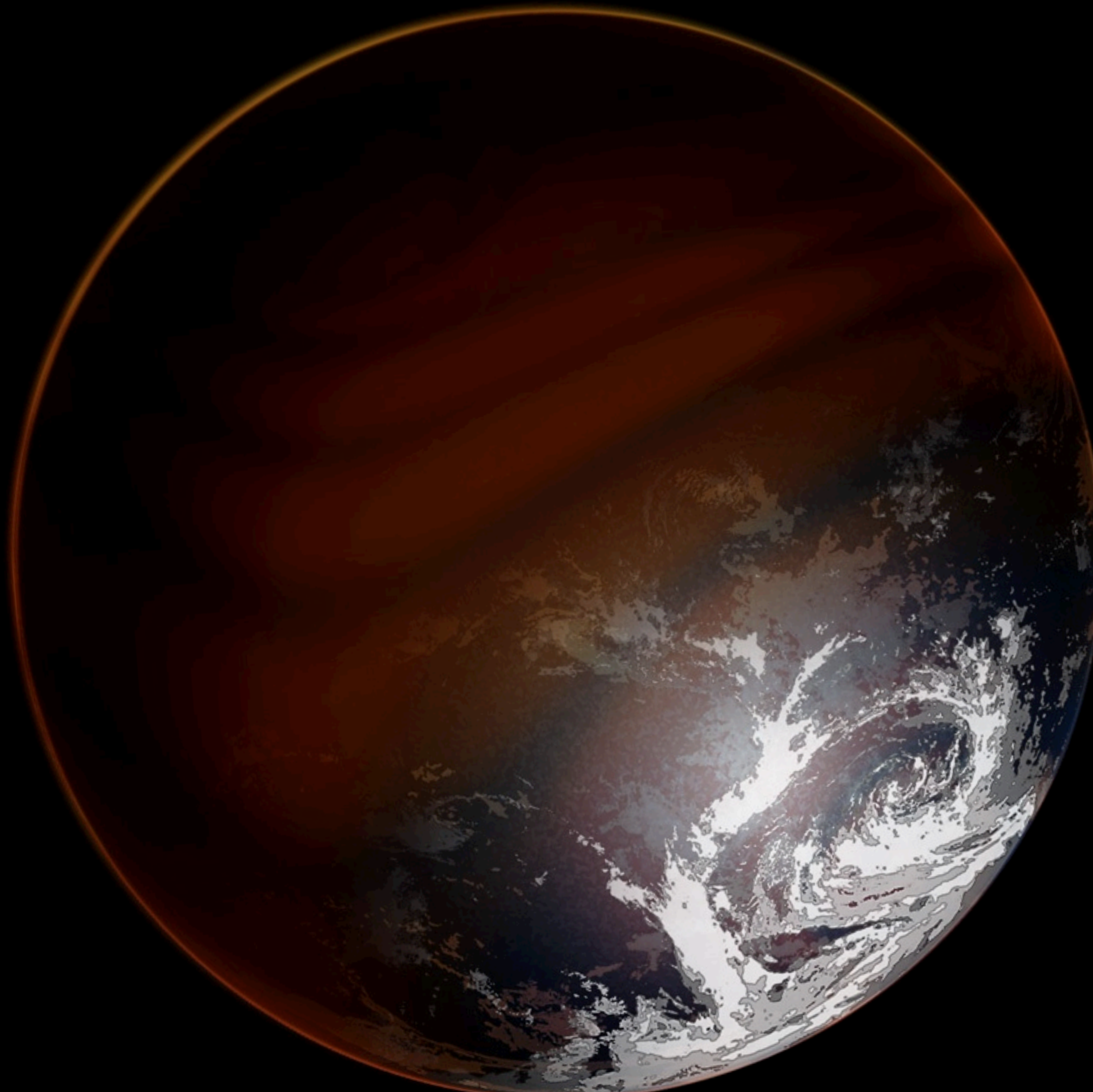
Learning to disentangle exoplanet signals from correlated noise

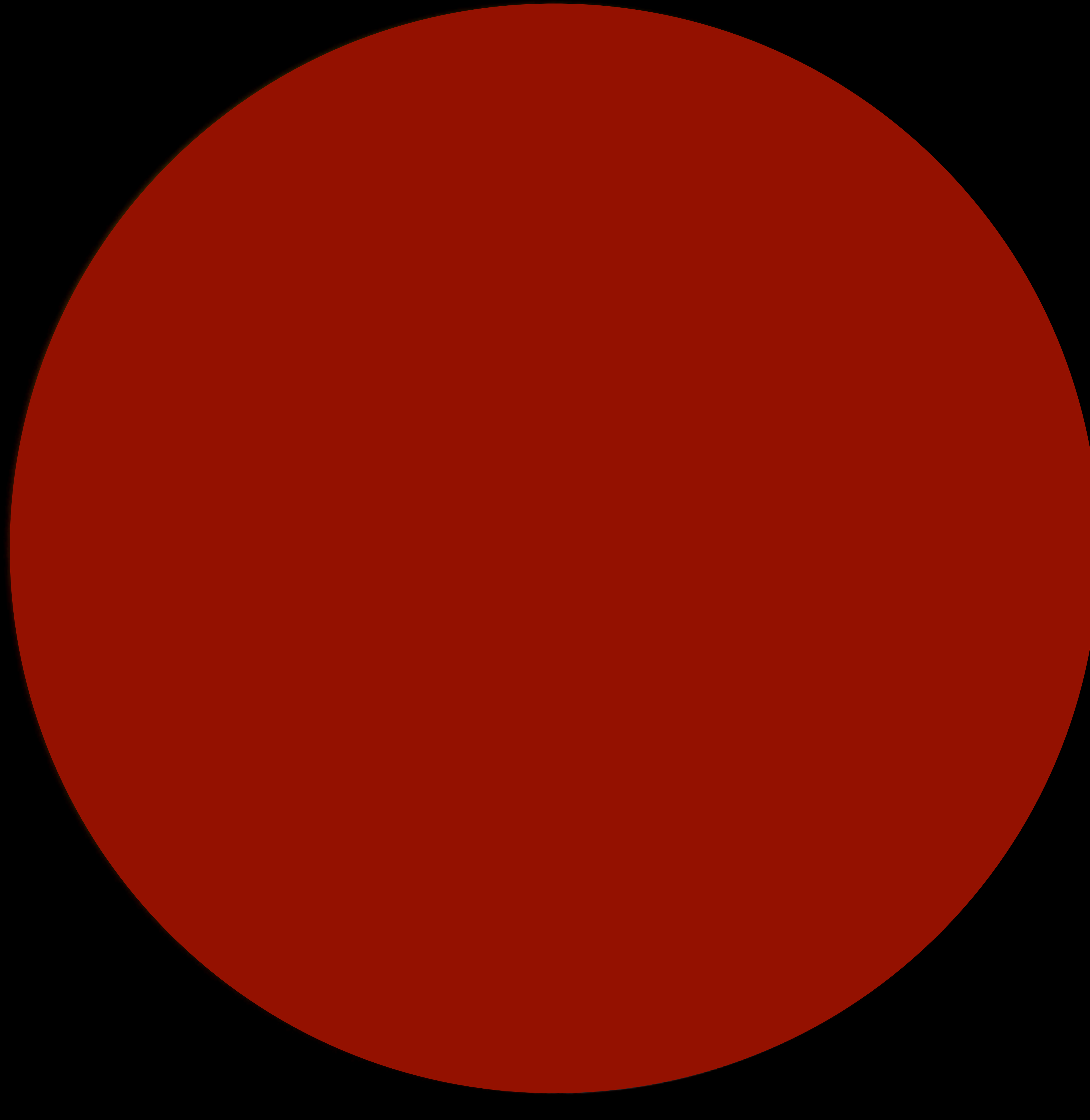
Suzanne Aigrain (University of Oxford)

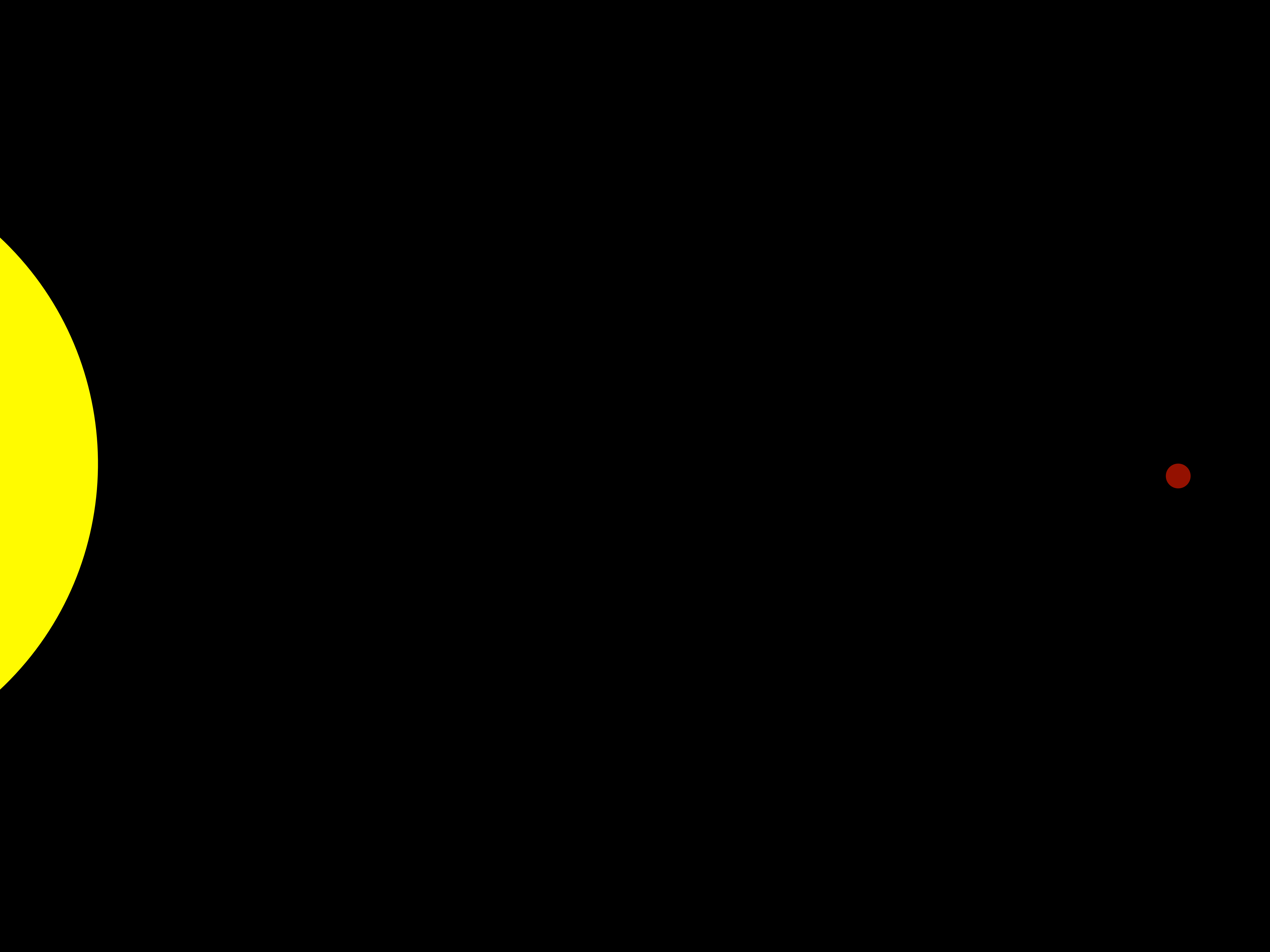
Learning to disentangle signals from correlated noise in exoplanet data

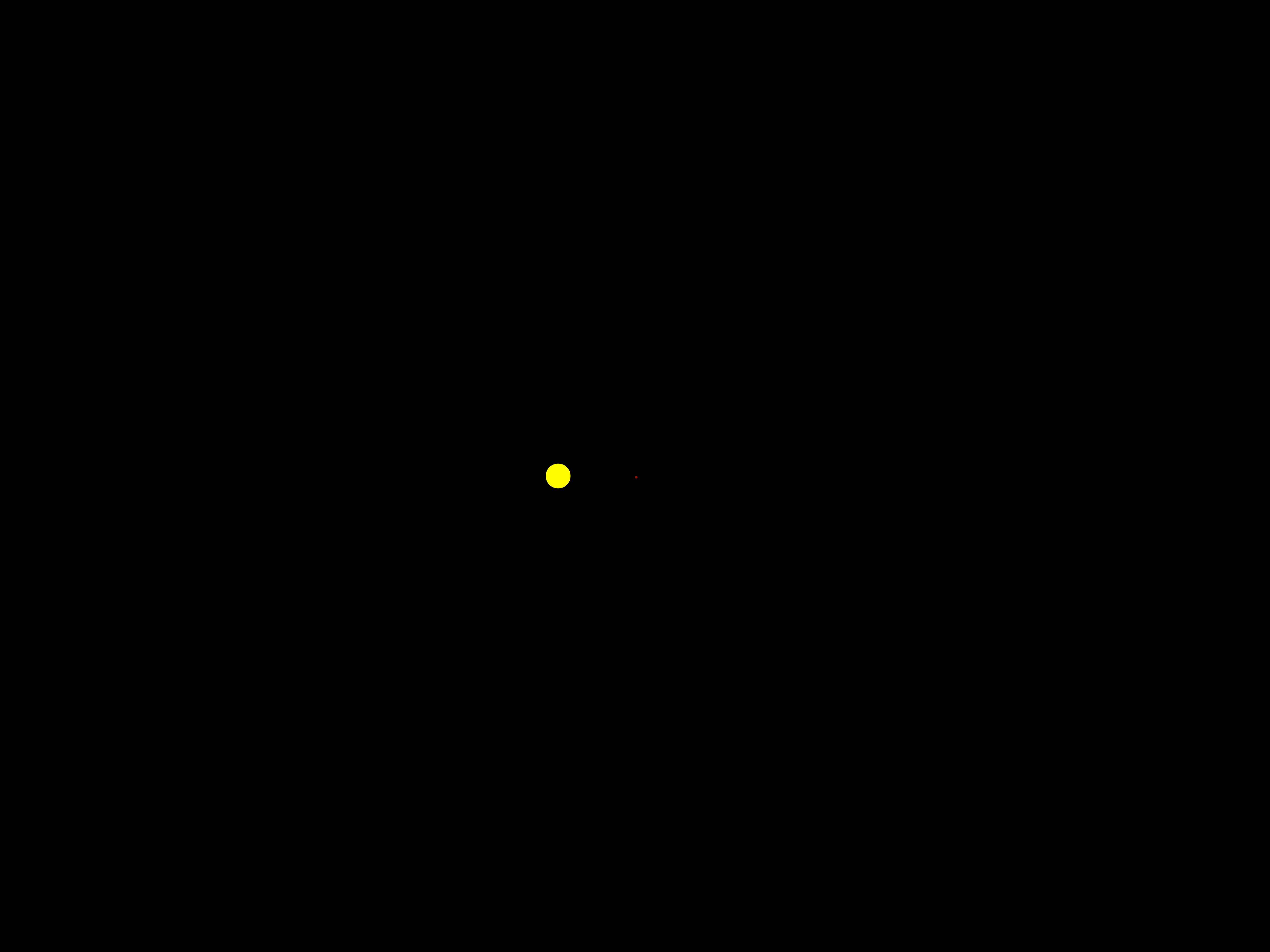
Suzanne Aigrain (University of Oxford)

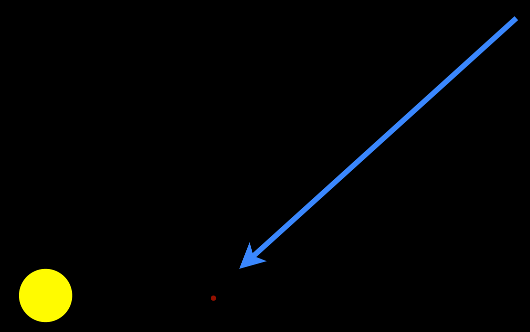
Neale Gibson, Amy McQuillan, Tom Evans
Stephen Roberts, Mike Osborne, Steven Reece
Frederic Pont, David Sing (Exeter)
Shay Zucker (Tel Aviv)

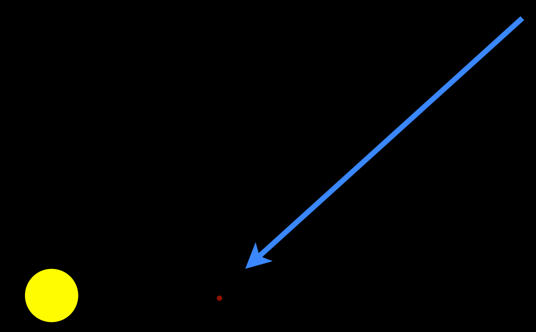






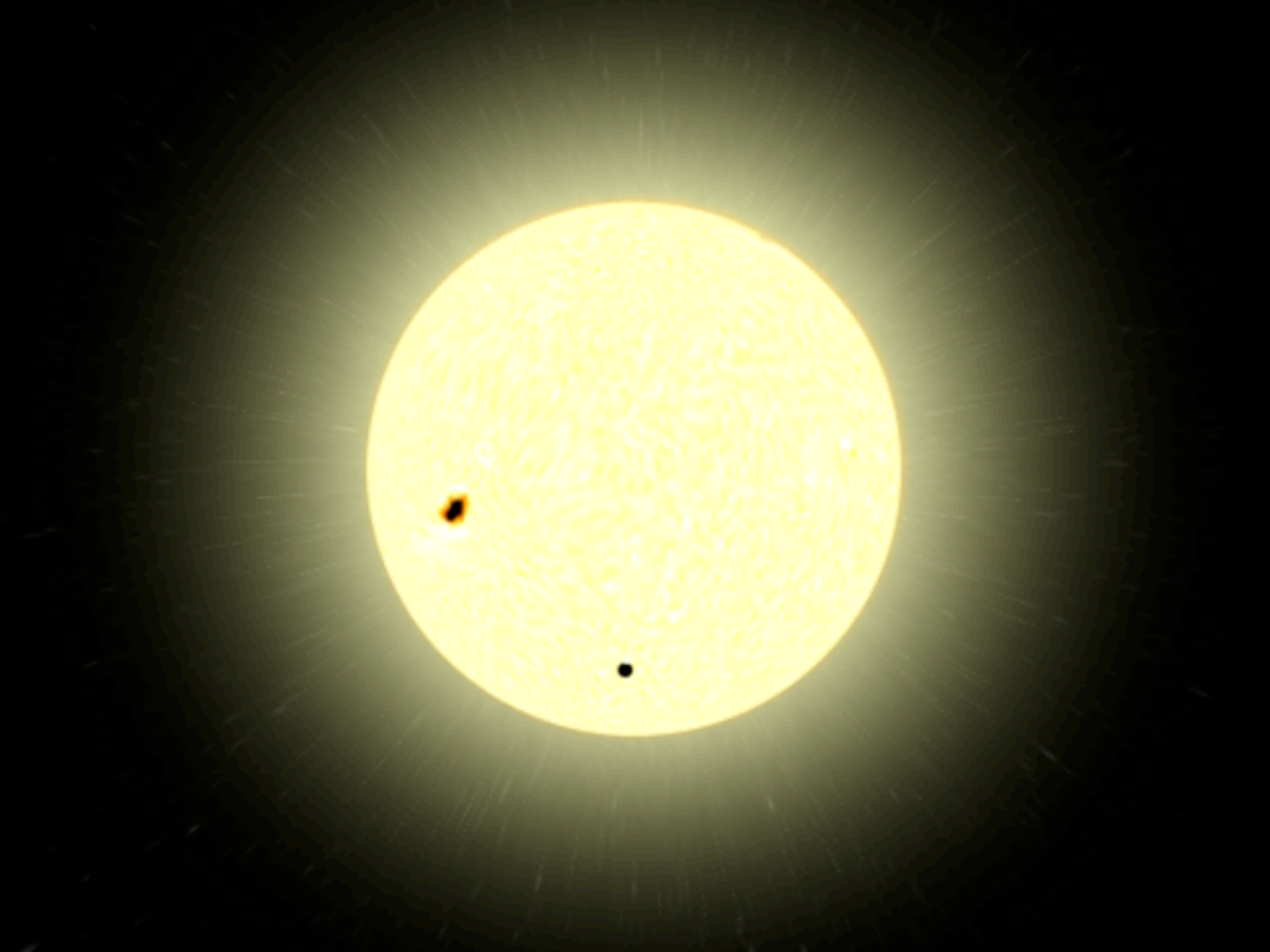


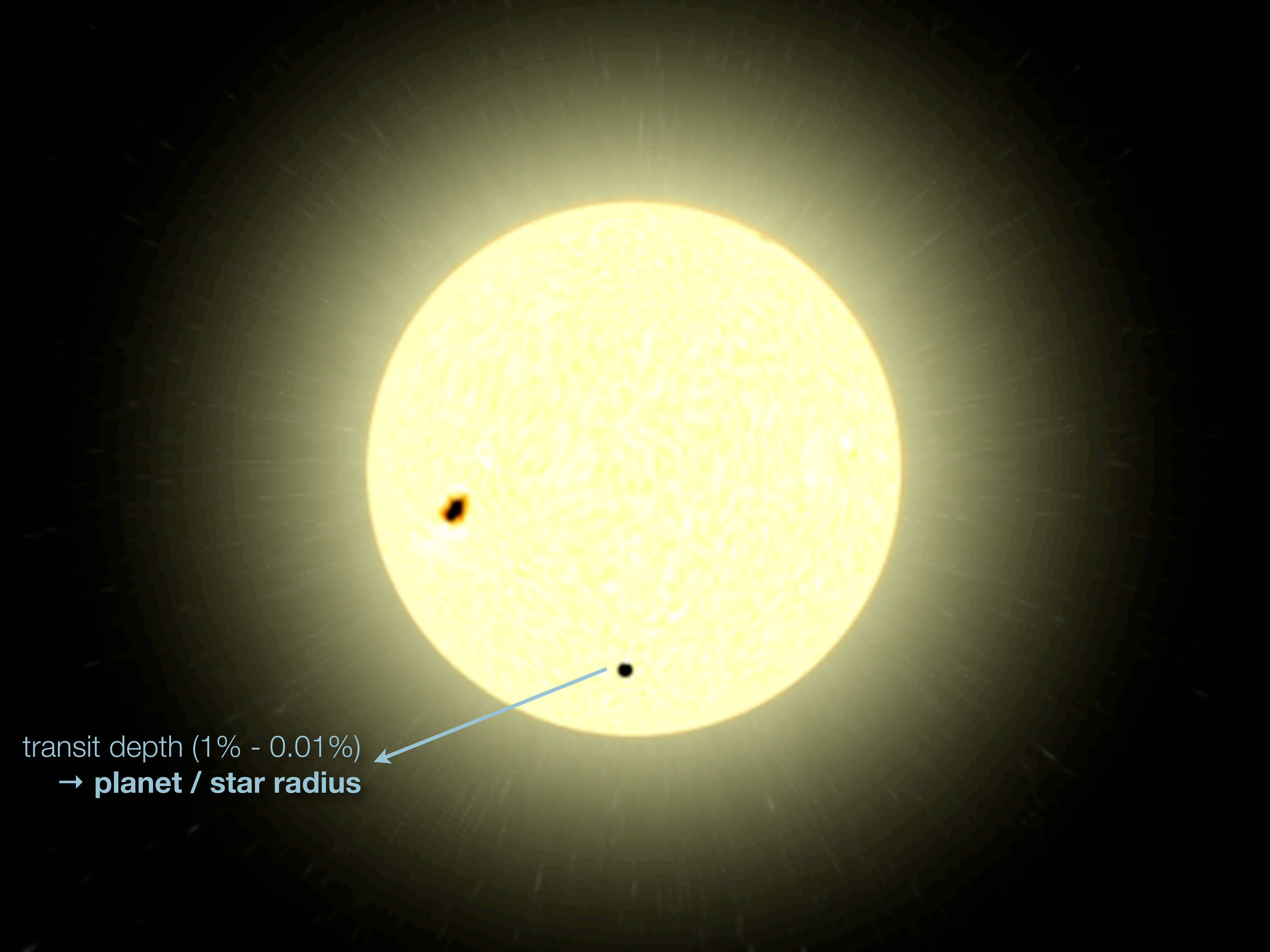


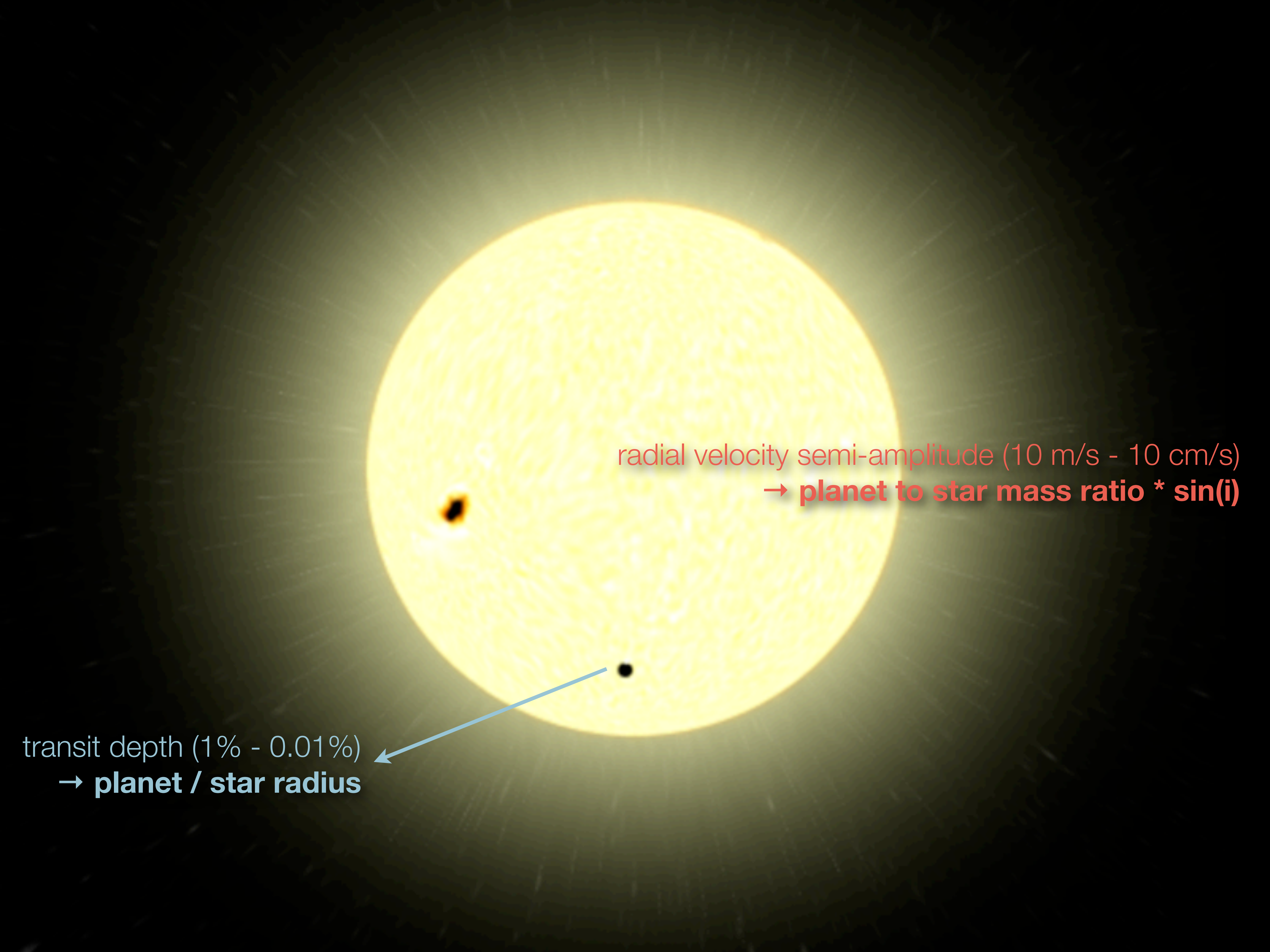


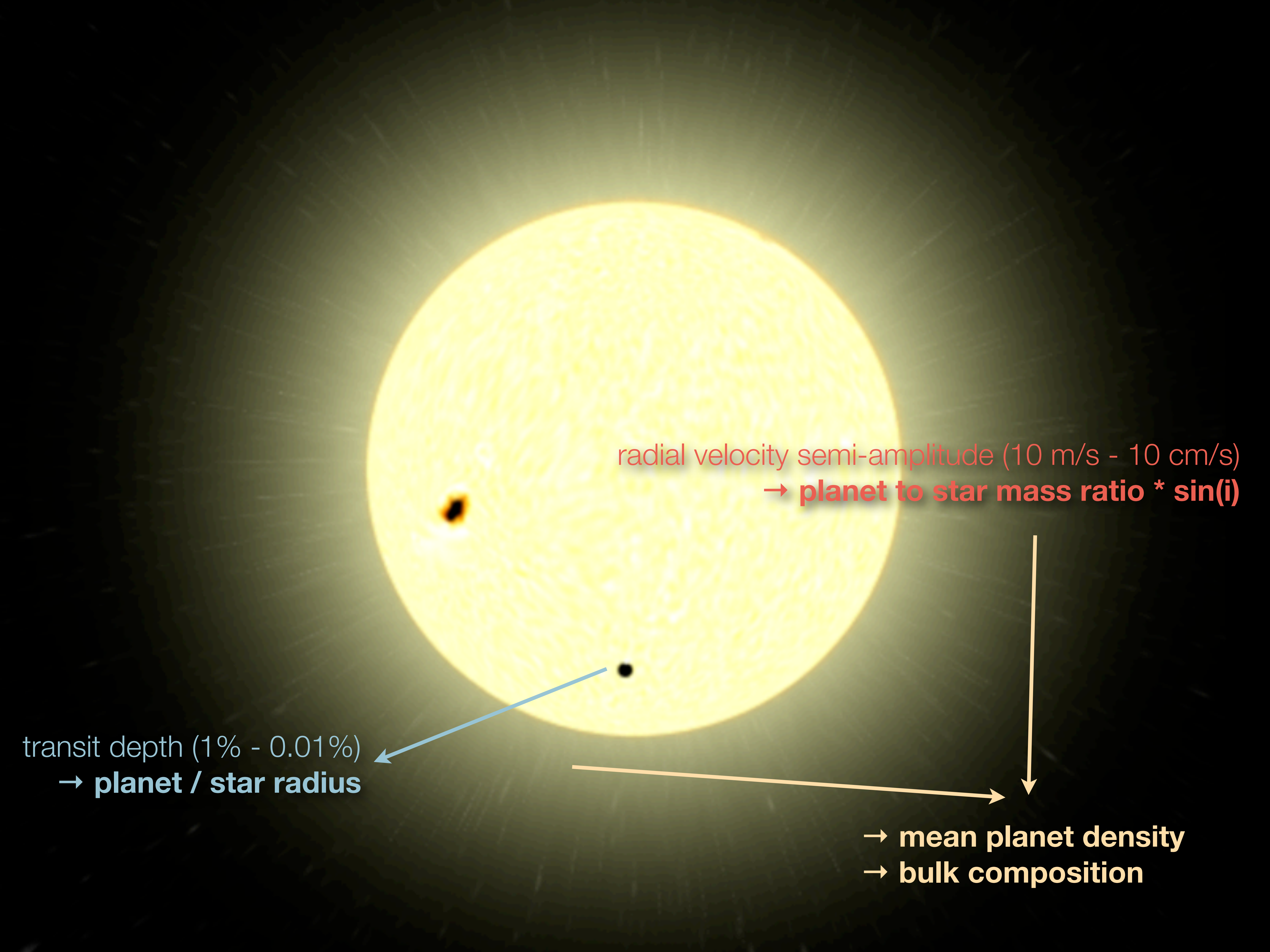
Long term goal for direct characterisation:
unresolved `images', low-resolution spectra

just achievable today for young giant planets at tens of AU
not before 2025 for Earth-like planets




transit depth (1% - 0.01%)
→ **planet / star radius**

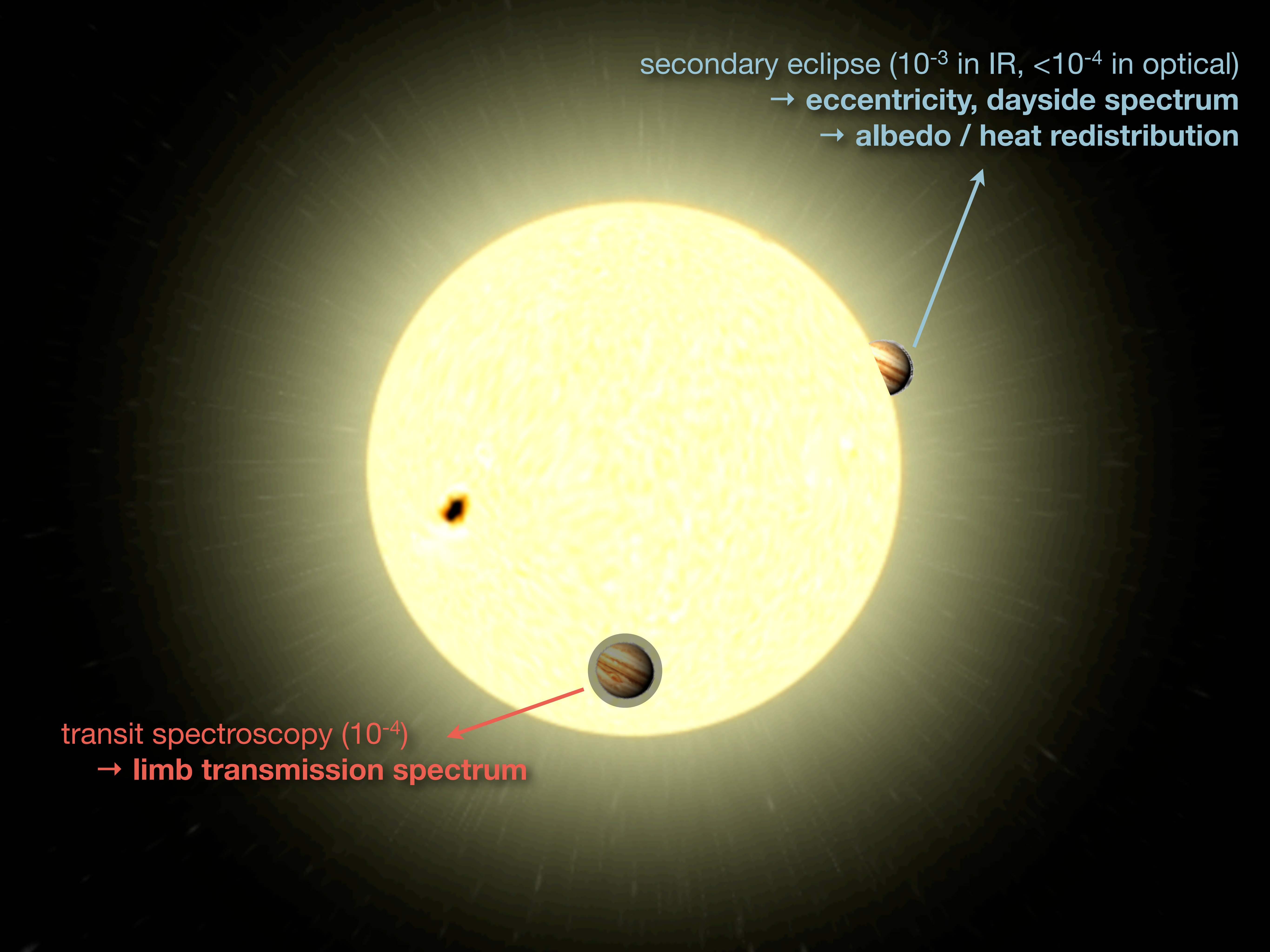

radial velocity semi-amplitude (10 m/s - 10 cm/s)
→ planet to star mass ratio * sin(i)
transit depth (1% - 0.01%)
→ planet / star radius


radial velocity semi-amplitude (10 m/s - 10 cm/s)
→ planet to star mass ratio * sin(i)
transit depth (1% - 0.01%)
→ planet / star radius
→ mean planet density
→ bulk composition

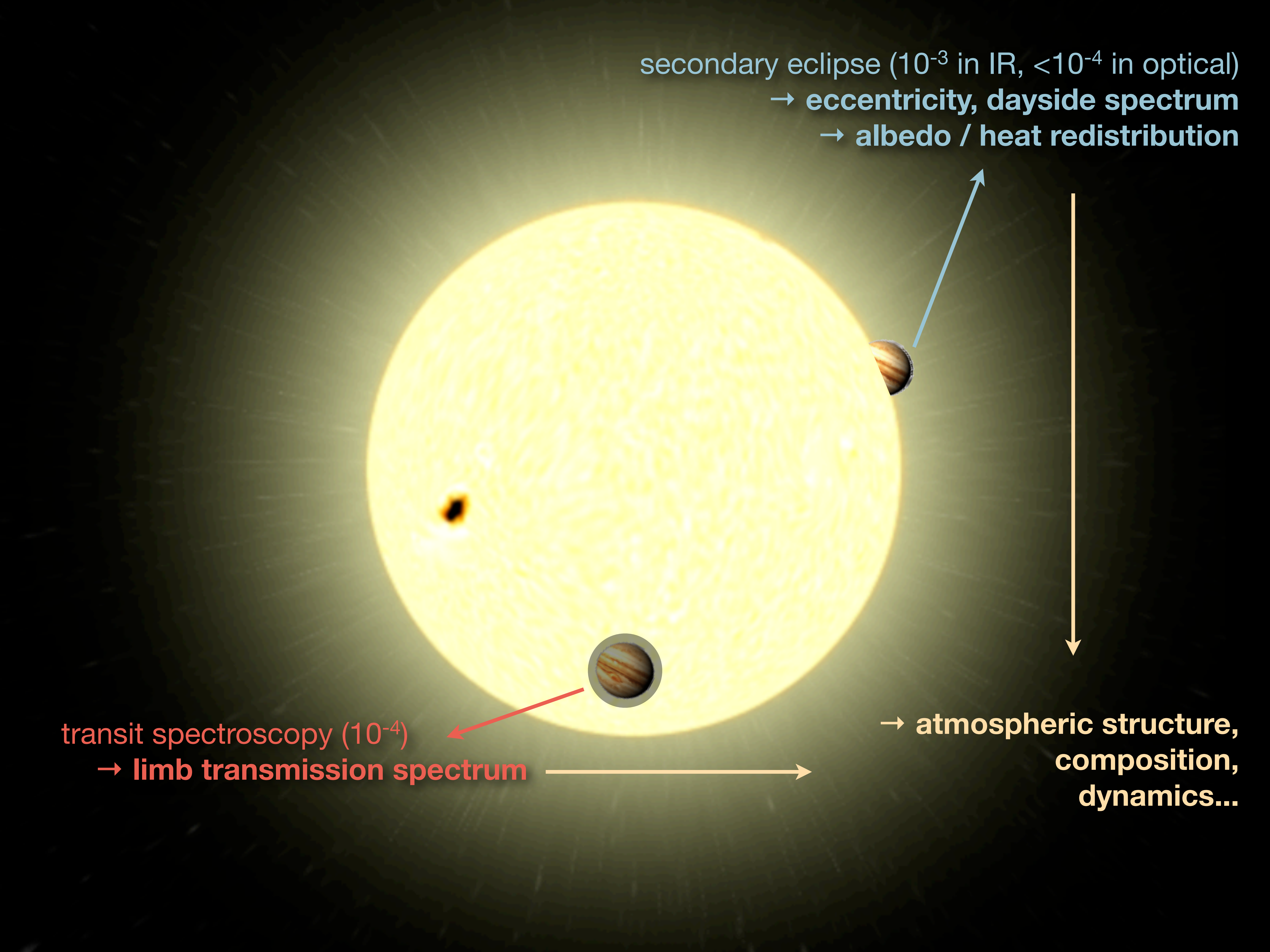
secondary eclipse (10^{-3} in IR, $<10^{-4}$ in optical)

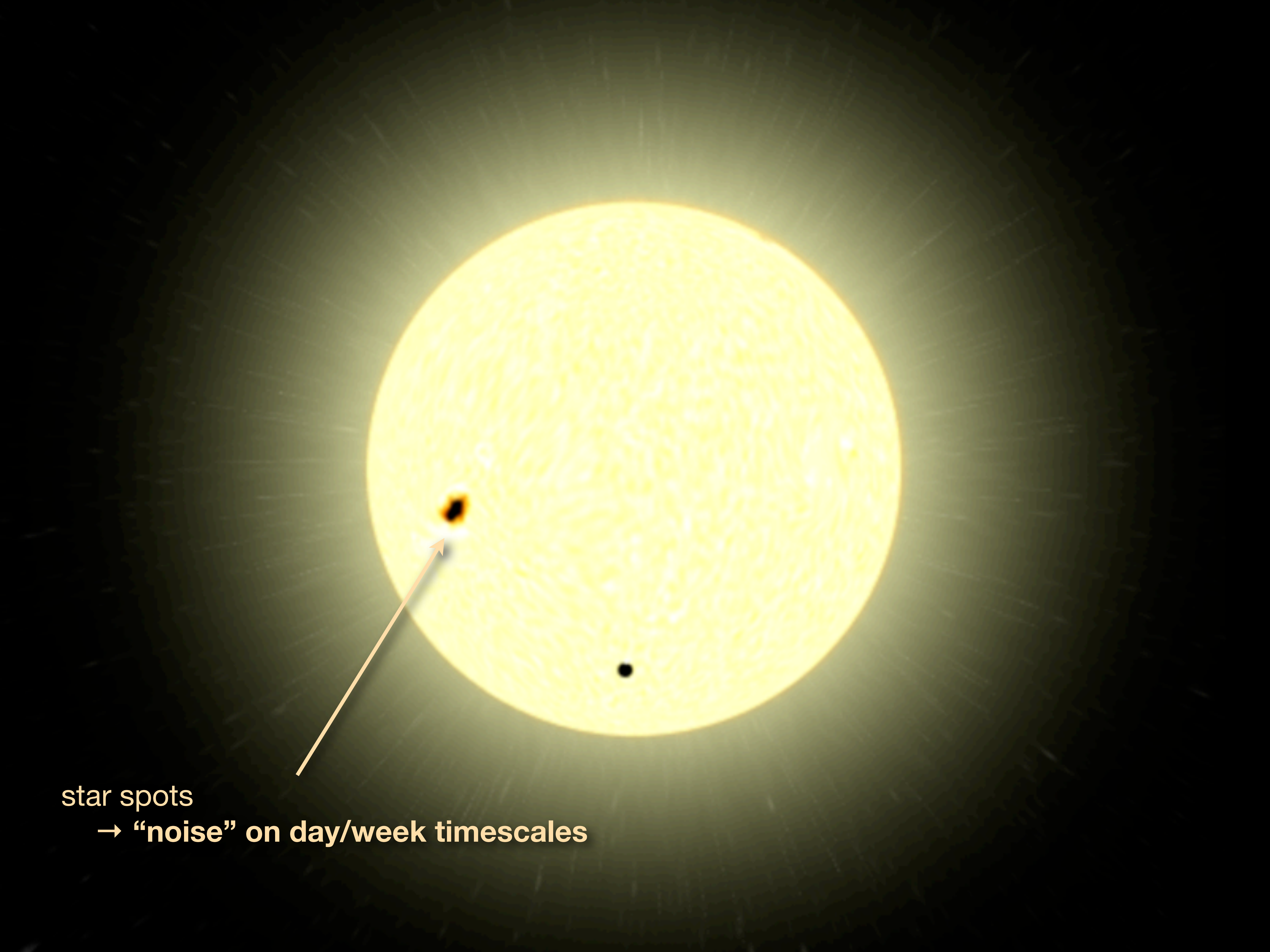
→ **eccentricity, dayside spectrum**

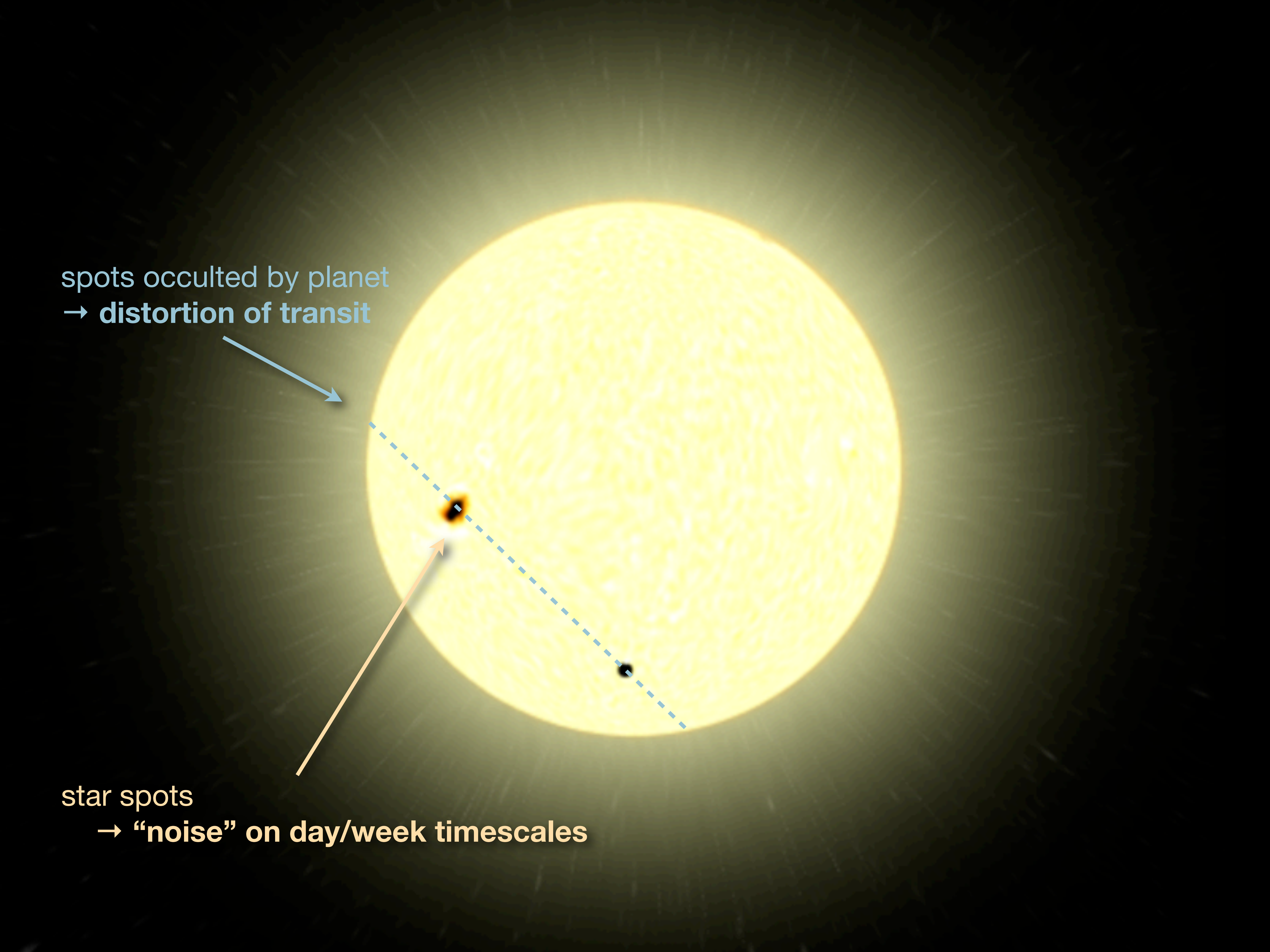
→ **albedo / heat redistribution**

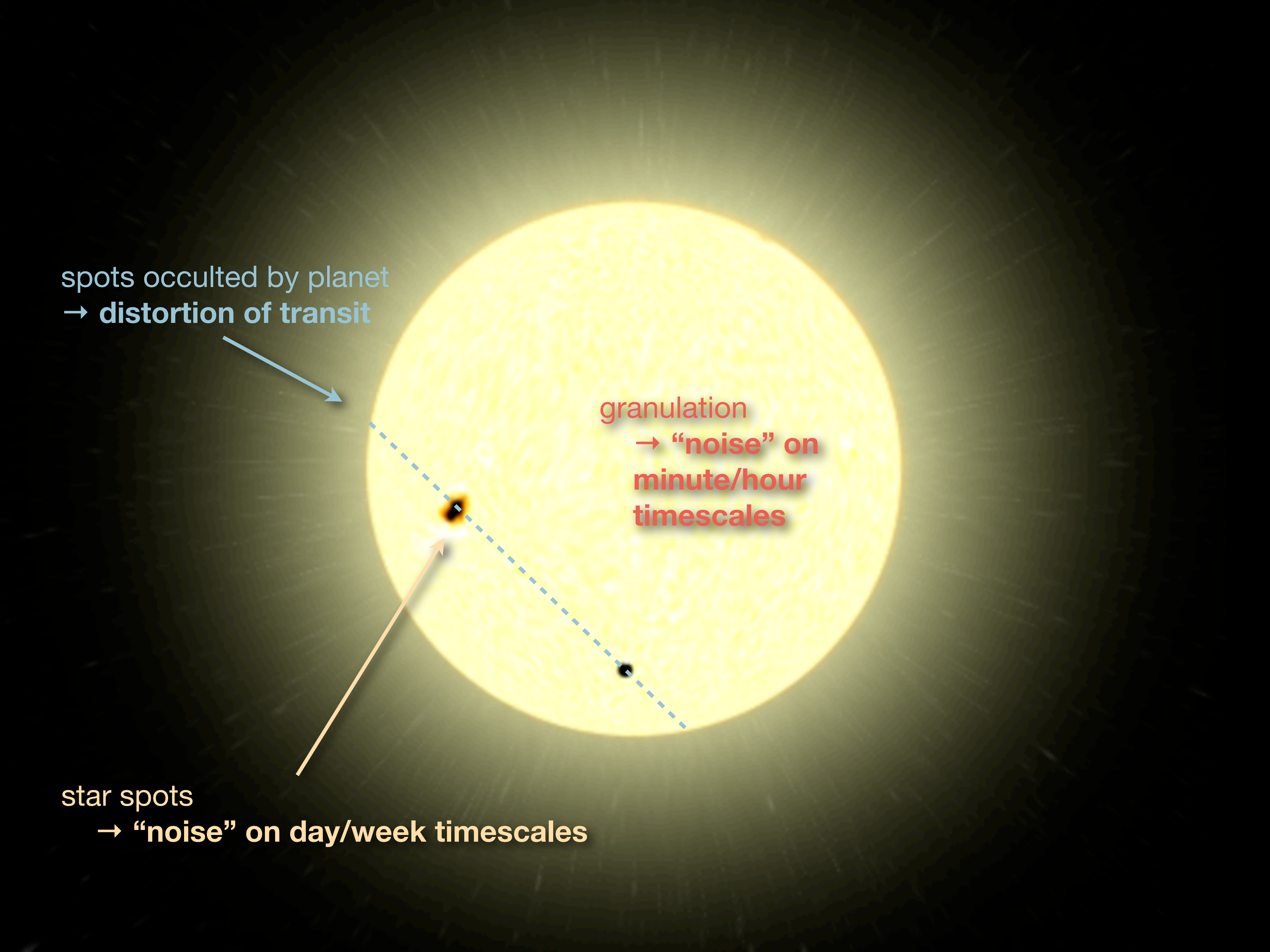


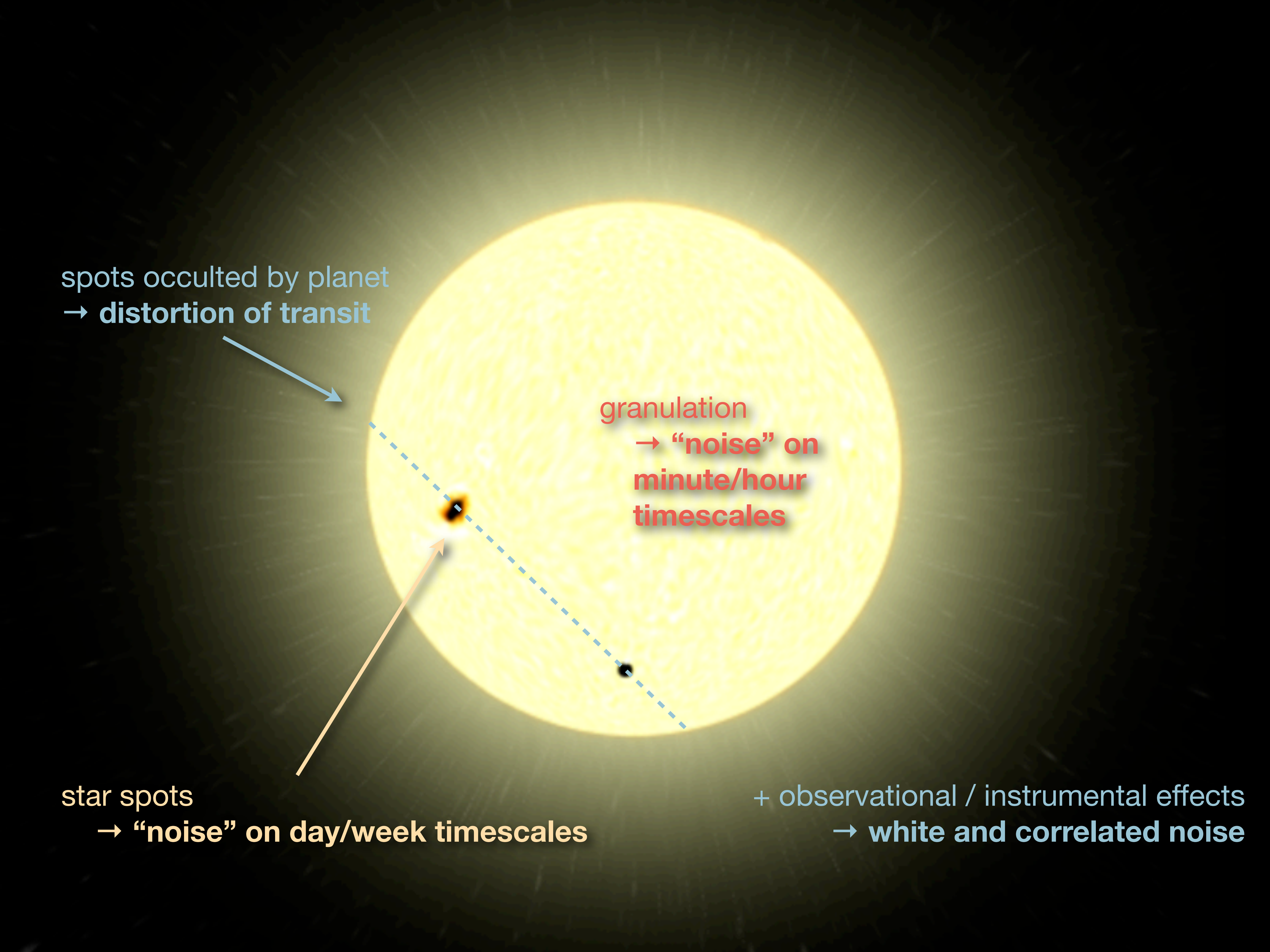
secondary eclipse (10^{-3} in IR, $<10^{-4}$ in optical)
→ **eccentricity, dayside spectrum**
→ **albedo / heat redistribution**
transit spectroscopy (10^{-4})
→ **limb transmission spectrum**

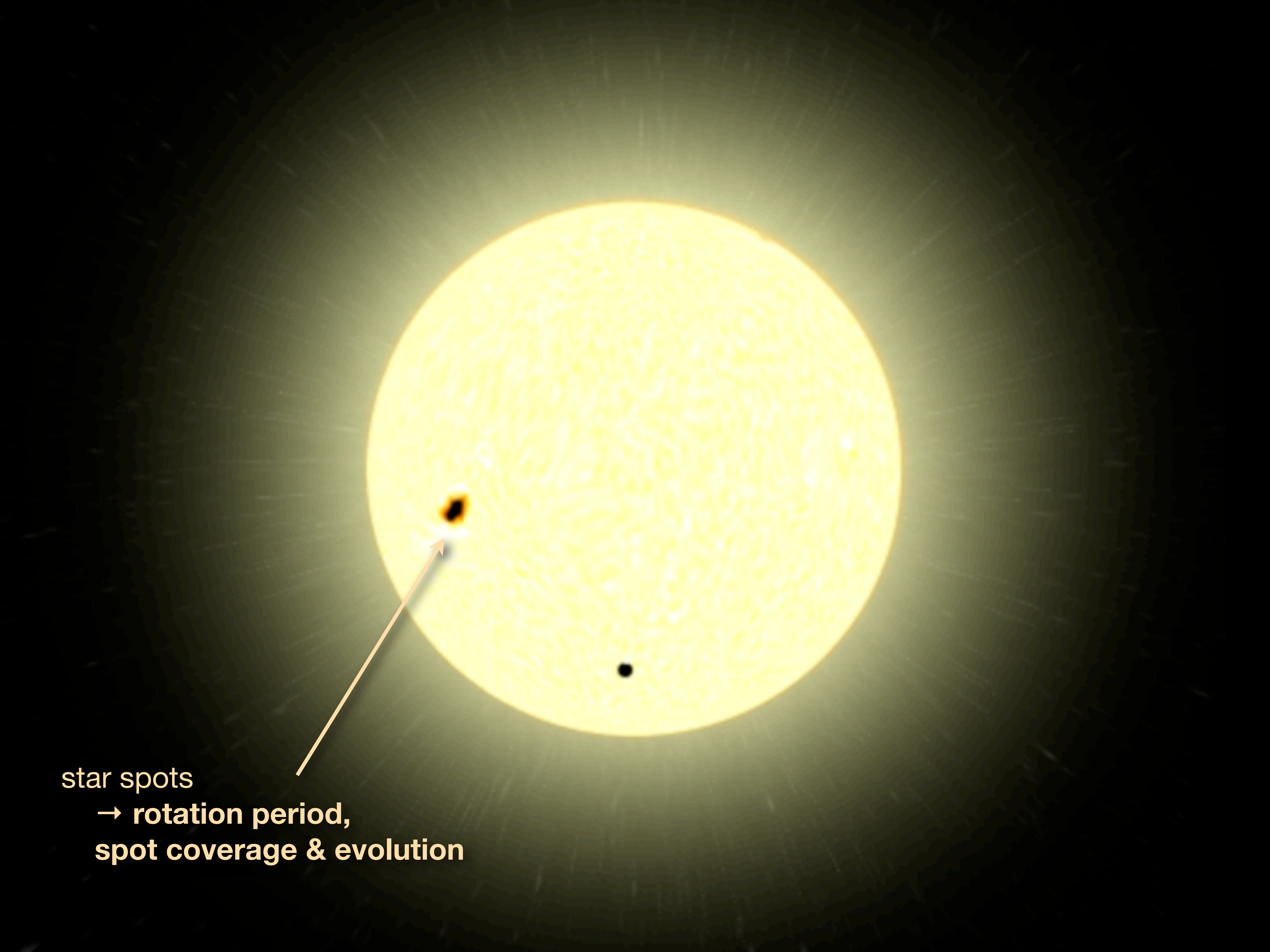

secondary eclipse (10^{-3} in IR, $<10^{-4}$ in optical)
→ eccentricity, dayside spectrum
→ albedo / heat redistribution
transit spectroscopy (10^{-4})
→ limb transmission spectrum
→ atmospheric structure,
composition,
dynamics...

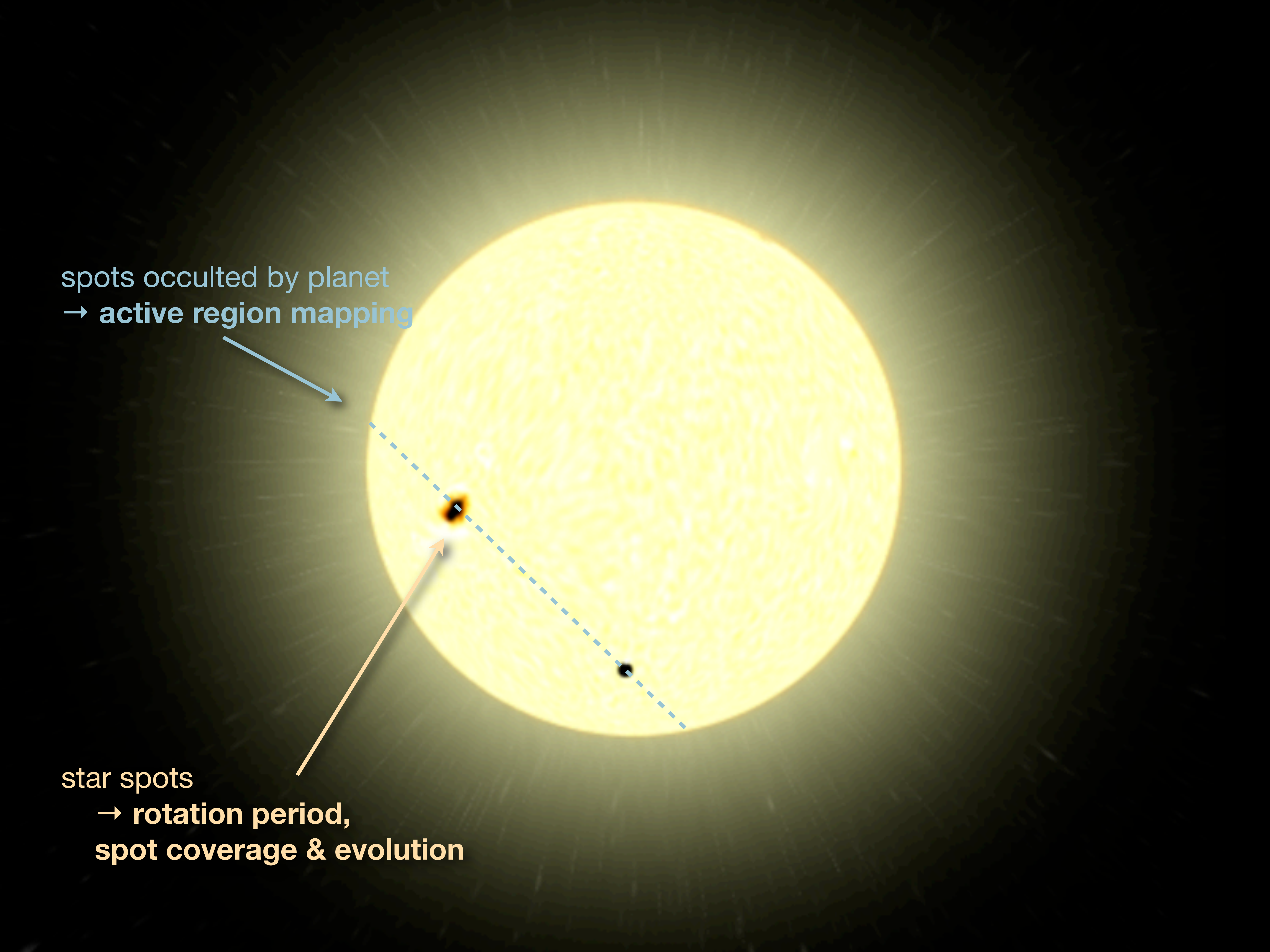

star spots
→ “noise” on day/week timescales

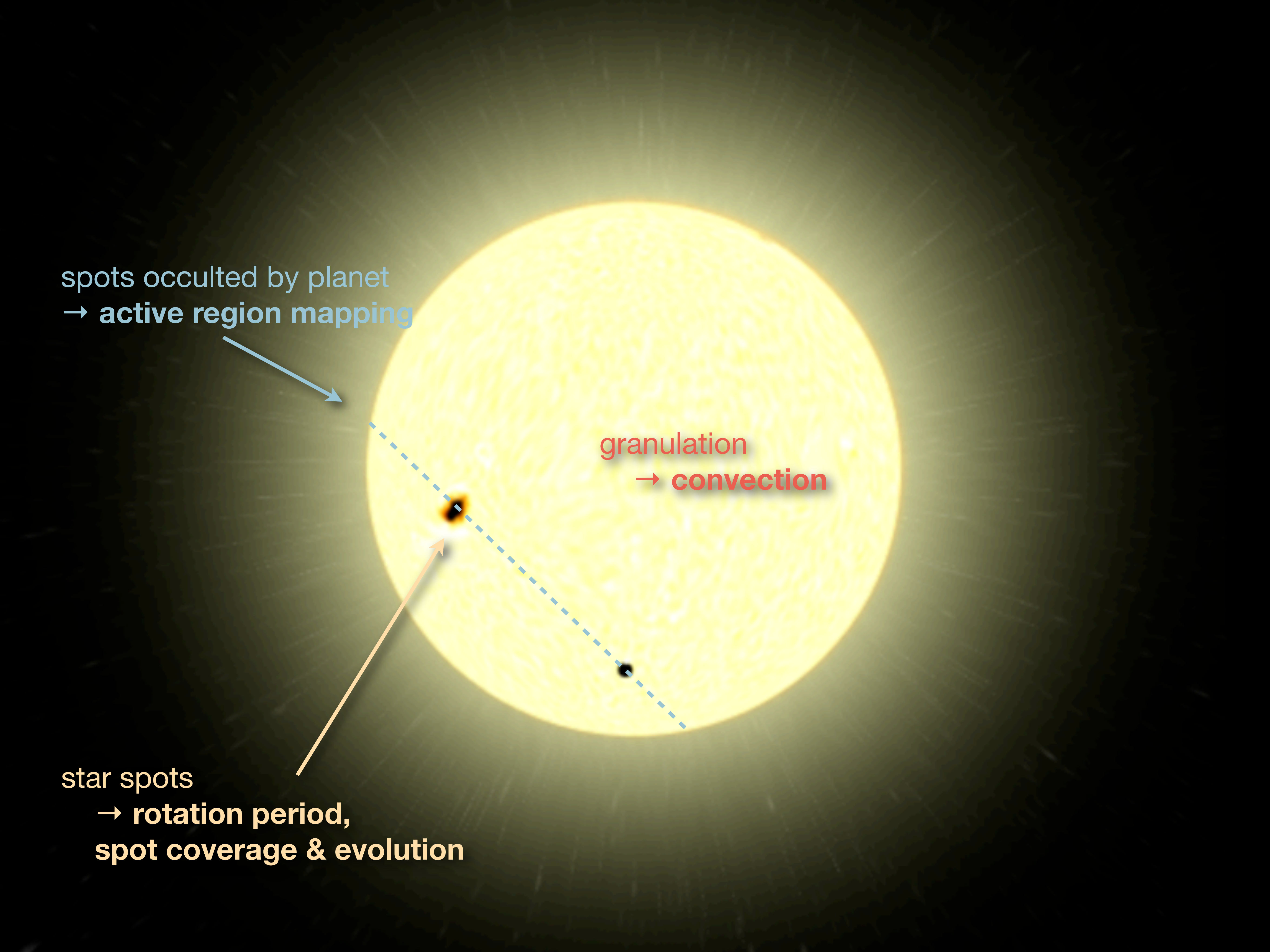

spots occulted by planet
→ distortion of transit
star spots
→ “noise” on day/week timescales


spots occulted by planet
→ distortion of transit
granulation
→ "noise" on minute/hour timescales
star spots
→ "noise" on day/week timescales


spots occulted by planet
→ distortion of transit
granulation
→ "noise" on minute/hour timescales
star spots
→ "noise" on day/week timescales
+ observational / instrumental effects
→ white and correlated noise


star spots
→ rotation period,
spot coverage & evolution

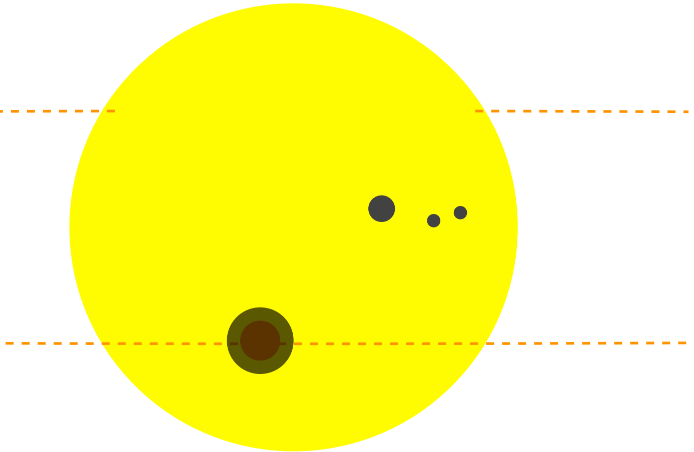

spots occulted by planet
→ active region mapping
star spots
→ rotation period,
spot coverage & evolution

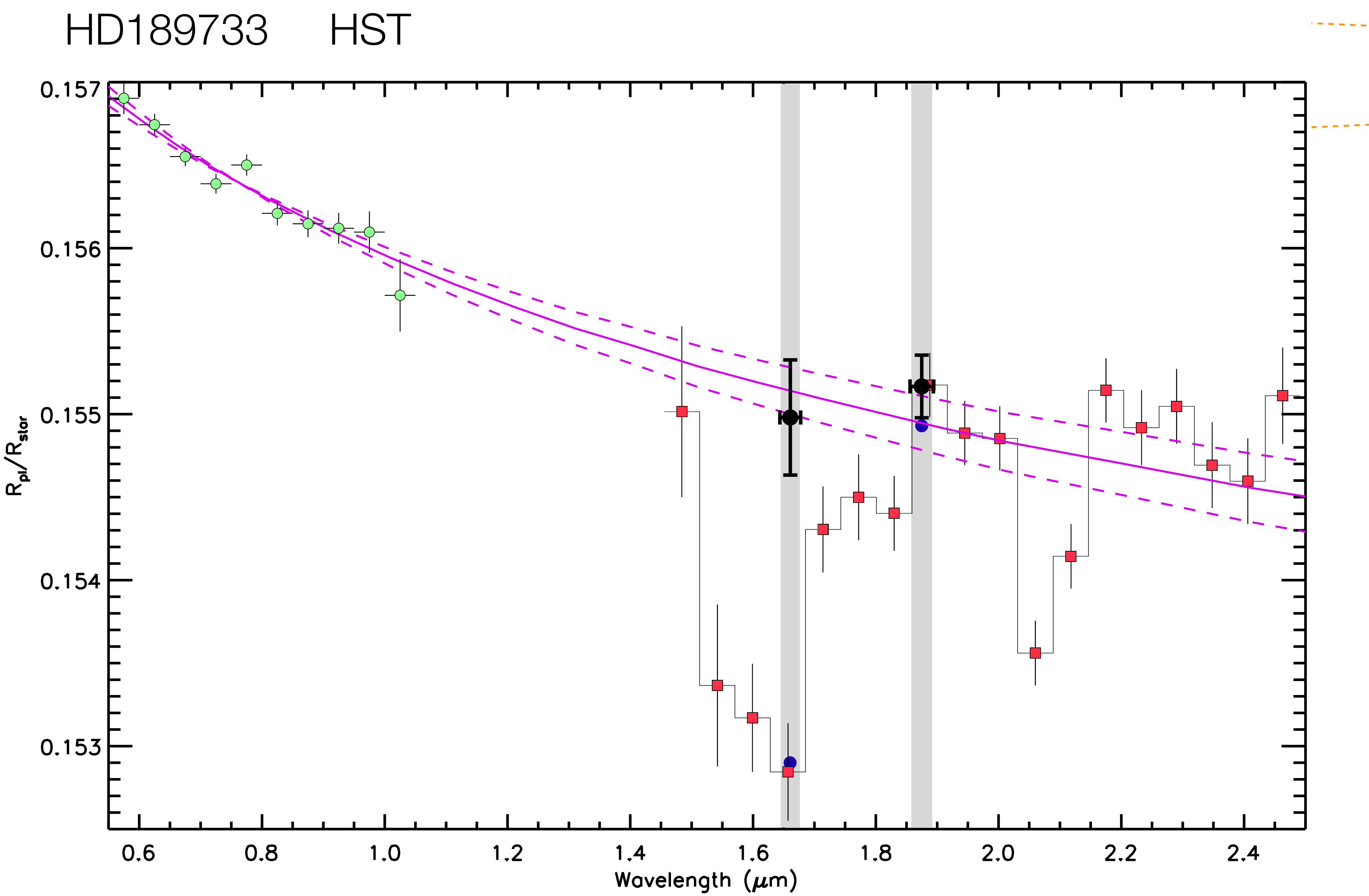

spots occulted by planet
→ active region mapping
granulation
→ convection
star spots
→ rotation period,
spot coverage & evolution

Transmission spectroscopy of exoplanet atmospheres with Gaussian Processes

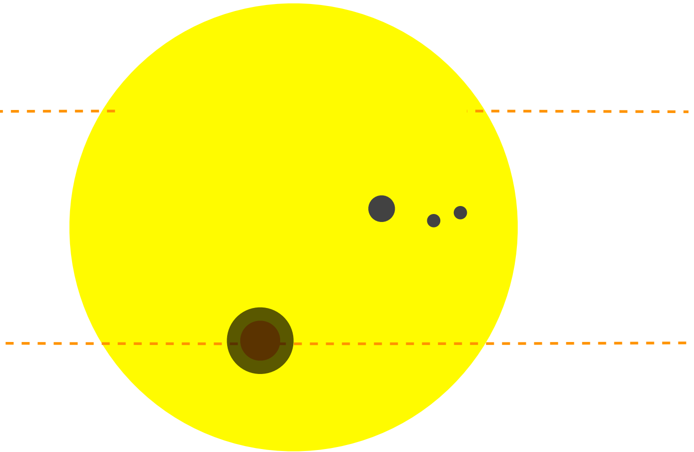
Gibson, Pont & Aigrain (2010), Gibson et al. (in prep)

Absorption spectra of hot Jupiters

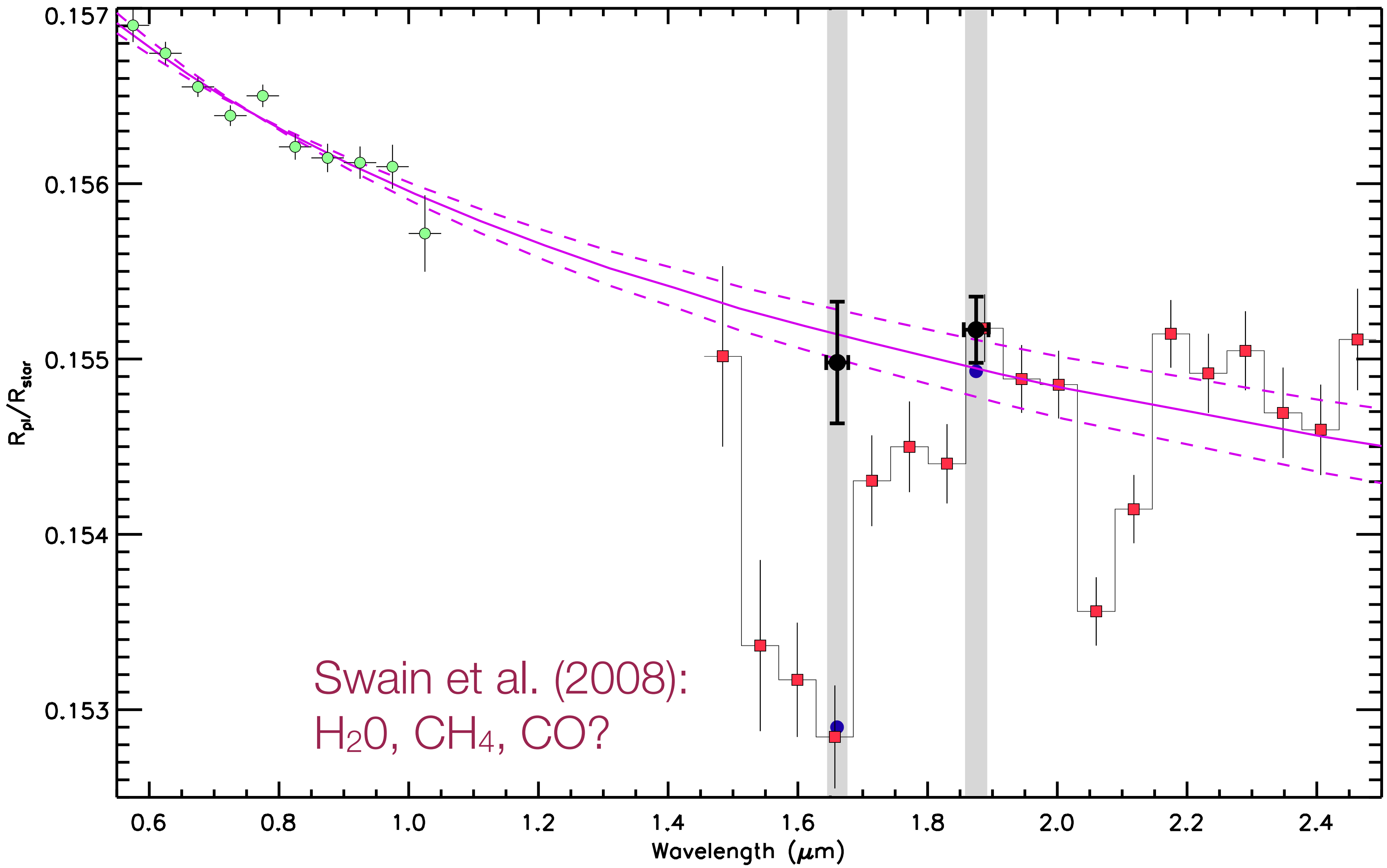




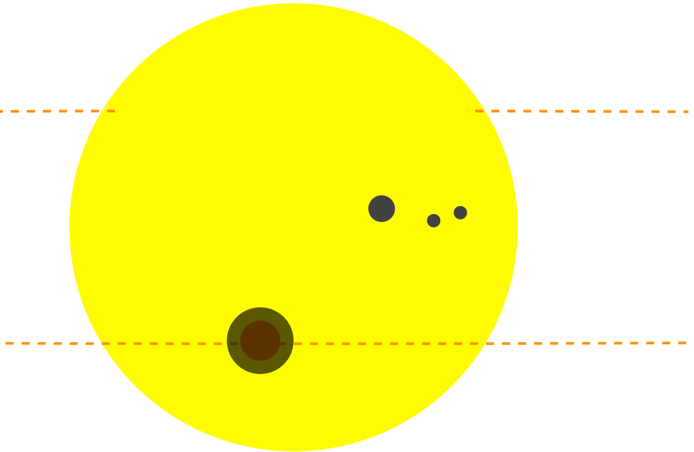
Absorption spectra of hot Jupiters

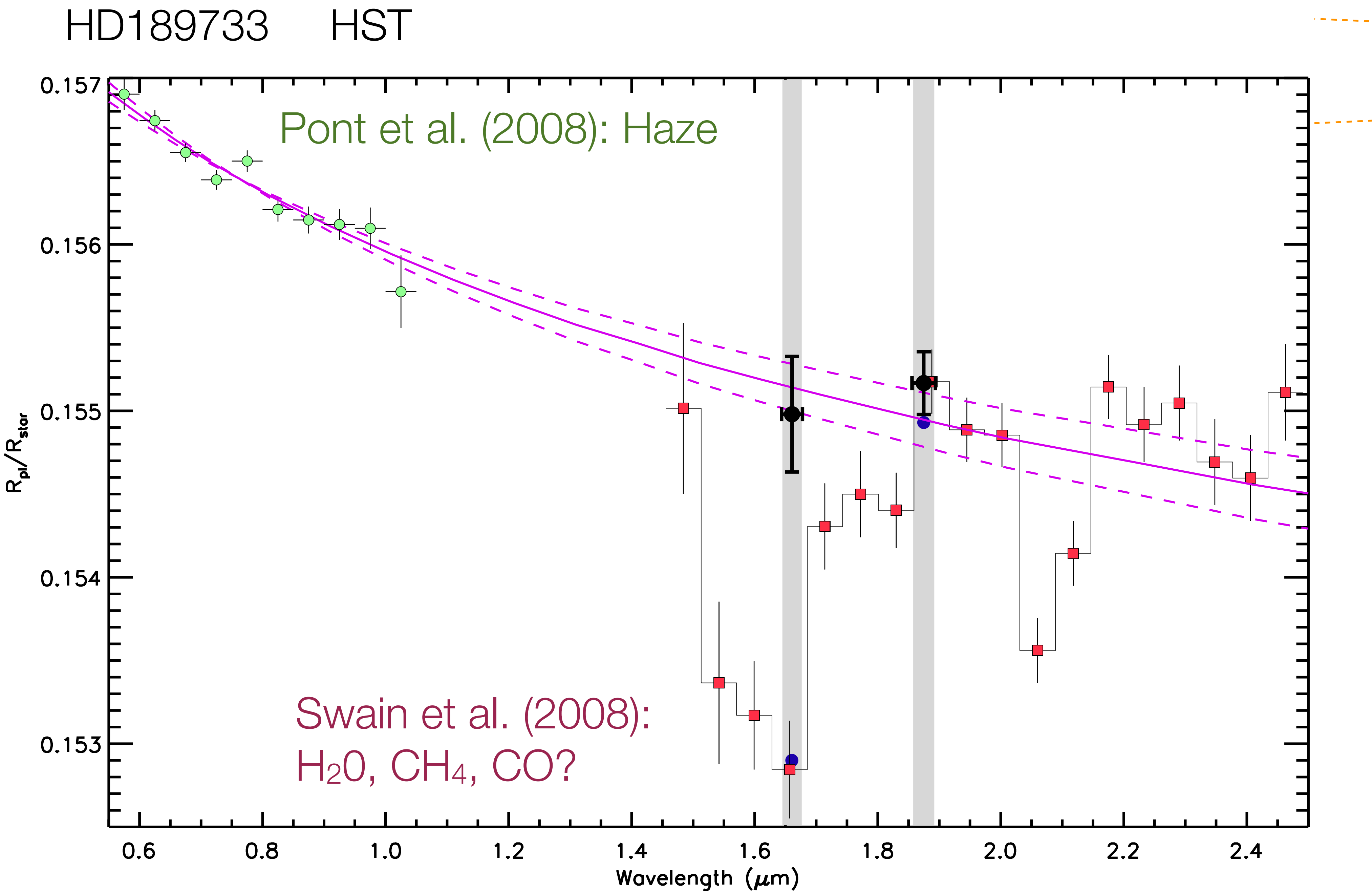


HD189733 HST

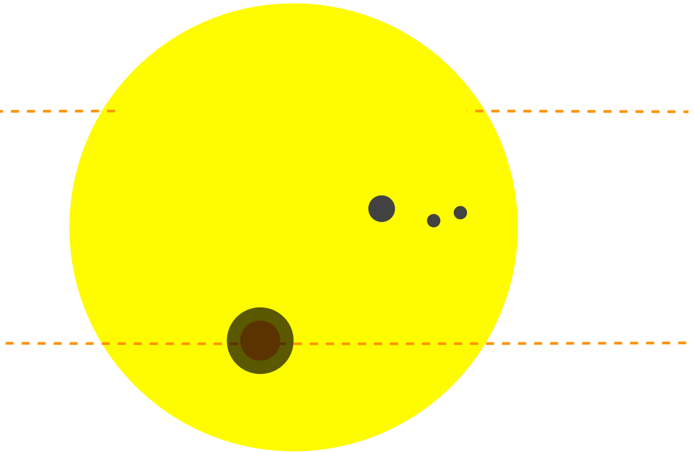


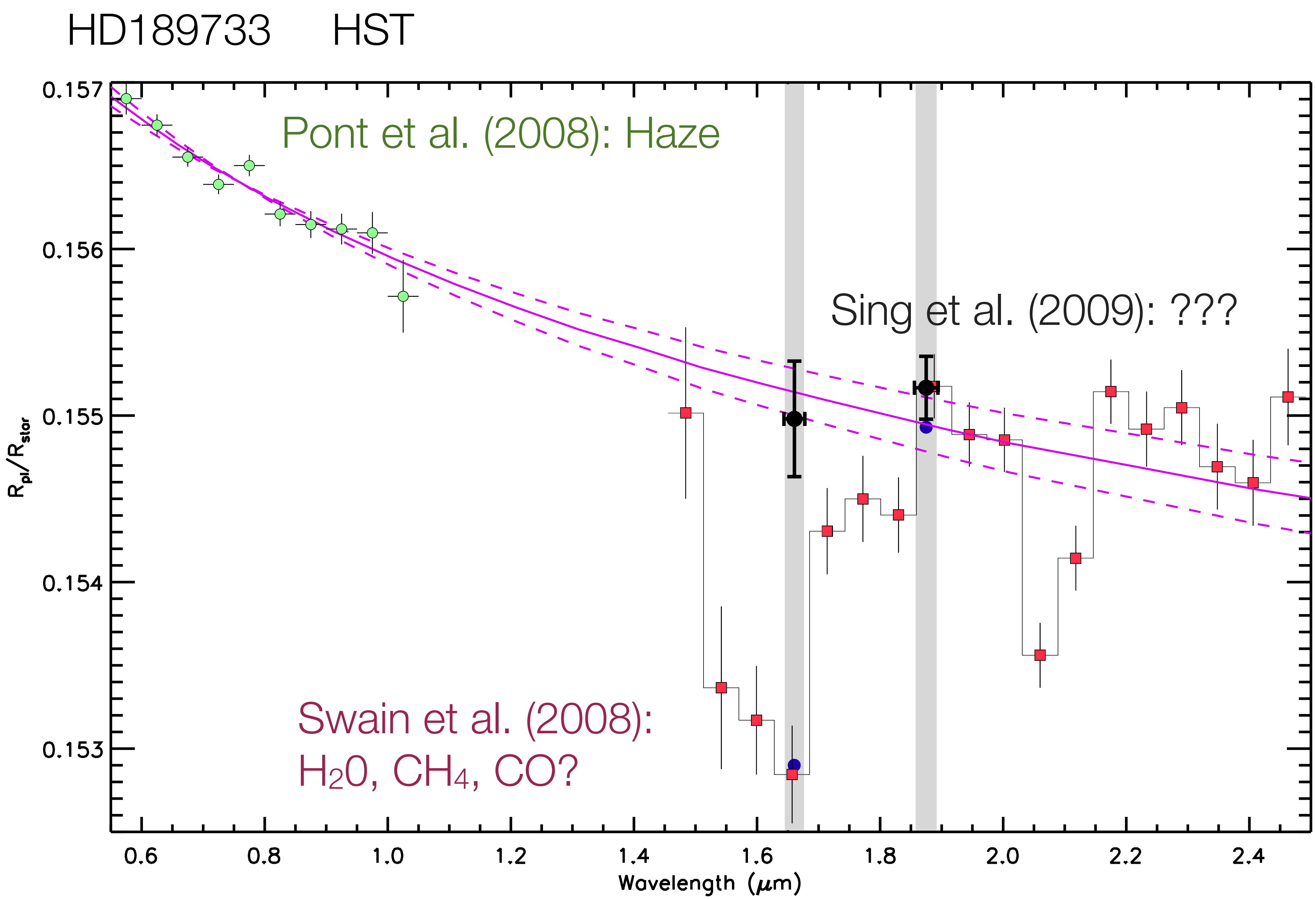
Absorption spectra of hot Jupiters



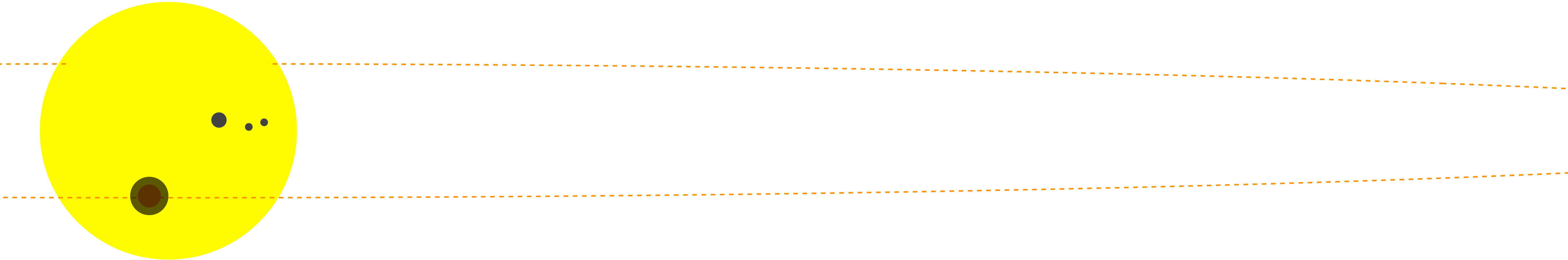


Absorption spectra of hot Jupiters

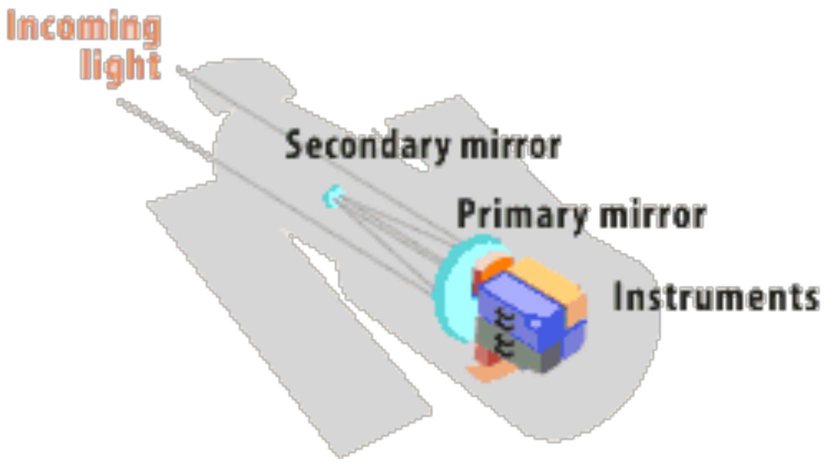




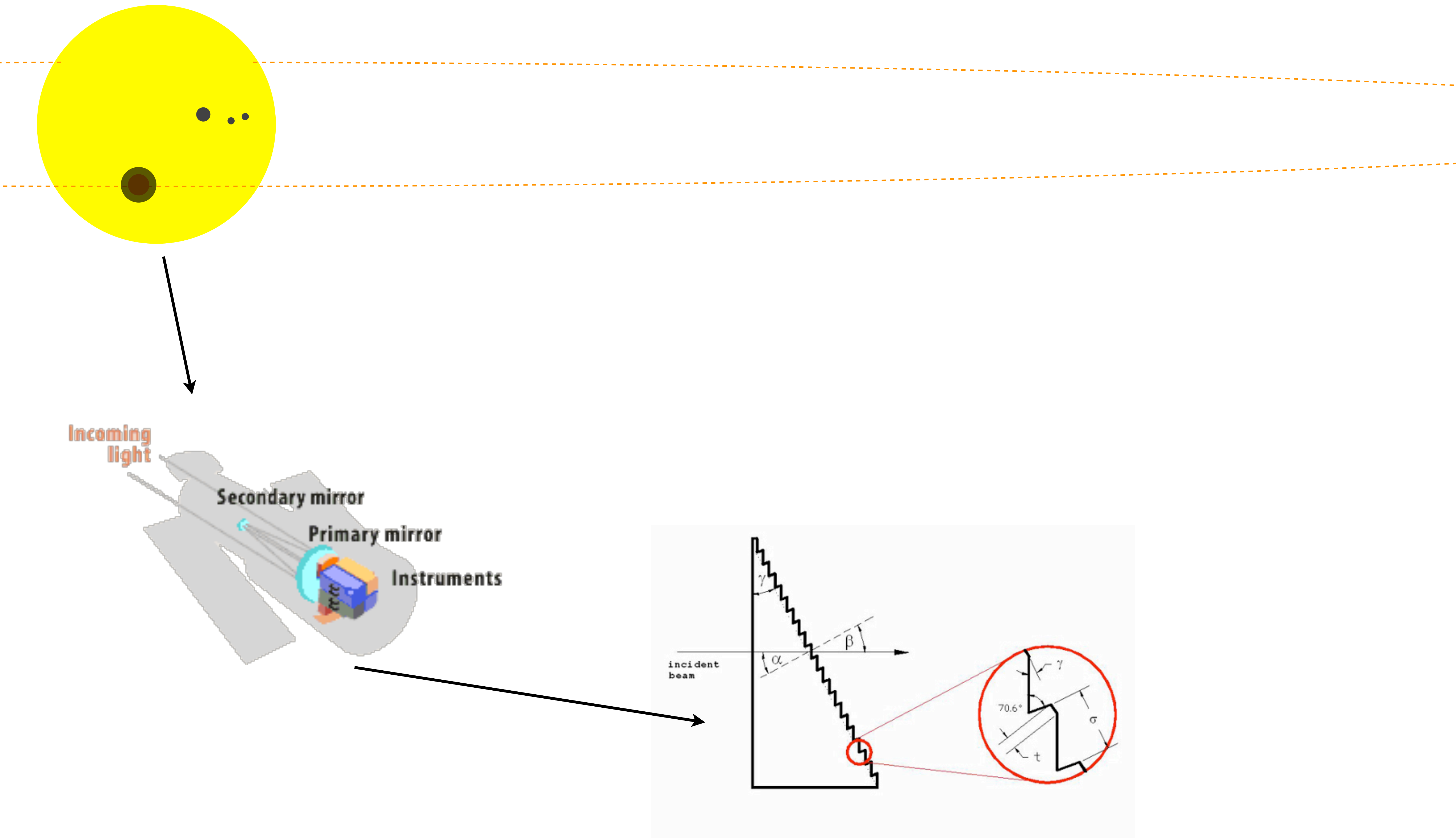
Near-IR transit observations with HST



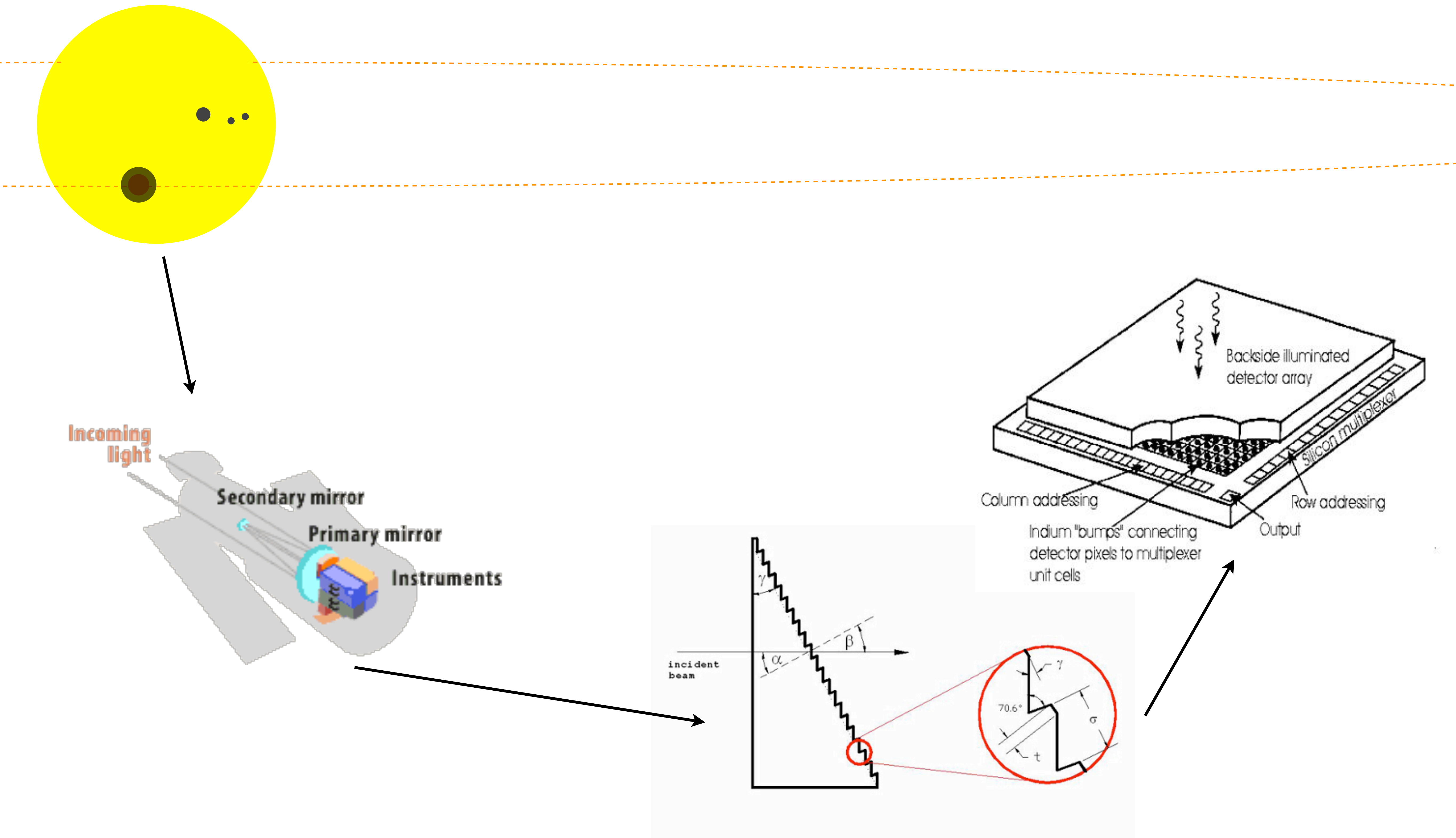
Near-IR transit observations with HST



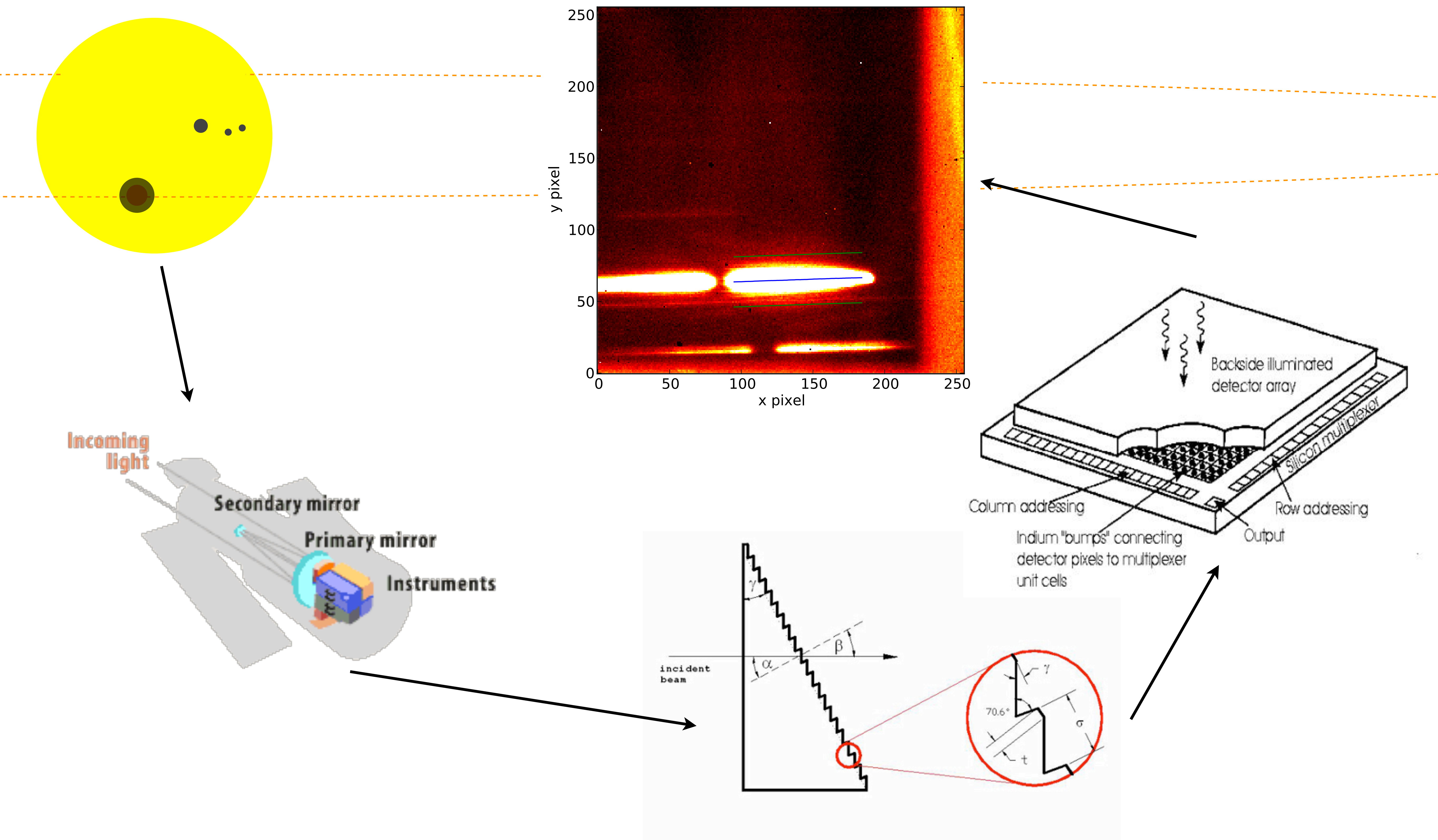
Near-IR transit observations with HST



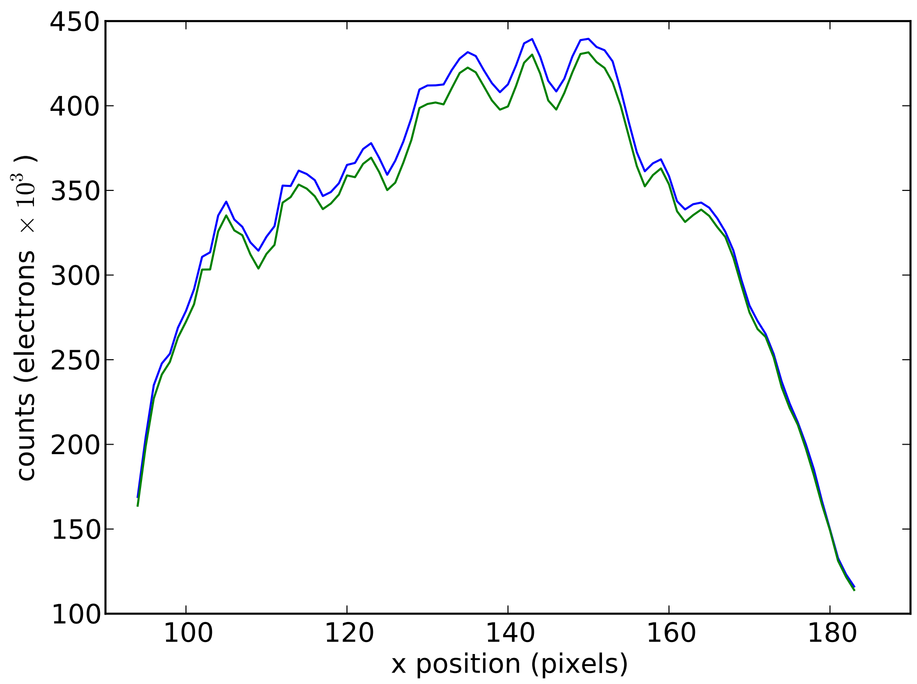
Near-IR transit observations with HST

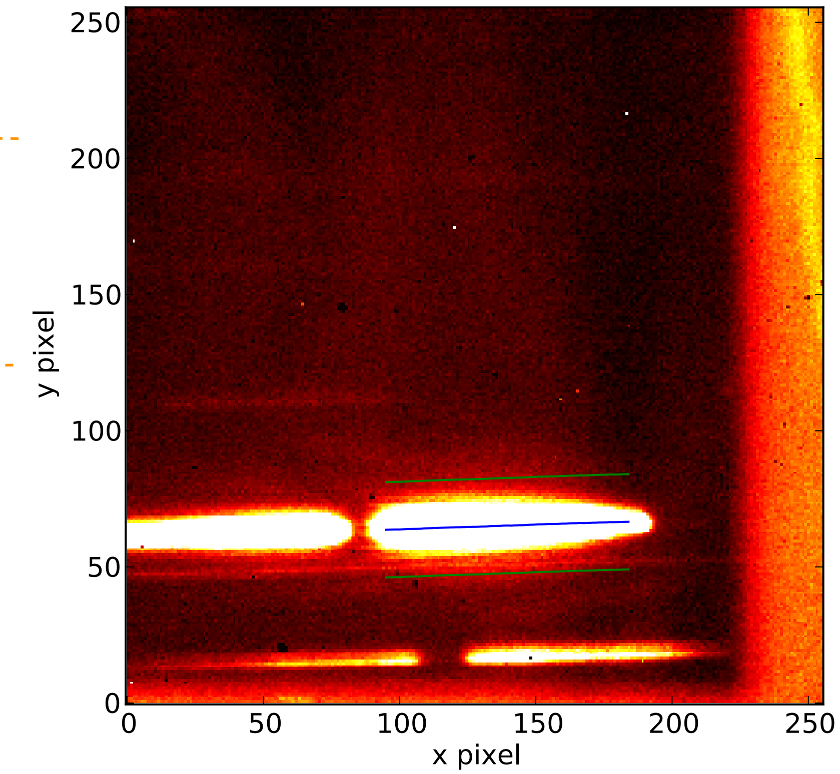


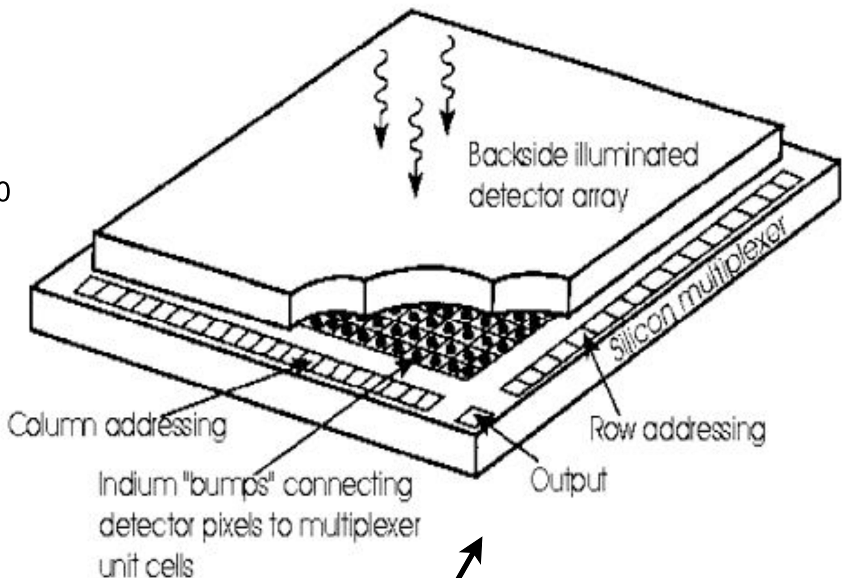
Near-IR transit observations with HST

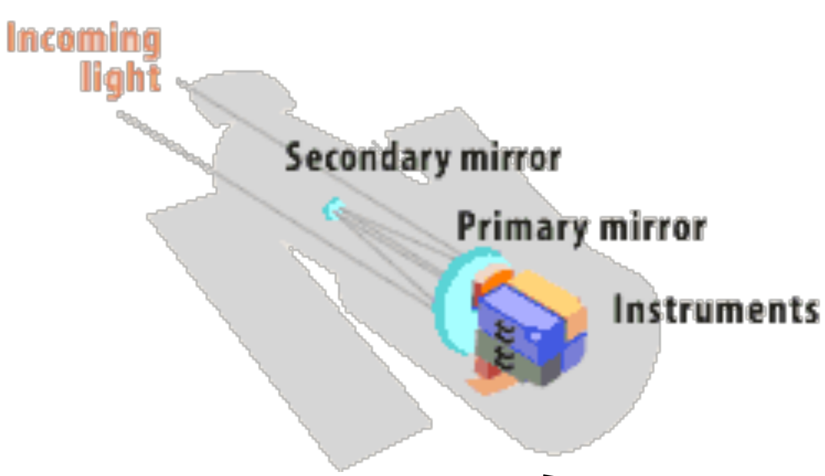


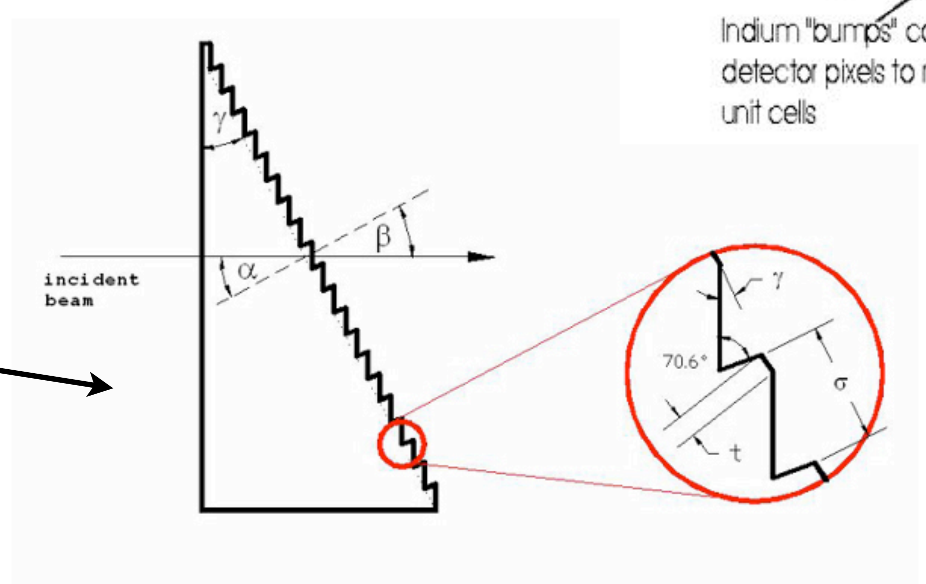
Near-IR transit observations with HST



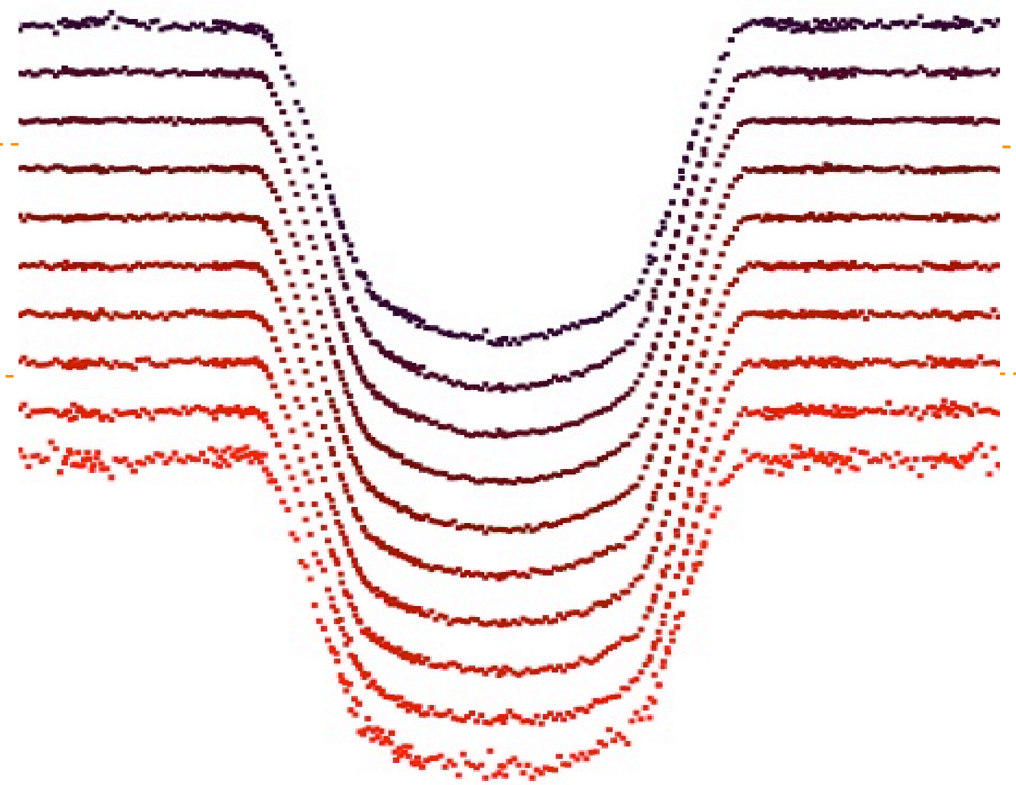


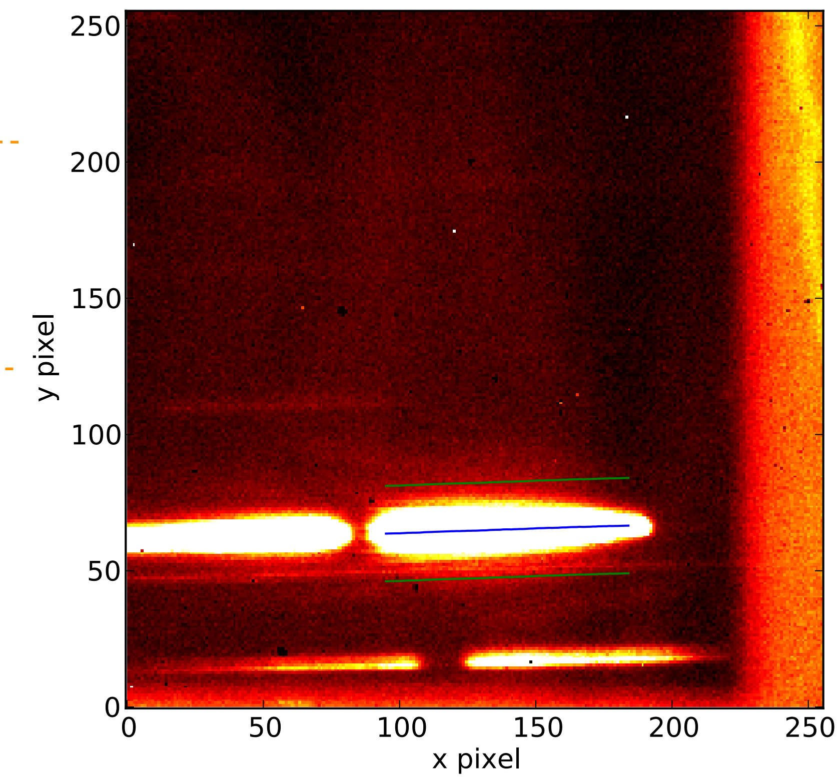


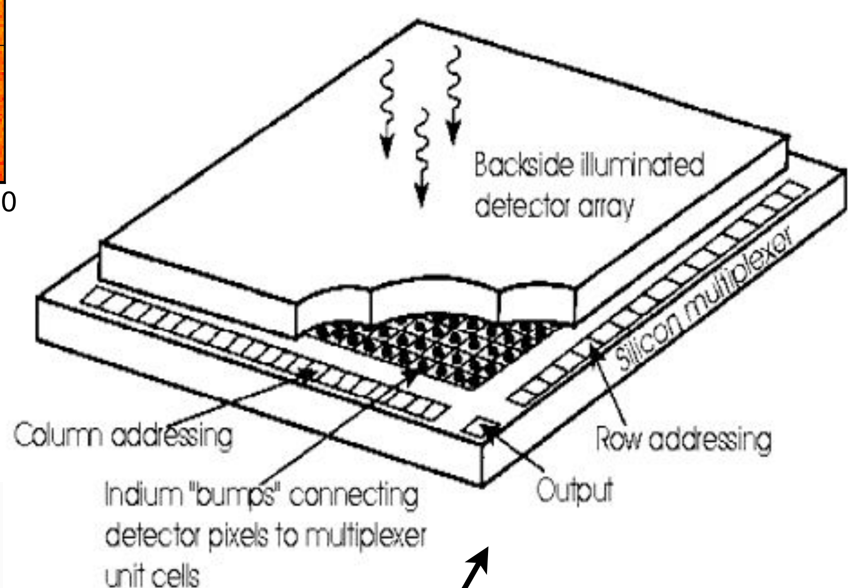


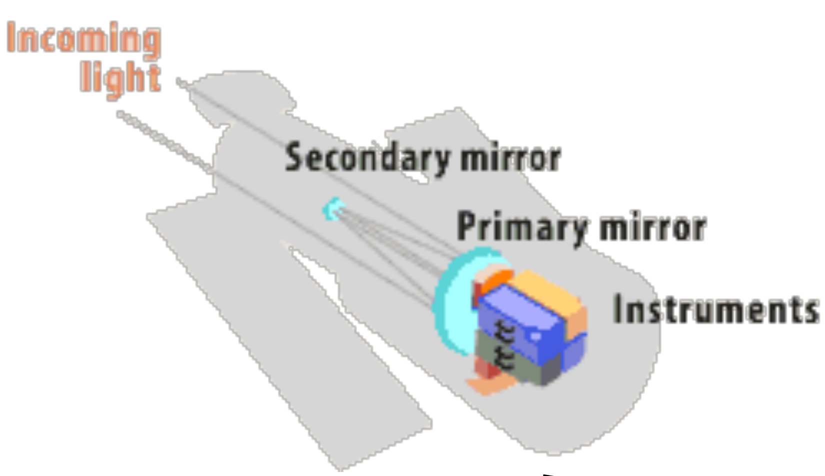


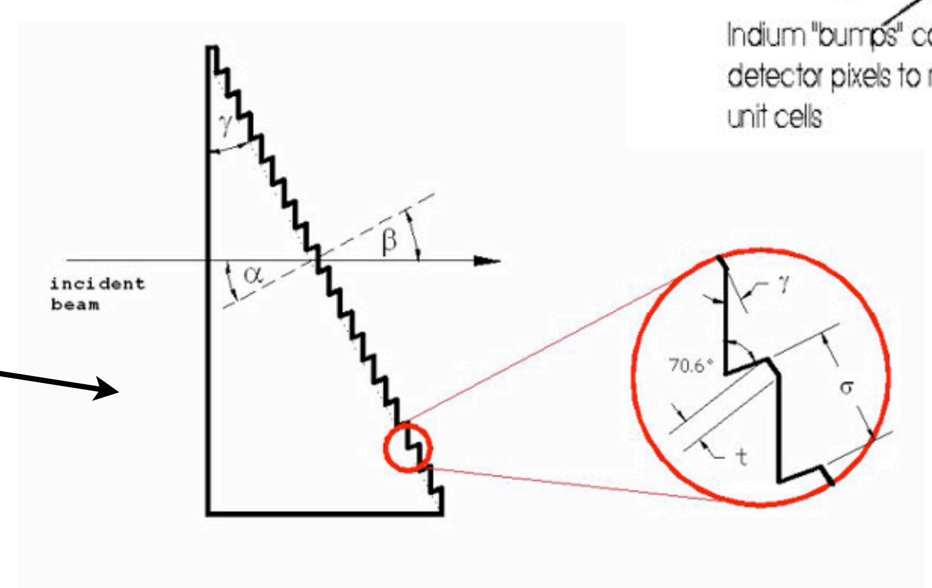
Near-IR transit observations with HST



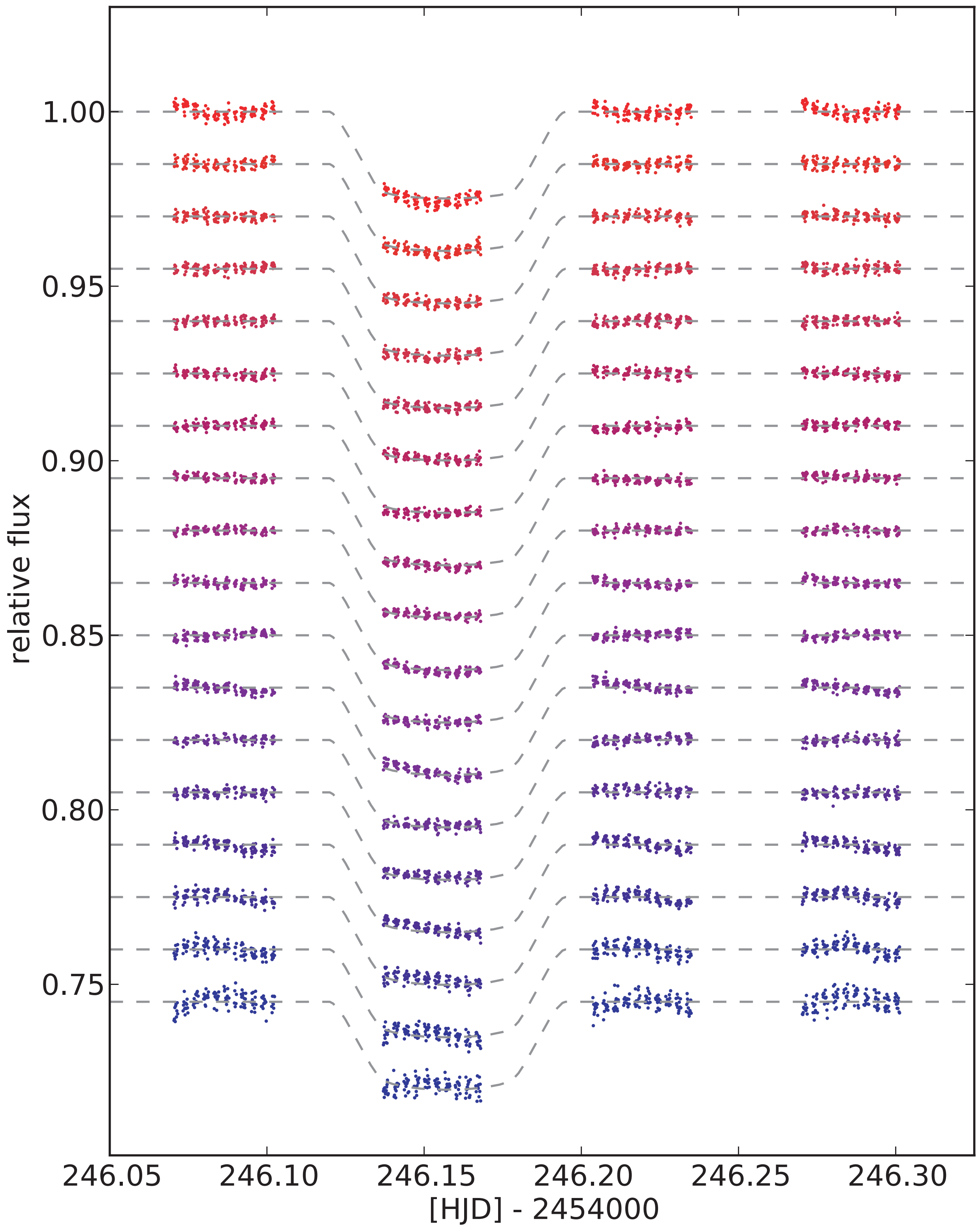




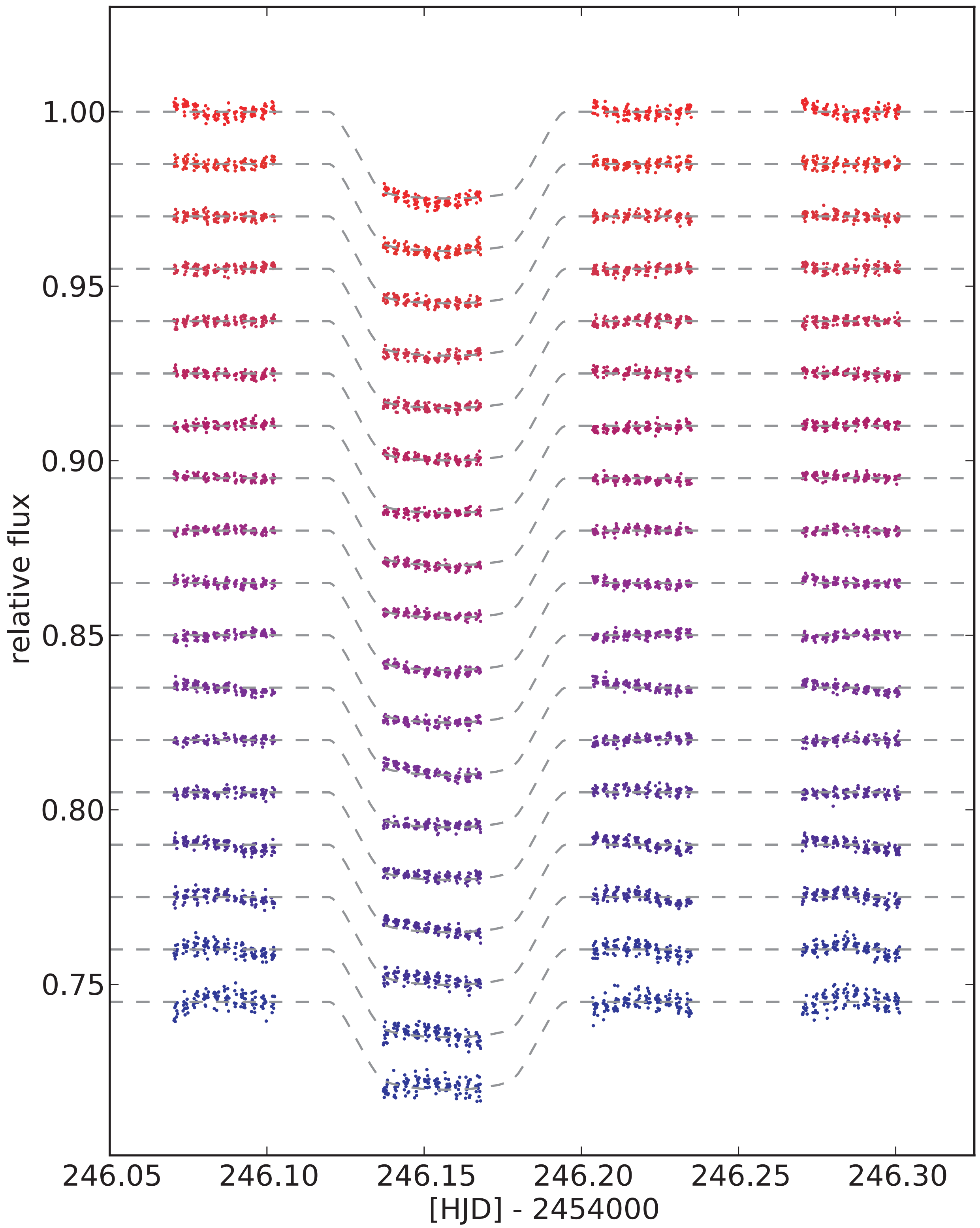


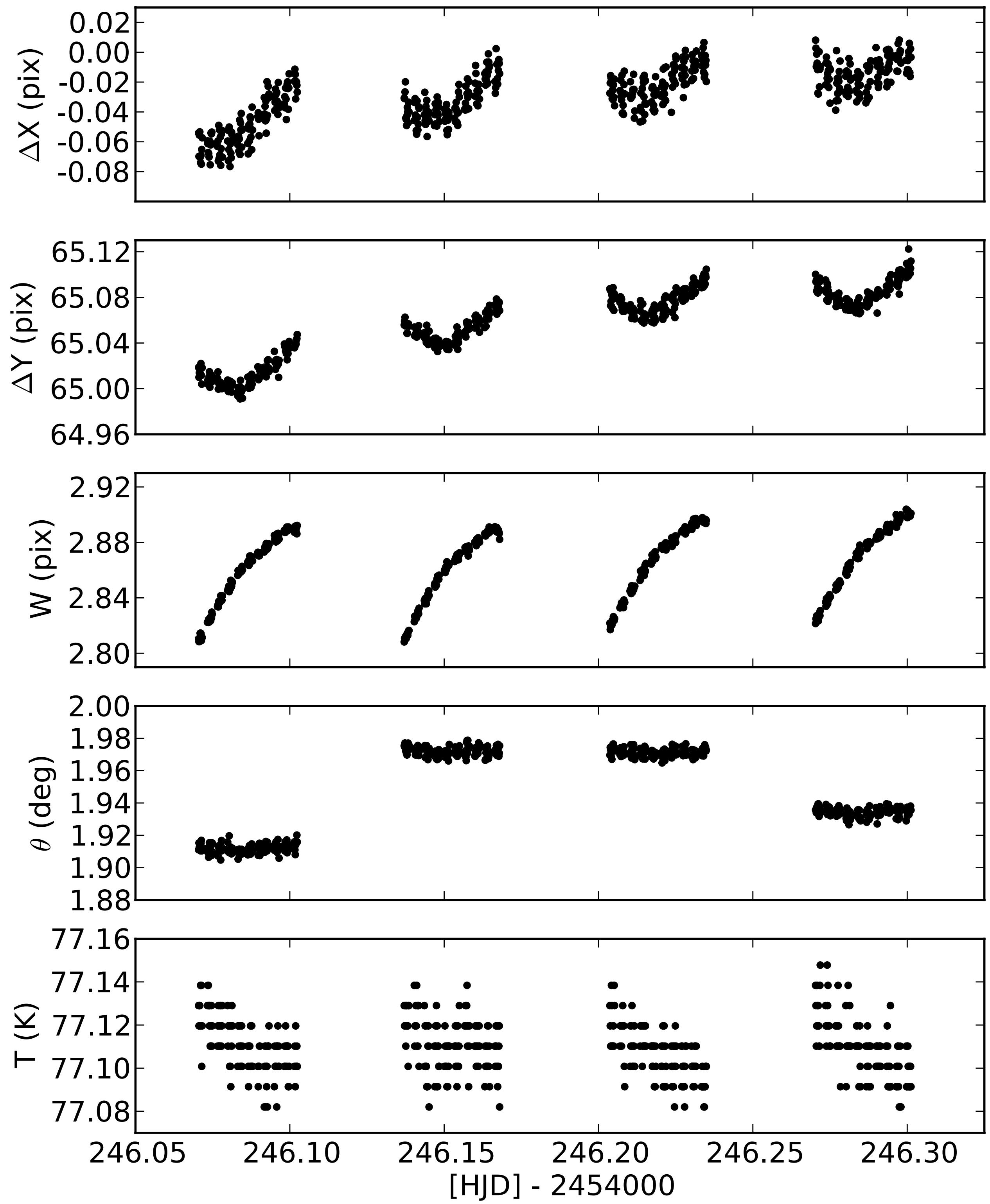


Systematics

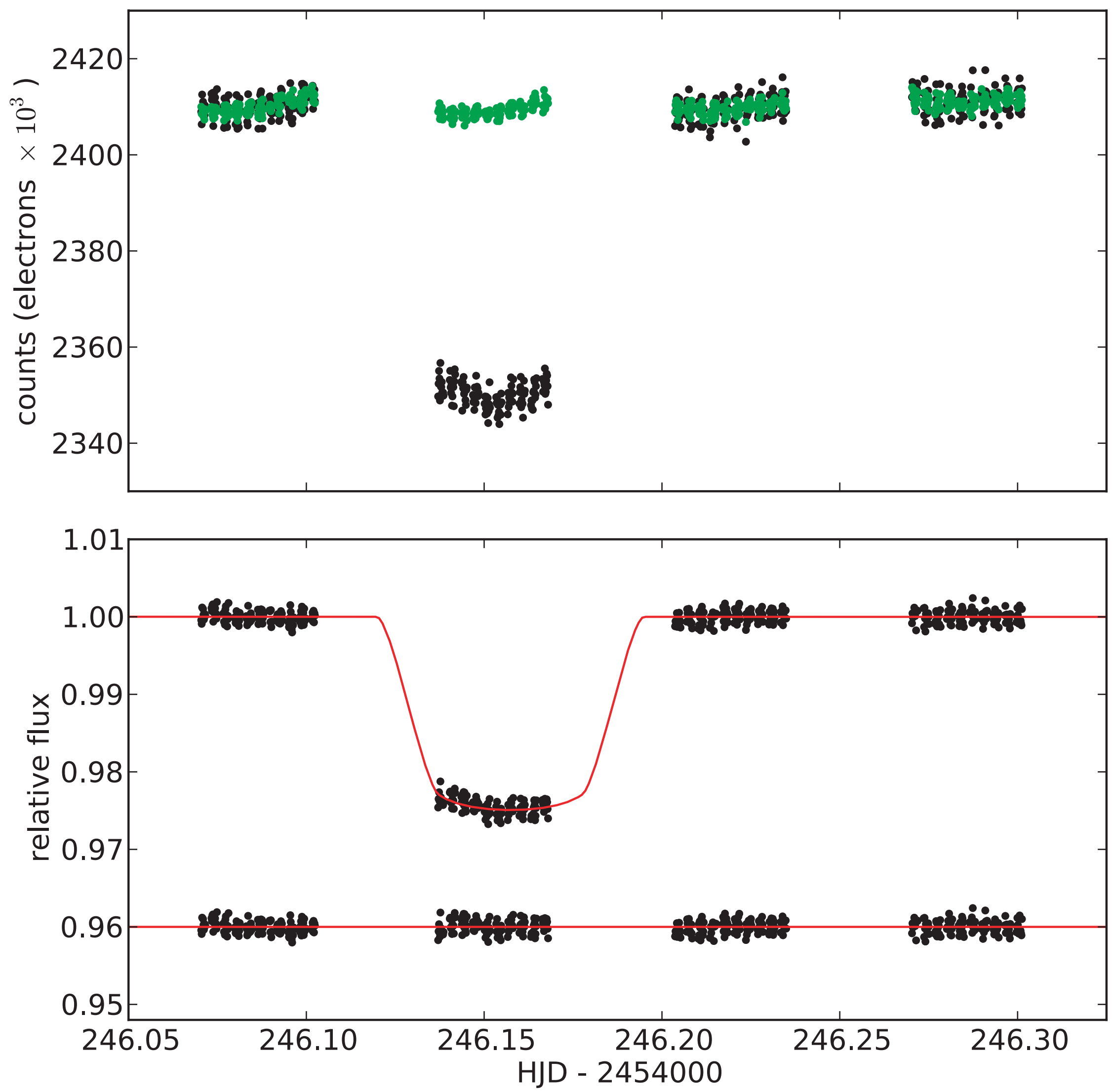


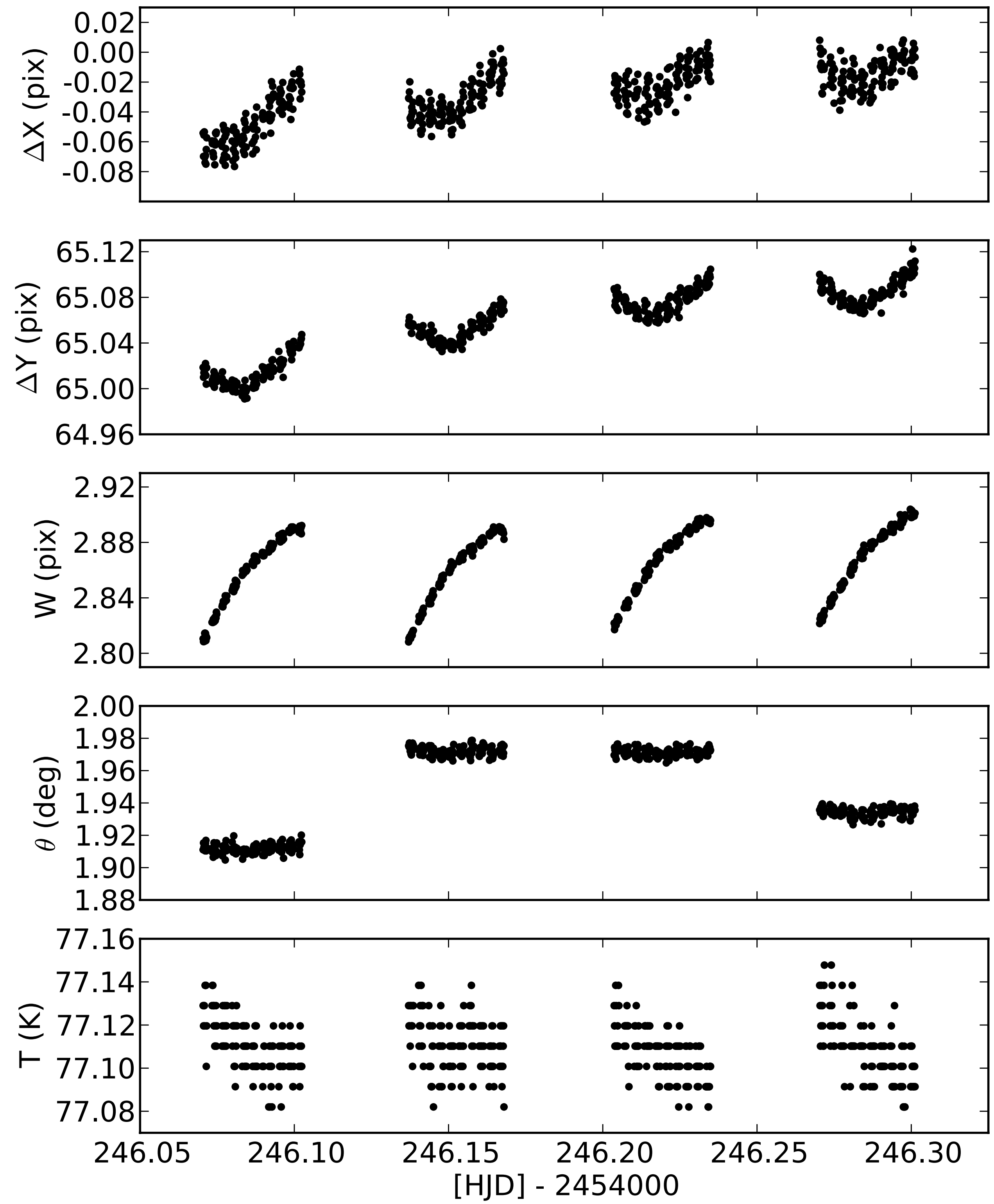
Systematics





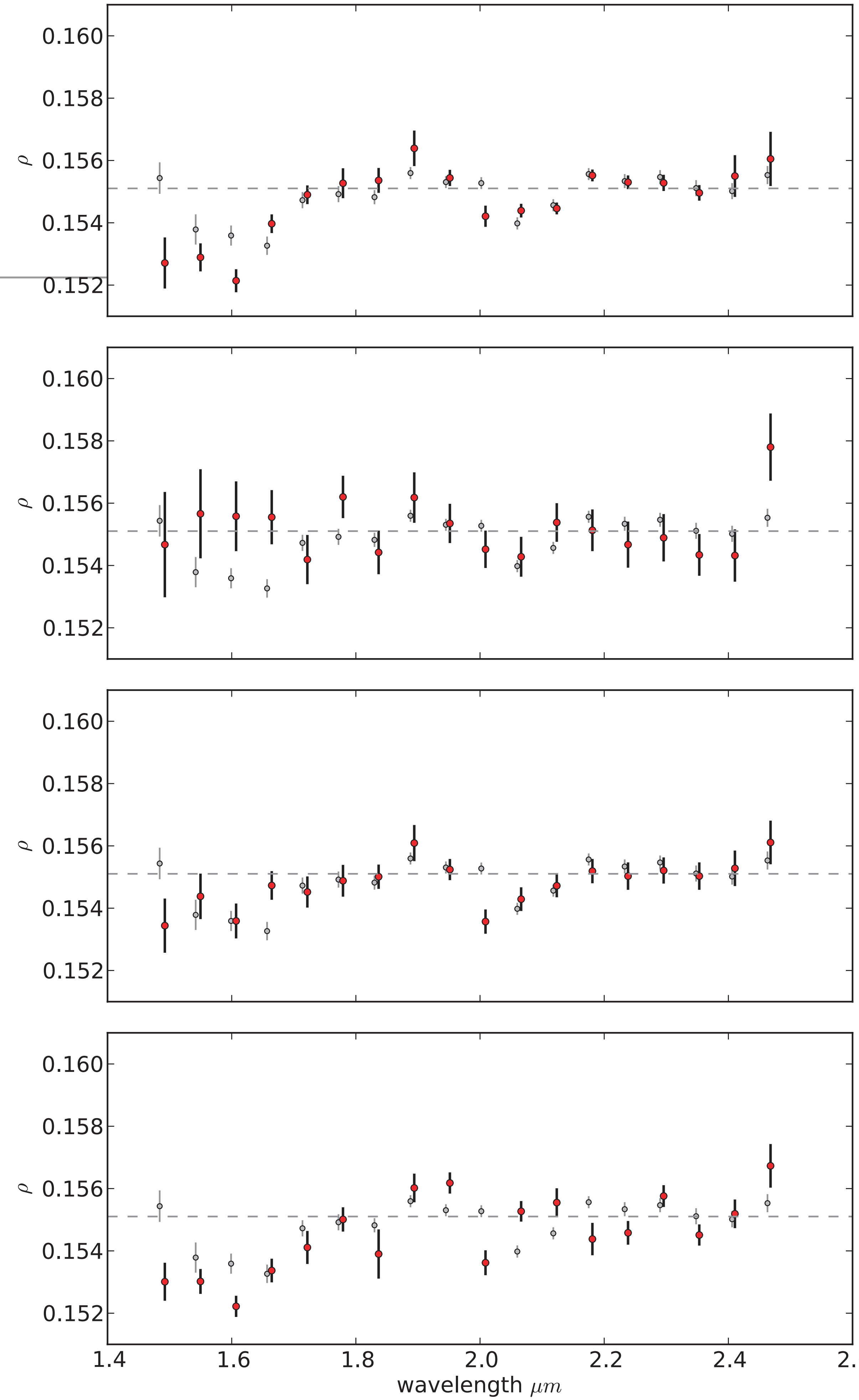
Standard procedure: linear decorrelation





Problems...

- Correlated noise remains in the residuals of the systematics correction
- Final spectrum is highly dependent on difficult choices
  - form of basis functions
  - what to include in the training data
- Uncertainty of systematics correction impacts final spectrum uncertainties
  - this is sometimes not propagated at all
  - if it is done, it is in an ad-hoc way



Gaussian processes

Gaussian processes

- $P(\boldsymbol{y} \mid \mathbf{X}, \boldsymbol{\Theta}) = \mathcal{N}(\boldsymbol{m}, \mathbf{K}), \quad \mathrm{K}_{ij} = k_{\Theta}(\mathbf{x}_i, \mathbf{x}_j)$

Gaussian processes

- $P(\boldsymbol{y} \mid \mathbf{X}, \mathbf{\Theta}) = \mathcal{N}(\boldsymbol{m}, \mathbf{K})$, $\mathrm{K}_{ij} = k_{\Theta}(\mathbf{x}_i, \mathbf{x}_j)$
- Probability distribution over random functions
  - use the data to learn the dependence of the systematics on the external parameters

Gaussian processes

- $P(\boldsymbol{y} \mid \mathbf{X}, \mathbf{\Theta}) = \mathcal{N}(\boldsymbol{m}, \mathbf{K}), \quad \mathrm{K}_{ij} = k_{\Theta}(\mathbf{x}_i, \mathbf{x}_j)$
- Probability distribution over random functions
  - use the data to learn the dependence of the systematics on the external parameters
- Bayesian
  - built-in complexity penalty
  - natural propagation of uncertainties to final spectrum

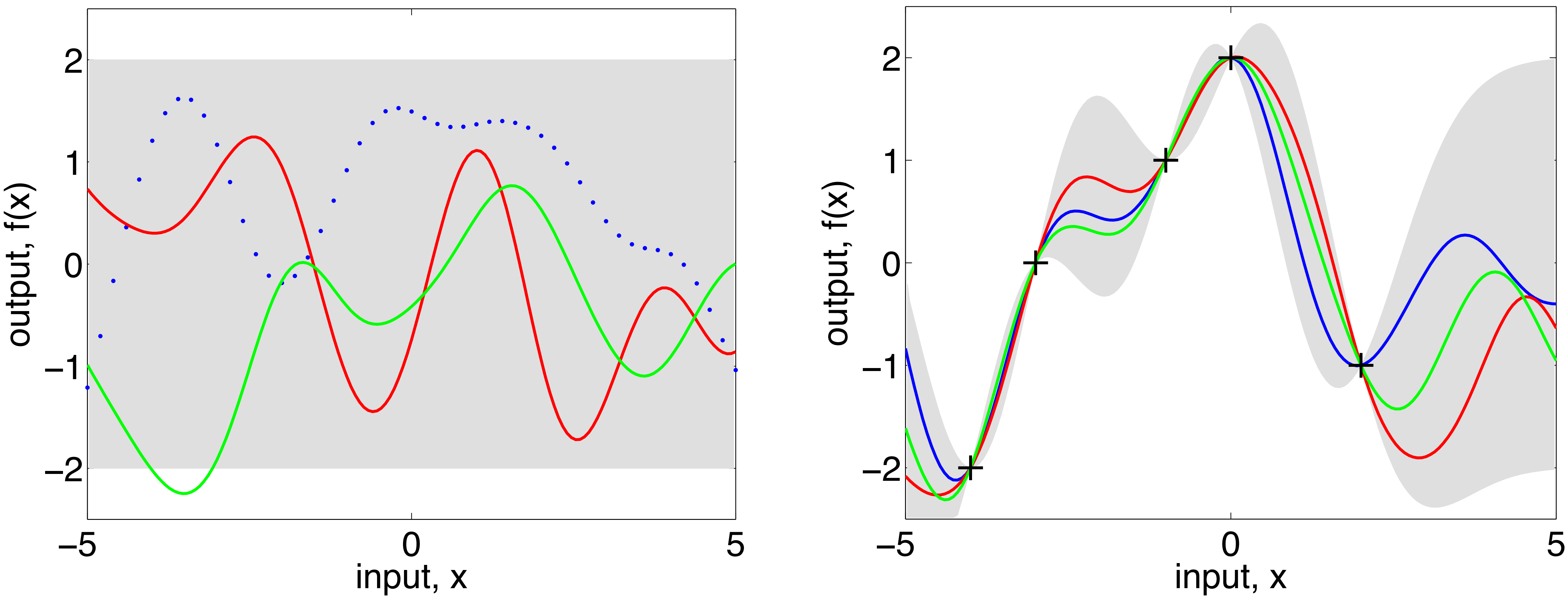
Gaussian processes

- $P(\boldsymbol{y} \mid \mathbf{X}, \boldsymbol{\Theta}) = \mathcal{N}(\boldsymbol{m}, \mathbf{K})$, $\quad \mathrm{K}_{ij} = k_{\Theta}(\mathbf{x}_i, \mathbf{x}_j)$
- Probability distribution over random functions
  - use the data to learn the dependence of the systematics on the external parameters
- Bayesian
  - built-in complexity penalty
  - natural propagation of uncertainties to final spectrum
- Simultaneous modelling of systematics and transit function
  - make optimal use of available data

Gaussian process regression

For a specific prior (k_Θ), and dataset, obtain
a predictive distribution for any new data point
a marginal likelihood (expresses goodness of fit of the regression)

Gaussian process regression



(a), prior (b), posterior

For a specific prior (k_Θ), and dataset, obtain

a predictive distribution for any new data point

a marginal likelihood (expresses goodness of fit of the regression)

GPs for transmission spectra

GPs for transmission spectra

- $P(\mathbf{y} \mid \mathbf{X}, \boldsymbol{r}, \boldsymbol{\Theta}) \sim \mathcal{N}(\boldsymbol{m}, \mathbf{K})$
  - targets $\boldsymbol{y}$ = array of relative flux measurements
  - inputs $\mathbf{X}$ = {time, pointing, satellite orbital phase, detector temperature, wavelength...}
  - mean function $\boldsymbol{m} = f_{\mathrm{transit}}\,(t, r_\lambda, \boldsymbol{\Theta}_{\mathrm{transit}})$
  - $K_{ij} = \Sigma_l\, k(X_{li}, X_{lj}, \boldsymbol{\Theta}_{\mathrm{covariance}})$

GPs for transmission spectra

- $P(\mathbf{y} \mid \mathbf{X}, \boldsymbol{r}, \boldsymbol{\Theta}) \sim \mathcal{N}(\boldsymbol{m}, \mathbf{K})$
  - targets $\boldsymbol{y}$ = array of relative flux measurements
  - inputs $\mathbf{X}$ = {time, pointing, satellite orbital phase, detector temperature, wavelength...}
  - mean function $\boldsymbol{m} = f_{\mathrm{transit}}\,(t, r_\lambda, \boldsymbol{\Theta}_{\mathrm{transit}})$
  - $K_{ij} = \Sigma_l\, k(X_{li}, X_{lj}, \boldsymbol{\Theta}_{\mathrm{covariance}})$
- We want to measure $\boldsymbol{r}$ (wavelength-dependent planet-to-star radius ratio)
  - marginalise over $\boldsymbol{\Theta}$ (all other hyper-parameters)

GPs for transmission spectra

- $P(\mathbf{y} \mid \mathbf{X}, \boldsymbol{r}, \boldsymbol{\Theta}) \sim \mathcal{N}(\boldsymbol{m}, \mathbf{K})$
  - targets $\boldsymbol{y}$ = array of relative flux measurements
  - inputs $\mathbf{X}$ = {time, pointing, satellite orbital phase, detector temperature, wavelength...} use only orbital phase and pointing
  - mean function $\boldsymbol{m} = f_{\mathrm{transit}}\,(t, r_{\lambda}, \boldsymbol{\Theta}_{\mathrm{transit}})$
  - $K_{ij} = \Sigma_l\, k(X_{li}, X_{lj}, \boldsymbol{\Theta}_{\mathrm{covariance}})$ use squared exponential
- We want to measure $\boldsymbol{r}$ (wavelength-dependent planet-to-star radius ratio)
  - marginalise over $\boldsymbol{\Theta}$ (all other hyper-parameters) use MCMC (adaptive sampling) and a fast machine

Gaussian process regression

work with residuals of mean function

input: X (inputs), $\mathbf{y}$ (targets), k (covariance function), σ_n^2 (noise level), $\mathbf{x}_*$ (test input)

2: $L := \text{cholesky}(K + \sigma_n^2 I)$

$\boldsymbol{\alpha} := L^\top \backslash (L \backslash \mathbf{y})$

4: $\bar{f}_* := \mathbf{k}_*^\top \boldsymbol{\alpha}$ $\Big\}$ predictive mean eq. (2.25)

$\mathbf{v} := L \backslash \mathbf{k}_*$

6: $\mathbb{V}[f_*] := k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{v}^\top \mathbf{v}$ $\Big\}$ predictive variance eq. (2.26)

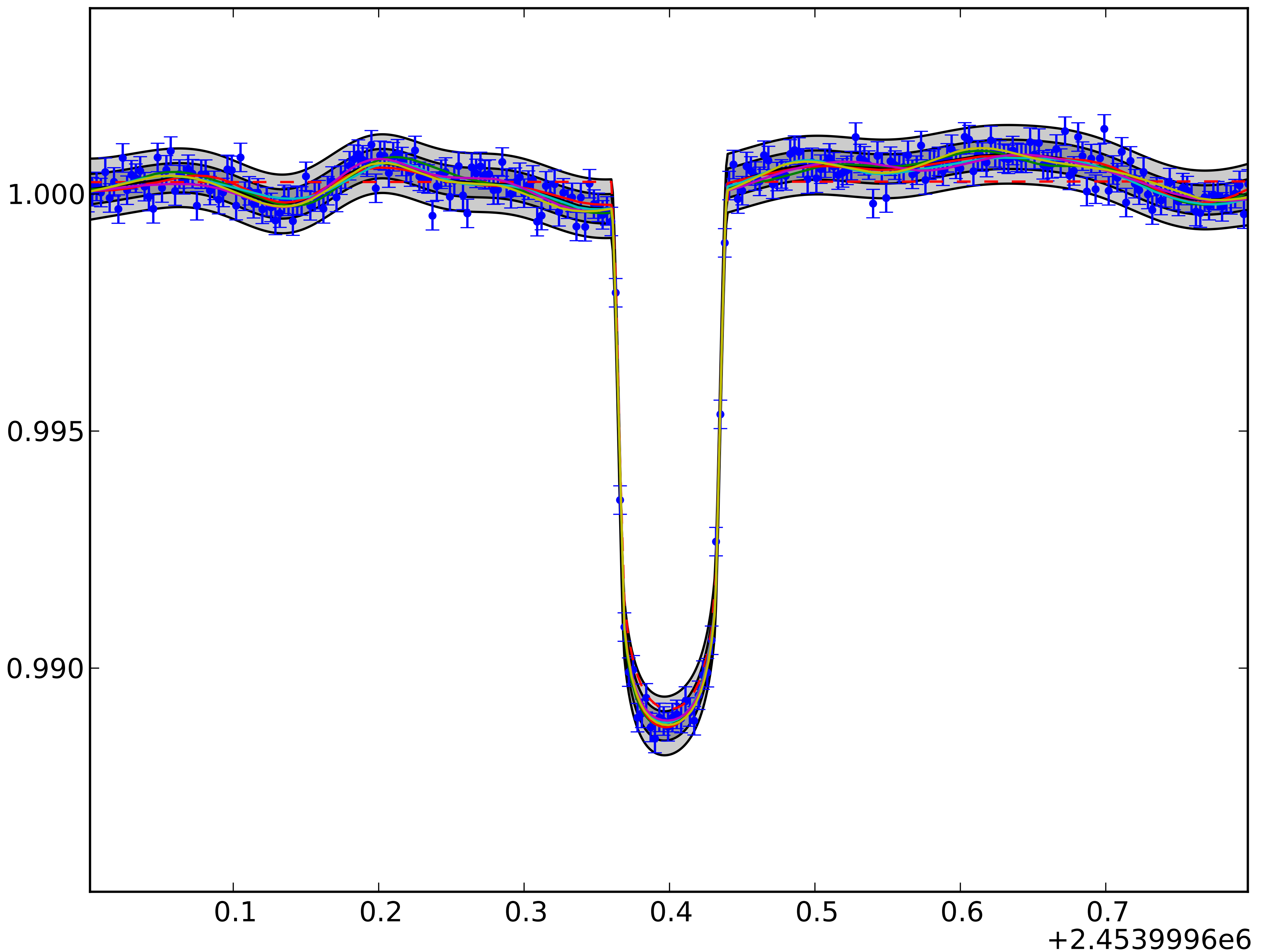
$\log p(\mathbf{y}|X) := -\frac{1}{2}\mathbf{y}^\top \boldsymbol{\alpha} - \sum_i \log L_{ii} - \frac{n}{2}\log 2\pi$ eq. (2.30)

8: **return**: $\bar{f}_*$ (mean), $\mathbb{V}[f_*]$ (variance), $\log p(\mathbf{y}|X)$ (log marginal likelihood)

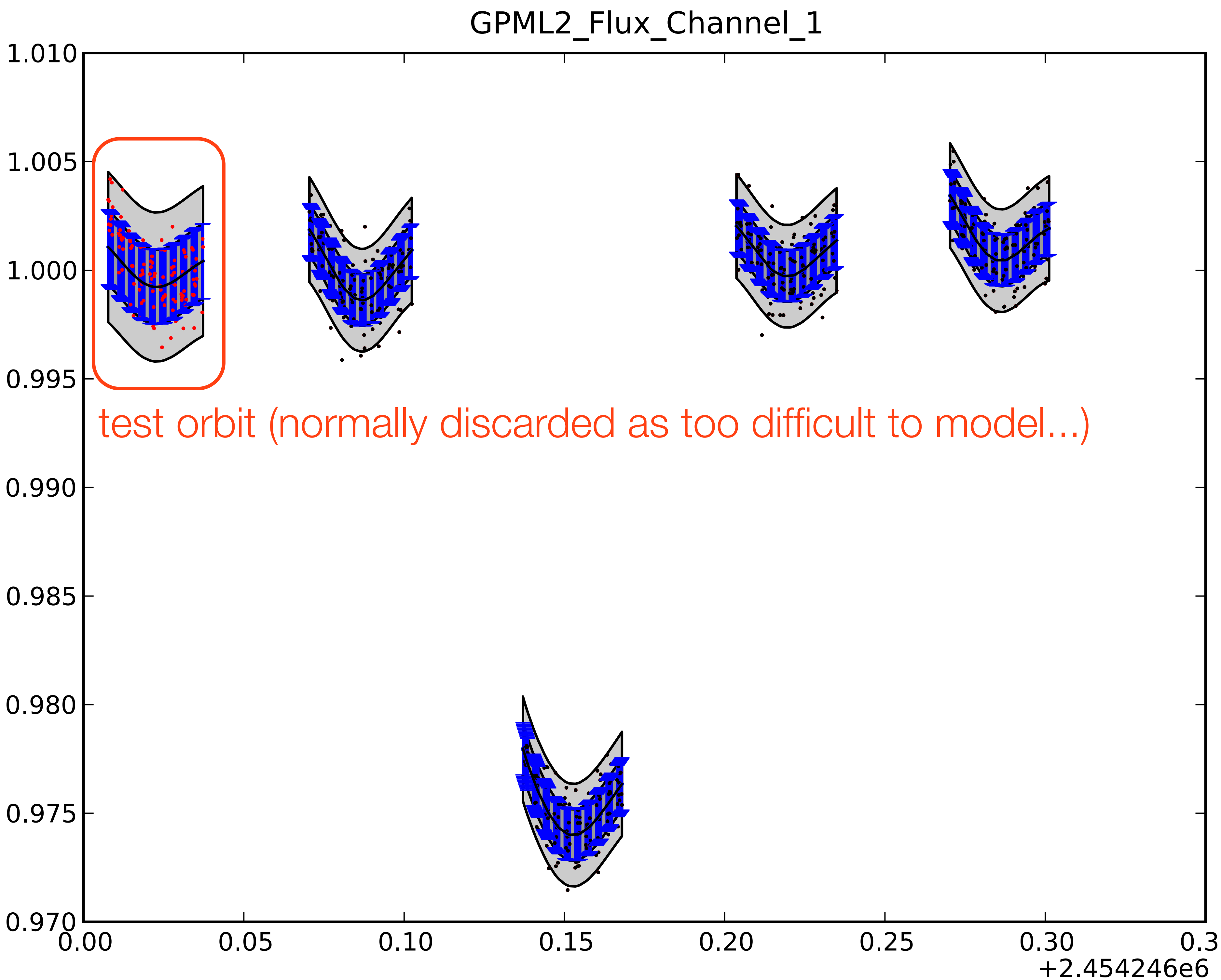
Algorithm 2.1: Predictions and log marginal likelihood for Gaussian process regression. The implementation addresses the matrix inversion required by eq. (2.25) and (2.26) using Cholesky factorization, see section A.4. For multiple test cases lines 4-6 are repeated. The log determinant required in eq. (2.30) is computed from the Cholesky factor (for large n it may not be possible to represent the determinant itself). The computational complexity is $n^3/6$ for the Cholesky decomposition in line 2, and $n^2/2$ for solving triangular systems in line 3 and (for each test case) in line 5.

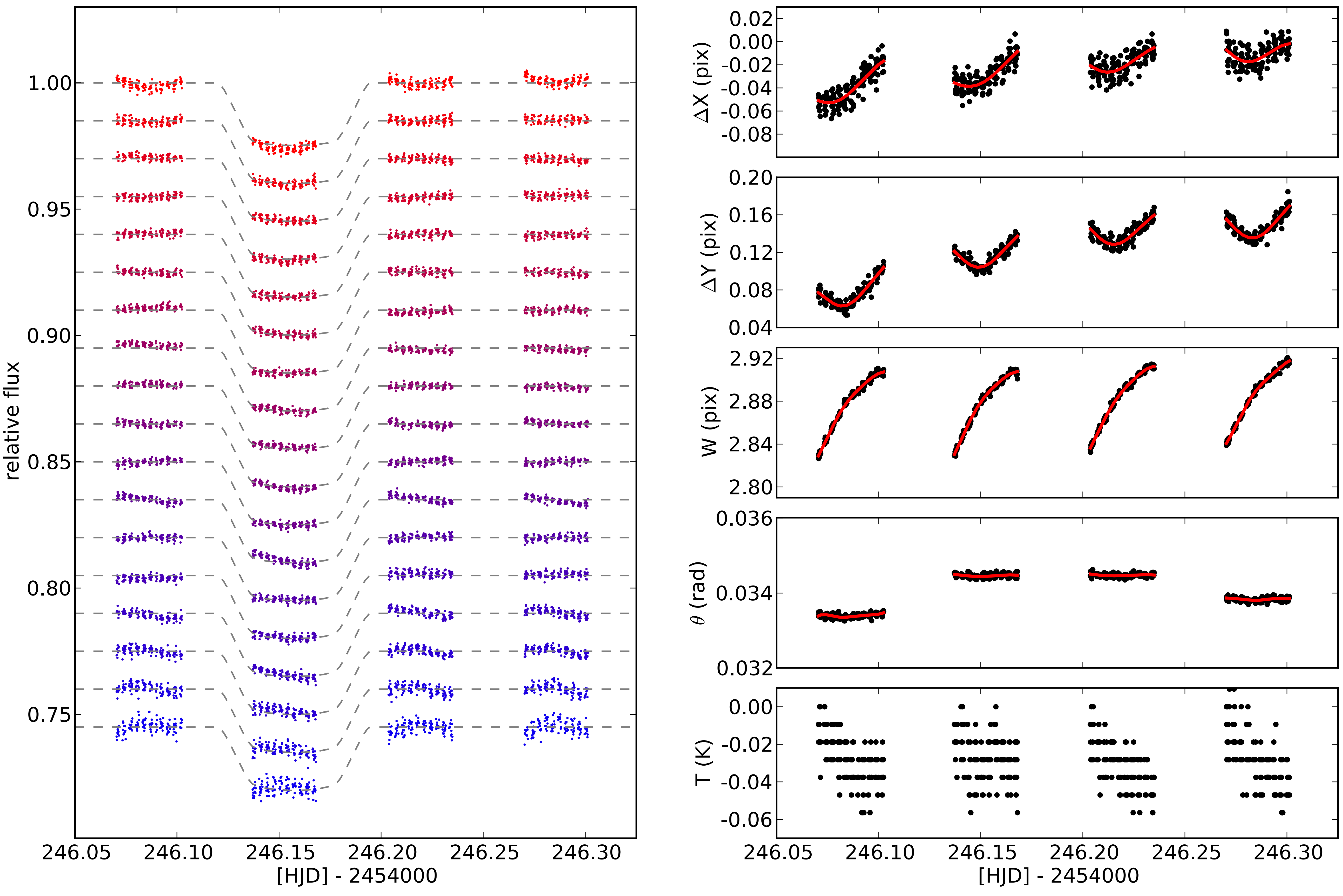
Rasmussen & Williams (2009)

Simulated example



Application to NICMOS HD189733b data




relative flux
[HJD] - 2454000
ΔX (pix)
ΔY (pix)
W (pix)
θ (rad)
T (K)
[HJD] - 2454000

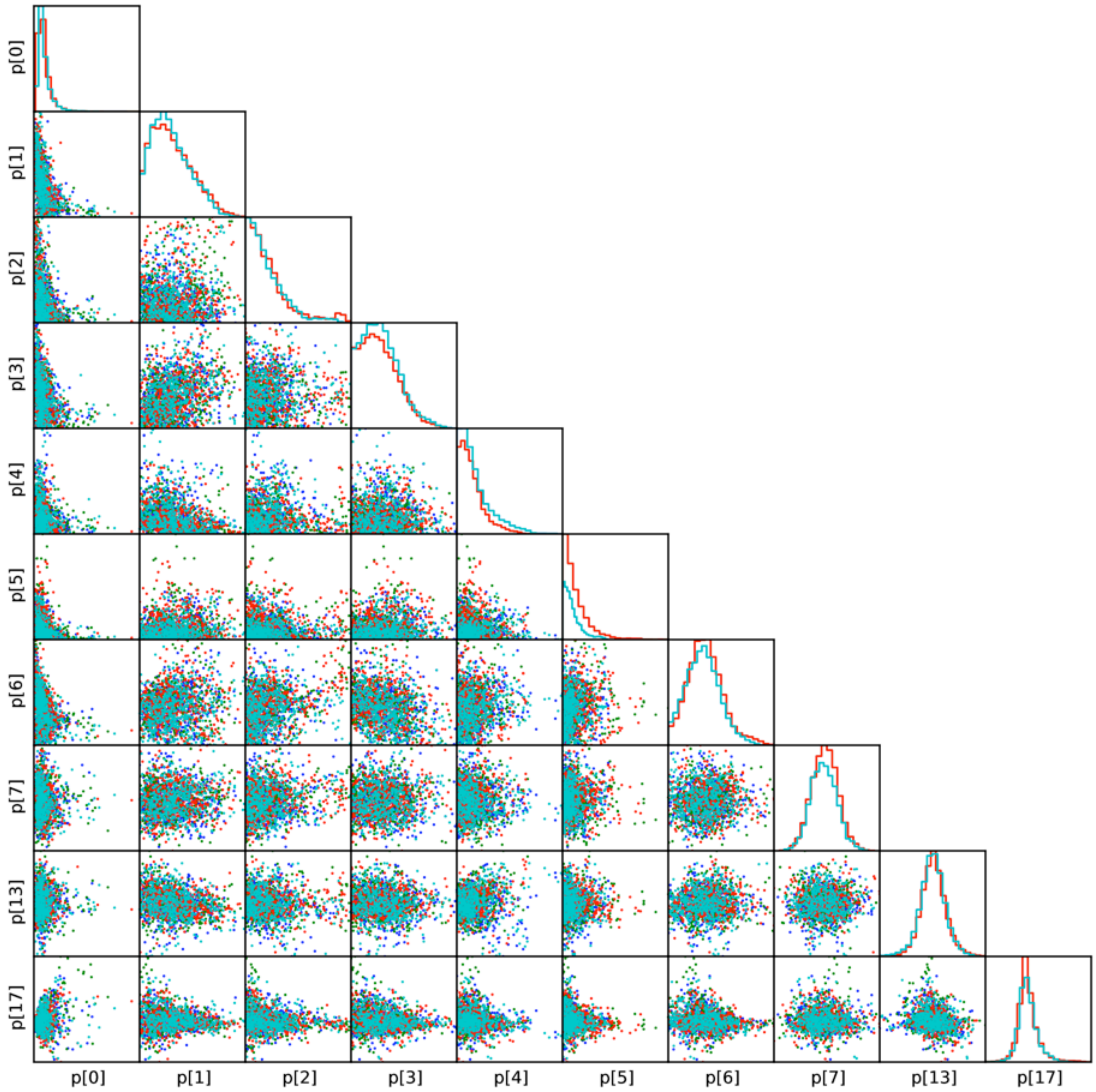
overall covariance

covariance
hyperparameters

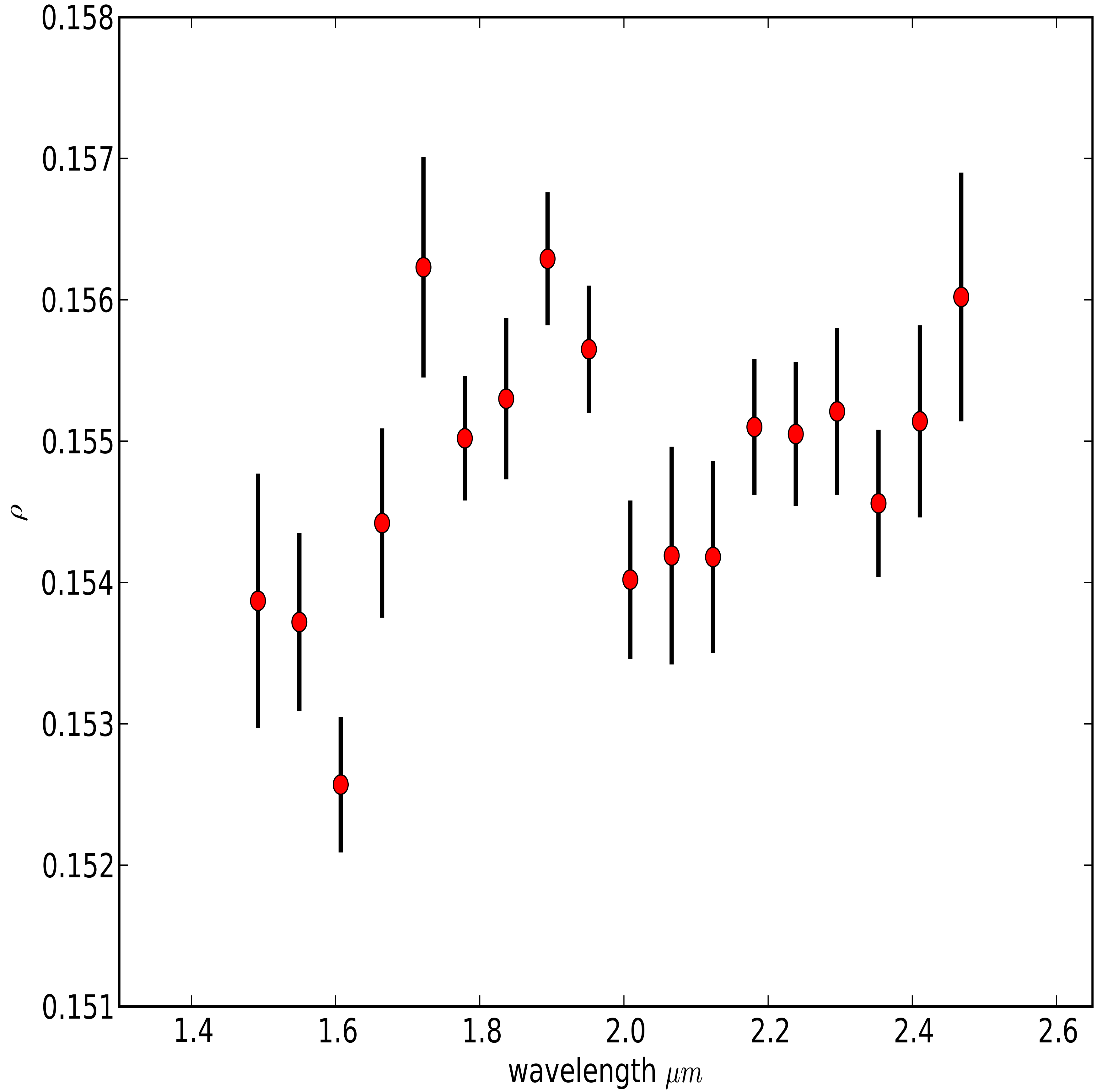
white noise variance

radius ratio

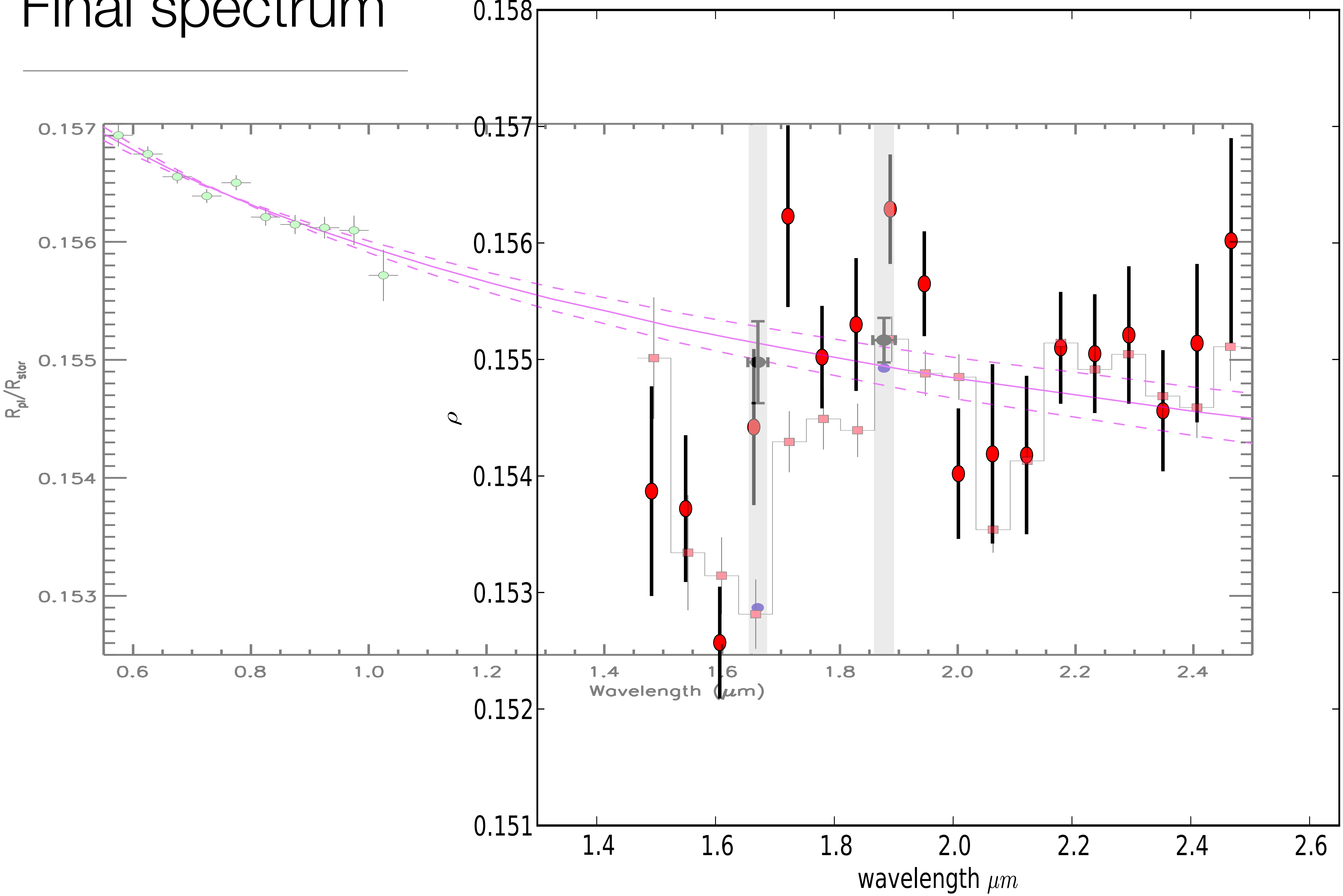
zero point



Final spectrum



Final spectrum

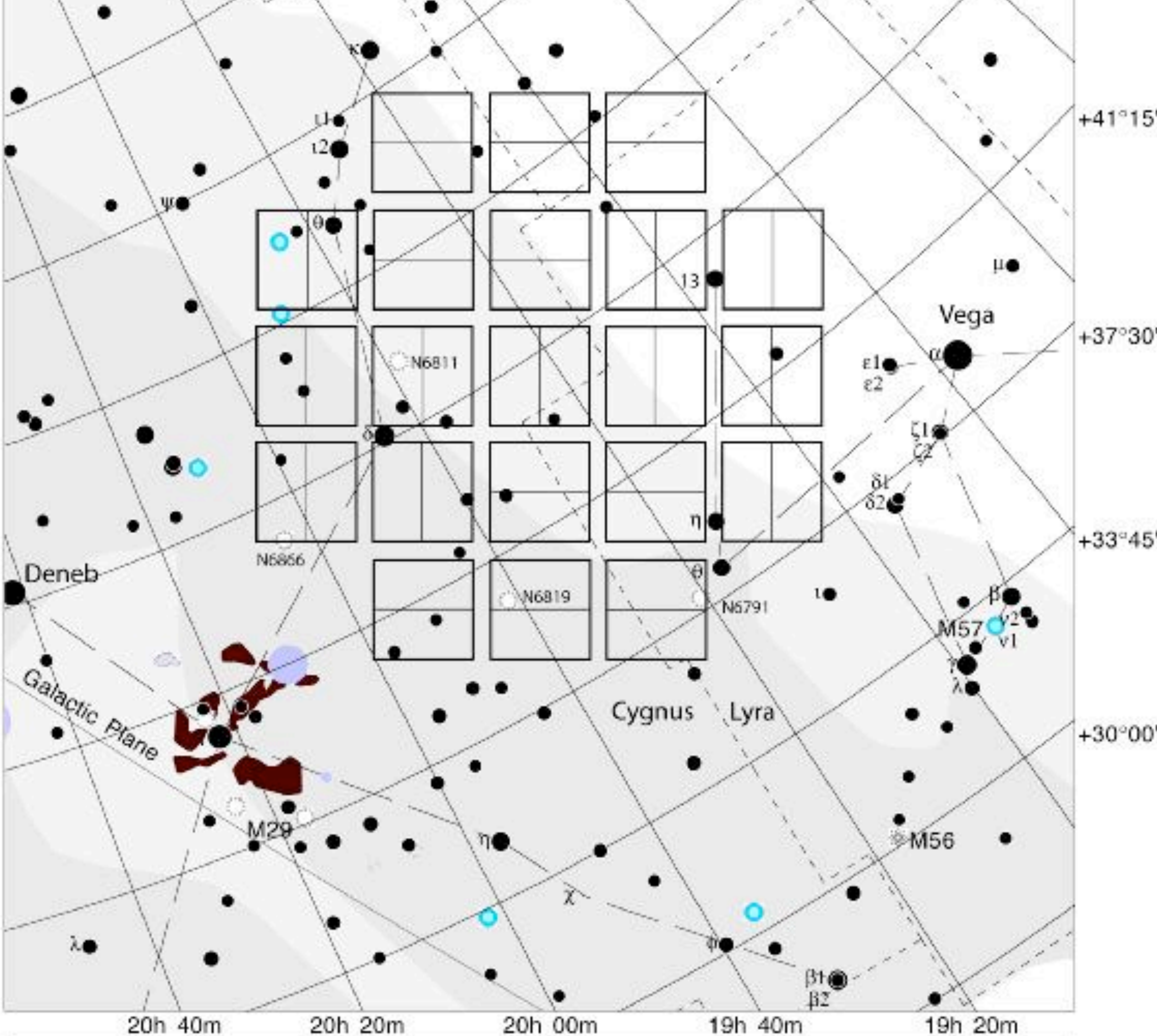


An entropy criterion for systematic trend discovery in large ensembles of time-series

Roberts et al (in prep.), McQuillan et al. (in prep)

Kepler

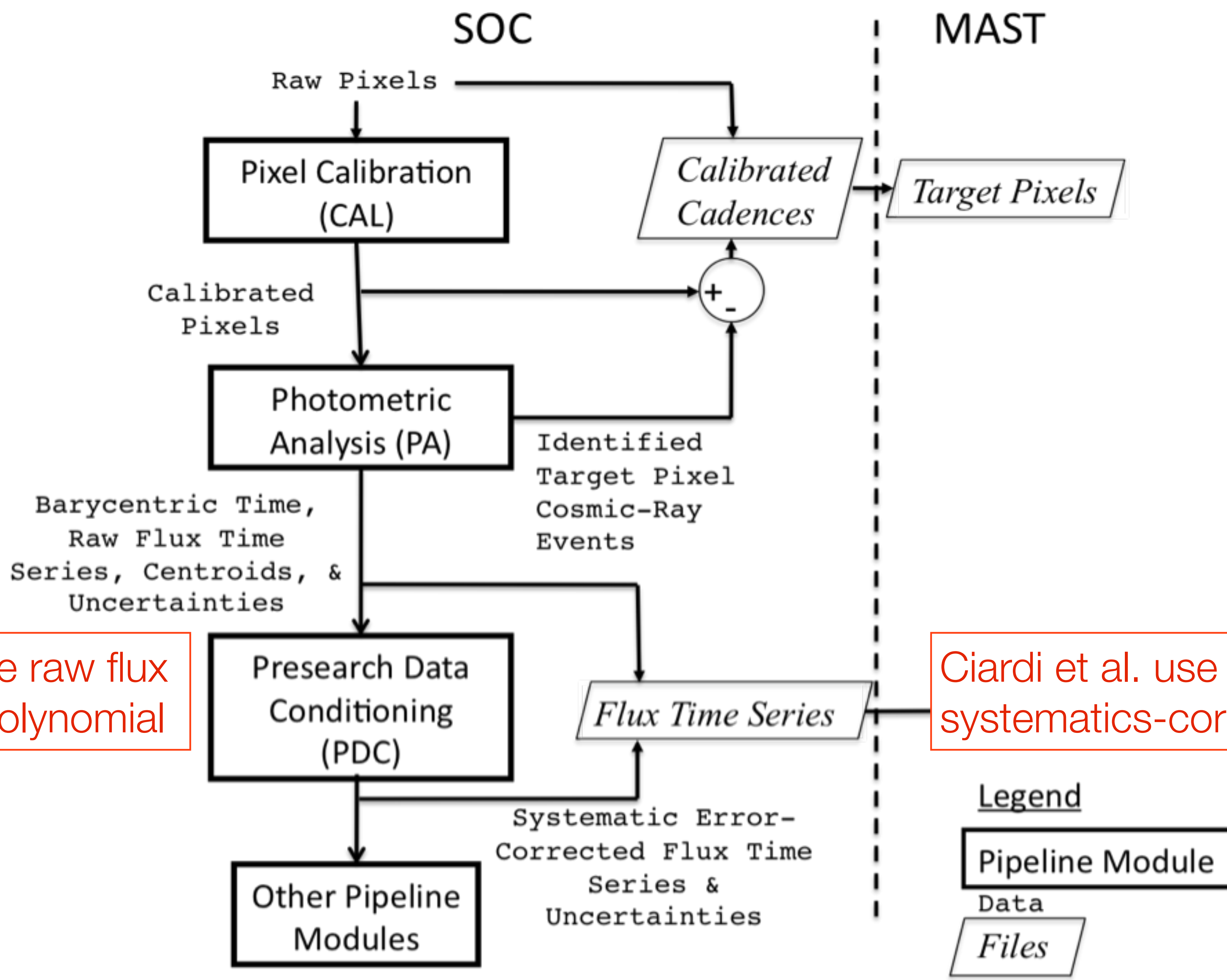
- 0.9m Schmidt telescope with 110 sq.deg. FOV at L2
- Simultaneously monitor > 10^5 stars every 30 min for 4 years
- On-board aperture photometry, precision down to ~10 ppm
- Primary goal: detection of transits of exoplanets, including Earth-like
- First 4 months of data now public
    - Gold mine for stellar science
    - My interest: activity and angular momentum evolution


+41°15'
+37°30'
+33°45'
+30°00'
20h 40m
20h 20m
20h 00m
19h 40m
19h 20m
Vega
Deneb
Galactic Plane
Cygnus
Lyra
M29
M56
M57
N6811
N6866
N6819
N6791

Kepler variability studies

- Based on quarter 1: 33 days
- Basri et al. (2010a,b)
  - >50% (80%) of Kepler (G-)dwarf targets are quieter than the Sun
    - Important implications for prospected for detecting Earth analogs
  - much larger fraction of significantly variable stars are periodic
- Ciardi et al. (2010):
  - different light curve preprocessing and variability statistics, similar variability fractions
  - tantalisingly bimodal variability distribution - old & young populations?

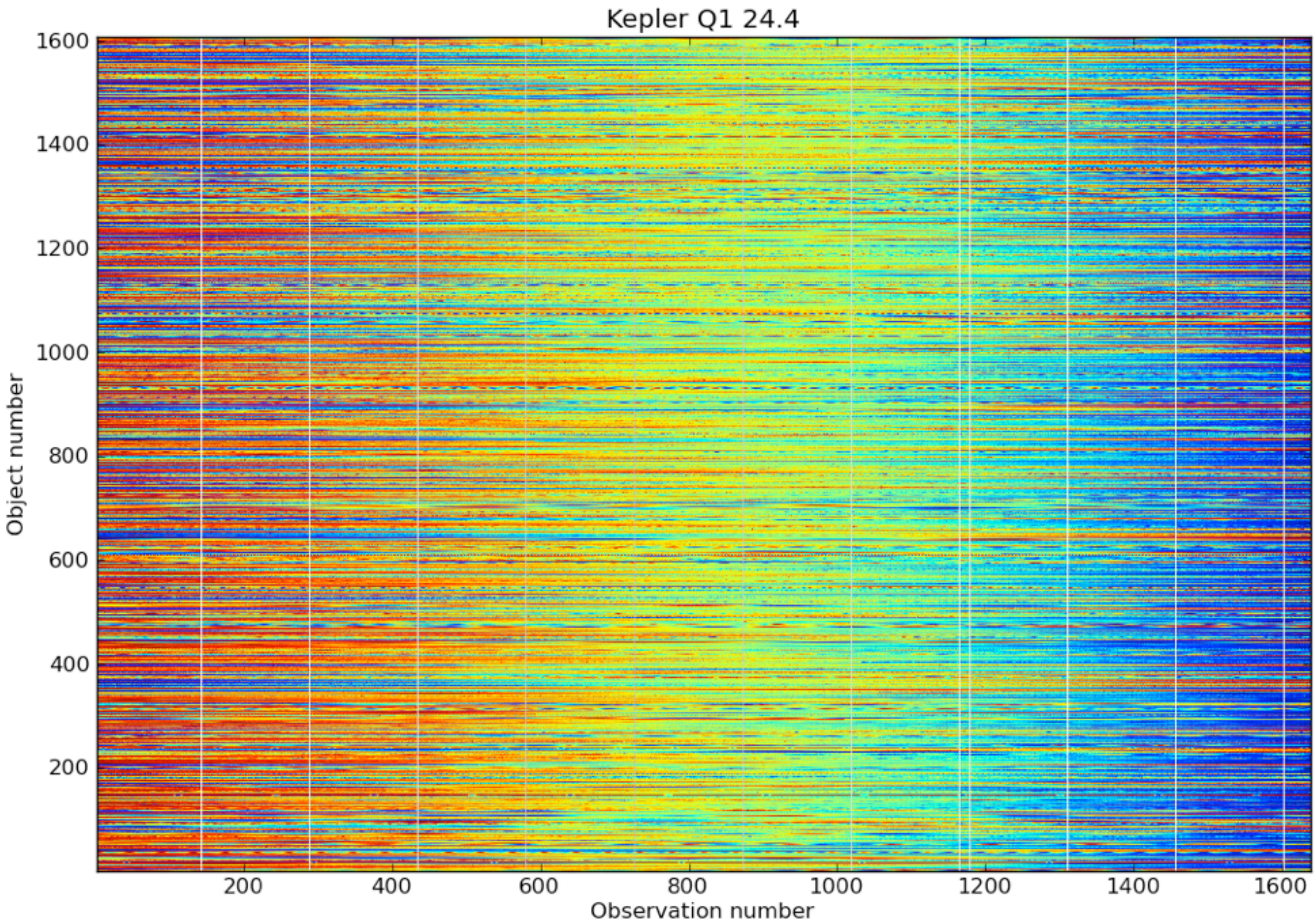
Kepler processing



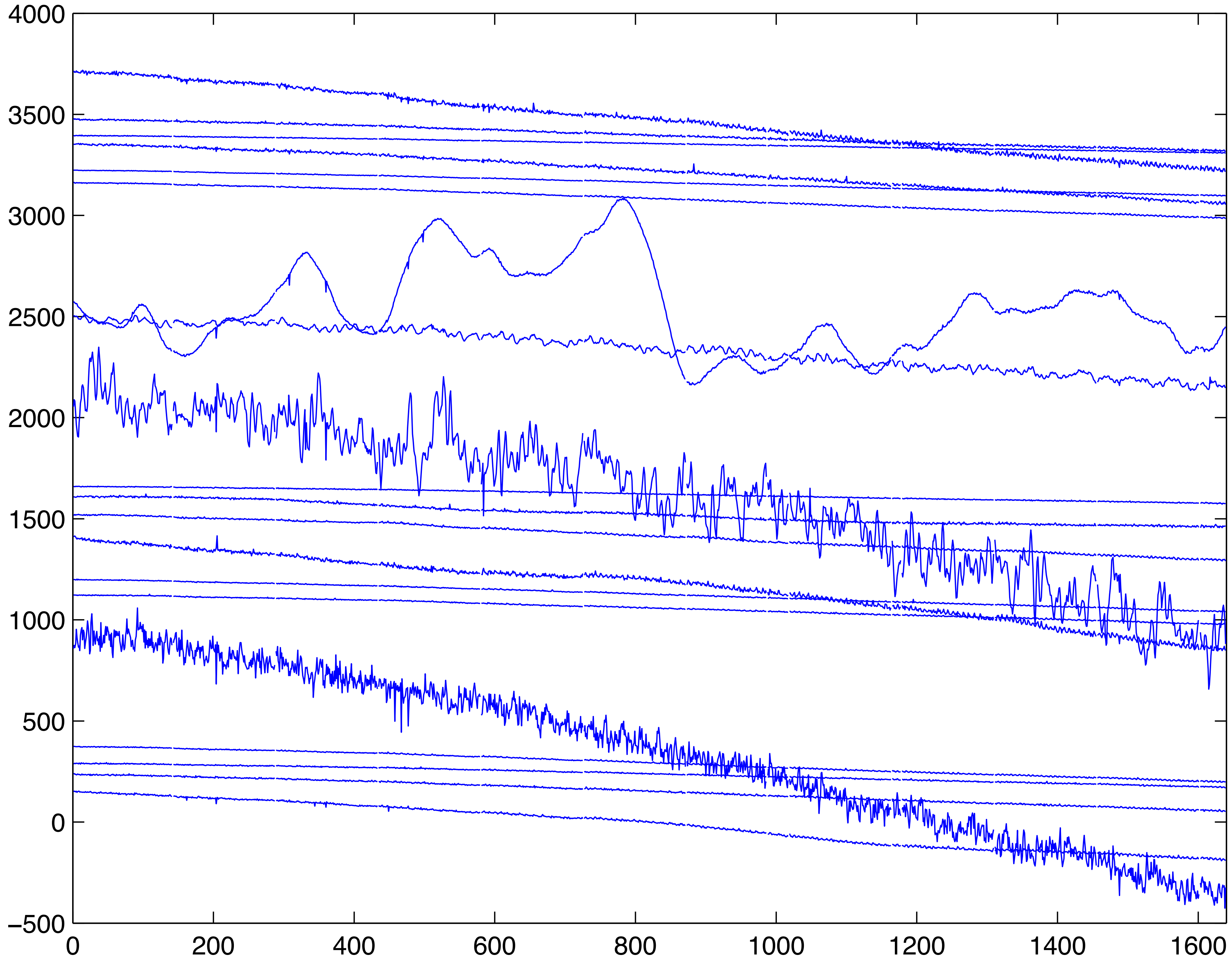
Basri et al. use raw flux + low-order polynomial

Ciardi et al. use systematics-corrected flux

Kepler systematics



Kepler systematics



Kepler systematics

- Attitude jitter / drift
    - variable stars among guide stars...
    - affects flux through intra- and inter-pixel sensitivity variations
- Temperature changes
- Background changes
- Detector degradation?
- Anything else?

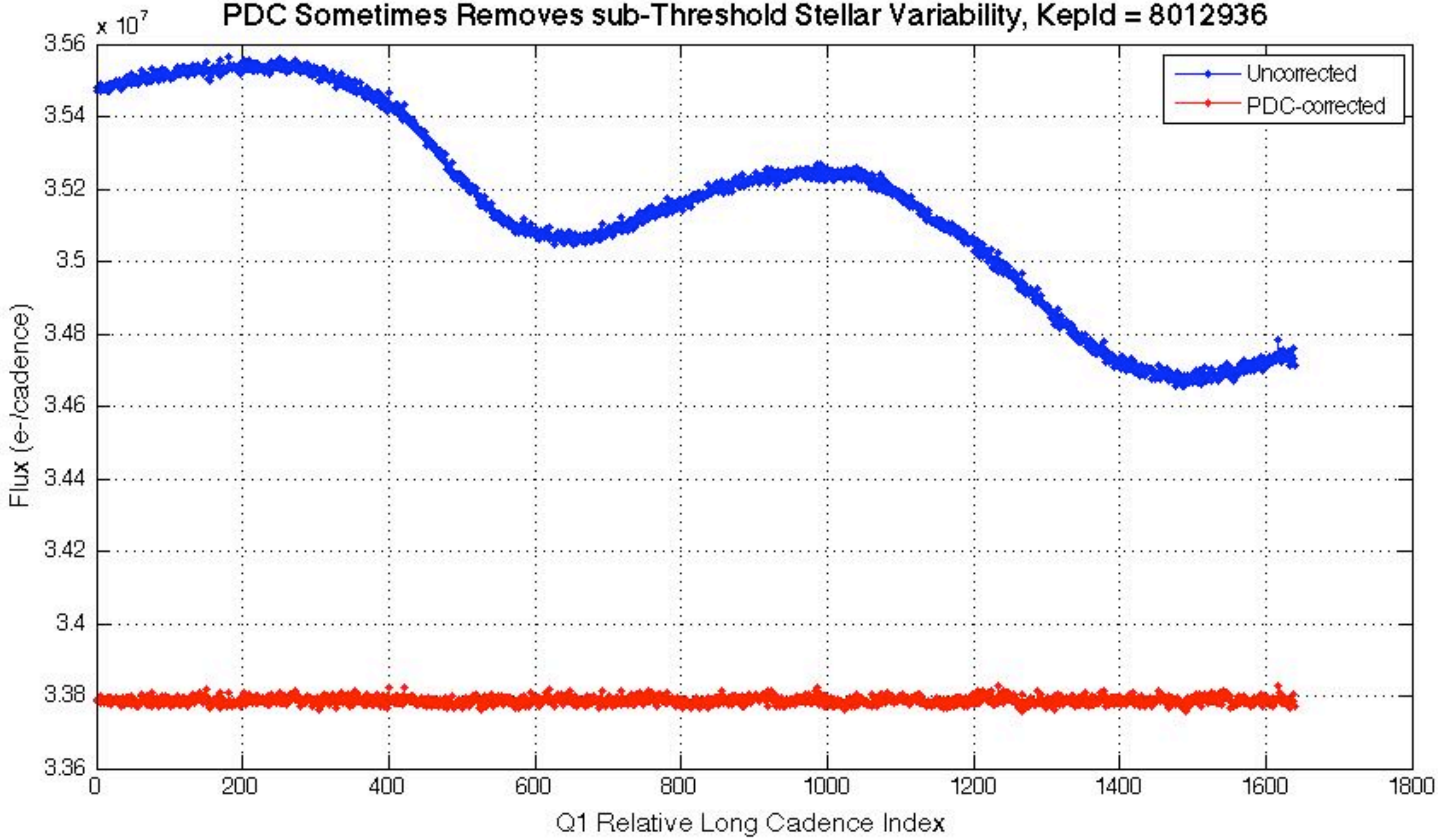
Linear basis models for systematic trends

- flux = sum(coeffs * common trend) + intrinsic variations + noise
- Problem: how to chose the basis?
  - Systematics: common to many stars. Helps but still under-constrained
  - Intrinsic variations: could be anything: generally treated as IID noise...

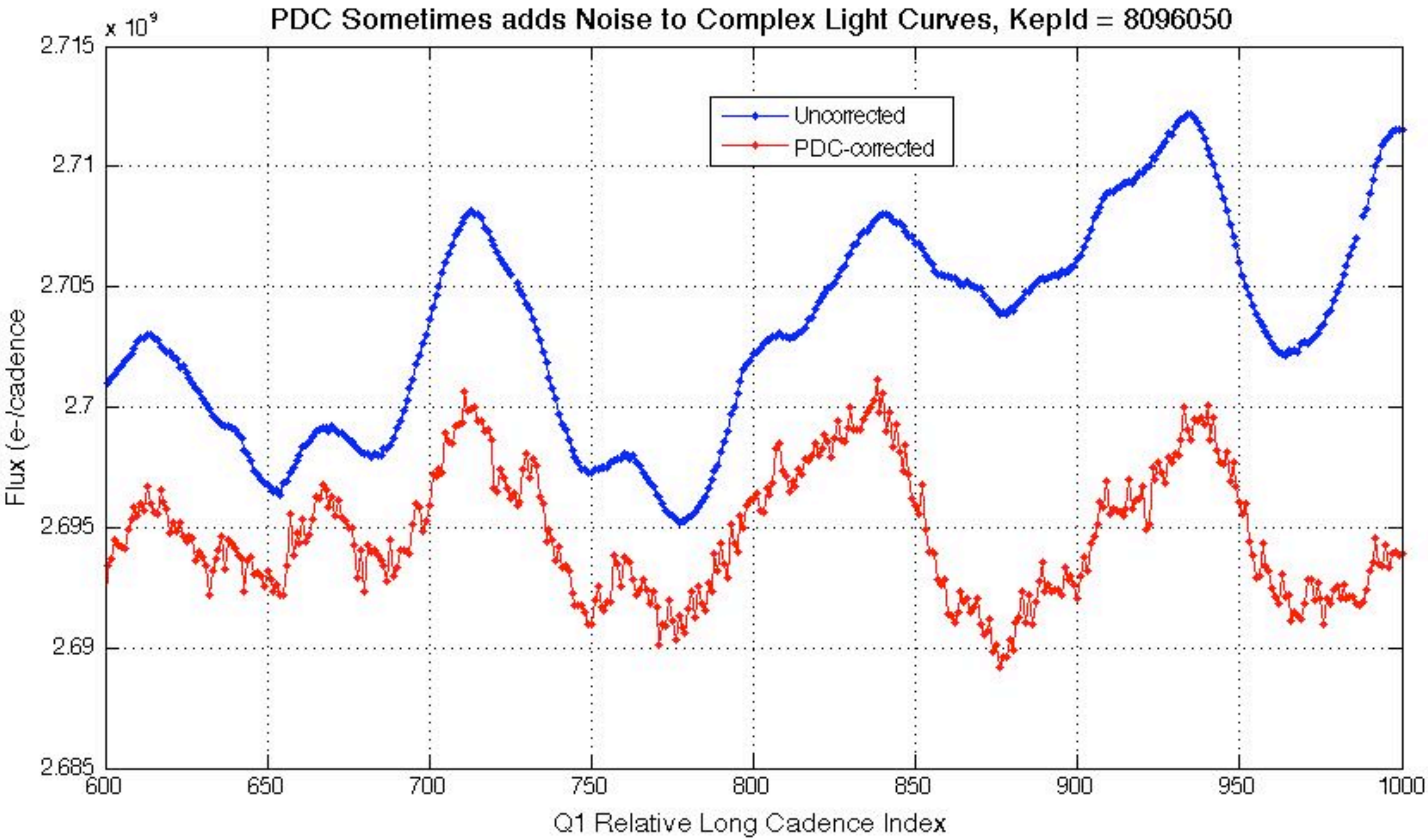
The Kepler pipeline systematics correction

- Correct for discontinuities due to known events
- Linear basis model using {x, y, roll, temperature, background} as basis
  - if a star appears variable after correction, attempt to remove the variability from original LC, then repeat systematics estimation
- If the correction made RMS >5% worse for a given star, revert to original
- Kepler science team warn this correction is optimised and tested for transit detection (timescales 2-12h) only

Problems...


PDC Sometimes Removes sub-Threshold Stellar Variability, KepId = 8012936
x 10^7
3.56
3.54
3.52
3.5
3.48
3.46
3.44
3.42
3.4
3.38
3.36
Flux (e-/cadence)
0
200
400
600
800
1000
1200
1400
1600
1800
Q1 Relative Long Cadence Index
Uncorrected
PDC-corrected

Problems...



PDC Sometimes adds Noise to Complex Light Curves, KepId = 8096050
x 10^9
2.715
2.71
2.705
2.7
2.695
2.69
2.685
Flux (e-/cadence
600
650
700
750
800
850
900
950
1000
Q1 Relative Long Cadence Index
Uncorrected
PDC-corrected

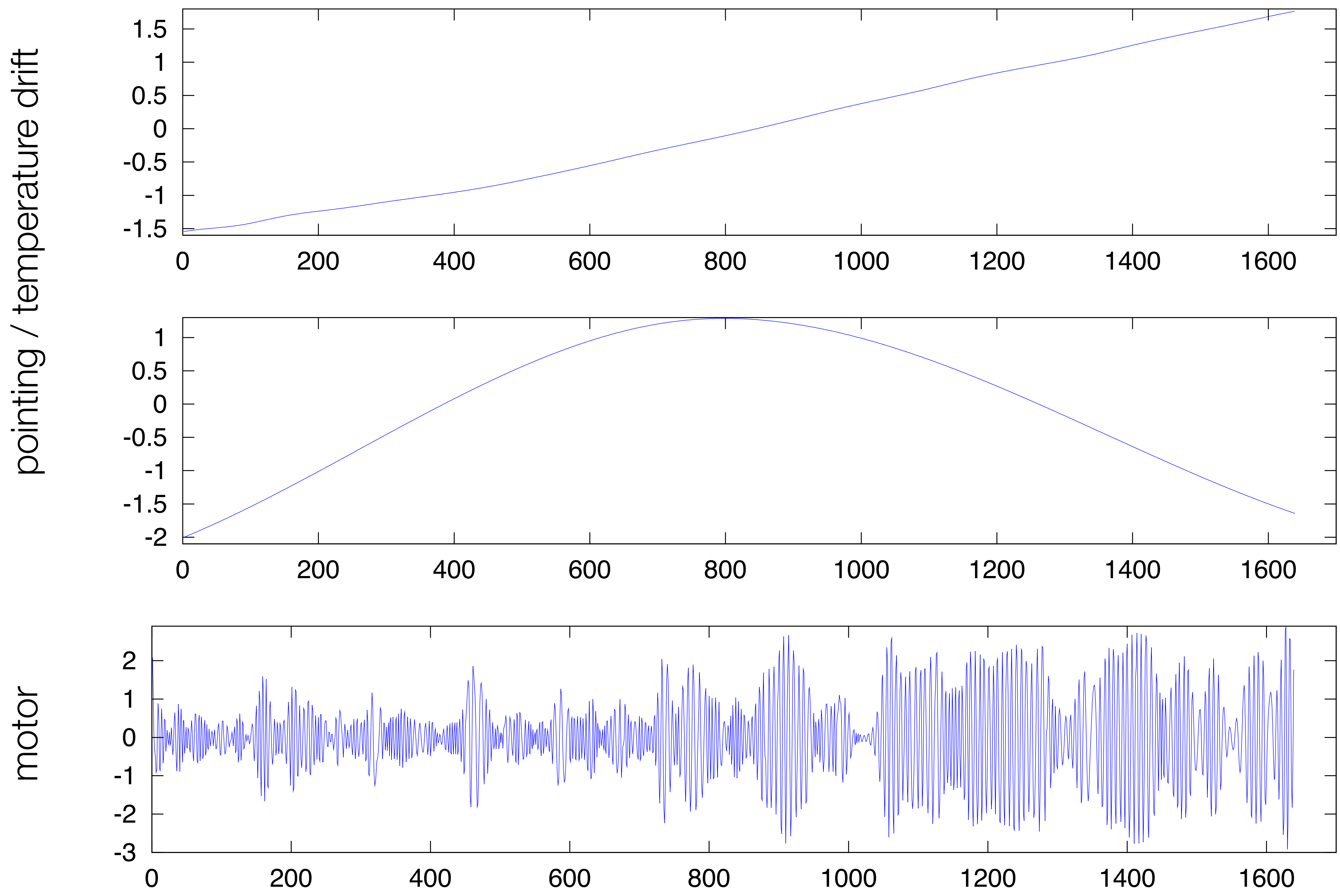
Alternative class of approaches

- Construct the basis from linear combinations of the light curves themselves
    - $LC_i = \Sigma_{j \neq i} \beta_{ij} LC_j + S_i + \text{noise}$
        - solve for the β's by ignoring S (!)
- Problems:
    - global variance is dominated by intrinsically variable stars
    - ~all stars are intrinsically variable at Kepler precision

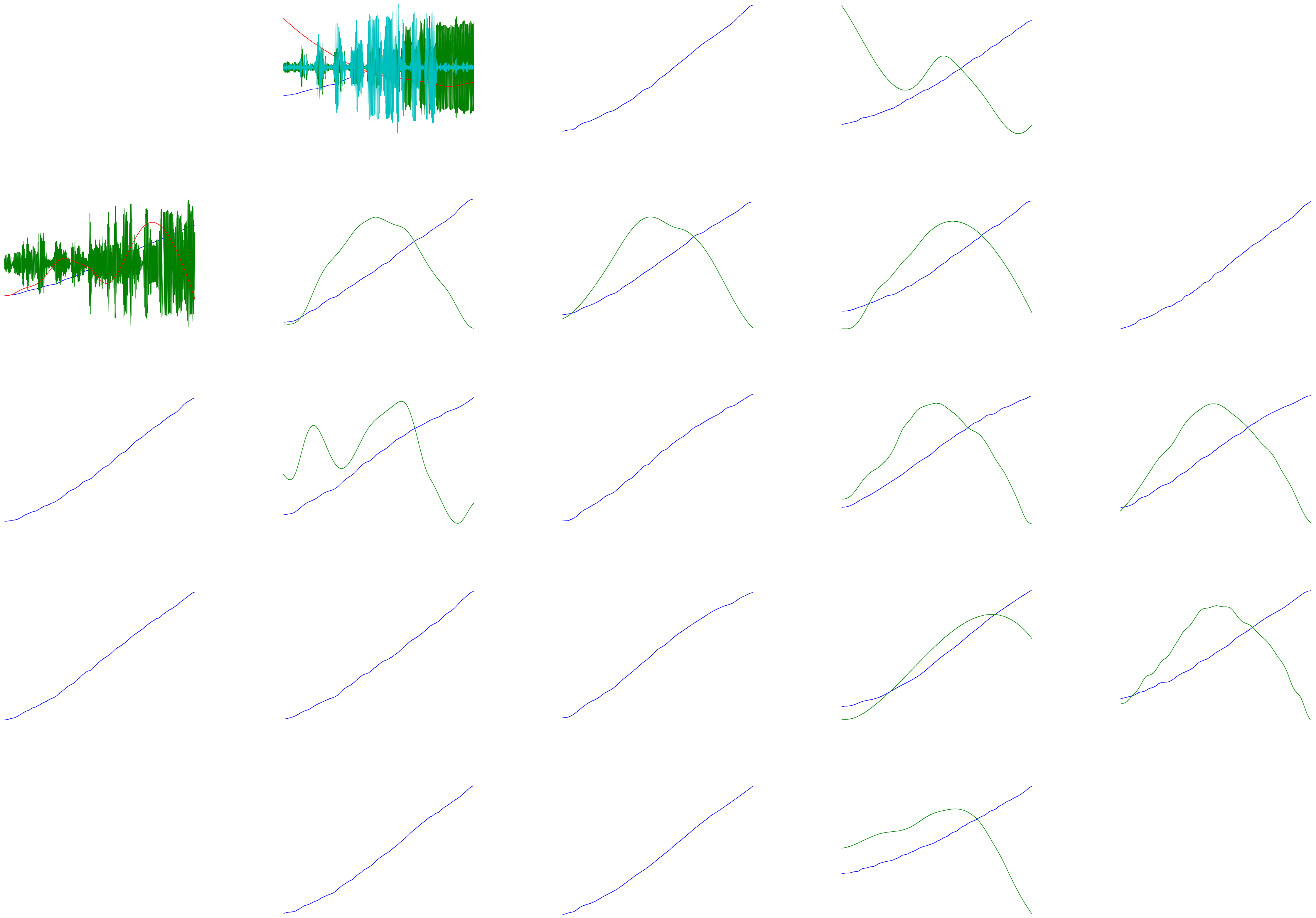
Our approach

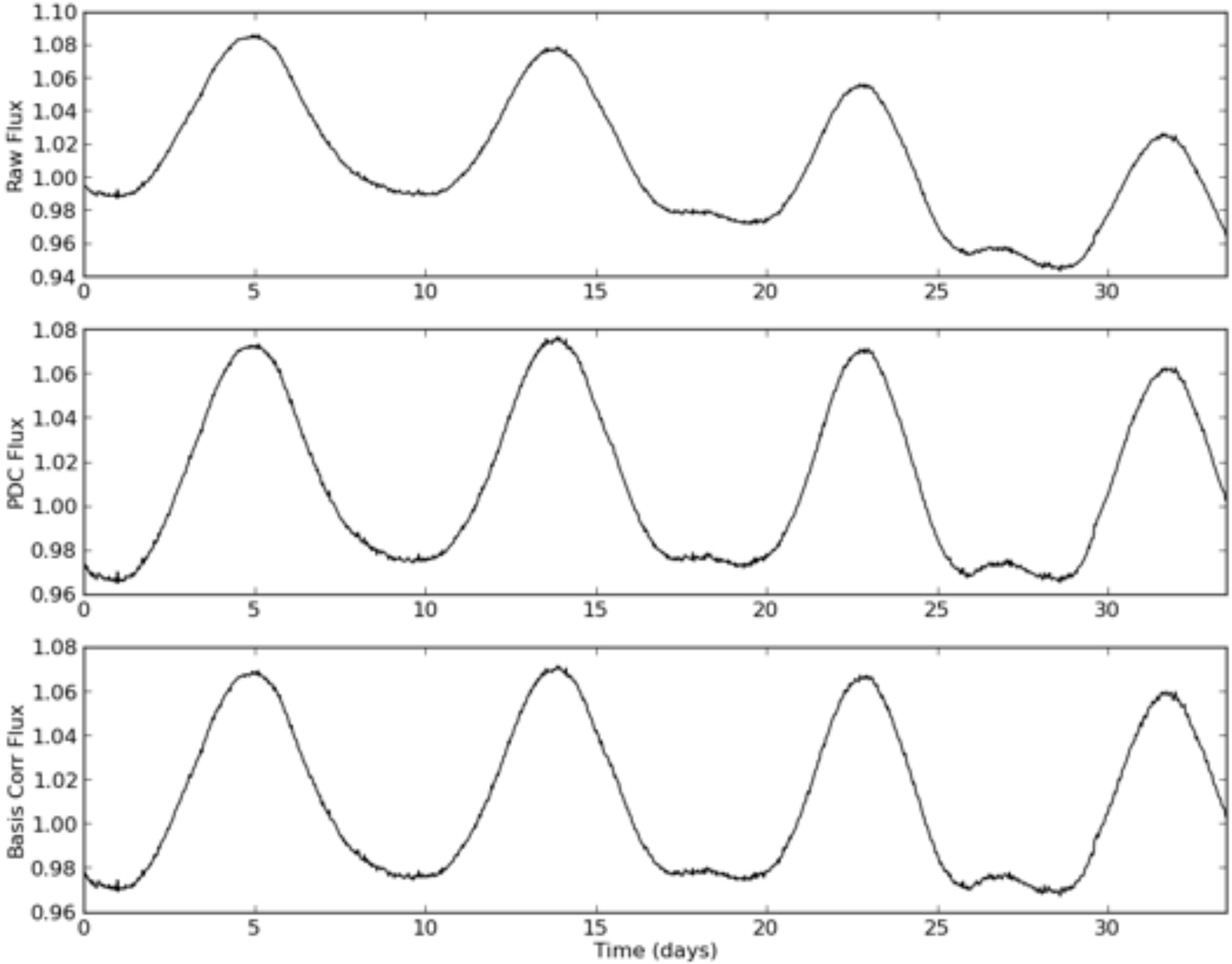
- Construct the basis from linear combinations of the light curves themselves
    - $LC_i = \Sigma_{j \neq i} \beta_{ij} LC_j + S_i +$ noise
        - we solve for the β's by ignoring S (still) but using Bayesian linear regression, which gives us probabilities p(β)
        - Each LC gives a putative systematic trend, specified by a column of β
- True systematic trends should be present in many LCs:
    - rank them by Shannon entropy $\mathcal{H}(\beta_i) = \Sigma_j \beta_{ij} P(\beta_{ij})$
    - combine highest $\mathcal{H}$ trends (details, details...)
- Repeat procedure until gain in explained global variance flattens out
- The few trends identified thus form final basis

Q1 global trends

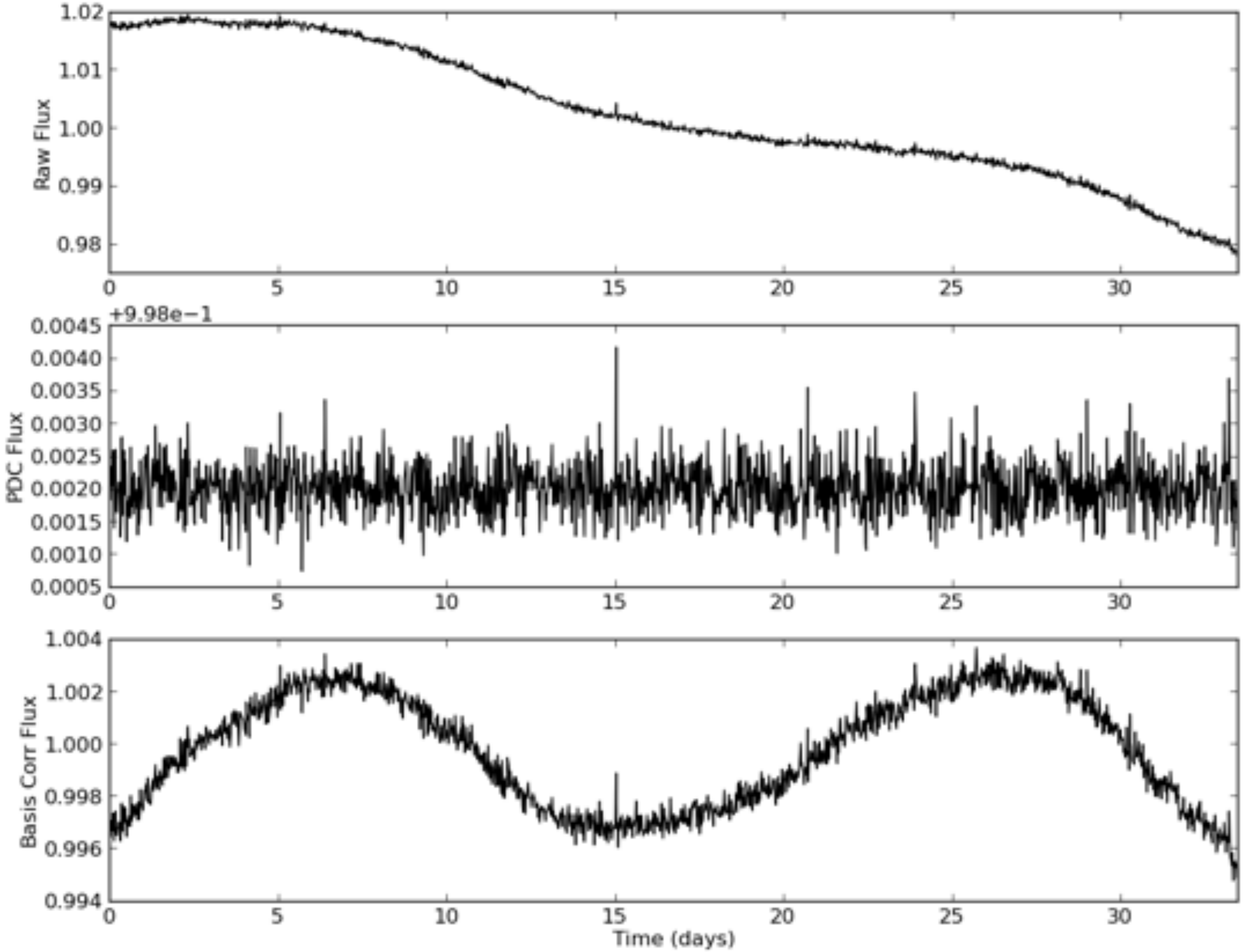


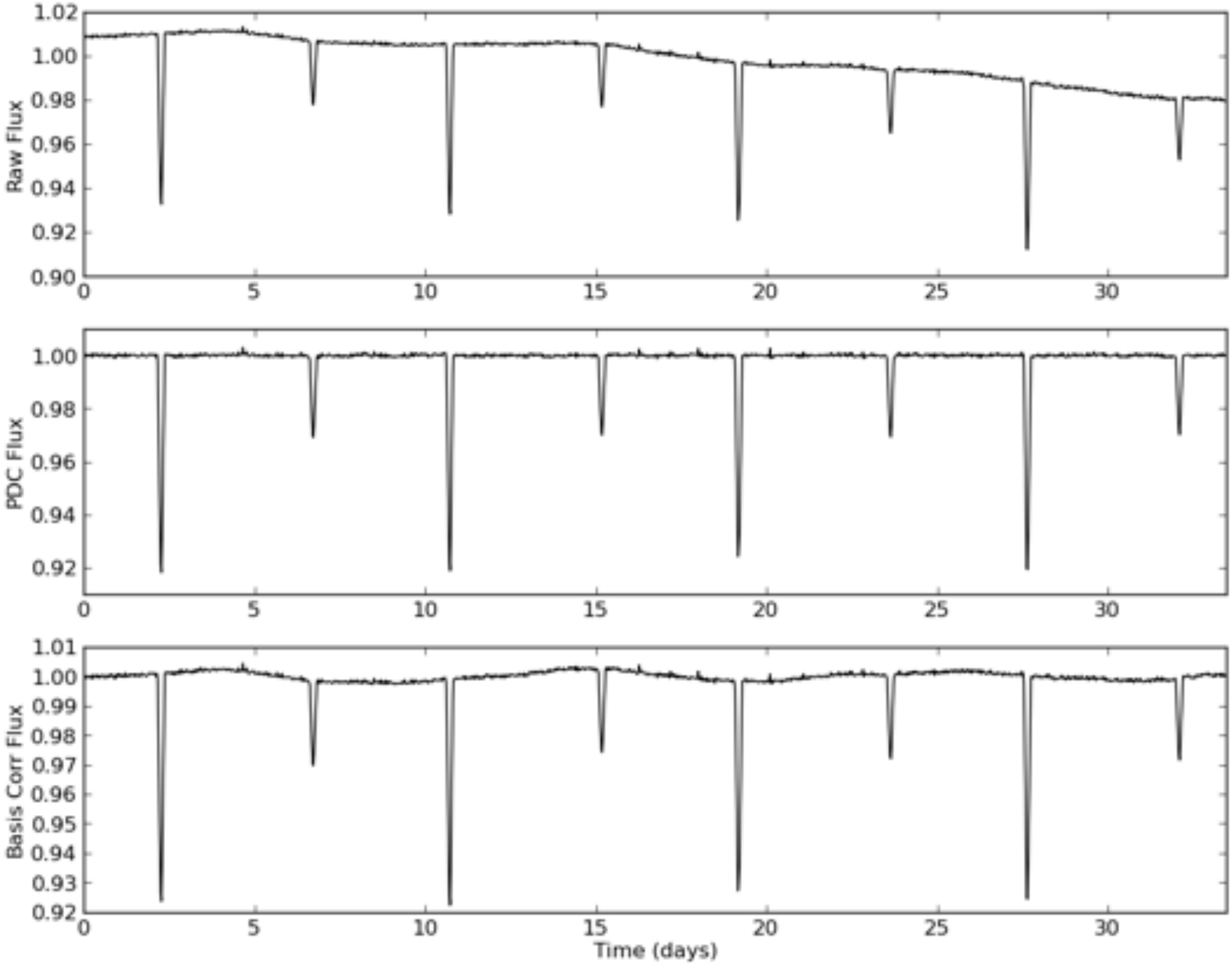
Module by module trends

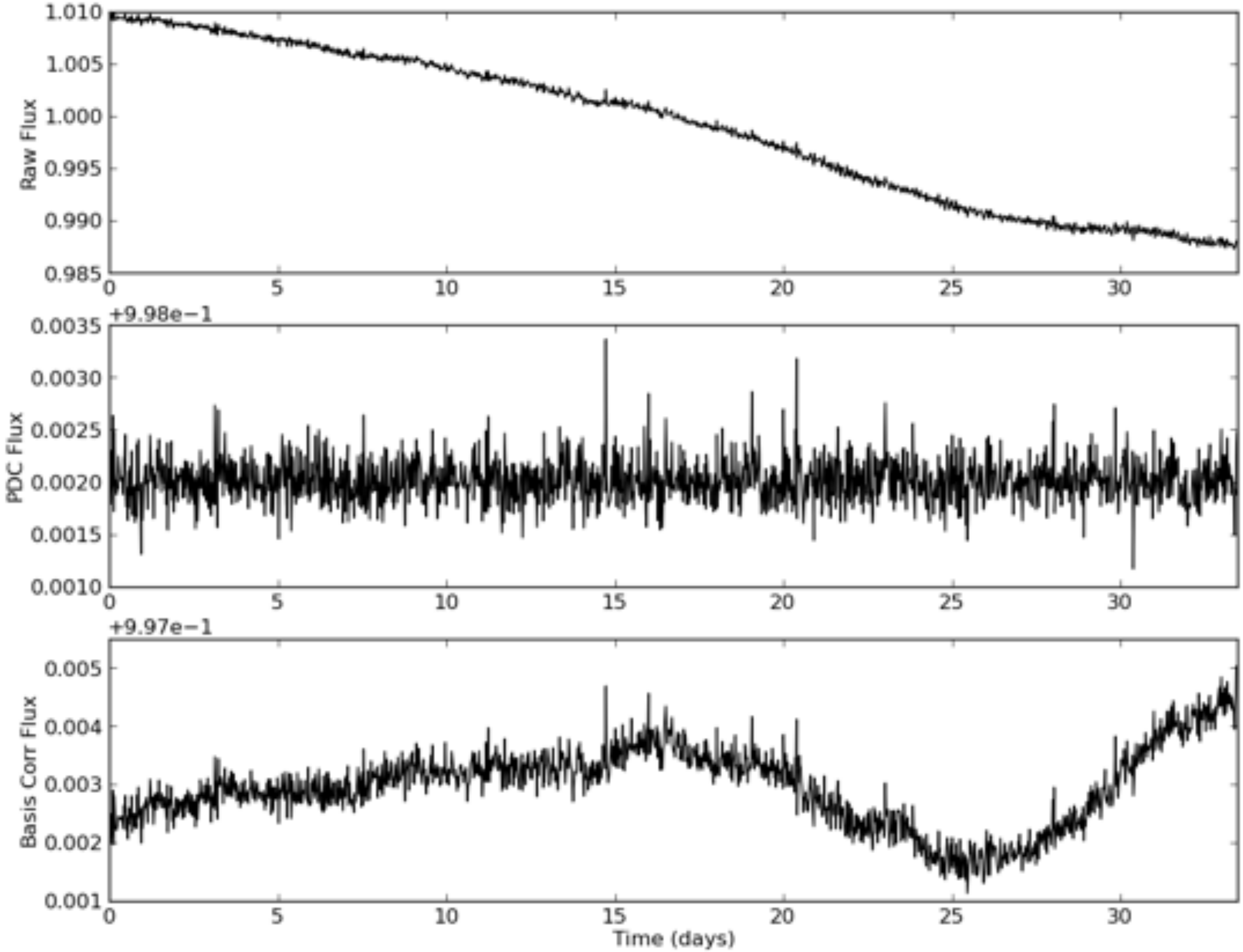




Raw Flux
PDC Flux
Basis Corr Flux
Time (days)

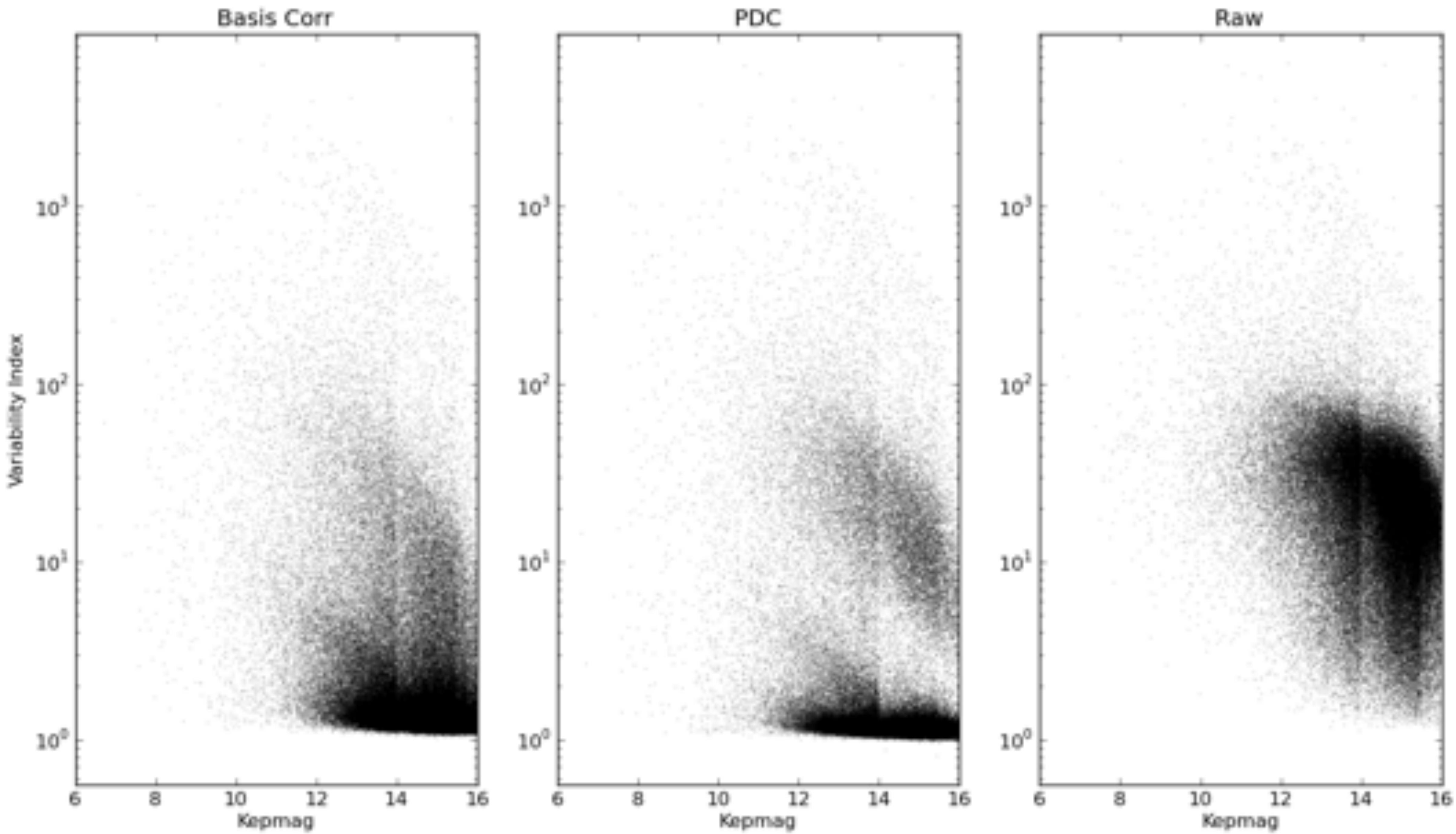

Raw Flux
PDC Flux
+9.98e−1
Basis Corr Flux
Time (days)


Raw Flux
PDC Flux
Basis Corr Flux
Time (days)

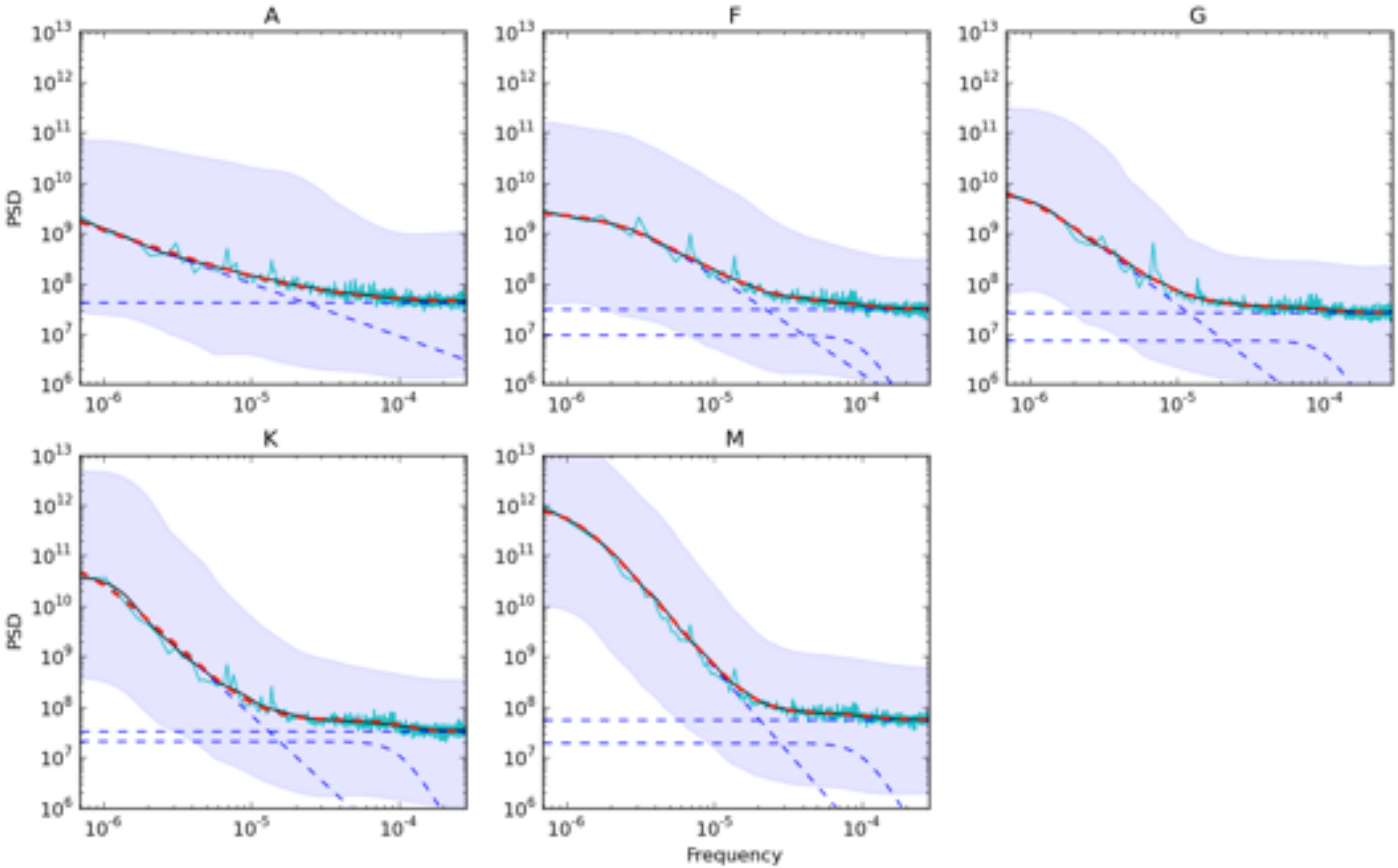


Raw Flux
1.010
1.005
1.000
0.995
0.990
0.985
0
5
10
15
20
25
30
+9.98e−1
PDC Flux
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
+9.97e−1
Basis Corr Flux
0.005
0.004
0.003
0.002
0.001
Time (days)

Revisiting the bimodality



Autoregressive models by spectral type



Summary

- It's not because we are attempting to measure signals that we know are barely detectable that we should not use Bayesian methods
  - quite the contrary!
- It's always worth finding a way to express the constraint we instinctively want to apply
- If you're interested in either using GPs for astronomical (time series) applications, or using Kepler data for variability studies, do talk to me!