A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts

Enrica Bellone James P. Hughes Peter Guttorp

NRCSE

Technical Report Series

NRCSE-TRS No. 021

A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts

Enrica Bellone[1,2], James P Hughes[3], Peter Guttorp[2]

[2]Department of Statistics 354322, University of Washington, Seattle, WA 98195.

[3]Department of Biostatistics 357232, University of Washington, Seattle, WA 98195.

Abstract

Nonhomogeneous hidden Markov models (NHMMs) provide a relatively simple framework for simulating precipitation at multiple rain gauge stations conditional on synoptic atmospheric patterns. Building on existing NHMMs for precipitation occurrences, we propose an extension to also include precipitation amounts. The model we describe assumes the existence of unobserved (or hidden) weather patterns, the weather states, which follow a Markov chain. The weather states depend on observable synoptic information and therefore serve as a link between the synoptic scale atmospheric patterns and the local scale precipitation. The presence of the hidden states simplifies the spatio-temporal structure of the precipitation process. We assume the temporal dependence of precipitation is completely accounted for by the Markov evolution of the weather state. The spatial dependence in precipitation can also be partially or completely accounted for by the existence of a common weather state. In the proposed model, occurrences are assumed to be conditionally spatially independent given the current weather state and, conditional on occurrences, precipitation amounts are modeled independently at each rain gauge as gamma deviates with gauge specific

[1]Email: bellone@stat.washington.edu

parameters.

We apply these methods to model precipitation at a network of 24 rain gauge stations in Washington state over the course of 17 winters. The first 12 years are used for model fitting purposes, while the last 5 serve to evaluate the model performance. The analysis of the model results for the reserved years suggests that the characteristics of the data are captured fairly well and points to possible directions for future improvements.

Keywords: hidden Markov model, precipitation amounts model, downscaling

1 Introduction

Stochastic models for precipitation have several important applications. For example, simulations from these models enter as input into flooding, runoff and crop growth models. General circulation models (GCMs) have been developed to realistically simulate atmospheric circulation patterns under different climate regimes and rainfall models can be used to downscale the effect of these atmospheric patterns to precipitation.

Historically, rainfall modeling has followed two main themes. Some models were constructed to incorporate physical principles (e.g. Hobbs and Locatelli 1978), while others gave a more statistical description of the data. Along the lines of the former approach, point process models were developed by Le Cam (1961), Waymire, Gupta, and Rodriguez-Iturbe (1984) and Goodall and Phelan (1991), based on the idea of *rain storm cells.* The basic entity is the *convective cell* – identified by birth time and location and the models describe a cluster process of N cells.

In the context of statistical descriptions of rainfall data, Gabriel and Neumann (1992) modeled precipitation occurrences as a first-order Markov chain. Their approach has been extended to allow seasonal differences (Stern and Coe 1984; Woolhiser 1992) by using time-varying parameters.

Recently, the idea of relating daily precipitation to synoptic atmospheric patterns has led to the development of *weather state* models. One motivation for including atmospheric variables in the model is the desire to assess the regional and local effects of global climate changes. General circulation models, which typically operate on grids on the order of 5° latitude × 5° longitude, can capture large-scale atmospheric patterns and determine the effect of changes in the atmosphere on those patterns. However, GCMs are not as adequate for reproducing local and regional phenomena, such as rainfall (Giorgi and Mearns 1991). Thus, there is a need for models that can downscale the GCM predictions of global climate to local precipitation patterns. Stochastic models for rainfall that do not include synoptic atmospheric information can not be used for this purpose, since they can only produce simulations under the current climate regime. In weather state models, synoptic atmospheric patterns are the basis for classifying each day into a weather state and precipitation is then modeled within each state via multivariate distributions. Different versions of these models have been proposed by, for example, Hay et al. (1991), Bardossy and Plate (1992), Hughes et al. (1993) and Bartholy et al. (1995).

Our goal is to obtain a model that allows simulation of precipitation amounts, conditional on the value of some synoptic atmospheric variables. We base our approach on nonhomogeneous hidden Markov models (NHMMs), a class of models introduced by Hughes and Guttorp (1994). NHMMs extend the hidden Markov models (HMMs) used by Zucchini and

Guttorp (1991) by incorporating synoptic atmospheric information. Nonhomogeneous hidden Markov models assume the existence of *weather states*, but they differ from the weather state models mentioned above in the way the states are defined. In weather state models, each day is classified a priori into a state, according to synoptic patterns. Precipitation does not affect the state definition. In NHMMs, instead, the states are identified as precipitation patterns that result from the model fitting procedure, while the role of synoptic atmospheric information is to influence the state transitions. In Hughes et al. (1994, 1999) nonhomogeneous Markov models (NHMM) are used to model precipitation occurrences. Here we extend this approach to precipitation amounts.

In Section 2 we describe the assumptions that define a NHMM, the parameterization we use and the methods we apply to obtain estimates of the model parameters. Section 2 also explains our approach to the problem of determining the model order and to the treatment of the atmospheric variables. Section 3 describes an application of our methods to precipitation amounts at a network of gauges in Washington state. In section 4, we conclude with a discussion.

2 Methods

2.1 Model Assumptions

The NHMM assumes the existence of a *hidden* or unobservable stochastic process, which can take on a discrete number of states. In the context of precipitation modeling, we interpret this process as the "state of the weather." The adjective "nonhomogeneous" derives from the assumption that the state of the weather at time t depends not only on the state of the weather at the previous time point, but also on the current value of some atmospheric variables. Thus, the state transition matrix varies in time with the atmospheric quantities. The assumptions for the hidden process can be summarized as:

$$P\left(S_t \mid S_1^{t-1}, \mathbf{X}_1^T\right) = P\left(S_t \mid S_{t-1}, \mathbf{X}_t\right), \tag{1}$$

where S_t is the weather state at time t and $\mathbf{X}_t$ is a vector of atmospheric variables at time t.

Assumption (1) asserts that, given the state of the weather at the previous time point and the current value of some atmospheric variables, the state of the weather at time t does not depend on any other history of states nor on any other past or future values of the atmospheric quantities.

The parameterization we adopt for $P(S_t \mid S_{t-1}, \mathbf{X}_t)$ is:

$$P\left(S_t \mid S_{t-1}, \mathbf{X}_t\right) \propto \gamma_{ij} \exp\left(-\frac{1}{2}(\mathbf{X}_t - \boldsymbol{\mu}_{ij})\boldsymbol{\Sigma}^{-1}(\mathbf{X}_t - \boldsymbol{\mu}_{ij})'\right), \tag{2}$$

where $\boldsymbol{\Sigma}$ is the variance-covariance matrix for the atmospheric data. The $\boldsymbol{\mu}_{ij}$ parameters represent the mean vectors of the atmospheric variables when the state of the weather at the previous time point was state i and the current state of the weather is j, while the γ_{ij} parameters can be interpreted as baseline transition probabilities. It is necessary to impose the constraints $\sum_j \gamma_{ij} = 1$ and $\sum_j \boldsymbol{\mu}_{ij} = \mathbf{0}$, in order to ensure identifiability of the parameters.

The other fundamental element in the NHMM is the *observed* stochastic process – in this context precipitation – which is assumed to be conditionally temporally independent, given the weather state. The hidden Markov model assumptions for the observed process can be summarized by:

$$f_{\mathbf{R}_t|\left(S_1^T,\mathbf{R}_1^{t-1},\mathbf{X}_1^T\right)}(\mathbf{r}) = f_{\mathbf{R}_t|S_t}(\mathbf{r}) \tag{3}$$

where $\mathbf{R}_t$ is the vector of precipitation amounts at a network of stations at time t. Thus, given the current weather state, precipitation is assumed independent from all the past precipitation values, all other past and future weather states and from any values of the atmospheric variables.

Assumptions (1) and (3) determine the temporal structure in the precipitation process. The definition of the spatial structure requires additional hypotheses. In the analysis presented here, we assume conditional spatial independence of both occurrences and amounts given the weather state; i.e. we hypothesize that all the dependence between rain gauges is induced by the common weather state. In the discussion we suggest possible extensions to this relatively simple dependence structure.

The parameterization for the observed process builds on the spatial independence model for precipitation occurrences of Hughes and Guttorp (1994). Amounts are introduced by modeling precipitation at each station, given the weather state, as a mixture of a point mass at zero and a gamma distribution. In other words, conditional on the current weather state and on the occurrences, we model the amounts at each gauge as a gamma distribution (with state specific parameters). The resulting parameterization is:

$$f_{\mathbf{R}_t|S_t=s}(\mathbf{r}) = \prod_{i=1}^{N} \left[\left(p_{si} \Gamma \left(r^i; \alpha_{si}, \beta_{si} \right) \right)^{1_{[r^i > c]}} \left(1 - p_{si}\right)^{1_{[r^i \leq c]}} \right], \tag{4}$$

where N is the number of rain stations, p_{si} is the precipitation probability at station i in state s, α_{si} and β_{si} are the gamma parameters for station i in state s, and c is a prespecified cutoff (i.e. amounts below c are treated as no precipitation).

In the model described by (2) and (4) the number of unconstrained parameters is

$$S(S-1)(M+1)+3SN,$$

where M is the number of atmospheric variables included in the model, S the number of weather states and N the number of rain gauges.

2.2 Parameter estimation

Parameter estimates are obtained by maximizing the likelihood. The likelihood of the observed data given the atmospheric variables is:

$$\begin{aligned} L(\boldsymbol{\theta}) &= f_{\mathbf{R}_1^T|\mathbf{X}_1^T=\mathbf{x}_1^T,\boldsymbol{\theta}}\left(\mathbf{r}_1^T\right) && (5)\\ &= \sum_{s_1,\ldots,s_T} f_{\left(\mathbf{R}_1^T,S_1^T\right)|\mathbf{X}_1^T=\mathbf{x}_1^T,\boldsymbol{\theta}}\left(\mathbf{r}_1^T,s_1^T\right) && (6)\\ &= \sum_{s_1,\ldots,s_T} P(S_1=s_1|\mathbf{X}_1) f_{\mathbf{R}_1|S_1=s_1}(\mathbf{r}_1)\prod_2^T P(S_t=s_t|S_{t-1},\mathbf{X}_t) f_{\mathbf{R}_t|S_t=s_t}(\mathbf{r_t}), && (7)\end{aligned}$$

where $\boldsymbol{\theta}$ is the vector of the model parameters. In rainfall modeling, the number of observation times, T, is usually large. But even for small T's, computation of the likelihood directly as in (7) is intractable. However, the calculation is possible using the forward-backward algorithm, originally developed by Baum (1972). The algorithm is recursive and successively moves the terms of the summations in (7) as far to the right as possible. Following this principle, the likelihood can be written as:

$$L(\boldsymbol{\theta}) = \delta(\mathbf{x}_1)\mathbf{B}(\mathbf{r}_1)\mathbf{A}(\mathbf{x}_2)\mathbf{B}(\mathbf{r}_2)...\mathbf{A}(\mathbf{x}_T)\mathbf{B}(\mathbf{r}_T)\mathbf{1}',$$

where $\mathbf{B}(\mathbf{r})$ is a $S \times S$ diagonal matrix, with $b_{ss}(r) = f_{\mathbf{R}_t|S_t=s}(\mathbf{r})$, $\mathbf{A}(\mathbf{x})$ is a $S \times S$ transition matrix with $a_{ij}(x) = P(S_t = j|S_{t-1} = i, \mathbf{X}_t = \mathbf{x})$, $\mathbf{1}'$ is a length S column vector of ones and $\delta(\mathbf{x})$ is a row vector of length S. The quantity $\delta(\mathbf{x})$ is the solution to $\delta(\mathbf{x})A(\mathbf{x}) = \delta(\mathbf{x})$ (i.e. it is the stationary distribution for $A(\mathbf{x})$).

To maximize the likelihood, we apply the EM algorithm. Hughes, Guttorp, and Charles (1999) give a detailed description of this procedure.

2.3 Model order

Fitting an NHMM to precipitation data involves the choice of a model order and of the atmospheric variables to be included. We first determine the order of the NHMM, i.e. the number of hidden weather states, and include the atmospheric variables afterwards. The choice of the number of hidden states is a non-trivial issue. Standard likelihood-based methods – such as the Akaike information criterion (AIC) (Akaike 1974) and the Bayesian information criterion (BIC) (see Kass and Raftery 1995 for a review) – rely upon assumptions that do not hold for the order selection problem. Nonetheless, the Bayesian information criterion, defined as:

$$\text{BIC} = -2 \text{ loglikelihood} + \log(\text{no. observations}) \text{ (no. free parameters)},$$

yields reasonable models in terms of interpretability and fit to the data (Hughes, Guttorp, and Charles 1999). Thus, BIC is one of the elements – but not the sole determining factor – that we use in choosing the number of weather states.

Models of different order can also be compared with respect to their capability of reproducing some key features in the observed data. For example, an important characteristic we try to match is the distribution of the "storm" durations at the different rain gauges, where "storm" is defined as a string of consecutive days when precipitation occurred.

Another consideration is the increase in the number of parameters induced by an increase in the number of weather states. Choosing too many states can lead to an intractable model in terms of computer time.

2.4 Atmospheric variables

Atmospheric data are used to help determine the current (hidden) weather state (see equations (1) and (2)). To reduce the number of model parameters, we prefer to include relatively few atmospheric variables in the model. However, synoptic scale atmospheric variables are typically available on regular grids and several grid nodes usually cover the region of interest. Thus, a method for summarizing the grid data into few values is needed. Our approach is based on the singular value decomposition (SVD) technique (von Storch and Zwiers 1998). For each atmospheric field $\mathbf{Y}$ we compute a matrix $\mathbf{C}$, with element c_{ij} given by the correlation between the precipitation process at station i and the atmospheric variable Y at node j. This matrix is decomposed using the SVD method, to obtain:

$$\mathbf{C} = \mathbf{U}\mathbf{W}\mathbf{V}^T. \tag{8}$$

Letting N denote the number of rain gauges and G the number of grid nodes, $\mathbf{U}$ is a $N \times N$ matrix, $\mathbf{V}$ is a $G \times N$ matrix and $\mathbf{W}$ is a diagonal $N \times N$ matrix. The SVD technique ensures that $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$ and the diagonal elements of $\mathbf{W}$, $w_1, \ldots, w_N$, are the singular values of the matrix $\mathbf{C}$, in non-increasing order. If we standardize the atmospheric variable Y at each node j separately and call the resulting field $\mathbf{Y}^{std}$, we can construct a summary of the original field by multiplying $\mathbf{Y}^{std}$ by the ith column of the matrix V, $V^{(i)}$. This summary variable explains $\frac{w_i^2}{\sum_k w_k^2}$ of the correlation between the precipitation process and the atmospheric field $\mathbf{Y}$. The number of summary variables needed to explain a certain portion of the correlation depends on the relative magnitude of the singular values.

Once the SVD procedure has been applied to each of the atmospheric fields under consideration, the decision on how many and which of the resulting summary variables are to be included in the model is based on BIC.

3 An application

We used the NHMM to analyze precipitation amounts at a network of rain gauges in Washington state. The precipitation dataset consists of daily precipitation amounts for the winters (November through March) 1973–1990, at the 24 rain gauges shown in Figure 1. These data were recorded by the National Weather Service and cooperators and corrected by the National Climatological Datacenter (NCDC) 'Validated Historical Daily Data' project. The 12 winters 1973–1985 were used for model fitting, while the winters 1985–1990 were reserved for model validation. The atmospheric data consists of daily geopotential height at 1000 and 850 mb, temperature at 850 mb and relative humidity at 1000 and 850 mb from the

NCAR/NCEP Reanalysis project, provided through the NOAA Climate Diagnostic Center. These variables are given on a $2.5°$ latitude $\times$ $2.5°$ longitude grid for the same period as the precipitation data. The area of interest spans 48 grid nodes.

`insert figure 1`

The model fitting procedure was hierarchical. The number of weather states was first determined by fitting HMMs with 2 through 7 states to the occurrence data. Several considerations contributed to the decision to include 6 states. The Bayesian information criterion suggested a "large" number of states, since BIC decreased monotonically as the number of states increased, actually pointing at the 7 state model. However, the 7th state did not seem to improve the fit of the model to the observed storm duration distribution or any other important feature of the data. Thus we focused the remainder of our model building efforts on the 6 state model.

Atmospheric variables were added to the 6 state model after performing the SVD decomposition on each of the 5 fields – geopotential height at 1000 and 850 mb, temperature at 850 mb and humidity at 1000 and 850 mb – to summarize the 48 grid values into a few quantities. Very few summary variables are sufficient to explain most of the correlation between each field and the precipitation process, as shown in Table 1.

`insert table 1`

As an example of the type of summary variables obtained with the SVD technique, Figure 2 shows a contour plot of the weights assigned to each grid node to form the first linear combination variable for geopotential height at 1000 mb. The weights are highest just at the northwest of Washington state, and decay in all directions as the nodes get further away from this region. The resulting summary variable can be interpreted as a weighted mean of the standardized 1000 mb geopotential height field.

`insert figure 2`

Several NHMMs with 6 states were fit using different combinations of the selected summary variables, and BIC was used to choose the best model. The model that minimizes BIC contains two atmospheric variables: the first summary variable for geopotential height at 1000 mb and the first summary variable for humidity at 850 mb.

A NHMM with 6 states and including the first summary variables for geopotential height at 1000 mb and relative humidity at 850 mb was then fit to the precipitation amounts. The 6 states identified by the NHMM correspond to the precipitation patterns in Figure 3a. States 1 and 6 are clear cut wet and dry respectively, for all the stations in the network. The other states correspond to intermediate patterns that reflect regional differences. Fitting the 6 state NHMM to occurrences only, leads to very similar patterns, indicating that the inclusion of amounts does not seem to substantially change the state definitions, in terms of precipitation probabilities. The 6 weather states also correspond to different amount distributions. For each state, Figure 3b shows the distribution of the positive precipitation amounts at Puyallap, in the South Puget region. Larger amounts correspond to the predominantly wet states, especially state 1 where the precipitation probability is large at all stations. In predominantly dry states, when precipitation occurs the amounts tend to be smaller. State 4, which is dry in Eastern Washington and relatively wet around Puget Sound, corresponds to smaller amounts with respect to the first three states, even at the stations where the precipitation probability remains fairly large.

`insert figure 3`

The Viterbi algorithm (Juang and Rabiner 1991) identifies the most probable sequence of states, so that each day is classified into one of the states defined by the NHMM. The resulting relative frequencies of the weather states are 14%, 15%, 16%, 18%, 20% and 17%. Averaging the geopotential height at 1000 mb field over all days classified into a particular state gives the predominant pattern associated with that state. The same procedure leads to the predominant 850 mb relative humidity pattern associated with each of the 6 weather states. One may compare these atmospheric patterns to the corresponding precipitation patterns in Figure 3a. Figure 4 shows the contour plots for geopotential height at 1000 mb and relative humidity at 850 mb for all the 6 states. State 6 is characterized by a high pressure system and low relative humidity over the Washington region, which correspond to low precipitation probability. In state 1, low pressure at the northwest of Washington state and high moisture over the entire region correspond to the high precipitation probability at all stations. The other atmospheric patterns are consistent with the observed precipitation patterns and suggest that some of the weather states might be regarded as “transition states”. For example, we find that state 5 typically transitions to either state 4 or state 6 with approximately equal probability.

`insert figure 4`

Some indications of how well the NHMM fits the data derives from the comparison between observed and model-based precipitation probabilities (Figure 5a), and between observed and model-based log odds ratios (Figure 5b). The precipitation probabilities are reproduced well, while the log odds ratios, which reflect the spatial correlation between occurrences at each pair of stations, are modeled less adequately, especially when the observed correlation is high. This indicates that the hypothesis of conditional spatial independence, given the weather state, may need to be modified. The common weather state seems to explain much of the correlation, but additional unexplained local spatial correlation remains. A similar conclusion is suggested by Figure 5c, which shows the Spearman correlation coefficient corresponding to the precipitation amounts at each station pair. A relatively low spatial correlation between amounts, as well as between occurrences, can be adequately captured by the common weather state, but when the correlation between gauges is strong, the weather state is not sufficient to account for all of it.

`insert figure 5`

Another issue is whether the gamma distribution is an appropriate choice to model the conditional distribution of precipitation amounts, given occurrence and the weather state. The fit varies from station to station; Figure 6 shows qqplots of observed versus model-based precipitation amounts at three representative stations from three geographical regions in Washington state. In general the distribution of precipitation amounts is best modeled at the stations in the South Puget area, while the North Puget stations correspond to the worst fit. The Eastern Washington region, which is the driest area, shows the largest variability in fit from gauge to gauge.

`insert figure 6`

Plots similar to those in Figures 5 and 6 were obtained using the reserved data. The final 6 state NHMM (which was fit using 1973–1985 data), together with geopotential height at

1000 mb and relative humidity at 850 mb for the 1985–1990 period, were used to generate precipitation amounts for the 1985–1990 winters. The SVD weights obtained previously were applied to form the summary atmospheric variables from the ’85–’90 geopotential height and humidity fields. The comparison of various statistics for the observed and generated ’85–’90 precipitation amounts indicates how well the model captures the characteristics of the reserved data. Figure 7a shows the observed versus model-based precipitation probabilities at all stations. The model underestimates the precipitation probability at most stations. Since the rain gauges are not independent, it is reasonable to expect, for any given 5 year realization, that the observed precipitation probabilities will be mostly smaller or mostly larger than the model-predicted precipitation probabilities. Therefore, it is not too surprising that most of the points in Figure 7a lie below the $y = x$ line. We have generated several 5 year realizations from the NHMM and compared the resulting ’observed’ precipitation probabilities with the ’model-based’ probabilities obtained by averaging over many sets of 5 year realizations. Even for these cases, where the observations are a realization from the model, the ’observed probabilities’ are typically mostly smaller or mostly larger than the ’model-based’ ones.

`insert figure 7`

Another characteristic of the reserved data that should be captured by the model is the conditional distribution of precipitation amounts, given occurrence. Figure 8 shows the qqplots of observed versus model-based precipitation amounts at the same stations as in Figure 6. The model seems to reproduce the distribution of observed precipitation amounts reasonably well overall, although the fit varies from station to station.

insert figure 8

4 Discussion

The model described in this paper can be used to generate simulations of precipitation amounts that incorporate synoptic atmospheric information. The hidden Markov model assumptions simplify the temporal and spatial structures to be parameterized, since the common *weather state* accounts for the temporal dependence and much of the spatial correlation between rain gauges. Several possible improvements to the model are currently under investigation, including more realistic spatial dependence structures and reduced parameterizations.

The conditional spatial independence structure adopted in the present application is relatively simple. Although this assumption captures most of the correlation between rain gauges, Figures 5 and 5c suggest the need to include some additional dependence in the model. We plan to investigate two alternative structures. The first step is to introduce dependence between precipitation *occurrences* and assume conditional spatial independence of *amounts* given occurrences and the weather state. The autologistic model of Hughes, Guttorp, and Charles (1999) can be adopted to describe the dependence of precipitation occurrences at different rain gauges. Precipitation amounts, conditional on occurrences, would be modeled independently at each gauge as in the previous sections. Thus the parameterization for the observed process becomes:

$$f_{\mathbf{R}_t|S_t=s}(\mathbf{r}) \propto \exp\left(\sum_{i=1}^{N} \lambda_{is} r^i + \sum_{j<i} \nu_{ijs} r^i r^j\right) \prod_{i=1}^{N} \Gamma\left(r^i; \alpha_{si}, \beta_{si}\right)^{1_{[r^i>c]}} \tag{9}$$

If the structure described by (9) still does not account for all the observed correlation between rain gauges, more complicated models which allow for interactions between the precipitation amounts at different stations will be considered. The spatial dependence between occurrences could still be described by the autologistic model and the amounts could be modeled jointly at all stations, through a multivariate gamma or exponential distribution.

The proposed modifications to the spatial dependence structure would increase the number of parameters, already large in the NHMM applied to the Washington state data. A possibility that will need to be investigated is the reduction of the number of parameters, both in the hidden and observed parts of the model. One reasonable modification of the hidden part is to have only one vector $\boldsymbol{\mu}_i$ of means of the atmospheric variables for each state i, regardless of the state of the system at the previous time point. In the observed part of the NHMM, one could specify some function of the precipitation amount parameters to have a common value at all stations within a sub-region.

Models like the NHMM can be used to study the effect of *climate variability*. Repeated GCM simulations under current climate conditions can constitute different realizations of the atmospheric fields included in the model. The NHMM can be used to generate occurrences and amounts for each realization, thereby downscaling the effect of the variability in the synoptic scale variables to precipitation. The effect of *climate change* is another issue that can be investigated using NHMMs. The output of GCM runs under altered climate conditions can serve as input into the downscaling model described here. Thus, the effects of the altered climate scenario could be downscaled to the local scale precipitation processes by generating precipitation occurrences and amounts from the NHMM. For this application of the NHMM to be valid, the relationship between the synoptic scale atmospheric variables and the local scale precipitation, as found under the model fitting conditions, would have to hold also under the altered climate. Charles, Bates, and Hughes (1999) discuss issues related to validation of downscaling models for studying climate change.

	gph 1000		gph 850		tem 850		hum 1000		hum 850	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
$\frac{w_i^2}{\sum_k w_k^2}$	0.95	0.03	0.96	0.03	0.84	0.11	0.91	0.07	0.96	0.03

Table 1: Percentage of correlation explained by the summary variables

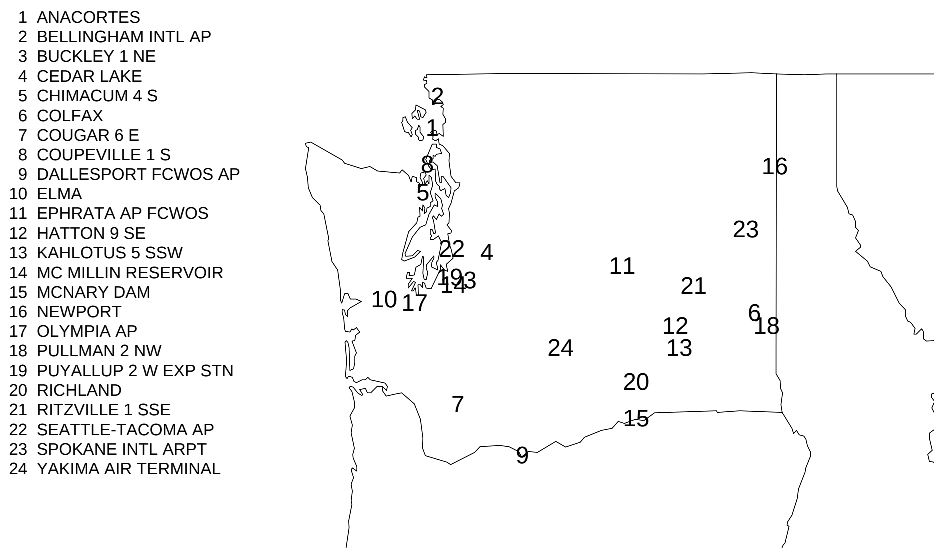


Figure 1: Map of the rain gauges

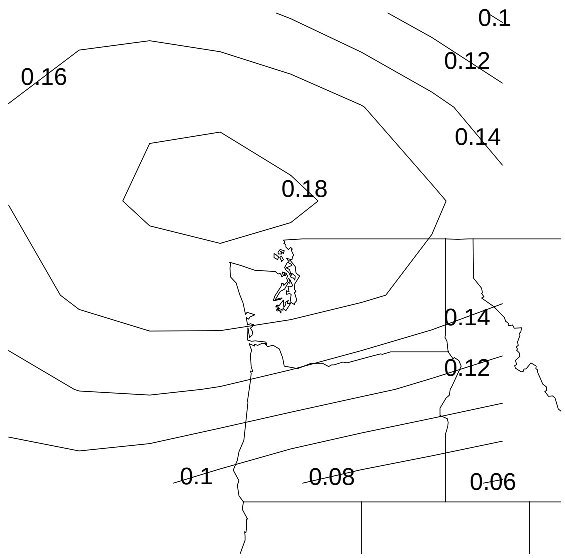
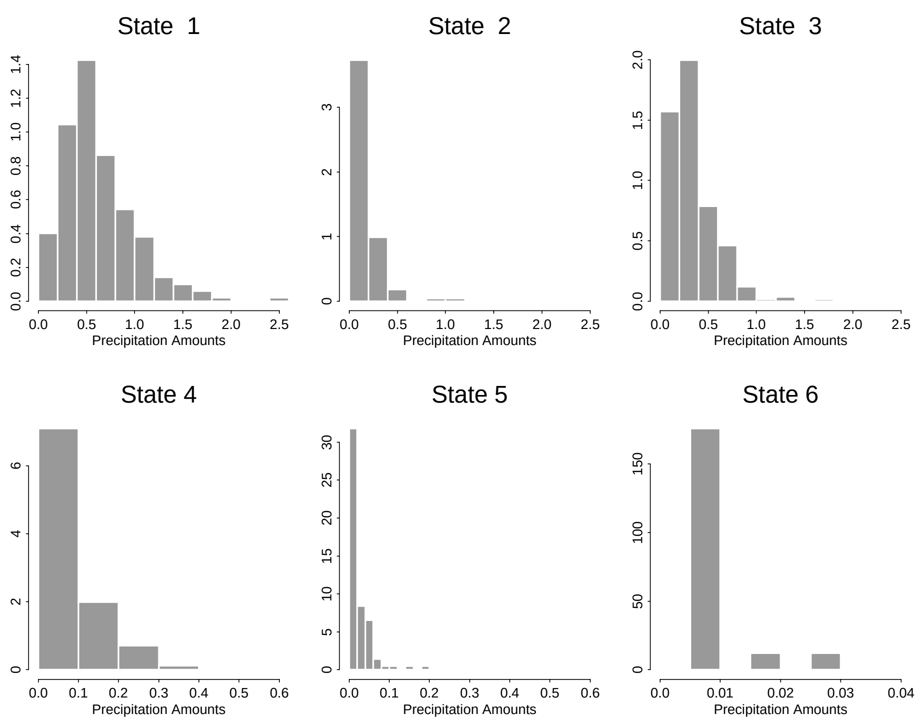


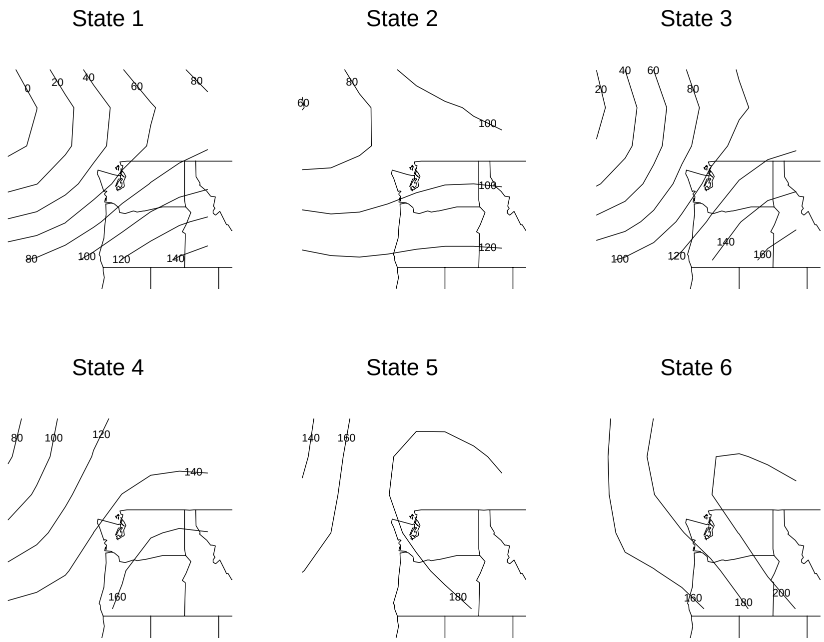
Figure 2: Contour plot of the weights for the first gph 1000 summary variable.

(a) Precipitation probabilities at the 24 rain gauges.

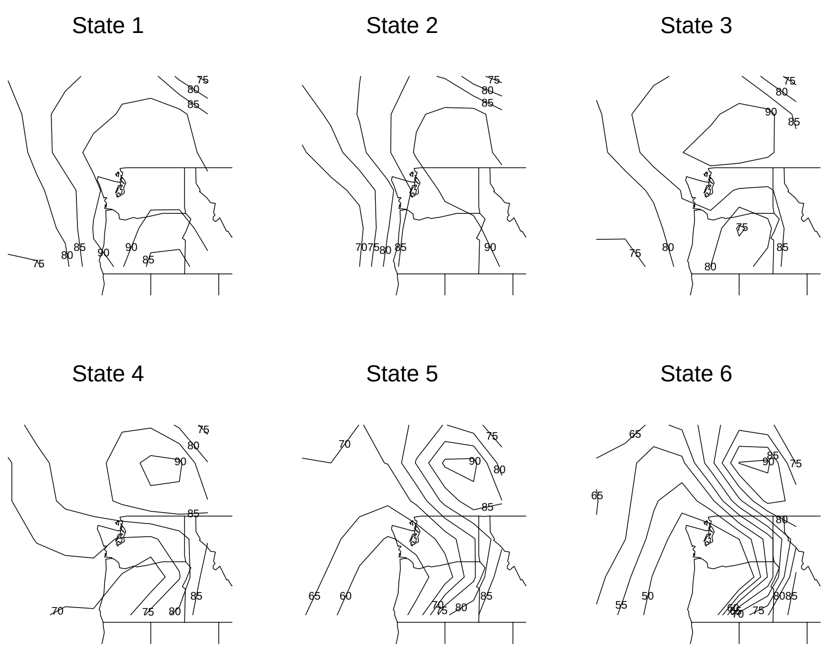


(b) Histograms of precipitation amounts at Puyallap (South Puget area)

Figure 3: Precipitation probabilities and histograms of amounts corresponding to the 6 weather states.

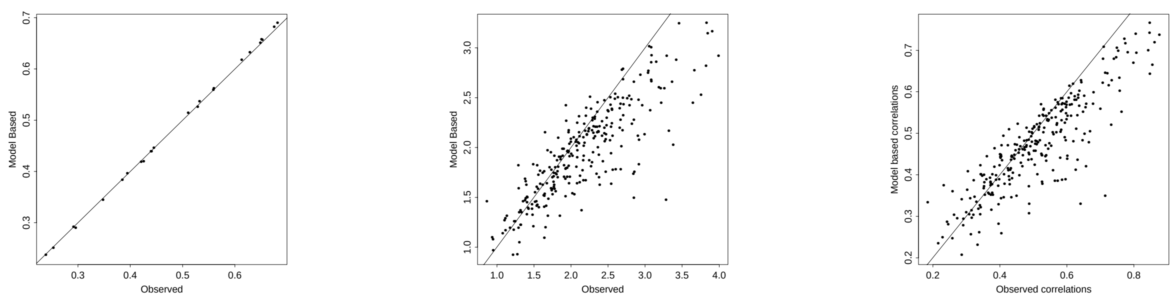


(a) geopotential height at 1000 mb



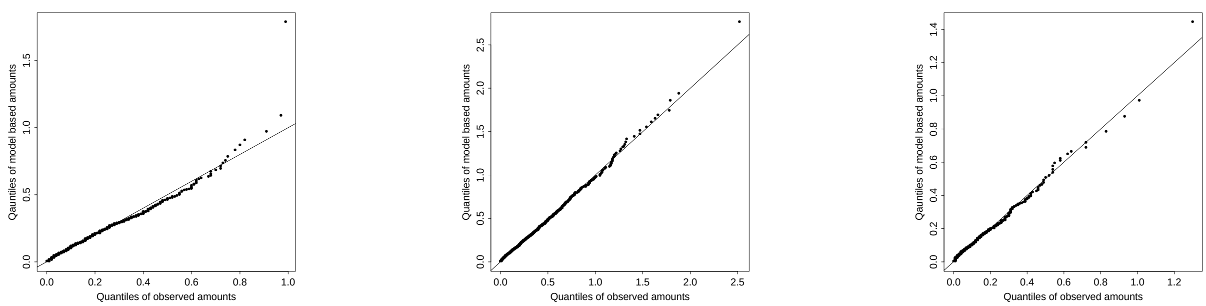
(b) humidity at 850 mb

Figure 4: Contour of the geopotential height at 1000 mb and humidity at 850 mb fields for each weather state.



(a) Precipitation Probability (b) Log odds ratio (c) Spearman coefficient

Figure 5: Observed versus model based precipitation probabilities, log odds ratios and correlations of positive amounts between all station pairs (Spearman coefficient). The model based quantities are obtained by simulating data from the 6 state NHMM for amounts.



(a) Coupeville (North Puget area)

(b) Puyallap (South Puget area)

(c) Yakima (Eastern Washington area)

Figure 6: Qqplots of observed versus model-based amounts at selected stations. The model based amounts are simulated from the 6 state NHMM.

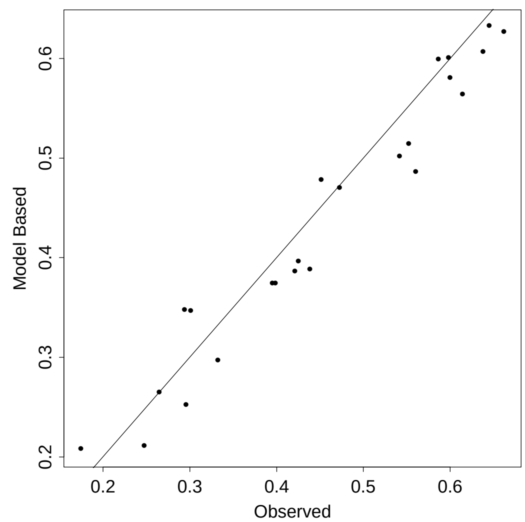
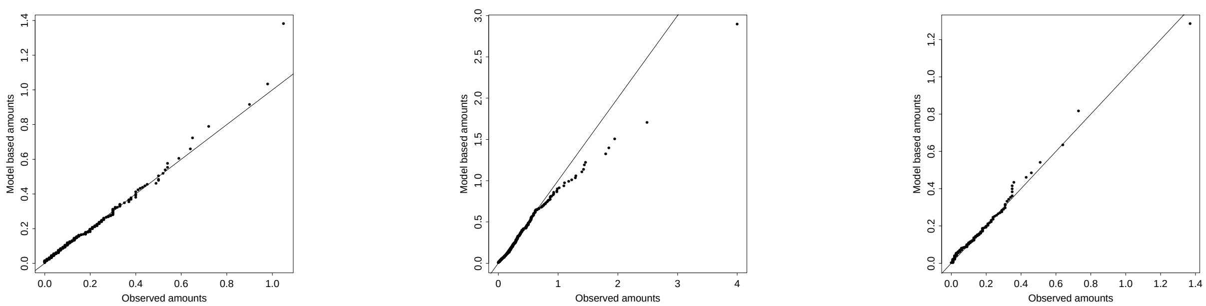


Figure 7: Observed versus model-based precipitation probabilities for the reserved period, 1985–1990



(a) Coupeville (North Puget area)

(b) Puyallap (South Puget area)

(c) Yakima (Easter Washington area)

Figure 8: Qqplots of observed versus model-based amounts at selected stations for the reserved period, 1985–1990.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*, 716–723.

Bardossy, A. and E. J. Plate (1992). Space-time models for daily rainfall using atmospheric circulation patterns. *Water Resources Research 28*, 1247–1259.

Bartholy, J., I. Bogardi, and I. Matyasovszky (1995). Effect of climate change on regional precipitaion in Lake Balaton watershed. *Theor. Appl. Climatol. 51*, 237–250.

Baum, L. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities 3*, 1–8.

Charles, S., B. Bates, and J. P. Hughes (1999). A spatio-temporal model for downscaling precipitation occurrence and amounts. *Journal of Geophysical Research*. To appear.

Gabriel, K. R. and J. Neumann (1992). A Markov chain model for daily rainfall occurrences at tel-aviv. *Q. J. R. Metereol. Soc. 88*, 85–90.

Giorgi, F. and L. O. Mearns (1991). Approaches to the simulation of regional climate change: a review. *Rev. Geophys. 29*, 191–216.

Goodall, C. R. and M. J. Phelan (1991). Edge-preserving smoothing and the assessment of point process models for Gate rainfall fields. In N. U. Prabhu and I. V. Basawa (Eds.), *Statistical Inference in Stochastic Processes*, pp. 35–66. Marce Dekker, inc.

Hay, L., J. McCabe, D. M. Wolock, and M. A. Ayers (1991). Simulation of precipitation by weather type analysis. *Water Resources Research 27*, 493–501.

Hobbs, P. V. and D. Locatelli (1978). Rainbands, precipitation cores and generating cells in a cyclonic storm. *Journal of Atmospheric Sciences 35*, 230–241.

Hughes, J. P. and P. Guttorp (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research 30*(5), 1535–1546.

Hughes, J. P., P. Guttorp, and S. Charles (1999). A nonhomogeneous hidden Markov model for precipitation occurrence. *Journal of the Royal Statistical Society 48*, 15–30.

Hughes, J. P., D. P. Lettenmaier, and P. Guttorp (1993). A stochastic approach for assessing the effects of changes in regional circulation patterns on local precipitation. *Water Resources Research 29*, 3303–3315.

Juang, B. H. and L. R. Rabiner (1991). Hidden Markov models for speech recognition. *Technometrics 33*, 251–272.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Le Cam, L. M. (1961). A stochastic description of precipitation. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Sattistics and Probability*, Volume 3, pp. 165–186. University of California press, Berkeley.

Stern, R. D. and R. Coe (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society A 147*, 1–34.

von Storch, H. and F. Zwiers (1998). *Statistical analysis on climate research.* New York: Cambridge University Press. In Print.

Waymire, E., V. K. Gupta, and I. Rodriguez-Iturbe (1984). A spectraltheory of rainfall intensity at the meso-β scale. *Water Resources Research 20*, 1453–1465.

Woolhiser, D. A. (1992). Modeling daily precipitation – Progress and problems. In P. Guttorp and A. Walden (Eds.), *Statistics in the Environmental & Earth Sciences*. Griffin, London.

Zucchini, W. and P. Guttorp (1991). A hidden Markov model for space-time precipitation. *Water Resources Research 27*(8), 1917–1923.