Biometrics ??, 1–??
?? 0000
DOI: 000

Web-based Supplementary Materials for: Registration for exponential family functional data

Julia Wrobel[1,*], Vadim Zipunnikov[2], Jennifer Schrack[3,4], and Jeff Goldsmith[1]

[1]Department of Biostatistics, Mailman School of Public Health, Columbia University

[2]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University

[3]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University

[4]Longitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging, National Institutes of Health

**email:* jw3134@cumc.columbia.edu

This paper has been submitted for consideration for publication in *Biometrics*

Supplementary Materials

2 August 2018

Appendix A: methods

Here we provide extra details of our methods.

Variational approximation to Bernoulli likelihood

Our method for binary functional principal components analysis uses a variational approximation to the Bernoulli likelihood given in Equation (6) of our manuscript. Recall that for subject i measured at time j observation $Y_i(t_{ij}) \sim Bernoulli(\mu_i(t_{ij}))$ where $\mu_i(t_{ij})$ is defined in Equation (3). For convenience we rearrange the usual formulation of the Bernoulli distribution to get the probability density function given in Equation (4) by

$$\begin{aligned} P\{Y_i(t_{ij})|\boldsymbol{c}_i\} &= \mu_i(t_{ij})^{Y_i(t_{ij})}\{1-\mu_i(t)\}^{1-Y_i(t_{ij})} \\ &= g^{-1}\{A_i(t_{ij})\}^{Y_i(t_{ij})}\left[1-g^{-1}\{A_i(t_{ij})\}\right]^{1-Y_i(t_{ij})} \\ &= g^{-1}\left[\{2Y_i(t_{ij})-1\}\,A_i(t_{ij})\right] \end{aligned}$$

where $A_i(t_{ij}) = \boldsymbol{\Theta}_\phi(t_{ij})\,(\boldsymbol{\alpha}_\Theta + \boldsymbol{\Psi}_\Theta \boldsymbol{c}_i)$. Equality holds in the last step because observations are limited to $\{0,1\}$ and when g is the logit function $g^{-1}(-z) = 1 - g^{-1}(z)$.

We use the variational approximation for logistic regression outlined by Jaakkola and Jordan (1997) and extended to binary PCA by Tipping and Bishop (1999), with additional basis expansion $\boldsymbol{\Theta}_\phi(t_{ij})$ embedded in $A_i(t_{ij})$ to allow for a functional data framework. This approximation is a lower bound on $P\{Y_i(t_{ij})|\boldsymbol{c}_i\}$, given by

$$
\begin{aligned}
P\{Y_i(t_{ij})|\boldsymbol{c}_i\} &= g^{-1}\left[\{2Y_i(t_{ij})-1\}\,A_i(t_{ij})\right] \\
&\geqslant g^{-1}\{\xi_i(t_{ij})\}\exp\left[\frac{\{2Y_i(t_{ij})-1\}\,A_i(t_{ij})-\xi_i(t_{ij})}{2}+\lambda\left\{\xi_i(t_{ij})\right\}\left\{A_i(t_{ij})^2-\xi_i(t_{ij})^2\right\}\right] \\
&= \widetilde{P}\{Y_i(t_{ij})|\boldsymbol{c}_i,\xi_i(t_{ij})\},
\end{aligned}
$$

where $\lambda(z)=\frac{0.5-g^{-1}(z)}{2z}$ and $\xi_i(t_{ij})$ is the variational parameter. Equality of the original distribution $P\{Y_i(t_{ij})|\boldsymbol{c}_i\}$ and the variational distribution is attained when $\widetilde{P}\{Y_i(t_{ij})|\boldsymbol{c}_i,\xi_i(t_{ij})\}$ is maximized with respect to $\xi_i(t_{ij})$, and the value of $\xi_i(t_{ij})$ at the maximum is $\{2Y_i(t_{ij})-1\}\,A_i(t_{ij})$.

Updating $\boldsymbol{\alpha}_\Theta$ and $\boldsymbol{\Psi}_\Theta$ for binary FPCA

We obtain parameter updates for binary FPCA by maximizing the variational likelihood given in Equation (7) of our manuscript. In Section (3.1.3) we obtain updates for $\boldsymbol{\alpha}_\Theta$ and $\boldsymbol{\Psi}_\Theta$ by reparameterizing Equation (7) such that $\boldsymbol{\Phi}=(\boldsymbol{\Psi}_\Theta^T,\boldsymbol{\alpha}_\Theta)^T$. This reparameterization leads to the variational log-likelihood below:

$$
\begin{aligned}
\widetilde{l}(\boldsymbol{Y},\boldsymbol{c}) &\propto \sum_{j=1}^{D_i}\sum_{i=1}^{I}\log\widetilde{P}\left\{Y_i(t_{ij})|\boldsymbol{c}_i,\xi_i(t_{ij})\right\}-\sum_i \boldsymbol{c}_i^T\boldsymbol{c}_i \\
&\propto \sum_i\left[\left\{Y_i(\boldsymbol{t}_i)-\frac{1}{2}\right\}^T A_i(\boldsymbol{t}_i)-\frac{1}{2}\xi_i(\boldsymbol{t}_i)\mathbb{1}_{D_i\times 1}+A_i^T(\boldsymbol{t}_i)diag\left[\lambda\left\{\xi_i(\boldsymbol{t}_i)\right\}\right]A_i(\boldsymbol{t}_i)\right] \\
&\propto \sum_i\left\{Y_i(\boldsymbol{t}_i)-\frac{1}{2}\right\}^T\left\{\boldsymbol{\Theta}_\phi(\boldsymbol{t}_i)\otimes\boldsymbol{s}_i^T\right\}vec(\boldsymbol{\Phi}) \\
&+ \sum_i vec(\boldsymbol{\Phi})^T(\boldsymbol{\Theta}_\phi\left\{\boldsymbol{t}_i\right\}^T\otimes\boldsymbol{s}_i\}\,diag\left[\lambda\left\{\xi_i(\boldsymbol{t}_i)\right\}\right]\left\{\boldsymbol{\Theta}_\phi(\boldsymbol{t}_i)\otimes\boldsymbol{s}_i^T\right\}vec(\boldsymbol{\Phi}).
\end{aligned}
$$

Maximizing with respect to $\boldsymbol{\Phi}$ gives estimates $\widehat{\boldsymbol{\Phi}}$.

Optimization constraints for the warping step

Section (3.3) refers to optimization constraints for the `R` function `constrOptim()` implemented in the warping step of our algorithm. We constrain inverse warping function to be monotonic with fixed endpoints, and these constraints are enforced through $\boldsymbol{\beta}_i$, the

warping function B-spline coefficients for each subject. To ensure that estimated inverse warping functions $\widehat{h}_i^{-1}$ span the same domain as chronological time t_i^*, we fix the outer coefficients $\beta_{i,1}$ and β_{i,K_h}. Thus in practice we estimate the $K_h - 2$ inner spline coefficients $\boldsymbol{\beta}_{i,inner} = (\beta_{i,2}, ..., \beta_{i,K_h-1})^T$.

To enforce monotonicity of the warping functions we must ensure $\beta_1 < \beta_2 < ... < \beta_{K_h-1}$. Using the notation from the `constrOptim()` function, we define a matrix ui and a vector ci such that

$$ui \times \boldsymbol{\beta}_{i,inner} - ci \geqslant 0. \tag{A.1}$$

This leads to a ui matrix of dimension $(K_h - 1) \times (K_h - 2)$ and a size $(K_h - 1)$ vector, ci, that take the forms:

$$ui = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ -1 & 1 & 0 & ... & 0 \\ 0 & -1 & 1 & \ddots & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & & \ddots & -1 & 1 \\ 0 & & & 0 & -1 \end{pmatrix}$$

and

$$ci = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}$$

such that

$$\begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ -1 & 1 & 0 & ... & 0 \\ 0 & -1 & 1 & & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & & & -1 & 1 \\ 0 & \cdots & \cdots & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_{i,2} \\ \\ \beta_{i,3} \\ \vdots \\ \\ \beta_{i,K_h-1} \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \\ -1 \end{pmatrix} > \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \\ 0 \end{pmatrix}.$$

Analytic gradient for exponential family registration

For the general exponential family case, this gradient is

$$\frac{dl(Y_i(\boldsymbol{t}_i^*), \boldsymbol{\beta}_i)}{d\boldsymbol{\beta}_i} = \frac{1}{\varphi} \sum_{j=1}^{D_i} \left\{ \left(Y_i(t_{ij}^*) - b' \left[g \left\{ \mu_i(t_{ij}) \right\} \right] \right) \times \Theta_h(t_{ij}^*)^T \Theta_\phi' \{ \boldsymbol{\Theta}_h(t_{ij}^*) \boldsymbol{\beta}_i \} \left(\boldsymbol{\alpha}_\Theta + \boldsymbol{\Psi}_\Theta \boldsymbol{c}_i \right) \right\}, \tag{A.2}$$

where $\boldsymbol{\Theta}_\phi'(\boldsymbol{t}_i)$ is a $D_i \times K_h$ matrix of first derivatives of the B-spline basis functions used to reconstruct $\boldsymbol{t}_i$, and $b' \Big[g \left\{ \mu_i(t_{ij}) \right\} \Big] = \mu_i(t_{ij}) = g^{-1} \left[\Theta_\phi \{ \boldsymbol{\Theta}_h(t_{ij}^*) \boldsymbol{\beta}_i \} (\boldsymbol{\alpha}_\Theta + \boldsymbol{\psi}_\Theta \boldsymbol{c}_i) \right]$. For the Bernoulli loss function $\varphi = 1$ and $g^{-1}(z) = \frac{1}{1+e^{-z}}$, so the gradient becomes

$$\frac{dl\{Y_i(\boldsymbol{t}_i^*), \boldsymbol{\beta}_i\}}{d\boldsymbol{\beta}_i} = \sum_{j=1}^{D_i} \left[\left(Y_i(t_{ij}^*) - \frac{1}{1 + e^{-\Theta_\phi \{ \boldsymbol{\Theta}_h(t_{ij}^*) \boldsymbol{\beta}_i \} (\boldsymbol{\alpha}_\Theta + \boldsymbol{\psi}_\Theta \boldsymbol{c}_i)}} \right) \times \Theta_h(t_{ij}^*)^T \Theta_\phi' \{ \boldsymbol{\Theta}_h(t_{ij}^*) \boldsymbol{\beta}_i \} \left(\boldsymbol{\alpha}_\Theta + \boldsymbol{\Psi}_\Theta \boldsymbol{c}_i \right) \right]. \tag{A.3}$$

Appendix B: simulations and analysis

Here we provide extra results from simulations and analysis of BLSA data.

Functional principal components for BLSA data

Figure (Web.App.1) shows the effects of the estimated principal component basis functions for the BLSA data after the registration process. The first principal component is a vertical shift around the population mean, $\alpha(t)$, indicating a higher or lower probability of being active. More interesting is the second principal component, which shows that some subjects have higher probability of activity earlier in the day, while others have higher probability of activity later in the day.

[Figure 1 about here.]

Optimizing parameters

As a sensitivity analysis we evaluate our method as a function of parameters K_ϕ and K_h. We evaluated all combinations of $K_\phi \in \{5, 10, 15\}$, $K_h \in \{3, 4, 5, 6\}$ and grid length $D \in \{50, 100, 200\}$ using the same simulation setup and performance metrics as in Section (4). Mean integrated squared errors are given in Figure (Web.App.2) and computation times across these simulation scenarios are given in Figure (Web.App.3).

[Figure 2 about here.]

[Figure 3 about here.]

Both MISE and computation time increase linearly with K_h. Mean integrated squared errors decrease slightly with increasing K_ϕ, and computation time slightly increases with increasing K_ϕ.

Sensitivity of BFPCA

Below we compare our binary functional principal components algorithm with that from Hall et al. (2008). Mean integrated squared errors are based on deviations from the population level mean.

[Figure 4 about here.]

Analysis of weekdays for BLSA

In our primary analysis we averaged across visits for subjects. Here we separate visits by day of the week and look at day-specific effects. Figure (Web.App.5) shows unregistered and registered binary and smooth curves for each day of the week. Our algorithm consistently identifies similar patterns across days of the week. Alignment may be slightly better on week days than weekends, which suggests an area for future exploration.

[Figure 5 about here.]

References

Hall, P., Müller, H.-G., and Yao, F. (2008). Modelling sparse generalized longitudinal observations with latent gaussian processes. *Journal of the Royal Statistical Society: Series B* **70,** 703–723.

Jaakkola, T. S. and Jordan, M. I. (1997). A variational approach to bayesian logistic regression models and their extensions. In *Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics*.

Tipping, M. E. and Bishop, C. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B* **61,** 611–622.

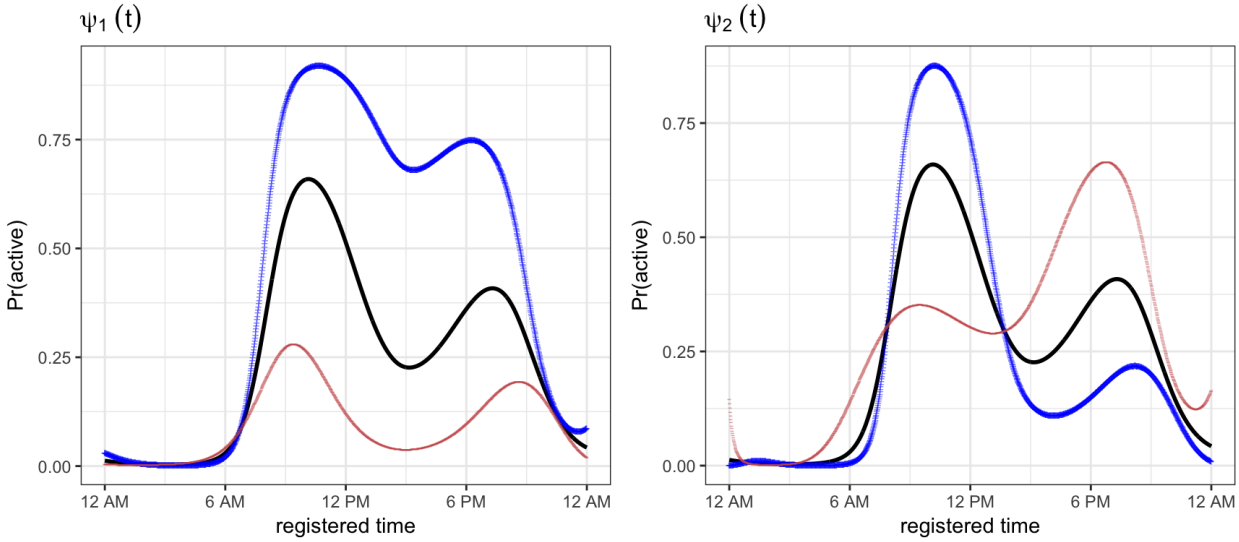

Figure Web.App.1. Estimated binary FPCA basis functions after registration process, illustrated by plotting $g^{-1}\left\{\alpha(t) \pm \psi_k(t)\right\}$ for basis functions $k \in \{1, 2\}$.

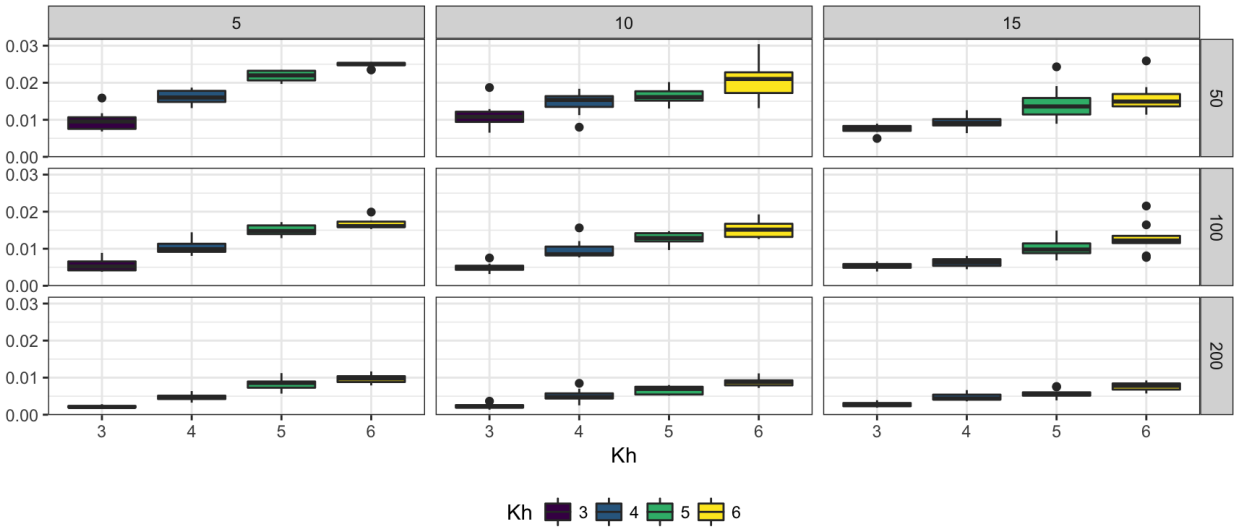

Figure Web.App.2. Parameter sensitivity across values of K_ϕ and K_h for *registr* method. Shown are mean integrated squared error (MISE) summaries across 10 datasets for each parameter scenario. Columns represent distinct values of K_ϕ and rows distinct grid lengths D.

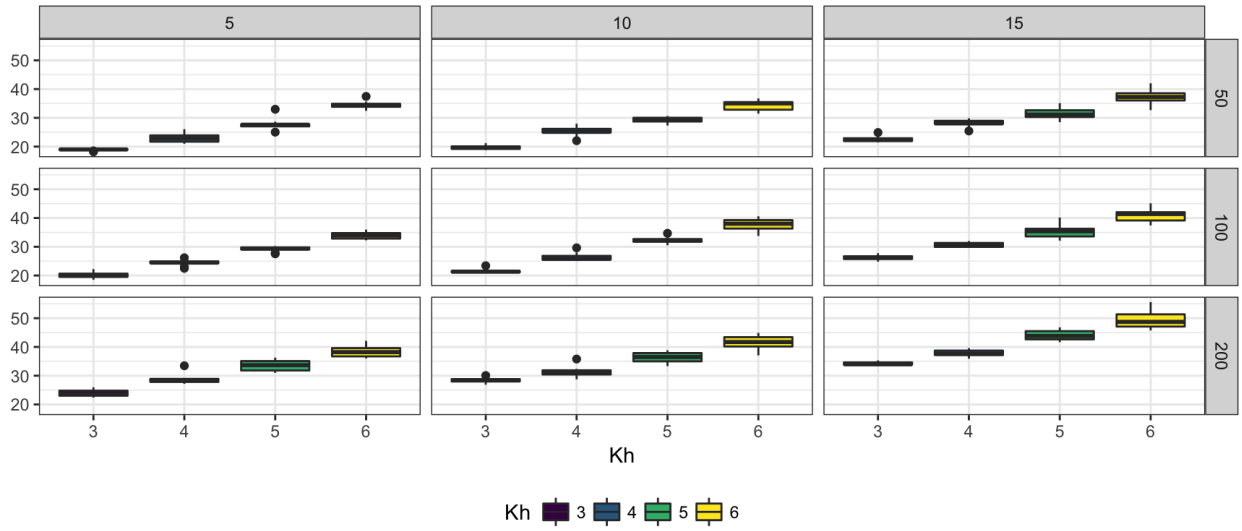

Figure Web.App.3. Parameter sensitivity across values of K_ϕ and K_h for *registr* method. Shown are boxplots of computation time (in seconds) across 10 datasets for each parameter scenario. Columns represent distinct values of K_ϕ and rows distinct grid lengths D.

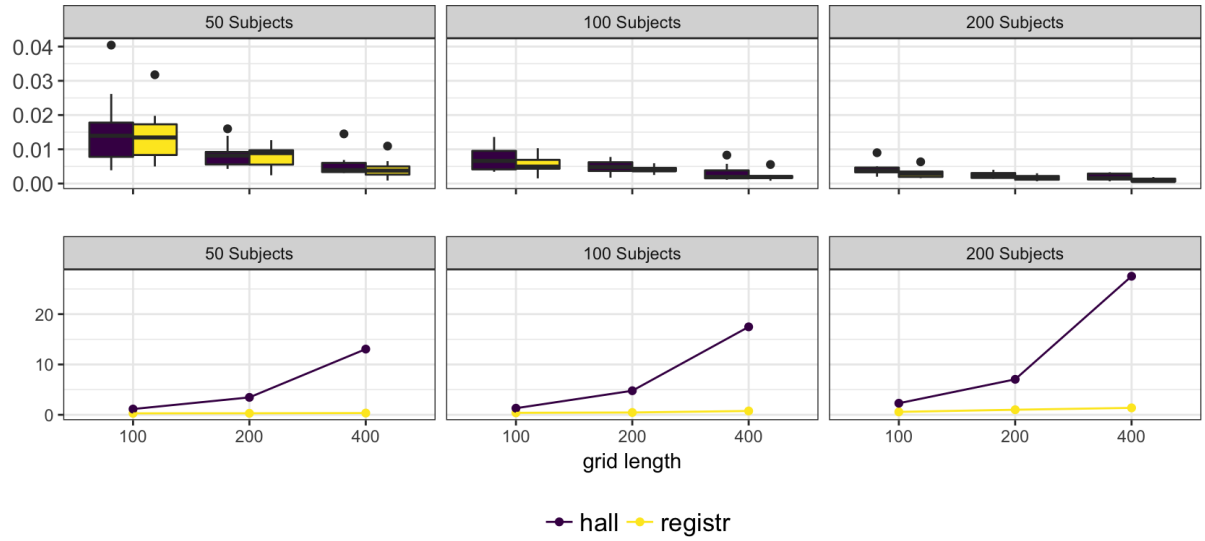

Figure Web.App.4. This figure shows mean integrated squared errors (top row) and median computation times (bottom row) for *hall* (in red) and *registr* (in green) methods across varying sample sizes and grid lengths. The columns, from left to right, show sample sizes 50, 100, and 200, respectively. Within each panel we compare grid lengths of 100, 200, and 400. Mean integrated squared errors are based on deviations from the population level mean.

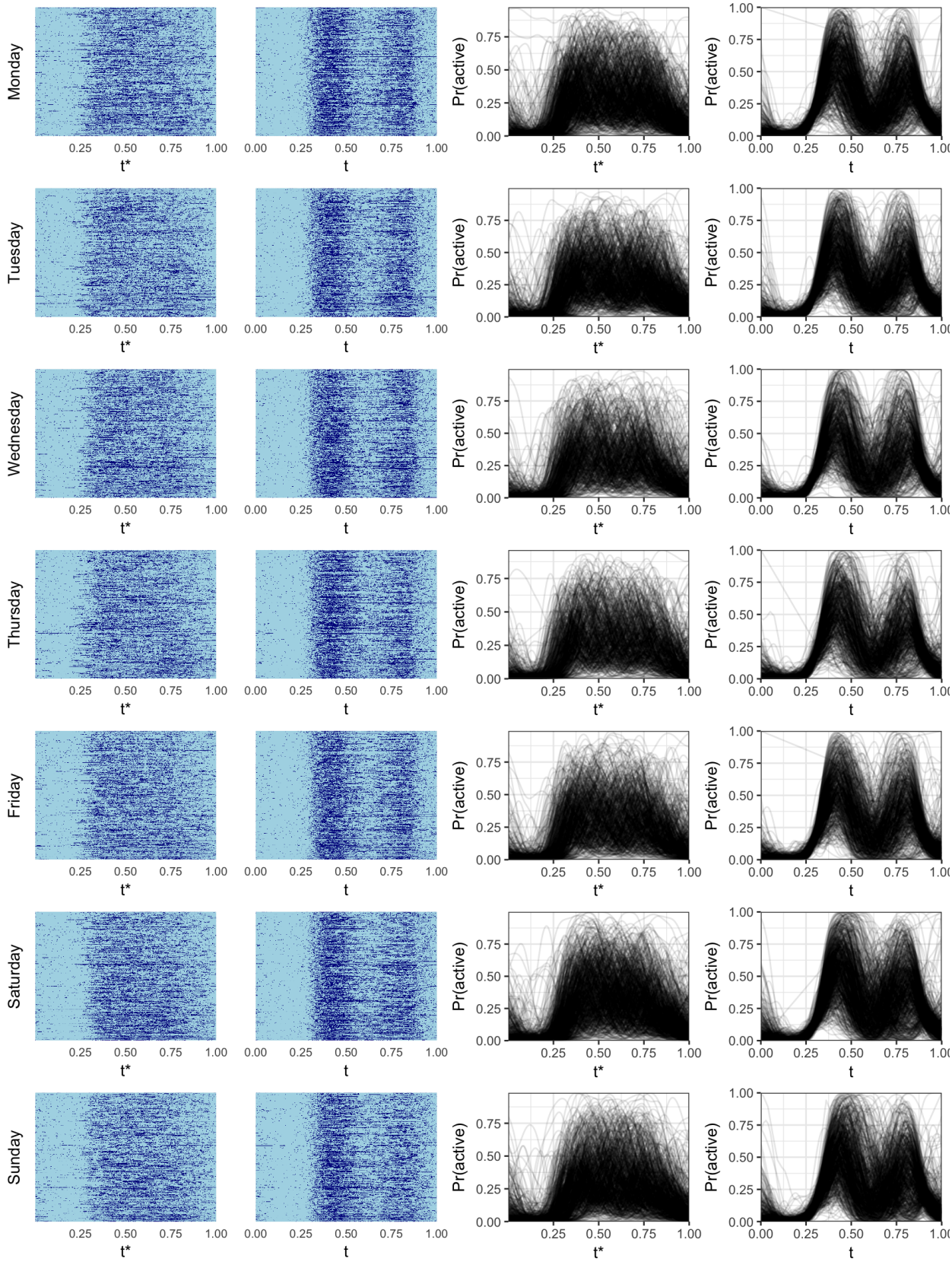

Figure Web.App.5. Analysis results for each day of the week.