Ecological Indicators 74 (2017) 62–72

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Estimating linear temporal trends from aggregated environmental monitoring data

Richard A. Erickson[a,*], Brian R. Gray[a], Eric A. Eager[b]

[a] Upper Midwest Environmental Sciences Center, US Geological Survey, La Crosse, WI, United States
[b] Department of Mathematics and Statistics, University of Wisconsin – La Crosse, La Crosse, WI, United States

ARTICLE INFO

Article history:
Received 25 June 2016
Received in revised form 25 October 2016
Accepted 28 October 2016

Keywords:
Aquatic invasive species
Long term monitoring
MARSS
Mississippi River
State-space model
Trend estimation

ABSTRACT

Trend estimates are often used as part of environmental monitoring programs. These trends inform managers (e.g., are desired species increasing or undesired species decreasing?). Data collected from environmental monitoring programs is often aggregated (i.e., averaged), which confounds sampling and process variation. State-space models allow sampling variation and process variations to be separated. We used simulated time-series to compare linear trend estimations from three state-space models, a simple linear regression model, and an auto-regressive model. We also compared the performance of these five models to estimate trends from a long term monitoring program. We specifically estimated trends for two species of fish and four species of aquatic vegetation from the Upper Mississippi River system. We found that the simple linear regression had the best performance of all the given models because it was best able to recover parameters and had consistent numerical convergence. Conversely, the simple linear regression did the worst job estimating populations in a given year. The state-space models did not estimate trends well, but estimated population sizes best when the models converged. We found that a simple linear regression performed better than more complex autoregression and state-space models when used to analyze aggregated environmental monitoring data.

1. Introduction

Environmental monitoring programs often seek to quantify temporal trends (e.g., annual linear changes over means) (Lindenmayer and Likens, 2010). These trends may be reflected in biotic observations (e.g., Weissteiner et al., 2011; Gröger et al., 2011; Peng et al., 2012; Murphy et al., 2014) or abiotic observations (e.g., Rodrigues et al., 2011; Mózner et al., 2012; van Puijenbroek et al., 2015). The results of trend analysis may provide insight and guidance for managers and inform them if action is needed or previous management has been successful (e.g., van Puijenbroek et al., 2015). Quantitative trend analysis is typically assumed to require the use of statistical models. However, data limitations may hamper efforts to estimate trends with complex models, even though these models may more accurately represent both the data collection methodology and actual ecological and environmental processes generating the data (Ward et al., 2014).

Environmental scientists and managers often face data limitations for time-series models. Data can be aggregated (van Puijenbroek et al., 2014). Aggregating or pooling data collapse the data by taking the mean of a subgroup of samples. For example, data can be aggregated spatially (e.g., all samples within a lake are pooled), temporally (e.g., all samples taken within a year are pooled), or taxonomically (e.g., all observations from a genus are pooled). Such aggregation melds variation in means (i.e., process variability) with sampling variation. Data may be aggregated or pooled for a wide range of reasons. Hierarchical structures may be difficult to model statistically (Gelman and Hill, 2007; Zar, 2010; Gotelli and Ellison, 2013). Conclusions may need to be drawn across sampling boundaries and borders (van Puijenbroek et al., 2014). Aggregated data, however, often lose valuable information about sample size, which is important for some models such as those assuming a binomial distribution for count data (Suter, 2001). Modeling approaches (e.g., state-space model) exist that can regain some of this lost information (Holmes et al., 2012).

Another limitation can be the duration of monitoring, which may limit the use of time-series models. Time-series models require a minimum time-series length that depends upon model complexity and data variability (Shumway and Stoffer, 2011) These datasets may also have high sampling variability. Researchers and managers are typically interested in variability in biological processes per se rather than in variability induced by sampling. Additionally,

* Corresponding author.
E-mail addresses: rerickson@usgs.gov, raerickson@gmail.com (R.A. Erickson).

http://dx.doi.org/10.1016/j.ecolind.2016.10.036
1470-160X/Published by Elsevier Ltd.

the processes that lead to changes in abundance and occurrence are typically assumed not to be additive. However, most statistical models contain some assumptions of linearity (Zar, 2010; Shumway and Stoffer, 2011; Ward et al., 2014; Auger-Méthé et al., 2015).

Methods exist such as state-space models that model both the underlying processes (e.g., ecological trend) and observation or sampling error. We distinguish observation and sampling error because aggregating data create sample means and treats the means as the "data", which are a step removed from the observation process. Complex models, such as state-space models, come at a cost of complexity that may or may not improve the model's ability to estimate trends (Ward et al., 2014). Besides state-space models, other models exist to estimate trends from aggregated time-series data. Simple linear regression allows for trends from aggregated data to be estimated over time. Auto-regressive models allow for temporal correlations while also estimating trends. Random walk models describe systems that vary by chance as well as "drift". State-space models may consider all of these features, but these come at the cost of possibly over fitting data (Shumway and Stoffer, 2011; Ward et al., 2014). Furthermore, state-space models may be challenging to formulate. Hence, their use in routine monitoring settings, where hundreds of species may be monitored and potential modeled, may be questionable (Auger-Méthé et al., 2015).

Herein, we seek to compare the above mentioned statistical models (i.e., linear regression, auto-regressive model, and state-space models) as methods for estimating trends in transformed means from a relative abundance observation time-series and aggregated discrete data. We assumed generating processes that were Bernoulli or negative binomial with means that varied by sampling year. This assumption matched the sampling method from which we sought to estimate trends. Simulations under these assumptions allowed us to evaluate the model's ability to recover parameters from a known set of values. We also used simulation experiments to compare the effects of different time-series lengths.

We then compared the performance of different parameter estimation approaches using fish and vegetation data from the Upper Mississippi River that was collected as part of the Long Term Resource Monitoring element of the Upper Mississippi River Restoration Program (LTRM).[1] The LTRM collects these data because the river's is "a nationally significant ecosystem and a nationally significant commercial navigation system" (33U.S.C.§652, 1983). The goals of the LTRM include supporting decision makers by "monitoring resource change" and developing "a better understanding of the ecology Mississippi River System" (http://www.umesc.usgs.gov/ltrmp/about_us_mission.html). Trend estimation is one method for estimating change, which may also provide insight into the ecology of the Mississippi River System. Understanding how different trend estimation methods perform will benefit programs such as LTRM because state-space models have not been previously evaluated over this sampling design (E. Ward, personal communication; Holmes et al., 2012, 2014; Ward et al., 2014).

2. Methods

We compared five methods for estimating temporal trends from two generating processes. Both of the simulated processes yielded average of random outcomes from an assumed probability distribution based upon Eqs. (2) and (4) which, in turn, were based upon sample data. The first generating process assumed an underlying trend that had samples drawn from a negative binomial distribution (Ickes et al., 2014, c.f. http://www.umesc.usgs.gov/reports_publications/ltrmp/fish/fish_methods.html). The second generating process assumed an underlying trend that had samples drawn from a Bernoulli distribution (Yin et al., 2000, c.f. http://www.umesc.usgs.gov/reports_publications/ltrmp/veg/srs_methods.html). We then compared the ability of five different models to recover trend estimates from these datasets: a simple linear regression, an auto-regressive model with trend, a state-space model with strong density dependence, a state-space model with estimated density dependence, and a state-space model with a random walk. The three state-space models make different assumptions about the abundance of the population relative to density limitations (Table 1). The random walk model assumes a growing or shrinking population (i.e., "drift" or trend) with no density limitations, the density dependent model estimates the effect of density on population growth, and the strong density dependent model assumes a population that has its growth strongly limited by density (Fig. 1). Last, we evaluated these five models on long-term monitoring data from the Upper Mississippi River.

Table 1
Parameters used in generating models. For the fish simulation type, $U = 0$ had $X_0 = -2$ and 1.5, $U = -0.05$ had $X_0 = 4$ and 1.5, $U = 0.05$ had $X_0 = -4$ and 1.5. Otherwise, all permutation combinations were used.

Parameter	Name	Values used	Simulation type
X_0	Intercept	logit({0.05, 0.5, 0.95})	Vegetation
X_0	Intercept	{−4.0, −2.0, 4.0})	Fish
B	AR term	{0, 0.5, 0.9}	Vegetation
B	AR term	{0, 0.9}	Fish
U	Trend (slope)	{−0.05, 0, 0.05}	Both
Q	Error (SD)	{0.025, 1}	Vegetation
Q	Error (SD)	{0.1, 1}	Fish
T	Length of simulation	{10, 20}	Vegetation
T	Length of simulation	{10, 25}	Fish
r	NB dispersion parameter	{10,000, 1/8}	Fish
n	Sample size	{30, 450}	Both

2.1. Generating models

We assumed our time-series data had a known generation process, specifically a time-series with auto-regression with a trend

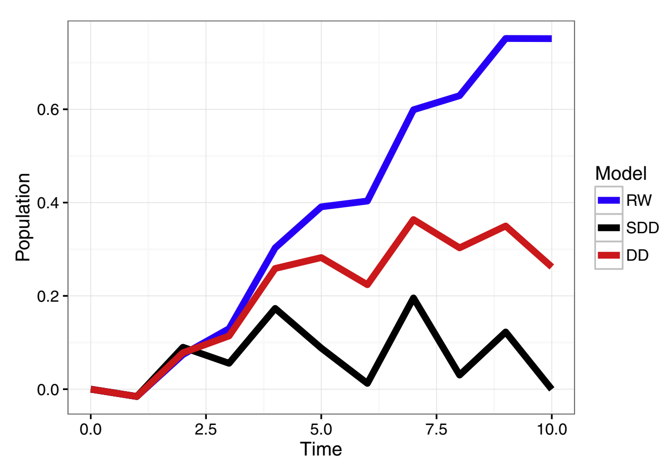

Fig. 1. Conceptual example of the random walk (RW) generating process, strong density dependence (SDD) generating process, and density dependence (DD) generating process. Note that the DD model population size is between the SDD and RW population sizes.

[1] The U.S. Army Corps of Engineers' Upper Mississippi River Restoration (UMRR) Program Long Term Resource Monitoring (LTRM) element is implemented by the U.S. Geological Survey, Upper Midwest Environment Sciences Center (UMESC), in cooperation with the five Upper Mississippi River System (UMRS) states of Illinois, Iowa, Minnesota, Missouri, and Wisconsin. The U.S. Army Corps of Engineers (Corps) provides guidance and has overall Program responsibility.