https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

A probabilistic view of extreme sea level events in the Baltic Sea

Seuri Basilio Kuosmanen[1] and Magnus Hieronymus[2]
[1,2]SMHI, Folkborgsvägen 17, 601 76 Norrköping, Sweden
Correspondence: Seuri Basilio Kuosmanen (Seuri.basilio@smhi.se)

Abstract. The importance of accounting for extreme sea level events is often an integrate part in any risk mitigation strategies that aims to protect present and future coastal infrastructure. The Extreme value theory (EVT) gives a probabilistic framework for studying such events. However, the conventional methods used in the application of EVT are often restrictive, since they are generally confined to location where there are sufficiently long tide gauge observation data, while simultaneously fail to obtain good estimates of lower probability events.

In this article, we use the Bayesian hierarchical modeling paradigm and the Block Maxima method in the EVT, to estimate the extreme sea level event that occur on average ones every e.g. 1000 years. Four novel models are presented in this study, and each of the models incorporates both missing values and spatial dependency structures to obtain estimates of such extreme events, with a varying complex dependency structure. In addition, two of the models (Hilbert and Latent) allow for the estimation of extreme sea level events at both gauged and ungauged locations.

The results of this study show that Hilbert and Latent obtain good estimates with a reduced uncertainty range for both higher and lower probability events. From the in and out-of-sample evaluation, it follows that the two model's outperformance the conventional method of combining maximum likelihood estimates and bootstrapping, when comparing the uncertainty range, for the estimates of extreme sea level events that occurs on average ones every e.g. 100 years and up to 100000 years.

1 Introduction

Between 1901 and 2018 the globally averaged mean sea level rose by 20 cm [15,25], and the rise is accelerating (Frederikse et al. 2020)

Moreover, projections for the current century show much higher rates of mean sea level rise than for the last century, even under very low emission scenarios (Kikstra et al. 2022). Mean sea level rise exacerbates the risk of coastal flooding, as it increases the baseline from which intermittent weather driven sea level extremes, such as storm surges, act. Coastal flooding can thus be thought of as a two component problem consisting of a slow varying mean sea level rise component and a fast varying, high amplitude short duration, extreme sea level component. Our focus here is on the latter component.

The impacts of extreme sea levels along coastlines are given mean sea level rise of growing concern, leading to, among other things, the need for improved risk mitigation strategies in coastal planning at different levels of government and time scales. The probabilistic understanding of extreme sea levels is a crucial part of any coastal planning aiming to account for the risk of such extreme events. Such undertanding is usually gathered by studying time series, with high temporal resolution,

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

from tide gauges. However, one common problem is that there are no tide gauge stations at or near the location of interest. Moreover, the number of observation years varies between tide gauge stations. Resulting in the problem of the sparsity of observable locations, respectively, and the non-uniformness of observation years between sites. Our assumption, which builds on the works of F.M.Calafat, and M.Marcos in (Calafat and Marcos 2020), and the works of O.Räty, M.Laine, and U.Leijala, J.Särkkä, and M.M.Johansson in (Räty et al. 2023), is that there is much to gain by exploiting the spatial dependency of extreme sea level rise in the statistical modeling of such extremes. In addition to levering dependency, we ask what the observed annual sea level maxima at a given location tells us about the corresponding unobserved sea level maxima for locations at varying distances.

Our study area encompasses the Baltic Sea and parts of the North Sea, see Fig.3.1.3. Both are relatively shallow shelf seas. The Baltic Sea is brackish and semi-enclosed, being connected to the open ocean by the narrow and shallow Danish Straits that act as a physical filter for high frequency sea level variability (M. Hieronymus, J. Hieronymus, and Arneborg 2017). The North Sea is much more open and thus more exposed to storm surges coming from the North Atlantic. Sea level extremes in the Baltic Sea are primarily wind driven, while those in the North Sea may also have a strong tidal component (M. Hieronymus, J. Hieronymus, and Arneborg 2017).

We utilize methods from Extreme value theory (EVT), Bayesian statistics, and Probabilistic programming to study and drive results for the above discussion. More specifically, we gain probabilistic insights about the extreme sea levels in the Baltic Sea by applying the Block Maxima approach and the Generalized Extreme Value distribution (GEV) in the EVT, for varying collections of tide gauge stations, all of which are available through GESLA (Haigh et al. 2023), together with the Markov Chain Monte Carlo simulations, respectively, the Bayesian hierarchical modeling paradigm, which allows for the smooth modeling of the GEV parameters using Latent Gaussian Process and Hilbert space approximated Gaussian Process, resulting in tight uncertainty bounds for the parameters, the theoretical support of the distribution, respectively, the r-year return levels, and thus decreased uncertainty for estimations at ungauged locations.

The article is structured as follows; Sect.2 outlines the general methodology used to derive the results. The data, and models used to derive the results are presented in Sect.3. The results are found in Sect.4, followed by conclusions and discussions in Sect.5.

2 Methods

In this section, we recall the methods used to derive the results in Sect.4. We first introduce some relevant notation and general results from the Extreme Value theory, and Bayesian Statistics, see (Coles et al. 2001), respectively, (Banerjee, Carlin, and Gelfand 2003) for a detailed presentation of the topics.

Recall that we are interested in estimating return levels (defined below in Eq.(3)) at any arbitrary location contained in the region of interest (the Baltic Sea and parts of the North Sea). In particular, we are interested in reducing the uncertainty for

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

the estimated return levels, where uncertainty is quantified using the $95\%-$ highest probability density HPD set (see Sect.2.1), which is sometimes referred to as the $95\%-$ credibility interval.

2.1 Extreme value theory and bayesian statistical modeling

2.1.1 The Block maxima method & extreme value theory

Let $S := \{s_i : i = 1,2,\ldots,n\}$ denote any finite collection of tide gauge stations (points) contained in the region of interest, where $s_i = i$ for every $s_i \in S$, or $s_i = (x_i, y_i)$ for every $s_i \in S$, where $(x,y) = (\text{lat}, \text{long})$ is the latitude-longitude coordinates of tide gauge stations $s_i \in S$.

Let $\zeta(s) := \{\zeta^{\alpha}(s) : \alpha \in I\}$ denote the normal (weak-)stationary sequences corresponding to the observed sea level at the tide gauge stations s, for any $s \in S$, with time interval I. Then, for any $s \in S$, we have that $z^k(s) := \max\limits_{\alpha_k \in I_k} \{\zeta^{\alpha_k}(s)\}$ is the random variable corresponding to the observed sea level maxima at s over the k^{th} block (time period), where $I(K) := \{I_k\}_{k=0}^{K}$ is the parition of I in to K blocks, with cardinality $\text{card}(I_k) = \text{card}(I'_k)$ for any $k, k', K \in \mathbb{N}$.

For every $s \in S$, assume that $z(s) := \{z^k(s)\}_{k=0}^{K}$ is the collection of K block maxima (annual sea level maxima) at the station s, for any $k, K \in \mathbb{N}$. Furthermore, suppose that $z(s)$ satisfies the $D(u_n)$ condition; see (Leadbetter, Lindgren, and Rootzén 2012).

Recall that given sufficently large k, if the intervals $I_k \in I(K)$ are defined such that the indices α_k are contained in the interior of the interval I_k, where α_k is the index of $z^k(s)$. Then $z(s)$ satisfies the $D(u_n)$ condition as $K \to \infty$; see Beirlant et al. 2006.

Given the above recollection, it suffices to define $I(K)$ as the collection of K shifted calendar years, due to the seasonality in occurrence of extreme sea level events over the region of region of interest; see (M. Hieronymus 2022).

Similarly, for every $k \in \mathbb{N}$, let $z^k(\mathbf{s}) := (z^k(s_1), z^k(s_2), \ldots, z^k(s_n)) \in \mathbb{R}^n$ denote the k^{th} block maxima component (annual sea level maxima component) over $\mathbf{s} := (s_1, s_2, \ldots, s_n)$, with $s_i \in S$, for all $i = 1,2,\ldots,n$. In particular, for any $K \in \mathbb{N}$, we denote by $z(\mathbf{s}) := \{z^k(\mathbf{s})\}_{k=0}^{K}$ the collection of K block maxima's components over $\mathbf{s}$. We call $z(\mathbf{s})$ the $(K \times n)-$dim Block maxima matrix, or Block maxima matrix.

Observe that for any $z^k(\mathbf{s})$, then $z^k(s_i)$ need not be independent of $z^k(s_j)$, whenever $s_i \neq s_j$, for any $s_i, s_j \in S$, and for every $k \in \mathbb{N}$. In particular, additional arguments or assumptions are needed for mutual independence between entries of $z^k(\mathbf{s})$.

From the Extreme Value theory, it follows that for any $i = 0,1,\ldots,n$, the entry $z(s_i)$ of $z(\mathbf{s})$ tends to a univariate GEV, as K tends to infinity, see (Coles et al. 2001). Therefore, given sufficiently large K, we may assume that each $z(s_i)$ of $z(\mathbf{s})$ is the realization of a univariate GEV. Hence, the random variable of interest $Z(s)$ follows a univariate GEV, for any $s \in S$, where $Z(s)$ corresponds to the annual sea level maxima at the station s. Therefore, we have that

$$Z(s) \sim \text{GEV}(\Lambda(s)) \quad \text{where } \Lambda(s) = (\mu(s), \sigma(s), \xi(s)), \tag{1}$$

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

with $\mu(s) \in \mathbb{R}, \sigma(s) \in \mathbb{R}_{>0}, \xi(s) \in \mathbb{R}$ denoting the location, scale, respectively, shape parameter of the distribution at the location $s \in S$.

Note that for any $s_i, s_j \in S$, with $s_i \neq s_j$, then $Z(s_i)$ is not necessarily independent of $Z(s_j)$, by a previous argument.

Now, recall that the cumulative distribution function (cdf), and the $r-$ return level (the $(1-p)-$quantile function) of the GEV are given by

$$G(x : \Lambda(s)) := \begin{cases} \exp\Big(-\Big(1+\xi(s)\big(\frac{x-\mu(s)}{\sigma(s)}\big)\Big)^{-1/\xi(s)}\Big), & \text{if } \xi(s) \neq 0, \text{ with } x \in \{x : 1+\xi(s)(x-\mu(s))/\sigma(s) > 0\} \\ \exp\Big(-\exp\Big(-\big(\frac{x-\mu(s)}{\sigma(s)}\big)\Big)\Big), & \text{if } \xi(s) = 0, \text{ with } x \in \mathbb{R}, \end{cases} \tag{2}$$

respectively,

$$z_p(s) = Q(p : \Lambda(s)) := \begin{cases} \mu(s) - \frac{\sigma(s)}{\xi(s)}\big(1-\big(-\log(1-p)\big)^{-\xi(s)}\big), & \text{if } \xi(s) \neq 0, \\ \mu(s) - \sigma(s)\log\big(1-\log(1-p)\big), & \text{if } \xi(s) = 0 \end{cases}, \tag{3}$$

for any $p = 1/r \in (0,1)$, with $r \in \mathbb{N}$ denoting the $r-$(year in our case) return period.

In particular, for every location $s \in S$, we have that during any given year, by definition, the probability of observing an annual sea level maxima event greater or equal to $z_p(s)$ is

$$p = \mathbb{P}\big[Z(s) \geq z_p(s)\big] = 1 - \mathbb{P}\big[Z(s) \leq z_p(s)\big] = 1 - G_s(z_p(s)) \text{ for } Z(s) \sim \text{GEV}(\Lambda(s)), \tag{4}$$

where $G_s(x) = G(x : \Lambda(s))$. Equivalently $z_p(s)$ is the extreme sea level event expected to occur once every $r = 1/p$ years, and that occurs every year with probability $p = 1/r$. Observe that the probability $p = 1/r$ is decreasing in r, which implies that we get further out in the tail of the distribution as $r \to \infty$. For the model simulation, we consider the return periods

$$r \in \{5, 10, 20, 30, 50, 100, 200, 500, 1000, 5000, 10000, 50000, 100000\}. \tag{5}$$

Let $\hat{\Lambda}_{ML}(s) = (\hat{\xi}_{ML}(s), \hat{\mu}_{ML}(s), \hat{\sigma}_{ML}(s))$ denote the maximum likelihood (ML) estimation of $\Lambda(s)$, for any $s \in S$. Then the maximum likelihood (ML) estimated return level is $\hat{z}_p(s) := Q(p : \hat{\Lambda}_{ML}(s))$, for any $s \in S$, with p and Q defined as above. We recall that if $\hat{\xi}_{ML}(s) \leq -1.0$, then there exist no consistent maximum likelihood estimates for any of the GEV parameters; see Remark 4 in (Dombry 2015). In particular, if $\hat{\xi}_{ML}(s) > -0.50$, then $\Lambda_{ML}(s)$ satifies asymptotic normality properties; see Proposition 3.3 in (Bücher and Segers 2017). Hence, we have that $\Lambda_{ML}(s)$ is consistent, with asymptotically normal confidence intervals, whenever $\hat{\xi}_{ML}(s) > -0.50$.

In addition to the above and in the spirit of completeness we note that Eq.(3) is continuous, for any $\xi(s) \in \mathbb{R}$. Moreover the support of the distribution GEV$(\Lambda(s))$ is defined by

$$\text{supp GEV}(\Lambda(s)) := \begin{cases} (-\infty, \infty), & \text{if } \xi(s) = 0, \text{ (Type I)} \\ [\mu(s) - \frac{\sigma(s)}{\xi(s)}, +\infty), & \text{if } \xi(s) > 0, \text{ (Type II)} \\ (-\infty, \mu(s) + \frac{\sigma(s)}{|\xi(s)|}], & \text{if } \xi(s) < 0, \text{ (Type III)} \end{cases} \tag{6}$$

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

where the Type I-III corresponds to the Gumbel, Frechet, respectively, Weibull families of extreme value distributions.

Observe that given the support of the distribution and the $r-$return levels in Eq.(6), respectively, Eq.(3). If $\xi(s) \neq 0$, then $z_p(s)$ is non-linear in both $\sigma(s)$ and $\xi(s)$, for increasing return periods $r \in \mathbb{N}$. Moreover, for any $\xi(s) \in \mathbb{R}_{<0}$, then $z_p(s) \to \sup\{\text{supp GEV}(\Lambda(s))\}$ as $r \to \infty$, where $\sup\{\text{supp GEV}(\Lambda(s))\}$ denotes the supremum (upper bound) of supp $\text{GEV}(\Lambda(s))$. Therefore, the return levels associated with greater return periods are closer to each other, whenever $\xi(s) < 0$. Conversely, if $\xi(s) \geq 0$, then $\sup\{\text{supp GEV}(\Lambda(s))\} = +\infty$, thus $z_p(s) \to +\infty$ as $r \to \infty$. In particular, $z_p(s)$ is concave, convex, and linearly about $-\log(1-p)$, for $\xi(s) < 0, \xi(s) > 0$, respectively, $\xi(s) = 0$. Therefore, if $\xi(s) \geq 0$, then the return levels associated with very larger return periods are not close to each other.

2.1.2 Bayesian statistical modeling

This section aims to give the reader a short recollection of Bayesian statistical modeling, for the more detailed presentation of the topic; see (Watanabe 2018) and (Robert et al. 2007) to name a few.

Suppose that $z := \{z^k\}_{k=0}^K$ is a collection of K independent realization of some random variable Z. We assume that the probability density function of Z is contained in $\mathcal{F} := \{p(z:\theta) : \theta \in \Theta\}$, for $\mathcal{F}$ the parametric family of some known probability distribution, where Θ denotes the parameter space.

Bayesian statistical modeling allows us to assign beliefs to the parameter θ independent of the data z, in the form of a prior distribution $p(\theta)$. We denote by $\mathcal{L}(z|\theta) := \prod_{k=0}^K \mathcal{L}(z^k|\theta)$ the likelihood function of z given the parameter θ. Then, by Bayes' Theorem, we have that the posterior distribution of θ conditioned on the data z, denoted by $p(\theta|z)$, is given by

$$p(\theta|z) = \frac{\mathcal{L}(z|\theta)p(\theta)}{\int_\Theta \mathcal{L}(z|\theta)p(\theta)d\theta} \tag{7}$$

$$\mathbb{E}_\theta[\mathcal{L}(z|\theta)] = \int\limits_\Theta \mathcal{L}(z|\theta)p(\theta)d\theta. \tag{8}$$

The determinant in Eq.(7) is the marginal likelihood. Note that the marginal likelihood is the expected value of the likelihood function $\mathcal{L}(z|\theta)$ over the prior distribution $p(\theta)$, see Eq.(8). Now, let z_* denote some future observations, with probability density function $p(z_*|\theta)$. Then the posterior predictive distribution of z_* conditioned on z, denoted by $p(z_*|z)$, is given by

$$p(z_*|z) = \int\limits_\Theta p(z_*|\theta)p(\theta|z)d\theta, \tag{9}$$

with $p(\theta|z)$ as in Eq.(7). Note that $p(z_*|z)$ is the expected value of $p(z_*|\theta)$ over the posterior distribution $p(\theta|z)$.

In Bayesian statistical modeling, there is an analog to the frequentist's confidence interval called the credibility (interval/region) set.

Let p denote the prior distribution of $\theta \in \Theta$, and consider the subset $C := \{\theta : p(\theta|Z = z)\} \subset \Theta$. Given any $\alpha \in (0,1)$, if

$$\mathbb{P}^p[C|Z = z] = \int\limits_C p(\theta|z)d\theta = 1 - \alpha, \tag{10}$$

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

then C is said to be an $100(1-\alpha\%)-$credible set, where $\mathbb{P}^p$ denotes the prior probability measure.

Given the above, let $C_k := \{\theta : p(\theta|Z=z) \geq k\}$, where k is the largest number, such that Eq.(10) holds. Then C_k is said to be the $100(1-\alpha\%)-$highest probability density set (HPD). Moreover, the $100(1-\alpha\%)-$HPD is the minimizing volume set (smallest) for all $100(1-\alpha\%)-$credible sets.

Observe that one can assign beliefs to the parameter(s) of the prior distribution $p(\theta)$. Let φ denote the parameter of prior distribution $p(\theta)$, and suppose that $p(\varphi)$ is the density function of φ. Then $p(\varphi)$ is the hyperprior distribution of the hyperparameter φ, with $\varphi \in \Phi$, where Φ denotes the hyperparameter space.

From here, it follows that Eq. (7), respectively, (9) can be expressed as

$$p(\theta,\varphi|z) \propto \mathcal{L}(z|\theta,\varphi)p(\theta,\varphi) \tag{11}$$

$$\propto \mathcal{L}(z|\theta)p(\theta|\varphi)p(\varphi) \tag{12}$$

and

$$p(z_*|z) = \int_{\Theta}\int_{\Phi} p(z_*|\theta,\varphi)p(\theta,\varphi|z)d\theta d\varphi, \tag{13}$$

where $\propto$ denotes being proportional to, which we employ to highlight that the marginal distribution of a hierarchical model is not usually analytically solvable.

2.2 Gaussian processes

We recall the definition of Gaussian processes, following (Dudley 2018). From there, we make some short remarks deemed relevant relevant to our simulation case.

Given any nonempty set $\mathcal{X}$, assume that $f : \mathcal{X} \to \mathbb{R}$ is some random function. If there exists a covariance function (positive-semidefinite, symmetric function) $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, and a mean function $\mu : \mathcal{X} \to \mathbb{R}$, such that $f(X) = (f(x_1),\ldots,f(x_n))^T \sim \mathcal{N}(\mu_{|X}, C_{|XX})$, where $f(X) \in \mathbb{R}^n$ and $X = (x_1,\ldots,x_n) \subset \mathcal{X}$, for any $n \in \mathbb{N}$, with mean vector $\mu_{|X} = (\mu(x_1),\ldots,\mu(x_n))^T \in \mathbb{R}^n$, and covariance matrix $C_{|XX} = \{c(x_i,x_j)\}_{i,j=1}^n \in \mathbb{R}^{n\times n}$. Then f is a Gaussian process, denoted by $f \sim GP(\mu, C)$, with $\mu(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$, and $C(\mathbf{x},\mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))]$, where C denotes the covariance operator. In particular, a Gaussian process is any collection of random variables, for which any finite subcollection have a joint Gaussian distribution.

The above implies that for any pair (μ, C), there exists a Gaussian process $f \sim GP(\mu, C)$, see (Dudley 2018). Moreover, it follows that $f \sim GP(\mu, C)$ is completely determined by its first and second-order order statistics (moments). Indeed, since $f \sim GP(\mu, C)$, it follows that the marginals of f follow a joint Gaussian distribution (multivariate normal distribution). Thus the marginals of f are determined by the first and second-order statistics of the joint Gaussian distribution. Therefore, verifying the claim that $f \sim GP(\mu, C)$ is completely determined by its first and second-order statistics. Hence, $f \sim GP(\mu, C)$ is completely determined by $f(\mu, C)$.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

From the previous arguments, one can deduce that the choice of the pair (μ, C) has implications on the behavior of the random function $f \sim GP(\mu, C)$. In particular, the smoothness of the random function f is induced by the smoothness of the covariance function C.

2.2.1 Conditional gaussian processes

In the spirit of completeness and clarity, we recall the Gaussian process regression, which is used to model each of the three GEV parameters in two of our models, see Sect.3.3.3 and 3.3.4. Moreover, the conditional Gaussian processes are used to obtain smooth parameters and return level surfaces over the region of interest, see Sect.4.3.

Suppose that $f \sim GP(\mu, C)$, with f, μ, C as in the previous section. Furthermore, assume that $y_i = f(x_i) + \epsilon_i$, for all $i = 1, \ldots, n$, with $\{\epsilon_i\}_{i=1}^n \overset{iid}{\sim} \mathcal{N}(0, \sigma_n^2)$. Let $D := \{X, Y\} = \{(x_i, y_i)\}_{i=0}^n$ denote the training data, then the Gaussian process f conditioned on the training data D is the Gaussian process $f|D \sim GP(\bar{\mu}, \bar{C})$, with

$$\bar{\mu}(x) = \mu(x) + C_{|xX}\left(C_{|XX} + \sigma_n^2 I_n\right)^{-1}(Y - \mu_X) \tag{14}$$

$$\bar{C}(x, x') = C(x, x') - C_{|xX}\left(C_{|XX} + \sigma_n^2 I_n\right)^{-1} C_{|Xx'}, \tag{15}$$

where $C_{|Xx} = C_{|xX}^T = (c(x_1, x), \ldots, c(x_n, x))^T$, and I_n is the $n \times n$ identity matrix. From here, it follows that the posterior distribution of f given the training data D is defined by

$$p(f|X, Y) = \frac{\mathcal{L}(Y|X, f)p(f)}{p(Y|X)}, \tag{16}$$

where $\mathcal{L}(Y|X, f), p(f)$, and $p(Y|X)$ denotes respectively, the likelihood of the training data D given f, the prior distribution of f, and the marginal distribution of D.

In particular, for all $m \in \mathbb{N}$, and for any collection of test data $Z_* := (z_{*1}, \ldots, z_{*m})$, with $Z_* \not\subset X$. Then, by the above argument, the random variable obtained by conditioning $f(Z_*)$ on the training data D have a joint Gaussian distribution $f(Z_*)|D \sim \mathcal{N}(\bar{\mu}_{|Z}, \bar{C}_{|ZZ})$, with $\bar{\mu}_{|Z}, \bar{C}_{|ZZ}$ defined as in Eq.(14), respectively, Eq.(15). Therefore, the posterior predictive distribution for $f_* := f(Z_*)$ conditioned on the test, respectively, training data Z_*, D is given by

$$p(f_*|Z_*, Y, X) = \int p(f_*|Z_*, f)p(f|Y, X)df. \tag{17}$$

Observe that to obtain the left-hand side of both Eq. (14) and Eq.(15), one needs to invert the gram matrix $C_{|XX} + \sigma_n^2 I_n$. One way to find the inverse and solve the matrix equations on the right- hand side of Eq.(14)-(15), is to apply the Cholesky decomposition. Recall that the covariance matrix is by definition positive-definite, symmetric matrix. Hence, the assumptions outlined in the Cholesky Factorization Theorem are satisfied. Therefore verifying the existence of such a decomposition. Given the Cholesky decomposition of the Gram matrix. Then both the Eq.(14) and Eq.(15), can be written using lower, and upper triangular matrices. Thus the new equations can be solved using forward and back substitution, for a detailed argument see (Williams and Rasmussen 2006).

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

2.2.2 Hilbert space approximation gaussian processes

Given any $f \sim GP(\mu, C)$, with covariance function c, where $f : \mathcal{X} \to \mathbb{R}$, for some nonempty $\mathcal{X} \subset \mathbb{R}^n$. Consider the eigenvalue problem for the negative Laplace operators, subject to the Dirichlet boundary conditions

$$-\Delta\phi = \lambda\phi, \text{ in } \Omega \quad \phi = 0 \text{ on } \partial\Omega, \tag{18}$$

where $\mathcal{X} \subset \Omega$, for $\Omega \subset \mathbb{R}^n$ compact and $\partial\Omega$ sufficently smooth. If c is staionary and isotropic, then there exists an approximation

$$C\phi(x) = \int c(x,x')\phi(x')dx' \approx \int \sum_{j}^{\infty} S(\sqrt{\lambda_j})\phi_j(x)\phi_j(x')dx', \text{ for any } x, x' \in \mathcal{X}, \tag{19}$$

where C and S denotes the covariance operator, respectively, the spectral density of c, for $\{\phi_j(x), \lambda_j,\}_{j=1}^{\infty}$ the eigenfunctions and eigenvalues of the negative Laplace operator; see (Solin and Särkkä 2020) for a sketch of the proof.

In particular, let $\tilde{c}_m(x,x')$ denote the $m-$term approximation of the RHS in Eq.(19), where $\tilde{c}_m(x,x')$ is defined by

$$\tilde{c}_m(x,x') = \sum_{j}^{m} S(\sqrt{\lambda_j})\phi_j(x)\phi_j(x'), \tag{20}$$

then $\tilde{c}_m(x,x')$ converges uniformaly in the limit to $c(x,x')$ as $m, L \to \infty$, where $\Omega \subset \prod_{i=1}^{n}[-L,L]^i$ and $[-L,L] \subset \mathbb{R}$. Therefore, given the Gaussian process regression $y_i = f(x_i) + \epsilon_i$ defined as in Sect.2.2.1, then posterior mean and covariance function converges uniformaly in the limit to the exact solution as $m, L \to \infty$; see (Solin and Särkkä 2020).

2.3 Latent spatial process models

For any $s = (x,y) \in S$, where $(x,y) = (\text{lat}, \text{long})$, then each of the entries in $\Lambda(s) = (\mu(s), \sigma(s), \xi(s))$ is assumed to vary smoothly about the $(\text{lat}, \text{long})-$coordinates. The entries of $\Lambda(s)$ are modeled using latent spatial processes, which we refer to as a latent spatial process model; see (Coles et al. 2001).

Given a latent spatial process model, then $Z(s_i)$ is conditionally independent of $Z(s_j)$, whenever $s_i \neq s_j$, for any $s_i, s_j \in S$, conditioned on the collection of parameters $\{\Lambda(s_i)\}_{i=0}^{n}$. Thus, we have that

$$Z(s_i)|(\Lambda(s_0), \Lambda(s_1), \ldots, \Lambda(s_n)) \overset{\text{iid}}{\sim} \text{GEV}\Lambda(s_i) \quad \text{for all } s_i \in S \text{ with } i = 0, 1, \ldots, n, \tag{21}$$

where each of the three parameters $\Lambda(s_i) = (\mu(s_i), \sigma(s_i), \xi(s_i))$ is the realization of a smooth varying random surface observed at the location $s_i \in S$, for every $s_i \in S$. So, for each of the three parameters, we assume that there exists necessarily a smooth parameter surface, which implies that the parameter values varies smoothly as a function of $s_i = (x,y)$.

Observe that the latent spatial process model does not account for extremal dependence between tide gauge stations. Moreover, the model setup implies that we will miss certain small-scale effects like wave set-up and bay effects.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

3 Data & models

3.1 Data

We use tide gauge time series data from the Global Sea Level Observing System 3 (GESLA-3), which is an open data source, see (Haigh et al. 2023). We restrict our attention to the Baltic Sea and parts of the North Sea. Moreover, we only consider data uploaded by Copernicus Marine Environment Monitoring Service (CMEMS), which we for simplicity referee to as GESLA data.

3.1.1 Data preprocessing

The GESLA data used for any type of extreme value analysis in the preceding sections is first preprocessed. The first step in the preprocessing workflow is the removal any time series data points which are flagged as do not use in analysis, respectively, interpolated values. This corresponds to a 0 in column 5, respectively, 2 in column 4 in the GESLA tide gauge time series dataset. Motivated by the recommendation of GESLA, respectively, the lack of available information regarding interpolation methods, with the possibility of non-uniformness about interpolation methods. From there, the time series data is detrended, using a 20-year centered moving average (CMA), for each station of the GESLA data contained in the region of interest. The use of a 20-year CMA is motivated by the need to account for both land up lift and mean sea level rise. For the interested reader, there is a time series analysis jupyter notebook; see Supplement.

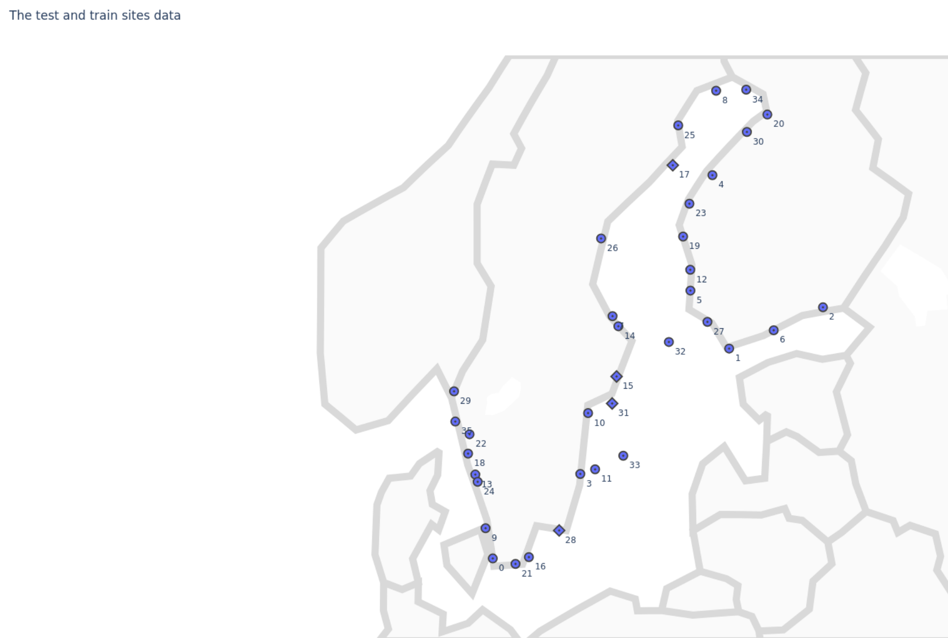
To handle the non-uniformness of the time-series frequency, at each location, we consider an observation frequency resolution of one minute. For any time-series data with courser frequency (usually hourly frequency), we simply add missing values to obtain a uniformness concerning the frequency of the time-series, which we motivate by the fact that we are only interested in the observed annual maxima, see Sect.2.1 and below, where a annum is defined as the shifted calender year (1 July yyyy to 30 June yyy(y+1)). From here, a Block maxima matrix (see Sect.2.1) is obtained by extracting the annual observed maxima, for each shifted calender year between 1850-2020, and for every station in the region of interest. The Block maxima matrix is a $(K \times N)$-matrix, where $K = 134$ is the number of observation years, and $N = 76$ is the number of stations, with 7001 of the 10184 being missing values (68.75%).

3.1.2 Train test split

The training data is a $(k \times n)$-matrix, where $k < K$ is the number of observation years, and $n < N$ is the number of stations and, for $K = 134$, and $N = 76$ the total number of observation years, respectively, stations in the region of interest. In particular, the training set for the region of interest is an $(79 \times 36)-$dim matrix with 864 (30.38%) of the 2844 entries being missing values, for the 79 shifted calendar years taken over the time period 1941-2220. Moreover, for any station in the training data, the total number of observed annual maxima is bounded below by 35, thus any station in the training data has at least 35, respectively, at most 79 observations over the time period 1941-2220.

The geographical positions of the tide gauge stations in the training data are illustrated in Fig.1.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025



(a) Tide gauge stations in the traning data

Figure 1. The plot shows the location of each of the tide gauge stations contained in the training set, described in Sect.3.1.2. The numbers indicate the stations column posistion. The circles, respectivley, diamonds corresponds to the stations not used, and used in the out-of-sample model validation step, see Sect.4.2.2.

3.1.3 Geographical spread

Due to both the large number of tide gauge stations and the varying oceanographic conditions prevailing at the different locations, we consider subsets of tide gauge stations. The aim of this is to obtain some geographical spread and to compare model results over a smaller number of stations. We consider the subsets listed in Table 1.

3.2 Baseline

The Baseline model is the standard maximum likelihood estimation method (MLM). For each station in the training data, Baseline obtains estimates of the three GEV parameters by solving the maximum likelihood equations; see (Bücher and Segers 2017). This implies that Baseline is not subject to priors, nor does Baseline account for any spatial dependency structure between tide gauge stations. In particular, Baseline only requires an adequate number of observations to obtain good maximum likelihood (ML) estimates for the GEV parameters.

From studying the kernel density estimate (KDE) plots, descriptive statistics, respectively, the max min values for each of the three GEV parameters, it follows that the lower bound for an adequate number is at least 20 observations; see Supplement.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

Subset name	Station indexing	Number of obs (Training)	Total number of obs
Skagerrak	13, 18, 22, 24, 29, 35	50, 53, 56, 41, 47, 79	51, 54, 57, 96, 48, 111
The coastline of Scania county (Scania)	0, 9, 16, 21, 28	79, 45, 38, 46, 79	92, 46, 39, 101, 135
The costline from Stockholm to Smaland	3, 10, 11, 15, 31, 33	60, 56, 57, 79, 66, 60	61, 56, 58, 133, 121, 61
The gulf of Finland	1, 2, 6	50, 50, 50	51, 51, 51
Bothnia bay south	5, 7, 12, 14, 19, 26, 27, 32	50,38,50,45,50,52,50,50	51,88,46,51,53,51,51
Bothnia bay north	4, 8, 17, 20, 23, 25, 30, 34	50,46,79,50,50,79,50,50	51,47,130,51,51,106,51,51
Top 4 stations with fewest NaNs	15, 17, 28, 31	79, 79, 79,66	135, 133, 130, 121

Table 1. The first column corresponds to the sub collection names referenced in the text and in the figures. The station indexing column shows the stations continued in each sub collection, for the indexing presented in Fig.1. The two last columns show the total number of observed annual maxima's in the training data, respectively, the overall data, and presented using the order in the Station indexing column.

In particular, for any given station $s \in S$, then Baseline uses all of the observed annual maxima's in $z(s) := \{z^k(s)\}_{k=0}^{K}$, for $K \leq 134$, with $K \in \mathbb{N}$. This implies that Baseline uses a larger Block maxima matrix, referred to as the Regional block maxima matrix. The Regional block maxima matrix is a $(134 \times 51)-$dim matrix, where 3947 of the total 6834 entries are missing values $(57,76\%)$.

The uncertainty of the Baseline estimated return levels are quantified using $95\%-$confidence intervals ($95\%-$CI), where $\hat{\xi}_{ML}(s)$ satisfies the asymptotic normality properties, and the ($95\%-$CI) are obtained using bootstrapping and 10000 resampling's.

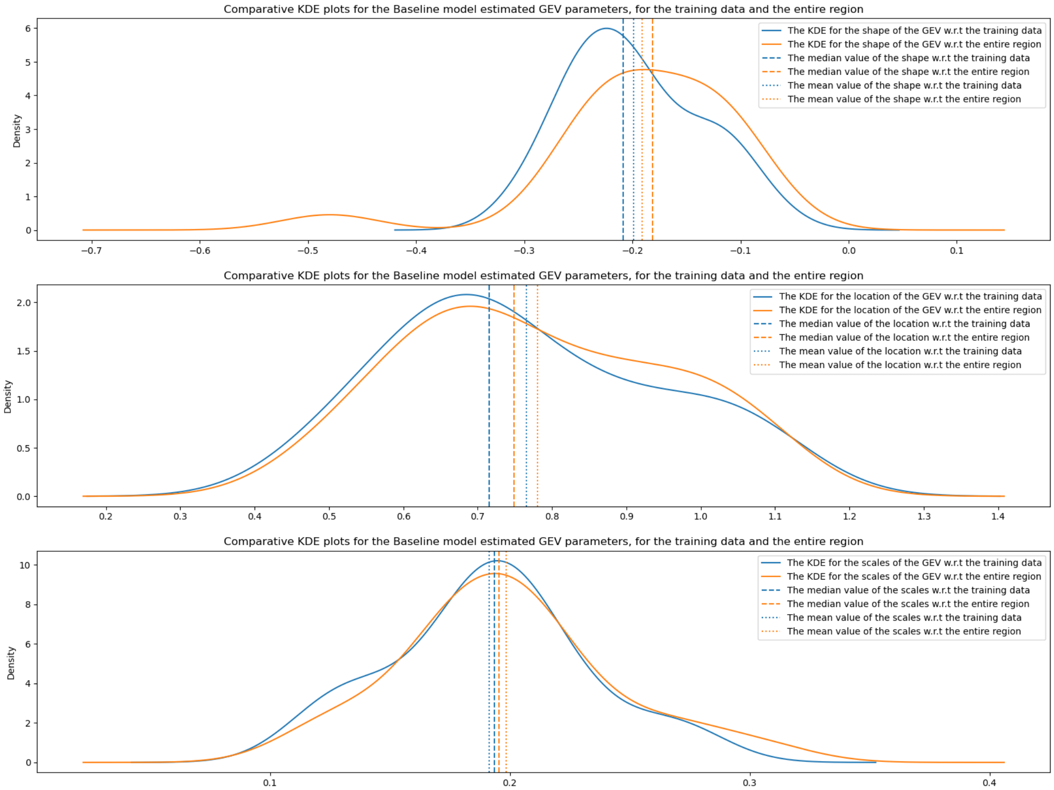
One way of studying the difference between the Regional block maxima matrix and the training data is to compare the KDE plots of the ML estimated GEV parameters. From the KDE plots in Fig.2, it follows that the training data is adequately capturing the Regional block maxima matrix.

3.3 Models

The models presented here are variations of the models in (Calafat and Marcos 2020) and (Räty et al. 2023). The key difference is that we consider a spatially varying GEV shape parameter for two of the models.

Each of the models are specified using PyMC, which is a probabilistic programming package in Python; see (Abril-Pla et al. 2023). For each of the models, we have applied the Bayesian workflow, and the MCMC diagnostics following the recommendations outlined in (Gelman et al. 2020), (Gabry et al. 2019), (Vehtari, Gelman, Simpson, et al. 2021), and (Palacios-Rodriguez, Bernardino, and Mailhot 2023). We use Markov Chain Monte Carlo simulations (MCMC) to obtain inference objects for each model. The sampling is conducted using the No-U-Turn Sampler (NUTS), which is available in PyMC. For the MCMC set-up, we use four chains on four cores, which implies four individual chains. For each chain, samples from 15000 draws, where the first 5000 draws are used for tuning (discarded from the actual sampling). Thus, each of the four individual chains uses 10000 samples, where sampling is done using NUTS, and with a target accept of 0.99-0.995.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025



(a) Baseline KDE plots

Figure 2. KDE plots for the asympotically normal ML estimated GEV paramters over the Regional block maxima matrix and training data.

Observe that the parameter specifications of the different priors used by the models are provided in the Supplement.

Recall the posterior density's in Sect.2.1.2, and observe that for each of the BHM models, we obtain posterior density's, for both the GEV parameters and $r-$year return level, with r as in Eq.(5).

3.3.1 Common

Let $\mathcal{N}$,$\mathcal{N}^+$, $|\mathcal{N}|$, denote the Normal, Half normal, respectively, Truncated normal distribution. Recall that $Z(s)$ denotes the random variable of interest corresponding to the annual sea-level maxima at $s \in S$.

For any $s_i, s_j \in S$, with $s_i \neq s_j$, then Common assumes that $Z(s_i)$ is independent of $Z(s_j)$, thus

$$Z(s_i) \overset{\text{iid}}{\sim} \text{GEV}\Lambda(s_i) \quad \text{for all } s_i \in S. \tag{22}$$

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

For every $s_i \in S$, and for each of the GEV parameters $\Lambda(s_i) = (\mu(s_i), \sigma(s_i), \xi(s_i))$, then Common assumes a hierarchical data structure with prior and hyperpriors as below

$$\mu(s_i) \overset{\text{iid}}{\sim} \mathcal{N}(\mu_\mu, \sigma_\mu) \quad \text{where } \mu_\mu \sim \mathcal{N},\ \sigma_\mu \sim \mathcal{N}^+, \tag{23}$$

$$\sigma(s_i) \overset{\text{iid}}{\sim} |\mathcal{N}|(\mu_\sigma, \sigma_\sigma) \quad \text{where } \mu_\sigma \sim |\mathcal{N}|,\ \sigma_\sigma \sim \mathcal{N}^+, \tag{24}$$

$$\xi(s_i) \overset{\text{iid}}{\sim} |\mathcal{N}|(0, \sigma_\xi) \quad \text{where } \sigma_\xi \sim \mathcal{N}^+. \tag{25}$$

In particular, Common assumes that each of the GEV parameters is drawn individually from a prior distribution. The parameters of the prior distributions are drawn from a collection of hyperpriors; see Supplement.

3.3.2 Separate

For any $s_i, s_j \in S$, then Separate assumes that $Z(s_i)$ is independent of $Z(s_j)$, see Eq.(22). For every $s_i \in S$, and for each of the GEV parameters $\Lambda(s_i) = (\mu(s_i), \sigma(s_i), \xi(s_i))$, the parameters are drawn independently from prior distributions

$$\mu(s_i) \overset{\text{iid}}{\sim} \mathcal{N}, \tag{26}$$

$$\sigma(s_i) \overset{\text{iid}}{\sim} |\mathcal{N}|, \tag{27}$$

$$\xi(s_i) \overset{\text{iid}}{\sim} |\mathcal{N}|. \tag{28}$$

Note that for each of the GEV parameters at any stations $s_i, s_j \in S$ are drawn independently of each other. However, Separate assumes that the parameters of the prior distributions are fixed; see Supplement. In particular, Separate assumes a nonhierarchical data structure, which implies that Separate has inherently less variability between stations.

Given the model formulations for Common (23), and Separate (26), then the two models have no way of conditioning the learned information to new locations. Therefore, there exists no "Bayesian way" of interpolating to new locations.

3.3.3 Hilbert

Hilbert is a latent spatial process model; see Sect.2.3. The latent spatial processes are Hilbert space approximated Gaussian processes; see Sect.2.2.2. The approximations are made in the $(x_1, x_2, x_3)-$Cartesian coordinate system, with $(x_1, x_2, x_3) = \Psi(\text{lat}, \text{lon})$, where Ψ denotes the Geographic to Cartesian coordinate transformation. The coordinate transformation is due to the lack of information pertaining the coordinate systems for the the different tide gauge stations.

Now, given any $s_i \in S$, with $s_i = (\text{lat}_i, \text{lon}_i)$, then Hilbert assumes that

$$Z(s_i)|(\Lambda(s_0), \Lambda(s_1), \ldots, \Lambda(s_n)) \overset{\text{iid}}{\sim} \text{GEV}\Lambda(s_i) \quad \text{for all } s_i \in S \text{ and for } i = 0, 1, \ldots, n, \tag{29}$$

with

$$\mu(s_i) = f_\mu(s_i) + \epsilon_\mu \quad \text{for } f_\mu \sim \mathcal{GP}(\mu_{f_\mu}(s_i), C_{f_\mu}(s_i,s_j)) \text{ and } \epsilon_\mu \sim \mathcal{N}(0, \sigma^2_{\epsilon_\mu}), \tag{30}$$

$$\sigma(s_i) = f_\sigma(s_i) + \epsilon_\sigma \quad \text{for } f_\sigma \sim \mathcal{GP}(\mu_{f_\sigma}(s_i), C_{f_\sigma}(s_i,s_j)) \text{ and } \epsilon_\sigma \sim \mathcal{N}^+(0, \sigma^2_{\epsilon_\sigma}), \tag{31}$$

$$\xi(s_i) = f_\xi(s_i) + \epsilon_\xi \quad \text{for } f_\xi \sim \mathcal{GP}(\mu_{f_\xi}(s_i), C_{f_\xi}(s_i,s_j)) \text{ and } \epsilon_\xi \sim \mathcal{N}(0, \sigma^2_{\epsilon_\xi}), \tag{32}$$

where

$$f_{(\cdot)} \sim \mathcal{GP}(\mu_{f_{(\cdot)}}(s_i), C_{f_{(\cdot)}}(s_i,s_j)), \tag{33}$$

denotes a Gaussian Processs (see Sect.2.2), with mean and covariance function $\mu_{f_{(\cdot)}}(s_i)$, respectively, ,$C_{f_{(\cdot)}}(s_i,s_j)$. Let δ_{ij} denote the Kronecker delta, and recall that $\delta_{ij} = 1$ whenever $i = j$, and $\delta_{ij} = 0$ whenever $i \neq j$, for any $i,j = 1,2,\ldots,n$.

For each of the three GP processes in Eq.(30)-(32), the mean function $\mu_{f_{(\cdot)}}$ is defined as a linear combination of

$$\mu_{f_{(\cdot)}}(s_i) = \mu_{f_{(\cdot)}}(\text{lat}_i, \text{lon}_i) := \beta_0^{f_{(\cdot)}} + \beta_1^{f_{(\cdot)}}(\text{lat}_i) + \beta_2^{f_{(\cdot)}}(\text{lon}_i), \tag{34}$$

where $\{\beta_1^{f_{(\cdot)}}, \beta_2^{f_{(\cdot)}}\} \overset{iid}{\sim} |\mathcal{N}|$ for each $f_{(\cdot)}$, with $\beta_0^{f(\xi)} = 0$ and $\beta_0^{f(\mu)}, \beta_0^{f(\sigma)} \sim \mathcal{N}, |\mathcal{N}|$

Here we use, for each of the three GP processes in Eq.(30)-(32), the Matérn $5/2$ covariance function is given by

$$C_{f_{(\cdot)}}(s_i,s_j) = \rho^2_{f_{(\cdot)}} \Big(1 + \frac{\sqrt{5}d(s_i,s_j)}{l_{(\cdot)}} + \frac{5d(s_i,s_j)^2}{3l^2_{(\cdot)}}\Big)\exp\Big(-\frac{\sqrt{5}d(s_i,s_j)}{l_{(\cdot)}}\Big) + \sigma^2_{\epsilon_{(\cdot)}}\delta_{ij}, \tag{35}$$

with $d(s_i,s_j)$ the euclidean distance function, where $l_{(\cdot)} \in \mathbb{R}_{>0}$, and $\rho_{f_{(\cdot)}} \sim N(0, \sigma^2_{\rho_{f_{(\cdot)}}})$ denotes the lenghtscale of the GEV parameter, respectively, the noise term of the covariance fucntion. For each of the three covariance functions, we assume deterministic length scales

$$l_{(\xi)} = 350, \tag{36}$$

$$l_{(\mu)} = 350, \tag{37}$$

$$l_{(\sigma)} = 350. \tag{38}$$

The spectral density (power spectrum) of the Matérn $5/2$ covariance function is

$$S(\boldsymbol{\omega}) = \rho^2 \frac{2^n \pi^{n/2} \Gamma(\frac{5+n}{2}) 2^{5/2}}{\Gamma(5/2) l^5_{(\cdot)}} \Big(\frac{5}{l^2_{(\cdot)}} + 4\pi^2 \boldsymbol{\omega}^T \boldsymbol{\omega}\Big)^{\frac{-5+n}{2}}, \tag{39}$$

with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n) \subset \mathbb{R}^{n=3}$, where $\Gamma(\cdot)$ denotes the gamma function.

The $3-$dim input space $\Omega = \prod_{i=1}^3 [-L_i, L_i]^i$, the total number of basis function m^*, and the proportional extension factor $c \geq 1$ is given by

$$m^* = \prod_{i=1}^3 m_i = 12 \times 11 \times 20 = 2640, \tag{40}$$

$$L_i = c \cdot S_i, \;\text{ for } S_i = \max_{i=1,2,3} \{|x_i| : (x_1,x_2,x_3) = \Psi(s),\, s \in \mathcal{S}\}, \quad (\text{with } c = 2.8). \tag{41}$$

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

3.3.4 Latent

Latent is a latent spatial process model, where the latent spatial processes are Guassian processes; see Sect.2.3, respectively, Sect.2.2.

For any $s_i \in S$, with $s_i = (\text{lat}_i, \text{lon}_i)$, then Latent assumes that

$$Y(s_i)|(\Lambda(s_0), \Lambda(s_1), \ldots, \Lambda(s_n)) \overset{\text{iid}}{\sim} \text{GEV}\Lambda(s_i) \quad \text{for all } s_i \in S \text{ with } i = 0, 1, \ldots, n, \tag{42}$$

with

$$\mu(s_i) = f_\mu(s_i) + \epsilon_\mu \quad \text{for } f_\mu \sim \mathcal{GP}(\mu_{f_\mu}(s_i), C_{f_\mu}(s_i, s_j)) \text{ and } \epsilon_\mu \sim \mathcal{N}, \tag{43}$$

$$\sigma(s_i) = f_\sigma(s_i) + \epsilon_\sigma \quad \text{for } f_\sigma \sim \mathcal{GP}(\mu_{f_\sigma}(s_i), C_{f_\sigma}(s_i, s_j)) \text{ and } \epsilon_\sigma \sim \mathcal{N}^+, \tag{44}$$

$$\xi(s_i) = f_\xi(s_i) + \epsilon_\xi \quad \text{for } f_\xi \sim \mathcal{GP}(\mu_{f_\xi}(s_i), C_{f_\xi}(s_i, s_j)) \text{ and } \epsilon_\xi \sim \mathcal{N}, \tag{45}$$

with

$$f_{(\cdot)} \sim \mathcal{GP}(\mu_{f_{(\cdot)}}(s_i), C_{f_{(\cdot)}}(s_i, s_j)), \tag{46}$$

as in Sect.3.3.3, for each of the GEV parameters.

4 Results

This section presents the results obtained from applying the methods and models outlined in Sect.2-3. Recall that the models use, up to a target accept specification, an equvivalent MCMC set up (see Sect.3.3). Recall that for each of the models, there exists a comprehensive model diagnostics notebook; see Supplement.

We first consider a model comparison analysis; see Sect.4.1. The model evaluation and validation are conducted in Sect.4.2, where we consider in-sample and out-of-sample validation, see Sect.4.2.1, respectively, Sect.4.2.2. Smooth parameter and return levels surfaces are presented in Sect.4.3.

To evaluate the models (outside the Bayesian workflow), we need to assign a ground truth. Given that we are primarily interested in the return levels, we assign the return levels obtained using the maximum likelihood estimator (MLM) to be our ground truth. However, we note that, especially for short gauges earlier work has demonstrated that even a single year added to a short observational record can give rise to substantial changes to the MLM estimator (M. Hieronymus and Kalén 2020). Our ground truth is thus not the real truth, but at least a somewhat independent estimate that is the current standard used to determine design heights and that should work reasonably well for longer gauges.

Given the large number of stations in the training data, we occasionally restrict our attention to some of the subsets outlined in Table 1. However, for any given plot, we add relevant remarks for the stations (subsets) not being presented in the plot. Moreover, for any plot about some subset, there exist, for every subset in Table 1, equivalent plots in the Jupyter notebook; see Supplement.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

4.1 Model comparison

First, we consider relevant statistics for the four models; see Sect.4.1.1. From there we study both the box-whiskers and violin plots for the $50, 1000-$year return levels, respectively, the shape parameter; see Sect.4.1.2.

4.1.1 Pareto smoothed importance sampling leave-one-out cross validation

The statistics presented in Table 2 are obtained from the Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO-CV) serve as a "goodness of fit" for the models; see (Vehtari, Gelman, and Gabry 2017). Table 2 is a model comparison table, where the model with the highest predictive accuracy is given the lowest rank. Predictive accuracy is defined as the sum of the expected log pointwise predictive density (elpd_{loo}), where the sum is taken over all of the data points in the training data, with each point being treated as an out-of-sample prediction, for every data point in the training data. The comparison emphasizes a high (elpd_{loo}). However, there are other important statistics. In particular, the predictive difficulties of future (test) data compared to training data are indicated by the effective number of parameters (p_{loo}), where the predictive capability is decreasing in (p_{loo}). Moreover, if $p_{loo} > p$ and $p_{loo} > N$, for the total number of parameters p and datapoints N, then the predictability of the model is questionable at best; see (Vehtari, Gelman, and Gabry 2017). The cavitate is that the reliability of PSIS-LOO-CV statistics is dependent on the Pareto smoothed importance sampling (PSIO). In fact, if $\hat{k} < 0.7$, where $\hat{k}$ denotes the estimated shape parameter of the generalized Pareto distribution, then the PSIS-LOO-CV statistics are reliable; see (Vehtari, Simpson, et al. 2024).

From the above discussion, it follows that the PSIS-LOO-CV statistics are reliable for Hilbert and Latent. Recall that the training data contains a total of 2844 datapoints, so $p_{\text{loo}} < N$ for each of the four models. Moreover, for any of the four models, there exists at least 3×36 parameters, which implies that $p \geq 108$, thus $p > p_{loo}$ for each of the four models; see Table 2.

	rank	(elpd_{loo}),	p_{loo}	($\text{elpd}_{\text{diff}}$),	SE	warning ($\hat{k} \geq 0.7$)
Hilbert	0	319.036085	39.267076	0.000000	29.657687	False
Latent	1	318.173279	35.403501	0.862807	29.810929	False
Common	2	310.592203	78.147082	8.443882	29.909770	True
Separate	3	309.916475	80.546255	9.119610	30.029866	True

Table 2. The PSIS-LOO-CV statistics for each of the four models, with respect to the training data. Fix any of the two pairs (Hilbert, Latent) and (Common, Separate), then the PSIS-LOO-CV statistics is similar for the two models. For the pairwise comparison, we see that the leveraging spatial dependency at the parameter level results in better PSIS-LOO-CV statistics. Note that the comparison is not necessarily intrinsic to the model types, since the PSIS-LOO-CV statistics depends on both the individual observations and the total number of observations. Instead the properties are emergent with respect to the training data.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

4.1.2 Comparing return levels and the shape parameter

From Sect.4.1.1, it follows that Hilbert and Latent gave the best PSIS-LOO statistics, see Table 2. Therefore, we will restrict our attention to the two latent spatial process models. However, we first consider some comparative plots for each of the four models.

The box and whiskers plots in Fig.3, show the $50, 1000-$year return levels, for each station in the training data. In both (a) and (b) of Fig.3, we see that the two spatial process models have similar position and sizes of the boxes. In particular, the lengths of the whiskers are similar for Hilbert and Latent. The analogue observations are valid with respect to Common and Separate, with the obvious exception holding for station 34. The comparabilities could be the consequences of similar (hyper) prior specifications.

For the station comparison, we see that there is a clear discrepancy between the sizes of the boxes, the median values, and the length, respectively, symmetry of the whiskers when comparing both Common and Separate to Hilbert or Latent (except for stations 5, 6, 12, 26, and 27). In particular, for both Common and Separate, the 95^{th} percentile whiskers are visibly longer than the ones for Hilbert and Latent. This implies that the posterior densities of the corresponding return levels have more mass, respectively, long tails to the right of the mean.

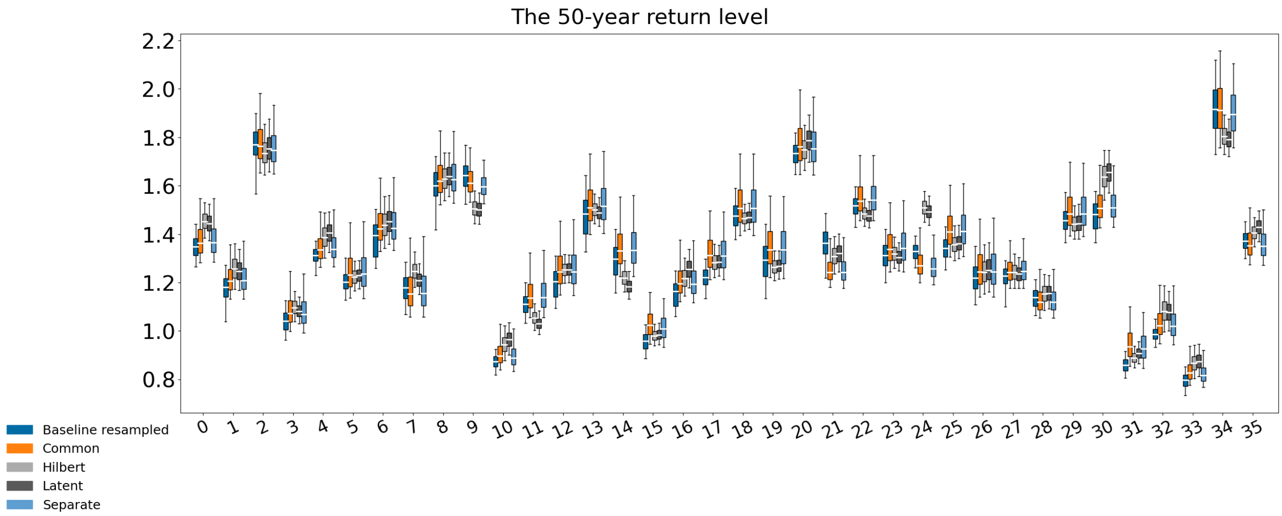
For every station in the training data, we plot the posterior densities of the GEV shape parameter, for each of the four models; see (a)-(b) of Fig.4. In both (a) and (b) of Fig.4, we see an example of the agreement in the (hyper) prior specification, for the pairs Hilbert and Latent in (a), respectively, Common and Separate in (b). In particular, fix any model in (a), then there exist only subtle differences between both the positioning and shape of the posterior densities, which implies that the (hyper) prior specification do not adequately capture the spatial variability of the shape parameter, for Hilbert and Latent. However, the posterior densities of (a) are approximately symmetric and have most of their mass centered around the mean. Therefore, for any of the stations in the training data. Then the volume of the $95\%-$HPD set about the shape parameter is smaller compared to the ones obtained from Common or Separate, since the posterior densities of the shape parameter for both Common and Separate are approximately uniform.

4.2 Model validation

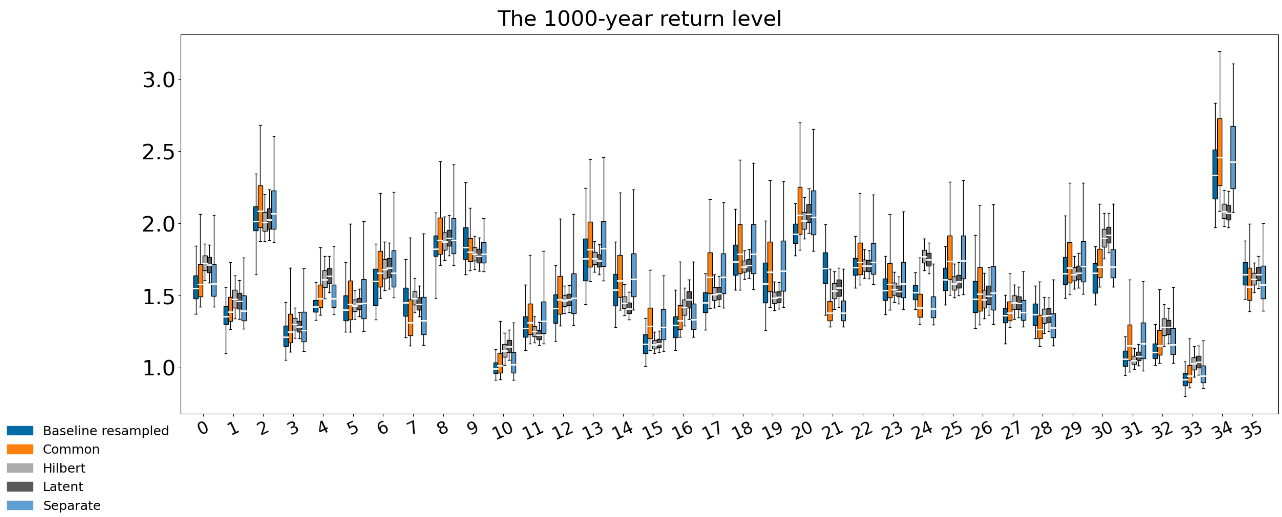
For model evaluation and validation, we consider the in, respectively, out-of-sample case; see Sect.4.2.1, respectively, 4.2.2. We recall the caveat given in the introduction of Sect.4, and assign Baseline to be our ground truth, where the uncertainty in Baseline is quantified using $95\%-$CI; see Sect.3.2. From there, we consider the posterior density plots of the GEV parameters, respectively, comparative return levels plots, for Hilbert and Latent. We use the $95\%-$HPD sets, see Sect.2.1.2, to quantify uncertainties with respect to Hilbert and Latent. Note that Baseline utilizes all available observations for any given station in the region of interest.

For each of the three GEV parameters, the $95\%-$HPD of the posterior density contains the corresponding MLE value for all stations in the training data. Note that we do not compare distributions to point estimates. Instead, the MLEs serve as reference

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025



(a) Comparative box and whisker plot for the posterior densitys of the 50-year return level



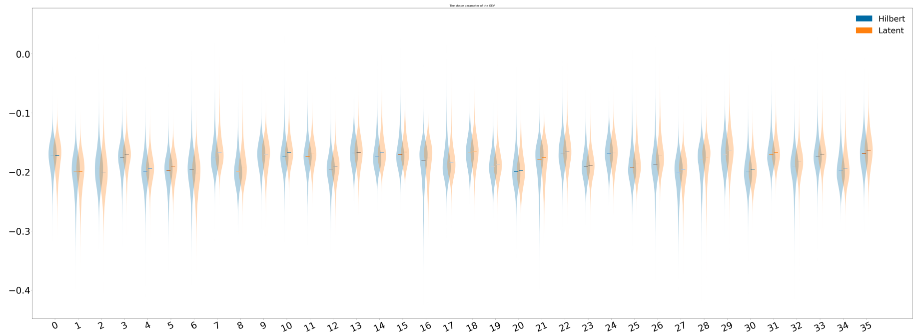
(b) Comparative box and whisker plot for the posterior densitys of the 1000-year return level

Figure 3. The $x,y-$axes correspond to different stations in the training data, respectively, in units of meters the $50,1000$-year return levels. The position of the whiskers is at the 5^{th}, respectively, 95^{th} percentile, while the white lines in the boxes represent the median value. Observe that the boxes are lexicographically ordered according to the names of the four models.

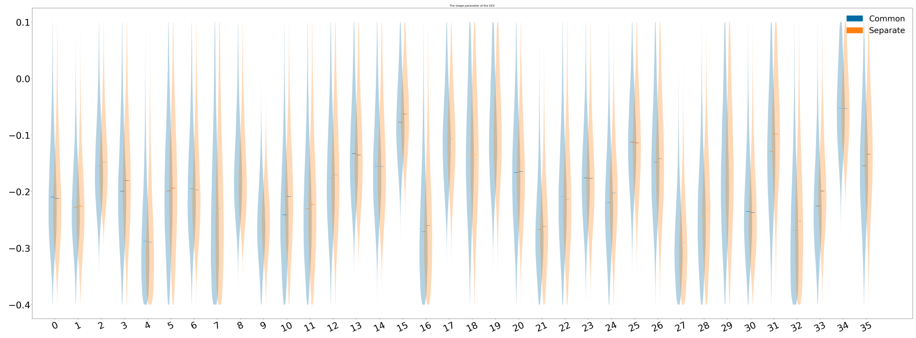
points. This allows for the study of how adequately the model captures the ML estimated GEV parameters, and some local variability. Given the (hyper)prior specifications made for the entire region of interest.

Recall that the return level plots are in the form of semi-log plots, where the $x,y-$axes are in a log scale corresponding to the $\log_{10}(10^r) = r$ year, respectively, linear scale corresponding to the sea levels in meters. The thick dashed line is the maximum

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025



(a) (Hilbert and Latent) The shape parameter of the GEV



(b) (Common and Separate) The shape parameter of the GEV

Figure 4. Violin plots for the shape parameter of the GEV at the different stations in the training data. The violin plot gives the shape of the probability distribution, obtained by using a kernel density estimator (KDE). In the above plots, we restrict our attention to the median value. From the shape of the distributions at each station, it follows that one can deduce which values are more likely to occur in the posterior distributions of the MCMC simulations. For a normally distributed sample, then the mass is centered around the mean value, with the mean and median being the same. Comparing this to the standard uniform distribution, where the mass is evenly distributed between 0 and 1.

of the observed annual sea level maxima at the given location. For both Hilbert and Latent, we use full lines to plot the return level curves, the black and read lines are the end points of the $95\%-$HPD sets, respectively, the mean values. For Baseline, we use thin dashed lines and the analogous color scheme.

Due to the large number of stations in the training, we restrict our attention to thethe Gulf of Finland, Scania, and the Top 4 subset, see Table 1. In particular, for the out-of-sample case, we consider only the Top 4 subset.

4.2.1 In-sample model validation

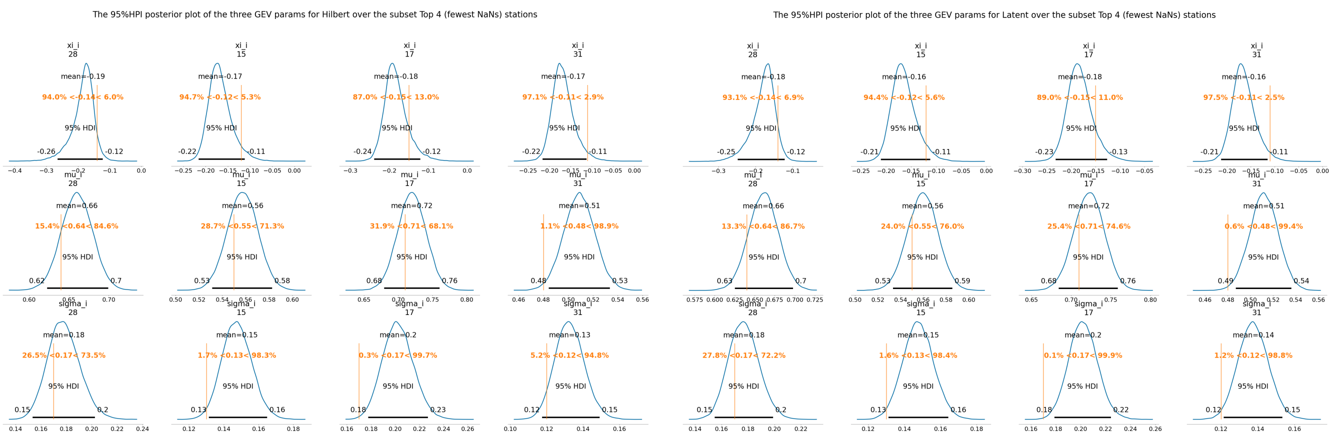
We first consider the Top 4 subset; see Fig.5, respectively, Fig.6 for the $95\%-$HPDs of the posterior densities and the return level curves. From there we consider the two models separately; see Fig.7-8.

From the posterior density plots; see (a)-(b) in Fig.5. We see that although the bounds of the $95\%-$HPD sets are tight, the MLE parameters do not always intersect the bulk of their corresponding $95\%-$HPD sets for each of the three GEV parameters.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

From the disscusion of the nonlinear part in the return levels; see the end of Sect.2.1.1. It follows that there exists an asymmetry between the preferred misalignment of the posterior densities and their corresponding MLEs. However, both models yield tight bounds with respect to the 95%−HPD sets for the different return levels. This is due to the tightness of the 95%−HPD posterior densities of the GEV parameters.

The comparative return level plots in (a) and (b) of Fig.6, give good indications of the overall performance. For stations 28, 15, and 31, the mean level curves of both Hilbert and Latent agrees with Baseline. In particular, for the 5 to 20 year return periods, the bounds of the 95%−HPD sets are approximately in line with the bounds of the 95%−CIs. Note that the mean level curve of both Hilbert and Latent at station 17, lies above Baseline. However, for each of the four stations, the return levels curves corresponding to the lower, respectively, upper bound of the 95%−HPD sets are closer to the mean level curve compared to Baseline for increased return periods, corresponding to the right tail of the distribution. This implies that both Hilbert and Latent produce better estimates (reduced uncertainty) compared to Baseline for both the bulk and the right tail of the distribution, with the latter corresponding to lower probability events.



(a) Hilbert in sample MLE comparison

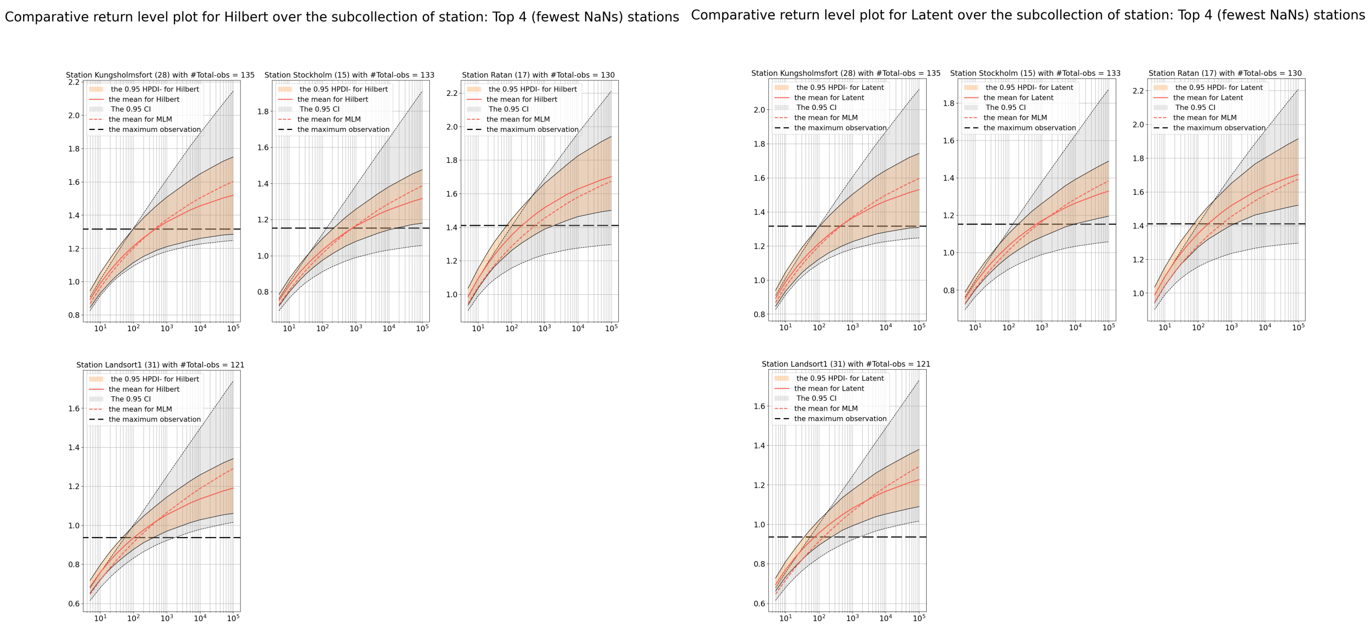
(b) Latent in sample MLE comparison

Figure 5. The 95%−HPD posterior densities of the three GEV parameters, for the four stations in the Top 4 subset. Fix any GEV parameter, then one can see that the models obtain similar posterior densities, for any of the four stations. In particular, for the location and scale parameter, then each of the two models obtain tight bounds with respect to the 95% HPD sets.

We restrict our attention to Hilbert and consider the Scania subset; see (a)-(b) in Fig.7. From (a), we see that MLEs are not consistent at stations 9 and 16; see the end of Sect.2.1.1. Therefore, we consider only stations 0, 21, and 28.

For the stations with consistent MLEs, we can observe that Hilbert struggles to capture some of the variation about the shape parameter, where the converse holds for both the scale and location parameter. In particular, for the shape parameter at station 21, we see that 99.9% of the 95%−HPD set lies to the right of the MLE. Therefore, if we compare Hilbert and Baseline, the maxima of all observed maxima at station 21, then the maxima is on average a 10000-year, respectively, 1000-year event. This implies that Hilbert underestimates the upper bound of the support for the GEV distribution at station 21.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025



(a) Hilbert vs Baseline (MLE) in-sample comparison

(b) Latent vs Baseline (MLE) in-sample comparison

Figure 6. Comparative return level plots, for the four stations in the Top 4 subset. From the plots, we see that both Hilbert and Latent have more symmetric uncertainty bounds. However, Baseline exhibits right-tailed asymmetry about the endpoints of the 95%-CI uncertainty bounds, with an increasing asymmetry about the x-axis. Moreover, from each of the return level plots in (a)-(b), we see that the respective model and Baseline both approximately agrees on the average return period for the observed maximum at each station.

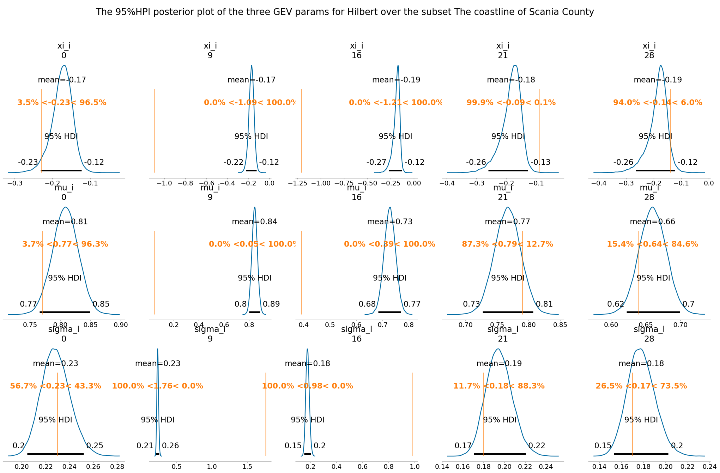
If we instead consider station 0, then we see that $\hat{\sigma}_{ML}$ agrees with the mean, while $\hat{\xi}_{ML}, \hat{\mu}_{ML}$ intersects the left tail of the corresponding $95\%-$HPD set. This illustrates some of the difficulties in specifying hyperpriors. In fact, the caveat of Sect.4 is applicable with respect to the KDE plots of Fig.2, where the lower MLE values for the shape parameters could be the consequence of not enough observations to adequately capture the tail of the distribution.

Consider the Gulf of Finland, and Latent; see Fig.8. Then, for each of the three stations, the shape parameter $\xi(s)$ has identical mean and similar bounds for the $95\%-$HPD sets; see (a) of Fig.8. Note that the MLE value $\hat{\xi}_{ML}(s)$ intersects more of the bulk here compared to Scania. In particular, given $\hat{\mu}_{ML}(s)$,$\hat{\sigma}_{ML}(s)$, then the two estimates align well with the mean of their corresponding posterior densities for each of the three stations. From there return level plots, we see that the uncertainty, analogues to Scania and Hilbert, is reduced for the lower probability events; see (b) of Fig.8.

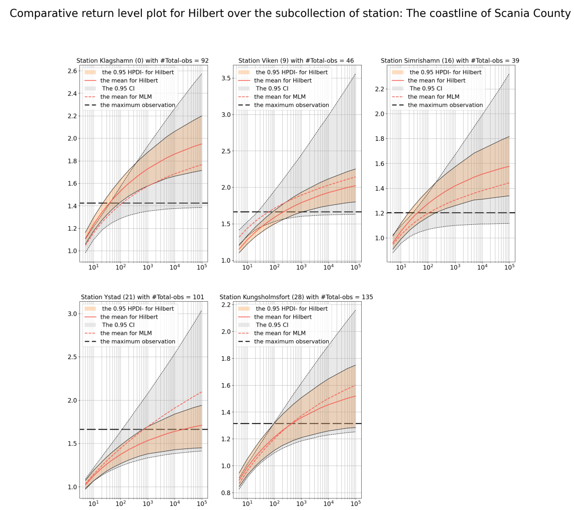
4.2.2 Leave-one-out model validation

For the out-of-sample case, we consider one at a time the stations in the Top 4 subset, where each station is treated as an out-of-sample prediction and is referred to as a leave-one-out (loo) station. For any loo station, we obtain posterior predictive densities for the GEV parameters and return levels using conditional GPs, see Sect.2.2.1. Observe that for any loo station (and

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

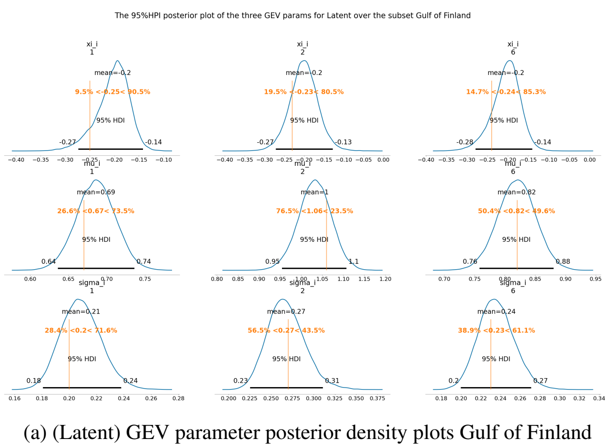


(a) (Hilbert) GEV parameter posterior density plots Scania

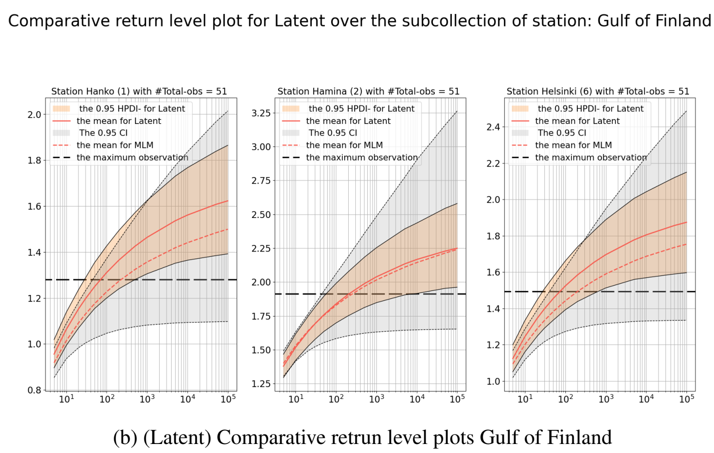


(b) (Hilbert) Comparative return level plots Scania

Figure 7. (Hilbert about the Scania subset). If we consider plots (b), then we see that Hilbert unlike Baseline is able to obtain tight uncertainty bounds at stations with lower number of observation years. In particular, we see that the bounds of the GEV parameters in (a) are similar to the ones in Fig.5, which indicates that Hilbert make us of the defined spatial dependency to estimate the parameters.



(a) (Latent) GEV parameter posterior density plots Gulf of Finland



(b) (Latent) Comparative retrun level plots Gulf of Finland

Figure 8. (Latent about the Gulf of Finland subset) If we compare the posterior densities in (a) to the ones in (a) of Fig.5, and to Hilbert's in (a) of Fig.7. Then we see that the mean and endpoints for both the location and scale parameter is larger, for the stations in the Gulf of Finland subset. From here, along with the MLE scale parameter intersecting the bulk of the distribution, one can deduce that the model adequately captures some of the variation present in the different subsets.

every out of sample station), then the position of the loo station relative to the station in the training data have implications on the conditional Gaussian processes, by definition of covariance functions, see Eq.(35).

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

From the comparative return level plots, see (a)-(d) of Fig.9. We see that Hilbert and Latent obtain roughly the same return levels and uncertainty estimates for (b) and (d), where (b), and (d) correspond to station 15, respectively, 31 in the Top 4 subset. Moreover, the mean return level curves for Latent and Baseline are approximately in agreement. However, if we compare the return level curves in Fig.9 to the ones in Fig.6. Then, for every loo station, we see an increased uncertainty about the return levels. In particular, we see an increase in the size of the $95\%-$HPD set, for the lower period return levels corresponding to the bulk of the distribution.

Now, we consider the $95\%-$HPD of the posterior predictive densities, for each of the GEV parameters, and for every loo station. From the plots in (a)-(d) of Fig.10, we see that the two models obtain similar posterior predictive densities for each of the GEV parameters. Moreover, we see that both the mean and the bounds of the posterior predictive densities vary between the different loo stations. In particular, we note that the loo stations in plots (b) and (d) corresponding to station 13, respectively, 15 are not that far apart; see Fig.3.1.3. This might explain some of the reduced uncertainty in their posterior predictive densities. Moreover, if we consider the loo station in (a) of Fig.9, corresponding to station 28 in the Top 4 subset. We see that the uncertainty is larger compared to the three other loo stations; see (c)-(d) of Fig.9. This strengthens the previous argument, since station 28 is farthest away from any station in the training data for any of the loo stations. Therefore, we see that the models indeed; make use of dependency to better estimate the GEV parameters at ungauged locations. However, we note that the models struggles to capture some of the local variability.

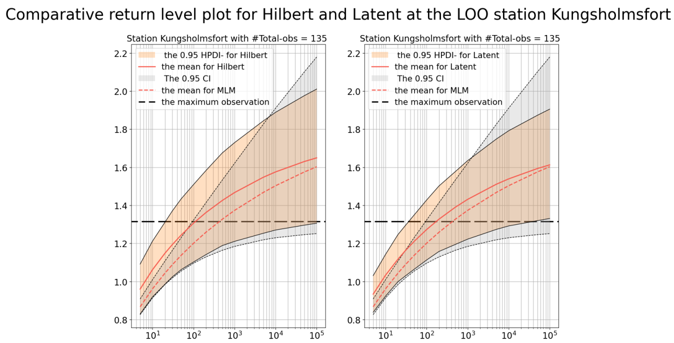
4.3 Parameter and return level surfaces

In this section, we present some of the benefits in using nonparametric (HS)GPs to represent the GEV parameters. In fact, from the formulation of the GEV parameters in Eq.(30)-(32), it follows that we can in theory model each of the random 2-dim GEV parameter surfaces as the graph of a smooth functions over the lat,long coordinate plane covering the region of interest. Thus, we obtain smooth return level surfaces for any return period. Note that in practice we can only obtain values for any finite collection of points. Therefore, the parameter and return level surfaces presented here are merely illustrations of the corresponding theoretically ones.

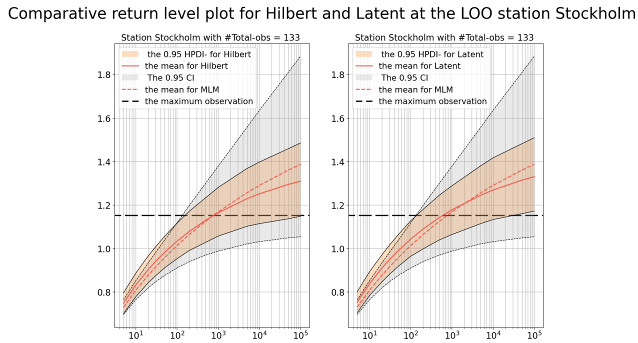
Following the remarks on the nonlinear at the end of Sect.2.1.1, we choose to highlight the scale and the shape parameters of the GEV; see (a)-(b) in Fig.11. For the return levels, we consider both the 10000- and 100000-year return levels; see (a)-(b) in Fig.12. Hilbert and Latent are plotted column-wise and in lexicographical order. Note that the mean value is plotted on the middle row. The lower and upper bounds of the $95\%-$HPD sets are plotted on the first, respectively, last row. The stations pressent in the traning data corresponds to the red dots in each plot.

If we consider the scale and shape parameter surfaces; see (a)-(b) of Fig.11. Then we see that the surfaces of Hilbert and Latent have similarities, for each row. In particular, we observe that both models assigne a negative sign to the shape paramater. From the discussion in Sect.4.2.2, it follows that areas with fewer training stations are not necessarily as accurate compared to the areas with more training stations. Therefore, we do not expect the models to capture well the posterior densities of the

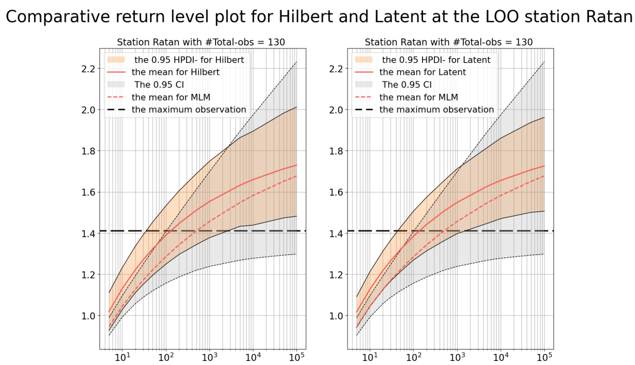
https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025



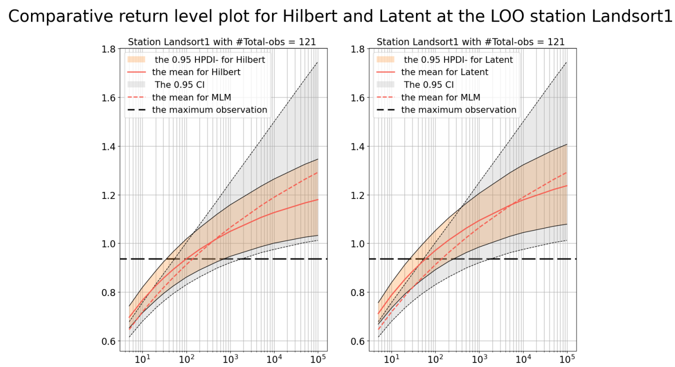
(a) Loo station (28)



(b) Loo station (15)



(c) Loo station (17)



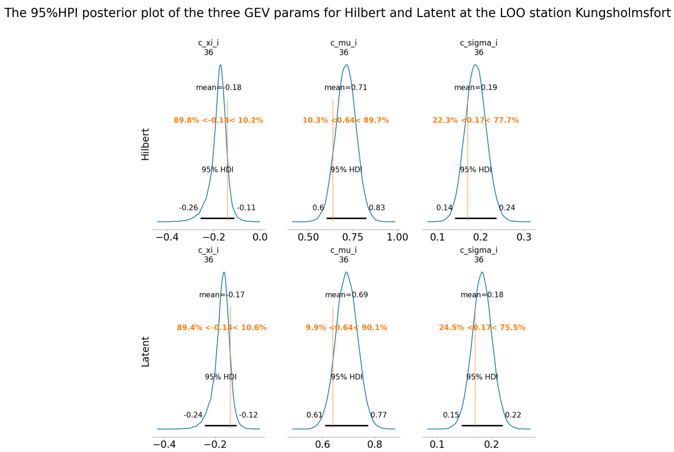
(d) Loo station (31)

Figure 9. Out-of-sample comparative return levels plots, with Hilbert (RHS) and Latent (LHS) in each of the plots (a)-(d). From the above plots, it follows that Hilbert in (a) gives large uncertanty bounds for any return period along the x-axies. In particular, the observation holds independent of model or station in the above plots.

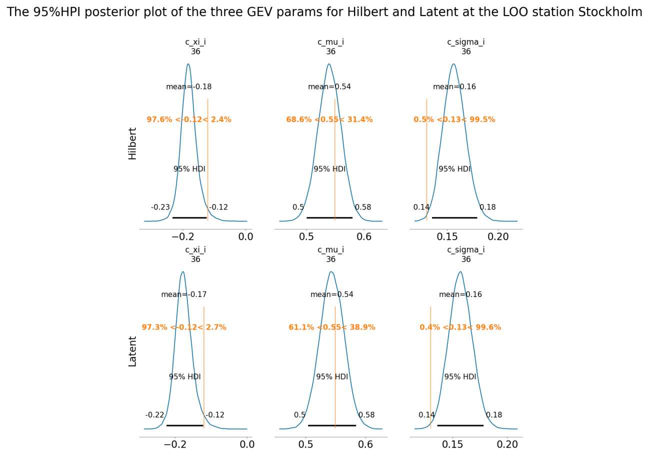
GEV parameters for the southern parts of the Baltic Sea and the western coastline of Norway. However, we note that for the case of the southern parts of the Baltic Sea, this is primarily due to lack of sufficiently long tide gauge stations.

Similarly, if we consider the return level plots in (a)-(b) of Fig.12, and compare the rows in (a) to the ones in (b), then we see that there is only a small difference between the models lower HPD, respectively, mean return level surfaces in (a) and (b). However, if we consider the higher HPD surfaces in (a) to the ones in (b), we see that the range is increased in (b) compared to (a). Recall that for a negative GEV shape parameters, then the support of the distribution is bounded from above, which implies that for increasingly large return periods the corresponding return levels are closer to each other. Therefore, the similarities are in general due to a negative shape parameter, for each of the stations in the training data.

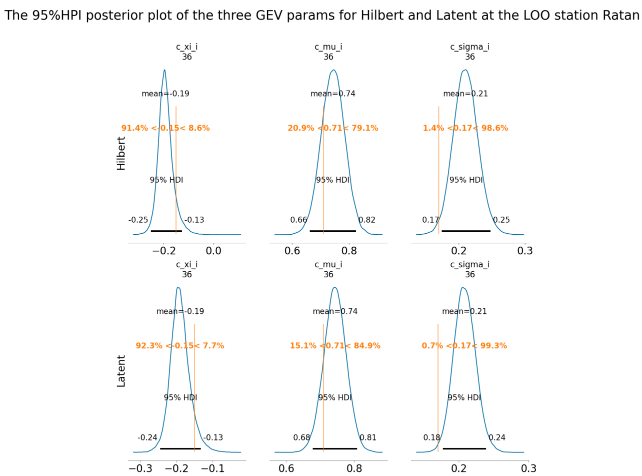
https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025



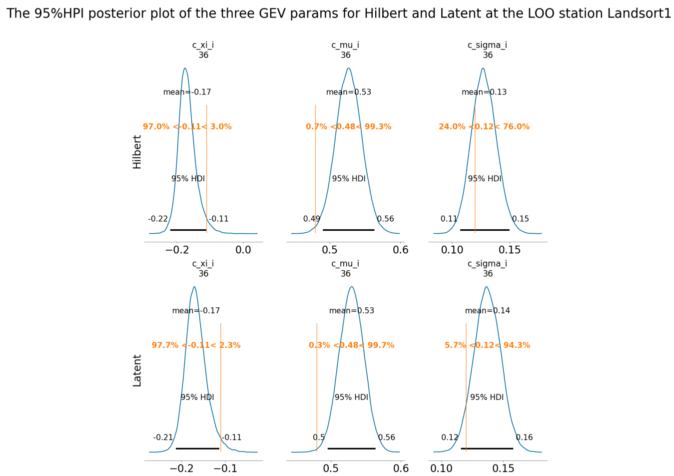
(a) Loo station (28) posterior density plots for Hilbert and Latent



(b) Loo station (15) posterior density plots for Hilbert and Latent



(c) Loo station (17) posterior density plots for Hilbert and Latent



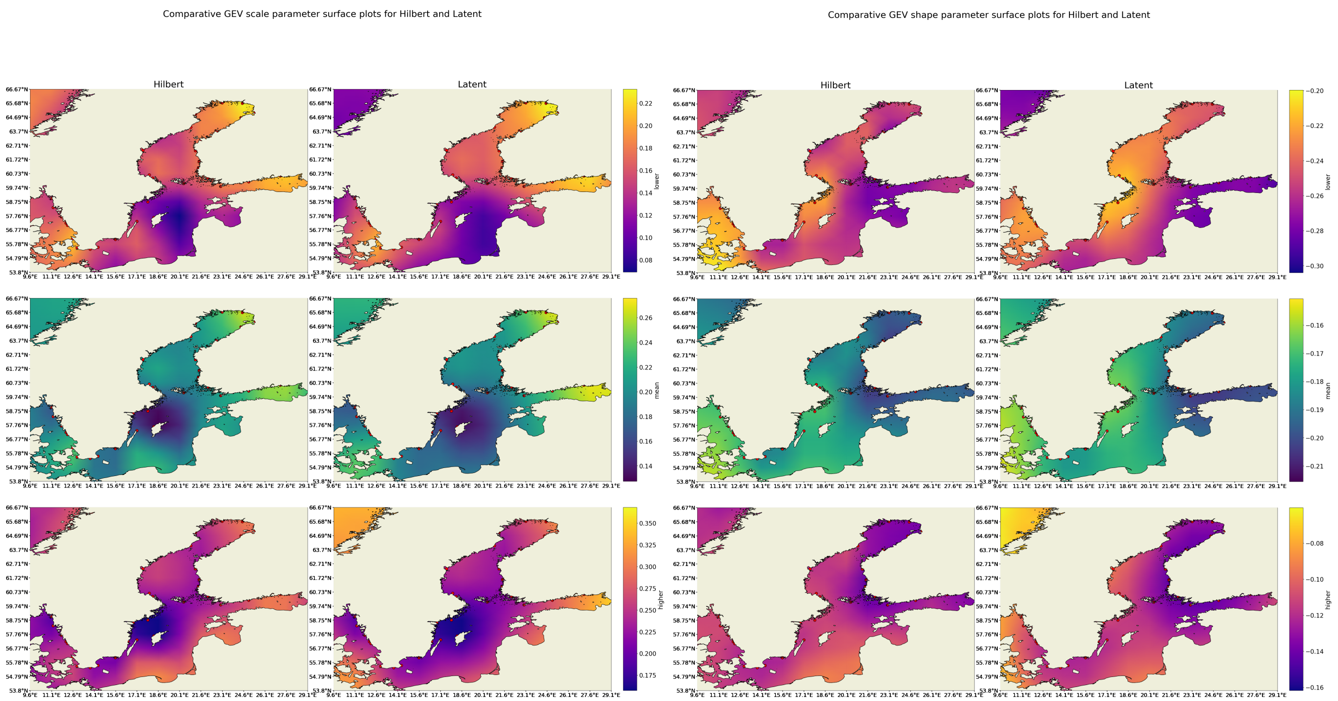
(d) Loo station (31) posterior density plots for Hilbert and Latent

Figure 10. Out-of-sample posterior predictive density plots, with Hilbert and Latent plotted on the first, respectively, second row in each of the plots (a)-(d). From the plots (a)-(b), we see that Hilbert and Latent obtain similar posterior predictive densities. However, in (a) we can see that the 95%-HPD sets of Latent are smaller compared to Hilbert's, thus Latent obtain better return level estimates; see (a) in Fig.9.

5 Conclusions

Several novel models and methods for producing extreme sea level statistics were tested for the Baltic and the North Sea. The best results were given by the Hilbert and Latent model. Both models make use of non-parametric (Hilbert space) Gaussian processes to capture large-scale spatial variability, for each of the GEV parameters. This method is not new; see (Calafat and Marcos 2020), (Räty et al. 2023) and (Davison, Padoan, and Ribatet 2012). The main differences in our work is the combination of coordinate dependent random functions, and random coefficients, where the priors for the coefficients of the mean functions

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025



(a) The scale parameter surface of Hilbert & Latent

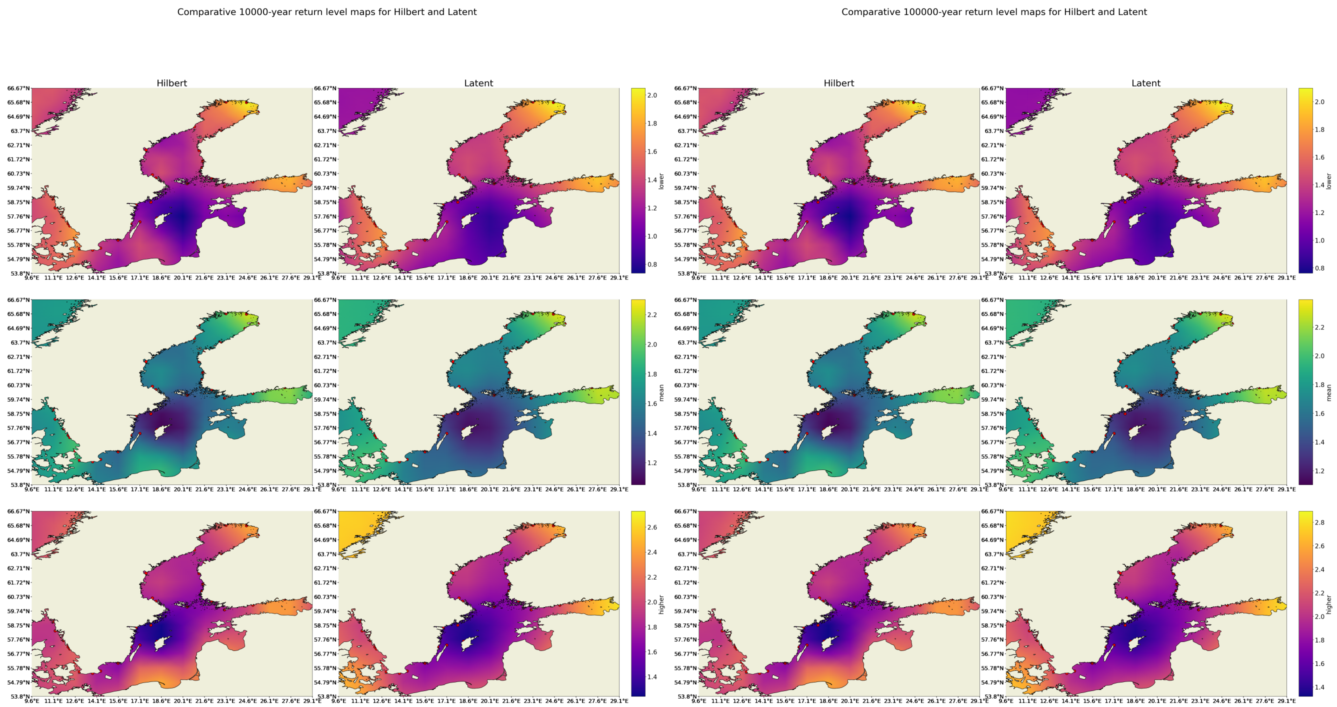
(b) The shape parameter surface Hilbert & Latent

Figure 11. The mean and endpoints of the 95%-HDI posterior surfaces of Hilbert, and Latent, about the scale (a) and shape (b) GEV paramter. From

are specified using the KDEs of the maximum likelihood estimated GEV parameters. In general, the specification of any (hyper) priors adds subjectivity to the model. The approach used here falls under Empirical Bayes methods, and is a variation of the Type II maximum likelihood prior. The benefit of using this method is that it is not subject to any additional data outside of the Block maxima matrix, while at the same time making use of a well-established estimate both in practice and theory. However, one of the main drawbacks with this approach is that its assignees equivalent weights to the MLEs, independent of the number of observations used to obtain the aforementioned estimates. In particular, this implies that lower (higher) values of the shape parameter get assigned equal weights as the higher (lower) values, although the lower (higher) values might be derived from a smaller collection of observations. Hilbert and Latent are both stationary and do not capture any kind of (tail) dependency at the marginal distribution level, for any pair of tide gauges stations (points in general). Therefore, they offer only improvements to the standard univariate approach.

The biggest advantage the new models have over those currently used in coastal spatial planning is that they, through the use of spatial covariances, can infer return levels also for ungaged locations. This, indeed, is a considerable advantage as a vast majority of all coastal municipalities do not have long tide gauge series with high temporal resolution to estimate return levels from. Moreover, the quality suggested by our leave-one-out experiment of the return level estimates for ungauged locations is

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025



(a) Hilbert and Latent 10000-year return level surface

(b) Hilbert and Latent 100000-year return level surface

Figure 12. Mean and upper/lower-HDI posterior surfaces for the $10000, 100000-$year return levels of Hilbert, and Latent.

extremely good. Differences of some centimetres on a 100000 year return level, suggests that from a planners point of view these curves are interchangable, and the important uncertainties will lie in the mean sea level projections for their location, not the extreme sea level component.

Another considerable advantage is the reduction in uncertainty around the central estimate for the return level. Both Hilbert and Latent give much tighter ranges than the MLM estimator, especially on the important right tail. This should give planners more confidence in the estimated return levels and better constrain future coastal spatial planning.

A further advantage is that we found the MLM estimator to fail for some of the Scania gauges, producing shape parameter estimates lower than negative one, a situation the estimator cannot handle. Hilbert and Latent, however, do not suffer from the same deficiency and can thus be used at all gauges.

Lastly, it seems prudent to mention also caveats. The biggest one is likely processes working on small spatial scales that might give rise to considerable sea level extremes, such as wave set-up (Soomere et al. 2020). Such local effects will not be captured by these models, unless they feature in the record of a local tide gauge. This also means that return levels inferred at ungaged locations will not reflect effects of processes working on much smaller scales than the chosen length scales in Eq.(37-

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

36) and/or effects that are not covariant with neighboring stations. Fortunately, our experiments suggest that such effects are seldom important in our region of interest, at least not in locations where tide gauges are placed.

Code and data availability. The code and that needed for reproducing this study are publicly available in Zenodo: https://doi.org/10.5281/zenodo.1502296. The repository contains all of the implemented PyMC models and Python code needed for running the models, along with the raw GESLA-3 dataset. In particular, we have made the results presented in the artical available online, in the form of jupyter notebooks. The raw GESLA-3 datasets used in the analysis are avilable in Zendoo: https://doi.org/10.5281/zenodo.5807461.

Supplement. The supplement related to this article is publicly available online at: https://doi.org/10.5281/zenodo.15019295

Author contributions. SBK and MH designed the study. SBK performed the analyses and prepared the figures. SBK prepared the manuscript with inputs from MH

Competing interests. Seuri Basilio Kuosmanen, and Magnus Hieronymus declares that they have no conflict of interest.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

References

Abril-Pla, Oriol et al. (2023). "PyMC: a modern, and comprehensive probabilistic programming framework in Python". In: *PeerJ Computer Science* 9, e1516.

Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand (2003). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.

Beirlant, Jan et al. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Bücher, Axel and Johan Segers (2017). "On the maximum likelihood estimator for the generalized extreme-value distribution". In: *Extremes* 20, pp. 839–872.

Calafat, F M and M Marcos (2020). "Probabilistic reanalysis of storm surge extremes in Europe". In: *Proceedings of the National Academy of Sciences* 117.4, pp. 1877–1883.

Coles, Stuart et al. (2001). *An introduction to statistical modeling of extreme values*. Vol. 208. Springer.

Davison, Anthony C, Simone A Padoan, and Mathieu Ribatet (2012). "Statistical modeling of spatial extremes". In.

Dombry, Clément (2015). "Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework". In.

Dudley, Richard M (2018). *Real analysis and probability*. Chapman and Hall/CRC.

Frederikse, Thomas et al. (2020). "The causes of sea-level rise since 1900". In: *Nature* 584.7821, pp. 393–397.

Gabry, Jonah et al. (2019). "Visualization in Bayesian workflow". In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 182.2, pp. 389–402.

Gelman, Andrew et al. (2020). "Bayesian workflow". In: *arXiv preprint arXiv:2011.01808*.

Haigh, Ivan D et al. (2023). "GESLA Version 3: A major update to the global higher-frequency sea-level dataset". In: *Geoscience Data Journal* 10.3, pp. 293–314.

Hieronymus, Magnus (2022). "A yearly maximum sea level simulator and its applications: A Stockholm case study". In: *Ambio* 51.5, pp. 1263–1274.

Hieronymus, Magnus, Jenny Hieronymus, and Lars Arneborg (2017). "Sea level modelling in the Baltic and the North Sea: The respective role of different parts of the forcing". In: *Ocean Modelling* 118, pp. 59–72.

Hieronymus, Magnus and Ola Kalén (2020). "Sea-level rise projections for Sweden based on the new IPCC special report: The ocean and cryosphere in a changing climate". In: *Ambio* 49.10, pp. 1587–1600.

Kikstra, Jarmo S et al. (2022). "The IPCC Sixth Assessment Report WGIII climate assessment of mitigation pathways: from emissions to global temperatures". In: *Geoscientific Model Development* 15.24, pp. 9075–9109.

Leadbetter, Malcolm R, Georg Lindgren, and Holger Rootzén (2012). *Extremes and related properties of random sequences and processes*. Springer Science & Business Media.

Palacios-Rodriguez, Fatima, Elena Di Bernardino, and Melina Mailhot (2023). "Smooth copula-based generalized extreme value model and spatial interpolation for extreme rainfall in Central Eastern Canada". In: *Environmetrics* 34.3, e2795.

Räty, O et al. (2023). "Bayesian hierarchical modelling of sea-level extremes in the Finnish coastal region". In: *Natural Hazards and Earth System Sciences* 23.7, pp. 2403–2418.

Robert, Christian P et al. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Vol. 2. Springer.

https://doi.org/10.5194/egusphere-2025-1257
Preprint. Discussion started: 2 April 2025

Solin, Arno and Simo Särkkä (2020). "Hilbert space methods for reduced-rank Gaussian process regression". In: *Statistics and Computing* 30.2, pp. 419–446.

Soomere, T. et al. (2020). "Variability of distributions of wave set-up heights along a shoreline with complicated geometry". In: *Ocean Science* 16.5, pp. 1047–1065. DOI: 10.5194/os-16-1047-2020. URL: https://os.copernicus.org/articles/16/1047/2020/.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry (2017). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC". In: *Statistics and computing* 27, pp. 1413–1432.

Vehtari, Aki, Andrew Gelman, Daniel Simpson, et al. (2021). "Rank-normalization, folding, and localization: An improved R ˆ for assessing convergence of MCMC (with discussion)". In: *Bayesian analysis* 16.2, pp. 667–718.

Vehtari, Aki, Daniel Simpson, et al. (2024). "Pareto Smoothed Importance Sampling". In: *Journal of Machine Learning Research* 25.72, pp. 1–58. URL: http://jmlr.org/papers/v25/19-556.html.

Watanabe, Sumio (2018). *Mathematical theory of Bayesian statistics*. chapman and hall/cRc.

Williams, Christopher KI and Carl Edward Rasmussen (2006). *Gaussian processes for machine learning*. Vol. 2. 3. MIT press Cambridge, MA.