University of British Columbia
Department of Statistics
Technical Report #256
April 2010

"Selecting a binary Markov model for a precipitation process"

by

Reza Hosseini[1], Nhu D Le[2] and James V Zidek[1]
[1] University of British Columbia
[2] British Columbia Cancer Agency & University of British Columbia

---

Department of Statistics, University of British Columbia

ABSTRACT

This paper uses rth-order categorical Markov chains to model the probability of precipitation. Several stationary and non-stationary high order Markov models are proposed and compared using BIC. The number of parameters increases exponentially by adding the Markov order. Several classes of high-order Markov models are proposed which their increase of number of parameters are modest. The theory of partial likelihood is used to estimate the parameters.

Keywords: binary Markov processes; precipitation model; Calgary precipitation; Markov model selection.

1 Introduction.

This paper studies the Markov order of the 0-1 precipitation process (PN from now on). Its genesis lies in agroclimate risk management and references to related work are given below. The paper develops a modeling strategy that can be used for such things as calculating the likelihood of a long sequence on non-precipitation days (drought). The same approach can be used for other climatological events such as extreme temperatures, as shown in Chapter 10 of [9]. Likelihoods that can be calculated with this approach can play a role in setting crop insurance premiums and managing irrigation programs, which are attaining increasing importance as the climate changes.

Many authors such as Anderson et al. in [2] and Barlett in [3] have developed techniques to test different assumptions about the order of the Markov chain. For example in [2], Anderson et al. develop a Chi-squared test to test that a Markov chain is of a given order against a larger order. In particular, with it we can test the hypothesis that a chain is 0th-order Markov against a 1st-order Markov chain, which in this case is testing independence against the usual (1st-order) Markov assumption. (This reduces simply to the well-known Pearson's Chi-squared test.) Hence, to "choose" the Markov order one might follow a strategy of testing 0th-order against 1st-order, testing 1st-order against 2nd-order and so on to rth-order against $(r + 1)$th-order, until the test rejects the null hypothesis and then choose the last r as the optimal order. However, some drawbacks are immediately seen with this method. To begin with, the choice of the significance level will affect our chosen order. Moreover, the method only works for chains with several independent observations of the same finite chain, a requirement that is certainly not met where spatially correlated series are observed at multiple sites. Finally this theory does not accommodate other explanatory variables such as for example the maximum daily temperature.

Issues like this have led researchers to think about other methods of order selection. Akaike in [1], using the information distance and Schwartz in [16] using Bayesian methods develop the AIC and BIC, respectively. Other methods and generalizations of the above methods have been proposed by some authors such as Hannan in [7], Shibita in [17] and Haughton in [8].

Many authors have studied the order of precipitation processes at different locations on Earth. Gabriel et al. in [6] use the test developed in Anderson et al. [2] to show that the precipitation in Tel-aviv is a 1st-order Markov chain. Tong in [18] used the AIC for Hong Kong, Honolulu and New York and showed that the process is 1st-order in Hong Kong and Honolulu but 0th-order in New York. In a later paper, [19], Tong and Gates use the same techniques for Manchester and Liverpool in England and also re-examined the Tel-aviv data. Chin in [4] studies the problem using AIC over 100 stations (separately) in the United States over 25 years. He concludes that the order depends on the season and geographical location.

Moreover, he finds a prevalence of first order conditional dependence in summer and higher orders in winter. Other studies have been done by several authors using similar techniques over other locations. For example, Moon et al. in [11] study this issue at 14 location in South Korea.

This paper investigates the Markov order for a cold-climate region. The Markov order of the precipitation in this region might be different due to a large fraction of precipitation being being in the form of snowfall. The paper also drops the homogeneity (stationarity) condition usually imposed in studying the Markov order. In fact the model proposed here can accommodate both continuous (here time and potentially geographical location and other explanatory variables) and categorical variables (e.g. precipitation occurred/not occurred on a given day).

An issue with increasing the order of a Markov chain is the exponential increase in number of parameters in the model. Here as a special case, we propose models that increase with the order of Markov chain by adding only 1 parameter. Other authors such as Raftery in [15] and Ching in [5] have proposed other methods to reduce the number of parameters. The dataset used in this study contains more than 100 years of daily precipitation for some stations. This allows us to look at some properties of the precipitation process such as stationarity more closely.

The models used here are an extension of the logistic regression for the independent data to dependent case. [14] investigates the estimation of the coefficients of rth order Markov chains with seasonal terms (non-homogenous) using partial likelihood and picking the model using BIC. It shows that the partial likelihood performs very well in picking the true model, the partial likelihood estimates are close to the true values and the distribution of the parameter estimates are close to normal distribution.

The paper begins in Section 2 with an exploratory analysis of precipitation data for the Province of Alberta, Canada. From there we turn in Section 3 to the construction of a family of Markov statistical models for the binary daily precipitation process. That leads in Section 4 to the model selection process. We give our conclusions in Section 7.

2 Exploratory data analysis.

This section explores daily precipitation data collected over the years from 1895 to 2006 at 48 stations distributed over Alberta, Canada. Although we investigated other locations and report some of that work here, for brevity we highlight those for Calgary, a site selected because of its comparatively long record of PN measurements. For a more detailed exploratory analysis of Alberta's climate variables, in particular similar plots for Banff see [12] and [13].

Figures 1 to 4 show plots for Calgary. Figure 1 plots the estimated 1st-order transition probabilities $\hat{p}_{11}$ (the probability of precipitation if precipitation occurs the day before) and $\hat{p}_{01}$ (the probability of precipitation if it does not occur the day before). These transition probabilities are estimated using the observed data. For example $\hat{p}_{11}$ for Jan 5th is estimated by n_{11}/n_1, where n_{11} is the number of pairs of days (Jan 4th, Jan 5th) with precipitation and n_1 is the number of Jan 5th's with precipitation during the available years. Figure 2 shows similar plots for estimated 2nd–order transition probabilities. Figure 3 gives the estimated annual probability of precipitation for Calgary computed by dividing the number of wet days of a year by the number of days in that year. The plot of the *logit* function and the transformed estimated probability of precipitation in Calgary are shown in Figure 4.

Before proceeding to develop statistical models in Section 3 for binary precipitation processes, we summarize the conclusions suggested by our exploratory data analyses. First, we find that that the binary PN process is not temporally stationary. in fact, Figure 1 shows the transition probabilities changing over time and season. Figure 1 also suggests the transition probabilities change continuously over time.

Although a lot of variation is seen in the higher order probabilities, a generally continuous trend is observed. We see a periodic trend for the transition probabilities over the course of the year; a simple periodic function seems appropriate for modeling these probabilities.

Figure 1 suggests p_{11} and p_{01} differ over the course of the year. Therefore a 0th-order Markov chain (independent) does not seem appropriate.

Figures 2 depict the plots for the pair $\hat{p}_{111}, \hat{p}_{011}$ and for $\hat{p}_{001}, \hat{p}_{101}$. They have considerable overlap over the course of the year. Therefore a 2nd-order Markov chain does not seem necessary.

Figure 3 shows the estimated probability of precipitation for different years, computed by averaging through the days of a given year. The median probability over all the available years is plotted to see the trends better. The probability of precipitation seems to differ from year-to-year. Moreover consecutive years seem to have similar probabilities. Hence assuming that different years are identically distributed and independent does not seem reasonable. The probability of precipitation seems to have increased over the past century for Calgary, possibly a manifestation of climate change.

Finally, Figure 4 shows the plot of the *logit* function applied to the estimated probabilities. Observe how the *logit* function transforms the values between 0 and 1 to a wider range in $\mathbb{R}$ and since it s an increasing function, its peaks occur at the same times as the original values.

3 Statistical model.

Let $X_0, X_1, X_2, \cdots$ be a discrete-time stochastic process, where X_t takes value in M_t, a finite subset of $\mathbb{R}$. Also assume strict positivity of the joint probabilities, which means every finite chain up to time t is possible. In order to model such

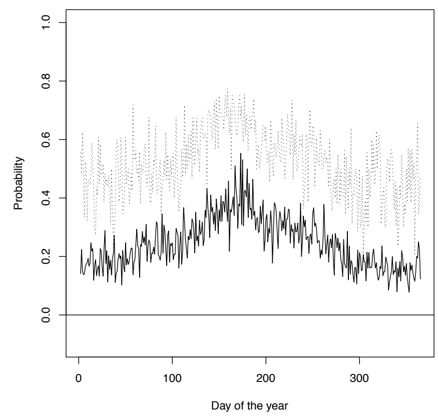

Figure 1: The transition probabilities for the Calgary site. The dotted line represents $\hat{p_{11}}$ (the estimated probability of precipitation if precipitation occurs the day before) and the dashed represents $\hat{p_{01}}$ (the estimated probability of precipitation if precipitation does not occur the day before.)

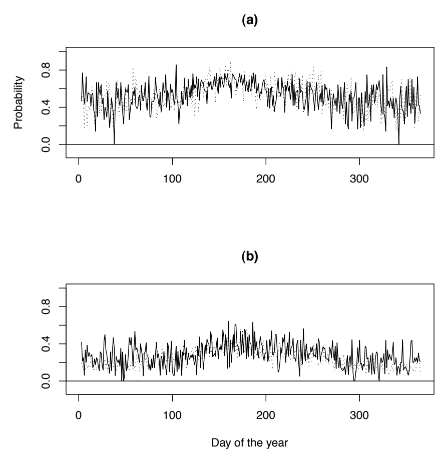

Figure 2: a) The solid curve represents $\hat{p}_{111}$ (the estimated probability of precipitation if during both two previous days precipitation occurs) and the dashed curve represents $\hat{p}_{011}$ (the estimated probability that precipitation occurs if precipitation occurs the day before and does not occur two days ago) for the Calgary site. b) The solid curve represents $\hat{p}_{001}$ (the estimated probability of precipitation occurring if it does not occur during the two previous days) and the dotted curve is $\hat{p}_{101}$ (the estimated probability that precipitation occurs if precipitation does not occur the day before but occurs two days ago) for the Calgary site.

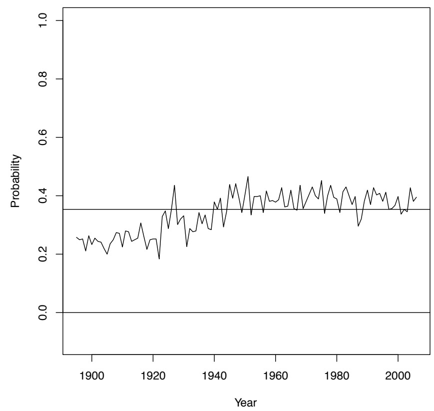

Figure 3: Calgary's estimated mean annual probability of precipitation calculated from historical data.

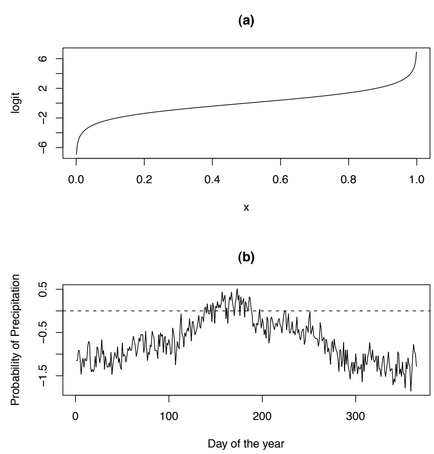

Figure 4: (a) The *logit* function: $logit(x) = log(x/(1-x))$. (b) The *logit* of the estimated probability of precipitation in Calgary for different days of the year.

processes it is more natural and desirable to use the conditional probabilities of the present given the past, thus requiring a parametric collection of functions to represent them. However one needs to introduce a statistical model for the conditional probabilities that are consistent in Kolmogorov's sense so that they actually correspond to a unique discrete-time categorical stochastic process. This requirement is fundamental for inference about these models. Meeting that requirement, we can start with a known stochastic process and follow the conceptual steps below to a generalization of logistic regression for categorical time series:

1. Start with a categorical stochastic process $X_0, X_1, \cdots$ where X_t takes a value in M_t. Fix an element m_t^0 in each M_t.

2. Find its joint densities: $p_t = P(X_0 = x_0, \cdots, X_t = x_t)$.

3. Find its conditional densities $P_t = P(X_t = x_t | X_{t-1} = x_{t-1}, \cdots, X_0 = x_0)$.

4. Find for each t:
$$h_t = \frac{P(X_t = x_t | X_{t-1} = x_{t-1}, \cdots, X_0 = x_0)}{P(X_t = m_t^0 | X_{t-1} = x_{t-1}, \cdots, X_0 = x_0)}.$$

5. Transform the result using the inverse of a bijective transformation $g : \mathbb{R} \to \mathbb{R}^+$ (for example $g(x) = \exp(x) \Rightarrow g^{-1}(x) = \log(x)$):
$$g_t(x_0, \cdots, x_t) = g^{-1}\{\frac{P(X_t = x_t | X_{t-1} = x_{t-1}, \cdots, X_0 = x_0)}{P(X_t = m_t^0 | X_{t-1} = x_{t-1}, \cdots, X_0 = x_0)}\},$$
$$(x_0, \cdots, x_{t-1}, x_t) \in M_0 \times \cdots \times M_{t-1} \times M_t'.$$

The above shows how to start with a stochastic process, fixed elements m_t^0 in each M_t and a prescribed function g to get a unique family of functions $\{g_t\}$. [9] shows the converse is also true, i.e. an arbitrary collection of functions g_t with the above relations correspond to a unique categorical process. Thus we have the following theorems, that are included for completeness.

Theorem 3.1 *Suppose* $M_0, M_1, \cdots \subset \mathbb{R},\ |M_t| = c_t < \infty,\ t = 0, 1, \cdots$. *Let* $P_0 : M_0 \to \mathbb{R}$ *be the density of a probability measure on* M_0 *and more generally for* $n = 1, \ldots,$ $P_n(x_0, x_1, \cdots, x_{n-1}, .)$ *be a positive probability density on* $M_n,\ \forall (x_0, \cdots, x_{n-1}) \in M_0 \times \cdots \times M_{n-1}$. *Then there exists a unique stochastic process (up to distributional equivalence) on a probability space* (Ω, Σ, P) *such that*
$$P(X_n = x_n | X_{n-1} = x_{n-1}, \cdots, X_0 = x_0) = P_n(x_0, x_1, \cdots, x_{n-1}, x_n).$$

Theorem 3.2 *Let $g : \mathbb{R} \to \mathbb{R}^+$ a bijection. For every choice of probability density p on $M = \{m_1, \cdots, m_n\}$, $n \geq 2$, there exists a unique function $f : M - \{m_1\} \to \mathbb{R}$, such that*

$$p(m_1) = \frac{1}{1 + \sum_{x \in M - \{m_1\}} h(x)}, \tag{1}$$

$$p(x) = \frac{h(x)}{1 + \sum_{x \in M - \{m_1\}} h(x)} \; x \neq m_1, \tag{2}$$

where $h = g \circ f$. Moreover, $h(x) = p(x)/p(m_1)$. Inversely, for an arbitrary function $f : M - \{m_1\} \to \mathbb{R}$, the p defined above is a density function.

Example Consider the binomial distribution with a trials and probability of success π and the transformation $g(x) = \exp x$. Then $M = \{0, 1, \cdots, a\}$. Let $m_1 = 0$. Then for $x \neq 0$

$$f(x) = g^{-1}(h(x)) = \log p(x)/p(0) = \log \binom{n}{x} p^x (1-p)^{n-x}/(1-p)^n =$$

$$\log \binom{n}{x} + x \log\{p/(1-p)\}.$$

The above theorems show how we can characterize discrete-time categorical stochastic processes using free functions g_t on a finite domain. In order to find a parametric form for these function Hosseini in [9] proved the following theorem

Theorem 3.3 *(Categorical Expansion Theorem) Suppose M_i is a finite subset of $\mathbb{R}$ with $|M_i| = c_i$, $i = 1, 2, \cdots, r$. Let $d_i = c_i - 1, M = \prod_{i=1,\cdots,r} M_i$ and consider the vector space of functions over $\mathbb{R}$, $V = \{f : M \to \mathbb{R}\}$ with the function addition as the addition operation of the vector space and the scalar product of a real number to the function as the scalar product of the vector space. Then this vector space is of dimension $C = \prod_{i=1,\cdots,r} c_i$ and $\{x_1^{i_1} \cdots x_r^{i_r}\}_{0 \leq i_1 \leq d_1, \cdots, 0 \leq i_r \leq d_r}$ forms a basis for it.*

Hosseini used the above theorems to find parametric forms for all discrete-time categorical stochastic processes and in particular rth-order Markov chains. In the following we give a specific example of stationary binary 3rd-order Markov chains with $M_t = \{0, 1, \}$ for clarity.

Example 3.1 *For the stationary binary (0-1) Markov chain of order $r = 3$ and $t \geq 3$ and the fixed transformation $g : \mathbb{R} \to \mathbb{R}^+$*
There exist unique α parameters:

$$g_t = g^{-1}\{\frac{P(X_t = 1|X_{t-1} = x_{t-1}, \cdots, X_0 = x_0)}{P(X_t = 0|X_{t-1} = x_{t-1}, \cdots, X_0 = x_0)}\}$$
$$= \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \alpha_3 x_{t-3}$$
$$+\alpha_{12} x_{t-1} x_{t-2} + \alpha_{23} x_{t-2} x_{t-3} + \alpha_{13} x_{t-1} x_{t-3}$$
$$+\alpha_{123} x_{t-1} x_{t-2} x_{t-3}.$$

Conversely every collection of arbitrary α's corresponds to a unique 3rd-order stationary binary (0-1) Markov chain. If we take $g : \mathbb{R} \rightarrow \mathbb{R}^+$, $g(x) = \exp(x)$ then $g^{-1}(x) = \log(x)$ in the above.

An advantage of this linear form is the ability to estimate the parameters using the partial likelihood as discussed by Kedem and Fokianos in [10]. Another advantage is its capacity to allow other linear terms needed to build non-stationary chains. For example, we can add $cos(\omega t)$ to model seasonality. In the theory of partial likelihood the covariate process is denoted by Z_{t-1}. We denote the 0-1 precipitation process by Y_t and discuss a few models in the following. For notational simplicity we denote Y_{t-i} by Y^i.

Examples of stationary Markov models of at most order 2:

- $Z_{t-1} = 1$:
  The probability of PN's occurrence does not depend on the previous days' states. In other words days are independent.

- $Z_{t-1} = (1, Y^1)$:
  The probability of PN today depends only on the day before and given the latter's value, it is independent of the other previous days.

- $Z_{t-1} = (1, Y^2)$:
  The probability of PN given the information for the day before yesterday is independent of other previous days, in particular yesterday! This does not seem reasonable.

- $Z_{t-1} = (1, Y^1, Y^2)$:
  This model includes both Y^1 and Y^2. One might suspect that it contains all relevant information and therefore that it is the most general 2nd-order Markov model. However, note that in the model, the transformed conditional probability is a linear combination of the past two states:

  $$logit\{P(Y = 1|Y^1, Y^2)\} = \alpha_0 + \alpha_1 Y^1 + \alpha_2 Y^2,$$

  which implies,

  $$logit\{P(Y = 1|Y^1 = 0, Y^2 = 0)\} = \alpha_0,$$

$$logit\{P(Y = 1|Y^1 = 1, Y^2 = 0)\} = \alpha_0 + \alpha_1,$$

$$logit\{P(Y = 1|Y^1 = 0, Y^2 = 1)\} = \alpha_0 + \alpha_2,$$

and

$$logit\{P(Y = 1|Y^1 = 1, Y^2 = 1)\} = \alpha_0 + \alpha_1 + \alpha_2.$$

We conclude that

$$\begin{aligned} & logit\{P(Y = 1|Y^1 = 1, Y^2 = 0)\} - logit\{P(Y = 1|Y^1 = 0, Y^2 = 0)\} \\ = \; & logit\{P(Y = 1|Y^1 = 1, Y^2 = 1)\} - logit\{P(Y = 1|Y^1 = 0, Y^2 = 1)\} \\ = \; & \alpha_1. \end{aligned}$$

In other words, the model implies that no matter what value Y^2 may have, the differences between the conditional probabilities given $Y^1 = 1$ and given $Y^1 = 0$ (in the *logit* scale) are the same.

- $Z_{t-1} = (1, Y^1Y^2)$:
  Among other things, this model implies that the conditional probabilities given $(Y^1 = 0, Y^2 = 1)$, $(Y^1 = 1, Y^2 = 0)$ or $(Y^1 = 0, Y^2 = 0)$ are the same.

- $Z_{t-1} = (1, Y^1, Y^2, Y^1Y^2)$:
  This is the full 2nd-order stationary Markov model with no restrictive assumptions as shown by Categorical Expansion Theorem.

Remark. The above explanations show that one must be careful about the assumptions made about any proposed model. Including or dropping various covariates can lead to implications that might be unrealistic.

[14] investigates the estimation of the coefficients of rth order Markov chains with seasonal terms (non-homogenous) using partial likelihood and picking the model using BIC. It shows that the partial likelihood performs very well in picking the true model, the partial likelihood estimates are close to the true values and the distribution of the parameter estimates are close to normal distribution.

3.1 Models for the 0-1 precipitation process.

In the light of Categorical Expansion Theorem 3.3, we know all the possible forms of rth-order Markov chains for binary data. Since this theorem gives us linear forms, time series following generalized linear models (TGLM) as discussed in [10] provides a method to estimate the parameters. However, it is worthwhile to study simpler models rather than the full model. For one thing, the full model contains a very large number of parameters, making their estimation difficult. Not only that, interpreting those parameters proves more difficult than in their more parsimonious alternatives.

We now introduce some processes that can be useful in developing simple models for the precipitation process:

- Y_t represents the occurrence of precipitation on day t. Here Y_t is a binary process with 1 denoting precipitation and 0 denoting its absence on day t.
- $N_{t-1}^l = \sum_{j=1}^l Y_{t-j}$ represents the number of PN days in the past l days.
- Seasonal processes (deterministic):

$$\cos(\omega t) \text{ and } \sin(\omega t),\ \omega = \frac{2\pi}{366}.$$

  We can also consider higher order terms in the Fourier series $\cos(\omega nt)$ and $\sin(\omega nt)$, where n is a natural number.

Some possibly interesting models present themselves when Z_{t-1} is a covariate process. The probability of precipitation today depends on the value of that covariate process, and those processes might include:

- $Z_{t-1} = (1, N_{t-1}^l)$. This model assumes that the probability of PN today only depends on the number of PN days during l previous days.
- $Z_{t-1} = (1, N_{t-1}^l, Y_{t-1})$. This model assumes that the PN occurrence today depends on the PN occurrence yesterday and the number of PN occurrences during l previous days.
- $Z_{t-1} = (1, \cos(\omega t), \sin(\omega t), N_{t-1}^l, Y_{t-1})$.
- $Z_{t-1} = (1, \cos(\omega t), \sin(\omega t), Y_{t-1})$.
- $Z_{t-1} = (1, Y_{t-1}, \cdots, Y_{t-r})$. This is a special case of Markov chain of order r. No interaction between the days is assumed. In this model increasing the order of Markov chain by one corresponds to adding one parameter to the model.
- $Z_{t-1} = (1, Y_{t-1}, \cdots, Y_{t-r}, Y_{t-1}Y_{t-2})$. In this model, the interaction between the previous day and two days ago is included.
- $Z_{t-1} = (1, \cos(\omega t), \sin(\omega t), Y_{t-1}, \cdots, Y_{t-r})$. In this model, two seasonal terms are added to the previous model.

4 Comparing the models using BIC

This section uses the methods developed previously to find appropriate models for the 0-1 PN process. We use the PN data for Calgary for just five years, from 2000 to 2004. This restriction is in part to ensure that the parameter estimates are current and hence, that the methodology developed here is of current value. In part if is to cut down computational time, recognizing that even with the restriction, five years yield quite a lot of daily process data. We compare several models using the BIC criterion. The partial likelihood is computed and then maximized using the "optim" function in "R".

Using the method called "Time Series Following Generalized Linear Model" as described by Kedem et al. in [10], applied to binary time series with the canonical link function, we have:

$$P(Y_t = 1|Z_{t-1}) = logit^{-1}(\alpha Z_{t-1}),$$

and,

$$P(Y_t = 0|Z_{t-1}) = 1 - logit^{-1}(\alpha Z_{t-1}).$$

We conclude that the log partial likelihood is equal to:

$$\sum_{t=1}^{N} \log P(Y_t|Z_{t-1}) =$$

$$\sum_{1 \le t \le N, Y_t=1} \log(logit^{-1}(\alpha Z_{t-1})) + \sum_{1 \le t \le N, Y_t=0} \log(1 - logit^{-1}(\alpha Z_{t-1})).$$

To ensure that the maximum picked by "optim" in the R package is close to the actual maximum, several initial values were chosen randomly until stability was achieved.

In order to find an optimal model to describe a binary (0-1) PN process, we can include several factors such as previous values of the process, seasonal terms, previous maximum temperature values and so on. We have done this comparison in several tables. The smallest BIC in the tables is shown by boldface.

Table 1 shows the constant process 1 and N^l, the number of wet days during l previous days, as predictors. Note that $N^1 = Y^1$. The BIC criterion in this case picks the simplest model which includes only the previous day. Hence a 1st-order Markov chain is chosen among these particular lth-order chains.

In Table 2 we see a refinement of the previous model, where now in addition to the number of days we add information about whether or not precipitation occurred during the previous day. More precisely it compares models with predictors: $(1, Y^l$ and $N^l)$, $l = 1, 2, \cdots, 30$. Since $Y^1 = N^1$ the first row is obviously an over–parameterized model. The smallest BIC corresponds to the model $(1, Y^1, N^{28})$. However, the primary gains are seen in going from N^1 to N^5 and in fact in going

Model: Z_{t-1}	BIC	parameter estimates
$(1, N^1)$	**2268.1**	$(-1.03, 1.26)$
$(1, N^2)$	2294.5	$(-1.09, 0.73)$
$(1, N^3)$	2293.4	$(-1.18, 0.56)$
⋮	⋮	⋮
$(1, N^{14})$	2340.5	$(-1.35, 0.15)$
$(1, N^{15})$	2342.6	$(-1.36, 0.14)$

Table 1: BIC values for models including N^l, the number of precipitation days during the past $l = 1, \dots, 15$ days for the Calgary site.

Model: Z_{t-1}	BIC	parameter estimates
$(1, Y^1, N^1)$	2275.6	(-1.04, -0.40, 1.67)
$(1, Y^1, N^2)$	2270.2	(-1.10, 0.94, 0.255)
$(1, Y^1, N^3)$	2258.3	(-1.21, 0.88, 0.279)
$(1, Y^1, N^5)$	2247.5	(-1.32, 0.91, 0.221)
$(1, Y^1, N^6)$	2248.2	(-1.34, 0.95, 0.187)
	⋮	
$(1, Y^1, N^{27})$	2246.2	(-1.60, 1.10, 0.062)
$(1, Y^1, N^{28})$	**2244.7**	(-1.62, 1.10, 0.061)
$(1, Y^1, N^{29})$	2245.4	(-1.62, 1.10, 0.059)
$(1, Y^1, N^{30})$	2246.2	(-1.62, 1.11, 0.057)

Table 2: BIC values for models including N^l, the number of wet days during the past l days and Y^1, the precipitation occurrence of the previous day for the Calgary site.

beyond $l = 5$ we see some inconsistency and little change. Thus we have confined our subsequent modeling to the case of $l \leq 5$. Note that the model $(1, Y^1, N^5)$ shows a subsequent improvement over $(1, Y^1)$. Hence by adding the number of PN days to the simple model $(1, Y^1)$, an improvement is achieved. However, we might expect seasonality to play an important role and this we now investigate.

Thus Table 3 compares models with predictors $(1, N^l, COS, SIN)$ that capture the seasonality in the precipitation pattern over a year through their harmonic terms. By including both the sine and the cosine, the model allows for phase changes.

Note its superiority over its simpler cousin $(1, Y^1)$, and even over $(1, Y^1, N^{28})$. Furthermore these results suggest that nothing is gained by counting the number of precipitation days beyond the previous day, once seasonality is included.

But what if in addition to seasonality and precipitation day counts, we add knowledge of what happened on the previous day? Table 4 includes Y^1, seasonal terms and N^l for $l = 1, 2, \cdots, 10$ as predictors. The model with predictors $(1, Y^1, N^5, COS, SIN)$ which includes a combination of seasonal terms and number of precipitation days has the smallest BIC so far. It seems that once it is freed of

Model: Z_{t-1}	BIC	parameter estimates
$(1, N^1, COS, SIN)$	**2222.5**	(-1.00, 1.10, -0.59, 0.100)
$(1, N^2, COS, SIN)$	2254.6	(-1.02, 0.59, -0.56, 0.097)
$(1, N^3, COS, SIN)$	2260.1	(-1.07, 0.44, -0.54, 0.096)
	⋮	
$(1, N^{10}, COS, SIN)$	2302.2	(-1.07, 0.14, -0.52, 0.107)

Table 3: BIC values for models including N^l, the number of wet days during the past $l = 1, \ldots, 10$ days and seasonal terms for the Calgary site.

Model: Z_{t-1}	BIC	parameter estimates
$(1, Y^1, N^1, COS, SIN)$	2230.0	(-1.00, -2.31, 3.41, -0.589, 0.0999)
$(1, Y^1, N^2, COS, SIN)$	2229.2	(-1.03, 0.977, 0.0997, -0.576, 0.0985)
$(1, Y^1, N^3, COS, SIN)$	2224.8	(-1.10, 0.895, 0.156, -0.546, 0.0946)
$(1, Y^1, N^4, COS, SIN)$	2222.1	(-1.14, 0.89, 0.147, -0.525, 0.0941)
$(1, Y^1, N^5, COS, SIN)$	**2221.7**	(-1.16, 0.922, 0.124, -0.515, 0.0934)
$(1, Y^1, N^6, COS, SIN)$	2223.3	(-1.16, 0.959, 0.0954, -0.517, 0.0946)
$(1, Y^1, N^7, COS, SIN)$	2223.7	(-1.17, 0.978, 0.0822, -0.513, 0.0947)
$(1, Y^1, N^8, COS, SIN)$	2224.7	(-1.16, 0.997, 0.0682, -0.515, 0.0945)
$(1, Y^1, N^9, COS, SIN)$	2225.5	(-1.16, 1.0129, 0.0582, -0.515, 0.0961)
$(1, Y^1, N^{10}, COS, SIN)$	2226.0	(-1.16, 1.026, 0.0502, -0.517, 0.0958)

Table 4: BIC values for models including N^l, the number of PN days during the past $l = 1, \ldots, 10$ days, Y^1, the precipitation occurrence of the previous day and seasonal terms for the Calgary site.

its responsibility to cover for Y^1, N^l is able to bring valuable additional information into play and thus improve the model, N^5 making the key contribution, presumably because this agrees roughly with the synoptic scales of meterological events that generate precipitation. However both the seasonal terms and the number of precipitation days prior to the day contribute to the statistical representation of "weather conditions". This is because climate has natural cycles throughout the year that can help predict weather on a particular day of the year. These natural cycles are modeled by the periodic functions COS and SIN. By looking at a short period prior to the current day (short-term past), we are able to learn about those weather conditions. At the same time, the best predictors (seasonal or short-term past) are important or necessary will depend on factors such as location.

We next investigate the effect of the order of a Markov process model used in conjunction with varying numbers of previous day indicators of precipitation or no-precipitation. More precisely Table 5 compares models with predictors ranging from $(1, Y^1)$ to $(1, Y^1, \cdots, Y^7)$. The first model is a 1st-order Markov chain and the last one is a 7th-order chain. The optimal choice proves to be : $(1, Y^1, Y^2, Y^3)$. Comparing this table to Table 2, we see that $(1, Y^1, N^3)$ is superior to $(1, Y^1)$,

Model: Z_{t-1}	BIC	parameter estimates
$(1, Y^1)$	2268.1	(-1.03, 1.27)
$(1, Y^1, Y^2)$	2270.2	(-1.11, 1.20, 0.23)
$(1, Y^1, Y^2, Y^3)$	**2263.3**	(-1.21, 1.19, 0.140, 0.410)
$(1, Y^1, \cdots, Y^4)$	2263.9	(-1.28, 1.16, 0.133, 0.334, 0.281)
$(1, Y^1, \cdots, Y^5)$	2268.5	(-1.32, 1.15, 0.121, 0.328, 0.232, 0.192)
$(1, Y^1, \cdots, Y^6)$	2335.4	(-1.34, 1.15, 0.0837, 0.357, 0.213, 0.135, 0.115)
$(1, Y^1, \cdots, Y^7)$	2286.7	(-1.51, 1.33, -0.113, 0.378, 0.418, 0.204, -0.0050, 0.214)

Table 5: BIC values for Markov models of different order with small number os parameters for the Calgary site.

Model: Z_{t-1}	BIC	parameter estimates
$(1, COS, SIN, Y^1)$	**2222.6**	(-1.0, -0.5, 0.1, 1.1)
$(1, COS, SIN, Y^1, Y^2)$	2229.1	(-1.0, -0.5, 0.1, 1.0, 0.1)
$(1, COS, SIN, Y^1, Y^2, Y^3)$	2230.4	(-1.1, -0.5, 0.1, 1.0, 0.02, 0.3)
$(1, COS, SIN, Y^1, \cdots, Y^4)$	2247.3	(-1.1, -0.5, 0.1, 1.0, 0.03, 0.2, 0.15)
$(1, COS, SIN, Y^1, \cdots, Y^5)$	2243.4	(-1.3, -0.4, 0.2, 1.4, -0.4, -0.1, 1.0, -0.15)
$(1, COS, SIN, Y^1, \cdots, Y^6)$	2501.6	(-1.2, -1.5, 0.4, 0.2, 0.8, 0.9, 0.9, -0.6, -0.2)
$(1, COS, SIN, Y^1, \cdots, Y^7)$	2447.3	(-1.1, -0.2, 0.07, 0.8, -0.02, 0.3, 0.4, -0.07, 0.4, -0.3)

Table 6: BIC values for Markov models with different order plus seasonal terms for the Calgary site.

$(1, Y^1, Y^2)$ and $(1, Y^1, Y^2, Y^3)$. Note that $(1, Y^1, N^3)$ is equivalent to $(1, Y^1, Y^2 + Y^3)$. Hence, including Y^2 and Y^3 and giving them the same weight is better than not including them, including one of them or including both of them.

Our Markov order analysis continues, but now with seasonality thrown in. Table 6 compares models with different Markov orders plus seasonal terms. The model $(1, Y^1, COS, SIN)$ is the winner. Hence, whether we include the seasonal terms or not, the model that only depends on the previous day is the winner.

Table 7 shows the results of further analysis of seasonality where we consider the possibility of other harmonics in the precipitation cycles over the year. It turns out that the model with $(1, Y^1, COS)$ is the optimal model so far - surprisingly one term seems sufficient to model the seasonal nature of the process. Table 8 compares all stationary 2nd-order Markov models. The smallest BIC corresponds to $(1, Y^1)$.

Table 9 compares all 2nd-order Markov chains with a seasonal COS term. The model $(1, Y^1, COS)$ proves to be the winner.

Table 10 also includes the maximum and minimum temperature of the day before, as predictors of some of the models which performed better in the above tables. We have also included the annual intercepts $A^1, \cdots, A^5$ for the 5 years under investigation. Finally, we have included the model $(1, Y^1, N^5, COS)$. This model has a combination of the seasonal term COS and the short-term past process N^5, which did best when combined with the seasonal terms and Y^1 in Table 4. It

Model: Z_{t-1}	BIC	parameter estimates
$(1, COS)$	2322.7	(-0.556, -0.717)
$(1, SIN)$	2424.3	(-0.523, 0.115)
$(1, COS, SIN)$	2327.3	(-0.568, -0.738, 0.119)
$(1, Y^1, COS)$	**2216.9**	(-1.00 , 1.10, -0.587)
$(1, Y^1, SIN)$	2273.9	(-1.03, 1.26, 0.0933)
$(1, Y^1, COS, SIN)$	2222.6	(-1.004, 1.102, -0.589, 0.100)
$(1, Y^1, COS, SIN, COS2)$	2229.7	(-1.00, 1.10, -0.586, 0.0998, 0.0247)
$(1, Y^1, COS, SIN, SIN2)$	2230.0	(-1.00, 1.10, -0.590, 0.101, 0.0125)
$(1, Y^1, COS, SIN, COS2, SIN2)$	2237.2	(-1.01, 1.11, -0.575, 0.0978, 0.0236, -0.0101)

Table 7: BIC values for models including seasonal terms and the occurrence of precipitation during the previous day for the Calgary site.

Model: Z_{t-1}	BIC	parameter estimates
(1)	2419.6	(-0.528)
$(1, Y^1)$	**2268.0**	(-1.04, 1.27)
$(1, Y^2)$	2392.8	(-0.76, 0.590)
$(1, Y^1, Y^2)$	2270.2	(-1.11, 1.20, 0.26)
$(1, Y^1Y^2)$	2335.5	(-0.78, 1.13)
$(1, Y^1, Y^1Y^2)$	2272.7	(-1.04, 1.11, 0.28)
$(1, Y^2, Y^1Y^2)$	2342.3	(-0.76, -0.11, 1.22)
$(1, Y^1, Y^2, Y^1Y^2)$	2277.7	(-1.10, 1.18, 0.23, 0.05)

Table 8: BIC values for 2nd-order Markov models for precipitation at the Calgary site.

Model: Z_{t-1}	BIC	parameter estimates
$(1, COS)$	2322.7	(-0.57, -0.74)
$(1, COS, Y^1)$	**2216.8**	(-1.0, -0.59, 1.10)
$(1, COS, Y^2)$	2317.4	(-0.71, -0.68, 0.37)
$(1, COS, Y^1Y^2)$	2223.5	(-0.76, -0.62, 0.90)
$(1, COS, Y^1, Y^2)$	2276.1	(-1.0, -0.57, 1.08, 0.10)
$(1, COS, Y^1, Y^1Y^2)$	2223.9	(-1.0, -0.58, 1.0, 0.12)
$(1, COS, Y^2, Y^1Y^2)$	2280.9	(-0.70, -0.63, -0.24, 1.0)
$(1, COS, Y^1, Y^2, Y^1Y^2)$	2231.0	(-1.0, -0.57, 1.0, 0.08, 0.04)

Table 9: BIC values for 2nd-order Markov models for precipitation at the Calgary site plus seasonal terms.

Model: Z_{t-1}	BIC	parameter estimates
$(1, COS, Y^1)$	**2216.8**	(-1.0, -0.59, 1.10)
$(1, Y^1, COS, MT^1)$	2221.7	(-0.84, 1.0, -0.74, -0.01)
$(1, Y^1, COS, mt^1)$	2224.2	(-1.0, 1.0, -0.65, -0.005)
$(1, Y^1, COS, MT^1, mt^1)$	2227.4	(-0.65, 0.99, -0.67, -0.025, 0.02)
$(1, Y^1, COS, A^1, \cdots, A^5)$	2241.2	(1.1, -0.5, -0.9, -1.2, -1.1, -1.0, -0.7)
$(1, Y^1, N^5, COS, MT^1)$	2297.3	(-2.1, 0.9, 0.4, 0.6, 0.2, 0.04)
$(1, Y^1, N^5, COS, SIN, MT^1, mt^1)$	2516.8	(1.4, 0.04, 0.2, 0.7, 0.8, -0.2, 0.3)
$(1, Y^1, N^5, COS, MT^1, mt^1)$	2393.9	(1.4, 0.7, -0.1, -0.5, 0.5, -0.1, 0.2)
$(Y^1, N^5, COS, MT^1, A^1, \cdots, A^5)$	2697.1	(1.2, -0.64, -2.0, -0.10, 2.0, 1.2, 2.2, 1.2, 1.8)
$(Y^1, N^5, COS, A^1, \cdots, A^5)$	2447.1	(0.1, 0.1, -0.7, -0.39, -0.01, -0.2, -0.9, -1)
$(1, Y^1, MT^1)$	2251.5	(-1.2, 1.3, 0.021)
$(1, Y^1, N^5, COS)$	**2215.8**	(-1.1, 0.9, 0.1, -0.5)
$(1, Y^1, N^5, COS, MT^1)$	2223.8	(-1.2, 0.9, 0.1, -0.4, 0.0)

Table 10: BIC values for models including several covariates as temperature, seasonal terms and year effect for precipitation at the Calgary site for the five years under investigation, 2000–2004.

turns out that including MT and mt does not improve the BIC as well as does the annual intercept terms. However, $(1, Y^1, N^5, COS)$ has the smallest BIC in all the models, which is a seasonal Markov chain of order 5 with only 4 parameters. Also the simpler model $(1, Y^1, COS)$ has a BIC close to that of $(1, Y^1, N^5, COS)$.

5 Changing the location and the time period.

This section investigates the robustness of our findings to changes in the five time period and monitoring site. More specifically we examine several models seen in the previous section by changing the five year time period from that studied there, to 1990 to 1994 in one case and the site from Calgary to Medicine Hat in another. Table 11 presents the results for the new time period, and Table 12 , for the new site.

For the new time period (Table 11) model $(1, Y^1, COS)$ has the smallest BIC. In fact particular the BIC for this model is smaller than the BIC for $(1, Y^1, N^5, COS)$, which has the smallest BIC for Calgary 2000–2004. However $(1, Y^1, COS)$ ranked second in the latter case, so in fact, the results change little. Including the maximum and minimum temperature to the model increases the BIC again as it did in the previous section.

When we change the location to Medicine Hat, (Table 12) the smallest BIC corresponds to $(1, Y^1, COS)$. However, several models have very similar BIC values. Once again, including the maximum and minimum temperature increases the BIC here.

In all the three cases Calgary 2000 – 2004, Calgary 1990–1994, and Medicine

Model: Z_{t-1}	BIC	parameter estimates
$(1, Y^1)$	2312.7	(-0.93, 1.3)
$(1, Y^1, Y^2)$	2318.8	(-0.97, 1.2, 0.13)
$(1, Y^1, COS)$	**2228.8**	(-0.86, 1.03, -0.71)
$(1, Y^1, N^5)$	2303.3	(-1.17, 1.012, 0.17)
$(1, Y^1, N^{10})$	2287.9	(-1.6, 1.01, 0.13)
$(1, Y^1, N^{15})$	2282.7	(-1.5, 1.04, 0.10
$(1, Y^1, COS, SIN)$	2231.9	(-0.85, 1.03, -0.71 , 0.15)
$(1, Y^1, N^5, COS)$	2236.4	(-0.86, 1.03, 0.004, -0.70)
$(1, Y^1, N^5, SIN)$	2307.8	(-1.2, 1.01, 0.16, 0.12)
$(1, Y^1, N^5, COS, SIN)$	2239.4	(-0.85, 1.03, -0.004, -0.72, 0.15)
$(1, Y^1, N^{10}, COS)$	2236.4	(-0.84, 1.03, -0.002, -0.72, 0.15)
$(1, Y^1, N^{10}, COS, SIN)$	2239.4	(-0.85, 1.03, -0.002, -0.72, 0.15)
$(1, Y^1, N^5, COS, MT^1)$	2244.3	(-0.43, 1.04, -0.096, -1.07, -0.021)
$(1, Y^1, N^5, COS, mt^1)$	2244.1	(-0.91, 1.01, 0.03, -0.58, 0.006)

Table 11: BIC values for several models for the binary process of precipitation in Calgary, 1990–1994

Model: Z_{t-1}	BIC	parameter estimates
$(1, Y^1)$	2202.9	(-1.138, 1.094)
$(1, Y^1, Y^2)$	2207.9	(-1.183, 1.051, 0.181)
$(1, Y^1, N^5)$	2203.6	(-1.275, 0.921, 0.119)
$(1, Y^1, N^{10})$	2228.9	(-0.858, 1.036, -0.712)
$(1, Y^1, N^{15})$	2200.5	(-1.420, 0.980, 0.065)
$(1, Y^1, N^{20})$	2202.5	(-1.421, 1.008, 0.048)
$(1, Y^1, COS)$	**2201.2**	(-1.134, 1.067, -0.224)
$(1, Y^1, COS, SIN)$	2202.9	(-1.132, 1.052, -0.225, 0.177)
$(1, Y^1, N^5, COS)$	2203.9	(-1.252, 0.924, 0.101, -0.201)
$(1, Y^1, N^5, SIN)$	2206.6	(-1.263, 0.922, 0.109, 0.158)
$(1, Y^1, N^5, COS, SIN)$	2206.6	(-1.239, 0.925, 0.091, -0.204, 0.163)
$(1, Y^1, N^{10}, COS)$	2201.9	(-1.336, 0.958, 0.073, -0.183)
$(1, Y^1, N^{10}, COS, SIN)$	2205.1	(-1.311, 0.958, 0.065, -0.187, 0.151)
$(1, Y^1, N^5, COS, MT^1)$	2306.5	(-1.455, 2.099, -0.130, 0.041, 0.004)
$(1, Y^1, N^5, COS, mt^1)$	2211.1	(-1.238, 0.937, 0.087, -0.267, -0.005)
$(1, Y^1, N^{15}, COS)$	2202.7	(-1.363, 0.981, 0.053, -0.175)

Table 12: BIC values for several models for precipitation occurrence in Medicine Hat, 2000-2004

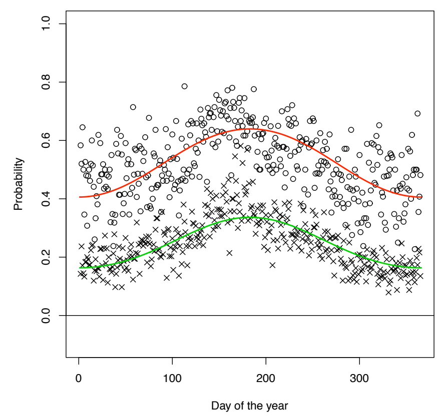

Figure 5: The circles represent $\hat{p_{11}}$ (the estimated probability of precipitation if precipitation occurs the day before) and the crosses represent $\hat{p_{01}}$ (the estimated probability of precipitation if precipitation does not occur the day before.) The fitted transition probabilities are also plotted.

Hat 2000–2004 $(1, Y^1, COS)$ is either optimal or the second to the optimal (using BIC) pointing to it as the best choice. To validate that choice, we repeated our analysis for Calgary over a long time period of close to 100 years. The same simple model $(1, Y^1, COS)$ proved optimal.

6 Applying the model

At first we use the Calgary data from 1900 to 2000 and fit the optimal model with covariates $Z_{t-1} = (1, Y^1, COS)$. Then we compare the fitted 1st-order transition probabilities with the calculated transition probabilities using the historical data. Figure 5 shows a good fit. However the peak of the precipitation season is somewhat underestimated.

Next we show how the approach developed above can be used in applications. This time we use the data for Calgary site from 1980 to 2000. Using BIC shows the superiority of the model $Z_{t-1} = (1, Y_{t-1}, 1)$ to the above discussed models again. We use the developed methodology to compute two probabilities:

- π_1 : The probability of no precipitation in the first week of October at the Calgary site.
- π_2 : The probability of at least 5 days without precipitation in the first week of October at the Calgary site.

The first day of October is the 275th day of the year in a leap year and the 274th day of the year in a non-leap year. We compute the probabilities for the week

Covariate	Theoretical sd	Experimental sd
1	0.033	0.034
Y^1	0.051	0.057
COS	0.037	0.032

Table 13: Theoretical and simulation estimated standard deviations for extremely cold process $PN(t)$ at the Calgary site.

between 274th day and 281th day which corresponds to the first week of October in a non-leap year. We prefer this option to computing the probability for the actual first week of October, since this corresponds better to the natural cycles. Of course with a little modification one could compute the probability for the first week of October, for example by introducing a probability of 1/4 for being in a leap year.

We compute the standard deviations once using simulations by generating chains from the fitted model with covariates $(1, Y^1, COS)$, and once by computing the partial information matrix, G_N, using partial likelihood theory. The results are given in Table 13. The variance-covariance matrix calculated using partial likelihood theory is given below:

$$\begin{pmatrix} 0.0011 & -0.0011 & 0.0000 \\ -0.0011 & 0.0026 & 0.0003 \\ 0.0000 & 0.0003 & 0.0013 \end{pmatrix}$$

We also find the variance-covariance matrix using simulations. To do that we generate 50 chains over time using the estimated parameters. The variance-covariance matrix using the simulations is given by:

$$\begin{pmatrix} 0.0011 & -0.0014 & -0.0001 \\ -0.0014 & 0.0033 & 0.0007 \\ -0.0001 & 0.0007 & 0.0012 \end{pmatrix}$$

We see that the simulated variance-covariance matrix has close values to the partial likelihood, all entries having the same sign. We also look at the distribution of the estimators using the 50 samples. Figure 6 shows the parameter estimates approximately follow a normal distribution.

To estimate the desired probabilities, we generate samples (10000) from the parameter space using the mean of the parameters and variance-covariance matrix from a multivariate normal. To fix ideas suppose we want to compute the probability of no frost between (and including) the 274th day and the 280th day of the year. For every vector of parameters, we then compute the probability of observing $(0, 0, 0, 0, 0, 0, 0)$ exactly once given it was dry on the 273th day and once it wet. In other words we compute

$$P(Y(274) = 0, \cdots, Y(281) = 0 | Y(273) = 1),$$

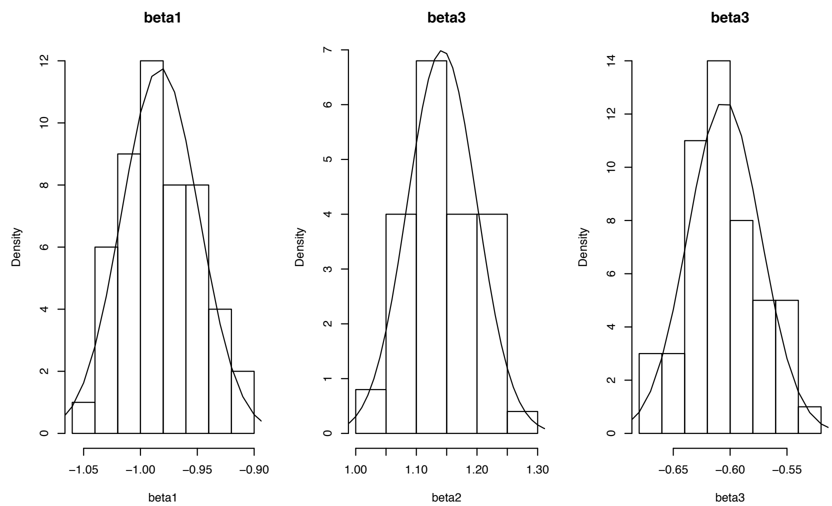

Figure 6: Normal curved fitted to the distribution of 50 samples of the estimated parameters.

and

$$P(Y(274) = 0, \cdots, Y(281) = 0|Y(273) = 0).$$

We also use the historical data to estimate $p_0 = P(Y(273) = 1)$. Then the desired probability would be

$$\begin{aligned} P(Y(274) = 0, \cdots, Y(281) = 0) &= \\ p_0 P(Y(274) = 0, \cdots, Y(281) = 0|Y(273) = 1) &+ \\ (1 - p_0) P(Y(274) = 0, \cdots, Y(281) = 0|Y(273) = 0) & \end{aligned}$$

Then in order to get a 95% confidence intervals we use $(q(0.025), q(1-0.025))$, where q is the (left) quantile function of the vector of the probabilities.

Using the historical data, we obtain $p_0 = P(Y(273) = 1) = 0.428$. Then for every parameter generated from the multivariate normal with mean and the above variance-covariance matrix we can estimate the two probabilities π_1 and π_2. We sample 10000 times from the multivariate normal, compute 10000 probabilities and take the 0.025th and 0.975th (left) quantiles to get the following confidence intervals for π_1 and π_2 respectively:

$$(0.115, 0.141),$$

and

$$(0.577, 0.624).$$

If we use the simulated variance-covariance matrix, we'll get the following confidence intervals for π_1 and π_2

$$(0.115, 0.141),$$

and

$$(0.579, 0.622),$$

which are very similar to the aforementioned intervals.

7 Summary and conclusions.

In all the three cases Calgary 2000–2004, Calgary 1990–1994, and Medicine Hat 2000–2004 $(1, Y^1, COS)$ is either optimal or the second to the optimal (using BIC) pointing to it as the best choice. To validate that choice, we repeated our analysis for Calgary over a long time period of close to 100 years. The same simple model $(1, Y^1, COS)$ proved optimal.

This seems surprising since the absence of the SIN term in the model means that the model does not allow for phase changes. However, the presence of the COS only with a negative coefficient, agrees with what we see in Figure 4. That figure shows that the likelihood of precipitation is low at the beginning of the year, maximized in the middle and finally low at the end. While the shape of the cosine may not be ideal, it captures the precipitation likelihood pattern very parsimoniously with just a single parameter, its amplitude, making it attractive by the BIC criterion.

The work reported in this paper has focussed on a single site to identify good models. However, the precipitation is a space-time field. Current work is directed at the problem of modeling the spatial dependence between sites, and mapping that field between monitoring sites to give an integrated picture of the overall risk of such things as drought.

Acknowledgement. We are indebted to E. Mékis from Environment Canada for providing the data on which the application in this paper is based. This work was partially supported by grants from the Natural Science and Engineering Research Council of Canada and the National Institute for Complex Data Structures.

References

[1] H. Akiake. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, pages 716–723, 1974.

[2] T. W. Anderson and L. A. Goodman. Statistical inference about markov chains. *Ann. Math. Statist.*, pages 89–110, 1957.

[3] M. S. Bartlett. The frequency goodness of fit test for probability chains. *Proc. Cambridge Philos. Soc.*, pages 86–95, 1951.

[4] E. H. Chin. Modeling daily precipitation process with markov chain. *Water resources research*, (6):949–956, 1977.

[5] W. K. Ching, E. S. Fung, and K. M. NG. Higher-order markov chain models for categorical data sequences. *Naval Research Logistics*, pages 557–574, 2004.

[6] K. R. Gabriel and J. Neumann. A markov chain model for daily rainfall occurance at tel aviv. *Quart. J. Roy. Met. Soc.*, pages 90–95, 1962.

[7] E. J. Hannan. The estimation of the order of an arma process. *Ann. Statist.*, pages 1071–1081, 1980.

[8] D. M. A. Haughton. On the choice of a model to fit data from an exponential family. *The Annals of Statistics*, (1):342–355, 1988.

[9] R. Hosseini. *Statistical Models for Agroclimate Risk Analysis.* PhD thesis, Department of Statistics, UBC, 2009.

[10] B. Kedem and K. Fokianos. *Regression Models for Time Series Analysis.* Wiley Series in Probability and Statistics, 2002.

[11] S. E. Moon, S. Ryo, and J. Kwon. International journal of climatalogy. pages 1009–116, 1993.

[12] Hosseini R., Zidek J., and Le N. An analysis of alberta's climate. part i: Non-homogenized data. Technical Report TR #245, Department of Statistics, UBC, 2009.

[13] Hosseini R., Zidek J., and Le N. An analysis of alberta's climate. part ii: Homogenized data. Technical Report TR #246, Department of Statistics, UBC, 2009.

[14] Hosseini R., Zidek J., and Le N. r-th order categorical markov chains. Technical Report TR #248, Department of Statistics, UBC, 2009.

[15] A. E. Raftery. A model for higher order markov chains. *J. R. Statist. B.*, (3):528–539, 1985.

[16] G. Schwartz. Estimating the dimension of a model. *Ann. Statist.*, pages 461–464, 1978.

[17] R. Shibita. Selection of the order of an autoregressive modelby akiake's information criterion. *Biometrika*, pages 117–126, 1976.

[18] H. Tong. Determination of the order of a markov chain by akiake's information criterion. *J. Appl. Prob.*, pages 488–497, 1975.

[19] H. Tong and P. Gates. On markov chain modelling to some weather data. *Journal of applied meteorology*, pages 1145–1151, 1976.