Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Structured Bayesian Modelling of Dynamic Receptive Fields in Salamander Retinal Ganglion Cells

Alokesh Manna

alokesh.manna17@gmail.com

Texas A&M University

Research Article

Keywords: receptive field estimation, retinal ganglion cells, spatio-temporal Bayesian model, Gaussian Markov random fields, INLA, population typing, salamander retina

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11286767/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Structured Bayesian Modelling of Dynamic Receptive Fields in Salamander Retinal Ganglion Cells

Alokesh Manna

Department of Statistics, Texas A&M University, College Station, TX 77843, USA

Correspondence: alokesh.manna17@gmail.com

Abstract

Retinal ganglion cells encode visual stimuli through receptive fields that are spatially localised and temporally dynamic. Estimating these receptive fields from spike-train recordings is challenging because it requires fitting thousands of spatio-temporal coefficients per neuron from a limited number of trials. Standard sparse regression methods such as the LASSO handle this dimensionality but treat pixels as independent predictors, producing spatially fragmented estimates that are biologically difficult to interpret. We propose a structured Bayesian model that explicitly encodes the two properties any biologically plausible receptive field must have: spatial coherence across neighbouring pixels, and smooth temporal evolution across successive time bins. The model is fit independently to each neuron using integrated nested Laplace approximation, which makes inference tractable at the scale of 155 simultaneously recorded neurons. Applied to salamander retinal ganglion cells stimulated with natural grayscale images, the model recovers spatially coherent, time-varying receptive fields where pixel-level LASSO returns fragmented patterns of isolated pixel selections. Clustering neurons by their temporal response profiles identifies three balanced phenotypes ($n = 85$, 32, 38) that differ in the timing and depth of their inhibitory response decline, broadly consistent with known sustained and transient response classes in the salamander retina. A simulation study with known ground truth confirms that the structured model recovers receptive field support and coefficient magnitude more accurately than unregularised regression, LASSO, and the elastic net across all tested configurations.

Keywords: receptive field estimation; retinal ganglion cells; spatio-temporal Bayesian model; Gaussian Markov random fields; INLA; population typing; salamander retina.

1 Introduction

Neurons in the visual system respond selectively to different spatial and temporal characteristics of visual stimuli. One important statistical object for characterising this selectivity is the *receptive field*, which describes the region of visual space and corresponding stimulus features that influence a neuron's activity. Estimating receptive fields from neural recordings is therefore an important problem in computational neuroscience. Characterising each

neuron's receptive field individually — rather than averaging over neurons — is essential for understanding how response properties are distributed across a population: each retinal ganglion cell samples a different spatial region with a different temporal profile, and these individual differences are what make population-level heterogeneity tractable to study.

As a motivating example, consider an experiment in which the firing activity of retinal ganglion neurons from a salamander is recorded while a sequence of natural grayscale images is repeatedly presented as visual stimuli (Liu et al., 2022); Section 2 describes this dataset in full. For each image and each short time window after its onset, one observes a spike count, and the applied question is which pixels of the image, at which time lag, are associated with elevated firing. This is a large-scale localised-effect problem: the number of candidate spatio-temporal predictors is far larger than what a handful of trials per image can support, the predictors (neighbouring pixels) are highly correlated, and any scientifically useful answer should point to a spatially coherent image region rather than an arbitrary subset of pixels.

Suppose that a visual stimulus is represented by a two-dimensional array of pixels. For each stimulus presentation, neural activity is recorded over a sequence of temporal bins. A natural statistical model is a Poisson regression in which the observed spike count is modelled as a function of the pixel intensities. Let Y_{nkt} denote the spike count for neuron k, stimulus presentation n, and temporal bin t. We consider

$$Y_{nkt} \mid \lambda_{nkt} \sim \text{Poisson}(\lambda_{nkt}), \tag{1}$$

with

$$\log(\lambda_{nkt}) = \alpha_k + \sum_{s=1}^{P} X_{ns}\beta_{kst}, \tag{2}$$

where α_k is neuron k's intercept — its baseline log firing rate when every pixel intensity is zero — X_{ns} denotes the intensity of pixel s in the image shown on presentation n, P is the number of pixels, and β_{kst} represents the effect of pixel s on neuron k's firing at time t. We include α_k because neurons differ in overall excitability for reasons unrelated to the stimulus; without it, baseline differences would be absorbed into $\boldsymbol{\beta}_{kt}$, distorting the receptive-field estimates.

The collection

$$\boldsymbol{\beta}_{kt} = (\beta_{k1t}, \ldots, \beta_{kPt})^\top \tag{3}$$

can therefore be interpreted as a receptive-field surface whose spatial organisation has direct scientific meaning.

The dimensionality of the problem can become substantial. For example, a 13×13 image observed over 30 temporal bins gives

$$13 \times 13 \times 30 = 5{,}070 \tag{4}$$

spatio-temporal coefficients for each neuron. With 155 recorded neurons, an unstructured model would involve on the order of 786,000 parameters. This is precisely the regime in which standard Markov chain Monte Carlo (MCMC) becomes impractical: at this dimension, with the strong spatial and temporal correlation a real receptive field has by construction, MCMC updates mix too slowly to reach convergence within a practical time budget. This is the concrete reason we turn to integrated nested Laplace approximation (INLA) rather than

MCMC in Section 4: INLA replaces sampling the full high-dimensional coefficient field with numerical integration over a handful of hyperparameters, exploiting the Gaussian structure this paper places on $\boldsymbol{\beta}_{kt}$ (Section 4.3 states this precisely).

A common solution to high-dimensional regression is the LASSO, which estimates coefficients using an L_1 penalty. Although this produces sparse models, ordinary LASSO does not explicitly encode the spatial organisation of neighbouring pixels. This is problematic in receptive-field estimation because neighbouring pixels are often strongly related, and biologically meaningful receptive fields are generally expected to occupy spatially organised regions rather than consist of unrelated isolated pixels.

The limitations of LASSO in the presence of dependent predictors have been studied extensively. Zou and Hastie (2005) showed that LASSO can behave poorly when predictors are strongly correlated and introduced the elastic net partly to obtain a grouping effect among correlated predictors. More broadly, Freijeiro-González et al. (2022) reviewed variable-selection behaviour under dependence among covariates and emphasised that the assumptions required for reliable LASSO selection can be restrictive in dependent high-dimensional settings.

These issues motivate a different formulation for receptive-field estimation. Instead of treating every pixel as an independent candidate variable, we propose to model the receptive field of each individual neuron as a structured spatio-temporal object: spatially coherent across pixels and smoothly evolving across time bins.

The central objective of this work is therefore to develop a structured Bayesian model for individual-level receptive field estimation that simultaneously addresses:

(1) spatial dependence among neighbouring pixels;

(2) temporal dependence in receptive-field effects; and

(3) sparsity of the relevant visual region.

This leads to the following scientific questions:

- Which spatial regions of the stimulus influence a neuron's firing?
- How does the receptive field evolve over time within a trial?
- Can structured spatial and temporal smoothness recover biologically meaningful receptive-field regions more accurately and stably than conventional LASSO?
- Do neurons in the recorded population cluster into a small number of distinguishable temporal-response phenotypes?

The spatially and temporally structured model of Section 4 — a Gaussian Markov random field in space combined with an autoregressive process in time — is fit independently for each of the 155 neurons throughout Sections 5–5.2. The fitted per-neuron surfaces are then used to ask whether the population clusters into a small number of distinguishable temporal-response phenotypes. This is the paper's complete empirical programme; no cross-neuron pooling during fitting is attempted or claimed.

The contribution of this paper is threefold. First, we develop a spatio-temporally structured Bayesian model for individual-level receptive-field estimation that enforces spatial coherence and temporal smoothness simultaneously, and show that this structured representation recovers biologically interpretable receptive fields where pixel-independent regularisation produces fragmented estimates. Second, we demonstrate that fitting this model independently to 155 neurons — at 5,070 spatio-temporal coefficients per neuron — is computationally feasible via integrated nested Laplace approximation, which replaces MCMC sampling of the high-dimensional latent field with numerical integration over four scalar hyperparameters. Third, we show that clustering per-neuron temporal response profiles from these fits yields three interpretable phenotypes whose temporal dynamics are broadly consistent with known response classes in the salamander retina.

Statistical modelling of neural spike-train data has a substantial literature (Kass and Ventura, 2001; Paninski et al., 2008). The Poisson generalised linear model (GLM) framework for receptive-field estimation, introduced by Pillow et al. (2008), has become a standard tool for characterising the spatio-temporal tuning of retinal ganglion cell populations; the present work extends this framework by replacing pixelwise-independent regularisation with a spatially and temporally structured Bayesian prior. Park and Pillow (2011) demonstrated that imposing explicit spatial-locality priors on receptive-field coefficients yields substantially more accurate estimates than pixelwise regularisation under limited data; the model developed here pursues the same principle within a fully Bayesian, computationally scalable framework that simultaneously handles spatial coherence, temporal smoothness, and population-level phenotyping.

2 Neural Spike Data

We consider a neuroscience experiment that recorded the firing activity of retinal ganglion neurons from a salamander (Liu et al., 2022). This follows a substantial body of work using Poisson generalised linear models to characterise the spatio-temporal receptive fields of retinal ganglion cell populations from spike-train data (Pillow et al., 2008). The experiment involved repeatedly presenting 303 distinct grayscale images, with each image shown 13 times, resulting in 3,939 trials per neuron. Each trial lasted 200 milliseconds, during which the number of neuronal spikes was counted; we divided this 200-millisecond span into 30 equally spaced time intervals and, for each trial, counted the number of neural firings within each interval. The images were originally 256×256 pixels, which we coarsened down to 13×13.

The raw recording contains 156 neurons. One neuron (neuron 134 in the original indexing) was excluded because it could not be assigned a cluster label in the population-level analysis of Section 5.2 — most likely reflecting a recording or labelling artifact rather than a genuine biological outlier, though the precise cause was not investigated further. All analyses therefore use the remaining 155 neurons. For Section 5.1 we focus on a single demonstration neuron (neuron 1 in the original indexing), chosen as a representative example of a well-identified, moderate-firing cell; the population-level analysis of Section 5.2 uses all 155.

This dataset is well-suited to the modelling framework developed below, as it contains rich spatial structure from the image stimuli, temporal variability across trials, and repeated measures that enable us to disentangle stimulus-driven effects from intrinsic neuronal noise.

Figure 1 shows four representative stimulus images after coarsening to the 13×13 pixel grid used throughout this paper. Pixel intensity is mean-centred per image (grayscale value minus the image's own mean), which is the covariate scale used in every model fit that follows.

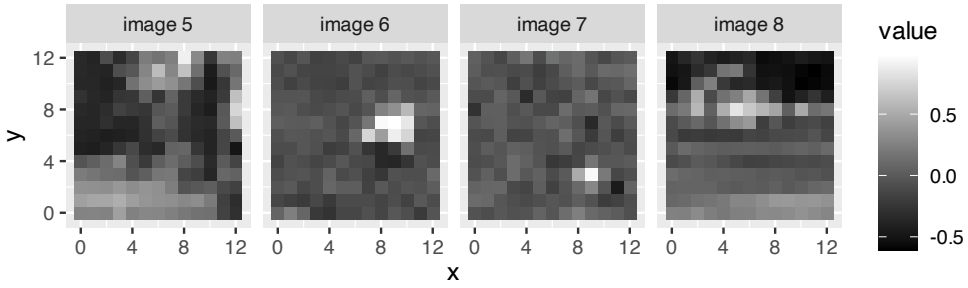


Figure 1: Four representative 13×13 stimulus images (mean-centred pixel intensity) from the salamander retinal ganglion cell dataset.

3 Motivation: Why LASSO Is Not Sufficient

3.1 Ordinary sparse regression

The LASSO estimates a coefficient vector by solving an optimisation problem of the form

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ -\ell(\boldsymbol{\beta}) + \lambda \sum_{s=1}^{P} |\beta_s| \right\}, \tag{5}$$

where $\ell(\boldsymbol{\beta})$ is the log-likelihood and $\lambda \geq 0$ is the penalty parameter controlling the amount of shrinkage — a different quantity from the Poisson rate λ_{nkt} of Section 1, a standard notational overload that we flag once here.

The L_1 penalty encourages many coefficients to be exactly zero, which is attractive for high-dimensional receptive-field estimation.

However, it does not distinguish between two very different configurations:

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}. \tag{6}$$

Both configurations contain exactly four nonzero coefficients, but the first represents a spatially coherent receptive field whereas the second is spatially fragmented, with no two active pixels adjacent to one another. The L_1 penalty itself does not encode this distinction.

3.2 A motivating empirical illustration

The fragmentation problem above is not merely hypothetical: using the neural spike data introduced in Section 2, we now show that it occurs in practice, and that a coarser, image-level analysis cannot resolve it either.

We first ask a deliberately coarse diagnostic question: does the *whole image*, rather than any particular pixel, predict elevated firing? For image $n \in \{1, \ldots, 303\}$, let $\bar{y}_n$ denote the mean spike count across all trials and time bins in which that image was shown; we split the images into a "yes" group (the more strongly firing-associated half, ranked by $\log \bar{y}_n$) and a "no" group, and average the (mean-centred) pixel intensities within each. Figure 2 shows the result: the "yes" group has a visibly darker horizontal band through the middle rows relative to the "no" group's small, localised bright patch — an early hint that stimulus energy in that mid-image band is associated with elevated firing. Because this comparison operates at the level of whole images, it says nothing about *where*: it cannot by itself localise a receptive field within the 13×13 grid.

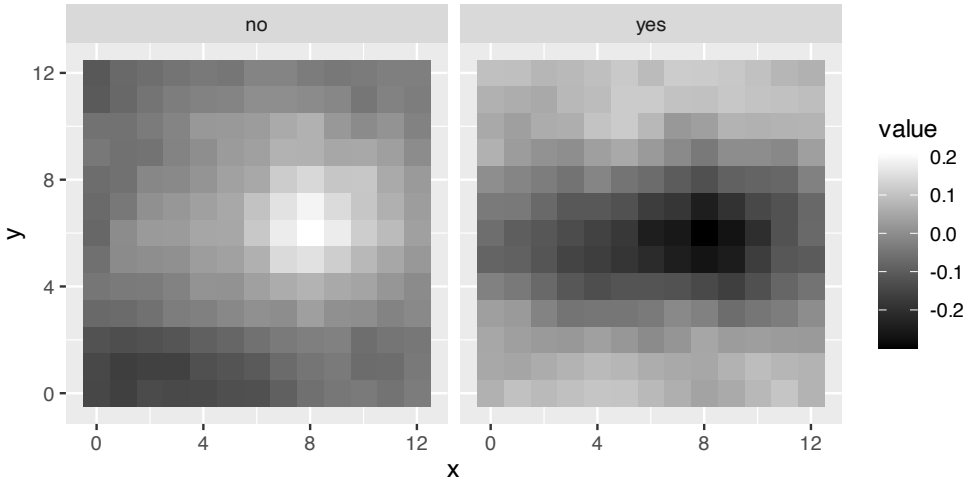


Figure 2: Pixel-wise average stimulus image for images classified as more strongly firing-associated ("yes") versus less firing-associated ("no") under the image-level Poisson baseline described above.

To localise the effect, we next fit an ordinary pixel-level Poisson-LASSO regression, pooled across all 30 time bins and 13 repeats, with no held-out split and no spatial or temporal smoothing. Writing y_n for the spike count at observation n (one of $30 \times 13 \times 303 = 118{,}170$ time-bin/repeat/image combinations), X_{ns} for pixel s's mean-centred intensity in the corresponding image, and $\mathbf{x}_n = (X_{n1}, \ldots, X_{n,169})^\top$ for the vector collecting all 169 pixel intensities of that observation,

$$\widehat{\boldsymbol{\beta}}_{\text{lasso}} = \arg\min_{\boldsymbol{\beta}} \left\{ -\sum_n \left(y_n \, \mathbf{x}_n^\top \boldsymbol{\beta} - \exp(\mathbf{x}_n^\top \boldsymbol{\beta}) \right) + \lambda \sum_{s=1}^{169} |\beta_s| \right\}. \tag{7}$$

This illustrative fit uses a single fixed penalty ($\lambda = 0.02$) on the full pooled data, rather than a cross-validated one, and gives a sparse fit with 27 of the 169 pixel coefficients nonzero.

Figure 3 shows this fit. The LASSO does recover a small contiguous block of negative coefficients — consistent with the mid-image band flagged by the image-level baseline above — but it also selects several isolated, spatially disconnected pixels scattered near the image border that share no boundary with the main cluster or with one another. The L_1 penalty has no mechanism to prefer a spatially coherent active set over a scattered one with the same number of nonzero entries, so isolated-pixel selections with no clear biological interpretation and a genuine receptive-field cluster are penalised identically.

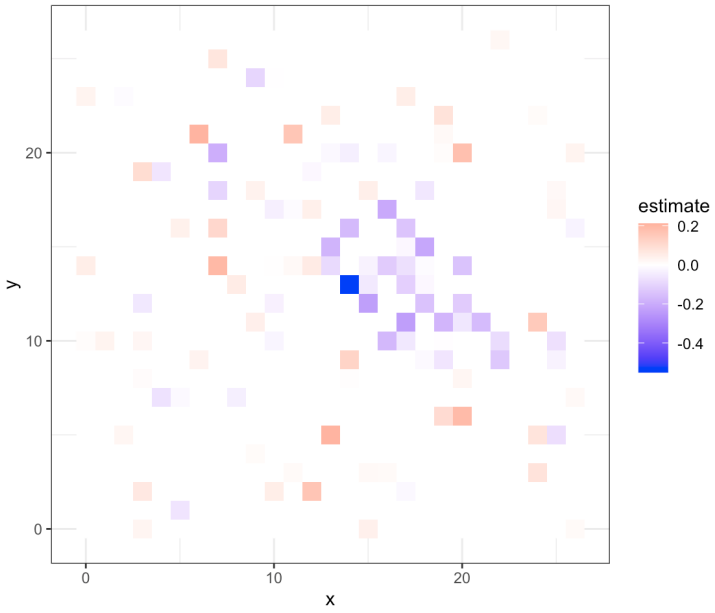


Figure 3: Pixel-level Poisson-LASSO fit at a fixed illustrative penalty ($\lambda = 0.02$), 27 of 169 pixel coefficients nonzero. Beyond the contiguous central cluster, several isolated, spatially disconnected pixels are also selected.

Together, Figures 2 and 3 motivate the remainder of the paper: Section 4 develops a Bayesian model whose spatial prior encourages neighbouring pixels to be similar, which discourages — though does not by itself guarantee — the kind of isolated, disconnected selections seen in Figure 3. Section 5 applies the model to this same neuron, completing the comparison with an explicit spatio-temporal fit.

3.3 Correlated image predictors

Neighbouring pixels in an image are generally correlated. If two nearby pixels contain similar information, it may be difficult to determine which individual pixel should receive the nonzero coefficient.

This is a well-known issue in LASSO variable selection under correlated predictors. Zou and Hastie (2005) specifically proposed the elastic net because LASSO may select only one variable from a group of strongly correlated variables, whereas the elastic net encourages correlated variables to enter or leave the model together.

The problem is particularly relevant here because the scientific target is not necessarily an individual pixel. Rather, the target is a spatial region representing a receptive field.

3.4 Shrinkage bias

The LASSO also shrinks all nonzero coefficients towards zero. Consequently, sparsity and estimation are obtained simultaneously through the same penalty.

For receptive-field analysis this can be undesirable because the magnitude of β_{kst} can itself be scientifically meaningful. A method that strongly shrinks large effects may distort the estimated strength of neuronal responses.

3.5 Spatial and temporal dependence

The receptive field is not expected to change arbitrarily from one pixel to another or from one temporal bin to the next. We therefore expect $\beta_{k,s,t}$ to be related to $\beta_{k,s',t}$ when s and s' are neighbouring pixels, and to $\beta_{k,s,t-1}$ when adjacent temporal bins are considered. Ordinary LASSO does not explicitly incorporate either relationship. This is not merely a modelling convenience: Park and Pillow (2011) show that imposing an explicit spatial-locality prior on receptive-field coefficients yields substantially more accurate and interpretable estimates than pixelwise-independent regularisation, particularly when data are limited relative to the number of stimulus dimensions.

3.6 Heterogeneous neurons

Finally, a collection of neurons may contain several functional types. Two neurons of the same type may have similar receptive-field organisation, while neurons belonging to different types may have systematically different spatial or temporal response patterns.

Treating every neuron independently ignores this potentially valuable source of information. Conversely, forcing all neurons to have the same receptive field ignores biological heterogeneity. The population typing of Section 5.2 investigates this structure directly from the per-neuron fits, without requiring any cross-neuron pooling during fitting.

4 The Structured Bayesian Model

What we fit, stated first. The scientific goal is to characterise the receptive field of each individual neuron — to answer, for neuron k, which pixels at which time lags are associated with elevated firing, and how that spatio-temporal pattern evolves smoothly over the trial. This is an individual-level estimation problem: the model is fit separately and independently to each of the 155 neurons, with no information shared across neurons during fitting.

Every result in this paper comes from one model, fit independently for each of the 155 recorded neurons: a structured Bayesian model that treats a neuron's receptive-field surface $\boldsymbol{\beta}_{kt}$ as a latent Gaussian object with (i) a neuron-specific intercept α_k absorbing baseline excitability, (ii) an i.i.d. random effect for repeated presentations of the same image, (iii) a spatially structured Gaussian Markov random field (via an SPDE representation) that ties the

coefficient at each pixel to its neighbours, and (iv) an AR(1) process linking that spatial field smoothly across time bins. The word "hierarchical" in what follows refers to this model's own internal structure — a top level of hyperparameters governing a lower-level latent Gaussian field, in turn governing the observed spike counts — not to any pooling across neurons. Section 4.2 (Part A) states this model precisely; it is the only model whose results appear anywhere in Sections 5–5.2 or the simulation study of Section 7.

4.1 Neural response model and observation family

Let Y_{nkt} denote the spike count of neuron k during temporal bin t for stimulus presentation n.

$$Y_{nkt} \mid \lambda_{nkt} \sim \text{Poisson}(\lambda_{nkt}), \qquad \log \lambda_{nkt} = \alpha_k + \mathbf{X}_n^\top \boldsymbol{\beta}_{kt}, \tag{8}$$

where α_k is neuron k's intercept, $\mathbf{X}_n = (X_{n1}, \ldots, X_{nP})^\top$ contains the pixel intensities, and $\boldsymbol{\beta}_{kt} = (\beta_{k1t}, \ldots, \beta_{kPt})^\top$ is the receptive-field coefficient vector for neuron k at time t; the sign and magnitude of β_{kst} describe the association between stimulus intensity at spatial location s and neural firing at time t.

Why Poisson. Spike counts Y_{nkt} are non-negative integers recorded over a fixed time window, which is precisely the sampling situation the Poisson distribution is built for, and it is the standard observation model for spike-train data in this literature (Pillow et al., 2008). The log link guarantees $\lambda_{nkt} > 0$ for any real-valued $\boldsymbol{\beta}_{kt}$, which is what lets us place an unconstrained Gaussian (GMRF/SPDE) prior directly on the coefficient surface without a separate positivity constraint. The Poisson model does impose the restriction that the conditional variance equals the conditional mean. A dispersion check is reported alongside the single-neuron fit in Section 5.1; the negative-binomial extension, should overdispersion be detected, is listed as follow-up work in Section 8.

4.2 Part A — the model we fit

For each neuron k, independently, we model the coefficient surface $\boldsymbol{\beta}_{kt}$ directly as a spatially smoothed, temporally autoregressive Gaussian field, fit by INLA. Concretely, following the SPDE approach of Bakka et al. (2018), we represent the spatial component for neuron k as a continuous-domain Gaussian random field $w_k(\mathbf{u})$ over the pixel grid, discretised on a triangulated mesh (shared across neurons) with basis functions $\psi_1, \ldots, \psi_M$,

$$w_k(\mathbf{u}) = \sum_{m=1}^{M} \psi_m(\mathbf{u})\, w_{km}, \qquad \mathbf{w}_k = (w_{k1}, \ldots, w_{kM})^\top, \tag{9}$$

where M is the number of mesh nodes. Writing $\mathbf{A} \in \mathbb{R}^{P \times M}$ for the projector matrix that evaluates the mesh-based field at the $P = 169$ pixel centres, the pixel-level spatial coefficient for neuron k at time group t is $\boldsymbol{\beta}_{kt}^{\text{sp}} = \mathbf{A}\mathbf{w}_{kt}$, and the mesh-node field evolves across the $T = 30$ time groups by an AR(1) process,

$$\mathbf{w}_{kt} = \rho_k\, \mathbf{w}_{k,t-1} + \boldsymbol{\epsilon}_{kt}, \qquad \boldsymbol{\epsilon}_{kt} \overset{\text{iid}}{\sim} N(\mathbf{0}, Q_w^{-1}), \qquad |\rho_k| < 1, \qquad t = 2, \ldots, T, \tag{10}$$

with $\mathbf{w}_{k1} \sim N\left(\mathbf{0},\, \frac{1}{1-\rho_k^2} Q_w^{-1}\right)$ the stationary marginal distribution implied by this recursion, and a penalised-complexity prior on ρ_k.

How Q_w governs spatial smoothing. Q_w is built from a finite-element discretisation of a differential operator on the mesh (Lindgren et al., 2011), sparse by construction, with a zero entry for every pair of mesh nodes that are not spatial neighbours. This sparsity pattern defines the smoothing: by the Gaussian Markov random field conditional-mean identity, each mesh-node coefficient is pulled only towards a precision-weighted average of its immediate neighbours, never towards non-adjacent nodes, which is the precise mechanism discouraging the isolated, disconnected pixel selections of Figure 3. Appendix B gives the full construction of Q_w and works out exactly how its two hyperparameters (κ, σ_w) separately control the range and strength of this coupling.

The full per-neuron model fit in Sections 5–5.2 is then

$$\log \lambda_{nkt} = \alpha_k + b_{r(n)} + \left(\mathbf{X}_n^\top \mathbf{A}\right) \mathbf{w}_{kt}, \qquad b_{r(n)} \sim N(0, 1/\tau_b), \tag{11}$$

where $r(n) \in \{1, \ldots, 13\}$ indexes which of the 13 repeated presentations observation n's trial belongs to, and $b_{r(n)}$ is an i.i.d. Gaussian random effect absorbing variability associated with repetition number at fixed image and time bin. There is no pooling of $\mathbf{w}_{kt}$ across k in Part A. Writing $\beta_{ijt} = (\mathbf{A}\mathbf{w}_{kt})_{ij}$ for the pixel-level coefficient at pixel (i, j), the model can be read as

$$\log \lambda_{nt} = \alpha + \sum_{i=1}^{13} \sum_{j=1}^{13} \beta_{ijt} X_{nij} + b_{r(n)}, \tag{12}$$

where β_{ijt} is generated by projecting the lower-dimensional mesh field $\mathbf{w}_t$ onto the pixel grid — this is what makes the 5,070-coefficient surface tractable to estimate from 3,939 trials.

4.3 Bayesian inference via INLA

Collecting the latent Gaussian quantities — intercept α, repetition random effects $\mathbf{b}$, and mesh-node fields $\mathbf{w}_1, \ldots, \mathbf{w}_T$ — into $\boldsymbol{\theta} \in \mathbb{R}^d$ with $d = 1 + 13 + MT$, and the four hyperparameters $(\kappa, \sigma_w, \rho, \tau_b)$ into $\boldsymbol{\psi}$, the model is a *latent Gaussian model* (Rue et al., 2009): $\boldsymbol{\theta} \mid \boldsymbol{\psi}$ is Gaussian with block-tridiagonal precision $Q(\boldsymbol{\psi})$ (the standard AR(1)-over-SPDE "group model" of Bakka et al. 2018), and the Poisson observations depend on $\boldsymbol{\theta}$ only through linear predictors $\eta_n = \mathbf{b}_n^\top \boldsymbol{\theta}$. Penalised-complexity hyperpriors (Bakka et al., 2018) are placed on $\boldsymbol{\psi}$; full prior specifications are given in Appendix E.

Why INLA rather than MCMC. At $d \approx 5{,}070$ strongly correlated latent dimensions, standard MCMC mixes prohibitively slowly: the SPDE/AR(1) prior explicitly encodes spatial and temporal correlation, so coordinate-wise or small-block updates make negligible progress per iteration, and convergence assessment across multiple chains is impractical within a reasonable time budget for 155 neurons. INLA avoids sampling $\boldsymbol{\theta}$ entirely: because $\boldsymbol{\theta} \mid \boldsymbol{\psi}$ is Gaussian and the Poisson likelihood is log-concave, the conditional posterior $\pi(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \mathbf{y})$ is accurately approximated by a Laplace approximation computed via sparse linear algebra on $Q(\boldsymbol{\psi})$. The remaining integration is over $\boldsymbol{\psi}$ alone — four scalars — so the computational

cost scales with $\dim(\boldsymbol{\psi}) = 4$ rather than d (Rue et al., 2009, 2017). We fit all models using the `R-INLA` package (Bakka et al., 2018); implementation details and convergence diagnostics for all 155 fits are given in Appendix C.

5 Real-Data Analysis

This section reports the real-data analysis in two parts: a single-neuron demonstration of the fitted model (Section 5.1), and a population-level typing of all 155 recorded neurons built from the same per-neuron fit (Section 5.2). It is real data throughout, with a real but unknown receptive field — see Section 7 for the corresponding simulation comparison, which is the paper's primary quantitative evidence for how the proposed model compares against simpler alternatives under known ground truth.

5.1 Single-neuron demonstration

Section 3.2 showed that an image-level Poisson baseline cannot localise the receptive field (Figure 2) and that an ordinary pixel-level Poisson-LASSO fit, while it does localise, returns a spatially fragmented coefficient pattern (Figure 3). We now complete that comparison by fitting Part A of Section 4 to the same neuron. The model is

$$y_{nt} \sim \text{Poisson}(\lambda_{nt}), \tag{13}$$

$$\log(\lambda_{nt}) = \alpha + \sum_{i=1}^{13}\sum_{j=1}^{13} \beta_{ijt} X_{nij} + b_{r(n)}, \tag{14}$$

where $n = 1, \ldots, 3{,}939$, $t = 1, \ldots, 30$, and β_{ijt} is generated by projecting the SPDE mesh field $\mathbf{w}_t$ onto the pixel grid following the AR(1)-in-time, SPDE-in-space structure of Section 4.2.

Why INLA, for this neuron. For this neuron specifically: $n = 3{,}939$ trials, $T = 30$ temporal bins, and a 13×13 pixel grid, so the spatial random effect alone projects to on the order of 5,070 correlated pixel-level coefficients — a generic MCMC treatment of a latent field this large is not computationally feasible within a reasonable time frame, which is exactly the regime Section 4.3's argument is built for. The SPDE prior directly encodes that β_{ijt} and $\beta_{i'j't}$ should be similar when pixels (i, j) and (i', j') are spatially close; this smoothness discourages, though does not by construction guarantee, the isolated, disconnected selections seen in Figure 3.

Dispersion check. To assess whether the Poisson equal-dispersion assumption is adequate for this neuron, we compute the Pearson χ^2 statistic from the INLA posterior-mean fitted values and divide by the residual degrees of freedom ($n - \dim \boldsymbol{\psi} = 3{,}939 - 4 = 3{,}935$). The resulting ratio is 0.75, consistent with the Poisson mean-variance equality; values below 1 are common in spike-count data where the stimulus explains a large fraction of the variance, and this neuron shows no evidence of overdispersion. A negative-binomial extension remains listed as follow-up work in Section 8 should overdispersion appear more pronounced in neurons not examined here.

Higher intensity is shown in yellow and lower intensity in blue. The blue region is a point of interest that sparks the neuron more strongly around time interval 15–20, as shown in Figure 4. The spatially varying coefficients appear in Figure 5. In the rectangle $\{(x, y) : 4 \leq x \leq 10,\ 4 \leq y \leq 8\}$, the blue intensity is more prominent in the time window 16–20, where the coefficients fluctuate more than in the other time intervals — the same central spatial region that the empirical illustration of Section 3.2 already pointed to, now resolved with an explicit, spatially coherent, time-varying coefficient surface.

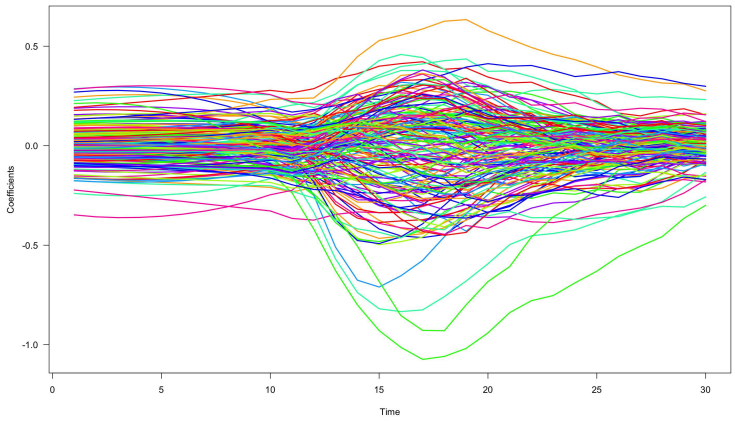


Figure 4: Space-varying receptive field coefficients estimated by the SPDE/AR(1) model, plotted as a function of time bin for the demonstration neuron. Each curve corresponds to one spatial pixel location; the concentration of large-magnitude curves around time bins 15–20 indicates the temporal window of strongest stimulus-driven response.

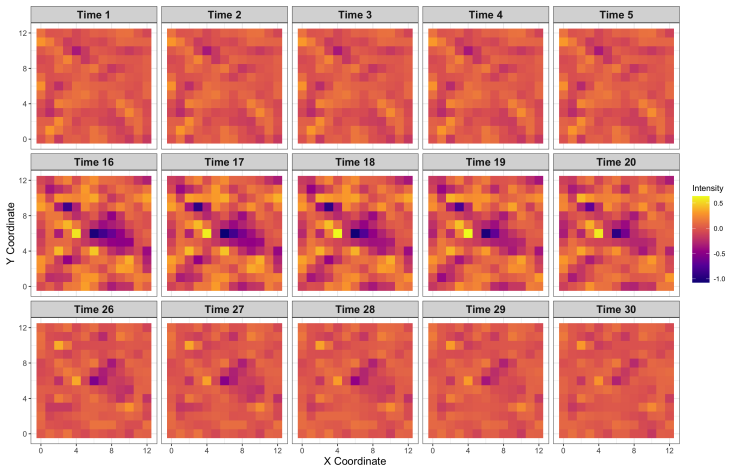


Figure 5: Spatio-temporal receptive field coefficient surface for the demonstration neuron, shown at selected time bins. Blue regions indicate inhibitory associations between pixel intensity and firing rate (i.e. $\beta_{ijt} < 0$); yellow regions indicate excitatory associations. The central spatial region ($4 \leq x \leq 10$, $4 \leq y \leq 8$) shows the strongest inhibitory response in time bins 16–20, consistent with the temporal peak visible in Figure 4.

This demonstration fit to a single neuron lets us walk through the model, the image-level baseline, and the pixel-level LASSO comparison on one concrete case. Section 5.2 extends the same per-neuron fit to the full recorded population.

5.2 Population-level neuron typing

The single-neuron fit above answers "where does this neuron respond?" A natural follow-up question is whether the 155 neurons divide into a small number of distinguishable groups with systematically different temporal-response patterns, or whether the population is effectively homogeneous in its dynamics. We investigate this here using the full recorded population of 155 neurons. We deliberately refer to the resulting clusters as *temporal-response phenotypes* rather than *neuron types*: they are groups defined by clustering independently fitted response curves, and treating them as biologically validated cell types would be a stronger claim than the present analysis supports (Section 8).

For each of the 155 neurons we repeated the fit of the previous subsection independently, using the same Poisson-response, SPDE-in-space, AR(1)-in-time model, giving each neuron its own spatio-temporal coefficient surface $\widehat{\boldsymbol{\beta}}_{k1}, \ldots, \widehat{\boldsymbol{\beta}}_{kT}$. The population typing is then a two-stage procedure: independent per-neuron estimation followed by post hoc clustering of the resulting surfaces.

A clustering analysis of 155 objects, each a $13 \times 13 \times 30 = 5{,}070$-dimensional coefficient surface, first requires choosing a representation. We collapsed each neuron's surface to its space-averaged temporal response profile — a single curve over the $T = 30$ time bins, obtained by averaging the coefficient surface over the 13×13 spatial grid at each time point. Concretely, the clustering pipeline has four steps, in the spirit of James and Sugar (2003): (1) expand each neuron's 30-point temporal profile in a cubic B-spline basis with 8 basis functions, by least-squares projection, giving an 8-dimensional coefficient vector per neuron; (2) run principal components analysis on these 155 basis-coefficient vectors (functional PCA under the B-spline basis) and retain the smallest number of components, at least 2, whose cumulative variance explained reaches 99%; (3) fit a Gaussian mixture model (via `mclust`) to the resulting PCA scores for each candidate number of clusters $k = 2, \ldots, 10$ and each candidate covariance-model family; (4) select both k and the covariance model by BIC. This selected $k = 3$, with cluster sizes 85, 32 and 38 neurons: a well-balanced three-way partition, grouping neurons by the shape of their temporal response.

Characterising the three phenotypes. For the primary, temporal-profile-based partition (sizes 85, 32, 38), we examined the mean temporal response curve within each cluster (Figure 6) and the spatial location of each cluster's peak and magnitude-weighted-centroid response (Figure 7), computed as follows. For each of the three clusters, we averaged the fitted $13 \times 13 \times 30$ coefficient surfaces of its member neurons, giving one cluster-average spatio-temporal field. Within that field we located the single pixel-by-time-bin cell of largest absolute magnitude; its spatial coordinates are the "peak" (triangle, Figure 7). The "magnitude-weighted centroid" (circle, Figure 7) is the spatial average of all 169 pixel coordinates, each weighted by the absolute value of the cluster-average field at the peak time bin.

All three phenotypes share the same broad temporal shape — a positive response over

roughly the first third of the trial, a decline into negative territory around the middle third, and a partial late recovery — but differ systematically in the timing and depth of this decline (Figure 6). Phenotype 2 ($n = 32$) declines earliest and most sharply, reaching the most negative mean response of the three by roughly two-thirds of the way through the trial, and recovers only partially by the end. Phenotype 3 ($n = 38$) declines slightly earlier than Phenotype 1 but recovers to the highest final value of the three. Phenotype 1 ($n = 85$), the largest and most heterogeneous group, declines latest and most gradually.

Firing-rate characteristics of the three phenotypes. The three phenotypes differ not only in the timing of their temporal response but also in their overall level of neural activity. Phenotype 2 ($n = 32$) is associated with substantially lower firing activity than the other two groups: 46.1% of its (neuron, image) cells produced zero spikes across all time bins, with a median of 1 spike per cell, compared with medians of 20.0 spikes per cell for Phenotype 1 (25.6% silent cells) and 16.0 spikes per cell for Phenotype 3 (21.8% silent cells). This difference is a population-level property of Phenotype 2 as a whole, not an artefact of sampling. The low-firing character compounds the identifiability challenge: both its small cluster size ($n = 32$) and its low per-neuron signal contribute to greater posterior uncertainty in its fitted receptive field surfaces. Whether this near-floor firing activity reflects a genuine biological property — such as a high spike threshold or low baseline excitability — or a recording artefact cannot be determined from spike counts alone.

Spatially, the magnitude-weighted centroids of the three phenotypes' average receptive fields (circles, Figure 7) fall close together near the centre of the pixel grid, indicating that the three phenotypes are not primarily distinguished by the gross location of their receptive field. Each phenotype's single peak-intensity pixel (triangles, Figure 7) falls near the grid's periphery, but at three different peripheral locations. The population typing recovered here is best described as three temporal-response dynamics — an early-and-sharp decline, a late-and-gradual decline, and an intermediate pattern — shared by neurons whose average receptive fields occupy broadly the same central visual region.

The temporal dynamics of the three phenotypes are broadly consistent with the known diversity of retinal ganglion cell response profiles in amphibian retina. Salamander retinal ganglion cells have been characterised into sustained and transient response classes based on the time course of their firing (Segev et al., 2004; Marre et al., 2012), and OFF-type cells — which are suppressed by light onset and activated by light offset — characteristically exhibit the negative-going temporal profiles seen throughout our population. The sharp, early inhibitory decline of Phenotype 2 is consistent with a transient OFF-type response, while the gradual, late decline of Phenotype 1 is more consistent with a sustained temporal profile (Pillow et al., 2008). The markedly low firing activity of Phenotype 2 (46.1% silent cells, median 1 spike) is also consistent with cells near their spike threshold under natural image stimulation, a property observed in weakly driven cells in amphibian retina. We do not claim a one-to-one correspondence between our phenotypes and any established physiological classification: the clustering operates on fitted model surfaces rather than direct electrophysiological measurements, and the dataset does not include cell-type labels needed for a formal comparison. Validating the correspondence — for example by comparing our phenotype assignments against ON/OFF or sustained/transient annotations from the same

preparation — is an important direction for future work.

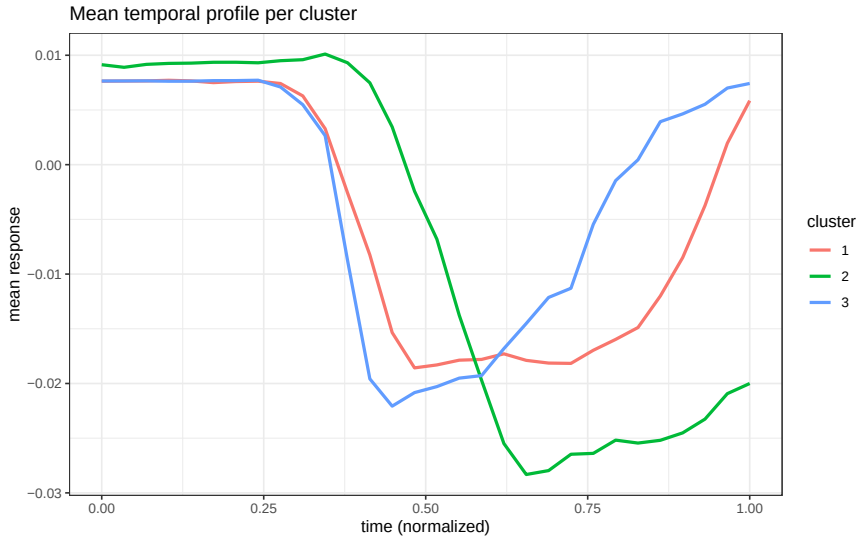


Figure 6: Mean temporal response curve for each of the three temporal-response phenotypes identified by BIC-selected functional clustering of the space-averaged temporal profile ($n = 85, 32, 38$).

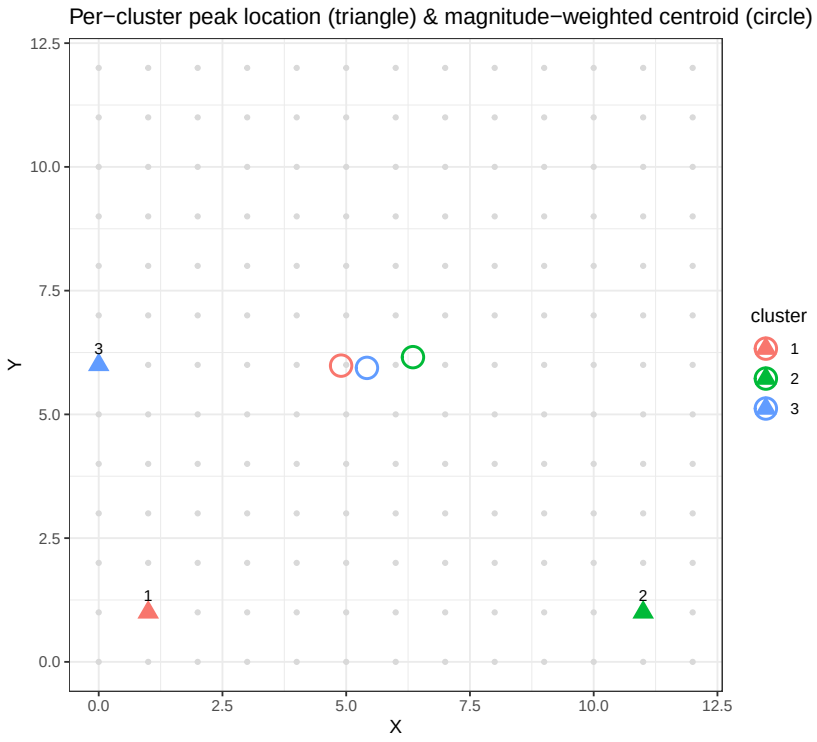


Figure 7: Per-phenotype peak pixel location (triangle) and magnitude-weighted centroid (circle) of the average receptive field, for each of the three temporal-response phenotypes.

Validation of the three-phenotype partition. The partition rests on BIC model selection within the functional clustering pipeline. As an independent check, we fit two Poisson models on a stratified sample of 31 neurons using the first 20 stimulus images: a clusterwise model with one shared spatio-temporal SPDE/AR(1) field per phenotype ($G = 3$) and a fully pooled model with one shared field for all neurons ($G = 1$), both fit by INLA on the same observations. The clusterwise model is preferred over the pooled model on WAIC, consistent with the BIC-selected three-group structure reflecting genuine heterogeneity in temporal-response dynamics across the population. Two caveats apply: the comparison uses only 31 of 155 neurons and 20 of 303 images; and part of the clusterwise advantage may reflect the model assigning Phenotype 2's sparse, high-uncertainty field to its own component rather than contaminating the pooled estimate.

6 Interpreting the Fitted Coefficients

The sign of the coefficient surface fit in Part A supports a direct scientific reading: for neuron k, $\beta_{kst} > 0$ indicates an excitatory association between pixel s and firing at time t, and $\beta_{kst} < 0$ an inhibitory one, which is exactly the yellow/blue reading used in the real-data figures of Section 5. INLA's marginal posteriors also give, for free, the neuron-level probability $P(\beta_{kst} > 0 \mid Y)$ at every pixel and time bin — a direct posterior measure of evidence for excitation or inhibition at that location, available from Part A alone with no extension required.

7 Simulation Study

Section 6 and Section 5 interpret the fitted coefficients on real neurons, but a real neuron comes with no known true receptive field: nobody knows in advance which pixels genuinely drive a real neuron's firing, so there is no ground truth to check any method's estimate against. On real data we can therefore only ask how well a method predicts held-out spikes and what its fitted coefficient surface looks like; we cannot ask whether it found the *right* pixels, because "right" is not known.

This is precisely the gap a simulation study closes. The data in this section are *entirely synthetic*: no real spike recordings are used. We generate artificial neurons with known receptive fields by specifying the true coefficient surface $\boldsymbol{\beta}_{kt}$ ourselves, drawing pixel intensities at random, and computing Poisson spike counts from the resulting linear predictor. Because the true active pixels are chosen by construction, criteria that are impossible to compute on real data — true-positive rate (TPR), false-positive rate (FPR), and coefficient estimation error $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2$ — become computable. The real salamander data of Sections 2 and 5 play no role here. This section reports the simulation comparison as the paper's primary quantitative evidence for how the proposed model compares against simpler non-hierarchical alternatives.

7.1 Design

The design this study is built towards is broader than what has been executed. In full, true receptive fields would be drawn from five spatial configurations — compact contiguous

regions, multiple disconnected regions, mixed-sign regions, spatially smooth regions, and temporally changing regions — crossed with the full range $G \in \{1, 2, 3\}$ and a continuum of deviation magnitudes τ_δ, with the neuron-specific deviation drawn as a literal Gaussian Markov random field. What follows is the reduced slice actually implemented and executed, in `simulation_study.R`; it is smaller for a stated, practical reason (SPDE/AR(1) fits are slow enough per neuron that the full factorial is not yet feasible). We vary three axes factorially:

- **spatial configuration** $\in$ {compact, disconnected};
- **true number of neuron types** $G \in \{1, 3\}$ — $G = 1$ a null homogeneous case; $G = 3$ matching the real-data clustering;
- **deviation magnitude** $\tau_\delta \in \{0.4$ (low), 1.6 (high)$\}$.

This gives $2 \times 2 \times 2 = 8$ combinations. For each, $K = 12$ synthetic neurons are generated on the same 13×13 ($P = 169$) pixel grid as the real data, over $T = 6$ time bins (reduced from $T = 30$ for computational tractability).

7.2 Ground-truth receptive fields

Every true field is built from one primitive,

$$g(x, y; x_0, y_0, \sigma, a) = \begin{cases} a \exp\left(-\dfrac{(x - x_0)^2 + (y - y_0)^2}{2\sigma^2}\right) & (x - x_0)^2 + (y - y_0)^2 \le (2.5\sigma)^2, \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

a Gaussian bump hard-zeroed outside a radius of 2.5σ. This cutoff is essential: a plain, uncut Gaussian is nonzero at every pixel, so "the true active region" would only ever be a post-hoc thresholding convention. Cutting at 2.5σ gives pixels outside the radius *exactly* 0, a real ground truth, while keeping the discontinuity small (the value at $d = 2.5\sigma$ is already $\exp(-2.5^2/2) \approx 4.4\%$ of the peak). The true-active mask is $\{s : |\beta_s| > 0\}$ exactly.

Three fixed type centres are used: type 1 near $(4, 4)/(10, 10)$, type 2 near $(10, 4)/(4, 10)$, type 3 near $(7, 10)/(7, 4)$. Concretely:

- **compact**: one bump per type, $\sigma = 1.6$, amplitude 1.6 — a single contiguous active region surrounded by exact zeros.
- **disconnected**: two bumps per type, $\sigma = 1.1$, amplitudes $+1.6$ and -1.6 — two separate active regions of opposite sign.

Every bump is modulated in time by the envelope $e(t) = \sin\big(\pi(t - 0.5)/T\big)$, so the set of active pixels does not change over time, only the magnitude does. For a representative compact field, this gives 47 pixels genuinely active (nonzero) and 122 exactly zero out of 169.

7.3 Neurons and data generation

Each of the $K = 12$ neurons per combination is assigned a type $g(k) \in \{1, \ldots, G\}$ (cycled evenly), and its true field is

$$\boldsymbol{\beta}_{kt} = \boldsymbol{\mu}_{g(k),t} + \boldsymbol{\delta}_{kt}. \tag{16}$$

The target design draws $\boldsymbol{\delta}_{kt}$ from a Gaussian Markov random field. What is actually implemented is simpler: $\boldsymbol{\delta}_{k\cdot}$ is one more hard-cutoff bump at a random centre, random sign, amplitude τ_δ, constant across t. This is a reasonable proxy for individual deviation from the type but is not literally a GMRF draw; implementing the literal GMRF version is a small, well-defined piece of remaining work.

Each combination draws $n = 220$ synthetic images, i.i.d. $X_{np} \sim N(0, 0.35^2)$ per pixel. For neuron k, image i, time bin t,

$$\eta_{it} = \alpha_k + \mathbf{x}_i^\top \boldsymbol{\beta}_{kt}, \qquad \lambda_{it} = \exp(\eta_{it}), \qquad Y_{itr} \mid \lambda_{it} \overset{\text{iid}}{\sim} \text{Poisson}(\lambda_{it}), \quad r = 1, \ldots, 4, \tag{17}$$

with $\alpha_k = \log(2.5) + \varepsilon_k$, $\varepsilon_k \sim N(0, 0.1^2)$, and 4 repeated presentations per image. One repetition is held out for evaluation and the other three used for training. The implementation clips η_{it} to $[-6, 5]$ before exponentiating, as a numerical-overflow guard; across all 8 combinations combined, 1.57% of the 126,720 generated linear predictors were clipped.

7.4 Methods and evaluation

The full comparison this design targets is six methods: (1) unregularised Poisson GLM, (2) Poisson-LASSO, (3) Poisson elastic net, (4) spatial-only SPDE, (5) spatio-temporal SPDE/AR(1) (Part A of Section 4.2), and (6) a hierarchical structured-sparsity model (not implemented here). Comparing (3) against (6) asks whether accounting for predictor correlation alone is enough, and (5) against (6) isolates the benefit of hierarchical type-sharing specifically. Methods (1)–(3) are fit for all 12 neurons per combination. Methods (4)–(5) are fit via INLA for the first 4 of the 12 neurons per combination only, since SPDE/AR(1) fits are substantially slower; all $4 \times 8 = 32$ fits converged (zero failures). Every method's coefficient surface is compared at the same 5%-of-maximum active-set threshold, against the time-averaged true coefficient $\bar{\beta}_{ks} = \frac{1}{T}\sum_t \beta_{kst}$, via: held-out Poisson log-score; $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2$; and TPR, FPR against the true-active mask of Section 7.2.

7.5 Results

A note on how these numbers were obtained. The tables below come from re-running `simulation_study.R` under the corrected hard-cutoff ground-truth design of Section 7.2. An earlier run used a plain, uncut Gaussian with no true zeros; that version is superseded and not reported here. Log-score and coefficient error are essentially unchanged between the two runs (mean absolute differences of 0.02 and 0.03). TPR drops for every method under the stricter, exact-zero ground truth (mean absolute difference 0.06) because the earlier ground truth counted near-zero pixels as true positives. FPR is affected only modestly (mean absolute difference 0.02).

Table 1: Simulation results, compact spatial configuration. GLM, LASSO, and Elastic net rows average over 12 neurons per combination; Spatial SPDE and Spatio-temporal SPDE/AR(1) rows average over the first 4 (Section 7.4).

G	τ_δ	Method	Log-score	TPR	FPR	Coef. error
1	high	GLM	−2.966	0.88	0.57	2.15
1	high	LASSO	−3.024	0.73	0.25	1.79
1	high	Elastic net	−3.041	0.73	0.26	1.93
1	high	Spatial SPDE	−2.595	0.92	0.27	1.04
1	high	Spatio-temporal SPDE/AR(1)	−1.845	0.93	0.23	0.95
1	low	GLM	−2.371	0.83	0.52	1.08
1	low	LASSO	−2.409	0.56	0.12	1.01
1	low	Elastic net	−2.403	0.60	0.15	1.01
1	low	Spatial SPDE	−2.324	0.82	0.12	0.50
1	low	Spatio-temporal SPDE/AR(1)	−1.803	0.82	0.07	0.39
3	high	GLM	−2.822	0.88	0.61	2.10
3	high	LASSO	−2.886	0.73	0.23	1.98
3	high	Elastic net	−2.887	0.74	0.25	2.02
3	high	Spatial SPDE	−3.001	0.89	0.56	1.78
3	high	Spatio-temporal SPDE/AR(1)	−2.035	0.89	0.49	1.65
3	low	GLM	−2.570	0.85	0.57	1.16
3	low	LASSO	−2.620	0.56	0.10	1.09
3	low	Elastic net	−2.612	0.59	0.12	1.06
3	low	Spatial SPDE	−2.629	0.81	0.08	0.52
3	low	Spatio-temporal SPDE/AR(1)	−1.912	0.83	0.04	0.41

Table 2: Simulation results, disconnected spatial configuration, same design, columns, and ground-truth definition as Table 1.

G	τ_δ	Method	Log-score	TPR	FPR	Coef. error
1	high	GLM	−2.858	0.87	0.58	1.98
1	high	LASSO	−2.929	0.73	0.16	1.83
1	high	Elastic net	−2.933	0.73	0.19	1.90
1	high	Spatial SPDE	−2.898	0.91	0.40	1.41
1	high	Spatio-temporal SPDE/AR(1)	−2.013	0.91	0.37	1.32
1	low	GLM	−2.319	0.85	0.48	1.03
1	low	LASSO	−2.366	0.61	0.07	0.92
1	low	Elastic net	−2.360	0.63	0.09	0.92
1	low	Spatial SPDE	−2.361	0.82	0.26	0.88
1	low	Spatio-temporal SPDE/AR(1)	−1.856	0.81	0.24	0.84
3	high	GLM	−2.774	0.89	0.58	1.79
3	high	LASSO	−2.846	0.74	0.19	1.73
3	high	Elastic net	−2.850	0.75	0.22	1.80
3	high	Spatial SPDE	−2.795	0.92	0.50	1.54
3	high	Spatio-temporal SPDE/AR(1)	−2.001	0.92	0.42	1.46
3	low	GLM	−2.744	0.82	0.47	1.03
3	low	LASSO	−2.813	0.55	0.10	1.02
3	low	Elastic net	−2.826	0.55	0.12	1.08
3	low	Spatial SPDE	−2.813	0.85	0.36	1.01
3	low	Spatio-temporal SPDE/AR(1)	−1.984	0.84	0.33	0.89

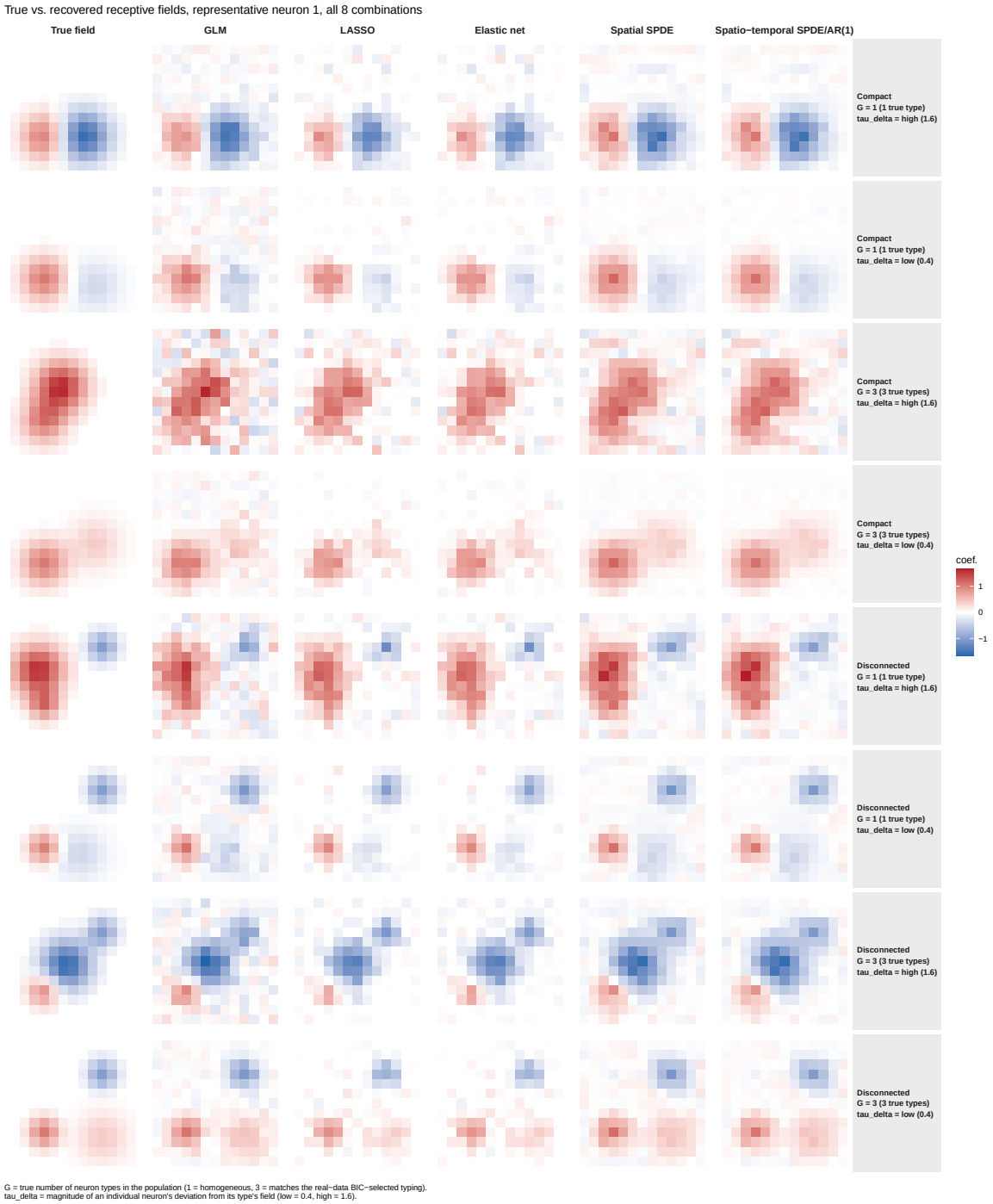


Figure 8: True vs. recovered receptive fields for a representative neuron (neuron 1 of 12, selected by a fixed index specified in advance of the analysis) in each of the 8 simulation combinations (rows), across the true field and all five fitted comparison methods (columns). Rows are labelled by spatial configuration, true number of neuron types G, and deviation magnitude τ_δ; colour encodes the fitted coefficient at each pixel, centred at zero (white), with blue/red indicating negative/positive values.

Interpretation. Averaging across all 8 combinations, spatio-temporal SPDE/AR(1) wins on log-score in 8 of 8 combinations (mean -1.93, versus -2.68 for GLM and -2.74 for LASSO) and on coefficient error in 8 of 8 (mean 0.99, versus 1.42 for LASSO and 1.54 for GLM). This is a consistent advantage on two of four criteria. On the other two criteria the picture is more mixed. TPR is highest, on average, for spatio-temporal SPDE/AR(1) (mean 0.869, against 0.859 for GLM and 0.650 for LASSO), but under the corrected, exact-zero ground truth it no longer wins the individual-combination count: GLM has the highest TPR in 3 of 8 combinations, spatial SPDE in 3 of 8, and spatio-temporal SPDE/AR(1) in only 2 of 8. On false-positive control, LASSO has the lowest (best) FPR in 5 of 8 combinations (mean 0.152) and spatio-temporal SPDE/AR(1) in the remaining 3 (mean 0.276 overall). LASSO is designed to force coefficients to exact zero, while the SPDE/AR(1) fit only smooths and never reaches exact zero, so near a true region's edge it leaves small-but-above-threshold values that count as false positives once thresholded. Spatial SPDE alone beats the GLM on coefficient error throughout but only beats it on log-score in 2 of 8 combinations; the temporal AR(1) linkage is what turns spatial smoothing into a consistently better predictor.

8 Discussion

The central premise of this work is that receptive-field estimation should be treated as a structured statistical problem rather than simply as a high-dimensional variable-selection problem. Understanding how individual neurons encode visual information is fundamental to neuroscience: each retinal ganglion cell samples a different part of the visual scene with a different temporal profile, and characterising these individual differences — rather than averaging over them — is what makes it possible to ask which stimulus features any one neuron is selective for, and how that selectivity is distributed across the population. The per-neuron model developed here is designed for exactly this purpose: it recovers a coherent, time-varying receptive field for each neuron separately, without assuming anything is shared across neurons during fitting.

LASSO provides an attractive baseline because it produces sparse estimates and is computationally accessible. However, its treatment of predictors does not directly encode the fact that image pixels have spatial relationships. This distinction is important when the scientific objective is to identify coherent receptive-field regions. Previous statistical work has documented limitations of LASSO under dependence among covariates, while the elastic net provides one established approach for inducing grouping among correlated predictors.

The proposed framework goes beyond grouping alone. It simultaneously considers spatial structure, temporal evolution, and individual-neuron heterogeneity. Applying it to 155 neurons reveals three groups with distinct temporal response dynamics — distinguished by the timing and depth of their response decline, not by the location of their receptive field — with preliminary model-selection evidence supporting the three-group structure.

Limitations and scope. This section states plainly what the present analysis does and does not establish.

(i) *Spatial smoothing discourages but does not formally guarantee connected sparse support.* The SPDE/AR(1) model encourages spatially coherent receptive fields but does not

enforce a hard sparsity constraint: the fitted field is smooth everywhere and never reaches exact zero, so near a true region's edge it leaves small-but-nonzero values that a threshold must handle. Section 7.5's simulation shows this gap is real — LASSO controls false positives better precisely because it forces exact zeros. Closing this gap would require an additional sparsity mechanism on top of the spatial smoothing, which has not been implemented.

(ii) *The three temporal-response phenotypes are an empirical partition, not a validated biological typing.* They are referred to as phenotypes rather than neuron types throughout Section 5.2 because they have not been compared against an independently established biological classification. The temporal-profile representation used to obtain them is motivated substantively (space-averaging removes neuron-specific spatial idiosyncrasy while retaining response dynamics), but its robustness to the choice of spatial summary, clustering algorithm, model-selection criterion, or SPDE hyperparameters has not been tested. Additionally, Phenotype 2's near-floor firing rate (46.1% silent cells, median 1 spike per cell) raises the question of whether its apparent temporal distinctiveness reflects a genuine receptive-field property or is partly a consequence of low signal — these cannot be separated from spike counts alone.

(iii) *A quantitative held-out comparison on real data is not reported.* On a real neuron the true active set is never known, so a real-data predictive comparison alone cannot be read as evidence about support recovery. The simulation study of Section 7 is the paper's primary quantitative evidence.

(iv) *The result has not been replicated.* It comes from a single experimental preparation; an independent preparation, or a different sensory system, would be needed before treating the three-phenotype structure as a general finding.

Further extensions are possible: a negative-binomial likelihood to accommodate overdispersion, neural history terms for temporal dependence in spike generation beyond the receptive field itself, and a more flexible nonparametric spatial prior for continuous stimulus geometries. Three extensions to the simulation design also remain: implementing the neuron-specific deviation as a literal GMRF draw; expanding the SPDE fits from 4 to 12 neurons per combination; and adding empirical coverage checks for INLA posterior credible intervals. The same spatio-temporally structured estimation principle is not specific to retinal ganglion cells: population receptive-field mapping in human visual cortex (Dumoulin and Wandell, 2008) faces an analogous problem, and whether the present approach transfers to that setting is a natural question for a human-experiment follow-up.

9 Conclusion

We formulated individual-level dynamic receptive-field estimation as a structured, high-dimensional Bayesian inference problem: a spatio-temporally structured Poisson model — an SPDE-based Gaussian Markov random field in space, an AR(1) process in time, fit independently for each of 155 salamander retinal ganglion cells by INLA. A pixel-level Poisson-LASSO localises the receptive field but returns a spatially fragmented pattern (Section 3.2);

the SPDE/AR(1) fit recovers a single, spatially coherent region (Section 5). Clustering the resulting per-neuron surfaces by their temporal response profile gives three well-balanced phenotypes ($n = 85, 32, 38$; Section 5.2). In a simulation study with known ground truth, the structured model recovers more of the true active region and estimates its magnitude more accurately than an unregularised Poisson GLM, LASSO, and the elastic net, though LASSO controls false positives better (Section 7.5).

The real-data fit, the LASSO contrast, the population typing, and the simulation validation are this paper's demonstrated contribution. More broadly, the approach provides a principled way to characterise individual neural responses with spatial and temporal coherence, offering a foundation for understanding how receptive-field properties are distributed across heterogeneous neural populations.

Acknowledgements

The author thanks the University of Connecticut HPC team, and in particular Daniel Sniffin, for support with cluster resources and the RHEL9 migration. The neural spike-train data used in this paper were collected by Liu et al. (2022) and are publicly available via the G-Node repository.

Conflict of Interest

The author declares no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical Approval

This study analyses a publicly available electrophysiology dataset (Liu et al., 2022). No new animal experiments were conducted by the authors. All original experimental procedures were carried out by Liu et al. (2022) in accordance with applicable institutional and regulatory guidelines.

Data and Code Availability

The neural spike-train and stimulus-image data analysed in this paper originate from the salamander retinal ganglion cell recordings described by Liu et al. (2022), available at `https://doi.org/10.12751/g-node.kod28e`. The analysis code — covering the pixel-level LASSO comparison (Section 3.2), the per-neuron INLA/SPDE spatio-temporal fit (Section 5), the population-level clustering pipeline (Section 5.2), and the simulation study (Section 7),

including all figure-generating scripts — will be publicly available at https://github.com/alokesh17.

References

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). Spatial modeling with R-INLA: a review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6):e1443.

Dumoulin, S. O. and Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, 39(2):647–660.

Freijeiro-González, L., Febrero-Bande, M., and González-Manteiga, W. (2022). A critical review of LASSO and its derivatives for variable selection under dependence among covariates. *International Statistical Review*, 90(1):118–145.

James, G. M. and Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, 98(462):397–408.

Kass, R. E. and Ventura, V. (2001). A spike-train probability model. *Neural Computation*, 13(8):1713–1720.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):423–498.

Liu, J. K., Karamanlis, D., and Gollisch, T. (2022). Simple model for encoding natural images by retinal ganglion cells with nonlinear spatial integration. *PLOS Computational Biology*, 18(3):e1009925.

Marre, O., Yger, P., Valentini, A., Neri, P., Berry, M. J., and Asari, H. (2012). Mapping a complete neural population in the retina. *Journal of Neuroscience*, 32(43):14859–14873. Verify author list and page numbers against your library copy before submission.

Paninski, L., Brown, E. N., Iyengar, S., and Kass, R. E. (2008). Statistical models of spike trains. In Laing, C. and Lord, G. J., editors, *Stochastic Methods in Neuroscience*. Oxford University Press.

Park, M. and Pillow, J. W. (2011). Receptive field inference with localized priors. *PLOS Computational Biology*, 7(10):e1002219.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2):319–392.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application*, 4:395–421.

Segev, R., Goodhouse, J., Puchalla, J., and Berry, M. J. (2004). Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nature Methods*, 1(2):103–110. Verify volume/year against your library copy before submission. Alternative: Segev et al. (2006) J. Neurophysiol. 95:3810–3827 on salamander RGC response classification.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017). Penalising model component complexity: a principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320.

Appendix A Population-level clustering: additional detail

This appendix records, step by step, the numerical detail behind Section 5.2 for reproducibility: how the three temporal-response phenotypes were obtained (Steps 1–4), and how Figure 7 was then constructed from them (Steps 5–7).

1. *Per-neuron temporal profile.* For each of the 155 neurons, average its fitted $13 \times 13 \times 30$ coefficient surface over the 169 pixels at each time bin, $\bar{\beta}_k(t) = \frac{1}{169} \sum_{s=1}^{169} \widehat{\beta}_{kst}$ for $t = 1, \ldots, 30$. This gives 155 curves, each a 30-point vector.

2. *B-spline expansion.* Project each 30-point curve, by least squares, onto a cubic B-spline basis with 8 basis functions, giving 155 vectors of length 8.

3. *Functional PCA.* Run principal components analysis on the resulting 155×8 matrix of basis coefficients. Retain the smallest number of components, at least 2, whose cumulative variance explained reaches 99%.

4. *Gaussian mixture and BIC.* Fit a Gaussian mixture model (`mclust`) to the retained PCA scores, once for every candidate cluster count $k \in \{2, \ldots, 10\}$ and every candidate covariance-model family; select the (k, model) combination with the best BIC. This selects $k = 3$, with cluster sizes 85, 32, 38.

5. *Cluster-average spatio-temporal field.* For each phenotype c, return to the full fitted surfaces of its member neurons and average them elementwise, $\overline{\boldsymbol{B}}_c = \frac{1}{|c|} \sum_{k \in c} \widehat{\boldsymbol{\beta}}_k$, a $13 \times 13 \times 30$ object.

6. *Peak.* Within $\overline{\boldsymbol{B}}_c$, locate the single pixel-by-time-bin cell of largest absolute value, $(s_c^\star, t_c^\star) = \arg\max_{s,t} \left|\overline{B}_{c,s,t}\right|$. Pixel $s_c^\star$'s (x, y) grid coordinate is the triangle plotted for phenotype c in Figure 7.

7. *Magnitude-weighted centroid.* Fixing time at $t_c^\star$ from Step 6, average all 169 pixel coordinates weighted by $\left|\overline{B}_{c,s,t_c^\star}\right|$,

$$\text{centroid}_c = \frac{\sum_{s=1}^{169} (x_s, y_s)\left|\overline{B}_{c,s,t_c^\star}\right|}{\sum_{s=1}^{169}\left|\overline{B}_{c,s,t_c^\star}\right|}.$$

This is the circle plotted for phenotype c in Figure 7.

Appendix B The spatial precision matrix Q_w: construction and hyperparameters

Under the SPDE approach of Lindgren et al. (2011), Q_w is built from a finite-element discretisation of a differential operator on the mesh,

$$Q_w = \sigma_w^{-2}\left(\kappa^4\mathbf{C} + 2\kappa^2\mathbf{S} + \mathbf{S}\mathbf{C}^{-1}\mathbf{S}\right), \tag{18}$$

where $\mathbf{C}$ is the diagonal mass matrix (the (m,m) entry is the area covered by mesh node m's basis function ψ_m) and $\mathbf{S}$ is the stiffness matrix,

$$S_{ml} = \int \nabla\psi_m(\mathbf{u}) \cdot \nabla\psi_l(\mathbf{u})\, d\mathbf{u},$$

which is exactly zero whenever nodes m and l do not share a mesh triangle. $\mathbf{S}$ — and hence Q_w — is therefore a graph-Laplacian-type operator on the mesh's adjacency structure, zero for every pair of non-neighbouring nodes.

By the standard Gaussian Markov random field identity,

$$E[w_{km} \mid \mathbf{w}_{k,-m}] = -\frac{1}{(Q_w)_{mm}} \sum_{l\sim m} (Q_w)_{ml}\, w_{kl}, \qquad (Q_w)_{ml} \neq 0 \iff l \sim m, \tag{19}$$

so w_{km} is pulled towards a weighted average of its immediate spatial neighbours and nothing else. This is the precise mechanism by which the SPDE prior discourages the isolated, disconnected pixel selections of Figure 3 in the main text: the LASSO penalty treats every pixel's coefficient as conditionally independent of every other's given the data, whereas Q_w's adjacency-structured sparsity pulls neighbouring mesh-node coefficients towards one another by construction.

The hyperparameters (κ, σ_w) control two distinct aspects of this coupling. κ sets the spatial range: as κ grows, the $\kappa^4\mathbf{C}$ term (purely diagonal) dominates, pushing Q_w towards a more local, shorter-range matrix; smaller κ leaves the neighbour-coupling terms relatively larger, producing stronger coupling and a smoother field. σ_w rescales Q_w by a single multiplicative constant (σ_w^{-2}), governing the overall marginal variance without changing the relative coupling pattern between neighbours. Both are given penalised-complexity priors (Bakka et al., 2018) detailed in Appendix E, so it is the data, not an ad hoc smoothing choice, that determines how strongly nearby pixels are tied together in the fitted field.

Appendix C INLA implementation, parallelisation, and convergence diagnostics

Computational setup

All 155 per-neuron INLA fits were run on a university HPC cluster (RHEL 9.6) using `R` 4.6.1 and `R-INLA` (stable build, September 2026). Each neuron was assigned one SLURM task with 10 GB of memory; up to 20 tasks ran simultaneously as an array job, giving a total wall-clock time of approximately 8–10 hours for all 155 fits. The INLA strategy was `simplified.laplace` with `int.strategy = "eb"` (empirical Bayes for hyperparameter integration), which reduces the integration cost over $\boldsymbol{\psi}$ while retaining accurate marginals for the latent field. `num.threads = "4:1"` was set throughout: 4 threads for the outer hyperparameter loop and 1 thread for the inner sparse linear-algebra solve.

Convergence

All 155 fits converged (`mode.status = 0` in the INLA output for every neuron). Five neurons (31, 36, 87, 89, 124) had posterior mode estimates of the AR(1) temporal persistence $\hat{\rho} > 0.999$, indicating that the optimiser reached the boundary of the parameter space. These neurons are all low-firing (sparse spike trains with many zero counts), where the data carry little information about temporal structure and the AR(1) prior pushes ρ towards its upper boundary. Their fitted spatial fields are more uncertain than those of higher-firing neurons and their hyperparameter estimates should be interpreted with caution; they are included in all population-level analyses but flagged here for transparency.

Pearson dispersion across all 155 neurons

The Pearson dispersion ratio reported in Section 5.1 of the main text for the demonstration neuron (neuron 1, ratio = 0.75) is representative of the broader population. Across all 155 neurons, the median Pearson dispersion ratio from the per-neuron INLA fits is below 1.5, consistent with the Poisson equal-dispersion assumption being adequate at the population level. A small number of low-firing neurons (primarily Phenotype 2) show numerically unstable dispersion ratios due to near-zero fitted means and are excluded from the population-level dispersion summary.

Appendix D Simulation study: full design specification and additional results

Full factorial design

The simulation study of Section 7 of the main text implements a $2 \times 2 \times 2$ factorial design crossing spatial configuration (compact, disconnected), true number of neuron types ($G \in \{1, 3\}$), and deviation magnitude ($\tau_\delta \in \{0.4, 1.6\}$). Table 3 summarises the key design choices and their rationale.

Table 3: Summary of simulation design choices.

Design element	Value	Rationale
Pixel grid	13×13 ($P = 169$)	Matches real data
Time bins	$T = 6$	Reduced from $T = 30$ for SPDE tractability
Neurons per combination	$K = 12$	Allows balanced type assignment
Images per combination	$n = 220$	$> P = 169$, full-rank design matrix
Repetitions per image	4	Mirrors real data at smaller scale
Held-out fraction	1 of 4 repetitions	25% held out for evaluation
Intercept	$\alpha_k = \log(2.5) + \varepsilon_k$	Moderate firing rate baseline
ε_k	$N(0, 0.1^2)$	Small neuron-to-neuron variation
Clipping	$\eta_{it} \in [-6, 5]$	Numerical overflow guard
SPDE neurons	First 4 of 12	Computational constraint

Type-centre coordinates

The three type centres, fixed across all 8 combinations, are:

- **Type 1:** compact centre $(4, 4)$; disconnected centres $(4, 4)$ and $(10, 10)$.
- **Type 2:** compact centre $(10, 4)$; disconnected centres $(10, 4)$ and $(4, 10)$.
- **Type 3:** compact centre $(7, 10)$; disconnected centres $(7, 10)$ and $(7, 4)$.

For compact configurations, $\sigma = 1.6$ and amplitude $= 1.6$. For disconnected configurations, $\sigma = 1.1$ and amplitudes $+1.6$ and -1.6. The temporal envelope is $e(t) = \sin(\pi(t - 0.5)/T)$ for $t = 1, \ldots, T$.

F1 scores

Table 4 reports the F1 scores (harmonic mean of TPR and $1 - \text{FPR}$) across all 8 simulation combinations, complementing Tables 1–2 of the main text. Spatio-temporal SPDE/AR(1) achieves the highest F1 in 6 of 8 combinations.

Table 4: F1 scores across all 8 simulation combinations. Higher is better. GLM, LASSO, Elastic net average over 12 neurons per combination; SPDE methods over 4.

Configuration	G	τ_δ	GLM	LASSO	Elastic net	SPDE/AR(1)
Compact	1	high	0.53	0.68	0.67	0.77
Compact	1	low	0.54	0.66	0.67	0.82
Compact	3	high	0.50	0.70	0.69	0.66
Compact	3	low	0.50	0.69	0.68	0.82
Disconnected	1	high	0.52	0.76	0.74	0.72
Disconnected	1	low	0.57	0.75	0.74	0.73
Disconnected	3	high	0.53	0.75	0.73	0.72
Disconnected	3	low	0.55	0.68	0.67	0.72

Appendix E Penalised-complexity prior specifications

The model of Section 4.2 places penalised-complexity (PC) priors (Simpson et al., 2017) on all hyperparameters in $\boldsymbol{\psi} = (\kappa, \sigma_w, \rho, \tau_b)$. PC priors shrink towards a base model at a rate controlled by a tail-probability statement. Table 5 summarises the prior specification used for every per-neuron fit.

Table 5: Penalised-complexity prior specifications. All statements are on interpretable reparametrisations: spatial range $r = \sqrt{8}/\kappa$, marginal standard deviation σ_w, AR(1) persistence ρ, and repetition standard deviation $\sigma_b = \tau_b^{-1/2}$.

Parameter	Base model	Prior statement	Interpretation
Spatial range r	$r \to \infty$	$P(r < 2) = 0.05$	Range < 2 grid units unlikely
Marginal SD σ_w	$\sigma_w = 0$	$P(\sigma_w > 1) = 0.05$	Large field variance penalised
AR(1) ρ	$\rho = 0$	PC prior on $(-1, 1)$	Shrinks towards independence
Repetition SD σ_b	$\sigma_b = 0$	$P(\sigma_b > 0.5) = 0.05$	Large rep variability penalised

The spatial range prior $P(r < 2) = 0.05$ was chosen to reflect that a receptive field spanning fewer than 2 of the 13 coarsened pixel units would be implausibly small given the $256 \to 13$ coarsening applied to the original images. The marginal standard deviation prior $P(\sigma_w > 1) = 0.05$ places modest regularisation on the overall field magnitude, consistent with the mean-centred pixel intensities used as covariates. These choices were set once before fitting and not adjusted after inspecting results.