Considering Uncertainty of Historical Ice Jam Flood Records in a Bayesian Frequency Analysis for the Peace-Athabasca Delta

Jared D. Smith[1], Jonathan R Lamontagne[2], and Martin Jasek[3]

[1]Department of Engineering Systems and Environment, University of Virginia
[2]Department of Civil and Environmental Engineering, Tufts University
[3]BC Hydro

June 25, 2023

Abstract

The Peace-Athabasca Delta in Alberta, Canada has numerous perched basins that are primarily recharged after large ice jams cause floods (an ecological benefit). Previous studies have estimated that such large floods are likely to decrease in frequency under various climate projections. However, there is a sizeable uncertainty range in these predicted flood probabilities, in part due to the short 60-year systematic record that contained few large ice jam floods. An additional 50 years of historical data are available from various sources, with expert-interpreted flood categories; however, these categorizations are uncertain in magnitude and occurrence. We developed a Bayesian framework that considers magnitude and occurrence uncertainties within a logistic regression model that predicts the annual probability of a large flood. The Bayesian regression estimates the joint distribution of parameters describing the effects of climatic factors and parameters that describe the probability that historical flood magnitudes were recorded as large (or not) when a truly large (or not) flood occurred. We compare four models for hindcasting and projecting large ice jam flood probabilities in future climates. The models consider: 1) historical data uncertainty, 2) no historical data uncertainty, 3) only the systematic record, and 4) the systematic record with a different model structure. Neglecting historical data uncertainty provides inaccurate estimates, while using only the systematic record provides wider prediction intervals than considering the full record with uncertain historical data. Thus, we demonstrate that including uncertain historical information can effectively extend the record length and improve flood frequency analyses.

Considering Uncertainty of Historical Ice Jam Flood Records in a Bayesian Frequency Analysis for the Peace-Athabasca Delta

Jared D. Smith[1*], Jonathan R. Lamontagne[2], Martin Jasek[3]

[1]University of Virginia Department of Engineering Systems and Environment, Charlottesville, VA, USA
[2]Tufts University Department of Civil and Environmental Engineering, Medford, MA, USA
[3]BC Hydro, Burnaby, BC, Canada

Key Points:

- We use a Bayesian logistic regression framework to estimate ice jam flood frequency while considering uncertainty in the historical record.
- We compare annual flood probabilities from a model trained with a systematic record to a model trained with additional historical data.
- Prediction intervals for projected climates are narrower when uncertain historical data are used.

*Currently employed at: U.S. Geological Survey, Reston, VA, USA

Corresponding author: Jared Smith, `jared.d.smith485@gmail.com`

Abstract

The Peace-Athabasca Delta in Alberta, Canada has numerous perched basins that are primarily recharged after large ice jams cause floods (an ecological benefit). Previous studies have estimated that such large floods are likely to decrease in frequency under various climate projections. However, there is a sizeable uncertainty range in these predicted flood probabilities, in part due to the short 60-year systematic record that contained few large ice jam floods. An additional 50 years of historical data are available from various sources, with expert-interpreted flood categories; however, these categorizations are uncertain in magnitude and occurrence. We developed a Bayesian framework that considers magnitude and occurrence uncertainties within a logistic regression model that predicts the annual probability of a large flood. The Bayesian regression estimates the joint distribution of parameters describing the effects of climatic factors and parameters that describe the probability that historical flood magnitudes were recorded as large (or not) when a truly large (or not) flood occurred. We compare four models for hindcasting and projecting large ice jam flood probabilities in future climates. The models consider: 1) historical data uncertainty, 2) no historical data uncertainty, 3) only the systematic record, and 4) the systematic record with a different model structure. Neglecting historical data uncertainty provides inaccurate estimates, while using only the systematic record provides wider prediction intervals than considering the full record with uncertain historical data. Thus, we demonstrate that including uncertain historical information can effectively extend the record length and improve flood frequency analyses.

1 Introduction

The Peace-Athabasca Delta (PAD) in Alberta, Canada has numerous perched basins (small lakes) that are primarily recharged with water and nutrients after large ice jam floods cause long duration flooding of the vast delta area (Timoney, 2013). Periodic flooding of the PAD has ecological benefits (Timoney, 2013) and allows for better navigation for the First Nations communities to access resources and utilize the land. Previous studies of the PAD region have estimated that large floods are likely to decrease in frequency under various climate projections (Lamontagne et al., 2021; Jasek et al., 2021; Das et al., 2020; Beltaos et al., 2008). The predicted future flood probabilities have a wide uncertainty range, partly due to a short 60-year systematic record with only seven large ice jam floods. An additional 50 years of historical information are available as expert-interpreted flood magnitudes based on traditional knowledge, historical written records, and proxy data, as summarized by Timoney (2009), but these categorizations are uncertain with respect to flood magnitude and occurrence. For example, a flood labeled as moderate could have been large, and a year labeled as having no flood could have actually contained a flood that was not recorded.

Wolfe et al. (2020) provide an excellent discussion of the uncertainties that could arise when assigning flood magnitudes based on available historical information. In brief, the uncertainties may be characterized by observer bias and spatial variability. Observer bias includes gaps in years with recorded information, and differences in descriptions of flood events from one location to another. Spatial variability includes inter-annual differences in the locations that were flooded, and intra-annual differences in proximity of the locations to that year's ice jam (Prowse & Conly, 2002). Timoney (2009) and Peterson (1995) used aggregate information of flooding events across the PAD to inform their assigned annual flood magnitudes, so gaps in records and varying spatial information leads to uncertainty in how much inundation of the PAD occurred. As a result, Prowse and Conly (2002) and Wolfe et al. (2020) note that the recorded flood magnitudes and occurrences could be incorrect. This is not the fault of the experts who interpreted the historical flood record, and is rather a feature of the available data. However, it is likely that the historical record contains some useful information, even though it is uncertain.

In this paper, we provide a method to use uncertain historical information with certain systematic records in an ice jam flood frequency analysis.

We present a Bayesian logistic regression methodology to account for the uncertainty in historical flood magnitude and occurrence data, and analyze the value of including uncertain historical information into predictions of future large ice jam flood probabilities. Our key research questions is: Compared to using only the systematic record, how does considering magnitude and occurrence uncertainty in historical flood data impact the estimated hindcasted and projected probabilities of a large ice jam flood?

There is a long history of using flood frequency analysis with historical records and paleo-flood information to obtain more precise estimates of flood magnitudes and their probability than can be gathered from relatively short systematic records (Kjeldsen et al., 2014; Payrastre et al., 2011; Benito et al., 2004; Stedinger & Cohn, 1986; Hosking & Wallis, 1986; Condie & Lee, 1982). The focus of the literature has primarily been on estimating flood magnitude quantiles, typically via estimation of the parameters of an extreme value probability distribution. Including historical information is generally done by augmenting the likelihood function to consider the historical flood magnitudes are censored information, or the occurrence of historical floods above a perception threshold follows a binomial distribution (e.g., Stedinger & Cohn, 1986). An additional critical need is the interpretation of historical information by local experts, but even those interpretations can be uncertain in the magnitude of historical floods (e.g., perception threshold values in Parkes & Demeritt, 2016). Recent studies have considered that the historical information may be uncertain, and use Bayesian frameworks to cleanly handle the data uncertainty (Reis & Stedinger, 2005; Salinas et al., 2016; Parkes & Demeritt, 2016). For example, Salinas et al. (2016) use a fuzzy membership approach to integrate imprecise descriptions of historical flood events within a Bayesian estimation of flood magnitude quantiles. Fuzzy membership allows for each historical data point to have non-zero probability of actually having been any of several flood magnitudes. In this study, our Bayesian framework also allows for historical data to have a different categorical flood magnitude than recorded, or to not have occurred. However, we seek to estimate the annual probability of a large flood, instead of estimating long-term quantile flood magnitudes. We are not aware of literature that addresses this application, nor the application to ice jam flood frequency analysis with uncertain historical data.

1.1 Background on Flood Generation Processes in the PAD

The PAD lies at the confluence of the Peace and Athabasca Rivers in northern Alberta, Canada (see Fig. 1). A series of typically northward-flowing channels connect the PAD to the Peace River, though the direction of flow can reverse during high water events on the Peace River. Open water floods on the Peace River are typically not capable of generating water levels that can recharge many of the PAD's highest elevation perched basins, called restricted basins. Restricted basins are instead recharged when temporary ice jams on the Peace, Athabasca, or the Slave Rivers form during the spring freshet and result in resistance to flow and resulting higher water levels that can be tens of kilometers long. These ice jams cause flow reversals (southward flow) within the channels that connect the Peace River to the PAD and can generate widespread flooding. Flooding of a restricted basin indicates that a large flood likely occurred in a given year, but there is still uncertainty in how widespread the flooding was in each year. Jasek (2019c) details a taxonomy that describes the conditions under which large, moderate and small floods are likely to occur in the PAD from Peace River ice jams (further discussed in Lamontagne et al., 2021). The large floods of interest tend to occur when winter snowpack is high in upstream tributary basins (high discharge potential) including the Smoky River, followed by a spring with rapid and sustained warming. These conditions along with high ice resistance in the PAD increase the likelihood of a dynamic, mechanical breakup of river ice that can form an ice jam along the Peace River whose impounded water could eventually flood the PAD. Lamontagne et al. (2021) explored the predictive power of a variety of climatic and riverine proxy data that relate to these physical flooding processes. In particular, winter snowpack at Beverlodge and Grande Prairie in the Smoky Basin are used as proxies of discharge potential, and winter degree day freezing and thawing at Forts Vermilion, Chipewyan, and Smith are used as proxies for ice resistance and speed of the warming respectively. We also rely on these proxy data and explore their use in prediction with the historical and systematic records, as explained in Section 2.

Since 1972, the Peace River has been partially regulated by the Bennett and Peace Canyon Dams that are located about 1200 km upstream (Fig. 1). While flow regulation does affect streamflow and hence stage at the time the Peace River freezes up in the fall and winter, Lamontagne et al. (2021) found that these freeze-up elevations have little to no predictive power for large ice jam flood occurrence in the systematic record after accounting for the effects of climatic factors.

2 Exploratory Data Analysis

We gathered candidate explanatory variables from several meteorological stations that are located across the Peace River Basin (Fig. 1). The two primary variables that were available for 1915-2020 were temperature and precipitation, from which we derived cumulative degree-days freezing (DDF), degree-days thaw (DDT) and snowpack variables, as in Lamontagne et al. (2021) and Jasek et al. (2021). We focus on climatic conditions because Lamontagne et al. (2021) found that they have the best predictive skill for large ice jam floods in the systematic record (1962-2020), and historical (1915-1962) Peace River streamflow and freeze-up elevation data were not consistently available before the 1960s.

The locations of the derived variables aim to capture processes that lead to large ice jam flood generation in the PAD (Jasek, 2019a, 2019b; Jasek et al., 2021). Grande Prairie, Beaverlodge, and Fort Vermillion represent conditions upstream of the PAD, Fort Chipewyan represents conditions within the PAD, and Fort Smith represents conditions downstream of the PAD. We computed winter DDF at Fort Vermillion, Fort Chipewyan, and Fort Smith stations, and accumulated winter snowpack at Grande Prairie and Beaverlodge stations during sustained DDF (i.e, snowpack was reset to 0 if DDF at Grande Prairie or Beaverlodge became negative, which would indicate melting conditions). As in Lamontagne et al. (2021), we created a complete precipitation record by using Beaverlodge data to fill in gaps in Grande Prairie data. One additional explanatory variable used in Lamontagne et al. (2021), called "melt test," could be computed back to 1915 and describes how rapidly thaw occurs in the medium elevations of the watershed in spring. Lamontagne et al. (2021) computed the melt test as the number of days to go from 40 DDT to 150 DDT at Grande Prairie, where smaller values indicate more rapid warming. 40 DDT is sufficient to initiate a dynamic break-up of ice on the Smoky River, and the remaining DDT sustains the high freshet flows on the Peace River to ultimately flood the PAD (Jasek, 2019c; Lamontagne et al., 2021).

Because these climate data are correlated, we applied a principal component analysis (PCA) to arrive at a reduced set of uncorrelated PCs for exploratory visualization and for possible use in regression models. Before using PCA, all variables were normalized to have a mean of 0 and standard deviation of 1. The first 2 PCs explain about 96% of the variance from the DDF and snowpack variables (melt test was not included in the best model). The first PC explains about 81% of the variance and it represents mostly winter DDF (positive values are smaller DDF and more snowpack). The second PC explains about 15% of the variance and it represents mostly snowpack (positive values are more snowpack and larger DDF).

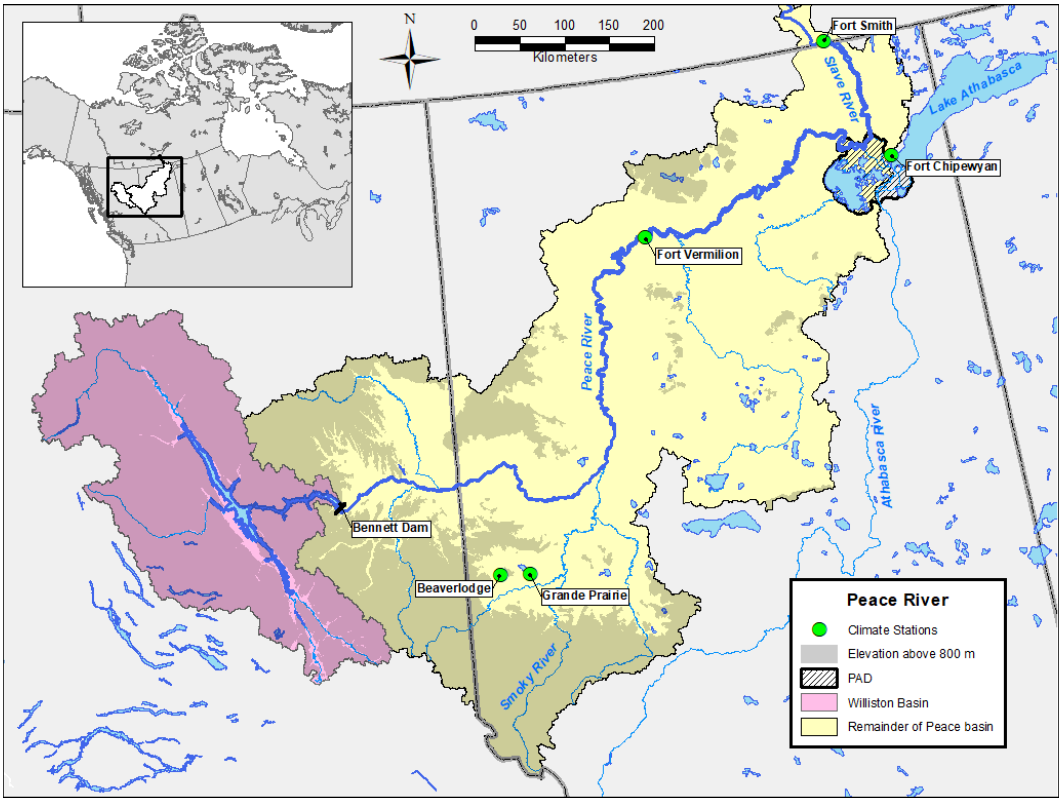

Figure 1. The Peace River watershed with locations of the Peace Athabasca Delta (PAD) and meteorological stations used for this analysis. The Williston Basin is the regulated portion of the basin. Modified from (Lamontagne et al., 2021).

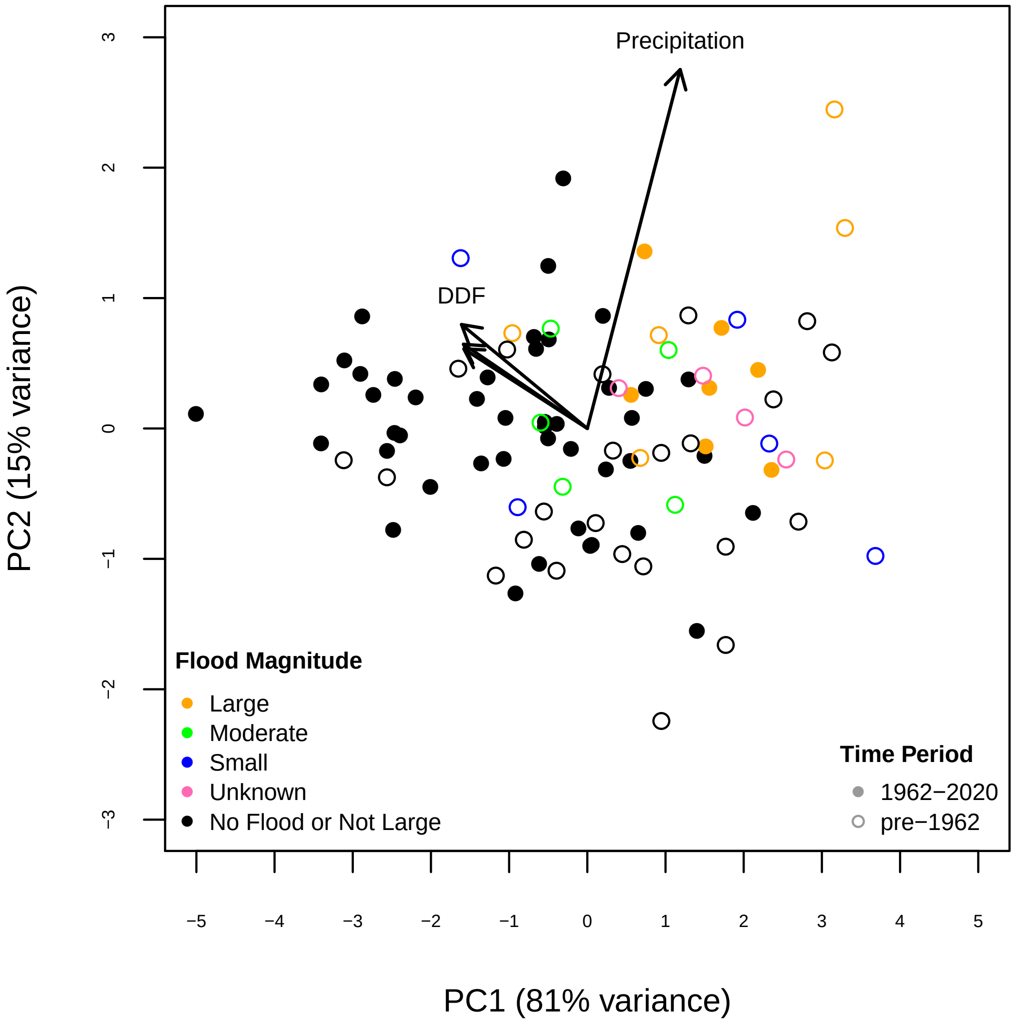

Figure 2. Flood magnitudes shown on each of the principal component (PC) axes that were used as predictor variables in the logistic regression. Only years from 1962-2020 have known magnitude and occurrence. The directions of positive snowpack and degree-days freezing are shown for reference.

Figure 2 plots the annual ice jam flood data for 1915-2020 on these PC1 and PC2 axes. Large ice jam floods in the systematic record tend to occur in colder winters with more snowpack. For the historical record, there is overlap of recorded small, moderate, and unknown magnitude floods in the space occupied by the large floods in the systematic record. There are also some years of the historical record with no recorded floods that occur in more extreme climatic conditions than large floods in the systematic record. These findings motivate testing a regression model that allows for historical data to have uncertain flood magnitudes and occurrences. A summary of the data used in this analysis is provided in Table 1.

3 Methods

Considering uncertainty in a predicted variable (e.g., large ice jam flood occurrence) is commonly used in the epidemiology literature when medical diagnostic tests are ap-

Table 1. Summary of available flood data and climatic factors. GP: Grande Prairie, BL: Beaverlodge, FC: Fort Chipewyan, FS: Fort Smith, FV: Fort Vermillion. Fort Smith does not have data for three years, and four dam filling years are not used for modelling.

Raw Variables	Interpreted or Derived Variables	Coverage
Historical Floods	6 Large, 5 Moderate, 5 Small, 4 Unknown, 23 No Flood	1915 - 1962
Systematic Floods	7 Large, 45 Not Large	1962 - 2020
GP, BL Precipitation	Snowpack	1915 - 2020
FC, FS, FV Temperature	Degree-days freezing, melt test	1915 - 2020

plied in small sample sizes and inferences must be made about the reliability of the test outcomes (e.g., McInturff et al., 2004). To consider uncertainty in the historical flood record, we adapted a Bayesian framework presented in (McInturff et al., 2004) that allows us to consider flood magnitude and occurrence uncertainties within a logistic regression model that predicts the annual probability of a large ice jam flood as a function of climatic variables and model parameters. Section 3.1 presents the standard logistic regression model and our adapted format that allows for considering flood magnitude and occurrence uncertainties, and Section 3.2 presents the Bayesian framework that we used to estimate the joint posterior distribution of model parameters and the distribution of each annual large ice jam flood probability. These estimated parameters are used to estimate annual probabilities in projected climate scenarios to year 2100, as described in Section 3.3.

3.1 Logistic Regression Model

The standard logistic regression model describes the probability of a Bernoulli random variable, Z

$$Z_i \sim \textsc{Bernoulli}(p_i = \Pr[Z_i = 1|\mathbf{X} = \mathbf{x_i}]) \tag{1}$$

where p_i is the probability that a large ice jam flood occurred, $Z_i = 1$, given conditions, $\mathbf{X}$, in the i^{th} year. In other words, we allow the annual probability of a large ice jam flood to change from year to year based on climatic conditions (for more details on logistic regression applied to ice jam flood frequency analysis, see Lamontagne et al., 2021). To compute values of the annual probability on [0, 1], we employ the logistic function in equation 2

$$p_i = f(\mathbf{x_i}, \boldsymbol{\beta}) = \frac{e^{\mathbf{x_i}\boldsymbol{\beta}}}{1 + e^{\mathbf{x_i}\boldsymbol{\beta}}} \tag{2}$$

where $\boldsymbol{\beta}$ are the true model coefficients to be estimated from the data. This equation may be linearized as shown in equation 3

$$\log\left(\frac{p_i}{1 - p_i}\right) = \mathbf{x_i}\boldsymbol{\beta} = \beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 ... x_{i,n}\beta_n \tag{3}$$

where the left hand side is the logistic function, β_0 is a constant, and $x_{i,a}$ and β_a are the a^{th} variables and model coefficients, respectively. The model coefficients and their statistical significance may be estimated using standard maximum likelihood approaches. As in Lamontagne et al. (2021), we maximize a penalized likelihood proposed by Firth

(1993) that reduces bias in the estimated $\boldsymbol{\beta}$ values when sample sizes are small, and penalizes less with increasing sample size. The correction is equivalent to Jeffrey's uninformative prior (Gelman, 2009), so maximum likelihood estimation with this likelihood function provides a Bayesian posterior mode, and the estimator is known as the maximum *a posteriori* (MAP) estimator. We used the R package logistf for Firth's maximum likelihood estimation (Heinze et al., 2020).

When the outcomes (e.g., large ice jam floods) are uncertain, we can adapt this regression framework with an additional random variable for the recorded outcome, Y

$$Y_i \sim \textsc{Bernoulli}(q_i = \Pr[Y_i = 1|\mathbf{X} = \mathbf{x_i}]) \tag{4}$$

where q_i is the probability that a large ice jam flood was recorded given conditions, X, in the i^{th} year. With this modification, additional model parameters are needed to estimate probability q_i to account for the possibility that a recorded large flood may not have actually been large (a false positive), and that a year without a recorded large flood actually may have had a large flood (a false negative). The parameters used to do this are called sensitivity, η, and specificity, θ, as presented in equations 5 and 6, respectively

$$\eta = \Pr[Y_i = 1|Z_i = 1] \tag{5}$$

$$\theta = \Pr[Y_i = 0|Z_i = 0]. \tag{6}$$

Each of these conditional probabilities describe the probability that the historical record is correctly reporting when a large ice jam flood did or did not occur. From the law of total probability, the probability q_i for the case of two flood categories (large and not large) is presented in equation 7

$$\begin{aligned} q_i &= \Pr[Y_i = 1|\mathbf{X} = \mathbf{x_i}] \\ &= \Pr[Y_i = 1|Z_i = 1]\Pr[Z_i = 1|\mathbf{x_i}] + \Pr[Y_i = 1|Z_i = 0]\Pr[Z_i = 0|\mathbf{x_i}] \\ &= \eta * p_i + (1-\theta) * (1-p_i) \end{aligned} \tag{7}$$

where q_i reduces to a function of the true (unknown) probability p_i of a large ice jam flood that we are interested in estimating. When there is full confidence in the recorded data, $\eta = 1$ (always record a large flood when one occurs) and $\theta = 1$ (always record that no large flood occurred when one does not occur), and $q_i = p_i$. Therefore, estimating the values of η and θ equates to estimating the fidelity of the historical record.

We could stop here and estimate regression coefficients, sensitivity, and specificity while assuming that all years recorded as not having a flood in the historical record had the same data generating process (i.e., are represented by one value of θ). However, our historical record includes other flood magnitudes that are not large. So, our implementation recognizes that the experts who labeled the flood data had additional information that provided those flood categories.

To account for four "not large" categories, C, we decomposed θ into components labeled as θ_M, θ_S, θ_U, and θ_N for moderate, small, unknown, and no flood, respectively. The decomposition is provided in equation 8

$$\begin{aligned}\theta_i &= \Pr[Y_i = 0|Z_i = 0] \\ &= \Pr[Y_i = 0|Z_i = 0, C_i = M]\Pr[C_i = M] \\ &\quad + \Pr[Y_i = 0|Z_i = 0, C_i = S]\Pr[C_i = S] \\ &\quad + \Pr[Y_i = 0|Z_i = 0, C_i = U]\Pr[C_i = U] \\ &\quad + \Pr[Y_i = 0|Z_i = 0, C_i = N]\Pr[C_i = N] \\ &= \theta_M \Pr[C_i = M] + \theta_S \Pr[C_i = S] + \theta_U \Pr[C_i = U] + \theta_N \Pr[C_i = N]\end{aligned} \tag{8}$$

where θ_i would be used in place of θ in equation 7 because each observation could have a different specificity according to its recorded flood category. Each of the θ_c values can be read as the probability that a large flood was not recorded given that a large flood really did not occur and the recorded flood category was the value of C_i (moderate, small, unknown, or no flood). While we could have used a multinomial model that estimates the probability of each of these flood categories, $\Pr[C_i = c]$, we decided to estimate their values based on the recorded data. In doing so, we assume that the experts who assigned the flood categories had full confidence that the assigned categories were not actually large floods, given the available historical information. When the recorded $Y_i = 0$, then the probability $\Pr[C_i = c]$ is 1 for the recorded category and 0 for all other categories. When the recorded $Y_i = 1$, then each $\Pr[C_i = c]$ is estimated as the proportion of each flood category when $Y_i = 0$ (i.e., θ is estimated as a weighted average of θ_M, θ_S, θ_U, and θ_N).

With the formulation described above, we can estimate the true probabilities that we are interested in, p_i, via estimating the probability of recording a large flood, q_i. The likelihood equation to maximize is the standard likelihood for a Bernoulli distributed random variable shown in equation 9

$$\mathscr{L}(\boldsymbol{\beta}, \eta, \boldsymbol{\theta_C}|\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{C}) = \prod_{i=1}^{N} [\eta_i p_i + (1 - \theta_i)(1 - p_i)]^{y_i} [(1 - \eta_i)p_i + \theta_i(1 - p_i)]^{1 - y_i} \tag{9}$$

where the first bracketed term on the right applies when a large flood was recorded, and the second bracketed term applies when a large flood was not recorded. For our study, $\eta_i = 1$ and $\theta_i = 1$ for the systematic record, η_i is a single parameter to be estimated for the historical record, and we use equation 8 to compute θ_i for each observation based on the vector of estimated parameters for each flood category, $\boldsymbol{\theta_C}$.

In summary, the parameters to be estimated in the Bayesian logistic regression are the $\boldsymbol{\beta}$ coefficients for the regression model variables, the four θ_C specificity parameters for non-large flood categories in the historical record, and the η specificity parameter for recorded large floods in the historical record.

3.2 Bayesian Framework

Constructing a Bayesian framework involves defining the prior for each of the logistic regression model coefficients and data uncertainty parameters in equation 9, and selecting a numerical integration solver to estimate the joint posterior distribution of model parameters. Defining Ω as the set of model parameters to be estimated, the posterior distribution of parameters may be written as shown in equation 10

$$\Pr[\Omega|\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{C}] \propto \mathscr{L}(\Omega|\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{C})\Pr[\Omega] \tag{10}$$

where the likelihood can also be written as $f(x|\Omega)$ for fixed values of x and variable values of Ω.

We tested several prior distribution shapes for the logistic regression model coefficients, $\boldsymbol{\beta}$: 1) using a normal distribution with a mean of 0 and standard deviation of 10, 2) using a normal distribution with mean equal to the estimated MAP values from the Firth logistic regression (Tab. 3), and 3) using a uniform distribution with a range of [-30, 30]. These priors provided similar results, so we decided to present results for normal distribution centered at 0.

Assigning prior distribution shapes to the sensitivity and specificity parameters amounts to making assumptions about how likely it is that a recorded flood magnitude or occurrence is accurate. Almost surely, each expert would come up with different distribution shapes to use for each of the flood categories, and could even use a different distribution for each year if support for a large flood varied from year to year. For demonstration purposes, in this paper we decided to use naive uniform priors for each of the flood categories to avoid unduly influencing the parameter values.

For Markov Chain Monte Carlo (MCMC) estimation of the posterior distribution, we employed the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm with archive (z) and snooker update (s) adaptations (DREAM$_{zs}$) (Laloy & Vrugt, 2012). We used seven independent chains and a sufficient number of steps in the chain to ensure convergence of the estimated posterior distribution, as evaluated by the Gelman-Rubin potential scale reduction factor (Gelman & Rubin, 1992). We evaluated autocorrelation of samples in each chain and selected a thinning rate that ensured essentially uncorrelated samples. About 10% of the total iterations were used to adapt the transition probabilities used within the DREAM algorithm, and all of those iterations were not saved as part of the chains. Other DREAM$_{zs}$ hyperparameter settings used the recommended default settings in the BayesianTools R package (Hartig et al., 2017). A table of hyperparameter values and MCMC diagnostic figures are provided in the supplement for each model.

To estimate a distribution of possible annual large ice jam flood probabilities, we used the final 1001/7 samples from each chain to construct the posterior distribution of parameters. We used estimated $\boldsymbol{\beta}$ values to estimate the true probability of a large ice jam flood, p_i, and used the η and $\boldsymbol{\theta_C}$ values to estimate the probability of recording a large flood, q_i. Samples of large floods, Z_i, and recorded large floods, Y_i, were then generated using a Bernoulli distribution with those estimated probabilities.

3.3 Projecting Large Ice Jam Flood Probability in Future Climates

We employed the results from forcing 6 Global Climate Models (GCMs) (HadGEM2, ACCESS, CanESM, CCSM4, CNRM-CM5, and MPI-ESM-LR) with two Representative Concentration Pathways (RCPs) (van Vuuren et al., 2011) to simulate future DDF and snowpack for our selected weather stations from 2020 - 2100. We used RCP4.5 and RCP8.5 to be consistent with the scenarios used in (Lamontagne et al., 2021). The projected explanatory variables were computed from downscaled estimates of temperature and precipitation at each of our stations (Cannon et al., 2015; Werner & Cannon, 2016). The DDF and snowpack variables were then transformed into the PC axes and used with each of the posterior samples of $\boldsymbol{\beta}$ to estimate a distribution of annual probabilities of a large ice jam flood. For this study, we present 20-year averages in the predicted mean, 25^{th}, and 75^{th} percentile of the projected annual large ice jam flood probability from 2020-2100. For brevity, we present results for 1 GCM and both RCPs, and the remaining GCM results are provided in the supplement.

3.4 Regression Model Selection

We aim to demonstrate the impact of neglecting uncertainty in the historical data when estimating annual large ice jam flood probabilities. To accomplish this, we consider a baseline model for which the parameters are estimated using all available data while assuming that the historical record is correct. While we could solve for the full posterior distribution for each model, as described in Section 3.2, we are simply interested in which combination of climatic factors result in the most preferable baseline model. So, we instead compute the MAP estimates of the regression model coefficients in equation 3 using maximum likelihood estimation. We compute the statistical significance of each coefficient, and compare model performance using the second-order corrected Akaike Information Criterion (AICc). The AICc is preferred over AIC for small sample sizes. We consider the model with the smallest AICc and most significant coefficients to be the baseline model structure. We refer to this model in tables and figures as"Historical Uncertainty Not Considered, Trained 1915-2020."

We compare the baseline model to three additional models. The first is a model that uses the same climatic variables but considers historical flood magnitude and occurrence uncertainty. We refer to this model as"Historical Uncertainty Considered, Trained 1915-2020." The second also uses the same climatic variables but is trained using only the systematic record. We refer to this model as"Best Model with Ft. Smith, Trained 1962-2020." This label is used because Fort Smith had 3 fewer years of record than the other stations, so model training and performance metrics for all models were compared for only the years that all stations have in common. For PCA models with the 1962-2020 data, we used the same normalization means and standard deviations as were used for the 1915-2020 data so that the resulting coefficients, β, assigned to the PCs are comparable to each other. The final model is the best model from Lamontagne et al. (2021), which is also trained on the systematic record and uses Fort Vermillion DDF and Grande Prairie / Beaverlodge snowpack as predictors. We refer to this model as"Lamontagne et al. Best Model, Trained 1962-2020." Each of these models are solved using the Bayesian framework described in Section 3.2. Results for each of these models are compared for the full record from 1915 - 2020.

4 Results

Table 2 contains logistic regression model coefficients and the AICc for 2 and 3 explanatory variable models that were trained on 1915-2020 data while assuming no uncertainty in the historical flood record. As in Lamontagne et al. (2021), we find that the most significant single predictor is snowpack. Without using PCA, none of the DDF variables are statistically significant at the 5% level in 2 or 3 parameter models, but they provide a competitive model according to the AICc. For the models that use PCs, the best model provides the lowest AICc and the most significant coefficients among the models tested. This model uses the first 2 PCs from a PCA with GP/BL snowpack, and DDF from FC, FV, and FS stations.

With the baseline model established, we estimated the parameters for each of the four models described in Section 3.4. The estimated MAP and 95% credible intervals (CIs) are provided for each regression model coefficient in Table 3. For the three PC models, each of them show statistical significance from 0 at the 5% level, but the most credible range is different for each model. There is a striking difference in CI width for the model that neglects uncertainty in the historical data and the other models' CIs. This model thinks the historical data are certain, and so it has higher confidence in the estimated parameter values. It is also biased towards lower coefficient values for PC1 and PC2 compared to the other PC models. The CIs for the model with uncertainty considered are wider than the CIs for the best model with Ft. Smith, and they are also more extreme in absolute value. This indicates that considering uncertainty in the historical

Table 2. Logistic regression models for 1915-2020, assuming no uncertainty in the flood record. Bold indicates statistical significance at the 5% level. GP/BL: Grande Prairie / Beaverlodge Snowpack, FC: Fort Chipewyan DDF, FS: Fort Smith DDF, FV: Fort Vermillion DDF, MT: melt test, PC: principal component, AICc: second-order corrected Akaike Information Criterion. The first 2 or 3 PCs are used for models with PCs.

# Variables	Models	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	AICc
2	GP/BL + FV	**-2.43**	**1.34**	-0.31		55
	GP/BL + FC	**-2.49**	**1.27**	-0.46		54.5
	GP/BL + FS	**-2.40**	**1.42**	-0.12		55.6
	PCs from GP/BL, FC, FS, FV	**-2.45**	**0.70**	**1.15**		54
	PCs from GP/BL, FC, FS, FV, MT	**-2.24**	**0.68**	0.11		60.7
3	GP/BL + FC + Interaction	**-2.40**	**1.26**	-0.47	0.1	55.2
	GP/BL + FC + MT	**-2.44**	**1.26**	-0.39	-0.17	54.5
	PCs from GP/BL, FC, FS, FV	**-2.43**	**0.70**	**1.11**	-0.35	55.8
	PCs from GP/BL, FC, FS, FV, MT	**-2.40**	**0.68**	0.16	**1.15**	53.8

Table 3. Maximum *a posteriori* (MAP) model coefficients and 95% credible intervals (CI).

Model	Coefficient	MAP	95% CI
Historical Uncertainty Considered	Intercept	-3.18	[-8.27, -2.04]
Trained 1915-2020	PC1	1.98	[1.16, 5.46]
	PC2	1.57	[0.47, 5.07]
Historical Uncertainty Not Considered	Intercept	-2.44	[-3.74, -1.79]
Trained 1915-2020	PC1	0.70	[0.31, 1.31]
	PC2	1.12	[0.34, 2.32]
Best Model with Ft. Smith	Intercept	-3.11	[-7.19, -1.86]
Trained 1962-2020	PC1	1.74	[0.91, 4.15]
	PC2	1.71	[0.36, 4.53]
Lamontagne et al. Best Model	Intercept	-4.7	[-10.95, -2.68]
Trained 1962-2020	DDF[a]	-1.65	[-4.03, -0.49]
	Snowpack[b]	2.31	[0.94, 5.69]

[a]Ft. Vermillion degree-days freezing (DDF)
[b]Grande Prairie / Beaverlodge snowpack

record provided even more evidence in support of these parameters being predictive of large ice jam flood occurrence.

4.1 Understanding the Model with Historical Uncertainty Considered

The estimated marginal posterior distributions of the parameters for the model that considers historical data uncertainty are plotted in Figure 3. The distribution of the sensitivity parameter indicates a relatively small 25% probability that the recorded flood magnitude is large when a large flood actually occurred. The distributions of the specificity parameter categories all indicate a relatively high probability that the recorded flood magnitude is not large when a large flood did not occur. The most certain of these is the years with no flood recorded, which should be expected. This is followed by years categorized as having a moderate flood, then years with recorded small floods, and finally years with unknown magnitude floods. Years with a categorized moderate flood could have a lower probability of being large than years with a categorized small flood

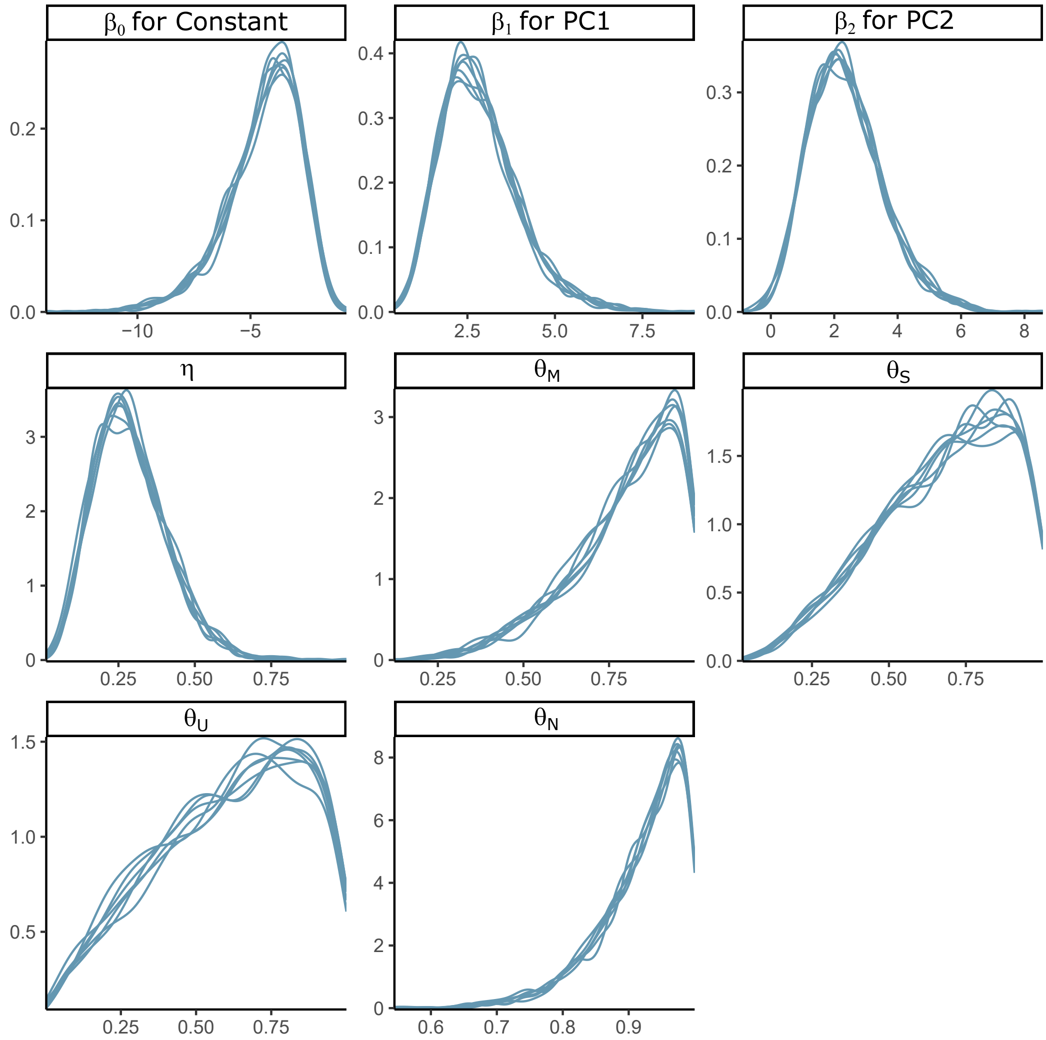

Figure 3. Marginal posterior distributions for each of the parameters in the logistic regression model that considered historical data uncertainty. The distributions from each of seven independent MCMC chains are overlain, showing good convergence. This figure was made using the bayesplot R package (Gabry & Mahr, 2018).

because the critical locations in the PAD that are used to assess moderate flood magnitudes are also the locations used to assess large floods (Timoney, 2009). In other words, the moderate labels could be expected to more certainly be moderate than the small labels, for which other unobserved factors could have resulted in the flood actually being large.

To explain the sensitivity and specificity results, it can help to consider the confusion matrix. The confusion matrix consists of true positives (a large flood is recorded when a large flood occurred), true negatives (no flood is recorded when a large flood did not occur) as well as the false positives and negatives. A low sensitivity can occur due to few true positives or many false negatives (recorded non-large floods that were likely large) or a combination of both. Similarly, a low specificity can occur due to few true negatives or many false positives. In our case, we do not know if a large ice jam flood

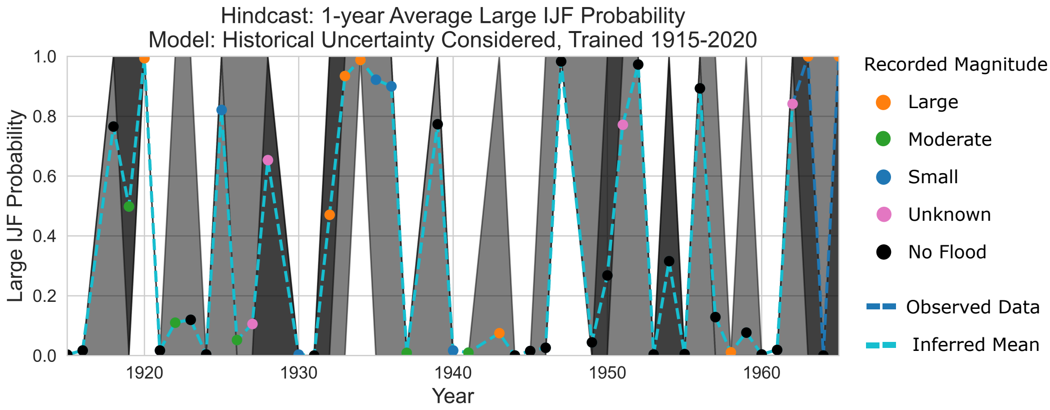

Figure 4. Annual average probability of an ice jam flood (IJF) provided as 50% (dark gray) and 90% (light gray) prediction intervals. The observed known data are connected by a dark blue line. The recorded uncertain historical data are plotted at the inferred mean value from the model that considered historical data uncertainty and they are connected by a light blue line. Point colors indicate the recorded flood magnitude.

truly did or did not occur, but for illustrative purposes we can estimate the sensitivity and specificity based on the estimated probability of a large ice jam flood.

Figure 4 provides a visual assessment of the recorded flood categories and predicted true probabilities of a large ice jam flood. From this plot, we see that three of the recorded large ice jam floods were in years with very high probability of a large ice jam flood, one was in a year with a roughly 50-50 chance, and the other two were in years with low probability. The expected number of true positives for these data would be about 3.5. For the false negatives, we see high estimated probability of a large ice jam flood for five years with no flood recorded, about three years with unknown magnitude floods recorded, and three years with small magnitude floods recorded. There is also one year with a recorded moderate flood with about a 50-50 chance of being large. The expected false negatives would be about 11.5. Therefore, the expected sensitivity would be $3.5/15 = 23\%$, which is very close to the mode of the estimated sensitivity parameter value in Figure 3.

The sensitivity and specificity results so far suggest that more large ice jam floods are likely to have happened than were recorded in the historical record, based on the fitted model and historical climate. One way to further visualize this is to plot the difference between the predicted probability of a large ice jam flood truly occurring, $\hat{p}_i$, and the the predicted probability of recording a large ice jam flood, $\hat{q}_i$. In Figure 5, we see that most of the differences are positive, which again supports that the record contains fewer large floods than our model estimates actually occurred.

4.2 Hindcasting and Projecting Results

Figure 6 shows the prediction intervals for each of the four models. For the historical record, each model's estimated large ice jam flood probability contains the inferred mean from the model that considers historical data uncertainty. Even though the inferred mean is not necessarily what truly occurred, it is a useful point of reference to compare models. Over the whole record, the model that ignores historical data uncertainty is consistently biased towards lower probability. Biased performance of this model in the systematic time period is an indicator that the historical record likely had some misclassified flood magnitudes or occurrences. The Best Model with Ft. Smith trained from 1962-

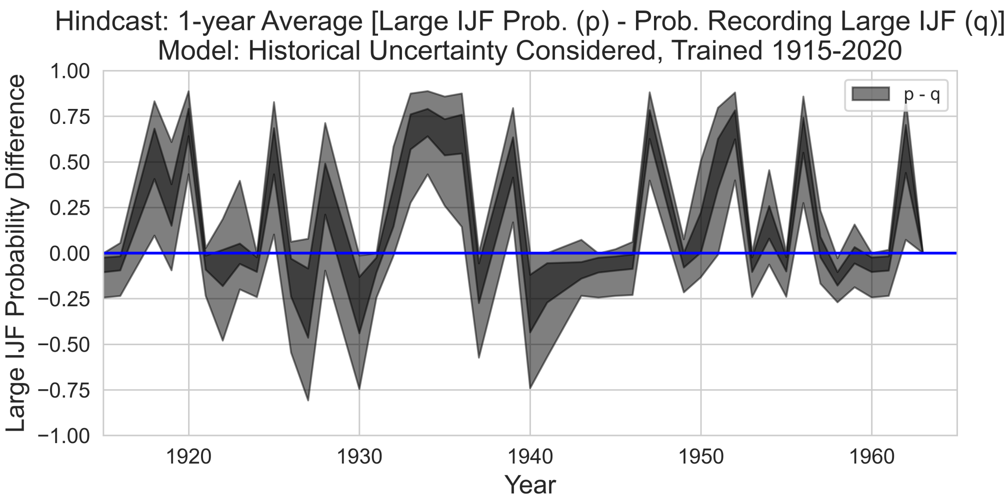

Figure 5. Annual average probability of an ice jam flood (IJF) and recording an IJF, provided as 50% (dark green/blue) and 90% (light green/blue) prediction intervals.

2020 was consistently biased higher. The best model from Lamontagne et al. (2021) provides similar estimates as the model that considers historical data uncertainty, even though it was trained with only the systematic record. Visually, each model has a similar predicted trend in large ice jam flood probability over time.

Figure 7 compares the prediction intervals for each of the GCM-RCP scenarios for each model. This figure plots probability in log space, so we also provide Table 4 with the interquartile ranges (IQRs) of the predictions at selected years to compare model results. Except for the model that ignores historical data uncertainty, we see an overall decreasing trend in the probability of a large ice jam flood from 2020-2100, consistent with Lamontagne et al. (2021). The model that considers historical data uncertainty has the most precise IQRs among the models considered. It also estimates the smallest probability of a large ice jam flood, which is in part due to the larger coefficient values estimated for PC1 (mostly temperature) as compared to the other models. To further support that point, Figure 8 provides the projected data on each of the PC axes. We see that projected climatic conditions explore a larger space in the PC domain than did the 1915-2020 record, and that the primary change is an increase in the temperature. As PC1 becomes more negative quicker than PC2 becomes more positive, the probability of a large ice jam flood decreases.

For the model that neglects historical data uncertainty, the projected probability of a large ice jam flood is similar to the probability from 1915-2020. This results from estimating a smaller coefficient for PC1 and PC2 due to historical recorded large floods being located in warmer winter conditions (Fig. 2, Fig. 8). These years were less likely to have a large ice jam flood occur according to the model that considers historical data uncertainty.

5 Discussion

The recent past is a small sample of the climatic conditions that a region may experience. For the PAD, the historical record contains nearly the same number of years as the systematic record, but the historical data are uncertain. We demonstrated that uncertain flood magnitudes and occurrences could be used to estimate annual probabil-

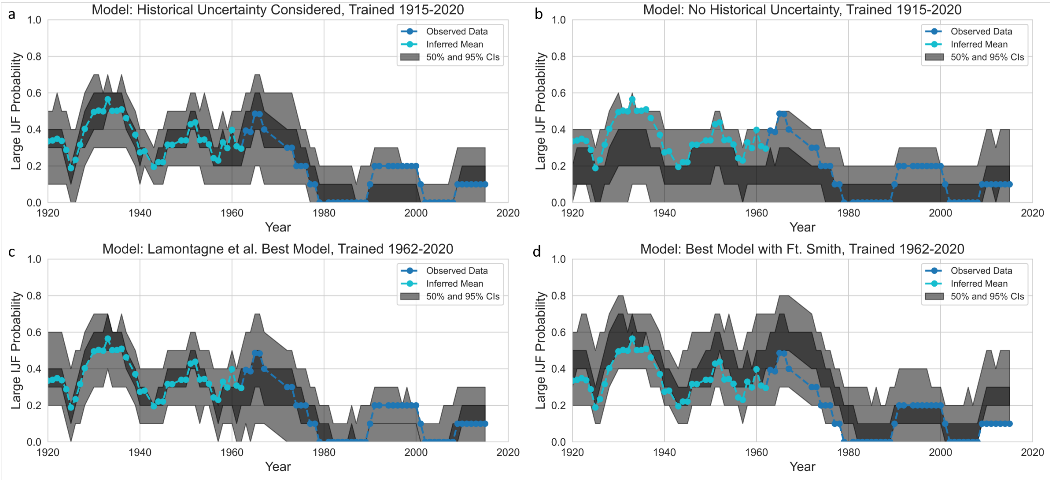

Figure 6. Centered 10-year average probability of an ice jam flood (IJF) provided as 50% (dark gray) and 90% (light gray) prediction intervals. The observed 10-year average is in blue and the 10-year average of the inferred mean from the model with historical uncertainty considered is in light blue.

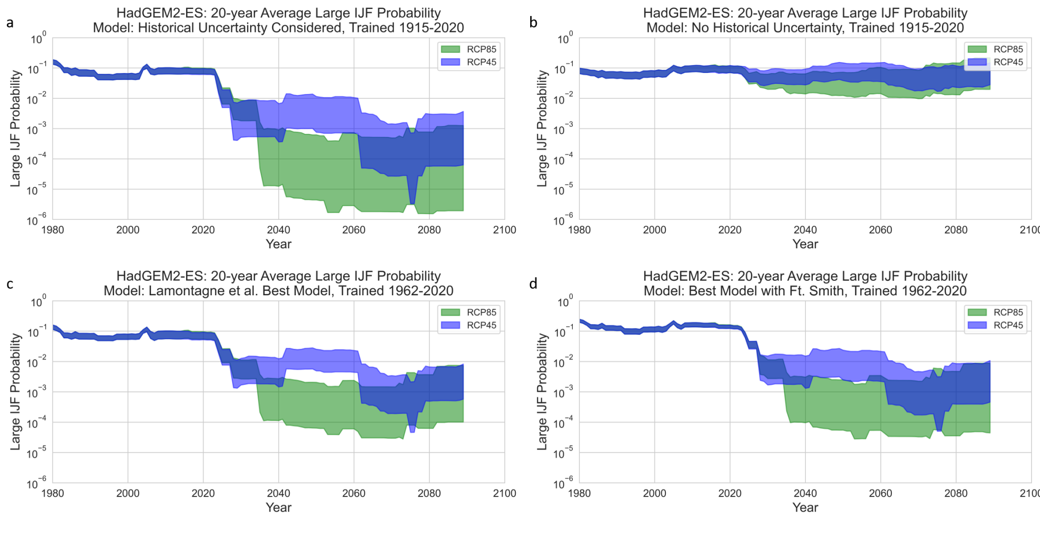

Figure 7. Centered 20-year average probability of a large ice jam flood (IJF) using the HadGEM2-ES GCM forced with RCP4.5 (blue) and RCP8.5 (green). Predictions are provided as the 50% prediction intervals. Projections begin in 2020. The y-axis shows probability in log-space, so a reduction by 1 tick mark corresponds to an order of magnitude reduction in probability.

Table 4. Interquartile ranges of large ice jam flood probability for projections in Figure 7

Model	Projection	2020	2040	2060	2080
Historical Uncertainty Considered Trained 1915-2020	RCP4.5	3.4×10^{-2}	6.3×10^{-3}	1.0×10^{-2}	3.0×10^{-3}
	RCP8.5	3.7×10^{-2}	9.0×10^{-4}	7.3×10^{-4}	7.7×10^{-4}
Historical Uncertainty Not Considered Trained 1915-2020	RCP4.5	5.1×10^{-2}	6.5×10^{-2}	1.1×10^{-1}	9.5×10^{-2}
	RCP8.5	4.8×10^{-2}	5.0×10^{-2}	9.1×10^{-2}	1.3×10^{-1}
Best Model with Ft. Smith Trained 1962-2020	RCP4.5	3.7×10^{-2}	1.4×10^{-2}	1.9×10^{-2}	8.1×10^{-3}
	RCP8.5	3.8×10^{-2}	3.0×10^{-3}	3.4×10^{-3}	4.7×10^{-3}
Lamontagne et al. Best Model Trained 1962-2020	RCP4.5	4.0×10^{-2}	1.1×10^{-2}	1.9×10^{-2}	6.1×10^{-3}
	RCP8.5	4.2×10^{-2}	2.6×10^{-3}	2.3×10^{-3}	3.6×10^{-3}

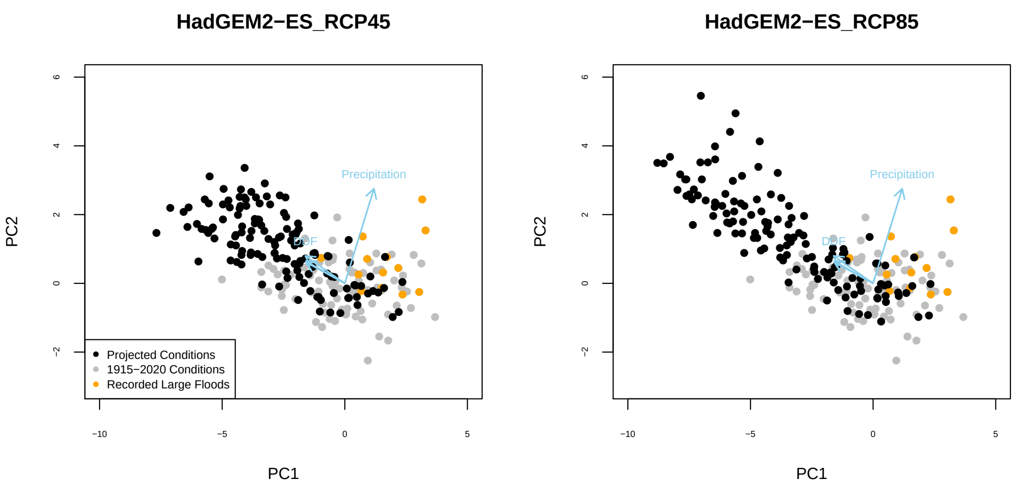

Figure 8. Projected future climate conditions and the 1915-2020 conditions plotted on the PC axes.

ities of large ice jam floods in the PAD. An interesting question is whether such historical information improved estimates of flood frequency. To answer that question, we can look to the bias and uncertainty variance of the estimated predictions. The climatic conditions in the historical time period overlap with the systematic time period and also include years with more extreme DDF and snowpack, as well as several additional years with recorded large ice jam floods. Our model that considers uncertainty in the historical record estimates that some of those large floods were likely truly large, and also estimates that some of the non-large-flood years may have had a large flood. As a result, we see model coefficients increase in absolute value relative to models that were trained on only the systematic record. This result suggests that using only the systematic record may bias coefficients to be smaller as a result of not sampling a full set of representative climatic conditions that lead to large ice jam flood generation in the PAD. As a result of this bias in the coefficients, the projected annual probabilities of large ice jam floods are larger for models trained only with the systematic record. For example, the 25^{th} quantile estimate in Figure 7 is a substantial one order of magnitude smaller for the model that considers historical data uncertainty. This result motivates using all available data in prediction models while appropriately handling uncertainty to better inform projections under climatic conditions that could differ from recent history.

Turning our discussion to assuming all available data are certain, we found that this results in a model that does worse in the systematic time period than the other three models (biased to lower annual probability of a large ice jam flood than was observed). While it is possible for any model to provide different predictions after new data are added, it is unlikely to see a drastic change in the model predictions like we observed when training the same model structure on the systematic record. So, the modeling assumptions (uncertainties) must be examined to understand why the predictions may differ and if there is a problem with our models. For our study, uncertainties can arise from structural assumptions (e.g., the linear combination of climatic factors being used as explanatory variables), parametric uncertainty in the values of the estimated coefficients for each of the factors, and nonstationarity in the system's response to the climatic factors (i.e., a different effect of snowpack and DDF on large ice jam flood generation over time, as modeled by the logistic regression). In this study, we accounted for parametric uncertainty by sampling from the Bayesian-estimated posterior distribution of parameters to estimate a distribution of ice jam flood probability in each year. This is the Bayesian analog to the bootstrapping approach presented in Lamontagne et al. (2021). We accounted for structural uncertainty by comparing two different combinations of climatic factors for the systematic record, and both models provided similar probability estimates. We also compared the influence of prior for the regression model coefficients and found similar results (in supplementary information). Our logistic regression model does assume stationarity in the system's response to the climatic factors; however, previous studies have argued that stationarity is an appropriate assumption for the PAD on the timescales that are being modeled (Beltaos, 2014; Lamontagne et al., 2021).

Ruling out these key modeling assumptions as causes for different results when data uncertainty is ignored, we can turn to the climatic explanatory variables as possible sources of error. These variables are derived based on temperature and precipitation records. All meteorological stations have measurement errors, but these are typically small and are not likely to change much over the period of record, although more error for older data would be expected as technology has improved over time. So, data uncertainty in the historical record is left as the best possible explanation for the drastic difference between our models that do and do not consider data uncertainty. This result highlights the danger of assuming all data are certain when it is known that data uncertainty exists.

Our evaluation of structural uncertainty revealed that the best model structure can change after adding more data. This result suggests that the climatic factors that we considered could vary in importance from year to year based on the processes that they affect for the PAD (Jasek, 2019a, 2019b). Thus, the small 1962-2020 sample may not be representative of how these factors influence the generation of large ice jam floods over the long run. When structural uncertainty is a concern, several competing model structures may be considered within a Bayesian framework using multi-model ensemble methods, like Bayesian model averaging (e.g., Duan et al., 2007). For this particular study, the differences in predicted probabilities are small and have similar trends for the structural models considered. Considering structural uncertainty in a formal way may not be necessary unless more precision is required (e.g., for risk estimation).

When allowing for data uncertainty, a critical question is the validity of the resulting predictions. For our model, we see that the predictions in the systematic record resemble those of the models that were trained using only the systematic time period. As reported in Lamontagne et al. (2021), our models predict a climatic-driven decline in large ice jam flood probability in the 1970s, which continues to present day. We also see that models that were trained with only the systematic record show predicted trends in the historical period that are similar to those predicted by the model that considered historical data uncertainty. In tandem, these results provide confidence for each of the models that we used, but how do the predicted trends in probability compare to paleolimnological evidence that has been collected across the PAD?

There are several instances in the historical time period for which our model estimates a higher probability of a large ice jam flood than was recorded. Sediment core data from oxbow lakes in the PAD that are summarized by Wolfe et al. (2020, 2006) reveal similar trends in likely flood events, particularly for lake PAD 15 magnetic susceptibilities before 1920, in the mid-1930s, and in the 1950s. Another lake that flooded more frequently, PAD 54, also reveals similar timing of flood events within the 5-yr uncertainty range suggested by (Wolfe et al., 2020). Although these are examples from just two lakes and additional lakes should be evaluated to support the concept of widespread flooding that is necessary for a large ice jam flood, this sediment core evidence provides support for our model results. Even so, it is important to note that these results are still probabilistic, and what is likely to have occurred based on our models and paleo evidence does not mean that it did occur. The comparison of statistical model results to in-situ data collected in the PAD should be further examined while considering the uncertainties present in both flood and paleo records (Wolfe et al., 2020). If warranted, paleo data can be integrated into the flood frequency analysis as well (e.g., Harden et al., 2021).

5.1 Limitations

For model building, we rely on expert-interpreted flood magnitude and occurrence information. We group years with the same flood category together and, in doing so, we assume that the same information was available for each year in that category. It is instead possible that each year had different information available based on written records and proxy evidence (Timoney, 2009), which could mean that one year in a category could be more certain than others. The framework that we presented is general and can be modified to have each individual year as their own category, at the expense of more parameters in the model that describe the uncertainty attributed to each year. For such models, a good rule of thumb is that the total number of parameters should be smaller than the total number of data points to avoid overfitting and to arrive at an explainable model. In supplementary material, we present model diagnostic results of a model that considers a sensitivity and a specificity parameter for each year of the historical record. We find that the years with the most extreme climatic conditions have marginal distributions that converge away from the uniform prior distribution, but other years still resemble a uniform shape. This is likely a result of too many parameters to estimate a reliable value for all individual years. However, it also suggests that if prior information were available for individual years, it could have high influence on the resulting model (i.e., the posterior distribution for sensitivity or specificity in each year may not deviate much from the prior distribution when too many parameters are used).

We do not explore the impact of the choice of prior distribution (uniform) for the sensitivity and specificity parameters in this study. Based on the converged values of sensitivity and specificity in Figure 3, our model was able to converge to a shape that is different from uniform for each of the flood categories, and therefore the choice of using a uniform prior likely does not have much influence on the resulting posterior model coefficients for the climatic factors. The choice of prior for sensitivity and specificity could be explored further in future work for which the uniform distribution could serve as a baseline model to which alternative beliefs about flood magnitude and occurrence uncertainty could be compared.

6 Conclusions

We presented a Bayesian logistic regression framework to account for uncertainties in historical flood magnitude and occurrence. We find that considering uncertainty in the historical record both reduces the minimum predicted probability of a large ice jam flood and narrows the uncertainty range in projected future climate conditions compared to neglecting the uncertainty or using only the systematic record. Therefore, our Bayesian framework allows for the use of a longer and less reliable record to obtain more precise estimates of ice jam flood probability.

7 Open Research

Version number v3.0.0 of the PAD_IceJamFloods GitHub repository that was used for statistical modeling and figure generation is preserved at https://doi.org/10.5281/zenodo.7484504, and is available via the MIT License and developed openly at https://github.com/jds485/PAD_IceJamFloods in the bayesian_regression folder. This repository also contains the flood and climate data used in this study.

Acknowledgments

This work was supported by BC Hydro. M. Jasek is an employee of BC Hydro. The lead authors were in control of all analysis and conclusions reached in this paper.

References

Beltaos, S. (2014, 12). Comparing the impacts of regulation and climate on ice-jam flooding of the peace-athabasca delta. *Cold Regions Science and Technology*, *108*, 49-58. doi: 10.1016/j.coldregions.2014.08.006

Beltaos, S., Prowse, T., Bonsal, B., Carter, T., MacKay, R., Romolo, L., ... Toth, B. (2008). *Climate impacts on ice-jam floods in a regulated northern river.* Springer Berlin Heidelberg. doi: 10.1007/978-3-540-75136-6_18

Benito, G., Lang, M., Barriendos, M., Llasat, M. C., Francés, F., Ouarda, T., ... Bobée, B. (2004, 3). Use of systematic, palaeoflood and historical data for the improvement of flood risk estimation. review of scientific methods. *Natural Hazards*, *31*, 623-643. doi: 10.1023/B:NHAZ.0000024895.48463.eb

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015, 9). Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, *28*, 6938-6959. doi: 10.1175/JCLI-D-14-00754.1

Condie, R., & Lee, K. A. (1982, 8). Flood frequency analysis with historic information. *Journal of Hydrology*, *58*, 47-61. doi: 10.1016/0022-1694(82)90068-3

Das, A., Rokaya, P., & Lindenschmidt, K.-E. (2020, 6). Ice-jam flood risk assessment and hazard mapping under future climate. *Journal of Water Resources Planning and Management*, *146*. doi: 10.1061/(ASCE)WR.1943-5452.0001178

Duan, Q., Ajami, N. K., Gao, X., & Sorooshian, S. (2007, 5). Multi-model ensemble hydrologic prediction using bayesian model averaging. *Advances in Water Resources*, *30*, 1371-1386. doi: 10.1016/j.advwatres.2006.11.014

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*, 27-38. doi: 10.1093/biomet/80.1.27

Gabry, J., & Mahr, T. (2018). bayesplot: Plotting for bayesian models [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=bayesplot` (R package version 1.6.0)

Gelman, A. (2009, 5). Bayes, jeffreys, prior distributions and the philosophy of statistics. *Statistical Science*, *24*. doi: 10.1214/09-STS284D

Gelman, A., & Rubin, D. B. (1992, 11). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*. doi: 10.1214/ss/1177011136

Harden, T. M., Ryberg, K. R., O'Connor, J. E., Friedman, J. M., & Kiang, J. E. (2021). *Historical and paleoflood analyses for probabilistic flood-hazard assessments - approaches and review guidelines.* U.S. Geological Survey. doi: https://doi.org/10.3133/tm4B6

Hartig, F., Minunno, F., & Paul, S. (2017). Bayesiantools: General-purpose mcmc and smc samplers and tools for bayesian statistics [Computer software manual]. Retrieved from `https://github.com/florianhartig/BayesianTools` (R package version 0.1.4)

Heinze, G., Ploner, M., & Jiricka, L. (2020). logistf: Firth's bias-reduced logistic regression [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=logistf` (R package version 1.24)

Hosking, J. R. M., & Wallis, J. R. (1986, 10). The value of historical data in flood frequency analysis. *Water Resources Research*, *22*, 1606-1612. doi: 10.1029/WR022i011p01606

Jasek, M. (2019a). An emerging picture of peace river break-up types that influence ice jam flooding of the peace-athabasca delta. part 1: The 2018 peace river break-up. cgu hs committee on river ice processes and the environment. In *20th workshop on the hydraulics of ice covered rivers. ottawa, canada.*

Jasek, M. (2019b). An emerging picture of peace river break-up types that influence ice jam flooding of the peace-athabasca delta part 2: Insights from the comparison of the 2014 and 2018 break-ups. In *The 20th workshop on the hydraulics of ice covered rivers, ottawa, canada.*

Jasek, M. (2019c). Peace river break-up and pad ice jam flooding flow chart. In *20th workshop on the hydraulics of ice covered rivers* (p. 1). Ottawa, ON, Canada: CGU HS Committee on River Ice Processes and the Environment. Retrieved from `http://www.cripe.ca/docs/proceedings/20/Jasek-2019c.pdf`

Jasek, M., Lamontagne, J., & Smith, J. D. (2021). Analysis of climatic and riverine factors influencing peace river ice jam flood frequency in the peace-athabasca delta: From history to future climate implications. In *21st workshop on the hydraulics of ice covered rivers* (p. 31). Saskatoon, Saskatchewan, Canada: CGU HS Committee on River Ice Processes and the Environment. Retrieved from `http://www.cripe.ca/docs/proceedings/21/Jasek-et-al-2021.pdf`

Kjeldsen, T., Macdonald, N., Lang, M., Mediero, L., Albuquerque, T., Bogdanowicz, E., ... Wilson, D. (2014, 9). Documentary evidence of past floods in europe and their utility in flood frequency estimation. *Journal of Hydrology*, *517*, 963-973. doi: 10.1016/j.jhydrol.2014.06.038

Laloy, E., & Vrugt, J. A. (2012, 1). High-dimensional posterior exploration of hydrologic models using multiple-try dream ¡sub¿(zs)¡/sub¿ and high-performance computing. *Water Resources Research*, *48*. doi: 10.1029/2011WR010608

Lamontagne, J. R., Jasek, M., & Smith, J. D. (2021, 12). Coupling physical understanding and statistical modeling to estimate ice jam flood frequency in the northern peace-athabasca delta under climate change. *Cold Regions Science and Technology*, *192*, 103383. doi: 10.1016/j.coldregions.2021.103383

McInturff, P., Johnson, W. O., Cowling, D., & Gardner, I. A. (2004, 4). Modelling risk when binary outcomes are subject to error. *Statistics in Medicine*, *23*, 1095-1109. doi: 10.1002/sim.1656

Parkes, B., & Demeritt, D. (2016, 9). Defining the hundred year flood: A bayesian approach for using historic data to reduce uncertainty in flood frequency estimates. *Journal of Hydrology*, *540*, 1189-1208. doi: 10.1016/j.jhydrol.2016.07.025

Payrastre, O., Gaume, E., & Andrieu, H. (2011, 8). Usefulness of historical information for flood frequency analyses: Developments based on a case study. *Water Resources Research*, *47*. doi: 10.1029/2010WR009812

Peterson, M. (1995). *Peace-athabasca delta flood history study. task f.1 of the peace-athabasca delta technical studies.*

Prowse, T. D., & Conly, F. M. (2002, 9). A review of hydroecological results of the northern river basins study, canada. part 2. peace-athabasca delta. *River Research and Applications*, *18*, 447-460. doi: 10.1002/rra.682

Reis, D. S., & Stedinger, J. R. (2005, 11). Bayesian mcmc flood frequency analysis with historical information. *Journal of Hydrology*, *313*, 97-116. doi: 10.1016/j.jhydrol.2005.02.028

Salinas, J. L., Kiss, A., Viglione, A., Viertl, R., & Blöschl, G. (2016, 9). A fuzzy bayesian approach to flood frequency estimation with imprecise historical information. *Water Resources Research*, *52*, 6730-6750. doi: 10.1002/2016WR019177

Stedinger, J. R., & Cohn, T. A. (1986, 5). Flood frequency analysis with historical and paleoflood information. *Water Resources Research*, *22*, 785-793. doi: 10.1029/WR022i005p00785

Timoney, K. P. (2009, 4). Three centuries of change in the peace–athabasca delta, canada. *Climatic Change*, *93*, 485-515. doi: 10.1007/s10584-008-9536-4

Timoney, K. P. (2013). *The peace-athabasca delta: Portrait of a dynamic ecosystem.* University of Alberta Press.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., … Rose, S. K. (2011, 11). The representative concentration pathways: an overview. *Climatic Change*, *109*, 5-31. doi: 10.1007/s10584-011-0148-z

Werner, A. T., & Cannon, A. J. (2016, 4). Hydrologic extremes – an intercomparison of multiple gridded statistical downscaling methods. *Hydrology and Earth System Sciences*, *20*, 1483-1508. doi: 10.5194/hess-20-1483-2016

Wolfe, B. B., Hall, R. I., Last, W. M., Edwards, T. W. D., English, M. C., Karst-Riddoch, T. L., … Palmini, R. (2006, 12). Reconstruction of multi-century flood histories from oxbow lake sediments, peace-athabasca delta, canada. *Hydrological Processes*, *20*, 4131-4153. doi: 10.1002/hyp.6423

Wolfe, B. B., Hall, R. I., Wiklund, J. A., & Kay, M. L. (2020, 9). Past variation in lower peace river ice-jam flood frequency. *Environmental Reviews*, *28*, 209-217. doi: 10.1139/er-2019-0047

Supporting Information for "Considering Uncertainty of Historical Ice Jam Flood Records in a Bayesian Frequency Analysis for the Peace-Athabasca Delta"

Jared D. Smith[1] *, Jonathan R. Lamontagne[2], Martin Jasek[3]

[1]University of Virginia Department of Engineering Systems and Environment, Charlottesville, VA, USA

[2]Tufts University Department of Civil and Environmental Engineering, Medford, MA, USA

[3]BC Hydro, Burnaby, BC, Canada

Contents of this file

1. Tables S1 and S2
2. Figures S1 to S26

Additional Supporting Information (Files uploaded separately)

1. Captions for Dataset S1 to S5

Corresponding author: J. D. Smith, Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA. (jared.d.smith485@gmail.com)

*Currently employed at: U.S. Geological Survey, Reston, VA, USA

Introduction

Table S1 provides the hyperparameter values we used in the DREAM$_{zs}$ algorithm. Table S2 provides interquartile ranges for all of the GCM and RCP scenarios for each of the regression models.

Figures S1 to S19 provide MCMC diagnostics for each of the 4 models tested. Figure S20 provides marginal distributions for a model that contains a sensitivity or specificity parameter for each of the historical years of record, and Figures S21 to S25 provide exploratory data analysis plots with historical large, moderate, small, and no flood years labeled to match the sensitivity and specificity numbers in Figure S20. Figure S26 provides the difference between the predicted probability of a large ice jam flood using the model that does not consider historical data uncertainty, and the predicted probability of recording a large ice jam flood using the model that does consider historical data uncertainty. These values should be similar because the model that does not consider historical data uncertainty is trying to match the historical data exactly as recorded. They may be different because of the influence of the systematic record.

The flood records and derived climatic variables are provided in Dataset S1 for 1915-2020. Datasets S2-S5 provide the downscaled climatic variables for our projected scenarios. All of these datasets are also available in our GitHub repository.

Dataset S1 File cleaned_dataLMSAllYears.csv provides the interpreted flood record and annual climatic variables for 1915-2020. Climatic variables: cumulative degree-days freezing for Fort Chipewyan, Fort Vermillian, and Fort Smith; Melt test; Beaverlodge snowpack, and the derived Grande Prairie / Beaverlodge variable that we used. Flood record: Flood (large floods), AllLM (all large or moderate floods), AllLMS (all large, moderate, or small floods), and FloodMag (the recorded flood category).

Dataset S2 File GCM_Temp_Smith.csv provides Fort Smith annual cumulative degree-days freezing estimates for the 12 GCM and RCP scenarios used in this study.

Dataset S3 File GCM_Temp_Verm.csv provides Fort Vermillion annual cumulative degree-days freezing estimates for the 12 GCM and RCP scenarios used in this study.

Dataset S4 File GCM_Temp_Chip.csv provides Fort Chipewyan annual cumulative degree-days freezing estimates for the 12 GCM and RCP scenarios used in this study.

Dataset S5 File GCM_Precip.csv provides Grande Prairie annual snowpack estimates for the 12 GCM and RCP scenarios used in this study.

DREAMzs Hyperparameter	Model with Historical Uncertainty Considered	Model with Historical Uncertainty Considered, Parameters for every historical year	All Other Models
Iterations	2,000,000	4,000,000	330,000
Burnin iterations	50,000	100,000	30,000
Adaptation iterations	50,000	100,000	30,000
Archive update frequency	50 iterations	100 iterations	30 iterations
Thinning frequency	100 iterations	200 iterations	30 iterations
Number of crossovers	3	3	3
Crossover update interval	10 iterations	10 iterations	10 iterations
Differential Evolution pairs	2	2	2
Snooker update probability	0.1	0.1	0.1
Ergodicity	0.05	0.05	0.05

Table S1. Table of DREAM$_{zs}$ hyperparameters.

Historical Uncertainty Considered, Trained 1915-2020	RCP 8.5				RCP 4.5			
	2020	2040	2060	2080	2020	2040	2060	2080
HadGEM2-ES	3.7E-02	8.9E-04	7.3E-04	7.7E-04	3.4E-02	6.3E-03	9.9E-03	3.0E-03
ACCESS1-0	3.4E-02	1.1E-02	3.9E-03	2.5E-04	3.8E-02	2.0E-02	4.5E-03	1.2E-02
CanESM2	4.0E-02	1.2E-02	3.8E-02	6.5E-03	3.4E-02	3.1E-02	3.6E-02	5.1E-02
CCSM4	3.6E-02	2.6E-03	6.9E-03	5.8E-04	3.6E-02	2.7E-02	4.2E-03	5.7E-03
CNRM-CM5	4.4E-02	1.9E-02	7.0E-03	1.1E-03	4.3E-02	8.9E-03	1.2E-02	3.2E-02
MPI-ESM-LR	3.4E-02	2.1E-03	4.9E-03	6.0E-04	4.6E-02	2.5E-03	1.6E-03	1.1E-03
Historical Uncertainty Not Considered, Trained 1915-2020	**RCP 8.5**				**RCP 4.5**			
	2020	2040	2060	2080	2020	2040	2060	2080
HadGEM2-ES	4.8E-02	5.0E-02	9.1E-02	1.3E-01	5.1E-02	6.5E-02	1.1E-01	9.5E-02
ACCESS1-0	5.1E-02	7.1E-02	6.8E-02	8.1E-02	5.0E-02	5.6E-02	7.0E-02	9.4E-02
CanESM2	5.7E-02	8.7E-02	1.2E-01	2.0E-01	5.8E-02	6.6E-02	8.8E-02	1.2E-01
CCSM4	5.3E-02	4.6E-02	6.1E-02	5.0E-02	4.7E-02	5.8E-02	5.6E-02	6.2E-02
CNRM-CM5	5.4E-02	5.8E-02	8.9E-02	8.3E-02	5.2E-02	3.7E-02	7.0E-02	5.6E-02
MPI-ESM-LR	5.4E-02	6.4E-02	7.6E-02	7.5E-02	7.0E-02	5.5E-02	5.2E-02	6.8E-02
Best Model with Ft. Smith, Trained 1962-2020	**RCP 8.5**				**RCP 4.5**			
	2020	2040	2060	2080	2020	2040	2060	2080
HadGEM2-ES	3.8E-02	3.0E-03	3.4E-03	4.7E-03	3.7E-02	1.4E-02	1.9E-02	8.2E-03
ACCESS1-0	3.6E-02	1.9E-02	8.4E-03	1.2E-03	4.0E-02	2.4E-02	1.1E-02	2.1E-02
CanESM2	4.8E-02	2.4E-02	5.6E-02	2.6E-02	3.8E-02	3.7E-02	5.0E-02	6.2E-02
CCSM4	3.9E-02	5.4E-03	1.3E-02	2.0E-03	3.8E-02	3.2E-02	7.1E-03	1.2E-02
CNRM-CM5	5.0E-02	2.6E-02	1.7E-02	4.1E-03	4.5E-02	1.2E-02	2.1E-02	3.3E-02
MPI-ESM-LR	3.7E-02	5.5E-03	1.1E-02	2.5E-03	5.1E-02	6.5E-03	4.6E-03	3.9E-03
Lamontagne et al. Best Model, Trained 1962-2020	**RCP 8.5**				**RCP 4.5**			
	2020	2040	2060	2080	2020	2040	2060	2080
HadGEM2-ES	4.2E-02	2.6E-03	2.3E-03	3.6E-03	4.0E-02	1.1E-02	1.9E-02	6.1E-03
ACCESS1-0	3.9E-02	1.7E-02	6.9E-03	7.4E-04	4.3E-02	2.5E-02	9.4E-03	2.0E-02
CanESM2	4.4E-02	1.8E-02	4.8E-02	2.4E-02	4.1E-02	3.3E-02	4.3E-02	5.2E-02
CCSM4	4.3E-02	4.1E-03	1.1E-02	1.7E-03	4.2E-02	2.7E-02	8.0E-03	9.7E-03
CNRM-CM5	5.0E-02	2.2E-02	1.6E-02	3.0E-03	4.9E-02	1.2E-02	1.8E-02	2.4E-02
MPI-ESM-LR	4.0E-02	4.1E-03	1.2E-02	1.7E-03	5.2E-02	5.2E-03	3.2E-03	2.7E-03

Table S2. Table of interquartile ranges for each of the GCM and RCP scenarios for each of the four regression models.

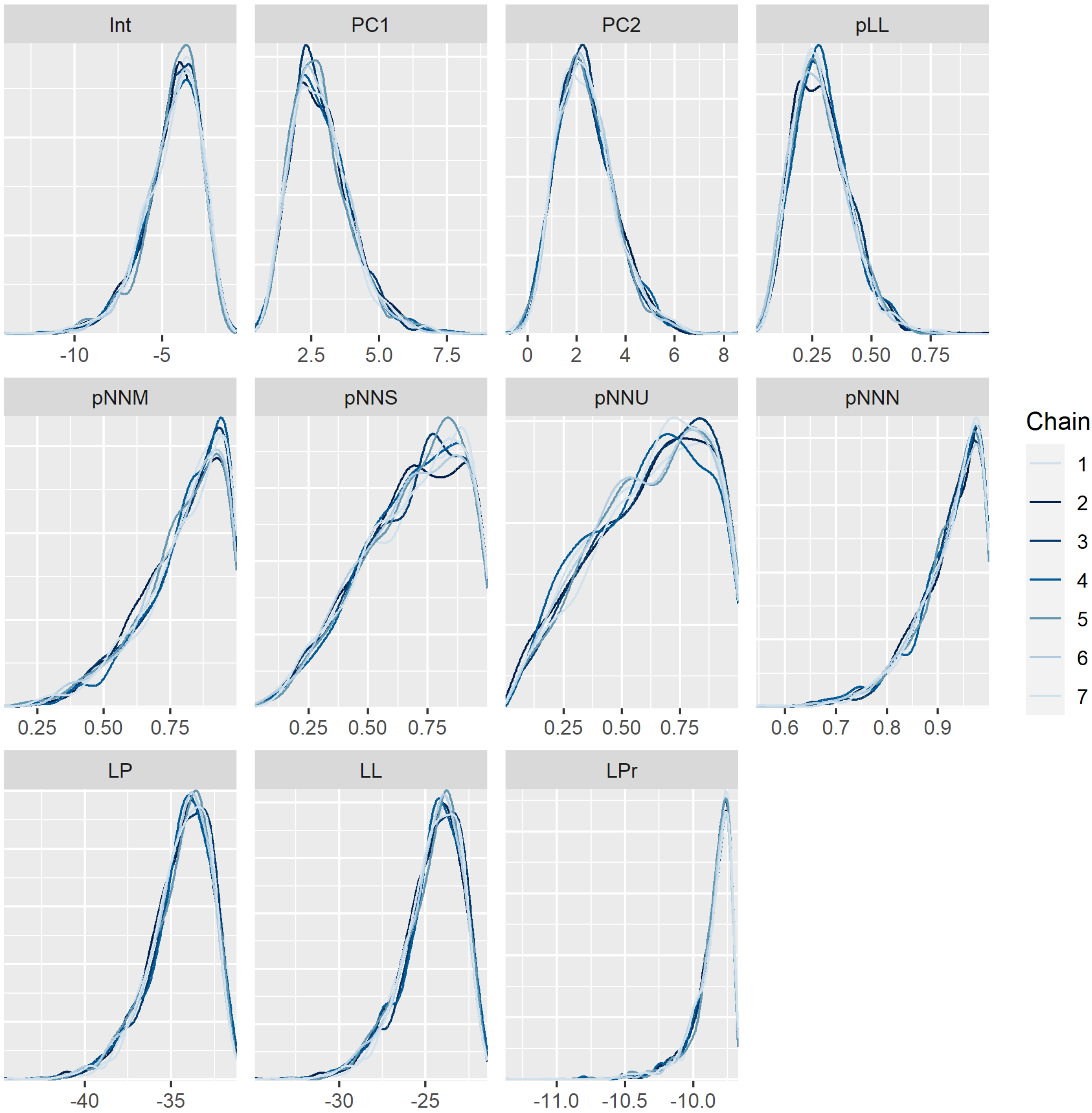

Figure S1. Marginal distributions for parameters and likelihoods in the model with historical data uncertainty considered. Int: $beta_0$ constant (intercept), PC: $beta_1$ and $beta_2$ for principal components, pLL: η probability of recording a large flood given that a large flood truly occurred, pNN#: θ_M probability of recording no large flood given that a large flood truly did not occur and moderate (M), small (S), unknown (U), or no flood (N) was the labeled category, LP: log posterior, LL: log likelihood, LPr: log prior.

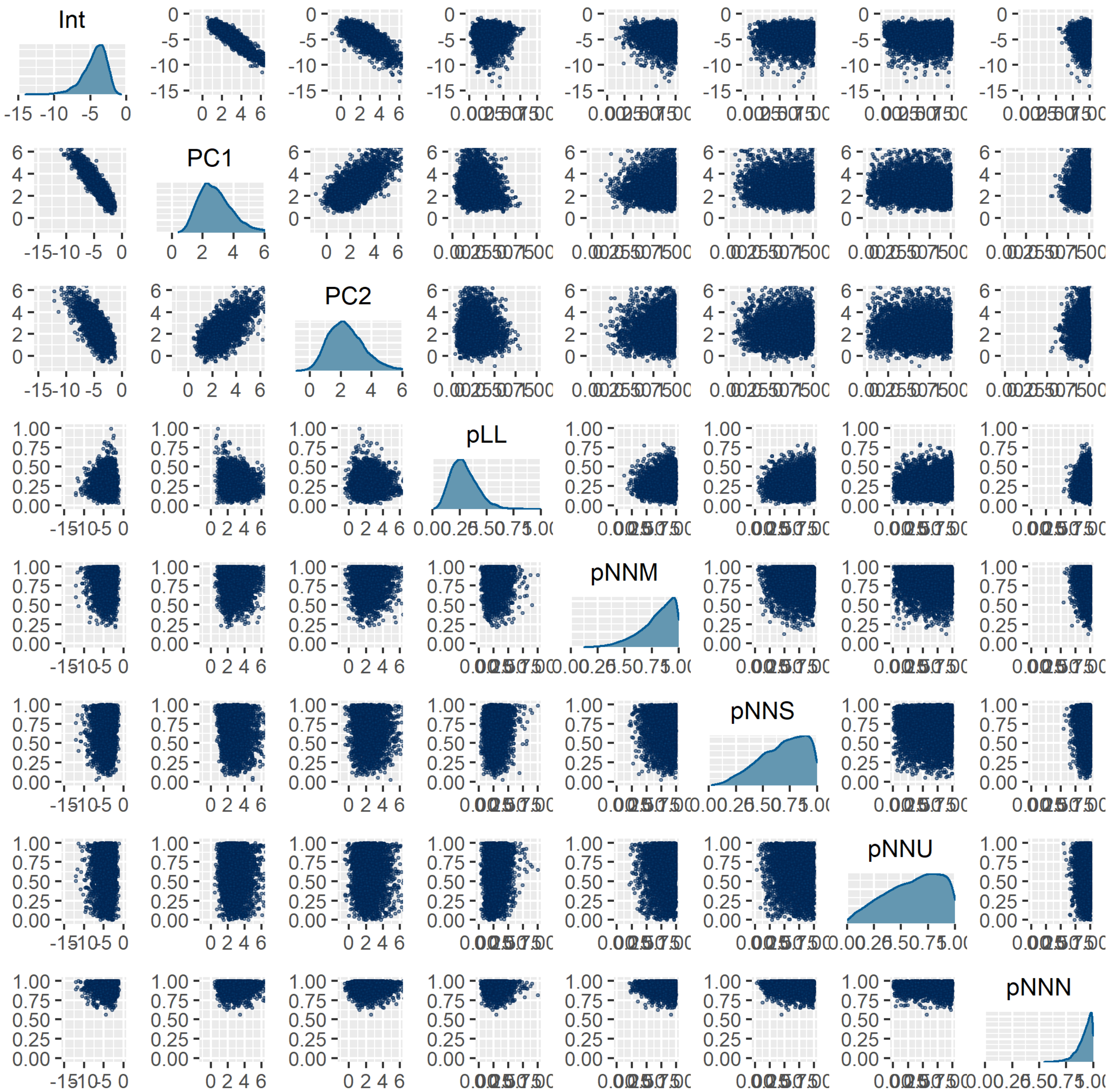

Figure S2. Scatterplot matrix and marginal distributions for parameters in the model with historical data uncertainty considered. Parameter names are the same as in Figure S1.

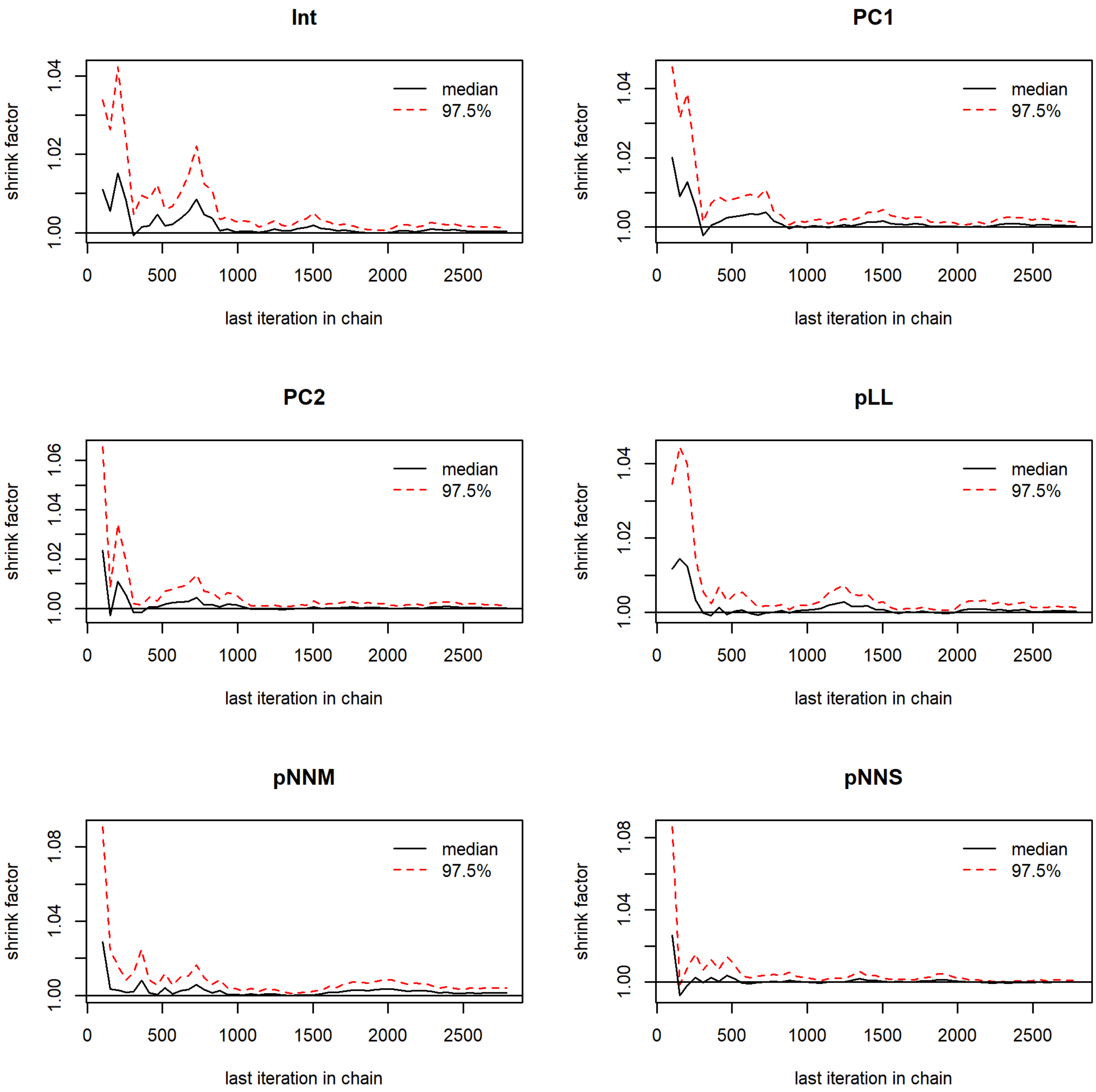

Figure S3. Gelman-Rubin shrink reduction for parameters in the model with historical data uncertainty considered. Parameter names are the same as in Figure S1.

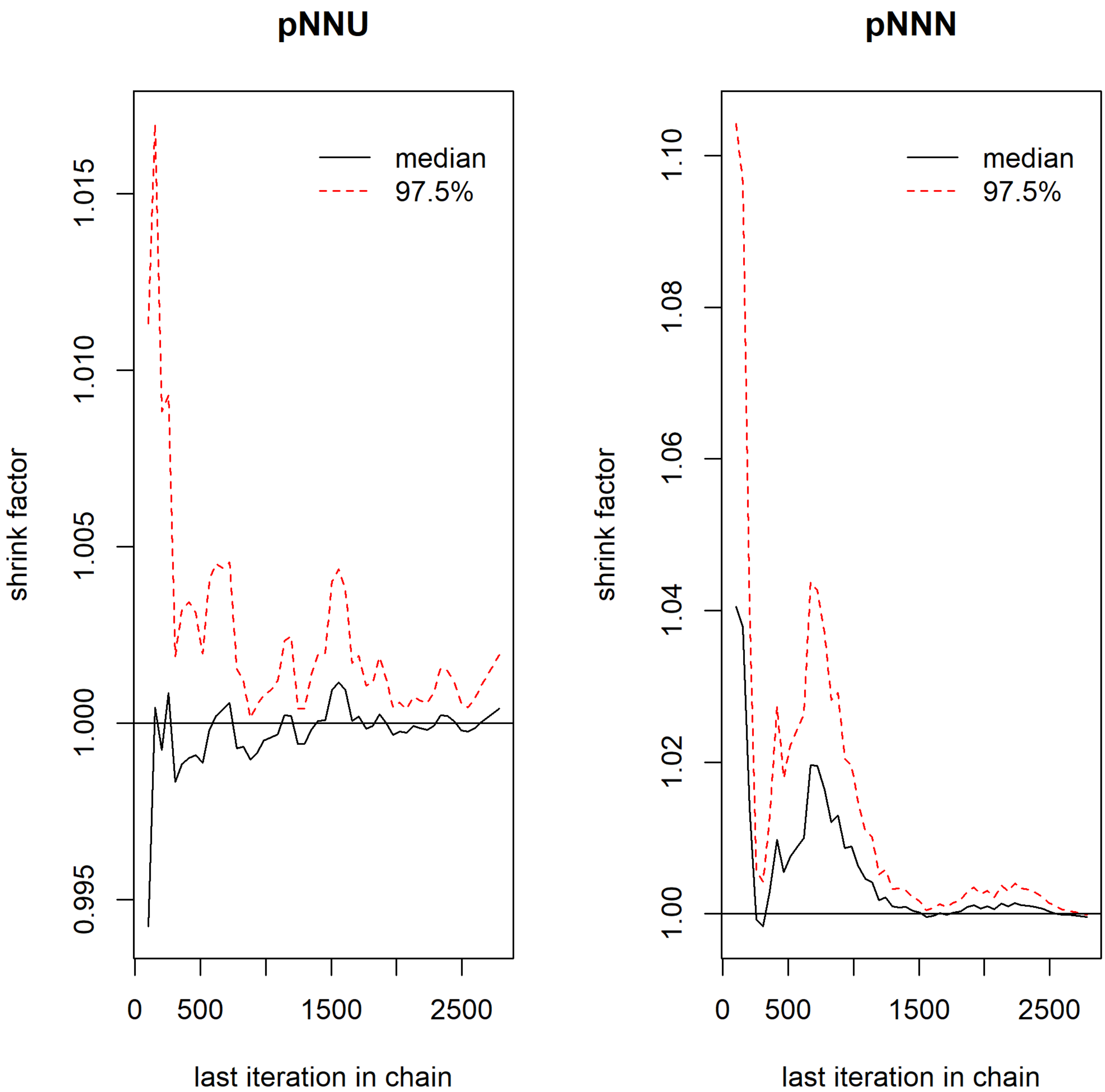

Figure S4. Gelman-Rubin shrink reduction for parameters in the model with historical data uncertainty considered. Parameter names are the same as in Figure S1.

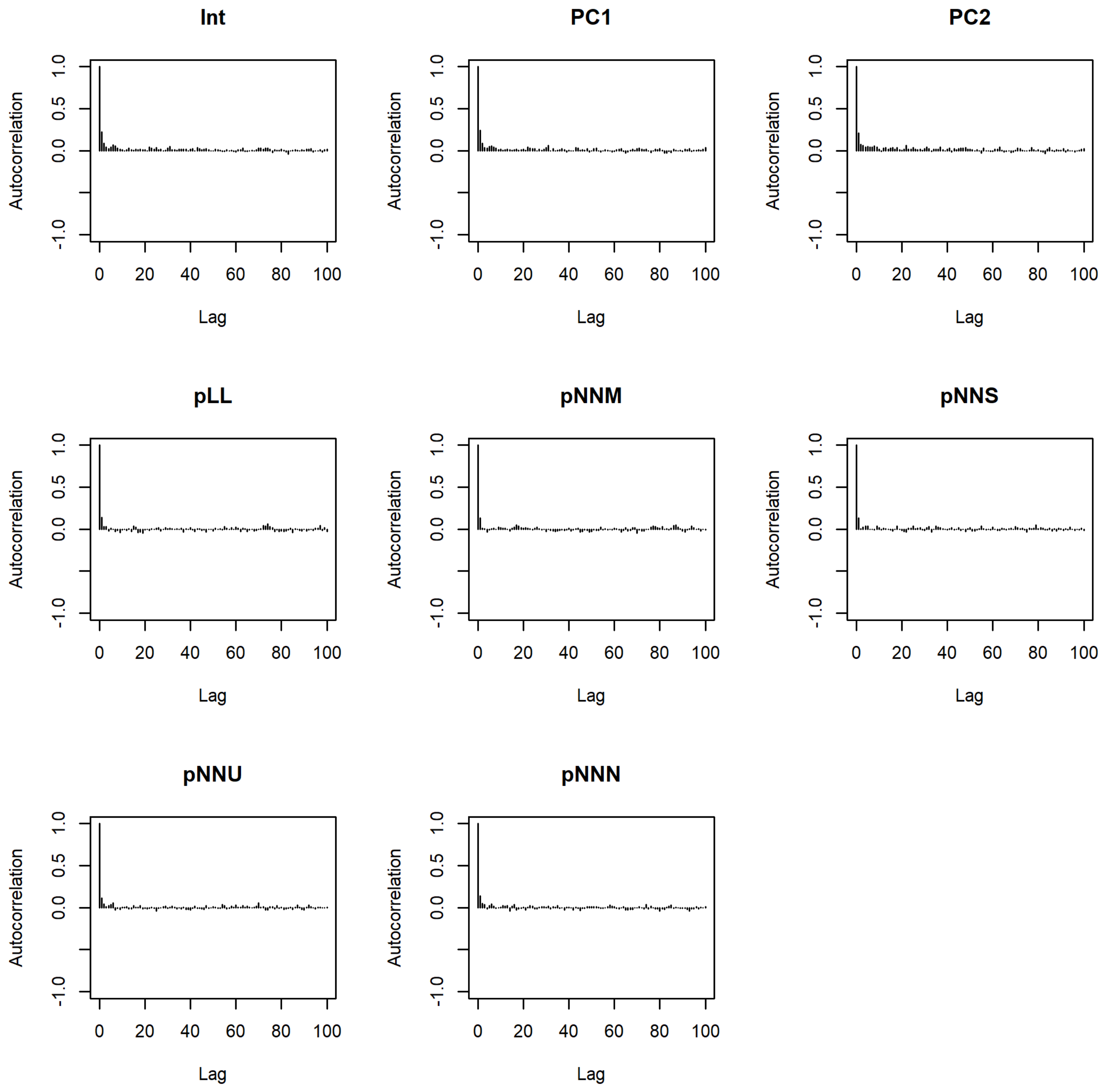

Figure S5. Autocorrelation in the selected MCMC chain samples for parameters in the model with historical data uncertainty considered. Parameter names are the same as in Figure S1.

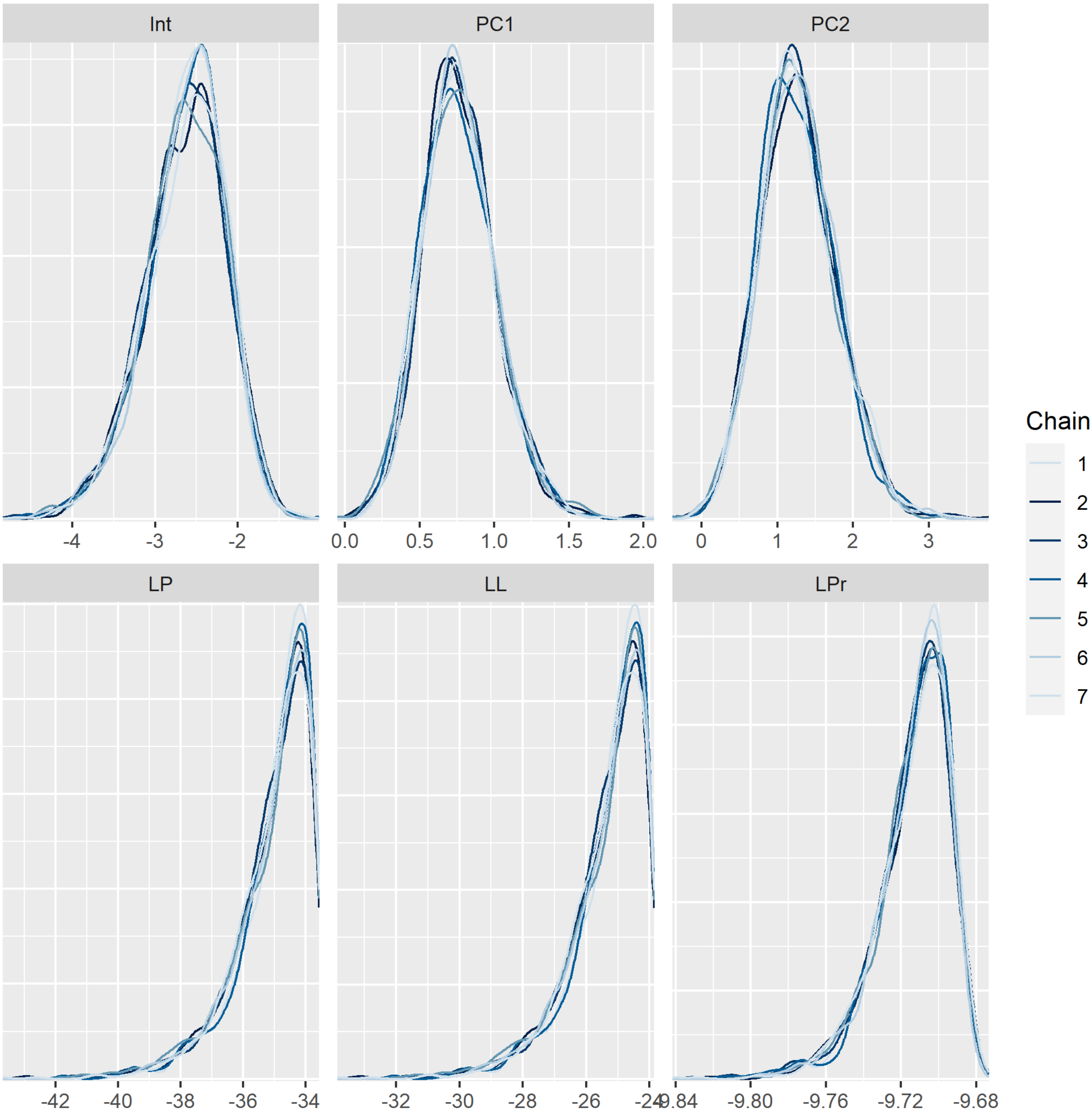

Figure S6. Marginal distributions for parameters and likelihoods in the model with no historical data uncertainty considered. Parameter names are the same as in Figure S1.

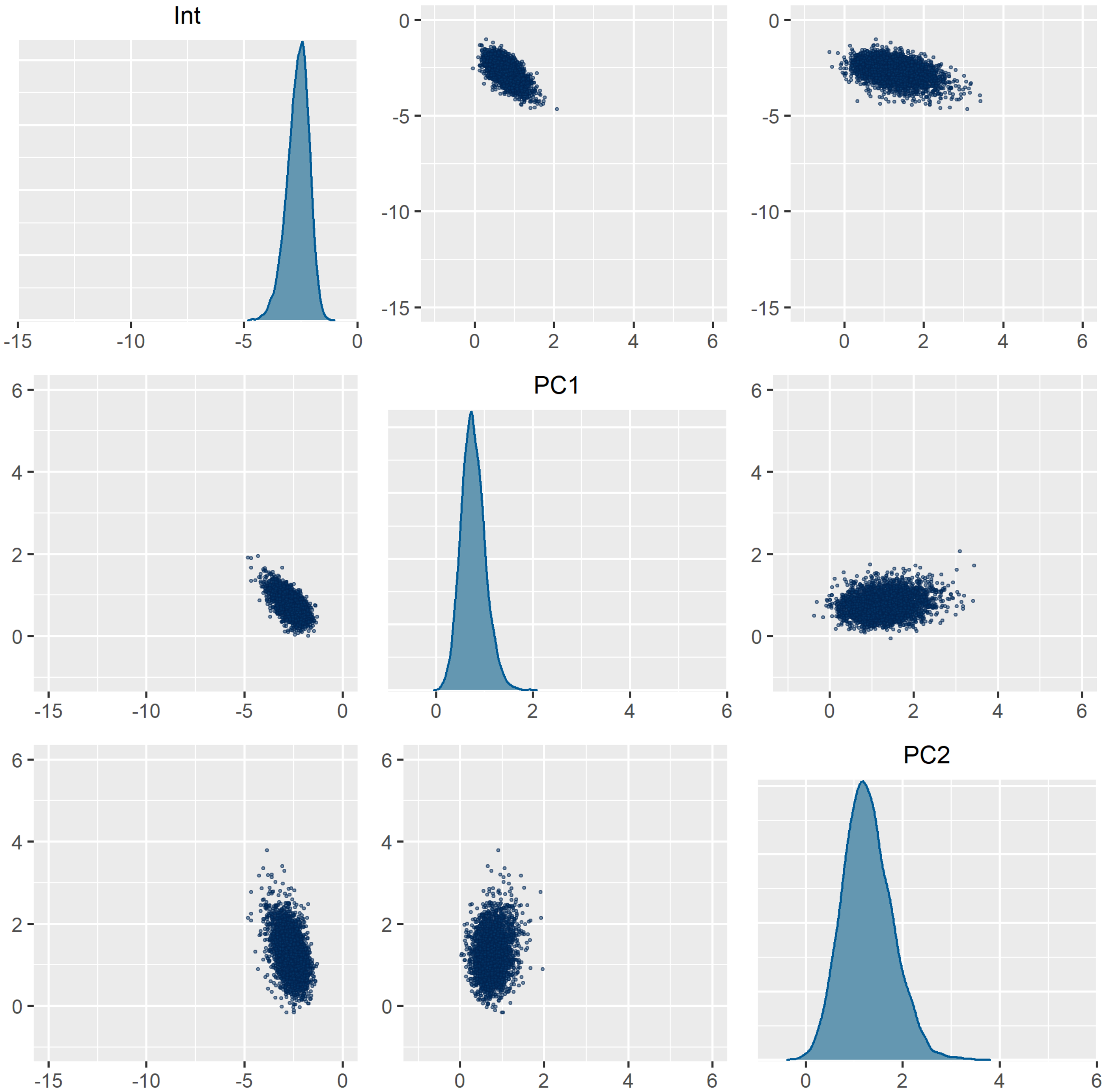

Figure S7. Scatterplot matrix and marginal distributions for parameters in the model with no historical data uncertainty considered. Parameter names are the same as in Figure S1.

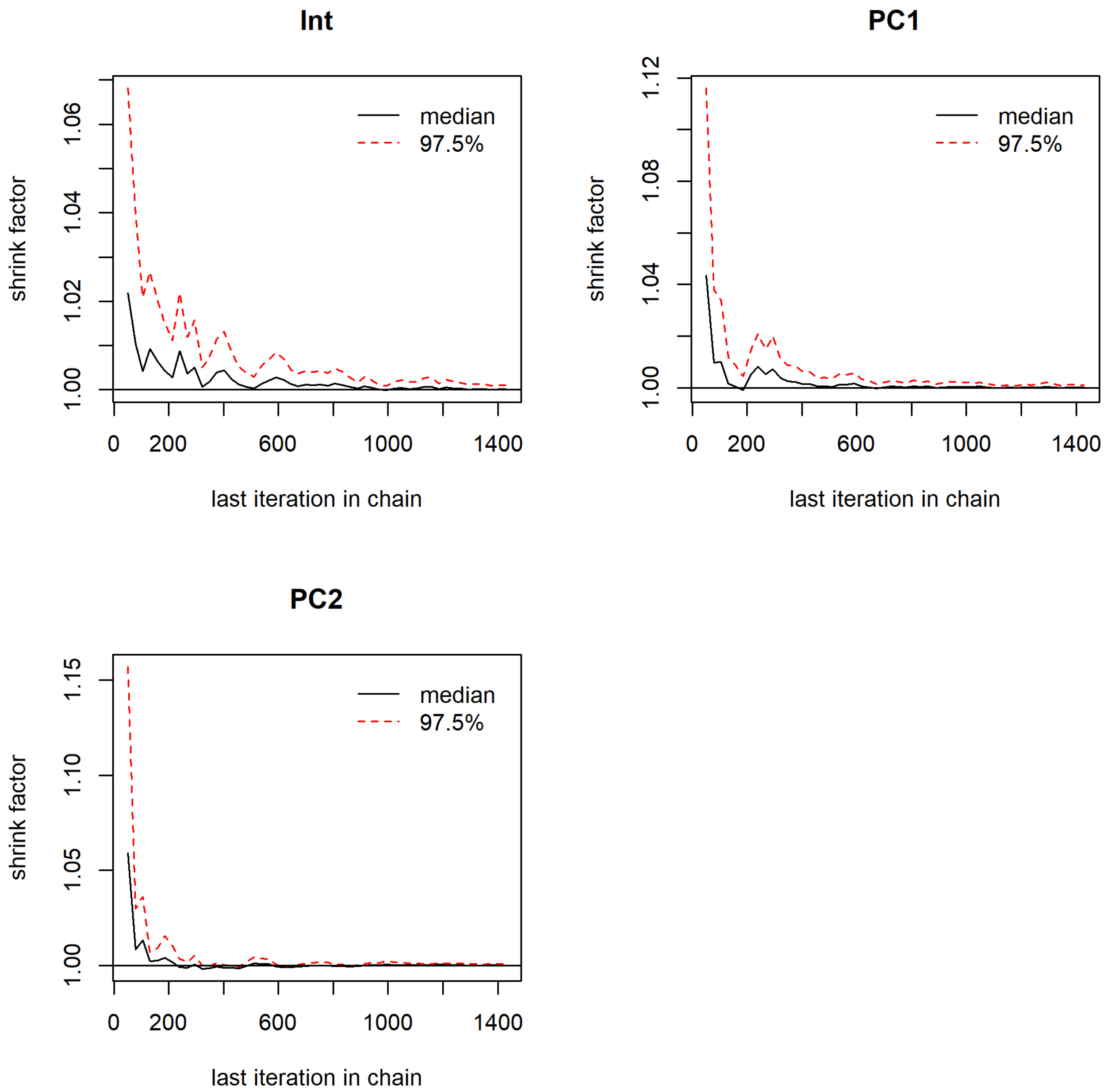

Figure S8. Gelman-Rubin shrink reduction for parameters in the model with no historical data uncertainty considered. Parameter names are the same as in Figure S1.

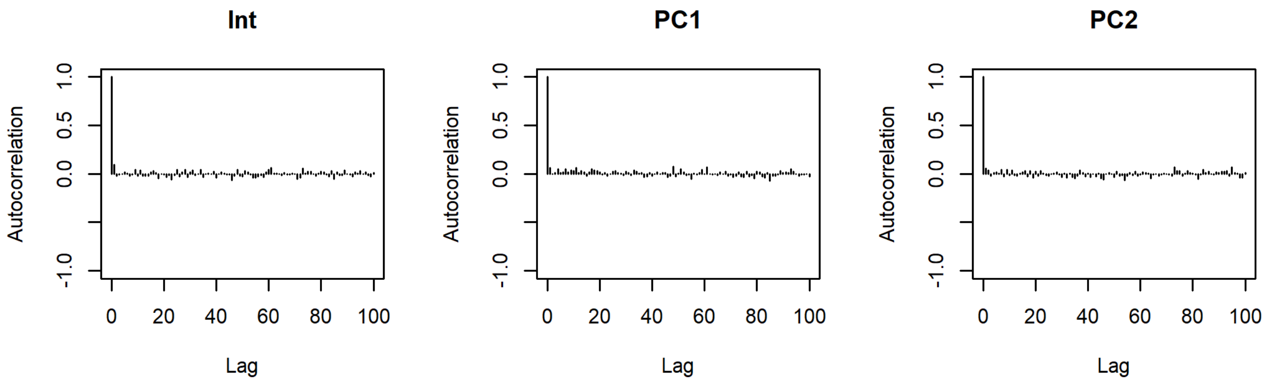

Figure S9. Autocorrelation in the selected MCMC chain samples for parameters in the model with no historical data uncertainty considered. Parameter names are the same as in Figure S1.

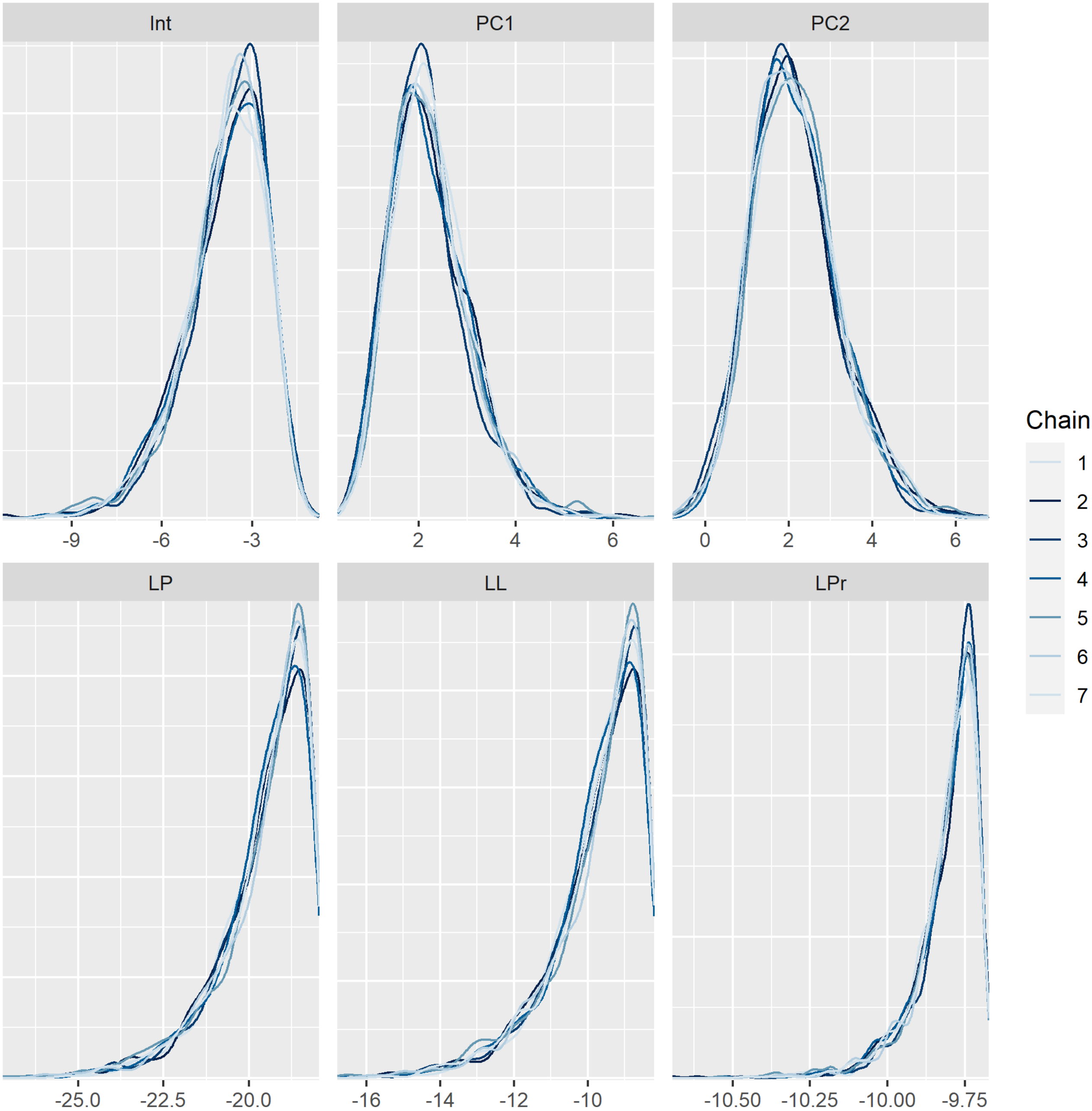

Figure S10. Marginal distributions for parameters and likelihoods for the best model with Fort Smith, trained on the 1962-2020 systematic record. Parameter names are the same as in Figure S1.

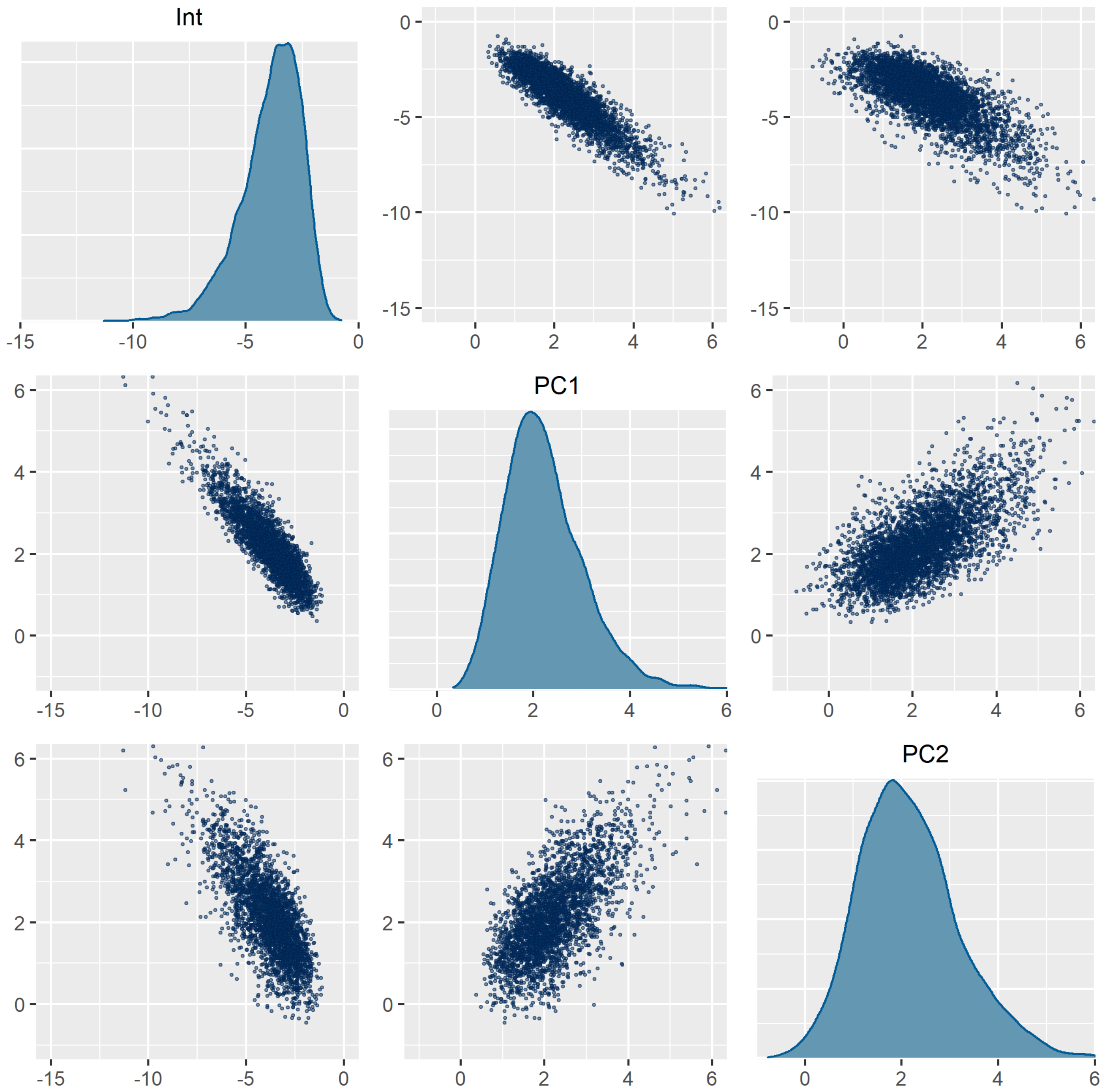

Figure S11. Scatterplot matrix and marginal distributions for parameters in the best model with Fort Smith, trained on the 1962-2020 systematic record. Parameter names are the same as in Figure S1.

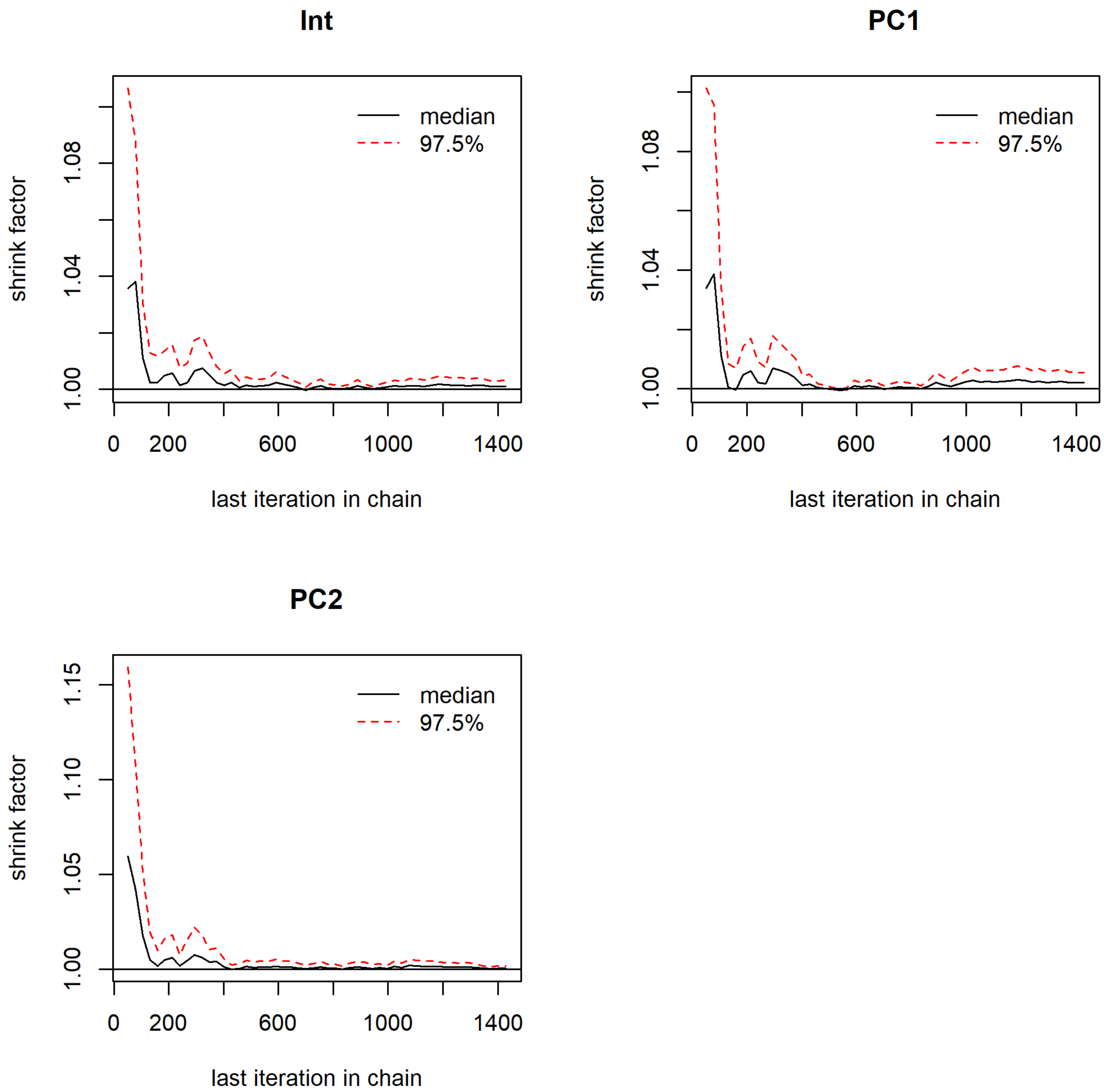

Figure S12. Gelman-Rubin shrink reduction for parameters in the best model with Fort Smith, trained on the 1962-2020 systematic record. Parameter names are the same as in Figure S1.

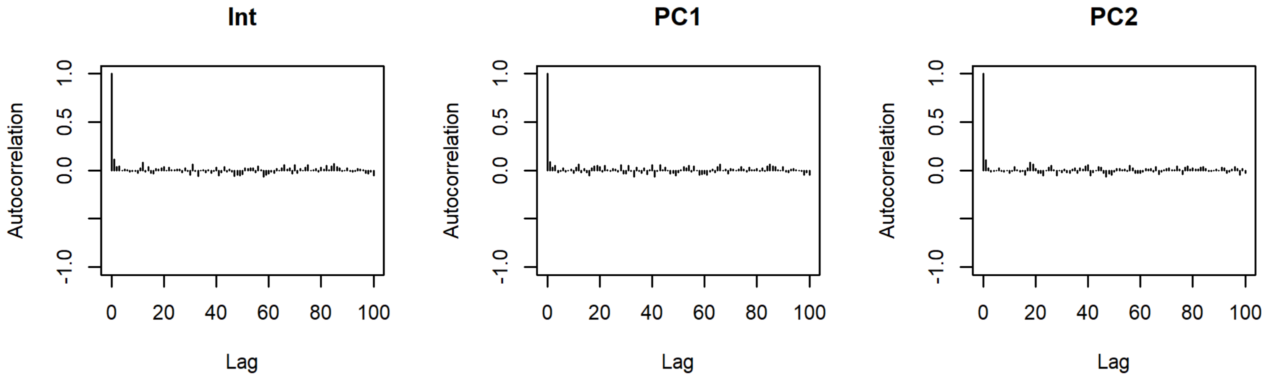

Figure S13. Autocorrelation in the selected MCMC chain samples for parameters in the best model with Fort Smith, trained on the 1962-2020 systematic record. Parameter names are the same as in Figure S1.

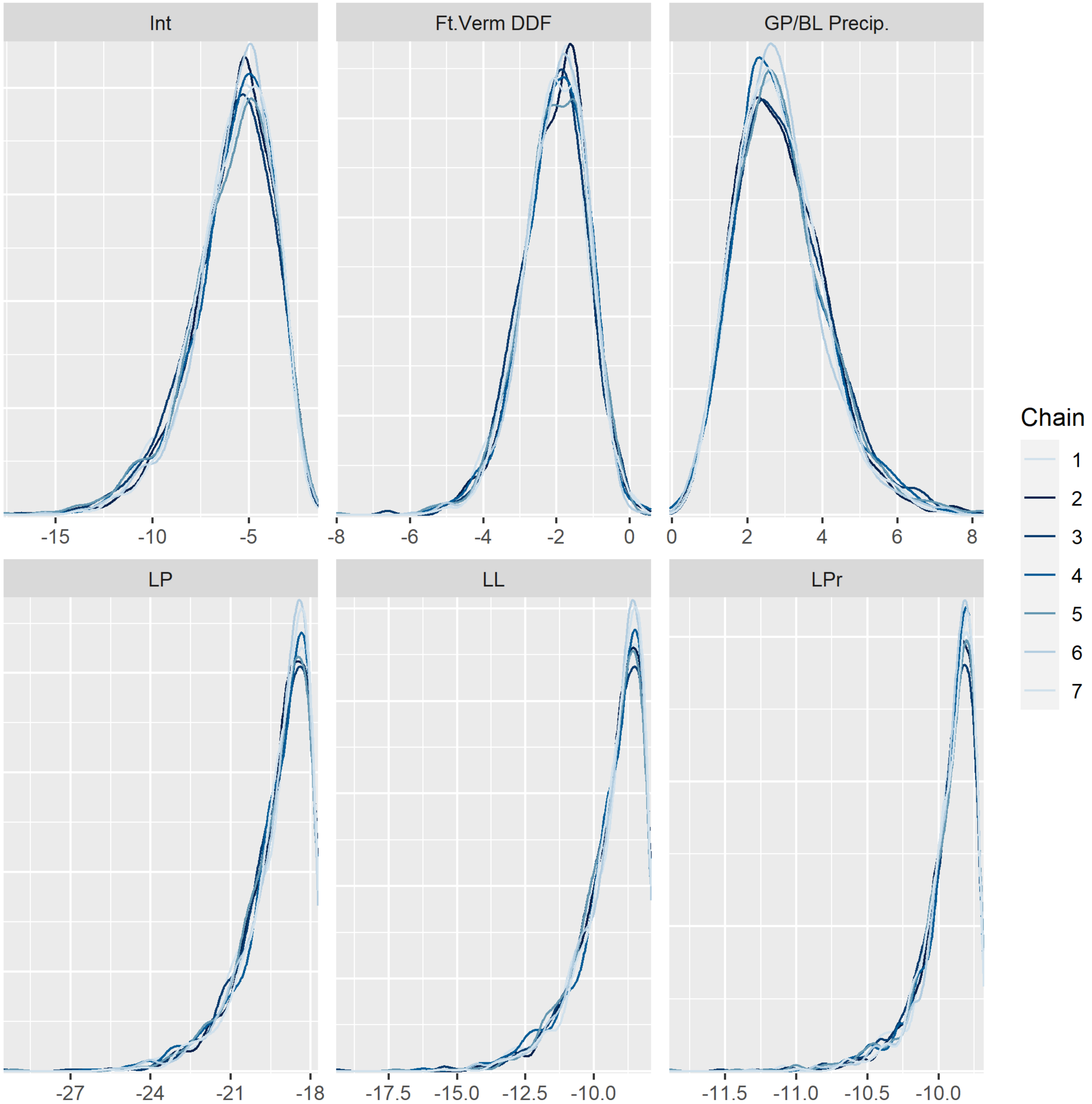

Figure S14. Marginal distributions for parameters and likelihoods for the Lamontagne et al. best model, trained on the 1962-2020 systematic record. Parameter names are the same as in Figure S1.

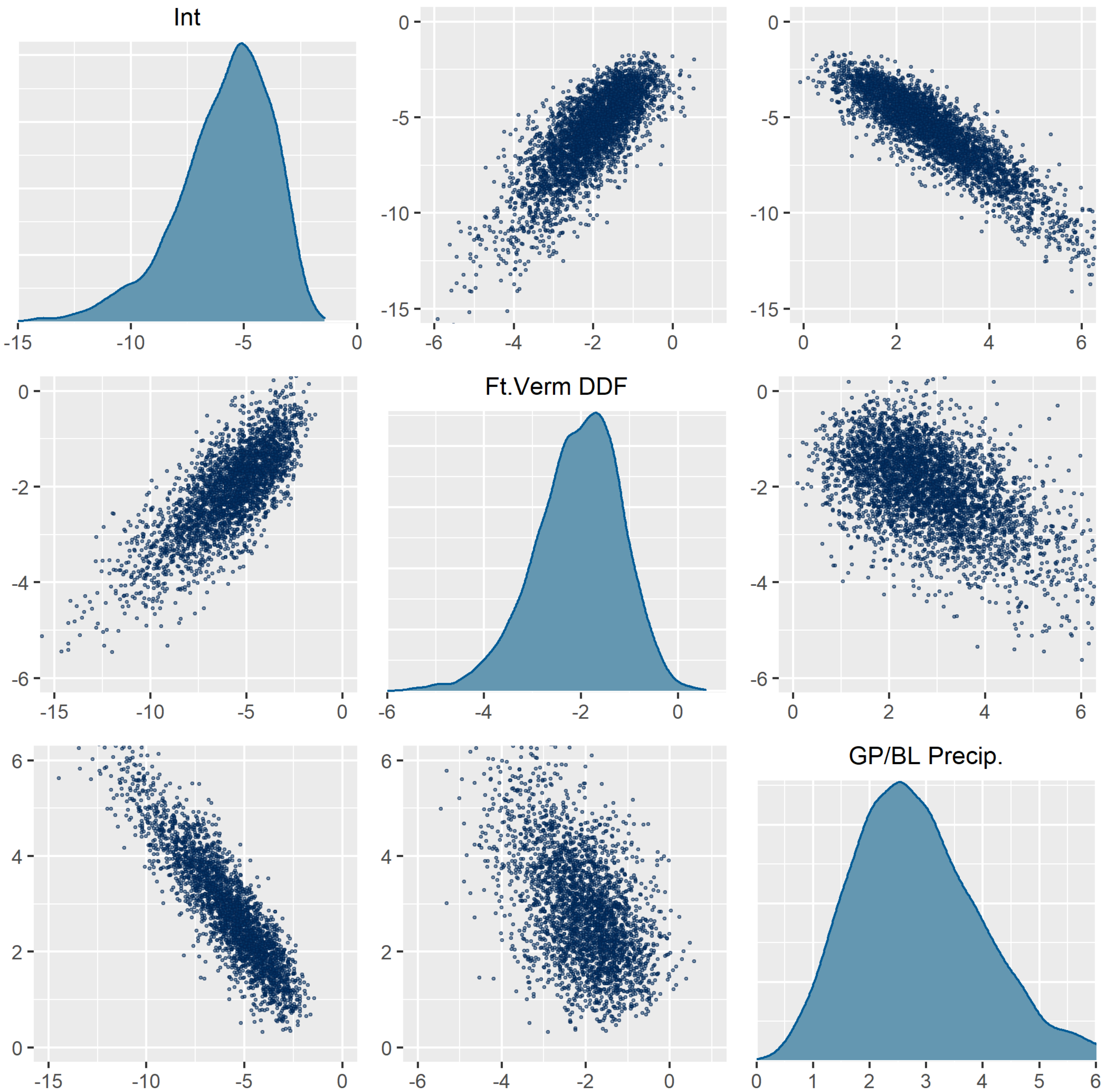

Figure S15. Scatterplot matrix and marginal distributions for parameters in the Lamontagne et al. best model, trained on the 1962-2020 systematic record. Parameter names are the same as in Figure S1.

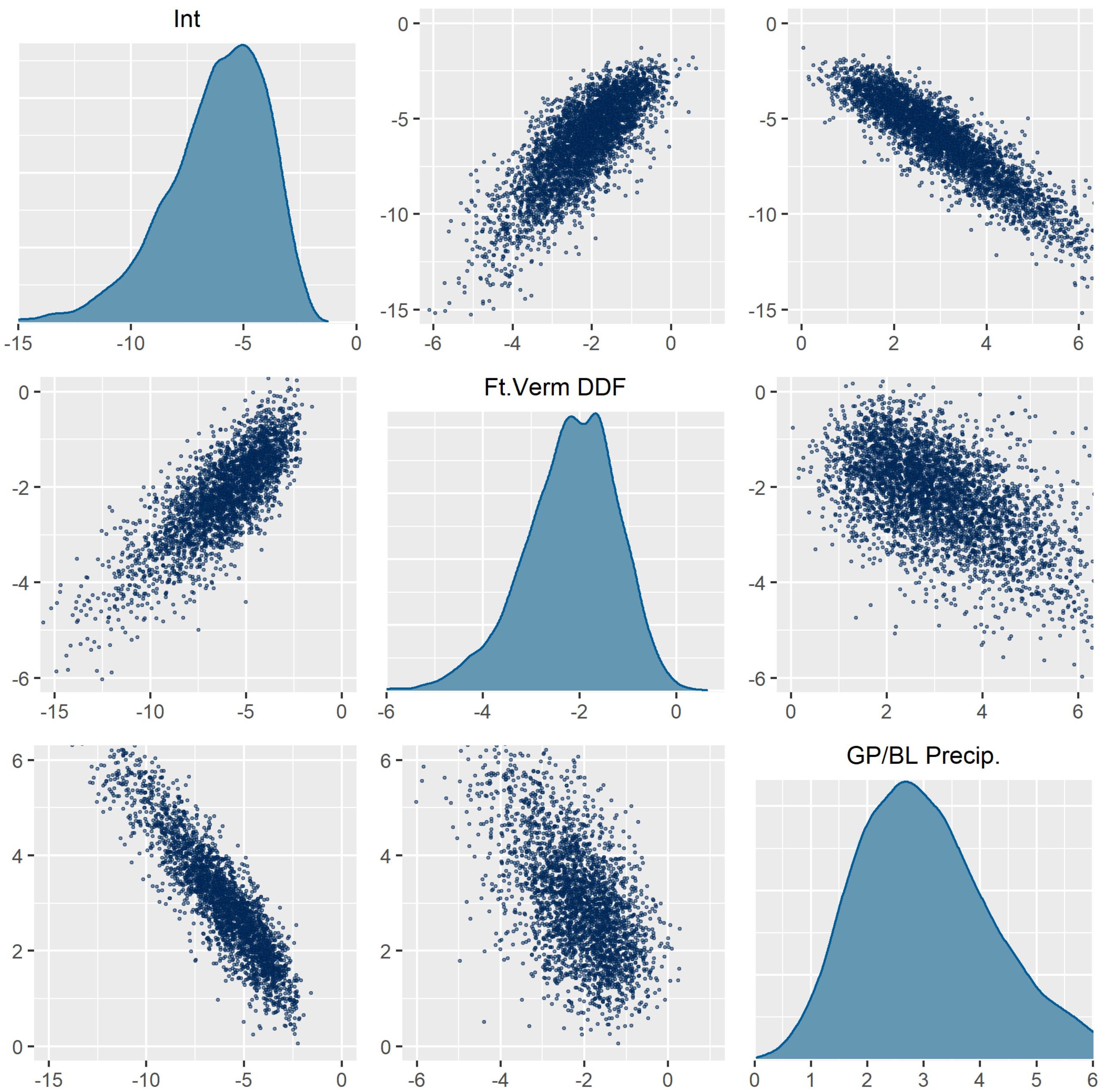

Figure S16. Scatterplot matrix and marginal distributions for parameters in the Lamontagne et al. best model, trained on the 1962-2020 systematic record using a multivariate normal prior for logistic regression model coefficients with mean equal to the estimated MAP values from Firth's logistic regression. Parameter names are the same as in Figure S1.

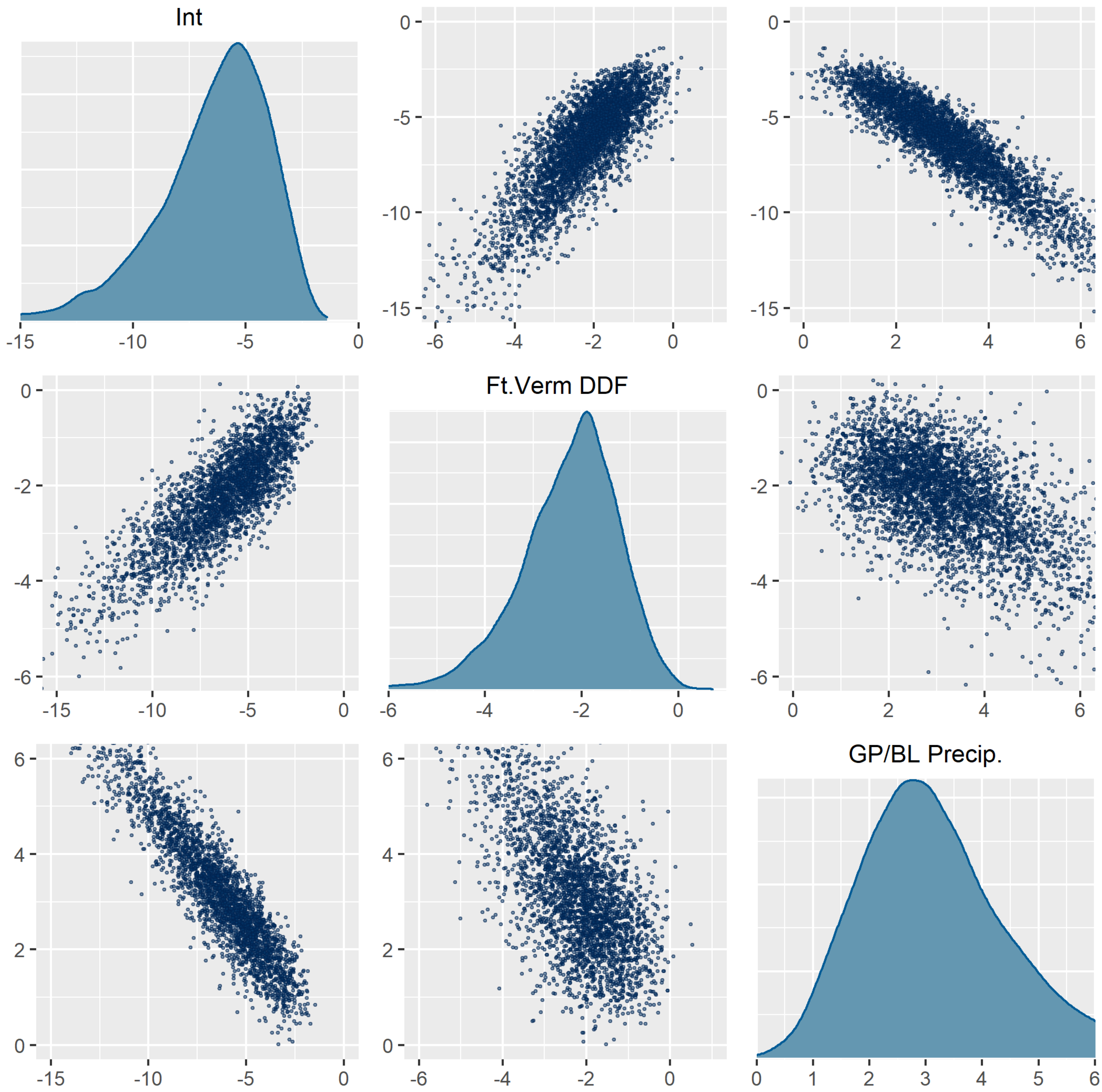

Figure S17. Scatterplot matrix and marginal distributions for parameters in the Lamontagne et al. best model, trained on the 1962-2020 systematic record using uniform priors for logistic regression model coefficients. Parameter names are the same as in Figure S1.

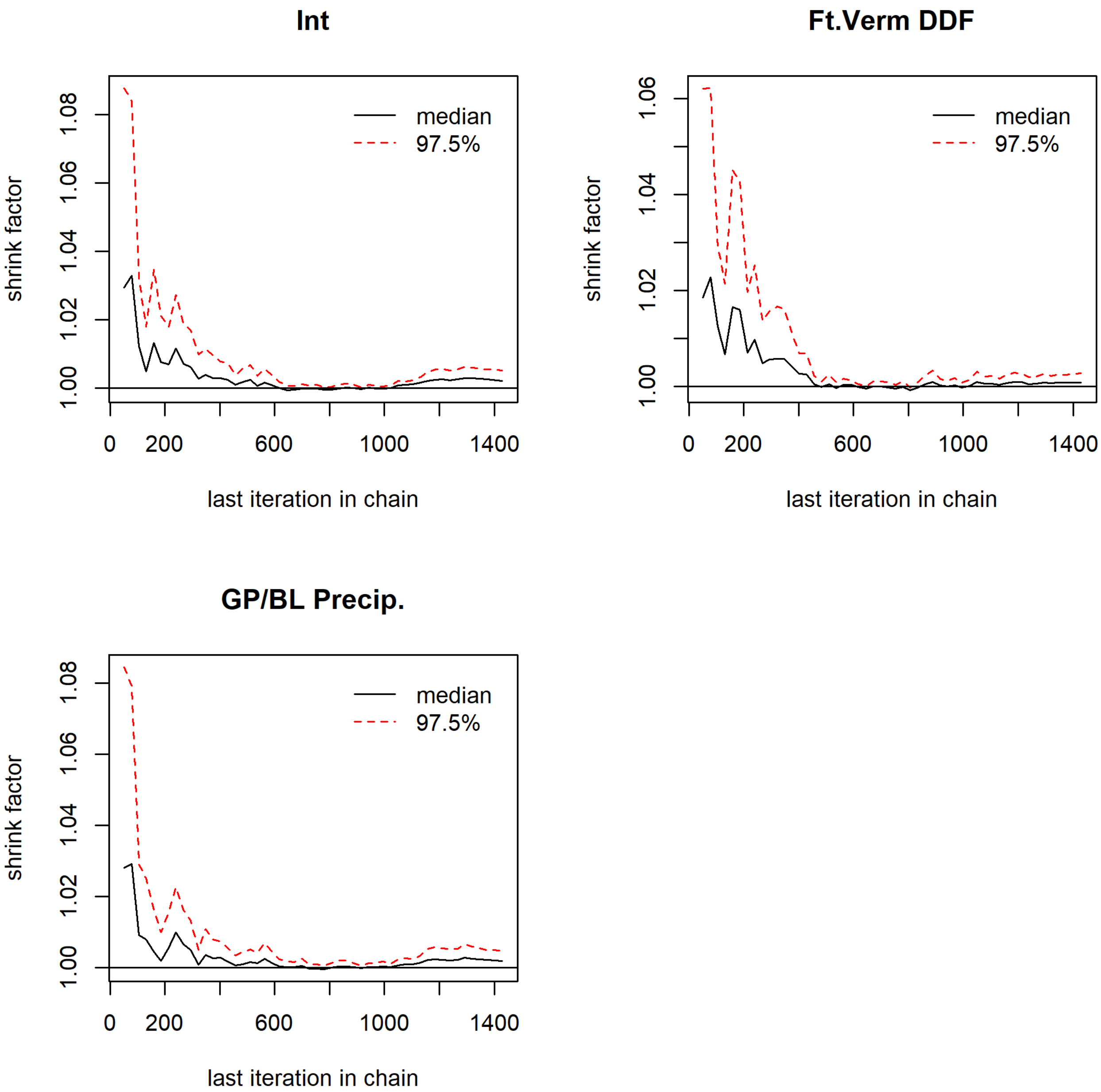

Figure S18. Gelman-Rubin shrink reduction for parameters in the Lamontagne et al. best model, trained on the 1962-2020 systematic record. Parameter names are the same as in Figure S1.

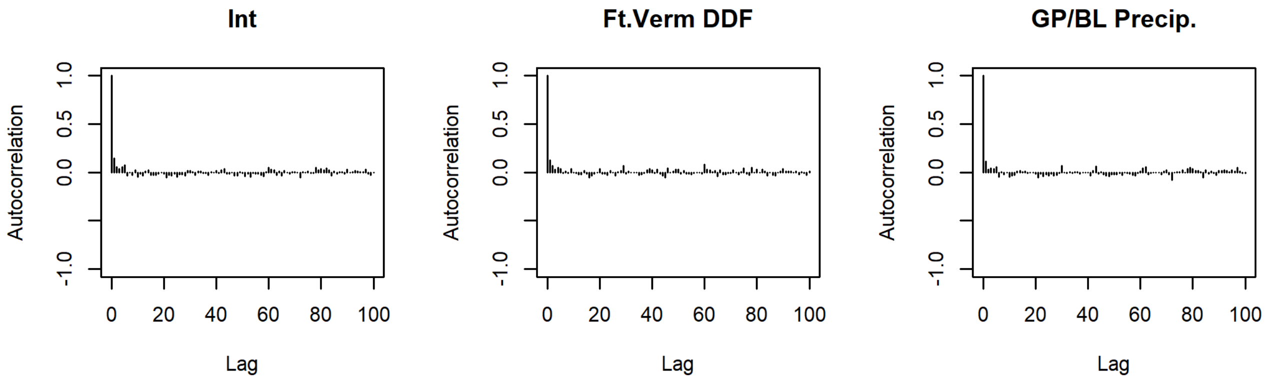

Figure S19. Autocorrelation in the selected MCMC chain samples for parameters in the Lamontagne et al. best model, trained on the 1962-2020 systematic record. Parameter names are the same as in Figure S1.

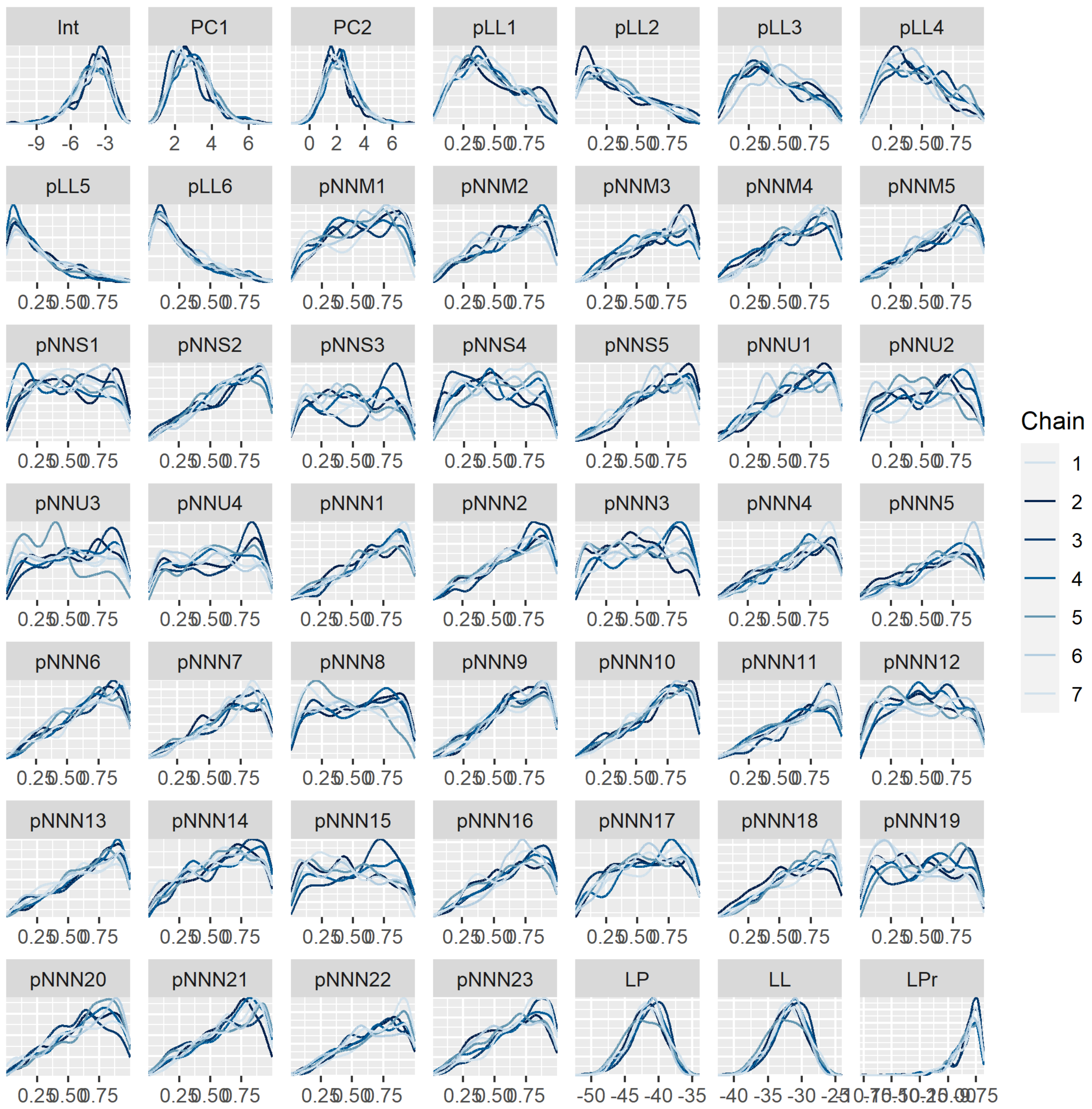

Figure S20. Marginal distributions for parameters and likelihoods for a model with a sensitivity or specificity parameter for each of the historical years of record. Parameter names are the same as in Figure S1, and numbers in each of these names match the numbers in Figures S21-S25.

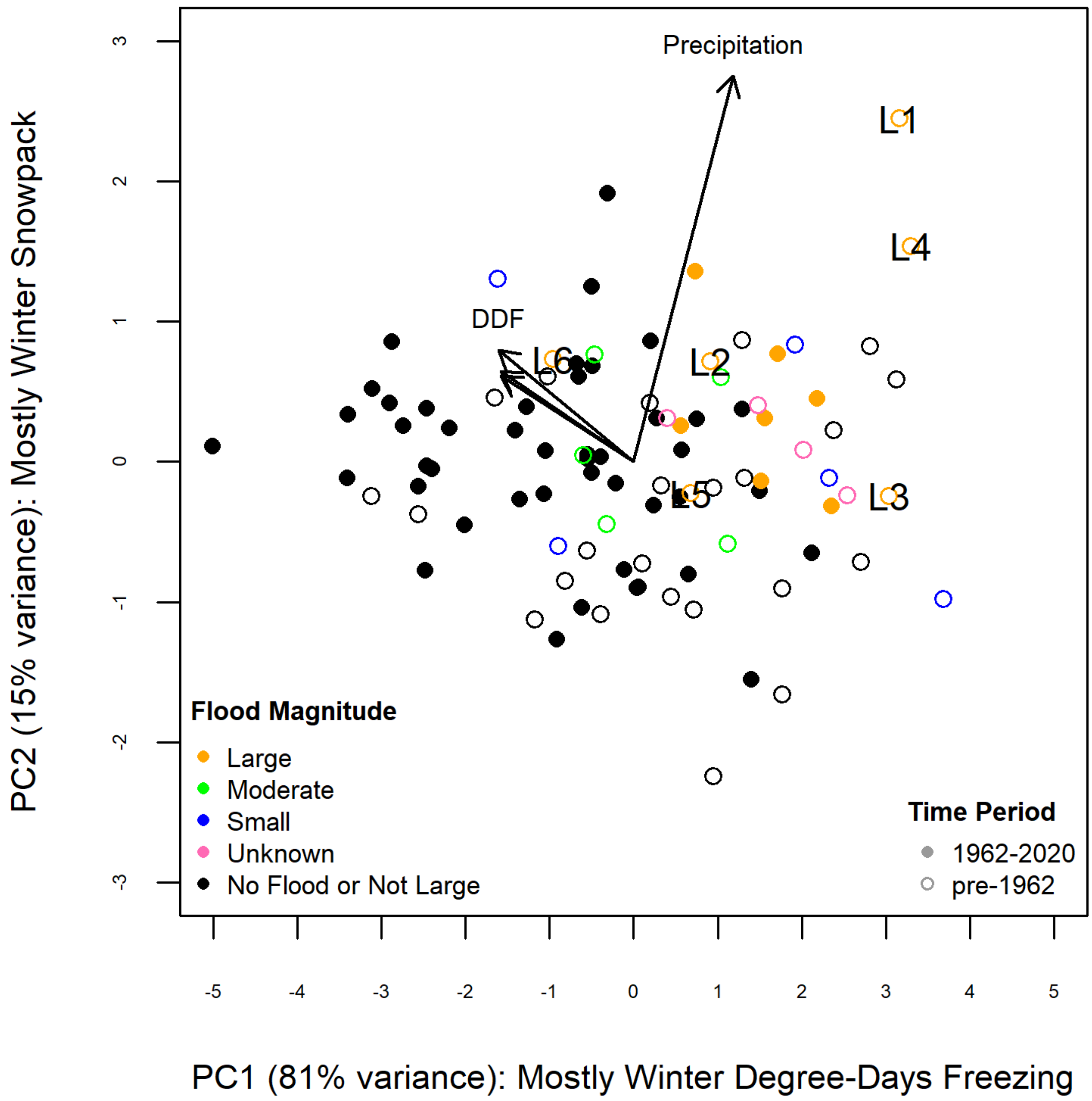

Figure S21. Flood magnitudes shown on each of the principal component (PC) axes that were used as predictor variables in the logistic regression. The directions of positive snowpack and degree-days freezing are shown for reference. Recorded large floods in the historical period are labeled with numbers that match Figure S20.

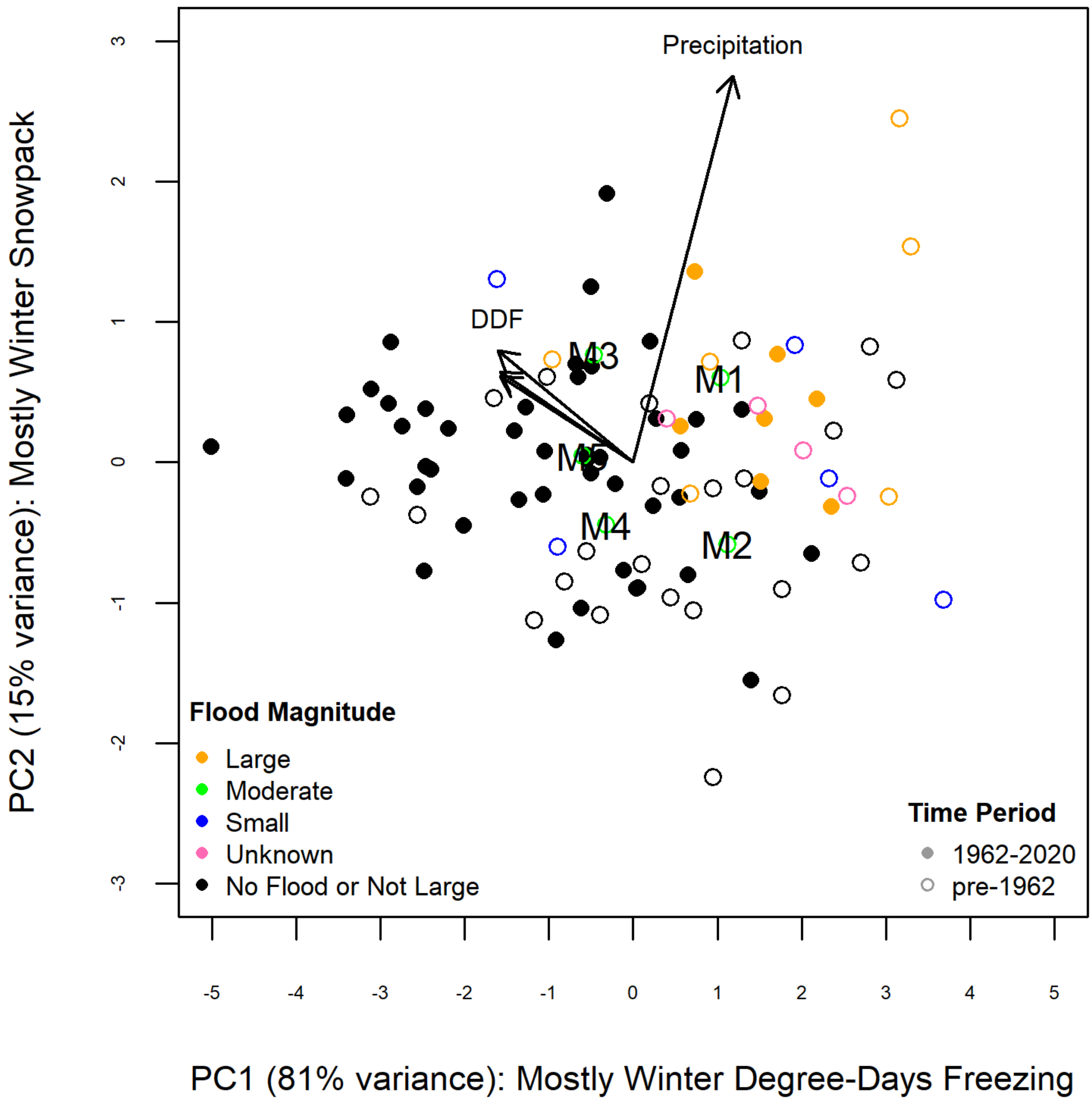

Figure S22. Flood magnitudes shown on each of the principal component (PC) axes that were used as predictor variables in the logistic regression. The directions of positive snowpack and degree-days freezing are shown for reference. Recorded moderate floods in the historical period are labeled with numbers that match Figure S20.

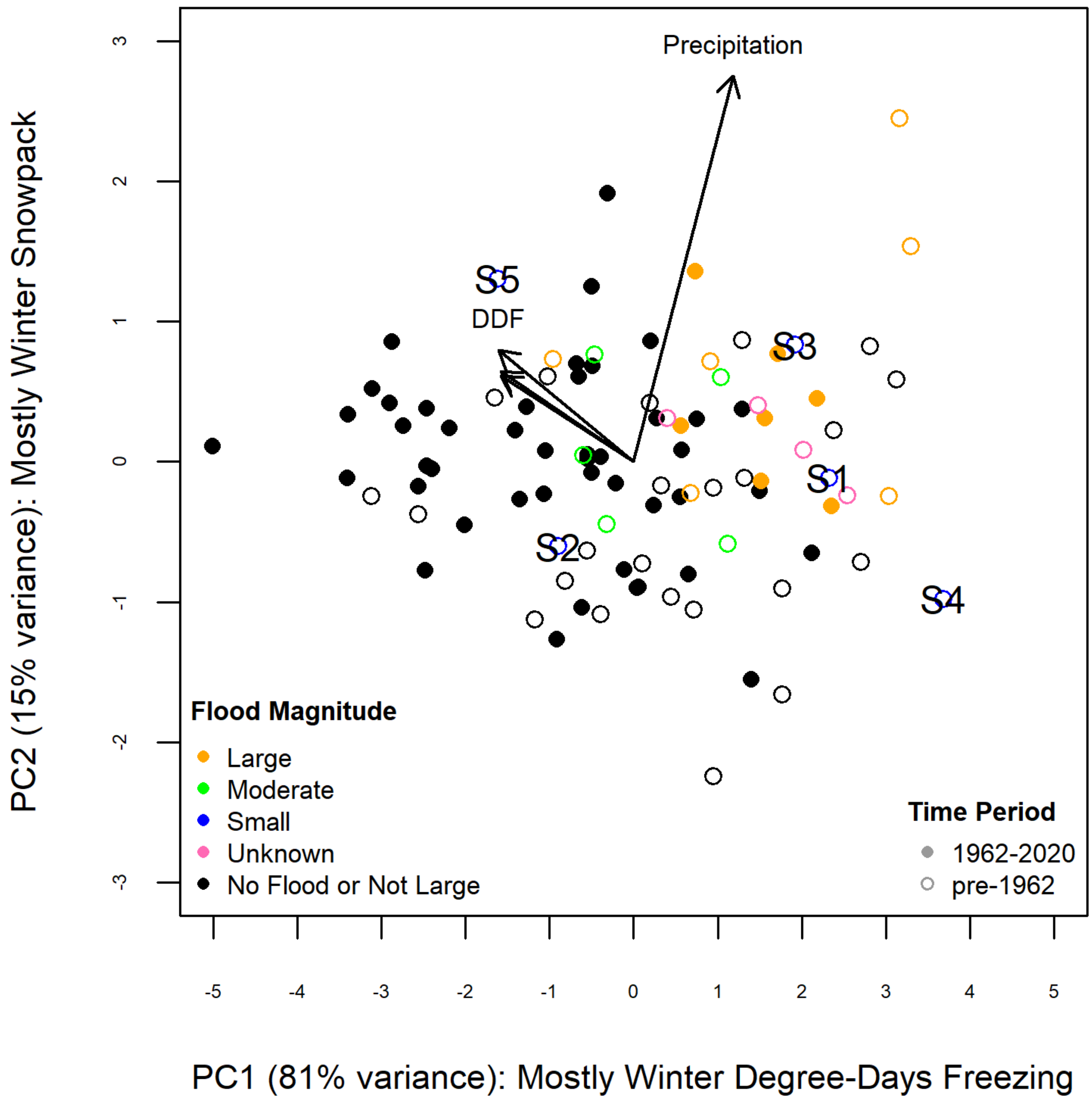

Figure S23. Flood magnitudes shown on each of the principal component (PC) axes that were used as predictor variables in the logistic regression. The directions of positive snowpack and degree-days freezing are shown for reference. Recorded small floods in the historical period are labeled with numbers that match Figure S20.

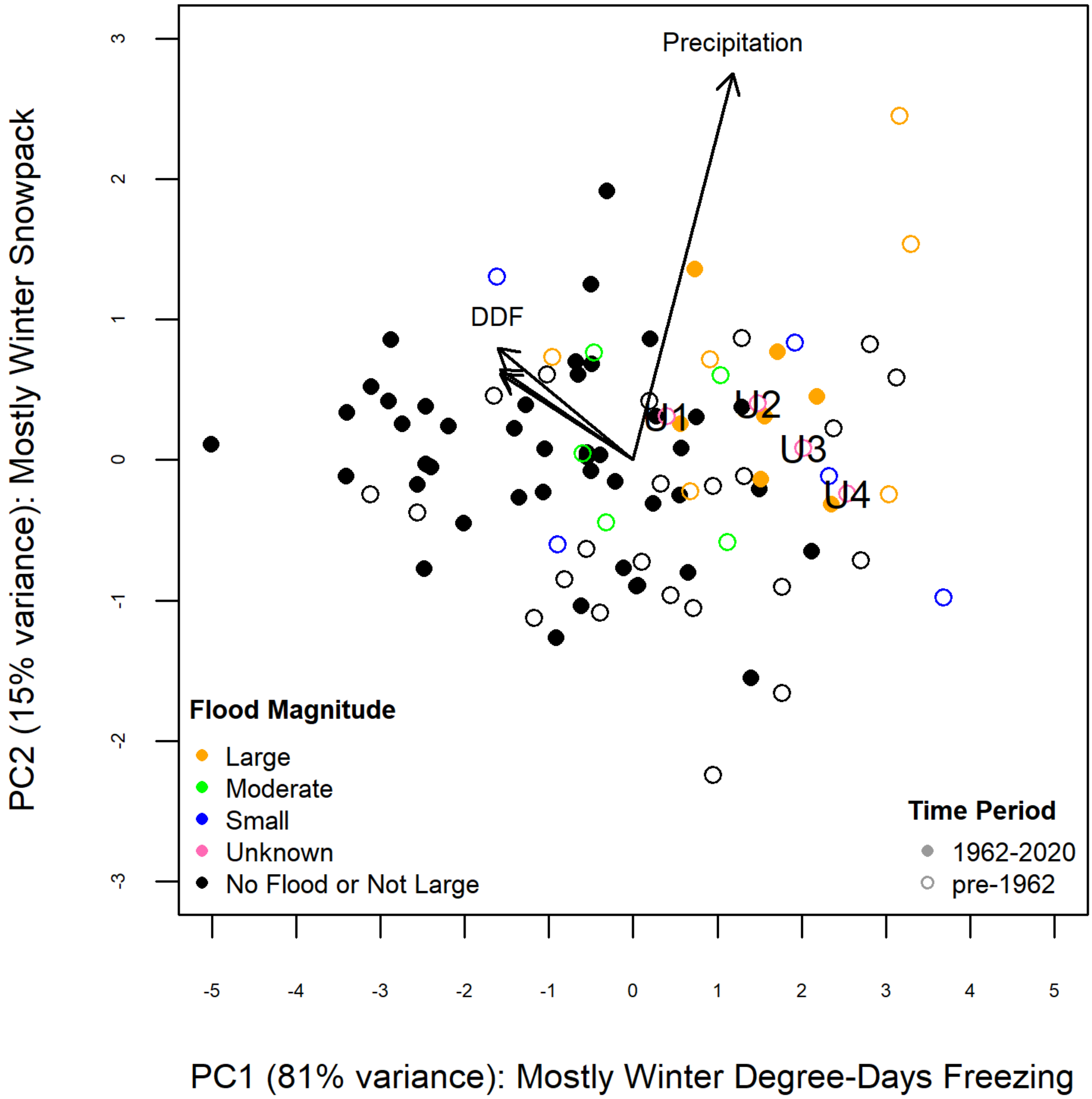

Figure S24. Flood magnitudes shown on each of the principal component (PC) axes that were used as predictor variables in the logistic regression. The directions of positive snowpack and degree-days freezing are shown for reference. Recorded unknown magnitude floods in the historical record are labeled with numbers that match Figure S20.

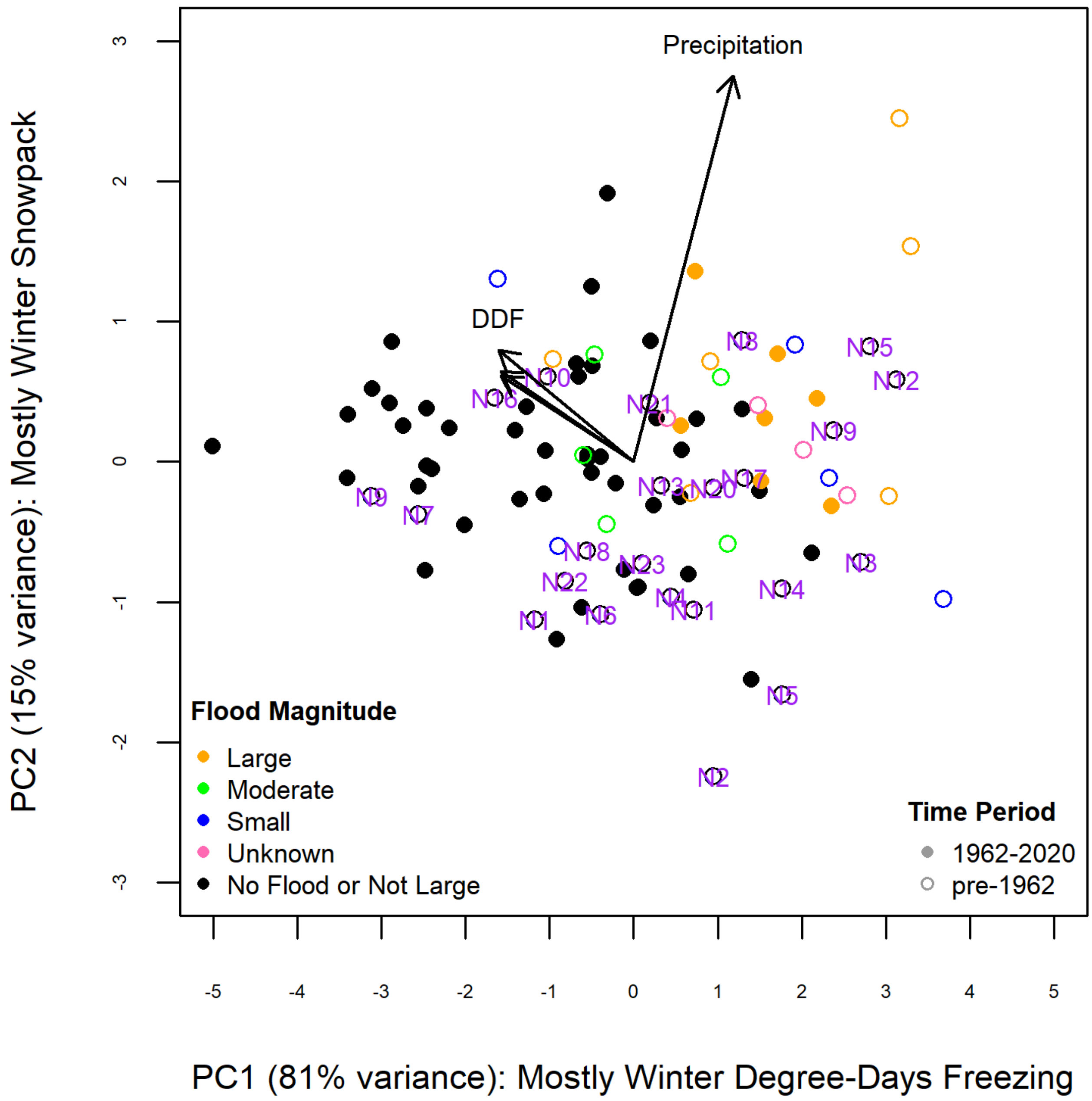

Figure S25. Flood magnitudes shown on each of the principal component (PC) axes that were used as predictor variables in the logistic regression. The directions of positive snowpack and degree-days freezing are shown for reference. Recorded years without floods in the historical record are labeled with numbers that match Figure S20.

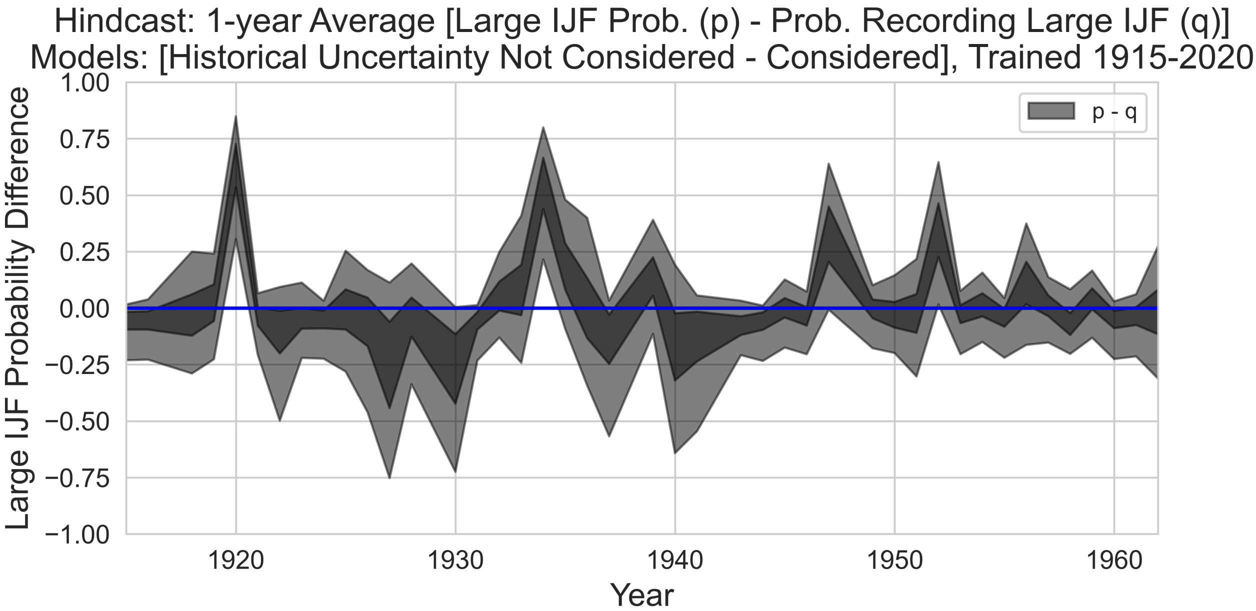

Figure S26. Difference between the predicted probability of a large ice jam flood using the model that does not consider historical data uncertainty, and the predicted probability of recording a large ice jam flood using the model that does consider historical data uncertainty. Significant positive differences occur only when p is close to 1, and q is unable to reach exactly 1 because because we use a single η parameter to describe all large floods.