Hybrid Statistical Forecasting of Daily Screen Time: A TV-Regularized GLM and Prophet Approach

Rae Song [21198617] Sophia Feng [20823681]

Abstract

Forecasting daily electronic screen time presents significant statistical challenges due to small sample sizes, strong weekly seasonality, and abrupt behavioral shifts driven by external context, such as academic deadlines. This study analyzes longitudinal daily screen-time data collected from two university students over a twelve-week academic period. We evaluate the limitations of standard forecasting baselines—Seasonal Naive, SARIMA, and standard Prophet—and propose three progressively expressive statistical frameworks to address non-stationarity and structural breaks. First, we introduce a Total-Variation Regularized Generalized Linear Model (TV-GLM) that incorporates Fourier seasonality and specific academic event indicators. Second, we enhance the Prophet forecasting procedure by augmenting it with exogenous regressors representing assignment urgency and "cramming" behavior. Finally, we construct a sequential Hybrid Prophet-GLM model that utilizes Prophet as a non-linear seasonal backbone while learning structured residual corrections via the TV-GLM. Our results demonstrate that the hybrid approach yields the highest predictive accuracy across both participants, achieving the lowest Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) compared to all other methods. By effectively combining smooth seasonal trends with sharp, event-driven corrections, the hybrid model offers a robust and interpretable solution for small-sample personal behavioral forecasting.

1 Introduction

Understanding and forecasting daily electronic device screen-time has become an important topic in behavioral analytics and digital well-being. Screen-time patterns exhibit strong weekly structure, sharp deviations associated with academic deadlines, and occasional abrupt changes in behavior, making the forecasting task both practically relevant and statistically challenging. In this study, we analyze the daily screen-time data of two university students collected over a twelve-week academic period. Each day's observation records total minutes spent on electronic devices (smartphone, laptop, and tablet), together with contextual indicators such as weekday/weekend status, reading week, exam week, and assignment deadlines. We use the first ten weeks for model training and the final two weeks for forecasting evaluation.

The dataset presents several statistical challenges typical of human behavioral time series: (i) a **small sample size** (fewer than 90 observations per participant), (ii) evident **non-stationarity** driven by the academic calendar, (iii) **heavy-tailed outliers** during periods of peak workload, and (iv) **abrupt structural changes** that smooth-trend models cannot express. These characteristics limit the effectiveness of classical forecasting tools such as Seasonal Naive, SARIMA, and standard Prophet, which either lack interpretability or struggle to capture sudden behavioral shifts.

To address these challenges, we begin by developing an interpretable **Total-Variation Penalized Generalized Linear Model (TV-GLM)** incorporating weekly Fourier features, lagged terms, and academic-calendar indicators. After comparing this model against standard forecasting baselines (Seasonal Naive, SARIMA, and standard Prophet), we observe that Prophet performs comparatively well, motivating a deeper exploration of Prophet-based approaches. This leads us to extend Prophet with exogenous academic features and, ultimately, to construct a hybrid model that combines Prophet's smooth seasonal structure with a correction term learned through our GLM framework.

Across both participants, this modeling progression yields substantial improvements in forecasting accuracy, with the final hybrid model achieving the lowest error among all methods while preserving interpretability through explicit associations with academic events. The results highlight the value of combining structured time-series decomposition with event-aware statistical modeling in small-sample behavioral forecasting tasks.

2 Background

Daily screen-time forecasting connects to several strands of prior work in time-series modeling, statistical change-point detection, and digital phenotyping. Practical seasonal forecasting tools such as Prophet capture weekly periodicity effectively but impose smooth trend components, which can obscure abrupt regime shifts commonly found in individual behavioral data (Taylor & Letham, 2018). Classical deep sequence models and attention-based architectures can achieve strong predictive accuracy when long histories and large datasets are available, but they are data-hungry and offer limited interpretability regarding when and why behavior changes (Lim et al., 2021).

Statistical change-point methods provide a complementary perspective. The fused-lasso lineage yields convex estimators capable of recovering piecewise-constant structure with formal guarantees (Tibshirani et al., 2005). Recent extensions adapt these penalties to generalized linear models and count data, enabling interpretable, date-level identification of behavioral shifts (Ohishi, 2024; Wang et al., 2022; Zhang et al., 2024). This is particularly relevant for screen-time, where usage patterns often change at specific academic events such as exams or deadlines.

Digital phenotyping research further shows that smartphone logs encode meaningful variation in daily stress, sleep patterns, and symptom trajectories (Aalbers et al., 2023; Currey et al., 2023). These studies highlight the importance of modeling individual-level behavior rather than relying solely on population-level averages.

A practical constraint in this project is the **limited availability of suitable external data**. Although large public smartphone-usage datasets exist, most provide only cross-sectional aggregates (one record per user) or irregular multi-user activity logs that cannot be aligned with the day-level forecasting task in this study. As a result, no external dataset could be incorporated without violating temporal structure or introducing incompatible measurement definitions. This motivates methods that remain statistically stable in extremely small-sample settings while still capturing both weekly seasonality and event-driven structural changes.

Taken together, these lines of work motivate the modeling approach in this project: combining event-aware regression models with mechanisms for detecting abrupt behavioral shifts, while benchmarking against established seasonal forecasting tools such as Prophet and SARIMA.

3 Methods

We observe daily total screen-time for two participants, Rae and Sophia, over $T^{(i)}$ consecutive days indexed by $t = 1, \ldots, T^{(i)}$ for person $i \in \{\text{Rae}, \text{Sophia}\}$. Let $y_t^{(i)} \in [0, 1440]$ denote the total minutes of screen-time on day t, and let $\text{dow}(t) \in \{0, \ldots, 6\}$ be the day-of-week. For each participant we use the first ten weeks as a training window and hold out the final two weeks for evaluation. All models are trained separately for each participant; in what follows we suppress the superscript (i) when it is clear from context.

3.1 Feature construction

Our goal is to build a small but expressive feature vector that captures weekly seasonality, academic events, assignment urgency, and short-term persistence. For each calendar date we engineer the following components.

Calendar and event indicators. From the calendar we compute

$$\mathrm{dow}_t = \mathrm{dow}(t), \qquad \mathrm{isWeekend}_t = \mathbf{1}\{\mathrm{dow}_t \geq 5\},$$

together with binary indicators

$$\mathrm{Exam}_t,\ \mathrm{Reading}_t,\ \mathrm{Assignment}_t \in \{0,1\}$$

flagging exam week, reading week, and whether an assignment is due on day t (coded as 1 on the due date and the three days leading up to the deadline).

To capture last-minute "cramming" behavior we further derive two urgency variables. If $\mathrm{Assignment}_t = 1$ then we set

$$\mathrm{UrgentPrep}_{t-1} = 1, \qquad \mathrm{UrgentDeadline}_t = 1,$$

and zero otherwise, subject to remaining within the observation window. The interaction terms

$$\mathrm{Weekend}{\times}\mathrm{Prep}_t = \mathrm{isWeekend}_t{\cdot}\mathrm{UrgentPrep}_t, \qquad \mathrm{Weekend}{\times}\mathrm{Deadline}_t = \mathrm{isWeekend}_t{\cdot}\mathrm{UrgentDeadline}_t$$

allow weekend usage to respond differently when assignments are imminent.

Weekly seasonality. We encode smooth weekly seasonality using a Fourier pair

$$\cos_t = \cos\left(\tfrac{2\pi\,\mathrm{dow}_t}{7}\right), \qquad \sin_t = \sin\left(\tfrac{2\pi\,\mathrm{dow}_t}{7}\right).$$

Lagged usage and local slope. Short-term persistence and momentum are summarized via one-day and one-week lags together with a finite-difference velocity:

$$\ell_t^{(1)} = y_{t-1}, \qquad \ell_t^{(7)} = y_{t-7}, \qquad v_t = y_{t-1} - y_{t-2}.$$

These lags are available once $t \geq 8$.

Final feature vector. Collecting all pieces, for $t \geq 8$ we define a $d = 13$ dimensional feature vector

$$\begin{aligned} x_t = \Big[&\cos_t,\ \sin_t,\ \mathrm{isWeekend}_t,\ \mathrm{Exam}_t,\ \mathrm{Reading}_t,\ \mathrm{Assignment}_t,\ \mathrm{UrgentPrep}_t,\ \mathrm{UrgentDeadline}_t, \\ &\mathrm{Weekend} \times \mathrm{Deadline}_t,\ \mathrm{Weekend} \times \mathrm{Prep}_t,\ \ell_t^{(1)},\ \ell_t^{(7)},\ v_t\Big]^\top \in \mathbb{R}^{13}. \end{aligned} \tag{1}$$

3.2 Proposed model 1: TV-regularized weighted GLM

Our first model treats y_t as a count and uses a Poisson GLM with a time-varying intercept path μ_t that is encouraged to be piecewise constant via a total-variation (TV) penalty. Conditioned on the linear predictor η_t we assume

$$y_t \mid \lambda_t \sim \mathrm{Poisson}(\lambda_t), \qquad \log \lambda_t = \eta_t = x_t^\top \beta + \mu_t, \tag{2}$$

where $\beta \in \mathbb{R}^{13}$ are global coefficients shared across the training window and μ_t captures slow baseline drifts and structural regime changes. To downweight ordinary days and emphasize exam and deadline periods, we introduce observation weights

$$w_t = \begin{cases} 3, & \text{Exam}_t = 1 \text{ or UrgentDeadline}_t = 1, \\ 1, & \text{otherwise.} \end{cases}$$

For a given participant we estimate $(\beta, \mu_{8:T})$ by minimizing the weighted negative log-likelihood with TV and ridge regularization:

$$\min_{\beta,\, \mu_{8:T}} \mathcal{L}_{\text{Pois}}(\beta, \mu) + \lambda_{\text{TV}} \sum_{t=9}^{T} |\mu_t - \mu_{t-1}| + \alpha \|\beta\|_2^2, \tag{3}$$

where

$$\mathcal{L}_{\text{Pois}}(\beta, \mu) = \sum_{t=8}^{T} w_t \big(\exp(\eta_t) - y_t \eta_t\big) \qquad \text{with} \quad \eta_t = x_t^\top \beta + \mu_t. \tag{4}$$

The TV term encourages μ_t to be piecewise constant; locations where μ_t jumps correspond to estimated change-points in baseline screen-time. The objective (3) is convex in (β, μ) and is solved with `cvxpy`; we tune the TV strength λ_{TV} by blocked, time-aware cross-validation on the training window while fixing the ridge parameter α to a small constant.

Forecasting. Given $(\hat{\beta}, \hat{\mu}_t)$ on the training window, we forecast the test period autoregressively. For each future day we form x_t using the most recent available lags (including previous predictions) and propagate the intercept by extrapolating the recent trend of $\hat{\mu}_t$:

$$\hat{\mu}_t = \bar{\mu}_{\text{last3}} + s_\mu \,(t - T_{\text{train}})\, 0.95^{(t - T_{\text{train}})},$$

where s_μ is the average slope over the last two weeks of training and $\bar{\mu}_{\text{last3}}$ is the mean of the last three intercept values. The GLM forecast is then

$$\hat{y}_t^{\text{GLM}} = \min\big\{\exp(x_t^\top \hat{\beta} + \hat{\mu}_t),\, 1440\big\}.$$

Baseline models. We compare the weighted TV-GLM against three standard seasonal baselines:

- **Seasonal naive.** $\hat{y}_t = y_{t-7}$ for all test days t, i.e. we repeat last week's same weekday.
- **Prophet** Prophet with weekly seasonality enabled and no additional regressors; we fit on the training window and use the built-in forecast for the test horizon.
- **SARIMA.** A seasonal ARIMA model $\text{ARIMA}(1,0,1) \times (0,1,0)_7$ fitted to the training series using `statsmodels`.

We evaluate all methods using mean absolute error (MAE) and root mean squared error (RMSE) on the held-out final two weeks. Table 1 and Table 2 report results for Rae and Sophia.

Table 1: Rae: comparison of TV-regularized weighted GLM with standard seasonal baselines.

Method	MAE	RMSE
Seasonal Naive	330.07	367.96
Prophet	197.83	247.37
SARIMA	262.52	320.40
Proposed 1: Weighted GLM	**138.77**	**179.68**

Table 2: Sophia: comparison of TV-regularized weighted GLM with standard seasonal baselines.

Method	MAE	RMSE
Seasonal Naive	414.00	467.24
Prophet	271.96	309.32
SARIMA	323.08	389.65
Proposed 1: Weighted GLM	**92.62**	**141.39**

Across both participants the TV-GLM substantially improves over all three baselines, reducing MAE by 30-50% relative to the strongest baseline. Even though our Weighted GLM model achieved the lowest MAE and RMSE, we took a further look to the results of all baseline results. Among the baselines, Prophet consistently performs best, which motivates using Prophet as a flexible seasonal backbone in our subsequent models.

3.3 Proposed model 2: Augmented Prophet

The second model keeps Prophet as the main forecasting engine but augments it with the same calendar and urgency features used in the GLM. In the standard Prophet formulation, the observed series is decomposed as

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t, \tag{5}$$

where $g(t)$ is a (piecewise) linear or logistic growth trend, $s(t)$ captures smooth seasonality via Fourier terms, $h(t)$ represents the effect of user-specified holidays or events, and ε_t is an error term. We configure Prophet with weekly seasonality and then add a vector of exogenous regressors z_t consisting of

$$z_t = \begin{matrix}[\text{Exam}_t,\ \text{Reading}_t,\ \text{Assignment}_t,\ \text{UrgentPrep}_t, \\ \text{UrgentDeadline}_t,\ \text{Weekend} \times \text{Deadline}_t,\ \text{Weekend} \times \text{Prep}_t]^\top.\end{matrix}$$

The augmented model can be written as

$$y_t = g(t) + s_{\text{weekly}}(t) + z_t^\top \gamma + \varepsilon_t, \tag{6}$$

where γ are regression coefficients estimated jointly with the standard Prophet components using its underlying Bayesian procedure.

We fit the augmented Prophet on the training window and forecast the final two weeks. Table 3 compares the augmented Prophet to the original Prophet and to the TV-GLM.

Table 3: Augmented Prophet versus other methods.

	Rae		Sophia	
Method	MAE	RMSE	MAE	RMSE
Seasonal Naive	330.07	367.96	414.00	467.24
Prophet	197.83	247.37	271.96	309.32
SARIMA	262.52	320.40	323.08	389.65
Proposed 1: Weighted GLM	**138.77**	**179.68**	**92.62**	**141.39**
Proposed 2: Aug Prophet	175.51	202.86	147.50	175.04

For both participants, augmenting Prophet with academic and urgency regressors significantly improves over the baseline Prophet configuration (roughly 10-30% lower MAE). Nevertheless, the TV-GLM still achieves the lowest error, suggesting that explicit lags and a TV-penalized intercept capture additional structure beyond what Prophet's smooth trend and seasonality can represent.

3.4 Proposed model 3: Hybrid Prophet-GLM

Our final method combines the strengths of the two previous approaches. Intuitively, Prophet provides a flexible nonlinear seasonal backbone, while the GLM excels at modeling sharp regime changes and lagged effects. We therefore use the augmented Prophet as a base forecaster and let a TV-GLM learn structured corrections to its residuals.

Training. Let $\hat{y}_t^{\text{Prop}}$ denote the in-sample fitted values from the augmented Prophet model (6) on the training window, and define residuals

$$r_t = y_t - \hat{y}_t^{\text{Prop}}.$$

Because r_t can be positive or negative, we treat residual modeling as a Gaussian regression problem. Using the same feature vector x_t as in (1) and weights w_t, we fit a TV-regularized residual GLM

$$r_t \approx x_t^\top \beta^{\text{res}} + \mu_t^{\text{res}}, \tag{7}$$

by minimizing the weighted least-squares objective

$$\min_{\beta^{\text{res}},\, \mu_{8:T}^{\text{res}}} \sum_{t=8}^{T} w_t \big(r_t - x_t^\top \beta^{\text{res}} - \mu_t^{\text{res}}\big)^2 \;+\; \lambda_{\text{TV}}^{\text{res}} \sum_{t=9}^{T} \big|\mu_t^{\text{res}} - \mu_{t-1}^{\text{res}}\big| \;+\; \alpha_{\text{res}} \big\|\beta^{\text{res}}\big\|_2^2. \tag{8}$$

Again, the TV penalty induces piecewise-constant segments in the residual intercept, which can be interpreted as structural misspecification of the Prophet backbone.

Hybrid forecasting. On the test window, we first obtain the augmented Prophet forecast $\hat{y}_t^{\text{Prop}}$ using the fitted Prophet model and the observed regressors z_t. In parallel, we forecast residuals using the GLM:

$$\hat{r}_t^{\text{GLM}} = x_t^\top \hat{\beta}^{\text{res}} + \hat{\mu}_t^{\text{res}},$$

where x_t uses autoregressive lags built from the hybrid's own past predictions, and $\hat{\mu}_t^{\text{res}}$ is extrapolated from the training intercept path in the same way as in the pure GLM case. The final hybrid prediction adds the GLM correction to the Prophet baseline,

$$\hat{y}_t^{\text{Hybrid}} = \hat{y}_t^{\text{Prop}} + \hat{r}_t^{\text{GLM}}, \qquad \hat{y}_t^{\text{Hybrid}} \leftarrow \min\{\max(\hat{y}_t^{\text{Hybrid}}, 0), 1440\}. \tag{9}$$

Hybrid performance. Table 4 summarizes the full comparison for both participants, including all baselines and our three proposed methods.

Table 4: Final performance comparison for Rae and Sophia (test window).

	Rae		Sophia	
Method	MAE	RMSE	MAE	RMSE
Seasonal Naive	330.07	367.96	414.00	467.24
Prophet	197.83	247.37	271.96	309.32
SARIMA	262.52	320.40	323.08	389.65
Proposed 1: Weighted GLM	138.77	179.68	92.62	141.39
Proposed 2: Augmented Prophet	175.51	202.86	147.50	175.04
Proposed 3: Hybrid (Prophet - GLM)	**109.57**	**143.33**	**87.49**	**128.62**

For both Rae and Sophia the hybrid model achieves the lowest MAE and RMSE, improving substantially over the best purely Prophet-based configuration and over the standalone TV-GLM. Qualitatively, the hybrid forecasts track the peaks and dips of the held-out weeks more closely than any individual component model: Prophet captures smooth weekly structure, while the residual GLM sharpens responses around exams and assignment deadlines and adjusts for short-term momentum via lagged features. This complementary combination motivates our hybrid approach as the final recommended model.

4 Discussion

The qualitative forecast traces in Figures 1 illustrate the behavior of all competing models on the held-out final two weeks. Across both participants the hybrid Prophet-GLM model most closely follows the observed day-to-day movements:

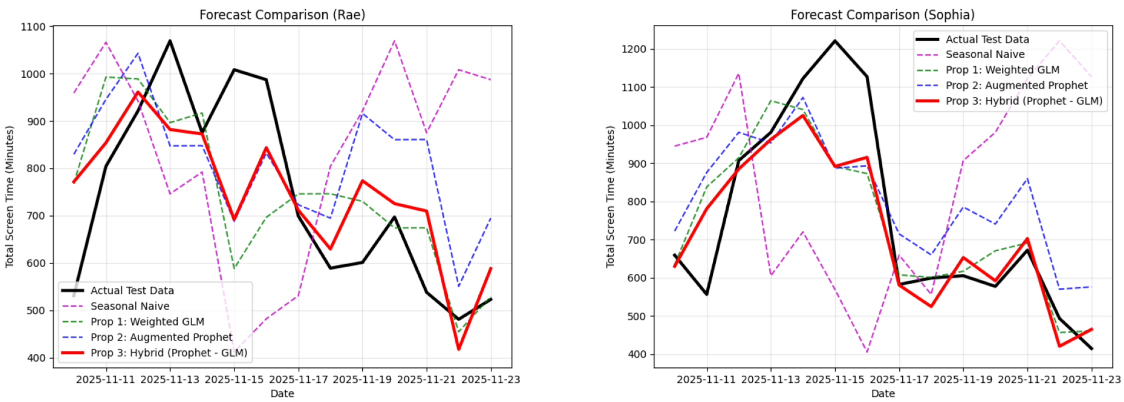

Figure 1: Forecast Comparison

From a quantitative standpoint, the hybrid model achieves MAE values of approximately 110 minutes for Rae and 87 minutes for Sophia. At first glance these errors may appear large, but they must be interpreted relative to each participant's baseline usage. Rae's mean daily screen time is 678 minutes and Sophia's is 710 minutes, so an MAE in the range of 80-110 minutes corresponds to only a 12-16% relative error. Given the extreme variability of human digital behavior and the very small sample size available for training (roughly ten weeks per participant), this level of accuracy is reasonable. Indeed, the hybrid model reduces error substantially relative to all baselines, often by 40-50%.

The challenges encountered in this project highlight why similar forecasting tasks are rarely studied using fine-grained, person-specific daily data. Unlike aggregate screen-time surveys or cross-sectional datasets, where each individual contributes a single summary statistic, our setting requires modeling a highly personalized, stochastic process. Behavioral patterns differ markedly across users: Rae and Sophia exhibit distinct volatility, weekly rhythms, stress responses, and sensitivity to academic deadlines. These idiosyncratic patterns limit the feasibility of training a single universal model and partly explain the lack of publicly available datasets with the same longitudinal structure. Generalizing from one user to another remains difficult without substantially larger multiuser time-series corpora.

4.1 Limitations and future work

The primary limitation of this study is the small sample size: two participants and twelve weeks per user. This restricts both the statistical power of model comparison and the stability of estimated effects, especially for models with nonlinear components such as Prophet. The scarcity of longitudinal datasets with comparable structure prevented us from augmenting training with external users or performing meaningful cross-person transfer learning. Larger multi-participant datasets, even with sparse or irregular observations, would enable hierarchical modeling, meta-learning, or regularized pooling strategies that share information across individuals while still allowing personalized dynamics.

Another limitation arises from unobserved covariates. Daily screen time is impacted by numerous latent factors (social activities, coursework load, mood, weather, sleep, app usage composition), none of which are captured in our dataset. Incorporating richer sensor streams or explicit academic calendars may significantly improve predictability.

Future work could explore (i) hierarchical Prophet or state-space extensions that borrow strength across users, (ii) neural temporal models with attention mechanisms that encode deadlines and behavioral cycles, (iii) Bayesian TV-GLMs that quantify uncertainty in change-point detection, and (iv) hybrid models that update dynamically as new data arrive online. More broadly, assembling larger and more diverse longitudinal screen-time datasets would substantially advance the empirical rigor of personal digital behavior modeling.

5 Conclusion

This project examined the problem of forecasting daily screen time for individual users using a combination of seasonal structure, academic schedule features, and autoregressive behavioral signals. Working with a small but detailed twelve-week dataset for two participants, we developed three increasingly expressive models: a TV-regularized weighted GLM, an augmented Prophet with academic and urgency regressors, and a sequential hybrid that combines Prophet's flexible seasonal backbone with a structured residual GLM. Across both participants, the hybrid approach achieved the lowest prediction error and most faithfully reproduced the dynamics of the held-out weeks, demonstrating the value of combining smooth global trends with sharp, feature-driven corrections.

Although absolute errors remain on the order of one to two hours, these represent modest relative deviations given each participant's high daily usage and the inherent variability of human digital behavior. The results highlight both the promise and the difficulty of personalized behavioral forecasting: individuals exhibit strong idiosyncrasies, yet meaningful structure can still be extracted from calendar cues, deadlines, and recent usage momentum. The methodological insights developed here, particularly the residual hybridization strategy, provide a foundation for future work on larger and richer longitudinal datasets, where personalization and generalization can be studied at scale.

References

Aalbers, G., Hendrickson, A. T., Vanden Abeele, M. M., & Keijsers, L. (2023). Smartphone-tracked digital markers of momentary subjective stress in college students: Idiographic machine learning analysis. *JMIR mHealth and uHealth, 11*. `https://doi.org/10.2196/37469`

Currey, D., & Torous, J. (2023). Digital phenotyping data to predict symptom improvement and mental health app personalization in college students: Prospective validation of a predictive model. *Journal of Medical Internet Research, 25*. `https://doi.org/10.2196/39258`

Lim, B., Arık, S., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting, 37*(4), 1748-1764. `https://doi.org/10.1016/j.ijforecast.2021.03.012`

Ohishi, M. (2024). Generalized fused lasso for grouped data in Generalized Linear Models. *Statistics and Computing, 34*(4). `https://doi.org/10.1007/s11222-024-10433-5`

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician, 72*(1), 37-45. `https://doi.org/10.1080/00031305.2017.1380080`

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(1), 91-108. `https://doi.org/10.1111/j.1467-9868.2005.00490.x`

Wang, X., Liu, B., Zhang, X., & Liu, Y. (2022). Efficient multiple change point detection for high-dimensional generalized linear models. *Canadian Journal of Statistics, 51*(2), 596-629. `https://doi.org/10.1002/cjs.11721`

Zhang, Z., Wang, X., Zhang, X., & Zhang, J. (2024). Simultaneously detecting spatiotemporal changes with penalized Poisson regression models. *Computational Statistics & Data Analysis, 212*, 108240. `https://doi.org/10.1016/j.csda.2025.108240`