ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/339292055

Modeling hydrological inflow persistence using paleoclimate reconstructions on the Québec-Labrador (Canada) Peninsula

Article *in* Water Resources Research · February 2020

DOI: 10.1029/2019WR025122

CITATIONS
0

READS
66

6 authors, including:

Bouchra R. Nasri
Université de Montréal
22 PUBLICATIONS 58 CITATIONS
SEE PROFILE

Etienne Boucher
Université du Québec à Montréal
41 PUBLICATIONS 292 CITATIONS
SEE PROFILE

L. Perreault
Hydro-Québec
80 PUBLICATIONS 1,548 CITATIONS
SEE PROFILE

Bruno N. Remillard
HEC Montréal - École des Hautes Études commerciales
138 PUBLICATIONS 3,010 CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Millennial multi-proxy reconstruction of summer PDSI for southern south america View project

Project Statistical post-processing of ensemble weather forecasts View project

All content following this page was uploaded by Bouchra R. Nasri on 24 February 2020.

The user has requested enhancement of the downloaded file.

Modeling hydrological inflow persistence using paleoclimate reconstructions on the Québec-Labrador (Canada) Peninsula

B. R. Nasri [1]

É. Boucher [2]

L. Perreault [3]

B. N. Rémillard [4]

D. Huard [5]

A. Nicault [6]

Members of the ARCHIVES-PERSISTENCE projects [7]

[1]Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montréal, Québec, CANADA, H3A 0B9

[2]Department of Geography, GEOTOP and Centre d'études nordiques, Université du Québec à Montréal, 1255 St-Denis, Montréal, Québec, CANADA, H2X 3R9

[3]Hydro-Québec Research Institute

[4]Department of Decision Sciences, HEC Montréal, 3000 chemin de la Côte-Sainte-Catherine, Montréal, Québec, CANADA, H3T 2A7

[5]Ouranos, 550 Rue Sherbrooke W. West Tower, 19th floor Montréal, Québec, H3A 1B9

[6]ECCOREV, FR 3098, CNRS/Aix-Marseille Université, Europôle Méditerranéen de l'Arbois, BP 80, 13545 Aix-en-Provence cedex 4, France

[7] Dominique Arseneault (Université du Québec à Rimouski), Christian Bégin (Geological Survey of Canada, Natural Resources Canada), Martine M. Savard (Geological Survey of Canada, Natural Resources Canada), Yves Bégin (Institut National de la Recherche Scientifique, centre Eau-Terre-Environnement), Pierre Francus (Institut National de la Recherche Scientifique, centre Eau-Terre-Environnement), Patrick Lajeunesse (Université Laval).

Corresponding author: Bouchra R. Nasri, `bouchra.nasri@mail.mcgill.ca`

Key Points:

- Reservoir inflows have been reconstructed from tree-ring proxies over the last 200 years for four large watersheds on the Québec-Labrador Peninsula.
- Gaussian hidden Markov models accurately describe the regime-switching behavior seen in the observed and reconstructed inflow time series.
- A formal goodness-of-fit test is used to estimate the number of regimes of the Gaussian hidden Markov models.
- The accuracy of annual inflow forecasts can be improved by extending observational time series with 200-year paleoclimatic reconstructions.

Abstract

Annual inflow forecasts are often based on historical time series, where every year is considered equally likely to reoccur. This process ignores the persistence of dry/wet conditions often observed in time series, behavior that is of utmost importance for hydroelectric energy producers. However, the modeling of persistence properties is challenging when only short time series are available for calibration. Here, we use Gaussian hidden Markov models to describe the regime-switching behavior, where the next year's inflow depends on the current estimated regime. For four large hydropower reservoirs on the Québec-Labrador Peninsula, a Gaussian hidden Markov model is calibrated on both a 30-year observational record and a 190-year paleoclimatic inflow reconstruction. Each reconstruction is a composite of three reconstruction methods drawing on five different tree-ring proxies (ring widths, minimal wood density, maximal wood density, δ^{13}C and δ^{18}O). The calibration on the reconstructed series finds two hydrological regimes, while the calibration on the observed data has only one regime for three out of four watersheds. Yearly hindcasts with the two calibrated Gaussian hidden Markov models suggest that for all four watersheds, extending the time series with reconstructions improves the model's predictive accuracy. This approach does not explicitly account for the differing accuracy of the observational and reconstructed time series or compare hidden Markov models to other models of persistence.

1 Introduction

In Canada, the provinces of Québec, Manitoba and British Columbia rely almost exclusively on hydropower generation to meet electricity demand [*National Energy Board*, 2017]. Because space and water heating constitute a large fraction of the electrical load, power demand peaks in winter during cold snaps [*Hydro-Québec Distribution*, 2014]. By contrast, water inflows peak a few months later with the snow melt. This timing mismatch between water inflows and energy demand can be compensated by using reservoirs to build up stocks in preparation for winter. The largest reservoirs can regulate flows over multiple years, providing some measure of resilience to prolonged droughts.

Management rules for reservoir operations are guided by the historical interplay between energy demand and the water regime, accounting for natural fluctuations around average hydrological conditions. For example, in the case of Hydro-Québec, Québec's provincial electric utility, reservoir levels are regulated by the Québec Energy Board [*Hydro-Québec Production*, 2018] to ensure that the electric utility has sufficient reserves to meet power demand in the event of prolonged low inflows. Persistent dry conditions pose risks not only for power generation, but also for groundwater availability, forest fires and ecosystems [*Diffenbaugh et al.*, 2015].

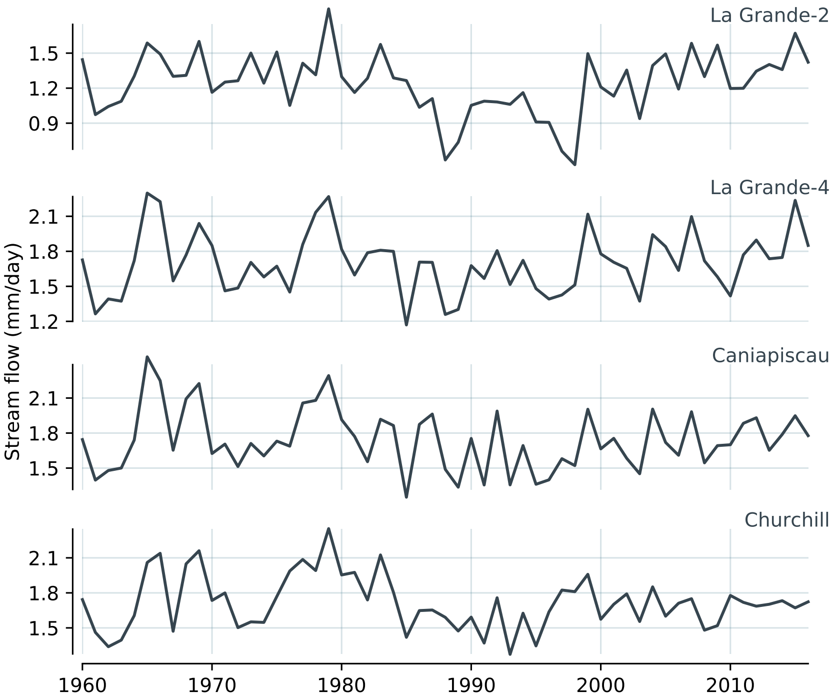


Figure 1. Observed 1960-2016 annual water supply time series for watersheds La Grande 2 (LG2), La Grande 4 (LG4), Caniapiscau (CAN) and Churchill Falls (CHU) located on the Québec-Labrador Peninsula, Canada.

In Québec-Labrador, reservoir inflows are based on historical streamflow records dating back to the 1950's, and management rules implicitly assume that any past year is equally likely to reoccur next year. This assumes independence and stationarity hypotheses that, as in many hydrologic time series, are partially falsified by autocorrelation and climate change. Indeed, several authors have noted that the behavior of hydroclimatic historical records often exhibits persistence in several distinct states with occasional transitions between these states; see e.g., *Thyer and Kuczera* [2000]. Figure 1 illustrates the historical annual water supplies measured for four important watersheds in the Québec-Labrador region. The geographical locations of these basins are presented in Figure 2. After examining these time series, one may suspect the presence of local nonstationarity. Dry and wet sequences appear to have occurred, which may suggest that the annual inflows of these basins exhibit distinct shifting regimes. The mid 1980's change-point corresponds to the beginning of the longest period of consecutive low flows in Hydro-Québec's historical water supply time series. Moreover, these time series also appear to exhibit abrupt changes in variability, for instance, in the early 1970's. These characteristics have led a number of authors to study the available hydroclimatic time series in the region by using different types of change-point and mixture models [*Perreault et al.*, 2000; *Perreault*, 2001; *Perreault et al.*, 2007; *Jandhyala et al.*, 2009; *Evin et al.*, 2011; *Merleau*, 2017, 2018].

Although the hypothesis of a stationary process for these time series should be questioned, the relatively short length of observation records limit our ability to adequately describe the naturally-occurring hydrological fluctuations, especially when regime lengths span a decade or more [*Wilhelm et al.*, 2018]. Indeed, the perspective gained from observed records is limited to a few decades. From an operational forecasting point of view, having a reliable model to describe the likelihood of prolonged wet or dry periods is valuable, and being able to extend the hydrological time series would improve our knowledge of long-term variability and persistence. Ideally, long time series (at least more than 100 years) that cover a broad spectrum of hydrological variability, from yearly to multidecadal variations, would be used to adequately characterize long-term persistence, possible nonstationarity, and feed into operational forecasting models. In addition, as shown in Thyer et al. [2006] in the context of Gaussian hidden Markov models and autoregressive models, longer series significantly reduce model and parametric uncertainties.

Where regimes cannot be accurately described and modeled from observed records, proxy data may be used to extend the length of the hydrological time series beyond the period covered by instruments [*Loaiciga et al.*, 1993]. In particular, moisture-sensitive tree-ring series provide annually resolved records that cover a broad spectrum of hydrological variability. Tree-ring widths have been used to reconstruct past hydrological conditions in arid regions in which the growth-limiting factor was water availability [*Stockton and Fritts*, 1973; *Smith and Stockton*, 1981; *Meko et al.*, 2001; *Woodhouse and Lukas*, 2006; *Woodhouse et al.*, 2013; *Nicault et al.*, 2008]. In boreal regions, although water availability is not the main factor limiting tree growth, recent research has shown that the use of a multiproxy approach (incorporating tree-ring widths, discrete markers of wood density and stable isotope fractionation of tree-ring cellulose) considerably strengthens hydrological reconstructions produced in high-latitude areas [*Nicault et al.*, 2014a; *Boucher et al.*, 2011a; *Brigode et al.*, 2016; *Boreux et al.*, 2009].

Here, we show how multicentury tree-ring-based reconstructions may be used to shed light on the characteristics of hydrological regimes in the context of intensive hydroelectric production. We use an extensive, multiproxy tree-ring network to explore both the spatial and temporal variability of hydrological regimes in large hydroelectric infrastructure, namely the La Grande (Québec, Canada) and the Churchill Falls (Newfoundland-Labrador, Canada) hydroelectric facilities. Our objectives are (i) to reconstruct water supplies to major hydroelectric generating stations over the past two centuries, (ii) to use those reconstructions to characterize regime properties (average flow, variability, duration) and model the probability of regime change under Gaussian hidden Markov models with a formal goodness-of-fit test, and (iii) to investigate the predictive ability of the selected Gaussian hidden Markov model for each basin, using scoring rules suited for probabilistic forecasts.

2 Methods and data

The reconstruction, conducted over four basins located in Québec and Labrador (Section 2.1), is based on both recent observations (1960–2000) of reservoir inflows and paleoclimatic reconstructions (1800–2000) made from a combination of three dendrochronological proxies (Section 2.2). The persistence analysis within the observed and reconstructed inflow time series until 1990 is analyzed through the prism of Gaussian HMMs (Section 2.3). The performance of these models when calibrated on observations and reconstructions are compared by performing hindcast experiments over the period 1991–2016.

2.1 Study area and hydrological data

Four watersheds draining into major hydropower reservoirs located in Québec and Labrador (Canada) are considered in this study to illustrate our approach: La Grande 2 (LG2), La Grande 4 (LG4), Caniapiscau (CAN) and Churchill Falls (CHU). The watersheds all drain large areas, ranging from 28,440 km^2 for the LG4 watershed to 69,141 km^2 for the CHU watershed. These basins feed major hydropower facilities, and in this context, strategic management decisions are based on hydrological historical time series and forecasts produced for these sites. Figure 2 illustrates the geographical location of the four watersheds considered in this study. Watersheds LG2, LG4 and CAN are parts of the La Grande water resources system, one of the largest hydropower systems in North America. The CAN watershed is the farthest upstream on the La Grande river. Each basin has a large reservoir and a power plant at their outlet. The CHU basin also has a reservoir at its outlet and a single power plant. The total installed capacity in these river basins constitutes approximately 30% of Hydro-Québec's total capacity. The streamflow regime of these four watersheds is dominated by a northern climate, which favors snow accumulation and low streamflow during winter (December to February), followed by high streamflow during spring.

Generation planners face a variety of decisions in operating these systems. Two issues that are common to every installation are safety and the respect of environmental laws and regulations. Since these watersheds have large reservoirs to store water, the other main concerns are long-term energy planning and optimization. Clearly, the future state of inflows plays a major role in the decisions, namely for long-term strategies to set energy safety margins. Long hydrological informative time series and a thorough knowledge of their statistical characteristics, such as persistence, are thus of paramount importance. Daily observed streamflow data for all four basins have been provided by Hydro-Québec for the 1960–2016 period. The corresponding annual inflow time series are presented in Figure 1. Using these observed data and dendrochonological tree-ring proxies, we reconstruct 200 years of annual inflows in order to overcome the lack of streamflow data for the last two centuries.

2.2 Tree-ring reconstruction of water supplies

2.2.1 Tree-ring proxies description

In total, 39 black spruce sites across the Québec-Labrador Peninsula were sampled for dendrochronological analysis (Table 1). All sites are located between 53°N and 56°N and are found primarily in open spruce-lichen woodlands, the most widely distributed forest ecosystem in Québec-Labrador's boreal zone [*Girard et al.*, 2008]. The locations are presented in Figure 2. A minimum of 10 trees were sampled at each site. Only dominant trees with a symmetrical shape that were free from major growth anomalies were selected. Collected cross sections were dried and finely sanded. At each of the 39 sites, tree-rings were cross-dated and measured (two or three radii) using a micrometer with an accuracy of 0.001 mm (Velmex Inc., Bloomfield, NY). The dating accuracy was validated with the COFECHA software [*Holmes*, 1983].

Wood densitometry measurements were performed on selected trees from 20 sites across the network (Table 1) based on standard procedures [*Schweingruber et al.*, 1978, 1996]. Only discs without anomalies (reaction wood, branches, rotten wood, etc.) were selected. Three wood samples per tree were cut precisely into 1 mm laths, placed in a Soxhlet apparatus with ethanol for resin extraction, and then X-rayed. To measure density, X-ray micrographs were analyzed on a DENDRO 2003 microdensitometer (Walesch, Switzerland). A cellulose acetate calibration wedge was used to convert the lightness measurement into density (g.cm^{-3}) values. The time series retained from this densitometry analysis provided time series of the maximum (MXD) and minimum (MND) wood density. The tree-ring width and density series were standardized using the age-band approach [*Briffa et al.*, 2001].

Analysis of carbon and oxygen stable isotopes (δ^{13}C and δ^{18}O) was performed at three sites (DA1, HM1 and POOL). At each site, four radii were selected and subsampled on each disc. Growth rings covering the 1800–2004 period were manually separated using stainless steel blades. For the 1940–2000 period, tree-rings were cut at an annual resolution. Before 1940, the resolution was biannual to reduce the number of analyses performed in periods where no climatic data were available. Separated wood material from the same year was pooled, ground and homogenized. α-cellulose was extracted following *Savard et al.* [2004] to remove components that could create artifacts in the δ^{13}C signal due to their proportion changes in the wood (e.g., resin lipids, lignin). δ^{13}C values were measured from the α-cellulose samples via elemental analysis (Carlo Erba) in a continuous-flow isotope ratio

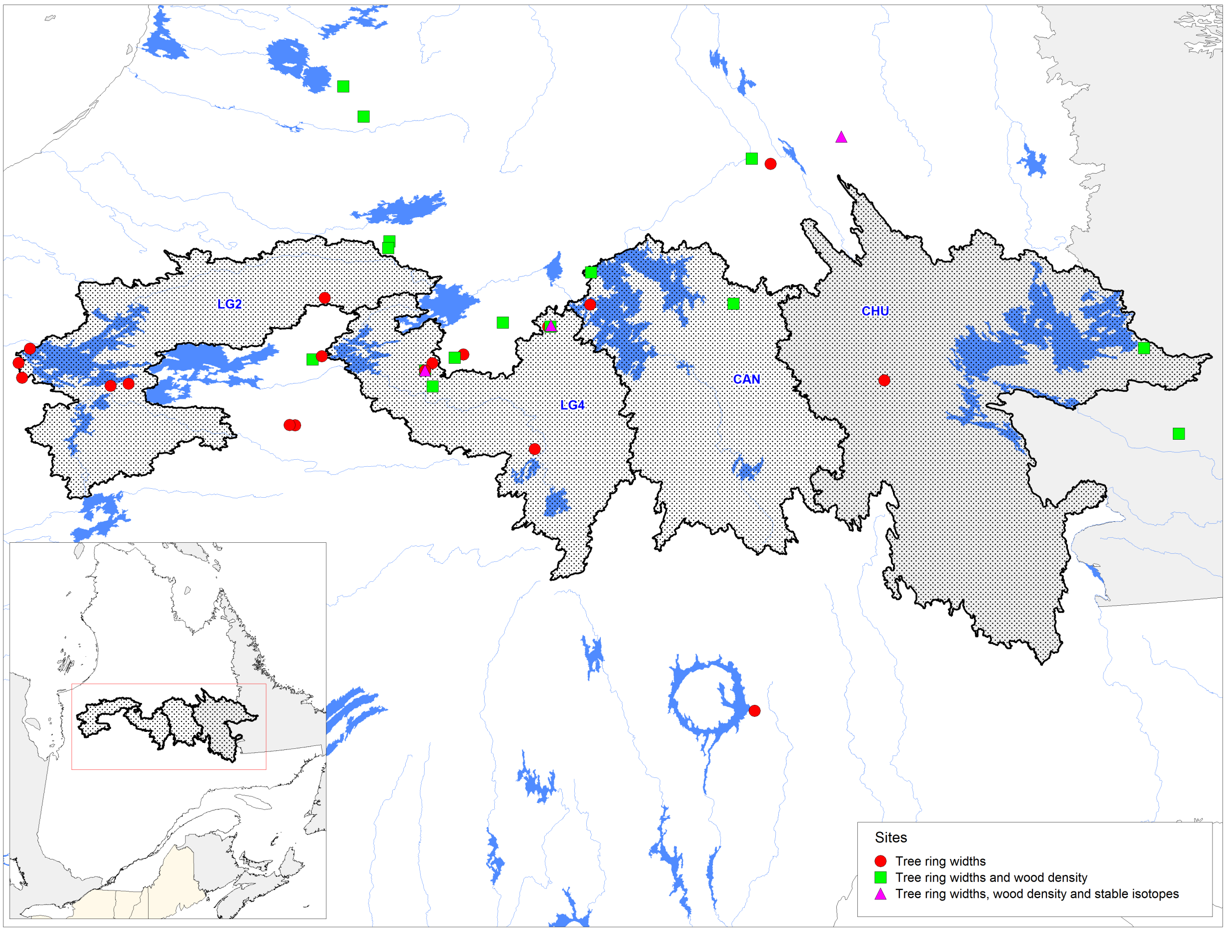


Figure 2. The locations of the four watersheds are highlighted in dark grey (LG2 [La Grande 2], LG4 [La Grande 4], CAN [Caniapiscau], CHU [Churchill Falls]), and a tree-ring multiproxy network is used to reconstruct the annual water supplies at these sites and all the basins managed by Hydro-Québec.

Table 1. Tree-ring chronologies from Québec-Labrador, extending back to at least AD 1800, used to reconstruct water supplies over the study area. RW: ring widths, MND: minimal wood density, MXD: maximal wood density, δ^{13}C and δ^{18}O stable isotope ratios.

ID	Site	Latitude	Longitude	Elevation (m)	Distance to sea (km)	Nb	EPS	Proxy				
								RW	MND	MXD	$\mathbb{C}$	$\mathbb{O}$
CANE	Caniapiscau East	54.44	-68.37	688	450	18	0.85	X	X	X		
CEA	Eaton Canyon	55.56	-68.12	175	305	28	0.91	X	X	X		
CORILE	Corvette 1	53.37	-74.05	324	323	21	0.84	X				
CORPL	Corvette 2	53.37	-74.12	324	328	29	0.94	X				
DA1M	DA1 1	53.86	-72.41	525	367	14	0.85	X	X	X	X	X
DA1R	DA1 2	53.86	-72.41	523	367	17	0.82	X				
DA1X	DA1 3	53.86	-72.41	523	367	15	0.76	X	X	X		
HER	Hervey	54.42	-70.27	530	450	17	0.82	X				
HH1	Hurault 1A	54.24	-70.82	541	432	20	0.87	X				
HM1	Hurault 1	54.25	-70.78	551	434	17	0.91	X	X	X	X	X
HM2	Hurault 2	54.24	-70.79	518	434	18	0.86	X	X	X		
HUR	Hurst	55.52	-67.86	419	307	13	0.74	X				
LAB17	Churchill N	53.97	-62.98	517	263	13	0.80	X	X	X		
LAB19	Trans Lab 1	53.29	-62.62	440	300	15	0.79	X	X	X		
LAB32	Goose-Bay	53.61	-60.89	265	200	30	0.83	X	X	X		
LAB35	Trans Lab 2	53.07	-61.63	372	273	14	0.83	X	X	X		
LAB42	Esker road	53.83	-66.40	490	400	16	0.83	X				
LAB56	Manic5	51.29	-68.12	465	168	13	0.89	X				
LAB65	Manic5-2	51.29	-68.12	462	173	16	0.83	X	X	X		
LECA	Clearwater 2	56.01	-73.75	327	205	19	0.87	X	X	X		
LJ2	Jourdin2	54.37	-73.79	445	261	13	0.80	X				
NFL1V	NFL1 V	53.52	-77.63	218	94	21	0.92	X				
NFL610	NFL610	53.75	-77.58	170	94	10	0.66	X				
NFLR1	NFL1C	53.63	-77.70	201	87	21	0.87	X				
NFLR2	NFL1D	53.57	-76.25	227	94	29	0.93	X				
NFT75	Trans-Taiga75	53.54	-76.48	210	173	10	0.76	X				
NIT	Nitchequon	53.29	-70.94	736	471	17	0.82	X				
POOL	Pool	55.72	-66.89	485	285	16	0.81	X	X	X	X	X
ROZM	Roz 2	54.84	-72.98	451	275	21	0.86	X	X	X		
ROZX	Roz 4	54.79	-72.99	451	275	21	0.82	X	X	X		
RT426	Transtaïga 426	53.97	-72.03	470	373	10	0.77	X	X	X		
RT485	Transtaïga 485	54.26	-71.42	447	393	16	0.83	X	X	X		
RT630	Transtaïga 680	54.67	-70.27	559	448	12	0.78	X	X	X		
T1	Tilly1	53.89	-73.89	432	294	22	0.71	X	X	X		
T4S	Tilly 4	53.92	-73.77	464	296	10	0.89	X				
THH	Thiers	53.74	-72.30	556	380	22	0.91	X	X	X		
TI26	TI26	54.00	-71.92	500	370	12	0.71	X				
TI41	TI41	53.92	-72.32	485	363	12	0.81	X				
TIDA1	TIDA1	53.86	-72.41	529	345	15	0.78	X				

mass spectrometer (CF-IRMS; Fisons Prism III). The external precision of the δ^{13}C ratios obtained on duplicate samples (treatment and analysis) was 0.08 ‰. All δ^{13}C values were corrected for the Suess effect and for changes in atmospheric [CO_2] (PIN correction; *McCarroll et al.* [2009]). The oxygen isotopic ratios (δ^{18}O) were measured with a pyrolysis-CF-IRMS (Delta plus XL), giving an external precision of 0.1‰.

2.2.2 Reconstruction methods

Previous annual water supplies to each of the four drainage basins (LG2, LG4, CAN, CHU) were reconstructed back to 1800. The reconstruction approach applied here was the same as that in *Nicault et al.* [2014a], and readers are referred to the original text for more detailed information on the methods. In short, three different statistical modelling approaches were used to perform annual hydrological reconstructions (Figure 3). Method 1 was based on the partial least squares approach, which represents an extension of the principal components regression [*Tenenhaus*, 1998]. For Method 1, an initial reconstruction was performed using the complete set of proxy series available in a 200-km radius around each hydroelectric power generating station. Method 2 used the same partial least square approach, but a selection of proxy series was performed based on the stepwise regression method. Only variables with P-values smaller than 0.01 were retained and included in the reconstruction (Table 2). Automatic selection of proxy records among the pool of available series was performed separately for each basin for tree-ring widths, MXD and stable isotope proxies to ensure that each proxy type was represented in the reconstructions. Selected proxy series were recombined into a single predictor matrix that was used as an input for the partial least square method. Method 3 was performed based on the best analogue method, which aimed at identifying, for each year i in the past for which no inflow value existed, the year k within the observed record that had the most similar proxy vector, according to an Euclidean distance metric [*Guiot et al.*, 2005; *Nicault et al.*, 2008; *Boucher et al.*, 2011b; *Guiot et al.*, 2010].

All reconstruction methods were calibrated with annual hydrological records from the 1961–2000 period, i.e., the maximal period covered by both tree-rings and hydrological data (Figure 3). Calibration (coefficient of determination: R^2, root-mean-squared error: RMSE) and validation (coefficient of determination for prediction: R_p^2, root-mean-squared error of prediction: $RMSE_p$) statistics were calculated based on a jackknife (Method 1 and Method 2) or a bootstrap procedure (Method 3).

The reconstruction produced by each of the three approaches were combined into a single, more robust reconstruction that accounts for shortcomings associated with each calibration method and proxy series selection. As shown in LeBlanc and Tibshirani [1996] and in Chapter 16 of Hastie et al. [2009], combining a collection of estimators can improve validation performance. Here, we applied the method proposed by *Nicault et al.* [2014a] to obtain the final reconstruction. First, for each year reconstructed, three Gaussian distributions were fitted based on the mean and standard deviation of each reconstruction (Figure 3). Second, 500 samples were randomly drawn from an equally weighted mixture of these Gaussian distributions. The composite reconstruction (COMP) corresponds to the mode of the mixture of distributions obtained for each year (Figure 3). The illustrated 90% confidence interval for the composite reconstruction is given by the 5^{th} and 95^{th} percentiles of the mixture of distributions. All analyses were performed in the R-project environment [*R Core Team*, 2017]. The validation statistics R_p^2 and RMSE$_p$ for COMP were computed using a jackknife method.

2.3 Time series modeling, persistence analysis, and forecasting

2.3.1 Persistence

In hydrological time series, persistence is often associated with long memory through an autoregressive fractionally integrated moving average model (ARFIMA); see, e.g., *Hosking* [1984]. In this case, long memory is measured by the fractional differentiation parameter of the ARFIMA model and is related to the Hurst exponent [*Mandelbrot and Wallis*, 1968]. This approach has largely been used to detect long memory effects in hydroclimatological time series [*Pelletier and Turcotte*, 1997; *Ault et al.*, 2013, 2014; *Koutsoyiannis*, 2005]. However, by definition, the Hurst exponent may exist without implying long memory [*Beran*, 1994]. An interesting alternative to describe persistence in time series is to use regime-switching models [*Hamilton*, 1990]. As shown in *Diebold and Inoue* [2001], regime-switching models can exhibit long memory. In addition, these models are easy to interpret and can easily be fitted to data. This type of model has also been used to detect long memory in hydroclimatological observed data [*Thyer and Kuczera*, 2000, 2003; *Evin et al.*, 2011] and to model climatic reconstructions from tree-ring time series [*Bracken et al.*, 2014; *Gennaretti et al.*, 2014]. Although the class of regime-switching models is large, we restrict our attention to the simple model of *Hamilton* [1990], which is also called a Gaussian hidden Markov model (HMM), since the annual reconstructed and observed inflows in our study are well fitted by this model. Note that this model is the same as that used by *Thyer and Kuczera* [2000, 2003]. In the Gaussian HMM setting, there are m hidden (non-observable) states or regimes, denoted τ_t for period t, and the observations in each regime (annual inflows) are distributed as an independent Gaussian distribution with its own mean μ_j and standard deviation σ_j, $j \in 1, \ldots, m$. The dynamics of regime switches are modeled by a Markov chain with transition matrix denoted by Q, where Q_{ij} is the probability that the next regime is j given the current regime i. In this model, persistence is measured by the number of switches between regimes.

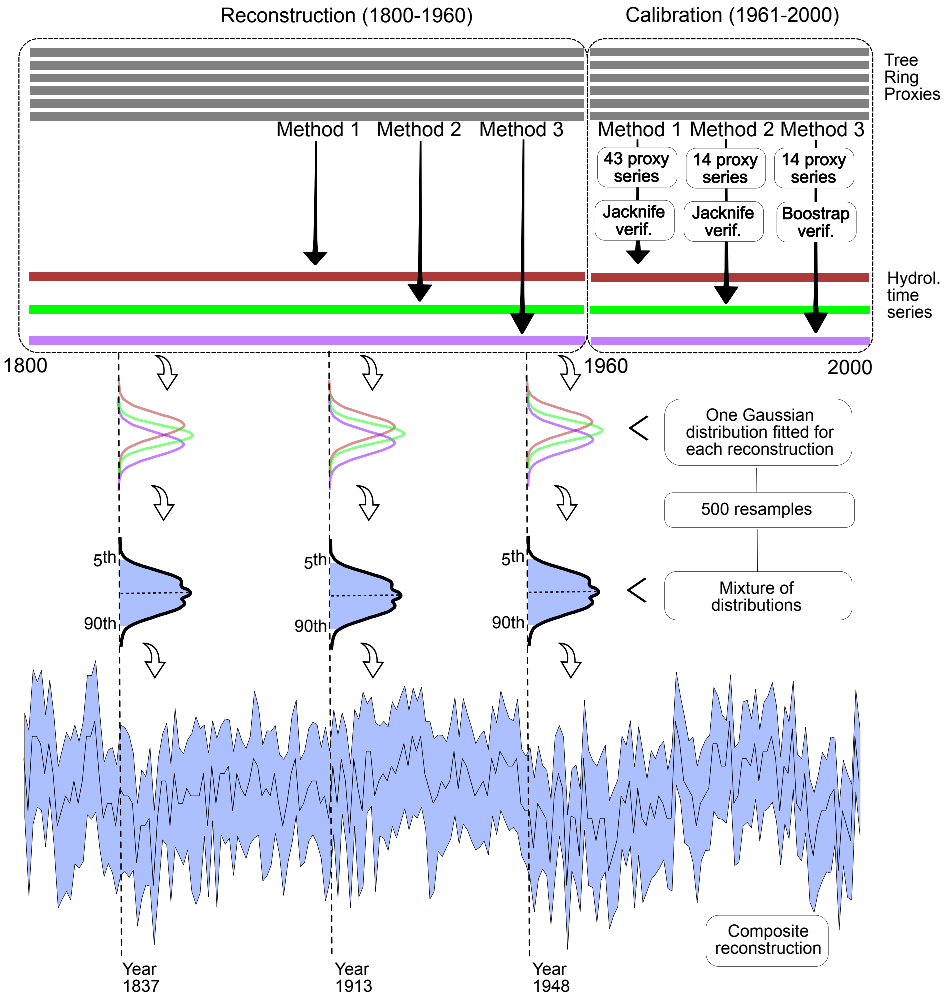


Figure 3. The hydrological reconstruction approach used to reconstruct historical water supplies back to 1800. An illustration of the approach is presented for LG2 only, but the method remains the same for LG4, CAN and CHU. The three reconstructions were then combined into a single composite reconstruction. 500 samples were drawn from a mixture of three Gaussian distributions fitted for each year and for each method. The reconstructed water supply value corresponds to the mode of the mixture distribution. The confidence intervals represent the 5^{th} and 95^{th} percentiles. The composite reconstruction is illustrated separately for 1837, 1913, and 1948.

Table 2. Available tree-ring proxies for each basin where annual water supplies were reconstructed. Method 1 (partial least square method) reconstruction used all available tree-ring proxies. Proxy series selected based on the stepwise regression approach for Method 2 (partial least squares) and Method 3 (best analogue method) are in bold.

ID	LG2	LG4	CAN	CHU
NFLR1	RW			
NFL1V	RW			
NFL610	RW			
NFT75	**RW**			
NFLR2	RW			
CORPL	RW	RW		
CORILE	**RW**	RW		
T1	RW, **MXD**, MND	**RW**, MXD, MND	**RW**, MXD, MND	
LJ2	RW	RW	RW	
T4S	**RW**	RW	**RW**	
LECA	RW, **MXD**, **MND**	RW, MXD, MND		
ROZX	**RW**, MXD, MND	**RW**, MXD, MND	**RW**, MXD, MND	
ROZM	RW, **MXD**, **MND**	RW, MXD, **MND**	RW, MXD, MND	
TIDA1	RW	RW		
DA1M	RW, MXD, MND, δ^{13}**C** , δ^{18}O	RW, MXD, **MND**, δ^{13}C , δ^{18}O	RW, MXD, **MND**, δ^{13}C , δ^{18}O	
DA1R	RW,	RW,	RW,	
DA1X	RW, MXD, **MND**	RW, **MXD**, **MND**	RW, **MXD**, MND	
TI41	RW	RW		
THH	**RW**, MXD, **MND**	RW, MXD, MND	RW	
RT426	RW, MXD, MND	**RW**, MXD, MND	RW	
TI26	RW	RW		
RT485	RW, MXD, MND	**RW**, MXD, MND	RW	
NIT	**RW**	RW	RW	
HH1		RW	RW	RW
HM2		RW, **MXD**, MND	RW, **MXD**, MND	RW, MXD, MND
HM1		**RW**, **MXD**, **MND**, δ^{13}C , δ^{18}O	**RW**, MXD, **MND**, δ^{13}C , δ^{18}**O**	RW, **MXD**, MND, δ^{13}C , δ^{18}O
RT630		RW	**RW**	RW
HER		RW	RW	RW
CANE		RW, MXD, MND	RW, MXD, MND	RW, MXD, MND
LAB56				RW
CEA		RW, MXD, MND	RW, MXD, MND	**RW**, MXD, MND
HUR		RW	RW	RW
POOL		RW, MXD, MND, δ^{13}C , δ^{18}O	RW, MXD, MND, δ^{13}C , δ^{18}O	RW, MXD, MND, δ^{13}C , δ^{18}O
LAB42			RW	RW
LAB17				RW, MXD, MND
LAB19				RW, MXD, MND
LAB35				RW, MXD, MND
LAB32				RW, MXD, MND

For example, if we take the simplest case of a Gaussian HMM with two regimes, the observations in Regime 1 (resp. Regime 2) follow a Gaussian distribution with mean μ_1 (resp. μ_2) and standard deviation σ_1 (resp. σ_2), and the transition probability matrix Q is given by Q_{11} (probability of remaining in Regime 1), Q_{22} (probability of remaining in Regime 2) and $Q_{12} = 1 - Q_{11}$ and $Q_{21} = 1 - Q_{22}$ (probabilities of switching from Regime 1 to Regime 2 and from Regime 2 to Regime 1). Figure 4 illustrates the dynamics of a two-regime Markov chain. Furthermore, on the right-hand side, we generated a Markov chain of length 200 with $Q_{11} = 0.98$ and $Q_{22} = 0.96$. In this case, we see that there are 5 switches. In the next two sections, we develop the estimation and forecasting procedures for the Gaussian HMM. All computations and estimations described next are done using the CRAN package *GaussianHMM1d* (https://CRAN.R-project.org/package=GaussianHMM1d) [*Nasri and Rémillard*, 2019b].

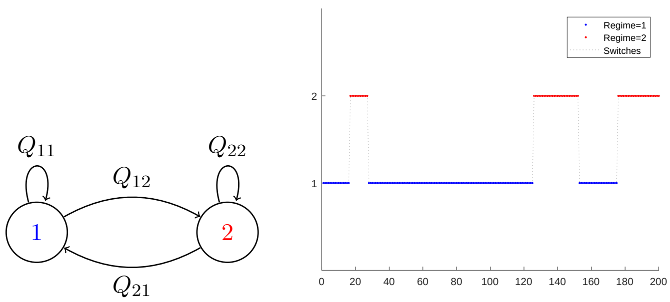


Figure 4. Simulation of a hidden Markov model with two regimes and 2×2 transition matrix Q where $Q_{11} = 0.98$, $Q_{12} = 0.02$, $Q_{21} = 0.04$, and $Q_{22} = 0.96$.

2.3.2 Gaussian HMM

Let Y be the variable of interest and let $y_1, \ldots, y_n$ be the observations for periods $t \in \{1, \ldots, n\}$. Further let $\tau_1, \ldots, \tau_n$ be the non-observable regimes, modeled by a homogeneous discrete-time Markov chain on $S = \{1, \ldots, m\}$ with transition probability matrix Q on $S \times S$. Given $\tau_1, \ldots, \tau_n$, the observations $y_1, \ldots, y_n$ are independent with densities $f_{\beta_{\tau_t}}$, $t \in \{1, \ldots, n\}$. Set $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m, Q)$, where in the Gaussian HMM, $\boldsymbol{\beta}_j = (\mu_j, \sigma_j)$ (the parameters of the Gaussian distribution), $j \in 1, \ldots, m$. Then, the joint density of $(\tau_1, \ldots, \tau_n)$ and $(y_1, \ldots, y_n)$ is given by

$$f_{\boldsymbol{\theta}}(\tau_1, \ldots, \tau_n, y_1, \ldots, y_n) = \left(\prod_{t=1}^{n} Q_{\tau_{t-1},\tau_t} \right) \times \prod_{t=1}^{n} f_{\boldsymbol{\beta}_{\tau_t}}(y_t), \tag{1}$$

where τ_0 is the first hidden state.

Since the regimes $\tau_1, \ldots, \tau_n$ are not observable, an easy way to estimate the parameters for a fixed number of regimes m is to use the Expectation-Maximization (EM) algorithm [*Dempster et al.*, 1977], which proceeds in two steps: the E step, during which

$$E_{y_1,\ldots,y_n}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}\right) = \mathbb{E}_{\boldsymbol{\theta}^{(k)}}\{\log f_{\boldsymbol{\theta}}(\tau_1, \ldots, \tau_n, y_1, \ldots, y_n)|Y_1 = y_1, \ldots, Y_n = y_n\} \tag{2}$$

is computed, and the M step, where we compute

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} E_{y_1,\ldots,y_n}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}\right), \tag{3}$$

for $k = 0, \ldots, N$. Here N, fixed by the user, is the maximum number of iterations allowable to reach the optimality tolerance (eps), also fixed by the user. In this paper, we chose $N = 10000$ and $eps = 10^{-4}$. The equations related to the EM algorithm for the Gaussian HHM are described in Appendix A. They are implemented in the function *EstHMM1d.R* of the package *GaussianHMM1d*. To choose an optimal number of regimes m based on a given dataset, we can use the formal goodness-of fit test proposed by *Rémillard* [2013], who suggests choosing the smallest m for which the P-value is greater than 5%. In the literature, the selection of the number of regimes is usually based on a maximum likelihood criterion, such as AIC or BIC, see, e.g., Bracken et al. [2016]. This selection procedure only compares models without any knowledge of their validity. Note that this goodness-of-fit test is described in Appendix B and is implemented in the function *GofHMM1d.R* of the package *GaussianHMM1d*.

There are two ways to estimate the probability of being in regime j at period t: we can consider only the observations up to period t, and compute $\eta_t(j) = P(\tau_t = j|Y_1 = y_1, \ldots, Y_t = y_t)$, $j \in \{1, \ldots, m\}$, using formulas (A.2)–(A.3), or we can consider all the observations and compute $\lambda_t(j) = P(\tau_t = j|Y_1 = y_1, \ldots, Y_n = y_n)$, using formula (A.6). In both cases, the estimated regime at period t, denoted by $\hat{\tau}_t$, is the regime with the largest probability, i.e., $\eta_t(\hat{\tau}_t) \geq \max_{j\in\{1,\ldots,m\}} \eta_t(j)$ (resp. $\lambda_t(\hat{\tau}_t) \geq \max_{j\in\{1,\ldots,m\}} \lambda_t(j)$). Generally λ_t is used for estimating the regimes while η_t is used for prediction purposes since computing λ_t requires all observations. Note that both η_t and λ_t can be calculated using the function *EstHMM1d.R*, while the regimes can be estimated using the function *EstRegime.R* of package *GaussianHMM1d*. After selecting the optimal number of regimes and estimating the parameters, persistence can be measured in terms of the number of switches during the observed period, using the estimated regimes $\hat{\tau}_1, \ldots, \hat{\tau}_n$. More precisely, the number of switches is defined as

$R_n = \sum_{t=2}^{n} \mathbb{I}(\hat{\tau}_{t-1} \neq \hat{\tau}_t)$. It can also be calculated with the function *EstRegime.R* of package *GaussianHMM1d*. The smaller R_n is, the more persistent the series. We can also approximate the long-term probability ν_j of being in each regime $j \in \{1, \ldots, m\}$ by using the transition matrix and the definition of the stationary distribution of a Markov chain. These probabilities represent the average percentage of time spent in each regime.

2.3.3 Forecasting using Gaussian HMM

Suppose that we observed $Y_1, \ldots, Y_t$ and we want to forecast Y_{t+k}. Then the conditional density $f_{t+k|1:t}$ of Y_{t+k} given $Y_1, \ldots, Y_t$ is expressed as a mixture of the Gaussian densities $f_{\boldsymbol{\beta}_i}$, viz.

$$f_{t+k|1:t}(y, \boldsymbol{\theta}) = \sum_{i=1}^{m} f_{\boldsymbol{\beta}_i}(y) \left\{ \sum_{j=1}^{m} \eta_t(j) \left(Q^k\right)_{ji} \right\}, \tag{4}$$

which is also a mixture of the Gaussian densities $f_{\boldsymbol{\beta}_i}$ with weights $P(\tau_{t+k} = i | Y_1 = y_1, \ldots, Y_t = y_t) = \sum_{j=1}^{m} \eta_t(j) \left(Q^k\right)_{ji}$, for $i \in \{1, \ldots, m\}$. The conditional distribution function $F_{t+k|1:t}$ of Y_{t+k} given $Y_1, \ldots, Y_t$ is then expressed as

$$F_{t+k|1:t}(y, \boldsymbol{\theta}) = \sum_{i=1}^{m} \Phi\left(\frac{y - \mu_i}{\sigma_i}\right) \left\{ \sum_{j=1}^{m} \eta_t(j) \left(Q^k\right)_{ji} \right\}. \tag{5}$$

where Φ is the cumulative distribution function of the standard Gaussian distribution. Using Equation (5) we can compute the conditional median and more generally the conditional quantile function as the inverse of the conditional distribution function $F_{t+k|1:t}$, for which there is no explicit expression; the inverse must be computed numerically. A 95% prediction confidence interval for Y_{t+k} is given by $\left[F^{-1}_{t+k|1:t}(.025), F^{-1}_{t+k|1:t}(.975)\right]$. Note that as k increases, the behavior of Y_{t+k} becomes independent of its past [*Rémillard*, 2013, p.382-383], leading to constant prediction intervals. This is due to the fact that if the Markov chain with transition matrix Q is ergodic, then the conditional distribution of Y_{t+k} given $Y_1, \ldots, Y_t$, converges, as $k \to \infty$, to the stationary distribution with density $f(y) = \sum_{i=1}^{m} \nu_i f_{\boldsymbol{\beta}_i}(y)$ and distribution function $F(y) = \sum_{i=1}^{m} \nu_i \Phi\left(\frac{y-\mu_i}{\sigma_i}\right)$. The next period forecast is obtained bi letting $k = 1$. Finally, note that formulas (4)–(5), including the conditional quantile function, are implemented in the functions *ForecastHMMPdf.R*, *ForecastHMMPdf.R*, *GaussianMixtureCdf.R*, and *GaussianMixtureInv.R* of the package *GaussianHMM1d*.

2.3.4 Application of Gaussian HMM to our case study

In this paper, we consider the Gaussian HMM for modeling inflows for four basins: LG2, LG4, CAN, and CHU. For each basin, we have two times series: reconstructed data from 1800 to 2000, and observed data from 1960 to 2016. A logarithm transformation of the inflows (observed and reconstructed) was applied to obtain observations that are more suitable for a mixture of normal probability distributions, since fitting Gaussian or a mixture of Gaussian to the original data which are positive can lead to inconsistencies since the probability of observing a negative value under these models is nonzero. Therefore data log-transform is needed in our case. Data up to 1990 will be used to choose the appropriate Gaussian HMM according to the methodology described in Section 2.3.2. More precisely, for both log-inflows from the observed data (1960-1990) and reconstructed data (1800-1990), we

- *set the number of regimes (starting at 1) and estimate the Gaussian HMM parameters;*
- *compute the P-value of the goodness-of-fit test;*
- *choose the model with the smallest number of regimes for which the P-value is larger than 5%;*
- *compute the probabilities λ_t according to formula* (A.6) *and estimate the regime $\hat{\tau}_t$;*
- *calculate the number of switches R_n.*

The observed inflows from 1991 to 2016 will be used to evaluate the performance of the chosen models in hindcast experiments, as described in Section 2.3.3. Hindcasts are forecasts for past events, which enable models to be compared to observations to assess their skill. Here, for both inflows models for the observed data and reconstructed data, we

- *compute 95% long-term prediction intervals for the inflows Y_t, $t \in \{1991, \ldots, 2016\}$ using Equation* (5) *and data up to 1990;*
- *compute 95% 1-year prediction intervals for the inflows Y_t, $t \in \{1991, \ldots, 2016\}$ using Equation* (5) *with $k = 1$ and past data $Y_1, \ldots, Y_{t-1}$; η_t is upgraded each time using formula* (A.3) .

To assess the performance of each model for the prediction, two criteria were used. They are both based on the fact that if a model is appropriate, then $F_{t|1:t-1}(Y_t, \boldsymbol{\theta})$ are i.i.d. uniform variables [Bai, 2003] for the true parameter $\boldsymbol{\theta}$. Then, using observed values $Y_{1991}, \ldots, Y_{2016}$, we define the empirical distribution function

$$\hat{D}(u) = \frac{1}{26} \sum_{t=1991}^{2016} \mathbf{1}\{F_{t|1:t-1}(Y_t, \boldsymbol{\theta}_n) \leq u\}, \quad u \in [0,1]. \tag{6}$$

Under the hypothesis that the model is appropriate, $\hat{D}$ should be uniformly distributed. Note that here, instead of using only one statistic, we use the full distribution of the predicted values. As a result, we define two scoring rules based on $\hat{D}$, namely the Kolmogorov-Smirnov (ks) and the Cramér-von Mises (cvm) statistics defined respectively by

$$ks = \max_{u \in [0,1]} \sqrt{26} \left| \hat{D}(u) - u \right| \tag{7}$$

and

$$cvm = 26 \int_0^1 \left\{ \hat{D}(u) - u \right\}^2 du. \tag{8}$$

These metrics, which can only be used for the one-period prediction since $F_{t+k|1:t-1}(Y_t, \boldsymbol{\theta})$ is not uniform [Bai, 2003], are negatively oriented, in the sense that smaller values of ks and cvm indicate more reliable probabilistic forecasts. Note that Equations (6)–(8) can be used for any dynamic model, not just Gaussian HMM. Finally, we use these statistics only as metric scores, not for goodness-of fit testing.

3 Results

After calibration and verification of the reconstructions from the individual proxies, the time series are combined into a single reconstruction for each basin (Section 3.1). Gaussian HMMs are then fitted to the reconstructions and to the recent observation time series (Section 3.2). These Gaussian HMMs are used to hindcast water inflows to compare the benefits, if any, of using longer, less accurate, annual inflow reconstructions in an operational setting (Section 3.3).

3.1 Calibration and validation of the reconstructions

Calibration statistics between tree-ring series and annual (Jan-Dec) inflow data (1960–2000) indicate that our proxy network and modeling approach can be used to reconstruct past water supplies beyond hydrological observations across the Québec-Labrador Peninsula. In LG2 and LG4, Method 1 yields the highest calibration R^2 statistics (0.74 and 0.79, respectively, see Table 3). In CAN and CHU, the highest calibration R^2 statistics are obtained by Method 2. The RMSE values (Table 3) are generally smaller than the standard deviations calculated for the 1960–2000 period for most methods (Table 4), which indicates that the reconstruction models are more accurate than the mean for prediction purposes. However, the verification R^2_p and RMSE_p statistics suggest that Method 1 generally has lower predictive skill (lower R^2_p, higher RMSE_p). By contrast, Method 3 has the best predictive skill, with the highest R^2_p and lowest RMSE_p. Combining the three reconstructions for each basin produces reconstructions with high calibration R^2 statistics: 0.70 (LG2), 0.75 (LG4), 0.64 (CAN) and 0.76 (CHU) (Table 3). Except for CHU, these statistics are well within the bounds of those of the three models used to compute the individual reconstructions, indicating that the composite reconstructions (COMP) integrate the strengths and possible weaknesses associated with each method and proxy selection. The four composite reconstructions extend the inflow records back to 1800 CE for each hydroelectric reservoir under study (Figure 5).

From Table 4, we see that the main descriptive statistics are comparable for the reconstructed and the observed data, with the exception of the standard deviations, which are about 33% larger for the

Table 3. Statistical results for the annual water inflows reconstructions: series_nb represents the number of series involved in each method, R^2 (resp., R_p^2) is the determination coefficient calculated on calibration data (resp., verification data), and RMSE (resp., $RMSE_p$) is the root-mean-square error calculated on calibration data (resp., verification data).

Statistics	series_nb	R^2	R_p^2	RMSE	$RMSE_p$	series_nb	R^2	R_p^2	RMSE	$RMSE_p$
Basin	LG2					LG4				
Method 1	43	0.74	0.38	0.14	0.22	60	0.79	0.20	0.13	0.25
Method 2	14	0.63	0.46	0.17	0.21	12	0.72	0.46	0.15	0.21
Method 3	14	0.53	0.53	0.18	0.17	12	0.50	0.54	0.20	0.18
COMP	-	0.70	0.69	0.15	0.16	-	0.75	0.75	0.14	0.14
Basin	CAN					CHU				
Method 1	50	0.68	0.40	0.15	0.20	61	0.71	0.45	0.17	0.20
Method 2	10	0.70	0.52	0.14	0.18	12	0.75	0.60	0.12	0.14
Method 3	10	0.69	0.66	0.14	0.18	12	0.68	0.68	0.15	0.15
COMP	-	0.64	0.65	0.17	0.18	-	0.76	0.77	0.12	0.13

Table 4. Descriptive statistics for each basin and for two time periods: observations (1960–2000) and reconstructions (1800–2000). NS= Not significant. *= *P*-value <0.05

		1960-2000		1800-2000	
Basin	Area (km^2)	Mean (Std.Dev.) (mm /day)	Trend	Mean (Std.Dev.) (mm /day)	Trend
LG2	30 989	1.20 (0.28)	NS	1.16 (0.21)	NS
LG4	28 440	1.67 (0.28)	NS	1.65 (0.21)	0.03*
CAN	37 330	1.73 (0.28)	NS	1.72 (0.21)	NS
CHU	69 141	1.73 (0.26)	NS	1.74 (0.21)	NS

observed data. However, these differences will have no significant impact in the results of the next two sections.

3.2 Regimes and persistence for the reconstructed and observed data

Following Section 2.3.4, we choose the appropriate Gaussian HMM for observed and reconstructed data. Based on the goodness-of-fit test proposed in *Rémillard* [2013] and described in Ap-

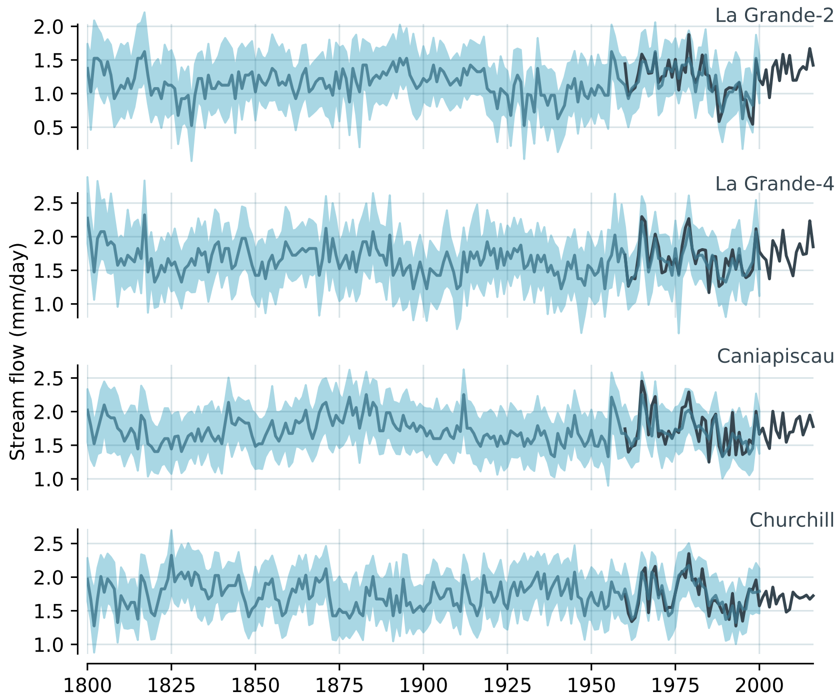


Figure 5. Composite reconstructions of annual water supply (mm/day) back to 1800 CE for each watershed. For each year reconstructed, the bold line corresponds to the mode of the joint distribution (see Figure 3) and the envelope represents the 90% bootstrap confidence interval. Observed annual water supplies are overlaid in black.

pendix B, the selected model for the reconstructed log-transformed data is a Gaussian HMM with two regimes. The P-values, listed in Table 5, are estimated with $B = 10000$ bootstrap samples, and all P-values for the reconstructed data and HMM with two regimes are larger than 5%. Note that for the LG4 watershed, we could not reject the Gaussian HMM with only one regime. Since the P-value is only slightly greater than 5%, the Gaussian HMM with two regimes is preferable. Therefore, the mean for the original reconstructed data for regime j is given by $\mu_{dat_j} = \exp\left(\mu_j + \sigma_j^2/2\right)$, $j \in \{1, 2\}$, and Regime 1 is defined as the wetter regime. Table 6 gives the estimated parameters for the selected models for each reconstructed time series. The behavior of the four studied stations is quite different in terms of persistence. In fact, as measured by the number of switches, the persistence decreases from western (LG2, $R_n = 6$; LG4, $R_n = 12$) to eastern watersheds (CAN, $R_n = 20$; CHU, $R_n = 30$). Similar behavior is observed in terms of the average time spent in the wetter regimes before switching as a function of Q_{11}, which is larger for LG2 ($Q_{11} = 0.975$) and LG4 ($Q_{11} = 0.937$) than for CAN ($Q_{11} = 0.867$) and CHU ($Q_{11} = 0.804$). Furthermore, for the LG2 and LG4 basins, as measured by ν_1, at least 60% of the time is spent in the wetter regime. For CAN, the opposite behavior is observed, i.e., 41% of the time is spent in Regime 1. For the CHU basin, the percentage of time spent in Regime 1 is not significantly different from that spent in Regime 2 (50%). The difference in the number of switches of the four studied time series is also displayed graphically in Figure 6.

Next, for the observed data, Table 5 shows that a Gaussian HMM model with one regime is selected for the LG4, CAN and CHU watersheds, while for LG2, two regimes are selected. Table 6 also shows the statistical results related to the observed data. Note that during the overlapping period, the reconstructed data and observed data for the LG2 basin behave similarly in terms of the number of switches.

Table 5. P-values in percentage for the Gaussian HMM for each basin and for two time periods: observations (1960–1990) and reconstructions (1800–1990), the symbol * indicates that the P-values for the HMM are very close to 5%; in this case, one might also use the model with one more regime.

	1960–1990		1800-1990		
Basin	HMM1	HMM2	HMM1	HMM2	HMM3
LG2	4.18	25.47	0.03	21.23	10.53
LG4	38.80	5.29	5.35	50.06*	55.22
CAN	89.09	70.59	1.75	8.11	38.85
CHU	20.39	74.80	1.54	17.20	67.06

Finally, to illustrate the fit of the reconstructed data with the proposed GHMM models, for each station, we estimated the density of the log data using the kernel method, we plotted the Gaussian densities for each of the two regimes, and we plotted the mixture of these two regimes using the weights ν_1, ν_2 given in Table 6 since the density obtained using the kernel method is an estimation of the stationary density. These results are displayed in Figure 7. We can see that the density estimated with the kernel method is always close to the density estimated by the mixture.

3.3 Hindcast experiments

In this section, we evaluate the predictive ability of Gaussian HMMs in hindcast experiments. We compare observation data with 1-year hindcasts and longer-term hindcasts based on the predictive distribution functions expressed by Equation (5). In the former case, we used the observed data to improve the hindcast. In addition, we compute the Kolmogorov-Smirnov and Cramér-von Mises scores

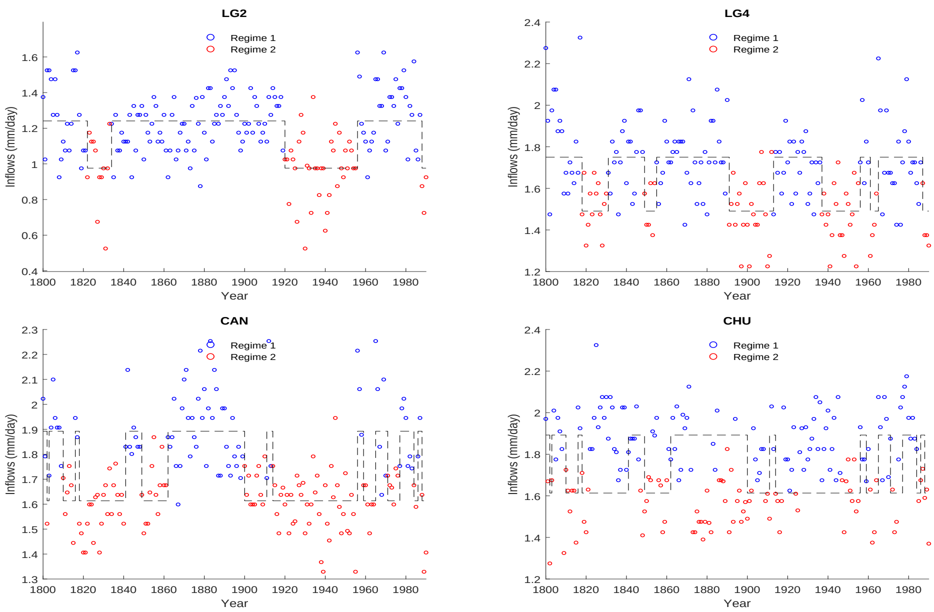


Figure 6. Estimated regimes for the four stations over 190 years. The dashed lines represent the regime jumps, where the horizontal parts are the mean of each regime.

to compare the quality of the 1-year hindcasts provided by the chosen models using the observed and reconstructed data.

The 95% prediction intervals for the period 1991-2016 are displayed in Figure 8. Based on the left-hand side graphs, we can see that the prediction intervals are almost constant after three years, which shows that the predictive distribution converges rapidly to the stationary distribution. By contrast, for 1-year hindcasts, the prediction intervals on the right-hand side for the Gaussian HMM model with two regimes vary considerably since we incorporate new information each year. Furthermore, the prediction intervals for the Gaussian HMM with only one regime are constant over time and are much less informative. The most important feature of Figure 8 is that the prediction intervals computed with the reconstructed data are more precise than those based on the observed data, mainly due to the fact that the reconstructed datasets are longer, and as a result, they help in building more accurate models.

Table 6. Estimated parameters for the Gaussian HMM with two regimes for the four reconstructed time series from 1800 to 1990 and the four observed data from 1960 to 1990. μ_j (mm/day) and σ_j (mm/day) are, respectively, the mean and standard deviation of regime j for the logarithmic-transformed series, $\mu_{dat\,j}$ (mm/day) is the mean of regime j for the series, Q_{jj} is the probability of staying in regime j, $\tau_j = \frac{Q_{jj}}{1-Q_{jj}}$ (year) is the average time spent in regime j before changing regimes, ν_j (%) is the proportion of time spent in regime j, and R_n is the number of switches.

	1960–1990					1800-1990							
Station	LG2		LG4	CAN	CHU	LG2		LG4		CAN		CHU	
Regime	1	2	1	1	1	1	2	1	2	1	2	1	2
μ_j	0.267	-0.165	0.511	0.564	0.556	0.206	-0.046	0.554	0.394	0.634	0.475	0.632	0.452
$\mu_{dat\,j}$	1.321	0.876	1.692	1.781	1.765	1.241	0.976	1.750	1.490	1.892	1.613	1.887	1.576
σ_j	0.151	0.261	0.171	0.160	0.150	0.135	0.211	0.101	0.097	0.084	0.074	0.073	0.082
Q_{jj}	0.962	0.996	–	–	–	0.975	0.950	0.937	0.904	0.867	0.916	0.836	0.804
τ_j	25.76	271.8	–	–	–	39.52	19.17	14.97	9.44	6.566	10.92	5.109	4.116
ν_j	86.63	13.3	–	–	–	73.21	26.78	63.43	36.56	40.99	59.00	55.48	44.51
R_n	2		–	–	–	6		12		20		30	

This conclusion is consolidated by the results displayed in Table 7 for 1-year forecasts, which show that the ks and cvm scores are all smaller for the model based on reconstructed data. In addition, we computed the mean absolute deviation (MAD) and the root mean square error (RMSE) between the predicted means and the observations from 1991 to 2016. The results are given in Table 8. The MAD values are all smaller for the reconstructed values. The same is true for the RMSE for all stations but LG4, but the difference between the two RMSE values is quite small (0.008). The overall conclusion from Tables 7 and 8 is that the model based on the reconstruction yields in general better predictions that the model based on the observed data.

Using the results of 1-year forecasts, we computed the 26 years predictive probabilities of being in Regime 2 (dry). Figure 10 shows these probabilities for the four basins. Note that the LG2 and LG4 basins spend respectively 16 and 18 years in the wet regime (probability of being in Regime 2 below 0.5), while CAN and CHU spend respectively 10 and 5 years in the wet regime. Over the predictive period, LG2 has only one regime switch, while the others have almost 9 regime switches. To assess the performance of the regime prediction, we performed 10,000 Monte Carlo simulations using

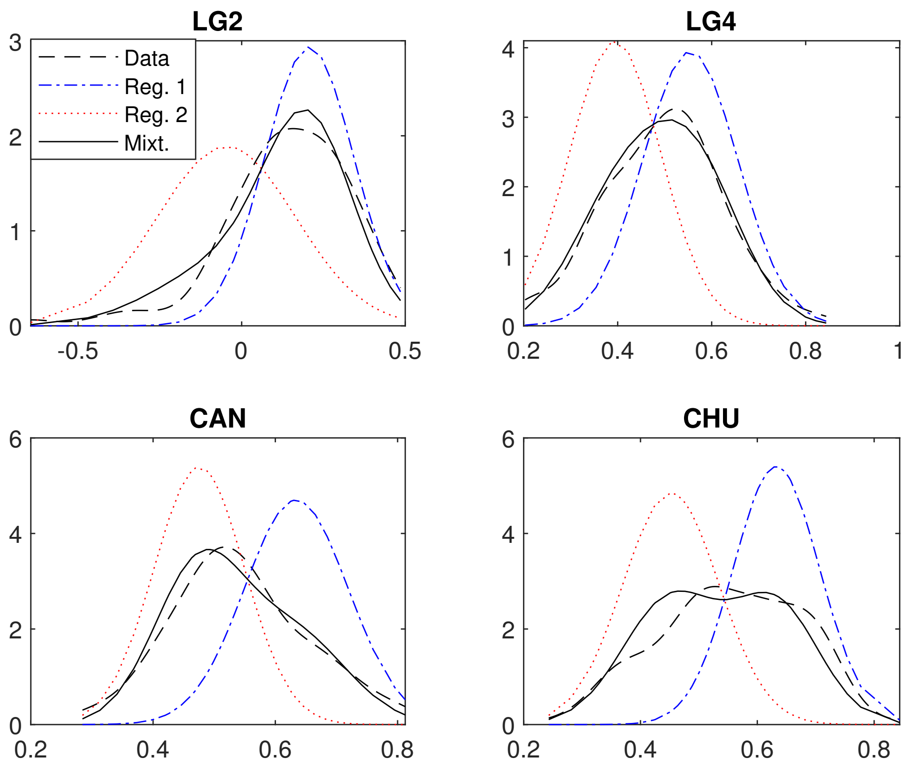


Figure 7. Estimated densities for the log data using the kernel method, together with the Gaussian densities for each of the two regimes and the stationary density.

the same estimated parameters obtained for the four basins over the period 1800–1990. We found that using a series of length 190 for the estimation of the parameters and with 26 observations used for the predictions, the regimes are predicted correctly about 90% of the time. It means, in our case, that on average less than 3 regimes might be incorrectly predicted. Finally, as a complement of information, we included in the Supplementary Material a video showing the evolution of the predictive densities for the four basins calculated using Equation (4) for the reconstructed and the observed data. We can see clearly that when there is a large probability of being in Regime 1 (resp. 2), the predictive density from the reconstructed data is shifted to the right (resp. left) and has basically mean μ_{dat} given in Table 6. For the observed data, we can see that for LG2, the predictive densities have similar behavior as for the reconstructed data, while for the other three basins the predictive densities from the observed data are constant since the best model in these cases are Gaussian HMM with one regime.

Finally, we computed the prediction densities for each model and each station for 2005 and 2015, together with the observed value for the year. These graphs are displayed in Figure 9.

Table 7. Scores for 1-year forecasts for the period 1991-2016 based on the selected models for reconstructed data (1800–1990) and for observed data (1960–1990).

Score	ks				cvm			
Station	LG2	LG4	CAN	CHU	LG2	LG4	CAN	CHU
Reconstructed	1.396	0.622	1.000	0.787	0.630	0.102	0.216	0.159
Observed	1.563	0.664	1.056	1.571	0.807	0.102	0.269	0.611

Table 8. Mean absolute deviation (MAD) and root mean square errors (RMSE) for 1-year forecasts for the period 1991-2016 based on the selected models for reconstructed data (1800–1990) and for observed data (1960–1990).

Score	MAD				RMSE			
Station	LG2	LG4	CAN	CHU	LG2	LG4	CAN	CHU
Reconstructed	0.192	0.175	0.187	0.125	0.237	0.232	0.223	0.166
Observed	0.208	0.179	0.190	0.138	0.251	0.224	0.226	0.190

4 Discussion

This study used an extensive and well-replicated multiproxy tree-ring network to produce the first, spatially explicit, hydrological reconstructions across the Québec-Labrador Peninsula. The reconstructions shed light on the fundamental properties of multidecadal hydrological variability in one of North America's largest hydroelectric facilities. The reconstructions result from the combination of three approaches based on partial least squares, stepwise partial least squares, and the best analogue method. Considered individually, all approaches yield satisfactory calibration (R^2, RMSE) and verification (R_p^2, RMSEp) statistics (Table 3). Combining the reconstructions into a single reconstruction (for each basin) produced 200-year time series that integrate the strengths and weaknesses of each approach, while explaining between 65% and 76% of the variance in the original water supply series. Hence, for each basin, the variance explained by the combined reconstructions appears to be well within the bounds of other works that used tree-rings to reconstruct streamflow in river systems used for hydroelectric production [*Woodhouse and Lukas*, 2006].

Our new 200-year reconstructed datasets were used to highlight and hindcast several hydrological regimes by using the Gaussian HMM. In fact, this type of model has been used in several studies to represent hydrological time series, including reconstructed data as in Bracken et al. [2016]. The advantage of such models is that they are not only able to detect hydrological regimes and classify the observations into regimes, but they can also produce short-term and long-term forecasts for the future regimes. In this study, we used the Gaussian HMM, which is the simplest regime-switching model. The originality of our statistical approach is the selection procedure of the number of regimes, which is based on a recent goodness-of-fit test proposed by [*Rémillard*, 2013]. Usually, the selection of the number of regimes is based on a maximum likelihood criterion, such as AIC or BIC, which only rank models without verifying if they are valid. In our approach, we test the validity of the models when choosing the number of regimes. However, as expected, detecting more than one regime requires longer datasets, which is true for any stochastic model. The power of the goodness-of-fit test has been studied in a similar context [Nasri et al., 2019] and the authors showed that the selection procedure based on P-values is valid and efficient.

The inference and model selection results presented in Section 3.2 for the reconstructed time series confirm that the annual inflows of the four basins exhibit persistence in several distinct states with occasional transitions between these states. Such information can be very useful for electric utilities, such as Hydro-Québec. For instance, these results can be used to define energy reliability criteria that could take into account possible switching hydrological regimes and events of prolonged low inflows. In this case, access to longer series (reconstructed data) helps to identify more accurate models and clearly improves forecasts. Moreover, taking into account new data reduces the uncertainty in the forecasts. The hindcast experiments performed in Section 3.3 with observed data from 1991 to 2016 clearly show that we can trust the reconstructed data and use them for short-term, as well as for medium-term predictions. In particular, for 1-year ahead forecasts, according to the ks and cvm scores, as well as the MAD and RMSE scores, the approach based on the reconstructed data outperforms the standard method which relies only upon the observed data. For instance, the cvm scoring rule shows a clear advantage for our approach for basins LG2, CAN and CHU. These results are in agreement with the conclusions of Thyer et al. [2006] who showed that the uncertainty around the estimation of parameters of Gaussian HMMs is quite large for short time series. As a result, the identification of the correct model is very difficult. In our case, for the observed data, we used only 31 observations, which is too small to perform an efficient calibration. However, based on the hindcast experiments, we showed that the reconstructed data can be combined with observed data to get more accurate and precise predictions. This result is perhaps the most important result from this study.

As an example of the usefulness of the proposed model, we can also predict future regimes, as illustrated in Figure 10, where we compute the probability of dry regimes for each station, using both reconstructed and new observed data. Note that the dry regime (Regime 2) is much less recurrent for stations LG2 and LG4 than for stations CAN and CHU, which is important information for the management of water resources. In fact, the ability to issue such medium-term forecasts with respect to hydroelectric generation at the beginning of the year is of particular interest to water resources planners and managers. Such knowledge can be used for making decisions about future releases during the winter, contributing to more proactive water management that may prove very useful in extreme dry or wet years. For example, the explicit integration of basin-specific regime properties in Table 6 could allow hydroelectric producers to make informed inflow predictions based on the current hydrological regime (Regime 1 or Regime 2), taking into account the flow characteristics (mean and variability of a given regime) and the regime-switching probabilities associated with each basin. In turn, the proposed model improves the quality of predictions by lowering the risk of mismatches between energy production and demand.

5 Conclusion

The objective of this study was threefold. First, we reconstructed 200 years of annual water-supplies at four basins in the Québec-Labrador Peninsula, which are among the largest in North America in terms of hydroelectric capacity. We used tree-ring proxies to extend the climatic series beyond recent 40 years observations of reservoir inflows. The reconstructed data were based on the combination of three statistical methods, as in *Nicault et al.* [2014a]. Second, for the reconstructed and observed data, we used Gaussian HMM to characterize the persistence in terms of regime switches. Two regimes were found for the reconstructed series, while only one regime was found for the observed data, in three out of four basins. As for the number of regime switches, we noticed that they increase significantly from west to east. In Quebec-Labrador, hydroclimate variability over decadal to multi-decadal time scales can, at least party, be related to ocean-atmosphere interactions occurring in the western North-Atlantic region [O'Reilly et al., 2017]. Indeed, oscillations in sea surface temperatures (SST) exhibit a significant persistence which has been shown to impact surrounding landmasses climate, most particularly low-frequency temperature variability over northeastern North-America. Whether this influence results from a direct thermodynamical influence or an indirect change in large-scale atmospheric circulation patterns, however, remains unclear. The analysis of such oscillatory modes nevertheless confirms their potential relevance for streamflow predictability in Quebec-Labrador region [Sveinsson et al., 2008a,b].

Third, we evaluated the predictive ability of the selected Gaussian HMM for each basin. The results showed that the predictions are better using the reconstructed data, when combined with the new observed data. This is mainly due to the fact that the reconstructed datasets are longer and reliable, allowing therefore efficient model selection and more accurate probabilistic forecasts. However, the expectations of water resource manager's are considerably higher. Performance over a longer reconstructed period and a rigorous assessment of the hindcast skill against other approaches are required. This assessment would include comparing different persistence models and providing a theoretical foundation connecting them to continental climate patterns. Indeed, future research may lead to millennial-scale reconstructions, taking into account the serial dependence of tree-ring proxies, which will enable us to produce more efficient reconstructions, and in particular reduce the difference between the variances of the reconstruction and the observed data. Also, longer reconstructed datasets will allow to consider other persistence models, such as the well-known ARFIMA models, which require very long datasets [*Bhardwaj and Swanson*, 2006]. Ideally, persistence models would also include climate change considerations, along with the uncertainty of climate sensitivity to greenhouse gases concentrations. Adding predictors that drive multidecadal variation to the model, such as large-scale climate indices, would certainly help to better explain the variability in regimes [Sveinsson et al., 2008a,b]. One way to do this would be to incorporate a probit model for the hidden regimes of the HMM [*Perreault et al.*, 2007; *Bracken et al.*, 2014]. Finally, in a forthcoming paper, we will develop goodness-of-fit tests to rigorously compare ARFIMA to HMM, and we will attempt to account for the uncertainties inherent to reconstruction procedures to provide a more robust foundation for risk analysis.

Acknowledgments

We would like to thank the Associate Editor and two referees for their comments that lead to an improved version of the manuscript. This research is partially supported by the Canadian Statistical Sciences Institute (CANSSI), the Groupe d'étude et de recherche en analyse des décisions (GERAD), the Fonds québécois de la recherche sur la nature et les technologies (FRQNT), the Natural Sciences and Engineering Research Council of Canada (NSERC), Ouranos, Hydro-Québec and Manitoba Hydro. We thank Frédéric Guay for his help with Figure 2. All data used in this paper, with the exception of observed inflows, are available at site `www.dendro-qc-lab.ca`. Inflow observations are proprietary to Hydro-Québec and are strategic information for the company. They can only be made available on request from Hydro-Québec.

References

Allen, D. T., V. M. Torres, J. Thomas, D. W. Sullivan, M. Harrison, A. Hendler, S. C. Herndon, C. E. Kolb, M. P. Fraser, A. D. Hill, B. K. Lamb, J. Miskimins, R. F. Sawyer, and J. H. Seinfeld (2013), Measurements of methane emissions at natural gas production sites in the United States, *Proceedings of the National Academy of Sciences*, *110*(44), 17,768–17,773, doi:10.1073/pnas.1304880110.

Ault, T. R., J. E. Cole, J. T. Overpeck, G. T. Pederson, S. St. George, B. Otto-Bliesner, C. A. Woodhouse, and C. Deser (2013), The continuum of hydroclimate variability in western North America during the last millennium, *Journal of Climate*, *26*(16), 5863–5878.

Ault, T. R., J. E. Cole, J. T. Overpeck, G. T. Pederson, and D. M. Meko (2014), Assessing the risk of persistent drought using climate model simulations and paleoclimate data, *Journal of Climate*, *27*(20), 7529–7549.

Axelson, J. N., D. J. Sauchyn, and J. Barichivich (2009), New reconstructions of streamflow variability in the South Saskatchewan River Basin from a network of tree-ring chronologies, Alberta, Canada, *Water Resources Research*, *45*(9).

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *The Review of Economics and Statistics*, 85(3):531–549.

Beran, J. (1994), *Statistics for Long-Memory Processes*, *Monographs on Statistics and Applied Probability*, vol. 61, Chapman and Hall, New York.

Bhardwaj, G., and N. R. Swanson (2006), An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series, *Journal of Econometrics*, *131*(1), 539 – 578.

Boreux, J.-J., P. Naveau, O. Guin, L. Perreault, and J. Bernier (2009), Extracting a common high frequency signal from northern Québec black spruce tree-rings with a Bayesian hierarchical model, *Climate of the Past*, *5*(4), 607–613.

Boucher, É., T. B. Ouarda, Y. Bégin, and A. Nicault (2011a), Spring flood reconstruction from continuous and discrete tree-ring series, *Water Resources Research*, *47*(7).

Boucher, E., J. Guiot, and E. Chapron (2011b), A millennial multi-proxy reconstruction of summer pdsi for Southern South America, *Climate of the Past*, *7*, 957–974.

Boucher, E., A. Nicault, D. Arseneault, Y. Bégin, and M. P. Karami (2017), Decadal variations in Eastern Canada's taiga wood biomass production forced by ocean-atmosphere interactions, *Scientific Reports*, *7*(1), 2457.

Bracken, C., B. Rajagopalan, and E. Zagona (2014), A hidden markov model combined with climate indices for multidecadal streamflow simulation, *Water Resources Research*, *50*(10), 7836–7846.

Bracken, C., Rajagopalan, B., and Woodhouse, C. (2016). A Bayesian hierarchical nonhomogeneous hidden Markov model for multisite streamflow reconstructions. *Water Resources Research*, 52(10):7837–7850.

Briffa, K. R., T. J. Osborn, F. H. Schweingruber, I. C. Harris, P. D. Jones, S. G. Shiyatov, and E. A. Vaganov (2001), Low-frequency temperature variations from a northern tree-ring density network, *Journal of Geophysical Research: Atmospheres*, *106*(D3), 2929–2941.

Brigode, P., F. Brissette, A. Nicault, L. Perreault, A. Kuentz, T. Mathevet, and J. Gailhard (2016), Streamflow variability over the 1881–2011 period in Northern Québec: comparison of hydrological reconstructions based on tree rings and geopotential height field reanalysis, *Climate of the Past*, *12*(9), 1785–1804.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc. Ser. B*, *39*, 1–38.

Diebold, F. X., and A. Inoue (2001), Long memory and regime switching, *Journal of Econometrics*, *105*(1), 131 – 159.

Diffenbaugh, N. S. D. L. Swain and D. Touma (2015), Anthropogenic warming has increased drought risk in California, *Proceedings of the National Academy of Sciences*, *112*(13), 3931–3936.

Evin, G., J. Merleau, and L. Perreault (2011), Two-component mixtures of normal, gamma, and gumbel distributions for hydrological applications, *Water Resources Research*, *47*(8).

Genest, C., B. Rémillard, and D. Beaudoin (2009), Omnibus goodness-of-fit tests for copulas: A review and a power study, *Insurance Math. Econom.*, *44*, 199–213.

Gennaretti, F., D. Arseneault, A. Nicault, L. Perreault, and Y. Bégin (2014), Volcano-induced regime shifts in millennial tree-ring chronologies from Northeastern North America, *Proceedings of the National Academy of Sciences*, *111*(28), 10,077–10,082.

Girard, F., S. Payette, and R. Gagnon (2008), Rapid expansion of lichen woodlands within the closed-crown boreal forest zone over the last 50 years caused by stand disturbances in Eastern Canada, *Journal of Biogeography*, *35*(3), 529–537.

Guiot, J., A. Nicault, C. Rathgeber, J.-L. Edouard, F. Guibal, G. Pichard, and C. Till (2005), Last-millennium summer-temperature variations in Western Europe based on proxy data, *The Holocene*, *15*(4), 489–500.

Guiot, J., C. Corona, et al. (2010), Growing season temperatures in Europe and climate forcings over the past 1400 years, *PloS one*, *5*(4), e9972.
https://www.overleaf.com/project/5bd71625cd89984c92d5e477

Hamilton, J. D. (1990), Analysis of time series subject to changes in regime, *J. Econometrics*, *45*(1-2), 39–70.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York, Second Edition.

Hoel, P. G., Port, S. C., and Stone, C. J. (1972). *Introduction to Stochastic Processes*. Houghton Mifflin Co., Boston, Mass. The Houghton Mifflin Series in Statistics.

Hofgaard, A., J. Tardif, and Y. Bergeron (1999), Dendroclimatic response of picea mariana and pinus banksiana along a latitudinal gradient in the Eastern Canadian boreal forest, *Canadian Journal of Forest Research*, *29*(9), 1333–1346.

Holmes, R. L. (1983), Computer-assisted quality control in tree-ring dating and measurement, *Tree-ring bulletin*.

Hosking, J. R. M. (1984), Modeling persistence in hydrological time series using fractional differencing, *Water Resources Research*, *20*(12), 1898–1908.

Hydro-Québec Distribution (2014), 2014 Québec Balancing Authority Area Comprehensive Review of Resource Adequacy, *Tech. rep.*, NPCC.

Hydro-Québec Production (2018), Annexe A - Respect du critère de fiabilité en énergie pour les approvisionnements provenant d'Hydro-Québec Production, *Tech. rep.*, Hydro-Québec.

Jandhyala, V. K., P. Liu, and S. B. Fotopoulos (2009), River stream flows in the Northern Québec Labrador region: A multivariate change point analysis via maximum likelihood, *Water resources research*, *45*(2).

Koutsoyiannis, D. (2005), Hydrologic persistence and the Hurst phenomenon, in *Water Encyclopedia*, edited by J. H. Lehr and J. Keeley, pp. 210–221, Wiley, New York.

Lara, A., A. Bahamondez, A. González-Reyes, A. A. Muñoz, E. Cuq, and C. Ruiz-Gómez (2015), Reconstructing streamflow variation of the Baker River from tree-rings in Northern Patagonia since 1765, *Journal of Hydrology*, *529*(P2), 511–523, doi:10.1016/j.jhydrol.2014.12.007.

LeBlanc, M. and Tibshirani, R. (1996). Combining estimates in regression and classification. *Journal of the American Statistical Association*, 91(436):1641–1650.

Loaiciga, H. A., L. Haston, and J. Michaelsen (1993), Dendrohydrology and long-term hydrologic phenomena, *Reviews of Geophysics*, *31*(2), 151–171.

Mandelbrot, B. B., and J. R. Wallis (1968), Noah, Joseph, and operational hydrology , *Water Resources Research*, *4*(5), 909–918.

McCarroll, D., M. H. Gagen, N. J. Loader, I. Robertson, K. J. Anchukaitis, S. Los, G. H. Young, R. Jalkanen, A. Kirchhefer, and J. S. Waterhouse (2009), Correction of tree-ring stable carbon isotope chronologies for changes in the carbon dioxide content of the atmosphere, *Geochimica et Cosmochimica Acta*, *73*(6), 1539–1547.

Meko, D. M., and C. A. Woodhouse (2005), Tree-ring footprint of joint hydrologic drought in Sacramento and Upper Colorado river basins, Western USA, *Journal of Hydrology*, *308*(1-4), 196–213.

Meko, D. M., M. D. Therrell, C. H. Baisan, and M. K. Hughes (2001), Sacramento river flow reconstructed to ad 869 from tree-rings, *Journal of the American Water Resources Association*, *37*(4), 1029–1039.

Merleau, J. (2017), Analyse probabiliste des moyennes historiques des apports énergétiques de certains systèmes hydriques du parc d'Hydro-Québec aux échelles temporelles mensuelle, saisonnière et annuelle, *Tech. Rep. IREQ-2017-XXX*, Hydro-Québec Research Institute.

Merleau, J. (2018), Analyse probabiliste des moyennes historiques des apports énergétiques de certains systèmes hydriques du parc d'Hydro-Québec aux échelles temporelles mensuelle, saisonnière et annuelle - Partie II, *Tech. Rep. IREQ-2018-XXX*, Hydro-Québec Research Institute.

Mundo, I. A., M. H. Masiokas, R. Villalba, M. S. Morales, R. Neukom, C. Le Quesne, R. B. Urrutia, and A. Lara (2012), Multi-century tree-ring based reconstruction of the Neuquén River streamflow, Northern Patagonia, Argentina, *Climate of the Past*, *8*(2), 815–829.

Nasri, B. R., and B. N. Remillard (2019), Copula-based dynamic models for multivariate time series, *Journal of Multivariate Analysis*, *172*, 102–121.

Nasri, B. R., and B. N. Rémillard (2019), *GaussianHMM1d*, R package version 1.0.1.

Nasri, B. R., Rémillard, B. N., and Thioub, M. Y. (2020). Goodness-of-fit for regime-switching copula models with application to option pricing. *The Canadian Journal of Statistics*, in press.

Nasri, B. R., Rémillard, B. N., and Thioub, M. Y. (2019). Goodness-of-fit for regime-switching copula models with application to option pricing. Technical report, SSRN Working Paper Series No. 3347348.

National Energy Board (2017), Canada's Renewable Power Landscape - Energy Market Analysis, *Tech. rep.*, National Energy Board, doi:10.1016/j.msea.2008.10.047.

Naulier, M., M. M. Savard, C. Bégin, J. Marion, D. Arseneault, and Y. Bégin (2014), Carbon and oxygen isotopes of lakeshore black spruce trees in Northeastern Canada as proxies for climatic reconstruction, *Chemical Geology*, *374*, 37–43.

Nicault, A., S. Alleaume, S. Brewer, M. Carrer, P. Nola, and J. Guiot (2008), Mediterranean drought fluctuation during the last 500 years based on tree-ring data, *Climate dynamics*, *31*(2-3), 227–245.

Nicault, A., E. Boucher, C. Bégin, J. Guiot, J. Marion, L. Perreault, R. Roy, M. Savard, and Y. Bégin (2014a), Hydrological reconstruction from tree-ring multi-proxies over the last two centuries at the Caniapiscau Reservoir, northern Québec, Canada, *Journal of Hydrology*, *513*, 435–445.

Nicault, A., E. Boucher, D. Tapsoba, D. Arseneault, F. Berninger, C. Bégin, J.-L. DesGranges, J. Guiot, J. Marion, S. Wicha, et al. (2014b), Spatial analysis of black spruce (picea mariana (mill.) bsp) radial growth response to climate in northern Québec–Labrador Peninsula, Canada, *Canadian Journal of*

Forest Research, *45*(3), 343–352.

O'Reilly, C. H., Woollings, T., and Zanna, L. (2017). The dynamical influence of the atlantic multidecadal oscillation on continental climate. *Journal of Climate*, 30(18):7213–7230.

Pelletier, J. D., and D. L. Turcotte (1997), Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment, *Journal of Hydrology*, *203*(1-4), 198–208.

Perreault, L. (2001), Analyse bayésienne de la stationnarité des apports naturels annuels agrégés des complexes Côte-Nord et La Grande, *Tech. Rep. 99-05-45-30*, Hydro-Québec Production.

Perreault, L., J. Bernier, B. Bobée, and E. Parent (2000), Bayesian change-point analysis in hydrometeorological time series. part 1. the normal model revisited, *Journal of Hydrology*, *235*(3-4), 221–241.

Perreault, L., R. Garçon, and J. Gaudet (2007), Modelling hydrologic time series using regime switching models and measures of atmospheric circulation (in french), *La Houille Blanche*, (6), 111–123.

R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rémillard, B. (2013), *Statistical Methods for Financial Engineering*, Chapman and Hall/CRC Financial Mathematics Series, Taylor & Francis.

Savard, M., C. Bégin, M. Parent, A. Smirnoff, and J. Marion (2004), The environmental impact of smelter So 2 emissions–a time and space perspective recorded by carbon isotope ratios in tree-ring cellulose, *Journal of Environmental Quality*, *33*, 13–26.

Schweingruber, F., H. Fritts, O. Bräker, L. Drew, and E. Schär (1978), The x-ray technique as applied to dendroclimatology, *Tree-Ring Bulletin*.

Schweingruber, F. H., et al. (1996), *tree-rings and Environment: Dendroecology.*, Paul Haupt AG Bern.

Smith, L. P., and C. W. Stockton (1981), Reconstructed stream flow for the Salt and Verde rivers from tree-ring data, *JAWRA Journal of the American Water Resources Association*, *17*(6), 939–947.

Stockton, C. W., and H. C. Fritts (1973), Long-term reconstruction of water level changes for lake Athabasca by analysis of tree-rings, *JAWRA Journal of the American Water Resources Association*, *9*(5), 1006–1027.

Sveinsson, O. G., Lall, U., Fortin, V., Perrault, L., Gaudet, J., Zebiak, S., and Kushnir, Y. (2008a). Forecasting spring reservoir inflows in Churchill Falls basin in Québec, canada. *Journal of Hydrologic Engineering*, 13(6):426–437.

Sveinsson, O. G., Lall, U., Gaudet, J., Kushnir, Y., Zebiak, S., and Fortin, V. (2008b). Analysis of climatic states and atmospheric circulation patterns that influence Québec spring streamflows. *Journal of Hydrologic Engineering*, 13(6):411–425.

Tenenhaus, M. (1998), *La régression PLS: théorie et pratique*, Editions technip.

Thyer, M., Frost, A. J., and Kuczera, G. (2006). Parameter estimation and model identification for stochastic models of annual hydrological data: Is the observed record long enough? *Journal of Hydrology*, 330(1):313–328.

Thyer, M., and G. Kuczera (2000), Modeling long-term persistence in hydroclimatic time series using a hidden state Markov model, *Water Resources Research*, *36*(11), 3301–3310.

Thyer, M., and G. Kuczera (2003), A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach, *Journal of Hydrology*, *275*(1-2), 12–26.

Wilhelm, B., J. A. B. Canovas, J. P. C. Aznar, L. Kämpf, T. Swierczynski, M. Stoffel, E. Støren, and W. Toonen (2018), Recent advances in paleoflood hydrology: From new archives to data compilation and analysis, *Water Security*, *3*, 1–8.

Woodhouse, C., D. Meko, D. Griffin, and C. Castro (2013), tree-rings and multiseason drought variability in the lower Rio grande basin, USA, *Water Resources Research*, *49*(2), 844–850.

Woodhouse, C. A., and J. J. Lukas (2006), Multi-century tree-ring reconstructions of Colorado streamflow for water resource planning, *Climatic Change*, *78*(2-4), 293–315.

A: Estimation of the HMM models

The EM algorithm for estimating parameters consists of two steps, expectation and maximization:

1. (E-Step) Compute the conditional probabilities.

$$\lambda_t(i) = P(\tau_t = i|Y_1, \ldots, Y_n) \quad \text{and} \quad \Lambda_t(i,j) = P(\tau_t = i, \tau_{t+1} = j|Y_1, \ldots, Y_n), \tag{A.1}$$

   for all $1 \le t \le n$ and $i, j \in \{1, \ldots, m\}$.

2. (M-Step) Estimate the new parameters.

First, a rough estimate of the parameters must be provided. Then, the two-step procedure is repeated until a stopping criterion is satisfied. The E-Step is described next for any densities, while the M-Step is stated only for Gaussian densities.

A.1 Conditional distribution of the regimes (E-Step)

First, define, for all $i \in \{1, \ldots, m\}$,

$$\begin{aligned}
\eta_0(i) &= 1/m, && \text{(A.2)}\\
\eta_t(i) &= \frac{f_{\boldsymbol{\beta}_i}(Y_t) \sum_{j=1}^m \eta_{t-1}(j) Q_{ji}}{\sum_{k=1}^m \sum_{j=1}^m f_{\boldsymbol{\beta}_k}(Y_t) \eta_{t-1}(j) Q_{jk}}, \quad t = 1, \ldots, n, && \text{(A.3)}\\
\bar{\eta}_n(i) &= 1/m, && \text{(A.4)}\\
\bar{\eta}_t(i) &= \frac{\sum_{k=1}^m \bar{\eta}_{t+1}(k) Q_{ik} f_{\boldsymbol{\beta}_k}(Y_{t+1})}{\sum_{\alpha=1}^m \sum_{k=1}^m \bar{\eta}_{t+1}(k) Q_{\alpha k} f_{\boldsymbol{\beta}_k}(Y_{t+1})}, \quad t = 1, \ldots, n-1. && \text{(A.5)}
\end{aligned}$$

Then, for all $i, j \in \{1, \ldots, m\}$, one can verify that

$$\begin{aligned}
\lambda_t(i) &= \frac{\eta_t(i)\bar{\eta}_t(i)}{\sum_{\alpha=1}^m \eta_t(\alpha)\bar{\eta}_t(\alpha)}, \quad t = 1, \ldots, n, && \text{(A.6)}\\
\Lambda_t(i,j) &= \frac{Q_{ij}\eta_t(i)\bar{\eta}_{t+1}(j) f_{\boldsymbol{\beta}_j}(Y_{t+1})}{\sum_{\alpha=1}^m \sum_{k=1}^m Q_{\alpha k}\eta_t(\alpha)\bar{\eta}_{t+1}(k) f_{\boldsymbol{\beta}_k}(Y_{t+1})}, \quad t = 1, \ldots, n-1, && \text{(A.7)}\\
\Lambda_n(i,j) &= \lambda_n(i) Q_{ij}. && \text{(A.8)}
\end{aligned}$$

We can now verify that Equations (A.6) and (A.7) are consistent. Indeed, for all $1 \le t \le n-1$,

$$\sum_{j=1}^m \Lambda_t(i,j) = \frac{\eta_t(i)\left(\sum_{j=1}^m Q_{ij}\bar{\eta}_{t+1}(j) f_{\boldsymbol{\beta}_j}(Y_{t+1})\right)}{\sum_{\alpha=1}^m \eta_t(\alpha)\left(\sum_{k=1}^m Q_{\alpha k}\bar{\eta}_{t+1}(k) f_{\boldsymbol{\beta}_k}(Y_{t+1})\right)} = \lambda_t(i), \tag{A.9}$$

using the definition of $\bar{\eta}_t$. Also, $\sum_{j=1}^m \Lambda_n(i,j) = \sum_{j=1}^m \lambda_n(i) Q_{ij} = \lambda_n(i)$. Similarly, for all $1 \le t \le n-1$, $\displaystyle\sum_{i=1}^l \Lambda_t(i,j) = \lambda_{t+1}(j)$.

A.2 Estimation for Gaussian regime-switching models (M-Step)

For the estimation procedure, we assume the densities $f_{\boldsymbol{\beta}_1}, \ldots, f_{\boldsymbol{\beta}_m}$ are Gaussian with means $(\mu_i)_{i=1}^m$ and covariance matrices $(A_i)_{i=1}^m$. The M-step consists of updating the parameters $(\nu_i)_{i=1}^m$, $(\mu_i)_{i=1}^m$, $(A_i)_{i=1}^m$ and Q according to

$$\begin{aligned}
\nu_i' &= \sum_{t=1}^n \lambda_t(i)/n, && \text{(A.10)}\\
\mu_i' &= \sum_{t=1}^n x_t w_t(i), && \text{(A.11)}\\
A_i' &= \sum_{t=1}^n (x_t - \mu_i')(x_t - \mu_i')^\top w_t(i), && \text{(A.12)}\\
Q_{ij}' &= \sum_{t=1}^n \Lambda_t(i,j) \Big/ \sum_{t=1}^n \lambda_t(i) = \frac{1}{n}\sum_{t=1}^n \Lambda_t(i,j) \Big/ \nu_i', && \text{(A.13)}
\end{aligned}$$

for all $i, j \in \{1, \ldots, m\}$ and where $w_t(i) = \lambda_t(i) \Big/ \sum_{m=1}^{n} \lambda_m(i)$. Note that ν' is not the stationary distribution for Q' since for any $j \in \{1, \ldots, m\}$,

$$\sum_{i=1}^{m} \nu'_i Q'_{ij} = \frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{m} \Lambda_t(i, j) = \frac{1}{n} \sum_{t=2}^{n+1} \lambda_t(j) = \nu'_j + \frac{\lambda_{n+1}(j) - \lambda_1(j)}{n} \neq \nu'_j. \tag{A.14}$$

However, $\max_{1 \leq j \leq l} \left| \sum_{i=1}^{m} \nu'_i Q'_{ij} - \nu'_j \right| \leq 1/n$. Hence, when n is large, ν' is close to the stationary distribution of Q'. In practice, we estimate the stationary distribution from Q', rather than ν', for consistency.

B: Goodness-of-fit test for the HMM

Suppose that $Y_1, \ldots, Y_n$ is a size n sample of a unidimensional vector drawn from a continuous distribution $\mathbf{P}$ belonging to a parametric family of univariate regime-switching models with m regimes. Formally, the hypothesis to be tested is $\mathcal{H}_0 : \mathbf{P} \in \mathcal{P} = \{\mathbf{P}_\theta; \theta \in \mathcal{O}\}$ $\quad vs \quad$ $\mathcal{H}_1 : \mathbf{P} \notin \mathcal{P}$. Under the null hypothesis, it follows that $V_t = F_{t|1:t-1}(Y_t, \boldsymbol{\theta})$ is independent and uniformly distributed over $(0, 1)$, where $F_{t|1:t-1}(\cdot, \boldsymbol{\theta})$ is the conditional distribution function for the true parameters $\boldsymbol{\theta} \in \mathcal{O}$, as defined by Equation (5).

B.1 Test statistics

Following *Nasri and Remillard* [2019a], define the empirical process

$$D_n(u) = \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\left(V_{n,t} \leq u\right), \quad u \in [0, 1], \tag{B.1}$$

where $V_{n,t} = F_{t|1:t-1}(Y_t, \boldsymbol{\theta}_n)$ and $\boldsymbol{\theta}_n$ is the consistent estimator of $\boldsymbol{\theta}$. Following, *Genest et al.* [2009], to test $\mathcal{H}_0$ against $\mathcal{H}_1$, it is suggested to use the Cramér-von Mises type statistic because it appears to be much more powerful and easier to compute than the Kolmogorov-Smirnov type statistic. The Cramér-von Mises type statistic is given by $S_n = B_n\left(V_{n,1}, \ldots, V_{n,n}\right) = n \int_0^1 \{D_n(u) - u\}^2 \, du$.

B.2 Parametric bootstrap

If a goodness-of-fit test is based on a statistic S_n of the observations $Y_1, \ldots, Y_n$ with distribution $\mathbf{P}_\theta$ for some unknown parameter θ estimated by θ_n, the parametric bootstrap approach consists of generating a large number B of sequences $Y_1^{(k)}, \ldots, Y_n^{(k)}$ with distribution $\mathbf{P}_{\theta_n}$, $k = 1, \ldots, B$, evaluatingthe goodness-of-fit statistic $S_n^{(k)}$ each time, and approximating the P-value as the percentage of values $S_n^{(k)}$ that are greater than S_n, assuming that the null hypothesis is rejected for large values of S_n. Hence, to perform the goodness-of-fit test, we use the following algorithm:

Algorithm 1 *For a given number of regimes m, obtain estimator* $\boldsymbol{\theta}_n$ *of* $\boldsymbol{\theta}$ *using the EM algorithm. Then, compute the statistic* $S_n = B_n(V_{n,1}, \ldots, V_{n,n})$ *using the pseudo-observations* $V_{n,t} = F_{t|1:t-1}(Y_t, \boldsymbol{\theta}_n)$, $t \in \{1, \ldots, n\}$. *Next, for* $k = 1, \ldots, B$, *with sufficiently large B, repeat the following steps:*

- *Generate a random sample* $Y_1^*, \ldots, Y_n^*$ *from a Gaussian HMM with parameter* $\boldsymbol{\theta}_n$.
- *Obtain the estimator* $\boldsymbol{\theta}_n^*$ *from* $Y_1^*, \ldots, Y_n^*$.
- *Compute the pseudo-observations* $V_{n,t}^* = F_{t|1:t-1}(Y_t^*, \boldsymbol{\theta}_n^*)$, $t \in \{1, \ldots, n\}$ *and calculate* $S_n^{(k)} = B_n\left(V_{n,1}^*, \ldots, V_{n,n}^*\right)$.

Then, an approximate P-value for the test based on the Cramér-von Mises statistic S_n *is given by*

$$\frac{1}{B}\sum_{k=1}^{B} \mathbb{1}\left(S_n^{(k)} > S_n\right). \tag{B.2}$$

The goodness-of-fit test methodology produces P-values from a Cramér-von Mises type statistic for a given number of regimes m. As suggested in *Rémillard* [2013], it makes sense to choose the optimal number of regimes, m^*, as the first m for which the P-value is larger than 5%.

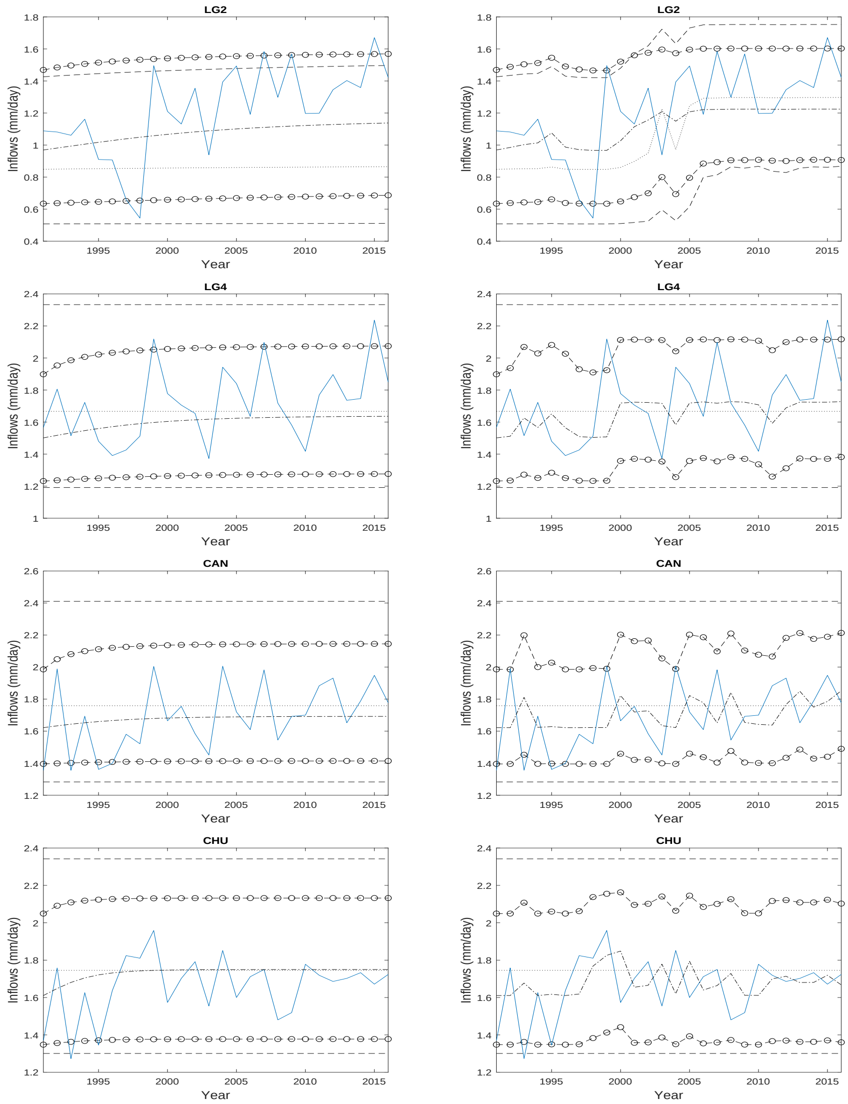
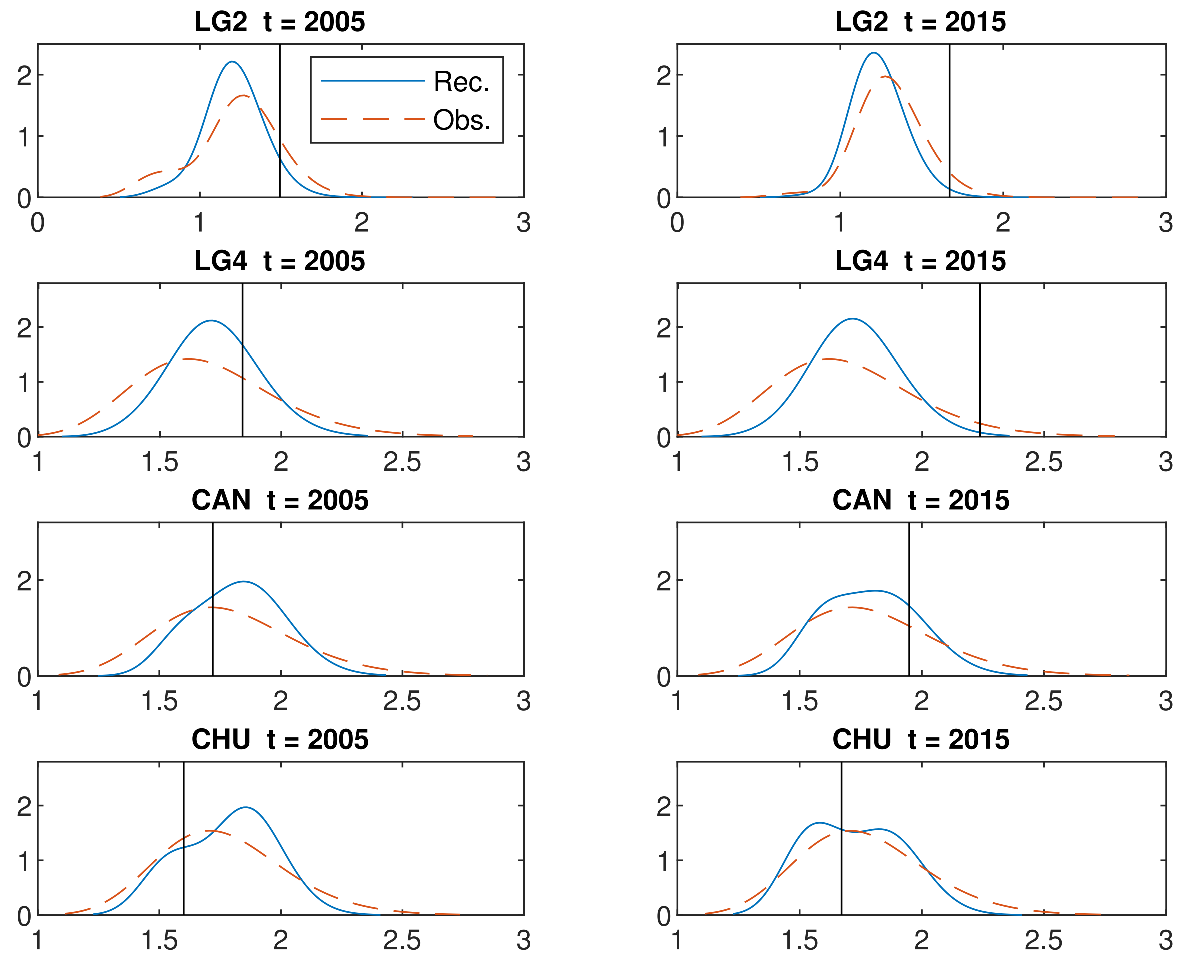
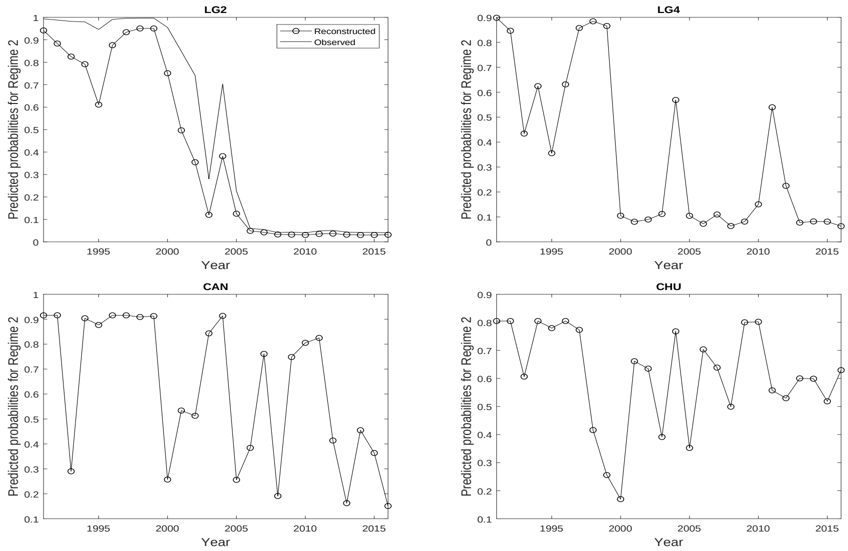


Figure 8. The graphs in the left column display the long-term predictions from 1991 to 2016, while the right column displays 1-year-ahead forecasts. The blue lines show the observed data, the dash-dot lines represent the predicted median from the reconstruction models, and the dotted lines represent the predicted median from the observed data models. Finally, the dashed lines are the 95% confidence intervals estimated by using the models from the observed data and the dashed lines with circles are the 95% confidence intervals estimated by using the models from the reconstructed data.



486 **Figure 9.** Plotted predictive densities for stations LG2, LG4, CAN, and CHU for 2005 and 2015, using the reconstructed
487 and observed data. The vertical lines represent the observed values for the year.



550 **Figure 10.** One-year-ahead predicted probabilities for the dry regime (Regime 2), estimated from the reconstructed data.
551 Predicted probabilities estimated from observed data are shown only for LG2.

View publication stats