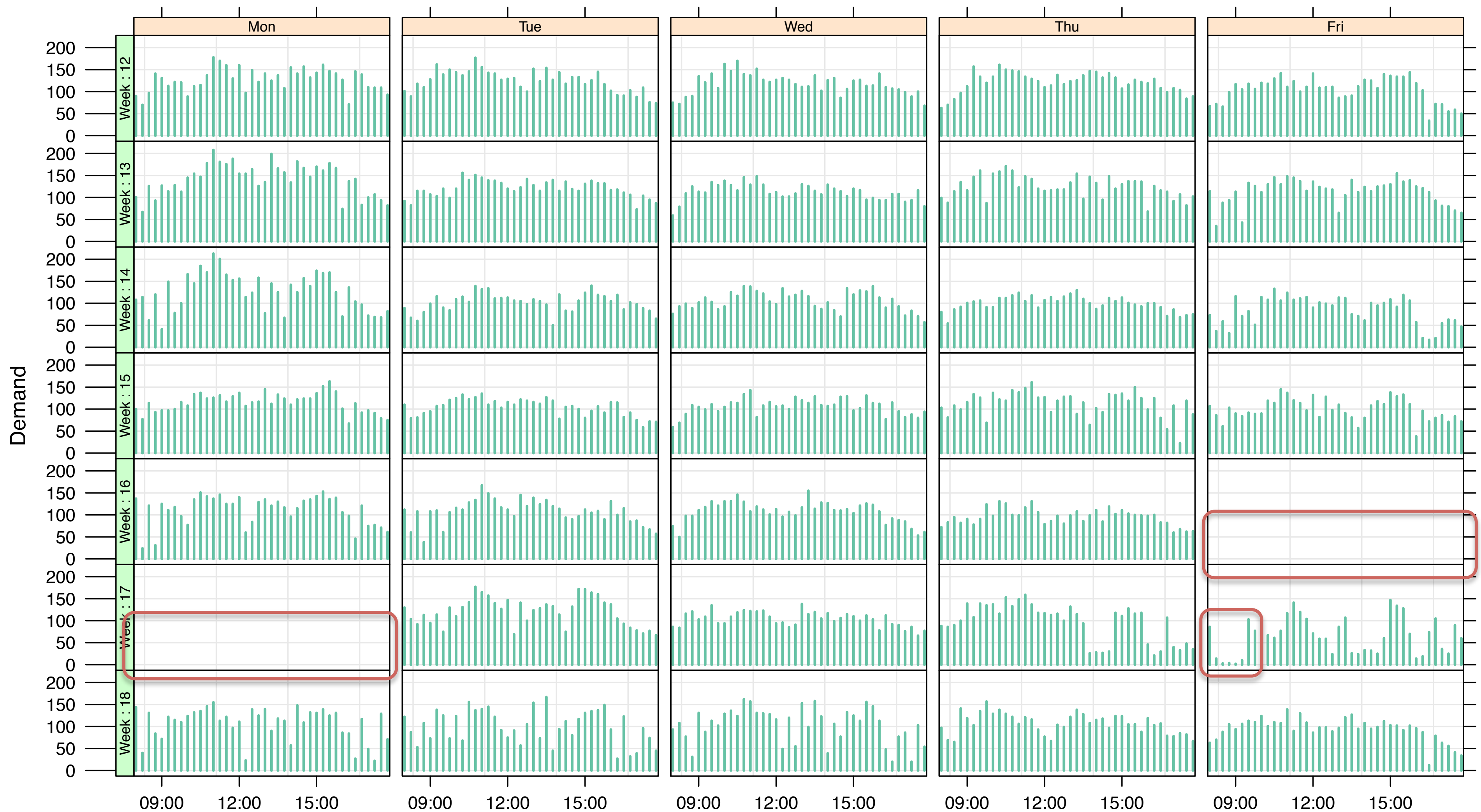
Hierarchical Bayesian State-Space Model for Call Center Arrival Rate Forecasting

Nicolas Chapados
Pierre L'Ecuyer

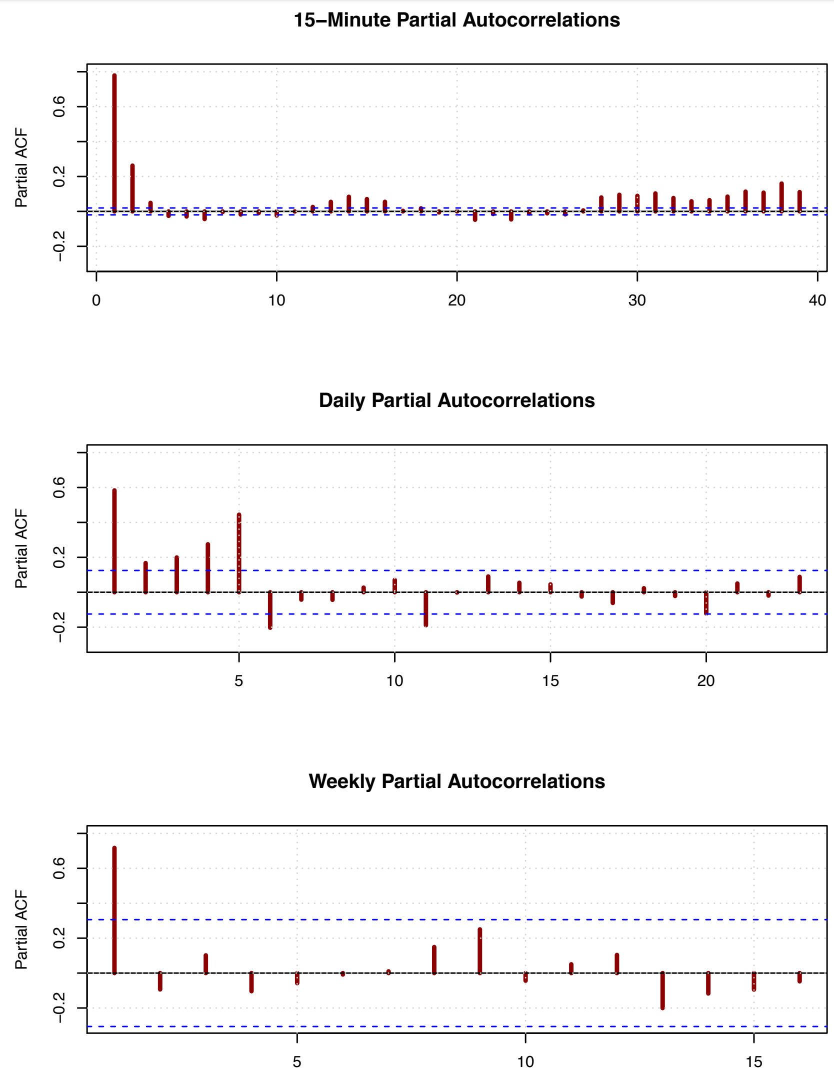
May 7th, 2014

Data Modeling Problem

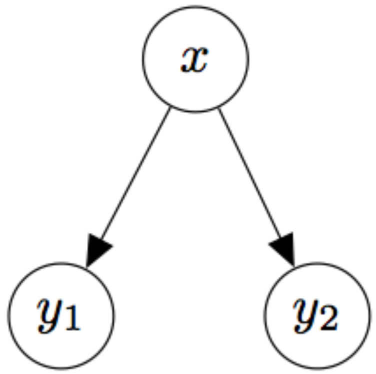


Stylized Facts

- Seasonalities
- Persistence at intraday, daily, weekly horizons
- Overdispersion wrt Poisson
  - ➢ Variance > Mean
- Special events: significant impact on demand
- Multiple call types (queues)
  - – 10 queues for HQ call center data

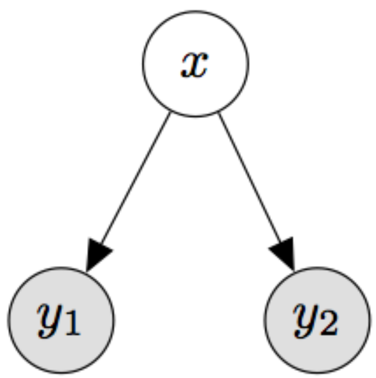


(Directed) Probabilistic Graphical Models



- *x*, y_1, y_2 represent **random variables**
- Graph expresses a **factorization** of the joint distribution, here as

$$P(x, y_1, y_2) = \underbrace{\boxed{P(x)}}_{\textbf{Prior}}\underbrace{\boxed{P(y_1 \mid x)P(y_2 \mid x)}}_{\textbf{Likelihood}}$$



- Shading: denotes observed variables
- **Inference** problem: determine **latent** *x* from observed *y*. Use **Bayes' Rule**

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \propto P(B \mid A)P(A)$$

$$\boxed{P(x \mid y_1, y_2) \propto P(y_1, y_2 \mid x)P(x) = P(y_1 \mid x)P(y_2 \mid x)P(x)}$$

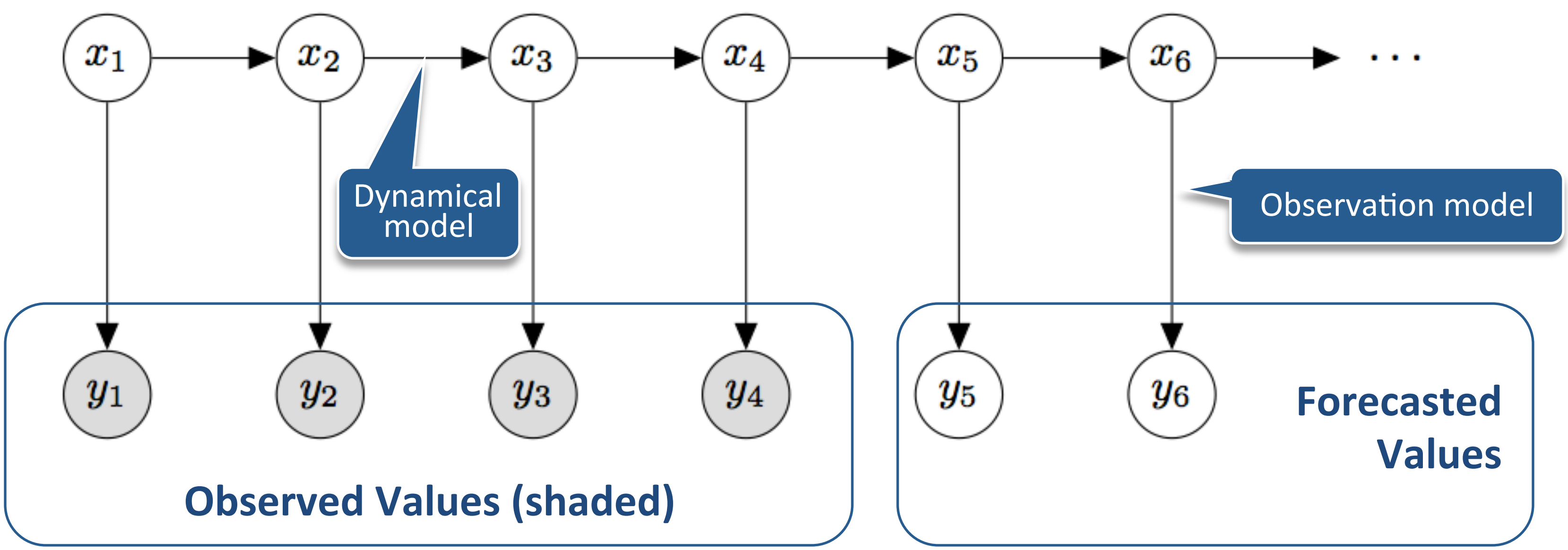
State-Space Modeling Basic Ideas

- To model $\{y_t\}$, we first invent a **latent** stochastic process $\{x_t\}$
  - Process $\{x_t\}$ is called the **state,** $x_t \in \mathbb{R}^k \text{ or } x_t \in \mathbb{Z}^k$
  - It is **Markovian**, i.e. $P(x_t \mid x_{t-1}, x_{t-2}, \ldots) = P(x_t \mid x_{t-1})$
  - Observations $\{y_t\}$ are probabilistically related to the latent process $\{x_t\}$

- Notion originates from engineering in the 1960's
  - Discrete- or continuous-time; this presentation: discrete-time only
- Known under different names in different fields
  - Kalman Filter, Linear Dynamical System (LDS)
    - Real-valued latent variables and observations, Linear-Gaussian dynamics
  - Hidden Markov Model (HMM), Markov Switching Model (MSM)
    - Discrete latent variables, arbitrary observations

State-Space Models for Time Series

- Observed values $\{y_t\}$ are modeled as a being "generated" from a **latent state** series $\{x_t\}$
- We first model the **dynamics** on x_t: how x_{t+1} depends on x_t
  - *Markovian dynamics*: the current state summarizes all the past until now
- We next model the **observations**: how y_t is generated from x_t
- Note that given x_t , y_t is independent of all others *y*'s!

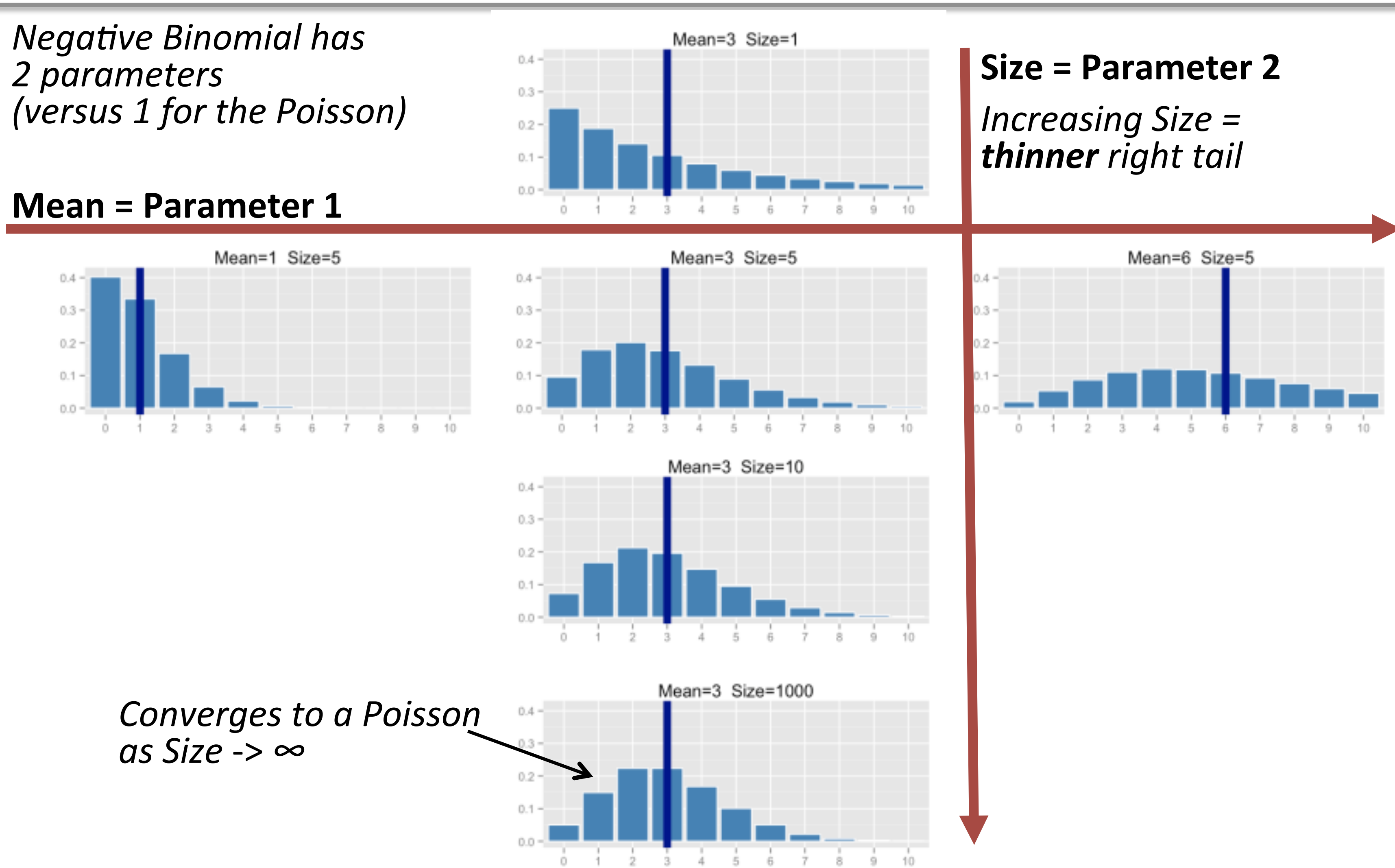


Benefits, i.e. Why Bother

- Encompass most classical time series forecasting approaches
  - For example: exponential smoothing and ARIMA models
- Easily deal with **missing data** in the history
- Easily allow for **explanatory variables**
- Naturally yield **distributional forecasts**
  - Rich measure of forecast uncertainty
  - Point forecasts not enough call center simulation tasks
- Useful **building block** from which more complex models naturally arise
  - Observations can be non-negative integers
  - Simple generalization to multivariate series, via multivariate state

Negative Binomial Distribution

Fatter-tailed generalization of the Poisson Distribution



State-Space Model for Call Arrivals

$$y_{t,h} \sim \text{Neg. Bin.}(\exp(\tilde{\eta}_{t,h}), \alpha),$$

$$\tilde{\eta}_{t,h} = \eta_{t,h} + \mathbf{x}'_{t,h}\boldsymbol{\theta},$$
$$\eta_{t,1} = \mu_t + \epsilon_{t,1},$$
$$\eta_{t,h} = \mu_t + \phi_I(\eta_{t,h-1} - \mu_t) + \epsilon_{t,h}, \quad h > 1,$$

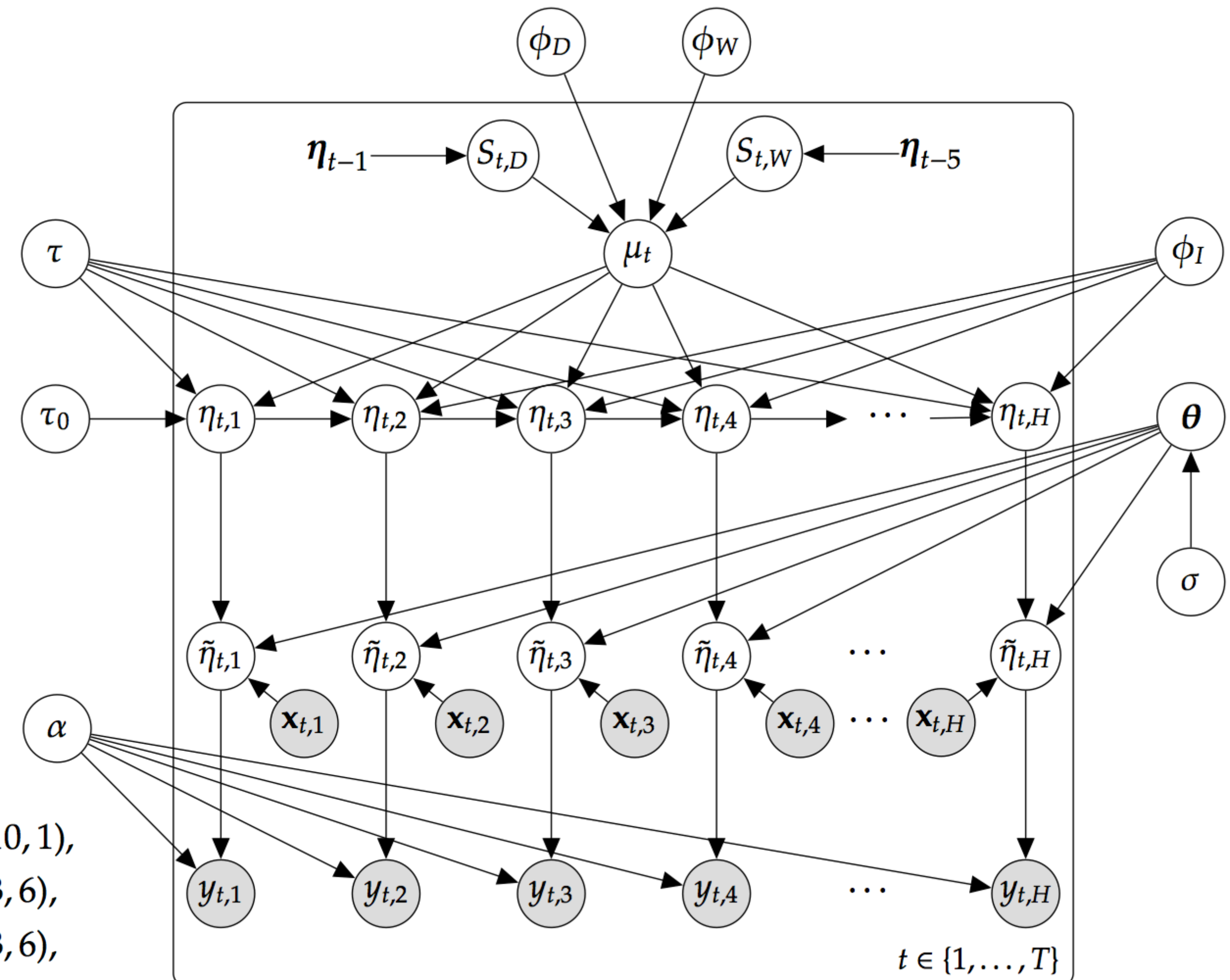
$$\mu_t = \omega + \frac{\phi_D}{H} S_{t,D} + \frac{\phi_W}{H} S_{t,W}$$

$$S_{t,D} = \begin{cases} \sum_{h=1}^{H} \eta_{t-1,h}, & \text{if } t > 1, \\ 0, & \text{otherwise,} \end{cases}$$

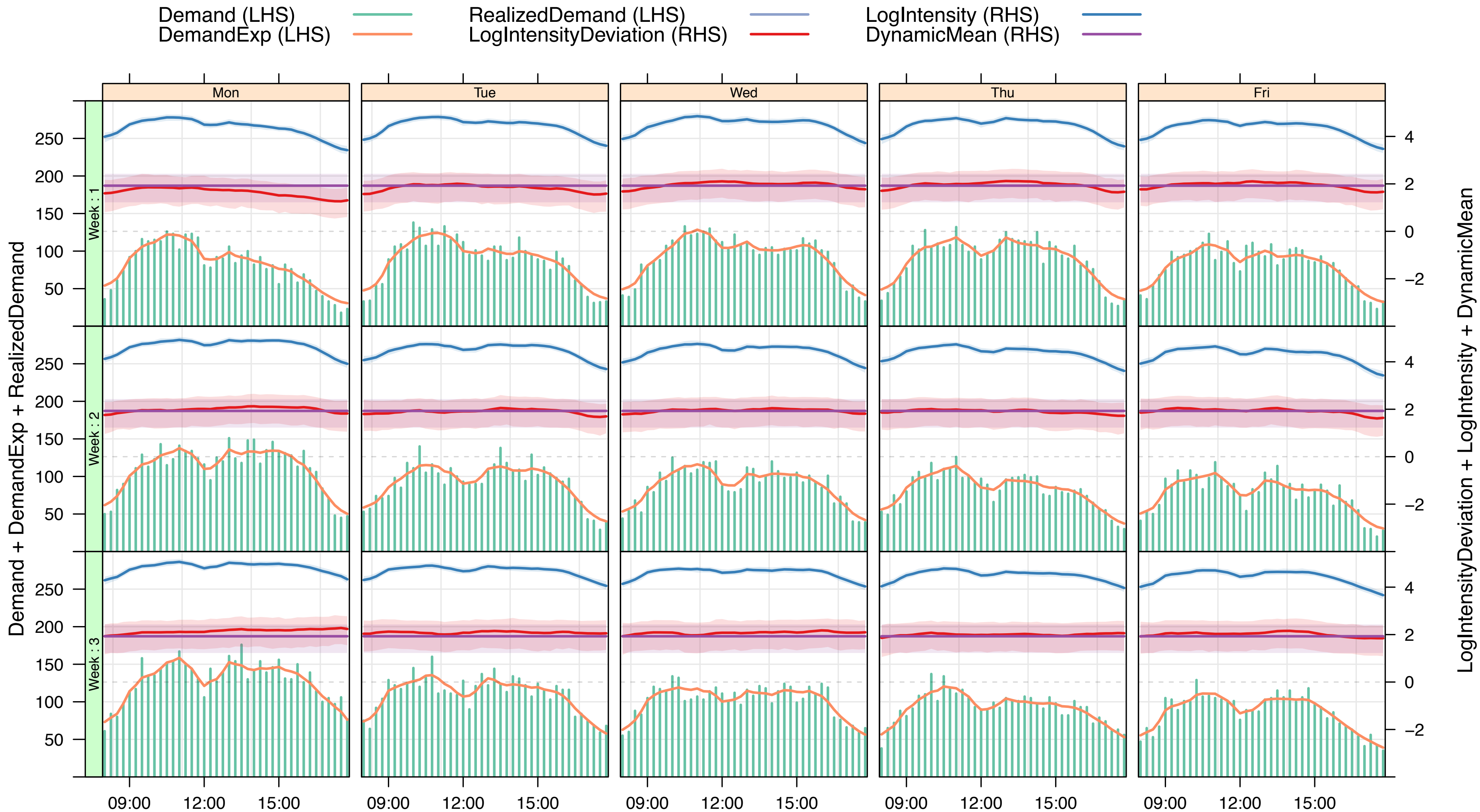
$$S_{t,W} = \begin{cases} \sum_{h=1}^{H} \eta_{t-5,h}, & \text{if } t > 5, \\ 0, & \text{otherwise,} \end{cases}$$

$$\epsilon_{t,1} \sim \mathcal{N}(0, \tau^2 + \tau_0^2),$$
$$\epsilon_{t,h} \sim \mathcal{N}(0, \tau^2), \qquad h > 1.$$

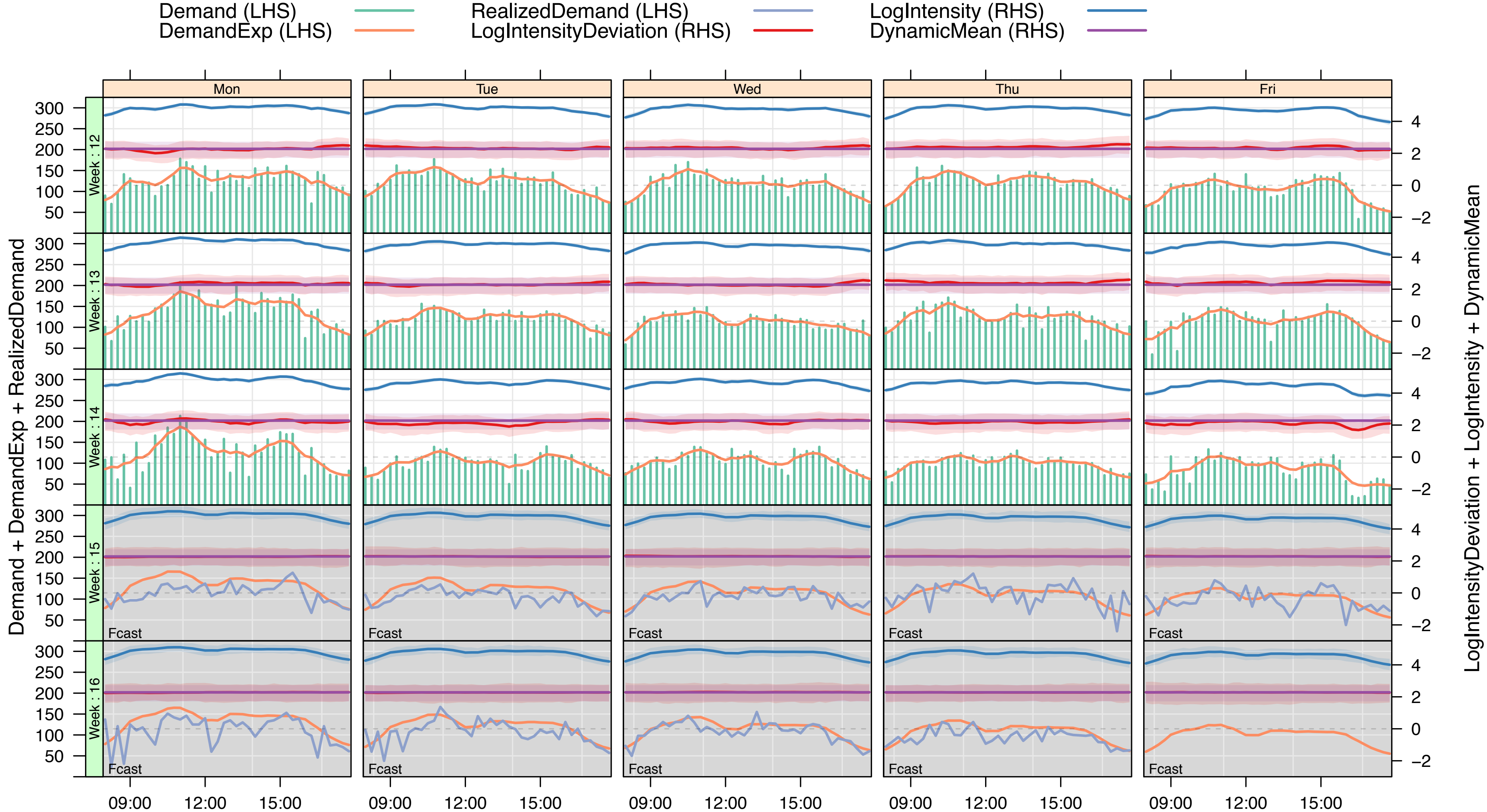
$$\alpha \sim \text{Exp.}(0.01), \qquad \phi_I \sim \text{Beta}(10, 1),$$
$$\tau_0 \sim \Gamma(1.1, 2), \qquad \phi_D \sim \text{Beta}(3, 6),$$
$$\tau \sim \Gamma(1.1, 2), \qquad \phi_W \sim \text{Beta}(3, 6),$$
$$\sigma \sim \Gamma(1.1, 5), \qquad \boldsymbol{\theta} \sim \mathcal{N}(0, \sigma).$$



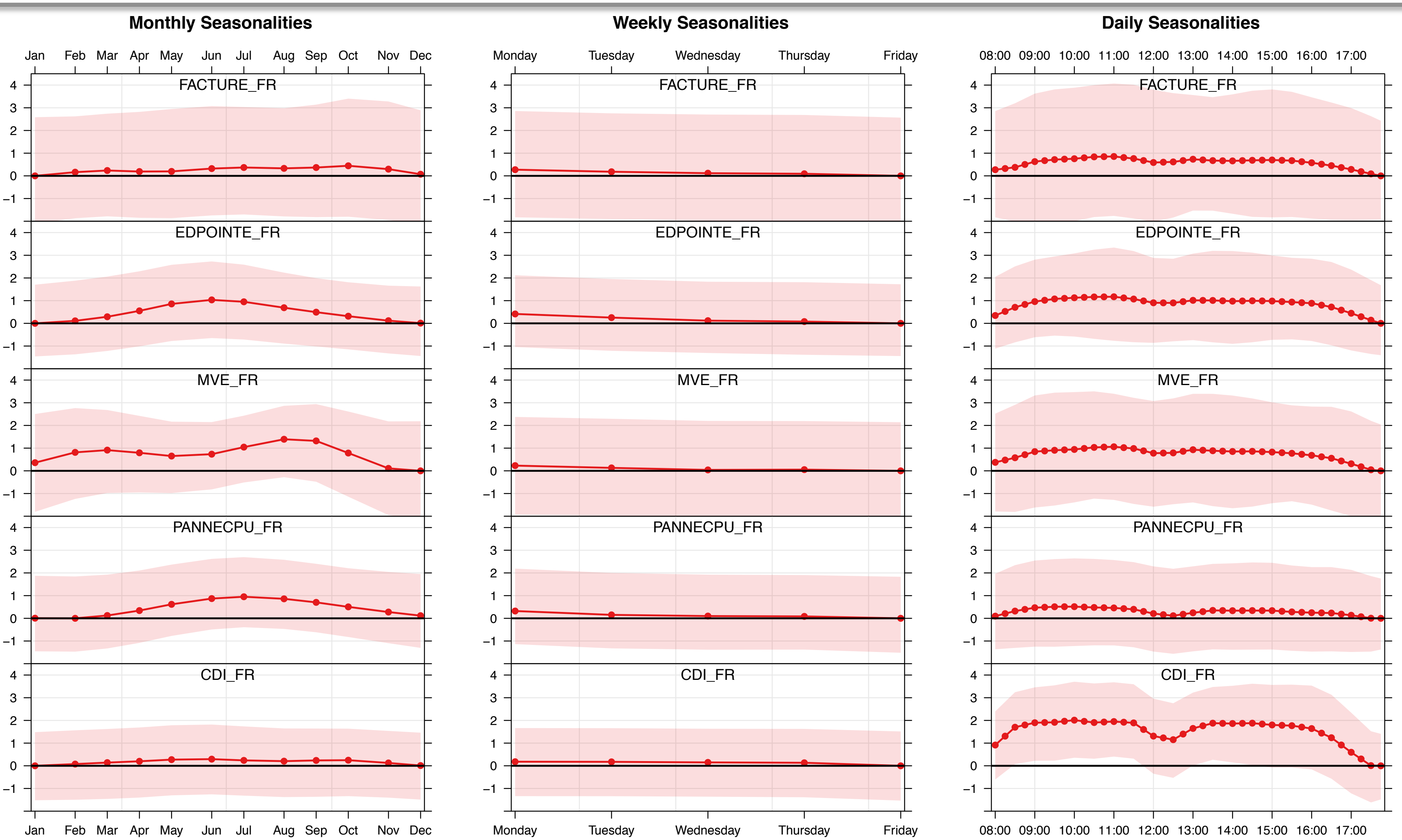
Data Fit Example



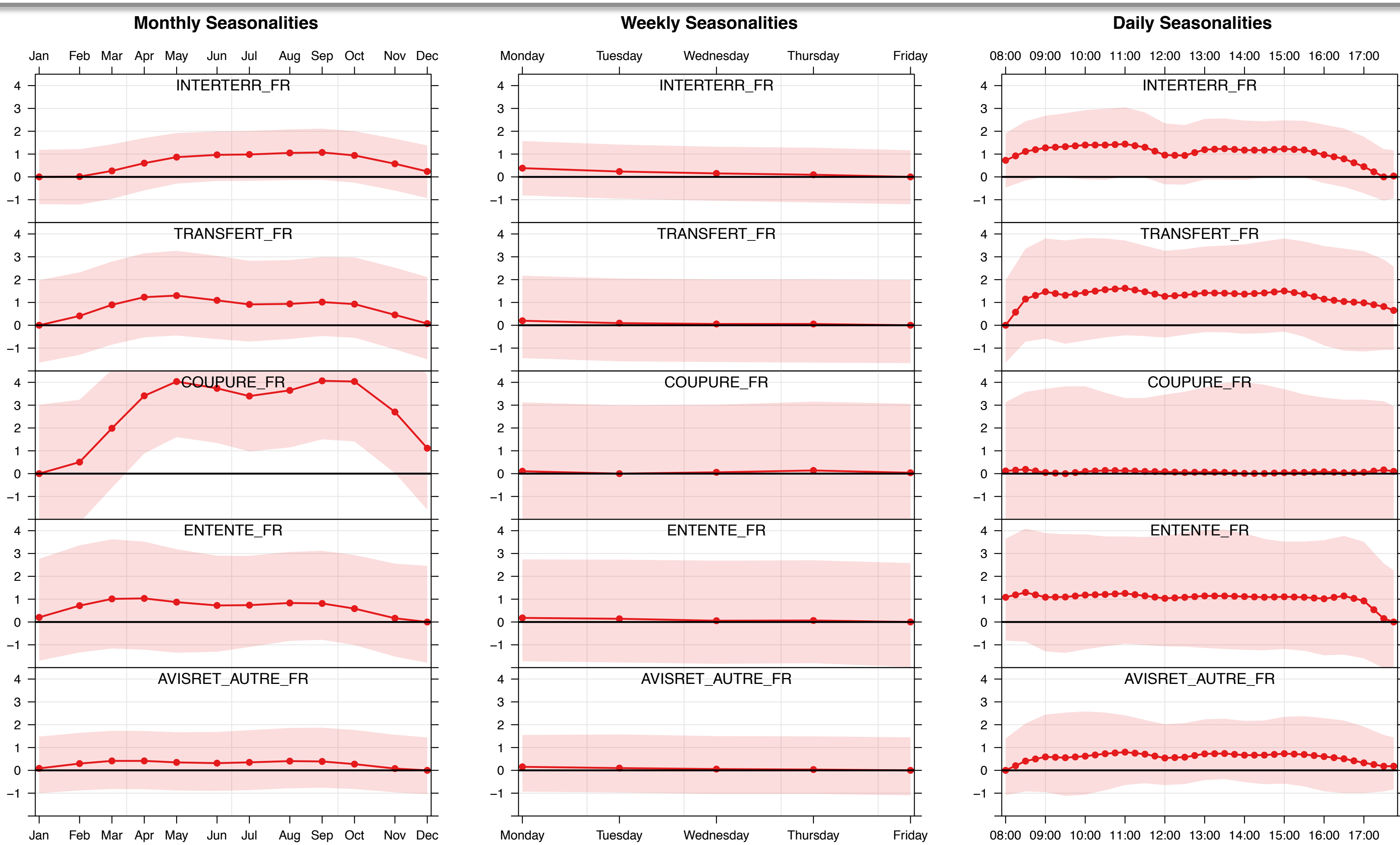
Fit and Forecast Example



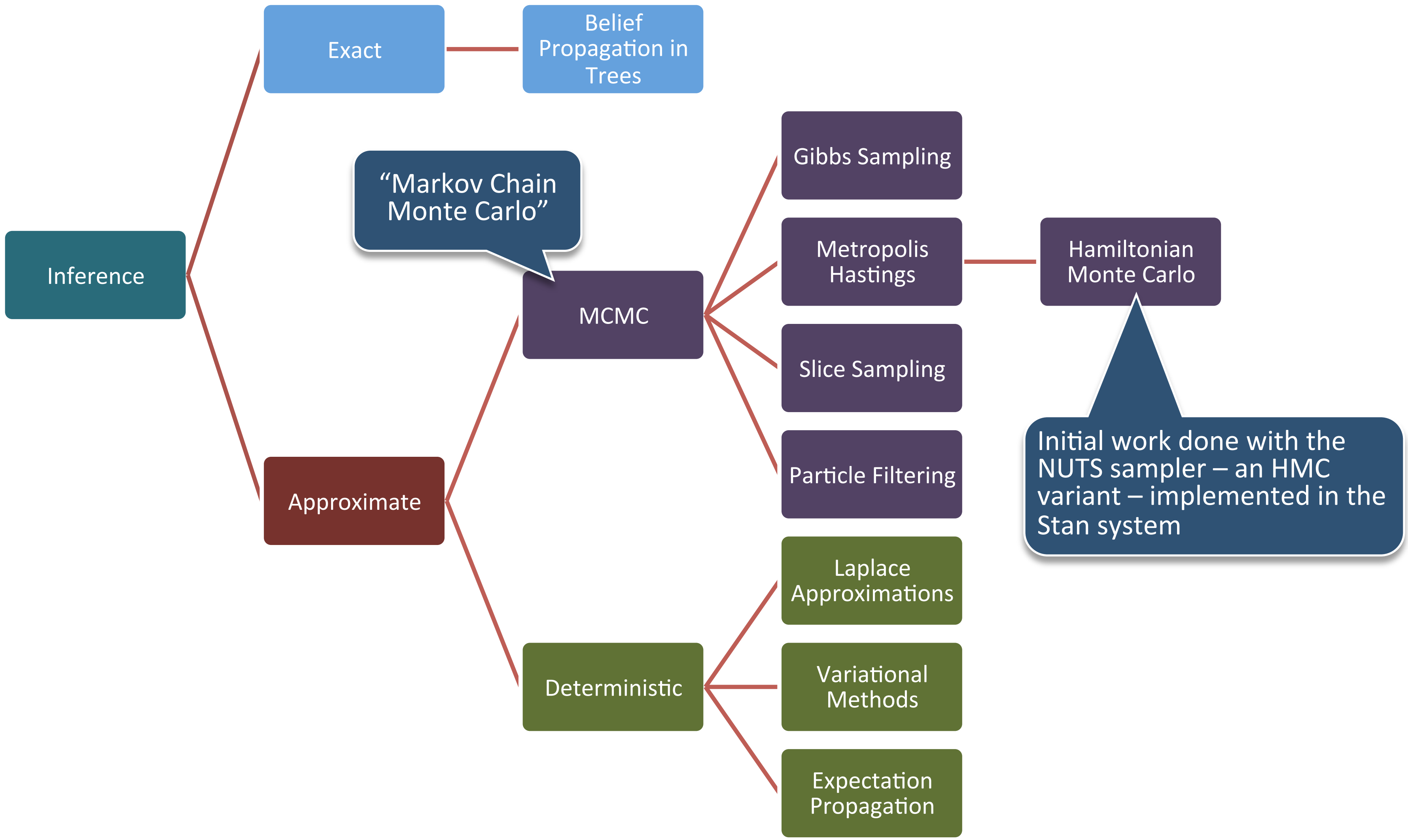
Seasonality Patterns



Seasonality Patterns (cont'd)

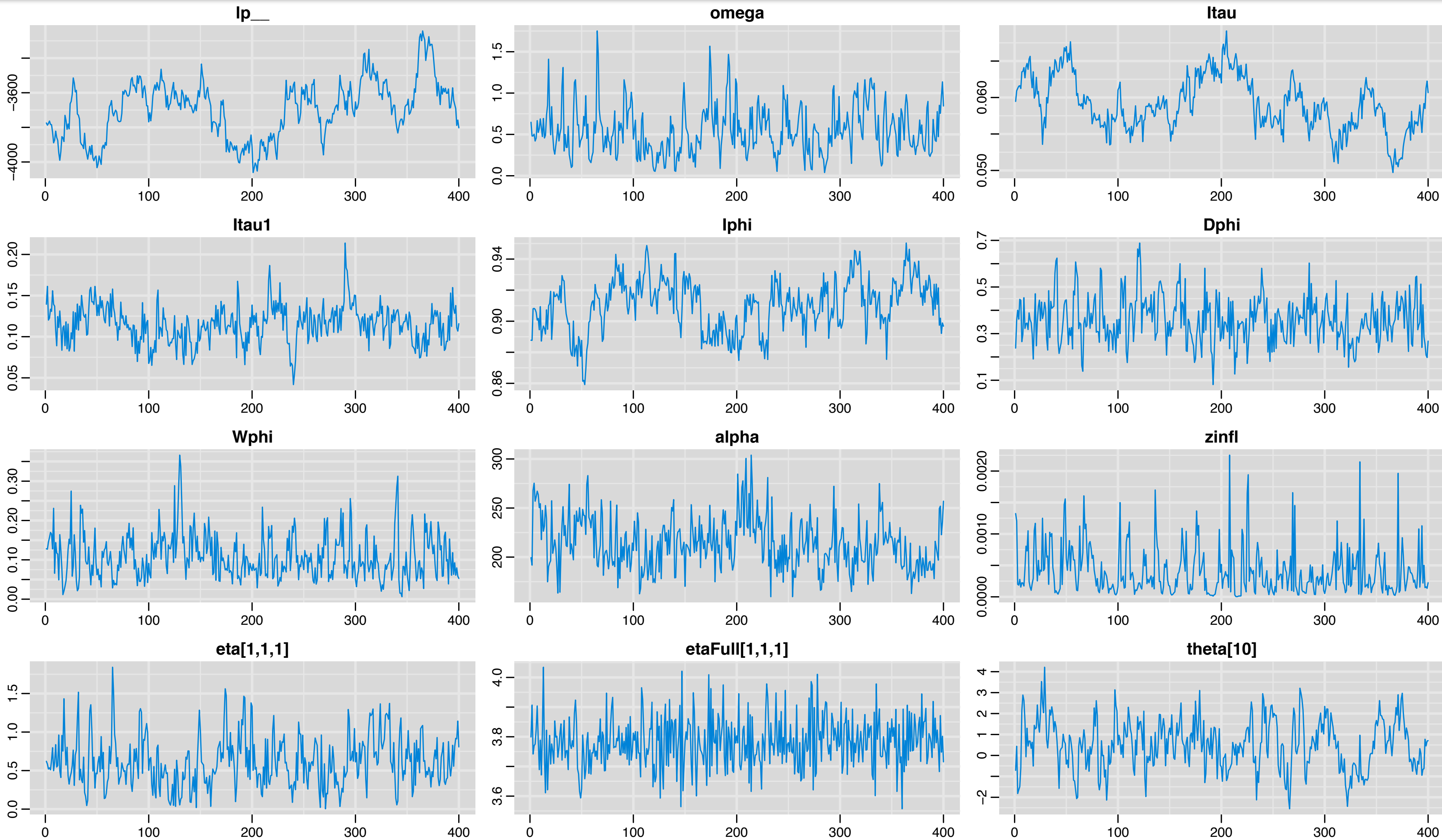


Landscape of Inference Methods

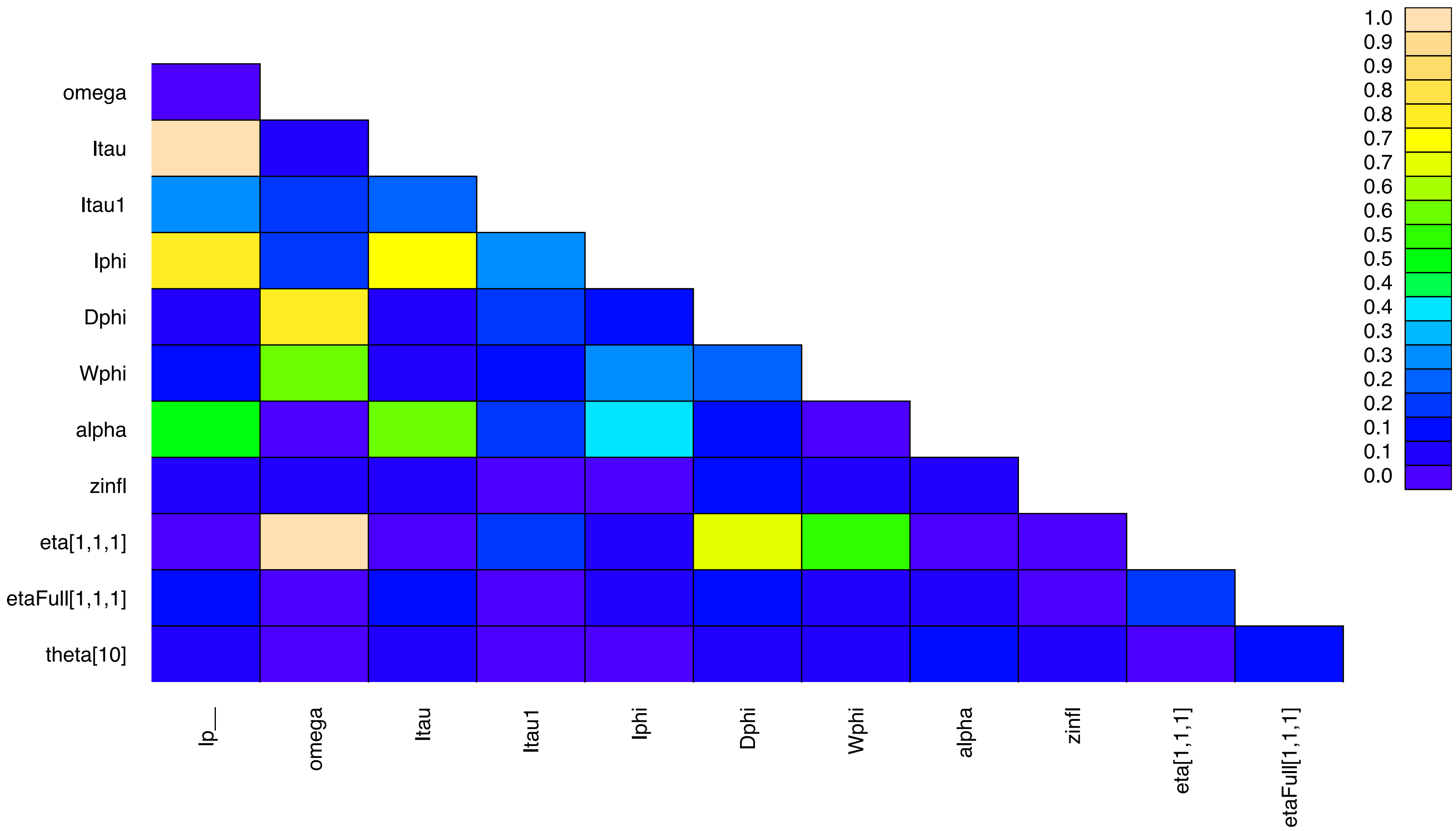


MCMC Mixing Example

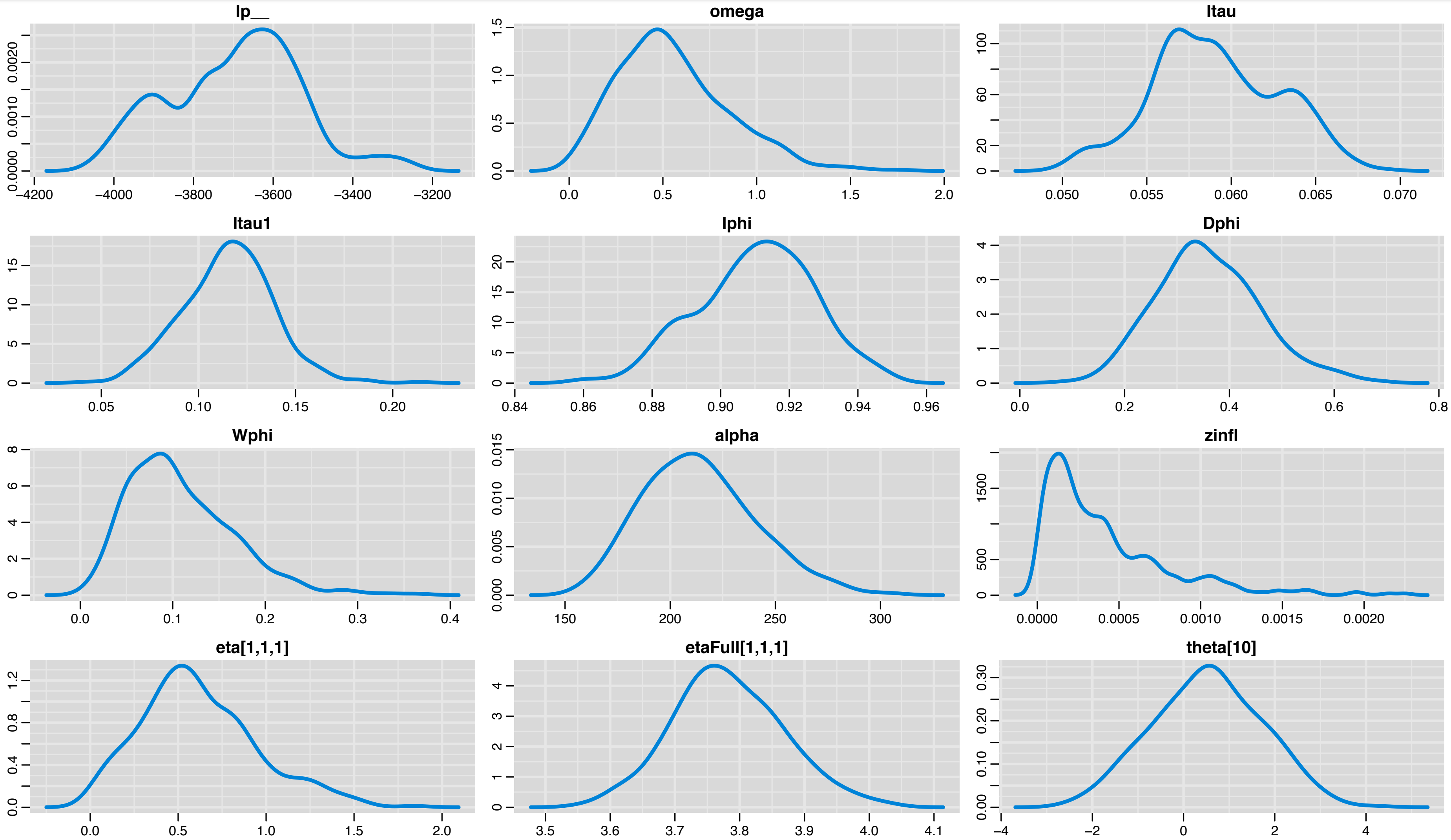
HMC+NUTS are very efficient



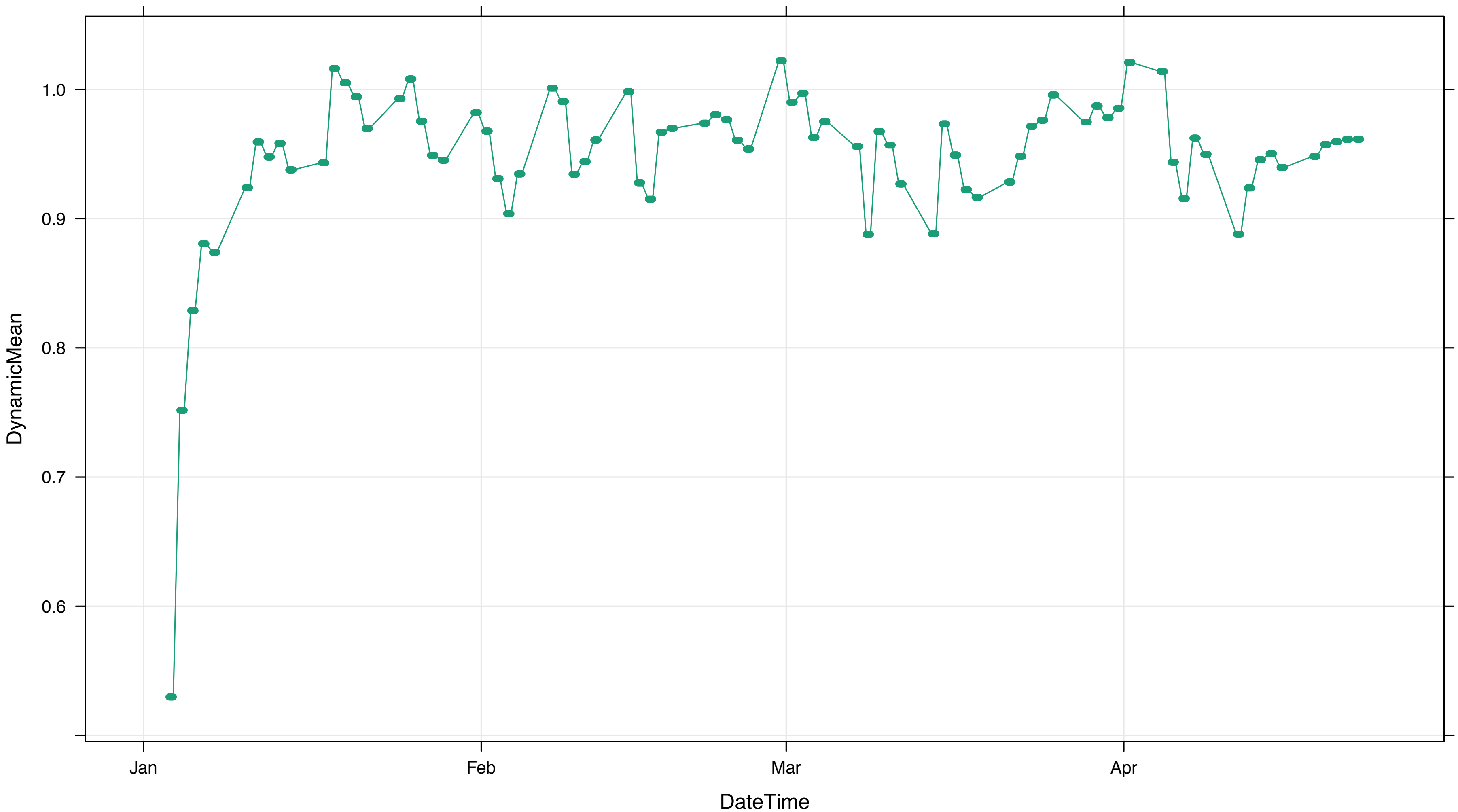
Sampling Correlations Example



Posterior Marginal Densities Example



Time Evolution of Process Dynamical Mean



Experimental Evaluation

- Seasonal regressors
  - Month of the year
  - Day of the week
  - 30-minute interval
  - 37 indicators in total
  - Independent (currently)

- Some smoothing applied so that we don't abruptly shift from one to the other

Benchmark model is Poisson generalized linear model:

$$y_{t,h} \sim \mathcal{P}(\exp(\eta_{t,h})),$$
$$\eta_{t,h} = \boldsymbol{\theta}'\mathbf{x}_{t,h}.$$

- Same calendar regressors
- Estimation criterion: maximum likelihood with L1-norm penalty
  - Drives to zero non-significant coefficients

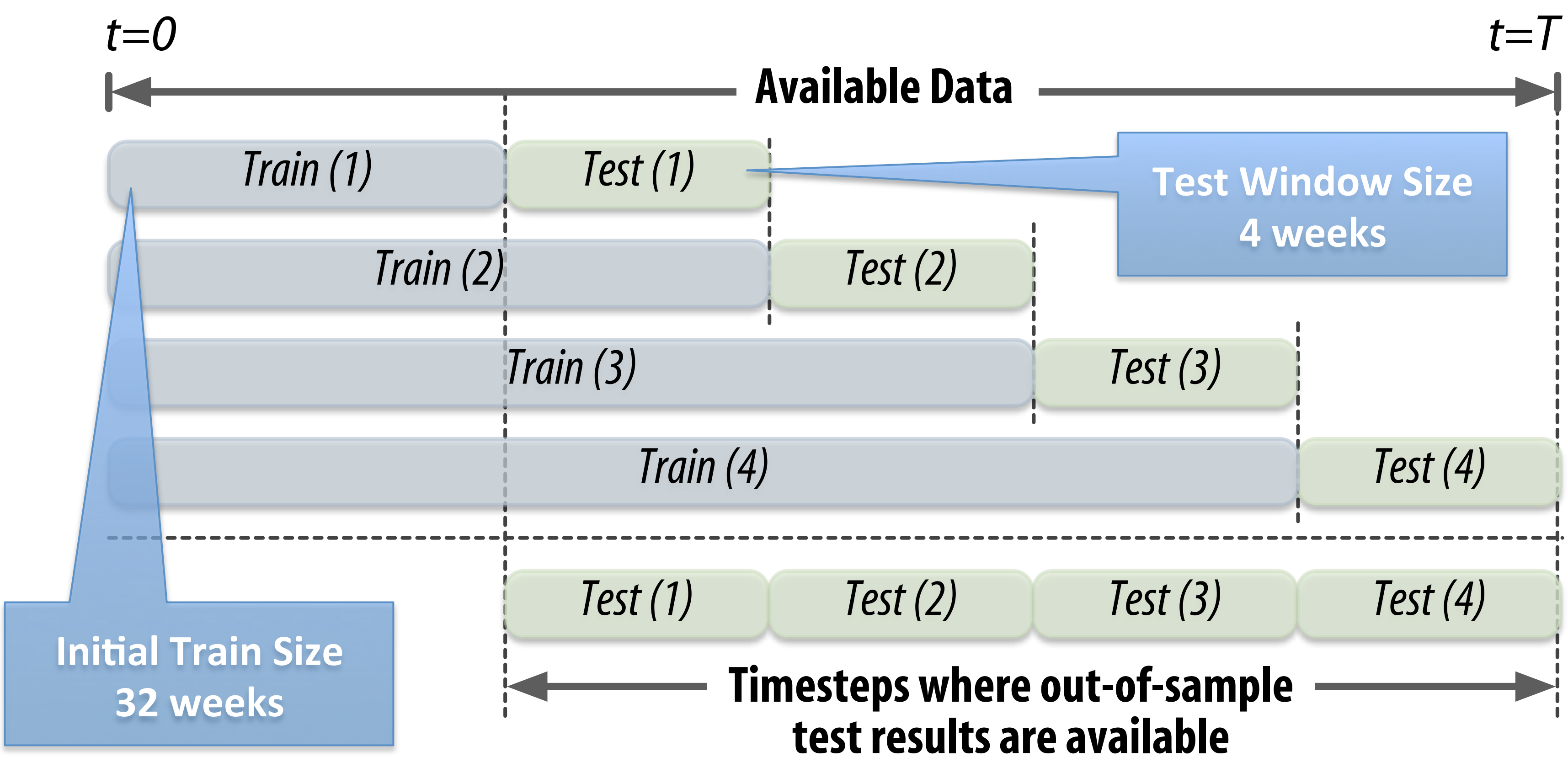
In-Sample Performance

Full NBSS model vs Poisson Regression (with full 52-week sample)

	NBSS						Poisson					
	RMSE	MAPE	99% Cov	99% Wid	95% Cov	95% Wid	RMSE	MAPE	99% Cov	99% Wid	95% Cov	95% Wid
FACTURE_FR	12,59	7,73	0,98	78,01	0,96	60,08	20,50	12,60	0,84	55,97	0,74	42,84
EDPOINTE_FR	7,98	11,81	0,99	43,52	0,96	33,37	13,57	18,31	0,88	37,73	0,78	28,97
MVE_FR	5,72	12,88	0,99	32,88	0,97	24,96	9,29	19,96	0,92	31,35	0,83	24,15
PANNECPU_FR	5,37	19,39	1,00	24,17	0,99	18,62	36,74	76,77	0,71	22,49	0,57	17,33
CDI_FR	4,09	24,37	0,99	21,98	0,96	16,97	4,67	27,81	0,96	20,10	0,90	15,60
INTERTERR_FR	3,68	23,38	0,99	19,83	0,95	15,35	4,23	26,87	0,98	19,17	0,91	14,87
TRANSFERT_FR	3,66	25,11	0,99	19,69	0,96	15,33	4,37	29,65	0,96	18,44	0,90	14,33
COUPURE_FR	4,14	25,44	0,99	23,05	0,96	14,47	5,98	41,50	0,92	17,46	0,85	13,63
ENTENTE_FR	4,32	26,50	0,99	24,00	0,97	20,77	5,85	36,84	0,87	17,83	0,77	13,82
AVISRET_AUTRE_FR	3,17	27,17	0,99	19,44	0,95	13,64	3,94	33,56	0,94	15,83	0,87	12,70

Sequential Validation

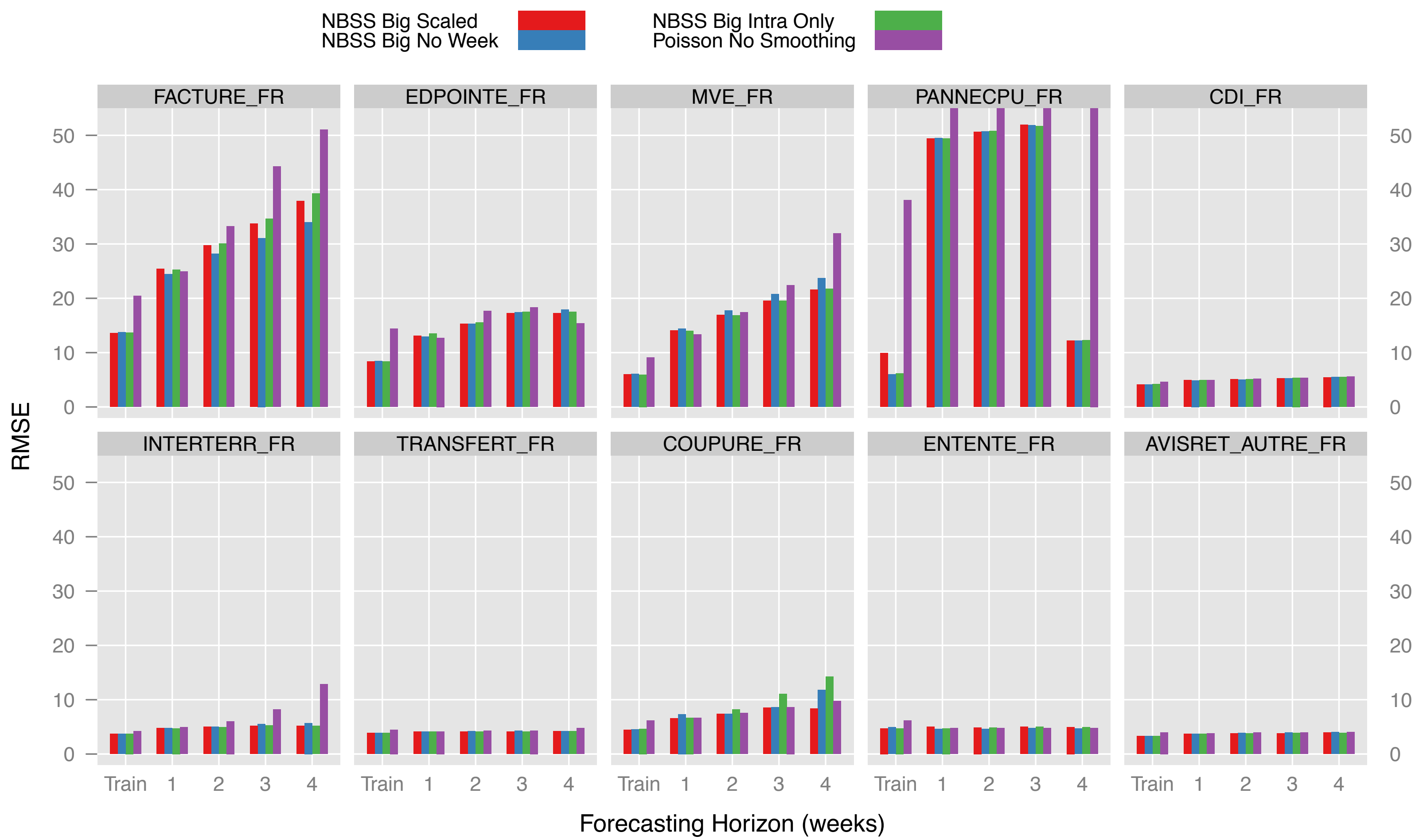
- Alternate model training and out-of-sample testing
- Simulate actions of a decision-maker acting in real time



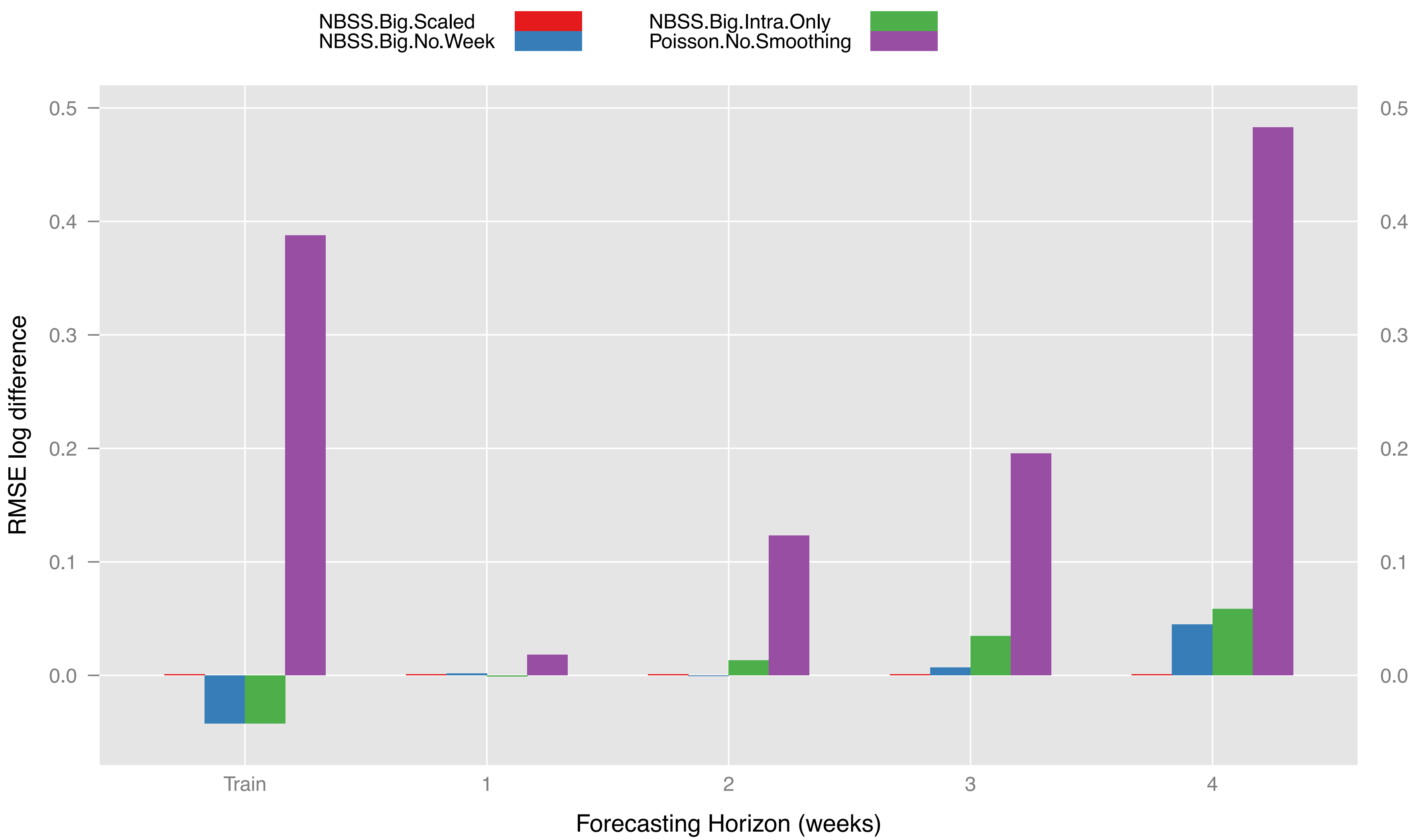
Basic Comparisons: Models

Model Name	Meaning
NBSS Big Scaled	Negative-Binomial State Space Model, daily and weekly persistence
NBSS Big No-Week	Negative-Binomial State Space Model, daily persistence only
NBSS Big Intra Only	Negative-Binomial State Space Model, no daily nor weekly persistence
Poisson No Smoothing	Poisson GLM benchmark

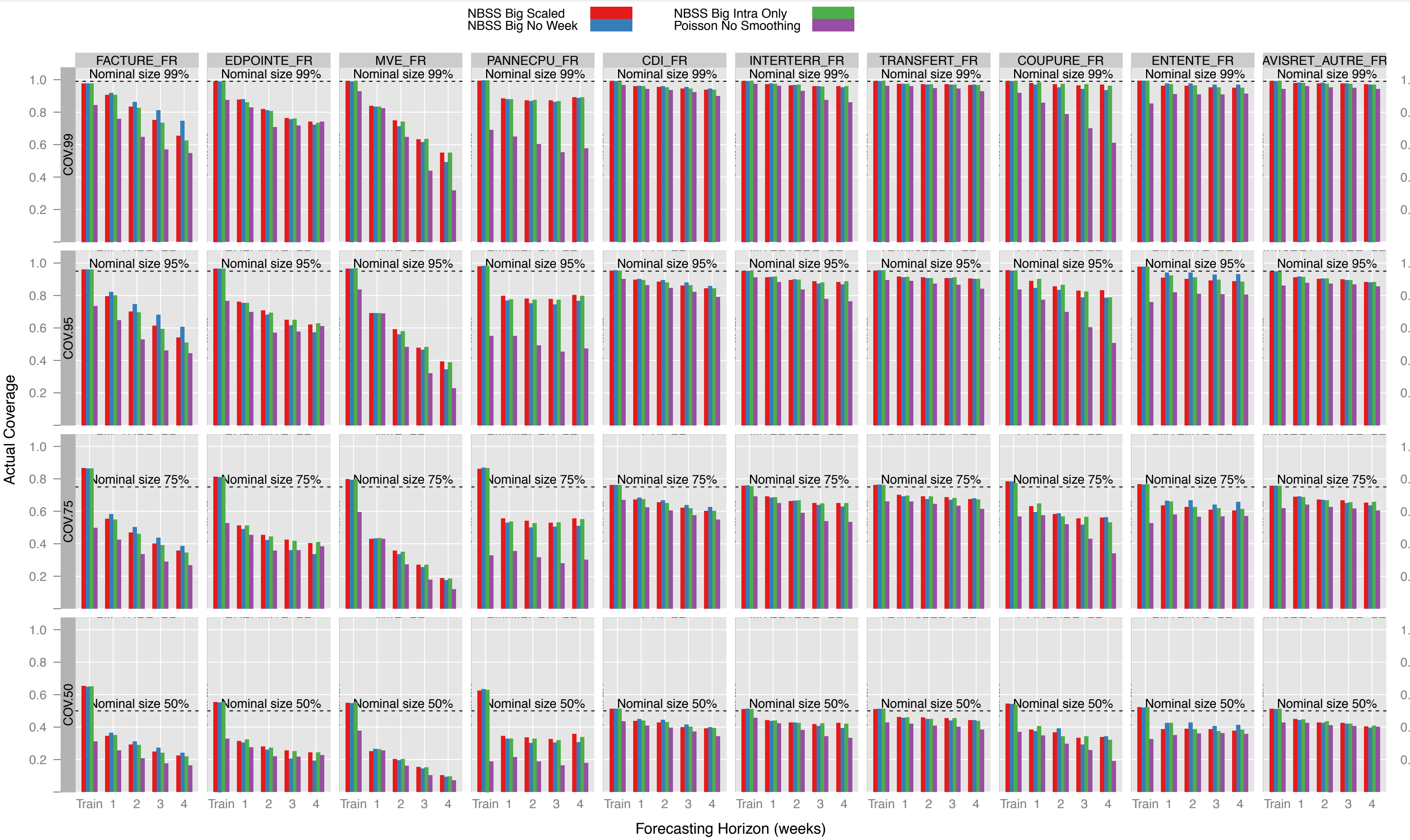
RMSE Performance: Absolute per Queue
(lower is better)



RMSE Performance: Relative
(Positive number: performance is worse than baseline)



Interval Coverage Performance



Generalizations and Conclusions

- Multiple queues: multidimensional latent state
- Small formulation changes:

$$y^i_{t,h} \sim \mathrm{Neg.\,Bin.}(\exp(\mathbf{b}_i \vec{\eta}_{t,h}), \alpha),$$
$$\vec{\eta}_{t,h} = \vec{\mu}_t + \mathbf{A}(\vec{\eta}_{t,h-1} - \vec{\mu}_t) + \vec{\epsilon}_{t,h}.$$

- Flexible model
  - Arbitrary seasonal and causal factors
  - Handle missing values
  - Fat tails
- Good performance
- Future work
  - Validation on more data and benchmark models
  - Other call centers
  - Adding special events