Hydrologic Change Detection Modelling Methods for Disturbed Forested Watersheds

by

Simon Guan Min Lin

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Civil Engineering (Water)

Waterloo, Ontario, Canada, 2022

Author’s Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Decades of paired catchment studies have provided insights on how forests regulate water redistribution following vegetation disturbances such as logging and wildfire. In these settings, changes in runoff characteristics are detected by comparing streamflow responses to those of undisturbed catchments nearby. While this is generally considered the best approach for assessing disturbance impacts, the method is prone to confounding factors. Challenges can often include issues relating to spatial scale, limited resources for monitoring, and climate trends that can confound subsequent analyses.

Process-based computer models can be used to address the shortcomings of empirically based paired catchment approaches. One way this is achieved is through simulating impacted catchments under no-disturbance conditions to provide a virtually identical control for comparison. In this thesis, we address two objectives, which are to 1) assess the capability of hydrologic models and hydrologic-vegetation growth models in simulating altered streamflow patterns of forested watersheds following disturbances, and 2) compare the utility of different change detection methods in describing the hydrologic impacts of forest disturbances using hydrologic signatures.

We applied and then evaluated two process-based models (Raven and Raven Robin) using their simulated weekly runoff ratios and calibrated parameter distributions to better understand the individual effects of climate variability and harvesting on streamflow at the Turkey Lakes Watershed in Ontario, Canada. Calibrated models achieved Kling Gupta Efficiency (KGE) scores higher than the climatological reference benchmark specific to their catchment-period scenario. However, most models became less effective outside of their calibration periods due largely to the effect of climate trends (the post-harvest period exhibited significantly warmer and drier climate conditions). Pre-harvest Raven models were an exception to this, likely owing to the canopy losses that initially buffered higher evapotranspiration demands driven by a warmer climate after harvesting. Despite maintaining their predictive skill, such models are subject to equifinality concerns as they do not incorporate changes to the vegetation cover induced by harvesting.

In assessing weekly runoff ratio distributions derived from both observed and simulated data, we found that all three catchments produced low magnitudes of runoff. For this reason, we separated our data into low-flow and high-flow time steps. Through comparisons of weekly runoff distributions describing high flows, we show that pre-harvest and post-harvest periods exhibited distinct

hydrologic responses across all three catchments. This resulted in models that simulated functionally different rainfall runoff responses when calibrated to different periods. However, these responses could only be assessed during high-flow weeks; catchments exhibited dry conditions for as much as 16 weeks during the three summer months of each water year – thus suggesting that the weekly runoff ratio is not an ideal change metric for use in the non-lake headwater catchments of the Turkey Lakes watershed. This guided our subsequent assessment of model ability in simulating low-flow conditions that were excluded from our initial analysis: models calibrated to drier post-harvest conditions were generally more successful at simulating dry weeks during the summertime. This was not the case for post-harvest Raven Robin models in catchment C31, which were trained to represent a regrowth forest with lower canopy densities after treatment. This decreased interception and evaporative losses in the model, making them less successful at simulating dry conditions.

We then compared parameter distributions between model ensembles calibrated to different periods in each catchment. Results from this analysis suggest that climate trends altered soil moisture dynamics in control catchments C32 and C35, promoting more infiltration, soil evaporation, altered baseflow regimes, and an increase in plant water uptake rates. Catchment C31 models experienced parameter distribution shifts that created a new evapotranspiration regime for its vegetation in addition to enhanced soil moisture storage. Overall, the significant differences detected between parameter distributions in this analysis suggest that both climate trends and clearcut harvesting have altered hydrologic processes in the watershed.

While our results demonstrate how models can be used for hydrologic change detection, improvements can be made to the employed methods. We recommend considering phenological timescales when designing model calibration experiments to address the non-stationarity of model parameters across time. This will better represent emergent processes throughout the various stages of hydrologic recovery, allowing for improved quantification of forest disturbance impacts on watershed hydrology.

Acknowledgements

I would like to thank Dr. James Craig and Dr. Nandita Basu for their supervision and support over the duration of this degree.

I would like to thank Dr. Ming Han for his assistance and support in setting up and operating the Robin(3PG) vegetation growth model, one of the central components to this thesis, as well as in resolving the endless bugs encountered along the way.

I would like to thank Dr. Jason Leach for providing me with all streamflow and forest growth data that went into the development of the process-based models contained within this thesis.

I would like to thank Dr. Bryan Tolson for his advice and support on model calibration and on how to evaluate model ensembles.

I would like to thank Dr. Juliane Mai for her patience in helping me set up the wonderfully designed EEE toolbox to conduct my efficient elementary effects sensitivity analysis and in teaching me several bash scripting tricks to automate my workflow.

I would like to thank the countless members of both the Hydrology Group and the Basu Lab for their shared insights, jokes, and (largely virtual) support - all of which have kept me motivated in the midst of the COVID-19 pandemic despite being physically far, far away.

Dedication

This thesis is dedicated to my parents, who dared to chase their dreams into an unknown side of the world and worked tirelessly to put me where I am today.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Understanding the nexus of forests, water security, and resource management

For thousands of years, human civilizations have relied on forests for food, shelter, and resources. Across the forest, plant-water interactions modulate both the timing and magnitude of catchment runoff, combating erosion and retaining water-borne nutrients in the landscape. For this reason, forests can yield reliable quantities of high-quality water, providing communities with an ideal source of drinking water. Today, forested watersheds supply an estimated 75% of all freshwater resources around the world (Millennium Ecosystem Assessment, 2005). However, the health of these forests is often also threatened by landscape disturbances including flooding, fire, harvesting, invasive species, and more recently, climate change. For communities relying on these forests for drinking water, addressing the implications of these landscape disturbances is essential for water security. In such contexts, forests are often considered a natural economic asset that is maintained as part of source water protection programs (Emelko et al., 2011).

In source water protection programs, forest management is a major component of protecting drinking water supplies. Forest management draws upon conservation, deforestation, and afforestation activities to achieve specific outcomes. However, achieving one outcome often comes at the expense of another ecosystem attribute, often hydrological function. For example, logging yields timber for wood products but destroys habitats and reduces water storage while re-planting trees increases evapotranspiration but creates local competition for soil moisture and so on. Additionally, hydrologic impacts manifest gradually according to phenological timescales, often over the course of decades. Thus, the challenge of forest management lies not only in balancing trade-offs to satisfy conflicting interests, but also in responding dynamically to cumulative and successive disturbance effects on the landscape. Further confounding these effects is climate variability, which may exhibit trends that make the hydrologic responses of forested watersheds more difficult to predict (e.g., Goodbrand et al., 2022).

Thus, effective forest management requires a comprehensive understanding of the linkages between climate, vegetation, and hydrology. This need has driven over a century of observations, small-scale field experiments, and numerical analyses. Review studies such as those by Hibbert

(1967), Bosch & Hewlett (1982), and Brown et al. (2005) have analyzed the aggregated findings of individual case studies from around the world, forming generalizations and empirical relationships about vegetation change impacts on the annual water yield at the catchment scale. Particularly, their conclusions unveil several nuances affecting hydrologic change and recovery including disturbance characteristics, vegetation type, and regional climate.

To understand the impacts of forest management and natural disturbances on forest hydrology, hydrologic change detection and hydrologic change analysis are conducted. The two processes are sequential: Hydrologic change detection first determines the presence and significance of changes to local hydrology following a disturbance. Change detection results are then interpreted in the following analysis, which discusses reasons for why changes have or have not occurred in relation to the disturbance. Thus, all studies that investigate hydrologic changes to a system involve both hydrologic change detection and hydrologic change analysis.

Central to both hydrologic change detection and hydrologic change analysis are quantitative metrics. In a forest hydrology context, interpretable metrics link local hydrology to the state of vegetation cover within the system. Such metrics are often hydrologic signatures - metrics that track specific aspects of streamflow and the overall water balance across various timescales. A primary example of this is the annual water yield, a metric frequently used in early studies. Since then, researchers have employed more signatures to evaluate the effects of vegetation change on different aspects of the hydrologic cycle. These additional metrics include but are not limited evapotranspiration, peak flows, low flows, flow duration, soil moisture, and runoff ratios.

Together with new hydrometric data sets and computer models, hydrologic signatures enable researchers today to conduct hydrologic change studies in a digital environment. With full control over which disturbances take place, researchers can use models to help disentangle concurrent disturbance effects by subjecting catchments to one disturbance at a time. Furthermore, the use of physically-based models specifically can help researchers draw process-level insights that cannot be directly observed in reality. However, models are not without limitations (e.g., simplifying assumptions, quality and availability data for calibration and validation, etc.), and must be scrutinized during both development and deployment to avoid erroneous results.

1.2 Thesis objectives and organization

The objectives of this thesis are to 1) assess the capability of hydrologic models and hydrologic-vegetation growth models in simulating altered streamflow patterns of forests following clearcut harvesting and 2) compare the utility of different change detection methods in describing the hydrologic impacts of forest disturbances via select hydrologic signatures. The value added by performing change detection using models in addition to observation data is assessed.

The organization of this thesis is as follows. Chapter 2 provides background on the paired catchment experiment, findings from studies that have employed it, and model-based hydrologic change detection methods as an alternate approach. Chapter 3 presents the case study site, process-based models developed for the study, calibration experiments, and hydrologic change detection methods. Chapter 4 presents and discusses the results of model calibration and hydrologic change detection analyses conducted in fulfillment of the two research objectives. The findings and limitations of this work are summarized in chapter 5.

Chapter 2
Background

This chapter presents background information in two parts. Section 2.1 provides an overview of the paired catchment experiment, insights of vegetation change impacts on hydrology derived from studies that make use of its analytical techniques, and its limitations. Section 2.2 introduces models and their applications in hydrologic change detection as an alternate approach to the techniques employed by the paired catchment experiment.

2.1 Paired catchment experiments: an empirical approach to hydrologic change detection

In this section the paired catchment experiment is first defined in section 2.1.1. In section 2.1.2, a summary of findings from "paired catchment studies" is presented. The limitations of paired catchment studies are discussed in section 2.1.3.

2.1.1 Definition, adaptation, and applications

Researchers and practitioners have long sought to understand vegetation impacts on catchment hydrology - particularly for forest management applications which often alter vegetation cover. This pursuit of knowledge has led to experimental treatments in small, measurable parcels of forests all over the world. The work of Bates & Henry (1928) is often credited by subsequent reviews as the first of its kind (Bosch & Hewlett, 1982; Hibbert, 1967; Stednick, 1996); this study monitored streamflow before and after clearing a stand of aspen and mixed spruce trees in an alpine catchment of the Colorado mountains in the United States to observe logging impacts on hydrology.

The evaluation technique demonstrated by Bates & Henry (1928) laid the foundation for what later became the "paired catchment experiment", a method that yields experimental data upon which researchers can derive quantitative insights about the impacts of vegetation change. At its core, the paired catchment experiment assumes that a relationship between two catchments exists indefinitely until one of them is significantly altered. In hydrologic studies, catchments exhibit similar hydrologic relationships when they feature similar drainage areas, geology, slope, aspect, elevation, land use, and vegetation cover.

When one catchment is altered, it is expected to respond differently to climate conditions and deviate from its original relationship with its paired, undisturbed reference (the 'control' catchment). Occasionally, a secondary control catchment may also be incorporated to investigate the effect of spatial variability. In all cases, observations are recorded across all catchments before and after a prescribed treatment is applied. Data collected before treatment are used to develop an initial relationship between the two catchments. One of the most common approaches for doing this is through linear regression, as demonstrated by Hornbeck et al. (1993) in their comparison of annual water yields between 11 catchments in northeastern USA. Developed relationships are then used to predict hydrologic responses of the altered catchment during the post-treatment period given the response of the control catchment. Differences between the predicted and actual value (observed in reality) are then calculated to describe the effect of treatment.

This analytical method has since been widely adopted for evaluating impacts on several catchment attributes including water quantity, water quality, and nutrient cycling (Stednick, 1996). Elements of this method have also been useful for assessing the impacts of unplanned forest disturbances such as wildfire. In practice, after-the-fact comparisons are often made in a less rigorous fashion with whatever hydrometric data are available to derive at best "circumstantial evidence" of disturbance effects (Hewlett, 1971). Given that not all cases employing this method meet the formal definitions of the original paired catchment experiment, the plethora of studies that exist to date are referred to in recent literature as "paired catchment studies" or simply studies that employ the "paired catchment approach" (Brown et al., 2005; Seibert & McDonnell, 2010; Zégre et al., 2010).

2.1.2 Insights on vegetation impact on hydrology: a century's worth of case studies

Over the past century, paired catchment studies have been conducted all over the world in a variety of climates, featuring several types of vegetation and disturbances. The collective insights from 39 studies were first summarized by Hibbert (1967), who made three generalizations about vegetation impacts on water yield:

- Reduction of forest cover increases water yield.
- Establishment of forest cover on sparsely vegetated land decreases water yield.
- Response to treatment is highly variable and, for the most part, unpredictable.

These generalizations were further evaluated by Bosch & Hewlett (1982), who compiled more paired catchment study results (bringing the review to a total of 94 studies) and updated sets of regression lines that mapped percent forest cover reductions to annual water yield increases after harvesting for a variety of vegetation types. From these regressions, Bosch & Hewlett (1982) determined that forest cover reductions of 20 percent or less rarely induced hydrologic changes that are detectable through streamflow measurements; this rule of thumb follows the logical assumption that treatment effects decrease with treatment extent, given that zero treatment should produce zero effect (McMinn & Hewlett, 1975). It should be noted however, that in these instances, hydrologic changes can still be present in other forms and at different scales despite exhibiting weak to non-existent streamflow signals.

Operational definitions of recovery from hydrologic changes were subsequently discussed by (Stednick, 1996). Initially, hydrologic recovery was first defined as "the return of annual water yield to pre-treatment levels" (Hibbert & Gottfried, 1987; Hornbeck et al., 1993; Stednick, 1995; Stednick & Kern, 1992), but was later amended to include additional hydrologic phenomena of interest referenced in other studies (Cheng, 1989; Harr, 1976; Harr et al., 1979; Hicks et al., 1991; Keppeler & Ziemer, 1990; Stednick, 1995; Whitehead & Robinson, 1993).

Stednick (1996) also highlighted factors affecting hydrologic recovery. Based on the findings of Swank et al. (1988), Burt & Swank (1992), and Stednick (1995), they identified climate variability and vegetation regrowth rates (indicated by the Leaf Area Index) as significant controls on streamflow responses to treatment. Stednick (1996) pointed out that vegetation regrowth exerted a dominant control when treatments were situated in high-rainfall areas, whereas climate variability had more of an influence in low-rainfall areas. Furthermore, the physical location of the treatment - relative to the contributing area of the catchment - was also noted as a factor affecting the variability in water yield changes reported between Bosch & Hewlett (1982), Troendle & King (1985), and Stednick (1995).

One decade later, Brown et al. (2005) presented an updated literature review. Here, they highlighted different types of vegetation changes (deforestation vs afforestation) and the respective timescales at which they can impact catchment hydrology. Their discussion concluded that:

- The annual water yield of a catchment takes more than 5 years to reach a new equilibrium following disturbances that permanently alter vegetation, with deforestation disturbances requiring less time to do so than afforestation disturbances.
- At sub-annual timescales, all catchments reviewed in the study exhibited the largest and smallest absolute volume changes during the wet and dry periods respectively.
- Seasonally distinct periods can sometimes account for all or none of the annual water yield changes and should be interpreted proportionally when discussing results on an annual basis.
- Flow duration curves provide a sense of flow magnitudes at various timescales and can be used to assess vegetation change impacts on high- and low-magnitude flows.

Despite these fruitful insights, the authors of this review paper also emphasize that their conclusions stem from studies conducted in "relatively small catchments" and call for methods to scale the findings to larger catchments where disturbance-impacted areas do not make up most of the domain, as is often the case in forest management applications.

2.1.3 Limitations of paired catchment studies

The paired catchment approach has yielded several insights about the impacts of vegetation change on catchment hydrology but also has several limitations frequently mentioned in discussions by past studies.

The first limitation is that of spatial scale. The paired catchment experiment is ideal for relatively small catchments: in practice, treatments can be uniformly applied over study areas up to 100 ha, above which measurement of catchment conditions (such as temperature and total precipitation) begins to lose accuracy (Bosch & Hewlett, 1982; Hewlett, 1971). Furthermore, in real-world applications involving large basins, a perfect control catchment becomes rather difficult to identify and may not actually exist because of spatial heterogeneity.

The second limitation is that of available resources to support monitoring activities. Finite amounts of funding, equipment, and personnel dictates both the spatial extent and duration of the data collected from paired catchment studies (Andréassian et al., 2003). Again, this confines experiments

to relatively small catchments but can also terminate data records before hydrologic recovery is observed.

Finally, one of the most challenging limitations in relation to short data records is capturing climate variability (Brown et al., 2005). Weather conditions occurring within a data record may not necessarily be representative of the entire range of the local climate and its full suite of associated hydrologic responses. Given a limited number of observations, reported hydrologic changes may still be well within the typical range of hydrologic responses.

2.2 Model-based hydrologic change detection

This section presents the case for using process-based models to detect hydrologic change as an alternative to observation-based methods in the paired catchment approach. A background on process-based hydrologic models is first presented in section 2.2.1, followed by an overview of model-based hydrologic change detection methods in section 2.2.2.

2.2.1 Process-based models

Process-based hydrologic models approximate real systems and how they respond to climate conditions. Such models are developed with a system of equations, most of which are physically based and related to specific hydrologic processes (e.g., snow melt, evapotranspiration). With an understanding of these equations, modellers can derive insights about properties of the real system when interpreting model outputs. This is particularly useful because some of these properties cannot be directly observed (e.g., hydraulic conductivity of deep soils). In other words, models can offer additional information about the system for every observed data point it simulates.

Furthermore, hydrologic models can be used to extend our understanding of the system under various conditions. This is the basis of hydrologic modelling applications, where streamflow predictions can be used to forecast floods in the short-term future (Kauffeldt et al., 2016), simulate flow regimes under potential climate scenarios in the long-term future (Dibike & Coulibaly, 2005), and more recently, detect and evaluate hydrologic changes associated with land use or land cover changes that have already occurred (e.g., Brandt et al., 1998; Lavabre et al., 1993; Schreider et al., 2002).

When analyzing simulated data, model uncertainty needs to be addressed. Model uncertainty is related to equifinality - the idea that many different models can all reasonably simulate observation

data (Beven, 1993). This is an inevitable artifact of model development, where subjective decisions are made about input forcing data, model structure, and model parameters (Beven, 2005; Butts et al., 2004; Chlumsky et al., 2021; Gupta et al., 2012). As such, one of the key contributions of this work is strategies for handling this uncertainty during model-based change detection analyses: although this uncertainty cannot be eliminated, it can be quantified by analyzing multiple calibrated model predictions (as opposed to that of a single best model) that each give unique results.

Ensemble model predictions provide a means for quantifying model uncertainty, which can be done using a variety of approaches (e.g., Beven & Binley (1992), Blasone et al. (2008), Kuczera & Parent (1998)). A model ensemble contains models that provide equally plausible yet differing approximations of the real system. These models are calibrated and meet or exceed a minimum performance benchmark (specified by the modeller). The benchmark is often a model evaluation metric that summarizes the agreement between simulated and observed data. Examples of common hydrologic model evaluation metrics include the Nash Sutcliffe Efficiency (NSE) score (Nash & Sutcliffe, 1970), root mean squared error (RMSE), percent bias, and Kling Gupta Efficiency (KGE) score (Gupta et al., 2009). In developing a model ensemble during model calibrations, "behavioral" models satisfy the performance benchmark and are retained while under-performing "non-behavioral" models are discarded. The constructed model ensemble can also be evaluated for its performance with reference to characteristics such as reliability and sharpness. These can be summarized by other evaluation metrics such as the continuous ranked probability score (Hersbach, 2000; Matheson & Winkler, 1976), average interval score, and prediction interval average width (Papacharalampous et al., 2020).

2.2.2 Model-based hydrologic change detection methods

Hydrologic models can be used in many ways to detect hydrologic change. The various methods, such as those proposed by Andréassian et al. (2003), Seibert et al. (2010), Seibert & McDonnell (2010), and Zégre et al. (2010), all begin with model calibration experiments: observed data are separated into distinct periods for calibration and change detection (herein referred to as the calibration period and the evaluation period). Ensembles of calibrated models are run in both periods, providing simulated data for analysis. These simulated data can be paired with observed data to quantify model error (model residuals), or with simulated data from other models to calculate relative differences between model predictions.

Model residuals and relative differences between models provide data used to detect hydrologic change. Both form sampling distributions that can be compared through statistical tests across periods or across models to detect hydrologic change. In this context, significance test results indicate that a catchment in reality exhibits distinct hydrologic responses between two periods - leading to different model representations of the system (depending on the period selected for calibration), as well as a decrease in predictive skill when models are applied to data not represented in the calibration period.

Model parameters can also inform the interpretation of hydrologic change. In another approach proposed by Seibert & McDonnell (2010), calibrated model parameters are compared between model ensembles. In this method, distributions of parameter values are generated when ensembles of models are calibrated to different periods. The parameter distributions between model ensembles are compared with statistical tests during the change detection analysis: significant test results here suggest that the system has distinct functional properties between periods, which may contribute to fundamentally different hydrologic responses.

In all cases, assumptions about the simulated data need to be considered when conducting statistical tests. Hydrologic data are serially correlated and often exhibit non-normal behavior as well as seasonality (Kundzewicz & Robson, 2004). In statistical tests, simulated data and those derived from it (e.g., model residuals and relative differences) should be sampled in a way that allows independence and constant variance (homoscedasticity) to be assumed. This may involve re-aggregating or re-sampling the data and, in some cases, adjusting for autocorrelation. The statistical test itself should ideally be a distribution-free method which allows for minimal assumptions to be made about the data but may in some cases (depending on how the data are sampled) still assume normality.

Numerous statistical tests can detect step changes. They evaluate differences between specific characteristics of data groups, sometimes even when the timing of the change point (e.g., forest disturbance event) is unknown. These include the median change point test (Pettitt, 1979; Siegel & Castellan, 1988) and the Worsley likelihood ratio test (Worsley, 1979) (looks at differences between means). Other methods that assume the change point is known include the Wilcoxon ranked-sign test (Wilcoxon, 1945) (looks at differences between two groups as a whole), Kruskal-Wallis test (Kruskal & Wallis, 1952) (tests for equality of means or variance), and the Kolmorgorov-Smirnov test

(Kolmorgorov, 1933; Smirnov, 1939) (looks at differences between empirical cumulative distributions).

Chapter 3
Methods

This chapter first describes the study site and the hydrologic change detection analyses previously conducted in section3.1, process-based models development in section 3.2, sensitivity analysis and calibration strategies in section 3.3, and hydrologic change detection analyses conducted in section 3.4.

3.1 The Turkey Lakes Watershed

This section first presents an overview of the Turkey Lakes watershed in section 3.1.1, followed by a summary of the previously conducted observation-based hydrologic change detection analyses in section 3.1.2.

3.1.1 Site Description

The Turkey Lakes Watershed is a 10.5 km^2 forest catchment situated approximately 60 km north of Sault Ste. Marie, Ontario. As part of the Great Lakes - St. Lawrence forest region of the Boreal Shield Ecozone, this mixed-age hardwood system consists of 90% sugar maples, 9% yellow birch, and 1% conifers (Jeffries et al., 1988; Rowe, J.S., 1972).

Canadian government agencies designated this watershed in the 1980s as an experimental study site for investigating the impacts of acid rain. Since its inception, the Turkey Lakes Watershed now hosts a variety of research investigating aquatic, ecologic, and hydrologic impacts of climate change, anthropogenic pollution, and land use alterations (Environment and Climate Change Canada, 2018). The scientific findings from these research activities at the Turkey Lakes watershed have then served to inform “high profile policy issues related to cumulative effects, drinking water treatability and ecosystem resilience to climate change” (Webster et al., 2021). The study presented herein hopes to contribute to this body of diverse research by evaluating the hydrologic impacts of a clearcut harvesting experiment at headwater catchment C31: harvesting took place during the fall of 1997, while forests at catchments C32 and C35 were left intact as spatial references (Buttle et al., 2018). These catchments are depicted in Figure 1.

The watershed records local climate conditions from an internal weather station but is also supplemented by Environment Canada weather stations in Sault Ste. Marie to the south and Lake

Superior National Park to the north. Local elevations range between 330 to 625 m above sea level, with a total relief of about 300 m (Leach et al., 2020). The watershed receives a mean annual precipitation of 1189 mm with 35% on average being snowfall, which accounts for 30% to 60% of the average annual runoff (Buttle et al., 2018; Nicolson, 1988; Semkin & Jeffries, 1988). This often occurs through shallow subsurface flows during snowmelt, given that deep snowpacks accumulate early in the winter months and effectively insulate the underlying soils to prevent permafrost formation (English et al., 1986).

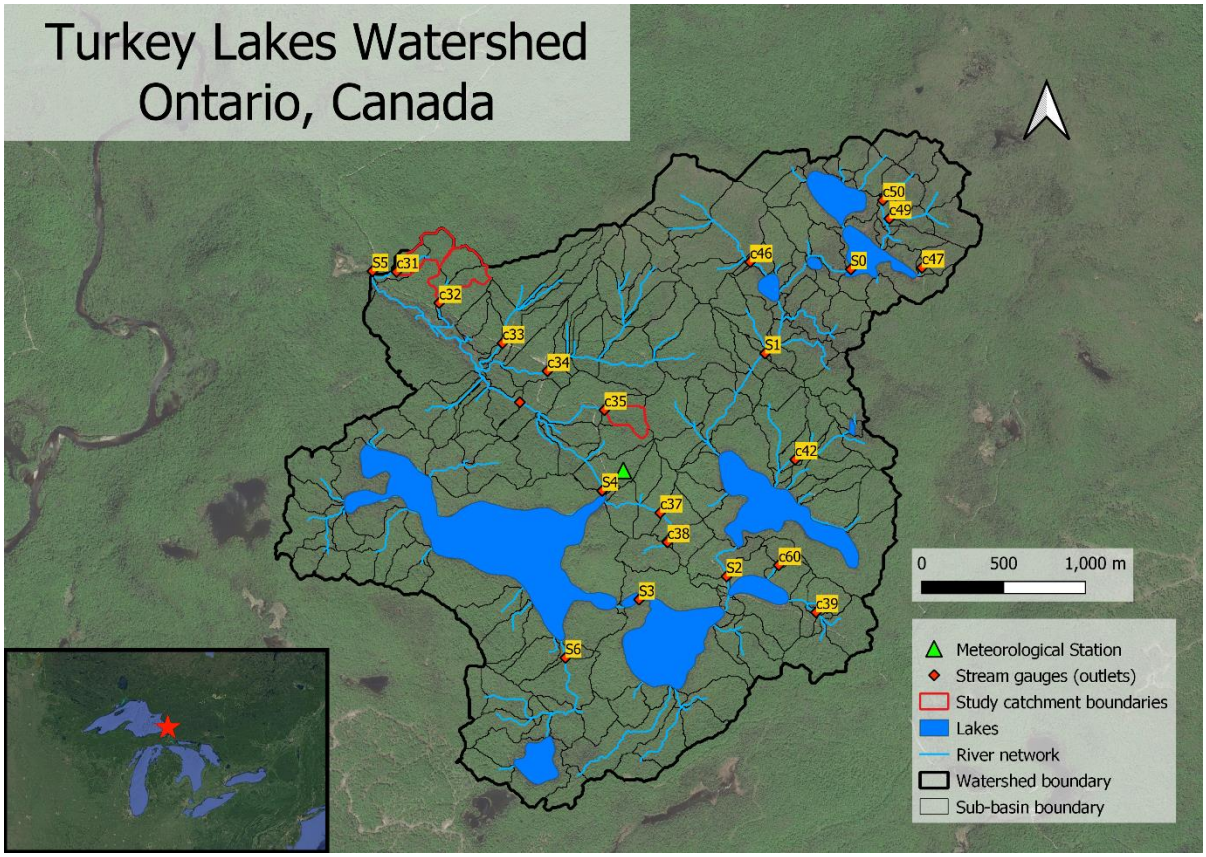

Figure 1. Turkey Lakes Watershed site map

Soils in the Turkey Lakes Watershed are characterized by a silty loam ablation till underlain by a sandy loam basal till that rests on top of the regional Precambrian metamorphic basalt bedrock. Topsoils contain more organic matter (orthic humo-ferric podzols) and are usually 0.05 m deep within the overall soil profile that extends 0.5 m to bedrock within headwater catchments (Buttle et al., 2019; Hazlett et al., 2001, 2011).

The Turkey Lakes stream network drains 4 lakes sequentially from the northeast to the south and then to the west. The watershed outlet connects to the Batchawana River, which ultimately discharges into Lake Superior. Sub-catchment outlets are outfitted with 90-degree v-notch weirs, from which stage-discharge relationships are developed in conjunction with stilling wells and water level loggers (Beall et al., 2001; Buttle et al., 2018). Streamflow data recorded at these weirs range from 1981 to 2012. These relationships are used to report daily streamflow values based on water level measurements. Reported flows represent most of the runoff leaving each sub-catchment but may not capture deep subsurface flows bypassing the underside of the weirs within the basal till (Leach et al., 2020). As such, runoff estimated from weirs likely underestimates actual streamflow losses from the site.

3.1.2 Summary of previous observation-based hydrologic change detection studies

Past studies at the Turkey Lakes watershed have evaluated clearcut harvesting effects on different aspects of the hydrologic cycle. Murray & Buttle (2003) compared 2 years of snow accumulation and snow melt on hillslopes of a clearcut catchment and an adjacent control catchment: they found that harvesting increased snow accumulation and daily melt rates to different extents, with slope aspect exerting a greater influence on the spatial variability of melt rates than harvesting. Buttle et al. (2018) assessed hydrologic responses to multiple silvicultural treatments including selection harvest, shelterwood harvest, and clearcut harvest. They found that clearcut harvesting induced the most "pronounced and prolonged response": runoff increased across all seasons with summertime flows showing the largest increase. All three treatment types were found to increase direct runoff and baseflow, with geochemical tracer experiments revealing the enhancement of surface and near-surface flows after vegetation losses. Following these findings, Buttle et al. (2019) investigated the specific impact of harvesting treatments on quickflow generation. Their findings suggest that 1) quickflow generation followed a threshold-based relationship governed by precipitation amounts and; 2) harvesting changed this relationship and enhanced the ability of a catchment to generate quickflow at lower precipitation amounts.

The findings of these past studies are based on interpretations of observed data. Although these studies have employed empirical modelling approaches such as piecewise regression, no analyses to date have used process-based models to evaluate the hydrologic impacts of forest disturbances at the Turkey Lakes watershed. As such, the study presented herein seeks to address this gap: process-based

models will be used to simulate and detect hydrologic changes in both clearcut and control catchments. The hope is that subsequent analyses with simulated data will draw insights that would otherwise not be readily derived from previous observation-based approaches.

3.2 Model descriptions

This section provides full descriptions of two process-based models developed to perform hydrologic change detection at the headwater catchments of the Turkey Lakes watershed. Section 3.2.1 presents the base hydrologic model, while section 3.2.2 presents its augmented configuration coupled with a vegetation growth model.

3.2.1 Turkey Lakes Watershed Hydrologic Model

Two process-based models were set up to simulate streamflow from non-lake catchments C31, C32, and C35 (see Figure 1 above): a modified Canadian Shield hydrological model in built in the Raven Hydrological Modelling Framework (Craig et al., 2020) (herein referred to as "Raven") and the same model augmented with vegetation growth routines through coupling with the Robin(3PG) model developed by Han (2022) (herein referred to as "Raven Robin"). Robin(3PG) is based upon the 3PG (Physiological Processes Predicting Growth) model developed by Landsberg & Waring (1997), but simulates evapotranspiration (ET) and leaf area index (LAI) at a daily time step. These vegetation state variables are passed as inputs to the Raven model, leading to changing hydrologic responses in Raven-based routines.

Both Raven and Raven Robin models share the same watershed discretization scheme across all three catchments, produced by the BasinMaker watershed delineation toolbox (Han et al., 2022). Drainage boundaries were first delineated using the MERIT Digital Elevation Model (90 m x 90 m spatial resolution) (Yamazaki et al., 2017). Next, catchment characteristics (land use, vegetation, soil, slope, and aspect) were intersected to form hydrologic response units (HRUs), with each representing one unique combination. Slopes were split into 3 profiles ($0°\sim5°$,$5°\sim20°$, and $20°\sim90°$), each with their own soils configuration. Aspects were characterized as either north- or south-facing bins. In these non-lake catchments, vegetation and land use types were treated as a homogenous hardwood forest emulating stands of sugar maples. This resulted in a total of 4, 3, and 4 HRUs for catchments C31, C32, and C35 respectively. These HRUs feature their own state variables (e.g., soil moisture and

canopy storage) that are tracked by Raven and Raven Robin models in each time step of the model simulation (Craig et al., 2020).

There are three distinct soils in the Turkey Lakes watershed. Based on previous field studies by Hazlett et al. (2001) the shallow subsurface is an ablation till containing primarily silty loam with organic matter concentrated towards the topsoil. This distinction is represented in the model by soil layers 1 and 2, named "forest floor" and "ablation till" respectively. The ablation till is underlain by a less conductive "basal till" (soil layer 3) featuring sandy silt in the deep subsurface.

Potential evapotranspiration (PET) is estimated at every daily time step using the Raven implementation of the Priestley-Taylor equation via Equation 1 (Priestley & Taylor, 1972):

$$PET = 1.26 \cdot \frac{1}{\rho_w \lambda_v} \cdot \left[\frac{\Delta}{\Delta + \gamma} \cdot R_n\right]$$

Equation 1

where R_n is the net incoming radiation (MJ/m2/d), ρ_w is the density of water (kg/m3), Δ is the slope of the saturated vapor pressure curve, and γ is the psychrometric constant that varies with atmospheric pressure p (Craig & the Raven Development Team, 2022):

$$\gamma = \frac{c_a}{0.622 \cdot \lambda_v} \cdot p$$

The latent heat of vaporization of water λ_v (MJ/kg) is estimated in Raven using air temperature, T (Craig & the Raven Development Team, 2022):

$$\lambda_v = 2.495 - 0.002361 \cdot T$$

Longwave radiation is estimated using the Stefan-Boltzmann law with corrections for the surface and atmospheric emissivity values. Here, forests are treated as black bodies (emissivity=1) that cover 100% of forested HRUs, which make up the entirety of all three catchments. All other forcing variables are directly informed by the Turkey Lakes Watershed climate data, including precipitation, shortwave radiation, wind speed, and relative humidity.

The forest canopy intercepts precipitation as a linear proportion of the leaf area index (LAI). Intercepted precipitation can then either evaporate at the maximum PET rate or fall instantaneously to the ground (represented as a gross canopy drip process) if the maximum canopy storage is exceeded.

Snowpack depletion occurs at the potential melt rate calculated from a modified degree-day approach used in the HBV (Hydrologiska Byråns Vattenbalansavdelning) model (Bergström & Singh, 1995) that features correction factors for forest coverage and aspect. Snowpacks are also allowed to refreeze following a degree-day relationship. Only snow water equivalent (SWE) and liquid water in the snowpack is simulated; density and albedo changes are not explicitly represented.

Green-Ampt infiltration delivers ponded water into the forest floor and then percolates into the ablation and basal till layers at a rate linearly proportional to the soil moisture as used in the GAWSER (Guelph All Weather Sequential Event Runoff) hydrologic model (Schroeter, 1989). Following the HBV model (Bergström & Singh, 1995), capillary rise occurs from the basal till to the ablation till at a rate linearly proportional to the saturation of the ablation till. Baseflow proportional to a power law of saturation above a threshold flushes groundwater from all three soil layers to surface water when saturation thresholds are exceeded, as calculated by Equation 2 (Craig & the Raven Development Team, 2022):

$$Q_{baseflow} = Q_{max}\left(\frac{\left(\frac{\phi}{\phi_{max}}\right) - S_{thres}}{1 - S_{thres}}\right)^{n}$$

Equation 2

where Q_{max} is the maximum baseflow [mm/d], ϕ is the moisture content [mm], ϕ_{max} is the maximum soil storage [mm], S_{thres} is the saturation threshold below which no baseflow occurs [-], and n is the power law coefficient [-].

Soil evaporation also occurs in the top two soil layers as a linear function of soil saturation and progresses sequentially starting from the top layer (Equation 3; (Craig & the Raven Development Team, 2022)):

$$M_{evap}^{upper} = PET \cdot \min\left(\frac{\phi^{upper}}{\phi_{tens}^{upper}}, 1\right)$$

$$M_{evap}^{lower} = \left(PET - M_{evap}^{upper}\right) \cdot \min\left(\frac{\phi^{lower}}{\phi_{tens}^{lower}}, 1\right)$$

Equation 3

where M_{evap}^{upper} and M_{evap}^{lower} are the upper and lower soil layer evaporation rates [mm/d] respectively, PET is the potential evapotranspiration rate [mm/d] obtained from Equation 1, ϕ is the soil water content [mm] and ϕ_{tens} is the maximum tension storage [mm], calculated as (Craig & the Raven Development Team, 2022):

$$\phi_{tens} = \phi_{max}(S_{fc} - S_{wilt})$$

where S_{fc} and S_{wilt} are the field capacity and wilting point saturations respectively.

Finally, diffusive wave routing delivers surface water to each catchment outlet via the delineated stream network.

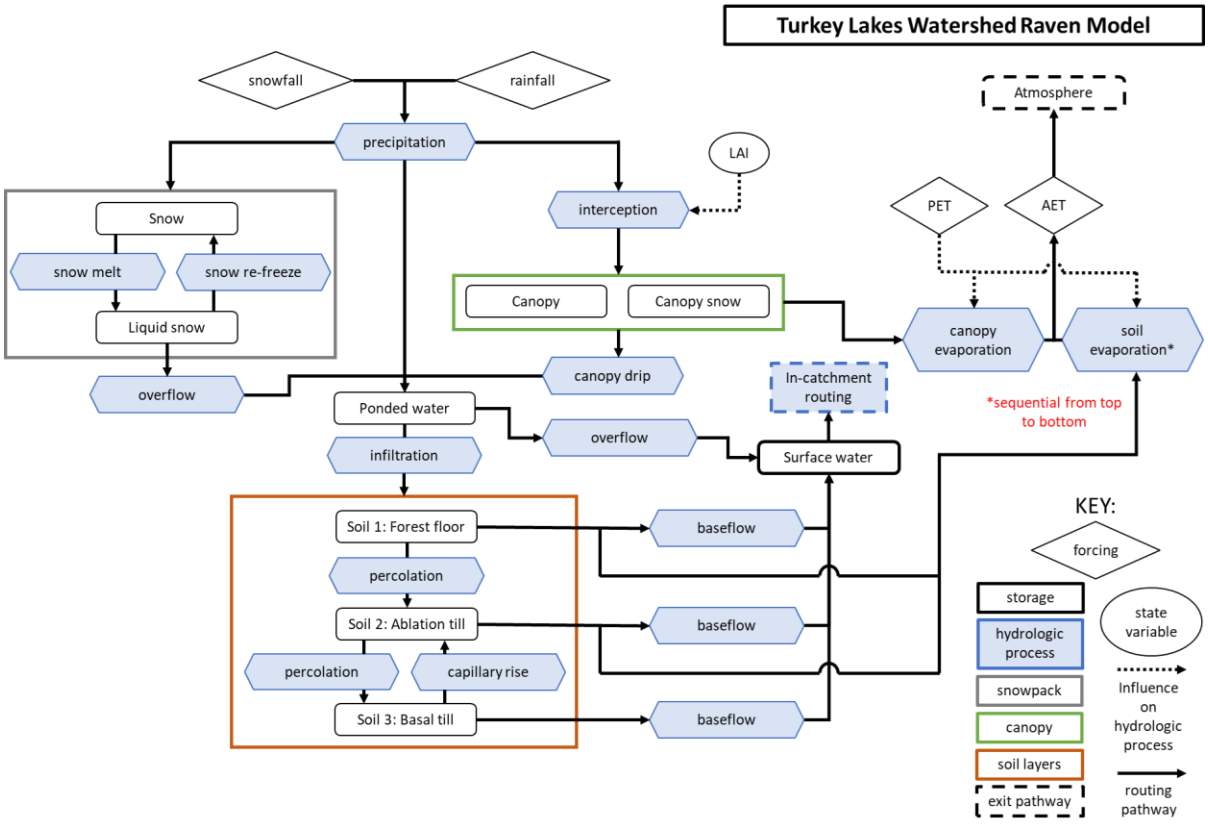

Figure 2. Turkey Lakes watershed base Raven hydrologic model schematic

3.2.2 Turkey Lakes Watershed coupled hydrologic-vegetation growth model

The base hydrologic model (Raven) was augmented in three ways with phenology routines through coupling with the Robin 3PG model (Robin) (Han, 2022). In this configuration, pre-harvest and post-harvest forests were represented in Robin as two vegetation types with distinct parameter sets.

First, Raven supplies the actual and potential evapotranspiration ratio (AET/PET) for every hydrologic response unit (HRU) to Robin as an indicator of vegetation water stress. With this value at every daily time step, Robin calculates an updated LAI value and passes this back to Raven for use in canopy-level processes (e.g., interception).

Secondly, the updated LAI is used to adjust the daily potential evapotranspiration rate PET_i through a correction coefficient k_c, as described by (Han, 2022)such that:

$$PET_{i,adj} = kc_{i,j} \times PET_i$$

$$kc_{i,j} = kc_{min,j} + \left(kc_{max,j} + kc_{min,j}\right) \times \frac{LAI_{i,j} - LAI_{min,j}}{LAI_{max,j} - LAI_{min,j}}$$

Equation 4

where $kc_{i,j}$ and $LAI_{i,j}$ are the PET correction coefficient and leaf area index of one hydrologic response unit (HRU) on day i for vegetation type j, while $kc_{min,j}$,$kc_{max,j}$, $LAI_{min,j}$, and $LAI_{max,j}$ are the corresponding minimum and maximum theoretical values.

Finally, this adjusted PET rate estimates AET from all soil layers based on the layer depth and root depth as described in the SWAT (Soil Water Assessment Tool) model (Arnold et al., 2012). This allows for soil water removal from all layers simultaneously rather than sequentially from the top two layers only as done in the base Raven model. The actual evapotranspiration $AET_{i,m}$ from soil layer m on day i is calculated using Equation 5 (Han, 2022):

$$AET_{i,m} = PET_{i,adj} \times rf_{i,m} \times \frac{\theta_{i,m} - \theta_{wilt,m}}{sf_j \times (\theta_{fc,m} - \theta_{wilt,m})}$$

Equation 5

where $rf_{i,m}$ is the root fraction occupying the soil layer m, $\theta_{i,m}$ is the soil water storage on day i, $\theta_{wilt,m}$ and $\theta_{fc,m}$ are the wilting point and field capacity soil water storage, and sf_j is the stress coefficient ranging from 0 to 1 for vegetation type j. A constant root depth of 2 meters was applied to both pre-harvest and post-harvest forest vegetation types.

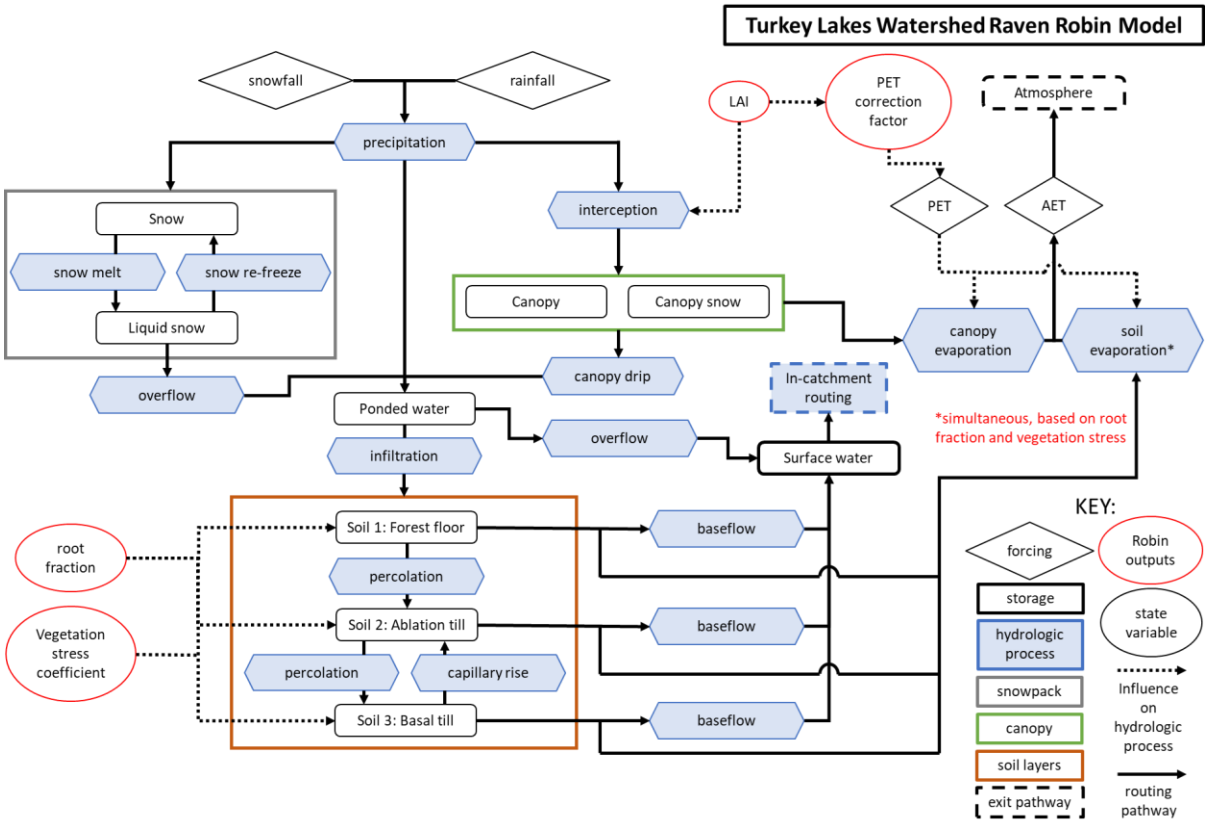

Figure 3. Turkey Lakes watershed base Raven Robin hydrologic-vegetation growth model schematic

3.3 Model calibration

This section describes the methods used in identifying calibration parameters (section 3.3.1) and in calibrating model ensembles (section 3.3.2).

3.3.1 Elementary Effects Sensitivity Analysis

A sensitivity analysis was conducted for both models in each catchment following the Efficient Elementary Effects (EEE) method of Cuntz et al. (2015) to identify calibration parameters in both

Raven and Robin: hydrology parameters (Raven) and vegetation parameters (Robin) were evaluated for their influence on the KGE of summertime (July to September) flow and Leaf Area Index (LAI) across the entire record period, respectively. Uniform random sampling was iteratively conducted for each parameter p_i based on a predetermined range to calculate its elementary effect on the model diagnostic $f(p_i)$ (Cuntz et al., 2015):

$$EE_i = \frac{f(p_i + dp_i) - f(p_i)}{\Delta}$$

Equation 6

where Δ is a fraction of the parameter perturbation, ranging between 0 and 1. A logarithmic curve was fitted to the ranked elementary effects, and a sensitivity threshold was defined at its inflection point. Parameters with elementary effects above this threshold were identified as informative to the corresponding model diagnostic and retained for model calibration, while parameters with elementary effects below this threshold were considered non-informative and fixed with default values. The number of calibration parameters retained for each model-catchment scenario are summarized in Table 1, while a full list of calibration parameters for each model ensemble can be found in Appendix B.

Table 1. Model parameter sensitivity analysis results for each model-catchment scenario

Catchment	Model	Parameters tested	Total retained for calibration
C31 (clearcut)	Raven	50	21
C31 (clearcut)	Robin	66	19
C32 (control)	Raven	50	22
C32 (control)	Robin	66	22
C35 (control)	Raven	50	24

Note: Raven model parameters are shared between both model types. As such, rows denoted as "Robin" only refer to the Robin-based model parameters in Raven Robin models. The total number of Raven Robin calibration parameters is then the sum of two rows (e.g., for catchment C31 this would be 21+19=40).

3.3.2 Turkey Lakes Watershed Calibration

Ensembles of Raven and Raven Robin models were constructed for each catchment period combination (e.g., C31 pre-harvest) through uncertainty-based calibrations using a pseudo-likelihood function. The Dynamically Dimensioned Search Approximation of Uncertainty (DDSAU) algorithm was used to calibrate models to summertime flows (July to September), and was selected for its simultaneous optimization of both ensemble reliability and sharpness (Shafii et al., 2015). The DDSAU algorithm calibrates a model multiple times using the original DDS algorithm developed by (Tolson & Shoemaker, 2008) with the same objective function in each trial.

In this study, objective functions in calibration experiments are meant to train models to simulate summertime rainfall-runoff responses. There is no specific preference between fitting simulated data to runoff volumes, peak flow magnitudes, or low flow magnitudes. As such, the Kling Gupta Efficiency (Gupta et al., 2009) score was chosen as the objective function because it quantifies the fit between simulated and observed data using multiple metrics. These include the correlation, variability, and bias:

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2}$$

Equation 7

where r is the Pearson correlation coefficient between observed (obs) and simulated (sim) data, while σ and μ are their standard deviations and means, respectively. The bounds of the KGE score ranges from negative infinity to 1, with a perfect model achieving a KGE score of 1. As shown by Knoben et al. (2019), the KGE score achieved by taking the mean flow of the observation time series in a calibration period as a benchmark model is -0.41, which can serve as a preliminary scoring threshold for screening sub-par models. The actual threshold used in this study is more stringent however, and is targeted to day-of-year flows as discussed below.

Behavioral models are identified in this analysis using benchmark models. Here, behavioral models have KGE scores above the climatological reference KGE score: This site and period-specific score is obtained by a benchmark model constructed using the average observed day-of-year flows Q_i^c; it is a 1-year long time series that repeats n times – once for every year contained within the model evaluation period (Equation 8):

$$Q_i^c = \sum_{n=1}^{n} Q_i \ for\ i = 1 \dots 365$$

Equation 8

The ensembles of behavioral models retained are then used to explicitly address model parameter uncertainty during the change detection analysis. It should be noted that model-based change detection analyses are informed by model predictive uncertainty, which consists of uncertainty in both the model parameters and in the model structure (Chlumsky et al., 2021). However, the assessment of model structural uncertainty is not addressed in the scope of this study.

To ensure behavioral Raven Robin models can adequately simulate vegetation state variables, the KGE score was penalized according to the mean absolute error (MAE) of the model in simulating the mean summertime Leaf Area Index during the calibration period (Equation 9). The KGE score was reduced anytime the MAE exceeded 10% of the maximum summertime LAI (about 0.5) such that:

$$KGE_{adj} = KGE_Q^{summer} - 0.2 \cdot \max(|MAE_{LAI}^{summer} - 0.5|, 0\ )$$

Equation 9

Twelve calibration scenarios were considered given the combination of catchments (one clearcut and two controls), models (Raven and Raven Robin), and periods (pre-harvest and post-harvest). Calibration scenarios in catchment C31 (clearcut) are designed to evaluate the concurrent impacts of harvesting and climate change on streamflow, while experiments in control catchment C32 only assess climate change impacts. Catchment C35 was included as a secondary control to investigate the spatial variability of climate change across the Turkey Lakes watershed, but was only simulated using Raven as vegetation data were only available in catchments C31 and C32 to calibrate Robin parameters. As a result, this forms a total of 10 calibration scenarios and thus 10 behavioral model ensembles for analysis.

Each of the 10 model ensembles was constructed with an initial target of 40 behavioral models due to time and computational constraints. Raven and Robin models were given budgets of 500 and 1000 iterations per calibration trial respectively to identify a behavioral model solution: A behavioral model is identified if the model solution achieves a KGE score above the climatological reference benchmark (which is specific to the catchment and period).

The number of models contained within each model ensemble (herein referred to as "model ensemble size") was standardized to provide equal sample sizes during change detection analyses. Following each calibration scenario, a trade-off curve between the mean continuous ranked probability score (MCRPS) and the model ensemble size was constructed. The MCRPS takes the average of the instantaneous continuous ranked probability score (CRPS) of a model ensemble evaluated at individual time steps. The CRPS itself describes the reliability of a model ensemble, and is defined as the difference between the cumulative distribution of model ensemble predictions and the empirical cumulative distribution of observations (Hersbach, 2000; Matheson & Winkler, 1976). It is defined as:

$$CRPS(P, \hat{x}) = \int_{-\infty}^{\infty} [F(x < \hat{x}) - H(x - \hat{x})]^2 dx$$

Equation 10

where $F(x < \hat{x})$ is the empirical cumulative distribution for the probability that a simulated value x will be less than the observation $\hat{x}$ and $H(x - \hat{x})$ is a Heaviside function representing the empirical cumulative distribution of the observed data, where:

$$H(x - \hat{x}) = \begin{cases} 0 \; if \; x < \hat{x} \\ 1 \; if \; x \geq \hat{x} \end{cases}$$

As an evaluative metric, the CRPS shares the same units as the predicted variable x, and decreases in value as the ability of a model ensemble to replicate the observed distribution improves; a perfect ensemble achieves a CRPS of 0 when all members predict exactly the observed value.

Each trade-off curve roughly converges to a minimum MCRPS as the ensemble size increases, thus providing a useful indication of the minimum required ensemble size to effectively describe model uncertainty. The standardized model ensemble size in this analysis is determined as the number of models required for all model ensembles to converge on their respective minimum MCRPS value.

3.4 Hydrologic change detection and analysis

The section presents a generalized change detection analysis workflow (section 3.4.1), followed by two different change detection methods (sections 3.4.2 and 3.4.3) to be employed in the hydrologic change detection analysis.

3.4.1 Workflow overview

Hydrologic change detection analysis generally follows a four-step process as described below.

Step one defines the basic premise of the analysis. A disturbance is identified in conjunction with hydrologic phenomena that may be altered as a result, forming a hypothesis of disturbance effects on hydrology. Dedicated metrics and timescales are then selected for the analysis that ensues to determine whether this hypothesis should be accepted (e.g., significant hydrologic change has occurred) or rejected (e.g., it is unclear whether significant change has occurred). Selected metrics here include signatures (which summarize hydrologic information) and test statistics that quantify the significance of the differences between calculated signatures.

Step two inventories all the necessary information, assigns pairings, and selects techniques for comparing paired data. Data are first categorized by catchment, disturbance period (before or after), and source (from observations or model outputs). Next, they are paired to achieve specified evaluation objectives, such as assessing baseline model quality (see Table 2). Sources of uncertainty associated with the data are noted to inform the selection of the statistical test used in evaluating differences between the pair of data.

Step three processes all the data and executes the evaluations. When simulated data are desired, this step involves generating model ensembles: model ensembles are required to explicitly address uncertainty arising from model formulation and parameterization during development. This is achieved through calibration experiments, which are designed according to the chosen objectives from Table 2. Following calibration trials, model outputs are processed alongside the observed data to derive metrics for comparisons. Finally, the chosen statistical tests are conducted using the derived metrics.

Step four collects the results from these statistical tests for interpretation and discussion. A significance threshold is established for each metric/signature of interest to determine the presence of hydrologic change.

This four-step process allows for one hypothesis to be evaluated each time and must be repeated for each unique combination of metrics and timescale.

In the analysis presented in this thesis, hydrologic change detection with behavioral model ensembles follows experiment 6 in Table 2 and uses two complementary comparisons (thus, iterating

through the process twice): the first comparison is a virtual paired-catchment experiment, where model ensembles calibrated to different periods are compared via their distributions of simulated runoff ratios. These findings are supplemented by the second comparison, which compares model ensembles via calibrated parameter distributions. The details of both methods are described next in sections 3.4.2 and 3.4.3.

Table 2. Purposes of comparing various data pairings

Experiment no.	Data source 1	Data source 2	Catchment	Period	Purpose of comparison
1	Observed	Observed	same	different	Before-after disturbance effects
2	Observed	Observed	different	same	Space-for-time disturbance effects* (spatial correlations)
3	Model	Observed	same	same	Baseline model quality assessment
4	Model	Observed	same	different	Disturbance effects on model error
5	Model 1	Model 2	same	same	Effect of model parameterization and structure on model quality
6	Model 1	Model 2	same	different*	Assess presence of changes to system conditions over time**

*basis of Paired Catchment studies

**refers to different calibration periods. Outputs spanning the entire record from both models are compared here.

3.4.2 Hydrologic change detection using runoff ratio distributions

One of the contributions of this thesis is the use of runoff ratio distribution comparisons as a mechanism of change detection. This analysis uses a 7-day mean runoff ratio to assess clearcut harvesting and climate change impacts on the rainfall-runoff responses of headwater catchments during the leaf-on period of the Turkey Lakes watershed (July to September). The 7-day timescale was selected to account for both the travel time through the catchment and the non-dichotomous nature of rainfall events that typically span 2-3 days and often overlap with each other, while the streamflow response emerges at the outlet 1-2 days later.

Hypotheses of changes to runoff ratios

Given the shift to a much drier climate at the Turkey Lakes watershed following the 1997 harvesting treatments, the hypothesis for this analysis is that climate and clear-cut harvesting both have significant effects on the catchment rainfall-runoff response. In the case of control catchments 32 and 35 it is expected that decreased rainfall will not fully saturate the landscape. Considering increased soil moisture demand by the vegetation in response to the resulting water stress that this creates, the hypothesis for control catchments C32 and C35 is that the runoff will disproportionately decrease with precipitation and ultimately decrease weekly runoff ratios.

In the case of clearcut catchment C31, it is expected that the concurrent climate shift and harvesting will initially have counteracting effects on the rainfall-runoff response: lower precipitation under drier climate conditions should decrease runoff, but this decrease may be compensated temporarily by reductions in evaporative losses that occur when the canopy is removed during harvesting. However, as vegetation recovery takes place, the emerging water demand by plants for regrowth will then cause ET to surge, thus decreasing the runoff response over time as an additive effect on top of drier climate conditions.

The hypothesis for catchment C31 then is that runoff ratios will initially increase due to canopy losses after harvesting, but will quickly decrease once climate and regrowth effects take place. Consequently, the change in runoff ratios when compared between the pre-harvest period and the entire post-harvest period will thus exhibit a weak signal due to the effect of time averaging. The hope is that the model can still effectively detect the magnitude of change due to harvesting despite the change in climate signal, and provide additional information not available from analyzing observational data alone.

Generating runoff ratio distributions

In this analysis, the runoff ratio is defined as the total weekly streamflow divided by the total weekly precipitation. Runoff ratios are calculated using a 7-day moving window average. This is done by first calculating the runoff ratio seven times - once using each day of the week i as the start of the 7-day period. The seven values are then averaged to produce the final estimate of the weekly runoff ratio for a particular week via Equation 11.

$$RR = \frac{1}{7}\sum_{i=1}^{7}\frac{\sum_{j=1}^{7} Q_{i,j}}{\sum_{j=1}^{7} P_{i,j}}$$

Equation 11

This additional aggregation produces runoff ratios that are centered around time steps seven days apart from each other to mitigate autocorrelation errors that would otherwise occur when sampling at a daily timescale. However, doing so does not fully render the samples as independent from each other.

At every weekly timestep, this experiment compares cumulative distributions of simulated weekly runoff ratios between behavioral model ensembles that are calibrated to different periods of data (pre-harvest or post-harvest). The simulated weekly runoff ratio is calculated using the same methodology as above, with both streamflow and precipitation values being simulated by the model (observed precipitation data is subject to gauge bias corrections, which is applied using calibrated model parameters for rainfall and snowfall respectively). Comparisons are made using the Kolmogorov-Smirnov test at every time step with a 5% significance level. Statistically significant differences here suggest that changes have occurred to the physical system that cannot be explained by model parameter uncertainty, which is one component of the model predictive uncertainty (model structural uncertainty is another component that is not assessed within the scope of this thesis).

3.4.3 Hydrologic change detection using model parameter distributions

Change detection results from comparing runoff ratio distributions will be supplemented by the second step of this analysis: comparisons of model parameter distributions stemming from model ensembles calibrated to different periods within each catchment. In this study, 5 sets of comparisons are made – each uses a pair of distinct model ensembles previously obtained from the 10 calibration scenarios described in section 3.3.2. Changes in parameter distributions across calibration periods for

a given catchment model ensemble may point to shifts in underlying runoff generation mechanisms over time.

This analysis acknowledges the presence of model equifinality and evaluates model ensembles as a whole, accepting multiple parameter sets as plausible descriptions of the real system. Collectively, the plausible values obtained through model calibration trials are assumed to form an unknown distribution for each parameter. In comparing two model ensembles, statistically significant differences in their parameter distributions suggest differences in the way certain processes influence the overall hydrologic behavior of the system. Conversely, given no significant differences in watershed function properties, we expect the parameter distributions from the model ensembles to be more or less the same between calibration periods. Differences between parameter distributions are evaluated in this analysis using the two-sample Kolmogorov-Smirnov test with a 5% significance level.

The hypothesis for both control catchments is that dry climate conditions will cause model calibrations to identify parameter sets that allow for more moisture retention in the landscape. The required shifts in parameter distributions to achieve this for both model ensembles are expected to be less prevalent in the case of catchment C31, owing to the initial compensatory harvesting effect previously described.

Chapter 4
Results and Discussion

This chapter is organized by thesis objectives. Section 4.1 presents results and discusses the ability of the calibrated process-based models in simulating hydrologic change within headwater catchments of the Turkey Lakes watershed. Section 4.2 discusses the results of the multiple hydrologic change detection analyses.

4.1 Model performance in simulating hydrologic change

This section assesses the ability of the calibrated models in simulating hydrologic change within 3 headwater catchments of the Turkey Lakes watershed, as needed to satisfy research objective 1. At first, the performance of individual models is evaluated in section 4.1.1, followed by an assessment of model ensemble performance in section 4.1.2.

4.1.1 Performance of individual models

Baseline model quality checks were conducted by comparing the daily Kling Gupta Efficiency (KGE) score for stream discharge during the summer period against the site- and period-specific climatological reference KGE score ($KGE_{clim,ref}$), as summarized in Table 3. The climatological reference KGE score used here is specific to both catchment and period, and provides a more stringent benchmark than the mean flow benchmark of -0.41 (Knoben et al., 2019). Thus, for every calibration scenario the goal is to reach a model solution that achieves a KGE score falling within the range of $KGE_{clim,ref} < KGE \leq 1$, where a KGE score of 1 represents a model that perfectly fits the observed data.

In each calibration period for all catchments, both models achieved KGE scores considerably higher than the corresponding climatological reference KGE after calibration (Figure 4): median KGE scores ranged between 0.72 and 0.81 for catchment C31 models, 0.70 and 0.86 for catchment C32 models, and 0.73 and 0.83 for catchment C35 models.

However, in almost all cases, the model performance significantly deteriorated when simulating the evaluation period. At a model ensemble level, this is captured by significant differences in the median KGE between periods based on two-sample Wilcoxon signed rank tests with a 5% significance level.

Table 3. Climatological reference KGE scores of summertime flows by catchment and period

Catchment	Pre-harvest	Post-harvest
C31 (clearcut)	0.349	0.291
C32 (control)	0.353	0.256
C35 (control)	0.386	0.263

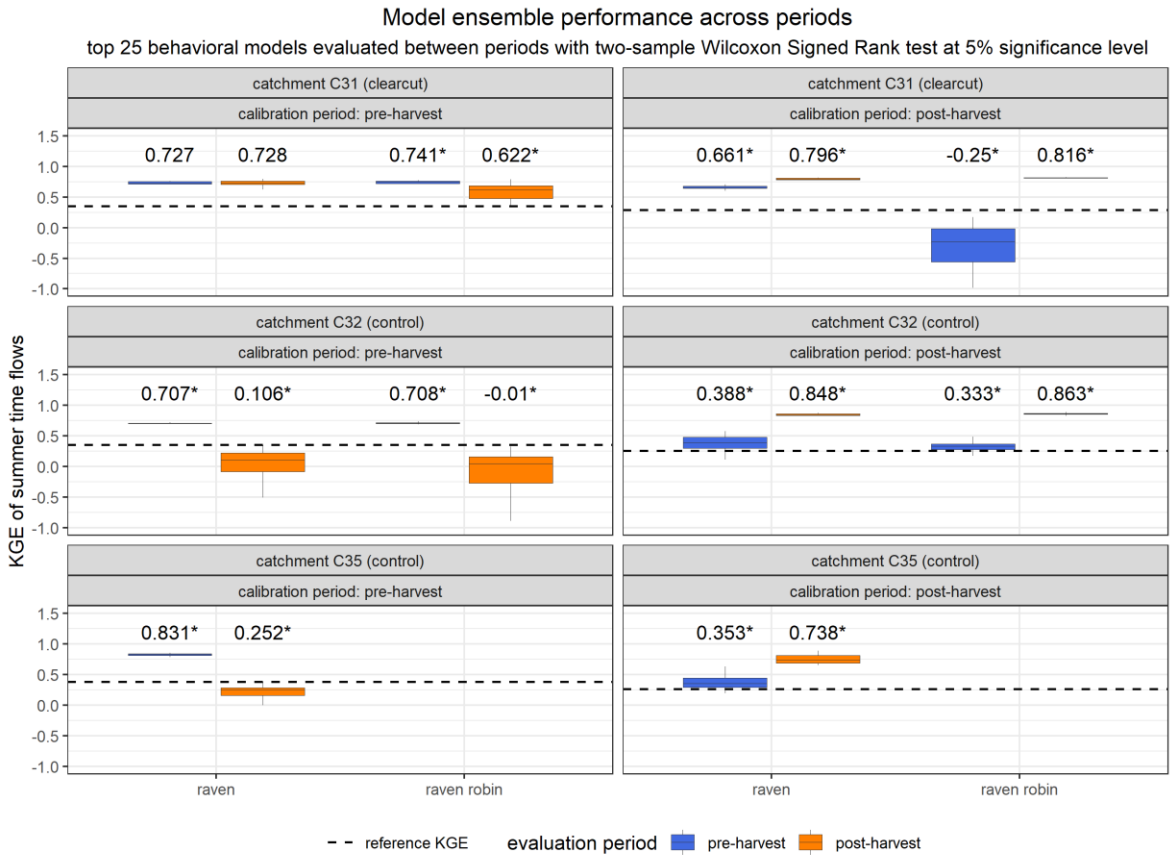

Figure 4. Distributions of KGE scores for model ensembles in pre-harvest and post-harvest periods

The degraded model performance at both control catchments C32 and C35 suggests that conditions have changed over time. This suggests that even in the absence of harvesting treatments, control catchments are exhibiting evolving hydrologic responses to rainfall events at the Turkey Lakes watershed. Based on previous studies, these changes in background conditions are largely owing to trends in climate: Buttle et al. (2018) found that the post-harvest climate conditions were warmer and drier than the pre-harvest period across the watershed. Specifically, through Mann-Kendall tests (at a 5% significance level with adjustments for autocorrelation) they identified a significant decrease in precipitation from 1981 to 2011 and an average increase of 0.06°C in mean annual temperatures.

Given these significant shifts in climate conditions, it is expected that calibrated models perform poorly in simulating streamflow during their evaluation periods. This is because models are trained to reproduce observed streamflow data under a particular range of climate conditions (which are used as inputs) during calibration. Thus, when climate inputs fall outside the range of conditions encountered during calibration, models become less effective at simulating the unprecedented streamflow responses observed.

An exception to the decline in model performance were Raven models calibrated to pre-harvest conditions in catchment C31. Pre-harvest Raven model ensembles consistently achieved KGE scores around 0.72 across both pre-harvest and post-harvest periods. Interestingly, the opposite calibration scenario does not hold true: Raven models calibrated to post-harvest conditions achieved higher scores (median KGE: 0.79) in the post-harvest period but deteriorated in performance when used to simulate pre-harvest conditions (median KGE: 0.66).

This suggests that calibrating Raven models to pre-harvest streamflow is sufficient for achieving model solutions that consistently fit streamflow in both pre-harvest and post-harvest periods. However, such models do not have representation for the clear-cut treatment applied to the catchment and thus be subject to model equifinality concerns when simulating post-harvest streamflow (i.e., they are likely getting the right fit to observed data for the wrong reasons).

4.1.2 Performance of model ensembles

Ensembles of models with distinct parameter sets were used to estimate model parameter uncertainty. Ensemble members were generated by calibrating each model multiple times using the DDS-AU algorithm (Shafii et al., 2015) and then retained or discarded from the ensemble based upon performance against the climatological reference benchmark. A calibration trial yields a behavioral

model if the model solution achieves a Kling Gupta Efficiency (KGE) score above the climatological reference benchmark (which is specific to the catchment and period as summarized in Table 3). While Raven model ensembles successfully identified behavioral models in every trial, Raven Robin models were unable to generate a sufficient number of behavioral models. Thus, an additional 40 trials were subsequently allocated for the Raven Robin model ensembles, which brought the minimum number of behavioral models identified up to 27 (Table 4).

Table 4. Number of behavioral models identified from Turkey Lakes watershed model calibration scenarios

Calibration period	Catchment	Raven	Raven Robin
Pre-harvest	C31 (clearcut)	40	36
Post-harvest	C31 (clearcut)	40	40
Pre-harvest	C32 (control)	40	36
Post-harvest	C32 (control)	40	27
Pre-harvest	C35 (control)	40	40

Based on the trade-off between the MCRPS and the model ensemble size for the 10 calibration scenarios (Figure 5), a sample size of 25 behavioral models was adopted for this analysis to ensure all ensembles achieved their minimum MCRPS. Given the small sample size, the results of the subsequent analysis should not be treated as statistically rigorous but may still provide useful insight into the influence of model uncertainty on change analysis.

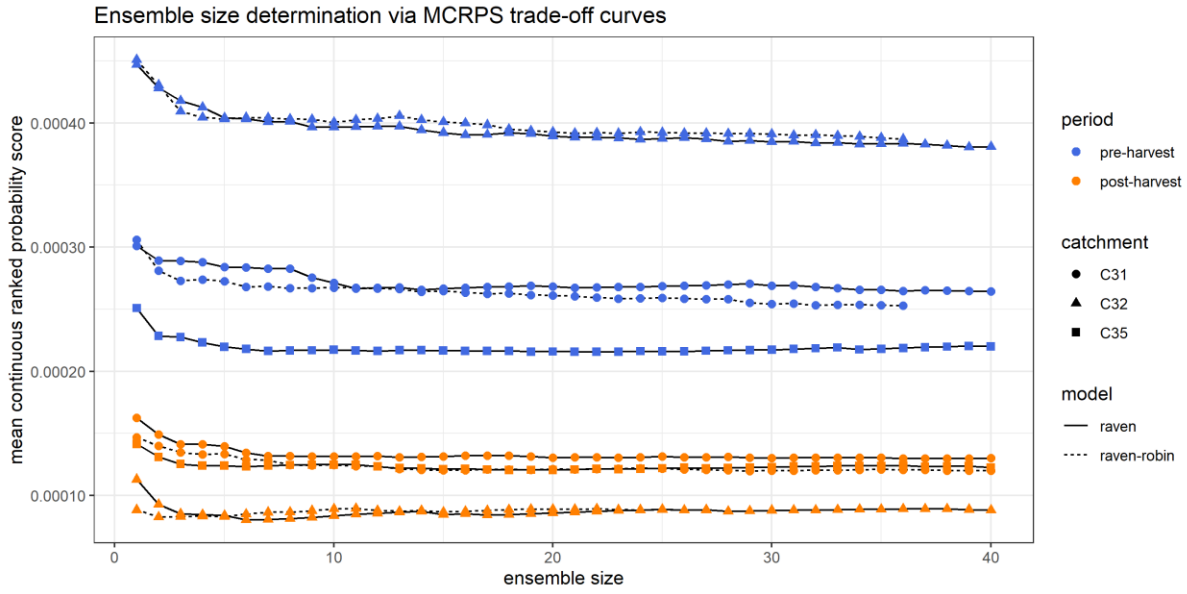

Figure 5. Turkey Lakes watershed model ensemble development: trade-off curves between mean continuous ranked probability score (MCRPS) and number of behavioral models retained in the ensemble (ensemble size)

4.2 Hydrologic change detection with multiple approaches

This section discusses the results of the hydrologic change detection analyses conducted using weekly runoff ratio distributions (section 4.2.1) and model parameter distributions (section 4.2.2) to fulfill research objective 2.

4.2.1 Change detection using weekly runoff ratio distributions

This section presents results from the hydrologic change detection analysis conducted using weekly runoff ratio distributions derived first from observation data (section 4.2.1.1), then from model simulations (section 4.2.1.2).

4.2.1.1 Change detection using only observation data

Here we assess the viability of weekly runoff ratios as a metric for hydrologic change detection at the headwater catchments of the Turkey Lakes watershed. This is done by first evaluating their ability to capture change based solely on observation data: weekly runoff ratios were derived from daily streamflow and daily total precipitation to detect hydrologic changes in the summer months (July to September) over time. Moving from pre-harvest to post-harvest periods, the weekly runoff ratios are assumed to be impacted by only two disturbances: climate variation and harvesting. It should be noted that these basins exhibit extremely small runoff ratios, with less than 1% of annual precipitation running off. This was found to be a continuous challenge during the analysis performed below.

Changes to control catchments C32 and C35

The watershed becomes warmer and drier over time, and this contributes to an overall decrease in runoff ratios in the post-harvest period. However, this decrease is more significant in the control catchments, where there is no change in the vegetation to counteract its effects. Visually, this is evident in Figure 6, where median summer runoff ratios for control catchments C32 and C35 occur between 10^{-5} and 10^{-4} in the pre-harvest period before dropping a few orders of magnitude to as low as 10^{-7} in the post-harvest period.

Decreased post-harvest runoff ratios in this watershed stem from disproportionate reductions of rainfall and runoff, which may be explained by threshold runoff generation behavior. The observed runoff measured at the catchment outlet mostly represents surface flow passing through the v-notch weirs. As such, they largely represent saturation excess and infiltration excess overland flows, which have been shown to emerge from a catchment only when precipitation amounts or soil moisture levels exceed critical threshold values (Mosley, 1979; Tani, 1997; Tromp-van Meerveld & McDonnell, 2006; Uchida et al., 2005; Whipkey, 1965). Given that infiltration rates of precipitation into the subsurface are non-linear, soil moisture levels – and thus also runoff – do not scale linearly with precipitation depth.

The variable contributing area model proposed by Hewlett & Hibbert (1967) also supplements the notion of threshold runoff generation behavior. Precipitation amounts affect the extent of saturated areas forming after a storm, which may become hydrologically connected to promote both subsurface flow and saturation excess overland flow that increases runoff in non-linear proportions. In their analysis of runoff responses to rainfall-dominated events, Buttle et al. (2019) used piecewise

regression to show that quickflow to precipitation ratios scaled linearly with precipitation at different rates below and above threshold precipitation amounts.

Changes to runoff ratios in clearcut catchment C31

In contrast to the control catchments that exhibit a decrease in the runoff ratio, catchment C31 runoff ratios do not initially change from pre-harvest conditions (Figure 6). This is owing to the canopy and vegetation losses, which drastically reduce evapotranspiration to allow for more streamflow generation. Evapotranspiration eventually increases again as vegetation regrowth occurs, effectively decreasing streamflow and the runoff ratio as a result. Despite being short-lived, the increases in runoff ratios due to harvesting combat the decreases caused by drying climate conditions (as seen in the control catchments) during the first half of the post-harvest period (water years 1998 to 2003). Ultimately this does not create a drastic observable shift in runoff ratios relative to pre-harvest magnitudes, but rather the lack thereof.

Dry climate conditions eventually prevail in water years 2004 to 2006, dominating the runoff response across all three catchments and generating virtually zero streamflow at the outlets. This results in runoff ratio values that are so small in magnitude that they are effectively zero, leading to the absence of boxplots in water years 2004 and 2005 for control catchments C32 and C35 (the y-axis is presented on a logarithmic scale). For small headwater catchments draining 4 to 7 ha areas under such dry conditions, this dryness challenges the viability of the proposed change detection analysis using distributions of weekly runoff ratios. A similar observation was made by Buttle et al. (2019), where they reported a decrease in the number of quickflow responses to post-harvest period rainfall events.

The shifts in runoff ratios can be further quantified by comparing weekly runoff ratio distributions between periods for each catchment (Figure 7): moving from the pre-harvest to post-harvest period, catchment C32 (primary control) saw the largest decrease in the median weekly runoff ratio (10^{-4} to $10^{-5.3}$), followed by catchment C35 (secondary control, $10^{-4.3}$ to just under 10^{-5}), and catchment C31 (clearcut treatment, $10^{-4.2}$ to $10^{-4.8}$).

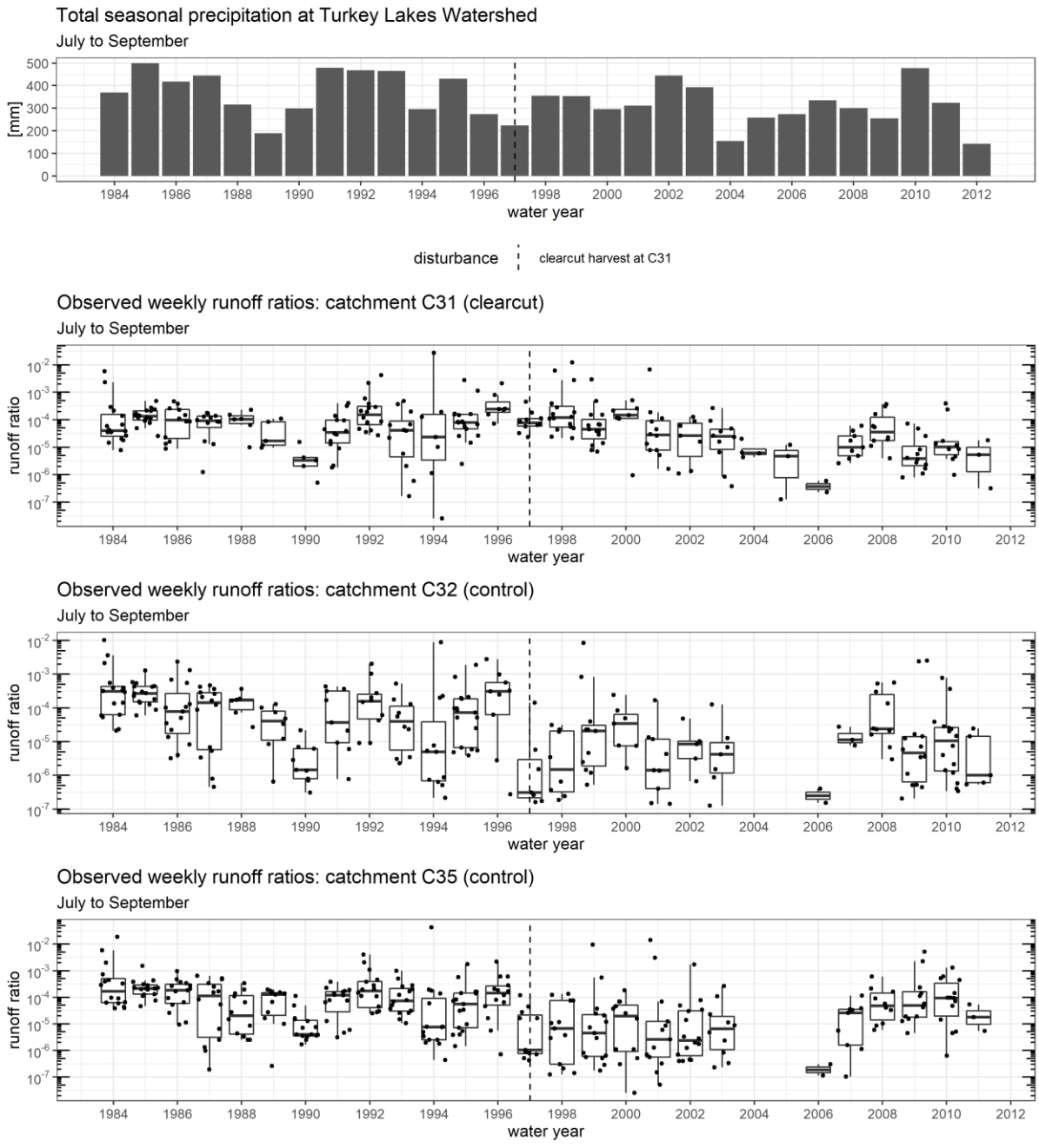

Figure 6. Time series of observed weekly runoff ratios at the Turkey Lakes watershed, overlain by total precipitation for the summer months (July to September)

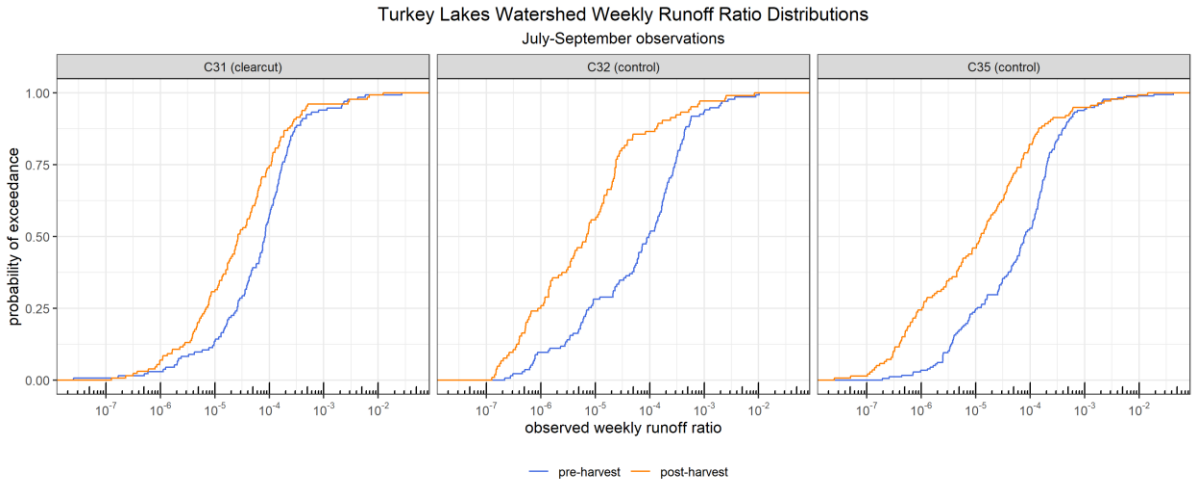

Figure 7. Cumulative distributions of observed weekly runoff ratios between pre-harvest (1984-1997) and post-harvest (1997 - 2012) periods for the clearcut and control catchments at the Turkey Lakes Watershed

4.2.1.2 Change Detection using simulated data

Data filtering: a prerequisite

From Figure 7 it is evident that runoff ratios above the upper 90th percentiles of the empirical cumulative distribution functions (CDFs) remain relatively unchanged between periods for all three catchments. The majority of differences in CDFs occur instead between the upper and lower quantiles, where runoff ratio magnitudes are lower than 10^{-3} and become both difficult to physically interpret (a runoff ratio of 10^{-3} means that 0.1% of rainfall turned into catchment runoff) and are subject to high potential for measurement errors that will impact the analysis.

To ensure the results from this change detection analysis (which compares simulated runoff ratios) are physically interpretable, a minimum runoff ratio threshold was introduced to split the data into two groups. Group 1 (higher flows) contains data from timesteps during which the observed weekly runoff ratio is greater than the minimum threshold. These data were used in the change detection analysis proposed in section 3.4.2. Group 2 (negligible flows) retains data from all other remaining timesteps where observed runoff ratios fall below the minimum threshold and grossly represent low-flow events (i.e., dry conditions). These data are used later to assess model performance when simulating low-flow runoff responses. In this approach, the success rate of model ensembles (out of

25 members) in correctly reporting low-flow conditions when they occur is assessed on a water year basis.

Several threshold values were tested to determine an appropriate minimum runoff ratio for subsetting the data. In each test, empirical CDFs of runoff ratios greater than the threshold were formed for comparison across periods in each catchment via the Kolmogorov-Smirnov test with a 5% significance level. The results are summarized in Table 5, which indicate that differences in runoff ratio distributions between pre-harvest and post-harvest periods are only significant for threshold values less than 10^{-4}. As such, this was selected as the minimum runoff ratio threshold to reach a compromise between maintaining sample sizes and interpretability of the values retained in group 1.

Although no significant differences are detected with this threshold (10^{-4}) when comparing the observation data, the hope is that comparing simulated distributions will unveil significant differences and detect hydrologic changes to the system that cannot be readily assessed from the data alone.

Table 5. Comparison of observed weekly runoff ratios between pre-harvest and post-harvesting periods above various minimum threshold values

Minimum runoff ratio Threshold	Catchment	Pre-harvest sample size	Post-harvest sample size	p-value
0 (none)	C31 (clearcut)	190	218	0.00011*
	C32 (control)	190	218	8.4×10^{-13}*
	C35 (control)	190	218	0.0*
10^{-5}	C31 (clearcut)	115	90	0.014*
	C32 (control)	97	46	2.5×10^{-8}*
	C35 (control)	134	75	3.1×10^{-4}*
10^{-4}	C31 (clearcut)	57	33	0.97
	C32 (control)	66	14	0.42
	C35 (control)	84	25	0.22
10^{-3}	C31 (clearcut)	8	5	0.18
	C32 (control)	9	3	0.49
	C35 (control)	10	7	0.81

Significant differences are annotated with an asterisk (*), whereby the p-values are less than 0.05 for the Kolmogorov-Smirnov test.

Change detection on data after filtering

At every time step retained in group 1, simulated weekly runoff ratio distributions were compared using the Kolmogorov-Smirnov (KS) test with a 5% significance level. Each distribution corresponds to a model ensemble calibrated to either pre-harvest or post-harvest conditions. Comparisons were

made between models of the same type (e.g., Raven vs Raven or Raven Robin vs Raven Robin), calibrated to different periods for each catchment.

Simplifying assumptions were made about the model error distributions. KS tests were conducted on a week-by-week basis to compare simulated runoff ratio distributions for time steps retained in data group 1. For a single week, the simulated weekly runoff ratios are subject to model error derived upon the same observed value (red dashed line in Figure 8 and Figure 9). The error distributions here are assumed to 1) contain independent samples (each model ensemble member provides only one prediction as opposed to several that may be serially correlated), 2) follow an unknown distribution (hence the use of the non-parametric KS test) and 3) possess constant variance (homoscedasticity given that simulated values represent a single runoff response as opposed to many scenarios across multiple timesteps with different flow magnitudes).

Each KS test yields one of two possible outcomes: simulates runoff ratio distributions are either significantly different or indifferent from each other as illustrated by examples in Figure 8 and Figure 9 respectively. Through the different examples depicted by Figure 9 it should be noted that similarities do not speak to model accuracy - both models can produce reasonable accurate predictions (week of 2001-07-10) or, in some other cases, exhibit equally poor performance (week of 1995-08-20).

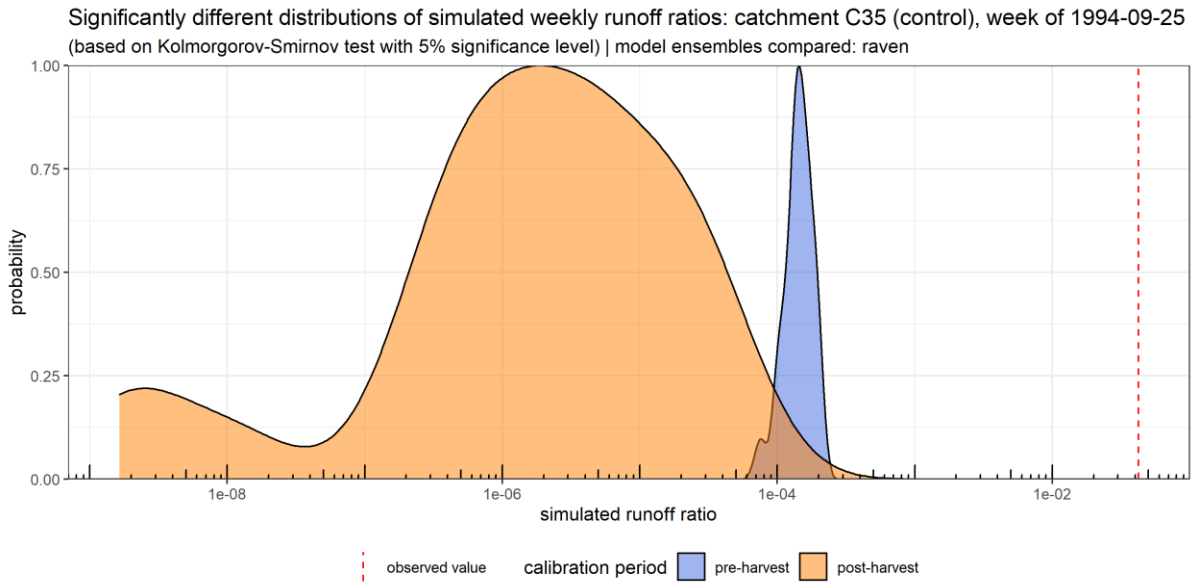

Figure 8. Example of significantly different distributions of simulated weekly runoff ratios at a given time step

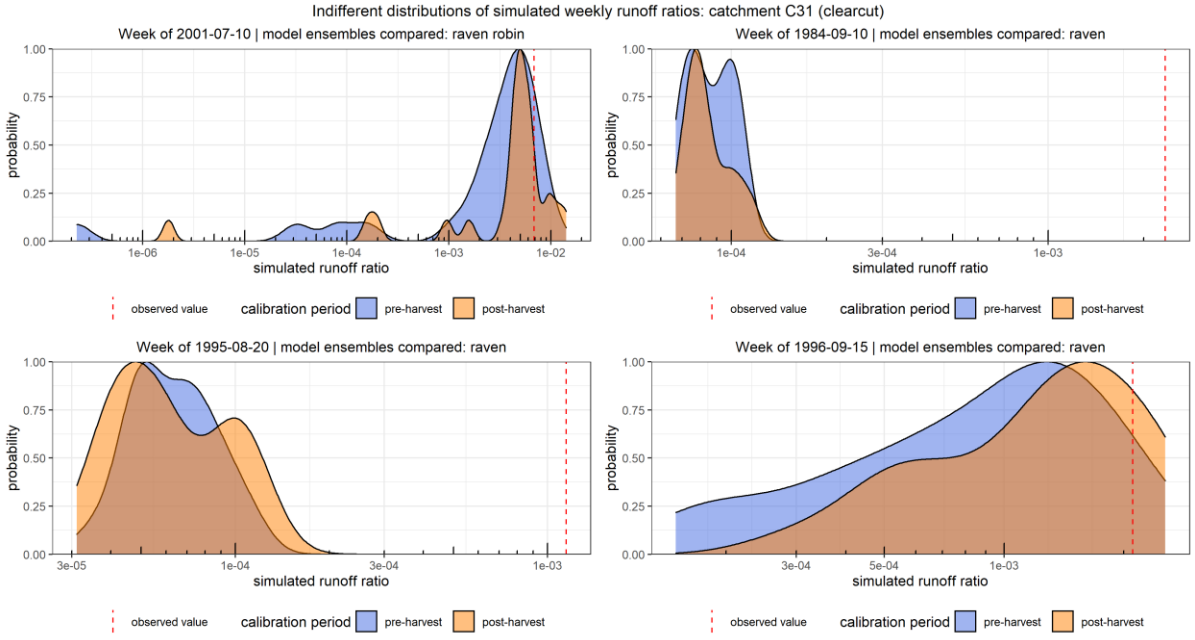

Figure 9. Examples of similar distributions of simulated weekly runoff ratios at a given time step

The four instances where KS tests reported insignificant differences all occurred in catchment C31 (Figure 9). One of these cases was for Raven Robin models operating in the post-harvest period while the other three cases were for Raven models simulating pre-harvest runoff events. In the case of the Raven Robin models, both pre-harvest and post-harvest calibrations lead to models that produced reasonable estimates of the observed runoff ratio: the central tendency of both ensembles of model predictions were within the same order of magnitude as the observed value. Conversely in the case of the Raven models, both calibrations lead to ensembles of predictions with errors that are large in magnitude and in some instances also highly variable (e.g., week of 1996-09-15).

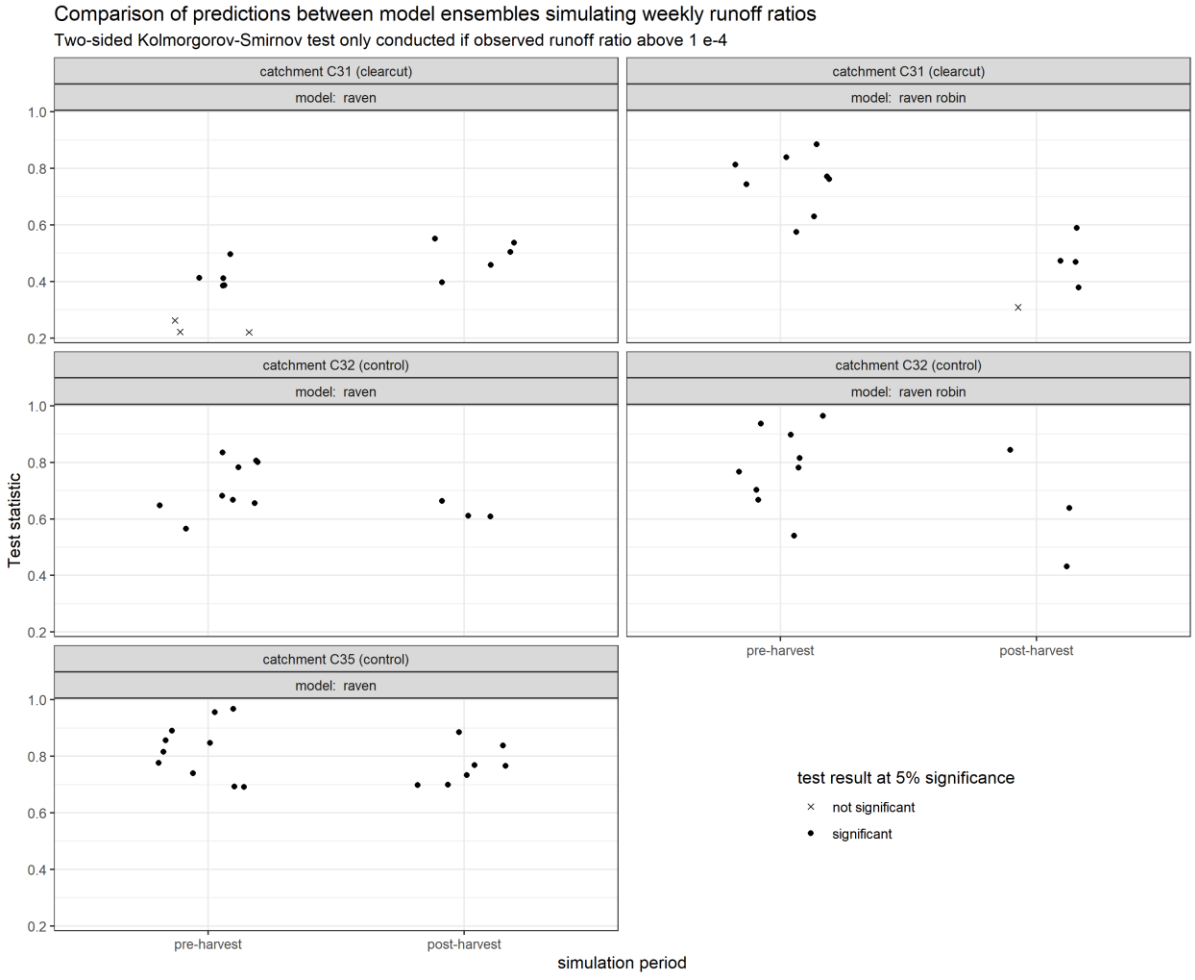

Figure 10. Two-sided Kolmogorov-Smirnov test statistics comparing adjusted distributions of weekly runoff ratios, simulated by model ensembles calibration to different periods

Aside from these four cases, KS test results largely suggest that pre-harvest and post-harvest periods have distinct hydrologic conditions in all three catchments. The majority of simulated runoff ratio distributions compared were found to be significantly different (Figure 10). This indicates that, when calibrated to streamflow from different periods, models produce significantly different runoff responses to large rainfall events that elicit observed runoff ratios of 10^{-4} or greater. These distinct conditions stem from unique combinations of climate variation and harvesting effects, which ultimately influence the optimal parameter set identified during calibration. Calibrated parameter sets can drive models to utilize their available runoff generation mechanisms to different extents. This will be further explored in section 4.2.2.

4.2.1.3 Comparison of low-flow events and dry condition frequencies

In this section we address data group 2, which contains runoff ratios at time steps where catchments exhibited dry conditions. These data were previously filtered out from the change detection analysis with runoff ratios that was originally proposed. Given that these low magnitude runoff ratios are difficult to physically interpret, we opt for a simpler assessment here by commenting on the ability of Raven and Raven Robin models to simulate dry conditions when they occur (as reported by observed data).

In this analysis, the number of weeks where dry conditions are observed is tallied for the summer (July to September) of each water year. A high frequency for a given water year suggests a dry summer with very little runoff while low frequencies represent wet summers exhibiting runoff ratios greater than 10^{-4}. For each dry week observed, a model successfully reports the dry condition if it simulates a runoff ratio less than the minimum threshold (10^{-4}). The percentage of models that achieve this across the entire model ensemble then defines the model success rate. Each model success rate belongs to a particular week, and the rates are summarized by water year as boxplots (Figure 11).

The calibration period had a large impact on the ability of models to simulate dry conditions during the summer months. According to observed data, summer months typically exhibit 10 to 15 dry weeks with very low flows (triangles on figure 11), with more wet (low frequency) years occurring in the pre-harvest period. In both control catchments C32 and C35, Raven and Raven Robin models calibrated to pre-harvest conditions had variable success in simulating these dry weeks throughout the entire record, occasionally achieving success rates lower than 25% (e.g., water year 1995). This is

owing to the wetter and cooler conditions in the Turkey Lakes watershed during the pre-harvest period, where dry weeks are less typical of an occurrence than in the post-harvest period. Thus, models calibrated to the pre-harvest period are less effective at correctly responding to climate inputs that create such dry conditions.

Less of this is true for Raven models in clearcut catchment C31, who had comparable success rates in simulating dry conditions regardless of calibration period. This is somewhat consistent with the previous results discussed in section 4.1.1, where pre-harvest Raven models were found to perform equally well across both periods; Post-harvest Raven models achieved only slightly lower success rates throughout the entire simulation period.

For Raven Robin models in clearcut catchment C31, calibrating to additional vegetation growth (Robin) parameters had a large impact on the model ability to simulate dry conditions. Post-harvest calibrations of Robin parameters aimed to represent evapotranspiration (ET) dynamics of the regrowth forest, which initially started with zero canopy coverage before recovering in the subsequent years after the clearcut harvesting treatment. Although the ET does gradually increase, full recovery of the forest to pre-harvest conditions by the end of the record (2012) remains inconclusive (Buttle et al., 2019). As such post-harvest Raven Robin models are expected to under-estimate ET losses during the pre-harvest period, thereby producing more streamflow and thus achieving lower success rates in simulating low-flow conditions (dry weeks).

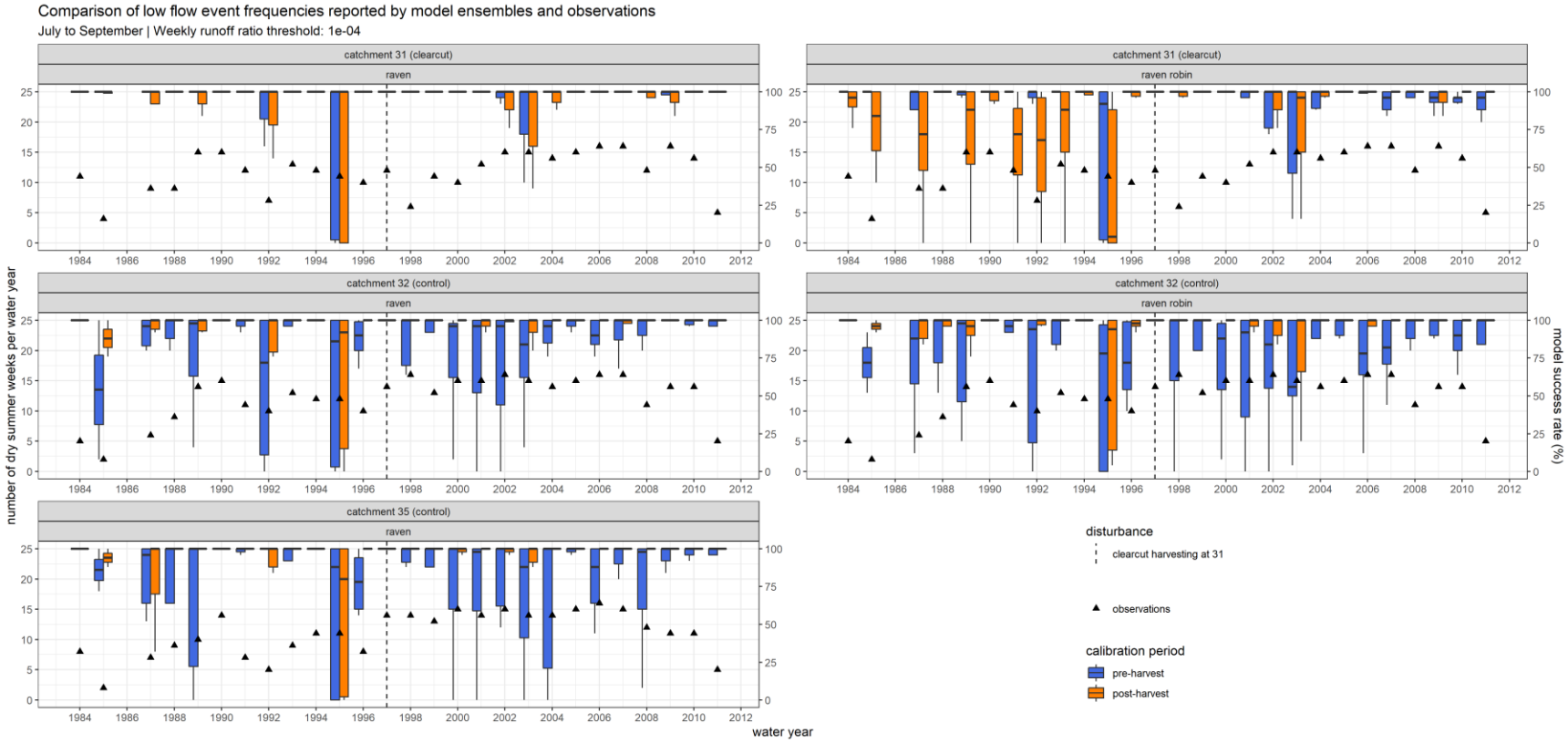

Figure 11. Model success rates in simulating low-flow events (dry conditions)

4.2.2 Change detection using parameter distributions

This chapter discusses change detection results from comparing parameter distributions across calibration periods. It is organized into two subsections: section 4.2.2.1 first defines each of the model parameters, of which significant distribution differences between model ensembles will be then discussed in section 4.2.2.2.

4.2.2.1 Overview of highlighted parameters

Calibration parameters were sorted into 4 groups based on the types of processes they describe: soil (21), snow (8), climate forcing (3), and vegetation (25). Parameter distributions were constructed for each behavioral model ensemble using parameter sets from their 25 members. Distributions were compared between models of the same type, trained to different periods in the same catchment.

A few calibration parameters were excluded from this analysis. Given that this change detection analysis focuses on the rainfall-dominated summer period of the water year (July to September), significant differences in snow-related calibration parameter distributions were excluded from discussion as they are expected to have no interpretable effect on rainfall-runoff responses during this time. The same applies to the only significant difference detected in the climate-related parameters, the snowfall gauge correction factor. The remaining parameters relating to soil and vegetation processes are first described here if differences in their distributions between calibration periods are significant for at least one model. Directions of the distribution shifts are then interpreted for each catchment.

Soils

Soil parameters in both models are stored in Raven. The parameters with significant differences in their distributions can be classified as relating to either water storage or subsurface flow. Storage-related parameters include the layer depth in meters, which dictates the absolute amount of available storage, and field capacity, which determines the level of soil saturation below which pore water becomes inaccessible for evaporation or baseflow.

Parameters related to subsurface flow govern baseflow rates via Equation 2, and include the saturated hydraulic conductivity in mm/d and the power law coefficient. An increase in this coefficient alters the curvature of the power-law relation, leading to lower baseflow rates for the majority of soil saturation states. Conversely, a decrease causes the equation to gradually mimic a linear relationship that scales baseflow rates to higher magnitudes at lower saturation states.

Vegetation

Vegetation parameters exist in both Raven and Robin model structures. However, significant distribution differences only emerged for some Robin parameters, which are further classified as LAI- or non-LAI related. The purpose of implementing Raven Robin in this analysis is to assess whether adding representation for vegetation growth dynamics has an impact on the model ability to simulate hydrologic change after a forest disturbance. As such, given that Robin augments Raven through the LAI state variable, only the LAI-related Robin parameters are discussed here. LAI-related Robin parameters that significantly changed between calibration periods include the minimum and maximum LAI, maximum evapotranspiration coefficient ($kc_{max,j}$), and vegetation stress factor (sf_j).

The maximum and minimum LAI form the upper and lower bounds between which the leaf area index is allowed to oscillate throughout the seasons. The LAI is used to account for the plant ET demand, and together with the coefficient $kc_{i,j}$ adjusts the potential evapotranspiration (PET) estimated by the Priestley-Taylor equation in Raven via Equation 1. This adjustment explicitly accounts for the plant ET demand not represented in the Priestley-Taylor equation, which is only dependent on temperature and radiation. Following LAI fluctuations across seasons, the coefficient $kc_{i,j}$ also oscillates between a maximum and minimum bound over time.

The vegetation stress factor is a species-dependent constant that characterizes the water uptake efficiency of the vegetation across all moisture conditions. Under water-limited conditions plants can reduce transpiration by closing their stomata, effectively retaining moisture for storage and repair. At the same time, the plant may increase its soil water uptake to meet its internal water deficit. Ranging between 0 and 1, the vegetation stress factor captures the average water uptake behavior when either of these phenomena are occurring and is inversely related to the degree of water stress: high water stress conditions are represented by low values, which consequently increase the uptake term in the expression for soil AET (third term in Equation 5). Although water uptake is responsive to surrounding conditions in reality, this parameter is implemented as a constant and calibrated to the entire range of soil moisture conditions.

4.2.2.2 Comparison of parameter distributions by catchment

Catchment C32 - primary control

In catchment C32, significant distribution differences between calibration periods were detected for both soil and vegetation parameters. Most of the changes in parameter values for both Raven and Raven Robin models shifted in the same direction. Moving from the pre-harvest period to the post-harvest period, both models saw significant increases in topsoil layer depth at the bottom of hillslopes and decreases in deep soil hydraulic conductivity.

Additional soil parameter distributions shifted in the same direction but were only significant for one of the two models (identified in parentheses): The topsoil layer experienced decreases in field capacity (Raven Robin) and increases in the PET correction factor which allows for more soil evaporation (Raven).

The power law coefficient governing the baseflow rate in Equation 2 also saw changes with mixed significance: coefficients decreased in the underlying shallow soil (Raven) but increased in the deep soil (Raven). Interestingly in the topsoil, shifts occurred in opposite directions: Raven models experienced a non-significant decrease while Raven Robin models exhibited a significant increase.

Both models suggest altered soil water dynamics in catchment C32. In Raven, a significant increase in both the depth and PET correction factor of the topsoil promotes more infiltration and near-surface evaporative losses. Infiltrated water is also routed to the bottom two soil layers, where baseflow rates have increased in the shallow soil (lower power coefficient) but decreased in the deep soil (higher power coefficient and lower hydraulic conductivity). Raven Robin models experienced the same parameter distribution shifts, with baseflow rates also decreasing in the topsoil (higher power coefficient).

Raven Robin models also exhibited shifts in vegetation parameter distributions. The minimum LAI increased in the post-harvest period under drier climate conditions, tightening the overall bounds within which LAI oscillates. Additionally, the vegetation stress factor significantly decreased when the model was calibrated to the post-harvest period, consequently allowing for more soil water uptake by plants for ET.

The vegetation parameter changes broadly signify an increase in plant water demand under drier climate conditions in the post-harvest period. This is underscored by the changes in soil parameter values, which drive both models to route more water into the subsurface for evaporative losses and baseflow, resulting in a dampened rainfall-runoff response.

Catchment C35 - secondary control

Catchment C35 flows were simulated with Raven to assess spatial effects of climate trends on control catchments between calibration periods. Only soil parameter distribution differences are discussed here given that no Raven Robin models containing vegetation growth parameters were run for this catchment.

The drier climate in the post-harvest period altered soil moisture dynamics in the same way as catchment C32. The topsoil experienced a significant increase in the PET correction factor and significant decreases in both the field capacity and the layer depth at the mid-slope position. In deep soils, the baseflow power coefficient significantly increased to reduce baseflow rates at lower saturation states.

Much of the significant parameter distribution shifts in catchment C35 models were found to be similar to those occurring in catchment C32 models. Despite the reduction in total available storage for topsoil in the mid-slope position, drier climate conditions consistently promote more infiltration and evaporative losses in the topsoil with concurrently slower baseflow rates in deep soils. Collectively, this again lowers the rainfall-runoff response of the catchment.

Catchment C31 - clearcut treatment

In contrast to the two control catchments, catchment C31 did not report as many significant changes to Raven parameter distributions. Only 3 parameters relating to soil storage experienced significant changes to their distributions across calibration periods (models reporting significant changes are once again notated in parentheses): a higher subsurface storage capacity was achieved at the bottom-slope position through increases in the layer thicknesses of topsoil (Raven Robin) and shallow soil (both). This is further extended by significant increases in the shallow soil porosity (Raven).

LAI-related parameter distributions in Raven Robin saw double the number of significant changes compared to catchment C32. Collectively, they impact soil AET estimates via the first and third terms in Equation 5.

The first term of Equation 5 is affected by parameters that influence the calculation of the PET correction coefficient $kc_{i,j}$: Both the minimum and maximum LAI values increased after harvesting, effectively shifting the bounds within which LAI can fluctuate towards a higher magnitude. The maximum value for the PET correction coefficient also significantly decreased, thus tightening the bounds within which $kc_{i,j}$ itself can oscillate. Together, these parameter distribution shifts create a new dynamic for maximum plant ET demand of the regrowth forest.

The third term of Equation 5 is affected by the vegetation stress factor. Just like in catchment C32, this parameter significantly decreased for catchment C31 Raven Robin models when comparing pre-harvest distributions to post-harvest distributions. This increases soil water uptake, suggesting that vegetation across both catchments want to access more soil moisture under drier climate conditions. This elevated water stress on the vegetation due to climate change is further exacerbated in catchment C31 by the additional soil water demand for growth as forest recovery takes place.

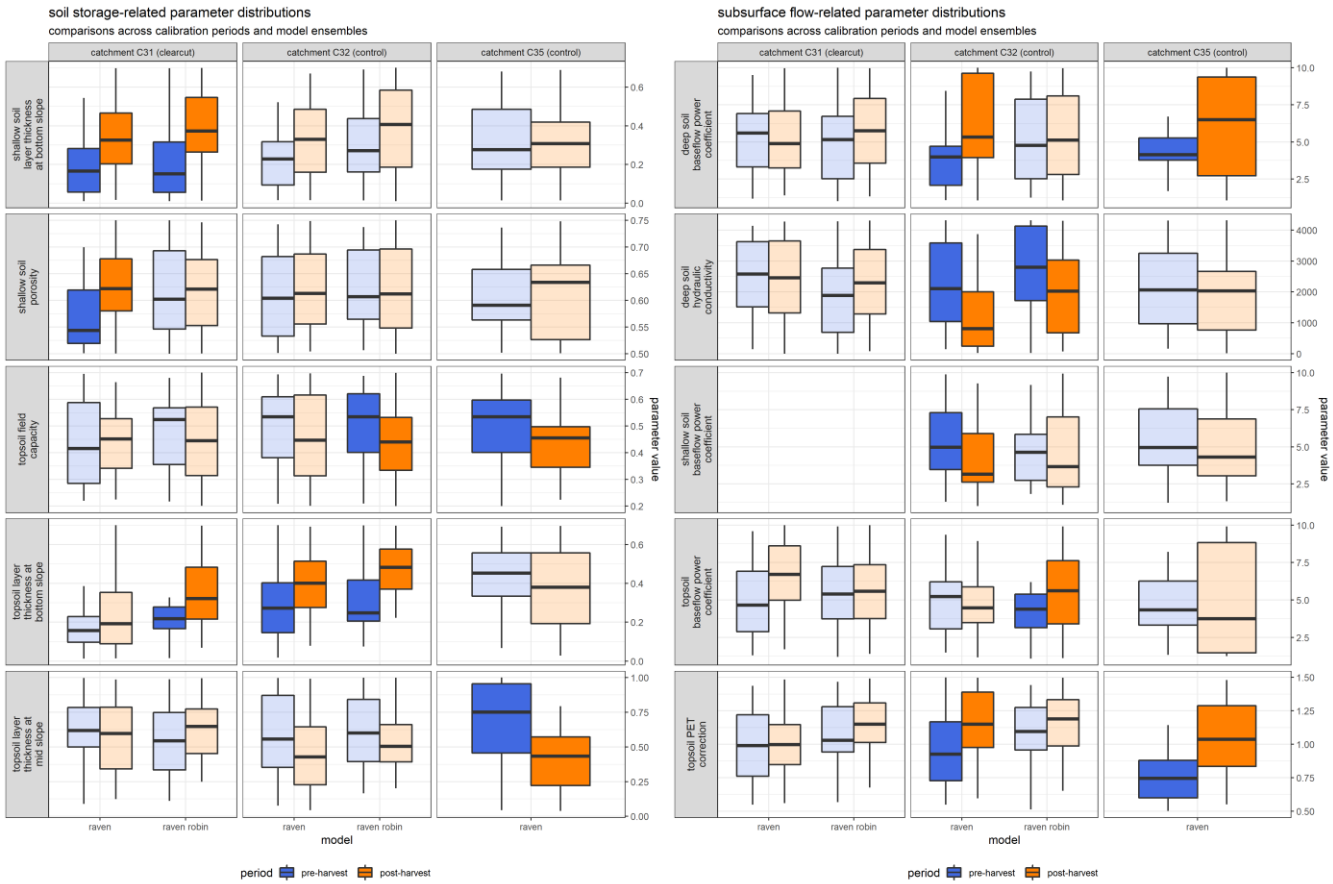

Figure 12. Soil parameter distributions across calibration periods, organized by those relating to soil water storage (left) and subsurface flow through soil layers (right).

Statistically significant shifts in distributions across ensembles (calibrated to different periods) based on the Kolmogorov-Smirnov test (5% significance level).

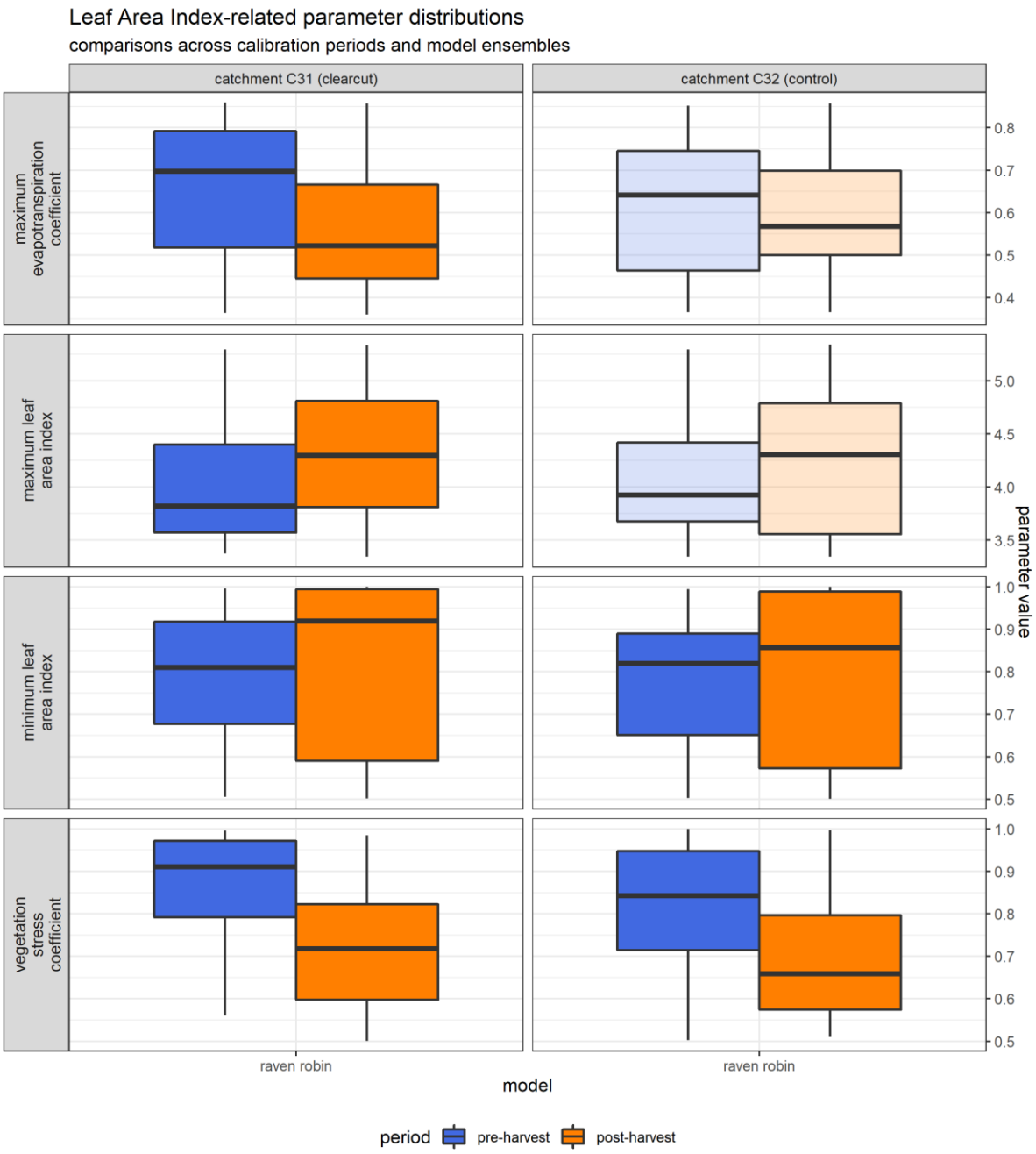

Figure 13. Leaf Area Index-related parameter distributions across calibration periods

Statistically significant shifts in distributions across ensembles (calibrated to different periods) based on the Kolmogorov-Smirnov test (5% significance level) are darkened, while non-significant shifts are illustrated with muted colors.

Chapter 5
Conclusions and future work

This thesis posed two research objectives relating to model-based hydrologic change detection analyses. The first objective was to assess the capability of hydrologic models and hydrologic-vegetation growth models in simulating streamflow patterns of forests following disturbances. The second objective sought to compare the utility of various hydrologic change detection methods in describing the hydrologic impacts of forest disturbances.

In this study, two process-based models were developed to simulate summertime streamflow at three forested headwater catchments of the Turkey Lakes watershed from 1981 to 2012. A custom hydrologic model was first developed in the Raven hydrologic modelling framework (Craig et al., 2020) (herein referred to as the "Raven model"). The Raven model was then coupled with the Robin(3PG) model developed by Han (2022) to create the "Raven Robin" model, which features vegetation growth routines. Robin(3PG) operates at a daily resolution and is based on the 3PG model developed by Landsberg & Waring (1997). Between the three catchments, control catchments C32 and C35 experienced warming and drying climate trends while catchment C31 was additionally treated with a clearcut harvest in the fall of 1997.

To fulfill the first research objective, we calibrated ensembles of models and assessed their ability to simulate summertime streamflow before and after forest disturbances. In both periods, calibrated Raven and Raven Robin models achieved daily Kling Gupta Efficiency (KGE) scores higher than their scenario-specific climatological reference KGE benchmarks, indicating that they could reasonably simulate observed streamflow. However, most models were less effective outside their calibration periods, suggesting that the two periods have distinct conditions. Pre-harvest Raven models in catchment C31 were an apparent exception to this, achieving consistent KGE scores across both periods. However, it was assumed that canopy losses from harvesting initially buffered the effect of climate trends, resulting in post-harvest streamflow that temporarily resembled the pre-harvest regime before vegetation regrowth took place. Although this allowed pre-harvest Raven models to perform well, their post-harvest streamflow simulations are subject to equifinality concerns given that they do not have representation for vegetation losses after harvesting. Thus, robust models that are capable of simulating streamflow responses to forest disturbances should be calibrated to both pre-

and post-harvest conditions, with explicit representation for altered watershed properties (such as vegetation cover after clearcut harvesting) if appropriate.

We calibrated ensembles of models in each catchment-period scenario to address model uncertainty in the hydrologic change detection analyses conducted for the second research objective. A standard model ensemble size of 25 members was determined to be ideal for all calibration scenarios; with 25 members, each model ensemble appeared to converge to a minimum mean continuous ranked probability score (MCRPS), which measures the difference between the cumulative distribution of model ensemble predictions and the empirical cumulative distribution of observations. Calibration trials yielded each of the 25 model members as a behavioral model solution that achieved a higher KGE score than the scenario-specific climatological reference benchmark.

In our first change detection method for fulfilling research objective 2, we used weekly runoff ratios to detect changes in the catchment rainfall-runoff response. Weekly runoff ratios were calculated as a 7-day moving window average using rainfall and streamflow data. In comparing the observed distributions between pre- and post-disturbance periods, we found that most of the differences occurred for runoff ratios below 10^{-3}. All three catchments produce low quantities of runoff: in every catchment more than 50% of runoff ratios were lower than 10^{-3}, making them difficult to physically interpret and subject to high potential for measurement errors that could impact the analysis. This suggests that the weekly runoff ratio was not an ideal change metric for evaluating forest harvesting and climate impacts on the hydrology of the headwater catchments in the Turkey Lakes watershed, given the low quantities of runoff that they produced at the weekly timescale. To address this concern, weekly timesteps were split into two groups based on whether the observed runoff ratio exceeded (high flows) or fell below (low flows) a minimum threshold of 10^{-4} assessed separately. Distributions of the simulated runoff ratio were only compared between model ensembles at time steps where high flows were observed; these distributions were compared using the Kolmorgorov-Smirnov test for models that were calibrated to different periods within the same catchment. Test results largely suggest that pre-harvest and post-harvest periods have distinct hydrologic conditions in all three catchments, causing the calibrated models to produce significantly different runoff responses.

To address the previously filtered data in our analysis using weekly runoff ratios, we compared the ability of models to simulate dry conditions when they occur as reported by observed data (dry

conditions typically occur for 10 to 15 weeks in the summer of every year). The calibration period was found to have a large impact on the effectiveness of the models in this regard: most pre-harvest models had variable success in simulating dry conditions as their training period featured a wetter and cooler climate. This was less so the case in catchment C31, where harvesting exerted a temporary buffering effect due to canopy losses that reduced evapotranspiration. Consequently, pre-harvest and post-harvest Raven models achieved similar success rates. Furthermore, post-harvest conditions required Raven Robin models to emulate a regrowth forest with less canopy coverage than the pre-harvest stands. This also inhibited the model's ability to simulate dry conditions as reduced evapotranspiration losses drive the model to produce more streamflow.

In our second hydrologic change detection method for fulfilling research objective 2, we compared calibrated model parameter distributions to detect process-level changes occurring within the system. Climate trends shifted Raven parameter distributions across calibration periods, creating altered soil water dynamics in control catchments C32 and C35: post-harvest parameters promoted more infiltration into the subsurface (soil layer thickness and porosity), increased soil evaporation (PET correction factors), and modified subsurface flow rates (baseflow power coefficient and hydraulic conductivity). Climate trends also impacted Raven Robin parameters, increasing plant water uptake (vegetation stress factor) and reducing variability in leaf density (minimum Leaf Area Index).

In catchment C31, clearcut harvesting buffered some of the climate trend impacts on soil parameters in the post-harvest period, but significantly changed vegetation parameters in Raven Robin models. Increases in soil storage were still detected (increased layer thickness and porosity) but accompanied this time with significant changes to Raven Robin parameters that created new plant evapotranspiration demand (Leaf Area Index bounds, PET adjustment coefficient) and plant water uptake dynamics (vegetation stress factor) for the regrowth forest. Overall, the hydrologic change detection analysis using model parameter distributions suggests that the distinct catchment conditions have implications for soil moisture dynamics as well as evapotranspiration regimes linking to vegetation cover.

The hydrologic change detection methods that we employed to fulfill research objective 2 provided different but complementary information about the forested systems in this study. In comparing distributions of hydrologic signatures (weekly runoff ratio) simulated between model ensembles, significant differences suggested that conditions are distinct between calibration periods - leading to

fundamentally different model responses. However, these significant differences were only reported for weeks where high flow conditions were observed: comparisons could not be made under dry conditions (which occurred for up to 16 weeks in the 3 summer months analyzed each year) given that the runoff ratios were difficult to physically interpret (all under 10^{-4}). Despite this shortcoming, the different high-runoff responses that were reported can be functionally explained by the results of the second method, where we compare calibrated parameter distributions between the same models: significant shifts in model parameter values point to process-level impacts exerted by disturbances to the system.

While our results demonstrate how models can be applied to simulate and detect hydrologic changes in forested systems, improvements can be made in future analyses with consideration for phenological timescales. The post-harvest calibration period in catchment C31 includes streamflow affected by initial vegetation losses as well as subsequent vegetation regrowth: calibrating models to the entire period yields "averaged" model solutions that simulate the mean hydrologic regime across multiple stages of vegetation recovery that may exhibit distinct hydrologic behavior. This was previously shown by Merz et al. (2011), who calibrated hydrologic models to 5-year periods over 40 years in 273 Austrian catchments: in their analysis, they found that root zone storage capacity gradually doubled over time with climate trends. Therefore, post-disturbance calibration periods for models should be broken down into sub-periods that match these stages to better quantify forest disturbance impacts on hydrology (which are dynamic) over time. In doing so, modellers will inevitably lean on the expertise and insights of those with knowledge of the study area and the vegetation species involved, particularly during data assimilation to extend change detection analyses.

Model-based change detection analyses are not perfect, but their results can at the very least speak to where additional observation data are most needed – effectively informing the design of future field programs which can, in turn, effectively improve the quality of models. Thus, it is the hope that this cyclical motivation can foster more collaboration between backgrounds and disciplines in not only detecting – but also inducing – positive hydrologic changes through effective forest management.

References

Andréassian, V., Parent, E., & Michel, C. (2003). A distribution-free test to detect gradual changes in watershed behavior: Detecting changes in watershed behavior. *Water Resources Research*, *39*(9). https://doi.org/10.1029/2003WR002081

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., Van Liew, M. W., Kannan, N., & Jha, M. K. (2012). SWAT: Model Use, Calibration, and Validation. *Transactions of the ASABE*, *55*(4), 1491–1508. https://doi.org/10.13031/2013.42256

Bates, C. G., & Henry, A. J. (1928). *Forest and streamflow experiments at Wagon Wheel Gap, Colorado* (Supplement No. 30; Monthly Weather Review, pp. 79–80). Government Printing Office.

Beall, F. D., Semkin, R. G., & Jeffries, D. S. (2001). Trends in the Output of First-Order Basins at Turkey Lakes Watershed, 1982-96. *Ecosystems*, *4*(6), 514–526.

Bergström, S., & Singh, V. (1995). The HBV Model. In *Computer Models of Watershed Hydrology* (pp. 443–476).

Beven, K. (2005). On the concept of model structural error. *Water Science and Technology*, *52*(6), 167–175. https://doi.org/10.2166/wst.2005.0165

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Blasone, R.-S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A., & Zyvoloski, G. A. (2008). Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Advances in Water Resources*, *31*(4), 630–648. https://doi.org/10.1016/j.advwatres.2007.12.003

Bosch, J. M., & Hewlett, J. D. (1982). A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *Journal of Hydrology*, *55*(1–4), 3–23. https://doi.org/10.1016/0022-1694(82)90117-2

Brandt, M., Bergström, S., & Gardelin, M. (1998). Modelling the effects of clearcutting on runoff—Examples from central Sweden. *Ambio*, *17*(5), 307–313.

Brown, A. E., Zhang, L., McMahon, T. A., Western, A. W., & Vertessy, R. A. (2005). A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. *Journal of Hydrology*, *310*(1–4), 28–61. https://doi.org/10.1016/j.jhydrol.2004.12.010

Burt, T. P., & Swank, W. T. (1992). Flow frequency responses to hardwood-to-grass conversion and subsequent succession. *Hydrological Processes*, *6*(2), 179–188. https://doi.org/10.1002/hyp.3360060206

Buttle, J. M., Beall, F. D., Webster, K. L., Hazlett, P. W., Creed, I. F., Semkin, R. G., & Jeffries, D. S. (2018). Hydrologic response to and recovery from differing silvicultural systems in a deciduous forest landscape with seasonal snow cover. *Journal of Hydrology*, *557*, 805–825. https://doi.org/10.1016/j.jhydrol.2018.01.006

Buttle, J. M., Webster, K. L., Hazlett, P. W., & Jeffries, D. S. (2019). Quickflow response to forest harvesting and recovery in a northern hardwood forest landscape. *Hydrological Processes*, *33*(1), 47–65. https://doi.org/10.1002/hyp.13310

Butts, M. B., Payne, J. T., Kristensen, M., & Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, *298*(1), 242–266. https://doi.org/10.1016/j.jhydrol.2004.03.042

Cheng, J. D. (1989). Streamflow changes after clear-cut logging of a pine beetle-infested watershed in southern British Columbia, Canada. *Water Resources Research*, *25*(3), 449–456. https://doi.org/10.1029/WR025i003p00449

Chlumsky, R., Mai, J., Craig, J. R., & Tolson, B. A. (2021). Simultaneous Calibration of Hydrologic Model Structure and Parameters Using a Blended Model. *Water Resources Research*, *57*(5). https://doi.org/10.1029/2020WR029229

Craig, J. R., Brown, G., Chlumsky, R., Jenkinson, R. W., Jost, G., Lee, K., Mai, J., Serrer, M., Sgro, N., Shafii, M., Snowdon, A. P., & Tolson, B. A. (2020). Flexible watershed simulation with the Raven hydrological modelling framework. *Environmental Modelling & Software*, *129*, 104728. https://doi.org/10.1016/j.envsoft.2020.104728

Craig, J. R., & the Raven Development Team. (2022). *Raven user's and developer's manual* (3.5). http://raven.uwaterloo.ca/

Cuntz, M., Mai, J., Zink, M., Thober, S., Kumar, R., Schäfer, D., Schrön, M., Craven, J., Rakovec, O., Spieler, D., Prykhodko, V., Dalmasso, G., Musuuza, J., Langenberg, B., Attinger, S., & Samaniego, L. (2015). Computationally inexpensive identification of noninformative model parameters by sequential screening. *Water Resources Research*, *51*(8), 6417–6441. https://doi.org/10.1002/2015WR016907

Dibike, Y. B., & Coulibaly, P. (2005). Hydrologic impact of climate change in the Saguenay watershed: Comparison of downscaling methods and hydrologic models. *Journal of Hydrology*, *307*(1), 145–163. https://doi.org/10.1016/j.jhydrol.2004.10.012

Emelko, M. B., Silins, U., Bladon, K. D., & Stone, M. (2011). Implications of land disturbance on drinking water treatability in a changing climate: Demonstrating the need for source water supply and protection strategies. *Water Resources Research*, *45*, 461–472. http://dx.doi.org/10.1016/j.watres.2010.08.051

English, M. C., Jeffries, D. S., Foster, N. W., Semkin, R. G., & Hazlett, P. W. (1986). A preliminary assessment of the chemical and hydrological interaction of acidic snowmelt water with the terrestrial portion of a Canadian shield catchment. *Water, Air, Soil, Pollution*, *31*, 27–34.

Environment and Climate Change Canada. (2018, April 23). *Turkey Lakes watershed study: Overview*. https://www.canada.ca/en/environment-climate-change/services/turkey-lakes-watershed-study/overview.html

Goodbrand, A., Anderson, A., Devito, K., & Silins, U. (2022). Untangling harvest-streamflow responses in foothills conifer forests: Nexus of teleconnections, summer-dominated precipitation, and storage. *Hydrological Processes*, *36*(2), e14479. https://doi.org/10.1002/hyp.14479

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., & Ye, M. (2012). Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, *48*(8). https://doi.org/10.1029/2011WR011044

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

Han, M. (2022). *Improving hydrological process representation in lake and forest dominated watersheds* [University of Waterloo]. https://uwspace.uwaterloo.ca/handle/10012/18441

Han, M., Shen, H., Tolson, B. A., Craig, J. R., Mai, J., Lin, S. G. M., Basu, N. B., & Awol, F. S. (2022). *Basinmaker 3.0: A GIS Toolbox for Distributed Watershed Delineation of Complex Lake-River Routing Networks* (SSRN Scholarly Paper No. 4135646). https://doi.org/10.2139/ssrn.4135646

Harr, R. D. (1976). *Forest practices and streamflow in Western Oregon* (PNW-GTR-49; Symposium on Watershed Management). US Forest Service.

Harr, R. D., Fredriksen, R. L., & Rothascher, J. (1979). *Changes in streamflow following timber harvest in Southwest Oregon* (Research Paper PNW-249). US Forest Service.

Hazlett, P. W., Curry, J. M., & Weldon, T. P. (2011). Assessing Decadal Change in Mineral Soil Cation Chemistry at the Turkey Lakes Watershed. *Soil Science Society of America Journal*, *75*(1), 287–305. https://doi.org/10.2136/sssaj2010.0090

Hazlett, P. W., Semkin, R. G., & Beall, F. D. (2001). Hydrologic Pathways during Snowmelt in First-order Stream Basins at the Turkey Lakes Watershed. *Ecosystems*, *4*(6), 527–535. https://doi.org/10.1007/s10021-001-0026-z

Hersbach, H. (2000). Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, *15*(5), 559–570. https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2

Hewlett, J. D. (1971). Comments on the catchment experiment to determine vegetal effects on water yield. *Water Resources Bulletin, 7*, 376–381.

Hewlett, J. D., & Hibbert, A. R. (1967). Factors affecting the response of small watersheds to precipitation in humid areas. In *Forest Hydrology* (Vol. 11, pp. 275–290). Pergamon Press.

Hibbert, A. R. (1967). Forest treatment effects on water yield. *Forest Hydrology*, 527–543.

Hibbert, A. R., & Gottfried, G. J. (1987). *Stormflow responses to forest treatments on two Arizona mixed conifer watersheds* (General Technical Report RM-149; Management of Subalpine Forests: Building on 50 Years of Research, pp. 189–193). US Forest Service.

Hicks, B. J., Beschta, R. L., & Harr, R. D. (1991). Long-term changes in streamflow following logging in western Oregon and associated fisheries implications. *Water Resources Bulletin*, *27*(2), 217–22610.

Hornbeck, J. W., Adams, M. B., Corbett, E. S., Verry, E. S., & Lynch, J. A. (1993). Long-term impacts of forest treatments on water yield: A summary for northeastern USA. *Journal of Hydrology*, *150*(2), 323–344. https://doi.org/10.1016/0022-1694(93)90115-P

Jeffries, D. S., Kelso, J. R. M., & Morrison, I. K. (1988). Physical, Chemical, and Biological Characteristics of the Turkey Lakes Watershed, Central Ontario, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, *45*(S1), s3–s13. https://doi.org/10.1139/f88-262

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., & Thielen, J. (2016). Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. *Environmental Modelling & Software*, *75*, 68–76. https://doi.org/10.1016/j.envsoft.2015.09.009

Keppeler, E. T., & Ziemer, R. R. (1990). Logging effects on streamflow: Water yield and summer low flows at Caspar Creek in northwestern California. *Water Resources Research*, *26*(7), 1669–1679. https://doi.org/10.1029/WR026i007p01669

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Kolmorgorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Instituto Italiano degli Attuari*, *4*, 83–91.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

Kuczera, G., & Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm. *Journal of Hydrology*, *211*(1), 69–85. https://doi.org/10.1016/S0022-1694(98)00198-X

Kundzewicz, Z. W., & Robson, A. J. (2004). Change detection in hydrological records—A review of the methodology / Revue méthodologique de la détection de changements dans les chroniques hydrologiques. *Hydrological Sciences Journal*, *49*(1), 7–19. https://doi.org/10.1623/hysj.49.1.7.53993

Landsberg, J. J., & Waring, R. H. (1997). A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest Ecology and Management*, *95*(3), 209–228. https://doi.org/10.1016/S0378-1127(97)00026-1

Lavabre, J., Torres, D. S., & Cernesson, F. (1993). Changes in the hydrological response of a small Mediterranean basin a year after a wildfire. *Journal of Hydrology*, *142*(1), 273–299. https://doi.org/10.1016/0022-1694(93)90014-Z

Leach, J. A., Buttle, J. M., Webster, K. L., Hazlett, P. W., & Jeffries, D. S. (2020). Travel times for snowmelt-dominated headwater catchments: Influences of wetlands and forest harvesting, and linkages to stream water quality. *Hydrological Processes*, *34*(10), 2154–2175. https://doi.org/10.1002/hyp.13746

Matheson, J. E., & Winkler, R. L. (1976). Scoring Rules for Continuous Probability Distributions. *Management Science*, *22*(10), 1087–1096.

McMinn, J. W., & Hewlett, J. D. (1975). First-Year Water Yield Increase After Forest Cutting: An Alternative Model. *Journal of Forestry*, *73*(10), 654–655. https://doi.org/10.1093/jof/73.10.654

Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, *47*(2). https://doi.org/10.1029/2010WR009505

Millennium Ecosystem Assessment (Ed.). (2005). *Ecosystems and human well-being: Wetlands and water synthesis: a report of the Millennium Ecosystem Assessment*. World Resources Institute.

Mosley, M. P. (1979). Streamflow generation in a forested watershed, New Zealand. *Water Resources Research*, *15*(4), 795–806. https://doi.org/10.1029/WR015i004p00795

Murray, C. D., & Buttle, J. M. (2003). Impacts of clearcut harvesting on snow accumulation and melt in a northern hardwood forest. *Journal of Hydrology*, 16.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nicolson, J. A. (1988). Water and Chemical Budgets for Terrestrial Basins at the Turkey Lakes Watershed. *Canadian Journal of Fisheries and Aquatic Sciences*, *45*(S1), s88–s95. https://doi.org/10.1139/f88-271

Papacharalampous, G., Tyralis, H., Koutsoyiannis, D., & Montanari, A. (2020). Quantification of predictive uncertainty in hydrological modelling by harnessing the wisdom of the crowd: A large-sample experiment at monthly timescale. *Advances in Water Resources*, *136*, 103470. https://doi.org/10.1016/j.advwatres.2019.103470

Pettitt, A. N. (1979). A Non-Parametric Approach to the Change-Point Problem. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(2), 126–135. https://doi.org/10.2307/2346729

Priestley, C., & Taylor, R. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, 81–92. https://doi.org/10.1175/1520-0493

Rowe, J.S. (1972). *Forest Regions of Canada* (Forest Service Publication No. 1300; p. 172). Canadian Forestry Service.

Schreider, S. Yu., Jakeman, A. J., Letcher, R. A., Nathan, R. J., Neal, B. P., & Beavis, S. G. (2002). Detecting changes in streamflow response to changes in non-climatic catchment conditions: Farm dam development in the Murray–Darling basin, Australia. *Journal of Hydrology*, *262*(1), 84–98. https://doi.org/10.1016/S0022-1694(02)00023-9

Schroeter, H. (1989). *GAWSER Training Guide and Reference Manual* [Manual]. Grand River Conservation Authority (GRCA).

Seibert, J., & McDonnell, J. J. (2010). Land-cover impacts on streamflow: A change-detection modelling approach that incorporates parameter uncertainty. *Hydrological Sciences Journal*, *55*(3), 316–332. https://doi.org/10.1080/02626661003683264

Seibert, J., McDonnell, J. J., & Woodsmith, R. D. (2010). Effects of wildfire on catchment runoff response: A modelling approach to detect changes in snow-dominated forested catchments. *Hydrology Research*, *41*(5), 378–390. https://doi.org/10.2166/nh.2010.036

Semkin, R. G., & Jeffries, D. S. (1988). Chemistry of Atmospheric Deposition, the Snowpack, and Snowmelt in the Turkey Lakes Watershed. *Canadian Journal of Fisheries and Aquatic Sciences*, *45*(S1), s38–s46. https://doi.org/10.1139/f88-265

Shafii, M., Tolson, B., & Shawn Matott, L. (2015). Addressing subjective decision-making inherent in GLUE-based multi-criteria rainfall–runoff model calibration. *Journal of Hydrology*, *523*, 693–705. https://doi.org/10.1016/j.jhydrol.2015.01.051

Siegel, S., & Castellan, N. J. (1988). *Non-parametric statistics for the Behavioral Sciences* (2nd ed.). McGraw-Hill.

Smirnov, N. V. (1939). Estimate of deviation between empirical distribution functions in two independent samples. *Moscow University Bulletin*, *2*(2), 3–16.

Stednick, J. D. (1995). Long term changes in streamflow following timber harvesting in the Oregon Coast Range: Water quantity. In *The Alsea Watershed: Hydrological and Biological Responses to Temperate Coniferous Forest Practices*. Springer.

Stednick, J. D. (1996). Monitoring the effects of timber harvest on annual water yield. *Journal of Hydrology*, *176*(1–4), 79–95.

Stednick, J. D., & Kern, T. J. (1992). Long term effects of timber harvesting in the Oregon Coast Range: The New Alsea Watershed Study (NAWS). In *Interdisciplinary Approaches to Hydrology and Hydrogeology* (pp. 502–510). American Institute of Hydrology.

Swank, W. T., Swift, L. W., & Douglass, J. E. (1988). Streamflow Changes Associated with Forest Cutting, Species Conversions, and Natural Disturbances. In W. T. Swank & D. A. Crossley (Eds.), *Forest Hydrology and Ecology at Coweeta* (pp. 297–312). Springer. https://doi.org/10.1007/978-1-4612-3732-7_22

Tani, M. (1997). Runoff generation processes estimated from hydrological observations on a steep forested hillslope with a thin soil layer. *Journal of Hydrology*, *200*(1), 84–109. https://doi.org/10.1016/S0022-1694(97)00018-8

Tolson, B. A., & Shoemaker, C. A. (2008). Efficient prediction uncertainty approximation in the calibration of environmental simulation models: Improving uncertainty analysis efficiency. *Water Resources Research*, *44*(4). https://doi.org/10.1029/2007WR005869

Troendle, C. A., & King, R. M. (1985). The Effect of Timber Harvest on the Fool Creek Watershed, 30 Years Later. *Water Resources Research*, *21*(12), 1915–1922. https://doi.org/10.1029/WR021i012p01915

Tromp-van Meerveld, H. J., & McDonnell, J. J. (2006). Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope. *Water Resources Research*, *42*(2). https://doi.org/10.1029/2004WR003778

Uchida, T., Tromp-van Meerveld, I., & McDonnell, J. J. (2005). The role of lateral pipe flow in hillslope runoff response: An intercomparison of non-linear hillslope response. *Journal of Hydrology*, *311*(1–4), 117–133. https://doi.org/10.1016/j.jhydrol.2005.01.012

Webster, K. L., Leach, J. A., Hazlett, P. W., Fleming, R. L., Emilson, E. J. S., Houle, D., Chan, K. H. Y., Norouzian, F., Cole, A. S., O'Brien, J. M., Smokorowski, K. E., Nelson, S. A., & Yanni, S. D. (2021). Turkey Lakes Watershed, Ontario, Canada: 40 years of interdisciplinary whole-ecosystem research. *Hydrological Processes*, *35*(4). https://doi.org/10.1002/hyp.14109

Whipkey, R. Z. (1965). Subsurface Stormflow from Forested Slopes. *International Association of Scientific Hydrology. Bulletin*, *10*(2), 74–85. https://doi.org/10.1080/02626666509493392

Whitehead, P. G., & Robinson, M. (1993). Experimental basin studies—An international and historical perspective of forest impacts. *Journal of Hydrology*, *145*(3), 217–230. https://doi.org/10.1016/0022-1694(93)90055-E

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, *1*(6), 80–83. https://doi.org/10.2307/3001968

Worsley, K. J. (1979). On the Likelihood Ratio Test for a Shift in Location of Normal Populations. *Journal of the American Statistical Association*, *74*(366a), 365–367. https://doi.org/10.1080/01621459.1979.10482519

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., & Bates, P. D. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, *44*(11), 5844–5853. https://doi.org/10.1002/2017GL072874

Zégre, N., Skaugset, A. E., Som, N. A., McDonnell, J. J., & Ganio, L. M. (2010). In lieu of the paired catchment approach: Hydrologic model change detection at the catchment scale: in lieu of the paired catchment approach. *Water Resources Research*, *46*(11). https://doi.org/10.1029/2009WR008601

Appendix A
Model Definition Files

Turkey Lakes watershed Raven hydrologic model setup

```
# ---------------------------------------------
# Raven Input file
# HBV-EC Turkey Lakes Watershed Emulation
# ---------------------------------------------
# --Simulation Details ------------------------
:StartDate              1981-01-01 00:00:00
:Duration               12050  # 1981-01-01 to 2013-12-29
:Method                         ORDERED_SERIES
:TimeStep               1.0
:RunName                calibration
:OutputDirectory        output/
#
# --Model Details -----------------------------
:Method                 ORDERED_SERIES
:Interpolation          INTERP_NEAREST_NEIGHBOR
:SoilModel              SOIL_MULTILAYER 3

#:VegetationProcesses Robin # toggle switch that connects Robin (off)

:Routing                ROUTE_DIFFUSIVE_WAVE
:CatchmentRoute         ROUTE_TRI_CONVOLUTION
:OW_Evaporation         PET_PRIESTLEY_TAYLOR
:RainSnowFraction       RAINSNOW_DATA

##Vegetation impacted processes
:SWCanopyCorrect        SW_CANOPY_CORR_NONE
:PrecipIceptFract       PRECIP_ICEPT_LAI
:Evaporation            PET_PRIESTLEY_TAYLOR
:PotentialMeltMethod    POTMELT_HBV

###about forcing data
:RelativeHumidityMethod         RELHUM_DATA
:SWRadiationMethod              SW_RAD_DATA
:LWRadiationMethod              LW_RAD_DEFAULT
:WindspeedMethod                WINDVEL_DATA

# --Hydrologic Processes-------------------------
:Alias          Forest_Floor    SOIL[0]
:Alias          Ablation_Till   SOIL[1]
:Alias          Basal_Till      SOIL[2]
#

:DefineHRUGroups        C31_C32_C35_HRUS  Non_C31_C32_C35_HRUS
:DisableHRUGroup        Non_C31_C32_C35_HRUS
:EvaluationPeriod       PRE-HARVEST 1984-01-01 1997-08-31
:EvaluationPeriod       POST-HARVEST 1997-09-01 2012-12-01
```

```
:EvaluationPeriod       POST-HARVEST_TRUNC 1998-01-01 2012-12-01

:HydrologicProcesses

  :SnowRefreeze         FREEZE_DEGREE_DAY     SNOW_LIQ     SNOW
  :Precipitation        PRECIP_RAVEN          ATMOS_PRECIP       MULTIPLE

  :CanopyDrip           CANDRIP_RUTTER        CANOPY             PONDED_WATER
  :CanopyEvaporation    CANEVP_MAXIMUM        CANOPY             ATMOSPHERE
  :CanopySnowEvap       CANEVP_MAXIMUM        CANOPY_SNOW        ATMOSPHERE
  :SnowBalance          SNOBAL_SIMPLE_MELT    SNOW               SNOW_LIQ
   :-->Overflow    RAVEN_DEFAULT    SNOW_LIQ      PONDED_WATER
  :Infiltration         INF_GREEN_AMPT        PONDED_WATER       MULTIPLE
  ### Forest_Floor
  :SoilEvaporation      SOILEVAP_SEQUEN       Forest_Floor       ATMOSPHERE
  :Percolation          PERC_GAWSER           Forest_Floor       Ablation_Till
  :Baseflow             BASE_THRESH_POWER     Forest_Floor       SURFACE_WATER
  ### Ablation_Till
  :Baseflow             BASE_THRESH_POWER     Ablation_Till      SURFACE_WATER
  :Percolation          PERC_GAWSER           Ablation_Till      Basal_Till
  ### Basal_Till
  :Baseflow             BASE_THRESH_POWER     Basal_Till         SURFACE_WATER
  :CapillaryRise        CRISE_HBV             Basal_Till         Ablation_Till
:EndHydrologicProcesses

# --Output Options------------------------------
:SilentMode
:WriteForcingFunctions
:EvaluationMetrics NASH_SUTCLIFFE  PCT_BIAS KLING_GUPTA RMSE
KLING_GUPTA_SUMMER KLING_GUPTA_NOT_SUMMER
:WriteMassBalanceFile
```

Turkey Lakes watershed Raven Robin hydrologic-vegetation growth model setup
Raven input file (*.rvi): model definition

```
# ---------------------------------------------
# Raven Input file
# HBV-EC Turkey Lakes Watershed Emulation
# ---------------------------------------------
# --Simulation Details -------------------------
:StartDate              1981-01-01 00:00:00
:Duration               12050  # 1981-01-01 to 2013-12-29
:Method                         ORDERED_SERIES
:TimeStep               1.0
:RunName                calibration
:OutputDirectory        output/
#
# --Model Details ------------------------------
:Method                 ORDERED_SERIES
:Interpolation          INTERP_NEAREST_NEIGHBOR
:SoilModel              SOIL_MULTILAYER 3

:VegetationProcesses Robin # toggle switch that connects Robin (on)

:Routing                ROUTE_DIFFUSIVE_WAVE
:CatchmentRoute         ROUTE_TRI_CONVOLUTION
:OW_Evaporation         PET_PRIESTLEY_TAYLOR
:RainSnowFraction       RAINSNOW_DATA

##Vegetation impacted processes
:SWCanopyCorrect        SW_CANOPY_CORR_NONE
:PrecipIceptFract       PRECIP_ICEPT_LAI
:Evaporation            PET_PRIESTLEY_TAYLOR
:PotentialMeltMethod    POTMELT_HBV

###about forcing data
:RelativeHumidityMethod         RELHUM_DATA
:SWRadiationMethod              SW_RAD_DATA
:LWRadiationMethod              LW_RAD_DEFAULT
:WindspeedMethod                WINDVEL_DATA

# --Hydrologic Processes-------------------------
:Alias          Forest_Floor    SOIL[0]
:Alias          Ablation_Till   SOIL[1]
:Alias          Basal_Till      SOIL[2]
#

:DefineHRUGroups        C31_C32_C35_HRUS  Non_C31_C32_C35_HRUS
:DisableHRUGroup        Non_C31_C32_C35_HRUS
:EvaluationPeriod       PRE-HARVEST 1984-01-01 1997-08-31
```

```
:EvaluationPeriod       POST-HARVEST 1997-09-01 2012-12-01
:EvaluationPeriod       POST-HARVEST_TRUNC 1998-01-01 2012-12-01

:HydrologicProcesses

  :SnowRefreeze         FREEZE_DEGREE_DAY       SNOW_LIQ     SNOW
  :Precipitation        PRECIP_RAVEN            ATMOS_PRECIP       MULTIPLE

  :CanopyDrip           CANDRIP_RUTTER          CANOPY             PONDED_WATER
  :CanopyEvaporation    CANEVP_MAXIMUM          CANOPY             ATMOSPHERE
  :CanopySnowEvap       CANEVP_MAXIMUM          CANOPY_SNOW        ATMOSPHERE
  :SnowBalance          SNOBAL_SIMPLE_MELT      SNOW               SNOW_LIQ
   :-->Overflow     RAVEN_DEFAULT     SNOW_LIQ       PONDED_WATER
  :Infiltration         INF_GREEN_AMPT          PONDED_WATER       MULTIPLE
  ### Forest_Floor
  :SoilEvaporation      SOILEVAP_ROOT_VEG       Forest_Floor       ATMOSPHERE
  :Percolation          PERC_GAWSER             Forest_Floor       Ablation_Till
  :Baseflow             BASE_THRESH_POWER       Forest_Floor       SURFACE_WATER
  ### Ablation_Till
  :Baseflow             BASE_THRESH_POWER       Ablation_Till      SURFACE_WATER
  :Percolation          PERC_GAWSER             Ablation_Till      Basal_Till
  ### Basal_Till
  :Baseflow             BASE_THRESH_POWER       Basal_Till         SURFACE_WATER
  :CapillaryRise        CRISE_HBV               Basal_Till         Ablation_Till
:EndHydrologicProcesses

# --Output Options------------------------------
:SilentMode
:WriteForcingFunctions
:EvaluationMetrics NASH_SUTCLIFFE  PCT_BIAS KLING_GUPTA RMSE
KLING_GUPTA_SUMMER KLING_GUPTA_NOT_SUMMER
:WriteMassBalanceFile
```

Robin input file (*.model): process algorithm definition

21

VEG_NAME	Forest	Forest2
LAI_MODEL	3PG_LAI	3PG_LAI
PHENOLOGY_MODEL	3PG_PNLG	3PG_PNLG
HT_MODEL	Ht_3PG	Ht_3PG
WATER_STRESS_MODEL	AETPET	AETPET
TEMP_STRESS_MODEL	Temp_3PG	Temp_3PG
VPD_STRESS_MODEL	VPD_3PG	VPD_3PG
TOTAL_STRESS_MASS_ACC_MODEL	MA_3PG	MA_3PG
OPTIMAL_GPP_MODEL	Beer_LAW	Beer_LAW
RESPIRATION_MODEL	R_Const_C	R_Const_C
MASS_DIS_ROOT_MODEL	BioA_Rt_3PG	BioA_Rt_3PG
MASS_DIS_LEAF_MODEL	BioA_LF_3PG	BioA_LF_3PG
MASS_DIS_SEED_MODEL	BioA_SD_3PG	BioA_SD_3PG
DBH_MODEL	DBH_3PG	DBH_3PG
BASAL_AREA_MODEL	BasA_3PG	BasA_3PG
STAND_VOlUME_MODEL	StVol_3PG	StVol_3PG
MASS_LOSS_ROOT_MODEL	WL_Rt_3PG	WL_Rt_3PG
MASS_LOSS_LEAF_MODEL	WL_LF_3PG	WL_LF_3PG
MASS_LOSS_VEG_NUM_MODEL	WL_NV_3PG	WL_NV_3PG
ROOT_DISTRIBUTION_MODEL	ROOT_SWAT	ROOT_SWAT
REPRODUCTION_MODEL	ReProd_3PG	ReProd_3PG

Robin input file (*.info): simulation information

```
1981            # read (ifile,*) Begin_Year
1               # read (ifile,*) Begin_juliday
12050           # read (ifile,*) ndays
38              # read (ifile,*) nsub
169             # read (ifile,*) nhru
2               # read (ifile,*) ncrop
88              # nseeding event
88              # nharvest event
2               # nrotation
88              # number of growing season
999             # max hru id
999             # max sub id
1               # NoNutrient
1981 01 01      # YEAR MONTH DAY
1               # 1 suppresses screen output (0 turns it back on)
###
./cropout/
0               # is the number of customized output options. 0 means no customized outputs.
TIMESERIES      # Output type
-1              # OutputVars -1 writes all supported variables
27 63 144       # HRU IDs
-1              # sub-basin IDs -1 means no sub-basins
-1              # timestep -1 means all model time steps
TIMESERIES
-1
-1
178
-1
TIMESERIES
-1
22 110 109
-1
-1
TIMESERIES
-1
-1
116
-1
```

Robin input file (*.inic): initial conditions

##### crop Initial condition		
13		
CropName	Forest	Forest2
w_total	305725.7438	4
w_leaf	5399.823839	1056
w_stems	1.358499E+05	36114.64968
w_seed	0	1
w_root	33000	500
ht	15	5
rootdep	5	5
dbh	17.5	20
basa_area	22	6.3
volume	176	1
yr_c	182	182
vg_number	1.095562E+03	201
stage	1	1

Robin input file (*.mgt): harvest and seeding schedule

####

88

mgtype index	Forest	Forest2
seeding 1	31	31
seeding 2	31	31
seeding 3	30	30
seeding 4	30	30
seeding 5	30	30
seeding 6	31	31
seeding 7	30	30
seeding 8	30	30
seeding 9	30	30
seeding 10	31	31
seeding 11	30	30
seeding 12	30	30
seeding 13	30	30
seeding 14	31	31
seeding 15	30	30
seeding 16	30	30
seeding 17	30	30
seeding 18	31	31
seeding 19	30	30
seeding 20	30	30
seeding 21	30	30
seeding 22	31	31
seeding 23	30	30
seeding 24	30	30
seeding 25	30	30
seeding 26	31	31
seeding 27	30	30
seeding 28	30	30
seeding 29	30	30
seeding 30	31	31
seeding 31	30	30
seeding 32	30	30
seeding 33	30	30
seeding 34	31	31
seeding 35	30	30
seeding 36	30	30
seeding 37	30	30
seeding 38	31	31
seeding 39	30	30
seeding 40	30	30
seeding 41	30	30
seeding 42	31	31
seeding 43	30	30

seeding	44	30	30
harvest	1	335	335
harvest	2	336	336
harvest	3	335	335
harvest	4	335	335
harvest	5	335	335
harvest	6	336	336
harvest	7	335	335
harvest	8	335	335
harvest	9	335	335
harvest	10	336	336
harvest	11	335	335
harvest	12	335	335
harvest	13	335	335
harvest	14	336	336
harvest	15	335	335
harvest	16	335	335
harvest	17	335	335
harvest	18	336	336
harvest	19	335	335
harvest	20	335	335
harvest	21	335	335
harvest	22	336	336
harvest	23	335	335
harvest	24	335	335
harvest	25	335	335
harvest	26	336	336
harvest	27	335	335
harvest	28	335	335
harvest	29	335	335
harvest	30	336	336
harvest	31	335	335
harvest	32	335	335
harvest	33	335	335
harvest	34	336	336
harvest	35	335	335
harvest	36	335	335
harvest	37	335	335
harvest	38	336	336
harvest	39	335	335
harvest	40	335	335
harvest	41	335	335
harvest	42	336	336
harvest	43	335	335
harvest	44	335	335

Robin input file (*.hrucrop): crop rotation schedule

subid	hruid	rotationname	temp_ave
2	11	R1	18
2	33	R1	18
2	34	R1	18
2	73	R1	18
2	74	R1	18
11	12	R1	18
11	35	R1	18
11	36	R1	18
11	75	R1	18
11	76	R1	18
16	13	R1	18
16	77	R1	18
16	78	R1	18
16	79	R1	18
29	1	R1	18
29	14	R1	18
29	37	R1	18
29	38	R1	18
29	39	R1	18
29	80	R1	18
36	2	R1	18
36	15	R1	18
36	40	R1	18
36	41	R1	18
36	81	R1	18
36	82	R1	18
38	3	R1	18
38	83	R1	18
38	84	R1	18
38	85	R1	18
38	86	R1	18
40	87	R1	18
40	88	R1	18
40	89	R1	18
40	90	R1	18
40	91	R1	18
41	92	R1	18
41	93	R1	18
41	94	R1	18
41	95	R1	18
50	16	R1	18
50	42	R1	18
50	43	R1	18
50	96	R1	18
53	10	R1	18

53	17	R1	18
53	44	R1	18
53	45	R1	18
59	4	R1	18
59	97	R1	18
59	98	R1	18
59	99	R1	18
59	100	R1	18
73	18	R1	18
73	46	R1	18
73	47	R1	18
73	101	R1	18
73	102	R1	18
93	19	R1	18
93	48	R1	18
93	49	R1	18
93	103	R1	18
103	5	R1	18
103	20	R1	18
103	50	R1	18
103	51	R1	18
106	104	R1	18
106	105	R1	18
106	106	R1	18
106	107	R1	18
106	108	R1	18
115	21	R1	18
115	52	R1	18
115	53	R1	18
116	22	R2	18
116	54	R2	18
116	109	R2	18
116	110	R2	18
119	23	R1	18
119	55	R1	18
119	111	R1	18
133	112	R1	18
133	113	R1	18
133	114	R1	18
133	115	R1	18
134	24	R1	18
134	56	R1	18
134	57	R1	18
134	116	R1	18
142	117	R1	18
142	118	R1	18
142	119	R1	18

142	120	R1	18
150	121	R1	18
150	122	R1	18
150	123	R1	18
150	124	R1	18
150	125	R1	18
163	6	R1	18
163	25	R1	18
163	58	R1	18
163	59	R1	18
163	60	R1	18
164	7	R1	18
164	126	R1	18
164	127	R1	18
164	128	R1	18
164	129	R1	18
166	130	R1	18
166	131	R1	18
166	132	R1	18
166	133	R1	18
167	134	R1	18
167	135	R1	18
167	136	R1	18
167	137	R1	18
170	138	R1	18
170	139	R1	18
170	140	R1	18
170	141	R1	18
170	142	R1	18
173	26	R1	18
173	61	R1	18
173	62	R1	18
173	143	R1	18
178	27	R1	18
178	63	R1	18
178	144	R1	18
181	145	R1	18
181	146	R1	18
181	147	R1	18
181	148	R1	18
195	28	R1	18
195	149	R1	18
195	150	R1	18
205	151	R1	18
205	152	R1	18
205	153	R1	18
221	8	R1	18

221	29	R1	18
221	64	R1	18
221	65	R1	18
221	154	R1	18
222	155	R1	18
222	156	R1	18
222	157	R1	18
222	158	R1	18
263	9	R1	18
263	30	R1	18
263	66	R1	18
263	67	R1	18
263	68	R1	18
263	159	R1	18
275	31	R1	18
275	69	R1	18
275	160	R1	18
275	161	R1	18
275	162	R1	18
276	163	R1	18
276	164	R1	18
276	165	R1	18
276	166	R1	18
276	167	R1	18
288	32	R1	18
288	70	R1	18
288	71	R1	18
288	72	R1	18
288	168	R1	18
288	169	R1	18

rotationname	cropname1	cropname2	cropname3	cropname4	cropname5
cropname6	cropname7	cropname8	cropname9	cropname10	
cropname11	cropname12	cropname13	cropname14	cropname15	
cropname16	cropname17	cropname18	cropname19	cropname20	
cropname21	cropname22	cropname23	cropname24	cropname25	
cropname26	cropname27	cropname28	cropname29	cropname30	
cropname31	cropname32	cropname33	cropname34	cropname35	
cropname36	cropname37	cropname38	cropname39	cropname40	
cropname41	cropname42	cropname43	cropname44	cropname45	
cropname46	cropname47	cropname48	cropname49	cropname50	
cropname51	cropname52	cropname53	cropname54	cropname55	
cropname56	cropname57	cropname58	cropname59	cropname60	
cropname61	cropname62	cropname63	cropname64	cropname65	
cropname66	cropname67	cropname68	cropname69	cropname70	
cropname71	cropname72	cropname73	cropname74	cropname75	
cropname76	cropname77	cropname78	cropname79	cropname80	

cropname81 cropname82 cropname83 cropname84 cropname85
cropname86 cropname87 cropname88

R2 Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest
Forest Forest Forest Forest Forest Forest Forest2 Forest2 Forest2 Forest2 Forest2
Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2
Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2
Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2
Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2
Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2
Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2 Forest2

R1 Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest
Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest
Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest
Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest
Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest
Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest
Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest
Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest Forest

Appendix B
Model Calibration Tables

Table 6 Raven model calibration parameters for catchment C31

catchment	process group	description	parameter	units	min	max
31	soils	topsoil depth at top of slope	D_FF_slp1	m	0.01	0.5
31	soils	topsoil depth at mid slope	D_FF_slp2	m	0.01	1
31	soils	topsoil depth at bottom slope	D_FF_slp3	m	0.01	0.7
31	soils	shallow soil depth at bottom slope	D_AT_slp3	m	0.01	0.7
31	soils	deep soil depth at mid slope	D_BT_slp2	m	0	20
31	soils	topsoil hydraulic conductivity	HYDCOND_FF	mm/d	864	8640
31	soils	shallow soil hydraulic conductivity	HYDCOND_AT	mm/d	0.0864	4320
31	soils	deep soil hydraulic conductivity	HYDCOND_BT	mm/d	0.0864	4320
31	soils	topsoil PET correction factor	PET_COR_FF	% [0..1]	0.5	1.5
31	soils	topsoil field capacity	FC_FF	% [0..1]	0.2	0.7
31	soils	shallow soil porosity	PORO_AT	% [0..1]	0.5	0.75
31	soils	deep soil wilting point	WILT_BT	% [0..1]	0.00001	0.1
31	soils	topsoil baseflow power coefficient	BASEFLOW_N_FF	[-]	1	10
31	soils	deep soil baseflow power coefficient	BASEFLOW_N_BT	[-]	1	10
31	vegetation	Ma ximum storage capacity for rain	MAX_CAPACITY	mm	0	10
31	climate	rain/snow halfway transition temperature	RAINSNOW_TEMP	C	-1	1
31	snow	water saturation fraction of snow	SNOW_SWI	% [0..1]	0.04	0.07
31	snow	maximum snow melt factor used in degree day models	MELT_FACTOR	mm/d/C	3	6
31	snow	minimum snow melt factor used in degree day models	MIN_MELT_FACTOR	mm/d/C	1	3
31	climate	gauge correction for rainfall	RAIN_CORRECTION	% [0.5..1.5]	0.8	1.2
31	climate	gauge correction for snowfall	SNOW_CORRECTION	% [0.5..1.5]	0.8	1.2

Table 7 Robin model calibration parameters for catchment C31

catchment	process group	description	units	parameter	min	max
31	vegetation	radiation use efficiency	g/MJ	rue	3	7
31	vegetation	base temperature for growth	C	base_temp	-5	10
31	vegetation	optimal temperature for growth	C	opt_temp	12	25
31	vegetation	fractional NPP allocation to roots during youth	% [0..1]	phi_r_min	0.05	0.7
31	vegetation	fractional NPP allocation to roots during maturity*	% [0..1]	phi_r_max	0.05	0.7
31	vegetation	minimum mortality rate	1/year	lambda_nv_min	0.00001	0.001
31	vegetation	maximum mortality rate*	1/year	lambda_nv_max	0.00001	0.001
31	vegetation	power coefficient in mortality rate function*	-	aplha_nv_power1	0.644	1
31	vegetation	fraction of stem loss in total mortality	% [0..1]	lambda_nv_stems	0	1
31	vegetation	loss of stem mass per tree @ 1000 trees/hectare*	kg/tree	w_stems_max_1000nv1	45	500
31	vegetation	minimum ET coefficient	-	kc_min	0.5	1
31	vegetation	maximum ET coefficient	-	kc_max	1	1.5
31	vegetation	LAI when kc = kc_min	-	laimin	0.5	1
31	vegetation	LAI when kc = kc_max		laimax	4.3	6.2
31	vegetation	specific leaf area at maturity	-	sla_min	5	25
31	vegetation	specific leaf area at age 0	-	sla_max	5	25
31	vegetation	coefficient in DBH and stem mass relationship	-	alpha_dbh_a	0.1	10
31	vegetation	power coefficient in DBH and stem mass relationship	-	alpha_dbh_n	1	5
31	vegetation	vegetation stress factor	-	alpha_et	0.5	1

Table 8. Raven model calibration parameters for catchment C32

catchment	process group	description	parameter	units	min	max
32	soils	topsoil depth at mid slope	D_FF_slp2	m	0.01	1
32	soils	topsoil depth at bottom slope	D_FF_slp3	m	0.01	0.7
32	soils	shallow soil depth at mid slope	D_AT_slp2	m	0.3	1.2
32	soils	shallow soil depth at bottom slope	D_AT_slp3	m	0.01	0.7
32	soils	deep soil depth at bottom slope	D_BT_slp3	m	0	20
32	soils	topsoil PET correction factor	PET_COR_FF	% [0..1]	0.5	1.5
32	soils	topsoil field capacity	FC_FF	% [0..1]	0.2	0.7
32	soils	shallow soil porosity	PORO_AT	% [0..1]	0.5	0.75
32	soils	shallow soil hydraulic conductivity	HYDCOND_AT	mm/d	0.0864	4320
32	soils	deep soil hydraulic conductivity	HYDCOND_BT	mm/d	0.0864	4320
32	soils	topsoil baseflow power coefficient	BASEFLOW_N_FF	[-]	1	10
32	soils	shallow soil baseflow power coefficient	BASEFLOW_N_AT	[-]	1	10
32	soils	deep soil baseflow power coefficient	BASEFLOW_N_BT	[-]	1	10
32	climate	rain/snow halfway transition temperature	RAINSNOW_TEMP	C	-1	1
32	snow	water saturation fraction of snow	SNOW_SWI	% [0..1]	0.04	0.07
32	snow	maximum snow melt factor used in degree day models	MELT_FACTOR	mm/d/C	3	6
32	snow	minimum snow melt factor used in degree day models	MIN_MELT_FACTOR	mm/d/C	1	3
32	snow	maximum refreeze factor used in degree day models	REFREEZE_FACTOR	mm/d/C	0.01	4
32	snow	HBV snowmelt aspect correction (AM in HBV-EC)	HBV_MELT_FACTOR	[-]	0.1	1
32	vegetation	maximum canopy storage capacity for rain	MAX_CAPACITY	mm	0	10
32	climate	gauge correction for rainfall	RAIN_CORRECTION	% [0.5..1.5]	0.8	1.2
32	climate	gauge correction for snowfall	SNOW_CORRECTION	% [0.5..1.5]	0.8	1.2

Table 9. Robin model calibration parameters for catchment C32

catchment	process group	description	units	parameter	min	max
32	vegetation	light extinction coefficient	-	extcoef	0.4	0.6
32	vegetation	radiation use efficiency	g/MJ	rue	3	7
32	vegetation	base temperature for growth	C	base_temp	-5	10
32	vegetation	optimal temperature for growth	C	opt_temp	12	25
32	vegetation	maximum temperature for growth	C	max_temp	25	48
32	vegetation	fractional NPP allocation to roots during youth	% [0..1]	phi_r_min	0.1	0.7
32	vegetation	fractional NPP allocation to roots during maturity	% [0..1]	phi_r_max	0.1	0.7
32	vegetation	minimum mortality rate	1/year	lambda_nv_min	0.00001	0.001
32	vegetation	maximum mortality rate	1/year	lambda_nv_max	0.00001	0.001
32	vegetation	power coefficient in mortality rate function	-	aplha_nv_power1	0.644	1
32	vegetation	fraction of stem loss in total mortality	% [0..1]	lambda_nv_stems	0	1
32	vegetation	loss of stem mass per tree @ 1000 trees/hectare*	kg/tree	w_stems_max_1000nv1	45	500
32	vegetation	power coefficient in self thinning rule	-	alpha_n_max	1.174	2.1788
32	vegetation	minimum ET coefficient	-	kc_min	0.5	1
32	vegetation	maximum ET coefficient	-	kc_max	1	1.5
32	vegetation	LAI when kc = kc_min	-	laimin	0.5	1
32	vegetation	LAI when kc = kc_max		laimax	4.3	6.2
32	vegetation	specific leaf area at maturity	-	sla_min	5	25
32	vegetation	specific leaf area at age 0	-	sla_max	5	25
32	vegetation	coefficient in DBH and stem mass relationship	-	alpha_dbh_a	0.1	10
32	vegetation	power coefficient in DBH and stem mass relationship	-	alpha_dbh_n	1	5
32	vegetation	vegetation stress factor	-	alpha_et	0.5	1

Table 10. Raven model calibration parameters for catchment C35

catchment	process group	description	parameter	units	min	max
35	soils	topsoil depth at mid slope	D_FF_slp2	m	0.01	1
35	soils	topsoil depth at bottom slope	D_FF_slp3	m	0.01	0.7
35	soils	shallow soil depth at mid slope	D_AT_slp2	m	0.3	1.2
35	soils	shallow soil depth at bottom slope	D_AT_slp3	m	0.01	0.7
35	soils	deep soil depth at mid slope	D_BT_slp3	m	0	20
35	soils	deep soil depth at bottom slope	D_BT_slp3	m	0	20
35	soils	topsoil PET correction factor	PET_COR_FF	% [0..1]	0.5	1.5
35	soils	topsoil field capacity	FC_FF	% [0..1]	0.2	0.7
35	soils	shallow soil field capacity	FC_AT	% [0..1]	0.2	0.7
35	soils	shallow soil porosity	PORO_AT	% [0..1]	0.5	0.75
35	soils	topsoil wilting point saturation	WILT_FF	% [0..1]	0.0001	0.1
35	soils	shallow soil hydraulic conductivity	HYDCOND_AT	mm/d	0.0864	4320
35	soils	deep soil hydraulic conductivity	HYDCOND_BT	mm/d	0.0864	4320
35	soils	topsoil baseflow power coefficient	BASEFLOW_N_FF	[-]	1	10
35	soils	shallow soil baseflow power coefficient	BASEFLOW_N_AT	[-]	1	10
35	soils	deep soil baseflow power coefficient	BASEFLOW_N_BT	[-]	1	10
35	climate	rain/snow halfway transition temperature	RAINSNOW_TEMP	C	-1	1
35	snow	water saturation fraction of snow	SNOW_SWI	% [0..1]	0.04	0.07
35	snow	maximum snow melt factor used in degree day models	MELT_FACTOR	mm/d/C	3	6
35	snow	minimum snow melt factor used in degree day models	MIN_MELT_FACTOR	mm/d/C	1	3
35	snow	maximum refreeze factor used in degree day models	REFREEZE_FACTOR	mm/d/C	0.01	4
35	vegetation	maximum canopy storage capacity for rain	MAX_CAPACITY	mm	0	10
35	climate	gauge correction for rainfall	RAIN_CORRECTION	% [0.5..1.5]	0.8	1.2
35	climate	gauge correction for snowfall	SNOW_CORRECTION	% [0.5..1.5]	0.8	1.2