TECHNOMETRICS
2025, VOL. 67, NO. 1, 23–31
https://doi.org/10.1080/00401706.2024.2365730

Taylor & Francis
Taylor & Francis Group

Drift versus Shift: Decoupling Trends and Changepoint Analysis

Haoxuan Wu[a], Toryn L. J. Schafer[b], Sean Ryan[a], and David S. Matteson[a]

[a]Department of Statistics and Data Science, Cornell University, Ithaca, NY, USA; [b]Department of Statistics, Texas A&M University, College Station, TX, USA

ABSTRACT

We introduce a new approach for decoupling trends (drift) and changepoints (shifts) in time series. Our locally adaptive model-based approach for robustly decoupling combines Bayesian trend filtering and machine learning based regularization. An over-parameterized Bayesian dynamic linear model (DLM) is first applied to characterize drift. Then a weighted penalized likelihood estimator is paired with the estimated DLM posterior distribution to identify shifts. We show how Bayesian DLMs specified with so-called shrinkage priors can provide smooth estimates of underlying trends in the presence of complex noise components. However, their inability to shrink exactly to zero inhibits direct changepoint detection. In contrast, penalized likelihood methods are highly effective in locating changepoints. However, they require data with simple patterns in both signal and noise. The proposed decoupling approach combines the strengths of both, that is, the flexibility of Bayesian DLMs with the hard thresholding property of penalized likelihood estimators, to provide changepoint analysis in complex, modern settings. The proposed framework is outlier robust and can identify a variety of changes, including in mean and slope. It is also easily extended for analysis of parameter shifts in time-varying parameter models like dynamic regressions. We illustrate the flexibility and contrast the performance and robustness of our approach with several alternative methods across a wide range of simulations and application examples.

ARTICLE HISTORY
Received July 2022
Accepted June 2024

KEYWORDS
Dynamic linear models; Posterior summary; Stochastic Volatility; Structural change; Trend filtering

1. Introduction

Complex nonstationary dynamic systems often exhibit both global (macro patterns) and local (micro fluctuations) features of inferential interest. Herein, we focus on making distinctions between drifts and shifts. Drift describes the micro-level evolution of a process and may appear as variation about gradual trends. In contrast, shifts represent macro-level changes in a process perceived as sharp discontinuities, rapid changes, or major breaks.

A commonly used approach for modeling drift in time series regression is the dynamic linear model (DLM). DLMs tend to be overparameterized models with at least one parameter per observation time. Therefore, we focus on Bayesian estimation of DLMs (Chan and Eisenstat 2018) with priors for selection or regularization. Continuous shrinkage priors regularize parameters to produce smoother and more reliable estimates for the dynamic features by "shrinking" small values closer to zero (Carvalho, Polson, and Scott 2009; Bitto and Fruhwirth-Schnatter 2019). Bayesian DLMs with shrinkage priors are excellent for capturing changes that occur smoothly over time. However, since the inference does not include exact zero values, characterizing shifts in the trend, such as changepoints, is not straightforward.

On the other hand, changepoint methods tend to be effective at capturing sudden breaks (Aminikhanghahi and Cook 2017). Common changepoint methods include likelihood ratio tests with cumulative statistics (Jeske et al. 2009; Fryzlewicz 2014), penalized likelihood approaches (Killick, Fearnhead, and Eckley 2012; Maidstone et al. 2017) and nonparametric distanced based metrics (Matteson and James 2014; James and Matteson 2014). While these methods have shown to be effective on well-behaved time series, they tend to struggle when we model systems characterized by drift and shift.

Given the complementary strengths of Bayesian DLMs and changepoint models, it is natural to explore an intersection of these methods. In this article, we propose a two step Bayesian method using a decoupled posterior summary that allows us to identify changepoints in any Bayesian DLM. First, a Bayesian DLM is fitted to filter the signal of the data from the noise components. We do not specify a particular structure for the DLM but rather will show the approach works for a wide range of structures. Second, a penalized loss on the posterior of the model imposes a sparse summary of changepoint locations.

The decoupled approach as presented by Hahn and Carvalho (2015) separated the processes of regression modeling and discrete inference of variable selection. In a similar vein, Florian Huber and Onorante (2021) applied the framework to a time-varying parameter model with a specification of the decoupled loss as introduced by Ray and Bhattacharya (2018). In this article, we extend the decoupled approach to nonstationary time series analysis and changepoint detection.

The decoupled approach provides two key advantages. First, the decoupled approach separates the estimation of the trend from the changepoint locations. As a result, we can fit a highly

CONTACT Haoxuan Wu *hw399@cornell.edu* Comstock Hall, 1198, 129 Garden Ave, Ithaca, NY 14853, USA.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/TECH*.

flexible Bayesian model to deal with the intricacies of the data such as outliers, heterogeneity and seasonality. Most existing changepoint algorithms struggle to deal with these components as they tend to significantly skew the distribution of the data and violate distributional assumptions. Second, by using a penalized loss on the posterior, the decoupled approach is able to provide uncertainty estimates for the number of changepoints selected. In turn, the decoupled approach can provide more insights into the selection process and the tradeoff between goodness-of-fit and the number of changepoints.

The article proceeds as follows. In Section 2, we introduce the decoupled approach for identifying changepoints in Bayesian DLMs. Section 3 and 4 illustrates the effectiveness of the decoupled approach in diverse sets of simulation scenarios and real-world datasets. We conclude with a discussion of key benefits. The supplementary materials details the loss derivation, methodology extensions, extended simulation results, and more real data applications.

2. Methodology

2.1. Decoupled Modeling

To introduce the decoupled approach, we start by introducing a standard Bayesian dynamic linear model (DLM). Suppose we observe a univariate time series $\boldsymbol{Y} = (y_1, \ldots, y_n)'$ and a predictor series $\boldsymbol{X} = (x_1, \ldots, x_n)'$, a Bayesian DLM can be formulated as follows:

$$\begin{aligned} y_t &= x_t\beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_{\epsilon,t}), \\ \triangle^D \beta_t &= \omega_t, \quad \omega_t \sim N(0, \sigma^2_{\omega}), \end{aligned} \tag{1}$$

where $\triangle^D(\cdot)$ is the degree D differencing operator with $D = 0$ defined as the identity function. In this setup, $\{\beta_t\}$ encodes the time-varying relationship between the predictor series $\{x_t\}$ and the response series $\{y_t\}$. The process $\{\epsilon_t\}$ models noise; $\{\sigma^2_{\epsilon,t}\}$ is modeled as potentially time varying; a heteroscedastic noise process gives additional flexibility with low computational cost, in practice. For now, we will assume only one predictor series. Later on, we will extend the framework to deal with multiple predictors.

Specifically, the random walk process, corresponding to $D = 1$, induces smooth estimates for $\{\beta_t\}$ when σ^2_{ω} is small. For well-behaved time series, a globally smooth estimate for $\{\beta_t\}$ provides sufficient inference. However, (1) does not include a mechanism for discrete inference applicable to time series characterized by shifts. Rather than adjust the priors, we chose to take a decoupled approach to summarize the posterior. The decoupled approach summarizes a relatively smooth estimate of $\{\beta_t\}$ with a penalized loss function that induces discrete inference. The discrete inference explored by Hahn and Carvalho (2015) was variable selection and we adapt the approach for discrete shift features such as abrupt changepoints. As we will show later on, for more noisy series, locally adaptive shrinkage priors may be necessary to induce sufficiently smooth estimation in time series with variable degrees of wiggliness.

To illustrate the connection between variable selection and changepoint detection, notice that the time-varying relationship $\{\beta_t\}$ can be seen as a discrete integration over the estimated increments $\{\omega_t\}$ and the initial values of $\{\beta_t\}$. In order for the coefficient function to be constant for some period of time, the increments must be zero. Therefore, shift detection is equivalent to estimating the nonzero increments analogous to estimating the nonzero coefficients in variable selection inference. For the decoupled approach, we fit the above model to the observed data via Gibbs sampling with the MCMC sampling scheme provided by the R package *dsp* from the methods in Kowal, Matteson, and Ruppert (2019) to estimate the posteriors for the coefficients which are dense and nonzero everywhere by model construction. Then, we choose a penalized loss function to summarize the posterior. Due to heteroscedastic noise in (1), we consider a weighted least squares loss function: $L^*_\lambda(\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{\beta}}) = \sum_{t=1}^n [w_t(\tilde{y}_t - x_t\widetilde{\beta}_t)]^2 + q_\lambda(\widetilde{\boldsymbol{\beta}})$, where $\widetilde{\boldsymbol{y}} = (\widetilde{y}_1, \ldots, \widetilde{y}_n)$ is the posterior prediction given (1), $\widetilde{\boldsymbol{\beta}} = (\widetilde{\beta}_1, \ldots, \widetilde{\beta}_n)$ is the penalized linear predictor, $q_\lambda(\cdot)$ is a penalty function to induce sparsity given a penalty parameter λ, and $\{w_t\}$ are the weights for each time-step. Details on $q_\lambda()$ and $\{w_t\}$ will be given in Section 2.2.

As in Hahn and Carvalho (2015), we first integrate $L^*_\lambda(\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{\beta}})$ over $\{\tilde{y}_t\}$ given $\{\beta_t, \sigma^2_{\epsilon,t}\}$ then integrate over $\{\beta_t, \sigma^2_{\epsilon,t}\}$ given $\{y_t\}$. This results in the decoupled loss as follows:

$$L^*_\lambda(\widetilde{\boldsymbol{\beta}}_t) = \sum_{t=1}^n w_t(x_t\bar{\beta}_t - x_t\widetilde{\beta}_t)^2 + q_\lambda(\widetilde{\boldsymbol{\beta}}_t), \tag{2}$$

where $\{\bar{\beta}_t\}$ denotes the posterior mean of the trend estimate from the Bayesian DLM (supplementary materials, Section C). Equation (2) can be thought of as a second level shrinkage on the underlying coefficients to induce hard thresholding (Hahn and Carvalho 2015). The loss function, parameterized by the penalty parameter λ, will be used to select changepoints from the posterior estimates of a Bayesian DLM.

2.2. Weights and Penalty Function

The choice of weighted least squares allows the approach to use the estimated variance from the Bayesian DLM to induce additional localized adaptivity. The weights adjust the penalty to time-varying volatility inherent in the data, inducing a smaller loss for time-steps with a larger variance and a larger loss for time-steps with a smaller variance.

For the weights $\{w_t\}$, the classic choice is inverse to the noise (Kiers 1997). In our case, since we have posterior estimates of the variance after sampling the Bayesian DLM, we set our weights to be

$$w_t = \overline{\sigma^{-2}_{\epsilon,t}}, \quad \text{for } t = 1, \ldots, n,$$

where $\overline{\sigma^{-2}_{\epsilon,t}}$ is the posterior mean for the precision at time t. As previously discussed, the weights induce additional robustness for change detection in heteroscedastic data.

The penalty term $q_\lambda(\widetilde{\boldsymbol{\beta}})$ will penalize the number of time-steps for which the Dth difference (i.e., D = 1 or 2) in $\widetilde{\boldsymbol{\beta}}$ is nonzero. Table 1 shows three possible choices for the penalty function. Ideally, the ℓ_0 penalty will be used to identify the optimal subset of time-steps in which the Dth difference are nonzero. However, the ℓ_0 penalty is difficult to estimate efficiently, making it infeasible for long time-series. One solution is to relax the ℓ_0 penalty to the ℓ_1 penalty. However, the ℓ_1 penalty induces shrinkage which tends to bias the result. This is due to the fact

Table 1. Choices of penalty function.

	ℓ_0	ℓ_1	Adaptive ℓ_1
Penalty function	$\lambda \sum_t \mathbb{I}_{\triangle^D \widetilde{\beta}_t \neq 0}$	$\lambda \sum_t \lvert \triangle^D \widetilde{\beta}_t \rvert$	$\lambda \sum_t \frac{1}{\lvert\psi_t\rvert} \lvert \triangle^D \widetilde{\beta}_t \rvert$
Motivation	Selection	Shrinkage and Selection	Combination of ℓ_0 and ℓ_1

Table defines various penalty functions, $q_\lambda(\cdot)$, for differenced coefficient vectors and their associated motivation. For example, the ℓ_0 penalty can be used for selecting nonzero increments of the differenced coefficient vectors.

that the penalty term increases linearly in relation to $\{\triangle^D \widetilde{\beta}_t\}$, resulting in the penalty favoring changepoints of low magnitude. As we will show in the simulations, using the ℓ_1 penalty directly will lead to significant over-estimation of changepoints.

A refined goal is then to identify a penalty function that combines the computational efficiency of the ℓ_1 penalty and the optional subset selection ability of the ℓ_0 penalty. As a result, we propose a version of the adaptive ℓ_1 penalty (Zou 2006) that pushes the ℓ_1 penalty closer to the ℓ_0 penalty. The resulting penalty can be written as follows:

$$q_\lambda(\widetilde{\boldsymbol{\beta}}) = \lambda \sum_t \frac{1}{|\psi_t|} |\triangle^D \widetilde{\beta}_t|, \tag{3}$$

where $\psi_t = \overline{\triangle^D \beta_t}$ for all t. Motivated by similar refinement in Hahn and Carvalho (2015), ψ_t at time t is the posterior mean of the Dth degree difference of β_t. This term can function as a normalizer which levels the impact of each changepoint regardless of the magnitude of the change at that time-step. In a time-step with a larger change in $\{\beta_t\}$, ψ_t will tend to be higher in magnitude, leading a changepoint to be penalized less. As a result, this weight term corrects some of the bias in the ℓ_1 penalty and gives better results for changepoint estimation. Optimization of (3) simplifies to using a standard penalized regression function; we used *glmnet* in R (Tay, Narasimhan, and Hastie 2023). A computational special case of optimization without a global penalty, λ, is discussed in (Ray and Bhattacharya 2018).

2.3. Selecting the Optimal Number of Changepoints

As seen in (2) and (3), we utilize a penalty function indexed by a parameter λ. The value of λ plays a critical role in the final selection of the number of changepoints. As λ approaches 0, there would be no enforcement of sparsity and every point will be treated as a changepoint. As λ approaches ∞, all $\{\triangle^D \widetilde{\beta}_t\}$ will be 0 and no changepoint will be detected. Typically, with a penalized loss function, cross-validation is used to select the penalty parameter. However, in the case for the proposed decoupled approach, since the loss is taken over the posterior estimate for the latent parameter $\{\beta_t\}$, the MCMC samples can be used in identifying the optimal set of changepoints.

First, minimization of the loss in (2) with our recommended penalty function can be solved via coordinate descent to produce a path of λ values corresponding to different number of changepoints. This path of solutions will express a direct tradeoff between goodness-of-fit and the number of changepoints. As the number of changepoints increases, the estimated solution $\widetilde{\boldsymbol{\beta}}$ will be closer to the posterior mean across time.

Second, for each λ in the corresponding solution path, we will compute the "projected posterior" (Woody, Carvalho, and Murray 2021) to quantify its uncertainty. The key idea behind the projected posterior is to project each MCMC draw from the Bayesian model onto the summary space defined by locations of changepoints. For a given value λ, let η denote the time indices which $\{\triangle^D \widetilde{\beta}_t \neq 0\}$ (i.e., the estimated changepoint locations). Initial points $1, \ldots, D$ are automatically included in every η as they are unpenalized. Let $\boldsymbol{\beta}^{(i)}$ denote the ith MCMC draw from the Bayesian model and $\boldsymbol{Z}$ denote the inverse of the Dth difference matrix. Let $\boldsymbol{Z}_\eta$ denote the subset of columns of $\boldsymbol{Z}$ indexed by a given η. The ith projected posterior is then given by:

$$\boldsymbol{\beta}_\eta^{(i)} = (\boldsymbol{Z}_\eta^{\mathbb{T}} \boldsymbol{Z}_\eta)^{-1} \boldsymbol{Z}_\eta^{\mathbb{T}} \boldsymbol{\beta}^{(i)}, \tag{4}$$

where $\mathbb{T}$ is the transpose operator. This projects $\boldsymbol{\beta}^{(i)}$ from each MCMC draw onto the best fitted model given the changepoint estimates. In summary, the "projected posterior" takes a set of changepoint locations and produces the best estimate of $\boldsymbol{\beta}$ for each of the MCMC draws given the changepoint locations. This, in turn, allows us to visualize a tradeoff between the number of changepoints and the corresponding fit for the posterior estimates.

Third, after deriving the projected posterior, we use a diagnostic tool to calculate a goodness-of-fit metric commonly based on amount of variation explained. Since we accounted for heteroscedasticity in the noise term of the Bayesian DLM, we propose using the following metric as an estimate for the amount of variation explained by the changepoints for the ith MCMC draw:

$$R_\eta^{2,(i)} \equiv 1 - \frac{\sum_{t=1}^n w_t (x_t \beta_t^{(i)} - x_t \beta_{\eta,t}^{(i)})^2}{\sum_{t=1}^n w_t (x_t \beta_t^{(i)} - x_t \mu_{\boldsymbol{\beta}^{(i)}})^2}$$

where $\mu_{\boldsymbol{\beta}^{(i)}}$ is the mean over t of the ith MCMC draw, $\boldsymbol{\beta}^{(i)}$. This metric is similar to R-squared in that it measures the amount of variation explained by the projected posterior $\boldsymbol{\beta}_\eta$ for each of the MCMC draws. However, the error for each time-step is multiplied by the corresponding weight value, giving time-steps with higher variances lower weights. This makes sense as we expect more uncertainty in regions of high noise volatility. In turn, this metric provides an estimate of variation explained for each of the MCMC draws. The higher the value of $R_\eta^{2,(i)}$, the better the fit of the "projected posterior" to the ith MCMC draw. For selecting the optimal value of λ, we will select the lowest number of changepoints which the upper 90% credible interval for $\tilde{E}[R_\eta^2]$ exceeds a certain threshold. We find this simple selection criterion to be quite effective in empirical settings and easy to visualize. Details on the threshold selection will be given in Section 3.

Figure 1 illustrates the decoupled approach on a simulated series with two changepoints in mean. The fit from a Bayesian DLM with random walk is very wiggly but captures the underlying trend for the most part (Figure 1 top-left). However, the model does not provide a clear identification of changepoints.

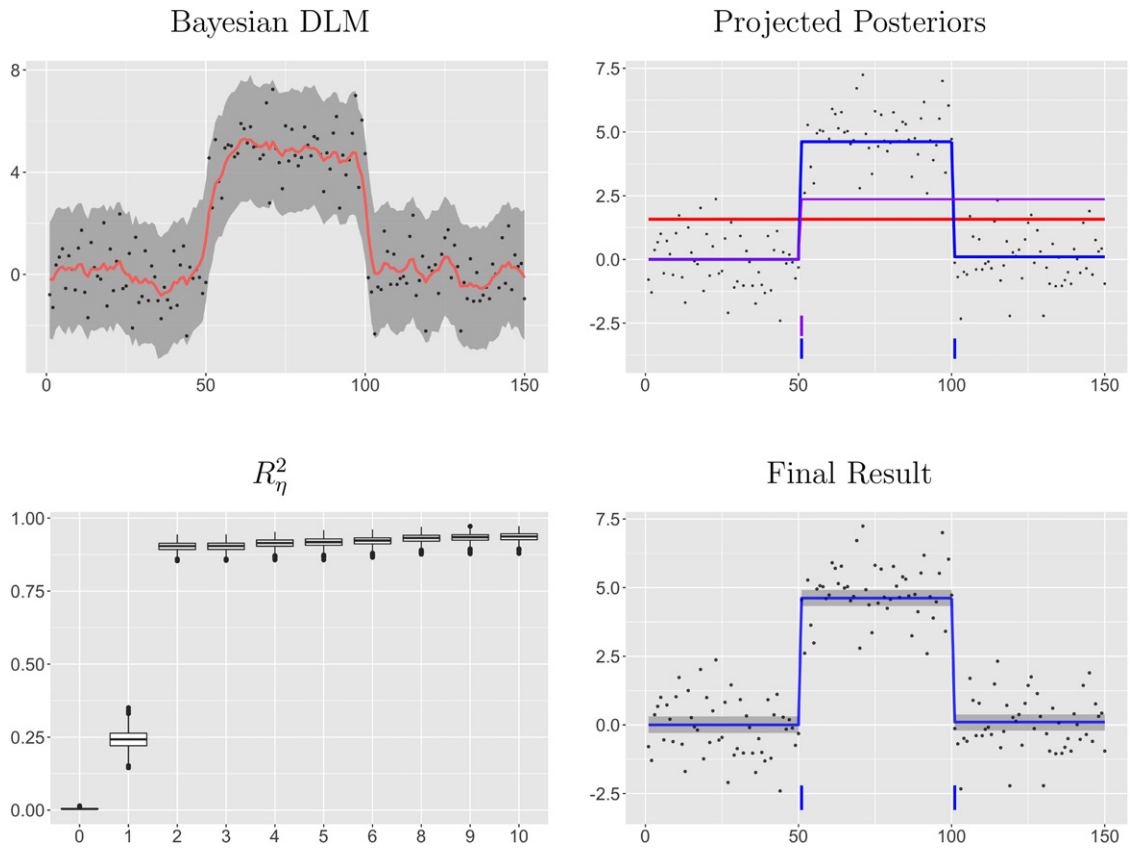

Figure 1. Illustrative Example of the Decoupled Approach with Random Walk: The top-left plot shows the resulting posterior mean of $\{\beta_t\}$ using a Bayesian DLM with random walk. 90% credible bands are shown as a ribbon. The top-right plot shows the mean of the projected posterior for 0, 1, and 2 number of changepoints. The bottom-left plots shows the distribution of R^2_η as a function of the number of changepoints. The bottom-right plots shows the final resulting fit from the decoupled approach and 90% bands corresponding to the projection.

The projected posterior for varying number of changepoints is shown in the top-right plot. For 0 changepoints, the projected posterior fits the global mean. For 1 changepoint, the projected posterior fits the first segment and combines the next 2 segments. For 2 changepoints, the projected posterior captures both true changepoints. This is reflected in the goodness-of-fit R^2_η. The metric shows large jumps from 0 to 1 changepoints and 1 to 2 changepoints, with marginal improvements afterward. We select 2 changepoints as the final result and plot the final projected posterior in the bottom-right plot. We additionally project all the posterior samples according to (4) for 2 changepoints (i.e., 3 nonzero values) and add the 90% credible bands. The narrow uncertainty bands of the projection reflect the shrinkage of the penalization as compared to the uncertainty from the DLM. As seen, we can turn a very wiggly fit of the Bayesian DLM to a clear separation of drifts and shifts.

2.4. Locally Adaptive Trends with Global-Local Shrinkage Priors

In the current model (1), we assume the coefficients $\{\beta_t\}$ follow a random walk with a constant variance σ^2_ω. While this setup can be sufficient for data with strong signal, this model tends to overfit in datasets with low signal-to-noise ratios. Shrinkage priors present a tradeoff between goodness-of-fit and smoothness of the underlying process; more shrinkage will typically result in a smoother underlying fit for the $\{\beta_t\}$ process. In this section, we will introduce the shrinkage priors for the decoupled approach.

As previously discussed in Section 1, various forms of shrinkage priors have shown to be effective in Bayesian DLMs. For this section, we will focus on the class of so called "global-local shrinkage priors" which have shown to be effective for Bayesian modeling (Bhadra et al. 2016). The prior on ω_t of (1) will be modified to $\omega_t \sim N(0, \tau^2_\omega \gamma^2_{\omega,t})$. This modification induces global-local shrinkage on the Dth difference of the coefficients for the predictor. The parameter τ^2_ω induces global shrinkage across all time-steps and the process $\{\gamma^2_{\omega,t}\}$ induces time-specific shrinkage for the coefficients. The two parameters combined shrink small deviations toward zero while allowing large signals to remain unchanged. This provides localized adaptivity while maintaining strong global shrinkage.

In time dependent data, an additional dependence in the latent shrinkage or selection process has been shown to improve estimation in a variety of techniques (Nakajima and West 2013; Kowal, Matteson, and Ruppert 2019; Wu and Matteson 2020; Rockova and McAlinn 2021). One example is the dynamic shrinkage process detailed in Kowal, Matteson, and Ruppert (2019). The shrinkage process uses an AR(1) structure on $\{\log(\tau^2_\omega \gamma^2_{\omega,t})\}$ to induce localized smoothing of high/low noise regions. The process is detailed as follows:

$$h_t \equiv \log(\tau^2_\omega \gamma^2_{\omega,t}), \quad h_t = u + \phi(h_{t-1} - u) + \xi_t,$$

where ϕ is a univariate autocorrelation parameter, $\xi_t \overset{\text{iid}}{\sim} Z(0.5, 0.5, 0, 1)$, in which $Z(\cdot)$ denotes the four parameter

$Z-$distribution, $Z(\alpha, \beta, \mu_z, \sigma_z)$, with density function

$$[z] = \{\sigma_z B(\alpha, \beta)\}^{-1} \exp\{(z - \mu_z)/\sigma_z\}^{\alpha} \\ [1 + \exp\{(z - \mu_z)/\sigma_z\}]^{-(\alpha+\beta)}, \; z \in \mathbb{R},$$

where $B(\cdot, \cdot)$ is the beta function. The distribution describes the log of an inverted beta random variable with parameters α and β. The parameters μ_z and σ_z allow for shifting and scaling. Due to the previous effectiveness of the model, we use this model for all simulations with first differences (D = 1) unless otherwise specified and refer to it as decoupled dynamic shrinkage (DC-DS).

Note that the decoupled approach is not restricted by the Bayesian DLM specification. A complex Bayesian model incorporating a variety of complexities such as covariates, heteroscedastic noise and nonstationary inputs can be fit to the data. The main recommendation is to fit a model that estimates fairly smooth coefficients for the predictors of interest. Then, the decoupled approach can adapt the inference part to identify key changepoints. This level of flexibility grants the decoupled approach the ability to work with more applications than previous existing changepoint algorithms. Extensions for dealing with multiple predictors and static parameters are shown in the supplementary materials, Section D.

3. Simulated Experiments

In this section, we illustrate the effectiveness and flexibility of the decoupled approach. The competing methods are the Pruned Exact Linear Time method (PELT, Killick, Fearnhead, and Eckley 2012) and Robust FPOP algorithm (R-FPOP, Fearnhead and Rigaill 2019). PELT identifies changepoint based on penalized cost function using a goodness-of-fit metric based on maximum negative likelihood for each segment and a penalty parameter on the number of changepoints. R-FPOP adapts the PELT penalty function using a biweight-loss in order to deal with outliers by establishing a maximum threshold for the impact of each time-step. The first simulation setting does not include an outlier process so therefore the R-FPOP method will not be considered a comparative method.

These two methods are similar to the decoupled approach in their utilization of a penalized cost function. However, unlike the decoupled approach, they use the data rather than the posterior of a Bayesian model. The comparisons will start on simple cases of changes in mean with Gaussian noise, then extend to more complicated scenarios adding in outliers and heterogeneity. For both competing methods, we will use the default parameters as used in the original papers. For the decoupled approach, we use a cutoff threshold of 0.9 for the lowest number of changepoints which the upper 90% credible interval for R^2_η exceeds. Full details of the parameters used for the Bayesian DLM and comparisons for other simulation settings are shown in the supplementary materials, Section D.

3.1. Comparison Metric Details

Five metrics are used to evaluate the results for simulations: Rand index, adjusted Rand index, precision, recall and F1-score. Rand index calculates a similarity score between the predicted partition and the true partition; the score ranges between 0 and 1 with 1 being a perfect match (Hubert and Arabie 1985). Adjusted Rand index provides an additional correction step to the Rand Index by accounting for random chance of a correct partition. Precision measures the proportion of true changepoints in the number of predicted changepoints while recall measures the proportion of all true changepoints detected by the models. F1-score calculates the harmonic mean between precision and recall. Since changepoints occur very rarely in the data, the F1-score is a good indicator for the accuracy of predictions (van den Burg and Williams 2020). We consider a predicted changepoint to be a true positive if it is within ± 5 of a true changepoint, with the caveat that each true changepoint can only match to at most one predicted changepoint.

3.2. Change in Mean with Gaussian Noise

For the first set of simulations, we start with a simple change in mean with standard Gaussian noise. We simulate data of length 200, with a changepoint in the middle of the data at location 100. We adjust different levels for the magnitude of change (MC) to understand the effectiveness of the algorithms with varying signal-to-noise ratios. We test 4 different magnitudes of change values of $\{1, 0.75, 0.5, 0.25\}$. We will compare the decoupled approach with a random walk state equation (DC-RW, model introduced in Section 2.1) and the decoupled approach with dynamic shrinkage (DC-DS, model introduced in Section 2.4) against PELT. PELT is a penalized likelihood changepoint algorithm which is designed to identify changes in this setting, making it a good baseline for comparison. We expect the decoupled approach to perform slightly worse than PELT in this setting as a tradeoff for increased flexibility. As we will show in the later simulations, the flexibility of the decoupled approach allows it to perform much better when the assumptions of homoscedasticity and Gaussian noise are violated.

As seen in Table 2, the decoupled approach with dynamic shrinkage performs slightly worse than PELT. With a signal-to-noise ratio of 1 to 1 (magnitude of change 1), both DC-DS and PELT perform similarly well with Rand average above 0.95, adjusted Rand average above 0.925 and F1-score above 0.8. As the signal-to-noise ratio reaches a low of 1–4 (magnitude of change 0.25), both changepoint algorithms can no longer distinguish the correct changepoint. For the magnitude change of 0.75 and 0.5, PELT performs slightly better in terms of F1-score. However, we still see a tradeoff of precision and recall between the two algorithms. For magnitude of change of 0.5, PELT has a higher precision while DC-DS has a higher recall. This shows that DC-DS has a tendency to slightly over-predict in low signal-to-noise ratio while PELT has a tendency to under-predict. A key note is that DC-DS maintains the highest Rand and adjusted Rand average in this settings, showing that DC-DS produces the partition closest to the true partition.

Comparing DC-RW against DC-DS, we can clearly see that using shrinkage priors in the Bayesian DLM significantly improves the performance of the decoupled approach. This is due to the fact that shrinkage priors induce smoother estimates of the underlying trend resulting in easier changepoint inference. This further supports the discussion in Section 2.4 of the advantages of the decoupled framework allowing for fitting of

Table 2. Single change in mean.

MC	Algorithms	Rand Avg.	Adj. Rand Avg.	Precision	Recall	F1-score
1	DC-RW	$0.795_{(0.013)}$	$0.590_{(0.027)}$	0.18	0.83	0.29
	DC-DS	$0.964_{(0.004)}$	$0.929_{(0.009)}$	0.78	0.83	0.80
	PELT	$\mathbf{0.968}_{(0.004)}$	$\mathbf{0.936}_{(0.007)}$	**0.82**	**0.87**	**0.84**
0.75	DC-RW	$0.689_{(0.016)}$	$0.379_{(0.031)}$	0.14	0.53	0.22
	DC-DS	$0.926_{(0.009)}$	$0.851_{(0.018)}$	**0.68**	0.56	0.61
	PELT	$\mathbf{0.930}_{(0.009)}$	$\mathbf{0.860}_{(0.021)}$	0.62	**0.67**	**0.64**
0.5	DC-RW	$0.576_{(0.012)}$	$0.156_{(0.024)}$	0.12	0.25	0.16
	DC-DS	$\mathbf{0.791}_{(0.017)}$	$\mathbf{0.582}_{(0.035)}$	0.23	**0.38**	0.29
	PELT	$0.755_{(0.021)}$	$0.512_{(0.041)}$	**0.38**	0.30	**0.33**
0.25	DC-RW	$0.498_{(0.000)}$	$0.000_{(0.000)}$	0.00	0.00	0.00
	DC-DS	$0.510_{(0.006)}$	$0.025_{(0.013)}$	0.02	0.01	0.01
	PELT	$0.498_{(0.000)}$	$0.000_{(0.000)}$	0.00	0.00	0.00

Table details results of the decoupled approach with random walk (DC-RW), decoupled approach with dynamic shrinkage (DC-DS), and PELT on simulated data with one change in mean of varying magnitudes (MC) and standard Gaussian noise. Rand average and adjusted Rand average measures the similarity between predicted partition and true partition. Standard error for Rand average and adjusted Rand average are given in subscripts. F1-score measures accuracy of changepoint detection through a comparison of precision and recall. Bolded values indicate best results for the metric in the column.

Table 3. Change in mean with outliers.

MC	Algorithms	Rand Avg.	Adj. Rand Avg.	Precision	Recall	F1-score
2	DC-DS	$\mathbf{0.977}_{(0.003)}$	$\mathbf{0.954}_{(0.007)}$	**0.88**	**0.91**	**0.89**
	PELT	$0.681_{(0.006)}$	$0.361_{(0.012)}$	0.06	0.82	0.11
	R-FPOP	$0.952_{(0.011)}$	$0.904_{(0.022)}$	0.86	0.83	0.84
1.5	DC-DS	$\mathbf{0.967}_{(0.005)}$	$\mathbf{0.934}_{(0.009)}$	**0.80**	**0.82**	**0.81**
	PELT	$0.689_{(0.007)}$	$0.376_{(0.014)}$	0.05	0.74	0.10
	R-FPOP	$0.860_{(0.020)}$	$0.721_{(0.040)}$	**0.80**	0.63	0.70
1	DC-DS	$\mathbf{0.935}_{(0.007)}$	$\mathbf{0.870}_{(0.015)}$	**0.68**	0.60	**0.63**
	PELT	$0.680_{(0.007)}$	$0.358_{(0.013)}$	0.04	0.59	0.08
	R-FPOP	$0.804_{(0.009)}$	$0.638_{(0.019)}$	0.35	**0.62**	0.44
0.5	DC-DS	$\mathbf{0.742}_{(0.018)}$	$\mathbf{0.486}_{(0.036)}$	**0.18**	0.32	**0.24**
	PELT	$0.676_{(0.006)}$	$0.350_{(0.013)}$	0.04	**0.48**	0.07
	R-FPOP	$0.741_{(0.017)}$	$0.484_{(0.035)}$	0.15	0.25	0.19

Table details decoupled approach with dynamic shrinkage (DC-DS), PELT, and R-FPOP on simulated data with one change in mean of varying magnitudes (MC) and outliers. Outliers are simulated using t-distributed noise with 2 degrees of freedom. Rand average and adjusted Rand average measures the similarity between the predicted partition and true partition. Standard error for Rand average and adjusted Rand average across simulations are given in subscripts. F1-score measures accuracy of changepoint detection through a comparison of precision and recall. Bolded values indicate best results for the metric in the column.

any appropriately complex Bayesian model. Due to the significant improvements of using shrinkage priors in the baseline case, we will use DC-DS as our main method from this point onward.

3.3. Change in Mean with Outliers

For the next set of simulations, we added outliers onto the same problem as Section 3.2 to illustrate the robustness of the methods. All other simulation settings will be kept the same as Section 3.2. As this is a more difficult problem, we increase the magnitude of changes to $\{2, 1.5, 1, 0.5\}$. Instead of Gaussian noise, we will use t-distributed noise with 2 degrees of freedom to simulate data with outliers.

The results of the simulation can be seen in Table 3. With the addition of outliers, the decoupled approach is able to achieve the best performance across all settings. By using a Bayesian DLM with dynamic shrinkage, the decoupled approach is robust to the presence of extreme outliers. Unsurprisingly, PELT, with no mechanism to deal with extreme values, is significantly influenced by outliers. This lead to PELT significantly over-predicting the number of changepoints. In the setting of magnitude of change of 2, DC-DS achieves an F1-score of 0.89 in comparison to 0.84 for R-FPOP. As the signal-to-noise ratio decreases from 2 to 1, we can see an increasing gap in adjusted Rand average and F1-score between DC-DS and R-FPOP. This indicates that DC-DS can produce more precise changepoint estimations and more accurate partitions. The advantage becomes more significant in the setting of magnitude of change of 1. DC-DS achieves an F1-score of 0.63 in comparison to 0.44 of R-FPOP. As magnitude of change approaches 0.5, the problem becomes too difficult for all algorithms and performance is comparable between DC-DS and R-FPOP.

3.4. Change in Mean in Presence of Heteroscedasticity

For the next set of simulations, we evaluate these algorithms in presence of heterogeneity. We simulate 100 series of length 200, with a changepoint in the middle of the data at location 100. However, instead of standard Gaussian noise or t-distributed noise, we generate noise using stochastic volatility of order 1 (Kim, Shephard, and Chib 1998) as follows:

$$\log(\sigma^2_{\epsilon,t}) = \mu_\epsilon + \phi_\epsilon[\log(\sigma^2_{\epsilon,t-1}) - \mu_\epsilon] + \xi_{\epsilon,t},$$
$$\xi_{\epsilon,t} \sim N(0, \sigma^2_\eta). \tag{5}$$

We set the following values: $\mu_\epsilon = 0$, $\phi_\epsilon = 0.9$, and $\sigma^2_{\epsilon,t} = 0.5$. This creates high auto-correlation which causes regions of high/low volatility which can occur frequently in real world data. We used 4 magnitude of change values of $\{0.5, 1, 1.5, 2\}$ to evaluate the algorithms' effectiveness in varying signal-to-noise ratios. The results are reported in Table 4. To be fair to PELT

Table 4. Change in mean with heterogeneity.

MC	Algorithms	Rand Avg.	Adj. Rand Avg.	Precision	Recall	F1-score
2	DC-DS	**0.982**$_{(0.004)}$	**0.964**$_{(0.008)}$	**0.89**	0.90	**0.89**
	PELT	0.834$_{(0.010)}$	0.667$_{(0.019)}$	0.11	**0.96**	0.20
	R-FPOP	0.944$_{(0.009)}$	0.889$_{(0.018)}$	0.73	0.85	0.79
1.5	DC-DS	**0.977**$_{(0.004)}$	**0.955**$_{(0.008)}$	**0.87**	**0.90**	**0.88**
	PELT	0.827$_{(0.010)}$	0.654$_{(0.020)}$	0.10	0.89	0.19
	R-FPOP	0.926$_{(0.011)}$	0.852$_{(0.022)}$	0.64	0.81	0.72
1	DC-DS	**0.931**$_{(0.009)}$	**0.862**$_{(0.020)}$	**0.60**	**0.64**	**0.62**
	PELT	0.800$_{(0.010)}$	0.599$_{(0.020)}$	0.07	0.61	0.13
	R-FPOP	0.835$_{(0.019)}$	0.671$_{(0.037)}$	0.39	0.50	0.44
0.5	DC-DS	**0.836**$_{(0.015)}$	**0.673**$_{(0.030)}$	**0.29**	**0.38**	**0.33**
	PELT	0.686$_{(0.013)}$	0.373$_{(0.026)}$	0.03	0.25	0.05
	R-FPOP	0.626$_{(0.018)}$	0.254$_{(0.036)}$	0.12	0.11	0.12

Table details decoupled approach with dynamic shrinkage (DC-DS), PELT, and R-FPOP on simulated data with one change in mean of varying magnitudes (MC) and stochastic volatility. Stochastic volatility is simulated using highly autocorrelated SV(1) model. Rand average and adjusted Rand average measures the similarity between predicted partition and true partition. Standard error for Rand average and adjusted Rand average are given in subscripts. F1-score measures accuracy of changepoint detection through a comparison of precision and recall. Bolded values indicate best results for the metric in the column.

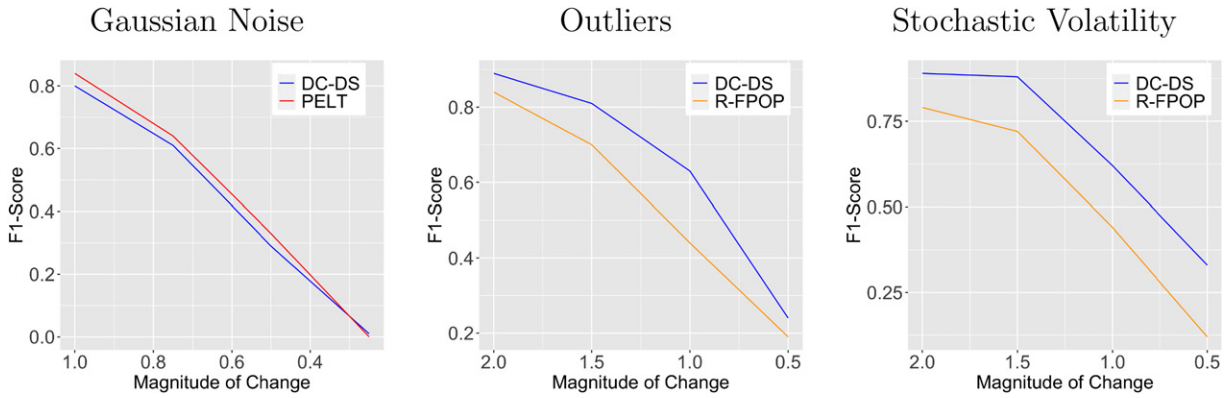

Figure 2. Change in Mean, comparison of F1-Scores. F1-score calculates the harmonic mean between precision and recall. The score ranges between 0 and 1 with 1 being a perfect prediction. The left plot shows F1-score of DC-DS against PELT in simulated data with Gaussian noise from Section 3.2. The middle plot shows F1-score of DC-DS against R-FPOP in simulated data with outliers from Section 3.3. The right plot shows F1-score of DC-DS against R-FPOP in simulated data with stochastic volatility from Section 3.4.

and R-FPOP, the algorithms are not intended to work in this setting. As a result, the performances are not reflective of the effectiveness of the algorithms.

Comparing results in Table 4 to Table 2 for magnitude of change 1, we see that the performance of all changepoint algorithms decreased in presence of stochastic volatility. This is to be expected as stochastic volatility makes detection of changepoints much more difficult. With the addition of stochastic volatility, DC-DS outperformed other competing changepoint methods in all settings. DC-DS achieves an F1-score of 0.89 for setting of magnitude of change of 2, an F1-score of 0.62 for setting of magnitude of change of 1 and an F1-score of 0.33 for setting of magnitude of change of 0.5. DC-DS achieves the most accurate partitions by having the highest adjusted Rand average and the best tradeoffs of precision/recall. This illustrates the robustness of the decoupled approach in dealing with heterogeneity.

Figure 2 summarizes the results in term of F1-score for Section 3. As seen in the plots, the decoupled approach is slightly worse in standard Gaussian noise settings but performs significantly better when outliers or stochastic volatility are added. This illustrates the tradeoff of the decoupled approach. By fitting a Bayesian DLM to the data first, the decoupled approach can be more locally adaptive to the complexities inherent in time series data. Outliers and heterogeneity are just two examples of the challenges that the decoupled approach can deal with. As long as the posterior estimates for the $\{\beta_t\}$ process remains relatively smooth, the decoupled loss can identify correct changepoint locations in a variety of complex scenarios.

4. Global Land Surface Air Temperature Anomaly

For an illustrative application we consider monthly global land surface air temperature anomaly with reference period 1951–1980 in 0.01 degrees Celsius from 1880 to 2018 (*https://data.giss.nasa.gov/gistemp/tabledata_v3/GLB.Ts.txt*). The urgency to detect sudden shifts in climate patterns has been growing amidst ongoing human-induced change. As shown, there are clear long term linear time trends underlying local annual trends in the data. Overall global temperatures appear increasing over time; however, three features make standard changepoint analysis difficult. First, there exists seasonal fluctuation in the data, and these seem somewhat irregular. This implies the local trend is not flat but rather a smooth curve fluctuating through the months. Second, there is differing levels of variability over time. Third, there may be anomalies throughout the data as a result of certain global events.

As seen in the top-left plot of Figure 3, the underlying signal fluctuates over time as a result of irregular cyclical patterns over the years; these patterns have less variability than the longer term approximately linear time trends. This results in a wiggly

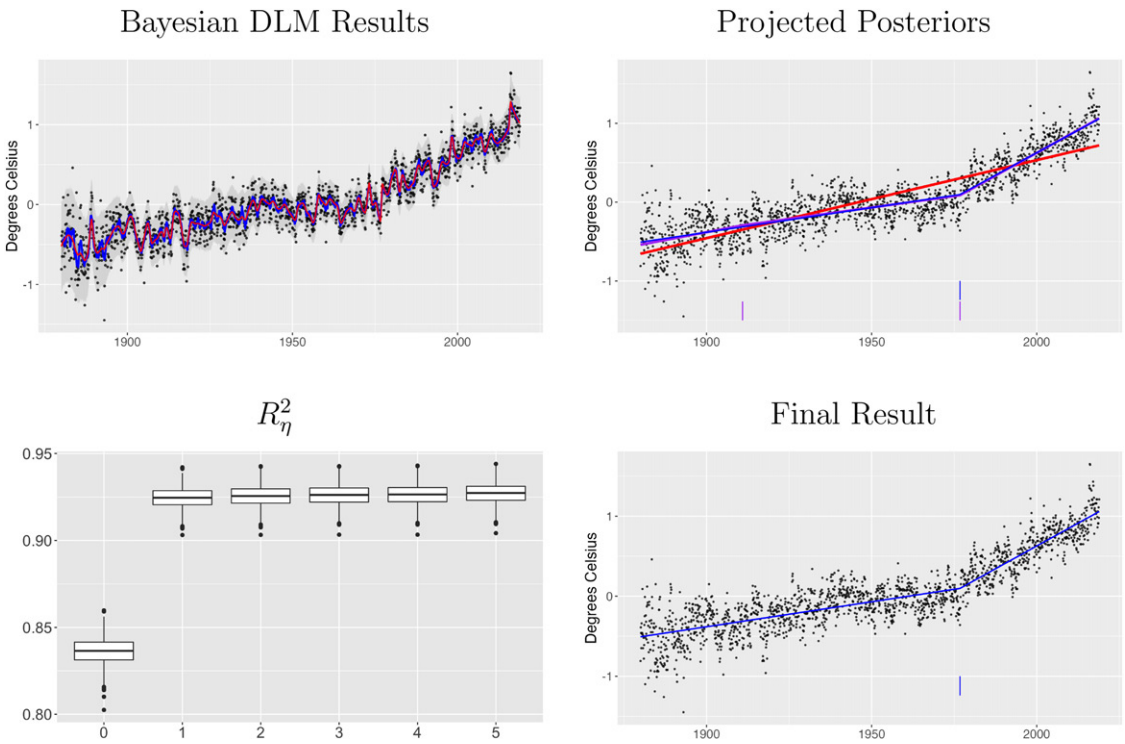

Figure 3. Monthly Global Land Surface Air Temperature Anomaly: The top-left figure shows the monthly average global land surface air temperature anomaly from 1880 to 2018 with 10 year moving average and Bayesian DLM fit. Additionally, the inner ribbon is 95% credible bands for $\{\beta_t\}$ and the outer ribbon is 95% credible bands for $\{\beta_t + \epsilon_t\}$ from the Bayesian DLM. The top-right figure shows the mean of "projected posterior" for $\{0, 1, 2\}$ changepoints. The predicted changepoint locations are shown by the vertical lines. The bottom left plot illustrates the distribution R^2_η for various number of changepoints. The bottom right plot shows the final result for the decoupled approach.

fit from the Bayesian dynamic linear model with $D = 2$. Using the decoupled approach, we can visualize different fits of the projected posterior. The top-right plot of Figure 3 illustrates the mean of the "projected posterior" for $\{0, 1, 2\}$ number of changepoints. As the number of changepoints increase, the fit becomes increasingly better. This is because increasing the number of changepoints essentially increases the degrees of freedom for the "projected posterior." We select 1 as the optimal number of changepoints as it's the simplest fit in which the upper 90% credible interval exceeds the 0.9 threshold. We estimate the single changepoint at November, 1976, after which there is a steeper long term slope. Our changepoint time aligns well with a recognized regime shift in the 1976–1977 winter originally determined by climate scientists in the 1990s based on multiple signals, but attributed to the North Pacific (Hare and Mantua 2000). Several changepoint analyses of temperature anomalies or multiple climate measures identify at least one shift in the 1970s decade (Alley et al. 2003; Ivanov and Evtimov 2010; Matyasovszky 2011; Yang and Song 2014). More real world applications involving changes in dynamic regression are shown in the supplementary materials, Section E.

5. Conclusion

In conclusion, this article proposes a decoupled approach for changepoint analysis that separates the processing of modeling and inference. As seen throughout the simulations and real world examples, the decoupled approach offers several key advantages over the competing method. First, by separating the process of modeling and inference, the decoupled approach allows for fitting of a highly complex Bayesian model to the underlying data while still allowing for reasonable inference of changepoints. This allows the decoupled approach to deal with many complexities inherent in time series. As the data becomes increasingly complex, the decoupled approach can adapt the Bayesian DLM to deal with these issues while maintaining the same inference process for changepoints. Additionally, Bayesian modeling frameworks for other challenging time series data such as data with varying degrees of sparsity can be used in the first stage of the decoupled approach as long as it gives estimates of the trend at the desired inference times.

Second, the decoupled approach is flexible in its ability to identify different types of changepoints. From the examples shown in the article and the supplementary materials, the decoupled approach has the ability to identify changes in mean, changes in regression coefficients and changes in higher order trends. Most other changepoint algorithms can only be used for one specific scenario. Lastly, the Bayesian decoupled approach maintains the ability to quantify uncertainty of parameters and derived quantities as compared to traditional changepoint algorithms. The flexibility of the decoupled approach allows the algorithm to be more adaptive to a wide variety of datasets and scientific conclusions.

Supplementary Materials

Narrative Supplement: Online supplement containing an in-depth loss derivation, various methodological extensions, a detailed explanation of the Bayesian DLM framework and more simulation/real world results.

Code: Rmd file containing the code for setting up and running the decoupled approach.

Acknowledgments

We would like to thank the editor, the associate editor and the referees for their thoughtful feedback and recommendations for helping improve the article.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

The authors gratefully acknowledge financial support from the National Science Foundation 1455172, 1934985, 1940124, 1940276, 2114143, and USAID 7200AA18CA00014.

References

Alley, R. B., Marotzke, J., Nordhaus, W. D., Overpeck, J. T., Peteet, D. M., Pielke, R. A., Pierrehumbert, R. T., Rhines, P. B., Stocker, T. F., Talley, L. D., and Wallace, J. M. (2003), "Abrupt Climate Change," *Science*, 299, 2005–2010. DOI:10.1126/science.1081056. Available at *https://www.science.org/doi/abs/10.1126/science.1081056*. [30]

Aminikhanghahi, S., and Cook, D. J. (2017), "A Survey of Methods for Time Series Change Point Detection," *Knowledge Information System*, 51, 339–367. [23]

Bhadra, A., Datta, J., Polson, N. G., and Willard, B. (2016), "Default Bayesian Analysis with Global-Local Shrinkage Priors," *Biometrika*, 103, 955–969. [26]

Bitto, A., and Fruhwirth-Schnatter, S. (2019), "Achieving Shrinkage in a Time-Varying Parameter Model Framework," *Journal of Econometrics*, 210, 75–97. [23]

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2009), "Handling Sparsity via the Horseshoe," *AISTATS*, 5, 73–80. [23]

Chan, J. C. C., and Eisenstat, E. (2018), "Bayesian Model Comparison for Time-Varying Parameter Vars with Stochastic Volatility," *Journal of Applied Econometrics*, 33, 509–532. [23]

Fearnhead, P., and Rigaill, G. (2019), "Changepoint Detection in the Presence of Outliers," *Journal of the American Statistical Association*, 114, 169–183. [27]

Florian Huber, G. K., and Onorante, L. (2021), "Inducing Sparsity and Shrinkage in Time-Varying Parameter Models," *Journal of Business & Economic Statistics*, 39, 669–683. DOI:10.1080/07350015.2020.1713796 [23]

Fryzlewicz, P. (2014), "Wild Binary Segmentation for Multiple Change-Point Detection," *Annals of Statistics*, 42, 2243–2281. [23]

Hahn, P. R., and Carvalho, C. M. (2015), "Decoupling Shrinkage and Selection in Bayesian Linear Models: A Posterior Summary Perspective," *Journal of American Statistical Association*, 110, 435–448. [23,24,25]

Hare, S. R., and Mantua, N. J. (2000), "Empirical Evidence for North Pacific Regime Shifts in 1977 and 1989," *Progress in Oceanography*, 47, 103–145. DOI:10.1016/S0079-6611(00)00033-1. Available at *https://www.sciencedirect.com/science/article/pii/S0079661100000331*. [30]

Hubert L., and Arabie, P. (1985), "Comparing Partitions," *Journal of Classification*, 2, 193–218. DOI:10.1007/BF01908075 [27]

Ivanov, M. A., and Evtimov, S. N. (2010), "1963: The Break Point of the Northern Hemisphere Temperature Trend During the Twentieth Century," *International Journal of Climatology*, 30, 1738–1746. DOI:10.1002/joc.2002. *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.2002*. [30]

James, N. A., and Matteson, D. S. (2014), "ecp: An r Package for Nonparametric Multiple Change Point Analysis of Multivariate Data," *Journal of Statistical Software*, 62, 1–25. [23]

Jeske, D. R., De Oca, V. M., Bischoff, W., and Marvasti, M. (2009), "Cusum Techniques for Timeslot Sequences with Applications to Network Surveillance," *Computational Statistics and Data Analysis*, 53, 4332–4344. [23]

Kiers, H. A. L. (1997), "Weighted Least Squares Fitting Using Ordinary Least Squares Algorithm," *Psychometrika*, 62, 251–266. [24]

Killick, R., Fearnhead, P., and Eckley, A. (2012), "Optimal Detection of Changepoints with a Linear Computational Cost," *Journal of American Statistical Association*, 107, 1590–1598. [23,27]

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with Arch Models," *Review of Economic Studies*, 65, 361–393. [28]

Kowal, D., Matteson, D., and Ruppert, D. (2019), "Dynamic Shrinkage Process," *Journal of the Royal Statistical Society*, Series B, 81, 781–804. [24,26]

Maidstone, R., Hocking, T., Rigaill, G., and Fearnhead, P. (2017), "On Optimal Multiple Changepoint Algorithms for Large Data," *Statistics and Computings*, 27, 519–533. [23]

Matteson, D. S., and James, N. A. (2014), "A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data," *Journal of the American Statistical Association*, 109, 334–345. [23]

Matyasovszky, I. (2011), "Detecting Abrupt Climate Changes on Different Time Scales," *Theoretical and Applied Climatology*, 105, 445–454. DOI:10.1007/s00704-011-0401-4 [30]

Nakajima, J., and West, M. (2013), "Bayesian Analysis of Latent Threshold Dynamic Models," *Journal of Business & Economic Statistics*, 31, 151–164. DOI:10.1080/07350015.2012.747847 [26]

Ray, P., and Bhattacharya, A. (2018), "Signal Adaptive Variable Selector for the Horseshoe Prior." [23,25]

Rockova, V., and McAlinn, K. (2021), "Dynamic Variable Selection with Spike-and-Slab Process Priors," *Bayesian Analysis*, 16, 233–269. DOI:10.1214/20-BA1199 [26]

Tay, J. K., Narasimhan, B., and Hastie, T. (2023), "Elastic Net Regularization Paths for All Generalized Linear Models," *Journal of Statistical Software*, 106, 1–31. DOI: 10.18637/jss.v106.i01 [25]

van den Burg, G. J. J., and Williams, C. K. I. (2020), "An Evaluation of Change Point Detection Algorithms," arXiv preprint arXiv:2003.06222. [27]

Woody, S., Carvalho, C. M., and Murray, J. S. (2021), "Model Interpretation through Lower-Dimensional Posterior Summarization," *Journal of Computational and Graphical Statistics*, 30, 144–161. [25]

Wu, H., and Matteson, D. S. (2020), "Adaptive Bayesian Change-point Analysis and Local Outlier Scoring," arXiv preprint arXiv:2011.09437. [26]

Yang, Y., and Song, Q. (2014), "Jump Detection in Time Series Nonparametric Regression Models: A Polynomial Spline Approach," *Annals of the Institute of Statistical Mathematics*, 66, 325–344. [30]

Zou, H. (2006), "The Adaptive Lasso and its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429. [25]