

**Hierarchical Continuous Time
Markov Chain Models
for
Threshold Exceedance**

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the requirements for the Degree

Doctor of Philosophy (Systems and Information Engineering)

by

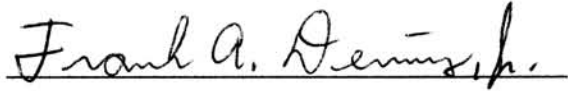
Frank Allen Deviney, Jr.

August 2009

© Copyright by
Frank Allen Deviney, Jr.
All rights reserved
August 2009

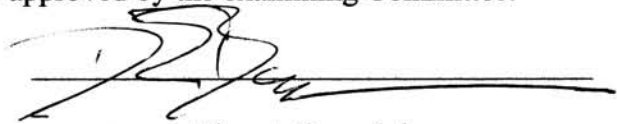
APPROVAL SHEET

The dissertation is submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy (Systems and Information Engineering)

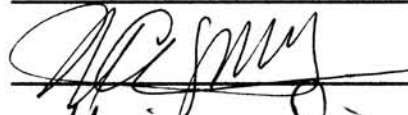


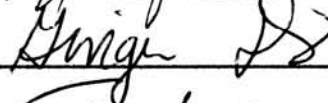
AUTHOR

This dissertation has been read and approved by the examining Committee:



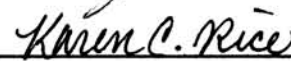
Dissertation advisor



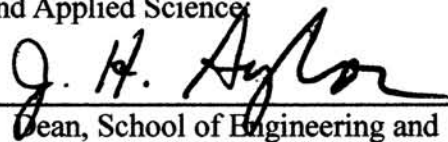








Accepted for the School of Engineering and Applied Science:



Dean, School of Engineering and
Applied Science

August 2009

Acknowledgements

In no particular order (other than the first), I would like to acknowledge:

- Karen Siegrist, the love of my life, whom I met after I had just started on this project. She probably thought I was never going to finish, and is now wondering what's going to happen next? As am I.
- The University for its decision and ongoing efforts to provide educational benefits to its employees. Hard to imagine where I would be if they had not had the vision.
- My family. Extended, four-footed, and otherwise.
- My friends, for arranging fishing trips. Sometimes I was on them, sometimes just participated vicariously.
- My high school friend and fellow math aficionado, Professor Jean Turner, who kept telling me, “the best thesis is a filed thesis”.
- Rick Webb and Professor Jim Galloway, for support as my employers.
- My committee, especially Professor Steve Patek and my advisor Professor Don Brown, for their support navigating the process.
- The SEAS Graduate Studies Committee, for providing a final motivational boost for getting it done.

Abstract

Thresholds have been defined for many water quality indicators (WQIs) which separate the measurement space of the indicator into two states, one of which, the exceedance or violation state, has undesirable consequences. Observations are often made at unevenly spaced intervals, are usually uncoordinated with the timing of state changes, and are usually made asynchronously at multiple locations. These typical observation protocols have hindered estimation of violation-state properties. To address this problem, six hierarchical two-state continuous-time Markov chain (CTMC) models were developed and tested. These allow estimation of duration, frequency, and limiting probabilities from asynchronous, uncoordinated, and unevenly spaced observations. Three of these models were developed for single Markov processes but can be modified to handle multiple processes. Three of the models were developed for multiple processes. Two of the models were homogeneous; the other four were non-homogeneous with sinusoidally varying components. Model parameters were estimated with Bayesian MCMC methods.

In each of three experiments, processes were simulated at high-frequency time steps. Asynchronous, infrequent, uncoordinated, and unevenly spaced observations of these processes were then extracted using protocols specified with varying observation

period length, quasi-regular observation interval, and violation-state observation probability. Models were estimated from the simulated observations, and compared on nominal parameter value recovery, predictive performance, and frequency and duration distribution error. Effects of process and observation protocol characteristics on recovery, prediction performance and distribution estimation error were measured.

In the first experiment, simulated observations of single-chain two-state CTMCs were made and modeled. First, choice of prior distribution model was evaluated. Uniform and Gamma priors were found to be roughly equivalent in terms of performance, and both were found to perform substantially better than a Jeffrey's prior. Next, recovery, prediction, and distribution estimation error were evaluated. Duration, frequency, and violation-state probability were overestimated. Lower distribution estimation error was associated with longer observation period and more observations. Lower prediction and distribution estimation error was associated with more non-homogeneous processes.

In the second experiment, observation and modeling of multiple correlated WQI processes was simulated by mimicking WQIs with dual correlated two-state continuous time Markov chains. Estimates were made both jointly and individually, using the homogeneous model from the first experiment modified for multiple chains. Duration, frequency, and long-term violation-state probability were overestimated. Joint and individual estimates produced nearly equal results. Positively correlated and relatively low transition-rate processes were more-accurately predicted. Several observation

characteristics were related to better prediction: greater event observation intensity, greater quasi-regular observation intensity, and longer observation period.

In the third experiment, two methods were compared for estimating threshold exceedance frequency and duration properties. One method was adapted from the Partial Duration Series (PDS) method popular in flood frequency analysis. The second method was based on three multiple-chain CTMC models. Simulations of WQI time series were generated using a sinusoidal model with autocorrelated errors adapted from the literature. Duration and violation-state probability were overestimated. Frequency was underestimated. A multiple-chain homogeneous CTMC model produced lower error estimates of frequency and duration than did the PDS method. Results were mixed for the two non-homogeneous multiple-chain CTMC models. For all CTMC models considered, more-positively correlated processes were easier to predict. Higher observation rates were found to improve predictive performance. The multiple-chain CTMC models were shown to be extend-able to allow for prediction of frequency and duration properties from watershed characteristics or to allow these properties to vary with time.

The bias in recovery of nominal parameter values seen in all three experiments appeared to be related to the observation protocol characteristics and not to the models or estimation method. The sources of this bias were not fully investigated.

Table of Contents

1	Introduction.....	13
2	Literature Review.....	17
2.1	Flood-frequency analysis and methods.....	17
2.2	Applications of flood-frequency methodology to WQIs.....	21
2.3	Markov Chains.....	22
2.4	Time series models.....	23
2.5	Bayesian inference.....	24
3	Continuous-time Markov models for threshold violations.....	26
3.1	Single-process CTMC – Model #1 (homogeneous).....	27
3.2	Single-process CTMC with varying rate – Model #2 (non-homogeneous).....	35
3.3	Single-process CTMC with varying limiting probabilities - Model #3 (non-homogeneous).....	37
3.4	Correlated Continuous Time Markov Chains - Model #4.....	39
3.5	Multiple-process CTMC Model #1 (homogeneous).....	43
3.6	Multiple-process CTMC Model #2 (non-homogeneous).....	44
3.7	Multiple-process CTMC Model #3 (non-homogeneous).....	46
4	Single-chain Markov models for threshold exceedance.....	48
4.1	Abstract.....	48
4.2	Introduction.....	49
4.2.1	Model #1: homogeneous two-state CTMC.....	51
4.2.2	Model #2: varying transition rates.....	54
4.2.3	Model #3: varying limiting probabilities.....	54
4.3	Method.....	57
4.3.1	Experimental design.....	57
4.3.2	Process Simulation.....	58
4.3.3	Observation simulation.....	60
4.3.4	Modeling simulation.....	61
4.3.5	Determination of metrics.....	64
4.3.6	Comparison of priors.....	66
4.3.7	Recovery of nominal parameter values.....	66
4.3.8	Evaluation of effect of characteristics on metrics.....	67
4.4	Results.....	68
4.4.1	Comparison of priors.....	69
4.4.2	Nominal parameter recovery.....	72
4.4.3	Predictive performance.....	79
4.4.4	Effects of process and measurement process characteristics on predictive performance.....	80
4.5	Discussion.....	86
4.6	Conclusions/Recommendations.....	92

5 Multiple-chain Markov models for threshold exceedance.....	96
5.1 Abstract.....	96
5.2 Introduction.....	97
5.2.1 Correlated CTMCs.....	99
5.2.2 The two-state CTMC Model.....	101
5.3 Method.....	104
5.3.1 Experimental design.....	105
5.3.2 Process simulation.....	106
5.3.3 Observation simulation.....	109
5.3.4 Modeling simulation.....	110
5.3.5 Determination of metric values.....	111
5.3.6 Methods of evaluation.....	113
5.4 Results.....	113
5.4.1 Recovery of nominal parameter values.....	114
5.4.2 Overall performance.....	119
5.4.3 Method comparison.....	120
5.4.4 Effect of simulation characteristics on AUC	125
5.5 Discussion.....	137
5.6 Conclusions.....	140
6 CTMC models for threshold exceedance in water-quality indicators.....	142
6.1 Abstract.....	142
6.2 Introduction.....	143
6.2.1 Partial duration series method.....	146
6.2.2 CTMC Model #1 (homogeneous).....	147
6.2.3 CTMC Model #2 (non-homogeneous).....	150
6.2.4 CTMC Model #3 (non-homogeneous).....	151
6.3 Method.....	154
6.3.1 Experimental design.....	154
6.3.2 Process simulation.....	155
6.3.3 Nominal property determination.....	160
6.3.4 Observation simulation	162
6.3.5 Modeling simulation	164
6.3.6 Metric determination.....	166
6.3.7 Evaluation method.....	168
6.4 Results.....	169
6.4.1 Recovery of nominal parameter values.....	169
6.4.2 Overall predictive performance.....	177
6.4.3 Intra-method comparison.....	178
6.4.4 Process and observation protocol characteristic effect on AUC.....	180
6.5 Discussion.....	188
6.6 Conclusions.....	196
7 Conclusions.....	199

8	References.....	203
9	Appendix A.....	209
9.1	Single-process CTMC – Model #1 (homogeneous).....	209
9.2	Single-process CTMC with varying rate – Model #2 (non-homogeneous).....	210
9.3	Single-process CTMC with varying limiting probabilities - Model #3 (non-homogeneous).....	211
9.4	Multiple-process CTMC – Model #1 (homogeneous), simple version.....	212
9.5	Multiple-process CTMC Model #1 (homogeneous), hierarchical version.....	213
9.6	Multiple-process CTMC Model #2 (non-homogeneous).....	214
9.7	Multiple-process CTMC Model #3 (non-homogeneous).....	215

Table of Figures

Figure 1: Two-state homogeneous CTMC.....	28
Figure 2: Simulation of a 2-state homogeneous CTMC versus time.....	29
Figure 3: CDF and PDF of the renewal interval of a two-state homogeneous CTMC.....	33
Figure 4: Prior distributions for rate parameters.....	35
Figure 5: Correlated dual process CTMC.....	40
Figure 6: Two-state CTMC.....	52
Figure 7: Prior distributions of rate parameters.....	63
Figure 8: Predicted recovery rate along steepest path gradient.....	75
Figure 9: Steepest ascent gradient vs. process/observation characteristics I.....	77
Figure 10: Steepest ascent gradient vs. process/observation characteristics II.....	78
Figure 11: Steepest ascent gradient vs. process/observation characteristics III.....	78
Figure 12: Nominal versus estimated distributions of duration and renewal period.....	80
Figure 13: Correlated dual process CTMCs.....	100
Figure 14: Distributions of process and observation protocol characteristics.....	108
Figure 15: Predicted recovery rate along steepest path gradient.....	117
Figure 16: Steepest ascent gradient vs. process/observation characteristics.....	119
Figure 17: AUC from individual determination vs. joint determination.....	121
Figure 18: DIC from individual determination vs. joint determination.....	122
Figure 19: Effective number of parameters.....	124
Figure 20: Observation characteristic interaction effects contours.....	129
Figure 21: Observation characteristic interaction contour, continued.....	131
Figure 22: Observation/process characteristic mixture interaction contours.....	133
Figure 23: Observation/process characteristic mixture interaction contours, continued.....	134
Figure 24: Process characteristics-only interaction effects contours.....	136
Figure 25: Two-state CTMC.....	148
Figure 26: Simulation of a sinusoid with error.....	156
Figure 27: Simulation (left) and ACF (right) of a sinusoid with error.....	157
Figure 28: Simulations of correlated WQIs.....	159
Figure 29: Nominal distributions of duration and renewal period for two processes.....	161
Figure 30: Simulated observations of state	163
Figure 31: Probability of recovery along steepest path.....	172
Figure 32: Steepest ascent gradient vs. process/observation characteristics I.....	174
Figure 33: Steepest ascent gradient vs. process/observation characteristics II.....	175
Figure 34: Steepest ascent gradient vs. process/observation characteristics III.....	176
Figure 35: Effect of error correlation on expected renewal interval.....	177
Figure 36: ISE distributions by process and method.....	179
Figure 37: Second-order effects I.....	184
Figure 38: Second-order effects II.....	186
Figure 39: Second-order effects III.....	188

Table of important symbols

Symbols	Description
λ, μ, γ	λ and μ are transition rates. γ is the renewal rate.
X, Y, Z	X and Y are random variables for the times spent in state '0' and state '1' of a two-state continuous time Markov chain. Z is a random variable for the renewal period.
S_n	The state of the process at observation n
P_0, P_1	Limiting probabilities for state '0' and state '1'
A, B	Coefficients of a sinusoid, as in $A \sin(\theta) + B \cos(\theta)$
K	A constant relating λ and μ , as in $\mu = K \lambda$
τ	Generally a precision for a normal distribution
ω	A radian

1 Introduction

At a very general level, managers often seek information in the form of assessments of the risk to valuable resources from identified threats. In the natural resource management field, and in particular in the aquatic natural resource field, threats to the biological resource arise from frequent and/or lengthy exposure to chemical or physical changes in the aquatic environment (Baldigo, Lawrence 2000, Baldigo, Murdoch 1997, Bulger, Cosby & Webb 2000, Davies et al. 1992, DeWalle, Swistock & Sharpe 1995, Laio et al. 2001, Sickie et al. 1996). A large number of chemical and physical water quality indicators (WQIs) are monitored and for many of them, thresholds have been determined for which violation of the threshold implies risk of significant ecological harm. Quantification of such risk provides the kind of information management seeks for decision support.

A map color-coded on a scale from green (no risk) to red (high risk) that conveys the risk of exposure to a given threat or combination of threats fulfills the conceptual information requirement for management. For example, a resource manager might ask, “Do I have any red areas, and if so, how much of my area is colored red?” Or more specifically, “What is the probability of a consecutive four-year sequence of springtime

violation events of at least three days duration occurring at least once in the next 50 years?” If this probability can be estimated then this map can be made.

Depending on the parameter, a violation period may be defined as either a period where the parameter level is above the threshold (an exceedance) or a period where the parameter is below the threshold (but not both). Risk arises not only from the overall long-term probability of violation, but also from the frequency and duration of such violations. For example, violations might occur on average once per year and last on average 6 months, or they might occur on average once per day and last on average 12 hours. Both scenarios yield the same long-term probability of violation, yet have potentially different ecological implications. Seasonal differences in violation risk are also important, depending on the extent to which they coincide with critical life stages of organisms.

If there were a high density of locations where WQIs were observed frequently at evenly spaced intervals of time, then methodology already developed for characterizing flood frequency and magnitude might be adapted for characterizing violation duration, frequency, and long-term probability. First of all, however, many parameters of interest are not practical to observe in situ. Typically what we have is a sparse network of sites where WQIs are observed at asynchronous, infrequent, uncoordinated (with threshold crossings), and unevenly spaced intervals. For example, whereas discharge is often observed at 15 minute intervals, WQIs are often observed on a monthly or weekly schedule possibly with occasional high-frequency observations made during high-water

events. It is unusual to find locations where the total number of WQI observations made in one year exceeds 100, except for those few locations where in-situ probes are used to measure a few parameters such as pH, dissolved oxygen, temperature, and conductivity.

Secondly, flood-frequency methods, and in general other existing methods for modeling threshold violation behavior are based on extreme value theory. These approaches typically assume that thresholds are important only with respect to the variance of the process, whereas defined thresholds are typically absolute and without respect to process variance. Violations of defined thresholds may in fact be rather commonplace, and within a region of interest, a specific threshold may have a different relationship with respect to the variance of the process at one location than with that at another.

With respect to regional assessment of risk, one might want to leverage information from surrounding sites by exploiting common characteristics or other factors leading to correlation among them. And while it is reasonable to think that time series of WQIs at different locations might be correlated, the lack of data collected simultaneously makes traditional estimates of correlation impractical. Nor is it then a simple matter to construct multivariate time-series models that leverage that correlation.

The broad objective of this dissertation was to develop and evaluate models for estimating the frequency and duration of violation events and their long-term probability, given arbitrary thresholds and availability only of asynchronous, infrequent, uncoordinated, and unevenly spaced observations. This broad objective was approached

through the pursuit of three subordinate objectives: 1) to develop and test models for single sites, 2) to develop and test models for multiple-sites, and 3) to develop and compare multiple-site models with an adapted flood-frequency approach.

Clearly the universe of possible models to consider is too large to undertake a complete assessment in one dissertation. While the notion that many important ecological processes are non-stationary in nature (Milly et al. 2008) is a compelling one, as a starting point only stationary, or to be more accurate seasonally stationary, models are considered in this work. It is expected that insights gained from this work will assist future researchers in developing models for this problem that address non-stationarity, and thereby facilitate attainment of the broad objective.

The remainder of this dissertation is organized as follows. In Chapter 2 relevant literature is reviewed related to the problem, which comes from the fields of stochastic processes, extreme value theory, experimental design, and flood-frequency estimation. In Chapter 3, a number of models and estimation methods are described. In Chapter 4 three models for single site processes are developed and tested. In Chapter 5 one of the single site models is extended to permit multiple-site models and the single-site model is compared to the multiple-site model. In Chapter 6 additional multiple site models are developed and are compared with an adaptation to the Peaks Over Threshold method from the flood-frequency literature. In Chapter 7 conclusions and contributions from the research are presented.

2 Literature Review

Relatively little work has been done in the specific target application area. However, considerable work has been done in the broader area of threshold violation. In particular, hydrologists have worked for decades to couple extreme value theory with the theory of stochastic processes, motivated by the need to be able to predict the likelihood of occurrence of disastrous but infrequent and random events within typically long management horizons such as the lifetime of a structure. For these analyses, hydrologists are generally blessed with the availability of relatively high-frequency, evenly spaced observations of water flow. In this section a few of the more-popular flood-frequency methods, Markov chains, time series methods, and other subjects related to the problem are reviewed.

2.1 Flood-frequency analysis and methods

A number of flood-frequency methods have been developed over the decades. In general, they build on theories of extreme values and Poisson processes to yield distributions for the frequency and magnitude of flood exceedance events. A popular

legacy method referred to as the Annual Maxima Series (AMS) method builds distributions from the annual maximum flow level observed. An increasingly popular newer method is commonly referred to as either the Partial Duration Series (PDS) or the Peaks Over Threshold (POT) method (Madsen, Pearson & Rosbjerg 1997, Madsen, Rasmussen & Rosbjerg 1997, Madsen, Rosbjerg 1997, Rosbjerg, Funder Rasmussen & Madsen 1991, Rosbjerg, Madsen & Rasmussen 1992, Wang 1991 and references therein). In this method a threshold of discharge is specified. Various suggestions have been made, but generally these thresholds end up being in the upper tails of the distribution for discharge. For example, (Rasmussen, Rosbjerg 1991) suggested the mean plus 3-3.5 times the standard deviation. In setting the threshold, the objective is generally to obtain 1-2 exceedance events per year. Up-crossings of the threshold demarcate the onsets of exceedances and down-crossings demarcate their ends. The magnitude of an exceedance may be defined as the peak exceedance value, the area above the threshold associated with the exceedance, or some other function of the exceedance period. The number of exceedance events per year is thought to be Poisson-distributed. The associated magnitudes are thought to have a generalized Pareto (GP) distribution. The GP distribution implies that annual maxima will have a generalized extreme value (GEV) distribution. This relates the PDS method to the AMS method. In practice data-reduction rules must be implemented because exceedance events do not generally obey the Poisson assumption of independence.

The Poisson probability of n events in time t is given by

$$P[N(t)=n]=\frac{(\lambda t)^n}{n!}e^{-\lambda t}, n=0,1,2,\dots \quad (1)$$

where λ is the rate per unit of time and the estimator of λ is

$$\hat{\lambda}=\frac{N}{t} \quad (2)$$

The GP distribution for discharge (Madsen, Rasmussen & Rosbjerg 1997, Madsen, Rosbjerg 1997) has one of two possible cumulative distribution functions (CDFs)

$$F(q)=1-e^{\frac{-q-q_0}{\alpha}}, \kappa=0 \quad (3)$$

$$F(q)=1-\left(1-\kappa\frac{q-q_0}{\alpha}\right)^{1/\kappa}, \kappa\neq 0 \quad (4)$$

where α is the scale parameter, κ is the shape parameter, and q_0 is the discharge threshold level. Several methods have been suggested for estimating the GP distribution parameters. The Method of Moments (Hosking, Wallis 1987) (MOM) estimators are

$$\hat{\alpha}=\frac{1}{2}\hat{\mu}\left(\frac{\hat{\mu}^2}{\hat{\sigma}^2}+1\right) \quad (5)$$

$$\kappa = \frac{1}{2} \left(\frac{\hat{\mu}^2}{\hat{\sigma}^2} - 1 \right) \quad (6)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the sample mean and sample variance, respectively. Other estimators include the probability-weighted moments (PWM) estimators and the maximum likelihood (ML) estimators (Hosking, Wallis 1987). The significance of a negative $\hat{\kappa}$ is that theoretically the range of duration is unbounded above, while a positive $\hat{\kappa}$ indicates a bounded distribution. The mean and variance are given by (Madsen, Rosbjerg 1997)

$$\mu = E[q - q_0] = \frac{\alpha}{1 + \kappa} \quad (7)$$

and

$$\sigma^2 = \text{var}[q - q_0] = \frac{\alpha^2}{(1 + \kappa)^2 (1 + 2\kappa)} \quad (8)$$

The median duration is given by the value of d that yields $0.5 = F(d)$. The formula for this value is:

$$\text{median} = \frac{\alpha}{\kappa} (1 - 2^{-\kappa}) \quad (9)$$

2.2 Applications of flood-frequency methodology to WQIs

Greb and Graczyk (1995) used flood-frequency methods to analyze continuously collected (every fifteen minutes for one year) dissolved oxygen (DO) data from two streams in Wisconsin. They applied the AMS method, using periods of one week instead of years. They found that confidence intervals and errors of CDFs of DO were substantially affected by sub-sampling at a rate of one or two samples per day. They demonstrated the utility of their analysis for predicting recurrence probabilities for low-DO conditions likely to cause mortality in smallmouth bass larvae.

Behera and Adams (2000) performed a frequency analysis by constructing PDFs of one year's event mean concentration data from urban runoff events from two sewer catchments in Toronto, Canada. For fifteen different water quality parameters, they found that event mean concentrations could be modeled using either the exponential, gamma, or lognormal distributions. They compared goodness of fit for these distributions using the Kolmogorov-Smirnov test. They extended their analysis to a regional subpopulation of sixteen waterfront outfalls along the Metropolitan Toronto Waterfront (the other fourteen were only sampled during the Fall) by an aggregation of data.

Deviney et al. (2006) extrapolated recurrence interval distribution models (essentially cumulative distribution functions) for episodic acidification in five catchments in Shenandoah National Park (SHEN) to 226 other catchments. Hourly concentration time series for the five sites were obtained using a transfer function to

predict hourly concentration from hourly discharge and intermittent concentrations. The AMS method, applied to these hourly predictions, was used to estimate parameters of exponential distributions for recurrence intervals for these sites. Regression of distribution parameters on physical catchment attributes was used to extrapolate recurrence interval distribution models to other catchments.

2.3 Markov Chains

Markov processes have been widely discussed in hydrology (Lu, Berliner 1999, Szilagyi, Balint & Csik 2006 and references therein). For more-general primers see (Ross 2006) or (Gallagher 1996). A continuous-time Markov chain (CTMC) is a Markov chain where the conditional distribution of some future state of the chain $X(t+s)$ is only dependent on the present state of the chain $X(s)$. If the transition probabilities from $X(s)$ to $X(t+s)$ are invariant with time (independent of s) we say that the process is *homogeneous*. If these probabilities vary with time we say the process is *nonhomogeneous*. A number of researchers have developed models for threshold violations that leverage extreme value theory and Markov chains (Smith, Tawn & Coles 1997 and references therein).

2.4 Time series models

Time series models are useful in threshold violation characterization as a preliminary step to create an evenly spaced time series for event extraction when missing data are present. The time series literature is deep; for example, see (Brockwell, Davis 1996, Brockwell, Davis 1991, Chatfield 1989, Hamilton 1994). Many time series models have been used in hydrology, including seasonal or periodic autoregressive integrated moving average (SARIMA or PARIMA) models or in some cases simpler versions such as SARMA or ARMA models (Chu, Katz & Ding 1995, Bender, Simonovic 1994, Vecchia 1985, Vecchia et al. 1983, Stedinger, Lettenmaier & Vogel 1985, Salas, Obeysekera 1992, Obeysekera, Salas 1986). Long-term persistence has been modeled using fractionally differenced ARIMA models (Montanari, Rosso & Taquq 1997). Non-normality has been addressed using gamma-autoregressive models (Fernandez, Salas 1990). Changes in chemical concentration have been modeled using intervention analysis (Rasmussen et al. 1993) and quality-control approaches (MacNally, Hart 1997).

State space models admit a wider class of processes than SARIMA models, although SARIMA models can also be formulated as state space models. In particular, state space models are not restricted to time-invariance to the same degree as SARIMA models (Brockwell, Davis 1996). State space models are also referred to as Dynamic Linear Models (West, Harrison 1997, Harrison, Stevens 1976).

Typically, the Kalman filter is used in state space applications to compute estimates of the state vector (Harvey 1990). These estimates are optimal if the

disturbances and initial state vector are normally distributed. The standard Kalman filter is a linear estimator. The Extended Kalman filter (EKF) , as its name suggests, admits non-linear models by basing estimation on the partial derivatives of the process function, the objective being to linearize the process in the region of the current state. The Unscented Kalman Filter (Wan, Van Der Merwe 2000) purports to reduce error in the EKF from the first-order approximation. Hydrological applications of these methods have been published for WQIs from the River Cam (Young 1999) and the Bedford Ouse river system (Whitehead, Beck & O'Connell 1981).

2.5 Bayesian inference

Bayesian models provide an approach to modeling uncertainty in threshold violation models. Techniques and software for performing Bayesian inference have improved dramatically over the last twenty years (Gelman et al. 2004, Gelman, Hill 2007). This is in no small part due to growth and availability of sufficient computational capability to implement algorithms for inference such as Gibbs sampling and Metropolis-Hastings using the Markov chain Monte Carlo (MCMC) technique (Gelman 1997). Bayesian inference is applicable in situations of complex, nonlinear models and sparse data (Scipione, Berliner 1992). Increasingly, the choice of Bayesian vs. frequentist inference is based on pragmatism rather than philosophy (Clark 2005). In particular, the BUGS (Bayesian inference Using Gibbs Sampling) software package provides the

capability to perform Bayesian inference for a wide range of potential models
(Spiegelhalter et al. 1994).

3 Continuous-time Markov models for threshold violations

In this section the development of several models for threshold violations and the method used for estimating model parameters are described. These are used throughout the dissertation to address the primary objectives. A requirement for model development was to be able to estimate the distributions of renewal time (the inverse of frequency), duration of threshold violation events, and limiting probabilities starting only from a set of asynchronous, infrequent, uncoordinated, and unevenly spaced observations (the AIUUS observations). Section 3.1 describes the development of a homogeneous two-state continuous-time Markov chain (CTMC) (Model #1). Sections 3.2 and 3.3 describe extensions of Model #1 that result in the specification of two non-homogeneous models. These three models are used in Chapter 4, and a variant of the homogeneous model is used in Chapter 5. Section 3.4 describes a four-state CTMC that can be used to simulate or model dual correlated CTMC processes. Sections 3.5, 3.6, and 3.7 describe modifications of the first three models to accommodate multiple processes. These three models are used in Chapter 6.

A homogeneous two-state CTMC is a reasonable first choice for a model of threshold exceedance. Homogeneous simply means that the transition rates of the process are invariant with time. Such a process might be expected where seasonal influences are not present. Seasonal influences are present, however, in many WQIs. Two possible modifications of the homogeneous model (among many such possible modifications) are entertained to allow for the presence of seasonal influences. These modifications result in non-homogeneous CTMCs (where the transition rates are a function of time). The rationale for these two modifications are provided in the appropriate sections.

3.1 *Single-process CTMC – Model #1 (homogeneous)*

A Bayesian model specification for a simple two-state homogeneous CTMC model (Figure 1) is described that allows estimation of model parameters and properties from AIUUS observations. Without loss of generality, define state “0” to represent the condition “below threshold” and state “1” the condition “above threshold”. The transition probabilities from “0” to “1” and from “1” to “0” are both 1, but the time spent in each state before the next transition is assumed to be exponentially distributed with potentially different transition rates λ (for state “0”) and μ (for state “1”). That is, upon entering state “0”, the process waits a random amount of time according to the $\text{Exp}(\lambda)$ distribution before transitioning to state “1”. Upon entering state “1”, the process waits a

random amount of time according to the $\text{Exp}(\mu)$ distribution before transitioning back to state “0”. This process model has nice properties that derive from the process parameters λ and μ that speak directly to the renewal rate and the duration of time spent in each state, as well as to the long-term probabilities of state residence.

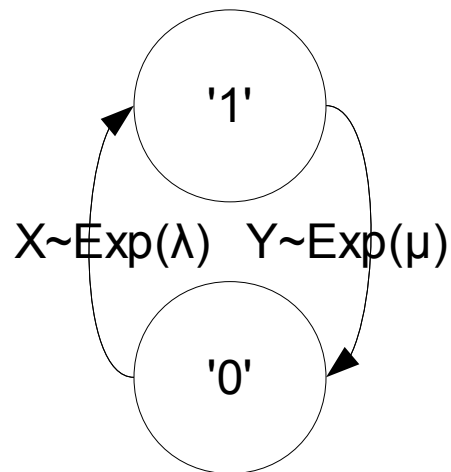


Figure 1: Two-state homogeneous CTMC

In fact, parameter estimation is simple when the process can be observed in such a way that the exact times of transitions between states are known (Figure 2). The rates can be estimated by averaging the lengths of time of excursions into each state. However, when the process is only observed at intermittent points in time not coincident with state changes (the points plotted in Figure 2), the estimation is not so easy. Given an appropriate observation schedule, the limiting probabilities (the long-term proportions of time spent in each state) can be estimated by averaging, but the rates cannot.

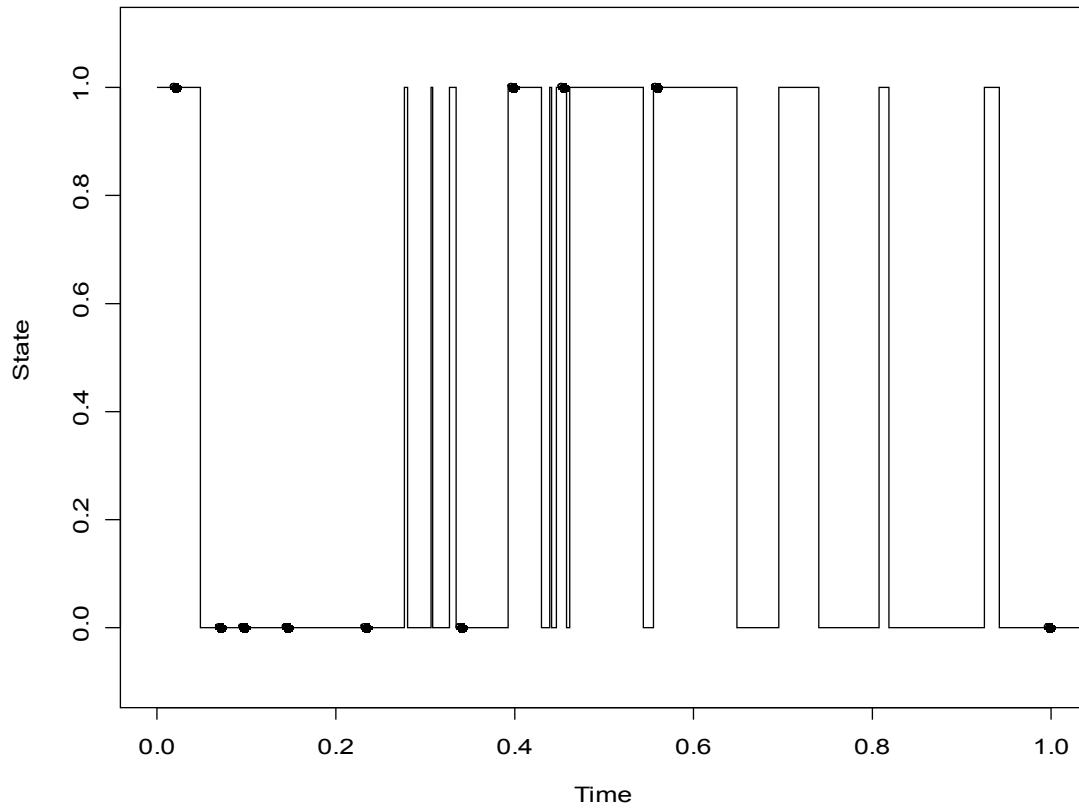


Figure 2: Simulation of a 2-state homogeneous CTMC versus time

Transition rates are $\lambda=20$ and $\mu=30$. Ten simulated observations made at random intervals are plotted as points.

Ross (1993) showed how the two-state homogeneous CTMC model and the Kolmogorov Backward Equation lead to equations that predict the probability of being in state “0” having been in state “0” or having been in state “1” some time previous. Ross gives these equations respectively as:

$$\begin{aligned}
P_{00}(t) &= \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)\Delta t} + \frac{\mu}{\lambda + \mu} \\
P_{10}(t) &= \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)\Delta t}
\end{aligned} \tag{10}$$

Complementary equations can be derived for $P_{11}(t)$ and $P_{01}(t)$. For two points in time indexed by $n-1$ and n , separated by an interval of time $\Delta t = t_n - t_{n-1}$, and with some re-arrangement of terms, these can be written:

$$\begin{aligned}
P_{01}(t_n) &= P_1 F_n \\
P_{11}(t_n) &= 1 - (1 - P_1) F_n \\
P_1 &= \frac{\lambda}{(\lambda + \mu)} \\
F_n &= 1 - e^{-(\lambda + \mu)\Delta t} \\
\Delta t &= t_n - t_{n-1}
\end{aligned} \tag{11}$$

Note that 1) P_1 is the limiting probability of being in state “1” given λ and μ , and 2) F_n is the CDF of an exponential random variable with rate $\lambda + \mu$. It should be noted that this rate is not the renewal rate.

The product $P_1 F_n$ is simply the probability of being in state “1” at time t_n having been in state “0” Δt time intervals previous. The expression

$1 - (1 - P_1) F_n$ is the probability of being in state “1” at time t_n having been in state “1” Δt time intervals previous. For small Δt , P_{01} will be close to 0 and

P_{11} will be close to 1. For large Δt , both P_{01} and P_{11} will be close to P_1 , the limiting probability of being in state “1”. That is, when observations are relatively close together, the previous state observed is somewhat predictive of the subsequent state

observed; it is likely to be the same as the previous state. When observations are relatively far apart, the previous state observed is not predictive of the subsequent state.

Note that any number of transitions may have occurred between observations. The value of these equations is that they express the behavior of the process in terms of observations made at arbitrary times not necessarily coincident with changes in state. All that is required is knowledge of the beginning and ending states and the time between. This model does not require that the time of state transitions be observed. This is advantageous as shall be seen because, in the water quality monitoring field, the exact times of state transitions are never observed.

If S_n is defined as the state of the system at the time t_n , then the probability that S_n is "1" can be written, having observed the state at $n-1$, as

$$Prob(S_n = '1') = \begin{cases} P_1 F_n & , S_{n-1} = '0' \\ 1 - (1 - P_1) F_n & , S_{n-1} = '1' \end{cases} \quad (12)$$

However, this can be written in a more general form as

$$Prob(S_n = '1') = P_1 F_n + (1 - F_n) Prob(S_{n-1} = '1') \quad (13)$$

which can be seen as a weighted average between P_1 , the limiting probability for state "1", and the state at $n-1$. When little time has elapsed, $F_n \approx 0$, and more weight is given to the previous state. When much time has elapsed, more weight is given to the limiting probability. This form also holds when S_{n-1} has not been observed but has been estimated as a probability from a previous observation.

A quantity of interest is the sum of the random variables representing the times spent in each state, since this sum represents the distribution of renewal times, or the times between the onsets of violation events. If X is the random variable representing the time spent in state “0” and Y is the random variable representing the time spent in state “1”, then the CDF of $Z = X + Y$ is given by

$$F_Z(z) = \begin{cases} \frac{\mu(1 - e^{(-\lambda z)}) - \lambda(1 - e^{(-\mu z)})}{\mu - \lambda}, & \lambda \neq \mu \\ 1 - e^{(-\lambda z)}(1 + \mu z), & \lambda = \mu \end{cases} \quad (14)$$

The PDF of Z is then

$$f_Z(z) = \begin{cases} \frac{\lambda\mu}{\mu - \lambda}(e^{(-\lambda z)} - e^{(-\mu z)}) & , \lambda \neq \mu \\ \mu\lambda z e^{(-\lambda z)} & , \lambda = \mu \end{cases} \quad (15)$$

The expected value of Z is

$$E[Z] = E[X] + E[Y] = \frac{1}{\lambda} + \frac{1}{\mu} \quad (16)$$

The renewal rate γ is the inverse of $E[Z]$, but Z is not exponential. See Figure 3 for an example CDF and PDF.

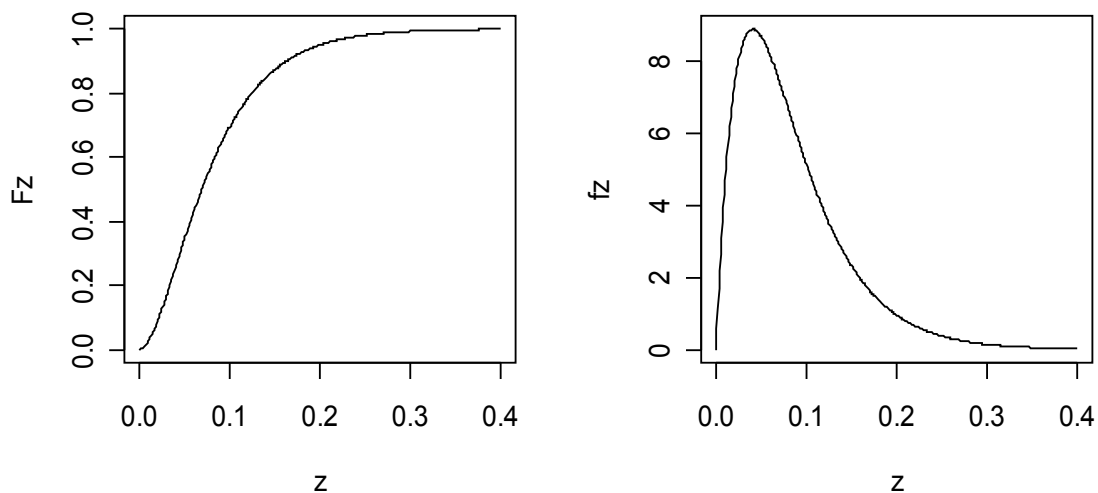


Figure 3: CDF and PDF of the renewal interval of a two-state homogeneous CTMC
 CDF on left. PDF on right. Transition rates: $\lambda=20$ and $\mu=30$.

Given observations of a two-state homogeneous CTMC that include the time of the observation and the state, the process parameters may be estimated using OpenBUGS (Thomas et al. 2006) in R and the following model specification:

$$\begin{aligned}
 \text{Prob}(S_n = '1') &\sim \text{Bernoulli}(P_1 F_n + (1 - F_n) \text{Prob}(S_{n-1} = '1')) \\
 \text{Prob}(S_{n-1} = '1') &= \begin{cases} 1, & S_{n-1} = '1' \\ 0, & S_{n-1} = '0' \end{cases} \\
 P_1 &= \lambda / (\lambda + \mu) \\
 F_n &= 1 - e^{-(\lambda + \mu) \Delta t_n} \\
 \Delta t_n &= t_n - t_{n-1}
 \end{aligned} \tag{17}$$

WinBUGS model code for this model is located in *Appendix A, Single-process CTMC – Model #1 (homogeneous)* and code for a multi-variate version is located in

Appendix A, Multiple-process CTMC – Model #1 (homogeneous), simple version.

Potential priors for λ and μ can be specified in multiple ways, including the following:

$$\begin{aligned} \lambda, \mu &\sim \text{Uniform}(min, max) \\ \lambda, \mu &\sim \text{Gamma}(shape, rate = 1/scale) \\ \lambda, \mu &\sim \text{Jeffreys}(min, max) \end{aligned} \quad (18)$$

The hyper-parameters *min* and *max* for the uniform and Jeffreys priors should be chosen to be far outside the rates specified by the nominal parameters λ and μ . The hyper-parameters *shape* and *rate* for the Gamma priors should be chosen to provide a relatively flat prior (although it will be weakly informative). The advantage of the Gamma prior is that it is unbounded above, unlike the uniform or Jeffreys priors, as implemented here. The Gamma distribution is also a natural conjugate prior for the exponential distribution rate parameter (Gelman et al. 2004). The differences between these priors are shown in Figure 4 over the range from 0.01 to 10,000 with the Gamma priors *shape* and *rate* set to 0.001.

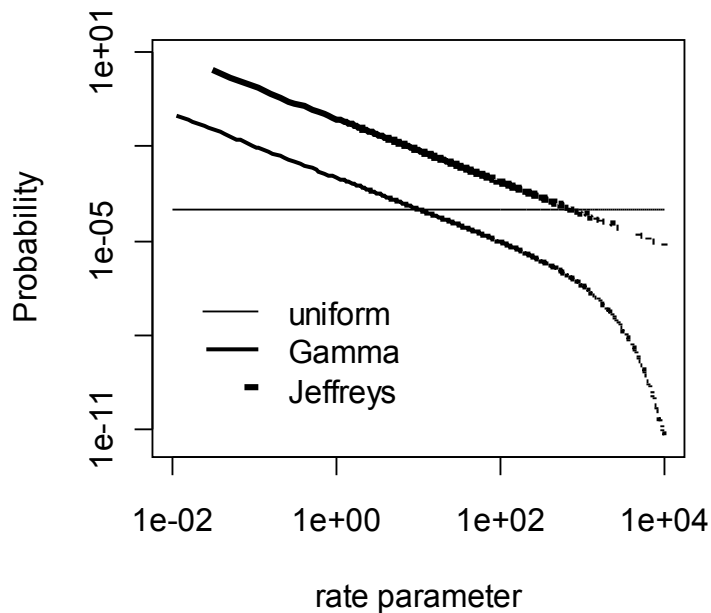


Figure 4: Prior distributions for rate parameters

3.2 Single-process CTMC with varying rate – Model #2 (non-homogeneous)

Many water quality indicator processes in the target application area exhibit annual and/or daily periodicity which can be approximated with a sinusoid. It therefore seems natural to explore modification of the CTMC discussed in Section 3.1 to allow the transition rates to vary with time. Two such modifications are described in this and Section 3.3. First, letting the transition rates evolve periodically as sinusoids while

holding the ratio between them (the limiting probabilities) constant is described. This corresponds to a process where the limiting probabilities are invariant with time, but the expected rates of transition from either state to the other are higher or lower depending on the time of year. In the target application area, this might be observed in a location where the pattern of precipitation varied seasonally or where there was substantial snowpack.

The Bayes specification for Model #2 follows:

$$\begin{aligned}
 Prob(S_n = '1') &\sim Bernoulli(P_1 F_n + (1 - F_n) Prob(S_{n-1} = '1')) \\
 Prob(S_{n-1} = '1') &= \begin{cases} 1, & S_{n-1} = '1' \\ 0, & S_{n-1} = '0' \end{cases} \\
 P_1 &= 1 / (1 + K) \\
 F_n &= 1 - e^{-(1+K)\lambda_n \Delta t_n} \\
 \lambda_n &= \lambda + A \sin(\theta_n) + B \cos(\theta_n) \\
 \Delta t_n &= t_n - t_{n-1} \\
 \mu_n &= K \lambda_n \\
 A &\sim Uniform(-\lambda, \lambda) \\
 B &\sim Uniform(-\sqrt{\lambda^2 - A^2}, \sqrt{\lambda^2 - A^2}) \\
 K &\sim LogNormal(0, 1)
 \end{aligned} \tag{19}$$

WinBUGS model code for this model is located in *Appendix A, Single-process CTMC with varying rate – Model #2 (non-homogeneous)*. Potential priors for λ include the following three:

$$\begin{aligned}
 \lambda, \mu &\sim Uniform(min, max) \\
 \lambda, \mu &\sim Gamma(shape, rate = 1/scale) \\
 \lambda, \mu &\sim Jeffreys(min, max)
 \end{aligned} \tag{20}$$

The priors for A and B are dependent on λ :

$$\begin{aligned} A &\sim \text{Uniform}(-\lambda, \lambda) \\ B &\sim \text{Uniform}(-\sqrt{\lambda^2 - A^2}, \sqrt{\lambda^2 - A^2}) \end{aligned} \quad (21)$$

The variable λ_n is a sinusoid with mean λ and amplitude $2\sqrt{A^2 + B^2}$. A and B are constrained so that λ_n is non-negative. The effect of K is to shrink or expand λ_n and μ_n by the same multiplicative factor at any point in time, thus maintaining constant the instantaneous limiting probabilities P_0 and P_1 but allowing the instantaneous renewal rate to vary. The instantaneous renewal rate γ_n is the inverse of the expected value of Z_n , which is given by

$$E[Z_n] = \frac{1}{\lambda_n} + \frac{1}{\mu_n} = \frac{K+1}{K\lambda_n} \quad (22)$$

3.3 Single-process CTMC with varying limiting probabilities - Model #3 (non-homogeneous)

In the second modification, the renewal rate was held constant but the limiting probabilities P_0 and P_1 were allowed to evolve periodically as sinusoids. This is a process where the distribution of times between events is invariant with time but the instantaneous transition rates vary, with an increase in transition rate from one state being balanced in some way by a decrease in transition rate from the other state. In hydrology, it might be expected to see this behavior in any periodic process in general, and perhaps

even more so in processes which have a diurnal cycle super-positioned on an annual cycle.

There are multiple ways to formulate a model which incorporates the desired properties. To hold the renewal rate constant means that the sum $\frac{1}{\lambda_n} + \frac{1}{\mu_n}$, or $E[Z]$, must be constant. Given values for P_{1n} and $E[Z]$, λ_n and μ_n may be specified according to the following:

$$\begin{aligned}\lambda_n &= \frac{\gamma}{1 - P_{1n}} \\ \mu_n &= \frac{\gamma}{P_{1n}} \\ \gamma &= \frac{1}{E[Z]}\end{aligned}\tag{23}$$

The Bayes specification for Model #3 follows:

$$\begin{aligned}
& \text{Prob}(S_n = '1') \sim \text{Bernoulli}(p_n) \\
p_n &= P_{1,n}F_n + (1 - F_n)\text{Prob}(S_{n-1} = '1') \\
\text{Prob}(S_{n-1} = '1') &= \begin{cases} 1, & S_{n-1} = '1' \\ 0, & S_{n-1} = '0' \end{cases} \\
F_n &= 1 - e^{-(\lambda_n + \mu_n)\Delta t} \\
\mu_n &= \gamma / P_{1,n} \\
\lambda_n &= \gamma / (1 - P_{1,n}) \\
P_{1,n} &= P_1 + A \sin(\theta) + B \cos(\theta) \\
P_1 &\sim \text{Beta}(0.999, 0.999) \\
P_0 &= 1 - P_1 \\
C &= \min(P_0, P_1) \\
A &\sim \text{Uniform}(-C, C) \\
B &\sim \text{Uniform}(-\sqrt{C^2 - A^2}, \sqrt{C^2 - A^2})
\end{aligned} \tag{24}$$

WinBUGS model code for this model is located in *Appendix A, Single-process CTMC with varying limiting probabilities - Model #3 (non-homogeneous)*.

3.4 Correlated Continuous Time Markov Chains - Model #4

There are many situations in water quality monitoring where observations of water quality indicators exist at multiple locations but are not observed simultaneously. It would be desirable to develop models capable of exploiting correlation between such processes, based on the two-state CTMC models already developed for single locations. As an example of one such model, suppose that for two locations A and B there exists the following state diagram (Figure 5) where the state is the ordered pair (S^A, S^B) . It is assumed that the probabilities of both elements changing at precisely the same time (that

is, from (0,0) to (1,1) or from (0,1) to (1,0)) are zero. If the transition rate from

to $(\sim S^A, S^B)$ or from (S^A, S^B) to $(S^A, \sim S^B)$ is independent of the

status of the non-changing element, then S^A and S^B are uncorrelated. That is, if

$\lambda_{A0}=\lambda_{A1}$ and $\lambda_{B0}=\lambda_{B1}$ and $\mu_{A0}=\mu_{A1}$ and $\mu_{B0}=\mu_{B1}$. S^A and S^B are

correlated if the transition rates for one element depend on the state of the other element.

That is, if at least one of the above equalities does not hold.

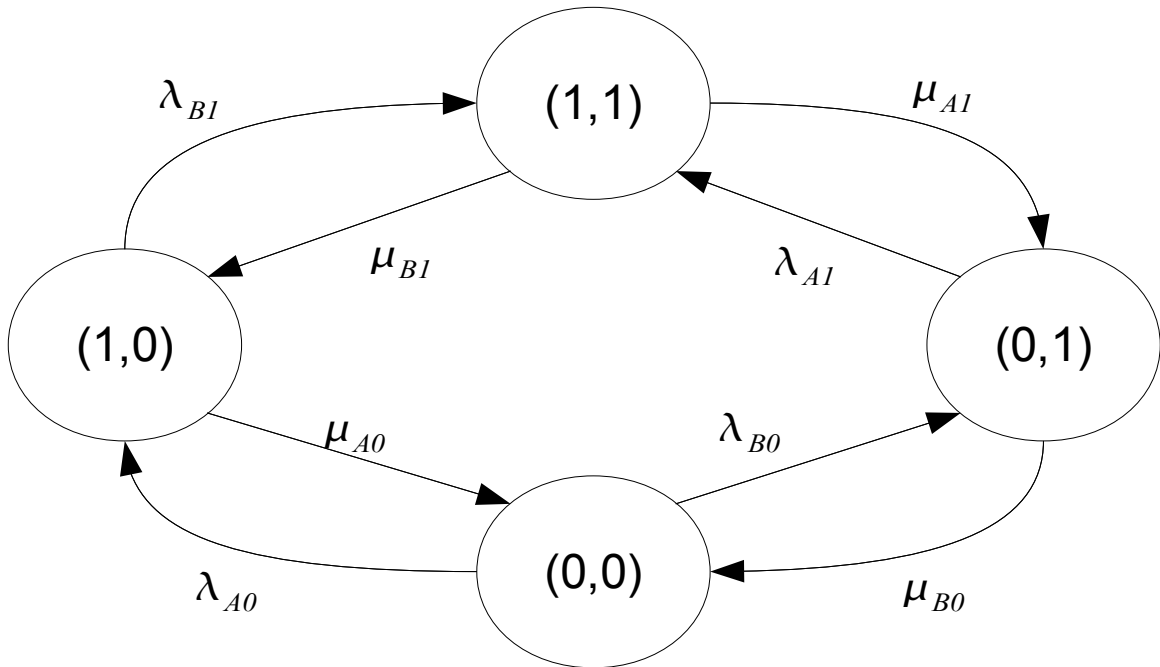


Figure 5: Correlated dual process CTMC

The measure of correlation ρ is given by

$$\rho = \frac{P[S^A = '1', S^B = '1'] - P[S^A = '1']P[S^B = '1']}{\sqrt{P[S^A = '1']P[S^A = '0']P[S^B = '1']P[S^B = '0']}} \quad (25)$$

The limiting probabilities are solved for using the balance equations, which assume that the rate of transitions into any state is equal to the rate out of the state, and using the property that the sum of the limiting probabilities is equal to one. Solving the following system of equations gives the limiting probabilities $P_{(0,0)}$, $P_{(1,0)}$, $P_{(1,1)}$, and $P_{(0,1)}$ for the chain in Figure 5.

$$\begin{bmatrix} -(\lambda_{A0} + \lambda_{B0}) & \mu_{A0} & 0 & \mu_{B0} \\ \lambda_{A0} & -(\mu_{A0} + \lambda_{B1}) & \mu_{B1} & 0 \\ 0 & \lambda_{B1} & -(\mu_{A1} + \mu_{B1}) & \lambda_{A1} \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} P_{(0,0)} \\ P_{(1,0)} \\ P_{(1,1)} \\ P_{(0,1)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (26)$$

The probabilities needed to calculate ρ can then be determined from the limiting probabilities:

$$\begin{aligned} P[S^A = '1', S^B = '1'] &= P_{(1,1)} \\ P[S^A = '1'] &= P_{(1,0)} + P_{(1,1)} \\ P[S^A = '0'] &= P_{(0,0)} + P_{(0,1)} \\ P[S^B = '1'] &= P_{(1,1)} + P_{(0,1)} \\ P[S^B = '0'] &= P_{(1,0)} + P_{(0,0)} \end{aligned} \quad (27)$$

This leads to the following Bayesian specification for Model #4.

$$\begin{aligned}
 S_n^A &\sim \text{bernoulli} \left(\begin{aligned} &(P_{10}^A F_{n0}^A + (1 - F_{n0}^A) P[S_{n-1}^A = '1']) P[S_{n-1}^B = '0'] \\ &+ (P_{11}^A F_{n1}^A + (1 - F_{n1}^A) P[S_{n-1}^A = '1']) P[S_{n-1}^B = '1'] \end{aligned} \right) \\
 S_n^B &\sim \text{bernoulli} \left(\begin{aligned} &(P_{10}^B F_{n0}^B + (1 - F_{n0}^B) P[S_{n-1}^B = '1']) P[S_{n-1}^A = '0'] \\ &+ (P_{11}^B F_{n1}^B + (1 - F_{n1}^B) P[S_{n-1}^B = '1']) P[S_{n-1}^A = '1'] \end{aligned} \right) \\
 P_{10}^A &= \lambda_{A0} / (\lambda_{A0} + \mu_{A0}) \\
 F_{n0}^A &= 1 - e^{-(\lambda_{A0} + \mu_{A0}) \Delta t} \\
 P_{11}^A &= \lambda_{A1} / (\lambda_{A1} + \mu_{A1}) \\
 F_{n1}^A &= 1 - e^{-(\lambda_{A1} + \mu_{A1}) \Delta t} \\
 P_{10}^B &= \lambda_{B0} / (\lambda_{B0} + \mu_{B0}) \\
 F_{n0}^B &= 1 - e^{-(\lambda_{B0} + \mu_{B0}) \Delta t} \\
 P_{11}^B &= \lambda_{B1} / (\lambda_{B1} + \mu_{B1}) \\
 F_{n1}^B &= 1 - e^{-(\lambda_{B1} + \mu_{B1}) \Delta t} \\
 \lambda_{ij}, \mu_{ij} &\sim \text{Gamma}(0.001, 0.001)
 \end{aligned} \tag{28}$$

3.5 Multiple-process CTMC Model #1 (homogeneous)

The specification for multiple-process Model #1 is as follows:

$$\begin{aligned}
 Prob(S_{j,n} = '1') &\sim Bernoulli(P_{j,1} F_{j,n} + (1 - F_{j,n}) Prob(S_{j,n-1} = '1')) \\
 Prob(S_{j,n-1} = '1') &= \begin{cases} 1, & S_{j,n-1} = '1' \\ 0, & S_{j,n-1} = '0' \end{cases} \\
 F_{j,n} &= 1 - e^{-(\lambda_j + \mu_j) \Delta t_n} \\
 \lambda_j &= C_j P_{j,1} \\
 \mu_j &= C_j (1 - P_{j,1}) \\
 \text{logit}(P_{j,1}) &\sim Normal(\hat{\beta}_j, \tau_\beta) \\
 \hat{\beta}_j &\sim Normal(\mu_\beta, \tau_{\mu_\beta}) \\
 \log(C_j) &\sim Normal(\hat{\xi}_j, \tau_\xi) \\
 \hat{\xi}_j &\sim Normal(\mu_\xi, \tau_{\mu_\xi}) \\
 \tau_\beta &\sim Gamma(0.5, 0.5) \\
 \tau_\xi &\sim Gamma(0.5, 0.5) \\
 \tau_{\mu_\beta} &\sim Gamma(0.5, 0.5) \\
 \tau_{\mu_\xi} &\sim Gamma(0.5, 0.5) \\
 \mu_\beta &\sim Normal(0, 0.0001) \\
 \mu_\xi &\sim Normal(0, 0.0001) \\
 \Delta t_n &= t_n - t_{n-1}
 \end{aligned} \tag{29}$$

where λ_j and μ_j are the transition rates from states '0' to '1' and '1' to '0' respectively for process j and $C_j = \lambda_j + \mu_j$. WinBUGS model code for this model is located in *Appendix A, Multiple-process CTMC Model #1 (homogeneous), hierarchical version*. By specifying the logit transformation of the limiting probability of process j being in state '1' ($P_{j,1}$) as having a normal distribution, the capability to add regressors to predict its value is facilitated. Similarly for the constant C_j , a log transformation modeled as having a normal distribution facilitates a similar capability. Since the range of possible

predictors is virtually unlimited, that feature is not explored here. The model as specified simply allows for different intercept parameters for each process.

The shape and inverse-rate parameters that are set here to 0.5 lead to reliable convergence when there are two processes but would need to be reduced for more than two processes. Another approach is to make the shape and inverse-rate hyper-parameters with prior distributions of their own, say for example a uniform prior over the range [0.001, 0.5]. That was not found to be necessary in this work.

3.6 Multiple-process CTMC Model #2 (non-homogeneous)

Multiple-process Model #2 lets the transition rates evolve periodically as sinusoids while holding the ratio between them (the limiting probabilities) constant. This corresponds to a process where the limiting probabilities are invariant with time, but the expected rates of transition from either state to the other are higher or lower depending on the time during the period. The Bayes specification for multiple-process Model #2 follows:

$$\begin{aligned}
\text{Prob}(S_{j,n} = '1') &\sim \text{Bernoulli}(P_{j,1} F_{j,n} + (1 - F_{j,n}) \text{Prob}(S_{j,n-1} = '1')) \\
\text{Prob}(S_{j,n-1} = '1') &= \begin{cases} 1, & S_{j,n-1} = '1' \\ 0, & S_{j,n-1} = '0' \end{cases} \\
F_{j,n} &= 1 - e^{-(1+K_j)\lambda_{j,n}\Delta t_n} \\
\lambda_{j,n} &= \lambda_j + A_j \sin(2\pi t_n) + B_j \cos(2\pi t_n) \\
\lambda_j &= C_j P_{j,1} \\
K_j &= (1 - P_{j,1}) / P_{j,1} \\
\text{logit}(P_{j,1}) &\sim \text{Normal}(\hat{\beta}_j, \tau_\beta) \\
\hat{\beta}_j &\sim \text{Normal}(\mu_\beta, \tau_{\mu_\beta}) \\
\log(C_j) &\sim \text{Normal}(\hat{\xi}_j, \tau_\xi) \\
\hat{\xi}_j &\sim \text{Normal}(\mu_\xi, \tau_{\mu_\xi}) \\
\tau_\beta &\sim \text{Gamma}(0.5, 0.5) \\
r_j &\sim \text{Uniform}(0, 1) \\
A_j &= r_j \lambda_j \cos(\omega) \\
B_j &= r_j \lambda_j \sin(\omega) \\
\tau_\xi &\sim \text{Gamma}(0.5, 0.5) \\
\tau_{\mu_\beta} &\sim \text{Gamma}(0.5, 0.5) \\
\tau_{\mu_\xi} &\sim \text{Gamma}(0.5, 0.5) \\
\mu_\beta &\sim \text{Normal}(0, 0.0001) \\
\mu_\xi &\sim \text{Normal}(0, 0.0001) \\
\omega &\sim \text{Uniform}(-\pi, \pi) \\
\Delta t_n &= t_n - t_{n-1}
\end{aligned} \tag{30}$$

WinBUGS model code for this model is located in *Appendix A, Multiple-process CTMC Model #2 (non-homogeneous)*. A number of other approaches for modeling the two sinusoidal amplitude variables A_j and B_j , including that of Section 3.2 were tried, but none were found to converge as reliably as the one specified above.

3.7 Multiple-process CTMC Model #3 (non-homogeneous)

In the second modification, the renewal rate is held constant but the limiting probabilities P_0 and P_1 are allowed to evolve periodically as sinusoids. This is a process where the distribution of times between events is invariant with time but the instantaneous transition rates vary, with an increase in transition rate from one state being balanced in some way by a decrease in transition rate from the other state.

There are multiple ways to formulate a model which incorporates the desired properties. To hold the renewal rate constant means that the sum $\frac{1}{\lambda_n} + \frac{1}{\mu_n}$, or $E[Z]$, must be constant. Given values for P_{1n} and $E[Z]$, λ_n and μ_n may be specified according to the following:

$$\begin{aligned}\lambda_n &= \frac{\gamma}{1 - P_{1n}} \\ \mu_n &= \frac{\gamma}{P_{1n}} \\ \gamma &= \frac{1}{E[Z]}\end{aligned}\tag{31}$$

The Bayes specification for multiple-process Model #3 follows:

$$\begin{aligned}
& \text{Prob}(S_{j,n} = '1') \sim \text{Bernoulli}(P_{j,1,n} F_{j,n} + (1 - F_{j,n}) \text{Prob}(S_{j,n-1} = '1')) \\
& \text{Prob}(S_{j,n-1} = '1') = \begin{cases} 1, & S_{j,n-1} = '1' \\ 0, & S_{j,n-1} = '0' \end{cases} \\
& F_{j,n} = 1 - e^{-(\lambda_{j,n} + \mu_{j,n}) \Delta t_n} \\
& \lambda_{j,n} = \gamma_j / (1 - P_{j,1,n}) \\
& \mu_{j,n} = \gamma_j / P_{j,1,n} \\
& P_{j,1,n} = P_{j,1} + A_j \sin(2\pi t_n) + B_j \cos(2\pi t_n) \\
& \gamma_j = C_j (1 - P_{j,1}) P_{j,1} \\
& \text{logit}(P_{j,1}) \sim \text{Normal}(\hat{\beta}_j, \tau_\beta) \\
& \hat{\beta}_j \sim \text{Normal}(\mu_\beta, \tau_{\mu_\beta}) \\
& \log(C_j) \sim \text{Normal}(\hat{\xi}_j, \tau_\xi) \\
& \hat{\xi}_j \sim \text{Normal}(\mu_\xi, \tau_{\mu_\xi}) \\
& \tau_\beta \sim \text{Gamma}(0.5, 0.5) \\
& \text{Amp}_{\max_j} = 0.5 - \sqrt{((P_{j,1} - 0.5)^2)} \\
& r_j \sim \text{Uniform}(0, 1) \\
& A_j = r_j \text{Amp}_{\max_j} \cos(\omega) \\
& B_j = r_j \text{Amp}_{\max_j} \sin(\omega) \\
& \tau_\xi \sim \text{Gamma}(0.5, 0.5) \\
& \tau_{\mu_\beta} \sim \text{Gamma}(0.5, 0.5) \\
& \tau_{\mu_\xi} \sim \text{Gamma}(0.5, 0.5) \\
& \mu_\beta \sim \text{Normal}(0, 0.0001) \\
& \mu_\xi \sim \text{Normal}(0, 0.0001) \\
& \omega \sim \text{Uniform}(-\pi, \pi) \\
& \Delta t_n = t_n - t_{n-1}
\end{aligned} \tag{32}$$

WinBUGS model code for this model is located in *Appendix A, Multiple-process*

CTMC Model #3 (non-homogeneous).

4 Single-chain Markov models for threshold exceedance

4.1 Abstract

Thresholds have been defined for many water quality indicators (WQIs) which separate the measurement space of the indicator into two states, one of which (the exceedance or violation state) has undesirable consequences. Observations of the indicator are often made at relatively infrequent, unevenly spaced intervals and are uncoordinated with the precise timing of changes in state. As a first attempt at constructing models to allow estimation of the process properties of frequency, duration, and long-term probability of violation from such observations, and to permit extrapolation to data-poor locations, three different two-state continuous-time Markov chain (CTMC) models based on the Kolmogorov Backward Equation were evaluated under simulated observation conditions. Two of the models were non-homogeneous CTMCs with sinusoidally varying components; the third was a standard homogeneous CTMC. WQIs were simulated as if they were known to follow CTMC models. The

observation process was simulated resulting in unevenly spaced observations. The modeling process was simulated using random model selection. Estimations of model parameters were made with OpenBUGS software and Bayesian MCMC methods. First, three different prior distributions for the rate parameters were compared. Next, a response surface analysis was conducted to assess nominal process model parameter recovery. Last, a principal components regression analysis was employed to evaluate the effects of process, observation, and modeling on various performance metrics. Uniform and Gamma priors were found to be roughly equivalent in terms of performance, and both were found to perform substantially better than a Jeffrey's prior. Duration, frequency, and long-term probability of violation were overestimated more often than not. Not surprisingly, lower estimation error was associated with longer observation period and more observations, regardless of observation type (violation-state or routine). However, estimation error was also associated with degree and type of process homogeneity.

4.2 Introduction

Aquatic resource professionals recognize that serious harm can come to aquatic ecosystems when water quality thresholds are violated too often or for excessively long periods of time (Baldigo, Murdoch 1997, Bulger, Cosby & Webb 2000, Davies et al. 1992, DeWalle, Swistock & Sharpe 1995, Laio et al. 2001, Sickle et al. 1996). However, for many WQIs the high-frequency observations necessary to identify and to time periods

of violation are not available. In addition, these thresholds are generally absolute and independent of the range of values of a specific indicator, so that in a regional context, a threshold may have a different relative relationship at each location where observations are made. These characteristics of the problem are inconsistent with the assumptions of current threshold violation methodologies used in hydrology, which are that observations have been made frequently and at evenly spaced time steps, and that thresholds are relative. So under these more realistic assumptions the distributions of return period and duration of threshold violation events have not been well-characterized.

Relatively little work has been done in this specific area. However, considerable work has been done in the broader area of threshold violation. In particular, hydrologists have worked for decades to couple extreme value theory with the theory of stochastic processes, motivated by the need to be able to predict the likelihood of occurrence of disastrous but infrequent and random flood events within typically long management horizons such as the lifetime of a structure. For these analyses, hydrologists are generally blessed with the availability of relatively high-frequency, evenly spaced observations of water flow.

Markov processes have been widely discussed in hydrology (Lu, Berliner 1999, Szilagyi, Balint & Csik 2006 and references therein). A number of researchers have developed models for threshold violations that leverage extreme value theory and Markov chains (Smith, Tawn & Coles 1997 and references therein). Others have used statistical (Deviney, Rice & Hornberger 2006) or process (Zhang, Arhonditsis 2008)

models to make time series predictions from which threshold violation properties can be estimated.

The purpose of this paper was to evaluate three different two-state continuous-time Markov chain (CTMC) models based on the Kolmogorov Backward Equation under simulated observation conditions. These models allow estimation of transition rates and limiting probabilities from uncoordinated and unevenly spaced observations. The evaluation consisted of assessments of nominal parameter recovery, of out-of-sample prediction performance, and of the effect of various process and observation characteristics on said performance. The purpose of these assessments was to establish the conditions under which the methodology could be applied, assuming that successful application would be a function of both process characteristics and of observation protocol characteristics.

In the following three sub-sections three two-state Markov chain models for threshold exceedance are presented that allow the estimation of frequency and duration properties from uncoordinated and unevenly spaced observations of a process. The first model is homogeneous while the second and third are non-homogeneous. The latter two models allow transition rates to vary periodically.

4.2.1 Model #1: homogeneous two-state CTMC

A homogeneous two-state CTMC is a reasonable first choice for a model of threshold exceedance. Let one state (say state '0') represent the below-threshold

condition and the other state (state '1') represent the above-threshold condition. Let state '0' represent the exceedance or violation state. Assume that upon entering one state, a random amount of time passes before the process transits to the other state. Once in the other state, a random amount of time passes before the process returns to the original state. Assume these random times have exponential distributions with rates λ and μ that are not necessarily equal. Such a process is depicted in Figure 6.

Some important properties of the process in Figure 6 are given in Equation (33), where $E[X]$ represents the expectation of the amount of time spent in state '0' on a typical excursion into that state, $E[Y]$ represents the same expectation for state '1', $E[Z]$ is the expectation of the time between entrances into either state, P_0 is the long-term proportion of time spent in state '0', and P_1 is the long-term proportion of time spent in state '1'.

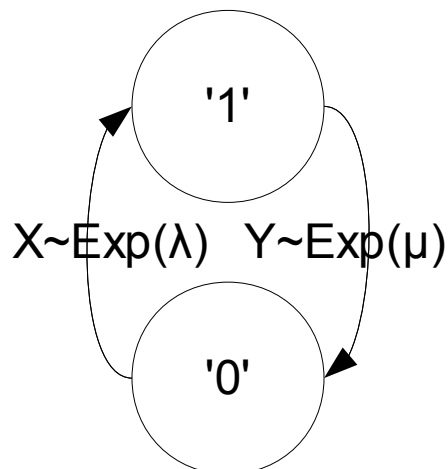


Figure 6: Two-state CTMC

$$\begin{aligned}
E[X] &= \frac{1}{\lambda} \\
E[Y] &= \frac{1}{\mu} \\
E[Z] &= E[X] + E[Y] \\
P_0 &= \frac{\mu}{\lambda + \mu} \\
P_1 &= \frac{\lambda}{\lambda + \mu}
\end{aligned} \tag{33}$$

Ross (1993) showed how the two-state homogeneous CTMC model and the Kolmogorov Backward Equation lead to equations that predict the probability of being in state “0” having been in state “0” or having been in state “1” some time previous. Ross gave these equations respectively as:

$$\begin{aligned}
P_{00}(t) &= \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)\Delta t} + \frac{\mu}{\lambda + \mu} \\
P_{10}(t) &= \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)\Delta t}
\end{aligned} \tag{34}$$

With some re-arrangement of terms, a Bayesian specification can be written as

$$\begin{aligned}
\text{Prob}(S_n = '1') &\sim \text{Bernoulli}(P_1 F_n + (1 - F_n) \text{Prob}(S_{n-1} = '1')) \\
\text{Prob}(S_{n-1} = '1') &= \begin{cases} 1, & S_{n-1} = '1' \\ 0, & S_{n-1} = '0' \end{cases} \\
P_1 &= \lambda / (\lambda + \mu) \\
F_n &= 1 - e^{-(\lambda + \mu)\Delta t_n} \\
\Delta t_n &= t_n - t_{n-1}
\end{aligned} \tag{35}$$

where λ and μ are the transition rates from '0' to '1' and '1' to '0' respectively. Prior distributions for λ and μ need only to be determined. Several choices are reasonable including exponential, Gamma, uniform, and Jeffreys.

4.2.2 Model #2: varying transition rates

Model #2 lets the transition rates evolve periodically as sinusoids while holding the ratio between them (the limiting probabilities) constant. This corresponds to a process where the limiting probabilities are invariant with time, but the expected rates of transition from either state to the other are higher or lower depending on the time during the period. A Bayes specification for Model #2 follows:

$$\begin{aligned}
 \text{Prob}(S_n = '1') &\sim \text{Bernoulli}(P_1 F_n + (1 - F_n) \text{Prob}(S_{n-1} = '1')) \\
 \text{Prob}(S_{n-1} = '1') &= \begin{cases} 1, & S_{n-1} = '1' \\ 0, & S_{n-1} = '0' \end{cases} \\
 P_1 &= 1/(1 + K) \\
 F_n &= 1 - e^{-(1+K)\lambda_n \Delta t_n} \\
 \lambda_n &= \lambda + A \sin(\theta_n) + B \cos(\theta_n) \\
 \Delta t_n &= t_n - t_{n-1} \\
 \mu_n &= K \lambda_n \\
 A &\sim \text{Uniform}(-\lambda, \lambda) \\
 B &\sim \text{Uniform}(-\sqrt{\lambda^2 - A^2}, \sqrt{\lambda^2 - A^2}) \\
 K &\sim \text{LogNormal}(0, 1)
 \end{aligned} \tag{36}$$

Other than the distributions given, a prior only needs to be specified for λ .

4.2.3 Model #3: varying limiting probabilities

Model #3 holds the renewal rate constant but lets the limiting probabilities P_0 and P_1 evolve periodically as sinusoids. This is a process where the distribution of event inter-arrival times is invariant but the instantaneous transition rates vary, with an

increase in transition rate from one state being balanced in some way by a decrease in transition rate from the other state.

There are multiple ways to formulate a model which incorporates the desired properties. To hold the renewal rate constant means that the sum $\frac{1}{\lambda_n} + \frac{1}{\mu_n}$, or $E[Z]$, must be constant. Given instantaneous values for P_{1n} and the constant $E[Z]$, λ_n and μ_n may be specified according to the following:

$$\begin{aligned}\lambda_n &= \frac{\gamma}{1 - P_{1n}} \\ \mu_n &= \frac{\gamma}{P_{1n}} \\ \gamma &= \frac{1}{E[Z]}\end{aligned}\tag{37}$$

A Bayes specification for Model #3 follows:

$$\begin{aligned}
& \text{Prob}(S_n = '1') \sim \text{Bernoulli}(p_n) \\
p_n &= P_{1,n}F_n + (1 - F_n)\text{Prob}(S_{n-1} = '1') \\
\text{Prob}(S_{n-1} = '1') &= \begin{cases} 1, & S_{n-1} = '1' \\ 0, & S_{n-1} = '0' \end{cases} \\
F_n &= 1 - e^{-(\lambda_n + \mu_n)\Delta t} \\
\mu_n &= \gamma / P_{1,n} \\
\lambda_n &= \gamma / (1 - P_{1,n}) \\
P_{1,n} &= P_1 + A \sin(\theta) + B \cos(\theta) \\
P_1 &\sim \text{Beta}(0.999, 0.999) \\
P_0 &= 1 - P_1 \\
C &= \min(P_0, P_1) \\
A &\sim \text{Uniform}(-C, C) \\
B &\sim \text{Uniform}(-\sqrt{C^2 - A^2}, \sqrt{C^2 - A^2})
\end{aligned} \tag{38}$$

For Model #3 a prior only needs to be specified for γ , the inverse of the expected inter-arrival interval.

The remainder of the paper proceeds as follows: In Section 4.3 methods are presented for process, observation, and modeling simulation. An experimental design is given for testing the performance of three different two-state Markov Chain models that allow the estimation of process properties from unevenly spaced data. The performance metrics used are introduced, and the methods used to 1) evaluate the relative performance of the models given different priors, 2) evaluate nominal parameter recovery performance, and 3) relate performance to process and observation protocol characteristics are described. In Section 4.4 analytical results are presented, and in Sections 4.5 and 4.6 discussion and conclusions are provided.

4.3 Method

In this Section the methods used to perform the experiment and evaluate the results are described. In all, 9,000 trials were run, 3,000 using uniform priors for the rate parameters, 3,000 using Gamma priors, and 3,000 using Jeffreys priors. The exponential prior was not used explicitly because the exponential and Gamma distributions are conjugates and the exponential is simply a degenerate form of the Gamma. Individual trials varied with respect to process and observation protocol characteristics, and with respect to the model chosen for process simulation and the model chosen for estimation. The same set of 3,000 simulated process observations was used for each set of prior distributions.

4.3.1 Experimental design

A design was created to simulate a range of processes, a range of process observation protocols, and to simulate modeling, or modeling selection. Choices for nominal parameter levels describing the processes, observation protocols, and model selection were generated using leaped Halton sequences (Kocis, Whiten 1997). To reduce the chance of correlation between parameters, the leaped Halton sequence for each parameter was generated using a different prime number as the base, and a common leap parameter of the next prime number greater than the maximum base value. Fifteen hundred sets of nominal parameter values were generated. For each parameter set, two

separate process simulations were generated. In total, 3000 scenarios were created, each using one of the three models to generate the simulated observations and one of the three models for parameter estimation.

4.3.2 Process Simulation

Processes were simulated using the three CTMC process models described in Section 4.2 . Simulations were designed to exhibit a range of process behaviors typical of and of interest in the water-quality monitoring field (Table 1). Parameters were set so that the mean duration of the violation state (state '0') was between 0.005 periods and 0.05 periods (roughly between 40 hours and 2.5 weeks). Parameters were set so that processes had a nominal mean renewal time of between 1 and 0.1 (between once per period and ten times per period, or roughly from annually to monthly). For Model #2 parameters were set so that the sinusoidal amplitude (peak to trough) ranged from zero to 2λ , the maximum value that preserves the constraint that instantaneous λ must be greater than zero. For Model #3 parameters were set so that the sinusoidal amplitude ranged from zero to $2\min(P_0, P_1)$, the maximum value that preserves the constraint that P_0 must be between zero and one. With amplitude set to zero, the non-homogeneous models reduce to the homogeneous model.

Table 1: Ranges of important process characteristics of CTMCs used in simulations

Applies to Model			Property	Range	Description
#1	#2	#3			
√	√	√	Mean duration $E[X]$	$E[X]: [0.005, 0.05]$ Set $\lambda=1/E[X]$	Expected length of time in the event state
√			Mean renewal interval $E[Z]$	$E[Z]: [0.1, 1.0]$ Set $\mu=\lambda/(\lambda E[Z]-1)$	Expected time between event onsets
	√		A, B	A: $[-\lambda, \lambda]$ B: $[-\sqrt{\lambda^2 - A^2}, \sqrt{\lambda^2 - A^2}]$	Determine amplitude and phase shift
		√	A, B	Set $P_0 = \mu / (\mu + \lambda)$ Set $P_1 = 1 - P_0$ Set $A_{max} = \min(P_0, P_1)$ A: $[-A_{max}, A_{max}]$ B: $[-\sqrt{A_{max}^2 - A^2}, \sqrt{A_{max}^2 - A^2}]$	Determine amplitude and phase shift

Simulated high-frequency observations (the HF observations) were spaced 0.0001 time units apart (10,000 data points per period). Compared to the solar year, that is roughly equivalent to 1 data point per hour. It was assumed that more-frequently simulated observations would not be necessary to capture state changes important in water-quality monitoring. In any case it is unusual for water-quality indicator observations to be made this frequently except by sondes located in-situ. These are typically only capable of measuring a few parameters and are not used to make observations at as many locations as grab sampling (collecting samples manually), and

not used at all for all but a few important water-quality indicators such as temperature, dissolved oxygen, pH, and conductivity.

4.3.3 Observation simulation

In the context of this experiment, the measurement process has two major components, observation and model selection. Both components were simulated by first simulating the observation of the HF process described in Section 4.3.2, and then simulating model selection. Three minor components were defined associated with the observation protocol (Table 2). For the first component *observation period length*, lengths of between 5 and 30 periods were specified, as the lower limit corresponds roughly to a generally accepted minimum time period for analysis of stochastic processes with a seasonal component (Hirsch, Slack & Smith 1982), and the upper limit corresponds roughly to the maximum length of observed data available in water-quality monitoring. The HF observations were sub-sampled by mimicking observation practices in water-quality monitoring; that is, a mixture of infrequent but approximately evenly spaced observations along with even more-infrequent bursts of high-frequency sampling during the violation state. This mimics many water-quality monitoring programs where samples are obtained during routine visits to the site on a weekly or monthly basis, and automated samplers are set up to collect samples every few hours during storm events that occur several times per year, but not on an otherwise predictable basis. The second component of the observation policy *quasi-regular observation interval* mimicked a

regular collection spaced at intervals of 0.01 to 1 period ($\sim 100 - 1$ observations per period), where the actual interval lengths were allowed to vary with a standard deviation equal to one one-hundredth of the mean interval length. For example, for a mean interval of one week (168 hours), the standard deviation would have been 1.68 hours. The third component of the observation policy *event observation probability* mimicked typical violation state observation protocols by “taking” samples while the process was in the violation state with a probability between zero and a probability which would yield on average 100 samples per year given the nominal transition rates specified for the simulation.

Table 2: Ranges of important process observation characteristics

Characteristic	Range	Description
Observation period length	[5, 30]	Total # of periods observed
quasi-regular observation interval	[0.01, 1.0]	Quasi-regular spacing. That is, spacing is equal \pm a small perturbation
Event observation probability	[0, 1]	Event observation probability. Probability is set such that the expected number of event samples per period is less than 100.

4.3.4 Modeling simulation

Model selection was simulated by selecting one of the three models at random for estimation of process parameters, irrespective of which model was used to generate the sub-sampled observations. Posterior distributions of model parameters were generated using the sub-sampled observations and OpenBUGS (Thomas et al. 2006), a Bayesian estimation software package that can be run from R using the BRugs package (Thomas et

al. 2006). All trials were run on the Cross-Campus Grid (XCG) at the University of Virginia (Morgan, Grimshaw 2007). Convergence was determined by requiring that the Gelman-Rubin statistic \hat{R} for all monitored parameters be less than 1.1, a threshold suggested by Gelman and Hill (2007). In addition, the slopes of all chains were required to be trend-free. To check this, a regression of chain values versus their iteration index, modeling residuals to be first-order autocorrelated, was performed. A conservative p-value of 0.01 divided by the total number of chains was used for rejection of the null hypothesis of no trend to compensate for the number of tests.

Estimations were performed using three chains. After burn-in, sufficient iterations were run to obtain 350 values from each chain, for a total of 1050 samples from the posterior distributions. Each estimation run was initialized with a relatively small burn-in period and thinning parameter. Then, if convergence was not attained, the thinning parameter was increased and the procedure ran again, starting from the last iteration's values of the previous run, and considering all iterations of the algorithm up to that point to be burn-in.

The priors for λ , μ , and γ were given respectively by one of the three following specifications:

$$\begin{aligned}
 &\lambda, \mu, \gamma \sim \text{Uniform}(0.01, 10000) \\
 &\text{or} \\
 &\lambda, \mu, \gamma \sim \text{Gamma}(0.001, 0.001) \\
 &\text{or} \\
 &\lambda, \mu, \gamma \sim \text{Jeffreys}(0.01, 10000)
 \end{aligned} \tag{39}$$

The hyper-parameters 0.01 and 10,000 for the uniform and Jeffreys priors were chosen to be far outside the rates specified by the nominal parameters (20 to 200 for λ and 1 to 17 for μ). The hyper-parameters (both 0.001) for the Gamma priors were chosen to provide a relatively flat, but weakly informative prior. The advantage of the Gamma prior is that it is unbounded above, unlike the uniform or Jeffreys priors, as implemented here. The Gamma distribution is also a natural conjugate prior for the exponential distribution rate parameter (Gelman et al. 2004). The differences between these priors are shown in Figure 7 over the range from 0.01 to 10,000. The vertical lines indicate the combined range of the two rate parameters.

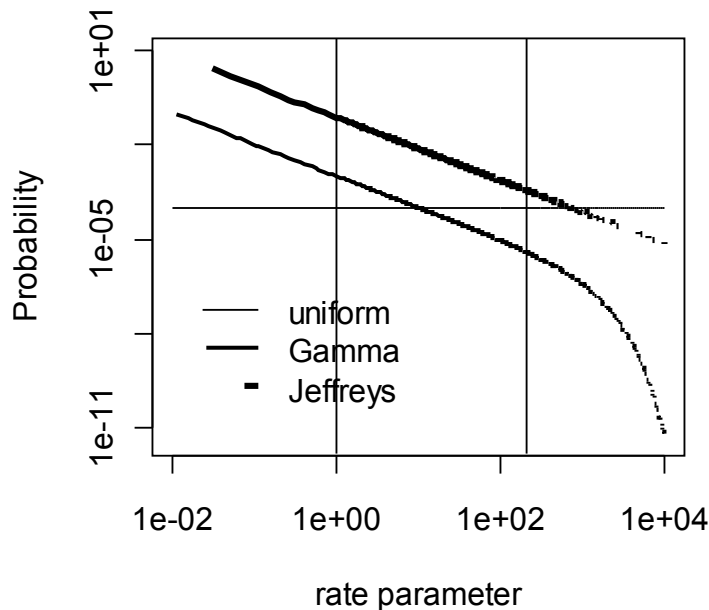


Figure 7: Prior distributions of rate parameters

4.3.5 Determination of metrics

Performance was evaluated on a set of metrics which included Integrated Squared Error, or ISE (Scott 1992, Bowman, Azzalini 1997) , Deviance Information Criteria, or DIC (Spiegelhalter et al. 2002), the area under the ROC curve, or AUC (Hanley, McNeil 1982), and a binary parameter recovery metric which indicated whether or not the nominal parameter value was within a 95% credible interval determined from the posterior distribution for the parameter.

To measure the recovery of these nominal parameter values by the estimation process, they were compared to their corresponding posterior distributions. As the posterior distributions might not be normally distributed, the percentile for each nominal value corresponding to the posterior distribution was calculated. That yielded a set of values between 0 and 100. A 0 indicates that the nominal lies to the left of the posterior (it was over-estimated) and a 100 indicates the nominal lies to the right of the posterior (it was under-estimated). A 95% credible interval for the posterior was defined as the percentiles between 2.5 and 97.5, and used to determine the value of the binary metric.

DIC and AUC are useful for evaluating performance on the prediction of future states of the process. The distributions of the renewal interval and of the duration with

respect to the violation state are the stochastic properties of primary interest, so the ISE was used as the index of performance for these properties.

The first half (the *training set*) of each set of sub-sampled observations was used for parameter estimation. The second half (the *test set*) was used for determination of the AUC metric. AUC was calculated using the ROCR package in R (Sing et al. 2005). Predictions of observed states were made using the maximum a posteriori (MAP) value drawn from the posterior distributions of parameter values and the times between observations in the test set.

The deviance information criterion, or DIC (Spiegelhalter et al. 2002) is a measure of predictive capability similar to the Akaike Information Criterion (AIC). It is the sum of a reward term based on the likelihood of the data given the model, and a penalty term based on the number of effective parameters. Like the AIC, smaller is better. DIC is returned by the OpenBUGS procedure.

ISE was evaluated using the nominal parameter values and the posterior values of the estimated parameters associated with the minimum deviance. For Model #1 the densities of duration and renewal period are easily derived from the rate parameters λ and μ . Duration is exponentially distributed with rate parameter λ . Renewal period has a distribution given by

$$f_z(z) = \begin{cases} \frac{\lambda \mu}{\mu - \lambda} (e^{-\lambda z} - e^{-\mu z}) & , \lambda \neq \mu \\ \mu \lambda z e^{-\lambda z} & , \lambda = \mu \end{cases} \quad (40)$$

However, for the other two models these densities are difficult to derive, if at all possible. These densities were therefore approximated for these two models. Points of the densities were determined by averaging numerically over the values the density could take over two time periods. The ISEs for the CTMC methods were then calculated from the nominal and CTMC densities for duration and renewal period.

4.3.6 Comparison of priors

Each set of sub-sampled observations was modeled three times with the same model, but each time using a different prior distribution for the parameters λ , μ , and γ . The paired Wilcoxon test for differences in location was used for each of the metrics on a pairwise basis. That is, uniform vs. Gamma, uniform vs. Jeffreys, and Gamma vs. Jeffreys.

4.3.7 Recovery of nominal parameter values

Recovery of nominal parameter values was assessed by first computing a metric indicating whether parameter recovery was obtained or not. The three primary process characteristics of interest assessed in this step were 1) the expected duration period $E[X]$, 2) the expected renewal interval $E[Z]$, and 3) the limiting probability for state "1" P_1 . The metric computed was determined from the percentile of the posterior distribution corresponding to the nominal parameter value. Percentile values between 2.5 and 97.5

were considered to be representative of cases where the nominal parameter value was recovered. This binary metric was then used as the dependent variable in a logistic regression with the process characteristics and observation protocol characteristics as regressors. This model allows assessment of the predicted recovery rate associated with different regions of the experimental design space, and through analysis of the steepest gradient of the response surface (Myers, Montgomery 1995), the direction in which greater or lesser recovery rate occurs. The 0.95 contour of the response surface represents the line at which the nominal value can be expected to be found within the 95% credible interval of the posterior distribution 95% of the time.

4.3.8 Evaluation of effect of characteristics on metrics

The principal components associated with the process definition characteristics and the observation protocol definition characteristics were used as independent regressors against suitable transformations of three of the four metrics. DIC was not included because DIC should only be used to compare models on the same dataset. The three characteristics defining the observation process (observation period length, quasi-regular observational interval, and event observation probability) coincide with the choices researchers and managers have when establishing a monitoring program. However, there are typically constraints involved, otherwise the default choice would be to make observations quasi-continuously for long periods of time. For example, a choice might need to be made between frequent observations for a short period of time or

infrequent observations over a longer period of time. Or, process characteristics could influence the way observations are made. PCA allowed the identification of combinations of these characteristics that would correspond to different observation strategies and different process characteristics.

Suitable transformations of the metrics were regressed against a subset of the principal components explaining a significant proportion of the variance represented by the process and observation characteristics. In each regression, components with insignificant coefficient scores were removed from the model. Test model identity was also included as a categorical regressor.

4.4 Results

In this Section, results are reported for the simulated observation and measurement of process properties. First, observations of processes were simulated from each of the models at high-frequency, evenly spaced time steps. Next, typical observation protocols were mimicked by sub-sampling the high-frequency time series. Model parameters were estimated and performance metrics calculated. Modeling results were compared using various priors for the rate parameters, and then the effects of process and observation protocol characteristics on the performance metrics were determined.

Of the 9000 trials performed, 8994 completed successfully. Some scenarios required multiple runs to achieve convergence; however, it was not determined which

required multiple runs because of the operating environment and which required multiple runs because of problems with the estimation process, such as poor starting values. The operating environment consisted of computers in public labs at the University of Virginia, which were subject to students and other users logging in and disrupting a scenario estimation in progress, requiring a re-start. The probability of this happening increased with running time, so scenarios requiring longer running time to reach convergence were more subject to interruption. The six scenarios that never reached convergence were run enough times that it appeared that the training data were simply insufficient to allow the estimation process to reach convergence.

4.4.1 Comparison of priors

The DICs, AUCs, duration ISEs, and renewal ISEs resulting from using different priors were compared using the Wilcoxon signed rank test. Since the estimation of AUC was based on performance on new data, use of this metric for comparison of priors has some aesthetic appeal over the other metrics, which are in-sample results. However, the AUC doesn't directly measure the error on two important process properties, duration and renewal distributions. The results (Table 3) indicate that the uniform prior outperformed both the Gamma and Jeffreys priors on AUC in a statistical sense. There was no statistical difference between the Gamma and the Jeffreys. The pseudomedians (Hollander, Wolfe 1973), which are robust estimators of the central locations, are all very small.

Table 3: Wilcoxon signed rank test on AUC – MAP (bigger is better)

Test	95 percent confidence interval:		(pseudo) median	p-value
	Lower	Upper		
Gamma - uniform	-1.36E-004	-7.08E-005	-6.90E-005	0.0000
Gamma - Jeffreys	-1.74E-005	5.26E-006	3.60E-005	0.0652
Uniform - Jeffreys	1.56E-005	1.35E-004	8.85E-005	0.0000

The DIC has been touted as a reasonable information-based performance criterion that, like the AIC and others, rewards good fit while penalizing model complexity (Gelman, Hill 2007, Spiegelhalter et al. 2002). Both the Gamma and uniform priors resulted in significantly lower DICs than the Jeffreys prior (Table 4), and the Gamma prior performed slightly better than the uniform. Comparisons of DIC are appropriate in this instance since the same training sets were used in each pair of comparisons made in the paired test. The reason for the much larger DIC values associated with the Jeffrey's prior is unknown; however, it is suspected to be related to the fact that the Jeffrey's prior is not a built-in prior in WinBUGS but has to be constructed using a “zeros trick”. This construction is known to result in highly correlated chains.

Table 4: Wilcoxon signed rank test on DIC (smaller is better)

Test	95 percent confidence interval		(pseudo) median	p-value
	Lower	Upper		
Gamma - uniform	-4.01E-002	-1.00E-005	-1.50E-002	0.0686
Gamma - Jeffreys	-1.91E+001	-1.86E+001	-1.89E+001	0.0000
Uniform - Jeffreys	-1.89E+001	-1.82E+001	-1.86E+001	0.0000

In the case of the duration ISE the Gamma and uniform priors resulted in significantly lower duration ISE than the Jeffrey's prior (Table 5). There was no significant difference between the uniform prior and the Gamma prior.

Table 5: Wilcoxon signed rank test on duration ISE (smaller is better)

Test	95 percent confidence interval:		(pseudo) median	p-value
	Lower	Upper		
Gamma - uniform	-4.99E-003	1.53E-002	5.09E-003	0.3181
Gamma - Jeffreys	-8.55E-001	-7.76E-001	-8.14E-001	0.0000
Uniform - Jeffreys	-8.39E-001	-7.60E-001	-7.98E-001	0.0000

The Jeffreys prior resulted in significantly lower renewal ISE than the Gamma prior. The Gamma prior resulted in significantly lower renewal ISE than the uniform prior (Table 6).

Table 6: Wilcoxon signed rank test on renewal ISE (smaller is better)

Test	95 percent confidence interval:		(pseudo) median	p-value
	Lower	Upper		
Gamma - uniform	-4.40E-003	-3.48E-004	-2.30E-003	0.0203
Gamma - Jeffreys	1.27E-001	1.45E-001	1.36E-001	0.0000
Uniform - Jeffreys	1.34E-001	1.52E-001	1.43E-001	0.0000

There appeared to be no clearly preferable prior, although the Jeffreys could be rationalized as the worst choice among the three based on the size of the pseudomedians resulting from tests involving the Jeffreys prior.

4.4.2 Nominal parameter recovery

There were three primary nominal attributes of interest: 1) $E[X]$ – the expected value of time spent in the violation state, 2) $E[Z]$ – the expected value of the renewal interval (time between re-entries to the violation state, or the inverse of frequency of occurrence), and 3) P_1 – the limiting probability for the non-violation state. Very few of the nominal parameter percentiles fell within the credible interval (Table 7). There was a tendency to overestimate the expected duration period $E[X]$ and underestimate the expected renewal period $E[Z]$ and the limiting probability P_1 . This is equivalent to overestimating the frequency and the long-term violation state probability. These patterns favor the conservationist. The estimates for $E[Z]$ fell in the credible interval (c.i.) more often than for the other two. The estimates for P_1 fell in the credible interval rarely. This is interesting since P_1 is a function of the other two ($P_1=(E[Z]-E[X])/E[Z]$).

Table 7: Counts of recovered nominal values across all priors

Group	# below c.i. (nominal was overestimated)	# inside c.i.	# above c.i. (nominal was underestimated)	total	% inside
E[X]	7044	1755	195	8994	19.51
E[Z]	213	2226	6555	8994	24.75
P1	0	189	8805	8994	2.1

To gain understanding of the relationship between recovery and process/observation characteristics, a basic response surface methodology (RSM) analysis was performed using logistic regression and a first-order model without interaction. Independent variables included the following (nominal values): E[X], P₁, E[Z], observation period, quasi-regular observation interval, and event observation probability. Variables were transformed for the analysis because they need to be on unbounded scales in order to make predictions at any point along the path of steepest ascent. The observation protocol properties were included to avoid masking effects. In a preliminary model fit, simulation model, test model, and variable name had significant effects but did not change magnitudes, signs, or significance of the other variables much. Choice of prior distribution had no effect. These other variables were not included in the final logistic regression because recovery was generally poor in the experimental design space and because they were not helpful in determining sufficient conditions for recovery. All variables were found to be highly significant (Table 8).

Table 8: RSM regression results

```

> summary(glm.RSM)

Call:
glm(formula = In ~ log_E_X + logit_P1 + log_E_Z + log_oP + log_rS +
     logit_eP, family = "binomial", data = merged.All)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6818  -0.6010  -0.4459  -0.2871   3.1267

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.71072    0.22570  12.010 < 2e-16 ***
log_E_X      2.88860    0.45468   6.353 2.11e-10 ***
logit_P1     2.04172    0.41790   4.886 1.03e-06 ***
log_E_Z     -2.65015    0.46738  -5.670 1.43e-08 ***
log_oP      -0.95406    0.03563 -26.778 < 2e-16 ***
log_rS      -0.53715    0.01770 -30.342 < 2e-16 ***
logit_eP    -0.13067    0.01569  -8.326 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 23232  on 26981  degrees of freedom
Residual deviance: 20681  on 26975  degrees of freedom
AIC: 20695

Number of Fisher Scoring iterations: 5

```

The mean vector of the transformed experimental design space is given in Table

9.

Table 9: Experimental design space mean vector

```

> mean.vector
      log_E_X logit_P1   log_E_Z   log_oP   log_rS   logit_eP
1 -3.740215  2.919782 -0.7446788  2.759322 -0.954003 -2.377568

```

The probability of recovery associated with this vector was 0.1226232. The unit vector in the direction of steepest ascent is given in Table 10.

Table 10: Unit vector in direction of steepest gradient

```
> unit.vector
  log_E_X logit_P1 log_E_Z log_oP log_rS logit_eP
1 0.6341031 0.4481966 -0.5817588 -0.2094356 -0.1179138 -0.02868556
```

Movement from any point in the direction of this vector results in a higher probability of recovery. Predicted recovery rate reaches 0.95 between 1 and 1.5 units along the steepest path from the mean vector (Figure 8), rising quickly from the predicted value associated with the mean vector.

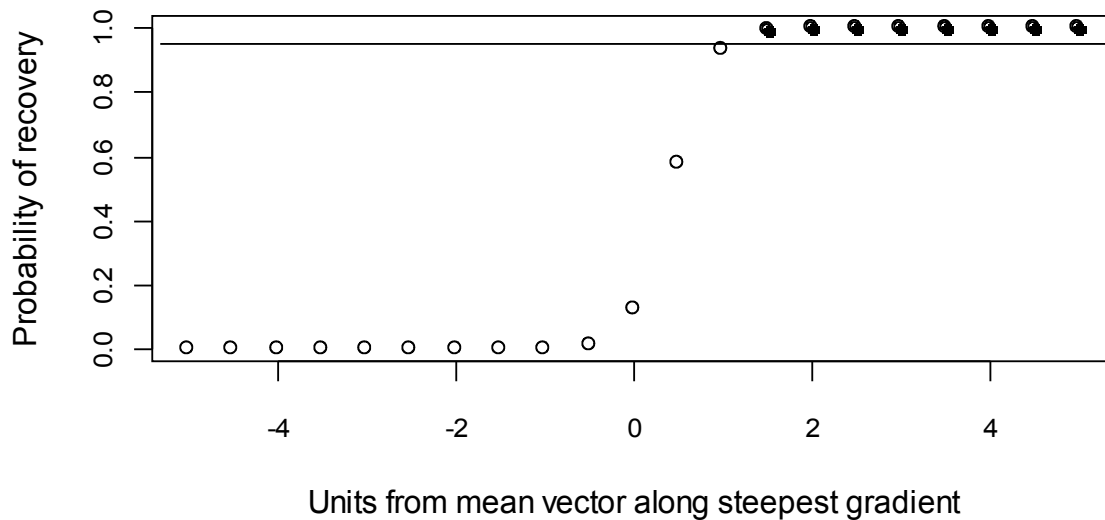


Figure 8: Predicted recovery rate along steepest path gradient

The following plots (Figure 9, Figure 10, and Figure 11) show the experimental design-space regions, two variables at a time, and plot points along the vector in either direction from the mean vector (the same points that were plotted in Figure 8). Solid

points indicate a probability of recovery in excess of 0.95. The area of predicted recovery rate greater than 0.95 includes the portion of the plot perpendicular to the steepest ascent line on either side of the solid points. These plots suggest that the experimental design space was largely outside of the space where recovery in excess of 95% could be expected. In particular, they suggest that the dimensions of the experimental design space should have been larger. Increasing the maximum of $E[X]$ would decrease the range of $E[Z]$ without adjusting its maximum value. There is also some suggestion that the observation period range should include smaller values, although this seems counterintuitive. The bottom line is that the experimental design space should be modified in a repeat experiment in order to obtain more scenarios where recovery was achieved. The objective would be to explore the region around the 0.95 contour of the response surface rather than to follow the gradient.

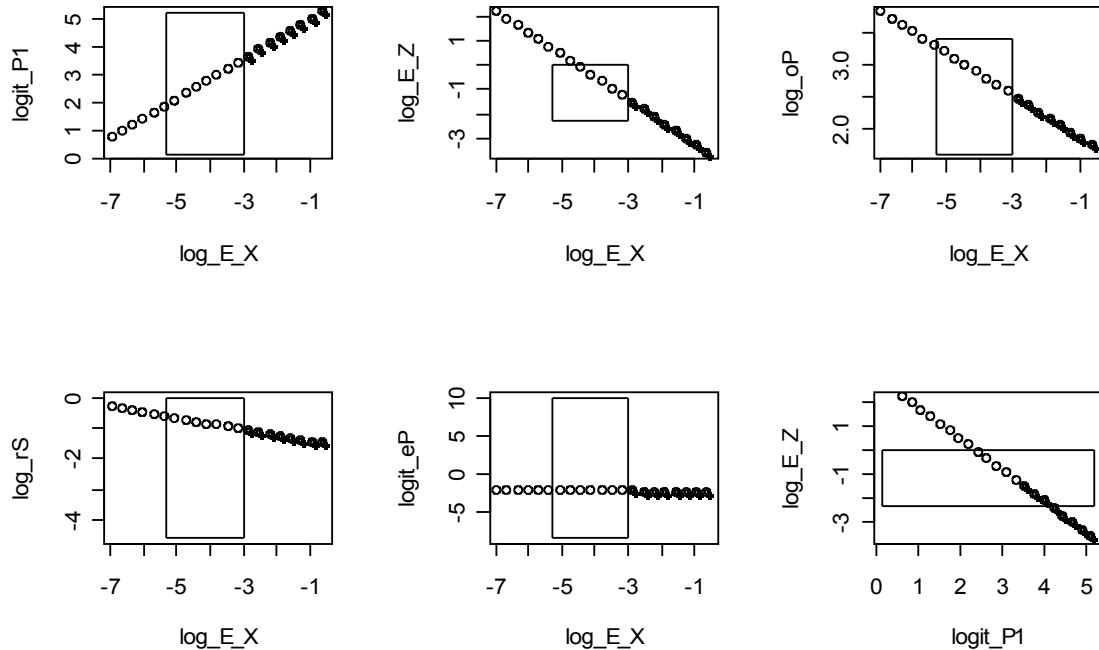


Figure 9: Steepest ascent gradient vs. process/observation characteristics I

There is not much graphical evidence (see particularly Figure 9 and Figure 11) to support changes to the ranges of event observation probability or to quasi-regular spacing interval, although one would think more-frequent observations would help in any scenario. Again, it should be remembered that a first order response surface model, when the input space is outside of the target response zone, is at best a crude approximation and is useful primarily for determining how to re-configure the experimental design space.

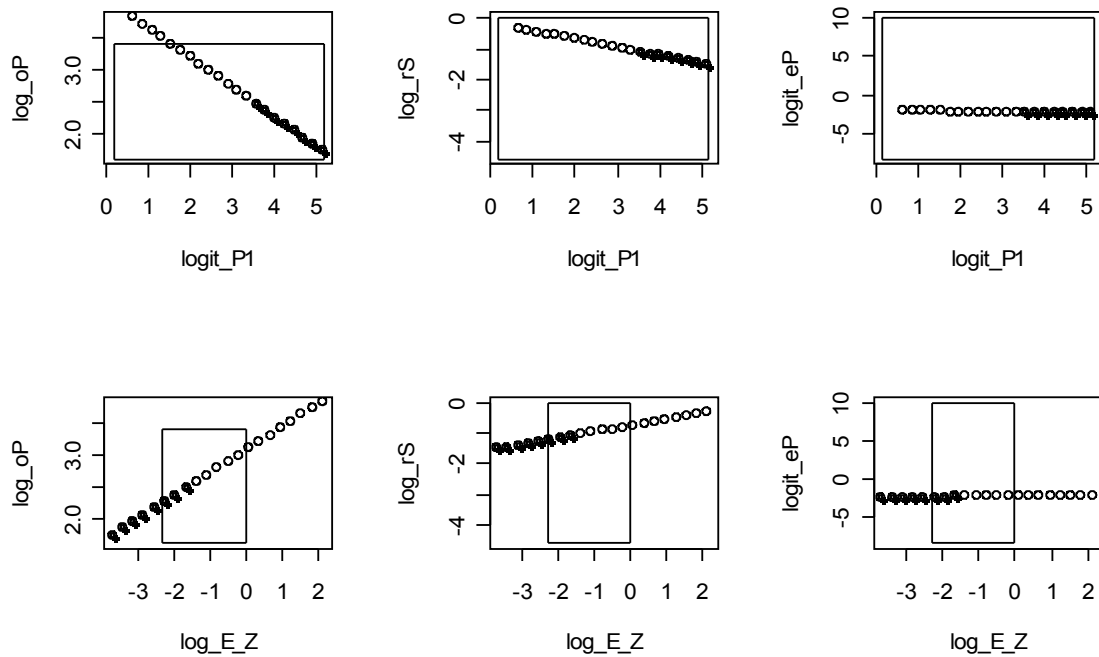


Figure 10: Steepest ascent gradient vs. process/observation characteristics II

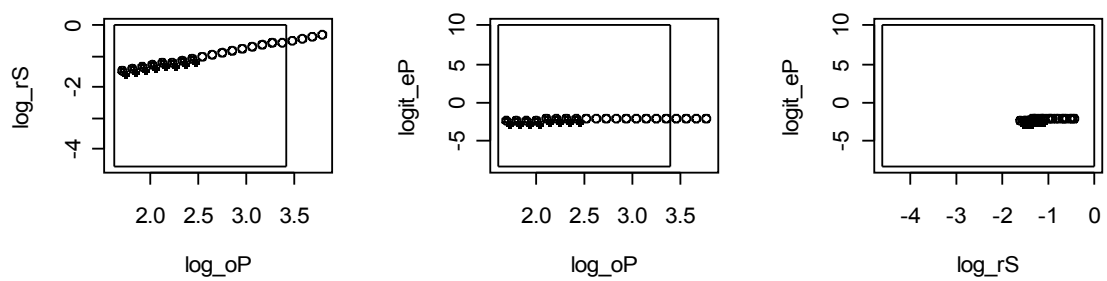


Figure 11: Steepest ascent gradient vs. process/observation characteristics III

4.4.3 Predictive performance

Predictive performance measured with AUC was very nearly equal for each of the three prior distributions tested (Table 11). The AUC was generally high in all three cases, exceeding 0.9 on over 84% of all successful evaluations and exceeding 0.8 on over 93% of them. In some cases, AUC was not successfully evaluated. The mean AUC values to four significant figures were 0.9388 (Gamma), 0.9374 (uniform), 0.9392 (Jeffreys). The maximum AUC values are 0.9973, 0.9973, and 0.9972 respectively. Less than one percent of the trials resulted in AUC values less than 0.5.

Table 11: Cumulative distribution of AUC values

AUC	>0.9	>0.8	>0.7	>0.6	>0.5	>0.4	>0.3	>0.2	>0.1	>0	N
Gamma	0.847	0.939	0.972	0.988	0.993	0.995	0.998	0.999	0.999	1.000	2998
Uniform	0.845	0.935	0.964	0.978	0.990	0.998	1.000	1.000	1.000	1.000	2998
Jeffrey's	0.847	0.939	0.971	0.988	0.994	0.997	0.999	0.999	1.000	1.000	2998

ISE can vary greatly. To give an idea of how ISE varies with distributional differences, Figure 12 shows the four most extreme cases for duration ISE and renewal ISE based on results using the Gamma prior and the two 95th percentile cases. For both duration ISE and renewal ISE, the minimum ISE cases appear to be very close matches between the nominal and estimated distributions. For the maximum ISE cases, the estimated distribution for duration ISE appears to be a reasonable match, yet its ISE is larger than that for the renewal ISE, which doesn't appear to be a close match at all.

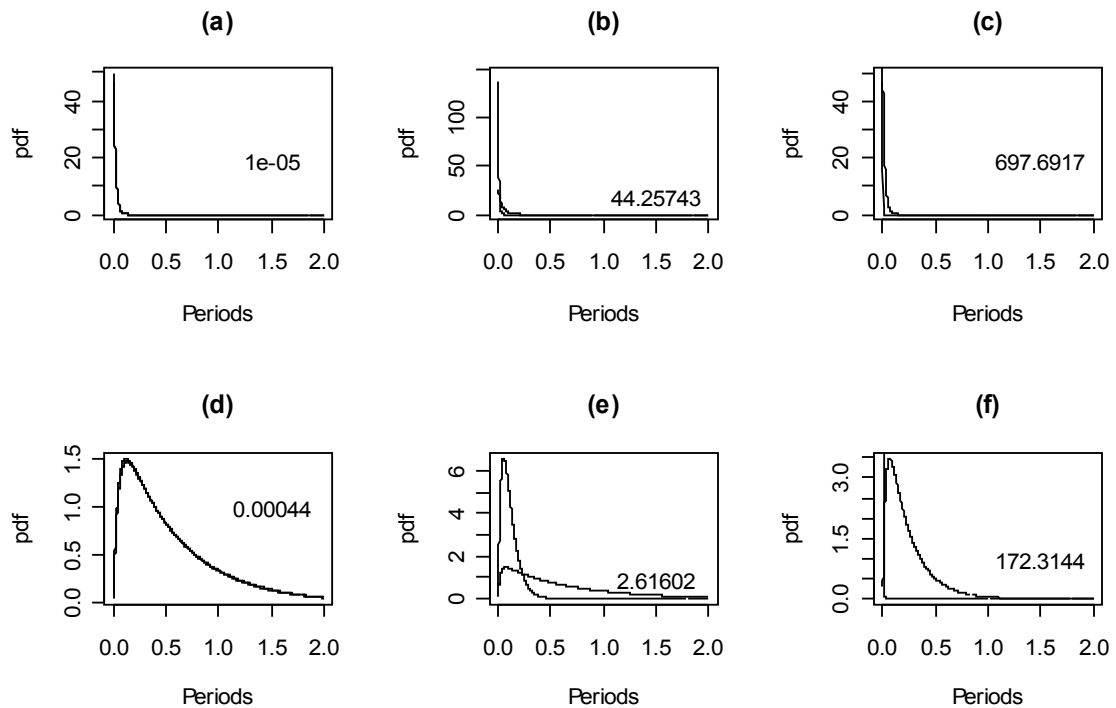


Figure 12: Nominal versus estimated distributions of duration and renewal period

Solid line is the nominal distribution. Dashed line is the estimated distribution. (a) minimum duration ISE (b) 95th percentile duration ISE (c) maximum duration ISE (d) minimum renewal ISE (e) 95th percentile renewal ISE (f) maximum renewal ISE.

4.4.4 Effects of process and measurement process characteristics on predictive performance

Using the results from the Gamma prior trials, principal components regression analysis was used to relate metric performance to process and observation protocol characteristics and to model selection. The process characteristic values used in the principal components analysis were the expected duration time, the expected renewal

time, and the magnitudes of the sinusoidal components of Model #2 and Model #3. The expected duration time and the expected renewal time are measures of the overall activity level within the system. The loadings from the principal components analysis are given in Table 12.

Table 12: PCA loadings for process and observation characteristics

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
obsPeriod				0.999			
regSpacing			-0.187		0.979		
eventProb	-0.568	0.417	0.126			0.120	-0.686
E_X	0.489		-0.623		-0.122	0.210	-0.551
E_Z	-0.344	0.442	-0.666			0.169	0.454
sqrt.amp.2	-0.378	-0.623				0.682	
sqrt.amp.3	0.421	0.477	0.341		0.115	0.670	0.131

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard deviation	1.37	1.09	1.01	1.00	1.00	0.762	0.583
Proportion of Variance	0.27	0.17	0.15	0.14	0.14	0.083	0.049
Cumulative Proportion	0.27	0.44	0.58	0.73	0.87	0.951	1.000

The interpretation of the components (Table 13) suggests that three of the components are process based, two are observation based, and two are a mix of the two. Component #1 is a mix interpreted as the experimental design constraint on the number of permissible violation-state observations per period. Thus, processes spending a lot of time in the violation state were limited to low violation-state observation probabilities. Processes spending little time in the violation state could have violation-state probabilities as high as one without exceeding the constraint. This is an equal effort component in that the interaction of the variables tends to maintain a constant periodic observation rate. Component #7 is a mix interpreted as a contrast between intensive

event observation and extensive event observation. High scores on this component correspond to both events of short duration and a low event observation probability, resulting in relatively few event observations.

Components #2, #3, and #6 are all process based components. Component #2 identifies the Model type. High scores are associated with a Model #3-like process, low scores with a Model #2-like process, and score near zero with Model #1 (or homogeneous processes). Component #3 is interpreted as differentiating between low transition rate processes (negative scores) and high transition rates (positive scores). Component #6 is interpreted as quantifying homogeneity, with more-negative scores indicating a more-homogeneous process and more-positive scores indicating a more non-homogeneous process (without regard to model type).

Components #4 and #5 are observation-based. Component #4 corresponds to observation period length (negative scores corresponding to shorter periods), and Component #5 corresponds to quasi-regular observation interval (negative scores corresponding to shorter (more intense) observation intervals).

Table 13: Interpretation of principal components

Compon ent	Interpretation	Low score	High score
#1	Violation state observation constraint (equal effort)	Shorter event duration, higher observation probability	Longer event duration, lower observation probability
#2	Non-homogeneity type	Like Model #2	Like Model #3
#3	Transition rates	Lower	Higher
#4	Observation period length	Shorter	Longer
#5	quasi-regular observation interval	Shorter	Longer
#6	Homogeneity	Homogeneous	Non-homogeneous
#7	Violation state observations	Longer duration, higher observation probability => more observations	Shorter duration, lower observation probability => less observations

AUC was logit-transformed for this regression. Component #4 was dropped from the regression because it was not statistically significant. The resulting model was statistically significant although it left a substantial proportion of the variance unexplained (Table 14). Higher AUC was associated with shorter duration, higher event observation probability, Model #3-like non-homogeneity, lower transition rates, shorter quasi-regular observation intervals, more non-homogeneous processes, and more event observations. Higher AUCs were obtained when Model #3 was the test model than when the test model was either of the other two. The intercept value is associated with an AUC of 0.9603 for datasets fit with Model #1. This rises to 0.9678 for datasets fit with Model

#3. If the dataset arose from a Model #3-like process indicated by positive scores on Components #2 and #6, the AUC could be expected to be even higher.

Table 14: Regression results for AUC

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.1861    0.0249  127.99 < 2e-16 ***
Comp.1        -0.2933    0.0105  -27.95 < 2e-16 ***
Comp.2         0.3372    0.0132   25.47 < 2e-16 ***
Comp.3        -0.2725    0.0143  -19.10 < 2e-16 ***
Comp.5         0.0691    0.0144    4.79 1.7e-06 ***
Comp.6         0.1850    0.0189    9.79 < 2e-16 ***
Comp.7        -0.5257    0.0247  -21.31 < 2e-16 ***
as.factor(testModel)2  0.0541    0.0352    1.53  0.12
as.factor(testModel)3  0.2173    0.0353    6.16 8.1e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.79 on 2986 degrees of freedom
Multiple R-squared:  0.447,    Adjusted R-squared:  0.445
F-statistic:  301 on 8 and 2986 DF,  p-value: <2e-16

```

The ISE statistics were log-transformed for these regressions. The implication of a log transformation is that the model becomes a multiplicative model. The multiplicative power of each variable is equal to the value of e raised to its coefficient value. If the coefficient value is equal to zero the multiplicative value is one, that is it does not change the dependent variable. Coefficient values greater than zero increase the dependent value; values less than zero decrease it.

Component #2 was removed from the regression for duration ISE because it was not statistically significant. Again this model was significant but left a substantial proportion of the log duration ISE variation unexplained. Low duration ISE was associated with longer event duration and lower event observation probability, lower

transition rates, longer observation period length, shorter quasi-regular observation intervals, more non-homogeneous behavior, and more event observations (Table 15).

Duration ISE statistics associated with estimations based on Model #3 were significantly higher than those associated with Model #1, which were significantly higher than those associated with Model #2. The mean duration ISE for Models #1, #2, and #3 was 5.777, 5.018, and 7.127 respectively.

Table 15: Regression results for duration ISE

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.7539	0.0311	56.42	< 2e-16	***
Comp.1	-0.3828	0.0131	-29.21	< 2e-16	***
Comp.3	0.5243	0.0178	29.43	< 2e-16	***
Comp.4	-0.0403	0.0180	-2.24	0.0249	*
Comp.5	0.9510	0.0180	52.87	< 2e-16	***
Comp.6	-0.1262	0.0236	-5.35	9.6e-08	***
Comp.7	0.1734	0.0308	5.63	2.0e-08	***
as.factor(testModel)2	-0.1408	0.0440	-3.20	0.0014	**
as.factor(testModel)3	0.2100	0.0440	4.77	1.9e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.98 on 2986 degrees of freedom
 Multiple R-squared: 0.609, Adjusted R-squared: 0.608
 F-statistic: 581 on 8 and 2986 DF, p-value: <2e-16

Components #4 and #5 were removed from the regression for renewal ISE as was the categorical variable for the model because they were not statistically significant.

Although this model was statistically significant, it only explained about 27% of the total variation in log renewal ISE. However, the coefficient values were highly significant.

Lower renewal ISE was associated with longer violation-state duration and lower violation-state observation probability, Model #2-like non-homogeneity, lower transition

rates, more non-homogeneous processes, and more event observations (Table 16). The renewal ISE associated with the intercept was 0.4218.

Table 16: Regression results for renewal ISE

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.8632	0.0213	-40.55	< 2e-16	***
Comp.1	-0.4077	0.0155	-26.27	< 2e-16	***
Comp.2	0.2320	0.0196	11.85	< 2e-16	***
Comp.3	0.2400	0.0211	11.37	< 2e-16	***
Comp.6	-0.1296	0.0279	-4.64	3.7e-06	***
Comp.7	0.4366	0.0365	11.96	< 2e-16	***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.2 on 2989 degrees of freedom
 Multiple R-squared: 0.273, Adjusted R-squared: 0.272
 F-statistic: 225 on 5 and 2989 DF, p-value: <2e-16

4.5 Discussion

In this paper three different two-state continuous-time Markov chain (CTMC) models based on the Kolmogorov Backward Equation were evaluated using simulations of infrequent, unevenly spaced, and uncoordinated observations. The evaluation consisted of assessments of prior distribution selection, of nominal parameter recovery, of out-of-sample prediction performance, and of the effect of various process and observation characteristics on said performance. The purpose of these assessments was to establish the conditions under which the methodology could be applied, assuming that successful application would be a function of both process characteristics and of observation protocol characteristics.

There was no strong consensus among the results as to the preferable prior, although the Jeffreys prior yielded fairly clearly inferior results. The Gamma distribution at least is a conjugate prior for the exponential distribution, the distribution of the duration periods associated with parameters λ and μ , and outperforms the uniform prior 2-to-1 on head-to-head comparisons. The uniform distribution outperforms the Gamma on the AUC, which is the only metric based on an out-of-sample evaluation. However, the AUC as used here may not be the ideal metric, as discussed previously.

Nominal parameter recovery was found to be biased towards overestimation of duration, frequency, and long-term violation state probability. This bias favors the conservationist. The lack of a large number of scenarios where recovery was obtained on all parameters of interest casts some doubt on the inferences that can be made from the other analyses. An analyst could get good performance on a cross-validation test of prediction performance and be misled into thinking that the true process characteristics had been adequately estimated. Good performance on traditional validation tests is necessary but not sufficient to obtain good recovery of true process parameter values. With real data the true value is usually not known, however.

Quasi-regular observation interval, event observation probability, expected violation-state duration, and expected renewal interval were set to ranges that would produce anywhere from ~40 to ~200 observations per period, which is a reasonable number for a long-term monitoring project with multiple sites. The recovery results

suggest that either the overall observation rate was insufficient, or that the preferential observation of the violation state biased the results.

A natural expectation is that those process and observation protocol characteristics that lead to better predictive performance as measured by higher AUC should lead to lower duration and renewal ISE. One indication of that would be that the component coefficients would differ in sign between the AUC regression and the ISE regressions. However, that was not the case for three of the six components where such a comparison could be made (component #4 was not included in the models for AUC or renewal ISE). Opposite signs were found for components #3 (transition rates), #6 (homogeneity), and #7 (violation-state observations), but not for components #1 (violation-state observation effort), #2 (model type), or #5 (quasi-regular observation interval). There was agreement between the coefficient signs for duration and renewal ISE for all four components shared by both (#1, #3, #6, and #7).

One possible explanation for this discrepancy is that the ISE statistics are based on a comparison between the nominal and estimated distributions for duration and renewal times, whereas the AUC is based on an out-of-sample prediction of observations made using the same protocol as the observations used in the estimation, and not from predictions of out-of-sample observations made at a high frequency, which would more faithfully replicate the properties of the nominal process. It is not clear how this difference could explain the results found. However, one could speculate that the out-of-sample observations simulated using an observation protocol that resulted in a sparse set

of observations, even if some were spaced relatively closely in time, might lack features of the full process that would be represented in a nominal distribution. That is, the observation process is a sort of filter that transforms the underlying process into a new process, with possibly different properties. However, it could be speculated that an observation process with invariant properties, applied to an underlying process with invariant properties, would result in a new process with invariant properties. This would explain the good performance on AUC.

Typically in a classification problem the costs due to Type I and Type II errors can be minimized by a judicious choice of the cutoff value used in prediction of cases of unknown class. The proper cutoff value can be determined from expected error costs and the same ROC curve used to calculate AUC, the metric used in much of this work. The conservationist, in the context of this experiment where the violation state is state “0”, prefers a false negative error to a false positive error. That is they assume the cost of a false prediction of violation is less than a false prediction of non-violation. The overall cost given such a preference can be minimized by using a cutoff value greater than what would be optimal given equal costs. In many classification procedures knowledge of the error cost structure can be built in to the model. It was not obvious how to do that in this case. However, one wonders if the choice of observation protocol did not have a similar, if difficult to quantify, effect. Given just quasi-regularly spaced observations, one might expect transition rates to be underestimated, resulting in overestimates of duration and underestimates of frequency, but one would expect long-term probability to be estimated

without bias, if not without precision. The fact that long-term violation state probability was overestimated could be a result of the choice to preferentially observe that particular state. In actual practice event observation cannot normally be governed to only observe one state; some event-based observation of both states generally occurs. Unresolved questions are whether this difference would result in less recovery bias, and whether there might be modifications to the observation protocol that would result in less bias.

While five periods (if periods are years) is pretty short for a typical long-term monitoring project, it is pretty long for a typical research project. Therefore these results may not extrapolate well to shorter research projects, although it is expected general tendencies would still hold. There also may be situations where an event occurring much less frequently than once a period could have a significant impact on some ecosystems. While expected duration and renewal periods were specified to be less than one period, these results could be adapted to both such situations by a suitable definition of the length of one period, however the usual caveats concerning extrapolation apply.

Does model selection matter? Clearly the two non-homogeneous two-state CTMCs models advanced here do not exhaust the possibilities. Model selection may not matter that much given the constraints imposed by the observation protocols considered, which generally result in very sparse observations of the process, making processes indistinguishable. Initially the regression results appear to be conflicting. However, a careful examination reveals consistent differences in predictive performance depending on the process generating the observations and the model selected. Components #2 and

#6 have been interpreted as describing process type and degree of homogeneity respectively. A reasonable approach would be to form subsets of the data corresponding to scores on these two components. Homogeneous processes should score near zero (say within the middle tercile) on component #2 and below zero on component #6. Non-homogeneous processes following Model #2 should score in the negative (say the bottom tercile) on component #2 and above zero on component #6. Non-homogeneous processes following Model #3 should score in the positive (say within the upper tercile) on component #2 and above zero on component #6.

Grouping the results following these criteria and performing signed rank tests on transformations of AUC, duration ISE, and renewal ISE indicates significant differences between groups in all three cases, and also indicates that in all three cases, predictive performance is worse for the homogeneous processes than for the non-homogeneous processes, and best for non-homogeneous processes following Model #3 (Table 17). Predictive performance for non-homogeneous processes following Model #2 is generally better than for homogeneous processes and worse than non-homogeneous processes following Model #3.

Table 17:

Metric	Group	95% confidence interval	(pseudo) median	p-value
logit(AUC)	#1 - #2	-0.15001 0.05933	-0.0445	0.399
	#1 - #3	-0.5043 -0.2952	-0.3994	2.831e-13
	#2 - #3	-0.4507 -0.2575	-0.3546	1.211e-12
log(duration ISE)	#1 - #2	-1.2982 0.1931	-0.5223	0.1549
	#1 - #3	-0.08295 1.19495	0.5417	0.08986
	#2 - #3	0.5167 1.7735	1.135	0.0002107
log(renewal ISE)	#1 - #2	0.05335 0.15981	0.1049	3.918e-05
	#1 - #3	0.04071 0.14320	0.08993	0.0002489
	#2 - #3	-0.05612 0.02015	-0.01724	0.3668

4.6 Conclusions/Recommendations

Two-state continuous time Markov chains are simple processes which are nonetheless useful for modeling a wide variety of phenomena. Non-homogeneous versions are a bit more complicated but extend the range of phenomena which can be modeled considerably beyond the homogeneous case. The objective of this experiment was to evaluate the performance of three two-state hierarchical CTMC models (one homogeneous and two non-homogeneous), as components of a measurement process that

includes observation and modeling, on simulated observations of WQIs. The modeling phase included a Bayesian estimation procedure.

This work began with a few simple equations relating observations of a two-state CTMC and established a baseline for development of a general hierarchical model for regional, or multi-process, threshold exceedance behavior of WQIs. Some insight into the general problem was established by conducting experiments with single-process simulations looking at not only the effect of model selection, but also the effect of various process observation protocols on the ability of the measurement process to make accurate and precise measurements of important process characteristics, in this case the distributions of time spent in the violation state and the frequency with which such excursions occur.

Three processes were simulated to represent WQIs: 1) a two-state homogeneous CTMC, 2) a two-state non-homogeneous CTMC with varying transition rates but constant limiting probabilities; and 3) a two-state non-homogeneous CTMC with varying limiting probabilities but constant expected renewal period.

In a preliminary step the impact of using three different prior distributions for the transition rate and renewal rate parameters was examined. The Wilcoxon signed rank test was used to make pairwise comparisons between performance metrics resulting from the use of uniform, Gamma, and Jeffrey's prior distributions. Gamma and uniform distributions used as priors for transition rate parameters were found to outperform a

Jeffreys prior. The under-performance of the Jeffrey's prior may be due to the implementation method required by the software.

Nominal parameter recovery analysis indicated that the experimental design space did not result in a large number of scenarios where nominal parameter values were recovered. Duration, frequency, and long-term violation state probability were all overestimated. First, this qualifies the other analyses. Second, it suggests that the experiment should be repeated in order to better define the suitable conditions under which the method may be used. Currently the optimal operating conditions of the method cannot be defined with any precision.

The Gamma prior results were selected and a principal components regression analysis performed on the effects of process and measurement process characteristics on the performance metrics. Differences were noted in the performance metrics simply related to the process characteristics. Homogeneous processes produced lower AUC and higher ISE metrics than the two non-homogeneous processes. The varying limiting-probability (Model #3) processes produced higher AUC and lower ISE metrics than the other two processes.

Observation protocols had a substantial impact on predictive performance and distribution estimation error. As would be expected, longer observation period and more observations, whether event-based or routine, lead to reduced ISE for violation-state duration and expected renewal period distributions. Results for AUC were inconsistent

with those for ISE. This inconsistency needs to be investigated further, but may be related to the recovery results.

Benchmarks were established for model estimation performance given processes arising from the models considered and observation protocols that would result in a sequence of infrequent and unevenly spaced observations. Predictive performance was good despite sparse observations of the process. Regardless of the model simulated or selected for estimation, AUC exceeded 0.8 in over 93% of the cases and over 0.9 in over 84% of cases. Duration ISE ranged from near 0 to over 697 with a 95th percentile value of 44.38. Renewal ISE ranged from near 0 to over 172 with a 95th percentile value of 2.61.

Future research should first determine the conditions under which good nominal parameter value recovery occurs. Additional research should test further extensions of these models in a multiple process framework, test them against simulated observations more closely resembling those of WQIs, test them against existing methods used in hydrology, and further develop the hierarchical levels of the models to allow for prediction of process parameters using physical characteristics predictive of WQIs.

5 Multiple-chain Markov models for threshold exceedance

5.1 Abstract

Thresholds have been defined for many water quality indicators (WQIs) that separate the measurement space of the indicator into two states, one of which, the exceedance or violation-state, has undesirable consequences. Observations of the indicator are often made at relatively infrequent and unevenly spaced intervals and are uncoordinated with the precise timing of changes in state. In addition, observations made at different locations are usually asynchronous as well. These typical observation protocols make estimation of frequency, duration, and long-term probability properties difficult. To address this problem, a hierarchical model was developed for multiple two-state CTMCs based on the Kolmogorov Backward Equation. Observation and modeling of WQI processes was simulated using dual correlated two-state continuous time Markov chains. Each process was given different stochastic properties. Simulated asynchronous, uncoordinated, and unevenly spaced observations of each process were made using observation protocols comprising a common observation period but different quasi-

regular observation intervals and violation-state observation probabilities. Model parameters for both processes were estimated from a training set of simulated observations both separately and jointly using a Bayesian MCMC method. A response surface analysis was conducted to assess nominal process model parameter recovery. Predictive performance was assessed on test datasets using the area under the ROC curve. Recovery was equally poor over the experimental design space regardless of method, and indicated a bias favoring conservationists: overestimation of violation-state properties (longer duration, higher frequency, and higher long-term violation-state probability). Both methods exhibited generally good predictive performance, but neither was superior to the other. Positively correlated and relatively low transition-rate processes were easier to predict. Several observation characteristics resulted in better prediction: event observation intensity, quasi-regular observation intensity, and length of observation period.

5.2 Introduction

Regulators have long used thresholds to define water-quality standards because serious harm can come to aquatic ecosystems when water-quality thresholds are violated too often or for excessively long periods of time. Typically, standards require that WQI concentrations remain on one side of the threshold (above or below) at least a specified percentage of time during some period such as a year or a defined season. Recently ecologists also have begun to look at thresholds as indicators of climate change (Lenton

et al. 2008, Rosenzweig et al. 2008, Keller, Yohe & Schlesinger 2008). Changes in the stochastic properties of a process with respect to a threshold could indicate change in climatic conditions. However, for many WQIs the high-frequency observations necessary to identify and to time periods of violation are not available. In some cases only a subset of locations within a region of interest may actually have observations. And although two WQI processes might be correlated, observations typically are not made simultaneously on both processes. All of this makes it difficult to determine violation-state properties for specific locations and to extrapolate from data-rich locations to data-poor locations.

Eventually it will be desirable to be able to model many processes at one time because often processes within a region (a population of processes) move together roughly as a group or set of groups, and it will be desirable to be able to take advantage of common characteristics among processes and to make predictions for processes within a region for which there is little or no water-quality data, perhaps using other characteristics common to all member processes in the region other than water quality itself. However, devising an all-inclusive experimental framework for testing models for such a problem is difficult.

Markov processes have been widely discussed in hydrology (Lu, Berliner 1999, Szilagyi, Balint & Csik 2006 and references therein). A number of researchers have developed models for threshold violations that leverage extreme value theory and Markov chains (Smith, Tawn & Coles 1997 and references therein). Others have used

statistical (Deviney, Rice & Hornberger 2006) or process (Zhang, Arhonditsis 2008) models to make time series predictions from which threshold violation properties can be estimated.

The objective of this paper, as a first step, was to evaluate models for dual process CTMCs under simulated observation conditions, and to compare performance under individual and joint estimation, leaving the additional complexities mentioned above for later work. The evaluation consisted of assessments of nominal parameter recovery, of out-of-sample prediction performance, and of the effect of various process and observation characteristics on said performance. The purpose of these assessments was to compare the two estimation methods used and to establish the operational characteristics of the methodology, assuming that performance would be a function of both process characteristics and of observation protocol characteristics.

Following in this section, a model for dual correlated CTMCs is presented which can be used for process simulation. Next the derivation of a two-state model for multiple CTMC processes is developed which will be used for estimation.

5.2.1 Correlated CTMCs

Suppose that for two locations A and B we have the following state diagram (Figure 13) where the state is the ordered pair (S^A, S^B) . Assume that the probabilities of both elements changing at precisely the same time (that is, from (0,0) to (1,1) or from (0,1) to (1,0)) are zero. If the transition rate from (S^A, S^B) to $(\sim S^A, S^B)$ or from

(S^A, S^B) to $(S^A, \sim S^B)$ is independent of the status of the non-changing element, then S^A and S^B are uncorrelated. That is, if $\lambda_{A0} = \lambda_{A1}$ and $\lambda_{B0} = \lambda_{B1}$ and $\mu_{A0} = \mu_{A1}$ and $\mu_{B0} = \mu_{B1}$. S^A and S^B are correlated if the transition rates for one element depend on the state of the other element. That is, if at least one of the above equalities does not hold.

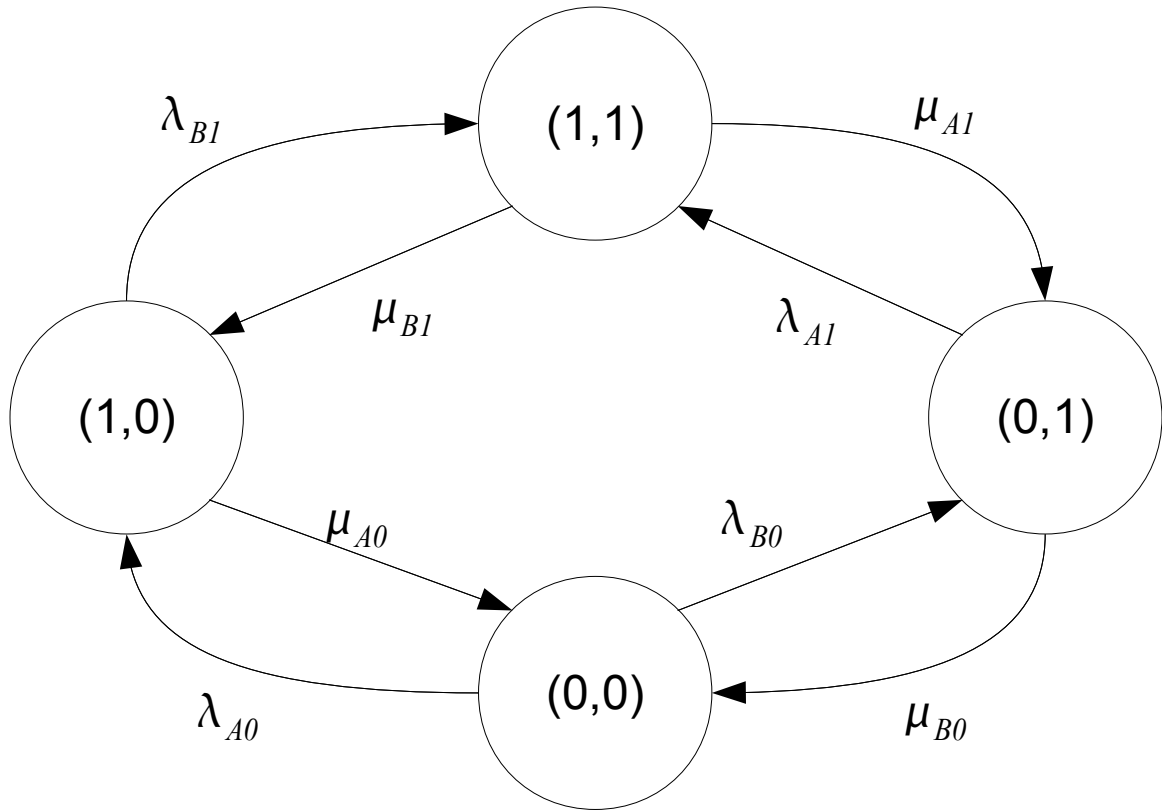


Figure 13: Correlated dual process CTMCs

The measure of correlation ρ between A and B is given by

$$\rho = \frac{P[S^A = '1', S^B = '1'] - P[S^A = '1']P[S^B = '1']}{\sqrt{P[S^A = '1']P[S^A = '0']P[S^B = '1']P[S^B = '0']}} \quad (41)$$

where the terms of the numerator and denominator can be determined from the balance equations and the limiting probabilities. Prediction equations for the state of each process are given in Equation (42).

$$\begin{aligned}
 Prob(S_n^A = '1') &= \left(\begin{aligned} &(P_{10}^A F_{n0}^A + (1 - F_{n0}^A) P[S_{n-1}^A = '1']) P[S_{n-1}^B = '0'] \\ &+ (P_{11}^A F_{n1}^A + (1 - F_{n1}^A) P[S_{n-1}^A = '1']) P[S_{n-1}^B = '1'] \end{aligned} \right) \\
 Prob(S_n^B = '1') &= \left(\begin{aligned} &(P_{10}^B F_{n0}^B + (1 - F_{n0}^B) P[S_{n-1}^B = '1']) P[S_{n-1}^A = '0'] \\ &+ (P_{11}^B F_{n1}^B + (1 - F_{n1}^B) P[S_{n-1}^B = '1']) P[S_{n-1}^A = '1'] \end{aligned} \right) \\
 P_{10}^A &= \lambda_{A0} / (\lambda_{A0} + \mu_{A0}) \\
 F_{n0}^A &= 1 - e^{-(\lambda_{A0} + \mu_{A0})\Delta t} \\
 P_{11}^A &= \lambda_{A1} / (\lambda_{A1} + \mu_{A1}) \\
 F_{n1}^A &= 1 - e^{-(\lambda_{A1} + \mu_{A1})\Delta t} \\
 P_{10}^B &= \lambda_{B0} / (\lambda_{B0} + \mu_{B0}) \\
 F_{n0}^B &= 1 - e^{-(\lambda_{B0} + \mu_{B0})\Delta t} \\
 P_{11}^B &= \lambda_{B1} / (\lambda_{B1} + \mu_{B1}) \\
 F_{n1}^B &= 1 - e^{-(\lambda_{B1} + \mu_{B1})\Delta t}
 \end{aligned} \tag{42}$$

5.2.2 The two-state CTMC Model

A two-state CTMC is a reasonable first choice for a model of threshold exceedance. Let one state (say state '0') represent the below-threshold condition and the other state (state '1') represent the above-threshold condition. Let state '0' represent the violation state. Assume that upon entering one state, a random amount of time passes before the process transits to the other state. Once in the other state, a random amount of

time passes before the process returns to the original state. Assume these random times have exponential distributions with rates λ and μ that are not necessarily equal.

Some important properties of the process are given in Equation (43), where $E[X]$ represents the expectation of the amount of time spent in state '0' on a typical excursion into that state, $E[Y]$ represents the same expectation for state '1', $E[Z]$ is the expectation of the time between entrances into either state, P_0 is the long-term proportion of time spent in state '0', and P_1 is the long-term proportion of time spent in state '1'.

$$\begin{aligned}
 E[X] &= \frac{1}{\lambda} \\
 E[Y] &= \frac{1}{\mu} \\
 E[Z] &= E[X] + E[Y] \\
 P_0 &= \frac{\mu}{\lambda + \mu} \\
 P_1 &= \frac{\lambda}{\lambda + \mu}
 \end{aligned} \tag{43}$$

Ross (1993) showed how the two-state homogeneous CTMC model and the Kolmogorov Backward Equations lead to equations that predict the probability of being in state "0" having been in state "0" or having been in state "1" some time previous.

Ross gives these equations respectively as:

$$\begin{aligned}
 P_{00}(t) &= \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)\Delta t} + \frac{\mu}{\lambda + \mu} \\
 P_{10}(t) &= \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)\Delta t}
 \end{aligned} \tag{44}$$

If we define S_n as the state of the system at the time t_n , then with some re-arrangement of terms we can write the probability that S_n is “1”, having observed the state at $n-1$, as

$$Prob(S_n = '1') = \begin{cases} P_1 F_n & , S_{n-1} = '0' \\ 1 - (1 - P_1) F_n & , S_{n-1} = '1' \end{cases} \quad (45)$$

where 1) P_1 is the limiting probability of being in state “1” given λ and μ , and 2)

F_n is the CDF of an exponential random variable with rate $\lambda + \mu$. Note that this rate is not the renewal rate. Equation (45) can be written in a more general form as

$$Prob(S_n = '1') = P_1 F_n + (1 - F_n) Prob(S_{n-1} = '1') \quad (46)$$

which can be seen as a weighted average between P_1 , the limiting probability for state “1”, and the state at $n-1$. Given independent observations of multiple two-state homogeneous CTMCs that include the time between observations and the previous and current states, the process parameters may be determined with Bayesian estimation methods using OpenBUGS software (Thomas et al. 2006) in R and the following joint model specification:

$$\begin{aligned} S_{j,n} &\sim \text{Bernoulli}(P_{j,1} F_{j,n} + (1 - F_{j,n}) S_{j,n-1}) \\ P_{j,1} &= \lambda_j / (\lambda_j + \mu_j) \\ F_{j,n} &= 1 - e^{-(\lambda_j + \mu_j) \Delta t_{j,n}} \\ \Delta t_{j,n} &= t_{j,n} - t_{j,n-1} \\ \lambda_j, \mu_j &\sim \text{Uniform}(0.01, 10000) \end{aligned} \quad (47)$$

Note that the only difference between the individual specification and the joint specification is the presence of the subscript j to indicate parameters for different processes. The choice of a uniform prior was justified in the preceding chapter. It is non-informative and results in maximum-likelihood estimates, theoretically. The choices of 0.01 and 10,000 for the limits of the distribution are justified as being well outside the expected rate values, which range from 1 to 200 times per period, as they correspond to rates equivalent to once per hundred periods and 10,000 times per period, respectively.

The rest of this paper proceeds as follows: In Section 5.3 the method used in this experiment is described. The method included three levels of simulation, model estimation, determination of performance metrics, and results analysis. In Section 5.4 results are presented. In Section 5.5 the results, implications of this work, and shortcomings are discussed. In Section 5.6 conclusions are presented.

5.3 Method

High-frequency (HF) observations of dual correlated CTMC processes with different stochastic properties were simulated. These HF observations were sub-sampled at asynchronous, uncoordinated and unevenly spaced (AUUS) intervals to produce a set of AUUS observations for both processes typical of real-world observations of WQIs made with respect to a threshold. These AUUS observations were divided into a training set and a test set. The process properties of models for both processes were estimated

both individually and jointly using Bayesian estimation methods. Predictions of the test set were made and performance metrics calculated. These results were evaluated to compare the individual versus the joint estimation methods. Recovery of nominal parameter values was assessed for bias and to determine the operational conditions of good recovery. Effects of process and observation protocol characteristics on predictive performance were also evaluated.

5.3.1 Experimental design

Nominal parameter value sets for individual scenarios of the experiment were determined using a combination of leaped Halton sequences (Kocis, Whiten 1997) and random selection. Initially 10,000 parameter sets were generated for the following design parameters: *observation period length*, *quasi-regular observation interval* for process A, *quasi-regular observation interval spacing* for process B, *event observation probability* for process A, *event observation probability* for process B, *expected violation-state duration* for process A given process B in state '0' ($1/\lambda_{A0}$ in Figure 13), *expected violation-state duration* for process A given process B in state '1' ($1/\lambda_{A1}$ in Figure 13), *expected renewal period* of process A given process B in state '0' ($\frac{1}{\lambda_{A0}} + \frac{1}{\mu_{A0}}$ in Figure

13), and expected renewal period of process A given process B in state '1' ($\frac{1}{\lambda_{A1}} + \frac{1}{\mu_{A1}}$ in Figure 13). Leaped Halton sequences were generated with the *sfsmisc* package in R.

Levels for the four parameters associated with process B were determined by sampling from uniform distributions about their equivalents for process A. The endpoints of these intervals were chosen so that the range of ratios between the two equivalent

parameters was the same. In other words, $\max\left(\frac{X_A}{X_B}, \frac{X_B}{X_A}\right) \sim Uniform(1, K)$.

5.3.2 Process simulation

Dual two-state correlated CTMC processes A and B were simulated for each scenario. Each process was simulated by generating simulated observations of the process at the rate of 10,000 per unit time. Compared to the solar year, this is roughly equivalent to once an hour. The expected duration in the violation state ($E[X_{S_A}|S_B]$) was chosen from a range of 0.005 to 0.05, corresponding roughly to a range of a few hours to a few weeks (Table 18). The expected renewal period ($E[Z_{S_A}|S_B]$) was chosen from a range of the expected duration up to 1 full period. From these values the transition rates necessary for the simulation and limiting state probabilities can be calculated.

Table 18: Levels of important process characteristics of dual CTMCs

Variable	Range	Description
$E[X_{S_A} S_B]$	0.005 to 0.05 time units	expected value of duration in the violation state of process A given the state of process B
$E[Z_{S_A} S_B]$	$E[X_{S_A} S_B]$ to 1 time units	expected value of the renewal period of process A given the state of process B
Correlation	-1 to 1	Correlation between process A and process B

It is difficult to specify transition rates for the two processes a priori that will yield a desired correlation value precisely, although it is easy enough to calculate the expected correlation either from a set of transition rate specifications or from simulated data. Parameter sets for 10,000 pairs of processes were generated and the correlation between each pair was calculated using Equation (41). Correlations spanned the range from negative one to one, but the vast majority of pairs had correlations near zero. One thousand of these pairs were then selected without replacement using an inclusion probability weight equal to the absolute value of the correlation between the pairs. These 1,000 were used for the subsequent estimation experiment. Replicate simulated observations were generated for each of the 1,000 parameter sets, yielding 2,000 scenarios. Characteristics retained their distributions except for correlation and event observation probability (Figure 14).

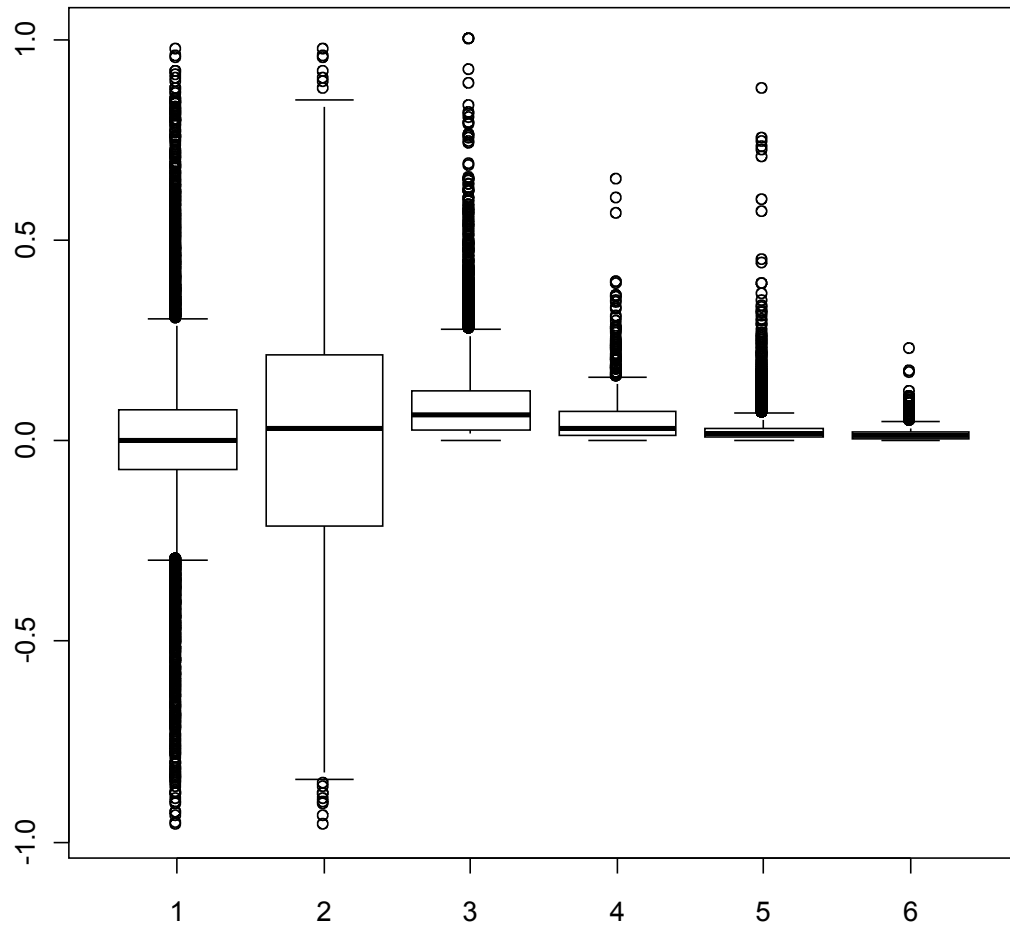


Figure 14: Distributions of process and observation protocol characteristics

Characteristic distributions before (1,3,5) and after (2,4,6) selection based on correlation. 1&2 – correlation between A and B, 3&4 – event observation probability – A, 5&6 – event observation probability – B.

5.3.3 Observation simulation

The observation process was simulated by selecting a single observation period length for both processes but a different quasi-regular observation interval and a different event observation probability (Table 19). The observation period length is the total number of units of time that the processes are to be “observed” in simulation. The quasi-regular observation interval is the interval at which quasi-regular observations are to be made, meaning the exact interval size is a random variable. The event observation probability is the probability that an observation will be made at a given observation opportunity during an excursion into the violation state. Synchronous simulated observation of both processes occurred only by chance and generally with very low probability. This mimics to some extent how observations are made in actual practice.

Table 19: Observation protocol characteristics

Characteristics	Range	Description
Observation period length	5 to 30 units	Total period length of observation
quasi-regular sampling interval	0.01 to 1 units	Quasi-regular spacing \sim $Normal\left(x, \left(\frac{x}{100}\right)^2\right)$
Event sampling probability	0 to 1, but constrained to limit the total number of event observations	Event sampling yielding on average ≤ 100 observations per period

5.3.4 Modeling simulation

The training set of each pair of processes was used to estimate process parameters by both estimating each process's parameters individually and by estimating them jointly. OpenBUGS (Thomas et al. 2006) and R were used to perform the estimations on the Cross-Campus Grid (XCG) at the University of Virginia (Morgan, Grimshaw 2007). OpenBUGS is a software package which implements the Gibbs sampling method of Monte Carlo Markov Chain (MCMC) Bayesian estimation. Models are prescribed using R-like syntax and functions. From random starting values, a vector of parameter values moves forward iteratively according to specified prior distributions and model specifications. The user may specify multiple *chains* of parameters values, whose *convergence* to a common distribution may be tracked using the Gelman-Rubin statistic (\hat{R}). Several model fit statistics are generated, including the *deviance information criterion*, or DIC.

Estimations were performed using three chains. After burn-in, sufficient iterations were run to obtain 350 values from each chain, for a total of 1050 samples from the posterior distributions. A ubiquitous problem in Bayesian MCMC estimation is determining a proper (minimum) burn-in period and a thinning parameter. Usually these choices are made while operating the software interactively. The necessary burn-in period and thinning parameter can vary widely from one estimation run to another, however. With a total of 6,000 estimations to perform, each requiring anywhere from 15 minutes to many hours, a systematic method of performing the estimation was needed.

Each estimation run was therefore initialized with a relatively small burn-in period and thinning parameter. Then, if convergence was not attained, the thinning parameter was increased and the procedure ran again, starting from the last iteration's values of the previous run. Essentially all iterations of the algorithm up to that point were considered to be burn-in. Following the recommendation of Gelman and Hill (2007), all parameters including the deviance were required to have an \hat{R} value less than 1.1 before considering the iterations to have converged. To guard against premature convergence, a test of chain slope was added, and it was required that no chain have a significant slope at a p-value $< 0.01/3p$, where p is the number of parameters. A linear model with autocorrelated errors was used because of the typical high serial correlation of chain values.

5.3.5 Determination of metric values

Recovery of nominal parameter values was assessed by first computing a binary metric indicating whether parameter recovery was obtained or not. The metric computed was determined from the percentile of the posterior distribution corresponding to the nominal parameter value. Percentile values between 2.5 and 97.5 were considered to be representative of cases where the nominal parameter value was recovered. Under optimal operating conditions, the nominal parameter value should be recovered 95% of the time, based on this metric.

The area under the ROC curve (AUC) is a normalized measure of predictive performance (Hanley, McNeil 1982). Determination of the AUC statistic was made in R using functions in the ROCR package (Sing et al. 2005). The functions require a vector of probabilities and a vector of categorical responses. To determine the overall AUC for the individual estimation method, the vectors of probabilities and responses from the test set predictions were concatenated and passed to the functions. AUC can vary between zero and one, where one indicates perfect prediction; however, a value of 0.5 corresponds to random guessing. Therefore, only values above 0.5 are acceptable.

The deviance information criterion, or DIC (Spiegelhalter et al. 2002) is a measure of predictive capability similar to the Akaike Information Criterion (AIC). It is the sum of a reward term based on the likelihood of the data given the model, and a penalty term based on the number of effective parameters. Like the AIC, smaller is better. DIC is returned by the OpenBUGS procedure. DIC was compared by summing the two DIC values from the individual estimations to compare with the single DIC value resulting from the joint estimation. This seemed reasonable given the definition of DIC and the fact that the joint model had exactly twice as many parameters as the individual model (in fact, the only real difference between the two was the number of processes being modeled).

5.3.6 Methods of evaluation

The three primary process characteristics of interest assessed in recovery analysis were the expected duration period $E[X]$, the expected renewal interval $E[Z]$, and the limiting probability for state “1” P_1 . The binary metric for recovery was then used as the dependent variable in a logistic regression with the process characteristics and observation protocol characteristics as regressors. This model allowed assessment of the predicted recovery rate associated with different regions of the experimental design space, and through analysis of the steepest gradient of the response surface (Myers, Montgomery 1995), the direction in which greater or lesser recovery changes at the fastest rate with respect to a unit vector of the regressors in the same direction.

AUC values for the individual and joint estimation results were evaluated for overall predictive capability and were compared using the Wilcoxon signed rank test. The effect of various experimental design-space characteristics on the AUC was evaluated using regression.

5.4 Results

Convergence was obtained on both individual estimations and on the joint estimation on 1999 of the 2000 scenarios. One trial converged on both individual estimations but failed on the joint estimation, despite many repeated attempts. For six of the 1999, the calculation of AUC failed because of problems with the test data set. That

left 1993 good scenarios. In the rest of this section recovery is assessed as a function of the experimental design space and is compared with respect to estimation method. Then a comparison of method is made based on AUC, followed by an assessment of process and observation protocol characteristic effect on AUC.

5.4.1 Recovery of nominal parameter values

There are three primary nominal attributes of interest: 1) $E[X]$ – the expected value of time spent in the violation state, 2) $E[Z]$ – the expected value of the renewal interval (time between re-entries to the violation state, or the inverse of frequency of occurrence), and 3) P_1 – the limiting probability for the non-violation state.

Recovery rates were nearly the same regardless of whether estimation was done individually or jointly (Table 20). Recovery rates were highest for $E[Z]$ and lowest for P_1 . $E[X]$ tended to be overestimated while $E[Z]$ tended to be underestimated (inferring that frequency was overestimated). This is consistent with the underestimation of P_1 as can be seen in Table 20. Since P_1 was underestimated, then P_0 was overestimated. This combination of estimation bias (duration, frequency and P_0 all overestimated) favors conservationists (over polluters).

Table 20: Counts of recovered nominal values given the uniform prior (combining A and B)

Group	method	below c.i. (nominal was overestimated)	Inside c.i.	above c.i. (nominal was underestimated)	Total	% inside c.i.
E[X]	individual	1231	580	182	1993	29.1
	joint	1226	586	181	1993	29.4
E[Z]	individual	303	883	807	1993	44.31
	joint	298	877	818	1993	44
P1	individual	1	113	1879	1993	5.67
	joint	1	112	1880	1993	5.62

To gain understanding of the relationship between recovery and process/observation characteristics, a basic response surface methodology (RSM) analysis (Myers, Montgomery 1995) was performed using logistic regression and a first-order model without interaction. Independent variables included the following (nominal values): E[X], P₁, E[Z], observation period, quasi-regular observation interval, and event observation probability. Variables were transformed for the analysis because they need to be on unbounded scales in order to make predictions at any point along the path of steepest ascent. The observation protocol properties were included to avoid masking effects. In a preliminary fit involving all six regressors plus estimation method (individual or joint), the terms for estimation method (p-value = 0.98267), event observation probability (p-value = 0.53002), and P₁ (p-value = 0.91355) were insignificant, so were dropped. The remaining four terms were all highly significant (Table 21).

Table 21: RSM regression results

```

> summary(glm.RSM)

Call:
glm(formula = In ~ log_E_X + log_E_Z + log_oP + log_rS, family =
"binomial",
    data = Results.All)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6910  -0.8004  -0.6099   1.0332   2.4704

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.26657    0.15023   15.09  <2e-16 ***
log_E_X      0.74022    0.03102   23.86  <2e-16 ***
log_E_Z     -0.39325    0.02716  -14.48  <2e-16 ***
log_oP      -0.68587    0.04410  -15.55  <2e-16 ***
log_rS      -0.33190    0.02328  -14.26  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13792  on 11957  degrees of freedom
Residual deviance: 12750  on 11953  degrees of freedom
AIC: 12760

Number of Fisher Scoring iterations: 4

```

Ignoring event observation probability and long-term probability P_1 , the mean vector of the transformed experimental design space is given in Table 22.

Table 22: Experimental design space mean vector

```

> mean.vector
      log_E_X  log_E_Z  log_oP  log_rS
1 -3.219765 -1.38963  2.75026 -0.955005

```

The probability of recovery associated with this vector was 0.2423817. The unit vector in the direction of steepest ascent is given in Table 23.

Table 23: Unit vector dimensions along steepest path gradient

```

> unit.vector
      log_E_X  log_E_Z  log_oP  log_rS
1  0.6534672 -0.3471581 -0.605487 -0.2929972

```

Movement from any point in the direction of this vector results in a higher probability of recovery. A predicted recovery rate of 0.95 was obtained at a distance of ~3.6 units from the mean vector in the positive direction of the steepest gradient (Figure 15). There were only three simulations where the predicted recovery rate was greater than 0.90 (all exceeded 0.95). All nine of the associated nominal parameters were recovered.

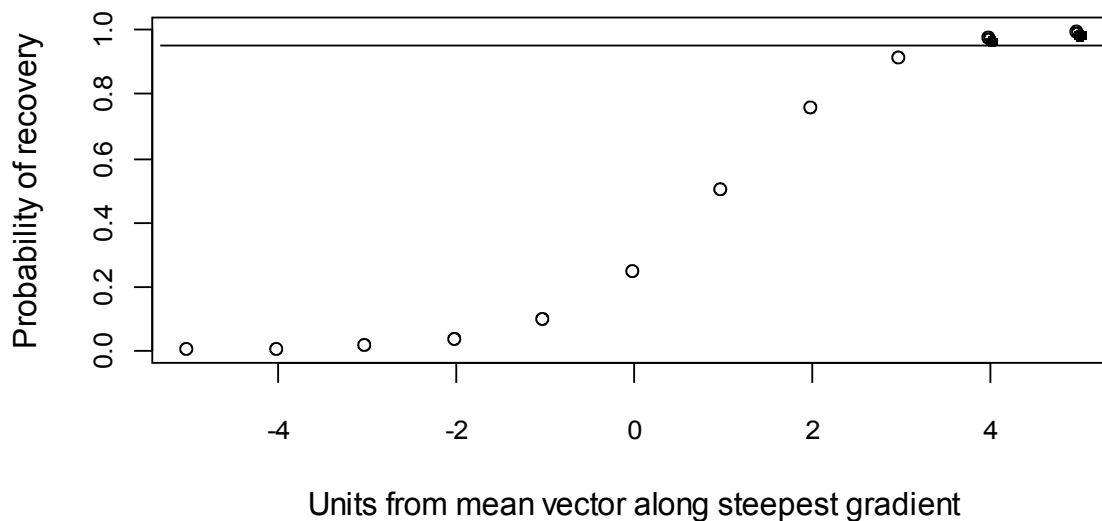


Figure 15: Predicted recovery rate along steepest path gradient

The experimental design space does not include a large proportion of the area where acceptable recovery rate would occur (Figure 16). The area of each plot where the predictive recovery rate would be greater than 0.95 lies to either side of the plotted solid points. The 0.95 contour is perpendicular to the plotted points between the last open circle and the first solid circle.

Very little of the experimental design space intersects with the over 0.95 recovery rate region except for $E[Z]$ and quasi-regular observation interval. Specifically it appears the experimental design space should be adjusted by increasing $E[X]$ and decreasing observation period, although this latter adjustment seems counterintuitive. The objective of a repeat experiment should be to explore the region on either side of the 0.95 contour, to better determine the region where acceptable recovery occurs. It is not really necessary that a large proportion of cases fall within the credible interval, but what is necessary is that the boundary between the experimental design spaces where recovery is acceptable on one side and not on the other be well-defined. This requires that the experimental design space be sufficiently large.

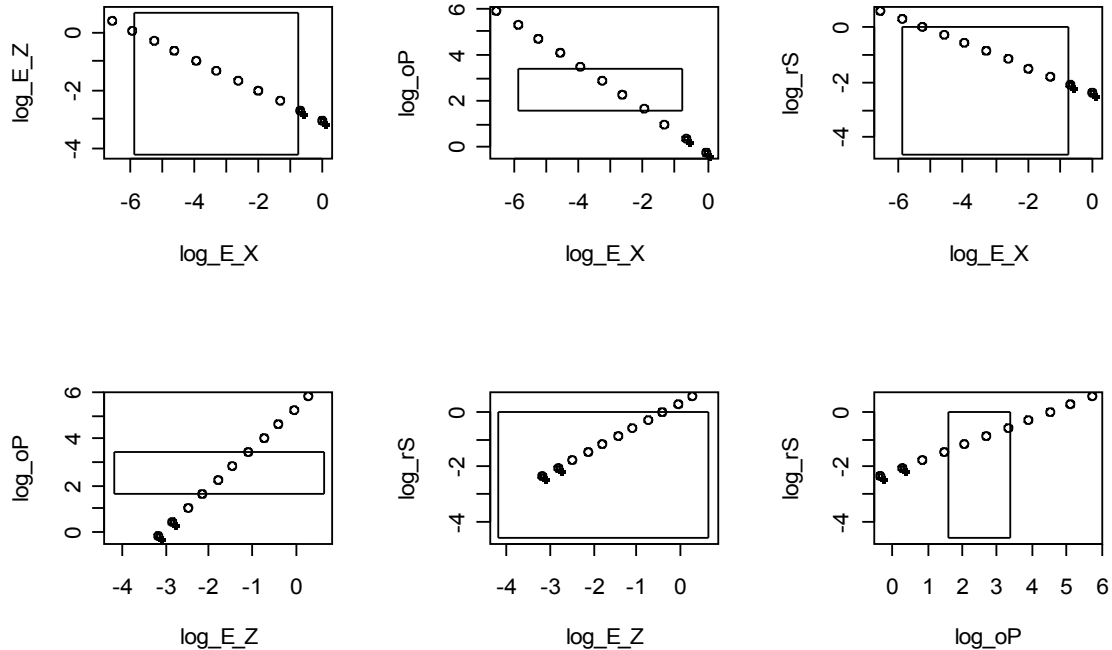


Figure 16: Steepest ascent gradient vs. process/observation characteristics

5.4.2 Overall performance

Prediction of test data was generally very good regardless of whether individual or joint estimation was used. In fact results were very nearly equal in both cases in terms of the distribution of AUC values. AUC was above 0.9 in 71% of cases and above 0.8 in 91% of cases regardless of whether individual or joint estimation was used (Table 24). The mean AUC under both methods was 0.92. The median under both methods was 0.94.

Table 24: Cumulative distribution of AUC values

Method	AUC					N
	>0.9	>0.8	>0.7	>0.6	>0.5	
Individual	0.71	0.91	0.96	0.98	1.00	1993
Joint	0.71	0.91	0.96	0.99	1.00	1993

5.4.3 Method comparison

A more-precise test of difference in AUC between methods was made using the Wilcoxon signed-rank test. The test was performed with the null hypothesis of no improvement. That is, the null hypothesis was that the AUC from joint estimation was less than or equal to the AUC from individual estimation. This hypothesis was rejected with a p-value = 0.007972. The 95% confidence interval for the median difference was (2.1e-05, ∞). The pseudomedian (Hollander, Wolfe 1973) difference was approximately 5.4e-05. Most of the results cluster tightly along the 1:1 line (indicating no difference between methods), with relatively few lying either above or below (Figure 17).

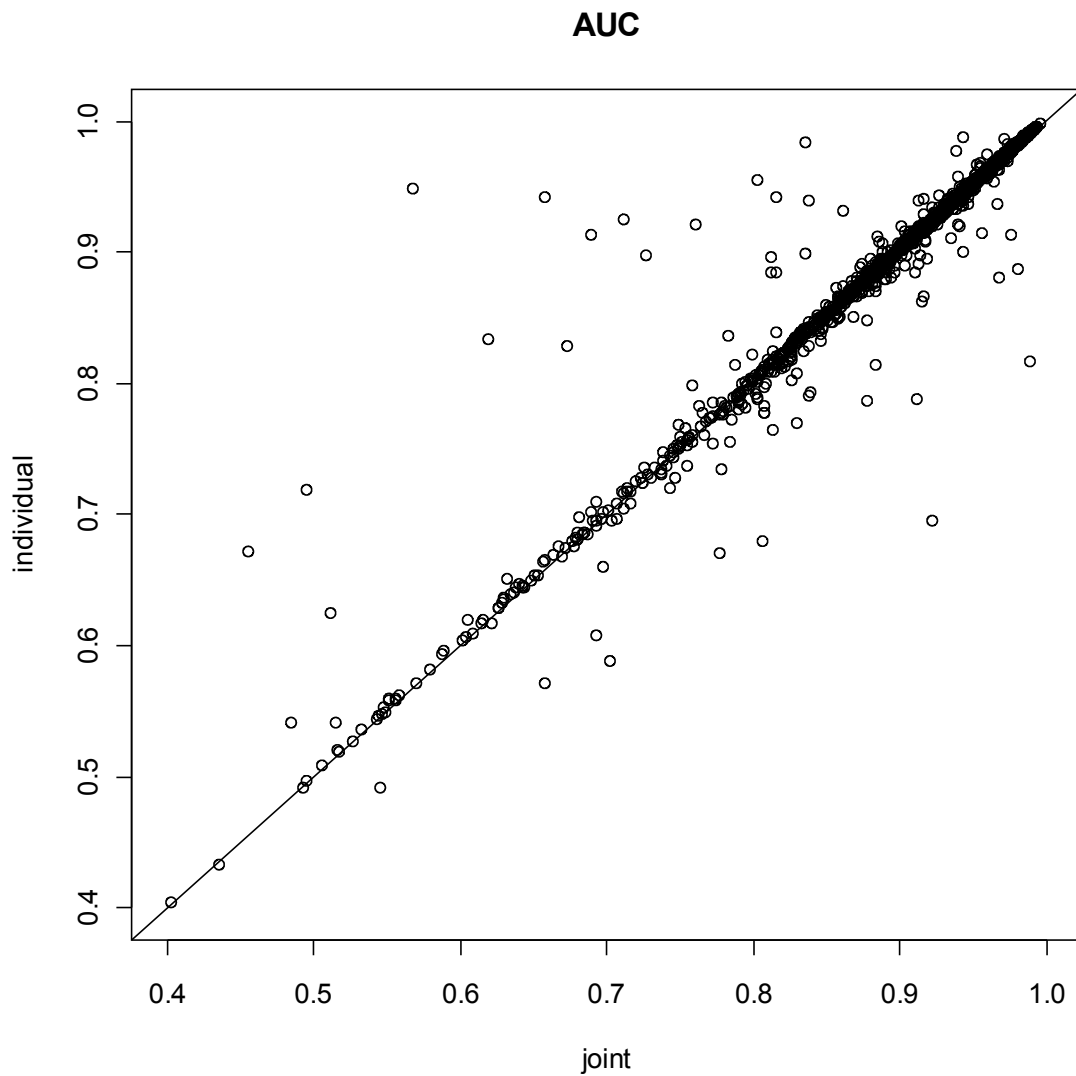


Figure 17: AUC from individual determination vs. joint determination

Most of the pairs of DIC values plot tightly along a 1:1 line (Figure 18); however, a substantial number plot along a 2:1 line. That is, the DIC for the joint estimation was

almost exactly twice that of the sum of the DICs of the individual estimations. It is suspected that this is a result of some underlying code error in WinBUGS.

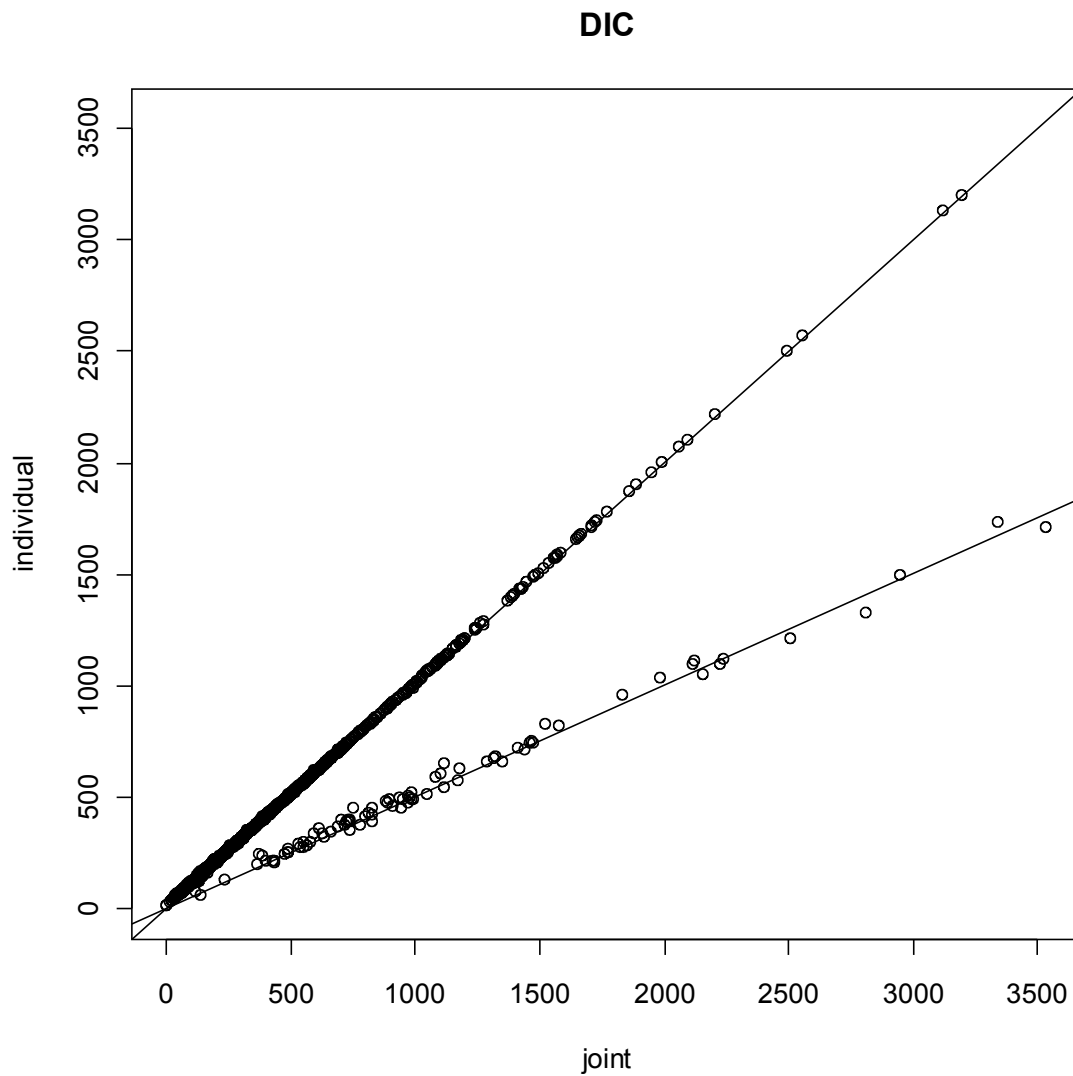


Figure 18: DIC from individual determination vs. joint determination

The *effective number of parameters*, or pD , is a measure of the amount of pooling taking place in the model estimation. Pooling in the joint model would provide some justification for using it. The computation of pD is unstable (Gelman, Hill 2007), but it is still interesting to compare the joint estimation pD to the sum of the pD of the two individual estimations (Figure 19). The instability can be seen in the number of cases with pD less than zero or greater than four, the actual number of parameters. Visually, there does not appear to be much difference between the two methods, however there is a statistically significant difference by the Wilcoxon signed-rank test ($p < 0.0001$) favoring the joint estimation method. This test was performed with the null hypothesis that the joint method was no better than the individual method.

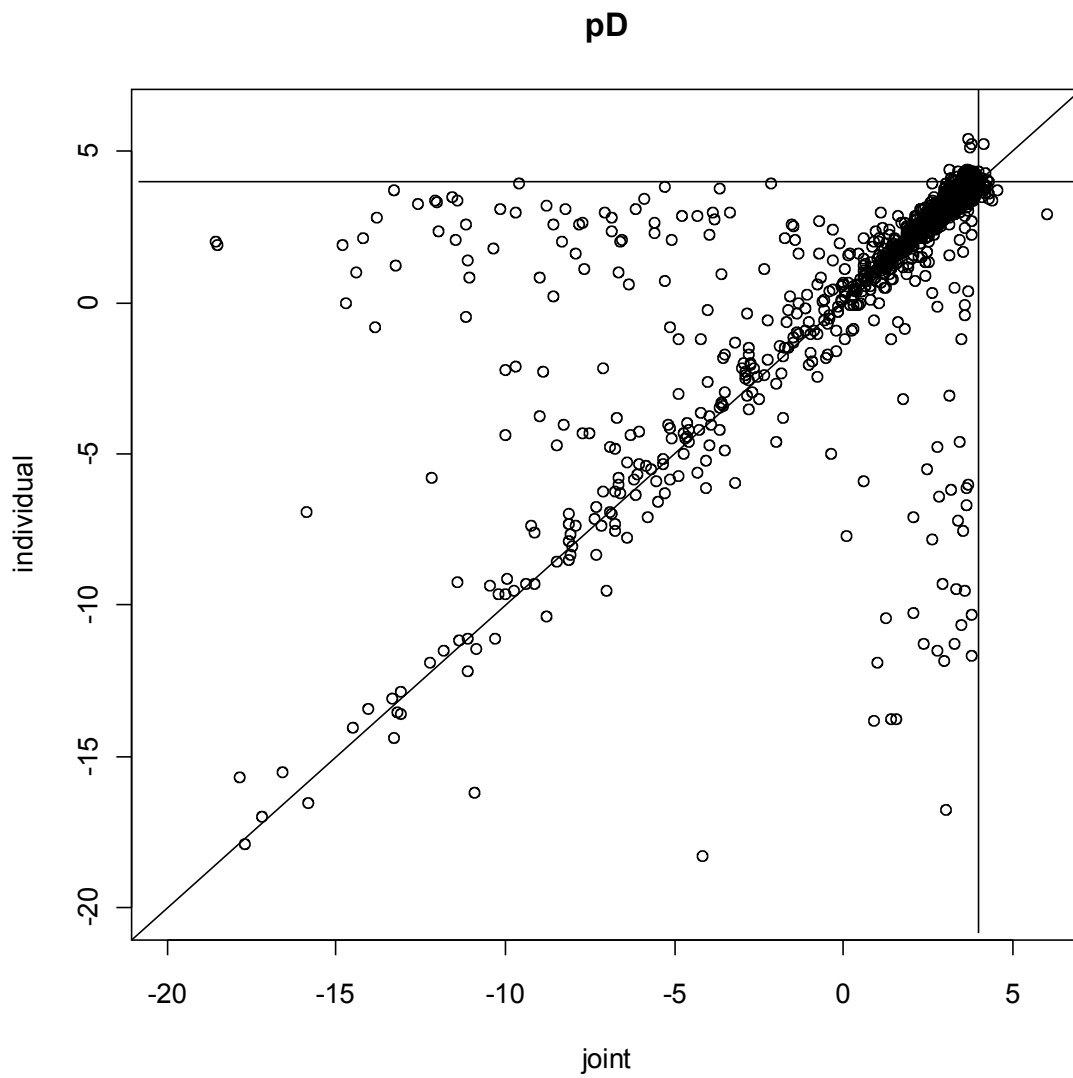


Figure 19: Effective number of parameters

5.4.4 Effect of simulation characteristics on AUC

Stepwise regression was used to determine the relationship between process and observation protocol characteristics (Table 25) and AUC. The process parameter values used in the regression analysis were the expected duration time in each state and the correlation between processes. The expected duration time is the inverse of the overall rate of transition out of the state, so this set of parameters describes the relative proportion of time spent in each state and the overall activity level within the system. The observation protocol parameters were the observation period, the two quasi-regular observation interval lengths, and the two violation-state observation probabilities. In order to mask the effect of the process names A and B, the variables were re-cast as minimums and maximums. For example the variable *min.EZ0011* indicates the minimum expected duration time between the two states (0,0) and (1,1).

Table 25: Variable definitions

Variable	Definition
oP	Observation period length
avg.rS	Average of the quasi-regular observation intervals
diff.rS	Absolute difference between the quasi-regular observation intervals
min.eP	Minimum event observation probability
max.eP	Maximum event observation probability
min.EZ0011	Minimum expected duration in state (0,0) or (1,1)
min.EZ0110	Maximum expected duration in state (0,0) or (1,1)
max.EZ0011	Minimum expected duration in state (1,0) or (0,1)
max.EZ0110	Maximum expected duration in state (1,0) or (0,1)
rho	Inter-process correlation

The logit transformation of the AUC was regressed against the characteristic values and against the estimation method as a categorical variable. In the stepwise regression, models were considered from all first-order effects up to a model including all second-order effects. The term for model selection was insignificant and was dropped from the model during the stepwise procedure. The final regression results are given in Table 26. All of the included second-order coefficients were statistically significant. All first-order coefficients were involved in one or more significant interactions. The model itself was significant.

Table 26: Regression results

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	9.885e-01	1.015e-01	9.735	< 2e-16	***
oP	-1.063e-02	4.276e-03	-2.485	0.012981	*
avg.rS	-1.455e+00	1.352e-01	-10.766	< 2e-16	***
diff.rS	9.655e-01	1.392e-01	6.938	4.62e-12	***
min.eP	2.465e+01	2.245e+00	10.983	< 2e-16	***
max.eP	8.323e+00	6.706e-01	12.412	< 2e-16	***
min.EZ0011	2.342e+00	1.932e+00	1.213	0.225363	
min.EZ0110	2.379e+01	2.925e+00	8.133	5.55e-16	***
max.EZ0011	9.802e-01	1.831e-01	5.352	9.18e-08	***
max.EZ0110	1.086e+01	5.152e-01	21.089	< 2e-16	***
rho	4.046e-02	9.651e-02	0.419	0.675050	
max.eP:max.EZ0110	-4.320e+01	1.870e+00	-23.108	< 2e-16	***
min.eP:max.eP	-1.488e+02	8.410e+00	-17.692	< 2e-16	***
min.eP:max.EZ0110	1.639e+02	1.535e+01	10.676	< 2e-16	***
max.EZ0011:rho	2.368e+00	2.470e-01	9.585	< 2e-16	***
avg.rS:min.eP	2.191e+01	3.277e+00	6.686	2.61e-11	***
min.EZ0110:max.EZ0110	-1.712e+02	1.636e+01	-10.464	< 2e-16	***
oP:avg.rS	4.312e-02	6.094e-03	7.076	1.75e-12	***
max.eP:min.EZ0110	1.214e+02	1.517e+01	8.000	1.62e-15	***
max.eP:min.EZ0011	1.018e+02	1.407e+01	7.238	5.44e-13	***
min.EZ0011:max.EZ0011	-5.368e+01	7.914e+00	-6.783	1.35e-11	***
avg.rS:max.eP	2.819e+00	7.040e-01	4.004	6.34e-05	***
min.eP:min.EZ0110	-3.140e+02	5.850e+01	-5.367	8.46e-08	***
min.EZ0011:rho	1.480e+01	3.912e+00	3.784	0.000157	***
oP:max.eP	-6.250e-02	1.797e-02	-3.478	0.000510	***
diff.rS:min.eP	-8.587e+00	2.480e+00	-3.462	0.000542	***
avg.rS:diff.rS	-8.115e-01	2.569e-01	-3.159	0.001594	**
oP:rho	1.794e-02	4.187e-03	4.285	1.87e-05	***
oP:min.EZ0110	3.792e-01	1.255e-01	3.021	0.002538	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5266 on 3957 degrees of freedom

Multiple R-squared: 0.7222, Adjusted R-squared: 0.7202

F-statistic: 367.4 on 28 and 3957 DF, p-value: < 2.2e-16

The logit transformation of AUC poses some difficulties for interpretation of the coefficient values. For example, the effect of a unit change in the minimum event observation probability $min.eP$ (the maximum change possible) on AUC depends on the levels of the other variables. As usual, coefficient magnitude is a function of scale and so

it also can be problematic to compare magnitudes directly. A common method is to evaluate the effect in the neighborhood of the means of the other variables. In this case several of the variables are constrained by the level of other variables, however the point is to establish some constant value near which to assess effect size.

Visualization of the effects of pairs of characteristics involved in second-order effects was accomplished through five sets of contour plots (Figures 20 through 24). Plots were grouped according to whether the two characteristics were 1) both observation protocol, 2) both process, or 3) one of each. In most cases the plots are self-explanatory and need little explanation, so focus is on overall interpretation. Contours were generated using a grid of points which in a few cases included points that were outside the range actually found in the experiment. Significant portions of most plots predict AUCs in excess of 0.90.

The plots in Figure 20 indicate that higher AUC results from higher event observation probability, shorter quasi-regular observation intervals, and longer observation period. The interaction between average quasi-regular observation interval and event observation probability (both min and max), is minimal, but does indicate the the effect of event observation probability is greater when quasi-regular observations are more spread out. Subplot [c] indicates that as observation period length increases, the effect of quasi-regular observation interval decreases.

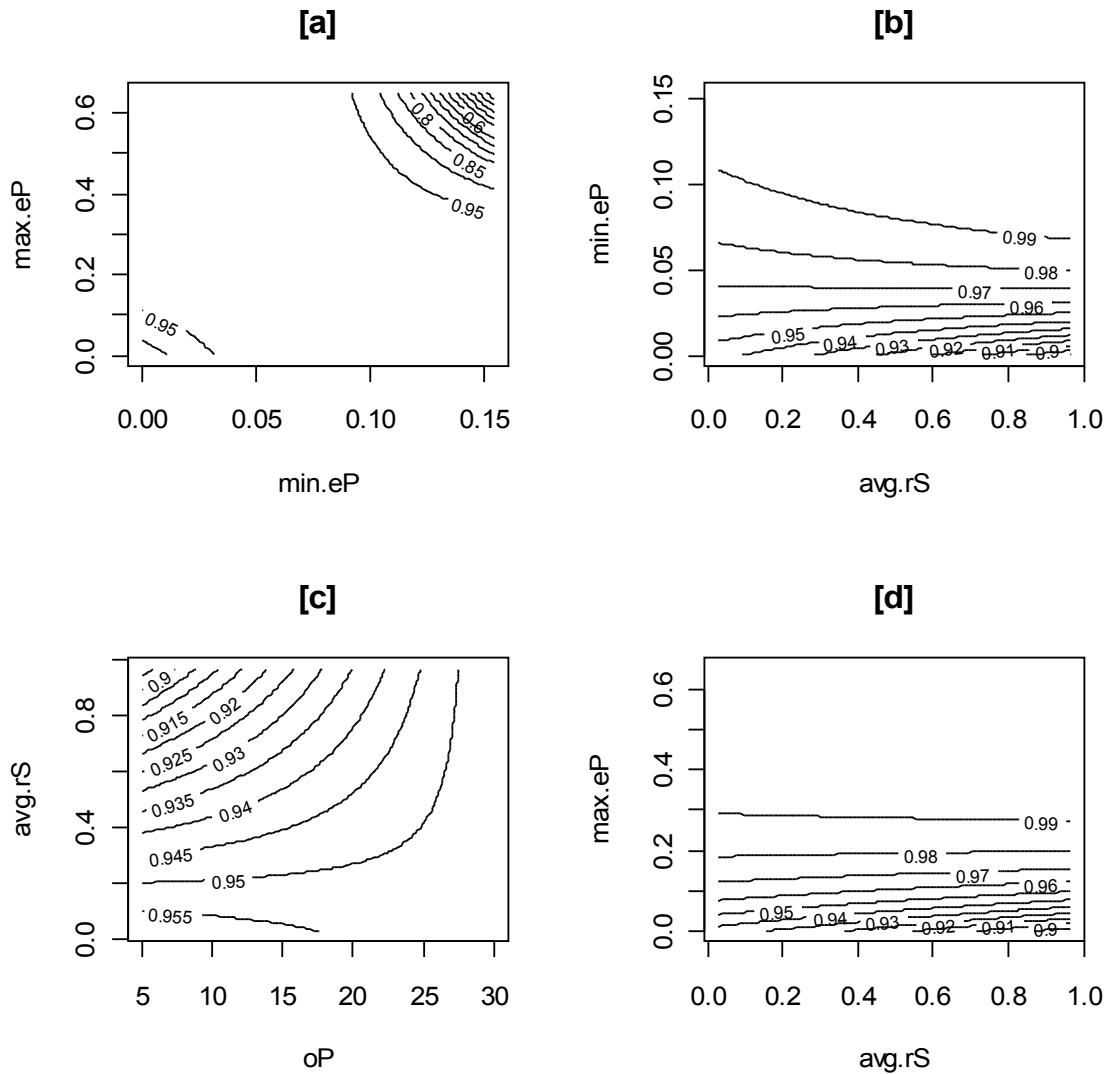


Figure 20: Observation characteristic interaction effects contours

Figure 21 also shows two interaction pairs where the interaction effect is slight. Subplot [a] indicates that as observation period length increases, the effect of the maximum event observation probability decreases. Subplot [b] requires some explanation. The *diff.rS* characteristic can be 0 for a wide variety of pairs of values for

the quasi-regular observation intervals, but can only be 1 in two cases, and in those two cases one of the intervals will be very short. In such a case the effect of event observation can be expected to be reduced, which is what is indicated in the plot. Subplot [c] also requires some explanation, as the two upper corners of the plots are not feasible. However it makes sense that higher AUC would result when the average is low and the difference high.

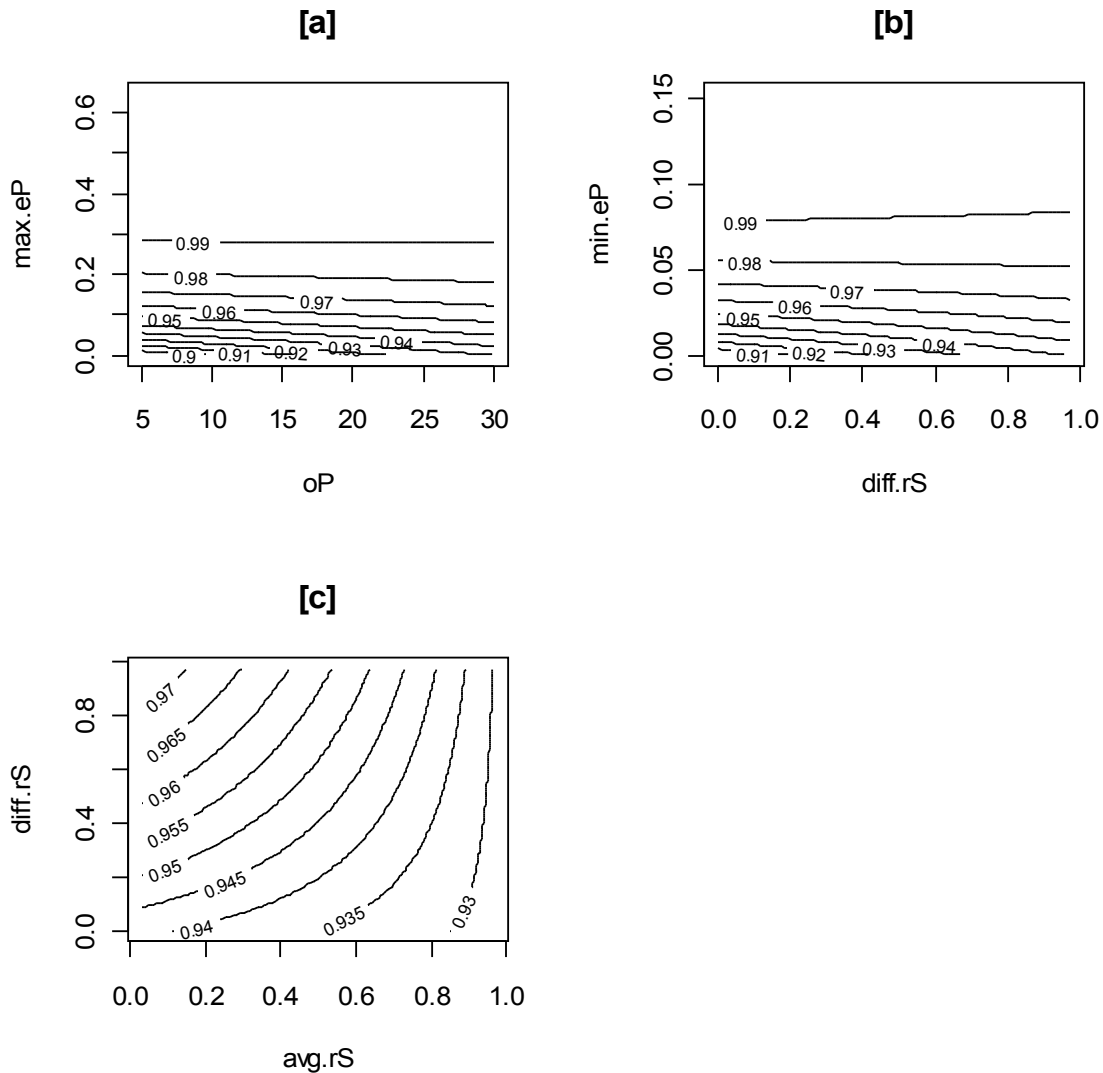


Figure 21: Observation characteristic interaction contour, continued

The next two figures (Figure 22 and Figure 23) contain plots where the x-axis is an observation characteristic and the y-axis is a process characteristic. In Figure 22 and subplot [a] of Figure 23, the y-axes correspond to expected duration times in the states. High values on these axes correspond to low transition rates. The x-axes are all event

observation probabilities. All of these plots essentially indicate the same thing, that high AUC results from relatively low transition rates, relatively high event observation probabilities, and that the interaction of these effects is positive. Conversely, high transition rates and low event observation probability would lead to low AUC.

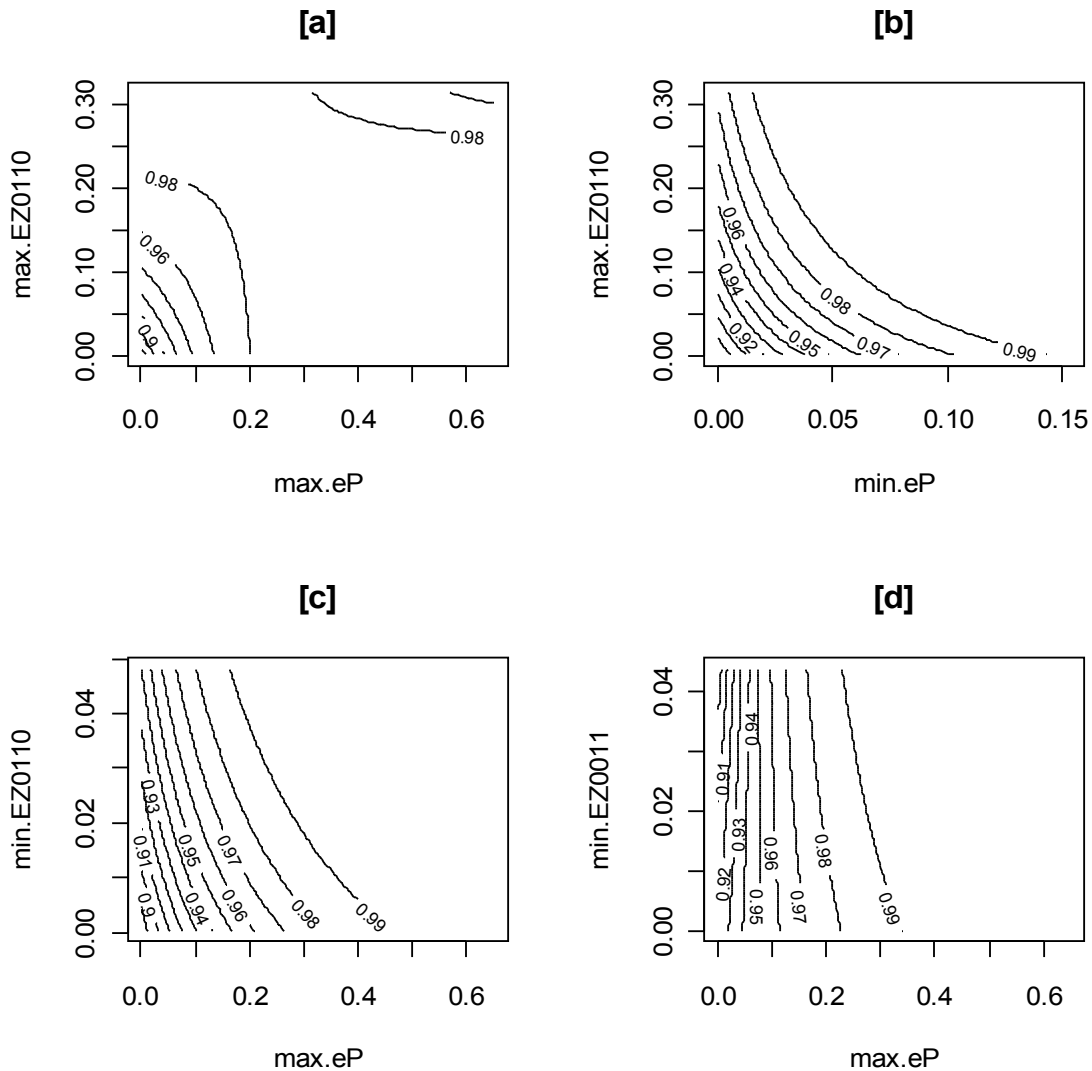


Figure 22: Observation/process characteristic mixture interaction contours

Figure 23 ([b] and [c]) show interactions involving observation period length. Both show increasing AUC with increasing observation period length, as long as inter-process correlation is above about -0.8. Subplot [b] indicates that 1) AUC increases with positive increase in interprocess-correlation and that there is positive interaction between

the two. Subplot [c] indicates, as before, that AUC is higher with lower transition rates out of the states (in this case the (0,1) and (1,0) states) and the interaction between this effect and observation period length is positive.

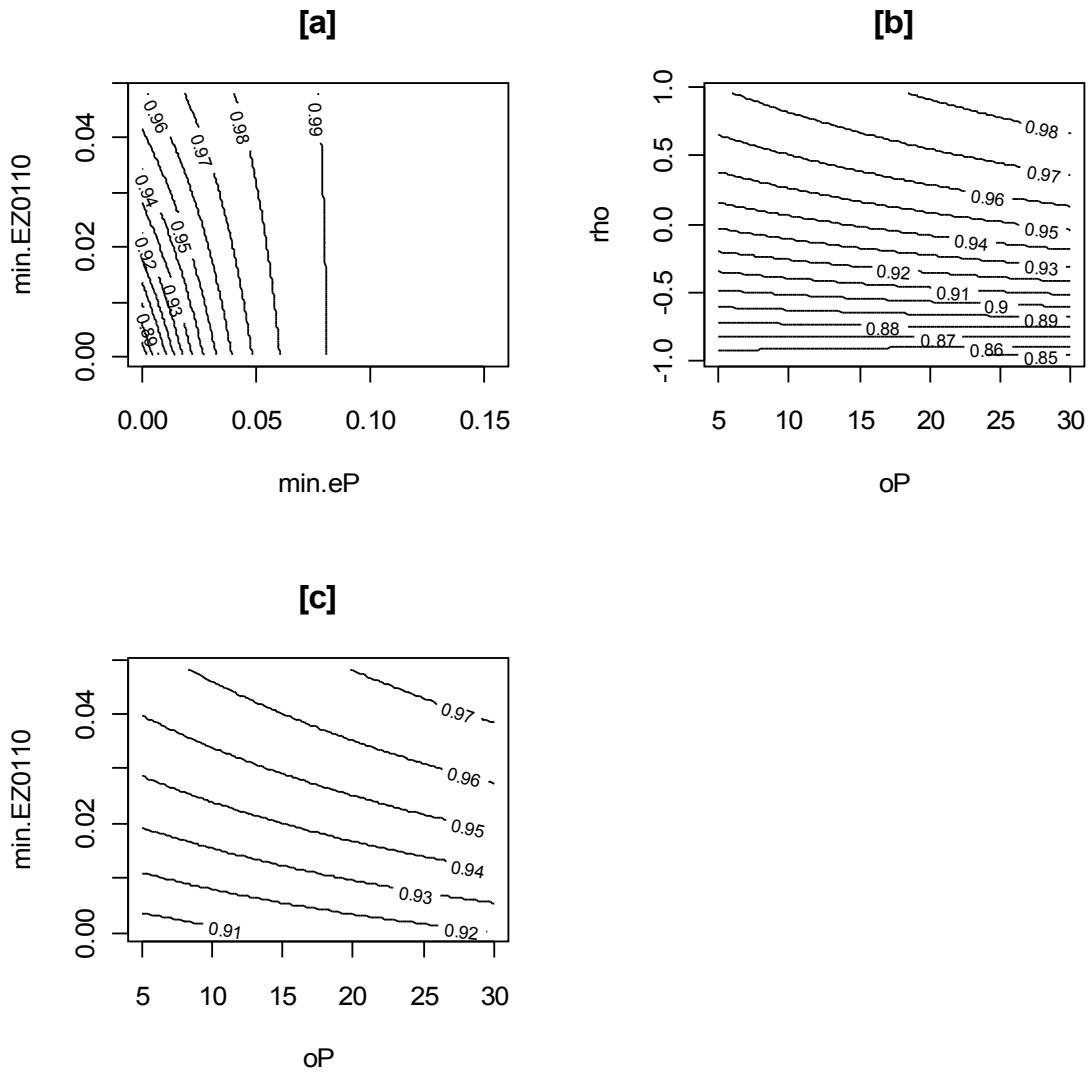


Figure 23: Observation/process characteristic mixture interaction contours, continued

Figure 24 contains interaction pairs where both characteristics were process characteristics. Interestingly there were fewer of these than there were of the observation protocol characteristics-only figures. Subplots [a] and [d] indicate that the effect of inter-process correlation was greatest when transition rates were lowest. The highest AUCs were predicted when transition rates were low and interprocess correlation positive. Subplot [b] can be misleading because of the difference in scale. Predicted AUC is actually more responsive to the x-axis variable *min.EZ0110*, although the contours are farther apart there. However, the highest predicted AUCs were obtained when the transition rates along both axes were low. Subplot [c] requires some explanation. In the actual data there were very few cases appearing in the upper right and lower right regions of the plot. Otherwise it would be expected that the highest predicted AUC would be in the upper right hand corner of the plot, which is not the case. The other thing that is not obvious in this plot is that the *min.EZ0011* value is usually that corresponding to the (0,0) state. Low transitions out of this state correspond roughly to increased numbers of event observations. In this case there may be some third-order effects which should be examined.

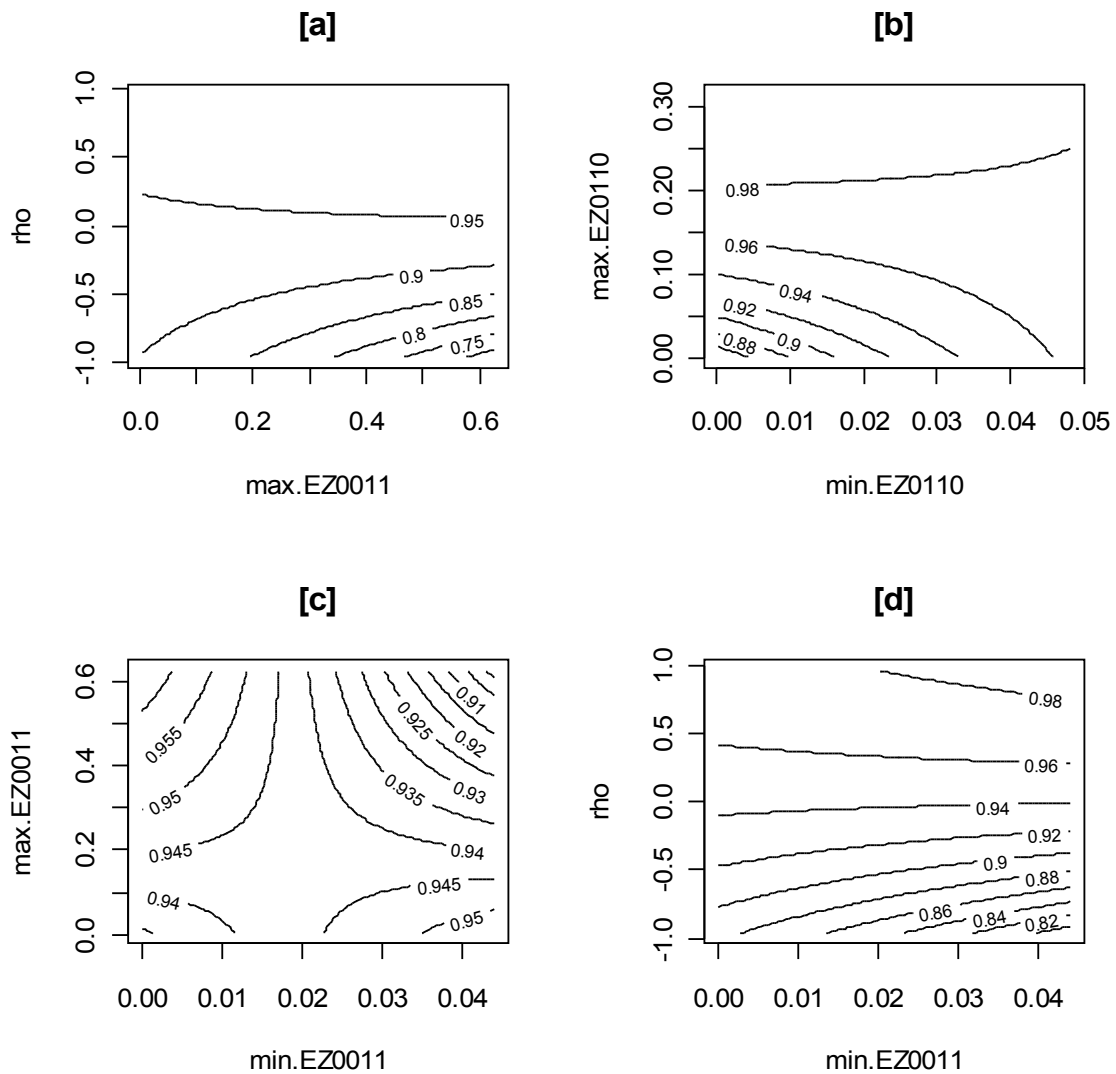


Figure 24: Process characteristics-only interaction effects contours

None of these results are non-intuitive or unexpected. The moderate value of the adjusted R^2 term indicates there is still some unexplained variance in the dependent variable despite the highly significant component coefficients.

5.5 Discussion

In this paper individual and joint estimation of models for dual process CTMCs was evaluated using simulations of infrequent, unevenly spaced, and uncoordinated observations. The evaluation consisted of assessments of nominal parameter recovery, of out-of-sample prediction performance, and of the effect of various process and observation characteristics on said performance. The purpose of these assessments was to compare the two estimation methods and to establish the conditions under which they could be applied, assuming that successful application would be a function of both process characteristics and of observation protocol characteristics.

In all of the comparisons between estimation methods, either no difference or a small but statistically significant difference was found between the two estimation methods. These significant differences were always in favor of the joint method. On the other hand, the joint method failed to converge in a few cases. However, choice of estimation method was clearly not a decisive factor with regards to quality of results. For a large number of processes a single joint estimation would seem to be logistically preferable to a large number of individual estimations, and at least these results do not argue against making that choice. In the experiment, each OpenBUGS run, whether individual or joint, was initialized with random values. In a real-world application, the results of individual estimates could be used as starting values for the joint method.

Both estimation methods were found to produce biased estimates of the three primary nominal parameters over the experimental design space considered. This space

was specified to represent process characteristics and observation protocol characteristics thought to be common in the water quality monitoring field. Under these assumptions the method can be expected to overestimate violation state duration, frequency, and long-term probability of occurrence. This bias favors the conservationist. It appears this bias could be reduced; however, additional experiments should be conducted to determine the conditions under which good recovery of true process parameter values could be expected. These results provide some guidance as to what these conditions might be and how additional experiments could be carried out.

Even given the biased recovery results, prediction of out-of-sample observations was excellent ($AUC > 0.9$). This suggests that parameter values for the process as observed were estimated accurately, otherwise the predictions should have been faulty. This in turn suggests that the processes as observed differed from the underlying processes in such a way as to produce estimates systematically biased with respect to the nominal values. Observation of the processes at regular intervals would not be expected to produce the pattern of over and underestimation seen. Therefore it seems likely that the source of the bias was the violation-state observation protocol. This suggests that recovery might be improved by a modification of that part of the observation protocol.

In this paper, dual correlated processes were simulated using a four-state model. Additional processes could be simulated using this approach, although the number of states necessary grows exponentially with the number of desired processes. In addition, it is suspected that finding correlated examples using the method used here would require

the generation of even greater numbers of test cases. Also, the pairwise correlations between processes would not necessarily be of the same sign and approximate magnitude, a complicating factor for analysis.

Although negatively correlated processes were included for completeness, WQIs in general are positively correlated, as they tend to be influenced similarly by regional-scale processes such as geologic weathering, atmospheric deposition, and climate. Since positively correlated processes resulted in higher AUCs, this bodes well for real-world WQI applications.

In general making synchronous or coordinated observations across a network of locations is impractical. Typically observers visit locations in sequence across a span of time. In trend-monitoring programs, assigning an observation to a day, month, or quarter, regardless of when within the period the observation was actually made, results in a dataset of pseudo-synchronous observations. This is done usually to meet the assumptions of the trend test model. The method developed here requires no such pseudo-alignment of observation times.

The model fit to the processes was not the same as the model used to generate the process observations. Despite that, predictive performance was more than acceptable, and was better when the processes were positively correlated or had relatively low transition rates. The four-state process could have been modeled precisely, but generalizing to more than two processes is complicated. Without a model or a practical way to make observations that could support inter-process correlation estimates, it seems

that pursuance of a hierarchical modeling approach is warranted. Behavior of WQIs at neighboring locations is often influenced by broader regional characteristics such as underlying geology or climate. The state of one process may be predictable from neighboring processes, but the overall probability of observation ensures few will be synchronous.

5.6 Conclusions

The objective of this paper was to evaluate individual and joint estimation of models for dual process CTMCs under simulated observation conditions. A method for generating simulations of dual correlated two-state CTMC processes using non-homogeneous Markov Chains was demonstrated. The method may be extended to more than two processes, although the effort grows exponentially with the number of processes. The simulations were used to mimic the behavior of WQIs with respect to a threshold that divides the range of the indicator into two states, one of which is associated with ecological risk and is referred to here as the violation state. Observation protocol characteristics were simulated to mimic common protocols in the water-quality monitoring field that result in observations which are asynchronous, infrequent, unevenly spaced, and uncoordinated with threshold crossings.

A joint model for multiple two-state CTMC processes was introduced and applied to simulated observations of a dual correlated two-state CTMC. The parameters of the model were estimated both jointly and individually using Bayesian estimation methods.

Both estimation methods predicted out-of-sample observations well, despite the fact that the model used in the estimation was simpler than the model used to generate the simulated observations. The joint estimation method showed a meager but statistically significant improvement over the individual estimation method.

In general, violation-state duration, frequency and long-term probability of occurrence were overestimated over most of the range of processes and observation protocols considered. This favors the conservationist with regard to risk. Processes and observations were simulated to represent typical problematic conditions in WQIs and typical observation protocols in the water quality field, which combine quasi-regular observation with event-based observation. Additional experiments are needed to better define conditions under which unbiased recovery could be expected.

Positively correlated processes or processes with lower transition rates resulted in better predictive performance than their counterparts. Within the range considered, all observation protocol characteristics were seen to have a significant effect on predictive performance. AUC responded positively to increases in observation period length and event observation probability, and to decreases in quasi-regular observation interval length. There were significant interaction effects between many pairs of characteristics.

6 CTMC models for threshold exceedance in water-quality indicators

6.1 Abstract

Thresholds have been defined for many water quality indicators (WQIs) which separate the measurement space of the indicator into two states, one of which, the exceedance or violation state, has undesirable consequences. Observations of the indicator are often made at relatively infrequent, unevenly spaced intervals and are almost always uncoordinated with the precise timing of changes in state. Observations of the indicator made at multiple locations are almost always asynchronous. These typical observation protocols make estimation of frequency, duration, and long-term probability properties difficult. To address this problem, three hierarchical CTMC models, one homogeneous and two non-homogeneous, were developed and evaluated with respect to each other and with respect to a method adapted from the partial duration series (PDS) method popular in flood-frequency analysis. Bayesian methods were used to estimate the parameters of the CTMC models. Simulations of WQI time series were generated using a sinusoidal model with autocorrelated errors adapted from the literature. These

simulations were sub-sampled to produce time series of asynchronous, uncoordinated and unevenly spaced observations similar to those found in practice. A response surface analysis was conducted to assess nominal process model parameter recovery in the CTMC models. A principal components regression analysis was employed to evaluate the effects of process, observation, and modeling on various performance metrics. Recovery was insufficient overall to establish optimal operating conditions with certainty. Duration and long-term probability of violation tended to be overestimated whereas frequency tended to be underestimated. The homogeneous CTMC model produced estimates of frequency and duration with lower error than did the estimates produced using the PDS method. Results were mixed for the two non-homogeneous CTMC models. More-positively correlated processes were easier to predict. Higher observation rates, regardless of type, were found to improve predictive performance. The CTMC models may be extended to allow for prediction of frequency and duration properties from watershed characteristics or altered to allow these properties to vary with time.

6.2 Introduction

Regulators have long used thresholds to define water-quality standards. Typically standards require that WQI concentrations remain on one side of the threshold (above or below) at least a specified percentage of time during some period such as a year or a defined season. Serious harm can come to aquatic ecosystems when water-quality thresholds are violated too often or for excessively long periods of time (Baldigo,

Murdoch 1997, Bulger, Cosby & Webb 2000, Davies et al. 1992, DeWalle, Swistock & Sharpe 1995, Laio et al. 2001, Sickle et al. 1996) . Recently ecologists have begun to look at thresholds as indicators of climate change (Lenton et al. 2008, Rosenzweig et al. 2008, Keller, Yohe & Schlesinger 2008). Changes in the stochastic properties of a process with respect to a threshold could indicate change in climatic conditions. However, for many WQIs the high-frequency observations necessary to identify and to time periods of violation are not available. In some cases only a subset of locations within a region of interest may have observations. Although two WQI processes might be correlated, observations typically are not made simultaneously on both processes.

Eventually it will be desirable to be able to model many processes at one time because often processes within a region (a population of processes) move together roughly as a group or set of groups, and it will be desirable to be able to take advantage of common characteristics among processes and to make predictions for processes within a region for which there is little or no water-quality data, perhaps using other characteristics common to all member processes in the region other than water quality itself. However, devising an all-inclusive experimental framework for such a problem is difficult.

The objective of this paper was to evaluate the performance of three hierarchical CTMC models for threshold exceedance behavior and associated Bayesian estimation methodology, and to compare them with a method (the partial duration series method) adopted from the flood-frequency analysis domain and similar to one that has been used

by other researchers (Madsen, Rasmussen & Rosbjerg 1997, Wang 1991, Deviney, Rice & Hornberger 2006). Three two-state CTMC models were developed that could be extended to include additional group-level predictor variables. The models allow estimation of three important process properties: 1) expected violation state duration time, 2) expected renewal period (the inverse of frequency), and 3) long-term proportion of time spent in the violation state. The models also allow the estimation of these properties from asynchronous, infrequent, uncoordinated and unevenly spaced observations of a process. The first model is homogeneous while the second and third are non-homogeneous. The latter two models allow transition rates to vary periodically, which may be necessary for locations where transition rates between states vary seasonally. All three models allow estimation of properties for multiple locations, or processes. Performance of these models was compared with a method adapted from the flood-frequency literature.

The performance evaluation consisted of assessments of nominal parameter recovery, of out-of-sample prediction performance, and of the effect of various process and observation characteristics on said performance. The purpose of these assessments was to establish the conditions under which the methodology could be applied, assuming that successful application would be a function of both process characteristics and of observation protocol characteristics.

In the next sub-section the partial duration series method is described. Following that three CTMC models are developed, starting with a simple two-state homogeneous model and concluding with two non-homogeneous models.

6.2.1 Partial duration series method

The partial duration series (PDS) method, sometimes referred to as the peaks over threshold (POT) method, is a flood-frequency method that assumes that onsets of exceedances of a high threshold by river discharge occur infrequently as a Poisson process. That is, the number of onsets per unit time, often one calendar or water year, is Poisson distributed. This is equivalent to assuming an exponential distribution of inter-arrival times. The distribution of the maximum discharge peak above threshold during independent exceedance events is currently thought to be generalized Pareto (GP), although historically other distributions have been suggested and investigated. The GP distribution for discharge (Madsen, Rasmussen & Rosbjerg 1997, Madsen, Rosbjerg 1997) has one of two possible cumulative distribution functions (CDFs):

$$\begin{aligned}
 F(q) &= 1 - \exp\left(\frac{-q - q_0}{\alpha}\right), \kappa = 0 \\
 F(q) &= 1 - \left(1 - \kappa \frac{q - q_0}{\alpha}\right)^{1/\kappa}, \kappa \neq 0
 \end{aligned}
 \tag{48}$$

where α is the scale parameter, κ is the shape parameter, and q_0 is the discharge threshold level. If κ is negative then the distribution is bounded above. Several methods have been

suggested for estimating the GP distribution parameters. The Method of Moments (Hosking, Wallis 1987) (MOM) estimators are

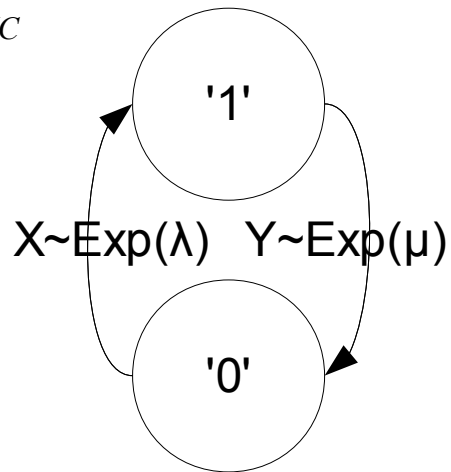
$$\begin{aligned}\hat{\alpha} &= \frac{1}{2} \hat{\mu} \left(\frac{\hat{\mu}^2}{\hat{\sigma}^2} + 1 \right) \\ \hat{\kappa} &= \frac{1}{2} \left(\frac{\hat{\mu}^2}{\hat{\sigma}^2} - 1 \right)\end{aligned}\tag{49}$$

Given estimates of α and κ and the derivatives of Equation (48), a distribution function for duration can be determined.

6.2.2 CTMC Model #1 (homogeneous)

A two-state CTMC is a reasonable first choice for a model of threshold exceedance. Let one state (say state '0') represent the below-threshold condition and the other state (state '1') represent the above-threshold condition. Let state '0' represent the violation state. Assume that upon entering one state, a random amount of time passes before the process transits to the other state. Once in the other state, a random amount of time passes before the process returns to the original state. Assume these random times have exponential distributions with rates λ and μ that are not necessarily equal. Such a process is depicted in Figure 25.

Figure 25: Two-state CTMC



Some important properties of the process in Figure 25 are given in Equation (50), where $E[X]$ represents the expectation of the amount of time spent in state '0' on a typical excursion into that state, $E[Y]$ represents the same expectation for state '1', $E[Z]$ is the expectation of the time between entrances into either state, P_0 is the long-term proportion of time spent in state '0', and P_1 is the long-term proportion of time spent in state '1'.

$$\begin{aligned}
 E[X] &= \frac{1}{\lambda} \\
 E[Y] &= \frac{1}{\mu} \\
 E[Z] &= E[X] + E[Y] \\
 P_0 &= \frac{\mu}{\lambda + \mu} \\
 P_1 &= \frac{\lambda}{\lambda + \mu}
 \end{aligned} \tag{50}$$

Ross (1993) showed how the two-state homogeneous CTMC model and the Kolmogorov Backward Equation lead to equations that predict the probability of being in state “0” having been in state “1” or having been in state “0” some time previously. Ross gives these equations respectively as:

$$\begin{aligned} P_{00}(t) &= \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)\Delta t} + \frac{\mu}{\lambda + \mu} \\ P_{10}(t) &= \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)\Delta t} \end{aligned} \quad (51)$$

With some re-arrangement of terms, a Bayesian specification for multiple sites following Equation (51) can be written as

$$\begin{aligned} \text{Prob}(S_{j,n} = '1') &\sim \text{Bernoulli}(P_{j,1} F_{j,n} + (1 - F_{j,n}) \text{Prob}(S_{j,n-1} = '1')) \\ \text{Prob}(S_{j,n-1} = '1') &= \begin{cases} 1, & S_{j,n-1} = '1' \\ 0, & S_{j,n-1} = '0' \end{cases} \\ F_{j,n} &= 1 - e^{-(\lambda_j + \mu_j)\Delta t_n} \\ \lambda_j &= C_j P_{j,1} \\ \mu_j &= C_j (1 - P_{j,1}) \\ \text{logit}(P_{j,1}) &\sim \text{Normal}(\hat{\beta}_j, \tau_\beta) \\ \hat{\beta}_j &\sim \text{Normal}(\mu_\beta, \tau_{\mu_\beta}) \\ \log(C_j) &\sim \text{Normal}(\hat{\xi}_j, \tau_\xi) \\ \hat{\xi}_j &\sim \text{Normal}(\mu_\xi, \tau_{\mu_\xi}) \\ \tau_\beta &\sim \text{Gamma}(0.5, 0.5) \\ \tau_\xi &\sim \text{Gamma}(0.5, 0.5) \\ \tau_{\mu_\beta} &\sim \text{Gamma}(0.5, 0.5) \\ \tau_{\mu_\xi} &\sim \text{Gamma}(0.5, 0.5) \\ \mu_\beta &\sim \text{Normal}(0, 0.0001) \\ \mu_\xi &\sim \text{Normal}(0, 0.0001) \\ \Delta t_n &= t_n - t_{n-1} \end{aligned} \quad (52)$$

where λ_j and μ_j are the transition rates from states '0' to '1' and '1' to '0' respectively for

process j and $C_j = \lambda_j + \mu_j$. By specifying the logit transformation of the limiting probability of process j being in state '1' ($P_{j,1}$) as having a normal distribution, the capability to add regressors to predict its value is facilitated. Similarly for the constant C_j , a log transformation modeled as having a normal distribution facilitates a similar capability. Since the range of possible predictors is virtually unlimited, that feature is not explored here. The model simply allows for different intercept parameters for each process. The shape and inverse-rate parameters that are set here to 0.5 lead to reliable convergence when there are two processes but would need to be reduced for more than two processes. Another approach is to make the shape and inverse-rate hyper-parameters with prior distributions of their own, say for example a uniform prior over the range [0.001, 0.5]. That was not found to be necessary for this paper.

6.2.3 CTMC Model #2 (non-homogeneous)

Model #2 lets the transition rates evolve periodically as sinusoids while holding the ratio between them (the limiting probabilities) constant. This corresponds to a process where the limiting probabilities are invariant with time, but the expected rates of transition from either state to the other are higher or lower depending on the time during the period. The Bayes specification for Model #2 follows:

$$\begin{aligned}
\text{Prob}(S_{j,n} = '1') &\sim \text{Bernoulli}(P_{j,1} F_{j,n} + (1 - F_{j,n}) \text{Prob}(S_{j,n-1} = '1')) \\
\text{Prob}(S_{j,n-1} = '1') &= \begin{cases} 1, & S_{j,n-1} = '1' \\ 0, & S_{j,n-1} = '0' \end{cases} \\
F_{j,n} &= 1 - e^{-(1+K_j)\lambda_{j,n}\Delta t_n} \\
\lambda_{j,n} &= \lambda_j + A_j \sin(2\pi t_n) + B_j \cos(2\pi t_n) \\
\lambda_j &= C_j P_{j,1} \\
K_j &= (1 - P_{j,1}) / P_{j,1} \\
\text{logit}(P_{j,1}) &\sim \text{Normal}(\hat{\beta}_j, \tau_\beta) \\
\hat{\beta}_j &\sim \text{Normal}(\mu_\beta, \tau_{\mu_\beta}) \\
\log(C_j) &\sim \text{Normal}(\hat{\xi}_j, \tau_\xi) \\
\hat{\xi}_j &\sim \text{Normal}(\mu_\xi, \tau_{\mu_\xi}) \\
\tau_\beta &\sim \text{Gamma}(0.5, 0.5) \\
r_j &\sim \text{Uniform}(0, 1) \\
A_j &= r_j \lambda_j \cos(\omega) \\
B_j &= r_j \lambda_j \sin(\omega) \\
\tau_\xi &\sim \text{Gamma}(0.5, 0.5) \\
\tau_{\mu_\beta} &\sim \text{Gamma}(0.5, 0.5) \\
\tau_{\mu_\xi} &\sim \text{Gamma}(0.5, 0.5) \\
\mu_\beta &\sim \text{Normal}(0, 0.0001) \\
\mu_\xi &\sim \text{Normal}(0, 0.0001) \\
\omega &\sim \text{Uniform}(-\pi, \pi) \\
\Delta t_n &= t_n - t_{n-1}
\end{aligned} \tag{53}$$

A number of other approaches for modeling the two sinusoidal amplitude variables A_j and B_j were tried but none were found to converge as reliably as the one specified above.

6.2.4 CTMC Model #3 (non-homogeneous)

In the second modification, the renewal rate is held constant but the limiting probabilities P_0 and P_1 are allowed to evolve periodically as sinusoids. This is a

process where the distribution of times between events is invariant with time but the instantaneous transition rates vary, with an increase in transition rate from one state being balanced in some way by a decrease in transition rate from the other state.

There are multiple ways to formulate a model which incorporates the desired properties. To hold the renewal rate constant means that the sum $\frac{1}{\lambda_n} + \frac{1}{\mu_n}$, or $E[Z]$, must be constant. The method used here was to model P_{1n} and γ , and from them calculate λ_n and μ_n , which may be specified according to the following:

$$\begin{aligned}\lambda_n &= \frac{\gamma}{1 - P_{1n}} \\ \mu_n &= \frac{\gamma}{P_{1n}} \\ \gamma &= \frac{1}{E[Z]}\end{aligned}\tag{54}$$

The Bayes specification for Model #3 follows:

$$\begin{aligned}
& \text{Prob}(S_{j,n} = '1') \sim \text{Bernoulli}(P_{j,1,n} F_{j,n} + (1 - F_{j,n}) \text{Prob}(S_{j,n-1} = '1')) \\
& \text{Prob}(S_{j,n-1} = '1') = \begin{cases} 1, & S_{j,n-1} = '1' \\ 0, & S_{j,n-1} = '0' \end{cases} \\
& F_{j,n} = 1 - e^{-(\lambda_{j,n} + \mu_{j,n}) \Delta t_n} \\
& \lambda_{j,n} = \gamma_j / (1 - P_{j,1,n}) \\
& \mu_{j,n} = \gamma_j / P_{j,1,n} \\
& P_{j,1,n} = P_{j,1} + A_j \sin(2\pi t_n) + B_j \cos(2\pi t_n) \\
& \gamma_j = C_j (1 - P_{j,1}) P_{j,1} \\
& \text{logit}(P_{j,1}) \sim \text{Normal}(\hat{\beta}_j, \tau_\beta) \\
& \hat{\beta}_j \sim \text{Normal}(\mu_\beta, \tau_{\mu_\beta}) \\
& \log(C_j) \sim \text{Normal}(\hat{\xi}_j, \tau_\xi) \\
& \hat{\xi}_j \sim \text{Normal}(\mu_\xi, \tau_{\mu_\xi}) \\
& \tau_\beta \sim \text{Gamma}(0.5, 0.5) \\
& \text{Amp}_{\max_j} = 0.5 - \sqrt{((P_{j,1} - 0.5)^2)} \\
& r_j \sim \text{Uniform}(0, 1) \\
& A_j = r_j \text{Amp}_{\max_j} \cos(\omega) \\
& B_j = r_j \text{Amp}_{\max_j} \sin(\omega) \\
& \tau_\xi \sim \text{Gamma}(0.5, 0.5) \\
& \tau_{\mu_\beta} \sim \text{Gamma}(0.5, 0.5) \\
& \tau_{\mu_\xi} \sim \text{Gamma}(0.5, 0.5) \\
& \mu_\beta \sim \text{Normal}(0, 0.0001) \\
& \mu_\xi \sim \text{Normal}(0, 0.0001) \\
& \omega \sim \text{Uniform}(-\pi, \pi) \\
& \Delta t_n = t_n - t_{n-1}
\end{aligned} \tag{55}$$

The remainder of the paper proceeds as follows: Section 6.3 details the methods used to set up, execute, and analyze the results of the experiment. Section 6.4 provides the results of a number of comparison and effects analyses. In Section 6.5 the results are discussed along with the limitations and other aspects of the methodology. Section 6.6 presents the major conclusions of the paper.

6.3 Method

To summarize this section, a design was created randomizing over the variables of interest. Simulated observations were generated at high-frequency time steps (the HF observations) and those data were used to derive nominal duration and renewal distributions. The HF observations were sub-sampled to create time series comparable to the asynchronous, infrequent, and unevenly spaced observational time series (the AIUS observations) typical of the field of water-quality monitoring. Time-series methods were used to predict an evenly spaced time series from the AIUS observations and the PDS method was used to estimate duration and renewal distributions from those series. For the CTMC models, posterior parameter distributions were then sampled and compared to their nominal values. Duration and renewal distributions were derived from the maximum a posteriori values of those posterior distributions. The error was measured between the distributions resulting from these four alternatives and the nominal renewal and duration distributions. The four alternatives were compared with respect to this error. The effect on this error of the components of process, observation, and modeling was evaluated. The three CTMC models were compared based on their predictive ability.

6.3.1 Experimental design

An experimental design was created using leaped Halton sequences (Kocis, Whiten 1997, Halton 1960) to establish the levels of the parameters varied during the

experiment. Leaped Halton sequences are deterministic processes having desirable quasi-random properties of uniformity and independence. A leaped Halton sequence can easily be extended at a later time should the size of the experiment need to be increased.

6.3.2 Process simulation

Hirsch and Slack (1984) and Hirsch et al. (1982) developed a non-parametric test for monotonic trend which they evaluated using simulations from a variety of time-series models for generic WQIs. One of the models they used was a sinusoid with error,

$$x_{ij} = (0.5)\epsilon_{ij} + \sin\left(\frac{\pi}{3} + \frac{\pi}{6}i\right), \epsilon_{ij} \sim N(0,1) \quad (56)$$

where $i=1,2,3,\dots,12$ indexes months and $j=1,2,3,\dots,n$ indexes years. A fifteen-period realization of this process and its ACF are presented in Figure 26.

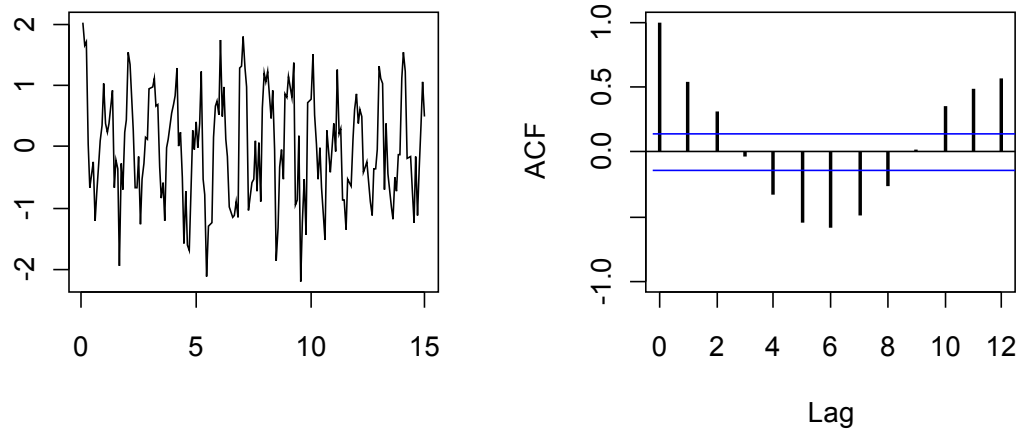


Figure 26: Simulation of a sinusoid with error

Simulated sinusoid has a period of 1 and an error variance of 1

The ACF plotted in Figure 26 indicates a seasonal pattern with a first order correlation of slightly more than 0.5. Hirsch and Slack (1984) reported that it is rare to see first-order correlation greater than 0.6 in WQI time series at a time step of one month. When the time step is 1/10,000 of a period, this correlation is maintained at a corresponding lag of 1/12 of 10,000, or approximately 833 (Figure 27).

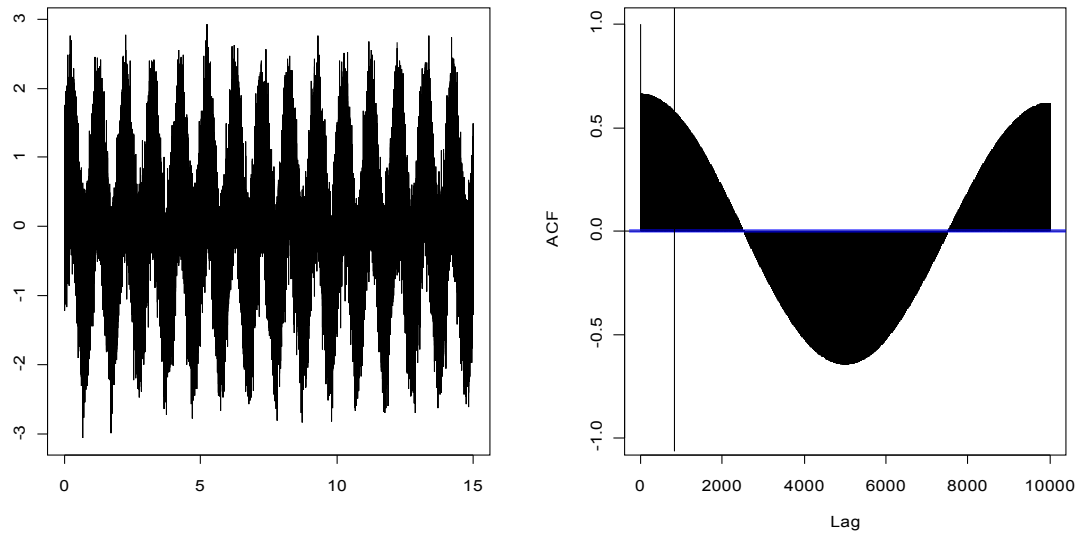


Figure 27: Simulation (left) and ACF (right) of a sinusoid with error

Multiple correlated time series can be generated by adding small but different constants to Equation (56), and correlating the errors, as in Equation (57). Ignoring the error term, this equation has an amplitude of two and a mean value of μ .

$$\begin{aligned}
 x_t &= \mu + \sin\left(\frac{\pi}{2500} + 2\pi t\right) + R_t \\
 R_t &= \epsilon_t + \phi R_{t-1} \\
 \epsilon_t &\sim N(0, \sigma^2)
 \end{aligned}
 \tag{57}$$

Two such time series A and B were simulated by first selecting a $\Delta\mu$ value from the interval (0,2) with uniform probability, where $\Delta\mu$ is the difference between

the mean values of A and B, that is $\Delta\mu = \mu_B - \mu_A$. Restricting $\Delta\mu$ to this range ensures that A and B overlap. The range of overlap is then approximately

$(\mu_B - 1, \mu_A + 1)$. A threshold was selected from this range with uniform probability.

The standard deviation of the error term was set to 0.1. The parameter ϕ was selected with uniform probability from the range (-1,1). These selections were made from their respective ranges using the leaped Halton sequence values for each parameter.

Two-state time series for A and B were constructed by applying the rule:

$$y_t = \begin{cases} 1, & x_t > K \\ 0, & x_t \leq K \end{cases}$$

For example, Figure 28 gives one period of two example time series based on nominal parameter set #135, where μ_A equaled -0.62656, μ_B equaled 0.62656, ϕ equaled 0.9661637, and the threshold was equal to -0.02315766. The correlation between these two series was 0.768395. The top figure shows both time series and the threshold. The middle figure shows the upper time series transformed to a two-state time series, and the lower figure shows the lower time series transformed to a two-state series. The threshold is coincidentally near zero.

For each of the 1500 sets of nominal process parameter values, two sets of simulations were generated for A and for B. That is, a total of four sets of simulated HF observations per set of nominal process parameters.

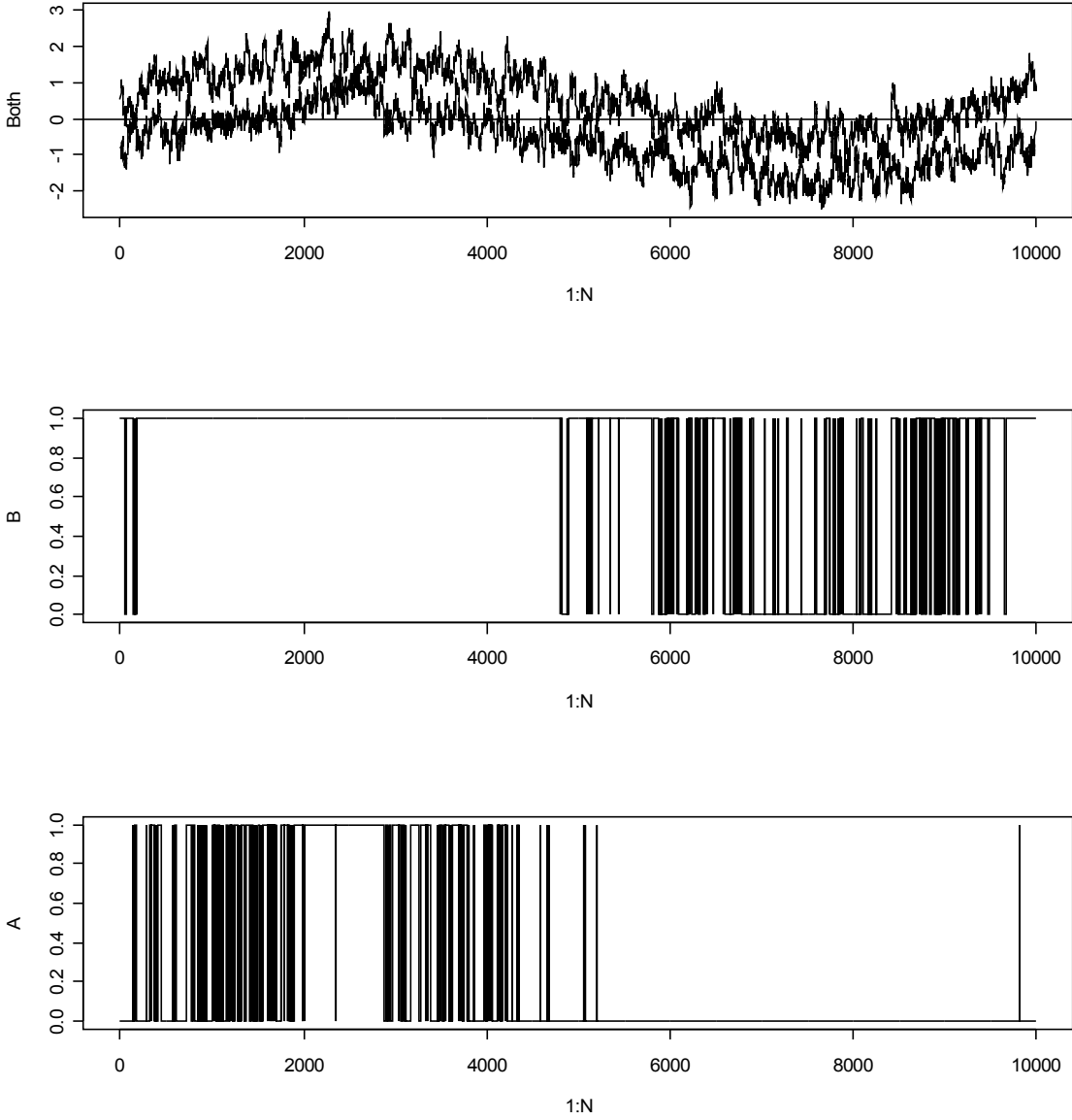


Figure 28: Simulations of correlated WQIs
As a time series of actual (simulated) values (top) and transformed to two-state processes (middle and bottom).

6.3.3 Nominal property determination

Duration and renewal interval data were extracted from the HF observations by measuring the lengths of excursions into the violation state (from threshold down-crossing to threshold up-crossing) and measuring the periods between excursions into the violation state (down-crossing to down-crossing). Nominal distributions for duration and renewal period for the nominal processes were determined empirically from the extracted duration and renewal interval data using the *sm.density* function of the *sm* package (Bowman, Azzalini 1997). Since these quantities are strictly positive, densities were estimated from log-transformed data, and then re-transformed back to the original units. Distributions for nominal parameter set #135 are given in Figure 29.

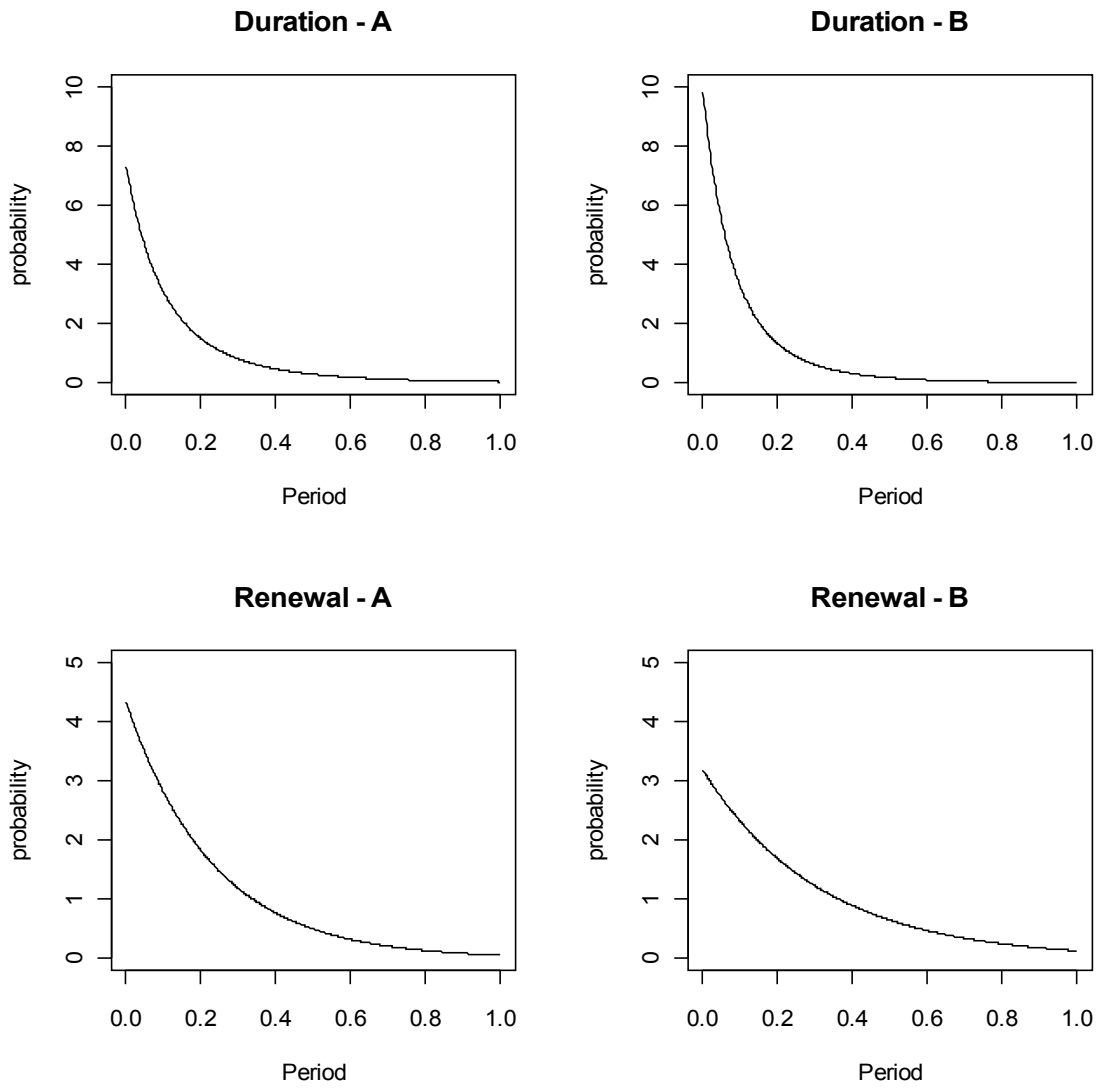


Figure 29: Nominal distributions of duration and renewal period for two processes

6.3.4 Observation simulation

Observation of these processes was simulated using the Halton sequence selection of five different parameters. First, *observation period length* was allowed to vary from five to 30 periods. This corresponds to the number of years of observations commonly seen in the field of water-quality monitoring. The same observation period length was used for both process A and process B (although commonly in the field, varying observation period lengths will be found). Second, *regular observation interval* was allowed to vary for each simulated process from 0.01 to 1 periods, with standard deviation of the interval equal to the mean interval length divided by 100. Typically in the field, observations are made weekly, monthly, quarterly, or annually. Third, *event sampling probability* was allowed to vary for each simulated process from zero to a probability which would produce on average no more than 100 observations per period. The purpose of this parameter was to mimic the practice of making observations during high-discharge periods. It is not unusual for research/monitoring programs to be resource-limited and have to vary the frequency of observation during these periods depending on their characteristic frequency and duration. The AIUS observations for parameter set #135 are shown in Figure 30.

Training and test sets were generated by doubling the observation period length and assigning simulated observations from the first half to the training set and the second half to the test set. Normally this resulted in different numbers of simulated observations

in each set. Some cases resulted in too few observations in the test set to calculate the AUC.

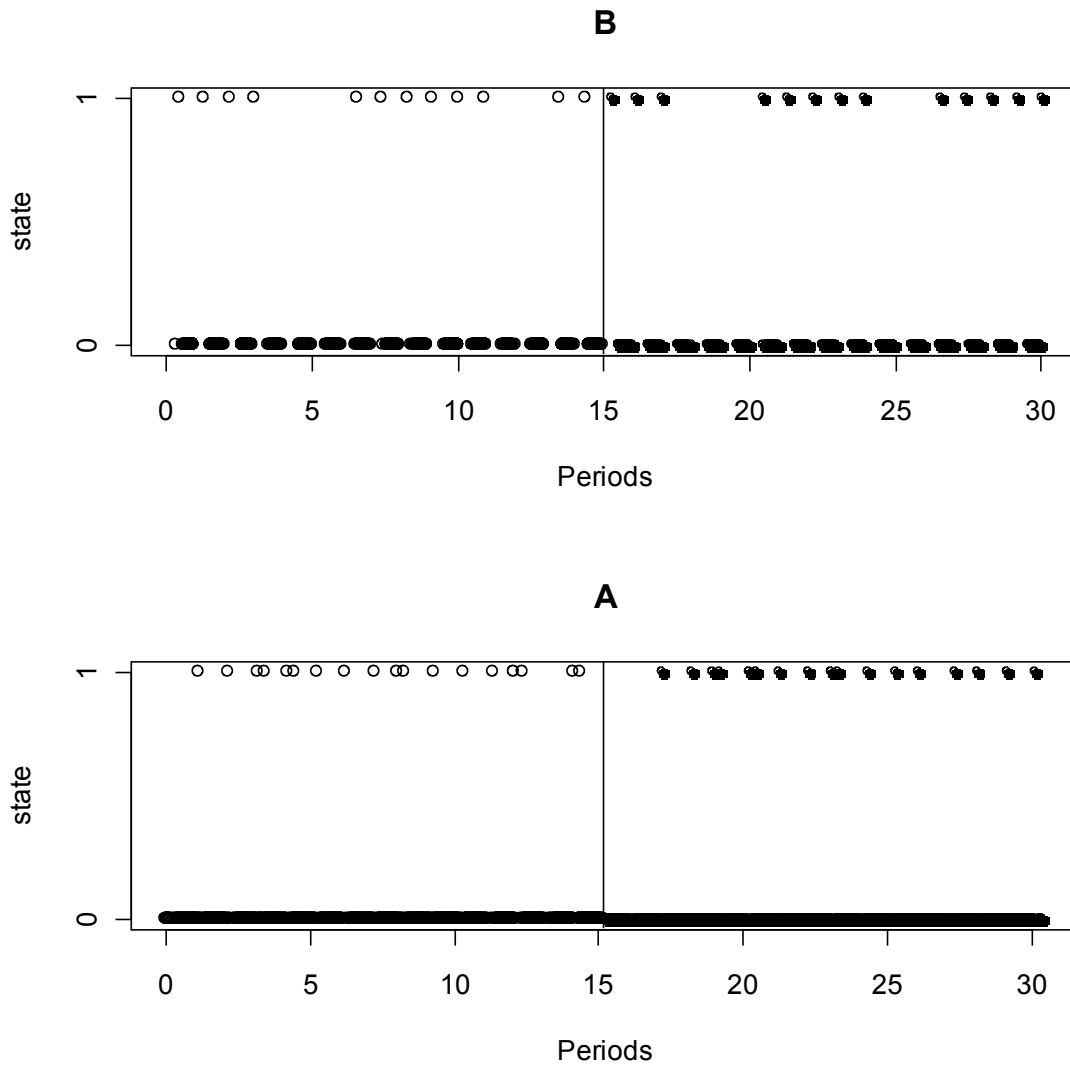


Figure 30: Simulated observations of state

Vertical line near period 15 separates training set from test set

6.3.5 Modeling simulation

Each set of AIUS observations was modeled using two methods. The first method combined the use of a time-series model to predict the process level and a variation on the Peaks-Over-Threshold (POT) method to estimate the distributions of duration and renewal period. The second method used was to estimate the same distributions using one of the three CTMC models. The selection of which model to use was determined from a Halton sequence as part of the experimental design.

The time series model estimated from the AIUS observations was similar to the model in Equation (57). However, the time step for modeling and prediction was determined from the minimum distance between the AIUS observations, which was in general less frequent than those in the HF observations. Observation times were shifted to align with the chosen time step. By choosing the minimum distance, no two observations aligned on the same time step, yet the number of time steps with missing data was greatly reduced and at least some observations at a lag of one was ensured. Parameter estimation was performed on the training data using SAS PROC ARIMA (SAS Institute Inc. 2000-2004) ; step-ahead predictions were made on the test data. The step-ahead predictions were compared to the threshold to determine periods of violation-state excursions and time between onsets of violation-state excursions.

The training set of each pair of processes was used to estimate process characteristics by both estimating each process's characteristics individually and by estimating them jointly. OpenBUGS (Thomas et al. 2006) and R were used to perform

the estimations on the Cross-Campus Grid (XCG) at the University of Virginia (Morgan, Grimshaw 2007). OpenBUGS is a software package which implements the Gibbs sampling method of Monte Carlo Markov Chain (MCMC) Bayesian estimation. Models are prescribed using R-like syntax and functions. From random starting values, a vector of parameter values moves forward iteratively according to specified prior distributions and model specifications. The user may specify multiple *chains* of parameters values, whose *convergence* to a common distribution may be tracked using the Gelman-Rubin statistic (\hat{R}). Several model fit statistics are generated, including the *deviance information criterion*, or DIC (Spiegelhalter et al. 2002).

Estimations were performed using three chains. After burn-in, sufficient iterations were run to obtain 350 values from each chain, for a total of 1050 samples from the posterior distributions. A ubiquitous problem in Bayesian MCMC estimation is determining a proper (minimum) burn-in period and a thinning parameter. Usually these choices are made while operating the software interactively. The necessary burn-in period and thinning parameter can vary widely from one estimation run to another. With a total of 3,000 estimations to perform, each requiring anywhere from 15 minutes to several hours, a systematic method of performing the estimation was needed. Therefore each estimation run was initialized with a relatively small burn-in period and thinning parameter. Then, if convergence was not attained, the thinning parameter was increased and the procedure ran again, starting from the last iteration's values of the previous run, and considering all iterations of the algorithm up to that point to be burn-in. Following

the recommendation of Gelman and Hill (2007), all parameters including the deviance were required to have an \hat{R} value less than 1.1 before considering the iterations to have converged. To guard against premature convergence, a test of chain slope was added, and it was required that no chain have a significant slope at a p-value $< 0.01/3p$, where p is the number of parameters. A linear model with autocorrelated errors was used because of the typical high serial correlation of chain values.

6.3.6 Metric determination

Three metrics were calculated for this experiment: 1) integrated squared error (ISE) (Scott 1992, Bowman, Azzalini 1997), 2) the area under the ROC curve (AUC) (Hanley, McNeil 1982), and 3) a binary parameter recovery metric which indicated whether or not the nominal parameter value was within a 95% credible interval determined from the posterior distribution for the parameter. The binary metric computed was determined from the percentile of the posterior distribution corresponding to the nominal parameter value. Percentile values between 2.5 and 97.5 were considered to be representative of cases where the nominal parameter value was recovered. Under optimal operating conditions, the nominal parameter value should be recovered 95% of the time, based on this metric.

Densities for the distributions of duration and renewal period based on time-series modeling were determined empirically from the timing and lengths of excursion periods derived from the predicted time series. The *duration ISE* and the *renewal ISE* for the

PDS method were calculated from the squared difference between the nominal and PDS densities. Similarly to the recovery metric, which provides information on how well parameter values were recovered, the ISE provides information on how well distributions were recovered.

Densities for the distributions of duration and renewal period for the CTMC methods were derived using the maximum a posteriori (MAP) values of the empirical posterior distributions of the model parameters. The MAP values were determined using the R *density* function. In the case of the homogeneous CTMC model, the distributions follow directly from the rate parameters λ and μ . In the case of the two non-homogeneous models, the theoretical forms of the distributions are complicated. The points of the densities were determined by averaging numerically over the values the density could take over two time periods. The ISEs for the CTMC methods were then calculated from the nominal and CTMC densities for duration and renewal period.

Determination of the AUC statistic was made in R using functions in the ROCR package (Sing et al. 2005). The functions require a vector of probabilities and a vector of categorical responses. AUC can vary between zero and one, where one indicates perfect prediction; however, a value of 0.5 corresponds to random guessing. Therefore, only values above 0.5 are acceptable. The AUC, while not indicative of recovery, is indicative of precision, especially when calculated in a cross-validation framework. High AUC scores indicate that the parameters of the observed process have been accurately

estimated, although if the observed process has different stochastic properties than the underlying process, then high AUC may not indicate good recovery.

6.3.7 Evaluation method

The three primary process characteristics of interest assessed in this step were the expected duration period $E[X]$, the expected renewal interval $E[Z]$, and the limiting probability for state “1” P_1 . The percentiles used to calculate the binary metric previously defined were also used to assess recovery bias. The binary metric was then used as the dependent variable in a logistic regression with the process characteristics and observation protocol characteristics as regressors. This model allowed assessment of the predicted recovery rate associated with different regions of the experimental design space, and through analysis of the steepest gradient of the response surface (Myers, Montgomery 1995), the direction in which greater or lesser recovery rate occurs. These analyses were performed with the results from the CTMC models.

The PDS and CTMC methods were compared using the Wilcoxon signed-rank paired test. Each set of AIUS observations was modeled using the PDS method and one of the three CTMC methods. The three CTMC methods were compared with each other using the Wilcoxon signed-rank test. In this case since the CTMC models were not applied to the same simulated observations, the test was not paired. The effect of the process, observation protocol and the modeling selection characteristics on performance metrics were determined using regression analysis.

6.4 Results

The 3,000 sets of simulated observations were all modeled with the time series/PDS method, and each modeled with one of the three CTMC models (Model #1 – 1002 times, Model #2 – 1000 times, Model #3 – 998 times), the specific model being selected as part of the experimental design. Convergence was not obtained for three of the sets modeled with Model #3. The ISE statistic could not be calculated for 237 of the scenarios from the time series/PDS results (actually it was infinity for these scenarios), and could not be calculated for six of the CTMC results. AUC could not be calculated for 12 of the scenarios modeled with Model #1, 10 of those modeled with Model #2, and eight of those modeled with Model #3.

6.4.1 Recovery of nominal parameter values

There are three primary nominal attributes of interest: 1) $E[X]$ – the expected value of time spent in the violation state, 2) $E[Z]$ – the expected value of the renewal interval (time between re-entries to the violation state, or the inverse of frequency of occurrence), and 3) P_1 – the limiting probability for the non-violation state. Over the range of the experimental design space, $E[X]$ and $E[Z]$ were overestimated and P_1 was underestimated (Table 27). This infers that frequency was underestimated and that P_0

was overestimated. The overestimation of $E[X]$ and P_0 favors the conservationist, whereas the underestimation of frequency favors the polluter.

Table 27: Counts of recovered nominal values across all priors

Group	# below c.i. (nominal was overestimated)	# inside c.i.	# above c.i. (nominal was underestimated)	total	% inside
E[X]	5664	332	0	5996	5.54
E[Z]	5554	437	5	5996	7.29
P1	152	934	4910	5996	15.58

To gain understanding of the relationship between recovery and process/observation characteristics, a basic response surface methodology (RSM) analysis was performed using logistic regression and a first-order model without interaction. Independent variables included the following (nominal values): $E[X]$, P_1 , $E[Z]$, observation period, quasi-regular observation interval, event observation probability, and the CTMC model as a categorical variable. Numerical variables were transformed for the analysis because they need to be on unbounded scales in order to make predictions at any point along the path of steepest ascent. The observation protocol properties were included to avoid masking effects. The six numerical regressor terms were all highly significant (Table 28), as was the term for CTMC model #3, which gave a significantly higher predicted recovery rate than the other two models.

Table 28: RSM regression results

```

> summary(glm.RSM.All)

Call:
glm(formula = OK ~ log_E_X + logit_P1 + log_E_Z + log_oP + log_rS +
     logit_eP + testModel, family = "binomial", data = Recovery.All)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4885  -0.4322  -0.2906  -0.1953   3.0501

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.91989    0.36392  -5.276 1.32e-07 ***
log_E_X      -0.87837    0.17116  -5.132 2.87e-07 ***
logit_P1     -0.35070    0.08113  -4.323 1.54e-05 ***
log_E_Z       1.43060    0.20722   6.904 5.06e-12 ***
log_oP       -1.00569    0.05556 -18.102 < 2e-16 ***
log_rS       -0.59537    0.02688 -22.149 < 2e-16 ***
logit_eP     -0.74145    0.02251 -32.940 < 2e-16 ***
testModel2   0.11434    0.07290   1.569  0.117
testModel3   0.62473    0.06835   9.140 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 11268.4 on 17987 degrees of freedom
Residual deviance: 9154.8 on 17979 degrees of freedom
AIC: 9172.8

Number of Fisher Scoring iterations: 6

```

The mean vector, disregarding CTMC model, of the transformed experimental design space is given in Table 29.

Table 29: Experimental design space mean vector

```

> mean.vector
      log_E_X      logit_P1      log_E_Z      log_oP      log_rS      logit_eP
1 -6.600725  0.002774700 -5.714353  2.758917 -0.9526023 -4.674129

```

The probability of recovery for Model #1 associated with this vector was 0.04566282. The unit vector in the direction of steepest ascent is given in Table 37.

Table 30: Unit vector along steepest ascent gradient

```
> unit.vector
      log_E_X  logit_P1  log_E_Z      log_oP      log_rS  logit_eP
1 -0.3985708 -0.1591329  0.649148 -0.4563432 -0.2701541 -0.3364411
```

Movement from any point in the direction of this vector results in a higher probability of recovery. A predicted recovery rate of 0.95 was obtained at a distance of ~2.7 units along the steepest path from the mean vector location (Figure 31).

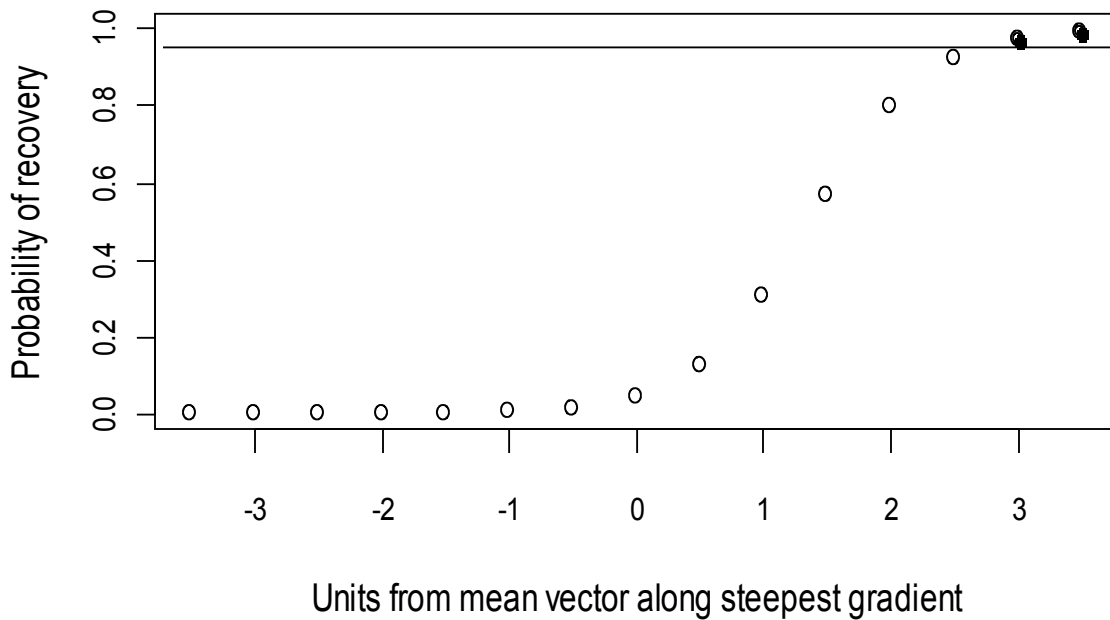


Figure 31: Probability of recovery along steepest path

The point along the steepest path corresponding to a recovery rate of 0.95 for Model #1 (Table 31) is within the ranges of four of the regressor variables, but not all six.

Only 15 scenarios had predicted recovery rates greater than 0.90. This made it difficult to quantify the conditions under which good recovery could be expected to occur.

Table 31: Experimental design space ranges and 95% recovery point along steepest path

	E[X]	P1	E[Z]	Observation period	Quasi-regular spacing	Event observation probability
minimum	0.0001	0.0390	0.0002	5.0008	0.0105	4.54E-005
maximum	0.0104	0.9403	0.0168	29.9789	0.9997	1.00E+000
95% contour point (Model #1)	0.0005	0.3943	0.0192	4.5712	0.1852	3.73E-003

Plots of the results (Figure 32, Figure 33, and Figure 34) indicate that the experimental design space is largely outside of the zone where acceptable recovery rates are found. To include more of the acceptable recovery rate area, the plots indicate that the upper limit of $E[Z]$ should be increased, and the lower limit of observation period decreased, although this latter indication is counter-intuitive. However, a first order response surface model is only a crude approximation, and it is likely that additional adjustments to the experimental design space dimensions might need to be made following a subsequent experiment with the adjusted input space. The objective of a subsequent experiment would be to explore the region on either side of the 0.95 contour.

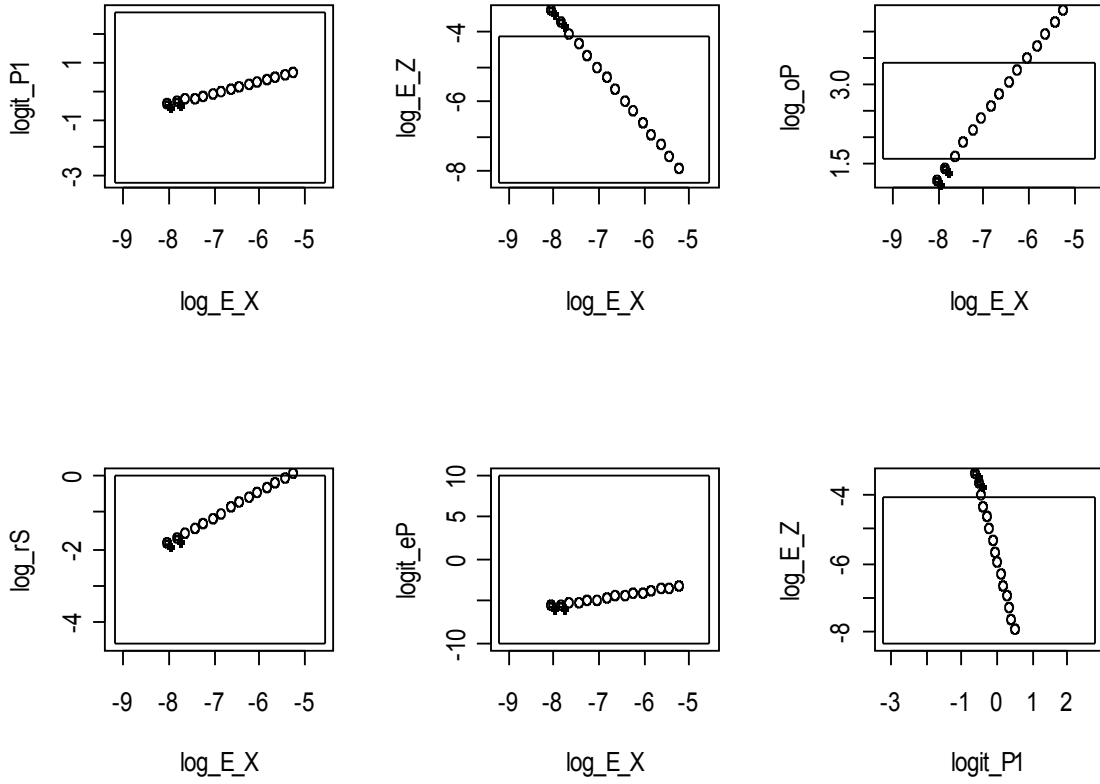


Figure 32: Steepest ascent gradient vs. process/observation characteristics I

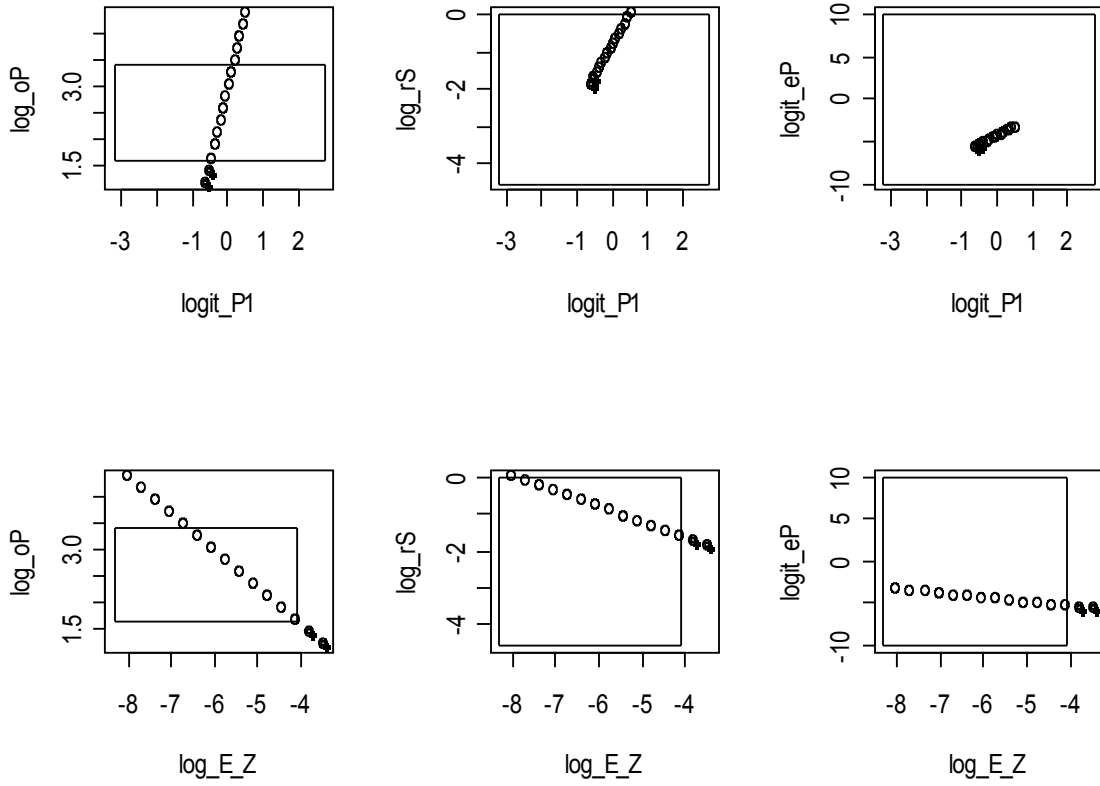


Figure 33: Steepest ascent gradient vs. process/observation characteristics II

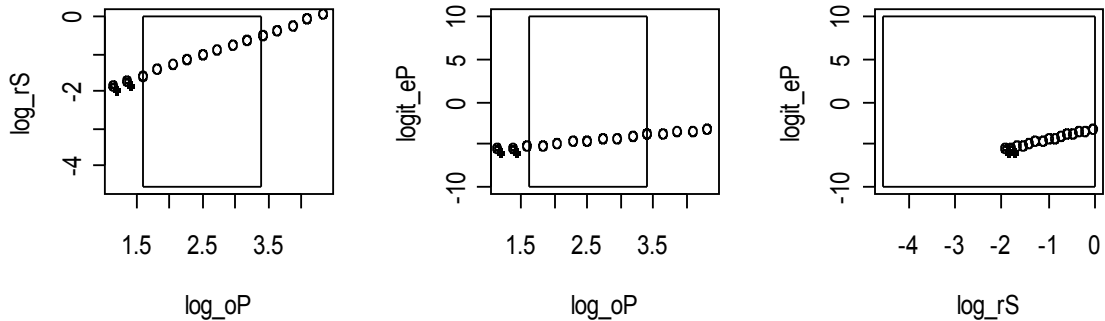


Figure 34: Steepest ascent gradient vs. process/observation characteristics III

It is not abundantly clear what adjustments would be made in the process simulation in order to obtain larger $E[Z]$ values, since the specification for the process model does not include $E[Z]$ directly. Figure 35 indicates that larger values of $E[Z]$ can only be obtained when the error correlation is very nearly 1, which is actually not unusual in water quality time series spaced at short intervals (i.e. one hour). If this is not enough to generate enough high-recovery rate scenarios, then it may be necessary to modify the span of some of the observation protocol characteristic ranges, or at least do more sampling from within restricted ranges of those variables. For example, it might be beneficial to sample the quasi-regular sampling interval preferentially from the lower end of the range (the end near 0.01 periods).

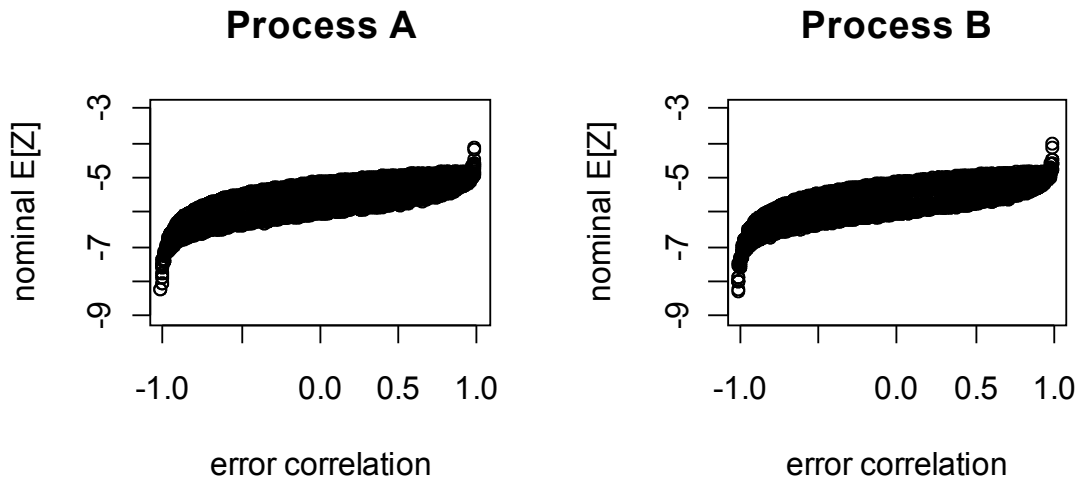


Figure 35: Effect of error correlation on expected renewal interval

6.4.2 Overall predictive performance

Over 90% of all AUC values exceeded 0.8 regardless of which of the three CTMC models was used to estimate process properties from the data (Table 32). Nearly all exceeded 0.7. The means for Models #1, #2, and #3 were 0.918, 0.902, and 0.908 respectively, while the medians were 0.936, 0.919, and 0.930.

Table 32: Cumulative distribution of AUC values

AUC	>0.9	>0.8	>0.7	>0.6	>0.5	N	Mean
Model #1	0.76	0.95	0.99	1.00	1.00	990	0.918
Model #2	0.63	0.92	0.97	0.99	1.00	990	0.902
Model #3	0.69	0.92	0.97	0.99	0.99	990	0.908

The previous results were confirmed by a Wilcoxon signed rank test for differences in population location. Both Model #1 and Model #3 had significantly greater AUC values than Model #2, but were not significantly different themselves (Table 33).

Table 33: Wilcoxon signed rank test on AUC – MAP (bigger is better)

Test	95 percent confidence interval:		Difference in location	p-value
	Lower	Upper		
Model #1 – Model #2	6.70E-003	1.50E-002	1.10E-002	0.0000
Model #1 – Model #3	-9.50E-004	6.78E-003	2.90E-003	0.1456
Model #2 - Model #3	-1.20E-002	-3.30E-003	-7.60E-003	0.0006

6.4.3 Intra-method comparison

Estimated distributions of the log of duration and renewal ISEs for each process A and B are shown in Figure 36. Visually it appears that the distributions from the CTMC estimations lie to the left of those from the time series estimations, indicating better predictive performance. However, the differences are small.

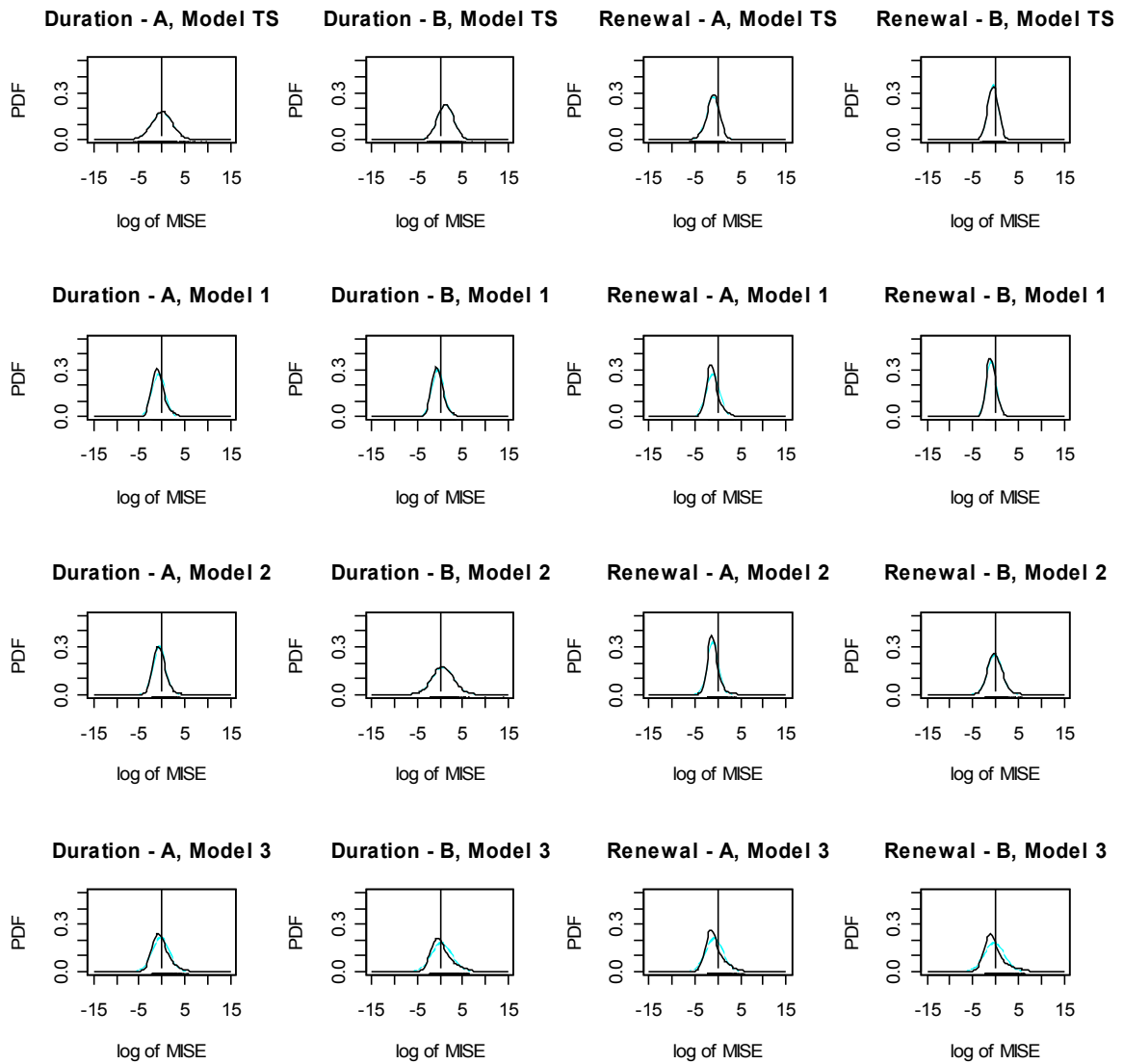


Figure 36: ISE distributions by process and method

Results of a Wilcoxon signed rank test of the paired differences between the time series method ISEs and the CTMC method ISEs were mixed. In all three duration ISE cases, the CTMC method resulted in significantly lower duration ISE than the time series method (Table 34). On renewal ISE the CTMC methods were superior in one case and

time-series methods were superior in two cases (Table 35). Model#1 was superior on both duration ISE and renewal ISE. Results were mixed on the other two models. For each test, only trials with valid values for all four ISE statistics were included.

Table 34: Wilcoxon signed rank test on duration ISE (smaller is better)

Test	N	95 percent confidence interval:		(pseudo) median	p-value
		Lower	Upper		
TS – Model #1	1914	1.20E+000	1.48E+000	1.33E+000	0.0000
TS – Model #2	1898	1.46E-001	3.19E-001	2.32E-001	0.0000
TS – Model #3	1900	2.40E-001	4.78E-001	3.61E-001	0.0000

Table 35: Wilcoxon signed rank test on renewal ISE (smaller is better)

Test	N	95 percent confidence interval:		(pseudo) median	p-value
		Lower	Upper		
TS – Model #1	1950	8.48E-002	1.24E-001	1.04E-001	0.0000
TS – Model #2	1922	-4.61E-001	-3.38E-001	-3.97E-001	0.0000
TS – Model #3	1934	-2.15E-001	-7.47E-002	-1.36E-001	0.0000

6.4.4 Process and observation protocol characteristic effect on AUC

Regression analysis was used to determine the effect on AUC of the process and observation characteristics and the model selected. The three characteristics describing the processes were the difference in the process means, the correlation between the processes, and the ϕ parameter. Difference in process means was dropped from the

regression because of its high negative correlation with the inter-process correlation. The six characteristics describing the observation protocol were the observation period, the regular spacing intervals, the event observation probabilities, and the threshold level. AUC was logit-transformed to produce a dependent variable with a more-normal distribution. A stepwise regression was used to select the model. Variable names appearing in the model and their definitions are given in Table 36.

Table 36: Characteristic definitions

Characteristic	Definition
obsPeriod	Number of periods processes were “observed” in simulation
regSpacing.A	Mean interval between regularly spaced observations – process A
regSpacing.B	Mean interval between regularly spaced observations – process B
eventProb.A	Probability of making an observation in the violation state – process A
eventProb.B	Probability of making an observation in the violation state – process B
threshold	Arbitrary threshold separating state '0' from state '1'
rho.y	Inter-process correlation
testModel	Indicator for model used in the estimation step (Model #1, #2, or #3)

The results of the regression are in Table 37. The adjusted R^2 is fairly low but the model is significant. All of the second-order terms are significant, although some of the first order terms are not. There are no insignificant first-order terms, other than for the selected model, that are not involved in second-order terms. Because of this fact, interpretation of effects was done graphically. Only one of the process characteristics, *rho.y*, was retained in the model, whereas all of the observation protocol characteristics were retained.

Table 37: Regression results

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.892e+00	1.111e-01	17.029	< 2e-16	***
obsPeriod	2.288e-03	3.736e-03	0.612	0.540366	
regSpacing.A	1.051e+00	1.526e-01	6.886	6.98e-12	***
regSpacing.B	7.156e-02	1.494e-01	0.479	0.632036	
eventProb.A	1.330e+01	5.633e+00	2.361	0.018296	*
eventProb.B	-1.461e+00	8.315e-01	-1.757	0.079030	.
threshold	-3.647e-01	1.260e-01	-2.895	0.003817	**
rho.y	1.137e+00	1.337e-01	8.500	< 2e-16	***
as.factor(testModel)2	-1.521e-01	3.257e-02	-4.669	3.17e-06	***
as.factor(testModel)3	-8.850e-03	3.264e-02	-0.271	0.786311	
eventProb.A:threshold	4.445e+01	5.302e+00	8.384	< 2e-16	***
regSpacing.A:rho.y	-2.224e+00	2.236e-01	-9.946	< 2e-16	***
regSpacing.B:threshold	-8.771e-01	1.414e-01	-6.203	6.33e-10	***
regSpacing.A:regSpacing.B	-8.723e-01	1.633e-01	-5.340	9.98e-08	***
regSpacing.A:eventProb.A	6.332e+01	9.090e+00	6.966	4.02e-12	***
regSpacing.A:threshold	7.534e-01	1.566e-01	4.811	1.58e-06	***
eventProb.B:rho.y	3.725e+01	3.919e+00	9.505	< 2e-16	***
eventProb.B:threshold	1.126e+01	1.987e+00	5.665	1.61e-08	***
eventProb.A:eventProb.B	-2.380e+02	4.545e+01	-5.237	1.75e-07	***
regSpacing.A:eventProb.B	-4.293e+00	1.203e+00	-3.568	0.000365	***
regSpacing.B:eventProb.B	2.693e+00	8.439e-01	3.191	0.001434	**
obsPeriod:regSpacing.B	-1.830e-02	6.451e-03	-2.836	0.004594	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 0.7226 on 2948 degrees of freedom (27 observations deleted due to missingness)					
Multiple R-squared: 0.2221, Adjusted R-squared: 0.2165					
F-statistic: 40.07 on 21 and 2948 DF, p-value: < 2.2e-16					

Because of the way the experiment was designed, process B was almost always above the threshold (state '1' or the non-violation state), whereas process A was almost always below threshold. Because of this, the constraint on the number of event observations per period affected process A much more than process B. The result was that *eventProb.A* only spanned a range of about [0, 0.073], whereas *eventProb.B* spanned the entire range [0,1]. The plots in Figure 37 indicate that higher AUC was associated

with higher event observation probability. Subplot [b] indicates a decline in AUC in the upper right-hand corner, however there aren't any observations in that region to support that inference; it is most likely an artifact. Whenever *eventProb.B* is greater than about 0.2 (on these plots), AUC is nearly 1.0, except for a small region in subplot [d].

Whenever inter-process correlation is above 0.2, AUC is generally above 0.95. More positive values of threshold also tend to result in higher AUC. The effect of a higher threshold is to increase the amount of time process B spends below threshold (both processes really, but proportionally more so for process B), which may in some cases increase the total number of violation-state observations made.

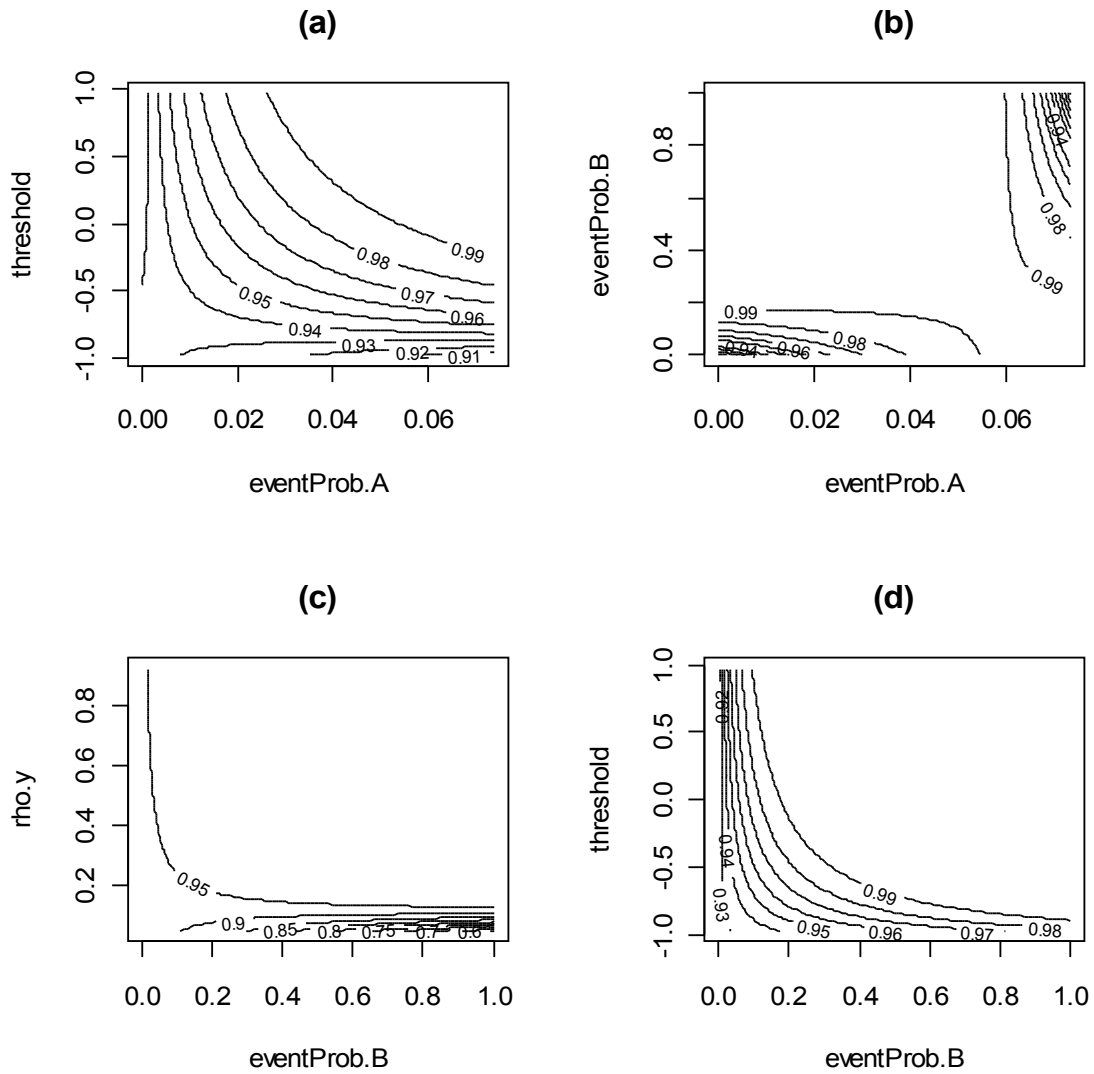


Figure 37: Second-order effects I

Most of the subplots in Figure 38 and Figure 39 are self-explanatory, however a few seem counter-intuitive at first glance. For example subplot [c], which maps the interaction between *regSpacing.A* and *threshold*, indicates that the relationship between

AUC and *regSpacing*. *A* reverses as *threshold* goes from negative to positive. To understand this it is necessary to understand the implication of *threshold*. When *threshold* is near +1 or -1 it must be that process A and B have a high degree of overlap. However, near +1 indicates that both processes are below threshold a great deal of the time, whereas near -1 indicates the opposite. When *threshold* is near 0 the degree of overlap may range from none to complete, and the proportion of time below threshold will be dissimilar for each process. When processes are below threshold more event observations are made, and so threshold is somewhat of a surrogate for event observation probability, although the two are not correlated. Subplot [a] reflects this effect of *threshold* as when inter-process correlation is low, there is little overlap, restricting values of *threshold* to be near zero.

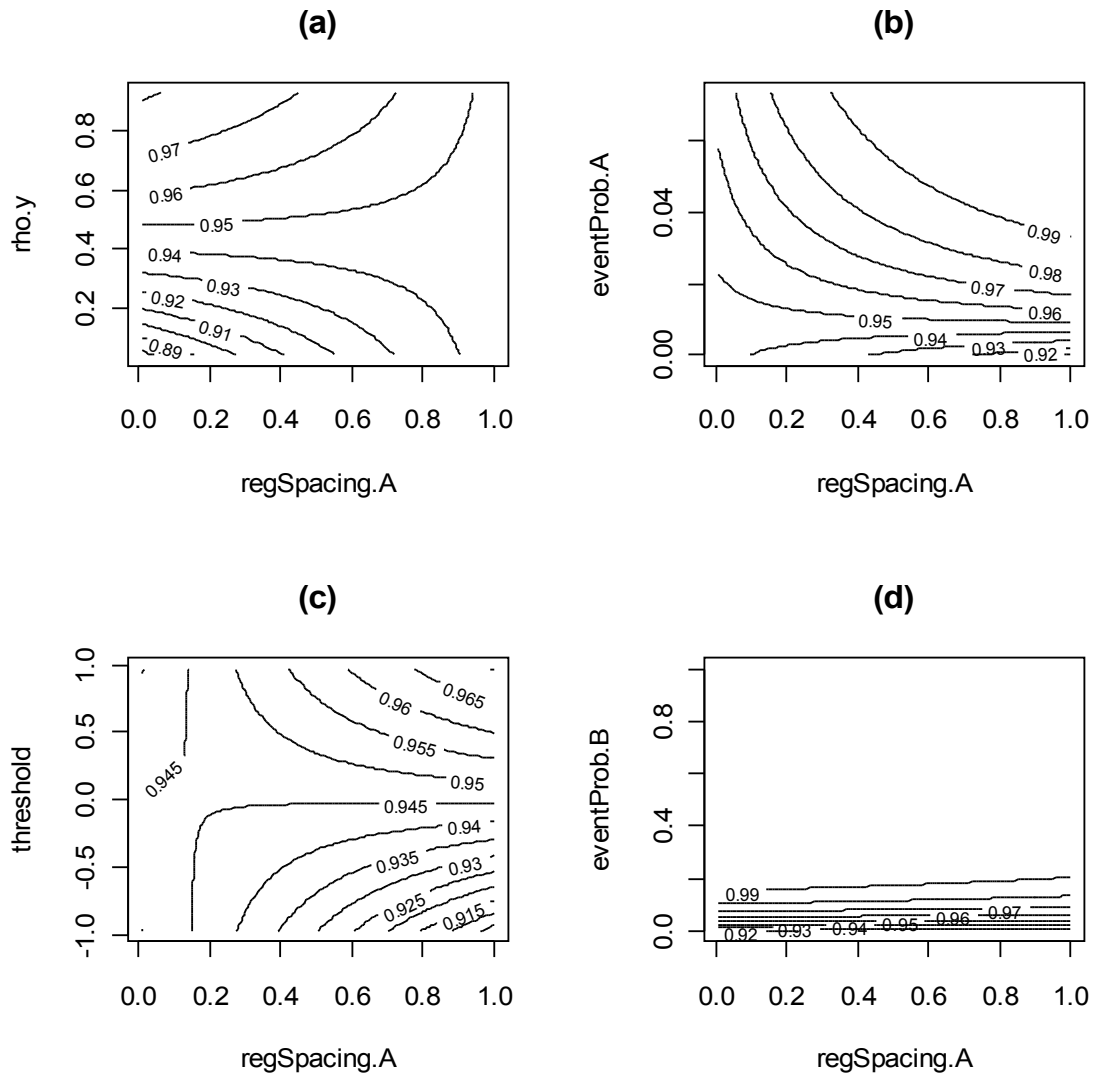


Figure 38: Second-order effects II

Subplot [d] of Figure 39 appears counterintuitive in that, for a fixed value of *regSpacing.B*, AUC decreases with increasing *obsPeriod*. However, the first-order coefficients for both characteristics were insignificant. Concentrating therefore on the

interaction, any line through the origin, other than one that is purely vertical or horizontal, represents a fixed number of observations over the observation period. For example, the line through the point (30,1) corresponds to 30 observations, regardless of the observation period. That is, for shorter observation periods the observation rate is more intense. One interpretation of this subplot is then that higher AUC is obtained with higher observation rates. Steeper sloped lines indicate fewer observations than less steep slopes. Another interpretation therefore is that higher AUC is obtained with more observations.

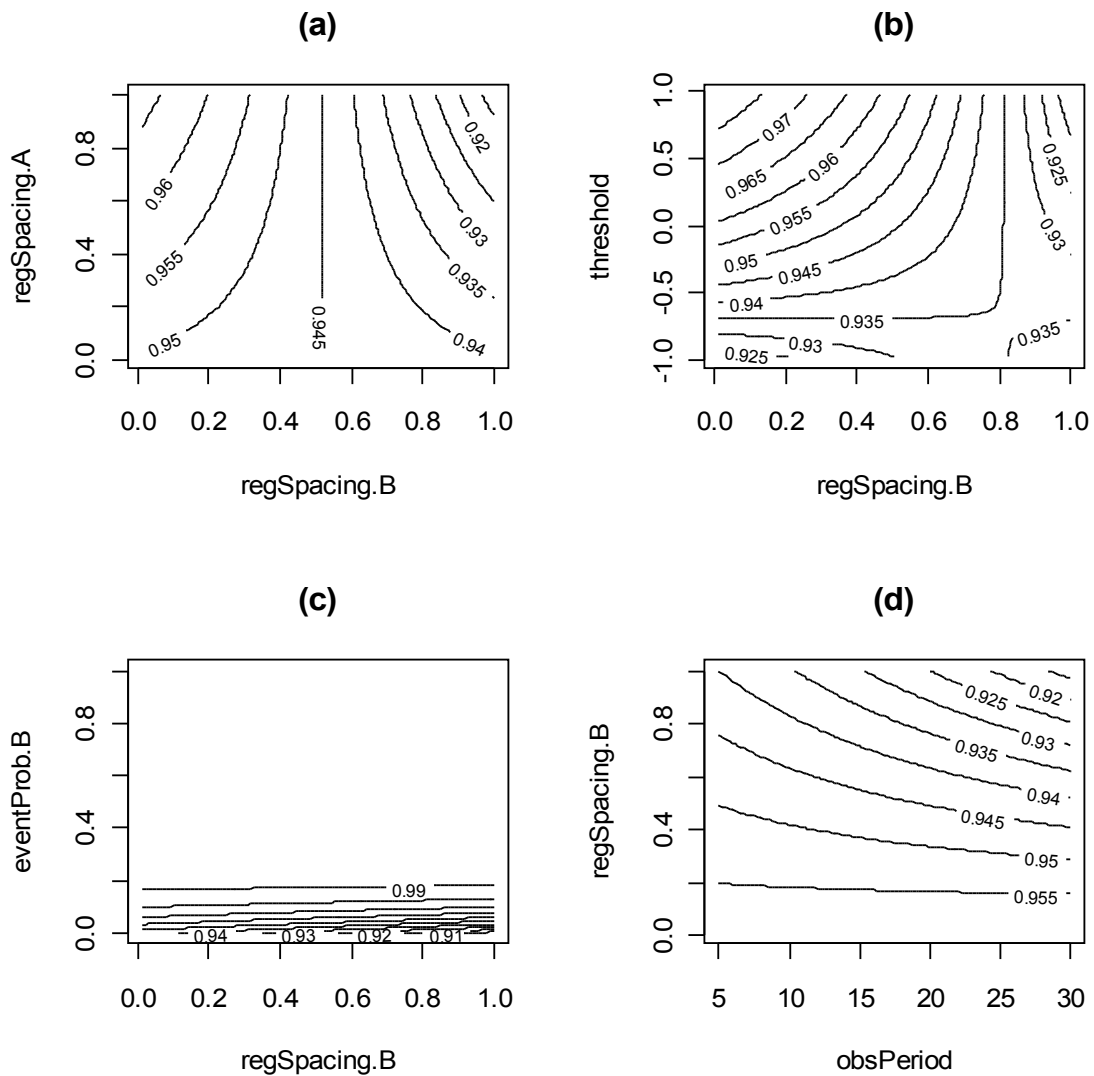


Figure 39: Second-order effects III

6.5 Discussion

The objective of this paper was to evaluate the performance of three hierarchical CTMC models for threshold exceedance behavior and compare them with a method

adopted from the flood-frequency analysis domain and similar to one that has been used by other researchers (Madsen, Rasmussen & Rosbjerg 1997, Wang 1991, Deviney, Rice & Hornberger 2006). All models and methods were applied to simulated asynchronous, unevenly spaced, and uncoordinated observations typical of the field of water-quality monitoring. The evaluation consisted of assessments of nominal parameter recovery, of out-of-sample prediction performance, of a comparison with the PDS method, and of the effect of various process and observation characteristics on said performance. The purpose of these assessments was to compare the three candidate models with each other and the PDS approach, and to establish the conditions under which the CTMC models could be used, assuming that successful application would be a function of both process characteristics and of observation protocol characteristics.

The CTMC methodology was found to produce biased parameter estimates over much of the experimental design space considered. This space was specified to represent process characteristics and observation protocol characteristics thought to be common in the water quality monitoring field. Under these assumptions the method can be expected to overestimate violation state duration and long-term probability of violation-state occurrence, which favors the conservationist, and to underestimate frequency, which favors the polluter. It appears these biases could be reduced; however, additional experiments should be conducted to determine the conditions under which good recovery of true process parameter values could be expected. The major source of bias appears to be the observation protocol characteristics.

In this experiment, the simulation of violation-state observation was made under the assumption that the observer would know that the process was in the violation-state. In practice this is not generally known. In some cases automated samplers can be activated by monitoring a different process, such as discharge, known to be correlated with the violation-state. To model this degree of reality in this experiment would have required the simulation of the different process as well, and this was not seriously considered. However to be more realistic the occasional high-frequency observation of the process should have some probability of occurring during non-violation-state periods.

Estimation of frequency (renewal period) and duration properties of the processes were made with the CTMC models and with a partial-duration series method adopted from the flood-frequency literature. The homogeneous CTMC model (Model #1) exhibited lower error on both duration and renewal period (frequency) distributions than the partial duration series method or either of the non-homogeneous CTMC models. The non-homogeneous CTMC models exhibited lower error on duration than the PDS method, but greater error on renewal period. Interestingly, Model #1 performed the worst on recovery. Since there was evidently some transformation of the underlying process made by the observation protocol, this is not necessarily a contradiction.

The three CTMC models were compared for predictive ability, and a regression analysis was performed to measure the relationship between predictive performance and process and observation protocol characteristics. The high median AUC values for all three models indicated that parameter estimation for the observed processes was

generally accurate. This suggests that the source of bias in recovery was because of characteristics of the observation protocol and not because of the models or estimation method. Additional experiments would be needed to tease out the exact causes of this bias and under what conditions it could be avoided. The homogeneous model exhibited better predictive performance, again in contrast to the recovery results. More-positively correlated processes were easier to predict. Not surprisingly, higher observation rates, regardless of type, were found to improve predictive performance. However, there was considerable unexplained variance associated with the model.

Unfortunately it is difficult at this point, without additional experimentation, to recommend one model or method over another.

In their seminal paper on the Seasonal Kendall Tau test for trend, Hirsch and Slack (1982) evaluated a non-parametric test for trend using six different trend-free models (reproduced in Table 38) for simulated monthly observations (12 per year) of univariate WQI processes. Trends of various magnitudes were added to these models to test for power. A modification of the NIS model from Table 38 was used in this paper because, in the author's experience, this model corresponds to a large number of actual WQI processes. The simulations from this model yielded processes with expected duration lengths that ranged from about 0.0001 to 0.01 periods, or from roughly one hour to one-half week, assuming one period equals one year. The expected return intervals ranged from about 0.0002 to 0.0168 periods, or from roughly two hours to one week (in frequency terms from roughly 60 times to 5000 times per period). These seem to be a bit

short of the range that would normally be thought to be of interest, particularly with respect to the expected return interval, where even an expected return interval of one year can be problematic. Additional investigations could consider additional models for water-quality processes from Table 38 that might yield processes with different properties. There are also a number of physical models for water-quality processes, however their use would be considerably more complicated.

Table 38: Trend-free models for monthly WQIs

1. Normal independent (NI)	$x_{ij} = \epsilon_{ij}$	(58)
2. Log normal independent (LNI)	$x_{ij} = \exp[0.83 \epsilon_{ij} - 0.35] - 1.0$	(59)
3. Normal independent with seasonal cycle (NIS)	$x_{ij} = (0.5) \epsilon_{ij} + \sin\left(\frac{\pi}{3} + \frac{\pi}{6} i\right)$	(60)
4. Normal autoregressive (NAR)	$x_{ij} = 0.2 [x_{ij}]_L + 0.98 \epsilon_{ij}$	(61)
5. Normal autoregressive-moving average (NARMA)	$x_{ij} = 0.75 [x_{ij}]_L + 0.97 \epsilon_{ij} - 0.57 [\epsilon_{ij}]_L$	(62)
6. Log normal, autoregressive with seasonal cycle (LNARS)	$x_{ij} = (0.5) \exp[0.22 [x_{ij}]_L + 0.80 \epsilon_{ij} - 0.35] - 0.71$	(63)

Note 1: L is a function for lag

On the one hand, the time-series model used for prediction was the same model used to generate the data. From that perspective it should have worked well. On the other hand, other potentially better time-series models, for example with more lagged terms in the error model, were not considered. In addition, there are other methods in hydrology used for modeling magnitude, frequency, and duration other than the PDS

method, such as the annual maximum series (AMS) method, which were not used. However the AMS method, in particular, is known to underestimate frequency compared to the PDS method (Madsen, Pearson & Rosbjerg 1997, Madsen, Rasmussen & Rosbjerg 1997). Even the PDS method may underestimate actual rates as it is common in actual practice to remove short-term excursions from the analysis in order to meet assumptions of independence between excursion events (Madsen, Pearson & Rosbjerg 1997, Madsen, Rasmussen & Rosbjerg 1997).

In a real-world application, it is expected that additional data will be available for prediction of group-level parameters of the hierarchical model. Which data would be useful, and how many groups and levels there are in the hierarchical model, would depend on the WQI being modeled. For example, for pH or acid-neutralizing capacity, geology is generally an important predictor of process level. For temperature, however, elevation, aspect, and latitude should be of more importance. Using Model #1 for illustration, the model might be modified as follows (showing only the relevant parts) in order to incorporate location-level predictors of transition rates and the limiting probability P_l :

$$\begin{aligned}
\text{logit}(P_{j,1}) &\sim \text{Normal}(\hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \dots + \hat{\alpha}_p x_p, \tau_{P1}) \\
\log(C_j) &\sim \text{Normal}(\hat{\xi}_0 + \hat{\xi}_1 x_1 + \dots + \hat{\xi}_p x_p, \tau_C) \\
\alpha_0 &\sim \text{Normal}(0, 0.0001) \\
&\vdots \\
\alpha_p &\sim \text{Normal}(0, 0.0001) \\
\xi_0 &\sim \text{Normal}(0, 0.0001) \\
&\vdots \\
\xi_p &\sim \text{Normal}(0, 0.0001)
\end{aligned} \tag{64}$$

In addition, it might be of interest to know if process properties are changing with time. Following is a modification to Model #3 which could be used to estimate a monotonic trend in the long-term proportion of time spent in state '1'. In this particular case, where there was an assumption of constant (but unknown) renewal period, a monotonic trend in P_l would result in monotonic trends in the transition rates, which would result in monotonic trend in the expected duration period in the violation state. The range of possible models built on this framework seems to be unlimited.

$$\begin{aligned}
\text{logit}(P_{j,1,n}) &= \text{logit}(P_{j,1}) + \hat{\alpha}_j t \\
\text{logit}(P_{j,1}) &\sim \text{Normal}(\hat{\beta}_j, \tau_\beta) \\
\hat{\alpha}_j &\sim \text{Normal}(0, \tau_{\hat{\alpha}})
\end{aligned} \tag{65}$$

One of the challenges to performing this experiment stemmed from the use of a grid of Windows machines for running scenarios. These machines were part of public computer labs at the University of Virginia (UVA). Most of them were set up to be able to run two scenarios simultaneously. Walk-in users logging in to a machine would cause

whatever scenarios were running on the machine at that time to abort. In addition, all public Windows computers were set to re-start around 4:00 a.m. each morning, which had the same effect. Although the job queuing system for the grid could automatically re-start such aborted jobs, it was found that trouble-shooting was facilitated if this was not allowed to occur. The upshot was that if all 3,000 jobs (scenarios) were loaded into the queue, at the end not all would have reached convergence. The result was that a significant amount of effort had to be devoted to job submission management, to ensure that completed scenarios were not re-run. In addition, although the overall runtime of the experiment was surely shortened, since the scenario output came back as 3,000 sets of output files, another significant effort was required to merge all of the results into datasets that could be analyzed. Of course, the inevitable coding errors and model changes required that the whole process be repeated.

A minor issue that arose was that of setting random number generator seeds. Typically this is done so that randomized experiments can be repeated exactly. Attempting to do this turned out to be a nightmare, and was therefore abandoned. First of all, with each scenario running on a different machine in potentially different order, it would have been necessary to specify a seed for each scenario. In addition, as the OpenBUGS software uses its own seed separate from the seed used by R, it would also have had to have its own seed. Even the generation of simulated observations was done on the grid, so each one of those scenarios would have required a seed. Not of insignificant importance was the observation that convergence appeared to be affected by

the choice of seed. Some seeds presumably lead to bad starting values and a failure to converge. Simply re-running a failed scenario often led to convergence. For all these reasons seeds were not set. Therefore these experiments cannot be repeated exactly. However, it is expected that conclusions made are robust enough to withstand a repeat of the experiment.

6.6 Conclusions

The objective of this paper was to evaluate the performance of three hierarchical CTMC models for threshold exceedance behavior and compare them with a method adopted from the flood-frequency analysis domain. Three hierarchical two-state CTMC models for threshold exceedance at multiple locations were introduced and tested. The first of these was homogeneous. The second was non-homogeneous with time-varying transition rates. The third was non-homogeneous with time-varying limiting probabilities. These CTMC models may be easily extended to allow for group-level prediction of process parameters using extraneous data, to allow for changes in process properties over time, and to allow extrapolation to other catchments where process observations are not available. The methodology investigated here allowed the use of all available data, despite observation schedule, to be used in an analysis.

Dual simulations of observations of correlated WQI processes were generated using a sinusoidal model with correlated errors adapted from the statistics of hydrology literature. Observation of these processes was simulated assuming a protocol of quasi-

regular observation intervals and random observation during excursions of the processes to one side of an arbitrary threshold.

Nominal parameter values were outside of the corresponding posterior distributions over much of the experimental design space. Duration and long-term probability of violation tended to be overestimated, which favors the conservationist. Frequency tended to be underestimated, which favors the polluter. Additional experiments are recommended to better define operational conditions which lead to good recovery of nominal parameter values.

Estimation of frequency (renewal period) and duration properties of the processes were made with the CTMC models and with a partial-duration series method adopted from the flood-frequency literature. The homogeneous CTMC model exhibited lower error on both duration and renewal period (frequency) distributions than the partial duration series method or either of the non-homogeneous CTMC models. The non-homogeneous CTMC models exhibited lower error on duration than the PDS method, but greater error on renewal period.

The three CTMC models were compared for predictive ability, and a regression analysis was performed to measure the relationship between predictive performance and process and observation protocol characteristics. Overall, predictive ability was excellent, with median AUC in excess of 0.90 regardless of CTMC model. The homogeneous model exhibited better predictive performance. More-positively correlated processes were easier to predict. Not surprisingly, higher observation rates, regardless of

type, were found to improve predictive performance. However, there was considerable unexplained variance associated with the model. The excellent predictive performance indicates that the bias in parameter estimates was more likely because of characteristics of the observation protocol than the estimation method.

7 Conclusions

Decisions and assessments in the field of aquatic-resources management are often made based on the perceived risk that a given water quality indicator (WQI) makes excursions to one side of a defined threshold that divides the WQI process domain into two states, one of which is recognized to have undesirable ecological consequences. The risky state is referred to as the exceedance or violation state. This risk is a function of the frequency of occurrence of violation state excursions, the duration of these excursions, and the long-term proportion of time spent in the violation state. Yet the methodology to assess such risk has not been well-developed, in part because of the cost of acquiring sufficient data to estimate risk by simply counting the frequency of and measuring the duration of violation events. This is particularly true if the problem of assessing risk across a region is considered, where observations may have been made following a number of different protocols. Data are often collected at multiple locations in a region; however, these observations are usually made at infrequent, asynchronous, unevenly spaced, and uncoordinated (with threshold crossing) times.

The objective of this dissertation was to address this problem with CTMC models and Bayesian methods. The work builds in a straightforward manner using simulations

based on simple single-process scenarios derived from continuous-time Markov chains, to multiple-process scenarios based on statistical time-series models of WQIs. To understand the performance of these models on real problems, formal simulation tests were performed over a realistic range of problem parameters. These parameters included both process and observation protocol characteristics intended to be representative of the water quality monitoring domain. Process characteristics were defined in terms related to duration, frequency, and long-term proportion. Observation protocol characteristics were defined in terms of total length of observation period, quasi-regular observation interval, and event observation probability, as many water quality monitoring protocols consist of a combination of approximately evenly spaced but infrequent observations made on a timed schedule and higher-frequency observations made on an approximately randomly spaced schedule and intended to occur during periods of threshold violation. Assessments were made of the true parameter value recovery and out-of-sample predictive capabilities of the methodology under these simulated conditions.

The major finding from these assessments was that under the range of process and observation protocol characteristics considered, estimation of true process parameter values was biased, for the most part in favor of the conservationist. In most cases duration, frequency, and long-term proportion of time spent in the violation state were overestimated.

In spite of this bias, predictive performance on simulated out-of-sample observations was excellent. This indicated that the estimation process worked well but

that the properties of the simulated observation process did not match the properties of the underlying process. This in turn suggested that the source of the bias mentioned above was not the models or the estimation process but was related to the characteristics of the observation protocols. Additional work is needed to determine if this bias stems from the models, the estimation process, or the observation protocols.

The proposed models and their evaluation provide significant theoretical, methodological, and practical contributions to the environmental systems analysis communities. From a theoretical standpoint, several new models of threshold violation have been developed that permit estimation of three very important properties of WQI processes, the frequency and duration of threshold violation events, and the long-term proportion of time spent in the violation state. Several of these models are univariate, and some are multivariate. The hierarchical nature of the multivariate models allows for additional predictors to be used in regional modeling efforts.

From a methodological standpoint, it has been shown that Bayesian methods can be used to estimate the model parameters and properties from the type of asynchronous, infrequent, uncoordinated, and unevenly spaced observations that are typical in water-quality monitoring. This was done with a combination of hierarchical two-state homogeneous and non-homogeneous Markov chain models, the Kolmogorov Backward Equation, and Bayesian inference. No other work has been found that has combined these elements for this or any other purpose. The methodology developed adds significantly to those currently available for estimating return period, duration, and long-

term probability of threshold violation events. This methodology enables the analysis of a large number of water-quality time series that have as yet gone unanalyzed for distributions of return period and duration of threshold violation events. The results of this work will change the way scientists, engineers, and managers think about and use the observations generated through water-quality monitoring.

From a practical standpoint, evaluations have been provided of the models and the estimation methods based on results analyses that have validated the approach and enabled prospective practitioners to use the methodology. Additional work would be useful to better define all operating conditions that result in unbiased estimates.

Recognizing these limitations, the results of the evaluations assist users of water quality data in assessing the usefulness of existing data, gives them a tool with which to analyze it, and gives them a basis upon which to plan future monitoring strategies. Guidance has also been given on model specification. Although the focus has been on water quality, it is expected that the evaluation methodology will be useful in other application areas as well.

8 References

- Baldigo, B.P. & Lawrence, G.B. 2000, "Composition of Fish Communities in Relation to Stream Acidification and Habitat in the Neversink River, New York", *Transactions of the American Fisheries Society*, vol. 129, no. 1, pp. 60-76.
- Baldigo, B.P. & Murdoch, P.S. 1997, "Effect of stream acidification and inorganic aluminum on mortality of brook trout (*Salvelinus fontinalis*) in the Catskill Mountains, New York", *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 54, no. 3, pp. 603-615.
- Behera, P.K., Li, J.Y. & Adams, B.J. 2000, "Characterization of Urban Runoff Quality: A Toronto Case Study", *Applied Modeling of Urban Water Systems.*, , no. 8, pp. 225-248.
- Bender, M. & Simonovic, S. 1994, "Time-Series Modeling for Long-Range Stream-Flow Forecasting", *Journal of Water Resources Planning and Management*, vol. 120, no. 6, pp. 857-870.
- Bowman, A.W. & Azzalini, A. 1997, *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, Oxford University Press.
- Brockwell, P.J. & Davis, R.A. 1996, *Introduction to time series and forecasting*, Springer, New York.
- Brockwell, P.J. & Davis, R.A. 1991, *Time series: theory and methods*, 2nd edn, Springer-Verlag, New York.
- Bulger, A.J., Cosby, B.J. & Webb, J.R. 2000, "Current, reconstructed past, and projected future status of brook trout (*Salvelinus fontinalis*) streams in Virginia.", *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 57, no. 7, pp. 1515-1523.
- Chatfield, C. 1989, *The analysis of time series : an introduction*, 4th edn, Chapman and Hall, London ; New York.

- Chu, P.S., Katz, R.W. & Ding, P. 1995, "Modelling and forecasting seasonal precipitation in Florida: A vector time-domain approach", *International Journal of Climatology*, vol. 15, no. 1, pp. 53-64.
- Clark, J.S. 2005, "IDEAS AND PERSPECTIVES: Why environmental scientists are becoming Bayesians", *Ecology Letters*, vol. 8, no. 1, pp. 2-14.
- Davies, T.D., Tranter, M., Wigington, P.J., Jr & Eshleman, K.N. 1992, ""Acidic episodes" in surface waters in Europe", *Journal of Hydrology (Amsterdam)*, vol. 132, no. 1-4, pp. 25-69.
- Deviney, F.A., Jr, Rice, K.C. & Hornberger, G.M. 2006, "Time series and recurrence interval models to predict the vulnerability of streams to episodic acidification in Shenandoah National Park, Virginia", *Water Resources Research Vol.42*, vol. 42, pp. 1-14.
- DeWalle, D.R., Swistock, B.R. & Sharpe, W.E. 1995, "Episodic flow-duration analysis: a method of assessing toxic exposure of brook trout (*Salvelinus fontinalis*) to episodic increases in aluminum", *Canadian journal of fisheries and aquatic sciences(Print)*, vol. 52, no. 4, pp. 816-827.
- Fernandez, B. & Salas, J.D. 1990, "Gamma-Autoregressive Models for Stream-Flow Simulation", *Journal of Hydraulic Engineering*, vol. 116, no. 11, pp. 1403-1414.
- Gallagher, R.G. 1996, *Discrete Stochastic Processes*, Springer.
- Gamerman, D. 1997, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, CRC Press.
- Gelman, A.B., Carlin, J.B., Stern, H.S. & Rubin, D.B. 2004, *Bayesian Data Analysis*, CRC Press.
- Gelman, A. & Hill, J. 2007, *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press, Cambridge ; New York.
- Greb, S.R. & Graczyk, D.J. 1995, "Frequency-duration analysis of dissolved-oxygen concentrations in two southwestern Wisconsin streams", *Water Resources Bulletin*, vol. 31, no. 3, pp. 431-438.
- Halton, J.H. 1960, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals", *Numerische Mathematik*, vol. 2, no. 1, pp. 84-90.
- Hamilton, J.D. 1994, *Time series analysis*, Princeton University Press, Princeton, N.J.

- Hanley, J.A. & McNeil, B.J. 1982, "The meaning and use of the area under a receiver operating characteristic (ROC) curve", *Radiology*, vol. 143, no. 1, pp. 29-36.
- Harrison, P.J. & Stevens, C.F. 1976, "Bayesian Forecasting", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 38, no. 3, pp. 205-247.
- Harvey, A.C. 1990, *Forecasting, structural time series models, and the Kalman filter*, Cambridge University Press, Cambridge ; New York.
- Hirsch, R.M. & Slack, J.R. 1984, "Nonparametric Trend Test for Seasonal Data with Serial Dependence", *Water Resources Research*, vol. 20, no. 6, pp. 727-732.
- Hirsch, R.M., Slack, J.R. & Smith, R.A. 1982, "Techniques of Trend Analysis for Monthly Water Quality Data", *Water Resources Research*, vol. 18, no. 1, pp. 107-121.
- Hollander, M. & Wolfe, D.A. 1973, "Nonparametric statistical methods, 503 pp", .
- Hosking, J.R.M. & Wallis, J.R. 1987, "Parameter and Quantile Estimation for the Generalized Pareto Distribution", *Technometrics*, vol. 29, no. 3, pp. 339-349.
- Keller, K., Yohe, G. & Schlesinger, M. 2008, "Managing the risks of climate thresholds: uncertainties and information needs", *Climatic Change*, vol. 91, no. 1, pp. 5-10.
- Kocis, L. & Whiten, W.J. 1997, "Computational investigations of low-discrepancy sequences", *ACM Transactions on Mathematical Software (TOMS)*, vol. 23, no. 2, pp. 266-294.
- Laio, F., Porporato, A., Ridolfi, L. & Rodriguez-Iturbe, I. 2001, "Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress II. Probabilistic soil moisture dynamics", *Advances in Water Resources*, vol. 24, no. 7, pp. 707-723.
- Lenton, T.M., Held, H., Kriegler, E., Hall, J.W., Lucht, W., Rahmstorf, S. & Schellnhuber, H.J. 2008, "Tipping elements in the Earth's climate system", *Proceedings of the National Academy of Sciences*, vol. 105, no. 6, pp. 1786-1793.
- Lu, Z. & Berliner, L.M. 1999, "Markov switching time series models with application to a daily runoff series", *Water Resources Research*, vol. 35, no. 2, pp. 523-534.
- MacNally, R. & Hart, B.T. 1997, "Use of CUSUM methods for water-quality monitoring in storages", *Environmental science & technology*, vol. 31, no. 7, pp. 2114-2119.
- Madsen, H., Pearson, C.P. & Rosbjerg, D. 1997, "Comparison of annual maximum series

- and partial duration series methods for modelling extreme hydrologic events. 2. Regional modelling", *Water Resources Research*, vol. 33, no. 4, pp. 759-769.
- Madsen, H., Rasmussen, P.F. & Rosbjerg, D. 1997, "Comparison of annual maximum series and partial duration series methods for modelling extreme hydrologic events. 1. At-site modelling", *Water Resources Research*, vol. 33, no. 4, pp. 747-757.
- Madsen, H. & Rosbjerg, D. 1997, "The partial duration series method in regional index-flood modelling", *Water Resources Research*, vol. 33, no. 4, pp. 737-746.
- Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P. & Stouffer, R.J. 2008, "CLIMATE CHANGE: Stationarity Is Dead: Whither Water Management?", *Science*, vol. 319, no. 5863, pp. 573-574.
- Montanari, A., Rosso, R. & Taqqu, M.S. 1997, "Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation", *Water Resources Research*, vol. 33, no. 5, pp. 1035-1044.
- Morgan, M.M. & Grimshaw, A.S. 2007, "Genesis II-Standards Based Grid Computing", IEEE Computer Society Washington, DC, USA, , pp. 611.
- Myers, R.H. & Montgomery, D.C. 1995, *Response surface methodology : process and product optimization using designed experiments*, Wiley, New York.
- Obeysekera, J.T.B. & Salas, J.D. 1986, "Modeling of Aggregated Hydrologic Time Series", *Journal of Hydrology JHYDA 7*, vol. 86, no. 3/4.
- Rasmussen, P.W., Heisey, D.M., Nordheim, E.V. & Frost, T.M. 1993, "Time-Series Intervention Analysis: Unreplicated Large-Scale Experiments" in *Design and Analysis of Ecological Experiments*, eds. S.M. Scheiner & J. Gurevitch, Chapman and Hall, New York, pp. 138-158.
- Rasmussen, P.F. & Rosbjerg, D. 1991, "Application of Bayesian principles in regional flood frequency estimation", , pp. 65.
- Rosbjerg, D., Funder Rasmussen, P. & Madsen, H. 1991, "Modelling of exceedances in partial duration series", *Natl Conf Publ Inst Eng Aust, Ie Aust, Barton*, vol. 3, no. 91 pt 22, pp. 755-760.
- Rosbjerg, D., Madsen, H. & Rasmussen, P.F. 1992, "Prediction in Partial Duration Series with Generalized Pareto-Distributed Exceedances", *Water Resources Research WRETAQ*, vol. p 3001-3010, pp. append.
- Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Menzel, A.,

- Root, T.L., Estrella, N. & Seguin, B. 2008, "Attributing physical and biological impacts to anthropogenic climate change", *Nature*, vol. 453, no. 7193, pp. 353-357.
- Ross, S.M. 2006, *Introduction to Probability Models*, Academic Press.
- Ross, S.M. 1993, *Introduction to probability models*, 5th edn, Academic Press, Boston.
- Salas, J.D. & Obeysekera, J.T.B. 1992, "Conceptual basis of seasonal streamflow time series models", *Journal of Hydraulic Engineering*, vol. 118, no. 8, pp. 1186-1194.
- SAS Institute Inc. 2000-2004, *SAS 9.1.3 Help and Documentation*, SAS Institute Inc., Cary, NC.
- Scipione, C. & Berliner, L. 1992, "Bayesian statistical inference in nonlinear dynamical systems", *Proceedings of the Bayesian Section of the American Statistical Association, American Statistical Association, Washington, DC*, .
- Scott, D.W. 1992, *Multivariate density estimation: theory, practice, and visualization*, Wiley-Interscience, New York.
- Sickle, J.V., Baker, J., Simonin, H., Baldigo, B., Kretser, W. & Sharpe, W. 1996, "Episodic Acidification of Small Streams in the Northeastern United States: Fish Mortality in Field Bioassays", *Ecological Applications*, vol. 6, no. 2, pp. 408-421.
- Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. 2005, "ROCR: visualizing classifier performance in R", *Bioinformatics*, vol. 21, no. 20, pp. 3940-3941.
- Smith, R.L., Tawn, J.A. & Coles, S.G. 1997, "Markov chain models for threshold exceedances", *Biometrika*, vol. 84, no. 2, pp. 249-268.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. 2002, "Bayesian measures of model complexity and fit (with discussion)", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 64, pp. 583-639.
- Spiegelhalter, D.J., Thomas, A., Best, N.G. & Gilks, W.R. 1994, "BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.30", *Cambridge: Medical Research Council Biostatistics Unit*, .
- Stedinger, J., Lettenmaier, D. & Vogel, R. 1985, "Multisite ARMA(1, 1) and Disaggregation Models for Annual Streamflow Generation", *Water Resources Research*, vol. 21, no. 4.
- Szilagyi, J., Balint, G. & Csik, A. 2006, "Hybrid, Markov Chain-Based Model for Daily Streamflow Generation at Multiple Catchment Sites", *Journal of Hydrologic*

Engineering, vol. 11, pp. 245.

Thomas, A., O Hara, B., Ligges, U. & Sturtz, S. 2006, "Making BUGS Open", *R News*, vol. 6, no. 1, pp. 12-17.

Vecchia, A.V. 1985, "Maximum Likelihood Estimation for Periodic Autoregressive Moving Average Models", *Technometrics*, vol. 27, no. 4, pp. 375-384.

Vecchia, A.V., Obeysekera, J.T.B., Salas, J.D. & Boes, D.C. 1983, "Aggregation and Estimation for Low-Order Periodic ARMA Models", *Water Resources Research*, vol. 19, no. 5.

Wan, E. & Van Der Merwe, R. 2000, "The unscented Kalman filter for nonlinear estimation", *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000.AS-SPCC.The IEEE 2000*, , pp. 153-158.

Wang, Q. 1991, "POT model described by the generalized Pareto distribution with Poisson arrival rate", *Journal of Hydrology(Amsterdam)*, vol. 129, no. 1, pp. 263-280.

West, M. & Harrison, J. 1997, *Bayesian forecasting and dynamic models*, 2nd edn, Springer, New York.

Whitehead, P., Beck, B. & O'Connell, E. 1981, "A Systems Model of Streamflow and Water Quality in the Bedford Ouse River System-- II. Water Quality Modelling", *Water research*, vol. 15, no. 10.

Young, P.C. 1999, "Nonstationary time series analysis and forecasting", *Progress in Environmental Science*, vol. 1, no. 1, pp. 3-48.

Zhang, W. & Arhonditsis, G.B. 2008, "Predicting the frequency of water quality standard violations using Bayesian calibration of eutrophication models", *Journal of Great Lakes Research*, vol. 34, no. 4, pp. 698-720.

9 Appendix A

9.1 Single-process CTMC – Model #1 (homogeneous)

```
model {
  for (i in 1:nObs) {
    stateEnd[i] ~ dbern(p[i])
    p[i] <- P1*F[i] + (1-F[i])*stateBeg[i]
    F[i] <- 1 - exp(-(lambda+mu)*delta_t[i])
  }
  P1 <- lambda/(lambda+mu)
  mu ~ dunif(0.01,10000)
  lambda ~ dunif(0.01,10000)
}
```

9.2 Single-process CTMC with varying rate – Model #2 (non-homogeneous)

```

model {
  for (i in 1:nObs) {
    stateEnd[i] ~ dbern(p[i])
    p[i] <- P1*F[i] + (1-F[i])*stateBeg[i]
    F[i] <- 1 - exp(-(1+K)*lambda.t[i]*delta_t[i])
    lambda.t[i] <- lambda + A*sin(theta[i]) + B*cos(theta[i])
  }
  P1 <- 1/(1+K)
  mu <- K*lambda
  K ~ dlnorm(0, 1)
  lambda ~ dunif(0.01,10000)
  ubdA <- lambda
  lbdA <- -ubdA
  pA ~ dunif(0, 1)
  A <- lbdA + pA*(ubdA-lbdA)
  ubdB <- sqrt(pow(ubdA,2) - pow(A,2))
  lbdB <- -ubdB
  pB ~ dunif(0, 1)
  B <- lbdB + pB*(ubdB-lbdB)
}

```

9.3 Single-process CTMC with varying limiting probabilities -

Model #3 (non-homogeneous)

```

model {
  for (i in 1:nObs) {
    stateEnd[i] ~ dbern(p[i])
    p[i] <- P1.t[i]*F[i] + (1-F[i])*stateBeg[i]
    F[i] <- 1 - exp(-(lambda[i]+mu[i])*delta_t[i])
    mu[i] <- gam/P1.t[i]
    lambda[i] <- gam/(1-P1.t[i])
    P1.t[i] <- P1 + A*sin(theta[i]) + B*cos(theta[i])
  }
  gam ~ dunif(0.01,10000)
  P1 ~ dbeta(0.999,0.999)
  P0 <- 1 - P1
  ubdA <- min(P0, P1)
  lbdA <- -ubdA
  pA ~ dunif(0, 1)
  A <- lbdA + pA*(ubdA-lbdA)
  ubdB <- sqrt(pow(ubdA,2) - pow(A,2))
  lbdB <- -ubdB
  pB ~ dunif(0, 1)
  B <- lbdB + pB*(ubdB-lbdB)
}

```

9.4 Multiple-process CTMC – Model #1 (homogeneous), simple version

```
model {
  for (i in 1:nObs) {
    stateEnd[i] ~ dbern(p[i])
    p[i] <- P1s[node[i]]*F[i] + (1-F[i])*stateBeg[i]
    F[i] <- 1 - exp(-(lambdas[node[i]]+mus[node[i]])*delta_t[i])
  }
  for (j in 1:J) {
    P1s[j] <- lambdas[j]/(lambdas[j]+mus[j])
    mus[j] ~ dunif(0.01,10000)
    lambdas[j] ~ dunif(0.01,10000)
  }
}
```

9.5 Multiple-process CTMC Model #1 (homogeneous), hierarchical version

```

model {
  for (i in 1:nObs) {
    stateEnd[i] ~ dbern(p[i])
    p[i] <- P1s[node[i]]*F[i] + (1-F[i])*stateBeg[i]
    F[i] <- 1 - exp(-(lambdas[node[i]]+mus[node[i]])*delta_t[i])
  }
  for (j in 1:J) {
    P1s[j] <- max(0, min(1, p1.raw[j]))
    lambdas[j] <- Cs[j] * P1s[j]
    mus[j] <- Cs[j] * (1 - P1s[j])

    logit(p1.raw[j]) <- b.0[j]
    b.0[j] ~ dnorm(b.0.hat[j], tau.b.0)I(-10,10)
    b.0.hat[j] ~ dnorm(mu.b, tau.b)

    log(Cs[j]) <- b.Cs.0[j]
    b.Cs.0[j] ~ dnorm(b.Cs.0.hat[j], tau.b.Cs.0)I(-10,10)
    b.Cs.0.hat[j] ~ dnorm(mu.Cs, tau.Cs)
  }
  tau.b.0 ~ dgamma(0.5, 0.5)
  tau.b.Cs.0 ~ dgamma(0.5, 0.5)
  mu.b ~ dnorm(0, 0.0001)
  tau.b ~ dgamma(0.5, 0.5)
  mu.Cs ~ dnorm(0, 0.0001)
  tau.Cs ~ dgamma(0.5, 0.5)
}

```

9.6 Multiple-process CTMC Model #2 (non-homogeneous)

```

model {
  for (i in 1:nObs) {
    stateEnd[i] ~ dbern(p[i])
    p[i] <- P1s[node[i]]*F[i] + (1-F[i])*stateBeg[i]
    F[i] <- 1 - exp(-(1+Ks[node[i]])*lambda.t[i]*delta_t[i])
    lambda.t[i] <- lambdas[node[i]] + As[node[i]]*sin(theta[i])
      + Bs[node[i]]*cos(theta[i])
  }
  for (j in 1:J) {
    P1s[j] <- max(0, min(1, p1.raw[j]))
    Ks[j] <- (1-P1s[j])/P1s[j]
    lambdas[j] <- Cs[j] * P1s[j]

    logit(p1.raw[j]) <- b.0[j]
    b.0[j] ~ dnorm(b.0.hat[j], tau.b.0)I(-10,10)
    b.0.hat[j] ~ dnorm(mu.b, tau.b)

    log(Cs[j]) <- b.Cs.0[j]
    b.Cs.0[j] ~ dnorm(b.Cs.0.hat[j], tau.b.Cs.0)I(-10,10)
    b.Cs.0.hat[j] ~ dnorm(mu.Cs, tau.Cs)

    r[j] ~ dunif(0,1)
    As[j] <- r[j] * lambdas[j] * cos(omega)
    Bs[j] <- r[j] * lambdas[j] * sin(omega)
  }
  tau.b.0 ~ dgamma(0.5, 0.5)
  tau.b.Cs.0 ~ dgamma(0.5, 0.5)
  mu.b ~ dnorm(0, 0.0001)
  tau.b ~ dgamma(0.5, 0.5)
  mu.Cs ~ dnorm(0, 0.0001)
  tau.Cs ~ dgamma(0.5, 0.5)
  omega ~ dunif(-3.141592594, 3.141592594)
}

```

9.7 Multiple-process CTMC Model #3 (non-homogeneous)

```

model {
  for (i in 1:nObs) {
    stateEnd[i] ~ dbern(p[i])
    p[i] <- P1.t[i]*F[i] + (1-F[i])*stateBeg[i]
    F[i] <- 1 - exp(-(lambda[i]+mu[i])*delta_t[i])
    mu[i] <- gams[node[i]]/P1.t[i]
    lambda[i] <- gams[node[i]]/(1-P1.t[i])
    P1.t[i] <- P1s[node[i]] + As[node[i]]*sin(theta[i])
      + Bs[node[i]]*cos(theta[i])
  }
  for (j in 1:J) {
    gams[j] <- Cs[j]*P1s[j]*(1-P1s[j])
    P1s[j] <- max(0, min(1, p1.raw[j]))
    P0[j] <- 1 - P1s[j]

    log(Cs[j]) <- b.Cs.0[j]
    b.Cs.0[j] ~ dnorm(b.Cs.0.hat[j], tau.b.Cs.0)I(-10,10)
    b.Cs.0.hat[j] ~ dnorm(mu.Cs, tau.Cs)

    logit(p1.raw[j]) <- b.0[j]
    b.0[j] ~ dnorm(b.0.hat[j], tau.b.0)I(-10,10)
    b.0.hat[j] ~ dnorm(mu.b, tau.b)

    maxamp[j] <- 0.5 - sqrt(pow(P1s[j]-0.5,2))
    r[j] ~ dunif(0,1)
    As[j] <- r[j] * maxamp[j] * cos(omega)
    Bs[j] <- r[j] * maxamp[j] * sin(omega)
  }
  tau.b.0 ~ dgamma(0.5, 0.5)
  tau.b.Cs.0 ~ dgamma(0.5, 0.5)
  mu.b ~ dnorm(0, 0.0001)
  tau.b ~ dgamma(0.5, 0.5)
  mu.Cs ~ dnorm(0, 0.0001)
  tau.Cs ~ dgamma(0.5, 0.5)
  omega ~ dunif(-3.141592594, 3.141592594)
}

```