HIDA Climate Datathon

TEAM BAYES
Sascha Ranftl, Christopher Albert, Gloria Wolkerstorfer

April 2020

1 Introduction

We want to infer the hidden hyperparameters θ from data $D = \{D_A, D_B\}$, $D_i = \{x_d^{(i)}, d^{(i)}\}, i = A, B$ and prior knowledge $\mathcal{C}$

The hyperparameters are in return a (non-linear) function of vulcanic and solar forcing. A change of vulcanic or solar forcing will then cause a change of the hparams. The splitting allows different correlation lengths in longitude and latitude.

Let the temperature field T be a 2D-Gaussian Process at each time-instance t. We normalize the data, i.e. subtract the time-average and simulation-average from the data. With $x = (x_1, x_2)^T$, x_1, x_2 being lateral and longitudinal coordinates, this leads us to a Zero-Mean-Function and a Covariance Function

$$p(T \mid x, \theta, \sigma, \mathcal{C}) = \mathcal{N}(T; \mu, \sigma^2 K; x) \tag{1}$$

$$\mu(x) = \vec{0} \tag{2}$$

$$K(x, x'; \theta) = K_1(x_1, x_1'; \theta_1) \otimes K_2(x_2, x_2'; \theta_2) \tag{3}$$

We move σ out of K because we can analytically integrate σ for full UQ. Note that we can split the proposition in equation 1, but not in the factorisation of the posterior equation. 9

From this we find the standard posterior for predictions

$$p(T \mid x, \theta, \sigma, D) = \mathcal{N}(T; \mu^*; , \sigma^2 K^*; x) \tag{4}$$

$$\mu^* = K(x, x_d) K(x_d, x_d)^{-1} \vec{d} \tag{5}$$

$$K^* = K(x, x) - K(x_d, x) K(x_d, x_d)^{-1} K(x_d, x) \tag{6}$$

where $x_d = (x_A, x_B)^T, x_A = (x_{A,1}, x_{A,2})^T$

which readily gives us predictions and associated prediction uncertainties via posterior mean and posterior covariance for fixed hparams.

We need

$$p(T \mid x, D) = \int d\theta d\sigma p(T \mid x, \theta, \sigma, D) p(\theta, \sigma \mid \cancel{x}, D) \tag{7}$$

$$= \int d\theta d\sigma \; p(T \mid x, \theta, \sigma, D) p(D_1 \mid \theta, \sigma) p(D_2 \mid \theta, \sigma) p(\theta \mid \mathcal{C}) p(\sigma \mid \mathcal{C}) \tag{8}$$

where we used

$$p(D \mid \theta, \sigma) = \prod_i p(D_i \mid \theta, \sigma) \tag{9}$$

$$p(\theta, \sigma \mid D) = p(D_1, D_2 \mid \theta, \sigma) p(\theta, \sigma \mid \mathcal{C}) \tag{10}$$

$$= p(D_1 \mid \theta, \sigma) \cdot p(D_2 \mid \theta, \sigma) \cdot p(\theta, \sigma \mid \mathcal{C}) \tag{11}$$

$$\log p(D_i \mid \theta, \sigma) = -\frac{d_i^T (K_i^*)^{-1} d_i}{2\sigma^2} - \frac{1}{2} \log \det(\sigma^2 K_i^*) - \frac{N_i}{2} \log 2\pi \tag{12}$$

We will use Jeffreys' prior for σ, i.e. $p(\sigma \mid \mathcal{C}) = \frac{1}{\sigma}$ and integrate analytically wrt σ so we end up with a student-t distribution to be integrated wrt θ. Then the marginal follows as (with Jeffreys prior for σ), $N = N_A + N_B$ the number of data points in D_A, D_B

$$p(\theta \mid D) = \int p(\theta, \sigma \mid D) d\sigma \tag{13}$$

$$= \frac{(2\pi)^{-(N/2)}}{2 \prod_i \sqrt{\det K_i}} \sqrt{\sum_i \frac{d_i^T (K_i^*)^{-1} d_i}{2}}^{N} \cdot \Gamma(N) \tag{14}$$

and the properly normalized 2nd moment conditioned on θ

$$\frac{\left\langle \sigma^2 \mid \theta \right\rangle}{\left\langle \sigma^0 \mid \theta \right\rangle} = \frac{1}{Z} \int \sigma^2 p(\theta, \sigma \mid D) d\sigma \tag{15}$$

$$= \frac{1}{2} \frac{1}{N_A + N_B - 1} \Big[\sum_{i=A,B} d_i^T (K_i^*)^{-1} d_i \Big] \tag{16}$$

Integrating this expression as well as θ wrt the θ-posterior yields the estimate for σ^2 for each time-instance.

$$\left\langle \sigma^2 \right\rangle = \frac{1}{2} \frac{1}{N_A + N_B - 1} \int \Big[\sum_{i=A,B} d_i^T (K_i^*)^{-1} d_i \Big] \; p(\theta \mid D) d\theta \tag{17}$$

$$\left\langle \theta \right\rangle = \int \theta \; p(\theta \mid D) d\theta \tag{18}$$

Use a particle approximation or Riemann numerical integral to solve above integrals. If need be due to limited computational resources, we may concentrate the posterior mass at point MAP, i.e.

$$p(\theta \mid D) \approx p(\theta^{MAP} \mid D) \cdot \delta(\theta - \theta^{MAP}) \tag{19}$$

$$\theta^{MAP} = \arg\max_{\theta} p(\theta \mid D) \tag{20}$$

and substitute the integration for an optimisation. The MAP-estimate of previous time-instance serves as an educated guess for the optimisation-starting-point of the next time-instance, effectively further speeding up.

We need to track the expected/MAP hparams σ, θ over time, i.e. compute for each time-instance.

We can further infer temperature predictions at unobserved locations and time-instances (just substitute $x \to t$), with the predictive mean and variance

$$\left\langle T(x)^n \right\rangle = \int T(x)^n p(T(x) \mid x, D) dT \tag{21}$$

$$= \int T(x)^n p(T(x), \theta, \sigma \mid x, D) dT d\theta d\sigma \tag{22}$$

$$= \int T(x)^n p(T(x) \mid \theta, \langle\sigma^2\rangle, D)\ p(\theta \mid D) dT d\theta \tag{23}$$

The substitution $\sigma^2 \to \left\langle \sigma^2 \mid \theta \right\rangle$ via integration is actually exact. Note that $\left\langle \sigma^2 \mid \theta \right\rangle$ still depends on θ, but will appear in the second moment only anyway. Since T is per assumption a Gaussian process, the integral wrt T results in a replacement of $T \to \mu^*$ for the mean, and $T(x) \times T(x')^T \to \langle \sigma^2 \mid \theta \rangle K^*$ for the covariance. We are still exact.

1.1 Model Comparison

Herein we shall proove that there is no evidence for solar activity to cause predicted climate change.

Let H_1 say there is volcanic activity, and H_2 say there is volcanic activity and solar activity. They shall be defined by their time-scales in the kernel structure, i.e. one time-scale in H_1 (2 /(1 for radial kernel) hparams) and two time-scales in H_2 (4/(2 for radial kernel) hparams). Obviously the latter ones prior volume is much larger.

Again $D = \{D_A, D_B\}$. A priori, both hypotheses shall be equally probable. Then

$$\frac{p(H_1 \mid D)}{p(H_2 \mid D)} = \frac{p(D \mid H_1)}{p(D \mid H_2)} \tag{24}$$

where with $q = 1, 2$

$$p(D \mid H_q) = \int p(\theta \mid D, H_q) p(\theta \mid H_q) d\theta \tag{25}$$

This has essentially already been solved above by computing $p(\theta \mid D)$. Again, a flat prior is suggested.

1.2 What goes into the computer

We only need to approximate the integrals wrt θ numerically. The basic ingredient is $p(\theta \mid D)$ as written down above, all other quantities have been computed already too.

Define a 2dim equidistant grid for $\theta_1 = \{\theta_1^{(k)}\}_{k=1}^{N_{\theta_1}}, \theta_2 = \{\theta_2^{(l)}\}_{l=1}^{N_{\theta_2}}$ in the region of the posterior mass and compute $p(\theta \mid D)$ on that grid. Then all the other quantities follow from these Riemannian integrals

$$p(\theta_1^{(k)}, \theta_2^{(l)} \mid D) = \textit{see above} \tag{26}$$

$$p(H_q \mid D) \approx \Delta\theta_1 \Delta\theta_2 \sum_{k,l} p(\theta_1^{(k)}, \theta_2^{(l)} \mid D) \tag{27}$$

$$\left\langle \sigma^2 \right\rangle \approx \frac{1}{2} \frac{\Delta\theta_1 \Delta\theta_2}{N_A + N_B - 1} \sum_{k,l} \Big[\sum_{i=A,B} d_i^T (K_i^*)^{-1} d_i \Big] \; p(\{\theta_1^{(k)}, \theta_2^{(l)}\} \mid D) \tag{28}$$

$$\left\langle \theta_j \right\rangle \approx \Delta\theta_1 \Delta\theta_2 \sum_{k,l} \theta_j \; p(\{\theta_1^{(k)}, \theta_2^{(l)}\} \mid D) \quad , \; j = 1, 2 \tag{29}$$

$$\left\langle T(x) \right\rangle \approx \Delta\theta_1 \Delta\theta_2 \sum_{k,l} \mu^*(x; \theta_1^{(k)}, \theta_2^{(l)}) \; p(\{\theta_1^{(k)}, \theta_2^{(l)}\} \mid D) \tag{30}$$

$$\left\langle T(x) \cdot T(x')^T \right\rangle \approx \Delta\theta_1 \Delta\theta_2 \sum_{k,l} \left\langle \sigma^2 \mid \theta_1^{(k)}, \theta_2^{(l)} \right\rangle K^*(x, x'; \theta_1^{(k)}, \theta_2^{(l)}) \; p(\{\theta_1^{(k)}, \theta_2^{(l)}\} \mid D) \tag{31}$$