($\lambda = 0$), the EDoF will be equal to the number of splines used in the smooth function and can therefore not be larger than n (Baayen and Linke, 2020). A visual representation of how P-splines work and the effect of λ is shown in Figure 2.10.

2.6.2. Model input

Lopushinsky (1986) found that the seasonal variations in sap flow velocity for coniferous trees (Douglas fir and ponderosa pine) were significantly related to T, R_{s} and vapour pressure deficit (D_{vp}) when soil moisture was not limiting. As soil moisture decreased during the summer, v_{sf} no longer followed the evaporative demand. Juice et al. (2016) had similar findings for deciduous trees (northern red oak).

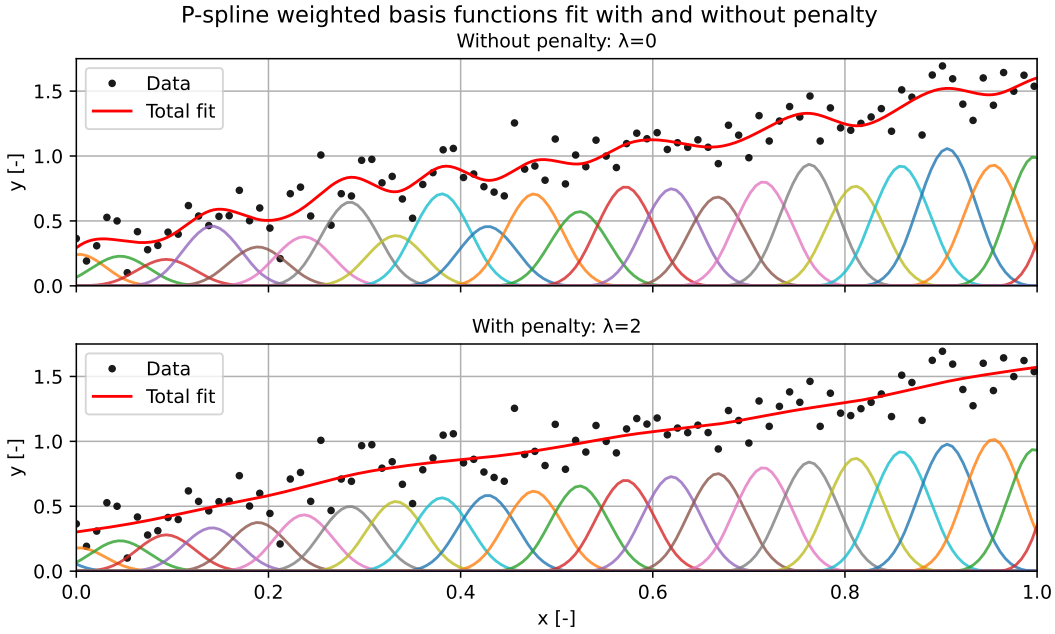


Figure 2.10: Illustration of the impact of the penalty on B-splines. Both plots have the same number of B-splines but a different penalty. The upper plot uses has a lower penalty than the lower, resulting in a smoother fit for the latter one.

They found that soil and air temperature explained the largest amount of variance in v_{sf}, with h, R_{s}, and D_{vp} also having a significant effect, but smaller.

Studies that used GAMs to analyse the relation between environmental variables and v_{sf} often considered these same variables (e.g., Luo et al. (2022), Tian et al. (2018), Flo et al. (2022) and (Ohana-Levi et al., 2020)). Additionally, several studies included a variable that was strongly linked to seasonality such as the Normalised Difference Vegetation Index (NDVI) and proved to be statistically significant.

In this study, similar environmental variables were considered as input for the sap flow prediction model, namely T, R_{s}, h, and θ at 10, 20, 40, and 60 cm depth. Seasonality was represented using the yearly normalised cumulative Growing Degree-Days ($T_{\text{CGDD,n}}$). This was the normalised version of the T_{CGDD}, which was also (indirectly) used to represent seasonality in the K_{v} methods for the hydrological models (see Sections 2.4.3 and 2.4.4). Yearly normalisation allowed $T_{\text{CGDD,n}}$ to serve as an indicator of phenological events. For example, leaf development started earlier in years with a warm spring, which was reflected by $T_{\text{CGDD,n}}$ reaching the same value earlier in the year.

Another considered variable was the vapour pressure deficit (D_{vp}), which is the difference between the amount of moisture in the air and the amount it could hold at saturation. It is a function of both T and h, and often used to describe E_{t}. Since D_{vp} is a function of T and h, including it could result in their over-representation. Therefore, D_{vp} was excluded.

Finally, a weighted average of θ (θ_{avg}) was considered for multiple reasons. The root architecture is species specific but is also influenced by other variables such as soil type, terrain characteristics, and competition with trees from a different species (Hackmann et al., 2025; Zhiyanski, 2014). This difference in root architecture also influences the trees ability for water uptake. For example, Douglas fir is more drought resistant than Norway spruce by accessing deeper water sources (Hackmann et al., 2025).

Due to uncertainties regarding root depths, a weighted average was calculated based on the thickness of the soil layer represented by each measurement depth. Data were available at 10, 20, 40, and 60 cm, and the weighted average was computed for the 0–70 cm soil profile. The 70 cm lower boundary was chosen because most of the root structure was located above this depth and to avoid over- or under-representing data from the 60 cm measurements relative to the other depths.

The depth intervals were defined based on the midpoints between adjacent sensors, such that the 10 cm measurement represented the 0–15 cm layer, the 20 cm measurement the 15–30 cm layer, the 40 cm measurement the 30–50 cm layer, and the 60 cm measurement the 50–70 cm layer. The weighted average (θ_{avg}) was then calculated as the sum of the products of each layer's mean θ and its thickness (Δz), divided by the total profile thickness (Equation 2.43).

$$\theta_{\text{avg}} = \frac{\sum_{i=1}^{n} (\theta_i \cdot \Delta z_i)}{\sum_{i=1}^{n} \Delta z_i} \tag{2.43}$$

The Spearman's rank correlation coefficient was used to determine how strongly $v_{\text{sf,n}}$ was correlated with each of the environmental variables. Spearman's rank correlation was chosen over the Pearson correlation coefficient because not all variables were linearly related to $v_{\text{sf,n}}$. Correlations were calculated for each tree across all considered variables, and the averages are shown in Figure 2.11. A detailed overview of coefficients per variable per tree is provided in Appendix B. The coefficients indicated that T had a strong positive correlation, while h had a strong negative correlation with $v_{\text{sf,n}}$. R_{s} had a moderate positive correlation, and $T_{\text{CGDD,n}}$ a moderate negative correlation. θ_{avg} and θ at different depths all exhibited similar weak negative correlations.

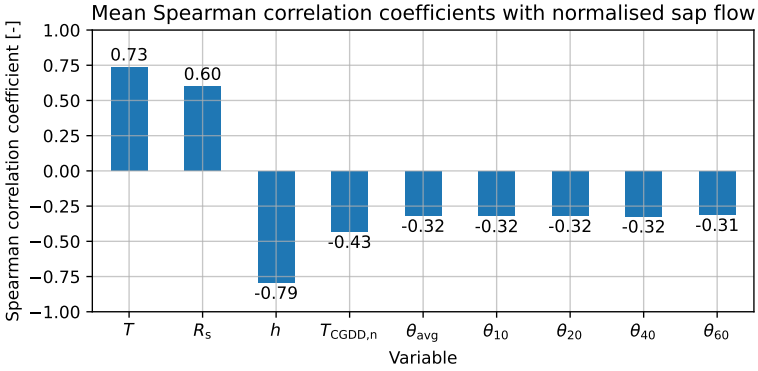


Figure 2.11: Spearman's rank correlation coefficient between normalised sap flow and temperature (T), incoming shortwave radiation (R_{s}), relative humidity (h), normalised accumulated Growing Degree-Days ($T_{\text{CGDD,n}}$), weighted average of volumetric water content (θ_{avg}) and θ at 10 (θ_{10}), 20 (θ_{20}), 40 (θ_{40}), and 60 (θ_{60}) cm depth.

Although the different θ variables showed low correlations, θ_{avg} was included in the model because of its limiting effect on transpiration, particularly during dry summer periods. Since θ_{avg} exhibited a similar correlation to depth-specific θ, no additional depth-specific θ's were included.

Hourly sap flow data were available for two growing seasons (2019 & 2020) for several trees. Although the hydrological model operated on daily time steps and therefore required daily $v_{\text{sf,n}}$ input, the sap flow prediction model was developed on an hourly scale. This provided more data points for calibration and validation and allowed the model to better capture diurnal dynamics that would otherwise be lost in daily averages. Consequently, the input data also had to be provided at an hourly frequency.

To conclude, the following variables were used as predictors for the GAM model:

1. Temperature (T)
2. Relative humidity (h)
3. Incoming shortwave radiation (R_{s})
4. Weighted average of volumetric water content (θ_{avg})
5. Normalised accumulated Growing Degree-Days ($T_{\text{CGDD,n}}$)

2.6.3. GAM implementation

While many published studies used the *mgcv* R package (Wood, 2011) for fitting GAMs, this study adopted the *pyGAM* Python package (Servén and Brummitt, 2018) version 0.9.1. This decision was primarily motivated by the author's familiarity with Python, which allowed for a consistent and streamlined workflow across data preprocessing, modelling, and visualization.

Moreover, *pyGAM* implements the same core concepts of GAMs as introduced by Hastie and Tibshirani (1986) and also supports the use of P-splines. Furthermore, *pyGAM* is partly based on Wood (2006), written by the author of the *mgcv* package. Although *mgcv* provides a wider range of features, *pyGAM* offers the flexibility and functionality needed for this study.

Input preparation

The hydrological model was semi-distributed, requiring distinct $v_{\text{sf,n}}$ inputs for the deciduous and coniferous parts. Accordingly, the sap flow prediction model produced two separate time series: one for deciduous trees and one for coniferous trees. Because of differences in phenology and physiology, each vegetation category was modelled separately.

The sap flow prediction model used five predictors (X) as input: T, h, R_{s}, θ_{avg}, and $T_{\text{CGDD,n}}$. The response variable was $v_{\text{sf,n}}$ (Y). *PyGAM* required both X and Y as input to estimate partial dependencies between Y and X.

The GAM did not use time series but individual data points. Each data point consisted of the value of Y and the corresponding values of X at which Y occurred. For each tree, every sap flow measurement was matched with the concurrent environmental conditions, with θ_{avg} estimated at the tree location using Equation 2.43 and the method described in Section 2.2. The resulting data points were defined as:

$$\text{Data point } i = (Y_i, X_i) = \left([v_{\text{sf,n},i}], [T_i, h_i, R_{\text{s},i}, \theta_{\text{avg},i}, T_{\text{CGDD,n},i}]\right)$$

Analysis showed that sap flow was observed at temperatures below 0 °C. This flow was likely caused by freeze-thaw cycles, which create pressure differences in the tree stem. Because stomata are closed at low temperatures, such flow likely did not contribute to transpiration. Therefore, data from days with minimum temperatures below 0 °C were excluded from the analysis.

Next, all data points for deciduous trees were aggregated into a single dataset, and those for coniferous trees into another. These datasets served as input for the respective models. Both were split randomly into training (80%) and testing (20%) subsets using a fixed random seed, ensuring reproducibility.

Structure & constraints

A GAM has three main components, namely the statistical distribution of the response variable, the link function, and the linear predictor. First, a Gamma distribution was chosen to capture the positive, continuous, and right-skewed nature of $v_{\text{sf,n}}$. Next, a log link function was used, which is a common combination with a Gamma distribution as it avoids the need for constraints on the linear predictor to ensure non-negative results (Dunn and Smyth, 2018).

To prevent internal model errors arising from the logarithmic link function, all $v_{\text{sf,n}}$ values below 10^{-5} were replaced by 10^{-5}. For deciduous trees, this adjustment affected 100 data points (0.047%), and for coniferous trees, 29 data points (0.046%), and was therefore considered negligible.

The linear predictor was built using P-spline terms (s) for all predictors to allow for non-linear relations between environmental variables and $v_{\text{sf,n}}$ (Equation 2.44). Here, the default number of 20 splines and a λ value of 0.6 were used for each term, and an intercept was added. Initially the model was created without constraints to detect the functional relationships and potential overfitting. Afterwards, shape constraints were imposed to limit overfitting, smoothening the curves to better reflect biological plausibility. Additionally, shape constraints gave more control on the extrapolation by influencing the direction of the tails. This was important as pyGAM uses a linear extrapolation when the value of a predictor falls outside the training range. For this reason, only the constrained models were used for the prediction of $v_{\text{sf,n}}$.

$$\eta_{\text{SF}} = \beta_0 + s_0(T) + s_1(h) + s_2(R_{\text{s}}) + s_3(\theta_{\text{avg}}) + s_4(T_{\text{CGDD,n}}) \tag{2.44}$$

Like the smoothness penalty parameter λ, the introduction of shape constraints increased the smoothness of the curves and reduced the EDoF. The constraints smoothed the curves sufficiently, eliminating the need for the computationally intensive grid search to optimise λ. The used parameters and their constraints are listed in Table 2.2.

Table 2.2: Overview of predictors with their corresponding unit and constraint as applied in the GAM structure.

Predictor	Unit	Constraint
T	[°C]	Concave
h	[%]	Monotonically decreasing
R_s	[W m^{-2}]	Monotonically increasing
θ_{avg}	[m^3 m^{-3}]	Concave
$T_{CGDD,n}$	[−]	Concave

A concave constraint was applied to T, as v_{sf} is known to increase with temperature up to an optimum, above which stomatal closure causes v_{sf} to decrease (Kubota, 2016). Although this behaviour was not observed in the data, a concave constraint was considered the most realistic option.

A monotonically decreasing constraint was applied to h, as v_{sf} decreases with higher humidity due to the reduced vapour pressure deficit. Conversely, a monotonically increasing constraint was applied to R_s, reflecting the expected increase in v_{sf} with higher radiation availability.

A low θ_{avg} was expected to result in a low v_{sf}, as transpiration is constrained by water availability. As θ_{avg} increases, v_{sf} also increases until water is no longer limiting. When the soil moisture becomes too high, oxygen is displaced, and above a certain anaerobiosis threshold, root water uptake may be assumed to cease (Feddes et al., 1976). For this reason, a concave shape constraint was imposed on θ_{avg}.

Transpiration is low in winter and high in summer, causing v_{sf} to increase with $T_{CGDD,n}$ at the start of the year. Towards the end of the year, at high $T_{CGDD,n}$, v_{sf} decreases again, producing a concave relationship. Therefore, a concave constraint was applied to $T_{CGDD,n}$.

Prediction procedure

Each tree category had its own constrained model, where each model had the previously described structure. These models were first fitted on their respective training datasets. The resulting models predicted $v_{sf,n}$ given the environmental conditions.

As both models were trained on all trees in their own dataset, the models were used to predict $v_{sf,n}$ for each tree in their datasets separately. The hydrological models needed $v_{sf,n}$ data for 2012–2022, so the GAMs were used to predict the hourly $v_{sf,n}$ within this period. Next, the mean $v_{sf,n}$ was calculated per hour per tree category, resulting in a single hourly $v_{sf,n}$ time series per category.

The hydrological models required daily values as inputs. Therefore, the predicted hourly $v_{sf,n}$ time series were resampled to a daily mean. Because higher $v_{sf,n}$ values typically occurred only during the afternoon and are much lower during the rest of the day, calculating daily means significantly reduces the maximum predicted values. Therefore, each year was normalised back to values between 0 and 1. The two resulting daily time series were used as input for the hydrological models that required $v_{sf,n}$ time series.

Model evaluation

Both the unconstrained and constrained models were evaluated. The first evaluation was based on the partial dependence, allowing verification of whether the constraints listed in Table 2.2 were consistent with the unconstrained partial dependencies and the expected relationships.

The models were fitted on their training datasets, allowing the testing datasets to be used for validation. The first method to evaluate the models was comparing the coefficient of determination (N_{R^2}) and

Root Mean Squared Error (N_{RMSE}) between the training and testing sets, providing an assessment of generalisability. A small difference in N_{R^2} and N_{RMSE} between the two sets indicates that the model captured the underlying relationships without overfitting, whereas large differences suggests overfitting or sensitivity to data variation.

Additionally, the Nash-Sutcliffe efficiency metric (N_{NSE}, Equation 2.34), its logarithmic version (N_{logNSE}), and N_{RMSE} were calculated for 2019, 2020, and 2019–2020 combined to assess model performance over these periods. The predictions were also visually compared with observations using one-to-one plots and the 2019 and 2020 time series. The one-to-one plots were used for both hourly and daily predictions to give insights into the potential difference in prediction between hourly and daily time series. The other evaluation methods were performed on the daily time series.

3 Results

In this chapter, the results are presented. First, the sap flow prediction model is evaluated, including its ability to reproduce observed dynamics and the role of environmental predictors. Next, the results of incorporating phenology into the conceptual hydrological model are shown. The performance of the conventional and modified models is assessed with respect to discharge and transpiration, using both evaluation metrics and visual inspection of the simulated time series. Additionally, seasonal contributions of coniferous and deciduous trees to total transpiration are examined.

3.1. Sap flow prediction model

Two sap flow prediction models were developed to estimate the normalized sap flow ($v_{\text{sf,n}}$) for deciduous and coniferous trees in periods without data availability. The GAMs were trained on data from the 2019 and 2020 growing seasons. In this section, the fitted models are described based on their partial dependencies. The performance of the models was evaluated using several metrics.

3.1.1. Normalised accumulated growing degree-day

The sap flow prediction model had five predictor variables: temperature T, relative humidity h, incoming shortwave radiation R_{s}, weighted average of the volumetric water content θ_{avg}, and normalised accumulated growing degree-day $T_{\text{CGDD,n}}$.

$T_{\text{CGDD,n}}$ was derived from the temperature time series and is shown in Figure 3.1. This figure shows that the curve was different for each year. In a colder year (e.g., 2013), the curve showed a delayed increase compared to other years, whereas in a warmer year (e.g., 2020), the increase occurred earlier.

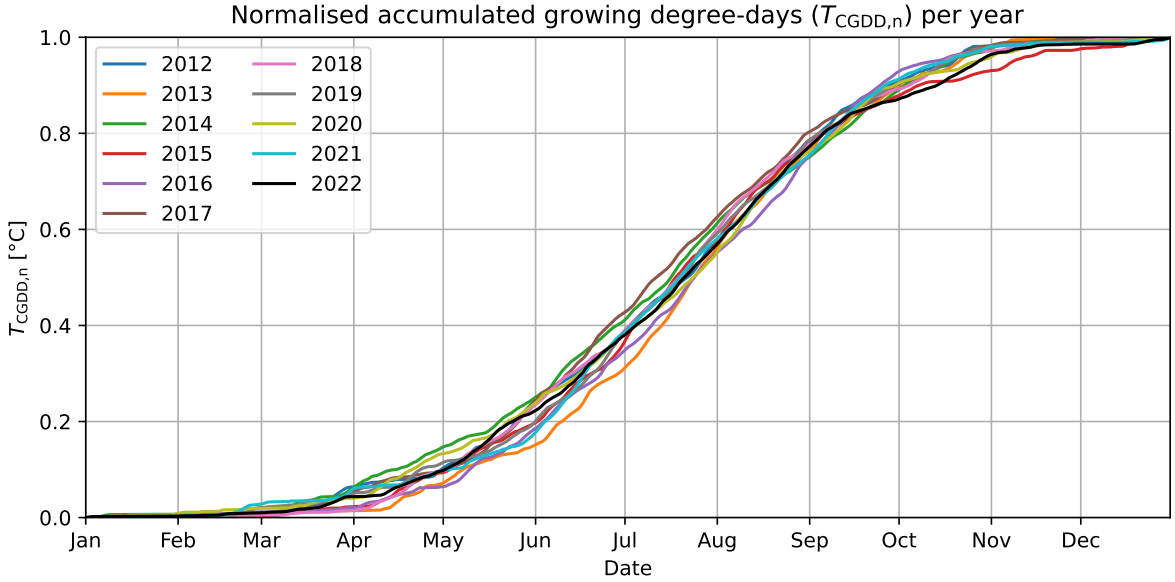
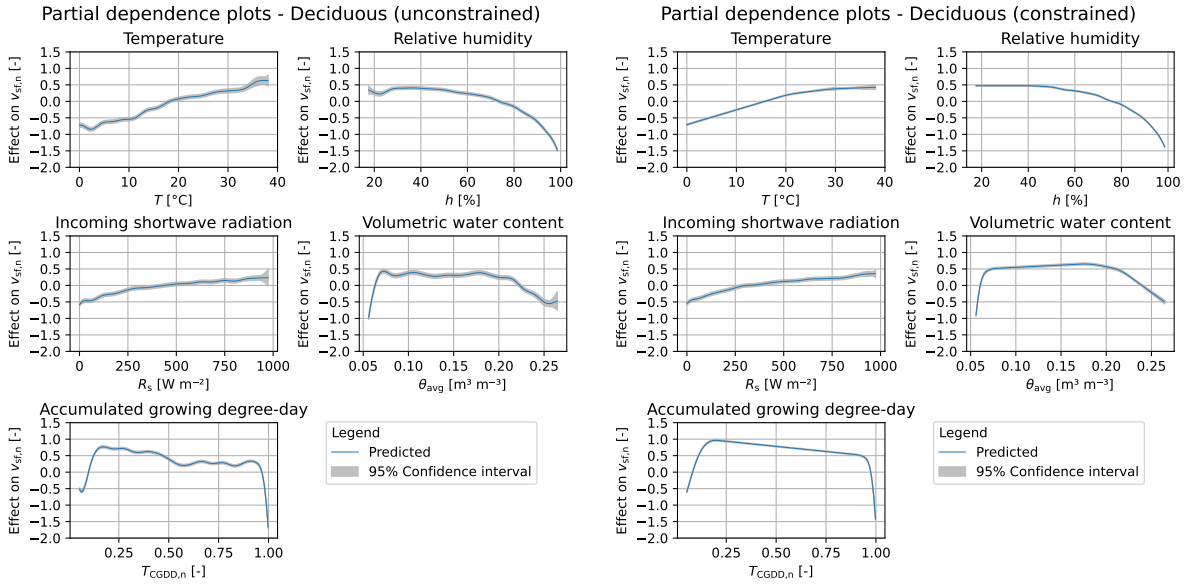


Figure 3.1: Normalised accumulated growing degree-day ($T_{\text{CGDD,n}}$) per year as used as input for the Sap Flow Prediction Model.

3.1.2. Partial dependencies

Separate GAM models were developed for the deciduous and coniferous trees. First, the models were fitted without shape constraints. Afterwards, shape constraints were introduced based on the unconstrained fits and the known physical relationships between v_{sf} and the predictor variables (Table 2.2).



(a) Fitted unconstrained deciduous GAM with an intercept value of -1.887 [−].

(b) Fitted constrained deciduous GAM with an intercept value of -2.719 [−].

Figure 3.2: Partial dependencies of the predictors for the unconstrained and constrained fitted GAMs for deciduous trees.

The unconstrained deciduous GAM (Figure 3.2a) showed that $v_{sf,n}$ increases with temperature but has a negative slope near zero temperature. In general, relative humidity showed a decrease in $v_{sf,n}$ for increasing h, reaching its lowest value at 100%. However, $v_{sf,n}$ decreased at 20% h and increased again when h was slightly higher. $v_{sf,n}$ showed a steady increase with an increasing incoming shortwave radiation. The partial dependence on θ_{avg} showed a strong increase in $v_{sf,n}$ when θ_{avg} was low and increasing. It then reached a plateau around 0.20 m^3 m^{-3}, after which $v_{sf,n}$ began to decrease again. Finally, $v_{sf,n}$ increases with $T_{CGDD,n}$ for low values of $T_{CGDD,n}$ and gradually decreases after reaching its peak $v_{sf,n}$. $v_{sf,n}$ showed a strong decrease for $T_{CGDD,n}$ values above 0.90.

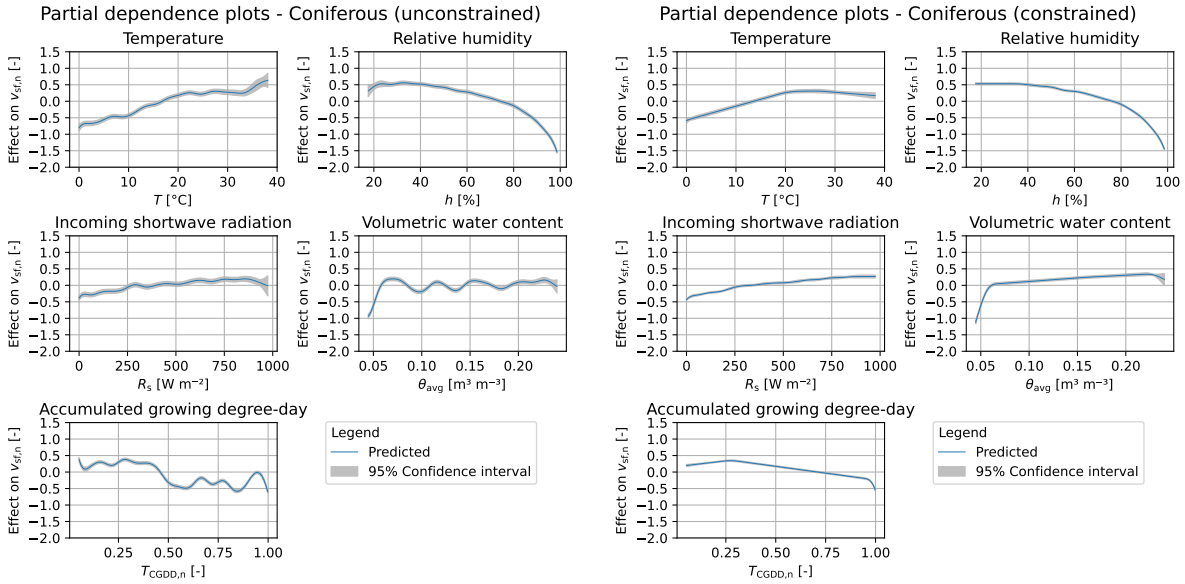
Some of the partial dependencies showed a wider range of the 95% confidence interval near the minimum or maximum variable values. Furthermore, the unconstrained GAM showed local fluctuations around the general trends ('wiggliness').

The constrained GAM (Figure 3.2b) exhibited fewer local fluctuations, and the increased range of the 95% confidence interval was reduced compared to the unconstrained GAM. When constrained, the small local irregularities (e.g., at T near 0 °C or h near 20%) disappeared. Once constrained, $v_{sf,n}$ slightly increased with θ_{avg} followed by a gradual decrease above 0.20 m^3 m^{-3}. After a strong increase of $v_{sf,n}$ for low $T_{CGDD,n}$, $v_{sf,n}$ showed a gradual decrease with increasing $T_{CGDD,n}$ and a sharp drop around 0.90.

The unconstrained GAM for coniferous trees (Figure 3.3a) showed similar trends as its deciduous counterpart. However, the local fluctuations were stronger. The partial dependence curve for T showed a plateau around 30 °C and increased further above 35 °C.

The θ_{avg} dependency plot exhibited the same initial increase in $v_{sf,n}$ as the unconstrained deciduous GAM at low θ_{avg}. However, whereas the deciduous model showed a decreasing trend in $v_{sf,n}$ above 0.20 m^3 m^{-3}, the coniferous model continued its upward trajectory.

Another difference was observed in the $T_{\text{CGDD,n}}$ dependency plot, which did not display clear seasonality like the deciduous GAM. For low $T_{\text{CGDD,n}}$ values, the coniferous model showed a slight increase in $v_{\text{sf,n}}$, followed by a strong decrease around 0.5, after which $v_{\text{sf,n}}$ remained approximately constant (ignoring local fluctuations). Additionally, the overall influence of $T_{\text{CGDD,n}}$ on $v_{\text{sf,n}}$ was smaller than in the deciduous model, as indicated by the reduced difference between its minimum and maximum values.



(a) Fitted unconstrained coniferous GAM with an intercept value of -1.403 [−].

(b) Fitted constrained coniferous GAM with an intercept value of -1.825 [−].

Figure 3.3: Partial dependencies of the predictors for the unconstrained and constrained fitted GAMs for coniferous trees.

3.1.3. Model performance

Model performance was first assessed by comparing fitting metrics between the training and test datasets to evaluate the models' generalisability. Prediction performance was subsequently evaluated using multiple metrics.

Model generalisability

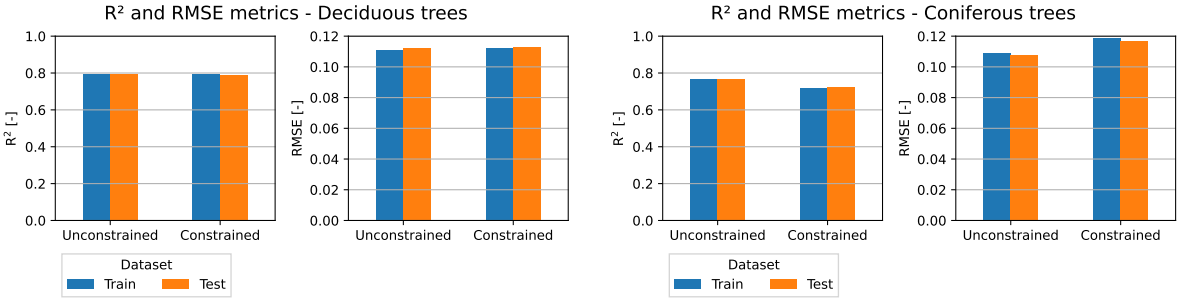
The GAMs were fitted on a training dataset, of which the N_{R^2} and N_{RMSE} fitting metrics were calculated. The same metrics were calculated when the fitted models were used on the unseen testing dataset. Figure 3.4 shows the comparison of these metrics. Based on these fitting metrics, the deciduous GAMs exhibited a better fit, as indicated by higher N_{R^2} values. N_{RMSE} values were very similar across models, with only the constrained coniferous GAM performing slightly lower.

Results for unconstrained and constrained models were very similar, with differences of less than 1% in N_{R^2} and N_{RMSE}.

Prediction Performance

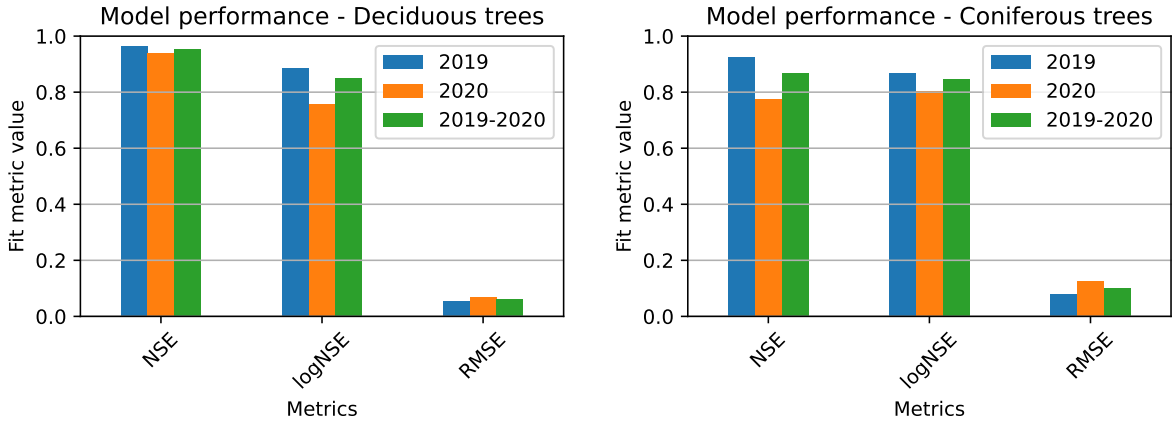
The constrained models were used for the prediction of $v_{\text{sf,n}}$. Their performance was evaluated based on N_{NSE}, N_{logNSE} and N_{RMSE}. Both models performed better in 2019 compared with 2020. For the deciduous model, N_{NSE} and N_{RMSE} were similar between the two years, while N_{logNSE} indicated lower performance in 2020 (see Figure 3.5a). This suggests that the model predicted high $v_{\text{sf,n}}$ peaks well (N_{NSE}) but was less accurate for lower $v_{\text{sf,n}}$ values (N_{logNSE}).

Overall, the coniferous model had a lower performance, scoring lower on both N_{NSE} and N_{logNSE} metrics and exhibiting higher N_{RMSE} (see Figure 3.5b). Additionally, the model's performance for 2020 was lower than for 2019 on all three metrics. The N_{NSE} score decreased significantly between 2019 and 2020 (0.93 to 0.78), indicating it performed lower on predicting the peaks in $v_{\text{sf,n}}$ for 2020. The model's



(a) Deciduous GAMs. (b) Coniferous GAMs.

Figure 3.4: The N_{R^2} and N_{RMSE} fitting metrics for unconstrained and constrained GAMs, both for train and test datasets.



(a) Deciduous GAMs. (b) Coniferous GAMs.

Figure 3.5: The N_{NSE}, N_{logNSE} and N_{RMSE} prediction performance metrics for the constrained deciduous and coniferous GAMs.

performance on low $v_{sf,n}$ also decreased (lower N_{logNSE}), but this difference (0.87 to 0.80) was smaller than for N_{NSE}.

The time series of observed and predicted $v_{sf,n}$ for 2019 are shown in Figure 3.6. Visually, both the predictions closely followed the observations, indicating that the models successfully reproduced them. The deciduous time series showed that the GAM predicted near-zero $v_{sf,n}$ between January and mid-March, as well as in November and December. $v_{sf,n}$ slightly increased between mid-March and mid-April, followed by a stronger increase. The highest values occurred toward the end of June, after which $v_{sf,n}$ gradually decreased.

The coniferous time series had higher $v_{sf,n}$ predictions at the start of the year compared to the deciduous time series. In sharp contrast to the deciduous model, higher $v_{sf,n}$ peaks already occurred in February and March. November and December showed low $v_{sf,n}$ values, but slightly higher than for its deciduous counterpart.

For 2020, the time series of observed and predicted $v_{sf,n}$ are shown in Figure 3.7. The deciduous observations and predictions showed a similar seasonality as 2019, although $v_{sf,n}$ reached higher values earlier in the year, showing a strong increase in April. From August onwards, predictions often slightly overestimated $v_{sf,n}$ but still captured the small fluctuations. Visually, the model closely followed the observations, indicating it was able to reproduce the observations.

As in 2019, the coniferous model predicted higher $v_{sf,n}$ values early in the year. However, the high February peaks observed in 2019 did not occur in 2020. The model significantly overestimated two $v_{sf,n}$ peaks in May. While it captured the small fluctuations at the peaks, the predicted peak magnitudes were

too high compared to the observations. However, directly after those peaks, $v_{sf,n}$ dropped, which the model predicted well. During June and July the model underestimated $v_{sf,n}$ multiple times. In contrast, the model overestimated $v_{sf,n}$ from mid-July to the end of September while accurately predicting October. Visually, predictions from the coniferous model generally followed the observations but often over- or underestimated $v_{sf,n}$ in 2020, leading to lower performance than the deciduous model for that year.

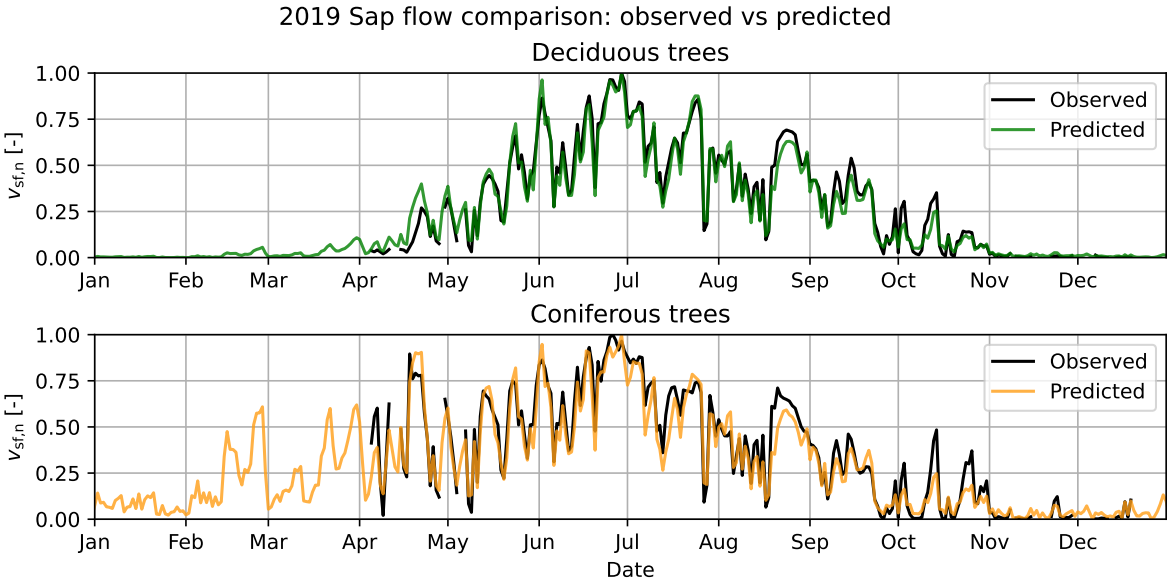


Figure 3.6: The 2019 time series with observed and predicted $v_{sf,n}$ for both coniferous and deciduous trees.

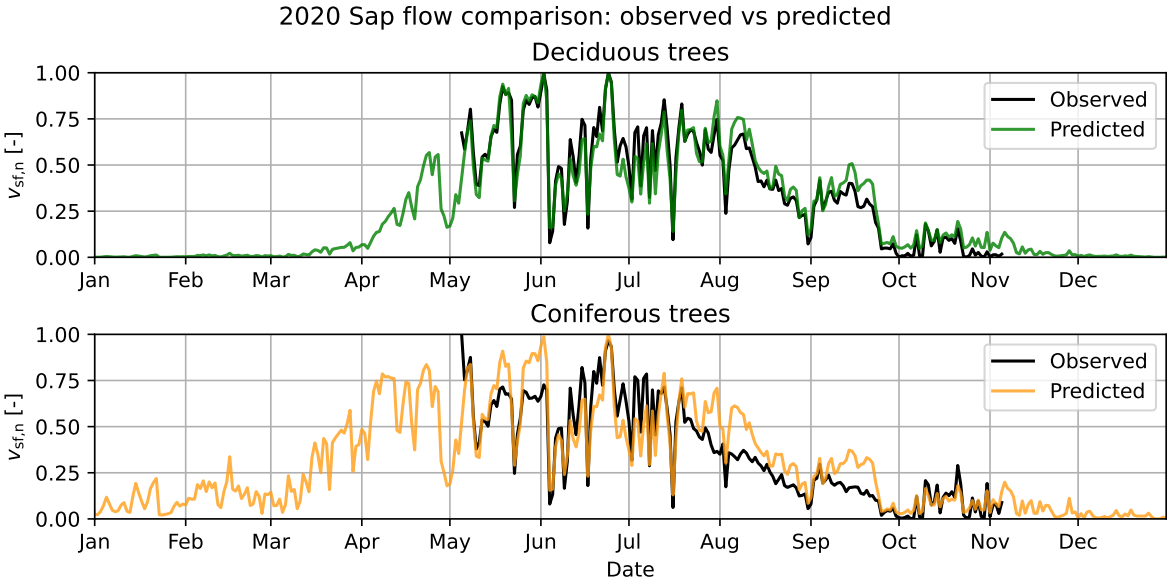
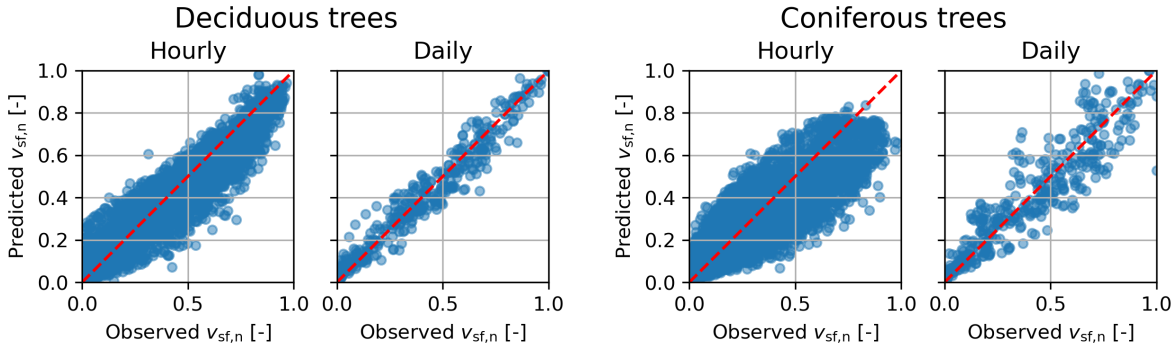


Figure 3.7: The 2020 time series with observed and predicted $v_{sf,n}$ for both coniferous and deciduous trees.

The observations and predictions were also compared using one-to-one plots, shown in Figure 3.8. The hourly plots show a larger spread compared to the daily plots. Visually, the deciduous model outperformed the coniferous model in terms of model spread for both hourly and daily frequency. The daily deciduous model, in particular, predicted values that closely followed the observations. Furthermore,

the point clouds were spread around the 1:1 line with no change in the range of the spread for higher values.



(a) Deciduous GAMs.

(b) Coniferous GAMs.

Figure 3.8: Comparison of the hourly and daily $v_{sf,n}$ predictions to observations using one-to-one plots for both deciduous and coniferous models. The red line indicates the 1:1 line, the theoretically perfect fit.

3.2. Phenology inclusion

The first and third phenology integration methods used normalised sap flow as input. The second and third methods used the K_v shape function as input. In this section, these additional inputs for the modified hydrological models are evaluated.

3.2.1. Sap flow

Two of the modified hydrological models included $v_{sf,n}$ data as input. This data were partly predicted using the developed sap flow prediction model based on data from the 2019 and 2020 growing seasons. The predictions were compared to observations in Section 3.1. The $v_{sf,n}$ time series for 2012–2022 are shown per year in Appendix C.

In general, $v_{sf,n}$ was higher for coniferous trees during the first months of the year (January–April). From May until October coniferous and deciduous time series were very similar, generally following the same fluctuations. During August and September, some years showed higher values for deciduous trees. During November and December $v_{sf,n}$ was near zero, although coniferous trees still showed small fluctuations.

3.2.2. Evaporation partitioning

Evaporation was partitioned between interception and transpiration using the K_v shape function and scaling factor $K_{v,max}$. The latter factor is discussed in more detail in Section 3.3.1. The K_v shape function was developed using the first and second derivatives of the logistic function fitted on the T_{CGDD} curve (Equations 2.25, 2.26 and 2.27). The fitted curve with its derivatives and final K_v shape functions are shown in Figure 3.9.

The fitted logistic function showed that warmer years (e.g., 2018 and 2022) had the highest T_{CGDD} at the end of the year. Accordingly, warmer years had a higher maximum value of their first derivative. The second derivatives generally had an earlier maximum for years that had a warmer spring, which resulted in an earlier start of the plateau for the K_v shape function. Warm years did not always result in an earlier growing season. For example, 2022 was a warmer year than 2014 on average, but 2014 had a warmer spring. As a result, its growing season likely started earlier, which was also indicated by the early maximum of the second derivative and the resulting plateau in the K_v shape function.

The end of the growing season was defined as the moment when the slope of the T_{CGDD} started to decrease (i.e., the moment when the second derivative reached its minimum). The moment at which this happened had less variance compared to the start of the growing season. The combination of start and end of the growing season defined its length, and is shown in Table 3.1.

In constructing the K_v shape function, the slope of the first derivative was used. This resulted in a smooth increasing partitioning during the development stages of deciduous trees. The late-season stage followed a smooth decreasing curve which ended near zero at the end of the year.

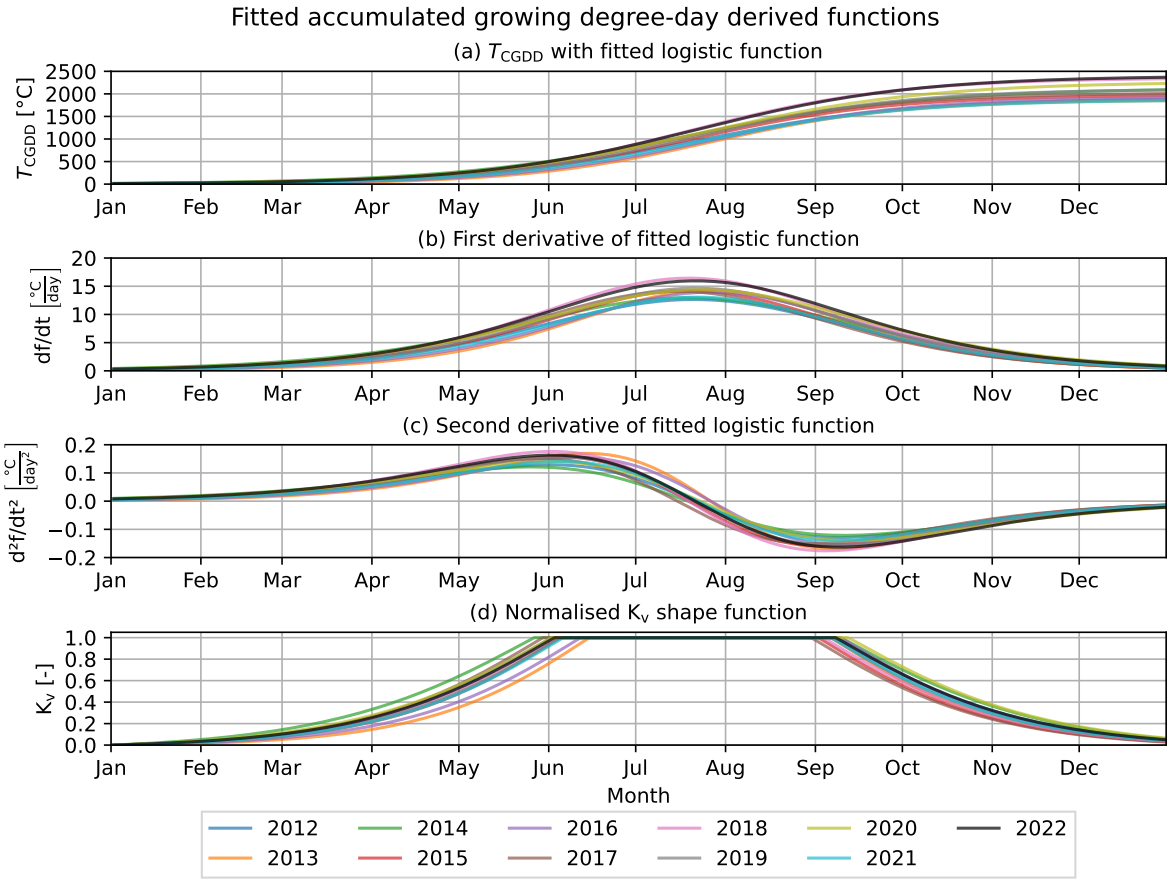


Figure 3.9: (a) An overview of each fitted logistic function on the T_{CGDD} curve. (b) The first derivative of the fitted T_{CGDD} curve. (c) The second derivative of the fitted T_{CGDD} curve. (d) The K_v shape function derived using T_{CGDD} and its derivatives.

Table 3.1: Start and end dates of the growing seasons with duration in days based on the K_v shape functions. The earliest start date, latest end date and longest duration are marked in **bold** text.

Year	Start	End	Length (d)
2012	30-May	07-Sep	100
2013	13-Jun	07-Sep	86
2014	**25-May**	10-Sep	**108**
2015	03-Jun	02-Sep	91
2016	09-Jun	09-Sep	92
2017	28-May	31-Aug	95

Year	Start	End	Length (days)
2018	01-Jun	04-Sep	95
2019	03-Jun	06-Sep	95
2020	30-May	**11-Sep**	104
2021	04-Jun	07-Sep	95
2022	01-Jun	08-Sep	99

3.3. Hydrological models

In addition to the conventional conceptual hydrological model, three modified semi-distributed models were developed using the three proposed phenology integration methods. All four models were first assessed by a sensitivity analysis, indicating how sensitive the models were to a change in value of calibrated parameters. Discharge and transpiration simulations were compared with observational data using several metrics.

3.3.1. Parameter selection & sensitivity analysis

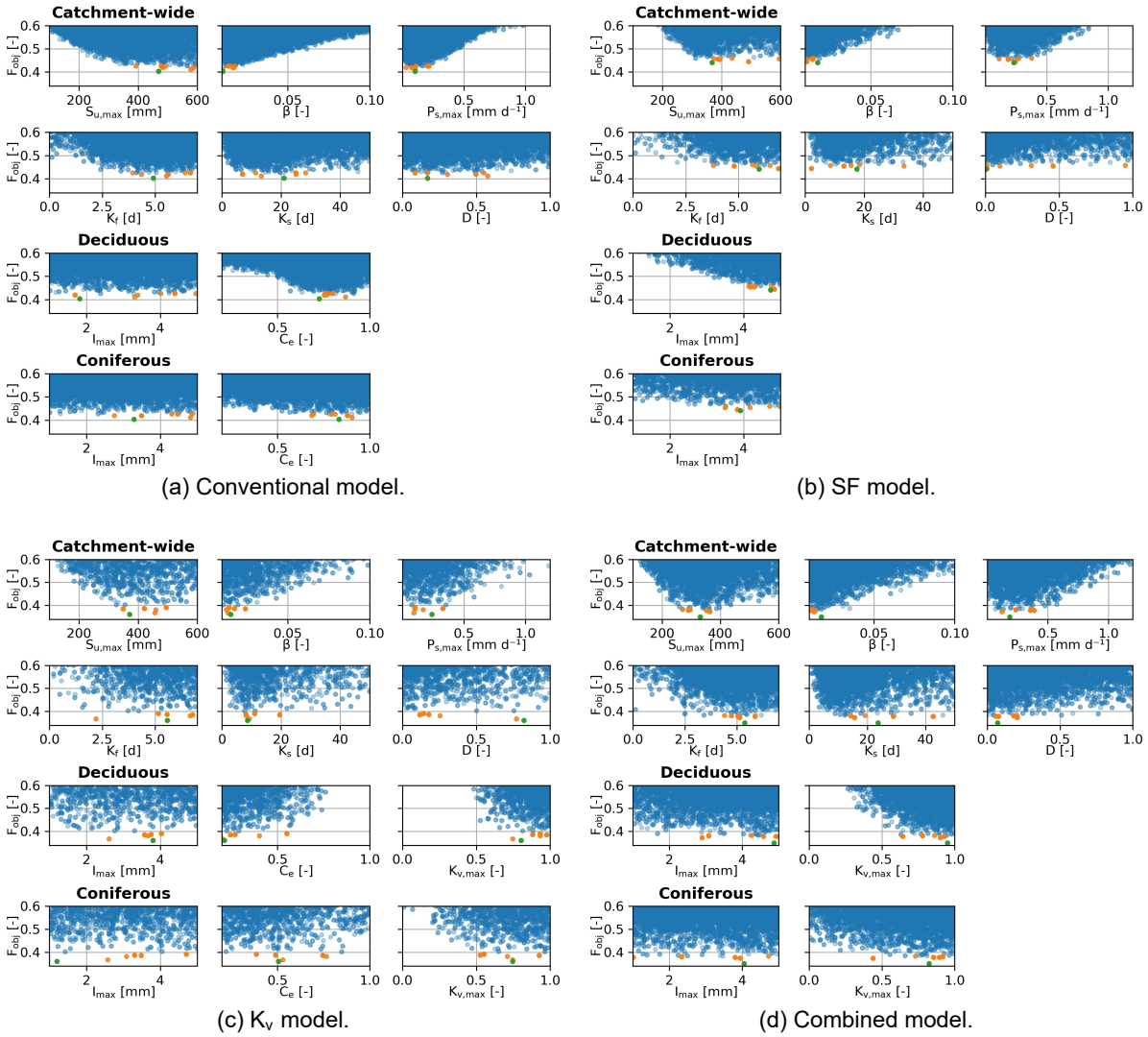
Each of the four models was calibrated using 90 000 Monte Carlo iterations. The parameter configuration with the highest performance (lowest average objective value F_{obj} of the calibration and validation period) was selected. These values are shown in Table 3.2. All evaluated parameter configurations that resulted in $F_{\text{obj}} \leq 0.6$ are shown in Figure 3.10.

Table 3.2: Overview of the selected model parameters estimated during calibration for each model, including units. Selected values follow from a sample size of 90 000 random parameter configurations based on a pre-defined realistic parameter range.

Parameter	**Unit**	**Selected value per model**			
		Conventional	***SF***	***K_v***	***Combined***
Catchment					
$S_{\text{u,max}}$	[mm]	469	368	371	330
β	[−]	0.010	0.018	0.015	0.018
$P_{\text{s,max}}$	[mm d^{-1}]	0.10	0.23	0.24	0.19
K_{f}	[d]	4.9	6.0	5.6	5.4
K_{s}	[d]	20.9	17.4	8.6	23.7
D	[−]	0.17	0.01	0.82	0.070
Deciduous					
I_{max}	[mm]	1.82	4.73	3.80	4.87
C_{e}	[−]	0.73	N/A	0.21	N/A
$K_{\text{v,max}}$	[−]	N/A	N/A	0.80	0.95
Coniferous					
I_{max}	[mm]	3.29	3.91	1.20	4.05
C_{e}	[−]	0.83	N/A	0.51	N/A
$K_{\text{v,max}}$	[−]	N/A	N/A	0.75	0.82

Most parameters had similar values between different models. Judging by only the table, one clear outlier is runoff partitioning coefficient D for the K_v model. Where the other models found around 0.1, the value found for the K_v model was 0.82, reversing the partitioning between preferential recharge R_{sr} and runoff R_{fr} (see Equations 2.16 and 2.17). However, the sensitivity plots for D show an approximately horizontal bottom boundary. This indicates that D had a low sensitivity with respect to F_{obj}, making it more likely that values differ between versions.

For the catchment-wide parameters, all models showed that $S_{\text{u,max}}$, β, $P_{\text{s,max}}$ and K_{f} were sensitive with respect to F_{obj}, while showing a lower sensitivity for K_{s} and D. The vegetation parameter I_{max} showed low sensitivity in all models for both deciduous and coniferous categories, except for the deciduous I_{max} in the SF model. Here, the SF model showed a strong sensitivity for the deciduous I_{max}. C_{e} was used in the conventional and K_v models and was sensitive for the deciduous part of the model, while showing a poor sensitivity for the coniferous part. $K_{\text{v,max}}$ was used in the K_v and combined models, showing a strong sensitivity for deciduous trees and moderate for coniferous trees.



(a) Conventional model.

(b) SF model.

(c) K_v model.

(d) Combined model.

Figure 3.10: Sensitivity analysis of calibrated parameters for the conventional and modified models. Parameter configurations with $F_{\text{obj}} \leq 0.6$ are shown in blue and the 10 best performing parameter configurations in orange. The selected (best performing) configuration is shown in green. Lower values of F_{obj} indicate better performance.

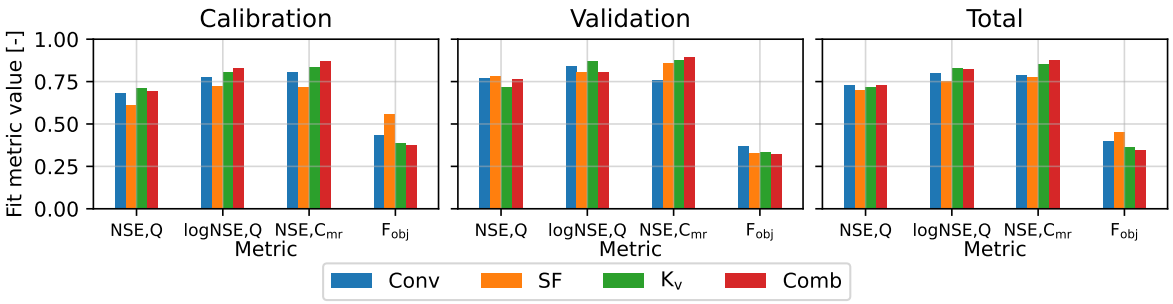
During calibration, F_{obj} was calculated for the calibration and validation periods. When both had an objective value below 1.0, the parameter configuration was saved. This resulted in 52 353 saved configurations for the conventional model, 16 390 for the SF model, 5 509 for the K_v model and 16 594 for the combined model. Although having the lowest number of saved configurations, the K_v model did provide several parameter configurations that performed better than the conventional model and was only outperformed by the combined model in terms of F_{obj}.

3.3.2. Discharge simulations

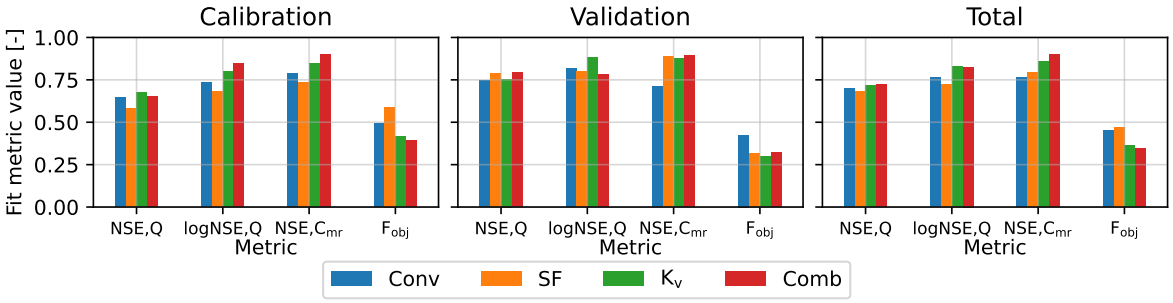
The discharge simulations were assessed using evaluation metrics (N_{NSE}, N_{logNSE}, and N_{RMSE}) and their resulting objective function value. Next, the discharge time series were visually compared with each other and observational data. An overview of atmospheric forcing (precipitation and potential evaporation) with the predicted discharges per model is shown per year and per season in Annex D, Figure D.1.

Evaluation metrics

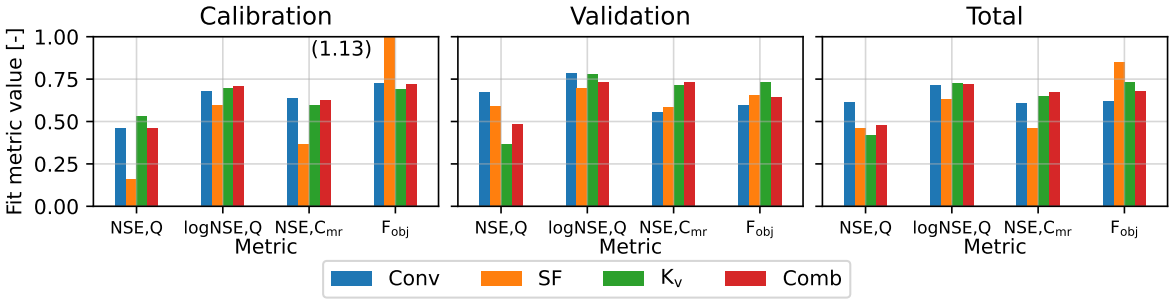
The discharge simulations were first evaluated using the NSE, logNSE and RMSE metrics and their objective function value. These metrics were calculated for the calibration (2015–2019), validation (2020–2022) and combined (2015–2022) periods, excluding the warm-up period (2012–2014) of the model. Additionally, these sets were split into the hydrological winter (October–March) and summer (April–September) semesters to assess seasonal performance using the same metrics. The results are shown in Figure 3.11. In general, models performed better during the validation period than the calibration period.



(a) Annual.



(b) Winter.



(c) Summer.

Figure 3.11: The annual and seasonal performances of the conventional model (Conv), Sap flow model (SF), K_v model (K_v) and the combined model (Comb) for the calibration, validation, and total period. Used performance metrics are N_{NSE} and N_{logNSE} on discharge (Q), N_{NSE} on the monthly runoff coefficient (C_{mr}) with an optimal performance value of 1. F_{obj} is the objective function value which had to be minimised. Annual: Jan–Dec, Winter: Oct–Mar, Summer: Apr–Sep.

Annually, the conventional model achieved similar performances as the K_v and combined models but performed lower on $N_{NSE,C_{mr}}$. The SF model was the lowest performing model during calibration but showed similar performances as the other models during the validation period. The conventional model performed lowest during validation. Overall, the combined model had the highest performances for the annual dataset.

During the winter period, the combined model generally performed highest over calibration, validation, and in total. The K_v model performed similar but slightly lower. The conventional model performed better than the SF model during the calibration period but lower during validation, although both showed higher performance during validation. The largest difference between the conventional and other models was during the validation period for $N_{NSE,C_{mr}}$.

The summer subset demonstrated lower performances for all models compared to the annual and winter periods. The models scored especially low on the $N_{NSE,Q}$ and $N_{NSE,C_{mr}}$ metrics. Generally, the SF model was the lowest performing model, specifically during calibration. In terms of the objective value, the conventional model was the highest performing model during the validation and combined sets, and showed a similar performance as the combined and K_v models during calibration.

The N_{RMSE} was calculated for the calibration, validation, and combined datasets. Like for the metrics discussed above, N_{RMSE} was evaluated for the hydrological winter and summer in addition to the annual time series to assess differences in seasonality. The results are shown in Figure 3.12. Here, models showed a higher N_{RMSE} during winter than during summer. A lower N_{RMSE} indicates a smaller error between the models and the discharge observations. During winter, models performed similarly over the calibration, validation, and total sets. In contrast, their N_{RMSE} was lower during the calibration period and higher during validation.

During summer, the conventional model performed equally (calibration) or better (validation) than the modified models. For the total summer set, the conventional model achieved the highest performance and the three modified models performed similarly to each other. All models were within 0.15 mm d^{-1} of each other.

During winter all models performed similarly. The SF model performed worse than the others during calibration. The SF and combined models showed the highest performances during validation, with the conventional and K_v performing lower. For the winter total set, all models were within 0.09 mm d^{-1} of each other.

For the annual set, models displayed better performances during calibration than during validation. During calibration, the conventional, K_v and combined models performed similar to each other, with the SF model performing lower. The SF model performed highest during validation, while the K_v model performed lowest. The conventional and combined model performed similarly but slightly lower than the SF model. Overall, all models were within 0.06 mm d^{-1} of each other, indicating similar performances with no model being significantly better than the others based on N_{RMSE}.

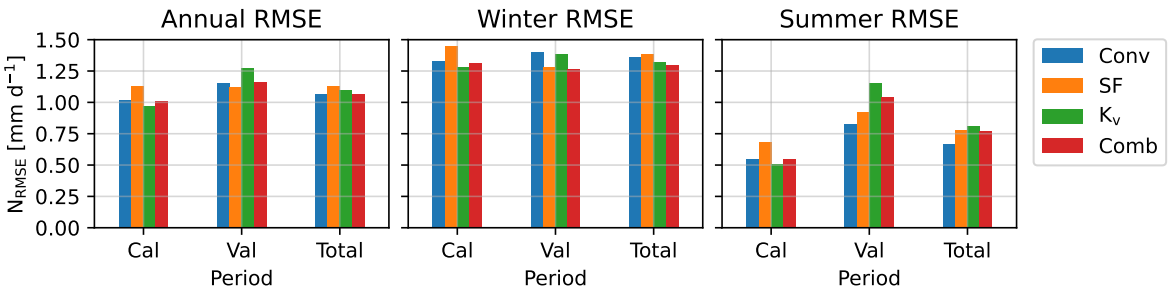


Figure 3.12: The RMSE for discharge predictions compared with observations for the calibration (cal), validation (val) and total period for each model. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model. Annual: Jan–Dec, Winter: Oct–Mar, Summer: Apr–Sep.

Discharge time series

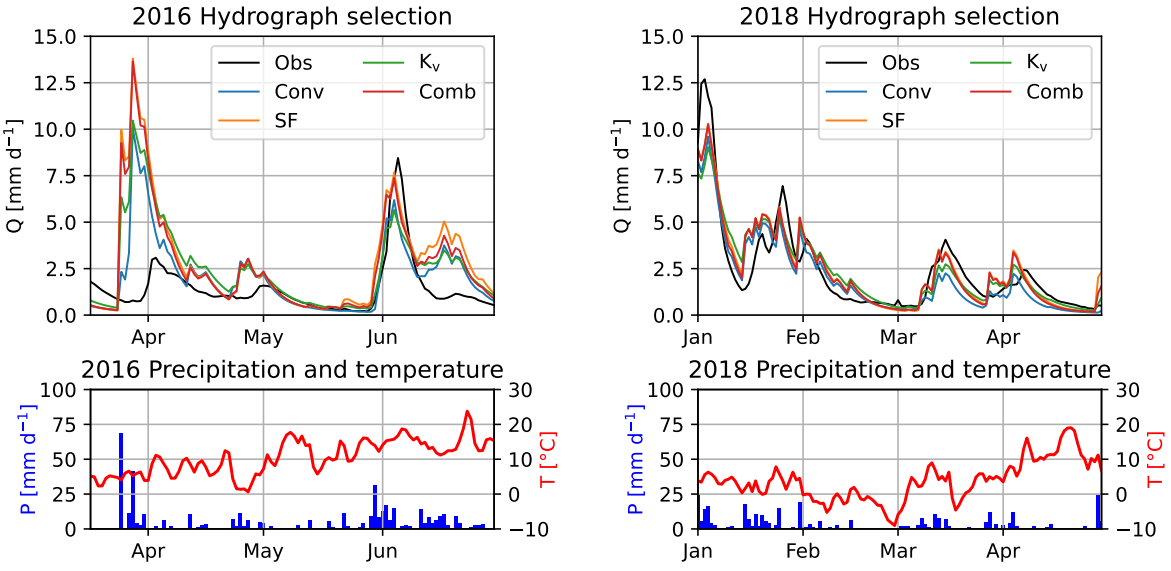
The discharge time series for 2015–2022 are shown per year in Appendix D. To compare years with each other, all figures were shown using the same scaling on the y-axis. For evaluation purposes, zoomed in sections are shown in Figures 3.13 and 3.14.

The models performed similarly during 2015 except at the end of September. Here, the conventional model simulated a peak in discharge where the modified models and observations did not. Two sequential peaks in January were both underestimated by all models. However, little precipitation was recorded during this period.

The reverse occurred in March/April 2016. Here, significant precipitation was registered, but no immediate peak was measured (Figure 3.13a). The models displayed a strong response to this precipitation, resulting in an overestimation compared to the observations. The peak in June 2016 was simulated well, but all models simulated a second peak after, which was not observed.

In 2017 the models were able to simulate discharges that closely matched the observations. No high discharge peaks occurred during this year. The largest discharge was observed in December, which was simulated quite well. Several small peaks were simulated during October and November by the conventional and SF model, which were not observed. The K_v and combined model correctly did not produce these peaks.

The discharge peaks during January 2018 were decently captured by the models. The modified models simulated the lower peaks during March and April better than the conventional model, which underestimated these (Figure 3.13b). The first peak in December was significantly overestimated by the conventional model while reproduced by the modified models. The second peak was simulated well by all models.



(a) 2016 hydrograph selection.

(b) 2018 hydrograph selection.

Figure 3.13: (a) Selection of two discharge peaks in 2016, with the first one missed by the observations. (b) The underestimation of discharge by the conventional model early 2018. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

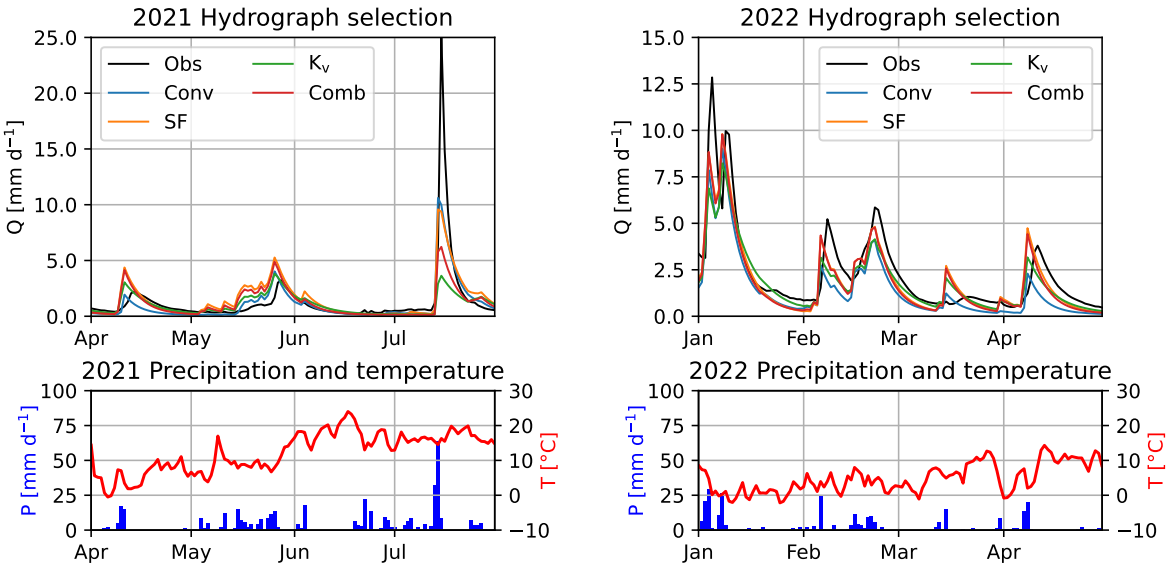
In 2019, a sharp discharge peak occurred in February, which was underestimated by all models. The models produced nearly identical responses to this event. A precipitation peak was observed, but not of the magnitude that was expected to produce a discharge peak this strong. Two smaller peaks in March and May were simulated well, with the latter one being simulated slightly better by the modified models compared to the conventional model. In October the conventional model first produced a peak that was not observed nor simulated by the other models. A second peak was produced by all models, which was not observed in this magnitude. The K_v and combined model were closer to the observations

than the conventional and SF model for this peak. November and December were simulated similarly by all models.

Higher discharges occurred in February and March 2020 with several peaks. The higher discharges were produced well by the models, but the peaks were underestimated. The remaining part of the year was simulated similarly for all models, with the conventional model overestimating a small peak in November.

In 2021, there were two peaks in January, which were not simulated by the models. During this time, no peaks in precipitation were registered. Lower peaks during March and April were underestimated by the conventional model, but overestimated by the modified models. All models overestimated the peak in May, with the conventional model's simulation being closest to the observations. A sharp discharge peak occurred in July which was underestimated significantly by all models (Figure 3.14a). The conventional and SF models produced similar results, which were closer to the observations than the simulations by the K_v and combined models. Here, the K_v model only simulated a small peak. A lower peak in November was overestimated by the SF model, simulated well by the conventional and combined models, and underestimated by the K_v model. Discharge in December was underestimated by all models.

The last simulated year was 2022. Here, peaks in January, February and April were underestimated by the models. The modified models produced simulations closer to the observations than the conventional model (Figure 3.14b). The observed peaks were not only higher, but also kept rising for a longer time. The observational peaks still increased, while the simulations already showed decreasing discharges.



(a) 2021 hydrograph selection.

(b) 2022 hydrograph selection.

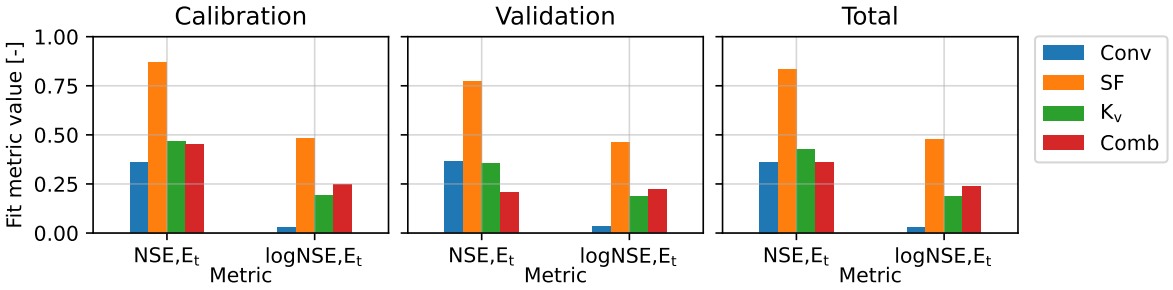
Figure 3.14: (a) Selection of three discharge peaks in 2021, with the first two being overestimated by the models and the last peak being underestimated. (b) The better estimation of the modified models compared to the conventional model for the peaks early 2022. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

3.3.3. Transpiration simulations

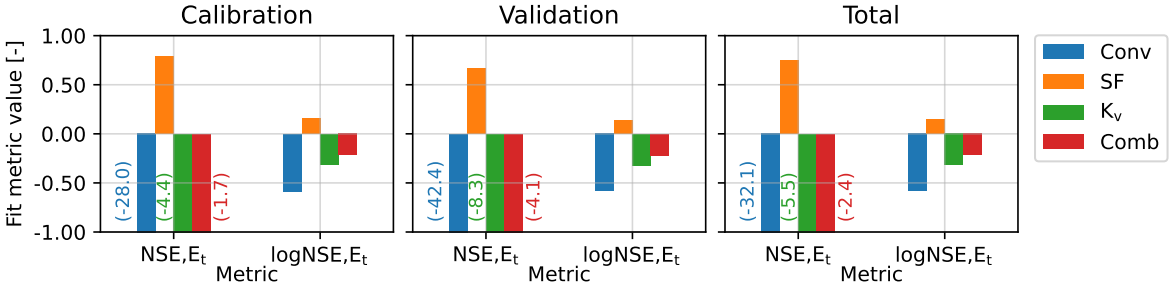
The transpiration (E_t) simulations were first assessed using evaluation metrics (N_{NSE}, N_{logNSE}, and N_{RMSE}), using transpiration estimates from the GLDAS-2.2 product as a reference. Next, the time series were visually compared with each other and GLDAS-2.2 transpiration estimates using the 7-day rolling mean. All model evaluation metrics were computed on the original daily values.

Evaluation metrics

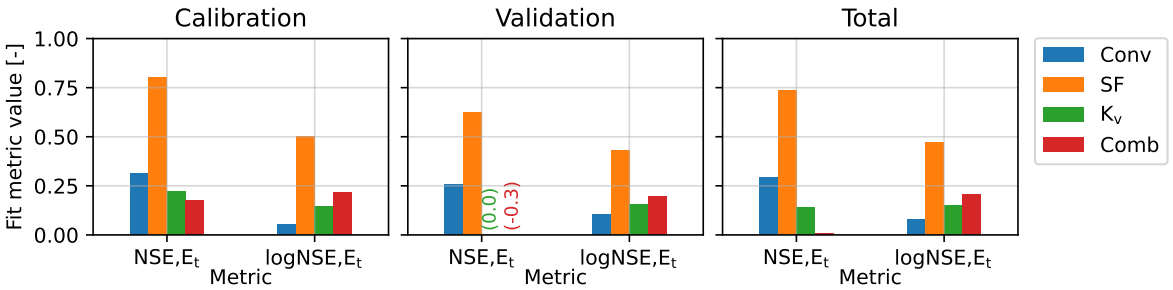
Using the same approach as during the discharge evaluation, transpiration simulations were first evaluated using the NSE, logNSE and RMSE metrics. These metrics were calculated for the calibration (2015–2019), validation (2020–2022) and combined (2015–2022) periods, excluding the warm-up period (2012–2014) of the model. Additionally, these sets were split into the hydrological winter (October–March) and summer (April–September) semesters to assess seasonal performance using the same metrics. The results are shown in Figure 3.15a.



(a) Annual.



(b) Winter. Values below -1 are displayed in text.



(c) Summer. Negative values are displayed in text.

Figure 3.15: The annual and seasonal model performances on transpiration of the conventional model (Conv), Sap flow model (SF), K_v model (K_v) and the combined model (Comb) for the calibration, validation, and total period. Annual: Jan–Dec, Winter: Oct–Mar, Summer: Apr–Sep.

The annual metrics displayed in Figure 3.15a showed that N_{NSE} scores were generally higher than N_{logNSE} scores. This indicates that the models performed better for higher (peak) transpiration than for lower rates. During calibration, the conventional model performed slightly lower on N_{NSE} than the K_v and combined models. The SF model was the only model that had a good N_{NSE} score during calibration, performing significantly better than the other models. During this period, the SF model performed

lower on the logNSE metric, but still considerably better than the other models. The simulation of the conventional model was slightly better than using the mean transpiration. All models scored lower during the validation years. The SF model performed significantly better on both metrics than the other models. The conventional model performed similarly to the K_v model on N_{NSE}, where the combined model performed lowest. The logNSE metrics showed a similar result as during calibration. For the total set, the SF model was the only good performing model in terms of N_{NSE}. The other models performed similarly to each other on the NSE metric. Again, all models scored lower on N_{logNSE} than on N_{NSE}. The SF model performed the highest, the K_v and combined models performed similarly to each other but lower, and the conventional model performed the lowest, similar to the mean.

The hydrological winter semester displayed negative values for both metrics during calibration, validation, and in total for the conventional, K_v, and combined models. These models performed lowest on N_{NSE} and higher on N_{logNSE} during all three periods. The SF model was the only model performing well based on the NSE metric. Its scores on N_{logNSE} were low, but still positive.

Performances were higher during the summer semester. During the calibration period, the conventional model performed higher than the K_v and combined models on N_{NSE}, but lower on N_{logNSE}. The SF model achieved a high performance on N_{NSE}. It performed lower on N_{logNSE} but significantly better than the other models. The models performed lower during validation, with the SF model still outperforming the other models on both metrics. The conventional model scored better on N_{NSE} than the K_v and combined models. The latter two performed equally or lower than when using the mean. All modified models performed better on N_{logNSE} than the conventional model. In total, the SF model had a good performance on N_{NSE} and lower on N_{logNSE}, but still better than the other models. The conventional model scored the lowest on N_{logNSE}, but performed better on N_{NSE} compared to the K_v and combined models.

The annual and seasonal N_{RMSE} values were calculated for the calibration, validation, and combined time series for all models. The results are shown in Figure 3.16. It shows that the SF model performed significantly better than the other models. Annually, the other models had double the N_{RMSE} values shown by the SF model, all three performing similarly. The conventional model had a significantly larger error during the winter period than the modified models. In contrast, the SF model had a low error during this time. During the summer, the conventional model performed better than the K_v and combined models. Again, the SF model had the lowest errors.

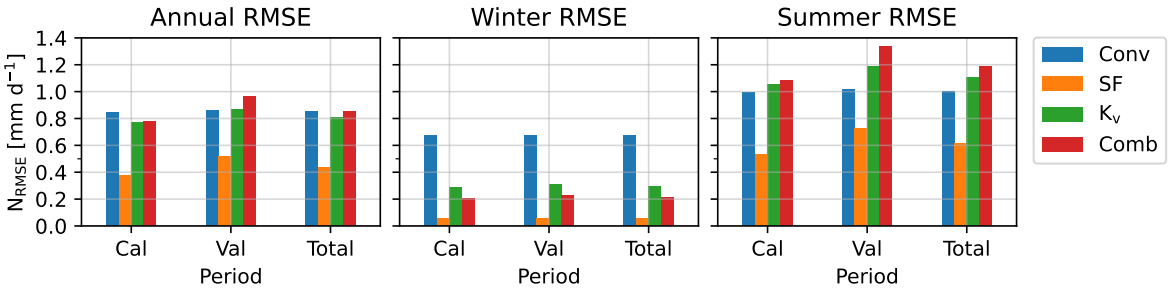


Figure 3.16: The N_{RMSE} for discharge predictions compared with observations for the calibration (cal), validation (val) and total period for each model. Conv = conventional model, SF = sap flow model, K_v = K_v model, comb = combined model. Annual: Jan–Dec, Winter: Oct–Mar, Summer: Apr–Sep.

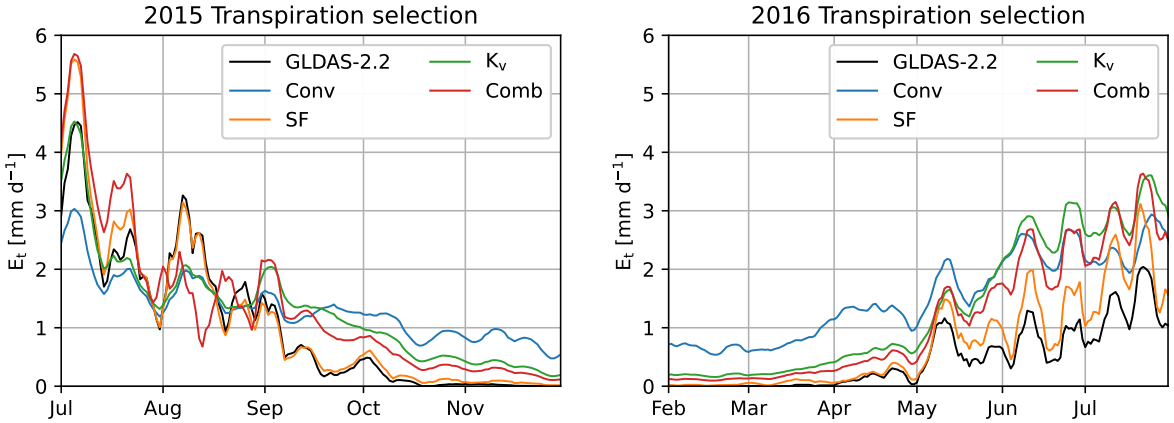
Transpiration time series

Time series were visually compared with each other and the reference GLDAS-2.2 transpiration estimates. For visualization purposes, the daily transpiration time series were smoothed using a 7-day moving average to reduce short-term fluctuations and highlight seasonal dynamics. The time series are shown per year in Annex E. Zoomed in selections from these time series are shown in Figures 3.17 and 3.18 for evaluation purposes.

During the first months of 2015, the conventional model simulated much higher transpiration rates than the modified models and reference data. The K_v and combined models simulated higher rates too.

The SF model was close to the reference data until April, but overestimated from May until July. During this period, the conventional model generally underestimated transpiration while the other models overestimated. The peak in July was best simulated by the K_v model. From August to December, the SF model was similar to the reference (see Figure 3.17a). The other models underestimated during August and overestimated in the remaining months.

Transpirations were lower during 2016 compared to 2015. All models overestimated throughout the year. Again, the conventional model overestimated the transpiration during the first months of the year (Figure 3.17b). Between May and September, the SF model was closest to the reference data, while the K_v and combined models showed the highest values. The models showed the same general shape and timing as the reference, but with values that were too high. From September onwards, the SF model closely resembled the reference.



(a) 2015 transpiration selection (7-days rolling mean).

(b) 2016 transpiration selection (7-days rolling mean).

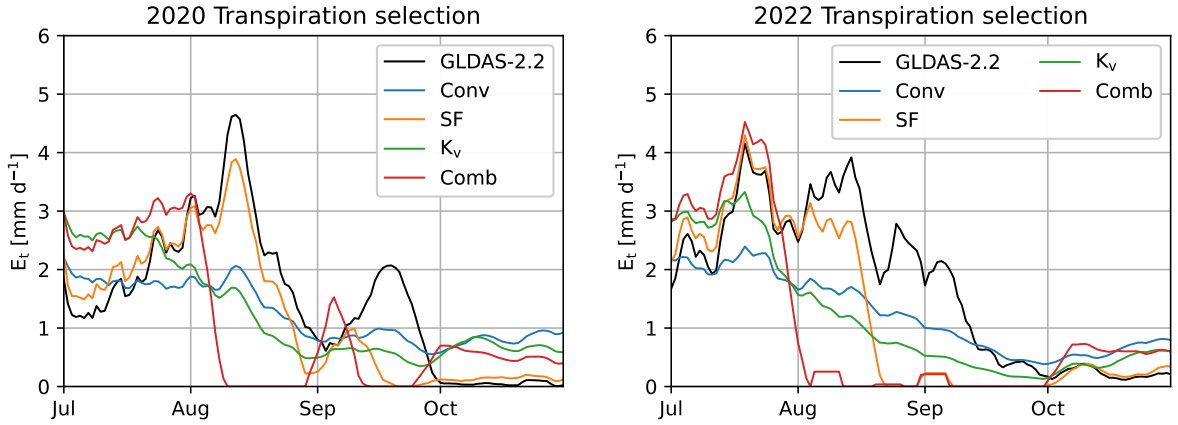
Figure 3.17: Two selections from the transpiration time series from 2015 and 2016 showing the general behaviour of the different models during different times of the year. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

The SF and combined models overestimated transpiration peaks in June 2017. In contrast, the conventional model simulated the first peak well but was insensitive to the second peak, underestimating it. The K_v was closer to the reference. From July onwards, the SF model was close to the reference data. The other models overestimated transpiration during this period.

Other years showed similar behaviour as described above for 2015–2017. Generally, the conventional model overestimated the transpiration during the winter. The K_v and combined models performed better during the first few months but also overestimated during the last few months of the year. The SF model was usually closest to the reference for these months. During the summer, the conventional model over- and underestimated peaks, seemingly not very sensitive to transpiration fluctuations displayed by the reference data and other models. The combined model generally overestimated during this time, while the K_v was often between the conventional and combined model. The SF model was usually closer to the reference but often overestimated during May and June. From July onwards, the SF model was generally similar to the reference.

The combined model showed anomalous behaviour during August and September in 2018, 2020 and 2022. Here, it showed events of strong decline, producing transpiration simulations zero as shown in Figure 3.18a. The same occurred for the SF model towards the end of September 2020. A more extreme example was 2022, shown in Figure 3.18b. Here, the combined model produced almost no transpiration during August and October. The SF model did the same but dropped to zero later. Both increased again in October in response to precipitation.

The semi-distributed structure of the models was based on the two tree categories (coniferous and deciduous). In terms of covered area, the deciduous part of the catchment was larger (76%) than the coniferous part (24%). Furthermore, each category had its own seasonality. Consequently, each part contributed differently to the total discharge and transpiration.



(a) 2020 transpiration selection (7-days rolling mean).

(b) 2022 transpiration selection (7-days rolling mean).

Figure 3.18: Two selections from the transpiration time series showing transpiration anomalies for the K_v and SF models in 2020 and 2022, both having periods of a zero transpiration simulation. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

The individual in contributions to the total transpiration during 2015 are shown in Figure 3.19 for each model. Most other years had similar behaviour as 2015, which was therefore used to describe the contribution differences between models in general. All models showed a larger contribution by the deciduous part of the model compared to the coniferous part. Furthermore, all models simulated higher transpiration rates for both parts during summer than during the winter. The contribution as a percentage of the total transpiration (fractional contribution) of each part is shown in Figure 3.20 for 2015.

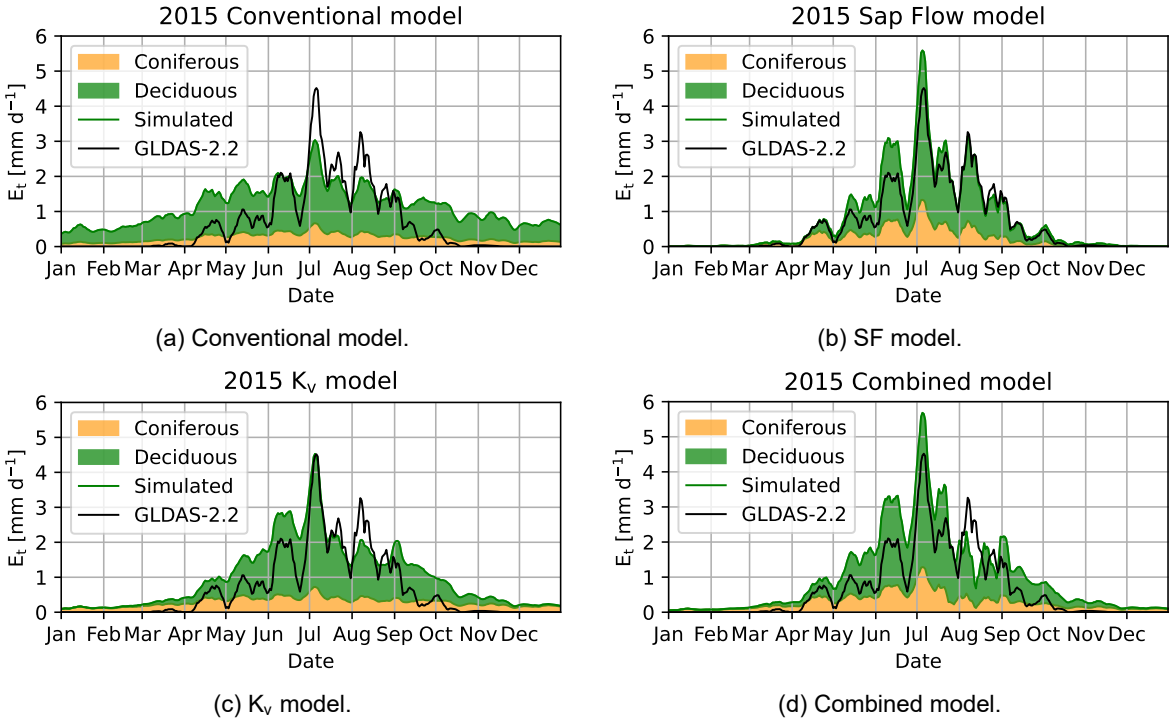
The conventional model simulated relatively high transpiration rates for the deciduous part during the winter. The fluctuations in transpiration were mainly captured by the deciduous part of the model, as the coniferous part showed a smoother time series. The relative contribution to the total transpiration did not change over time as shown in Figure 3.20a, indicating no seasonal differences in relative contribution.

In contrast to the conventional model, the SF model captured the transpiration fluctuations in both parts of the model. The deciduous part showed a larger contribution during the summer. The coniferous part of the model contributed more during April than the deciduous part, even though it covered a significantly smaller part of the catchment. Coniferous trees had the highest relative contribution during January to March. In April the ratio changed, as deciduous trees started to contribute more.

The K_v model had a relatively constant contribution of the coniferous part, but slightly higher during the summer. The deciduous part contributed less during the winter and more during the summer. This is also shown in Figure 3.20c, displaying clear seasonality. The relative contribution of coniferous trees increased slightly during July and August, which was still during the plateau of the K_v shape function. The plateau ends early September, which is also when the contribution by the deciduous trees decreased.

The combined model displayed similar behaviour as the K_v. The main difference was that the coniferous trees showed more fluctuations during the summer for the combined model. Where the K_v model showed an increasing coniferous contribution during July, the combined model showed fluctuations in contribution. These fluctuations were not visible in the fractional contribution figure of the SF model (Figure 3.20b).

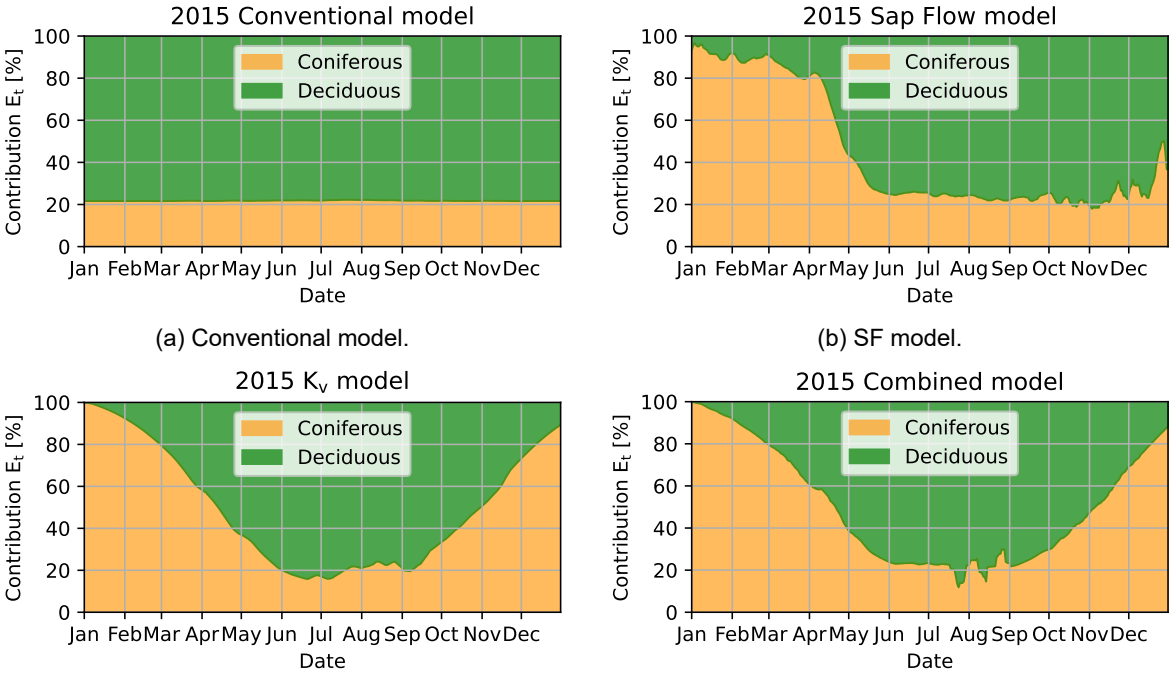
Noticeable was that the relative contributions for the modified models were not continuous between December and the next year's January. The K_v and combined models showed a small jump while this difference is larger for the SF model.



(a) Conventional model. (b) SF model.

(c) K_v model. (d) Combined model.

Figure 3.19: The contribution of the coniferous and deciduous parts of the models to the total transpiration estimate for 2015. GLDAS-2.2 is the reference transpiration for the total catchment.



(a) Conventional model. (b) SF model.

(c) K_v model. (d) Combined model.

Figure 3.20: The fractional contribution of the coniferous and deciduous parts to the total transpiration per model for 2015.

4 Discussion

This chapter discusses the main findings of the study in relation to the research objectives. First, the performance and limitations of the sap flow prediction model are considered, with particular attention to the role of environmental predictors and model constraints. Next, the approaches used to incorporate phenology into the hydrological models are evaluated, highlighting how differences between deciduous and coniferous trees were captured. Finally, the conventional and modified hydrological models are compared in terms of their ability to simulate discharge and transpiration, and the implications of these results for modelling heterogeneous forest catchments are addressed.

4.1. Sap flow prediction model

This section discusses the developed sap flow prediction model. First, it discusses the normalised accumulated growing degree-day, which was derived from temperature data. Next, the partial dependencies are discussed, which show the relation between individual environmental predictors and predicted $v_{\text{sf,n}}$. Finally, the model's performance is discussed.

4.1.1. Normalised accumulated growing degree-day

To predict $v_{\text{sf,n}}$, $T_{\text{CGDD,n}}$ was needed as input. By using the normalised version of T_{CGDD}, it became possible to use the general shape as an indicator for seasonality. Warmer years such as 2014 and 2020 reached higher $v_{\text{sf,n}}$ earlier in the year, indicating an earlier start of the growing season.

Without scaling, 2018 and 2022 were the years with the highest T_{CGDD} at the end of the year but also in June, reaching 500 °C relatively soon. However, when scaled, this was not evident as both years performed close to the average compared with the other years. Due to the scaling effect, years with a high T_{CGDD} at the end of the year were flattened more than colder years with a lower maximum T_{CGDD}. As warm years were flattened more, some cases occurred where the warmer spring did not clearly show in $T_{\text{CGDD,n}}$, limiting the capacity of $T_{\text{CGDD,n}}$ to predict the timing of phenological events such as budburst and leaf-out.

T_{CGDD} is potentially a more suitable indicator of phenological timing in spring, but its variability towards the end of the year could reduce the accuracy of GAM predictions. A possible improvement could involve the use of T_{CGDD} as an indicator of the start of the growing season, and using a different indicator to predict the end of the growing season.

4.1.2. Partial dependencies

The partial dependencies of the unconstrained GAMs generally showed expected trends and relations between the environmental variables and $v_{\text{sf,n}}$. Accordingly, the shape constraints matched the general shapes of the unconstrained GAMs, indicating that these were the right constraints. This was further supported by the small difference in fitting performance metric values between the unconstrained and constrained models.

Based on literature, $v_{\text{sf,n}}$ was expected to increase with temperature up to an optimum of 20–25 °C, followed by a decline as stomatal closure limited transpiration and sap flow. However, this pattern was not evident in the partial dependencies derived from the GAMs. Scatterplots between temperature and $v_{\text{sf,n}}$ (not shown) likewise failed to display the anticipated response. The reason for this deviation remains unclear.

The relative humidity partial dependency plots for all GAMs showed a decrease in $v_{\text{sf,n}}$ with increasing h, suggesting that sap flow and transpiration decline as relative humidity rises. This was expected, since higher h corresponds to a lower vapour pressure deficit, which limits transpiration.

Incoming shortwave radiation generally exhibited a monotonically increasing relationship with $v_{\text{sf,n}}$ for the deciduous GAMs. The unconstrained coniferous GAM followed the same overall pattern but displayed local deviations, particularly at the highest R_{s} values (above 900 W m^{-2}), where a moderate negative slope appeared. This coincided with wider 95% confidence intervals, likely due to the limited number of observations in this range. As this behaviour was absent in the deciduous GAMs and unsupported by literature on stomatal closure under high R_{s}, a monotonically increasing shape constraint was imposed. This adjustment removed the unexpected negative slope and enforced a more realistic model response.

The weighted averaged volumetric water content (θ_{avg}) showed itself to be strongly limiting $v_{\text{sf,n}}$ at low values, which was expected and in line with literature. For slightly higher θ_{avg} conditions, its influence on $v_{\text{sf,n}}$ was approximately constant, showing a slight increase for higher θ_{avg} values. For this range of values, the unconstrained coniferous GAM showed strong local fluctuations, likely caused by the smaller dataset compared to the deciduous dataset, resulting in overfitting. Above 0.20 m^3 m^{-3} θ_{avg}, the deciduous GAMs showed a strong negative slope. This behaviour was also observed by Tian et al. (2018) for spruce trees, a coniferous species, while not observed in this study for coniferous trees. The coniferous GAMs showed a limited negative slope above 0.20 m^3 m^{-3} θ_{avg}. The negative slopes indicate that $v_{\text{sf,n}}$ decreased for higher θ_{avg} values. Following Feddes et al. (1976), this can be explained by the decreasing oxygen levels in the soil for higher soil moisture rates, which limit water uptake by the roots. The negative slopes at high θ_{avg} values supported the choice of a concave shape constraint, as this increased the smoothness while maintaining the ability to show the general trends for both coniferous and deciduous trees.

The partial dependence plots for normalised accumulated growing degree-days ($T_{\text{CGDD,n}}$) revealed the expected seasonal pattern for deciduous trees. At low $T_{\text{CGDD,n}}$ values, before the growing season, $T_{\text{CGDD,n}}$ had a negative contribution to $v_{\text{sf,n}}$, followed by a positive slope as the season progressed. After the curve reached its maximum, $v_{\text{sf,n}}$ gradually declined, with a sharp decrease observed towards the end of the year, corresponding to leaf shedding. This strong seasonality was not observed for coniferous trees. These trees exhibited a slight increase in $v_{\text{sf,n}}$ at low $T_{\text{CGDD,n}}$ values, followed by a gradual decline, similar to deciduous trees. In contrast to the deciduous trees, no distinct decrease occurred at the end of the growing season. This pattern was consistent with the assumption that coniferous trees display limited seasonality.

4.1.3. Model performance

The models' performance was evaluated based on their generalisability and ability to predict $v_{\text{sf,n}}$, which was compared to observational data.

Model generalisability

The R^2 and RMSE metrics (N_{R^2} and N_{RMSE}, respectively) were comparable between training and testing, with differences of less than 1%. This suggests that the models effectively captured the underlying data patterns without overfitting, as their performance remained consistent on unseen data. This also indicated that

It is important to note that the original data were time series. Randomly splitting data points from the series into two sets may produce very similar datasets, as conditions at a given time are often similar to those just before or after. Consequently, assigning one point to set A and the surrounding points to set B may result in comparable model performance. Alternative splitting strategies, such as using alternating days, could provide a more distinct separation between datasets and improve model evaluation.

Prediction performance

The prediction performance of the constrained GAMs was evaluated based on three metrics. The NSE metric mainly indicates how well the model performs on matching the peaks with the observations. The logNSE uses a natural logarithmic transformation on the data before applying N_{NSE}, making it a better

indicator of performance on low $v_{sf,n}$ predictions. N_{RMSE} is the average difference between predictions and observations, and is an indication of the model's accuracy.

The constrained GAMs both showed a lower model performance for 2020 compared to 2019. As described in Section 3.1.3, the deciduous GAM overall had a better performance compared to the coniferous model. This is possibly the result of the smaller available data set for coniferous trees. Where the deciduous dataset contained 27 trees, the coniferous dataset contained 6 trees. Furthermore, the observed deciduous trees were spread across the catchment, while coniferous trees were all located at the north or south borders of the catchment. Therefore, the coniferous dataset was more susceptible to spatial biases. Both GAMs had a lower performance during 2020 compared to 2019. Furthermore, both models had higher N_{NSE} values compared to N_{logNSE}, indicating that the models performed better in predicting higher $v_{sf,n}$ values.

A potential explanation for the difference between the 2019 and 2020 model performances is that the models had more data from 2019 as more data were available that year. This could have biased the GAMs more towards 2019. If this was the case, it would also imply that the GAMs were potentially not as generalizable as previously suggested.

Both GAMs had N_{NSE} and N_{logNSE} scores above 0.80 for the total observational data range (2019–2020). Following the model performance categorisation suggested by Ritter and Muñoz-Carpena (2013), this indicates a good model performance over this period.

The time series showed that the models were able to reproduce the observations well, except for the coniferous model in 2020 which often under- or overestimated $v_{sf,n}$. The coniferous model for 2020 showed it was able to still capture small fluctuations at peaks during May, while underestimating the magnitude. This would indicate that the model was able to perform precisely, but not accurately during those circumstances. The deciduous model showed near zero $v_{sf,n}$ during the winter and early spring, which was expected as no leaves were present during that time. This means that it was able to capture this seasonality without being trained on data from these months. The coniferous model showed less seasonality, which was expected because these trees do not shed their leaves.

The one-to-one plots indicated that the deciduous model predicts values closer to the observations. Furthermore, the daily predictions exhibited less variability than the hourly time series. The point clouds were distributed evenly around the 1:1 line, indicating no systematic tendency to under- or overestimate values. In addition, the approximately constant scatter across the full range of observed values indicates homoscedasticity of the model residuals, suggesting that predictive uncertainty remained approximately constant across the range of $v_{sf,n}$ values.

4.2. Phenology inclusion

Three methods were designed to incorporate tree phenology in the hydrological models, which were based on two approaches. The first method involved direct inclusion of sap flow dynamics. The second method indirectly included sap flow dynamics by partitioning evaporation based on seasonality, which was derived from temperature.

4.2.1. Sap flow

Coniferous trees showed higher $v_{sf,n}$ values during the first few months of the year, which was expected as deciduous trees were still in dormancy during that period. When deciduous trees came out of dormancy, their sap flow increased. During May, both tree categories started to show very similar behaviour. This could indicate that the deciduous trees had fully developed their leaves.

Compared with the deciduous trees, the coniferous trees exhibited lower $v_{sf,n}$ values during August and September in some years. A possible explanation is the influence of soil moisture, as during these months θ_{avg} is typically low. Oak trees generally have a deeper root system than beech and the coniferous species (Fabiani et al., 2022). Due to their access to deeper water, oaks are less drought-sensitive which could have influenced the deciduous sap flow, resulting in higher $v_{sf,n}$ compared to coniferous trees. Although theoretically plausible, the partial dependence plots for θ_{avg} from the GAMs (Figures 3.2 and 3.3) did not confirm this hypothesis.

During November and December, $v_{\text{sf,n}}$ was near zero for deciduous trees, as expected because they had shed their leaves by this time. Consequently, no sap flow was anticipated. For the same period, coniferous trees also exhibited low $v_{\text{sf,n}}$ values but showed some fluctuations, indicating that limited sap flow, and thus transpiration, occurred. This is plausible because, unlike deciduous trees, conifers retain their leaves, allowing continued, albeit limited, photosynthesis when conditions are suitable.

4.2.2. Evaporation partitioning

The derived K_v shape functions captured an earlier start of the growing season in years with a warmer spring, as expected. The end of the growing season indicated by the function typically occurred in early September, coinciding with the end of summer. By basing the function on temperature, which is strongly linked to phenological timing, the growing season was estimated dynamically, allowing for interannual variation. This approach is more realistic than using fixed calendar days and, although a rough estimate, likely represents an improvement over static evaporation partitioning.

This method roughly indicated if the growing season started earlier or was delayed compared to an average year, but the exact timing was not determined using the K_v shape function. A more precise timing, e.g., using T_{CGDD} values, would likely show slightly different timing of the plateau. Additionally, the gradual increasing/decreasing slope of the K_v shape function is a conceptualisation of changing partitioning over time and was not based on specific literature. Therefore, this approach could, and is recommended to, be improved in future research.

4.3. Hydrological models

In addition to the conventional model, three modified hydrological models were developed. First, parameter selection after calibration and the sensitivity analysis are discussed. This is followed by the evaluation of discharge and transpiration simulations produced by the models.

4.3.1. Parameter selection & sensitivity analysis

The sensitivity analysis showed that several parameters had a poor sensitivity with respect to the objective function. This could indicate that the processes related to these parameters are unrealistic or that the objective function was not suitable to assess these parameters. Furthermore, poor sensitivity indicates the issue of equifinality, which is a common problem in hydrology. It increased the uncertainty in the model's correspondence with reality. Other parameters did show a strong sensitivity, indicating a significant influence on the model's internal processes. The sensitive parameters generally found similar values across the four models.

A noticeable difference was observed in the sensitivity of C_e between coniferous and deciduous trees. In both the conventional and K_v models, C_e for deciduous trees exhibited strong sensitivity, whereas for coniferous trees it was weak. This suggests that deciduous trees may be more sensitive to limitations in available soil moisture than coniferous trees, which is contrary to the general expectation that, on average, deciduous trees are more drought-resistant (Jiménez-Rodríguez et al., 2024).

Another significant difference between the conventional and K_v models was that the conventional model found higher C_e values during calibration, while the K_v found lower values. A possible explanation is that the K_v compensated for the multiplication by the $K_{v,a}$ factor through lower C_e values. A low C_e value caused the min term in Equation 2.31 to return a value of 1 more often. Therefore, E_t would have followed the K_v shape function more closely.

There is a large difference in the number of parameter configurations that passed the thresholds to save the configuration. The conventional model had over three times as many saved combinations (58%) as the SF and combined models (both 18%). Under the same conditions, the K_v model saved 6% of its configurations. While the conventional model produced configurations with F_{obj} below the threshold, it was not able to reach the same low objective values as the combined and K_v models. This indicates that the conventional model was more robust but was still outperformed by the K_v and combined models in terms of best found objective values.

Because the modified models are more sensitive to parameter configuration, they would require longer calibration (larger Monte Carlo sample size) to find their optimal combination compared to the conventional model.

4.3.2. Discharge simulations

The discharge simulations generated by the conventional and modified models were evaluated based on performance metrics and a visual inspection of the resulting discharge time series.

Evaluation metrics

In general, the models exhibited higher performance during the validation period than during calibration, which is counter-intuitive. Normally, a model trained on a calibration set would perform similarly or worse on the validation set. However, this can be explained by the model selection method used in this study. All parameter configurations that resulted in a F_{obj} below a set threshold during calibration were tested during validation. The parameter configuration that resulted in the highest performance over both calibration and validation was selected. Therefore, it is possible that the selected parameter configuration was not the highest performing during the calibration period, but was compensated for by a higher performance during validation. This selection method ensured that the model was generalizable (good performance over both periods), but consequently can perform higher during validation than calibration.

No significant differences in precipitation or potential evaporation were observed between the calibration and validation periods. However, the validation period spanned three years, compared with five years for calibration. The selected models may have performed better during validation due to the smaller dataset, potentially compensating for lower performance during calibration. As a result of the selection method, the chosen models could have been biased toward the validation set.

The models showed higher performances during winter than during summer. Models scored lower on $N_{\text{NSE,Q}}$ and $N_{\text{NSE,C}_{\text{mr}}}$, specifically during this time compared to the winter. This is likely caused by the low discharge during summer. With low discharge, the few peaks that did occur had a relatively large influence on these metrics due to their sensitivity. The $N_{\text{logNSE,Q}}$ is less sensitive to peaks as it typically is stronger influenced by the low values. Therefore, the low base discharge had a lower impact on $N_{\text{logNSE,Q}}$ compared to $N_{\text{NSE,Q}}$ and $N_{\text{NSE,C}_{\text{mr}}}$.

Generally, the K_v and combined models had higher performances than the conventional model for $N_{\text{NSE,C}_{\text{mr}}}$. This was expected as $N_{\text{NSE,C}_{\text{mr}}}$ is an indirect measure of actual evaporation (E_a). Because the modifications were designed to improve E_t and the partitioning between E_i and E_t, E_a should theoretically be more realistic. Consequently, $N_{\text{NSE,C}_{\text{mr}}}$ should indicate an improvement with respect to the conventional model. The SF model performed lower than the conventional model for this metric during calibration but higher during validation.

The models showed similar performances compared to each other in terms of annual N_{RMSE}. The combined and K_v were the highest performing models during winter. During summer, the conventional model was the highest performing model with the modified models showing similar performances, but slightly lower than the conventional model. The N_{RMSE} was much lower during the summer for all models, which was the result of the low discharge during this period compared to the winter.

Discharge time series

The comparison of observed and simulated discharges indicated that the models generally produced similar results. Sharp and high peak discharges were consistently underestimated by all models, whereas lower peaks were captured more accurately. During the summer, the modified models occasionally simulated low discharges that were not observed, a phenomenon less frequent in the conventional model.

Generally, the modified models produced slightly better discharge simulations between November and April compared to the conventional model. In contrast, the conventional model performed slightly higher between May and September compared to the modified models. This is in line with the calculated performance metrics for the hydrological summer and winter semesters.

In some cases, high observed discharge peaks could not be explained by precipitation, suggesting that rainfall occurred in the WEC but was not captured by the meteorological station, a common issue in hydrology. Conversely, the models occasionally produced discharge peaks in response to heavy rainfall that did not appear in the observations, indicating that the recorded precipitation at the station missed portions of the catchment.

4.3.3. Transpiration simulations

As also done for discharge, the transpiration simulations generated by the conventional and modified models were evaluated based on performance metrics and a visual inspection of the resulting transpiration time series. Additionally, the individual contributions of the coniferous and deciduous parts of the models to the total transpiration were evaluated.

Evaluation metrics

The transpiration simulations were compared with GLDAS-2.2 data as a reference. However, this dataset was generated using multiple land surface models and does not represent direct observations. While GLDAS-2.2 provides a useful benchmark for evaluating model outputs, it is subject to uncertainties stemming from both the underlying models and their input data. Consequently, differences between the transpiration simulations and GLDAS-2.2 may reflect uncertainties in the reference dataset as well as in the hydrological models.

The SF model achieved the highest scores for N_{NSE} and N_{logNSE}, annually as well as seasonally. Annually, the conventional model performed similar to the K_v and combined models based on N_{NSE}, but performed lower for N_{logNSE}. This indicates that the conventional model was not able to simulate low transpiration rates well (winter). During the summer, the conventional model performed higher on N_{NSE} but lower on N_{logNSE} compared to the K_v and combined models.

All models performed worse in winter than in summer, with all except the SF model yielding negative N_{NSE} and N_{logNSE} values. Negative values indicate that the simulations were less accurate than simply using the mean transpiration over the given period. This outcome was driven by (near-)zero transpiration during winter: when models simulated non-zero transpiration, it strongly affected the N_{NSE} and N_{logNSE} scores.

The N_{RMSE} further confirmed that the SF model was superior to the other models for simulating transpiration. Like for N_{NSE} and N_{logNSE}, the SF model performed higher for each time period. During summer the conventional model performed slightly higher than the K_v and combined models, but much lower during the winter. This was because the conventional model produced relatively high transpiration rates during the winter, while in reality these were (near-)zero.

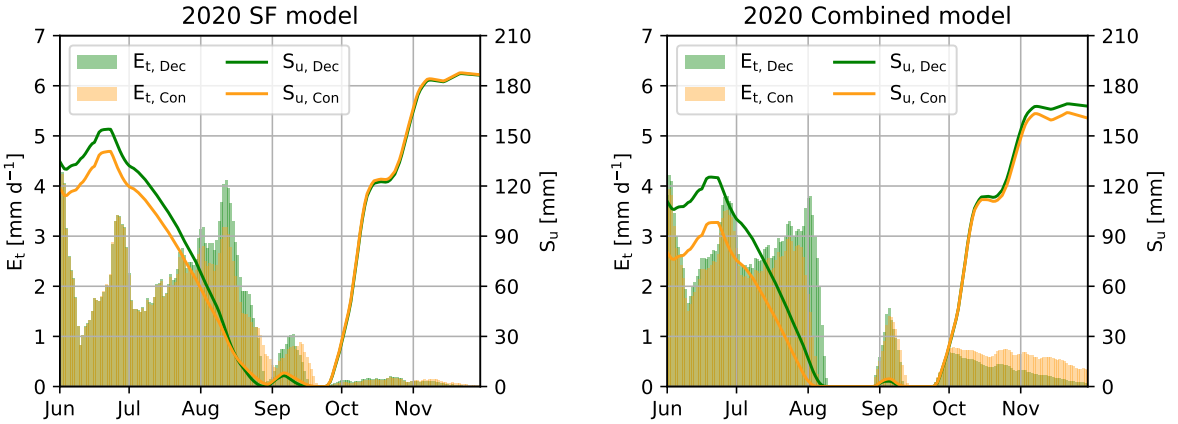
Transpiration time series

The time series analysis supported the metrics-based evaluation, indicating that transpiration was best simulated by the SF model. Its simulations generally matched the reference data, particularly between August and April. From May to July, the SF model often overestimated transpiration, but still remained closer to the reference than the other models. The SF model relied on sap flow data generated by the sap flow prediction model, which had been trained on 2019–2020 data. Importantly, its performance was not noticeably better during those years than in others, suggesting that the sap flow prediction model was generalizable.

The other modified models simulated winter transpiration more accurately than the conventional model, although they generally still overestimated, particularly in October and November. They also captured fluctuations in transpiration more effectively, whereas the conventional model showed fewer fluctuations and often failed to reproduce transpiration peaks.

During summer, the combined model typically produced the highest transpiration rates, overestimating the transpiration. During transpiration peaks in June and July, the K_v model simulated lower rates than the SF and combined models, generally falling between those models and the conventional model.

The SF and combined models showed anomalous behaviour during August and September in 2018, 2020 and 2022. Here, both models simulated transpiration values of zero while this should not happen at these moments. Analysis of the internal fluxes and storages of these models indicate that the unsaturated reservoir (UR) depleted at these moments, as shown in Figure 4.1. This issue did not occur for the conventional and K_v models. These models used Equations 2.15 and 2.31 respectively, calculating transpiration based on the relative soil moisture and potential evaporation. Here, transpiration decreased with a decreasing relative soil moisture, resulting in a slower depletion of UR. As the SF and combined models used methods not directly limited by relative soil moisture, their UR depleted faster, resulting in the zero transpiration simulations.



(a) 2020 Transpiration anomaly (7-days rolling mean) for SF model.

(b) 2020 Transpiration anomaly (7-days rolling mean) for combined model.

Figure 4.1: The 2020 transpiration anomalies found in the SF and combined models. Transpiration simulations were zero when the unsaturated reservoir (UR) was empty (S_u = 0), and returned when S_u increased. Transpiration and storages are shown for both deciduous and coniferous model parts per unit of area.

In terms of fractional contributions by the coniferous and deciduous parts of the models, the conventional model showed no seasonal differences between the coniferous and deciduous components. It also predicted transpiration during winter because E_t was calculated solely as a function of relative soil moisture and E_p. As a result, seasonality was not accounted for, leading to overestimation of E_t during winter, as clearly shown in the results.

The SF model showed a strong increase in the deciduous contribution in April, which persisted until the end of the year. This increase coincides with leaf development in spring, indicating a realistic representation of this process.

The K_v and combined model both showed a strong seasonality, with coniferous trees contributing 100% of the transpiration at the start of the year. This contribution gradually decreased as the deciduous trees became more dominant. From September onwards, the deciduous contribution decreased again. This shape strongly corresponds with the K_v shape function, which is expected since both models use this shape for the partitioning of E_p in the deciduous part of the model.

The analysis of coniferous and deciduous contributions to total transpiration showed that these were not continuous. The relative contributions on 31 December differed from those on 1 January, resulting in an artificial jump. Although transpiration was near zero at that time and thus the effect on the simulations was negligible, such discontinuities are undesirable because they are unrealistic. These jumps arise from different causes.

The jump in contribution shown by the SF model can be explained by the method used to predict normalised sap flow. The prediction model relied on $T_{CGDD,n}$ as a seasonal indicator but was trained only on data from April to mid-December, and thus did not cover the full year. As a result, it could not derive a cyclic partial dependence for $T_{CGDD,n}$, leading to a discontinuity between values of 1 and 0 (31 December to 1 January). If sap flow data were available year-round, the prediction model could establish cyclic seasonality and remove this jump.

The K_v model's jump was the result of the K_v shape function. This shape function started every year at zero but did not end at zero. This was caused by the method used to derive the shape function. A logistic function (Equation 2.25) was fitted to the T_{CGDD} curve, using a parameter t_0 to define the midpoint of the shape. The actual midpoint of the year occurs after 183 days (2 July), whereas the midpoints determined during fitting were typically around day 200 (19 July). As a result, the theoretical maximum of the T_{CGDD} curve was not reached by 31 December, leaving the slope of the K_v shape function nonzero at year-end. This creates a slight discontinuity, producing the observed jump.

The combined model used both the K_v and SF models approaches and therefore its jump has to be explained by both reasons mentioned above. However, visually the discontinuity seems to be mainly influenced by the K_v shape function (Figure 3.20c).

4.3.4. Model evaluations

The analysis of simulated discharge showed that the combined model was the highest performing model during winter and annually. The K_v model showed very similar performances as the combined model, but overall slightly lower. The SF model performed the lowest on discharge simulations, with the conventional model performing higher. During the summer, however, the conventional model was the highest performing model. Here, the combined model scored slightly lower, followed by the K_v model. Here, the SF had the lowest performance again. It should be emphasised that, although the metrics suggest some models performed better than others, no model was significantly superior or inferior; all produced good and very similar discharge simulations.

Analysis of the models' internal transpiration simulation was in sharp contrast with the discharge simulations. The SF model consistently outperformed the others, achieving higher scores across all metrics, both annually and seasonally. Its advantage was most pronounced in winter, when other models generally performed worse than using the mean (negative N_{NSE} or N_{logNSE} values), while the SF model maintained good N_{NSE} scores. This low performance by the other models during winter was explained by the overestimation of transpiration, as transpiration was typically near-zero during this time. The conventional model typically performed worst of the four models.

The sharp contrast between the discharge and transpiration simulation scores was surprising. It was expected that more accurate transpiration simulations would improve the internal water balance, leading to higher discharge performance. However, the highest performing model on transpiration (SF model) performed worst on discharge. This suggests that other internal model processes compensated for inaccuracies in transpiration estimates.

5 Conclusion

The objective of this study was to incorporate tree phenology into a conceptual hydrological model for a heterogeneous forest catchment by using environmental variables to represent sap flow dynamics. To address this, a sap flow prediction model was developed to extend limited observations, and three methods were tested to integrate phenological information into a conceptual hydrological model. It was hypothesised that incorporating tree phenology would improve discharge and transpiration simulations compared to a conventional model without vegetation dynamics.

The central questions for this research were as follows:

1. How can environmental variables representing sap flow dynamics be used to include tree phenology in a conceptual hydrological model for a catchment with heterogeneous tree coverage?
2. How can sap flow be predicted based on environmental variables?
3. How do conceptual hydrological models that include phenology information perform on discharge and transpiration simulations compared to the conventional conceptual hydrological model?

5.1. Sap flow prediction model

The sap flow prediction model was developed using a Generalized Additive Model (GAM) framework. Separate models were created for coniferous and deciduous trees, each predicting normalised sap flow ($v_{\text{sf,n}}$) for its respective category. The models used environmental variables as inputs: temperature (T), relative humidity (h), incoming shortwave radiation (R_{s}), weighted average volumetric water content over depth (θ_{avg}), and normalised accumulated growing degree-day ($T_{\text{CGDD,n}}$), with the latter serving as a seasonal indicator. The model successfully captured sap flow dynamics in both tree categories and demonstrated generalizability, achieving similar performance on training and testing datasets.

Analysis of the fitted models showed that T and R_{s} were positively correlated with $v_{\text{sf,n}}$. Higher temperatures had a slightly negative correlation with $v_{\text{sf,n}}$ for coniferous trees, indicating they experienced stress at higher temperatures. Deciduous trees did not show this behaviour, indicating they were less sensitive to higher temperatures.

Both tree categories responded similarly to h, which was negatively correlated with $v_{\text{sf,n}}$, consistent with its role in regulating atmospheric demand. Furthermore, h had a larger influence on $v_{\text{sf,n}}$ than T or R_{s}, confirming its importance in sap flow prediction.

Responses to θ_{avg} were comparable for both tree types. Low soil moisture strongly reduced $v_{\text{sf,n}}$ due to water stress, while high soil moisture limited sap flow in deciduous trees, likely due to reduced oxygen availability. Coniferous trees were less sensitive to high soil moisture.

The models showed a clear difference in seasonality between coniferous and deciduous trees. Based on the relations between $v_{\text{sf,n}}$ and $T_{\text{CGDD,n}}$, it can be concluded that deciduous trees had a strong seasonality. In contrast, coniferous trees had limited seasonality.

Overall, the sap flow prediction model provides a reliable method to extend sap flow records. Therefore, it can be concluded that $v_{\text{sf,n}}$ can be predicted using a GAM structure based on T, h, R_{s}, θ_{avg}, and $T_{\text{CGDD,n}}$.

5.2. Hydrological model

To incorporate tree phenology into the hydrological modelling framework, the $FLEX_I$ model was first adapted from a lumped into a semi-distributed structure, representing the two dominant vegetation types in the Weierbach Experimental Catchment: deciduous and coniferous trees. This model was referred to as the conventional model.

Three methods were developed to represent phenological dynamics in this model. The first method (SF model) relied directly on normalised sap flow, replacing the conventional relative soil moisture term in the transpiration equation. The second method (K_v model) used the temperature-based accumulated growing degree-day (T_{CGDD}) as an indirect measure of sap flow seasonality. It dynamically partitioned potential evaporation into interception evaporation and transpiration. The third method (combined model) combined the first two methods.

The performance on discharge simulations of the modified models was generally comparable to the conventional model. During the hydrological winter (Oct–Mar), the K_v and combined models had the highest performances based on discharge, followed by the conventional model. The SF model had the lowest performance during this period. However, during summer the conventional model performed better than the modified models. In terms of root mean square error, all models had similar performances during summer, winter and annually. Overall, all models performed well in simulating discharge, with no model being significantly superior or inferior to the others.

In contrast, the inclusion of phenological dynamics had a significant impact on transpiration simulations. The conventional model consistently overestimated transpiration during winter months, when deciduous trees were leafless and no transpiration took place. The modified models were able to correct this unrealistic behaviour, simulating (near-)zero transpiration outside the growing season of deciduous trees. During the winter, the SF model simulated coniferous trees more accurately than the K_v and combined models, which overestimated coniferous transpiration.

Among the three approaches, the SF model captured sharper seasonal transitions, such as the rapid increase in transpiration during leaf-out, while the K_v and combined models represented more gradual seasonal dynamics. During summer, the transpiration simulations of the SF model were closest to the reference data, whereas the K_v and combined models tended to overestimate transpiration at both the start and the end of the growing season. These differences reflect the contrasting strengths of direct sap flow driven versus temperature-driven approaches.

The inclusion of phenological dynamics also revealed a more realistic partitioning of transpiration between coniferous and deciduous trees. Whereas the conventional model assigned fixed contributions to each tree category, the modified models captured the dynamic shift in contributions as deciduous trees enter and exit the growing season. This demonstrates that the inclusion of phenology improves the internal realism of the model, making it better aligned with ecological processes, even if the discharge simulations remain broadly unchanged.

Overall, the results indicate that incorporating tree phenology enhances the ability of conceptual hydrological models to represent transpiration seasonality and vegetation dynamics. While this did not translate into major improvements in discharge simulation, it did improve the realism of the model. Therefore, it reduces the issue of equifinality, making the model more likely to produce realistic simulations under changing conditions.

Among the three approaches, the SF method performed best, particularly in capturing transpiration seasonality. Because this method relies on the sap flow prediction model developed in this study, it highlights the added value of directly linking environmental drivers to sap flow dynamics when incorporating tree phenology into conceptual hydrological models.

References

Allen, R. G. et al. (1998). *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*. Vol. 56. FAO, Rome, Italy. ISBN: 92-5-104219-5.

Anderegg, W. R. L., J. M. Kane and L. D. L. Anderegg (Jan. 2013). 'Consequences of widespread tree mortality triggered by drought and temperature stress'. In: *Nature Climate Change* 3.1. Publisher: Nature Publishing Group, pp. 30–36. ISSN: 1758-6798. DOI: 10.1038/nclimate1635.

Baayen, R. H. and M. Linke (2020). 'An introduction to the generalized additive model'. In: *A practical handbook of corpus linguistics*, pp. 563–591.

Bearup, L. A. et al. (June 2014). 'Hydrological effects of forest transpiration loss in bark beetle-impacted watersheds'. In: *Nature Climate Change* 4.6. Publisher: Nature Publishing Group, pp. 481–486. ISSN: 1758-6798. DOI: 10.1038/nclimate2198.

Beven, K. (Mar. 2001). 'How far can we go in distributed hydrological modelling?' In: *Hydrology and Earth System Sciences* 5.1. Publisher: Copernicus GmbH, pp. 1–12. ISSN: 1027-5606. DOI: 10.5194/hess-5-1-2001.

Blöschl, G. (Mar. 2001). 'Scaling in hydrology'. In: *Hydrological Processes* 15.4, pp. 709–711. ISSN: 0885-6087, 1099-1085. DOI: 10.1002/hyp.432.

Bouaziz, L. J. E. et al. (Mar. 2021). 'Behind the scenes of streamflow model performance'. In: *Hydrology and Earth System Sciences* 25.2. Publisher: Copernicus GmbH, pp. 1069–1095. ISSN: 1027-5606. DOI: 10.5194/hess-25-1069-2021.

Breil, M., A. Weber and J. G. Pinto (June 2023). 'The potential of an increased deciduous forest fraction to mitigate the effects of heat extremes in Europe'. In: *Biogeosciences* 20.12. Publisher: Copernicus GmbH, pp. 2237–2250. ISSN: 1726-4170. DOI: 10.5194/bg-20-2237-2023.

Coenders-Gerrits, A. M. J. et al. (Feb. 2014). 'Uncertainties in transpiration estimates'. In: *Nature* 506.7487. Publisher: Nature Publishing Group, E1–E2. ISSN: 1476-4687. DOI: 10.1038/nature12925.

Copernicus Climate Change Service, Climate Data Store (2023). *ERA5 hourly data on single levels from 1940 to present*. Accessed on 07-May-2025. DOI: 10.24381/cds.adbb2d47.

Crimmins, M. A. and T. M. Crimmins (2019). 'Does an Early Spring Indicate an Early Summer? Relationships Between Intraseasonal Growing Degree Day Thresholds'. In: *Journal of Geophysical Research: Biogeosciences* 124.8, pp. 2628–2641. ISSN: 2169-8961. DOI: 10.1029/2019JG005297.

Daley, M. J. et al. (Oct. 2007). 'Water use by eastern hemlock (Tsuga canadensis) and black birch (Betula lenta): implications of effects of the hemlock woolly adelgid'. In: *Canadian Journal of Forest Research* 37.10. Publisher: NRC Research Press, pp. 2031–2040. ISSN: 0045-5067. DOI: 10.1139/X07-045.

Dantec, C. F. et al. (Nov. 2014). 'Chilling and heat requirements for leaf unfolding in European beech and sessile oak populations at the southern limit of their distribution range'. In: *International Journal of Biometeorology* 58.9. Publisher: Springer Science and Business Media LLC, pp. 1853–1864. ISSN: 0020-7128, 1432-1254. DOI: 10.1007/s00484-014-0787-7.

Daw, A. et al. (2020). 'Physics-Guided Architecture (PGA) of Neural Networks for Quantifying Uncertainty in Lake Temperature Modeling'. In: *Proceedings of the 2020 SIAM International Conference on Data Mining (SDM)*, pp. 532–540. DOI: 10.1137/1.9781611976236.60.

Dix, M. J. and D. P. Aubrey (Mar. 2021). 'Recalibrating Best Practices, Challenges, and Limitations of Estimating Tree Transpiration Via Sap Flow'. In: *Current Forestry Reports* 7.1, pp. 31–37. ISSN: 2198-6436. DOI: 10.1007/s40725-021-00134-x.

Dunn, P. K. and G. K. Smyth (2018). *Chapter 11: Positive Continuous Data: Gamma and Inverse Gaussian GLMs*. Ed. by P. K. Dunn and G. K. Smyth. New York, NY: Springer, pp. 425–456. ISBN: 978-1-4419-0118-7. DOI: 10.1007/978-1-4419-0118-7_11.

Eilers, P. H. C. and B. D. Marx (May 1996). 'Flexible smoothing with B-splines and penalties'. In: *Statistical Science* 11.2. Publisher: Institute of Mathematical Statistics, pp. 89–121. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/ss/1038425655.

Eilers, P. H. C. and B. D. Marx (2002). 'Generalized linear additive smooth structures'. In: *Journal of Computational and Graphical Statistics* 11.4, pp. 758–783. ISSN: 1061-8600.

Esri (2022). *ACT (Luxembourg2022) [aerial imagery basemap]*. Sources: Esri, ACT, Maxar, Earthstar Geographics, and the GIS User Community. Captured 10 August 2022. Retrieved via QuickMapServices plugin in QGIS (2025).

Fabiani, G. (Dec. 2022). 'Tree water use across landscape and time'. PhD thesis. Ginevra Fabiani, Belvaux, Luxembourg.

Fabiani, G. et al. (2022). 'Transpiration patterns and water use strategies of beech and oak trees along a hillslope'. In: *Ecohydrology* 15.2, e2382. ISSN: 1936-0592. DOI: 10.1002/eco.2382.

Feddes, R. A. et al. (Sept. 1976). 'Simulation of field water uptake by plants using a soil water dependent root extraction function'. In: *Journal of Hydrology* 31.1, pp. 13–26. ISSN: 0022-1694. DOI: 10.1016/0022-1694(76)90017-2.

Fenicia, F. et al. (Feb. 2006). 'Is the groundwater reservoir linear? Learning from data in hydrological modelling'. In: *Hydrology and Earth System Sciences* 10.1. Publisher: Copernicus GmbH, pp. 139–150. ISSN: 1027-5606. DOI: 10.5194/hess-10-139-2006.

Fenicia, F. et al. (2008). 'Understanding catchment behavior through stepwise model concept improvement'. In: *Water Resources Research* 44.1. ISSN: 1944-7973. DOI: 10.1029/2006WR005563.

Flo, V. et al. (July 2022). 'Vapour pressure deficit is the main driver of tree canopy conductance across biomes'. In: *Agricultural and Forest Meteorology* 322, p. 109029. ISSN: 0168-1923. DOI: 10.1016/j.agrformet.2022.109029.

Fu, Y. H. et al. (2019). 'Daylength helps temperate deciduous trees to leaf-out at the optimal time'. In: *Global Change Biology* 25.7, pp. 2410–2418. ISSN: 1365-2486. DOI: 10.1111/gcb.14633.

Gao, H. et al. (May 2014). 'Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China'. In: *Hydrology and Earth System Sciences* 18.5. Publisher: Copernicus GmbH, pp. 1895–1915. ISSN: 1027-5606. DOI: 10.5194/hess-18-1895-2014.

Gimenez, C., M. Gallardo and R. Thompson (2005). 'PLANT–WATER RELATIONS'. In: *Encyclopedia of Soils in the Environment*. Ed. by D. Hillel. Oxford: Elsevier, pp. 231–238. ISBN: 978-0-12-348530-4. DOI: https://doi.org/10.1016/B0-12-348530-4/00459-8.

Good, S. P., D. Noone and G. Bowen (July 2015). 'Hydrologic connectivity constrains partitioning of global terrestrial water fluxes'. In: *Science* 349.6244. Publisher: American Association for the Advancement of Science, pp. 175–177. DOI: 10.1126/science.aaa5931.

Hackmann, C. A. et al. (June 2025). 'Root water uptake depth in temperate forest trees: species-specific patterns shaped by neighbourhood and environment'. In: *Plant Biology*. ISSN: 1438-8677. DOI: 10.1111/plb.70058.

Hamon, W. R. (1961). 'Estimating potential evapotranspiration'. In: *Journal of the Hydraulics Division* 87.3, pp. 107–120.

Hamon, W. R. (1963). *Computation of Direct Runoff Amounts From Storm Rainfall*. International Association of Scientific Hydrology Publication 63, pp. 52–62.

Hastie, T. and R. Tibshirani (1986). 'Generalized Additive Models'. In: *Statistical Science* 1.3. Publisher: Institute of Mathematical Statistics, pp. 297–310. ISSN: 0883-4237.

Hersbach, H. et al. (2023). *ERA5 hourly data on single levels from 1940 to present*. Accessed on 07-May-2025. DOI: 10.24381/cds.adbb2d47.

Hissler, C. et al. (2021). 'The Weierbach experimental catchment in Luxembourg: A decade of critical zone monitoring in a temperate forest - from hydrological investigations to ecohydrological perspectives'. In: *Hydrological Processes* 35.5, e14140. DOI: https://doi.org/10.1002/hyp.14140.

Hodges, T. and J. Ritchie (1991). 'The CERES-Wheat phenology model'. In: *Predicting crop phenology*, pp. 133–141.

Hrachowitz, M. and M. P. Clark (Aug. 2017). 'HESS Opinions: The complementary merits of competing modelling philosophies in hydrology'. In: *Hydrology and Earth System Sciences* 21.8. Publisher: Copernicus GmbH, pp. 3953–3973. ISSN: 1027-5606. DOI: 10.5194/hess-21-3953-2017.

Jaime, L., E. Batllori and F. Lloret (Feb. 2024). 'Bark beetle outbreaks in coniferous forests: a review of climate change effects'. In: *European Journal of Forest Research* 143.1, pp. 1–17. ISSN: 1612-4677. DOI: 10.1007/s10342-023-01623-3.

Jarvis, P. G. (1976). 'The Interpretation of the Variations in Leaf Water Potential and Stomatal Conductance Found in Canopies in the Field'. In: *Philosophical Transactions of the Royal Society of*

London. Series B, Biological Sciences 273.927. Publisher: The Royal Society, pp. 593–610. ISSN: 0080-4622.

Jarvis, P. G. and K. G. McNaughton (1986). 'Stomatal Control of Transpiration: Scaling Up from Leaf to Region'. In: ed. by A. MacFadyen and E. D. Ford. Vol. 15. Advances in Ecological Research. ISSN: 0065-2504. Academic Press, pp. 1–49. DOI: https://doi.org/10.1016/S0065-2504(08)60119-1.

Jayathilake, D. I. and T. Smith (July 2020). 'Understanding the role of hydrologic model structures on evapotranspiration-driven sensitivity'. In: *Hydrological Sciences Journal* 65.9. Publisher: Taylor & Francis, pp. 1474–1489. ISSN: 0262-6667. DOI: 10.1080/02626667.2020.1754421.

Jiménez-Rodríguez, C. D. et al. (2024). 'Detecting Vegetation Stress in Mixed Forest Ecosystems Through the Joint Use of Tree-Water Monitoring and Land Surface Modeling'. In: *Water Resources Research* 60.8, e2024WR037194. ISSN: 1944-7973. DOI: 10.1029/2024WR037194.

Juice, S. M. et al. (2016). 'Ecosystem warming increases sap flow rates of northern red oak trees'. In: *Ecosphere* 7.3, e01221. ISSN: 2150-8925. DOI: 10.1002/ecs2.1221.

Kim, K. B., H.-H. Kwon and D. Han (Jan. 2018). 'Exploration of warm-up period in conceptual hydrological modelling'. In: *Journal of Hydrology* 556, pp. 194–210. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2017.11.015.

Klaus, J. et al. (May 2024). *Detailed sap flow monitoring data at Weierbach catchment, Luxembourg*. Version v01. DOI: 10.5281/zenodo.11381618.

Kool, D. et al. (Jan. 2014). 'A review of approaches for evapotranspiration partitioning'. In: *Agricultural and Forest Meteorology* 184, pp. 56–70. ISSN: 0168-1923. DOI: 10.1016/j.agrformet.2013.09.003.

Koutsoyiannis, D. (Jan. 2012). 'Clausius–Clapeyron equation and saturation vapour pressure: simple theory reconciled with practice'. In: *European Journal of Physics* 33.2. Publisher: IOP Publishing, p. 295. ISSN: 0143-0807. DOI: 10.1088/0143-0807/33/2/295.

Kubota, C. (2016). 'Growth, Development, Transpiration and Translocation as Affected by Abiotic Environmental Factors'. In: *Plant Factory*. Elsevier, pp. 151–164. ISBN: 978-0-12-801775-3. DOI: 10.1016/b978-0-12-801775-3.00010-x.

Köstner, B. M. M. et al. (Sept. 1992). 'Transpiration and canopy conductance in a pristine broad-leaved forest of Nothofagus: an analysis of xylem sap flow and eddy correlation measurements'. In: *Oecologia* 91.3, pp. 350–359. ISSN: 1432-1939. DOI: 10.1007/BF00317623.

l'Agriculture, A. des Services Techniques de (2025). *Wetterdaten – Agrarmeteorologie*. https://www.agrimeteo.lu/Agrarmeteorologie/Wetterdaten/Download-Grafik. Accessed: 2025-03-05.

Li, B. et al. (2019). 'Global GRACE Data Assimilation for Groundwater and Drought Monitoring: Advances and Challenges'. In: *Water Resources Research* 55.9, pp. 7564–7586. ISSN: 1944-7973. DOI: 10.1029/2018WR024618.

Li, B. et al. (2020). *GLDAS Catchment Land Surface Model L4 daily 0.25 x 0.25 degree GRACE-DA1, Version 2.2*. DOI: 10.5067/TXBMLX370XX8.

Liu, Z. et al. (2019). 'Conceptual Hydrological Models'. In: *Handbook of Hydrometeorological Ensemble Forecasting*. Springer, Berlin, Heidelberg, pp. 389–411. ISBN: 978-3-642-39925-1. DOI: 10.1007/978-3-642-39925-1_22.

Lopushinsky, W. (Aug. 1986). 'Seasonal and diurnal trends of heat pulse velocity in Douglas-fir and ponderosa pine'. In: *Canadian Journal of Forest Research* 16.4. Publisher: NRC Research Press, pp. 814–821. ISSN: 0045-5067. DOI: 10.1139/x86-144.

Loritz, R. et al. (Sept. 2022). 'Leveraging sap flow data in a catchment-scale hybrid model to improve soil moisture and transpiration estimates'. In: *Hydrology and Earth System Sciences* 26.18. Publisher: Copernicus GmbH, pp. 4757–4771. ISSN: 1027-5606. DOI: 10.5194/hess-26-4757-2022.

Luo, Y. et al. (Aug. 2022). 'Evergreen broadleaf greenness and its relationship with leaf flushing, aging, and water fluxes'. In: *Agricultural and Forest Meteorology* 323, p. 109060. ISSN: 0168-1923. DOI: 10.1016/j.agrformet.2022.109060.

Martínez-Carreras, N. et al. (Dec. 2016). 'Storage controls on the generation of double peak hydrographs in a forested headwater catchment'. In: *Journal of Hydrology* 543, pp. 255–269. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2016.10.004.

McMaster, G. S. and W. W. Wilhelm (Dec. 1997). 'Growing degree-days: one equation, two interpretations'. In: *Agricultural and Forest Meteorology* 87.4, pp. 291–300. ISSN: 0168-1923. DOI: 10.1016/S0168-1923(97)00027-0.

Mert, O. (2021). 'The Impact of Vegetation on the Partitioning of Evaporation in Conceptual Hydrological Models'. MA thesis. TU Delft.

Monteith, J. L. (1965). 'Evaporation and environment'. In: *Symposia of the society for experimental biology*. Vol. 19. Cambridge University Press (CUP) Cambridge, pp. 205–234.

Nash, J. E. and J. V. Sutcliffe (Apr. 1970). 'River flow forecasting through conceptual models part I — A discussion of principles'. In: *Journal of Hydrology* 10.3, pp. 282–290. ISSN: 0022-1694. DOI: 10.1016/0022-1694(70)90255-6.

Nelder, J. A. and R. W. M. Wedderburn (1972). 'Generalized Linear Models'. In: *Journal of the Royal Statistical Society. Series A (General)* 135.3. Publisher: [Royal Statistical Society, Oxford University Press], pp. 370–384. ISSN: 0035-9238. DOI: 10.2307/2344614.

Netherer, S. et al. (June 2021). 'Interactions among Norway spruce, the bark beetle Ips typographus and its fungal symbionts in times of drought'. In: *Journal of Pest Science* 94.3, pp. 591–614. ISSN: 1612-4766. DOI: 10.1007/s10340-021-01341-y.

Nijzink, J. et al. (May 2025). 'CAMELS-LUX: Highly Resolved Hydro-Meteorological and Atmospheric Data for Physiographically Characterized Catchments around Luxembourg'. In: *Earth System Science Data Discussions*. Publisher: Copernicus GmbH, pp. 1–34. DOI: 10.5194/essd-2024-482.

Ohana-Levi, N. et al. (Feb. 2020). 'Evaluation of within-season grapevine evapotranspiration patterns and drivers using generalized additive models'. In: *Agricultural Water Management* 228. ISSN: 0378-3774. DOI: 10.1016/j.agwat.2019.105808.

Oudin, L. et al. (Mar. 2005). 'Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling'. In: *Journal of Hydrology* 303.1, pp. 290–306. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2004.08.026.

Page, J. (Jan. 2012). 'Chapter IIA-1 - The Role of Solar-Radiation Climatology in the Design of Photovoltaic Systems'. In: *Practical Handbook of Photovoltaics (Second Edition)*. Ed. by A. McEvoy, T. Markvart and L. Castañer. Boston: Academic Press, pp. 573–643. ISBN: 978-0-12-385934-1. DOI: 10.1016/B978-0-12-385934-1.00017-9.

Pfister, L. et al. (2017). 'Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments'. In: *Hydrological Processes* 31.10, pp. 1828–1845. ISSN: 1099-1085. DOI: 10.1002/hyp.11134.

Pierik, A. C. (2022). 'The impact of an additional phenology model on the performance of conceptual hydrological models'. MA thesis. TU Delft.

Polgar, C. A. and R. B. Primack (2011). 'Leaf-out phenology of temperate woody plants: from trees to ecosystems'. In: *New Phytologist* 191.4, pp. 926–941. ISSN: 1469-8137. DOI: 10.1111/j.1469-8137.2011.03803.x.

Poyatos, R. et al. (June 2021). 'Global transpiration data from sap flow measurements: the SAPFLUXNET database'. In: *Earth System Science Data* 13.6. Publisher: Copernicus GmbH, pp. 2607–2649. ISSN: 1866-3508. DOI: 10.5194/essd-13-2607-2021.

Rahmat, A. et al. (2018). 'Runoff Characteristcs of Forest Watersheds Under Different Forest Managements'. In: *Reviews in Agricultural Science* 6, pp. 119–133. DOI: 10.7831/ras.6.119.

Renaud, V. et al. (Aug. 2011). 'Comparison between open-site and below-canopy climatic conditions in Switzerland for different types of forests over 10 years (1998−2007)'. In: *Theoretical and Applied Climatology* 105.1, pp. 119–127. ISSN: 1434-4483. DOI: 10.1007/s00704-010-0361-0.

Ritter, A. and R. Muñoz-Carpena (Feb. 2013). 'Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments'. In: *Journal of Hydrology* 480, pp. 33–45. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2012.12.004.

Schoppach, R. et al. (Sept. 2021). 'Species-specific control of DBH and landscape characteristics on tree-to-tree variability of sap velocity'. In: *Agricultural and Forest Meteorology* 307, p. 108533. ISSN: 0168-1923. DOI: 10.1016/j.agrformet.2021.108533.

Servén, D. and C. Brummitt (Mar. 2018). *pyGAM: Generalized Additive Models in Python*. DOI: 10.5281/zenodo.1208723. URL: https://doi.org/10.5281/zenodo.1208723.

Shao, W. et al. (2023). 'A modified Jarvis model to improve the expressing of stomatal response in a beech forest'. In: *Hydrological Processes* 37.8, e14955. ISSN: 1099-1085. DOI: 10.1002/hyp.14955.

Tanguy, M. et al. (June 2018). 'Historical gridded reconstruction of potential evapotranspiration for the UK'. In: *Earth System Science Data* 10.2. Publisher: Copernicus GmbH, pp. 951–968. ISSN: 1866-3508. DOI: 10.5194/essd-10-951-2018.

Tian, Q. et al. (Jan. 2018). ‘Growing Season Stem Water Status Assessment of Qinghai Spruce through the Sap Flow and Stem Radial Variations in the Qilian Mountains of China’. In: *Forests* 9.1. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 2. ISSN: 1999-4907. DOI: 10.3390/f9010002.

Wang, H. et al. (Aug. 2020). ‘Improving the Jarvis-type model with modified temperature and radiation functions for sap flow simulations’. In: *Journal of Hydrology* 587, p. 124981. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2020.124981.

Wei, Z. et al. (2017). ‘Revisiting the contribution of transpiration to global terrestrial evapotranspiration’. In: *Geophysical Research Letters* 44.6, pp. 2792–2801. ISSN: 1944-8007. DOI: 10.1002/2016GL072235.

Wiederhold, P. R. (Jan. 2012). *Water Vapor Measurement: Methods and Instrumentation*. Boca Raton: CRC Press. 384 pp. ISBN: 978-0-429-09824-6. DOI: 10.1201/9781466551978.

Wood, S. N. (Feb. 2006). *Generalized Additive Models: An Introduction with R*. New York: Chapman and Hall/CRC. 416 pp. ISBN: 978-0-429-09315-9. DOI: 10.1201/9781420010404.

Wood, S. N. (Jan. 2011). ‘Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models’. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73.1, pp. 3–36. ISSN: 1369-7412. DOI: 10.1111/j.1467-9868.2010.00749.x.

Zhao, L. et al. (Apr. 2013). ‘Evapotranspiration estimation methods in hydrological models’. In: *Journal of Geographical Sciences* 23.2, pp. 359–369. ISSN: 1861-9568. DOI: 10.1007/s11442-013-1015-9.

Zhiyanski, M. (2014). ‘Seasonal dynamics of fine root biomass in selected forest stands.’ In: *Silva Balcanica* 15 (2), pp. 5–15. ISSN: 1311-8706.

A Data availability

Table A.1: Sap flow data availability is shown for 2019 and 2020. The ”valid” columns report the percentage of available data for each tree over its recording period. Percentages indicating low availability are colour-coded: orange (50–80% available) and red (<50% available).

Tree	Start 2019	End 2019	Valid 2019	Start 2020	End 2020	Valid 2020
B1	2019-04-05	2019-12-19	82.0 %	2020-05-05	2020-11-05	96.0 %
B2	2019-04-05	2019-12-19	62.1%	2020-05-05	2020-11-05	90.7 %
B3	2019-04-05	2019-12-19	87.2 %	2020-05-05	2020-11-05	88.2 %
B4	2019-04-05	2019-12-18	82.8 %	2020-05-05	2020-11-05	94.0 %
B5	2019-04-05	2019-12-19	84.9 %	2020-05-05	2020-11-02	88.1 %
B6	2019-04-05	2019-12-19	61.1%	2020-05-05	2020-11-04	93.0 %
B7	2019-04-05	2019-12-19	87.9 %	2020-05-05	2020-11-05	90.1 %
B8	2019-04-05	2019-11-17	77.7%	2020-05-05	2020-11-05	93.3 %
B9	2019-04-05	2019-12-19	81.9 %	2020-05-05	2020-11-05	96.3 %
B10	2019-04-05	2019-12-19	66.8%	2020-05-05	2020-11-05	95.5 %
B11	2019-04-05	2019-12-19	61.7%	-	-	-
B12	2019-04-05	2019-12-19	87.4 %	2020-05-06	2020-11-05	96.6 %
B13	2019-05-03	2019-12-19	60.9%	2020-05-06	2020-11-01	87.0 %
B14	2019-05-03	2019-12-19	84.4 %	-	-	-
O1	2019-04-05	2019-11-07	82.0 %	2020-05-05	2020-11-05	86.8 %
O2	2019-04-05	2019-12-19	87.7 %	2020-05-05	2020-11-05	97.3 %
O3	2019-04-05	2019-12-19	62.4%	2020-05-05	2020-11-05	99.1 %
O4	2019-04-05	2019-12-19	74.8%	2020-05-05	2020-11-05	99.6 %
O5	2019-04-05	2019-12-19	70.6%	2020-05-05	2020-11-04	91.9 %
O6	2019-04-19	2019-10-15	51.3%	2020-05-05	2020-11-05	99.6 %
O7	2019-04-05	2019-11-21	46.1%	2020-05-05	2020-11-05	96.3 %
O8	2019-04-05	2019-12-07	69.4%	2020-05-05	2020-11-05	98.6 %
O9	2019-04-05	2019-12-19	81.2 %	2020-05-05	2020-11-05	93.3 %
O10	2019-04-05	2019-12-19	82.2 %	2020-05-05	2020-11-05	99.6 %
O11	2019-04-05	2019-12-19	87.3 %	2020-05-06	2020-11-05	95.7 %
O12	2019-04-05	2019-12-19	86.8 %	-	-	-
O14	2019-05-03	2019-12-19	89.0 %	-	-	-
S1	2019-04-05	2019-12-19	94.4 %	2020-05-06	2020-11-03	77.0%
S2	2019-04-05	2019-12-19	55.0%	-	-	-
S3	2019-04-05	2019-12-19	67.0%	2020-05-06	2020-11-05	99.6 %
S4	2019-07-09	2019-12-19	72.3%	2020-05-05	2020-11-05	97.9 %
S5	2019-05-03	2019-12-19	85.5 %	2020-05-05	2020-11-05	99.6 %
S6	2019-05-03	2019-12-19	71.8%	2020-05-05	2020-11-05	98.4 %
D1	2019-07-09	2019-12-19	67.2%	2020-05-06	2020-11-05	99.4 %
D2	2019-05-03	2019-12-19	71.9%	-	-	-
D3	2019-05-03	2019-12-19	94.4 %	2020-05-06	2020-11-04	81.9 %

B
Pearson correlation coefficients

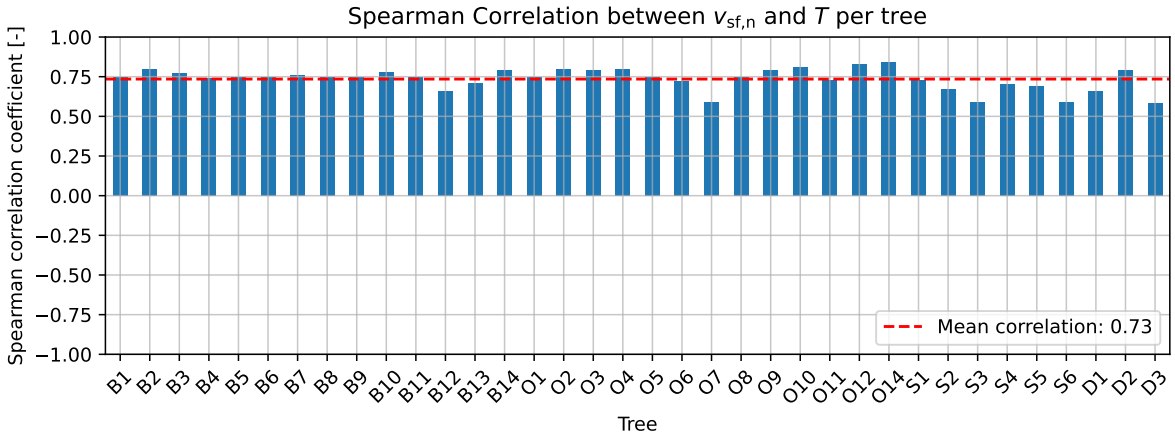


Figure B.1: Spearman correlation coefficient between normalised sap flow ($v_{sf,n}$) and temperature (T) for each tree.

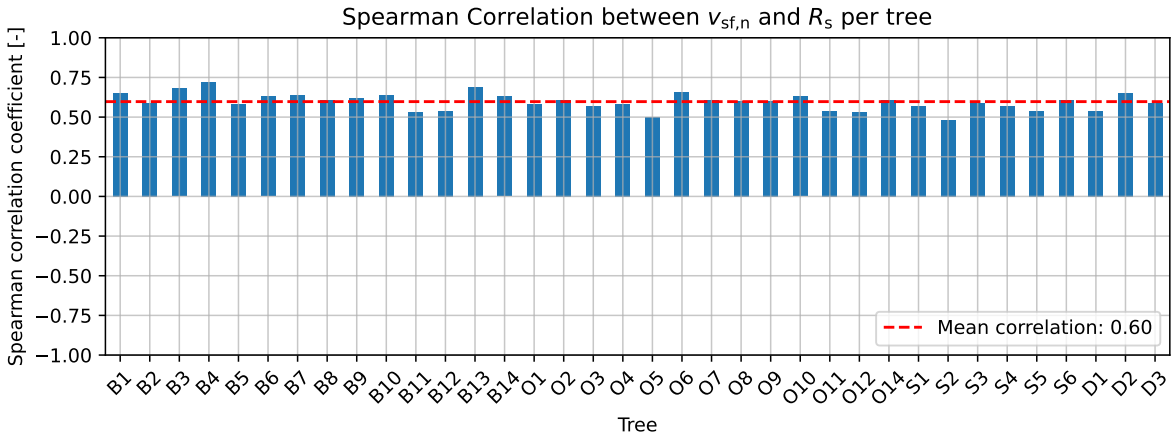


Figure B.2: Spearman correlation coefficient between normalised sap flow ($v_{sf,n}$) and incoming shortwave radiation (R_s) for each tree.

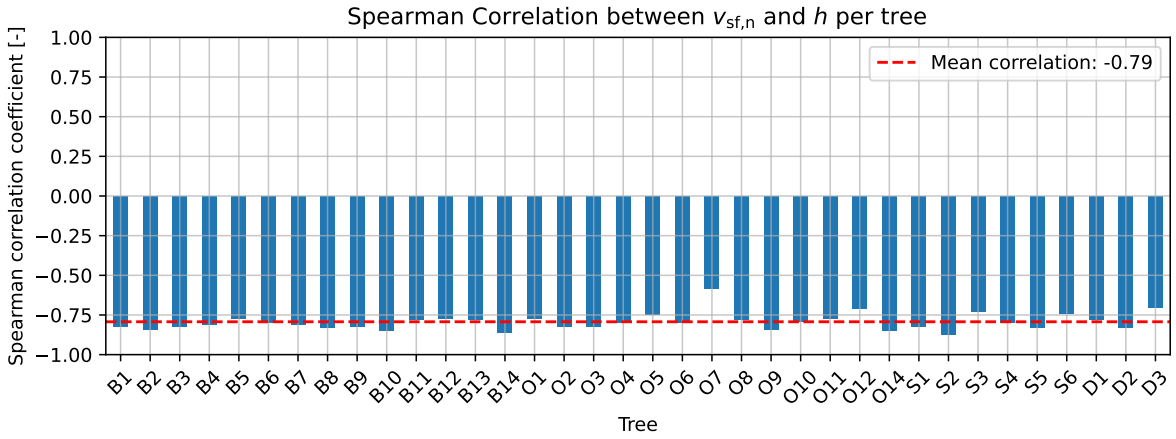


Figure B.3: Spearman correlation coefficient between normalised sap flow ($v_{sf,n}$) and relative humidity (h) for each tree.

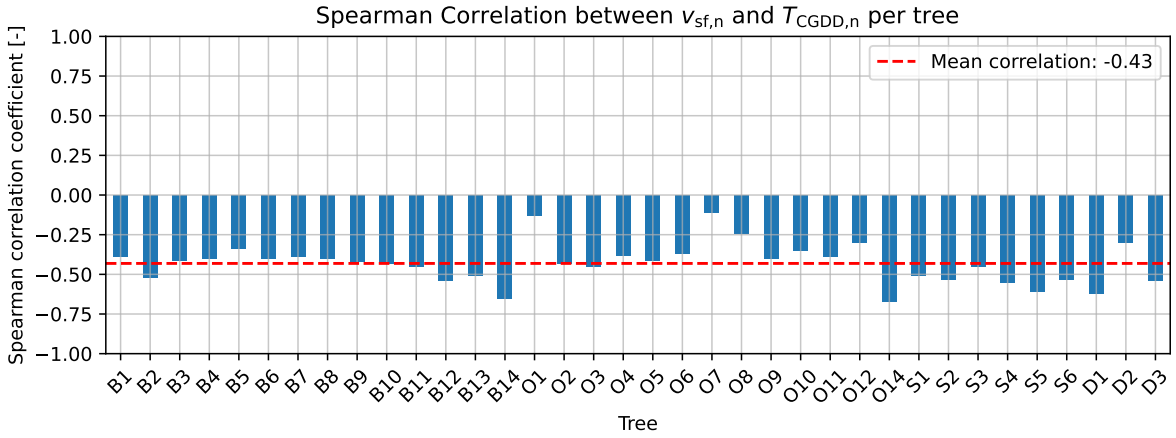


Figure B.4: Spearman correlation coefficient between normalised sap flow ($v_{sf,n}$) and normalised accumulated Growing Degree-Days (T_{CGDD}) for each tree.

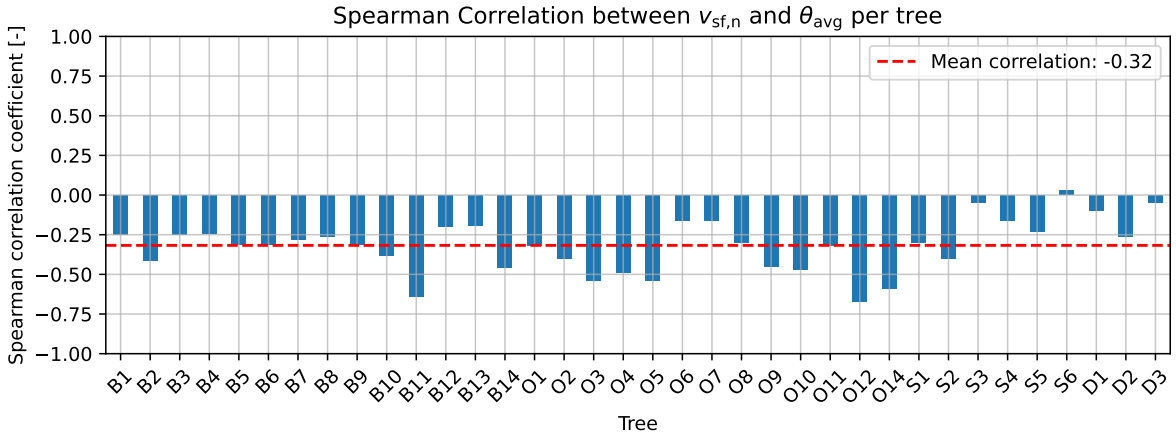


Figure B.5: Spearman correlation coefficient between normalised sap flow ($v_{sf,n}$) and weighted average of volumetric water content (θ_{avg}) for each tree.

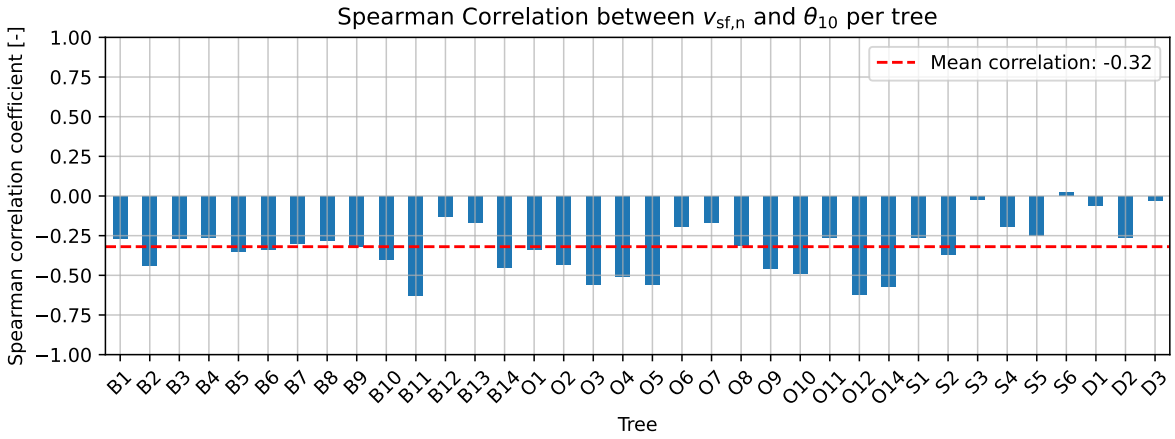


Figure B.6: Spearman correlation coefficient between normalised sap flow ($v_{sf,n}$) and volumetric water content at 10 cm depth (θ_{10}) for each tree.

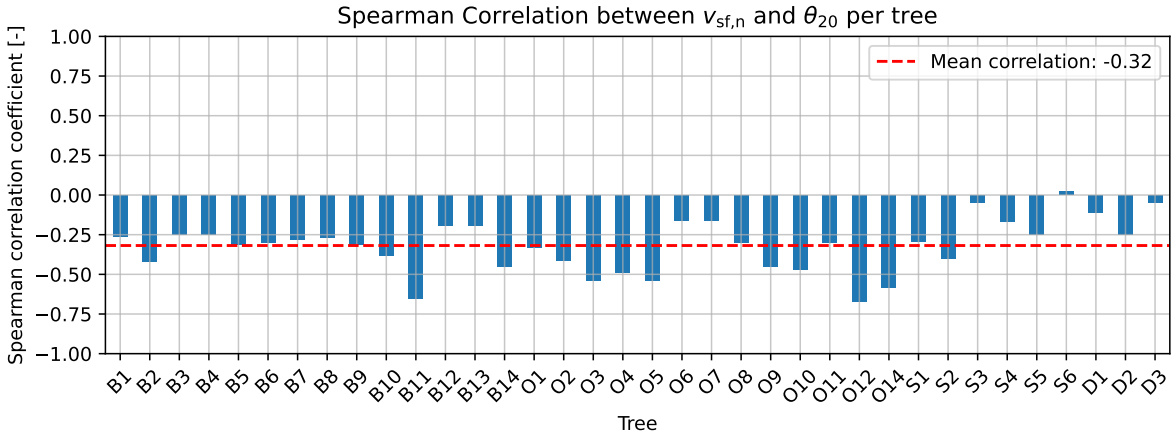


Figure B.7: Spearman correlation coefficient between normalised sap flow ($v_{sf,n}$) and volumetric water content at 20 cm depth (θ_{20}) for each tree.

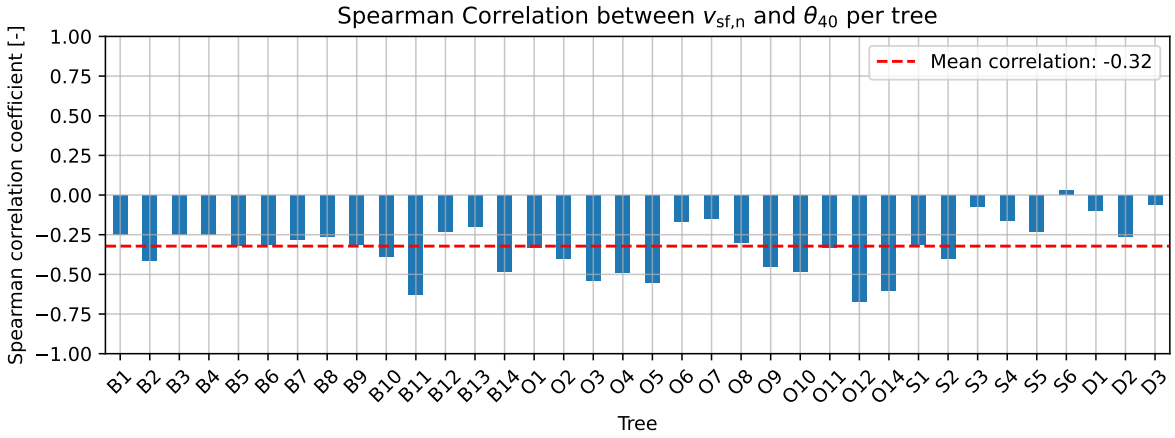


Figure B.8: Spearman correlation coefficient between normalised sap flow ($v_{sf,n}$) and volumetric water content at 40 cm depth (θ_{40}) for each tree.

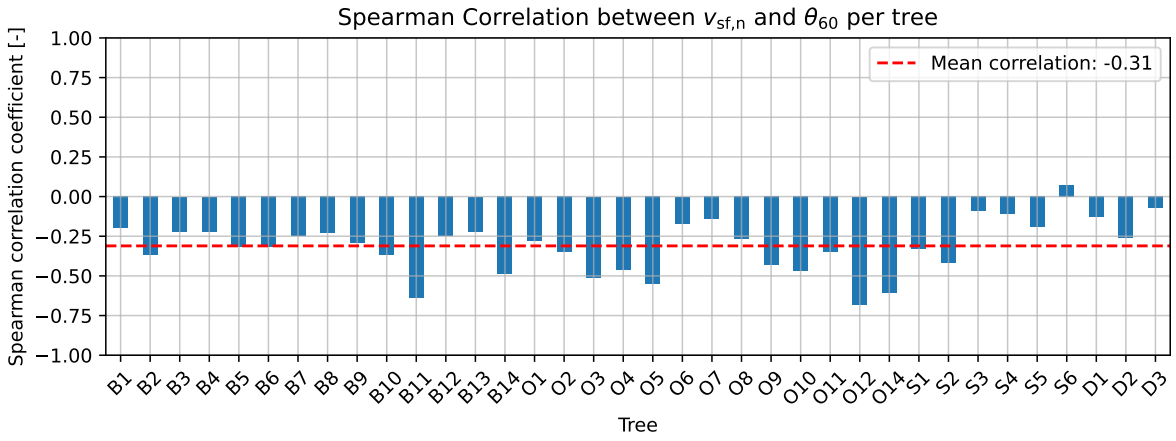


Figure B.9: Spearman correlation coefficient between normalised sap flow ($v_{sf,n}$) and volumetric water content at 60 cm depth (θ_{60}) for each tree.

C
Normalised sap flow time series

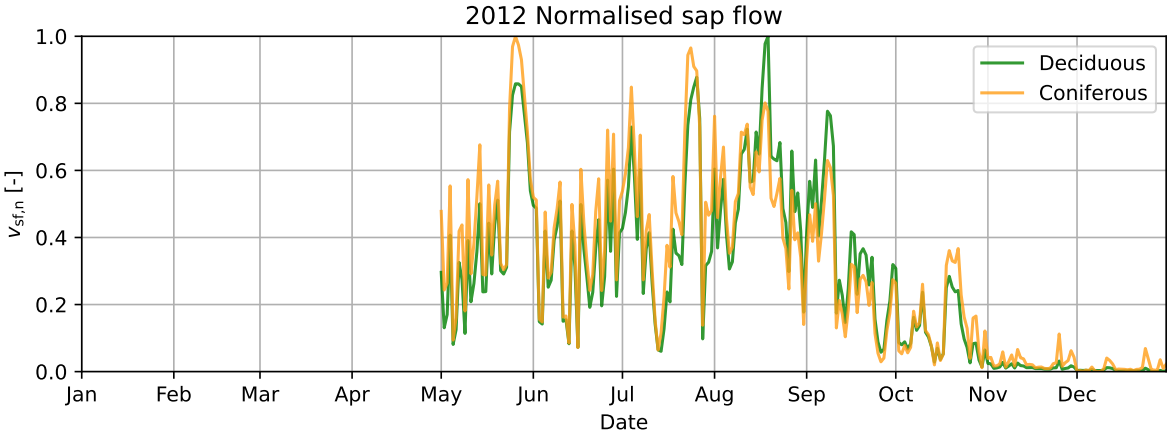


Figure C.1: Time series of the normalised sap flow ($v_{sf,n}$) for 2012. Note: No predictions between January and April were possible due to lack of data.

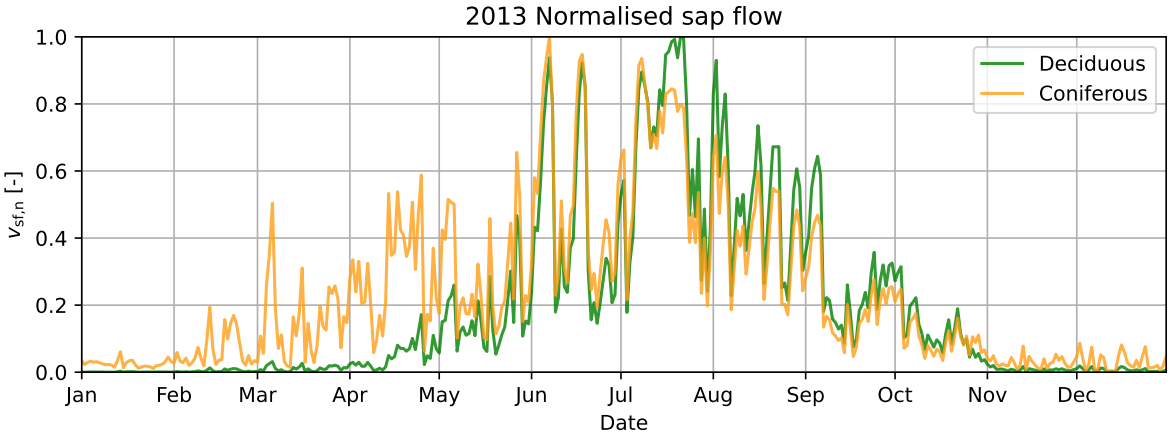


Figure C.2: Time series of the normalised sap flow ($v_{sf,n}$) for 2013.

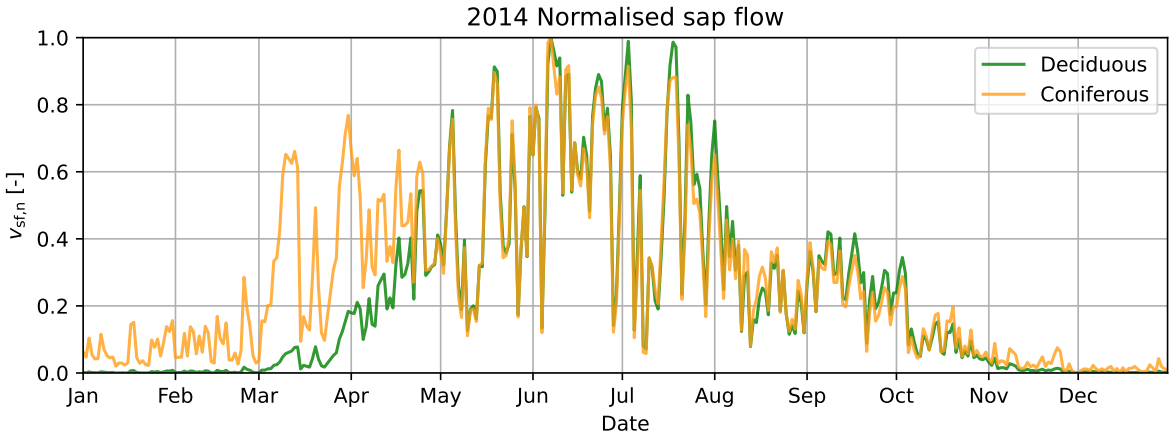


Figure C.3: Time series of the normalised sap flow ($v_{sf,n}$) for 2014.

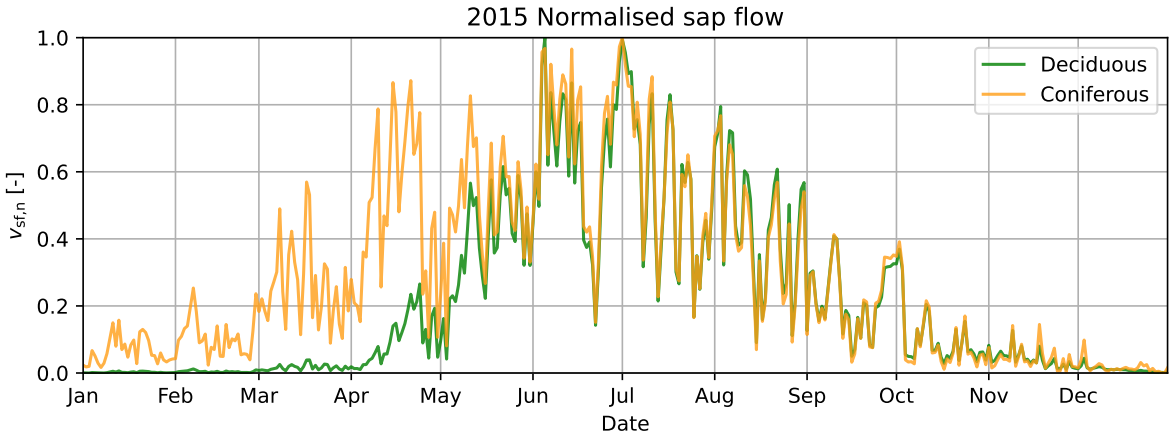


Figure C.4: Time series of the normalised sap flow ($v_{sf,n}$) for 2015.

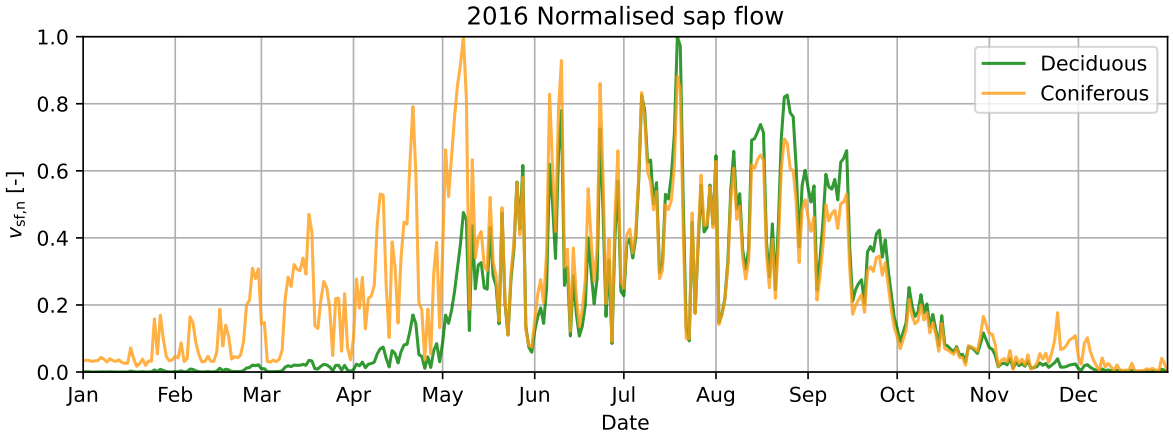


Figure C.5: Time series of the normalised sap flow ($v_{sf,n}$) for 2016.

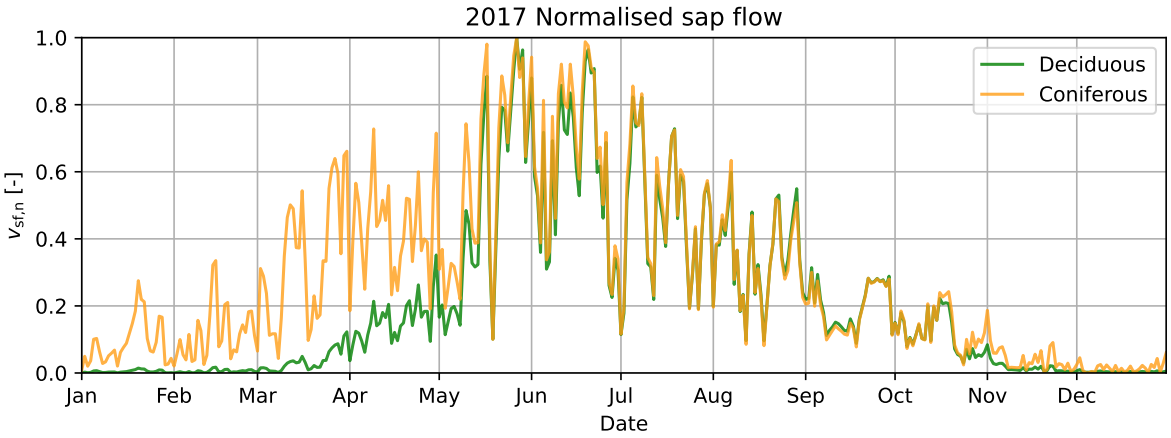


Figure C.6: Time series of the normalised sap flow ($v_{sf,n}$) for 2017.

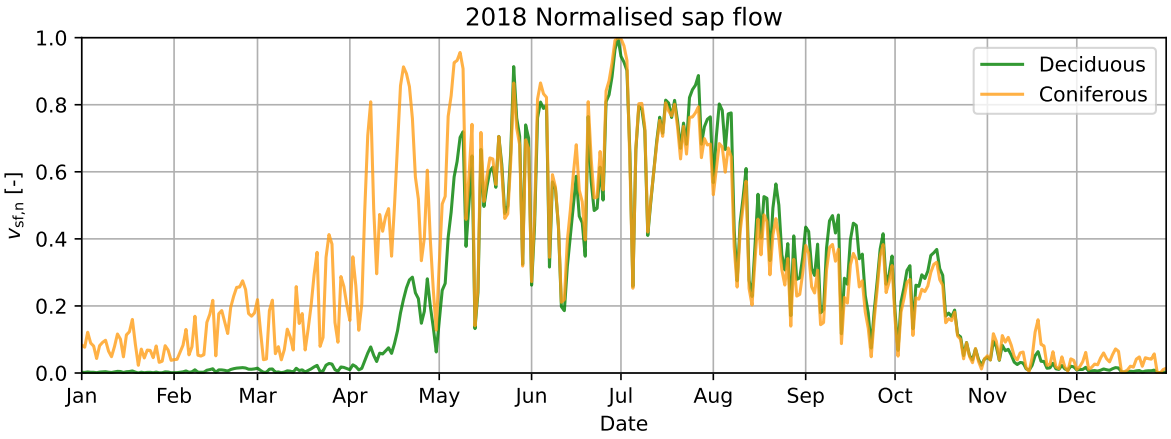


Figure C.7: Time series of the normalised sap flow ($v_{sf,n}$) for 2018.

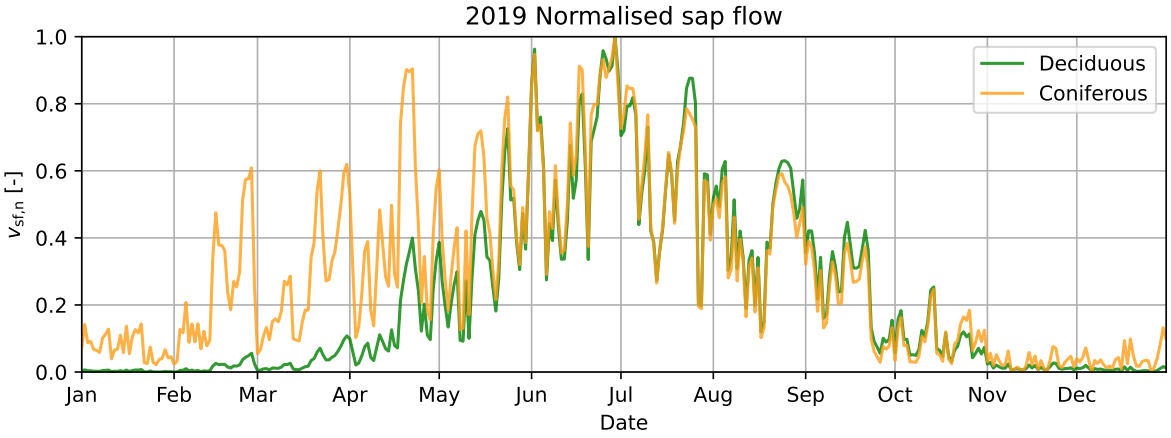


Figure C.8: Time series of the normalised sap flow ($v_{sf,n}$) for 2019.

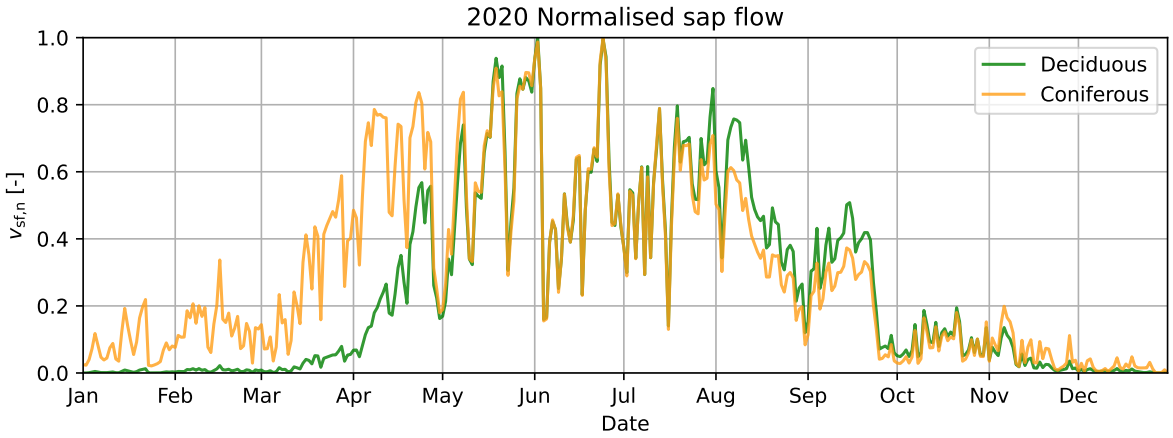


Figure C.9: Time series of the normalised sap flow ($v_{sf,n}$) for 2020.

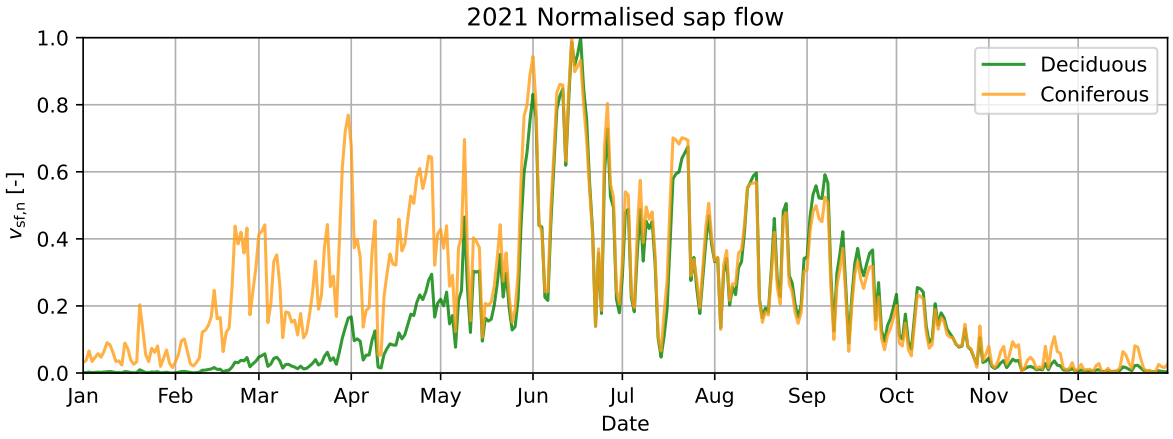


Figure C.10: Time series of the normalised sap flow ($v_{sf,n}$) for 2021.

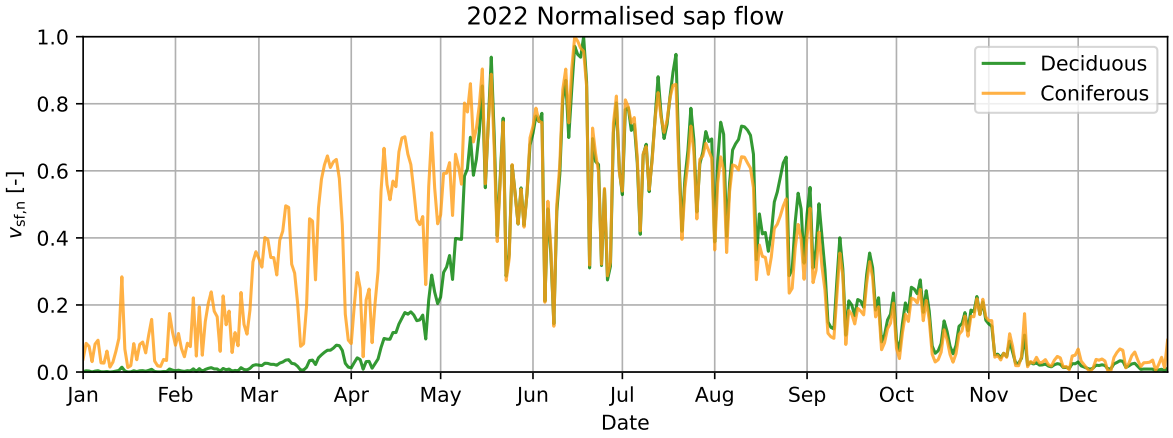
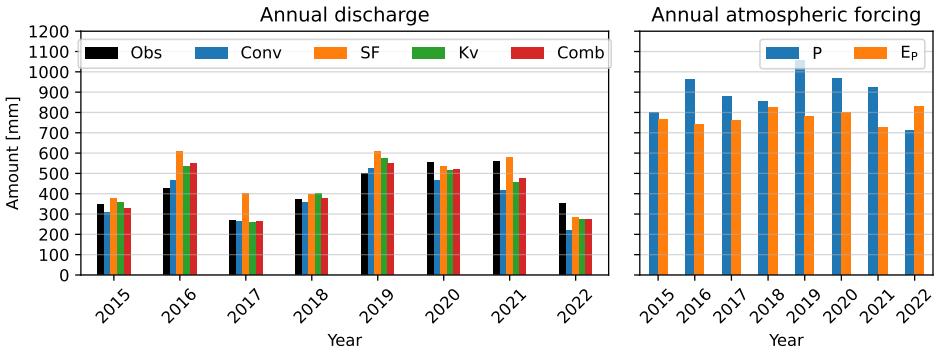
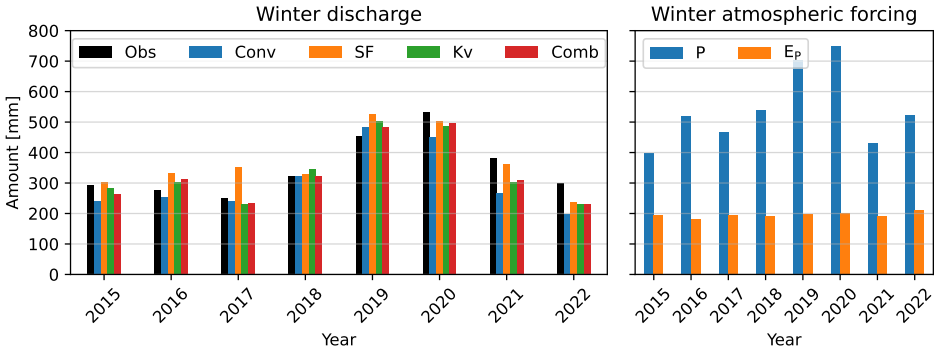


Figure C.11: Time series of the normalised sap flow ($v_{sf,n}$) for 2022.

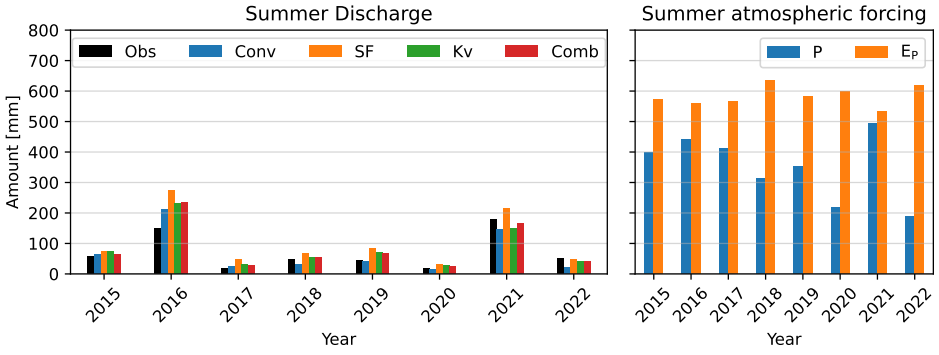
D
Discharge figures



(a) Annual (Jan–Dec).



(b) Winter (Oct–Mar).



(c) Summer (Apr–Sep).

Figure D.1: Overview of total precipitation (P) and potential evaporation (E_p) with total observed and simulated discharge per model. Shown annually (Jan–Dec) at the top, during hydrological winter (Oct–Mar) in the middle and during the summer (Apr–Sep) at the bottom.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

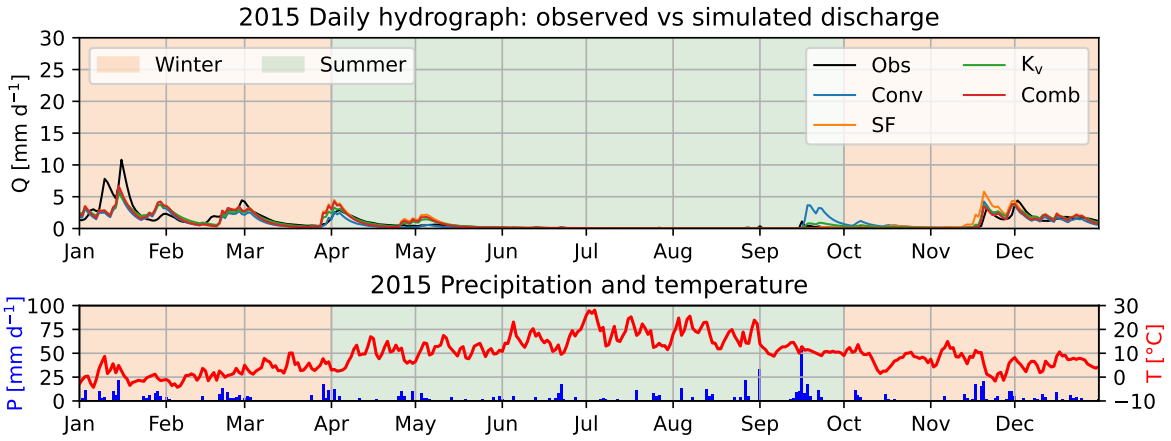


Figure D.2: Hydrograph of the observed and simulated discharges per model in the upper panel for 2015. The bottom panel shows the precipitation (P) and temperature (T) over this period. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

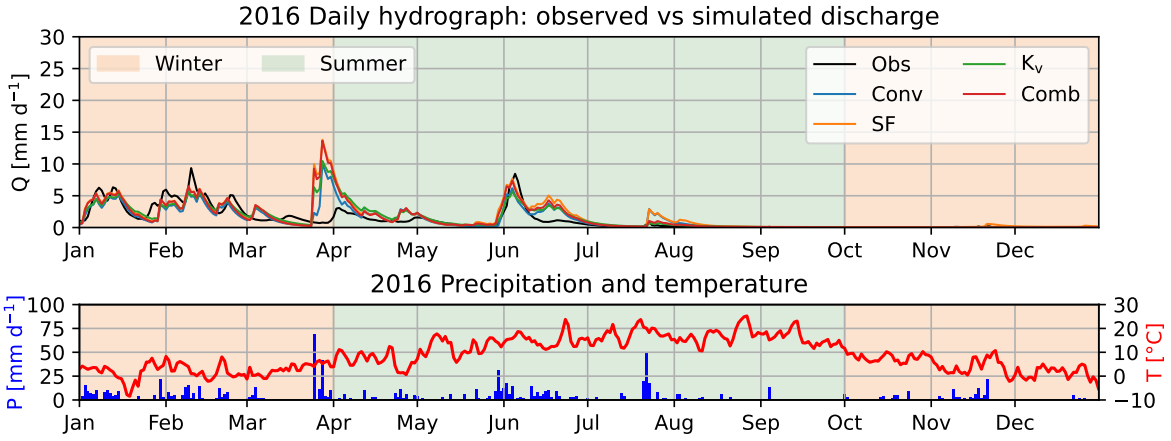


Figure D.3: Hydrograph of the observed and simulated discharges per model in the upper panel for 2016. The bottom panel shows the precipitation (P) and temperature (T) over this period. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

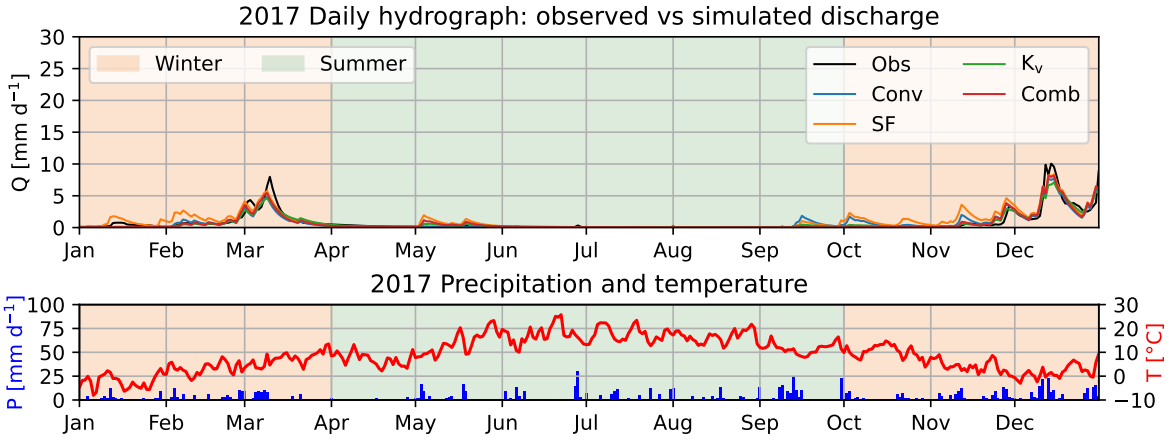


Figure D.4: Hydrograph of the observed and simulated discharges per model in the upper panel for 2017. The bottom panel shows the precipitation (P) and temperature (T) over this period. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

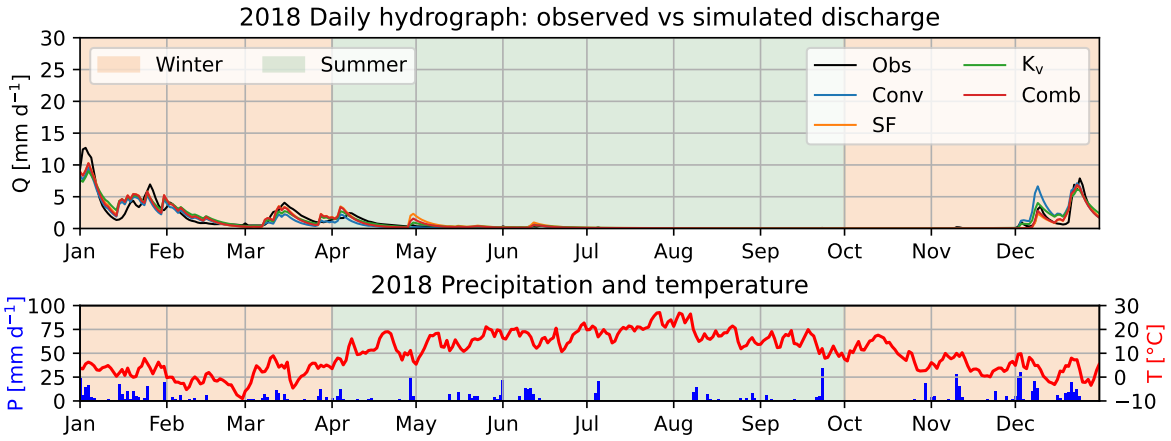


Figure D.5: Hydrograph of the observed and simulated discharges per model in the upper panel for 2018. The bottom panel shows the precipitation (P) and temperature (T) over this period. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

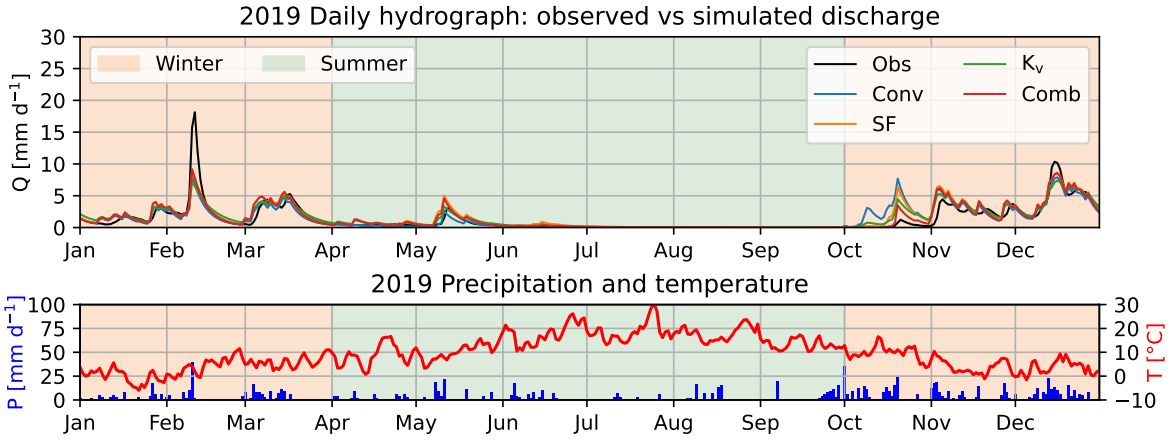


Figure D.6: Hydrograph of the observed and simulated discharges per model in the upper panel for 2019. The bottom panel shows the precipitation (P) and temperature (T) over this period. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

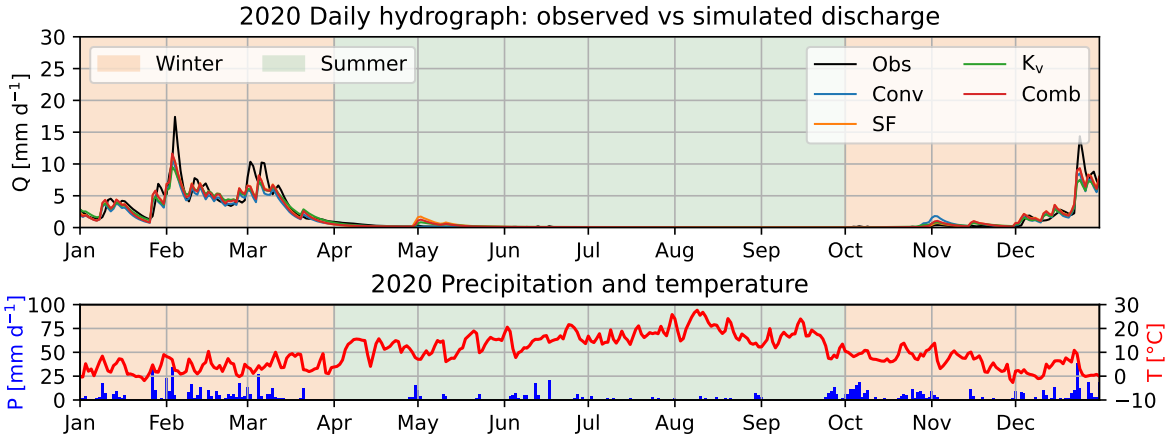


Figure D.7: Hydrograph of the observed and simulated discharges per model in the upper panel for 2020. The bottom panel shows the precipitation (P) and temperature (T) over this period. Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

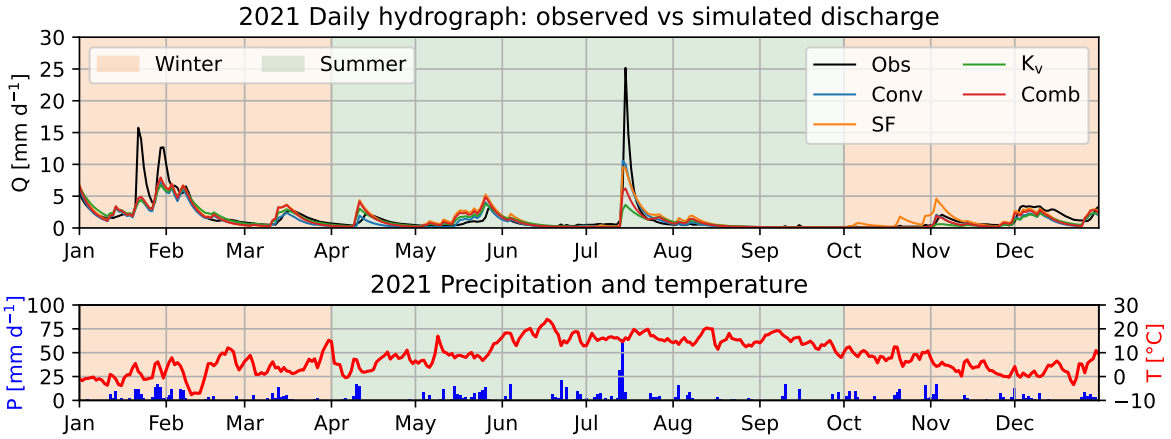


Figure D.8: Hydrograph of the observed and simulated discharges per model in the upper panel for 2021. The bottom panel shows the precipitation (P) and temperature (T) over this period.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

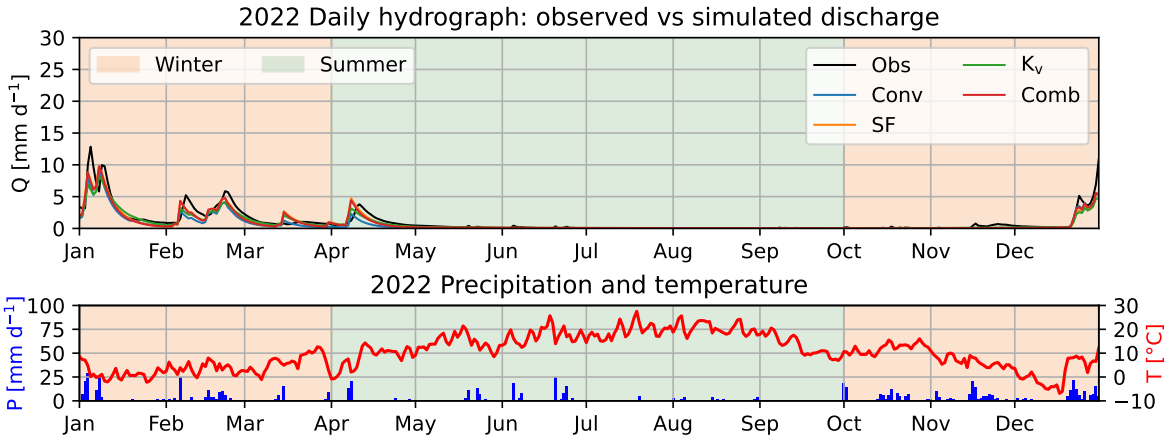


Figure D.9: Hydrograph of the observed and simulated discharges per model in the upper panel for 2022. The bottom panel shows the precipitation (P) and temperature (T) over this period.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

E
Transpiration Figures

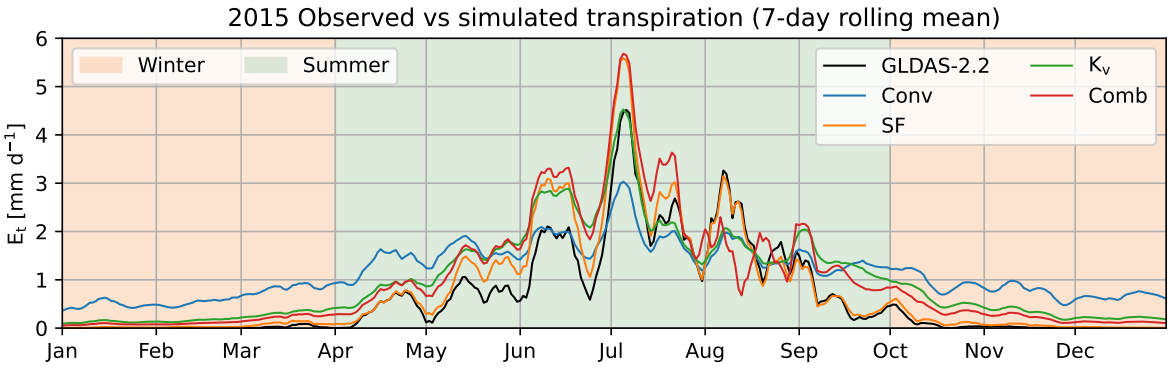


Figure E.1: The 7-day rolling mean transpiration time series for 2015 with GLDAS-2.2 transpiration data as reference.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

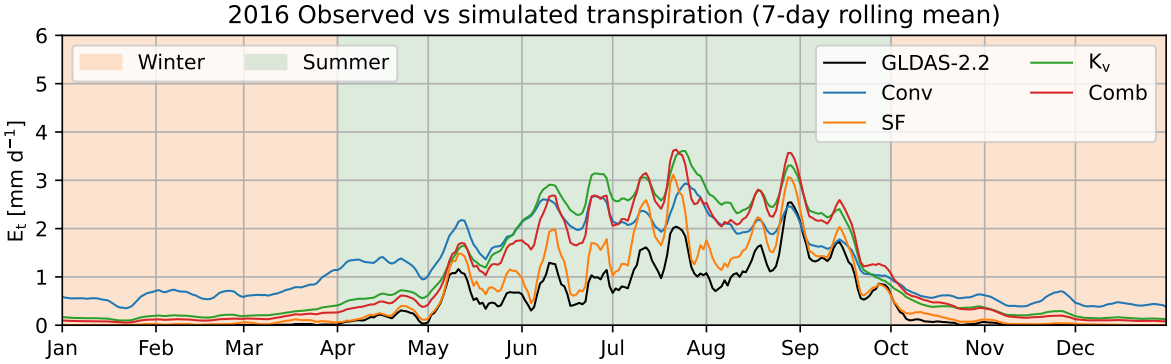


Figure E.2: The 7-day rolling mean transpiration time series for 2016 with GLDAS-2.2 transpiration data as reference.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

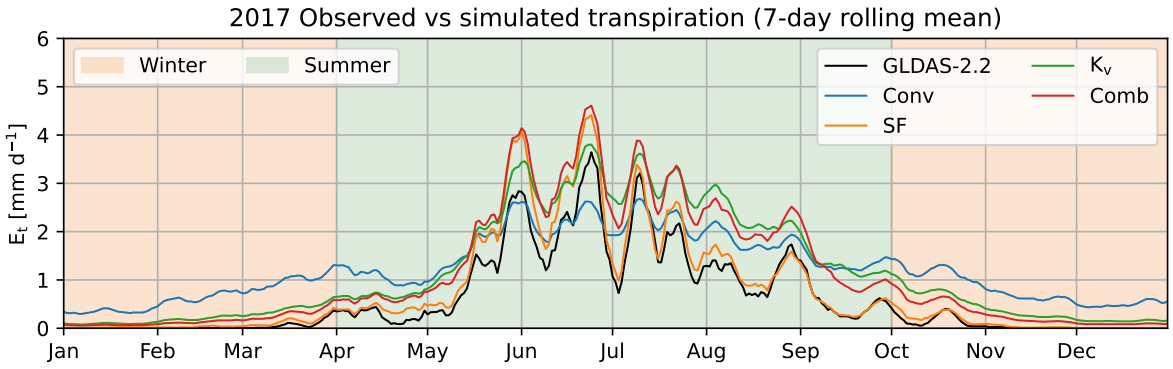


Figure E.3: The 7-day rolling mean transpiration time series for 2017 with GLDAS-2.2 transpiration data as reference.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

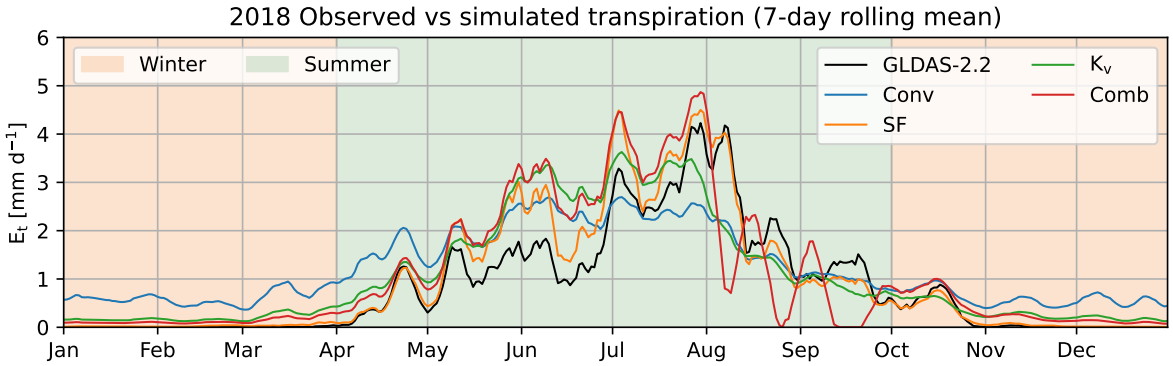


Figure E.4: The 7-day rolling mean transpiration time series for 2018 with GLDAS-2.2 transpiration data as reference.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

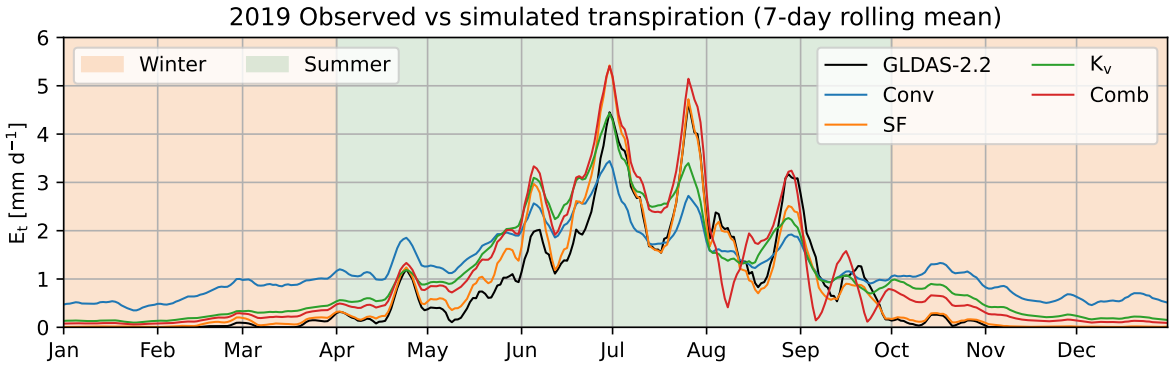


Figure E.5: The 7-day rolling mean transpiration time series for 2019 with GLDAS-2.2 transpiration data as reference.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

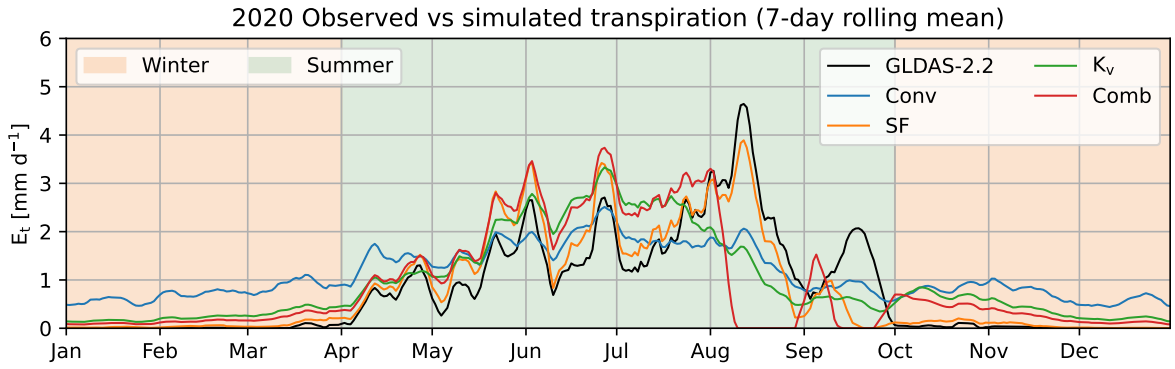


Figure E.6: The 7-day rolling mean transpiration time series for 2020 with GLDAS-2.2 transpiration data as reference.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

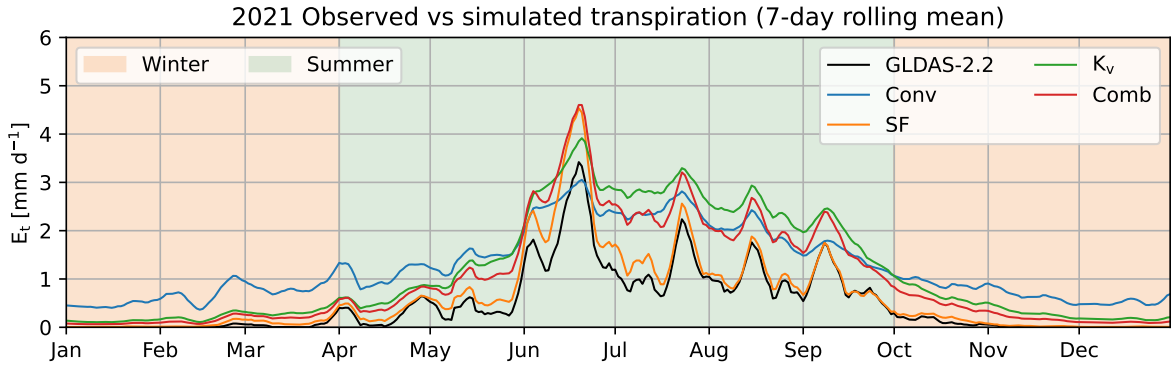


Figure E.7: The 7-day rolling mean transpiration time series for 2021 with GLDAS-2.2 transpiration data as reference.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.

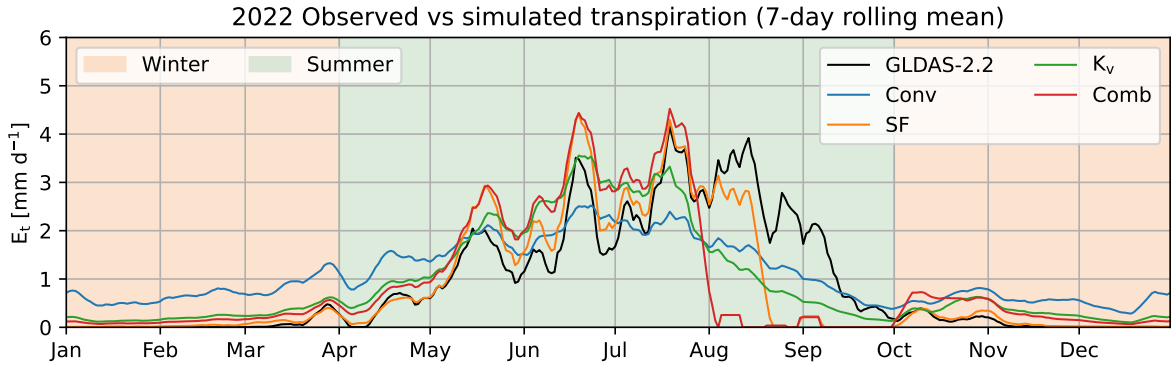


Figure E.8: The 7-day rolling mean transpiration time series for 2022 with GLDAS-2.2 transpiration data as reference.
Conv = conventional model, SF = sap flow model, K_v = K_v model, Comb = combined model.