In lieu of the paired catchment approach: Hydrologic model change detection at the catchment scale

Nicolas Zégre,[1] Arne E. Skaugset,[2] Nicholas A. Som,[3] Jeffrey J. McDonnell,[2] and Lisa M. Ganio[3]

Received 2 September 2009; revised 22 July 2010; accepted 6 August 2010; published 23 November 2010.

[1] The paired catchment approach has been the predominant method for detecting the effects of disturbance on catchment-scale hydrology. Notwithstanding, the utility of this approach is limited by regression model sample size, variability between paired catchments, type II error, and the inability of locating a long-term suitable control. An increasingly common practice is to use rainfall-runoff models to discern the effect of disturbance on hydrology, but few hydrologic model studies (1) consider problems associated with model identification, (2) use formal statistical methods to evaluate the significance of hydrologic change relative to variations in rainfall and streamflow, and (3) apply change detection models to undisturbed catchments to test the approach. We present an alternative method to the paired catchment approach and improve on stand-alone hydrologic modeling to discern the effects of forest harvesting at the catchment scale. Our method combines rainfall-runoff modeling to account for natural fluctuations in daily streamflow, uncertainty analyses using the generalized likelihood uncertainty estimation method to identify and separate hydrologic model uncertainty from unexplained variation, and GLS regression change detection models to provide a formal experimental framework for detecting changes in daily streamflow relative to variations in daily hydrologic and climatic data. We include statistical analyses of climate variation and a two-part evaluation to explore model performance and account for unexplained variation. Evaluations consisted of applying our method to a control catchment and to a period prior to harvesting in a treated catchment to demonstrate that our method was capable of capturing the absence of land use change in an undisturbed catchment and capturing the absence of land use change during a period of no disturbance in the harvested catchment. In addition, we explore the sensitivity of our method to model identification, number of simulations, and likelihood thresholds for model identification. We show that an increase in the number of model simulations does not necessarily result in increased change detection performance. Our method is a potentially useful alternative to the paired catchment approach where reference catchments are not possible and to stand-alone hydrologic modeling for detecting the effects of land use change on hydrology.

Citation: Zégre, N., A. E. Skaugset, N. A. Som, J. J. McDonnell, and L. M. Ganio (2010), In lieu of the paired catchment approach: Hydrologic model change detection at the catchment scale, *Water Resour. Res.*, *46*, W11544, doi:10.1029/2009WR008601.

1. Introduction

[2] The impact of forest management on catchment-scale hydrology has remained a central research interest in water resources for several decades [e.g., *Bates*, 1921; *Bosch and Hewlett*, 1982; *Moore and Wondzell*, 2005]. There is little doubt of the relationship between forests and hydrology [*Eisenbies et al.*, 2007], but the sustainability of water resources depends on understanding both the process-level changes in runoff generation, storage, and movement of water through catchments and the ability to detect changes that occur following disturbance. Here we focus on the second: the ability to detect catchment-level changes in hydrologic behavior following disturbance.

[3] In this study we present a potentially useful alternative to the paired catchment approach to detect the effects of disturbance on catchment hydrology. The proposed change detection method combines hydrologic modeling, uncertainty analysis, and time series regression to isolate the effects of forest harvesting from the large natural variability of daily streamflow. The specific objectives of this study are (1) to use hydrologic modeling to account for natural fluctuations in daily streamflow, (2) to use model uncertainty analysis to identify and separate model uncertainty from unexplained variation, (3) to provide a formal experimental framework for

[1] Division of Forestry and Natural Resources, West Virginia University, Morgantown, West Virginia, USA.
[2] Department of Forest Engineering, Resources, and Management, Oregon State University, Corvallis, Oregon, USA.
[3] Department of Forest Ecosystems and Society, Oregon State University, Corvallis, Oregon, USA.

0043-1397/10/2009WR008601

detecting change in time series data, (4) to explore the sensitivity of model change detection to different likelihood thresholds, and (5) to explore the influence of simulation size on change detection. There are a variety of terms used interchangeably in statistical and hydrologic modeling literature. We define the following terms for clarity of our presentation: model uncertainty is the variations in streamflow simulations resulting from errors in model structure, model identification, and observed input data; statistical uncertainty is the variation of statistical model parameters given the parameters are estimated using samples of data and the lack of model fit to observed response and explanatory variables (residuals); and residuals, $\hat{\epsilon}_j$, are the difference between observations and predicted values ($y_j - \hat{y}_j$) from regression.

2. Literature Review

[4] Discerning the hydrologic responses of catchments to natural and anthropogenic disturbances are based either on the paired catchment approach (section 2.1) or on the use of rainfall-runoff hydrologic models (section 2.2), but there is no consensus on the most satisfactory approach [*Hewlett*, 1971; *Schnorbus and Alila*, 2004; *Alila et al.*, 2009]. Below we review how each method is used to detect change at the catchment scale.

2.1. Methods Based on Paired Catchment Approach

[5] The paired catchment approach [*Hewlett*, 1971] has been the predominant method for detecting the effects of forest management of hydrology, starting with first experiment at Wagon Wheel Gap, Colorado, in 1910 [*Bates*, 1921]. Reviews of this approach are given by *Bosch and Hewlett* [1982], *Stednick* [1996], *Moore and Wondzell* [2005]. The paired catchment approach establishes statistical relationships for catchment outlet responses (e.g., peak flow) between two paired catchments during a calibration period, where both catchments are undisturbed. Ideally, catchments are similar in size, locale, and share similar land use, climate, and physiographic attributes. Following calibration, land use treatments are applied to one catchment (treated), while the other catchment remains unchanged (control) and hydrologic differences between them are indicative of treatment effect. The control catchment serves as a climatic standard to account and correct experimental results for meteorological influences.

[6] The standard approach for detecting change in the paired catchment design is through ordinary least-squares (OLS) regression. If observations are temporally scaled such that they are independent, the regression model is

$$y_j = \beta_0 + x_j\beta_1 + \epsilon_j, \tag{1}$$

where y_j is the observed response variable at time j from the treated catchment; x_j is observed explanatory variable at time j from the control catchment; β's are coefficients to be estimated by regression; and ϵ_j is error at time j, where $\epsilon \sim N(0, \sigma^2)$.

[7] This approach includes the following steps: (1) fit regression model (equation (1)) to predisturbance observations from each catchment (calibration); (2) use calibration model to compute prediction intervals and to predict streamflow in the treated catchment based on observations from the control during the postdisturbance period; (3) calculate model residuals from equation (1) as $y_j - (\hat{y}_j|x_j)$, where y_j and $(\hat{y}_j|x_j)$ are observed and predicted values for time j; and (4) perform change detection on residuals by evaluating the proportion of postdisturbance residuals that exceed prediction limits. A disproportionate number of postdisturbance residuals exceeding prediction limits is an indication of treatment effects [*Harr et al.*, 1979]. For example, using 95% prediction intervals, significant changes are detected if >5% of the postdisturbance residuals exceed the prediction limits.

[8] Important among several limitations of this approach are the relatively few samples used for model development and chronological pairing of events. Regression change detection requires normally distributed and temporally independent residuals, homoscedastic variance, and linearity between response and explanatory variables [*Ramsey and Schafer*, 2002]. In many cases these approaches are applied to hydrologic response variables measured at annual [*Harris*, 1977] or storm-based [*Rothacher*, 1973] time steps when observations are likely to meet the requirements of regression. However, small sample size in model development can inflate type I and type II error potentially obfuscating the true effects of disturbances on hydrologic behavior. *Alila et al.* [2009] discussed problems that arise from chronological pairing of events and demonstrated how inappropriate pairing and statistical analysis resulted in incorrect estimates of changes in flood magnitude because neither the pairing or tests account for changes in variance or flood frequency. Chronological pairing of floods is difficult because storms do not always coincide in time, duration, intensity, or spatial extent [*Thomas and Megahan*, 1998] between the paired catchments.

[9] Advances in computational power and measurement technology have permitted the development of regression models using data collected at great frequencies [e.g., *Jackson et al.*, 2001; *Swank et al.*, 2001] but only a few studies address problems associated with autocorrelation. For example, *Watson et al.* [2001] used OLS regression to study the effects of forest harvesting on monthly streamflow by adjusting residuals using a first-order (AR1) autocorrelation filter prior to computing prediction limits. Generalized least-squares (GLS) regression was used by *Gomi et al.* [2006] to fit preharvest daily stream temperatures in British Columbia and by *Zégre* [2009] to daily streamflow models in Oregon to evaluate the effects of forest harvesting in headwater catchments. In both studies, model residuals were adjusted to account for up to three orders (AR3) of autocorrelation.

[10] The primary objective of time series modeling is to detect and describe all sources of variation in the given sequence of observations [*Machiwal and Jha*, 2008]. However, a trade-off between the frequency of streamflow observations used to develop time series models and the ability to detect changes in hydrology was reported by *Zégre* [2009]. *Zégre* [2009] was not able to detect changes in runoff following clear-cut harvesting (31% to 65% basin harvested) in several experimental headwater catchment using daily GLS models. However, significant increases in runoff were detected using monthly streamflow regression models. It was hypothesized that the inability to detect changes in runoff, when it was plausible to expect such increases (area harvested >20% [*Bosch and Hewlett*, 1982]), was related to not fully accounting for all sources of variation between paired catchments. The ability to confidently detect change following disturbance using regression in the paired catchment approach is a function of discerning the degree of intercatchment and intracatchment variability of processes and timing between two

paired catchments. *Zégre* [2009] attributed the unaccounted-for variation to climatic heterogeneity (e.g., rainfall distribution) and to the timing of runoff processes (e.g., catchment residence time) that desynchronized measured streamflow at the catchment outlets.

2.2. Methods Based on Hydrologic Models

[11] It is often impractical to identify and establish suitable control catchments with respect to time, expense, location, and market pressures. An increasingly common practice is to use hydrologic models to discern the effects of disturbance on catchment hydrology. Hydrologic models can be a useful alternative to the paired catchment to detect changes under various disturbances, climatic and physiographic conditions, and spatial scales [*Andréassian et al.*, 2003]. While hydrologic models have been widely used to evaluate hydrologic response to land use change, few change detection studies consider hydrologic model uncertainty or incorporate hydrologic modeling into formal statistical frameworks for change detection. Further, we are not aware of studies that have explored the influence of model simulation size on model change detection.

[12] Two common applications of hydrologic models in land use/land change studies are to (1) detect changes in hydrology by comparing simulated and observed streamflow [e.g., *Lørup et al.*, 1998] and (2) use the model to simulate changes in hydrology following disturbance [e.g., *Post et al.*, 1996]. The first approach (referred to as virtual control) consists of calibrating a before-treatment model based on observed rainfall and streamflow for use as a virtual control catchment during the postdisturbance period to reconstruct streamflow. Treatment effects are discerned as the differences between observed and simulated streamflow. An underlying assumption is that catchment behavior is stationary in both periods, but this assumption is seldom tested. Several examples of this application follow. (1) *Brandt et al.* [1988] used a semidistributed rainfall-runoff model to simulate the effects of clear-cut harvesting in small catchments in Sweden by calibrating the model to observed streamflow prior to harvesting and used the calibrated parameters and precipitation after clear-cutting to simulate streamflow. The study did not include statistical change detection or uncertainty analysis. (2) *Lavabre et al.* [1993] used a lumped rainfall-runoff model to assess the hydrologic response of a small Mediterranean catchment to wildfire. They calibrated a two-parameter model using predisturbance monthly streamflow and used it reconstruct the streamflow following wildfire. Change detection was assessed by calculating the residual difference between observed and simulated streamflow during the postfire period and did not include statistical methods or analyses of model uncertainty. (3) *Bowling et al.* [2000] used a distributed mechanistic model to evaluate the effects of forest harvesting and forest roads on peak flow events. The model was calibrated using preharvest climate, streamflow, and land cover conditions and used to reconstruct stormflow response during postharvest periods. Hydrologic changes were assessed as the difference between model simulations based on a single optimized parameter set and observed runoff for each storm and tested using the Mann-Kendall nonparametric test. Model uncertainty was not demonstrated in this study. (4) *Schreider et al.* [2002] used a lumped rainfall-runoff model to detect the streamflow response to farm-dam development in nine catchments in Australia. They developed regional models with several parameter sets calibrated in different agricultural catchments to simulate streamflow at the beginning of their study and used precipitation data to simulate streamflow over the entire study period based on the regional models. Residual errors were calculated as the difference between observed streamflow and modeled streamflow, simulated from different models. Though this study described variation in simulated streamflows vis-á-vis varying model parameters, it did not provide statistical tests to evaluate the credibility of the trend in residuals.

[13] The second approach consists of calibrating different models that resemble streamflow during different periods. In this case, a model is calibrated to predisturbance streamflow conditions using observed streamflow and precipitation, and a new model is calibrated to postdisturbance conditions using observed streamflow and precipitation from the postdisturbance period. In the virtual reference discussed above, the model is used to reconstruct streamflow, where as this second approach uses the model to simulate changes in hydrology that result from disturbance. Two examples follow. (1) *Kuczera et al.* [1993] tested the structure of two lumped parameter catchment-scale models by forcing the models to describe the effects of strip thinning of mountain ash on water yield. Both models were calibrated during a pretreatment period with 100% forest cover. The model parameters that describe interception and evapotranspiration were modified to account for the effect of strip thinning on water yield. Although this approach does not provide formal statistical tests to detect change, it does provide an interesting means of generating hypotheses concerning process-level changes in catchment hydrology. (2) *Post et al.* [1996] used a lumped rainfall-runoff model to simulate the hydrologic response of Picaninny Creek to clear-cut harvesting. Daily rainfall, air temperature, and streamflows from 36 years were used to calibrate seven preharvest models, one during-harvest model, and 10 postharvest models. Changes in hydrology were discerned by comparing preharvest and postharvest model parameter sets and volumetric changes in streamflow.

[14] *Seibert and McDonnell* [2010] and *Seibert et al.* [2010] combined both approaches discussed above to detect and model forest harvesting and wildfire effects on streamflow using a rainfall-runoff model applied to the H. J. Andrews (Oregon) and Entiat (Washington) Experimental Forests. Also introduced in these studies is the use of a model-to-model approach that compares runoff simulated with parameter sets calibrated for periods before and after land cover changes. This third approach allows assessment of integrated catchment behavior rather than single parameters [*Seibert and McDonnell*, 2010]. In both studies, Monte Carlo methods were used to generate large populations of parameter sets and estimate parameter uncertainty. The nonparametric Wilcoxon rank-sum test was used to test for changes.

3. Conceptual Framework

[15] In this section we present a general conceptual framework of our proposed method. Our approach differs from previous methods in that we combine hydrologic modeling, uncertainty analysis, and regression change detection to isolate the effects of forest harvesting from input errors, model identification, and the large natural variability attributed to

daily streamflow. We use a hydrologic model in the virtual reference approach to reconstruct streamflow.

3.1. Model Selection and Generation of Realizations

[16] The selection of model structure is dictated by the specific research objectives. For example, if the goal is to generate hypotheses of how streamflow generation processes change following disturbance, a more complex, mechanistic model is appropriate. If the objective is to detect volumetric changes in streamflow at the outlet of a catchment, a simple rainfall-runoff model is more appropriate. The critical point is to select a model that captures the temporal variation and stability of the rainfall-runoff relationship in a specific catchment.

[17] Iterative methods can be used to find an optimal set(s) of model parameters and sensitivity analysis can be used to evaluate the influence of individual parameters on model performance. However, uncertainties that arise from model identification, structure, boundary conditions, and measurement error make the search for a optimal parameter set unrealistic or impractical in most hydrologic systems [*McMichael et al.*, 2006]. Iterative methods, such as Monte Carlo, can be used to randomly sample across parameter ranges to generate a large number of model realizations that are used to explore parameter sensitivity and generate populations of acceptable models.

3.2. Selection of Acceptable Models

[18] Performance of model realizations is evaluated on the basis of a likelihood-based measure of the agreement between observed and simulated streamflow. Models that produce acceptable simulations based on the defined likelihood measure are retained as behavioral models. Acceptance of model realizations based on a likelihood criteria assumes that population of models represents the likely behavior of the catchment under scrutiny. By focusing our analysis on a population of behavioral models rather than a single optimized parameter set, we acknowledge the coexistence of alternative models that perform equally well when compared to observed calibration data.

3.3. Determination of Uncertainty Bounds

[19] Simulation uncertainty bounds can be placed around observed streamflow to estimate how well the population of behavioral models represent the systems under consideration. Simulated streamflows for each time step are ranked in ascending order to give a cumulative distribution of streamflow for each time step. Quantiles are then calculated based on the cumulative distributions to represent model uncertainty at each time step. For example, simulated streamflow from the upper 97.5th and lower 2.5th percentiles per day would be used to calculate 95% uncertainty bounds for a daily streamflow model. By plotting observed daily streamflow and the upper and lower simulations for each time step, we create uncertainty bounds that show when the hydrologic model over- or undersimulates relative to observed streamflow.

3.4. Discernment of Hydrologic Changes

[20] Unlike the paired catchment approach which relies on formal statistical approaches to detect change, hydrologic modeling studies seldom include statistical tests to evaluate the significance of hydrologic change. Where present, the focus has been on nonparametric methods that offer limited information other than differences in central tendency (e.g., Wilcoxon test), distribution, or changes in variance (e.g., Kolmogorov-Smirnov) [*Helsel and Hirsch*, 1992]. The final step in our approach is to use regression methods to evaluate the significance of hydrologic change relative to variations of measured rainfall and runoff time series and uncertainty in modeled streamflow. To achieve this we extend equation (1) to allow for temporally autocorrelated residuals from our daily streamflow models and use behavioral model simulations as explanatory variables in our regression models (Figure 1) to ascertain the significance of hydrologic change.

4. Methods

4.1. Study Area and Hydrometric Data

[21] Fenton and DeMersseman Creeks are headwater catchments of the Hinkle Creek Paired Watershed Study located on the foothills of the west slope of the southern Oregon Cascades Mountains, USA (Figure 2). Fenton Creek (referred to as treated catchment) is a 0.23 km^2 catchment with elevations that range from 615 to 815 m and slopes that range from 12% to 30%. DeMersseman Creek (control) is a 1.56 km^2 catchment with elevations that range that from 650 to 1,260 m and slopes that range from 14% to 54%. Hinkle Creek is located in the rain-snow transitional zone with a climate dominated by frontal Pacific storms from November to May and dry, warm conditions during the remainder of the year. Mean annual precipitation at 839 m is approximately 1,800 mm, with a mean annual temperature of 8.5°C. Fenton Creek is entirely forested while DeMersseman Creek is mostly forested, with approximately 3% of catchment area in roads. In both catchments, slopes are forested by 60 year old, harvest-regenerated stands of Douglas fir (*Pseudotsuga menziesii*). Riparian vegetation in higher order stream networks is composed mainly of overstory species such as red alder (*Alnus rubra*) and understory species such as huckleberry (*Vaccinium parvifolium*) and sword fern (*Polystichum munitum*) while low-order headwater catchments are dominated by Douglas fir.

[22] Contemporary forest harvesting methods, defined by the Oregon Forest Practice Rules, were used to clear-cut harvest trees in the treated catchment, while the control catchment remained untouched. Harvesting started in mid-July 2005 and continued through January 2006 and removed trees from 65% (0.15 km^2) of the watershed area in Fenton Creek.

[23] Streamflow was recorded at 10 min intervals from November 2003 through January 2008 at the outlet of each catchment using modified Parshall flumes equipped with automated Druck 1830 pressure transducers and Campbell Scientific CR 10x dataloggers. Climate data were measured at a micrometeorological station located between the North Fork and South Fork experimental basins (Figure 2). Air temperature, relative humidity, precipitation, wind speed, and photosynthetic active radiation were recorded at 10 min intervals from December 2003 through January 2008. Long-term climate data from a nearby meteorological station (Idleyld Park COOP ID 354126) were used to supplement climatic data prior to monitoring in the experimental catchments.

4.2. Hydrologic Model Description

[24] Rainfall-runoff models have been widely used to detect catchment-scale changes in hydrology [e.g., *Brandt et al.*, 1988; *Post et al.*, 1996]. We chose a model with

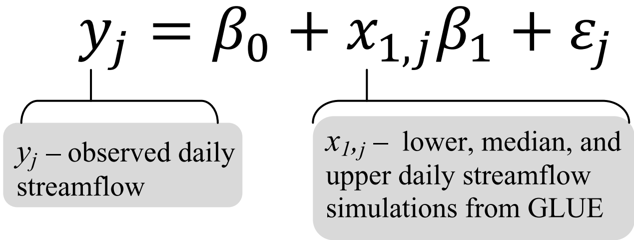

Figure 1. Regression model construct for detecting effects of forest harvesting and characterizing the range of model uncertainty on daily streamflow. GLS regression models were developed using median and 95th percentile streamflow simulations identified using GLUE based on 850,000 Monte Carlo simulations. Daily observed streamflow were used as the response variable and simulated streamflow were used as the explanatory variable in three separate GLS regression models.

(1) readily available, low data requirements and whose (2) structure lent itself to efficiently address model sensitivity and uncertainty analysis.

[25] We used the HBV-EC model [*Hamilton et al.*, 2000], a modified version of the original HBV model [*Bergström*, 1995], to simulate daily streamflow during a preharvest period and to reconstruct streamflows following a period of forest harvesting. HBV-EC is a partially distributed conceptual model that simulates streamflow using daily time series of precipitation, air temperature, and long-term monthly potential evaporation [*Bergström*, 1995]. Basins are divided into subcatchments based on elevation and land use zones to represent lateral climatic and vegetation gradients across the catchment. The HBV model consists of three components (Table 1): a snow routine for snow accumulation and melt based on the degree-day method; a soil routine that controls the proportion of rainwater and snowmelt that generates excess water after considering soil moisture and evaporation requirements; and a runoff transfer routine that consists of an upper, nonlinear reservoir that represents fast discharge and a lower linear reservoir that represents slow discharge or baseflow. Detailed descriptions of the HBV model can be found, for example, in work by *Bergström* [1995] and *Seibert* [1997].

4.2.1. Model Identification, Calibration, and Uncertainty Analysis Using the Generalized Likelihood Uncertainty Estimation Method

[26] Experimental catchments were delineated using the EnSim Hydrologic modeling platform [*Canadian Hydraulics Centre*, 2006] using the Hinkle Creek, OR USGS 10-m digital elevation model. Catchments were divided into five 40 m elevation zones. A record of climate data for 1,618 days was distributed into three periods; a model spin-up period (123 days), model calibration period (533 days), and post-harvest period (962 days). Observed time series data of daily streamflow were used for model calibration and evaluation.

[27] Regional sensitivity analysis (RSA) [*Hornberger and Spear*, 1981] and the generalized likelihood uncertainty estimation (GLUE) [*Beven and Binley*, 1992] were used to characterize parameter sensitivity and model uncertainty. GLUE has been reviewed by *Beven* [2002] and is widely used in the hydrologic modeling literature [e.g., *Freer et al.*, 1996; *McMichael et al.*, 2006] due to its ease of implementation, computational efficiency, and flexibility of likelihood measure [*Blasone et al.*, 2008]. GLUE is a Bayesian-like parameter estimation method that rejects the concept that there is a globally optimal parameter set and that multiple parameter sets have an equal likelihood of providing acceptable simulations of the system in question. GLUE has been criticized for not being formally Bayesian [*Vrugt et al.*, 2009], subjective [*Blasone et al.*, 2008], and fails to produce uncertainty bounds that capture the precision of estimated parameters and differences between simulations and observations [*Stedinger et al.*, 2008]. However, recent studies have demonstrated similar estimates of model and parameter uncertainty using GLUE and formal Bayesian uncertainty approaches [*Vrugt et al.*, 2009; *Jin et al.*, 2010]. GLUE allows the conditional assessment of model uncertainty based on prior knowledge of model acceptability to identify distributions of behavioral models. Behavioral models are subjectively selected on the basis of likelihood criteria. We chose GLUE because of its flexibility in defining the likelihood definition that is used to define behavioral models that allows for having no assumptions about the error of our hydrologic model [*Jin et al.*, 2010].

[28] Behavioral models are used to construct probability distributions of model parameters and uncertainty bounds [*Beven and Binley*, 1992]. GLUE rescales the likelihood weights from behavioral models to yield cumulative distributions of streamflow simulations at the respective time step [*Freer et al.*, 1996]. Uncertainty bounds are calculated using the specified streamflow percentiles. Although GLUE provides estimates for uncertainty due to parameter sampling, it does not provide insight on uncertainty due to model structure or errors in input variables. Therefore the GLUE methodology is bound to underestimate total uncertainty.

[29] The Monte Carlo analysis toolbox (MCAT) [*Wagener et al.*, 2004] was used to implement a variation of RSA [*Freer et al.*, 1996] and GLUE to assess the impacts of parameter sensitivity and uncertainty on daily streamflow simulations using the HBV-EC model. Initial models were identified by generating 850,000 Monte Carlo simulations for each catchment conditioned on uniform parameter distributions for 15 parameters (Table 1). With little prior information about the likely distributions of each parameter, we choose to sample from independent uniform samples across parameter ranges [*Beven*, 2002]. Uncertainty bounds were based on the 95th percentiles (2.5th and 97.5th) streamflow simulations and conditioned on the posterior distributions for sensitive parameters identified from RSA and GLUE.

[30] The *Nash and Sutcliffe* [1970] likelihood measure, R_{eff}, was used to identify our populations of behavioral models. The specified level of rejection criteria is arbitrary. However, it is important to select a threshold to identify a population of behavioral models that captures the range of natural variability and model uncertainty. We used three different R_{eff} thresholds (0.3, 0.4, and 0.6) to evaluate the impact of behavioral model definition on our ability to detect change in the treated catchment only. However, we focus our discussion of the effects of forest harvesting on behavioral models defined by $R_{\mathrm{eff}} > 0.4$. R_{eff} is calculated by

$$R_{\mathrm{eff}} = 1 - \frac{\sum\left(Q_{\mathrm{obs}_j} - Q_{\mathrm{sim}_j}\right)^2}{\sum\left(Q_{\mathrm{obs}_j} - \overline{Q_{\mathrm{obs}}}\right)^2} \tag{2}$$

where Q_{obs_j} is observed streamflow on day j, $\overline{Q_{\mathrm{obs}}}$ is the mean observed streamflow, and Q_{sim_j} is the simulated streamflow

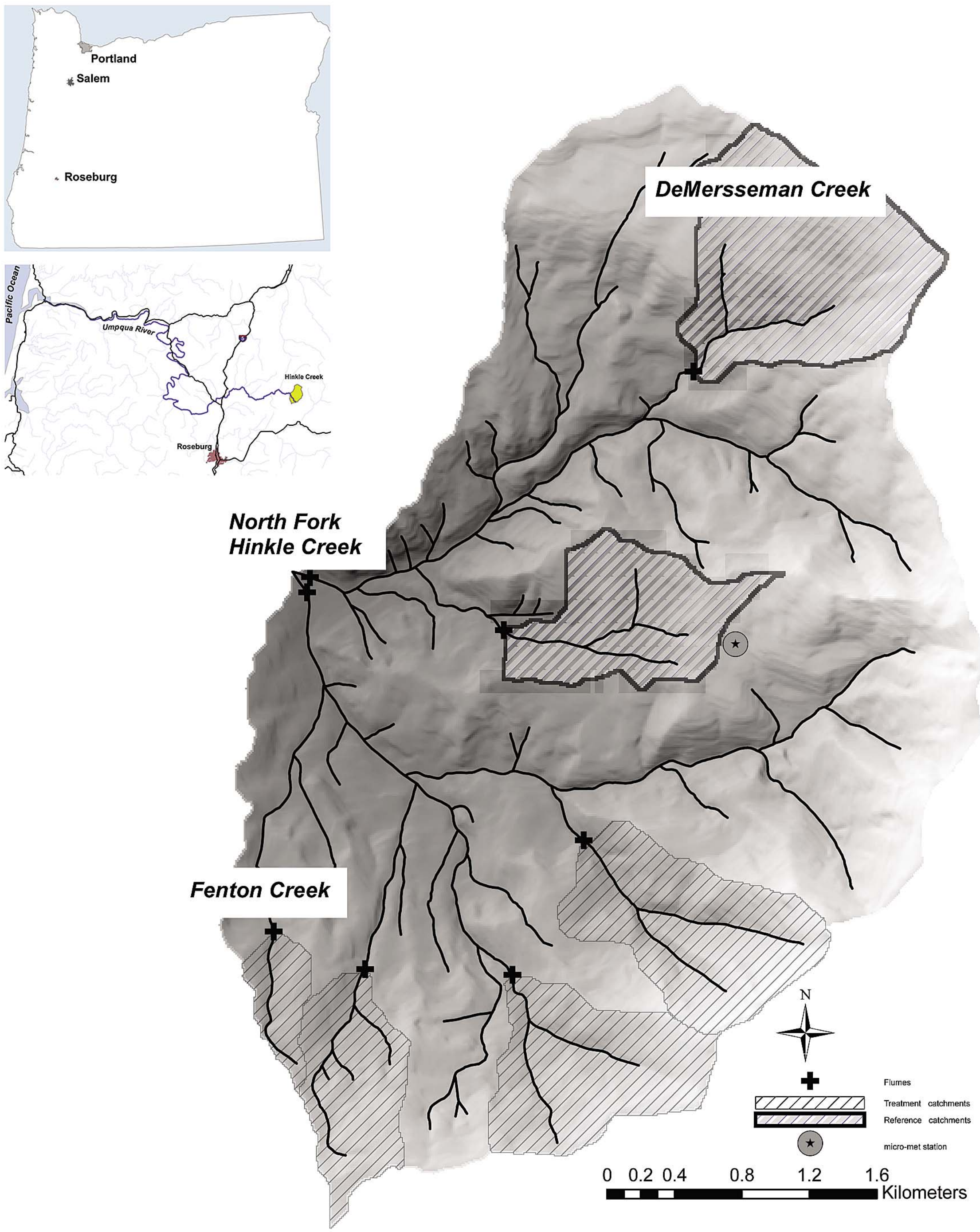

Figure 2. Location of the Hinkle Creek Paired Watershed Study (latitude 43°25′25″, longitude 123°02′17″). Fenton Creek is the treated catchment; DeMersseman Creek is the control catchment.

on day j from the hydrologic model. R_{eff} ranges from minus infinity to 1, with higher values indicating better agreement between observed and simulated streamflow [*Legates and McCabe*, 1999].

4.3. Detecting the Effects of Forest Harvesting on Daily Streamflow

[31] Our method of change detection relies heavily on the approach utilized in paired catchment studies. Whereas paired catchment studies fit regression models for preharvest data between control and treated catchments, we instead developed preharvest regression models between daily observed streamflow and simulated streamflow from the GLUE-generated uncertainty percentiles. To fit these statistical models we used GLS, because in situations where autocorrelated residuals occur, GLS regression is appropriate as it allows the independence assumption required of equation (1) to be relaxed and incorporates this correlation structure into the estimation of linear model parameters [*Myers*, 1990]. GLS

Table 1. Initial and Posterior Model Parameters From 850,000 Monte Carlo Simulations Using HBV-EC for Daily Streamflow[a]

			Initial Range		Posterior Range			
					Treated		Control	
Parameter	Explanation	Unit	Min	Max	Min	Max	Min	Max
Routing routine								
KF	Fast reservoir component	–	0	1	0.54	0.88	0.28	0.82
KS	Slow reservoir component	–	0	1	0.12	0.83	0.05	0.98
FRAC	Fraction of runoff directed to fast reservoir	–	0.1	1	0.12	0.9	0.19	0.9
α	Fast reservoir exponent	–	0	1	0.42	0.99	0.2	0.96
Soil routine								
FC	Soil field capacity	mm	100	400	120	360	150	380
β	Controls relationship between soil infiltration and soil water release	–	0	2	–	–	–	–
LP	Soil moisture content threshold where evaporation becomes limited	–	0	1	–	–	–	–
Snow routine								
DC	Snowmelt factor for summer solstice	°C mm^{-1} d^{-1}	0	4	0.16	3.8	0.9	3.1
CRFR	Controls rate at which liquid refreezes into snowpack	°C mm^{-1} d^{-1}	0	4	0.1	4	0.4	3.7
Climate								
RFCF	Rainfall correction factor	–	1	2	–	–	–	–
SFCF	Snowfall correction factor	–	1	1.5	1.2	1.5	1	1.5
PGRADL	Fractional increase in precipitation with elevation	–	0.001	0.01	–	–	–	–
TLAPSE	Temperature lapse rate	°C m^{-1}	0.006	0.01	–	–	–	–
TT	Threshold air temperature	°C	−1.5	2.5	–1.2	0.36	–0.76	2.4
ETF	Temperature anomaly correction of potential evapotranspiration	–	0	1	–	–	–	–

[a]Posterior parameter ranges presented for sensitive parameters are based on modified RSA [*Freer et al.*, 1996].

also allows us to use Akaike information criteria (AIC) [*Padmanabhan and Rao*, 1982] to determine the order of autoregressive (AR) models used to address the correlation structure in model variance [*Salas*, 1993].

[32] The results from Monte Carlo simulations can be processed in the regression framework in different ways. A first approach (approach a) is based on fitting an individual regression model to simulations for each behavioral model and aggregate the responses of all models to evaluate change. A second approach (approach b) aggregates simulations from all behavioral models into median and percentile streamflows that are then used in regression analyses to evaluate change. These approaches differ in that approach a gets more at the variations attributed to the sampling distributions of the different regression models, whereas approach b gets at the uncertainty of hydrologic model output and the influence of hydrologic model uncertainty on our ability to detect change. The impact of regression model variation on change detection have been discussed elsewhere [*Loftis et al.*, 2001]; here we use approach b to focus on the impact of hydrologic model uncertainty within the context of regression change detection.

[33] Preharvest regression models were developed using daily streamflow from 1 February 2004 to 17 July 2005 (533 days). Sinusoidal trigonometric terms can be included as covariates in time series regression to account for temporal dependencies such as lag or seasonality between responses but did not improve model performance and were not included in our models. This is attributed to climatic forcing in the rainfall-runoff model and the use of a single catchment instead of paired observations used in the traditional paired catchment approach. The median model was used to reconstruct streamflow during the postharvest period (18 July 2005 to 17 January 2008), without harvesting effects. The lower and upper models were used to explore how a hydrologic model that consistently under- or overestimates streamflow affects our ability to detect change using regression. These three model outputs were then used as predictors to estimate postharvest streamflow without harvesting effects using regression parameters estimated from the preharvest GLS regression models (Figure 1). The residuals from these statistical models were then used to discern treatment effects as described below.

[34] To evaluate the statistical evidence of treatment effects, we relied on the method described and applied by *Watson et al.* [2001] and *Gomi et al.* [2006] for paired catchment studies. This method begins by filtering model residuals using the AR time series model that leaves the estimated posttreatment innovations, $\hat{\mu}_j$:

$$\hat{\mu}_j = \hat{\epsilon}_j - \hat{\phi}_1\hat{\epsilon}_{j-1} - \hat{\phi}_2\hat{\epsilon}_{j-2} - \ldots - \hat{\phi}_k\hat{\epsilon}_{j-k}, \tag{3}$$

where $\hat{\phi}_i$ is the estimated autocorrelation coefficient of error term at lag k. From equation (1), let $\epsilon_j = [\epsilon_1, \epsilon_2, \epsilon_3, \ldots, \epsilon_j, \ldots, \epsilon_J]'$, where J is the total number of time steps. The jth residual, ϵ_j, is estimated as $\hat{\epsilon}_j = (y_j - \hat{y}_j)$, and ϵ_{j-k} is the residual error term k days before day j; ϕ_i is the estimated autocorrelation coefficient of error term at lag k. Obtaining innovations is important because under the null hypothesis of no treatment effects, innovations are independent and randomly distributed and $\hat{\mu}_j \sim N(0, \sigma^2)$, with the same σ^2 as estimated during the pretreatment phase of the experiment. The 95% prediction intervals, PI, at time j are calculated by

$$PI_j = 0 \pm 1.96\sqrt{\mathrm{Var}(\hat{\mu}_j)} \quad (4)$$

where $\mathrm{Var}(\hat{\mu}_j)$ is the prediction variance incorporating uncertainty associated with predicting future values, estimated linear model, and autocovariance parameters. Comparing these posttreatment innovations to the prediction intervals established based on the pretreatment period allows us to ascertain if a significant number of these innovations fall above our prediction limits, thus signaling a significant increase in runoff following forest harvest activities.

4.4. Change Detection Evaluation

[35] Control catchments are used in the paired catchment design to account for climatic variation between experimental catchments overtime. However, the use of controls in modeling studies are often overlooked; therefore the modeler must assume that catchment behavior would be the same overtime if there were no disturbances. Unaccounted-for variation from climatic nonstationarity or unknown natural shifts in hydrologic behavior can obfuscate or elevate the effects of disturbance, resulting in the rejection of the null hypothesis, H_0, when in fact it is true (type I error) or accepting H_0 when it is false (type II error).

[36] To assess the significance of climatic variability between the preharvest and postharvest periods, total precipitation and mean temperature were calculated for each month within each experimental phase. Monthly values of precipitation and temperature were temporally autocorrelated, and we therefore relied on GLS to compute t tests on total monthly precipitation and mean monthly temperature between these two periods. To model the temporal autocorrelation we considered a suite of AR models and chose AR structures based on AIC values of the GLS model fits.

[37] We used a two-part evaluation to explore model performance and account for unexplained variation. Specifically, we applied our method to a control catchment to demonstrate that our method was capable of capturing the absence of land use change in an undisturbed catchment (herein referred to as evaluation 1) and to a period prior to harvesting in the treated catchment (evaluation 2) to demonstrate that our method was capable of capturing the absence of land use change during a period of no land use change in this catchment. For evaluation 1, we imposed a hypothetical treatment period (Figure 3) similar to the period of harvesting in the treated catchment, to partition streamflow records into two testable populations. Since harvesting activities were hypothetical in the control catchment, we expected no difference between preharvest and postharvest periods. The detection of significant changes in the control would indicate errors in model structure, statistical errors, and/or unaccounted-for variation in climatic or hydrologic input data. For evaluation 2, a subset of the preharvest data (450 days; 1 February 2004 to 25 April 2005) for the treated catchment was used to develop new regression models based on the lower, median, and upper uncertainty streamflow time series. Change detection was performed on the remaining 80 days from the original preharvest data set (533 days) to show that our method was capable of capturing the absence of land use change. The two-part evaluation serves to test the assumption that, in the absence of disturbance, catchment behavior is similar through time.

5. Results

[38] Observed streamflow and median simulated streamflows from GLUE and precipitation for the study period are shown in Figure 3. Observed and simulated streamflow are consistently greater during the postharvest periods for the treated and control catchments. The most precipitation and highest peak discharge during the 2 year study occurred during the 2006 water year (October 1–September 31), which was the first year following harvesting in the harvest catchment. The two largest peak flow measured in the North Fork had recurrence periods of approximately 8 and 3 years (8 and 7 m^3 s^{-1}, respectively). Most of the storms had subannual recurrence periods.

5.1. Model Identification, Calibration, and Uncertainty Analysis

[39] Monte Carlo simulations were used to generate 850,000 candidate models for model sensitivity and uncertainty analyses. Of the 15 model parameters sampled, 9 showed sensitivity: KF, KS, FRAC, α, FC, DC, CRFR, SFCF, and TT (Table 1). The temperature threshold parameter TT and slow reservoir component KS showed the greatest sensitivity for predicting streamflow in our catchments.

[40] Uncertainty analysis was estimated using GLUE by randomly sampling parameter values conditioned on uniform parameter distributions for the nine sensitive parameters. Of the 850,000 models, 97,577 and 65,914 were retained for the treated and control catchments based on a rejection criteria $R_{\mathrm{eff}} > 0.4$ (Table 2). 146,824 and 1,332 behavioral models were identified for the treated catchment based on $R_{\mathrm{eff}} > 0.3$ and 0.6, respectively (the impact of different levels of R_{eff} on detecting change are discussed in section 6.2.3). Uncertainty bounds were calculated using the 2.5th and 97.5th simulated streamflow percentiles. Observed streamflow fell within the uncertainty bounds 81% and 73% of days during calibration period in the treated and control catchments, respectively (Figure 4). Observed streamflow that exceeded the 95th percentile uncertainty bounds accounted for approximately 4% (31 mm) and 5% (91 mm) of total annual streamflow in the treated and control catchments.

[41] Generally, the HBV-EC model under-simulated (observed streamflow >97.5th bounds) streamflow in the treated catchment during March and May, with the greatest errors by volume occurring in March 2005 (9 mm/~1%). In the control, under-simulated errors occurred during May and June, with the greatest errors by volume (49 mm/~3%) in June 2004 (Figure 5). Errors associated with over-simulation were consistently less, with the greatest error by volume occurring in May 2005 (6.7 mm/<1%) in the harvest catchment and in March 2005 (16 mm/1%) in the control (Figure 5).

5.2. Change Detection Results and Evaluation

[42] To ascertain the impact of climatic variability on changes in daily streamflow, we used GLS to compute t tests on total monthly precipitation and mean monthly temperature between the preharvest and postharvest periods. GLS models for precipitation and temperature exhibited, respectively,

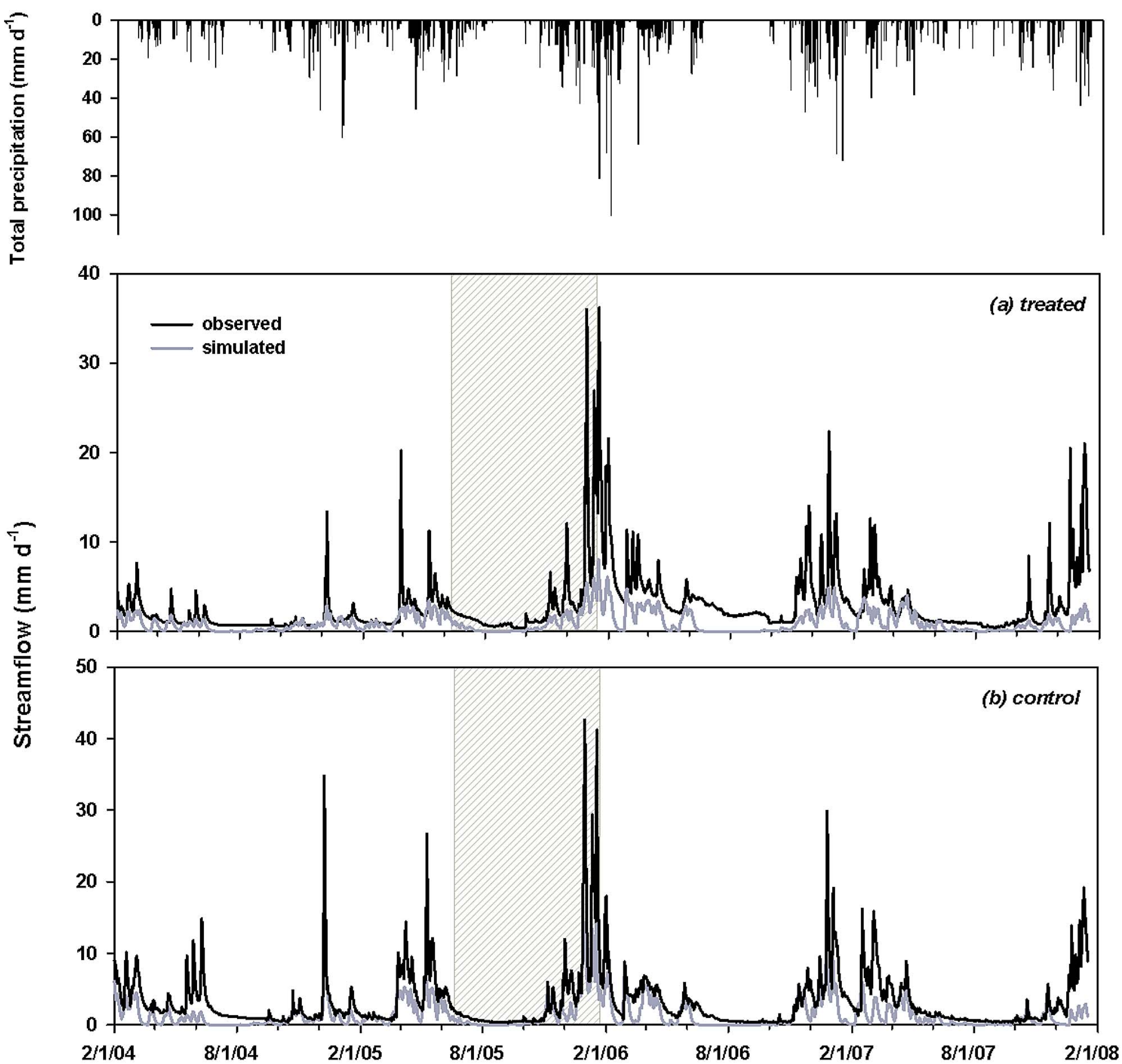

Figure 3. Daily precipitation and observed and simulated streamflow for (a) treated and (b) control catchments in the Hinkle Creek experimental catchments. Simulated streamflows based on median simulations from GLUE. Actual (treated) and hypothetical (control) harvesting periods shaded.

fourth (AR 4) and third order (AR 3) autoregressive variance structures. There was not statistical evidence to suggest that either total monthly precipitation (p value = 0.2910) or mean monthly temperature (p value = 0.2376) differed between the preharvest and postharvest periods.

[43] Posterior parameter distributions derived from RSA and GLUE were used to reconstruct daily streamflow in the treated catchment and to simulate daily streamflow in the control catchment. Three regression models were developed for each catchment using median simulations and the lower and upper uncertainty percentiles. AIC were used to identify parsimonious autocorrelation structures for each model. All regression models exhibited a second-order autoregressive (AR 2) variance structure; s_e of regression models ($R_{\mathrm{eff}} > 0.4$) were 0.003 and 0.04 for treated and control catchments, respectively. Results of the regression and autoregressive modeling are summarized in Table 2.

5.2.1. Detecting the Effects of Forest Harvesting in the Treated Catchment

[44] Nine percent of the postharvest median model innovations exceeded the 95% prediction intervals, suggesting statistically significant increases in daily streamflow (Table 2 and Figure 6). Lower and upper models, based on the 2.5th and 97.5th percentiles, were used to characterize the range of detectability under model uncertainty and similarly showed 9% of postharvest innovations exceededing the prediction limits (Table 2).

[45] Estimated changes in streamflow following forest harvesting were calculated as the residual difference between observed and predicted streamflow. Deviations between

Table 2. Relationship between Nash-Sutcliffe [*Nash-Sutcliffe*, 1970] R_{eff} Rejection Thresholds, Population of Behavioral Models, Standard Error of GLS Models, and Change Detection Under Three Different Levels of Hydrologic Model Uncertainty for the Treated and Control Catchments[a]

		GLS Model[d] (%)		
Threshold[b]	Behavioral Models[c]	Lower	Median	Upper
Treated				
0.30	146,824	8 (0.004)	8 (0.004)	8 (0.004)
0.40	97,577	9 (0.003)	9 (0.003)	9 (0.003)
0.60	1,332	10 (0.003)	10 (0.003)	10 (0.003)
Control				
0.40	65,914	4 (0.04)	4 (0.04)	4 (0.04)

[a]Statistically significant changes are detected when >5% of the postdisturbance innovations exceed the 95% prediction intervals.
[b]R_{eff} rejection criterion threshold.
[c]The number of behavioral models resulting from specified rejection criterion threshold.
[d]Proportion calculated as the ratio of postharvest innovations >95% prediction intervals to total population of postharvest innovations. Standard error of GLS regression model, s_e, is given in parentheses.

observed and predicted streamflow following harvesting varied by day, season, and year (Figure 7). Maximum daily streamflow increased by as much as 31 mm for each model. Average seasonal increases based on all models were greatest during winter (485 mm), followed by spring (146 mm), fall (114 mm), and then summer (100 mm) (Table 3).

5.2.2. Evaluation of the Change Detection Models

[46] A two-part evaluation was used to assess change detection model performance. The objective was to assess if our method was capable of capturing the absence of land use changes. Evaluation 1 consisted of applying our method to a control catchment, where no harvesting took place; evaluation 2 was applied during a period prior to harvesting in the treated catchment. Following the hypothetical harvest period in the control catchment, approximately 4% of innovations from evaluation 1 exceeded 95% prediction intervals for the median, lower, and upper models, respectively (Figure 8 and Table 2). Evaluation 2 consisted of applying our method to a period prior to harvesting in the treated catchment; 3% of the innovations exceeded the prediction limits for each model (Figure 9). In both cases, innovations were not significantly different from zero. Therefore, we have no reason to reject the null hypothesis that preharvest and the hypothetical postharvest innovations differ in the control catchment or during the evaluation period *prior* to harvesting in the treated catchment.

6. Discussion

6.1. Sources of Variation

[47] In general, the HBV-EC model was better suited for simulating streamflow in the treated catchment than the control. In the treated catchment, observed streamflow fell within the uncertainty bounds approximately 81% of the time during calibration compared to 73% in the control catchment. Deviations between simulated and observed streamflows in each catchment were attributed to input errors, model structure, and identification.

6.1.1. Model Inputs and Observed Data

[48] Input error is likely attributed to errors in measured streamflow, particularly during summer lowflows and large peak flows, and the short observed streamflow and climate time series used for model calibration (~1.5 years) and simulation (~2.5 years). Large measurement error during lowflow conditions are attributed to flume design and installation; flumes in Hinkle Creek were designed to measure hydrology when a larger portion of streamflow volume flows in the channel. Accurate streamflow measurements were more difficult during lowflow conditions when a larger proportion of streamflow volume is transported in the substrate below the base of the flume thereby bypassing pressure transducers. Uncertainty is additionally attributed to the relatively short model calibration period. A short calibration period can increase model uncertainty by not fully capturing the range of hydrologic and climatic variability that drive model outcome and exacerbate parameter identification. However, there is no rule concerning the length of the calibration period; the data should cover a range of significant events to find stable model parameters [*Bergström*, 1991]. An exploration of uncertainty due to the length of our calibration period is beyond the scope of this paper. We contend, however, that a longer preharvest model calibration period would exert little influence over the results of this study. The influence of different levels of hydrologic model uncertainty on change detection is discussed in section 6.2.3.

[49] By assessing the evidence that climate was similar between the preharvest and postharvest periods, we are able to attribute changes in streamflow response in the treated catchment to harvesting or other sources of unexplained variation rather than to climatic variations during the period of this study. However, climate data were measured at a single station at an elevation of 839 m and used to simulate streamflow in both catchment that have considerably different hypsometry and watershed area. Elevation in the treated catchment ranges from 615 m at the outlet to 815 m at the drainage divide, with 100% of the total catchment area (0.23 km^2) below the climate station. In contrast, elevations and area in the control catchment range from 658 to 1,260 m, with 85% of the 1.56 km^2 catchment area above the climate station. The parameter PGRADL is used to adjust precipitation for elevation difference but showed little sensitivity for either catchment. Additional heterogeneity is introduced in the distribution of elevation bands, precipitation, and catchment attributes. For example, aspect in the treated catchment is predominantly north facing, whereas it is predominantly southwest facing in the control.

6.1.2. Model Structure and Identification

[50] We used the HBV-EC model due to its relatively simple structure and low data requirements. Though the structure of this model has at best a moderate capacity to simulate hydrologic processes, we found the model to be appropriate for conducting uncertainty analyses and detecting changes at the catchment scale. However, model structure and identification are not without error. For example, the HBV-EC model systematically undersimulated streamflow during spring and summer periods when hydrology is dominated by lowflow hydrology and during winter when large peak flows are dominated by rain-on-snow hydrology [*Jones and Post*, 2004]. Lowflow hydrology in the western Cascades are sustained by baseflow that may be desynchronized from antecedent soil moisture conditions [*Moore and Wondzell*, 2005].

[51] Alternatively, the snow routine in the HBV model is more appropriate for high mountain regimes where snow accumulates during winter and melts during spring and

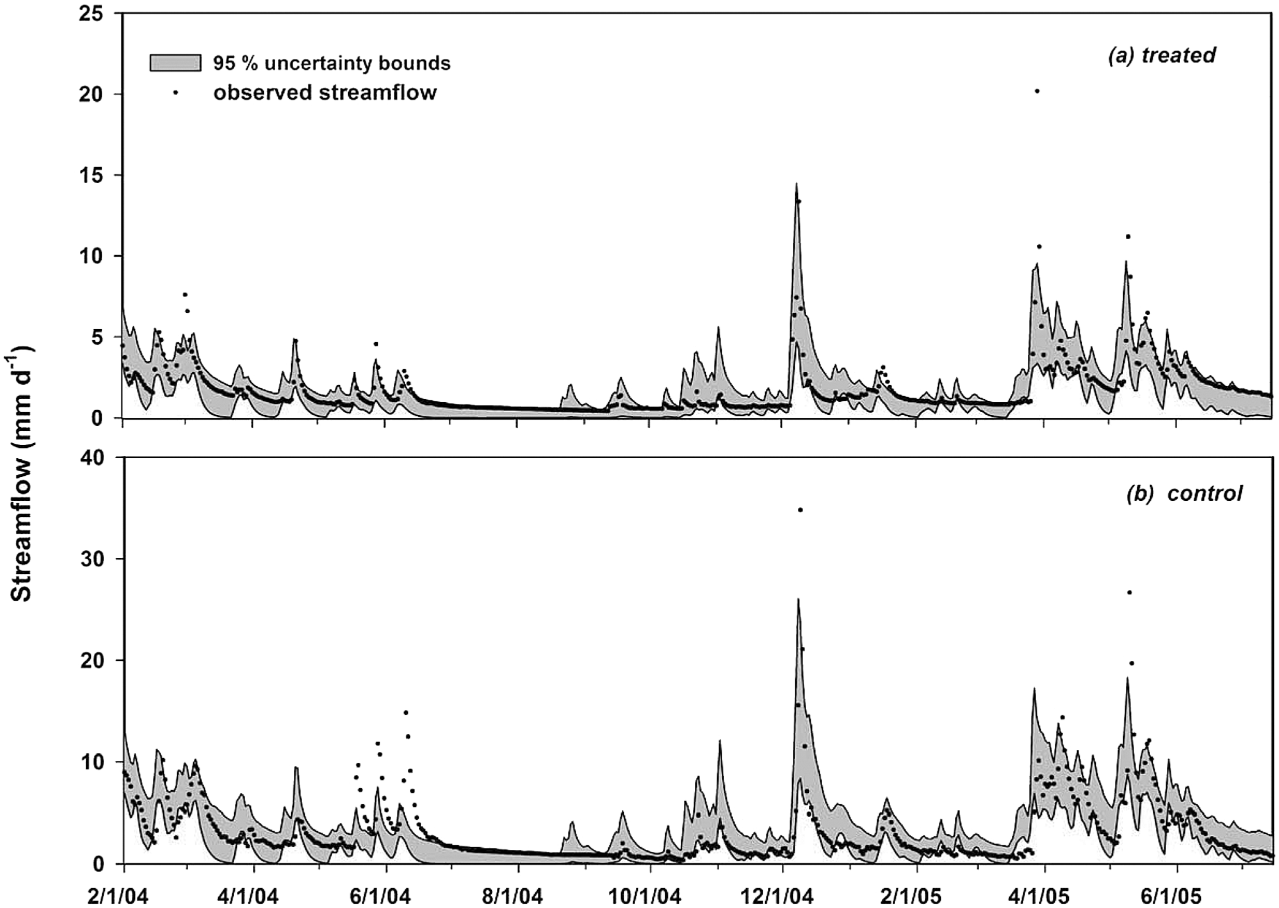

Figure 4. The 95th percentile uncertainty bounds (GLUE) for simulated streamflows and observed streamflow during model calibration period for (a) treated and (b) control catchments. GLUE analysis based on 850,000 simulations. The observed streamflow fell within uncertainty bounds 81 and 73% of days during the calibration period for the treated and control catchments and accounted for respectively 4% (31 mm) and 5% (91 mm) of total annual streamflow during the calibration period.

summer. It is not well suited to modeling rain-on-snow events, especially at a daily time step. Snowfall was not implicitly measured in this study. Rather, the model classifies all precipitation as either rain or snow based on threshold value TT and uses a fixed temperature lapse rate (TLAPSE) that incorrectly predicts snowfall rather than rain and melt at higher elevations when storms are associated with weather inversions [*Moore*, 1993]. This effect of temperature misclassification on simulated streamflow can be seen in Figures 4 and 5, where under- and oversimulation predominantly occur during spring.

[52] The aggregation of processes and interactions of parameters in the HBV model structure makes it difficult to rely on field observations to adequately inform model identification and calibration. Therefore, we relied on sensitivity analysis to define parameter space and conditioned Monte Carlo sampling on uniform distributions. We chose uniform distributions due to the lack of information regarding internal behavior [*Wagener and Kollat*, 2007] and process observations of the examined catchments as well as the ease of implementing such a distribution. The use of uniform sampling ensures that each set of chosen parameter values is evaluated as a set, implicitly reflecting interactions and insensitivities of model parameters, thereby avoiding the need to sample from multivariate sets of correlated distributions [*Freer et al.*, 1996]. Although not assessed in this study, additional field measurements such as isotopic hydrograph separation [*Sklash and Farvolden*, 1979] could be used to further describe and identify parameter distributions to more appropriately model catchment processes, thereby increasing model performance.

[53] TT and KS were identified as critical parameters for simulating daily streamflow in our catchments. *Seibert* [1997] reported similar sensitivity for TT in two Swedish catchments, and *Hamilton et al.* [2000] for KS for a Canadian catchment. The cumulative distributions of TT (not shown) for the control catchment showed higher spread than the treated catchment, indicating higher model sensitivity [*Wagener and Kollat*, 2007]. Differences in parameter distributions and sensitivity of TT values in this study are explained by elevational differences between catchments. Optimized parameter values for TT ranged from −1.2 to 0.36 for the treated and −0.76 to 2.4 for the control and are similar to ranges reported in other studies. For example, the lower and upper tails of our TT distributions are larger than those reported by *Hamilton et al.* [2000] (−0.20 to 0.80) but smaller

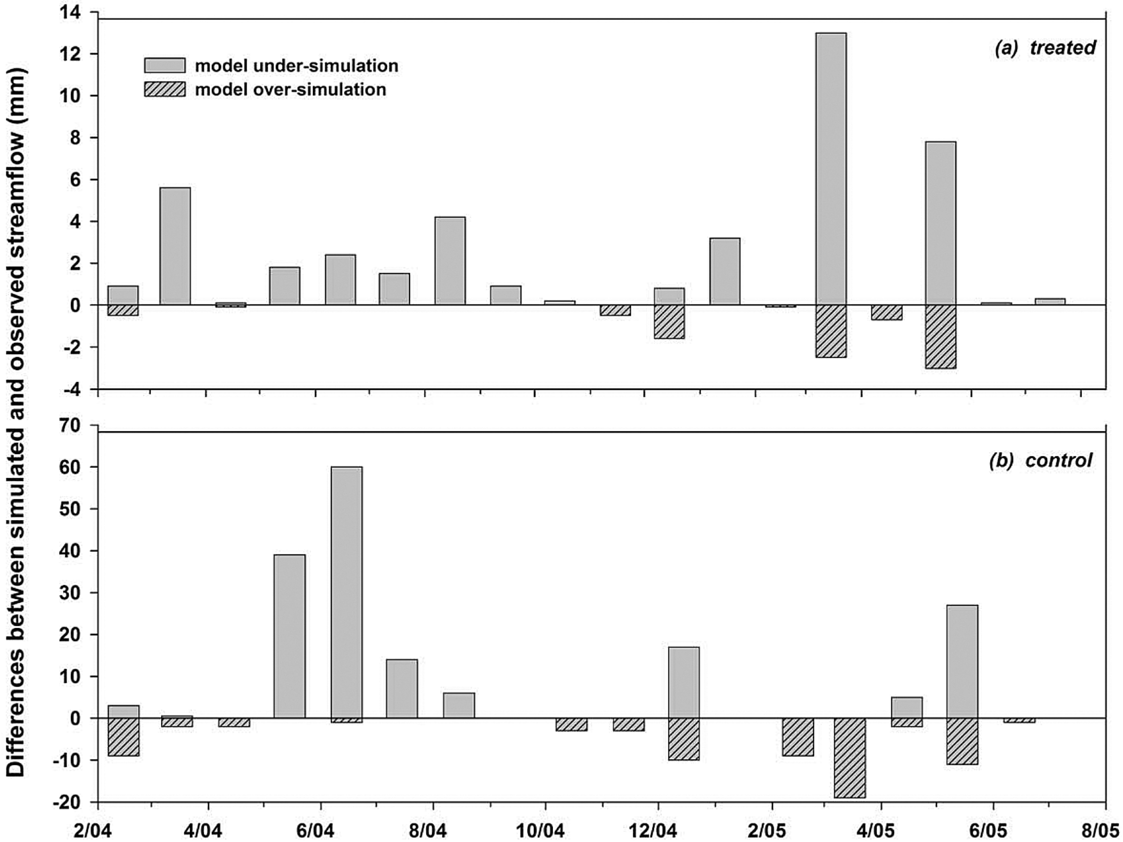

Figure 5. Monthly simulation errors for (a) treated and (b) control catchments during calibration period. Error defined by streamflow observations not contained within the 95th percentile uncertainty bounds. Errors are calculated as the difference between observed streamflow and nearest uncertainty bound totaled by month. Undersimulation occurs when observed streamflow is above upper 97.5th percentile bound; oversimulation occurs when observed streamflow is below lower 2.5th percentile bound.

than those reported by *Seibert* [1997] (−2.5 to 2.5) and *Seibert and McDonnell* [2010] (−1.5 to 2.5). *Hundecha and Bárdossy* [2004] and *Seibert* [1999] showed positive sensitivity for model parameters that controlled soil water balance (parameter FC) and runoff routing (KS). The differences in parameter ranges between these studies and our study are explained by the different methods of parameter optimization, errors in calibration data in each catchment, and regional differences in climate and catchment processes.

[54] The percentage of streamflow observations falling within the estimated uncertainty bounds and the proportion of behavioral models identified in our study are similar to other GLUE-based studies [*Freer et al.*, 1996; *McMichael et al.*, 2006]. Previous studies have explored the influence of the choice of likelihood measure and/or threshold definition on behavioral model identification and width of uncertainty bounds with varying results. For example, *Freer et al.* [1996] showed that different likelihood measures had little effect on the width of uncertainty bounds because the procedure retains only those models considered behavioral with different likelihood measures having common sets of simulations. Alternatively, *McMichael et al.* [2006] showed that different likelihood measures and thresholds can alter the number of behavioral models as well as the particular models selected from Monte Carlo simulations. The impact of different likelihood measures on change detection using our method was not explored, but we would expect minimal changes in the width of uncertainty bounds and the overall outcome of our study (see below). The effect of different rejection criterion thresholds and number of simulations on change detection are discussed in sections 6.2.3 and 6.2.4.

6.2. Hydrologic Model Change Detection

[55] Figure 3 shows the importance of using a (1) control catchment and a (2) formal statistical framework to evaluate the significance of hydrologic changes in modeling studies. In this figure we see that our hydrologic models consistently undersimulated streamflow compared to observed streamflow. The most elementary way to use a hydrologic model for discerning the impact of disturbance on hydrology is to simply compare observed and simulated streamflow. For this case, we would conclude that streamflow increased in both catchments, though harvesting only took place in the treated catchment. However, results from our regression analyses show that postharvest increases in streamflow were significantly greater than the population of preharvest streamflow in the treated catchment, whereas significant changes were not detected in the control catchment where no harvesting took place. The use of a formal statistical framework allows us to evaluate the significance of hydrologic changes relative to variations in observed and simulated

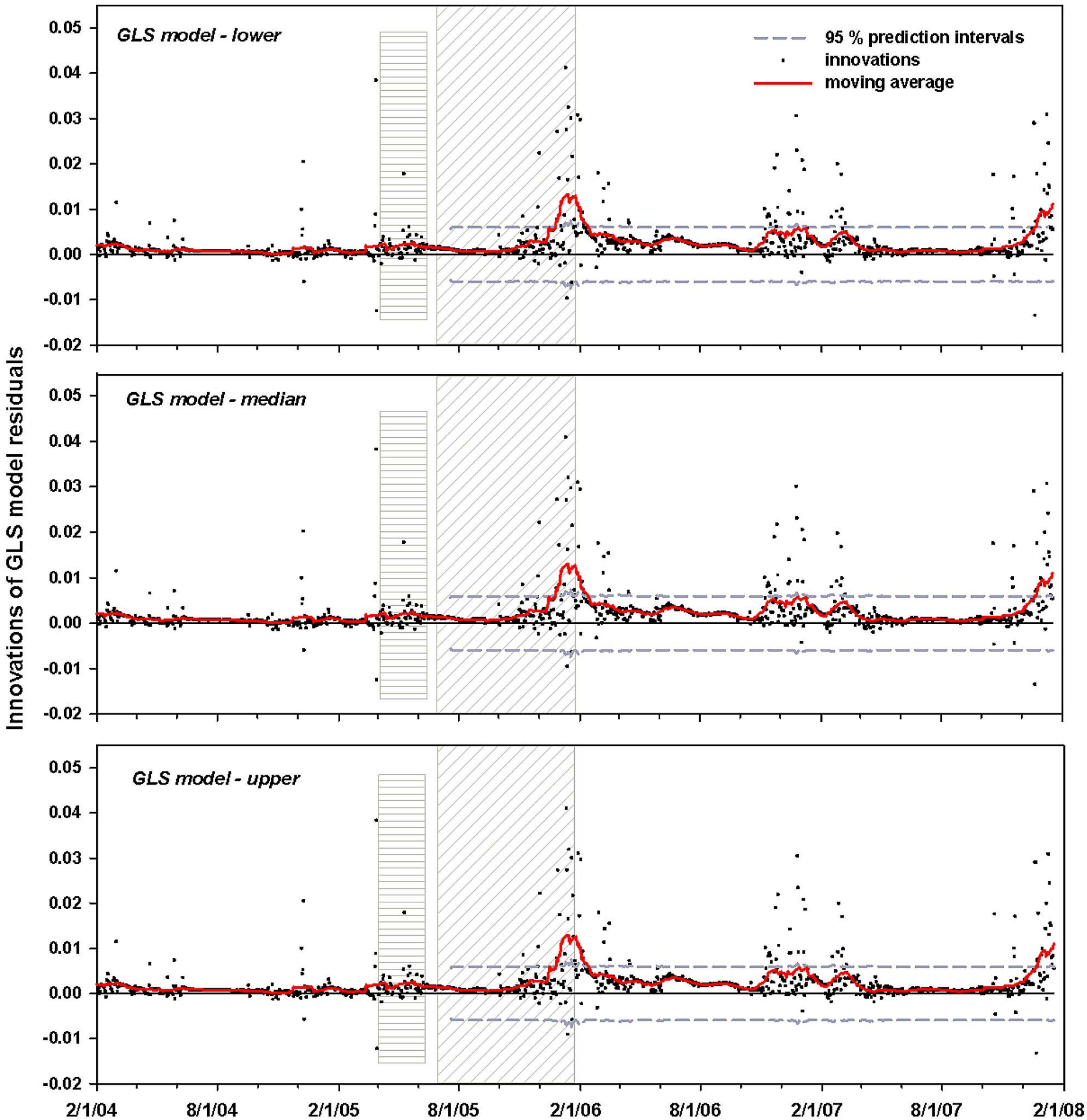

Figure 6. Time series of innovations of GLS model residuals for the treated catchment using streamflow from the 2.5th, 50th, and 97.5th uncertainty simulations (model evaluation period horizontally shaded, harvest period diagonally shaded). Statistically significant increases were detected for each level of hydrologic model uncertainty; postharvest innovations exceed the 95% prediction intervals 9% of the time for all models. Monthly moving average shown to facilitate visualization of trend.

streamflow rather than by simply comparing observed and simulated streamflow.

6.2.1. Detecting the Absence of Land Use Change

[56] Control catchments are used in the traditional paired catchment approach to account for climatic variation but are seldom used in modeling studies. By applying our method to two periods of no land use change, our modeled catchments serve as climatic controls. Through the application of our change detection method to the control catchment (evaluation 1) and during a period of no disturbance in the treated catchment (evaluation 2), we were able to critically evaluate overall change detection model performance and, more importantly, account for additional sources of variation. Significant changes were not detected in the control catchment at anytime during the experiment, nor during the period prior to harvesting in the treated catchment. Evaluation results corroborate the *t* tests performed on the climate data (section 4.4) that climate variations in these catchments are similar between the preharvest and postharvest periods. Thus, in these catchments during the study period, natural variations in

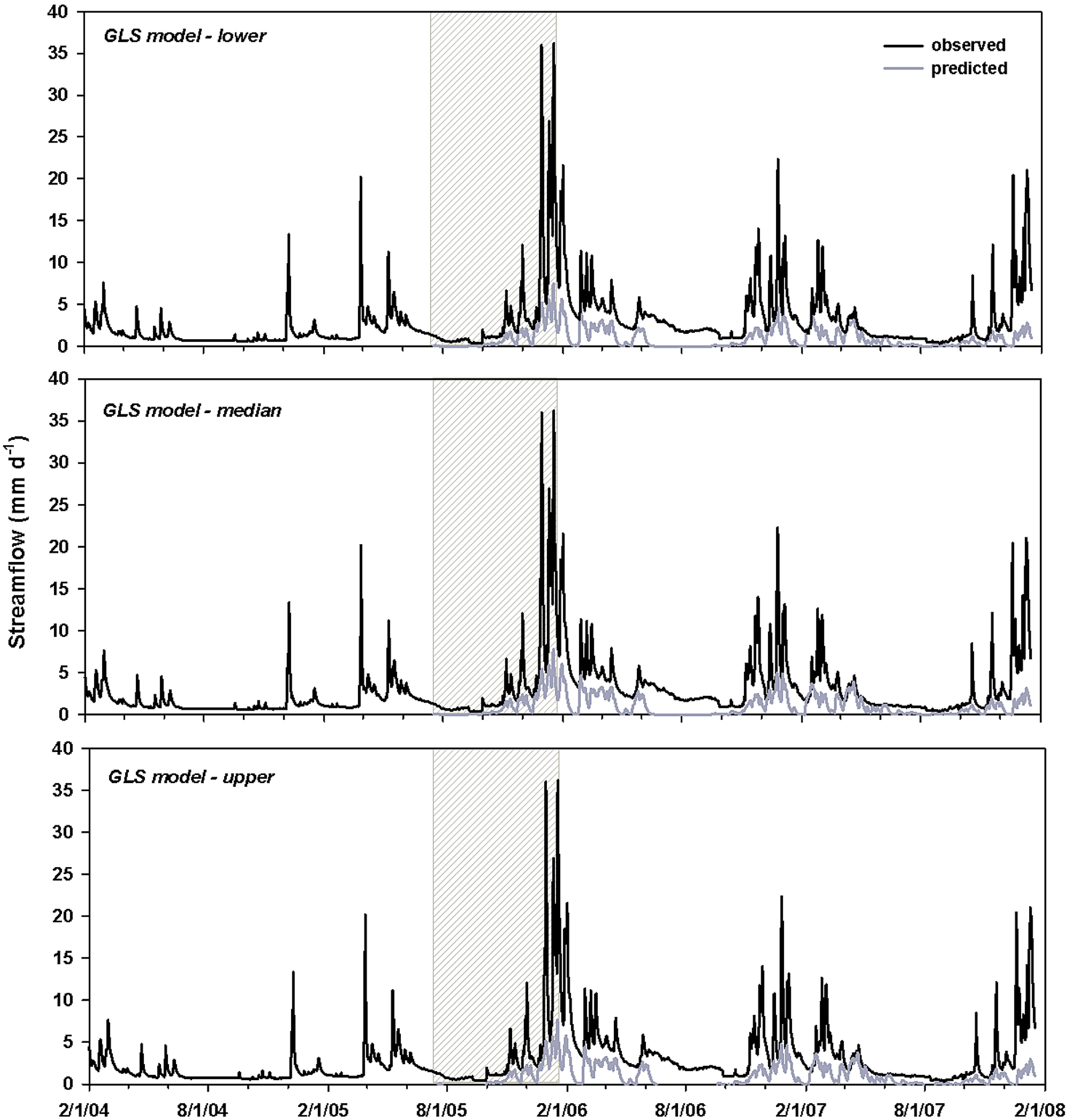

Figure 7. Observed and predicted daily streamflow for the treated catchment using the 2.5th, 50th, and 97.5th uncertainty simulations in GLS regression models. Daily streamflow is predicted using parameters coefficients from linear regression model developed using simulations from GLUE (explanatory variable) and observed streamflow (response variable)(harvesting period shaded).

climate do not contribute additional variation to the response signal detected following harvesting in the treated catchment.

6.2.2. Detecting and Estimating Changes in Streamflow Following Forest Harvesting

[57] Independent innovations from GLS regression models were used to detect streamflow changes while the estimated effects of forest harvesting on streamflow were calculated as the residual difference between observed and predicted streamflow. Statistically significant increases in streamflow were detected following clear-cut harvesting 65% of the treated catchment area. Increases in streamflow exhibit a distinct seasonal trend with the largest increases during winter, followed by spring, fall, and then summer (Table 3). Forest harvesting increases streamflow by reducing canopy interception and transpiration, modifying soil moisture depletion [*Hewlett and Hibbert*, 1961], while seasonal changes in hydrology are attributed to antecedent soil moisture conditions and effective precipitation. Streamflow increases in this study are consistent with similar studies in the Pacific Northwest that have shown greatest increases in streamflow during winter periods. For example, *Rothacher* [1970] showed that approximately 80% of water yield

Table 3. Streamflow Changes Following Harvesting in the Treated Catchment[a]

Season and Water Year[c]	GLS Model[b] (mm) Lower	Median	Upper	Average (mm)
Winter				
2005	ND[d]	ND	ND	ND
2006	751	727	744	741
2007	420	401	414	412
2008	305	301	306	304
Average	492	476	488	485
Spring				
2005	ND	ND	ND	ND
2006	235	222	229	229
2007	69	56	64	63
2008	ND	ND	ND	ND
Average	152	139	147	146
Summer				
2005	50	49	53	51
2006	185	185	190	187
2007	60	60	65	62
2008	ND	ND	ND	ND
Average	98	98	102	100
Fall				
2005	ND	ND	ND	ND
2006	92	89	93	91
2007	164	161	164	163
2008	88	86	90	88
Average	115	112	116	114

[a]GLS regression models were developed between observed streamflow and the lower, median, and upper streamflow simulations identified by GLUE using 850,000 simulations. Estimated effects were calculated as the residual difference between observed and GLS-predicted streamflow.

[b]Model respectively median, 2.5th percentile, and 97.5th percentiles.

[c]Seasons defined as follows: winter is months 12, 1, 2, and 3; spring is months 4, 5, and 6; summer is months 7, 8, and 9; and fall is months 10 and 11.

[d]ND means no data (beyond study period).

volume increases occurred during wet October to March following 100% harvesting of a 0.96 km^2 catchment in western Oregon.

[58] For each regression model, statistically significant increases in postharvest innovations were detected, but estimates in streamflow varied considerably. As previously described, we use the median model to detect and describe the effects of harvesting on streamflow, while the lower and upper models are used to explore how varying levels of hydrologic model uncertainty impact our ability to detect and estimate change. The variability in estimated harvesting effects on streamflow is expected and related to the different models. If we use the upper, 97.5th percentile simulations, which by definition represents overprediction of the hydrologic model [*Beven and Binley*, 1992], we would expect lower streamflow estimates. Alternatively, the lower 2.5th percentile model generally overestimates streamflow changes. Streamflow values in Table 3 illustrate the influence of hydrologic model over- and undersimulation on estimated changes in streamflow.

6.2.3. What Is the Influence of Different Likelihood Measure Thresholds on Detecting Change?

[59] To explore the sensitivity of our method to different likelihood thresholds, we extended our analysis to include simulations using better fitting ($R_{\mathrm{eff}} > 0.60$) and poorer fitting ($R_{\mathrm{eff}} < 0.30$) behavioral models. The selected level of behavioral model definition is somewhat arbitrary; one might expect that the definition of the rejection threshold may affect the width and locations of the uncertainty bounds predicted by GLUE.

[60] For each set of behavioral models, the proportion of postharvest innovations that exceeded the 95% prediction intervals was significantly greater than the 5% (Table 2) expected from natural variation alone, suggesting statistically significant changes in streamflow. Though the estimates of streamflow change depends on the specified level of model uncertainty, the ability to detect changes in postharvest innovations based on different levels of uncertainty remains consistent.

[61] These results are not unexpected for two reasons: (1) the statistical test used in this study is performed specifically on model innovations, the random noise component of the time series model [*Chatfield*, 2004], rather than on simulated and observed streamflow. The effect of this is that much of the model variation due to hydrologic model uncertainty is incorporated in the regression model and is reflected in the estimates of regression model parameters. The standard error (s_e) of GLS regression models differs little between different likelihood thresholds; s_e ranges from 0.003 and 0.004 for models based on $R_{\mathrm{eff}} > 0.60$ and <0.30, respectively. (2) The general lack of sensitivity to behavioral model definition is also related to the fact that only a small number of simulations will achieve the higher level thresholds; the majority of parameters sets will fall in the tails of the cumulative distributions of behavioral models and have little affect of the location of uncertainty bounds calculated by GLUE [*Lamb et al.*, 1998]. However, it is likely that variations due to poorer fitting models will eventually overwhelm the ability to detect change with increasing greater levels of model uncertainty. Furthermore, different likelihood measures (e.g., RMSE) could change the location of uncertainty bounds and our ability to detect change, though this was not explored in this study. In the range of likelihood thresholds evaluated in this study, our method is functionally robust to the varying levels of hydrologic model uncertainty and provides a suitable framework for detecting change using hydrologic models.

6.2.4. Does the Number of Simulations Affect Change Detection?

[62] No study we are aware of has explored the sensitivity of model change detection to the number of simulations used to characterize model uncertainty and, in our case, identify uncertainty time series used in GLS-based change detection. Though several studies focusing specifically on uncertainty methodology have reported simulations on the order of 10^6 [*Brazier et al.*, 2000], the majority of uncertainty studies are on the order of 10^4 (e.g., 15,000 [*Blasone et al.*, 2008], 100,000 [*Jin et al.*, 2010], 150,000 [*Vrugt et al.*, 2009], and 500,000 [*Seibert*, 1997]). There is a propensity to generate an increasingly larger set of model runs to explore parameter space, yet how the definition and population of behavioral models affects model change detection is not known.

[63] To assess the influence of hydrologic model population size on the ability to detect change using our method, we generated two additional populations of 350,000 and 50,000 simulations and repeated GLUE ($R_{\mathrm{eff}} > 0.40$) and GLS. In both cases, statistically significant increases in daily streamflow were detected following harvesting in the treated catchment, with 9% (350,000) and 8% (50,000) of postharvest innovations exceeding 95% prediction limits (Table 4).

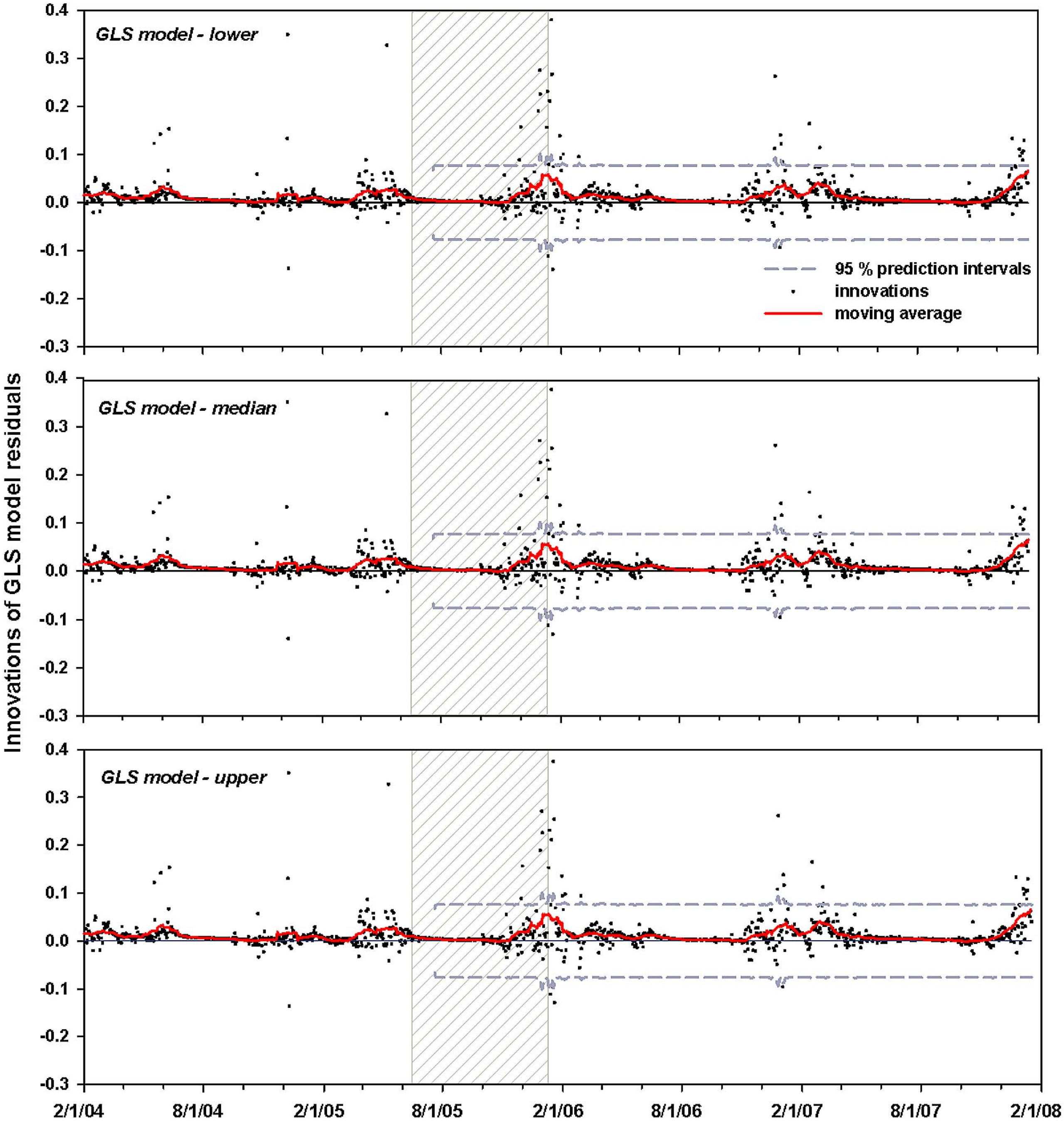

Figure 8. Time series of innovations of GLS model residuals for the control catchment using streamflow from the 2.5th, 50th, and 97.5th uncertainty simulations (hypothetical harvest period diagonally shaded). Changes in hydrology were not detected for any level of hydrologic model uncertainty in the control catchment; postharvest innovations exceed the 95% prediction intervals due to natural variation 4% of the time for each model. Monthly moving average shown to facilitate visualization of trend.

These results were identical and similar to the original analysis using 850,000 simulations. s_e for each model were identical across simulation sizes and regression model parameters, $\hat{\beta}_0$ and $\hat{\beta}_1$, exhibited little change between models (Table 4). Differences are so small between three population sizes because 50,000 simulations was large enough to capture variations in hydrologic model output in our catchments during the study period.

6.2.5. Overcoming Limitations of the Paired Catchment Approach

[64] Though the paired catchment study continues to be the predominant method for detecting and estimating hydrologic changes following disturbance, it has its limitations. (1) As

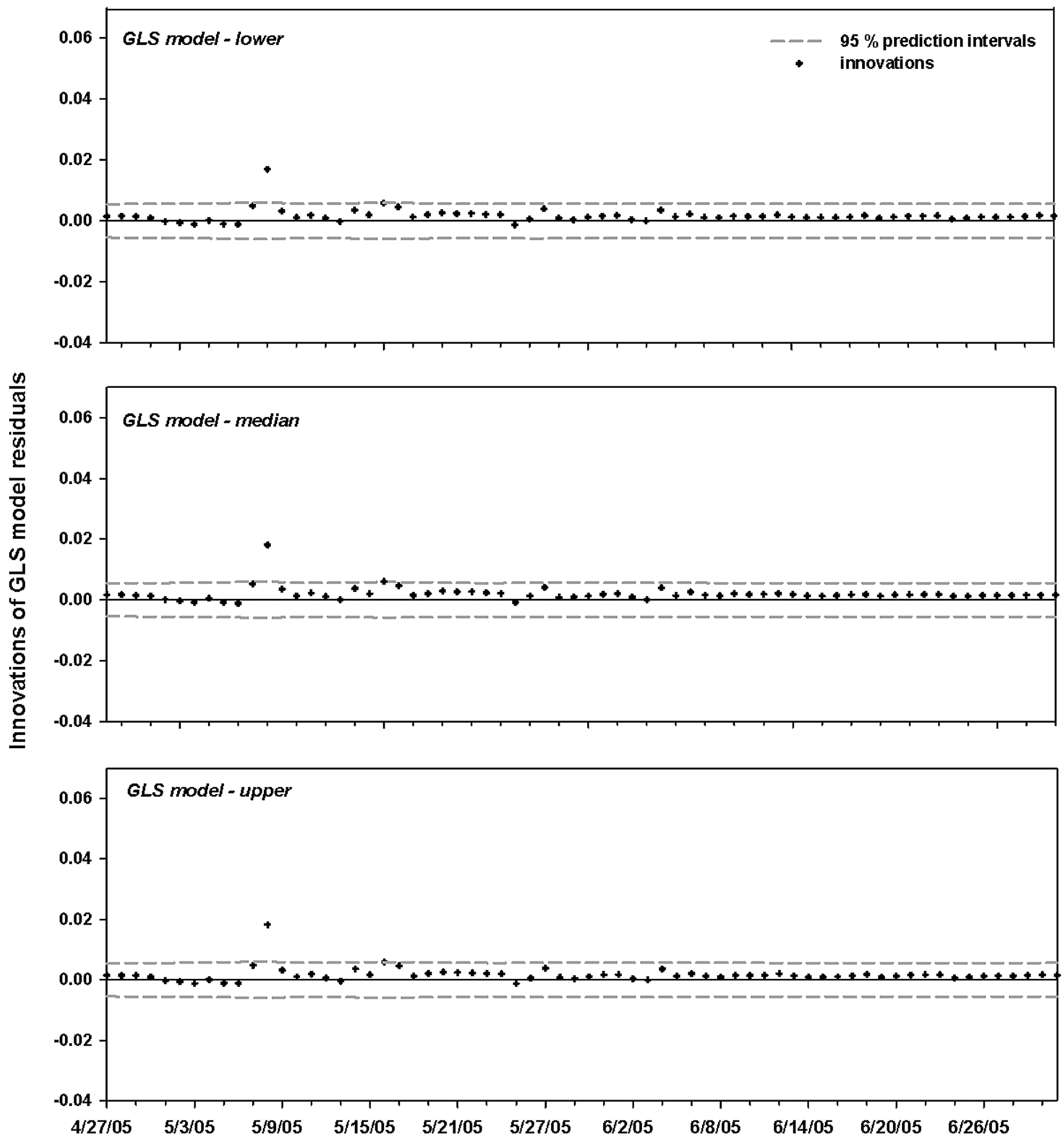

Figure 9. Time series of innovations of GLS model applied prior to harvesting in the treated catchment. This evaluation is used to test whether our method is capable of not detecting change in a period with no disturbance. Changes in hydrology were not detected during the preharvest period in the treated catchment for any level of hydrologic model uncertainty; postharvest innovations exceed the 95% prediction intervals due to natural variation 3% of the time for each model (the figure shows evaluation period with respect to study period.)

the name suggests, this approach requires a control catchment that serves as a climatic reference for the duration of the experiment. However, suitable control catchments are seldom available due to harvesting schedules, market pressures, and the identification of similarity in catchment characteristics to warrant pairing. (2) Long calibration periods and large sample sizes are needed to reliably detect changes unless variance is small and changes are small relative to the mean [*Loftis et al.*, 2001]. Though time series can be used to increase sample size, time series data can increase temporal variations in chronologically paired observations, thereby increasing the width of prediction limits and obfuscating the the effects of disturbance on streamflow [*Zégre*, 2009]. (3) Chronological pairing of flood events and inappropriate

Table 4. GLS Regression and AR Modeling Results for the Treated and Control Catchments[a]

Model[b]	d.f	k	ϕ_1	ϕ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	s_e	Percent Postharvest Innovations[c]
				Treated				
850,000								
Lower	531	2	0.79	−0.14	0.002	1.5	0.003	9
Median	531	2	0.76	−0.17	0.001	1.5	0.003	9
Upper	531	2	0.76	−0.16	0.001	1.5	0.003	9
350,000								
Lower	531	2	0.75	−0.17	0.005	1.6	0.003	9
Median	531	2	0.76	−0.17	0.001	1.5	0.003	9
Upper	531	2	0.75	−0.16	0.001	1.5	0.003	9
50,000								
Lower	531	2	0.77	−0.16	0.002	1.3	0.003	8
Median	531	2	0.76	−0.16	0.002	1.1	0.003	8
Upper	531	2	0.76	−0.16	0.002	1.1	0.003	8
				Control				
850,000								
Lower	531	2	2	0.87	−0.21	0.02	1.69	4
Median	531	2	2	0.87	−0.21	0.02	1.69	4
Upper	531	2	2	0.87	−0.21	0.02	1.69	4

[a]Coefficients $\hat{\beta}_0$, $\hat{\beta}_0$ are estimated by regression, k is the order of autocorrelation, $\hat{\phi}_1$ and $\hat{\phi}_2$ are estimated autocorrelation coefficients selected by AIC, and s_e is the standard error of the estimate. Eight hundred fifty thousand simulations were used for change detection in the treated and control catchments. To assess the influence of population size on the ability to detect change using our method, two additional populations of 350,000 and 50,000 were generated for the treated catchment. With selected significance level of $\alpha = 0.05$, 5% of estimated innovations should fall outside of the 95% predictions intervals due to random error. Statistically significant changes are noted if greater then 5% of the innovations exceed prediction intervals.

[b]Shown is the number of Monte Carlo simulations.

[c]Postharvest innovations exceedance of 95% prediction limits; postharvest period in the control catchment is hypothetical for change detection evaluation (1).

use of statistical analysis can result in incorrect estimates of changes in flood magnitude because neither the pairing or statistical tests account for changes in variance or flood frequency. Problems with chronological pairing of floods results from differences in timing, duration, intensity, or spatial extent [*Thomas and Megahan*, 1998] between the paired catchments.

[65] Our change detection method is a useful alternative to the traditional paired catchment approach as it addresses these issues. By using a rainfall-runoff model to reconstruct streamflow, we generate a virtual control that can be used in lieu of establishing an actual control on the ground. An individual catchment therefore serves as the treated and control catchment thereby reducing spatial and temporal variations exhibited between paired catchments. This is especially useful when using daily streamflow observations that tend to be desynchronized between catchments. Our method overcomes the problems associated with chronological pairing in the paired catchment approach [*Alila et al.*, 2009] as our method focuses on meteorological pairing between modeled and observed streamflow responses within the same catchment. As a result, paired observations of daily streamflow from the same catchment do not exhibit lag as shown by the lack of statistical evidence to include sinusoidal covariates. Therefore, error is primarily associated with how well the rainfall-runoff model mimics the behavior of the catchment under consideration. Further, our approach remains in accordance with the original principals of ANCOVA [*Alila et al.*, 2009] as our alternative hypothesis is that "postharvest innovations exceed prediction limits"; prediction limits are based on the variance of predisturbance models.

7. Conclusion

[66] In this study we present an alternative to the paired catchment approach to detect the effects of disturbance on catchment hydrology. Our method combines hydrologic modeling, uncertainty analysis, and regression time series modeling to isolate the effects of forest harvesting from input errors, model identification, and the large natural variability attributed to daily streamflow. Specifically, we used hydrologic modeling to account for natural fluctuations in daily streamflow and GLUE to identify and separate uncertainty from unexplained variation. Though GLUE does not provide insight on structural or input uncertainty, we demonstrate the use of GLUE to estimate model uncertainty and the effect of uncertainty on change detection. Furthermore, we provide a formal experimental framework to evaluate the significance of hydrologic change relative to variations in rainfall and streamflow.

[67] This study is unique from other modeling studies in that we evaluate the stationarity of climate records using GLS to compute t tests in autocorrelated observations to assess the impact of climate variations on detecting change and apply our change detection method to a control catchment (evaluation 1) and to a period prior-to forest harvesting in a treated catchment (evaluation 2) to demonstrate that our method was capable of capturing not only land use changes when they occur but also the absence of land use change. These exercises serve to critically evaluate overall change detection model performance and, more importantly, account for additional sources of variation that may ordinarily obfuscate or inflate the true effects of disturbance. In addition, we show that our method was robust under different levels of model uncertainty defined by three levels of R_{eff} and show that an increase in the number of model simulations does not necessarily result in increased change detection performance.

[68] The proposed method is a potentially useful alternative to the paired catchment approach for detecting change using highly variable daily hydrology data and stand-alone hydrologic modeling. Further, this approach can be useful to construct virtual control catchments when it is impossible to establish reference catchments on the ground. Though developed and tested for evaluating the effects of forest harvesting on hydrology, the proposed method may be applicable to studies evaluating and forecasting change in water resources resulting from fire, insect denudation, urbanization, and directional climate change.

[69] **Acknowledgments.** Funding for this research was provided by the Giustina Innovative Research Grants Program through the College of Forestry, Oregon State University, and the Oregon State University Watershed Research Cooperative. The authors acknowledge the ongoing commitment of Roseburg Forest Products to the Hinkle Creek Paired Watershed Study. We thank R. Dan Moore of the University of British Columbia for thoughtful reviews and contributions to this paper. Additionally, the authors thank David Hutchinson of Environment Canada for EnSim Hydrologic support and Levi Kilcher (Oregon State University) and Matthew Thompson (U.S. Forest Service) for programming and technical language support.

References

Alila, Y., P. K. Kuras, M. Schnorbus, and R. Hudson (2009), Forests and floods: A new paradigm sheds light on age-old controversies, *Water Resour. Res.*, *45*, W08416, doi:10.1029/2008WR007207.

Andréassian, V., E. Parent, and C. Michel (2003), A distribution-free test to detect gradual changes in watershed behavior, *Water Resour. Res.*, *39*(9), 1252, doi:10.1029/2003WR002081.

Bates, C. (1921), First results in the Streamflow Experiment, Wagon Wheel Gap, Colorado, *J. For.*, *19*(4), 402–408.

Bergström, S. (1991), Principles and confidence in hydrologic modelling, *Nord. Hydrol.*, *22*(2), 123–136.

Bergström, S. (1995), The HBV Model, in *Computer Models of Watershed Hydrology*, edited by V. Singh, pp. 443–470, Water Resour. Publ., Highlands Ranch, Colo.

Beven, K. (2002), Uncertainty and the detection of structural changes in models of environmental systems, in *Environmental Foresight and Models: A Manifesto*, edited by M. Beck, pp. 227–250, Elsevier, Oxford, U. K.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298.

Blasone, R.-S., H. Madsen, and D. Rosbjerg (2008), Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling, *J. Hydrol.*, *353*, 18–32.

Bosch, J. M., and J. D. Hewlett (1982), A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration, *J. Hydrol.*, *55*(1), 3–23.

Bowling, L. C., P. Storck, and D. P. Lettenmaier (2000), Hydrologic effects of logging in western Washington, United States, *Water Resour. Res.*, *36*(11), 3223–3240.

Brandt, M., S. Bergstrom, and M. Gardelin (1988), Modelling the effects of clearcutting on runoff-Examples from central Sweden, *Ambio*, *17*, 307–313.

Brazier, R. E., K. J. Beven, J. Freer, and J. S. Rowan (2000), Equifinality and uncertainty in physically based soil erosion models: Application of the GLUE methodology to WEPP (the Water Erosion Prediction Project) for sites in the UK and USA, *Earth Surf. Processes Landforms*, *25*(8), 825–845.

Canadian Hydraulics Centre (2006), EnSim Hydrologic, Natl. Res. Counc., Ottawa.

Chatfield, C. (2004), *The Analysis of Time Series: An Introduction*, 6th ed., Chapman and Hall, Boca Raton, Fla.

Eisenbies, M., W. Aust, J. Burger, and M. Adams (2007), Forest operations, extreme flooding events, and considerations for hydrologic modeling in the Appalachians: A review, *For. Ecol. Manage.*, *242*(2–3), doi:10.1016/j.foreco.2007.01.051.

Freer, J., K. Beven, and B. Ambroise (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, *32*(7), 2161–2174.

Gomi, T., R. D. Moore, and A. S. Dhakal (2006), Headwater stream temperature response to clear-cut harvesting with different riparian treatments, coastal British Columbia, Canada, *Water Resour. Res.*, *42*, W08437, doi:10.1029/2005WR004162.

Hamilton, A. S., D. G. Hutchinson, and R. D. Moore (2000), Estimating winter streamflow using conceptual streamflow model, *J. Cold Reg. Eng.*, *14*(4), 158–175.

Harr, D. R., R. Fredriksen, and J. Rothacher (1979), Changes in streamflow following timber harvest in Southwestern Oregon, *Tech. Rep. PNW-249*, For. Serv., U.S. Dep. of Agric., Washington, D. C.

Harris, D. (1977), Hydrologic changes after logging in two small Oregon coastal watersheds, *U.S. Geol. Surv. Tech. Rep. Water Supply Pap., 2037*.

Helsel, D., and R. M. Hirsch (1992), *Statistical Methods in Water Resources*, Elsevier, Amsterdam.

Hewlett, J. D. (1971), Comments on the catchment experiment to determine vegetal effects on water yield, *Water Resour. Bull.*, *7*(2), 376–381.

Hewlett, J. D., and A. R. Hibbert (1961), Increases in water yield after several types of forest cutting, *Q. Bull. Int. Assoc. Sci. Hydrol*,, *6*, 5–16.

Hornberger, G., and R. Spear (1981), Approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, *12*(1), 7–18.

Hundecha, Y., and A. Bárdossy (2004), Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *J. Hydrol.*, *292*(1–4), 281–295.

Jackson, C., C. Sturm, and J. Ward (2001), Timber harvest impacts on small headwater stream channels in the coast ranges of Washington, *J. Am. Water Resour. Assoc.*, *37*(6), 1533–1550.

Jin, X., C.-Y. Xu, Q. Zhang, and V. Singh (2010), Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model, *J. Hydrol.*, *383*, 1–4, doi:10.1016/j.jhydrol.2009.12.028.

Jones, J. A., and D. A. Post (2004), Seasonal and successional streamflow response to forest cutting and regrowth in the northwest and eastern United States, *Water Resour. Res.*, *40*, W05203, doi:10.1029/2003WR002952.

Kuczera, G., G. P. Raper, N. S. Brah, and M. D. Jayasuriya (1993), Modelling yield changes after strip thinning in a mountain ash catchment: An exercise in catchment model validation, *J. Hydrol.*, *150*(2–4), 433–457.

Lamb, R., K. Beven, and S. Myrabø (1998), Use of spatially distributed water table observations to constrain uncertainty in a rainfall runoff model, *Adv. Water Resour.*, *22*(4), doi:10.1016/S0309-1708(98)00,020-7.

Lavabre, J., D. S. Torres, and F. Cernesson (1993), Changes in the hydrological response of a small Mediterranean basin a year after a wildfire, *J. Hydrol.*, *142*(1), 273–299.

Legates, D. R., and G. J. McCabe (1999), Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, *35*(1), 233–241.

Loftis, J. C., L. H. MacDonald, S. Streett, H. K. Iyer, and K. Bunte (2001), Detecting cumulative watershed effects: the statistical power of pairing, *J. Hydrol.*, *251*(1–2), 49–64.

Lørup, J. K., J. C. Refsgaard, and D. Mazvimavi (1998), Assessing the effect of land use change on catchment runoff by combined use of statistical tests and hydrological modelling: Case studies from Zimbabwe, *J. Hydrol.*, *205*(3–4), 147–163.

Machiwal, D., and M. Jha (2008), Comparative evaluation of statistical tests for time series analysis: Application to hydrological time series, *Hydrol. Sci. J.*, *53*(2), 353–366.

McMichael, C. E., A. S. Hopeb, and H. A. Loaiciga (2006), Distributed hydrological modelling in California semi-arid shrublands: MIKESHE model calibration and uncertainty estimation, *J. Hydrol.*, *317*, 307–324.

Moore, R. (1993), Application of a conceptual streamflow model in a glacierized drainage basin, *J. Hydrol.*, *150*(1), 151–168.

Moore, R. D., and S. Wondzell (2005), Physical hydrology and the effects of forest harvesting in the Pacific Northwest: A review, *J. Am. Water Resour. Assoc.*, *41*(4), 763–784.

Myers, R. H. (1990), *Classical and Modern Regression with Applications*, PWS-Kent, Boston, Mass.

Nash, J., and J. Sutcliffe (1970), River flow forecasting through conceptual models. Part I: a discussion of principles, *J. Hydrol.*, *10*, 282–290.

Padmanabhan, G., and A. R. Rao (1982), Order selection of AR Models of Hydrologic Time Series, *Nord. Hydrol.*, *13*(2), 93–104.

Post, D., A. Jakeman, I. G. Littlewood, P. G. Whitehead, and M. D. A. Jayasuriya (1996), Modelling land-cover-induced variations in hydrologic response: Picaninny Creek, Victoria, *Ecol. Modell.*, *86*(2–3), 177–182.

Ramsey, F., and D. Schafer (2002), *The Statistical Sleuth: A Course in Methods of Data Analysis*, 2nd ed., Duxbury, Pacific Grove, Calif.

Rothacher, J. (1970), Increases in water yield following clear-cut logging in the Pacific northwest, *Water Resour. Res.*, *6*(2), 653–658.

Rothacher, J. (1973), Does harvest in west slope Douglas-fir increase peak fow in small forest streams?, *Tech. Rep. Res. Pap. PNW-163*, Pac. Northwest Res. Stn., For. Serv., U.S. Dep. of Agric., Portland, Oreg.

Salas, J. (1993), Analysis and modeling of hydrologic time series., in *Handbook of Hydrology*, edited by D. Maidement, pp. 19.1–19.72, McGraw-Hill, New York.

Schnorbus, M., and Y. Alila (2004), Forest harvesting impacts on the peak flow regime in the Columbia Mountains of southeastern British Columbia: An investigation using long-term numerical modeling, *Water Resour. Res.*, *40*, W05205, doi:10.1029/2003WR002918.

Schreider, S. Y., A. J. Jakeman, R. A. Letcher, R. J. Nathan, B. P. Neal, and S. G. Beavis (2002), Detecting changes in streamflow response to changes in non-climatic catchment conditions: farm dam development in the Murray-Darling basin, Australia, *J. Hydrol.*, *262*(1–4), 84–98.

Seibert, J. (1997), Estimation of Parameter Uncertainty in the HBV Model, *Nord. Hydrol.*, *28*(4/5), 247–262.

Seibert, J. (1999), Regionalisation of parameters for a conceptual rainfall-runoff model, *Agric. For. Meteorol.*, *98–99*, 279–293.

Seibert, J., and J. McDonnell (2010), Land-cover impacts on streamflow: Change-detection modelling approach that incorporates parameter uncertainty, *Hydrol. Sci. J.*, *55*(3), 316–332.

Seibert, J., J. McDonnell, and R. Woodsmith (2010), Effects of wildfire on catchment runoff responses: A modelling approach to detect changes in

snow-dominated forested catchments, *Hydrol. Res.*, *41*(5), 378–390, doi:10.2166/nh.2010.036.
Sklash, M. G., and R. N. Farvolden (1979), The role of groundwater in storm runoff, *43*(1–4), 45–65.
Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, *44*, W00B06, doi:10.1029/2008WR006822.
Stednick, J. D. (1996), Monitoring the effects of timber harvest on annual water yield, *J. Hydrol.*, *176*(1–4), 79–95.
Swank, W. T., J. M. Vose, and K. J. Elliott (2001), Long-term hydrologic and water quality responses following commercial clearcutting of mixed hardwoods on a southern Appalachian catchment, *For. Ecol. Manage.*, *143*(1–3), 163–178.
Thomas, R. B., and W. F. Megahan (1998), Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon: A second opinion, *Water Resour. Res.*, *34*(12), 3393–3403.
Vrugt, J. A., C. J. F. ter Braak, H. V. Gupta, and B. A. Robinson (2009), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, *Stochastic Environ. Res. Risk Assess.*, *23*(7), 1011–1026, doi:10.1007/s00,477-008-0274-y.
Wagener, T., and J. Kollat (2007), Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox, *Environmental Modell. Software*, *22*(7), doi:10.1016/j.envsoft.2006.06.017.
Wagener, T., H. S. Wheater, and M. J. Lees (2004), Monte Carlo analysis toolbox, Pa. State Univ., University Park.
Watson, F. G. R., R. Vertessy, T. McMahon, B. Rhodes, and I. Watson (2001), Improved methods to assess water yield changes from paired-catchment studies: application to the Maroondah catchments, *For. Ecol. Manage.*, *143*(1–3), 189–204.
Zégre, N. P. (2009), Local and downstream effects of contemporary forest harvesting on streamflow and sediment yield, Ph.D. thesis, Oreg. State Univ., Corvallis.

L. M. Ganio and N. A. Som, Department of Forest Ecosystems and Society, Oregon State University, 321 Richardson Hall, Corvallis, OR 97331, USA.

J. McDonnell and A. Skaugset, Department of Forest Engineering, Resources, and Management, Oregon State University, 204 Peavy Hall, Corvallis, OR 97331, USA.

N. Zégre, Division of Forestry and Natural Resources, West Virginia University, PO Box 6125, Morgantown, WV 26506, USA. (nicolas.zegre@mail.wvu.edu)