# Supplementary Material: State Space Expectation Propagation

## A. Nomenclature

Vectors: bold lowercase. Matrices: bold uppercase.

| Symbol | Description |
|---|---|
| $n$ | Number of time steps |
| $m$ | Number of latent functions / processes |
| $s$ | State dimensionality |
| $d$ | Output dimensionality |
| $t \in \mathbb{R}$ | Time (input) |
| $\mathbf{r}$ | Space (input, of arbitrary dimension) |
| $k$ | Time index, $t_k$, $k = 1, \ldots, n$ |
| $\mathbf{y}_k \in \mathbb{R}^d$ | Observation (output) |
| $\mathbf{y} \in \mathbb{R}^{d \times n}$ | Collection of outputs, $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$ |
| $\boldsymbol{\theta}$ | Vector of model (hyper)parameters |
| $\kappa(t, t')$ | Covariance function (kernel) |
| $\mathbf{K}(t, t')$ | Multi-output covariance function |
| $\mu(t)$ | Mean function |
| $\boldsymbol{\mu}(t)$ | Multi-output mean function |
| $\boldsymbol{\sigma}_k$ | Measurement noise |
| $\boldsymbol{\Sigma}_k$ | Measurement noise covariance |
| $f(t) : \mathbb{R} \to \mathbb{R}$ | Latent function (Gaussian process) |
| $\mathbf{f}(t) : \mathbb{R} \to \mathbb{R}^m$ | Vector of latent functions, $\mathbf{f}_k = \mathbf{f}(t_k)$ |
| $\mathbf{f} \in \mathbb{R}^{m \times n}$ | Collection of latents, $(\mathbf{f}(t_1), \ldots, \mathbf{f}(t_n))$ |
| $\mathbf{H}_k \in \mathbb{R}^{m \times s}$ | State→function mapping |
| $\mathbf{h}(\mathbf{f}_k, \boldsymbol{\sigma}_k)$ | Measurement model $(\mathbb{R}^m, \mathbb{R}^d) \to \mathbb{R}^d$ |
| $\mathbf{x}(t) : \mathbb{R} \to \mathbb{R}^s$ | State vector, $\mathbf{f}(t) = \mathbf{H}_k \mathbf{x}(t)$ |
| $\mathbf{x}_k \in \mathbb{R}^s$ | State variable, $\mathbf{x}_k = \mathbf{x}(t_k) \sim \mathrm{N}(\mathbf{m}_k, \mathbf{P}_k)$ |
| $\mathbf{F} \in \mathbb{R}^{s \times s}$ | Feedback matrix (continuous) |
| $\mathbf{L} \in \mathbb{R}^{s \times v}$ | Noise effect matrix (continuous) |
| $\mathbf{Q}_\mathrm{c} \in \mathbb{R}^{v \times v}$ | White noise spectral density (continuous) |
| $\mathbf{A}_k \in \mathbb{R}^{s \times s}$ | Dynamic model (discrete) |
| $\mathbf{q}_k \in \mathbb{R}^s$ | State space process noise (discrete) |
| $\mathbf{Q}_k \in \mathbb{R}^{s \times s}$ | Process noise covariance (discrete) |
| $\mathbf{P}_\infty \in \mathbb{R}^{s \times s}$ | Stationary state covariance (prior) |
| $\mathbf{m}_k \in \mathbb{R}^{s \times 1}$ | State mean |
| $\mathbf{P}_k \in \mathbb{R}^{s \times s}$ | State covariance |
| $\mathbf{K}_k \in \mathbb{R}^{s \times d}$ | Kalman gain |
| $\mathbf{G}_k \in \mathbb{R}^{s \times s}$ | Smoother gain |
| $\mathbf{J}_\mathbf{f} \in \mathbb{R}^{d \times m}$ | Jacobian of $\mathbf{h}$ w.r.t $\mathbf{f}_k$ |
| $\mathbf{J}_{\boldsymbol{\sigma}} \in \mathbb{R}^{d \times d}$ | Jacobian of $\mathbf{h}$ w.r.t $\boldsymbol{\sigma}_k$ |
| $\alpha$ | EP power / fraction |
| $\mathcal{L}_k$ | log-normaliser of true posterior update |
| $q_k^{\text{site}}(\mathbf{f}_k)$ | EP site (approximate likelihood) |
| | $q_k^{\text{site}}(\mathbf{f}_k) \sim \mathrm{N}(\boldsymbol{\mu}_k^{\text{site}}, \boldsymbol{\Sigma}_k^{\text{site}})$ |
| $q_k^{\text{cav}}(\mathbf{f}_k)$ | EP cavity (leave-one-out posterior) |
| | $q_k^{\text{cav}}(\mathbf{f}_k) \sim \mathrm{N}(\boldsymbol{\mu}_k^{\text{cav}}, \boldsymbol{\Sigma}_k^{\text{cav}})$ |

## B. Gaussian Filtering

Given observation model $p(\mathbf{y}_k \,|\, \mathbf{f}_k) = \mathrm{N}(\mathbf{y}_k \,|\, \mathbf{f}_k, \boldsymbol{\Sigma}_k)$ for $\mathbf{f}_k = \mathbf{H}_k \mathbf{x}_k$, along with current filter predictions $p(\mathbf{x}_k \,|\, \mathbf{y}_{1:k-1}) = \mathrm{N}(\mathbf{x}_k \,|\, \mathbf{m}_k^{\text{pred}}, \mathbf{P}_k^{\text{pred}})$, the Kalman filter update equations are,

$$
\begin{aligned}
\boldsymbol{\mu}_k &= \mathbf{H}_k \mathbf{m}_k^{\text{pred}}, \\
\mathbf{S}_k &= \mathbf{H}_k \mathbf{P}_k^{\text{pred}} \mathbf{H}_k^\top + \boldsymbol{\Sigma}_k, \\
\mathbf{C}_k &= \mathbf{P}_k^{\text{pred}} \mathbf{H}^\top, \\
\mathbf{K}_k &= \mathbf{C}_k \mathbf{S}_k^{-1}, \\
\mathbf{m}_k &= \mathbf{m}_k^{\text{pred}} + \mathbf{K}_k (\mathbf{y}_k - \boldsymbol{\mu}_k), \\
\mathbf{P}_k &= \mathbf{P}_k^{\text{pred}} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top.
\end{aligned}
\tag{22}
$$

For nonlinear measurement models, $\mathbf{y}_k = \mathbf{h}(\mathbf{f}_k, \boldsymbol{\sigma}_k)$, letting $\boldsymbol{\mu}_k^{\text{cav}} = \mathbf{H}_k \mathbf{m}_k^{\text{pred}}$ and $\boldsymbol{\Sigma}_k^{\text{cav}} = \mathbf{H}_k \mathbf{P}_k^{\text{pred}} \mathbf{H}_k^\top$, the statistical linear regression equations for the general Gaussian filtering methods are,

$$
\begin{aligned}
\boldsymbol{\mu}_k &= \iint \mathbf{h}(\mathbf{f}_k, \boldsymbol{\sigma}_k) \\
&\qquad \times \mathrm{N}(\mathbf{f}_k \,|\, \boldsymbol{\mu}_k^{\text{cav}}, \boldsymbol{\Sigma}_k^{\text{cav}}) \mathrm{N}(\boldsymbol{\sigma}_k \,|\, \mathbf{0}, \boldsymbol{\Sigma}_k) \, \mathrm{d}\mathbf{f}_k \, \mathrm{d}\boldsymbol{\sigma}_k, \\
\mathbf{S}_k &= \iint (\mathbf{h}(\mathbf{f}_k, \boldsymbol{\sigma}_k) - \boldsymbol{\mu}_k)(\mathbf{h}(\mathbf{f}_k, \boldsymbol{\sigma}_k) - \boldsymbol{\mu}_k)^\top \\
&\qquad \times \mathrm{N}(\mathbf{f}_k \,|\, \boldsymbol{\mu}_k^{\text{cav}}, \boldsymbol{\Sigma}_k^{\text{cav}}) \mathrm{N}(\boldsymbol{\sigma}_k \,|\, \mathbf{0}, \boldsymbol{\Sigma}_k) \, \mathrm{d}\mathbf{f}_k \, \mathrm{d}\boldsymbol{\sigma}_k, \\
\mathbf{C}_k &= \iint (\mathbf{f}_k - \boldsymbol{\mu}_k^{\text{cav}})(\mathbf{h}(\mathbf{f}_k, \boldsymbol{\sigma}_k) - \boldsymbol{\mu}_k)^\top \\
&\qquad \times \mathrm{N}(\mathbf{f}_k \,|\, \boldsymbol{\mu}_k^{\text{cav}}, \boldsymbol{\Sigma}_k^{\text{cav}}) \mathrm{N}(\boldsymbol{\sigma}_k \,|\, \mathbf{0}, \boldsymbol{\Sigma}_k) \, \mathrm{d}\mathbf{f}_k \, \mathrm{d}\boldsymbol{\sigma}_k.
\end{aligned}
\tag{23}
$$

Note that in the additive noise case, $\mathbf{h}(\mathbf{f}_k, \boldsymbol{\sigma}_k) = \tilde{\mathbf{h}}(\mathbf{f}_k) + \boldsymbol{\sigma}_k$, these can be simplified to,

$$
\begin{aligned}
\boldsymbol{\mu}_k &= \int \tilde{\mathbf{h}}(\mathbf{f}_k) \mathrm{N}(\mathbf{f}_k \,|\, \boldsymbol{\mu}_k^{\text{cav}}, \boldsymbol{\Sigma}_k^{\text{cav}}) \, \mathrm{d}\mathbf{f}_k, \\
\mathbf{S}_k &= \int \left[ (\tilde{\mathbf{h}}(\mathbf{f}_k) - \boldsymbol{\mu}_k)(\tilde{\mathbf{h}}(\mathbf{f}_k) - \boldsymbol{\mu}_k)^\top + \mathrm{Cov}[\mathbf{y}_k \,|\, \mathbf{f}_k] \right] \\
&\qquad \times \mathrm{N}(\mathbf{f}_k \,|\, \boldsymbol{\mu}_k^{\text{cav}}, \boldsymbol{\Sigma}_k^{\text{cav}}) \, \mathrm{d}\mathbf{f}_k, \\
\mathbf{C}_k &= \int (\mathbf{f}_k - \boldsymbol{\mu}_k^{\text{cav}})(\tilde{\mathbf{h}}(\mathbf{f}_k) - \boldsymbol{\mu}_k)^\top \mathrm{N}(\mathbf{f}_k \,|\, \boldsymbol{\mu}_k^{\text{cav}}, \boldsymbol{\Sigma}_k^{\text{cav}}) \, \mathrm{d}\mathbf{f}_k.
\end{aligned}
\tag{24}
$$

for $\boldsymbol{\sigma}_k \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_k = \mathrm{Cov}[\mathbf{y}_k \,|\, \mathbf{f}_k])$. Note that we include the case where $\boldsymbol{\Sigma}_k$ is a nonlinear function of $\mathbf{f}_k$, which occurs in our approximations to discrete likelihoods presented in App. I. Here we have used $\tilde{\mathbf{h}}(\mathbf{f}_k) = \mathbb{E}[\mathbf{y}_k \,|\, \mathbf{f}_k]$.

## C. Closed-form Site Updates in Sec. 3.2

Here we derive in full the closed form site updates after analytical linearisation in Sec. 3.2. Plugging the derivatives from Eq. (15) into the updates in Eq. (10) we get,

$$\begin{aligned}\boldsymbol{\mu}_k^{\text{site}} &= \boldsymbol{\mu}_k^{\text{cav}} + \left(\mathbf{J}_{\mathbf{f}_k}^\top \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1} \mathbf{J}_{\mathbf{f}_k}^\top \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{v}_k, \\ \boldsymbol{\Sigma}_k^{\text{site}} &= -\alpha \boldsymbol{\Sigma}_k^{\text{cav}} + \left(\mathbf{J}_{\mathbf{f}_k}^\top \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1},\end{aligned} \tag{25}$$

where $\mathbf{v}_k = \mathbf{y}_k - \mathbf{h}(\boldsymbol{\mu}_k^{\text{cav}}, \mathbf{0})$. By the matrix inversion lemma, and letting $\mathbf{R}_k = \mathbf{J}_{\boldsymbol{\sigma}_k} \boldsymbol{\Sigma}_k \mathbf{J}_{\boldsymbol{\sigma}_k}^\top$,

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_k^{-1} &= \mathbf{R}_k^{-1} - \\ &\mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k} \left((\alpha \boldsymbol{\Sigma}_k^{\text{cav}})^{-1} + \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1} \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1},\end{aligned} \tag{26}$$

so that

$$\mathbf{J}_{\mathbf{f}_k}^\top \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{J}_{\mathbf{f}_k} = \mathbf{W}_k - \mathbf{W}_k \left((\alpha \boldsymbol{\Sigma}_k^{\text{cav}})^{-1} + \mathbf{W}_k\right)^{-1} \mathbf{W}_k, \tag{27}$$

where $\mathbf{W}_k = \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}$. Applying the matrix inversion lemma for a second time we obtain

$$\begin{aligned}&\left(\mathbf{J}_{\mathbf{f}_k}^\top \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1} \\ &= \mathbf{W}_k^{-1} - \mathbf{W}_k^{-1} \mathbf{W}_k \Bigg( \mathbf{W}_k \mathbf{W}_k^{-1} \mathbf{W}_k \\ &\qquad - \left((\alpha \boldsymbol{\Sigma}_k^{\text{cav}})^{-1} + \mathbf{W}_k\right) \Bigg)^{-1} \mathbf{W}_k \mathbf{W}_k^{-1} \\ &= \mathbf{W}_k^{-1} + \alpha \boldsymbol{\Sigma}_k^{\text{cav}} \\ &= \left(\mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1} + \alpha \boldsymbol{\Sigma}_k^{\text{cav}}.\end{aligned} \tag{28}$$

We can also write

$$\begin{aligned}\left(\mathbf{J}_{\mathbf{f}_k}^\top \hat{\boldsymbol{\Sigma}}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1} \mathbf{J}_{\mathbf{f}_k}^\top \hat{\boldsymbol{\Sigma}}_k^{-1} &= \left(\left(\mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1} + \alpha \boldsymbol{\Sigma}_k^{\text{cav}}\right) \\ &\times \mathbf{J}_{\mathbf{f}_k}^\top \left(\mathbf{R}_k + \alpha \mathbf{J}_{\mathbf{f}_k} \boldsymbol{\Sigma}_k^{\text{cav}} \mathbf{J}_{\mathbf{f}_k}^\top\right)^{-1}.\end{aligned} \tag{29}$$

Together the above calculations give the approximate site mean and covariance as

$$\begin{aligned}\boldsymbol{\Sigma}_k^{\text{site}} &= \left(\mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1}, \\ \boldsymbol{\mu}_k^{\text{site}} &= \boldsymbol{\mu}_k^{\text{cav}} + \\ &\left(\boldsymbol{\Sigma}_k^{\text{site}} + \alpha \boldsymbol{\Sigma}_k^{\text{cav}}\right) \mathbf{J}_{\mathbf{f}_k}^\top \left(\mathbf{R}_k + \alpha \mathbf{J}_{\mathbf{f}_k} \boldsymbol{\Sigma}_k^{\text{cav}} \mathbf{J}_{\mathbf{f}_k}^\top\right)^{-1} \mathbf{v}_k.\end{aligned} \tag{30}$$

## D. Analytical Linearisation in EP ($\alpha = 1$) Results in an Iterated Version of the EKF

Here we prove the result given in Sec. 3.2: a single pass of the proposed EP-style algorithm with analytical linearisation (*i.e.* a first order Taylor series approximation) is exactly equivalent to the EKF. Plugging the closed form site updates, Eq. (16), with $\alpha = 1$ (since the filter predictions can be interpreted as the cavity with the *full* site removed), into our modified Kalman filter update equations, Eq. (11), we get a new set of Kalman updates in which the latent noise terms are determined by scaling the observation noise with the Jacobian of the state. Crucially, on the first forward pass the Kalman prediction is used as the cavity such that $\boldsymbol{\Sigma}_k^{\text{cav}} = \mathbf{H}_k \mathbf{P}_k^{\text{pred}} \mathbf{H}_k^\top$:

$$\begin{aligned}\mathbf{S}_k &= \boldsymbol{\Sigma}_k^{\text{cav}} + \left(\mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1}, \\ \mathbf{K}_k &= \mathbf{P}_k^{\text{pred}} \mathbf{H}_k^\top \mathbf{S}_k^{-1}, \\ \mathbf{m}_k &= \mathbf{m}_k^{\text{pred}} + \mathbf{K}_k \mathbf{S}_k \mathbf{J}_{\mathbf{f}_k}^\top \left(\mathbf{R}_k + \mathbf{J}_{\mathbf{f}_k} \boldsymbol{\Sigma}_k^{\text{cav}} \mathbf{J}_{\mathbf{f}_k}^\top\right)^{-1} \mathbf{v}_k, \\ \mathbf{P}_k &= \mathbf{P}_k^{\text{pred}} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top.\end{aligned} \tag{31}$$

where $\mathbf{R}_k = \mathbf{J}_{\boldsymbol{\sigma}_k} \boldsymbol{\Sigma}_k \mathbf{J}_{\boldsymbol{\sigma}_k}^\top$. This can be rewritten to explicitly show that there are two innovation covariance terms, $\mathbf{S}_k$ and $\hat{\mathbf{S}}_k$, which act on the state mean and covariance separately:

**Linearised update step:**

$$\begin{aligned}\hat{\mathbf{S}}_k &= \boldsymbol{\Sigma}_k^{\text{cav}} + \left(\mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1}, \\ \mathbf{S}_k &= \mathbf{J}_{\mathbf{f}_k} \boldsymbol{\Sigma}_k^{\text{cav}} \mathbf{J}_{\mathbf{f}_k}^\top + \mathbf{R}_k, \\ \hat{\mathbf{K}}_k &= \mathbf{P}_k^{\text{pred}} \mathbf{H}_k \hat{\mathbf{S}}_k^{-1}, \\ \mathbf{K}_k &= \mathbf{P}_k^{\text{pred}} \mathbf{H}_k^\top \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{S}_k^{-1}, \\ \mathbf{m}_k &= \mathbf{m}_k^{\text{pred}} + \mathbf{K}_k \mathbf{v}_k, \\ \mathbf{P}_k &= \mathbf{P}_k^{\text{pred}} - \hat{\mathbf{K}}_k \hat{\mathbf{S}}_k \hat{\mathbf{K}}_k^\top.\end{aligned} \tag{32}$$

Now we calculate the inverse of $\hat{\mathbf{S}}_k$:

$$\begin{aligned}\hat{\mathbf{S}}_k^{-1} &= \left(\boldsymbol{\Sigma}_k^{\text{cav}} + \left(\mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1}\right)^{-1} \\ &= \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k} - \\ &\quad \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k} \left(\boldsymbol{\Sigma}_k^{\text{cav}\,-1} + \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1} \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\end{aligned} \tag{33}$$

and the inverse of $\mathbf{S}_k$:

$$\begin{aligned}\mathbf{S}_k^{-1} &= \left(\mathbf{J}_{\mathbf{f}_k} \boldsymbol{\Sigma}_k^{\text{cav}} \mathbf{J}_{\mathbf{f}_k}^\top + \mathbf{R}_k\right)^{-1} \\ &= \mathbf{R}_k^{-1} - \\ &\quad \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k} \left(\boldsymbol{\Sigma}_k^{\text{cav}\,-1} + \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1} \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{R}_k^{-1}\end{aligned} \tag{34}$$

which shows that

$$\hat{\mathbf{S}}_k^{-1} = \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{S}_k^{-1} \mathbf{J}_{\mathbf{f}_k}, \tag{35}$$

and hence, recalling that $\mathbf{R}_k = \mathbf{J}_{\mathbf{r}_k} \boldsymbol{\Sigma}_k \mathbf{J}_{\mathbf{r}_k}^\top$, Eq. (32) simplifies to give exactly the extended Kalman filter updates:

**EKF update step:**

$$\begin{aligned}
\mathbf{S}_k &= \mathbf{J}_{\mathbf{f}_k} \mathbf{H}_k \mathbf{P}_k^{\text{pred}} \mathbf{H}_k^\top \mathbf{J}_{\mathbf{f}_k}^\top + \mathbf{J}_{\boldsymbol{\sigma}_k} \boldsymbol{\Sigma}_k \mathbf{J}_{\boldsymbol{\sigma}_k}^\top, \\
\mathbf{K}_k &= \mathbf{P}_k^{\text{pred}} \mathbf{H}_k^\top \mathbf{J}_{\mathbf{f}_k}^\top \mathbf{S}_k^{-1}, \\
\mathbf{m}_k &= \mathbf{m}_k^{\text{pred}} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{h}(\mathbf{H}_k \mathbf{m}_k^{\text{pred}}, \mathbf{0})), \\
\mathbf{P}_k &= \mathbf{P}_k^{\text{pred}} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top.
\end{aligned} \tag{36}$$

## E. General Gaussian Filter Site Updates in Sec. 3.3

Here we derive in full the site updates after statistical linear regression in Sec. 3.3. The Gaussian likelihood approximation results in,

$$\begin{aligned}
\mathcal{L}_k &= \log \mathbb{E}_{q_k^{\text{cav}}} \left[ \mathrm{N}^\alpha(\mathbf{y}_k \,|\, \boldsymbol{\mu}_k + \mathbf{C}_k^\top (\boldsymbol{\Sigma}_k^{\text{cav}})^{-1} (\mathbf{f}_k - \boldsymbol{\mu}_k^{\text{cav}}), \mathbf{R}_k) \right] \\
&= c + \log \mathrm{N}\big(\mathbf{y}_k \,|\, \boldsymbol{\mu}_k, \alpha^{-1} \tilde{\boldsymbol{\Sigma}}_k\big),
\end{aligned} \tag{37}$$

where $q_k^{\text{cav}} = \mathrm{N}(\mathbf{f}_k \,|\, \boldsymbol{\mu}_k^{\text{cav}}, \boldsymbol{\Sigma}_k^{\text{cav}})$, $\mathbf{R}_k = \mathbf{S}_k - \mathbf{C}_k^\top (\boldsymbol{\Sigma}_k^{\text{cav}})^{-1} \mathbf{C}_k$ and $\tilde{\boldsymbol{\Sigma}}_k = \mathbf{R}_k + \alpha \mathbf{C}_k^\top (\boldsymbol{\Sigma}_k^{\text{cav}})^{-1} \mathbf{C}_k$ for $\boldsymbol{\mu}_k$, $\mathbf{S}_k$ and $\mathbf{C}_k$ given in Eq. (23) with $\boldsymbol{\mu}_k^{\text{cav}} = \mathbf{H}_k \mathbf{m}_k^{\text{pred}}$, $\boldsymbol{\Sigma}_k^{\text{cav}} = \mathbf{H}_k \mathbf{P}_k^{\text{pred}} \mathbf{H}_k^\top$. Taking the derivatives of this log-Gaussian w.r.t. the cavity mean, we get

$$\begin{aligned}
\boldsymbol{\nabla} \mathcal{L}_k &= \frac{\partial \mathcal{L}_k}{\partial \boldsymbol{\mu}_k^{\text{cav}}} = \alpha \boldsymbol{\Omega}_k^\top \tilde{\boldsymbol{\Sigma}}_k^{-1} \mathbf{v}_k, \\
\boldsymbol{\nabla}^2 \mathcal{L}_k &= \frac{\partial^2 \mathcal{L}_k}{\partial \boldsymbol{\mu}_k^{\text{cav}} \partial (\boldsymbol{\mu}_k^{\text{cav}})^\top} = -\alpha \boldsymbol{\Omega}_k^\top \tilde{\boldsymbol{\Sigma}}_k^{-1} \boldsymbol{\Omega}_k,
\end{aligned} \tag{38}$$

where $\mathbf{v}_k = \mathbf{y}_k - \boldsymbol{\mu}_k$ and

$$\begin{aligned}
\boldsymbol{\Omega}_k &= \frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{\mu}_k^{\text{cav}}} \\
&= \iint \mathbf{h}(\mathbf{f}_k, \boldsymbol{\sigma}_k) (\boldsymbol{\Sigma}_k^{\text{cav}})^{-1} (\mathbf{f}_k - \boldsymbol{\mu}_k^{\text{cav}}) \mathrm{N}(\mathbf{f}_k \,|\, \boldsymbol{\mu}_k^{\text{cav}}, \boldsymbol{\Sigma}_k^{\text{cav}}) \\
&\qquad \times \mathrm{N}(\boldsymbol{\sigma}_k \,|\, \mathbf{0}, \boldsymbol{\Sigma}_k) \, \mathrm{d}\mathbf{f}_k \, \mathrm{d}\boldsymbol{\sigma}_k.
\end{aligned} \tag{39}$$

As in Sec. 3.2, to ensure consistency we have assumed here that the derivative of $\tilde{\boldsymbol{\Sigma}}_k$ is zero, despite the fact that it depends on $\boldsymbol{\mu}_k^{\text{cav}}$.

Plugging the derivatives from Eq. (38) into the updates in Eq. (10) we get,

$$\begin{aligned}
\boldsymbol{\mu}_k^{\text{site}} &= \boldsymbol{\mu}_k^{\text{cav}} + \left( \boldsymbol{\Omega}_k^\top \tilde{\boldsymbol{\Sigma}}_k^{-1} \boldsymbol{\Omega}_k \right)^{-1} \boldsymbol{\Omega}_k^\top \tilde{\boldsymbol{\Sigma}}_k^{-1} \mathbf{v}_k, \\
\boldsymbol{\Sigma}_k^{\text{site}} &= -\alpha \boldsymbol{\Sigma}_k^{\text{cav}} + \left( \boldsymbol{\Omega}_k^\top \tilde{\boldsymbol{\Sigma}}_k^{-1} \boldsymbol{\Omega}_k \right)^{-1}.
\end{aligned} \tag{40}$$

## F. Avoiding Numerical Issues When Computing the Cavity

Computing the cavity distribution in Eq. (12) involves the subtraction of two PSD covariance matrices. The result is not guaranteed to be PSD and not guaranteed to be invertible, which can lead to numerical issues. If $\mathbf{f}_k$ is one-dimensional, then no such issue occurs. In the higher-dimensional case issues can be avoided by discarding the cross-covariances such that Eq. (12) involves only element-wise subtraction of scalars. If using cubature to perform moment matching / linearisation, then this results in a loss of accuracy. However, for the Taylor series approximations (EKF / EKS / EEP) the cross-covariances are discarded anyway.

An alternative approach which does not trade off accuracy is to instead compute the cavity by taking the product of the forward and backward filtering distributions, an approach known as two-filter smoothing (Särkkä, 2013), and then include a fraction $(1 - \alpha)$ of the local site. This method only involves *products* of PSD matrices which is more numerically stable. We did not implement this approach here.

## G. Marginal Likelihood Calculation During Filtering

The marginal likelihood, $p(\mathbf{y} \,|\, \boldsymbol{\theta})$, is used as an optimisation objective for hyperparameter learning. The marginal likelihood can be written as a product of conditional terms (dropping the dependence on $\boldsymbol{\theta}$ for notational convenience),

$$p(\mathbf{y}) = p(\mathbf{y}_1) \, p(\mathbf{y}_2 \,|\, \mathbf{y}_1) \, p(\mathbf{y}_3 \,|\, \mathbf{y}_{1:2}) \prod_{k=4}^{n} p(\mathbf{y}_k \,|\, \mathbf{y}_{1:k-1}). \tag{41}$$

Each term can be computed via numerical integration during the Kalman filter by noticing that,

$$\begin{aligned}
p(\mathbf{y}_k \,|\, \mathbf{y}_{1:k-1}) &= \int p(\mathbf{y}_k \,|\, \mathbf{x}_k, \mathbf{y}_{1:k-1}) p(\mathbf{x}_k \,|\, \mathbf{y}_{1:k-1}) \, \mathrm{d}\mathbf{x}_k \\
&= \int p(\mathbf{y}_k \,|\, \mathbf{f}_k = \mathbf{H}\mathbf{x}_k) p(\mathbf{x}_k \,|\, \mathbf{y}_{1:k-1}) \, \mathrm{d}\mathbf{x}_k.
\end{aligned} \tag{42}$$

The first component in the integral is the likelihood, and the second term is the forward filter prediction.

The Taylor series methods (EKF / EKS / EEP) aim to avoid numerical integration, and hence use an alternative approximation to the marginal likelihood based on the linearised model, as shown in Algorithm 2.

## H. The EEP Algorithm

**Algorithm 2** EEP: Extended Expectation Propagation, a globally iterated Extended Kalman filter with power EP-style updates such that linearisation is performed w.r.t. the cavity mean.

**Input:** $\{t_k, \mathbf{y}_k\}_{k=1}^n$, $\mathbf{A}_k$, $\mathbf{Q}_k$, $\mathbf{P}_\infty$, $\mathbf{\Sigma}_k$ — data, discrete state space model and obs. noise
$\mathbf{h}$, $\mathbf{H}_k$, $\mathbf{J}_\mathbf{f}$, $\mathbf{J}_{\boldsymbol{\sigma}}$, $\alpha$ — measurement model, Jacobian and EP power
$\mathbf{m}_0 \leftarrow \mathbf{0}$, $\mathbf{P}_0 \leftarrow \mathbf{P}_\infty$, $\mathbf{e}_{1:n} = \mathbf{0}$ — initial state
**while** not converged **do** — iterated EP-style loop
  **for** $k = 1$ **to** $n$ **do** — forward pass (FILTERING)
    $\mathbf{m}_k \leftarrow \mathbf{A}_k\,\mathbf{m}_{k-1}$ — predict mean
    $\mathbf{P}_k \leftarrow \mathbf{A}_k\,\mathbf{P}_{k-1}\,\mathbf{A}_k^\top + \mathbf{Q}_k$ — predict covariance
    **if** has label $\mathbf{y}_k$ **then**
      $\mathbf{\Sigma}_k^{\text{cav}} \leftarrow \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^\top$ — predict = forward cavity
      $\boldsymbol{\mu}_k^{\text{cav}} \leftarrow \mathbf{H}_k \mathbf{m}_k$
      $\mathbf{v}_k \leftarrow \mathbf{y}_k - \mathbf{h}(\boldsymbol{\mu}_k^{\text{cav}}, \mathbf{0})$ — residual
      $\mathbf{J}_{\mathbf{f}_k} \leftarrow \mathbf{J}_\mathbf{f}|_{\boldsymbol{\mu}_k^{\text{cav}}, \mathbf{0}}$; $\mathbf{J}_{\boldsymbol{\sigma}_k} \leftarrow \mathbf{J}_{\boldsymbol{\sigma}}|_{\boldsymbol{\mu}_k^{\text{cav}}, \mathbf{0}}$ — evaluate Jacobians
      **if** first iteration **then**
        $\mathbf{\Sigma}_k^{\text{site}} \leftarrow \left(\mathbf{J}_{\mathbf{f}_k}^\top \left(\mathbf{J}_{\boldsymbol{\sigma}_k} \mathbf{\Sigma}_k \mathbf{J}_{\boldsymbol{\sigma}_k}^\top\right)^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1}$ — match moments to get site covariance...
        $\boldsymbol{\mu}_k^{\text{site}} \leftarrow \boldsymbol{\mu}_k^{\text{cav}} + \left(\mathbf{\Sigma}_k^{\text{site}} + \mathbf{\Sigma}_k^{\text{cav}}\right) \mathbf{J}_{\mathbf{f}_k}^\top \left(\mathbf{J}_{\boldsymbol{\sigma}_k} \mathbf{\Sigma}_k \mathbf{J}_{\boldsymbol{\sigma}_k}^\top + \mathbf{J}_{\mathbf{f}_k} \mathbf{\Sigma}_k^{\text{cav}} \mathbf{J}_{\mathbf{f}_k}^\top\right)^{-1} \mathbf{v}_k$ — and site mean ($\alpha = 1$)
      **end if**
      $\mathbf{S}_k \leftarrow \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^\top + \mathbf{\Sigma}_k^{\text{site}}$ — innovation
      $\mathbf{K}_k \leftarrow \mathbf{P}_k \mathbf{H}_k^\top \mathbf{S}_k^{-1}$ — Kalman gain
      $\mathbf{m}_k \leftarrow \mathbf{m}_k + \mathbf{K}_k(\boldsymbol{\mu}_k^{\text{site}} - \boldsymbol{\mu}_k^{\text{cav}})$ — update mean
      $\mathbf{P}_k \leftarrow \mathbf{P}_k - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top$ — update covariance
      $\mathbf{E}_k \leftarrow \mathbf{J}_{\boldsymbol{\sigma}_k} \mathbf{\Sigma}_k \mathbf{J}_{\boldsymbol{\sigma}_k}^\top + \mathbf{J}_{\mathbf{f}_k} \mathbf{\Sigma}_k^{\text{cav}} \mathbf{J}_{\mathbf{f}_k}^\top$
      $\mathbf{e}_k \leftarrow \frac{1}{2} \log |2\pi \mathbf{E}_k| + \frac{1}{2} \mathbf{v}_k^\top \mathbf{E}_k^{-1} \mathbf{v}_k$ — energy
    **end if**
  **end for**
  **for** $k = n - 1$ **to** $1$ **do** — backward pass (SMOOTHING)
    $\mathbf{G}_k \leftarrow \mathbf{P}_k\,\mathbf{A}_{k+1}^\top \left(\mathbf{A}_{k+1}\,\mathbf{P}_k\,\mathbf{A}_{k+1}^\top + \mathbf{Q}_{k+1}\right)^{-1}$ — smoothing gain
    $\mathbf{m}_k \leftarrow \mathbf{m}_k + \mathbf{G}_k \left(\mathbf{m}_{k+1} - \mathbf{A}_{k+1}\,\mathbf{m}_k\right)$ — update
    $\mathbf{P}_k \leftarrow \mathbf{P}_k + \mathbf{G}_k \left(\mathbf{P}_{k+1} - \mathbf{A}_{k+1}\,\mathbf{P}_k\,\mathbf{A}_{k+1}^\top - \mathbf{Q}_{k+1}\right) \mathbf{G}_k^\top$
    **if** has label $\mathbf{y}_k$ **then**
      $\mathbf{\Sigma}_k^{\text{cav}} \leftarrow \left((\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^\top)^{-1} - \alpha\big(\mathbf{\Sigma}_k^{\text{site}}\big)^{-1}\right)^{-1}$ — remove site to get cavity covariance...
      $\boldsymbol{\mu}_k^{\text{cav}} \leftarrow \mathbf{\Sigma}_k^{\text{cav}} \left((\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^\top)^{-1} \mathbf{H}_k \mathbf{m}_k - \alpha\big(\mathbf{\Sigma}_k^{\text{site}}\big)^{-1} \boldsymbol{\mu}_k^{\text{site}}\right)$ — and cavity mean
      $\mathbf{J}_{\mathbf{f}_k} \leftarrow \mathbf{J}_\mathbf{f}|_{\boldsymbol{\mu}_k^{\text{cav}}, \mathbf{0}}$; $\mathbf{J}_{\boldsymbol{\sigma}_k} \leftarrow \mathbf{J}_{\boldsymbol{\sigma}}|_{\boldsymbol{\mu}_k^{\text{cav}}, \mathbf{0}}$ — evaluate Jacobians
      $\mathbf{v}_k \leftarrow \mathbf{y}_k - \mathbf{h}(\boldsymbol{\mu}_k^{\text{cav}}, \mathbf{0})$ — residual
      $\mathbf{\Sigma}_k^{\text{site}} \leftarrow \left(\mathbf{J}_{\mathbf{f}_k}^\top \left(\mathbf{J}_{\boldsymbol{\sigma}_k} \mathbf{\Sigma}_k \mathbf{J}_{\boldsymbol{\sigma}_k}^\top\right)^{-1} \mathbf{J}_{\mathbf{f}_k}\right)^{-1}$ — match moments to get site covariance...
      $\boldsymbol{\mu}_k^{\text{site}} \leftarrow \boldsymbol{\mu}_k^{\text{cav}} + \left(\mathbf{\Sigma}_k^{\text{site}} + \alpha \mathbf{\Sigma}_k^{\text{cav}}\right) \mathbf{J}_{\mathbf{f}_k}^\top \left(\mathbf{J}_{\boldsymbol{\sigma}_k} \mathbf{\Sigma}_k \mathbf{J}_{\boldsymbol{\sigma}_k}^\top + \alpha \mathbf{J}_{\mathbf{f}_k} \mathbf{\Sigma}_k^{\text{cav}} \mathbf{J}_{\mathbf{f}_k}^\top\right)^{-1} \mathbf{v}_k$ — and site mean
    **end if**
  **end for**
**end while**
**Return:** $\mathbb{E}[\mathbf{f}(t_k)] = \mathbf{H}_k \mathbf{m}_k$; $\mathbb{V}[\mathbf{f}(t_k)] = \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^\top$ — posterior marginal mean and variance
$\log p(\mathbf{y} \mid \boldsymbol{\theta}) \simeq -\sum_{k=1}^n \mathbf{e}_k$ — log marginal likelihood

# I. Continuous Measurement Model Approximations for Sec. 4

The next subsections provide further details of the model formulations used in the experiments (*i.e.*, how to write down approximative measurement models for common tasks such as heteroscedastic noise modelling, Poisson likelihoods, or logistic classification).

## I.1. Heteroscedastic Noise Model

The heteroscedastic noise model contains one GP for the mean, $f^{(1)}$, and another for the time-varying observation noise, $f^{(2)}$, both with Matern-$3/2$ covariance functions. The GP priors are independent,

$$\begin{aligned} f^{(1)}(t) &\sim \mathcal{GP}(0, \kappa(t,t')), \\ f^{(2)}(t) &\sim \mathcal{GP}(0, \kappa(t,t')), \end{aligned} \tag{43}$$

and the likelihood model is

$$\mathbf{y} \,|\, \mathbf{f}^{(1)}, \mathbf{f}^{(2)} \sim \prod_{k=1}^{n} \mathrm{N}(y_k \,|\, f_k^{(1)}, [\phi(f_k^{(2)})]^2). \tag{44}$$

The corresponding state space observation model is

$$\mathbf{h}(\mathbf{f}_k, \sigma_k) = f_k^{(1)} + \phi(f_k^{(2)})\sigma_k, \tag{45}$$

where $\sigma_k \sim \mathrm{N}(0,1)$ and $\phi(f) = \log(1+\exp(f-\frac{1}{2}))$. The Jacobians w.r.t. the (two-dimensional) latent GPs $\mathbf{f}_k$ and the noise variable $\sigma_k$ are,

$$\mathbf{J}_{\mathbf{f}}(\mathbf{f}_k, \sigma_k) = \frac{\partial \bar{\mathbf{h}}}{\partial \mathbf{f}_k^\top} = \left[1,\, \phi'\big(f_k^{(2)}\big)\sigma_k\right], \tag{46a}$$

$$\mathbf{J}_{\boldsymbol{\sigma}}(\mathbf{f}_k, \sigma_k) = \frac{\partial \bar{\mathbf{h}}}{\partial \sigma_k^\top} = \phi\big(f_k^{(2)}\big), \tag{46b}$$

where the derivative of the softplus is the sigmoid function:

$$\phi'(f) = \frac{1}{1+\exp(-f+\frac{1}{2})}. \tag{47}$$

In practice a problem arises when using the above linearisation. Since the mean of $\sigma_k$ is zero, the Jacobian w.r.t. $f^{(2)}$ disappears when evaluated at the mean regardless of the value of $f^{(2)}$. This means that the second latent function is never updated, which results in poor performance, as shown in Table 1. We found that statistical linearisation suffers from a similar issue, providing little importance to the latent function that models the noise, which highlights a potential weakness of linearisation-based methods.

Fig. 7 plots a breakdown of the different components in the posterior for the motorcycle crash data set.

## I.2. Log-Gaussian Cox Process

For a log-Gaussian Cox process, binning the data into subregions and assuming the process has locally constant intensity in these subregions allows us to use a Poisson likelihood, $p(\mathbf{y} \,|\, \mathbf{f}) \approx \prod_{k=1}^{n} \mathrm{Poisson}(\mathbf{y}_k \,|\, \exp(\mathbf{f}(\hat{t}_k)))$, where $\hat{t}_k$ is the bin coordinate and $\mathbf{y}_k$ the number of data points in it. However, the Poisson is a discrete probability distribution and the EKF and EEP methods requires the observation model to be differentiable. Therefore we use a Gaussian approximation, noticing that the first two moments of the Poisson distribution are equal to the intensity $\boldsymbol{\lambda}_k = \exp\big(\mathbf{f}_k\big)$.

We have a GP prior over $\mathbf{f}$:

$$\mathbf{f}(t) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}(t,t')), \tag{48}$$

and the approximate Gaussian likelihood is

$$p(\mathbf{y} \,|\, \mathbf{f}) = \prod_{k=1}^{n} \mathrm{N}\left(\mathbf{y}_k \,|\, \exp(\mathbf{f}_k),\, \mathrm{diag}[\exp(\mathbf{f}_k)]\right), \tag{49}$$

where $\mathrm{diag}[\exp(\mathbf{f}_k)]$ is a diagonal matrix whose entries are the elements of $\exp(\mathbf{f}_k)$. This implies the following state space observation model:

$$\mathbf{h}(\mathbf{f}_k, \boldsymbol{\sigma}_k) = \exp(\mathbf{f}_k) + \mathrm{diag}[\exp(\mathbf{f}_k/2)]\, \boldsymbol{\sigma}_k, \tag{50}$$

where $\boldsymbol{\sigma}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$. The EKF and EEP algorithms require the Jacobian of $\mathbf{h}(\mathbf{f}_k, \boldsymbol{\sigma}_k)$ with respect to $\mathbf{f}_k$ and $\boldsymbol{\sigma}_k$, which are given by,

$$\begin{aligned} \mathbf{J}_{\mathbf{f}}(\mathbf{f}_k, \boldsymbol{\sigma}_k) &= \frac{\partial \mathbf{h}}{\partial \mathbf{f}_k^\top} \\ &= \mathrm{diag}\left[\exp(\mathbf{f}_k) + \frac{1}{2}\mathrm{diag}[\exp(\mathbf{f}_k/2)]\boldsymbol{\sigma}_k\right], \end{aligned} \tag{51a}$$

$$\mathbf{J}_{\boldsymbol{\sigma}}(\mathbf{f}_k, \boldsymbol{\sigma}_k) = \frac{\partial \mathbf{h}}{\partial \boldsymbol{\sigma}_k^\top} = \mathrm{diag}[\exp(\mathbf{f}_k/2)]\,. \tag{51b}$$

## I.3. Bernoulli (Logistic Classification)

As in standard GP classification we place a GP prior over the latent function $f$, and use a Bernoulli likelihood by mapping $f$ through the logistic function $\psi(f) = \frac{1}{1+\exp(-f)}$,

$$f(t) \sim \mathcal{GP}(0, \kappa(t,t')), \tag{52a}$$

$$\mathbf{y} \,|\, \mathbf{f} \sim \prod_{k=1}^{n} \mathrm{Bern}(\psi(f(t_k))). \tag{52b}$$

As with the Poisson likelihood, we wish to approximate the Bernoulli with a distribution that has continuous support. We form a Gaussian approximation whose mean and variance are equal to that of the Bernoulli distribution, which has mean $\mathbb{E}[y \,|\, f] = \psi(f)$, and variance $\mathrm{Var}[y \,|\, f] = \psi(f)(1-\psi(f))$, giving:

$$\mathbf{y} \,|\, \mathbf{f} \sim \prod_{k=1}^{n} \mathrm{N}\big(\, y_k \,|\, \psi(f_k),\, \psi(f_k)(1-\psi(f_k))\,\big). \tag{53}$$

Therefore the approximate state space observation model is

$$h(f_k, \sigma_k) = \frac{1}{1+\exp(-f_k)} + \frac{\exp(f_k/2)}{1+\exp(f_k)}\sigma_k, \quad (54)$$

and the Jacobians are

$$\mathbf{J}_f(f_k, \sigma_k) = \frac{\partial h}{\partial f_k} = \frac{\exp(f_k)}{(1+\exp(f_k))^2} + \frac{\exp(f_k/2) - \exp(3f_k/2)}{2(1+\exp(f))^2}\sigma_k, \quad (55a)$$

$$\mathbf{J}_\sigma(f_k, \sigma_k) = \frac{\partial h}{\partial \sigma_k} = \frac{\exp(f_k/2)}{1+\exp(f_k)}. \quad (55b)$$

### I.4. Bernoulli (Probit Classification)

The Probit likelihood can be constructed similarly to the Logistic model above, by simply swapping the logistic function for the Normal CDF: $\psi(f) = \Phi(f) = \int_{-\infty}^{f} \mathrm{N}(x \,|\, 0, 1)\, \mathrm{d}x$.

## J. Supplementary Figures for Sec. 4

### J.1. Rainforest

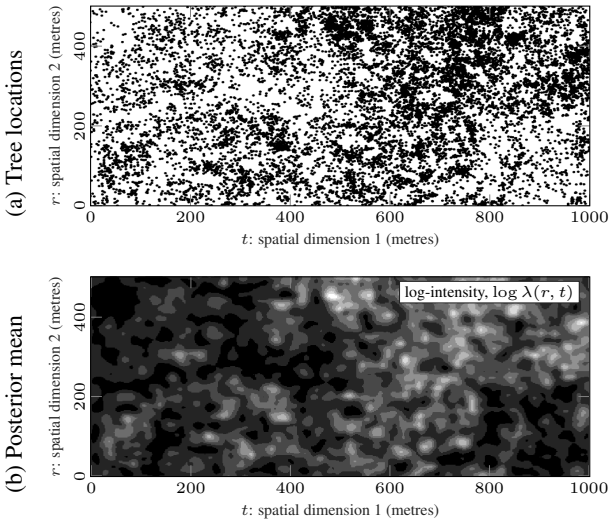


*Figure 5.* The data (a) are 12,929 tree locations in a rainforest. They are binned into a grid of $500 \times 250$ and we apply a log-Gaussian Cox process using EEP for inference. The posterior log-intensity is shown in (b).

### J.2. Audio

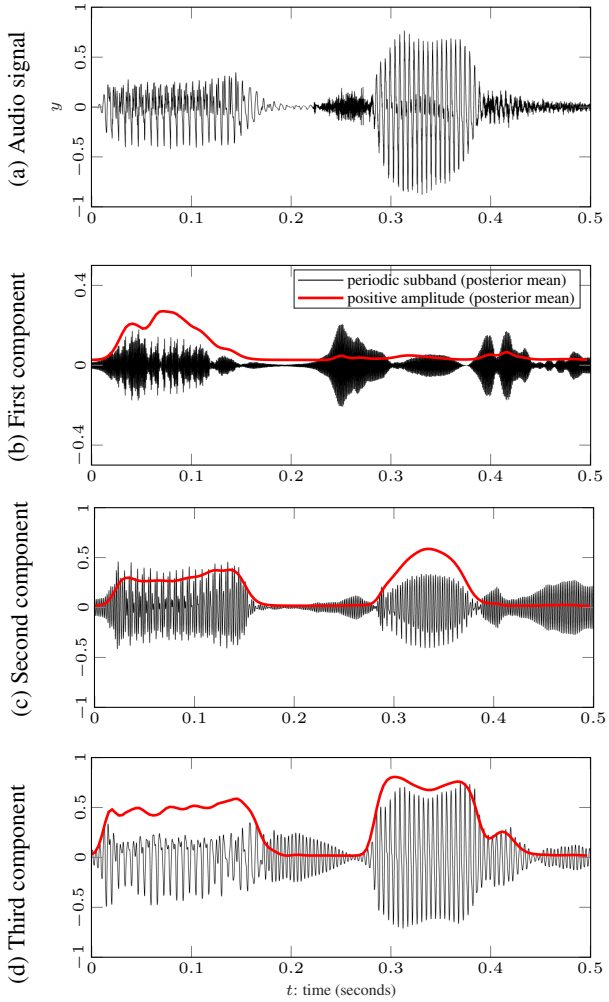


*Figure 6.* Analysis of a recording of female speech (a), duration 0.5 seconds, sampled at 44.1 kHz, $n = 22{,}050$. The three-component GP prior is overly simple given the true harmonic structure of the data, but the model is able to uncover high-, medium-, and low-frequency behaviour (b)-(d) along with their positive amplitude envelopes (shown in red). Only the posterior means are shown. The posterior for the signal (not shown) is produced by multiplying the periodic components by their amplitudes and summing the three resulting signals (see Sec. 4 for more details regarding the model). The sub-components have been rescaled for visualisation purposes.

## J.3. Motorcycle

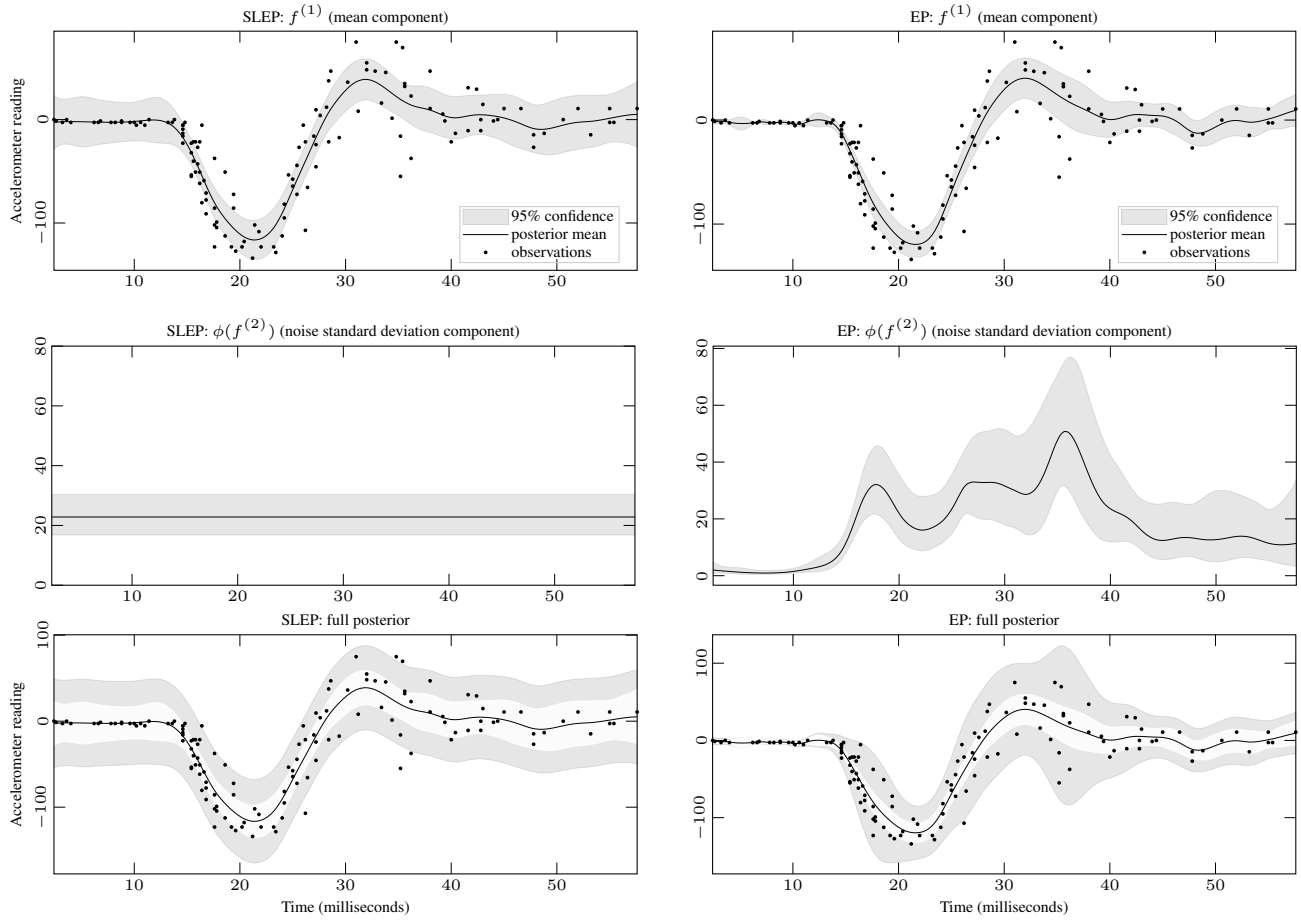


*Figure 7.* Model components for the motorcycle crash experiment. **Left** is the SLEP method (with Gauss-Hermite cubature, *i.e.* GHEP) and **right** is the EP equivalent. The linearisation-based methods fail to incorporate the heteroscedastic noise, whereas EP captures rich time-varying behaviour. The **top** plots are the posterior for $f^{(1)}(t)$ (the mean process), the **middle** plots show the posterior for $f^{(2)}(t)$ (the observation noise process), and the **bottom** plots show the full model.