https://doi.org/10.1016/j.rse.2019.04.034

Detect change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm

Kaiguang Zhao[1,2*], Michael A. Wulder[3], Tongxi Hu[2], Ryan Bright[4], Qiusheng Wu[5], Haiming Qin[6], Yang Li[2], Elizabeth Toman[2], Bani Mallick[6], Xuesong Zhang[7], and Molly Brown[8]

1. Ohio Agricultural Research and Development Center, School of Environment and Natural Resources, The Ohio State University, Wooster, OH 44691, USA

2. School of Environment and Natural Resources, Environmental Science Graduate Program, The Ohio State University, Columbus, OH 43210, USA

3. Canadian Forest Service, Natural Resources Canada, Victoria, British Columbia V8Z 1M5, Canada

4. Norwegian Institute of Bioeconomy Research (NIBIO), 1431 Ås, Norway

5. Department of Geography, Binghamton University, State University of New York, Binghamton, NY 13902, USA

6. Department of Statistics, Texas A&M University, College Station, USA

7. Joint Global Change Research Institute, Pacific Northwest National Laboratory, University of Maryland, College Park, MD, 20740, USA

8. Department of Geographical Sciences, University of Maryland, MD 20771, College Park, USA

* To whom all correspondence should be addressed (zhao.1423@osu.edu).

Abstract: Satellite time-series data are bolstering global change research, but their use to elucidate land surface or vegetation dynamics is sensitive to algorithmic choices. Different algorithms often give inconsistent or sometimes conflicting interpretations of the same data. This lack of consensus has adverse implications and can be mitigated via ensemble modeling, an algorithmic paradigm that combines many competing models rather than choosing only a single "best" model. Here we report one such time-series decomposition algorithm for deriving nonlinear ecosystem dynamics across multiple timescales—A Bayesian Estimator of Abrupt change, Seasonal change, and Trend (BEAST). As an ensemble algorithm, BEAST quantifies the relative usefulness of individual decomposition models, leveraging all the models via Bayesian model averaging. We tested it upon simulated, Landsat, and MODIS data. BEAST reliably detected changepoints, seasonality, and trends in the data; it derived realistic nonlinear trend signals and credible uncertainty measures (e.g., occurrence probability of changepoints over time)—some information difficult to derive by conventional single-best-model algorithms but critical for interpretation of ecosystem dynamics and detection of low-magnitude disturbances. The combination of many models enabled BEAST to alleviate model misspecification, address algorithmic uncertainty, and reduce overfitting. BEAST is generically applicable to time-series data of all kinds, serving to improve robustness in uncovering true time-series dynamics. It offers a new analytical option for changepoint detection and nonlinear trend analysis and will help exploit environmental time-series data for probing patterns and drivers of ecosystem dynamics.

Keywords: *Changepoint; Bayesian changepoint detection; Disturbance ecology; Breakpoint; Trend analysis; Time series decomposition; Bayesian model averaging; Disturbances; Nonlinear dynamics; Regime shift*

1. Introduction

Ecosystems are changing constantly, driven by natural forcings and human activities in complex ways. Disentangling the complexity to build predictive biospheric sciences is a defining theme of global change research *(Franklin et al. 2016)*—a goal hard to attain without reliable capabilities of monitoring lands over time *(Pettorelli et al. 2014; Su et al. 2016; Zhao and Jackson 2014)*. To date, such spatiotemporal data come primarily from satellites *(Hu et al. 2017; Jetz et al. 2016)*. Satellite time-series data, such as decades of Landsat, MODIS, or AVHRR imagery, have proven particularly valuable for elucidating patterns and drivers of land and ecosystem dynamics *(Hawbaker et al. 2017; Li et al. 2018; Zhu and Woodcock 2014)*.

Despite existing successes in satellite time-series analyses, challenges remain. A notable issue pertains to the diverging findings from the use of satellite data in addressing the same problem. For example, there is controversy regarding how the Amazon forests respond to basin-wide droughts; some satellite analyses suggested a green-up but others not *(Huete et al. 2006; Samanta et al. 2010)*. Inconsistencies like this are attributed partly to different algorithms and perspectives taken for data processing and analysis *(Liu et al. 2018; Shen 2011; Tewkesbury et al. 2015)*. A preponderance of satellite time-series analyses take a statistical modeling perspective, seeking a so-called best model out of many candidates to decompose time series into vegetation dynamics such as trends and abrupt changes *(Cai et al. 2017; Jonsson and Eklundh 2002)*. This single-best-model paradigm is broadly embraced by practitioners *(Powell et al. 2010; Zhao et al. 2018)*, but its use for seeking mechanistic understandings of ecosystems is not necessarily safe *(Chen et al. 2014; Grossman et al. 1996)*.

Mechanistic interpretations of time-series data are sensitive to choices of statistical algorithms or models. When fitting a linear model to decades of AVHRR data, a greening trend in vegetation was inferred and was attributed to global warming *(Myneni et al. 1997)*. If using a piecewise linear model with one changepoint instead, a greening was observed only for the first period whereas a browning for the second, generating new explanations of climate-biosphere interactions *(Wang et al. 2011)*. If piecewise models with multiple changepoints were fitted, the conclusion would change again, giving alternative speculations on drivers of ecosystem changes *(Jong et al. 2012)*. Similar studies with diverging findings abound *(Alcaraz-Segura et al. 2010; Yu et al. 2010)*. Extrapolation from such findings is at stake if applied blindly to validate predictive models and inform resource management.

Inconsistent or contradicting insights gained from different models are a common problem of the single-best-model paradigm. The "best" models are often selected to optimize certain criteria such as AIC and BIC. Depending on data quality and the choices of optimization algorithms and model selection criteria, many "best" models are possible for the same time series *(Banner and Higgs 2017; Cade 2015)*. The usefulness of these models is not dichotomous. Favoring one against others is an over-simplifying strategy that often overlooks the utility of alternative models and ignores model uncertainties. Model selection in the single-best-model paradigm is also complicated by the subjectivity in specifying data analysis models and the inability of simple models to represent complex time-series signals. Model structures with increased complexity are more likely to capture variations in satellite data at multiple timescales, but they are also more likely to overfit the data and their estimation entails sophisticated statistical techniques.

Many problems difficult to tackle by conventional methods can now be addressed by turning to Bayesian statistics—an inferential paradigm that can treat both model parameters and structures probabilistically and offer a unified framework to address uncertainties of various forms *(Denison 2002; Ellison 2004; Finley et al. 2007; Zhao et al. 2008; Zhou et al. 2017)*. Unlike conventional criterion-based methods that choose only a single best model, the Bayesian paradigm can embrace all candidate models, evaluate how probable each of them is a true one, and synthesize the many models into an average model *(Denison 2002; Thomas et al. 2018; Zhao et al. 2013)*. This scheme is known as

Bayesian model averaging (BMA). It belongs to a category of multi-model techniques broadly called ensemble learning. Consideration of many models helps BMA to capture model uncertainty, alleviate model misspecification, and improve flexibilities and generalizability in modeling complex data. These advantages of BMA have been exemplified in numerous case studies across disciplines *(Banner and Higgs 2017; Raftery et al. 2005; Zhang and Zhao 2012; Zhao et al. 2013)*. Despite all the benefits of Bayesian inference or BMA, its use for satellite time-series analysis remains rather limited, with enormous potential to tap.

This study seeks to reliably decipher time-series data for via Bayesian modeling. Our aim is (1) to introduce a generic Bayesian time-series decomposition algorithm for changepoint detection and nonlinear trend analysis, and (2) to demonstrate its applications to satellite data for tracking land and ecosystem nonlinear dynamics. We term the algorithm BEAST—a Bayesian Estimator of Abrupt change, Seasonality, and Trend. BEAST features many advantages over conventional non-Bayesian algorithms. Foremost, it forgoes the single-best-model paradigm and applies the Bayesian ensemble modeling technique to combine numerous competing models and generate a rich set of information unobtainable from non-Bayesian algorithms. BEAST can quantify various sources of uncertainties, detect abrupt changes of any magnitude, and uncover complex nonlinear dynamics from time-series data. But due to the Bayesian computation needed, its applications to high-resolution imagery over large areas may be constrained by computer power.

In what follows, we further justify the value of Bayesian statistics for time-series analysis (Sect 2; Fig. 1), then detail the formulation of our BEAST algorithm (Sect 3; Figs. 2-3), and test the capabilities of BEAST using both simulated and real data (Sect 4 &5; Figs. 4-11). We also discuss the many features of BEAST as contrasted to existing time-series decomposition algorithms, and explain how ensemble learning and Bayesian modeling help to make BEAST a useful tool to capture, monitor, and derive land surface dynamics from satellite data (Sect 6).

2. **Why use Bayesian statistics?**

We begin with extra backgrounds on how time-series data have been conventionally decomposed in non-Bayesian frameworks. Their potential weaknesses are then detailed to justify the needs for Bayesian algorithms. Below, our presentation focuses on time series of Normalized Difference Vegetation Index (NDVI)—a spectral variable measuring land surface greenness or vegetation vigor (Fig. 1a). But the reasoning applies equally to non-NDVI or non-satellite data, such as LAI, albedo, climate, streamflow, and social-ecological indicators.

Ecologically speaking, a NDVI time series captures landscape dynamics at three major timescales *(Kennedy et al. 2014)*: (1) seasonality or periodic variations as forced by intra-annual climatic variations or phenological drivers; (2) gradual changes as driven by long-term environmental trends, chronic disturbances, or successional dynamics; and (3) abrupt changes associated with severe disturbances, sudden recoveries, regime shifts, or altered management practices (e.g., fire, insect, logging, weeding, urbanization, re-vegetation, extreme weather, crop rotation, and climate shift). In this decomposition, the time series is treated as the sum of the first two components—seasonal and trend signals (Fig. 1b). The third component—abrupt changes—do not stand out alone but is embedded in seasonality and trends as changepoints (Fig. 1b, blue vertical bars).

Mathematically speaking, the search for ecological interpretations of a time series reduces to finding the relationship between NDVI (*y*) and time (*t*) from the observed data at n points of time $\mathcal{D} = \{t_i, y_i\}_{i=1,\ldots,n}$ via a statistical model $\hat{y}(t) = f(t)$. The model generally treats the time series $y(t_i)$ as an addition of seasonal $S(\cdot)$ and trend $T(\cdot)$ signals(Fig. 1):

$$\boldsymbol{\hat{y}(t_i) = f(t_i; \Theta) = S(t_i; \Theta_s) + T(t_i; \Theta_T)}, \mathbf{i=1,\ldots,n} \qquad \mathbf{(1)}$$

where the parameters Θ_s and Θ_T specify the seasonal and trend signals; they also encode the abrupt changes implicitly. By analogy to linear regression, the time *t* and data *y* are independent and dependent variables, respectively; Θ_s and Θ_T are parameters to be estimated from the data $\mathcal{D}$.

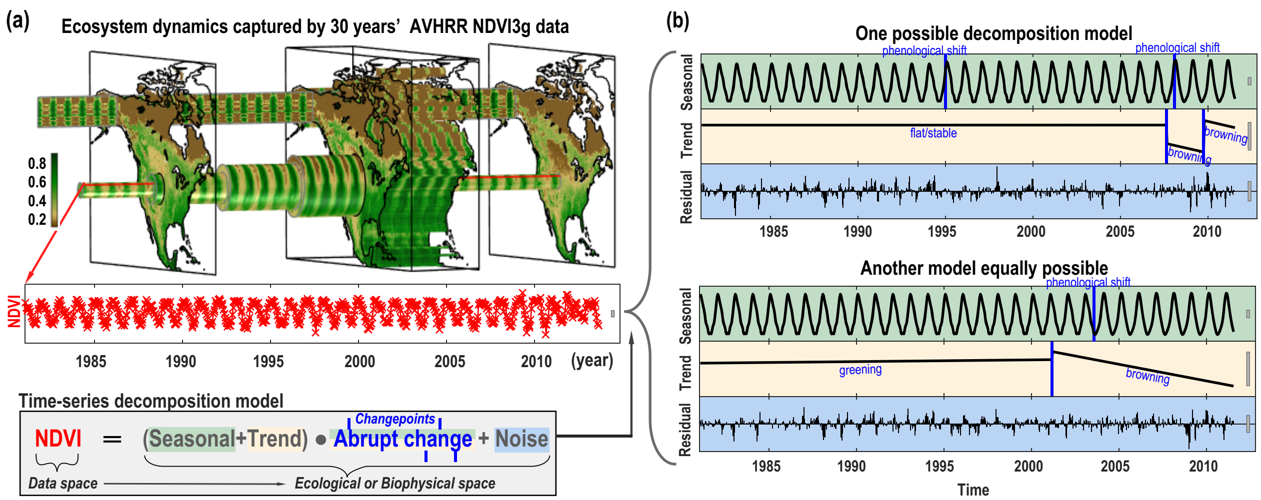

Fig. 1. Tracking land surface dynamics from space is treated here as a time-series decomposition problem. (**a**) A 3D volumetric view of 30 years of AVHRR NDVI data depicts ecosystem dynamics at three timescales: seasonality, trend (e.g., climate-driven responses or successional dynamics), and abrupt change (e.g., disturbance or changepoint). Algorithmically speaking, decomposition of a time series into these three components is a model selection problem, seeking an "optimal" model structure that best fits the time series. (**b**) But the use of different inferential procedures or selection criteria yields different or even contradictory decompositions, with adverse implications. For example, two "optimal" models in (b) can fit the same time series of (a) almost equally well, but with inconsistent decompositions and ecological interpretations. Vertical blue bars denote changepoints in seasonal dynamics or trends. The equal plausibility of the two "best" models highlights an inherent weakness of many existing satellite time-series analyses for studying ecosystem changes.

By decomposing a time series with Eq. 1, we seek to answer the following questions:

(1) How many changepoints occur and when? Changepoints indicates any abrupt changes in trend/seasonal signals (*Jamali et al. 2015)*. By "abrupt", we refer to not only sudden NDVI jumps (e.g., forest clearing or quick recovery) but also any turning points or breakpoints at which trend or seasonal signals start to deviate from the previous regular trajectories. This definition is broader and more inclusive than that assumed by other algorithms. As examples, a smooth recovery from tree stand-clearing is often associated with only one changepoint by many algorithms, but in our definition, the recovery trajectory may have many changepoints related to different succession stages

or rates of recovery. A subtle transition in vegetation dynamics caused by a shift in climate regime is rarely considered as a changepoint by many algorithms, but in our definition it is.

(2) What is the underlying trend? A trend is not just a linear line but can be a complex nonlinear trajectory interspersed with changepoints. The transient trend trajectory at changepoints are rarely true discontinuous jumps but rather quasi-continuous sharp transitions. Detection of trends with high fidelity is critical for inferring subtle drivers of ecosystem dynamics (e.g., climatic effects).

(3) What is the underlying seasonal signal? A seasonal signal may be also interspersed by changepoints. Seasonal changepoints do not necessarily coincide with trend changepoints. Detection of seasonal changepoints helps to identify potential drivers of phenology changes.

Any uncertainties or errors in inferring the model Eq. 1 will be translated to those in answering these questions, thereby engendering contradictory or wrong ecological insights into ecosystem dynamics.

Existing methods to infer the model or relationship f come in many fashions *(Brooks et al. 2014; Kennedy et al. 2010; Zhu and Woodcock 2014).* Often, the trend is parameterized and approximated by linear, piecewise-linear, or polynomial models *(Browning et al. 2017)*. The seasonal signal is modeled via flexible basis functions, such as Fourier curves and wavelets *(Brooks et al. 2012; Jiang et al. 2010; Martínez and Gilabert 2009; Shu et al. 2017)*. Another alternative is to ignore seasonal signals by fitting a trend model to a sub-time series (e.g., summertime NDVI only) *(Wang et al. 2011)*. Moreover, abrupt NDVI changes are implicitly encoded in the parameters Θ_T and Θ_S. These changepoints also need to be inferred from the data $\mathcal{D}$ *(Chen et al. 2014)*. Such diverse options for model configurations lead to a large or even infinite number of candidate models for analyzing the same time series. Conventional methods aim to seek the "best" model and discard others based on selection criteria, such as mean square error, Cp, AIC, anomaly threshold, or subjective criteria *(Chen et al. 2014; Wang et al. 2011)*.

These conventional methods have potential weaknesses that were not always articulated in previous studies (Fig. 1). First, vegetation dynamics normally shows a nonlinear trend *(Burkett et al. 2005; Jentsch et al. 2007)*, which is not guaranteed to be adequately approximated by a single linear, piecewise-linear, or polynomial model. Second, many conventional analyses make too restrictive model assumptions. For example, prior studies often assumed a prescribed number of changepoints or a fixed harmonic order in seasonality *(Lu et al. 2004; Wang et al. 2011)*, which is too arbitrary a choice. Third, the true model for NDVI dynamics is essentially unknown so that model misspecification is inevitable *(Kennedy and O'Hagan 2001)*. The use of misspecified or wrong models is of little concern for those applications on retrievals of biophysical variables *(Shmueli 2010; Zhao et al. 2018)*, but it becomes problematic for ecological interpretation of NDVI data simply because different models imply contrasting or contradicting hypotheses. Such model uncertainties are typically ignored by non-Bayesian approaches.

Fourth, even for the same class of model type, the final model chosen is sensitive to not only model selection criteria but also data noises, thus opening up possibilities for inconsistent interpretations. The hyper-sensitivity of model inference to data noises has been widely recognized and reported in the literature of various disciplines *(Grossman et al. 1996; Oreskes et al. 1994; Zhao et al. 2013)*. Fifth, the number of all possible models is often enormous, making it computationally infeasible to evaluate all of them. Instead, efficient approaches, such as forward sweep, greedy searching, and genetic algorithms, come into play to evaluate a finite selection of models with regards to some optimization schemes, but these schemes tend to find sub-optimal solutions *(Denison 2002)*. Sixth, diagnostic statistics generated using conventional approaches are inadequate for answering many practical questions (e.g., what is the probability of seeing an abrupt change in the year 2002?).

These potential limitations of conventional statistical modeling can be alleviated by switching from the conventional single-best-model paradigm to the Bayesian paradigm (Fig. 2). Advantages of the Bayesian paradigm are being demonstrated by a growing body of theoretical and empirical evidence *(Denison 2002; Rankin et al. 2017; Reiche et al. 2015; Zhao et al. 2013; Zhou et al. 2017)*. Foremost, Bayesian inference treats both model parameters and structures as random and therefore characterizes them explicitly and probabilistically. As such, Bayesian inference tends not to deem any single model configuration as the true model, but instead recognizes the relevance and usefulness of all the potential models *(Kennedy and O'Hagan 2001; Zhao et al. 2013)*. Specifically, each model is assigned a probability being the true model (Fig. 2); this probability can be learned from data and then used as an informative weight to synthesize all the models into a weighted average model. This Bayesian model averaging scheme is flexible enough to approximate complex relationships that cannot be represented by individual models (Fig. 2). It also alleviates the adverse consequences of model misspecification and tackles model uncertainty *(Zhao et al. 2013)*.

Computationally, Bayesian inference or BMA is implemented via stochastic sampling algorithms known as Markov Chain Monte Carlo (MCMC) *(Denison 2002; Green 1995)*. MCMC sampling helps to effectively explore the enormous model space at a reasonable computation cost. The use of MCMC circumvents analytical intractability and enables the Bayesian paradigm to handle the complexity that conventional methods cannot handle. MCMC also generates various sample-based statistics to test hypotheses that are difficult to tackle using the conventional paradigm.

Bayesian statistics can aid in inferring the model of Eq. 1, due especially to its additive nature: a time series is the sum of seasonal and trend signals, with changepoints being inseparable parts of them. Inference of the three— trend, seasonality, and changepoints,—is not separable. Any estimation error in one will be leaked to bias the estimation of others. It is unlikely to correctly detect changepoints if the trend or seasonality is not well modelled. Trend analysis and changepoint detection are two sides of the same goal. It is also impossible to estimate true decomposition uncertainties if not accounting for model misspecification simultaneously for the three components.

Therefore, reliable time-series decomposition requires sufficiently approximating the nonlinearity of both trend and seasonality and simultaneously incorporating model uncertainties of all sorts. These issues are explicitly tackled by Bayesian inference, as detailed next.

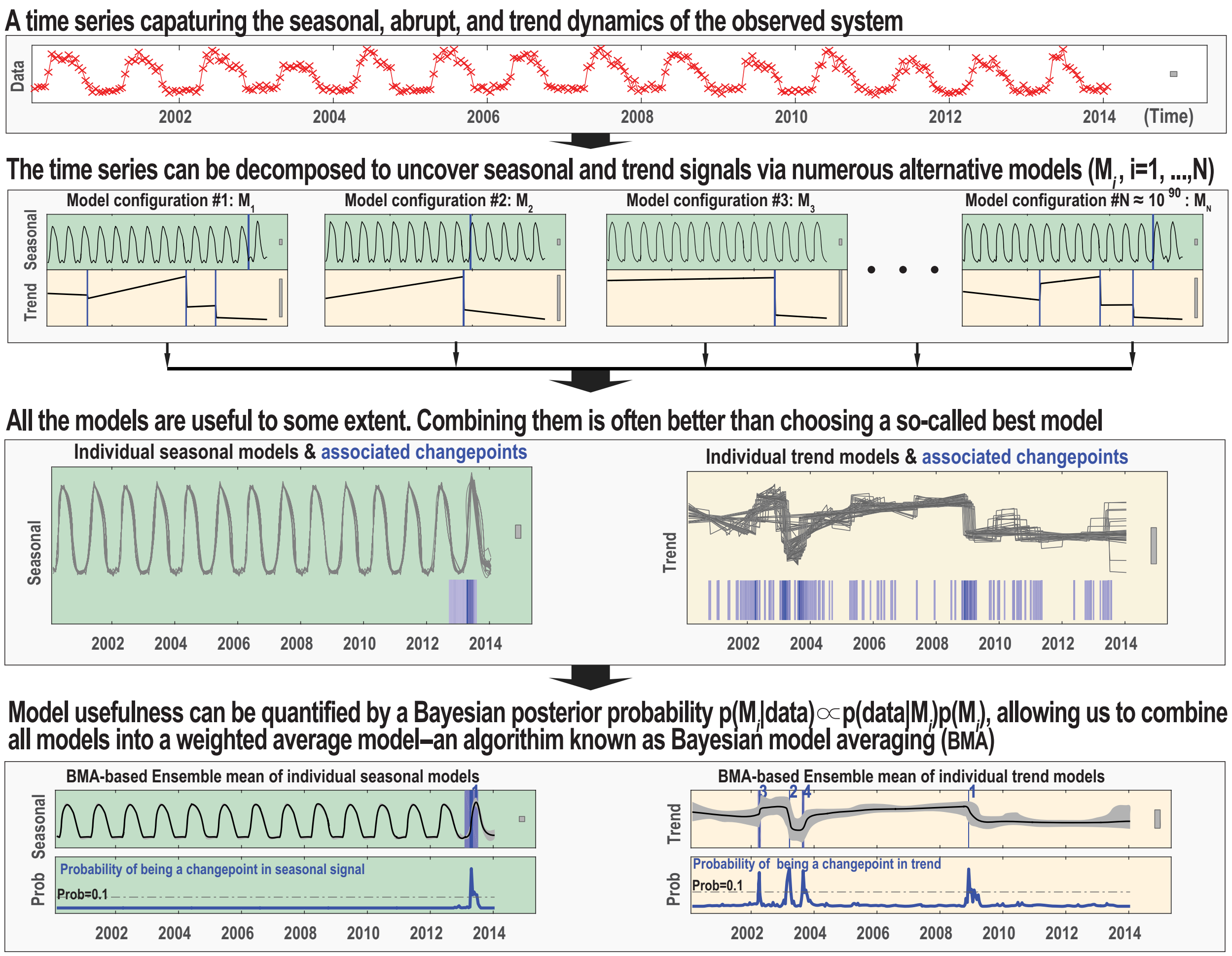

Fig. 2. Illustration of BEAST—a Bayesian ensemble time-series decomposition algorithm. Our modeling philosophy is that a time series can be fitted by numerous competing models, all of which are wrong but useful to some degree. Conventional methods choose the "best" model, ignoring model uncertainty or misspecification and opening up room for fortuitous conclusions (Fig. 1). As a remedy, BEAST quantifies the relative usefulness of individual models (i.e., model structures) and incorporates all the models into the inference via Bayesian model averaging. This ensemble learning makes BEAST a universal approximator of complex nonlinear trends and allows BEAST to account for uncertainties difficult to consider by non-Bayesian methods. For example, model uncertainty is explicitly addressed (e.g., gray envelope around the fitted seasonal or trend signals are 95% credible intervals). BEAST not only

detects the changepoints but also quantifies their probabilities of being true changepoints, providing confidence measures to guide informative interpretation of satellite time-series data.

3. BEAST: Bayesian estimator of abrupt change, seasonality & trend

This section describes the formulation and implementation of our BEAST time-series decomposition algorithm. The description is inevitably mathematical. Readers not interested in technical specifics may skip to Section 4 while re-visiting Figs. 2 &3 or Section 2 for an overview of the concept and capabilities of BEAST. The implemented software is available as both a Matlab library and an R package (*to be released upon acceptance of this ms*).

3.1 Parametric form of BEAST for time-series decomposition

Our analysis considers a time series $\mathcal{D} = \{t_i, y_i\}_{i=1,\ldots,n}$ to be composed of three components—seasonality, trend, abrupt changes—plus noise (Fig. 1b), which is formulated as a rewriting of Eq. 1:

$$\boldsymbol{y_i} = \boldsymbol{S}(\boldsymbol{t_i}; \boldsymbol{\Theta_s}) + \boldsymbol{T}(\boldsymbol{t_i}; \boldsymbol{\Theta_T}) + \boldsymbol{\varepsilon_i}. \qquad \textbf{(2)}$$

Here, we assume the noise ε to be Gaussian, capturing the remainder in the data not explained by the seasonal $S(\cdot)$ and trend $T(\cdot)$ signals. Following the common practice, we adopted general linear models to parameterize $S(\cdot)$ and $T(\cdot)$ *(Jiang et al. 2010; Verbesselt et al. 2010b)*. Abrupt changes are implicitly encoded in the parameters Θ_s and Θ_T of the seasonal and trend signals.

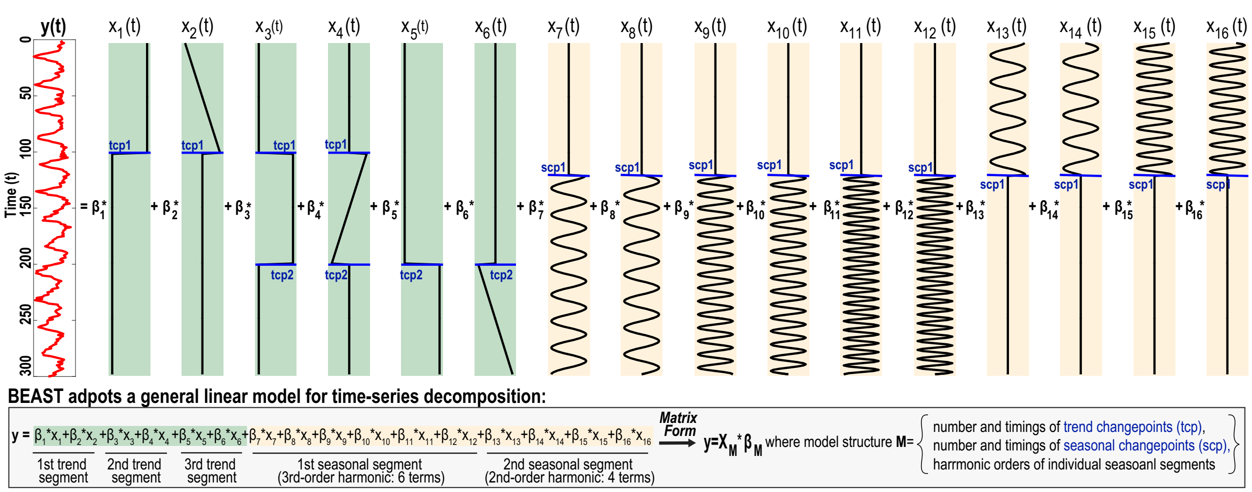

Fig. 3. How does BEAST decompose time series? BEAST is an additive model $y = \mathrm{x_M}\beta_\mathrm{M}$, formulating a time series y(t) as the linear combination of many basis functions $\mathrm{x_M}$ (e.g., line segments x_1-x_6 or harmonics x_7-x_{16} that are zero-valued except for the active segments). These basis terms are specified by the model structure parameters M: numbers and locations of seasonal/trend changepoints (i.e., horizontal blue bars such as tcp1 and scp1) and seasonal harmonic orders (e.g., 3rd and 2nd in this example). The aim is to infer not only the coefficients β_M for a given M but also the model structure M itself (i.e., changepoint and harmonic orders). The combinatorics of all possible changepoints and harmonic orders gives an enormous model space with numerous candidate basis terms, making it computationally impossible to pinpoint the true best model. BEAST infers M by randomly traversing the model space via Bayesian model selection, so it is essentially a Bayesian general linear regression model. (The time axis is oriented vertically for ease of displaying.)

Specifically, the seasonal signal $S(t)$ is approximated as a piecewise harmonic model, defined with respect to p knots in time at ξ_k, k=1,…p (Fig. 3). These knots divide the time series into (p+1) intervals $[\xi_k, \xi_{k+1}]$, j=0,…,p, where $\xi_0 = t_0$ and $\xi_{p+1} = t_n$ are the starting and ending times of the data. The model is specified for each of the (p+1) segments $[\xi_k, \xi_{k+1}]$, k=0,…,p, taking the form of

$$S(t) = \sum_{l=1}^{L_k} a_{k,l}\sin(\frac{2\pi lt}{P}) + b_{k,l}\cos(\frac{2\pi lt}{P}) \text{ for } \xi_k \le t < \xi_{k+1}, k = 0, \dots, p$$

Here, P is the period of the seasonal signal (i.e., one year in our cases); L_k is the harmonic order for the k-th segment and is an unknown segment-specific parameter; and the coefficients $\{a_{k,l}, b_{k,l}\}_{l=1,\dots,L_k}$ are the parameters for sins and cosines. This harmonic model is non-continuous as a whole; the knots ξ_k indicate the changepoints at which abrupt seasonal changes may occur. Both the total number of changepoints p and their timings $\{\xi_k\}_{k=1,\dots,p}$ are unknown parameters to be estimated. In short, we need the following parameters to fully specify the seasonal harmonic curve:

$$\Theta_s = \{p\} \cup \{\xi_k\}_{k=1,\dots,p} \cup \{L_k\}_{k=0,\dots,p} \cup \{a_{k,l}, b_{k,l}\}_{k=0,\dots,p;l=1,\dots,L_k}$$

which includes the number and timings of changepoints, the harmonic orders for all the (p+1) segments, and the coefficients of all the harmonic terms. All have to be estimated.

The trend $T(t)$ is modeled as a piecewise linear function with respect to m knots at τ_j, j=1,…m (Fig. 3), which divide the time span into (m+1) intervals $[\tau_j, \tau_{j+1}]$, j=0,…,p, with $\tau_0 = t_0$ and $\tau_{m+1} = t_n$ being the start and end of the time series. The trend knots or changepoints τ_j are not necessarily the same as the seasonal changepoints ξ_k. The trend over each interval simply is a line segment (Fig. 2), defined by coefficients a_j and b_j:

$$T(t) = a_j + b_j t \quad \text{for } \tau_j \le t < \tau_{j+1}, j = 0, \dots, m$$

Similar to the seasonal signal, the number of changepoints m and their timings $\{\tau_j\}_{j=1,\dots,m}$ are unknown parameters to estimate. Hence, the full set of parameters specifying the trend T is

$$\Theta_T = \{m\} \cup \{\tau_j\}_{j=1,\dots,m} \cup \{a, b_j\}_{j=0,\dots,m}$$

which comprises the number and timings of trend changepoints and the intercepts and slopes of individual line segments.

Both sets of the parameters, Θ_s and Θ_T, need to be estimated from the data $\mathcal{D}$. For ease of presentation, we re-classified the parameters Θ_T and Θ_s into two groups (Fig. 3): $\{\Theta_T, \Theta_s\} = \{\mathrm{M}, \boldsymbol{\beta}_\mathrm{M}\}$. The first group M refers to model structure, including numbers and timings of trend and seasonal changepoints, and seasonal harmonic orders:

$$\mathrm{M} = \{m\} \cup \{\tau_j\}_{j=1,\dots,m} \cup \{p\} \cup \{\xi_k\}_{k=1,\dots,p} \cup \{L_k\}_{k=0,\dots,p}.$$

The second group $\boldsymbol{\beta}_\mathrm{M}$ is the segment-specific coefficient parameters used to determine exact shapes of the trend and seasonal curves once the model structure M is given. Collectively, $\boldsymbol{\beta}_\mathrm{M}$ is denoted by

$$\boldsymbol{\beta}_\mathrm{M} = \{a, b_j\}_{j=0,\dots,m} \cup \{a_{k,l}, b_{k,l}\}_{k=0,\dots,p;l=1,\dots,L_k}.$$

The subscript M indicates the dependence of $\boldsymbol{\beta}_\mathrm{M}$ on model structure M.

After this re-grouping, the original general linear model Eq. 2 becomes a familiar form:

$$\boldsymbol{y(t_i) = \mathrm{x}_\mathrm{M}(t_i)\beta_\mathrm{M} + \varepsilon_i} \qquad \textbf{(3)}$$

where $\mathbf{x}_{\mathrm{M}}(t_i)$ and $\boldsymbol{\beta}_{\mathrm{M}}$ are dependent variables and associated coefficients, respectively. Again, the subscript M suggests that the exact form of $\mathbf{x}_{\mathrm{M}}$ and the coefficients in $\boldsymbol{\beta}_{\mathrm{M}}$ both depend on the model structure M (e.g., numbers and timings of changepoints). For example, column vectors of the design matrix $\mathbf{x}_{\mathrm{M}}(t_i)$ are associated with individual segments of the piecewise linear and harmonic models (Fig. 3), with the number of coefficients in $\boldsymbol{\beta}_{\mathrm{M}}$ being $2(m+1)+2\sum_{k=0}^{p} L_k$.

As revealed in the re-formulated model of Eq. 3, the inference of vegetation dynamics now reduces to a model selection problem—determining an appropriate model structure M, including the numbers and timings of changepoints and the harmonic orders. Identifying an optimal model structure M for our problem is analogous to choosing the best subset of variables for simple linear regression. Once a model structure M is selected, its coefficients $\boldsymbol{\beta}_{\mathrm{M}}$ are straightforward to estimate. However, unlike simple linear regression, the number of possible model structures for Eq. 3 is extremely large. Even for a time series of moderate length (e.g., n >100), it takes billions of years' computation to enumerate all possible models for finding the best one that optimizes certain criteria (e.g., BIC). We circumvented this problem by resorting to Bayesian inference, as described next.

3.2 Bayesian formulation of BEAST

We extended the general linear model of Eq. 2 or 3 to build a Bayesian model for detecting abrupt change, seasonality, and trend from time-series data. In the Bayesian modeling, all the unknown parameters are considered random, including model structure M, coefficients $\boldsymbol{\beta}_{\mathrm{M}}$, and data noise σ^2. Given a time series $\mathcal{D}=\{t_i, y_i\}_{i=1,\ldots,n}$, the goal is to get not just the best possible values of these parameters but more importantly, their posterior probability distribution $p(\beta_{\mathrm{M}}, \sigma^2, \mathrm{M}|\mathcal{D})$. By Bayes' theorem, this posterior is the product of a likelihood and a prior model:

$$\boldsymbol{p(\beta_{\mathrm{M}}, \sigma^2, \mathrm{M}|\mathcal{D}) \propto p(\mathcal{D}|\beta_{\mathrm{M}}, \sigma^2, \mathrm{M})\pi(\beta_{\mathrm{M}}, \sigma^2, \mathrm{M}).} \qquad \textbf{(4)}$$

Here, the likelihood $p(\mathcal{D}|\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2, \mathrm{M})$ is the probability of observing the data $\mathcal{D}$ given the model parameters $\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2$, and M. Its form is governed by the general linear model $y=\mathbf{x}_{\mathrm{M}}\boldsymbol{\beta}_{\mathrm{M}}+\varepsilon$ in Eq. 3.

Owing to the normality of error ε, the likelihood is simply Gaussian $p(\mathcal{D}|\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2, \mathrm{M}) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{N}(y_i; \mathbf{x}_{\mathrm{M}}(t_i)\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2)$.

To complete our Bayesian formulation, what remains is to specify the second term of Eq. 4, $\pi(\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2, \mathrm{M})$, which is called the prior distribution of the model parameters. By definition, we have

$$\pi(\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2, \mathrm{M}) = \pi(\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2|\mathrm{M})\pi(\mathrm{M}).$$

Therefore, it suffices to elicit the conditional prior $\pi(\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2|\mathrm{M})$ and the model prior $\pi(\mathrm{M})$ separately. The priors encode our existing knowledge or beliefs in possible values of the model parameters. Because of a lack of such general knowledge beforehand, our choices are flat priors, close to being non-informative. First, for $\pi(\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2|\mathrm{M})$, we considered the normal-inverse Gamma distribution and introduced an extra dispersion hyperparameter v into it to further reflect our vague knowledge of the magnitude of model coefficients $\boldsymbol{\beta}_{\mathrm{M}}$. Second, for the prior on model structure $\pi(\mathrm{M})$, we assumed that the numbers of changepoints are any nonnegative integers that are equally probable a prior. The exact formula of our priors are detailed in Appendix A.

Given our likelihood and prior models, the posterior of the model parameters becomes

$$\boldsymbol{p}\left(\beta_{\mathbf{M}}, \boldsymbol{\sigma}^2, \boldsymbol{v}, \mathbf{M} \middle| \boldsymbol{\mathcal{D}}\right) \propto \prod_{i=1}^{n} \boldsymbol{N}(\boldsymbol{y}_i; \mathbf{x}_{\mathbf{M}}(\boldsymbol{t}_i)\boldsymbol{\beta}_{\mathbf{M}}, \boldsymbol{\sigma}^2)\, \boldsymbol{\pi}_{\beta}\left(\beta_{\mathbf{M}}, \boldsymbol{\sigma}^2, \boldsymbol{v} \middle| \mathbf{M}\right)\boldsymbol{\pi}(\mathbf{M}). \tag{5}$$

Its complete formulation after incorporating each component prior is expanded and presented in Appendix A, with more technical details explained there for interested readers.

3.3 Monte Carlo-based Inference

The posterior distribution $p\left(\boldsymbol{\beta}_{\mathbf{M}}, \sigma^2, v, \mathbf{M} \middle| \boldsymbol{\mathcal{D}}\right)$ of Eq. 5 encodes all the information essential for inferring ecosystem dynamics. But it is analytically intractable, so we resorted to MCMC sampling to generate a realization of random samples for posterior inference. The MCMC sampling algorithm we used is a hybrid sampler that embeds a reverse-jump MCMC sampler (RJ-MCMC) into a Gibbs sampling framework, as briefly described below.

The Gibbs framework samples the following three conditional posterior distributions in alteration for a total of N iterations.

$$
\begin{aligned}
&p\left(\mathbf{M}^{(i+1)}\middle|v^{(i)},\boldsymbol{\mathcal{D}}\right);\\
&p\left(\boldsymbol{\beta}_{\mathbf{M}}^{(i+1)},\sigma^{2^{(i+1)}}\middle|v^{(i)},\mathbf{M}^{(i+1)},\boldsymbol{\mathcal{D}}\right);\\
&p\left(v^{(i+1)}\middle|\boldsymbol{\beta}_{\mathbf{M}}^{(i+1)},\sigma^{2^{(i+1)}},\mathbf{M}^{(i+1)},\boldsymbol{\mathcal{D}}\right);
\end{aligned}
\tag{6}
$$

These three conditional posteriors permit generating the (i+1)-th sample $\{\mathbf{M}^{(i+1)},\boldsymbol{\beta}_{\mathbf{M}}^{(i+1)},\sigma^{2^{(i+1)}},v^{(i+1)}\}$ from the previous sample $\{\mathbf{M}^{(i)},\boldsymbol{\beta}_{\mathbf{M}}^{(i)},\sigma^{2^{(i)}},v^{(i)}\}$. In particular, the second and third conditional posteriors are a normal-inverse Gamma distribution and a Gamma distribution (Appendix A), which are easy to sample. In contrast, the first conditional posterior $p\left(\mathbf{M}^{(i+1)}\middle|v^{(i)},\boldsymbol{\mathcal{D}}\right)$ is difficult to sample because it is defined only up to an unknown proportionality constant (Appendix A) and also because the dimension of M varies from one model to another. These two difficulties were tackled by using the RJ-MCMC algorithm *(Denison 2002; Green 1995)*. Details about RJ-MCMC are available in Zhao et al. *(2013)* and not given here.

3.4 **Posterior inference of changepoints, seasonality, and trends**

The preceding MCMC algorithm generates a chain of posterior samples of length N $\{\mathbf{M}^{(i)},\boldsymbol{\beta}_{\mathbf{M}}^{(i)},\sigma^{2^{(i)}},v^{(i)}\}_{i=1,\cdots,\mathrm{N}}$. The chain captures all the information essential for inference of land dynamics, including trends, seasonal variations, and abrupt changes (Fig. 2). In particular, the sampled model structure $\mathbf{M}^{(i)}$, such as timings of changepoints and seasonal harmonic orders, can be directly translated into the model's covariates $\mathbf{x}_{\mathcal{M}^{(i)}}(t)$ (Fig.3), with their associated coefficients being $\boldsymbol{\beta}_{\mathbf{M}}^{(i)}$. Each sampled model $\mathbf{M}^{(i)}$ gives one estimate of the land dynamics, $\mathbf{x}_{\mathbf{M}^{(i)}}(t)\cdot\boldsymbol{\beta}_{\mathbf{M}}^{(i)}$. Combining the individual estimates provides not only a final BMA estimate but also uncertainty measures. The BMA estimate of time-series dynamics is the averaging of all the sampled models:

$\hat{y}(t) \approx \frac{\sum_{i=1}^{N} \mathbf{x}_{\mathbf{M}^{(i)}}(t)\boldsymbol{\beta}_{\mathbf{M}}^{(i)}}{N}$. The associated uncertainty is given as a sample-based variance estimate: $\widehat{var}[\hat{y}(t)] \approx \frac{\sum_{i=1}^{N} [\mathbf{x}_{\mathbf{M}^{(i)}}(t)\boldsymbol{\beta}_{\mathbf{M}}^{(i)} - \hat{y}(t)]^2}{N-1}$.

Although each single model $\mathbf{M}^{(i)}$ is a piecewise model, the combination of all the individual models enables the BMA estimate $\hat{y}(t)$ to approximate arbitrary nonlinear signals. Moreover, because the covariates $\mathbf{x}_{\mathbf{M}^{(i)}}(t)$ and model coefficients $\boldsymbol{\beta}_{\mathbf{M}}^{(i)}$ are simply a coalescing of the individual elements of the trend and seasonal signals, these elements can be separated to recover the trend and seasonal components, respectively (Fig. 3).

More interestingly, the sampled model structure $\{\mathbf{M}^{(i)}\}_{i=1,\cdots,\mathrm{N}}$, which is $\{m^{(i)}, \tau_{k=1,\ldots,m^{(i)}}^{(i)}, p^{(i)}, \xi_{k=1,\ldots,p^{(i)}}^{(i)}, L_{k=0,\ldots,p^{(i)}}^{(i)}\}$, allows making inference and testing hypothesis related to abrupt changes and land disturbances. Specifically, the chain $\{m^{(i)}\}_{i=1,\cdots,\mathrm{N}}$ or $\{p^{(i)}\}_{i=1,\cdots,\mathrm{N}}$ gives an empirical distribution of the number of changepoints in the trend or seasonal signals; therefore, the mean total numbers of trend and seasonal changepoints can be estimated as $\overline{m} = \frac{\sum_{i=1}^{N} m^{(i)}}{N}$ and $\overline{p} = \frac{\sum_{i=1}^{N} p^{(i)}}{N}$. For the seasonal signal, the chain of $\{L_{k=0,\ldots,p^{(i)}}^{(i)}\}_{i=1,\cdots,\mathrm{N}}$ can be used to compute the average harmonic order $\overline{L}(t)$ needed to sufficiently approximate the seasonality for any given time t:

$$\overline{L}(t) = \frac{\sum_{i=1}^{N} L_{k_i}^{(i)}}{\mathrm{N}}, \quad subject\ to\ t \in [\xi_{k_i}^{(i)}, \xi_{k_i+1}^{(i)}].$$

The use of differing harmonic orders for different times or intervals is a strength distinguishing BEAST from those algorithms that choose a pre-set, fixed order uniformly for the seasonal signal.

In addition, the chains $\{\tau_{k=1,\ldots,m^{(i)}}^{(i)}, \xi_{k=1,\ldots,p^{(i)}}^{(i)}\}_{i=1,\cdots,\mathrm{N}}$ indicate the exact timings at which the trend or seasonal changepoints occurred for the sampled individual models. From these chains, we can estimate the probability that a changepoint occurs at a time t_s or within a interval $[t_s, t_e]$ by counting the frequency of the sample $\{\tau_{k=1,\ldots,m^{(i)}}^{(i)}\}_{i=1,\cdots,\mathrm{N}}$ containing the time t_s or falling into $[t_s, t_e]$:

$$p(changepoint\ at\ t_s\ or\ within\ [t_s, t_e]\ |\boldsymbol{\mathcal{D}}) \approx \frac{\#of\mathcal{M}^{(i)}\ that\ includes\ t_s\ or\ falls\ into\ [t_s, t_e]}{\mathrm{N}}.$$

Likewise, given an estimated changepoint, we can derive its credible interval. We can also calculate many more sophisticated statistics, such as what is the conditional probability of observing a changepoint in trend at a time if another changepoint has already occurred somewhere, and what is the joint probability of observing a changepoint in trend at one time and a seasonal changepoint at another time? All these sample-based statistics serve as important measures for statistical diagnostics such as uncertainty analysis and hypothesis testing. For example, a changepoint with an estimated occurrence probability of 3% is less likely to represent a true abrupt change.

3.5 Software Implementation

We implemented BEAST in the C programming language. The core is the MCMC sampler of Eq. 6, an iterative process involving heavy matrix computation such as matrix multiplication and inversion. We tested several matrix libraries and found that Intel's MKL was the fastest. We also implemented an R and a Matlab interface to BEAST: an R package named "Rbeast" is forthcoming. To facilitate algorithm assessment, we further developed a toolkit "trackEcoDyn" (Fig. 10). It offers a graphical user interface (GUI) that allows interactively running BEAST and more importantly, manually analyzing and interpreting Landsat time series data in reference to other image sources (e.g., Landsat images, and aerial photos). The tool is automatically linked with Google Earth and its high-resolution historical imagery, facilitating visually cross-checking land histories among multiple sources. The purpose of trackEcoDyn is to aid in interpreting Landsat time series and collecting ground-reference data for algorithm assessment, as used below in our second case study.

4. Examples

Three examples are given below to illustrate the basic usage and typical outputs of BEAST. To highlight its differences from existing methods, we also compared BEAST to a community-endorsed algorithm called bfast *(Verbesselt et al. 2010a)*. Bfast is a criterion-based method seeking a single best model, as opposed to the Bayesian inference with BEAST. Bfast and BEAST adopt the same general linear model form, thus allowing us to isolate the effects of inference paradigms and remove other confounding effects on the algorithm comparison. As shall be seen below, despite the use of the same parametric model, BEAST and bfast disagreed on decomposition results.

4.1 Example 1: A simulated time series

In Example 1, we simulated a time series of length n=774 with a period of P=24 (Fig. 4a1). In the simulation, the true reference seasonal signal has two seasonal changepoints (scp), giving three seasonal segments; the seasonality was simulated using different orders of harmonics for individual segments (Fig. 4-b1). The true trend has two trend changepoints (tcp), giving three trend segments: The first two are piecewise-linear; the third is a slow-varying nonlinear continuous signal with no abrupt jumps (Fig 3a). We chose such a continuous trend for the third segment because this is often the case for real ecosystem dynamics and the performances of conventional methods for such nonlinear trends were rarely evaluated in the remote sensing time-series literature.

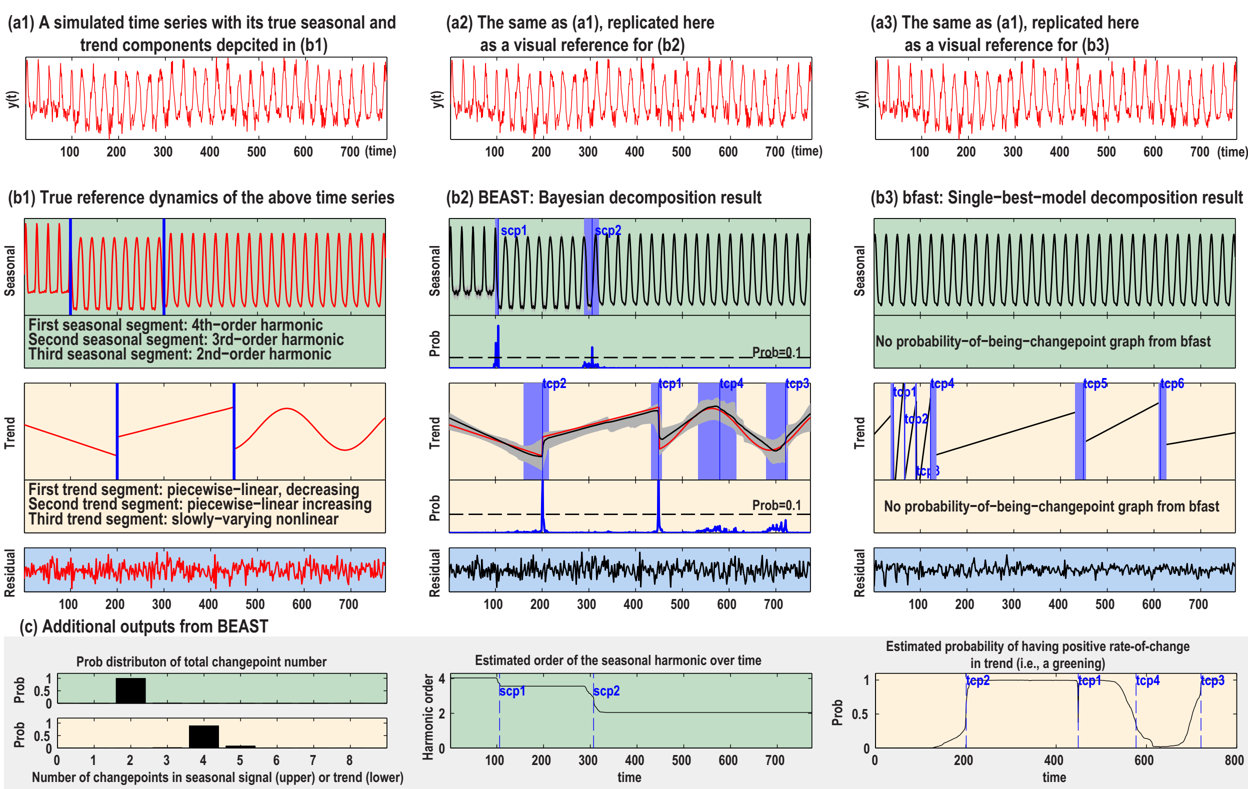

Fig. 4. Example 1: Use of a simulated time series (a1-a3) to illustrate BEAST. The true dynamics underlying the time series (b1) were uncovered by BEAST accurately (b2). Specific information estimated by BEAST includes, but is not limited to, seasonal and trend signals, seasonal and trend changepoints (scp or tcp, as denoted by vertical blue bars), and harmonic orders of individual seasonal segments (c, middle). BEAST also provided an array of useful uncertainty diagnostic statistics, such as credible intervals of the estimated signals (i.e., gray envelopes), the probability of observing a scp or tcp at any given time, the probability distribution of total numbers of scp or tcp (c, left), and the probability of having a positive rate-of-change in trend (c, right). For comparisons, the results from the single-best-model algorithm "bfast" are given in (b3). Bfast detected no scp and six tcps.

Use of BEAST and bfast is sensitive to the specification of two hyperparameters: maximum number of changepoints (M_{max})—an upper limit imposed on how many changepoints are allowed in a single model; minimum separation interval (h) —the minimum distance in time allowed between two neighboring changepoints in a single model. In this example, we chose M_{max} =8 and h= 24 (one period). (In bfast, h is expressed as the ratio of the interval to the time series length.)

BEAST uncovered the true dynamics from the simulated time series with high fidelity. The detected signals correlated well with the true references [r=0.998 (seasonal) and 0.956 (trend), n=774]. BEAST not only successfully pinpointed the two true scps but also correctly identified the differing harmonic orders for the three seasonal segments (Fig. 4c, middle). In the trend, BEAST precisely detected the two reference tcps associated with the piecewise-linear segments. For the third nonlinear trend segment, BEAST additionally detected 2.2≈2 tcps to capture the sinuous nonlinearity. Because changepoints are defined as any timings at which the trend deviates from its previous linear trajectory (Sect 2), in theory, the nonlinear trend segment of this example is fraught with changepoints through the time. This theoretical expectation aligns well with the BEAST-estimated probability of changepoint occurrence (Fig. 4b2), wherein the estimated probability curve was often nonzero with many small peaks over the third trend segment. All the probabilities were small, indicating the very low likelihood of identifying high-intensity abrupt changes in this nonlinear trend segment, except at the two turning points of the sinuosity. In contrast to BEAST, bfast detected no scps and six tcps (Fig. 4b2).

BEAST also produced a rich set of uncertainty measures useful to guide the interpretation of inferred dynamics (Figs. 4b2 & 4c). As examples, the synthesis of individual models allows BEAST to generate uncertainties that incorporate both data noises and model missspecification. The inferred trend signal in Fig. 4b2 was not identical to the true signal, but the envelopes of 95% uncertainty intervals enclosed the true signal almost completely, attesting to the utility and reliability of the estimated credible intervals (Fig. 4b2). BEAST tells not only the most likely timings and numbers of tcp or scps but also the probability of observing a scp or tcp for any given time as well as the probability of detecting a certain total number of scps or tcps (Fig. 4c, left). In this example, the probabilities of having 2 scps were 0.9963, leaving only a probability of 0.0037 to find other numbers of scps and suggesting high confidence in pinpointing the two scps.

Likewise, BEAST can derive the probability distribution of harmonic orders needed to adequately model a seasonal segment. Another output important for ecological remote sensing is pertinent to the rate of change in trend. For example, BEAST infers not only the sign of the change (e.g., a greening or browning) but also the probability of having a greening or browning at any time (Fig. 4c, right). In essence, for all parameters of interest, BEAST infers not only the most likely values but also the associated error bars and even more, the associated probability distributions, the latter of which are generally impossible to estimate by non-Bayesian algorithms.

4.2 Example 2: A MODIS NDVI time series

Example 2 is based on 9-years' MODIS NDVI data at a forest site in Australia (Fig. 5), which has been used by Verbesselt et al. *(2010a)* to test bfast. Despite being familiar to large audiences, its true underlying seasonal and trend dynamics are unknown, except that we know that the site experienced droughts in 2001 and 2002 and was harvested in 2004. With all trees removed, the 2004 harvest should have altered both the NDVI trend and seasonality. It remains untested whether the drought effects are detectable from this time series. To run BEAST and bfast, we used a maximum changepoint number of M_{max} =10 and a minimum inter-changepoint distance of h= .5 year.

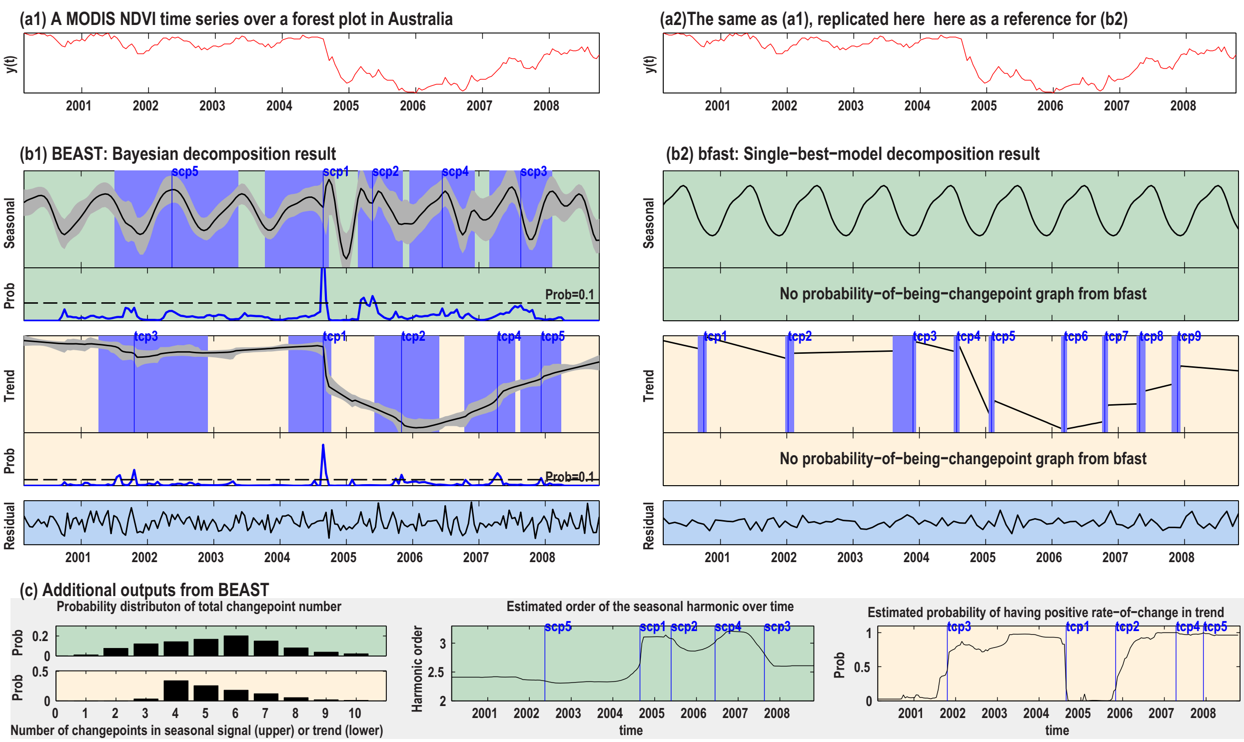

Fig. 5. Example 2: Use of a MODIS NDVI time series in the bfast R package to illustrate the use of BEAST. The true underlying seasonal and trend signals are unknown, except that we know that this site experienced droughts in 2001 and 2002 and was clear-cut in 2004. BEAST detected 5 scps and 5 tcps, uncovering not only the abrupt changes from the 2004 clear-cut but also the subtle disturbances associated with the 2001 drought. Phenological changes resulting from the 2004 clear-cut and the subsequent recovery and forest management activities were captured by a total of four scps (i.e., scp 1 to 4). For comparisons, bfast found no scp and 9 tcps detected (b2).

BEAST unveiled both the large-magnitude and subtle changepoints (Fig. 5b1). On average, it detected 5.2≈5 tcps and 5.3≈5 scps. One of the five tcps is attributed to the drought (i.e., tcp#3) and the other four attributed to the 2004 harvest and the consequent post-harvest recovery. The five scps are evidenced in the estimated seasonal trajectory, the seasonal changepoint probability graph, and the harmonic-order graph. In contrast, bfast detected no scps and 9 tcps.

BEAST estimated a more parsimonious trend than bfast (i.e., 5.2 < 9). Despite the parsimony, the BEAST trend captured a complex nonlinear dynamics (Fig. 6b1). As examples, the low-intensity stresses of the 2001 and 2002 droughts were noticeable in the trend. The effect of the 2001 drought was found more severe and was associated with tcp#3 in Fig. 5.1. The rapid recovery past the year 2006 was uncovered by BEAST as a continuous nonlinear trajectory, which contrasts with the bfast-detected discontinuous trajectory that has jumps with a browning trend after the year 2008. Another salient difference pertains to shifts in seasonality. With the 5 scps detected (Fig. 5b1), BEAST was able to capture the phenological shifts caused by the 2002 drought (scp#5), the 2004 logging (scp#1), and the post-harvest recovery (scp#2, 3, &4). In contrast, bfast detected no scp and uncovered a stable seasonal trajectory (Fig. 5b2), suggesting no phenological change before and after the harvest.

4.3 Example 3: CO2 time series at Manua Loa

Example 3 is intended to demonstrate the use of BEAST as a generic algorithm. We considered 45 years' bi-weekly atmospheric CO2 measurements from the year 1974 to 2018 at Manua Loa (Fig. 7a1). The true trend or seasonal CO2 dynamics are unknown, except that we know there was a rising trend due to human activities and there was a regime shift in the Earth system in the 1980s (*Reid et al. 2016)*, which should be reflected in the seasonal CO2 dynamics. This knowledge provides valuable information to assess the validity of the decomposition results. For both BEAST and bfast, we chose M_{max} =10 and h= 1 year.

Decomposition results of BEAST and bfast appeared visually similar, but the exact dynamics uncovered by the two differed greatly (Fig. 6). On average, BEAST detected 1 seasonal changepoint (scp) and 7 trend changepoints (tcps); bfast detected no scp and 10 tcps. One of the tcps—detected by both BEAST and bfast—occurred around the year 1977, marking a heightened increase in CO2 and coinciding with the start of China's economic reform (Fig. 7b1). This finding is the first time that the carbon footprint of an economic policy has ever been directly pinpointed in a station-based CO2 time series. Exact drivers for other tcps need close scrutiny in future studies.

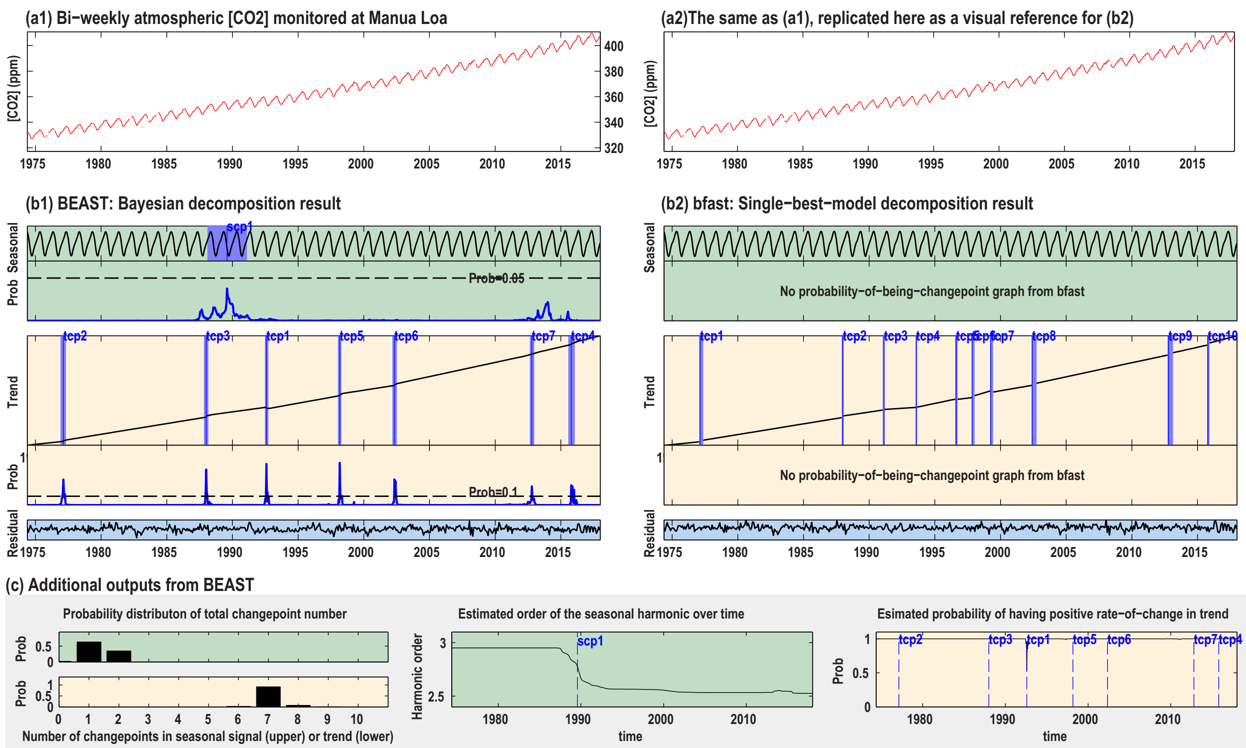

Fig. 6. Example 3: Use of 45 years' atmospheric CO2 data at Manua Loa (a1-a2) to illustrate BEAST for generic applications. On average, BEAST detected one seasonal changepoint (scp) and 7 trend changepoints (tcps). The true seasonal or trend CO2 dynamics are unknown, but the BEAST decomposition is consistent with known drivers. The CO2 trend shifted to a faster rising trajectory around the detected 1977 tcp [i.e., tcp2 in (b1)], coinciding with the end of China's Cultural Revolution and the start of its economic reform. More interestingly, the detected scp around the year 1989 (b1, top; c, middle) is consistent with the growing body of evidence that the Earth system saw a systematic regime shift in the 1980s. For comparisons, the bfast results detected no scp and 10 tcps.

The most likely seasonal changepoint (scp) detected by BEAST was found around the year 1989, an abrupt change not detected by bfast (Fig. 6b2). This scp is not too sharp a one but a gradually transient one spanning multiple years in the 1980s (Fig. 6c, middle). Its presence was indicated in the scp probability graph (Fig. 6b1). Its occurrence was also evident in the trajectory of estimated seasonal harmonic orders (Fig. 6c, middle). The detection of this 1989 scp is non-trivial, showing that the global carbon cycle was subject to a regime shift in the 1980s (Figs. 6b1 and 6c). The validity of this shift is supported by a converging body of observational and modeling evidence (*Reid et al. 2016)*. A comparison of the BEAST seasonal trajectories before and after 1989 indicates an intensified global carbon cycle over time with a strengthened carbon sink. As a rough estimate, the amplitude in seasonal CO2 variation increased from 6.35 (pre-1989) to 6.58 ppm (post-1989), a 3.6% increase. The magnitude of peak global carbon sink—estimated as the temporal derivative of the seasonal CO2 trajectories—was enhanced from 26.5 to 27.6 ppm/year. The post-1989 seasonal dynamics also showed some enhanced springtime carbon sink, an advancing in peak sink, and a slight increase in autumn carbon source, all consistent with the recognized effects of global warming on ecosystem productivity *(Piao et al. 2008*).

5. Case Studies and Results

To evaluate BEAST and further exemplify its usefulness for remote sensing applications, we conducted three case studies using either simulated or real data (Figs. 7-11). These case studies were targeted at different aspects of BEAST; each chose a differing type of strategies or reference data for algorithm assessment:

(1) Case study 1 used simulated data with true reference dynamics precisely known. The aim is to test how BEAST can uncover the true reference trend signals, an aspect critical for ecological remote sensing but seldom tested before. A secondary aim is to quantify how the performance of BEAST responds to data noises and relative magnitude of trend signals.

(2) Case study 2 used dense stacks of Landsat imagery. Ground-reference data on disturbances and changepoints were visually derived from interpretation of multisource imagery following a protocol similar to Cohen et al. (*2011*). The aim is to evaluate the ability of BEAST in detecting disturbances from high-resolution data; trend signals is not evaluated here due to the impossibility of obtaining true reference trend signals.

(3) Case study 3 used MODIS EVI data at 250-m resolution over a region where the extents and timings of two large-scale disturbance events are known. Independent reference data were obtained from aerial photos or Landsat imagery. The aim is to determine whether BEAST can help to reveal the disturbance patterns from the MODIS data and also to assess the utility of the probabilistic information derived by BEAST.

5.1 Case study 1: Simulated data

Simulated time-series data were generated by additively combining synthetic trend and seasonal signals, abrupt changes, and random noises. The trends considered were piecewise linear, with coefficients randomly sampled from a Gaussian distribution; the seasonal signals were piecewise harmonics, with the order randomly sampled between 1 and 5. The simulation was based on varying levels of data noises (2% to 20%), relative trend-to-seasonal signal strength (5% to 50%), and changepoint number (0 to 10). For each combination, we replicated 1000 times with the time-series length randomly chosen between 200 and 500, with a total of 110,000 time series generated. The use of such well-controlled data is not only appealing but also necessary for algorithm evaluation because ground-truthing is rarely available at temporal and spatial scales commensurate with the satellite data.

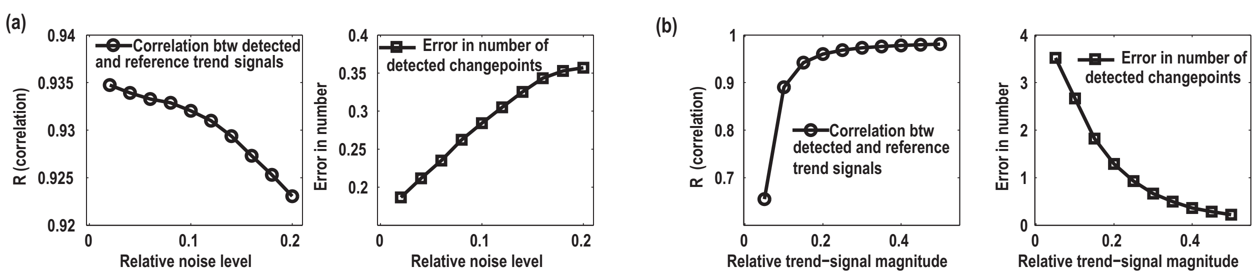

Fig. 7. Case study 1: Assessment of BEAST upon 110,000 simulated reference time series. Two performance metrics are plotted here—correlation between BEAST-detected and true trend signals and error in the number of detected changepoints. Positive errors indicate underestimates of the true changepoint numbers. Shown here are the performances of BEAST at different levels of data noises (a) and relative trend-signal magnitudes (b). Each data point plot here represents the averaging over 110,000/10=11,000 time series.

BEAST unveiled true trend signals accurately (Fig. 7). When tested upon the simulated data, the estimated trends matched the true signals closely, with a correlation coefficient averaging 0.931. Even for the noisy NDVI simulation with a noise magnitude of 20% (i.e., a signal-to-noise ratio of 5.0), BEAST could detect the true trend signals well; the correlation averaged 0.923 (Fig. 7a). In contrast, the estimation of trends showed more sensitivity to relative magnitudes of the trend to seasonal signals (Fig. 7b). For example, when the magnitudes of trends in simulated data were 5% of those of seasonal signals, the correlation between the BEAST-detected and true trends was 0.67 ($p<<0.001$); if the relative trend magnitude increased to 10%, the correlation rose to 0.89 ($p<<0.001$).

Similarly, BEAST detected changepoints reliably, irrespective of the data noise levels considered (Fig. 7a). However, when the true trend signals became weak and dwarfed by the seasonal signals, detection of trend changepoints became difficult or impossible (Fig. 7b)—a data quality problem that no algorithms can resolve. Therefore, the true changepoint numbers are increasingly underestimated as the trend signal becomes weaker. The problem with weak trends also explains the consistent underestimation pattern for all noise levels (Fig. 7a). The error depicted there at a given noise level is the average over all possible levels of relative trend magnitude; therefore, this error is contributed and dominated by the underestimation associated with those time series with weak trend signals.

5.2 Case Study 2: Dense Landsat Stack

In Case study 2, we acquired 495 images of Landsat TM5 or Landsat 7 ETM+ (WRS2 Path 24/Row 37) over the Southern US. We corrected the images radiometrically and atmospherically into surface reflectance via the LEDAPS framework and the FMask cloud masking algorithm (*Schmidt et al. 2013; Zhu et al. 2015*); we then computed NDVI and stacked the results. The number of clear-sky dates in the stack averaged 191 across the scene. To assess BEAST, we randomly sampled 200 time series across the scene. This sample was interpreted independently by three analysts to manually identify all potential changepoints to their full capacities using our GUI-based tool "trackEcoDyn" (Sect 3.5 or Fig. 8). Any inconsistency among the three was reconciled if an agreement could be reached and otherwise was simply not considered as changepoints. We used the final consensus set as ground-reference data to assess how well BEAST detects changepoints in Landsat-type images. The focus here is on evaluating changepoints rather than trend signals, due to the impossibility of obtaining true reference trend signals.

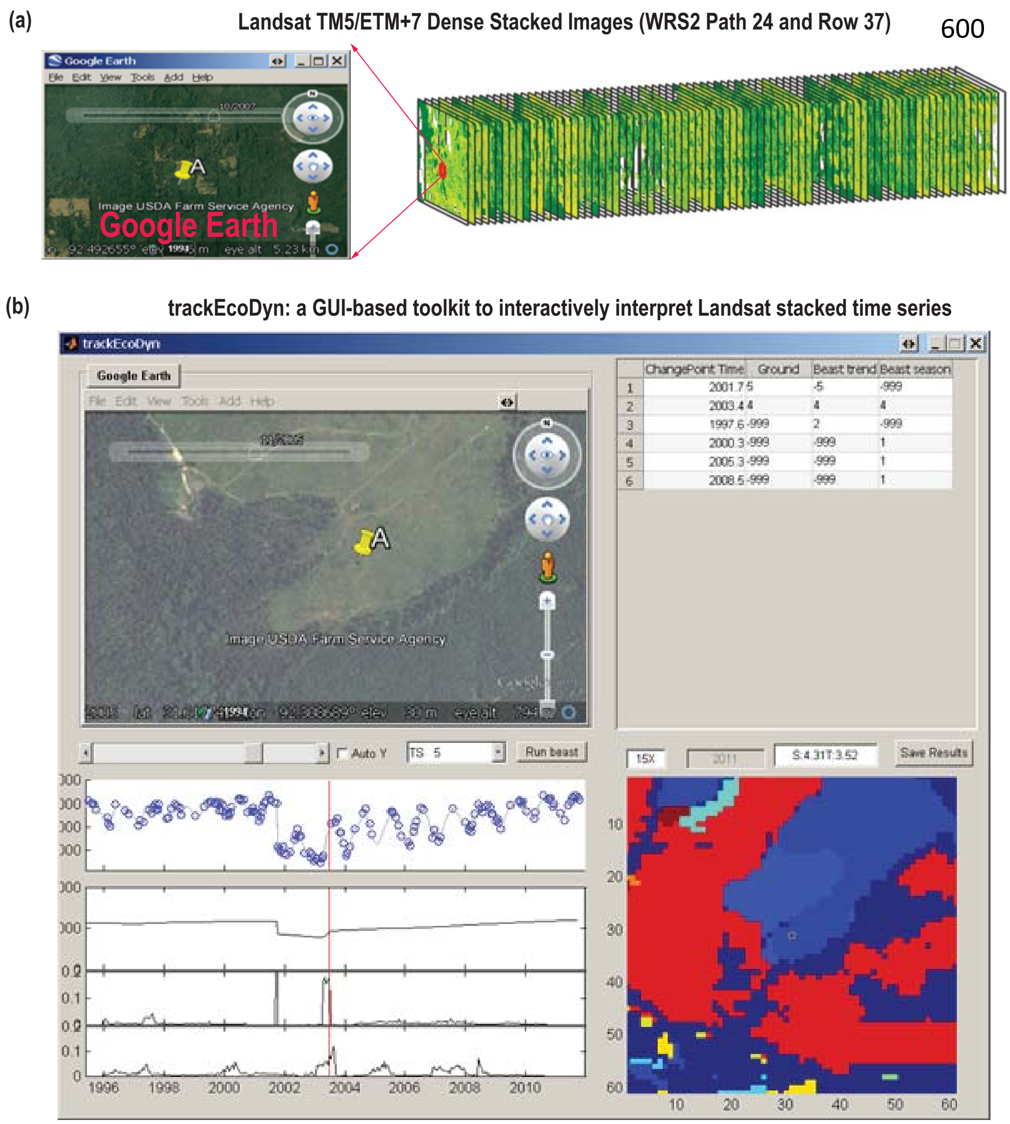

Fig. 8. Case study 2: Algorithm assessment based on dense Landsat stack of 495 scenes (WRS2 Path 24/Row 37) over the Southern US. Shown in (a) is just a subset of the full scene. (b) To collect independent data of land

changes, we developed a GUI-based toolkit "trackEcoDyn" through the mixed use of Matlab and C to interactively analyze and interpret Landsat time series. It integrates BEAST and ingests multiple external data sources. Google Earth is also synchronized automatically. This toolkit was built here to help image analysts manually and visually collect reference data for assessing the BEAST algorithm, but it can be equally applied to visually interpret or automatically analyze other spatiotemporal data.

BEAST has detected most of the changepoints in the ground-reference data, though with seemingly non-negligible omission and commission errors. In the ground reference for the 200 time series, there were a total of 368 changepoints pinpointed via visual interpretation, including 190 disturbance-type events (i.e., declining NDVI) and 178 recovery-type events (i.e., rising NDVI). These reference changepoints were resolved to individual years. As for comparisons, BEAST detected 217 disturbance-type and 197 recovery-type changepoints, all of which were resolved to the sub-monthly level. An automatic matching of the event years showed that the omission and commission errors of BEAST were 17.7% (i.e., a producer accuracy of 82.3%), and 26.8% (i.e., a user accuracy of 73.2%). Examined for disturbance-type events only, the omission and commission errors are 9.5% (i.e., 18/190) and 20.7% (i.e., 45/217); for recovery-type events, the omission and commission errors are 26.6% and 33.5%. It appears that BEAST had larger commission errors than omission errors.

The assessment metrics reported above, especially the commission errors, are underestimates of the true capabilities and accuracies of BEAST. As a further evaluation, we manually paired and compared the BEAST results with the ground-reference data. The 18 omission errors out of the 190 disturbance-type reference changepoints, as labelled by the automatic matching, were not always true errors. Six of the 18 were not true omissions because BEAST correctly detected them to the sub-monthly level in years different from but immediately adjacent to the years in the ground references; the BEAST-detected timings were more accurate. Likewise, the commission errors reported above are not always true algorithmic errors (Fig. 9). At least three of the 45 disturbance-type changepoints labeled as commission errors are associated with data anomaly due to cloud contamination: we expected BEAST to detect these anomalies as changepoints, although they are not ecologically meaningful. More importantly, many other commission errors are unlikely to be true errors because BEAST detected all kinds of changepoints of varying intensity but the ground-reference data included only those visually conspicuous to the analysts. For example, in the recovery trajectory of a forest plot (Fig. 9), BEAST identified three changepoints that demarcated contrasting succession stages of the recovery, but the analysts could pinpoint only one changepoint.

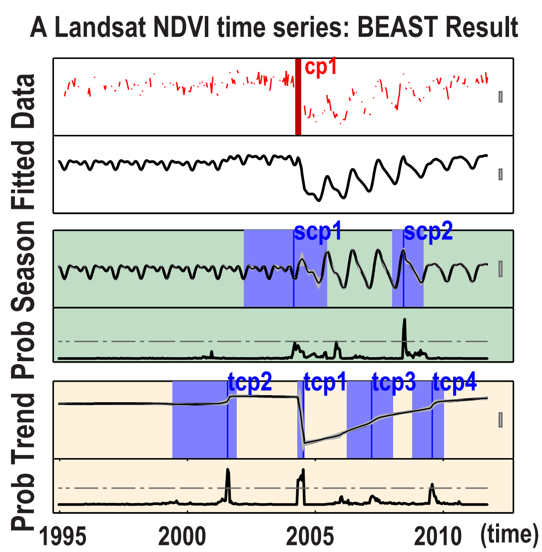

Fig. 9. Case study 2: BEAST decomposition of a Landsat time series at a forest pixel as an example to illustrate the artificial discrepancy in changepoint detection between BEAST and visual interpretation. BEAST found all types of changepoints—a total of four tcps, being abrupt or gradual. However, when interpreting visually, the three experts pinpointed only one changepoint (i.e., cp1--the sudden NDVI drop to forest logging). Hence, commission errors of BEAST in detecting changepoints in reference to the visually-interpreted ground reference are not always true errors. Illustrated also here is the capability of BEAST for filling data gaps in the Landsat time series.

5.3 Case Study 3: MODIS EVI

The study area chosen here is part of the Shawnee State Forest, Ohio, USA (Fig. 12a). This region has been disturbed or frequently managed across various scales. In particular, the region was struck by an ice storm in Feb 2003 and a fire in April 2009—Ohio's largest recorded wildfire (Eidenshink et al. 2007). The data we examined were EVI data from MODIS's 16-day L3 data at 250-m resolution from year 2002 to 2014. To better characterize the disturbance patterns of the ice storm and fire, we also complied Landsat-5 Thematic Mapper images at 30-m spatial resolution collected before or after the disturbances. The perimeters of the regions disturbed by the ice storm and fire were manually delineated from aerial photos (i.e., white polygons in Fig 10.b). We also calculated the Normalized Burn Ratio from the pre-and post-fire Landsat images and took the difference—dNBR—as an indicator of burn severity. These high spatial-resolution images and information provide independent reference data to evaluate and interpret the MODIS results.

When applied to the MODIS data, BEAST uncovered spatiotemporal patterns of vegetation dynamics that were consistent with the known disturbance history (Fig. 10). In particular, the two major landscape-scale disturbances, the 2003 ice storm and 2009 fire, were detected successfully. The estimated disturbance timings matched closely with the true dates (Fig. 10c). The BEAST-detected locations and extents of the disturbances closely resembled those patterns revealed by the post-disturbance Landsat images as well as those manually derived from independent high-resolution imagery. Within the perimeter of the disturbed regions, BEAST depicted the spatial heterogeneity in disturbance magnitude. One such output is the probability of being a true changepoint. For example, when tested for the burned region within the 2009 fire rim, the BEAST-estimated changepoint probability strongly correlated with the independent Landsat-based dNBR ($r=0.66$, $n=288$, p-value $<<0.001$) (Fig. 11a). Such probabilistic outputs enable BEAST to characterize not only those large-magnitude disturbances but also all other disturbances over a continuous range of magnitude. The probabilistic results should be more informative and practically useful than the mere reporting of binary outcomes about occurrence or not.

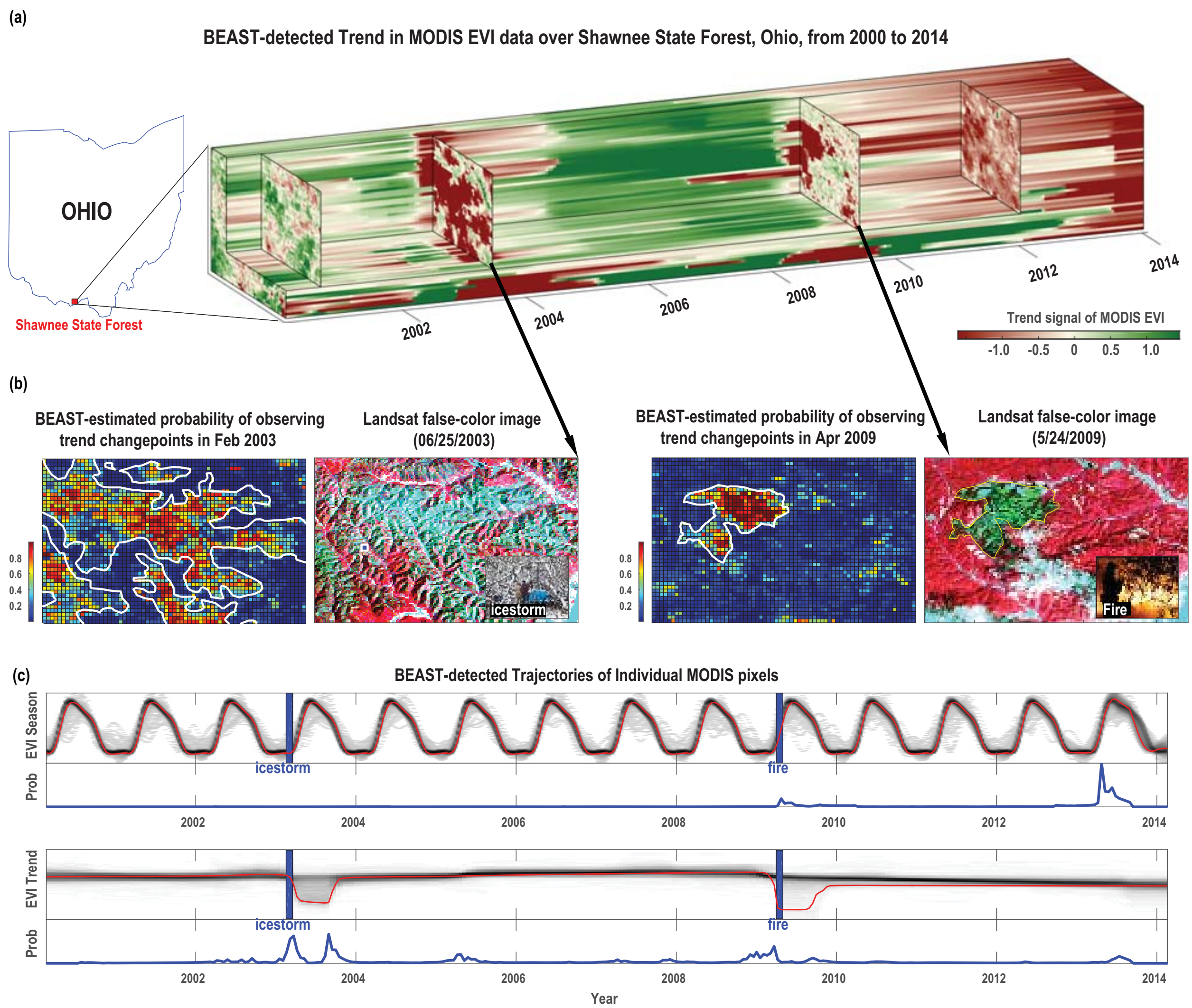

Fig. 10. Case study 3: MODIS EVI data from 2001 to 2014 over the Shawnee State Forest, Ohio where the forests have been disturbed by many natural events and anthropogenic activities, for example, including an ice storm in February 2003 and a fire in April 2009. Shown in (a) is a 3D volumetric view of the spatiotemporal patterns of forest trend dynamics detected by BEAST; therein, brown areas indicate spatiotemporal locations where the forest ecosystem is of low vitality. The ice storm and fire disturbance events are singled out to illustrate the detected probability of changepoint occurrences. (b) Post-disturbance Landsat images together with manually-derived disturbed regions (i.e., white polygons) are accompanied as visual references to assess the spatial patterns of MODIS-based disturbances. Shown in (c) are density plots to depict individual trajectories of detected seasonal and

trend dynamics for all the MODIS pixels: the darker the color, the higher the trajectory density. Overlaid on the density plots are red solid curves to indicate the mean trajectories averaged over all the pixels. The “Prob” subplots show the mean changepoint-occurrence probabilities averaged over all the pixels (blue curves). The true timings of the ice storm and fire are indicated by vertical blue bars.

Ecologically speaking, the seasonal and trend dynamics uncovered by BEAST were compatible with true vegetation responses to ice storm and fire. Sudden drops in NDVI were detected by BEAST at the starts of the ice storm and fire, followed by rapid continuous transient transitions for post-disturbance recovery. When uncovering seasonal dynamics, BEAST detected no seasonal changepoints for the 2003 ice storm but some changepoints over part of the region for the 2009 fire (Fig. 10c). This result is corroborated by the contrasting damage severity of the two disturbances: the ice storm caused branch breakage and infrequent treefall; the fire was more destructive and sometimes stand-replacing. The severe fire damages shifted the phenology at some disturbed pixels (Fig. 10c).

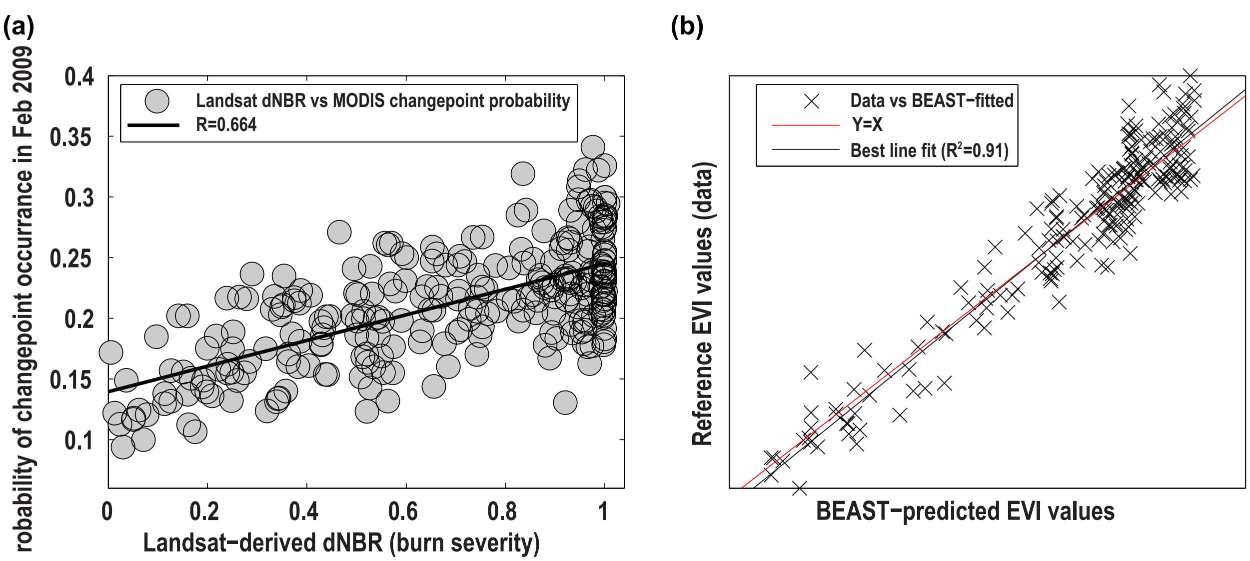

Fig. 11. Case study 3: (a) BEAST-estimated probabilities of changepoints occurring in Feb 2009 for 288 MODIS pixels within the fringe of the wildfire are correlated significantly with independently

derived Landsat burn severity dNBR. (b) True reference EVI values vs. BEAST-estimated values for a selected MODIS time series based on the leave-one-out cross-validation.

From a regression standpoint, BEAST fitted the MODIS time series well and made accurate predictions within time-series data gaps. Averaged over all the pixels of the region, the correlation between the actual and fitted time series was 0.943 (Fig. 11a). The strong correlation highlights the predictive power of BEAST and supports its potential use as a gap-filling method. This capability is further confirmed by cross-validation. For example, in Fig. 11b, a leave-out-out cross-validation tested upon a MODIS time series showed that the estimated missing values matched the true reference values with high fidelity (R2=0.91).

6. Discussion

Leveraging the rapid growth of satellite time-series data to uncover the vagaries of landscape change is an area seeing a surge in algorithm development *(Cohen et al. 2017; Wulder et al. 2012; Zhu 2017)*. This advance leads to many new ecosystem dynamics products but, at the same time, opens new research gaps *(Cohen et al. 2017)*. The problem examined here is how to improve algorithmic robustness and characterize algorithmic uncertainty. Cohen et al. (2017) stressed this problem by noticing considerable discrepancies among seven algorithms, with a pixel-level agreement of only 0.2% in detected disturbances and a 1500% difference in estimated disturbance areas. Ensemble learning is touted as a remedy *(Cohen et al. 2018; Healey et al. 2018)* but was only partially explored. The development of BEAST helps to bridge the gap by offering a generic tool that incorporates an ensemble of models into time-series analysis. Our case studies provide experimental evidence on the efficacy of BEAST in detecting abrupt change, seasonality, and trend.

6.1 What BEAST can and can't do?

BEAST is a Bayesian regression method to isolate periodic and trend signals from a time series and to pinpoint abrupt shifts in the two isolated signals. It is intended primarily for trend analysis and changepoint detection, targeted at questions like those elicited in Section 2. Are there any increasing or decreasing trends, any changepoints, or any phenological shifts? What is the rate of change at a given time? Is the detected greening trend real? What is the probability of observing 3 changepoints between 2001 and 2015, or both a seasonal and a tend changepoint in August 2009? Interpretations, connotations, validation of the answers to these questions are context-specific, depending on the goals of applications *(Cohen et al. 2017; Wulder et al. 2012)*. BEAST is applicable to any real-valued variables, such as LAI, temperature, soil moisture, gravity, and other biological or even socioeconomic data; therefore, translating the results into insights is contingent on the natures of the data and problems at hand.

BEAST detects temporal dynamics but can't attribute drivers. Is a detected greening due to a warming climate, post-disturbance recovery, or reduced grazing? Is a detected forest loss caused by fire, insect, hurricane, logging, or urbanization? Is a shift in phenology due to climate change, altered management, crop rotation, succession, or land conversion? BEAST can't answer these attribution questions directly. To a lesser extent, even a simple question like whether a detected NDVI drop corresponds to a forest or grass loss can't be answered unless we know that the site observed is a forest or a grassland. To answer the questions, we need to combine BEAST further with other algorithms and ancillary information. For example, BEAST can be used to map extents, timings, and severity of gypsy moth infestation if we know that it is the disturbance agent. For applications on mapping both changes and drivers, BEAST should be interfaced with a classifier that is trained empirically to relate BEAST-derived metrics with land-change classes or causative agents *(Cohen et al. 2017)*. The training and validation of the classifier can follow the good practices recommended for mapping land cover *(McRoberts 2011; Olofsson et al. 2014; Olofsson et al. 2013*).

We envision that BEAST is particularly useful for three related but subtly different areas in remote sensing. One area concerns ecosystem dynamics; the aim is to track vegetation changes over time and understand their drivers. Current use of satellite data for such purposes is fraught with debates, for example, regarding how climate change has affected long-term vegetation growth, how global warming alters land surface phenology, and how extreme weather impacts forests. The second area pertains to mapping land disturbances and land conversion over time. Despite recent advances, the existing algorithms diverged greatly and produced inconsistent disturbance maps. With its proven analytical capability, BEAST should be able to provide new perspectives into these two areas. A third area of applications is to apply BEAST to fill temporal gaps in satellite data. BEAST can fit a nonlinear curve to data with gaps and estimate the missing values.

6.2 BEAST vs. existing methods

Numerous time-series methods have been introduced for applications in remote sensing or other disciplines *(Brockwell and Davis 2016; Hamilton 1994; Zhu 2017)*. Many of them were developed under various names, such as trend analysis, seasonal decomposition, changepoint or breakpoint analysis, signal segmentation, regime shift detector, anomaly detection, and structural change *(Brockwell and Davis 2016; Denison 2002; Hamilton 1994; Harvey 1990)*. Rigorous comparisons of BEAST to the existing methods are complicated by the sheer number and diversity of algorithms and, to some extent, a lack of consensus on nomenclature. For ease of comparisons, our discussion below focuses only on two aspects of BEAST: trend analysis and changepoint detection.

BEAST extends conventional trend analyses in several ways. The majority of existing analyses—based mostly on NDVI—examined linear trends by fitting a global line to the data without considering seasonality, if any *(Brando et al. 2010; Myneni et al. 1997; Piao et al. 2006)*. BEAST applies flexible basis functions to fit both linear and nonlinear trends and disentangle trends from seasonality. Some recent trend analysis methods attempted to address nonlinearity using piecewise linear models, but with a prescribed number of changepoints *(Chen et al. 2014; Wang et al. 2011)*. A landmark study in this category is Wang et al. *(2011)* that applied a piecewise linear model with one changepoint to

AVHRR data for new insights into climate-ecosystem interactions. Indeed, it is the one-changepoint model of Wang et al. *(2011)* that motivated our algorithm development. BEAST goes beyond by making the changepoint number an unknown parameter and letting the data tell what it is. Statistically speaking, existing analyses were mostly based on frequentist methods, seeking only the "best" model; BEAST employs Bayesian model averaging, embracing all candidate models rather than selecting just one.

What distinguishes BEAST most from the existing trend analysis methods is its capability of inferring nonlinear dynamics. BEAST provides a universal approximator of any arbitrarily complex trends. In contrast, most existing methods derive only linear or piecewise linear trends *(Wang et al. 2011)*. True drivers of ecosystem dynamics are unlikely to be purely linear or piecewise-linear over time but rather complex and nonlinear. For example, plant successional stages are known to largely follow a nonlinear recovery trajectory *(Burkett et al. 2005)*. Long-term climate trends are confirmed to be inherently nonlinear *(Franzke 2014)*. With its better approximation power, BEAST is more likely to find these true nonlinear trends than the existing methods. Improved fitting of trends can help with changepoint detection because errors in fitting trends may be translated into errors in changepoint detection.

For changepoint detection, existing algorithms are mostly heuristically-based, involving the testing or optimization of criteria *(Cohen et al. 2017; Zhu 2017)*. For example, several well-known algorithms, such as LandTrendr, VCT, and CCDC, rely on locally-based heuristic rules by checking if some deviation metrics meet certain pre-set thresholds. They often iteratively analyze the time series piece by piece or step by step *(Huang et al. 2010; Kennedy et al. 2010; Zhu et al. 2012)*. In contrast, BEAST is a parametric regression method. It does not require any threshold testing or criterion optimization but, instead, fits a global model to decompose the whole time series in one step and uncover changepoint , trend, and seasonality altogether. As another key difference, many existing algorithms are hard detectors in that their outputs are limited to either 1 or 0—a changepoint or not; BEAST is a soft/fuzzy detector capable of estimating the occurrence probability of changepoints over time *(Cohen et al. 2017; Huang et al. 2010)*. This difference is analogous to that between hard and soft/fuzzy classifiers. To our knowledge, BEAST is the first fuzzy time-series algorithm ever developed for remote sensing applications.

Of the existing algorithms, bfast is the one that shares the most commonality with BEAST *(Verbesselt et al. 2010a; Verbesselt et al. 2010b)*. The two have almost identical parametric models except that bfast fixes the seasonal harmonic order to a constant of 3 or other constants but BEAST treats it as an unknown to be estimated for individual seasonal segments. This difference seems minor but has substantial effects, partially explaining why BEAST detected more seasonal changepoints. The varying harmonic order gives a flexible representation of seasonality and helps BEAST to capture subtle variations difficult to represent by a fixed-order seasonal model (Fig. 6b1 vs Fig. 6b2). The BEAST or bfast model is additive. If parts of the true seasonal dynamics are not captured by the seasonal model S(t), these seasonal parts will be squeezed into the trend model T(t) or noises. As a result, seasonal abrupt changes may be confused with trend changepoints. Likewise, parts of the true tend, if not adequately captured by T(t), will leak to contaminate the estimation of seasonality. This is why we strived to make BEAST a flexible approximator of any arbitrary trends. In short, reliable detection of changepoints, especially those subtle ones, requires the accurate modeling of not only the trend or seasonal component alone but both altogether.

Another key difference between BEAST and bfast or other algorithms lies in parameter estimation. Bfast treats the model parameters as unknown constants. BEAST treats them as random variables; its inferential goal is not only the best values of the parameters--number and timing of changepoints, harmonic orders, and coefficients--but also their probability distributions. BEAST tells not only a detection of 3 tcp but also a 71% probability for having 3 tcps, a 20% for 2 tcps, or a 5% for 1 tcp. Put differently, bfast seeks a single best model but BEAST embraces numerous models in terms of a probability distribution over the model space. This is likened to the difference between CART and Random Forests *(Friedman et al. 2001)*. Bfast is like CART that finds only one decision tree; BEAST is like Random Forests that uses many trees. As shown in the ecology and machine learning literature, Random Forests is less likely to overfit and more likely to find ecologically-meaningful relationships than does CART *(Breiman 2001a)*. Similarly, as an ensemble modeling algorithm, BEAST tends to generate more flexible and interpretable results.

6.3 Ensemble learning: One plus one is more than two

BEAST is based on ensemble learning. Most other algorithms for satellite time-series analysis are not, except two recent algorithms in Cohen et al. *(2018)* and Healey et al. *(2018)*. But BEAST and these two are not comparable. BEAST is a regression method for time-series decomposition wherein ensemble learning is internalized into the Bayesian formulation. In contrast, the other two algorithms are some classifiers that ingest the pool of multiple model outputs as predictors to classify disturbance agents. More generally, ensemble learning comes in many other fashions, but the core is to combine many models or algorithms into a better one *(Friedman et al. 2001)*. Experimental evidence is unequivocal about the effectiveness and superiority of ensemble learning, compared to the single-best-model paradigm *(Friedman et al. 2001)*. Recent years also saw a growing urge for better leveraging the paradigm of ensemble modeling or multi-model inference—a voice that is being heard in many scientific disciplines and is reinforced again here.

Why does ensemble learning help with our time-series analysis? The answer lies in a familiar example: the IPCC relies on many climate models instead of any single model to augment confidence in climate prediction *(Solomon 2007)*. What is implicit here is George Box's aphorism "all models are wrong", a creed that, if held, may help little with practical modeling but, if ignored, can engender unwarranted epistemological debate *(Beven 2010)*. All remote sensing models, including radiative transfer models in operational use and our BEAST algorithm, are also wrong in the sense that they are always simplifications and approximations of the true processes *(Schowengerdt 2006)*. This is connoted by the fact that remote sensing is fraught with the use of different algorithms or models to decipher the same linkage or functional relationship *(Cohen et al. 2017)*.

Since all models are wrong, the consideration of many models as in BEAST can reduce the chance of deviating too far from the unknown truth, compared to the choice of just a "best" model. When ranking models in terms of usefulness metrics such as AIC and BIC, the "best" ranked model is not guaranteed to be closer to the truth than other models of lower ranks *(Shmueli 2010; Zhao et al. 2013)*. As a rule of thumb, if two models have an AIC difference of <2.0, there is no strong evidence that one should be favored over the other *(Burnham and Anderson 2003)*. Similar rules exist for other model selection criteria. In the current context, each candidate model is uniquely specified by the model structure parameters, such as numbers and timings of changepoints. The entire model space may comprise quadrillions of candidates or more. Not surprisingly, there can be numerous competing models (e.g., millions) that are statistically indistinguishable from the "best" model in terms of AIC or BIC, a phenomenon called model equalfinality *(Beven 2010)*. The lack of strong statistical power to discriminate some models against others makes it safer to use the many models than a "best" model.

Even if not all models are wrong and the true model is in the space of candidate models, ensemble learning can still be more robust than single-best-model algorithms *(Friedman et al. 2001; Wintle et al. 2003; Zhao et al. 2013)*. Even in a simple scenario of linear regression, Zhao et al. (2013) showed that many single-best-model regression procedures failed to recover the true linear model. More generally, no model selection criterion guarantees finding the true model. The failure is primarily due to two factors. First, the space of candidate models is so enormous that optimization may be trapped at local minima, failing to find the real optimal model. Second, even if the real optimal model is luckily found, it may still not be the true model: optimality is not equivalent to truth. The true model may have worse AIC or BIC values than other models, for example, due to data noises or multicollinearity *(Friedman et al. 2001; Grossman et al. 1996)*. These difficulties justify the use of ensemble learning even if we can correctly parameterize the true model, let alone when we can't.

How can BEAST uncover arbitrary nonlinear vegetation dynamics, given that individual trend models are just piecewise-linear? A rigorous mathematical answer to this is beyond the current scope. Intuitively speaking, the averaging of many piecewise irregular functions will smooth out the irregularity and mold them into a more flexible function *(Friedman et al. 2001)*. Indeed, for almost all practical applications, the use of ensemble averaging is more flexible in fitting nonlinear functions than any individual models. A familiar example again is Random Forests: each tree is a discontinuous partition-based function, but the averaging of many trees is able to approximate complex functions *(Cogger 2010)*. This is what we call here as "the making of a stronger model from many weak models" or "one plus one is more than two" *(Friedman et al. 2001)*. It is this property that enables BEAST to detect realistic vegetation trend dynamics.

6.4 Bayesian statistical modeling: To explain or predict?

BEAST fits a Bayesian regression model or a function curve to match the observed time series, with time as the independent variable. In this regard, BEAST is the same as the many existing statistical models relating remote sensing predictors to biophysical variables. However, their modeling purposes are not the same *(Shmueli 2010)*. Most of the statistical models are calibrated to minimize differences between fitted and observed land variables and then estimate the variables for new data unused in the calibration, that is, to predict *(Breiman 2001b; Zhao et al. 2018)*. The purpose of BEAST is not to minimize the fitted-vs-observed differences or predict values at a new time but to identify the right mechanisms underlying the observed time series, that is, to explain *(Dashti et al. 2019; Thomas et al. 2018).* Models that predict well may not explain well, and vice versa *(Shmueli 2010)*. Black-box machine learning models are such examples. Although the explain-vs-predict divide is not dichotomous, future studies will garner more successes in designing robust statistical algorithms to obtain useful vegetation dynamics information if we pay attention to the explaining nature of such modeling efforts.

Bayesian modeling is of great value in mining complex data to find meaningful and interpretable relationships *(Finley et al. 2008).* A confirmatory example is our BEAST algorithm. Bayesian inference is powerful particularly because of its explicit consideration of various sources of uncertainty *(Denison 2002; Kennedy and O'Hagan 2001)*. The time-series problem at hand is a difficult one fraught with uncertainties, due to some apparent conflict. On one hand, model misspecification is inevitable (in theory, true vegetation dynamics are unlikely piecewise-linear or –harmonic); on the other hand, our aim is to use the misspecified model to capture the true vegetation dynamics (the model should explain well). This conflict will be subtly translated into uncertainties in the model parameters and structure. A full characterization of these uncertainties is practically impossible with the conventional single-best-model paradigm, especially because it ignores model uncertainty *(Beven 2010; Kennedy and O'Hagan 2001)*. These uncertainties, however, can be formalized and treated rigorously and systematically by the Bayesian paradigm.

Despite the exceptional power of Bayesian modeling, its use in remote sensing remains limited. Many factors contribute to this. Philosophically, the "subjective reasoning" label of Bayesian inference may deter many researchers from considering it seriously *(Denison 2002)*: who wants to sound subjective in the science enterprise? This concern is unwarranted, given the rising acceptance of Bayesian statistics in essentially all fields after decades of philosophical debate *(Ellison 2004; Friedman et al. 2001)*. Even if its utility is realized, there is habitual inertia to overcome because conventional statistical methods still find the dominant use. Moreover, the use of Bayesian methods is often hampered more by practical factors, such as a dearth of easy-to-use Bayesian statistical software, the unfamiliarity of these methods to the larger community, the inherent complexity of

Bayesian modeling, a lack of formal training in Bayesian statistics, and often enough, daunting computation costs of Bayesian methods. Nonetheless, we hope that the demonstrated value of BEAST provides an impetus to encourage future remote sensing applications of Bayesian techniques. We envision that Bayesian techniques are particularly appealing in cases that data are complex or noisy, characterization of uncertainty is pivotal, computation is not constraining, and the modeling purpose is to explain (e.g., uncover the truth or test theories) rather than predict.

6.5 How to validate time-series decomposition algorithms

What should be validated or evaluated? The goal is to quantify the degrees to which estimated dynamics and changepoints represent the truth. Because changepoints are parts of trend/seasonal signals, validation of changepoints underpins that of trend or seasonal dynamics. In particular, validation of changepoints should cover all the model structure parameters—numbers and timings of changepoints, and harmonic orders. Changepoints here refer to abrupt shifts in both seasonality and trend, be positive or negative in direction, that span a continuous spectrum of magnitude. The changepoints referred to here are consistent with those in Wang et al. *(2011)* and Browning et al. *(2017)*. This range goes beyond the consideration of only large-magnitude disturbances (e.g., land conversion, forest loss, and fire) that are characteristic of many Landsat-based algorithms *(Cohen et al. 2017; Huang et al. 2010)*. Estimated trend or seasonal dynamics, such as the sign and the rate of change, should be compared to true dynamics.

A full evaluation of time-series algorithms such as bfast and BEAST is critical but, often enough, difficult due to a lack of ground-truth *(Lu et al. 2004)*. Take bfast as an example. In the MODIS data of Example 2, only the 2004 trend changepoint (tcp) was originally used to test bfast (Fig. 6) because the 2014 harvest was known precisely. The other tcps and scps were not assessed yet and are hard to test (Fig. 6b2). These tests are about whether the bfast-detected changepoints are real or artificial. A more difficult test is about how many true tcps and scps have been detected by bfast.

An even more difficult task is to test the veracity of the estimated trend or seasonal signal. (In Fig. 6b2, is the post-harvest browning after tcp#9 a real trend or an algorithmic artifact?) These validation questions are difficult, if not impossible, to answer. The same difficulties apply to validation of BEAST. Similar difficulties were also noted in Browning et al. *(2017)*, which, though considering only 15 MODIS time series, is the first study to assess bfast with field-based vegetation dynamics data. But even for that study, the validation was partial, limited to the testing of user accuracy for changepoints.

The aforementioned difficulties are manifested in other fashions. Current controversies surrounding satellite-derived ecosystem dynamics could have been safely dismissed if the associated time-series analyses had been validated. Even a seemingly simple question, like "is the NDVI trend a greening or browning?", has been debated and largely unverifiable. Such issues are not confined to remote sensing. Other fields, such as statistics, ecology, and econometrics, are also fraught with time series analyses of the same nature as ours. We are unaware of any studies in these fields that conducted full validation of changepoint algorithms when tested upon real-world data. For example, many algorithms were applied to the river flow data of the Nile at Aswan (Balke 1993; Betken 2017; Denison 2002; Wu and Zhao 2007). Most of them gave similar but essentially different decompositions. Even for this well-studied data, it is hard to test whether the algorithms uncovered the true river flow dynamics.

Because we normally have no access to all the ground-truth needed, we recommend eight practical strategies for algorithm validation. First, a simple yet powerful strategy is to validate algorithms against synthetic data. If an algorithm cannot recover the known true dynamics from the synthetic data to which it is tailored, it unlikely applies well to real data. This test is the first filter that a useful algorithm must pass. Use of synthetic data also permits full assessments under various conditions (e.g., different noise levels) (Section 5.1). Second, algorithms can be validated qualitatively (not equal to "subjectively") with respect to some general known patterns. This is another filter that useful algorithms must pass. If they detect a declining trend in the air CO2 data, a browning for forest recovery, or no changepoints for a frequently-disturbed region, there must be some problems with the algorithms. Third, validation can be done using well-established knowledge, such as ecological principles and empirical evidence. One example is the confirmation of the 2004 scp in the MODIS time series because forest clearing is known to change phenology. Another is the 1980s regime shift in the air CO2 time series. Passing this test will enhance users' confidence in the algorithm. Fourth, as a relative evaluation, an algorithm can be compared to other algorithms. Fifth, cross-validation is another effective strategy, especially for those algorithms that apply parametric models to approximate time series.

Sixth, validation can be done using known individual events (e.g., disturbance or land cover change). One example is the evaluation of bfast upon the harvest/planting years over a region in Australia. Our test of BEAST for the fire and ice storm events in Ohio was another example. Seventh, validation can be done using reference data derived from independent sources. One example is through photo-interpretation with TimeSync; another example is our tool trackEcoDyn, which is functionally similar to TimeSync. As a caveat, such reference data are subject to errors and uncertainties *(Cohen et al. 2010)*, as explained in Figure 11. Eighth, validation can be done using proxy data of all kinds. One example is the use of climate variables and field-based vegetation composition and biomass in Browning et al. *(2011)* to assess NDVI time series. Another example is our use of dNBR to assess the BEAST-derived changepoint probability. As a third example, to test if surface evaporation sees an abrupt change at a time, we can check air temperature or moisture as proxies.

Overall, none of these validation strategies is complete and perfect. A compromise is to rely on as many strategies as possible. Indeed, to test BEAST, we employed all the eight strategies, each emphasizing a differing aspect of BEAST. But in many existing satellite time-series analyses, the practice was to focus only on one aspect of the algorithms (e.g., large-magnitude changepoint only, trend only, or phenology only). Obviously, a comprehensive strategy that embraces more aspects of the algorithms should be preferred because the various components of time-series decomposition do not stand on their own but rather are linked: any errors in one component will be leaked to degrade the estimation of others. Without such comprehensive evaluations, it becomes inevitable that ecological interpretations of satellite time-series decomposition are laden with inconsistency or controversies.

6.6 Caveats and future research

Several caveats are noted. First, BEAST detects anomalies and trends but doesn't attribute the drivers. If data are contaminated by spurious errors (e.g., clouds) or systematic biases (e.g., gradual sensor degradation) *(Wang et al. 2012)*, these outliers and drifts can be misconstrued as true signals. To reduce such commission errors, data artifacts should be removed or suppressed beforehand. Second, BEAST makes inference via Monte Carlo sampling and therefore, requiring more computation than many other algorithms *(Kennedy and O'Hagan 2001)*. Applications of BEAST to massive high-resolution data, such as the Landsat archive at the global scale, will demand daunting computation. The recommended use of BEAST is for global coarse-resolution or local high-resolution (e.g., Landsat coverage of a county).

Third, BEAST explicitly quantifies how likely each point of time is a changepoint . The resultant probability appears indicative of disturbance severity (Fig. 11) and also captures low-magnitude disturbances that may be missed by other algorithms. Interpretation of the probabilities is contingent data quality. An abrupt change will have a lower detection probability if data get noisier. All else being equal, the higher the signal-to-noise ratio, the larger the estimated probability of the same disturbance. The interpretation is also confounded by sub-pixel heterogeneity. A changepoint detected with a 5% probability at a pixel, for example, may be due to either a low-magnitude disturbance across the whole pixel or contrastively, from a severe disturbance over a small fraction of the pixel. BEAST can't distinguish the two cases. The confounding can be resolved by turning to finer-grained data *(Roy et al. 2014; Zhao et al. 2018)*.

Fourth, the scale matters. When detecting changepoints, BEAST is scale-dependent. Consider two adjacent pixels, one with a sudden NDVI drop and the other with a rise of the same magnitude at the same time. If applied separately, BEAST will detect a changepoint for each pixel. But if the two pixels are combined into one, the two abrupt changes cancel out and the changepoint disappears at the aggregated scale. This scaling effect is an inherent characteristic of all algorithms. On the contrary, we speculate that BEAST is scale-invariant when uncovering trends or seasonal dynamics. That is, applying BEAST to many pixels and then aggregating the individual detected trends should give the same overall trend as that obtained by first aggregating the individual pixels into a large pixel and then applying BEAST to the aggregated pixel. This nice property is attributed to the additive nature of general linear models *(Zhao et al. 2009)*. The scale-invariance permits the use of BEAST across scales to infer trends without introducing artificial discrepancies , thereby facilitating fusion of multi-resolution data. For applications concerning only trends not changepoints, the use of BEAST at aggregated scales will also lessen the computation needed.

Fifth, BEAST is applicable to any real-valued data. However, it is a univariate method and can't decompose multiple time series simultaneously or leverage the inter-correlatedness of the many time series (e.g., multispectral bands). Extending BEAST into a multivariate algorithm is conceptually easy but the implementation is complex—a future topic to be explored. Other extensions are also possible. Here we tested BEAST upon only dense time series to track both trend and seasonality. It can be revised to handle sparser non-periodic time series (e.g., annual Landsat data with one observation per year) by simply suppressing the seasonal component in its formulation. BEAST can also be extended to handle data collected at irregular time intervals or data with duplicate measurements at a single time. As an unsupervised decomposition algorithm, BEAST can't classify disturbance agents *(Kennedy et al. 2015)*; therefore, another extension is to embed a supervised classifier into BEAST for simultaneously detecting changepoints and classifying disturbance types.

Last, we highlighted the unique features of BEAST but our intent is not to favor or discriminate one algorithm against others. All the algorithms have their own niches and offer different perspectives. Algorithmically speaking, there is no panacea for inferring true dynamics from noisy data *(Breiman 2001b)*. The validity of the diverse or conflicting perspectives, therefore, needs to be judged based on domain-specific knowledge and high-fidelity ground-truthing. Because BEAST is the first ensemble-based fuzzy time series decomposition algorithm ever developed for remote sensing applications and also because it is able to recover complex dynamics and characterize various types of uncertainty, its use can engender new insights not obtainable by other algorithms. Future studies may further test the utility of BEAST for various data, problems, and geographic regions. One example is the analysis of AHVRR or MODIS data to detect disturbances and nonlinear long-term dynamics and determine how ecosystems have been driven by climate change and human activities, an area still fraught with many conflicting findings. Overall, BEAST serves a useful tool to derive observational information from satellite data, as a way to complement field surveys, controlled experiments, and computer models in quantifying ecosystem responses to environmental changes.

7. Summary

We presented a Bayesian algorithm—BEAST—for decomposition of time series into three contrasting components: abrupt change, periodic change, and trend. BEAST helps to leverage the increasing availability of multisource satellite time-series data for detecting land disturbances and tracking nonlinear ecosystem dynamics. Compared to many existing algorithms, BEAST explicitly addresses model uncertainties via ensemble learning, thereby alleviating inter-algorithm inconsistencies to some extent. Such inconsistencies were widely recognized and, if not addressed, would result in diverging or conflicting interpretations of the same data. Conceptually, BEAST combines many individual weak models into a better model via Bayesian model averaging. Mathematically, BEAST is rigorously formulated, with its key equations being analytically tractable. Practically, BEAST can estimate probabilities of changepoint occurrence, detect not only large but also low-magnitude disturbances, and uncover complex nonlinear trend dynamics, all of which are difficult to obtain by single-best-model algorithms. BEAST is generically applicable to not only remote sensing data but other environmental, ecological, or socioeconomic time-series data. Our initial experiments confirm the utility of BEAST. We envision that its use will offer new satellite-based insights into patterns and drivers of ecosystem dynamics.

Appendix A.

Here we described more on the specification of the prior $\pi(\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2, \mathrm{M}) = \pi(\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2|\mathrm{M})\pi(\mathrm{M})$ for BEAST.

First, we chose a normal-inverse Gamma distribution as the prior of model coefficients $\boldsymbol{\beta}_{\mathrm{M}}$ and variance σ^2 conditional on model configuration M:

$$\boldsymbol{\pi}\left(\boldsymbol{\beta}_{\mathbf{M}}, \boldsymbol{\sigma}^2 \middle| \mathbf{M}\right) = \boldsymbol{\pi}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_{\mathbf{M}} \middle| \boldsymbol{\sigma}^2, \mathbf{M}\right) \boldsymbol{\pi}_{\boldsymbol{\sigma}^2}(\boldsymbol{\sigma}^2) = \mathbb{N}\left(\boldsymbol{\beta}_{\mathbf{M}}; \mathbf{0}_{\mathbf{M}}, \boldsymbol{\sigma}^2 \boldsymbol{v} \mathbf{I}_{\mathbf{M}}\right) \cdot \mathbb{IG}(\boldsymbol{\sigma}^2; \underline{\boldsymbol{a}}, \underline{\boldsymbol{b}}).$$

where the conditional prior $\pi_{\beta}\left(\beta_{\mathrm{M}} \middle| \sigma^2, \mathrm{M}\right)$ is a Gaussian distribution $\mathbb{N}\left(\beta_{\mathrm{M}}; 0_{\mathrm{M}}, \sigma^2 v \mathrm{I}_{\mathrm{M}}\right)$; the prior π_{σ^2} is an inverse-gamma distribution $\mathbb{IG}(\sigma^2; \cdot, \cdot)$ that is independent of the model configuration M and is specified by two scalar hyperparameters $\underline{a}$ and $\underline{b}$. To parameterize the Gaussian prior $\pi_{\beta}(\cdot)$, we set its prior mean to zeros 0_{M} , a justifiable choice if the covariates are centered beforehand; the prior covariance we choseK is the ridge prior $\sigma^2 v \mathrm{I}_{\mathrm{M}}$. The subscript "M" in the zero-mean vector 0_{M} and the identity matrix I_{M} indicates that their dimensions depend on the model structure M. Moreover, in the prior covariance for $\pi_{\beta}(\cdot)$, v is a scalar hyperparameter. Judicious values for v are not available in advance; therefore, we also treated v as random and further assigned it an inverse-gamma prior

$\pi_v(v)=\mathbb{IG}(v;\underline{c},\underline{d})$ with two hyperparameters $\underline{c}$ and $\underline{d}$. This prior $\pi_v(v)$ is a hyperprior because it is elicited at a level deeper than β_{M} . Consequentially, the full conditional prior of Eq. 5 is refurnished as

$$\pi\left(\beta_{\mathrm{M}},\sigma^2,v\middle|\mathrm{M};\underline{a},\underline{b},\underline{c},\underline{d}\right)$$
$$=\pi_{\beta_{\mathrm{M}}}\left(\beta_{\mathrm{M}}\middle|\sigma^2,v,\mathrm{M}\right)\pi_{\sigma^2}\left(\sigma^2|\underline{a},\underline{b}\right)\pi_v\left(v|\underline{c},\underline{d}\right)$$

where the hyperparameters $\underline{a}$, $\underline{b}$, $\underline{c}$, and $\underline{d}$ are underlined and made explicit for the respective priors.

Second, the prior on model structure $\pi(\mathrm{M})$ is chosen to be vague in order to reflect a lack of prior knowledge on when and how many abrupt changes occur in an observed time series. Because of the separate parameterization for the trend and seasonal signals, it is reasonable to independently elicit the model priors for the trend and season signals:

$$\pi(\mathrm{M})=\pi\left(\{m\}\cup\{\tau_j\}_{j=1,\ldots,m}\right)\pi\left(\{p\}\cup\{\xi_k\}_{k=1,\ldots,p}\cup\{L_k\}_{k=0,\ldots,p}\right).$$

Firstly, the prior for the trend can be decomposed as

$$\pi\left(\{m\}\cup\{\tau_j\}_{j=1,\ldots,m}\right)=\pi\left(\{\tau_j\}_{j=1,\ldots,m}|m\right)\pi(m).$$

As a way to encode the vagueness of these model priors, we assume that the number of changepoints, m, takes any integer with an equal probability a priori. Meanwhile, we impose a constraint on the maximum number of changepoints allowable in a trend signal, as denoted by $\underline{\mathrm{m}_{max}}$, which helps to preclude over-complicated models. The prior $\pi(m)$ is therefore a uniform distribution over $\{0,1,..\ \underline{\mathrm{m}_{max}}\}$:

$$\pi(m)=\begin{cases}1/(\underline{\mathrm{m}_{max}}+1) & if\ 0\le m\le\ \underline{\mathrm{m}_{max}}\\ 0 & if\ m>\ \underline{\mathrm{m}_{max}}\end{cases}$$

Further, given a total of m changepoints, their locations, $\{\tau_j\}_{j=1,\ldots,m}$ are assumed to take random values from the points of observation time $\{t_i\}_{i=1,\ldots,n}$. This choice again represents a non-informative prior. As a practical constraint, we assume that any consecutive changepoints should be separated apart by at least a time interval $\underline{\mathrm{T}}$. Put together, the conditional prior for changepoint locations is

$$\pi\left(\left\{\tau_j\right\}_{j=1,\dots,m}|m\right) \propto \begin{cases} 1 & if \max\limits_{(i,j)}\left|\tau_i - \tau_j\right| < \underline{T} \\ 0 & \text{otherwise} \end{cases}.$$

Secondly, the prior on the seasonal model structure can be re-written as

$$\pi\left(\{p\} \cup \{\xi_k\}_{k=1,\dots,p} \cup \{L_k\}_{k=0,\dots,p}\right) = \pi\left(\{\xi_k\}_{k=1,\dots,p}|p\right)\pi(p)\prod\nolimits_{k=0}^{p}\pi(L_k)$$

where the priors on the number and locations of changepoints, $\pi(p)$ and $\pi\left(\{\xi_k\}_{k=1,\dots,p}|p\right)$, take the same forms as those of the trend signal, except that the maximum number of changpoints allowable is $\underline{\mathrm{p}_{max}}$ rather than $\underline{\mathrm{m}_{max}}$ and that the minimum separable distance between adjacent changepoints is $\underline{\mathrm{W}}$ rather than $\underline{\mathrm{T}}$. Similarly, the prior on the order of the piecewise harmonic model, $\pi(L_k)$, is also considered non-informative in that L_k randomly takes any value between pre-defined lower and upper limits of the allowable orders ($\underline{L_{min}}$ and $\underline{L_{max}}$):

$$\pi(L_k) = \begin{cases} 1/(\underline{L_{max}} - \underline{L_{min}} + 1) & if\ \underline{L_{min}} \le L_k \le \underline{L_{max}} \\ 0 & \text{otherwise} \end{cases}.$$

In the prior above, the model parameters $\{\beta_{\mathbf{M}}, \sigma^2, v, \mathbf{M}\}$ are of inferential interest and are all considered random. In contrast, the ten underlined hyperparameters $\{\underline{a}, \underline{b}, \underline{c}, \underline{d}, \underline{\mathrm{m}_{max}}, \underline{\mathrm{p}_{max}}, \underline{T}, \underline{W}, \underline{L_{min}}, \underline{L_{max}}\}$ are treated as fixed and should be pre-specified, although it is permissible to additionally treat them as random variables by further eliciting hyperpior distributions at higher levels in a manner similar to the treatment of v. There are no general rules on how to specify the values of these hyperparameters. The setup in this study was chosen as $\underline{a} = \underline{b} = 0.01$, $\underline{c} = \underline{d} = 0.02$, $\underline{T} = \underline{W} = 1\ year$, $\underline{L_{min}} = 0$, $\underline{L_{max}} = 10$, $\underline{\mathrm{m}_{max}} = \underline{\mathrm{p}_{max}} = \max(\mathrm{n/P}, 30)$ with n and P being the total number of observations and the period of the NDVI signal, respectively. Such choices for the inverse gamma priors are almost equivalent to non-informative priors for practical purposes, reflecting our vague knowledge on σ^2 or v a priori. Preliminary trials with various datasets suggest that the resulting predictive performances are insensitive to the settings of these hyperparameters as long as $\underline{\mathrm{m}_{max}}$, $\underline{\mathrm{p}_{max}}$, and $\underline{L_{max}}$ assumes a moderately large value (e.g., $\underline{\mathrm{m}_{max}} >$ 15, and $\underline{\mathrm{L}_{max}} > 6$), $\{\underline{a}, \underline{b}, \underline{c}, \underline{d}\}$ take small values, and the data are standardized beforehand.

As a recap of the Bayesian formulation for BEAST, the likelihood Eq. 4 and the priors Eqs.5 and 7 combine to reach the full posterior of our formulation according to Eq. 3:

$$p(\beta_{\mathbf{M}}, \sigma^2, v, \mathrm{M}|\boldsymbol{\mathcal{D}}) \propto$$

$$\prod_{i=1}^{n} N(y_i; \mathbf{x}_{\mathrm{M}}(t_i)\boldsymbol{\beta}_{\mathrm{M}}, \sigma^2)\, \pi_{\beta}(\beta_{\mathbf{M}}|\sigma^2, v, \mathbf{M})\pi_{\sigma^2}(\sigma^2|\underline{a}, \underline{b})\pi_v(v|\underline{c}, \underline{d})\, \pi\left(\{\tau_j\}_{j=1,\dots,m}, m|\underline{\mathrm{m}_{max}}, \underline{\mathrm{T}}\right) \pi\left(\{\xi_k\}_{k=1,\dots,p}, p|\underline{p_{max}}, \underline{W}\right) \prod_{k=0}^{p} \pi(L_k|p, L_{min}, L_{max}).$$

It can be further factored into three conditional posteriors:

$$p(M|v, \boldsymbol{\mathcal{D}}) \propto\ p(\boldsymbol{\mathcal{D}}|v, M) \cdot \pi(M);$$

$$p(\boldsymbol{\beta}_M, \sigma^2|v, M, \boldsymbol{\mathcal{D}}) = \mathbb{N}\left(\boldsymbol{\beta}_M;\ \mathbf{V}_M^* \mathbf{X}_M^{\mathrm{T}}\mathbf{y}, \sigma^2\mathbf{V}_M^*\right) \cdot \mathbb{IG}\left(\sigma^2; \underline{a} + \tfrac{n}{2}, \underline{b} + \left[\mathbf{y}^T\mathbf{y} - \mathbf{y}^T \mathbf{X}_M \mathbf{V}_M^* \mathbf{X}_M^{\mathrm{T}}\mathbf{y}\right]\right);$$

$$p(v|\boldsymbol{\beta}_M, \sigma^2, M, \boldsymbol{\mathcal{D}}) = \mathbb{IG}\left(v; \underline{c} + \frac{p_M}{2}, \underline{d} + \frac{\sum_{k=1}^{p_M} \beta_{k,M}^2}{2}\right)$$

where we have $\mathbf{V}_M^* = \left(v^{-1}\mathrm{I}_{\mathrm{M}} + \mathbf{X}_M^{\mathrm{T}} \mathbf{X}_M\right)^{-\mathbf{1}}$ and p_M is the total number of coefficients collected for all segments of the piecewise linear and harmonic models. These three conditional posteriors were sampled iteratively to simulate a chain of posterior samples using our hybrid Gibbs MCMC sampler.

References

Alcaraz-Segura, D., Chuvieco, E., Epstein, H.E., Kasischke, E.S., & Trishchenko, A. (2010). Debating the greening vs. browning of the North American boreal forest: differences between satellite datasets. *Global Change Biology, 16*, 760-770

Balke, N.S. (1993). Detecting level shifts in time series. *Journal of Business & Economic Statistics, 11*, 81-92

Banner, K.M., & Higgs, M.D. (2017). Considerations for assessing model averaging of regression coefficients. *Ecological Applications, 27*, 78-93

Betken, A. (2017). Change point estimation based on Wilcoxon tests in the presence of long-range dependence. *Electronic Journal of Statistics, 11*, 3633-3672

Beven, K. (2010). *Environmental modelling: An uncertain future?* : CRC Press

Brando, P.M., Goetz, S.J., Baccini, A., Nepstad, D.C., Beck, P.S., & Christman, M.C. (2010). Seasonal and interannual variability of climate and vegetation indices across the Amazon. *Proceedings of the National Academy of Sciences*, 200908741

Breiman, L. (2001a). Random forests. *Machine learning, 45*, 5-32

Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science, 16*, 199-231

Brockwell, P.J., & Davis, R.A. (2016). *Introduction to time series and forecasting*. springer

Brooks, E.B., Thomas, V.A., Wynne, R.H., & Coulston, J.W. (2012). Fitting the multitemporal curve: A Fourier series approach to the missing data problem in remote sensing analysis. *IEEE transactions on Geoscience and Remote Sensing, 50*, 3340-3353

Brooks, E.B., Wynne, R.H., Thomas, V.A., Blinn, C.E., & Coulston, J.W. (2014). On-the-fly massively multitemporal change detection using statistical quality control charts and Landsat data. *IEEE transactions on Geoscience and Remote Sensing, 52*, 3316-3332

Browning, D.M., Maynard, J.J., Karl, J.W., & Peters, D.C. (2017). Breaks in MODIS time series portend vegetation change ‐ verification using long ‐ term data in an arid grassland ecosystem. *Ecological Applications*

Burkett, V.R., Wilcox, D.A., Stottlemyer, R., Barrow, W., Fagre, D., Baron, J., Price, J., Nielsen, J.L., Allen, C.D., & Peterson, D.L. (2005). Nonlinear dynamics in ecosystem response to climatic change: case studies and policy implications. *Ecological complexity, 2*, 357-394

Burnham, K.P., & Anderson, D.R. (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media

Cade, B.S. (2015). Model averaging and muddled multimodel inferences. *Ecology, 96*, 2370-2382

Cai, Z., Jönsson, P., Jin, H., & Eklundh, L. (2017). Performance of Smoothing Methods for Reconstructing NDVI Time-Series and Estimating Vegetation Phenology from MODIS Data. *Remote Sensing, 9*, 1271

Chen, B., Xu, G., Coops, N.C., Ciais, P., Innes, J.L., Wang, G., Myneni, R.B., Wang, T., Krzyzanowski, J., & Li, Q. (2014). Changes in vegetation photosynthetic activity trends across the Asia–Pacific region over the last three decades. *Remote Sensing of Environment, 144*, 28-41

Cogger, K.O. (2010). Nonlinear multiple regression methods: a survey and extensions. *Intelligent Systems in Accounting, Finance and Management, 17*, 19-39

Cohen, W.B., Healey, S.P., Yang, Z., Stehman, S.V., Brewer, C.K., Brooks, E.B., Gorelick, N., Huang, C., Hughes, M.J., & Kennedy, R.E. (2017). How Similar Are Forest Disturbance Maps Derived from Different Landsat Time Series Algorithms? *Forests, 8*, 98

Cohen, W.B., Yang, Z., Healey, S.P., Kennedy, R.E., & Gorelick, N. (2018). A LandTrendr multispectral ensemble for forest disturbance detection. *Remote Sensing of Environment, 205*, 131-140

Dashti, H., Glenn, N.F., Ustin, S., Mitchell, J.J., Qi, Y., Ilangakoon, N.T., Flores, A.N., Silván-Cárdenas, J.L., Zhao, K., & Spaete, L.P. (2019). Empirical Methods for Remote Sensing of Nitrogen in Drylands May Lead to Unreliable Interpretation of Ecosystem Function. *IEEE transactions on Geoscience and Remote Sensing*

Denison, D.G. (2002). *Bayesian methods for nonlinear classification and regression*. John Wiley & Sons

Eidenshink, J., Schwind, B., Brewer, K., Zhu, Z., Quayle, B., & Howard, S. (2007). 1145801. A project for monitoring trends in burn severity. *Nutrition and cancer, 58*, 28-34

Ellison, A.M. (2004). Bayesian inference in ecology. *Ecology letters, 7*, 509-520

Finley, A.O., Banerjee, S., & Carlin, B.P. (2007). spBayes: an R package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software, 19*, 1

Finley, A.O., Banerjee, S., Ek, A.R., & McRoberts, R.E. (2008). Bayesian multivariate process modeling for prediction of forest attributes. *Journal of Agricultural, Biological, and Environmental Statistics, 13*, 60

Franklin, J., Serra-Diaz, J.M., Syphard, A.D., & Regan, H.M. (2016). Global change and terrestrial plant community dynamics. *Proceedings of the National Academy of Sciences, 113*, 3725-3734

Franzke, C.L. (2014). Warming trends: Nonlinear climate change. *Nature Climate Change, 4*, 423

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning*. Springer series in statistics New York

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika, 82*, 711-732

Grossman, Y., Ustin, S., Jacquemoud, S., Sanderson, E., Schmuck, G., & Verdebout, J. (1996). Critique of stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. *Remote Sensing of Environment, 56*, 182-193

Hamilton, J.D. (1994). *Time series analysis*. Princeton university press Princeton, NJ

Harvey, A.C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press

Hawbaker, T.J., Vanderhoof, M.K., Beal, Y.-J., Takacs, J.D., Schmidt, G.L., Falgout, J.T., Williams, B., Fairaux, N.M., Caldwell, M.K., & Picotte, J.J. (2017). Mapping burned areas using dense time-series of Landsat data. *Remote Sensing of Environment, 198*, 504-522

Healey, S.P., Cohen, W.B., Yang, Z., Brewer, C.K., Brooks, E.B., Gorelick, N., Hernandez, A.J., Huang, C., Hughes, M.J., & Kennedy, R.E. (2018). Mapping forest change using stacked generalization: An ensemble approach. *Remote Sensing of Environment, 204*, 717-728

Hu, T., Zhao, T., Shi, J., Wu, S., Liu, D., Qin, H., & Zhao, K. (2017). High-Resolution Mapping of Freeze/Thaw Status in China via Fusion of MODIS and AMSR2 Data. *Remote Sensing, 9*, 1339

Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z., & Vogelmann, J.E. (2010). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment, 114*, 183-198

Huete, A.R., Didan, K., Shimabukuro, Y.E., Ratana, P., Saleska, S.R., Hutyra, L.R., Yang, W., Nemani, R.R., & Myneni, R. (2006). Amazon rainforests green ‐ up with sunlight in dry season. *Geophysical research letters, 33*

Jamali, S., Jönsson, P., Eklundh, L., Ardö, J., & Seaquist, J. (2015). Detecting changes in vegetation trends using time series segmentation. *Remote Sensing of Environment, 156*, 182-195

Jentsch, A., Kreyling, J., & Beierkuhnlein, C. (2007). A new generation of climate ‐ change experiments: events, not trends. *Frontiers in Ecology and the Environment, 5*, 365-374

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R., Kattge, J., Latimer, A.M., & Moorcroft, P. (2016). Monitoring plant functional diversity from space. *Nature plants, 2*

Jiang, B., Liang, S., Wang, J., & Xiao, Z. (2010). Modeling MODIS LAI time series using three statistical methods. *Remote Sensing of Environment, 114*, 1432-1444

Jong, R., Verbesselt, J., Schaepman, M.E., & Bruin, S. (2012). Trend changes in global greening and browning: contribution of short ‐ term trends to longer ‐ term change. *Global Change Biology, 18*, 642-655

Jonsson, P., & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE transactions on Geoscience and Remote Sensing, 40*, 1824-1832

Kennedy, M.C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63*, 425-464

Kennedy, R.E., Andréfouët, S., Cohen, W.B., Gómez, C., Griffiths, P., Hais, M., Healey, S.P., Helmer, E.H., Hostert, P., & Lyons, M.B. (2014). Bringing an ecological view of change to Landsat ‐ based remote sensing. *Frontiers in Ecology and the Environment, 12*, 339-346

Kennedy, R.E., Yang, Z., Braaten, J., Copass, C., Antonova, N., Jordan, C., & Nelson, P. (2015). Attribution of disturbance change agent from Landsat time-series in support of habitat monitoring in the Puget Sound region, USA. *Remote Sensing of Environment, 166*, 271-285

Kennedy, R.E., Yang, Z., & Cohen, W.B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sensing of Environment, 114*, 2897-2910

Li, L., Vrieling, A., Skidmore, A., Wang, T., & Turak, E. (2018). Monitoring the dynamics of surface water fraction from MODIS time series in a Mediterranean environment. *International Journal of Applied Earth Observation and Geoinformation, 66*, 135-145

Liu, D., Toman, E., Fuller, Z., Chen, G., Londo, A., Zhang, X., & Zhao, K. (2018). Integration of historical map and aerial imagery to characterize long-term land-use change and landscape dynamics: An object-based analysis via Random Forests. *Ecological Indicators, 95*, 595-605

Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing, 25*, 2365-2401

Martínez, B., & Gilabert, M.A. (2009). Vegetation dynamics from NDVI time series analysis using the wavelet transform. *Remote Sensing of Environment, 113*, 1823-1842

McRoberts, R.E. (2011). Satellite image-based maps: Scientific inference or pretty pictures? *Remote Sensing of Environment, 115*, 715-724

Myneni, R.B., Keeling, C., Tucker, C.J., Asrar, G., & Nemani, R.R. (1997). Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature, 386*, 698

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., & Wulder, M.A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment, 148*, 42-57

Olofsson, P., Foody, G.M., Stehman, S.V., & Woodcock, C.E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment, 129*, 122-131

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science, 263*, 641-646

Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H., & Turner, W. (2014). Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology, 51*, 839-848

Piao, S., Ciais, P., Friedlingstein, P., Peylin, P., Reichstein, M., Luyssaert, S., Margolis, H., Fang, J., Barr, A., & Chen, A. (2008). Net carbon dioxide losses of northern ecosystems in response to autumn warming. *Nature, 451*, 49

Piao, S., Mohammat, A., Fang, J., Cai, Q., & Feng, J. (2006). NDVI-based increase in growth of temperate grasslands and its responses to climate changes in China. *Global Environmental Change, 16*, 340-348

Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., & Ohmann, J.L. (2010). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment, 114*, 1053-1068

Raftery, A.E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review, 133*, 1155-1174

Rankin, B.M., Meola, J., & Eismann, M.T. (2017). Spectral Radiance Modeling and Bayesian Model Averaging for Longwave Infrared Hyperspectral Imagery and Subpixel Target Identification. *IEEE transactions on Geoscience and Remote Sensing, 55*, 6726-6735

Reiche, J., de Bruin, S., Hoekman, D., Verbesselt, J., & Herold, M. (2015). A Bayesian approach to combine Landsat and ALOS PALSAR time series for near real-time deforestation detection. *Remote Sensing, 7*, 4973-4996

Reid, P.C., Hari, R.E., Beaugrand, G., Livingstone, D.M., Marty, C., Straile, D., Barichivich, J., Goberville, E., Adrian, R., & Aono, Y. (2016). Global impacts of the 1980s regime shift. *Global Change Biology, 22*, 682-703

Roy, D.P., Wulder, M., Loveland, T.R., Woodcock, C., Allen, R., Anderson, M., Helder, D., Irons, J., Johnson, D., & Kennedy, R. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment, 145*, 154-172

Samanta, A., Ganguly, S., Hashimoto, H., Devadiga, S., Vermote, E., Knyazikhin, Y., Nemani, R.R., & Myneni, R.B. (2010). Amazon forests did not green‐up during the 2005 drought. *Geophysical research letters, 37*

Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., & Gao, F. (2013). Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description. In: US Geological Survey

Schowengerdt, R.A. (2006). *Remote sensing: models and methods for image processing*. Academic press

Shen, M. (2011). Spring phenology was not consistently related to winter warming on the Tibetan Plateau. *Proceedings of the National Academy of Sciences, 108*, E91-E92

Shmueli, G. (2010). To explain or to predict? *Statistical science, 25*, 289-310

Shu, L., Jiang, Q., Zhang, X., & Zhao, K. (2017). Potential and limitations of satellite laser altimetry for monitoring water surface dynamics: ICESat for US lakes. *International Journal of Agricultural and Biological Engineering, 10*, 154-165

Solomon, S. (2007). *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*. Cambridge University Press

Su, Y., Guo, Q., Xue, B., Hu, T., Alvarez, O., Tao, S., & Fang, J. (2016). Spatial distribution of forest aboveground biomass in China: Estimation through combination of spaceborne lidar, optical imagery, and forest inventory data. *Remote Sensing of Environment, 173*, 187-199

Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A., & Fisher, P.F. (2015). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment, 160*, 1-14

Thomas, R.Q., Jersild, A.L., Brooks, E.B., Thomas, V.A., & Wynne, R.H. (2018). A mid‐century ecological forecast with partitioned uncertainty predicts increases in loblolly pine forest productivity. *Ecological Applications, 28*, 1503-1519

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010a). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment, 114*, 106-115

Verbesselt, J., Hyndman, R., Zeileis, A., & Culvenor, D. (2010b). Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment, 114*, 2970-2980

Wang, D., Morton, D., Masek, J., Wu, A., Nagol, J., Xiong, X., Levy, R., Vermote, E., & Wolfe, R. (2012). Impact of sensor degradation on the MODIS NDVI time series. *Remote Sensing of Environment, 119*, 55-61

Wang, X., Piao, S., Ciais, P., Li, J., Friedlingstein, P., Koven, C., & Chen, A. (2011). Spring temperature change and its implication in the change of vegetation growth in North America from 1982 to 2006. *Proceedings of the National Academy of Sciences, 108*, 1240-1245

Wintle, B.A., McCarthy, M.A., Volinsky, C.T., & Kavanagh, R.P. (2003). The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology, 17*, 1579-1590

Wu, W.B., & Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69*, 391-410

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., & Woodcock, C.E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment, 122*, 2-10

Yu, H., Luedeling, E., & Xu, J. (2010). Winter and spring warming result in delayed spring phenology on the Tibetan Plateau. *Proceedings of the National Academy of Sciences, 107*, 22151-22156

Zhang, X., & Zhao, K. (2012). Bayesian neural networks for uncertainty analysis of hydrologic modeling: a comparison of two schemes. *Water resources management, 26*, 2365-2382

Zhao, K., & Jackson, R.B. (2014). Biophysical forcings of land‐use changes from potential forestry activities in North America. *Ecological Monographs, 84*, 329-353

Zhao, K., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment, 113*, 182-196

Zhao, K., Popescu, S., & Zhang, X. (2008). Bayesian learning with Gaussian processes for supervised classification of hyperspectral data. *Photogrammetric Engineering & Remote Sensing, 74*, 1223-1234

Zhao, K., Suarez, J.C., Garcia, M., Hu, T., Wang, C., & Londo, A. (2018). Utility of multitemporal lidar for forest and carbon monitoring: Tree growth, biomass dynamics, and carbon flux. *Remote Sensing of Environment, 204*, 883-897

Zhao, K., Valle, D., Popescu, S., Zhang, X., & Mallick, B. (2013). Hyperspectral remote sensing of plant biochemistry using Bayesian model averaging with variable and band selection. *Remote Sensing of Environment, 132*, 102-119

Zhou, T., Popescu, S.C., Lawing, A.M., Eriksson, M., Strimbu, B.M., & Bürkner, P.C. (2017). Bayesian and Classical Machine Learning Methods: A Comparison for Tree Species Classification with LiDAR Waveform Signatures. *Remote Sensing, 10*, 39

Zhu, Z. (2017). Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS Journal of Photogrammetry and Remote Sensing, 130*, 370-384

Zhu, Z., Wang, S., & Woodcock, C.E. (2015). Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment, 159*, 269-277

Zhu, Z., & Woodcock, C.E. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment, 144*, 152-171

Zhu, Z., Woodcock, C.E., & Olofsson, P. (2012). Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment, 122*, 75-91