Supporting Information

for

A Spatial Bayesian Latent Factor Model for Image-on-Image Regression

by

Cui Guo, Jian Kang and Timothy D. Johnson

Web Appendix A: Full Model

Our hierarchical full model is:

$$\text{Level 1}: \quad Z_i(v) = U(v) + \sum_{m=1}^{M} \theta_{im} b_m(v) + e_i(v)$$

$$\text{Level 2}: \quad \theta_{im} = \sum_{k=1}^{K} \lambda_{mk} \eta_{ik} + \zeta_{im}$$

$$\text{Level 3}: \quad \eta_{ik} = \sum_{v' \in \mathcal{R}} \widetilde{X_i}(v') \beta_k(v') + \epsilon_{ik}$$

$$\text{where} \quad \widetilde{X_i}(v') = \sum_{p=1}^{P} \gamma_p X_{ip}(v'), \quad \beta_k(v') = \sum_{m=1}^{M} \alpha_{km} b_m(v')$$

Combining the two sub-models on Level 1 and 2, we get

$$Z_i(v) = U(v) + \sum_{m=1}^{M} \left[\sum_{k=1}^{K} \lambda_{mk}\eta_{ik} + \zeta_{im} \right] b_m(v) + e_i(v)$$

$$= U(v) + \sum_{m=1}^{M}\sum_{k=1}^{K} \lambda_{mk}\eta_{ik}b_m(v) + \sum_{m=1}^{M} \zeta_{im}b_m(v) + e_i(v)$$

$$= U(v) + \sum_{k=1}^{K} \left[\sum_{m=1}^{M} \lambda_{mk}b_m(v) \right]\eta_{ik} + \tilde{\zeta}_i(v) + e_i(v)$$

$$= U(v) + \sum_{k=1}^{K} \tilde{\lambda}_k(v)\eta_{ik} + \tilde{\zeta}_i(v) + e_i(v), \tag{1}$$

where the spatially varying prediction effect on response voxel v from predictor voxel v' is

$$\tilde{\lambda}_k(v) = \sum_{m=1}^{M} \lambda_{mk}b_m(v), \qquad \tilde{\zeta}_i(v) = \sum_{m=1}^{M} \zeta_{im}b_m(v).$$

By plugging equations on Level 3 in equation (1), we get

$$Z_i(v) = U(v) + \sum_{k=1}^{K} \tilde{\lambda}_k(v)\eta_{ik} + \tilde{\zeta}_i(v) + e_i(v)$$

$$= U(v) + \sum_{k=1}^{K} \tilde{\lambda}_k(v) \left[\sum_{v'\in\mathcal{R}} \left(\sum_{p=1}^{P} \gamma_p X_{ip}(v') \right) \beta_k(v') \right] +$$

$$\sum_{k=1}^{K} \tilde{\lambda}_k(v)\epsilon_{ik} + \tilde{\zeta}_i(v) + e_i(v)$$

$$= U(v) + \sum_{p=1}^{P} \gamma_p \sum_{v'\in\mathcal{R}} \left[\sum_{k=1}^{K} \tilde{\lambda}_k(v)\beta_k(v') \right] X_{ip}(v') + \tilde{\epsilon}_i(v) + \tilde{\zeta}_i(v) + e_i(v)$$

$$= U(v) + \sum_{p=1}^{P} \gamma_p \sum_{v'\in\mathcal{R}} \psi(v, v')X_{ip}(v') + \tilde{\epsilon}_i(v) + \tilde{\zeta}_i(v) + e_i(v),$$

2

where

$$\boldsymbol{\psi}(v,v') = \sum_{k=1}^{K} \tilde{\lambda}_k(v)\beta_k(v')$$

$$= \sum_{k=1}^{K}\sum_{m=1}^{M} \lambda_{mk}b_m(v)\beta_k(v')$$

$$= \sum_{m=1}^{M} \Big[\sum_{k=1}^{K}\lambda_{mk}\beta_k(v')\Big]b_m(v)$$

$$= \sum_{m=1}^{M} \Big[\sum_{k=1}^{K}\lambda_{mk}\Big(\sum_{m'=1}^{M}\alpha_{km'}b_{m'}(v')\Big)\Big]b_m(v)$$

$$= \sum_{m=1}^{M}\sum_{m'=1}^{M} \Big[\sum_{k=1}^{K}\lambda_{mk}\alpha_{km'}\Big]b_{m'}(v')b_m(v)$$

$$= \sum_{k=1}^{K}\Big\{\Big[\sum_{m=1}^{M}\lambda_{mk}b_m(v)\Big] \times \Big[\sum_{m'=1}^{M}\alpha_{km'}b_{m'}(v')\Big]\Big\}$$

and

$$\tilde{\epsilon}_i(v) = \sum_{k=1}^{K}\tilde{\lambda}_k(v)\epsilon_{ik} = \sum_{m=1}^{M}\sum_{k=1}^{K}\lambda_{mk}\epsilon_{ik}b_m(v).$$

Web Appendix B: Full Conditional Posterior Distributions

Elements in the lower triangular part of the working loading matrix, denoted $\boldsymbol{\lambda}_m^* = \{\lambda_{m1}^*, \lambda_{m2}^*, \ldots, \lambda_{mq}^*\}^T$ with $q = \min(m, K)$, are sampled from a multivariate normal distribution that

$$\pi(\boldsymbol{\lambda}_m^*|\cdot) \sim \mathrm{N}_q(\mathrm{Mean}, \mathrm{Cov}),$$

$$\mathrm{Mean} = \mathrm{Cov} \times \sigma_\zeta^{-2} \sum_{i=1}^N \theta_{im} \boldsymbol{\eta}_{i_m}^*, \quad \mathrm{Cov} = \Big[\sum_{i=1}^N \big(\sigma_\zeta^{-2} \boldsymbol{\eta}_{i_m}^* \boldsymbol{\eta}_{i_m}^{*T} \big) + \sigma_\lambda^{-2} \boldsymbol{I}_q \Big]^{-1},$$

where $\boldsymbol{\eta}_{i_m}^* = (\eta_{i1}^*, \eta_{i2}^*, \ldots, \eta_{iq}^*)^T$, the corresponding latent vector to $\boldsymbol{\lambda}_m^*$. Other working loading matrix elements in the upper triangular part are set to be 0.

The subject-specific working latent vector $\boldsymbol{\eta}_i^* = (\eta_{ik}^*)_{K \times 1} = (\eta_{i1}, \eta_{i2}, \ldots, \eta_{iK})^T$ is sampled from a posterior multivariate normal distribution that

$$\pi(\boldsymbol{\eta}_i^*|\cdot) \sim \mathrm{N}_K\big(\mathrm{Mean}, \mathrm{Cov}\big),$$

$$\mathrm{Mean} = \mathrm{Cov} \times \Big\{ \sigma_\zeta^{-2} \boldsymbol{\Lambda}^{*T} \boldsymbol{\theta}_i + \sigma_\epsilon^{-2} \boldsymbol{\Phi} \big(\boldsymbol{\mu}_i^* + \boldsymbol{\beta}^{*T} \tilde{\boldsymbol{X}}_i \big) \Big\}, \quad \mathrm{Cov} = \big(\sigma_\zeta^{-2} \boldsymbol{\Lambda}^{*T} \boldsymbol{\Lambda}^* + \sigma_\epsilon^{-2} \boldsymbol{\Phi} \big)^{-1},$$

where $\boldsymbol{\Lambda}^* = (\lambda_{mk}^*)_{M \times K}$, $\boldsymbol{\theta}_i = (\theta_{im})_{M \times 1}$, $\boldsymbol{\Phi} = \mathrm{diag}\{\phi_1^2, \phi_2^2, \ldots, \phi_K^2\}$, $\boldsymbol{\mu}_i^* = (\mu_{ik}^*)_{K \times 1}$, $\boldsymbol{\beta}^* = \{\beta_k^*(v)\}_{|\mathcal{R}| \times K}$, $\tilde{\boldsymbol{X}}_i = \{\tilde{X}_i(v)\}_{|\mathcal{R}| \times 1}$ and let $|\mathcal{R}|$ represent the number of voxels in $\mathcal{R}$. Similarly, the extra K-length intercept vector $\boldsymbol{\mu}_i^*$ is sampled from a multivariate normal distribution that

$$\pi(\boldsymbol{\mu}_i^*|\cdot) \sim \mathrm{N}_K \Big\{ \sigma_\epsilon^{-2} \boldsymbol{\Phi} \big(\sigma_\epsilon^{-2} \boldsymbol{\Phi} + \sigma_\mu^{-2} \boldsymbol{I} \big)^{-1} \big(\boldsymbol{\eta}_i^* - \boldsymbol{\beta}^{*T} \tilde{\boldsymbol{X}}_i \big), \quad \big(\sigma_\epsilon^{-2} \boldsymbol{\Phi} + \sigma_\mu^{-2} \boldsymbol{I}_M \big)^{-1} \Big\}.$$

The working basis coefficient vector $\boldsymbol{\alpha}_k^*$ for approximating $\boldsymbol{\beta}_k^*$ has the following full conditional posterior distribution that

$$\pi(\boldsymbol{\alpha}_k^*|\cdot) \sim \mathrm{N}_M(\mathrm{Mean}, \mathrm{Cov}),$$

$$\mathrm{Mean} = \mathrm{Cov} \times \sigma_\epsilon^{-2} \phi_k^2 \boldsymbol{b}^T \Big(\sum_{i=1}^N \big(\eta_{ik}^* - \mu_{ik}^* \big) \tilde{\boldsymbol{X}}_i \Big), \quad \mathrm{Cov} = \Big[\sigma_\epsilon^{-2} \phi_k^2 \boldsymbol{b}^T \big(\sum_{i=1}^N \tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{X}}_i^T \big) \boldsymbol{b} + \sigma_\alpha^{-2} \boldsymbol{I}_M \Big]^{-1},$$

where $\boldsymbol{\alpha}_k^* = (\alpha_{mk})_{M \times 1}$ and $\boldsymbol{b} = \{b_m(v)\}_{|\mathcal{R}| \times M}$.

The reciprocal of diagonal element ϕ_k^2 in the working matrix $\boldsymbol{\Phi}$ is sampled from a Gamma full conditional posterior distribution:

$$\pi(\phi_k^{-2}|\cdot) \sim G\left(a_\phi + \frac{N}{2}, \quad b_\phi + \frac{1}{2}\sigma_\epsilon^{-2}\sum_{i=1}^N \left(\eta_{ik}^* - \mu_{ik}^* - \tilde{\boldsymbol{X}}_i^T\boldsymbol{\beta}_k^*\right)^2\right).$$

Here we list the full conditional posterior distributions for other parameters:

$$\pi(\boldsymbol{U}|\cdot) \sim \mathrm{N}_{|\mathcal{R}|}\left(\frac{\sigma_e^{-2}}{N\sigma_e^{-2}+\sigma_u^{-2}}\sum_{i=1}^{N}\left(\boldsymbol{Z}_i - \boldsymbol{b}\boldsymbol{\theta}_i\right),\ \left(N\sigma_e^{-2}+\sigma_u^{-2}\right)^{-1}\boldsymbol{I}_{|\mathcal{R}|}\right),$$

$$\pi(\sigma_u^{-2}|\cdot) \sim G\left(a_u + \frac{1}{2},\ b_u + \frac{1}{2}\boldsymbol{U}^T\boldsymbol{U}\right),$$

$$\pi(\boldsymbol{\theta}_i|\cdot) \sim \mathrm{N}_M\left(\left[\sigma_e^{-2}\boldsymbol{b}^T\boldsymbol{b}+\sigma_\zeta^{-2}\boldsymbol{I}_M\right]^{-1}\left(\sigma_e^{-2}\boldsymbol{b}^T(\boldsymbol{Z}_i-\boldsymbol{U})+\sigma_\zeta^{-2}\boldsymbol{\Lambda}^*\boldsymbol{\eta}_i^*\right),\right.$$

$$\left.\left[\sigma_e^{-2}\boldsymbol{b}^T\boldsymbol{b}+\sigma_\zeta^{-2}\boldsymbol{I}_M\right]^{-1}\right),$$

$$\pi(\sigma_e^{-2}|\cdot) \sim G\left(a_e + \frac{QN}{2},\quad b_e + \frac{1}{2}\sum_{i=1}^{N}\left(\boldsymbol{Z}_i - \boldsymbol{U} - \boldsymbol{b}\boldsymbol{\theta}_i\right)^T\left(\boldsymbol{Z}_i - \boldsymbol{U} - \boldsymbol{b}\boldsymbol{\theta}_i\right)\right),$$

$$\pi(\sigma_\zeta^{-2}|\cdot) \sim G\left(a_\zeta + \frac{MN}{2},\quad b_\zeta + \frac{1}{2}\sum_{i=1}^{N}\left(\boldsymbol{\theta}_i - \boldsymbol{\Lambda}^*\boldsymbol{\eta}_i^*\right)^T\left(\boldsymbol{\theta}_i - \boldsymbol{\Lambda}^*\boldsymbol{\eta}_i^*\right)\right),$$

$$\pi(\sigma_\epsilon^{-2}|\cdot) \sim G\left(a_\epsilon + \frac{KN}{2},\quad b_\epsilon + \frac{1}{2}\sum_{i=1}^{N}\left[\boldsymbol{\eta}_i^* - \boldsymbol{\mu}_i^* - \boldsymbol{\beta}^{*T}\tilde{\boldsymbol{X}}_i\right]^T\boldsymbol{\Phi}\left[\boldsymbol{\eta}_i^* - \boldsymbol{\mu}_i^* - \boldsymbol{\beta}^{*T}\tilde{\boldsymbol{X}}_i\right]\right),$$

$$\pi(\gamma_p|\cdot) \sim \mathrm{Bernoulli}\left(\frac{c_1}{c_0+c_1}\right),$$

$$c_1 = \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{K}\sigma_\epsilon^{-2}\phi_k^2\left(\tau_{ik,-p} - \boldsymbol{\beta}_k^{*T}\boldsymbol{X}_{ip}\right)^2\right\} \times w,$$

$$c_0 = \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{K}\sigma_\epsilon^{-2}\phi_k^2\tau_{ik,-p}^2\right\} \times (1-w),$$

$$\tau_{ik,-p} = \eta_{ik}^* - \mu_{ik}^* - \boldsymbol{\beta}_k^{*T}\sum_{h=1,h\neq p}^{P}\gamma_h\boldsymbol{X}_{ih},$$

$$\pi(w|\cdot) \sim \mathrm{Beta}\left(a_w + \sum_{p=1}^{P}\gamma_p,\ b_w + P - \sum_{p=1}^{P}\gamma_p\right),$$

where $\boldsymbol{U} = \{U(v)\}_{|\mathcal{R}|\times 1}$ and $\boldsymbol{Z}_i = \{Z_i(v)\}_{|\mathcal{R}|\times 1}$ and $\boldsymbol{\beta}_k^* = \{\beta_k^*(v)\}_{|\mathcal{R}|\times 1}$ and $\boldsymbol{X}_{ip} = \{X_{ip}(v)\}_{|\mathcal{R}|\times 1}$ and $\boldsymbol{\gamma} = (\gamma_p)_{P\times 1}$.

The time complexity of the Gibbs sampler for each updating scheme within the MCMC algorithm.

- Update $\boldsymbol{U}$: $O(NM|\mathcal{R}|)$.

- Update σ_u^{-2}: $O(N|\mathcal{R}|)$.

- Update $\boldsymbol{\theta}_i$ for $i = 1, \ldots N$: $O(NM^2|\mathcal{R}|)$.

- Update σ_e^{-2}: $O(NM|\mathcal{R}|)$.

- Update σ_ζ^{-2}: $O(NMK)$.

- Update σ_ϵ^{-2}: $O(NK|\mathcal{R}|)$.

- Update γ_p, for $p = 1, \ldots, P$: $O(NKP|\mathcal{R}|)$.

- Update ω: $O(P)$.

This implies that the total time complexity within each iteration of the MCMC algorithm is $O(NM^2|\mathcal{R}|) + O(NKP|\mathcal{R}|)$.

Web Appendix C: Estimations and Predictions from Gibbs Sampler

Let $\hat{\boldsymbol{\Theta}}^{(t)}$ be the set of posterior samples of all parameters in every iteration t given the dataset $(\boldsymbol{Z}, \boldsymbol{X})$, that $\hat{\boldsymbol{\Theta}}^{(t)} = (\hat{\boldsymbol{\alpha}}^{(t)}, \hat{\boldsymbol{\beta}}^{(t)}, \hat{\boldsymbol{\gamma}}^{(t)}, \hat{\boldsymbol{\eta}}^{(t)}, \hat{\boldsymbol{\Lambda}}^{(t)}, \hat{\boldsymbol{\theta}}^{(t)}, \hat{\boldsymbol{U}}^{(t)}, \hat{\sigma}_\epsilon^{(t)}, \hat{\sigma}_\zeta^{(t)}, \hat{\sigma}_e^{(t)})$.

In every iteration t, the posterior estimations of outcome image at voxel v for subject i in training set that $i \in \mathcal{I}_i^{\mathrm{tr}}$ is defined as

$$[\hat{Z}_i(v)]^{(t)} = \hat{U}^{(t)}(v) + \sum_{m=1}^{M} \hat{\theta}_{im}^{(t)} b_m(v) + \hat{e}_i^{(t)}(v), \quad i \in \mathcal{I}_i^{\mathrm{tr}}$$

where

$$\hat{\theta}_{im}^{(t)} = \sum_{k=1}^{K} \hat{\lambda}_{mk}^{(t)} \hat{\eta}_{ik}^{(t)} + \hat{\zeta}_{im}^{(t)}$$

$$\hat{\eta}_{ik}^{(t)} = \sum_{v \in \mathcal{R}} \left[\sum_{p=1}^{P} \hat{\gamma}_p^{(t)} X_{ip}(v) \right] \hat{\beta}_k^{(t)}(v) + \hat{\epsilon}_{ik}^{(t)}$$

$$\hat{e}_i^{(t)}(v) \sim N(0, \hat{\sigma}_e^{(t)}), \quad \hat{\zeta}_{im}^{(t)}(v) \sim N(0, \hat{\sigma}_\zeta^{(t)}), \quad \hat{\epsilon}_i^{(t)}(v) \sim N(0, \hat{\sigma}_\epsilon^{(t)}).$$

For a total of T iterations, the posterior mean estimation of outcome image at voxel v for subject $i \in \mathcal{I}_i^{\mathrm{tr}}$ from the training set is

$$\hat{Z}_i(v) = \sum_{t=T/2+1}^{T} [\hat{Z}_i(v)]^{(t)}, \quad i \in \mathcal{I}_i^{\mathrm{tr}}.$$

Given the new predictor images $\{\boldsymbol{X}_{j1}, \boldsymbol{X}_{j2}, \ldots, \boldsymbol{X}_{jP}\}$ $(\boldsymbol{X}_{jp} = \{X_{jp}(v)\}_{|\mathcal{R}| \times 1})$ of subject $j \in \mathcal{I}_j^{\mathrm{ts}}$ in the test set, the prediction of the corresponding outcome image $\boldsymbol{Z}_j = \{Z_j(v)\}_{|\mathcal{R}| \times 1}$ at voxel v in each iteration t is defined as

$$[\hat{Z}_j(v)]^{(t)} = \hat{U}^{(t)}(v) + \sum_{m=1}^{M} \hat{\theta}_{jm}^{(t)} b_m(v) + \hat{e}_j(v)^{(t)}, \quad j \in \mathcal{I}_j^{\mathrm{ts}}.$$

where

$$\hat{\theta}_{jm}^{(t)} = \sum_{k=1}^{K} \hat{\lambda}_{mk}^{(t)} \hat{\eta}_{jk}^{(t)} + \hat{\zeta}_{jm}^{(t)}$$

$$\hat{\eta}_{jk}^{(t)} = \sum_{v \in \mathcal{R}} \Big[\sum_{p=1}^{P} \hat{\gamma}_{p}^{(t)} X_{jp}(v) \Big] \hat{\beta}_{k}^{(t)}(v) + \hat{\epsilon}_{jk}^{(t)}$$

$$\hat{e}_{j}^{(t)}(v) \sim N(0, \hat{\sigma}_{e}^{(t)}), \;\; \hat{\zeta}_{jm}^{(t)}(v) \sim N(0, \hat{\sigma}_{\zeta}^{(t)}), \;\; \hat{\epsilon}_{j}^{(t)}(v) \sim N(0, \hat{\sigma}_{\epsilon}^{(t)}).$$

In particular, $\hat{U}^{(t)}(v)$, $\hat{\lambda}_{mk}^{(t)}$, $\gamma_{p}^{(t)}$, $\hat{\beta}_{k}^{(t)}(v)$, $\hat{\sigma}_{e}^{(t)}$, $\hat{\sigma}_{\zeta}^{(t)}$ and $\hat{\sigma}_{\epsilon}^{(t)}$ are posterior samples from training set in iteration t. Therefore, the final prediction of outcome image at voxel v for a subject j in test test $(j \in \mathcal{I}_{j}^{\text{ts}})$ is calculated as

$$\hat{Z}_{j}(v) = \sum_{t=T/2+1}^{T} [\hat{Z}_{j}(v)]^{(t)}.$$

Web Appendix D: Graphical Representation for the Simulation Study Generating Models

Web Figure. 1: Graphical representation of the generating models used in simulation study scenario 3. Dashed squares contain working parameters used in parameter expansion method. Shaded squares represent the transformation mechanism from working to original inferential parameters. The true number of latent factors K is set to be 5.
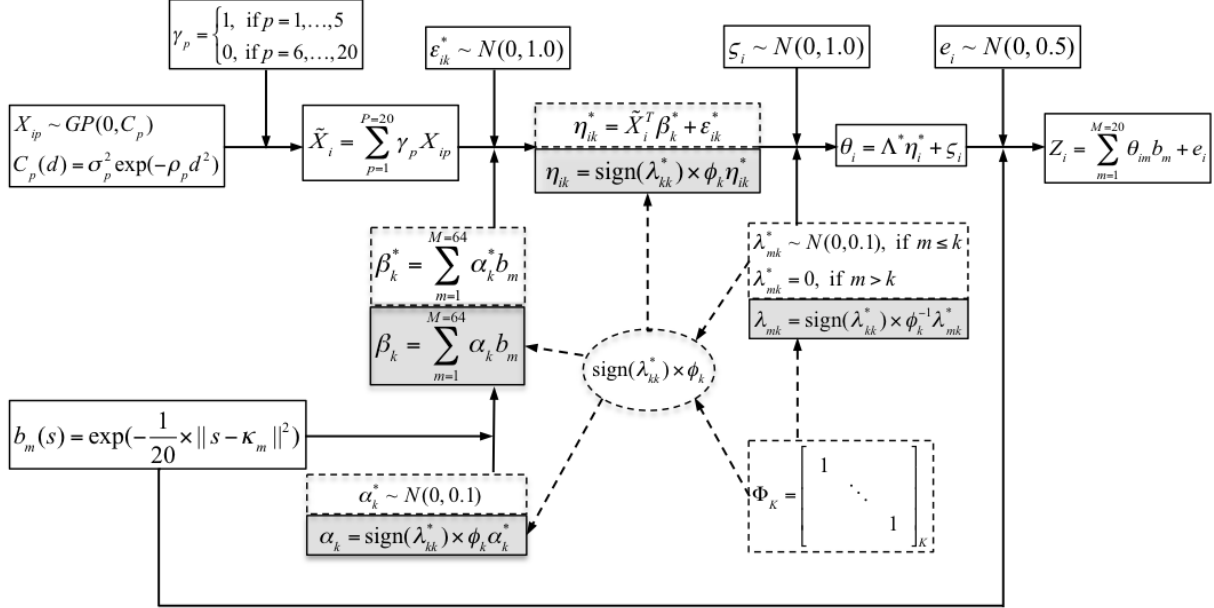
# Web Appendix E: Additional Results for the Analysis of HCP Data

Web Table. 1: Selected feature images in application analysis based on posterior estimations of γ and varied thresholds for the left (1) and right (2) amygdala regions, respectively.

(1) Left amygdala region

Threshold	NO. of Features	Selected Feature (Index)
0.0	10	1, 2, 6, 13, 15, 21, 22, 28, 29, 31
0.1	8	1, 2, 13, 15, 21, 22, 28, 29
0.3	7	1, 2, 13, 15, 21, 22, 28
0.4	5	1, 2, 13, 15, 28
0.5	4	**2**, **13**, **15**, **28**
0.8	2	**13**, **15**
0.9	0	None

(2) Right amygdala region

| Threshold	NO. of Features	Selected Feature (Index)
0.0	10	1, 2, 4, 13, 15, 21, 22, 28, 29, 31
0.1	9	1, 2, 4, 13, 15, 21, 22, 28, 31
0.2	6	1, 13, 15, 21, 22, 28
0.3	5	1, 13, 21, 22, 28
0.4	4	1, 21, 22, 28
0.5	2	**21**, **28**
0.6	1	**28**
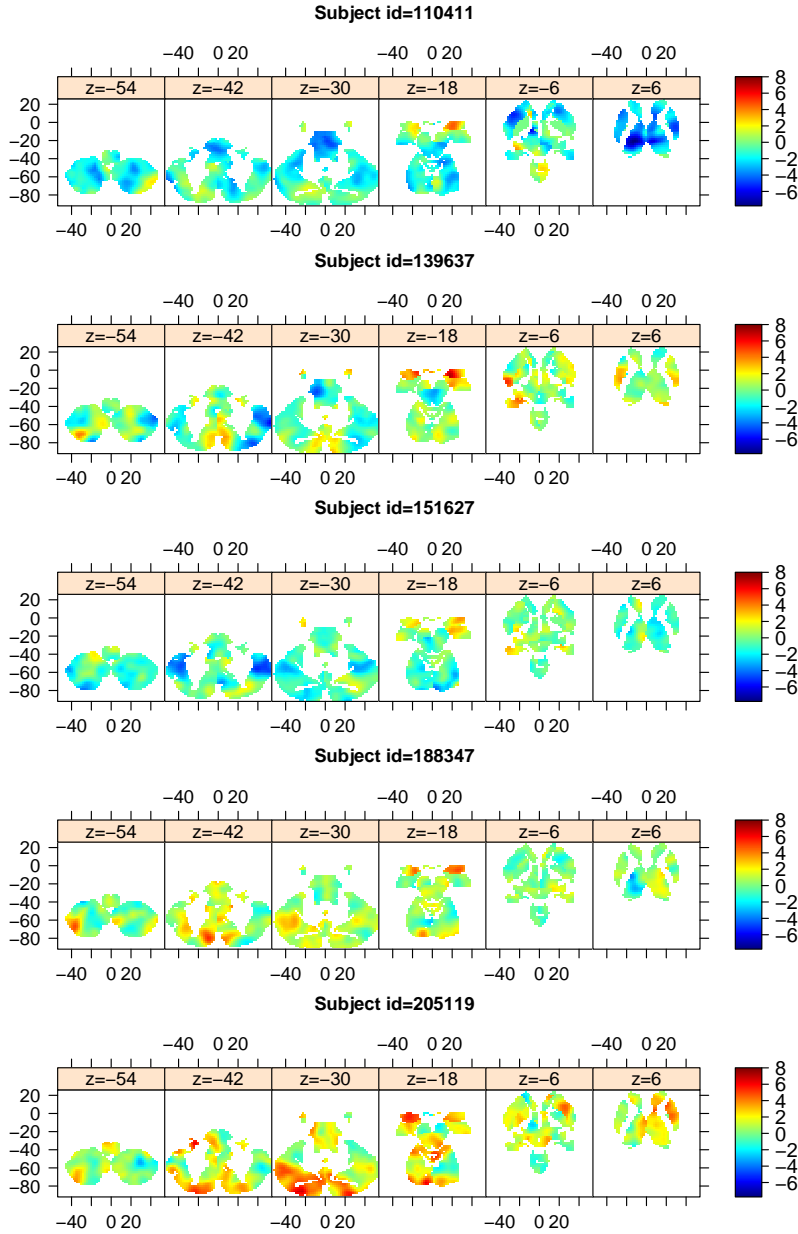
Web Figure. 2: Examples of outcome images (whole-brain faces-shapes contrast maps in EMOTION domain) from 5 subjects. Maps are shown at six different axial slices. All maps are plotted on the same color scale.

12
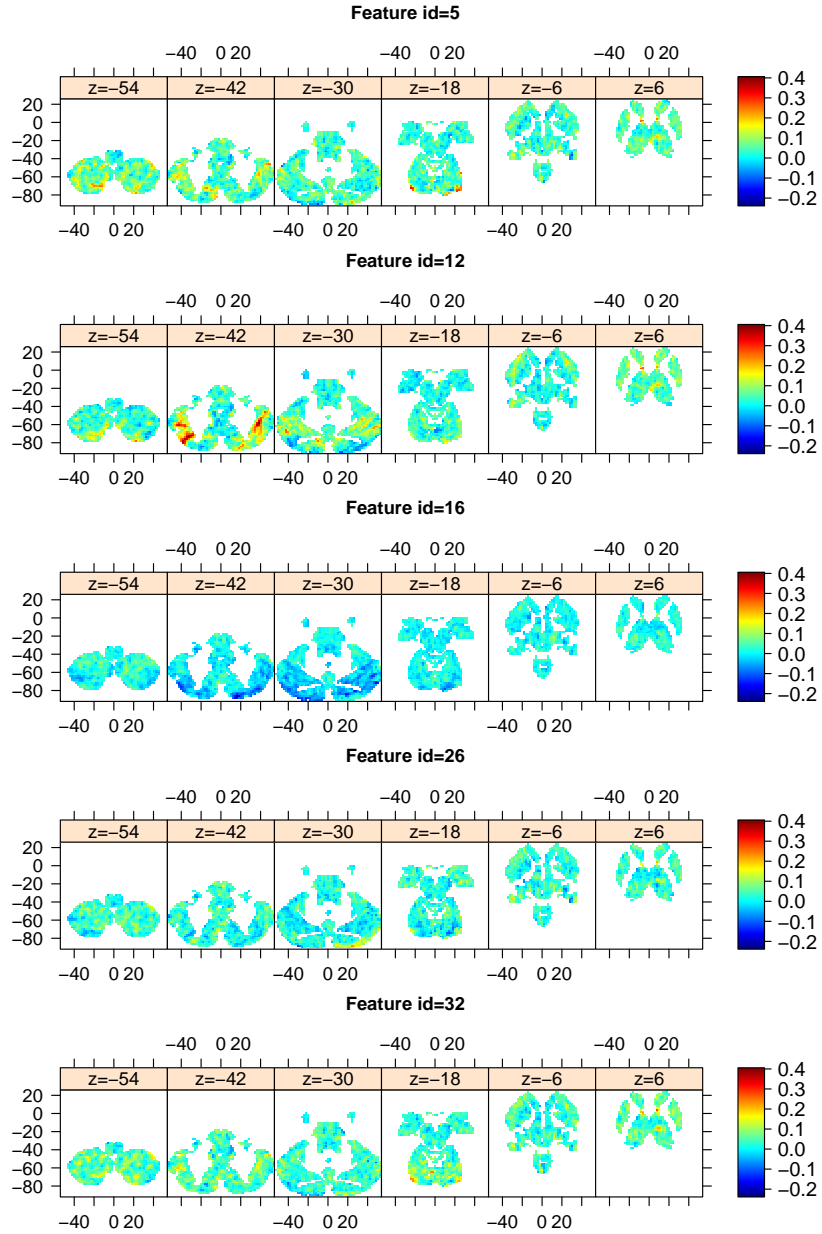
Web Figure. 3: Examples of six predictor maps (whole-brain sub-cortical seed maps) from a single subject (id = 110411), shown at six different axial slices. All maps are plotted on the same color scale.

13
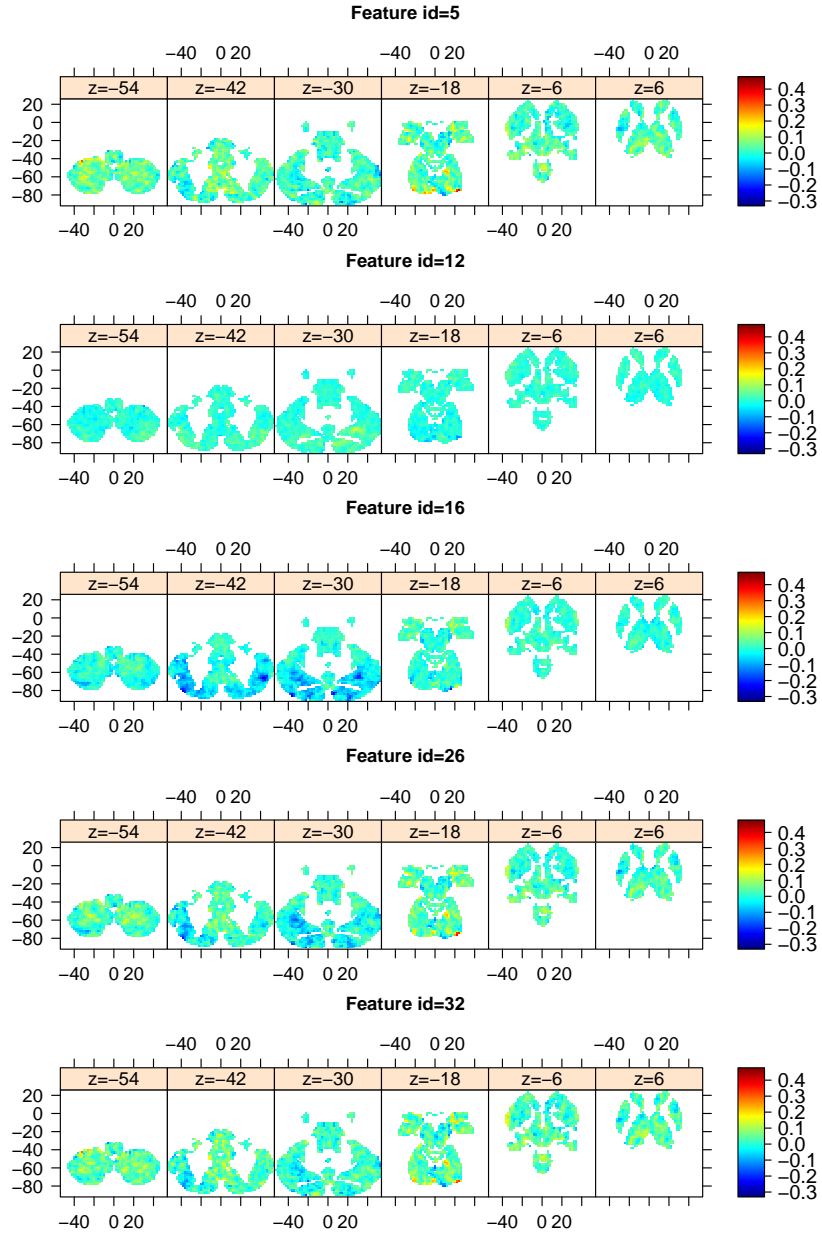
Web Figure. 4: Examples of six feature whole-brain maps (sub-cortical seed maps) from a single subject (id = 139637), shown at six different axial slices. All maps are plotted on the same color scale.

14
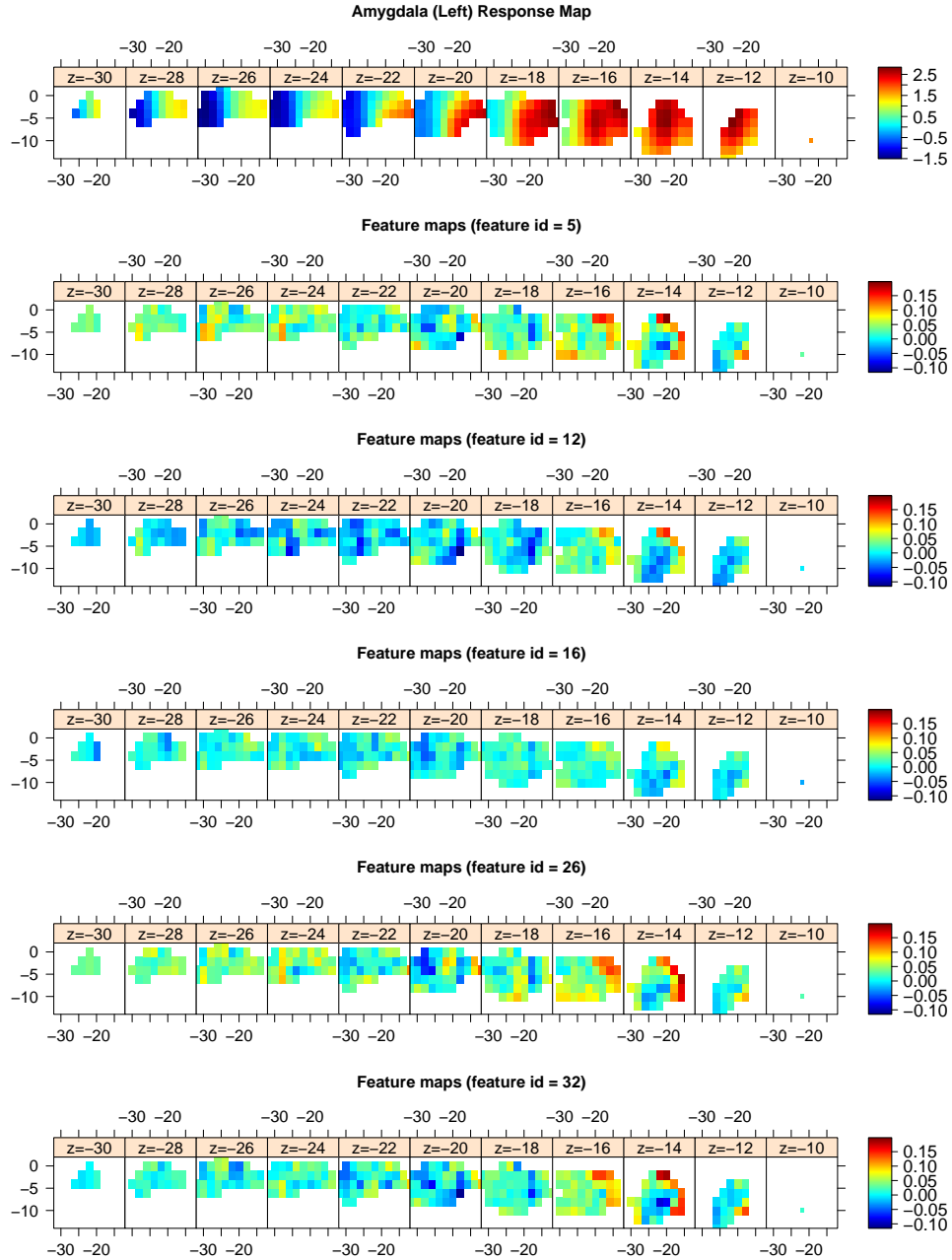
Web Figure. 5: Example images of outcome (faces-shapes contrast maps in EMOTION domain) and predictor (five sub-cortial seed maps) within the left amygdala region from a single subject (id = 110411), shown at axial slices.
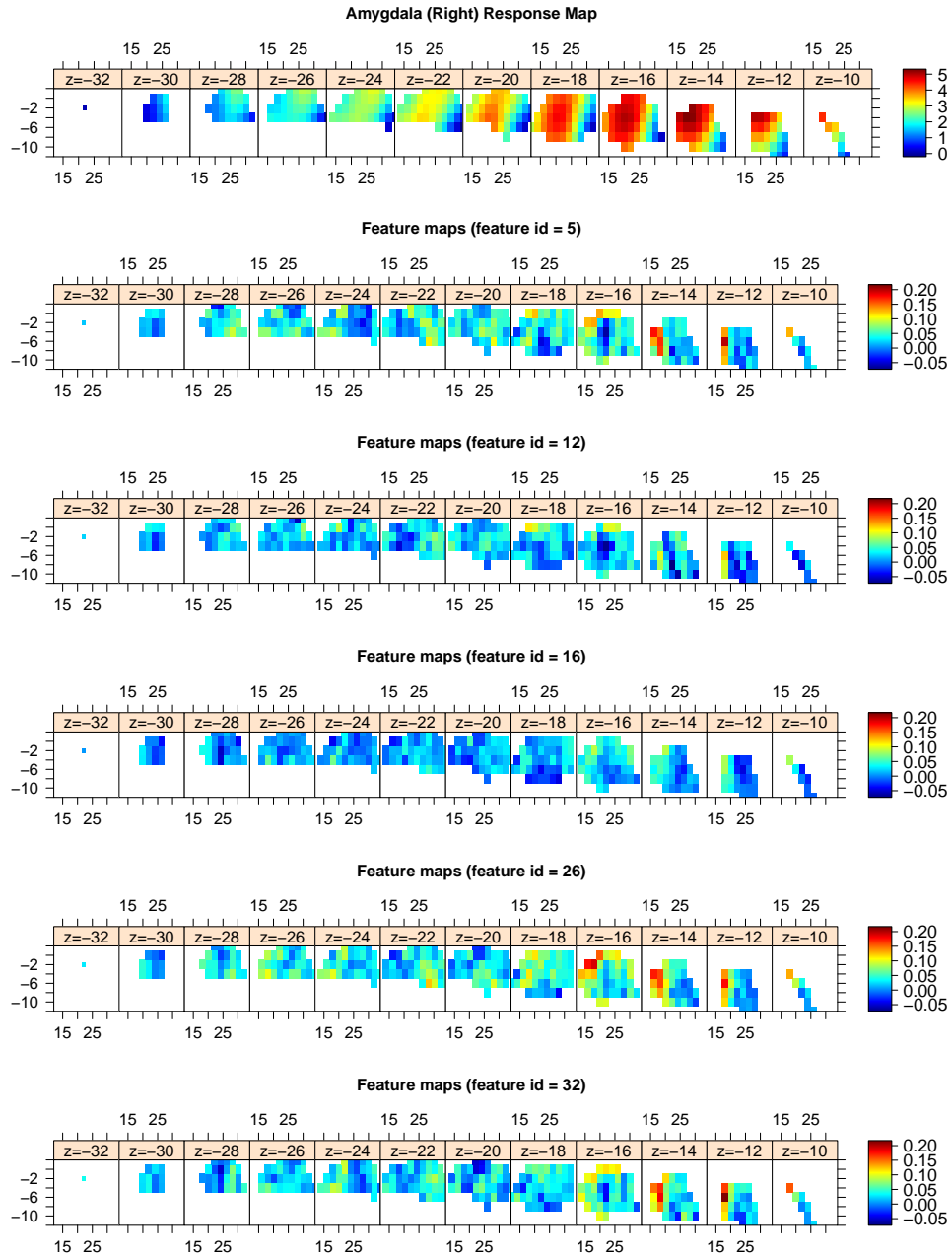
Web Figure. 6: Example images of outcome (faces-shapes contrast maps in EMOTION domain) and predictor (five sub-cortial seed maps) within the right amygdala region from a single subject (id = 110411), shown at axial slices.

16
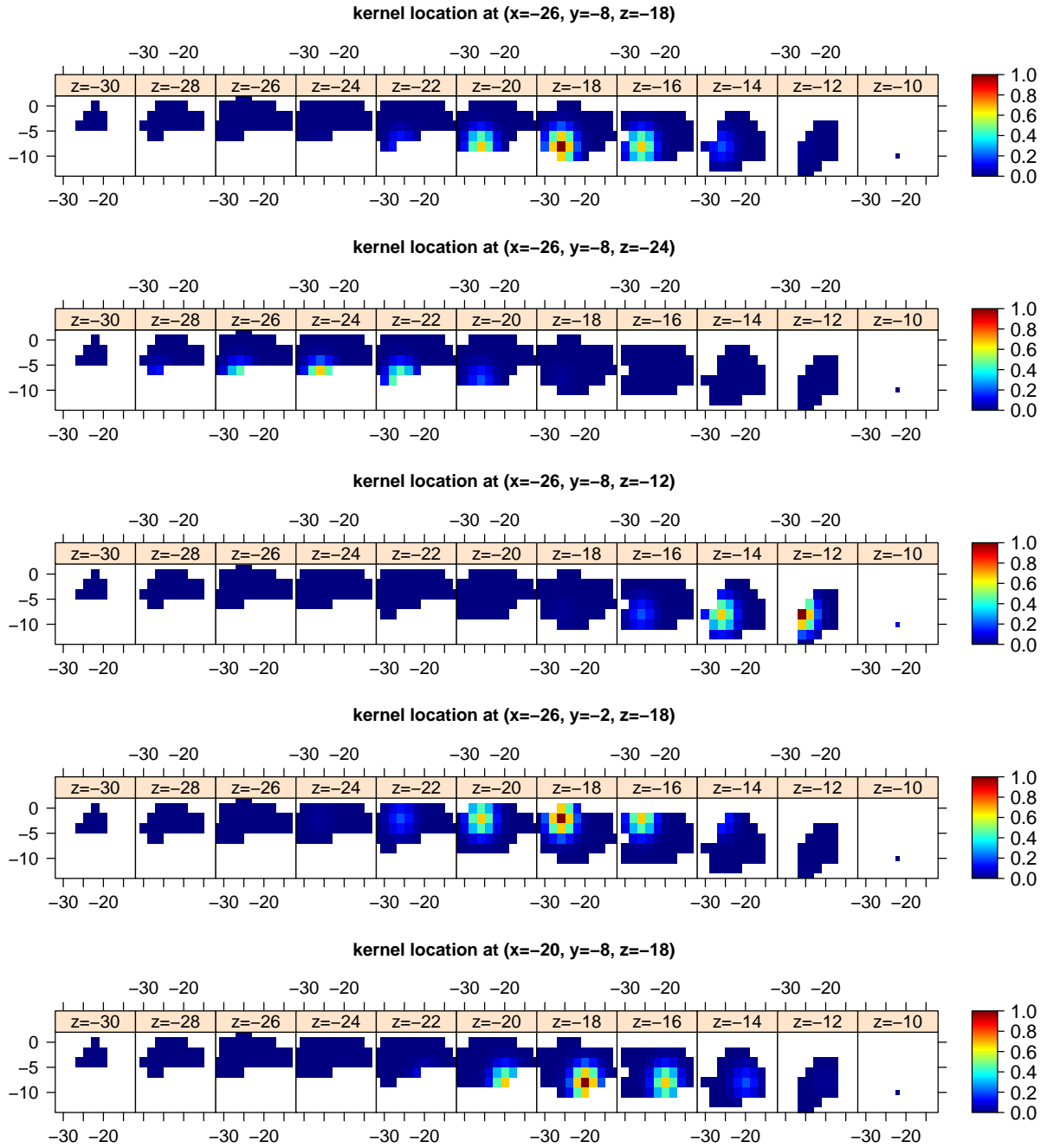
Web Figure. 7: Maps of basis functions with bandwidth value $b = 1/10$ and five different kernel locations for application analysis within the left amygdala region, shown at axial slices.

17
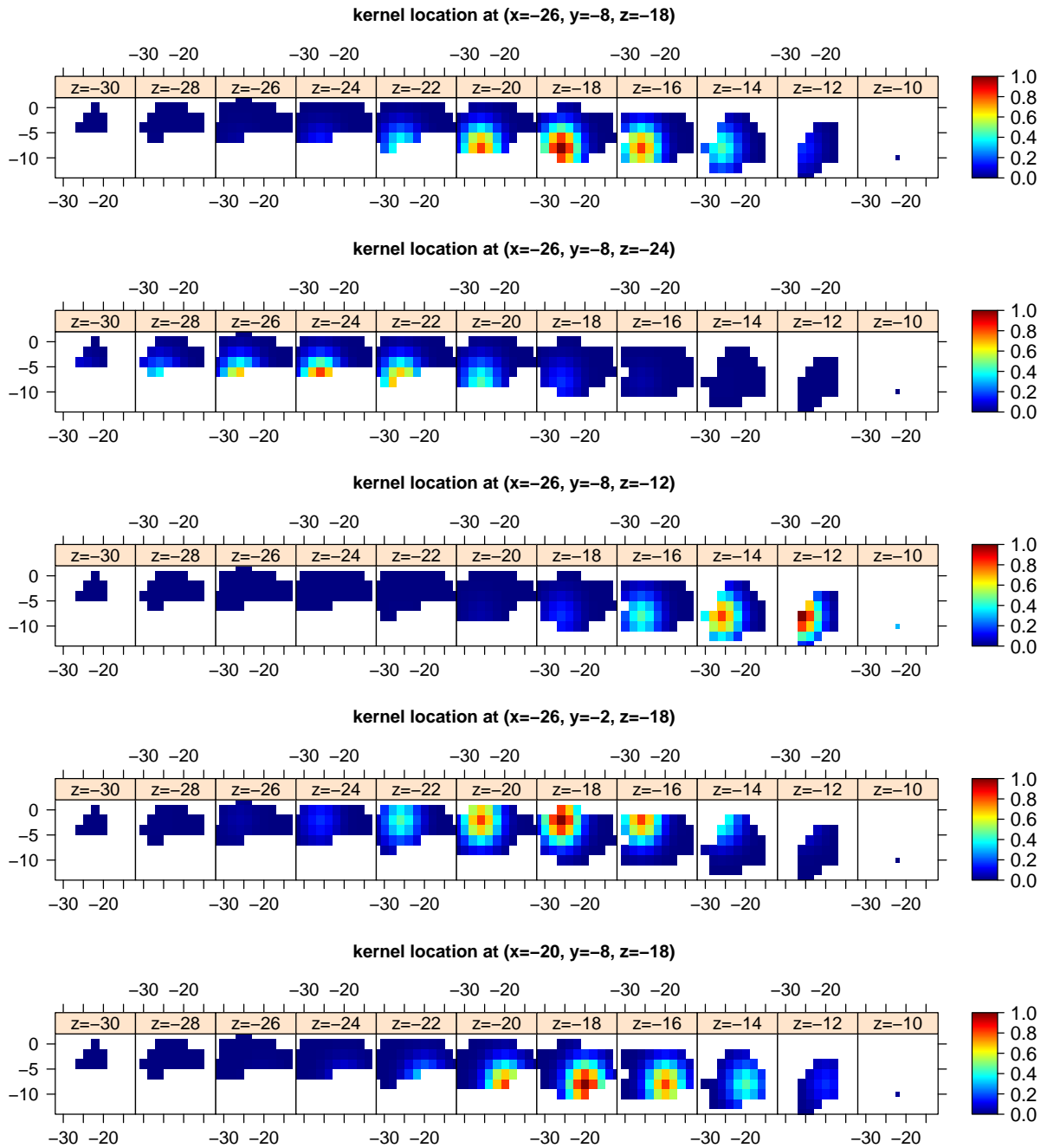
Web Figure. 8: Maps of basis functions with bandwidth value $b = 1/20$ and five different kernel locations for application analysis within the left amygdala region, shown at axial slices.

18
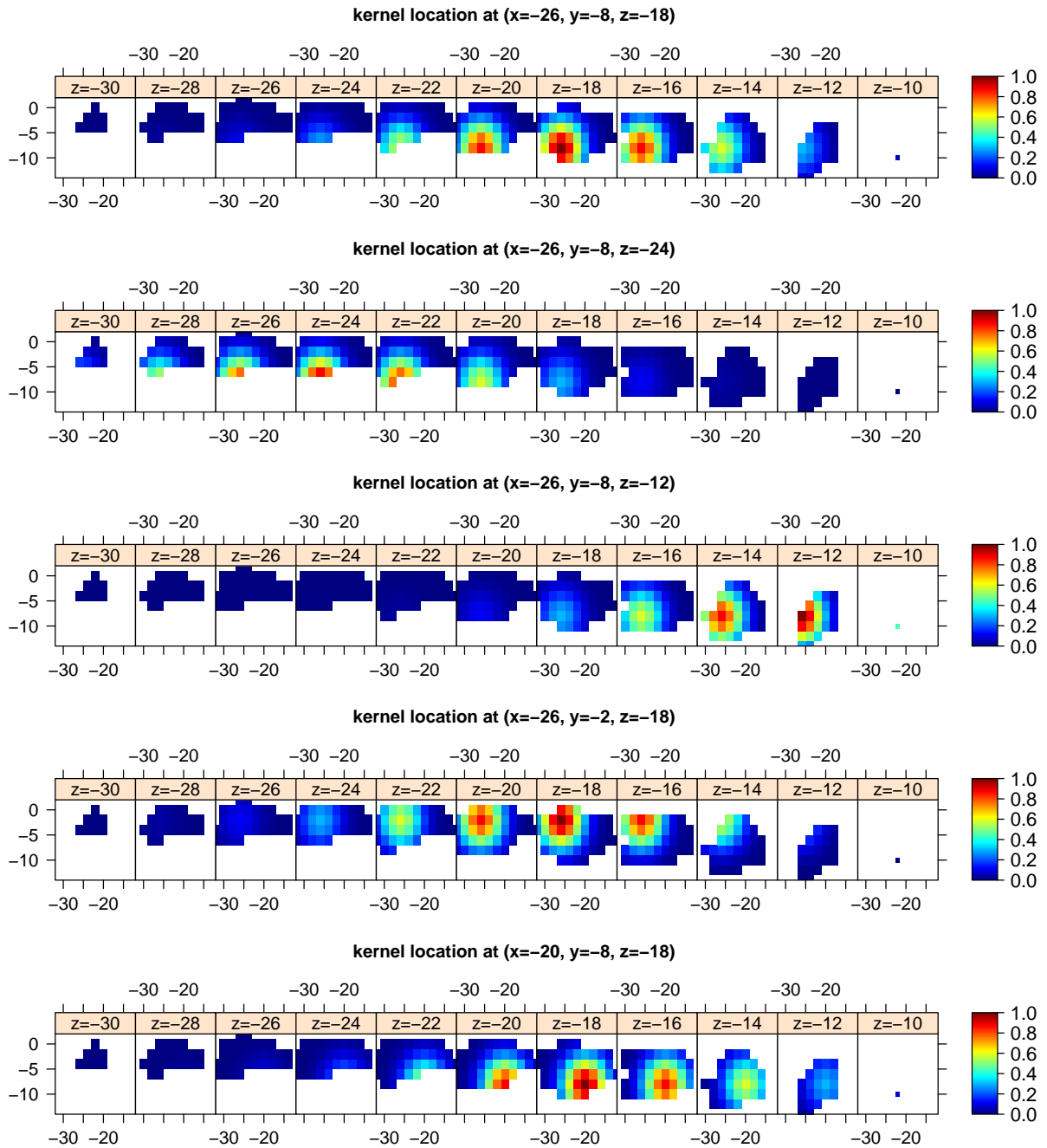
Web Figure. 9: Maps of basis functions with bandwidth value $b = 1/30$ and five different kernel locations for application analysis within the left amygdala region, shown at axial slices.

19

Web Table. 2: Results of independent parcels analysis in Application study using SBLF and linear regression models.

Parcel Name	NO. of Voxels	Optimal K	Mean Squared Error		R-Squared	
			SBLF	Linear	SBLF	Linear
ACCUMBENS LEFT	135	8	0.031	0.151	0.941	0.759
ACCUMBENS RIGHT	140	6	0.037	0.194	0.926	0.692
AMYGDALA LEFT	315	9	0.063	0.562	0.932	0.600
AMYGDALA RIGHT	332	10	0.069	0.651	0.935	0.621
BRAIN STEM	3472	2	0.107	1.349	0.921	0.068
CAUDATE LEFT	728	5	0.055	0.660	0.949	0.405
CAUDATE RIGHT	766	8	0.061	0.691	0.946	0.412
DIENCEPHALON VENTRAL LEFT	706	5	0.085	1.169	0.935	0.257
DIENCEPHALON VENTRAL RIGHT	712	10	0.088	1.117	0.930	0.246
HIPPOCAMPUS LEFT	764	5	0.060	0.818	0.949	0.380
HIPPOCAMPUS RIGHT	795	5	0.063	0.828	0.948	0.371
PALLIDUM LEFT	297	5	0.061	0.370	0.917	0.563
PALLIDUM RIGHT	260	6	0.062	0.389	0.913	0.495
PUTAMEN LEFT	1060	6	0.071	0.776	0.943	0.376
PUTAMEN RIGHT	1010	7	0.066	0.730	0.940	0.363
THALAMUS LEFT	1288	4	0.076	0.829	0.931	0.324
THALAMUS RIGHT	1248	6	0.078	0.764	0.927	0.346
Averaged Values Across Parcels			0.067	0.709	0.934	0.428

SBLF: our proposed spatial Bayesian latent factor model.

Web Appendix F: Sensitivity Analysis

The posterior inferences and prediction accuracy results are generally not very sensitive to mild changes of different hyperparameters in our model. To demonstrate this point, we perform a sensitivity analysis for hyperparameters σ_α^2, σ_μ^2, σ_λ^2, σ_e^{-2}, σ_ζ^{-2} and σ_ϵ^{-2} based on a small sample of the HCP data (50 training samples and 10 testing samples of images in the left amygdala region).

The three hyperparameters, σ_α^2, σ_μ^2 and σ_λ^2, were chosen to fix at 1.0 in the data application. For each of these three parameters, we consider three alternative values 2.0, 0.5 and 0.2. In current data application, the prior distributions placed on σ_e^{-2}, σ_ζ^{-2} and σ_ϵ^{-2} are all Gamma$(1, 1)$. We consider two additional settings Gamma$(1.0, 10.0)$ and Gamma$(10.0, 1.0)$.

For each setting of the hyperparameters, we run MCMC for 20,000 iterations with 10,000 burn-in. The number of latent factors is fixed as five. We report the MSE and MSPE as the evaluation criteria for the model fitting and prediction accuracy, respectively.

Web Tables 3 and 4 indicate that the estimated MSE almost did not change accross different settings of the hyperparameters, while the MSPE shows a slight but negligible difference. Therefore, we conclude that the results are not sensitive to the choices of hyperparameters.

Web Table. 3: MSE and MSPE (mean and sd) for different fixed values of the hyperparameters σ_α^2, σ_μ^2 and σ_λ^2. The values specified in the data application are in bold.

σ_α^2	MSE	MSPE
2.0	0.053 (0.094)	1.246 (1.693)
1.0	**0.053 (0.094)**	**1.229 (1.614)**
0.5	0.052 (0.093)	1.202 (1.594)
0.2	0.052 (0.093)	1.177 (1.553)

σ_u^2	MSE	MSPE
2.0	0.053 (0.094)	1.230 (1.627)
1.0	**0.053 (0.094)**	**1.229 (1.614)**
0.5	0.053 (0.094)	1.233 (1.623)
0.2	0.053 (0.094)	1.228 (1.616)

σ_λ^2	MSE	MSPE
2.0	0.053 (0.094)	1.254 (1.709)
1.0	**0.053 (0.094)**	**1.229 (1.614)**
0.5	0.052 (0.093)	1.194 (1.573)
0.2	0.052 (0.093)	1.177 (1.553)

Web Table. 4: MSE and MSPE (mean and sd) for different hyperprior specifications (a gamma prior) of σ_e^{-2}, σ_ζ^{-2} and σ_ϵ^{-2}. The values specified in the data application are in bold.

σ_e^{-2}	shape	rate	MSE	MSPE
	1.0	10.0	0.053 (0.094)	1.228 (1.616)
	1.0	**1.0**	**0.053 (0.094)**	**1.229 (1.614)**
	10.0	1.0	0.053 (0.094)	1.230 (1.601)

σ_ζ^{-2}	shape	rate	MSE	MSPE
	1.0	10.0	0.053 (0.093)	1.231 (1.615)
	1.0	**1.0**	**0.053 (0.094)**	**1.229 (1.614)**
	10.0	1.0	0.053 (0.094)	1.228 (1.618)

σ_ϵ^{-2}	shape	rate	MSE	MSPE
	1.0	10.0	0.053 (0.093)	1.239 (1.601)
	1.0	**1.0**	**0.053 (0.094)**	**1.229 (1.614)**
	1.0	0.1	0.053 (0.095)	1.210 (1.623)