Modeling **Correlated Noise** in Stellar Spectra with **Gaussian Processes**

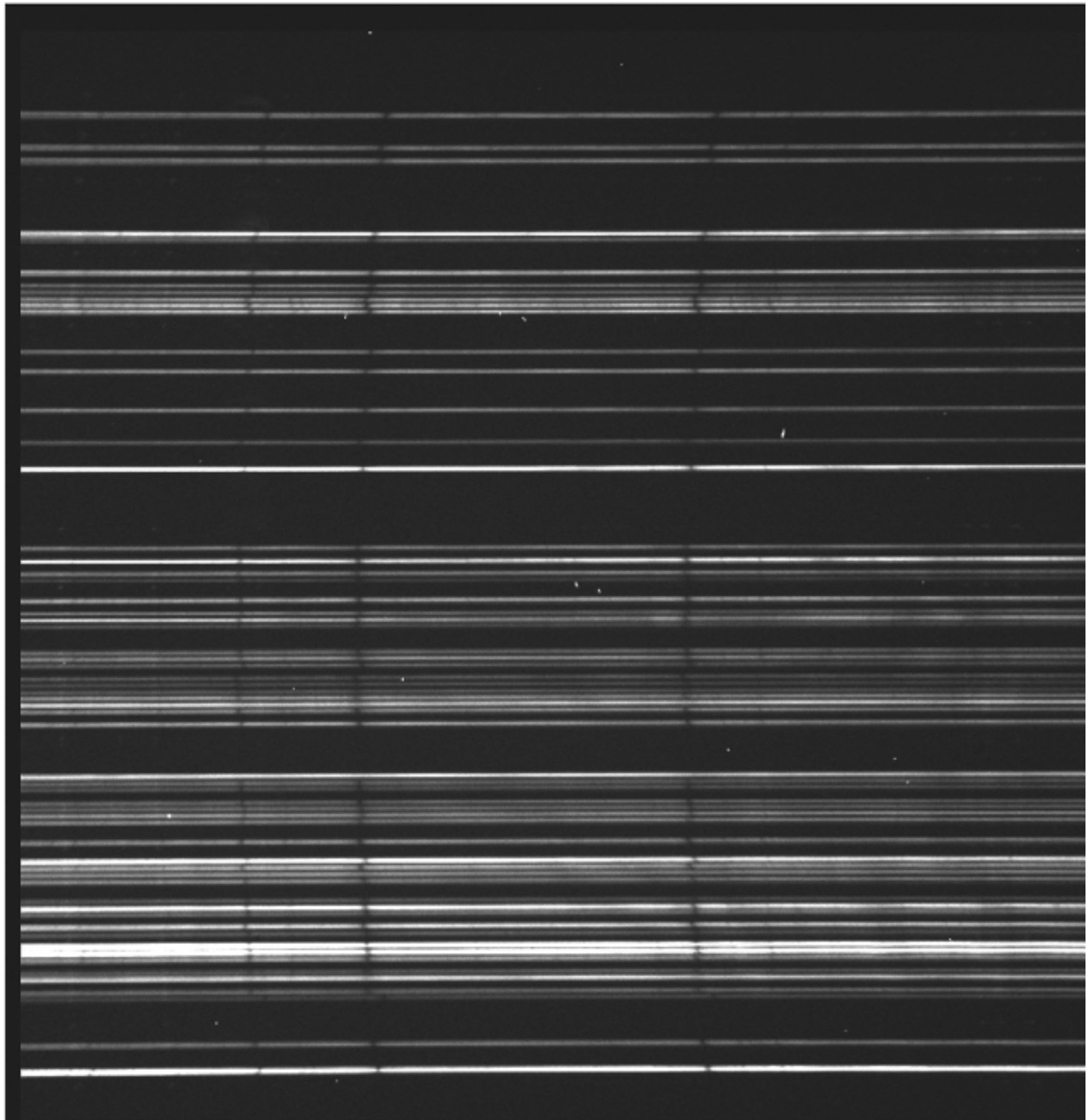
Gal Matijevič
Leibniz-Institut für Astrophysik Potsdam (AIP)
Supercomputing and E-Science

AIP

1 The problem

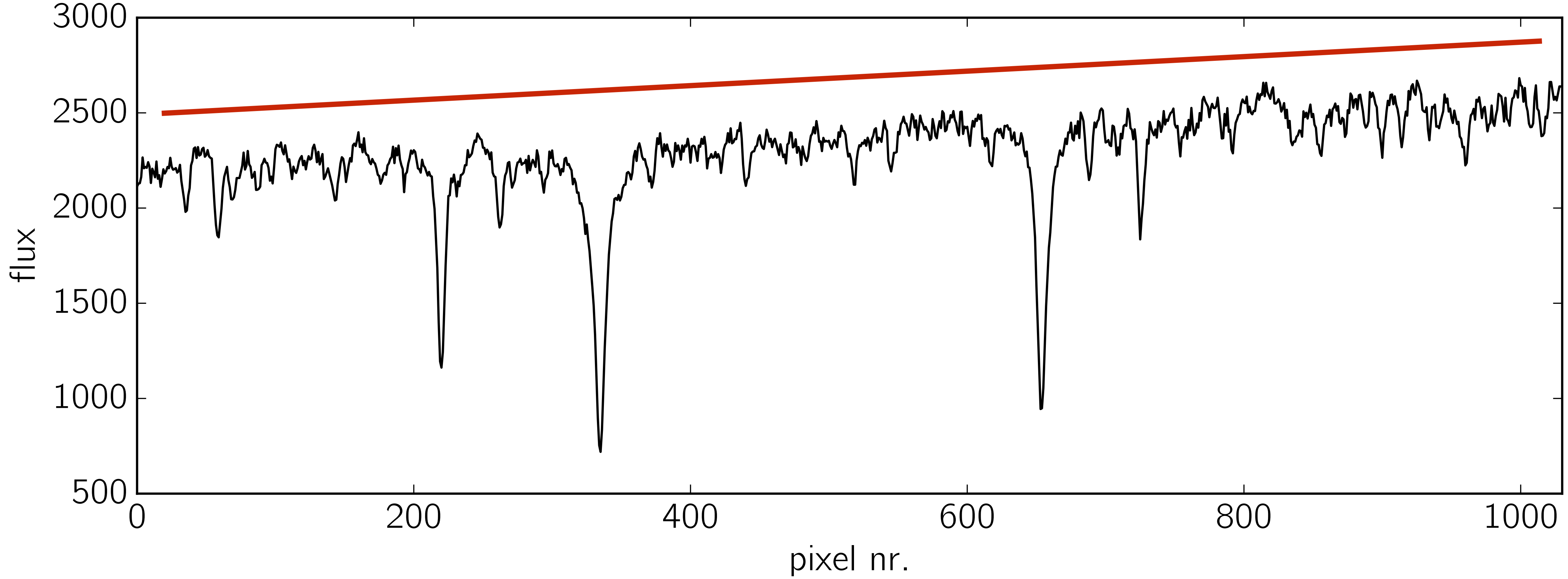
AIP

RAVE Survey CCD field

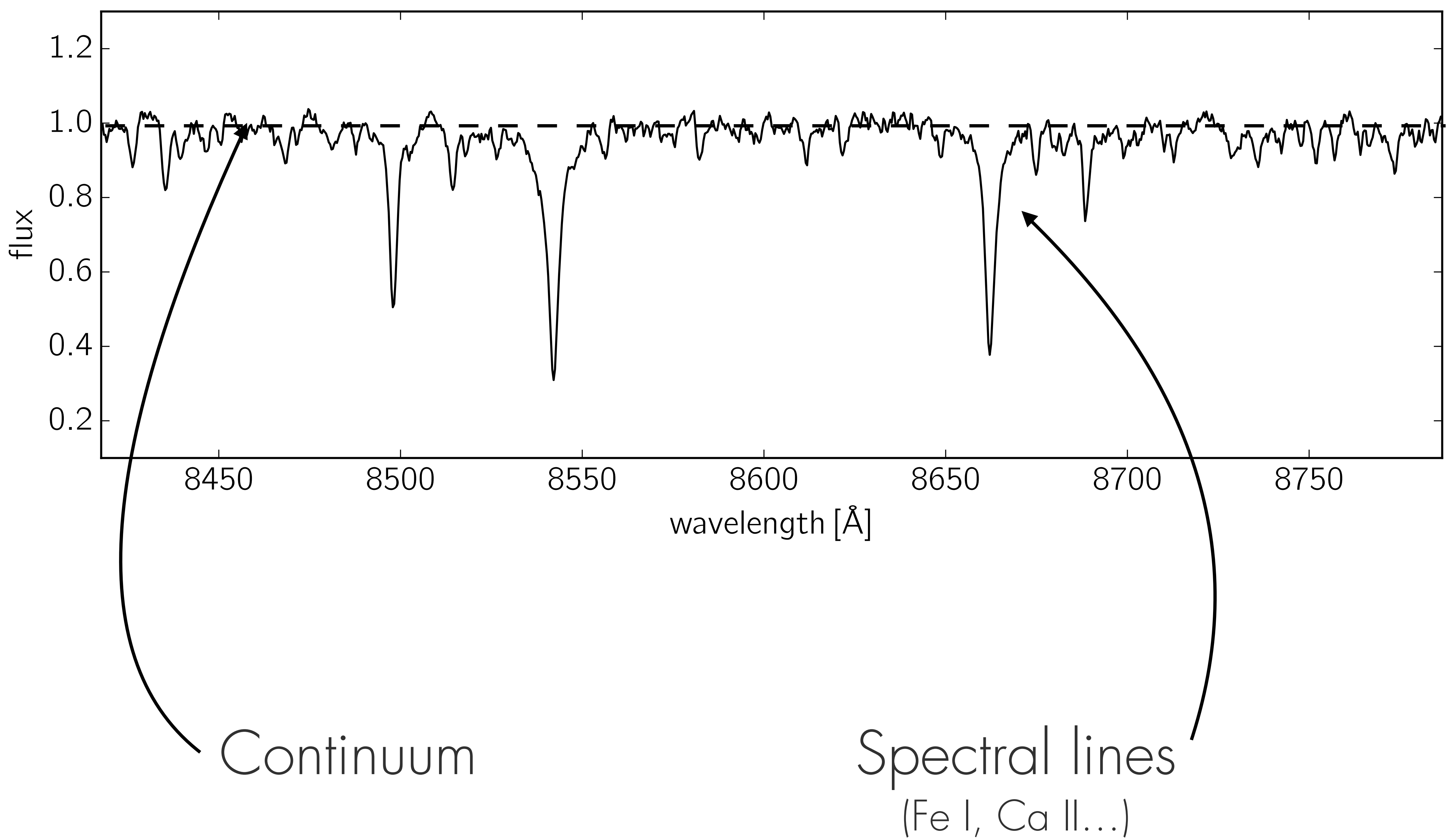


AIP

trend

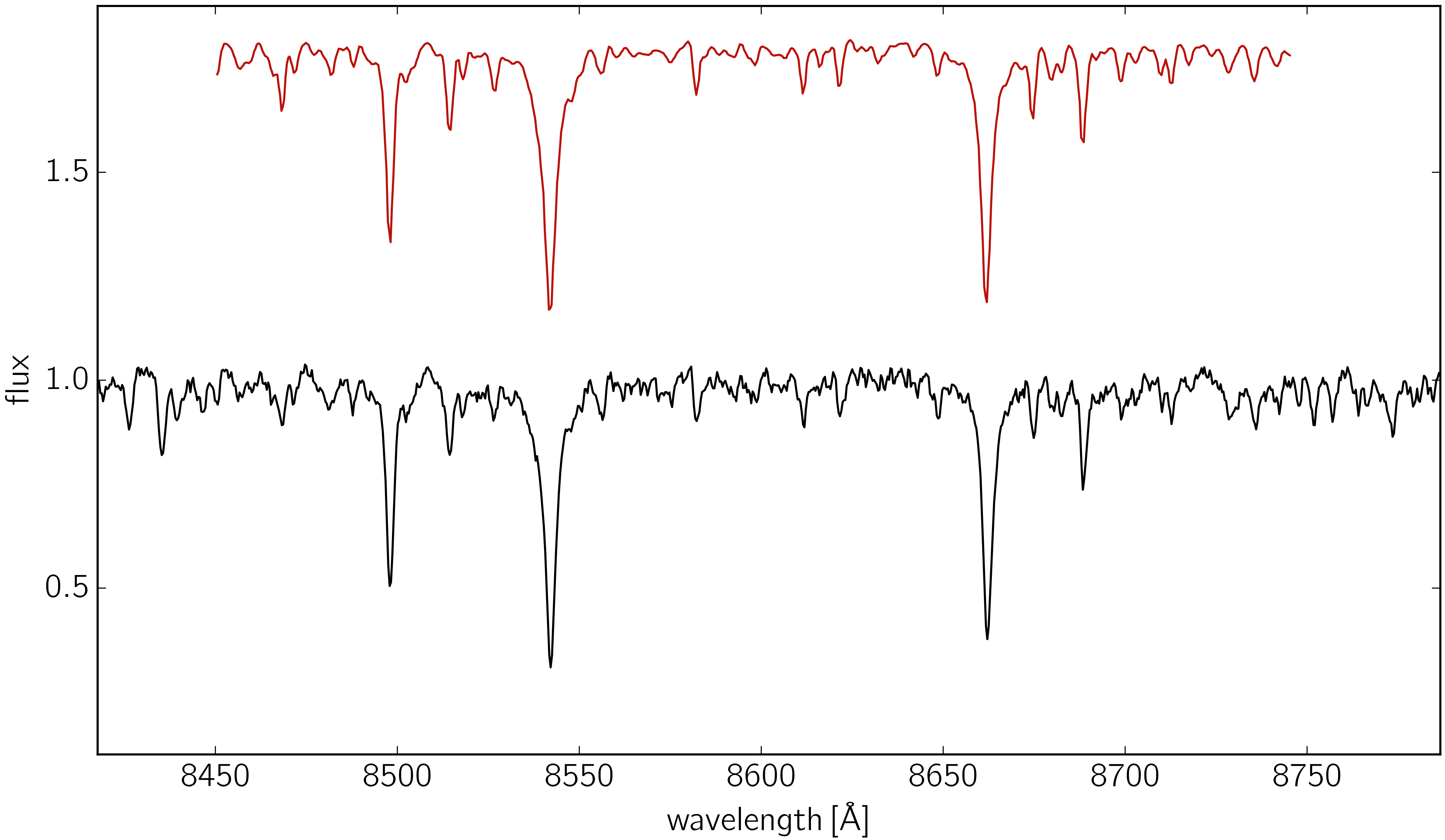

3000
2500
2000
1500
1000
500
flux
0
200
400
600
800
1000
pixel nr.

Giant star

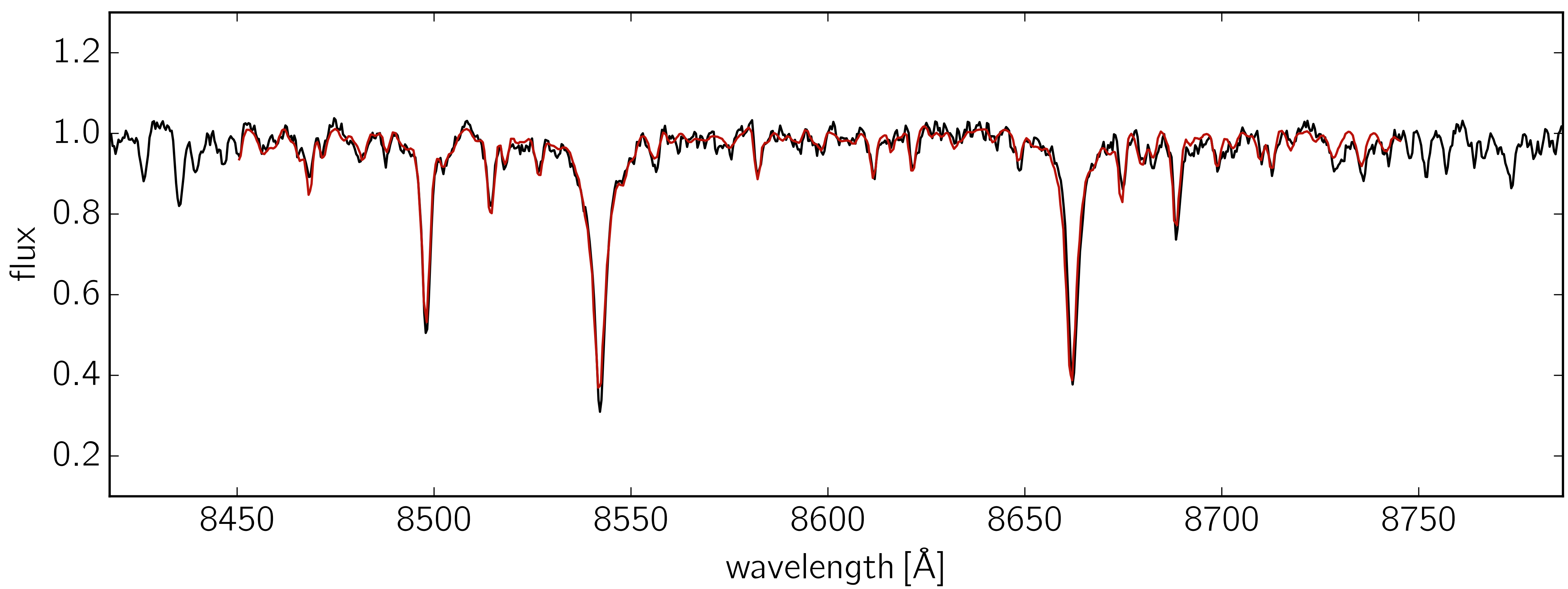


1.2
1.0
0.8
0.6
0.4
0.2
flux
8450
8500
8550
8600
8650
8700
8750
wavelength [Å]
Continuum
Spectral lines
(Fe I, Ca II...)

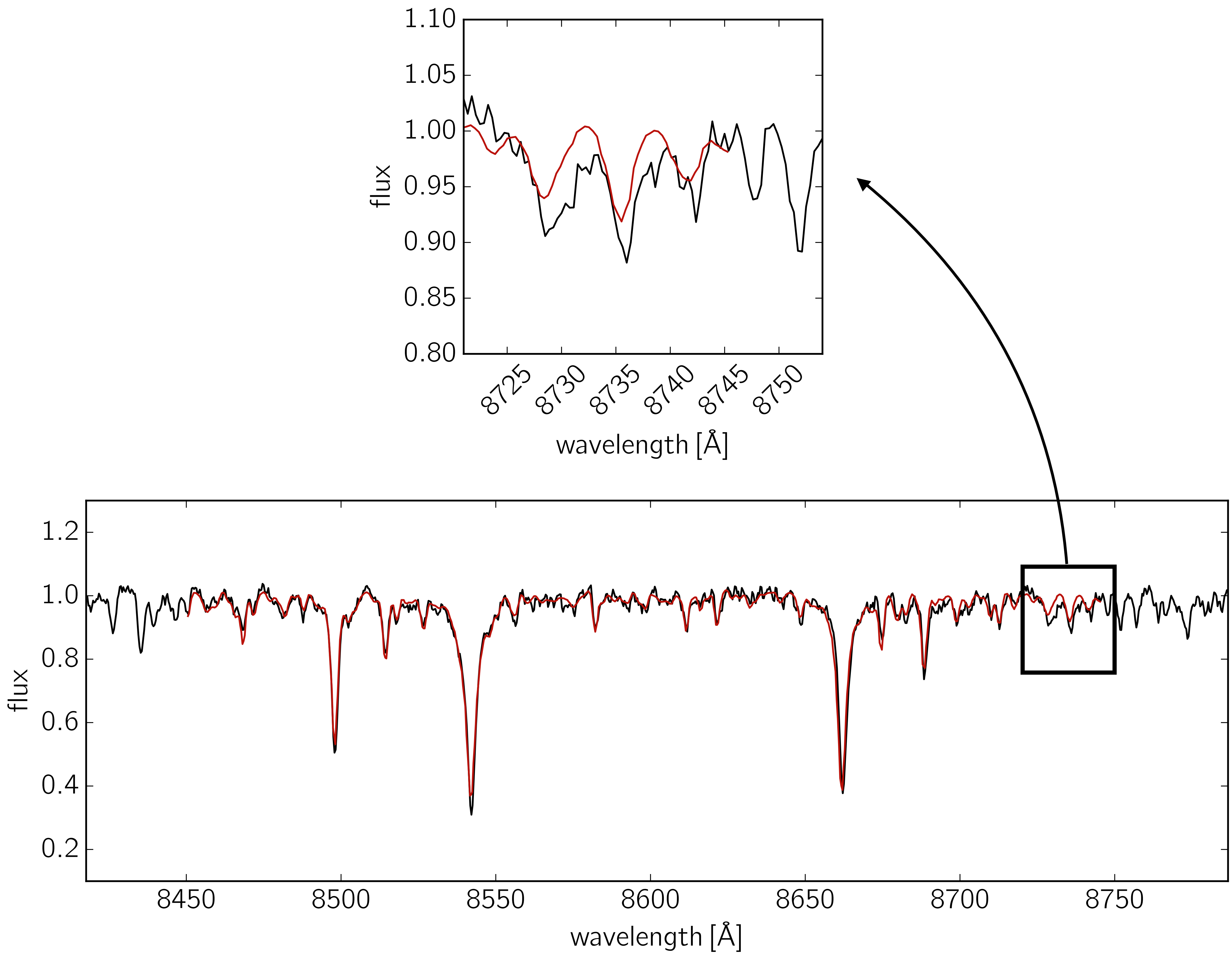
Best-fit model



AIP


1.2
1.0
0.8
0.6
0.4
0.2
flux
8450
8500
8550
8600
8650
8700
8750
wavelength [Å]

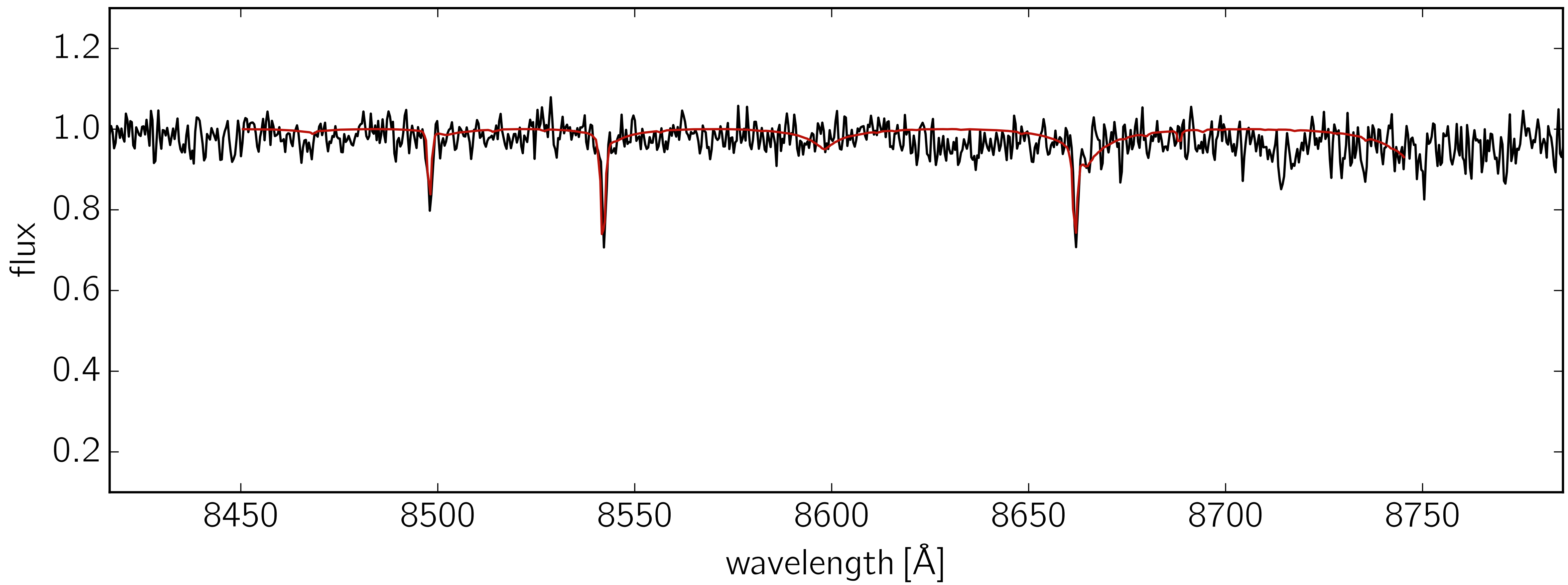
AIP



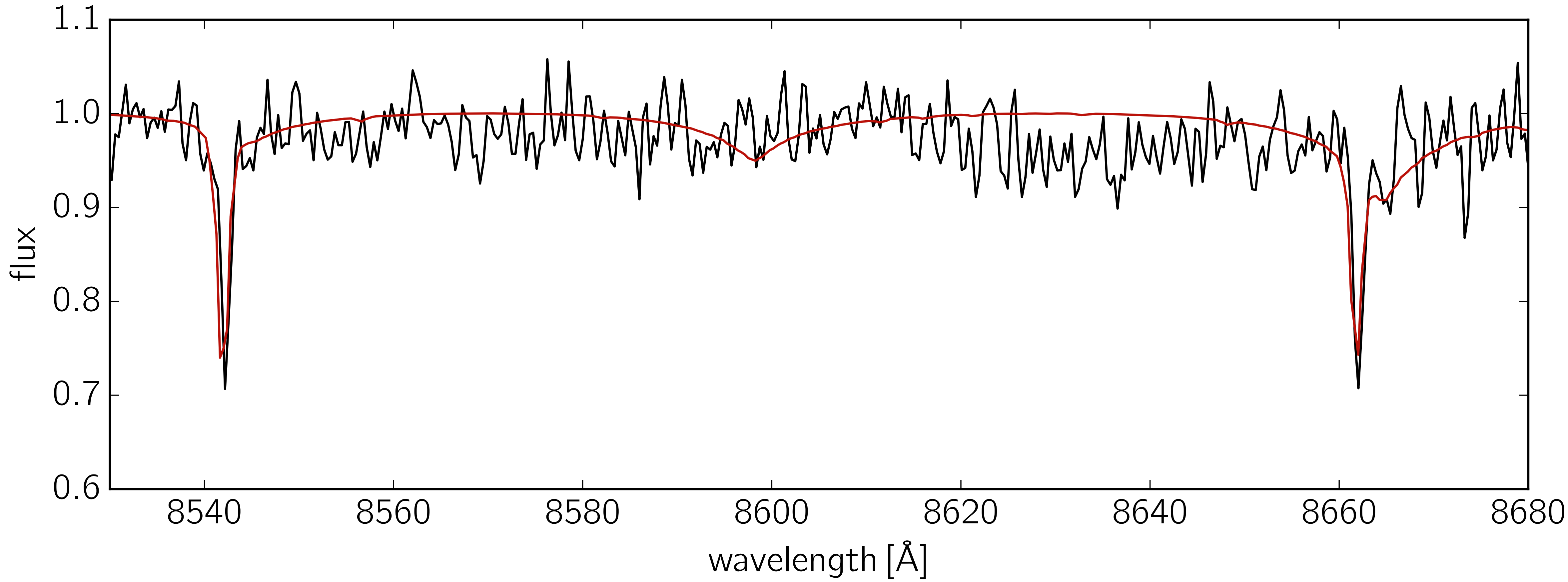
1.10
1.05
1.00
0.95
0.90
0.85
0.80
flux
8725
8730
8735
8740
8745
8750
wavelength [Å]
1.2
1.0
0.8
0.6
0.4
0.2
flux
8450
8500
8550
8600
8650
8700
8750
wavelength [Å]

AIP

Extremely metal-poor star

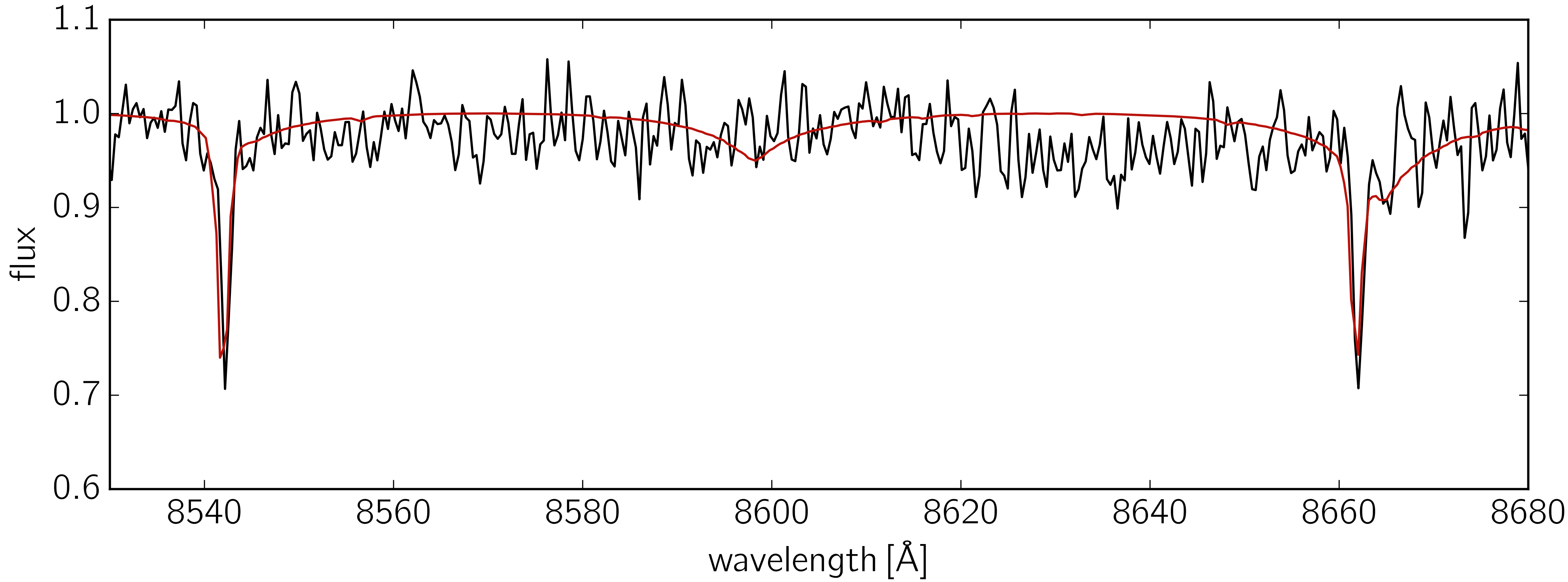


Extremely metal-poor star



AIP

Extremely metal-poor star



Noise is not IID!

AIP

How do we fit the
lines and the
continuum variations
simultaneously ?

2 Gaussian processes

Gaussian processes

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

$$y(x) \sim \mathcal{N}(m_{\boldsymbol{\theta}}(x), K_{\boldsymbol{\alpha}}(x, \boldsymbol{\sigma}))$$

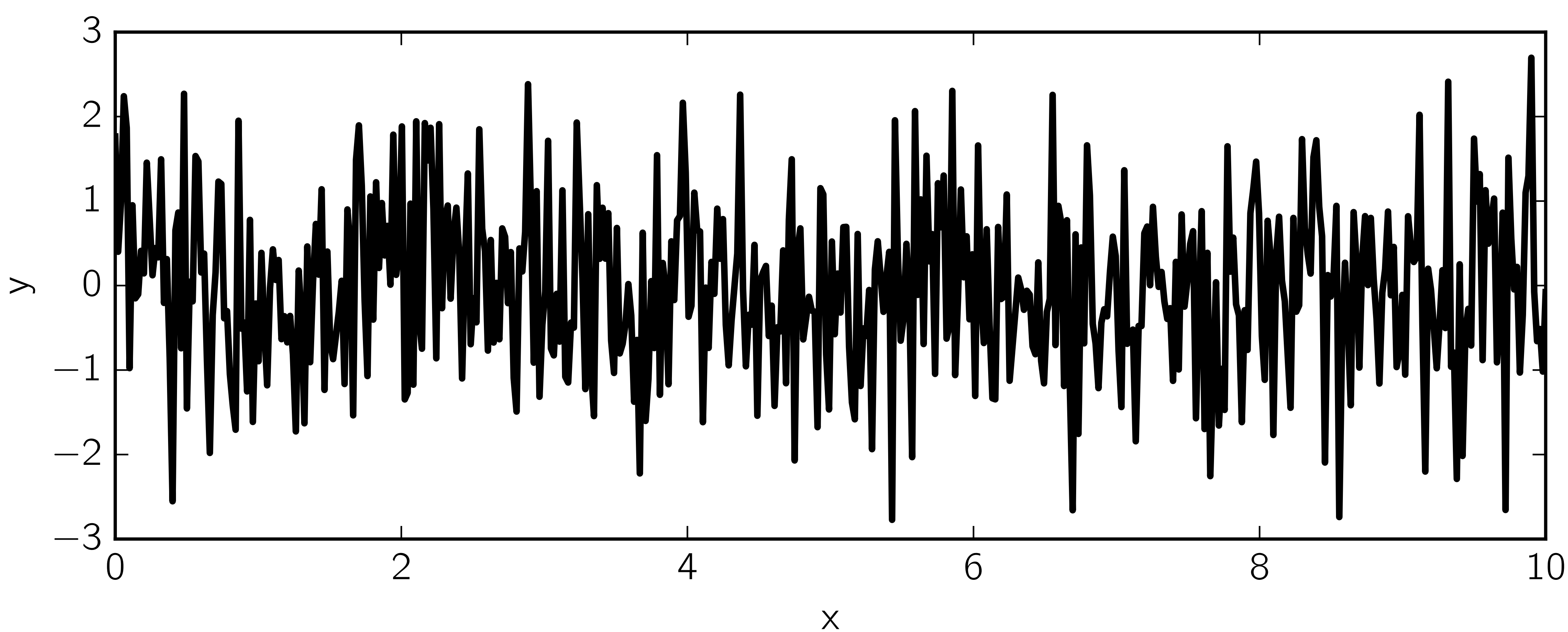
$$[K_{\boldsymbol{\alpha}}(x, \boldsymbol{\sigma})]_{ij} = k_{\boldsymbol{\alpha}}(x_i, x_j) + \sigma_i^2 \delta_{ij}$$

AIP

Gaussian processes

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

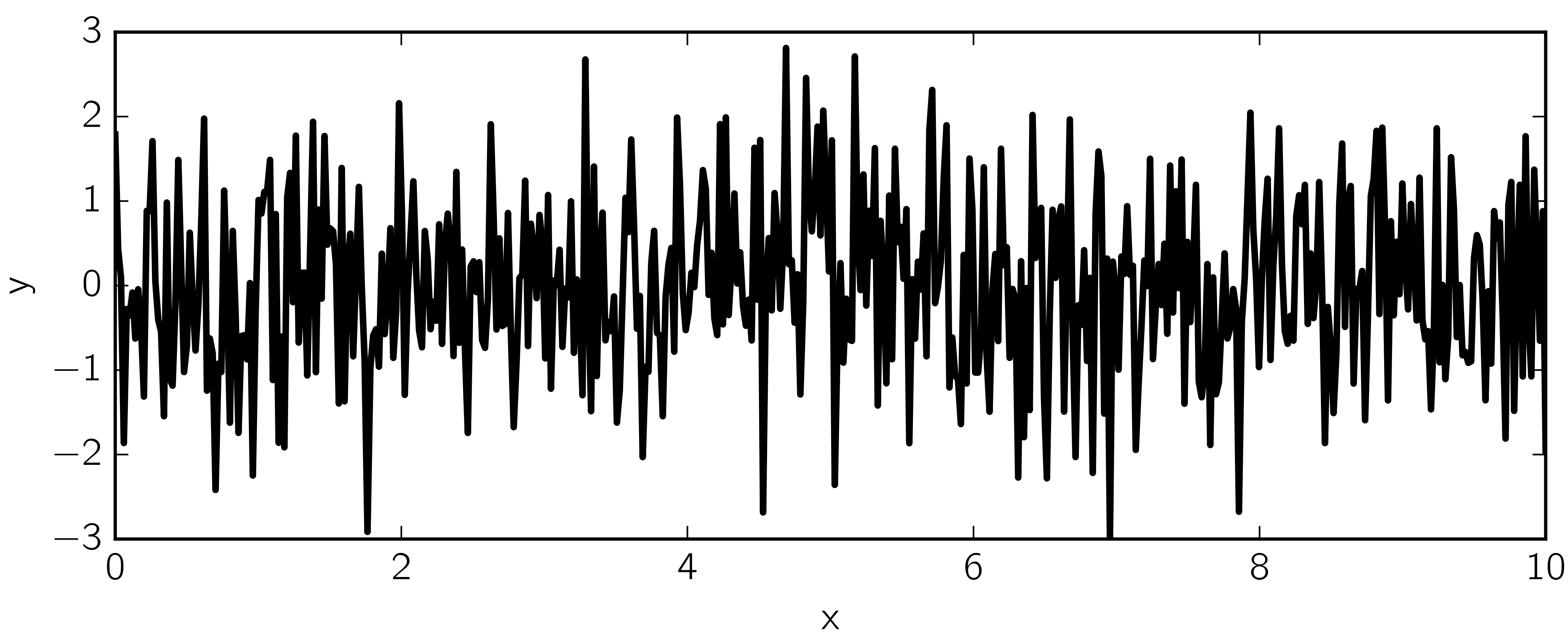
$$k_{\boldsymbol{\alpha}}(x_i, x_j) = \sigma\delta_{ij}$$



Gaussian processes

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

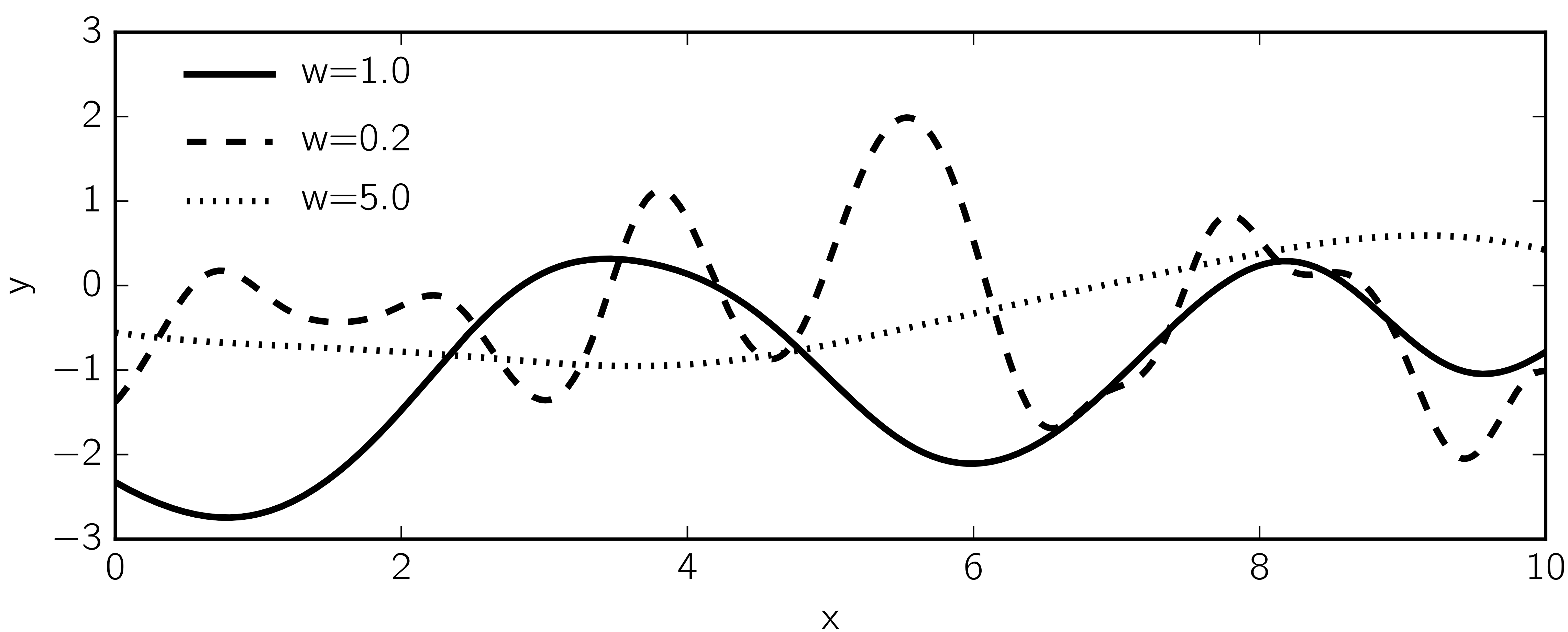
$$k_{\boldsymbol{\alpha}}(x_i, x_j) = \sigma\delta_{ij}$$



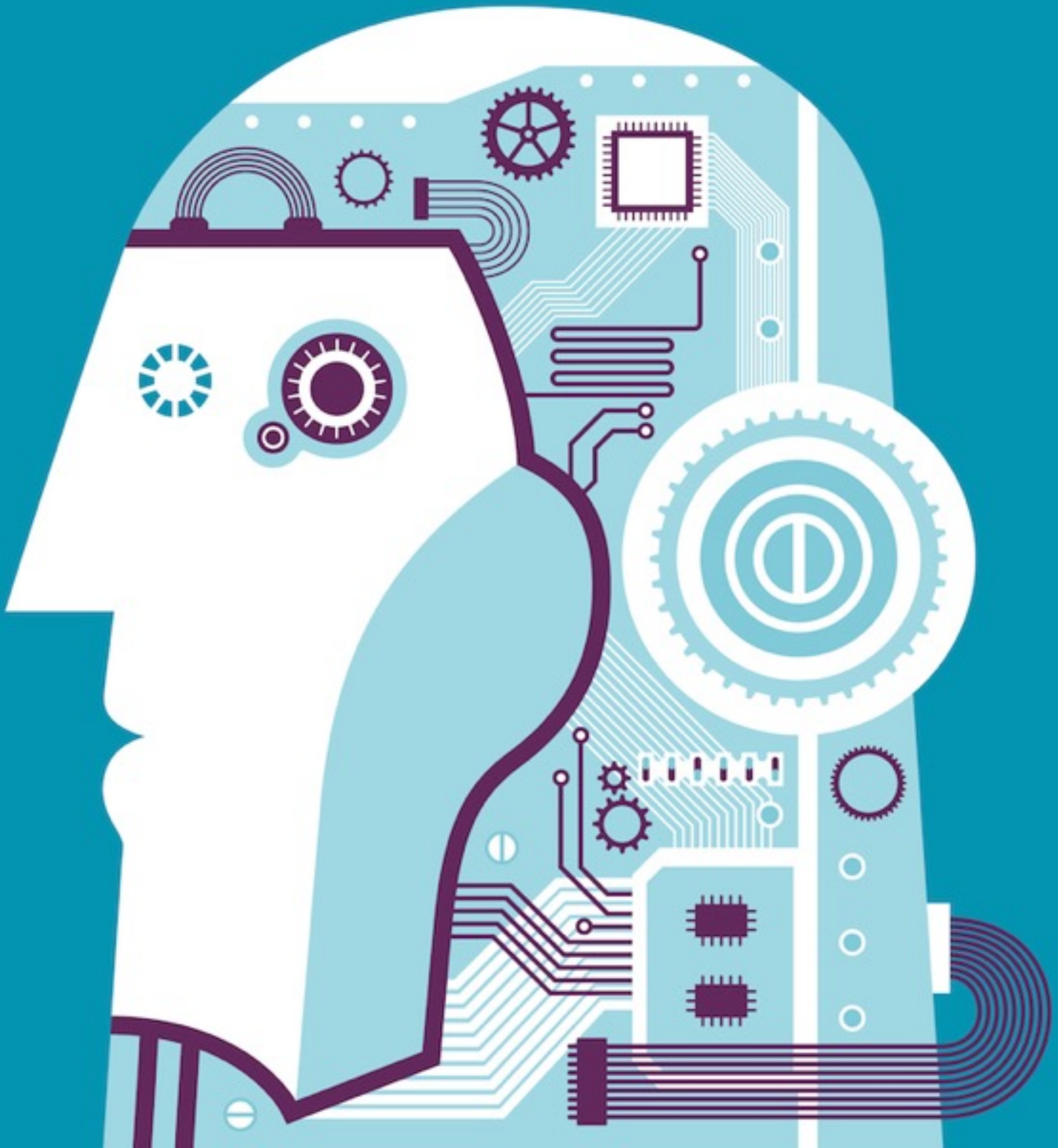
Gaussian processes

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

$$k_{\boldsymbol{\alpha}}(x_i, x_j) = A^2 \exp\left(-\frac{(x_i - x_j)^2}{2w^2}\right)$$



Machine learning…



Conditional probability

$$\begin{bmatrix} \mathrm{f} \\ \mathrm{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_{\boldsymbol{\theta}}(\mathrm{x}) \\ m_{\boldsymbol{\theta}}(\mathrm{x}_*) \end{bmatrix}, \begin{bmatrix} K(\mathrm{x},\mathrm{x}) + \boldsymbol{\sigma}^2 I & K(\mathrm{x},\mathrm{x}_*) \\ K(\mathrm{x}_*,\mathrm{x}) & K(\mathrm{x}_*,\mathrm{x}_*) \end{bmatrix} \right)$$

$$\Downarrow$$

$$\mathrm{f}_* | \mathrm{f}, \mathrm{x}, \mathrm{x}_* \sim \mathcal{N}\left(\bar{\mathrm{f}}_*, \mathrm{cov}(\mathrm{f}_*)\right)$$

$$\bar{\mathrm{f}}_* = \mathrm{m}_{\boldsymbol{\theta}}(\mathrm{x}_*) + K(\mathrm{x}_*,\mathrm{x}) \left[K(\mathrm{x},\mathrm{x}) + \boldsymbol{\sigma}^2 I\right]^{-1} (\mathrm{f} - \mathrm{m}_{\boldsymbol{\theta}}(\mathrm{x}))$$

$$\mathrm{cov}(\mathrm{f}_*) = K(\mathrm{x}_*,\mathrm{x}_*) - K(\mathrm{x}_*,\mathrm{x}) \left[K(\mathrm{x},\mathrm{x}) + \boldsymbol{\sigma}^2 I\right]^{-1} K(\mathrm{x},\mathrm{x}_*)$$

$$\Downarrow$$

$$\ln p(\mathrm{f}|\mathrm{x}, \boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = -\frac{1}{2} (\mathrm{f} - \mathrm{m}_{\boldsymbol{\theta}}(\mathrm{x}))^T \left[K_{\boldsymbol{\alpha}} + \boldsymbol{\sigma}^2 I\right]^{-1} (\mathrm{f} - \mathrm{m}_{\boldsymbol{\theta}}(\mathrm{x})) - \frac{1}{2} \ln |K_{\boldsymbol{\alpha}} + \boldsymbol{\sigma}^2 I| - \frac{N}{2} \ln 2\pi$$

Conditional probability

$$\chi^2 = \sum_{i=1}^{N} \frac{(f_i - m_i^{\boldsymbol{\theta}})^2}{\sigma_i^2}$$

$$\ln p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\sigma},\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{r}^T \mathrm{C}^{-1}\boldsymbol{r} - \frac{1}{2}\ln\det\mathrm{C} - \frac{N}{2}\ln(2\pi)$$

↑

$$\ln p(\mathrm{f}|\mathrm{x},\boldsymbol{\sigma},\boldsymbol{\theta},\boldsymbol{\alpha}) = -\frac{1}{2}\left(\mathrm{f} - \mathrm{m}_{\boldsymbol{\theta}}(\mathrm{x})\right)^T \left[K_{\boldsymbol{\alpha}} + \boldsymbol{\sigma}^2 I\right]^{-1}\left(\mathrm{f} - \mathrm{m}_{\boldsymbol{\theta}}(\mathrm{x})\right)$$

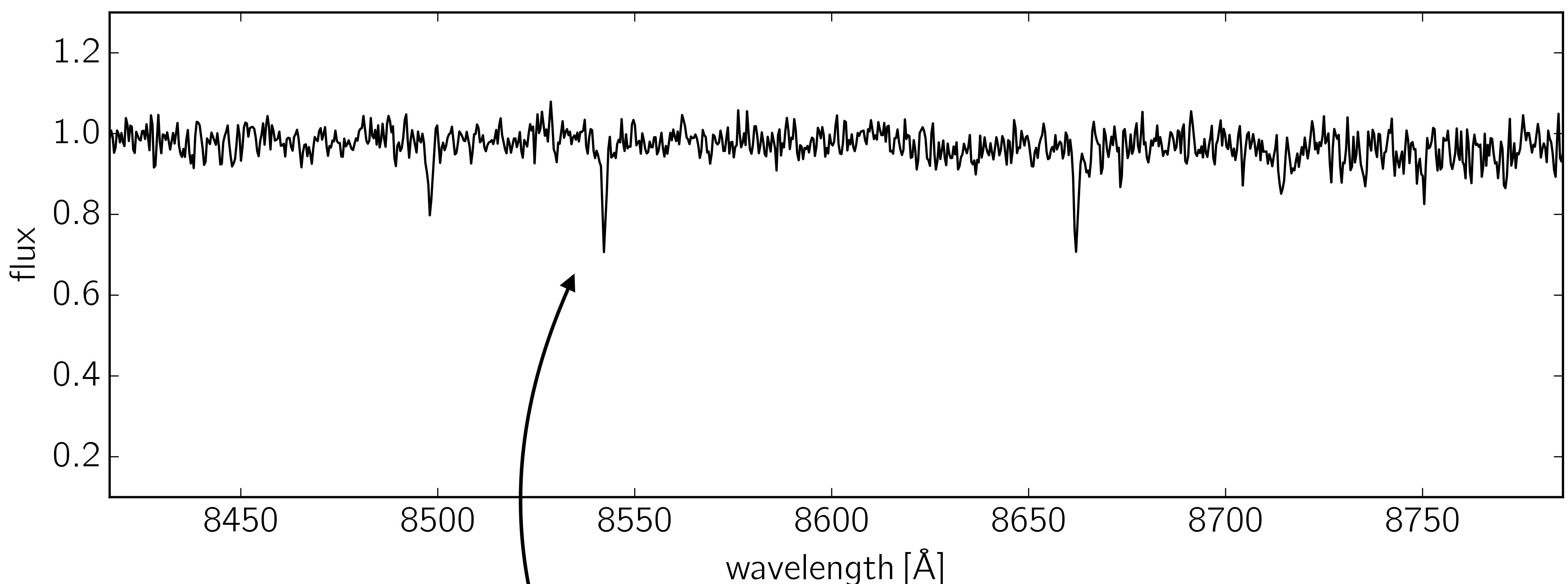
$$-\frac{1}{2}\ln\left|K_{\boldsymbol{\alpha}} + \boldsymbol{\sigma}^2 I\right| - \frac{N}{2}\ln 2\pi$$

3

The solution

AIP

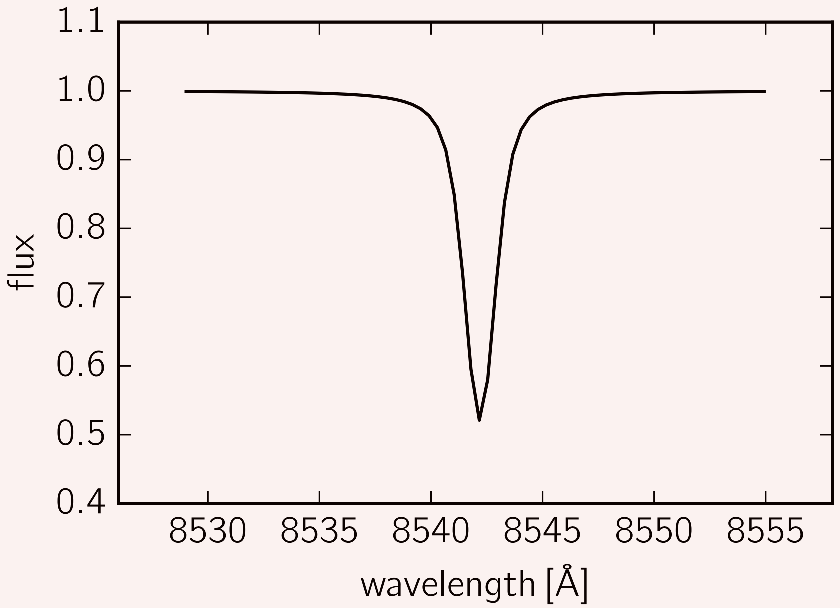
Extremely metal-poor star



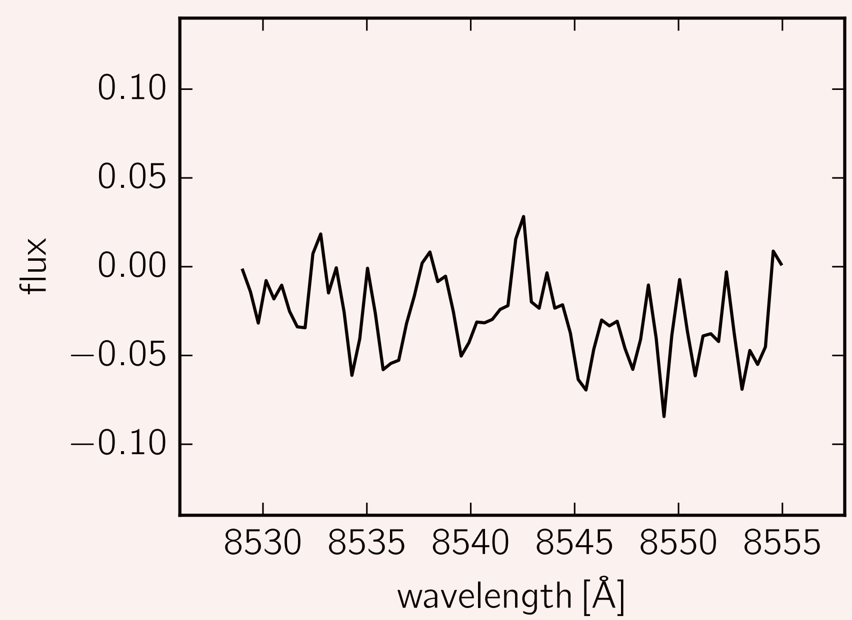
Let’s fit one of the lines!

(Fit = run an MCMC chain on model and noise parameters)

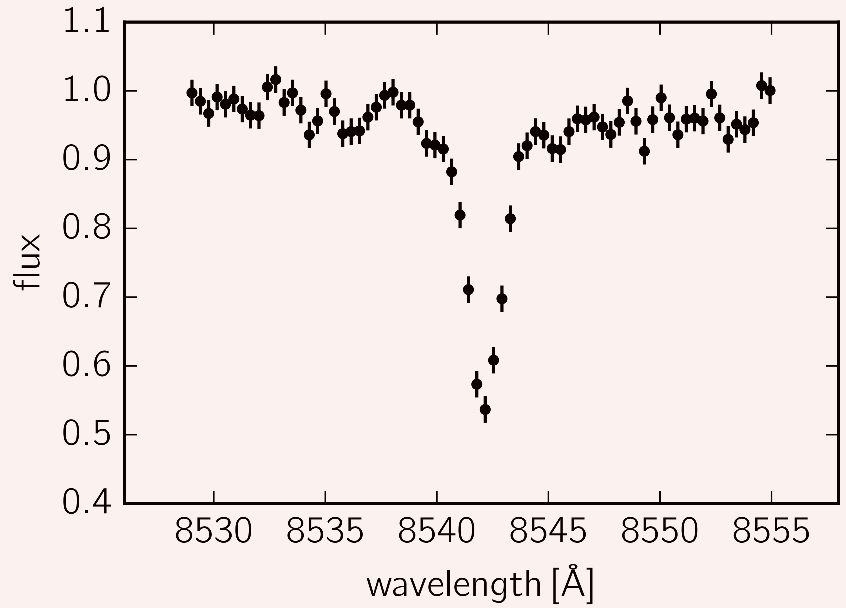
Line



Noise

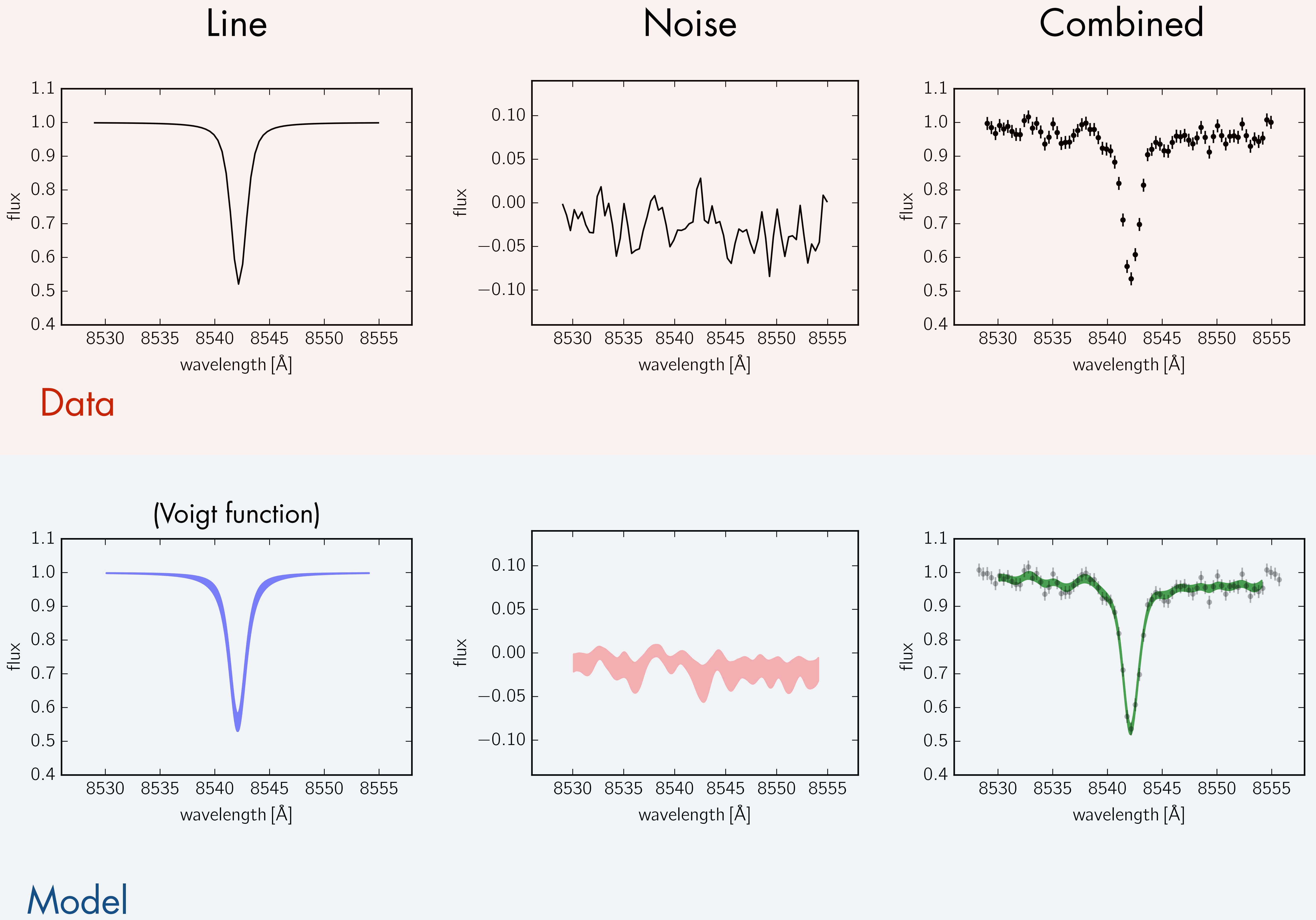


Combined

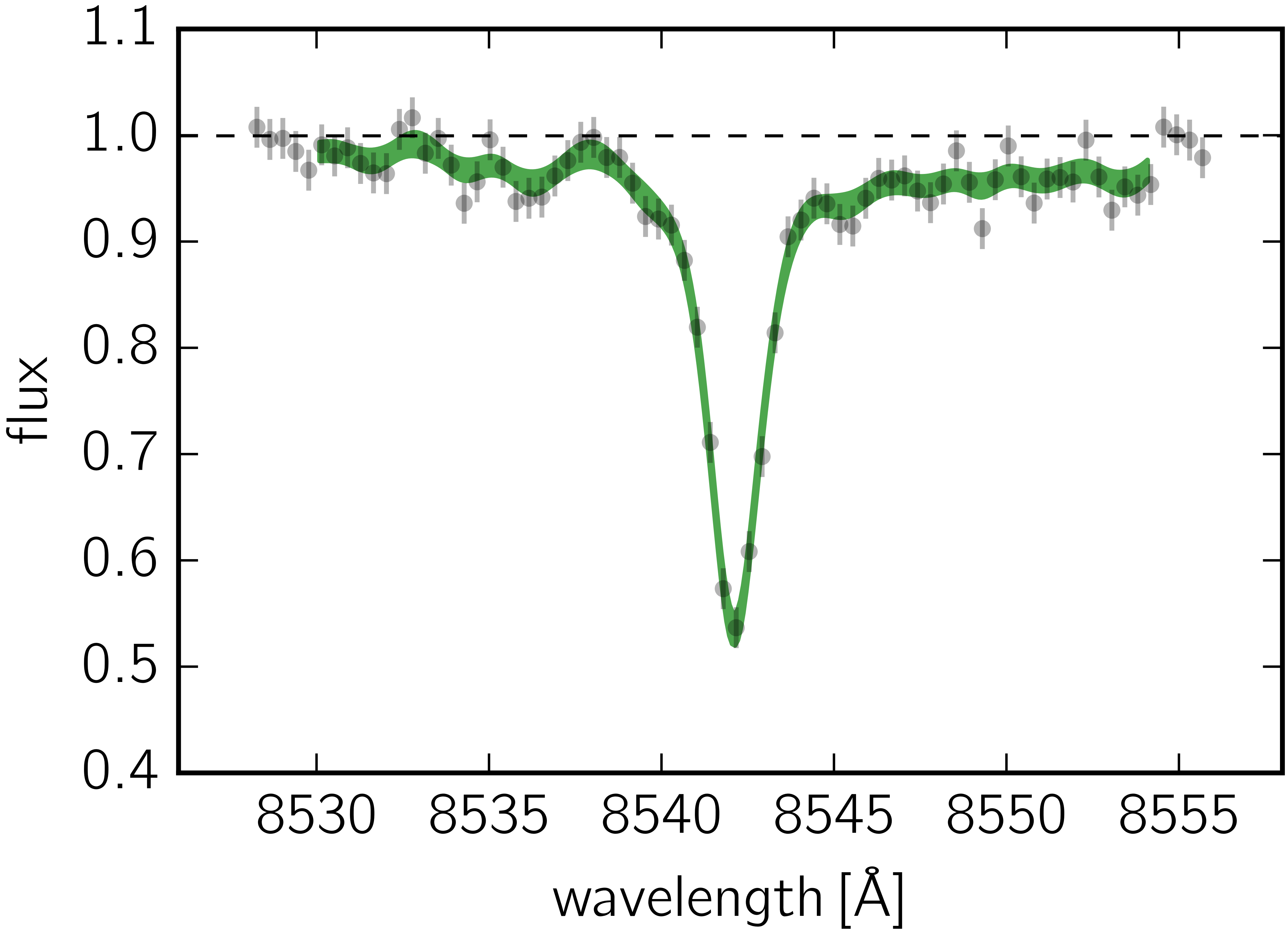


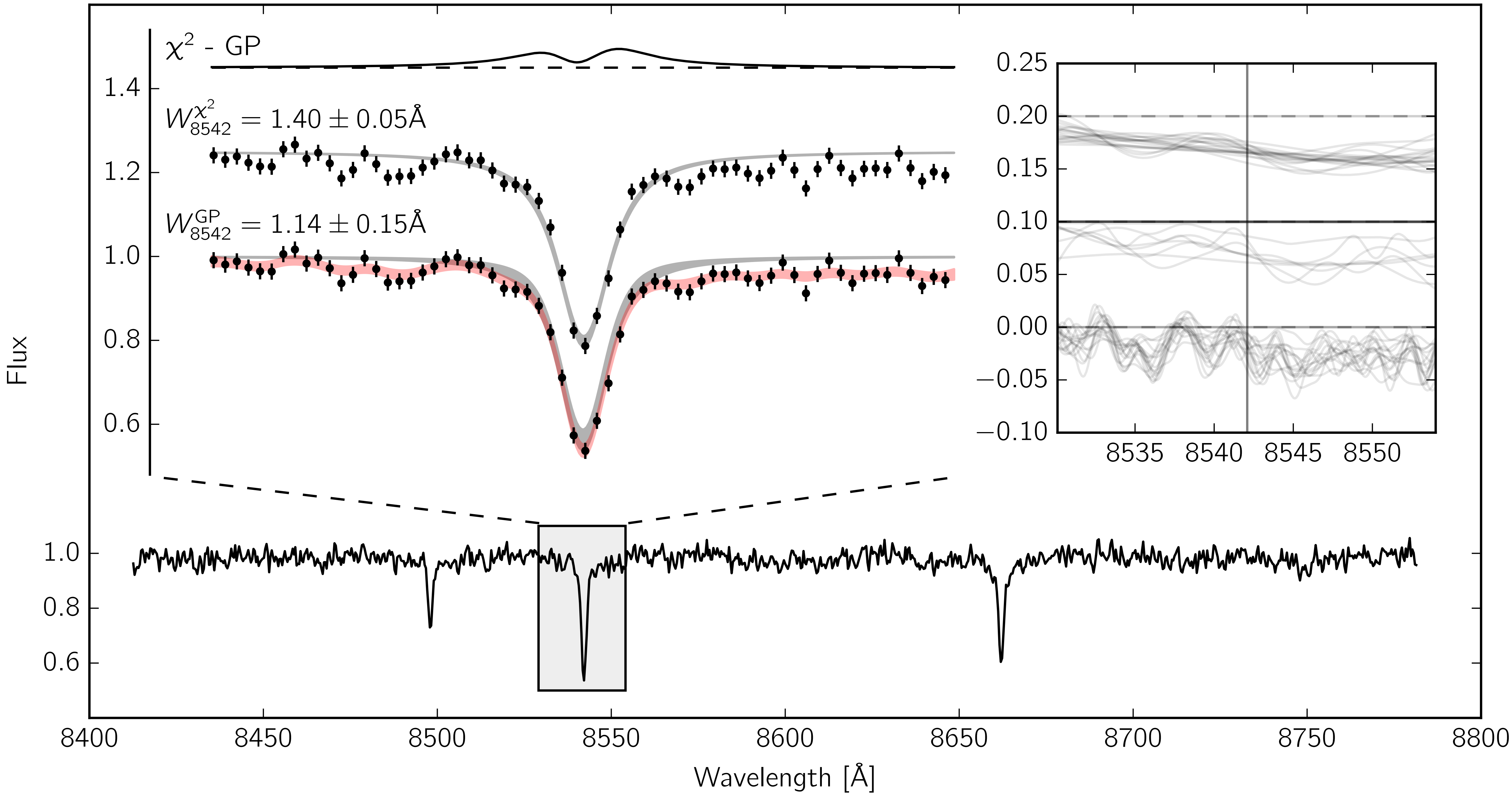
Data

AIP



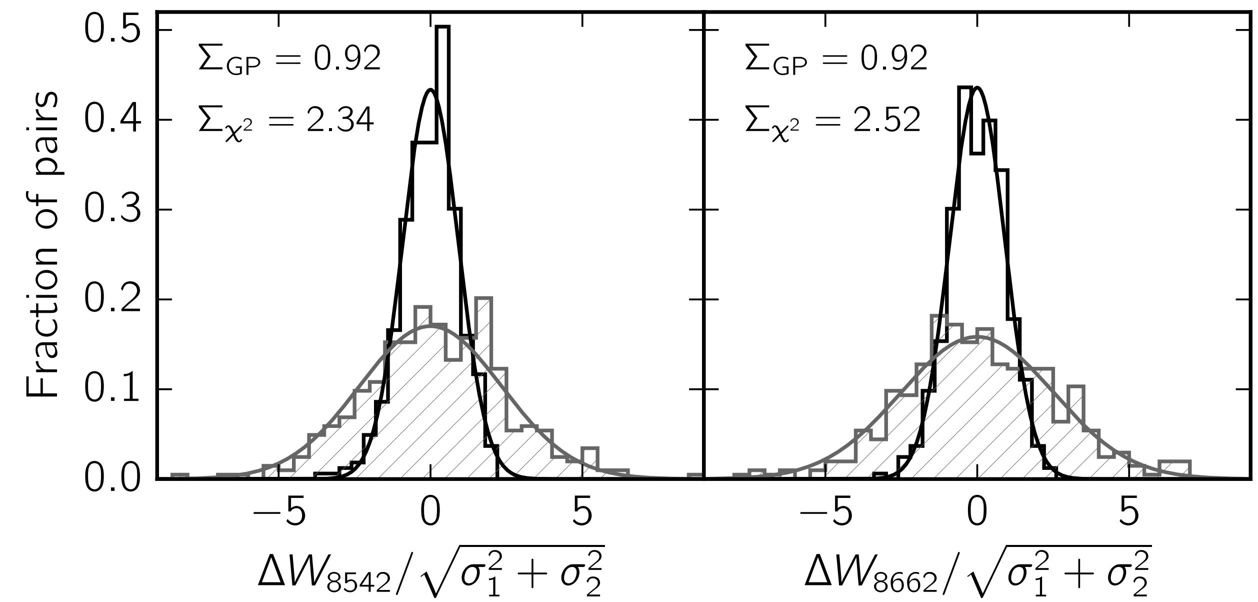
Line
Noise
Combined
flux
wavelength [Å]
8530 8535 8540 8545 8550 8555
1.1 1.0 0.9 0.8 0.7 0.6 0.5 0.4
0.10 0.05 0.00 −0.05 −0.10
Data
(Voigt function)
Model


1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
flux
8530
8535
8540
8545
8550
8555
wavelength [Å]

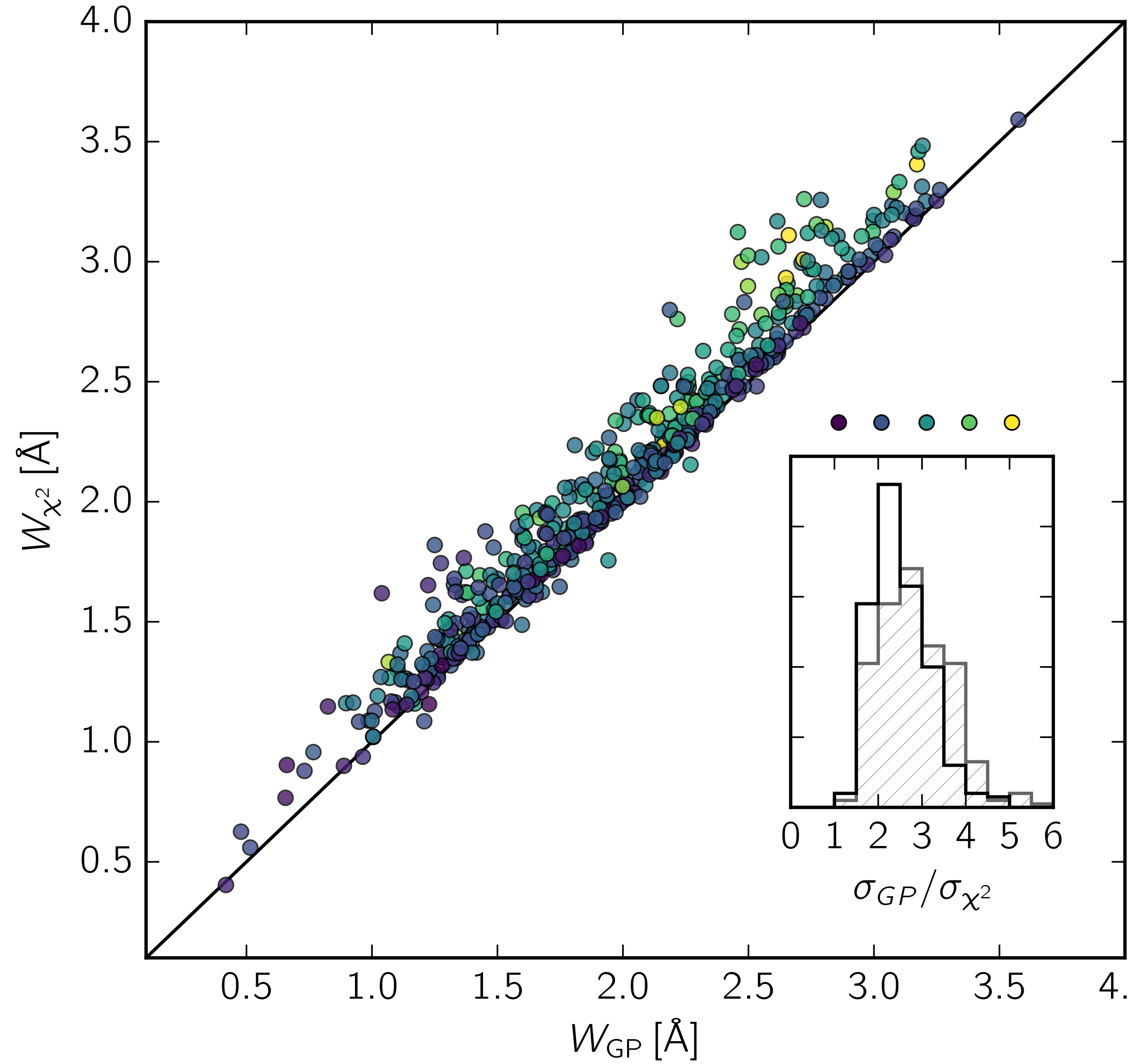


χ² - GP
$W_{8542}^{\chi^2} = 1.40 \pm 0.05$Å
$W_{8542}^{GP} = 1.14 \pm 0.15$Å
Flux
Wavelength [Å]

Repeated observations
(pairwise comparison)



Equivalent widths
GP vs. χ^2



AIP

Rasmussen & Williams 2006, MIT Press

Gaussian Processes for Machine Learning

gaussianprocess.org

dan.iel.fm/george, scikit-learn.org,
github.com/SheffieldML/GPy

AIP