Appendix to “Fast approximate Bayesian inference of HIV indicators using PCA adaptive Gauss-Hermite quadrature”

Adam Howes[*] Alex Stringer[†] Seth R. Flaxman[‡] Jeffrey W. Eaton[§]

Contents

[*]Department of Mathematics, Imperial College London
[†]Department of Statistics and Actuarial Science, University of Waterloo
[‡]Department of Computer Science, Oxford University
[§]Harvard T.H. Chan School of Public Health, Harvard University

S1 Simplified Naomi model description

In this section we describe the simplified Naomi model (Eaton et al. 2021) in complete detail.

S1.1 Background

S1.1.1 Indexing

Consider the most recent national household survey with HIV testing which has taken place in the country of interest. Let $x \in \{1, \ldots, n\}$ be district, $s \in \{\text{F}, \text{M}\}$ be sex, and $a \in \{0\text{-}5, 5\text{-}10, \ldots, 75\text{-}80, 80+\}$ be five-year age groups. Each district is located within a correspodning Spectrum (Stover et al. 2010) region R_x. As short-hand, we write $a = l$ to refer to the age group with lower bound l, e.g. $a = 20$ for $a = 20\text{-}25$.

We index the known quantity population size $N_{x,s,a}$ by district, sex and age-band, as well as the following unknown quantities:

- HIV prevalence $\rho_{x,s,a} \in [0, 1]$,
- ART coverage $\alpha_{x,s,a} \in [0, 1]$,
- annual HIV incidence rate $\lambda_{x,s,a} > 0$, and
- proportion of HIV positive persons recently infected $\kappa_{x,s,a} \in [0, 1]$.

Sometimes data are observed at an aggregate level, rather than the more granular modelled level. We use $\{\cdot\}$ to generically refer to a aggregate set over which an observation is made, e.g. $\{a\} = \{15\text{-}19, \ldots, 45\text{-}49\}$ for the adult age range 15-49. The notation $\sum_{\{x\}}$ is used as shorthand for $\sum_{x \in \{x\}}$. Analogous notion is used for sums over $\sum_{a \in \{a\}}$ and $\sum_{s \in \{s\}}$.

S1.1.2 Structured random effects

We use structured random effects to enable partial pooling of information across units assessed as being similar, such as those belonging to neighbouring districts or adjacent age groups. Let $\mathbf{u}$ be a generic random effect, with length $\dim(\mathbf{u}) > 1$. Three types of structured random effects are used in the model:

1. We specify the first order auto-regressive model by $\mathbf{u} \sim \text{AR1}(\sigma, \phi)$ such that

$$u_1 \sim \left(0, \frac{1}{1 - \rho^2}\right), \tag{1}$$

$$u_i = \rho u_{i-1} + \epsilon_i, \quad i = 2, \ldots, \dim(\mathbf{u}), \tag{2}$$

   where $\epsilon_i \sim \mathcal{N}(0, 1)$ are independent and identically distributed (IID) Gaussian random variables, $\sigma > 0$ is the marginal standard deviation, and $|\rho| < 1$ is the (lag-one) correlation parameter.

2. We use $\mathbf{u} \sim \text{ICAR}(\sigma)$ to refer to the Besag intrinsic conditional auto-regressive model (ICAR) (Besag, York, and Mollié 1991) with full conditionals

$$u_i \,|\, \mathbf{u}_{-i} \sim \mathcal{N}\left(\frac{\sum_{j:j\sim i} u_j}{n_{\delta i}}, \frac{\sigma^2}{n_{\delta i}}\right), \tag{3}$$

   where $\mathbf{u}_{-i}$ is $\mathbf{u}$ with the ith element removed i.e. $(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_{\dim(\mathbf{u})})$, $j \sim i$ if the units i and j are defined as adjacent, $n_{\delta i} = |\{j : j \sim i\}|$ is the total number of adjacent units, and $\sigma > 0$ is the marginal standard deviation. We follow recommendations of Freni-Sterrantino, Ventrucci, and Rue (2018) on scaling of precision matrices, disconnected adjacency graph components, and islands.

3. For the reparameterised Besag-York-Mollie model [BYM2; Simpson et al. (2017)], we write $\mathbf{u} \sim \text{BYM2}(\sigma, \phi)$, where $\mathbf{u}$ is comprised of a spatially structured ICAR component $\mathbf{v}$ with proportion $\phi \in (0, 1)$ and spatially unstructured IID component $\mathbf{w}$ with proportion $1 - \phi$, both scaled to have generalised variance equal to one, and $\sigma > 0$ is the marginal standard deviation such that

$$\mathbf{u} = \sigma\left(\sqrt{\phi} \cdot \mathbf{v} + \sqrt{1 - \phi} \cdot \mathbf{w}\right). \tag{4}$$

S1.1.3 Complex survey design

We assume the household survey was run according to a complex survey design where each individual j in the population U has non-zero probability $\pi_j \in (0, 1)$ of appearing in the sample $S \subseteq U$. Suppose we observe an outcome $\theta_j \in \{0, 1\}$ for $j \in S$. Let $w_j = 1/\pi_j \times 1/\omega_j$ be design weights, where ω_j is a non-response factor, then a survey weighted mean is given by

$$\hat{\theta} = \frac{\sum_{j \in S} w_j \theta_j}{\sum_{j \in S} w_j}. \tag{5}$$

We account for survey weighting in the variance via the Kish effective sample size (Kish 1965)

$$m^\theta = \frac{\left(\sum_{j \in S} w_j\right)^2}{\sum_{j \in S} w_j^2}. \tag{6}$$

The observed number of indicator cases is then $y^\theta = m^\theta \cdot \hat{\theta}$. To make these calculations we used the `survey` R package (Lumley 2004)

S1.2 Process specification

	Model component	Latent field	Hyperparameter
S1.2.1	HIV prevalence	$22 + 5n$	9
S1.2.2	ART coverage	$25 + 5n$	9
S1.2.3	HIV incidence rate	$2 + n$	3
S1.2.4	ANC testing	$2 + 2n$	2
S1.2.5	ART attendance	n	1
	Total	$51 + 14n$	24

Table S1: The numer of latent field parameters and hyperparameters in each model section.

We now describe the hyperparameter and latent field process specification for the model. Whereas in the main text process and likelihood specifications are given together, here we consider the likelihood equations separately (Section S1.3). Table S1 gives the number of latent field parameters and hyperparameters in each component of the model. In the case of Malawi then $n = 32$ such that the total number of latent field parameters is $51 + 14 \cdot 32 = 499$ and the total number of hyperparameters is 24.

S1.2.1 HIV prevalence

We model HIV prevalence $\rho_{x,s,a} \in [0, 1]$ on the logit scale using the linear predictor

$$\text{logit}(\rho_{x,s,a}) = \beta_0^\rho + \beta_S^{\rho,s=\text{M}} + \mathbf{u}_a^\rho + \mathbf{u}_a^{\rho,s=\text{M}} + \mathbf{u}_x^\rho + \mathbf{u}_x^{\rho,s=\text{M}} + \mathbf{u}_x^{\rho,a<15} + \boldsymbol{\eta}_{R_x,s,a}^\rho. \tag{7}$$

Table S2 provides a description of the terms included in Equation 7.

The two BYM2 random effects $\mathbf{u}_x^\rho$ and $\mathbf{u}_x^{\rho,s=\text{M}}$ are comprised of the following unit scaled spatially structured $\{\mathbf{v}_x^\rho, \mathbf{v}_x^{\rho,s=\text{M}}\}$ and spatially unstructured $\{\mathbf{w}_x^\rho, \mathbf{w}_x^{\rho,s=\text{M}}\}$ components, respectively

$$\mathbf{u}_x^\rho = \sigma_X^\rho \left(\sqrt{\phi_X^\rho} \cdot \mathbf{v}_x^\rho + \sqrt{1 - \phi_X^\rho} \cdot \mathbf{w}_x^\rho\right), \tag{8}$$

$$\mathbf{u}_x^{\rho,s=\text{M}} = \sigma_{XS}^\rho \left(\sqrt{\phi_{XS}^\rho} \cdot \mathbf{v}_x^{\rho,s=\text{M}} + \sqrt{1 - \phi_{XS}^\rho} \cdot \mathbf{w}_x^{\rho,s=\text{M}}\right). \tag{9}$$

We use half-normal priors for the standard deviation terms

$$\{\sigma_A^\rho, \sigma_{AS}^\rho, \sigma_X^\rho, \sigma_{XS}^\rho, \sigma_{XA}^\rho\} \sim \mathcal{N}^+(0, 2.5), \tag{10}$$

Term	Distribution	Description
β_0^ρ	$\mathcal{N}(0, 5)$	Intercept
$\beta_s^{\rho,s=\text{M}}$	$\mathcal{N}(0, 5)$	The difference in logit prevalence for men compared to women
$\mathbf{u}_a^\rho$	$\text{AR1}(\sigma_A^\rho, \phi_A^\rho)$	Age random effects for women
$\mathbf{u}_a^{\rho,s=\text{M}}$	$\text{AR1}(\sigma_{AS}^\rho, \phi_{AS}^\rho)$	Age random effects for the difference in logit prevalence for men compared to women age a
$\mathbf{u}_x^\rho$	$\text{BYM2}(\sigma_X^\rho, \phi_X^\rho)$	Spatial random effects for women
$\mathbf{u}_x^{\rho,s=\text{M}}$	$\text{BYM2}(\sigma_{XS}^\rho, \phi_{XS}^\rho)$	Spatial random effects for the difference in logit prevalence for men compared to women in district x
$\mathbf{u}_x^{\rho,a<15}$	$\text{ICAR}(\sigma_{XA}^\rho)$	Spatial random effects for the ratio of paediatric prevalence to adult women prevalence
$\boldsymbol{\eta}_{R_x,s,a}^\rho$	$-$	Fixed offsets specifying assumed odds ratios for prevalence outside the age ranges for which data are available

Table S2: Terms included in the linear predictor for HIV prevalence (Equation 7).

uniform priors for the AR1 correlation parameters

$$\{\phi_A^\rho, \phi_{AS}^\rho\} \sim \mathcal{U}(-1, 1), \tag{11}$$

and beta priors for the BYM2 proportion parameters

$$\{\phi_X^\rho, \phi_{XS}^\rho\} \sim \text{Beta}(0.5, 0.5). \tag{12}$$

S1.2.2 ART coverage

We model ART coverage $\alpha_{x,s,a} \in [0, 1]$ on the logit scale using the linear predictor

$$\text{logit}(\alpha_{x,s,a}) = \beta_0^\alpha + \beta_S^{\alpha,s=\text{M}} + \mathbf{u}_a^\alpha + \mathbf{u}_a^{\alpha,s=\text{M}} + \mathbf{u}_x^\alpha + \mathbf{u}_x^{\alpha,s=\text{M}} + \mathbf{u}_x^{\alpha,a<15} + \boldsymbol{\eta}_{R_x,s,a}^\alpha \tag{13}$$

with terms and priors analogous to the HIV prevalence process model in Section S1.2.1 above.

S1.2.3 HIV incidence rate

We model HIV incidence rate $\lambda_{x,s,a} > 0$ on the log scale using the linear predictor

$$\log(\lambda_{x,s,a}) = \beta_0^\lambda + \beta_S^{\lambda,s=\text{M}} + \log(\rho_x^{\text{15-49}}) + \log(1 - \omega \cdot \alpha_x^{\text{15-49}}) + \mathbf{u}_x^\lambda + \boldsymbol{\eta}_{R_x,s,a}^\lambda. \tag{14}$$

Table S3 provides a description of the terms included in Equation 14.

We model the proportion recently infected among HIV positive persons $\kappa_{x,s,a} \in [0, 1]$ as

$$\kappa_{x,s,a} = 1 - \exp\left(-\lambda_{x,s,a} \cdot \frac{1 - \rho_{x,s,a}}{\rho_{x,s,a}} \cdot (\Omega_T - \beta_T) - \beta_T\right), \tag{15}$$

where $\Omega_T \sim \mathcal{N}(\Omega_{T_0}, \sigma^{\Omega_T})$ is the mean duration of recent infection, and $\beta_T \sim \mathcal{N}^+(\beta_{T_0}, \sigma^{\beta_T})$ is the false recent ratio. We use an informative prior for Ω_T based on the characteristics of the recent infection testing algorithm. For PHIA surveys this is $\Omega_{T_0} = 130$ days and $\sigma^{\Omega_T} = 6.12$ days, and further there is assumed to be no false recency, such that $\beta_{T_0} = 0.0$ and $\sigma^{\beta_T} = 0.0$.

S1.2.4 ANC testing

HIV prevalence $\rho_{x,a}^{\text{ANC}}$ and ART coverage $\alpha_{x,a}^{\text{ANC}}$ among pregnant women are modelled as being offset on the logit scale from the corresponding district-age indicators $\rho_{x,F,a}$ and $\alpha_{x,F,a}$ according to

$$\text{logit}(\rho_{x,a}^{\text{ANC}}) = \text{logit}(\rho_{x,F,a}) + \beta^{\rho^{\text{ANC}}} + \mathbf{u}_x^{\rho^{\text{ANC}}} + \boldsymbol{\eta}_{R_x,a}^{\rho^{\text{ANC}}}, \tag{16}$$

$$\text{logit}(\alpha_{x,a}^{\text{ANC}}) = \text{logit}(\alpha_{x,F,a}) + \beta^{\alpha^{\text{ANC}}} + \mathbf{u}_x^{\alpha^{\text{ANC}}} + \boldsymbol{\eta}_{R_x,a}^{\alpha^{\text{ANC}}}. \tag{17}$$

Term	Distribution	Description
β_0^λ	$\mathcal{N}(0,5)$	Intercept term proportional to the average HIV transmission rate for untreated HIV positive adults
$\beta_S^{\lambda, s=\text{M}}$	$\mathcal{N}(0,5)$	The log incidence rate ratio for men compared to women
$\rho_x^{\text{15-49}} = \frac{\sum_{s\in\{\text{F,M}\}}\sum_{a=15}^{45} N_{x,s,a}\cdot\rho_{x,s,a}}{\sum_{s\in\{\text{F,M}\}}\sum_{a=15}^{45} N_{x,s,a}}$	–	The HIV prevalence among adults 15-49 calculated by aggregating age-specific HIV prevalences
$\alpha_x^{\text{15-49}} = \frac{\sum_{s\in\{\text{F,M}\}}\sum_{a=15}^{45} N_{x,s,a}\cdot\rho_{x,s,a}\cdot\alpha_{x,s,a}}{\sum_{s\in\{\text{F,M}\}}\sum_{a=15}^{45} N_{x,s,a}\cdot\rho_{x,s,a}}$	–	The ART coverage among adults 15-49 calculated by aggregating age-specific ART coverages
$\omega = 0.7$	–	Average reduction in HIV transmission rate per increase in population ART coverage fixed based on inputs to the Estimation and Projection Package (EPP) model
$\mathbf{u}_x^\lambda$	$\mathcal{N}(0,\sigma^\lambda)$	IID spatial random effects with $\sigma^\lambda \sim \mathcal{N}^+(0,1)$
$\boldsymbol{\eta}_{R_x,s,a}^\lambda$	–	Fixed log incidence rate ratios by sex and age group calculated from Spectrum model output

Table S3: Terms included in the linear predictor for HIV incidence (Equation 14). Note that the only source age structure for this part of the model are $\boldsymbol{\eta}_{R_x,s,a}^\lambda$. As Spectrum assumes that there are no new infections in children aged 5-9 or 10-14, or adults aged over 80, the posterior over new infections in these age groups is exactly zero, by definition. We remove these identically zero posteriors from any later inference comparisons.

Term	Distribution	Description
$\beta^{\theta^{\text{ANC}}}$	$\mathcal{N}(0,5)$	Intercept giving the average difference between population and ANC outcomes
$\mathbf{u}_x^{\theta^{\text{ANC}}}$	$\mathcal{N}(0,\sigma_X^{\theta^{\text{ANC}}})$	IID district random effects with $\sigma_X^{\theta^{\text{ANC}}} \sim \mathcal{N}^+(0,1)$
$\boldsymbol{\eta}_{R_x,a}^{\theta^{\text{ANC}}}$	–	Offsets for the log fertility rate ratios for HIV positive women compared to HIV negative women and for women on ART to HIV positive women not on ART, calculated from Spectrum model outputs for region R_x

Table S4: Terms included in the linear predictors for ANC testing, with $\theta \in \{\rho, \alpha\}$ (Equation 7).

Table S4 provides a description of the terms included in Equation 16 and Equation 17.

In the full Naomi model, for adult women 15-49 the number of ANC clients $\Psi_{x,a} > 0$ are modelled as

$$\log(\Psi_{x,a}) = \log(N_{x,\text{F},a}) + \psi_{R_x,a} + \beta^\psi + \mathbf{u}_x^\psi,$$

where $N_{x,\text{F},a}$ are the female population sizes, $\psi_{R_x,a}$ are fixed age-sex fertility ratios in Spectrum region R_x, β^ψ are log rate ratios for the number of ANC clients relative to the predicted fertility, and $\mathbf{u}_x^\psi \sim \mathcal{N}(0,\sigma^\psi)$ are district random effects. Here we fix $\beta^\psi = 0$ and $\mathbf{u}_x^\psi = \mathbf{0}$ such that $\Psi_{x,a}$ are simply constants.

S1.2.5 ART attendance

Let $\gamma_{x,x'} \in [0,1]$ be the probability that a person on ART residing in district x receives ART in district x'. We assume that $\gamma_{x,x'} = 0$ for $x \notin \{x, \text{ne}(x)\}$ such that individuals seek treatment only in their residing district or its neighbours $\text{ne}(x) = \{x' : x' \sim x\}$, where $\sim$ is an adjacency relation, and $\sum_{x'\in\{x,\text{ne}(x)\}} \gamma_{x,x'} = 1$. To model $\gamma_{x,x'}$ for $x \sim x'$ we use a multinomial logistic regression model, based on the log-odds ratios

$$\tilde{\gamma}_{x,x'} = \log\left(\frac{\gamma_{x,x'}}{1-\gamma_{x,x'}}\right) = \tilde{\gamma}_0 + \mathbf{u}_x^{\tilde{\gamma}}, \tag{18}$$

where $\tilde{\gamma}_0 = -4$ is a fixed intercept, and $\mathbf{u}_x^{\tilde{\gamma}} \sim \mathcal{N}(0, \sigma_X^{\tilde{\gamma}})$ are district random effects with $\sigma_X^{\tilde{\gamma}} \sim \mathcal{N}^+(0, 2.5)$. Note that Equation 18 does not depend on x', such that $\gamma_{x,x'}$ is only a function of x. Choice of $\tilde{\gamma}_0 = -4$ implies a prior mean on $\gamma_{x,x'}$ of 1.8%, such that $(100 - 1.8 \times \text{ne}(x))$% of ART clients in district x obtain treatment in their home district, a-priori. We fix $\tilde{\gamma}_{x,x} = 0$ and recover the multinomial probabilities using the softmax

$$\gamma_{x,x'} = \frac{\exp(\tilde{\gamma}_{x,x'})}{\sum_{x^\star \in \{x, \text{ne}(x)\}} \exp(\tilde{\gamma}_{x,x^\star})}. \tag{19}$$

Given the total number of PLHIV on ART $A_{x,s,a} = N_{x,s,a} \cdot \rho_{x,s,a} \cdot \alpha_{x,s,a}$, the number of ART clients who reside in district x and obtain ART in district x' are $A_{x,x',s,a} = A_{x,s,a} \cdot \gamma_{x,x'}$, and the total attending ART facilities in district x' are

$$\tilde{A}_{x',s,a} = \sum_{x \in \{x', \text{ne}(x')\}} A_{x,x',s,a}. \tag{20}$$

S1.3 Likelihood specification

S1.3.1 Household survey data

For HIV prevalence, ART coverage and recent HIV infections, denoted by $\theta \in \{\rho, \alpha, \kappa\}$, the most recent household survey furnishes weighted observations $\hat{\theta}_{\{x\},\{s\},\{a\}}$ with respective Kish effective sample sizes $m^\theta_{\{x\},\{s\},\{a\}} \in \mathbb{R}$, and observed number of cases

$$y^\theta_{\{x\},\{s\},\{a\}} = m^\theta_{\{x\},\{s\},\{a\}} \cdot \hat{\theta}_{\{x\},\{s\},\{a\}} \in \mathbb{R}. \tag{21}$$

To model these observations, we use the following three binomial working likelihoods:

1. The number of people living with HIV (PLHIV) is

$$y^\rho_{\{x\},\{s\},\{a\}} \sim \text{xBin}(m^\rho_{\{x\},\{s\},\{a\}}, \rho_{\{x\},\{s\},\{a\}}), \tag{22}$$

   where

$$\rho_{\{x\},\{s\},\{a\}} = \frac{\sum_{\{x\}} \sum_{\{s\}} \sum_{\{a\}} N_{x,s,a} \cdot \rho_{x,s,a}}{\sum_{\{x\}} \sum_{\{s\}} \sum_{\{a\}} N_{x,s,a}} \tag{23}$$

   has denominator given by the total population size.
2. The number of PLHIV on ART is

$$y^\alpha_{\{x\},\{s\},\{a\}} \sim \text{xBin}(m^\alpha_{\{x\},\{s\},\{a\}}, \alpha_{\{x\},\{s\},\{a\}}), \tag{24}$$

   where

$$\alpha_{\{x\},\{s\},\{a\}} = \frac{\sum_{\{x\}} \sum_{\{s\}} \sum_{\{a\}} N_{x,s,a} \cdot \rho_{x,s,a} \cdot \alpha_{x,s,a}}{\sum_{\{x\}} \sum_{\{s\}} \sum_{\{a\}} N_{x,s,a} \cdot \rho_{x,s,a}} \tag{25}$$

   has denominator given by the total PLHIV.
3. The number of PLHIV recently infected

$$y^\kappa_{\{x\},\{s\},\{a\}} \sim \text{xBin}(m^\kappa_{\{x\},\{s\},\{a\}}, \kappa_{\{x\},\{s\},\{a\}}), \tag{26}$$

   where

$$\kappa_{\{x\},\{s\},\{a\}} = \frac{\sum_{\{x\}} \sum_{\{s\}} \sum_{\{a\}} N_{x,s,a} \cdot \rho_{x,s,a} \cdot \kappa_{x,s,a}}{\sum_{\{x\}} \sum_{\{s\}} \sum_{\{a\}} N_{x,s,a} \cdot \rho_{x,s,a}} \tag{27}$$

   has denominator given by the total PLHIV.

The generalised binomial $y \sim \text{xBin}(m, p)$ used above is defined for $y, m \in \mathbb{R}^+$ with $y \leq m$ such that

$$\log p(y) = \log \Gamma(m+1) - \log \Gamma(y+1) - \log \Gamma(m-y+1) + y \log p + (m-y) \log(1-p), \tag{28}$$

where the gamma function Γ is such that $\forall n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$.

S1.3.2 ANC testing data

We include ANC testing data for the year of the most recent survey. Let $W^{\text{ANC}}_{\{x\}}$ be the number of ANC clients, $X^{\text{ANC}}_{\{x\}}$ the number of those with ascertained status, $Y^{\text{ANC}}_{\{x\}}$ the number of those with positive status (either known or tested) and $Z^{\text{ANC}}_{\{x\}}$ the number of ANC clients already on ART prior to first ANC, such that

$$W^{\text{ANC}}_{x} \geq X^{\text{ANC}}_{x} \geq Y^{\text{ANC}}_{x} \geq Z^{\text{ANC}}_{x}, \tag{29}$$

for all $x \in \{x\}$. When ANC testing data are only available for part of a given year, we denote $m^{\text{ANC}} \in \{1, \ldots, 12\}$ the number of months of reported data reflected in counts for that year. The observed number of HIV positive and already on ART among ANC clients is modelled by

$$Y^{\text{ANC}}_{\{x\}} \sim \text{Bin}\left(X^{\text{ANC}}_{\{x\}}, \rho^{\text{ANC}}_{\{x\},\{15,\ldots 45\}}\right), \tag{30}$$

$$Z^{\text{ANC}}_{\{x\}} \sim \text{Bin}\left(Y^{\text{ANC}}_{\{x\}}, \alpha^{\text{ANC}}_{\{x\},\{15,\ldots 45\}}\right), \tag{31}$$

where prevalence and ART coverage are aggregated by the number of pregnant women $\Psi_{x,a}$

$$\rho^{\text{ANC}}_{\{x\}\{a\}} = \frac{\sum_{\{x\}} \sum_{\{a\}} \Psi_{x,a} \cdot \rho^{\text{ANC}}_{x,a}}{\sum_{\{x\}} \sum_{\{a\}} \Psi_{x,a}}, \tag{32}$$

$$\alpha^{\text{ANC}}_{\{x\}\{a\}} = \frac{\sum_{\{x\}} \sum_{\{a\}} \Psi_{x,a} \cdot \rho^{\text{ANC}}_{x,a} \cdot \alpha^{\text{ANC}}_{x,a}}{\sum_{\{x\}} \sum_{\{a\}} \Psi_{x,a} \cdot \rho^{\text{ANC}}_{x,a}}. \tag{33}$$

S1.3.3 Number receiving ART

Let $\dot{A}_{\{x\},\{s\},\{a\}}$ be data for the number receiving ART

$$\dot{A}_{\{x\},\{s\},\{a\}} = \sum_{\{x\}} \sum_{\{s\}} \sum_{\{a\}} \sum_{x \sim x', x = x'} \dot{A}_{x',x,s,a}.$$

We model the unobserved numbers of ART clients travelling from x' to x as

$$\dot{A}_{x',x,s,a} \sim \text{Bin}(N_{x',s,a}, \pi_{x',x,s,a})$$

where $\pi_{x',x,s,a} = \rho_{x',s,a} \cdot \alpha_{x',s,a} \cdot \gamma_{x',x,s,a}$. This likelihood is approximated using a normal for the sum of binomials by

$$\dot{A}_{\{x\},\{s\},\{a\}} \sim \mathcal{N}(\tilde{A}_{\{x\},\{s\},\{a\}}, \sigma^{\tilde{A}}_{\{x\},\{s\},\{a\}})$$

where the mean is

$$\tilde{A}_{\{x\},\{s\}\{a\}} = \sum_{\{x\}} \sum_{\{s\}} \sum_{\{a\}} \sum_{x \sim x', x = x'} N_{x',s,a} \cdot \pi_{x',x,s,a},$$

and the variance is

$$\left(\sigma^{\tilde{A}}_{\{x\},\{s\}\{a\}}\right)^2 = \sum_{\{x\}} \sum_{\{s\}} \sum_{\{a\}} \sum_{x \sim x', x = x'} N_{x',s,a} \cdot \pi_{x',x,s,a} \cdot (1 - \pi_{x',x,s,a}).$$

S1.4 Identifiability constaints

If data are missing, some parameters are fixed to default values to help with identifiability. In particular:

1. If survey data on HIV prevalence or ART coverage by age and sex are not available then we set $\mathbf{u}^{\theta}_{a} = 0$ and $\mathbf{u}^{\theta}_{a,s=\text{M}} = 0$ and use the average age-sex pattern of from the Spectrum offset $\boldsymbol{\eta}^{\theta}_{R_x,s,a}$. For the Malawi example considered in the main text HIV prevalence and ART coverage data are not available for those aged 65+. As a result, there are $|\{0\text{-}4, \ldots, 50\text{-}54\}| = 13$ age groups included for the age random effects.

2. If no ART data, either survey or ART programme, are available but data on ART coverage among ANC clients are available, the level of ART coverage is not identifiable, but spatial variation is identifiable. In this instance, overall ART coverage is determined by the Spectrum offset, and only area random effects are estimated such that $\text{logit}\,(\alpha_{x,s,a}) = \mathbf{u}_x^\alpha + \boldsymbol{\eta}_{R_x,s,a}^\alpha$.
3. If survey data on recent HIV infection are not included in the model, then $\beta_0^\lambda = \beta_S^{\lambda,s=\mathrm{M}} = 0$ and $\mathbf{u}_x^\lambda = \mathbf{0}$. The sex ratio for HIV incidence is determined by the sex incidence rate ratio from Spectrum in the same years and the incidence rate in all districts is modelled assuming the same average HIV transmission rate for untreated adults, but varies according to district estimates of HIV prevalence and ART coverage.

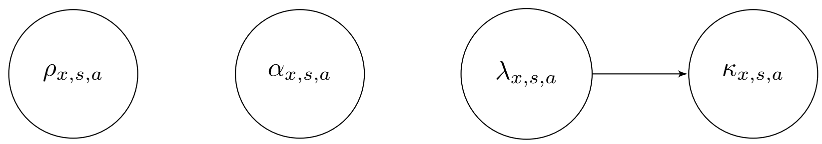

Figure S1: Directed acyclic graph describing the simplified Naomi model (work in progress).

S1.5 Implementation

The C++ `TMB` code for the negative log-posterior of the simplified Naomi model is available from the GitHub repository `athowes/naomi-aghq`. For ease of understanding, Table S5 provides correspondence between the mathematical notation used in Section S1 and the variable names used in the `TMB` code, for all hyperparameters and latent field parameters. For further reference on the `TMB` software see Kristensen (2021).

Variable name	Notation	Type	Size	Domain	ρ input?	α input?	λ input?
`logit_phi_rho_x`	$\mathrm{logit}(\phi_X^\rho)$	Hyper	1	$\mathbb{R}$	✓		
`log_sigma_rho_x`	$\log(\sigma_X^\rho)$	Hyper	1	$\mathbb{R}$	✓		
`logit_phi_rho_xs`	$\mathrm{logit}(\phi_{XS}^\rho)$	Hyper	1	$\mathbb{R}$	✓		
`log_sigma_rho_xs`	$\log(\sigma_{XS}^\rho)$	Hyper	1	$\mathbb{R}$	✓		
`logit_phi_rho_a`	$\mathrm{logit}(\phi_A^\rho)$	Hyper	1	$\mathbb{R}$	✓		
`log_sigma_rho_a`	$\log(\sigma_A^\rho)$	Hyper	1	$\mathbb{R}$	✓		
`logit_phi_rho_as`	$\mathrm{logit}(\phi_{AS}^\rho)$	Hyper	1	$\mathbb{R}$	✓		
`log_sigma_rho_as`	$\log(\sigma_{AS}^\rho)$	Hyper	1	$\mathbb{R}$	✓		
`log_sigma_rho_xa`	$\log(\sigma_{XA}^\rho)$	Hyper	1	$\mathbb{R}$	✓		
`logit_phi_alpha_x`	$\mathrm{logit}(\phi_X^\alpha)$	Hyper	1	$\mathbb{R}$		✓	
`log_sigma_alpha_x`	$\log(\sigma_X^\alpha)$	Hyper	1	$\mathbb{R}$		✓	
`logit_phi_alpha_xs`	$\mathrm{logit}(\phi_{XS}^\alpha)$	Hyper	1	$\mathbb{R}$		✓	
`log_sigma_alpha_xs`	$\log(\sigma_{XS}^\alpha)$	Hyper	1	$\mathbb{R}$		✓	
`logit_phi_alpha_a`	$\mathrm{logit}(\phi_A^\alpha)$	Hyper	1	$\mathbb{R}$		✓	
`log_sigma_alpha_a`	$\log(\sigma_A^\alpha)$	Hyper	1	$\mathbb{R}$		✓	
`logit_phi_alpha_as`	$\mathrm{logit}(\phi_{AS}^\alpha)$	Hyper	1	$\mathbb{R}$		✓	
`log_sigma_alpha_as`	$\log(\sigma_{AS}^\alpha)$	Hyper	1	$\mathbb{R}$		✓	
`log_sigma_alpha_xa`	$\log(\sigma_{XA}^\alpha)$	Hyper	1	$\mathbb{R}$		✓	
`OmegaT_raw`	Ω_T	Hyper	1	$\mathbb{R}$			✓
`log_betaT`	$\log(\beta_T)$	Hyper	1	$\mathbb{R}$			✓
`log_sigma_lambda_x`	$\log(\sigma^\lambda)$	Hyper	1	$\mathbb{R}$			✓
`log_sigma_ancrho_x`	$\log(\sigma_X^{\rho^{\mathrm{ANC}}})$	Hyper	1	$\mathbb{R}$			
`log_sigma_ancalpha_x`	$\log(\sigma_X^{\alpha^{\mathrm{ANC}}})$	Hyper	1	$\mathbb{R}$			
`log_sigma_or_gamma`	$\log(\sigma_X^{\tilde{\gamma}})$	Hyper	1	$\mathbb{R}$			
`beta_rho`	$(\beta_0^\rho, \beta_s^{\rho,s=\mathrm{M}})$	Latent	2	$\mathbb{R}^2$	✓		
`beta_alpha`	$(\beta_0^\alpha, \beta_S^{\alpha,s=\mathrm{M}})$	Latent	2	$\mathbb{R}^2$		✓	
`beta_lambda`	$(\beta_0^\lambda, \beta_S^{\lambda,s=\mathrm{M}})$	Latent	2	$\mathbb{R}^2$			✓
`beta_anc_rho`	$\beta^{\rho^{\mathrm{ANC}}}$	Latent	1	$\mathbb{R}$			
`beta_anc_alpha`	$\beta^{\alpha^{\mathrm{ANC}}}$	Latent	1	$\mathbb{R}$			
`u_rho_x`	$\mathbf{w}_x^\rho$	Latent	n	$\mathbb{R}^n$	✓		
`us_rho_x`	$\mathbf{v}_x^\rho$	Latent	n	$\mathbb{R}^n$	✓		
`u_rho_xs`	$\mathbf{w}_x^{\rho,s=\mathrm{M}}$	Latent	n	$\mathbb{R}^n$	✓		
`us_rho_xs`	$\mathbf{v}_x^{\rho,s=\mathrm{M}}$	Latent	n	$\mathbb{R}^n$	✓		
`u_rho_a`	$\mathbf{u}_a^\rho$	Latent	10	$\mathbb{R}^{10}$	✓		
`u_rho_as`	$\mathbf{u}_a^{\rho,s=\mathrm{M}}$	Latent	10	$\mathbb{R}^{10}$	✓		
`u_rho_xa`	$\mathbf{u}_x^{\rho,a<15}$	Latent	n	$\mathbb{R}^n$	✓		
`u_alpha_x`	$\mathbf{w}_x^\alpha$	Latent	n	$\mathbb{R}^n$		✓	
`us_alpha_x`	$\mathbf{v}_x^\alpha$	Latent	n	$\mathbb{R}^n$		✓	
`u_alpha_xs`	$\mathbf{w}_x^{\alpha,s=\mathrm{M}}$	Latent	n	$\mathbb{R}^n$		✓	
`us_alpha_xs`	$\mathbf{v}_x^{\alpha,s=\mathrm{M}}$	Latent	n	$\mathbb{R}^n$		✓	
`u_alpha_a`	$\mathbf{u}_a^\alpha$	Latent	13	$\mathbb{R}^{13}$		✓	
`u_alpha_as`	$\mathbf{u}_a^{\alpha,s=\mathrm{M}}$	Latent	10	$\mathbb{R}^{10}$		✓	
`u_alpha_xa`	$\mathbf{u}_x^{\alpha,a<15}$	Latent	n	$\mathbb{R}^n$		✓	
`ui_lambda_x`	$\mathbf{u}_x^\lambda$	Latent	n	$\mathbb{R}^n$			✓
`ui_anc_rho_x`	$\mathbf{u}_x^{\rho^{\mathrm{ANC}}}$	Latent	n	$\mathbb{R}^n$			
`ui_anc_alpha_x`	$\mathbf{u}_x^{\alpha^{\mathrm{ANC}}}$	Latent	n	$\mathbb{R}^n$			
`log_or_gamma`	$\mathbf{u}_x^{\tilde{\gamma}}$	Latent	n	$\mathbb{R}^n$			

Table S5: Correspondence between mathematical notation and variable names used in our `TMB` code. The total number of hyperparameters is 24, and the total number of latent field parameters is $51 + 14n$, where n is the number of districts. We use the notation ✓ to refer to direct dependence of the parameter on the variable, ✗ to refer to no dependence, and a blank entry to refer to dependence conditional on the data.

S2 Model assessment

S2.1 Additional figures

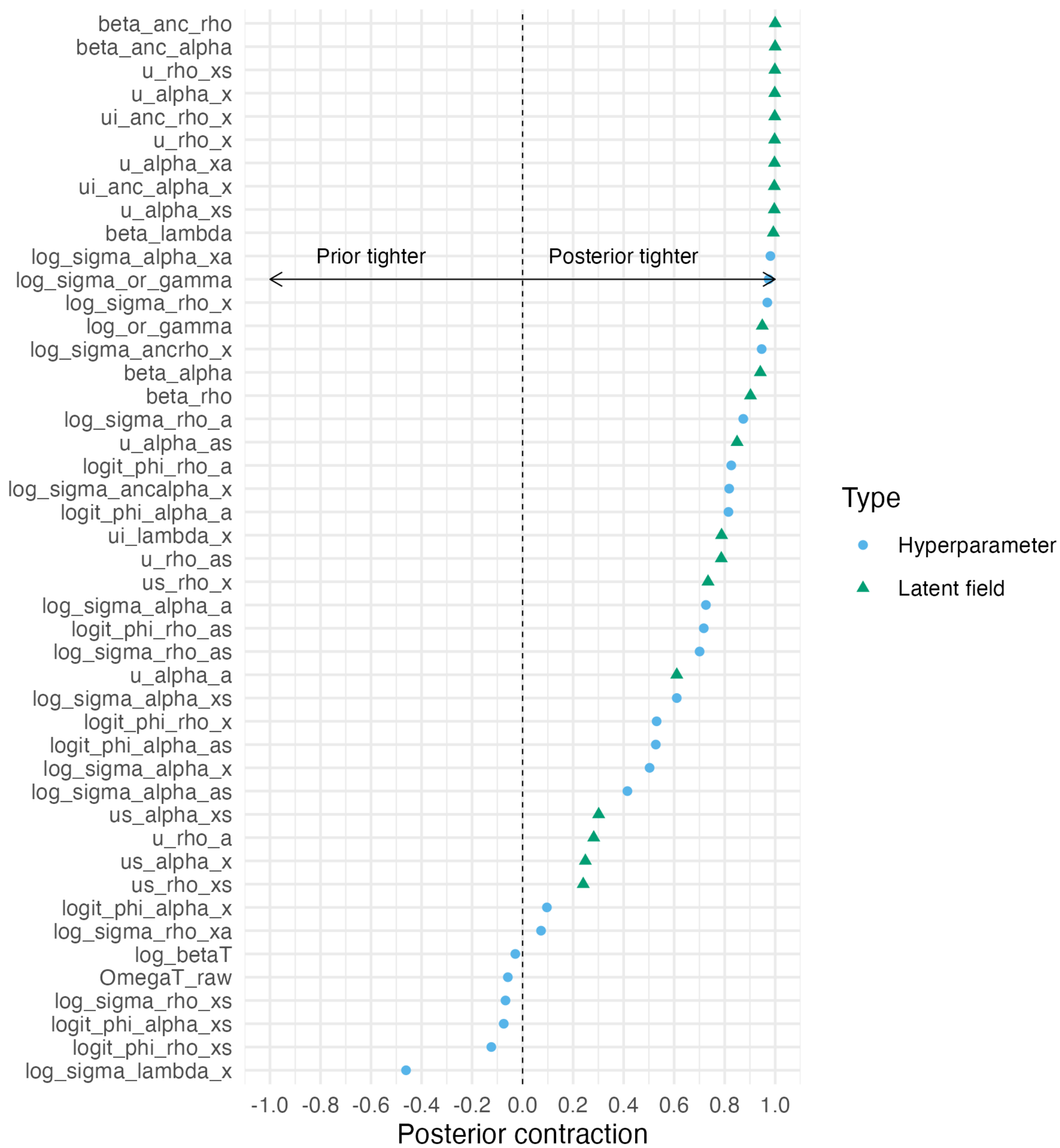

Figure S2: A posterior contraction of 1 corresponds to a Dirac delta function. A posterior contraction of 0 corresponds to no change in standard deviation between prior and posterior. A posterior contraction less than 0 corresponds to a wider posterior than prior.

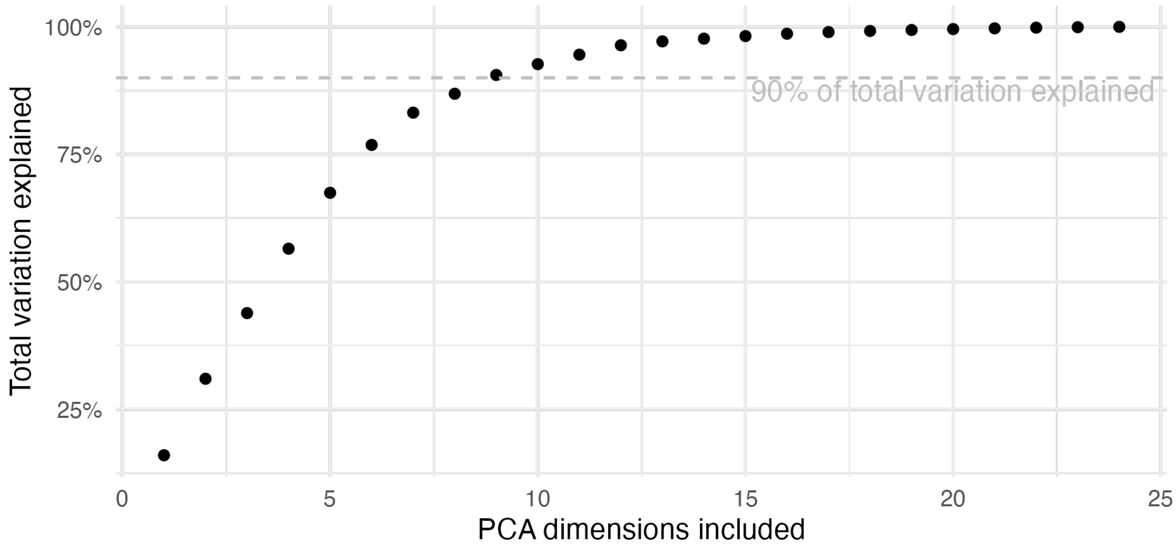

Figure S3: Most variation in the inverse curvature could be explained using far fewer than the full 24 dimensions.

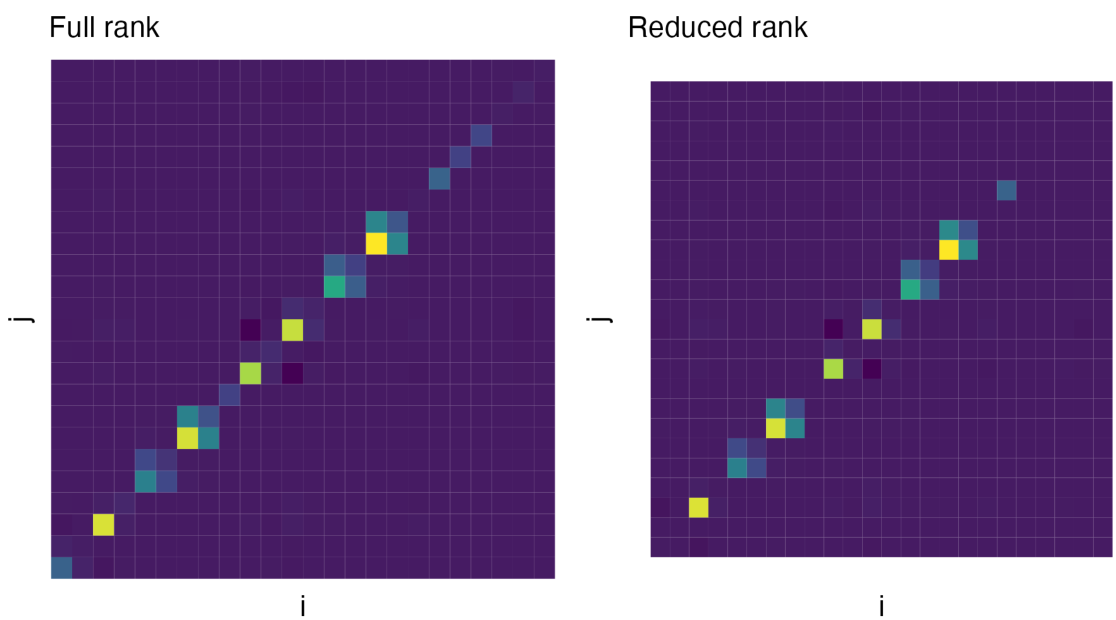

Figure S4: The reduced rank ($s = 8$) matrix approximation to the covariance matrix is visually similar to the full rank original.

S3 AGHQ and PCA-AGHQ details

S3.1 Additional figures

S3.2 Normalising constant estimation

Add here plots and description of the estimated normalising constant for different PCA-AGHQ settings.

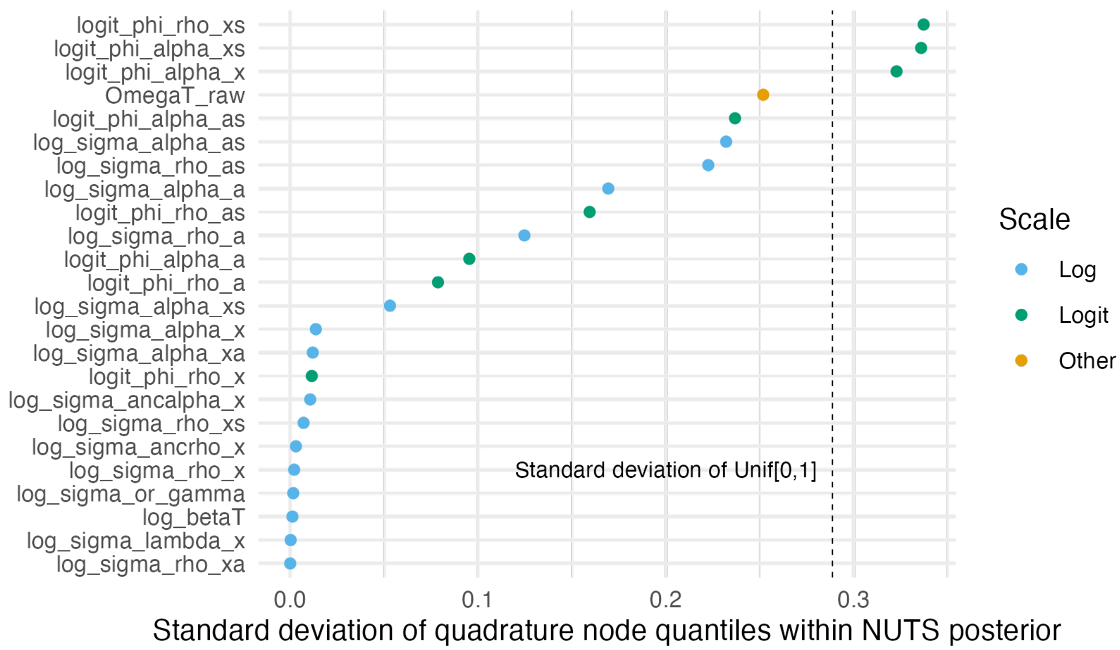

Figure S5: Standard deviations of the quantiles of the quadrature nodes within the NUTS posterior draws varied substantially, in accordance with the marginal standard deviations shown in Figure S6.

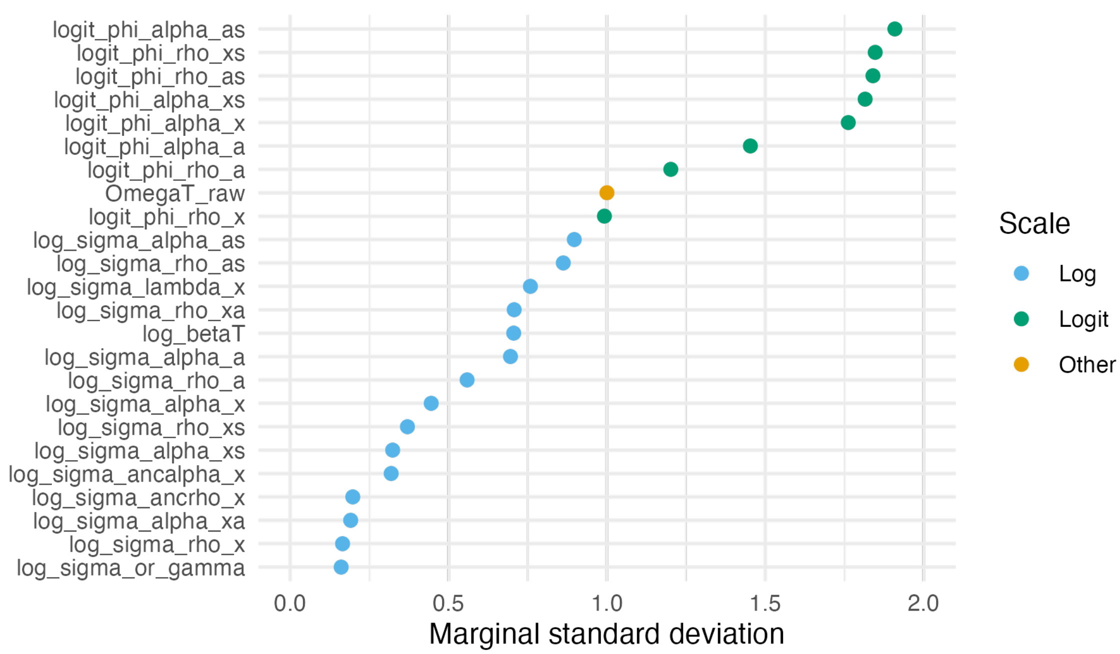

Figure S6: Hyperparameters on the logit-scale had systematically higher marginal standard deviations than those on the log-scale. This is because variation on the real scale is compressed by the inverse logit and expanded by the inverse log (exponential).

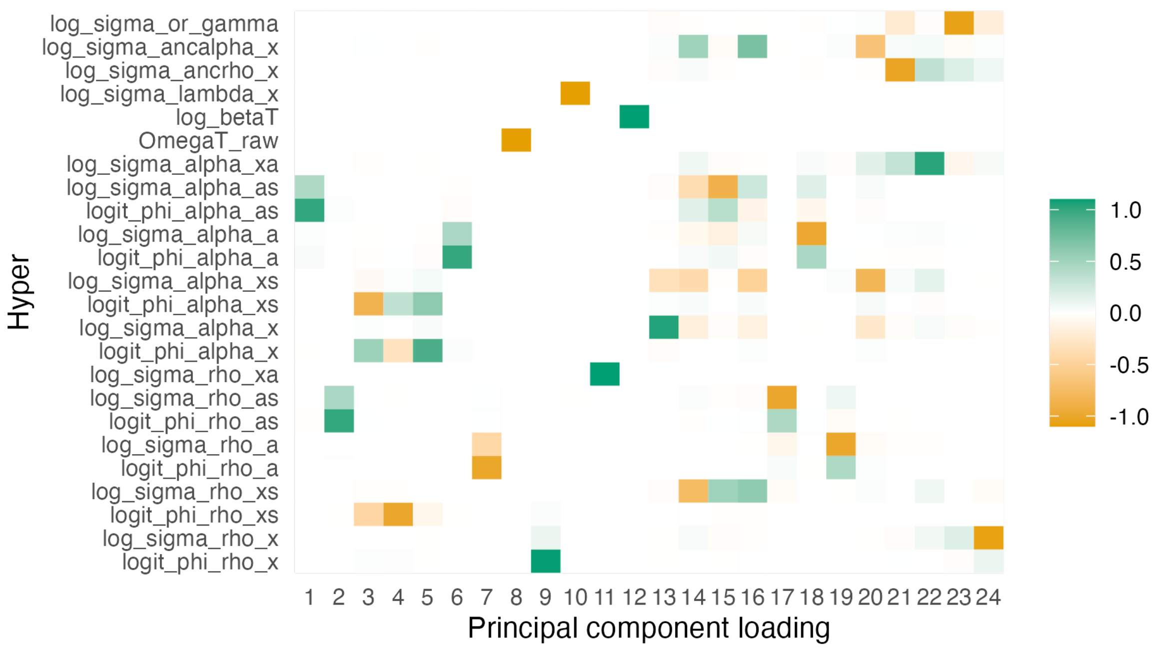

Figure S7: Principal component loadings based on the eigendecomposition of the inverse curvature.

S4 Inference comparison

In this section we give more detailed inference comparison results for each parameter or group of parameters. For any parameter with length greater than one, the results presented are an average.

S4.1 Point estimates

S4.1.1 Posterior mean

Parameter	TMB	PCA-AGHQ	Difference	Percent Difference
Output				
alpha_t1_out	0.0080	0.0063	−0.0018	−21.9%
lambda_t1_out	0.00035	0.00033	−0.000020	−5.6%
rho_t1_out	0.0012	0.0012	−0.000017	−1.4%
Latent				
beta_alpha	0.071	0.077	0.0057	8.0%
beta_anc_alpha	0.012	0.0096	−0.0028	−22.9%
beta_anc_rho	0.0097	0.013	0.0031	32.1%
beta_lambda	0.083	0.071	−0.012	−14.5%
beta_rho	0.19	0.17	−0.021	−10.9%
log_or_gamma	0.078	0.074	−0.0046	−5.9%
u_alpha_a	0.026	0.067	0.041	160.0%
u_alpha_as	0.086	0.070	−0.016	−18.9%
u_alpha_x	0.019	0.014	−0.0047	−24.7%
u_alpha_xa	0.018	0.016	−0.0020	−11.0%
u_alpha_xs	0.020	0.015	−0.0048	−24.1%
u_rho_a	0.26	0.19	−0.068	−26.1%
u_rho_as	0.089	0.16	0.068	76.6%
u_rho_x	0.018	0.017	−0.0011	−5.7%
u_rho_xs	0.019	0.016	−0.0034	−17.7%
ui_anc_alpha_x	0.013	0.012	−0.0011	−8.5%
ui_anc_rho_x	0.011	0.011	−0.00018	−1.6%
ui_lambda_x	0.015	0.020	0.0048	31.5%
us_alpha_x	0.14	0.10	−0.043	−30.1%
us_alpha_xs	0.15	0.092	−0.061	−40.1%
us_rho_x	0.065	0.064	−0.00072	−1.1%
us_rho_xs	0.056	0.050	−0.0060	−10.7%

S4.1.2 Posterior standard deviation

Parameter	TMB	PCA-AGHQ	Difference	Percent Difference
Output				
alpha_t1_out	0.0032	0.0035	0.00031	9.8%
lambda_t1_out	0.00016	0.00018	0.000022	14.0%
rho_t1_out	0.00047	0.00044	−0.000031	−6.7%
Latent				
beta_alpha	0.42	0.094	−0.32	−77.5%
beta_anc_alpha	0.0066	0.0080	0.0014	20.9%
beta_anc_rho	0.0014	0.0013	−0.00014	−10.0%
beta_lambda	0.021	0.025	0.0040	19.0%

beta_rho	0.42	0.14	−0.28	−67.1%
log_or_gamma	0.023	0.027	0.0034	14.6%
u_alpha_a	0.42	0.049	−0.37	−88.4%
u_alpha_as	0.40	0.12	−0.28	−71.1%
u_alpha_x	0.0098	0.010	0.00025	2.6%
u_alpha_xa	0.0082	0.0085	0.00034	4.2%
u_alpha_xs	0.024	0.023	−0.00078	−3.2%
u_rho_a	0.44	0.11	−0.33	−74.7%
u_rho_as	0.39	0.16	−0.23	−59.9%
u_rho_x	0.0069	0.0064	−0.00046	−6.7%
u_rho_xs	0.013	0.010	−0.0028	−21.6%
ui_anc_alpha_x	0.024	0.023	−0.00017	−0.7%
ui_anc_rho_x	0.0069	0.0076	0.00077	11.2%
ui_lambda_x	0.055	0.053	−0.0014	−2.6%
us_alpha_x	0.12	0.075	−0.041	−35.3%
us_alpha_xs	0.12	0.056	−0.069	−55.3%
us_rho_x	0.055	0.056	0.0012	2.2%
us_rho_xs	0.045	0.042	−0.0034	−7.4%

S4.2 Distribution tests

S4.2.1 Kolmogorov-Smirnov

	Parameter	TMB	PCA-AGHQ	Difference
	beta_alpha	0.093	0.077	−0.016
	beta_lambda	0.070	0.096	0.026
	beta_rho	0.069	0.073	0.004
	log_or_gamma	0.056	0.055	−0.001
	u_alpha_a	0.055	0.058	0.003
	u_alpha_as	0.103	0.083	−0.020
	u_alpha_x	0.117	0.118	0.001
	u_alpha_xa	0.069	0.069	0.000
	u_alpha_xs	0.092	0.081	−0.011
	u_rho_a	0.070	0.082	0.012
	u_rho_as	0.071	0.065	−0.006
	u_rho_x	0.069	0.070	0.001
	u_rho_xs	0.150	0.150	0.000
	ui_anc_alpha_x	0.076	0.074	−0.001
	ui_anc_rho_x	0.056	0.055	−0.001
	ui_lambda_x	0.114	0.045	−0.069
	us_alpha_x	0.088	0.085	−0.004
	us_alpha_xs	0.094	0.078	−0.016
	us_rho_x	0.084	0.080	−0.005
	us_rho_xs	0.041	0.044	0.003
Average	—	0.082	0.077	−0.005

S4.3 Pareto-smoothed importance sampling

S4.4 Maximum mean discrepancy

S5 MCMC convergence and suitability

We assessed MCMC convergence and suitability using a range of graphical and numerical tests. The largest scale reduction factor $\hat{R}$ was 1.02 (Figure S8). As such, all values of $\hat{R} < 1.05$. Even thinning by a factor of 20, samples were not obtained very efficiently, resulting in the majority of effective sample size (ESS) ratios being below 0.5, with some as low as 0.1 (Figure S9). As a result, the number of obtained ESS varied substantially by parameter (Figure S10): the minimum was 208.1, 2.5% quantile was 318.2, median was 1231, 97.5% quantile was 2776.3 and maximum was 4250.7. Traceplots for the parameters with the lowest ESS (`log_sigma_alpha_xs`) and highest $\hat{R}$ (the 10th index of `ui_lambda_x`) are shown in Figure S11. There were no divergent transitions.

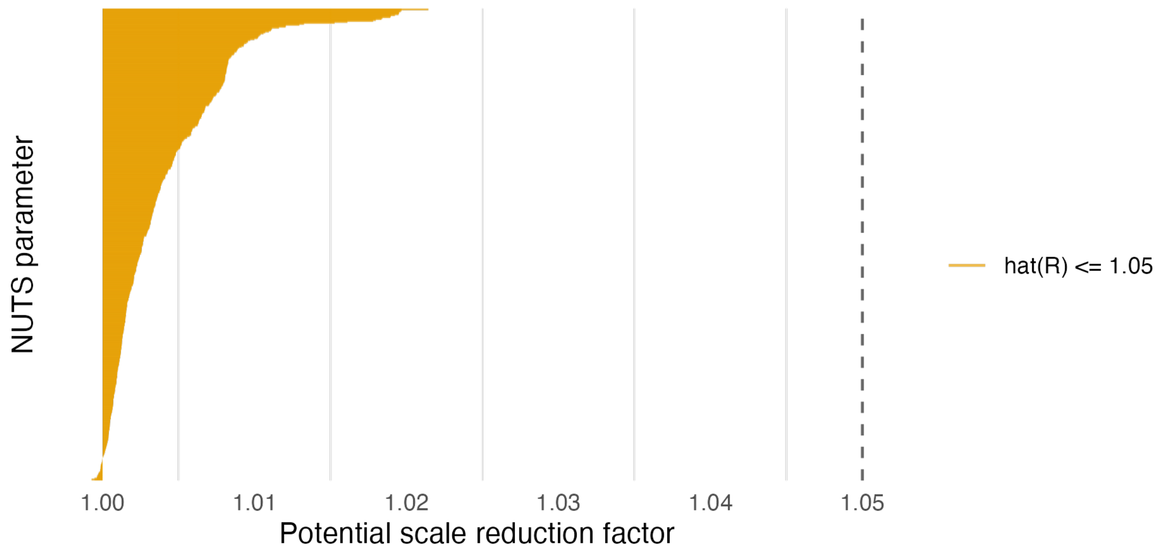

Figure S8: The potential scale reduction factor compares between- and within- estimates of univariate parameters. It is recommended only to use NUTS results if the value is less than 1.05, which it is for all parameters.

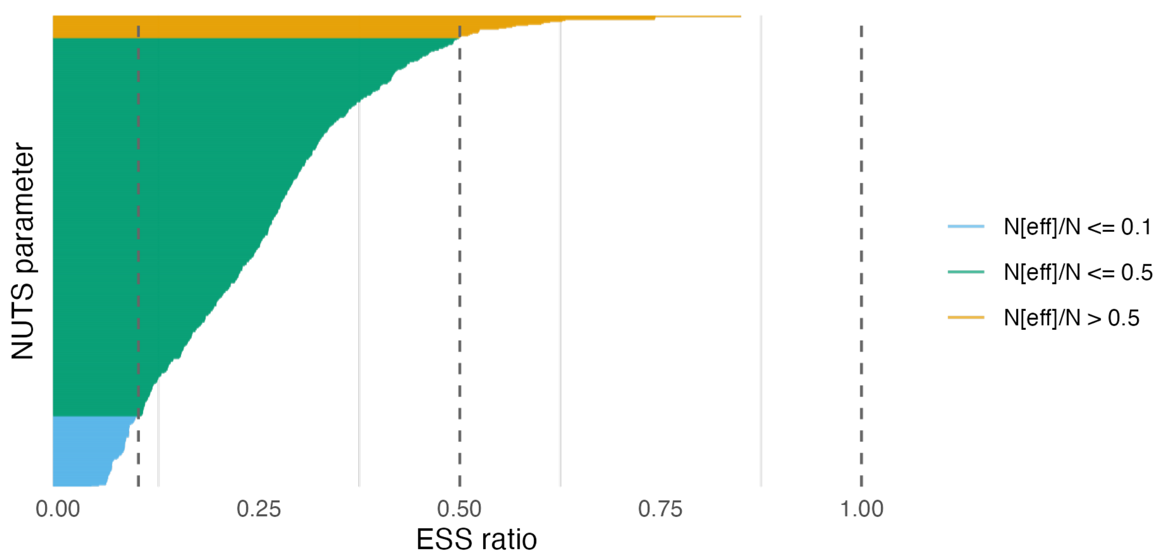

Figure S9: The efficiently, as measured by the ESS ratio, of the NUTS sampler was poor.

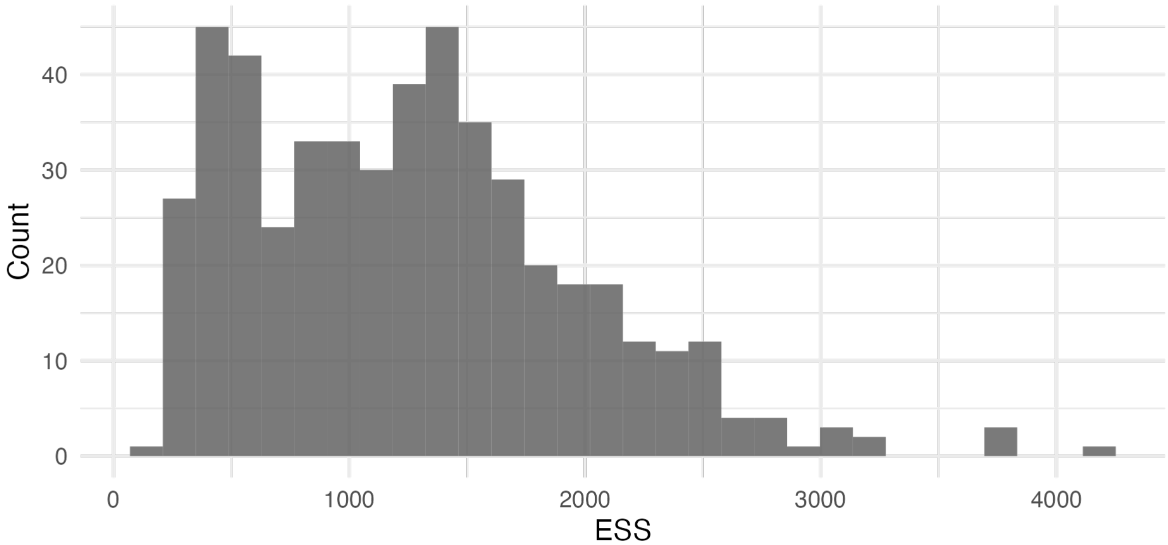

Figure S10: The effective number of samples we obtained varied substantially between parameters.

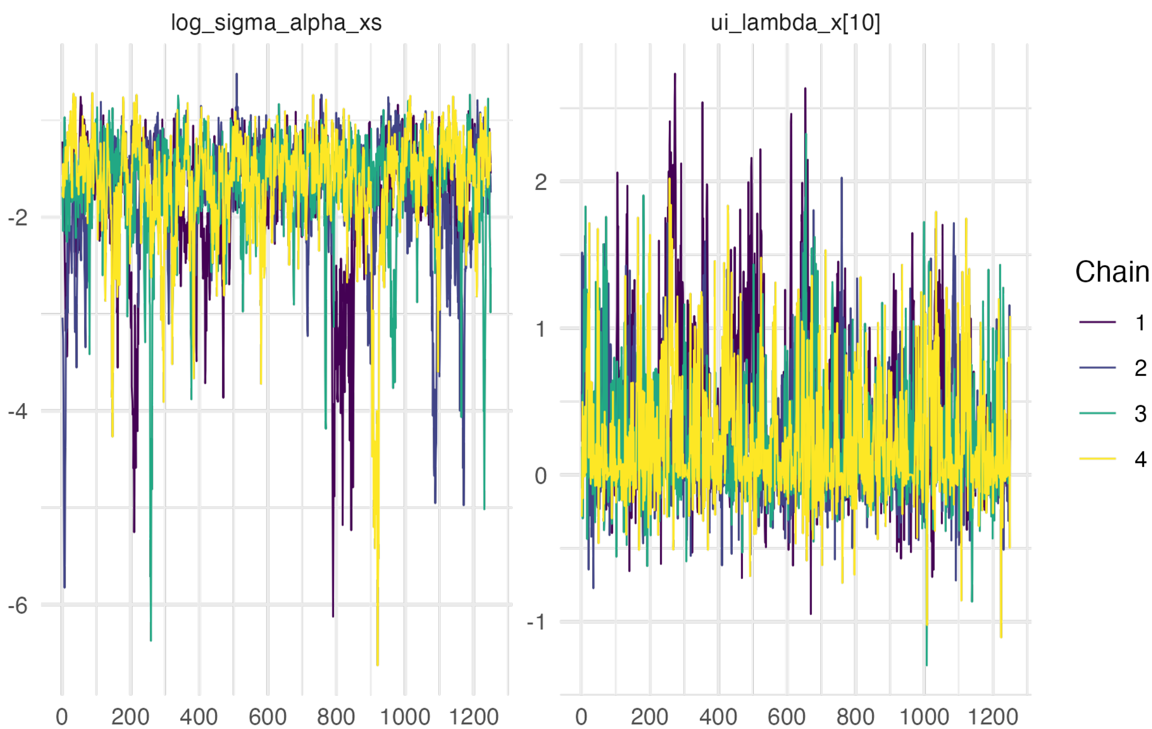

Figure S11: Traceplots of the parameters with the lowest ESS and highest scale reduction factor.

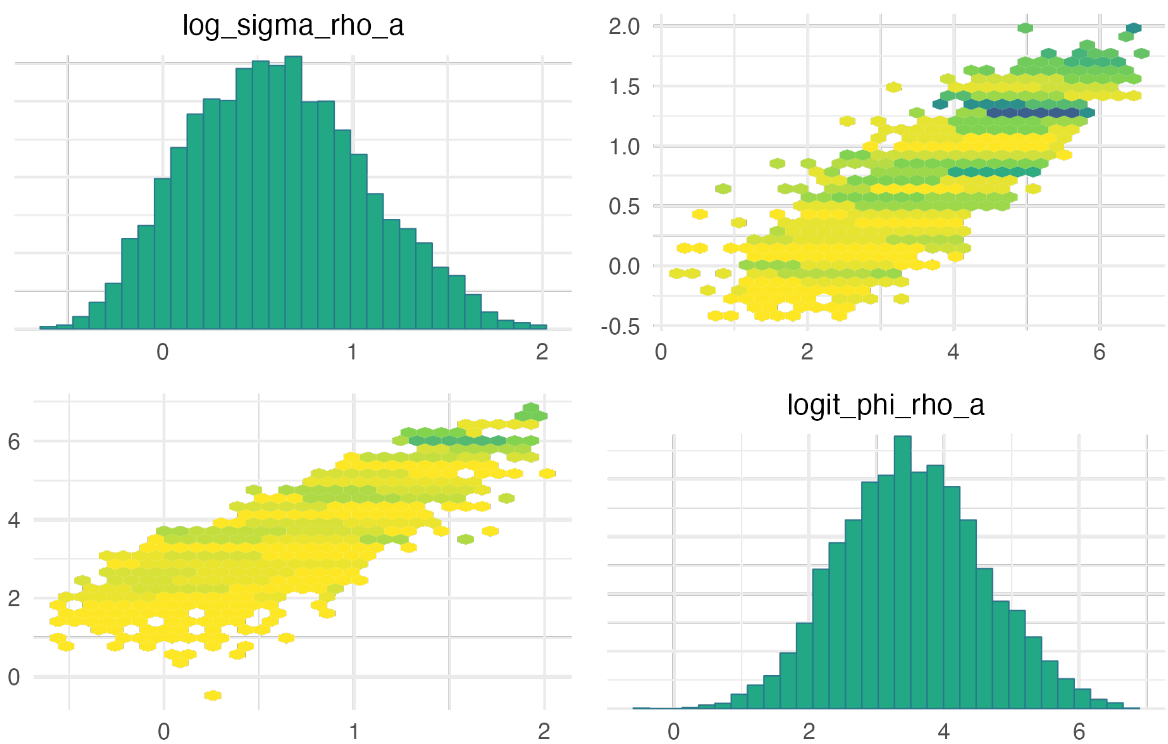

Figure S12: Variation between units can be explained either by high correlation and high variance, or by low correlation and low variance. As such, the AR1 hyperparameters had correlated posteriors. It is this correlation which we made use of with PCA-AGHQ.

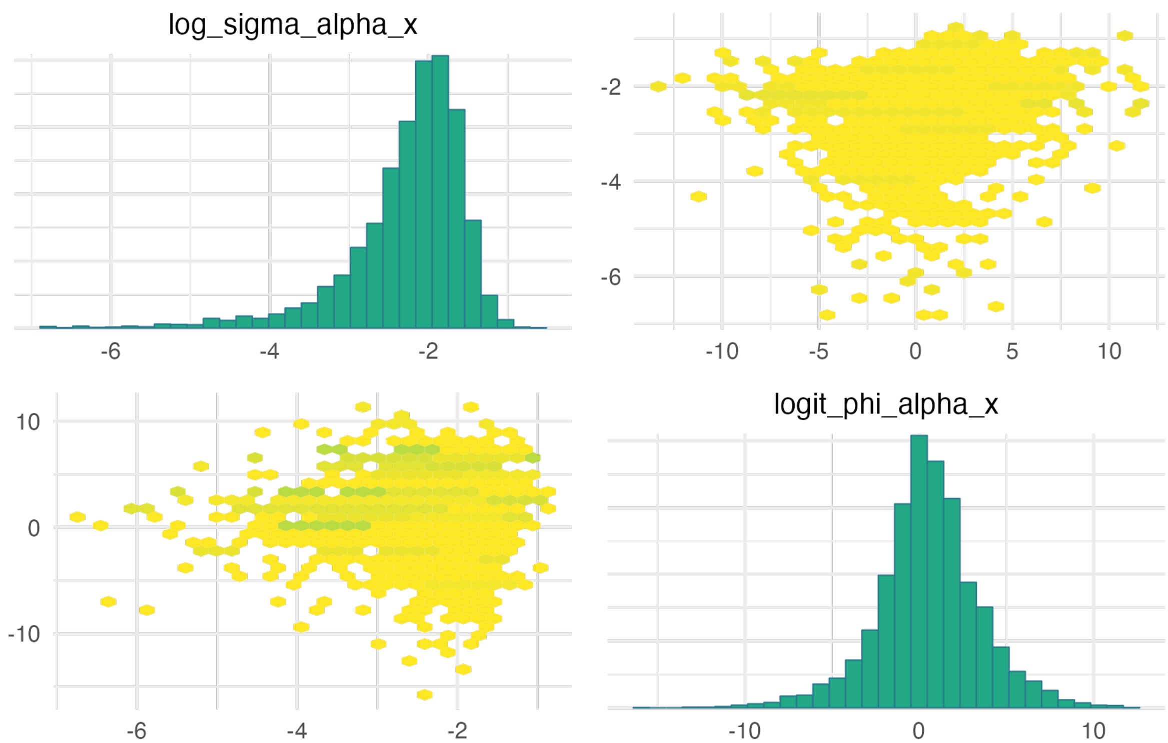

Figure S13: In contrast to the AR1 hyperparameters, the BYM2 hyperparameter posteriors are much less correlated.

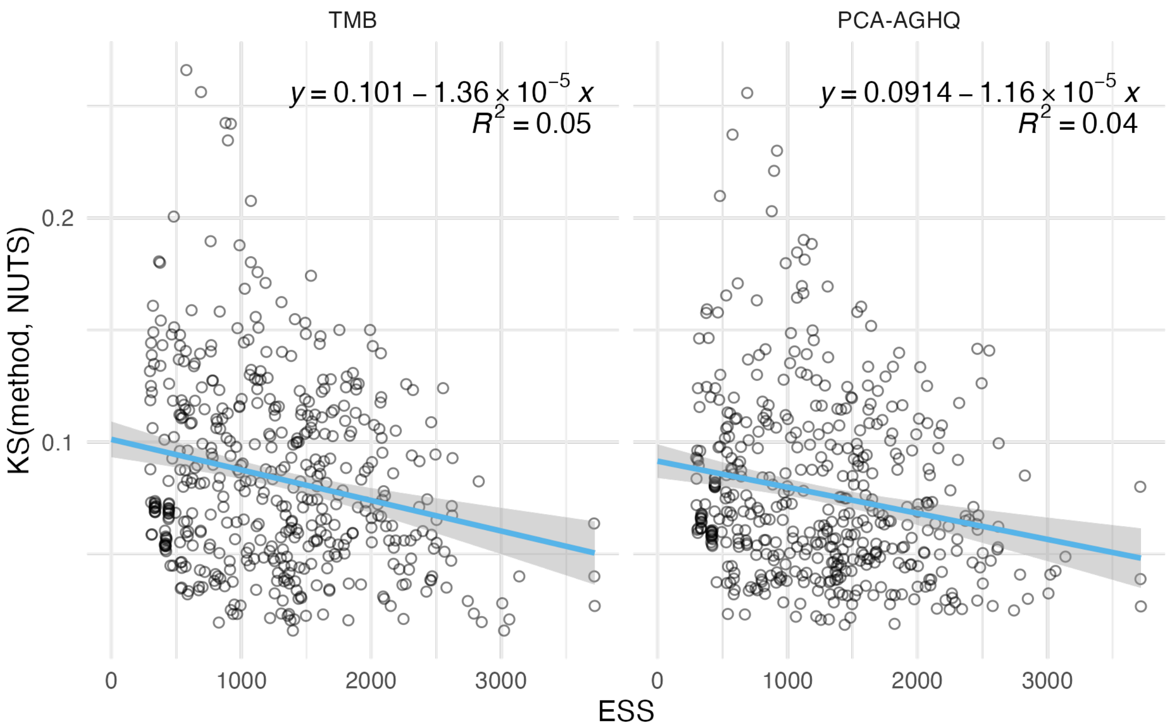

Figure S14: The ESS was negatively correlated with KS test statistic for both TMB and PCA-AGHQ. This could be explained in two ways. First, harder to sample parameters could be more difficult to estimate with approximate methods. Second, as more effective samples are collected, the NUTS posterior could become closer to the approximate posteriors.

S6 Algorithm using Laplace latent field marginals

1. Calculate the mode, Hessian at the mode, lower Cholesky, and Laplace approximation

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \tilde{p}_{\texttt{LA}}(\boldsymbol{\theta}, \mathbf{y}), \tag{34}$$

$$\hat{\mathbf{H}} = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \log \tilde{p}_{\texttt{LA}}(\boldsymbol{\theta}, \mathbf{y})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{35}$$

$$\hat{\mathbf{H}}^{-1} = \hat{\mathbf{L}}\hat{\mathbf{L}}^\top, \tag{36}$$

$$\tilde{p}_{\texttt{LA}}(\boldsymbol{\theta}, \mathbf{y}) = \frac{p(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})}{\tilde{p}_{\texttt{G}}(\mathbf{x} \,|\, \boldsymbol{\theta}, \mathbf{y})}\Big|_{\mathbf{x}=\hat{\mathbf{x}}(\boldsymbol{\theta})}, \tag{37}$$

   where $\tilde{p}_{\texttt{G}}(\mathbf{x} \,|\, \boldsymbol{\theta}, \mathbf{y}) = \mathcal{N}(\mathbf{x} \,|\, \hat{\mathbf{x}}(\boldsymbol{\theta}), \hat{\mathbf{H}}(\boldsymbol{\theta})^{-1})$ is a Gaussian approximation to $p(\mathbf{x} \,|\, \boldsymbol{\theta}, \mathbf{y})$ with mode and precision matrix given by

$$\hat{\mathbf{x}}(\boldsymbol{\theta}) = \arg\max_{\mathbf{x}} \log p(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}), \tag{38}$$

$$\hat{\mathbf{H}}(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial\mathbf{x}\partial\mathbf{x}^\top} \log p(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})|_{\mathbf{x}=\hat{\mathbf{x}}(\boldsymbol{\theta})}. \tag{39}$$

2. Generate a set of nodes $\mathbf{u} \in \mathcal{Q}(m, k)$ and weights $\omega : \mathbf{u} \to \mathbb{R}$ from a Gauss-Hermite quadrature rule with k nodes per dimension. Adapt these nodes based on the mode and lower Choleksy via $\boldsymbol{\theta}(\mathbf{u}) = \hat{\boldsymbol{\theta}} + \mathbf{L}\mathbf{u}$. Use this quadrature rule to calculate the normalising constant $\tilde{p}_{\texttt{AQ}}(\mathbf{y})$ as follows

$$\tilde{p}_{\texttt{AQ}}(\mathbf{y}) = \sum_{\mathbf{u}\in\mathcal{Q}(m,k)} \tilde{p}_{\texttt{LA}}(\boldsymbol{\theta}(\mathbf{u}), \mathbf{y})\omega(\mathbf{u}). \tag{40}$$

3. For $i \in [N]$ generate l nodes $x_i(\mathbf{v})$ via a Gauss-Hermite quadrature rule $\mathbf{v} \in \mathcal{Q}(1, l)$ adapted based on the mode $\hat{\mathbf{x}}(\boldsymbol{\theta})_i$ and standard deviation $\sqrt{\text{diag}[\hat{\mathbf{H}}(\boldsymbol{\theta})^{-1}]_i}$ of the Gaussian marginal. A value of $l \geq 4$ is recommended to enable B-spline interpolation. For $x_i \in \{x_i(\mathbf{v})\}_{\mathbf{v}\in\mathcal{Q}(1,l)}$ and $\boldsymbol{\theta} \in \{\boldsymbol{\theta}(\mathbf{u})\}_{\mathbf{u}\in\mathcal{Q}(m,k)}$ calculate the modes and Hessians

$$\hat{\mathbf{x}}_{-i}(x_i, \boldsymbol{\theta}) = \arg\max_{\mathbf{x}_{-i}} \log p(\mathbf{y}, x_i, \mathbf{x}_{-i}, \boldsymbol{\theta}), \tag{41}$$

$$\hat{\mathbf{H}}_{-i,-i}(x_i, \boldsymbol{\theta}) = -\frac{\partial^2}{\partial\mathbf{x}_{-i}\partial\mathbf{x}_{-i}^\top} \log p(\mathbf{y}, x_i, \mathbf{x}_{-i}, \boldsymbol{\theta})|_{\mathbf{x}_{-i}=\hat{\mathbf{x}}_{-i}(x_i,\boldsymbol{\theta})}, \tag{42}$$

   where optimisation to obtain $\hat{\mathbf{x}}_{-i}(x_i, \boldsymbol{\theta})$ can be initialised at $\hat{\mathbf{x}}(\boldsymbol{\theta})_{-i}$.

4. For $x_i \in \{x_i(\mathbf{v})\}_{\mathbf{v}\in\mathcal{Q}(1,l)}$ calculate

$$\tag{43}$$

   where

$$\tilde{p}_{\texttt{LA}}(x_i, \mathbf{y}) = \sum_{\mathbf{u}\in\mathcal{Q}(m,k)} \tilde{p}_{\texttt{LA}}(x_i, \boldsymbol{\theta}(\mathbf{u}), \mathbf{y})\omega(\mathbf{u}).$$

   and

$$\tilde{p}_{\texttt{LA}}(x_i, \boldsymbol{\theta}, \mathbf{y}) = \frac{p(x_i, \mathbf{x}_{-i}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{p}_{\texttt{G}}(\mathbf{x}_{-i} \,|\, x_i, \boldsymbol{\theta}, \mathbf{y})}\Big|_{\mathbf{x}_{-i}=\hat{\mathbf{x}}_{-i}(x_i,\boldsymbol{\theta})}.$$

   Although Equation **??** can be calculated using the estimate of the evidence given in Equation 40 it is more numerically accurate, and requires little extra computation, to use the estimate

$$\tilde{p}_{\texttt{AQ}}(\mathbf{y}) = \sum_{\mathbf{v}\in\mathcal{Q}(1,l)} \tilde{p}_{\texttt{LA}}(x_i(\mathbf{v}), \mathbf{y})\omega(\mathbf{v})$$

5. Given $\{x_i(\mathbf{v}), \tilde{p}_{\texttt{AQ}}(x_i(\mathbf{v}) \,|\, \mathbf{y})\}_{\mathbf{v}\in\mathcal{Q}(1,l)}$ create a spline interpolant to each posterior marginal on the log-scale. Samples, and thereby relevant posterior marginal summaries, may be obtained using inverse transform sampling.

References

Besag, Julian, Jeremy York, and Annie Mollié. 1991. "Bayesian image restoration, with two applications in spatial statistics." *Annals of the Institute of Statistical Mathematics* 43 (1): 1–20.

Eaton, Jeffrey W., Laura Dwyer-Lindgren, Steve Gutreuter, Megan O'Driscoll, Oliver Stevens, Sumali Bajaj, Rob Ashton, et al. 2021. "Naomi: A New Modelling Tool for Estimating HIV Epidemic Indicators at the District Level in Sub-Saharan Africa." *Journal of the International AIDS Society* 24 (S5): e25788. https://doi.org/https://doi.org/10.1002/jia2.25788.

Freni-Sterrantino, Anna, Massimo Ventrucci, and Håvard Rue. 2018. "A Note on Intrinsic Conditional Autoregressive Models for Disconnected Graphs." *Spatial and Spatio-Temporal Epidemiology* 26: 25–34.

Kish, Leslie. 1965. "Survey Sampling."

Kristensen, Kasper. 2021. "The Comprehensive TMB Documentation." https://kaskr.github.io/adcomp/_book/Introduction.html.

Lumley, Thomas. 2004. "Analysis of Complex Survey Samples." *Journal of Statistical Software* 9: 1–19.

Simpson, Daniel, Håvard Rue, Andrea Riebler, Thiago G Martins, and Sigrunn H Sørbye. 2017. "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors." *Statistical Science* 32 (1): 1–28.

Stover, J, P Johnson, T Hallett, M Marston, R Becquet, and IM Timaeus. 2010. "The Spectrum Projection Package: Improvements in Estimating Incidence by Age and Sex, Mother-to-Child Transmission, HIV Progression in Children and Double Orphans." *Sexually Transmitted Infections* 86 (Suppl 2): ii16–21.