Identification of Seasonality in Internet Traffic to Support Control of Online Advertising

Hao He[1] and Niklas Karlsson[2]

Abstract— **Feedback control is widely applied to the campaign management in online advertising. Learning the pattern of user traffic on Internet plays an important role in solving the control problem. In this paper, we focus on characterizing the seasonality, e.g., *time of day* (TOD) pattern of Internet user traffic for individual ad campaign. We model the seasonality using a truncated Fourier series with a set of amplitude and phase parameters. These seasonality parameters are estimated in a Bayesian framework using a *minimum mean square error* (MMSE) estimator, with their prior distribution learnt from historical data of a large number of campaigns. The proposed Bayesian method is shown to be robust and renders sensible seasonality for campaigns of disparate noise levels.**

I. INTRODUCTION

Online advertising is a fast growing industry and in order to deliver the campaign budget smoothly over time, as desired by the advertiser, it is critical to implement feedback control in the campaign management system. An early paper on feedback control applied to online advertising is available in [1], wherein several important challenges are outlined but detailed solutions are omitted. A more comprehensive and up-to-date overview of the control problem is available in [2]. The fact that the plant is unknown, dynamic, periodic, nonlinear, and in general discontinuous is a characteristic property of online advertising processes and is a fundamental challenge in the development of feedback control solutions. One of the main issues is that the user traffic to different web sites is volatile with stochastic effects as well as with trends and seasonality. In this paper, the Internet user traffic is represented by the number of impressions available during a fixed period of time, where an impression is one view of an ad. The seasonality, in particular, is dramatic and unless it is carefully accounted for during the feedback control, the advertiser may end up paying an unnecessary high price for impressions during hours of the day when the available number of impressions is low.

The idea of shaping the reference signal for ad campaign control is documented in [3]. In [4], the author proposed a control system to regulate a periodic plant subject to significant load disturbances and measurement noise, but with negligible dynamics between control input and output. The idea was further explored and the work was continued in [5]. However, there is no robust method proposed in all work on how to identify the seasonality of the plant for the reference signal. In this paper, we focus on developing an algorithm for seasonality identification of the periodic plant described in Section II. The more accurate the seasonality model is, the better a control system can be applied to the ad campaign.

In Verizon Media Demand Side Platform (DSP), feedback control is implemented on a campaign level. Thus, seasonality of the plant needs to be estimated for each ad campaign. In the meanwhile, campaigns in the same network (a group of timezones) exhibit strong similarity, due to the fact that the Internet traffic pattern, driven by human activity routine, is closely associated with the time zone. A good seasonality identification algorithm needs to be able to generate sensible model for each individual campaign based on its historical data and the network prior. The historical impression data can be of remarkably different noise level across campaigns, adding to the challenge of inference task.

There is a large literature on identifying patterns in time series data with seasonal component [6]. Traditionally, the autoregressive integrated moving average (ARIMA) model has been one of the most widely used linear models in time series forecasting [7]. Different variations of ARIMA combined with neural network have been proposed lately in [8], [9]. Fourier transformation is also widely utilized for extracting seasonal component of the time series [10]. However, in all previous work, seasonality parameters were estimated solely based on the historical data and can be highly unreliable when the data is extremely noisy, which is a common issue for small size campaigns in online advertising.

To have a more reliable algorithm to identify seasonality on campaign level, we propose a model in which robustness is introduced in two levels. A robust linear regression method is used to handle outliers in the impression data for initial assessment of seasonality parameters. Further, on a higher level we bring in an extra source of information into consideration, which is the network prior. Such Bayesian strategy benefits all campaigns in general, especially for campaigns with highly noisy data.

II. BACKGROUND

In [5], the ad optimization problem is turned into a control problem and solved using a periodic control system, where the periodicity of the plant is assumed to be a priori known.

Consider a plant which locally around the operating point can be described by a linear time-periodic model having insignificant dynamics but dramatic seasonality. The seasonality is in form of the T-periodic plant gain $K_p(1+h(t)) > 0$, where K_p is the constant component of the plant gain, and

[1]Hao He is with Demand Platforms R&D at Verizon Media, 701 1st Ave, Sunnyvale, California 94089 hao.he@verizonmedia.com

[2]Niklas Karlsson is with Demand Platforms R&D at Verizon Media, 701 1st Ave, Sunnyvale, California 94089 niklas.karlsson@verizonmedia.com

$h(t)$ is the seasonal component, $h(t) \in C$, $h(t) > -1$, and $\int_t^{t+T} h(\tau)d\tau = 0$. The plant maps a bid adjustment control signal u to an ad spend rate y. Assume the system is subject to a load disturbance $v_\ell(t)$ and measurement noise $v_m(t)$ entering the plant as indicated in Fig. 1. Measurement noise

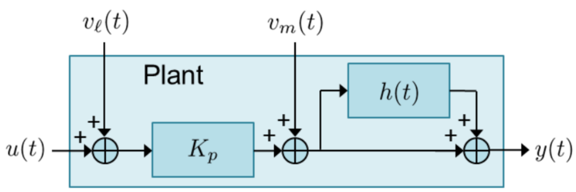

Fig. 1. Block diagram of the plant model.

is here interpreted as the random deviation of the spend rate away from its expected value and is caused by the stochastic behavior of Internet traffic. Note, by definition $\mathrm{E}\, v_m = 0$. Load disturbance is the reflection of a dynamic competitive landscape among advertisers, meaning that competing advertisers adjust their bids up and down over time.

The input-output relationships of the plant are mathematically described by

$$y \;=\; \big(1 + h(t)\big)\Big(K_p\big(u + v_\ell\big) + v_m\Big), \tag{1}$$

We may model the $u \mapsto y$ relationship as non-dynamic by ensuring the dynamics of the control system is dominant. Plant gain K_p may evolve dynamically, but this dynamic is in general slow and is disregarded in this paper. In [4], the authors proposed a periodic controller for the plan describe above. The closed-loop system is proven to be globally asymptotically stable, and the closed form solution of the state (and other signals) is derived. It is shown in simulations how the proposed controller outperforms a corresponding standard non-periodic feedback controller. The paper is based on the strong assumption that the controller has a perfect knowledge of the plant seasonality $h(t)$. The identification of the plant seasonality is a challenging task, as described in Section I and is the goal of this paper. In the following section, we formulate the problem mathematically and introduce the training data set.

III. Problem Formulation

The available impression data from Verizon Media forecasting system, shown in Fig. 2, demonstrates significant TOD pattern, meaning that a similar sine-shaped pattern repeat itself every 24 hours on top of a slowly varying trend component. The seasonality of plant is due to such

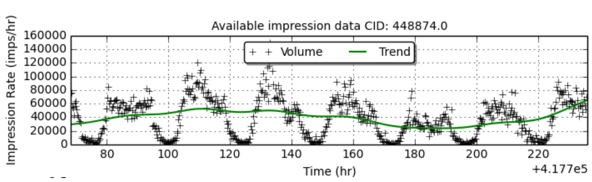

Fig. 2. Impression rate data sampled at every 0.25 hours.

seasonality of the available impression, and thus in this work, we estimate $h(t)$ based on the impression data. We use $f(t)$ to represent the impression volume per unit time (or impression rate) at time instant t and we model the time series in a linear form as follows

$$f(t) = f_{trend}(t)\,(1 + h(t))\,, \tag{2}$$

where the trend component $f_{trend}(t)$ is constant or at most slowly varying. And the seasonality component $h(t)$ represents a periodic function. The periodic function $h(t)$ is not limited to the TOD pattern, it can also include any periodic terms, such as *time of week* (TOW) pattern, which may not be as significant. We model the seasonality component $h(t)$ using a truncated Fourier series as follows

$$\begin{aligned} h(t) \;&=\; \sum_{i_d=1}^{I_d} \beta_{i_d} \sin\left(\frac{2 i_d \pi}{T_d} t + \phi_{i_d}\right) \\ &+\; \sum_{i_w=1}^{I_w} \beta_{i_w} \sin\left(\frac{2 i_w \pi}{T_w} t + \phi_{i_w}\right) + \epsilon(t), \end{aligned} \tag{3}$$

where $I_d \in \mathbb{N}$ and $I_w \in \mathbb{N}$ represent the numbers of terms corresponding to the period of a day, i.e., $T_d = 24$ (in hours), and the period of a week, i.e., $T_w = 168$ (in hours), respectively. In the rest of this paper, time unit is always in hours, unless otherwise stated. The index i_d represents the i_d-th component of day period term with an amplitude parameter β_{i_d} and a phase parameter ϕ_{i_d}. The index i_w represents the i_w-th component of week period term with an amplitude parameter β_{i_w} and a phase parameter ϕ_{i_w}. The term $\epsilon(t)$ represents a noise component, which is usually modeled as a Gaussian distributed random variable. The total number of sinusoidal terms in Eq (3) is $I = I_d + I_w$. We define two vector parameters $\boldsymbol{\beta} := [\beta_1, \ldots, \beta_{I_d}, \ldots, \beta_I]^\intercal$ and $\boldsymbol{\phi} := [\phi_1, \ldots, \phi_{I_d}, \ldots, \phi_I]^\intercal$ to represent all unknown seasonality parameters.

This paper focuses on estimating the seasonality parameters $\boldsymbol{\beta}, \boldsymbol{\phi}$ based on impression rate sampled according to a time vector $\boldsymbol{t} := [t_1, \ldots, t_M]^\intercal \in \mathbb{R}^{M\times 1}$, where $M \in \mathbb{N}$ represents the total number of samples. We assume that the time vector is evenly spaced with sampling interval $\Delta \in (0, 24)$, i.e., $t_m - t_{m-1} = \Delta$, for $m = 2, \ldots, M$. The observed impression rate data, or impression data for short, can be represented by a vector $\boldsymbol{n}_I := [n_I(t_1), \ldots, n_I(t_M)]^\intercal \in \mathbb{N}^{M\times 1}$.

IV. Methodology

The proposed algorithm has three steps. The first step, in IV-A, is to estimate and remove the trend component from impression data. The second step, in IV-B, is to have an initial estimate of the seasonality parameters and the campaign noise level. The last step, in IV-C, is to estimate seasonality parameters in a Bayesian framework where the network prior is taken into account.

A. Trend estimation

Impression data $\boldsymbol{n}_I$, according to the model in Eq (2), contains both trend component and seasonal component. In order to model the seasonal component, we need to first estimate and remove the trend component from the raw impression data $\boldsymbol{n}_I$. Trend component is estimated using smoothing spline [11], [12], [13], [14], due to its flexibility in adjusting the smoothness of the resulting curve. Using cubic B-spline, $f_{trend}(t)$ can be written as

$$f_{trend}(t) = \sum_{i=1}^{n+4} \theta_{2,i} B_{i,4}(t), \tag{4}$$

where $B_{i,4}(t)$ is cubic B-spline basis function. In our case, we choose every training data points as a knot, i.e., $n = M$. The trend parameters $\boldsymbol{\theta}_2 := [\theta_{2,1}, \theta_{2,2}, \ldots, \theta_{2,M+4}]^\intercal \in \mathbb{R}^{(M+4)\times 1}$ are estimated by minimizing the following penalized residual sum of squares

$$\begin{aligned} J(f_{trend}) &= \sum_{m=1}^{M} (n_I(t_m) - f_{trend}(t_m))^2 \\ &+ \lambda_{trend} \int_{t_{low}}^{t_{high}} \left(f''_{trend}(t) \right)^2 dt, \end{aligned} \tag{5}$$

where $\lambda_{trend} \geq 0$ is a tuning parameter, often called the smoothing parameter, the higher the value λ_{trend}, the smoother the resulting fit. To solve the above minimization problem, we first rewrite the equations in a linear form. $[f_{trend}(t_1), \ldots, f_{trend}(t_M)]^\intercal = \boldsymbol{N}_2\boldsymbol{\theta}_2$, where $\boldsymbol{N}_2 \in \mathbb{R}^{M\times(M+4)}$ is the coefficient matrix that can be obtained based on the property of B-Spline and the uniform knot sequence. The second term in Eq. (5) can be rewritten as follows,

$$\int_{t_{low}}^{t_{high}} \left(f''_{trend}(t) \right)^2 dt = \boldsymbol{\theta}_2^T \boldsymbol{\Omega}_2 \boldsymbol{\theta}_2, \tag{6}$$

where the coefficient matrix $\boldsymbol{\Omega}_2 \in \mathbb{R}^{(M+4)\times(M+4)}$ is obtained by taking the second derivative of $\boldsymbol{N}_2$. By minimizing the objective function in Eq (5), the parameters of trend component can be derived as follows

$$\hat{\boldsymbol{\theta}}_2 = \left(\boldsymbol{N}_2^\intercal \boldsymbol{N}_2 + \lambda_{trend}\boldsymbol{\Omega}_2\right)^{-1} \boldsymbol{N}_2^\intercal \boldsymbol{n}_I. \tag{7}$$

The estimated trend component are thus obtained as $[\hat{f}_{trend}(t_1), \ldots, \hat{f}_{trend}(t_M)]^\intercal = \boldsymbol{N}_2\hat{\boldsymbol{\theta}}_2$.

According to Eq (2), knowing the estimated trend, the measured seasonality at time t, denoted by $y(t)$, can be written as

$$y(t) = \frac{n_I(t)}{\hat{f}_{trend}(t)} - 1, \forall t \in \boldsymbol{t}. \tag{8}$$

We further define a vector of measured seasonality $\boldsymbol{y} = [y(t_1), \ldots, y(t_M)]^\intercal \in \mathbb{R}^{M\times 1}$. The detailed description of trend estimation and removal is provided in Algorithm 1. It has to be noted that since impression volume can not be negative, a constrained optimization problem is solved to obtain trend parameters, if the solution in Eq (7) generates negative trend component.

Algorithm 1 Trend Estimation and Removal

Parameters: λ_T, a, K
Input signals: $\boldsymbol{t}, \boldsymbol{n}_I$
Output signals: $\boldsymbol{y}$
Calculate $\boldsymbol{N}_2, \boldsymbol{\Omega}_2$ to transform (5) to the following: $J(\boldsymbol{\theta}_2) = (\boldsymbol{n}_I - \boldsymbol{N}_2\boldsymbol{\theta}_2)^\intercal (\boldsymbol{n}_I - \boldsymbol{N}_2\boldsymbol{\theta}_2) + \lambda_{trend}\boldsymbol{\theta}_2^\intercal\boldsymbol{\Omega}_2\boldsymbol{\theta}_2$
$\hat{\boldsymbol{\theta}}_2 = \left(\boldsymbol{N}_2^\intercal\boldsymbol{N}_2 + \lambda_{trend}\boldsymbol{\Omega}_2\right)^{-1} \boldsymbol{N}_2^\intercal\boldsymbol{n}_I$
$\hat{\boldsymbol{f}}_{trend} = \boldsymbol{N}_2\hat{\boldsymbol{\theta}}_2$
if $\hat{\boldsymbol{f}}_{trend}$ has negative values **then** ▷ If minimum of all elements in estimated trend vector is negative
    Solve the following constrained optimization problem
    $\hat{\boldsymbol{\theta}}_2 = \arg\min_{\boldsymbol{\theta}_2 \geq 0} (\boldsymbol{n}_I - \boldsymbol{N}_2\boldsymbol{\theta}_2)^\intercal (\boldsymbol{n}_I - \boldsymbol{N}_2\boldsymbol{\theta}_2) + \lambda_{trend}\boldsymbol{\theta}_2^\intercal\boldsymbol{\Omega}_2\boldsymbol{\theta}_2$
$\boldsymbol{y} = \frac{\boldsymbol{n}_I}{\hat{\boldsymbol{f}}_{trend}} - \boldsymbol{1}$
return $\boldsymbol{y}$

B. Seasonality estimation in a non-Bayesian framework

In this subsection, we focus on obtaining an initial estimate of seasonality parameters in a non-Bayesian framework based on the measured seasonality $\boldsymbol{y}$ computed in the previous subsection. According to Eq (3), by writing things in a linear form and stacking all M equations together, we have the following

$$\boldsymbol{y} = \boldsymbol{N}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{9}$$

where $\boldsymbol{N} \in \mathbb{R}^{M\times 2I}$ is the coefficient matrix and $\boldsymbol{\theta} := [\theta_1, \theta_2, ..., \theta_{2I}]^\intercal \in \mathbb{R}^{2I\times 1}$ is the intermediate seasonality parameter which has a one to one relationship with the original seasonality parameters $\boldsymbol{\beta}, \boldsymbol{\phi}$, as follows

$$\theta_i = \begin{cases} \beta_{(i+1)/2}\cos\phi_{(i+1)/2} & \text{if } i \text{ is odd} \\ \beta_{i/2}\sin\phi_{i/2} & \text{if } i \text{ is even.} \end{cases} \tag{10}$$

And the coefficient matrix $\boldsymbol{N}$ is given as the following

$$\boldsymbol{N} := \begin{bmatrix} \boldsymbol{N}(t_1) \\ \boldsymbol{N}(t_2) \\ \vdots \\ \boldsymbol{N}(t_M) \end{bmatrix} \in \mathbb{R}^{M\times 2I},$$

where $\boldsymbol{N}(t_m) := [N_1(t_m), \ldots, N_{2I}(t_m)] \in \mathbb{R}^{M\times 1}$ and each element $N_i(t_m), i \in \{1, 2, \ldots, 2I\}$ in $\boldsymbol{N}(t_m)$ can be expressed as follows
For $i \in \{1, 2, \ldots, 2I_d\}$,

$$N_i(t_m) = \begin{cases} \sin(\frac{\pi(i+1)/t_m}{T_d}) & \text{if } i \text{ is odd} \\ \cos(\frac{\pi i/t_m}{T_d}), & \text{if } i \text{ is even.} \end{cases} \tag{11}$$

For $i \in \{2I_d + 1, 2I_d + 2, \ldots, 2I\}$,

$$N_i(t_m) = \begin{cases} \sin(\frac{\pi(i-2I_d+1)/t_m}{T_w}) & \text{if } i \text{ is odd} \\ \cos(\frac{\pi(i-2I_d)/t_m}{T_w}) & \text{if } i \text{ is even.} \end{cases} \tag{12}$$

In this subsection we assume that $\boldsymbol{\theta}$ is deterministic but unknown. Regular least squares estimator [15] may have poor performance in presence of outliers. Thus, we use a robust linear regression method, which uses iteratively reweighted least squares with a specific weighting function (bisquare) [16]. Within each iteration, a new weighted least square (with the weights from previous iteration) is formed and solved. And a new weight is calculated for each data point based on its corresponding residual in the current iteration.

At step k, the estimate $\hat{\boldsymbol{\theta}}^{(k)}$ is obtained by :

$$\hat{\boldsymbol{\theta}}^{(k)} = \min_{\boldsymbol{\theta}} \sum_{m=1}^{M} w^{(k-1)}(t_m)(y(t_m) - \boldsymbol{N}(t_m)\boldsymbol{\theta})^2, \quad (13)$$

where $w^{(k-1)}(t_m)$ is the weight calculated at step $k-1$ for the data point sampled at time t_m. And the new weight $w^{(k)}(t_m)$ is based on the corresponding residuals from step k. For $m = 1, \ldots, M$

$$w^{(k)}(t_m) = f\left(y(t_m) - \boldsymbol{N}(t_m)\boldsymbol{\theta}^{(k)}\right), \quad (14)$$

where $f(\cdot)$ represents a specific weighting function which assigns higher weights for data points with larger residuals and lower weights for those with smaller residuals.

The algorithm terminates when the stopping criteria is satisfied (for example, when the difference between estimates of two consecutive iterations is small enough). This approach is robust to the outliers by iteratively assigning incrementally lower weights to those outliers.

The pseudo code of the above estimator is given in Algorithm 2. It may happen that the estimated seasonality $\hat{\boldsymbol{\theta}}$ results in negative impression volume at some t which is not physically acceptable. To guarantee that the estimator generates an acceptable seasonality model, the constraint in line 12 of Algorithm 2 is added to the optimization problem.

$$\min_{\boldsymbol{\theta}} \sum_{m=1}^{M} w(t_m)\left(y(t_m) - \boldsymbol{N}(t_m)\boldsymbol{\theta}\right)^2 \quad (15)$$
$$s.t. \;\; \boldsymbol{N}(t_m)\boldsymbol{\theta} + 1 \geq 0, \;\forall m \in 1, \ldots, M.$$

The constrained minimization problem can be solved by sequential least square programing.

After we obtain an initial estimate of $\hat{\boldsymbol{\theta}}$, we are able to calculate the estimated residual vector as $\hat{\boldsymbol{\epsilon}} = \boldsymbol{y} - \boldsymbol{N}\hat{\boldsymbol{\theta}}$. Then the estimated variance $\hat{\sigma}^2$ is obtained from $\hat{\boldsymbol{\epsilon}}$ as follows

$$\hat{\sigma}^2 = \text{var}(\hat{\boldsymbol{\epsilon}}, \boldsymbol{w}) = \frac{M}{(M-1)} \frac{\sum_{m=1}^{M} w_m(\hat{\epsilon}_m - \bar{\epsilon})^2}{\sum_{m=1}^{M} w_m}, \quad (16)$$

where $\bar{\epsilon} = \frac{\sum_{m=1}^{M} w_m \hat{\epsilon}_m}{\sum_{m=1}^{M} w_m}$ represents the weighted mean of all residuals terms. In practice, in dealing with extremely noisy data which may even contain faulty data points, weight adjustment is needed before we can have a good estimate of σ^2. An adjustment factor $w_{adj} \in [0, 1]$ is used to assign a lower weight to data points that are likely to be faulty. The estimated noise variance, $\hat{\sigma}^2$, implies the noise level of campaign. The smaller the $\hat{\sigma}^2$, the cleaner the impression data.

C. Bayesian framework

Knowing the prior distribution of $\boldsymbol{\theta}$, we can formulate the estimation problem of seasonality parameters in a Bayesian framework [15], [17]. Unlike the previous subsection where we take $\boldsymbol{\theta}$ as deterministic, now we model $\boldsymbol{\theta}$ as random with certain prior distribution.

We solve the Bayesian estimation problem by minimizing the *Mean Square Error* (MSE)

$$\text{MSE} = E\left[\left(\hat{\boldsymbol{\theta}}_B(\boldsymbol{y}) - \boldsymbol{\theta}\right)^{\mathsf{T}}\left(\hat{\boldsymbol{\theta}}_B(\boldsymbol{y}) - \boldsymbol{\theta}\right)\right], \quad (17)$$

Algorithm 2 Initial Seasonality Estimation

```
1: Parameters: w_adj, a, K, a_r, K_r
2: Input signals: y
3: Output signals: σ̂²
4: Calculate matrix N to transform the problem to the following: y = Nθ
5: R = robustfit(y, N)          ▷ Solve y = Nθ using robust regression
6: θ̂ = R.parameters
7: w = R.weights
8: Assign lower weights to zero volume data using adjust factor w_adj
9: if θ̂ is physically not acceptable then
10:     Solve the following constrained optimization problem
11:     θ̂ = arg min_θ ||√W y − √W Nθ||²
12:     s.t. Nθ + 1 > 0                                 ▷ W = diag(w)
13: ε̂ = y − Nθ̂
14: σ̂² = var(ε̂, w)  ▷ Calculate the weighted variance of all elements in the vector
15: Return σ̂²
```

where the expectation is taken with respect to the prior distribution of $\boldsymbol{\theta}$ and the subscript B in $\hat{\boldsymbol{\theta}}_B$ is to indicate that the estimate is obtained in Bayesian framework. Using the MSE as risk function, the Bayesian estimate of the unknown parameter is the mean of the posterior distribution.

$$\hat{\boldsymbol{\theta}}_B = \text{E}\left[\boldsymbol{\theta}|\boldsymbol{y}\right] = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}. \quad (18)$$

We assume that the residual $\boldsymbol{\epsilon} := [\epsilon_1, \ldots, \epsilon_M]^{\mathsf{T}} \in \mathbb{R}^{M\times 1}$ in the seasonality model in Eq (9) is i.i.d and Gaussian distributed, i.e., $\epsilon_m \sim \mathcal{N}(0, \sigma^2), \forall m = 1, \ldots, M$. The noise variance, σ^2, is unknown but we have an estimate of it, $\hat{\sigma}^2$, from the previous subsection.

We model the prior distribution of $\boldsymbol{\theta}$ as multivariate Gaussian with mean $\boldsymbol{\mu_\theta}$ and covariance matrix $\boldsymbol{\Sigma_\theta}$, Eq (18) can be further derived as [15]

$$\hat{\boldsymbol{\theta}}_B = \boldsymbol{\mu_\theta} + \frac{\boldsymbol{\Sigma_\theta}\boldsymbol{N}^{\mathsf{T}}}{\boldsymbol{N}\boldsymbol{\Sigma_\theta}\boldsymbol{N}^{\mathsf{T}} + \hat{\sigma}^2\mathbb{I}}(\boldsymbol{y} - \boldsymbol{N}\boldsymbol{\mu_\theta}), \quad (19)$$

where $\mathbb{I}$ is a $M \times M$ identity matrix. Once $\hat{\boldsymbol{\theta}}_B$ is obtained, we can transform it back to $\hat{\boldsymbol{\beta}}_B$ and $\hat{\boldsymbol{\phi}}_B$ according to their one-to-one mapping relationship. The detailed implementation of the Bayesian estimator is in Algorithm 3. It can be seen from Eq. (19) that as $\hat{\sigma}^2$ increases, the prior information has higher weight while the measurements $\boldsymbol{y}$ becomes less valuable in the final evaluation of $\hat{\boldsymbol{\theta}}_B$. Intuitively, when the impression data is more noisy, the final estimate relies more on prior information. On the contrary, if the impression data is well behaved, prior information plays a smaller role. The advantage of having such a Bayesian framework is to fully utilize the prior information, especially when the other source of information, i.e., the impression data, is noisy and thus not informative about the seasonality model. The proposed algorithm generates a reasonable seasonality model while acknowledging the difference among campaigns, as will be shown in the following section.

V. Experimental Results

The TOD pattern, among all seasonal components, including day of week, are of particular interest, since the reference signal is pre-set according to the daily budget. Thus, in the

Algorithm 3 Bayesian Estimation of Seasonality

Parameters: $\boldsymbol{\mu}_\theta^{us}, \boldsymbol{\Sigma}_\theta^{us}, \boldsymbol{\mu}_\theta^{euro}, \boldsymbol{\Sigma}_\theta^{euro}, \boldsymbol{\mu}_\theta^{jap}, \boldsymbol{\Sigma}_\theta^{jap}$
Input signals: $\boldsymbol{y}, \hat{\sigma}^2, \boldsymbol{N}$
Output signals: $\hat{\boldsymbol{\beta}}_B, \hat{\boldsymbol{\phi}}_B$
Choose prior distribution $(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$ according to ad service ID
if No enough data points for estimation **then**
    $\hat{\boldsymbol{\theta}}_B = \boldsymbol{\mu}_\theta$ ▷ Use network prior mean
else

$$\hat{\boldsymbol{\theta}}_B = \boldsymbol{\mu}_{\boldsymbol{\theta}} + \frac{\boldsymbol{\Sigma}_\theta \boldsymbol{N}^\intercal}{(\boldsymbol{N}\boldsymbol{\Sigma}_\theta \boldsymbol{N}^\intercal + \hat{\sigma}^2 \mathbb{I})} (\boldsymbol{y} - \boldsymbol{N}\boldsymbol{\mu}_\theta)$$

    if $\hat{\boldsymbol{\theta}}_B$ is physically not acceptable **then**
        $\hat{\boldsymbol{\theta}}_B = \boldsymbol{\mu}_\theta$ ▷ Use network prior mean
Transform $\hat{\boldsymbol{\theta}}_B$ to $\hat{\boldsymbol{\beta}}_B, \hat{\boldsymbol{\phi}}_B$ ▷ $(\hat{\boldsymbol{\beta}}_B, \hat{\boldsymbol{\phi}}_B) = g^{-1}(\hat{\boldsymbol{\theta}}_B)$
return $\hat{\boldsymbol{\beta}}_B, \hat{\boldsymbol{\phi}}_B$

experiment, we set the number of terms corresponding to the period of a week in Eq. (3) be 0, i.e., $I_w = 0$. We choose $I_d = 2$ based on the observation that most TOD patterns exhibit at most 2 peaks. The selection of the hyper parameters may influence the performance of our model, but is beyond the scope of this paper.

In Verizon Media forecasting system, the raw impression data is provided with sampling interval $\Delta = 0.25$ hours. According to Nyquist-Shannon sampling theorem [18], if a function $x(t)$ contains no frequencies higher than B hertz, it is completely determined by giving its ordinates at a series of points spaced $1/(2B)$ seconds apart. The highest frequency of the continuous time series model for seasonality in Eq (3) is $B = \frac{I_d}{T_d}$, and thus, the condition $\Delta < \frac{T_d}{2I_d} = 6$ hours allows all seasonality information to be preserved in discrete impression data $\boldsymbol{n}_I$. We choose a feasible $\Delta = 4$ so that equal number of data points are sampled for each day. An example of the resulting 4-hour impression data is plotted using crosshair symbol in the first subplot of Fig. 4.

The estimated trend component shown by the green curve in the first subplot of Fig. 4. Once the trend component from impression data $\boldsymbol{n}_I$ is removed according to Eq (8), the measured seasonality $\boldsymbol{y}$ can be obtained, as shown by the dots in the second subplot of Fig. 4. A non-Bayesian or initial estimate of seasonality is shown by the blue curve in the second and third subplots of Fig. 4.

Before estimating the seasonality parameters $\boldsymbol{\theta}$ in a Bayesian framework, we infer its prior distribution using historical data of all campaigns. We learn a specific prior distributions of $\boldsymbol{\theta}$ for each network, such as U.S., Japan and Europe. Taking U.S. as an example, we first select K well behaved campaigns, $K = 1117$ in our case. Then, for each campaign an estimate of the seasonality parameter $\boldsymbol{\theta}$ is generated using the non-Bayesian method discussed in Subsection IV-B. At last, a multivariate normal distribution is fitted to training data $(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_K)$, as plotted in Fig. 3, and the parameters, mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}^{us}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{us}$ are obtained using *maximum likelihood estimation* (MLE), where the superscript represents a network information encoded in an ad service ID. The distribution $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{us}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{us})$ will be used as the prior of $\boldsymbol{\theta}$ for campaigns with ad service ID being U.S.

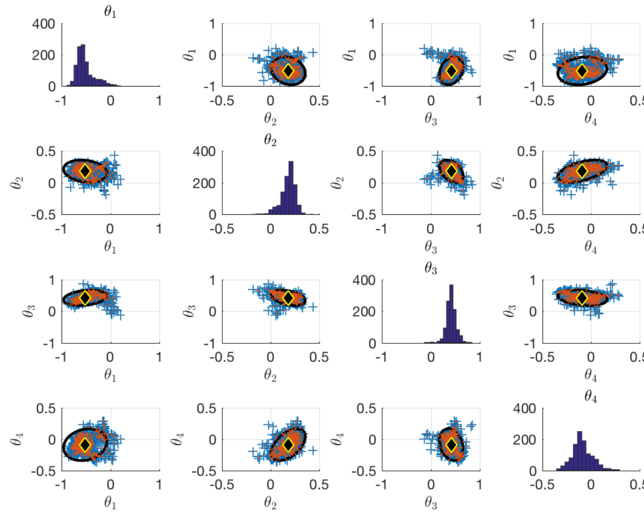

Fig. 3. Scatter plots and histograms of seasonality parameters $\boldsymbol{\theta}$ in U.S.

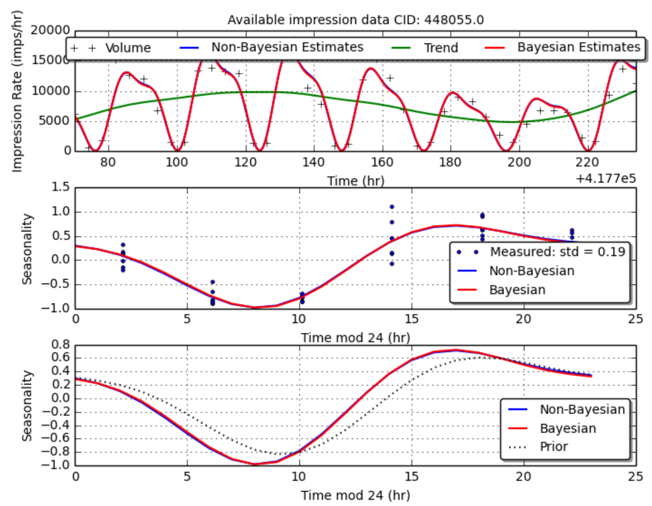

Fig. 4. Case one: well behaved campaign.

In the first case, Fig. 4, the data is well behaved, as can be seen from the raw impression data in the first subplot, and the TOD pattern is strong and easily recognizable. Thus, as shown in the second and third subplots, the Bayesian estimated TOD pattern (red line) is close to initial (non-Bayesian) estimates (blue line), though deviates from the network prior (dotted black line in the third subplot). In conclusion, when the impression data of the campaign is of low noise level, the Bayesian estimation algorithm assigns high weight or credibility to the impression data and low weight or credibility to network prior. Thus, the eventual estimated seasonality follows the impression data.

In the second case, Fig. 5, the noise level of the campaign are higher compared to the first case, as can be seen from both the raw impression data in the first subplot and the estimated noise variance $\sqrt{\hat{\sigma}^2} = 0.431$. The measured seasonality in the second subplot are not of clear pattern. The Bayesian estimated seasonality in the third subplot is different from the initial (non-Bayesian) estimated seasonal-

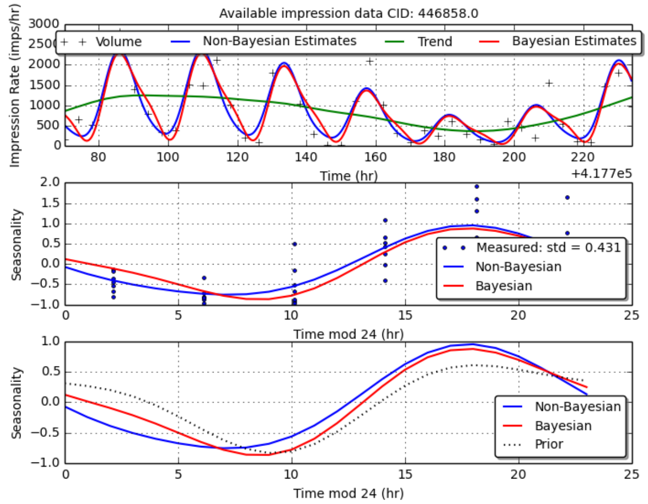

Fig. 5. Case two: slightly noisy campaign.

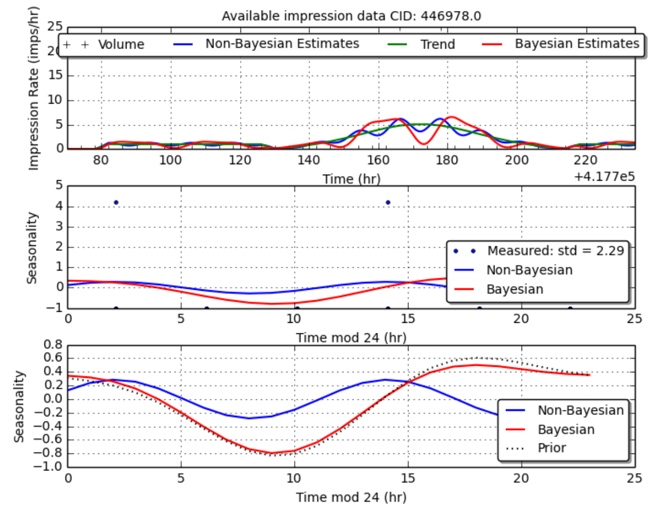

Fig. 6. Case three: extremely noisy campaign.

ity and is pulled towards the network prior. So, the algorithm assigns a slightly higher weight to the prior distribution when the noise level is higher. It strikes a balance between network prior and impression data when the impression data is moderately credible.

In the last case, Fig. 6, the campaign is extremely noisy. No pattern can be seen from the raw impression data in the first subplot, and the estimated noise variance is high. The initial (non-Bayesian) estimated seasonality is almost flat, as can be seen from the third subplot. But the Bayesian estimate majorly follows the prior information instead. So, when impression data is not informative at all, our proposed algorithm heavily relies on network prior and still renders a physically sensible TOD pattern. We applied our approach to more than 1000 campaigns across different networks. The above three examples cover all cases we met, in every one of which a reasonable TOD pattern is generated, demonstrating the flexibility and robustness of our proposed Bayesian estimation method. Such robustness is crucial in an automated campaign management system.

VI. Conclusions and Future work

In this paper, the problem of seasonality identification in feedback control system for campaign management was formulated as a Bayesian estimation problem. A solution composed of three main algorithms was proposed using the network prior distributions learnt from a large number of campaigns. The proposed Bayesian estimation method performed well in estimating TOD pattern of individual campaigns. It was robust and works for campaign with any noise level by adaptively assigning credibility to the prior distribution and the impression data. In the future, we will investigate the impact of some hyper parameters, such as sampling interval (Δ), number of day and week terms (I_d, I_w), on the overall system performance. We also plan to further investigate and analyze the confidence interval for estimated TOD pattern to facilitate the consumption of the model based its confidence level.

References

[1] N. Karlsson and J. Zhang, "Applications of feedback control in online advertising," in *2013 American Control Conference*, June 2013, pp. 6008–6013.

[2] N. Karlsson, "Control problems in online advertising and benefits of randomized bidding strategies," *European Journal of Control*, vol. 30, pp. 31–49, 2016.

[3] ——, "Systems and methods for shaping a reference signal in advertising," Nov. 16 2010, US Patent 7,835,938. [Online]. Available: http://www.google.ch/patents/US7835938

[4] ——, "Control of periodic systems in online advertising," in *2018 IEEE 57th Conference on Decision and Control (CDC)*, Dec 2018.

[5] ——, "Adaptive optimization and control in online advertising," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, Dec 2019, submitted.

[6] S. Hylleberg, Ed., *Modelling Seasonality*. Oxford University Press, 1992. [Online]. Available: https://EconPapers.repec.org/RePEc:oxp:obooks:9780198773184

[7] P. J. Brockwell, R. A. Davis, and M. V. Calder, *Introduction to time series and forecasting*. Springer, 2002, vol. 2.

[8] F.-M. Tseng, H.-C. Yu, and G.-H. Tzeng, "Combining neural network model with seasonal time series arima model," *Technological Forecasting and Social Change*, vol. 69, no. 1, pp. 71–87, 2002.

[9] D. Ö. Faruk, "A hybrid neural network and arima model for water quality time series prediction," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 4, pp. 586–594, 2010.

[10] C. A. Sims, "Seasonality in regression," *Journal of the American Statistical Association*, vol. 69, no. 347, pp. 618–626, 1974.

[11] T. Hastie, R. Tibshirani, and J. Friedman, "Overview of supervised learning," in *The elements of statistical learning*. Springer, 2009, pp. 9–41.

[12] J. O. Ramsay, "Monotone regression splines in action," *Statist. Sci.*, vol. 3, no. 4, pp. 425–441, 11 1988. [Online]. Available: http://dx.doi.org/10.1214/ss/1177012761

[13] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. Springer, 2009.

[14] C. H. Reinsch, "Smoothing by spline functions," *Numerische mathematik*, vol. 10, no. 3, pp. 177–183, 1967.

[15] S. M. Kay, *Fundamentals of statistical signal processing, volume I: estimation theory*. Prentice Hall, 1993.

[16] P. J. Huber, *Robust statistics*. Springer, 2011.

[17] E. L. Lehmann and G. Casella, *Theory of point estimation*. Springer Science & Business Media, 2006.

[18] F. Miller, A. Vandome, and J. McBrewster, *Nyquist-Shannon Sampling Theorem: Aliasing, Sine Wave, Signal Processing, Nyquist Rate, Nyquist Frequency, Sampling Rate, Shannon-Hartley Theorem, Whittaker-Shannon Interpolation Formula, Reconstruction from Zero Crossings*. Alphascript Publishing, 2010.