Machine Learning Research Group
UNIVERSITY OF OXFORD

mind foundry

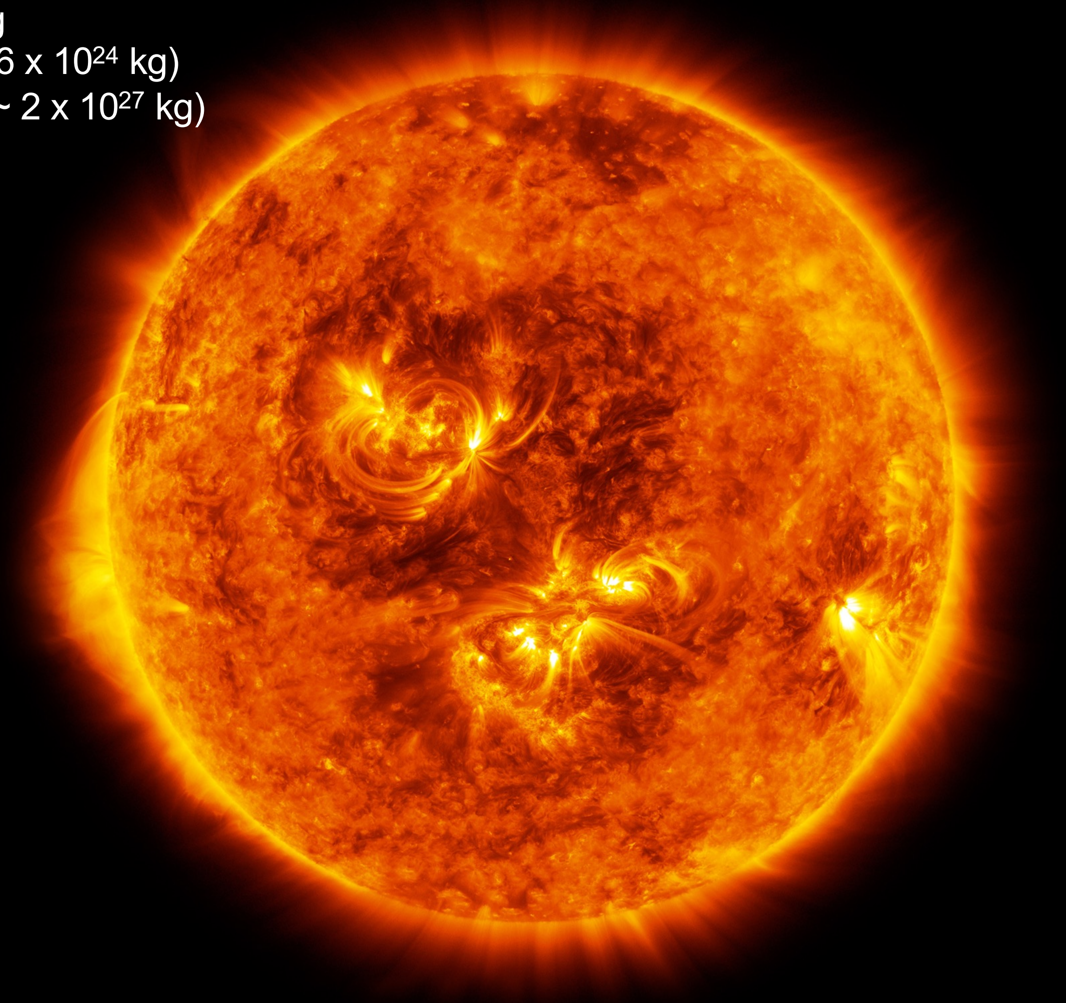
Finding Earth v2: adventures with Gaussian processes and exoplanets

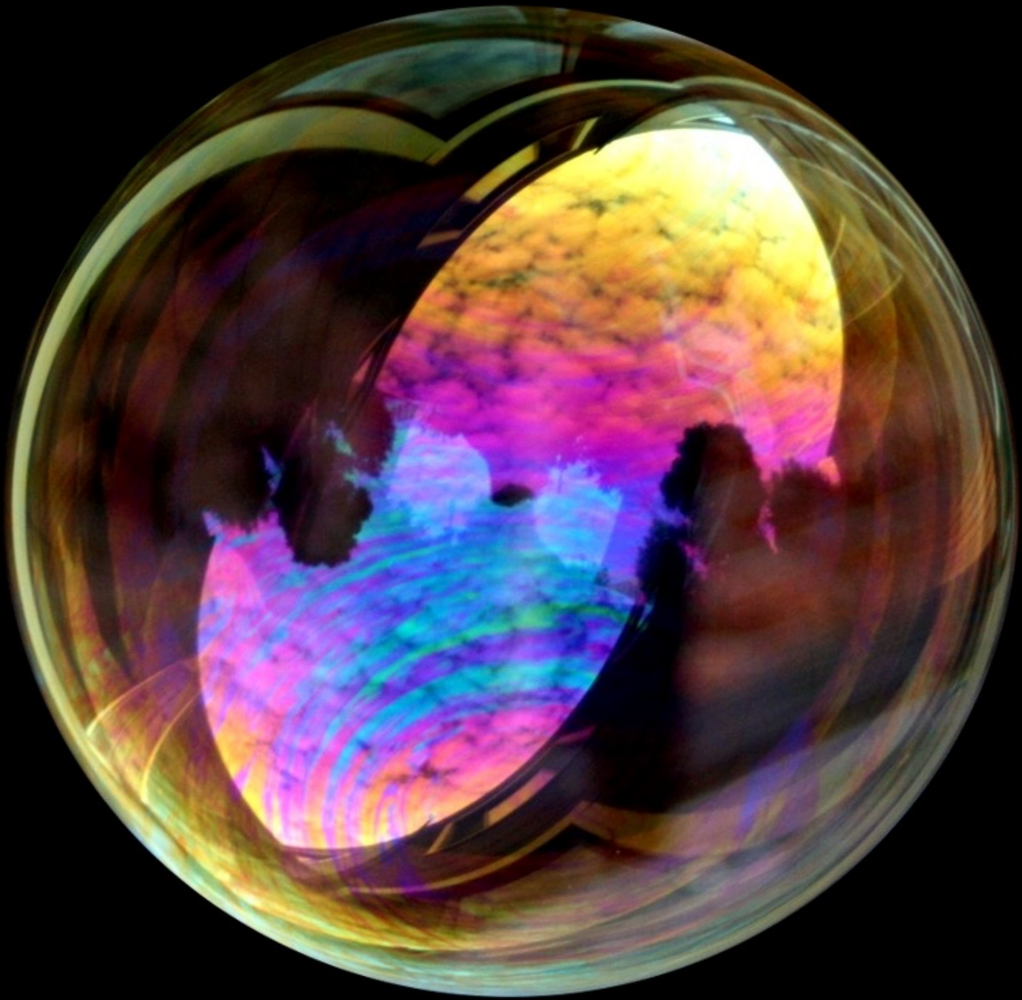
Stephen Roberts

University of Oxford & Mind Foundry

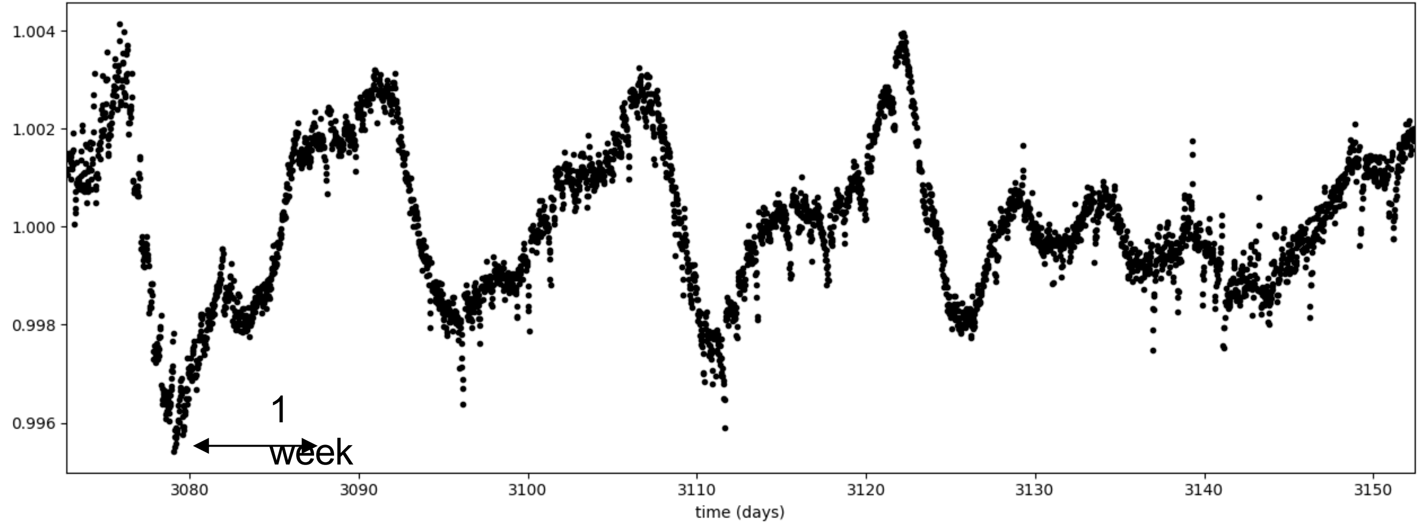
Stars

~ 10^{31} kg
(earth ~ 6×10^{24} kg)
(Jupiter ~ 2×10^{27} kg)





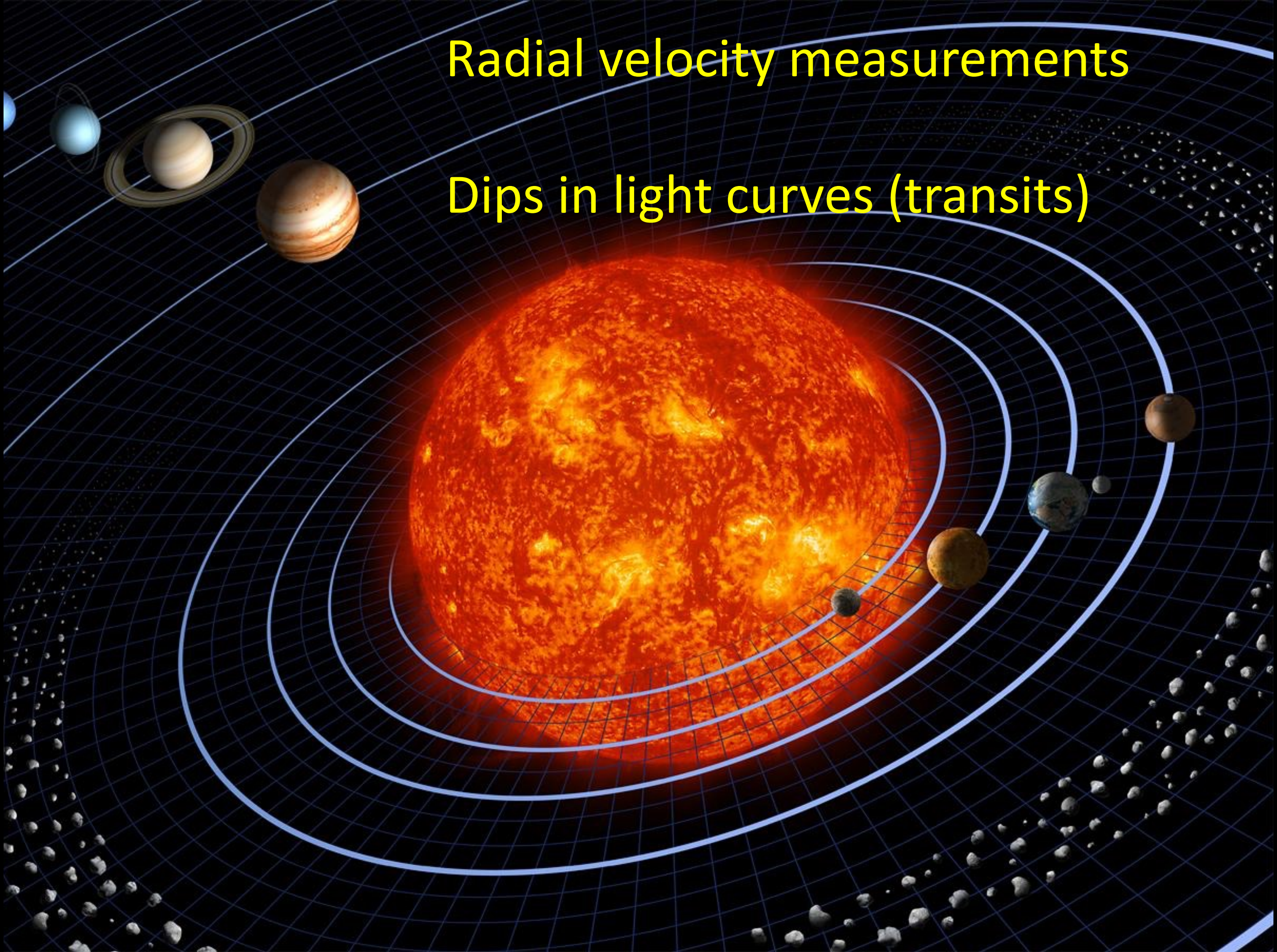
Starlight is not constant

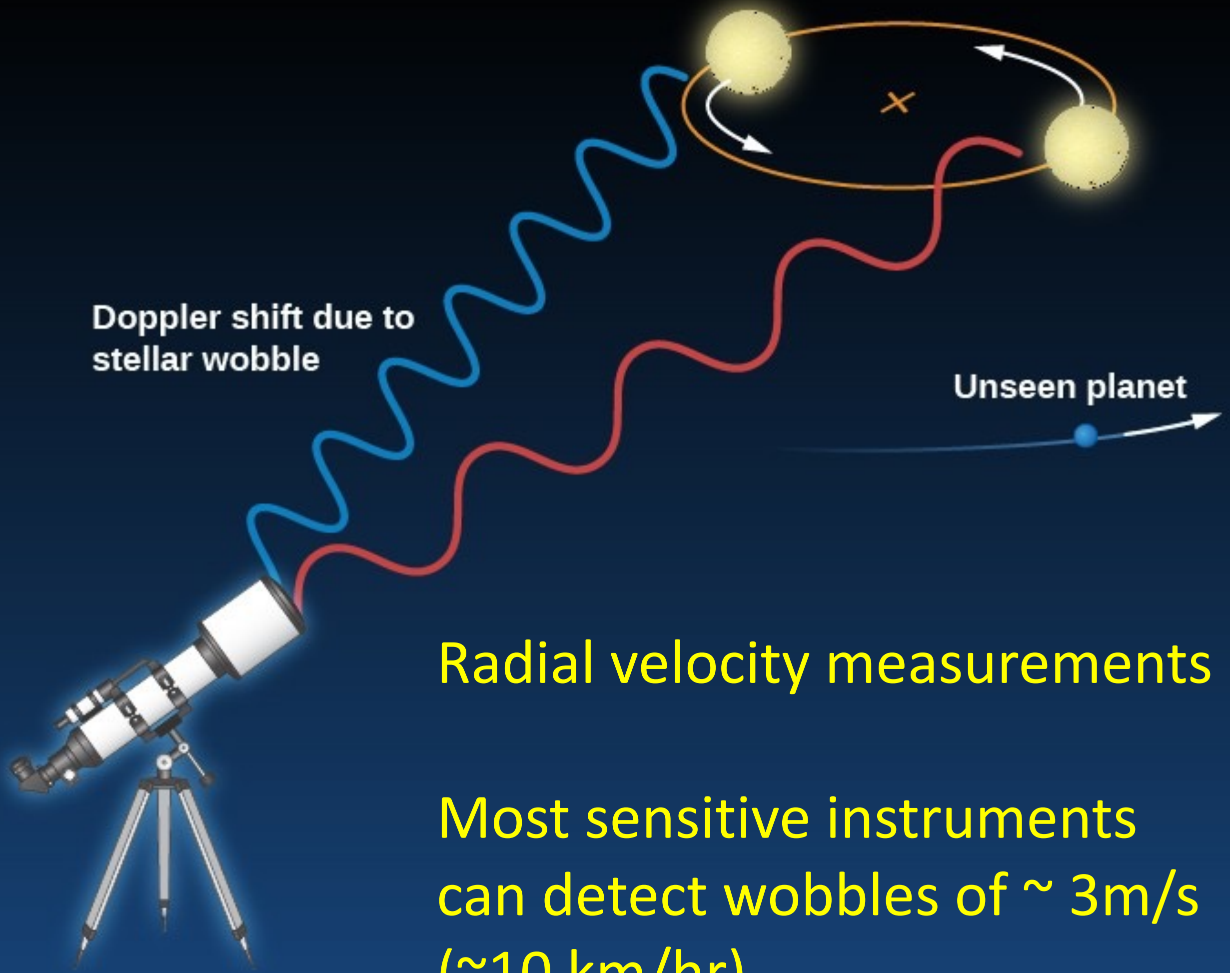


How to find a Planet

Radial velocity measurements

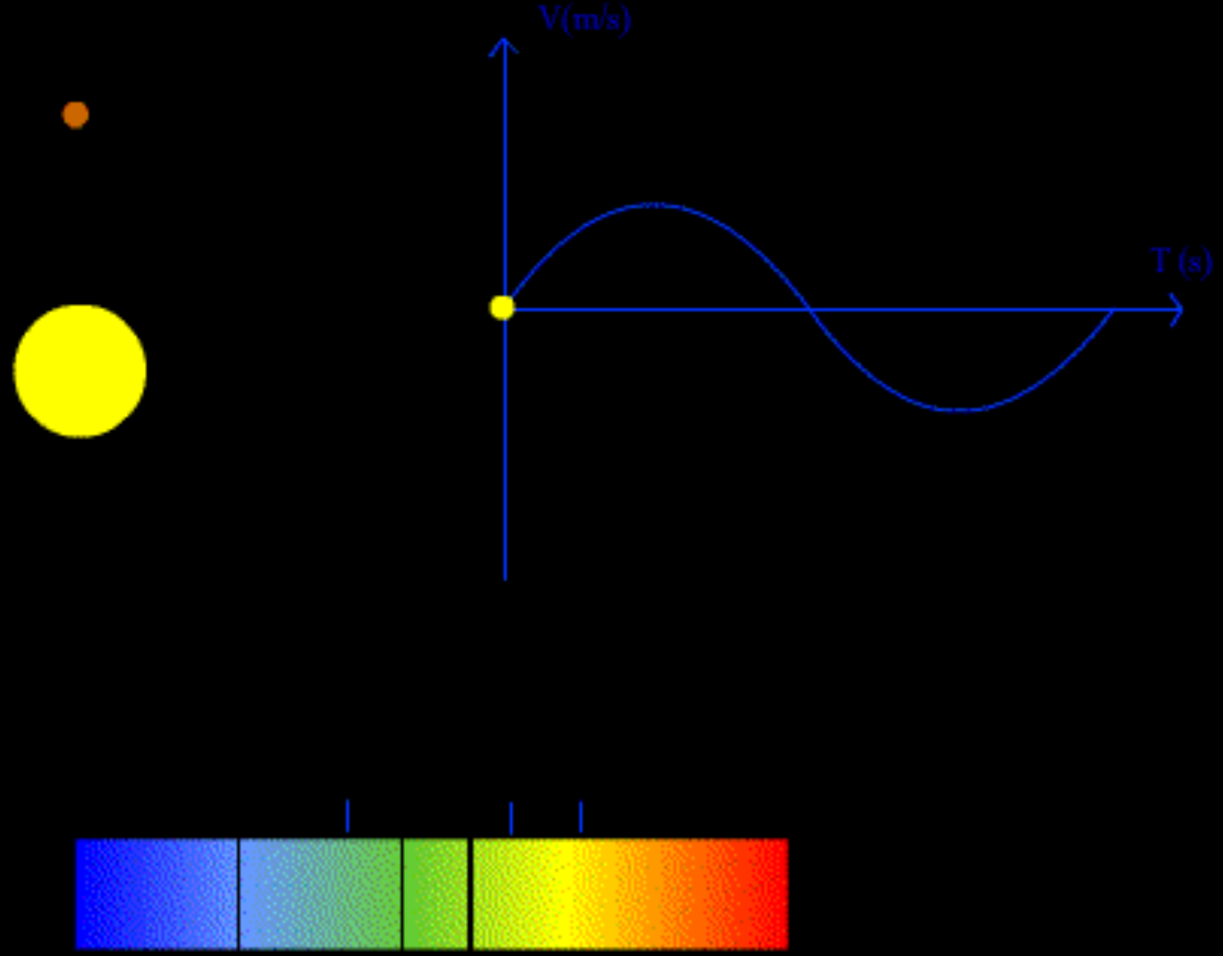
Dips in light curves (transits)



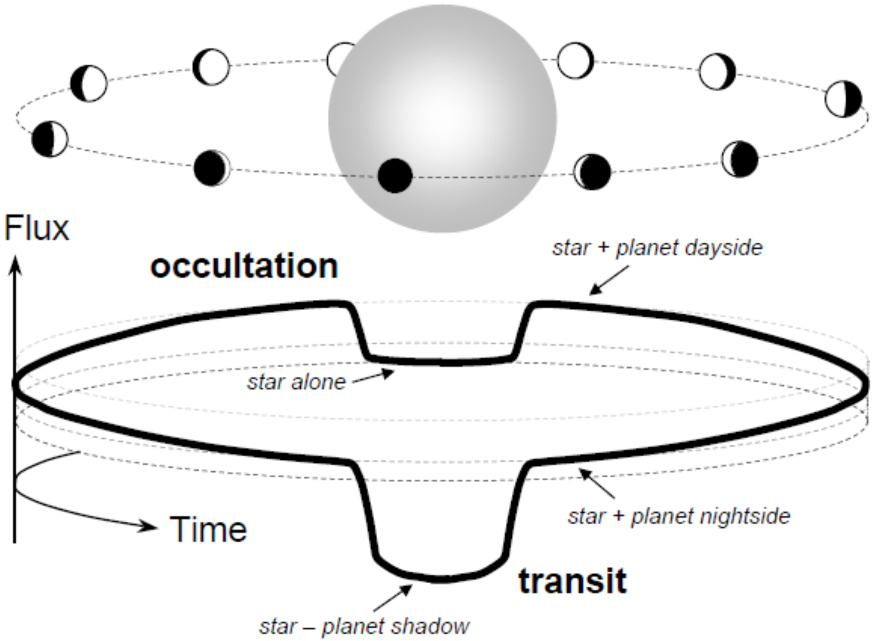


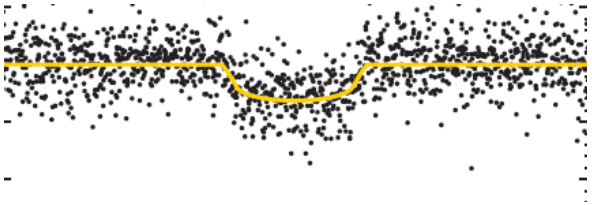
Radial velocity measurements

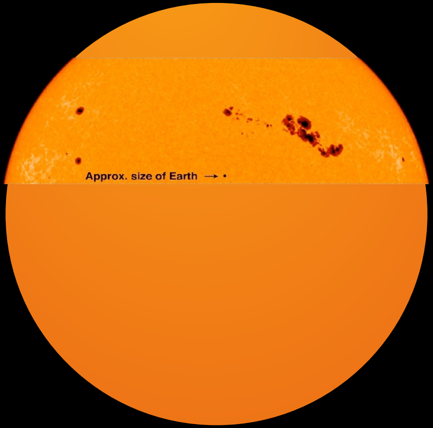
Most sensitive instruments can detect wobbles of ~ 3m/s (~10 km/hr)

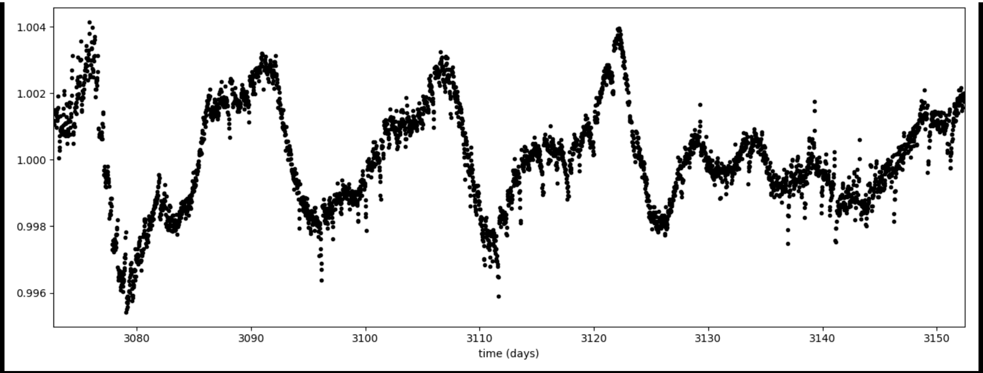

V(m/s)
T (s)

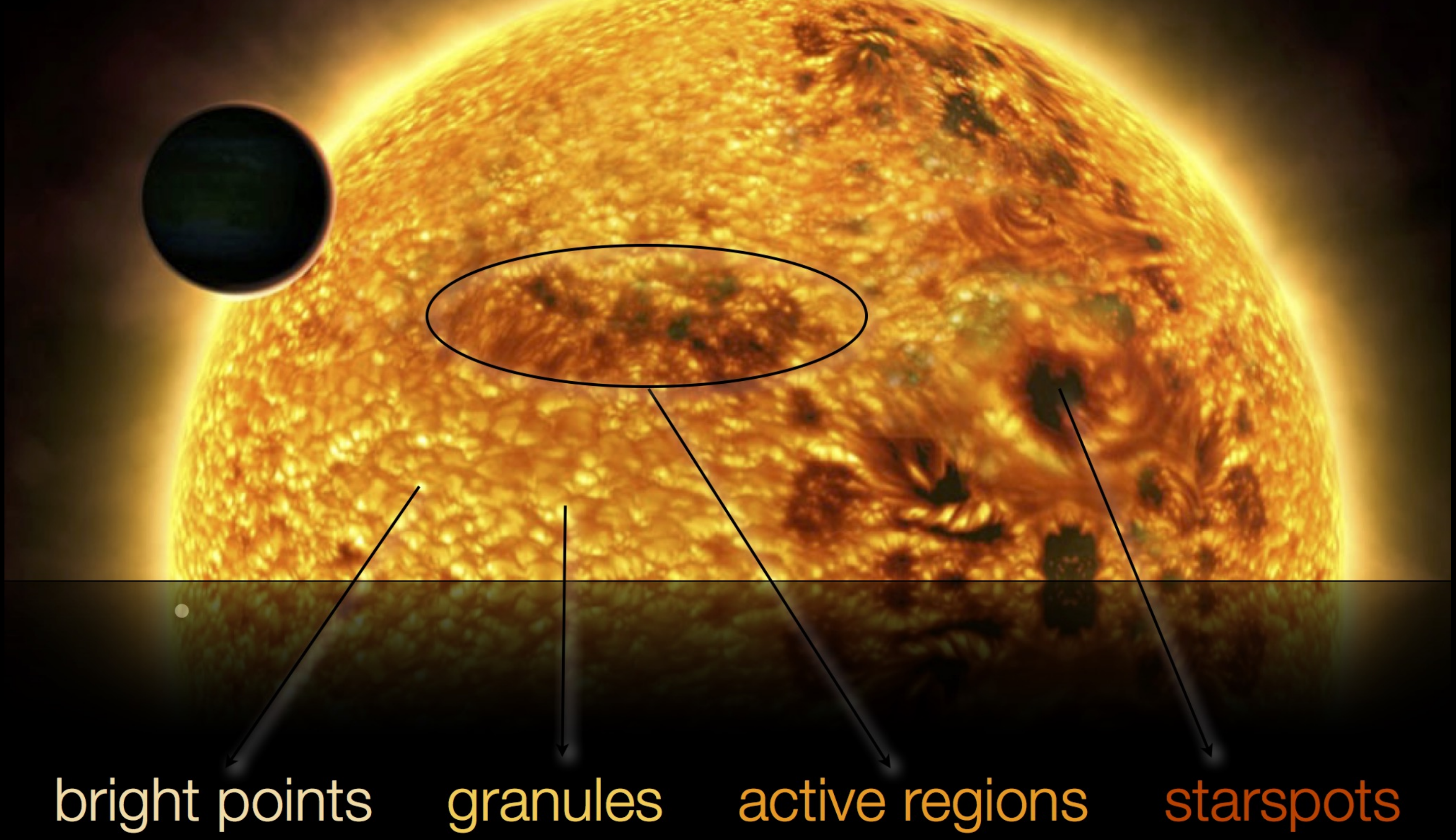
Transits


Flux
occultation
star + planet dayside
star alone
Time
star + planet nightside
transit
star – planet shadow

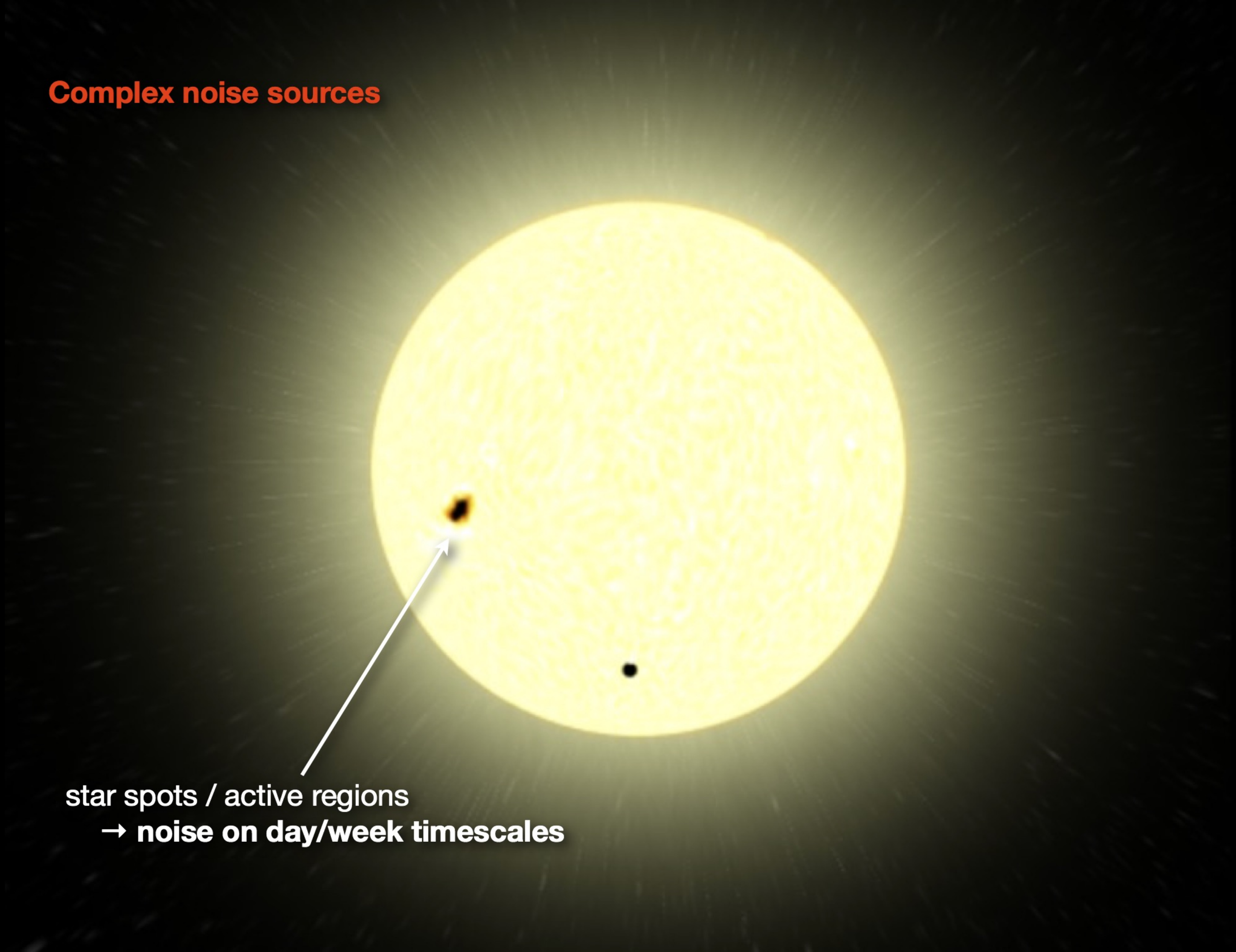



Approx. size of Earth

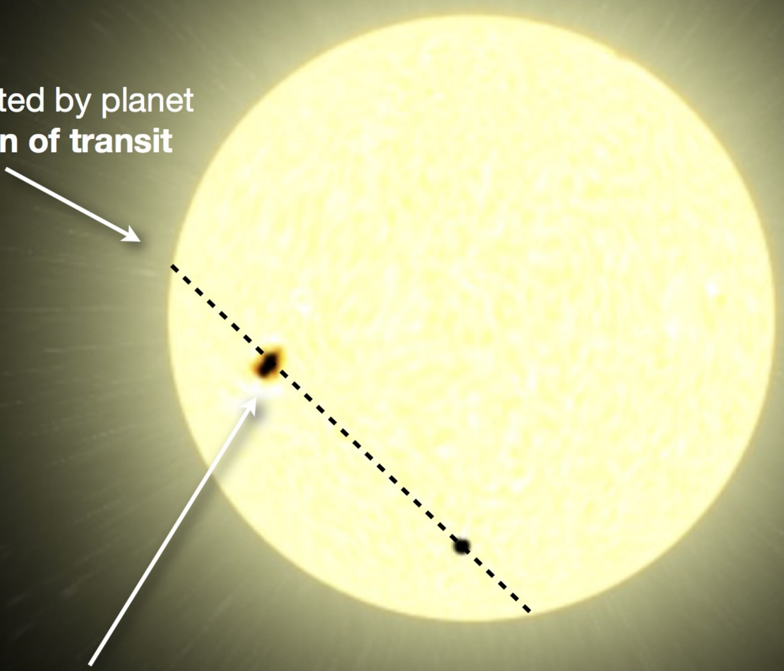

1.004
1.002
1.000
0.998
0.996
3080
3090
3100
3110
3120
3130
3140
3150
time (days)


bright points
granules
active regions
starspots

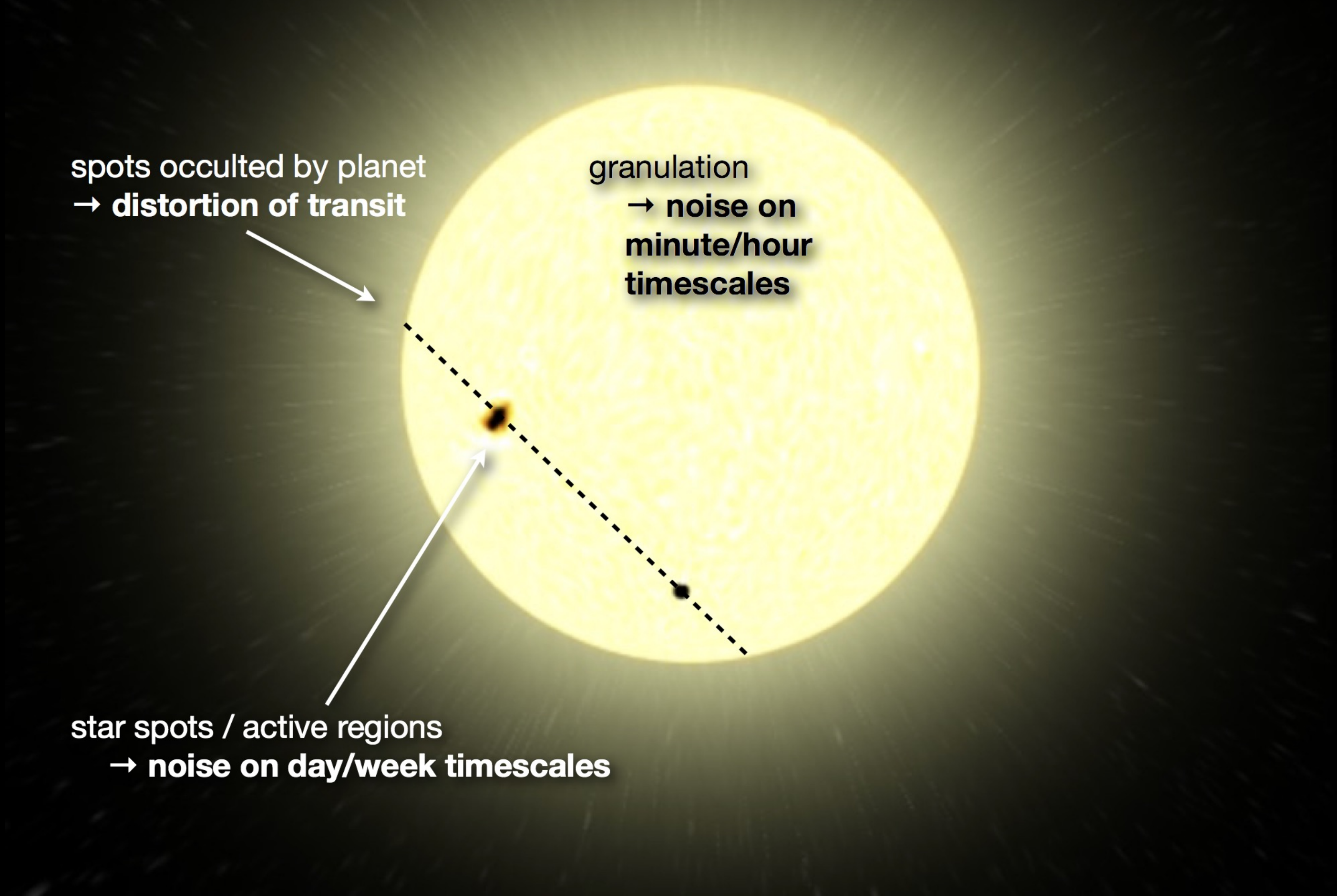
Complex noise sources


Complex noise sources
star spots / active regions
→ noise on day/week timescales

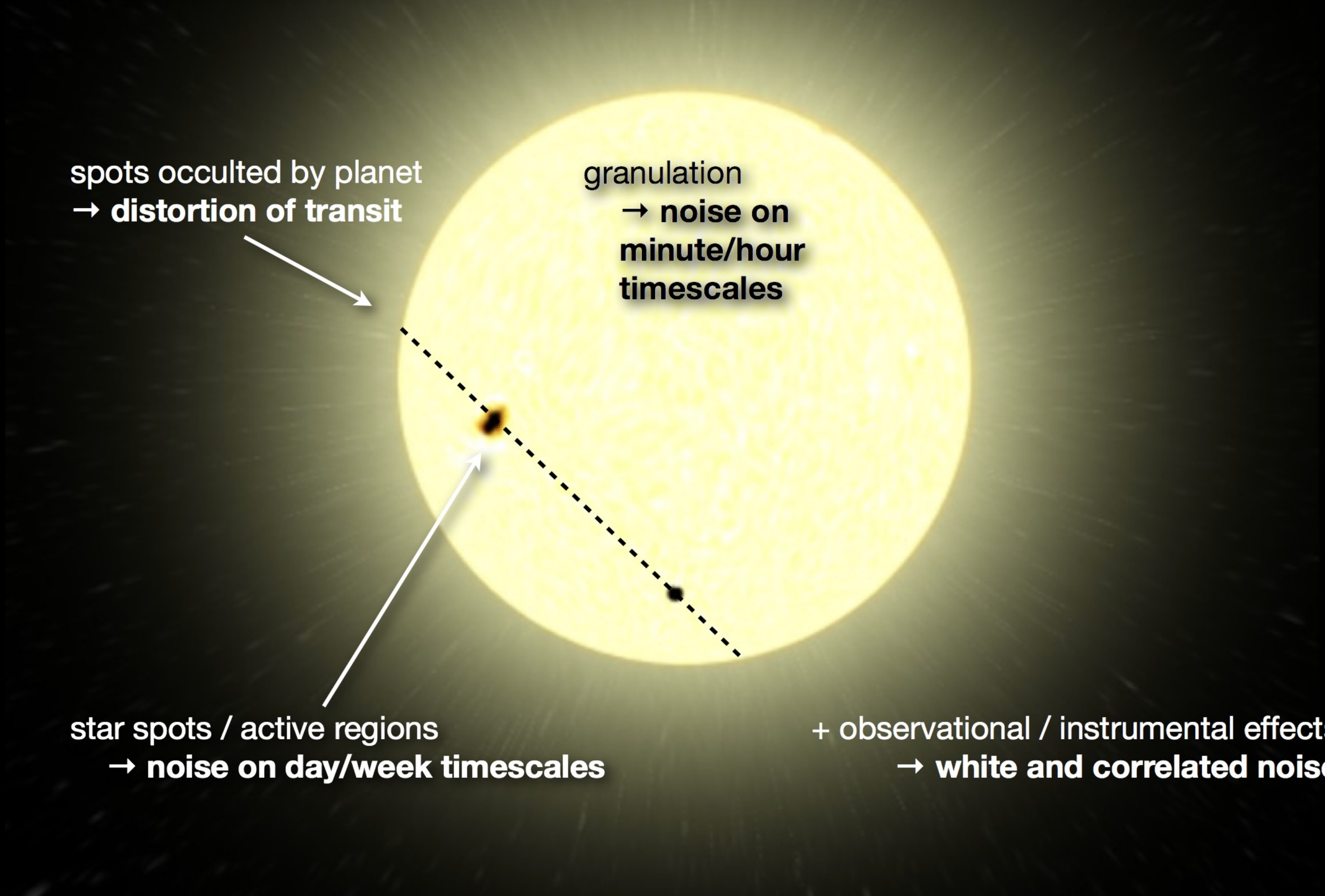
Complex noise sources



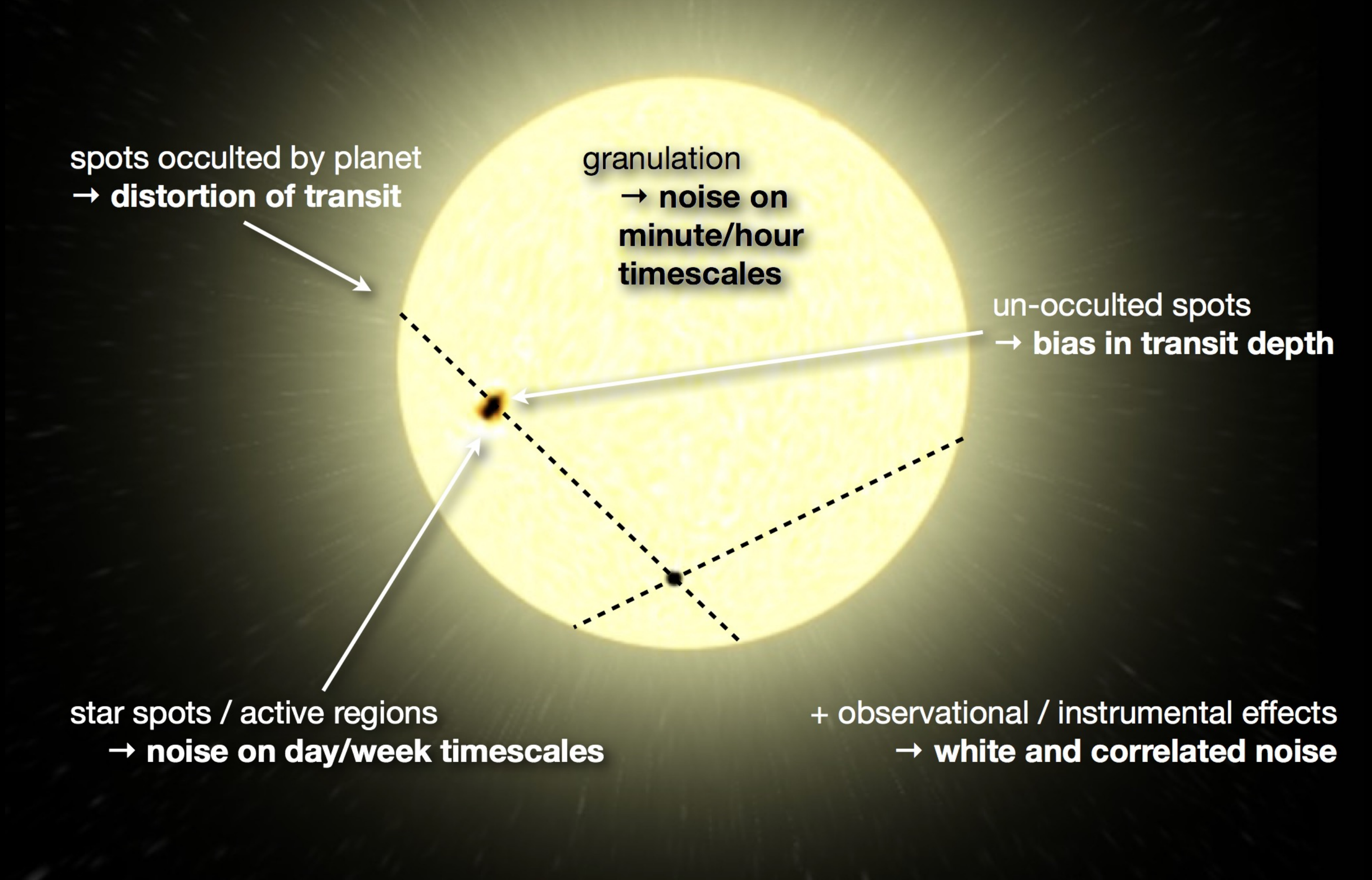
Complex noise sources



Complex noise sources



Complex noise sources



The problems

Data is not perfectly synchronous
Missing data

Abrupt dislocations & saturations
Power cycling
Heating / cooling cycles
Vibrational modes
Non-homogenous calibrations

Sun-spots & Stellar rotation
Stellar activity >> exoplanet signals

The problems

Data is not perfectly synchronous

Missing data

Abrupt dislocations & saturations

Power cycling

Heating / cooling cycles

Vibrational modes

Non-homogenous calibrations

Sun-spots & Stellar rotation

Stellar activity >> exoplanet signals

THERE ARE LOTS OF WOBBLY PROBLEMS AND NOISY BITS!

Kepler space telescope

(2009 – 2014, then...)

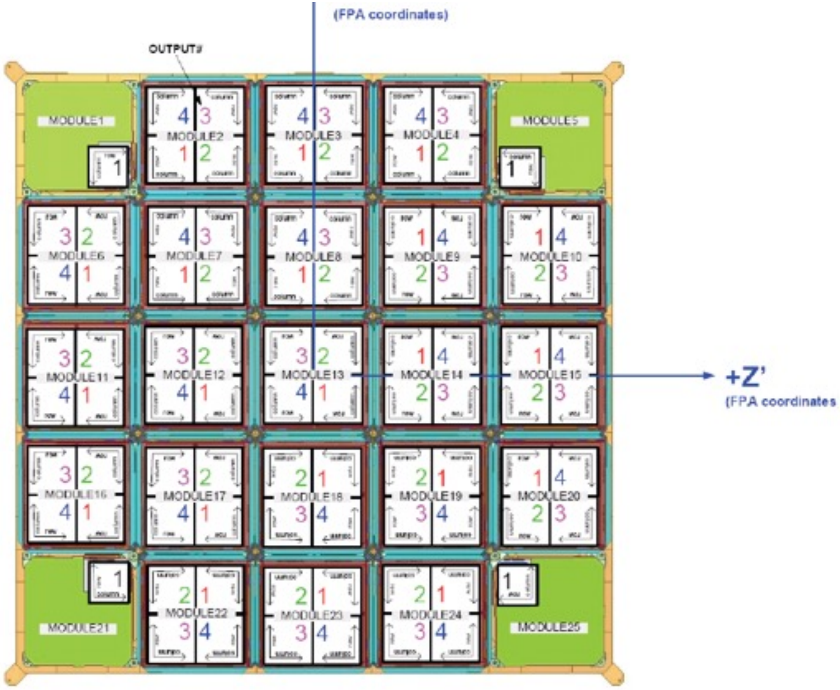
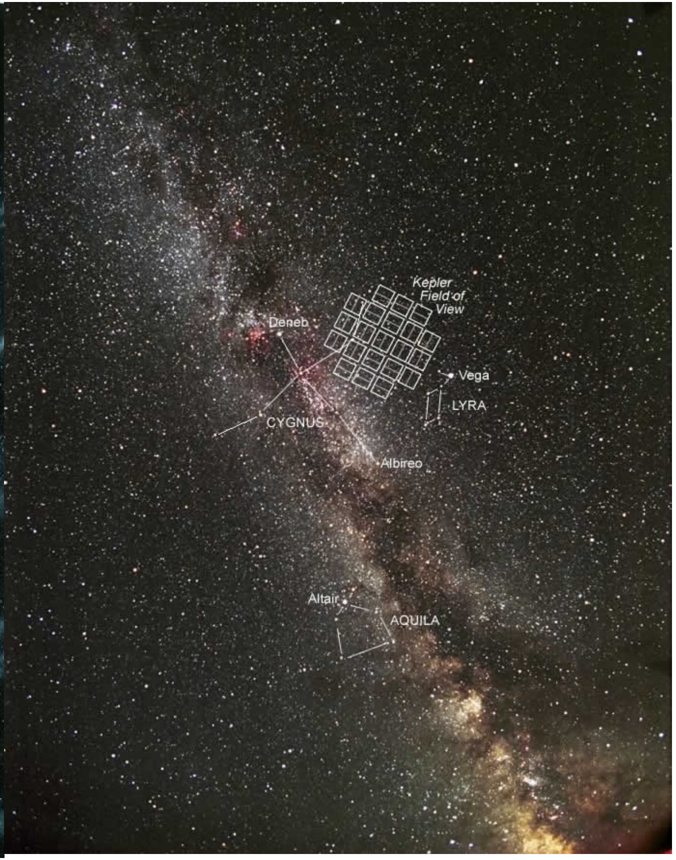
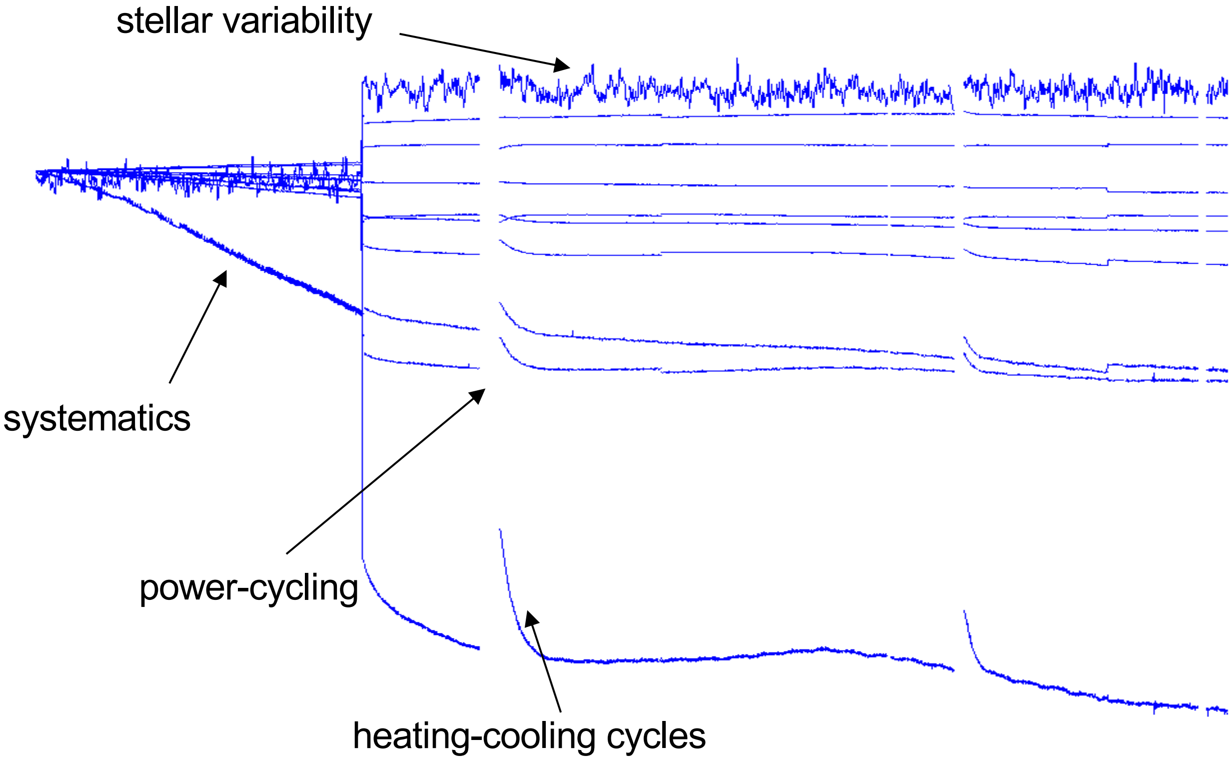


Figure 24: Focal plane layout, labeling modules and outputs (1-4), and the directions of rows and columns. Note that the focal plane is symmetric under 90 degree rotations, with the exception of the central module, module 13. Modules 1, 5, 21, and 25 are FGS modules.

Kepler Field of View



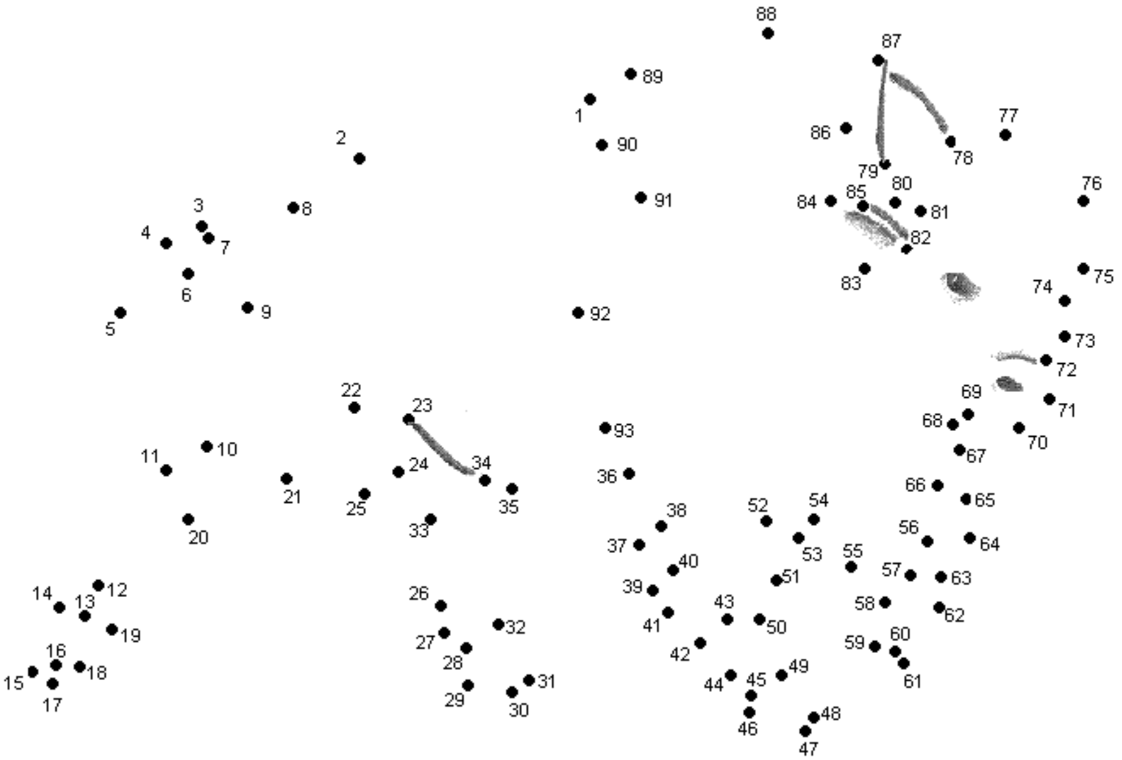
Kepler data


stellar variability
systematics
power-cycling
heating-cooling cycles

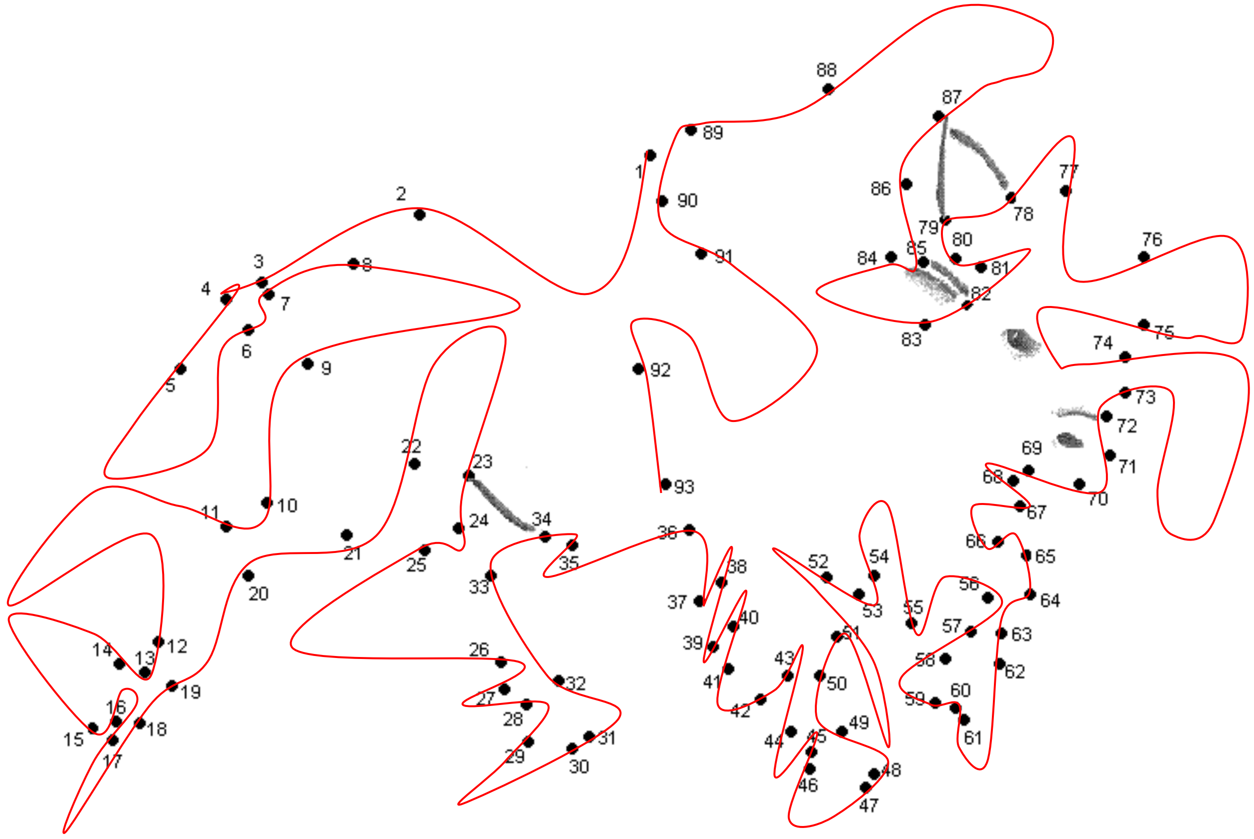
How can we solve all this?

Fit functions to sparse, noisy data!

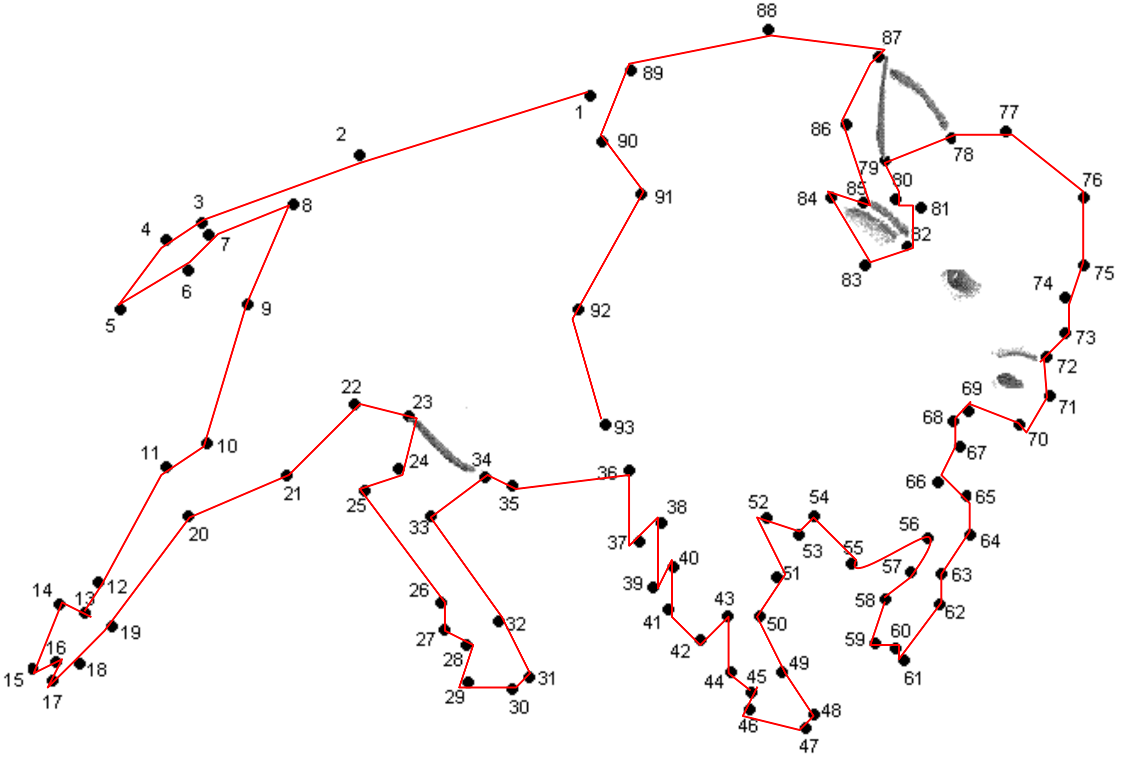
Fitting functions: a dot-to-dot is an inference problem.



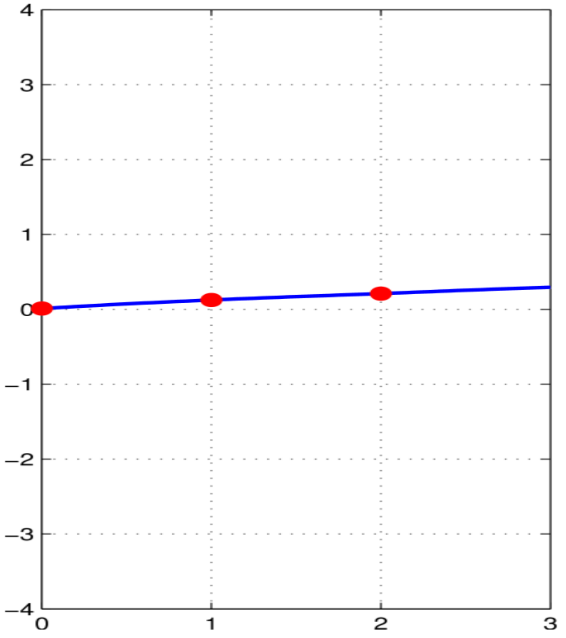
with many different solutions...

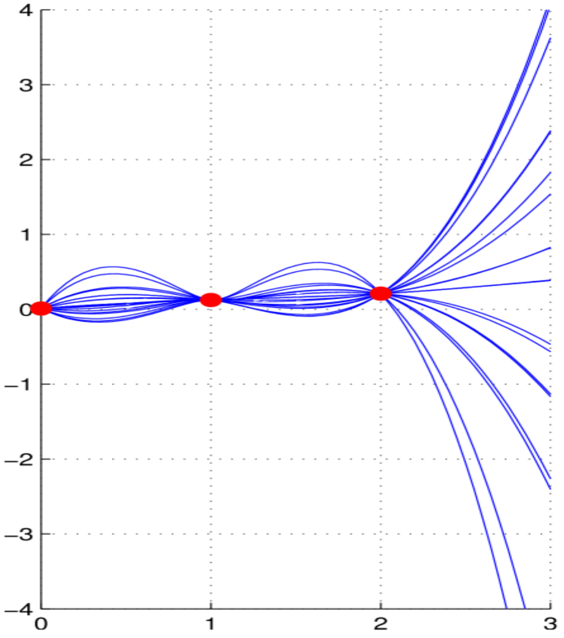


...prior information allows us to discriminate between solutions



The right model?



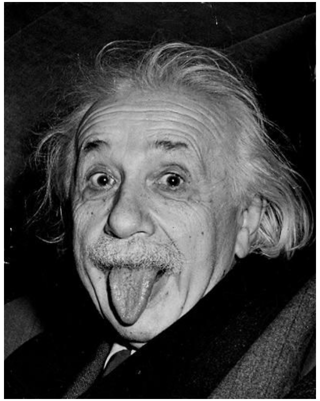


All these models explain the data equally well...

Occam's Razor

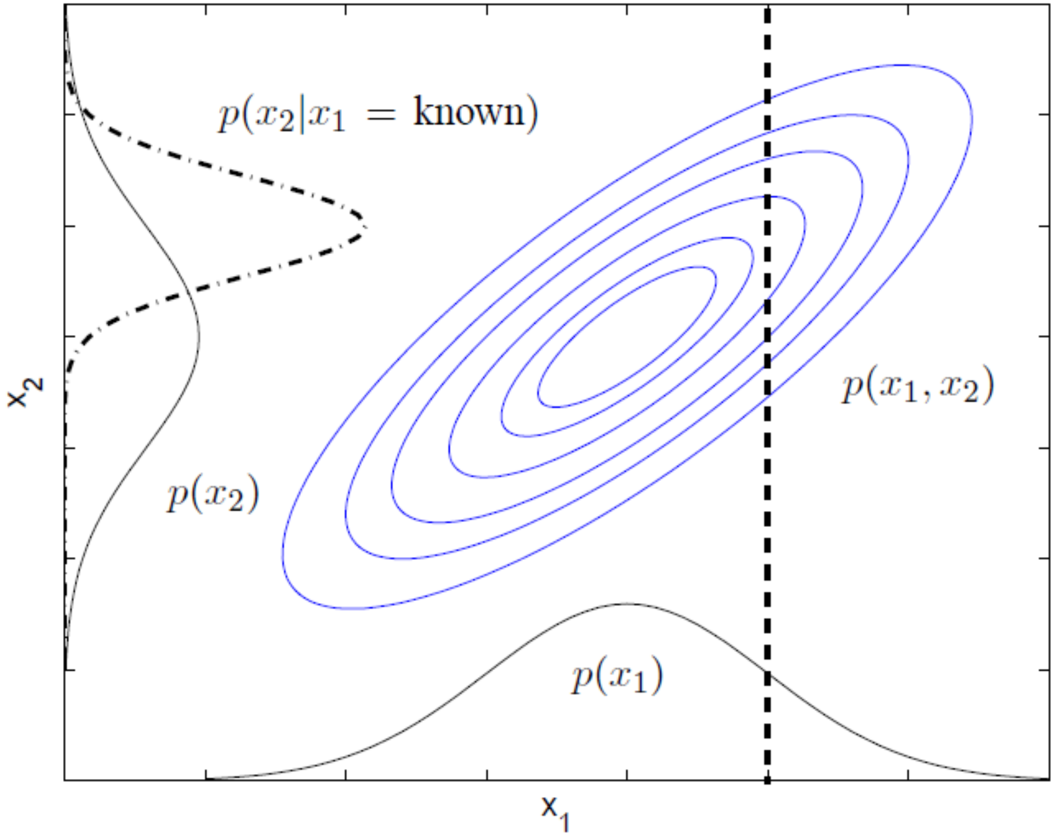
- *Numquam ponenda est pluralitas sine necessitate* - “Plurality must never be posited without necessity”

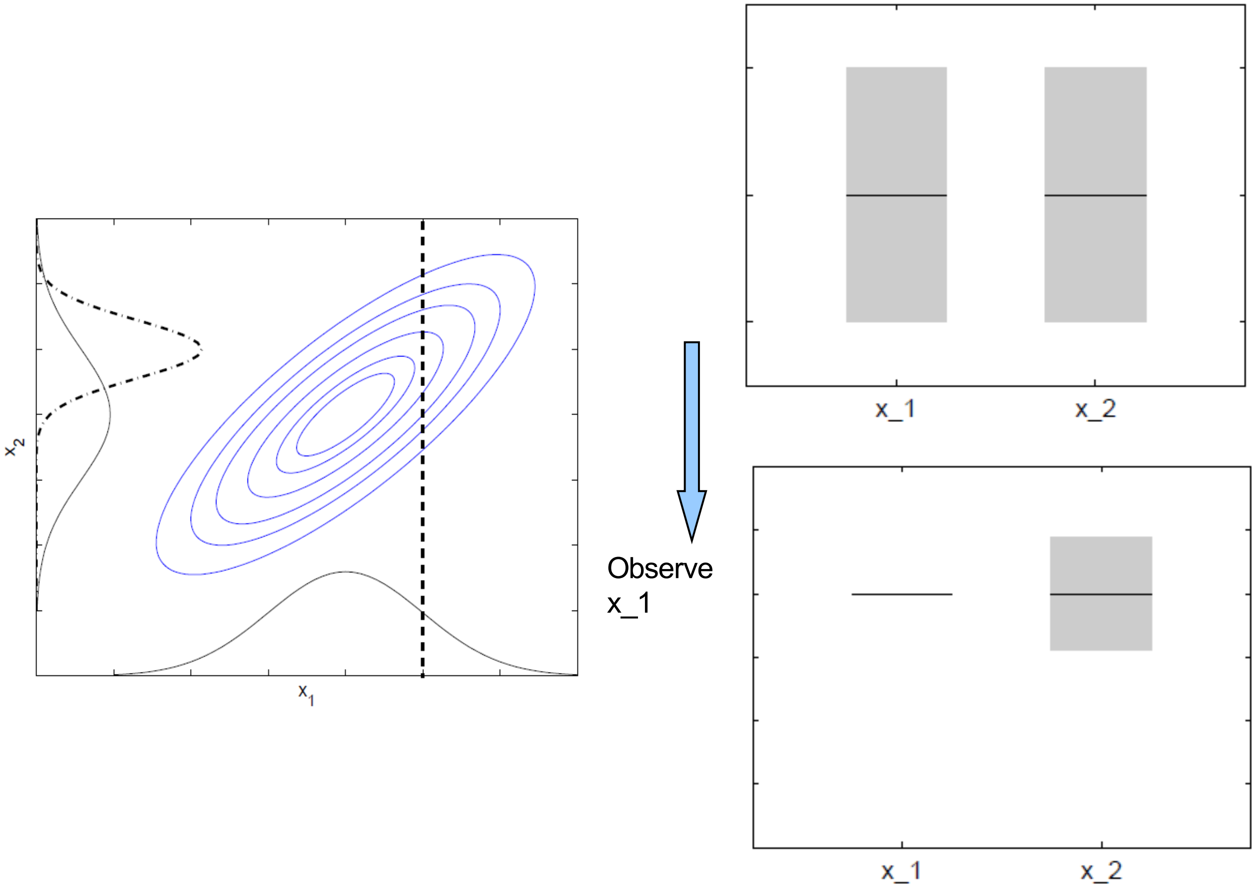
- "Everything should be kept as simple as possible, but no simpler."



Gaussian Processes

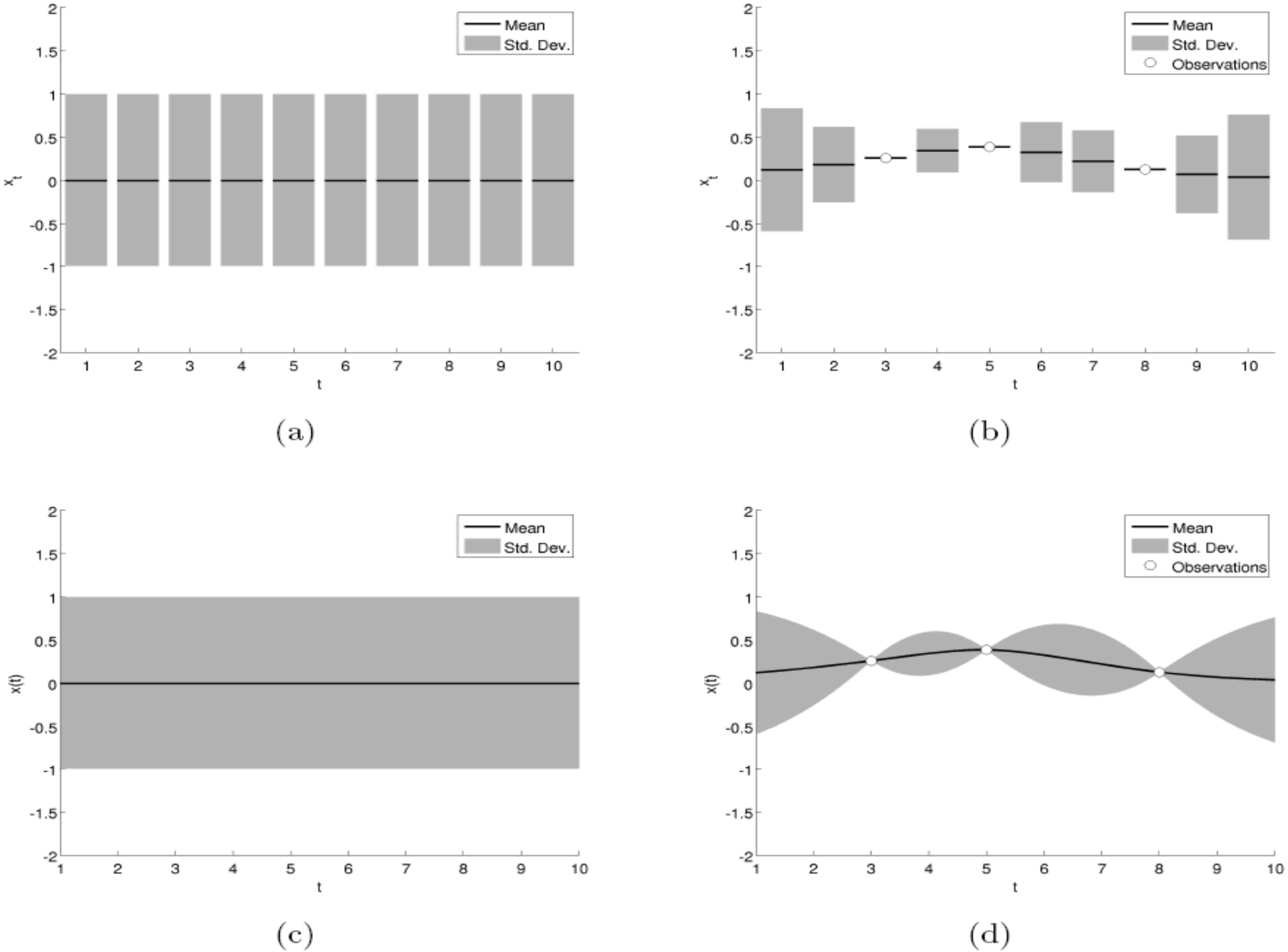
The humble (but useful) Gaussian



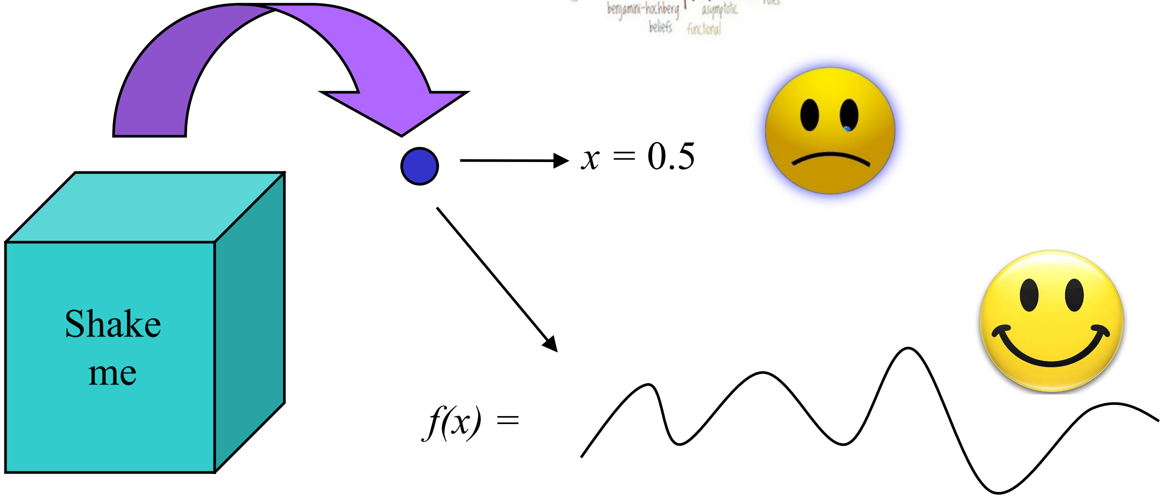


x_2
x_1
Observe
x_1
x_1
x_2
x_1
x_2

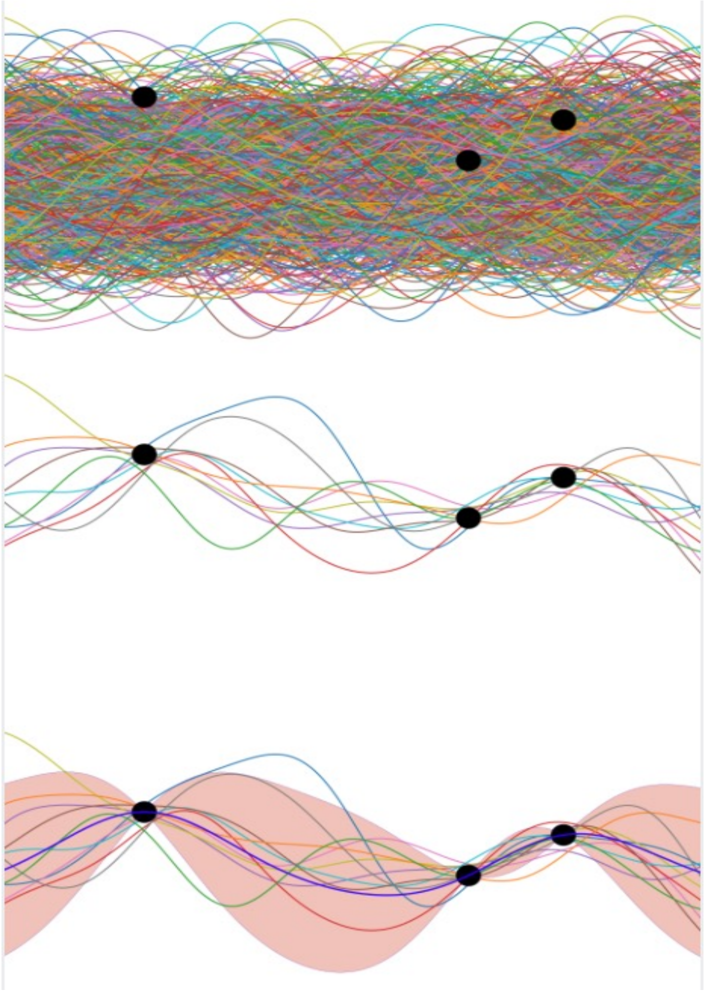
Extend to continuous variable



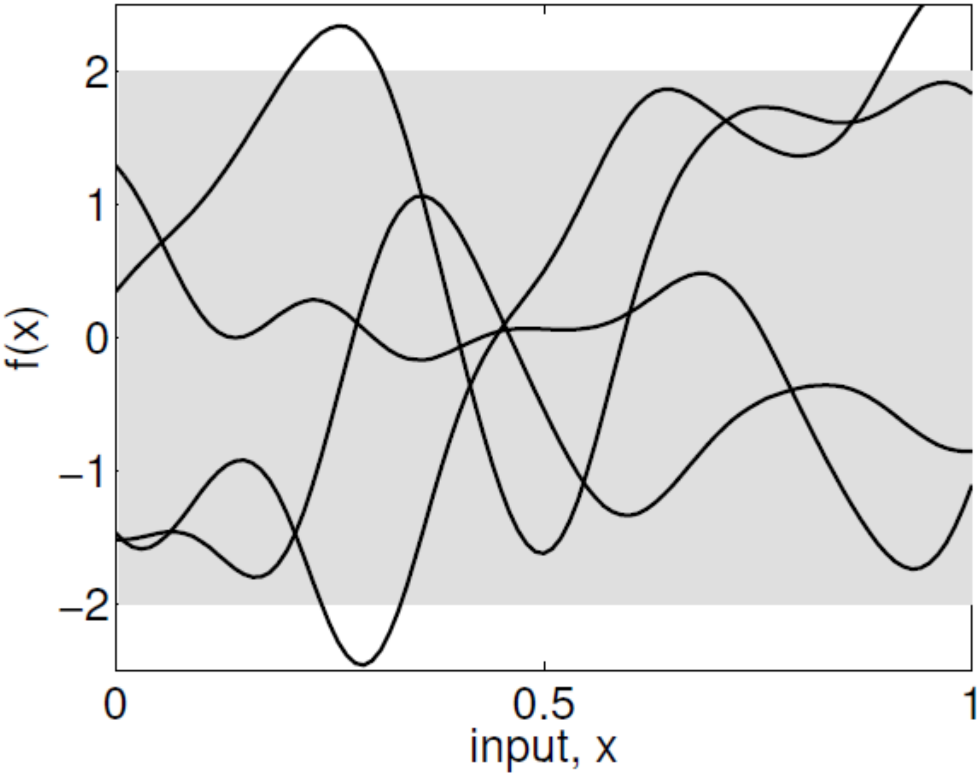
Probabilities over functions not samples



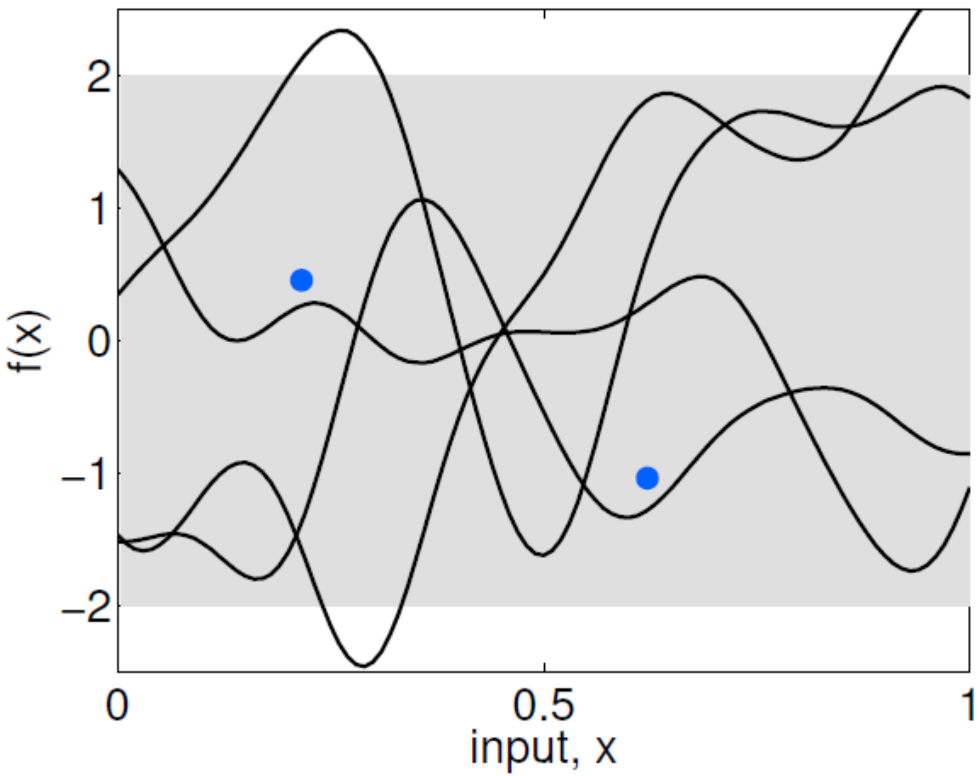
Observed data helps resolve which functions are useful



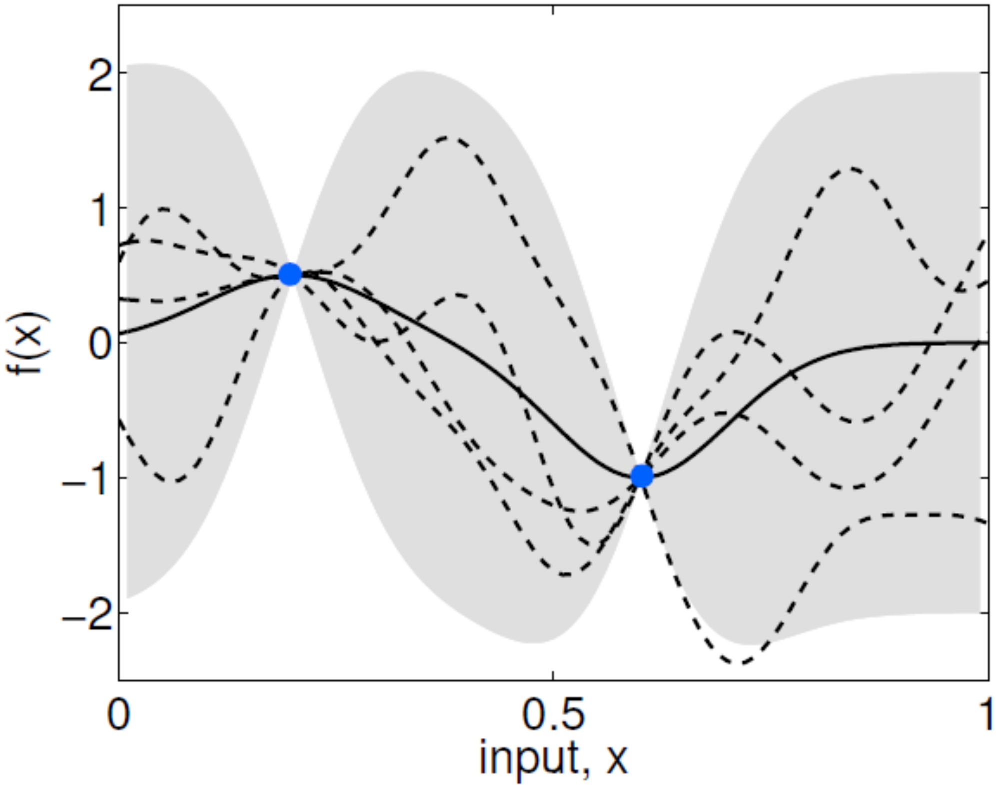
The learning process: we start with ignorance



Observe some data



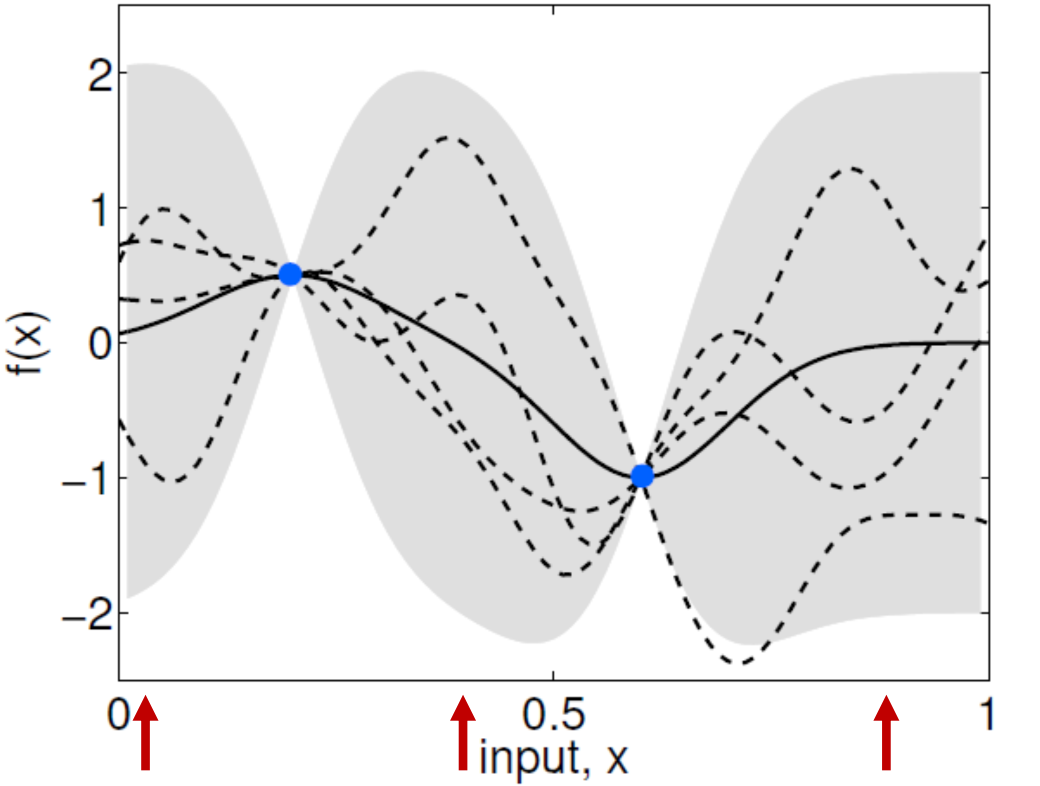
Condition on data – we do *work* to reduce *entropy*



Knowledge – here observing *data*, maximally shrinks our ignorance.

Seeing subsequent close by data provides less knowledge

What would I like to know next?



So these are good places to gather knowledge

Knowledge – here observing *data*, maximally shrinks our ignorance.

Seeing subsequent close by data provides less knowledge

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**

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

$$p(\mathbf{y}_*) = \mathcal{N}(\mathbf{m}_*, \mathbf{C}_*) \qquad \begin{aligned} m_* &= \mu(x_*) + \mathbf{K}(x_*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}(\mathbf{y} - \boldsymbol{\mu}(\mathbf{x})), \\ \sigma_*^2 &= K(x_*, x_*) - \mathbf{K}(x_*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{K}(\mathbf{x}, x_*). \end{aligned}$$

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? (This is a decision, which we can automate)

An example

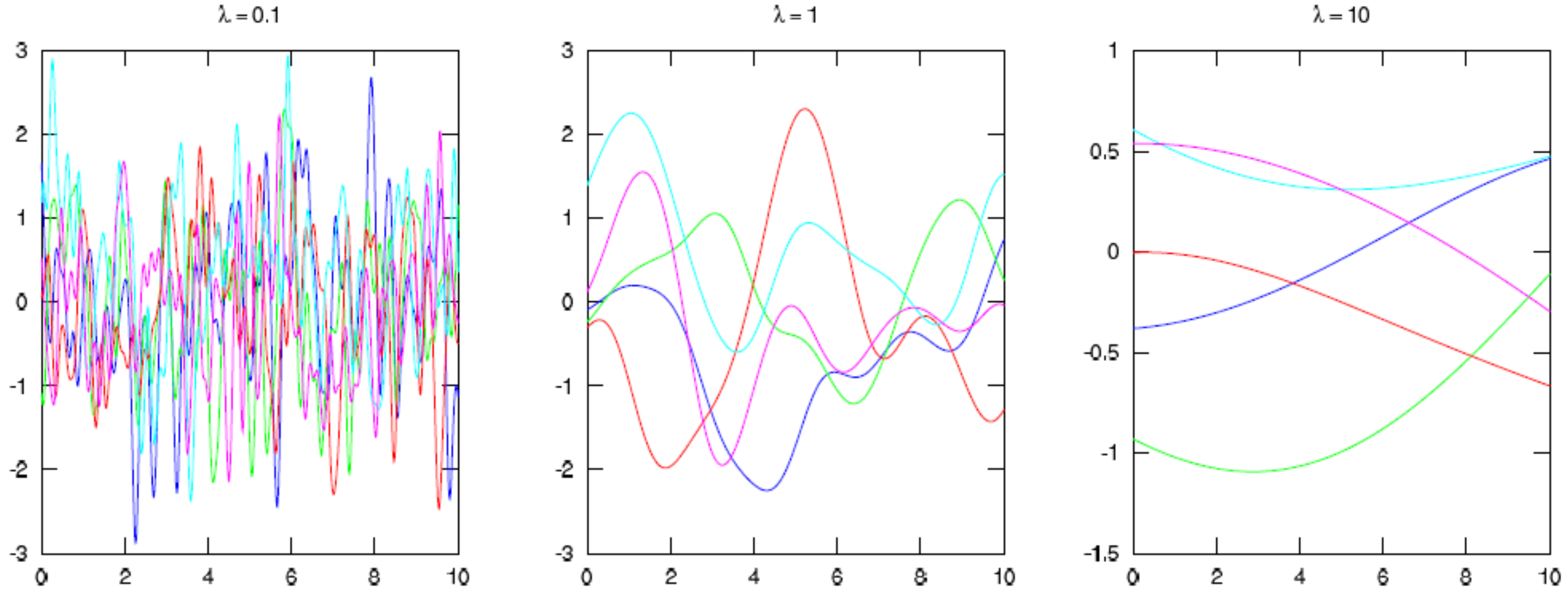
$$k(x_i, x_j) = h^2 \exp\left[-\left(\frac{x_i - x_j}{\lambda}\right)^2\right]$$

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

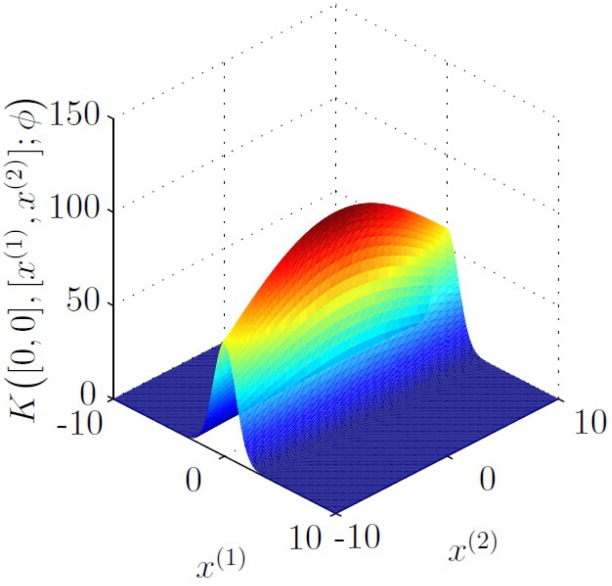
In a the Bayesian setting we work under, these scales, along with any noise process statistics, are *hyper-parameters* of the GP

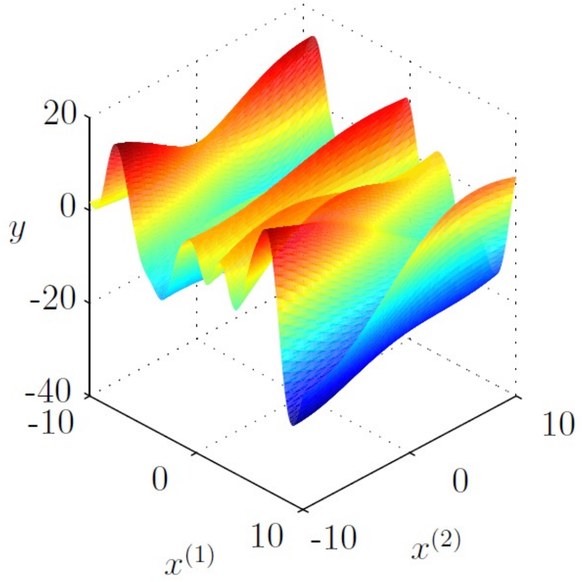
Differing length scales



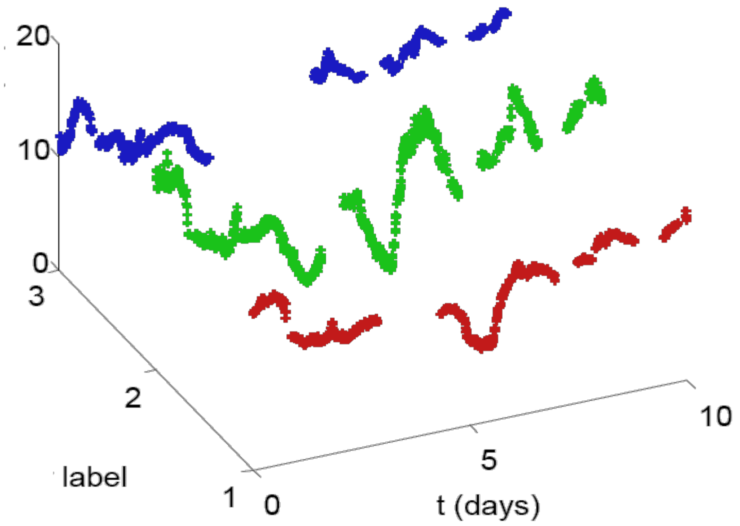
We commonly possess prior expectations that the function should be smooth. If we know something of the dynamics then this can inform our covariance functions accordingly

We can modify covariance functions to accommodate multiple input dimensions





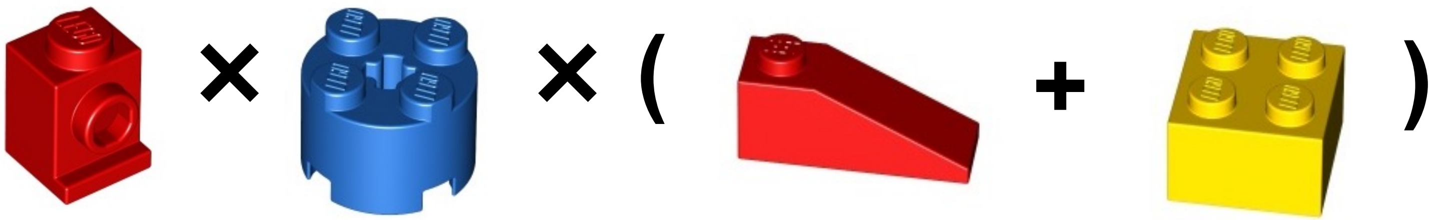
Multiple outputs, reframe the problem as having a single output, and an additional *label* input specifying the output.



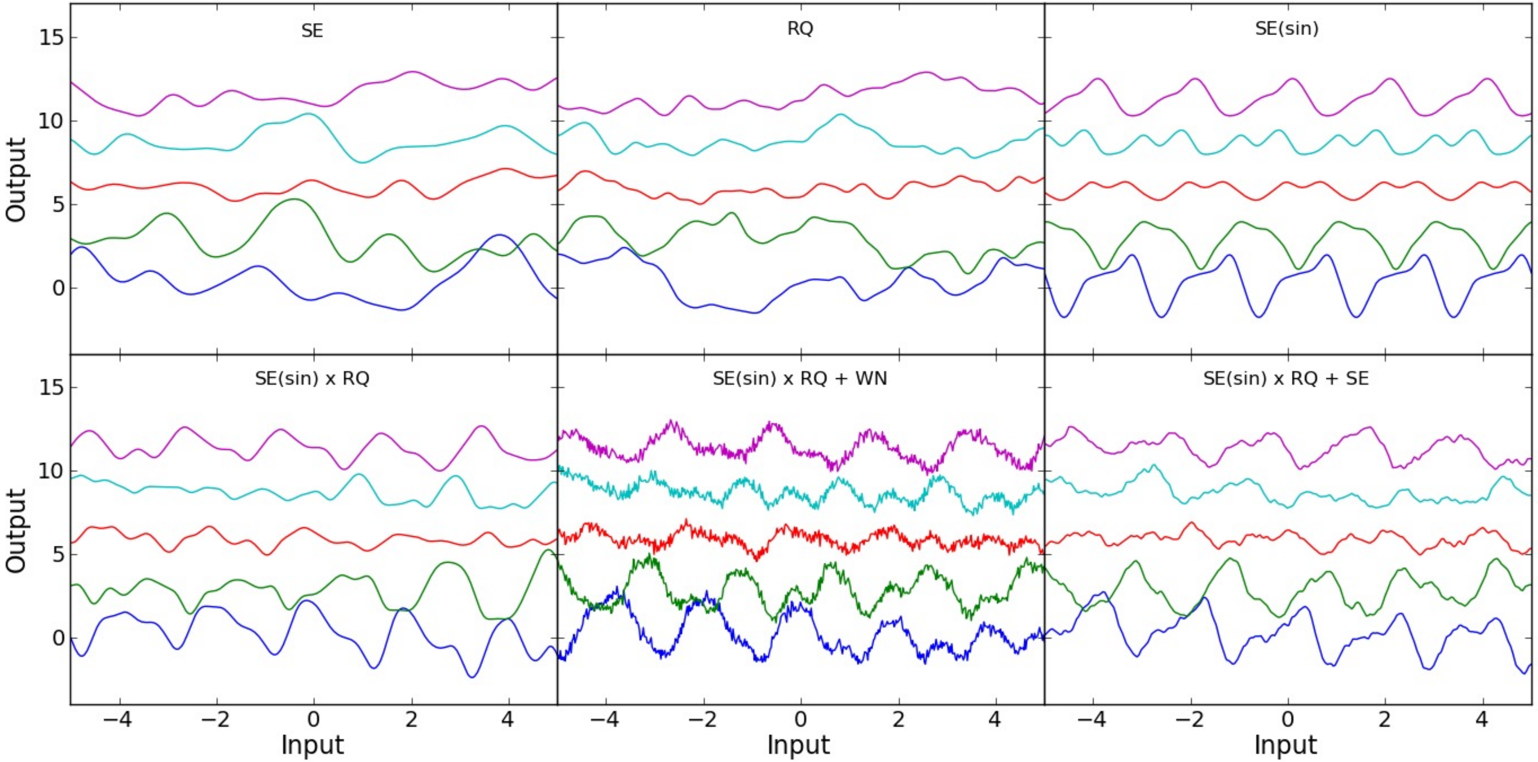
Hence we do not need simultaneous observations of all outputs.

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

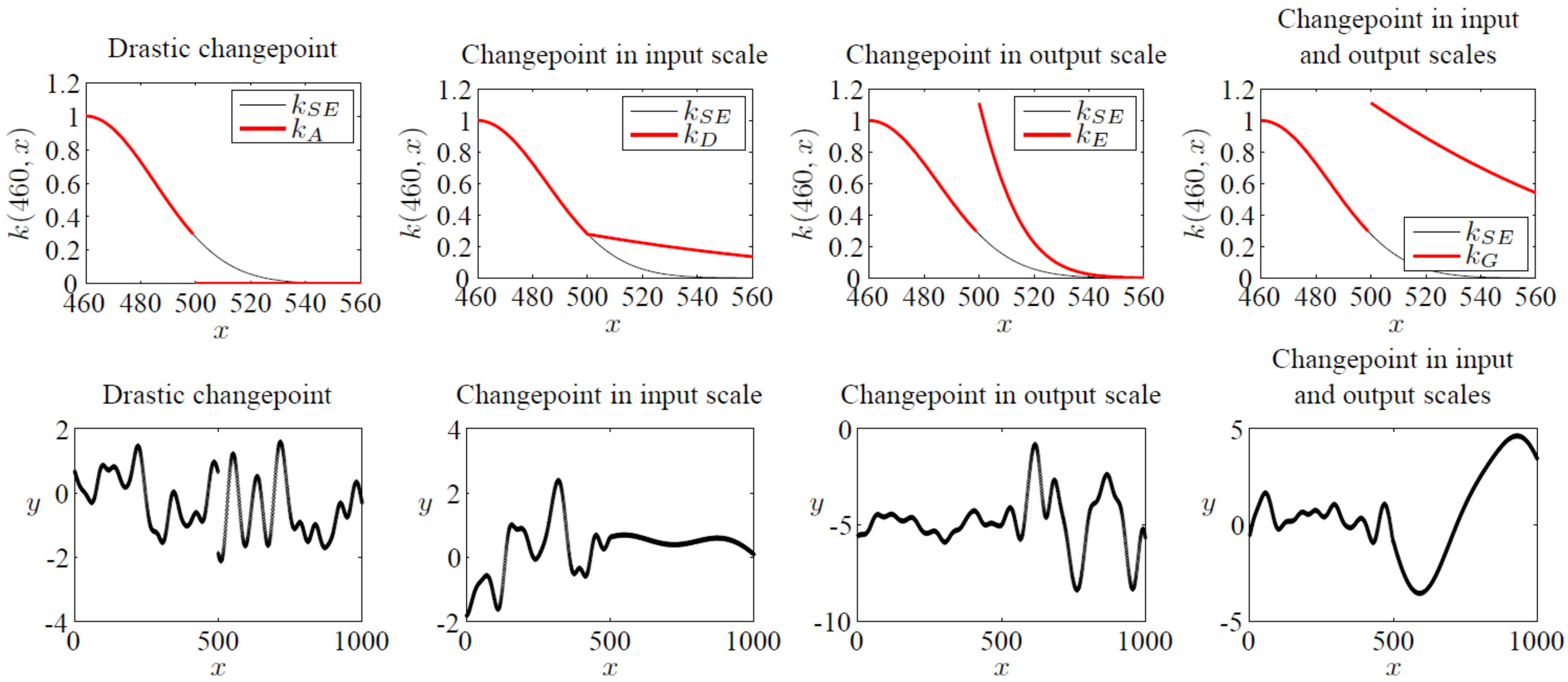
e.g.



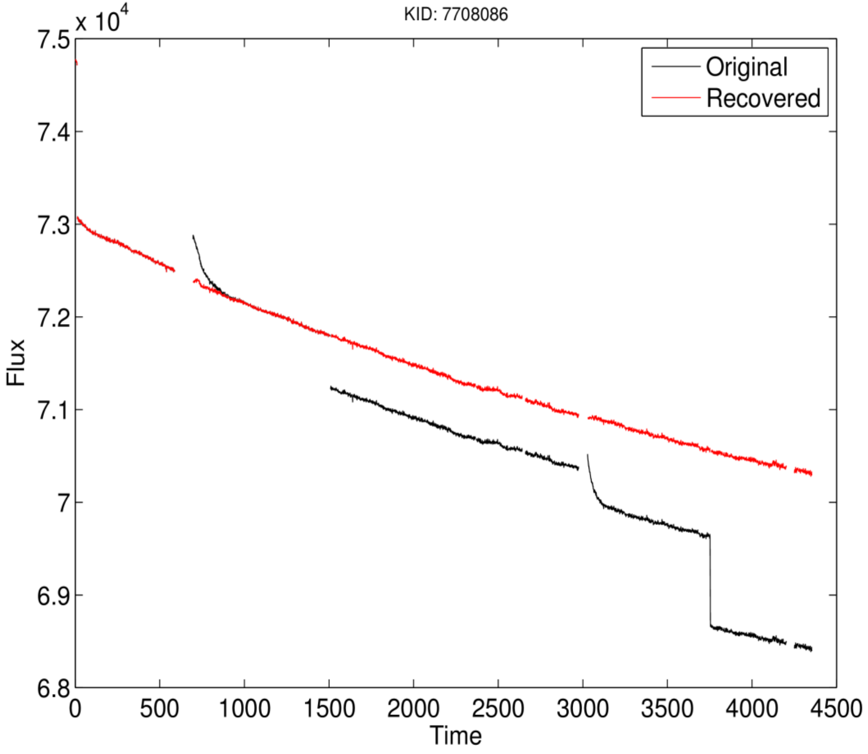
Kernel functions

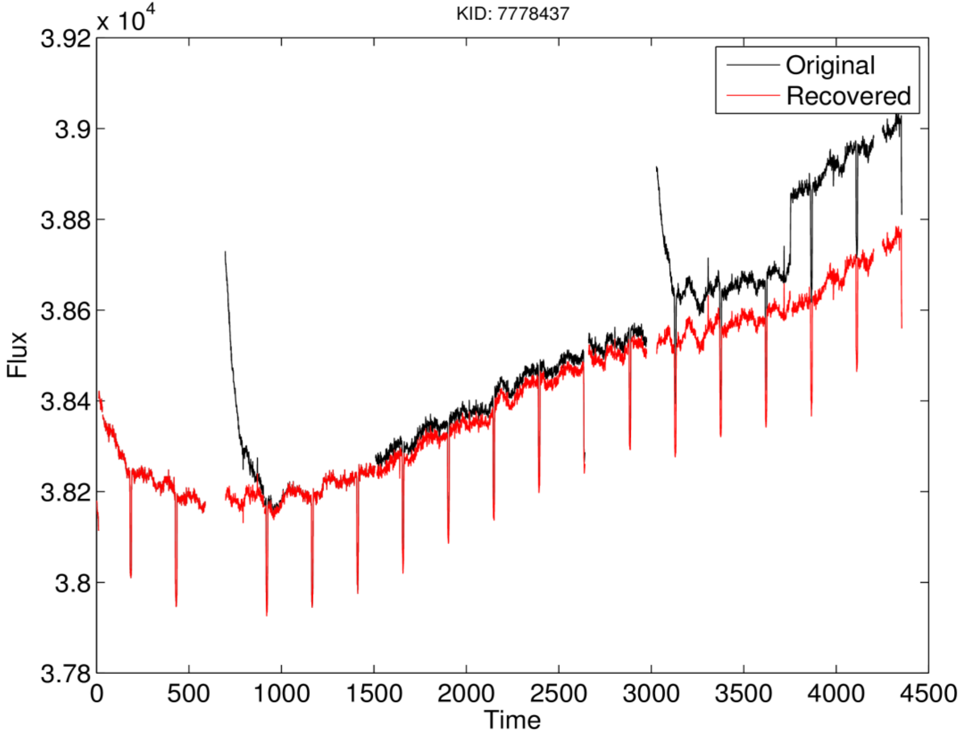


Covariances for e.g. changepoints, faults and sets.

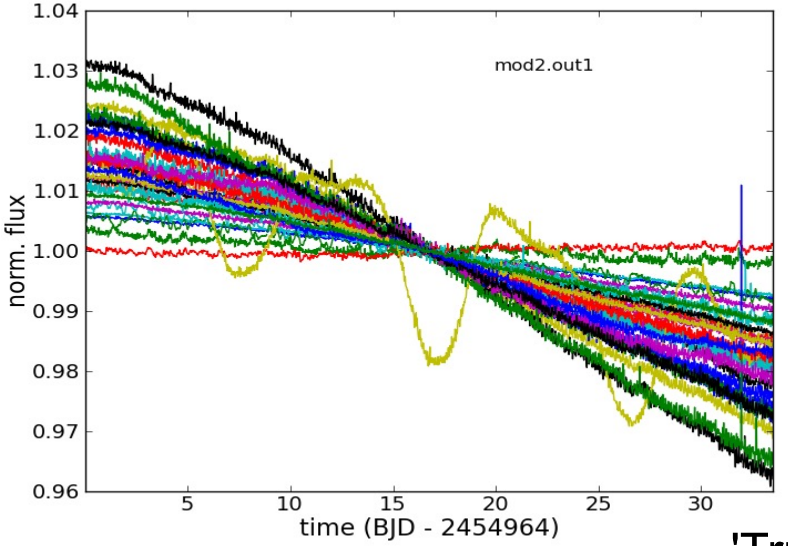


Gaussian process: jumps and coolings





Systematics



Latent trend discovery

>300,000 light curves

Variational Bayes for hyper-parameters

Entropic prior & shrinkage

'True' curves

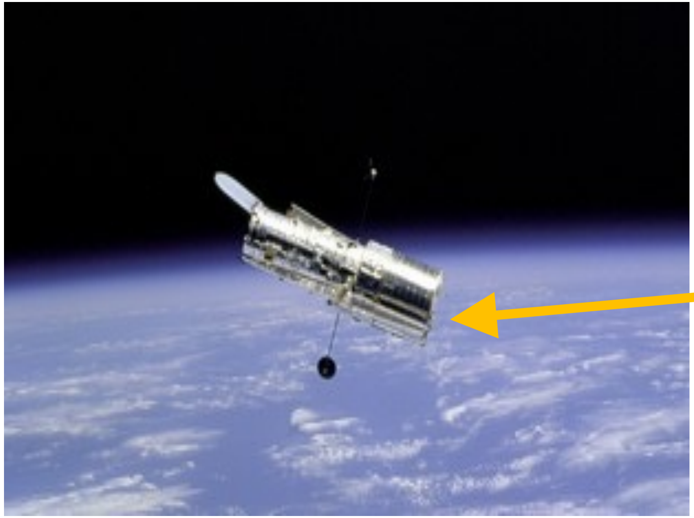
Systematic trends

$$d_i = s_i + \sum_{j=1}^{J} a_{ij} u_j + \epsilon_i$$

Observed curves

Noise component

Kepler space telescope



(2009 – 2014, then...)

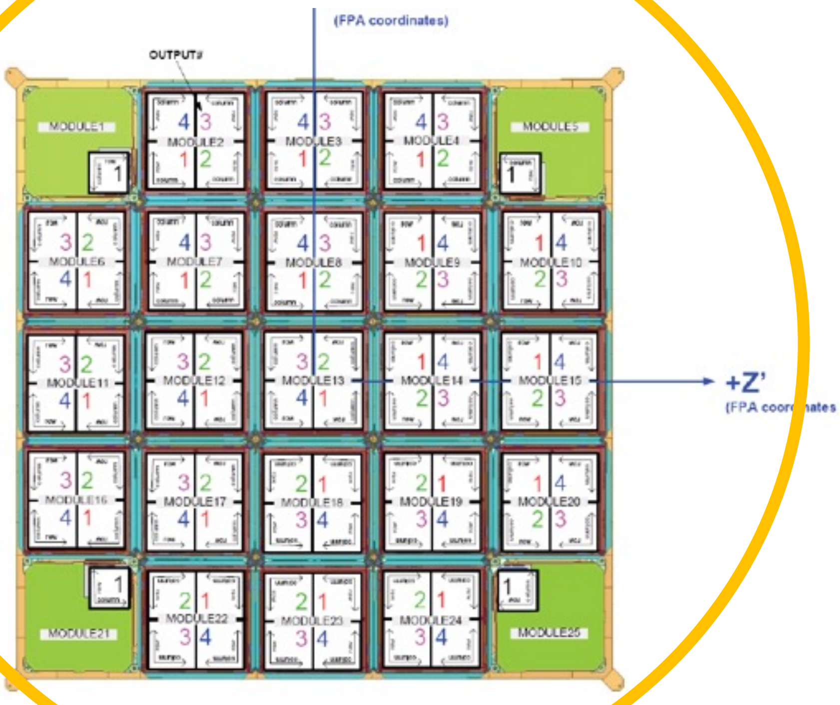
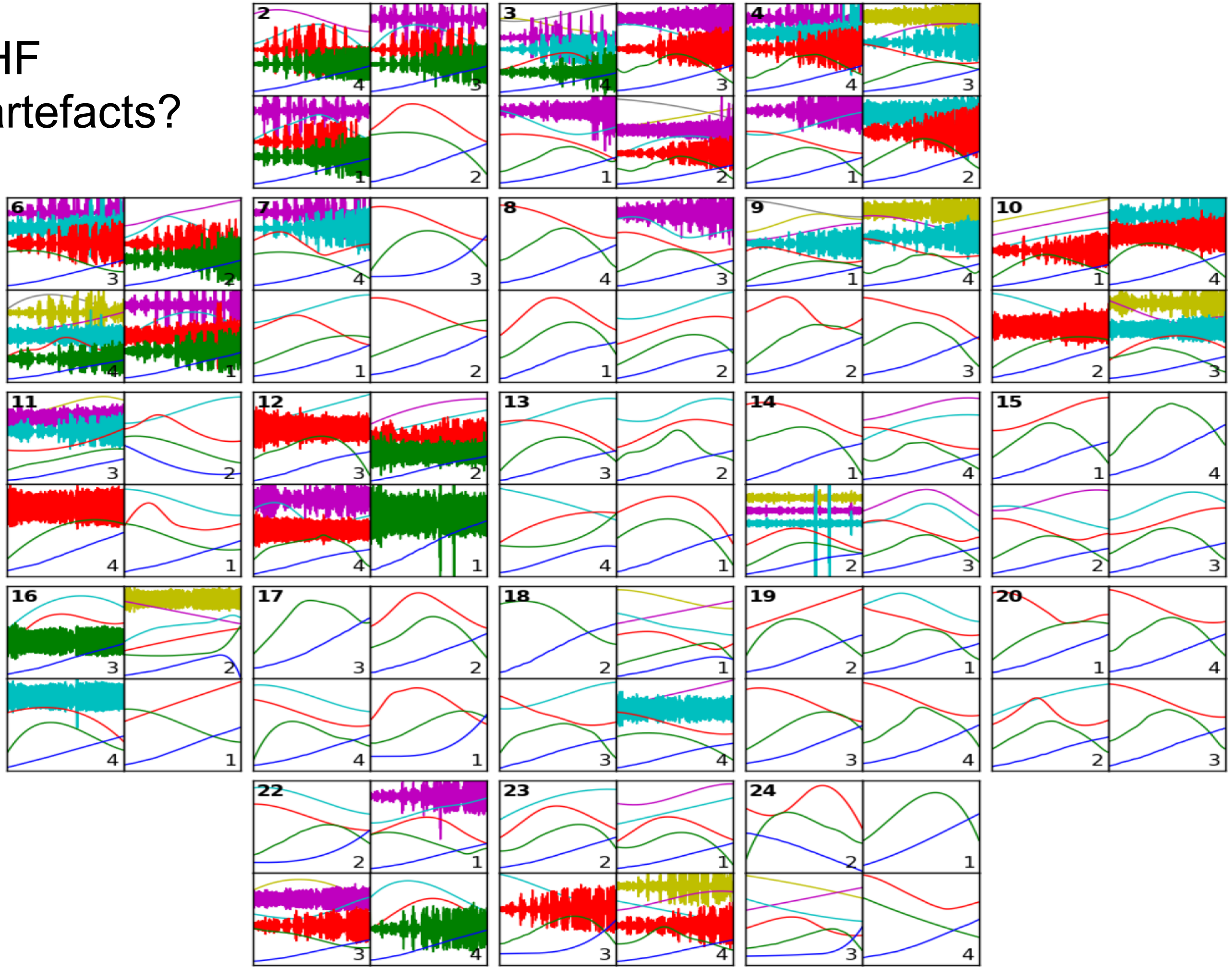
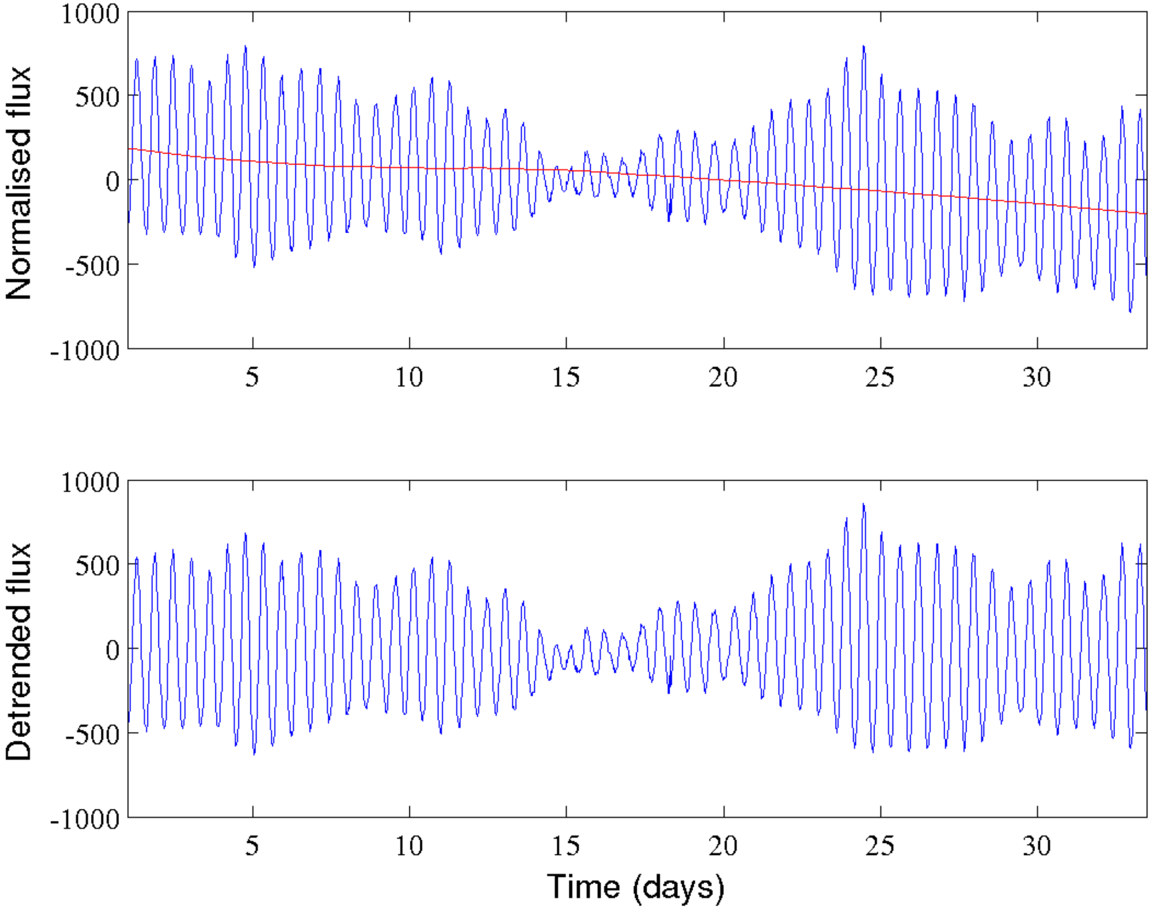


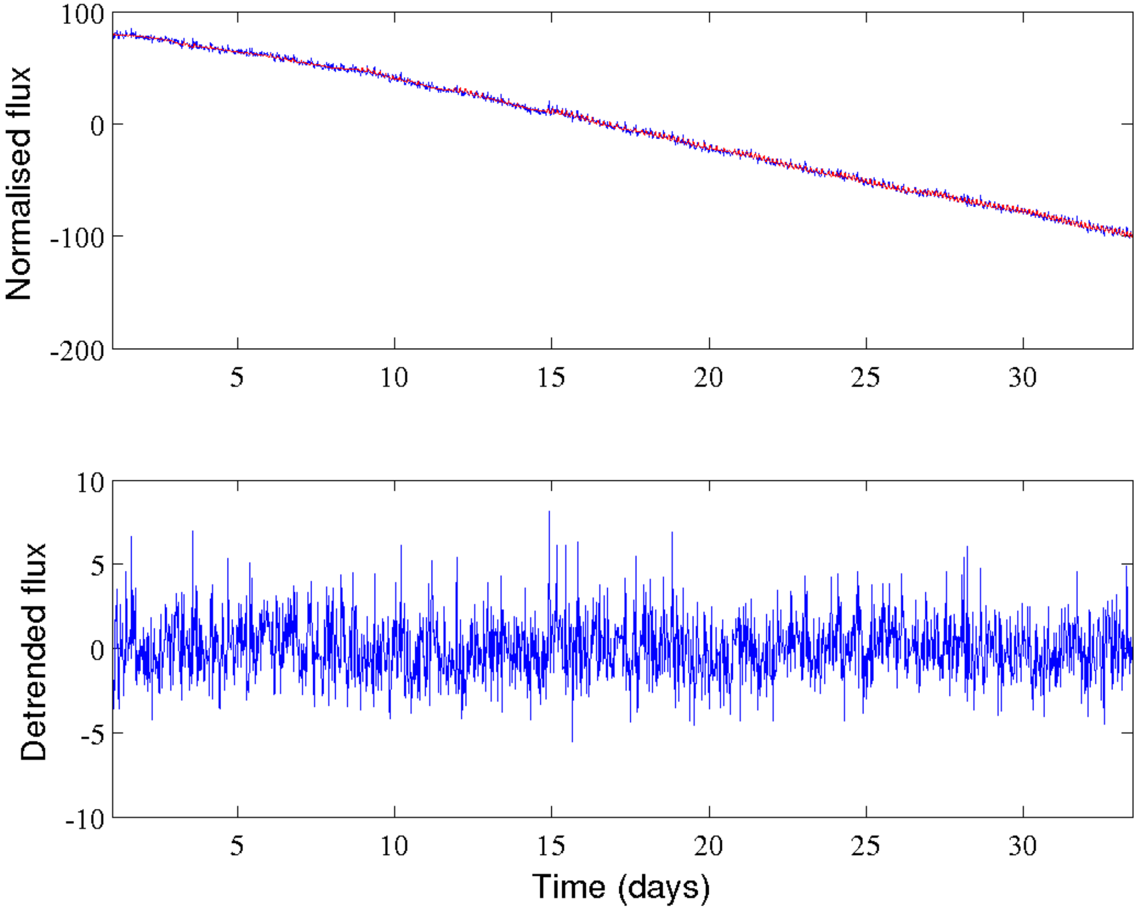
Figure 24: Fo... plane layout, labeling modules and outputs (1-4) ... the directions of rows and columns. Note tha... focal plane is symmetric under 90 de... rotations, with the exception of the central module, modu... Modules 1, 5, 21, and ... are FGS modules.

HF artefacts?

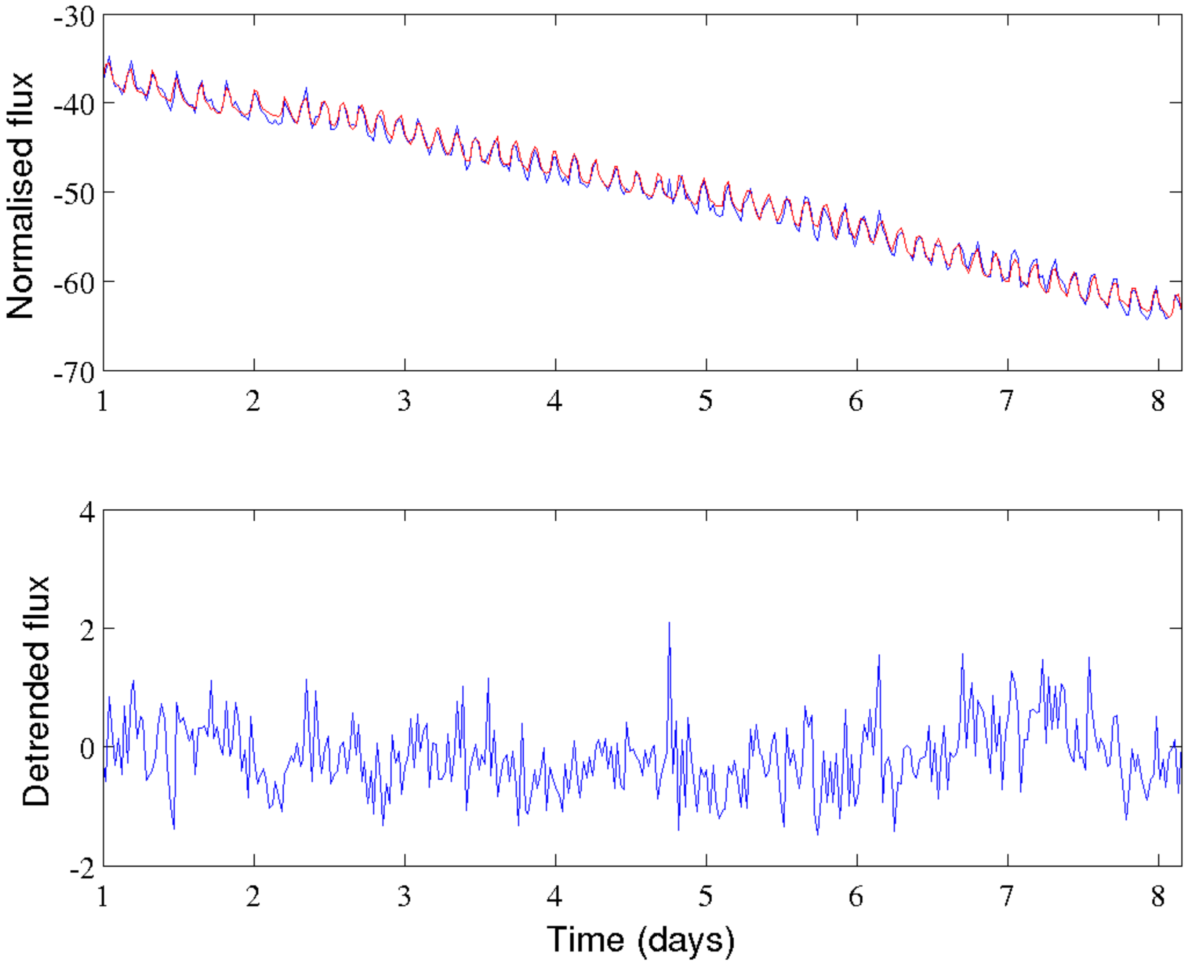




Normalised flux
1000
500
0
-500
-1000
5
10
15
20
25
30
Detrended flux
1000
500
0
-500
-1000
5
10
15
20
25
30
Time (days)



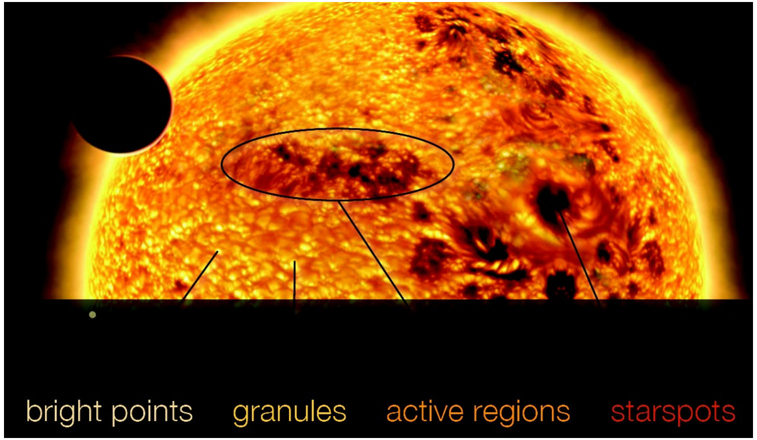
100
0
-100
-200
Normalised flux
5
10
15
20
25
30
10
5
0
-5
-10
Detrended flux
5
10
15
20
25
30
Time (days)

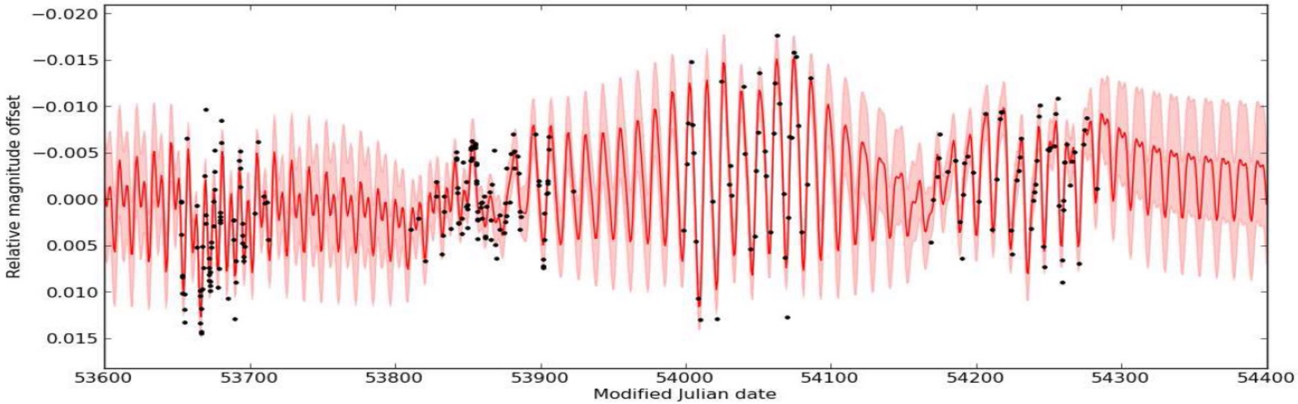


Normalised flux
-30
-40
-50
-60
-70
1
2
3
4
5
6
7
8
Detrended flux
4
2
0
-2
Time (days)

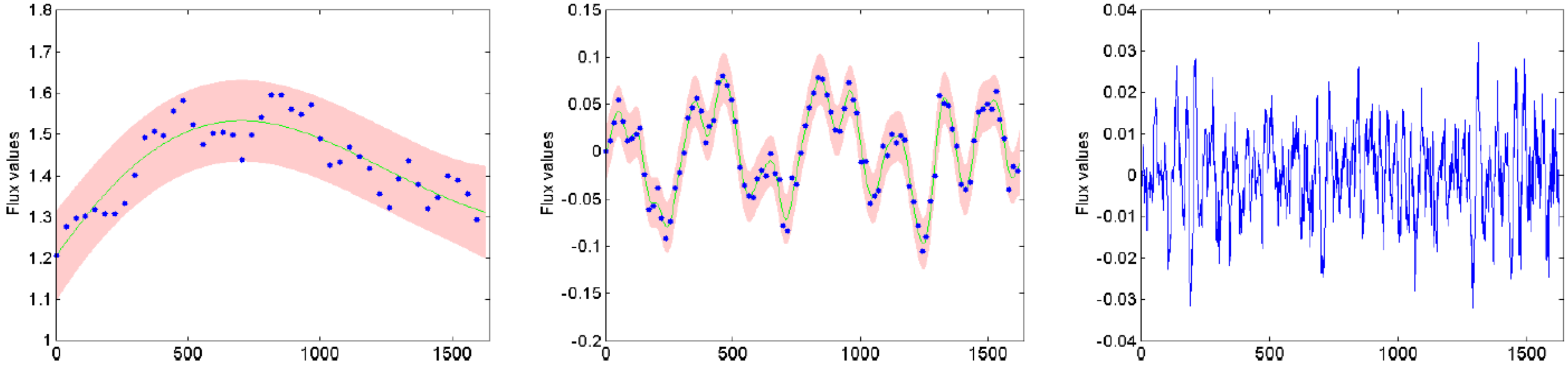
Quasi-periodic model for stellar flux

Problem is that stellar flux is highly variant... star-spots and stellar rotations... so first we need to model the quasi-periodic flux measurements





Stellar variability

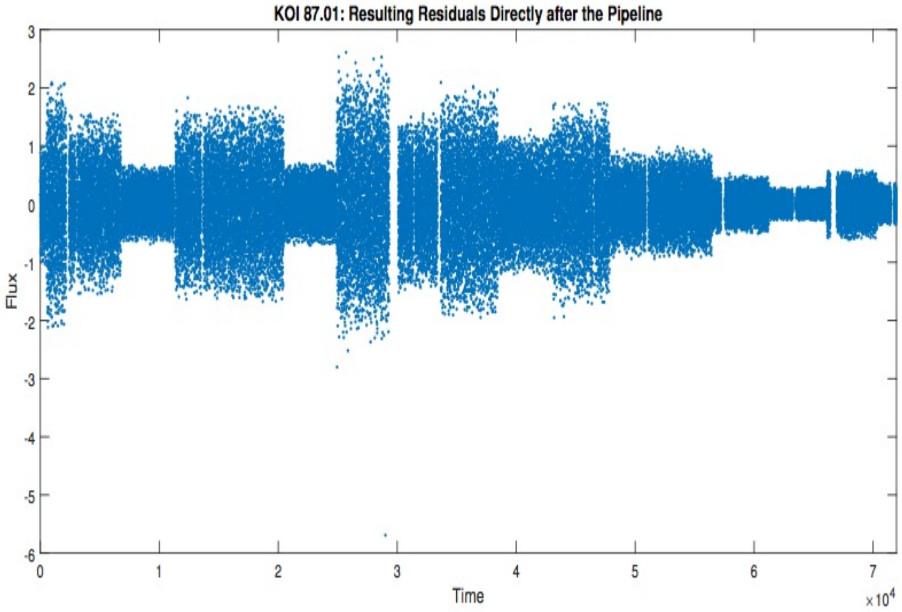


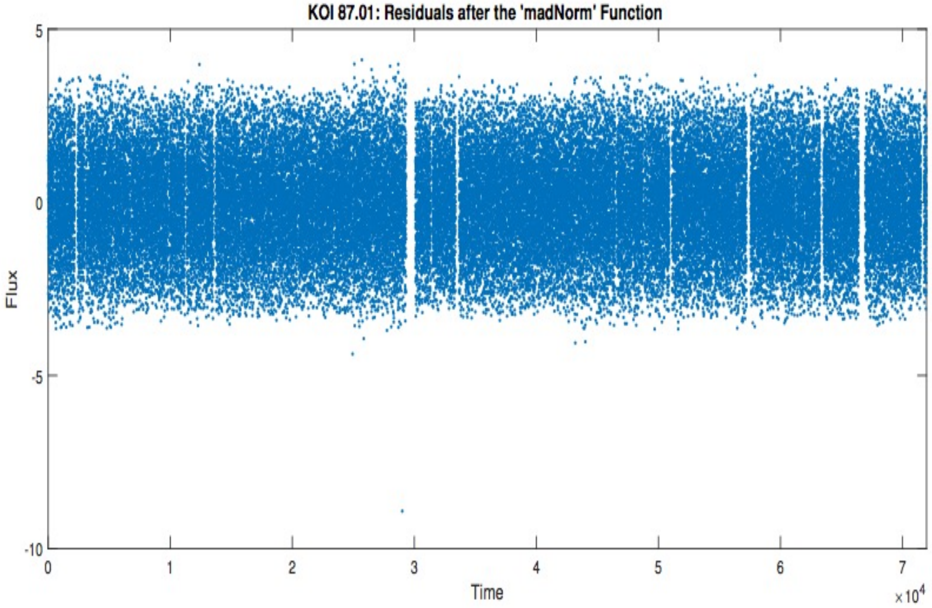
GP with sum of long term kernel and quasi-periodic kernel

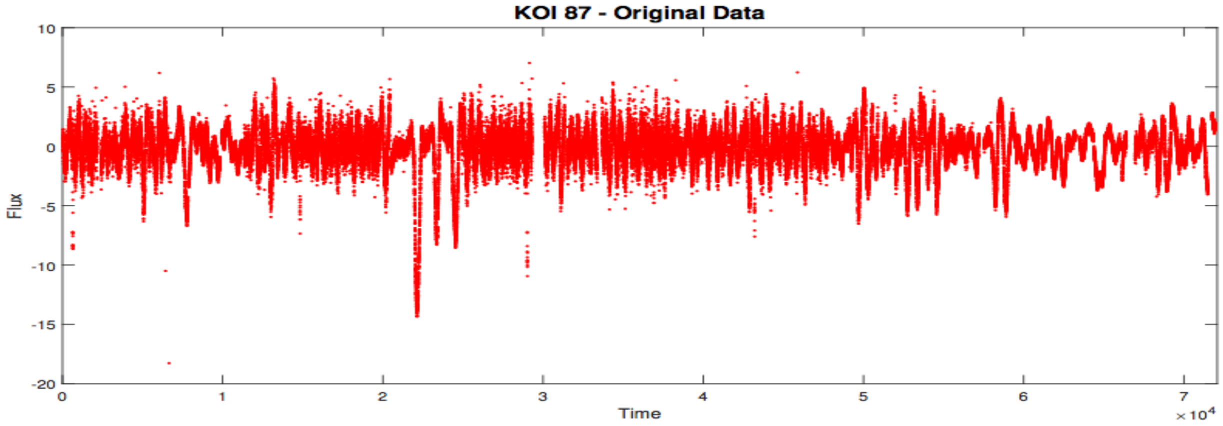
Astrophysical priors

Systematic removal of stellar variability

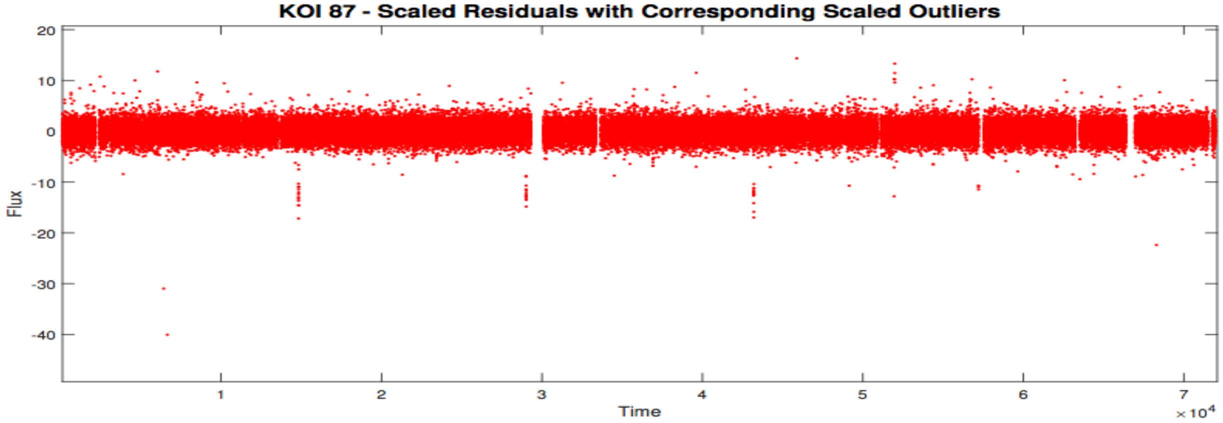
Joining observational data (multiple years)



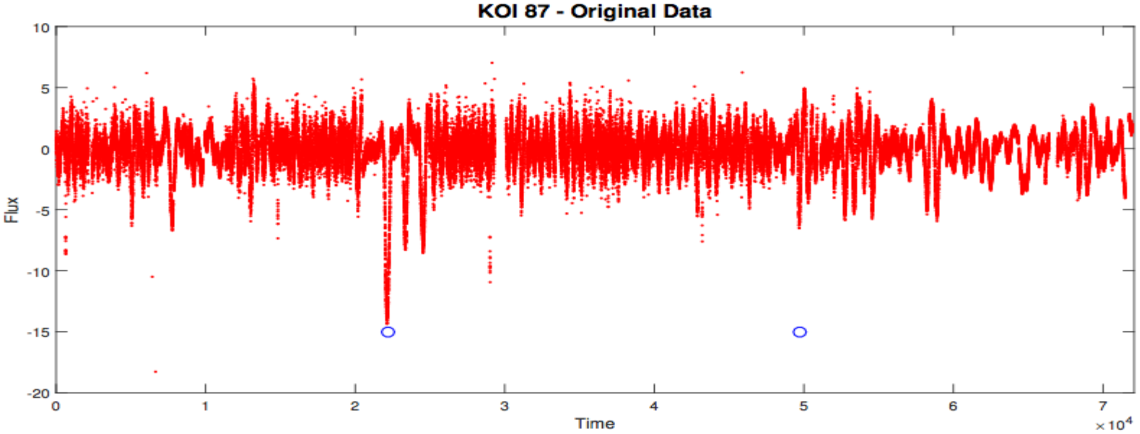




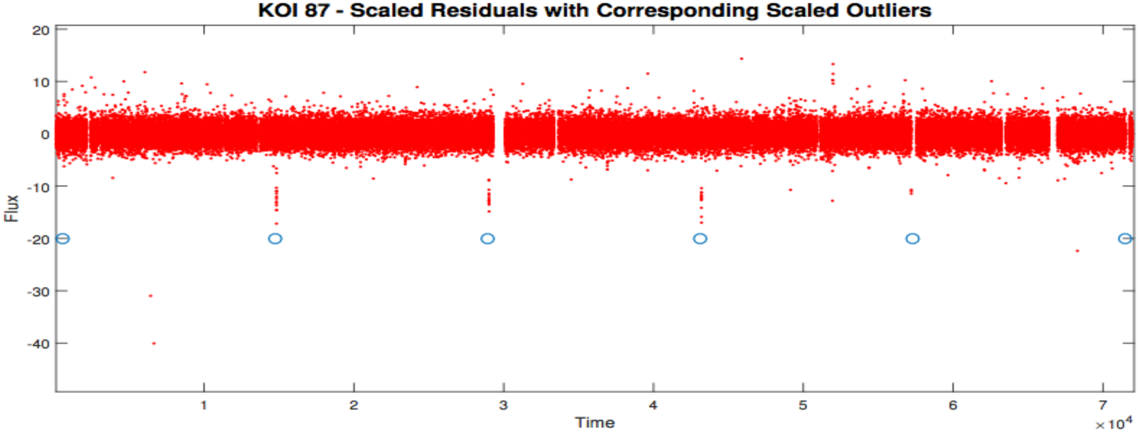
(a) Original data for KOI 87 before preprocessing in the pipeline.



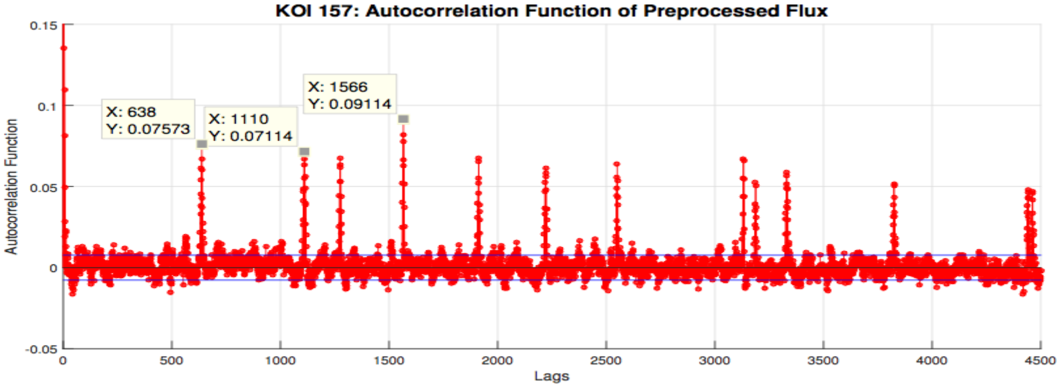
(b) The result for KOI 87 after preprocessing in the pipeline and reinserting the scaled outliers.



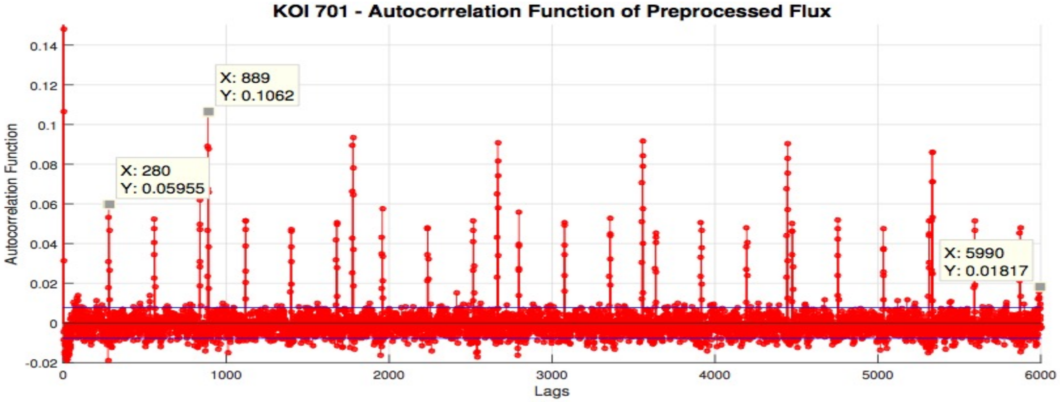
(a) Original data for KOI 87 before preprocessing in the pipeline. The blue circles correspond to the predicted transits according to the BLS algorithm.



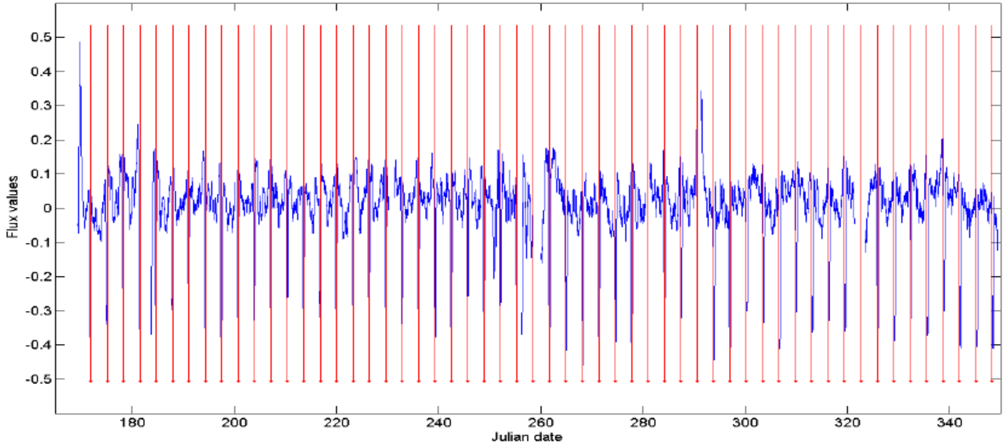
(b) The result for KOI 87 after preprocessing in the pipeline and reinserting the scaled outliers.The blue circles correspond to the predicted transits according to the BLS algorithm.



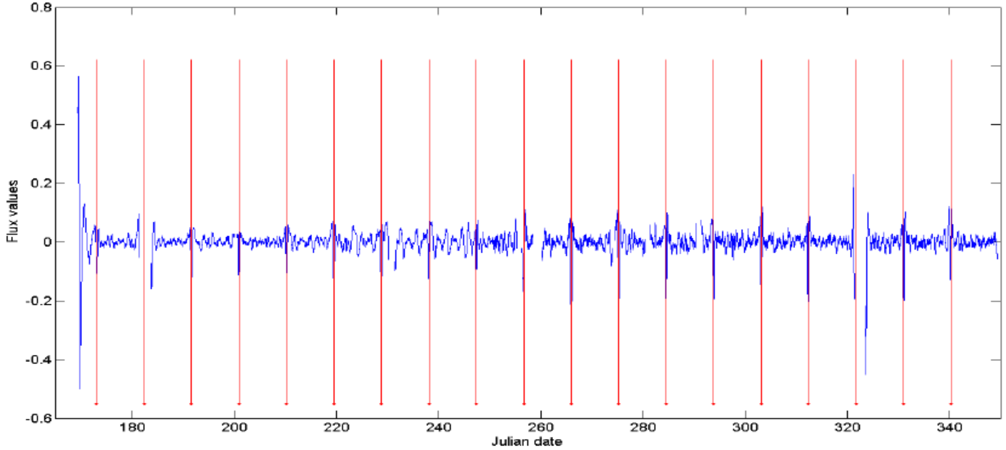
(c) Labelled peaks correspond to the planets KOI 157.01, KOI 157.02 and KOI 157.03.



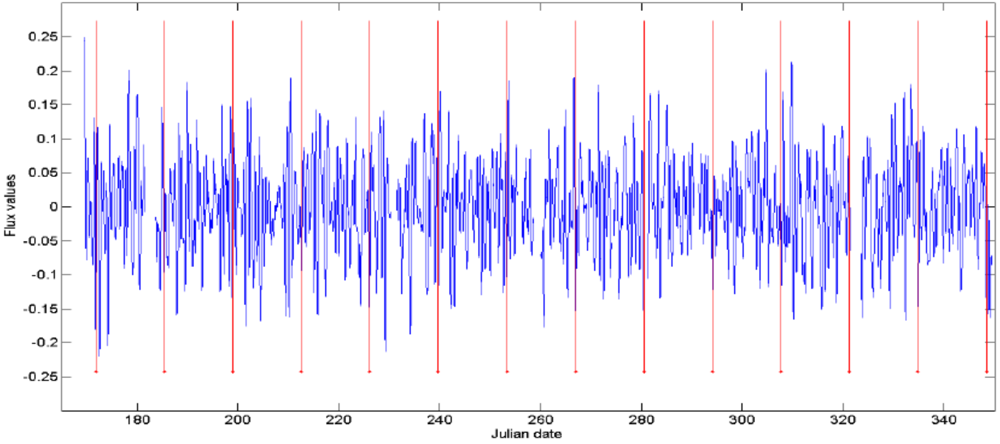
(e) Labelled peaks correspond to the planets KOI 701.01, KOI 701.02 and KOI 701.03.



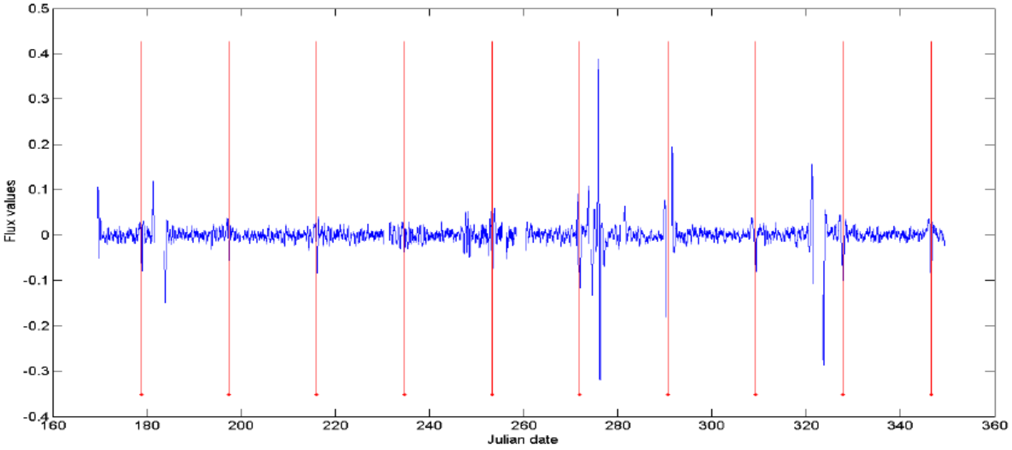
KID 11853905. Period 3.2076 days (confirmed 3.213). KS statistic 0.852. Kepler 4b



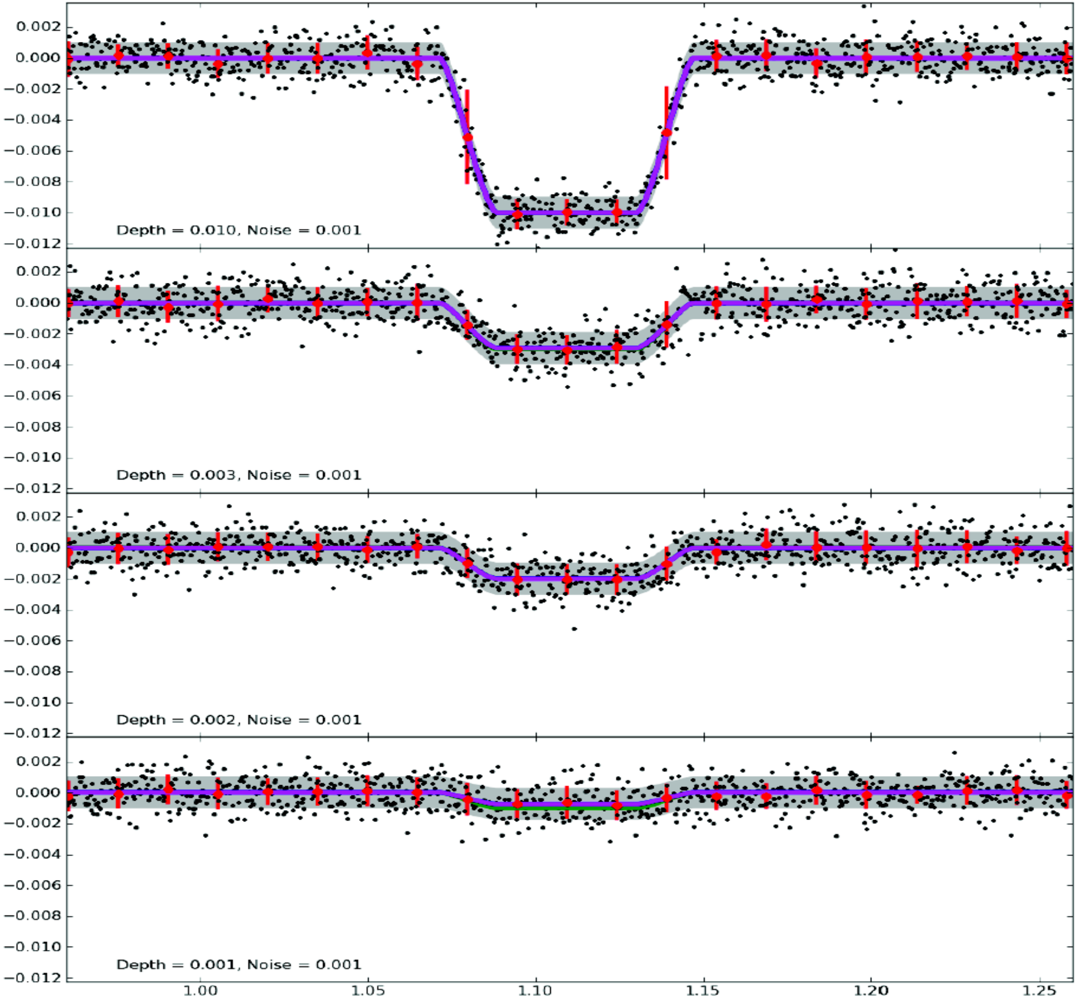
KID 2571238. Period 9.296 days (confirmed 9.287). KS statistic 0.972. Kepler 19b.



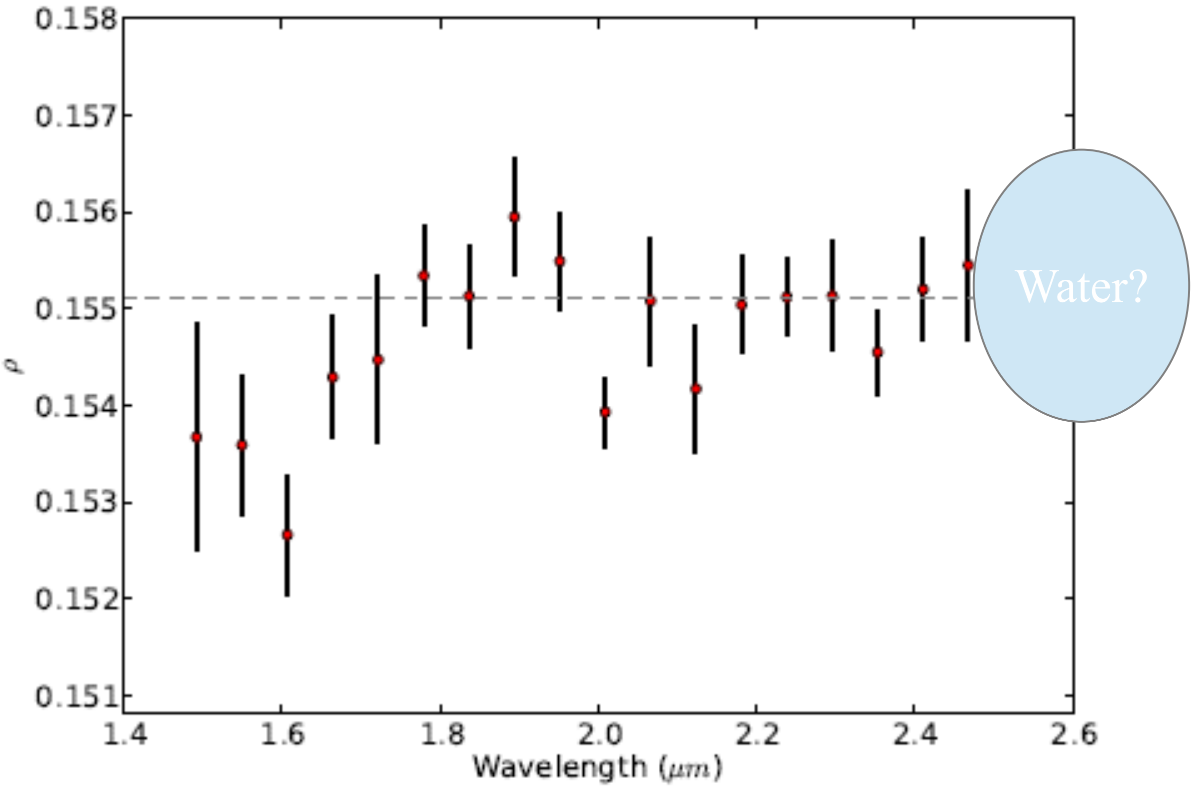
KID 1433399. Period 13.587 days. KS statistic 0.825. Potential candidate.

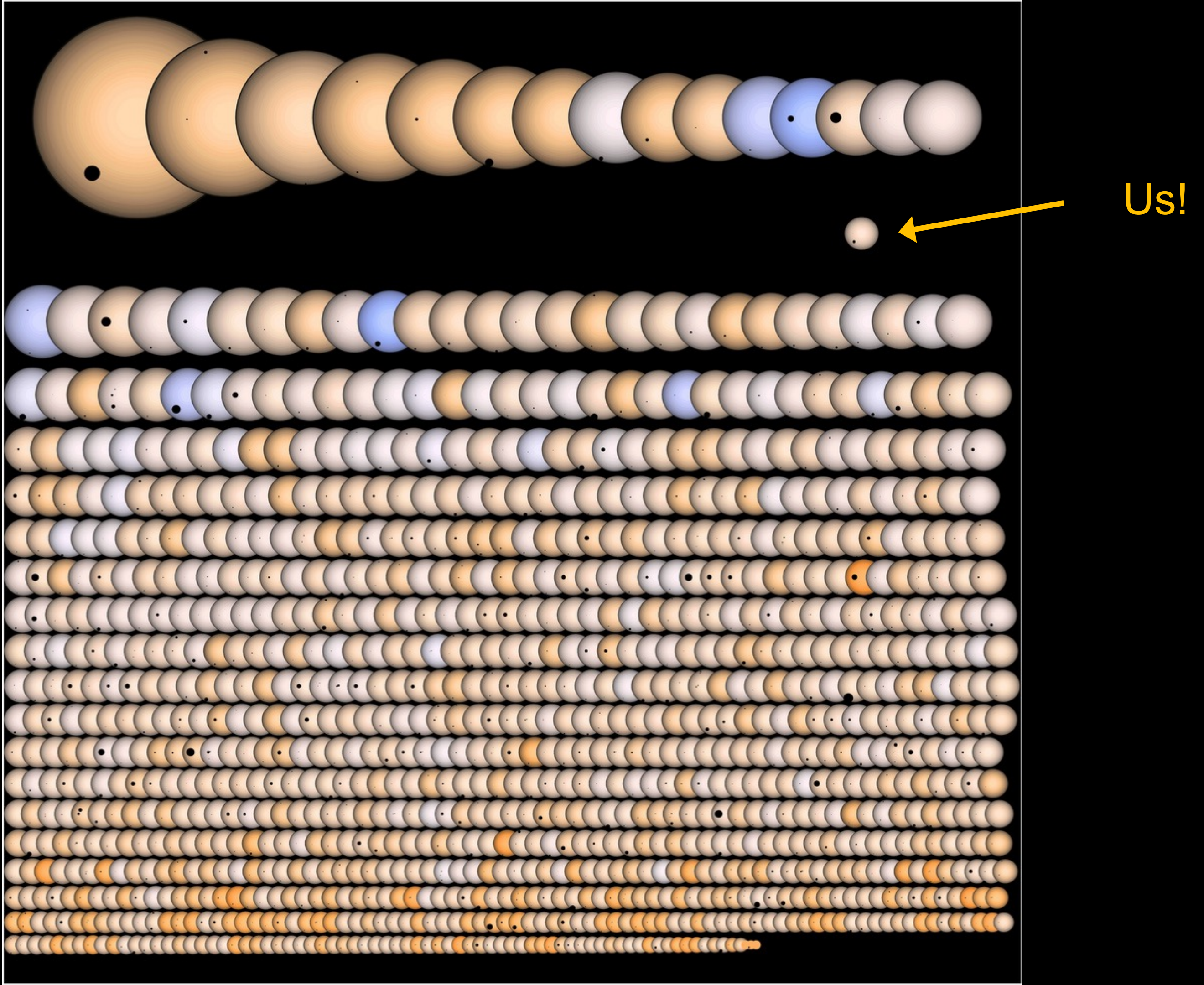


: KID 2010607. Period 18.633 days. KS statistic 0.972. Potential candidate.


0.002
0.000
−0.002
−0.004
−0.006
−0.008
−0.010
−0.012
Depth = 0.010, Noise = 0.001
Depth = 0.003, Noise = 0.001
Depth = 0.002, Noise = 0.001
Depth = 0.001, Noise = 0.001
1.00
1.05
1.10
1.15
1.20
1.25

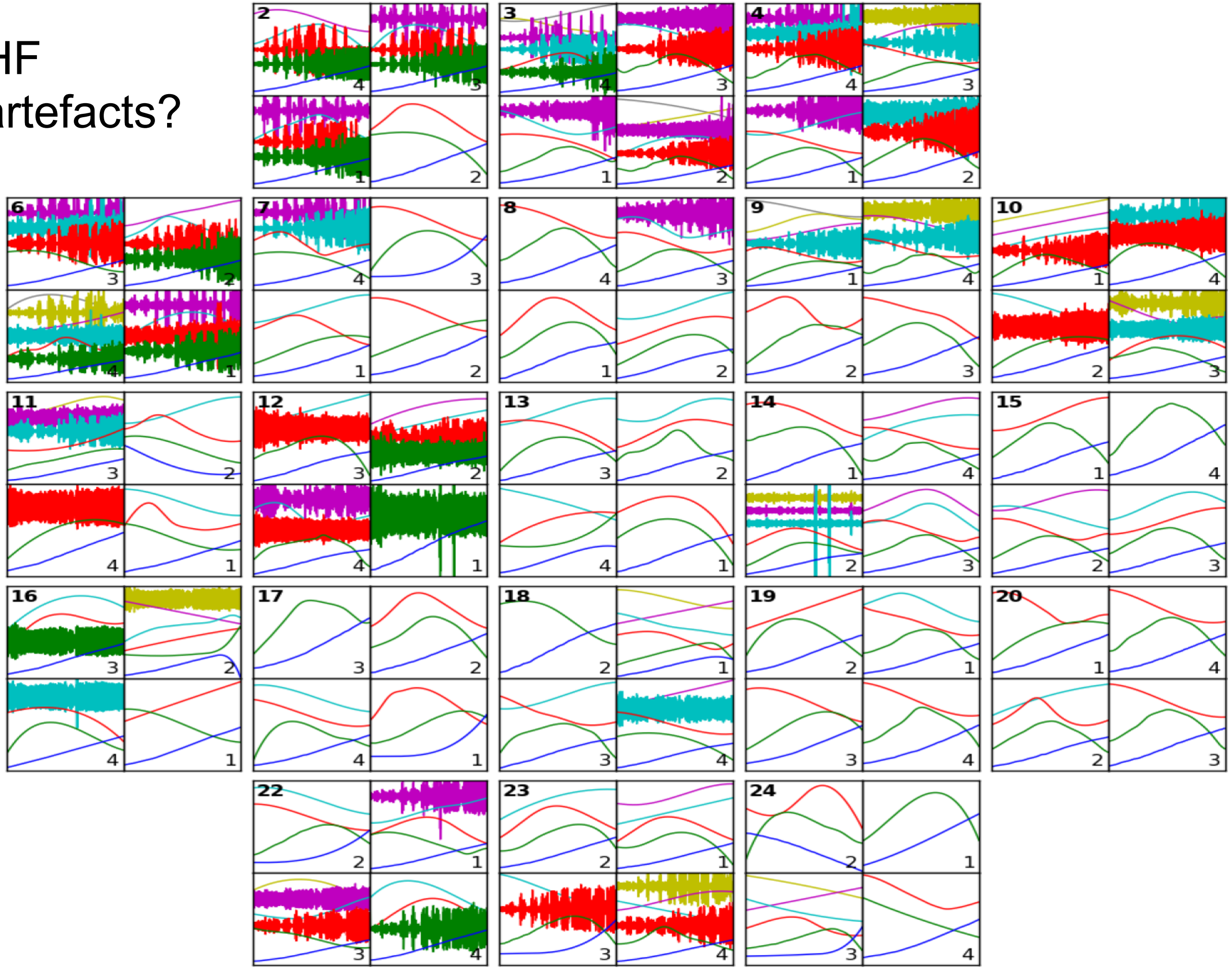
Planet size changes with wavelength



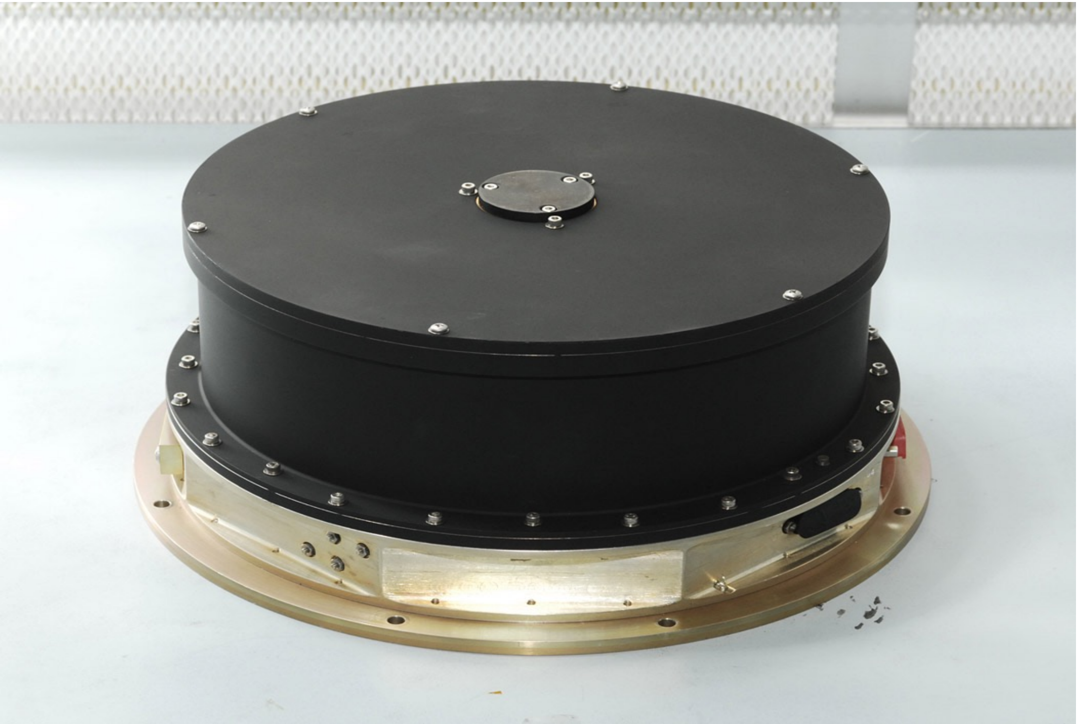

Us!


Kepler 186f

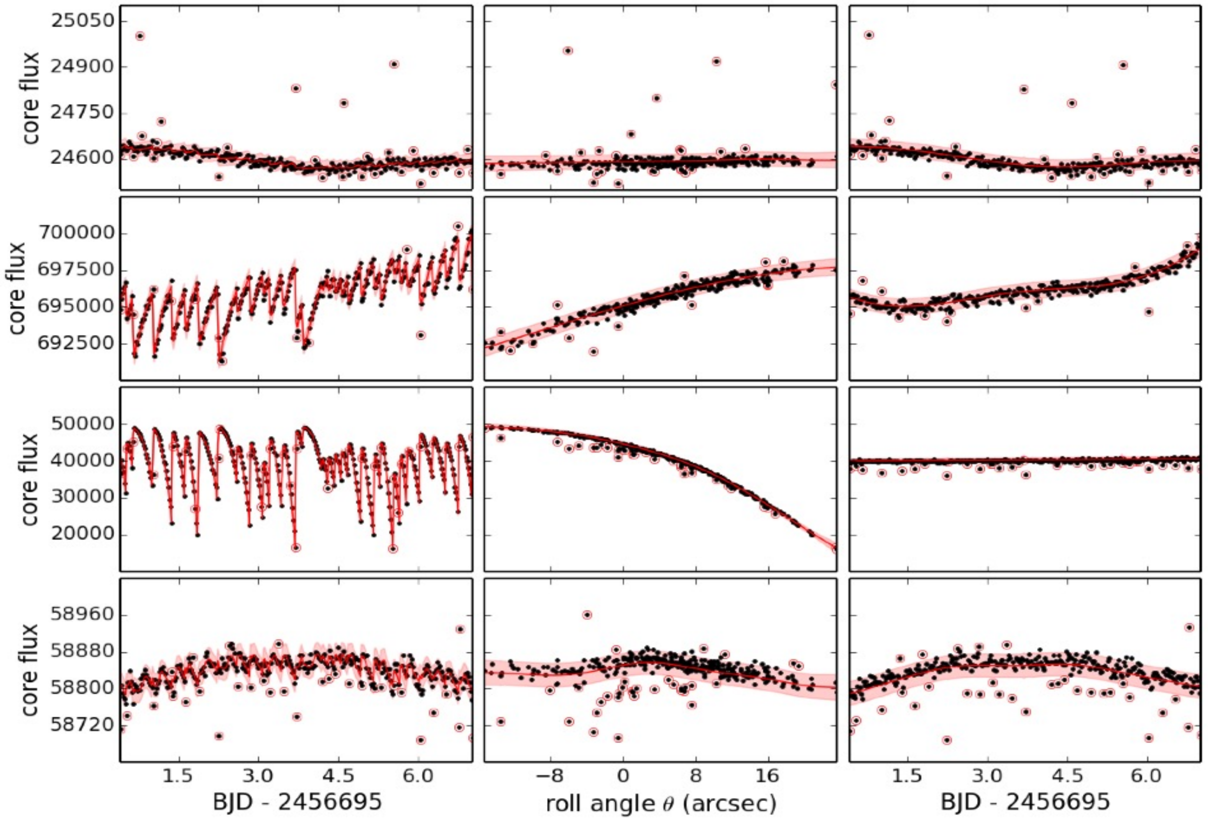
HF artefacts?



Reaction wheel failure



After the failure of reaction wheels, Kepler has moved to the K-2 mission- even larger systematics – requiring star by star analysis



Concluding remarks

- Kepler analysis is a case example of

- large data set
- asynchronous data
- unknown corruptions of unknown number
- K2 requires expansion of iteration to remove outliers
- valuable outcome in terms of science goals

- Bayesian non-parametrics (GPs) have proved invaluable

The future of science?

Automated systems can review data, extract meaning & find explanations faster & at scales that we can only dream of

Is the era of human science ending?

Has the era of the Automated Scientist begun?

With particular thanks to

Mike Osborne, Suzanne Aigrain, Aris Karastergiou, Chris Lintott, Matt Jarvis, Edwin Simpson, Steve Reece, Adam Cobb, Ivan Kiskin

and, of course, **machine learning algorithms...**

Ghost in the time series: no planet for Alpha Cen B

V. Rajpaul,[1★] S. Aigrain,[1] and S. Roberts[2]


Questions?