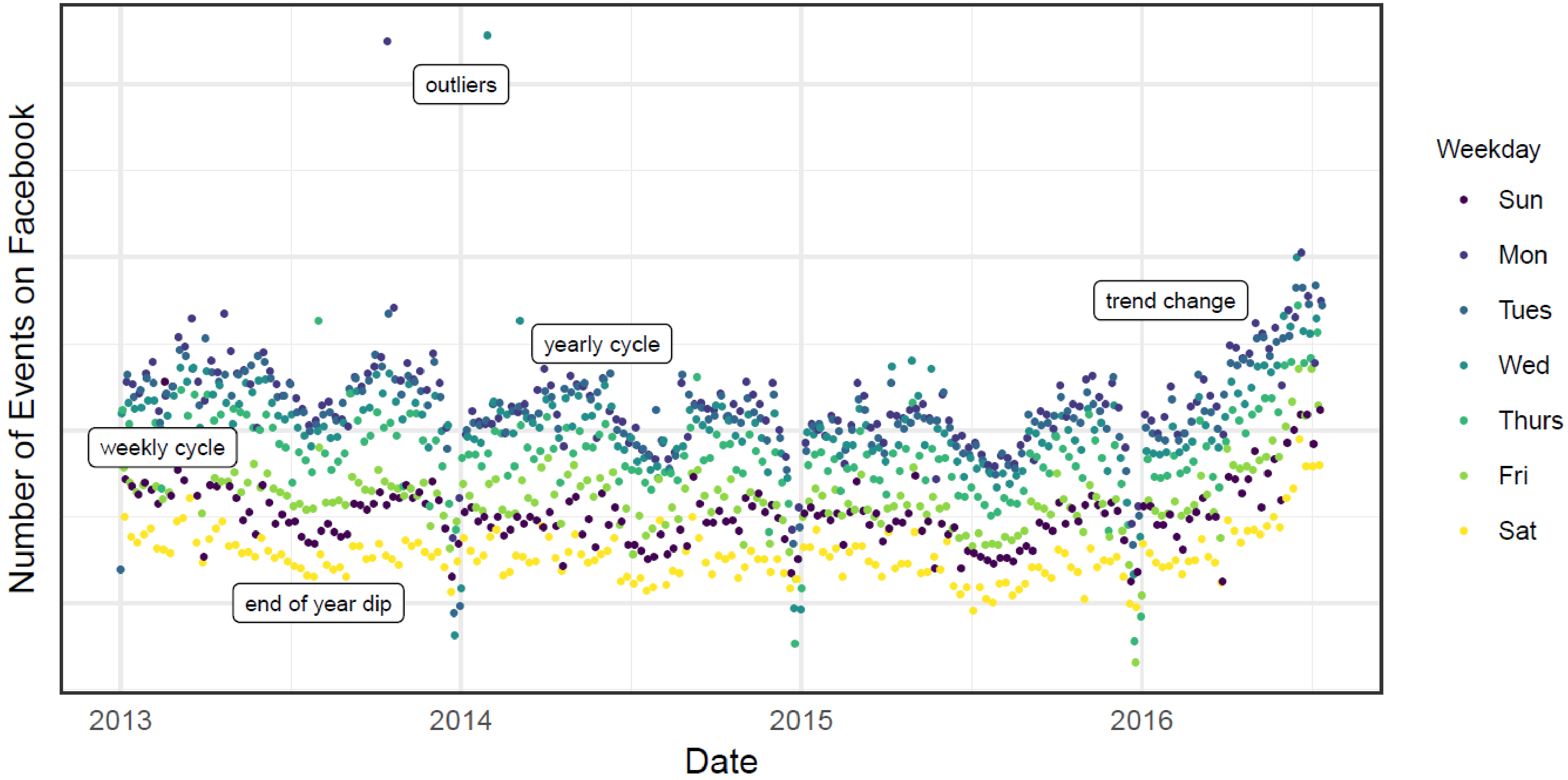
Time series modelling with the "Prophet" [2017]

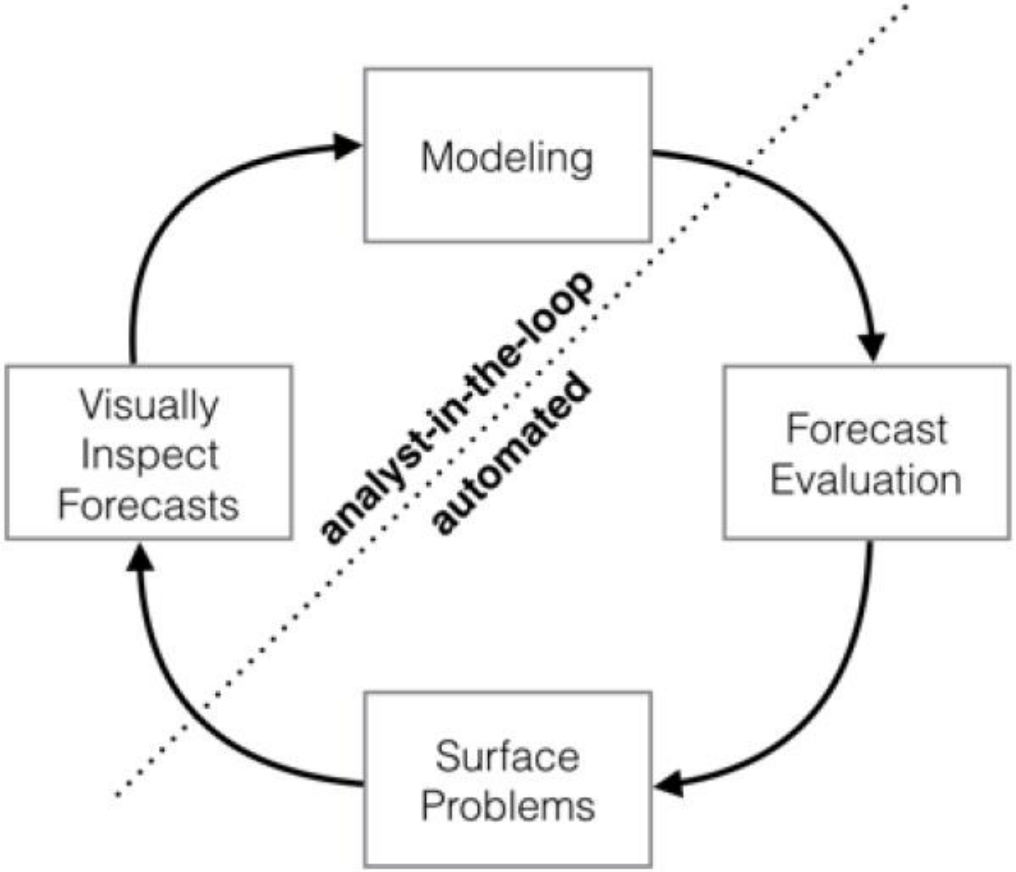
Forecasting at Scale

Sean J. Taylor*†
Facebook, Menlo Park, California, United States
sjt@fb.com

and

Benjamin Letham†
Facebook, Menlo Park, California, United States
bletham@fb.com


outliers
yearly cycle
trend change
weekly cycle
end of year dip
Number of Events on Facebook
2013
2014
2015
2016
Date
Weekday
Sun
Mon
Tues
Wed
Thurs
Fri
Sat


Modeling
Forecast Evaluation
Surface Problems
Visually Inspect Forecasts
analyst-in-the-loop
automated

Input to the model

- 'Past' training samples $\{t_k, y_k\}$
- Information about change point locations $\{s_j\}_1^S$ - optional
- Information about seasonality $P = 7$ (weekly), $P = 365.25$ (yearly) - optional
- Information about special types of days/events (e.g holidays) - optional.
- For any type of special day i, there is the set D_i of days of this type. So, for any time sample (past or future) we have a vector

$$Z(t) = \left[1_{t \in D_1}, \ldots, 1_{t \in D_L}\right].$$

Model

- Additive model

$$y(t) = \underbrace{g(t)}_{\text{trend}} + \underbrace{s(t)}_{\text{seasonality}} + \underbrace{h(t)}_{\text{special events}} + \underbrace{\varepsilon(t)}_{\text{noise}}$$

- Each component has set of parameters that needs to be ‘learnt’ through maximization of likelihood.

- Combination of Bayesian data analysis and constructive approximation.

- Suppose that the model has parameters θ. Bayes’ rule allows to find ‘optimal’ parameters under prior assumptions on $p(\theta)$.

- Posterior density $$p(\theta|y) = \frac{p(\theta)\,p(y|\theta)}{p(y)},$$

- Unnormalized posterior density $p(\theta|y) \propto p(\theta)p(y|\theta)$

- We can maximize likelihood of posterior.

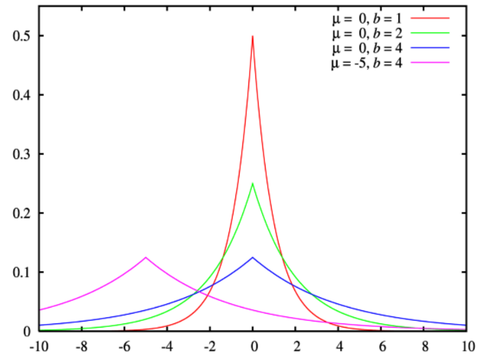
- The forecast is a distribution, at each t there is a mean with some level of certainty.

Trend model (nonlinear approximation)

- Logistic growth model with fixed capacity

$$\tilde{g}(t) = \frac{C}{1+e^{-k(t-m)}},\ k \text{ growth rate, } m \text{ offset.}$$

- The Prophet model allows capacity & growth to change.
- we model an expected continuous capacity $C(t)$, e.g. a linear function. Could be provided explicitly, e.g. world population estimates.
- Growth allowed to change at times $\{s_j\}_{j=1}^{S}$, with changes $\{\delta_j\}_{j=1}^{S}$.
- Prior $\delta_j \sim \text{Laplace}(0,\tau)$, allows negative growth change.



- Rate at time t is modeled as a piecewise constant

$$k + \sum_{j, t \geq s_j} \delta_j = k + \langle a(t), \delta \rangle, \quad a_j(t) = 1_{t \geq s_j}.$$

- The locations $\{s_j\}_{j=1}^{S}$ can be provided as input or found through automatic selection with a sparse prior such as: one per month.

- To achieve continuity of g, the offset parameter also needs to change at the change points.

- So the trend model is

$$g(t) = \frac{C(t)}{1 + e^{-(k + \langle a(t), \delta \rangle)(t - (m + \langle a(t), \gamma \rangle))}},$$

with

$$\gamma_j = \left(s_j - m - \sum_{l<j} \gamma_l \right) \left(\frac{\delta_j}{k + \sum_{l \le j} \delta_l} \right)$$

Proof (of continuity of g) Denote $X := k + \sum_{l<j} \delta_l$,

$Y := s_j - m - \sum_{l<j} \gamma_l$

For continuity we need $\underset{t \to (s_j)_-}{XY} = \underbrace{(X + \delta_j)(Y - \gamma_j)}_{t \to (s_j)_+}$.

$$XY = \left(X + \delta_j \right)\left(Y - \gamma_j \right) \Leftrightarrow \gamma_j = \frac{Y\delta_j}{X + \delta_j} \Leftrightarrow$$

$$\gamma_j = \left(s_j - m - \sum_{l<j} \gamma_l \right) \left(\frac{\delta_j}{k + \sum_{l \leq j} \delta_l} \right)$$

Seasonality model

- We use a Fourier series approximation

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi nt}{P} \right) + b_n \cos\left(\frac{2\pi nt}{P} \right) \right)$$

- There are $2N$ parameters to learn $\beta = [a_1, b_1, \ldots, a_n, b_n]$.
- For yearly the authors chose $N = 10$, for weekly $N = 3$. It provides a low-pass filter
- Smoothing prior $\beta_j \sim N(0, \sigma^2)$.

Special events model

- D_i set of special days of type i. So, we have a vector

$$Z(t) = \left[1_{t\in D_1}, \ldots, 1_{t\in D_L}\right].$$

- Observe this can be specified for forecasted time points.
- Special events regression $\kappa = \{\kappa\}_{i=1}^{L}$, prior $\kappa_i \sim N\left(0, \nu^2\right)$,

$$h(t) = \langle Z(t), \kappa \rangle.$$

- Sometimes one places a ‘window’ around the special days, making surrounding days also special.

Listing 1: Example Stan code for our complete model.

```
model {
  // Priors
  k ~ normal(0, 5);
  m ~ normal(0, 5);
  epsilon ~ normal(0, 0.5);
  delta ~ double_exponential(0, tau);
  beta ~ normal(0, sigma);

  // Logistic likelihood
  y ~ normal(C ./ (1 + exp(-(k + A * delta) .* (t - (m + A * gamma)))) +
             X * beta, epsilon);
  // Linear likelihood
  y ~ normal((k + A * delta) .* t + (m + A * gamma) + X * beta, sigma);
}
```