Incorporating Dependencies in Spectral Kernels for Gaussian Processes – Supplementary Material

Kai Chen[1,2,3,5], Twan van Laarhoven[3,4], Jinsong Chen[1,2,5], and Elena Marchiori[3]

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, PR China
js.chen@siat.ac.cn
[2] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, PR China
[3] Institute for Computing and Information Sciences, Radboud University, The Netherlands
{kai,elenam}@cs.ru.nl
[4] Faculty of Management, Science and Technology, Open University of The Netherlands, The Netherlands
twan.vanlaarhoven@ou.nl
[5] Shenzhen Engineering Laboratory of Ocean Environmental Big Data Analysis and Application, Shenzhen, 518055, PR China

1 Derivation of Posterior Covariance and Correlation

We can construct the posterior covariance [2] of $\boldsymbol{f}_i^*$ and $\boldsymbol{f}_j^*$ as

$$\mathbb{V}\left(\boldsymbol{f}_i^* + \boldsymbol{f}_j^* | \boldsymbol{f}_i, \boldsymbol{f}_j\right) = \mathbb{V}\left(\boldsymbol{f}_i^* | \boldsymbol{f}_i, \boldsymbol{f}_j\right) + \mathbb{V}\left(\boldsymbol{f}_j^* | \boldsymbol{f}_i, \boldsymbol{f}_j\right) + 2\,\mathrm{Cov}\left(\boldsymbol{f}_i^*, \boldsymbol{f}_j^* | \boldsymbol{f}_i, \boldsymbol{f}_j\right) \tag{1}$$

with

$$\mathbb{V}\left(\boldsymbol{f}_i^* + \boldsymbol{f}_j^* | \boldsymbol{f}_i, \boldsymbol{f}_j\right) = K_{i+j}^{**} - {\mathbf{k}_{i+j}^*}^\top K_{i+j}^{-1} \mathbf{k}_{i+j}^* \tag{2}$$

where $K_{i+j} = K_i + K_j$, $K_{i+j}^{**} = K_i^{**} + K_j^{**}$, and $\mathbf{k}_{i+j}^* = \mathbf{k}_i^* + \mathbf{k}_j^*$. Therefore we have the posterior cross covariance [2] between $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ as

$$\mathrm{Cov}\left(\boldsymbol{f}_i^*, \boldsymbol{f}_j^* | \boldsymbol{f}_i, \boldsymbol{f}_j\right) = -{\mathbf{k}_i^*}^\top K_{i+j}^{-1} \mathbf{k}_j^* \tag{3}$$

$$\rho_{ij}^* = \frac{\mathrm{Cov}\left(\boldsymbol{f}_i^*, \boldsymbol{f}_j^* | \boldsymbol{f}_i, \boldsymbol{f}_j\right)}{\left(\mathbb{V}\left(\boldsymbol{f}_i^* | \boldsymbol{f}_i, \boldsymbol{f}_j\right) \mathbb{V}\left(\boldsymbol{f}_j^* | \boldsymbol{f}_i, \boldsymbol{f}_j\right)\right)^{1/2}} \tag{4}$$

where $\mathrm{Cov}\left(\boldsymbol{f}_i^*, \boldsymbol{f}_j^* | \boldsymbol{f}_i, \boldsymbol{f}_j\right)$ definitely is not equal to zero and ρ_{ij}^* denotes the normalized posterior correlation of components $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$. The i-th SM component is of cause independent from the j-th SM component when $\rho_{ij}^* = 0$ otherwise they are dependent. Thus we think all SM, additive, and manually designed compositional kernels should have some kind of dependency from their nonzero posterior covariance. We will show such nonzero ρ^*_{ij} between components $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ in SM in the experiment section. Note that Equation (1) tells us any linearly combined kernels in GPs, on the prediction stage, should be implicitly impacted by the posterior cross covariance between components.

2 Proof of Positive Semi-definite for GCSM

Using the Fourier transform, before making the spectral density symmetric, we have, by the distributivity of the convolution operator, that the GCSM kernel is:

$$\begin{aligned} k_{\text{GCSM'}}(\tau) &= \mathcal{F}^{-1}_{s\to\tau}\left[\sum_{i=1}^{Q}\sum_{j=1}^{Q}\hat{k}^{i\times j}_{\text{GCSM}}(\mathbf{s})\right](\tau) \\ &= \sum_{i=1}^{Q}\sum_{j=1}^{Q} c_{ij}\exp\left(-2\pi^2\tau^\top\Sigma_{ij}\tau\right)\left(\cos\left(2\pi\tau^\top\boldsymbol{\mu}_{ij}\right)+i\sin\left(2\pi\tau^\top\boldsymbol{\mu}_{ij}\right)\right) \end{aligned} \tag{5}$$

where Q is the number of auto-convolution components in the GCSM, i is the imaginary unit, $c_{ij} = w_{ij}a_{ij}$ is the cross contribution incorporating cross weight and cross amplitude to encodes the significance of dependency between components in GCSM.

Kernel $k_{\text{GCSM}}(\tau)$ is definitely positive semi-definite if and only if its spectral density $\hat{k}_{\text{GCSM}}(\mathbf{s})$ is positive semi-definite [1,3]. Here, given any finite set of non-zero vectors $[\mathbf{z}_1, ..., \mathbf{z}_N]^\top \in \mathbb{C}^{N\times P}$ with complex entries, $\mathbf{s} \in \mathbb{R}^P$, we have

$$\begin{aligned} \sum_{n=1}^{N}\mathbf{z}_n\left(\sum_{i=1}^{Q}\sum_{j=1}^{Q}\hat{k}^{i\times j}_{\text{GCSM}}(\mathbf{s})\right)\mathbf{z}_n^\dagger &= \sum_{n=1}^{N}\left(\left(\sum_{i=1}^{Q}\mathbf{z}_n\hat{g}_{\text{GCSM}i}(\mathbf{s})\right)\cdot\left(\sum_{j=1}^{Q}\overline{\mathbf{z}_n\hat{g}_{\text{GCSM}j}(\mathbf{s})}\right)\right) \\ &= \sum_{n=1}^{N}\left|\left(\sum_{i=1}^{Q}\mathbf{z}_n\hat{g}_{\text{GCSM}i}(\mathbf{s})\right)\right|^2 \geq 0, \end{aligned} \tag{6}$$

where $\mathbf{z}_n^\dagger$ denotes the conjugate transpose of $\mathbf{z}_n$. Thus the sum of cross spectral densities satisfy the positive definite condition. Therefore the proposed GCSM kernel $k_{\text{GCSM}}(\tau)$ must be positive semi-definite.

In particular, from the definitions of SM and GCSM, for the diagonal elements of the trained kernel matrix ($\tau = 0$), we have

$$\gamma_{ij}(\tau = 0) = a_{ij} \tag{7}$$

$$k_{\text{SM}}(\tau = 0) = \sum_{i=1}^{Q} w_i \tag{8}$$

$$k_{\text{GCSM}}(\tau = 0) = \sum_{i=1}^{Q}\sum_{j=1}^{Q} c_{ij} \tag{9}$$

The diagonal values of kernel matrix in SM are not affected by the hyper-parameters $\boldsymbol{\mu}$, Σ. Instead in GCSM the diagonal values are affected by all hyper-parameters.

Since the complex conjugate of a positive semi-definite kernel, and the sum of two positive semi-definite kernels is still positive semi-definite, the symmetrized $k_{\text{GCSM}}(\tau) = (k_{\text{GCSM'}}(\tau) + \overline{k_{\text{GCSM'}}(\tau)})/2$ is also positive semi-definite.

3 Additional Experimental Results

Posterior functions of SM are with all those affected by the dependency among SM components, especially the uncertainty who affected by the posterior covariance. From Figures 1, 2, 3 and also posterior covariance analysis, SM kernel doesn't hold the conditional independence. Thus for arbitrary SM component we have $\boldsymbol{f}_i^*|\boldsymbol{f}_i \neq \boldsymbol{f}_i^*|\boldsymbol{f}_i + \boldsymbol{f}_j$. This conclusion can be applied on most of linearly combined kernels in GP. An important finding is that the posterior dependency helps correct trends and introduces much uncertainty.

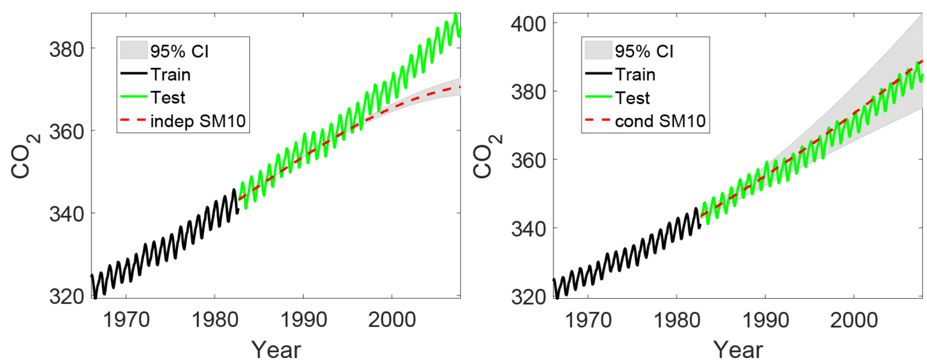

Fig. 1. Left: $\boldsymbol{f}_{10}^*|\boldsymbol{f}_{10}$; right: a conditional $\boldsymbol{f}_{10}^*|\sum_{i=1}^{10}\boldsymbol{f}_i$

References

1. Bochner, S.: Lectures on Fourier Integrals.(AM-42), vol. 42. Princeton University Press (2016)
2. Duvenaud, D.: Automatic model construction with Gaussian processes (Doctoral thesis). Ph.D. thesis (2014). https://doi.org/https://doi.org/10.17863/CAM.14087
3. Stein, M.: Interpolation of spatial data: some theory for kriging. (1999)

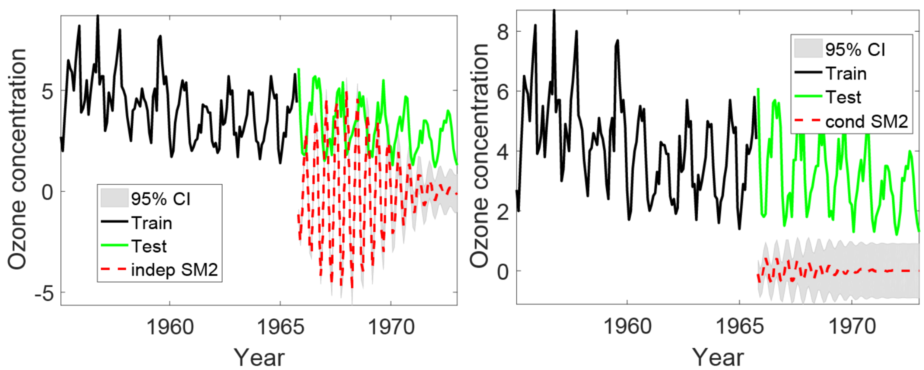

Fig. 2. Left: $\boldsymbol{f}_2^*|\boldsymbol{f}_2$; right: $\boldsymbol{f}_2^*|\sum_{i=1}^{10}\boldsymbol{f}_i$

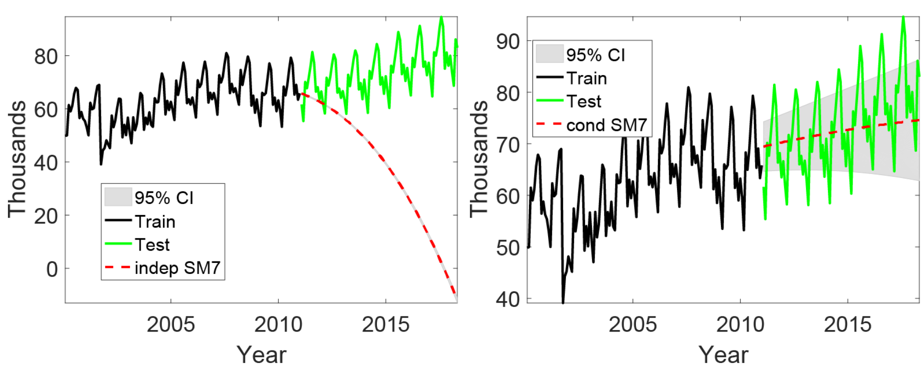

Fig. 3. Left: $\boldsymbol{f}_7^*|\boldsymbol{f}_7$; right: $\boldsymbol{f}_7^*|\sum_{i=1}^{10}\boldsymbol{f}_i$