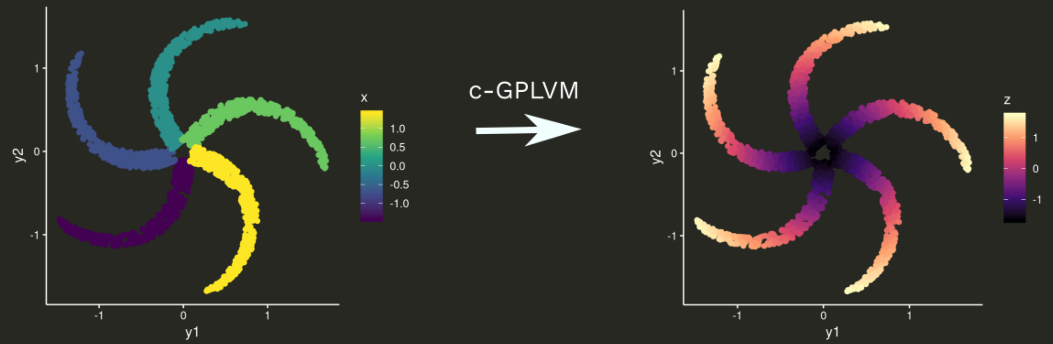
Decomposing feature-level variation with Covariate Gaussian Process Latent Variable Models

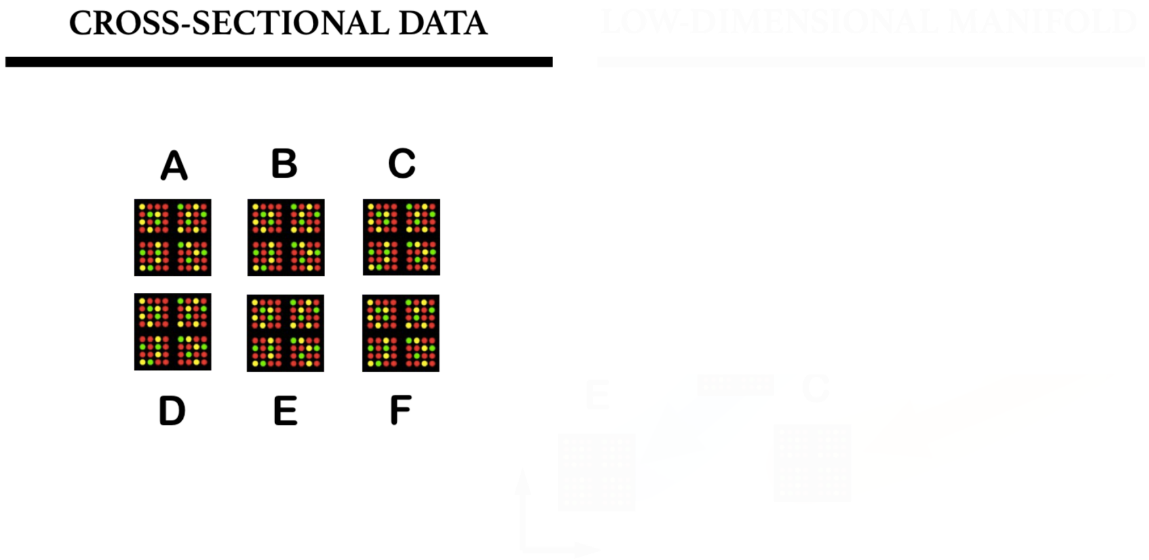


Kaspar Märtens, Kieran Campbell, Christopher Yau

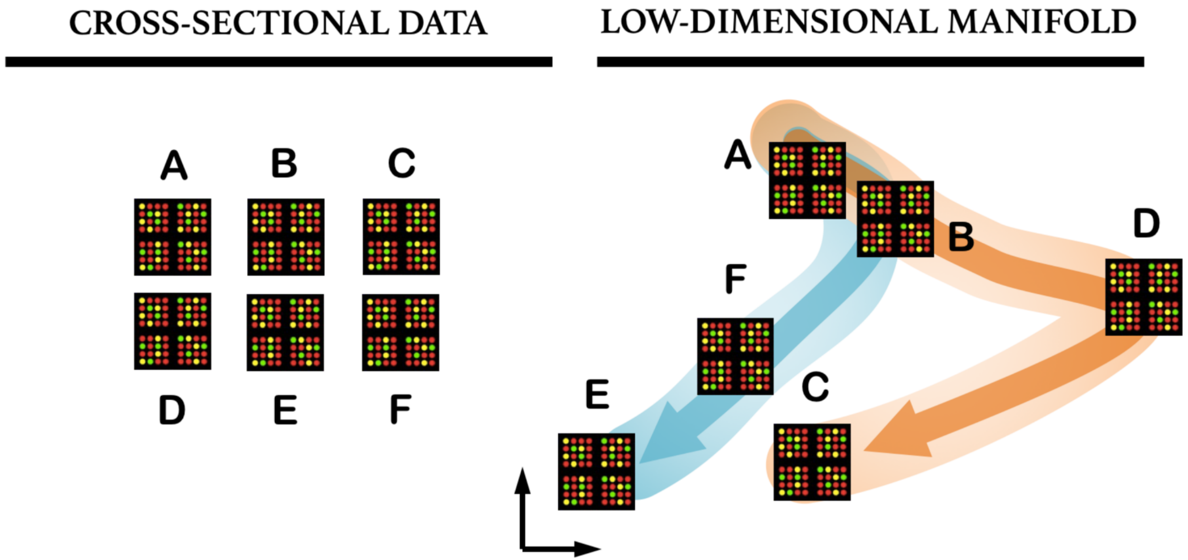
@kasparmartens

kasparmartens.rbind.io

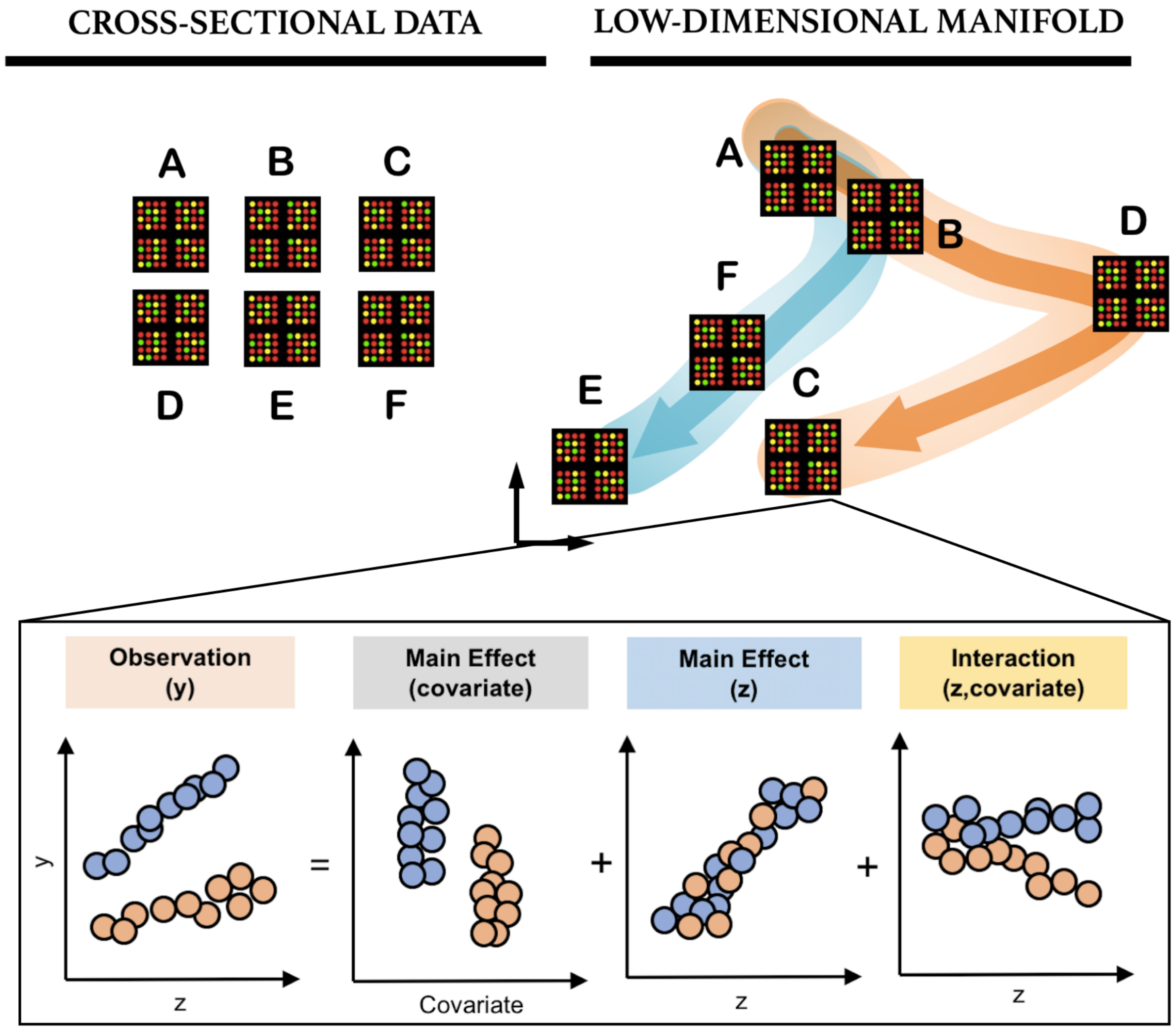
Motivation: disease progression modelling



Motivation: disease progression modelling



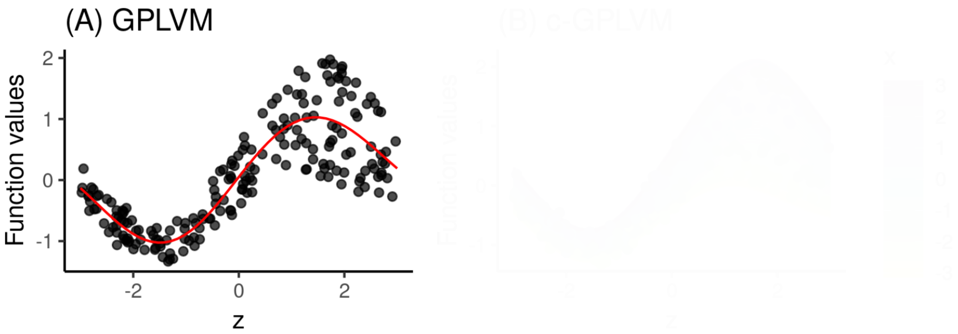
Motivation: disease progression modelling



Covariate-GPLVM

GPLVM maps latent $\mathbf{z}_i \sim \mathcal{N}(0, 1)$ to observed data $\mathbf{Y}$ using GP mappings

$$y_i^{(j)} = f^{(j)}(\mathbf{z}_i) + \varepsilon_{ij}$$



Covariate-GPLVM

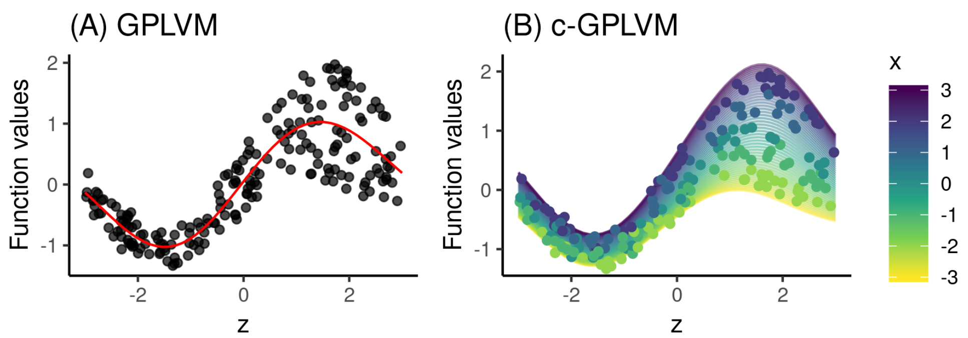
GPLVM maps latent $\mathbf{z}_i \sim \mathcal{N}(0,1)$ to observed data $\mathbf{Y}$ using GP mappings

$$y_i^{(j)} = f^{(j)}(\mathbf{z}_i) + \varepsilon_{ij}$$

Covariate-GPLVM extends GPLVM by:

1. Incorporating covariates $\mathbf{x}$

$$y_i^{(j)} = f^{(j)}(\mathbf{x}_i, \mathbf{z}_i) + \varepsilon_{ij}$$



Covariate-GPLVM

GPLVM maps latent $\mathbf{z}_i \sim \mathcal{N}(0, 1)$ to observed data $\mathbf{Y}$ using GP mappings

$$y_i^{(j)} = f^{(j)}(\mathbf{z}_i) + \varepsilon_{ij}$$

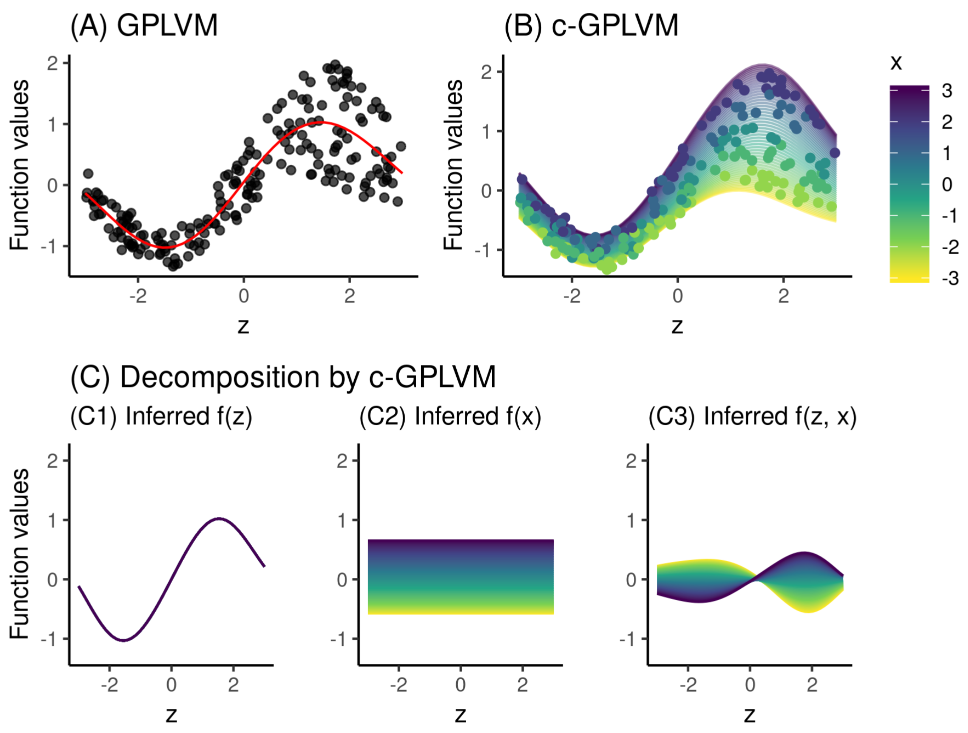
Covariate-GPLVM extends GPLVM by:

1. Incorporating covariates $\mathbf{x}$

$$y_i^{(j)} = f^{(j)}(\mathbf{x}_i, \mathbf{z}_i) + \varepsilon_{ij}$$

2. Providing a feature-level decomposition

$$y_i^{(j)} = \mu^{(j)} + f_z^{(j)}(\mathbf{z}) + f_x^{(j)}(\mathbf{x}) + f_{zx}^{(j)}(\mathbf{z}, \mathbf{x}) + \varepsilon_{ij}$$



Feature-level decomposition

$$y_i^{(j)} = \mu^{(j)} + f_z^{(j)}(\mathbf{z}) + f_x^{(j)}(\mathbf{x}) + f_{zx}^{(j)}(\mathbf{z}, \mathbf{x}) + \varepsilon_{ij}$$

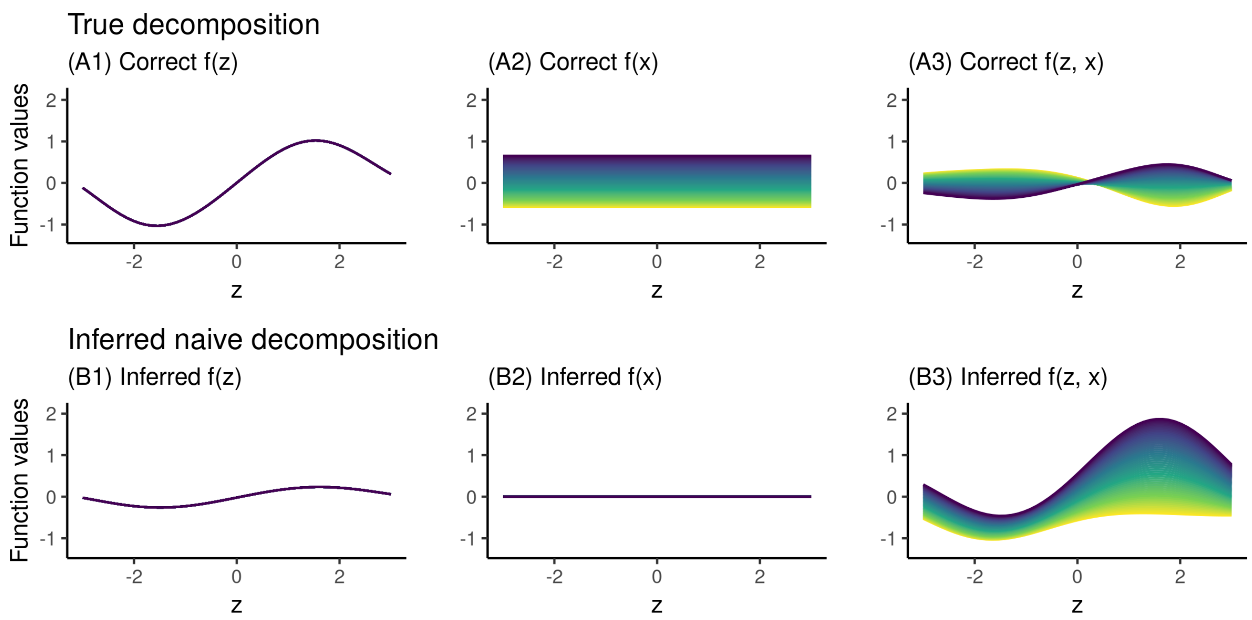
Readily available for *linear models*, otherwise challenging:

Feature-level decomposition

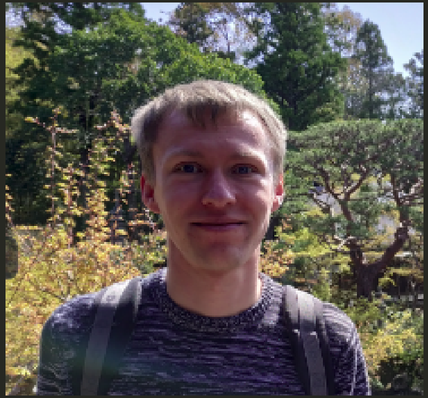
$$y_i^{(j)} = \mu^{(j)} + f_z^{(j)}(\mathbf{z}) + f_x^{(j)}(\mathbf{x}) + f_{zx}^{(j)}(\mathbf{z}, \mathbf{x}) + \varepsilon_{ij}$$

Readily available for *linear models*, otherwise challenging:

- Naive decompositions (with standard GP priors) can lead to misleading conclusions
- With appropriate *functional constraints* we learn an **identifiable non-linear decomposition**



Decomposing feature-level variation with Covariate Gaussian Process Latent Variable Models



Kaspar Märtens

Kieran Campbell

Christopher Yau

@kasparmartens

kasparmartens.rbind.io

Poster #261

Decomposing feature-level variation with Covariate GPLVMs

Kaspar Märtens[1] Kieran R Campbell[234] Christopher Yau[56]

Motivation

The interpretation of complex high-dimensional data typically requires the use of dimensionality reduction techniques (such as GPLVM) to extract low-dimensional representations ***z***. However,

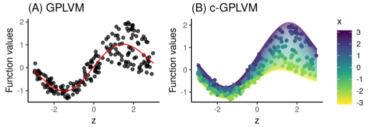
- In many problems these representations may not be sufficient to aid interpretation on their own, and it would be desirable to interpret ***z*** in terms of the original features themselves.
- Often there is external **covariate** information ***x*** that we would like to incorporate.

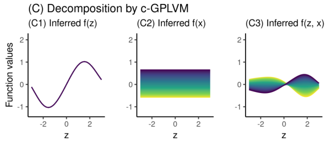
GPLVM is a non-linear dimensionality reduction method where latent $z_i \sim \mathcal{N}(0,1)$ are mapped to data ***Y*** using GP mappings

$$y_i^{(j)} = f^{(j)}(\mathbf{z}_i) + \varepsilon_i$$

Our goal is twofold:

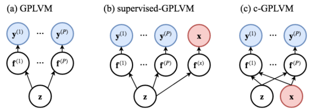
1. to incorporate covariates ***x*** within the GPLVM
2. to characterise how feature-level variation depends on latent representations ***z***, external covariates ***x***, and non-linear interactions between them



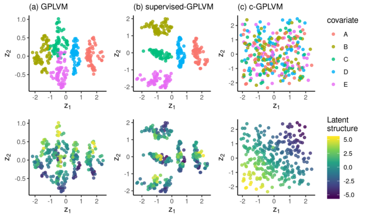


How to incorporate covariates

In principle, covariates can be incorporated within GPLVM in various ways

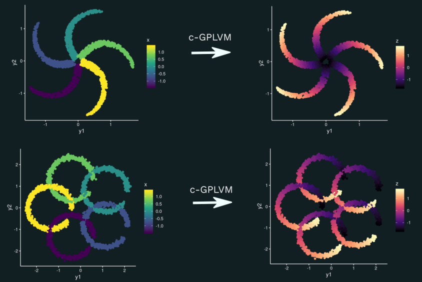


which lead to different behaviour. Our goal is to learn a *covariate-adjusted* latent ***z***, as illustrated below.



We propose a hybrid Gaussian Process model (c-GPLVM) to achieve two goals:

Goal 1. Incorporate covariates ***x*** within the GPLVM so that the latent low-dimensional ***z*** would be covariate-adjusted



We achieve this by defining GP mappings on the joint space of ***z*** and ***x***

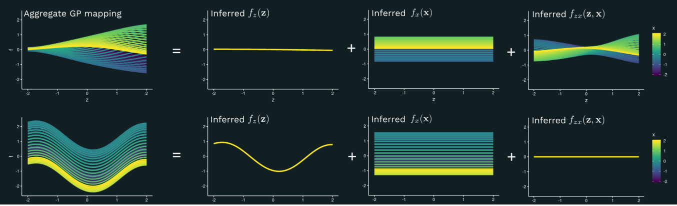
$$y_i^{(j)} = f^{(j)}(\mathbf{z}_i, \mathbf{x}_i) + \varepsilon_i$$

Goal 2. Interpretable feature-level decomposition

Classical statistics	Machine Learning
$y_i^{(j)} = \alpha^{(j)}\mathbf{z}_i + \beta^{(j)}\mathbf{x}_i + \gamma^{(j)}\mathbf{z}_i\mathbf{x}_i + \varepsilon_i$	$y_i^{(j)} = f^{(j)}(\mathbf{z}_i, \mathbf{x}_i) + \varepsilon_i$
+ Interpretable	+ Flexible model
- Not very flexible	- "Black box"

Covariate-GPLVM combines the best of both worlds. We learn an *identifiable* decomposition

$$y_i^{(j)} = f_z^{(j)}(\mathbf{z}_i) + f_x^{(j)}(\mathbf{x}_i) + f_{zx}^{(j)}(\mathbf{z}_i, \mathbf{x}_i) + \varepsilon_i$$



Identifiability

We would like to separate out the additive contributions of ***z*** and ***x***, i.e. we would like to learn the decomposition

$$y_i^{(j)} = f_0^{(j)} + f_z^{(j)}(\mathbf{z}_i) + f_x^{(j)}(\mathbf{x}_i) + f_{zx}^{(j)}(\mathbf{z}_i, \mathbf{x}_i) + \varepsilon_i$$

However, with independent GP priors this decomposition is not identifiable and it could lead to misleading conclusions. To turn this into a well-defined variance decomposition, we introduce the following functional constraints $\int_a^b f_z(z)dz = 0$, $\int_a^b f_x(x)dx = 0$, $\int_a^b f_{zx}(z,x)dz = 0$, and $\int_a^b f_{zx}(z,x)dx = 0$.

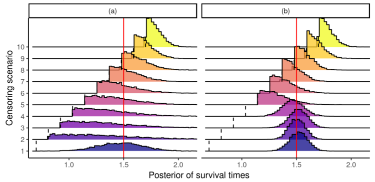
Under the GP prior on f_0, f_z, f_x, f_{zx}, conditioning on the above functional constraints is straightforward. The conditional distribution is still a GP, but with a modified kernel (Durrande et al 2013). For the squared exponential kernel, this can be calculated in closed form.

As a result, we have a joint GP prior over f_0, f_z, f_x, f_{zx} with the following properties:

- Their supports are non-overlapping
- The functional subspaces are orthogonal in L2

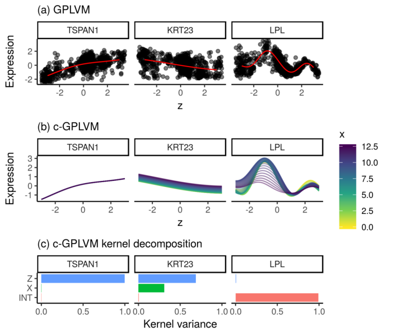
Censored covariates

The c-GPLVM framework can be extended to the case where covariate ***x*** is partially observed, e.g. when ***x*** = (potentially censored) patient survival times.



Survival-adjusted cancer modelling

On TCGA breast cancer gene expression data, using censored survival times as ***x***, we demonstrate how c-GPLVM lets us discover three distinct gene behaviours:



Affiliations

[1] Department of Statistics, University of Oxford
[2] Department of Statistics, University of British Columbia
[3] BC Cancer Agency
[4] UBC Data Science Institute
[5] The Alan Turing Institute
[6] Institute of Cancer and Genomic

Acknowledgements

KM is supported by a UK EPSRC doctoral studentship. KRC is supported by a Banting postdoctoral fellowship from the Canadian Institutes of Health Research, and postdoctoral fellowships from the Canadian Statistical Sciences Institute and the UBC DSI. CY is supported by a UK Medical Research Council Research Grant and by The Alan Turing Institute.