Latent Force Models

Mauricio Alvarez[1] David Luengo[1,2] Neil D. Lawrence[1]

[1] School of Computer Science, University of Manchester.
[2] Dep. Teoría de Señal y Comunicaciones, Universidad Carlos III de Madrid.

Data driven paradigm

- Traditionally, the main focus in machine learning has been model generation through a *data driven paradigm*.

- Combine a data set with a flexible class of models and, through regularization, make predictions on unseen data.

- Problems
  - Data is scarce relative to the complexity of the system.
  - Model is forced to extrapolate.

Mechanistic models

- Models inspired by the underlying knowledge of a physical system are common in many areas.
- Description of a well characterized physical process that underpins the system, typically represented with a set of differential equations.
- Identifying and specifying all the interactions might not be feasible.
- A mechanistic model can enable accurate prediction in regions where there may be no available training data

Hybrid systems

- We suggest a *hybrid approach* involving a mechanistic model of the system augmented through machine learning techniques.
- Dynamical systems (e.g. incorporating first order and second order differential equations).
- Partial differential equations for systems with multiple inputs.

Latent variable model: definition

- Our approach can be seen as a type of latent variable model.

$$\mathbf{Y} = \mathbf{F}\mathbf{W} + \mathbf{E},$$

where $\mathbf{Y} \in \mathbb{R}^{N \times Q}$, $\mathbf{F} \in \mathbb{R}^{N \times R}$, $\mathbf{W} \in \mathbb{R}^{R \times Q}$ $(R < Q)$ and $\mathbf{E}$ is a matrix variate white Gaussian noise with columns $\mathbf{e}_{:,q} \sim \mathcal{N}(\mathbf{0}, \Sigma)$.

- In PCA and FA the common approach to deal with the unknowns is to integrate out $\mathbf{F}$ under a Gaussian prior and optimize with respect to $\mathbf{W}$.

Latent variable model: alternative view

- Data with temporal nature and Gaussian (Markov) prior for rows of **F** leads to the Kalman filter/smoother.

- Consider a joint distribution for $p(\mathbf{F}|\mathbf{t})$, $\mathbf{t} = [t_1 \dots t_N]^\top$, with the form of a Gaussian process (GP),

$$p(\mathbf{F}|\mathbf{t}) = \prod_{r=1}^{R} \mathcal{N}\left(\mathbf{f}_{:,r}|\mathbf{0}, \mathbf{K}_{f_{:,r},f_{:,r}}\right),$$

  The latent variables are random functions, $\{f_r(t)\}_{r=1}^{R}$ with associated covariance $\mathbf{K}_{f_{:,r},f_{:,r}}$.

- The GP for **Y** can be readily implemented. In [**?**] this is known as a semi-parametric latent factor model (SLFM).

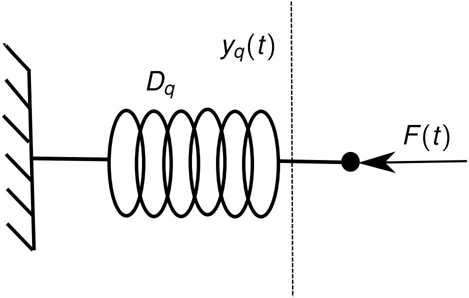
Latent force model: mechanistic interpretation (1)

- We include a further dynamical system with a *mechanistic* inspiration.
- Reinterpret equation $\mathbf{Y} = \mathbf{F}\mathbf{W} + \mathbf{E}$, as a force balance equation

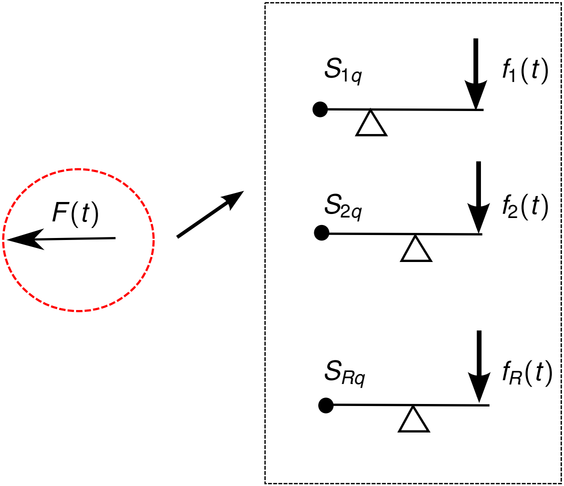
$$\mathbf{Y}\mathbf{D} = \mathbf{F}\mathbf{S} + \widetilde{\mathbf{E}},$$

where $\mathbf{S} \in \mathbb{R}^{R \times Q}$ is a matrix of sensitivities, $\mathbf{D} \in \mathbb{R}^{Q \times Q}$ is diagonal matrix of spring constants, $\mathbf{W} = \mathbf{S}\mathbf{D}^{-1}$ and $\widetilde{\mathbf{e}}_{:,q} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{D}^{\top}\boldsymbol{\Sigma}\mathbf{D}\right)$.

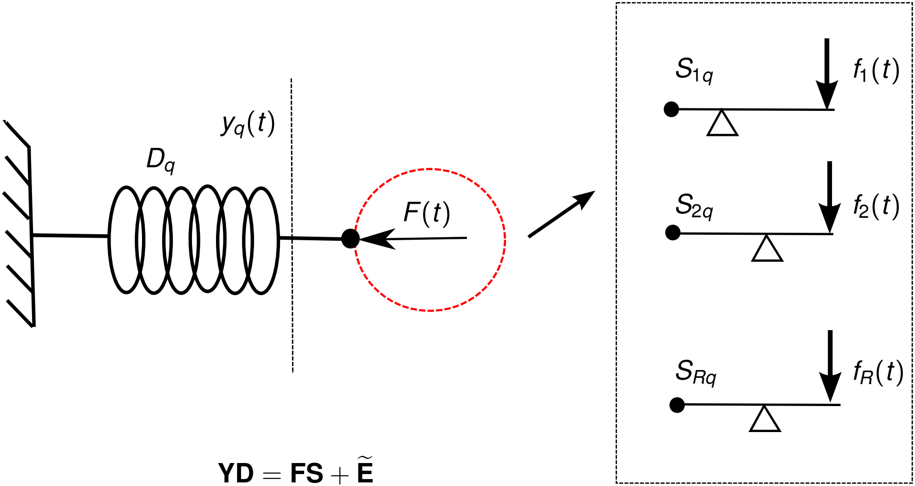
Latent force model: mechanistic interpretation (2)



Latent force model: mechanistic interpretation (2)



Latent force model: mechanistic interpretation (2)



$$\mathbf{YD} = \mathbf{FS} + \widetilde{\mathbf{E}}$$

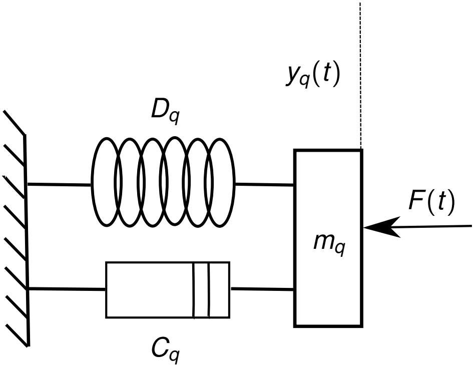
Latent force model: extension (1)

- The model can be extended including dampers and masses.

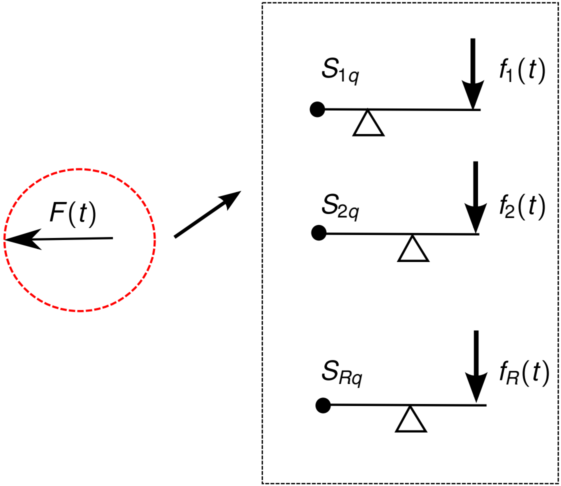
- We can write
$$\mathbf{YD} + \dot{\mathbf{Y}}\mathbf{C} + \ddot{\mathbf{Y}}\mathbf{M} = \mathbf{FS} + \widehat{\boldsymbol{E}},$$
where

$\dot{\mathbf{Y}}$	is the first derivative of **Y** w.r.t. time
$\ddot{\mathbf{Y}}$	is the second derivative of **Y** w.r.t. time
$\mathbf{C}$	is a diagonal matrix of damping coefficients
$\mathbf{M}$	is a diagonal matrix of masses
$\widehat{\boldsymbol{E}}$	is a matrix variate white Gaussian noise.

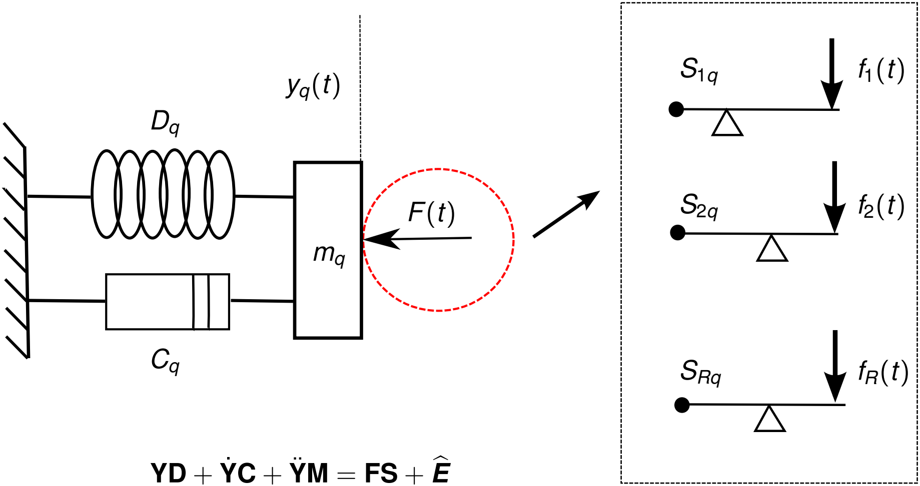
Latent force model: extension (2)



Latent force model: extension (2)



Latent force model: extension (2)



$$\mathbf{YD} + \dot{\mathbf{Y}}\mathbf{C} + \ddot{\mathbf{Y}}\mathbf{M} = \mathbf{FS} + \widehat{\boldsymbol{E}}$$

Latent force model: properties

- This model allows to include behaviors like inertia and resonance.
- We refer to these systems as *latent force models* (LFMs).
- One way of thinking of our model is to consider puppetry.

Second Order Dynamical System

Using the system of second order differential equations

$$m_q \frac{\mathrm{d}^2 y_q(t)}{\mathrm{d}t^2} + C_q \frac{\mathrm{d}y_q(t)}{\mathrm{d}t} + D_q y_q(t) = \sum_{r=1}^{R} S_{rq} f_r(t),$$

where

$f_r(t)$	latent forces
$y_q(t)$	displacements over time
C_q	damper constant for the q-th output
D_q	spring constant for the q-th output
m_q	mass constant for the q-th output
S_{rq}	sensitivity of the q-th output to the r-th input.

Second Order Dynamical System: solution

Solving for $y_q(t)$, we obtain

$$y_q(t) = \frac{B_q}{D_q} + \sum_{r=1}^{R} L_{rq}[f_r](t),$$

where the linear operator is given by a convolution:

$$L_{rq}[f_r](t) = \frac{S_{rq}}{\omega_q} \exp(-\alpha_q t) \int_0^t f_r(\tau) \exp(\alpha_q \tau) \sin(\omega_q (t-\tau)) \mathrm{d}\tau,$$

with $\omega_q = \sqrt{4D_q - C_q^2}/2$ and $\alpha_q = C_q/2$.

Second Order Dynamical System: covariance matrix

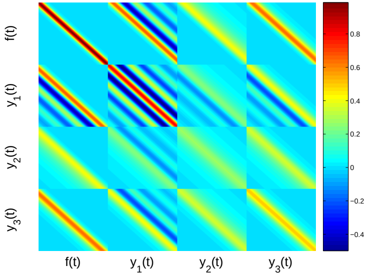
Behaviour of the system summarized by the damping ratio:

$$\zeta_q = \frac{1}{2} C_q / \sqrt{D_q}$$

$\zeta_q > 1$ overdamped system
$\zeta_q = 1$ critically damped system
$\zeta_q < 1$ underdamped system
$\zeta_q = 0$ undamped system (no friction)

Example covariance matrix:

$\zeta_1 = 0.125$ underdamped
$\zeta_2 = 2$ overdamped
$\zeta_3 = 1$ critically damped



Second Order Dynamical System: samples from GP

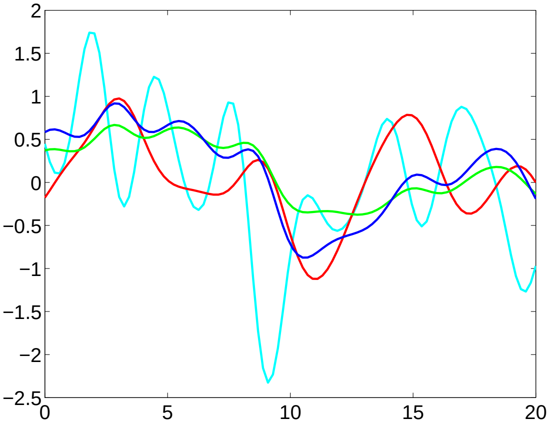


Figure: Joint samples from the ODE covariance, *cyan*: $f(t)$, *red*: $y_1(t)$(underdamped) and *green*: $y_2(t)$ (overdamped) and *blue*: $y_3(t)$ (critically damped).

Second Order Dynamical System: samples from GP

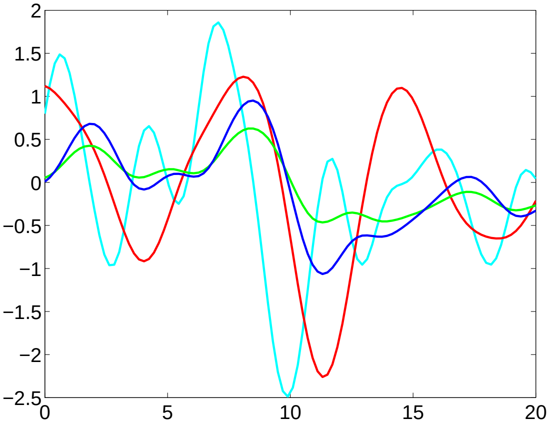


Figure: Joint samples from the ODE covariance, *cyan*: $f(t)$, *red*: $y_1(t)$(underdamped) and *green*: $y_2(t)$ (overdamped) and *blue*: $y_3(t)$ (critically damped).

Second Order Dynamical System: samples from GP

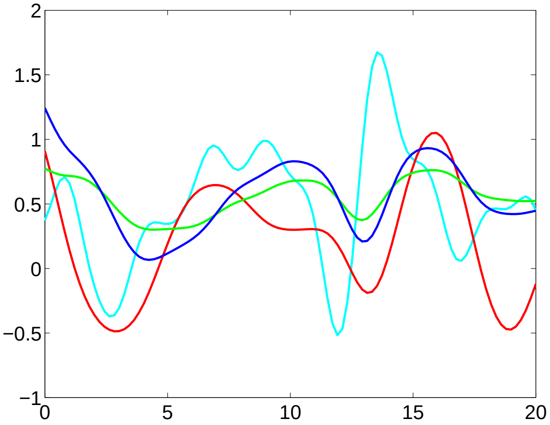


Figure: Joint samples from the ODE covariance, *cyan*: $f(t)$, *red*: $y_1(t)$(underdamped) and *green*: $y_2(t)$ (overdamped) and *blue*: $y_3(t)$ (critically damped).

Second Order Dynamical System: samples from GP

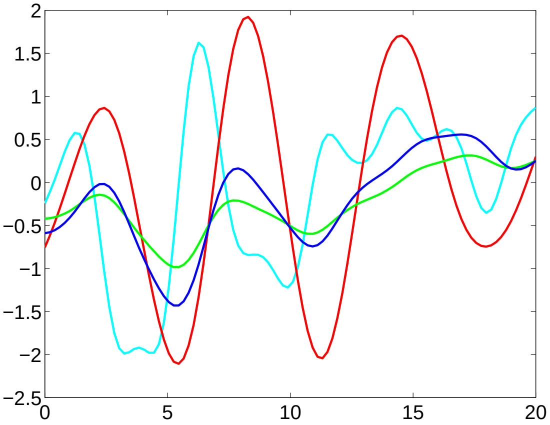


Figure: Joint samples from the ODE covariance, *cyan*: $f(t)$, *red*: $y_1(t)$(underdamped) and *green*: $y_2(t)$ (overdamped) and *blue*: $y_3(t)$ (critically damped).

Motion Capture Data (1)

- CMU motion capture data, motions 18, 19 and 20 from subject 49.
- Motions 18 and 19 for training and 20 for testing.

Motion Capture Data (2)

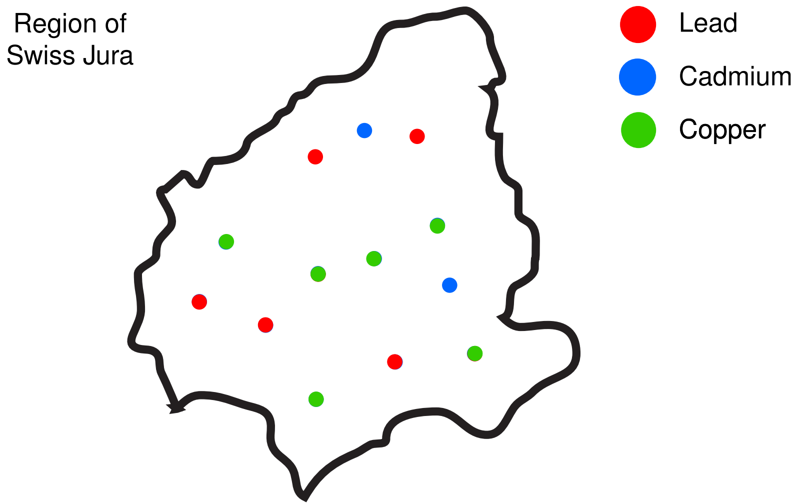
- The data down-sampled by 32 (from 120 frames per second to 3.75).
- We focused on the subject's left arm.
- For testing, we condition only on the observations of the shoulder's orientation (motion 20) to make predictions for the rest of the arm's angles.

Motion Capture Results

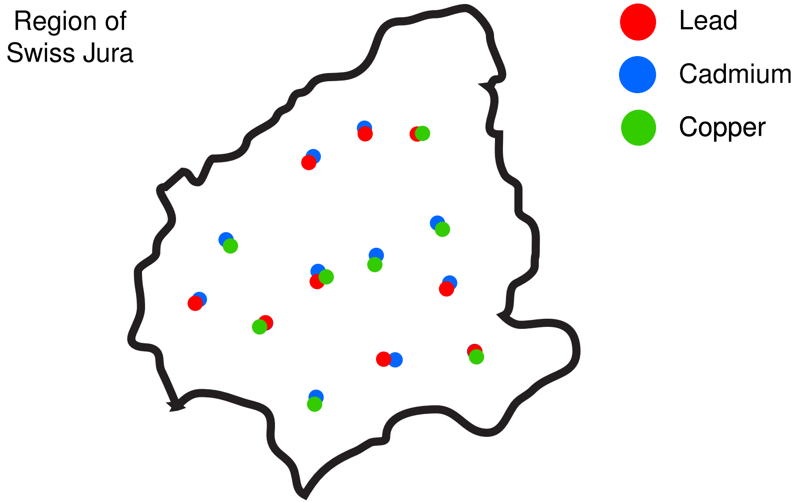
Root mean squared (RMS) angle error for prediction of the left arm's configuration in the motion capture data. Prediction with the latent force model outperforms the prediction with regression for all apart from the radius's angle.

Angle	Latent Force Error	Regression Error
Radius	4.11	**4.02**
Wrist	**6.55**	6.65
Hand X rotation	**1.82**	3.21
Hand Z rotation	**2.76**	6.14
Thumb X rotation	**1.77**	3.10
Thumb Z rotation	**2.73**	6.09

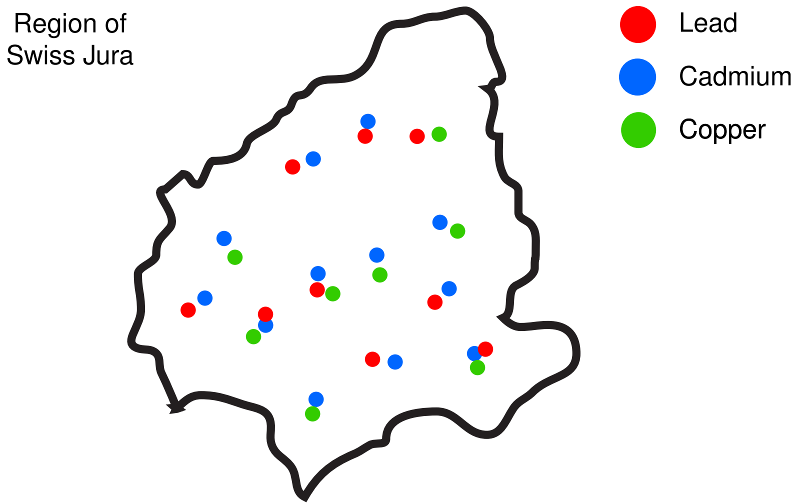
Diffussion in the Swiss Jura



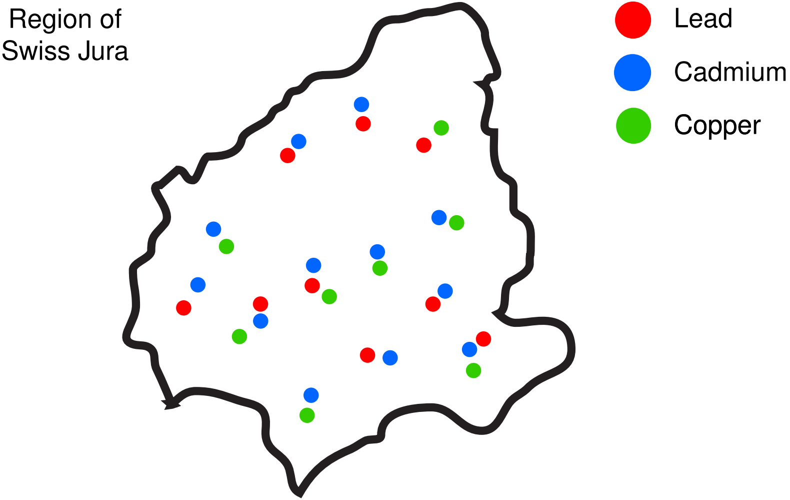
Diffussion in the Swiss Jura



Diffussion in the Swiss Jura



Diffussion in the Swiss Jura



Diffusion equation

- A simplified version of the diffusion equation is

$$\frac{\partial y_q(\mathbf{x},t)}{\partial t} = \sum_{j=1}^{d} \kappa_q \frac{\partial^2 y_q(\mathbf{x},t)}{\partial x_j^2},$$

  where $y_q(\mathbf{x},t)$ are the concentrations of each pollutant.

- The solution to the system is then given by

$$y_q(\mathbf{x},t) = \sum_{r=1}^{R} S_{rq} \int_{\mathbb{R}^d} f_r(\mathbf{x}') G_q(\mathbf{x},\mathbf{x}',t) \mathrm{d}\mathbf{x}'$$

  where $f_r(\mathbf{x})$ represents the concentration of pollutants at time zero and $G_q(\mathbf{x},\mathbf{x}',t)$ is the Green's function given as

$$G_q(\mathbf{x},\mathbf{x}',t) = \frac{1}{2^d \pi^{d/2} T_q^{d/2}} \exp\left[-\sum_{j=1}^{d} \frac{(x_j - x_j')^2}{4T_q}\right],$$

  with $T_q = \kappa_q t$.

Prediction of Metal Concentrations

- Prediction of a *primary variable* by conditioning on the values of some *secondary variables*.

Primary variable	**Secondary Variables**
Cd	Ni, Zn
Cu	Pb, Ni, Zn
Pb	Cu, Ni, Zn
Co	Ni, Zn

- Comparison bewteen diffusion kernel, independent GPs and “ordinary co-kriging”.

Metals	IGPs	GPDK	OCK
Cd	0.5823±0.0133	**0.4505±0.0126**	0.5
Cu	15.9357±0.0907	**7.1677±0.2266**	7.8
Pb	22.9141±0.6076	**10.1097±0.2842**	10.7
Co	2.0735±0.1070	1.7546±0.0895	**1.5**

Discussion

- Hybrid approach for the use of simple mechanistic models with Gaussian processes.
- Convolution processes not for multi-ouput regression but to augment data-driven models with characteristics of physical systems.
- Gaussian process as meaningful prior distributions.
- Other applications considered:
  - Bioinformatics: transcription factor networks [AISTATS'09].
  - Financial time series: foreign currency exchange [Learning'09].

Acknowledgments

DL has been partly financed by Comunidad de Madrid (project PRO-MULTIDIS-CM, S-0505/TIC/0233), and by the Spanish government (CICYT project TEC2006-13514-C02-01 and researh grant JC2008-00219). MA and NL have been financed by a Google Research Award and EPSRC Grant No EP/F005687/1 “Gaussian Processes for Systems Identification with Applications in Systems Biology”. MA received a travel grant from Nokia.

References I

Yee Whye Teh, Matthias Seeger, and Michael I. Jordan.
Semiparametric latent factor models.
In Robert G. Cowell and Zoubin Ghahramani, editors, *AISTATS 10*, pages 333–340, Barbados, 6-8 January 2005. Society for Artificial Intelligence and Statistics.