Hierarchical Bayesian modeling of multi-region brain cell count data

Sydney Dimmock[a], Benjamin M. S. Exley[b], Gerald Moore[c], Lucy Menage[d], Alessio Delogu[d], Simon R. Schultz[c], E. Clea Warburton[b], Conor Houghton[a], and Cian O'Donnell[a,e]

We can now collect cell-count data across whole animal brains quantifying recent neuronal activity, gene expression, or anatomical connectivity. This is a powerful approach since it is a multi-region measurement, but because the imaging is done post-mortem, each animal only provides one set of counts. Experiments are expensive and since cells are counted by imaging and aligning a large number of brain sections, they are time-intensive. The resulting datasets tend to be under-sampled with fewer animals than brain regions. As a consequence, these data are a challenge for traditional statistical approaches. We demonstrate that hierarchical Bayesian methods are well suited to these data by presenting a 'standard' partially-pooled Bayesian model for multi-region cell-count data and applying it to two example datasets. For both datasets the Bayesian model outperformed standard parallel t-tests. Overall, the Bayesian approach's ability to capture nested data and its rigorous handling of uncertainty in under-sampled data can substantially improve inference for cell-count data.

Bayesian analysis | Cell-counts | Hierarchical modeling | Neuroscience data

Significance Statement

Cell-count data is important for studying neuronal activation and gene expression relating to the complex processes in the brain. However, the difficulty and expense of data collection means that such datasets often have small sample sizes. Many routine analyses are not well-suited, especially if there is high variability among animals and surprising outliers in the data. Here we describe a multilevel, mixed effects Bayesian model for these data and show that the Bayesian approach improves inferences compared to the usual approach for two different cell-count datasets with different data characteristics.

In studying the brain we are often confronted with phenomena that involve specific subsets of neurons distributed across many brain regions. Computations, for example, are performed by neuronal networks connecting cells in different parts of the brain and, in development, neurons in different anatomical regions of the brain share the same lineage. An example of each of these types will be considered here, but the challenge is very general: how to measure and analyze multi-region neuronal data with cellular resolution.

In a typical cell-count experiment, gene expression is used to tag the specific cells of interest with a targeted indicator (1). The brain is sliced, an entire stack of brain sections from a single animal is imaged, aligning and registering the images to a standardised brain atlas such as the Allen mouse atlas (2–5), segmenting the images into anatomical regions and counting the labeled cells in each region. The resulting dataset consists of labeled cell counts across each of ∼10–100 brain regions. This technology is being deployed to address questions in a broad range of neuroscience subfields, for example: memory (6–8), neurodegenerative disorders (9), social behavior (10), and stress (11).

Cell-counts are often compared across groups of animals which differ by an experimental condition such as drug treatment, genotype or behavioural manipulation. However, the expense and difficulty of the experiment means that the number of animals in each group is often small, ten is a typical number but fewer than ten is not uncommon. This means that these data are undersampled: the dimensionality of the data, which corresponds to the number of brain regions, is much larger than the number of samples, which usually corresponds to the number of animals (Figure 1**A**). Current statistical methods are not well-suited to these nested wide-but-shallow datasets. Furthermore, due to the complicated preparation and imaging procedure, there is often missing data and there is variability derived from experimental artifacts.

In cell count data there are two obvious sources of noise. The first of these is easy to describe, if a region has an rate that determines how likely a cell is to be marked for counting, then the actual number of marked cells is sampled from a Poisson distribution. The second source of noise is the animal to animal variability of the rate itself and this depends on diverse features of the individual animal and the experiment that are often unrelated to the phenomenon of interest. The challenge is to control for outliers and 'poor' data points whose rate is noisy, while extracting as much information as possible about the underlying process. Dealing with outliers is often an opaque and *ad hoc* procedure, it is also a binary decision, a point is either excluded so it does not contribute, or included, noise and all. This

Author affiliations: [a]School of Engineering Mathematics and Technology, University of Bristol, Michael Ventris Building, Woodland Road, Bristol, BS8 1UB, United Kingdom; [b]School of Physiology, Pharmacology and Neuroscience, University of Bristol, Biomedical Sciences Building, University Walk, Bristol, BS8 1TD, United Kingdom; [c]Centre for Neurotechnology and Dept of Bioengineering, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom; [d]Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 9NU, United Kingdom; [e]School of Computing, Engineering & Intelligent Systems, Ulster University, Derry/Londonderry, BT48 7JL, Northern Ireland, United Kingdom

No competing interests.

[1]E-mail: sd14814bristol.ac.uk

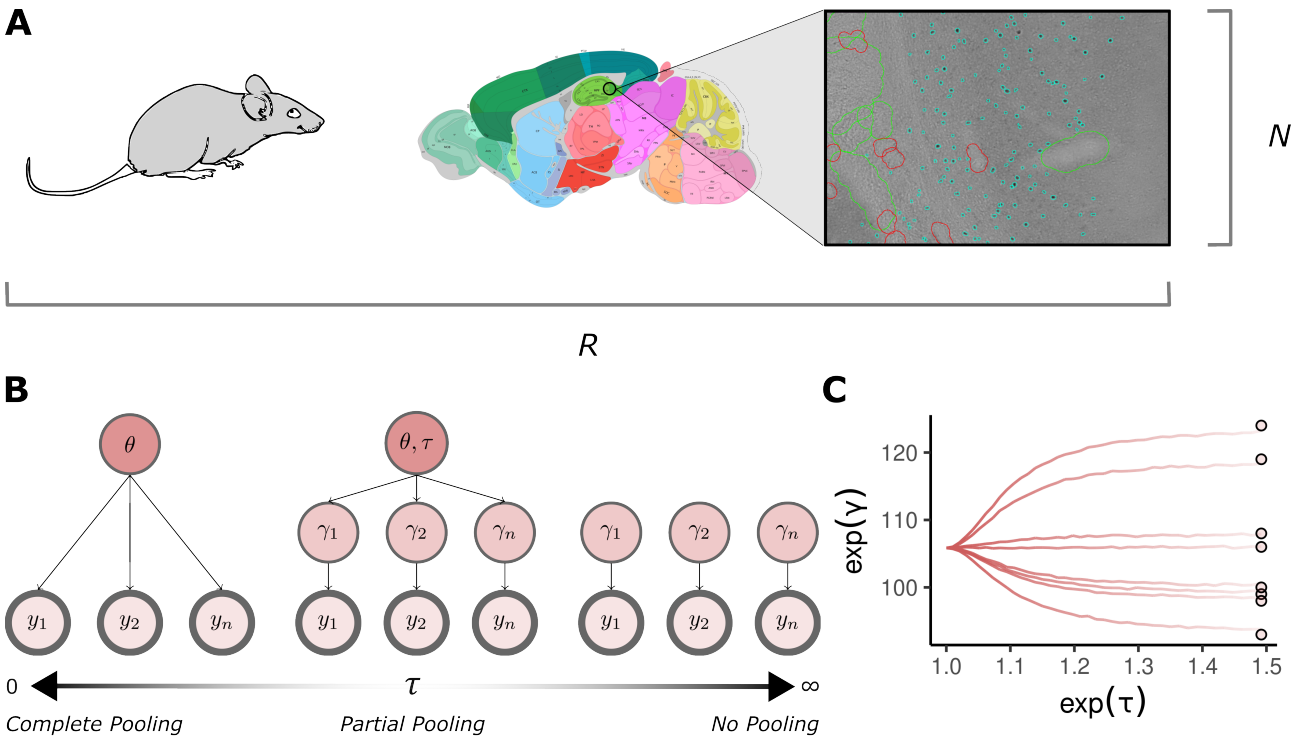

Fig. 1. Introduction. **A**: Each of N animals produce a cell-count from a total of R brain regions of interest. Cell-count data is typically undersampled with $N \ll R$. Scientists analyse the brain sections from experiment for positive signals. Here an example section is shown where teal points mark cells expressing the immediate early gene *c-Fos*. The final cell-count is equal to the sum of these individual items (sagittal brain map taken from the Allen mouse brain atlas: https://mouse.brain-map.org. **B**: Partial pooling is a hierarchical structure that jointly models observations from some shared population distribution. It is a spectrum that depends on the value of the population variance τ. When $\tau = 0$ there is no variation in the population and each individual observation is modeled as a conditionally independent estimate of some fixed population mean θ (complete pooling). As τ tends to infinity observations do not combine inferential strength but inform an independent estimate γ_i (no pooling). In between the two extremes combine. Each observation can contribute to the population estimate while simultaneously supporting a local one to effectively model the variance in the data. Observed quantities have been highlighted with a thicker stroke in the graphical model. **C**: An example of partial pooling on simulated count data. As the population standard deviation increases on the x-axis the individual estimates $\exp(\gamma_i)$ trace a path from a completely pooled estimate to an unpooled estimate. Circular points give the raw data values. Parameters are exponentiated because the outcomes are Poisson and so parameters are fit on the $\log$ scale.

is where partial pooling helps. Partial pooling allows for the simultaneous estimation of individual parameters with the population distribution that describes them. This helps the data to self-regularise and elegantly balances the contribution of informative and weak observations to parameter values.

In recent years Bayesian approaches to data analysis have become powerful alternatives to classical frequentist approaches (12–14), including for some types of neuroscience data: such as neurolinguistics (15), neural coding (16, 17), synaptic parameters (18–21) and neuronal-circuit connectivity (22, 23). A Bayesian approach is particularly well suited to cell-count data but has not previously been applied to this problem.

A Bayesian approach formalizes the process of scientific inference, it distinguishes the data and a probabilistic mathematical model of the data. This model has a likelihood which gives the probability of the observed data for a given set of model parameters. The model often has a hierarchical structure which we compose to reflect the structure of the experiment and the investigators hypothesis of how the data depends on experimental condition. This hierarchy determines a set of *a priori* probabilities for the parameter values. The result of Bayesian inference is a probability distribution for these model parameters given the data, termed the posterior.

There are three advantages of a Bayesian approach that we want to emphasis, (1) while traditional multi-level models also allow a hierarchy (24), Bayesian models are more flexible and the role of the model is clearer, (2) since the result of Bayesian inference is a probability distribution over model parameters, it indicates not just the fitted value of a parameter but the uncertainty of the parameter value. Finally (3) Bayesian models tend to make more efficient use of data and therefore improving statistical power.

Here, our aim is to introduce a 'standard' Bayesian model for cell count data. We illustrate the application of this model to two datasets, one related to neural activation and the other to developmental lineage. In both cases the Bayesian model produces clearer results than the classical frequentist approach.

Materials and Methods

Data. To illustrate our approach we consider two example applications, one which counts cells active in regions of the recognition memory circuit of rat during a familiarity discrimination task, and the other which examines the distribution of a specific interneuron type in the mouse thalamus.

Case study one - transient neural activity in the recognition memory circuit. The recognition memory network, Figure 3, is a

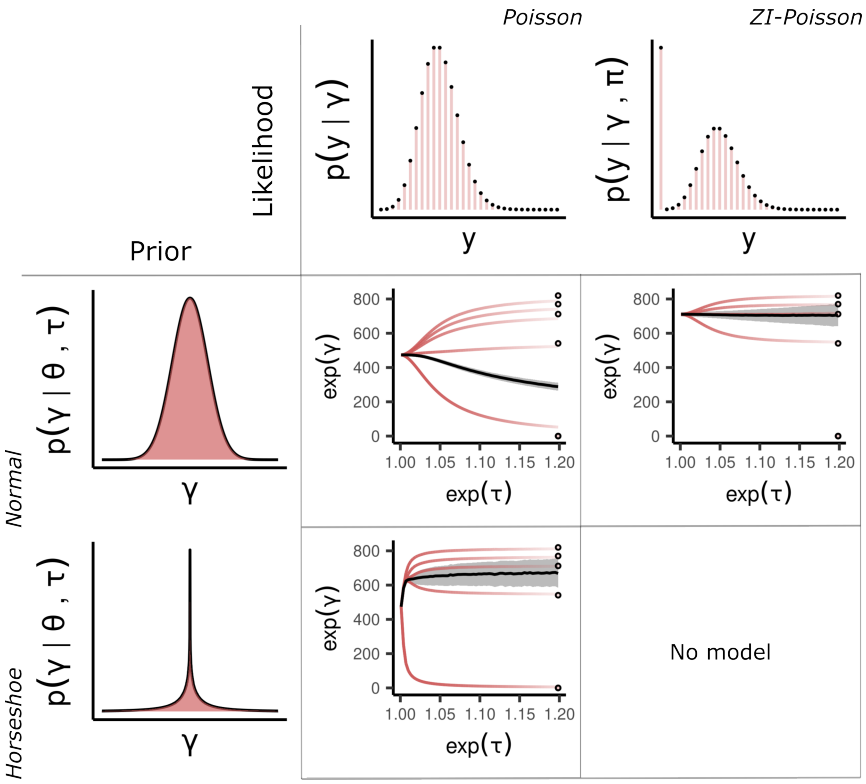

Fig. 2. Methods. A table of partial pooling behaviour for different likelihood and prior combinations. Rows are the two prior choices for the population distribution, and columns the two distributions for the data. Within each cell the expectation of the marginal posterior $p(\exp(\gamma_i)|\theta, \tau, y)$ is plotted as a function of τ. The solid black line is the expectation of the marginal posterior $p(\theta|\tau, y)$ with one standard deviation highlighted in grey. **Top left**: Combining a normal prior for the population with a Poisson likelihood is unsatisfactory in the presence of a zero observation. The zero observations influence the population mean in an extreme way owing to their high importance under the Poisson likelihood. **Bottom left**: By changing to a horseshoe prior the problematic zero observations can escape the regularisation machinery. However, regularisation of the estimates with positive observations is much less impactful. **Top right**: A zero-inflated Poisson likelihood accounts for the zero observations with an added process, reducing the burden on the population estimate to compromise between extreme values. **Bottom right**: No model.

distributed network which has been well studied using a variety of behavioural tasks. It includes the hippocampus (HPC) and perirhinal cortex (PRH), shown to deal with object spatial recognition and familiarity discrimination respectively (25–27); medial prefrontal cortex (MPFC), concerned with executive functions such as decision making but also with working memory, and the temporal association cortex (TE2) used for acquisition and retrieval of long-term object-recognition memories (28). The nucleus reuniens (NRe) is also believed to be an important component of the circuit (25, 29), owing to its reciprocal connectivity with both MPFC and HPC (30). In previous studies, lesions of the NRe have been shown to significantly impair long-term but not short-term object-in-place recognition memory (29).

The data analysed in this case study were collected to investigate the role of the NRe in the recognition memory circuit through contrasting the neural activation for animals with a lesion in the NRe with neural activation for animals with a sham surgery. The immediate early gene *c-Fos* is rapidly expressed following strong neural activation and is useful as a marker of transient neural activity. Animals in the experiment performed a familiarity discrimination task (single-item recognition memory), discriminating novel or familiar objects with or without an NRe lesion and the number of cells that expressed *c-Fos* was counted in regions across the recognition memory circuit. The two-by-two experimental design allocated ten animals to each of the four experiment groups $\{\text{sham}, \text{lesion}\} \times \{\text{novel}, \text{familiar}\}$ and cell counts were recorded from a total of 23 brain regions. The visual cortex V2C and the motor cortex M2C were taken as control regions as they were not expected to show differential *c-Fos* expression in response to novel or familiar objects (31).

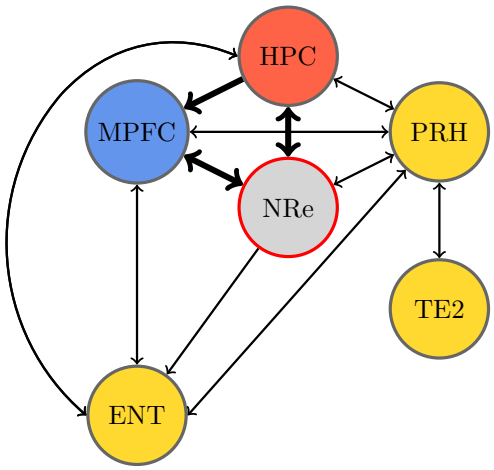

Fig. 3. Recognition memory circuit. Schematic of the recognition memory network adapted from (31). Bold arrows show the assumed two-way connection between the medial prefrontal cortex and the hippocampus facilitated by the NRe. Colours highlight the HPC (red), MPC (blue) and specific areas of the rhinal cortex (yellow). The NRe was lesioned in the experiment.

Case study two - Ontogeny of inhibitory neurons in mouse thalamus and hypothalamus. The second dataset comes from a study, (32), that, in part, counted the number of inhibitory interneurons in the thalamocortical regions of mouse. *Sox14* is a gene associated with inhibitory neurons in subcortical areas. It is required for the development and migration of local inhibitory interneurons in the dorsal lateral geniculate nucleus (LGd) of the thalamus (32, 33). Consequently, *Sox14* is useful for identifying discrete neuronal populations in the thalamus and hypothalamus (or in the diencephalon) (32–34). A heterozygous (HET) knock-in mouse line $Sox14^{GFP/+}$

to mark *Sox14* expressing neurons with green fluorescent protein (GFP), was compared to a homozygous knock-out (KO) mouse line $Sox14^{GFP/GFP}$ engineered to block expression of the endogenous *Sox14* coding sequence (35). Each animal produced two samples, one for each hemisphere giving a total of ten data points, six belonging to HET (three animals), and four to KO (two animals). Each observation is 50 dimensional, corresponding to fifty individual brain regions in each hemisphere.

Hierarchical modeling. Our goal in both cases is to quantify group differences in the data. We start with a 'standard' hierarchical model; this model can be further refined or changed to improve the analysis and to more clearly match the structure of a particular experiment, we will give an example of this for our second dataset. At the bottom of the model are the data themselves, the cell counts y_i. The index i runs over the full set of samples, in this case $23 \times 10 \times 4$ datapoints in the first study, and $50 \times 6 + 50 \times 4$ in the second. The basic assumption the model makes is that this count is derived from an underlying propensity, $\lambda_i > 0$, which depends on brain region and, potentially, group:

$$y_i \sim \text{Poisson}(\lambda_i) \quad [1]$$

Hence, in this case, the propensity λ_i is the mean of the Poisson distribution, and the statistical model describes the dependence of this parameter on brain region and animal. Since λ_i is strictly positive a log-link function is used:

$$\log \lambda_i = \theta_{r[i],g[i]} + \gamma_i + E_i \quad [2]$$

where

$$\gamma_i \sim \text{Normal}(0, \tau_{r[i],g[i]}) \quad [3]$$

and we have used 'array notation' (36), mapping the sample index i to properties of the sample, so $r[i]$ returns the region index of observation i, and similar for $g[i]$ but for groups and animals.

We use partial pooling. Thus, ignoring E_i for now, the rate has been split between two terms, $\theta_{r[i],g[i]}$ is the *fixed effect* which is constant across animals and $\gamma_{r[i],a[i]}$ is the *random effect* which captures the animal to animal variability. While θ_{rg} models the mean for log-cell-count for each region, given the condition; γ_i models variation around this mean. For this reason, γ_i is assumed to follow a normal distribution with zero mean. The regression term may appear over-parameterized, without θ_{rg} the γ_i could 'do the work' of matching the data. However, the model is regularized by a prior; observations with a weak likelihood will have their random effect γ_i shrunk towards the population location. The amount of regularization depends on the variation in the population, a quantity that is estimated from each likelihood. This is how partial pooling works as an adaptive prior for 'similar' parameters (Figure 1**B**). The data 'pools' some evidence while still allowing for individual differences in sample.

The final term is the exposure E_i. Cell counts may be recorded from sections with different areas. The exposure term scales the parameters in the linear model as the recording area increases (13). In our model, the exposure is equal to the logarithm of the recording area; this value is available as part of the experimental data.

The set of parameters $\tau_{r,g}$ models the population standard deviations of the noise for each region r and animal group g.

Table 1. Parameter table for the hierarchical model.

Parameter	Description
E_i	Exposure.
κ_i	Horseshoe inflation.
π	Zero inflation.
γ_i	Random effect for observation i.
θ_{rg}	Fixed effect for region r in group g.
τ_{rg}	Scale of random effects for region r in group g.

When working on the log scale, priors for these parameters are typically derived in terms of multiplicative increases. Since the parameters are positive they are assigned a half-normal distribution

$$\tau_{r,g} \sim \text{HalfNormal}(\log(s)) \quad [4]$$

with an appropriately chosen scale $s > 1$. For our analyses we used $s = 1.05$ because this gives a HalfNormal distribution with 95% of its density in the interval $[0, \log(1.1)]$. This translates into an approximate 10% variation around $\exp(\theta)$ at the upper end, which is a moderately informative prior, reflecting our belief that within-group animal variability is small relative to between-group variability. This regularization also helps model inference when the datasets are undersampled.

Horseshoe prior. Cell count data often has outliers, for example due to experimental artifacts. Since by default the likelihood does not account for these outliers, they may cause substantial changes in fitted parameter values. This is demonstrated in Figure 2, where a careless application of the Poisson distribution on data with several zero counts has a large influence on the posterior distribution. There are two general options for dealing with outliers: either modeling them in the likelihood or in the prior. Although the likelihood option is preferred as it is more direct, see our zero-inflation model below, it can be hard to design because it requires knowledge of the outlier generation process. The alternative is via a flexible prior such as the horseshoe (37, 38). This more generic option may be suitable as a default approach if the outliers are poorly understood.

The horseshoe prior is a hierarchical prior for sparsity. It introduces an auxiliary parameter κ_i that multiplies the population scale τ. This construction allows surprising observations far from the bulk of the population density to escape regularisation.

$$\gamma_i \sim \text{Normal}(0, \tau_{r[i],g[i]} \times \kappa_i) \quad [5]$$
$$\kappa_i \sim \text{HalfNormal}(1). \quad [6]$$

An example of this is given in Figure 2 as the bottom left cell of the 2×2 table of models. The horseshoe prior often uses a Cauchy distribution but in our case the heavy tail causes problems for the sampling algorithm, see SI section 8.

Zero inflation. A particular trait of the second dataset is that there are a large number of zero data points ($\sim 6\%$). Although a zero observation is always possible for a Poisson distribution, for plausible values of the propensity, zeros should be rare. It is likely that for some regions the experiment has not worked as expected and the zeros show that something has 'gone wrong' and that the readings are not well described by a Poisson distribution. Here we extend the model to include this possibility. This is a useful elaboration of the

'standard model', in the standard model the horseshoe prior ensures that these anomalous readings only have a small effect on the result but it is more informative to extend the model to include them. While this particular extension is specific to these data, it also serves as an example of how a standard Bayesian model can serve as a starting for point for an investigation of the data.

The zero-inflated Poisson model is intended to model a situation were there are zeros unrelated to the Poisson distribution, in this case this might, for example, be the result of an error in the automated registration process that identifies regions and counts their cells. It is a mixture model, if

$$y_i \sim \text{ZIPoisson}(\pi, \lambda_i) \quad [7]$$

there is a probability π that $y_i = 0$ and a $1-\pi$ probability that y_i follows a Poisson distribution with rate λ_i. Importantly this means there are two ways in which y_i can be zero, through the Bernoulli process parameterized by π, or through the Poisson distribution. This has the effect of 'inflating' the probability mass at zero with the additional parameter π giving the proportion of extra zeros in the data that could not be explained by the standard Poisson distribution. This distribution can be visualized in Figure 2 and further mathematical details are described in SI 2C.

Model inference. Posterior inference was performed with the probabilistic programming language Stan (39), using its custom implementation of the No-U-Turn (NUTS) sampler (40, 41). For each model, the posterior was sampled using four chains for 8000 iterations, with half of these being attributed to the warm-up phase; giving a total of 16000 samples from the posterior distribution.

Results

We describe differences in estimated counts between groups in terms of $\log_2$-fold changes. Fold changes are useful because they prevent differences that are small in absolute magnitude from being masked by regions with high overall expression. Our results compare Bayesian highest density intervals (HDI) with the confidence interval (CI) from an uncorrected Welch's t-test. The Bayesian HDI is calculated from the posterior distribution and is the smallest width interval that includes a chosen probability, here 0.95 (to correspond to $\alpha = 0.05$), and summarizes the meaningful uncertainty over a parameter of interest.

Case study one - transient neural activity in the recognition memory circuit. Results for the first dataset are presented in Figure 4 with Figure 4**A** plotting cell-count differences between the novel and familiar conditions without lesion and Figure 4**B** with lesion. These data were collected to investigate the role of different hippocampal and adjacent cortical regions in memory. However, some regions of interest, such as the intermediate dendate gyrus (IDG), and the dorsal subiculum (DSUB) look under-powered: in the sham animals, for both regions there is a markedly non-zero difference in expression between the novel and familiar conditions but a wide confidence interval overlapping zero makes the evidence unreliable (orange bars Figure 4**A**).

In contrast, the Bayesian estimates (green bars Figure 4) produce a clear result. For a number of brain regions in Figure 4**A**, sham-novel animals have higher expression than sham-familiar ones. These differences disappear in Figure 4**B** with lesion-novel and lesion-familiar animals showing roughly equal cell counts. This indicates that the difference is only present when the NRe is intact.

Case study two - Ontogeny of inhibitory interneurons of the mouse thalamus. The estimated $\log_2$-fold difference in GFP expressing cells between the two genotypes, for each of the fifty brain regions, are plotted in Figure 5. This include the purple and pink 95% HDI from the horseshoe and zero-inflated Poisson models along with the 95% confidence interval arising from a t-test in orange.

The use of the zero-inflated Poisson distribution uncovers some interesting differences when compared to the t-test confidence intervals. One example of this is the result for tuberomammillary nucleus, dorsal part (TMd). For this region the HET animals have high GFP expression across both hemispheres, yet animal three has a reading of zero for both hemispheres. This injects variability into the standard deviation of the HET group. Consequently, the pooled standard deviation used in the t-test is large and almost certainly guarantees a non-significant result. Furthermore, the sample mean of this region looks nothing like zero, but also nothing like the other two animals with positive counts. In addition to the wide interval, the sign of the difference does not agree with the data. Similarly, the medial preoptic nucleus (MPN) also suffers from poor estimation. Once again, this region contains a single HET animal for which the reading from both hemispheres is zero. The zero-inflated Poisson produces a posterior distribution of the appropriate sign with small uncertainty.

Despite the large difference in interval estimation between the HDI and CI for many brain regions, as the data becomes stronger from the perspective of the frequentist p-value towards the right-hand side of the second row in Figure 5, they become much more compatible. The variation within groups is very small for these regions; further regularization is not necessary and so the impact of partial pooling has been reduced. The sample estimate of the t-test has 'caught up' to the regularized estimate because the signal is strong.

Discussion

We have presented a standard workflow for Bayesian analysis of multi-region cell-count data. We propose a likelihood and appropriate priors with a nested hierarchical structure reflecting the structure of the experiment. We applied this to two distinct example datasets and demonstrate that they capture more fruitfully the characteristics of the data when compared to field-standard frequentist analyses.

For both case studies, the Bayesian uncertainty intervals are more precise than the confidence intervals. These confidence intervals tend to be quite wide on these data because of the small sample size and because of violations to their parametric model assumptions. Our workflow uses a horseshoe prior, along with the partial pooling, this allows our model to deal effectively with outliers. Furthermore, for the data sizes presented here, a full Bayesian inference using Stan does not require long computation time, or even particularly high performance hardware. Modern multi-core

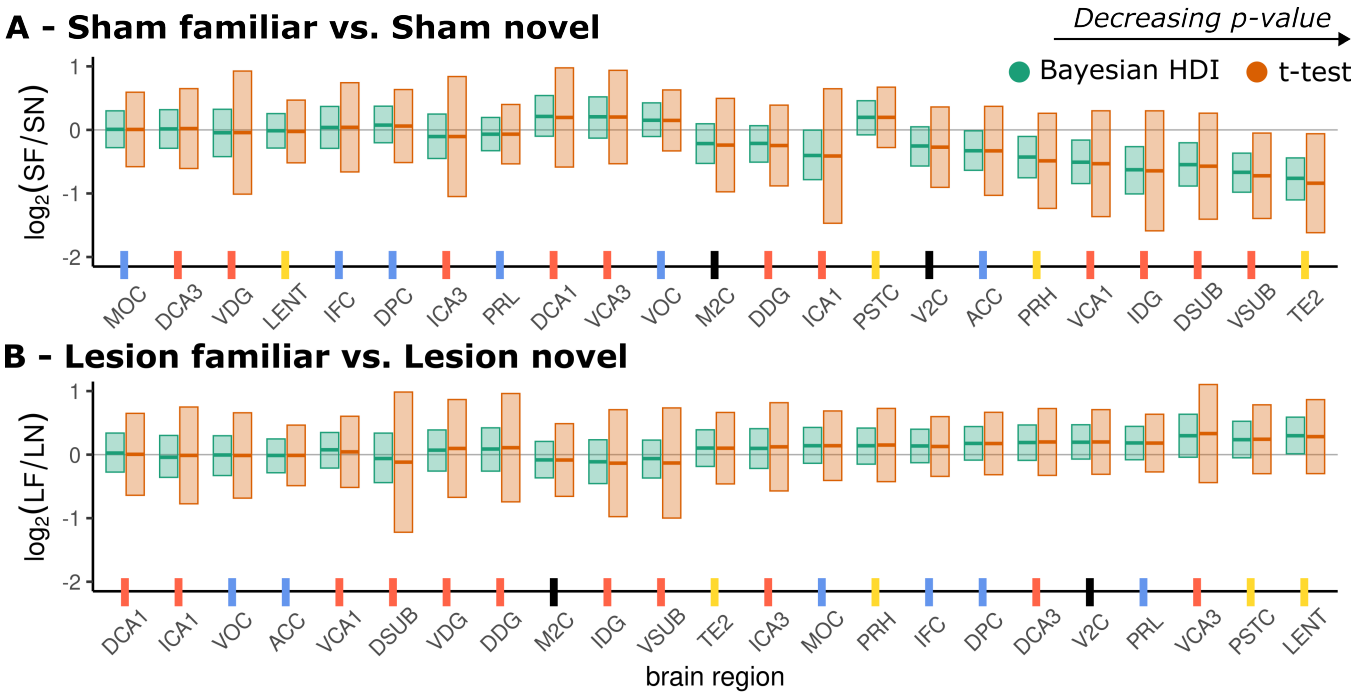

Fig. 4. Results—Case study 1. $\log_2$-fold differences for each surgery type. The 95% Bayesian HDI is given in green, and the 95% confidence interval calculated from a Welch's t-test in orange. Horizontal lines within the intervals mark the posterior mean of the Bayesian results, and the raw data means in the t-test case. The x-axis is ordered in terms of decreasing p-value from the significance test and ticks have been color-paired with the nodes in the recognition memory circuit diagram Figure 3. Black ticks are not present in the circuit because they are the control regions in the experiment. **A**: $\log_2$-fold differences between sham-familiar (SF) and sham-novel (SN) groups. **B**: $\log_2$-fold differences between lesion-familiar (LF) and lesion-novel (LN) groups.

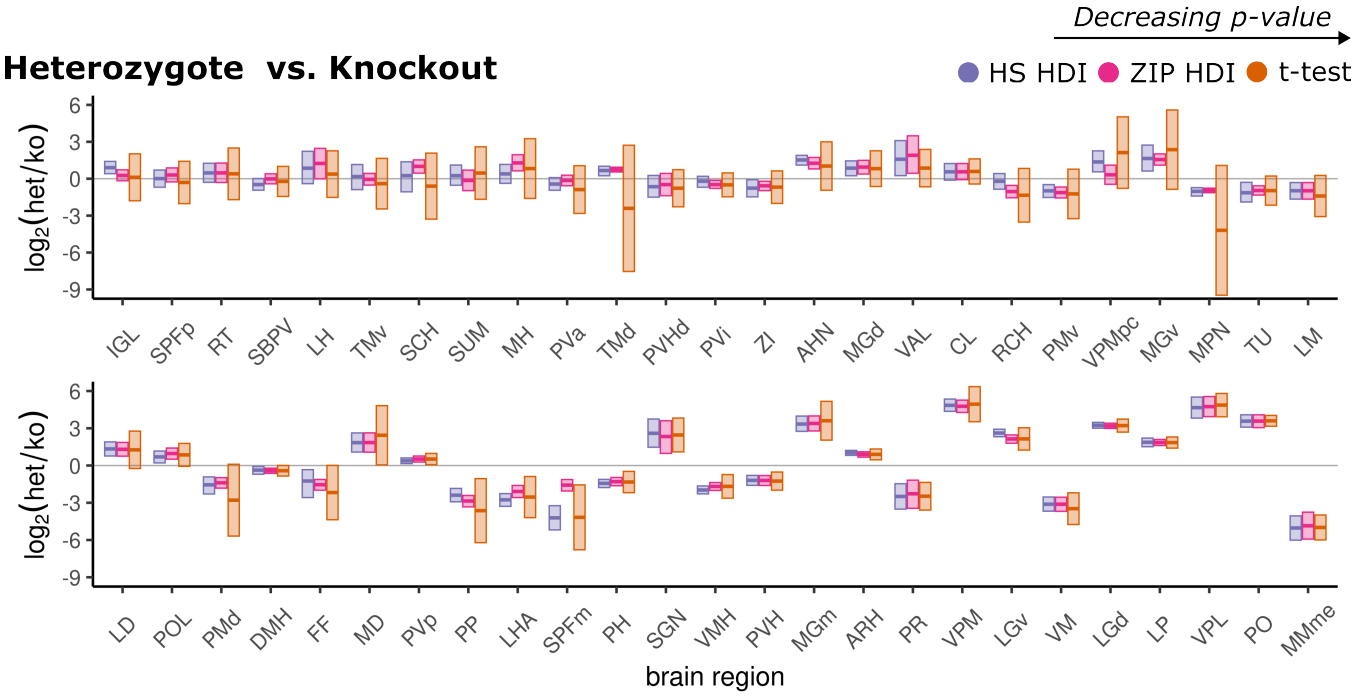

Fig. 5. Results—Case study 2. $\log_2$ fold differences in GFP positive cells between mouse genotypes, heterozygous (HET) and knockout (KO), for each of the fifty recorded brain regions spread across two rows. The 95% Bayesian HDI is given in purple and pink for the Bayesian horseshoe and zero-inflated model. The 95% confidence interval calculated from a Welch's t-test in orange. Horizontal lines within the intervals mark the posterior mean of the Bayesian results and the data estimate for the t-test. The x-axis is ordered in terms of decreasing p-value from the significance test.

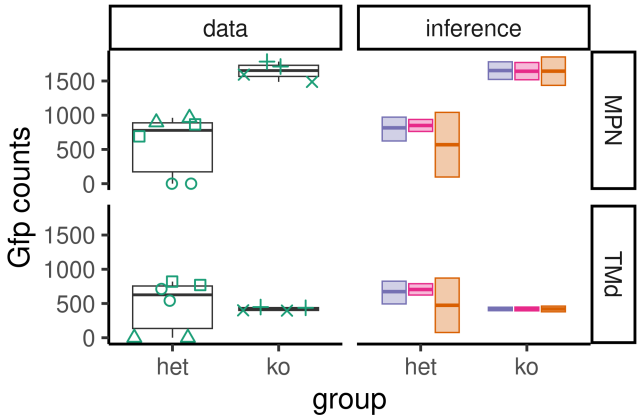

Fig. 6. Zero count observations. On the left under data: boxplots with medians and interquartile ranges for the raw data for two example brain regions. The shape of each point pairs left and right hemisphere readings in each of the five animals. For both example regions there is a heterozygous animal with zero readings in both hemispheres. On the right under inference: HDIs and confidence intervals are plotted. The Bayesian estimates are not strongly influenced by the zero-valued observations and have means close to the data median. This explains the advantage of the Bayesian results over the confidence interval.

laptop processors are quite sufficient for this task. Fitting a model typically takes less than an hour.

The workflow we have exhibited here is intended as a standard approach. We believe that, without extension, it will provide a robust model for cell-count data. However, we also suggest that the standard workflow can be a useful first step for a more comprehensive, extended, model when one is required. We have given an example of this for the second dataset where the anomalous zeros prompted us to change the likelihood to a zero-inflated Poisson. There are other possibilities, for example, zero inflation is not the only way to handle an anomaly in the number of zeros: the hurdle model is an alternative, (42). This is not a mixture model, instead it restricts the probability of zeros to some value π with the probabilities for the positive counts coming from a truncated Poisson distribution. The hurdle mode can deflate as well as inflate the probability mass at zero. This did not match the situation in the data we considered but might for other datasets. Another extension might involve tighter priors based on previous experiments. This is likely to be very relevant for cell-count data since these experiments are rarely performed in isolation and so prior information can be leveraged from a history of empirical results.

One obvious elaboration of our model would replace normal distributions with multivariate normal distributions. This would have two advantages. Firstly, correlations are difficult to estimate for under-sampled data. Including correlation matrix priors provides extra information, for example, based on anatomical connectivity, that can aid the statistical estimation of other parameters. Secondly, it would more closely match our understanding of the experiment: we know that activity is likely to be correlated across regions and so it is apposite to include that directly in the model. Unfortunately the problem of finding a suitable prior for the correlation proved insurmountable: the standard Lewandowski-Kurowicka-Joe distribution (43) which has been useful in lower-dimensional situations is too regularising here. This is an area where further work needs to be done.

It is important to highlight that a mixed effects model is not a uniquely Bayesian construction. Indeed, any model that tries to include more sophistication through hierarchical structures, Bayesian or otherwise is useful. However, non-Bayesian models can be complicated and opaque, they are also often more restrictive, for example they often assume normal distributions, and circumventing these restrictions can make the models even less transparent. A Bayesian approach is, at first, unfamiliar; this can make it seem more obscure than better established methods, but, in the long run, Bayesian models are typically clearer and do not involve so many different assumptions and so many fine adjustments.

Code and data availability

The code necessary to run the models presented in this manuscript can be found at our Github respository: https://github.com/SydneyJake/Hierarchical_Bayesian_Cell_Counts. The data for case study one on nucleus reunion lesion are available from https://doi.org/10.5281/zenodo.12787211 (44). The data from case study two on *Sox14* expressing neurons are available from https://doi.org/10.5281/zenodo.12787287 (45).

ACKNOWLEDGMENTS. We are grateful to Andrew Dowsey and Matthew Nolan for useful discussion and helpful suggestions. We would like to acknowledge funding from the EPSRC (EP/R513179/1 Doctoral Training Partnership PhD studentship to SD and EP/W024020/1 to SS), BBSRC (BB/L02134X/1 to ECW and BB/R007020/1 to AD), Wellcome Trust (206401/Z/17/Z to ECW), Leverhulme Trust (RF-2021-533 to CJH), and MRC (MR/S026630/1 to COD)

1. T Kawashima, H Okuno, H Bito, A new era for functional labeling of neurons: activity-dependent promoters have come of age. *Front. Neural Circuits* **8**, 37 (2014).
2. ES Lein, et al., Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
3. SW Oh, et al., A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
4. TL Daigle, et al., A suite of transgenic driver and reporter mouse lines with enhanced brain-cell-type targeting and functionality. *Cell* **174**, 465–480 (2018).
5. JA Harris, et al., Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195–202 (2019).
6. WB Kim, JH Cho, Encoding of discriminative fear memory by input-specific ltp in the amygdala. *Neuron* **95**, 1129–1146 (2017).
7. J Haubrich, K Nader, Network-level changes in the brain underlie fear memory strength. *eLife* **12**, RP88172 (2023).
8. KE Dorst, et al., Hippocampal engrams generate variable behavioral responses and brain-wide network states. *J. Neurosci.* **44** (2024).
9. T Liebmann, et al., Three-dimensional study of alzheimer's disease hallmarks using the idisco clearing method. *Cell Reports* **16**, 1138–1152 (2016).
10. Y Kim, et al., Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell Reports* **10**, 292–305 (2015).
11. V Bonapersona, et al., The mouse brain after foot shock in four dimensions: Temporal dynamics at a single-cell resolution. *Proc. Natl. Acad. Sci.* **119**, e2114002119 (2022).
12. A Gelman, et al., *Bayesian data analysis.* (CRC press), (2013).
13. R McElreath, *Statistical rethinking: A Bayesian course with examples in R and Stan.* (Chapman and Hall/CRC), (2018).
14. R van de Schoot, et al. *Nat. Rev. Methods Primers* **1**, 1–26 (2021).
15. S Dimmock, C O'Donnell, CJ Houghton, Bayesian analysis of phase data in EEG and MEG. *eLife* **12**, e84602 (2023).
16. EN Brown, LM Frank, D Tang, MC Quirk, MA Wilson, A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neurosci.* **18**, 7411–7425 (1998).
17. K Zhang, I Ginzburg, BL McNaughton, TJ Sejnowski, Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.* **79**, 1017–1044 (1998).
18. RJ Moran, et al., Bayesian estimation of synaptic physiology from the spectral responses of neural masses. *Neuroimage* **42**, 272–284 (2008).
19. RP Costa, PJ Sj"ostr"om, MC Van Rossum, Probabilistic inference of short-term synaptic plasticity in neocortical microcircuits. *Front. Comput. Neurosci.* **7**, 75 (2013).
20. AD Bird, MJ Wall, MJ Richardson, Bayesian inference of synaptic quantal parameters from correlated vesicle release. *Front. Comput. Neurosci.* **10**, 116 (2016).

21. O Bykowska, et al., Model-based inference of synaptic transmission. *Front. Synaptic Neurosci.* **11**, 21 (2019).
22. Y Mishchencko, JT Vogelstein, L Paninski, A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data. *The Annals Appl. Stat.* pp. 1229–1261 (2011).
23. F Cinotti, MD Humphries, Bayesian mapping of the striatal microcircuit reveals robust asymmetries in the probabilities and distances of connections. *J. Neurosci.* **42**, 1417–1435 (2022).
24. E Aarts, M Verhage, JV Veenvliet, CV Dolan, S Van Der Sluis, A solution to dependency: using multilevel analysis to accommodate nested data. *Nat. Neurosci.* **17**, 491–496 (2014).
25. GR Barker, EC Warburton, When is the hippocampus involved in recognition memory? *J. Neurosci.* **31**, 10721–10731 (2011).
26. A Ennaceur, N Neave, JP Aggleton, Neurotoxic lesions of the perirhinal cortex do not mimic the behavioural effects of fornix transection in the rat. *Behav. Brain Res.* **80**, 9–25 (1996).
27. G Norman, M Eacott, Impaired object recognition with increasing levels of feature ambiguity in rats with perirhinal cortex lesions. *Behav. Brain Res.* **148**, 79–91 (2004).
28. JWT Ho, et al., Contributions of area te2 to rat recognition memory. *Learn. & Mem.* **18**, 493–501 (2011).
29. GR Barker, EC Warburton, A critical role for the nucleus reuniens in long-term, but not short-term associative recognition memory formation. *J. Neurosci.* **38**, 3208–3217 (2018).
30. WB Hoover, RP Vertes, Collateral projections from nucleus reuniens of thalamus to hippocampus and medial prefrontal cortex in the rat: a single and double retrograde fluorescent labeling study. *Brain Struct. Funct.* **217**, 191–209 (2012).
31. BMS Exley, "The role of the nucleus reuniens of the thalamus in the recognition memory network," Master's thesis, School of Physiology, Pharmacology & Neuroscience, University of Bristol (2019).
32. P Jager, et al., Dual midbrain and forebrain origins of thalamic inhibitory interneurons. *eLife* **10**, e59272 (2021).
33. P Jager, et al., Tectal-derived interneurons contribute to phasic and tonic inhibition in the visual thalamus. *Nat. Commun.* **7**, 13579 (2016).
34. B Golding, et al., Retinal input directs the recruitment of inhibitory interneurons into thalamic visual circuits. *Neuron* **81**, 1057–1069 (2014).
35. A Delogu, et al., Subcortical visual shell nuclei targeted by iprgcs develop from a sox14+-gabaergic progenitor and require sox14 to regulate daily activity rhythms. *Neuron* **75**, 648–662 (2012).
36. A Gelman, J Hill, *Data analysis using regression and multilevel/hierarchical models.* (Cambridge University Press), (2006).
37. CM Carvalho, NG Polson, JG Scott, The horseshoe estimator for sparse signals. *Biometrika* **97**, 465–480 (2010).
38. J Piironen, A Vehtari, Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron. J. Stat.* **11** (2017).
39. B Carpenter, et al., Stan: A probabilistic programming language. *J. Stat. Softw.* **76** (2017).
40. M Betancourt, Identifying the optimal integration time in Hamiltonian Monte Carlo. *arXiv:1601.00225* (2016).
41. MD Hoffman, A Gelman, , et al., The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
42. JG Cragg, Some statistical models for limited dependent variables with application to the demand for durable goods. *Econom. J. Econom. Soc.* pp. 829–844 (1971).
43. D Lewandowski, D Kurowicka, H Joe, Generating random correlation matrices based on vines and extended onion method. *J. Multivar. Analysis* **100**, 1989–2001 (2009).
44. B Exley, S Dimmock, C Warburton, Exley_warburton_nre_lesion_cell_count_data (2024).
45. M Gerald, D Sydney, L Menage, A Delogu, S Schultz, Moore_schultz_sox14_expressing_neurons (2024).
46. A Gelman, DB Rubin, Inference from iterative simulation using multiple sequences. *Stat. Sci.* pp. 457–472 (1992).
47. A Vehtari, A Gelman, D Simpson, B Carpenter, PC B"urkner, Rank-normalization, folding, and localization: an improved r for assessing convergence of mcmc (with discussion). *Bayesian Analysis* **16**, 667–718 (2021).

1. Full model

Here we present the complete mathematical model for each of of the three models applied in the main text. In all cases, the exposure term is only necessary for the first case study. Area was not available for the second so the exposure term was left out.

A. Poisson model.

$$y_i \sim \text{Poisson}(\lambda_i) \quad [8]$$
$$\log \lambda_i = E_i + \gamma_i \quad [9]$$
$$\theta_{rg} \sim \text{Normal}(5, 2) \quad [10]$$
$$\gamma_i \sim \text{Normal}\Big(\theta_{r[i],g[i]}, \tau_{r[i],g[i]}\Big) \quad [11]$$
$$\tau_{rg} \sim \text{HalfNormal}(\log(1.05)) \quad [12]$$

B. Horseshoe model.

$$y_i \sim \text{Poisson}(\lambda_i) \quad [13]$$
$$\log \lambda_i = E_i + \gamma_i \quad [14]$$
$$\theta_{rg} \sim \text{Normal}(5, 2) \quad [15]$$
$$\gamma_i \sim \text{Normal}\Big(\theta_{r[i],g[i]}, \kappa_i \times \tau_{r[i],g[i]}\Big) \quad [16]$$
$$\tau_{rg} \sim \text{HalfNormal}(\log(1.05)) \quad [17]$$
$$\kappa_i \sim \text{HalfNormal}(1) \quad [18]$$

C. Zero-inflated model.

$$y_i \sim \text{ZIPoisson}(\lambda_i, \pi) \quad [19]$$
$$\pi \sim \text{Beta}(1, 5) \quad [20]$$
$$\log \lambda_i = E_i + \gamma_i \quad [21]$$
$$\theta_{rg} \sim \text{Normal}(5, 2) \quad [22]$$
$$\gamma_i \sim \text{Normal}\Big(\theta_{r[i],g[i]}, \tau_{r[i],g[i]}\Big) \quad [23]$$
$$\tau_{rg} \sim \text{HalfNormal}(\log(1.05)) \quad [24]$$

2. Distributions

A. Zero inflated Poisson distribution. The zero inflated Poisson distribution is a mixture distribution with mixing parameter π. The distribution is formally defined below. If,

$$p_Y(y) \equiv \text{ZIPoisson}(\pi, \lambda) \quad [25]$$

then

$$p_Y(y) = \begin{cases} \pi + (1-\pi)\exp(-\lambda) & \text{if} \quad y = 0 \\ (1-\pi) f(y) & \text{otherwise} \end{cases} \quad [26]$$

where

$$f(y) \equiv \text{Poisson}(\lambda) \quad [27]$$

Equivalently, the mixture can be specified with an indicator function.

$$p_Y(y) = \mathbf{1}_0(y)\pi + (1-\pi) f(y) \quad [28]$$

B. Half Normal distribution. The HalfNormal distribution coincides with a zero-mean normal distribution truncated at zero. It has a single scale parameter σ. If,

$$p_Y(y) \equiv \text{HalfNormal}(\sigma) \quad [29]$$

then

$$p_Y(y) = \frac{\sqrt{2}}{\sigma\pi} \exp\left(-\frac{y^2}{2\sigma^2}\right) \quad [30]$$

is the probability density function.

3. Additional methods

A. Fold differences. In our results we present differences between experimental groups in terms of $\log_2$-fold differences. We calculate this as follows. The parameter of interest $\theta_{r,g=i}$ is modeled on the natural log scale owing to the log-link function necessary for the Poisson regression. At the average animal the difference,

$$\theta_{r,g=i} - \theta_{r,g=j} = \log \frac{\exp(\theta_{r,g=i})}{\exp(\theta_{r,g=j})} \quad [31]$$

$$= \log \Delta_{ij} \quad [32]$$

is the natural log of the ratio of the expected counts. To obtain $\log_2$-fold differences we simply change the base by multiplying $\log \Delta_{ij}$ by $\log_2(e)$.

B. Data transformations. To facilitate a comparison with the Bayesian intervals in terms of $\log_2$-fold differences, it was necessary to add one to any zero counts before applying the t-test.

C. Non-centered parameterisation. Hierarchical models can produce geometry that is difficult for the sampler to explore. Fortunately, there exists a simple reparameterisation known as non-centering that can remedy this problem. In our model, instead of sampling γ_i directly, we sample the parameter $\tilde{\gamma}_i$ instead, and use it to reconstruct γ_i. That is, sample $\tilde{\gamma}_i$ from a standard normal distribution,

$$\tilde{\gamma}_i \sim \text{Normal}(0, 1) \quad [33]$$

and reconstruct γ_i as a deterministic function of sampled values of $\tilde{\gamma}_i$ and $\tau_{r[i],g[i]}$.

$$\gamma_i = \theta_{r[i],g[i]} + \tau_{r[i],g[i]} \times \tilde{\gamma}_i \quad [34]$$

This removes the frustrating joint behaviour between of $(\gamma_i, \tau_{r[i],g[i]})$, and promotes efficient sampling.

D. Preprocessing - case study 1. In these data some animals produced more than on reading per brain region. Before fitting our model these were summed together to produce a single count; the exposure term was also properly adjusted to correctly reflect the area of the recording site.

4. Software, packages and libraries

`R` Libraries	Version	Description
`rstan`	`2.26.3`	complete Stan library
`cmdstanr`	`0.5.2`	lightweight Stan library
`HDInterval`	`0.2.2`	calculating HDI in R
`ggplot2`	`3.4.1`	plotting
`bayesplot`	`1.9.0`	plotting
`tidyverse`	`1.3.1`	`tibble, tidyr, readr, purr, dplyr, stringr, forcats`

- `R` version `4.2.1`–"Funny-looking-kid".
- Computation was performed locally on a Dell XPS 13 7390 laptop[*][†]
- Panels composed using `inkscape` version `1.2.2`.

[*] Intel i7-10510U @ 1.80 GHz, 16 GB of RAM.
[†] Ubuntu `20.04.4` LTS.

5. Glossary

Here we include a glossary of terms for the brain regions in each case study.

Term	Definition
ACC	Anterior cingulate cortex.
DCA1/3	Dorsal CA1/3
DDG	Dorsal dendate gyrus.
DPC	Dorsal peduncular cortex.
DSUB	Dorsal subiculum.
HPC	hippocampus.
ICA1/3	Intermediate CA1/3
IDG	Intermediate dendate gyrus.
IFC	Infralimbic cortex.
LENT	Lateral entorhinal cortex.
MOC	medial orbital cortex.
MPFC	Medial prefrontal cortex
M2C	Motor cortex M2
NRe	Nucleus reuniens
PRL	prelimbic cortex.
PRH	Perirhinal cortex
PSTC	postrhinal cortex.
TE2	Temporal association cortex
VCA1/3	Ventral CA1/3
VDG	Ventral dendate gyrus.
VOC	Ventral orbital cortex.
VSUB	Ventral subiculum.
V2C	Visual cortex V2

Term	Definition
AHN	Anterior hypothalamic nucleus
ARH	Arcuate hypothalamic nucleus
CL	Central lateral nucleus of the thalamus
DMH	Dorsomedial nucleus of the hypothalamus
FF	Fields of Forel
IGL	Intergeniculate leaflet of the lateral geniculate complex
LD	Lateral dorsal nucleus of thalamus
LM	Lateral mammillary nucleus
LGv	Ventral part of the lateral geniculate complex
LGd	Dorsal part of the lateral geniculate complex
LH	Lateral habenula
LHA	Lateral hypothalamic area
LP	Lateral posterior nucleus of the thalamus
MD	Mediodorsal nucleus of thalamus
MGd	Medial geniculate complex, dorsal part
MGv	Medial geniculate complex, ventral part
MGm	Medial geniculate complex, medial part
MH	Medial habenula
MMme	Medial mammillary nucleus, median part
MPN	Medial preoptic nucleus
PH	Posterior hypothalamic nucleus
PMd	Dorsal premammillary nucleus
PMv	Ventral premammillary nucleus
PO	Posterior complex of the thalamus
POL	Posterior limiting nucleus of the thalamus

Term	Definition
PP	Peripeduncular nucleus
PR	Perireunensis nucleus
PVa	Periventricular hypothalamic nucleus, anterior part
PVH	Paraventricular hypothalamic nucleus
PVHd	Paraventricular hypothalamic nucleus, descending division
PVi	Periventricular hypothalamic nucleus, intermediate part
PVp	Periventricular hypothalamic nucleus, posterior part
RCH	Retrochiasmatic area
RT	Reticular nucleus of the thalamus
SBPV	Subparaventricular zone
SCH	Suprachiasmatic nucleus
SGN	Suprageniculate nucleus
SPFm	Subparafascicular nucleus, magnocellular part
SPFp	Subparafascicular nucleus, parvicellular part
SUM	Supramammillary nucleus
TMd	Tuberomammillary nucleus, dorsal part
TMv	Tuberomammillary nucleus, ventral part
TU	Tuberal nucleus
VAL	Ventral anterior-lateral complex of the thalamus
VMH	Ventromedial hypothalamic nucleus
VM	Ventral medial nucleus of the thalamus
VPL	Ventral posterolateral nuclues of the thalamus
VPM	Ventral posteromedial nucleus of the thalamus
VPMpc	Ventral posteromedial nucleus of the thalamus, parvicellular part
ZI	Zona incerta

6. Sampler diagnostics

The basic Poisson model (Equation 1) was sampled excellently. Measures of sampling performance such as $\hat{R}$ (46, 47), and effective sample size, were all satisfactory (SI figure 1). Similarly, for the zero-inflated model (Equation 7), no problems were observed for any of the diagnostics (SI figure 2). Contrasting with this, the horseshoe model (Equation 6), exhibited some signs of fitting problems (SI figure 3). Divergences were not observed, and given the longer chain length this is reassuring evidence against biased computation. However for many parameters the effective sample size is much lower than we would like to see. This is reflected in the trace plots: the sampler is not making large jumps across the parameter space implying high auto-correlation and low effective sample size. Unfortunately, the horseshoe is notoriously hard to fit and we resort to brute-force methods such as increasing the number of iteration and reducing the step size of the sampler to improve the inference. In the following three plots diagnostics have been summarised with the following three items:

- **A**: The performance of the sampler is illustrated by plotting $\hat{R}$ (R-hat, $\hat{R} \approx 1$ ideal) against the ratio of the effective number of samples (larger is better) for each parameter in the model. Points represent individual parameters in the model, and have further been colour coded by their type. For example, all $\theta_{r,g}$ are coloured in green. Points have also been scaled based on how numerous the parameters are so the more numerous parameters have smaller dots, the less numerous, larger.
- **B**: A histogram comparing the marginal energy distribution π_E, and the transitional energy distribution $\pi_{\Delta E}$ of the Hamiltonian. Ideally, these distributions should match each other closely if the posterior distribution has been properly explored by the sampler.
- **C**: For each parameter type the parameter with the 'poorest' mixing (largest $\hat{R}$) are presented with a post-warmup trace plot that overlays the ordered sequence of samples from each of the four chains. Corresponding points in **A** are marked with a black border and zero transparency.

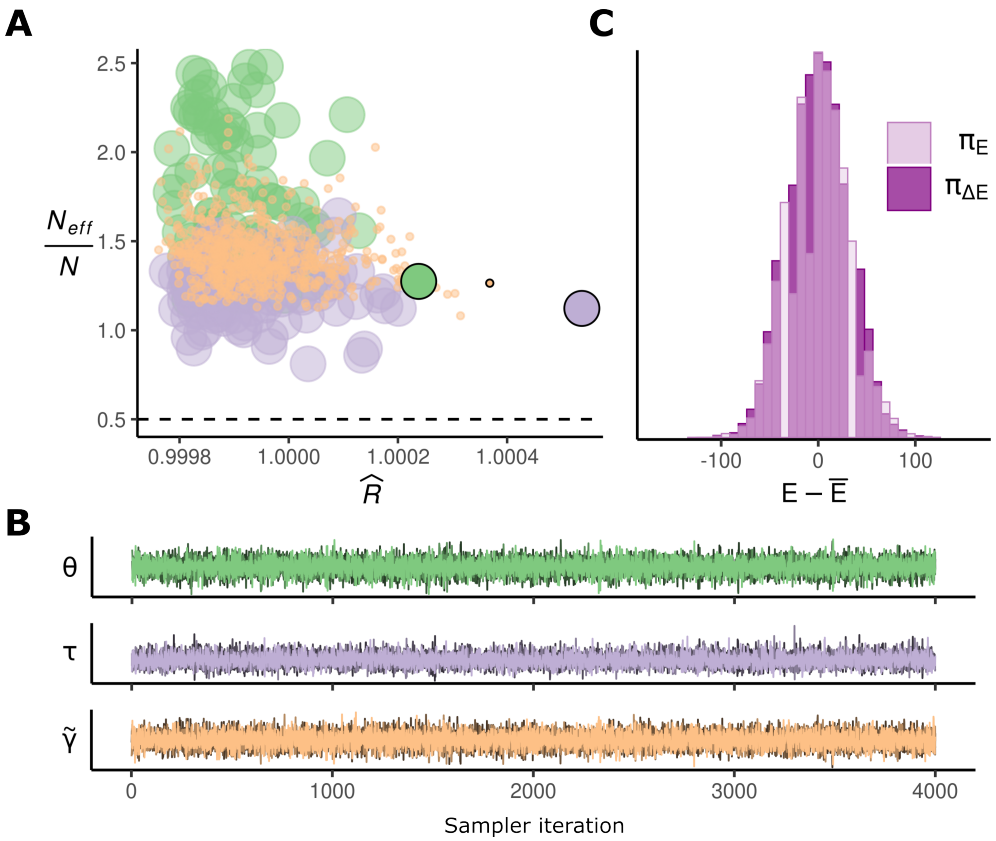

Fig. 7. Diagnostics - Poisson. Standard Poisson model, case study - 1.

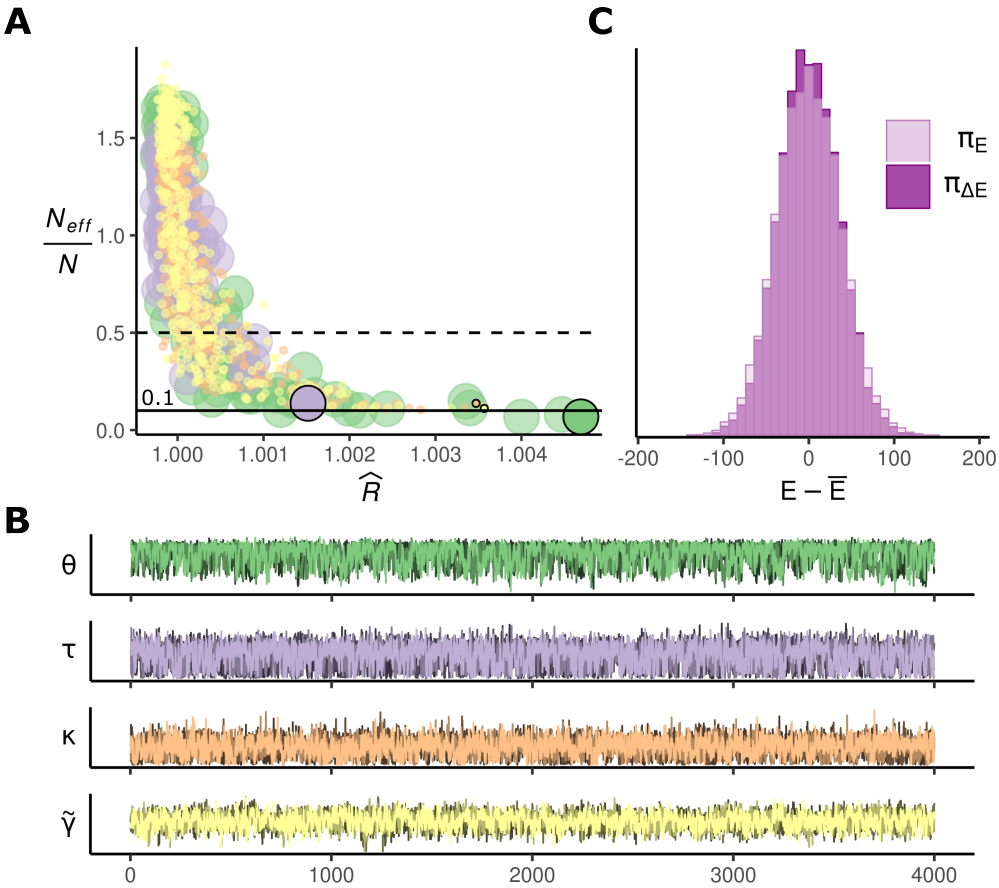

Fig. 8. Diagnostics - Horseshoe. Horseshoe model, case study - 2.

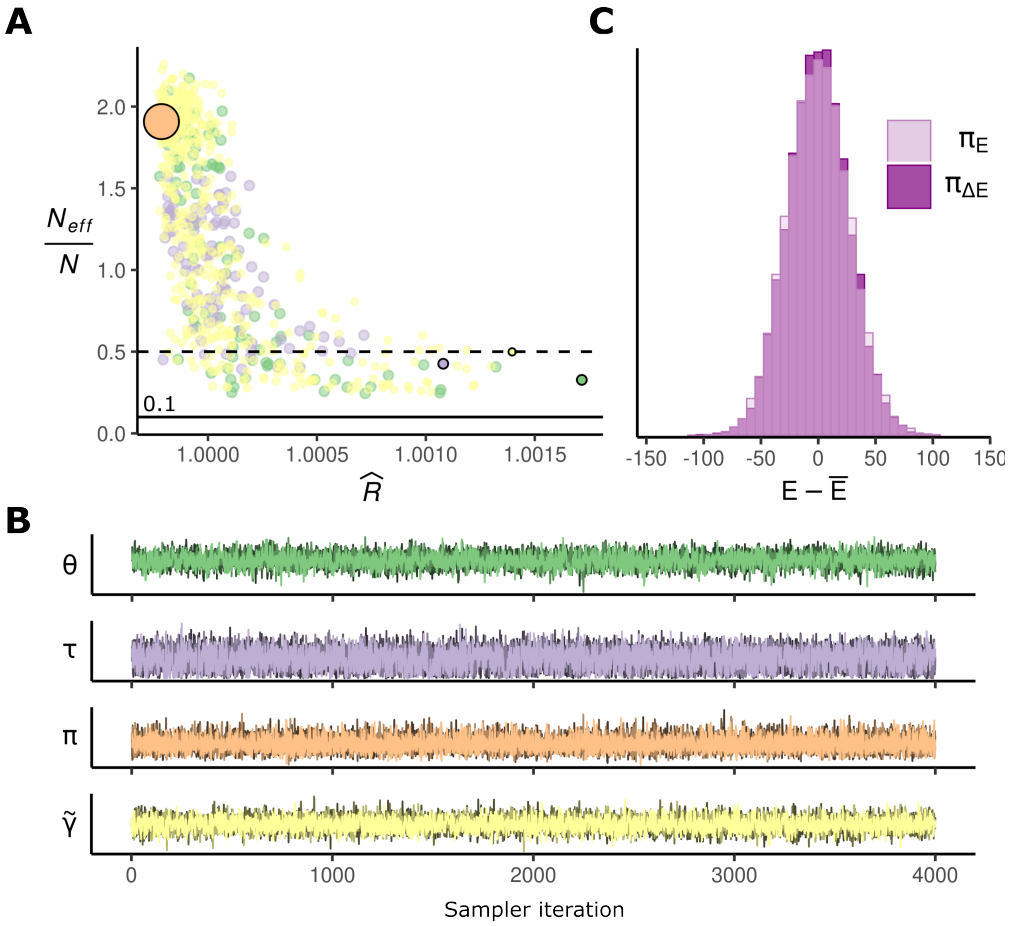

Fig. 9. Diagnostics - ZIPoisson. Zero-inflated Poisson, case study - 2.

7. Posterior predictive checking

Posterior predictive checks use the posterior predictive distribution,

$$p(y^{\text{rep}}|y) = \int p(y^{\text{rep}}|\theta)p(\theta|y)\,d\theta \qquad [35]$$

where $p(\theta|y)$ is the posterior distribution, and $p(y^{\text{rep}}|\theta)$ is the data distribution for y^{rep} that follows the same form as the likelihood for y. If we can verify that the posterior predictive distribution can generate replicate datasets with similar statistics to the observed data, then we might conclude that our model is consistent with the observed data and useful for answering questions about it. In practice a Monte-carlo approach is used to approximate statistics of the posterior predictive distribution. For example, if T is the test statistic of interest such as the sample mean, then

1. For $1, \ldots, S$.
   (a) sample $\theta_s \sim p(\theta|y)$
   (b) sample $y_s^{\text{rep}} \sim p(y^{\text{rep}}|\theta_s)$
   (c) calculate $T(y_s^{\text{rep}})$, where T is the statistic of interest
2. return $\{T(y_1^{\text{rep}}), \ldots, T(y_S^{\text{rep}})\} \approx p(T(y^{\text{rep}})|y)$

The posterior predictive checks that follow examine two important statistics of count data. 1) The standard deviation of the data (dispersion), as figure **a**. 2) The proportion of zeroes in the data (zero-inflation), as figure **B**. The value of the test statistic applied to the observed data $T(y)$ is plotted as a solid purple line, and the distribution of test statistics $\{T(y_1^{\text{rep}}), \ldots, T(y_S^{\text{rep}})\}$ as a purple histogram.

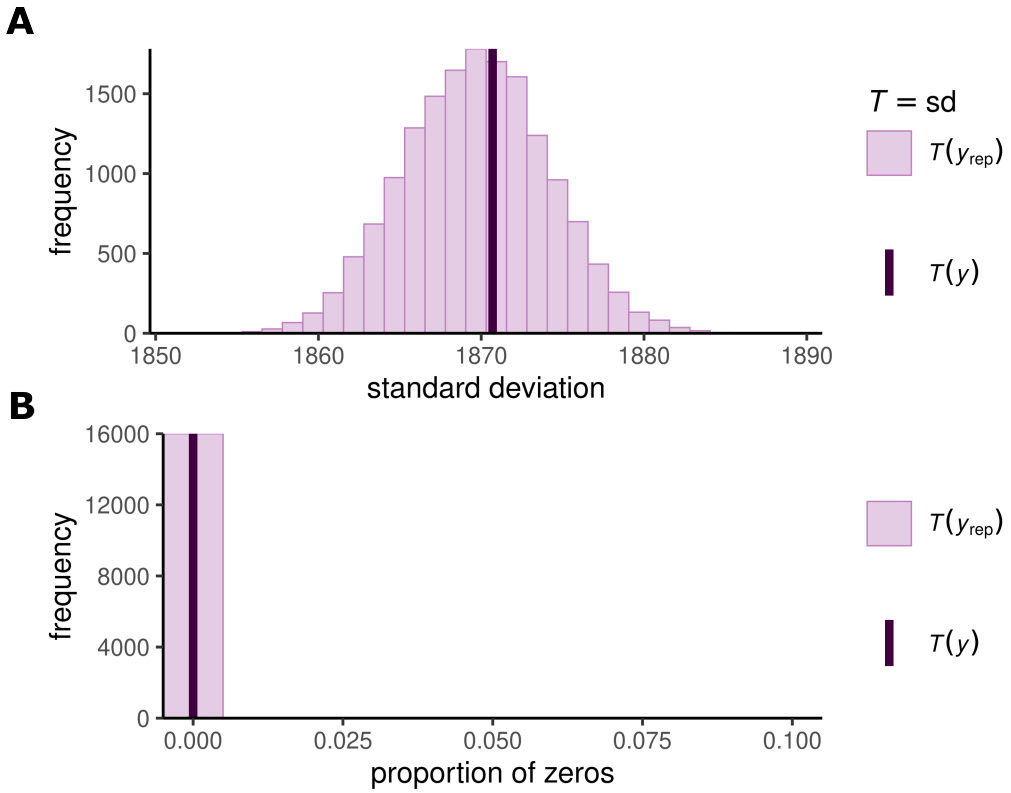

Fig. 10. PPC - Poisson. Posterior predictive check for the standard Poisson model in case study - 1. **A**: The proportion of zeroes in the data matches the proportion of zeroes in posterior predictive samples. This proportion is zero. **B**: The distribution of standard deviations computed over a number of posterior predictive datasets (histogram), aligns with the standard deviation of the data.

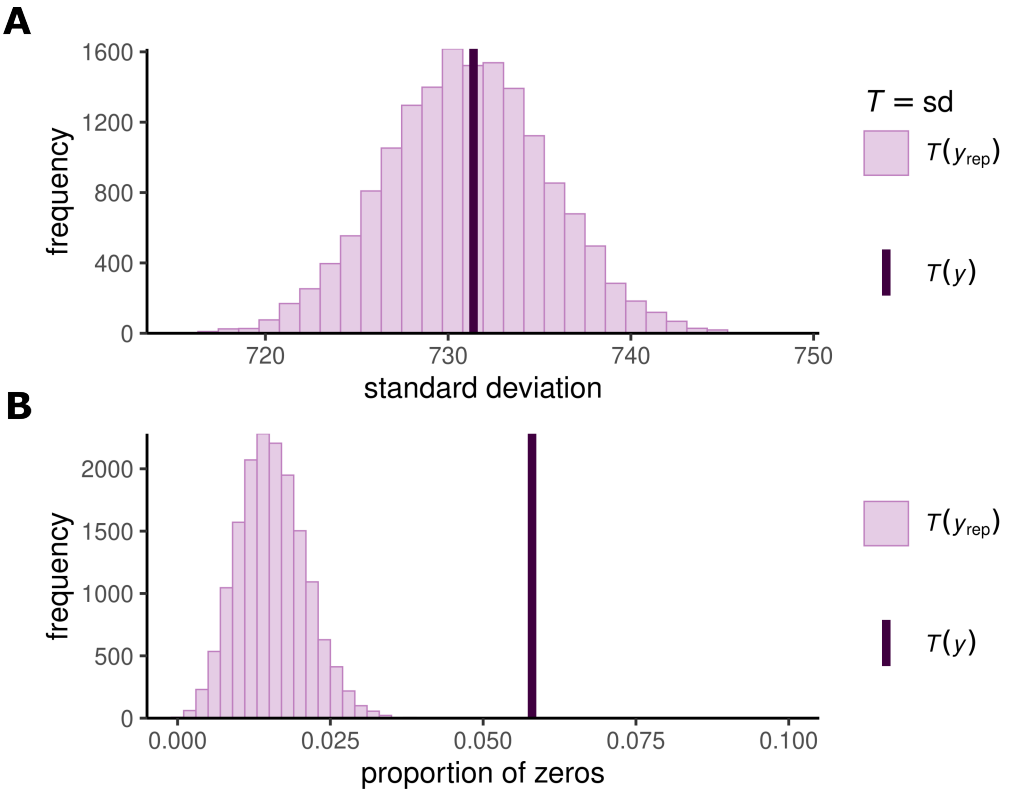

Fig. 11. PPC - Horseshoe. Horseshoe model, case study - 2. Posterior predictive check for the standard horseshoe model in case study - 2. **A**: The proportion of zeroes in the data is larger than those found in posterior predictive datasets. This makes sense, because the likelihood is still a Poisson distribution. **B**: The distribution of standard deviations computed over a number of posterior predictive datasets (histogram), aligns with the standard deviation of the data.

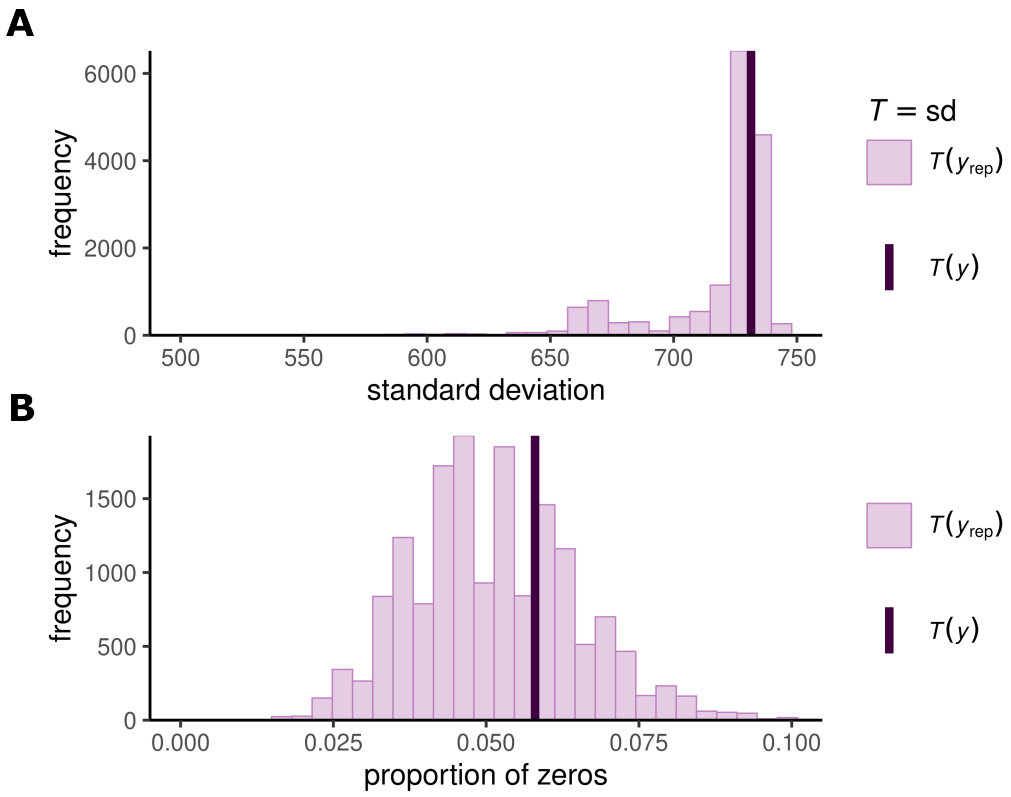

Fig. 12. PPC - ZIPoisson. Zero-inflated Poisson, case study - 2. **A**: The proportion of zeroes in the data matches the proportion of zeroes in posterior predictive samples. **B**: The distribution of standard deviations computed over a number of posterior predictive datasets (histogram), aligns with the standard deviation of the data.

8. Horseshoe densities

In our model a horseshoe prior was used to allow some γ_i, typically those informed by $y_i = 0$, to escape regularisation by partial pooling. However, we encountered many problems with the default parameterisation that assigns a HalfCauchy density to the individual inflation parameters κ_i. In Figure 13**A**, the proportional conditional posterior density

$$, p(\tilde{\gamma}_i, \kappa_i | \theta, \tau, y_i) \propto p(y_i | \tilde{\gamma}_i, \theta, \tau, \kappa_i) p(\tilde{\gamma}_i) p(\kappa_i) \quad [36]$$

where θ and τ have been fixed, is plotted when y_i lies close to the population mean (left), or when it equals zero (right). Note that the x-axis is $\tilde{\gamma}_i$ and not γ_i because the model samples a non-centered parameterisation. That is,

$$\tilde{\gamma}_i = \frac{\gamma_i - \theta}{\tau \kappa_i}, \quad \tilde{\gamma}_i \sim \text{Normal}(0, 1) \quad [37]$$

captures the deviations of γ_i around the population mean θ.

Figure 13**B** plots samples from the marginal posterior $p(\tilde{\gamma}, \kappa_i | y_i)$, when fitting to the data $\mathbf{y} = \{770, 820, 713, 541, 0, 0\}$. Each data point corresponds to one of the six plots in Figure 13**B**. This small example dataset is in fact the cell-counts for region TMd recorded from the heterozygote group in case study 2. The similarities in geometry can be readily seen with Figure 13**A**. A large number of divergences were produced by the sampler as demonstrated by the number of pink points compared to the non-divergent transitions in blue. For fixed τ the values $\tilde{\gamma}$ can take increase or decrease with smaller or larger κ respectively. The HalfCauchy places an extremely long right tail over κ that frustrates this relationship resulting in a posterior density that is difficult to sample from.

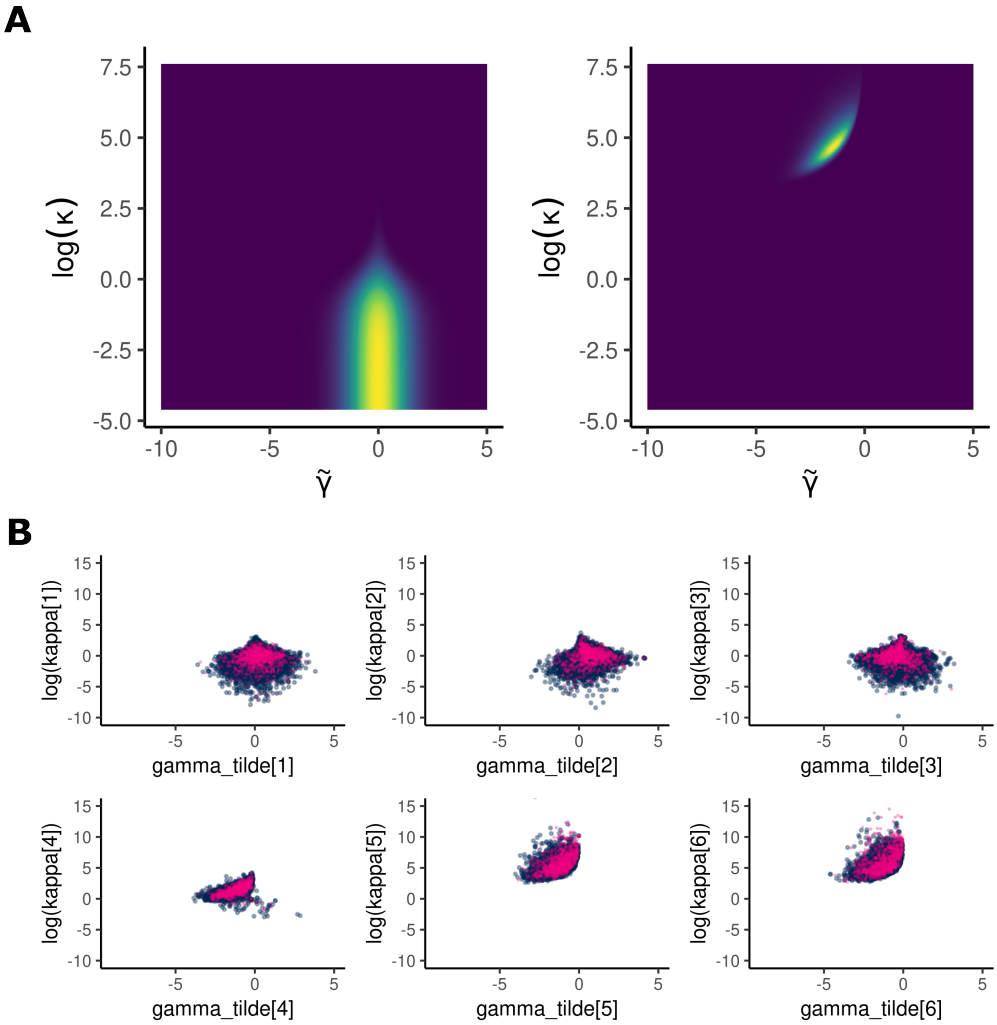

Fig. 13. Horseshoe densities. **A**: Conditional posterior. **B**: MCMC pair plots. Divergent samples are coloured in pink, non-divergent in blue.

Modified horseshoe. For a Poisson model with parameters modeled on the log scale we consider the Cauchy parameterisation to be too extreme. In light of this we opted for a pragmatic approach, a modification to the original horseshoe by replacing the HalfCauchy distribution with a HalfNormal distribution. The modified horseshoe cuts off the top of the funnels by restricting κ to produce pleasant posterior geometry Figure 13**A**, **B**. The modified horseshoe is much easier to sample from, but with the cost of a much more constraining prior over γ_i.

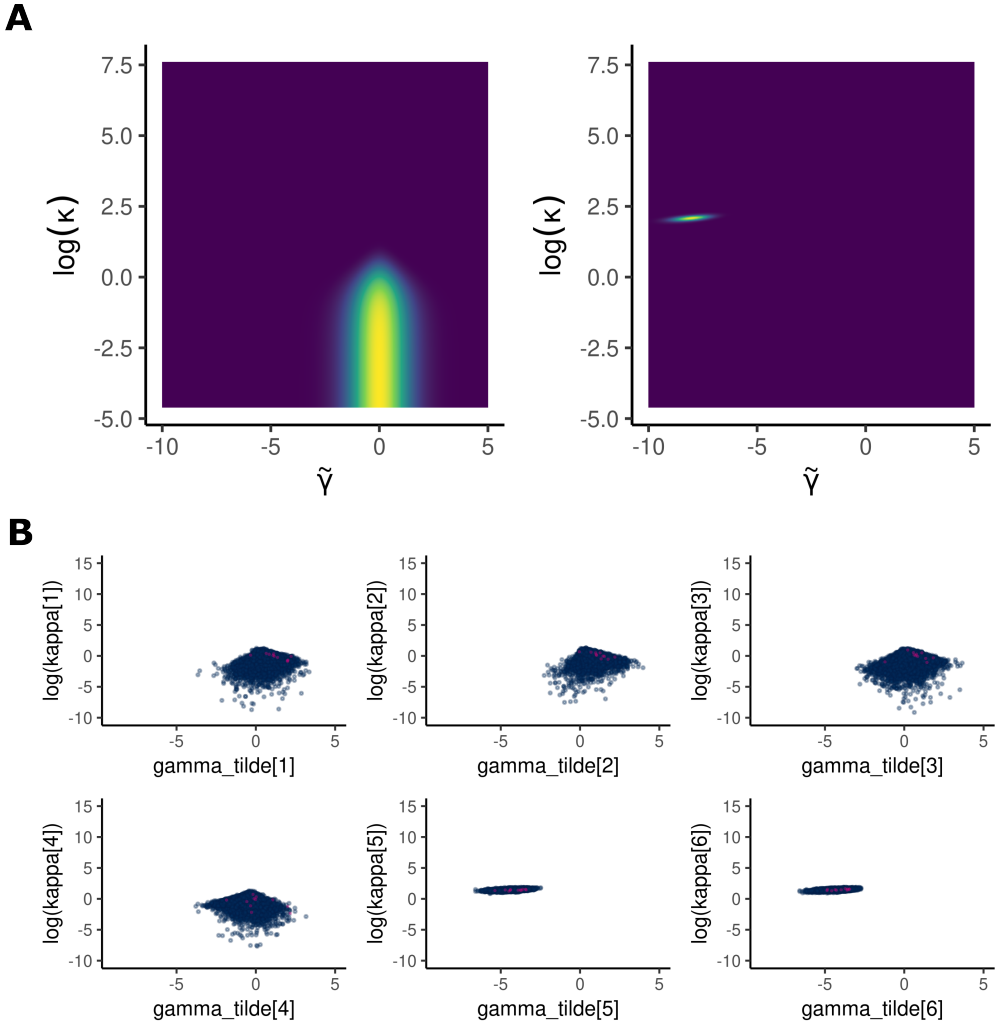

Fig. 14. Modified horseshoe densities. **A**: The conditional posterior $p(\tilde{\gamma}, \kappa|\theta, \tau, y)$ when $y = 0$ (left) and $y \neq 0$ (right). **B**: MCMC pair plots of samples from the marginal posterior density $p(\tilde{\gamma}, \kappa|\mathbf{y})$.