

# Important references

Monday, September 30, 2024 4:24 PM

As much as I don't want to distract you from publication workstream, there is a paper that at least is worth attention and whose experiments on MO data may be a good reference: <https://openreview.net/forum?id=6N8TW504aa>  
Graphical Multioutput Gaussian Process with Attention | OpenReview  
Integrating information while recognizing dependence from multiple data sources and enhancing the predictive performance of the multi-output regression are challenging tasks. Multioutput Gaussian P...

The GGM concept and possible extension is partially included.

<https://arxiv.org/pdf/1211.0358>  
Deep Gaussian Processes

Do you have an implementation of HGP in pytorch? If not, can we try the deepGP in gpytorch?  
[https://docs.gpytorch.ai/en/v1.13/examples/05\\_Deep\\_Gaussian\\_Processes/index.html](https://docs.gpytorch.ai/en/v1.13/examples/05_Deep_Gaussian_Processes/index.html)  
Deep GP and Deep Sigma Point Processes — GPyTorch 1.13 documentation

Clicking in one link on the above, there is this "In this example, we will demonstrate how to construct deep GPs that can model vector-valued functions (e.g. multitask/multi-output GPs)." I think this is HGP+MO in gpytorch!?



Another implementation sent on 12:01 above which may be worth looking into (correction from my previous comment) due to:

Multi Output Spectral Mixture (MOSM) [5]  
Convolutional Model (CONV) [4]  
Cross-Spectral Mixture (CSM) [3]

The 5 papers, mostly the first 3, that are very much what we have been trying to do:

I. Latent Variable Multi-output Gaussian Processes for Hierarchical Datasets, <https://arxiv.org/abs/2308.16822>, 2023 [arXiv, no official publication, derivative from the next two papers]

<https://github.com/ChunchaoPeter/HMOGP-LV>

II. Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters, <http://www.biomedcentral.com/1471-2105/14/252>, 2013  
<https://notebook.community/SheffieldML/notebook/compbio/Hierarchical>  
[https://github.com/jameshensman/GPclust/blob/master/notebooks/MOHGP\\_demo.ipynb](https://github.com/jameshensman/GPclust/blob/master/notebooks/MOHGP_demo.ipynb)

III. Computationally Efficient Convolved Multiple Output Gaussian Processes,  
<https://www.jmlr.org/papers/v12/alvarez11a.html>, 2011

IV. Transfer Learning with Gaussian Processes for Bayesian Optimization,  
<https://proceedings.mlr.press/v151/tighineanu22a/tighineanu22a.pdf>, 2022

V. Hierarchical Gaussian Processes model for multi-task learning, <https://doi.org/10.1016/j.patcog.2017.09.021>, 2018 [not sure if the modeling difference with II above]

<https://github.com/ChunchaoPeter/HMOGP-LV.git>

#### Efficient Modeling of Latent Information in Supervised Learning using Gaussian Processes

Dai's paper 2017, LVMOGP

<https://arxiv.org/abs/1705.09862>

To enable efficient inference, we propose a new model which assumes the covariance matrix can be decomposed as a Kronecker product of the covariance matrix of the latent variables  $K_h$  and the covariance matrix of the inputs  $K_x$

[https://gpy.readthedocs.io/en/devel/\\_modules/GPy/models/gp\\_multiout\\_regression\\_md.html](https://gpy.readthedocs.io/en/devel/_modules/GPy/models/gp_multiout_regression_md.html)

[https://notebook.community/mikecroucher/notebook/GPy/sparse\\_gp\\_regression](https://notebook.community/mikecroucher/notebook/GPy/sparse_gp_regression)

Kernels for Vector-Valued Functions: a Review

<https://arxiv.org/pdf/1106.6251>

The index kernel in gpytorch seems not much discussed/mentioned in the literature. I found one FYI

Index kernel and "multi-group GP"?

- <https://arxiv.org/pdf/2110.08411>, 2021, <https://github.com/andrewcharlesjones/multi-group-GP>
- Not published

On a second thought, the idea of HGP and deep GP are different. HGP refers to parameter dependency. Deep GP refers to hierarchical GPs. I'll have to look into that example.

HGP implementation in gpytorch?

I am reading the original paper: <https://arxiv.org/pdf/1211.0358>. It seems a deep version of GPLVM but not an deep version of HGP.

<https://www.sciencedirect.com/science/article/pii/S0925231220317410>

I.Heterogeneous input space

1.Hierarchical Inducing Point Gaussian Process for Inter-domain Observations,

<https://proceedings.mlr.press/v130/wu21b.html>, <https://github.com/cunningham-lab/hipgp?tab=readme-ov-file>, 2021

a.Inter-domain!? ↗ heterogeneous input space

Learning Multi-Task Gaussian Process Over Heterogeneous Input Domains, <https://arxiv.org/pdf/2202.12636>, 2022

2 things are interesting in this post/notebook

[https://bigaidream.gitbooks.io/subsets\\_ml\\_cookbook/content/bayes/gp/coregionalized\\_regression\\_gpy.html](https://bigaidream.gitbooks.io/subsets_ml_cookbook/content/bayes/gp/coregionalized_regression_gpy.html)

<https://arxiv.org/abs/2212.04450>

<sup>1</sup> We may have to test on the "benchmark" dataset soon: "substantially improved performance over eight popular

1. we may have to test on the benchmark dataset soon. substantially improved performance over eight popular multiple output (multi-task) Gaussian process models and three multivariate volatility models on benchmark datasets, including a 1000 dimensional gene expression dataset", in <https://arxiv.org/pdf/1110.4411>
  1. also "These are often called 'multi-task' learning or 'multiple output' regression models. Capturing correlations between outputs (response variables) can be used to make better predictions"
2. my attempt to formulate "non-independent" latent GPs is, likely, done in 2002: "Coregionalization by linear combination of nonorthogonal components. Mathematical Geology, 34 (4):405–419, 2002", <https://link.springer.com/article/10.1023/A:1015078911063>

OMG, I see my equation(s)! The four summations, which is (much) nicely explained in the author of the slide deck and my GP go to person N. Lawrence <https://www.jmlr.org/papers/volume12/alvarez11a/alvarez11a.pdf>

Actually, one more: assuming you are still onboard on the peptide solubility enhancement topic, the seminar, whose recording should be available soon, provides a good background [https://teams.microsoft.com/l/meetup-join/19%3ameeting\\_OWMzYWJIMjgtMGVkNC00YjQ3LWFkNjItOWQxOTM5ODBjNzgx%40thread.v2/0?context=%7b%22Tid%22%3a%22a00de4ec-48a8-43a6-be74-e31274e2060d%22%2c%22Oid%22%3a%2233a5de9f-46d4-4a42-8aa5-fa6539dad39f%22%7d](https://teams.microsoft.com/l/meetup-join/19%3ameeting_OWMzYWJIMjgtMGVkNC00YjQ3LWFkNjItOWQxOTM5ODBjNzgx%40thread.v2/0?context=%7b%22Tid%22%3a%22a00de4ec-48a8-43a6-be74-e31274e2060d%22%2c%22Oid%22%3a%2233a5de9f-46d4-4a42-8aa5-fa6539dad39f%22%7d)