# Latent Forces Models: related work

Mauricio A. Álvarez

Latent Force Models Workshop, Sheffield, June 2013

# Contenido

# Different Views of the LFM (I)

- Gaussian processes over latent forces plus linear differential equations lead to multiple output Gaussian process.

- This multi-output Gaussian process uses a covariance function that encodes the interactions between the different mechanistic models.

- Marginal likelihood to estimate the hyperparameters $\theta$ of the covariance function equals estimating the parameters of differential equations.

# Different Views of the LFM (II)

- The related work can be seen from different perspectives.

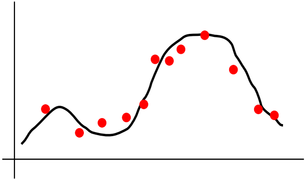
- We focus on three
  - Gaussian processes for multiple outputs.
  - Parameter estimation in differential equations.
  - Gaussian processes for systems identification.

# Contenido

# Valid covariance functions for multiple outputs (I)

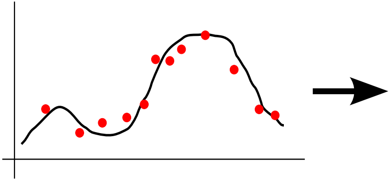
Gaussian process priors for multiple outputs have been thoroughly studied in the spatial analysis and geostatistics literature (Higdon, 2002; Boyle and Frean, 2005; Journel and Huijbregts, 1978; Cressie, 1993; Goovaerts, 1997; Wackernagel, 2003).

# Valid covariance functions for multiple outputs (II)



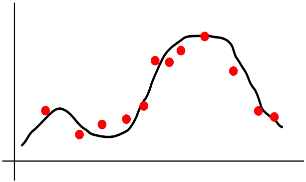
$\mathcal{D}^1 = \{(\mathbf{x}_i^1, y_i^1) | i = 1, \ldots, N_1\}$

# Valid covariance functions for multiple outputs (II)



$\mathcal{D}^1 = \{(\mathbf{x}_i^1, y_i^1) | i = 1, \ldots, N_1\}$

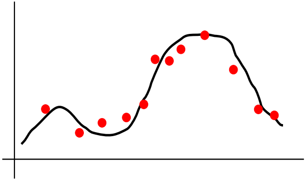
# Valid covariance functions for multiple outputs (II)



$\mathcal{D}^1 = \{(\mathbf{x}_i^1, y_i^1) | i = 1, \ldots, N_1\}$

$\longrightarrow k^1(\mathbf{x}_i, \mathbf{x}_j)$

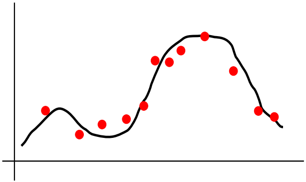
# Valid covariance functions for multiple outputs (II)



$\rightarrow$ $k^1(\mathbf{x}_i, \mathbf{x}_j)$ $\rightarrow$

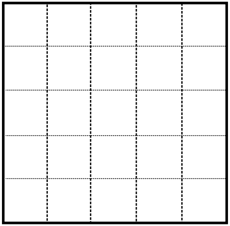
$\mathcal{D}^1 = \{(\mathbf{x}_i^1, y_i^1) | i = 1, \ldots, N_1\}$

# Valid covariance functions for multiple outputs (II)

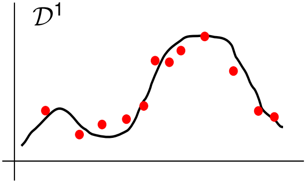


$\mathcal{D}^1 = \{(\mathbf{x}_i^1, y_i^1) | i = 1, \ldots, N_1\}$

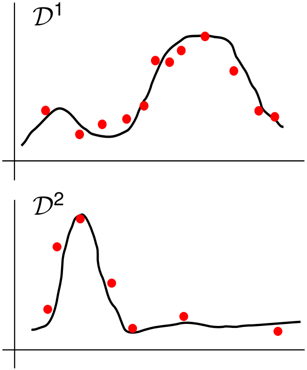
$\rightarrow k^1(\mathbf{x}_i, \mathbf{x}_j) \rightarrow \mathbf{K}^1 =$



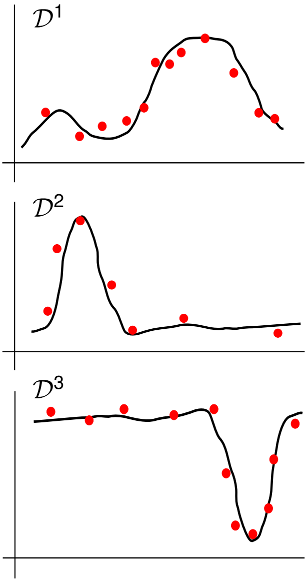
# Valid covariance functions for multiple outputs (II)



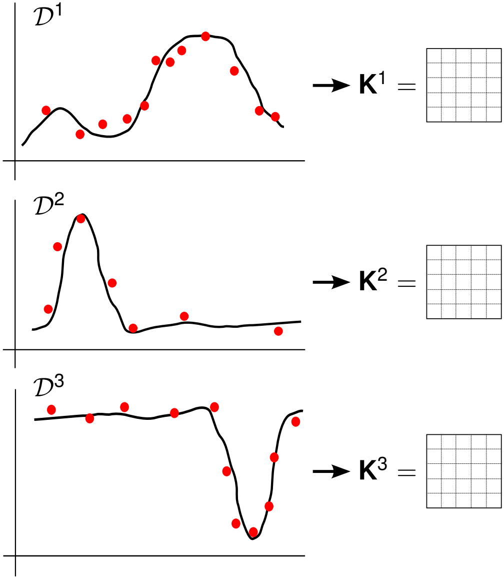
# Valid covariance functions for multiple outputs (II)



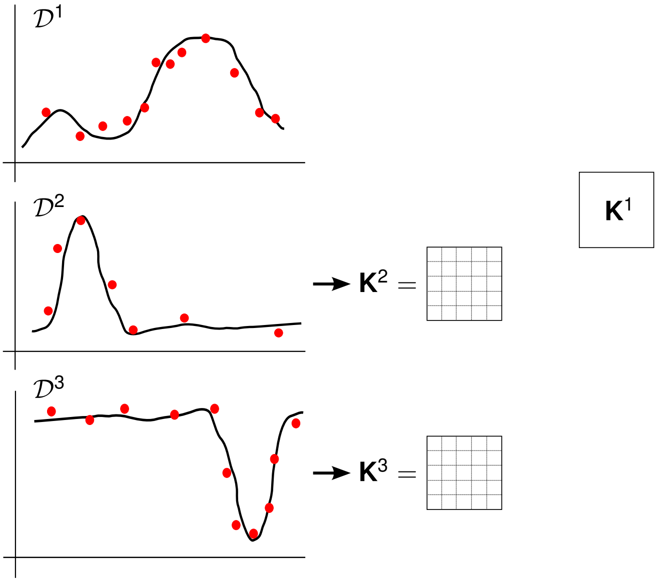
# Valid covariance functions for multiple outputs (II)



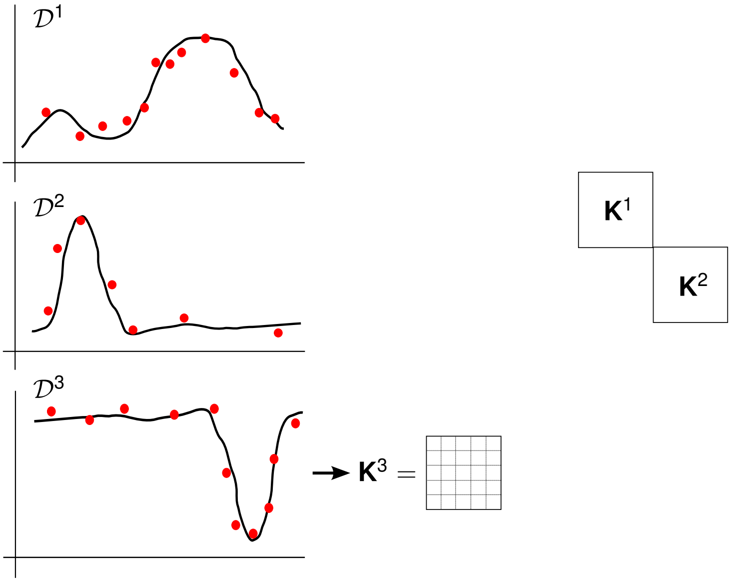
# Valid covariance functions for multiple outputs (II)



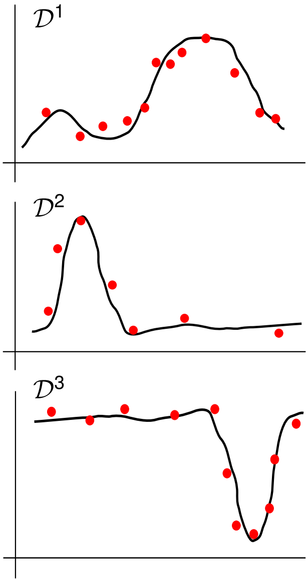
# Valid covariance functions for multiple outputs (II)

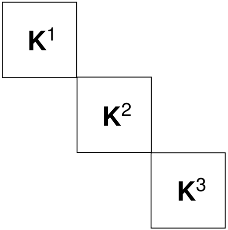


# Valid covariance functions for multiple outputs (II)

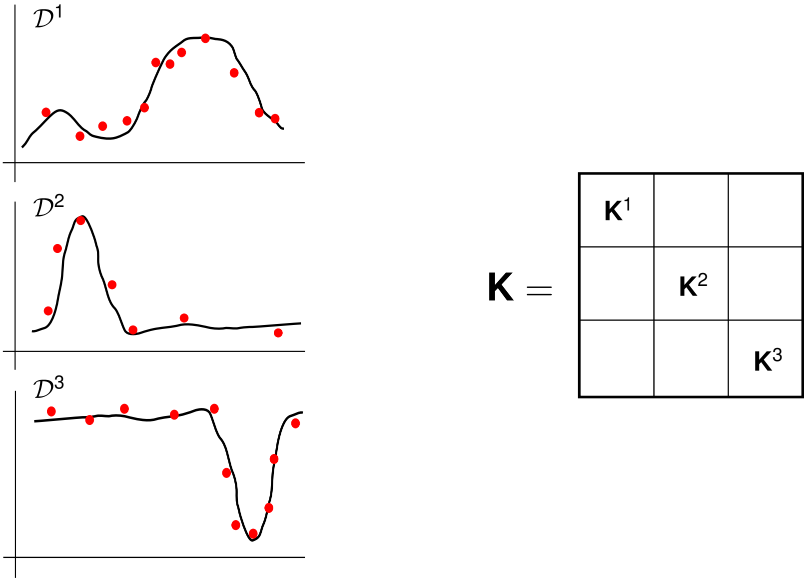


# Valid covariance functions for multiple outputs (II)

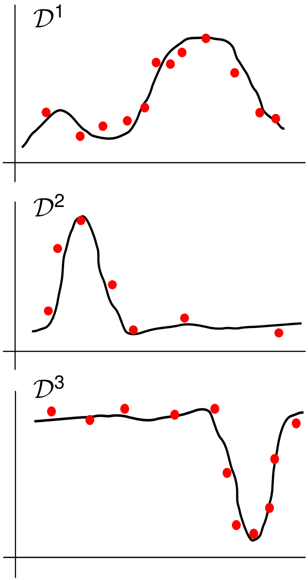


# Valid covariance functions for multiple outputs (II)



# Valid covariance functions for multiple outputs (II)




Joint covariance

$\mathbf{K} =$

| | | |
|---|---|---|
| $\mathbf{K}^1$ | ? | ? |
| ? | $\mathbf{K}^2$ | ? |
| ? | ? | $\mathbf{K}^3$ |

$\mathbf{K}$ be a valid covariance matrix

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

**Linear Model of Coregionalization** (Journel and Huijbregts, 1978; Goovaerts, 1997; Wackernagel, 2003).

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

**Linear Model of Coregionalization** (Journel and Huijbregts, 1978; Goovaerts, 1997; Wackernagel, 2003).

$$G^i_{d,q}(\mathbf{x}-\mathbf{z}) = a^i_{d,q}\delta(\mathbf{x}-\mathbf{z})$$

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

**Linear Model of Coregionalization** (Journel and Huijbregts, 1978; Goovaerts, 1997; Wackernagel, 2003).

$$G^i_{d,q}(\mathbf{x}-\mathbf{z}) = a^i_{d,q}\delta(\mathbf{x}-\mathbf{z})$$

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q} a^i_{d,q}a^i_{d',q}k_q(\mathbf{z},\mathbf{z}').$$

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

**Semiparametric Latent Factor Model** (Teh et al., 2005)

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

**Semiparametric Latent Factor Model** (Teh et al., 2005)

$$G^i_{d,q}(\mathbf{x}-\mathbf{z}) = a^i_{d,q}\delta(\mathbf{x}-\mathbf{z}), \quad R_q = 1$$

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

**Semiparametric Latent Factor Model** (Teh et al., 2005)

$$G^i_{d,q}(\mathbf{x}-\mathbf{z}) = a^i_{d,q}\delta(\mathbf{x}-\mathbf{z}), \quad R_q = 1$$

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q} a^1_{d,q}a^1_{d',q}k_q(\mathbf{z},\mathbf{z}').$$

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

**Intrinsic Coregionalization Model** (Goovaerts, 1997) or **Multi-task Gaussian Processes** (Bonilla et al., 2008)

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

**Intrinsic Coregionalization Model** (Goovaerts, 1997) or **Multi-task Gaussian Processes** (Bonilla et al., 2008)

$$G^i_{d,q}(\mathbf{x}-\mathbf{z}) = a^i_{d,q}\delta(\mathbf{x}-\mathbf{z}), \quad Q = 1$$

# A general framework

Assume we have $D$ outputs, $\{f_d(\mathbf{x})\}_{d=1}^{D}$. The covariance between $f_d(\mathbf{x})$ and $f_{d'}(\mathbf{x}')$ follows (Higdon, 2002; Boyle and Frean, 2005; Álvarez et al., 2012)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

Some particular cases:

**Intrinsic Coregionalization Model** (Goovaerts, 1997) or **Multi-task Gaussian Processes** (Bonilla et al., 2008)

$$G^i_{d,q}(\mathbf{x}-\mathbf{z}) = a^i_{d,q}\delta(\mathbf{x}-\mathbf{z}), \quad Q = 1$$

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{R_q} a^i_{d,1}a^i_{d',1}k_1(\mathbf{z},\mathbf{z}').$$

# Modifications and Extensions

Some extensions of the following covariance have been summarized by Calder and Cressie (2007)

# Modifications and Extensions

Some extensions of the following covariance have been summarized by Calder and Cressie (2007)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

# Modifications and Extensions

Some extensions of the following covariance have been summarized by Calder and Cressie (2007)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

**Covariance of the latent process**: exponentiated quadratic (EQ) covariance, OU covariance, Brownian motion covariance, to name a few.

# Modifications and Extensions

Some extensions of the following covariance have been summarized by Calder and Cressie (2007)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

**Covariance of the latent process**: exponentiated quadratic (EQ) covariance, OU covariance, Brownian motion covariance, to name a few.

**Non-stationarity** In general, we write $G^i_{d,q}(\mathbf{x},\mathbf{z},\theta_g)$, and $k_q(\mathbf{x},\mathbf{z},\theta_q)$.

# Modifications and Extensions

Some extensions of the following covariance have been summarized by Calder and Cressie (2007)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

**Covariance of the latent process**: exponentiated quadratic (EQ) covariance, OU covariance, Brownian motion covariance, to name a few.

**Non-stationarity** In general, we write $G^i_{d,q}(\mathbf{x},\mathbf{z},\theta_g)$, and $k_q(\mathbf{x},\mathbf{z},\theta_q)$.

$$G^i_{d,q}(\mathbf{x},\mathbf{z},\theta_g(\mathbf{x},\mathbf{z})),$$

# Modifications and Extensions

Some extensions of the following covariance have been summarized by Calder and Cressie (2007)

$$k_{f_d,f_{d'}}(\mathbf{x},\mathbf{x}') = \sum_{q=1}^{Q}\sum_{i=1}^{R_q}\int_{\mathcal{X}} G^i_{d,q}(\mathbf{x}-\mathbf{z})\int_{\mathcal{X}} G^i_{d',q}(\mathbf{x}'-\mathbf{z}')k_q(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'\mathrm{d}\mathbf{z}.$$

**Covariance of the latent process**: exponentiated quadratic (EQ) covariance, OU covariance, Brownian motion covariance, to name a few.

**Non-stationarity** In general, we write $G^i_{d,q}(\mathbf{x},\mathbf{z},\theta_g)$, and $k_q(\mathbf{x},\mathbf{z},\theta_q)$.

$$G^i_{d,q}(\mathbf{x},\mathbf{z},\theta_g(\mathbf{x},\mathbf{z})), \qquad k_q(\mathbf{x},\mathbf{z},\theta_q(\mathbf{x},\mathbf{z}))$$

# Discrete convolutions

- Murray-Smith and Pearlmutter (2005) introduced the idea of transforming a Gaussian process prior using a discretized version of the integral operator of the equation before.

- Applications mentioned include fusing the information from multiple sensors or for solving inverse problems in reconstruction of images (Shi et al., 2005).

# Contenido

# Differential equations and statistics

- Differential equations are the cornerstone of a diverse range of engineering fields and applied sciences.

- Combination with probabilistic models and use within machine learning and statistics has been less explored.

- *Functional data analysis* (Ramsay and Silverman, 2005).

# Classical point of view in frequentist statistics

- Given a differential equation with unknown coefficients $\{\mathbf{A}_m\}_{m=0}^{M}$, how do we use data to fit those parameters?

- Classical approaches to fitting parameters $\boldsymbol{\theta} \ (= \{\mathbf{A}_m\}_{m=0}^{M})$ of differential equations observed data include (Ramsay et al., 2007; Brewer et al., 2008)
  - Numerical approximations of initial value problems.
  - collocation methods

# Numerical approximations

- Iterative process given an initial set of parameters $\theta_0$ and initial conditions $\mathbf{y}_0$,
- A numerical method is used to solve the differential equation.
- The parameters of the differential equation are then optimized by minimizing an error criterion between the approximated solution and the observed data.

# Collocation methods (I)

- The solution of the differential equation is approximated using a set of basis functions, $\{\phi_i(t)\}_{i=1}^J$, this is $y(t) = \sum_{i=1}^J \beta_i \phi_i(t)$.

- The basis functions must be sufficiently smooth so that to compute $\mathcal{D}^m y(t) = \sum \beta_i \mathcal{D}^m \phi_i(t)$.

- Collocation methods also use an iterative procedure for fitting the additional parameters involved in the differential equation.

- Minimization of an error criteria is used to estimate the parameters of the differential equation.

# Collocation methods (II)

- Principal differential analysis (PDA) (Ramsay, 1996) is one example of a collocation method in which the basis functions are *splines*.

- Parameters are obtained by minimizing the squared residuals of the higher order derivative $\mathcal{D}^M y(t)$, and the weighted sum of derivatives $\{\mathcal{D}^m y(t)\}_{m=0}^{M-1}$.

# Collocation methods (III)

- Collocation method augmented with Gaussian process priors (Graepel, 2003).
- The solution of the differential equation $\mathcal{D}_0^M y(t)$ is assumed to follow a Gaussian process prior with covariance $k_{\mathcal{D}_{0,t}^M,\mathcal{D}_{0,t'}^M}(t,t')$
- An approximated solution $\widetilde{y}(t)$ can then be computed through the expansion $\widetilde{y}(t)=\sum_{n=1}^N \alpha_n k_{\mathcal{D}_{0,t}^M,\mathcal{D}_{0,t'}^M}(t,t_n)$.
- $\alpha_n$ is an element of the vector $(\mathbf{K}_{\mathcal{D}_{0,t}^M,\mathcal{D}_{0,t'}^M}+\sigma_y^2\mathbf{I}_N)^{-1}\widehat{\mathbf{y}}$, and $\widehat{\mathbf{y}}$ is a vector of noisy observations of $\mathcal{D}_0^M y(t)$.

# Contenido

# Systems Identification

- In control engineering, systems identification refers to a set of techniques used for representing a dynamical system by a mathematical model (mostly a linear model).

- A detailed description of the dynamical system is usually unknown, and parameters of the surrogate model are estimated from measured data.

# GPs for systems identification

- Gaussian processes have been used as models for systems identification (Solak et al., 2003; Kocijan et al., 2005; Calderhead et al., 2009; Thompson, 2009).

- A non-linear dynamical system is linearized around an equilibrium point by means of a Taylor series expansion (Thompson, 2009), $y(t) = \sum_{j=0}^{\infty} \frac{y^{(j)}(a)}{j!}(t-a)^j$, with $a$ the equilibrium point.

- Covariates correspond to the terms $(t-a)^j$ and the derivatives $y^{(j)}(a)$ to regression coefficients.

- The derivatives are assumed to follow a Gaussian process prior with a covariance function that is obtained as $k^{(j,j')}(t,t')$.

# State space models

- Representation of the possible nonlinear relationships between the latent space and between the latent space and the observation space (Ko et al., 2007; Deisenroth et al., 2009; Turner et al., 2010).

- Latent process is a GP (Hartikainen and Särkkä, 2011; Hartikainen et al., 2012), improving computational complexity.

# References I


Álvarez, M. A., Rosasco, L., Lawrence, N. D., 2012. Kernels for vector-valued functions: a review. Foundations and Trends® in Machine Learning 4 (3), 195–266.

Bonilla, E. V., Chai, K. M., Williams, C. K. I., 2008. Multi-task Gaussian process prediction. In: Platt, J. C., Koller, D., Singer, Y., Roweis, S. (Eds.), NIPS. Vol. 20. MIT Press, Cambridge, MA.

Boyle, P., Frean, M., 2005. Dependent Gaussian processes. In: Saul, L., Weiss, Y., Bouttou, L. (Eds.), NIPS. Vol. 17. MIT Press, Cambridge, MA, pp. 217–224.

Brewer, D., Barenco, M., Callard, R., Hubank, M., Stark, J., 2008. Fitting ordinary differential equations to short time course data. Philosophical Transactions of the Royal Society A 366, 519–544.

Calder, C. A., Cressie, N., August 2007. Some topics in convolution-based spatial modeling. In: Proceedings of the 56th Session of the International Statistics Institute.

Calderhead, B., Girolami, M., Lawrence, N. D., 2009. Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (Eds.), NIPS. Vol. 21. MIT Press, Cambridge, MA, pp. 217–224.

Cressie, N. A. C., 1993. Statistics for Spatial Data. John Wiley & Sons (Revised edition), USA.

Deisenroth, M. P., Huber, M. F., Hanebeck, U. D., 2009. Analytic moment-based Gaussian process filtering. In: Proceedings of the 26th International Conference on Machine Learning (ICML 2009). Omnipress, pp. 225–232.

Goovaerts, P., 1997. Geostatistics For Natural Resources Evaluation. Oxford University Press, USA.

Graepel, T., August 21-24 2003. Solving noisy linear operator equations by Gaussian processes: Application to ordinary and partial differential equations. In: Fawcett, T., Mishra, N. (Eds.), Proceedings of the Twentieth International Conference on Machine Learning. AAAI Press, Washington, DC, USA, pp. 234–241.

Hartikainen, J., Särkkä, S., 2011. Sequential inference for latent force models. In: Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence. UAI 2011. pp. 311–318.

Hartikainen, J., Seppänen, M., Särkkä, S., 2012. State-space inference for non-linear latent force models with application to satellite orbit prediction. In: Proceedings of the 29th International Conference on Machine Learning. ICML 2012.

Higdon, D. M., 2002. Space and space-time modelling using process convolutions. In: Anderson, C., Barnett, V., Chatwin, P., El-Shaarawi, A. (Eds.), Quantitative methods for current environmental issues. Springer-Verlag, pp. 37–56.

Journel, A. G., Huijbregts, C. J., 1978. Mining Geostatistics. Academic Press, London.

Ko, J., Klein, D. J., Fox, D., Haehnel, D., 2007. GP-UKF: Unscented Kalman filters with Gaussian process prediction and observation models. In: Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Diego, CA, USA, pp. 1901–1907.

Kocijan, J., Girard, A., Banko, B., Murray-Smith, R., 2005. Dynamic systems identification with Gaussian processes. Mathematical and Computer Modelling of Dynamical Systems 11 (4), 411–424.

# References II


Murray-Smith, R., Pearlmutter, B. A., 2005. Transformation of Gaussian process priors. In: Winkler, J., Niranjan, M., Lawrence, N. (Eds.), Deterministic and Statistical Methods in Machine Learning. LNAI 3635, Springer-Verlag, pp. 110–123.

Ramsay, J. O., 1996. Principal differential analysis: Data reduction by differential operators. Journal of the Royal Statistical Society. Series B (Methodological) 58 (3), 495–508.

Ramsay, J. O., Hooker, G., Campbell, D., Cao, J., 2007. Parameter estimation for differential equations: a generalized smoothing approach. Journal of the Royal Statistical Society. Series B (Methodological) 69 (5), 741–796.

Ramsay, J. O., Silverman, B. W., 2005. Functional Data Analysis, 2nd Edition. Springer Series in Statistics, New York, NY, (USA).

Shi, J. Q., Murray-Smith, R., Titterington, D., Pearlmutter, B., 2005. Learning with large data sets using filtered Gaussian process priors. In: Murray-Smith, R., Shorten, R. (Eds.), Proceedings of the Hamilton Summer School on Switching and Learning in Feedback systems. LNCS 3355, Springer-Verlag, pp. 128–139.

Solak, E., Murray-Smith, R., Leithead, W. E., Leith, D. J., Rasmussen, C. E., 2003. Derivative observations in Gaussian process models of dynamic systems. In: Becker, S., Thrun, S., Obermayer, K. (Eds.), NIPS. Vol. 15. MIT Press, Cambridge, MA, pp. 1033–1040.

Teh, Y. W., Seeger, M., Jordan, M. I., 6-8 January 2005. Semiparametric latent factor models. In: Cowell, R. G., Ghahramani, Z. (Eds.), AISTATS 10. Society for Artificial Intelligence and Statistics, Barbados, pp. 333–340.

Thompson, K. R., 2009. Implementation of Gaussian process models for nonlinear system identification. Ph.D. thesis, Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow, UK.

Turner, R., Deisenroth, M. P., Rasmussen, C. E., 13-15 May 2010. State-space inference and learning with Gaussian processes. In: Teh, Y. W., Titterington, M. (Eds.), Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR W&CP 9, Chia Laguna, Sardinia, Italy, pp. 868–875.

Wackernagel, H., 2003. Multivariate Geostatistics. Springer-Verlag Heidelberg New york.