

Scalable Bayesian maximum entropy for large environmental data sets through Vecchia graphical inference

Ying-Fan Lin

Chung Yuan Christian University

Hwa-Lung Yu

hlyu@ntu.edu.tw

National Taiwan University

Research Article

Keywords: Bayesian maximum entropy, Vecchia factorization, soft data, mixture propagation, sequential updating, groundwater monitoring

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11148258/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Scalable Bayesian maximum entropy for large environmental data sets through Vecchia graphical inference

Ying-Fan Lin¹ and Hwa-Lung Yu^{2*}

¹Department of Civil Engineering, Chung Yuan Christian University,
Taoyuan, Taiwan.

^{2*}Department of Bioenvironmental Systems Engineering, National
Taiwan University, Taipei, Taiwan.

*Corresponding author(s). E-mail(s): hlyu@ntu.edu.tw;
Contributing authors: yflin1110@cycu.edu.tw;

Abstract

Conventional Bayesian maximum entropy (BME) becomes computationally costly when large environmental data sets contain both exact observations and probability distributions. Prediction then requires operations on complete covariance matrices and integration over soft values. VISTA-BME combines Vecchia graphical inference with Gaussian expectation propagation and mixture propagation. Including every predecessor reproduced the matched BME posterior within 2×10^{-10} in dimensionless fields. Complete propagation of mixture components reproduced analytic Gaussian mixture reference posteriors in nine cases with two to four simultaneous bimodal likelihoods. An automatic rule selected mixture propagation in 12 of 45 cases with one soft location and six of nine cases with several soft locations, while reproducing every reference mode count. By comparison, one local Gaussian approximation gave a median Wasserstein-1 distance of 0.0341. In the Biscayne aquifer comparison, measured wall time and peak memory decreased by factors of 19.98 and 2391, respectively, relative to BME. The corresponding factors at 8,000 Gaussian observations were 2.49 and 25.1. A United States Geological Survey application reconstructed 178,776 latent nodes indexed by well, month, and grid location on one workstation. Predictions for

2025 varied little across seven covariance settings, although raw interval coverage ranged from 0.440 to 0.499. Split calibration increased coverage from 0.463 to 0.890. These results extend BME assimilation of explicit soft likelihoods to larger data sets and distinguish errors due to sparse approximation, likelihood approximation, and interval calibration from computational cost.

Keywords: Bayesian maximum entropy, Vecchia factorization, soft data, mixture propagation, sequential updating, groundwater monitoring

1 Introduction

Environmental monitoring networks often combine measurements with different descriptions of uncertainty. A screened water level can be treated as a hard datum whose value is known at the scale of the analysis. A detection limit, quality control range, indirect estimate, or ensemble distribution instead provides soft data expressed as a probability distribution over plausible values. Reducing these distributions to single values removes asymmetry, tails, and distinct modes. Bayesian maximum entropy (BME) places hard and soft data within one probabilistic model of a spatiotemporal random field (Christakos 1990, 2000; Serre and Christakos 1999). This treatment has been used to combine monitoring measurements with model results (LoBuglio et al. 2007) and, through BME filtering, to estimate groundwater parameters sequentially (Yu et al. 2020). At the base of the Biscayne aquifer, for example, measured chloride, specific conductance, and censoring limits provide different information about saltwater intrusion (Zhang and Renshaw 2024, 2025).

Several approaches avoid complete covariance calculations for large spatial data, among them predictive processes, stochastic partial differential equations, multiresolution approximations, nearest-neighbour Gaussian processes, and variational inducing variables (Banerjee et al. 2008; Lindgren et al. 2011; Katzfuss 2017; Datta et al. 2016; Titsias 2009). Vecchia factorization evaluates a location from a limited set of preceding observations (Vecchia 1988; Stein et al. 2004). Later work examined

21 ordering and grouping, developed sparse prediction with these precision matrices,
 22 and extended Vecchia regression to data sets with many observations and covariates
 23 (Guinness 2018; Katzfuss and Guinness 2021; Katzfuss et al. 2020; Cao et al. 2022).
 24 Dynamic nearest-neighbour models and streaming Gaussian process updates carry
 25 this sparse calculation into temporal and sequential settings (Datta et al. 2016; Bui
 26 et al. 2017). Expectation propagation, integrated nested Laplace approximation, and
 27 Vecchia–Laplace calculations provide complementary treatments for likelihoods out-
 28 side the Gaussian family (Minka 2001; Nickisch and Rasmussen 2008; Rue et al. 2009;
 29 Zilber and Katzfuss 2021).

30 These numerical developments have not yet been assembled around the operational
 31 BME posterior and explicit probability distributions for soft data. Restricting spatial
 32 dependence changes the representation of the general knowledge model, while approxi-
 33 mating each soft distribution changes the specific knowledge supplied at an observation
 34 location. A single Gaussian can match the first two moments of a skewed distribu-
 35 tion yet suppress modes that remain relevant after nearby measurements enter the
 36 calculation. The implementation must also accommodate measurements that arrive
 37 sequentially. Ensemble Kalman filters address sequential state and parameter esti-
 38 mation, including groundwater applications (Evensen 2003; Hendricks Franssen and
 39 Kinzelbach 2008). Tree models and graph neural networks estimate nonlinear spatial
 40 responses from covariates and monitoring networks (Chen and Guestrin 2016; Bai and
 41 Tahmasebi 2023; Rajaei et al. 2019; Wunsch et al. 2021); their standard regression
 42 inputs are point values, whereas a BME soft datum supplies a probability distribution
 43 at a location. Large spatial comparisons also show that point error, interval cover-
 44 age, time, and memory describe separate aspects of performance (Heaton et al. 2019).
 45 A scalable BME calculation requires separate tests of posterior agreement, the error
 46 introduced by each approximation, predictive calibration, computational time, and
 47 memory.

48 VISTA-BME denotes Vecchia Inference with Soft Data Treatment and Assimila-
 49 tion for Bayesian Maximum Entropy. It retains the operational BME update and its
 50 distinction between hard and soft data. Ordered conditional factorization limits the
 51 preceding locations used in each spatial calculation. Gaussian expectation propagation
 52 treats bounded, censored, skewed, and heavy-tailed soft distributions. When distinct
 53 modes affect posterior probabilities or threshold estimates, mixture expectation prop-
 54 agation represents the soft distribution as a Gaussian mixture and carries selected
 55 component combinations through the posterior calculation. An automatic rule uses
 56 the mode count and the GEP variance residual to make this choice at each soft loca-
 57 tion, then increases the component count until the density and posterior changes meet
 58 specified tolerances. Repeated observations at fixed monitoring locations update accu-
 59 mulated parameters for soft data without rebuilding the graph. Synthetic experiments
 60 compare the posterior with traditional BME and direct integration. The Biscayne and
 61 United States Geological Survey applications examine the contribution of soft data,
 62 predictive calibration, wall time, and memory as the spatial and temporal data sets
 63 increase in size.

64 2 Methodology

65 2.1 Operational Bayesian updating and the general knowledge 66 base

67 Conventional BME distinguishes the general knowledge base from the specific knowl-
 68 edge base (Christakos 1990, 2000; Serre and Christakos 1999). The attribute values
 69 to be estimated are collected in $\mathbf{x}_k$ at space and time locations $\mathbf{p}_k$. Observations with
 70 values treated as certain are the hard data $\mathbf{x}_h$ at locations $\mathbf{p}_h$. Observations whose
 71 values are described by probability distributions are the soft data $\mathbf{x}_s$ at locations $\mathbf{p}_s$.
 72 The joint density $f_G(\mathbf{x}_k, \mathbf{x}_h, \mathbf{x}_s)$ expresses the general knowledge base. The specific

73 knowledge base contains the soft information functions $f_S(x_{s,j})$. The implemented
 74 specific knowledge model assumes a product of factors across soft locations,

$$f_S(\mathbf{x}_s) = \prod_{j=1}^{n_s} f_S(x_{s,j}). \quad (1)$$

75 where n_s is the number of soft locations. The operational BME posterior for the values
 76 at the prediction locations is

$$f_K(\mathbf{x}_k | \mathbf{x}_h, f_S) = \frac{\int_{\mathbb{R}^{n_s}} f_G(\mathbf{x}_k, \mathbf{x}_h, \mathbf{x}_s) f_S(\mathbf{x}_s) d\mathbf{x}_s}{\int_{\mathbb{R}^{n_s}} f_G(\mathbf{x}_h, \mathbf{x}_s) f_S(\mathbf{x}_s) d\mathbf{x}_s}. \quad (2)$$

77 where f_K is the posterior knowledge density and $f_G(\mathbf{x}_h, \mathbf{x}_s)$ is the marginal general
 78 knowledge density after integrating over $\mathbf{x}_k$. A normalized probability density at one
 79 location is the conventional BME form of $f_S(x_{s,j})$. The implementation also accepts
 80 an observation likelihood proportional to $p(y_j | x_{s,j})$, because a constant multiplier
 81 cancels between the numerator and denominator of Eq. (2). This likelihood form
 82 admits interval and censored observations whose information is normalized after mul-
 83 tiplication by the general knowledge density. VISTA-BME retains Eq. (2); its sparse
 84 operators change how the specified general and specific knowledge are evaluated.

85 For the calculations reported here, the complete random vector is ordered as $\mathbf{x}_{hsk} =$
 86 $[\mathbf{x}_h^\top, \mathbf{x}_s^\top, \mathbf{x}_k^\top]^\top$ at the corresponding locations $\mathbf{p}_{hsk} = [\mathbf{p}_h^\top, \mathbf{p}_s^\top, \mathbf{p}_k^\top]^\top$. Its mean is

$$\mathbf{m}_{hsk} = E(\mathbf{x}_{hsk}) = \mathbf{m}_\phi(\mathbf{p}_{hsk}, \mathbf{U}), \quad (3)$$

87 where $\mathbf{U}$ contains the available covariates and ϕ contains the mean function param-
 88 eters. A covariance function with parameters θ determines the covariance matrix Σ .

89 The general knowledge model used in every comparison is

$$f_G(\mathbf{x}_{hsk}) = f_N(\mathbf{x}_{hsk} \mid \mathbf{m}_{hsk}, \mathbf{\Sigma}), \quad (4)$$

90 where $f_N(\cdot \mid \mathbf{m}, \mathbf{\Sigma})$ denotes a multivariate Gaussian density with mean $\mathbf{m}$ and covari-
 91 ance $\mathbf{\Sigma}$. Equation (2) permits other general knowledge models; Eq. (4) states the model
 92 evaluated here. Throughout this article, BME denotes the traditional calculation using
 93 this same f_G , the complete covariance matrix, and the same soft information. This
 94 matched definition isolates the effect of the numerical formulation. The derivation
 95 retains a nonzero deterministic mean. The implementation subtracts this mean before
 96 constructing the Vecchia graph, applies the sparse equations to the residual field, and
 97 adds the mean to the posterior prediction. These operations are algebraically equiva-
 98 lent for a fixed $\mathbf{m}_\phi$; the current implementation does not estimate ϕ jointly within the
 99 expectation propagation iterations. With Gaussian soft data, Eq. (2) reduces to condi-
 100 tional Gaussian equations and is equivalent to Gaussian process regression under the
 101 same mean, covariance, and observation error model (Rasmussen and Williams 2006).

102 2.2 From the joint chain rule to sparse conditional factorization

103 Place the $n = n_h + n_s + n_k$ variables in $\mathbf{x}_{hsk}$ in a fixed order. The chain rule conditions
 104 each variable on every value that precedes it,

$$f_G(\mathbf{x}_{hsk}) = \prod_{i=1}^n f_G(x_i \mid \mathbf{x}_{1:(i-1)}), \quad (5)$$

105 where $\mathbf{x}_{1:(i-1)}$ contains all variables before x_i . Vecchia factorization replaces the com-
 106 plete conditioning vector with a set of preceding variables (Vecchia 1988; Stein et al.

2004; Guinness 2018; Katzfuss and Guinness 2021),

$$f_G(\mathbf{x}_{hsk}) \simeq \widetilde{f}_G(\mathbf{x}_{hsk}) = \prod_{i=1}^n f_G(x_i \mid \mathbf{x}_{g(i)}), \quad (6a)$$

$$g(i) \subseteq \{1, \dots, i-1\}, \quad (6b)$$

$$|g(i)| \leq m, \quad (6c)$$

where $g(i)$ is the predecessor set for location i and m is its maximum size. Setting $g(i) = \{1, \dots, i-1\}$ recovers the complete chain rule. A finite m gives a sparse representation of the general knowledge model while leaving the operational BME relation in Eq. (2) unchanged.

For the Gaussian general knowledge model, each conditional factor in Eq. (6) is

$$f_G(x_i \mid \mathbf{x}_{g(i)}) = f_N(x_i \mid m_{i|g(i)}, \Sigma_{i|g(i)}), \quad (7)$$

where the conditional mean $m_{i|g(i)}$ predicts x_i from the predecessor values and the innovation variance $\Sigma_{i|g(i)}$ is the variance that remains after this prediction. They are

$$m_{i|g(i)} = m_i + \Sigma_{i,g(i)} \Sigma_{g(i),g(i)}^{-1} (\mathbf{x}_{g(i)} - \mathbf{m}_{g(i)}), \quad (8a)$$

$$\Sigma_{i|g(i)} = \Sigma_{i,i} - \Sigma_{i,g(i)} \Sigma_{g(i),g(i)}^{-1} \Sigma_{g(i),i}. \quad (8b)$$

where $\Sigma_{i,g(i)} \Sigma_{g(i),g(i)}^{-1}$ contains the conditional weights. The diagonal matrix $\mathbf{D} = \text{diag}(\Sigma_{1|g(1)}, \dots, \Sigma_{n|g(n)})$ collects the innovation variances. The diagonal of Σ contains the original prior marginal variances. Marginal variances implied by a truncated Vecchia graph may differ from them.

Collect the conditional innovations in $\boldsymbol{\epsilon} = \mathbf{L}(\mathbf{x}_{hsk} - \mathbf{m}_{hsk})$. Row i of the unit lower triangular matrix $\mathbf{L}$ contains $-\Sigma_{i,g(i)} \Sigma_{g(i),g(i)}^{-1}$ in columns $g(i)$, and ϵ_i has variance

124 $\Sigma_{i|g(i)}$. Summation of the Gaussian quadratic terms gives

$$\widetilde{f}_G(\mathbf{x}_{hsk}) \propto \exp\left[-\frac{1}{2}(\mathbf{x}_{hsk} - \mathbf{m}_{hsk})^\top \widetilde{\Lambda}(\mathbf{x}_{hsk} - \mathbf{m}_{hsk})\right], \quad (9a)$$

125

$$\widetilde{\Lambda} = \mathbf{L}^\top \mathbf{D}^{-1} \mathbf{L}. \quad (9b)$$

126 where $\widetilde{\Lambda}$ is the Vecchia precision matrix. Each row of $\mathbf{L}$ contains at most $m + 1$
 127 nonzero entries. Including every predecessor gives $\widetilde{\Lambda} = \Sigma^{-1}$ to numerical roundoff.
 128 With a finite predecessor set, the precision used in the calculation is $\widetilde{\Lambda}$. For brevity, Λ
 129 denotes this computational precision below; it equals Σ^{-1} only when all predecessors
 130 are included.

131 2.3 Conditioning on hard data and Gaussian soft data

132 Partition the complete vector into the hard data $\mathbf{x}_h$ and the uncertain variables
 133 $\mathbf{x}_{ks} = [\mathbf{x}_s^\top, \mathbf{x}_k^\top]^\top$. Applying the same partition to the precision matrix in Eq. (9b) gives
 134 the conditional Gaussian model in information form (Rasmussen and Williams 2006;
 135 Minka 2001),

$$\Lambda = \begin{bmatrix} \Lambda_{h,h} & \Lambda_{h,ks} \\ \Lambda_{ks,h} & \Lambda_{ks,ks} \end{bmatrix}, \quad (10a)$$

136

$$\Lambda_{ks|h} = \Lambda_{ks,ks}, \quad (10b)$$

137

$$\boldsymbol{\eta}_{ks|h} = \Lambda_{ks,ks} \mathbf{m}_{ks} - \Lambda_{ks,h}(\mathbf{x}_h - \mathbf{m}_h), \quad (10c)$$

138

$$f_G(\mathbf{x}_{ks} | \mathbf{x}_h) \propto \exp\left[-\frac{1}{2} \mathbf{x}_{ks}^\top \Lambda_{ks|h} \mathbf{x}_{ks} + \boldsymbol{\eta}_{ks|h}^\top \mathbf{x}_{ks}\right]. \quad (10d)$$

139 where $\Lambda_{ks|h}$ is the precision of $\mathbf{x}_{ks}$ conditional on the hard observations and $\boldsymbol{\eta}_{ks|h}$ is its
 140 linear information parameter. The corresponding conditional mean $\mathbf{m}_{ks|h}$ is obtained
 141 from the sparse linear system $\Lambda_{ks|h} \mathbf{m}_{ks|h} = \boldsymbol{\eta}_{ks|h}$. The binary selection matrices $\mathbf{I}_{s,ks}$
 142 and $\mathbf{I}_{k,ks}$ extract the soft and prediction variables from $\mathbf{x}_{ks}$. The BME posterior before

143 marginalization over the soft variables is

$$f_K(\mathbf{x}_{ks} | \mathbf{x}_h, f_S) \propto f_G(\mathbf{x}_{ks} | \mathbf{x}_h) \prod_{j=1}^{n_s} f_S(x_{s,j}). \quad (11)$$

144 where its marginal over $\mathbf{x}_s$ gives the prediction density in Eq. (2).

145 For a Gaussian soft datum, the mean $\tilde{m}_{s,j}$ and variance $\tilde{\sigma}_{s,j}^2$ define

$$f_S(x_{s,j}) = f_N(x_{s,j} | \tilde{m}_{s,j}, \tilde{\sigma}_{s,j}^2). \quad (12)$$

146 where $\tilde{\sigma}_{s,j}^{-2}$ is its precision and $\tilde{m}_{s,j}\tilde{\sigma}_{s,j}^{-2}$ is its precision weighted mean. Collecting the
 147 means in $\tilde{\mathbf{m}}_s$ and the variances in the diagonal matrix $\tilde{\Sigma}_s$ gives

$$\Lambda_{ks|hs}^{\text{BME}} = \Lambda_{ks|h} + \mathbf{I}_{s,ks}^T \tilde{\Sigma}_s^{-1} \mathbf{I}_{s,ks}, \quad (13a)$$

$$\Lambda_{ks|hs}^{\text{BME}} \mathbf{m}_{ks|hs}^{\text{BME}} = \eta_{ks|h} + \mathbf{I}_{s,ks}^T \tilde{\Sigma}_s^{-1} \tilde{\mathbf{m}}_s, \quad (13b)$$

$$\mathbf{m}_{ks|hs}^{\text{BME}} = (\Lambda_{ks|hs}^{\text{BME}})^{-1} \left(\eta_{ks|h} + \mathbf{I}_{s,ks}^T \tilde{\Sigma}_s^{-1} \tilde{\mathbf{m}}_s \right). \quad (13c)$$

150 where $\Lambda_{ks|hs}^{\text{BME}}$, $\mathbf{m}_{ks|hs}^{\text{BME}}$, and $\Sigma_{ks|hs}^{\text{BME}} = (\Lambda_{ks|hs}^{\text{BME}})^{-1}$ are the BME posterior precision, mean,
 151 and covariance of $\mathbf{x}_{ks}$ after using all hard and soft data. The prediction mean and
 152 covariance are $\mathbf{I}_{k,ks} \mathbf{m}_{ks|hs}^{\text{BME}}$ and $\mathbf{I}_{k,ks} \Sigma_{ks|hs}^{\text{BME}} \mathbf{I}_{k,ks}^T$. Adding the diagonal soft precision
 153 preserves the sparsity pattern of the Vecchia graph.

154 2.4 Non-Gaussian soft data

155 Interval, censored, skewed, heavy tailed, multimodal, and tabulated soft distributions
 156 generally leave the integral over $\mathbf{x}_s$ in Eq. (2) without a closed form. Traditional BME
 157 evaluates this integral numerically when it combines general knowledge with site-
 158 specific uncertain information (Christakos 2000; Kolovos et al. 2002). VISTA-BME

159 uses either one Gaussian approximation or a Gaussian mixture, allowing the precision
 160 calculation in Eq. (13) to be reused.

161 **2.4.1 Gaussian expectation propagation**

162 Gaussian expectation propagation (GEP) represents the contribution of the soft func-
 163 tion at location j by an unnormalized Gaussian form. The natural parameters used
 164 below add directly to the posterior precision and linear information vector in Eq. (13).
 165 For positive site precision, they give the same Gaussian mean and variance used for a
 166 Gaussian soft datum in Eq. (12).

$$\widetilde{f}_{S_j}^{(t)}(x_{s,j}) \propto \exp\left[-\frac{1}{2}\tau_j^{(t)}x_{s,j}^2 + \nu_j^{(t)}x_{s,j}\right], \quad (14)$$

167 where j indexes the soft location and t indexes the GEP iteration. The scalar $\tau_j^{(t)}$
 168 is the site precision, whereas $\nu_j^{(t)}$ is the linear natural parameter, equal to precision
 169 multiplied by the site mean. When $\tau_j^{(t)} > 0$, these parameters correspond to the
 170 Gaussian quantities introduced in Eq. (12) through

$$\widetilde{\sigma}_{s,j}^{2(t)} = \left(\tau_j^{(t)}\right)^{-1}, \quad (15a)$$

$$\widetilde{m}_{s,j}^{(t)} = \nu_j^{(t)} / \tau_j^{(t)}. \quad (15b)$$

172 The initialization $\tau_j^{(0)} = \nu_j^{(0)} = 0$ makes the factor in Eq. (14) equal to one, so the
 173 site contributes no soft information. Equations (15a) and (15b) apply only to positive
 174 precision; no normalized site variance is assigned at initialization. The complete pos-
 175 terior remains proper because its precision includes the general knowledge model and
 176 any other observations.

177 Collecting the site parameters in $\boldsymbol{\tau}^{(t)}$ and $\boldsymbol{\nu}^{(t)}$, the current Gaussian posterior is
 178 calculated as

$$\boldsymbol{\Lambda}_{ks|hs}^{\text{BME},(t)} = \boldsymbol{\Lambda}_{ks|h} + \boldsymbol{I}_{s,ks}^{\top} \text{diag}(\boldsymbol{\tau}^{(t)}) \boldsymbol{I}_{s,ks}, \quad (16a)$$

$$\boldsymbol{\Lambda}_{ks|hs}^{\text{BME},(t)} \boldsymbol{m}_{ks|hs}^{\text{BME},(t)} = \boldsymbol{\eta}_{ks|h} + \boldsymbol{I}_{s,ks}^{\top} \boldsymbol{\nu}^{(t)}. \quad (16b)$$

179 where $\boldsymbol{\Lambda}_{ks|hs}^{\text{BME},(t)}$, $\boldsymbol{m}_{ks|hs}^{\text{BME},(t)}$, and $\boldsymbol{\Sigma}_{ks|hs}^{\text{BME},(t)} = (\boldsymbol{\Lambda}_{ks|hs}^{\text{BME},(t)})^{-1}$ are the current BME poste-
 180 rior precision, mean, and covariance. They are the iteration-specific forms of $\boldsymbol{\Lambda}_{ks|hs}^{\text{BME}}$,
 181 $\boldsymbol{m}_{ks|hs}^{\text{BME}}$, and $\boldsymbol{\Sigma}_{ks|hs}^{\text{BME}}$ in Eq. (13). The marginal at soft location j is $q_j^{(t)} = f_N(x_{s,j} |$
 182 $\mu_{s,j}^{(t)}, v_{s,j}^{(t)})$, where $\mu_{s,j}^{(t)}$ and $v_{s,j}^{(t)}$ are the corresponding mean and variance entries. The
 183 marginal variances are obtained from selected columns of the inverse precision matrix
 184 using the same sparse factorization; the implementation evaluates these columns in
 185 batches of at most 64. Removing the site factor at location j gives the cavity density,
 186

$$q_j^{\setminus j,(t)}(x_{s,j}) = f_N(x_{s,j} | a_j^{(t)}, b_j^{(t)}), \quad (17)$$

187 where $a_j^{(t)}$ and $b_j^{(t)}$ are the cavity mean and variance. Their natural parameters are

$$\left(b_j^{(t)}\right)^{-1} = \left(v_{s,j}^{(t)}\right)^{-1} - \tau_j^{(t)}, \quad (18a)$$

$$a_j^{(t)} \left(b_j^{(t)}\right)^{-1} = \mu_{s,j}^{(t)} \left(v_{s,j}^{(t)}\right)^{-1} - \nu_j^{(t)}. \quad (18b)$$

188 The calculation stops if the leave one out precision is at most 10^{-12} . Multiplying this
 189 density by the original soft information gives
 190

$$f_j^{\star(t)}(x_{s,j}) = \frac{f_S(x_{s,j}) q_j^{\setminus j,(t)}(x_{s,j})}{\int_{-\infty}^{\infty} f_S(u) q_j^{\setminus j,(t)}(u) du}. \quad (19)$$

191 where u is the integration variable. The tilted mean $\bar{x}_{s,j}^{\star(t)}$ and variance $V_{s,j}^{\star(t)}$ of $f_j^{\star(t)}$
 192 determine the raw replacement parameters,

$$\tau_j^{\text{raw}} = \left(V_{s,j}^{\star(t)}\right)^{-1} - \left(b_j^{(t)}\right)^{-1}, \quad (20a)$$

$$\nu_j^{\text{raw}} = \bar{x}_{s,j}^{\star(t)} \left(V_{s,j}^{\star(t)}\right)^{-1} - a_j^{(t)} \left(b_j^{(t)}\right)^{-1}. \quad (20b)$$

194 where the proposed precision is $\tau_j^{\text{prop}} = \max(\tau_j^{\text{raw}}, 0)$. A numerical upper bound of
 195 $10^8/\sigma_{\text{soft},j}^2$, where $\sigma_{\text{soft},j}^2$ is a variance scale calculated from the original soft distribu-
 196 tion, prevents nearly degenerate updates. When this bound is active, the proposed
 197 linear parameter is reset to τ_j^{prop} multiplied by the tilted mean; for interval or censored
 198 information, that mean is first limited to the stated bounds. Otherwise, $\nu_j^{\text{prop}} = \nu_j^{\text{raw}}$.
 199 With damping coefficient α , the applied update is

$$\tau_j^{(t+1)} = (1 - \alpha)\tau_j^{(t)} + \alpha\tau_j^{\text{prop}}, \quad (21a)$$

$$\nu_j^{(t+1)} = (1 - \alpha)\nu_j^{(t)} + \alpha\nu_j^{\text{prop}}. \quad (21b)$$

201 where $0 < \alpha \leq 1$ (Minka 2001; Nickisch and Rasmussen 2008). All locations are
 202 updated from the same joint posterior before the replacements are applied together.
 203 This parallel form of expectation propagation is described by Cseke and Heskes (2011)
 204 and is the algorithm implemented in the software. A nonpositive raw precision is set to
 205 zero, and activation of the upper precision bound is reported. Either bound prevents
 206 exact moment matching at that location.

207 Interval and censored moments follow from truncated normal expressions. Moments
 208 for skew normal, lognormal, and Student- t functions are calculated by Gauss–Legendre
 209 integration with 512 nodes (Golub and Welsch 1969); repeating the calculation with
 210 256 nodes checks numerical integration. Refined interpolation and Simpson integra-
 211 tion are used for tabulated functions. Tables 3 and 4 give the damping coefficient,

212 stopping tolerance, and iteration limit. The numerical diagnostics include convergence,
 213 the largest change in either natural parameter, residuals in the matched moments,
 214 integration error, and active precision bounds.

215 2.4.2 Mixture expectation propagation

216 Mixture expectation propagation (MEP) represents a soft distribution as a Gaussian
 217 mixture and carries combinations of its components through Eq. (13). A single Gaus-
 218 sian preserves two posterior moments but cannot retain distinct modes. Gaussian
 219 sums provide an established representation of general probability densities (Alspach
 220 and Sorenson 1972). VISTA-BME partitions a normalized soft density into intervals
 221 of equal probability, then matches the probability, mean, and variance within each
 222 interval. A supplied Gaussian mixture can enter the calculation directly.

223 The index $j = 1, \dots, n_s$ identifies the soft location and C_j is the number of
 224 Gaussian components used at that location. The conversion uses the finite interval
 225 $\mathcal{I}_j = [a_{j,0}, a_{j,C_j}]$. For a normalized input density, the probability within this interval
 226 and the omitted tail probability are

$$Z_j^{\text{tr}} = \int_{a_{j,0}}^{a_{j,C_j}} f_{S,j}(u) \, du, \quad (22a)$$

$$\epsilon_j^{\text{tail}} = 1 - Z_j^{\text{tr}}. \quad (22b)$$

228 The density used to construct the components is normalized over this interval,

$$f_{S,j}^{\text{tr}}(x) = \frac{f_{S,j}(x) \mathbb{I}\{a_{j,0} \leq x \leq a_{j,C_j}\}}{Z_j^{\text{tr}}}, \quad (23)$$

229 where $\mathbb{I}$ is the indicator function. With equal probability interval edges $a_{j,0}, \dots, a_{j,C_j}$,
 230 the component quantities are

$$w_{j,c} = \int_{a_{j,c-1}}^{a_{j,c}} f_{S,j}^{\text{tr}}(x) \, dx, \quad (24a)$$

$$\tilde{m}_{j,c} = w_{j,c}^{-1} \int_{a_{j,c-1}}^{a_{j,c}} x f_{S,j}^{\text{tr}}(x) \, dx, \quad (24b)$$

$$\tilde{\sigma}_{j,c}^2 = w_{j,c}^{-1} \int_{a_{j,c-1}}^{a_{j,c}} (x - \tilde{m}_{j,c})^2 f_{S,j}^{\text{tr}}(x) \, dx, \quad (24c)$$

233 where $w_{j,c}$, $\tilde{m}_{j,c}$, and $\tilde{\sigma}_{j,c}^2$ are the weight, mean, and variance of component c . Their
 234 sum approximates the soft density,

$$f_{S,j}^{\text{tr}}(x) \simeq \sum_{c=1}^{C_j} w_{j,c} f_N(x \mid \tilde{m}_{j,c}, \tilde{\sigma}_{j,c}^2), \quad (25a)$$

$$\sum_{c=1}^{C_j} w_{j,c} = 1, \quad (25b)$$

236 where normalization over the finite interval gives weights that sum to one. The approx-
 237 imation is checked through its component count, interval, omitted tail probability,
 238 and Jensen–Shannon divergence from the normalized density on that interval (Lin
 239 1991). The conversion accepts finite intervals, skew normal, lognormal, Student- t , and
 240 tabulated densities. GEP is used for censoring at one bound because its normaliza-
 241 tion depends on the field prior. The comparison with one complex distribution fixed
 242 $C_j = 10$ so that GEP and MEP used the same component budget. The automatic
 243 calculation in Section 2.4.3 selects C_j from density and target posterior tolerances.

244 The full Cartesian product of component choices is $\mathcal{A}$. The vector $\mathbf{c} =$
 245 $(c_1, \dots, c_{n_s}) \in \mathcal{A}$ selects one component at each soft location represented by a mix-
 246 ture. In w_{j,c_j} , the first subscript j identifies the soft location and c_j is the component
 247 selected by vector $\mathbf{c}$ at that same location. The prior weight of the selected combination

248 is

$$\pi_{\mathbf{c}} = \prod_{j=1}^{n_s} w_{j,c_j}. \quad (26)$$

249 The normalization contributed by the general knowledge model is

$$Z_{\mathbf{c}} = \int_{\mathbb{R}^{n_s}} f_G(\mathbf{x}_s | \mathbf{x}_h) \prod_{j=1}^{n_s} f_N(x_{s,j} | \tilde{m}_{j,c_j}, \tilde{\sigma}_{j,c_j}^2) d\mathbf{x}_s, \quad (27)$$

250 where $f_G(\mathbf{x}_s | \mathbf{x}_h)$ is the marginal general knowledge density of the soft variables after
 251 conditioning on the hard data. The normalized posterior weight among the propagated
 252 combinations $\mathcal{C} \subseteq \mathcal{A}$ is

$$w_{\mathbf{c}}^{\text{BME}} = \frac{\pi_{\mathbf{c}} Z_{\mathbf{c}}}{\sum_{\mathbf{d} \in \mathcal{C}} \pi_{\mathbf{d}} Z_{\mathbf{d}}}. \quad (28)$$

253 Every component combination gives a Gaussian posterior from Eq. (13). Its prediction
 254 mean and covariance are

$$\mathbf{m}_k^{\text{BME},(\mathbf{c})} = \mathbf{I}_{k,ks} \mathbf{m}_{ks}^{\text{BME},(\mathbf{c})}, \quad (29a)$$

255

$$\Sigma_k^{\text{BME},(\mathbf{c})} = \mathbf{I}_{k,ks} \Sigma_{ks}^{\text{BME},(\mathbf{c})} \mathbf{I}_{k,ks}^{\top}. \quad (29b)$$

256 The posterior at the prediction locations is

$$f_K(\mathbf{x}_k | \mathbf{x}_h, f_S) \simeq \sum_{\mathbf{c} \in \mathcal{C}} w_{\mathbf{c}}^{\text{BME}} f_N(\mathbf{x}_k | \mathbf{m}_k^{\text{BME},(\mathbf{c})}, \Sigma_k^{\text{BME},(\mathbf{c})}), \quad (30)$$

257 The corresponding mixture mean and covariance include variation between component
 258 means,

$$\mathbf{m}_k^{\text{BME}} = \sum_{\mathbf{c} \in \mathcal{C}} w_{\mathbf{c}}^{\text{BME}} \mathbf{m}_k^{\text{BME},(\mathbf{c})}, \quad (31a)$$

259

$$\Sigma_k^{\text{BME}} = \sum_{\mathbf{c} \in \mathcal{C}} w_{\mathbf{c}}^{\text{BME}} \left[\Sigma_k^{\text{BME},(\mathbf{c})} + \delta_{\mathbf{c}} \delta_{\mathbf{c}}^{\top} \right], \quad (31b)$$

260 where $\delta_{\mathbf{c}} = \mathbf{m}_k^{\text{BME},(c)} - \mathbf{m}_k^{\text{BME}}$. Combinations are ordered by prior weight. The limit
 261 C_{max} controls computational cost and equals 128 unless stated otherwise. When $\mathcal{C} =$
 262 $\mathcal{A}$, Eqs. (28)–(30) reproduce the complete posterior of the represented mixture. A
 263 finite $\mathcal{C}$ renormalizes the included combinations and introduces an approximation. The
 264 simultaneous bimodal experiment used limits of 2, 4, and 8, as well as the complete
 265 set of 4 to 16 combinations. Included probability mass, omitted probability mass, and
 266 the effective number of posterior components quantify this selection. Explicit GEP
 267 and MEP calls remain available when the representation is prescribed. The automatic
 268 calculation described next applies the same diagnostics to select between them.

269 **2.4.3 Automatic choice between the two representations**

270 The automatic calculation begins with the converged GEP posterior. Let t_* denote
 271 its final iteration and let M_j denote the number of resolved modes in the original
 272 soft distribution at location j . The relative variance residual compares the final GEP
 273 marginal variance $v_{s,j}^{(t_*)}$ with the tilted variance $V_{s,j}^{*(t_*)}$ from the same site and iteration,

$$r_{V,j} = \frac{|v_{s,j}^{(t_*)} - V_{s,j}^{*(t_*)}|}{V_{s,j}^{*(t_*)}}. \quad (32)$$

274 The local representation indicator is

$$s_j = \begin{cases} 1, & M_j > 1 \text{ and } r_{V,j} > 10^{-2}, \\ 0, & \text{otherwise,} \end{cases} \quad (33)$$

275 where $s_j = 1$ assigns location j to MEP and $s_j = 0$ retains its converged GEP site.
 276 Gaussian mixtures and tabulated densities are evaluated on a deterministic grid. A
 277 local maximum must have a prominence of at least 0.05 times the maximum density
 278 to count as a resolved mode. Gaussian, interval, censored, skew normal, lognormal,

Table 1 Implemented choice of representation for soft information.

Representation	Automatic condition	Calculation and reported diagnostic
GEP	One resolved mode, or $r_{V,j} \leq 10^{-2}$	Retain the converged Gaussian site; report GEP convergence and moment residuals.
MEP	More than one resolved mode and $r_{V,j} > 10^{-2}$	Propagate Gaussian components only at switched locations; report selected counts, posterior convergence across counts, and omitted combination mass.
Largest tested mixture	No tested component count satisfies all four tolerances	Return the state with the largest tested component count and flag component selection as unresolved.

and Student- t forms enter this step as distributions without multiple modes. This rule applies MEP only where multimodality remains relevant to the GEP variance.

A switched density is converted and tested with 2, 4, 6, 8, 10, 12, 16, 20, 24, and 32 components. Supplied Gaussian mixtures keep their original components. For converted densities, successive hybrid posteriors are compared. The selected component state is the first one for which the maximum Jensen–Shannon divergence is at most 0.03, the change in posterior mean divided by posterior standard deviation is at most 0.01, the relative posterior variance change is at most 0.02, and the change in dominant mode divided by posterior standard deviation is at most 0.05. Dominant modes are evaluated at all targets when there are no more than 32 and at 32 evenly spaced target indices otherwise. The selection is specific to the compiled target set because three of the four criteria concern its posterior. If the state with 32 components does not satisfy every criterion, the calculation returns this state and marks component selection as unresolved. The maximum number of propagated component combinations remains $C_{\max} = 128$.

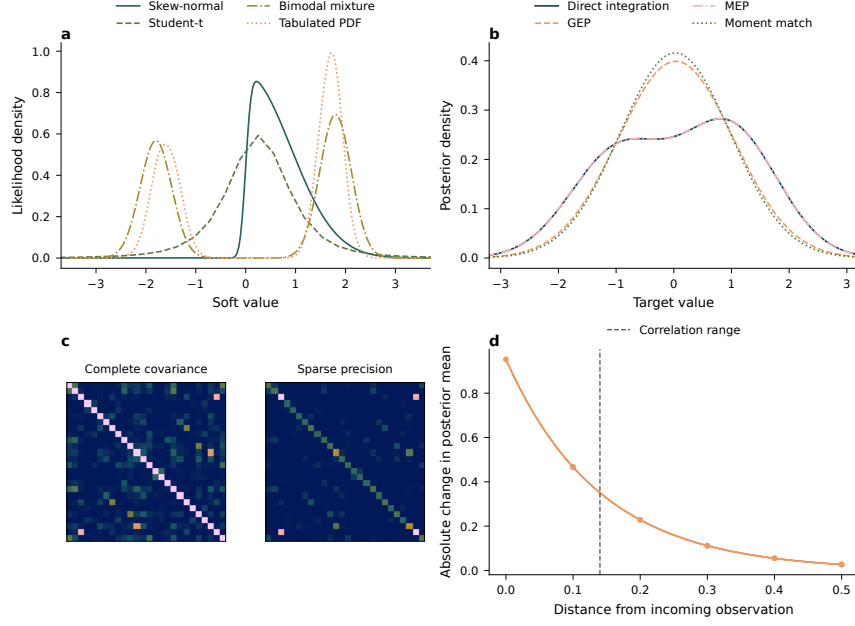


Fig. 1 Complex likelihoods and sparse updating: (a) normalized likelihood families used in the density comparison; (b) target posterior for the strongest bimodal case at normalized distance 0.08, with direct integration giving the comparison curve; (c) complete prior covariance and normalized absolute Vecchia precision with four conditioning neighbours for the same field of 26 locations; and (d) absolute change in the posterior mean after one noisy observation in an exponential field of 101 locations, with the dashed line marking the specified correlation range.

2.5 Sequential assimilation at fixed monitoring locations

Updates at fixed nodes reuse the Gaussian parameters for soft data in Eq. (12). Let the l th observation assigned to monitoring location i have mean $\tilde{m}_{s,i}^{(l)}$ and variance $\tilde{\sigma}_{s,i}^{2(l)}$. The index set $\mathcal{K}_i$ contains the observations assigned to that location, and $N_i = |\mathcal{K}_i|$ is their count. The product of the independent Gaussian factors is proportional to a Gaussian whose parameters accumulate as

$$\tilde{\sigma}_{s,i}^{-2(N_i)} = \sum_{l \in \mathcal{K}_i} \tilde{\sigma}_{s,i}^{-2(l)}, \quad (34a)$$

$$\tilde{m}_{s,i}^{(N_i)} \tilde{\sigma}_{s,i}^{-2(N_i)} = \sum_{l \in \mathcal{K}_i} \tilde{m}_{s,i}^{(l)} \tilde{\sigma}_{s,i}^{-2(l)}, \quad (34b)$$

where the omitted proportionality constant cancels in the BME ratio. A new observation at an existing node changes its accumulated precision, precision weighted mean, and count. The Vecchia graph and $\mathbf{\Lambda}_{ks|h}$ remain available for reuse. A coordinate not already represented in the graph requires insertion into the ordering and extension or reconstruction of the graph. These equations describe repeated assimilation at fixed locations. Forecasting a changing groundwater system additionally requires a state transition model. When the latent field is large, preconditioned conjugate gradients (PCG) solve Eq. (13c) for $\mathbf{m}_{ks|hs}^{\text{BME}}$ (Hestenes and Stiefel 1952).

2.6 Evaluation sequence and common measures

Table 2 summarizes seven evaluation stages. Direct integration or BME supplies the comparison whenever its cost remains feasible. Wasserstein-1 distance measures the mass and displacement required to transform one univariate density into another (Peyré and Cuturi 2019). Jensen–Shannon divergence measures their symmetric information difference (Lin 1991). Exceedance error compares posterior probabilities above a specified threshold, and mode count records whether distinct alternatives remain present. All density measures are dimensionless in the standardized synthetic experiments.

For N_v independent validation observations, let y_i be the observed value and μ_i the posterior mean. Root mean squared error (RMSE) and mean absolute error (MAE) are

$$\text{RMSE} = \left\{ \frac{1}{N_v} \sum_{i=1}^{N_v} (y_i - \mu_i)^2 \right\}^{1/2}, \quad (35a)$$

$$\text{MAE} = \frac{1}{N_v} \sum_{i=1}^{N_v} |y_i - \mu_i|, \quad (35b)$$

321 where both errors have the units of the modeled variable. If $[L_i, U_i]$ is the stated central
 322 predictive interval, empirical coverage and mean interval width are

$$\text{Coverage} = \frac{1}{N_v} \sum_{i=1}^{N_v} \mathbb{I}(L_i \leq y_i \leq U_i), \quad (36a)$$

$$\text{Width} = \frac{1}{N_v} \sum_{i=1}^{N_v} (U_i - L_i), \quad (36b)$$

323
 324 where the indicator function is $\mathbb{I}$. Coverage is dimensionless and is interpreted against
 325 the nominal interval probability; width has the modeled units and measures sharpness.
 326 Coverage and width are interpreted together (Gneiting et al. 2007).

327 The continuous ranked probability score (CRPS) compares the predictive cumula-
 328 tive distribution with the step function at the observation. Its units match the modeled
 329 variable, and smaller values indicate a distribution that is both better calibrated and
 330 more concentrated near the observation. For a Gaussian prediction (Hersbach 2000;
 331 Gneiting and Raftery 2007),

$$\text{CRPS}(\mathcal{N}(\mu, \sigma^2), y) = \sigma \left[z \{2\Phi(z) - 1\} + 2\phi(z) - \pi^{-1/2} \right], \quad (37a)$$

$$z = \frac{y - \mu}{\sigma}, \quad (37b)$$

333 where the observed value is y , the predictive standard deviation is σ , the standardized
 334 residual is z , and the standard normal cumulative distribution and density are Φ and
 335 ϕ , respectively.

Table 2 Executed analysis sequence and comparison calculation for each question.

Stage	Design	Comparison calculation and reported measures
Hard data	Ten realizations with 100 observations; 40 targets; 4-32 conditioning neighbours; one map with 1,024 targets	BME; error in posterior mean and variance
Common soft data	Gaussian, interval, and likelihoods with right censoring; 24 cases with intervals at 1-8 locations	Direct integration at one location or conditional Monte Carlo; moments, density, convergence
Complex soft data	Five families, three severities, three target distances; 45 cases across families and settings	Direct integration in one dimension; Wasserstein-1, Jensen-Shannon, tails, modes, time
Automatic representation	The 45 cases with one soft location and 9 cases with two to four multimodal locations; mode and variance switch followed by component refinement	Direct or analytic mixture posterior; selected method, component count, posterior convergence, modes, Wasserstein-1, time
Biscayne soft data	71 hard locations; 9 continuous likelihoods; 10 below 100 mg/L; 6 above 2000 mg/L; five spatial folds; 2,592 targets	USGS 1000 mg/L isochlor as an external reference; RMSE, CRPS, coverage, Brier score, field change, time, memory, posterior difference
USGS prediction	128,228 approved observations from 2024; 293 wells; five folds defined by spatial blocks	Independent prediction wells; RMSE, CRPS, coverage, width, time
USGS retrospective fields	1,519,998 approved observations from 2020-2025; 1,709 wells; 774 regional grid locations; 72 months	Posterior depth, anomaly, marginal uncertainty, and well histories
Computational scale	Matched calculations from 250 to 8,000 observations; 123,048 well and month nodes; first one million observations	Sparse error, PCG residual, construction and solution time, throughput, and memory

Table 3 Numerical settings for the synthetic and Biscayne analyses. EP entries give the damping coefficient, stopping tolerance for the largest change in a site parameter, and maximum iterations.

Analysis	Graph and variance	EP settings	Likelihood and mixture
Numerical checks at one site	All predecessors; exact marginal variance	$1.0; 10^{-12}; 20$	Analytic moments for a truncated normal; direct integration reference
Multiple interval likelihoods	All predecessors; exact marginal variance	$0.7; 10^{-8}; 100$	Analytic moments for a truncated normal; conditional Monte Carlo reference
One complex likelihood	All predecessors; exact marginal variance	$0.7; 10^{-8}; 80$	512 quadrature nodes; 256-node check; $C_j = 10$
Simultaneous bimodal likelihoods	All predecessors; exact marginal variance	$0.65; 10^{-9}; 120$	2–4 sites; caps 2, 4, 8, and all 4–16 combinations
Biscayne field and local comparison	24 predecessors in field; all hard data in local comparison	Field: $0.7; 10^{-6}; 80$. Local: $0.7; 10^{-8}; 100$	512 and 256 quadrature nodes; 25 bounded distributions in local comparison

Table 4 Numerical controls for the USGS analyses. EP entries give damping coefficient, stopping tolerance, and maximum iterations. The retrospective and terminal calculations used Gaussian monthly likelihoods and did not require EP.

Analysis	Graph and variance	EP settings	Likelihood and uncertainty
USGS spatial prediction	32 predecessors; exact variance for graphs with at most 3,000 free nodes	$0.7; 10^{-6}; 50$	Empirical intervals at 20% of training locations
USGS retrospective and terminal tests	16 predecessors; stochastic marginal variance	Not used	32 Rademacher probes for fields and calibration; 8 per covariance sensitivity setting

2.6.1 Hard data and sparse conditioning

Each of ten Gaussian field realizations contained 100 hard observations and 40 prediction targets. We compared calculations with 4, 8, 16, and 32 conditioning neighbours with the BME posterior. A separate field with 80 observations and 1,024 targets tested

340 the spatial pattern of the posterior mean and variance. Complete conditioning and
341 21 numerical checks covered exact reconstruction, repeated observations, and extreme
342 interval limits.

343 **2.6.2 Common soft likelihoods**

344 Calculations at one location covered Gaussian likelihoods, intervals, and likelihoods
345 with right censoring; analytic expressions or direct integration supplied the comparison. The experiment at several locations comprised 24 cases with 1, 2, 4, or 8 interval
346 observations. Conditional Monte Carlo used at least 50,000 draws in each case. Mean
347 error, variance error, density agreement, convergence, and quadrature stability were
348 recorded separately.
349

350 **2.6.3 Complex soft likelihoods**

351 The first complex likelihood experiment combined five families, three severities, and
352 three target distances. Its 45 cases each contained one soft location. The likelihoods
353 were skew normal, Student- t , lognormal, bimodal Gaussian mixtures, or irregular tabulated densities. Direct integration in one dimension supplied Wasserstein-1 distance,
354 Jensen-Shannon divergence, exceedance probability, posterior modes, and calculation time. The second experiment placed multimodal soft data at two, three, or four
355 locations under three separations. These nine cases contained 4, 8, or 16 possible
356 combinations of Gaussian components. We used analytic conditioning for every combination as the exact comparison. VISTA-MEP propagated at most 2, 4, 8, or all
357 combinations, and VISTA-GEP used the same coordinates, covariance, and likelihoods. The analysis recorded included and omitted prior mass, effective number of
358 posterior components, density error, moments, modes, time, and memory.
359
360
361
362

363 2.7 Biscayne field evaluation with soft chloride likelihoods

364 The 2022 USGS release for Miami-Dade County contains chloride measurements,
365 specific conductance, electromagnetic logs, summary tables, and an approximate
366 1000 mg/L chloride isochlor at the base of the Biscayne aquifer (Zhang and Ren-
367 shaw 2024, 2025). Sparse well coverage limits the mapped position of this interface.
368 Coincident horizontal coordinates were combined by retaining the maximum mea-
369 sured chloride, leaving 71 independent hard locations. The response was the decimal
370 logarithm of chloride concentration in mg/L.

371 Soft information comprised nine continuous likelihoods, 10 chloride observations
372 below 100 mg/L, and six above 2000 mg/L. For specific conductance from 714 to
373 6572 $\mu\text{S}/\text{cm}$, the USGS conversion relates chloride concentration c to conductance s
374 through $c = 0.3243s - 131.44$ (Zhang and Renshaw 2024). A Water Quality Portal
375 retrieval provided 64 paired chloride and conductance measurements through 2020
376 (National Water Quality Monitoring Council 2026b). The logarithmic residuals about
377 the conversion had a standard deviation of 0.0242968. We represented their distri-
378 bution with a Gaussian kernel density estimate using Scott’s bandwidth rule (Scott
379 1992). Shifting this density to the chloride concentration calculated from each conduc-
380 tance measurement produced one soft likelihood at each of nine locations. Likelihoods
381 censored at the stated bounds represented the remaining 16 observations outside the
382 valid conversion range.

383 Five spatial folds withheld complete wells for independent prediction. Within a
384 fold, any soft location within 0.10 km of a withheld well was removed. The other four
385 folds supplied hard observations for estimating a constant mean and an exponential
386 covariance with a nugget. Scores were calculated for all 71 withheld wells and sepa-
387 rately for the 17 wells within 10 km of soft data. Prediction covered 2,592 locations in
388 the southern corridor with 24 conditioning neighbours. The official 1000 mg/L isochlor

389 was reserved as an external spatial comparison and excluded from parameter estima-
390 tion and prediction. A horizontal map transect followed the median northing of the
391 soft locations.

392 Expectation propagation for the complete field converged after 22 iterations with-
393 out a bound on local precision. The largest relative discrepancy between quadrature
394 resolutions was 0.00143666. A separate calculation compared BME and VISTA-BME
395 at one target. Both methods used the same 12 nearest hard observations and 25
396 bounded, tabulated probability distributions; the two censored ranges were repre-
397 sented as 10–100 and 2000–20000 mg/L. Three runs in separate processes supplied
398 the median and observed range. This calculation isolates likelihood integration at one
399 target; the complete Biscayne field has a separate timing result.

400 2.8 USGS prediction and calibration

401 The USGS OGC service supplied approved groundwater observations for parameter
402 72019 (U.S. Geological Survey 2026a). Complete extracts for 2024 covered a North-
403 east box from 76°W to 72°W and 38°N to 42°N and a Southwest box from 112°W
404 to 104°W and 30°N to 38°N. Values were converted from feet to metres. The eligi-
405 bility rules included 374 monitoring location identifiers at 293 horizontal coordinates.
406 Among these coordinates, 47 contained 128 identifiers in total, with as many as six
407 at one coordinate. Aggregation was required where screen elevation was unavailable.
408 Consequently, the modeled variable summarizes the horizontal monitoring network;
409 a potentiometric surface for a single aquifer or screen interval would require further
410 stratification. Complete 0.5° spatial blocks assigned independent prediction wells to
411 five folds.

412 Within every training fold, a deterministic 20% subset of locations supplied an
413 empirical interval between the 10th and 90th percentiles. The other locations supplied
414 hard medians. The same fold assignments were used for BME, a Vecchia Gaussian

process (Vecchia 1988; Katzfuss et al. 2020), VISTA-BME with 32 conditioning neighbours, a static EnKF with 256 members (Evensen 2003; Hendricks Franssen and Kinzelbach 2008), XGBoost (Chen and Guestrin 2016), and a GNN with 16 neighbours (Bai and Tahmasebi 2023). The static EnKF drew an ensemble from the fitted covariance and assimilated hard medians and soft midpoints; empirical interval width determined the variance assigned to each soft observation. XGBoost predictors comprised the coordinates and inverse distance summaries from 12 neighbours. The model used 600 trees of depth five, a learning rate of 0.035, row and column fractions of 0.85 and 0.90, and a weight of 0.35 for soft locations. Separate models for squared error and the 0.05 and 0.95 quantiles supplied its mean and interval. The GNN used a symmetric graph with 16 neighbours and two graph layers of 64 units. Its settings were dropout 0.15, AdamW with learning rate 0.01 and weight decay 10^{-4} , random masking of 20%, at most 800 epochs with a patience of 100 epochs, and 64 dropout draws for predictive variance. These implementations provide static interpolation comparisons; a state transition model or explicit likelihood assimilation would constitute a different design. Four further calculations considered a stationary coordinate trend, an expanded covariate mean, regional residual covariances, and their combination. Covariates included official well depth, land surface altitude, aquifer descriptors, topographic setting (U.S. Geological Survey 2026b), annual NASA POWER precipitation (National Aeronautics and Space Administration 2026a), and county groundwater withdrawals (U.S. Geological Survey 2020). Each training fold independently determined imputation, scaling, category levels, covariance fitting, and calibration.

Interval recalibration also followed the five spatial folds. For a target fold, the remaining four supplied the calibration scores. The absolute scheme used the 90th percentile of the absolute residual. The normalized scheme used the 90th percentile of the ratio between absolute residual and model standard deviation. This is a split conformal calculation (Angelopoulos and Bates 2023); normalization permits the resulting

interval width to vary with the modeled standard deviation. Conformalized quantile regression is a separate adaptive approach based on conditional quantiles (Romano et al. 2019). Differences in RMSE and CRPS between field models were summarized from 5,000 paired bootstrap resamples of the 69 spatial blocks (Efron 1979).

2.9 USGS retrospective field

The retrospective field used the complete local acquisition for approved parameter 72019 from 2020 through 2025 (U.S. Geological Survey 2026a). Conversion of 1,519,998 values from feet to metres and monthly averaging at 1,709 wells produced 123,048 nodes indexed by well and month. Observations were available at 98,026 of these nodes. For well i in month t , the temporal anomaly was

$$a_{i,t} = y_{i,t} - \text{median}_{u \in 2020:2025}(y_{i,u}), \quad (38)$$

where $a_{i,t}$ is the temporal anomaly and $y_{i,t}$ is the monthly mean depth to water. The value $y_{i,u}$ is the corresponding monthly mean for month u between 2020 and 2025. Subtracting the median at each well separates temporal change from persistent differences in depth among wells.

After the well nodes, the fixed ordering contained 774 prediction locations within the convex hulls of monitoring locations in the Northeast and Southwest boxes. Repetition over 72 months added 55,728 nodes indexed by grid location and month. Together, the wells and grid locations formed 178,776 latent nodes over 72 months. Each sparse conditional distribution used 16 earlier neighbours. Spatial coordinates were normalized over the continental extent of the wells, and one month equalled 0.65 units of normalized temporal distance. The covariance used a nugget of 0.08, partial sill of 0.92, and normalized practical range of 0.18. Monthly sampling variance determined observation variance within limits of 0.0001 and 0.02909 m².

465 Marginal posterior variance was estimated from a Rademacher calculation derived
 466 from stochastic trace estimation (Hutchinson 1990). This calculation avoids formation
 467 of the complete inverse. Denote the retrospective BME posterior precision by $\Lambda_{\text{ret}}^{\text{BME}}$,
 468 the independent Rademacher vector for probe r by $\mathbf{z}_r$, and the number of probes by
 469 R . The covariance diagonal is estimated as

$$\hat{\mathbf{d}} = R^{-1} \sum_{r=1}^R \mathbf{z}_r \odot (\Lambda_{\text{ret}}^{\text{BME}})^{-1} \mathbf{z}_r, \quad (39)$$

470 where $\hat{\mathbf{d}}$ estimates the posterior covariance diagonal and $\odot$ denotes elementwise multi-
 471 plication. The field calculation used $R = 32$. On a smaller graph, comparison with an
 472 exact inverse gave a median relative absolute error of 0.000475 and a 90th percentile of
 473 0.00601. Reported posterior variances treat the well baseline as fixed. Stochastic error
 474 was larger at prediction grid locations. The uncertainty maps therefore display the
 475 well nodes and mark estimates with a relative Monte Carlo standard error above 0.5.

476 A terminal period test distinguished retrospective reconstruction from temporal
 477 prediction. The first design estimated each well baseline, anomaly scale, observation
 478 variance, and sparse posterior from 70,455 observed cells between January 2020 and
 479 December 2024. All 13,871 observed cells at 1,266 wells during 2025 were reserved for
 480 testing. Seven calculations used the same 91,152 nodes indexed by well and month,
 481 ordering, 16 conditioning neighbours, and likelihood values. One covariance setting
 482 changed in each calculation: temporal weights of 0.35, 0.65, and 1.00; practical ranges
 483 of 0.12, 0.18, and 0.28; or nuggets of 0.04, 0.08, and 0.16 with total variance fixed
 484 at one. Eight Rademacher probes per setting measured interval sensitivity. A second
 485 calculation fitted observations through 2023, used 15,893 observed cells from 2024
 486 for split calibration, and tested 13,596 cells from 2025. Its 32 probes supplied raw,

absolute split conformal, and normalized split conformal intervals. Temporal exchangeability remains approximate, so these results describe calibration and carry no causal interpretation.

2.10 Computational scaling

The direct computational comparison assigned identical synthetic coordinates, Gaussian hard values, covariance parameters, and 128 targets to both methods. Observation count increased from 250 to 8,000. BME conditioned on every observation through the complete covariance matrix, while VISTA-BME used 32 neighbors. Three runs for each method and size were made in separate processes. Their medians and observed ranges describe wall time and peak resident memory. For the separate USGS scale, the number of occupied nodes indexed by well and month determined the storage required by one complete covariance matrix.

The scale analysis read the first one million observations in a continental acquisition covering 2020–2025. Aggregation by 1,709 wells and 72 calendar months produced 123,048 possible combinations and 97,986 occupied cells. Because this sequence samples the acquisition file for a computational test, observation counts near its end do not represent a census of USGS availability.

2.11 Use of generative artificial intelligence tools

OpenAI ChatGPT and Codex assisted with source discovery, software development, figure and table code, manuscript organization and drafting, language revision, and checks of reproducibility and integrity. The authors defined the scientific questions and methods, supplied and interpreted the data, ran the repository scripts, checked references and numerical outputs against the source records, and reviewed the manuscript. Responsibility for its accuracy and integrity remains with the authors.

511 3 Results and Discussion

512 3.1 Posterior agreement and sparse approximation with hard 513 data

514 Complete conditioning reconstructed the Gaussian covariance precision and matched
515 the posterior mean and marginal variance from the traditional BME calculation within
516 2×10^{-10} . The 21 numerical checks extended this comparison to moments at one
517 location, a bimodal mixture, extreme intervals, and sufficient statistics for repeated
518 observations at one location.

519 Increasing the number of conditioning neighbors reduced the error at every tested
520 level. For 4, 8, 16, and 32 neighbors, the median RMSEs of the posterior mean were
521 0.0669, 0.0218, 0.00509, and 0.000708. The median variance RMSE fell from 0.0101 to
522 2.17×10^{-6} . In the separate field with 80 observations and 1,024 targets, 32 neighbors
523 gave a posterior mean RMSE of 0.00361 and a posterior standard deviation RMSE
524 of 0.000389. The signed differences were concentrated around particular observation
525 configurations (Fig. 2).

Table 5 Numerical comparisons for Gaussian, interval, and sequential calculations.

Experiment	Cases	Result
Complete Gaussian conditioning set	numerical verification	posterior agreement within 2×10^{-10}
Interval at one location	15	maximum mean and variance errors 2.22×10^{-16} and 8.88×10^{-16}
Intervals at multiple locations	24	all converged; maximum mean RMSE 0.00438 and variance RMSE 0.00927
Sequential observations at one fixed location	100 updates	maximum mean and variance errors 4.00×10^{-15} and 1.11×10^{-15}

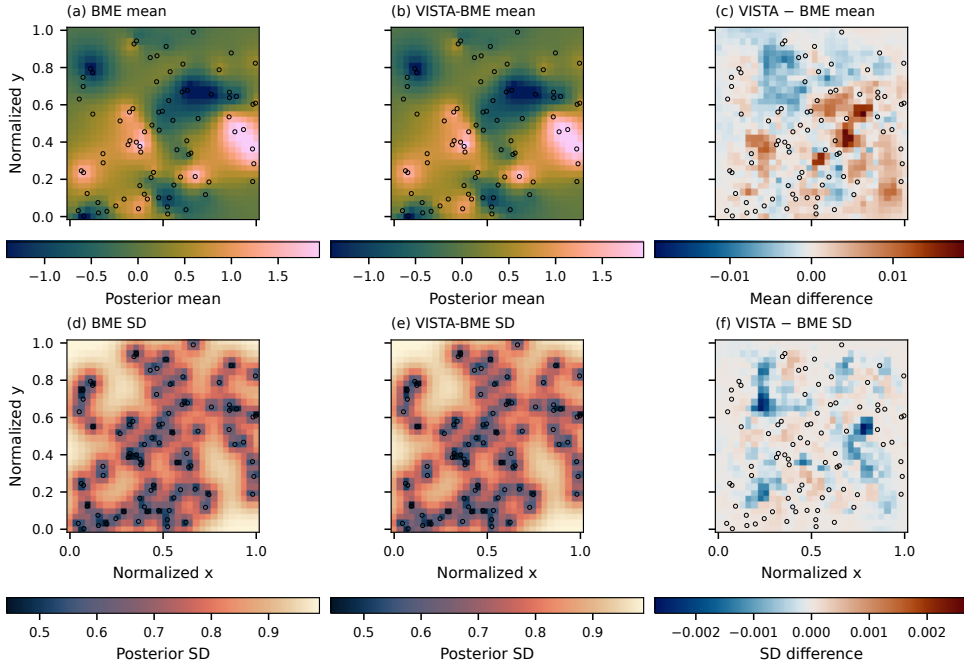


Fig. 2 BME and VISTA-BME with 32 conditioning neighbours for the reproducible field with 80 observations: posterior means on a shared scale for (a,b), their signed difference for (c), posterior standard deviations on a shared scale for (d,e), and their signed difference for (f). Open circles mark the same observations in all panels; 1,024 targets are compared.

526 3.2 Common soft likelihoods: accuracy and convergence

527 The 15 interval calculations at one location converged within two iterations and repro-
528 duced the analytic moments of the truncated Gaussian distribution to numerical
529 roundoff. Across the 24 cases with several locations, median posterior mean RMSE
530 against conditional Monte Carlo was 0.00317, 0.00236, 0.00338, and 0.00285 for 1, 2,
531 4, and 8 interval locations. The corresponding variance RMSEs were 0.00444, 0.00387,
532 0.00496, and 0.00781. All runs converged without a local precision bound; each Monte
533 Carlo comparison used at least 50,000 accepted draws.

534 The largest mean discrepancy occurred with four interval locations. GEP produced
535 a target mean of 0.3391 and standard deviation of 0.9099; conditional Monte Carlo
536 gave 0.3451 and 0.9077. The largest variance discrepancy occurred with eight interval
537 locations. GEP gave 0.01593 and 0.9808 for the two reported quantities, compared
538 with 0.01348 and 0.9862 from Monte Carlo. The target density curves in Fig. 3 remain
539 closely aligned despite these moment differences.

540 Across the common-likelihood grid at one location, GEP had a median Wasserstein-
541 1 distance of 1.03×10^{-5} , median absolute mean error of 2.26×10^{-9} , and median
542 absolute variance error of 1.50×10^{-10} . The moment-matched Gaussian had a median
543 Wasserstein-1 distance of 2.56×10^{-6} in the 18 cases where that comparison was
544 defined. Under the tested severities and distances, GEP reproduced the bounded and
545 censored likelihood results closely.

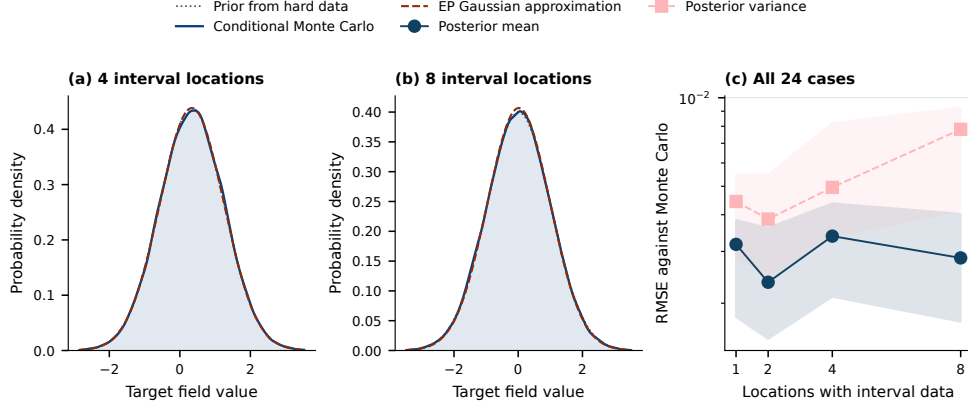


Fig. 3 Validation with multiple interval observations: target marginals for (a) the case with four interval locations and the largest posterior mean RMSE and (b) the case with eight interval locations and the largest posterior variance RMSE, followed by (c) median RMSE of the posterior mean and variance over 24 cases. Gray dotted curves are priors calculated from hard data, blue curves are conditional Monte Carlo densities with Scott’s bandwidth, and red dashed curves are GEP approximations; bands in (c) span the observed minima and maxima.

3.3 Complex soft likelihoods: density shape and cost

The experiment with one density at a time distinguished smooth unimodal likelihoods from distributions with persistent multiple modes (Table 7). Across 45 cases, the Gaussian with matched moments had a median Wasserstein-1 distance of 0.00276. GEP reduced this distance to 9.40×10^{-5} , and MEP reduced it to 6.60×10^{-5} . MEP gave the smaller Wasserstein-1 and Jensen–Shannon errors in 23 paired cases. The result varied with likelihood family and tested setting.

The nine simultaneous bimodal cases tested whether the calculation kept separate modes (Fig. 4a,b). With two to four soft sites, GEP had a median Wasserstein-1 distance of 0.0341 and a maximum of 0.308. MEP with all 4 to 16 combinations reproduced the exact Gaussian mixture posterior to numerical roundoff. Its maximum Wasserstein-1 distance was 4.34×10^{-15} ; maximum absolute errors in posterior mean

558 and variance were 2.72×10^{-15} and 4.22×10^{-15} . Complete MEP also had the smaller
559 median time in these small cases (Table 7). Limiting the propagated set discarded up
560 to 0.815 of the prior component mass, and density error grew with discarded mass.
561 Complete propagation kept every represented component, whereas the component
562 limit set the measured balance between density fidelity and cost.

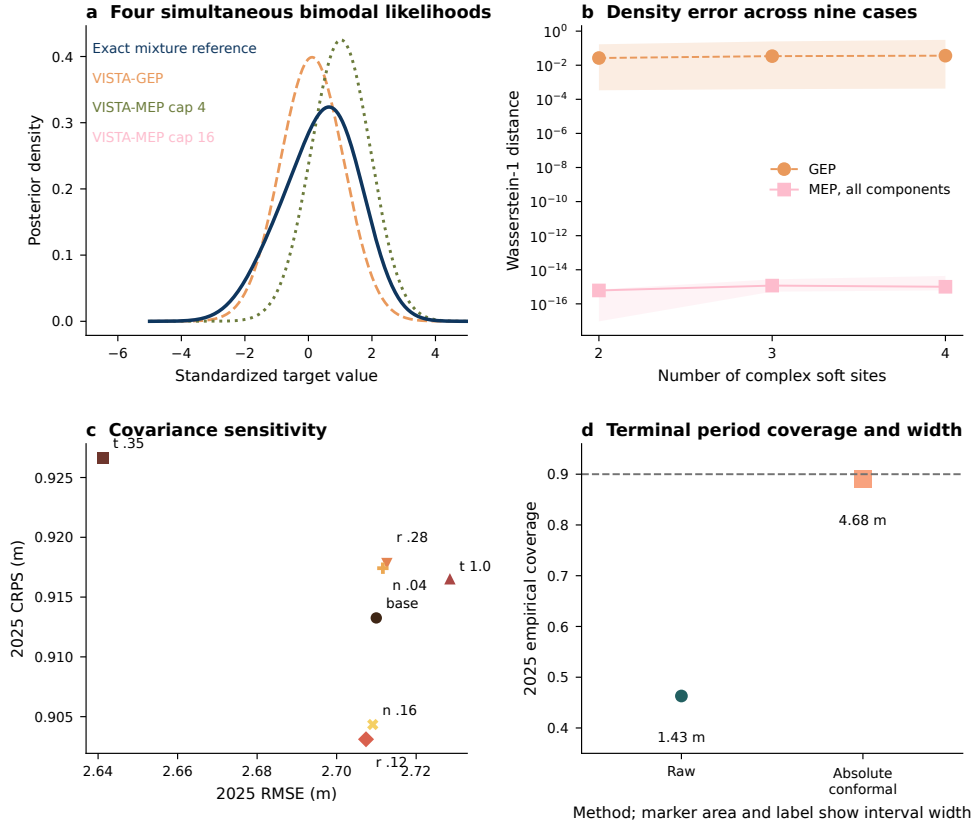


Fig. 4 Additional validation of complex soft data and groundwater prediction for a terminal period: (a) exact and approximate target densities for four simultaneous strongly bimodal soft likelihoods; (b) Wasserstein-1 distance for two to four simultaneous likelihoods over three separations, with bands spanning the observed range; (c) 2025 point prediction sensitivity after fitting through 2024, where labels identify temporal weight (t), practical range (r), and nugget (n); and (d) raw and absolute split conformal intervals after fitting through 2023, calibrating on 2024, and testing on 2025. Marker area and labels in (d) give mean interval width, and the dashed line marks 0.90 coverage.

563 GEP gave its smallest errors for the smooth families. Median Wasserstein-1 dis-
564 tances were 2.38×10^{-5} for skew normal, 5.79×10^{-6} for Student- t , and 1.48×10^{-5}
565 for lognormal likelihoods. MEP medians for these families ranged from 6.60×10^{-5} to

566 2.93×10^{-4} . The relative difference between the 512-node and 256-node integrations
567 stayed near 10^{-9} for these moments.

568 Strong bimodality marked the setting in which a single Gaussian site removed
569 one of the posterior modes. For bimodal mixtures, GEP had a median Wasserstein-1
570 distance of 2.44×10^{-3} and median relative variance residual of 0.212; the most severe
571 posterior became unimodal. MEP kept both components and reproduced the analytic
572 mixture at one location with a median Wasserstein-1 distance of 6.45×10^{-16} . For
573 irregular tabulated likelihoods, MEP reduced the within-family median from $1.79 \times$
574 10^{-3} with GEP to 1.67×10^{-4} . Table 7 gives the time for one soft location. With
575 several mixture locations, the component limit determines how many combinations
576 are propagated.

577 The automatic rule in Eq. (33) was evaluated with the same complete predecessor
578 sets and reference posteriors (Table 6). Among 45 cases with one soft location, it used
579 GEP in all 27 smooth cases and in the weakest severity of both multimodal families.
580 The two higher severities selected MEP, giving 33 GEP and 12 MEP calculations.
581 Tabulated densities selected 4, 10, or 32 components according to target distance and
582 severity; all component searches satisfied the four tolerances. The automatic posterior
583 reproduced the reference mode count in every case. Its median Wasserstein-1 distance
584 was 1.48×10^{-5} and its maximum was 0.0106. Median calculation time was 9.79 ms;
585 the 12 switched cases had a median of 43.8 ms because each calculation included the
586 initial GEP result and successive mixture refinements.

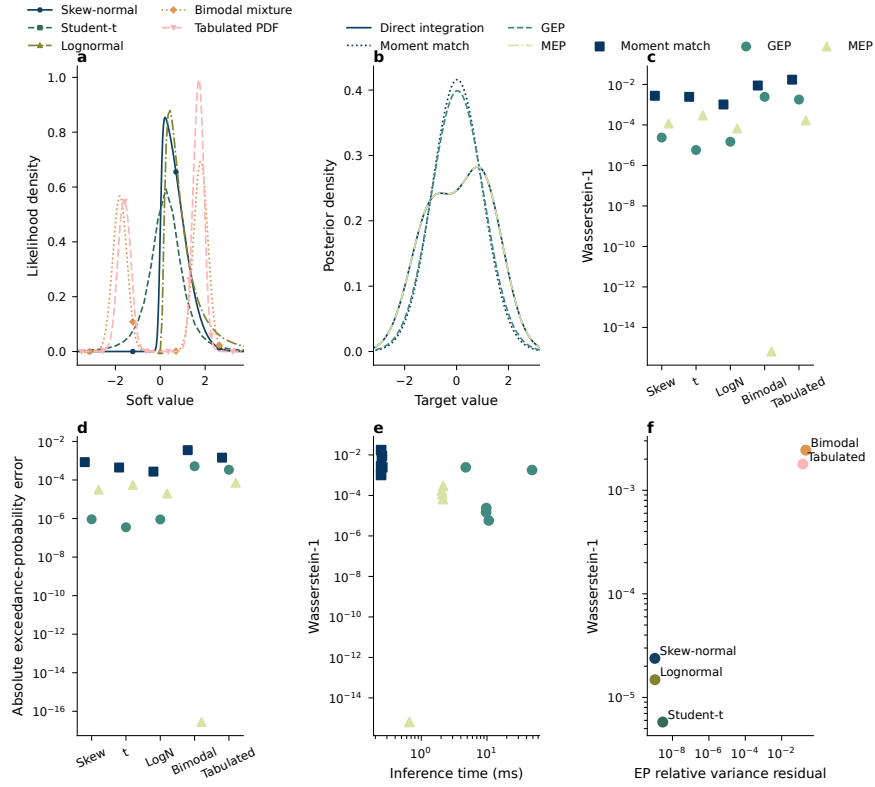
587 The nine calculations with two to four simultaneous bimodal locations selected
588 GEP for the weakest separation and MEP for the two larger separations at every site
589 count. The automatic posterior again reproduced every reference mode count. Median
590 Wasserstein-1 distance was 2.72×10^{-15} , the maximum was 4.26×10^{-4} , and median
591 time was 14.6 ms. No component search was unresolved, and the largest omitted prior
592 component mass was 1.11×10^{-16} .

Table 6 Validation of automatic GEP and MEP selection with complete predecessor sets.

Design	Cases	GEP	MEP	Unresolved	Correct modes	Median W_1	Maximum W_1	Median time (ms)
One soft location	45	33	12	0	1.00	1.48×10^{-5}	0.0106	9.79
Two to four locations	9	3	6	0	1.00	2.72×10^{-15}	4.26×10^{-4}	14.6

Table 7 Median results for complex likelihoods over five families, three severities, and three target distances.

Method	Cases	W_1	JS	Mean error	Tail error	Time (ms)
Moment-matched Gaussian	45	2.76×10^{-3}	1.04×10^{-6}	2.76×10^{-3}	6.02×10^{-4}	0.249
GEP	45	9.40×10^{-5}	2.84×10^{-9}	9.95×10^{-9}	2.09×10^{-5}	9.98
MEP	45	6.60×10^{-5}	6.94×10^{-10}	4.03×10^{-5}	1.77×10^{-5}	2.06

**Fig. 5** Accuracy and cost for complex soft data: (a) strongest tested likelihoods, (b) posterior density for the strongest bimodal case, (c) median Wasserstein-1 distance, (d) absolute error in exceedance probability over three severities and three target distances, (e) density error against measured calculation time for each likelihood family, and (f) variance residual from the local Gaussian approximation against density error by family. Direct integration at one location supplies the comparison in (b)–(f).

593 **3.4 Biscayne field: soft likelihood value and computational cost**

594 Adding the 25 soft likelihoods changed the mapped chloride field locally, with a max-
595 imum posterior mean change of 0.815050 on the logarithmic chloride scale (Fig. 6).
596 The four panels trace the same hydrologic question: how conductance uncertainty and
597 censored chloride observations shift the inferred saltwater boundary. The observation
598 map identifies the inputs, the posterior field shows the resulting surface, the signed
599 difference locates the change, and the transect gives the corresponding profile.

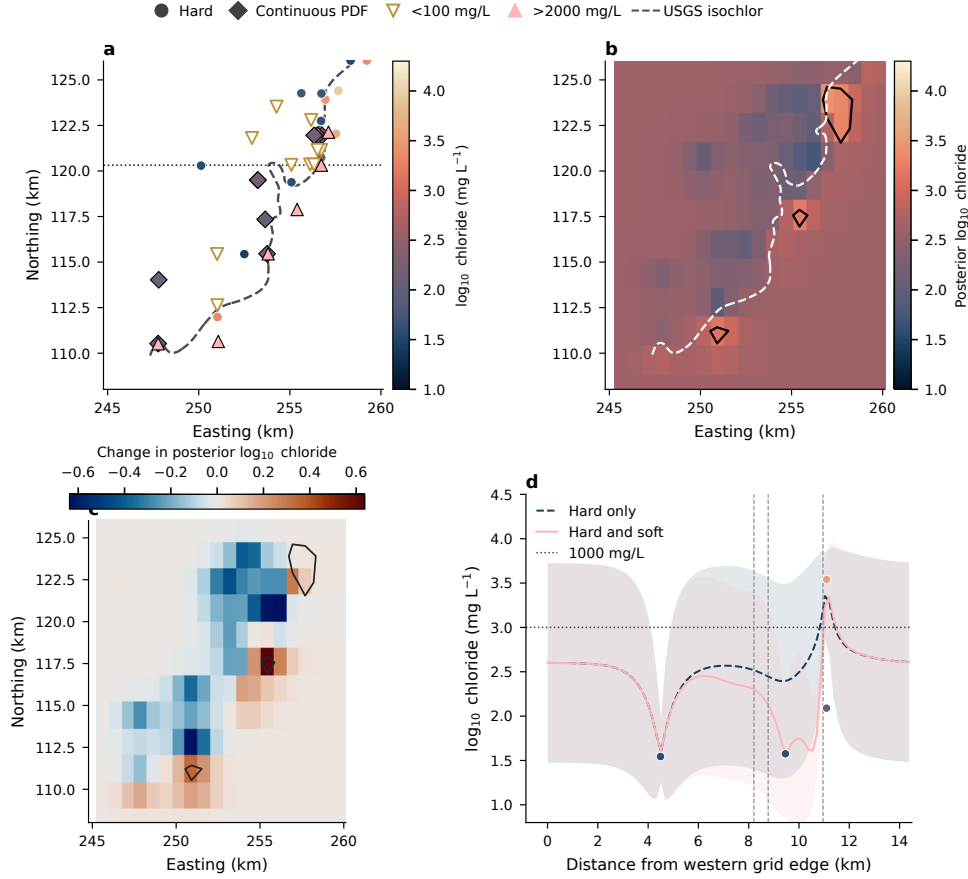


Fig. 6 Biscayne aquifer chloride in 2022: (a) hard measurements, continuous conductance likelihoods, censored likelihoods, the mapped USGS 1000 mg/L isochlor, and the transect location; (b) posterior mean from all hard and soft data, with posterior and USGS isochlors shown by solid and dashed lines; (c) change in posterior mean after adding the soft likelihoods, with the posterior isochlor; and (d) posterior means and 90% marginal intervals along the transect. Points in (d) are nearby hard measurements, and vertical dashed lines mark intersections with the USGS isochlor. Data are from the USGS Biscayne aquifer release and the Water Quality Portal (Zhang and Renshaw 2024; National Water Quality Monitoring Council 2026b).

600 At the 17 independent wells within 10 km of soft data, hard data alone gave
 601 RMSE 1.013, CRPS 0.663, 0.765 coverage for the 90% interval, mean width 2.224, and

602 Brier score 0.256 for exceedance of 1000 mg/L. The full likelihood gave RMSE 1.008
603 and CRPS 0.654, with coverage still 0.765, mean width reduced to 2.121, and Brier
604 score 0.264. RMSE and CRPS decreased by 0.52% and 1.37%. Across 10,000 paired
605 bootstrap resamples of withheld wells, the 95% intervals for the RMSE and CRPS
606 differences were -0.122 to 0.139 and -0.103 to 0.112 on the logarithmic scale. Each
607 interval crossed zero. The likelihoods narrowed the empirical intervals while preserving
608 their coverage; the Brier score for the threshold probability increased. Across all 71
609 wells, RMSE and CRPS changed by less than 0.4%.

610 The complete VISTA-BME field converged in 22 iterations. The largest rela-
611 tive difference between the two quadrature resolutions was 0.00144. The one-target
612 comparison isolates the computational effect of likelihood integration (Fig. 7). VISTA-
613 BME became faster once at least four soft probability distributions were present. With
614 12 hard observations, one target, and 25 distributions, wall time fell by a factor of
615 19.98 and peak resident memory by a factor of 2391 relative to BME. Posterior mean
616 and variance differences were 0.0117 and 0.00216 on the logarithmic chloride scale.
617 The figure gives the measured time and memory; the Gaussian scaling experiment
618 extends the comparison over observation counts.

Table 8 Spatial prediction at USGS wells omitted from fitting in five folds. Values are fold means; time is the fold median.

Method	RMSE (m)	CRPS (m)	Coverage 0.90	Width (m)	Time (s)
BME	31.108	16.418	0.600	27.153	0.006
Vecchia GP	31.147	16.452	0.607	27.163	0.060
VISTA-BME	31.156	16.519	0.604	27.135	0.980
XGBoost	30.826	17.774	0.116	10.210	0.405
GNN	32.924	20.402	0.075	3.832	0.539
EnKF	137.024	103.905	0.027	9.730	0.037

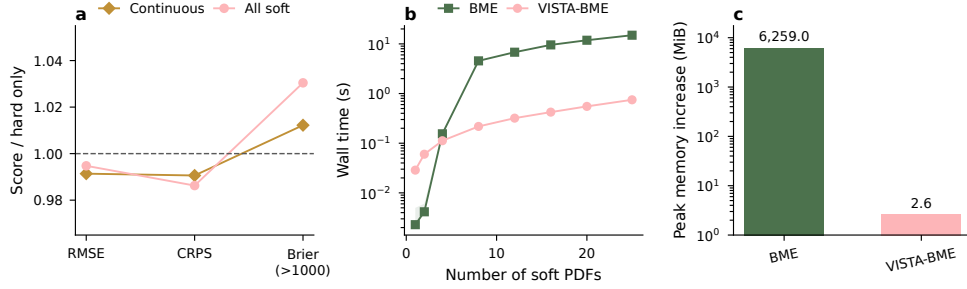


Fig. 7 Biscayne predictive and computational comparisons: (a) scores at 17 independent wells within 10 km of soft data, divided by the score from hard data alone, where values below one indicate improvement; (b) measured wall time for one target when BME and VISTA-BME use the same tabulated probability distributions, with ranges from three repeats; and (c) peak increase in resident memory with all 25 soft probability distributions.

3.5 USGS prediction: comparison and calibration

At independent prediction wells, BME, Vecchia GP, and VISTA-BME had similar point accuracy (Table 8). Their RMSE values ranged from 31.108 to 31.156 m, and their CRPS values ranged from 16.418 to 16.519 m. XGBoost gave the smallest RMSE, 30.826 m, with nominal 90% intervals covering 0.116 of the independent observations. GNN coverage was 0.075. These interval widths provide little uncertainty information in this comparison. The static EnKF with 256 members gave an RMSE of 137.0 m under the specified interpolation design.

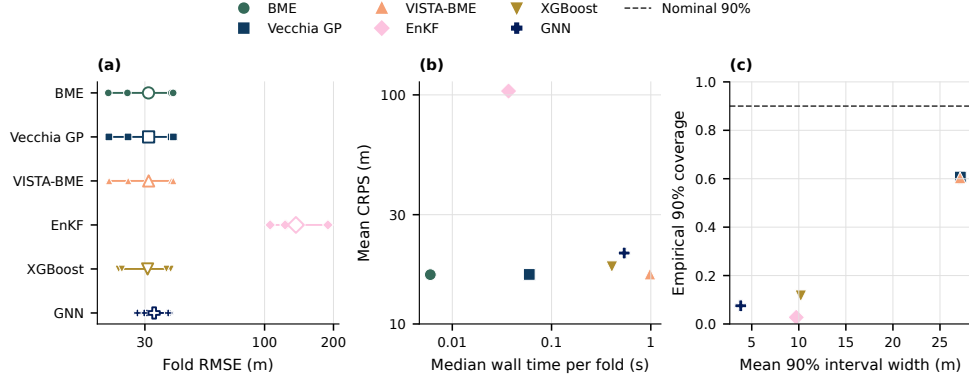


Fig. 8 USGS comparison over five spatial folds: (a) RMSE distributions with individual folds, medians, and observed ranges; (b) mean CRPS against median wall time on logarithmic axes; and (c) empirical 90% coverage against mean interval width, with the dashed line marking nominal coverage. The same markers identify each method throughout.

627 The additional field specifications separated point accuracy from interval width.
628 Official covariates reduced RMSE to 30.556 m and gave uncalibrated coverage of 0.412.
629 Regional covariance increased coverage to 0.655 and reduced CRPS to 16.087 m; RMSE
630 was 31.756 m. Combining the two gave RMSE 30.780 m, CRPS 16.261 m, coverage
631 0.641, and mean width 38.924 m. Relative to the stationary sparse model, the combined
632 differences were -0.554 m in RMSE and -0.219 m in CRPS. Spatial-block bootstrap
633 intervals were -3.638 to 2.500 m and -2.798 to 2.122 m, so the average differences
634 remain uncertain across the sampled blocks.

635 Nested normalized conformal calibration raised coverage for the combined model
636 to 0.891 and changed CRPS to 16.185 m. Mean width of the 90% interval increased to
637 101.113 m. Absolute conformal calibration gave 0.901 coverage with width 100.926 m.
638 This increase in width points to unresolved variation among aquifers and well screens
639 as a larger uncertainty source than the finite conditioning error measured in the
640 synthetic comparison.

3.6 USGS retrospective fields, 2020–2025

The absolute depth field varied much more among wells than the monthly change at an individual well (Fig. 9). Separate colour scales retain the spatial contrast within each region. Across the four displayed dates, predicted depth ranged from -2.15 to 40.70 m in the Northeast and from -0.95 to 164.73 m in the Southwest. Negative source values can represent water levels above the local land surface. With fixed scales, the four fields appear nearly stationary because regional contrasts exceed the submeter changes in regional means. Absolute depth consequently emphasizes persistent spatial differences, while smaller temporal changes are easier to read in the anomaly field.

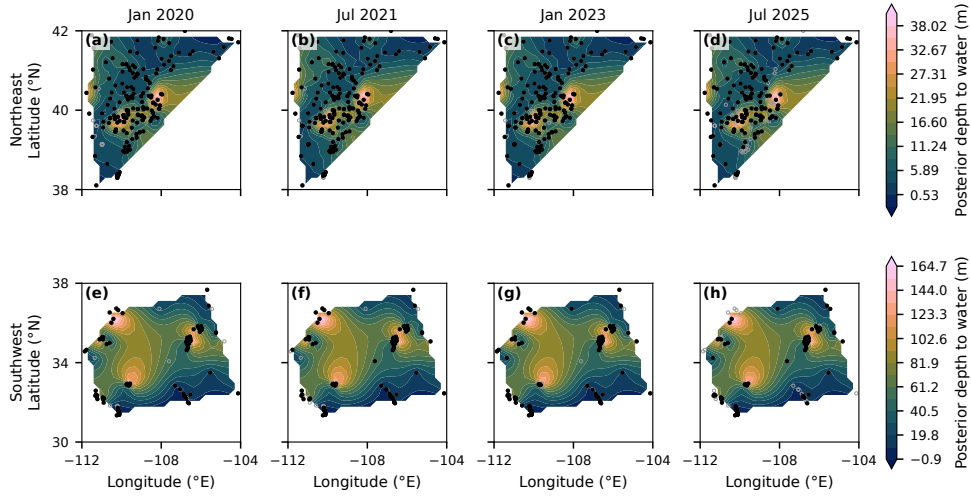


Fig. 9 Posterior depth to water in the Northeast for (a) January 2020, (b) July 2021, (c) January 2023, and (d) July 2025, with the corresponding Southwest fields in (e)–(h). Black points mark wells observed during the displayed month; open gray points mark other wells in the monitoring network. Separate fixed colour scales are used for the two regions.

Subtracting the median at each well places temporal change on a common scale (Fig. 10). Northeast mean anomalies on the four dates were -0.387 , 0.180 , -0.233 ,

652 and 0.340 m; the corresponding Southwest values were -0.448 , 0.170 , -0.219 , and
 653 0.288 m. Local anomalies ranged from -2.260 to 2.742 m in the Northeast and from
 654 -3.613 to 7.922 m in the Southwest. The January maps had negative regional means
 655 and the July maps had positive means. The displayed sequence describes temporal
 656 contrast; attribution to recharge, pumping, or seasonal forcing requires explanatory
 657 variables and a groundwater flow model.

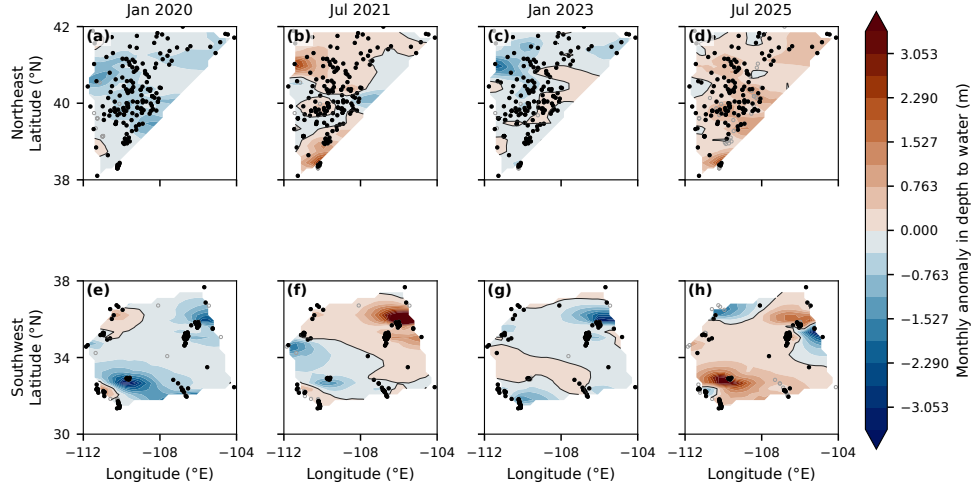


Fig. 10 Monthly depth to water anomaly relative to the 2020 to 2025 median at the same well for the Northeast in (a)–(d) and the Southwest in (e)–(h), shown for January 2020, July 2021, January 2023, and July 2025. The black contour marks zero anomaly. Black and open gray points indicate wells with and without observations in the displayed month, respectively. Colours are saturated at the symmetric 99.5th percentile to retain a common scale.

658 The representative records connect monthly observations to the reconstructed pos-
 659 terior series (Fig. 11). At the two frequently measured wells, posterior means follow
 660 the observed monthly variation closely. At the intermittent wells, neighbouring wells
 661 and adjacent months fill the gaps, and the interval widens when a local observation is

absent. The output table retains a posterior mean and standard deviation for every well and month, including months without an observation.

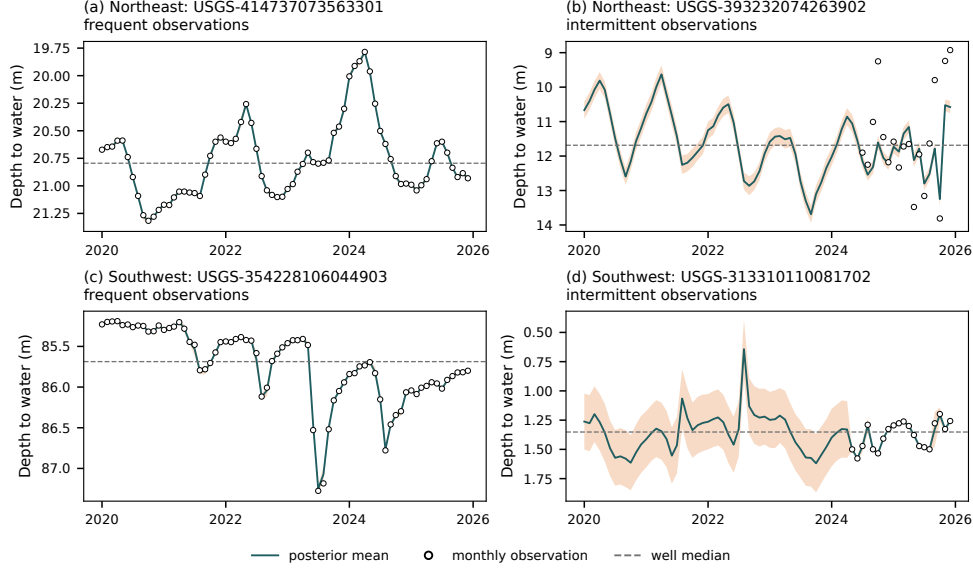


Fig. 11 Observed and reconstructed depth to water for a frequently measured Northeast well (a), an intermittently measured Northeast well (b), a frequently measured Southwest well (c), and an intermittently measured Southwest well (d). Solid curves give posterior means, shaded regions give central 90% intervals, open circles give monthly observations, and dashed lines give the 2020 to 2025 median at each well. Depth axes increase downward.

Posterior uncertainty was smaller at months with observations (Fig. 12). Northeast median posterior standard deviation was 0.0228 m for observed cells and 0.1814 m for unobserved cells. Southwest medians were 0.0106 and 0.1624 m. Median relative Monte Carlo standard errors were 0.0826 and 0.0496, respectively. Crosses mark the few displayed estimates above 0.5, allowing the numerical error of the variance approximation to be read separately from posterior uncertainty.

670 The terminal period test separated point sensitivity from interval calibration
671 (Fig. 4c,d). Fitting through 2024, the baseline setting predicted 13,871 observed cells
672 in 2025 with RMSE 2.710 m, CRPS 0.913 m, raw 90% coverage 0.467, and mean width
673 1.291 m. Across seven covariance settings, RMSE ranged from 2.641 to 2.728 m and
674 CRPS from 0.903 to 0.927 m. Raw coverage ranged from 0.440 to 0.499 while mean
675 width ranged from 1.079 to 1.432 m. Point predictions changed less than interval cal-
676 ibration across the tested covariance choices. The eight-probe sensitivity calculations
677 produced 5.2% to 6.1% negative diagonal estimates before clipping; their coverage
678 values are consequently numerical sensitivity diagnostics.

679 The split terminal calculation fitted through 2023, calibrated on 15,893 observed
680 cells from 2024, and tested 13,596 cells from 2025. Raw coverage was 0.463 with mean
681 width 1.432 m. Absolute split conformal calibration raised coverage to 0.890 and width
682 to 4.684 m; CRPS decreased from 1.069 to 1.000 m. The normalized score was omitted
683 because 32 probes produced negative diagonal estimates before clipping for 10.0% of
684 latent nodes, making ratios based on the model standard deviation unstable. The two
685 results separate interval undercoverage from instability in the variance estimator. The
686 absolute calibration result remains a calibration diagnostic, not a causal forecasting
687 guarantee.

Table 9 Scale and calculation time for the complete 2020 to 2025 retrospective field calculation.

Quantity	Value
Approved observations	1,519,998
Well and month nodes	123,048
Grid and month nodes	55,728
Joint latent nodes	178,776
Graph construction	16.403 s
Posterior mean	1.039 s
Marginal variance	63.495 s

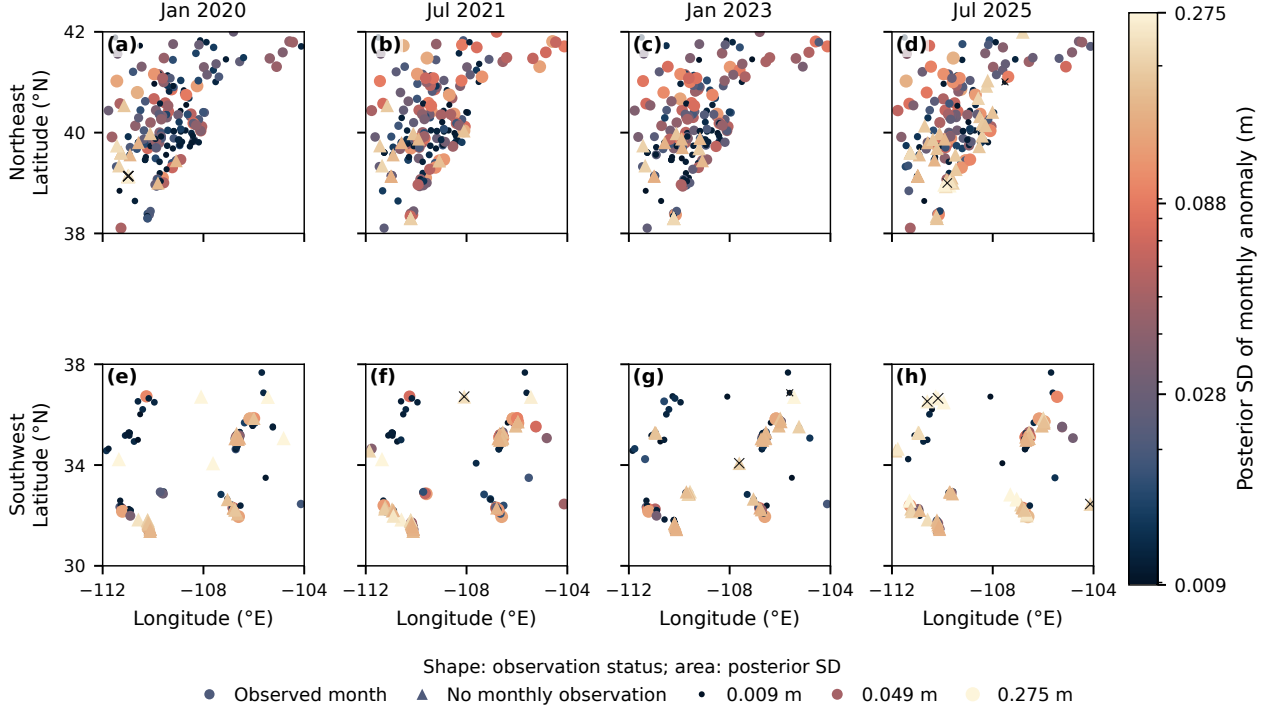


Fig. 12 Posterior standard deviation at well and month nodes for the Northeast in (a)–(d) and the Southwest in (e)–(h), shown for January 2020, July 2021, January 2023, and July 2025. Circles identify wells observed during the month and triangles identify wells without a monthly observation. Marker area and colour both increase with posterior standard deviation; crosses mark variance estimates with relative Monte Carlo standard error above 0.5. The logarithmic scale is shared by both regions, and uncertainty in the empirical well median is outside this calculation.

688 The sparse graph contained 3,039,056 nonzero entries. Construction took 16.403 s,
689 solution for the posterior mean took 1.039 s, and the marginal variance calculation
690 with 32 probes took 63.495 s (Table 9). A complete covariance matrix for 178,776 latent
691 nodes would require 238.1 GiB in double precision before factorization or temporary
692 arrays. VISTA-BME produced all monthly posterior means and marginal uncertainty
693 estimates on one workstation without storing that matrix.

694 Online Resource 1 contains 72 frames from January 2020 through December 2025.
695 It uses the anomaly definition and colour scale in Fig. 10. The sequence reconstructs
696 observed spatial and temporal patterns by covariance interpolation. Responses to
697 pumping or recharge require a groundwater state transition model and validation in
698 future periods.

699 **3.7 Computational scale and observation history**

700 The USGS graph for 123,048 combinations of wells and months contained 2,091,680
701 nonzero entries with 16 conditioning neighbors. Construction took 8.709 s and the
702 posterior mean solution took 1.057 s; final process memory increased by 58.8 MiB.
703 The calculation expanded 1,709 wells across 72 months.

704 At fixed monitoring locations, aggregation reduced the first million observations
705 to 1,709 counts, precisions, and linear parameters before the spatial solve. Figure 14
706 reports the CPU and GPU throughput for this calculation of sufficient statistics. We
707 constructed and solved a separate graph with 100,000 irregular locations in 5.838 s;
708 process memory increased by 54.6 MiB. The two calculations separate accumulation
709 of repeated measurements from spatial graph construction and solution.

710 BME had the shorter time through 2,000 observations in the matched calculation.
711 VISTA-BME became faster at 4,000 observations. At 8,000 observations, its wall time
712 was lower by a factor of 2.49 and its peak resident memory by a factor of 25.1. Across
713 the matched cases, the sparse posterior mean had a median RMSE of 0.00373 relative

714 to BME; the largest absolute difference at one target was 0.0258. Figure 13 shows the
 715 measured times, memory, and repeat-to-repeat variation.

716 The field calculation lies beyond the measured BME range (Fig. 13). One com-
 717 plete covariance matrix for the 97,986 occupied well and month cells requires at least
 718 71.5 GiB in double precision. The matrix for all 123,048 latent locations requires
 719 112.8 GiB. Factorization and temporary arrays would require additional memory.
 720 At the USGS scale, these explicit storage requirements define the BME comparison.
 721 VISTA-BME constructed the sparse graph in 8.709 s, solved for the posterior mean
 722 in 1.057 s, and increased final process memory by 58.8 MiB.

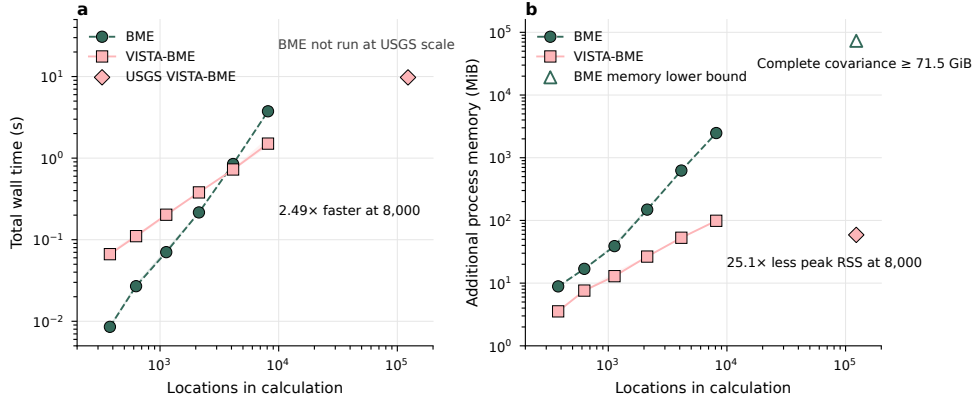


Fig. 13 Total wall time and additional process memory for BME and VISTA-BME with 32 conditioning neighbours: (a) median time and the measured range over three runs in separate processes for identical Gaussian observations and 128 targets, with the diamond giving the measured USGS graph and solution time; and (b) the corresponding peak increase in resident memory, with the open triangle giving the minimum memory for the complete covariance matrix among 97,986 occupied USGS well and month cells. The matched comparison reports peak resident memory, whereas the USGS diamond reports the final increase after the calculation with 16 conditioning neighbours. Both axes are logarithmic. The execution time for BME was not measured at the USGS scale.

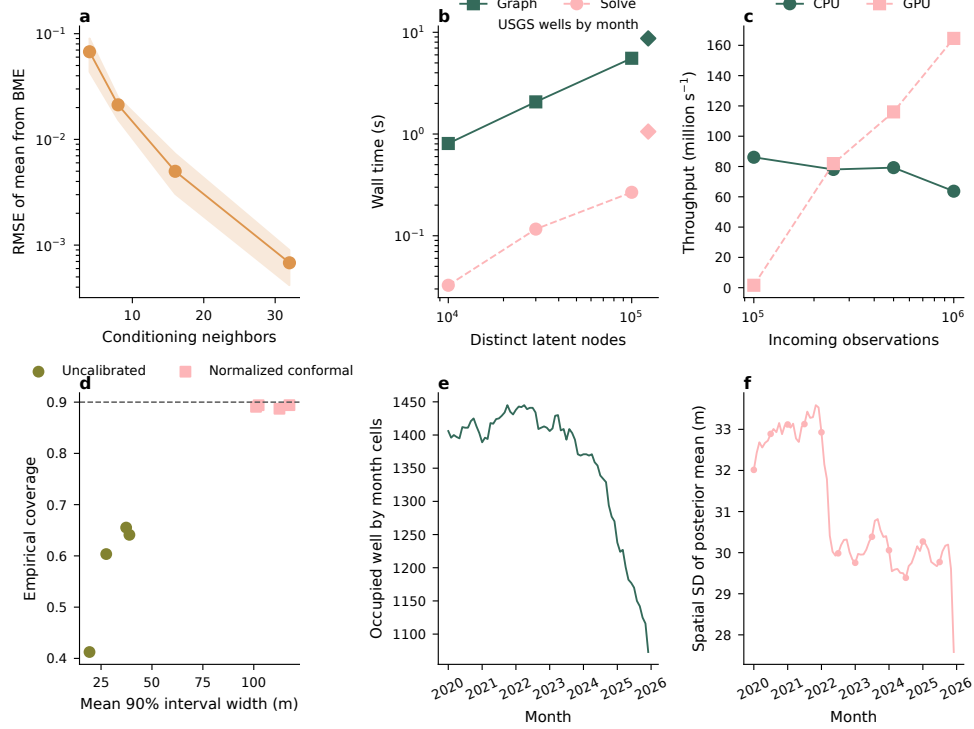


Fig. 14 Sparse accuracy, computational scale, and the USGS field calculation: (a) error in the posterior mean relative to BME over ten realizations, with shading from the 5th to the 95th percentile; (b) construction and solution time for synthetic locations, with diamonds marking 123,048 combinations of USGS wells and months; (c) throughput when likelihood terms are aggregated at fixed locations; (d) coverage and width before and after nested normalized conformal calibration; (e) monthly occupied cells in the sampled sequence of one million observations; and (f) spatial standard deviation of the inferred monthly well field from 2020–2025. The declines at the end of (e) and (f) reflect coverage of the sampled sequence and are not interpreted as hydrologic trends.

3.8 BME agreement and sparse approximation

Complete conditioning makes the sparse precision an alternative numerical representation of the matched continuous BME calculation; it reproduces the traditional BME

726 result to numerical roundoff. Finite conditioning sets introduce a measurable approx-
727 imation, and its error decreases monotonically as the number of neighbours increases.
728 The neighbour count thus controls both approximation accuracy and computational
729 cost.

730 3.9 Common and complex soft likelihoods

731 Gaussian expectation propagation reproduced the tested interval and censoring results,
732 as well as the smooth skewed densities and densities with heavy tails. The complex
733 likelihood cases changed when distinct modes remained after conditioning. A single
734 Gaussian site then preserved moments but could remove the shape governing threshold
735 probabilities or alternative states. Explicit mixture propagation kept that shape when
736 required. Both accuracy and calculation time varied with the likelihood shape.

737 The automatic rule made this distinction without assigning every multimodal input
738 to MEP. It used GEP for the weakest bimodal and tabulated distributions because
739 their variance residuals did not exceed 10^{-2} , and it selected MEP for the two higher
740 severities. Across the 45 cases with one soft location and nine cases with several soft
741 locations, every automatic component search met the density and posterior tolerances
742 and every posterior reproduced the reference mode count. The additional component
743 search increased median time to 43.8 ms among the 12 switched cases with one soft
744 location. This cost is incurred only after both the mode and variance criteria indicate
745 that the GEP representation is inadequate.

746 Direct integration covered one complex likelihood at nine combinations of severity
747 and target distance within each family. The multiple-location experiment included as
748 many as eight interval observations. Complete MEP reproduced the analytic posterior
749 of each represented Gaussian mixture for the tested cases with two to four sites and at
750 most 16 component combinations. Density error increased with prior mass omitted by
751 a finite component limit; omitted mass and density error therefore delimit the tested

approximation range. The current calculation uses a product of soft factors across locations. Shared conversion or measurement errors require joint factors or latent error variables and a corresponding extension of the likelihood model.

3.10 Field information from soft chloride data

The Biscayne aquifer case contains three separate results: field interpretation, withheld-well prediction, and computation. Soft likelihoods changed the mapped chloride field locally and modestly reduced RMSE and CRPS near soft locations. The 90% interval became narrower while empirical coverage stayed the same; the Brier score for the 1000 mg/L threshold increased. Paired bootstrap intervals crossed zero for changes in RMSE, CRPS, Brier score, and coverage, so the predictive differences are descriptive. The likelihoods retain conductance uncertainty and censoring in the field calculation, and the map locates their influence on the posterior.

The matched BME calculation supplies the computational comparison. With 25 soft distributions, VISTA-BME reduced wall time by a factor of 19.98 and peak memory by a factor of 2391. The simultaneous posterior mean difference was 0.0117 on the logarithmic chloride scale, quantifying the finite sparse approximation associated with these savings.

3.11 Prediction, calibration, and operational scale

Point prediction accuracy was comparable among BME, Vecchia GP, and VISTA-BME in the USGS comparison. XGBoost reduced RMSE by about 0.33 m relative to VISTA-BME, but its nominal 90% intervals covered 0.116 of the independent observations. VISTA-BME retains explicit soft distributions and permits repeated updates across a large latent field. Its general knowledge model accommodates mean functions from covariates or groundwater model outputs and covariance functions defined by hydrostratigraphy.

777 Adding covariates or regional covariance gave modest average changes, with boot-
778 strap intervals crossing zero. Conformal calibration reached nominal coverage after
779 interval widths increased to approximately 101 m. The result points to unresolved
780 variation among aquifers, well screens, and regional vertical datums. The network con-
781 tained 374 eligible monitoring identifiers at 293 horizontal coordinates; 128 identifiers
782 were combined at 47 coordinates. Such aggregation can mix aquifers and screen inter-
783 vals. The maps describe horizontal summaries of the monitoring network. Aquifer and
784 screen stratification remains necessary where metadata permit.

785 The retrospective calculation from 2020 to 2025 extends the comparison from inde-
786 pendent spatial prediction to field reconstruction. Absolute depth maps retain regional
787 contrasts. Anomaly maps isolate monthly change from the median over six years at
788 each well, and well histories separate observed from unobserved months. The terminal
789 period test found stable point error across covariance settings and pronounced raw
790 undercoverage. Absolute split calibration approached nominal coverage in 2025. Nor-
791 malized calibration was unsuitable because stochastic variance estimates were unstable
792 at extrapolation locations. These fields and well histories provide a retrospective mon-
793 itoring record. Causal forecasts require a groundwater state transition model and
794 validation against future pumping and recharge responses.

795 **3.12 Relation to scalable Bayesian spatial inference**

796 Scalable spatial methods obtain sparsity in different ways. Integrated nested Laplace
797 approximation commonly uses a Gaussian Markov random field whose sparse precision
798 follows from a stochastic partial differential equation representation of a Matérn field
799 (Rue et al. 2009; Lindgren et al. 2011). Vecchia approximations and nearest neighbour
800 Gaussian processes instead construct sparse conditional relations from covariance val-
801 ues for selected predecessor locations (Vecchia 1988; Datta et al. 2016; Katzfuss and

Guinness 2021). VISTA-BME uses this ordered conditioning within the BME general knowledge model. Predictive processes, inducing variables, and multiresolution approximations reduce dimension through basis representations (Banerjee et al. 2008; Titsias 2009; Katzfuss 2017). These constructions have different approximation errors, so the measured decrease in error with predecessor count in Fig. 14a applies to the present Vecchia calculation.

The treatment of soft information provides a second distinction. Vecchia-Laplace combines ordered conditioning with a Laplace approximation for observations outside the Gaussian family (Zilber and Katzfuss 2021). GEP uses repeated moment matching between each original soft function and its leave one out Gaussian distribution (Minka 2001; Nickisch and Rasmussen 2008). This operation accommodates the interval, censored, skewed, heavy tailed, and tabulated functions tested here without differentiating their logarithms. MEP retains separate Gaussian components when the posterior remains multimodal. Published comparisons of Laplace and expectation propagation concern particular likelihoods and data designs (Kuss and Rasmussen 2005; Nickisch and Rasmussen 2008; Cseke and Heskes 2011); they do not establish a universal ordering between the approximations.

The empirical study compares VISTA-BME with traditional BME, a Vecchia Gaussian process, a static ensemble Kalman filter, XGBoost, and a graph neural network under the stated interpolation designs. Matched field experiments with INLA and Vecchia-Laplace remain future work and would require identical general knowledge, soft information, prediction folds, and interval definitions. Within the implemented comparisons, the experiments quantify agreement with traditional BME, approximation error, and computational savings. Comparative claims are limited to these methods and data designs.

The current implementation estimates the mean and covariance parameters before sparse inference and treats them as fixed during the GEP and MEP updates. The

829 reported posterior intervals therefore omit uncertainty in these parameters. Expec-
 830 tation propagation has no general convergence guarantee; damping, the change in
 831 site parameters, moment residuals, and active precision bounds identify convergence
 832 in each calculation (Minka 2001; Cseke and Heskes 2011). Every GEP iteration also
 833 requires marginal variances at the non-Gaussian soft locations, obtained from selected
 834 solves using the sparse factorization. This cost increases with the number of such loca-
 835 tions. MEP adds a product of component choices across switched locations, so $C_{\max}$
 836 bounds both computation and the probability mass included in the mixture. These
 837 diagnostics state the implemented scope without attributing the observed interval
 838 undercoverage to a single omitted source of uncertainty.

839 **3.13 Future development of general knowledge models**

840 Future work will extend the general knowledge model to spatial processes outside the
 841 Gaussian family. A Student- t process can introduce joint behaviour with heavy tails
 842 while retaining conditional prediction from a covariance kernel (Shah et al. 2014).
 843 Copula processes offer another route by separating marginal distributions from spatial
 844 dependence (Wilson and Ghahramani 2010). Finite mixtures of sparse graphical fields
 845 could represent several hydrogeologic regimes; posterior component weights and diag-
 846 nostics for omitted probability mass would extend the mixture calculations developed
 847 here. Joint likelihood factors or latent error variables will be needed when soft infor-
 848 mation sources share conversion or measurement errors. A categorical BME extension
 849 could connect sparse graphical inference with probability tables from the maximum
 850 entropy formulation. Each extension requires a reference calculation in low dimension,
 851 a check using all predecessors, analysis of errors from finite predecessor sets, calibration
 852 and tail diagnostics, and separate measurements of time and memory. Vecchia–Laplace
 853 methods remain a relevant comparison for latent Gaussian models with non-Gaussian

854 observations (Zilber and Katzfuss 2021); they address the likelihood calculation, while
855 the general knowledge model remains Gaussian.

856 4 Conclusions

857 VISTA-BME combines Vecchia graphical inference with explicit soft likelihoods in a
858 scalable BME calculation. Complete predecessor conditioning reproduced the matched
859 traditional BME posterior. The approximation from finite predecessor sets decreased
860 as the number of predecessors increased. Gaussian expectation propagation accurately
861 represented the tested bounded, censored, and skewed likelihoods, including likelihoods
862 with heavy tails. Complete mixture propagation preserved distinct modes in the tested
863 cases with two to four simultaneous bimodal likelihoods. The implemented automatic
864 rule used GEP at locations where multimodality did not produce a variance residual
865 above 10^{-2} and selected MEP for the higher severities. It reproduced every reference
866 mode count in the 54 cases used to validate automatic selection. The component limit
867 quantified the associated balance between density fidelity and computational cost.

868 The field applications separated computational gain from predictive calibration.
869 With 25 soft likelihoods in the Biscayne comparison, VISTA-BME reduced measured
870 wall time by a factor of 19.98 and peak memory by a factor of 2391 relative to BME;
871 the posterior mean difference was 0.0117 on the logarithmic chloride scale. At 8,000
872 Gaussian observations, VISTA-BME was 2.49 times faster and used 25.1 times less
873 peak memory. The USGS calculation produced posterior means for 123,048 combina-
874 tions of wells and months and 55,728 combinations of grid locations and months. Point
875 accuracy was comparable among the three Gaussian spatial methods. The 2025 test
876 showed that interval calibration requires a separate assessment and correction. VISTA-
877 BME thus provides a route for BME assimilation of explicit probability distributions
878 when complete covariance storage becomes prohibitive, together with diagnostics for
879 sparse error, likelihood error, calibration, time, and memory.

References

- Angelopoulos, A.N., Bates, S.: Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning* **16**(4), 494–591 (2023) <https://doi.org/10.1561/22000000101>
- Alspach, D.L., Sorenson, H.W.: Nonlinear bayesian estimation using gaussian sum approximations. *IEEE Transactions on Automatic Control* **17**(4), 439–448 (1972) <https://doi.org/10.1109/TAC.1972.1100034>
- Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H.: Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B* **70**(4), 825–848 (2008) <https://doi.org/10.1111/j.1467-9868.2008.00663.x>
- Bui, T.D., Nguyen, C.V., Turner, R.E.: Streaming sparse gaussian process approximations. In: *Advances in Neural Information Processing Systems*, vol. 30 (2017)
- Bai, T., Tahmasebi, P.: Graph neural network for groundwater level forecasting. *Journal of Hydrology* **616**, 128792 (2023) <https://doi.org/10.1016/j.jhydrol.2022.128792>
- Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794 (2016). <https://doi.org/10.1145/2939672.2939785>
- Cao, J., Guinness, J., Genton, M.G., Katzfuss, M.: Scalable gaussian-process regression and variable selection using vecchia approximations. *Journal of Machine Learning Research* **23**(348), 1–30 (2022)
- Cseke, B., Heskes, T.: Approximate marginals in latent gaussian models. *Journal of Machine Learning Research* **12**(13), 417–454 (2011)

- 903 Christakos, G.: A bayesian/maximum-entropy view to the spatial estimation problem.
 904 Mathematical Geology **22**, 763–777 (1990) <https://doi.org/10.1007/BF00890661>
- 905 Christakos, G.: Modern Spatiotemporal Geostatistics. Oxford University Press, New
 906 York (2000)
- 907 Datta, A., Banerjee, S., Finley, A.O., Hamm, N.A.S., Schaap, M.: Nonseparable
 908 dynamic nearest neighbor gaussian process models for large spatiotemporal data
 909 with an application to particulate matter analysis. The Annals of Applied Statistics
 910 **10**(3), 1286–1316 (2016) <https://doi.org/10.1214/16-AOAS931>
- 911 Datta, A., Banerjee, S., Finley, A.O., Gelfand, A.E.: Hierarchical nearest-neighbor
 912 gaussian process models for large geostatistical datasets. Journal of the American
 913 Statistical Association **111**(514), 800–812 (2016) [https://doi.org/10.1080/016214](https://doi.org/10.1080/01621459.2015.1044091)
 914 [59.2015.1044091](https://doi.org/10.1080/01621459.2015.1044091)
- 915 Efron, B.: Bootstrap methods: Another look at the jackknife. The Annals of Statistics
 916 **7**(1), 1–26 (1979) <https://doi.org/10.1214/aos/1176344552>
- 917 Evensen, G.: The ensemble kalman filter: Theoretical formulation and practical imple-
 918 mentation. Ocean Dynamics **53**, 343–367 (2003) [https://doi.org/10.1007/s10236-0](https://doi.org/10.1007/s10236-003-0036-9)
 919 [03-0036-9](https://doi.org/10.1007/s10236-003-0036-9)
- 920 Gneiting, T., Balabdaoui, F., Raftery, A.E.: Probabilistic forecasts, calibration and
 921 sharpness. Journal of the Royal Statistical Society: Series B **69**, 243–268 (2007)
 922 <https://doi.org/10.1111/j.1467-9868.2007.00587.x>
- 923 Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation.
 924 Journal of the American Statistical Association **102**, 359–378 (2007) [https://doi.or](https://doi.org/10.1198/016214506000001437)
 925 [g/10.1198/016214506000001437](https://doi.org/10.1198/016214506000001437)

- 926 Guinness, J.: Permutation and grouping methods for sharpening gaussian process
 927 approximations. *Technometrics* **60**(4), 415–429 (2018) [https://doi.org/10.1080/00](https://doi.org/10.1080/00401706.2018.1437476)
 928 401706.2018.1437476
- 929 Golub, G.H., Welsch, J.H.: Calculation of gauss quadrature rules. *Mathematics of*
 930 *Computation* **23**(106), 221–230 (1969) [https://doi.org/10.1090/S0025-5718-69-996](https://doi.org/10.1090/S0025-5718-69-99647-1)
 931 47-1
- 932 Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., Ger-
 933 ber, F., Gramacy, R.B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D.W.,
 934 Sun, F., Zammit-Mangion, A.: A case study competition among methods for ana-
 935 lyzing large spatial data. *Journal of Agricultural, Biological and Environmental*
 936 *Statistics* **24**, 398–425 (2019) <https://doi.org/10.1007/s13253-018-00348-w>
- 937 Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble
 938 prediction systems. *Weather and Forecasting* **15**(5), 559–570 (2000) [https://doi.or](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2)
 939 [g/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2)
- 940 Hendricks Franssen, H.J., Kinzelbach, W.: Real-time groundwater flow modeling with
 941 the ensemble kalman filter: Joint estimation of states and parameters and the filter
 942 inbreeding problem. *Water Resources Research* **44**(9), 09408 (2008) [https://doi.or](https://doi.org/10.1029/2007WR006505)
 943 [g/10.1029/2007WR006505](https://doi.org/10.1029/2007WR006505)
- 944 Hestenes, M.R., Stiefel, E.: Methods of conjugate gradients for solving linear systems.
 945 *Journal of Research of the National Bureau of Standards* **49**(6), 409–436 (1952)
 946 <https://doi.org/10.6028/jres.049.044>
- 947 Hutchinson, M.F.: A stochastic estimator of the trace of the influence matrix for
 948 laplacian smoothing splines. *Communications in Statistics—Simulation and Com-*
 949 *putation* **19**(2), 433–450 (1990) [https://doi.org/10.1080/0361091900881286](https://doi.org/10.1080/03610919008812866)
 950 6

- 951 Katzfuss, M.: A multi-resolution approximation for massive spatial datasets. Journal
952 of the American Statistical Association **112**(517), 201–214 (2017) [https://doi.org/](https://doi.org/10.1080/01621459.2015.1123632)
953 [10.1080/01621459.2015.1123632](https://doi.org/10.1080/01621459.2015.1123632)
- 954 Kolovos, A., Christakos, G., Serre, M.L., Miller, C.T.: Computational Bayesian max-
955 imum entropy solution of a stochastic advection-reaction equation in the light of
956 site-specific information. Water Resources Research **38**(12), 1318 (2002) [https:](https://doi.org/10.1029/2001WR000743)
957 [//doi.org/10.1029/2001WR000743](https://doi.org/10.1029/2001WR000743)
- 958 Katzfuss, M., Guinness, J.: A general framework for vecchia approximations of gaus-
959 sian processes. Statistical Science **36**, 124–141 (2021) <https://doi.org/10.1214/19-S>
960 [TS755](https://doi.org/10.1214/19-S)
- 961 Katzfuss, M., Guinness, J., Gong, W., Zilber, D.: Vecchia approximations of gaussian-
962 process predictions. Journal of Agricultural, Biological and Environmental Statistics
963 **25**, 383–414 (2020) <https://doi.org/10.1007/s13253-020-00401-7>
- 964 Kuss, M., Rasmussen, C.E.: Assessing approximate inference for binary gaussian
965 process classification. Journal of Machine Learning Research **6**(57), 1679–1704
966 (2005)
- 967 LoBuglio, J.N., Characklis, G.W., Serre, M.L.: Cost-effective water quality assessment
968 through the integration of monitoring data and modeling results. Water Resources
969 Research **43**(3), 03435 (2007) <https://doi.org/10.1029/2006WR005020>
- 970 Lin, J.: Divergence measures based on the shannon entropy. IEEE Transactions on
971 Information Theory **37**(1), 145–151 (1991) <https://doi.org/10.1109/18.61115>
- 972 Lindgren, F., Rue, H.v., Lindström, J.: An explicit link between gaussian fields
973 and gaussian markov random fields: The stochastic partial differential equation
974 approach. Journal of the Royal Statistical Society: Series B **73**(4), 423–498 (2011)

975 <https://doi.org/10.1111/j.1467-9868.2011.00777.x>

976 Minka, T.P.: Expectation propagation for approximate bayesian inference. In: Pro-
977 ceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, pp.
978 362–369 (2001)

979 National Aeronautics and Space Administration: POWER Data Access Viewer and
980 Application Programming Interface. Accessed 18 August 2026 (2026). [https://po-](https://power.larc.nasa.gov/docs/services/api/)
981 [wer.larc.nasa.gov/docs/services/api/](https://power.larc.nasa.gov/docs/services/api/)

982 National Water Quality Monitoring Council: Water Quality Portal. Accessed 18
983 August 2026 (2026). <https://doi.org/10.5066/P9QRKUVJ> . [https://www.waterq-](https://www.waterqualitydata.us/beta/portal_userguide/)
984 [ualitydata.us/beta/portal_userguide/](https://www.waterqualitydata.us/beta/portal_userguide/)

985 Nickisch, H., Rasmussen, C.E.: Approximations for binary gaussian process classifica-
986 tion. *Journal of Machine Learning Research* **9**, 2035–2078 (2008)

987 Peyré, G., Cuturi, M.: Computational optimal transport. *Foundations and Trends in*
988 *Machine Learning* **11**(5–6), 355–607 (2019) <https://doi.org/10.1561/22000000073>

989 Rajaei, T., Ebrahimi, H., Nourani, V.: A review of the artificial intelligence methods
990 in groundwater level modeling. *Journal of Hydrology* **572**, 336–351 (2019) [https:](https://doi.org/10.1016/j.jhydrol.2018.12.037)
991 [//doi.org/10.1016/j.jhydrol.2018.12.037](https://doi.org/10.1016/j.jhydrol.2018.12.037)

992 Rue, H.v., Martino, S., Chopin, N.: Approximate bayesian inference for latent gaussian
993 models by using integrated nested laplace approximations. *Journal of the Royal*
994 *Statistical Society: Series B* **71**(2), 319–392 (2009) [https://doi.org/10.1111/j.1467-](https://doi.org/10.1111/j.1467-9868.2008.00700.x)
995 [-9868.2008.00700.x](https://doi.org/10.1111/j.1467-9868.2008.00700.x)

996 Romano, Y., Patterson, E., Candès, E.J.: Conformalized quantile regression. In:
997 *Advances in Neural Information Processing Systems*, vol. 32 (2019)

- 998 Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT
999 Press, Cambridge, Massachusetts (2006). <https://gaussianprocess.org/gpml/>
- 1000 Serre, M.L., Christakos, G.: Modern geostatistics: Computational BME analysis in the
1001 light of uncertain physical knowledge—the Equus Beds study. *Stochastic Environ-*
1002 *mental Research and Risk Assessment* **13**(1–2), 1–26 (1999) [https://doi.org/10.100](https://doi.org/10.1007/s004770050029)
1003 [7/s004770050029](https://doi.org/10.1007/s004770050029)
- 1004 Scott, D.W.: *Multivariate Density Estimation: Theory, Practice, and Visualization*.
1005 John Wiley & Sons, New York (1992). <https://doi.org/10.1002/9780470316849>
- 1006 Stein, M.L., Chi, Z., Welty, L.J.: Approximating likelihoods for large spatial data
1007 sets. *Journal of the Royal Statistical Society: Series B* **66**(2), 275–296 (2004) [https:](https://doi.org/10.1046/j.1369-7412.2003.05512.x)
1008 [//doi.org/10.1046/j.1369-7412.2003.05512.x](https://doi.org/10.1046/j.1369-7412.2003.05512.x)
- 1009 Shah, A., Wilson, A.G., Ghahramani, Z.: Student-t processes as alternatives to gaus-
1010 sian processes. In: Kaski, S., Corander, J. (eds.) *Proceedings of the Seventeenth*
1011 *International Conference on Artificial Intelligence and Statistics. Proceedings of*
1012 *Machine Learning Research*, vol. 33, pp. 877–885. PMLR, Reykjavik, Iceland (2014).
1013 <https://proceedings.mlr.press/v33/shah14.html>
- 1014 Titsias, M.: Variational learning of inducing variables in sparse gaussian processes. In:
1015 *Proceedings of the Twelfth International Conference on Artificial Intelligence and*
1016 *Statistics. Proceedings of Machine Learning Research*, vol. 5, pp. 567–574 (2009).
1017 <https://proceedings.mlr.press/v5/titsias09a.html>
- 1018 U.S. Geological Survey: Estimated Groundwater Withdrawals from Principal Aquifers
1019 in the United States: County-Level Data for 2015. Accessed 18 August 2026 (2020).
1020 <https://doi.org/10.5066/P9EI0KMR> . [https://www.usgs.gov/data/estimated-gro](https://www.usgs.gov/data/estimated-groundwater-withdrawals-principal-aquifers-united-states-county-level-data-2015)
1021 [undwater-withdrawals-principal-aquifers-united-states-county-level-data-2015](https://www.usgs.gov/data/estimated-groundwater-withdrawals-principal-aquifers-united-states-county-level-data-2015)

- 1022 U.S. Geological Survey: USGS Water Data APIs: OGC APIs. Accessed 17 August
1023 2026 (2026). <https://api.waterdata.usgs.gov/docs/ogcapi/>
- 1024 U.S. Geological Survey: Water Services Site Service: Site Service Details. Accessed 18
1025 August 2026 (2026). [https://waterservices.usgs.gov/docs/site-service/site-service-](https://waterservices.usgs.gov/docs/site-service/site-service-details/)
1026 [details/](https://waterservices.usgs.gov/docs/site-service/site-service-details/)
- 1027 Vecchia, A.V.: Estimation and model identification for continuous spatial processes.
1028 Journal of the Royal Statistical Society: Series B **50**(2), 297–312 (1988) [https:](https://doi.org/10.1111/j.2517-6161.1988.tb01729.x)
1029 [//doi.org/10.1111/j.2517-6161.1988.tb01729.x](https://doi.org/10.1111/j.2517-6161.1988.tb01729.x)
- 1030 Wilson, A.G., Ghahramani, Z.: Copula processes. In: Advances in Neural Information
1031 Processing Systems 23, pp. 2460–2468 (2010)
- 1032 Wunsch, A., Liesch, T., Broda, S.: Groundwater level forecasting with artificial neural
1033 networks: A comparison of long short-term memory, convolutional neural networks,
1034 and nonlinear autoregressive networks with exogenous input. Hydrology and Earth
1035 System Sciences **25**(3), 1671–1687 (2021) [https://doi.org/10.5194/hess-25-1671-2](https://doi.org/10.5194/hess-25-1671-2021)
1036 [021](https://doi.org/10.5194/hess-25-1671-2021)
- 1037 Yu, H.-L., Ku, S.-C., Lee, C.-H., Tseng, H.-T., Lee, S.-Y.: STAMPS: Spatial and
1038 Temporal Analysis and Mapping Python Suite. Software repository. Accessed 31
1039 August 2026 (2018). <https://github.com/NTU-STEMPLab/stamps>
- 1040 Yu, H.-L., Wu, Y.-Z., Cheung, S.Y.: A data assimilation approach for groundwa-
1041 ter parameter estimation under bayesian maximum entropy framework. Stochastic
1042 Environmental Research and Risk Assessment **34**, 709–721 (2020) [https://doi.org/](https://doi.org/10.1007/s00477-020-01795-z)
1043 [10.1007/s00477-020-01795-z](https://doi.org/10.1007/s00477-020-01795-z)
- 1044 Zilber, D., Katzfuss, M.: Vecchia–laplace approximations of generalized gaussian pro-
1045 cesses for big non-gaussian spatial data. Computational Statistics & Data Analysis

1046 **153**, 107081 (2021) <https://doi.org/10.1016/j.csda.2020.107081>

1047 Zhang, J.Z., Renshaw, C.K.: Shapefile and summary tables for the approximate inland
 1048 extent of saltwater intrusion at the base of the Biscayne aquifer in 2022, Miami-
 1049 Dade County, Florida. U.S. Geological Survey (2024). [https://doi.org/10.5066/P1](https://doi.org/10.5066/P13TSEEA)
 1050 3TSEEA . [https://www.usgs.gov/data/shapefile-and-summary-tables-approxima](https://www.usgs.gov/data/shapefile-and-summary-tables-approximate-inland-extent-saltwater-intrusion-base-biscayne)
 1051 [te-inland-extent-saltwater-intrusion-base-biscayne](https://www.usgs.gov/data/shapefile-and-summary-tables-approximate-inland-extent-saltwater-intrusion-base-biscayne)

1052 Zhang, J.Z., Renshaw, C.K.: Approximate inland extent of saltwater intrusion at the
 1053 base of the biscayne aquifer, miami-dade county, florida, 2022. Scientific Investiga-
 1054 tions Map 3541, U.S. Geological Survey (2025). <https://doi.org/10.3133/sim3541> .
 1055 <https://pubs.usgs.gov/publication/sim3541/full>

1056 **Statements and Declarations**

1057 **Funding**

1058 This study was supported by the National Science and Technology Council (NSTC),
 1059 Taiwan, through Grant Nos. NSTC 114-2221-E-002-014-MY3, NSTC 114-2119-M-002-
 1060 029, and NSTC 115-2221-E-033-018-MY3.

1061 **Competing interests**

1062 The authors have no relevant financial or non-financial interests to disclose.

1063 **Author contributions**

1064 Ying-Fan Lin developed the software, designed and conducted the analyses, prepared
 1065 the figures, and wrote the initial manuscript. Hwa-Lung Yu contributed to the BME
 1066 formulation, methodological interpretation, and manuscript revision. Both authors
 1067 reviewed and approved the manuscript.

1068 **Data availability**

1069 Online Resource 1 contains the 72-frame groundwater anomaly animation. Online
1070 Resource 2 is the manuscript reproducibility bundle and contains the analysis scripts,
1071 derived result tables, figure scripts, caption sources, QA reports, tests, and a com-
1072 plete artifact inventory. It includes posterior tables for every well and month and
1073 every regional grid location and month, representative well series, variance validation
1074 results, Biscayne predictions, validation with soft data at several sites, validation over
1075 a terminal period, and computational measurements. The original chloride tables, con-
1076 ductance summaries, electromagnetic logs, shapefile, and mapped isochlor are available
1077 from the USGS data release and map (Zhang and Renshaw 2024, 2025). The large raw
1078 USGS acquisitions remain outside the bundle and can be obtained again from the cited
1079 official services. Acquisition and analysis scripts specify parameter codes, geographic
1080 bounds, approval filters, pagination, unit conversion, fold assignments, random seeds,
1081 and numerical settings.

1082 **Code availability**

1083 VISTA-BME version 0.2.0, its tests, installation instructions, and Codex skill are
1084 publicly available at <https://github.com/aar246860/vista-bme>. The archived software
1085 revision is identified by commit `9c47ae4`. The BME reference calculations reported
1086 here were performed with the STAMPS package (Yu et al. 2018). The release provides
1087 continuous BME and sparse graphical operators, typed soft likelihoods, diagnostics for
1088 expectation propagation and Gaussian mixture conversion, automatic GEP or MEP
1089 selection, predictions that preserve mixture components, and updates at fixed nodes
1090 from sufficient statistics. Users can import the package through `vista_bme`. Online
1091 Resource 2 supplies the manuscript analysis scripts and derived results. The interactive
1092 demonstration is available at <https://vista-bme-live.yflin.chatgpt.site>.

1093 Supplementary Information

1094 Online Resource 1 (`ESM_1.mp4`) is an animation with 72 frames of monthly groundwater
 1095 level anomalies from January 2020 through December 2025. Each frame shows the
 1096 posterior anomaly relative to the 2020–2025 median at the same well and represents
 1097 a retrospective covariance reconstruction.

1098 Online Resource 2 (`ESM_2.zip`) is the manuscript reproducibility bundle. It con-
 1099 tains the analysis scripts, derived result tables, figure code, caption sources, QA
 1100 records, tests, and artifact inventory needed to reproduce the reported calculations
 1101 from the documented source acquisitions.

1102 Notation

1103 The symbols used in the methodology are summarized below. Each symbol is also
 1104 defined where it first enters the calculation.

Table 10: Notation used in the BME and VISTA-BME calculations.

Symbol	Definition
$\mathbf{x}_k, \mathbf{x}_h, \mathbf{x}_s$	Attribute values at prediction locations, hard data whose values are treated as certain, and uncertain attribute values represented by soft data.
$\mathbf{p}_k, \mathbf{p}_h, \mathbf{p}_s$	Space and time locations of the prediction values, hard data, and soft data; their counts are n_k , n_h , and n_s .
$\mathbf{x}_{hsk}, \mathbf{x}_{ks}$	Stacked vectors $[\mathbf{x}_h^\top, \mathbf{x}_s^\top, \mathbf{x}_k^\top]^\top$ and $[\mathbf{x}_s^\top, \mathbf{x}_k^\top]^\top$.
f_G, f_S, f_K	Densities representing the general knowledge base, the combined soft information in the specific knowledge base, and the BME posterior.
$f_S(x_{s,j})$	Probability density or proportional observation likelihood that supplies soft information at location j ; the implemented specific knowledge model uses $f_S(\mathbf{x}_s) = \prod_j f_S(x_{s,j})$.
$\mathbf{m}_{hsk}, \mathbf{m}_\phi, \mathbf{U}, \phi$	Prior mean of the complete random vector, mean function, covariates, and mean function parameters.

Symbol	Definition
Σ, θ, Λ	Covariance matrix of the general knowledge model, covariance parameters, and precision used in the calculation. The computational Λ equals $\tilde{\Lambda}$ for finite predecessor sets and Σ^{-1} when all predecessors are included.
$g(i), m$	Predecessor set for location i and its maximum size.
$m_{i g(i)}, \Sigma_{i g(i)}$	Conditional mean and innovation variance of x_i given the predecessor values.
$L, D, \tilde{\Lambda}$	Unit lower triangular conditional factor, diagonal matrix of innovation variances, and sparse Vecchia precision $\tilde{\Lambda} = L^T D^{-1} L$.
$\Lambda_{ks h}, \eta_{ks h}, m_{ks h}$	Precision, linear information parameter, and mean of x_{ks} conditional on the hard data.
$I_{s,ks}, I_{k,ks}$	Binary matrices that extract the soft and prediction variables from x_{ks} .
$\tilde{m}_{s,j}, \tilde{\sigma}_{s,j}^2, \tilde{m}_s, \tilde{\Sigma}_s$	Mean and variance of a Gaussian soft datum, and their vector and diagonal covariance matrix.
$\tau_j^{(t)}, \nu_j^{(t)}, \tilde{m}_{s,j}^{(t)}, \tilde{\sigma}_{s,j}^{2(t)}$	Site precision, linear natural parameter, and the corresponding Gaussian mean and variance at soft location j during GEP iteration t . The moment form applies when $\tau_j^{(t)} > 0$.
$\sigma_{\text{soft},j}^2$	Variance scale calculated from the original soft distribution and used to set the upper numerical bound on site precision.
α	Damping coefficient used in the GEP update.
$\Lambda_{ks h,s}^{\text{BME}}, m_{ks h,s}^{\text{BME}}, \Sigma_{ks h,s}^{\text{BME}}$	BME posterior precision, mean, and covariance of x_{ks} after all hard and soft data are used.
$\Lambda_{ks h,s}^{\text{BME},(t)}$	BME posterior precision during GEP iteration t .
$m_{ks h,s}^{\text{BME},(t)}, \Sigma_{ks h,s}^{\text{BME},(t)}$	BME posterior mean and covariance during GEP iteration t .
$q_j^{(t)}, \mu_{s,j}^{(t)}, v_{s,j}^{(t)}$	Marginal Gaussian posterior at soft location j during iteration t , with its mean and variance.
$q_j^{\setminus j,(t)}, a_j^{(t)}, b_j^{(t)}$	Cavity Gaussian density obtained by removing the site factor at location j , with its mean and variance.
$f_j^{*(t)}, \bar{x}_{s,j}^{*(t)}, V_{s,j}^{*(t)}$	Tilted density obtained by combining the original soft information at location j with its cavity density, with the tilted mean and variance.
$M_j, r_{V,j}, s_j$	Number of resolved modes, relative GEP variance residual, and indicator that assigns soft location j to MEP.
$\mathcal{I}_j, Z_j^{\text{tr}}, \epsilon_j^{\text{tail}}, f_{S,j}^{\text{tr}}$	Finite conversion interval, probability within that interval, omitted tail probability, and density normalized over the interval.

Symbol	Definition
$C_j, w_{j,c}, \tilde{m}_{j,c}, \tilde{\sigma}_{j,c}^2$	Number of Gaussian components used for soft distribution j and the weight, mean, and variance of component c .
$\mathbf{c}, \mathcal{A}, \mathcal{C}, \pi_{\mathbf{c}}, Z_{\mathbf{c}}$	One combination of mixture components, the full Cartesian product, the propagated subset, its prior weight, and its normalization under the general knowledge model.
$w_{\mathbf{c}}^{\text{BME}}, C_{\max}, \delta_{\mathbf{c}}$	Posterior weight of a component combination, maximum number of propagated combinations, and difference between its prediction mean and the mixture mean.
$N_i, \mathcal{K}_i, \tilde{\sigma}_{s,i}^{-2(N_i)}, \tilde{m}_{s,i}^{(N_i)}$	Number and index set of observations received at location i , accumulated precision, and accumulated Gaussian mean.
$\Lambda_{\text{ret}}^{\text{BME}}, R, \mathbf{z}_r, \hat{\mathbf{d}}$	Retrospective BME posterior precision, number of Rademacher probes, probe vector, and estimated diagonal of the posterior covariance matrix.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1.mp4](#)
- [ESM2.zip](#)