Thin and deep Gaussian processes

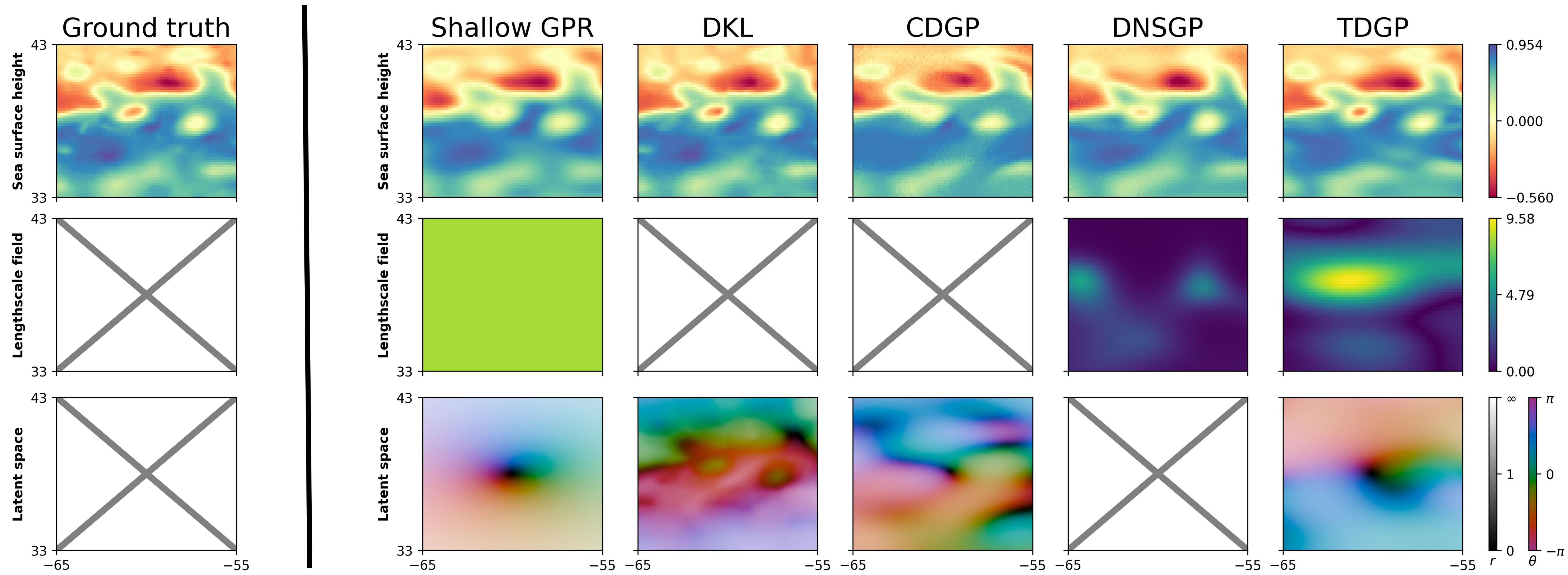
Daniel Augusto de Souza[1], Alexander Nikitin[2], S. T. John[2], Magnus Ross[3], Mauricio A Álvarez[3], Marc Peter Deisenroth[1], João P. P. Gomes[4], Diego Mesquita[5], César L. C. Mattos[4]

[1]University College London [2]Aalto University [3]University of Manchester [4]Universidade Federal do Ceará [5]Fundação Getúlio Vargas

TLDR:

1. Current hierarchical Gaussian Process methods learn one of the following:
   - latent mappings - reduce dimensionality (CDGP);
   - lengthscale fields - easily interpretable (DNSGP).
2. We:
   - Propose a method that learns both!
     - While also avoid pathologies of both CDGP and DNSGP.
   - Prove that our method extends the traditional CDGP framework.

Benchmark: Sea surface height – North Atlantic



Stationary kernels

A kernel k is stationary if:

$$\begin{aligned} k(a,b) &= k(a-b,0) \\ &= \pi_k\big((a-b)\Delta^{-1}(a-b)^T\big) \\ &= \pi_k\big((Wa-Wb)(Wa-Wb)^T\big) \end{aligned}$$

From stationary to non-stationary

There are two well known ways to get non-stationary kernels:

1. $k_{\mathrm{NS}}(a,b) = k\big(\tau(x),\tau(y)\big)$, if $\tau(x)$ follows a GP distribution, we obtain compositional deep GP (CDGP) model
2. $k_{\mathrm{PA}}(a,b) \propto \sqrt{\frac{\sqrt{|\Delta(a)|}\sqrt{|\Delta(b)|}}{|\Delta(a)+\Delta(b)|}}\,\pi_k\left((a-b)\left[\frac{\Delta(a)+\Delta(b)}{2}\right]^{-1}(a-b)^T\right)$, if $\Delta(a)$ follows a warped GP distribution, we obtain a deeply non-stationary GP (DNSGP) model.

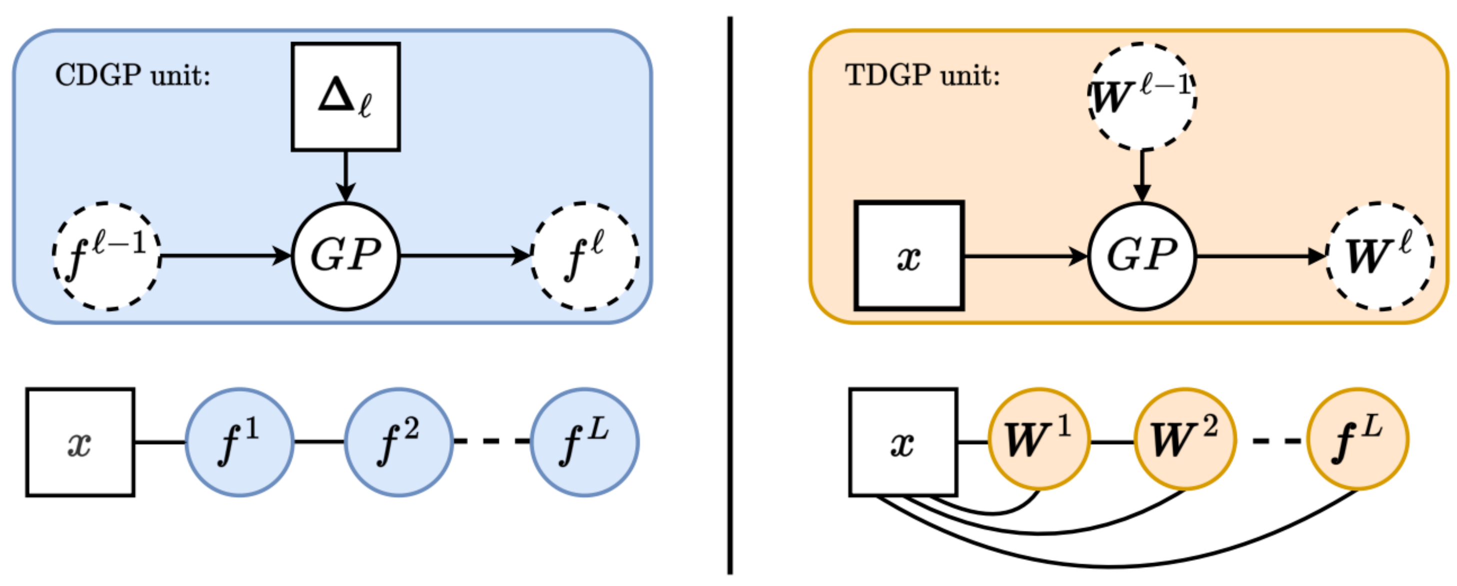
We choose a hybrid approach:

$$\begin{aligned} k_{\mathrm{TD}}(a,b) &= k\big(W(a)\cdot a, W(b)\cdot b\big) \\ &= \pi_k\Big(\big(W(a)\,a - W(b)\,b\big)\big(W(a)\,a - W(b)\,b\big)^T\Big) \end{aligned}$$

Previous issues

- CDGP models have pathologies when hidden layers have zero mean function:
  - therefore, they don't have inductive bias for dimensionality reduction;
- DNSGP models have well known interpretability problems:
  - The presence of the expressions with the lengthscale outside the quadratic term harms their interpretability;
  - The quadratic term in the kernel violates the triangle inequality.

Graphical models



UCL
Aalto University
MANCHESTER 1824
The University of Manchester
UFC
FGV