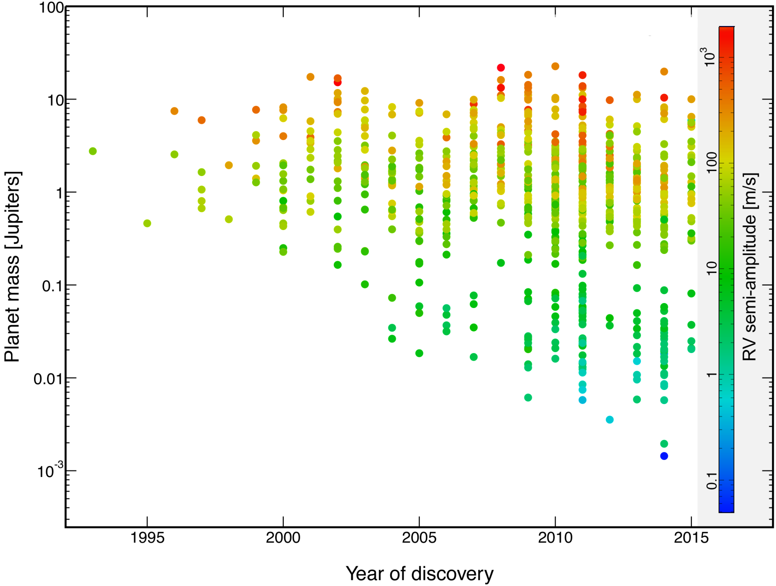
Gaussian processes for modelling stellar activity and detecting planets

Vinesh Rajpaul[1], **Suzanne Aigrain**[1], **Stephen Roberts**[2]

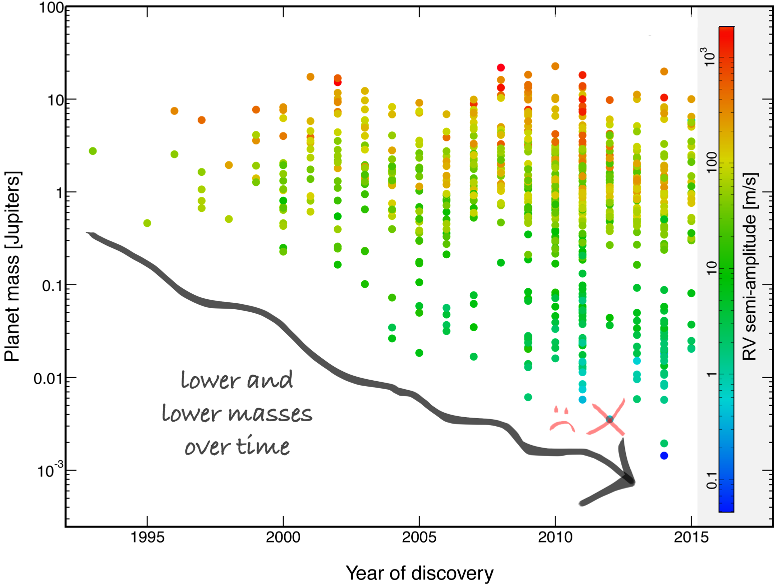
[1] Oxford Astrophysics
[2] Oxford Pattern Recognition and Machine Learning Group

31 March 2016

Doppler spectroscopy and exoplanets



Doppler spectroscopy and exoplanets



Stellar nuisance signals

- $\Delta\mathrm{RV}_{\mathrm{nuisance}} \gtrsim$ or $\gg \Delta\mathrm{RV}_{\mathrm{planet}}$

Stellar nuisance signals

- $\Delta\mathrm{RV}_{\mathrm{nuisance}} \gtrsim$ or $\gg \Delta\mathrm{RV}_{\mathrm{planet}}$
- Short-term ($\sim$ minutes, hours) signals due to oscillation, granulation

Stellar nuisance signals

- $\Delta\mathrm{RV}_{\mathrm{nuisance}} \gtrsim$ or $\gg \Delta\mathrm{RV}_{\mathrm{planet}}$
- Short-term ($\sim$ minutes, hours) signals due to oscillation, granulation
- Longer time-scale signals (days to years) associated with **rotationally-modulated activity**

Stellar nuisance signals

Rotationally-modulated activity signals...

Stellar nuisance signals

Rotationally-modulated activity signals...

- Are **stochastic** (active regions appear randomly)

Stellar nuisance signals

Rotationally-modulated activity signals...

- Are **stochastic** (active regions appear randomly)
- Characterised by **time-scales similar to those associated with planets** (days to years)

Stellar nuisance signals

Rotationally-modulated activity signals...

- Are **stochastic** (active regions appear randomly)
- Characterised by **time-scales similar to those associated with planets** (days to years)
- **Quasi-periodic** (periodic stellar rotation + evolving active regions + long-term activity cycles)

Stellar nuisance signals

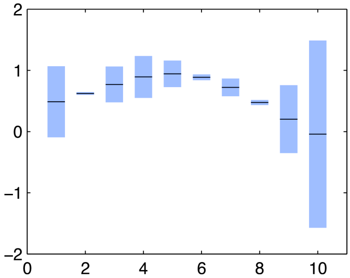
Rotationally-modulated activity signals...

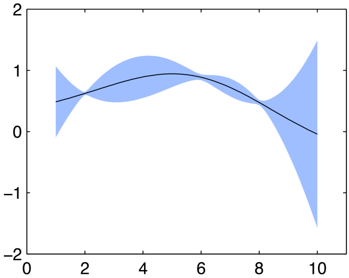
- Are **stochastic** (active regions appear randomly)
- Characterised by **time-scales similar to those associated with planets** (days to years)
- **Quasi-periodic** (periodic stellar rotation + evolving active regions + long-term activity cycles)
- Characterised by some **degree of smoothness** (active regions don't change instantaneously)

So what **is** a Gaussian process?

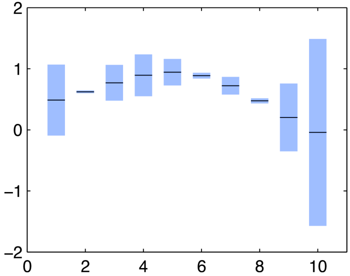
(And why should we care?)

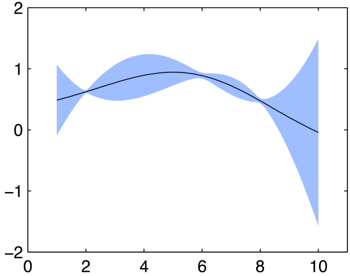
Introducing GPs





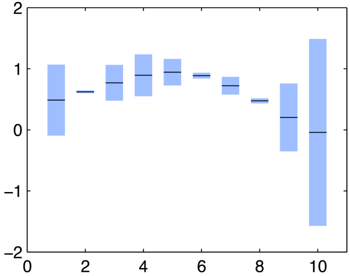
Introducing GPs

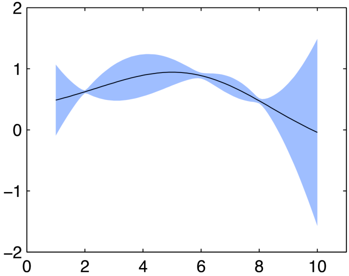




- An **infinite-dimensional Gaussian distribution** (cf. multivariate Gaussian)

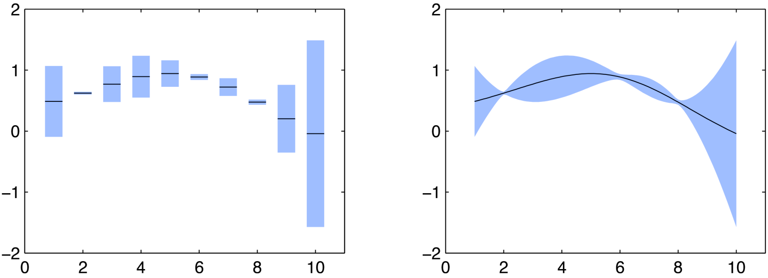
Introducing GPs





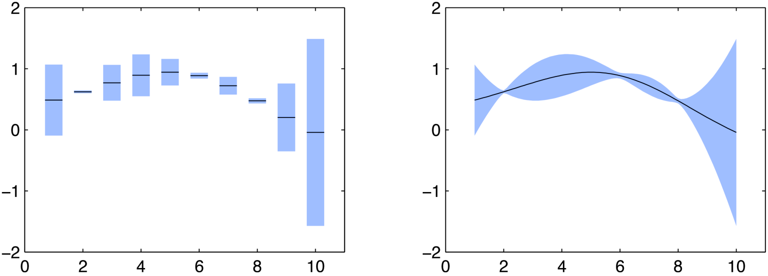
- An **infinite-dimensional Gaussian distribution** (cf. multivariate Gaussian)
- A **distribution over functions** (cf. distribution over vectors)

Introducing GPs



- An **infinite-dimensional Gaussian distribution** (cf. multivariate Gaussian)
- A **distribution over functions** (cf. distribution over vectors)
- Fully specified by a **mean function** and **covariance function** (cf. mean vector, covariance matrix)

Introducing GPs



- An **infinite-dimensional Gaussian distribution** (cf. multivariate Gaussian)
- A **distribution over functions** (cf. distribution over vectors)
- Fully specified by a **mean function** and **covariance function** (cf. mean vector, covariance matrix)
- In practice, we **parametrise** the mean and covariance functions (instead of writing down an infinite number of values!)

Why GPs?

Why GPs?

- Flexible yet principled, data-driven way to perform **Bayesian inference about functions** $\rightarrow$ rigorous treatment of uncertainty, model comparison, etc.

Why GPs?

- Flexible yet principled, data-driven way to perform **Bayesian inference about functions** $\rightarrow$ rigorous treatment of uncertainty, model comparison, etc.

- Obey lots of **convenient analytical identities**

Why GPs?

- Flexible yet principled, data-driven way to perform **Bayesian inference about functions** → rigorous treatment of uncertainty, model comparison, etc.
- Obey lots of **convenient analytical identities**
- Can **model functions by parametrising covariance between data points** → signal variance, evolution time-scales, (quasi)periodicities, smoothness, noise levels, stationarity, isotropy, etc.

Towards GP regression

$$y_i = f(t_i) + \varepsilon_i \quad , f \sim \mathcal{GP}(\mathbf{m}, \mathbf{K})$$

$$\mathbf{m} = m(\mathbf{t}; \phi)$$

mean function
deterministic components
(e.g. planets)

$$\mathbf{K}_{ij} = k(t_i, t_j; \theta)$$

covariance function
stochastic processes/stuff we can't
parametrise directly (e.g. activity)

hyper-parameters

$$\log \mathcal{L}(\theta, \phi) \propto (\mathbf{y} - \mathbf{m})^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{y} - \mathbf{m}) \quad + \log \det \mathbf{K}$$

Towards GP regression

$$y_i = f(t_i) + \varepsilon_i \quad , f \sim \mathcal{GP}(\mathbf{m}, \mathbf{K})$$

$$\mathbf{m} = m(\mathbf{t}; \phi)$$

mean function
deterministic components
(e.g. planets)

$$\mathbf{K}_{ij} = k(t_i, t_j; \theta)$$

covariance function
stochastic processes/stuff we can't
parametrise directly (e.g. activity)

hyper-parameters

$$\log \mathcal{L}(\theta, \phi) \propto (\mathbf{y} - \mathbf{m})^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{y} - \mathbf{m}) \quad + \log \det \mathbf{K}$$

Towards GP regression

$$y_i = f(t_i) + \varepsilon_i \quad , f \sim \mathcal{GP}(\mathbf{m}, \mathbf{K})$$

$$\mathbf{m} = m(\mathbf{t}; \phi)$$

mean function
deterministic components
(e.g. planets)

$$\mathbf{K}_{ij} = k(t_i, t_j; \theta)$$

covariance function
stochastic processes/stuff we can't parametrise directly (e.g. activity)

hyper-parameters

$$\log \mathcal{L}(\theta, \phi) \propto (\mathbf{y} - \mathbf{m})^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{y} - \mathbf{m}) \quad + \log \det \mathbf{K}$$

Towards GP regression

$$y_i = f(t_i) + \varepsilon_i \quad , f \sim \mathcal{GP}(\mathbf{m}, \mathbf{K})$$

mean function
deterministic components
(e.g. planets)

$$\mathbf{m} = m(\mathbf{t}; \phi)$$

covariance function
stochastic processes/stuff we can't
parametrise directly (e.g. activity)

$$\mathbf{K}_{ij} = k(t_i, t_j; \theta)$$

hyper-parameters

$$\log \mathcal{L}(\theta, \phi) \propto (\mathbf{y} - \mathbf{m})^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{y} - \mathbf{m}) \;\; + \log \det \mathbf{K}$$

Towards GP regression

$$y_i = f(t_i) + \varepsilon_i \quad , f \sim \mathcal{GP}(\mathbf{m}, \mathbf{K})$$

$$\mathbf{m} = m(\mathbf{t}; \phi)$$

mean function
deterministic components
(e.g. planets)

$$\mathbf{K}_{ij} = k(t_i, t_j; \theta)$$

covariance function
stochastic processes/stuff we can't
parametrise directly (e.g. activity)

hyper-parameters

$$\log \mathcal{L}(\theta, \phi) \propto (\mathbf{y} - \mathbf{m})^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{y} - \mathbf{m}) \quad + \log \det \mathbf{K}$$

Example: quasi-periodic covariance function

$$k(\tau) \propto \exp\left(\frac{-\tau^2}{2\lambda_e^2}\right)\exp\left(\frac{-\sin^2(\pi\tau/P)}{2\lambda_P^2}\right),$$

evolutionary time-scale (λ_e)

roughness/ structure per period (λ_P)

characteristic period (P)

where $\tau := t - t'$

time between any 2 points

Example: quasi-periodic covariance function

$$k(\tau) \propto \exp\left(\frac{-\tau^2}{2\lambda_e^2}\right)\exp\left(\frac{-\sin^2(\pi\tau/P)}{2\lambda_P^2}\right),$$

evolutionary time-scale

roughness/ structure per period

characteristic period

where $\tau := t - t'$

time between any 2 points

Example: quasi-periodic covariance function

$$k(\tau) \propto \exp\left(\frac{-\tau^2}{2\lambda_e^2}\right)\exp\left(\frac{-\sin^2(\pi\tau / P)}{2\lambda_P^2}\right),$$

evolutionary time-scale

roughness/ structure per period

characteristic period

where $\tau := t - t'$

time between any 2 points

Example: quasi-periodic covariance function

$$k(\tau) \propto \exp\left(\frac{-\tau^2}{2\lambda_e^2}\right)\exp\left(\frac{-\sin^2(\pi\tau/P)}{2\lambda_P^2}\right),$$

evolutionary time-scale (λ_e)

roughness/structure per period (λ_P)

characteristic period (P)

where $\tau := t - t'$

time between any 2 points (τ)

Example: quasi-periodic covariance function

$$k(\tau) \propto \exp\left(\frac{-\tau^2}{2\lambda_e^2}\right)\exp\left(\frac{-\sin^2(\pi\tau/P)}{2\lambda_P^2}\right),$$

λ_e: evolutionary time-scale

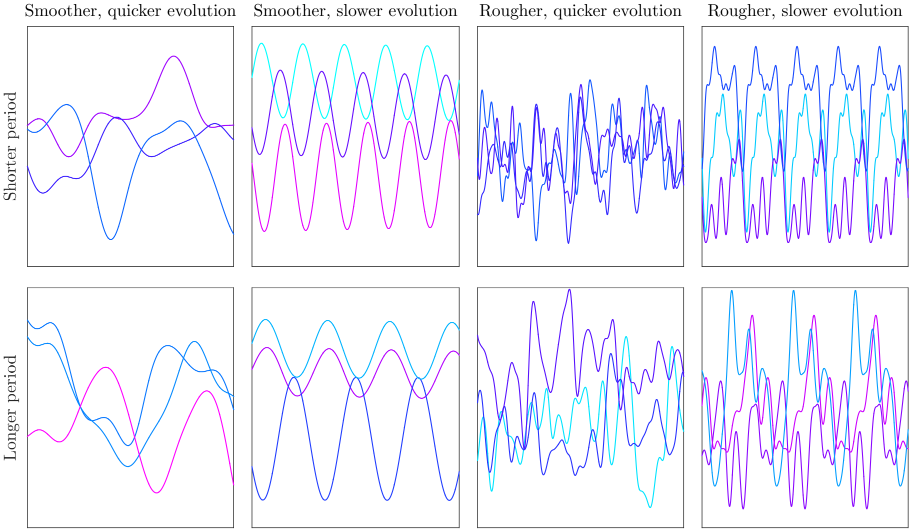
λ_P: roughness/ structure per period

P: characteristic period

where $\tau := t - t'$

τ: time between any 2 points

Function draws: quasi-periodic covariance function



Why not a parametric model?

[...] fitting sine waves at the rotational period of the star and the significant harmonics [...]

The global model fitted on the RVs is therefore:

$$
\begin{aligned}
\text{subset 2008} \quad : \quad & lin0 + lin1 \cdot JDB_{2008} + lin2 \cdot JDB_{2008}^2 + A_{RV-Rhk} \cdot RHK_{low\,freq,2008} \\
\text{subset 2009} \quad : \quad & lin0 + lin1 \cdot JDB_{2009} + lin2 \cdot JDB_{2009}^2 + A_{RV-Rhk} \cdot RHK_{low\,freq,2009} \\
& + A11s \cdot sin(\frac{2\pi}{P1}) \cdot JDB_{2009} + A11c \cdot cos(\frac{2\pi}{P1}) \cdot JDB_{2009} \\
& + A12s \cdot sin(\frac{2\pi}{P1/2}) \cdot JDB_{2009} + A12c \cdot cos(\frac{2\pi}{P1/2}) \cdot JDB_{2009} \\
\text{subset 2010} \quad : \quad & lin0 + lin1 \cdot JDB_{2010} + lin2 \cdot JDB_{2010}^2 + A_{RV-Rhk} \cdot RHK_{low\,freq,2010} \\
& + A21s \cdot sin(\frac{2\pi}{P2}) \cdot JDB_{2010} + A21c \cdot cos(\frac{2\pi}{P2}) \cdot JDB_{2010} \\
& + A23s \cdot sin(\frac{2\pi}{P2/3}) \cdot JDB_{2010} + A23c \cdot cos(\frac{2\pi}{P2/3}) \cdot JDB_{2010} \\
& + A24s \cdot sin(\frac{2\pi}{P2/4}) \cdot JDB_{2010} + A24c \cdot cos(\frac{2\pi}{P2/4}) \cdot JDB_{2010} \\
\text{subset 2011} \quad : \quad & lin0 + lin1 \cdot JDB_{2011} + lin2 \cdot JDB_{2011}^2 + A_{RV-Rhk} \cdot RHK_{low\,freq,2011} \\
& + A31s \cdot sin(\frac{2\pi}{P3}) \cdot JDB_{2011} + A31c \cdot cos(\frac{2\pi}{P3}) \cdot JDB_{2011} \\
& + A32s \cdot sin(\frac{2\pi}{P3/2}) \cdot JDB_{2011} + A32c \cdot cos(\frac{2\pi}{P3/2}) \cdot JDB_{2011} \\
& + A33s \cdot sin(\frac{2\pi}{P3/3}) \cdot JDB_{2011} + A33c \cdot cos(\frac{2\pi}{P3/3}) \cdot JDB_{2011}
\end{aligned}
$$

→ 23 free parameters (before adding a planet)

Activity model for Alpha Cen B, from Dumusque *et al.* (2012)

Example: quasi-periodic covariance function

$$k(\tau) \propto \exp\left(\frac{-\tau^2}{2\lambda_e^2}\right)\exp\left(\frac{-\sin^2(\pi\tau/P)}{2\lambda_P^2}\right),$$

evolutionary time-scale (λ_e)

roughness/structure per period (λ_P)

characteristic period (P)

where $\tau := t - t'$

time between any 2 points (τ)

Recap

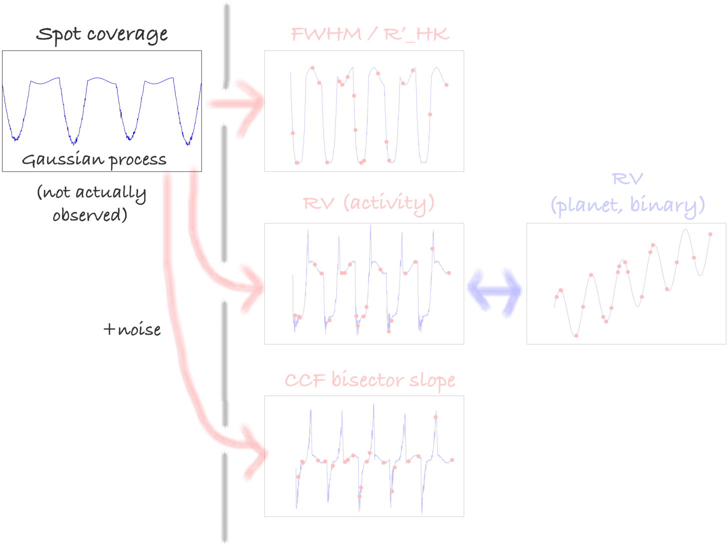
GPs $\rightarrow$ easy Bayesian inference about functions

Now, on to the science!

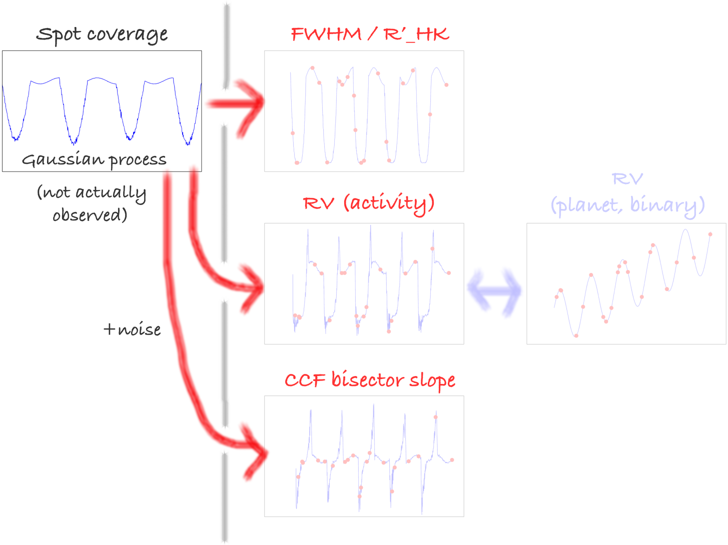
① GPs to **disentangle** activity and planetary signals

a.k.a. model ALL available data

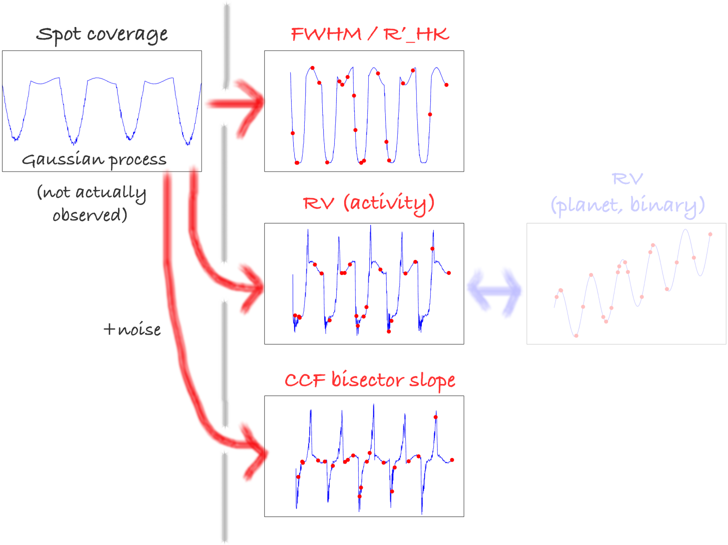
Disentangling activity and planetary signals



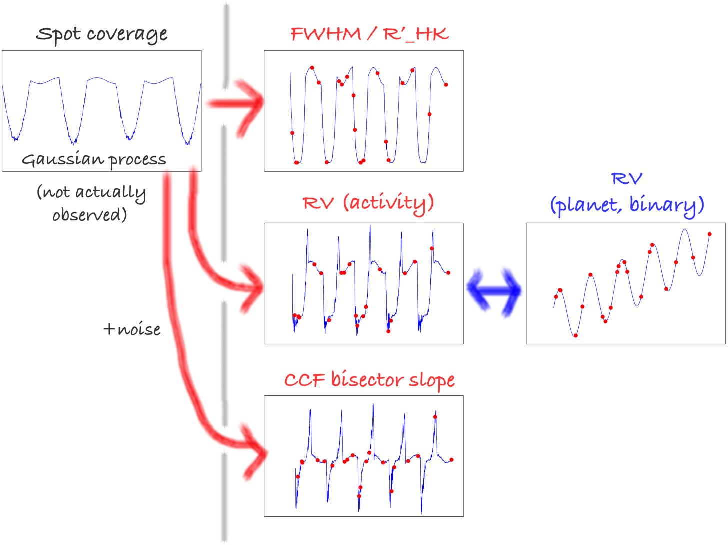
Disentangling activity and planetary signals



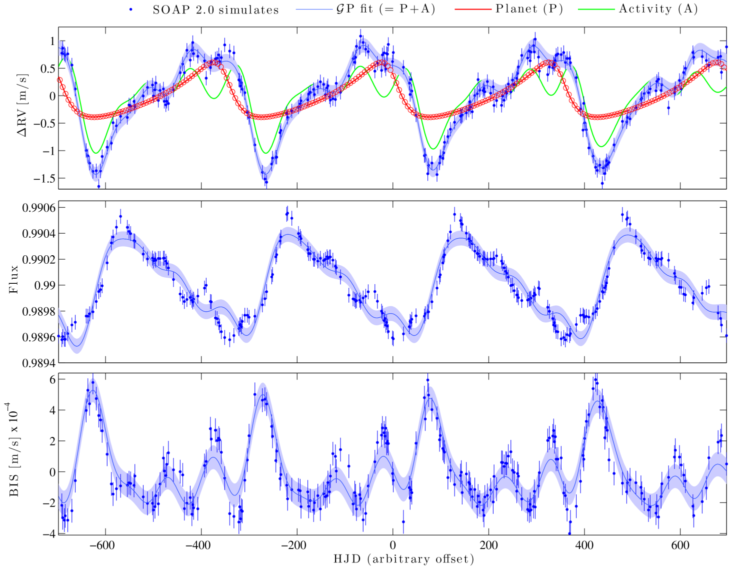
Disentangling activity and planetary signals



Disentangling activity and planetary signals



It works as intended



How to link different observables?

How to link different observables?

- Affine transformations (Σ, $\int dt$, $\frac{d}{dt}$, etc.) of a GP yields another GP

How to link different observables?

- Affine transformations (Σ, $\int dt$, $\frac{d}{dt}$, etc.) of a GP yields another GP
- $\rightarrow$ Can use e.g. the FF' method, which says $\Delta \mathrm{RV}$ related to photometric flux and its first derivative (Aigrain *et al.*, 2012, MNRAS)

How to link different observables?

- Affine transformations (Σ, $\int dt$, $\frac{d}{dt}$, etc.) of a GP yields another GP
- $\rightarrow$ Can use e.g. the FF' method, which says $\Delta \mathrm{RV}$ related to photometric flux and its first derivative (Aigrain *et al.*, 2012, MNRAS)
- More details on the physics:

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **452,** 2269–2291 (2015) doi:10.1093/mnras/stv1428

A Gaussian process framework for modelling stellar activity signals in radial velocity data

V. Rajpaul,[1★] S. Aigrain,[1] M. A. Osborne,[2] S. Reece[2] and S. Roberts[2]

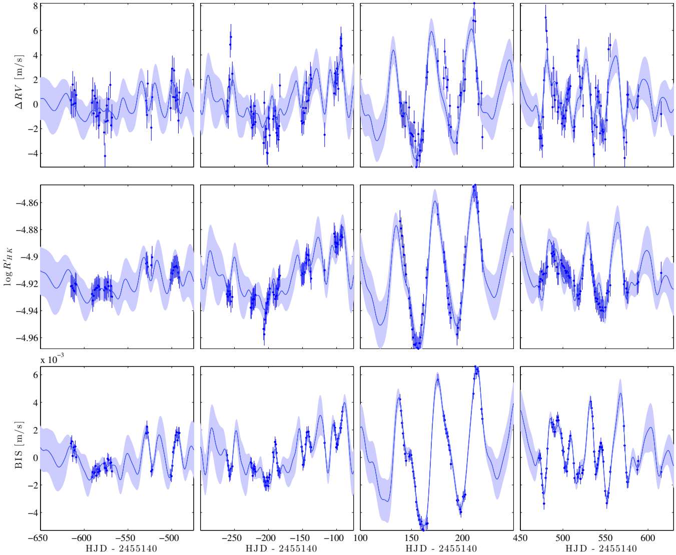
[1] *Subdepartment of Astrophysics, Department of Physics, University of Oxford, Oxford OX1 3RH, UK*
[2] *Pattern Recognition and Machine Learning Group, Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK*

Accepted 2015 June 24. Received 2015 June 22; in original form 2015 March 16

Some science results

1. Proof of concept: accurately **recover already-published exoplanet parameters**, e.g. for Gliese 15 A, CoRoT-7 (Rajpaul *et al.*, 2015; MNRAS)
2. New result: **discovery of HD175607 b**, most metal-poor G dwarf with an orbiting sub-Neptune; $P_{\mathrm{star}} \approx P_{\mathrm{planet}}$ (Mortier *et al.*, 2015; A&A)
3. New result: **demonstration that Alpha Cen Bb is a false positive** (Rajpaul *et al.*, 2016; MNRAS Letters)

Four seasons of Alpha Cen B data



② GPs as a powerful **simulation tool**

GPs as a simulation tool

GPs as a simulation tool

- Can **train GPs on real data** containing noise and astrophysical nuisance signals, instrumental effects and other systematics, uneven sampling, etc.

GPs as a simulation tool

- Can **train GPs on real data** containing noise and astrophysical nuisance signals, instrumental effects and other systematics, uneven sampling, etc.

- Can then randomly generate **realistic synthetic data**, with arbitrary sampling, with same covariance properties as the real data

GPs as a simulation tool

- Can **train GPs on real data** containing noise and astrophysical nuisance signals, instrumental effects and other systematics, uneven sampling, etc.
- Can then randomly generate **realistic synthetic data**, with arbitrary sampling, with same covariance properties as the real data
- Can use this to

GPs as a simulation tool

- Can **train GPs on real data** containing noise and astrophysical nuisance signals, instrumental effects and other systematics, uneven sampling, etc.
- Can then randomly generate **realistic synthetic data**, with arbitrary sampling, with same covariance properties as the real data
- Can use this to
  - study detection limits, optimal observing strategies, etc.

GPs as a simulation tool

- Can **train GPs on real data** containing noise and astrophysical nuisance signals, instrumental effects and other systematics, uneven sampling, etc.
- Can then randomly generate **realistic synthetic data**, with arbitrary sampling, with same covariance properties as the real data
- Can use this to
  - study detection limits, optimal observing strategies, etc.
  - identify **artefacts** introduced through **fitted models**

GPs as a simulation tool

- Can **train GPs on real data** containing noise and astrophysical nuisance signals, instrumental effects and other systematics, uneven sampling, etc.
- Can then randomly generate **realistic synthetic data**, with arbitrary sampling, with same covariance properties as the real data
- Can use this to
  - study detection limits, optimal observing strategies, etc.
  - identify **artefacts** introduced through **fitted models**
  - identify **artefacts** associated with **discrete sampling**

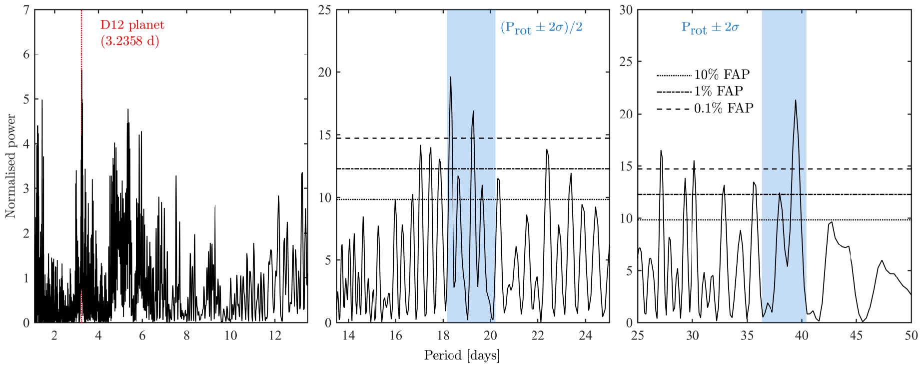
GPs as a simulation tool

- Can **train GPs on real data** containing noise and astrophysical nuisance signals, instrumental effects and other systematics, uneven sampling, etc.
- Can then randomly generate **realistic synthetic data**, with arbitrary sampling, with same covariance properties as the real data
- Can use this to
  - study detection limits, optimal observing strategies, etc.
  - identify **artefacts** introduced through **fitted models**
  - identify **artefacts** associated with **discrete sampling**

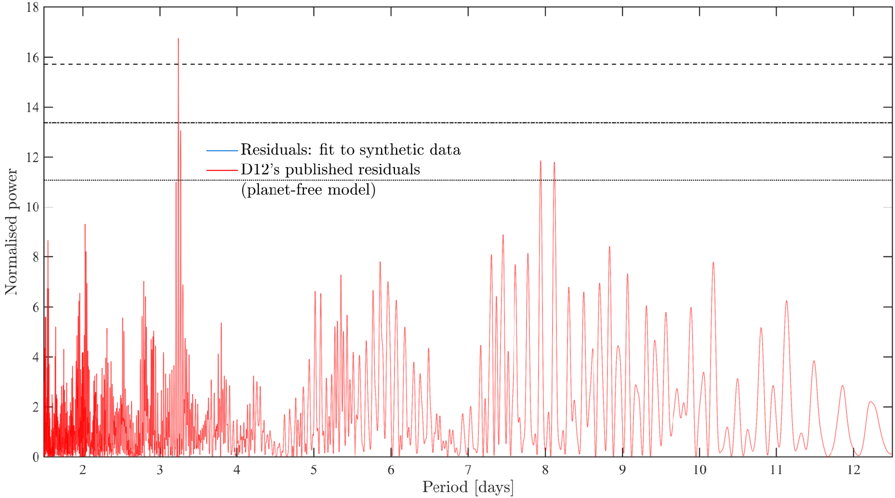
The example of Alpha Cen B

Power spectrum of the window function

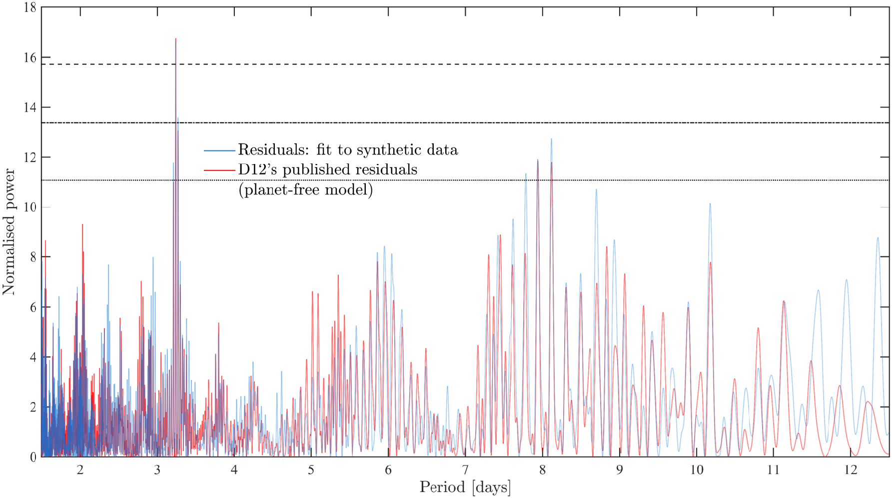
(Rajpaul et *al.*, 2016; MNRAS Letters)



The example of Alpha Cen B



The example of Alpha Cen B



③ GPs to study **periodic phenomena**

a.k.a. we can do better than Lomb-Scargle periodograms

Beyond the Lomb-Scargle periodogram

Beyond the Lomb-Scargle periodogram

- Lomb-Scargle method/LSSA is convenient, easy to understand, and popular...but restrictive, and often inappropriate

Beyond the Lomb-Scargle periodogram

- Lomb-Scargle method/LSSA is convenient, easy to understand, and popular...but restrictive, and often inappropriate
- What about signals that are non-sinusoidal, quasi-periodic, contain correlated noise, etc...?

Beyond the Lomb-Scargle periodogram

- Lomb-Scargle method/LSSA is convenient, easy to understand, and popular...but restrictive, and often inappropriate
- What about signals that are non-sinusoidal, quasi-periodic, contain correlated noise, etc...?
- Drop-in replacement: GP periodogram (with Angus *et al.*)

Quasi-periodic covariance function (again)

$$k(\tau) \propto \exp\left(\frac{-\tau^2}{2\lambda_e^2}\right)\exp\left(\frac{-\sin^2(\pi\tau/P)}{2\lambda_P^2}\right),$$

evolutionary time-scale (λ_e)

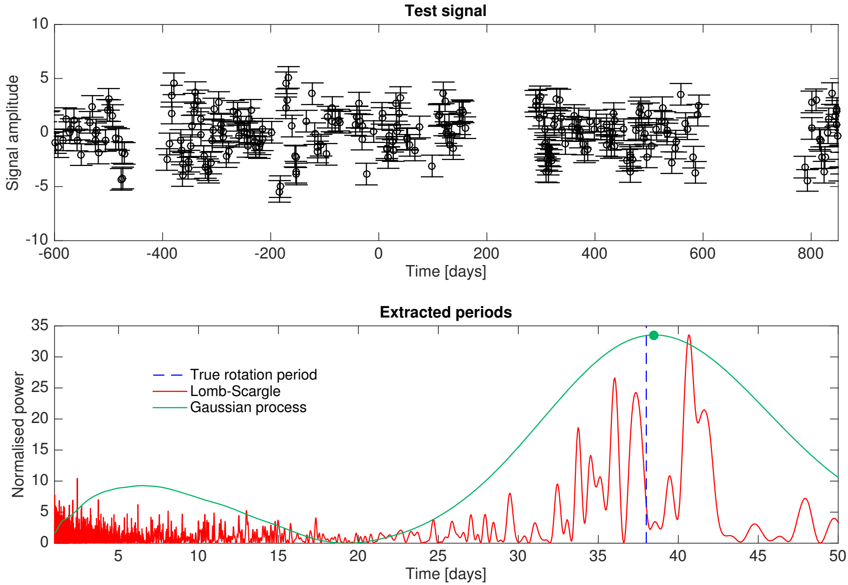
roughness/structure per period (λ_P)

characteristic period (P)

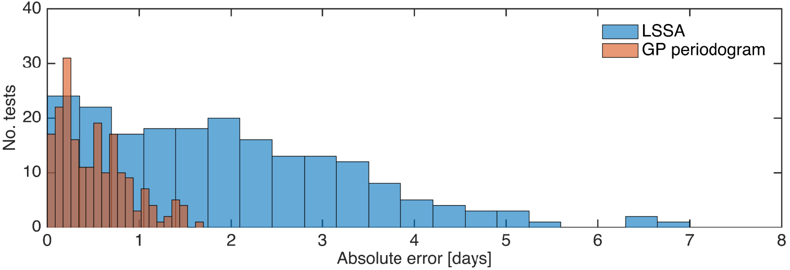
where $\tau := t - t'$

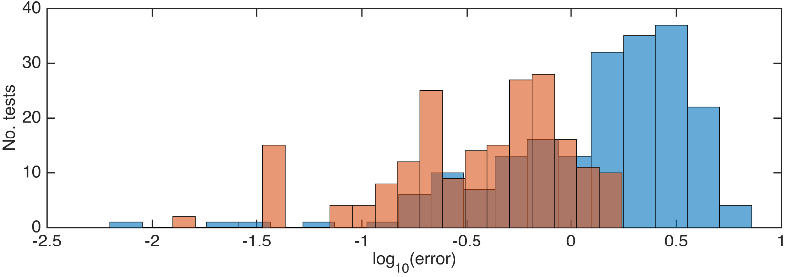
time between any 2 points

GP periodogram



GP periodogram





One step-further: differential rotation

- Assume distribution of quasi-periodic signals, each generated by quasi-periodic kernel. Say $f(P') \sim \mathcal{N}(P, \sigma)$, or $f(P') \sim \mathcal{U}(P - \sigma, P + \sigma)$

- Integrate over this distribution to get new covariance kernel

- Characterise differential rotation via posterior distribution of P and σ

- Early results are promising...watch this space (injection tests currently underway)

④ GPs to extract **precise RVs** directly from spectra

towards < 10 cm/s precisions ?

Using GPs to extract RVs from spectra

- Given spectra of a star observed N times, model all N spectra (lines, continuum) as draws from a **single underlying GP** $+\boldsymbol{\Delta}\lambda$

Using GPs to extract RVs from spectra

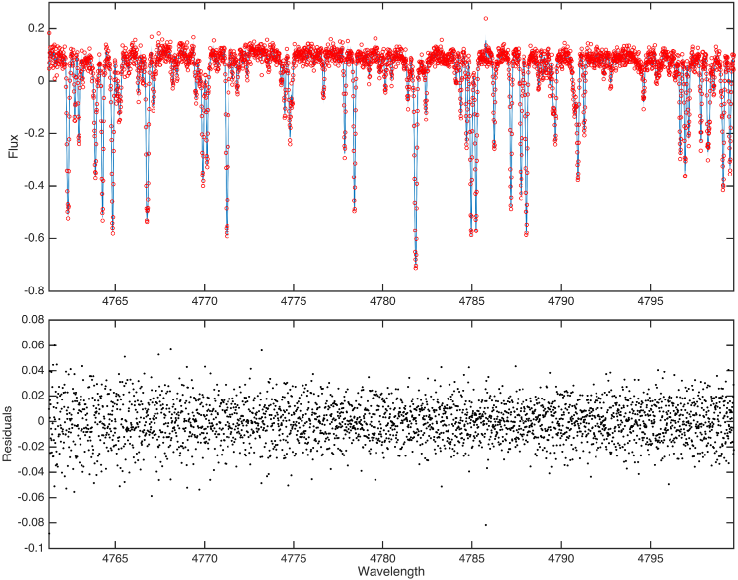
- Given spectra of a star observed N times, model all N spectra (lines, continuum) as draws from a **single underlying GP** $+\boldsymbol{\Delta}\lambda$

- Treat the GP models as **noise-free interpolants** of the spectra

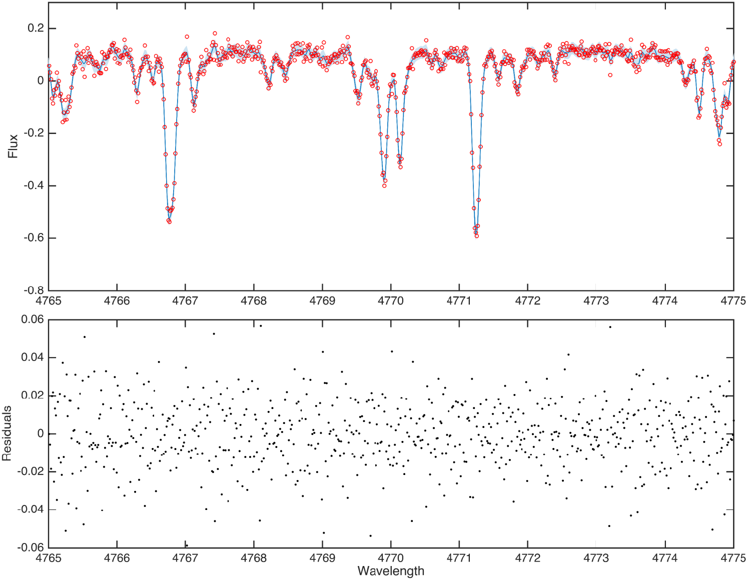
Using GPs to extract RVs from spectra

- Given spectra of a star observed N times, model all N spectra (lines, continuum) as draws from a **single underlying GP** $+\boldsymbol{\Delta}\lambda$
- Treat the GP models as **noise-free interpolants** of the spectra
- Fit $(N-1)$ RV shifts/translations to **maximise alignment**
- Optional: **unshift** all spectra to obtain high resolution "master" spectrum

Ultra high-precision RVs



Ultra high-precision RVs



Some early results

- Where HARPS-N DRS obtains ~ 43 cm/s precision, we manage ~ 37 cm/s precision $(\mathrm{m}_v \sim 6.5)$

Some early results

- Where HARPS-N DRS obtains ~ 43 cm/s precision, we manage ~ 37 cm/s precision ($\mathrm{m}_v \sim 6.5$)
- Fainter stars? Results pending (~ 1 m/s improvement?)...

Some early results

- Where HARPS-N DRS obtains ~ 43 cm/s precision, we manage ~ 37 cm/s precision $(\mathrm{m}_v \sim 6.5)$
- Fainter stars? Results pending (~ 1 m/s improvement?)...
- Beyond RVs: extract better measures of stellar activity?

Summary

Summary

- GPs represent a flexible, principled, data-driven way to perform inference in the presence of

Summary

- GPs represent a flexible, principled, data-driven way to perform inference in the presence of
  - correlated noise and nuisance signals; and/or

Summary

- GPs represent a flexible, principled, data-driven way to perform inference in the presence of
    - correlated noise and nuisance signals; and/or
    - functions we can't parametrise directly

Summary

- GPs represent a flexible, principled, data-driven way to perform inference in the presence of
  - correlated noise and nuisance signals; and/or
  - functions we can't parametrise directly
- Useful addition to the toolbox of anyone trying to detect or characterise exoplanets

Summary

- GPs represent a flexible, principled, data-driven way to perform inference in the presence of
    - correlated noise and nuisance signals; and/or
    - functions we can't parametrise directly
- Useful addition to the toolbox of anyone trying to detect or characterise exoplanets
- Some recent applications on which I've worked
    1. GPs to disentangle activity and planets
    2. GPs as a powerful simulation tool
    3. GPs to characterise periodic phenomena e.g. stellar rotation
    4. GPs for extracting high-precision RVs from spectra

Thanks!
Any questions?

get in
touch

vinesh.rajpaul@
astro.ox.ac.uk