Ecological Indicators 84 (2018) 765–792

Contents lists available at ScienceDirect

Ecological Indicators

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

Research paper

Evaluating statistical models to measure environmental change: A tidal turbine case study

Hannah L. Linder[a,*], John K. Horne[b]

[a] Oceana, 1350 Conneticut Ave NW, Washington, DC, 20036, USA
[b] University of Washington, School of Aquatic and Fishery Sciences, Box 355020, Seattle, WA, 98195, USA

ARTICLE INFO

Keywords:
Before-After monitoring
Intervention analysis
Regression
Statistical model evaluation
Marine renewable energy
Time-series
Nonparametric
Semi-parametric
State-space models
Power analysis
Forecast

ABSTRACT

Statistical models are routinely used in Before-After monitoring studies to detect, quantify, and forecast environmental change caused by natural or anthropogenic disturbances. For monitoring programs that have not established standard statistical procedures, an evaluation of models' abilities to measure change over a range of scenarios is vital to develop best practices for analyzing monitoring data. A comprehensive evaluation was developed and applied using Marine Renewable Energy (MRE) tidal turbine site as a case study of a developing industry with no standard monitoring methods. Before-After monitoring datasets that contained change were simulated using normal and non-normal empirical baseline data collected at a MRE site. Thirteen regression models from three classes were evaluated: 6 generalized regression models, 4 time series models, and 3 nonparametric models. Intervention Analysis was used to fit models to five change scenarios that included three amplitudes of change and a lag in the onset of change. A power analysis was used to evaluate model ability to detect change. Residual error diagnostics were used to quantify model fit and forecast accuracy. Parametric models that did not include lagged dependent variables were the most capable at detecting change. A comparison of the fit and forecast metric results indicated that deterministic time series models in conjunction with semi-parametric generalized regression models provided a robust and informative quantification of change. Nonparametric models most accurately forecasted change. The results provide insight on model behavior, which is used to recommend specific models to measure change in the case study data. These recommendations form best practices for analyzing monitoring data, which enables comparisons among monitoring sites and reduces uncertainty when quantifying environmental effects.

1. Introduction

The need to detect and measure change is ubiquitous in all environmental monitoring, with a common goal of preventing or minimizing detrimental effects on natural resources (e.g., Ferretti, 1997; Lovett et al., 2007; Ingersoll et al., 2013). Criteria used to define change and the resulting estimates of amplitudes and shapes of change affect inferences about the observed ecosystem (Underwood 1992; Nuno et al., 2015). Monitoring programs often fail to accurately measure change, either natural or anthropogenic, due to high costs of sampling, data analyses, time-constraints, and unclear objectives (Gitzen et al., 2012; Lindenmayer et al., 2012). To ensure efficient and accurate measures of change, it is critical to define environmental monitoring objectives and to develop best practices to achieve those objectives.

Statistical models are typically used to fulfill objectives of environmental monitoring that include detecting, quantifying, and forecasting change. Here, a statistical model is defined as a mathematical algorithm used to estimate response variable(s) assuming a probabilistic distribution with a given mean and variance (cf. Gitzen et al., 2012). We specifically investigate regression models; statistical models that use a set of explanatory variables to estimate expected response variable value(s). Statistical regression models are conventionally used in Before-After Intervention Analysis (IA) studies to detect hypothesized environmental change in operational monitoring (i.e., post-disturbance) data against a null hypothesis of no change from baseline conditions (Schmitt and Osenberg, 1996; Stewart-Oaten and Bence, 2001). A model's measure of change can then be used to assess perceived effects of a disturbance and to inform decisions on future monitoring and mitigation plans (Schmitt and Osenberg, 1996; Hewitt et al., 2001; Shumchenia et al., 2012). Forecasting, a third objective of monitoring, is used to develop future monitoring, operational modifications, and/or mitigation efforts. Model forecasts may predict future trends and provide advance warning that threshold conditions are approaching (Magurran et al., 2010; Dornelas et al., 2013; Lindenmayer

* Corresponding author.
E-mail address: hlinder33@gmail.com (H.L. Linder).

http://dx.doi.org/10.1016/j.ecolind.2017.09.041
Received 9 March 2017; Received in revised form 20 September 2017; Accepted 24 September 2017
1470-160X/ © 2017 Elsevier Ltd. All rights reserved.

et al., 2012).

Choice of a statistical model will affect estimates of change, and consequently, the operational modification, mitigation efforts, and management of a disturbance on the environment (Thomas and Martin, 1996; Nuno et al., 2015). Evaluations of statistical models to identify models that accurately quantify pattern have been repeatedly conducted in environmental fields including fisheries (e.g., Olden and Jackson, 2002), air pollution (e.g., Jerrett et al., 2004), and agriculture (e.g., Michel and Makowski 2013). Previous model evaluation studies have often been limited to a specific class of models, a single dataset, and/or a single monitoring objective. To be complete, a model evaluation must encompass a wide range of model classes and scenarios of Before-After change to identify accurate and robust models. A thorough model evaluation is especially needed for developing industries such as marine renewable energy (MRE) that do not have standards for the analysis of baseline or operational monitoring data.

Statistical models used to measure change in MRE Before-After studies are commonly limited to generalized regression models (see definition in Linder et al., 2017). To meet the monitoring objective of detecting change, hypothesis tests (i.e., ANOVA, Mann-Whitney *U* test) (e.g., Hammar et al., 2013) or indicator variables in generalized regression models (e.g., Bergström et al., 2013; Petersen et al., 2011) have been used. Generalized regression models used to quantify change by estimating trends and patterns in MRE monitoring studies have included: linear regression (e.g., Hammar et al., 2013; ORPC, 2014), Generalized Linear (Mixed) Models (GL(M)Ms) (e.g., Embling et al., 2013; Stenberg et al., 2015), and Generalized Additive (Mixed) Models (GA(M)Ms) (e.g., Tollit and Redden, 2013). These models are also used to predict data patterns in MRE Before-After monitoring studies (e.g., Mackenzie et al., 2013; Waggitt et al., 2014; Warwick-Evans et al., 2016). The limited number and low diversity of statistical models previously used highlights the need to evaluate other candidate models capable of measuring change in MRE monitoring data.

Objectives of the current study are to evaluate statistical models for their ability to detect, quantify, and forecast change, and to recommend best practices for environmental monitoring using MRE baseline data as a case study. At this time, Mackenzie et al. (2013) is the only study that evaluated three models used to detect shifts in abundances and distributions of marine animals at a MRE site. In the current study, active acoustic data are used in a scenario analysis to evaluate model ability to measure five different forms of change caused by hypothesized effects of MRE development. Generalized regression models, time series models, and nonparametric models, previously evaluated in a MRE baseline monitoring study (Linder et al., 2017), are all capable of measuring change over time (Chandler and Scott, 2011), and encompass a broad range of advantages and constraints for meeting monitoring objectives. Models within these three classes are evaluated, and results of the model evaluation are used to recommend statistical model(s) that are able to detect, quantify, and forecast change in monitoring data.

2. Methods

2.1. Case study rationale

MRE development, which includes tidal, current, wave, and offshore wind, provides a timely case study of anthropogenic disturbance and exemplifies the need for efficient and effective environmental monitoring methods. Although MRE technology is rapidly expanding, projects within the United States are primarily in demonstration stages rather than commercial scale operations, such as the Burbo Bank offshore wind farm (http://assets.dongenergy.com/DONGEnergyDocuments/Project%20Summary_BurboBank.pdf) and the MeyGen tidal project (http://www.meygen.com/) in Europe. This development lag is attributed, in part, to the sampling requirements for both baseline and operational monitoring in the US. To comply with the National Environmental Policy Act (NEPA), an environmental assessment prior to project installation is mandatory to obtain permits for any MRE project (Federal Energy Regulatory Commission, 2008; US Department of Energy, 2009; Portman, 2010). Plans for post-installation, compliance monitoring are commonly required for commercial leases of MRE sites (Bureau of Ocean Energy Management, 2016). At this time, there are no specific monitoring requirements under these regulations, yet the standardization of environmental monitoring has been cited as critical for expediting MRE permitting and advancing MRE development (Dubbs et al., 2013; Copping et al., 2016). There have been limited efforts to develop sample design and data collection guidelines for MRE sites (e.g., McCann 2012; Copping et al., 2014), but these guidelines do not suggest standard analytic models established from a quantitative evaluation of methods. To address this omission, we use MRE monitoring data to conduct a robust model evaluation and then recommend analytic models to measure change and ensure accurate and efficient environmental monitoring.

2.2. Case study data

The case study data used to establish baseline conditions is described in Linder et al. (2017). Active acoustic data from Admiralty Inlet, WA was collected from a tidal turbine pilot project site considered by the Snohomish County Public Utility District No. 1. The site is located ~1 kilometer off Admiralty head shoreline, Puget Sound Washington (48.18° N, −122.73°W), at a depth of ~60 m (Public Utility District No. 1 of Snohomish County, 2012). This study uses acoustic backscatter (i.e., reflected energy) data from a seabed mounted BioSonics DTX echosounder operating at 120 kHz. The echosounder sampled the water column at 5 Hz for 12 min every 2 h between May 9 and June 9, 2011. Backscatter data were restricted to the lower 26 m from the bottom (twice the proposed turbine height). Acoustic backscatter is used as a proxy for nekton (i.e., macro-invertebrates and fish that move independently of fluid motion) density within the water column (Maclennan et al., 2002).

Active acoustic data can be used to derive metrics that characterize nekton densities and spatiotemporal distributions (e.g., Urmy et al., 2012). Two of these metrics: mean volume backscattering strength (Sv, dB re 1 m^{-1}), and an aggregation index (AI, m^{-1}) were used in this study to represent normal (i.e., Sv) and non-normal (i.e., AI) data distributions found in MRE monitoring data and contain properties common to all monitoring data (Chandler and Scott, 2011; Gitzen et al., 2012). Each time series data set consists of 336 data points occurring every 2 h from May 9 through June 9, 2011 (cf. Linder et al., 2017). Both metrics are continuous, display periodic autocorrelation (Jacques 2014), and are trend-stationary (i.e., statistical data properties are constant over time, assuming that the periodicity and trend are associated with deterministic environmental variables). Sv data index nekton density and are normally distributed while AI values measure nekton patchiness, are right-skewed, and composed primarily of low values with high amplitude spikes (cf. Linder et al., 2017). The terms *low state* and *high state* will be used to refer to the two amplitudes of aggregation index values.

2.3. Evaluation approach

Performance of candidate models were evaluated using simulated Before-After change scenarios generated from MRE baseline case study data, with an indicator variable added to measure change from baseline conditions. The case study data are assumed representative of empirical monitoring data, and evaluated models are assumed representative of models capable of measuring change. Candidate models were evaluated on their ability to detect, quantify, and forecast change. A model's ability to detect change was assessed using a power analysis conducted by comparing interpolation accuracy (i.e., prediction of data within the range of the empirical data) by the model with or without indicator

variable(s). This binary definition of power to detect change aligns with traditional hypothesis testing used to detect change at a predetermined significance level (Morrison, 2007). Model ability to quantify the amplitude and shape of change was assessed using fit and forecast accuracy metrics that when combined, provide an assessment of model behavior (Shmueli 2010). Fit and forecast accuracy metrics are defined using residual methods that calculate differences between observed and model estimated values. The evaluation is designed to provide an equal comparison across model classes, and ensures optimal performance from each model. Results were then used to recommend model(s) capable of accurately detecting, quantifying, and forecasting change. All analyses were conducted in the R v.3.1.2 statistical software environment (R Core Development Team, 2014).

2.4. Baseline simulation models

A univariate, autoregressive state-space model (SSM) and hidden Markov model (HMM) were used to simulate baseline conditions from the empirical case study data to generate Before-After change scenarios. A SSM was used to simulate nekton density data, representative of normally distributed MRE monitoring data. The SSM with fixed low measurement error (i.e., SSM-P) was previously recommended for characterizing MRE baseline nekton density data (cf. Linder et al., 2017). A HMM was used to simulate baseline nekton aggregation index data, representative of non-normally distributed MRE monitoring data. The model estimates the probability of the data being within any given number of states at a specified time, which is applicable to the multi-state AI data.

The SSM-P was structured using a 10% fixed measurement error of the total error. Process error (i.e., natural variability in the true state of the population) was estimated as a parameter. We assumed that measurement error in linear backscatter data collected from a stationary echosounder results from calibration and hydrographic conditions. These sources of error have been suggested to be a maximum of 5% each of the total error (Simmonds and MacLennan, 2005). Within the SSMs the B parameter estimates mean-reversion: $B = 1$ indicates a non-stationary random walk and $B < 1$ indicates mean-reversion in the data (Holmes et al., 2014). No SSM parameters, including B, were fixed prior to parameterizing the model. The state-space simulation model was fit to the empirical Sv data using the R statistical package MARSS v.3.9 (Holmes et al., 2014). Detailed information on model parameters and structure are described in Holmes et al. (2012, 2014).

The HMM assumed a log-link Gamma distribution for the AI data and was fit using the R statistical package msm v.1.5.2 (Jackson, 2011), which requires a prior estimate of Gamma distribution parameters for each data state, and a transition intensity (i.e., transition rate from state to state) matrix (Q) to produce model estimates. The "fitdistr" function from the MASS v.7.3-45 R statistical package (Venables and Ripley, 2002) was used to estimate initial Gamma parameters of visually identified low ($< 0.1\ m^{-1}$) and high state ($\geq 0.1\ m^{-1}$) AI data. To produce preliminary estimates of the Q matrix, random numbers were generated to populate the matrix such that the rows sum to zero and diagonal elements equal $q_{ii} = -\sum_{j \neq i} q_{ij}$, where i and j represent two different states (Jackson 2011). A HMM model that included these initial Gamma and Q matrix parameters, and all independent predictor variables of the AI data within each aggregation state, was used to re-calculate Gamma and Q matrix parameters (that were fixed in the model selection) to produce the most accurate parameterized simulation model. The expected state of data at time t, either low or high, was estimated by applying the Viterbi algorithm (Viterbi, 1967) to the parameterized model using the "viterbi.msm" function. Detailed information on model parameters and structure can be found in Cappé et al. (2005), Zucchini and MacDonald (2009), and Jackson (2011).

2.5. Simulation model selection

The same ancillary environmental covariates (i.e., daily tidal range (m), tidal speed (m/s), Julian day of year, and a 24 h Fourier series) used to structure candidate models in the baseline evaluation study (Linder et al., 2017) were used to optimize the structure of the SMM and HMM. All possible combinations of covariates and specified two-way interactions (i.e., Julian day-tidal speed, Julian day-tidal range, tidal speed-tidal range, and tidal speed- 24 h period) produced 24 versions of each candidate model (see Linder et al., 2017). Akaike Information Criterion (AIC) was used to select the most accurate version of each model. AIC is a commonly recommended model selection tool that can efficiently select the most parsimonious SSM and HMM (Burnham and Anderson, 2002; Johnson and Omland, 2004). All combinations of the 24 covariates were evaluated (576 model versions) within the two-part SSM and HMM equations because combinations of covariates may equally affect both the process and the observation of the response in the SSM, and the low and high state of aggregation index data in the HMM. The versions of the SSM and HMM with the lowest AIC and complete convergence were chosen to simulate the normal and non-normal baseline monitoring data (Table A1).

2.6. Change scenarios

Change scenarios were developed based on perceived effects of MRE development, but the scenarios are representative of change that may occur in monitoring data from any natural or anthropogenic disturbance. Before-After monitoring study design and Intervention Analysis (Box and Tiao 1975) literature indicated that change may occur in the mean and/or the variance of response variables as a

Table 1
Change scenarios modeled in the "After" portion of "Before-After" simulated nekton density and aggregation index data. Change scenarios were developed from hypothesized effects of MRE development stressors on the fish receptor.

Stressor	Predicted Effect on Fish	Shape of Change	Changed Statistical Property	Change Function
Noise	Decrease	Step shift	Mean	$-\mu\delta$
Device Operation Movement	Variable increase and decrease	Periodic shift	Variance	$+/-\mu X$
Device Presence (FAD)	Increase	Linear trend	Mean	$\left(\frac{2\mu\delta}{n-1}\right)t_{i-1}$
Device Presence (FAD)	Increase	Nonlinear trend	Mean	$\frac{2\mu\delta}{1+4450e^{-0.05t_{i-1}}}$
Chemical Spill	Decrease	Step-change + nonlinear return to baseline (i.e., Abrupt trend)	Mean	$\left(\frac{1}{1+4450e^{-0.05t_{i-1}}}-1\right)2\mu\delta$

Notes: μ is the mean of simulated "before" data, δ is the amplitude of change in the mean, t is time-point in the data series, X is an arbitrarily chosen proportion ofμ. The logistic function used to simulate nonlinear change was structured to have an inflection point at 50% of the total mean change at $t = 168$. To simulate the MRE device movement change, the μX function was added to the half of the data that corresponded to tidal speed < 1.07 m/s and subtracted from the half of the data that corresponded to a tidal speed > 1.07 m/s to produce a change in variance rather than mean.

discrete step, a linear, or a nonlinear change (Box and Tiao, 1975; Underwood, 1994; Scheiner and Gurevitch, 2001). Five possible change scenarios in nekton densities and aggregations were hypothesized to result from MRE development: a step shift in mean from device noise (e.g., Inger et al. 2009; Boehlert and Gill, 2010), a change in variance resulting from a shift caused by device operation (e.g., Polagye et al., 2011; Klure et al., 2012; McCann 2012), a linear or nonlinear gradual change in mean from the presence of a device acting as a Fish Aggregation Device (FAD) (e.g., Klure et al., 2012; Shields and Payne, 2014), and a step reduction in mean plus nonlinear change back to baseline conditions in response to a chemical spill (e.g., Polagye et al., 2011; Boehlert et al., 2013; Fodrie et al., 2014) (Table 1).

Two change attributes, amplitude and lag, were varied to evaluate the sensitivity of a model's ability to measure change. A range of amplitudes were used in all 5 change scenarios to test the detection of change including: 10% of the mean, 25% of the mean, and the 5th or 95th percentiles of the data distribution (Munkittrick et al., 2009). Current MRE monitoring guidelines do not specify the deviation from a baseline that constitutes significant change, because there are no vetted biological thresholds available. For changes in variance, change amplitudes of 10%, 25% and 2 times the standard deviation were used. A two week lagged period was used to examine the detectability of a lagged onset of change within the FAD scenarios (cf. Matsumoto et al., 1981; Barnabé and Barnabé-Quet, 2000; Kramer et al., 2015). To explore sensitivity to lagged change, scenarios were simulated at the 10% and 95th percentile values in the linear and nonlinear FAD change scenarios.

A function was added to simulated baseline data to create the desired shape and amplitude of change in the after (i.e., post-installation) data (Table 1). Parameterized covariates and error distributions of each baseline simulation model were kept constant while only the added function produced the change in baseline conditions (e.g., Benedetti-Cecchi, 2001; Mackenzie et al., 2013; Vanermen et al., 2015). Changes to simulated baseline nekton density data were applied to the linear form of acoustic backscatter and then the data were transformed to logarithmic units (i.e. dB re 1 m^{-1}). "After" data were generated using a different random seed for error distribution than that used for the baseline data in an effort to simulate natural variability between the two monitoring periods. Change in lag scenarios was simulated in the same manner, except that the change was only added to the second half of the after data. The Before-After simulated datasets consisted of 672 data point (336 data points before and after the change, matching the Admiralty Inlet MRE tidal turbine baseline dataset).

2.7. Candidate models

The three classes of models evaluated included: generalized regression models, time series regression models, and nonparametric regression models (Table 2). Characteristics and differences among the three model classes are detailed in Linder et al., 2017. Candidate models that were evaluated on their ability to characterize data in the previous study (Tables A2 and A3, Linder et al., 2017), and are now evaluated on their ability to detect, quantify, and forecast change. Since the primary goal of environmental monitoring is to measure change relative to baseline conditions (Shumchenia et al., 2012), it is important to use the same methods in baseline and operational monitoring (McCann, 2012). The SSM-P candidate model (Linder et al., 2017) is also used to generate normally distributed baseline data for the change scenarios. Any potential bias in the SSM evaluation is minimized by using AIC, because it is a different model selection metric than the 10-fold cross-validation metric used to parameterize candidate models. An HMM was not included as a candidate model in the current evaluation to reduce bias produced from using a HMM as a simulation model. The HMM and SSM are both members of the state-space model class, so the SSM evaluation results should indicate advantages or constraints of using a HMM model for measuring change in MRE monitoring data.

Table 2
Description of properties and structure of evaluated candidate regression models fit to changes scenarios in normal and non-normal time-series data.

Model	Class	Form	Parametric/ Nonparametric	Error Components	Error Distribution	Auto-correlation Structure
Linear	GR	Linear	Parametric	Observation error	Normal	None
Generalized least squares (GLS)	GR	Linear	Parametric	Observation error	Normal	Residual correlation
Generalized linear model (GLM)	GR	Linear	Parametric	Observation error	Gamma (identity)	None
Generalized linear mixed model (GLMM)	GR	Linear	Parametric	Observation error	Gamma (identity)	Residual correlation
Generalized additive model (GAM)	GR	Nonlinear	Semi-parametric	Observation error	Gamma (identity)	None
Generalized additive mixed model (GAMM)	GR	Nonlinear	Semi-parametric	Observation error	Gamma (identity)	Residual correlation
Univariate autoregressive state-space model (SSM)	Time series	Linear	Parametric	Process and Observation error	Normal	AR-1 lagged variable
Regression – autoregressive moving average model (Reg-ARMA)	Time series	Linear	Parametric	Observation error	Normal	ARMA error
Regression – autoregressive moving average – generalized autoregressive conditional heteroscedasticity model (Reg-ARMA-GARCH)	Time series	Linear	Parametric	Observation error	Skewed-student-t	ARMA error; GARCH residual variance
Random forest (RF)	NP	Nonlinear	Nonparametric	N/A	None	Lagged variables
Support vector regression (SVR)	NP	Nonlinear	Nonparametric	N/A	None	Lagged variables

Note: The evaluated model classes are generalized regression (GR), time series, or nonparametric (NP) models.

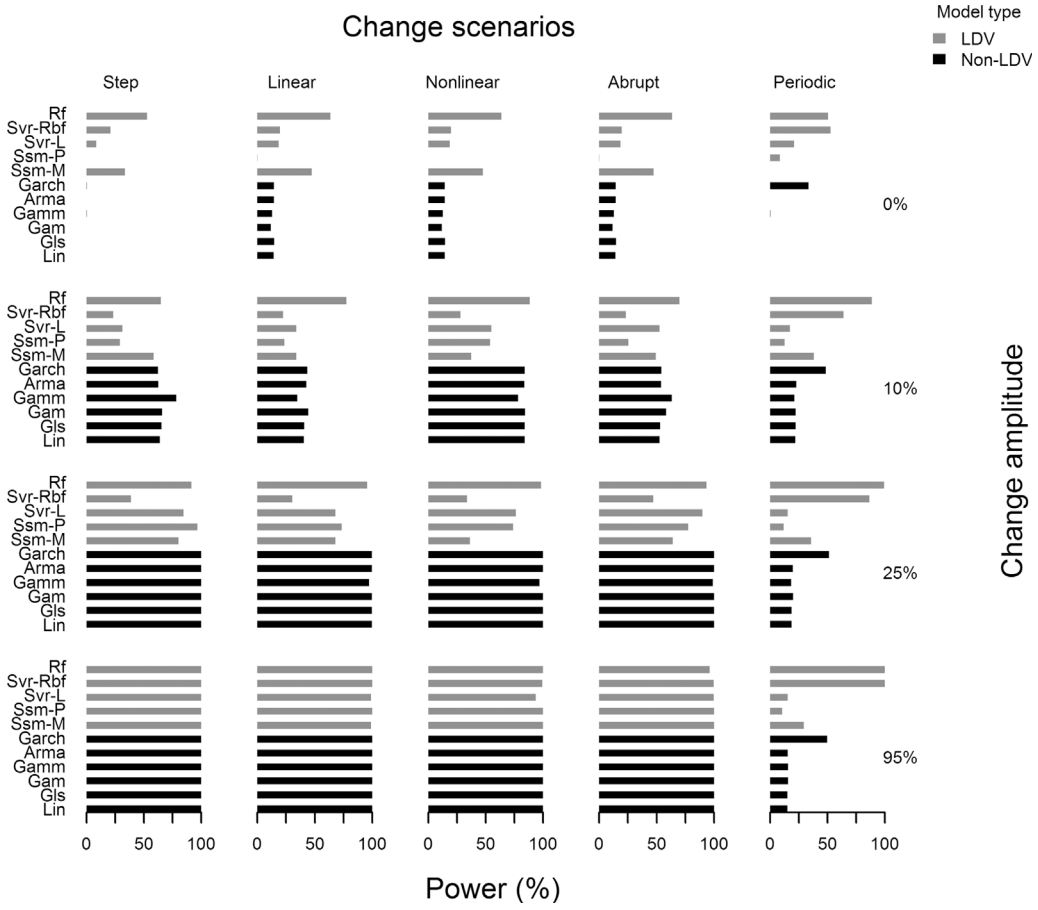

Fig. 1. Power of candidate models to detect change in simulated normally distributed Before-After MRE monitoring data. Power is calculated across varying forms: Step, Linear, Nonlinear, Abrupt, and Periodic; and amplitudes 0%, 10%, 25%, 95%. Models that included lagged dependent variables (i.e., LDV models) are shown in grey, and the models without lagged dependent variables (i.e., non-LDV models) are shown in black.

2.8. Intervention analysis

In this study, Intervention Analysis (IA) is adapted to measure post-disturbance change in the serially correlated, case study data. In traditional IA an ARIMA model is fit to baseline data, and then that ARIMA model is fit to the entire dataset (pre- and post- intervention) with an added indicator, or intervention function, to model change in baseline conditions (Biglan et al., 2000; Lagarde, 2011). This approach was used to measure change in simulated Before-After datasets by adding intervention functions to all baseline parameterized candidate models.

Step and ramp intervention functions (Lewis-Beck et al., 2004; Cryer and Chan, 2008) were added to candidate models. The step function is used to add a step change in the mean or variance. A ramp function models linear change in slope. The step (I_2) and ramp (I_3) intervention functions were added to each baseline parameterized candidate model as:

$$y = a + BX + b_2 I_2 + b_3 I_3 \quad (1)$$

where BX is a matrix of all parameterized covariates used in the baseline model, a is an intercept, I_2 is an indicator variable equal to 0 before the MRE development and 1 after. I_3 is an indicator variable equal to 0 before the MRE development, and an indexed data count after the intervention: $t - n_{before}$; where n is equal to the number of data points before the MRE development. Only I_2 was included in Step and Periodic change scenarios. I_2 plus I_3 was added to all models to measure change in trend (i.e., Linear, Nonlinear, and Abrupt change scenarios).

Indicator variables were added directly to all candidate models using the parametrized model specifications and structure defined in Linder et al. (2017). The GA(M)Ms and SMMs require additional specification of intervention variables. Within the GA(M)Ms the ramp function (I_3) was structured as an interaction between time (i.e., an indexed data count) as a thin-plate smoother spline and the parametric I_2 variable (Wood, 2006) to permit nonlinear change in trend. Within the SMMs the step intervention was only included in the observation equation to align with the parameterization of the intervention function in all other candidate models that do not include an explicit process equation. The ramp intervention was included in the SSMs by altering the trend parameter (u) to be estimated as two separate values before and after the intervention to allow for a change in trend.

2.9. Power analysis

A power analysis was used to evaluate model ability to detect change. Power can be interpreted as the ability to accurately detect a difference between the null and alternative hypotheses, which is equivalent to a change in the data. Mackenzie et al. (2013) suggested the use of a power analysis to quantify the ability of statistical models to detect change in MRE monitoring data.

A 10-fold cross-validation (10-fold CV) model selection (cf. Hastie et al., 2009) was used to conduct the power analysis, with Root-Mean-Squared-Error (RMSE) used as the metric of model accuracy. A RMSE value closer to 0 indicates a more accurate interpolation of the data (Bennett et al., 2013). The 10-fold CV model selection was applied to each candidate model with and without the indicator variable(s). If the IA version of a model produced a lower average RMSE than the baseline version, then it was considered a more accurate interpolation of the data, and the more appropriate model choice (Anderson and Burnham, 2002; Gitzen et al., 2012). The selection of the IA model form was defined as the model "detecting" change. The 10-fold CV model selection was applied to 1000 simulated Before-After datasets for every change scenario, and the proportion of the 1000 simulated datasets that each candidate model detected change was used to quantify the power of each model. An additional "no change" or 0% change scenario was

included in the power analysis to quantify the false detection of change (i.e., Type I error) by candidate models. To ensure datasets were not identical, different random error seeds were used to create each dataset's error distribution.

2.10. Model fit and forecast accuracy metrics

Root-Mean-Squared Error (RMSE) and Mean-Absolute-Scaled Error (MASE) (Chai and Draxler, 2014) were used to evaluate model ability to quantify pattern and to forecast change. The RMSE fit metric was calculated using the first 90% of the data. The remaining 10% of the data (72 data points) was used to estimate forecasting ability using the RMSE statistic and the MASE statistic:

$$\text{MASE} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{\frac{1}{n-1}\sum_{i=2}^{n} |y_i - y_{i-1}|} \quad (2)$$

where y_i is the observed i^{th} value, $\hat{y}_i$ is the model predicted value, and n is the total sample size.

A MASE value < 1 indicates a better average performance than the naïve in-sample forecast, which uses the observed value from point $t-1$ as the forecast of point t (Hyndman and Koehler, 2006). The MASE metric tends to be less sensitive to outliers compared to RMSE (Ward et al., 2014). Using both metrics provides a more representative description of model performance (Chai and Draxler, 2014). RMSE and MASE evaluation metrics were averaged from 30 simulated datasets of every change scenario, each using a different random error seed, to ensure robust evaluation metrics.

To identify models that accurately quantify and forecast change, models were ranked from most to least accurate using average fit and forecast metrics for each scenario. It may be misleading to suggest one model outperformed another solely based on rank position if their average performances are similar. Traditional statistical testing used to quantify performance similarity is not appropriate because p-values are affected by the number of simulations run (White et al., 2014). After an initial ranking, the Empirical Cumulative Distribution Function (ECDF) and associated Kolmogorov-Smirnov D-statistic of the 30 metric values produced for each scenario were used to group models with similar accuracy (Bennett et al., 2013).

D-statistic values, calculated using the "ks.test" function in the R package stats v.3.1.2 (R Core Development Team, 2014), equal the maximum distance between each model's individual Empirical Cumulative Distribution Function (ECDF) and the cumulative ECDF of all other models. For this study, a more positive D-statistic indicates a more accurate fit or forecast of a model relative to all other candidate models. To illustrate by example, if the D-statistic of a model is +0.5, then that model has a 50% greater cumulative probability of producing the RMSE or MASE value at the point of observed maximum distance than all other candidate models. D-statistic values were first inspected to identify similar performing models, and then followed by a visual inspection

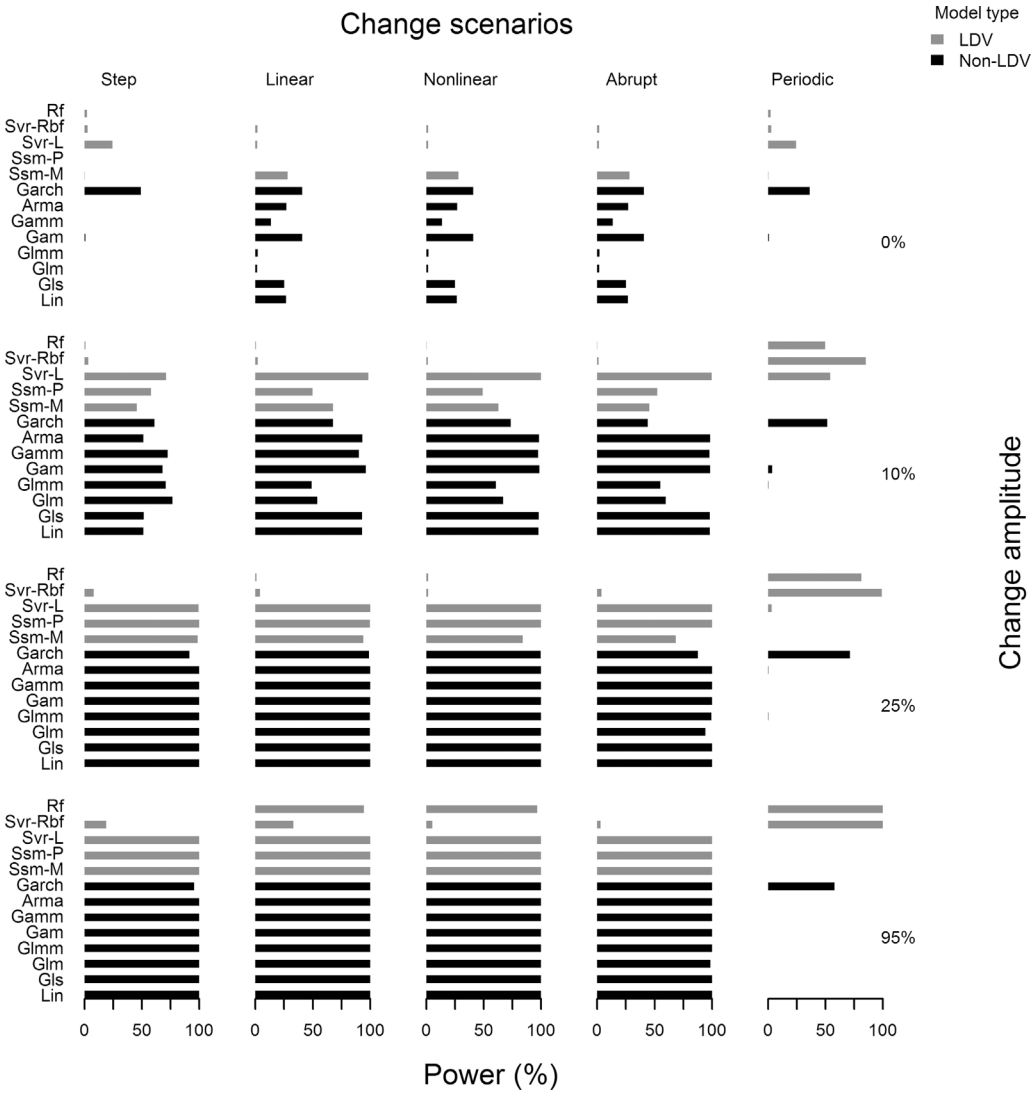

Fig. 2. Power of candidate models to detect change in simulated non-normally distributed Before-After MRE monitoring data. Power is calculated across varying forms: Step, Linear, Nonlinear, Abrupt, and Periodic; and amplitudes 0%, 10%, 25%, 95%. Models that included lagged dependent variables (i.e., LDV models) are shown in grey, and the models without lagged dependent variables (i.e., non-LDV models) are shown in black.

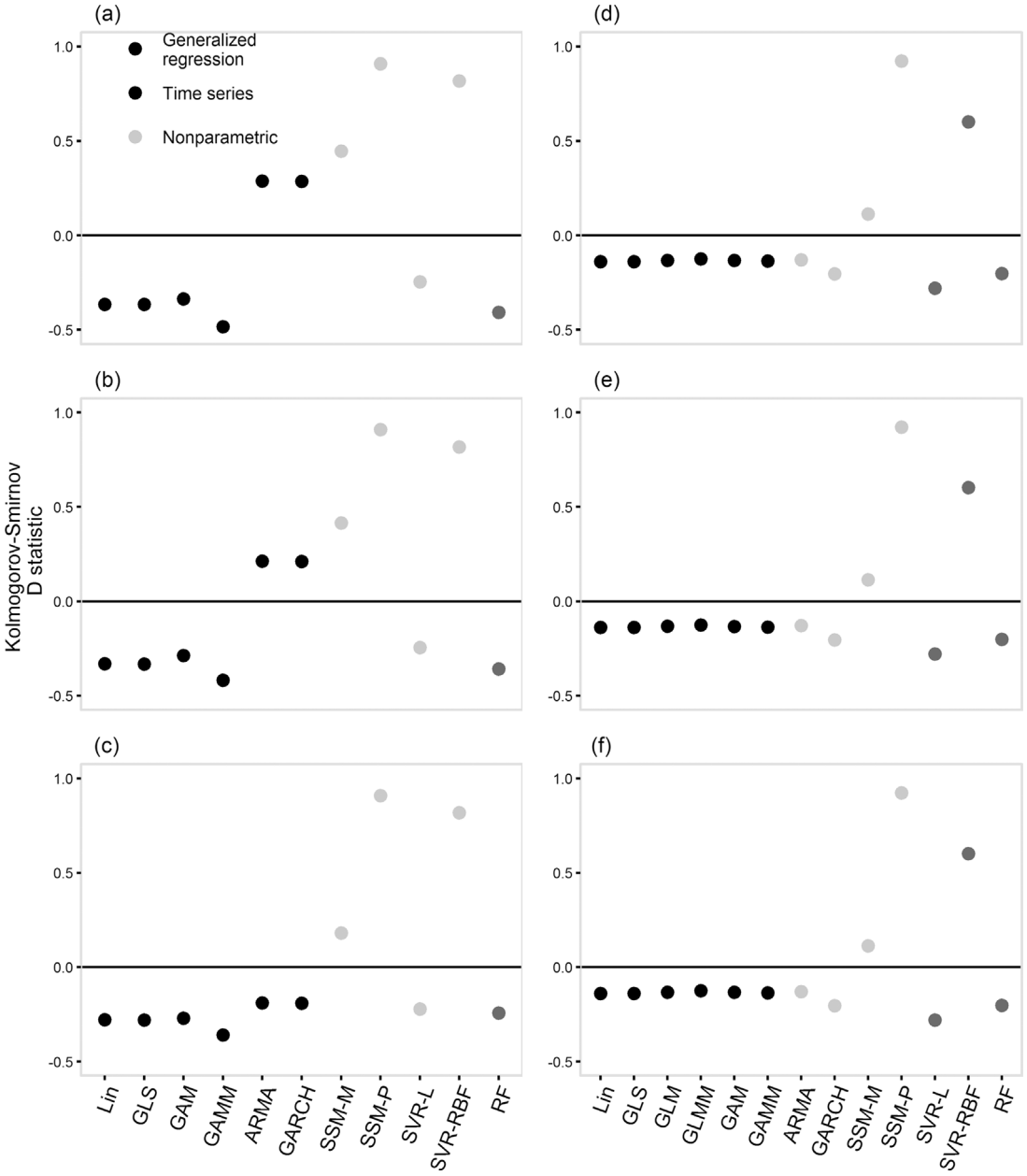

Fig. 3. Kolmogorov-Smirnov D statistic values of model fit for all change scenarios in normally and non-normally distributed data. The D statistic was calculated from the Empirical Cumulative Distribution Function (ECDF) of the root-mean-squared-error (RMSE) values of model fit across all amplitudes of: step mean-change (a), trended mean-change (b), and variance-change (c) in the normal data, and step mean-change (d), trended mean-change (e), and variance-change (f) in the non-normal data.

of ECDF plots to ensure accurate grouping of similarly performing models. If two model's ECDF plots did not overlap for > 50% of the cumulative distribution, then the models were ranked separately. After similarly performing models were grouped, all candidate models were re-ranked using a dense ranking method (cf. Cichosz, 2015).

To standardize labels and to avoid ambiguity, terms are defined and used to group results. The Linear, Nonlinear, and Abrupt change scenarios are referred to as "trended" change scenarios. The 95% amplitude of change is used to refer to both the 95% mean-change in all Step and trended scenarios, as well as the 2 standard deviations change (i.e., largest amplitude of variance-change) in all Periodic scenarios. The lagged change scenarios are not discussed for the power and fit accuracy results, because they did not advance the model evaluation findings.

3. Results

A total of 40,000 datasets (20,000 datasets each for normal and non-normal simulation models) were simulated to calculate power, fit, and forecast metrics. To estimate power, 4,800,000 model runs (2,200,000 for the normal data and 2,600,000 for the non-normal data) were completed. To calculate each fit and forecast metric values 6270 normal and 7410 non-normal model runs were completed. Appendices include all model results ranked by average fit and forecast metric, and associated D-statistic values, for all change scenarios in the normal (Appendix B) and non-normal (Appendix C) data.

3.1. Power

Overall, models with lagged dependent variables (i.e., SSM-M, SSM-P, RF, SVR-L, and SVR-RBF) exhibited lower power than non-LDV

models when detecting 10–25% mean-change compared to non-LDV models (Figs. 1 and 2). All non-LDV models except RF have lower power in the normal 10–25% mean-change scenarios (Fig. 1). Although RF has high power to detect change in these scenarios, it also has a high detection of false change in the 0% change scenarios relative to all other candidate models (Fig. 1). In nearly all non-normal mean change scenarios RF and SVR-RBF have lower power to detect change (Fig. 2), and have the highest number of lagged dependent variables fit to AI data (Table A3). This result is consistent with the observation that inclusion of lagged predictor variables has been shown to reduce explanatory power of other variables (Achen, 2000). In both the normal and non-normal data, SSM-P had equal or greater power than SSM-M for the 25–95% mean-change scenarios (Figs. 1 and 2).

Non-lagged dependent variable (non-LDV) models (i.e., Linear regression, GLS, ARIMA, GARCH, and GA(M)Ms) generally had the greatest power to detect change in mean-change scenarios, but their ability to detect change was dependent on baseline predictors and model assumptions. For example, in the normal data change scenarios GAMM was the only non-LDV model without the day-tidal range covariate (Table A2), and it generally displayed the least similar power results relative to all other non-LDV models at the 10% and 25% amplitude of all mean-change scenarios (Fig. 1). In the non-normal data, GL(M)Ms, which included day as a covariate, had lower power in the 0% and 10% trended change scenarios, and greater power in the 10% Step change scenario relative to the other non-LDV models (Fig. 2).

RF and SVR-RBF models generally had the greatest power to detect a change in variance for all Periodic scenarios in both the normal and non-normal data, and were the only models to consistently increase power as the amplitude of change increased (Figs. 1 and 2). RF had the greatest power for all Periodic scenarios, followed by SVR-RBF, and reg-ARMA-GARCH in the normal data. All other models exhibited less than 25% power for all Periodic scenarios. Patterns in model power to detect change in variance within non-normal data are very similar to those observed in the normal data. The main difference is that the SVR-RBF model displayed higher power for the Periodic 10–25% scenarios than the RF model.

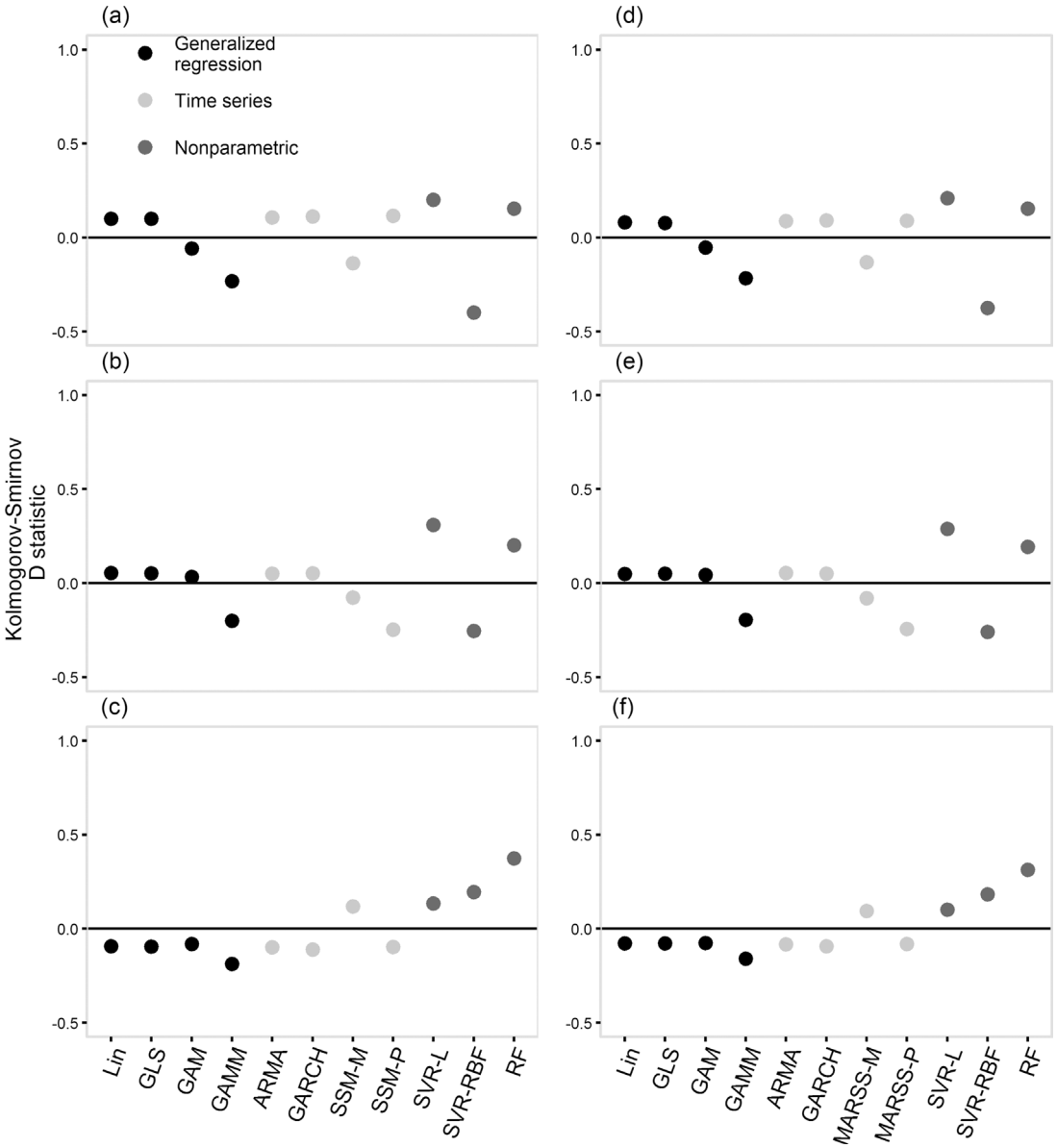

Fig. 4. Kolmogorov-Smirnov D statistics of model forecast for all change scenarios in normally distributed data. The D statistic was calculated from the Empirical Cumulative Distribution Function (ECDF) of the mean-absolute-scaled-error (MASE) values of model forecast across all amplitudes of: step mean-change (a), trended mean-change (b), and variance-change (c), and the root-mean-squared-error (RMSE) values of model forecast across all amplitudes of: step mean-change (d), trended mean-change (e), and variance-change (f).

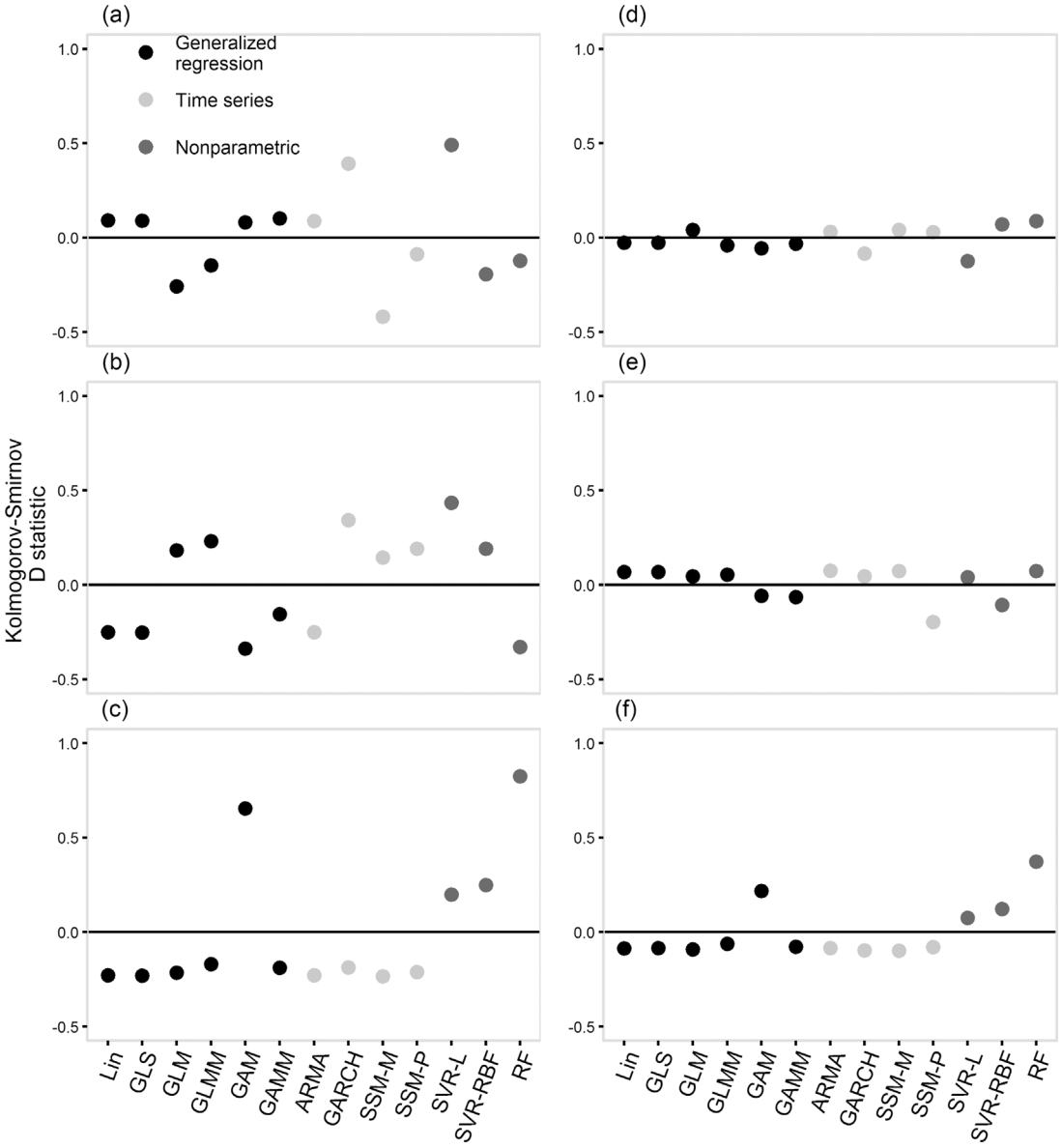

Fig. 5. Kolmogorov-Smirnov D statistic values of model forecast for all change scenarios in non-normally distributed data. The D statistic was calculated from the Empirical Cumulative Distribution Function (ECDF) of the mean-absolute-scaled-error (MASE) values of model forecast across all amplitudes of: step mean-change (a), trended mean-change (b), and variance-change (c), and the root-mean-squared-error (RMSE) values of model forecast across all amplitudes of: step mean-change (d), trended mean-change (e), and variance-change (f).

3.2. Fit

Accuracy of model fit is consistent across all change scenarios in normal and non-normal data. SSM-P, SVR-RBF, and SSM-M had the top 3 most accurate fits to change scenarios based on RMSE (Fig. 3). In the normal data, the other time series models (i.e., reg-ARMA and reg-ARMA-GARCH) were grouped as the 4th most accurate models for all mean-change scenarios (Table B1). RF outperformed SSM-M to rank 3rd in fit accuracy in the 95% Periodic change scenario in both the normal and non-normal data (Tables B1 and C1).

Greater differences existed among models that performed poorly in fitting change scenarios. Among normal data fits, GAMM was least able to accurately fit all change scenarios, and the RF model was generally considered the 2nd to worst ranked fit model for the Step and trended scenarios (Fig. 3). Among the non-normal data model fits, SVR-L, RF, and reg-ARMA-GARCH were generally the worst fit models for the mean-change scenarios (Fig. 3).

Exceptions to observed performance patterns illustrate the relative improvement in ranked performance of GA(M)Ms in nonlinear trend scenarios. In Abrupt and Nonlinear 95% change scenarios in the normal data, the GAM was ranked highest relative to its performance in all other scenarios (4th and 5th best fit) (Table B1). In the Abrupt 95% scenario GAM, followed by GAMM, outperformed all other candidate models except for the top 3 performers (i.e., SSM-P, SVR-RBF, SSM-M) (Table B1). The majority of candidate models performed similarly in change scenarios in the non-normal data except in the Nonlinear 95% scenario, in which GAM and GAMM were identified distinctly as the 4th and 5th best performers (Table C1).

3.3. Forecast

MASE and RMSE forecast results were similar for the normal data and are grouped unless results differed. Interpretation of the non-normal data forecast results differed depending on use of the RMSE or the MASE metric and will be discussed using the MASE metric, unless otherwise specified. As the amplitude of change increased, the range among model forecasting performances became wider. The relative performance of candidate models was generally most consistent in the

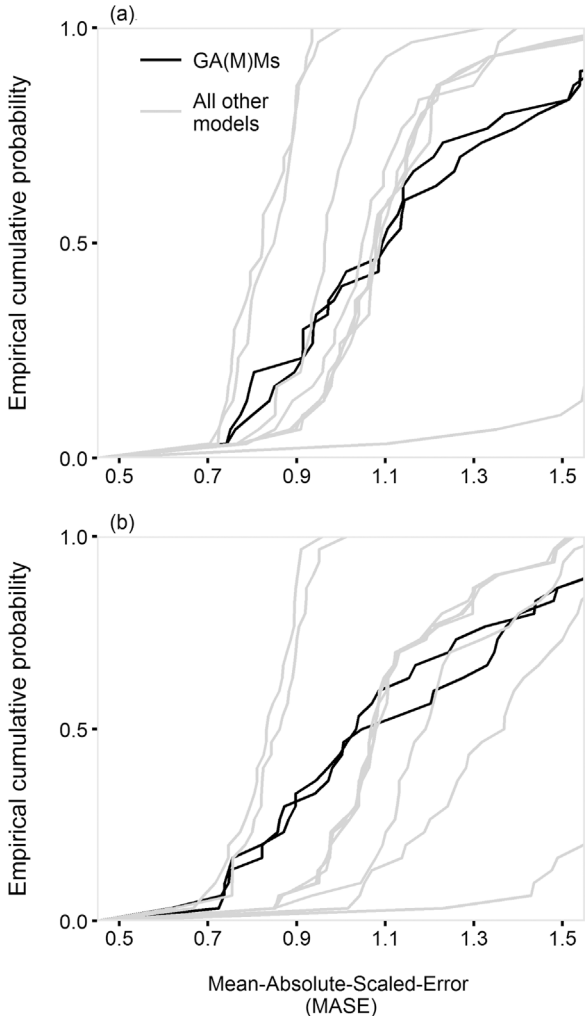

Fig. 6. Empirical cumulative probability of the mean-absolute-scaled-error (MASE) values of model forecasts, with the GAM(M) forecasts highlighted for the Abrupt (a) and Nonlinear (b) 95% change scenarios in normally distributed data. MASE plots are representative of the pattern also displayed in the corresponding root-mean-squared-error (RMSE) cumulative probability plots (not shown).

Step scenarios.

The SVR-L model was generally the most accurate forecast model for all mean-change scenarios in both the normal and non-normal data (Figs. 4 and 5). The second most accurate forecast model varied between the normal and non-normal data. In the normal data, RF was the best forecast model in the Abrupt 95% mean-change scenario, and the second best forecast model for almost all other mean-change scenarios. In the non-normal data, reg-ARMA-GARCH was generally the second top performing forecast model. For both data types, RF was the best forecast model for all variance-change scenarios. In the non-normal data, GL(M)Ms generally performed similarly and had better forecast accuracy than other parametric/semi-parametric regression models for all trended mean-change scenarios (Fig. 5).

In both the normal and non-normal data, GA(M)Ms displayed relatively weak forecasting ability. In general, SVR-RBF and GAMM were the worst forecasting models for the normal data (Fig. 4). In the non-normal data GAM was a poor forecast model for the trended mean-change scenarios (Fig. 5), and both GA(M)Ms were ranked among the 3 worst forecast models for Abrupt and Nonlinear 25–95% scenarios (Table C3). For both data types, SSM-P was a poor forecast model for the 95% trended change scenarios. In the normal data model evaluations, SSM-M was also a poor forecast model for the 95% trended change scenarios, with SSM-P having less accurate forecasts than SSM-M (Tables B3 and B4).

The trended 95% mean-change scenarios were exceptions to the general forecast patterns for both model evaluations of normal and non-normal data. The only normal data scenario in which SVR-L and RF were not the top two performers was the Linear 95% scenario, where deterministic parametric models were all grouped as the top performing forecast models (Tables B3 and B4). In the normal data, the GA(M)Ms demonstrated a wide ECDF of both RMSE and MASE values in the Abrupt and Nonlinear 95% change scenarios, indicating the presence of outliers in model predictions (Fig. 6). Top performing forecast models in the Nonlinear and Abrupt 95% change scenarios in the non-normal data were inconsistent with the general pattern of top performing models in all other change scenarios (Table C3).

The main contrast between the lagged and non-lagged change scenarios in both the normal and non-normal data was between the top 95% change forecasting models. Unlike the 95% non-lagged scenarios, GA(M)Ms were the top forecast models for the 95% lagged scenarios for both data types (Figs. 7 and 8). In the normal data the SVR-L and RF models were also the best forecast models in the 95% lagged scenarios, whereas in the non-normal data, the RF and reg-ARMA were the top-

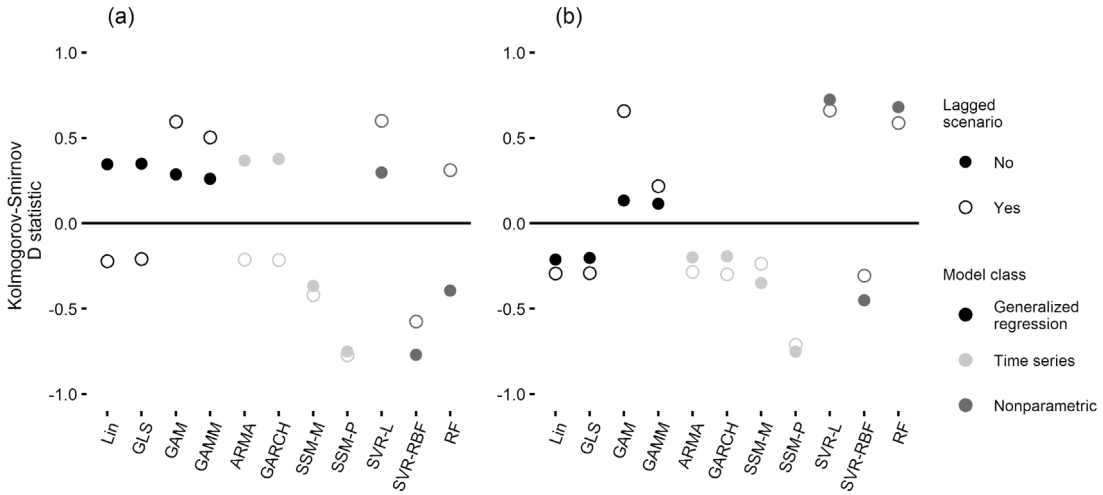

Fig. 7. Kolmogorov-Smirnov D statistic values of model forecast for the Linear (a) and Nonlinear (b) 95% change scenarios in normally distributed data. The D statistic was calculated from the Empirical Cumulative Distribution Function (ECDF) of the mean-absolute-scaled-error (MASE) values of model forecast in both the non-lagged scenarios (solid circles) and corresponding lagged scenarios (open circles). MASE plots are representative of the similar pattern displayed in the corresponding root-mean-squared-error (RMSE) cumulative probability plots (not shown).

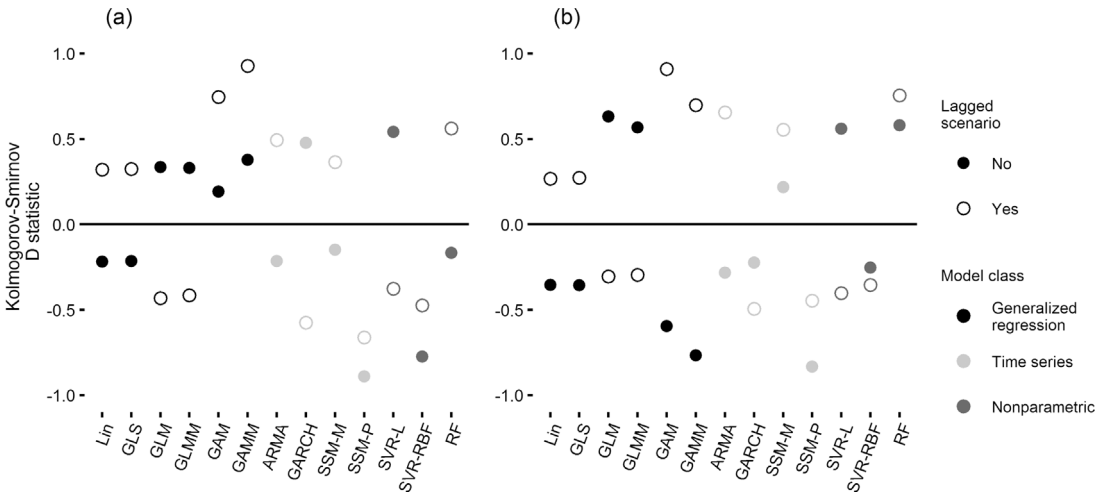

Fig. 8. Kolmogorov-Smirnov D statistic values of model forecast for the Linear (a) and Nonlinear (b) 95% change scenarios in non-normally distributed data. The D statistic was calculated from the Empirical Cumulative Distribution Function (ECDF) of the mean-absolute-scaled-error (MASE) values of model forecast in both the non-lagged scenarios (solid circles) and corresponding lagged scenarios (open circles). MASE plots are representative of the similar pattern displayed in the corresponding root-mean-squared-error (RMSE) cumulative probability plots (not shown).

ranked forecast models (Fig. 8).

Model forecast performance was similar for all models in all non-normal data change scenarios based on RMSE. An inspection of the D-statistic values and corresponding ECDFs of forecasting performance showed that differences among models were only distinct in the Periodic and 95% trended mean-change scenarios. The distinct model performances in these scenarios were generally consistent with model performances previously highlighted using MASE metric results. A relevant exception is that SVR-L and reg-ARMA-GARCH were the two worst forecasting models for all Step scenarios based on RMSE, although they were generally the top forecasting models based on MASE (Tables C3 and C4).

4. Discussion

4.1. Modeling to meet monitoring objectives

A plethora of ecological studies have highlighted strengths and weaknesses of modeling approaches for data inference and prediction (e.g., Thomas, 1996; Ward et al., 2014; Bell and Schlaepfer, 2016). The current evaluation is the first to quantify strengths and weaknesses of three model classes that include parametric, semi-parametric, and nonparametric statistical models used to measure change in MRE monitoring data, and to recommend model classes able to detect, quantify, and forecast change (see below). The evaluation also illustrates that the ability of a model to measure change depends on the monitoring objective. Deterministic, parametric models (e.g., linear regression, including ANOVA) are capable of detecting change, while the more flexible time-series models and semi-parametric models are advantageous for quantifying change. To forecast change, the most accurate results were produced by nonparametric models. By quantifying model behavior across a representative range of change scenarios in both normal and non-normal data, and tailoring recommendations to each monitoring objective, the evaluation ensures that requirements of each monitoring objective are met.

4.1.1. Power to detect change

Deterministic parametric/semi-parametric models are recommended for detecting change in monitoring data. This recommendation is consistent with previous use of these models to detect change in MRE monitoring studies (e.g., Hammar et al., 2013; Tollit and Redden, 2013; Stenberg et al., 2015). In support of previous use, non-LDV models (i.e., parametric and semi-parametric deterministic models) generally had the greatest power to detect change in mean-change scenarios, but their ability to detect change was dependent on baseline predictors and model assumptions. In the current study, appropriate model covariates are identified in an initial exploration of data characteristics and subsequent model selection (cf. Linder et al., 2017). Non-LDV models without day as a covariate had a higher rate of false change detection (i.e., type I error) compared to GL(M)Ms in non-normal data model evaluations. Conversely, this finding suggests that a model that identifies a baseline trend when it is not present may fail to detect environmental change (i.e., Type II error). Failure to detect change in monitoring data may result in environmental impacts at MRE or other monitoring sites (Fairweather, 1991; Legg and Nagy, 2006). A Type I error that falsely estimates environmental change can result in increased costs associated with implementing additional monitoring and/or mitigation measures to comply with monitoring regulations (Field et al., 2005; Levine et al., 2014). A Type II error is often considered more significant due to costs of long-term environmental damage added to the intervention costs that are associated with Type I error (e.g., Peterman 1990; Field et al., 2004). The evaluation was critical to identify model classes with the greatest power to assess risk of Type I or Type II error in model estimates of change.

Localized nonparametric models are recommended for detecting, quantifying, and forecasting change in variance. The RF and SVR-RBF model are the only two models capable of measuring change in variance as these models have a localized structure, which does not assume a constant variance (Taddy et al., 2011). The evaluation results highlight the inability of deterministic parametric models to measure change in variance. It is imperative to measure a change in variance rather than just the mean, because altered variability of populations may indicate perturbations in a population (Chapman et al., 1995). The Reg-ARMA-GARCH model may have also detected variance change through its estimate of heteroskedastic residual error, but this is not reflected as a change in the model's fit or forecasted predicted values. The Reg-ARMA-GARCH predicted values are based on a mean model equation that does not include residual variance estimates. Model predictions were the only criterion used to evaluate the ability to measure change in the effort to maintain an equal comparison across all parametric and nonparametric models.

4.1.2. Quantifying change

Deterministic time-series models (i.e., reg-ARMA and reg-ARMA-GARCH) are recommended for initial quantification of change in monitoring data. These models have not previously been used in MRE

monitoring studies, although ARMA and ARMA-GARCH models have been used to forecast wind speeds (e.g., Taylor et al., 2009; Liu et al., 2011). The current study illustrates that the reg-ARMA and reg-ARMA-GARCH models generally quantify change similar to other parametric, deterministic models, and provide additional information about change in monitoring data by quantifying autocorrelated error. A Reg-ARMA and/or Reg-ARMA-GARCH model may account for non-stationarity by estimating high autocorrelation in the ARMA error structure. A high estimate of autocorrelated error results in a better fit to nonlinear data (Granger and Newbold, 1974; Hyndman and Athanasopoulos, 2014). To illustrate by example, reg-ARMA excels in fit and forecast in the lagged 95% change scenarios in non-normal data relative to other parametric models.

If the estimate of autocorrelated error by deterministic time series models indicates nonlinear change, then the use of GA(M)Ms are recommended to quantify change. This recommendation is consistent with previous MRE monitoring studies that use GA(M)Ms to measure nonlinear trends (e.g., Petersen et al., 2011; Tollit and Redden, 2013), and the limited evaluation of smoother models in a previous MRE model evaluation study (Mackenzie et al., 2013). Semi-parametric models (i.e., GA(M)Ms) excel in fit and forecast in lagged scenarios of change in normal and non-normal data, indicating their advantage for quantifying nonlinear change. However, the inability of GA(M)Ms to outperform other models in general fit and forecast accuracy highlights the instability of GA(M)M estimates in autocorrelated time-series data (cf. Wood, 2006, 2015). As a result, these models are not recommended for an initial estimate of change, but are considered advantageous for quantifying nonlinear change.

Recommendation of both deterministic time-series models and GA(M)Ms to quantify change will provide a robust assessment of both linear and nonlinear change, and consequently reduces uncertainty in model choice and subsequent effects of change in the data. Accurate assessment of change is imperative for a monitoring program because it determines the validity of hypothesized effects of change from baseline conditions (Treweek, 2009), and may be used to inform management and mitigation decisions (Treweek 2009; Lindenmayer et al., 2012). Quantifying confidence around the detection of change provides managers an objective measure to assess risk of alternative environmental management decisions (Vos et al., 2000). The use of a parametric/semi-parametric models best aligns with these monitoring goals.

Although SSMs have previously been recommended for baseline data characterization (i.e., Linder et al., 2017), the inability to measure trended change prevents them from being recommended as a primary method for quantifying change in monitoring data. SSMs, especially SSM-P, excel in fitting change scenarios in normal and non-normal data, but were unable to accurately forecast trended change in monitoring data relative to all other models. Because the B parameter is not fixed at 1, the u trend parameter estimates mean level rather than trend (cf. Holmes et al., 2014). Consequently, both baseline parameterized SSM-P and SSM-M attribute trended change to stochastic variability. However, SSMs can be parameterized to estimate hypothesized trends across time (Ward et al., 2010; Holmes et al., 2014). Therefore, an initial estimate of change using deterministic time-series and semi-parametric models may be used to specify an appropriate parameterization of a state-space model that is used to quantify change attributed to natural variability (i.e., process error) and systemic change in baseline conditions (Dornelas et al., 2013). The partitioning of total error into process and observation components may reduce bias and improve accuracy in change estimates (e.g., Lindley 2003; Ward et al., 2010).

4.1.3. Forecasting change

Nonparametric models are recommended for forecasting change when accuracy in prediction rather than an estimate of casual relationships is a primary objective. Nonparametric models, specifically SVR-L and RF, generally excelled in forecasting change scenarios relative to all other evaluated models. These results are not surprising, as these models are regularly cited for their ability to predict both classification (e.g., Cutler et al., 2007; Grilli and Shumchenia, 2015) and time-series data (Thissen et al., 2003; Kane et al., 2014). Consistent with this study, SVR-L has been shown to be a more accurate forecast model for trended time-series data than SVR-RBF (e.g., Crone et al., 2006). Forecasting change in monitoring data does not require parametric measures of casual relationships (Shmueli, 2010), and the flexibility of nonparametric models can provide robust predictions of noisy and dynamic ecosystem data compared to parametric models (Perretti et al., 2013). Selection of accurate forecast models is imperative when informing management on timing or conditions of change, and when pre-emptive modifications to operations will reduce or eliminate mitigation measures to compensate for negative environmental effects (Clark et al., 2001; Lindenmayer et al., 2012).

The relatively weaker forecasting performance of the nonparametric models in the lagged change scenarios highlights the sensitivity of SVR model performance to the model's kernel parameters (Lorena et al., 2011), and the inability of a Random Forest model to extrapolate outside the training data range (Kacprzyk and Pedrycz, 2015). Re-tuning the baseline nonparametric models using after (e.g., MRE post-construction) monitoring data may further improve forecast estimates. Based on the ability of semi-parametric regression models to provide relatively accurate forecasts in highly nonlinear scenarios, semi-parametric models may be used to quantify and forecast estimates and associated confidence intervals of nonlinear change to help optimize forecasts of nonparametric models. For instance, if a GAMM model estimates the change as highly nonlinear, then the SVR baseline model may require a re-tuning of its parameters to provide an accurate forecast.

4.2. MRE monitoring model recommendations

Within recommended model classes, specific models are suggested for detecting, quantifying, and forecasting change in the MRE case study data (Fig. 9). Model recommendations for the non-normal aggregation data are based on the constraints and results of this study, but may produce biased estimates since the models do not fit the high-state spikes of the aggregation data (Barry and Elith, 2006). The general similarity of all RMSE model results for the non-normal data illustrates the inability of any model to excel at predicting aggregation spikes. Models that had a lower MASE value tended to forecast closer to the low-state non-normal data, because the MASE metric is less sensitive to outliers (i.e., high-state non-normal data) than the RMSE metric (Hyndman and Koehler, 2006). To evaluate model ability to measure change in the mean and variance of the AI data, the rate and occurrence of aggregation spikes was not altered from baseline conditions. Therefore, forecast model recommendations for the non-normal data are based on MASE metric results, which highlight model ability to fit to the dominant state of aggregation and form of change. Nekton aggregation data may be transformed using a Box-Cox normal data transformation (Box and Cox, 1964) to produce unbiased estimates when used with the recommended models. Although no model was able to accurately measure aggregation spikes, the evaluation provided insight into model behavior when fit to highly skewed data.

4.2.1. Detecting mean change

A Reg-ARMA model is recommended for detecting mean change in nekton density data (Fig. 9) because it assumes a normal distribution, includes environmental covariates that were used as predictor variables, and accounts for autocorrelation. This recommendation is consistent with the recommendation for quantifying change in normal MRE monitoring data.

A GLMM is recommended for detecting change in nekton aggregation data (Fig. 9). Non-LDV models were previously recommended to detect change in MRE monitoring data, and within this group GLMM does not assume normality, appropriately accounts for the day baseline trend in the simulated data, and estimates autocorrelation.

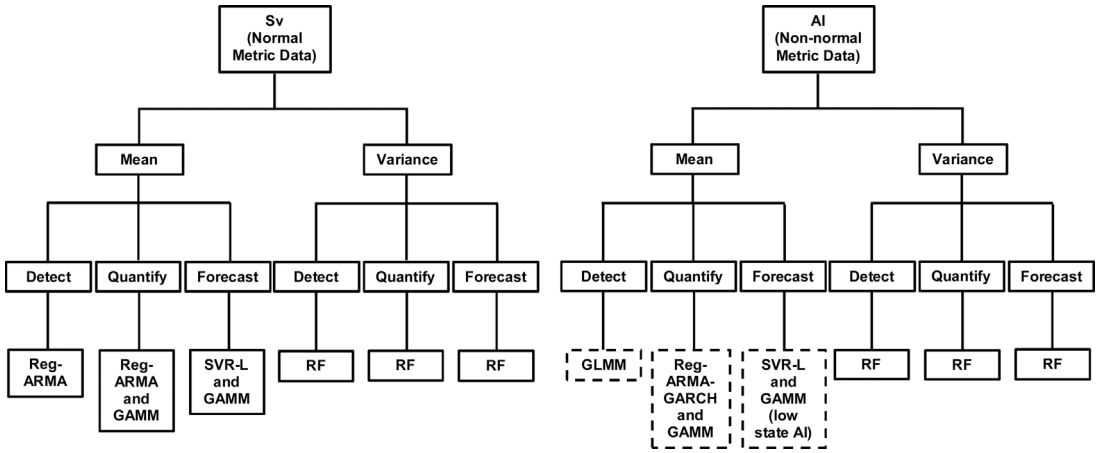

Fig. 9. Schematic of recommended models to detect, quantify, and forecast change in MRE monitoring data. Models are recommended for detecting, quantifying, or forecasting change in the statistical property of mean or variance of either Sv data (representative of normally distributed data) or AI monitoring data (representative of non-normally distributed data). The dashed boxes indicate potential bias associated with recommended models based on their inability to fit the spikes in aggregation index data.

4.2.2. Quantifying mean change

A reg-ARMA model is recommended for an initial inspection of mean change in nekton density data, followed by a GAMM to provide nonlinear change estimates (Fig. 9). A reg-ARMA model is less prone to convergence issues compared to a GAMM, and produces more accurate trend estimations for all scenarios, with the exception of a nonlinear trend. If the reg-ARMA model indicates potential non-stationarity through the estimation of highly autocorrelated error, then there may be a nonlinear change. The GAMM can then be used to estimate the shape and amplitude of nonlinear trend.

The reg-ARMA-GARCH model is recommended for an initial estimate of mean change in non-normal data, followed by a GAMM to further estimate nonlinear change (Fig. 9). The reg-ARMA-GARCH model is not recommended for detecting change in non-normal data, because it is less reliable than a GLMM to detect change. The reg-ARMA-GARCH model has been shown to have convergence difficulties when including an indicator variable in the mean model equation, especially when outliers are present in the data (e.g., Doornik and Ooms, 2005, 2008). These convergence difficulties may have caused the model's inconsistent ability to detect change in the no-change scenarios, and lower power in some of the 25–95% change scenarios relative to other deterministic, parametric models. The reg-ARMA-GARCH model generally exceled in forecasting low-state non-normal data, and provides estimates of autocorrelated error and spikes in aggregation data through the heteroskedastic residual variance equation. Therefore, if appropriately parameterized after an initial inspection of change using the GLMM model, the reg-ARMA-GARCH model is recommended for quantifying change. The model may have less convergence difficulties and produce more accurate estimates with a transformation of the data to a normal distribution.

4.2.3. Forecasting mean change

SVR-L is recommended for forecasting change in mean nekton density data (Fig. 9). Overall, the SVR-L model was the most accurate forecast model for all mean-change scenarios. Additionally, a GAMM may be used to provide a semi-parametric assessment of nonlinear forecasts.

Similar to the normal data recommendations, a SVR-L model in conjunction with a GAMM may be used to forecast change in low-state, non-normal data (Fig. 9). The SVR-L model is most able to accurately forecast mean change in low-state, non-normal data for almost all scenarios, with the exception of nonlinear change. No model was able to capture the spiked structure of the non-normal data, and therefore no model can accurately forecast spikes in non-normal data.

4.2.4. Variance change

The Random Forest regression model is recommended for detecting, quantifying, and forecasting change in variance in both nekton density and aggregation data (Fig. 9). The RF model is recommended over the SVR-RBF model because it had greater forecast accuracy for both datasets. The RF model is more robust in its parameter estimates compared to a SVR-RBF model, which is highly sensitive to tuning and prone to overfitting (Eitrich and Lang, 2006; Ben-Hur and Weston, 2010; Lorena et al., 2011). The measure of variable importance in a RF model can be used to interpret the relative importance of an indicator variable in the model.

5. Conclusion

The model evaluation ensures standardization of regression models that are capable of detecting, quantifying, and forecasting change in operational monitoring data, and collectively forms best practices for the analysis of monitoring data. Standardization of MRE monitoring protocols, including the choice of analytic model, will reduce time and costs of MRE permitting (Dubbs et al., 2013) and subsequent monitoring. Standard monitoring methods also enable comparisons among sites or at different times, which will reduce uncertainty when quantifying environmental effects (Froján et al., 2016). Insight gained from the evaluation on model ability to measure change is generally applicable to any monitoring data with similar properties as the Admiralty Inlet MRE case study. We offer that this evaluation approach can be used as a template to assess the ability of a suite of models to measure change.

The next logical step in developing standardized monitoring methods is to extend the results of the current evaluation, using data with a known point of anthropogenic disturbance, to scenarios of unknown change during operational monitoring. The need for real-time monitoring at MRE tidal or current turbine, wave, and offshore wind commercial sites is current in Europe and imminent in North America (e.g., Rush et al., 2014; Horne et al., 2016). As real-time, operational monitoring is conducted during an anthropogenic disturbance such as MRE operations, there will be unknown points of change (Anderson and Thompson, 2004), and the ability of recommended models to detect and quantify change when disturbances are not known will need to be evaluated. We expect that the demand for initial characterization and operational monitoring will continue to increase as additional countries develop MRE, ocean observing, and other anthropogenic-influenced aquatic sites.

Acknowledgements

We thank Dale Jacques for management and processing of the Admiralty Inlet acoustic backscatter data, Jim Faulkner for assisting with the development of the Hidden Markov Model, and Andrea Copping and Timothy Essington for review of the model evaluation framework. Funding was provided by the National Oceanographic Partnership Program, the US Bureau of Ocean Energy Management, and the National Science Foundation's Sustainable Energy Pathways Program. This work was in part facilitated though the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington.

Appendix A. Parameterized simulation models and candidate models for the normal (Sv) and non-normal (AI) metric data.

Table A1
Parameterized normal data simulation model (MARSS) and non-normal data simulation model (HMM), including selected environmental predictors and error distributions.

Model	Environmental Predictors	Error Distribution
SSM-P	Process Eq: Day, Tidal Range, Day:Tidal Range Observation Eq: Fourier Series	Normal
HMM	Low State: Day, Fourier Series, Tidal Range, Tidal Speed High State: Day, Fourier Series	Gamma (log)

Table A2
Candidate models parameterized using the empirical baseline nekton density (Sv) data with selected environmental predictors and autocorrelation structure for each candidate model. All models assumed a normal distribution.

Model	Environmental Predictors	Autocorrelation Structure (AR,MA)
LM	Day, Fourier Series, Tidal Range, Day: Tidal Range	NA
GLS	Day, Fourier Series, Tidal Range, Day: Tidal Range	(1,0)
GAM	Day, Fourier Series, Tidal Range†, Day:Tidal Range†	NA
GAMM	Day, Fourier Series, Tidal Range†	(1,0)
Reg-ARIMA	Day, Fourier Series, Tidal Range, Day: Tidal Range	(1,0)
Reg-ARMA-GARCH	Mean Eq: Day, Fourier Series, Tidal Range, Day: Tidal Range Variance Eq: Day, Fourier Series, Tidal Range,	ARMA:(1,0) GARCH:(2,3)
SSM-M	Process Eq: Tidal Range, Tidal Speed, Tidal Speed-Tidal Range Observation Eq: Day, Fourier Series, Tidal Speed	(1,0)
SSM-P	Process Eq: Day, Fourier Series, Tidal Range Observation Eq: Day, Tidal Range, Day-Tidal Range	(1,0)
RF	All Environmental Predictors	(14,0)
SVR-L	Fourier Series	(13,0)
SVR-RBF	Day, Fourier Series, Tidal Range, Tidal Speed	(1,0)

Note: The evaluation includes a MARSS model with fixed low observation error (SSM-P) and fixed low process error (SSM-M). A SVR model with a linear kernel (SVR-L) and a nonlinear Radial Basis Function kernel (SVR-RBF) were evaluated. The number of autoregressive (AR) and moving-average (MA) variables in model autocorrelation structures is shown in parenthesis as (AR, MA). The specified (AR, MA) structure of the nonparametric models indicates the number of lagged dependent variables included in the parameterized models. Environmental predictors are listed in alphabetical order of main effects followed by interactions. The Tidal Range† predictor is parametric in the GAM(M)s.

Table A3
Baseline candidate models parameterized using the empirical nekton aggregation (AI) data with selected environmental predictors, autocorrelation structure, and error distribution for each candidate model.

Model	Environmental Predictors	Autocorrelation Structure (AR,MA)	Error Distribution
LM	Fourier Series	NA	Normal
GLS	Fourier Series	(1,0)	Normal
GLM	Day, Fourier Series	NA	Gamma (identity)
GLM	Day, Tidal Range	(1,0)	Gamma (identity)
GAM	Fourier Series, Tidal Speed	NA	Gamma (identity)
GAMM	Fourier Series	(2,0)	Gamma (identity)
Reg-ARIMA	Fourier Series	(1,2)	Normal
Reg-ARMA-GARCH	Mean Eq: Day, Fourier Series, Tidal Range Variance Eq: Fourier Series	ARMA: (1,0) GARCH: (2,0)	Skewed-student-t
SSM-M	Process Eq: Day Observation Eq: Day, Fourier Series	(1,0)	Normal
SSM-P	Observation Eq: Fourier Series	(1,0)	Normal
RF	Fourier Series, Tidal Speed	(2,0)	NA
SVR-L	Day, Fourier Series, Tidal Range, Day: Tidal Range, Tidal Range: Fourier Series	(1,0)	NA
SVR-RBF	Day, Fourier Series, Tidal Speed, Day: Tidal Speed, Tidal Speed: Fourier Series	(13,0)	NA

Note: The evaluation includes a MARSS model with fixed low observation error (SSM-P) and fixed low process error (SSM-M). A SVR model with a linear kernel (SVR-L) and a nonlinear Radial Basis Function kernel (SVR-RBF) were evaluated. The number of autoregressive (AR) and moving-average (MA) variables in model autocorrelation structures is shown in parenthesis as (AR, MA). The specified (AR, MA) structure of the nonparametric models indicates the number of lagged dependent variables included in the parameterized models. The environmental predictors are listed in alphabetical order of main effects followed by interactions.

Appendix B. Ranked model results for the normal (Sv) metric data.

Table B1
Ranked in-sample model fit results and associated average Root-Mean-Squared-Error (RMSE), and Kolmogorov-Smirnov D-Statistic (D-stat) across 10%, 25%, and 95% amplitudes of change for all change scenarios.

	10% Change Amplitude				25% Change Amplitude				95% Change Amplitude			
Change Scenario	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat
Abrupt	1	SSM-P	0.047	0.909	1	SSM-P	0.047	0.909	1	SSM-P	0.048	0.909
	2	SVR-RBF	2.844	0.818	2	SVR-RBF	2.815	0.818	2	SVR-RBF	2.805	0.818
	3	SSM-M	3.620	0.445	3	SSM-M	3.578	0.458	3	SSM-M	3.618	0.564
	4	Reg-ARMA	3.711	0.270	4	Reg-ARMA	3.671	0.355	4	Reg-ARMA	3.800	0.297
	4	Reg-ARMA-GARCH	3.711	0.264	4	Reg-ARMA-GARCH	3.671	0.352	4	Reg-ARMA-GARCH	3.801	0.294
	5	SVR-L	3.787	−0.282	5	SVR-L	3.762	−0.385	5	GAM	3.828	0.279
	6	GAM	3.862	−0.300	6	GAM	3.818	−0.361	6	GAMM	3.944	−0.352
	7	Linear reg	3.873	−0.330	6	Linear reg	3.832	−0.352	7	RF	3.998	−0.448
	7	GLS	3.873	−0.333	6	GLS	3.832	−0.355	8	SVR-L	4.037	−0.539
	8	RF	3.876	−0.415	7	RF	3.870	−0.576	9	Linear reg	4.047	−0.497
	9	GAMM	3.951	−0.467	8	GAMM	3.904	−0.482	9	GLS	4.048	−0.500
Linear	1	SSM-P	0.046	0.909	1	SSM-P	0.047	0.909	1	SSM-P	0.047	0.909
	2	SVR-RBF	2.868	0.818	2	SVR-RBF	2.825	0.818	2	SVR-RBF	2.840	0.818
	3	SSM-M	3.641	0.436	3	SSM-M	3.585	0.512	3	SSM-M	3.674	0.421
	4	Reg-ARMA	3.737	0.282	4	Reg-ARMA	3.675	0.327	4	Reg-ARMA	3.729	0.321
	4	Reg-ARMA-GARCH	3.737	0.279	4	Reg-ARMA-GARCH	3.675	0.324	4	Reg-ARMA-GARCH	3.729	0.318
	5	SVR-L	3.802	−0.224	5	SVR-L	3.747	−0.267	5	SVR-L	3.852	−0.273
	6	RF	3.880	−0.327	6	GAM	3.820	−0.348	6	GAM	3.885	−0.348
	7	GAM	3.890	−0.336	7	RF	3.828	−0.364	7	Linear reg	3.894	−0.355
	7	Linear reg	3.896	−0.342	7	Linear reg	3.836	−0.391	7	GLS	3.894	−0.358
	7	GLS	3.897	−0.345	7	GLS	3.837	−0.394	8	RF	3.925	−0.488
	8	GAMM	3.955	−0.461	8	GAMM	3.907	−0.582	9	GAMM	3.963	−0.497
Nonlinear	1	SSM-P	0.047	0.909	1	SSM-P	0.047	0.909	1	SSM-P	0.049	0.909
	2	SVR-RBF	2.806	0.818	2	SVR-RBF	2.826	0.818	2	SVR-RBF	2.820	0.818
	3	SSM-M	3.557	0.470	3	SSM-M	3.573	0.476	3	SSM-M	3.613	0.558
	4	Reg-ARMA	3.664	0.255	4	Reg-ARMA	3.681	0.252	4	Reg-ARMA	3.788	0.209
	4	Reg-ARMA-GARCH	3.664	0.252	4	Reg-ARMA-GARCH	3.681	0.248	4	Reg-ARMA-GARCH	3.788	−0.209
	5	SVR-L	3.742	−0.264	5	SVR-L	3.754	−0.252	5	GAM	3.829	−0.264
	6	GAM	3.799	−0.345	6	GAM	3.826	−0.364	6	SVR-L	3.858	−0.279
	6	Linear reg	3.809	−0.367	7	RF	3.839	−0.361	7	RF	3.904	−0.370
	6	GLS	3.809	−0.370	7	Linear reg	3.843	−0.400	8	GAMM	3.928	−0.327
	6	RF	3.814	−0.348	7	GLS	3.843	−0.406	9	Linear reg	4.034	−0.527
	7	GAMM	3.869	−0.433	8	GAMM	3.896	−0.530	9	GLS	4.035	−0.530
Periodic	1	SSM-P	0.044	0.909	1	SSM-P	0.043	0.909	1	SSM-P	0.035	0.909
	2	SVR-RBF	2.924	0.818	2	SVR-RBF	2.956	0.818	2	SVR-RBF	3.202	0.818
	3	SSM-M	3.806	0.348	3	SSM-M	3.869	0.424	3	RF	4.213	0.715
	4	Reg-ARMA	3.911	0.200	4	Reg-ARMA	4.016	−0.212	4	SSM-M	4.604	0.458
	4	Reg-ARMA-GARCH	3.913	−0.200	4	Reg-ARMA-GARCH	4.020	−0.215	5	SVR-L	4.764	−0.285
	5	SVR-L	3.979	−0.273	5	SVR-L	4.066	−0.236	6	Reg-ARMA	4.871	−0.348
	6	RF	4.012	−0.294	6	RF	4.096	−0.267	7	Reg-ARMA-GARCH	4.884	−0.358
	7	GAM	4.048	−0.318	7	GAM	4.150	−0.333	8	Linear reg	4.965	−0.421
	7	Linear reg	4.056	−0.312	7	Linear reg	4.155	−0.370	8	GLS	4.965	−0.424
	7	GLS	4.056	−0.315	7	GLS	4.155	−0.373	8	GAM	4.967	−0.415
	8	GAMM	4.112	−0.418	8	GAMM	4.210	−0.503	9	GAMM	5.030	−0.512
Step	1	SSM-P	0.047	0.909	1	SSM-P	0.047	0.909	1	SSM-P	0.047	0.909
	2	SVR-RBF	2.827	0.818	2	SVR-RBF	2.854	0.818	2	SVR-RBF	2.810	0.818
	3	SSM-M	3.624	0.367	3	SSM-M	3.600	0.485	3	SSM-M	3.564	0.542
	4	Reg-ARMA	3.684	0.273	4	Reg-ARMA	3.701	0.364	4	Reg-ARMA	3.664	0.336
	4	Reg-ARMA-GARCH	3.685	0.270	4	Reg-ARMA-GARCH	3.701	0.361	4	Reg-ARMA-GARCH	3.664	0.333
	5	SVR-L	3.750	−0.248	5	SVR-L	3.769	−0.252	5	SVR-L	3.773	−0.318
	6	GAM	3.832	−0.330	6	GAM	3.861	−0.406	6	GAM	3.812	−0.333
	7	RF	3.844	−0.370	7	Linear reg	3.874	−0.418	7	Linear reg	3.827	−0.358
	7	Linear reg	3.850	−0.373	7	GLS	3.875	−0.424	7	GLS	3.828	−0.361
	7	GLS	3.851	−0.376	8	RF	3.878	−0.470	8	RF	3.878	−0.482
	8	GAMM	3.911	−0.476	9	GAMM	3.934	−0.518	8	GAMM	3.891	−0.470

Table B2
Ranked in-sample model fit results and associated average Root-Mean-Squared-Error (RMSE), and Kolmogorov-Smirnov D-Statistic (D-stat) for the 10% and 95% Linear and Nonlinear change scenarios and corresponding lagged change scenarios.

		Non-Lagged Change				Lagged Change			
Change Amplitude	Change Scenario	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat
10%	Linear	1	SSM-P	0.046	0.909	1	SSM-P	0.047	1.000
		2	SVR-RBF	2.868	0.818	2	SVR-RBF	2.842	0.900
		3	SSM-M	3.641	0.436	3	SSM-M	3.602	0.510
		4	Reg-ARMA	3.737	0.282	4	Reg-ARMA	3.699	0.333
		4	Reg-ARMA-GARCH	3.737	0.279	4	Reg-ARMA-GARCH	3.699	0.330
		5	SVR-L	3.802	−0.224	5	SVR-L	3.761	−0.267
		6	RF	3.880	−0.327	6	GAM	3.844	−0.437
		7	GAM	3.890	−0.336	6	RF	3.844	−0.397
		7	Linear reg	3.896	−0.342	7	Linear reg	3.858	−0.470
		7	GLS	3.897	−0.345	7	GLS	3.858	−0.473
		8	GAMM	3.955	−0.461	8	GAMM	3.923	−0.577
	Nonlinear	1	SSM-P	0.047	0.909	1	SSM-P	0.047	1.000
		2	SVR-RBF	2.806	0.818	2	SVR-RBF	2.816	0.900
		3	SSM-M	3.557	0.470	3	SSM-M	3.569	0.553
		4	Reg-ARMA	3.664	0.255	4	Reg-ARMA	3.680	0.373
		4	Reg-ARMA-GARCH	3.664	0.252	4	Reg-ARMA-GARCH	3.680	0.370
		5	SVR-L	3.742	−0.264	5	SVR-L	3.733	−0.237
		6	GAM	3.799	−0.345	6	RF	3.820	−0.407
		6	Linear reg	3.809	−0.367	6	GAM	3.821	−0.370
		6	GLS	3.809	−0.370	6	Linear reg	3.833	−0.433
		6	RF	3.814	−0.348	6	GLS	3.834	−0.437
		7	GAMM	3.869	−0.433	7	GAMM	3.894	−0.560
95%	Linear	1	SSM-P	0.047	0.909	1	SSM-P	0.047	1.000
		2	SVR-RBF	2.840	0.818	2	SVR-RBF	2.817	0.900
		3	SSM-M	3.674	0.421	3	SSM-M	3.589	0.560
		4	Reg-ARMA	3.729	0.321	4	Reg-ARMA	3.731	0.303
		4	Reg-ARMA-GARCH	3.729	0.318	4	Reg-ARMA-GARCH	3.731	0.293
		5	SVR-L	3.852	−0.273	5	SVR-L	3.789	−0.267
		6	GAM	3.885	−0.348	6	GAM	3.823	−0.307
		7	Linear reg	3.894	−0.355	7	RF	3.841	−0.353
		7	GLS	3.894	−0.358	8	GAMM	3.930	−0.513
		8	RF	3.925	−0.488	9	Linear reg	3.953	−0.533
		9	GAMM	3.963	−0.497	9	GLS	3.953	−0.537
	Nonlinear	1	SSM-P	0.049	0.909	1	SSM-P	0.047	1.000
		2	SVR-RBF	2.820	0.818	2	SVR-RBF	2.836	0.900
		3	SSM-M	3.613	0.558	3	SSM-M	3.612	0.553
		4	Reg-ARMA	3.788	0.209	4	Reg-ARMA	3.774	−0.237
		4	Reg-ARMA-GARCH	3.788	−0.209	4	Reg-ARMA-GARCH	3.775	−0.240
		5	GAM	3.829	−0.264	5	SVR-L	3.821	−0.237
		6	SVR-L	3.858	−0.279	6	GAM	3.831	−0.260
		7	RF	3.904	−0.370	7	RF	3.879	−0.333
		8	GAMM	3.928	−0.327	8	GAMM	3.983	−0.487
		9	Linear reg	4.034	−0.527	9	Linear reg	3.994	−0.530
		9	GLS	4.035	−0.530	9	GLS	3.995	−0.533

Table B3
Ranked model forecast results and associated average Mean-Absolute-Scale-Error (MASE), and Kolmogorov-Smirnov D-Statistic (D-stat) across 10%, 25%, and 95% amplitudes of change for all change scenarios.

	10% Change Amplitude				25% Change Amplitude				95% Change Amplitude			
Change Scenario	Rank	Model	MASE	D-stat	Rank	Model	MASE	D-stat	Rank	Model	MASE	D-stat
Abrupt	1	SVR-L	0.816	0.358	1	SVR-L	0.834	0.303	1	RF	0.825	0.700
	2	RF	0.828	0.312	1	RF	0.837	0.315	2	SVR-L	0.840	0.645
	3	SSM-P	0.863	0.136	2	SSM-M	0.886	0.152	3	SVR-RBF	0.976	0.355
	3	SSM-M	0.863	0.182	2	Linear reg	0.887	0.091	4	SSM-M	1.074	−0.170
	4	Reg-ARMA	0.881	−0.091	2	GLS	0.887	0.103	5	Reg-ARMA	1.109	−0.236
	4	Reg-ARMA-GARCH	0.882	−0.097	2	Reg-ARMA	0.888	0.091	5	GLS	1.109	−0.239
	4	GLS	0.882	−0.109	2	Reg-ARMA-GARCH	0.889	0.094	5	Linear reg	1.110	−0.236
	4	Linear reg	0.882	−0.124	3	SSM-P	0.899	−0.121	5	Reg-ARMA-GARCH	1.117	−0.242
	5	GAM	0.893	−0.158	4	SVR-RBF	0.931	−0.200	6	GAM	1.168	−0.158
	6	SVR-RBF	0.916	−0.270	5	GAM	0.939	−0.167	6	GAMM	1.185	−0.185
	7	GAMM	0.954	−0.391	6	GAMM	0.977	−0.358	7	SSM-P	1.697	−0.803
Linear	1	SVR-L	0.820	0.324	1	SVR-L	0.820	0.273	1	Reg-ARMA	0.841	0.367
	2	RF	0.831	0.233	2	RF	0.841	0.212	1	Reg-ARMA-GARCH	0.841	0.376
	3	SSM-P	0.851	0.182	3	Reg-ARMA	0.858	0.079	1	GLS	0.842	0.348
	4	Reg-ARMA-GARCH	0.871	0.094	3	Reg-ARMA-GARCH	0.859	0.085	1	Linear reg	0.843	0.345
	4	GLS	0.872	0.106	3	GLS	0.859	0.088	2	GAM	0.857	0.285
	4	Reg-ARMA	0.872	0.097	3	Linear reg	0.860	0.088	3	SVR-L	0.865	0.297
	4	Linear reg	0.873	0.094	3	SSM-P	0.861	0.097	4	GAMM	0.919	0.261
	5	SSM-M	0.877	−0.236	4	SSM-M	0.869	−0.061	5	RF	1.002	−0.394
	6	GAM	0.887	−0.097	5	GAM	0.881	−0.130	6	SSM-M	1.156	−0.367
	7	SVR-RBF	0.926	−0.312	6	SVR-RBF	0.926	−0.324	7	SSM-P	1.569	−0.752
	8	GAMM	0.929	−0.415	7	GAMM	0.939	−0.315	8	SVR-RBF	1.592	−0.770
Nonlinear	1	SVR-L	0.796	0.267	1	SVR-L	0.815	0.300	1	SVR-L	0.829	0.724
	2	RF	0.813	0.176	2	RF	0.839	0.188	2	RF	0.853	0.679
	3	SSM-P	0.825	0.164	3	SSM-M	0.852	0.200	3	GAM	1.118	0.133
	4	GAM	0.837	0.088	3	Reg-ARMA-GARCH	0.854	0.124	3	Reg-ARMA	1.120	−0.200
	4	Reg-ARMA	0.840	0.064	3	GLS	0.854	0.121	3	GLS	1.120	−0.203
	4	Reg-ARMA-GARCH	0.840	0.067	3	Linear reg	0.855	0.124	3	Linear reg	1.122	−0.212
	4	GLS	0.840	0.070	3	Reg-ARMA	0.855	0.127	3	Reg-ARMA-GARCH	1.123	−0.194
	4	Linear reg	0.840	0.094	4	SSM-P	0.861	0.067	3	GAMM	1.132	0.115
	4	SSM-M	0.840	0.091	5	GAM	0.878	−0.100	4	SSM-M	1.227	−0.352
	5	SVR-RBF	0.903	−0.312	6	GAMM	0.937	−0.270	5	SVR-RBF	1.351	−0.452
	5	GAMM	0.909	−0.318	7	SVR-RBF	0.940	−0.400	6	SSM-P	1.702	−0.752
Periodic	1	RF	0.797	0.270	1	RF	0.871	0.230	1	RF	0.765	0.758
	1	SVR-L	0.804	0.291	2	SVR-L	0.886	0.155	2	SVR-RBF	0.873	0.579
	2	Reg-ARMA-GARCH	0.832	0.121	2	SSM-M	0.893	0.167	3	SVR-L	0.957	0.270
	2	SSM-P	0.840	−0.106	3	SVR-RBF	0.915	−0.130	4	SSM-M	0.962	0.333
	2	Reg-ARMA	0.842	−0.115	3	GAM	0.918	−0.082	5	GAM	1.040	−0.221
	2	GAM	0.842	−0.127	3	Reg-ARMA-GARCH	0.919	−0.082	6	GAMM	1.050	−0.252
	2	GLS	0.842	−0.079	3	GLS	0.922	−0.073	6	Linear reg	1.052	−0.258
	2	Linear reg	0.842	−0.073	3	Linear reg	0.922	−0.076	6	GLS	1.052	−0.252
	2	SSM-M	0.853	−0.155	3	SSM-P	0.923	−0.085	6	SSM-P	1.053	−0.255
	3	SVR-RBF	0.876	−0.233	3	Reg-ARMA	0.924	−0.088	6	Reg-ARMA	1.055	−0.258
	4	GAMM	0.877	−0.167	4	GAMM	0.944	−0.209	7	Reg-ARMA-GARCH	1.063	−0.303
Step	1	SVR-L	0.811	0.221	1	SVR-L	0.790	0.233	1	SVR-L	0.806	0.255
	1	RF	0.816	0.179	2	RF	0.801	0.215	2	RF	0.823	0.158
	2	SSM-P	0.837	0.142	3	SSM-P	0.811	0.103	2	SSM-P	0.827	0.170
	2	Reg-ARMA-GARCH	0.839	0.121	3	Reg-ARMA-GARCH	0.811	0.109	2	Reg-ARMA	0.828	0.152
	2	Reg-ARMA	0.839	0.118	3	Reg-ARMA	0.812	0.091	2	Reg-ARMA-GARCH	0.828	0.164
	2	GLS	0.840	0.103	3	Linear reg	0.812	0.091	2	Linear reg	0.829	0.158
	2	Linear reg	0.840	0.106	3	GLS	0.812	0.073	2	GLS	0.829	0.167
	3	GAM	0.856	−0.121	4	GAM	0.827	−0.082	3	GAM	0.835	0.109
	4	SSM-M	0.860	−0.185	5	SSM-M	0.838	−0.121	4	SSM-M	0.857	−0.191
	5	GAMM	0.879	−0.221	6	GAMM	0.862	−0.242	5	GAMM	0.879	−0.297
	6	SVR-RBF	0.897	−0.345	7	SVR-RBF	0.896	−0.312	6	SVR-RBF	0.983	−0.624

Table B4
Ranked model forecast results and associated average Root-Mean-Squared-Error (RMSE), and Kolmogorov-Smirnov D-Statistic (D-stat) across 10%, 25%, and 95% amplitudes of change for all change scenarios.

	10% Change Amplitude				25% Change Amplitude				95% Change Amplitude			
Change Scenario	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat
Abrupt	1	SVR-L	3.810	0.361	1	SVR-L	3.963	0.239	1	RF	3.931	0.682
	2	RF	3.880	0.315	1	RF	3.980	0.279	2	SVR-L	4.005	0.642
	3	SSM-P	4.028	0.155	2	Linear reg	4.154	0.091	3	SVR-RBF	4.635	0.279
	3	SSM-M	4.046	0.161	2	GLS	4.155	0.112	4	SSM-M	5.050	−0.218
	4	Reg-ARMA	4.108	−0.109	2	Reg-ARMA	4.156	0.124	5	Reg-ARMA	5.159	−0.242
	4	GLS	4.108	−0.094	2	Reg-ARMA-GARCH	4.160	0.121	5	GLS	5.162	−0.239
	4	Reg-ARMA-GARCH	4.111	−0.106	2	SSM-M	4.177	0.112	5	Linear reg	5.166	−0.236
	4	Linear reg	4.111	−0.139	3	SSM-P	4.236	−0.100	5	Reg-ARMA-GARCH	5.193	−0.270
	5	GAM	4.158	−0.148	4	GAM	4.361	−0.136	6	GAM	5.362	−0.167
	6	SVR-RBF	4.304	−0.267	5	SVR-RBF	4.419	−0.221	6	GAMM	5.436	−0.142
	7	GAMM	4.468	−0.436	6	GAMM	4.534	−0.348	7	SSM-P	7.525	−0.806
Linear	1	SVR-L	3.900	0.300	1	SVR-L	3.838	0.230	1	Reg-ARMA-GARCH	3.986	0.376
	2	RF	3.963	0.245	2	RF	3.941	0.173	1	Reg-ARMA	3.986	0.373
	3	SSM-P	4.039	0.130	3	Reg-ARMA	3.996	0.091	1	GLS	3.991	0.367
	4	Reg-ARMA-GARCH	4.114	0.088	3	Reg-ARMA-GARCH	3.997	0.094	1	Linear reg	3.996	0.364
	4	Reg-ARMA	4.115	0.085	3	GLS	4.000	0.094	2	GAM	4.055	0.309
	4	GLS	4.116	0.082	3	Linear reg	4.007	0.076	3	SVR-L	4.081	0.294
	4	Linear reg	4.122	−0.061	3	SSM-P	4.018	0.103	4	GAMM	4.350	0.248
	5	SSM-M	4.154	−0.161	4	SSM-M	4.041	0.109	5	RF	4.743	−0.358
	6	GAM	4.165	−0.073	5	GAM	4.105	−0.106	6	SSM-M	5.331	−0.352
	7	GAMM	4.355	−0.342	6	SVR-RBF	4.320	−0.342	7	SSM-P	6.877	−0.748
	7	SVR-RBF	4.390	−0.355	7	GAMM	4.348	−0.258	8	SVR-RBF	7.311	−0.779
Nonlinear	1	SVR-L	3.712	0.321	1	SVR-L	3.829	0.252	1	SVR-L	3.885	0.715
	2	RF	3.806	0.203	2	RF	3.916	0.167	2	RF	4.044	0.621
	3	SSM-P	3.850	0.139	3	Reg-ARMA-GARCH	3.992	0.109	3	GAM	5.114	0.142
	4	GAM	3.911	0.103	3	SSM-M	3.992	0.158	4	GAMM	5.193	0.118
	4	Linear reg	3.920	0.067	3	GLS	3.994	0.094	5	Reg-ARMA	5.199	−0.164
	4	GLS	3.922	0.064	3	Reg-ARMA	3.997	0.106	5	GLS	5.199	−0.170
	4	Reg-ARMA	3.923	0.061	3	Linear reg	3.998	0.100	5	Linear reg	5.210	−0.176
	4	Reg-ARMA-GARCH	3.923	0.082	4	SSM-P	4.032	−0.073	5	Reg-ARMA-GARCH	5.212	−0.179
	4	SSM-M	3.928	0.109	5	GAM	4.085	−0.091	6	SSM-M	5.573	−0.315
	5	SVR-RBF	4.227	−0.421	6	GAMM	4.340	−0.285	7	SVR-RBF	6.281	−0.464
	5	GAMM	4.235	−0.300	7	SVR-RBF	4.362	−0.373	8	SSM-P	7.387	−0.712
Periodic	1	RF	3.991	0.288	1	RF	4.438	0.285	1	RF	4.722	0.845
	2	SVR-L	4.026	0.218	2	SVR-L	4.511	0.155	2	SVR-RBF	5.406	0.564
	3	Reg-ARMA-GARCH	4.158	0.103	2	SSM-M	4.579	0.182	3	SSM-M	5.782	0.306
	3	SSM-P	4.192	0.094	3	SVR-RBF	4.669	0.115	4	SVR-L	5.822	0.261
	3	GLS	4.202	−0.103	3	Reg-ARMA-GARCH	4.680	−0.103	5	GAM	6.180	−0.224
	3	Reg-ARMA	4.203	−0.097	3	GAM	4.687	−0.079	6	Linear reg	6.231	−0.242
	3	Linear reg	4.203	−0.106	3	SSM-P	4.701	−0.118	6	GLS	6.232	−0.245
	3	GAM	4.203	−0.115	3	GLS	4.701	−0.112	6	SSM-P	6.236	−0.255
	3	SSM-M	4.286	−0.173	3	Linear reg	4.703	−0.118	6	Reg-ARMA	6.245	−0.261
	4	GAMM	4.359	−0.170	3	Reg-ARMA	4.707	−0.130	7	Reg-ARMA-GARCH	6.268	−0.273
	5	SVR-RBF	4.380	−0.258	4	GAMM	4.822	−0.258	7	GAMM	6.286	−0.279
Step	1	SVR-L	3.744	0.206	1	SVR-L	3.732	0.270	1	SVR-L	3.778	0.273
	2	RF	3.787	0.167	2	RF	3.771	0.161	2	RF	3.858	0.185
	3	SSM-P	3.885	0.130	3	SSM-P	3.837	0.094	3	SSM-P	3.906	0.164
	3	Reg-ARMA-GARCH	3.891	0.088	3	Reg-ARMA	3.839	0.085	3	Reg-ARMA	3.909	0.152
	3	Reg-ARMA	3.892	0.091	3	Reg-ARMA-GARCH	3.839	0.091	3	Reg-ARMA-GARCH	3.909	0.155
	3	GLS	3.894	0.094	3	Linear reg	3.842	0.097	3	Linear reg	3.910	0.152
	3	Linear reg	3.896	0.082	3	GLS	3.843	0.088	3	GLS	3.911	0.142
	4	GAM	3.955	−0.130	4	GAM	3.906	−0.103	4	GAM	3.940	0.118
	5	SSM-M	3.977	−0.142	5	SSM-M	3.942	−0.142	5	SSM-M	4.026	−0.173
	6	GAMM	4.068	−0.185	6	GAMM	4.065	−0.279	6	GAMM	4.127	−0.248
	7	SVR-RBF	4.131	−0.330	7	SVR-RBF	4.174	−0.282	7	SVR-RBF	4.559	−0.679

Table B5
Ranked forecast model fit results and associated average Mean-Absolute-Scaled-Error (MASE), and Kolmogorov-Smirnov D-Statistic (D-stat) for the 10% and 95% Linear and Nonlinear change scenarios and corresponding lagged change scenarios.

		Non-Lagged Change				Lagged Change			
Change Amplitude	Change Scenario	Rank	Model	MASE	D-stat	Rank	Model	MASE	D-stat
10%	Linear	1	SVR-L	0.820	0.324	1	SVR-L	0.815	0.323
		2	RF	0.831	0.233	1	RF	0.816	0.267
		3	SSM-P	0.851	0.182	2	SSM-P	0.838	0.170
		4	Reg-ARMA-GARCH	0.871	0.094	3	Reg-ARMA-GARCH	0.849	0.083
		4	GLS	0.872	0.106	3	Reg-ARMA	0.849	0.090
		4	Reg-ARMA	0.872	0.097	3	GLS	0.849	0.097
		4	Linear reg	0.873	0.094	3	Linear reg	0.850	0.090
		5	SSM-M	0.877	−0.236	3	GAM	0.855	−0.073
		6	GAM	0.887	−0.097	4	SSM-M	0.877	−0.157
		7	SVR-RBF	0.926	−0.312	5	GAMM	0.924	−0.430
		8	GAMM	0.929	−0.415	5	SVR-RBF	0.934	−0.330
	Nonlinear	1	SVR-L	0.796	0.267	1	SVR-L	0.813	0.270
		2	RF	0.813	0.176	2	RF	0.825	0.290
		3	SSM-P	0.825	0.164	2	SSM-P	0.832	0.203
		4	GAM	0.837	0.088	3	Reg-ARMA	0.846	0.103
		4	Reg-ARMA	0.840	0.064	3	Reg-ARMA-GARCH	0.846	0.100
		4	Reg-ARMA-GARCH	0.840	0.067	3	GLS	0.847	0.097
		4	GLS	0.840	0.070	3	Linear reg	0.847	0.093
		4	Linear reg	0.840	0.094	4	SSM-M	0.858	−0.120
		4	SSM-M	0.840	0.091	4	GAM	0.872	0.087
		5	SVR-RBF	0.903	−0.312	5	GAMM	0.906	−0.333
		5	GAMM	0.909	−0.318	6	SVR-RBF	0.937	−0.460
95%	Linear	1	Reg-ARMA	0.841	0.367	1	GAM	0.994	0.653
		1	Reg-ARMA-GARCH	0.841	0.376	2	GAMM	1.084	0.553
		1	GLS	0.842	0.348	3	SVR-L	1.084	0.660
		1	Linear reg	0.843	0.345	4	RF	1.371	0.343
		2	GAM	0.857	0.285	5	Reg-ARMA-GARCH	1.470	−0.237
		3	SVR-L	0.865	0.297	5	Reg-ARMA	1.479	−0.233
		4	GAMM	0.919	0.261	5	GLS	1.482	−0.230
		5	RF	1.002	−0.394	5	Linear reg	1.491	−0.243
		6	SSM-M	1.156	−0.367	6	SSM-M	1.811	−0.463
		7	SSM-P	1.569	−0.752	7	SVR-RBF	1.999	−0.633
		8	SVR-RBF	1.592	−0.770	8	SSM-P	2.200	−0.850
	Nonlinear	1	SVR-L	0.829	0.724	1	GAM	1.250	0.723
		2	RF	0.853	0.679	2	SVR-L	1.320	0.727
		3	GAM	1.118	0.133	3	RF	1.363	0.647
		3	Reg-ARMA	1.120	−0.200	4	GAMM	1.955	0.240
		3	GLS	1.120	−0.203	5	SSM-M	2.198	−0.260
		3	Linear reg	1.122	−0.212	6	Reg-ARMA	2.326	−0.313
		3	Reg-ARMA-GARCH	1.123	−0.194	6	GLS	2.334	−0.320
		3	GAMM	1.132	0.115	6	Reg-ARMA-GARCH	2.350	−0.330
		4	SSM-M	1.227	−0.352	6	Linear reg	2.356	−0.323
		5	SVR-RBF	1.351	−0.452	7	SVR-RBF	2.396	−0.337
		6	SSM-P	1.702	−0.752	8	SSM-P	2.927	−0.783

Table B6
Ranked forecast model fit results and associated average Root-Mean-Squared-Error (RMSE), and Kolmogorov-Smirnov D-Statistic (D-stat) for the 10% and 95% Linear and Nonlinear change scenarios and corresponding lagged change scenarios.

		Non-Lagged Change				Lagged Change			
Change Amplitude	Change Scenario	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat
10%	Linear	1	SVR-L	3.900	0.300	1	RF	3.849	0.307
		2	RF	3.963	0.245	1	SVR-L	3.849	0.353
		3	SSM-P	4.039	0.130	2	SSM-P	3.952	0.183
		4	Reg-ARMA-GARCH	4.114	0.088	3	GLS	3.999	0.100
		4	Reg-ARMA	4.115	0.085	3	Reg-ARMA-GARCH	4.000	0.103
		4	GLS	4.116	0.082	3	Reg-ARMA	4.000	0.110
		4	Linear reg	4.122	−0.061	3	Linear reg	4.001	0.103
		5	SSM-M	4.154	−0.161	3	GAM	4.021	0.113
		6	GAM	4.165	−0.073	4	SSM-M	4.131	−0.250
		7	GAMM	4.355	−0.342	5	GAMM	4.339	−0.460
		7	SVR-RBF	4.390	−0.355	5	SVR-RBF	4.353	−0.403
	Nonlinear	1	SVR-L	3.712	0.321	1	SVR-L	3.787	0.353
		2	RF	3.806	0.203	2	RF	3.855	0.383
		3	SSM-P	3.850	0.139	3	SSM-P	3.915	0.167
		4	GAM	3.911	0.103	4	Reg-ARMA	3.977	0.103
		4	Linear reg	3.920	0.067	4	Reg-ARMA-GARCH	3.978	0.110
		4	GLS	3.922	0.064	4	GLS	3.981	0.097
		4	Reg-ARMA	3.923	0.061	4	Linear reg	3.984	0.093
		4	Reg-ARMA-GARCH	3.923	0.082	5	SSM-M	4.039	−0.157
		4	SSM-M	3.928	0.109	5	GAM	4.098	−0.140
		5	SVR-RBF	4.227	−0.421	6	GAMM	4.260	−0.357
		5	GAMM	4.235	−0.300	7	SVR-RBF	4.387	−0.500
95%	Linear	1	Reg-ARMA-GARCH	3.986	0.376	1	GAM	4.599	0.703
		1	Reg-ARMA	3.986	0.373	2	SVR-L	4.993	0.700
		1	GLS	3.991	0.367	3	GAMM	5.014	0.513
		1	Linear reg	3.996	0.364	4	RF	6.286	0.317
		2	GAM	4.055	0.309	5	Reg-ARMA-GARCH	6.539	0.247
		3	SVR-L	4.081	0.294	5	Reg-ARMA	6.571	0.243
		4	GAMM	4.350	0.248	5	GLS	6.577	0.237
		5	RF	4.743	−0.358	5	Linear reg	6.609	0.223
		6	SSM-M	5.331	−0.352	6	SSM-M	7.959	−0.513
		7	SSM-P	6.877	−0.748	7	SVR-RBF	9.004	−0.777
		8	SVR-RBF	7.311	−0.779	8	SSM-P	9.188	−0.790
	Nonlinear	1	SVR-L	3.885	0.715	1	GAM	5.791	0.677
		2	RF	4.044	0.621	2	SVR-L	5.938	0.760
		3	GAM	5.114	0.142	3	RF	6.138	0.653
		4	GAMM	5.193	0.118	4	GAMM	8.259	0.227
		5	Reg-ARMA	5.199	−0.164	5	SSM-M	9.279	−0.260
		5	GLS	5.199	−0.170	6	Reg-ARMA	9.604	−0.293
		5	Linear reg	5.210	−0.176	6	GLS	9.624	−0.297
		5	Reg-ARMA-GARCH	5.212	−0.179	6	Reg-ARMA-GARCH	9.691	−0.313
		6	SSM-M	5.573	−0.315	6	Linear reg	9.700	−0.307
		7	SVR-RBF	6.281	−0.464	7	SVR-RBF	10.327	−0.417
		8	SSM-P	7.387	−0.712	8	SSM-P	11.747	−0.763

Appendix C. Ranked model results for the non-normal (AI) metric data.

Table C1
Ranked in-sample model fit results and associated average Root-Mean-Squared-Error (RMSE), and Kolmogorov-Smirnov D-Statistic (D-stat) across 10%, 25%, and 95% amplitudes of change for all change scenarios.

	10% Change Amplitude				25% Change Amplitude				95% Change Amplitude			
Change Scenario	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat
Abrupt	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923
	2	SVR-RBF	0.043	0.641	2	SVR-RBF	0.045	0.723	2	SVR-RBF	0.046	0.577
	3	SSM-M	0.054	0.136	3	SSM-M	0.056	0.164	3	SSM-M	0.055	0.195
	4	GLMM	0.055	−0.126	4	GAM	0.058	−0.151	4	GAM	0.058	−0.144
	4	GAM	0.055	−0.144	4	GLMM	0.058	−0.146	4	GAMM	0.058	−0.146
	4	Reg-ARMA	0.056	−0.141	4	GAMM	0.058	−0.159	5	GLMM	0.059	−0.162
	4	GAMM	0.056	−0.151	4	Reg-ARMA	0.059	−0.149	5	Reg-ARMA	0.059	−0.159
	4	GLM	0.056	−0.144	4	Linear reg	0.059	−0.174	5	Linear reg	0.059	−0.172
	4	Linear reg	0.056	−0.156	4	GLS	0.059	−0.177	5	GLS	0.059	−0.172
	4	GLS	0.056	−0.159	4	GLM	0.059	−0.172	6	GLM	0.059	−0.187
	5	RF	0.057	−0.197	5	Reg-ARMA-GARCH	0.060	−0.259	7	Reg-ARMA-GARCH	0.060	−0.231
	5	Reg-ARMA-GARCH	0.057	−0.233	6	RF	0.061	−0.354	8	RF	0.061	−0.274
	6	SVR-L	0.058	−0.269	7	SVR-L	0.061	−0.328	8	SVR-L	0.061	−0.300
Linear	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923
	2	SVR-RBF	0.046	0.605	2	SVR-RBF	0.047	0.626	2	SVR-RBF	0.045	0.600
	3	SSM-M	0.057	0.128	3	SSM-M	0.058	−0.126	3	SSM-M	0.055	0.159
	4	GLMM	0.058	−0.131	4	GLMM	0.059	−0.138	4	GLMM	0.058	−0.144
	4	GAM	0.058	−0.138	4	GAM	0.059	−0.133	4	Reg-ARMA	0.058	−0.151
	5	Reg-ARMA	0.058	−0.138	5	Reg-ARMA	0.060	−0.149	4	GAM	0.058	−0.146
	5	GLM	0.058	−0.133	5	GLM	0.060	−0.156	4	GLM	0.058	−0.162
	5	Linear reg	0.058	−0.151	5	GAMM	0.060	−0.162	4	Linear reg	0.058	−0.156
	5	GLS	0.058	−0.154	5	Linear reg	0.060	−0.167	4	GLS	0.058	−0.159
	5	GAMM	0.058	−0.146	5	GLS	0.060	−0.169	4	GAMM	0.058	−0.164
	6	RF	0.060	−0.246	6	RF	0.061	−0.228	5	Reg-ARMA-GARCH	0.059	−0.208
	6	Reg-ARMA-GARCH	0.060	−0.244	6	Reg-ARMA-GARCH	0.061	−0.254	6	SVR-L	0.061	−0.259
	7	SVR-L	0.061	−0.300	7	SVR-L	0.062	−0.308	7	RF	0.061	−0.310
Nonlinear	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923
	2	SVR-RBF	0.045	0.621	2	SVR-RBF	0.046	0.613	2	SVR-RBF	0.043	0.738
	3	SSM-M	0.056	0.144	3	SSM-M	0.056	0.138	3	SSM-M	0.054	0.331
	4	GLMM	0.057	−0.131	4	GAM	0.057	−0.131	4	GAM	0.057	0.213
	4	GAM	0.057	−0.133	4	GLMM	0.058	−0.128	5	GAMM	0.058	0.177
	5	Reg-ARMA	0.057	−0.144	5	GAMM	0.058	−0.138	6	Reg-ARMA	0.058	0.162
	5	GLM	0.057	−0.156	5	Reg-ARMA	0.058	−0.133	7	RF	0.060	−0.203
	5	GAMM	0.058	−0.159	5	GLM	0.058	−0.144	8	Linear reg	0.063	−0.279
	5	Linear reg	0.058	−0.149	5	Linear reg	0.058	−0.149	8	GLS	0.063	−0.282
	5	GLS	0.058	−0.151	5	GLS	0.058	−0.151	8	SVR-L	0.063	−0.310
	6	RF	0.059	−0.187	6	Reg-ARMA-GARCH	0.060	−0.200	9	Reg-ARMA-GARCH	0.065	−0.336
	7	Reg-ARMA-GARCH	0.059	−0.231	6	RF	0.060	−0.210	10	GLMM	0.066	−0.359
	8	SVR-L	0.060	−0.279	7	SVR-L	0.061	−0.244	11	GLM	0.067	−0.456
Periodic	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923
	2	SVR-RBF	0.042	0.831	2	SVR-RBF	0.040	0.841	2	SVR-RBF	0.041	0.846
	3	RF	0.059	−0.151	3	RF	0.058	0.344	3	RF	0.062	0.769
	3	SSM-M	0.059	−0.149	4	SSM-M	0.062	−0.172	4	SSM-M	0.091	−0.231
	4	GAM	0.060	−0.177	5	GAM	0.063	−0.179	5	Reg-ARMA	0.093	−0.238
	5	GLMM	0.060	−0.197	6	GLMM	0.064	−0.192	6	GLM	0.093	−0.254
	6	GLM	0.061	−0.215	6	GLM	0.064	−0.192	7	GLMM	0.094	−0.264
	6	Reg-ARMA	0.061	−0.213	6	Reg-ARMA	0.064	−0.203	8	Linear reg	0.094	−0.269
	6	GAMM	0.061	−0.233	6	SVR-L	0.064	−0.190	8	GLS	0.094	−0.272
	6	Linear reg	0.061	−0.236	7	Linear reg	0.064	−0.228	8	GAMM	0.094	−0.285
	6	GLS	0.061	−0.238	7	GLS	0.064	−0.228	9	Reg-ARMA-GARCH	0.094	−0.292
	7	SVR-L	0.062	−0.249	7	GAMM	0.064	−0.223	10	GAM	0.097	−0.400
	8	Reg-ARMA-GARCH	0.062	−0.295	8	Reg-ARMA-GARCH	0.065	−0.249	10	SVR-L	0.097	−0.485
Step	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923
	2	SVR-RBF	0.044	0.613	2	SVR-RBF	0.044	0.644	2	SVR-RBF	0.046	0.597
	3	SSM-M	0.055	0.149	3	SSM-M	0.055	−0.126	3	SSM-M	0.056	0.174
	4	GLMM	0.056	−0.133	4	GLMM	0.056	−0.138	4	GLMM	0.057	−0.133
	5	GAM	0.056	−0.151	5	Reg-ARMA	0.057	−0.136	5	Reg-ARMA	0.058	−0.131
	5	Reg-ARMA	0.056	−0.154	5	GAM	0.057	−0.144	5	GAM	0.058	−0.144
	5	GLM	0.057	−0.146	5	GLM	0.057	−0.149	5	GLM	0.058	−0.149
	5	GAMM	0.057	−0.169	5	GAMM	0.057	−0.149	5	GAMM	0.058	−0.151
	5	Linear reg	0.057	−0.159	5	Linear reg	0.057	−0.154	5	Linear reg	0.058	−0.154
	5	GLS	0.057	−0.162	5	GLS	0.057	−0.156	5	GLS	0.058	−0.156
	6	RF	0.057	−0.182	6	Reg-ARMA-GARCH	0.058	−0.223	6	Reg-ARMA-GARCH	0.059	−0.244
	7	Reg-ARMA-GARCH	0.058	−0.246	6	RF	0.058	−0.223	7	SVR-L	0.060	−0.336
	8	SVR-L	0.059	−0.356	7	SVR-L	0.059	−0.269	8	RF	0.061	−0.326

Table C2
Ranked in-sample model fit results and associated average Root-Mean-Squared-Error (RMSE), and Kolmogorov-Smirnov D-Statistic (D-stat) for the 10% and 95% Linear and Nonlinear change scenarios and corresponding lagged change scenarios.

		Non-Lagged Change				Lagged Change			
Change Amplitude	Change Scenario	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat
10%	Linear	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923
		2	SVR-RBF	0.046	0.605	2	SVR-RBF	0.047	0.633
		3	SSM-M	0.057	0.128	3	SSM-M	0.058	−0.118
		4	GLMM	0.058	−0.131	4	GLMM	0.059	−0.138
		4	GAM	0.058	−0.138	4	GAM	0.059	−0.136
		5	Reg-ARMA	0.058	−0.138	4	Reg-ARMA	0.059	−0.141
		5	GLM	0.058	−0.133	4	GLM	0.059	−0.146
		5	Linear reg	0.058	−0.151	4	GAMM	0.060	−0.149
		5	GLS	0.058	−0.154	4	Linear reg	0.060	−0.154
		5	GAMM	0.058	−0.146	4	GLS	0.060	−0.156
		6	RF	0.060	−0.246	5	RF	0.061	−0.218
		6	Reg-ARMA-GARCH	0.060	−0.244	5	Reg-ARMA-GARCH	0.061	−0.210
		7	SVR-L	0.061	−0.300	6	SVR-L	0.062	−0.259
	Nonlinear	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923
		2	SVR-RBF	0.045	0.621	2	SVR-RBF	0.045	0.577
		3	SSM-M	0.056	0.144	3	SSM-M	0.056	0.128
		4	GLMM	0.057	−0.131	4	GLMM	0.057	−0.126
		4	GAM	0.057	−0.133	4	GAM	0.058	−0.136
		5	Reg-ARMA	0.057	−0.144	4	Reg-ARMA	0.058	−0.138
		5	GLM	0.057	−0.156	4	GLM	0.058	−0.133
		5	GAMM	0.058	−0.159	4	GAMM	0.058	−0.159
		5	Linear reg	0.058	−0.149	4	Linear reg	0.058	−0.154
		5	GLS	0.058	−0.151	4	GLS	0.058	−0.156
		6	RF	0.059	−0.187	5	RF	0.060	−0.215
		7	Reg-ARMA-GARCH	0.059	−0.231	5	Reg-ARMA-GARCH	0.060	−0.246
		8	SVR-L	0.060	−0.279	6	SVR-L	0.061	−0.303
95%	Linear	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923
		2	SVR-RBF	0.045	0.600	2	SVR-RBF	0.045	0.718
		3	SSM-M	0.055	0.159	3	SSM-M	0.057	0.262
		4	GLMM	0.058	−0.144	4	GAM	0.059	0.149
		4	Reg-ARMA	0.058	−0.151	5	GAMM	0.060	−0.146
		4	GAM	0.058	−0.146	6	Reg-ARMA	0.061	−0.159
		4	GLM	0.058	−0.162	7	RF	0.062	−0.203
		4	Linear reg	0.058	−0.156	8	GLMM	0.063	−0.233
		4	GLS	0.058	−0.159	8	Linear reg	0.063	−0.223
		4	GAMM	0.058	−0.164	8	GLS	0.063	−0.226
		5	Reg-ARMA-GARCH	0.059	−0.208	9	GLM	0.063	−0.262
		6	SVR-L	0.061	−0.259	10	SVR-L	0.065	−0.318
		7	RF	0.061	−0.310	11	Reg-ARMA-GARCH	0.066	−0.338
	Nonlinear	1	SSM-P	0.001	0.923	1	SSM-P	0.001	0.923
		2	SVR-RBF	0.043	0.738	2	SVR-RBF	0.045	0.682
		3	SSM-M	0.054	0.331	3	SSM-M	0.056	0.223
		4	GAM	0.057	0.213	4	GAM	0.059	−0.159
		5	GAMM	0.058	0.177	5	GAMM	0.060	−0.167
		6	Reg-ARMA	0.058	0.162	5	Reg-ARMA	0.060	−0.174
		7	RF	0.060	−0.203	5	RF	0.061	−0.154
		8	Linear reg	0.063	−0.279	6	GLMM	0.063	−0.244
		8	GLS	0.063	−0.282	6	Linear reg	0.063	−0.254
		8	SVR-L	0.063	−0.310	6	GLS	0.063	−0.256
		9	Reg-ARMA-GARCH	0.065	−0.336	6	GLM	0.063	−0.277
		10	GLMM	0.066	−0.359	7	Reg-ARMA-GARCH	0.066	−0.372
		11	GLM	0.067	−0.456	8	SVR-L	0.066	−0.382

Table C3
Ranked model forecast results and associated average Mean-Absolute-Scaled-Error (MASE), and Kolmogorov-Smirnov D-Statistic (D-stat) across 10%, 25%, and 95% amplitudes of change for all change scenarios.

	10% Change Amplitude				25% Change Amplitude				95% Change Amplitude			
Change Scenario	Rank	Model	MASE	D-stat	Rank	Model	MASE	D-stat	Rank	Model	MASE	D-stat
Abrupt	1	SVR-L	0.965	0.536	1	SVR-L	0.957	0.587	1	SVR-L	1.118	0.474
	2	Reg-ARMA-GARCH	1.022	0.451	2	Reg-ARMA-GARCH	1.026	0.462	2	GLM	1.142	0.474
	3	SSM-P	1.054	0.390	3	SSM-P	1.046	0.415	2	SVR-RBF	1.147	0.423
	4	SVR-RBF	1.157	0.228	4	GLMM	1.097	0.362	3	SSM-M	1.183	0.433
	4	GLMM	1.165	0.182	4	SVR-RBF	1.101	0.351	3	GLMM	1.186	0.446
	5	GLM	1.196	0.154	4	SSM-M	1.106	0.364	4	Reg-ARMA-GARCH	1.227	0.341
	6	SSM-M	1.228	0.138	4	GLM	1.114	0.326	5	RF	1.510	−0.321
	7	GAMM	1.288	−0.197	5	Linear reg	1.333	−0.390	6	Linear reg	1.552	−0.364
	7	Linear reg	1.315	−0.277	5	Reg-ARMA	1.333	−0.385	6	GLS	1.552	−0.362
	7	GLS	1.316	−0.279	5	GLS	1.333	−0.387	6	SSM-P	1.553	−0.310
	7	Reg-ARMA	1.319	−0.274	6	RF	1.420	−0.390	6	Reg-ARMA	1.553	−0.367
	8	GAM	1.488	−0.490	7	GAMM	1.538	−0.615	7	GAM	1.837	−0.556
	8	RF	1.524	−0.510	8	GAM	1.564	−0.608	8	GAMM	1.963	−0.685
Linear	1	SVR-L	0.951	0.546	1	SVR-L	0.970	0.551	1	SVR-L	0.953	0.541
	2	Reg-ARMA-GARCH	1.008	0.459	2	Reg-ARMA-GARCH	1.013	0.490	2	Reg-ARMA-GARCH	0.995	0.477
	3	SSM-P	1.069	0.344	3	SVR-RBF	1.035	0.464	3	GAMM	1.100	0.377
	3	SVR-RBF	1.069	0.336	4	SSM-P	1.060	0.400	4	GLMM	1.119	0.331
	4	GAMM	1.138	0.210	5	GAMM	1.166	0.197	4	GLM	1.122	0.336
	5	GLMM	1.182	0.133	6	SSM-M	1.170	0.203	5	GAM	1.247	0.192
	6	GLM	1.212	−0.172	6	GLMM	1.179	0.164	6	RF	1.312	−0.167
	7	SSM-M	1.249	−0.213	7	GLM	1.207	−0.185	7	Reg-ARMA	1.315	−0.215
	8	Linear reg	1.318	−0.290	8	Linear reg	1.318	−0.336	7	Linear reg	1.316	−0.218
	8	GLS	1.318	−0.285	8	GLS	1.318	−0.341	7	GLS	1.317	−0.215
	8	Reg-ARMA	1.320	−0.292	8	Reg-ARMA	1.319	−0.372	8	SSM-M	1.347	−0.149
	9	GAM	1.400	−0.354	8	GAM	1.329	−0.382	9	SVR-RBF	2.370	−0.774
	10	RF	1.642	−0.556	9	RF	1.507	−0.590	10	SSM-P	3.196	−0.890
Nonlinear	1	SVR-L	1.005	0.567	1	SVR-L	0.976	0.462	1	GLM	1.338	0.631
	2	Reg-ARMA-GARCH	1.064	0.451	2	SVR-RBF	1.038	0.413	2	RF	1.468	0.579
	3	SSM-P	1.098	0.446	2	Reg-ARMA-GARCH	1.050	0.397	3	GLMM	1.468	0.567
	4	SVR-RBF	1.128	0.364	2	SSM-P	1.053	0.392	4	SVR-L	1.521	0.559
	5	GLMM	1.229	0.154	3	SSM-M	1.165	0.274	5	SSM-M	1.873	0.218
	6	GAMM	1.229	0.169	3	GLMM	1.186	0.256	6	Reg-ARMA-GARCH	1.989	−0.223
	7	SSM-M	1.262	−0.154	4	GLM	1.218	0.221	7	Reg-ARMA	2.036	−0.282
	7	GLM	1.263	−0.190	5	Linear reg	1.402	−0.338	8	Linear reg	2.106	−0.354
	8	Reg-ARMA	1.360	−0.344	5	Reg-ARMA	1.403	−0.331	8	GLS	2.106	−0.356
	8	GLS	1.361	−0.367	5	GLS	1.403	−0.341	9	SVR-RBF	2.148	−0.254
	8	Linear reg	1.361	−0.369	6	GAMM	1.485	−0.333	10	GAM	2.851	−0.595
	9	GAM	1.519	−0.510	7	RF	1.590	−0.482	11	GAMM	3.443	−0.767
	9	RF	1.551	−0.562	8	GAM	1.707	−0.664	12	SSM-P	4.326	−0.833
Periodic	1	RF	0.934	0.692	1	RF	0.855	0.890	1	RF	0.660	0.923
	2	GAM	1.030	0.467	2	GAM	1.039	0.721	2	GAM	1.012	0.846
	3	SVR-L	1.044	0.474	3	SVR-L	1.162	0.472	3	SVR-RBF	1.358	0.654
	4	SVR-RBF	1.066	0.451	4	SVR-RBF	1.173	0.454	4	SVR-L	1.441	0.526
	5	GLMM	1.187	−0.200	5	GLMM	1.307	−0.187	5	GLMM	1.597	−0.254
	6	Reg-ARMA-GARCH	1.198	−0.205	6	GLM	1.349	−0.249	6	GAMM	1.647	−0.282
	7	GAMM	1.211	−0.213	6	SSM-M	1.352	−0.246	6	SSM-M	1.653	−0.326
	8	GLM	1.224	−0.282	6	GAMM	1.362	−0.269	6	GLM	1.654	−0.331
	8	SSM-P	1.229	−0.244	6	Reg-ARMA-GARCH	1.365	−0.285	7	SSM-P	1.663	−0.313
	8	Linear reg	1.236	−0.269	7	SSM-P	1.377	−0.282	7	Reg-ARMA-GARCH	1.664	−0.318
	8	GLS	1.236	−0.267	8	Linear reg	1.385	−0.305	7	Reg-ARMA	1.673	−0.344
	8	Reg-ARMA	1.238	−0.277	8	GLS	1.385	−0.305	7	GLS	1.673	−0.338
	8	SSM-M	1.241	−0.308	8	Reg-ARMA	1.385	−0.303	7	Linear reg	1.673	−0.341
Step	1	SVR-L	0.957	0.497	1	SVR-L	0.932	0.482	1	SVR-L	0.930	0.572
	2	Reg-ARMA-GARCH	1.012	0.410	2	Reg-ARMA-GARCH	0.968	0.397	2	Reg-ARMA-GARCH	0.979	0.462
	3	RF	1.112	0.128	3	GAMM	1.092	0.103	3	GAMM	1.098	0.144
	3	SVR-RBF	1.116	0.113	3	GLS	1.096	0.087	3	GAM	1.105	0.131
	3	GAMM	1.120	0.126	3	Linear reg	1.096	0.085	3	Linear reg	1.105	0.133
	3	GLS	1.128	−0.103	3	Reg-ARMA	1.098	−0.090	3	GLS	1.105	0.131
	3	Linear reg	1.128	−0.105	3	GAM	1.102	0.090	3	Reg-ARMA	1.107	0.128
	3	GAM	1.128	0.090	3	SSM-P	1.108	−0.085	4	SSM-P	1.117	−0.115
	3	Reg-ARMA	1.128	−0.100	4	RF	1.123	−0.144	5	GLMM	1.145	−0.136
	3	SSM-P	1.136	−0.133	5	SVR-RBF	1.142	−0.177	6	RF	1.171	−0.231
	4	GLMM	1.194	−0.221	6	GLMM	1.155	−0.208	6	GLM	1.175	−0.231
	5	GLM	1.225	−0.300	7	GLM	1.187	−0.313	7	SSM-M	1.264	−0.428
	6	SSM-M	1.303	−0.500	8	SSM-M	1.270	−0.477	8	SVR-RBF	1.272	−0.446

Table C4
Ranked model forecast results and associated average Root-Mean-Squared-Error (RMSE), and Kolmogorov-Smirnov D-Statistic (D-stat) across 10%, 25%, and 95% amplitudes of change for all change scenarios.

	10% Change Amplitude				25% Change Amplitude				95% Change Amplitude			
Change Scenario	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat
Abrupt	1	SSM-M	0.082	0.095	1	SSM-M	0.079	0.118	1	SSM-M	0.087	0.164
	1	GLM	0.082	0.100	1	SVR-RBF	0.079	0.087	1	Reg-ARMA-GARCH	0.088	0.144
	1	SVR-RBF	0.082	0.095	1	Linear reg	0.079	0.121	1	SVR-RBF	0.088	0.141
	1	Linear reg	0.082	0.072	1	GLS	0.079	0.123	1	Linear reg	0.088	0.136
	1	GLS	0.082	0.069	1	Reg-ARMA	0.079	0.118	1	GLS	0.088	0.133
	1	Reg-ARMA	0.082	0.067	1	GLM	0.080	0.090	1	Reg-ARMA	0.089	0.131
	1	GLMM	0.082	−0.056	1	GLMM	0.080	0.064	1	GLM	0.089	0.123
	1	GAMM	0.083	−0.069	1	Reg-ARMA-GARCH	0.080	0.062	1	SVR-L	0.089	0.095
	1	SSM-P	0.083	−0.082	2	RF	0.081	−0.090	1	RF	0.089	0.118
	1	Reg-ARMA-GARCH	0.084	−0.087	2	SVR-L	0.082	−0.072	1	GLMM	0.090	0.087
	1	GAM	0.084	−0.087	2	GAMM	0.082	−0.097	2	GAM	0.095	−0.231
	1	RF	0.084	−0.108	2	GAM	0.083	−0.123	3	GAMM	0.096	−0.321
	1	SVR-L	0.085	−0.108	3	SSM-P	0.085	−0.187	4	SSM-P	0.104	−0.413
Linear	1	GLM	0.081	0.103	1	GLM	0.082	0.105	1	GAM	0.081	0.182
	1	SSM-M	0.081	0.100	1	Linear reg	0.082	0.105	1	GLM	0.081	0.167
	1	SVR-RBF	0.081	0.121	1	GLS	0.082	0.103	1	GLMM	0.081	0.164
	1	GLMM	0.081	0.067	1	Reg-ARMA	0.082	0.108	1	Reg-ARMA	0.082	0.185
	1	Linear reg	0.081	0.100	1	GLMM	0.082	0.085	1	Linear reg	0.082	0.190
	1	GLS	0.081	0.097	1	SSM-M	0.083	0.108	1	GLS	0.082	0.187
	1	Reg-ARMA	0.081	0.103	1	GAMM	0.083	0.082	1	GAMM	0.082	0.154
	1	GAMM	0.082	−0.074	1	GAM	0.083	0.097	2	Reg-ARMA-GARCH	0.083	0.144
	1	SSM-P	0.082	0.062	2	Reg-ARMA-GARCH	0.085	−0.090	3	SVR-L	0.085	0.131
	1	Reg-ARMA-GARCH	0.083	−0.077	2	SVR-RBF	0.085	−0.092	4	RF	0.089	−0.138
	2	GAM	0.083	−0.092	2	SSM-P	0.086	−0.133	5	SSM-M	0.092	−0.164
	3	SVR-L	0.085	−0.146	2	SVR-L	0.087	−0.179	6	SVR-RBF	0.125	−0.523
	4	RF	0.087	−0.200	3	RF	0.087	−0.185	7	SSM-P	0.147	−0.823
Nonlinear	1	GLS	0.085	0.103	1	GLM	0.075	0.110	1	RF	0.086	0.433
	1	Linear reg	0.085	0.100	1	SSM-M	0.075	0.113	2	SVR-L	0.088	0.395
	1	Reg-ARMA	0.085	0.095	1	GLMM	0.075	0.103	3	SSM-M	0.092	0.249
	1	GLM	0.085	0.087	1	Reg-ARMA-GARCH	0.076	0.103	4	Reg-ARMA	0.095	0.290
	1	SSM-M	0.085	0.077	1	Linear reg	0.077	0.105	4	GLM	0.096	0.233
	1	GLMM	0.086	0.069	1	GLS	0.077	0.103	4	Reg-ARMA-GARCH	0.096	0.279
	1	RF	0.087	0.087	1	Reg-ARMA	0.077	0.100	5	Linear reg	0.098	0.249
	1	SVR-RBF	0.087	−0.067	2	SVR-RBF	0.077	0.105	5	GLS	0.098	0.246
	1	GAMM	0.087	−0.077	2	SVR-L	0.078	−0.105	6	GLMM	0.100	0.185
	1	SSM-P	0.087	−0.105	3	GAMM	0.079	−0.151	7	SVR-RBF	0.119	−0.364
	1	GAM	0.087	−0.105	4	SSM-P	0.080	−0.149	8	GAM	0.122	−0.367
	1	Reg-ARMA-GARCH	0.088	−0.115	5	GAM	0.082	−0.208	9	GAMM	0.145	−0.646
	2	SVR-L	0.090	−0.146	5	RF	0.083	−0.195	10	SSM-P	0.187	−0.838
Periodic	1	RF	0.080	0.221	1	RF	0.081	0.513	1	RF	0.087	0.854
	2	SVR-RBF	0.085	0.126	2	GAM	0.089	0.259	2	GAM	0.111	0.628
	2	SVR-L	0.086	0.097	3	SVR-RBF	0.091	0.228	3	SVR-RBF	0.128	0.369
	2	GAM	0.086	0.097	3	SVR-L	0.091	0.269	4	SVR-L	0.141	−0.195
	3	GLMM	0.089	−0.067	4	GLMM	0.096	−0.108	5	GLMM	0.142	−0.156
	3	GLM	0.090	−0.072	5	GAMM	0.097	−0.123	6	GAMM	0.145	−0.169
	3	GAMM	0.090	−0.072	5	GLM	0.097	−0.131	7	SSM-P	0.145	−0.179
	3	SSM-P	0.090	−0.059	5	SSM-P	0.097	−0.136	7	Reg-ARMA-GARCH	0.146	−0.182
	3	Reg-ARMA-GARCH	0.090	−0.082	5	SSM-M	0.097	−0.149	7	Reg-ARMA	0.146	−0.205
	3	Linear reg	0.090	−0.072	5	Reg-ARMA-GARCH	0.097	−0.149	7	GLS	0.146	−0.203
	3	GLS	0.090	−0.072	5	Linear reg	0.097	−0.144	7	Linear reg	0.146	−0.208
	3	Reg-ARMA	0.090	−0.069	5	GLS	0.097	−0.141	7	GLM	0.146	−0.269
	3	SSM-M	0.090	−0.126	5	Reg-ARMA	0.097	−0.144	7	SSM-M	0.147	−0.290
Step	1	RF	0.073	0.162	1	RF	0.076	0.177	1	SVR-RBF	0.076	0.105
	2	SVR-RBF	0.076	0.087	1	SVR-RBF	0.077	0.105	1	GLM	0.076	0.079
	3	GLM	0.078	0.072	2	GLM	0.078	0.067	1	SSM-M	0.076	0.072
	3	SSM-M	0.078	−0.062	2	SSM-M	0.078	0.062	1	SSM-P	0.076	0.064
	3	GLMM	0.078	0.062	2	SSM-P	0.078	0.062	1	Reg-ARMA	0.077	0.062
	3	SSM-P	0.078	−0.033	2	GLS	0.078	0.046	1	Linear reg	0.077	0.059
	3	Linear reg	0.078	−0.041	2	Linear reg	0.078	0.051	1	GLS	0.077	0.056
	3	GLS	0.078	−0.044	2	Reg-ARMA	0.078	0.051	1	GLMM	0.077	−0.059
	3	Reg-ARMA	0.078	−0.044	2	GLMM	0.079	−0.067	1	GAMM	0.077	−0.044
	3	GAMM	0.078	−0.069	2	GAMM	0.079	−0.064	1	GAM	0.077	−0.064
	3	GAM	0.078	−0.067	2	GAM	0.079	−0.105	1	RF	0.078	−0.118
	4	Reg-ARMA-GARCH	0.080	−0.092	3	Reg-ARMA-GARCH	0.080	−0.118	2	Reg-ARMA-GARCH	0.079	−0.100
	5	SVR-L	0.081	−0.149	4	SVR-L	0.081	−0.149	3	SVR-L	0.079	−0.128

Table C5
Ranked model forecast results and associated average Mean-Absolute-Scaled-Error (MASE), and Kolmogorov-Smirnov D-Statistic (D-stat) for the 10% and 95% Linear and Nonlinear change scenarios and corresponding lagged change scenarios.

		Non-Lagged Change				Lagged Change			
Change Amplitude	Change Scenario	Rank	Model	MASE	D-stat	Rank	Model	MASE	D-stat
10%	Linear	1	SVR-L	0.951	0.546	1	Reg-ARMA-GARCH	0.824	0.477
		2	Reg-ARMA-GARCH	1.008	0.459	1	SVR-L	0.826	0.462
		3	SSM-P	1.069	0.344	2	SSM-P	0.905	0.292
		3	SVR-RBF	1.069	0.336	2	SVR-RBF	0.906	0.336
		4	GAMM	1.138	0.210	3	GAMM	0.931	0.218
		5	GLMM	1.182	0.133	4	GLMM	0.936	0.187
		6	GLM	1.212	−0.172	5	GLM	0.964	0.133
		7	SSM-M	1.249	−0.213	6	GLS	1.010	−0.318
		8	Linear reg	1.318	−0.290	6	Linear reg	1.010	−0.313
		8	GLS	1.318	−0.285	6	Reg-ARMA	1.011	−0.315
		8	Reg-ARMA	1.320	−0.292	7	SSM-M	1.035	−0.313
		9	GAM	1.400	−0.354	8	GAM	1.139	−0.505
		10	RF	1.642	−0.556	9	RF	1.259	−0.636
	Nonlinear	1	SVR-L	1.005	0.567	1	Reg-ARMA-GARCH	0.840	0.423
		2	Reg-ARMA-GARCH	1.064	0.451	2	SVR-RBF	0.895	0.174
		3	SSM-P	1.098	0.446	2	SVR-L	0.896	0.246
		4	SVR-RBF	1.128	0.364	2	GLMM	0.900	0.190
		5	GLMM	1.229	0.154	2	SSM-P	0.905	0.182
		6	GAMM	1.229	0.169	3	GLM	0.925	0.131
		7	SSM-M	1.262	−0.154	4	GAMM	0.951	0.110
		7	GLM	1.263	−0.190	5	Linear reg	0.955	−0.174
		8	Reg-ARMA	1.360	−0.344	5	GLS	0.955	−0.169
		8	GLS	1.361	−0.367	5	Reg-ARMA	0.955	−0.167
		8	Linear reg	1.361	−0.369	6	SSM-M	0.988	−0.267
		9	GAM	1.519	−0.510	7	GAM	1.052	−0.341
		9	RF	1.551	−0.562	8	RF	1.162	−0.528
95%	Linear	1	SVR-L	0.953	0.541	1	GAMM	0.989	0.905
		2	Reg-ARMA-GARCH	0.995	0.477	2	GAM	1.446	0.708
		3	GAMM	1.100	0.377	3	RF	1.783	0.538
		4	GLMM	1.119	0.331	4	Reg-ARMA	1.860	0.477
		4	GLM	1.122	0.336	5	SSM-M	2.361	0.338
		5	GAM	1.247	0.192	6	GLS	2.466	0.305
		6	RF	1.312	−0.167	6	Linear reg	2.472	0.303
		7	Reg-ARMA	1.315	−0.215	7	SVR-L	2.979	−0.390
		7	Linear reg	1.316	−0.218	8	GLMM	3.347	−0.436
		7	GLS	1.317	−0.215	8	GLM	3.361	−0.446
		8	SSM-M	1.347	−0.149	9	SVR-RBF	3.896	−0.510
		9	SVR-RBF	2.370	−0.774	10	Reg-ARMA-GARCH	4.364	−0.623
		10	SSM-P	3.196	−0.890	11	SSM-P	4.535	−0.710
	Nonlinear	1	GLM	1.338	0.631	1	GAM	1.442	0.882
		2	RF	1.468	0.579	2	RF	2.153	0.690
		3	GLMM	1.468	0.567	3	GAMM	2.178	0.646
		4	SVR-L	1.521	0.559	4	Reg-ARMA	2.790	0.600
		5	SSM-M	1.873	0.218	5	SSM-M	3.367	0.510
		6	Reg-ARMA-GARCH	1.989	−0.223	6	GLS	4.436	−0.297
		7	Reg-ARMA	2.036	−0.282	6	Linear reg	4.450	−0.303
		8	Linear reg	2.106	−0.354	7	GLMM	4.992	−0.351
		8	GLS	2.106	−0.356	7	GLM	5.020	−0.362
		9	SVR-RBF	2.148	−0.254	8	SVR-RBF	5.405	−0.395
		10	GAM	2.851	−0.595	9	SVR-L	5.558	−0.446
		11	GAMM	3.443	−0.767	10	SSM-P	5.832	−0.490
		12	SSM-P	4.326	−0.833	11	Reg-ARMA-GARCH	6.134	−0.528

Table C6
Ranked model forecast results and associated average Root-Mean-Squared-Error (RMSE), and Kolmogorov-Smirnov D-Statistic (D-stat) for the 10% and 95% Linear and Nonlinear change scenarios and corresponding lagged change scenarios.

		Non-Lagged Change				Lagged Change			
Change Amplitude	Change Scenario	Rank	Model	RMSE	D-stat	Rank	Model	RMSE	D-stat
10%	Linear	1	GLM	0.081	0.103	1	SVR-RBF	0.067	0.162
		1	SSM-M	0.081	0.100	1	SSM-M	0.067	0.097
		1	SVR-RBF	0.081	0.121	1	GLM	0.067	0.105
		1	GLMM	0.081	0.067	1	Linear reg	0.068	0.072
		1	Linear reg	0.081	0.100	1	GLS	0.068	0.069
		1	GLS	0.081	0.097	1	Reg-ARMA	0.068	0.072
		1	Reg-ARMA	0.081	0.103	1	GLMM	0.068	0.077
		1	GAMM	0.082	−0.074	2	SSM-P	0.069	−0.077
		1	SSM-P	0.082	0.062	2	GAMM	0.069	−0.074
		1	Reg-ARMA-GARCH	0.083	−0.077	2	GAM	0.069	−0.108
		2	GAM	0.083	−0.092	2	RF	0.069	−0.128
		3	SVR-L	0.085	−0.146	2	Reg-ARMA-GARCH	0.069	−0.123
		4	RF	0.087	−0.200	3	SVR-L	0.072	−0.187
	Nonlinear	1	GLS	0.085	0.103	1	SSM-M	0.067	0.110
		1	Linear reg	0.085	0.100	1	RF	0.068	0.154
		1	Reg-ARMA	0.085	0.095	1	SVR-RBF	0.068	0.092
		1	GLM	0.085	0.087	1	GLS	0.068	0.085
		1	SSM-M	0.085	0.077	1	Linear reg	0.068	0.082
		1	GLMM	0.086	0.069	1	Reg-ARMA	0.068	0.082
		1	RF	0.087	0.087	1	GLM	0.068	0.074
		1	SVR-RBF	0.087	−0.067	1	GAM	0.069	0.062
		1	GAMM	0.087	−0.077	1	GLMM	0.069	−0.062
		1	SSM-P	0.087	−0.105	1	SSM-P	0.069	−0.072
		1	GAM	0.087	−0.105	1	GAMM	0.069	−0.082
		1	Reg-ARMA-GARCH	0.088	−0.115	2	Reg-ARMA-GARCH	0.071	−0.131
		2	SVR-L	0.090	−0.146	3	SVR-L	0.073	−0.236
95%	Linear	1	GAM	0.081	0.182	1	GAMM	0.067	0.756
		1	GLM	0.081	0.167	2	GAM	0.072	0.767
		1	GLMM	0.081	0.164	3	RF	0.096	0.464
		1	Reg-ARMA	0.082	0.185	3	Reg-ARMA	0.098	0.408
		1	Linear reg	0.082	0.190	4	SSM-M	0.116	0.287
		1	GLS	0.082	0.187	4	GLS	0.118	0.287
		1	GAMM	0.082	0.154	4	Linear reg	0.118	0.285
		2	Reg-ARMA-GARCH	0.083	0.144	5	SVR-L	0.135	−0.333
		3	SVR-L	0.085	0.131	6	GLMM	0.148	−0.354
		4	RF	0.089	−0.138	6	GLM	0.149	−0.387
		5	SSM-M	0.092	−0.164	7	SVR-RBF	0.171	−0.523
		6	SVR-RBF	0.125	−0.523	8	Reg-ARMA-GARCH	0.185	−0.574
		7	SSM-P	0.147	−0.823	9	SSM-P	0.191	−0.687
	Nonlinear	1	RF	0.086	0.433	1	GAM	0.086	0.787
		2	SVR-L	0.088	0.395	2	RF	0.108	0.638
		3	SSM-M	0.092	0.249	2	GAMM	0.111	0.564
		4	Reg-ARMA	0.095	0.290	3	Reg-ARMA	0.127	0.567
		4	GLM	0.096	0.233	4	SSM-M	0.149	0.423
		4	Reg-ARMA-GARCH	0.096	0.279	5	GLS	0.186	−0.226
		5	Linear reg	0.098	0.249	5	Linear reg	0.186	−0.228
		5	GLS	0.098	0.246	6	GLMM	0.207	−0.310
		6	GLMM	0.100	0.185	6	GLM	0.208	−0.315
		7	SVR-RBF	0.119	−0.364	7	SVR-RBF	0.222	−0.397
		8	GAM	0.122	−0.367	8	SVR-L	0.228	−0.392
		9	GAMM	0.145	−0.646	9	SSM-P	0.238	−0.433
		10	SSM-P	0.187	−0.838	10	Reg-ARMA-GARCH	0.249	−0.497

References

Achen, C., 2000. Why lagged dependent variables can suppress the explanatory power of other independent variables. In: Paper Presented at the Annual Meeting of the Society for Political Methodology. UCLA, California, USA.

Anderson, M.J., Thompson, A.A., 2004. Multivariate control charts for ecological and environmental monitoring. Ecol. Appl. 14, 1921–1935.

Barnabé, G., Barnabé-Quet, R., 2000. Ecology and Management of Coastal Waters: The Aquatic Environment. Springer, London, UK.

Barry, S., Elith, J., 2006. Error and uncertainty in habitat models. J. Appl. Ecol. 43, 413–423.

Bell, D.M., Schlaepfer, D.R., 2016. On the dangers of model complexity without ecological justification in species distribution modeling. Ecol. Modell. 330, 50–59.

Ben-Hur, A., Weston, J., 2010. A user's guide to support vector machines. Methods Mol. Biol. 609, 223–239.

Benedetti-Cecchi, L., 2001. Beyond BACI: optimization of environmental sampling designs through monitoring and simulation. Ecol. Appl. 11, 783–799.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Model. Softw. 40, 1–20.

Bergström, L., Sundqvist, F., Bergström, U., 2013. Effects of an offshore wind farm on temporal and spatial patterns in the demersal fish community. Mar. Ecol. Prog. Ser. 485, 199–210.

Biglan, A., Ary, D., Wagenaar, A.C., 2000. The value of interrupted time-series experiments for community intervention research. Prev. Sci. 1, 31–49.

Boehlert, G.W., Gill, A.B., 2010. Environmental and ecological effects of ocean renewable energy development: a current synthesis. Oceanography 23, 68–81.

Boehlert, G.W., Braby, C., Bull, A.S., Helix, M.E., Henkel, S., Klarin, P., Schroeder, D.,

2013. Proceedings of the Oregon Marine Renewable Energy Environmental Sciences Conference. Technical Report, U.S. Bureau of Ocean Energy Management, Corvallis, OR, USA.
Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B (Methodol.) 26, 211–252.
Box, G.E.P., Tiao, G.C., 1975. Intervention analysis with applications to economic and environmental problems. J. Am. Stat. Assoc. 70, 70–79.
Bureau of Ocean Energy Management, 2016. Guidelines for Information Requirements for a Renewable Energy Construction and Operations Plan (COP). U.S. Department of the Interior, Bureau of Ocean Energy Management, Office of Renewable Energy Programs, Washington, D.C., USA.
Burnham, K., Anderson, D., 2002. Model Selection and Multi-Model Inference a Practical Information-theoretic Approach, 2nd ed. Springer, New York, NY, USA.
Cappé, O., Moulines, E., Rydén, T., 2005. Inference in Hidden Markov Models. Springer, New York, NY, USA.
Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250. http://dx.doi.org/10.5194/gmd-7-1247-2014.
Chandler, R., Scott, E.M., 2011. Statistical Methods for Trend Detection and Analysis in the Environmental Sciences. Wiley, West Sussex, U.K.
Chapman, M.G., Underwood, A.J., Skilleter, G.A., 1995. Variability at different spatial scales between a subtidal assemblage exposed to the discharge of sewage and two control assemblages. J. Exp. Mar. Biol. Ecol. 189, 103–122.
Cichosz, P., 2015. Data Mining Algorithms: Explained Using R. John Wiley & Sons Inc., West Sussex; Malden MA.
Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W.H., Wall, D.H., Wear, D., 2001. Ecological forecasts: an emerging imperative. Science 293, 657–660.
Copping, A., Hanna, L., Hutchinson, I., 2014. Best Practices for Monitoring Environmental Effects of Marine Energy Devices. Aquatera Limited, Orkney, UK, and Pacific Northwest National Laboratory, Richland, WA, USA.
Copping, A., Sather, N., Hanna, L., Whiting, J., Zydlewski, G., Staines, G., Gill, A., Hutchison, I., O'Hagan, A., Simas, T., Bald, J., Sparling, C., Wood, J., Masden, E., 2016. Annex IV 2016 State of the Science Report: Environmental Effects of Marine Renewable Energy Development Around the World. Pacific Northwest National Laboratory, Richland, Washington, USA.
Crone, S.F., Guajardo, J., Weber, R., 2006. A study on the ability of support vector regression and neural networks to forecast basic time series patterns. In: Bramer, M. (Ed.), Artificial Intelligence in Theory and Practice. Springer, New York, NY, USA, pp. 149–158.
Cryer, Jonathan D., Chan, Kung-sik., 2008. Time Series Analysis: With Applications in R, 2nd edition. Springer, New York, NY, USA.
Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.
Doornik, J., Ooms, M., 2005. Outlier Detection in GARCH Models. Economics Group, Nuffield College, University of Oxford, Oxford, UK.
Doornik, J.A., Ooms, M., 2008. Multimodality in GARCH regression models. Int. J. Forecast. 24, 432–448.
Dornelas, M., Magurran, A.E., Buckland, S.T., Chao, A., Chazdon, R.L., Colwell, R.K., Curtis, T., Gaston, K.J., Gotelli, N.J., Kosnik, M.A., McGill, B., McCune, J.L., Morlon, H., Mumby, P.J., Øvreås, L., Studeny, A., Vellend, M., 2013. Quantifying temporal change in biodiversity: challenges and opportunities. Proc. R. Soc. B (Biol. Sci.) 280, 20121931.
Dubbs, L., Keeler, A.G., O'Meara, T., 2013. Permitting, risk and marine hydrokinetic energy development. Electr. J. 26, 64–74.
Eitrich, T., Lang, B., 2006. Efficient optimization of support vector machine learning parameters for unbalanced datasets. J. Comput. Appl. Math. 196, 425–436.
Embling, C.B., Sharples, J., Armstrong, E., Palmer, M.R., Scott, B.E., 2013. Fish behaviour in response to tidal variability and internal waves over a shelf sea bank. Prog. Oceanogr. 117, 106–117.
Fairweather, P., 1991. Statistical power and design requirements for environmental monitoring. Mar. Freshwater Res. 42, 555–567.
Federal Energy Regulatory Commission, 2008. Licensing Hydrokinetic Pilot Projects. Federal Energy Regulatory Commission. Office of Energy Projects, Washington DC.
Ferretti, M., 1997. Forest health assessment and monitoring – issues for consideration. Environ. Monit. Assess. 48, 45–72.
Field, S.A., Tyre, A.J., Jonzén, N., Rhodes, J.R., Possingham, H.P., 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. Ecol. Lett. 7, 669–675.
Field, S.A., Tyre, A.J., Possingham, H.P., 2005. Optimizing allocation of monitoring effort under economic and observational constraints. J. Wildl. Manage. 69, 473–482.
Fodrie, F.J., Able, K.W., Galvez, F., Heck, K.L., Jensen, O.P., López-Duarte, P.C., Martin, C.W., Turner, R.E., Whitehead, A., 2014. Integrating organismal and population responses of estuarine fishes in Macondo spill research. Bioscience 64, 778–788.
Froján, C.B., Cooper, K.M., Bolam, S.G., 2016. Towards an integrated approach to marine benthic monitoring. Mar. Pollut. Bull. 104, 20–28.
Gitzen, R.A., Millspaugh, J.J., Cooper, A.B., Licht, D.S., 2012. Design and Analysis of Long-term Ecological Monitoring Studies. Cambridge University Press, Cambridge, UK.
Granger, C.W.J., Newbold, P., 1974. Spurious regressions in econometrics. J. Econometr. 2, 111–120.
Grilli, A.R., Shumchenia, E.J., 2015. Toward wind farm monitoring optimization: assessment of ecological zones from marine landscapes using machine learning algorithms. Hydrobiologia 756, 117–137.
Hammar, L., Andersson, S., Eggertsen, L., Haglund, J., Gullström, M., Ehnberg, J., Molander, S., 2013. Hydrokinetic turbine effects on fish swimming behaviour. PLoS One 8, e84141.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second edition. Springer, New York, NY, USA.
Hewitt, J.E., Thrush, S.E., Cummings, V.J., 2001. Assessing environmental impacts: effects of spatial and temporal variability at likely impact scales. Ecol. Appl. 11, 1502–1516.
Holmes, E.E., Ward, E.J., Wills, K., 2012. Marss: multivariate autoregressive state-space models for analyzing time-series data. R Journal 4, 11–19.
Holmes, E.E., Ward, E.J., Scheuerell, M.D., 2014. Analysis of Multivariate Time-series Using the MARSS Package, Version 3.9 edition. NOAA Fisheries, Northwest Fisheries Science Center, Seattle, WA, USA. https://cran.r-project.org/web/packages/MARSS.
Horne, J.K., Hytnen, R., Maxwell, S., Ham, K., Maxwell, A., Condiotty, J., 2016. Fully utilizing the acoustic record for biological monitoring and ecological applications. J. Acoust. Soc. Am. 139 (2173-2173).
Hyndman, R., Athanasopoulos, G., 2014. Forecasting: Principles and Practice. OTexts.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22, 679–688.
Inger, R., Attrill, M.J., Bearhop, S., Broderick, A.C., James Grecian, W., Hodgson, D.J., Mills, C., Sheehan, E., Votier, S.C., Witt, M.J., Godley, B.J., 2009. Marine renewable energy: potential benefits to biodiversity? An urgent call for research. J. Appl. Ecol. 46, 1145–1153.
Ingersoll, T.E., Sewall, B.J., Amelon, S.K., 2013. Improved analysis of long-term monitoring data demonstrates marked regional declines of bat populations in the eastern United States. PLoS One 8, e65907.
Jackson, C., 2011. Multi-state models for panel data: the msm package for R. J. Stat. Softw. 38, 1–29.
Jacques, D., 2014. Describing and Comparing Variability of Fish and Macrozooplankton Density at Marine Hydrokinetic Energy Sites. University of Washington, Seattle, Washington, USA.
Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., Giovis, C., 2004. A review and evaluation of intraurban air pollution exposure models. J. Exp. Sci. Environ. Epidemiol. 15, 185–204.
Johnson, J.B., Omland, K.S., 2004. Model selection in ecology and evolution. Trends Ecol. Evol. 19, 101–108. http://dx.doi.org/10.1016/j.tree.2003.10.013.
Kacprzyk, J., Pedrycz, W., 2015. Springer Handbook of Computational Intelligence. Springer, Berlin, Germany.
Kane, M.J., Price, N., Scotch, M., Rabinowitz, P., 2014. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinf. 15, 276.
Klure, J., Hampton, T., McMurray, G., Boehlert, G., Henkel, S., Copping, A., Kramer, S., Chwaszczewski, R., Fresh, K., 2012. West Coast Environmental Protocols Framework: Baseline and Monitoring Studies. Pacific Energy Ventures LLC, Portland, Oregon, USA.
Kramer, S., Hamilton, C., Spencer, G., Ogston, H., 2015. Evaluating the Potential for Marine and Hydrokinetic Devices to Act as Artificial Reefs or Fish Aggregating Devices, Based on Analysis of Surrogates in Tropical, Subtropical, and Temperate U.S. West Coast and Hawaiian Coastal Waters. H.T. Harvey & Associates, Los Gatos, California, USA.
Lagarde, M., 2011. How to do (or not to do) … Assessing the impact of a policy change with routine longitudinal data. Health Policy Plan. 27, 76–83.
Legg, C.J., Nagy, L., 2006. Why most conservation monitoring is, but need not be, a waste of time. J. Environ. Manage. 78, 194–199.
Levine, C.R., Yanai, R.D., Lampman, G.G., Burns, D.A., Driscoll, C.T., Lawrence, G.B., Lynch, J.A., Schoch, N., 2014. Evaluating the efficiency of environmental monitoring programs. Ecol. Indic. 39, 94–101.
Lewis-Beck, Michael S., Bryman, A., Liao, T.F., 2004. The Sage Encyclopedia of Social Science Research Methods. Sage, Thousand Oaks, California, USA.
Lindenmayer, D.B., Gibbons, P., Bourke, M., Burgman, M., Dickman, C.R., Ferrier, S., Fitzsimons, J., Freudenberger, D., Garnett, S.T., Groves, C., Hobbs, R.J., Kingsford, R.T., Krebs, C., Legge, S., Lowe, A.J., Mclean, R., Montambault, J., Possingham, H., Radford, J., Robinson, D., Smallbone, L., Thomas, D., Varcoe, T., Vardon, M., Wardle, G., Woinarski, J., Zerger, A., 2012. Improving biodiversity monitoring. Austral Ecol. 37, 285–294.
Linder, H.L., Horne, J.K., Ward, E.J., 2017. Evaluating models that characterize baseline conditions: ecological monitoring at marine renewable energy sites. Ecol. Indic. 83, 178–191.
Lindley, S.T., 2003. Estimation of population growth and extinction parameters from noisy data. Ecol. Appl. 13, 806–813.
Liu, H., Erdem, E., Shi, J., 2011. Comprehensive evaluation of ARMA–GARCH(-M) approaches for modeling the mean and volatility of wind speed. Appl. Energy 88, 724–732.
Lorena, A.C., Jacintho, L.F.O., Siqueira, M.F., Giovanni, R.D., Lohmann, L.G., de Carvalho, A.C.P.L.F., Yamamoto, M., 2011. Comparing machine learning classifiers in potential distribution modelling. Expert Syst. Appl. 38, 5268–5275.
Lovett, G.M., Burns, D.A., Driscoll, C.T., Jenkins, J.C., Mitchell, M.J., Rustad, L., Shanley, J.B., Likens, G.E., Haeuber, R., 2007. Who needs environmental monitoring? Front. Ecol. Environ. 5, 253–260.
Mackenzie, M.L., Scott-Hayward, L.A., Oedekoven, C.S., Skov, H., Humphreys, E., Rexstad, E., 2013. SStatistical Modelling of Seabird and Cetacean Data: Guidance Document. Guidance Document. University of St. Andrews, St. Andrews, UK.
Maclennan, D.N., Fernandes, P.G., Dalen, J., 2002. A consistent approach to definitions and symbols in fisheries acoustics. ICES J. Mar. Sci. 59, 365–369.
Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D.A., Scott, E.M., Smith, R.I., Somerfield, P.J., Watt, A.D., 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends

Ecol. Evol. 25, 574–582.
Matsumoto, W.M., Kazama, T.K., Aasted, D.C., 1981. Anchored fish aggregating devices in Hawaiian waters. Mar. Fish. Rev. 43, 1–13.
McCann, J., 2012. Developing Environmental Protocols and Modeling Tools to Support Ocean Renewable Energy and Stewardship. U.S. Department of the Interior, Bureau of Ocean Energy Management, Office of Renewable Energy Programs, Herndon, Virginia, USA.
Michel, L., Makowski, D., 2013. Comparison of statistical models for analyzing wheat yield time series. PLoS One 8, e78615.
Morrison, L.W., 2007. Assessing the reliability of ecological monitoring data: power analysis and alternative approaches. Nat. Areas J. 27, 83–91.
Munkittrick, K.R., Arens, C.J., Lowell, R.B., Kaminski, G.P., 2009. A review of potential methods of determining critical effect size for designing environmental monitoring programs. Environ. Toxicol. Chem./SETAC 28, 1361–1371.
Nuno, A., Milner-Gulland, E.J., Bunnefeld, N., 2015. Detecting abundance trends under uncertainty: the influence of budget, observation error and environmental change. Anim. Conserv. 18, 331–340.
ORPC Maine LLC, 2014. Cobscook Bay Tidal Energy Project: 2013 Environmental Monitoring Report. Ocean Renewable Power Company (ORPC), Portland, Maine, USA.
Olden, J.D., Jackson, D.A., 2002. A comparison of statistical approaches for modelling fish species distributions. Freshwater Biol. 47, 1976–1995.
Perretti, C.T., Sugihara, G., Munch, S.B., 2013. Nonparametric forecasting outperforms parametric methods for a simulated multispecies system. Ecology 94, 794–800.
Peterman, R.M., 1990. Statistical power analysis can improve fisheries research and management. Can. J. Fish. Aquat. Sci. 47, 2–15.
Petersen, I.K., MacKenzie, M.L., Rexstad, E., Wisz, M.S., Fox, A.D., 2011. Comparing Pre- and PostConstruction Distributions of Long-tailed Ducks Clangula Hyemalis in and Around the Nysted Offshore Wind Farm, Denmark: A Quasi-Designed Experiment Accounting for Imperfect Detection, Local Surface Features and Autocorrelation. University of St. Andrews, St. Andrews, UK.
Polagye, B., Van Cleve, B., Copping, A., Kirkendall, K., 2011. Environmental Effects of Tidal Energy Development: Proceedings of a Scientific Workshop. U.S. Department of Commerce, NOAA Technical Memorandum NMFS F/SPO-116.
Portman, M.E., 2010. Marine renewable energy policy. Oceanography 23, 98.
Public Utility District No. 1 of Snohomish County, 2012. Admiralty Inlet Tidal Project Final Monitoring and Mitigation Plans. Public Utility District No. 1 of Snohomish County, Snohomish, Washington, USA.
R Core Development Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rush, B., Joslin, J., Stewart, A., Polagye, B., 2014. Development of an adaptable monitoring package for marine renewable energy projects part I: conceptual design and operation. In: Proceedings: 2nd Marine Energy Technology Symposium. Seattle, Washington, USA.
Scheiner, S.M., Gurevitch, J., 2001. Design and Analysis of Ecological Experiments, 2nd ed. Oxford University Press, Oxford, UK.
Schmitt, R.J., Osenberg, C.W., 1996. Detecting Ecological Impacts: Concepts and Applications in Coastal Habitats. Academic Press, San Diego, California, USA.
Shields, M.A., Payne, A.I.L., 2014. Marine Renewable Energy Technology and Environmental Interactions. Springer, New York, NY, USA.
Shmueli, G., 2010. To explain or to predict? Stat. Sci. 25, 289–310.
Shumchenia, E.J., Smith, S.L., McCann, J., Carnevale, M., Fugate, G., Kenney, R.D., King, J.W., Paton, P., Schwartz, M., Spaulding, M., Winiarski, K.J., Shumchenia, E.J., Smith, S.L., McCann, J., Carnevale, M., Fugate, G., Kenney, R.D., King, J.W., Paton, P., Schwartz, M., Spaulding, M., Winiarski, K.J., 2012. An adaptive framework for selecting environmental monitoring protocols to support ocean renewable energy development, an adaptive framework for selecting environmental monitoring protocols to support ocean renewable energy development. Sci. World J. 2012, e450685.
Simmonds, E.J., MacLennan, D.N., 2005. Fisheries Acoustics: Theory and Practice, 2nd. Blackwell Science, Oxford, UK.
Stenberg, C., Støttrup, J., van Deurs, M., Berg, C., Dinesen, G., Mosegaard, H., Grome, T., Leonhard, S., 2015. Long-term effects of an offshore wind farm in the North Sea on fish communities. Mar. Ecol. Prog. Ser. 528, 257–265.
Stewart-Oaten, A., Bence, J.R., 2001. Temporal and spatial variation in environmental impact assessment. Ecol. Monogr. 71, 305–339.
Taddy, M.A., Gramacy, R.B., Polson, N.G., 2011. Dynamic trees for learning and design. J. Am. Stat. Assoc. 106, 109–123.
Taylor, J.W., McSharry, P.E., Buizza, R., 2009. Wind power density forecasting using ensemble predictions and time series models. IEEE Trans. Energy Convers. 24, 775–782.
Thissen, U., van Brakel, R., de Weijer, A.P., Melssen, W.J., Buydens, L.M.C., 2003. Using support vector machines for time series prediction. Chemom. Intell. Lab. Syst. 69, 35–49.
Thomas, L., Martin, K., 1996. The importance of analysis method for breeding bird survey population trend estimates. Conserv. Biol. 10, 479–490.
Thomas, L., 1996. Monitoring long-term population change: why are there so many analysis methods? Ecology 77, 49–58.
Tollit, D., Redden, A., 2013. Passive Acoustic Monitoring of Cetacean Activity Patterns and Movements in Minas Passage: Pre-Turbine Baseline Conditions. Acadia University, Wolfville, NS, Canada, and SMRU Consulting, Vancouver, BC, Canada.
Treweek, J., 2009. Ecological Impact Assessment. Wiley Hoboken, NJ, USA.
US Department of Energy, 2009. Report to Congress on the Potential Environmental Effects of Marine and Hydrokinetic Energy Technologies. US Department of Energy, Office of Energy Efficiency and Renewable Energy, Washington, D.C., USA.
Underwood, A.J., 1992. Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. J. Exp. Mar. Biol. Ecol. 161, 145–178.
Underwood, A.J., 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. Ecol. Appl. 4, 4–15.
Urmy, S.S., Horne, J.K., Barbee, D.H., 2012. Measuring the vertical distributional variability of pelagic fauna in Monterey Bay. ICES J. Mar. Sci. 69, 184–196.
Vanermen, N., Onkelinx, T., Verschelde, P., Courtens, W., de Walle, M.V., Verstraete, H., Stienen, E.W.M., 2015. Assessing seabird displacement at offshore wind farms: power ranges of a monitoring and data handling protocol. Hydrobiologia 756, 155–167.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Fourth edition. Springer New York, NY, USA.
Viterbi, A., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans. Inf. Theory 13, 260–269.
Vos, P., Meelis, E., Ter Keurs, W.J., 2000. A framework for the design of ecological monitoring programs as a tool for environmental and nature management. Environ. Monit. Assess. 61, 317–344.
Waggitt, J.J., Bell, P.S., Scott, B.E., 2014. An evaluation of the use of shore-based surveys for estimating spatial overlap between deep-diving seabirds and tidal stream turbines. Int. J. Mar. Energy 8, 36–49.
Ward, E.J., Chirakkal, H., González-Suárez, M., Aurioles-Gamboa, D., Holmes, E.E., Gerber, L., 2010. Inferring spatial structure from time-series data: using multivariate state-space models to detect metapopulation structure of California sea lions in the Gulf of California, Mexico. J. Appl. Ecol. 47, 47–56.
Ward, E.J., Holmes, E.E., Thorson, J.T., Collen, B., 2014. Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. Oikos 123, 652–661.
Warwick-Evans, V.C., Atkinson, P.W., Robinson, L.A., Green, J.A., 2016. Predictive modelling to identify near-shore, fine-scale seabird distributions during the breeding season. PLoS One 11.
White, J.W., Rassweiler, A., Samhouri, J.F., Stier, A.C., White, C., 2014. Ecologists should not use statistical significance tests to interpret simulation model results. Oikos 123, 385–388. http://dx.doi.org/10.1111/j.1600-0706.2013.01073.x.
Wood, S., 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC, Boca Raton, Florida USA.
Wood, S., 2015. Mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation. R package version 1 (8-3). https://cran.r-project.org/web/packages/mgcv.
Zucchini, W., MacDonald, I.L., 2009. Hidden Markov Models for Time Series: An Introduction Using R. Chapman & Hall/CRC, Boca Raton, Florida USA.