(Sequential) Non-parametric Bayesian inference for Time Series Modelling

Stephen Roberts, Mike Osborne, Steve Reece, Mark Ebden, Roman Garnett, Neale Gibson, Suzanne Aigrain, Tom Evans, Aris Karastergiou, Alex Rogers

Pattern Analysis & Machine Learning Research Group.

University of Oxford

The lecture aims

- *Bayesian inference* has profound impact in the principled handling of uncertainty in *practical* computation

- What this lecture aims to do:
  - Give a conceptual overview of Bayesian inference applied to real-world problems in time-series modelling
  - Introduce **Gaussian Processes**

- What it does not aim to do:
  - Give endless equations – these are important and elegant, but are in publications and texts.

PART I : Bayesian basics, a gentle conceptual overview

Met Office got it wrong over ban on flights

By Caroline Gammell
David Millward
and Bruno Waterfield

THE Met Office was blamed last night for triggering the "unnecessary" six-day closure of British airspace that has cost airlines, passengers and the economy more than £1.5 billion.

The government agency was accused of using a scientific model based on "probability" rather than fact to forecast the spread of the volcanic ash cloud that made Europe a no-fly zone and ruined the plans of more than 2.5 million trav-

£1m flotilla

Britain is to spend up to £1 million using three Royal Navy ships to rescue British travellers stranded by the ash cloud, despite thousands of spaces available on cross-Channel ferries and the Eurostar.

Two carriers, HMS Ark Royal and HMS Ocean, and the landing ship HMS Albion are being sent to rescue passengers

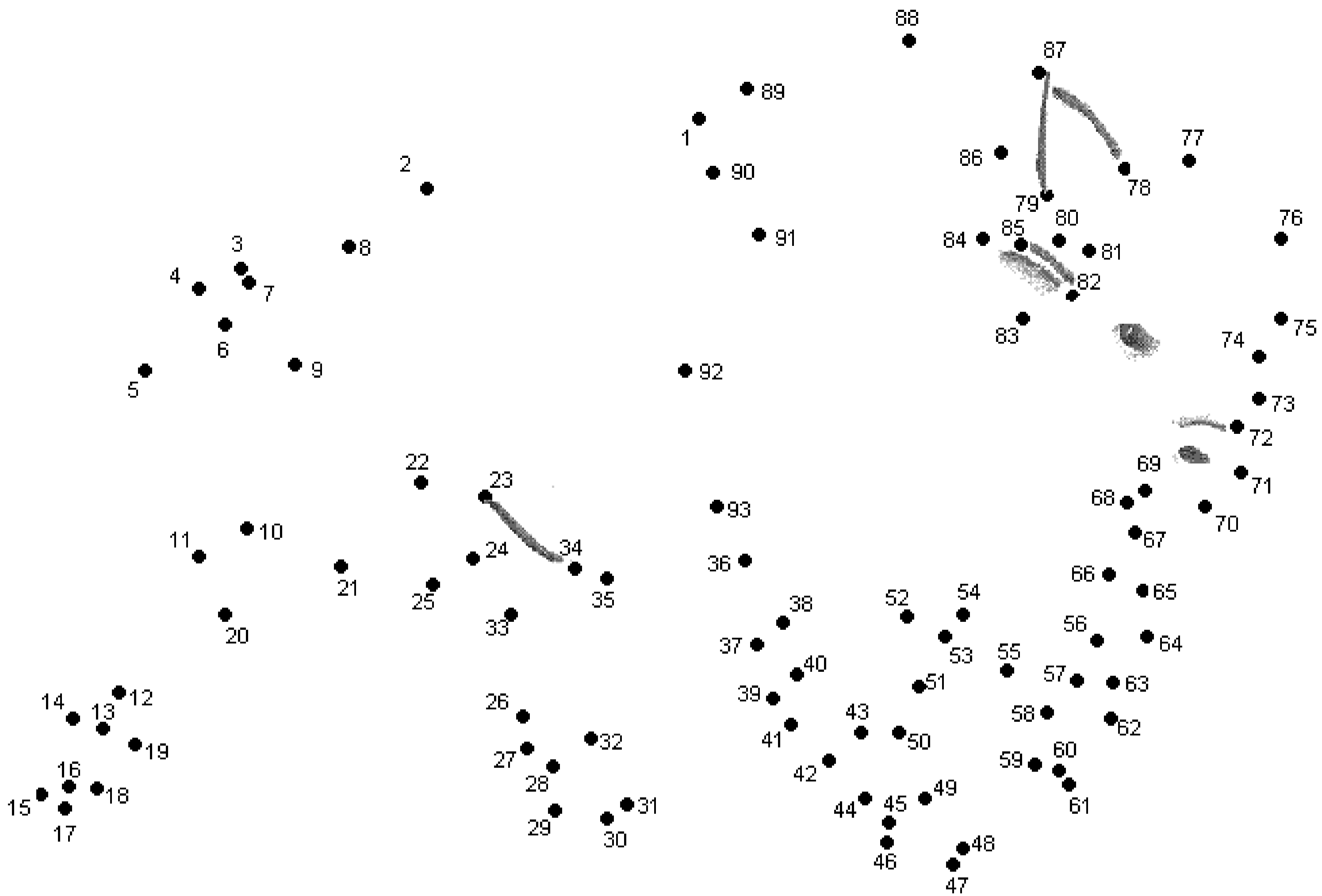
probability rather than actual things happening."

Mr Ruete said the commission had to intervene to allow airlines to make test flights in order to check the VAAC data

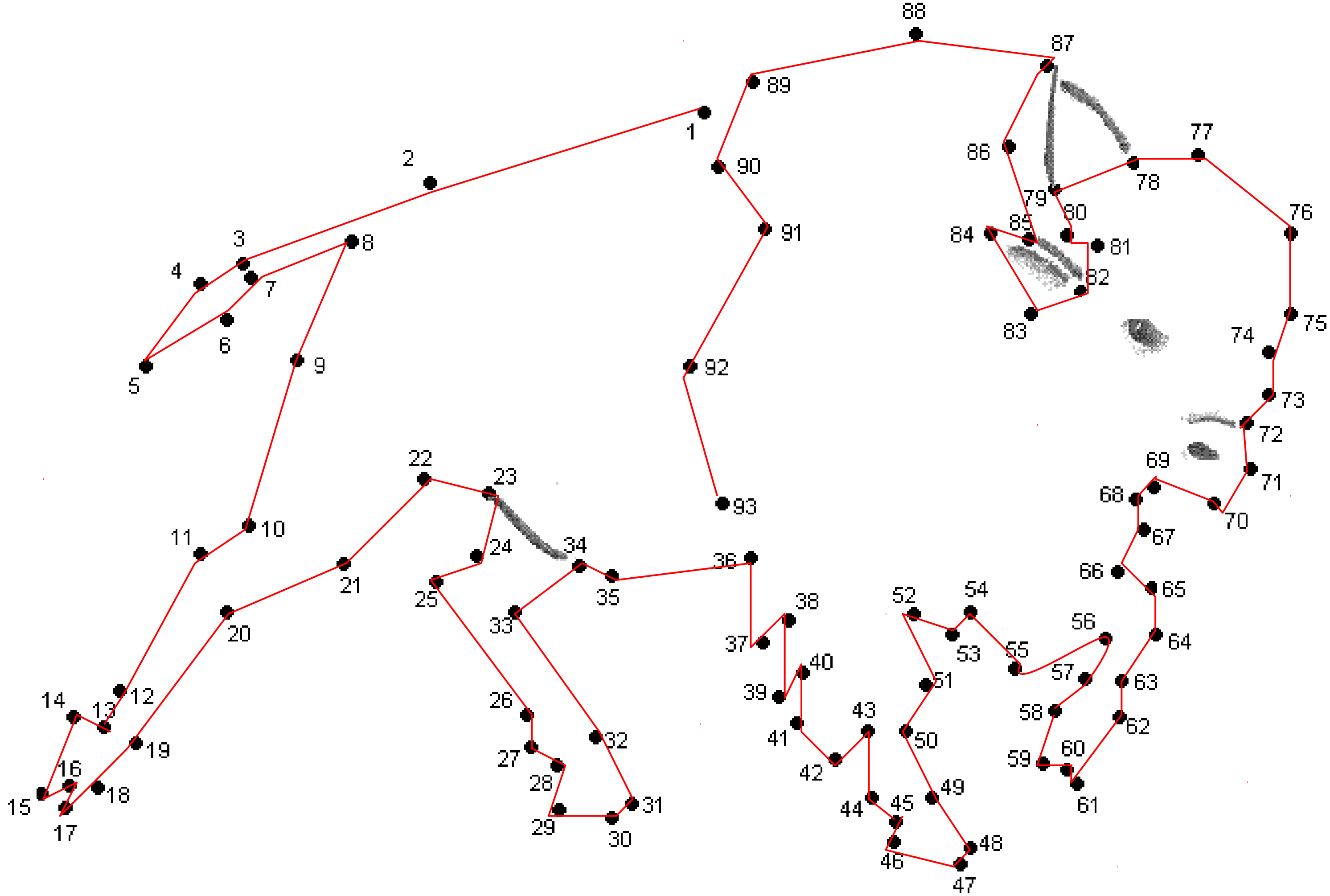
"unnecessary" six-day closure of British airspace that has cost airlines, passengers and the economy more than £1.5 billion.

The government agency was accused of using a scientific model based on "probability" rather than fact to forecast the spread of the volcanic ash cloud that made Europe a no-fly zone and ruined the plans

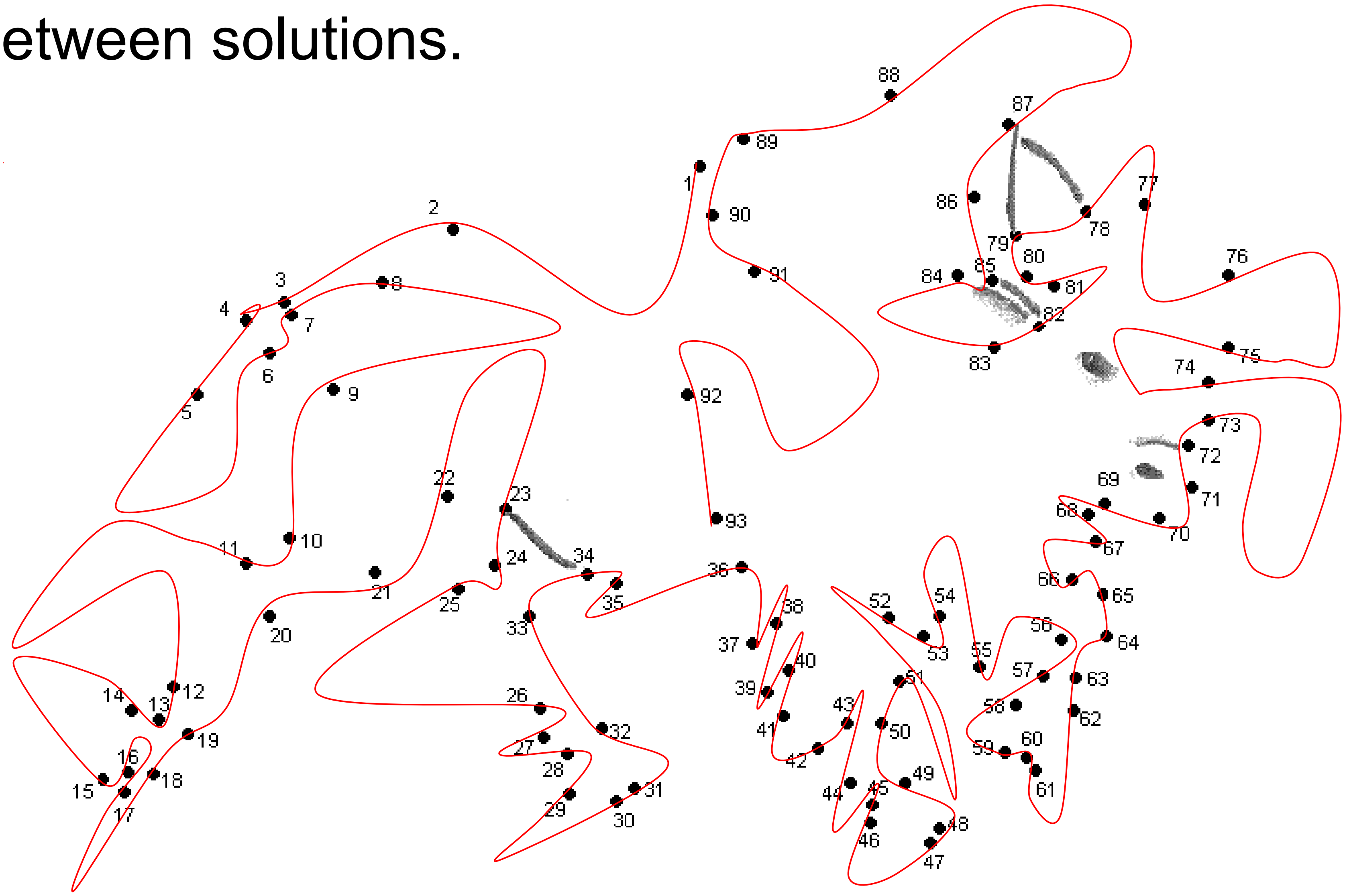
A dot-to-dot is an inference problem.



A dot-to-dot is a problem with many possible solutions.



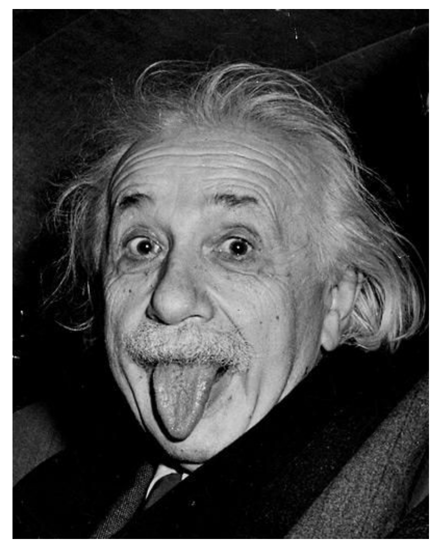
Our prior information allows us to discriminate between solutions.



Occam's Razor

Numquam ponenda est pluralitas sine necessitate - “Plurality must never be posited without necessity”

"Everything should be kept as simple as possible, but no simpler."



Core methodology

- Bayesian modelling allows for explicit incorporation of all *desiderata*
- Effort focused not only on theory development, but algorithmic implementations that are *timely & practical* for *real-world, real-time scenarios*

- Single, under- and over-arching philosophy…

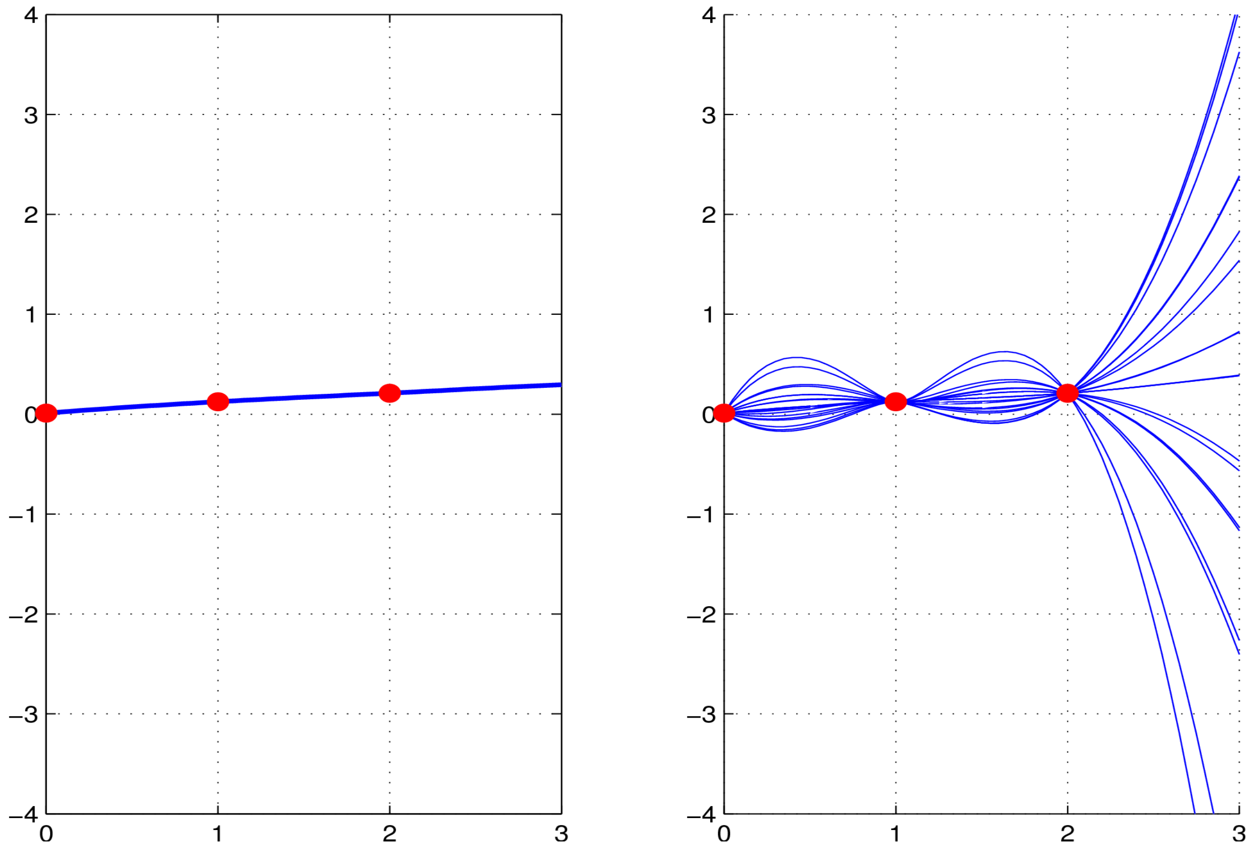
  "one method to rule them all… and in the darkness bind them"

What does this buy us?

- *Uncertainty* at all levels of inference is naturally *taken into account*
- *Optimal fusion* of information: subjective, objective
- Handling *missing values*
- Handling of *noise*
- *Principled* inference of *confidence* and *risk*
- *Optimal d*ecision making

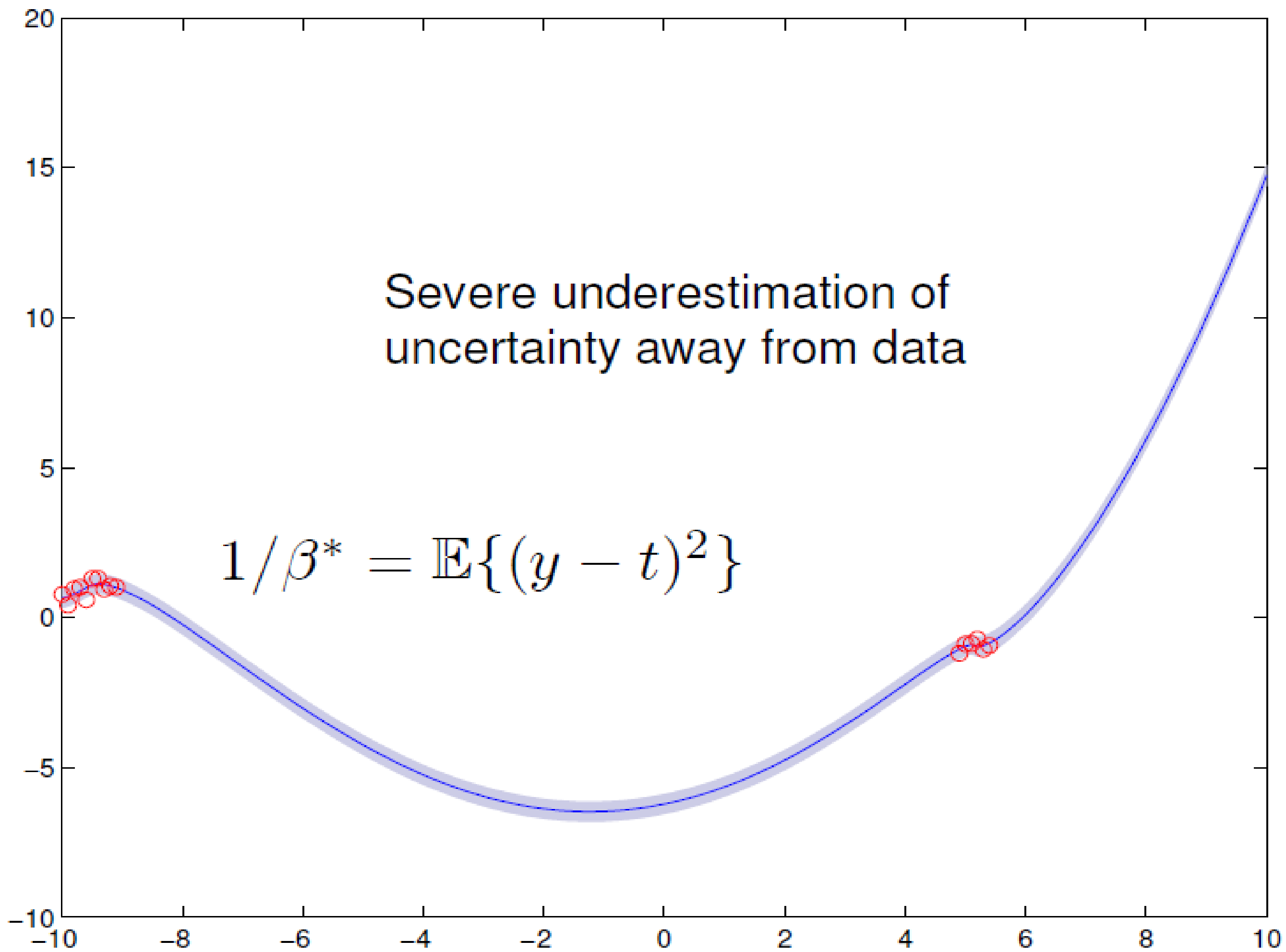
PART II: more details

The right model?

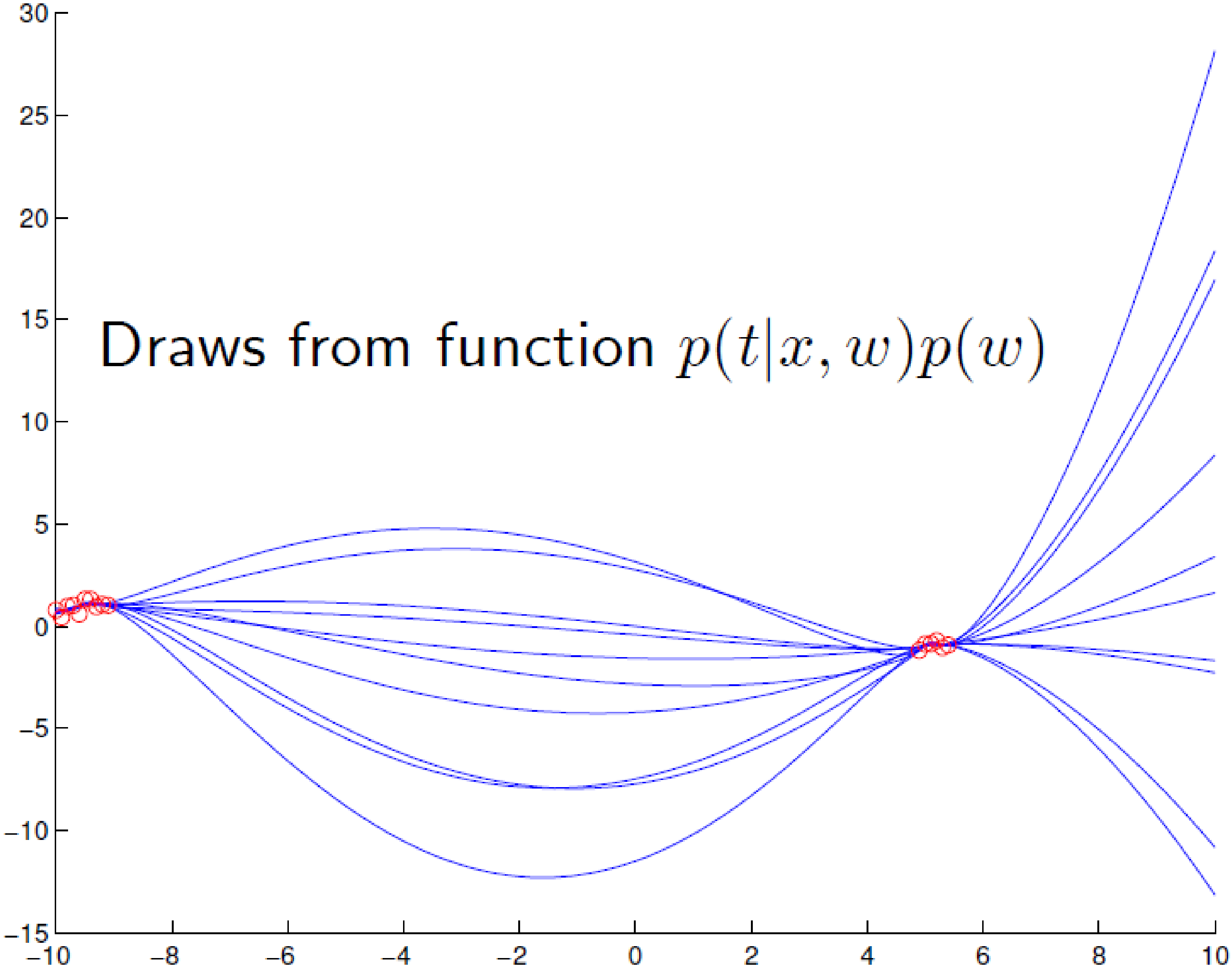


All these models explain the data equally well...

Maximum-likelihood solution



Draws from posterior



Bayesian marginal integral

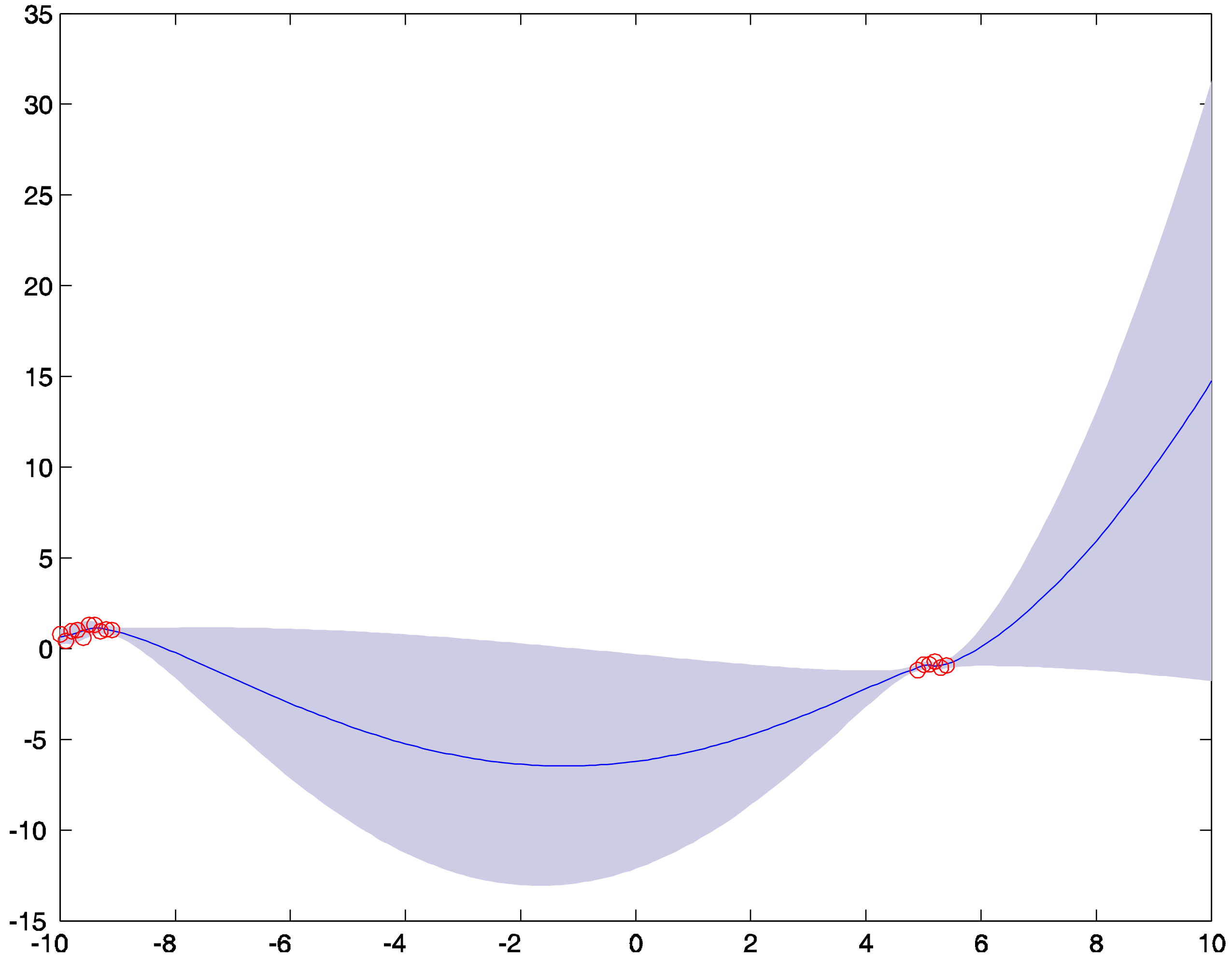
- Key equation of Bayesian inference

$$\boxed{p(t|x) = \int_{w,\beta} p(t|x,w,\beta)p(w)p(\beta)\mathrm{d}w\mathrm{d}\beta}$$

- Expend computational effort *integrating* not *optimising*
- Obtain full *predictive distribution*
- Maximum likelihood II: integrate out parameters, re-estimate hyperparameters

$$p(t|x,\beta^*) = \int_{w} p(t|x,w,\beta^*)p(w)\mathrm{d}w$$

Bayesian solution

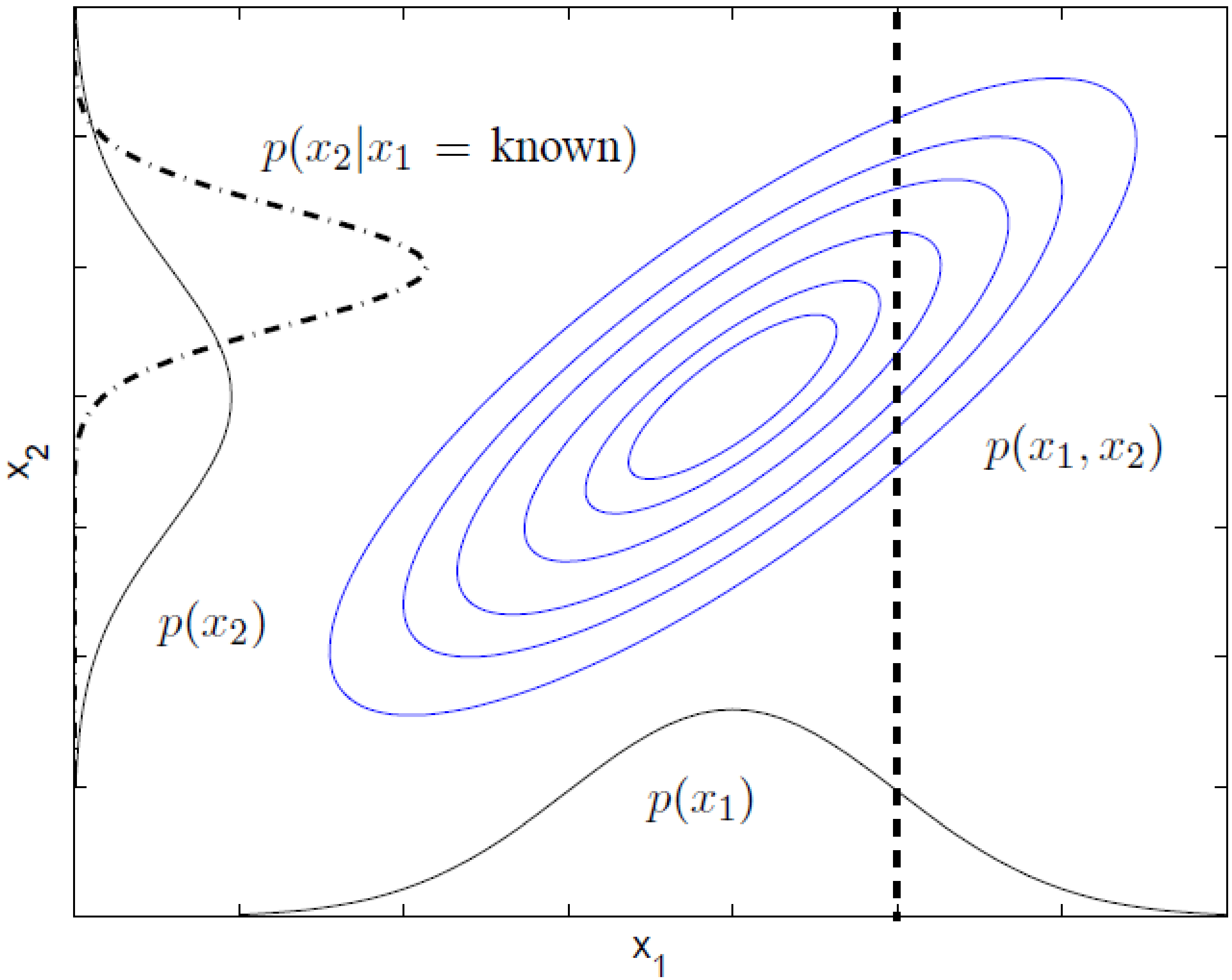


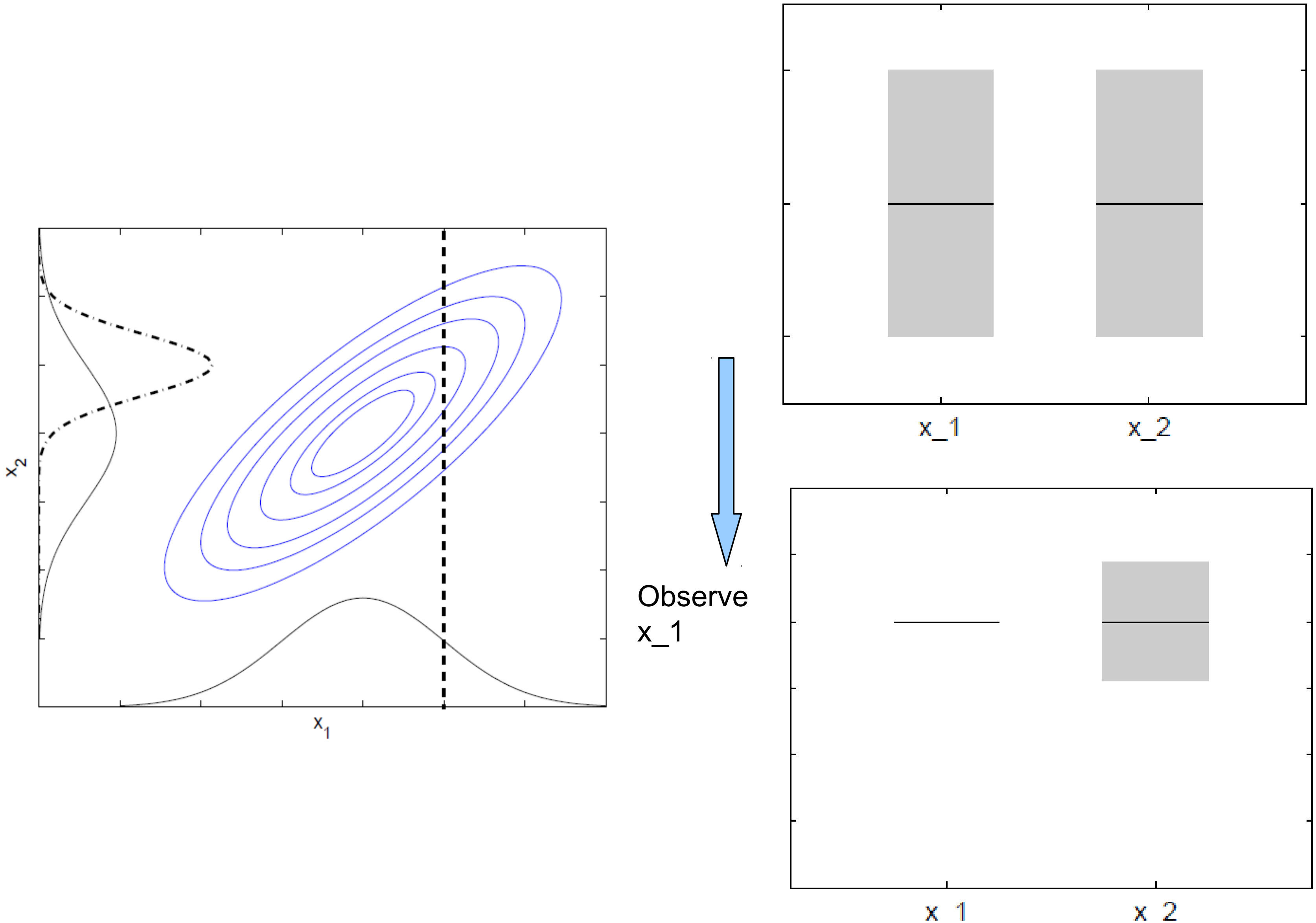
Complexity...

- Even when the models explain the data equally well, we somehow are urged to favour those that are “simpler”
- What we really want is somehow to work with probabilities over functions...
- Amazingly, Bayesian non-parametrics allows us to do just this!
- We now delve into the world of *Gaussian Processes*

Part III : Gaussian Processes

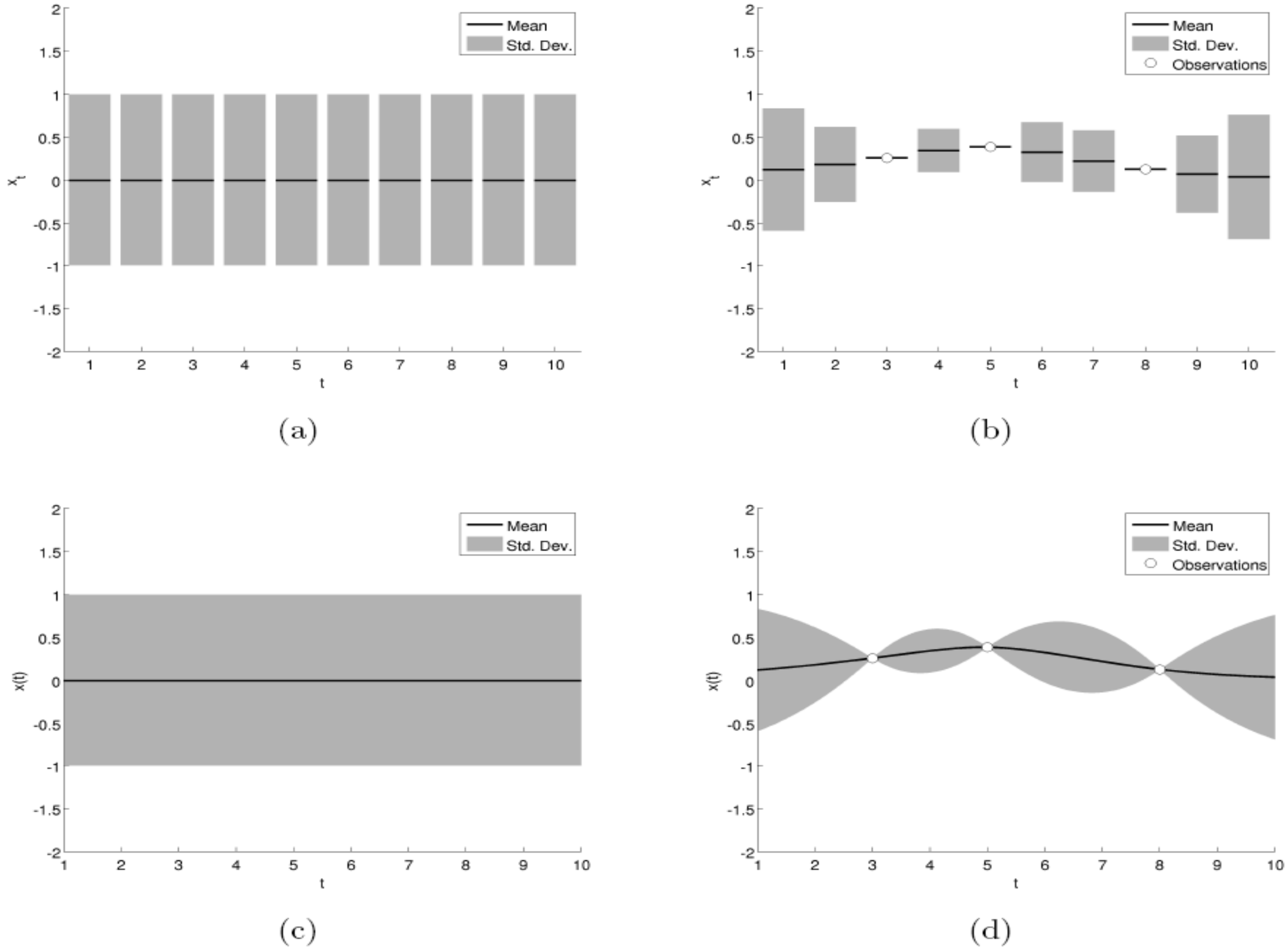
The humble (but useful) Gaussian



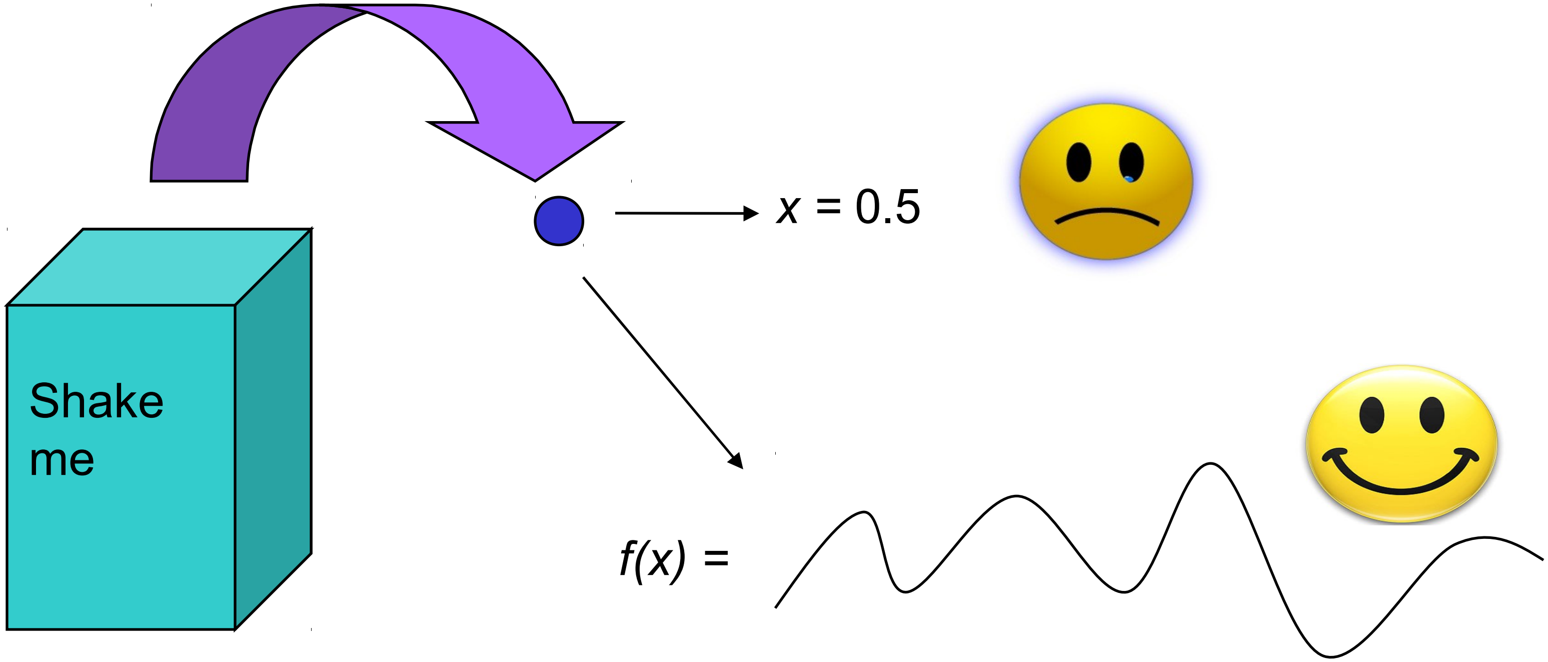


x_2
x_1
Observe
x_1
x_1
x_2
x_1
x_2

Extend to continuous variable



Probabilities over functions not samples

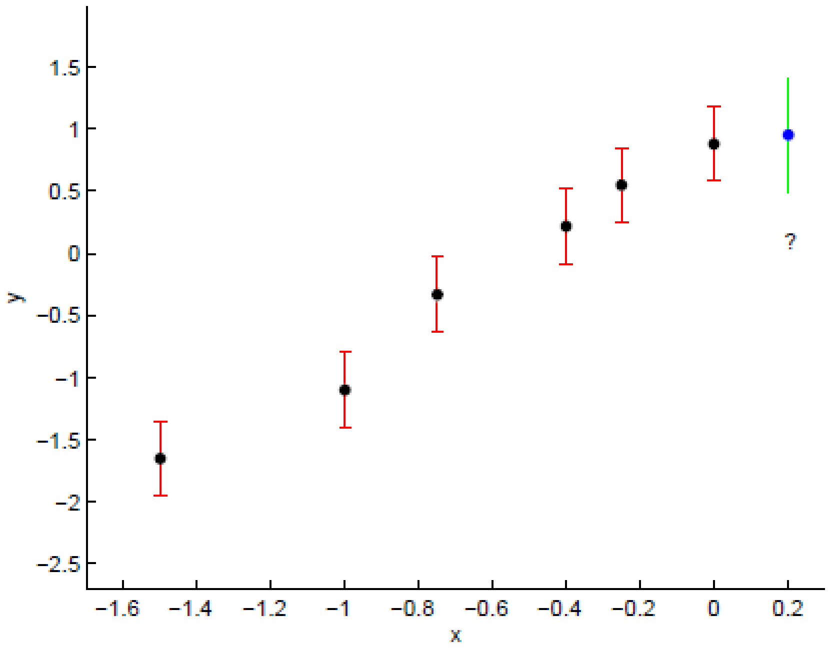


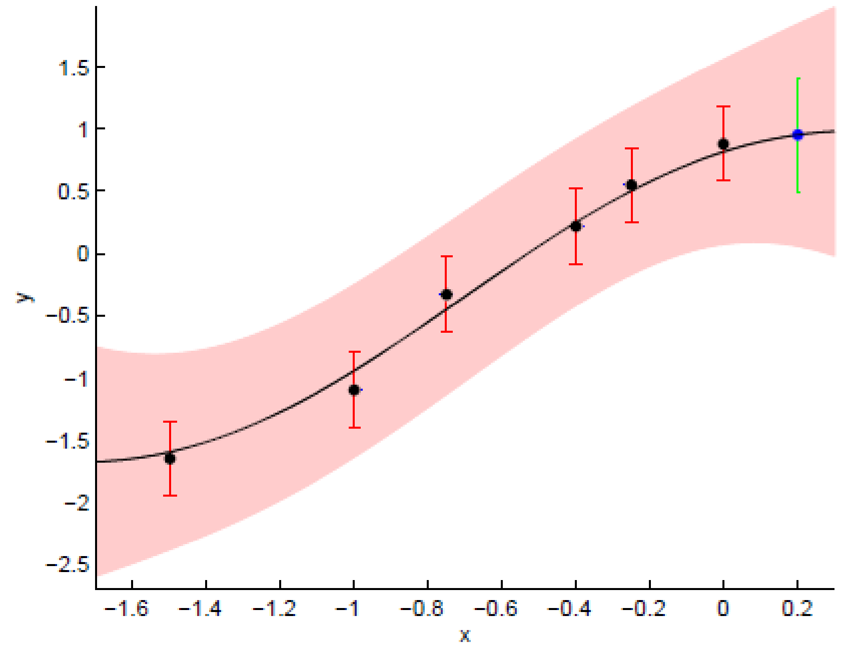
A “X” *process* is a distribution over a function space such that the pdf at any evaluation of the function are conditionally “X” distributed.

-Dirichlet Process [infinite state HMM]

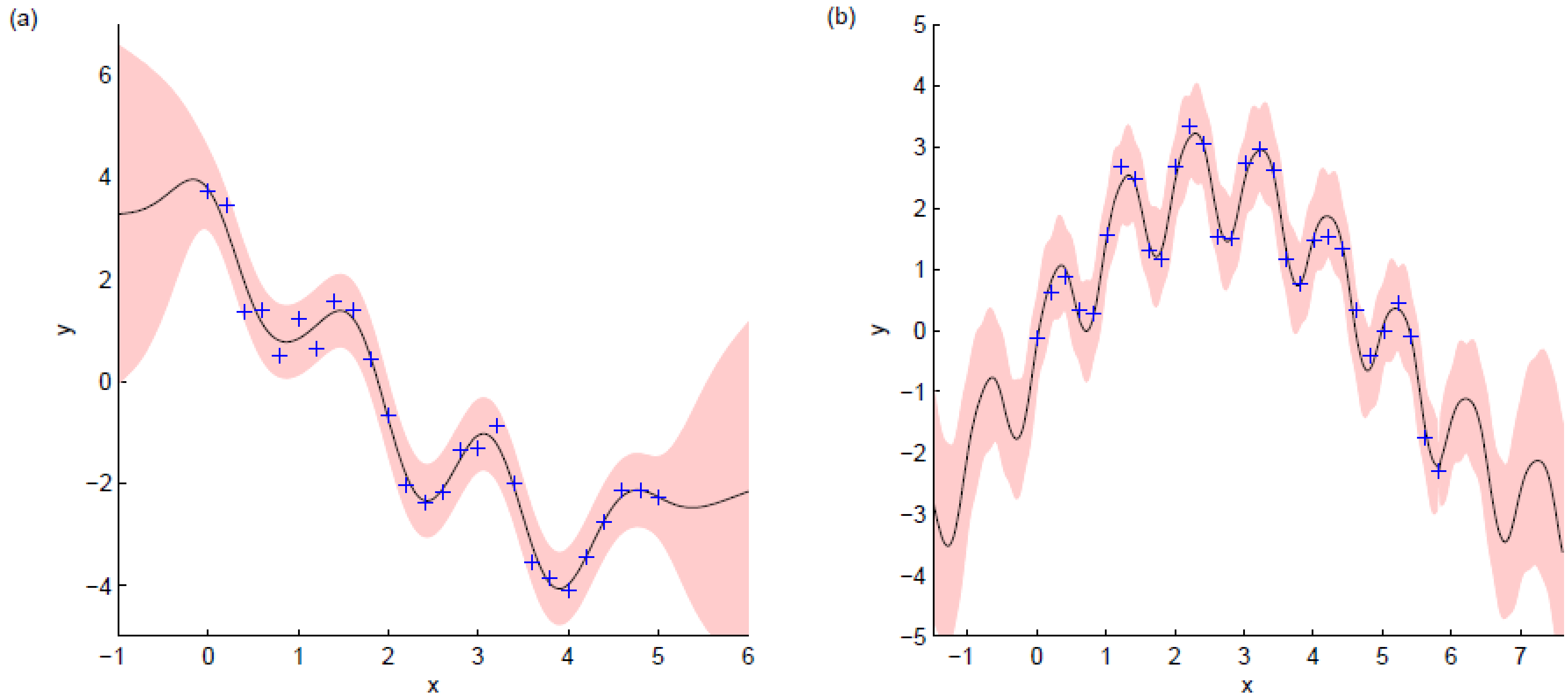
-Indian Buffet Process [infinite binary strings] etc etc.

Simple regression modelling





Less simple regression



The Gaussian process model

- See the GP via the distribution

$$\boxed{p(\mathbf{y}(\mathbf{x})) = \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}))}$$

- If we observe a set (x,y) and want to infer *y** at *x**

$$p\left(\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}(\mathbf{x}) \\ \mu(x_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{x}, \mathbf{x}) & \mathbf{K}(\mathbf{x}, x_*) \\ \mathbf{K}(x_*, \mathbf{x}) & k(x_*, x_*) \end{bmatrix}\right)$$

$$p(\mathbf{y}_*) = \mathcal{N}(\mathbf{m}_*, \mathbf{C}_*)$$

$$m_* = \mu(x_*) + \mathbf{K}(x_*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}(\mathbf{y} - \boldsymbol{\mu}(\mathbf{x})),$$
$$\sigma_*^2 = K(x_*, x_*) - \mathbf{K}(x_*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{K}(\mathbf{x}, x_*).$$

The beating heart...

What about these covariances though?

$$\mathbf{K}(\mathbf{x},\mathbf{x}) = \begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_n) \\ \vdots & \vdots & \vdots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & \cdots & k(x_n,x_n) \end{pmatrix}$$

Achieved using a *kernel function*, which describes the relationship between two points

What form should this take though?

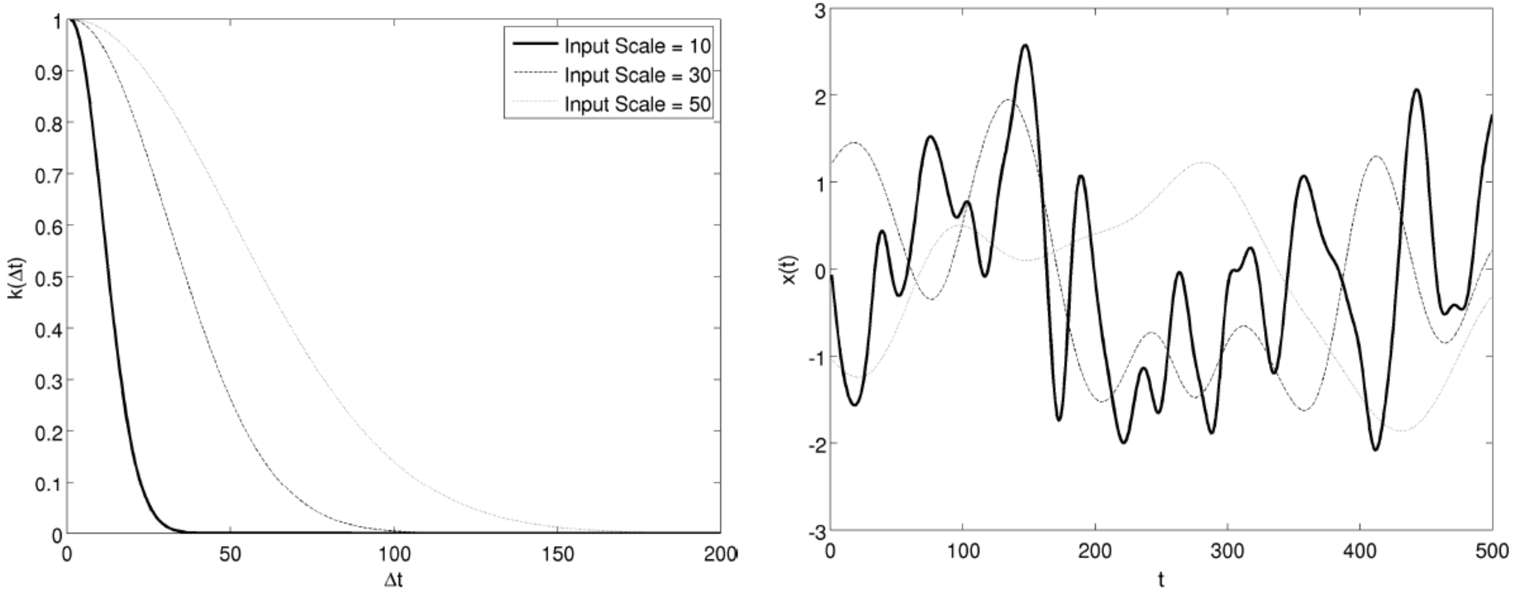
An example

$$k(x_i, x_j) = h^2 \exp\left[-\left(\frac{x_i - x_j}{\lambda}\right)^2\right]$$

What is this based upon?

- Intrinsic smoothness (infinitely differentiable)
- amplitude of expected functions is controlled by *h*
- typical scale of variations in time (correlation “length”) controlled by λ

Covariance functions



We commonly possess prior expectations that the function should be smooth. If we know something of the dynamics then this can inform our covariance functions accordingly

Covariances

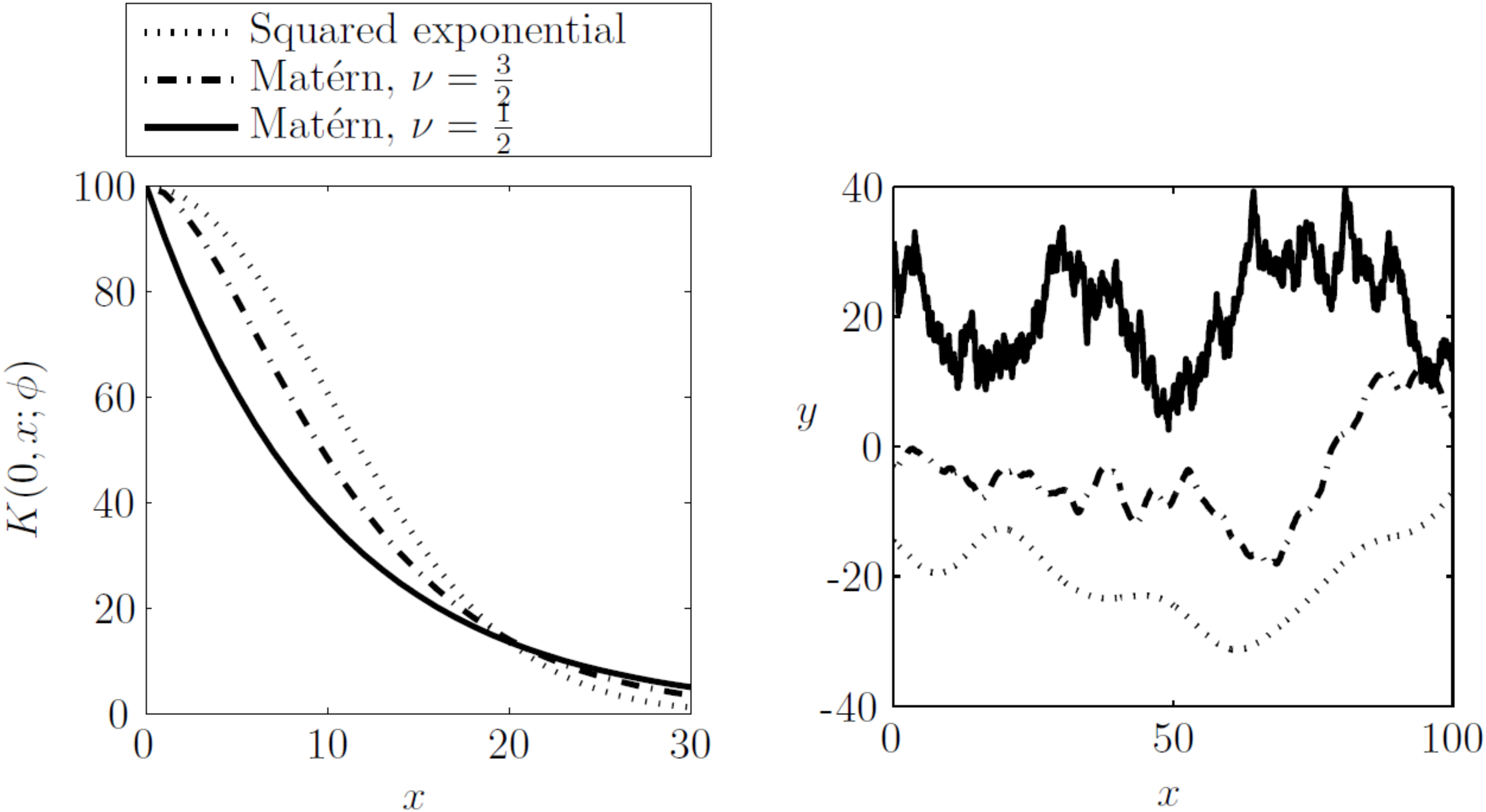
There are a huge number of covariance functions (in spite of the requirement that they be positive semi-definite) appropriate for modelling functions of different types

Elegantly, all continuous Markov time series models (AR, ARMA, ARIMA, GARCH, KF....) can be recast as special cases of Gaussian Processes

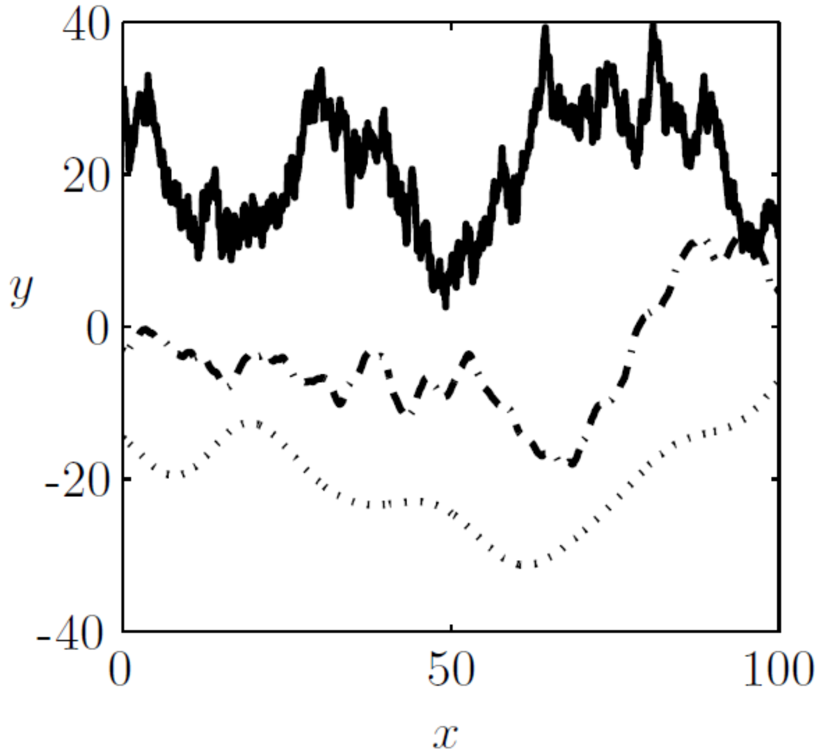
S. Reece and S. Roberts (2010). The Near Constant Acceleration Gaussian Process Kernel for Tracking. IEEE Signal Processing Letters.

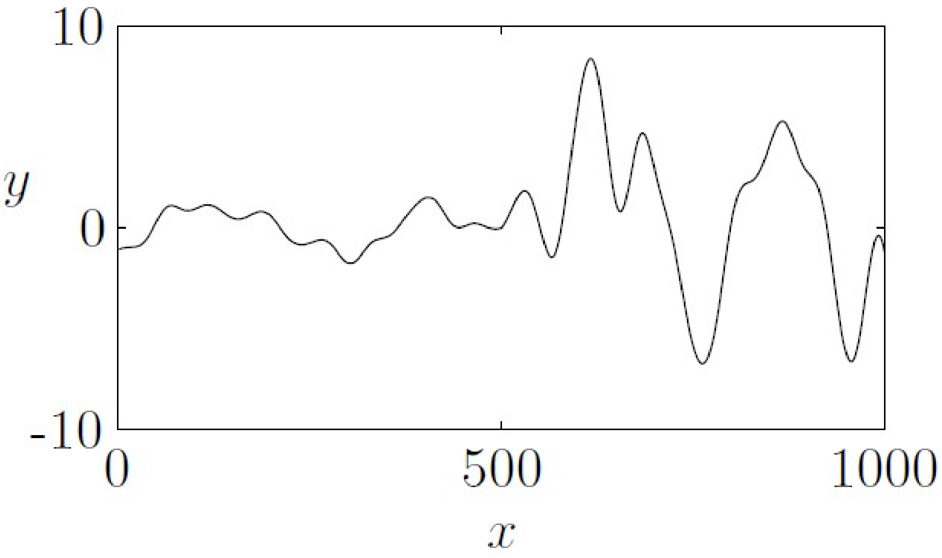
S. Reece and S. J. Roberts (2010). An Introduction to Gaussian Processes for the Kalman Filter Expert. Proceedings of Fusion 2010.

The squared exponential and Matérn covariances allow us to model functions of various degrees of smoothness..



We often want distances that are stationary (a function of x_1-x_2), implying that the function looks similar throughout its domain.



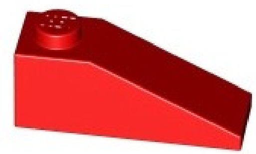
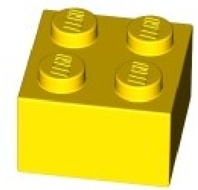


stationary functions

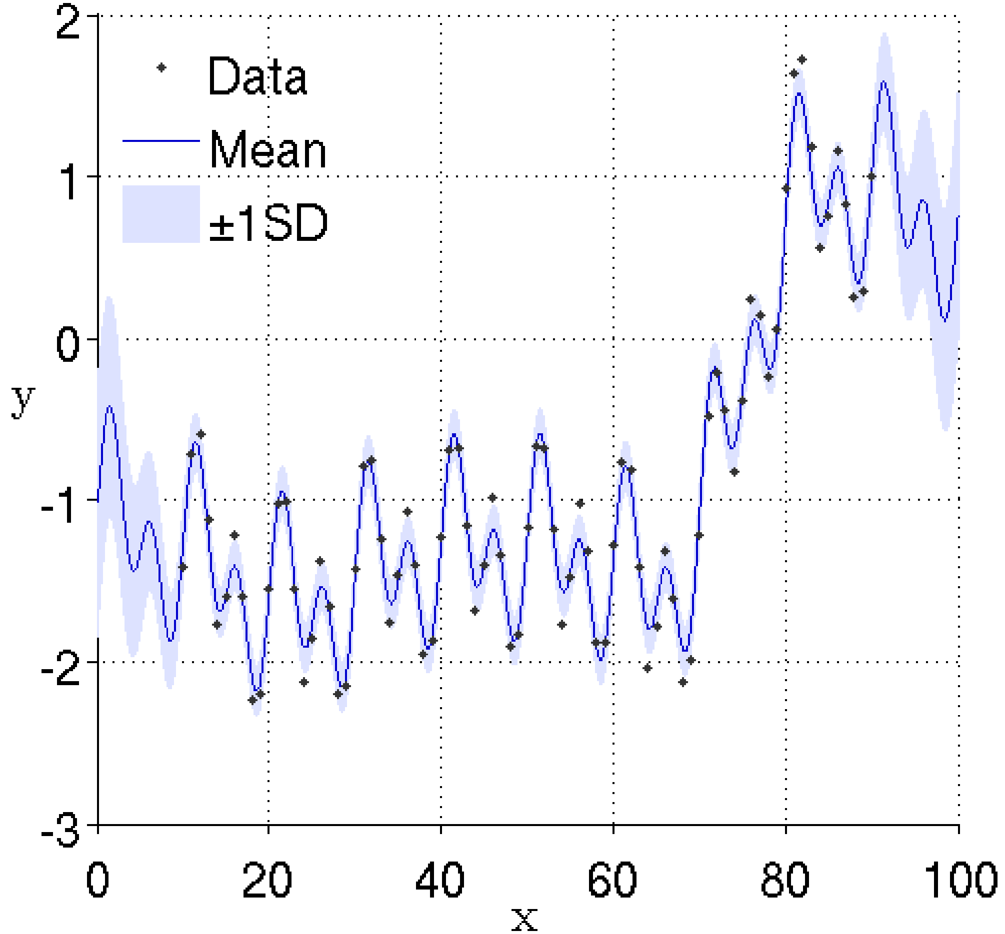
non-stationary function

We can create new covariance functions by adding or multiplying other covariance functions.

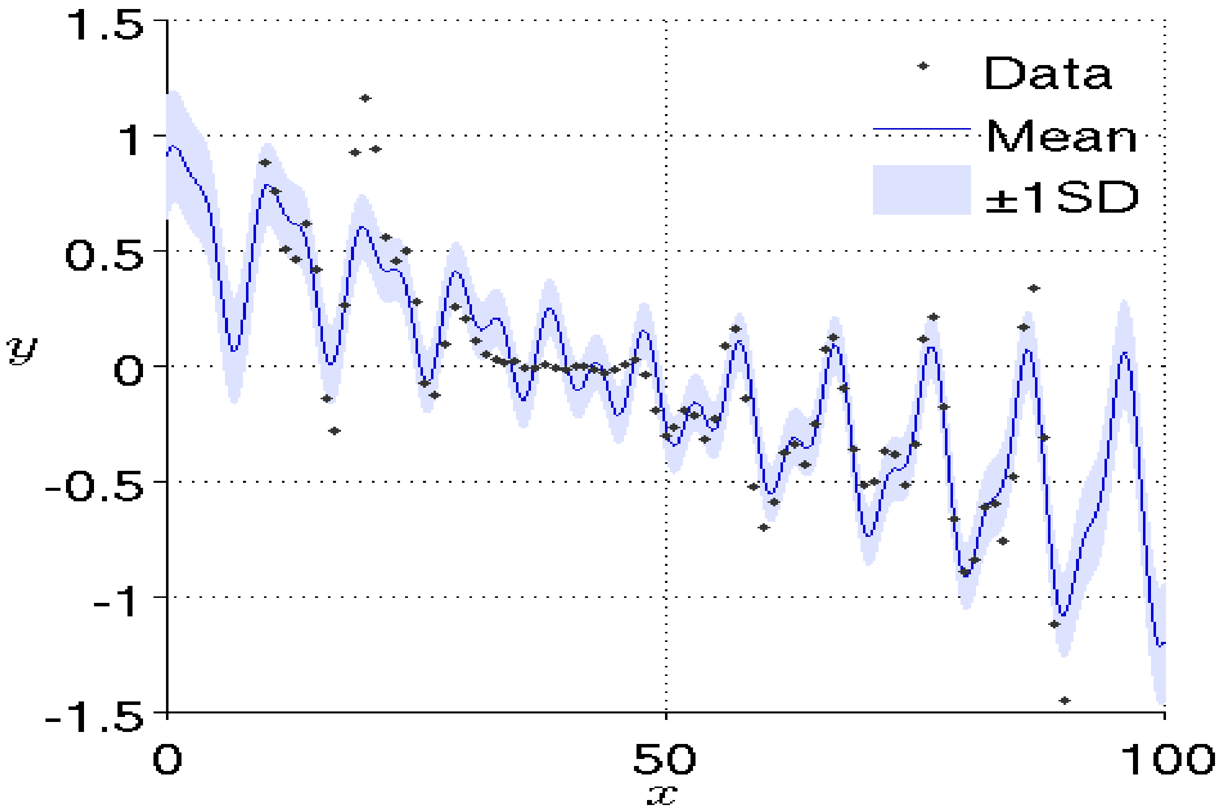
e.g.

 ×  × ( + )

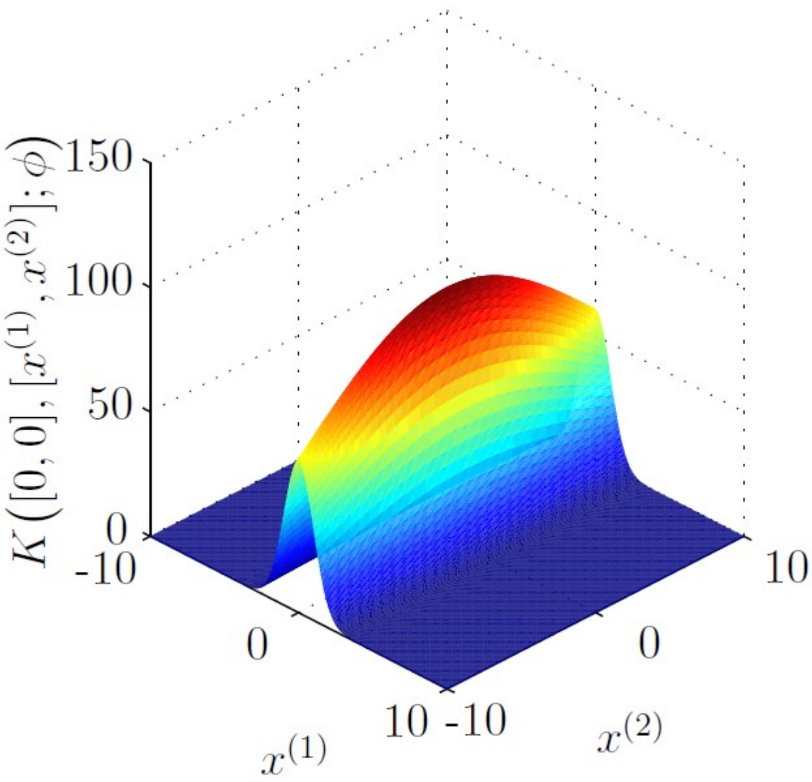
When a function is the sum of two independent functions, use a covariance that is the sum of the covariances for those two functions.

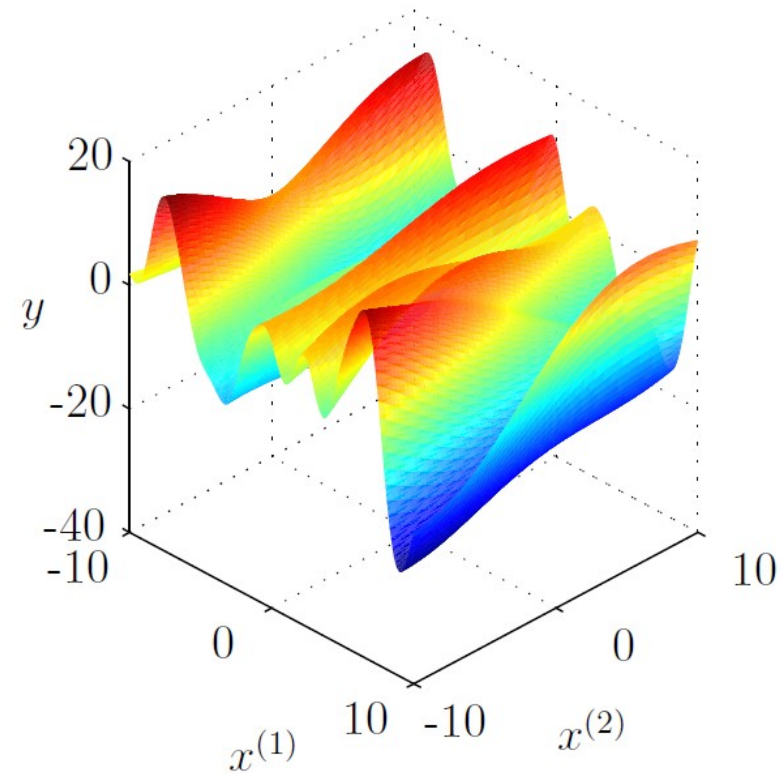


When a function is the product of two independent functions, use a covariance that is (almost) the product of the covariances for those two functions.

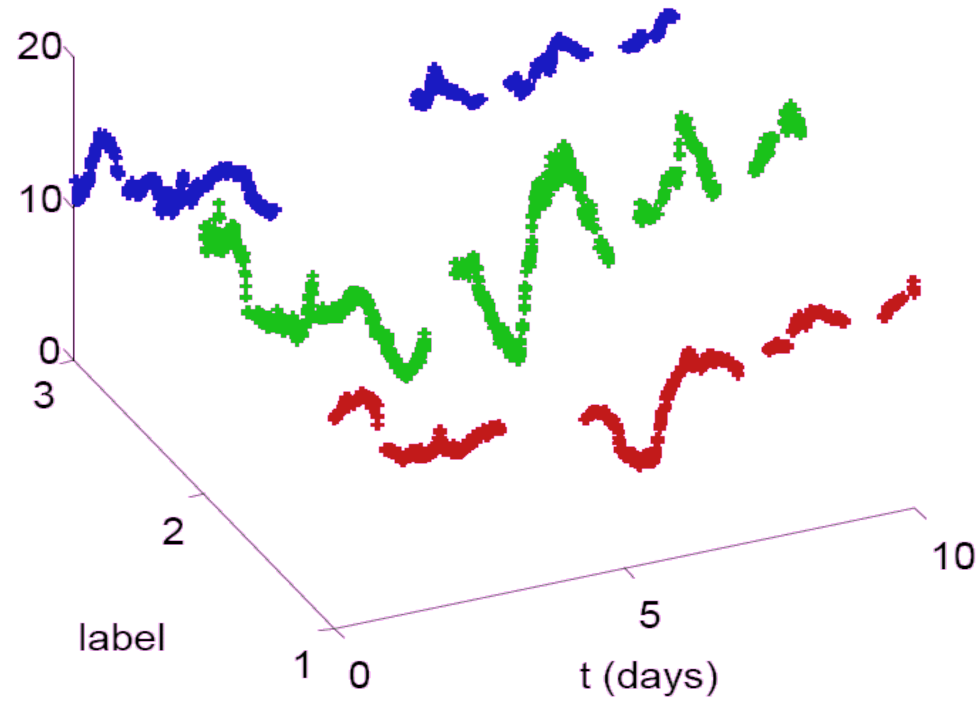


We can modify covariance functions to accommodate multiple input dimensions, using



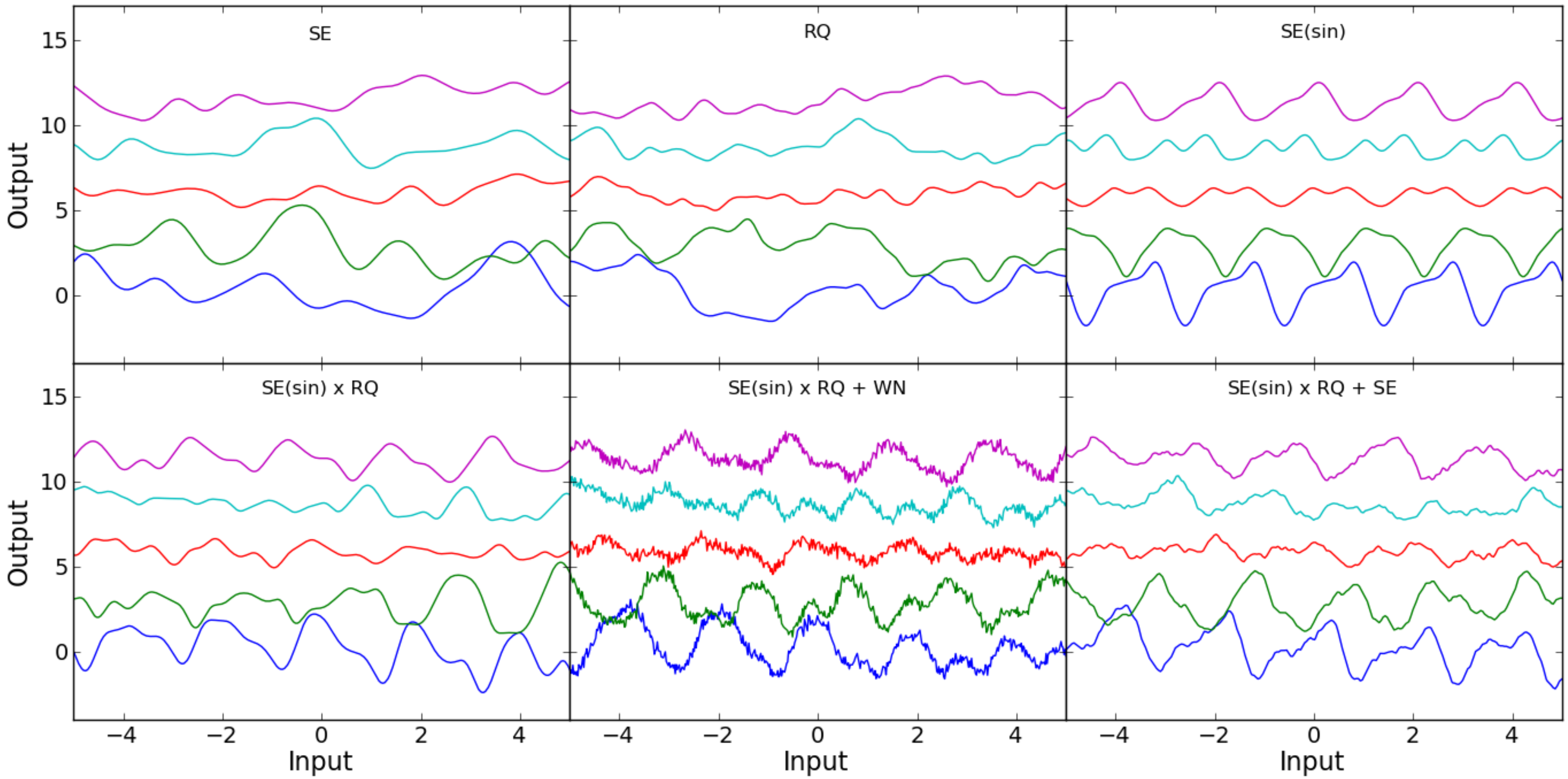


If there are multiple outputs, reframe the problem as having a single output, and an additional *label* input specifying the output.

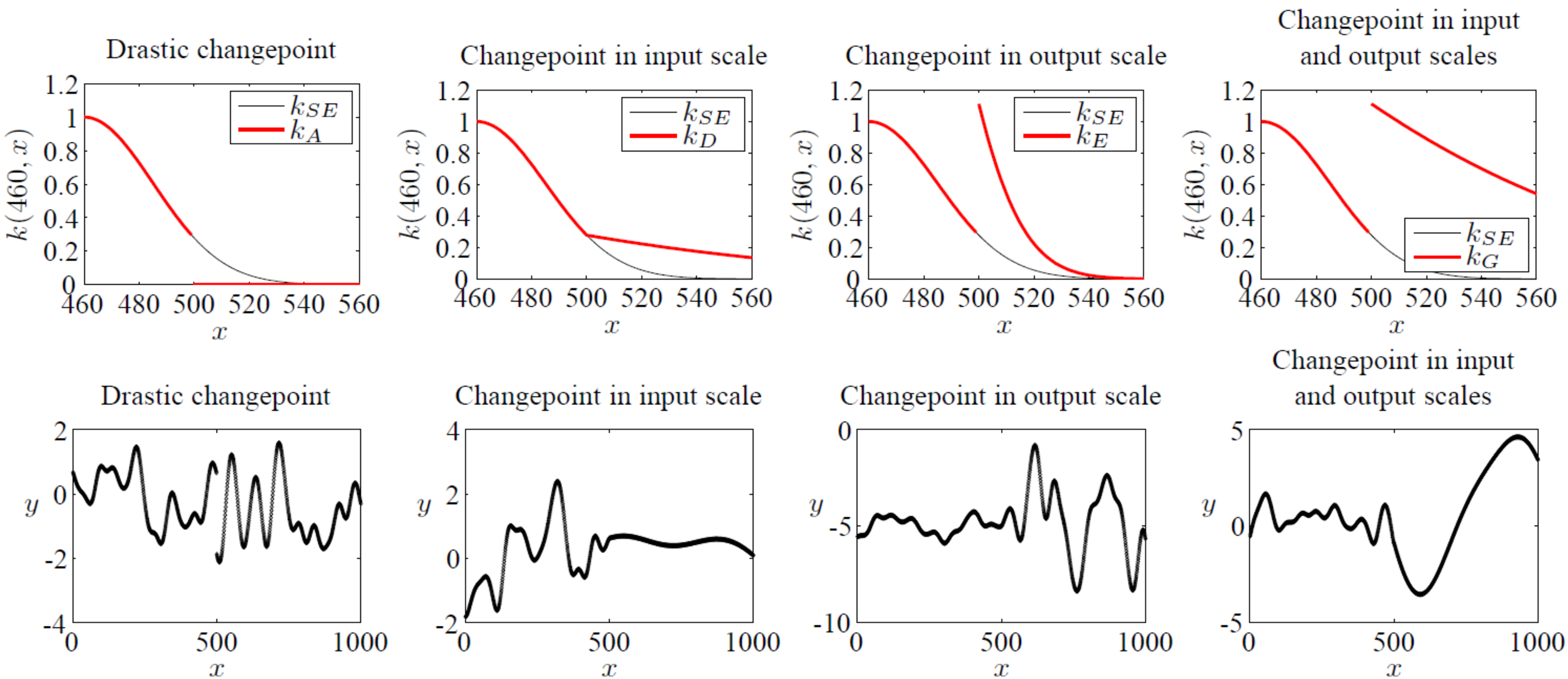


Hence we do not need simultaneous observations of all outputs.

Periodic & quasi-periodic



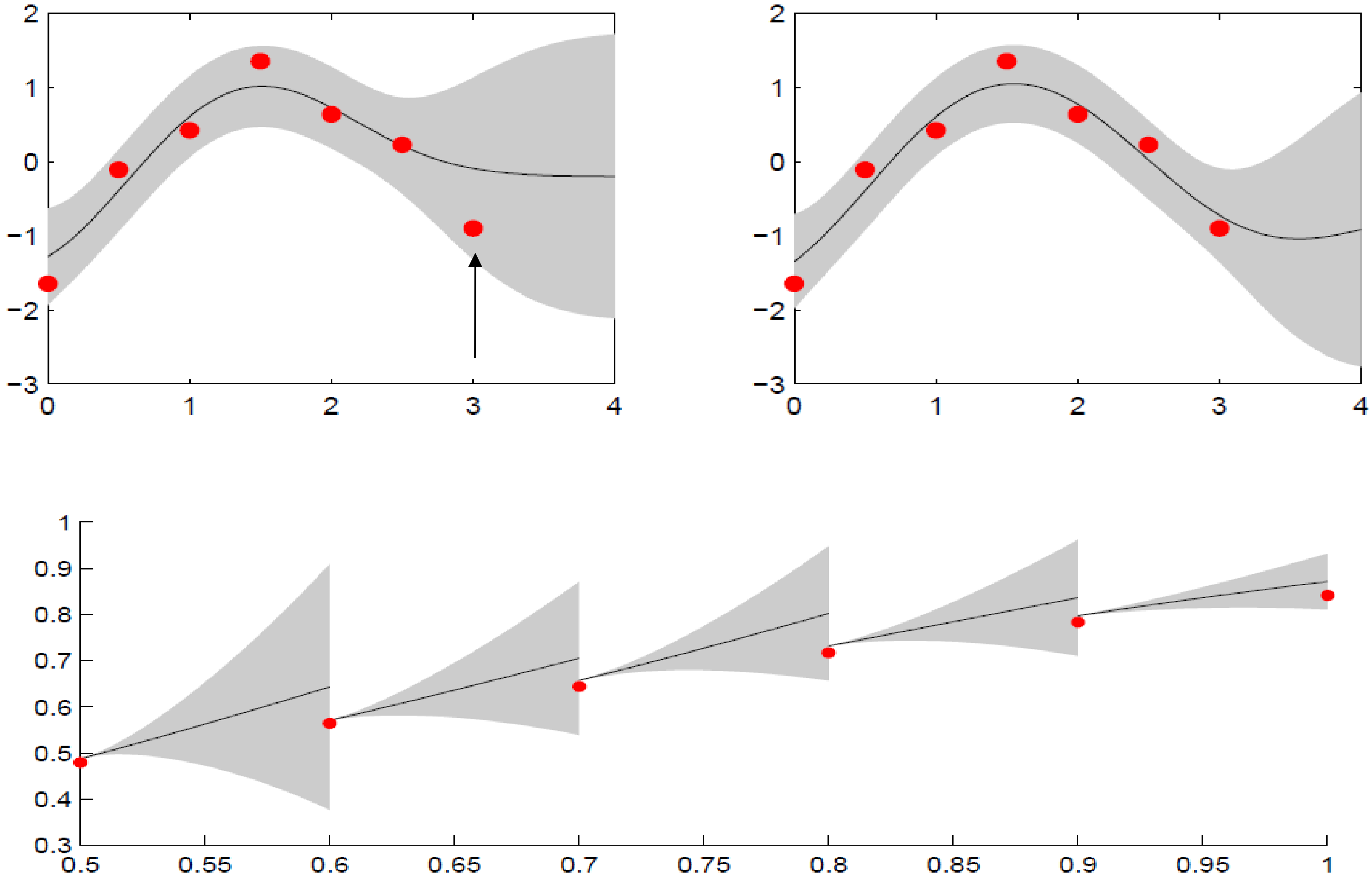
Many other modifications are possible, to build covariances allowing for e.g. changepoints, faults and sets.



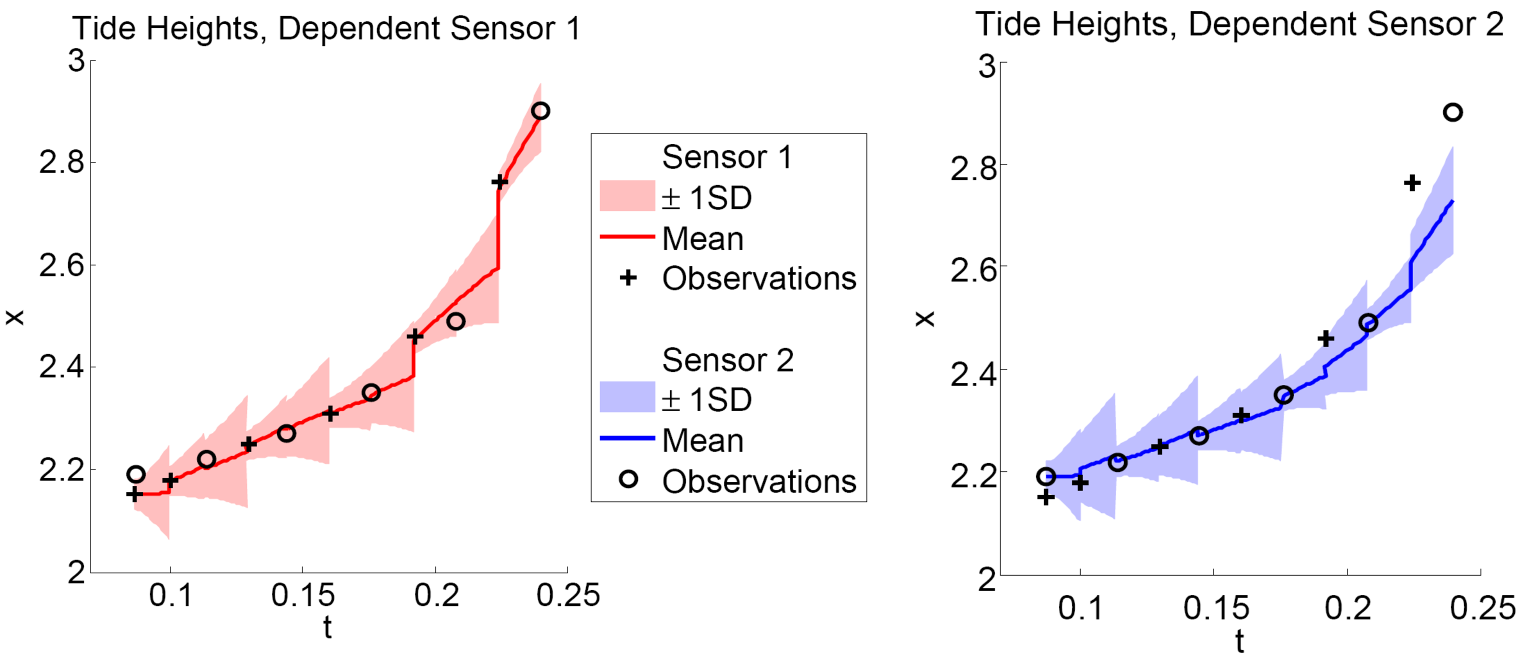
R. Garnett, M. A. Osborne, S. Reece, A. Rogers and S. J. Roberts (2010). Sequential Bayesian Prediction in the Presence of Changepoints and Faults. The Computer Journal, March 2010.

Part IV : some examples

In a sequential setting



Active data selection



The Gaussian process can decide for itself which sensor to observe, and when, by determining which observation will be most informative.

M.A. Osborne, S.J. Roberts, A. Rogers, and N.R. Jennings (2011). Real-Time Information Processing of Environmental Sensor Network Data using Bayesian Gaussian Processes Transactions on Sensor Networks

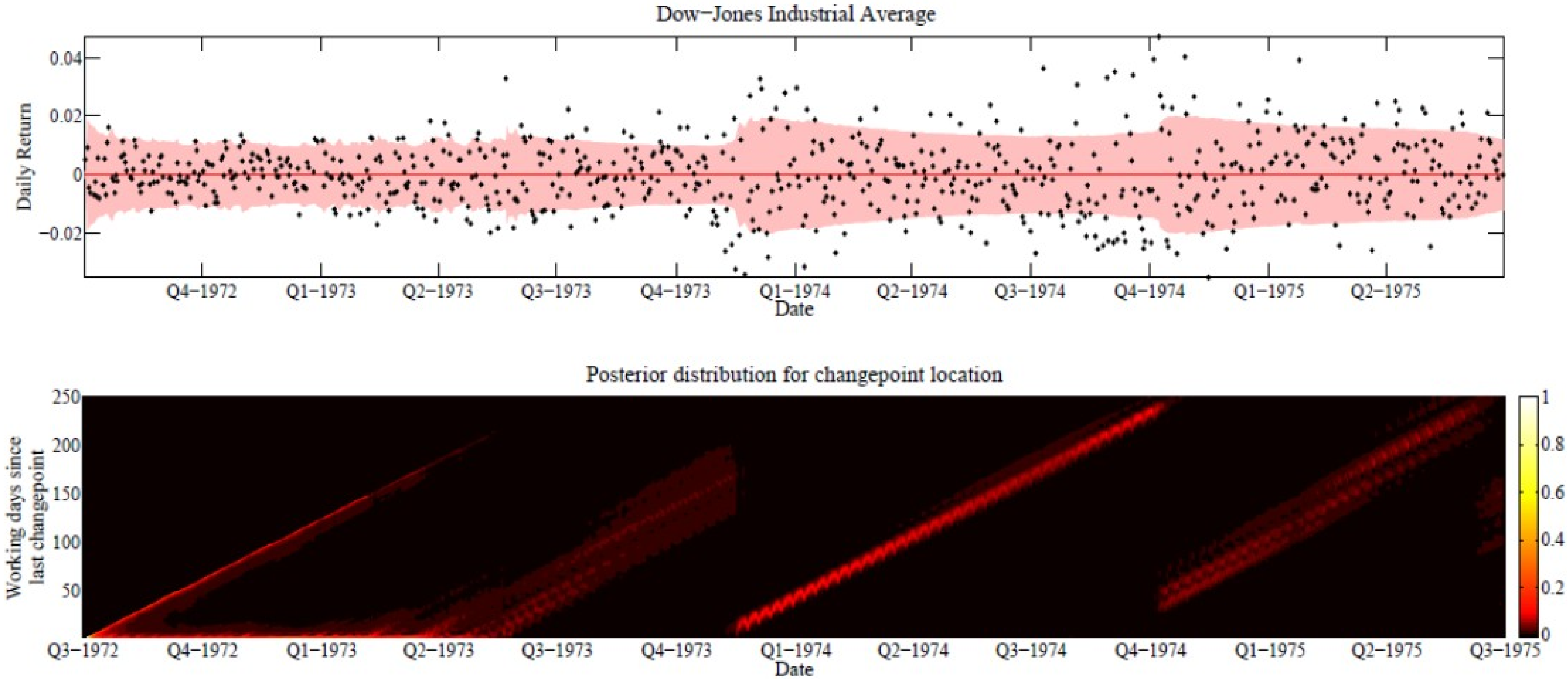
Demonstration

http://www.aladdinproject.org/situation/

Changepoints

R. Garnett, M. A. Osborne, S. Reece, A. Rogers and S. J. Roberts (2010). Sequential Bayesian Prediction in the Presence of Changepoints and Faults, The Computer Journal.

Dow-Jones data



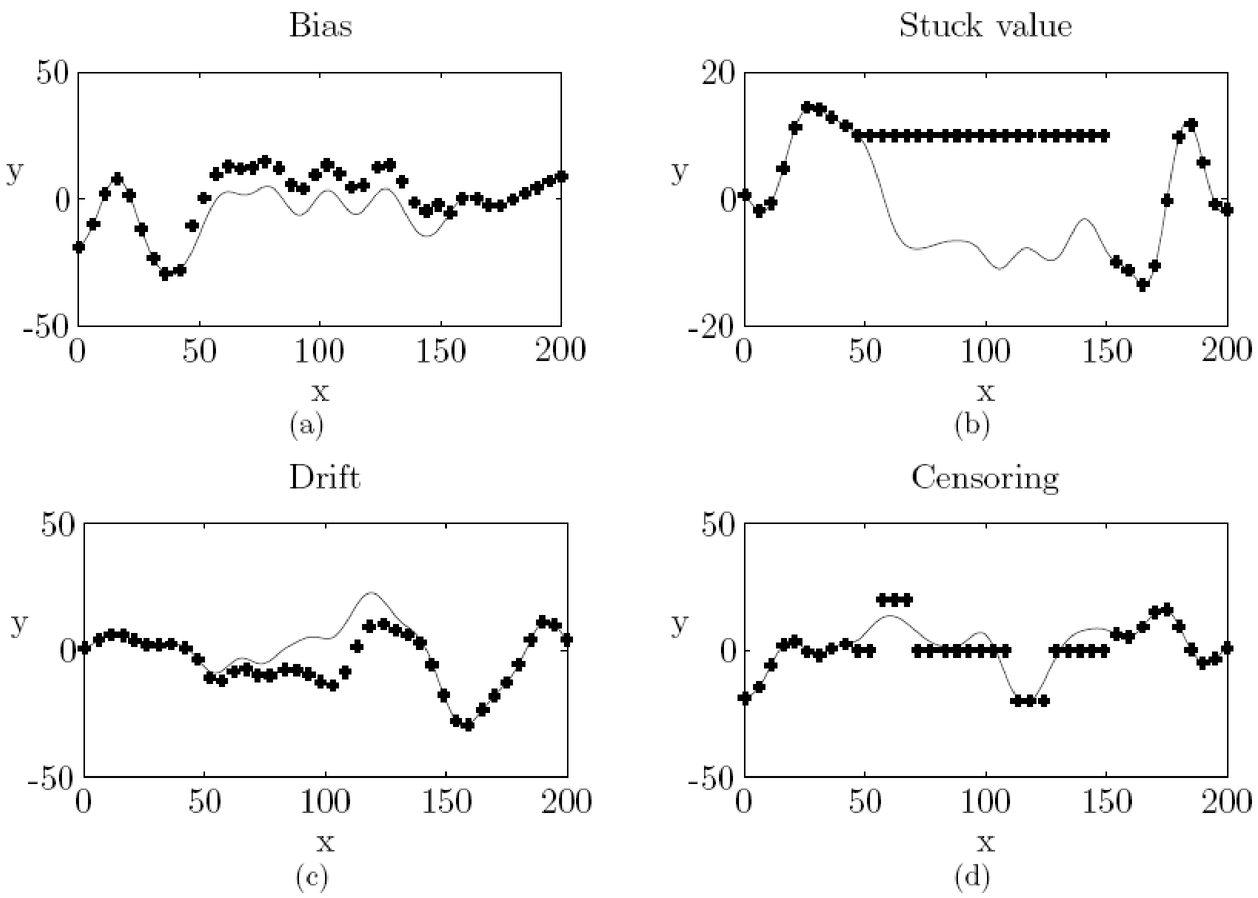
Watergate scandal

OPEC embargo

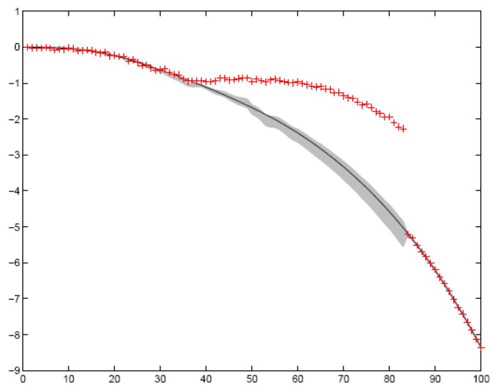
Nixon resigns

Faults & fault recovery

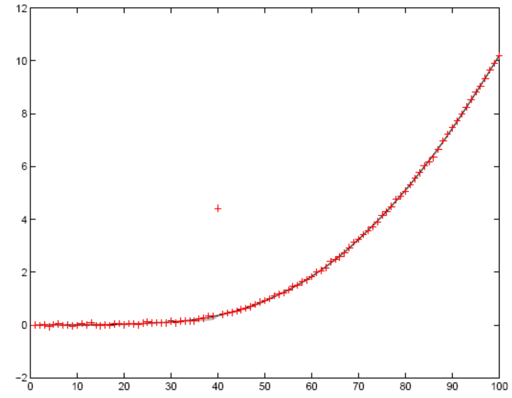
Potential fault types



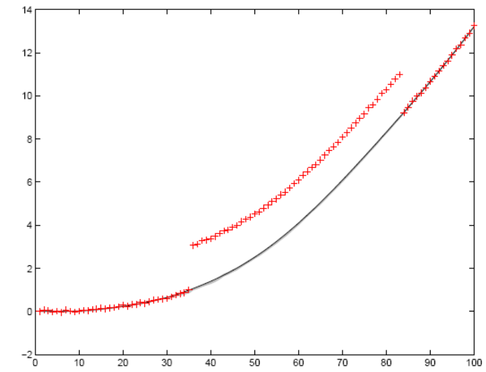
Posterior distribution over the fault type



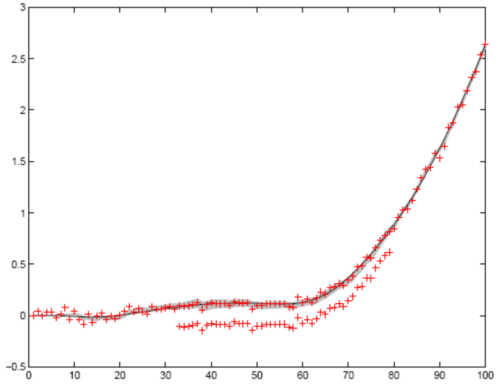
Drift



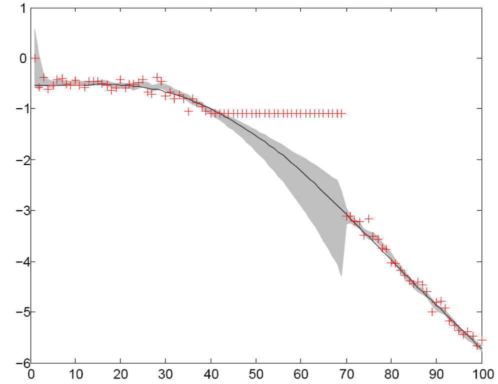
Spike



Bias

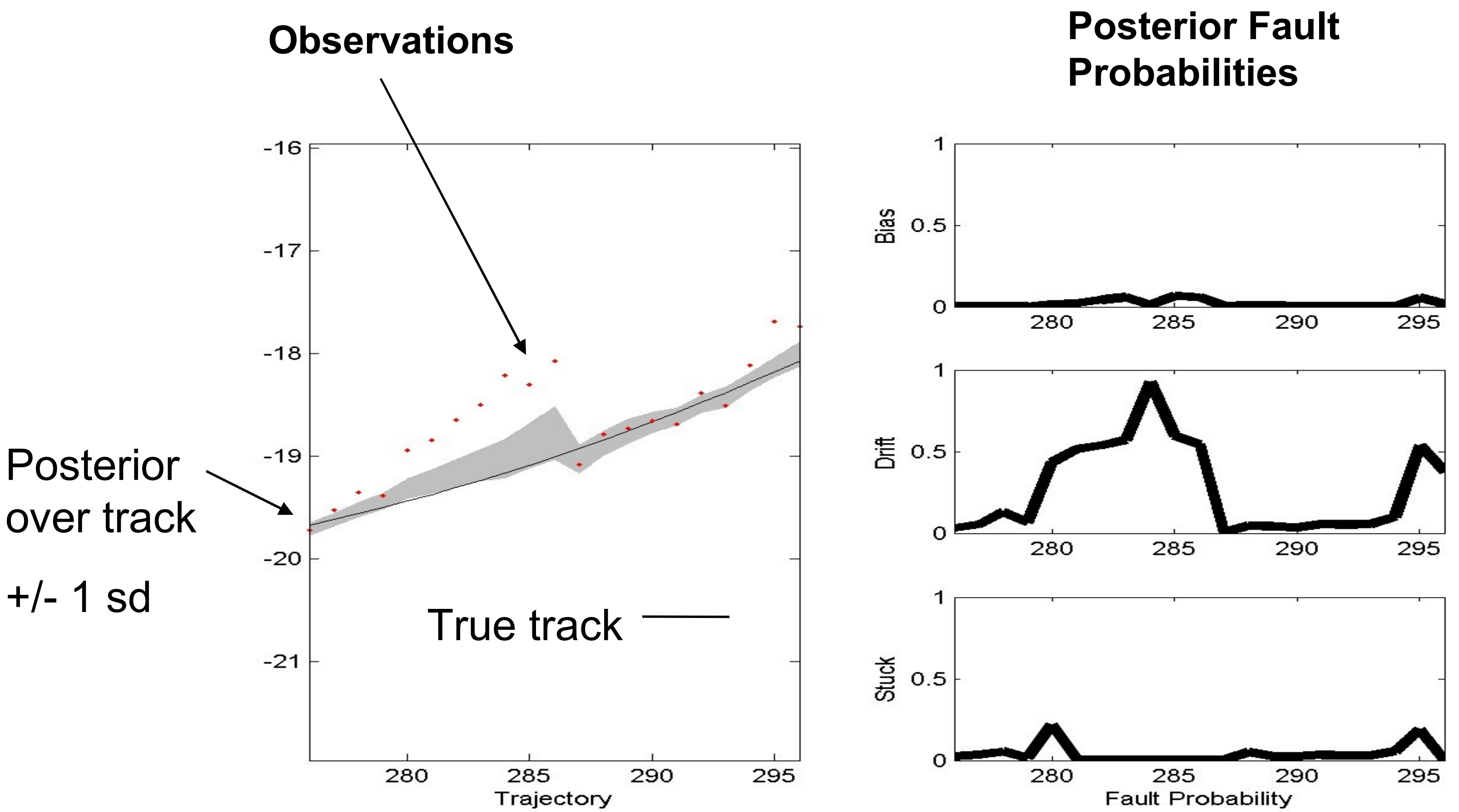


Echo



Stuck

Can track and fault recover sequentially



Faults - demonstration

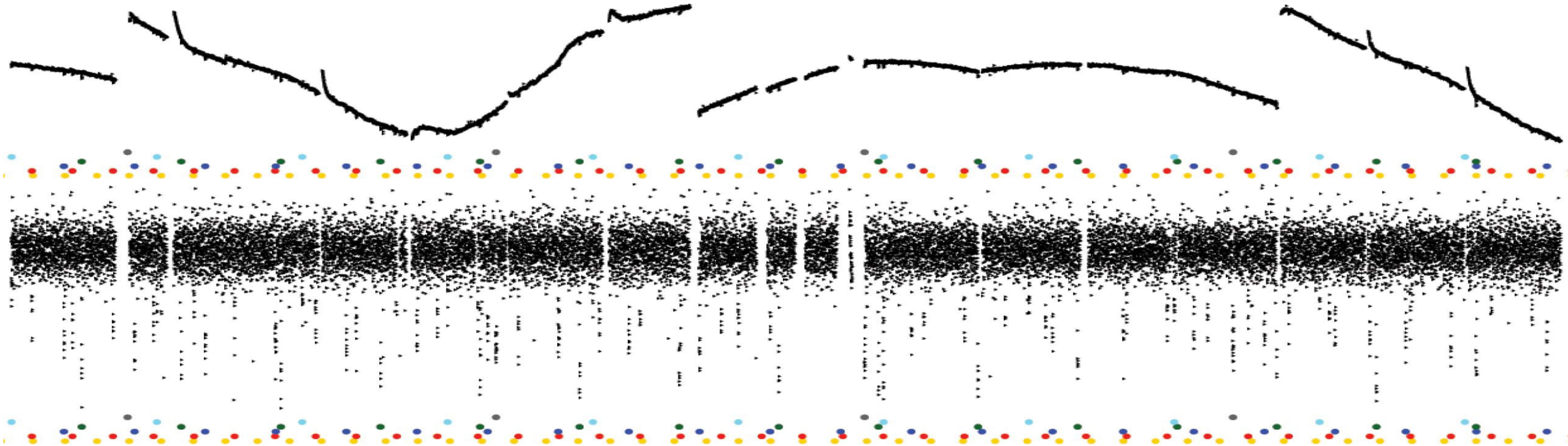
Recent work – applications to time-domain Astronomy

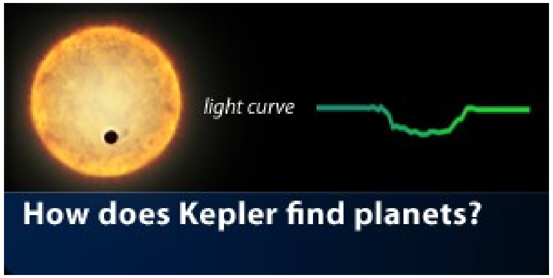
W. Armour, A. Karastergiou, M. Giles, C. Williams, A. Magro, K. Zagkouris, S. Roberts, S. Salvini, F. Dulwich and B. Mort (2011). A GPU-based survey for millisecond radio transients using ARTEMIS. Proceedings of ADASS XXI.

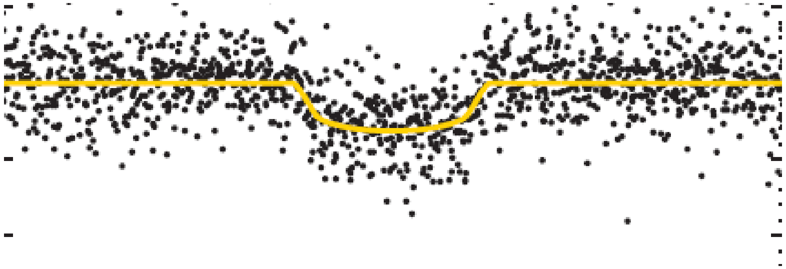
A. McQuillan, S. Aigrain, S. Roberts (2011). Statistics of Stellar Variability from Kepler - I: Revisiting Quarter 1 with an Astrophysically Robust Systematics Correction. Astronomy and Astrophysics.

N. P. Gibson, S. Aigrain, S. Roberts, T. M. Evans, M. Osborne and F. Pont (2011). A Gaussian process framework for modelling instrumental systematics: application to transmission spectroscopy. Monthly Notices of the Royal Astronomical Society.

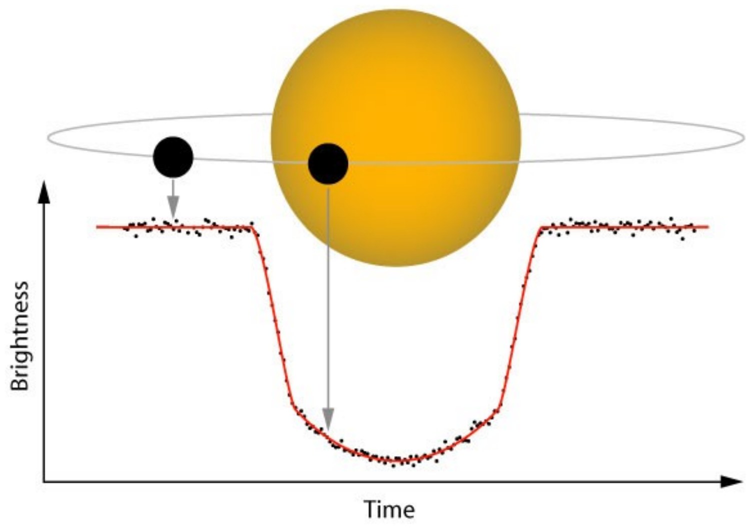
Light curves

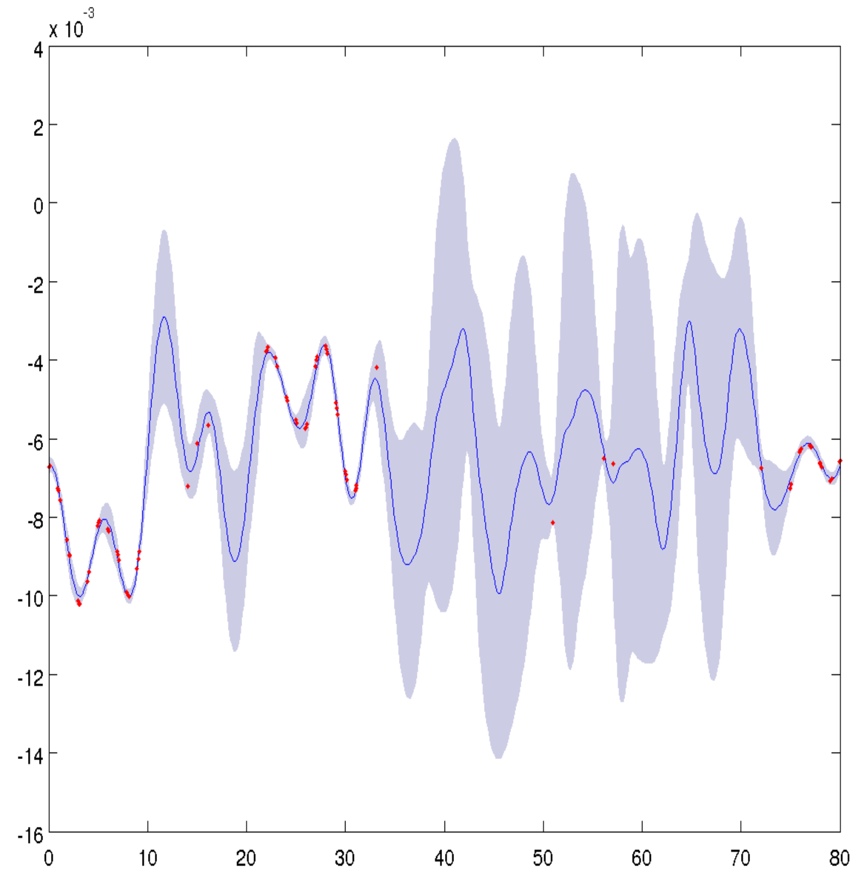






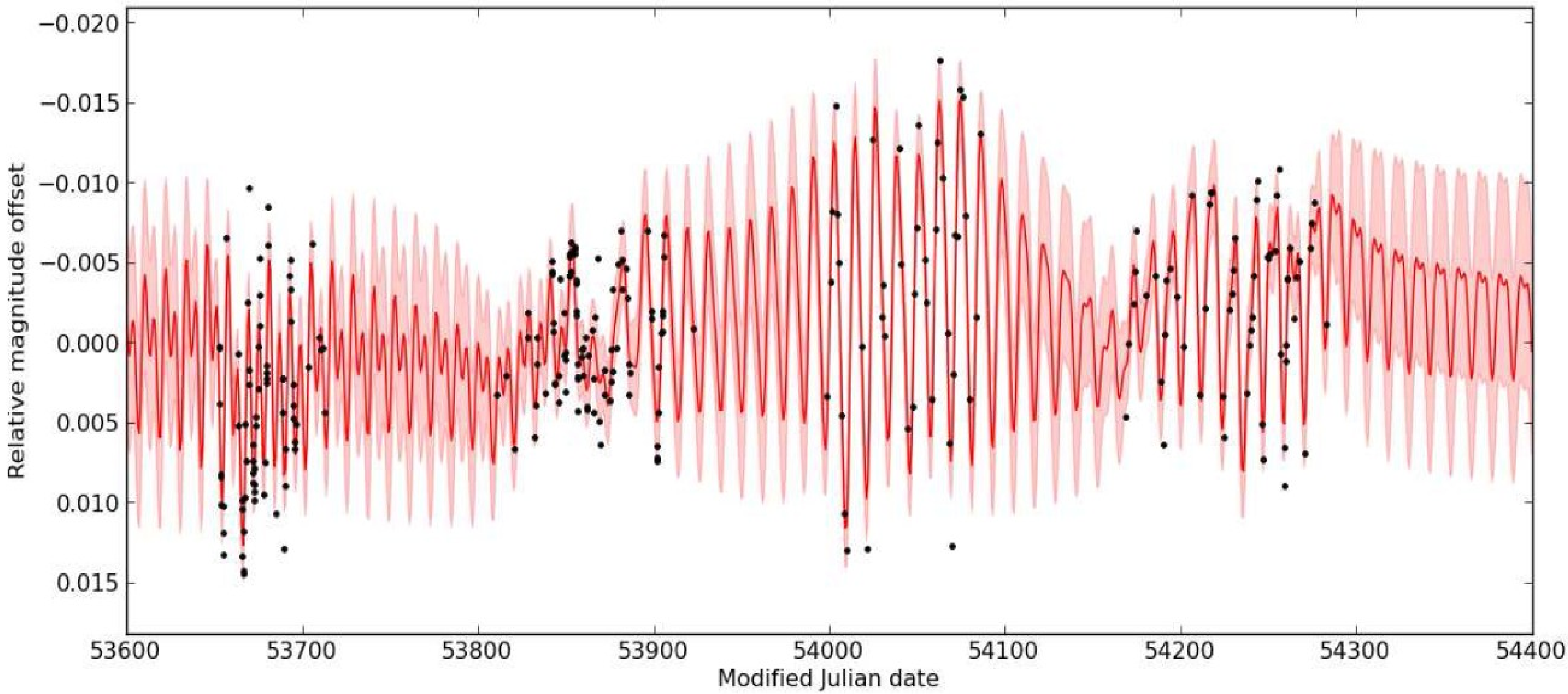
Quasi-periodic model for stellar flux



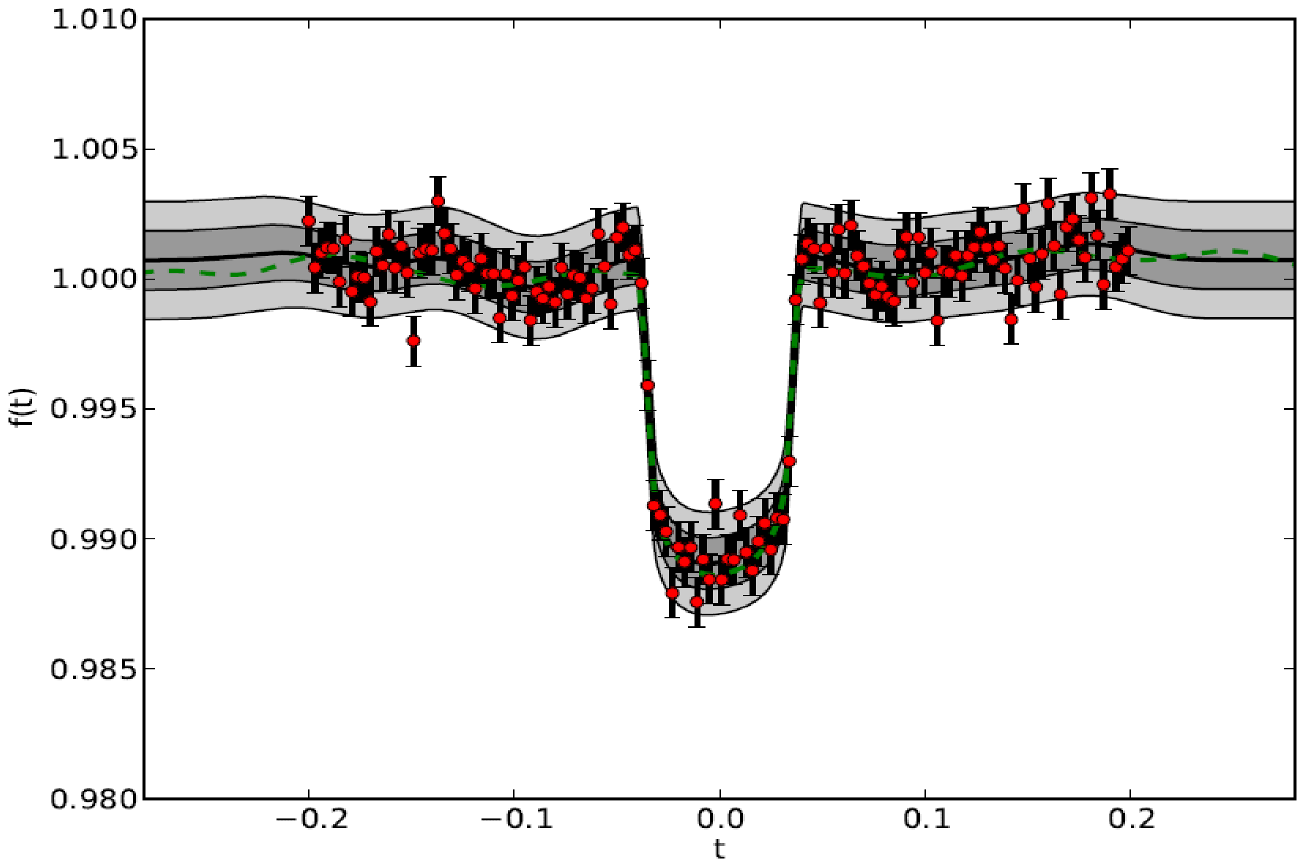


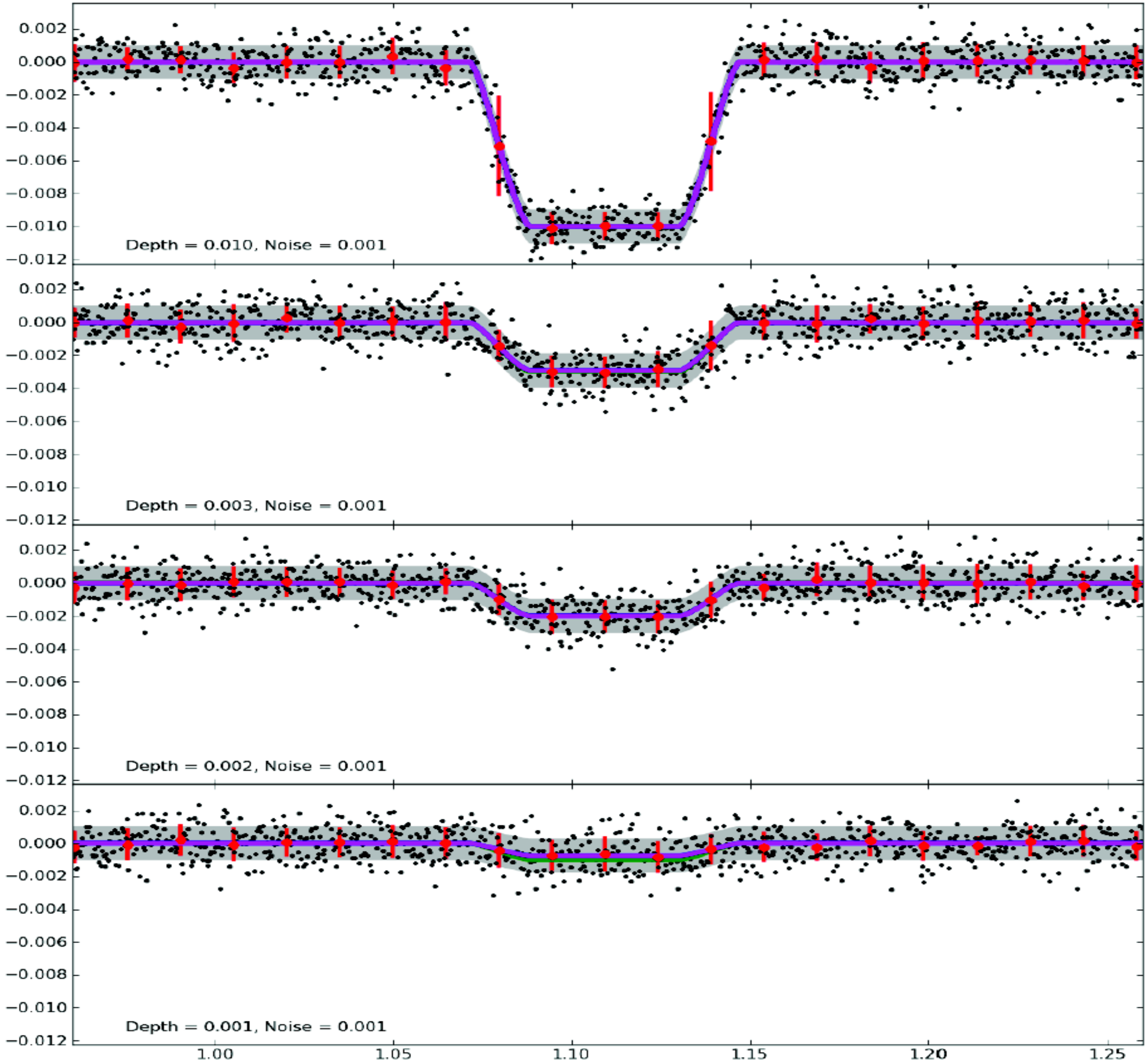
Problem is that stellar flux is highly variant... star-spots and stellar rotations... so first we need to model the quasi-periodic flux measurements... but there aren't many of them!

Quasi-periodic Gaussian Process regression to photometric observations of the well-known planet-host star HD 189733



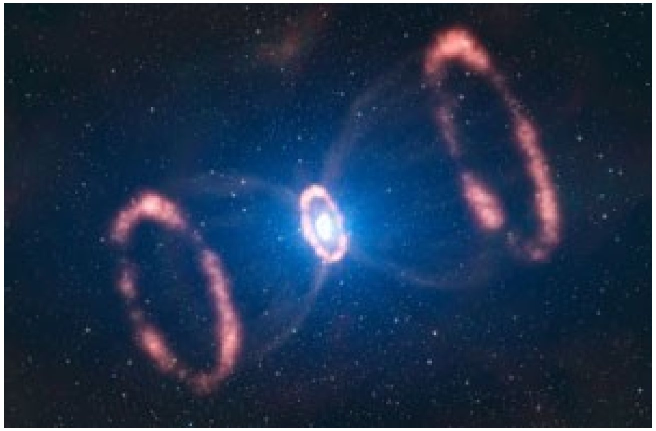
Exoplanet transit light curve. The data is fitted with a GP with an exoplanet transit mean function and a squared exponential covariance kernel to model the correlated noise process and the effects of external state variables





Royal Society 2012

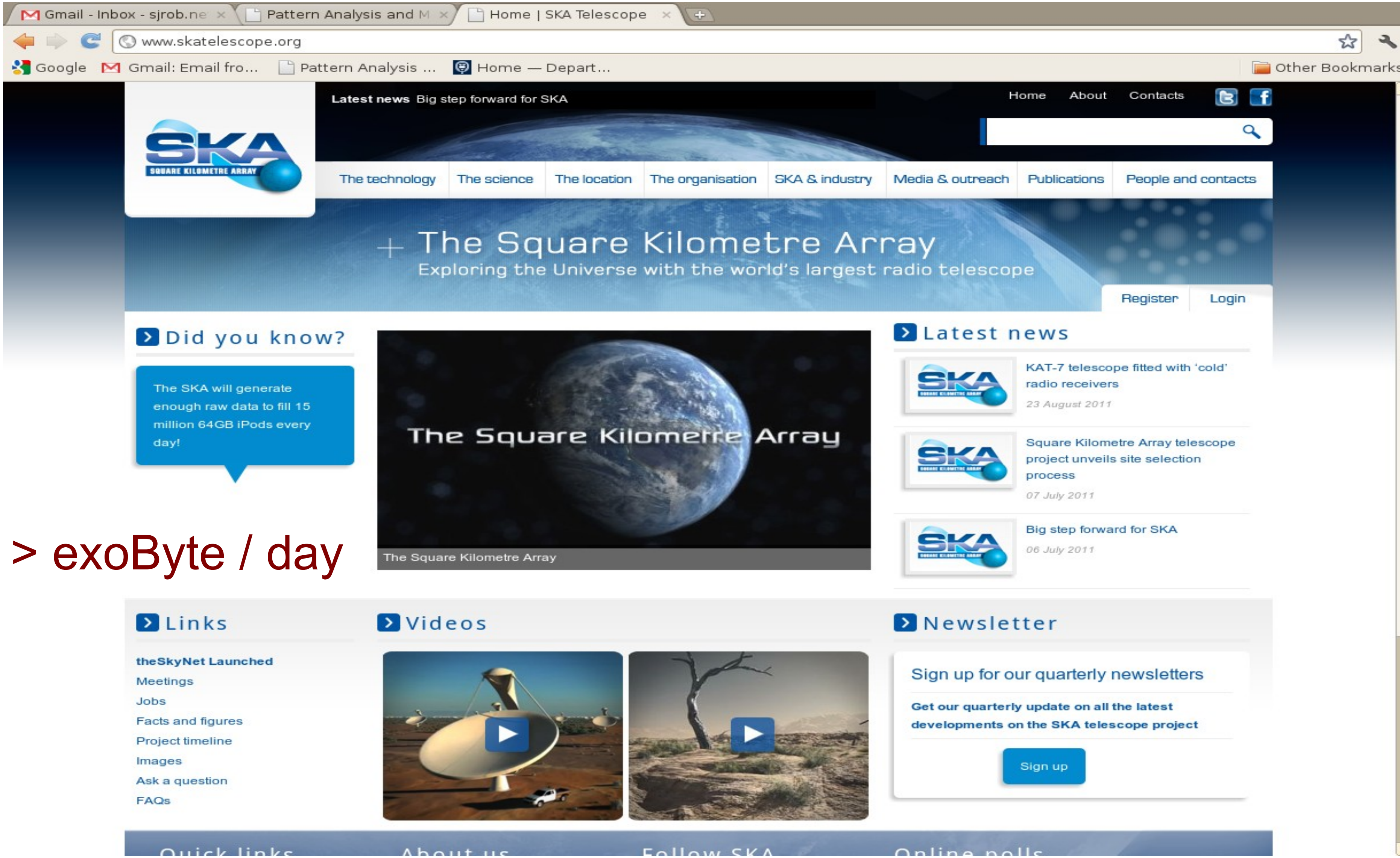
Transient phenomena: Radio Surveys



- Transient objects
  - Pulsars
  - Supernovae
  - ?
- ”Needle in haystack” problems
  - Computationally demanding – 100s of TBs of data each night

One of our major challenges is *scalability* in Bayesian modelling...

W. Armour, A. Karastergiou, M. Giles, C. Williams, A. Magro, K. Zagkouris, S. Roberts, S. Salvini, F. Dulwich and B. Mort (2011). A GPU-based survey for millisecond radio transients using ARTEMIS. Proceedings of ADASS XXI.


Gmail - Inbox - sjrob.ne
Pattern Analysis and M
Home | SKA Telescope
www.skatelescope.org
Google
Gmail: Email fro...
Pattern Analysis ...
Home — Depart...
Other Bookmarks
Latest news Big step forward for SKA
Home
About
Contacts
SKA
SQUARE KILOMETRE ARRAY
The technology
The science
The location
The organisation
SKA & industry
Media & outreach
Publications
People and contacts
The Square Kilometre Array
Exploring the Universe with the world's largest radio telescope
Register
Login
Did you know?
The SKA will generate enough raw data to fill 15 million 64GB iPods every day!
The Square Kilometre Array
The Square Kilometre Array
Latest news
KAT-7 telescope fitted with 'cold' radio receivers
23 August 2011
Square Kilometre Array telescope project unveils site selection process
07 July 2011
Big step forward for SKA
06 July 2011
Links
theSkyNet Launched
Meetings
Jobs
Facts and figures
Project timeline
Images
Ask a question
FAQs
Videos
Newsletter
Sign up for our quarterly newsletters
Get our quarterly update on all the latest developments on the SKA telescope project
Sign up

> exoByte / day

Questions?

The prior mean function is the function our inference will default to far from observations.