TDL.ORG TEXAS DIGITAL LIBRARY

Journal of Regulatory Science

http:\\journalofregulatoryscience.org

Regulatory Science

Journal of Regulatory Science 04 (2016) 10–17

Regulatory Science: Mathematical vs. Statistical Models*

Richard B. Shepard[a,**]

[a]*2404 SW 22nd St. Troutdale, OR 97060 USA*

Abstract

Natural resource companies do not object to environmental regulations that are consistent and support predictability. Consistency and predictability are critical for decision making under conditions of uncertainty. Natural ecosystems are inherently variable across a broad range of temporal and spatial scales; climate change, drought, and societal desires for sustainability make people more aware of this variability. The science used for development and enforcement of environmental regulations has not kept pace with developments in ecological theory and the analytical tools capable of describing, characterizing, classifying, and predicting natural ecosystems as well as distinguishing natural variability from anthropogenic changes.

Because natural resource industries (agriculture, energy, mining) provide the base for all economic and societal activities it is critical that environmental statutes and regulations be regularly updated to use the most technically sound and legally defensible scientific knowledge and tools.

Mathematical models were the tools of choice when environmental statutes and regulations were introduced, perhaps because they were successfully applied to static components of the built environment such as buildings and bridges. While their limitations for highly variable natural ecosystems were accepted then, there is now no benefit to not replacing them with statistical models.

This paper describes limitations in policy and regulatory decision-making based on mathematical models and explains how the appropriate statistical models avoid the subjectivity and rigidity of the former. Changing the basis of determining and justifying policy and environmental regulations is consistent with the concepts of regulatory science applied to human health.

Keywords: statistics, models, prediction, water quality, natural variability

1. Introduction

Natural resource companies do not object to environmental regulations that are consistent and predictable. Consistency and predictability are critical for decision making under conditions of uncertainty. Natural ecosystems are inherently variable across a broad range of temporal and spatial scales; climate change, drought, and societal desires for sustainability make people more aware of this variability. The science used for development and enforcement of environmental regulations has not kept pace with developments in ecological theory and analytical tools capable of describing, characterizing, classifying, and predicting natural ecosystems and distinguishing natural variability from anthropogenic changes.

Because natural resource industries (agriculture, energy, mining) provide the base for all economic and societal structures it is critical that environmental statutes and regulations be regularly updated to use the most technically sound and legally defensible scientific knowledge and tools.

The origin of the term "regulatory science" is unknown. According to Wikipedia it was likely coined sometimes in the late 1970s in an undated memorandum prepared by A. Alan Moghissi[1] who was describing scientific issues that the newly formed EPA faced. During that period the EPA was forced to meet legally mandated deadlines to make decisions, and this required reliance on science less rigorous than conventional scientific development because of time constraints. One definition of regulatory science is the application of science to support policy, notably regulatory objectives. The Institute for Regulatory Science describes regulatory science as the idea that societal decisions and public communications must be based on Best Available Science and Metrics for Evaluation of Scientific Claims derived from it; i.e., as the scientific and technical foundations upon which regulations are based. Regulatory science is distinguished from regulatory affairs and regulatory law. The former is focused on the regulations' scientific underpinnings and concerns while the latter refer to the administrative or legal

**Corresponding author: Richard B. Shepard, Tel: 503-667-4517; FAX: 503-667-8863. E-mail address: rshepard@appl-ecosys.com, Address: 2404 SW 22nd St. Troutdale, OR 97060 USA

aspects of regulation (i.e., the regulations' promulgation, implementation, compliance, and enforcement). There is growing awareness and support of regulatory science in academia and some federal regulatory agencies as attempts to broaden awareness of this relatively new science continue (e.g., [2, 3]).

One aspect of regulatory science is how natural ecosystems are modeled to characterize and classify them and to forecast future states under conditions of uncertainty. With climate change apparently accelerating, long-term drought in the west and widespread focus on sustainability there is enough uncertainty that regulatory staff rely on what has been done before rather than to seek more appropriate methods.

When environmental regulations in the US were first written the most common tools for analyzing complex systems such as natural ecosystems were mathematical models[1]. Four of these models still being used are described and their limitations for use in policy-making and regulatory environments explained.

Now that abundant computing power is widely available, and statistical models appropriate for analysis of environmental data are abundant and available at no cost, they are the tools of choice for supporting environmental policies and regulations. Statistical models[2] are fit to the available data, are based on sound and proven mathematics, and when model results are interpreted using ecological theory the results are technically sound and legally defensible. The robustness and lack of subjectivity in statistical analyses help regulators make decisions more quickly and with confidence that the decisions are justified.

2. Analyzing Natural Environments

2.1. Mathematical models

Mathematical models grow out of equations that define how a system changes from one state to the next (differential equations) and/or how one variable depends on the value or state of other variables (state equations). They also can be divided into either numerical models or analytical models[4]. The model structure is determined by creating equations that express what is believed to be the relationships between a response variable and explanatory variables. Therefore, mathematical models require input data be fit to the fixed equations of the model. Often, these complex models require very large data sets as input which are costly (in time and money) or not possible to acquire so it is common to estimate or assume values for rates and constants.

Four mathematical models commonly used to inform operational, regulatory, and policy decisions are HSPF, QUAL2E, PITLAKQ, and BLM. The complexity and comprehensive inclusiveness of these models require very large amounts of data if they are to incorporate inherent natural variability. These models are great research tools to increase understanding of the mechanisms and dynamics of the systems they model, but they are inappropriate for operational, regulatory, or policy use because time and cost constraints limit the quantity of input data and because the output is determined by the structure of the equations.

Hydrologic Simulation Program–Fortran (HSPF)

One of the first environmental mathematical models is the Hydrologic Simulation Program – Fortran (HSPF). The model was developed in the early 1960's as the Stanford Watershed Model. In the 1970's, water-quality processes were added. Development of a Fortran version incorporating several related models using software engineering design and development concepts was funded by the EPA in the late 1970's. Development continues by the USGS. HSPF simulates hydrologic and associated water quality processes on pervious and impervious land surfaces, in streams, and in well-mixed impoundments for extended periods of time. The model contains hundreds of process algorithms developed from theory, laboratory experiments, and empirical relations obtained from instrumented watersheds. Dozens of data types are required as inputs to the model.

HSPF uses continuous rainfall and other meteorologic records to compute streamflow hydrographs and pollutographs. The model simulates interception of soil moisture, surface runoff, interflow, base flow, snowpack depth and water content, snowmelt, evapotranspiration, ground-water recharge, dissolved oxygen, biochemical oxygen demand (BOD), temperature, pesticides, fecal coliforms, sediment detachment and transport, sediment routing by particle size, channel routing, reservoir routing, constituent routing, pH, ammonia, nitrite-nitrate, organic nitrogen, orthophosphate, organic phosphorus, phytoplankton, and zooplankton. The model can be configured to simulate one or many pervious or impervious unit areas discharging to one or many river reaches or reservoirs. Frequency-duration analysis can be done for any time series. Any time step from 1 minute to 1 day that divides equally into 1 day can be used. Any period from a few minutes to hundreds of years may be simulated. HSPF is generally used to assess the effects of land-use change, reservoir operations, point or nonpoint source treatment alternatives, flow diversions, etc. Programs, available separately, support data preprocessing and postprocessing for statistical and graphical analysis of data saved to the watershed data management file.

Data requirements include meteorologic records of precipitation and estimates of potential evapotranspiration for watershed simulation. Air temperature, dewpoint temperature, wind, and solar radiation are required for snowmelt. Air temperature, wind, solar radiation, humidity, cloud cover, tillage practices, point sources, and (or) pesticide applications may be required for water-quality simulation. Physical measurements and related parameters are required to describe the land area, channels, and reservoirs. When data are not available, constants and rates need to be estimated by the user.

[1] In the content of this article, the mathematical models used for US environmental regulations are deterministic models that are typically composed of variables and relationships.

[2] Statistical models, while belonging to the overall category of mathematical models, are commonly distinguished from mathematical models by using random variables with a probability distribution to represent the relationship between observations on model states.

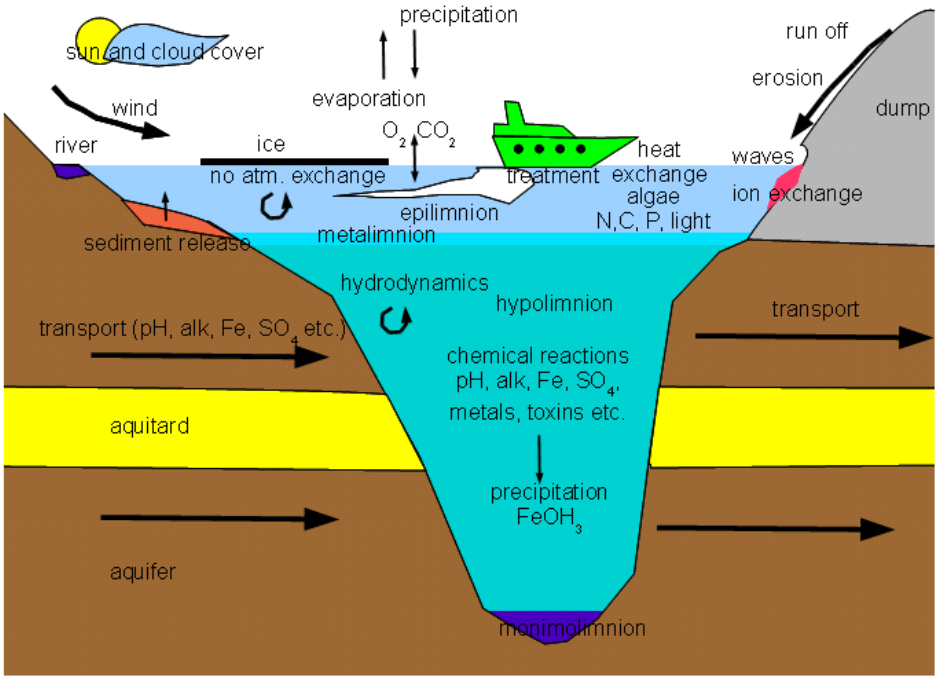

Figure 1: The physical and chemical processes modeled by PITLAKQ

Enhanced Stream Water Quality Model (QUAL2E)

A second frequently used model is the USGS's Enhanced Stream Water Quality Model, QUAL2E [5]. This is a comprehensive one-dimensional stream water quality model. It simulates the major reactions of nutrient cycles, algal production, benthic and carbonaceous demand, atmospheric re-aeration and their effects on the dissolved oxygen balance. In addition, the model includes a heat balance for computation of temperature and mass balances for conservative minerals, coliform bacteria, and non-conservative constituents such as radioactive substances. QUAL2E is intended as a water quality planning tool for developing total maximum daily loads (TMDLs) and can also be used in conjunction with field sampling for identifying the magnitude and quality characteristics of nonpoint sources. QUAL2E has been explicitly developed for steady flow and steady wasteload conditions and is therefore a "steady state model" although temperature and algae functions can vary on a diurnal basis. Although the core of the model has not changed since 1987, there have been some modifications on the interfaces and other associated tools to assist the users and the evaluation will discuss all the available versions of QUAL2E.

The conceptual representation of a stream used in the QUAL2E formulation is a stream reach that has been divided into a number of subreaches or computational elements equivalent to finite difference elements. For each computational element, a hydrologic balance in terms of flow, a heat balance in terms of temperature, and a materials balance in terms of concentration is written. Both advective and dispersive transports are considered in the materials balance. The model uses a finite-difference solution of the advective-dispersive mass transport and reaction equations and it specifically uses a special steady-state implementation of an implicit backward difference numerical scheme which gives the model an unconditional stability.

QUAL2E requires some degree of modeling sophistication and expertise on the part of a user. The user must supply more than 100 individual inputs, some of which *require considerable judgment to estimate*. The input data can be grouped into three categories: a stream/river system, global variables and forcing functions. The first group, input data for the stream/river system, describes the stream system into a format the model can read. The general variable group describes the general simulation variables such as units, simulation type, water quality constituents and some physical characteristics of the basin. The forcing functions are user-specified inputs that drive the system being modeled. The input data values depend on the type of simulation and the number of state variables used.

Pit Lake Hydrodynamic and Water Quality Model (PITLAKQ)

PITLAKQ couples the models CE-QUAL-W2 and PHREEQC and adds new functionality to account for the pit lake requirements. It includes the most important processes in pit lakes. For example, several sources of acidity such as erosion or release from submerged sediments and spatially distributed groundwater inflow help to better represent pit lake conditions. Furthermore, PITLAKQ can account for the effects of water treatment on water quality. The two-dimensional model setup with one vertical and one horizontal dimension allows having sinks and sources with defined spatial locations.

PITLAKQ models hydrodynamics, transport, heat ex-

change, wind impact, ice cover, tributary inflow, atmospheric exchanges of O_2 and CO_2, precipitation, evaporation, groundwater exchange, groundwater flow and transport, erosion (both mass transport and water quality impacts), algae and nutrients, chemical lake reactions, mineral precipitation, sediment release, deliberate treatment (on defined spatial and temporal schedules), alkalinity of sinks and sources, coupling of all processes, and user process additions and modifications (Figure 1)

Biotic Ligand Model (BLM)

The biotic ligand model (BLM) is a numerical model that couples chemical speciation calculations with toxicological information to predict toxicity of an aquatic metal on a particular species. This approach was proposed as an alternative to expensive toxicological testing, and the EPA incorporated the BLM into the 2007 revised aquatic life ambient freshwater quality criteria for copper. Research BLMs for silver, nickle, lead, and zinc are also available, and many other BLMs are under development. Current BLMs are limited to "one metal, one organism" considerations. Although the BLM generally is an improvement over previous approaches to determining water quality criteria, there are several challenges in implementing the BLM, particularly at mined and mineralized sites. These challenges include: (1) historically incomplete datasets for BLM input parameters, especially dissolved organic carbon (DOC), (2) several concerns about DOC, such as DOC fractionation in iron- and aluminum-rich systems and differences in DOC quality that result in variations in metal-binding affinities, (3) water-quality parameters and resulting metal-toxicity predictions that are temporally and spatially dependent, (4) additional influences on metal bioavailability, such as multiple-metal toxicity, dietary metal toxicity, and competition among organisms or metals, (5) potential importance of metal interactions with solid or gas phases and/or kinetically controlled reactions, and (6) tolerance to metal toxicity observed for aquatic organisms living in areas with elevated metal concentrations [6].

In 2015, the Oregon Department of Environmental Quality (DEQ) conducted an analysis of the copper Biotic Ligand Model (BLM) in preparation for replacing the state's aquatic life water quality standard for copper based on water hardness with a statewide adoption of the BLM. DEQ conducted this analysis in response to EPA's 2013 disapproval of the copper criteria Oregon adopted in 2004. The disapproved criteria were EPA's 1995 nationally recommended dissolved copper criteria for freshwater, which are dependent on the hardness of water. The EPA 1995 copper standard is still effective in most states. In 2007, EPA updated its national recommendation for copper, which uses the BLM to derive freshwater aquatic life criteria. The BLM requires 11 input parameters to derive criteria based on site- specific water chemistry. The EPA has indicated that Oregon's adoption of the BLM would remedy their disapproval action [7].

Because of the number of model input parameters, a major objective of the Oregon DEQ analysis was to evaluate methods to *estimate values for missing model inputs*. A method for estimating geochemical ion concentrations using specific conductance measurements was adopted. DEQ also adopted an approach to simplify large geographic scales by combining EPA Level-III Ecoregions into four physiographic BLM assessment regions for evaluating potential regional estimates of BLM parameters or criteria where model data are insufficient or absent. Unfortunately, this does not fit the generally accepted definition of "site specific" and causes model outputs to miss inherent natural variability. There were a limited number of locations and sampling events that had measured data for all of the required BLM input parameters. Therefore, to derive BLM criteria, estimates of missing parameters will frequently be required. Also, DEQ's analysis verified that BLM criteria calculations are most sensitive to DOC and pH. Consequently, estimating values for DOC or pH results in significant uncertainty in the accuracy of BLM criteria. DEQ's analysis indicates there are no routinely collected surrogate parameters that can be used to accurately estimate DOC or pH.

2.2. Statistical models

The descriptions of the four mathematical models reveal common characteristics that should concern the regulated public, consultants and attorneys who assist them in obtaining permits and demonstrating compliance, regulators, and policy makers. For example:

- Each requires extensive input data (a large number of variables and many replicate values over time from each location). This requirement works for academic and government scientists who can devote the time and effort to obtain sufficient data so the model run yields non-trivial results. But, this does not work in a regulatory context where business needs and regulatory staff require decisions in shorter time frames.

- Each of the mathematical models requires estimates or assumptions for missing data and the spatial scales are likely to be too large or two small for regulatory decisions about a particular project or local population of fish.

- Each numeric model has a static structure (QUAL2E has not changed since 1987) to which highly variable data need to be fit. It is not surprising when it is discovered that model output does not mach measurements and observations in the stream, river, lake, or reservoir of interest.

Statistical models address concerns such as characterization of the data (estimates of the expected value, variance, skewness, etc.), estimation of the probabilistic future behavior of a system based on past behavior, extrapolation or interpolation of data based on probability, error estimation of observations, or spectral analysis of data. Unlike mathematical models, the statistical model is fit to the existing data. It is possible to try several such models and mathematically determine which one best fits the data.

There are two main statistical paradigms, or approaches, and each has a role in the analysis of environmental data.

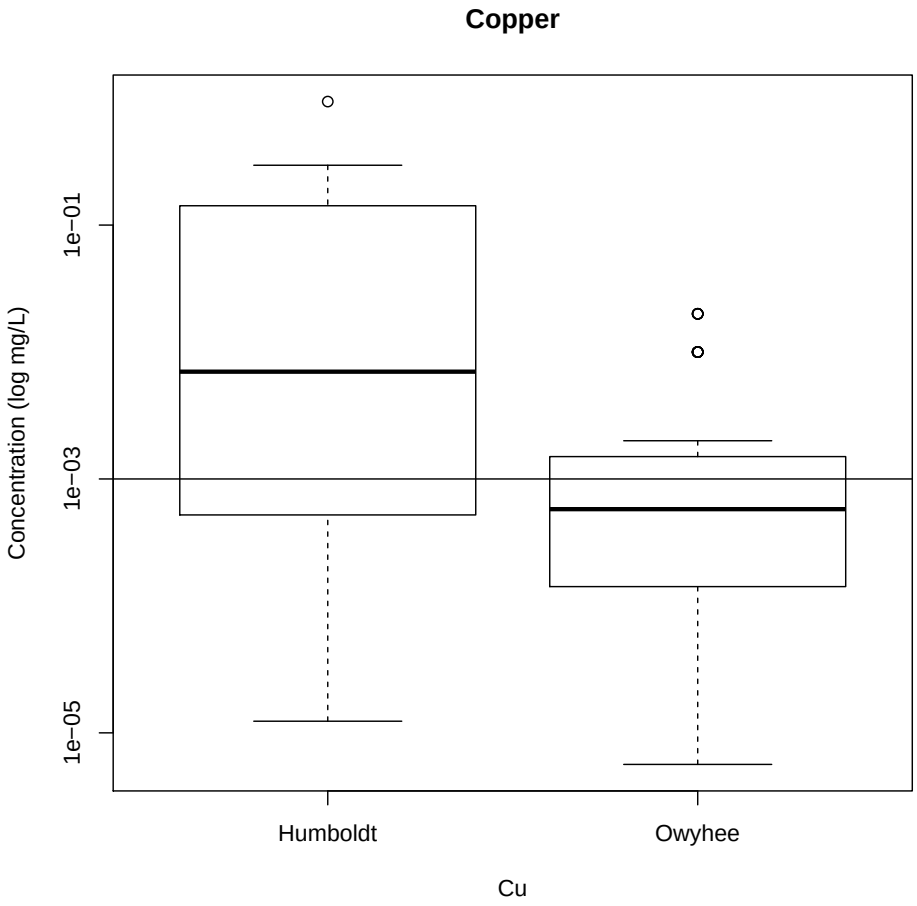

Figure 2: Boxplots of copper concentrations from samples taken from the Humboldt and Owyhee Rivers in northeastern Nevada between 1980–2013. See the text for interpretation of the data

The *frequentist* paradigm is the most familiar because it is taught in all basic statistics courses as part of a science or business curriculum. This paradigm is based on the expected frequencies of values if the population is described by a specific probability distribution. There are two types of frequentist analysis: null hypothesis significance testing and information theoretic.

The null hypothesis approach tests two hypothesis, the null and an alternative, and calculates the probability of observing the data or more extreme values if the null hypothesis is true. The alternative hypothesis is not tested. The standard acceptance threshold is 5%; the familiar p= 0.05. Unless environmental data have been suitably transformed, the null hypothesis significance testing approach has limited use with environmental data. When the probability is greater than 0.05 the null hypothesis is not accepted but only not rejected. The *information-theoretic* approach uses maximum likelihood estimation (MLE) to evaluate which of multiple hypotheses is best fit by the data. The MLE approach provides greater flexibility than does the frequentist approach in analyzing environmental data. Neither type of null hypothesis significance testing can answer questions about biological data, such as calculating the probability of a species being present at a habitat regardless of being observed there on any given site visit.

The *Bayesian* paradigm incorporates existing knowledge into the prediction of future conditions. While this might seem counter-intuitive or flaky it is a robust approach well suited to many environmental data sets, particularly biological ones. Perhaps it is more easily accepted when we understand that we subjectively use this approach to make decisions about our actions. We fish and hunt where we have been successful in the past and we select commuting routes and times to avoid congestion and delays we have experienced in the past. What Thomas Bayes did was to put this knowledge in a mathematical probability format.

Data characterization

The first thing the environmental data analyst does with a data set is to describe and characterize it visually and numerically. Environmental data differ from business, financial, and social data that are often statistically analyzed. Therefore, environmental data characterization is needed to determine both what further analyses can validly be done and which models

best fit the data.

Geochemical data in water, sediments, soils, and rocks are continuous data with true zeros but they are not normally distributed. That is, the familiar bell-shaped curve cannot describe their possible values. Chemical concentrations cannot be less than 0.0 (regardless of units of measure), frequently have long tails on the right side from infrequent high values, and can have values below laboratory detection limits. They also tend to be collected at irregular periods. Geochemical data represent parts of a whole which is expressed in the units used to describe their concentration in a medium: mg/L (milligrams per liter), ppm (parts per million), ppb (parts per billion), and similar. Biologic data are either integer counts or presence-absence records.

The most useful visual characterization of geochemical data is the box-and-whisker plot, commonly called a boxplot. Figure 2 is a boxplot of several hundred copper concentrations from each of two river systems (the Humbolt and Owyhee Rivers in northeastern Nevada).

There are five components to each boxplot: the minimum and maximum values represented by the whiskers (the short horizontal lines at the end of the dotted vertical lines), the range of the middle 50% of values (the second and fourth quartiles) represented by the box, and the median concentration (the heavy horizontal line in the box). The open circles are outlying values that might (or might not) have significance. The boxplots show the median, range, and variability of the measured values. The thin horizontal line across the entire figure at 0.001 mg/L) is the analytical laboratory's method detection limit; values below that line were not directly measured[8]. Boxplots present a full description and characterization of the data that is easily understood by all viewers. From policy and regulatory perspectives the most visible difference in the two rivers raises the question why they are so different. What explanatory variables caused the differences in copper concentrations in the two rivers, and do differences reflect inherent variability or anthropogenic influences?

One very important characterization of environmental chemical and biological data sets is variability. Not only does understanding inherent natural variability enhance decisions about policy and regulations, it also identifies which class of statistical models are most likely to produce technically sound and legally defensible results. Statistical models of all three paradigms are based on probability distributions and there are many from which to select. Wikipedia lists 34 discrete probability distributions and 106 continuous distributions; even more exist and each has specific parameters and applicability to a wide range of data types and characteristics. This is why the most appropriate model can be fit to any data set, unlike mathematical models that have fixed equations to which the data must be fit.

Answering questions

There are two common type of environmental data: measured and nominal (named). A broad range of statistical models for comparing environments and forecasting changes are available to analyze these mixed effects data sets. Measured quantities of physical, chemical, and biological data are familiar to everyone. However, explanatory named variables can be important in explaining differences in response variables. Common examples of nominal variables include season, soil type, stream name, vegetation type. These variables are common in data applicable to the Clean Water Act, Endangered Species Act, National Environmental Policy Act, Superfund sites, and solid waste disposal areas. When nominal variables are incorporated into additive mixed effects models the amount of the observed response variability they explain is quantified.

The class of statistical models used to evaluate cause and effect is that of regressions. There are linear and non-linear regression models, including survival models (see, for example, [9]) and less-widely known regression types such as quantile regression. Linear regressions are commonly applied to environmental chemical and biological data; for example species-habitat relationships of Greater sage-grouse or Lahontan cutthroat trout. Applying the correct type of regression model to the available data and the concern to be addressed provide strong justification for policy, regulatory, and operational decisions.

The generalized regression model, $Y = \alpha + \beta X$, is a straight line relating the response (dependent) variable Y on the vertical axis against the explanatory (independent) variable X on the horizontal axis. The constant α is the Y-intercept (i.e., the predicted response variable value when the explanatory variable is zero) and the constant β is the slope of the line that represents the mean (average) population size for a given habitat size.

When there are abundant data a single regression line representing the mean response value across the range of explanatory variable values provides limited value. It is better to apply a quantile regression that describes relationships of various response variable levels to the range of explanatory variables. Figure 3 shows data from an aquatic environmental risk assessment.

Notice the different slopes of the two regression lines: as the stressor value increases the highest 10 percent of the response variables decreases more quickly than does the average response. For a detailed explanation of the increased insights provided by quantile regression see Figure 1 in [10]. They studied the relationship of Lahontan cutthroat trout densities (the Y variable) as a function of the ratio of stream width to depth (the X variable) over 7 years and 13 streams. The mean response line is almost flat while the higher percentiles of fish density decreased with increasing width-to-depth ratios and the lower percentiles increased with increasing width-to-depth ratios. This shows that basing policy or regulatory decisions on a single, mean regression line will not necessarily reflect reality. Quantile regressions have also been used to quantify ranges of biotic responses to arsenic in soils [11] and sage-grouse to habitat densities [11].

Forecasting future states to assess change can be accomplished using regression models by using the regression line, or lines, relating the response variable to the explanatory variable and by using multivariate regression models that can accommodate one or more response or explanatory variables, or multiple variables in both categories (e.g., [12, 13]). There is also the

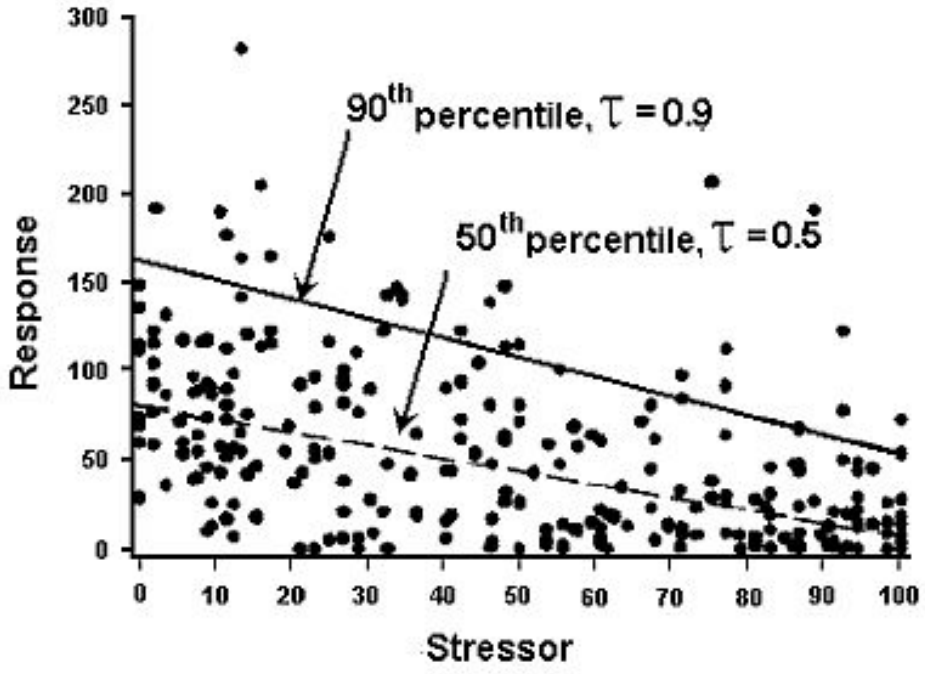

Figure 3: Quantile regression plot illustrating different relationships of a response variable to a stressor variable for the highest 10^{th} percentiles of the response variable compared to the mean of the response variable

more familiar time series approach to analyzing change over time and predicting future values.

Environmental data tend to have unequally spaced collection periods. Many time series models assume equally spaced data and they produce incorrect results with environmental data. There are many suitable models for time series analysis of environmental data based on the frequentist, information-theoretic, and Bayesian statistical paradigms (e.g., [14–17]). Time series analyses extend past plotting the data. Seasonal trends need to be removed before any overall trend can be determined and multiple time series can be compared to determine both how and why they differ (or are similar).

There is another aspect of environmental data that is a comparatively new concept: that measurements and observations of environmental chemicals and biota are components of the whole, not the complete set of chemicals or species. When fish are censored as an aquatic life beneficial use only some of the species present are counted. Geochemicals are most frequently found as multi-chemical compounds, not free ions, and represent only a portion of the whole chemical composition of the water sample. The units of concentration (for example, mg/L) reflect this reality. By definition, 1 liter of water weighs 1 kilogram (1,000 milligrams) so the total weight of all chemical elements has an upper bound of 1,000 mg. This means that if the concentration of sodium chloride (NaCl) increases, some other constituent (or multiple constituents) must decrease in concentration by the same weight. This dynamic is addressed by the compositional data analysis class of statistical models.

While compositional data was recognized early in the last century it was not until mathematical geochemists took advantage of the work of [18] and developed the statistical models that provide the proper way to look at geochemical and public[3] data to produce more realistic results. Biotic surveys such as those common under the ESA and in species-habitat relationships also benefit from being analzed as components of a larger whole. By doing this spatial and temporal variation in the ratios of the components provides mathematically and ecologically sound insights into the dynamics of the organisms and the ecosystems in which they live. Analyzing the function (energy transfer and nutrient cycling) of aquatic macroinvertebrate communities as compositions is more reliable and accurate than are structural comparisons based on taxonomic identifications and provides a robust alternative to chemical maximum concentration limits as a method of setting water quality standards for aquatic life[19]. This is a powerful tool for developing policies and regulations because it permits separation of anthropogenic changes from inherent variability of aquatic ecosystems.

3. Conclusions

Too many environmental policy and regulatory decisions are based on the output of mathematical models. These models use fixed mathematical equations to describe how an ecosystem behaves or how a variable changes. These assume that we know these behaviors and that they are relatively static. Ecologists know this is not the case; natural ecosystems change constantly and at different scales. For example, sand bars in the lower Columbia River move about 1 meter per day during the

[3]Economic, political, and social data collected by governments.

late summer low flow period, even when the dams are not releasing water through hydroelectric generating turbines, and the water temperature of headwater mountain streams can vary by as much as 20°C per day. Fixed-equation mathematical models of natural ecosystems also require many input variables and have rates and constants that need to be estimated, assumed, or approximated. These factors make them vulnerable to challenge by project or regulation opponents. Because of the complexities of mathematical models and the large amounts of data required they take a lot of time, effort and money to yield a run. While state-of-the-art 40-50 years ago they are not as useful for informing environmental policies and regulations today. Mathematical models are effective with static systems such as buildings and bridges, but not with highly variable systems such as the natural world.

Statistical models are based on probability distributions, each of which assumes certain data characteristics. There are many probability distributions whose assumptions are met with data from specific environments. This makes the choice of statistical model and the analytical process both technically sound and legally defensible. The three statistical paradigms include analytical models that incorporate variability at different scales in the data and allow separation of anthropogenic from natural changes. Predictions of future states and tests of reality with those predictions are key to compliance with regulations under the Clean Water, Endangered Species, and National Environmental Policy Acts, among others. While mathematical models define the causal variables for observed effects in their equations, statistical models examine a set of potential explanatory variables (singly and in various combinations) to determine which best fit the available data.

Policies and regulations can be supported by demonstrably best available science by taking advantage of current knowledge in environmental data analytical methods and our understanding of ecological theory applicable to all natural ecosystems.

4. Article Information

The article was received May the 3rd, 2016, in revised form July the 19th, 2016 and available on-line September the 15th, 2016.

References

[1] [link].
URL http://www.nars.org/

[2] A. Moghissi, S. Straja, B. Love, D. Bride, R. Stough. Innovation in regulatory science: evolution of a new scientific discipline. 16 (2014) 155-165.

[3] A. Moghissi, N. Gurudas, S. Pei, D. McBride, N. Swetnam. Scientific ethics: emphasizing regulatory science requirements. 17 (2015) 61-73.

[4] [link].
URL http://serc.carleton.edu/introgeo/mathstatmodels/index.html

[5] USEPA 1997. The enhanced stream water quality model qual2e and qual2e-uncas: Documentation and user model. Technical Report EPA/600-3-87/007, Environmenal Reseearch Laboratory, US Environmental Protection Agency.

[6] K. Smith, L. Balistrierib, A. Todda, Using biotic ligand models to predict metal toxicity in mineralized systems, Applied Geochemistry 57 (2015) 55–72.

[7] J. McConaghie, A. Matzke. Technical support document: An evaluation to derive statewide copper criteria using the biotic ligand model. Technical report, Oregon Department of Environmental Quality. 2016.

[8] [link].
URL http://www.appl-ecosys.com/publications/censored-chemical-analyses.pdf

[9] F. Harrell Jr.. Regression Modeling Strategies. Springer. 2010.

[10] B. Cade, B. Noon, A gentle introduction to quantile regression for ecologists, Frontiers in Ecology and Environment 1(8) (2003) 412–420.

[11] B. Cade, Estimating equivalence with quantile regression, Ecological Applications 21(1) (2011) 281–289.

[12] B. Everitt, T. Hothorn. An Introduction to Applied Multivariable Analysis With R. Springer. 2011.

[13] P. Mielke Jr., K. Berry. Permutation Methods: A Distance Function Approach, Springer Series in Statistics. Springer. 2001.

[14] J. Bobbin, F. Recknagel . Mining water quality time series for predictive rules of algal blooms by genetic algorithms. In Proc. of the Int. Conference MODSIM 99, 1999.

[15] A. Eckner. A framework for the analysis of unevenly spaced time series data. In development; not yet submitted for publication. 2014.

[16] R. Krzysztofowicz, Bayesian models of forecasted time series, Water Resources Bulletin 21(5) (1985) 805–814.

[17] D. Lettenmaier, Detection of trends in water quality data from records with dependent observations, Water Resources Research 12(5) (1976) 1037–1046.

[18] J. Aitchison. The statistical analysis of compositional data, Monographs on statistics and applied probability. Chapman & Hall. 1986.

[19] [link].
URL http://www.appl-ecosys.com/publications/biota-to-set-wq-standards.pdf