BAYESIAN INFERENCE: APPLICATION TO ENVIRONMENTAL MANAGEMENT AND DECISION-MAKING

FINAL REPORT

Scientific Advisory Board - Ecological Processes Standing Committee (EPSC)

Chair – Dr. Michael P. Weinstein, New Jersey Institute of Technology
Dr. Carolyn Bentivegna, Seton Hall University
Mr. Paul Bovitz, Weston Solutions
Mr. Charles R. Harman, AMEC Environment & Infrastructure
Dr. Robert A. Hoke, DuPont, Haskell Global Centers
Dr. Ursula Howson, Monmouth University
Dr. Jonathan Kennen, US Geological Survey
Dr. Zeyuan Qiu, New Jersey Institute of Technology

NJDEP Partners

Mr. Joseph Bilinski
Dr. Mihaela Enache
Dr. Robert Hazen
Dr. Leo Korn
Dr. Nicholas A. Procopio

A Report to the Scientific Advisory Board (SAB), New Jersey Department of Environmental Protection

25 April 2014

ECOLOGICAL PROCESSES STANDING COMMITTEE CONTACT INFORMATION

Carolyn S. Bentivegna, PhD
Seton Hall University
South Orange, NJ
973-275-2113
carolyn.bentivegna@shu.edu

Michael P. Weinstein, PhD (Chair)
New Jersey Institute of Technology
Newark, NJ 07043
(973) 467-1737
mweinstein_fishguy@verizon.net

Paul Bovitz
Weston Solutions, Inc
205 Campus Drive
Edison, NJ 08837
(732) 417-5815
Paul.Bovitz@WestonSolution.com

Charles R. Harman, PWS
AMEC Environment & Infrastructure
Somerset, NJ 08873
(732) 302-9500, x 27
charles.harman@amec.com

Robert A. Hoke, PhD
DuPont
Haskell Global Centers for Health and Environ Sci
Newark, DE 19711
302-451-4566
robert.a.hoke@usa.dupont.com

Ursula Howson, PhD
Department of Biology
Monmouth University
West Long Branch, NJ 07764
(732) 571-4432 (v)
uhowson@monmouth.edu

Jonathan G. Kennen, PhD
US. Geological Survey
354 Princeton Pike, Suite 110
Lawrenceville, New Jersey 08628
(609) 771-3948
jgkennen@usgs.gov

Zeyuan Qiu, PhD
New Jersey Institute of Technology
Newark, NJ 070102
(973) 596-5357
Zeyuan.qiu@njit.edu

PREFACE

No one on the Ecological Processes Standing Committee of the New Jersey Department of Environment Protection's Scientific Advisory Board (SAB) is a professional statistician, nor intimately knowledgeable of Bayesian methods. Consequently, there may be pitfalls when preparing a report on Bayesian inference, such as was undertaken here:

Unlike many professional statisticians, many empirical scientists are not able to use the method of analysis that best fits their particular problem. In addition, their understanding of [frequentist] *statistics in which they have been trained has been widely found to be rather abysmal. As John Tukey* [we should all recognize this name] *said in 1964, "Most uses of the classical tools of statistics have been, are, and will be, made by those who know not what they do". The situation has led many to argue for an educational reform in statistical training for empirical scientists, and for increased emphasis on translating between frequentist and Bayesian measures of evidence ... the latter is proving to be particularly useful in many fields ... Reasons for this increased popularity in Bayesian method are not hard to spot. Much of modern research, particularly in the life sciences, is based on the synthesis of multiple categories of evidence. Data coming from many different studies have to be integrated in order to assess the empirical evidence for a new theory, and Bayesian statistics lends itself very well to this.*

Robert van Hulst 2013

The EPSC is not faint-hearted, and has eagerly accepted the inherent challenges of preparing this report. We approached it as we would any scientific endeavor; i.e., to learn as much as possible about the subject and its methods before integrating the information across disciplines, To validate our results, we also enlisted the editorial assistance of several professional statisticians and empirical scientists who were well-acquainted with Bayesian methods. Any factual errors herein, however, are solely those of the EPSC.

ACKNOWLEDGEMENTS

The members of the Ecological Processes Standing Committee (EPSC) thank NJDEP staff and Dr. Ed Green (Rutgers University) for their presentations on various subjects related to *Bayesian Inference: Application to Environmental Management and Decision-Making*. A comprehensive library of relevant publications was made available to the Standing Committee by Dr. Nick Procopio, who also provided valuable review comments on the various drafts of the report. Ed Green and Peter Morin (Rutgers University) also reviewed a final draft of the report. We thank the USGS for logistical support at their Ewing and Lawrenceville facilities.

Sections of this report were prepared by: M.P. Weinstein, Z. Qiu; J. Kennen and C. Harman. Dr. Leo Korn kindly provided the text for Box 2. Editorial review was provided by: C. Bentivegna, P. Bovitz, E. Green, L. Korn, P. Morin, and N. Procopio

EXECUTIVE SUMMARY

Implementation of GIS based Bayesian inference into New Jersey Department of Environmental Protection's (NJDEP) environmental monitoring "toolkit" has the potential to substantially improve monitoring efficiency and reduce long-term monitoring costs while simultaneously improving prediction capacity. In this context, the Scientific Advisory Board, Ecological Processes Standing Committee (EPSC) was tasked with addressing the overarching question: should the NJDEP evaluate and/or test Bayesian-based statistical analysis methods with the potential to substantially improve monitoring efficiency and outcomes in the near future? If so, what alternative methods should be evaluated?

Two elements distinguish the Bayesian approach; first is the quantification of prior beliefs about a parameter in the form of a probability distribution, and the use of those beliefs in the actual data analysis; and second, the acceptance of the "likelihood principle" and the concomitant rejection of all sample-space probabilities from inferential conclusions about the parameter. Bayesian inference differs from classical, or frequentist inference in four general ways:

1. Frequentists estimate the probability of the data (B) having occurred given a specific hypothesis (P[B|H]), whereas Bayesian inference provides a quantitative measure of probability of a hypothesis being true in light of the available data (P[H|B]);
2. Frequentist inference defines probability in terms of infinite relative frequencies of events, whereas Bayesian inference defines it as the *individual's* degree of belief in the likelihood of an event;
3. Bayesian inference uses *prior knowledge* along with the sample data while frequentist inferences uses *only the sample data*; and
4. Bayesian inference treats model parameters as *random variables* whereas frequentist inference considers them to be estimates of *true fixed values*.

Several Bayesian analytical approaches were evaluated: (1) Bayesian *hierarchical modeling* that can be used to predict and causal inferences from experiments and observational studies that display multi-level structure; (2) Bayesian *networks* that are a suitable means for performing integrated ecological modeling; and, (3) Bayesian *retrospective analysis*, a procedure that aims to economize the long term costs of maintaining a network of monitoring site while minimizing the loss of information. The EPSC also presented a case study to introduce *kriging*, by itself not a Bayesian method, but with a Bayesian variation, can extend the method to analyze the performance of the underlying kriging predictor. In each instance, the EPSC presented case studies outlining the issues, the results, and the conclusions in applied Bayesian approaches. Finally, the EPSC provided a brief summary of available continuous or near-continuous data gathering devices in the region (Appendix II).

Several advantages of Bayesian inference were identified:

- Combined with the complexity inherent in most ecosystems, and the severity of environmental issues confronting managers and decision makers, many agencies and organizations have sought to explore new spatial analytical techniques that provide

timely, valid information to assist problem solving, and effective environmental management decisions. The Bayesian approach has the potential to do just this.

- The essential approach of the Bayesian method is to address the question: can we get better estimates of the mean by collecting additional data from populations, other than just a single or a few population? For both the Bayesian and empirical Bayesian, the answer is yes.
- Bayesian methods have the potential to substantially streamline data gathering and consequently reduce monitoring costs;
- A common issue with frequentist statistical approaches are related to weak experimental design in terms of replication, spatial and temporal confounding issues, and the nature of the 'treatment' itself. Even with a 'gradient-based' approach that stratifies data collection into separate categories cannot fully overcome the replication problem, and it is usually financially or physically prohibitive to sample with sufficient replication to detect significant differences among test variables. Bayesian methods can mitigate some of these difficulties because the approach is "inherently flexible", i.e., models are constructed to conform to the requirements of the data, whereas standard statistical approaches must force the data to comply with the requirements of a relatively small number of model types. Bayesian models that help obviate the problem of replication and are finding increasing uses in ecological applications. They appear particularly suited to dealing with the complexities of spatiotemporal variation in ecology, and allow for the construct of "far more complex models" than is possible with traditional statistical approaches; and
- Unlike most integrated modeling exercises, Bayesian networks are probabilistic, rather than deterministic, expressions to describe relationships among variables. This is an essential and desirable characteristic of an ecosystem model if predictions are to guide decision making. The Bayesian network approach has proven to be a suitable means for performing integrated ecological modeling because their graphical structure explicitly represents cause and effect assumptions among system variables that might not be tractable with alternative modeling approaches such as deterministic point estimate modeling. In general, Bayesian networks are not sensitive to imprecision in the input probabilities and can, therefore, be classified as "robust tools".

In this review, we do not mean to imply that Bayesian methods are not without disadvantage (as are most other methods). Among these are 1) computational challenges, even currently available software (Appendix III) may be difficult to use; and 2) the requirement to condition the hypothesis on the data; and the potential lack of objectivity therein, because different results will be obtained using different priors. However, as discussed in the Introduction to this report, while noting that the primary criticisms of the Bayesian approach stems from "overenthusiastic" application of "uninformative priors", the EPSC agrees with the general Bayesian community of practitioners that current data gathering practices using modern instruments produces voluminous data that not improve the ability of Bayesian statistics to connect disparate inferences, but also ensures that "genuine" informed priors become the rule rather than the exception.

We concluded that Bayesian methods are a potentially powerful tool for statistical analysis of complex data sets. The use of priors is a positive action in that it allows for prior knowledge or

perspectives to inform a current model. Additionally, Bayesian inference includes uncertainty in the probability model which should yield more realistic predictions. Bayesians employ only observed data, as opposed to frequentists who use both observed and hypothetical data. Bayesian networks are probabilistic rather than deterministic expressions to describe relationships among variables. Bayesians suggest that this is an essential and desirable characteristic of an ecosystem model if predictions are to guide decision making.

COMMITTEE RECOMMENDATIONS

Based on our findings regarding Bayesian statistics as outlined in this report, the EPSC has compiled the following recommendations:

- Because a comparison and evaluation of how other states and entities use Bayesian methods is well beyond the scope of this report, the EPSC recommends that NJDEP adopt a two phase approach to the question; (1) develop a survey instrument to ascertain what other states and entities are doing in this arena; and (2) depending on the outcome of step (1), convene a workshop of technical personnel from selected state and federal resource agencies (USGS, NOAA, and USEPA), and selected academic institutions, to address the general theme: *The Use of Bayesian Inference to Address Environmental Monitoring and Management Challenges.*
- DEP should work with colleges and universities to develop a one-week continuing education curriculum in applied aspects of Bayesian-inference for NJDEP scientists. The curriculum should be relevant to statewide monitoring programs and should be compatible with ArcGIS programs.
- NJDEP, in conference with their in-house and state university statisticians, geo-spatial modelers, and ecologists should identify a "training data set" from their vast monitoring programs to compare model characteristics and output (robustness and efficiency, potential bias and flexibility) and performance capacity among frequentist and Bayesian methods.
- Similarly, and with the same approach, conduct a "sensitivity analysis" on existing NJDEP monitoring data sets using retrospective analysis to examine the relationships among sampling locations, sampling frequency, and resource allocation, to enhance the quality of information produced; e.g., by using real-time, remotely collected data from the Department's data logger array.
- The Department should make its existing library on Bayesian literature available to the user community upon request.
- The Department should issue a request for proposals (RFP) to academic institutions in New Jersey for the study of practical applications of Bayesian methods that address state environmental management and ecological issues.
- Promote the practical application of Bayesian inference as an additional, oft desirable, tool in the Department's analytical toolkit.
- Use Bayesian inference, and the content of this report, to encourage the broader use of statistics in the Department's development of study designs and quantitative data analysis in fulfilling its regulatory mandates.

1. INTRODUCTION

1.1 Ecosystem Complexity, Forecasting, Informed Decision-Making and Bayesian Inference

> *It is increasingly clear that much of the workings of the world, and the challenges and opportunities these workings entail for a transition to sustainability lie in the interactions among environmental issues and human activities that have previously been treated as largely separate and distinct…**in the next decade we will see research/[education] and problem-solving shift in focus from single issues to multiple interacting stresses***
>
> US National Research Council 2002

The above quote is a striking parallel and foretelling of the need for this report:

> *Reasons for [the] increased popularity in Bayesian method are not hard to spot. Much of modern research, particularly in the life sciences, is based on the synthesis of multiple categories of evidence. Data coming from many different studies have to be integrated in order to assess the empirical evidence for a new theory, and Bayesian statistics lends itself very well to this.*
>
> Robert van Hulst 2013

For scientists to truly inform policy, they must provide predictive links between management actions and ecosystem responses (Borsuk *et al.* 2004). Not only does dealing with environmental change rest with a capacity to anticipate and prepare for an uncertain future (Clark *et al.* 2001), but reducing uncertainty is a necessary prerequisite to making forecasts that provide *useful* information. Moreover, experimental and observational data that extend to landscapes or regions, and sustained monitoring are a foundation for forecasting sustainable ecosystems. A relevant example for the New Jersey Department of Environmental Protection (NJDEP) is the relationship between broad-scale habitat loss and predictions of extinction risk (Clark *et al.* 2001). The use of geospatial statistics and methods at the landscape scale can contribute to reducing uncertainty in the decision-making process, and is the framework for this exercise. But the discipline is vast, and a comprehensive treatment of the subject is well beyond the scope of this report. The NJDEP has requested that the emphasis of this effort be focused on Bayesian statistical inference (and allied methods) and their utility in enhancing state-wide monitoring efforts designed to inform management decisions. It should be noted at the outset that Bayesian inference provides a natural framework for the inclusion of parameter uncertainty in spatial prediction (Diggle and Lophaven 2006).

Bayes Theorem

First introduced in 1763, Bayes Theorem is an algorithm for combining *prior experience* with current evidence (Bayes 1763; Gotelli and Ellison 2004; Efron 2013). When scientists want to estimate the probability of an event using non-Bayesian methods they often begin by assuming no prior knowledge of that probability, and estimate it by conducting a large number of

experimental and/or observational trials. In contrast the Bayesian approach builds on the notion that investigators have a predisposed belief [1] of the probability of an event (even before any trials are conducted) and that these *prior probabilities* may be based on previous experience, intuition or model predictions (Gotelli and Ellison 2004).

Priors may be *informative* or *uninformative*. The former expresses specific, definitive information about a variable; e.g., there are 12 eggs in a dozen. An uninformative prior, on the other hand, expresses only general information about a variable; e.g., the variable is positive, or it is less than some value. Efron (2012) cautions that most criticisms of the of the Bayesian approach stem from overenthusiastic application of "uninformative priors", and suggested that the latter be employed parsimoniously. But for introductory purposes, we can note that current data gathering practices using modern instruments "pumps out results in fire hose quantities" (Efron 2012), producing prodigious data that "bear on complex webs of interrelated questions". Efron (2012) suggests further that "in this new scientific era, the ability of Bayesian statistics to connect disparate inferences counts heavily in its favor", and in general terms, ensures that "genuine" informed priors become the rule rather than the exception. The author contends further, that we can not only estimate the relevant priors directly from the data, but that this *empirical Bayes* approach (Box 1) derives from experiments involving a large number of parallel situations carrying within them their own prior distributions (Robbins 1955). In the broadest sense, the primary goal of the method is to use a Bayesian approach without necessarily *fully* specifying the prior, either by estimating it or by its parameters (ver Hoef 1996).

Box 1

The term "empirical Bayes" was coined by Robbins (1955). Broadly, the goal is to use Bayesian methods *without fully specifying the prior*, either by estimating the prior or its parameters. For example, say we want to monitor trends in the abundance of individuals in a population over several years by drawing a random sample of individuals from that population, and multiple replicate samples are taken in a given year. Then sampling theory dictates that μ, the sample mean, has an associated estimate of variance, or dispersion around that value. The essential approach of the method is to address the question: can we get better estimates of the mean by collecting additional data from those populations, other than just the *ith* one? For both the Bayesian and empirical Bayesian, the answer is yes.

To conclude, Howson and Urbach (1991) suggested that balancing empirical and prior factor outcomes becomes a necessary prerequisite to evaluating competing hypotheses. Although neither factor by itself is decisive, the authors recommend that the prior plausibility always be checked against the empirical test results as a standard procedure.

1.2 Statistical Analysis

> *...to acknowledge the subjectivity inherent in the interpretation of data is to recognize the central role of statistical analysis as a formal mechanism by which new evidence can be integrated with existing knowledge...*
>
> Berger and Berry 1988

[1] Prior beliefs generally represent some amalgamation of information that is available before data collection (Wolfson *et al.* 1996).

Statistical analysis generally attempts to give "objectivity" to conclusions derived from a set of experimental data. But, reaching sensible conclusions from analysis of these same data may require, and in fact most often does, *subjective* input (Berger and Berry 1988). The authors argue that not only are standard statistical methods based on subjective input, but that the source of subjectivity depends on the intentions of the investigator. Probability (p) values are a particular case in point; e.g., the probability (provided the null hypothesis is true) that the test statistic would have been more extreme than the actual observed value of the statistic. In other words, a p value includes the probability of data that *didn't occur* (i.e., other sample-space[2] probabilities). Thus, subjectivity often arises from the producer (the scientist doing the study), rather than the consumer (others in the community of scientists, decision makers, etc.) (Berger and Berry 1988).

There are two fundamental approaches for dealing with subjectivity. Classical statisticians, or "frequentists" [3], consider the thoughts of the investigator about data that might have been observed, but were not (once again, the data that didn't occur!), *as being relevant*. In an alternative paradigm, "Bayesian" statisticians consider that only the actual data are relevant to the inferences drawn from an experiment. Moreover, Bayesian statistics may be used to compare alternative hypotheses, or models, in a single framework (Carpenter 1990). Such analysis employs a data set (Y) and a set of candidate models (M_i) chosen to represent distinct alternative explanations, mechanisms, or policy options (Walters 1986; Carpenter 1990). Prior to analysis, the investigator assigns a prior probability P_i that each model is correct.

The use of Bayesian methods, and its potential to reduce monitoring costs, is the subject of this report, specifically how the Bayesian approach might be incorporated into spatial and other analyses used by the NJDEP. In this framework, the Ecological Processes Standing Committee (EPSC) was tasked with addressing the following questions:

1.3 General Questions:

Should the NJDEP evaluate and/or test new, Bayesian-based, statistical analysis methods with the potential to substantially improve monitoring efficiency and outcomes in the near future? If so, what alternatives should be evaluated?

1.4 Specific Questions:

1. Implementation of GIS based Bayesian methods into DEP's monitoring "toolkit" could substantially improve monitoring efficiency and potentially reduce monitoring costs while simultaneously improving prediction capacity. Is there sufficient scientific consensus on the reliability of Bayesian modeling approaches? What research has been conducted to compare results calculated from standard methods with those produced using Bayesian models? Are other methods available and recommended for evaluation?

---

[2] The sample space, denoted by Ω, is the set of all possible outcomes in an experiment. The frequentist paradigm estimates probabilities as the relative frequency of outcomes based on *an infinitely large number of trials*.

[3] *Frequentist*, or classical inference assumes that there is a *true*, *fixed value* for each variable of interest (e.g., density of individuals in cities), and that the expected value of this parameter is an average value derived from repeated *random* sampling (Ellison 1996).

2. Have other States or entities evaluated and employed similar Bayesian modeling techniques to improve monitoring outcomes and potentially reduce costs? What NJDEP programs would benefit most from the incorporation of such advanced techniques?

3. Data loggers and other real time monitoring devices (Appendix I) are becoming more accurate and affordable. Data loggers can reduce sampling time and potentially increase data accuracy and reliability. However, depending on the frequency of data collection, much more data can be collected than should be analyzed using normal statistical methods. What types of data reduction techniques are necessary in combination with the increased use of deployable long-term data loggers in order to maintain adherence to appropriate statistical assumptions of independence and autocorrelation?

Three related **Tasks** have been identified for this effort:

- Conduct a literature review that describes the "state of the science" for Bayesian methods; how is it done, strengths and weaknesses, modeling approaches used, statistical analyses used, etc. NJDEP has conducted a first round literature search and the materials have been posted on their web-site. We are being asked to review this literature and prepare a 'primer' on the topic (**Task 1**);
- Survey who is doing what and where; locally, regionally and perhaps even globally if there are sterling examples available establish 'several case studies'; and prepare a summary document (**Task 2**);
- How might new technology; e.g., data loggers, be incorporated into the mix?
- Rather than going out and compiling model data sets from existing programs for future "ground-truthing" and validation (and to develop a frame of reference for comparison to existing approaches and improving them - both on a scientific and/or cost effectiveness basis), NJDEP prefers that we start with published case studies and synthesize them as a surrogate 'validation' step (**Task 3**).

2. COMPARING FREQUENTIST VERSUS BAYESIAN INFERENCE

> *In all of the sciences, statistics is the common language used to interpret our measurements and to test and discriminate among our hypotheses… probability*[4] *is the foundation of statistics*
>
> Gotelli and Ellison 2004

The frequentist paradigm (Box 2) estimates probabilities as the relative frequency of outcomes based on an infinitely large number of trials. Thus, to estimate the probability of a given phenomenon (e.g., births per 100 females in the population), frequentists begin by *assuming no prior knowledge of the probability of an event*, and then establish that probability on the basis of a large number of samples. In contrast, Bayesian inference[5] is underpinned by a formula for conditional probability based on *prior knowledge* (experience) of the likelihood of an outcome. These *prior probabilities*, as they are known, may be based on previous experience, intuition, or

[4] "Probability" is the likely outcome of an event, a process with a well-defined beginning and end.
[5] Named after Thomas Bayes, best known for his 1763 essay, "Towards Solving a Problem in the Doctrine of Chances".

modeling predictions. As in classical statistics, prior probabilities are ultimately derived from trials and experiments (Gotelli and Ellison 2004).

Bayes' Theorem may be stated as:

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \qquad (1)$$

where the quantity $P(A/B)$ is the probability of A given event B. The term $P(A)$ is referred to as the *prior probability*; i.e., the investigator's subjective prior beliefs, or the probability expected *before* the experiment is conducted (Dennis 1996; Ellison 1996). It should be noted that this parameter is not a random variable; rather it is an *unknown* variable that is selected by the investigator as quantifying a "best guess", emphasizing the subjective nature of the approach. Ellison (1996) discusses three interpretations of the prior probability: (1) a frequency distribution whose parameters are based on the use of existing data; (2) an objective measure *of what one can believe* about the parameter or distribution of interest; (3) a subjective measure of *what the investigator actually believes.* Most frequentists will likely use the first two interpretations of $P(A)$ when framing hypotheses and designing experiments. The remaining term in the numerator, $P(B/A)$, or the probability of B given A, is described as the *likelihood function* for the parameter of interest (Box and Tiao 1973).

Box 2

A **probability measure** is defined on a sample space Ω. Any probability measure P(A) must satisfy three axioms: (1) If A is a subset of Ω, then P(A) is non-negative; (2) If two subsets of Ω are non-overlapping then $P(A+B)=P(A)+P(B)$; and (3) P(Ω) = 1. As an example, in a coin toss: Ω = {heads, tails} and P(heads) + P(tails) = P(head or tails) = 1. These three axioms may be used to determine other mathematical properties of probability measures.

In **classical probability** a particular probability measure is deduced through an exercise of reason. Example: *The probability of rolling a die and getting a 1 is 1/6 because there are six equally likely outcomes.*

Frequentist probability views a probability measure as a converging value of sample proportions as the sample size get very large. This view allows for the interpretation of sampling statistics as estimates of population statistics. Example: *As the die is rolled more and more times, the observed proportion of 1s converges to the probability of rolling a 1.*

Subjective probability is used primarily by Bayesians. Rather than describing the probability of a sample statistic, it describes an individual's assessment of an event as a prior probability. This allows one to consider a hypothesis to have a certain probability (or a probability distribution) of being correct; e.g., stating the hypothesis that *"the probability that this coin is fair is 0.8"*.

Both frequentist probabilities and subjective probabilities must satisfy the axioms of probability measures in order to avoid internal contradictions.

The Bayesian approach uses a step-wise procedure to calculate the probability that a hypothesis is true for a *given set* of data; for example in a coin tossing exercise (Appendix II). *The principal of including only the actual data in the analysis (P[B/A]) and excluding consideration of all other sample-space probabilities is referred to as the "likelihood principle".* The goal is to produce final probabilities for testing hypotheses, or what are commonly referred to as *posterior probabilities* that reflect how the investigator's subjective beliefs have been altered by the actual data (Dennis 1996). Berger and Berry (1998) note that obtaining the final probability of a hypothesis in light of the experimental data requires that the investigator first *specify the probabilities of the hypothesis before or apart from the experimental data*; these 'initial probabilities' are, as noted above, also called *prior probabilities*. It also identifies a "final" probability for the hypothesis being tested. In the end, scientists with different prior beliefs draw their own conclusions from

the data, using their own priors. Consensus emerges when most scientists' priors become swamped by large amounts of data; i.e., when their posterior beliefs become nearly identical (Dennis 1996).

Thus, there are two key elements that distinguish the Bayesian approach; first is the quantification of prior beliefs about a parameter in the form of a probability distribution, and the use of those beliefs in the actual data analysis; and second, the acceptance of the "likelihood principle" and the concomitant rejection of all sample-space probabilities from inferential conclusions about the parameter (Dennis 1996). Thus, Bayesian inference differs from classical, frequentist inference in four ways (Ellison 2004):

1. Frequentists estimate the probability of the data having occurred given a specific hypothesis (*P[B/H]*), whereas Bayesian inference provides a quantitative measure of probability of a hypothesis being true in light of the available data (*P[H/B*]);
2. Frequentist inference defines probability in terms of infinite relative frequencies of events, whereas Bayesian inference defines it as the *individual's* degree of belief in the likelihood of an event;
3. Bayesian inference uses *prior knowledge* along with the sample data while frequentist inferences uses *only the sample data*; and
4. Bayesian inference treats model parameters as *random variables* whereas frequentist inference considers them to be estimates of *true fixed values*.

As noted in the Introduction, the presence of the prior distribution is the source of much controversy in Bayesian modeling, i.e., it can be seen as making the analysis "overly subjective" (McCarthy 2007). However, as McCarthy (2007) also notes that "when little information exists concerning a parameter, one is able to assign a so-called minimally informative or 'vague' prior distribution." Such a prior has only slight effects on the posterior distribution "… [and] familiar analyses (e.g., ANOVA or regression) carried out using Bayesian methods and vague priors for the parameters (e.g., regression slope) will usually come up with a similar distribution for that parameter as the almost universally used frequentist model".

3. BAYESIAN STATISTICAL DESIGN

Spatially focused analytical procedures are essential tools for understanding the distribution and interactions of biota among themselves and with physico-chemical drivers in the environment. Combined with the complexity inherent in most ecosystems, and the severity of environmental issues confronting managers and decision makers, many agencies and organizations have sought to explore new spatial analytical techniques that provide timely, valid information to assist problem solving, and effective environmental management decisions (Little *et al.* 1997). The New Jersey Department of Environmental Protection is clearly interested in exploring and evaluating these methods as a potential addition to its analytical toolkit.

A central challenge in applying any of these methods, however, is the prediction of a spatial surface over a region using data that are imperfect measurements, but nonetheless assumed to be adequate estimates of a parameter at a limited number of sampling locations (Diggle and Lophaven 2005). In practice, several attributes of spatial variation must be recognized (Burrough 1995):

- Spatial variation of ecological attributes vary *continuously* within a spatial unit that cannot be described by simple regression polynomials;
- Rather, short range variation of an attribute, observed as a set of observation points, vary in a *correlated manner*, at least at the scale at which the observations have been made; i.e., *sample points closer together tend to be more similar (display spatial autocovariance) than points further apart*; and
- Statistical properties of spatially correlated variation are the *same, or uniform*, within the whole area of study; this is referred to as *statistical stationarity*.

Thus, models of spatial variation should contain at least three structural components; 1) the average value of the variable within a defined area; e.g., breast height diameter of trees in a forest stand; 2) spatially correlated, gradual variation; and 3) uncorrelated random variation.

In these circumstances a stochastic linear model is often applied, with the assumption that the spatial surface of interest is underpinned by an unobserved *stationary Gaussian process* consisting of random values associated with a range of time, space, or space/time such that each random variable has a *normal distribution*[6]**.** Finally, statistical stationarity, or *uniformity*, is a necessary prerequisite of the approach. The latter can be understood by considering a series of observation points laid out at equal intervals along a linear transect. Say that a soil property is estimated at each sampling point x. Formally, we can say that if the joint distribution of the n random variables $S(x_1)$ …. $S(x_n)$ is the same as the joint distribution of $S(x_1 + h)…S(x_n + h)$ for all $x_1 … x_n$. Put simply, the statistical properties of the series are not affected by moving the sample points a distance h from point x_i to point x_i+h. In practice, it is usual to replace the above definition of a stationary Gaussian process, with second-order or weak stationarity in which the mean is constant, the autocovariance depends only on distance among sampling locations, and the variance is finite and constant.

Bayesian inference, including methods for geostatistical analysis using the linear Gaussian model (e.g., Kitanidis 1988; Handcock and Stein 1993) offers a useful framework for including the effects of parameter uncertainty in spatial predictions. The method was extended by Diggle *et al.* (1998) who embraced generalized linear models with an unobserved Gaussian process in the linear predictor (Diggle and Lophaven 2005; see Case Study 2).

Other common issues with frequentist statistical approaches; e.g., those that might be used with hydrogeomorphic assessments[7] are related to weak experimental design in terms of replication, spatial and temporal confounding issues, and the nature of the 'treatment' itself (e.g., flow characteristics) (Webb *et al.* 2009). Even with a 'gradient-based' approach that stratifies flow according to stream-bed slope cannot fully overcome the replication problem noted, and it is usually financially or physically prohibitive to sample with sufficient replication to detect significant differences between flow and response (Webb *et al.* 2009). These authors suggested that Bayesian methods may be able to mitigate some of these difficulties because the approach is

[6] Very often environmental variables are not normally distributed, and it is customary to first examine the distribution of sampling variables to check for normality, and if the distributions are skewed, apply a transformation to normalize the data.

[7] The provision of environmental flows is critical to, for example, maintaining ecological integrity of regulated river systems where there are 'competing' flows for ecosystem and anthropogenic uses (e.g., agricultural uses).

"inherently flexible" such that models are constructed to conform to the requirements of the data, whereas "standard statistical approaches[8] must force the data to comply with the requirements of a relatively small number of model types" (see also McCarthy 2007). One type of Bayesian approach, i.e., the use of hierarchical models (see Section 4.1) may help obviate the problem of replication and are experiencing increased usage in ecological applications. They appear particularly well-suited to dealing with the complexities of spatiotemporal variation in ecology, and allow for the construct of "far more complex models" than is possible with traditional statistical approaches (Clark 2005).

4. BAYESIAN APPLICATIONS: SELECTED CASE STUDIES

- *In the interest of space, succinctness, and to "reinforce" the EPSC recommendations that follow at the end of these collective case studies, analytical findings will be presented in summary format. The basic idea is to provide sufficient information to demonstrate the utility of Bayesian methods, without making any value judgments vis-à-vis the subject(s) of the study.*

4.1 Case Study 1. Environmental Flows and the Bayesian Hierarchical Model

Title: Detecting ecological responses to flow variation using Bayesian hierarchical models (Webb *et al.* 2009)

Natural phenomena displaying hierarchical or multilevel structure commonly occur in environmental studies. Hierarchical modeling, a generalization of regression methods, can be used to predict and causal inferences from experiments and observational studies that display multilevel structure (Kreft and De Leeuw 1998; Snijders and Bosker 1999; Raudenbush and Bryk 2002; Hox 2002; Gelman 2006). The linear multilevel regression model, is an example of a hierarchical model,

$$y_{ij} = \beta_{0(0)} + \sum_l \beta_l x_{ij(l)} + \beta_{0j} + \sum_k \beta_{kj} x_{ijk} + \epsilon_{ij} \qquad (2)$$

where ij is the i^{th} observation of j^{th} group, $x_{ij(l)}$ for the global variables while x_{ijk} for the variables within each levels and ϵ_{ij} is the error term. β_l is used to show the global effect and β_{kj} for in-level effect, where both $\beta_{\cdot}s$ are parameters of the hierarchical model.

If we assume the parameters to be random and estimate a prior distribution $\pi(\lambda_{\cdot})$ for $\beta_{\cdot}$, then the effects of each variables is based on the posterior distribution $\pi(\lambda_{\cdot}|x, y)$ for $\beta_{\cdot}$ after the data $\{x, y\}$ are collected. There are two challenges to invoking a Bayesian solution to multilevel data: (1) estimating a satisfactory prior and (2) calculating a solution based on posterior distribution. For prior estimation, the "conjugate prior" is most popular solution because of simplicity of calculation (Raiffa and Schlaifer 1961); but an uninformative prior (Bernardo 1979; Berger

[8] Non-Bayesian, multilevel (hierarchical) modeling is also an increasingly popular approach to modeling hierarchically-structured data, generally outperforming classical regression in predictive accuracy. An important feature of multilevel models is their ability to separately estimate the predictive effects of an individual predictor and its group-level mean. These models, however, are not discussed further in this report.

1990; Lehmann 1998) can also be used when prior information is limited. Bayesian calculation based on the conjugate prior uses the kernel distribution[9] (Raiffa and Schlaifer 1961), otherwise the Markov Chain Monte Carlo (MCMC) simulation can be used (Asmussen and Glynn 2007). It should be noted that empirical Bayesian methods (Robbins 1985) may also be used (see Box 1).

In most cases, nonhierarchical models are inappropriate for hierarchical data analysis: with few parameter estimates they generally fit large datasets poorly, whereas when many parameter estimates are available, they tend to "over-fit" such data, leading to inferior predictions for the analysis of new data (Gelman *et al.* 2003). More importantly, the hierarchical approach tends to unify the seemingly disparate methods of frequentist and Bayesian analysis (Efron and Morris 1973; Greenland, 2000).

Bayesian hierarchical models have increasingly been applied to assess stream biological responses to landscape changes (Reckhow, 1996; Rechhow *et al*., 2009; Kashuba et al., 2009; Qian *et al*., 2010) because of its apparent advantages over other conventional statistical methods (e.g., Riva-Murray *et al*., 2010). Traditional regression methods assume that such relationships are constant across space and use global estimates that assess the average conditions for a study area. However, stream physical, chemical and biological conditions are not only affected by local factors such as land use intensity parameters and climate conditions, but they also vary across distinct physiographic provinces (Kennen, 1999) and streamflow regimes as emphasized by Kennen *et al.* (2007, 2008). The hierarchical model allows both local and grouped variables across different spatial and temporal scales to be used to assess stream response, i.e. stream integrity parameters. For example, Kashuba *et al.* (2009) used a multilevel statistical model to assess in-stream invertebrate responses to urbanization and important climate parameters (e.g., precipitation and air temperature) while simultaneously explaining differences in the rates at which invertebrate assemblages respond to urbanization across nine metropolitan regions in the U.S. The hierarchical model overcomes the drawbacks of the traditional statistical models and achieves a balance between treating the data from different groups as completely independent (unpooled) and treating the data from different groups as replicates (completely pooled) through partial pooling (Rechhow *et al*., 2009). As such, the global estimates of the model parameters are the weighted average of the group-specific estimates (*borrowing*) and the group-specific estimates are *shrunk* toward the global estimates. The degree of shrinkage depends on the group-specific uncertainties such as the sample size and variability of the observations (Reckhow *et al.,* 2009; Kashuba *et al.*, 2009). The multilevel models can potentially be modified to handle commonly available longitudinal data on land use and stream biological conditions (Hox, 2002).

Environmental Flows

Environmental flows, as defined by the Brisbane Declaration (2007), describe the quantity, timing, and quality of water flows required to sustain freshwater and estuarine ecosystems and the human livelihoods and well being that depend on these ecosystems. Water is necessary to sustain freshwater ecosystems and their services (e.g., fisheries, recreation, and tourism) that they provide to people. A comprehensive understanding of how water availability influences the

[9] A kernel distribution is a nonparametric representation of the probability density function (pdf) of a random variable. It is used when a parametric distribution cannot properly describe the data, or when the investigator wants to avoid making assumptions about the distribution of the data.

ability of a watershed, river, riparian, and estuarine ecosystem to provide those services is the basis for informed water management including decisions about allocating water among various users. Improvements in water management can only be achieved when there is a scientific understanding of where and when water is available, and we ensure that river systems have adequate base flows to support both people and ecosystem needs. The Bayesian hierarchical modeling approach can be used to improve this understanding and is especially applicable in the detection of important associations between stream flows, including managed flows, and biophysical responses in rivers (i.e., environmental flows). Properties unique to the hierarchical approach – "borrowing strength" and "shrinkage" – mean that conclusions can be greatly strengthened in data-poor situations but will be almost unaffected when data are plentiful (as opposed to the tendency to "over-fit" such data, as mentioned above). Webb *et al.* (2009) stress that the flexibility of Bayesian modeling allows formulation of realistic models, which can be tested for generality using all available data from any source (e.g., routine river health monitoring data, or a particular flow experiment). Models can be updated as new knowledge and data become available via an iterative cycle of development and testing. The advantages appear obvious given that environmental flow monitoring programs often require a large investment of public money. Management agencies need to be convinced their investments in environmental flows and the monitoring of ecological outcomes are cost-effective and worthwhile activities.

The Issue

Environmental flows represent a critical part of maintaining ecological integrity in regulated river systems (Poff *et al*., 1997; Tharme, 2003; Arthington *et al*., 2006). However, providing flow to the environment may be economically costly due to foregone consumptive benefits (e.g. agriculture; Qureshi *et al*., 2007), and also socially divisive due to the inevitable self-interest of consumptive and environmental water users (Schofield & Burt, 2003). Recent droughts across the U.S. and around the world have highlighted the tensions that can exist between allocating water for people and water for nature. Given these tensions, it is important to demonstrate the ecological benefits of environmental flows, particularly in regions with fully or over-allocated water resources such as southeast Australia. In this paper, the authors assessed effects of flow on (1) salinity in the Glenelg and Wimmera Rivers and (2) abundance of a fish – Australian smelt, *Retropinna semoni* (Weber 1895) in the Thomson River. This study was designed to evaluate the utility of Bayesian hierarchical models (BHMs) for assessing the environmental effects of flow; i.e., the analyses sought to identify a link between variation in flow and ecosystem response by using a Bayesian hierarchical approach to improve their capacity to detect important associations in the data. In doing so, the authors tested hypotheses underlying environmental flow recommendations. The data were taken from existing monitoring programs funded by the local catchment management authorities. The salinity analysis had a rich data set however the data for the Australian smelt analysis was relatively poor. It was stipulated that the Bayesian hierarchical approach to the problem would mitigate such difficulties because BHMs are, as noted previously, inherently flexible (Clark 2005) and allow for the construction of far more complex models than is possible with traditional statistical approaches.

Models and their Implementation

Salinity Model

Daily salinity and discharge data were available for six sites on the Glenelg River and four sites on the Wimmera River for the period 1 January 2000 to April 2007. Environmental flow recommendations have been made for all sites and it is hypothesized that recommended summer low-flow rates should maintain water quality. However, discharges over the summer period (1 December–31 May) often fall well short of the recommended levels.

An initial examination of the data revealed a negative linear relationship between salinity and flow, and high temporal autocorrelation in the data series. The authors analyzed these data as a linear regression of salinity against flow with extra terms to account for temporal autocorrelation. Because it is hypothesized that recommended minimum environmental flow rates should maintain water quality, the authors used the summer low-flow recommendations to scale the discharge data from each site. This produced a flow metric related directly to the environmental flow recommendations, and was also comparable among sites – a characteristic that facilitates hierarchical treatment of the data. Data cleaning and data transformation were based on background knowledge on data structure. For example, salinity data were $\log_{10}$-transformed and discharge data were 4th root transformed. Salinity was modeled at each site as a linear function (with intercept α_s and slope β_s) of transformed standardized discharge. Potential autocorrelation of the data was accommodated using the Cochrane-Orcutt transformation (Ciongdon, 2006). The model parameter of primary interest for the salinity analysis β_s is the rate at which salinity changed proportional to discharge at the site level. It was modeled hierarchically and the prior distribution of β_s was chosen as a minimally informative prior (see

Box 3

Statistical Model: Salinity

For each site, the data were analyzed with the following model:

$$\log(y_{i(S)}) \sim N(\mu_{i(S)}, \sigma^2_{[y]}) \qquad (3)$$

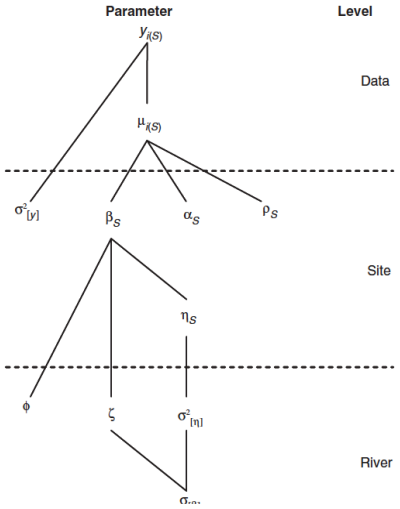
$$\mu_{i(S)} = \alpha_S + \beta_S \sqrt[4]{\frac{Q_{i(S)}}{L_S}} + \rho_S(\log(y_{i-1(S)}) - \alpha_S - \beta_S \sqrt[4]{\frac{Q_{i-1(S)}}{L_S}}) \qquad (4)$$

where $y_{i(S)}$ is the salinity data point for day i in the time-series at site S. N refers to a normal distribution, μ is the mean of the modeled salinity datum and $\sigma^2_{[y]}$ is the variance of the modeled salinity distributions at that site. Salinity was modeled at each site as a linear function of transformed standardized discharge with intercept α_S and slope β_S. The subscript S denotes that these are site-scale parameter. $Q_{i(S)}$ is the daily discharge and L_S is the summer low-flow recommendation for that site. Autocorrelation of the data was accommodated with the Cochrane-Orcutt transformation (Congdon 2006) that adjusted each μ_i by the difference between the previous data point and the line of best fit at that point, scaled by the standard autocorrelation coefficient ρ. Under the Bayesian framework, the investigators attempt to use minimally informative prior distribution for parameters, but the construction process requires some extra parameters in the model, namely:

$$\beta_S = \phi + \zeta \cdot \eta_S$$
$$\zeta \sim N(0, A^2)$$
$$\eta_S \sim N(0, \sigma^2_{[\eta]})$$
$$\sigma_{[\beta]} = |\zeta| \cdot \sigma_{[\eta]}$$

where ϕ is the overall mean of the distribution of β_S values, and ζ and η_S dictate the deviation of individual β_S values from ϕ. A is a 'scale' parameter, and corresponds to the median of the prior standard deviation. The posterior standard deviation, $\sigma_{[\beta]}$, of the distribution of β_s values was calculated as shown.

The hierarchical structure of this model is illustrated below.



Box 3 for more detail).

Smelt Model

Fish abundance data were available for 18 sites spread across six reaches of the Thomson River. There were three years of data (2005, 2006, 2007), with one bank-mounted or boat electrofishing sample taken per site each year (late March–early April), The analysis focused on Australian smelt because it was the most dominant species comprising 58% of the total abundance and because it is one of the species that the environmental flow program was designed to protect. In general, adult smelt are not considered to be particularly sensitive to flow changes, however, their eggs and larvae are highly associated with low-flow environments and the authors hypothesize that the number of fish recruited to the adult population will be a function of the amount of slow-flow habitat in the river over the summer period.

Based on the above ecological background and collected data structure, the smelt model was conceived as being multilayered (Box 4). The data were effectively adjusted for day-of-sampling discharge and turbidity effects before being aggregated at the reach-level and passed on to the main analysis. After transformation, a regression model with intercept λ_T and slope π_T was derived (Box 4). π_T, the regression slope of the relation

Box 4
Statistical Model: Smelt
The transformed site-level abundance data modeled as:

$$\log(y_{i[ST]}) \sim N(\mu_{i(st)}, \sigma^2_{[y]}) \quad (5)$$
$$\mu_{i(ST)} = \theta_{i(ST)} + \delta \log(Tu_{i(ST)}) + \gamma \log(Q_{i(T)})$$

where $y_{i[ST]}$ is the abundance data i=1…3 within each site (S) within each reach (T), $\mu_{i(ST)}$ is the mean of the modeled abundance for sample i, $\sigma^2_{[y]}$ is the variance of this distribution and $\theta_{i(ST)}$ is the mean of the modeled abundance once the effects of turbidity and discharge have been taken into account. $Tu_{i(ST)}$ and $Q_{i(T)}$ are turbidity and discharge on the day of sampling, respectively. These data were log-transformed to maximize the spread of explanatory variables in the analysis. The parameter δ and γ are coefficients for turbidity and flow covariates. The adjusted site-level data were aggregated at the reach scale such that

$$\theta_{i(ST)} \sim N(\varphi_i(T), \sigma^2_{[\theta]}) \quad (6)$$

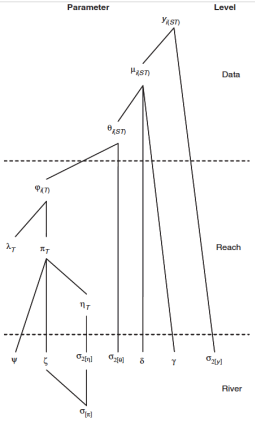
where $\varphi_i(T)$ is the mean of the distribution of site-level observations for the reach during year I, and other subscripts and parameters follow the naming conventions above. Within each reach, the reach-level mean abundance for each year were regressed against the average discharge for that summer period scaled against the low-flow recommendation:

$$\varphi_{i(T)} = \lambda_T + \pi_T \frac{\overline{\log(Q)_{i(T)}}}{\log(L_T)} \quad (7)$$

For this model, λ_T and π_T are regression intercept and slope parameters, respectively, $\overline{\log(Q)_{i(T)}}$ is the average log-transformed summer discharge over the 12 months preceding sampling and L_T is the summer low-flow recommendation for that reach. Same as in salinity model, the authors construct the prior as follows:

$$\pi_T = \psi + \zeta \cdot \eta_T$$
$$\zeta \sim N(0, A^2)$$
$$\eta_T \sim N(0, \sigma^2_{[\eta]})$$
$$\sigma_{[\pi]} = |\zeta| \cdot \sigma_{[\eta]}$$

where, ψ is the overall mean of π_T values, $\sigma_{[\pi]}$ is the standard deviation of this distribution and ζ, η and A have the same meanings as for the salinity model. The hierarchical structure of the model is illustrated below.



between reach-level smelt abundance and proportional achievement of the summer low-flow recommendation was modeled hierarchically among the reaches, using the same proportional achievement of the summer low-flow recommendation was the main objective for the smelt analysis. The regression slope π_T was modeled hierarchically among the reaches, using the same minimally informative prior as used in the salinity study (see Box 4 for additional details)

Implementation

The models were written using the Markov Chain, Monte- Carlo (MCMC) based Bayesian analysis using the WINBUGS 1.4.2 (http://www.mrc-bsu.cam.ac.uk/bugs; Lunn *et al*., 2000)[10]. To test the effect of using a hierarchical approach, the author also ran non-heirarchical versions of the models, by assigning minimally informative prior distributions to the parameters at a lower level ("site" for the salinity model, "reach" for the smelt model).

Study Results

Salinity Model

The distributions for β_s are characterized by different medians and credible intervals for different sites (Fig 1). The slopes were negative (i.e., increased discharge was correlated with reduced salinity) at three sites on the Glenelg River (G3, G4, G5) and three sites on the Wimmera River (W2, W3, W4); however, discharge was correlated with increased salinity at two locations on the Glenelg river (G2, G6) and appeared to have no effect at one site in each river (G1, W1). The precision of the β_s estimates was generally satisfactory, with the exception of W2. Hierarchical and non-hierarchical models produced results that were very similar. The posterior predictive checks indicated a very close fit between the hierarchical model and data, with high autocorrelation between log-transformed salinity and 'fake data' generated by the model simulation (r = 0.997).

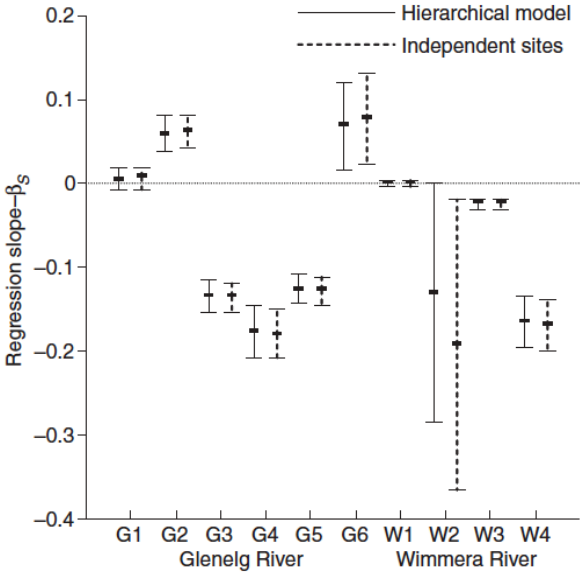


Fig. 1 Results for the salinity analysis. Graph shows β_S values at the different sites for the two model types, as dictated by the key. The median values are plotted, along with the 95% credible interval for the parameter distribution.

Smelt Model

Results of the hierarchical model runs showed that all reaches had probabilities for π_T that were positive between 0.1 and 0.3, indicating that increasing discharge was associated with a reduced number of smelt at the reach scale (Table 1). For the non-hierarchical model output, where the reaches were treated independently, there was a far larger range of probabilities, as well as, wider

[10] A brief survey of available "free" software for Bayesian analyses appears in Appendix III

intervals for π_T (especially for reaches 3 and 4b) (Fig. 2). The results also show that both higher turbidity and discharge on the day of sampling appeared to reduce the number of fish captured. The distribution of probability values for the simulated data showed a reasonable fit of the hierarchical model to the data.

Table 1. Results for smelt model

	Pr (parameter value >0)	
	Hierarchical model	Independent reaches model
π_T		
R2	0.11	0.05
R3	0.30	0.86
R4a	0.25	0.42
R4b	0.26	0.30
R5	0.21	0.28
Covariates		
δ	0.01	0.02
γ	0.11	0.11

Table shows the probabilities that parameter values are >0 for the reach-scale regression slopes and for the covariate effects of turbidity and flow.

Discussion and Conclusion

For salinity, the model results argue against the hypothesis that summer low flow should help to maintain water quality at all sites and instead compel the authors to consider mechanisms that may differ among sites. Also because of the unrealistic estimation of absolute changes in salinity at site W2, it is probably not advisable to consider absolute effects for a model where autocorrelation is so apparent.

For smelt, the hierarchical model provided some support for the hypothesis that higher than recommended summer discharges in the Thomson River led to a reduction in the abundance of smelt. These findings support the hypotheses of Milton & Arthington (1985) and Pusey *et al.* (2004) that higher flow during the larval and juvenile period might be expected to reduce smelt abundance in streams. Elevated summer discharges in the Thomson River occur principally to deliver irrigation water downstream. As such, there will always be tension between consumptive water use and environmental needs. Given that smelt are a common and widespread species in southern Australia, the authors conclude that it is unlikely that the findings reflect a significant enough ecological impact for existing flow rules to be reassessed, especially given the high levels of uncertainty associated with the predictions.

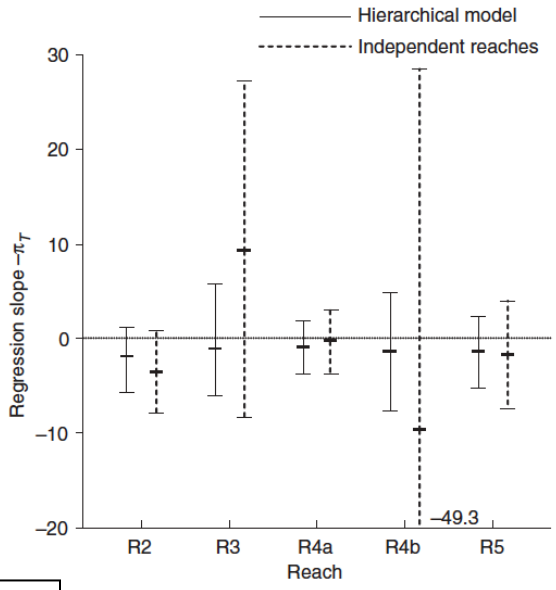


Fig. 2 Results for the smelt analysis. Graph shows π_T values for the different reaches for the two model types as dictated by the key. Plotting options are as for Fig. 9. For the error bar that extends beyond the range of the *y*-axis, the 2.5th percentile is given by the number adjacent to the error bar.

This study has demonstrated that Bayesian hierarchical modeling provides a rigorous statistical framework that can be used to detect the effects of flow on environmental variables. Unique to the hierarchical approach, the properties of borrowing strength and shrinkage mean that conclusions will be greatly strengthened in data-poor situations but will be almost unaffected when data are plentiful. Moreover, the Bayesian approach makes fitting hierarchical models relatively straightforward. Environmental flow monitoring programs such as VEFMP require a large investment of public funds, and the authors of the study believe that Bayesian hierarchical modeling approaches can maximize the benefits of such programs.

4.2 Case Study 2. Retrospective Design

Title: Bayesian Geostatistical Design (Diggle and Lophaven 2006)

In their study, Diggle and Lophaven (2006) referred to the set of sampling locations they used as the study *design*. They noted further that "designs that are efficient for parameter estimation are not necessarily efficient for spatial prediction given the model parameters." Accordingly, their approach was to develop a Bayesian design that integrated and balanced these competing considerations (Box 5). The idea was to find designs that were relatively efficient for spatial prediction while making "proper" allowance for the effects of parameter uncertainty; i.e., use parameter estimation as a means to the primary end of spatial prediction, rather than as just an end in itself.

Although two specific designs were distinguished; 1) *retrospective design* that arises when sampling locations are to be deleted from, or added, to an existing design[11]; and 2) *prospective design* that arises when the goal is to create a design in advance of data collection, only an example of the former will be summarized herein.

Box 5

A central concern of geostatistics is to predict a spatial surface of a region, A, using data that consist of potentially imperfect measurements, Y_i: i= 1,…, n, on the surface of a finite set of *sampling locations*, $x_i \in A$:i = 1,…, n. As noted in the introductory sections to this report, a widely used stochastic model for such data is the linear Gaussian model (LGM), which assumes the Y_i are normally distributed, with mean $E[Y_i] = S(x_i)$ and common variance τ^2.

The LGM in the author's study is represented by $[S,Y] = [S][Y|S]$; i.e., the model specifies the marginal distribution of the unobserved random field $S=\{S(x):x\in R^2\}$ and the conditional distribution of $Y=(Y_1, \ldots, Y_n)$ given S, with model parameters $\theta = (\beta,\sigma^2,\tau^2)$ a vector of unknown parameters. Under the Bayesian paradigm, unknown model parameters are also treated as random variables, and the symbolic representation is extended to $[S,Y,\theta]= [S|\theta][Y|S,\theta][\theta]$, where $[\theta]$ denotes the *prior* for θ.

The required distribution for predictive inference is $[S|Y]$, which is evaluated as:

$$[S|Y] = \int [S|Y,\theta][\theta|Y]\mathrm{d}\theta \qquad (9)$$

Thus, the Bayesian predictive distribution is a weighted average of classical, or "plug-in, predictive distribution $[S|Y,\theta]$ with different values weighted according to their *posterior* probabilities.

Study Results and Synthesis

The *retrospective* design criterion used by the authors was based upon the spatially averaged prediction variance,

$$\bar{v} = \int_A Var\{S(x)|Y\}\mathrm{d}x \qquad (8)$$

determined from the $[S|Y]$ and the posterior distribution of θ given in equation (9) (Box 5), and approximated by Monte Carlo simulation (Diggle *et al.* 2003). In the example that follows, the method was used to calculate a Monte Carlo approximation to the retrospective design criterion by repeated sampling from $[S|Y]$ where S represents the values of $S(x)$ at locations x on a spatial surface to represent A.

Monitoring Salinity in the Kattegat Basin

The Kattegat Basin (Figs. 3 a,b) is a relatively shallow coastal transition zone between the Baltic and North Sea that is degraded by eutrophication. The current water quality monitoring program at the site is comprised of 70 stations (Fig. 3b). Salinity is one of the water quality variables that is

[11] Retrospective designs are often incorporated into environmental monitoring, with the aim of economizing the long term costs of maintaining a network of monitoring site while minimizing the consequent cost of information.

routinely monitored in the study area, and serves partly as a surrogate for water column stratification that may affect dissolved oxygen concentrations below the halocline.

An initial geostatistical analysis showed a north-south trend in the salinity data, and for the design evaluations it was therefore assumed that a linear trend surface would be included in the chosen model,

$$\mu(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \qquad (10)$$

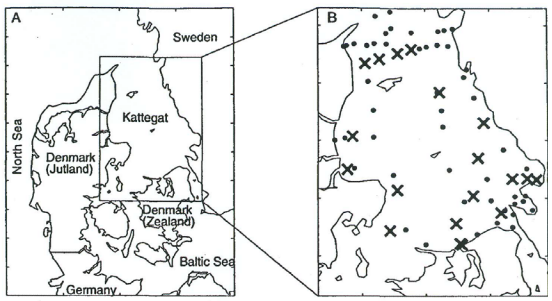


Fig. 3. (A) The Kattegat as a transitional sea between the North Sea and the Baltic Sea. (B) Locatic of the 70 monitoring stations (• and ×) measuring salinity in the Kattegat, and of the 20 stations (which constitute the final design.

Fig 3 a,b

where $x = (x_1, x_2)$. The north-south trend was expected because high salinity water from the North Sea flows into the northern part of Kattegat Basin, whereas water entering from the Baltic Sea to the south has a relatively low salinity. From the data, the authors estimated $\hat{\beta}$= (-69.0, -0.049, 0.020), and set the ratio $\nu^2 = \tau^2/\sigma^2$ at an estimated value of 0.42[12]. A uniform *prior* was selected for ϕ on an interval from 10 to 100 km, while for $(\beta, \sigma^2|\phi)$, a "diffuse" prior proportional to $1/\sigma^2$ was used. Lastly, the spatially averaged prediction variance was derived from equation (10) using 95 locations in a regular grid covering the Kattegat Basin at a spacing of 15 km, and using direct simulation of 1000 independent draws from the posterior. The resulting network design of 20 monitoring stations (Fig. 3b) consisted largely of well-separated stations, but with some pairs of closely located sites reflecting a compromise between designing for prediction and designing for estimation.

4.3 Bayesian Networks

Case Study 3 – Water Quality (Eutrophication)

Title: A Bayesian Network of Eutrophication Models for Synthesis, Prediction and Uncertainty Analysis (Borsuk *et al.* 2004)

Bayesian networks have proven to be a suitable means for performing integrated ecological modeling because their graphical structure explicitly represents cause and effect assumptions among system variables that might not be tractable using alternative modeling approaches such as deterministic point estimate modeling (Borsuk *et al.* 2004). The latter often consists of attempts to combine data from individual projects into a single predictive framework, usually by attempts to simulate relevant physico-chemical and biological processes at *pre-determined* model scales (Fitz *et al.* 1996). These authors note, however, that the most predictable relationships among sets of variables may emerge at numerous spatial, temporal and/or functional scales (Levin 1992), and that current scientific knowledge might be best served where *regular patterns of behavior emerge*, rather than at a scale that is identical for all processes (see also Jorgensen 1995). Thus, methods are required that: 1) allow representations at multiple scales and in a variety of forms, depending on available information; 2) assess how uncertainties in each

[12] The authors note that fixing ν^2 generally has a small impact on the chosen design.

component of the model translate to uncertainty in the final predictions (Reichert and Omlin 1999); and 3) allow models to be easily updated as knowledge and policy needs evolve (Walters 1986).

The basic idea of Bayesian network models ("influence diagrams", or "belief" networks as they are sometimes called) is that the uncertainty of the problem is described by the means of probabilities. As a general "rule of thumb", narrow probability distributions reflect high-quality information or good controllability (Kuikka *et al.* 1999). Probabilities can either be unconditional (i.e., not dependent on other variables) or conditional in which case the value of a variable depends on at least one of the other model variables. Conditional probabilities enable the modeling of "level of determinism; i.e., a poor knowledge or poor control is modeled by weak conditional probabilities and vice versa (Varis and Kuikka 1999). Stated in "management" terms, Bayesian networks focus on the relationship between action and knowledge, and so encourage the investigator to examine the options for actively managing an uncertain system and to conduct systematic studies on how information can support management. Pradham *et al.* (1966) have demonstrated that Bayesian networks are not sensitive to imprecision in the input probabilities and can, therefore, be classified as "robust tools".

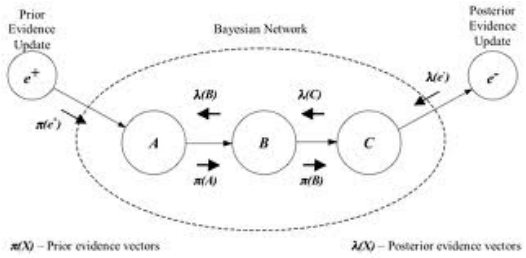


Fig. 4

The Bayesian network begins with a graphical depiction of relationships among the most important variables in the system (Fig. 4). Only where arrows occur is there a conditional anticipated dependence between one variable and another (shown by the circles). The graphical construct is important because it provides the basis for determining the degree of decomposition to be used in the subsequent construction of mathematical models (Varis and Kuikka 1997). This approach greatly facilitates the modeling process by allowing separate sub-models to be developed for each relationship defined by an arrow. The graphical network describes the probabilistic relationship among the system variables that resolves the joint distribution of all variables into a series of marginal and conditional probabilities[13] (Borsuk *et al.* 2004). Thus, unlike most integrated modeling exercises, Bayesian networks are probabilistic, rather than deterministic, expressions to describe relationships among variables. Clark *et al.* 2001, suggest that this is an essential and desirable characteristic of an ecosystem model if predictions are to guide decision making.

A Bayesian network was developed as an organizing structure for a set of previously published functional models used to study eutrophication impacts in the Neuse River estuary, North Carolina. Each model used was independently capable of real-time solutions in a network setting (e.g., Varis and Kuikka 1997). The authors claim that such an approach "can be expected to lead to greater exploitation of the representational and computational advantages of Bayesian networks, as well as more effective use of available scientific knowledge" (Borsuk *et al.* 2004).

[13] *Marginal* probability is the probability of one variable taking on a specific value irrespective of the values of the others; i.e., it gives the probabilities of various values of the variables without reference to the values of the other variables. A *conditional* probability, on the other hand, is the probability that an event will occur, when another event is known to occur or to have occurred, and is commonly denoted by $P(A|B)$; i.e., it gives probabilities contingent upon the values of the other variables.

The Issue

The Neuse River experiences severe consequences of eutrophication manifested in excessive algal blooms, low dissolved oxygen concentrations, declining shellfish populations, extensive fish kills, and toxic water conditions associated with blooms of *Pfiesteria piscida* (Fig. 5). Population growth and land development, animal-farming practices, storm water runoff, municipal wastewater, fertilizer use, and atmospheric deposition are implicated as the causative factors in nitrogen (N) enrichment, as has been the case nationally. In efforts to implement watershed based pollution control to limit N loading to the estuary, the state of North Carolina has implemented Total Maximum Daily Loads (TMDL's) as a strategy to address the issue (USEPA 1997; 1999). Typically, TMDLs are based on the output from a deterministic simulation model that predicts water quality characteristics, such as chlorophyll and/or DO concentrations at relatively fine spatial and temporal scales (USEPA 1999). Although these parameters are useful for agency personnel, they are often not terribly relevant to lay decision-makers, and the general public, whose interests largely focus on fish kills, harmful algal blooms (human health impacts), and shellfish mortality, etc. Moreover, the authors of the study suggest that at the scale employed in most simulation models, the ecological processes associated with these attributes are too complex or stochastic to be easily characterized mathematically.

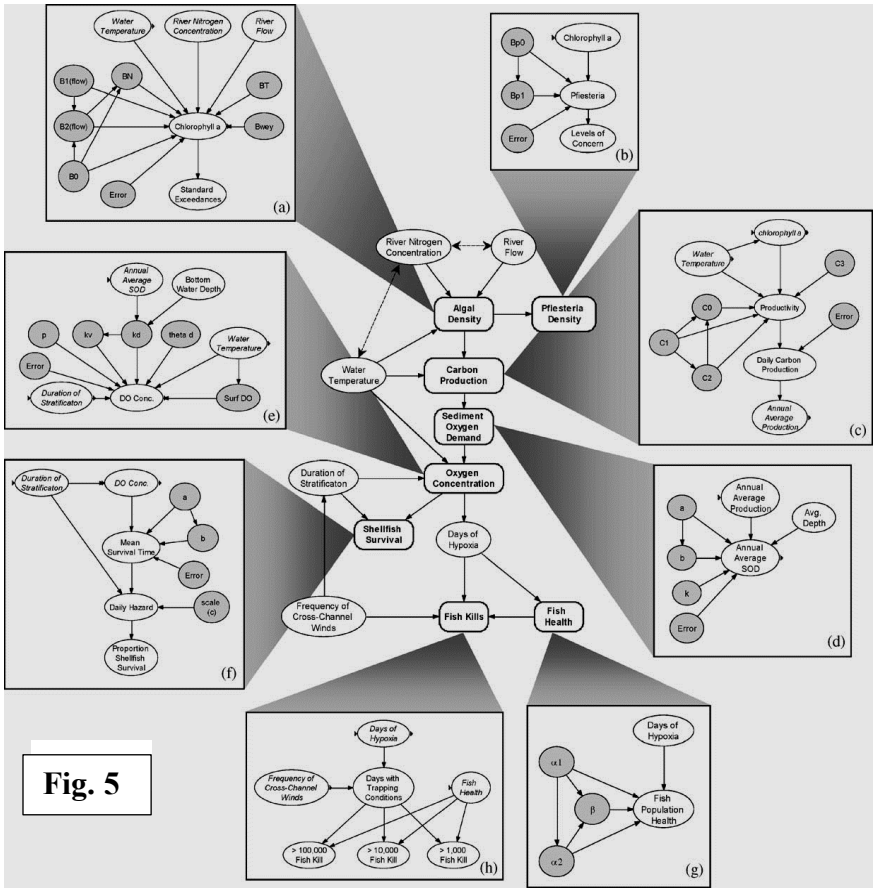


Fig. 5

In such cases, the study authors comment that "the aggregate causal relationships (Box 6) are well known and smaller scale dynamics might be better captured probabilistically[14]; therefore, flexible modeling tools that link processes occurring at multiple scales, might lead to better TMDL decisions by more directly addressing stakeholder concerns." (Borsuk *et al.* 2004).

[14] In the *Bayesian interpretation of probability*, the conditioning event (B) is interpreted as evidence for the conditioned event (A); i.e., $P(A)$ is the probability of A *before* accounting for evidence E, and $P(A|E)$ is the probability of A having accounted for evidence E.

Development of Causal Structure

The graphical structure (Fig. 5) of the Neuse eutrophication network was developed in two stages by 1) determining the attributes of the estuarine system for which decision-makers would like to see predictions; and 2) linking these attributes to N inputs using a causal network diagram. Because decisions of public officials should represent the views of the public, the attributes of concern were elicited from a set of stakeholders who cared about the health of the Neuse estuary. Input to the network diagram for the stage 2 activities was also solicited from an independent panel of estuarine research scientists. Thus, rather than forcing officials to extrapolate from traditional water quality variables to broader ecological attributes (Table 2), the goal of the investigators was to predict them directly using alternative model types integrated into a Bayesian network. As noted in Table 2, the selection of attributes demonstrated the broad public concern beyond those attributes generally predicted by traditional simulation models (Borsuk *et al.* 2004).

Network Development for the Neuse River Estuary Study

An "iterative" process was used to build the network diagram linking N inputs to meaningful attributes by first conducting a comprehensive survey of relevant scientific literature. Then using the primary attributes of the stakeholder process, it became a relatively straightforward process to identify the nodes preceding each attribute in the causal chain, ultimately back to the original model inputs including N loadings (Box 6):

Box 6

Methods to Quantify Conditional Relationships

Having established the primary causal relationships leading from N inputs to publicly meaningful ecosystem attributes (X), the next step was to quantify the conditional probability distribution of the attributes by:

$$X = f(p, \theta, \varepsilon) \qquad (11)$$

Where p is the set of immediate causes (or "parents") of X, θ is a vector of parameters of the function relating to p and X and ε is an error term ("random noise"). The simplest interpretation of this equation is that causal relationships representing physical mechanisms can be described by mathematical functions. If the arguments of the function are treated as random variables, then probability distributions can be assigned. From a Bayesian perspective, the distribution of the parameter set θ represents knowledge uncertainty of the parameter values that can be derived from a combination of prior judgement (see section 2) and statistical inference (Bernado and Smith 1994). The error term ε represents effects of endogenous factors that, out of choice or ignorance, have not been exclusively included in the model (Pearl 2000; see section 3). A common assumption is that ε is an independent, Gaussian distributed random variable with a mean of 0 and known (specified) variance.

Parameter estimates and probability distributions were developed for each sub-model shown in Fig. 5, and then integrated into the entire cohesive network. The functional relationship used for each variable and its parents in the sub-models follow:

- Algal density = f(water temperature, river flow, N concentration;
- *Pfiesteria* abundance = f(algal density);
- Carbon production = f(algal density)
- Sediment oxygen demand = f(algal carbon production);
- Bottom water oxygen concentration = f(sediment oxygen demand);
- Shellfish survival = f(bottom water oxygen concentration;
- Fish population health = f(bottom water oxygen concentration);
- Fish kills = f(fish population health, bottom water oxygen concentration);

The fully integrated Bayesian network comprised of the conditional probabilistic relationships described above was implemented in *Analytica* (Appendix II), a commercially available software program for evaluating graphical probability models (Lumina 1997). Uncertainty distributions were propagated throughout the network using Monte Carlo or Latin Hypercube sampling.

- The study authors then held a series of meetings with researchers to get their input on the causal diagram (Borsuk 2001). The process successfully produced a refined network linking causes and effects that represented the current opinion(s) of the scientist's panel;
- For purposes of completeness all relevant attributes proffered by the scientists were used in the graphical diagram resulting in a conceptualization with 35 nodes and 55 arrows. The authors noted though that "clearly some simplification was necessary to make the problem tractable and keep it consistent with available data; [in particular], … when the recommended variables were stochastic or uncontrollable and must be described by marginal distributions themselves [see footnote 14], then their inclusion is not very useful for informing management decisions" (Borsuk *et al.* 2004);
- Thus, to design the most parsimonious yet realistic model, each node in the network was reviewed to determine if the variable was either: controllable, predictable, or observable at the scale of the management problem. If not, the node was removed from the network.

Table 2 – Ecosystem attributes of concern to Neuse River stakeholders

Water Quality
Oxygen levels
Chlorophyll-a levels
Taste
Odor
Water clarity
Sandy bottom
Algal toxins
Biological Quality
Algal blooms
Fish/shellfish number and health
Species diversity
Human-induced fishkills
Submerged aquatic vegetation
Human Health
Fecal coliform
Toxic microorganisms

As a result of this iterative process, the conceptual model of the network was reduced to 14 nodes and 17 arrows (center insert in Fig. 5)

Study Results

To predict the effect of a substantial reduction in N inputs to the Neuse estuary, the marginal distribution of riverine N concentration was multiplied by one-half (i.e., a 50% reduction). All other functions and marginal nodes (e.g., Chl-a or DO concentrations) were held constant, and new probability distributions at the 50% reduction level of N were computed for the ecological variable of interest.

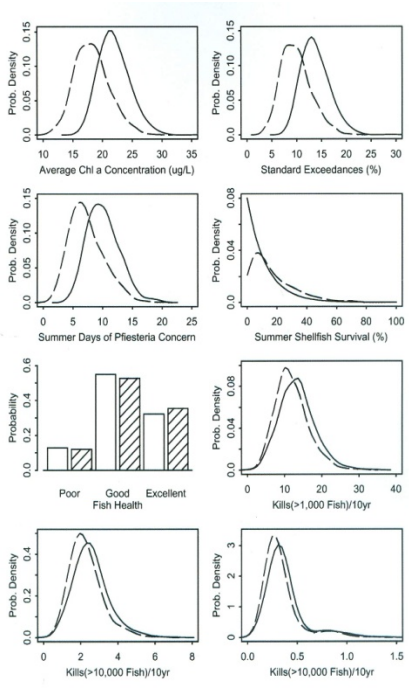


Fig. 6

Fig. 6

The baseline scenario of no N reduction is shown in Fig. 6 as solid curves, or in the instance of "Fish Health" as a histogram with solid bars. Scenarios for 50% reduction in N loadings as dashed lines or diagonally striped bars. The authors presented a summary for their model runs with and without N reduction (Table 3). Interestingly, the authors noted that the relatively minor response of most ecological attributes was revealed by looking at the trends in carbon production and days of summertime hypoxia (Fig. 7). While carbon production was expected to decrease by

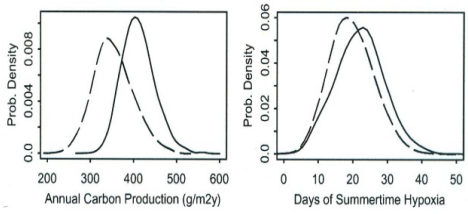


Fig. 7

about 15% in response to reduced algal stimulation, the effect was dampened further down the causal chain, so that the reduction in the number of days of resulting hypoxia was only 11% (Borsuk *et al.* 2004).

Synthesis and Conclusions

The primary goal of the Neuse River estuary study was to develop a model that more realistically represented current knowledge of the ecosystem, especially the linkage between N loading and the ecosystem attributes that reflected the interests of the public and state decision-makers. Because uncertainty is inherent at all levels

Table 3 Baseline Conditions, Predicted Responses to no N Reduction	Predicted Responses to 50% N Reduction
Average annual chlorophyll-a concentration in the estuary is expected to be slightly above 20 μg l^{-1} (90% CI: 18.3-26.5 μg l^{-1}); and the North Carolina standard of 40 μg l^{-1} is likely exceeded on more than 10% of the days (90% CI: 9.8 – 18.8%)	Average annual chlorophyll-a concentration decreased approximately 20%; however, the estimate was accompanied by a slight increase in uncertainty as suggested by the calculated CI (14.1-23.3 μg l^{-1}); predicted Chl-a exceedances were reduced to an average value of < 10%
Pfiesteria cell densities at levels of concern are expected to occur between 6 and 15 days (90% CI) during summer	*Pfiesteria* cell densities of concern decreased somewhat to between 3 and 13 days
Summer shellfish survival is predicted to be low, with a mean value of 12% (90% CI: 1 – 38%)	Shellfish survival rates showed a non-zero mode near 10% with a mean value of 17% (90% CI: 3-46%)
Under the baseline scenario (no N reduction), the most likely state of fish population health is "good", with a probability of 0.55, while "excellent" has a probability of 0.32, and "poor" of 0.13. This sub-model predicts between 6 and 21 kills in 10 years (90% CI) involving more than 1000 fish, between 1 and 4 kills involving 10,000+ fish, and an average between 0,2 and 0.8 involving 100,000 fish	Fish health increased modestly with probabilities of 0.12, 0.53 and 0.35 for "poor", "good" and "excellent" health, respectively; similarly, fish kill probabilities of all sizes decreased, but not substantially

of knowledge acquisition, Bayesian networks are used to represent that uncertainty using conditional probability distributions. The study authors noted further that "most commonly used aquatic ecosystem models have not undergone a rigorous uncertainty analysis [Reckhow 1994]. In this sense the Bayesian network was not seen as a replacement for other models in current use, but rather as an integrator of all forms of knowledge, whether expressed as a process-based description, a data-based relationship, or a quantification of expert judgment". Further, "to the extent that an existing simulation model appropriately represents our level of understanding about the functioning of the system, that model can be used as the basis for a set of relationships in the network." Thus, probabilistic predictions can give stakeholders and decision-makers a realistic appraisal of the chances of achieving desired outcomes – an outcome critical to the decision making process.

Because there was no single scale at which scientists have studies the Neuse system, there was no single scale at which all model relationships could be developed, [instead] a kernel characteristic of the Bayesian network was its ability to integrate sub-models at disparate scales

[Box 7] … [and,] therefore, when process models were used as an expression of knowledge in the Bayesian network, they were applied at a considerably more aggregate scale [of functional relationships among variables]". For some variables in the Neuse network, suitable "hard data" did not exist for quantifying functional relationships[15] among variables]." In such cases, the study team relied on "formally elicited judgement of a discipline-specific scientist". The study results demonstrated that ecological improvement was likely to result from N reductions, but the predictive uncertainty [in this study] arising from natural variation and lack of knowledge was high; in general, the less observable, less frequent and further down the casual chain a variable was the greater the predictive uncertainty. In such cases, and if the subject variable was of high interest to stakeholders, then a compromise would be necessary between achieving policy relevance and predictive precision; a task that the authors suggest was best performed by decision-makers[16].

Box 7

Integrating Sub-Models at Disparate Scales

While sufficient data was available to characterize a model relating the distribution of daily bottom water oxygen concentrations to sediment oxygen demand, temperature and duration of stratification, there was not enough site-specific data from the Neuse system to predict future changes in sediment oxygen demand in response to reductions in carbon loading. Thus, a model was developed using *cross-system* data from a number of estuaries to predict annual average carbon loading in the Neuse.

This average demand was assumed to represent the steady state mean, the short-term fluctuations around which could be predicted from water temperature changes using the oxygen dynamics model (Fig. 5). Expected changes in this mean rate of oxygen demand in response to carbon load reductions were then predicted from the cross-system model. This technique, known as "variable speed splitting" (Walters and Korman 1999), and was proposed by the authors as a useful general method for cross-scale modeling.

It was recommended further, that choosing the various scales of representation in a model should be a dynamic and iterative process.

This discussion would not be complete without consideration of "goodness of fit" statistics used for model testing. Usually, most of these statistics relate to deterministic or single valued predictions. But when predictions are expressed probabilistically, as in Bayesian networks, different methods are required for evaluation. Borsuk *et al.* (2004) note that methods have been developed for probabilistic weather predictions, and suggest that these are equally applicable to the ecological modeling domain. Most of these serve to characterize different attributes of the joint distribution of predictions and observations (e.g., see Murphy and Winkler 1987). They note further that various factorizations of the joint distribution provide different measures of prediction quality. They cite as a case in point, the question, "how often did different observations occur when a particular probabilistic prediction was given?"

The Bayesian network approach is not without shortcomings. Chief among these is its inability to explicitly represent system feedbacks. Because they are defined as "directed acyclic graphs", relations are either one-way causal influence at a particular time and place, or are integrated as *net* influences on steady-state conditions. The danger is that insufficiently capturing dynamic

[15] Although ecosystem data are often abundant, they are not necessarily sufficient at the spatial and temporal scale required by detailed simulation models.

[16] Scientific predictions only provide estimates of ecosystem response, which then require societal value judgements concerning costs and benefits in order to reach a rational decision.

aspects of system behavior can lead to unexpected consequences that are not adequately captured by the probabilistic predictions (Jorgensen 1999; Jorgensen *et al.* 2002).

Bayesian networks also do not improve the ability to represent structural uncertainty in ecological models. As with other modeling approaches, in addition to parameter uncertainty and natural variation that are captured by probability distributions, network models are subject to uncertainty in their inherent *causal structure*. This unaccounted for source of uncertainty implies that the real uncertainty in model prediction will be greater than that suggested by the model itself (Reichert and Omlin 1997).

4.4 Case Study 4 – Fisheries Management

Title: Modeling Environmentally Driven Uncertainties in Baltic Cod (*Gadus morhua*) Management by Bayesian Influence Diagrams (Kuikka *et al.* 1999).

Like most environmental management challenges, it is no surprise that uncertainty has become a centerpiece of fisheries assessment, and like the issues discussed above for water quality management, the goal of agency personnel and decision-makers has been to acquire a working understanding of uncertainty and its potential impact on management decisions. The role of scientists, of course, is to provide the tools that capture this uncertainty, and ultimately have science inform policy. The essential task in fisheries management then is to choose the correct action, with uncertainty reflected in precautionary approaches and risk assessment (FAO 1995).

Kuikka *et al.* (1999) comment "the recognition of uncertain information might substantially change our perception of the present state of resources and their exploitation compared with modeling by deterministic point estimate models [a thesis clearly shared above by Borsuk *et al.* 2004]. Ludwig (1996a and 1996b) concludes that, "maximum likelihood methodology, being based on point estimate parameters, grossly underestimates the risk of stock collapse." From the decision-making point of view, then, it is essential to analyze the sensitivity of management decisions to uncertainty. Some management strategies may be more information robust than others and some may be able to diminish future uncertainties more effectively than others.

With this background, Kuikka et al.(1999) set out to analyze the effects of uncertainty on the management of the Baltic Sea cod fishery, specifically by using a Bayesian network as meta-models to combine different sources of uncertain information, including the predictions proffered by three different recruitment models. Detailed descriptions of these models is beyond the scope of this study, but one of them, the "environmental Ricker model" merits further mention because it was the most sensitive to water quality deterioration, a major driving variable due to eutrophication impacts on the fishery (see below). The general approach is similar to that described for the Neuse River eutrophication meta-model by Borsuk et al.(2004). Decision variables (attributes) were selected and along with related objective functions (in the Neuse system, for example, the number of fish kills in a ten year period), that are either minimized or maximized by the decision attributes. As in the North Carolina study, the dependence of attributes on each other is described as conditional probabilities.

The Issue

Like many fisheries, Baltic cod management lacks sufficient knowledge of causalities, state variables, and fisheries parameters – fishing mortality, stock dynamics, stock size and catch rates (the latter, often misreported!). Combined with highly variable (unstable) hydrographic conditions exacerbated by low bottom dissolved oxygen concentrations, addressing uncertainty in both the assessment and implementation phases makes for a difficult management challenge.

Cod is the dominant piscivore in the Baltic Sea, but they are unique among cod populations in their ability to reproduce at salinities as low as 11‰. Yearly average catch of the species is about 200,000 kt. Both gill nets and trawls are used in the capture fishery, with the former used to obtain about 40% of the catch biomass. Current data on spawning volume have shown wide yearly fluctuations believed due to complications from eutrophication and the unpredictable influx of saline water (that ameliorates this problem). Over the past several decades these fluctuations have increased in magnitude further complicating future uncertainties.

In their study, Kuikka *et al.* (1999) modeled the relationship between the adult spawning stock and the number of recruits (young, immature fish) produced annually. The "decision" issue in their analysis was whether or not a change in gear mesh size would benefit the fishery[17]. Several alternative hypotheses and models were tested in a Bayesian network to identify the assumptions that "really mattered" from a management point of view. As expected, the simulation results obtained from the various models differed, but as the authors stated in a subsequent review, "this does not mean that the recommended actions should differ" (Varis and Kuikka 1999).

Study Results

The modeling task was to estimate the effects of two decision variables (mesh size and fishing mortality) on the interest variables (stock biomass, spawning volume, catch, and recruitment risk) and to demonstrate how sensitive the advised actions were to the relevant scientific uncertainties (Kuikka et al.1999). The analysis consisted of three main tasks: 1) using a model based on length distribution, calculate (deterministically) the selection process of the two gear types (trawl and gill net); 2) estimate the effects of alternative hypotheses about the recruitment process on the variables that decision makers and stakeholders tend to focus on, e.g., catch and stock biomass using an age structured model and Monte Carlo simulations[18]; and 3) use the conditional probability estimates in a Bayesian network meta-model to combine the uncertainty estimates of the different models. This last step required an exhaustive sensitivity analysis of the role of different sources of information and hypotheses from the point of view of decision-making (see below). Population parameters and the selection process used are summarized in Table 4 and Box 8.

[17] A large mesh size would allow smaller individuals to escape the trawl, thus reducing overall fishing mortality on the population.

[18] This step produced probability distributions of interest variables (e.g., catch and biomass) that were conditional on the recruitment hypothesis tested.

Bayesian Network Construction

Probability distributions derived from the Monte Carlo simulations (Box 8) were used as input data to construct the network diagram (Fig. 8). The Hugin software (http://www.hugin.dk/) (Appendix I) based on Lauritzen's algorithim (Lauritzen and Speigelhalter 1988) was used, an approach that the authors suggest "is computationally very effective, as it allowed totally interactive use of the model, despite the relatively large number of separate Monte Carlo simulations."

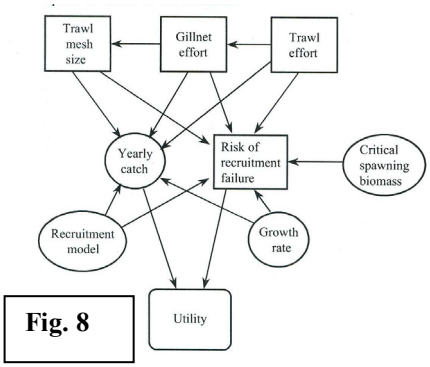


Fig. 8

Table 4. Input values for model simulations

Natural Mortality	0.2
Fishing Mortality of *trawl* fishery for totally recruited length class (1/year)	0.50
Fishing mortality of *gillnet* fishery for maximally recruited length class (1/year)	0.65
Length of fish at beginning of maturation year (cm)	40
Length-weight relationship $W = aL^b$	
a	0.0102
b	3.00
Survival of trawl escapees (%)	100
Mean length of 1-year-old fish (cm)	16.4
CV of length of 1 year old fish	0.35
Gillnet catch as a proportion of total catch (%)	40

The network diagram (Fig. 8) combines the input and output values of the Monte Carlo simulations:

- Prior variables (Table 5) included *growth rate*, *recruitment model* (using the three models described in Box 8), and *critical spawning biomass*. Because water quality in the Baltic Sea was expected to deteriorate in the future, the environmental Ricker model was assigned the highest *prior* probability;
- Decision variables included *mesh size of the trawl* (120 and 140 mm), and the *gillnet* and *trawl* generated fishing mortalities;
- The information in the nodes *yearly catch* and *risk for recruitment* were assigned conditional probabilities; e.g., if mesh size of the trawl used was 140 mm, *F* factor for the trawl fishery was assigned 1.0, *F* factor for the gill net fishery was assigned 0.8, growth was assumed to be "slow" (Table 4), and the recruitment model chosen was "random", then the probability of a yield falling between 100,000 and 150,000 kt was 0.3.

Table 5. Probabilities of priors used in configuring the Bayesian Network diagram

Prior Variable	Alternatives	Probability
Recruitment Model	Random	0.1
	Ricker	0.3
	Environmental Ricker	0.6
Growth Rate	Fast	0.7
	Slow	0.3
Critical Spawning Biomass (kt)	240,000	0.7
	480,000	0.3

- To address subjectivity in the assignment of prior probabilities (historical data were not used) an extensive sensitivity analyses was performed by testing the preferred decisions to the prior probabilities for the three recruitment models against models that 1) combined both risk and catch as the objective function, and 2) with models having only catch as the objective function.

Synthesis and Conclusions

A potential problem in dealing with uncertainty is that managers may get the impression that all relevant uncertainties have been addressed in a particular situation. Although the investigators in this study focused on the effects of recruitment uncertainty on management decisions, they commented that there are "obviously other variations in natural mortality rates; e.g., degree of cannibalism[19], interactions with other populations (e.g., competing for resources), and the survival of trawl escapees (thought to be less than 100%) could affect management decisions." Moreover, density dependent responses in the population to stock size, mortality rates, and available resources were not considered (Kuikka *et al.* 1999). The authors warn that underestimation of uncertainties like these may result in too overoptimistic projections of reaching a desired state of the fishery.

In addition to uncertainties related to the ecosystem at hand, there were also uncertainties related to management objectives. One example was the inclusion or exclusion of recruitment risk, another was the subjectivity associated with desired stock size. The author's choice to largely constrain uncertainty to considerations of growth and recruitment, however, was based on the assumption that an intermediate term (versus long-term) management strategy for Baltic cod would focus mainly on new mesh size rules. Their approach using selection curves and the modeling of length distributions was a response to this

Box 8

Population Parameters and Selection Process

- Fishing mortality rate (F) at a given length class was calculated by multiplying the fishing mortality of fully recruited length classes by the length-specific gear retention rate:

  $F(l,g) = F(g)*r(l,g)$, where $F(l,g)$ is the mortality rate for length-class l in gear g, $F(g)$ is the fishing mortality rate of the fully recruited length group in gear g, and $r(l,g)$ is the retention rate of length class l in gear g ($r[l,g] \leq 1$). Historical data were used for these estimates. Gear and age specific F values were calculated using the length distribution model and the retention data.

- Recruitment was modeled in three ways: 1) randomly from historic recruitment values reported in Sparholt (1996); 2) with the Ricker model – $R = A*SSB*\exp(-B*SSB)$ --with inputs for spawning stock biomass (SSB), and coefficients A and B from Sparholt (1995); and 3) with the environmental Ricker model that tracks the fluctuations caused by influxes of high salinity water that ameliorates the effect of eutrophication by increasing the volume of water suitable for spawning.

- Monte Carlo simulations were use to produce the conditional probability distributions of variables "of interest" (e.g., yield and probability that biomass was below a certain critical threshold). Simulations were conducted separately for all combinations of priors (growth rate, recruitment model [random, Ricker and environmental Ricker], and critical spawning biomass) and interest variables (Figure 5 and Table 4). With the exception of the environmental Ricker model where 600 iterations were run, each simulation consisted of 400 iterations among 110 years for each of the 300 simulations (2 growth rates * 3 recruitment models * 2 mesh sizes * 5 gillnet F values*5 trawl F values). Six hundred iterations were used with the environmental Ricker model due to the unstable behavior of the model. The number of iterations chosen in each case yielded asymptotic 'expected values' and variance. The authors reported that long simulations were need due to autocorrelation effects. Lastly, the probability distribution of the 90th year was chosen to describe the long term interests on managers and fishers.

[19] Cannibalism of young-of-year by adults in the spawning population is a frequently observed phenomenon in fish stocks, and may relate to density-dependent population controls.

requirement. A key finding of their study was that assumptions of environmental variability make a significant difference to conclusions regarding fishery management. The combination of declining trends in water quality (portended by the environmental Ricker model) and high exploitation rates was "dangerous for stock and fishery alike." In these scenarios, more age classes would be required to buffer environmentally
poor years with poor recruitment. If the fishery was, at present, heavily dependent on recruitment, as ascertained by this study, both fishers and managers should be ready to adjust the "total allowable catch" (TAC) *on short notice* after a prediction of poor recruitment. Other relevant comments worth stating in this summary include:

- The study outcome suggests that, with better management decisions vis-à-vis gear selection and increased mesh size, the mean yearly catch could increase by 66,000 kt;
- Increasing the mesh size to 140 mm would also markedly reduce the frequency of dangerously low spawning biomass and the need for very low TACs;
- The finding that a larger mesh size is beneficial irrespective of the assumed recruitment process is important because it implies rob**u**stness in the approach; i.e., that the key conclusions remain valid despite considerable uncertainty on the future state of the environment, and management goals would be achieved only by reducing fishing mortality rates;
- Although knowledge of the recruitment process is generally considered to be key in successful management of fish stocks, this study demonstrated that the need for knowledge of this process can be diminished simply by shifting the management strategy, in this case, by increasing mesh size. Such a measure would not remove all uncertainties, of course, but would serve as an "insurance fee" against their negative impact. In the Baltic Sea cod fishery, this fee would be the short-term catch loss after a mesh size change.

Thus, decision analysis of the type performed here can identify management strategies that are less sensitive to scientific uncertainty and implementation errors. Meta-modeling of structural uncertainties using a Bayesian network was shown to be a demonstrably efficient tool in the search for robust management strategies.

5. KRIGING

5.1 Case Study 5. Groundwater Salinity

Title: Spatial and Temporal Mapping of Groundwater Salinity Using Ordinary Kriging and Indicator Kriging: The Case Study of Bafra Plain, Turkey (Arslan 2012)

Introduction

By itself, kriging is not a direct Bayesian approach, but the latter can be used to improve estimates of uncertainty in the covariance function (see below). We use this case study to introduce kriging as a useful method, and then summarize the use of a Bayesian paradigm to analyze the performance of the kriging predictor (see section 5.2).

Geospatial kriging depends on a combination of mathematical and statistical models. The addition of a statistical model that includes probability separates kriging from deterministic methods; i.e., you associate some probability with your predictions when the values are not perfectly predictable from a statistical model alone.

In general, kriging is an interpolation technique that can construct statistically optimal predictions for data at *unobserved locations* using relatively small, spatially explicit sets of samples (Little *et al.* 1997; Arslan 2012). The predictions ("kriges") at a given location (encompassed by a grid of locations over the geographic area of interest) are calculated on the basis of weighted average of the sample values (sample points near the prediction location are given larger weights than those that occur further away), with weights usually assigned as the straight line (Euclidean) distance between actual sample sites and the target location[20] Weights are determined empirically by 'semivariogram analysis" (below). The latter models the similarity of sample values, in pairs, as a function of distance (or "lags") between the sampling sites (Little *et al*.1997).

For geostatistical data, trends can be expressed by following formula:

$$S(x) = \mu(x) + \varepsilon(x) \qquad (12)$$

where $S(x)$ is the variable of interest, decomposed into a deterministic trend $\mu(x)$ and random, autocorrelated errors of the form $\varepsilon(x)$. The symbol x simply refers to a particular place or location. No matter how complicated the trend in the model is $\mu(x)$ will not be predicted perfectly. In this case, some assumptions about the error term $\varepsilon(x)$ are made; namely that they are expected to be 0 (on average) and that the autocorrelation between $\varepsilon(x)$ and $\varepsilon(x + h)$ does not depend on the actual the location s but only on the displacement h between the two. Variations on equation (2) form the basis for all the different types of kriging; but in the interest of space, only three variations are be briefly summarized here:

1. If the trend is a simple constant; i.e., $\mu(x) = m$ for all locations x, and if μ is unknown, then the applicable approach is referred to as *ordinary kriging*;
2. It is also possible to perform transformations on $S(x)$. For example, it can be changed to an indicator variable, where it is 0 if $S(x)$ is below some value; e.g., 0.12 ppm for ozone concentration, or 1 if it is above that value. Predictions that the $S(x)$ is above the threshold value employ the proportion of values the fall within specific class intervals and incorporate the uncertainty of variable values at unsampled locations. This approach is referred to as *indicator kriging*; and
3. Whenever the trend is completely known (i.e., all parameters and covariates are known), whether constant or not, the model used forms the basis for *simple kriging*.

There are no hard and fast rules on choosing the "best" semivariogram model. Usually, the investigator examines the empirical semivariogram and chooses a model that is judged appropriate. Validation and cross-validation to estimate the accuracy of the interpolated values

[20] If the variable of interest, e.g., the concentration of a contaminants in sediments occurs in a riverine environment, the distance between two sites may be alternatively measured by the distance through water pathways (Little *et al*.1997)

is also employed. In practice, ordinary kriging (without inclusion of covariates) is the model used most often to calculate predictions.

The Semivariogram

The semivariogram is defined as one-half the average squared difference between points separated by distance h, as follows:

$$\gamma(h) = \frac{1}{2|N(h)} \sum_{N(h)} (x_i - x_j)^2 \qquad (13)$$

where $N(h)$ us the set of all pairwise Euclidean distances i - j = h, $|N(h)$ is the number of distinct pairs in N(h), and x_i and x_j are data values at spatial locations i and j, respectively. In this formulation, h represents a distance measure with magnitude only. If two locations, say x_i and x_{i+h}, are close to one another in terms of the distance measure of h, then you can expect them to be similar, so the difference in their values, $x_i - x_{i+h}$, will be small. As x_i and x_{i+h} get further apart, they become less similar, so the difference in their values will become larger (Fig. 9). In this typical semivariogram the variance of the difference increases with distance, so that the semivariogram itself can be thought of as a dissimilarity function. Several terms are associated with this function: a) the height that the semivariogram reaches when it levels off is called the *sill*. It is usually composed of two parts a discontinuity at the origin referred to as the *nugget* effect (comprised of measurement error and microscale variation), and the partial sill, which when added together give the sill. The distance at which the semivariogram becomes asymptotic is called the *range*.

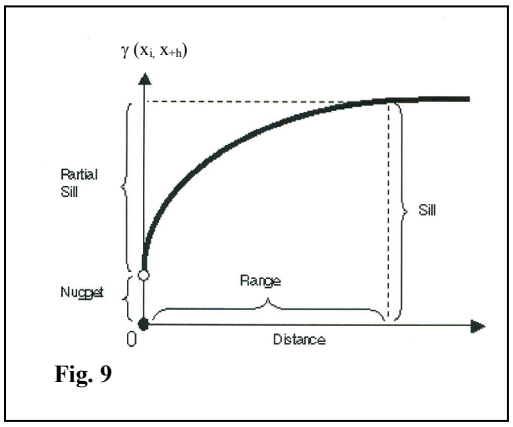


Fig. 9

Kriging has appeal for investigators because it requires minimal user intervention, generates optimal predictions under a given set of assumptions, is robust with respect to model choice and mis-specification of the mean function, and provides standard errors for the user (Cressie and Zimmerman 1992). It is inherently cost effective because it can be used to develop data (surface) layers for Geographical Information System (GIS) analysis of parameters that are expensive or time consuming to measure in large numbers over an entire study area (Vernberg *et al.* 1992). By identifying spatial patterns and interpolating values at unsampled locations, geostatistical analysis plays an important role in environmental management by providing estimated input parameters at regular grid points from measurements taken at randomly chosen locations (Arslan 2012).

Geostatistics provides useful techniques for managing spatially distributed data such as soil and groundwater pollution. In this case study, soil salinization due to elevated salinity in irrigation water, and its effect on crop production, was examined.

The Issue

In Turkey, large areas are affected by irrigation- related groundwater problems. The Bafra Plain, Right Bank Irrigation Area, covering about 1% of the Turkey's total irrigated area, is one of the largest irrigation and drainage projects in the nation. Excessive use of irrigation water, seepage from canals, inefficient irrigation methods, and inadequate or malfunctioning drainage systems have led to groundwater loss and elevated salinity threatening the sustainability of farming in the region. Geostatistics and Geographical Information System (GIS) technology was employed by the author to analyze the spatial distribution and seasonal variability of groundwater salinity in the Bafra Plain over a seven year period between 2004 and 2010 (Box 9) (Arslan 2012).

Box 9

Groundwater salinity was monitored over a seven year period (2004-2010) at 97 locations mapped with GPS coordinates. In areas where no measurements were taken, values were interpolated with ordinary and indicator kriging, and the data obtained were used to construct spatial variability and probability maps of the distribution of groundwater salinity.

Data were analyzed in four stages: 1) data were tested for normality in each study year; 2) descriptive statistics for annual groundwater salinity levels were generated - μ, σ, min, max; 3) trend analysis was used to identify the best predictive model from among 11 different semivariogram models tested; and 4) kriging techniques were employed to estimate or predict salinity concentrations at un-sampled locations.

For the semivariogram,

$$\gamma(h) = \frac{1}{2n} \sum_i^n (S_i - S_{i+h})^2 \qquad (14)$$

where $\gamma(h)$ is the estimated or "experimental" semi-variance for all pairs at lag distance h, $S(x_i)$ is the water quality value at point I; $S(x_{i+h})$ is the water quality value of all other points separated by a discrete distance h; x_i are the geo-referenced positions where the $S(x_i)$ were measured, and n represents the number of pairs of observations separated by distance h. *Ordinary kriging* was used to generate predictive maps for annual groundwater salinity and to interpolate groundwater salinity for un-sampled locations. *Indicator kriging* was used to obtain data to plot seasonal groundwater probability maps. Finally, prediction performance was assessed by cross validation. For a model to provide accurate predictions, the standardized mean error should be close to 0, the root mean square error should be as small as possible, and the root mean square error should be close to 1 (ESRI 2008).

Study Results and Synthesis

The descriptive statistics obtained in the study suggested that the data were not normally distributed; therefore, values were log-transformed prior to calculation of semivariance.

Ordinary Kriging

Spatial distribution of groundwater salinity was classified according to nugget-to-sill ratios (Fig. 1), with ratios of < 25% suggesting "strong" spatial dependence, 25-75% suggesting "moderate" spatial dependence, and > 75% indicating weak spatial dependence (Cambardella *et al.* 1994). Findings for the nugget to sill ratios of the semivariograms produced in this study demonstrated that groundwater salinity had a moderate spatial structure for all years tested, with similar annual values ranging from 6748 m (nugget ratio, 44.31) to 12,682 m (nugget ratio, 51.08). Cross-validation results suggested the groundwater mean error to be close to 0; i.e., between, – 0.0054 and – 0.0810, and the root mean square error term ranging from 1.205 to 3.033 among the seven years of study. The author noted that "healthy" irrigation waters usually have electrical conductivity (EC) threshold value of < 2.50 dS m^{-1}, but when waters with higher conductivity values are used, crop production can suffer (Richards 1954). In this study, spatial variation of salinity was distributed over five "thematic" classes: non-saline (< 2.25 dS m^{-1}), saline (5.0 – 7.5 dS m^{-1}), very saline (7.5 – 10.0 dS m^{-1}) and very high saline (> 10.0 dS m^{-1}) regions. As shown in Figs. 10 and 11 a,b salinity in the region increased with decreases in elevation, an observation that was "explained" by the down-slope movement of salts through precipitation and irrigation.

To address this problem, the General Directorate of State Hydraulic Works partially completed construction of an irrigation and drainage network, with 175 km of drainage canals opened in a 6000 ha area in the southern portion of the Bafra Plain. The observation that soil salinity decreased somewhat between 2004 and 2010 was attributed to this effort. However, as much of the region remains threatened by ongoing salinization processes, a recommendation was made for continued monitoring in the future.

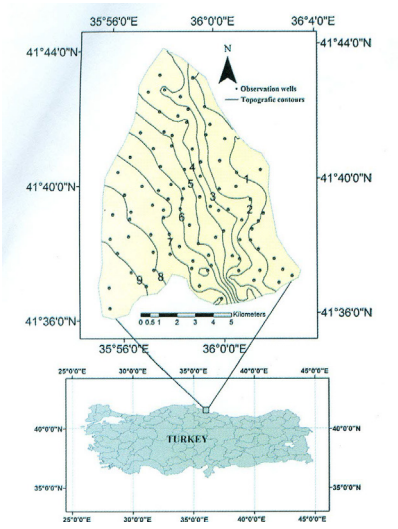
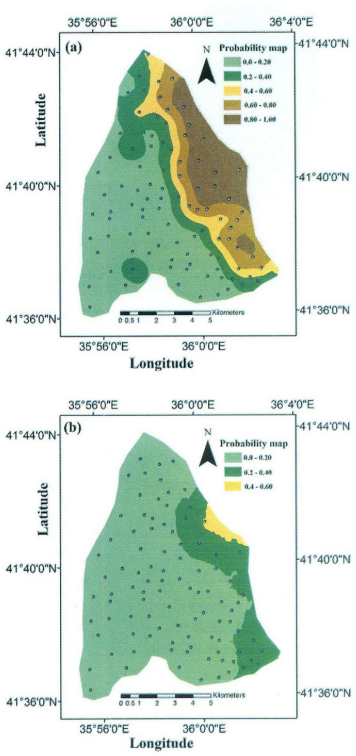


Fig. 10

Indicator Kriging

Indicator kriging was used to generate groundwater salinity probability maps for the years 2004-2010. At each sampling location, measurements were taken on a continuous scale and converted to discrete indicator variables given a value of either '1' or '0', with the former indicating a value below the threshold level (in this case 5.0 dS m^{-1}, or "moderately saline") for groundwater electrical conductivity. Probability maps generated by the author are shown in Fig. 11 a,b). Calculations for nugget-to-sill ratios suggested that groundwater salinity displayed moderate spatial structure for all years tested, with similar annual values ranging from 9,651 m to 12,682 m. Cross validation analysis resulted in a groundwater salinity mean error close to 0 (between, -0.0047 and 0.0046), and a mean square error term ranging from 0.8391 to 1.2150.



Despite a trend of decreasing EC over the seven year period, nearly 14% of the total area displayed the highest probability (0.8-1.0) of exceeding the threshold EC value of < 2.50 dS m^{-1}. Although, with one exception, none of the area locations exhibited the highest probability (0.8-1.0) threshold value, parts of the area continued to show a strong probability (0.6 – 0.8) of exceeding the threshold (ranging from 0.3 to 11.8% in individual years.

On the basis of the seven year program, Arslan 2012 concluded that:

Fig. 11 a,b

- Groundwater salinity displayed a tendency to increase toward the north and east of the Bafra Plain (Fig. 12 a,b);

- Over time overall groundwater salinity tended to decrease from 2004 to 2010 (Fig. 3);
- Ordinary kriging showed that in 2004, 31% of the area had non-acceptable levels of salinity, and moreover, the remaining 68% of the area was potentially at risk of excessive salinity;
- While salinity was found to have decreased substantially by 2010, the entire area "remains problematic", with 9% of the area above unacceptable levels of salinity and 71% at risk of excessive salinity;
- Indicator kriging showed (Fig. 12a) that in 2004, 13.6% of the area, mainly in the northern part of the plain had the highest probability (0.8-1.0) of exceeding the threshold for acceptable salinity levels;
- While the seven year decrease in salinity was reflected in lower probability values, the observation that 6% of the area still showed strong tendencies toward exceedances (0.6-0.8), led the author to comment that this trend "remained very alarming".

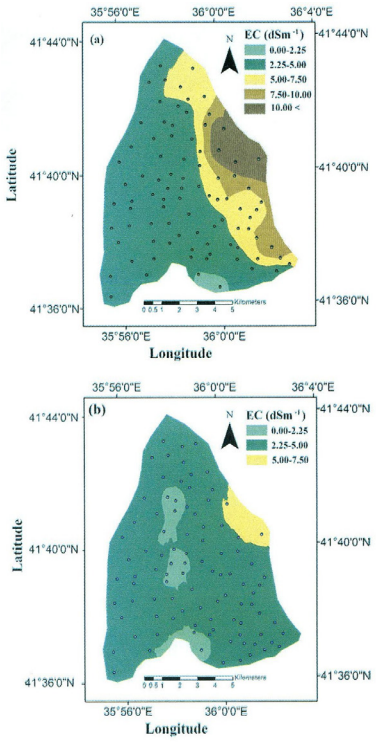


Fig. 12 a (2004), b (2010)

5.2 A Bayesian Analysis of Kriging

In kriging, the statistical model is seldom exactly known and is usually estimated from the very same data from which later predictions are made (Handcock and Stein 1993). In this article, the author's assessed the effect of model uncertainty on model predictions. Where substantial previous knowledge existed in model development, the authors suggested that a Bayesian approach could improve the accuracy of the model predictions. They focus on a parametric representation of the covariance structure, "as its direct interpretation is of interest".

As before, trends in the data can be expressed by:

$$S(x) = \mu(x) + \varepsilon(x) \qquad (12)$$

where $S(x)$ is the variable of interest, decomposed into a deterministic trend $\mu(x)$ and random, autocorrelated errors of the form $\varepsilon(s)$. The symbol s simply refers to a particular place or location. In assessing uncertainty in kriging, the quality of the prediction is determined by the distribution of the prediction error $\varepsilon_\theta(s_0) = S(x_0) - S_\theta(x_0)$. The kriging predictor *is the best linear unbiased predictor of the form* $S_\theta(x_0) = \lambda(\theta)'S$; i.e., the unbiased linear combination of the observations that minimize the variance of the prediction error. The authors note, however, that the underlying kriging procedure is motivated by sampling considerations, producing point predictions and associated measures of uncertainty for those predictions based on sampling distributions unconditional on the observed S. However, kriging when the mean is of a known

regression form can be given a Bayesian interpretation. It does so in the following summary manner:

- Traditionally, it is assumed that the covariance function is known exactly and the investigator has little knowledge of β, the vector of unknown regression coefficients, prior to analyzing the data;
- The underlying kriging approach usually presumes ignorance about β and the unrelatedness of β to the behavior of the covariance function;
- Under these assumptions, an appropriate *prior* distribution has pr($\beta|\alpha$, θ) locally uniform; and because β is a location parameter, the form of the prior used by the authors has the form pr(α, β. θ) $\propto$ pr(θ)/α.

Depending on the influence of θ the spread and location of pr($S(x_0)|\theta$, S), the Bayesian predictive distribution might be wider or narrower that the "plug-in" predictive distribution. Additionally, the Bayesian predictive distribution typically has no simple analytic form and must be determined numerically. The difference between the "plug-in" and Bayesian predictive distributions represents the *difference in inference* between the traditional kriging approach and the full Bayesian approach. Handcock and Stein (1993) comment, "a better approach [to dealing with uncertainty] would be to base inference on the Bayesian predictive distribution by taking into account the uncertainty about the covariance function expressed in the likelihood surface and ignored by point estimates of the covariance function. It allows the performance of the usual plug-in predictive distribution based on an estimated covariance structure to be critiqued with in a wider framework. [Our] results also suggest that fitting the empirical correlation function by eye[21] may lead to plug-in predictive distributions that differ markedly from the Bayesian predictive distribution."

This approach can be readily integrated into empirical Bayesian kriging (EBK), which compatible with ArcGIS software[22] (Kivoruchko 2012). Classical kriging assumes that the estimated semivariogram is the "true" semivariogram of the observed data; having been derived from a Gaussian distribution with the correlation structure defined by the estimated semivariogram. *This assumption rarely holds true in practice* and requires that action be taken to make the statistical model more realistic. EBK differs from classic kriging (and the aforementioned issue) by estimating the semivariogram model in several steps:

- A semivariogram model is estimated from the data;
- Using this semivariogram, a new value is simulated at each of the input data locations; and
- A new semivariogram model is estimated from the simulated data.

[21] One of the most common methods for fitting a covariance model to data is to match by eye a theoretical curve to the empirical correlation plot of the detrended observations.

[22] EBK can be implemented with the ArcGIS 10.1 Geostatistical Analyst extension

A weight for the new semivariogram is then calculated using Bayes' rule[23]; that is, a step to show *how likely* the observed data can be generated from the semivariogram.

VI. SUMMARY AND RECOMMENDATIONS

> *Data coming from many different studies have to be integrated in order to assess the empirical evidence for a new theory, and Bayesian statistics lends itself very well to this…Working scientists have noticed this, and many are using these tools now. With the increasing statistical literacy of empirical scientists and the growing availability of Bayesian computer software, the future of Bayes rule, along with that of other approaches to inference, seems well assured*
>
> Van Hulst 2013

Because it is probable that subjects of management interest (e.g., cause and effect between water pollution and ecosystem health) have been treated multiple times by the community of scientists, most Bayesians argue that more progress can be made if we chose to make use of data that *already exist* to frame our hypotheses. Bayesians argue further that we can make more progress by specifying the observed difference, and then using our data to extend earlier results of other investigators. Bayesian analysis does this, as well as, quantifies the probability of the observed difference. This is the most important difference between Bayesian and frequentist methods that can be described in six steps:

1. Specify the hypothesis;
2. Specify parameters as random variables;
3. Specify the prior probability distribution;
4. Calculate the likelihood;
5. Calculate the posterior probability distribution; and
6. Interpret the results.

This approach is valuable in several respects. If the motivation of the modeling effort is *prediction*, what counts most in addition to the model's predictive capability is a reasonable estimate of uncertainty in the model predictions (Omlin and Reichert 1999). In the case of *point parameter* estimates, as we have seen in this report, a single value is chosen for each model parameter and the uncertainty of this value is estimated from the local properties of the deviations for the model results from the data at this point in parameter space. *Regional* estimation, on the other hand, makes estimates of parameter distributions instead of values. In the case of good identifiable parameters, a narrow probability distribution is obtained, whereas the existence of one or more non-identifiable parameters (which is often the case) leads to a wide distribution of these and correlated parameters if more precise prior knowledge is not available (Omlin and Reichert 1999). Thus regional, rather than point estimation techniques may be more advantageous in situations where the parameters of the system are not identifiable because the data used for parameter estimation are often sparse relative to the model's complexity. Reichert

[23] The essence of Bayes' rule as discussed in Section 2 is to provide a mathematical rule explaining how you should change your existing beliefs in the light of new evidence. In other words, it allows scientists to combine new data with their existing knowledge or expertise.

and Omlin (1997) suggest that Bayesian techniques may be more applicable in these common situations, because they are based on probability theory and they explicitly consider probability distributions for *prior* knowledge and for the measurement process (in the likelihood function).

The problem of underestimation of prediction errors with parsimonious, identifiable models is a general problem of forecasting, and is especially important for modeling environmental systems for two reasons (Omlin and Reichert 1999); (1) the complexity of natural systems often requires model simplification and modeling of only a small component of the entire system[24], and (2) as state previously, "fire hose" quantities of data are available in the literature. For the latter reason, knowledge on processes that become important during the prediction period may be available, although it might not be contained in the data evaluated for model selection and parameter estimation. Omlin and Reichert (1999) suggest that "it is exactly for this reason that Bayesian techniques can be advantageously applied to these situations … [and] are very useful for the estimation of the uncertainty of model predictions for environmental systems".

Bayesian statistical design has recently experienced a resurgence of interest in the techniques and their application to environmental management problems. Although frequentism is a more cautious and self-critical philosophy, better able to withstand skeptical scrutiny from scientists, Bayesian methods are easier to explain and understand than their frequentist counterparts (Efron 2013). Efron goes further with an elegant comment, "there are two potent arrows in the statisticians quiver, and there is no need to hunting armed with only one."

We adopt Efron's (2012; 2013) comments by noting, 1) current data gathering practices using modern instruments create prodigious data sets that "bear on complex webs of interrelated questions", and the 2) in this new scientific era, the ability of Bayesian statistics to "connect disparate inferences counts heavily in its favor", and in general terms, ensures that *genuine* informed priors become the rule rather than the exception.

In its review of existing literature, the EPSC has identified several advantages of the Bayesian inference:

- Combined with the complexity inherent in most ecosystems, and the severity of environmental issues confronting managers and decision makers, many agencies and organizations have sought to explore new spatial analytical techniques that provide timely, valid information to assist problem solving, and effective environmental management decisions. The Bayesian approach has the potential to do just this;
- The essential approach of the Bayesian method is to address the question: can we get better estimates of the mean by collecting additional data from those populations, other than just the *ith* one? For both the Bayesian and empirical Bayesian, the answer is yes;
- As we have seen in Case Study 2, Retrospective Design, Bayesian methods have the potential to substantially reduce monitoring costs;

[24] The danger here is that in the predictive interval aspects of the system become important that are not described in the model of the subsystem.

- Other common issues with frequentist statistical approaches; e.g., those that might be used with hydrogeomorphic assessments[25] are related to weak experimental design in terms of replication, spatial and temporal confounding issues and the nature of the 'treatment' itself (in this case, flow characteristics). Even with a 'gradient-based' approach that stratifies flow according to stream-bed slope cannot fully overcome the replication problem, and it is usually financially or physically prohibitive to sample with sufficient replication to detect significant differences between flow and response. Bayesian methods can mitigate some of these difficulties because the approach is "inherently flexible", i.e., models are constructed to conform to the requirements of the data, whereas "standard statistical approaches must force the data to comply with the requirements of a relatively small number of model types" (see also McCarthy 2007). The use of Bayesian hierarchical models, for example, may help obviate the problem of replication and are finding increasing uses in ecological applications. They appear particularly suited to dealing with the complexities of spatiotemporal variation in ecology, and allow for the construct of "far more complex models" than is possible with traditional statistical approaches; and
- Unlike most integrated modeling exercises, Bayesian networks are probabilistic, rather than deterministic, expressions to describe relationships among variables. This is an essential and desirable characteristic of an ecosystem model if predictions are to guide decision making. The Bayesian network approach has proven to be a suitable means for performing integrated ecological modeling because their graphical structure explicitly represents cause and effect assumptions among system variables that might not be tractable using alternative modeling approaches such as deterministic point estimate modeling.

  The latter often consists of attempts to combine data from individual projects into a single predictive framework, usually by attempts to simulate relevant physico-chemical and biological processes at *pre-determined* model scales. However, the most predictable relationships among sets of variables may emerge at numerous spatial, temporal and/or functional scales, and that current scientific knowledge might be best served where *regular patterns of behavior emerge*, rather than at a scale that is identical for all processes. Methods are required that: 1) allow representations at multiple scales and in a variety of forms, depending on available information; 2) assess how uncertainties in each component of the model translate to uncertainty in the final predictions; and 3) allow models to be easily updated as knowledge and policy needs evolve. The Bayesian network approach meets this requirement. The basic idea is that the uncertainty of the problem is described by the means of probabilities. As a general "rule of thumb", narrow probability distributions reflect high-quality information or good controllability. Probabilities can either be unconditional (i.e., not dependent on other variables) or conditional in which case the value of a variable depends on at least one of the other model variables. Conditional probabilities enable the modeling of "level of determinism; i.e., a poor knowledge or poor control is modeled by weak conditional probabilities and vice versa. Stated in "management" terms, Bayesian networks focus on the relationship between action and knowledge, and so encourage the investigator to

[25] The provision of environmental flows is critical to, for example, maintaining ecological integrity of regulated river systems where there are 'competing' flows for ecosystem and anthropogenic uses (e.g., agricultural uses).

examine the options for actively managing an uncertain system and to conduct systematic studies on how information can support management. In general, Bayesian networks are not sensitive to imprecision in the input probabilities and can, therefore, be classified as "robust tools".

In this review, we do not mean to imply that Bayesian methods are not without disadvantage (as are most other methods). Among these are 1) computational challenges, even currently available software (Appendix III) may be difficult to use; 2) the requirement to condition the hypothesis on the data; and the potential lack of objectivity, because different results will be obtained using different priors. However, as discussed in the Introduction to this report, Efron (2012) while cautioning that primary criticisms of the Bayesian approach stem from overenthusiastic application of "uninformative priors", also suggested that current data gathering practices using modern instruments produces voluminous data that improve the ability of Bayesian statistics to connect disparate inferences and ensures that "genuine" informed priors become the rule rather than the exception.

On the basis of the full discourse herein, the EPSC makes the following recommendations to the SAB:

- Because a comparison and evaluation of how other states and entities use Bayesian methods is well beyond the scope of this report, the EPSC recommends that NJDEP adopt a two phase approach to the question; (1) develop a survey instrument to ascertain what other states and entities are doing in this arena; and (2) depending on the outcome of step (1), convene a workshop of technical personnel from selected state and federal resource agencies (USGS, NOAA, and USEPA), and selected academic institutions, to address the general theme: *The Use of Bayesian Inference to Address Environmental Monitoring and Management Challenges.*
- DEP should work with colleges and universities to develop a one-week continuing education curriculum in applied aspects of Bayesian-inference for NJDEP scientists. The curriculum should be relevant to statewide monitoring programs and should be compatible with ArcGIS programs.
- NJDEP, in conference with their in-house and state university statisticians, geo-spatial modelers, and ecologists should identify a "training data set" from their vast monitoring programs to compare model characteristics and output (robustness and efficiency, potential bias and flexibility) and performance capacity among frequentist and Bayesian methods.
- Similarly, and with the same approach, conduct a "sensitivity analysis" on existing NJDEP monitoring data sets using retrospective analysis to examine the relationships among sampling locations, sampling frequency, and resource allocation, to enhance the quality of information produced; e.g., by using real-time, remotely collected data from the Department's data logger array.
- The Department should make its existing library on Bayesian literature available to the user community upon request.
- The Department should issue a request for proposals (RFP) to academic institutions in New Jersey for the study of practical applications of Bayesian methods that address state environmental management and ecological issues.

- Promote the practical application of Bayesian inference as an additional, oft desirable, tool in the Department's analytical toolkit.
- Use Bayesian inference, and the content of this report, to encourage the broader use of statistics in the Department's development of study designs and quantitative data analysis in fulfilling its regulatory mandates.

VI. REFERENCES CITED

Arslan, H. 2012. Spatial and temporal mapping of groundwater salinity using ordinary kriging and indicator kriging: the case of Bafra Plain, Turkey. *Agricultural Water Management* 113:57-63.

Asmussen, S., & Glynn, P. W. (Eds.). (2007). *Stochastic simulation: Algorithms and analysis* (Vol. 57). New York: Springer.

Berger, J. O. 1990. Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference* 25:303-328.

Berger, J.O. and D.A. Berry 1988. Statistical analysis of illusion and objectivity. *American Scientist* 76:159-165.

Bernardo, J. M. 1979. Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 113-147.

Bernado, J.M. and A.F.M. Smith. 1994. *Bayesian Theory*. Wiley, New York, NY.

Borsuk, M.E. 2001. *A Graphical Probability Network Model to Support Water Quality Decision Making for the Neuse River Estuary, North Carolina*. Ph.D. Dissertation, Duke University, Durham, NC.

Borsuk, M.E., C.A. Stow, and K.H. Reckhow. 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecological Modelling 173:219-239.

Brisbane Declaration. 2007. International River Symposium and International Environmental Flows Conference, held in Brisbane, Australia, on 3-6 September 2007.

Burrough, P.A. 1995. 7. Spatial aspects of ecological data. In Jongman, R.H.G., C.J.F. Ter Braak and O.F.R. (Eds.) *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge, UK.

Box, G.E.P. and G.C. Tiao 1973. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

Cambardella, C.A., T.B. Moorman, J.M. Novak, T.B. Parkin, D.L. Karlen, R.F. Turco and A.E. Konopka. 1994. Field-scale variability of soil properties in central Iowa soils. *Soil Science Society of America Journal* 58:1501-1511.

Carpenter, S.R. 1990. Large-scale perturbations: opportunities for innovation. *Ecology* 71:2038-2043.

Clark, J.S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8:2-14.

Clark, J.S., S.R. Carpenter, M. Barber, *et al*. 2001. Ecological forecasts: an emerging imperative. *Science* 293:657-660.

Cressie, N. and D.L. Zimmerman. 1992. On the stability of the geostatistical method. *Mathematical Geology* 24:45-59.

Dennis, B. 1996. Discussion: should ecologists become Bayesians? *Ecological Applications* 6:1095-1103.

Diggle, P.J. and S. Lophaven. 2006. Bayesian geostatistical design. *Scandanavian Journal of Statistics* 33:53-64.

Diggle, P.J., P.J. Ribeiro, and O. Christensen. 2003. An introduction to model-based geostatistics. In J. Moller (Ed.) *Spatial Statistics and Computational Methods. Lecture Notes in Statistics 173*, Springer, NY.

Efron, B. 2013. Bayes theorem in the 21st century. *Science* 340:1177-1178.

Efron, B., and Morris, C. 1973. Stein's estimation rule and its competitors—an empirical Bayes approach. *Journal of the American Statistical Association.* 68:117-130.

Efron, B. 2013. A statistically significant future for Bayes' rule (response to van Hulst). *Science* 341:343.

Ellison, A.M. 2004. Bayesian inference in ecology. *Ecology Letters* 7:509-520.

ESRI. 2008. *Using Arc GIS Geostatistical Analyst*. Environmental Systems Research Institute, Redlands, CA, 300 pp.

FAO. 1995. Precautionary approach to fisheries. 1. Guidelines for the precautionary approach to capture fisheries and species introductions. *FAO Fisheries Technical Paper* No. 350/1.

Fitz, H.C., E.B. DeBellevue, R. Costanza *et al.* 1996. Development of a general ecosystem model for a range of scales and ecosystems. *Ecological Modeling* 88:263-295.

Gelman A. Multilevel (hierarchical) modeling: what it can and cannot do. 2006. *Technometrics* 48:432-435 .

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis*. CRC press.

Gotelli, N.J. and A.M. Ellison. 2004. *A Primer of Ecological Statistics*. Sinauer Associates, Sunderland, MA.

Greenland, S. 2000. Principles of multilevel modelling. *International Journal of Epidemiology* 29:158-167.

Howson, C. and P. Urbach. 1991. Bayesian reasoning in science. *Nature* 350:371-374.

Hox, J. 2002. Multilevel Analysis: Techniques and Applications, Mahwah, NJ: Lawrence Erlbaum Associates.

Jorgensen, S.E. 1995. State of the art in of ecological modeling in limnology. *Ecological Modeling* 78:101-115.

Jorgensen, S.E. 1999. State of the art ecological modeling with emphasis on the development of structural dynamic models. *Ecological Modeling*120:75-96.

Jorgensen, S.E., J. Marques, and S.N. Nielsen. 2002. Structural changes in an estuary, described by models and using exergy as an orientor. *Ecological Modeling* 158:233-240.

Kennen, J.G., 1999. Relation of macroinvertebrate community impairment to catchment characteristics in New Jersey streams. *Journal of the American Water Resources Association* 35: 939-955.

Kennen, J.G., Henriksen, J.A., and Nieswand, S.P., 2007, Development of the Hydroecological Integrity Assessment Process for Determining Environmental Flows for New Jersey Streams: U.S. Geological Survey Scientific Investigations Report 2007-5206, 55p.

Kennen, J.G., Kauffman, L.J., Ayers, M.A., and Wolock, D.M. 2008. Use of an integrated flow model to estimate ecologically relevant hydrologic characteristics at stream biomonitoring sites. *Ecological Modelling* 211:57-76

Kreft, I., and De Leeuw, J. 1998. *Introducing Multilevel Modeling*, London: Sage.

Kuikka, S., M. Hilden, H. Gislason, S. Hansson, H. Sparholt and O. Varis. 1999. Modeling environmentally driven uncertainties in Baltic cod (*Gadus morhua*) management by Bayesian influence diagrams. *Canadian Journal of Fisheries and Aquatic Sciences* 56:629-641.

Levin, S.A. 1992. The problem of pattern and scale in ecology. *Ecology* 73:1943-1967.

Little, L.S., D. Edwards and D.E. Porter. 1997. Kriging in estuaries: as the crow flies, or as the fish swims? *Journal of Experimental Marine Biology and Ecology* 213:1-11.

Lauritzen, S.L. and D.J. Speigelhalter. 1988. Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society Series B (Methodology) 50:157-224.

Legendre, P. and Legendre, L. 1998. *Numerical Ecology*, 2nd Edit. Elsevier Science BV, Amsterdam, Netherlands.

Lehmann, E. L., and Casella, G. 1998. *Theory of Point Estimation* (Vol. 31). Springer.

Ludwig, D. 1996a. Uncertainty and the assessment of extinction probabilities. *Ecological Applications* 6:1067-1076.

Ludwig, D. 1996b. The distribution of population survival times. *American Naturalist* 147:506-526.

Maritz, J.S. and T. Lwin. 1989. *Empirical Bayes Methods*, 2nd Edition. Chapman and Hall, London, UK.

McCarthy, M.A. 2007. *Bayesian methods for Ecology*. Cambridge University Press, Cambridge, NY.

NRC (National Research Council). 2002. Our common journey: A transition toward sustainability. Washington, DC: National Academy.

Omlin, M,, & Reichert, P. 1999. A comparison of techniques for the estimation of model prediction uncertainty. *Ecological Modelling* 115:45-59.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK.

Raiffa, H., & Schlaifer, R. 1961. *Applied Statistical Decision Theory* (Harvard Business School Publications).

Raudenbush, S. W., and Bryk, A. S. 2002. *Hierarchical Linear Models* (2nd ed.), Thousand Oaks, CA: Sage.

Reckhow, K.H. 1994. Water quality simulation modeling and uncertainty analysis for rish assessment and decision-making. *Ecological Modeling* 72:1-20.

Reichert, P., and M. Omlin. 1997. On the usefulness of over parameterized ecological models. *Ecological Modeling* 95:289-299.

Reichert, P., and M. Omlin. 1999. A comparison of techniques for the estimation of model prediction uncertainty. *Ecological Modeling* 115:45-59.

Richards, R.S. 1954. *Diagnosis and Improvement of Saline and Alkaline Soils.* USDA Agriculture Handbook, 60, Washington, DC.

Robbins, H. 1985. An empirical Bayes approach to statistics (pp. 41-47). Springer New York.

Robins, H. 1955. An empirical Bayes approach to statistics. *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability* 1:157-163.

Snijders, T. A. B., and Bosker, R. J. 1999. *Multilevel Analysis.* London: Sage.

Sparholt, H. 1995. Using the MSVPA/MSFOR model to estimate the right-hand side of the Ricker curve for Baltic cod. *ICES Journal of Marine Science* 52:819-826.

Sparholt, H. 1996. Causal correlation between recruitment and spawning stock size of central Baltic cod.

USEPA (Office of Water). 1997. *Compendium of Tools for Watershed Assessment and TMDL Development*. EPA-841-B-97-006. Environmental Protection Agency, Washington, D.C.

USEPA (Office of Water). 1999. Protocol for Developing Nutrient TMDLs. EPA-841-B-99-007. Environmental Protection Agency, Washington, D.C.

Van Hulst, R. 2013. A statistically significant future for Bayes' rule. *Science* 341:343.

Varis, O. and S. Kuikka. 1997. Joint use of multiple environmental assessment models by a Bayesian meta-model: the Baltic salmon case. *Ecological Modeling* 102:341-351.

Varis, O. and S. Kuikka. 1999. Learning Bayesian decision analysis by doing: lessons from environmental and natural resource management. *Ecological Modeling* 119:177-195.

Ver Hoef, J.M. 1996. Parametric empirical Bayes methods for ecological applications. *Ecological Applications* 6:1047-1055.

Vernberg, F.J., W.B. Vernberg, E. Blood *et al.* 1992. Impact of urbanization and high-salinity estuaries in southeastern United States. *Netherlands Journal of Sea Research* 30:239-248.

Walters, C.J. 1986. *Adaptive Management of Renewable Resources*. Macmillan, New York, NY.

Walters, C.J., and J. Korman. 1999. Cross-scale modeling of riparian ecosystem responses to hydrologic management. *Ecosystems* 2:411-421.

Webb, J.A., M.J. Stewardson, and W.M. Koster. 2009. Detecting ecological responses to flow variation using Bayesian hierarchical models. *Freshwater Biology* doi:10.1111/j.1365-2427.2009.02205.x

Wolfson, L.J., J.B. Kadane, and M.J. Small. 1996. Bayesian environmental policy decisions: two case studies. *Ecological Applications* 6:1056-1066.

APPENDIX I

DATA LOGGERS AND OTHER REAL-TIME MONITORING

This section is intended to give a brief summary of available continuous or near-continuous data gathering devices in the region. It does not provide and exhaustive review of these assets, but is simply discussed here in terms of the question posed to the EPSC in our original mandate: what types of data reduction techniques are necessary in combination with the increased use of deployable long-term data loggers in order to maintain adherence to appropriate statistical assumptions of independence and autocorrelation? We address that question in this overview.

NJ Department of Environmental Protection Water Quality and Biomontoring Program

NJDEP's water quality monitoring effort addresses ambient conditions of the state's fresh, marine and ground water resources (Fig 13):

Fresh Water and Ground Water. The Bureau is responsible for monitoring the ambient conditions of the state's Fresh and ground water resource monitoring includes regular sampling through a statewide network consisting of 115 surface water monitoring stations, 760 benthic macroinvertebrate biological stream monitoring stations, 100 fish assemblage biological stream monitoring stations, and 150 ground water stations. In addition, the bureau operates the Department's biological monitoring laboratory where bioassay and macroinvertebrate, fish and algal studies are regularly conducted. The bureau is also responsible for implementing the state's Ambient Lakes Monitoring Program. For ground water, New Jersey has developed and now maintains a cooperative network (NJDEP & USGS) consisting of 150 wells screened at the water table that are sampled 30 per year on a 5-year cycle. Parameters measured include conventionals, nutrients, VOCs, radioactivity, and pesticides.

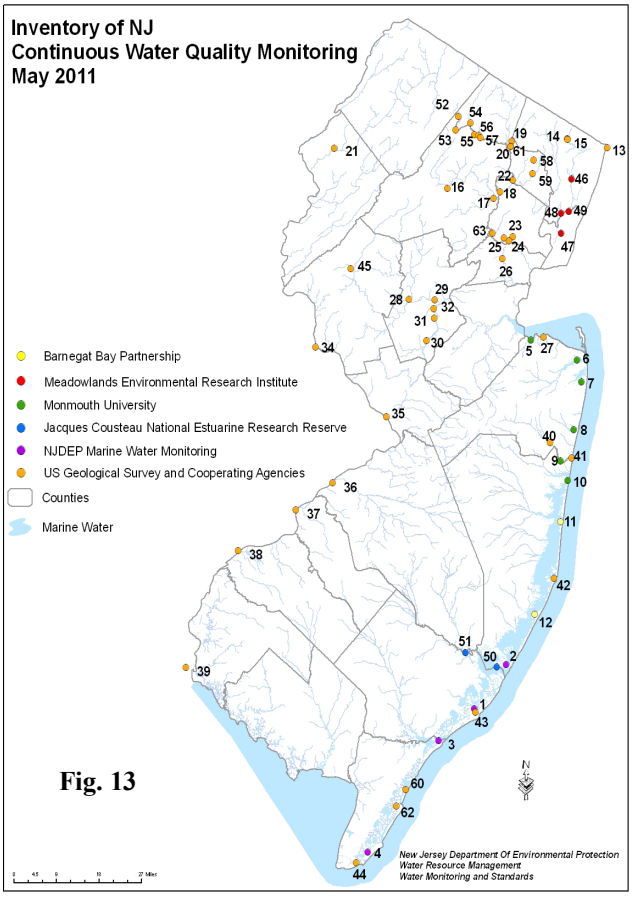


Fig. 13

Marine Waters. NJDEP conducts water quality monitoring to classify approximately 700,000 acres of marine and estuarine shellfish waters. For the National Shellfish Sanitation Program (NSSP), NJDEP collects approximately 15,000 ambient water samples per year from a network

of more than 2,500 monitoring stations throughout the State's coastal waters. These stations are sampled between five (5) and twelve (12) times per year. As part of the NSSP, NJDEP also conducts coastal phytoplankton monitoring every summer in New Jersey's bay and near-shore ocean waters. NJDEP also monitors the condition of the State's coastal waters by measuring basic water quality (dissolved oxygen, nutrients and water clarity) at 260 locations on a quarterly basis. EPA's National Coastal Assessment (NCA) research program is performed in partnership with NJDEP and includes measurements of sediment chemistry, sediment toxicity and the benthic community annually at about 50 locations in New Jersey's estuarine waters.

After undergoing QA/QC protocols, all data are integrated into the department's assessment database for use in preparation of Integrated Water Quality Monitoring and Assessment reporting as well as the addition of new external water monitoring data (e.g., volunteer monitoring) to STORET through development of a common data exchange element.

Mid-Atlantic Regional Association Coastal Ocean Observing System (MARACOOS)

Perhaps the most comprehensive and integrated regional use of "real-time" monitoring devices in New Jersey and the region is administered by the Mid-Atlantic Regional Association Coastal Ocean Observing System (MARACOOS), an entity established in 2004 as part of the U.S. Integrated Ocean Observing System (IOOS). Since then MACOORA created the framework in which the Mid-Atlantic's coastal ocean user community identified its five highest priority regional themes:*(1) Maritime Safety, (2) Ecosystem Based Management, (3) Water Quality, (4) Coastal Inundation,* and *(5) Offshore Energy*. MACOORA established the Mid-Atlantic Regional Coastal Ocean Observing System (MARCOOS) to provide the necessary ocean observing, data management, and forecasting capacity to systematically address the prioritized regional themes. Operations include an industry-funded coastal weather network, primary and back-up satellite data acquisition centers, a triple-nested multistatic HF Radar network, an accelerating autonomous underwater glider capability, and mission-specific statistical and dynamical ocean forecast models (Fig. 14). A subset of MACOORA assets are shown in the Figure consisting of the following elements:

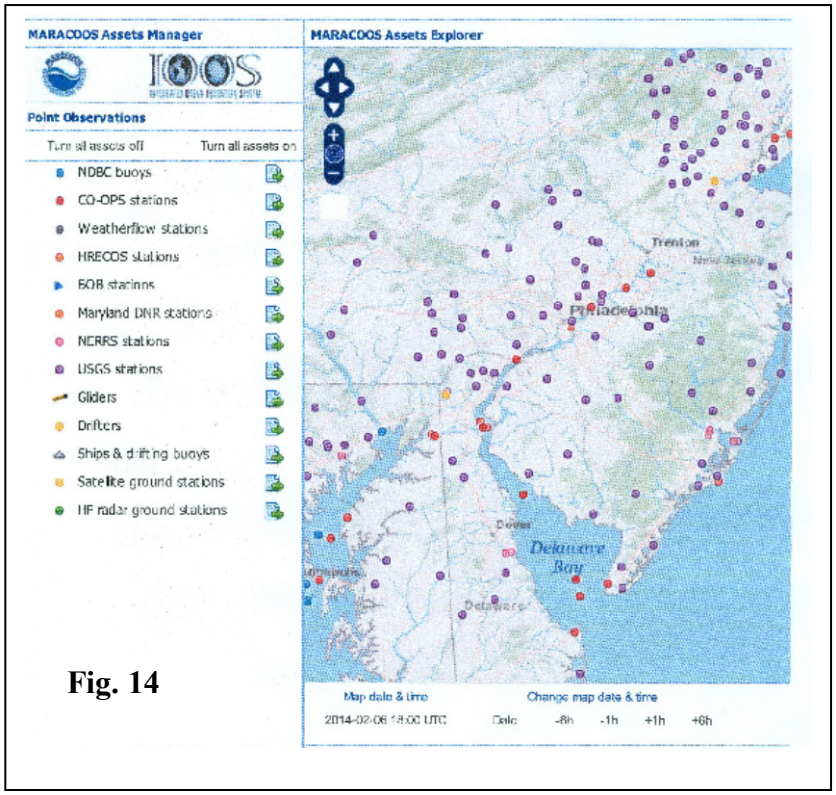


Fig. 14

- National Data Buoy Center (NDBC)

Among the data collected are sea level pressure, wind speed and direction, air temperature, sea surface temperature, significant wave height and dew point.

- Center for Operational Oceanographic Products and Services (CO-OPS)
  CO-OPS provides the national infrastructure, science, and technical expertise to monitor, assess, and distribute tide, current, water level, sea level trends, and other coastal oceanographic products and services that support the National Oceanographic and Atmospheric Administration's (NOAA's) mission. The Center also provides access to the Physical Oceanographic and Real-Time Measurement System (PORTS) that measures and disseminates observations and predictions of water levels, currents, salinity, and meteorological parameters (e.g., winds, atmospheric pressure, air and water temperatures) that mariners need to navigate safely.
- Hudson River Environmental Conditions Observing System (HRECOS)
  In 2008, the Hudson River Environmental Conditions Observing System (HRECOS) was established to provide high frequency geographically distributed real-time data between Albany and New York Harbor. It builds upon existing monitoring and observing activities on the Hudson River estuary, including the Hudson River NERR System-Wide Monitoring Program (SWMP), the US Geological Survey, and the NYS DEC Rotating Integrated Basin Studies. Data collected include: pH, DO, specific conductance, turbidity, water elevation, water temperature, air temperature, dew point, precipitation, and salinity
- National Estuarine Research Reserve System (NERRS)
  The NERRS established a System-Wide Monitoring Program (SWMP) in 1995 to develop quantitative measurements of short-term variability and long-term changes in the water quality, biological systems, and land-use / land-cover characteristics of estuaries and estuarine ecosystems. WMP currently has three major components that focus on: (1) water quality and weather; (2) biological monitoring; and (3) watershed, habitat and land use mapping. Abiotic parameters include nutrients, temperature, salinity, pH, dissolved oxygen, and in some cases, contaminants. Biological monitoring includes measures of biodiversity, habitat, and population characteristics. Watershed and land use classifications provide information on types of land use by humans and changes in land cover associated with each reserve. By using standard operating procedures for each component across all reserves, SWMP data help establish the NERRS as a system of national reference sites, as well a network of sentinel sites for detecting and understanding the effects of climate change in coastal regions."
- USGS Stations
  The United States Geological Survey (USGS) uses a variety of sensors for continuous measurement of many field parameters and chemical constituents, but six of the most commonly used are stream flow, temperature, specific conductance, DO, pH, and turbidity recorders.
- Satellite Ground Stations
  MARACOOS members own several satellite ground stations that have been directly downloading data from environmental satellites since 1992 and displaying the data on the web in real-time since 1994. These satellites deliver data to the ground stations more than 10 times per day. There are numerous types of products that can be generated for the ocean, land and atmosphere. MARACOOS products focus on ocean products, the most popular of which is the Sea Surface Temperature (SST) product which is displayed in real time for multiple areas throughout our MARACOOS study region (Cape Hatteras to Cape

May). Additional products include chlorophyll amounts, true color imagery and water turbidity.

The voluminous data collected by MARACOOS and allied groups generally undergo stringent quality control measures to ensure their validity. An example of data quality control measures for NDBC sites are summarized in Box 9.

Box 9

NDBC Data Quality Control

The objective behind the quality control procedures and techniques is that all sensors should perform within stated accuracies. The quality control procedures used by NDBC fall into two categories: completely automated and those that involve a man-machine mix. The completely automated procedures are performed at NWSTG for real-time messages used for operational forecasts and warnings. The other procedures are performed at NDBC for data that is submitted for archival.

The real-time automated procedures, performed at NWSTG, check to eliminate gross errors. Transmission parity error, range limit, and time continuity checks are performed. Relational checks, such as examining the wind gust to wind speed ratio, are performed to check the quality of both measurements. Another check ensures that the battery voltage is adequate for barometric pressure measurements.

At NDBC stricter range and time continuity limits are performed. Measurements from duplicate sensors are compared to ensure that they track together. NDBC also uses a man-machine mix of quality checks, such as graphical procedures which relate wind speed and spectral wave energy. Other man-machine procedures involve time series plots, spectral wave curves, and computerized weather maps.

When sensor or system degradation is detected, the affected data are removed before posting on the NDBC Web site or archival. The real-time processing at NWSTG is instructed to not release data from the degraded sensor. For more detailed information regarding NDBC's quality control techniques, the reader is referred to: NDBC Technical Document 09-02, *Handbook of Automated Data Quality Control Checks and Procedures*.

Data Reduction to Meet Appropriate Statistical Assumptions

The data management flowchart (Fig. 15) provides a brief summary of the methods employed to extract and utilize data that are recorded from data loggers, but can also serve to manage any large data set, however collected. A two step process is typically employed to manage, curate and store data. The "metadata" (Fig. 14) step will not be discussed in detail here, suffice it to say that metadata represent "data about data" and serve as descriptors of key attributes of the data set; e.g., where and how the data were collected, who collected it, description of the organization of the data file, etc. At the highest level of metadata organization, the investigator(s) employ rigorous quality assurance and quality control (QA/QC) procedures to identify potential errors and outliers in the data. After checking and flagging outliers to determine whether they are valid data or the result of errors (e.g., in collection or transcription); the investigator(s) move on to exploratory data analysis and the production of summary statistics. At this stage, the question is raised whether the data will require transformation to better understand and communicate patterns therein; or in order to "meet the mathematical assumptions" of the statistical procedures employed (Gotelli and Ellison 2004). The latter may include whether the data represent random, independent variables, and/or were sampled from a specified distribution, in this case, a normal distribution.

One final step that may (or may not) be taken before the data are subjected to statistical analysis is to subject the data to an ordination procedure that can help identify and extract patterns in the data that are not readily observable to the investigator. Ordination techniques are used to order (ordinate) multivariable data, by creating new variables, called principal axes along which samples are scored or ordered. Often the approach may result in a useful simplification of

patterns in complex multivariate data sets. Used in this way, ordination is a data reduction technique. Ecologists, in particular use five different types of ordination: principal component analysis, factor analysis, correspondence analysis, and non-metric multidimensional scaling, each with corresponding strengths and weaknesses (for a review, see Legendre and Legendre 1998).

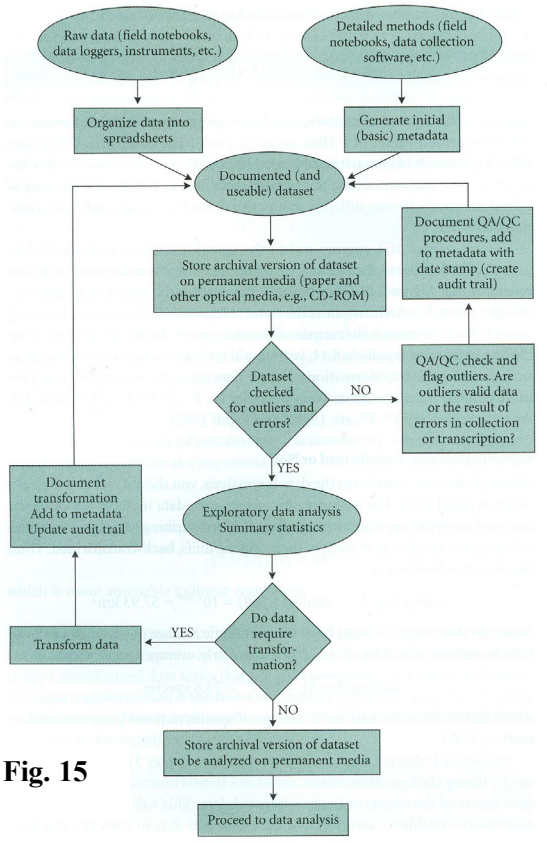


Fig. 15

APPENDIX II

Coin Tossing Exercise (adapted from www.icos.ethz.ch)

Challenge: For a given coin, what is the probability of tossing a head?

Bayesian inference relates this probability to a *prior measure of belief*. It, for example, uses prior knowledge of coins to initially assume P(heads) = 0.5. If necessary this model[1] can be revised in response to further observations (e.g., a series of coin tosses counting the total number of H and T).

Bayes' Rule

Bayes theorem, equation 1, was written as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Since P(B) =P(BA) + P(B$\bar{A}$) = P(B|A) P(A) + P(B|$\bar{A}$) P($\bar{A}$), and

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

This modified equation is referred to as *Bayes' Rule*

Coin Toss Example:

Let A = "coin is fair"; $\bar{A}$ = coin is "double headed"
Assume that the prior is: P(A) = 0.9
Trial 1. The coin is tossed and it comes up "heads" (H)

$$P(H|A)\,P(A) = 0.5*0.9 = 0.45$$
$$P(H|\,\bar{A})\,P(\bar{A}) = 1.0*0.1 = 0.1$$

$$P(A|H) = \frac{P(H|A)P(A)}{P(H|A)P(A) + P(H|\bar{A})P(\bar{A})} = \frac{0.45}{0.45+0.1} \approx 0.82$$

Note that we have used the observation H to update the model (i.e., belief in fairness of the coin).

Trials 2 and 3. The coin is tossed twice more it comes up heads each time (HH and HHH).
Second toss:

$$P(HH|A)\,P(A) = 0.25*0.9 = 0.225$$
$$P(HH|\,\bar{A})\,P(\bar{A}) = 1.0*0.1 = 0.1$$

$P(A|HH) = \frac{0.225}{0.225+0.1} \approx 0.69$; and for the 3rd toss $P(A|HHH) \approx 0.53$

Trial 4: The fourth toss also results in a head:

$P(HHHH|A)\ P(A) = 0.0625*0.9 = 0.05625$

$P(A|HHHH) = \frac{0.05625}{0.05625+0.1} = 0.36$

After 4th head in a row, the investigator is likely to be suspicious that the coin is "double-headed", or otherwise more biased, than fair!

APPENDIX III

Software Applications

There are many software applications or "packages" available for running Bayesian analyses, however for this brief summary we will concentrate on those that are in the public domain and are considered to be "free" software for Bayesian Statistical modeling. Bayesian statistical methods have become widely used for data analysis and modeling in recent years and there has been a substantive proliferation of software for doing Bayesian analysis including complex Bayesian networks, time-series, and hierarchical models to teaching and discovery packages for learning elementary Bayesian statistics. In general, classical statistical modeling methods such as regression and classification are considered to be special cases of Bayesian models. This means that there are Bayesian methods available for most classical modeling approaches such as linear and non-linear regressions, GLM, hierarchical regressions, neural network models, etc. For example, there is software available that supports Bayesian regression and classification models based on neural networks and Gaussian processes, and Bayesian density estimation and clustering using mixture models. For a fairly comprehensive overview of the various free software packages available for doing Bayesian analysis, the reader is directed to http://ksvanhorn.com/bayes/free-bayes-software.html. Probably the most popular software for running Bayesian statistical analysis is BUGS / WinBUGS (Bayesian Inference Using Gibbs Sampling). This software uses Markov Chain Monte Carlo (i.e., MCMC) methods to do a full range of complex Bayesian statistical analysis and can be downloaded for free from the BUGS Project web page. This software is highly flexible, however, one of the most useful features in WinBUGS 1.4 is the ability to interface or link WinBUGS programs from within other programs such as R, SAS, Matlab and even Excel. For example, the "R" package R2WinBUGS provides a set of functions (scripts) to call WINBUGS on a Windows system and returns the output simulations to R making it very useful for those interested in running Bayesian in the R environment. The R environment is being increasingly used by applied researchers interested in Bayesian statistics because of the ease at which one can code algorithms to sample from posterior distributions as well as well as the sheer number of tools available for Bayesian inference modeling. For example, R packages such as MCMCpack – provides model-specific Markov chain Monte Carlo algorithms for wide range of models, BayesTree – implements Bayesian Additive Regression Trees, and MCMCglmm – fits Generalized Linear Mixed Models using MCMC methods. For a full review of Bayesian modeling methods available in R, please refer to http://cran.r-project.org/web/views/Bayesian.html. For those familiar with the R environment and are interested in learning the basics of Bayesian statistical inference modeling, the R package LearnBayes contains functions for summarizing one and two parameter posterior distributions and MCMC algorithms. As would be expected, there are a wealth of resources available to Bayesian practitioners and this committee encourages further exploration of the software and methods available that support Bayesian modeling techniques.