Animal	Time-Invariant	Non-Clustering	Clustering
1	0.0111	0.0502	0.0562
2	0.0739	0.1664	0.1589
3	0.0055	0.2251	0.2152
4	0.0605	0.2260	0.1942
7	0.0873	0.1777	0.1724
8	0.0638	0.1781	0.1599
9	0.0733	0.2256	0.1939
10	0.0272	0.0795	0.1359
11	0.0283	0.0797	0.0866
12	0.0358	0.0796	0.0881

Table 3: Hold-out RFE predicting the Thalamus region with the novel environment dataset

Animal	Time-Invariant	Non-Clustering	Clustering
1	0.0128	0.0518	0.0597
2	0.1116	0.2110	0.2047
3	0.0132	0.2424	0.2265
5	0.0822	0.1576	0.1689
6	0.0987	0.1669	0.1651
7	0.1065	0.2014	0.1966
8	0.0904	0.2163	0.2016
9	0.0903	0.2513	0.2211
10	0.0355	0.0888	0.1467
11	0.0391	0.0912	0.1000
12	0.0313	0.0724	0.1177

Table 4: Hold-out RFE predicting the Thalamus region with the novel environment dataset

A Supplemental VB Updates and Equations

The VB update for $q(\mathbf{A}_k^\star)$ is dependent only on $\boldsymbol{x}_b^{-k}$ and $\boldsymbol{y}_b^{-k}$, $q(\alpha)$, and $q(\gamma)$, and the distribution over $\mathbf{A}_k^\star$ is judged to be $q(\mathbf{A}_k) = \mathcal{N}(\text{vec}(\mathbf{A}_k); \text{vec}(\hat{\boldsymbol{a}}_{k1}, \ldots, \hat{\boldsymbol{a}}_{kB}), \boldsymbol{\Lambda}_k)$.

First, let $\boldsymbol{\Lambda}_k = \boldsymbol{\Sigma}^{-1}$ and let $\boldsymbol{\Lambda}_0$ the the prior precision induced by the autoregressive prior, which is a block tri-diagonal matrix. Let $\boldsymbol{\Lambda}_{kbb'}$ be the block in $\boldsymbol{\Lambda}_k$ that corresponds to bins b and b'. Each $\boldsymbol{\Lambda}_{kbb}$ is updated by $\boldsymbol{\Lambda}_{kbb} = \Lambda_{0bb} + \hat{\tau}\mathbf{R}_{kbb}$. The mean of the variational distribution can be acquired by either solving $\boldsymbol{\Lambda}_k \text{vec}(\hat{\boldsymbol{a}}_{k1}, \ldots, \hat{\boldsymbol{a}}_{kB}) = \text{vec}(\boldsymbol{v}_{k1}^{-k}, \ldots, \boldsymbol{v}_{kB}^{-k})$, or using the forward filtering backwards smoothing algorithm [15], both of which can be solved in $\mathbb{O}(L^3 B)$ because of the structure of the precision matrix.

The necessary calculations for the novel terms in Eq. 7 are shown below. The hyper parameter α is approximated as a point estimate, and previous experiments have shown that point estimates low-dimensional hyperparameters don't have a significant impact on the variational lower bound [24].

$$\begin{aligned}\mathbb{E}_q[\log p(A_k|\alpha,\gamma) - \log q(A_k)] &= \text{const} + .5\log|\boldsymbol{\Lambda}_k| + \frac{\hat{\gamma}}{2(1-\alpha^2)}||\boldsymbol{a}_{k1}||_2^2 \\ &- \frac{\hat{\gamma}}{2}\sum_{b=2}^{B}(||\hat{\boldsymbol{a}}_{kb} - \hat{\alpha}\hat{\boldsymbol{a}}_{k,b-1}||_2^2 + \hat{\alpha}^2\text{cov}(\boldsymbol{a}_{k,b-1}, \boldsymbol{a}_{k,b}))\end{aligned} \tag{8}$$

Because the matrix $\boldsymbol{\Lambda}_k$ is a block diagonal matrix, the determinant can be calculated in $\mathcal{O}(L^3 B)$.

For Eq. 7, we also need to calculate:

$$\mathbb{E}_q[p(\boldsymbol{x}|\mathbf{A}_{1,\ldots,K}, \boldsymbol{\zeta}, \tau)] = \text{const} - \frac{\hat{\tau}}{2}\sum_b \left(||\boldsymbol{x}_b - \sum_k \boldsymbol{y}'_{kb} * \hat{\boldsymbol{a}}_{kb}||_2^2 + \sum_k T_b \text{trace}(\mathbf{R}_{kb}\text{cov}(\boldsymbol{a}_{kb}))\right) \tag{9}$$

The rest of the equations are standard and can be found in [12, 15].

Animal	Time-Invariant	Non-Clustering	Clustering
1	0.1016	0.1518	0.1612
2	0.1300	0.2257	0.2182
3	0.0327	0.3429	0.3863
4	0.1286	0.3470	0.3184
5	0.2146	0.2880	0.2965
6	0.1260	0.1904	0.1949
7	0.1204	0.2205	0.2143
8	0.2362	0.3765	0.3618
9	0.0861	0.3765	0.3444
10	0.1605	0.1864	0.2556
11	0.0325	0.0774	0.0869
12	0.0715	0.1111	0.1229

Table 5: Hold-out RFE predicting the Thalamus region with the novel environment dataset

Animal	Time-Invariant	Non-Clustering	Clustering
1	0.0991	0.1489	0.1583
2	0.1012	0.1918	0.1848
3	0.0624	0.2598	0.2564
4	0.1061	0.3086	0.2825
5	0.1081	0.1763	0.1811
6	0.1048	0.1683	0.1690
7	0.1158	0.2143	0.2088
8	0.1410	0.2569	0.2427
9	0.0594	0.2723	0.2460
10	0.0962	0.1323	0.1891
11	0.0333	0.0803	0.0884
12	0.0568	0.0960	0.1042

Table 6: Hold-out RFE predicting the Thalamus region with the novel environment dataset

B INLA for the non-clustering dynamic model

To fit the non-clustering dynamic model in the INLA [16] framework, the full model likelihood is rewritten with $z_j = \delta_j$, and $\zeta_j = 1$ (so that $\mathbf{A}_j = \mathbf{A}_j^*$). This likelihood can be written as:

$$p(\boldsymbol{x}, \Theta) \quad = \quad \prod_{b=1}^{B} \left[p(\boldsymbol{x}_b|\Theta)\right] \prod_{j=1}^{J} \left[p(\mathbf{A}_j^*|\alpha, \gamma)\right] p(\alpha, \gamma) \tag{10}$$

After reformulation, this observation likelihood can be combined with the prior to get the MAP estimate for a given $\{\alpha, \gamma\}$ by standard Kalman filter EM methods. INLA calculates an approximate posterior jointly over the hyperparameters with the latent variables integrated out, $p(\alpha, \gamma|\boldsymbol{x})$. To approximate the marginal posterior, this requires the calculation of:

$$p(\alpha, \gamma|\boldsymbol{x}) \propto \left. \frac{p(\boldsymbol{x}, \{\mathbf{A}_1^*, \ldots, \mathbf{A}_J^*\}, \alpha, \gamma)}{p(\{\mathbf{A}_1^*, \ldots, \mathbf{A}_J^*\}|\boldsymbol{x}, \alpha, \gamma)} \right|_{\{\mathbf{A}_1^*, \ldots, \mathbf{A}_J^*\} = \{\mathbf{A}_1^*, \ldots, \mathbf{A}_J^*\}_{(\alpha,\gamma)}^{MAP}} \tag{11}$$

Letting $\tilde{\mathbf{R}}_b$ be a block matrix with block row j represented by $[\mathbf{R}_{j1b}, \ldots, \mathbf{R}_{jJb}]$, $\tilde{\boldsymbol{\nu}}_b = \text{vex}([\boldsymbol{\nu}_{1b}, \ldots, \boldsymbol{\nu}_{Jb}])$, and $\tilde{\boldsymbol{d}}_b = \text{vex}([\boldsymbol{a}_{1b}^*, \ldots, \boldsymbol{a}_{Jb}^*])$, the full measurement model can be equivalently written:

$$\log p(\boldsymbol{x}_b|\boldsymbol{y}_{jb}, \boldsymbol{d}_{jb}, j = \{1, \ldots, J\}) \simeq \text{const} - \frac{\tau T_b}{2} (\tilde{\boldsymbol{d}}_b^T \tilde{\mathbf{R}}_b \tilde{\boldsymbol{d}}_b - 2\tilde{\nu}_b^T \tilde{\boldsymbol{d}}_b) \tag{12}$$

After reformulation, this observation likelihood can be combined with the prior to get the MAP estimate for a given $\{\alpha, \gamma\}$ by standard Kalman filter EM methods. To calculate the full posterior over $\{\alpha, \gamma\}$, the INLA

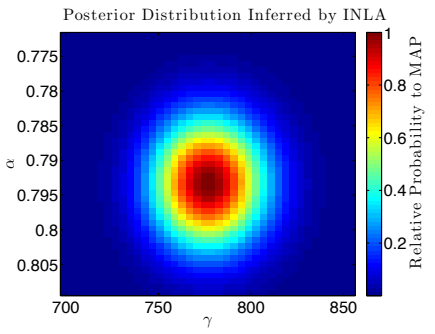

Figure 7: This is the posterior found from a single cell in INLA

method is used which allows for analysis of model uncertainty and also the calculation of the Bayesian MMSE estimator. INLA requires the calculation of:

$$p(\alpha, \gamma|\boldsymbol{x}) \propto \left. \frac{p(\boldsymbol{x}, \tilde{\mathbf{D}}, \alpha, \gamma)}{p(\tilde{\mathbf{D}}|\boldsymbol{x}, \alpha, \gamma)} \right|_{\tilde{\mathbf{D}}=\tilde{\mathbf{D}}^{MAP}_{(\alpha,\gamma)}} \tag{13}$$

$\tilde{\mathbf{D}}^{MAP}$ is found by forward filtering-backwards smoothing, and the conditional posterior can be calculated by $p(\tilde{\mathbf{D}} = \tilde{\mathbf{D}}^{MAP}(\alpha, \gamma)|\boldsymbol{x}, \alpha, \gamma) = p(\tilde{d}_B|\boldsymbol{x}, \alpha, \gamma) \prod_{b=1}^{B-1} p(\tilde{d}_b|\tilde{d}_{b+1}|\boldsymbol{x}, \alpha, \gamma) = \text{const} + (1/2) \log |\mathbf{\Phi}_B^{-1}| + \sum_{b=1}^{B-1} (1/2) \log |\mathbf{\Phi}_B^{-1} + \gamma \mathbf{I}|$, where $\mathbf{I}$ denotes the identity matrix, $\mathbf{\Phi}_1^{-1} = ((1 + \alpha^2)^{-1} \gamma \mathbf{I} + \tau T_1 \tilde{\mathbf{R}}_1)$, and $\mathbf{\Phi}_b^{-1} = ((\hat{\gamma}\mathbf{I} + \hat{\alpha}^{-2} \mathbf{\Phi}_{k-1}^{-1}) + \tau T_b \tilde{\mathbf{R}}_k)$. Speed improvements on this calculations are achieved by using the near sparsity of $\tilde{\mathbf{R}}_b$, and linear algebra identities are utilized so that only $\mathbf{\Phi}_b^{-1}$ needs to be calculated and store and the inverse is never calculated.

The procedure for calculating the approximate posterior is as follows. First, the MAP for $\{\alpha, \gamma\}$ is found by the EM maximization, and the curvature around the MAP point is estimated. A set of grid points is defined based upon the SVD of the curvature using the smart gridding algorithm shown in [16]. This quantity is evaluated at grid points around $\{\alpha, \gamma\}_{MAP}$ determined by the approximate curvature around the MAP estimate, and sample points are normalized to get the posterior estimate. Letting i denote i^{th} sample point, then Bayesian MMSE estimator can be evaluated as:

$$\tilde{\mathbf{D}}^{MMSE} = \sum_i \hat{p}(\alpha^{(i)}, \gamma^{(i)}|\boldsymbol{x}) \mathbb{E}_{p(\tilde{D}|\boldsymbol{x}, \alpha^{(i)}, \gamma^{(i)})}[\tilde{\mathbf{D}}] = \sum_i \hat{p}(\alpha^{(i)}, \gamma^{(i)}|\boldsymbol{x}) \tilde{\mathbf{D}}^{MAP}_{(\alpha^{(i)}, \gamma^{(i)})} \tag{14}$$

Additionally, the time-invariant model that [11] used for a single neuron and is generalized to multi-neuron setup is a special case of this inference when $B = 1$.