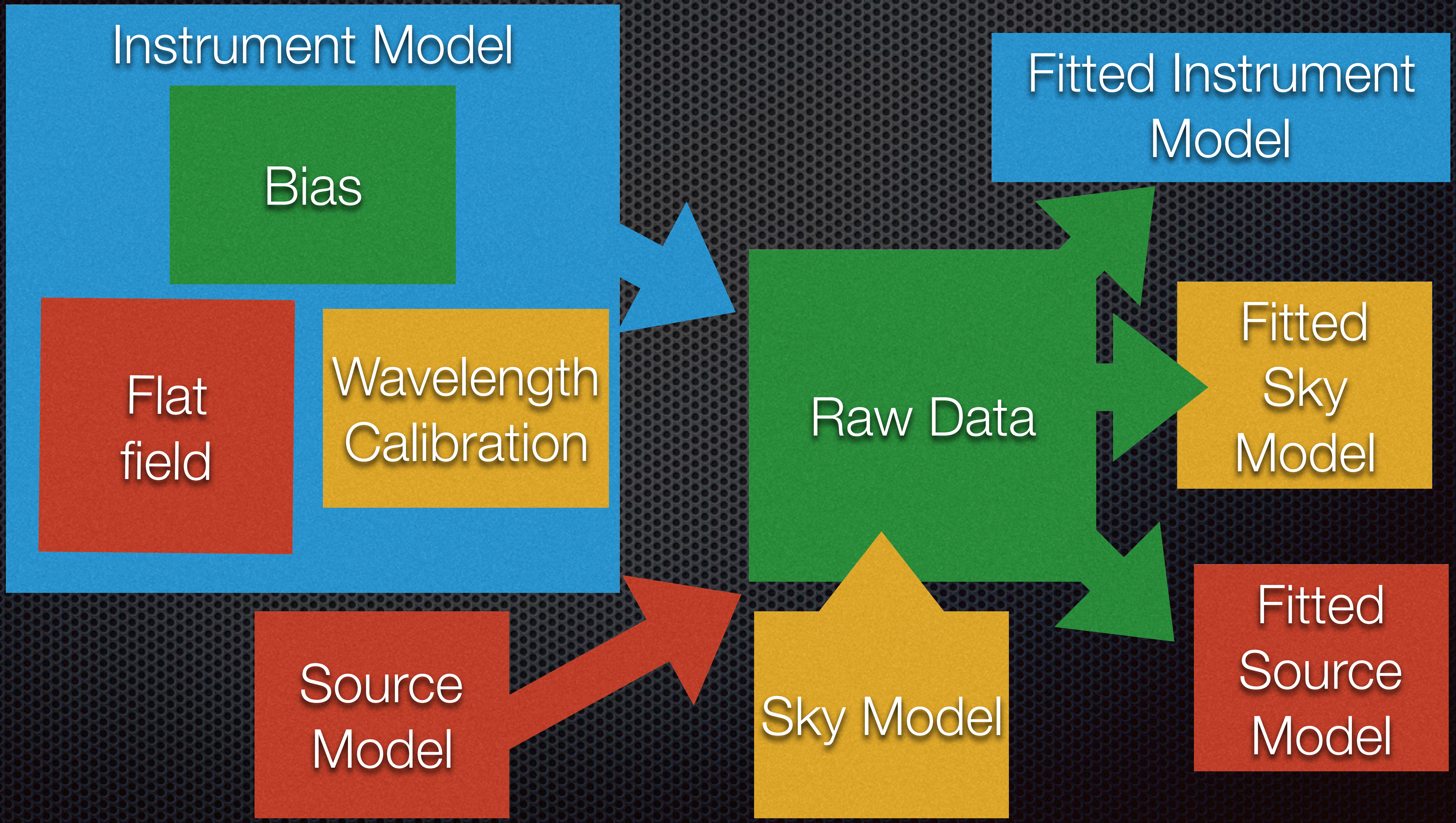
Beyond Spectral Extraction

Modeling of “raw” spectral data

Wolfgang Kerzendorf
ESO Fellow
as part of the Science Data Groups

How we currently think about data reduction

How we might want to think about data reduction



Why?????

There were way more arrows!

There will be 100s of Parameters

Are you mad?

Billions upon billions of Parameters

Billions upon billions of Parameters

- Throughput for each Pixel
  - (millions of Parameters)

Billions upon billions of Parameters

Throughput for each Pixel
(millions of Parameters)

Intensity for Source on
each spectral channel
(thousands)

Billions upon billions of Parameters

Throughput for each Pixel
(millions of Parameters)

Intensity for Source on
each spectral channel
(thousands)

Intensity of night
sky lines
(hundreds)

Billions upon billions of Parameters

Throughput for each Pixel
(millions of Parameters)

Intensity for Source on
each spectral channel
(thousands)

Intensity of night
sky lines
(hundreds)

transformation between pixel space
and wavelength, slit positions
(polynomials, >10 of parameters)

Priors, Computing Power and smart Algorithms

- Computing power virtually limitless

2000 (SDSS) vs 2022 (LSST)

by Geert Barentsen

2000 (SDSS) vs 2022 (LSST)

- • 81x more pixels;

by Geert Barentsen

2000 (SDSS) vs 2022 (LSST)

- • 81x more pixels;
- • 190x faster I/O;

by Geert Barentsen

2000 (SDSS) vs 2022 (LSST)

- • 81x more pixels;
- • 190x faster I/O;
- • 2565x more storage;

by Geert Barentsen

2000 (SDSS) vs 2022 (LSST)

- 81x more pixels;
- 190x faster I/O;
- 2565x more storage;
- 120 578x faster internet;

by Geert Barentsen

2000 (SDSS) vs 2022 (LSST)

- • 81x more pixels;
- • 190x faster I/O;
- • 2565x more storage;
- • 120 578x faster internet;
- • 1 202 277x faster supercomputers

by Geert Barentsen

Priors, Computing Power and smart Algorithms

Priors, Computing Power and smart Algorithms

- Computing power virtually limitless

Priors, Computing Power and smart Algorithms

- Computing power virtually limitless
- Many parameters BUT many informative priors

Priors, Computing Power and smart Algorithms

- Computing power virtually limitless
- Many parameters BUT many informative priors
  - Instruments relatively stable

Priors, Computing Power and smart Algorithms

- Computing power virtually limitless
- Many parameters BUT many informative priors
  - Instruments relatively stable
  - calibration plan for observations

Priors, Computing Power and smart Algorithms

- Computing power virtually limitless
- Many parameters BUT many informative priors
  - Instruments relatively stable
  - calibration plan for observations
- Parallel optimizers in many dimensions

The possibilities

The possibilities

- A robust way to obtain science products including uncertainties

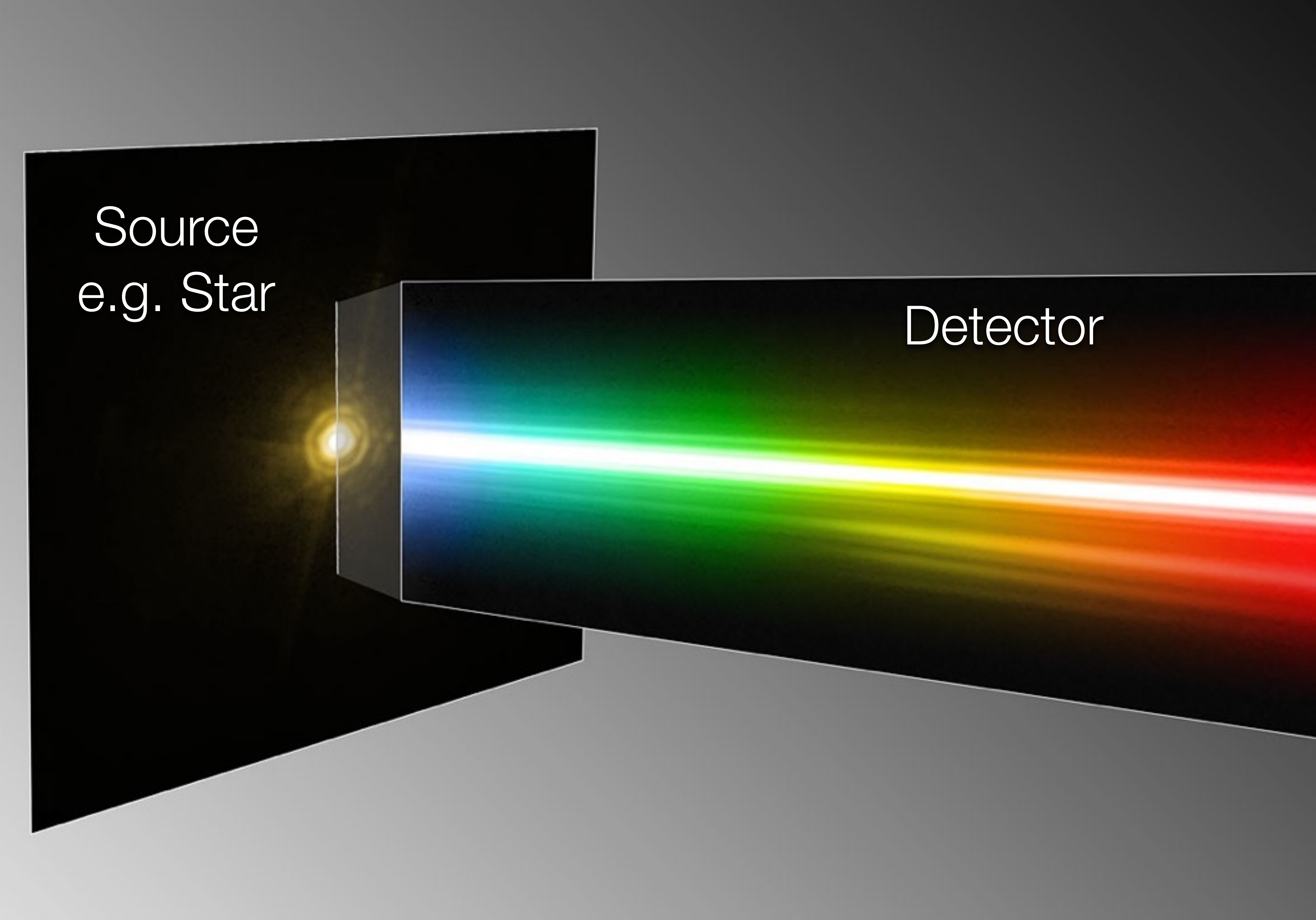
The possibilities

- A robust way to obtain science products including uncertainties
- A way to monitor instrument health

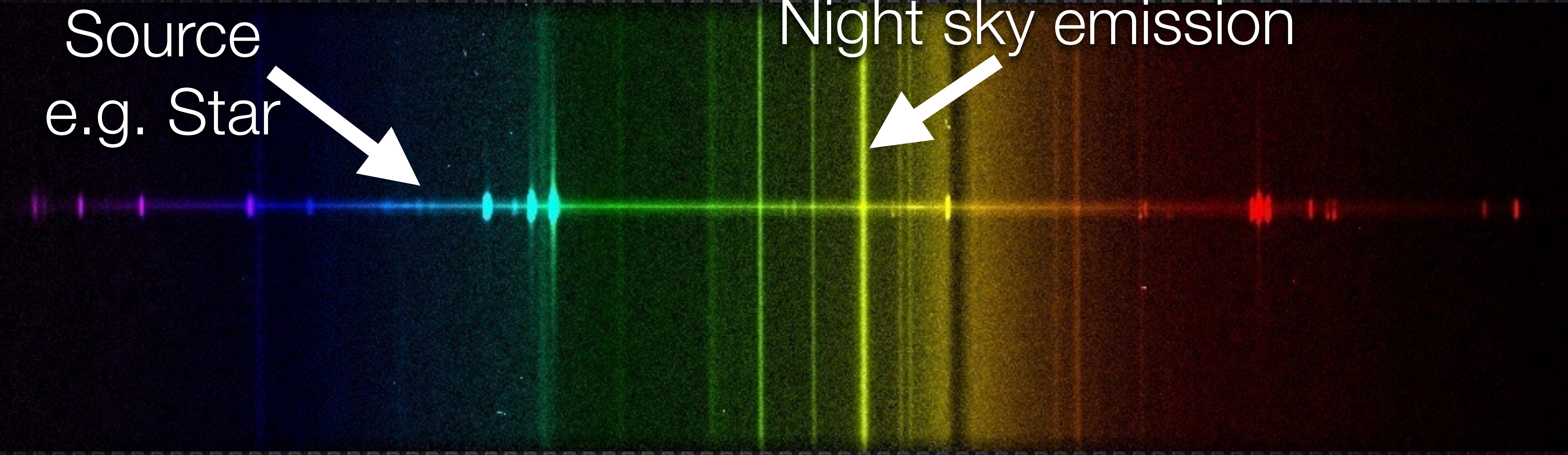
The possibilities

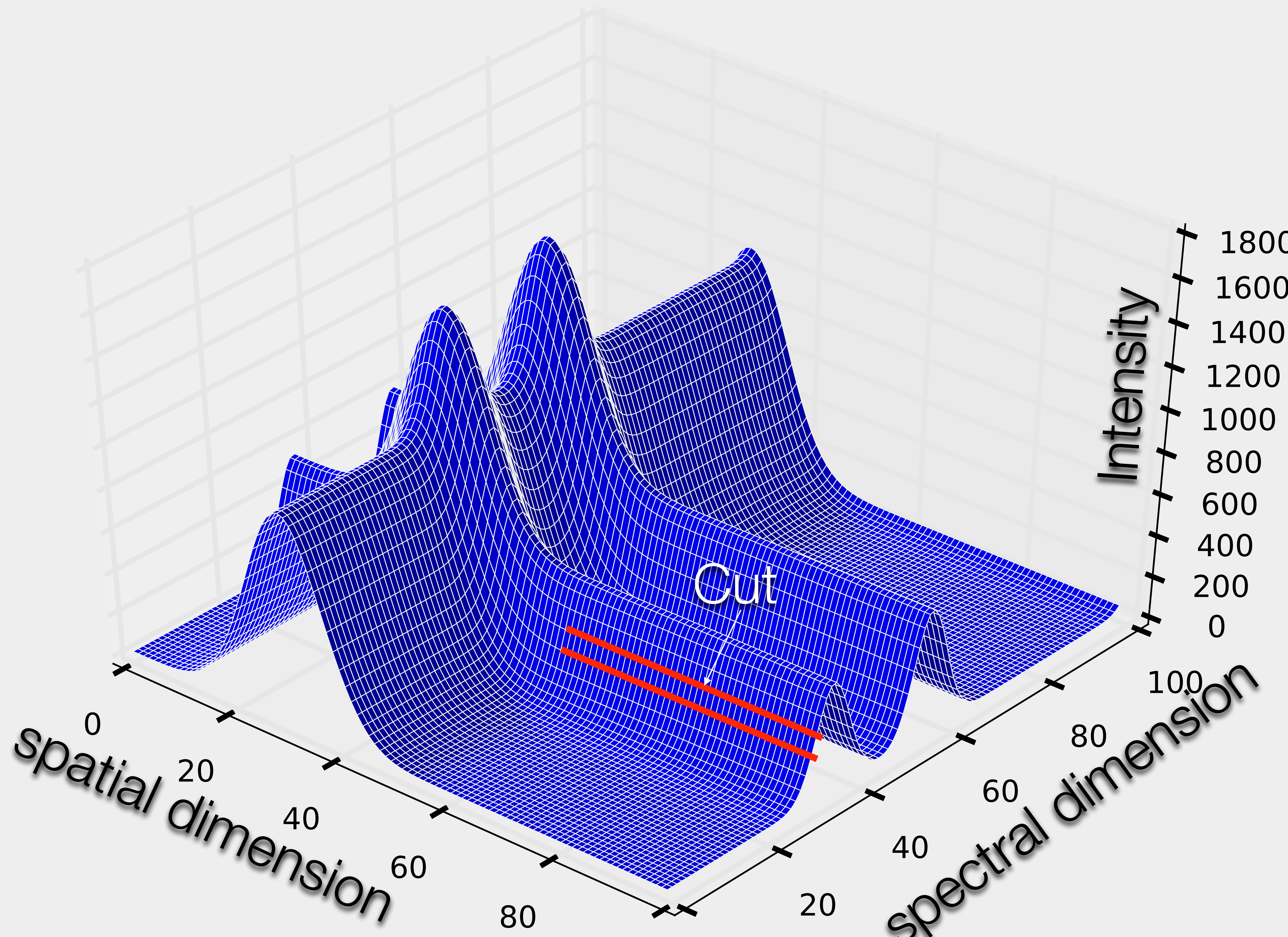
- A robust way to obtain science products including uncertainties
- A way to monitor instrument health
- Easy rejection of cosmics, bad pixels, etc.

Example to model longslit data

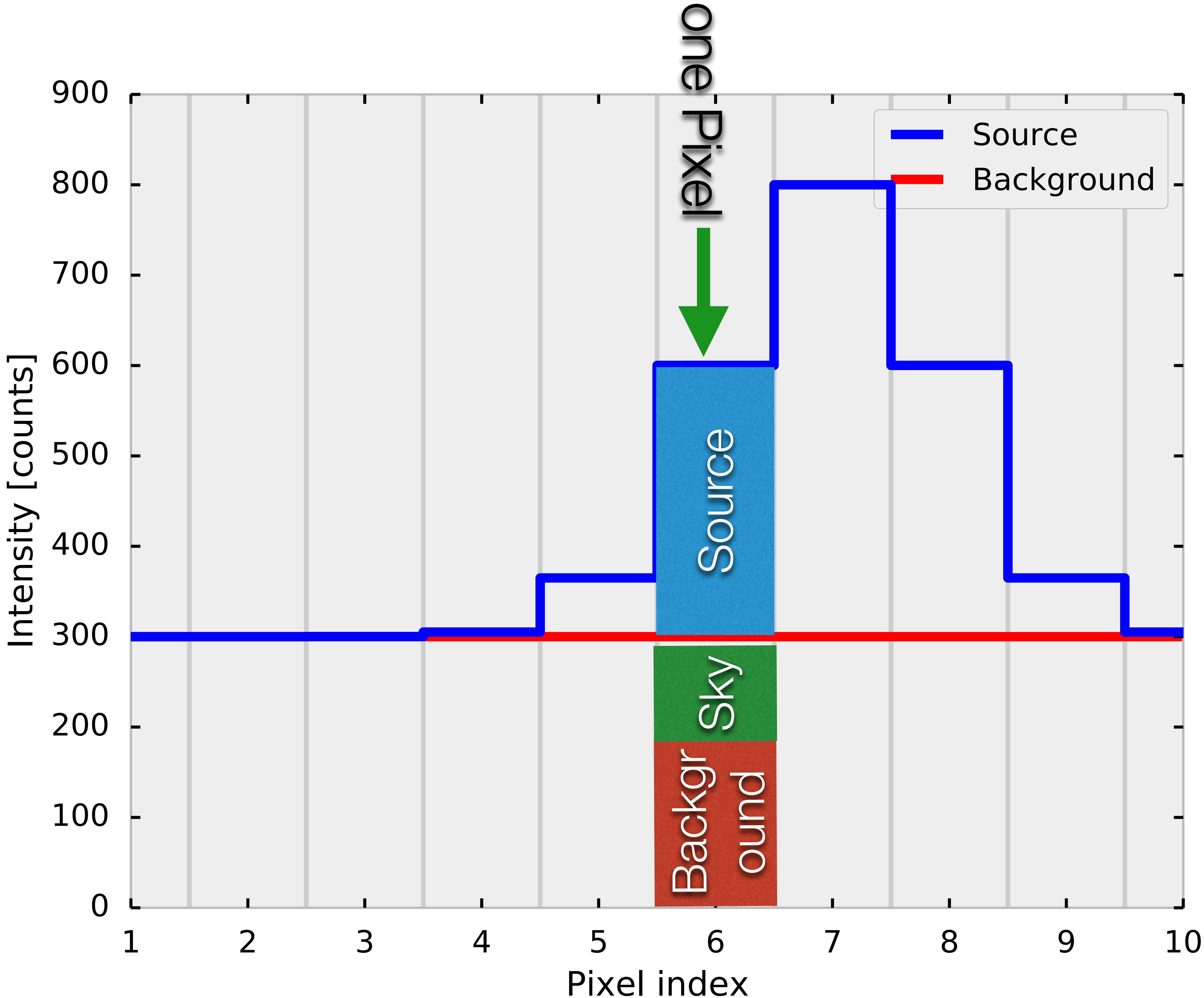


https://briankoberlein.com/wp-content/uploads/eso1002d.jpg


Source
e.g. Star
Night sky emission



Cut
Intensity
0
200
400
600
800
1000
1200
1400
1600
spatial dimension
0
20
40
60
80
spectral dimension
20
40
60
80
100



one Pixel
Source
Background
Source
Sky
Backgr
ound
Intensity [counts]
0
100
200
300
400
500
600
700
800
900
1
2
3
4
5
6
7
8
9
10
Pixel index

A linear model for one pixel

$$
\begin{aligned}
&\text{pixel value}_i = \\
&c_{\text{background}} \times 1 + \\
&c_{\text{sky}} \times 1 + \\
&c_{\text{trace amplitude}} \times \underbrace{\exp\left(-(f(x_i, y_i) \mapsto s_{\text{slit pos}} - s_{\text{trace pos}})/(\text{FWHM}/2.355)^2\right)}_{m_{\text{trace},i}}
\end{aligned}
$$

non-linear components

A linear model for one pixel

$$\text{pixel value}_i =$$
$$c_{\text{background}} \times 1 +$$
$$c_{\text{sky}} \times 1 +$$
$$c_{\text{trace amplitude}} \times m_{\text{trace},i}$$

A linear model for one row

$$\underbrace{\begin{pmatrix} \text{pixel value}_1 \\ \text{pixel value}_2 \\ \text{pixel value}_3 \\ \vdots \\ \text{pixel value}_n \end{pmatrix}}_{b} = \underbrace{\begin{pmatrix} 1 & 1 & m_1 \\ 1 & 1 & m_2 \\ 1 & 1 & m_3 \\ \vdots & & \\ 1 & 1 & m_n \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} c_{\text{background}} & c_{\text{sky}} & c_{\text{trace amplitude}} \end{pmatrix}}_{c}$$

Linear Least squares

finding c that

$$\min ||Ac - b||$$

By solving this equation

$$(A^T A)c = A^T b$$

Matrix operations are fast!!

Fit the non-linear instrument model

$$f(x, y) \mapsto s_{\mathrm{slitposition}}, \lambda_{\mathrm{wavelength}}$$

including parameters such as:
polynomial for wavelength solution
polynomial for trace width
Parameter for linespread function
etc

Solve Linear least squares for each row (total ~ few ms)

Non-linear components and linear components

Non-linear components

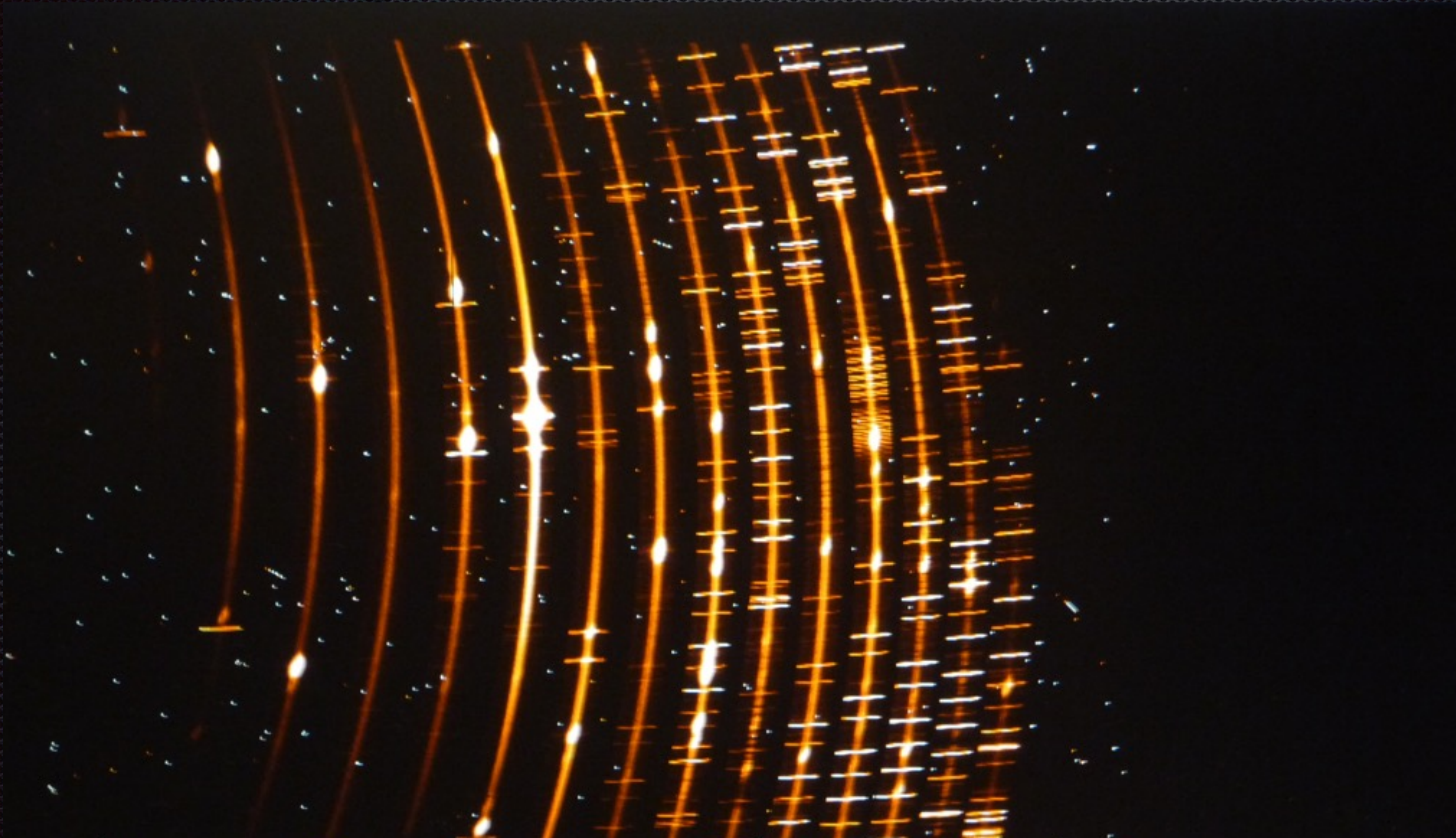
- position of source
- sigma of trace

Linear Components

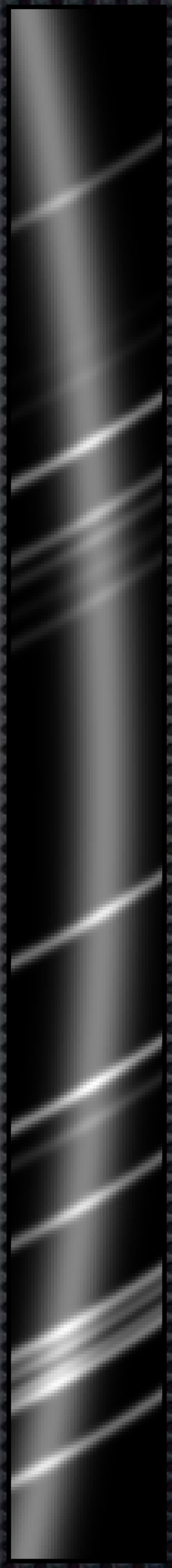
Amplitude of spectra at each row
background level at each row

Non - linear Fitter

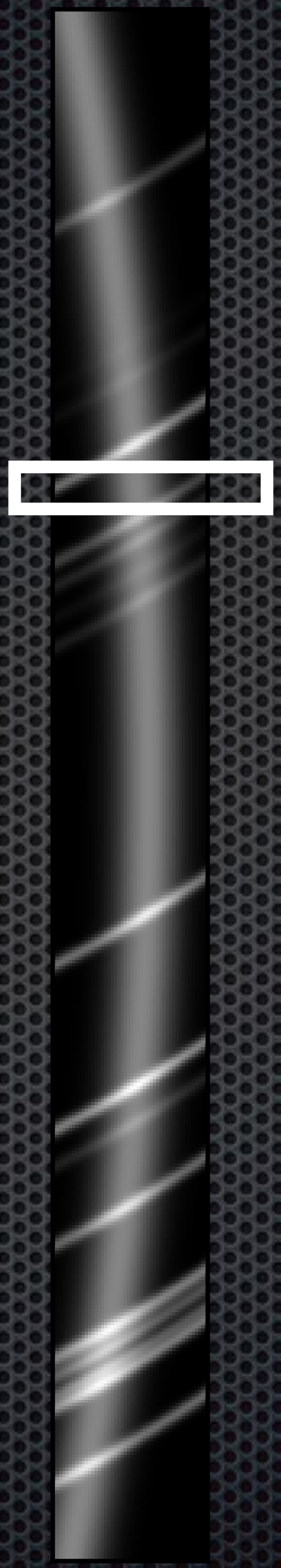
XShooter - a good prototype for modeling



Background + Sky + Source



Looking at one row



Looking at one row

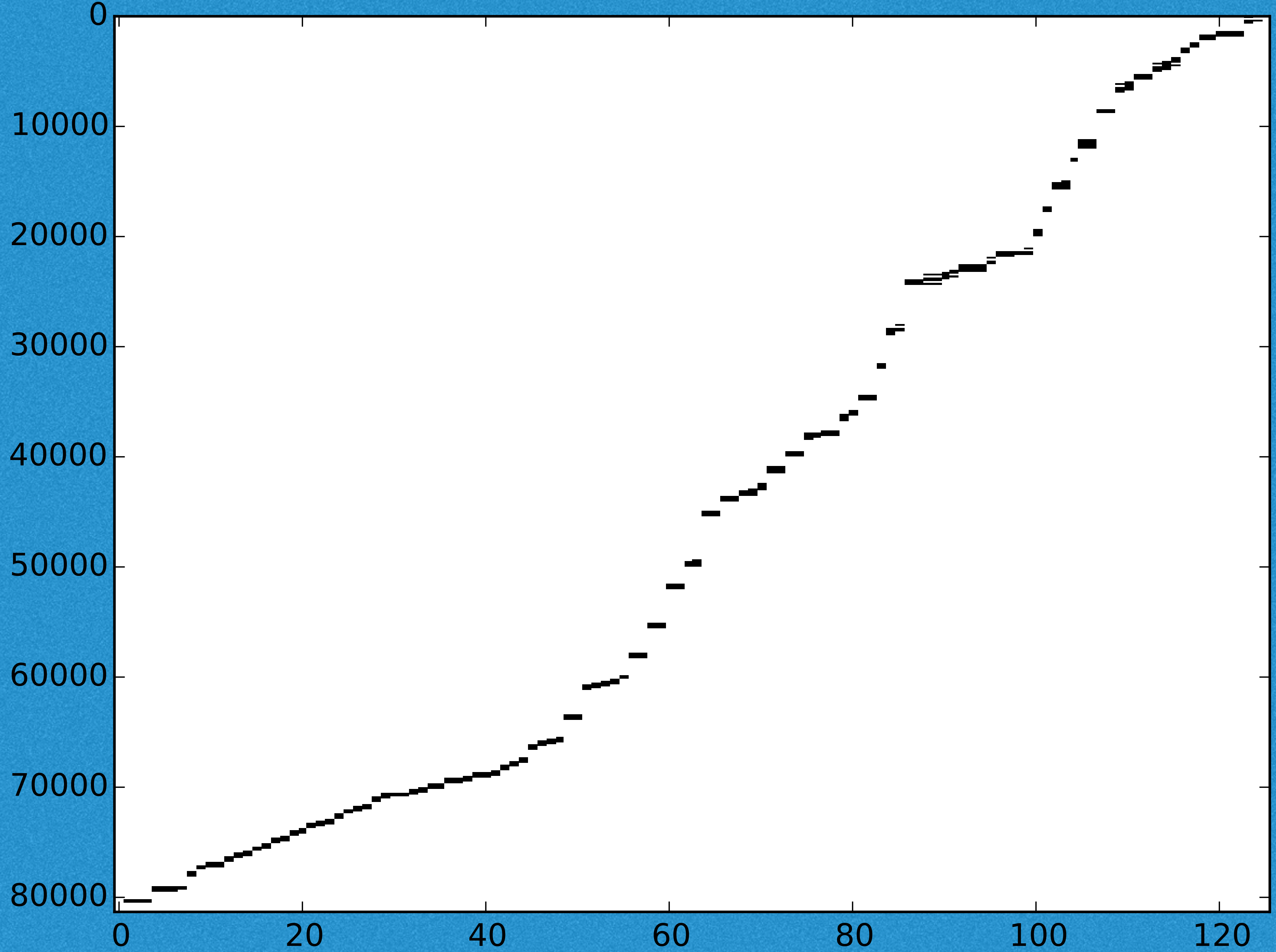
is not that sensible anymore

Looking at all pixels at the same time???

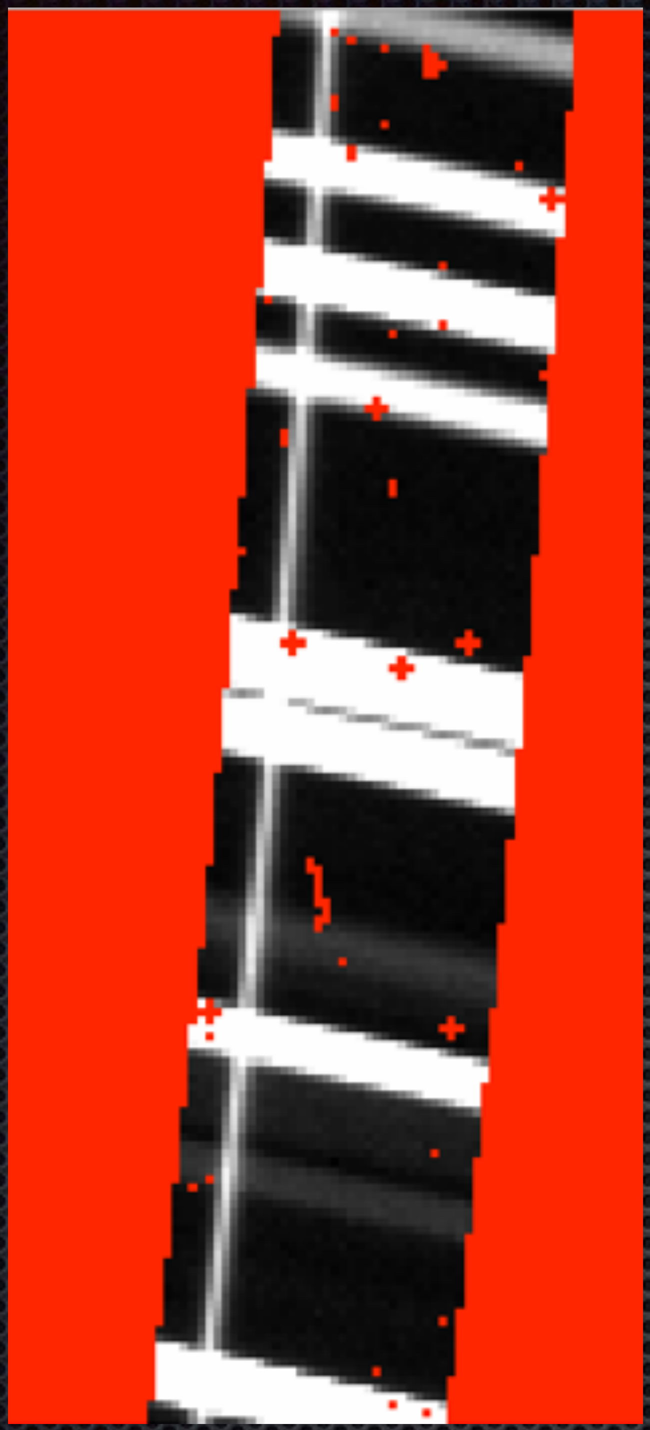
$$\underbrace{\begin{pmatrix} \text{pixel value}_1 \\ \text{pixel value}_2 \\ \text{pixel value}_3 \\ \vdots \\ \text{pixel value}_n \end{pmatrix}}_{b} = \underbrace{\begin{pmatrix} 1 & 1 & m_1 \\ 1 & 1 & m_2 \\ 1 & 1 & m_3 \\ \vdots & & \\ 1 & 1 & m_n \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} c_{\text{background}} & c_{\text{sky}} & c_{\text{trace amplitude}} \end{pmatrix}}_{c}$$

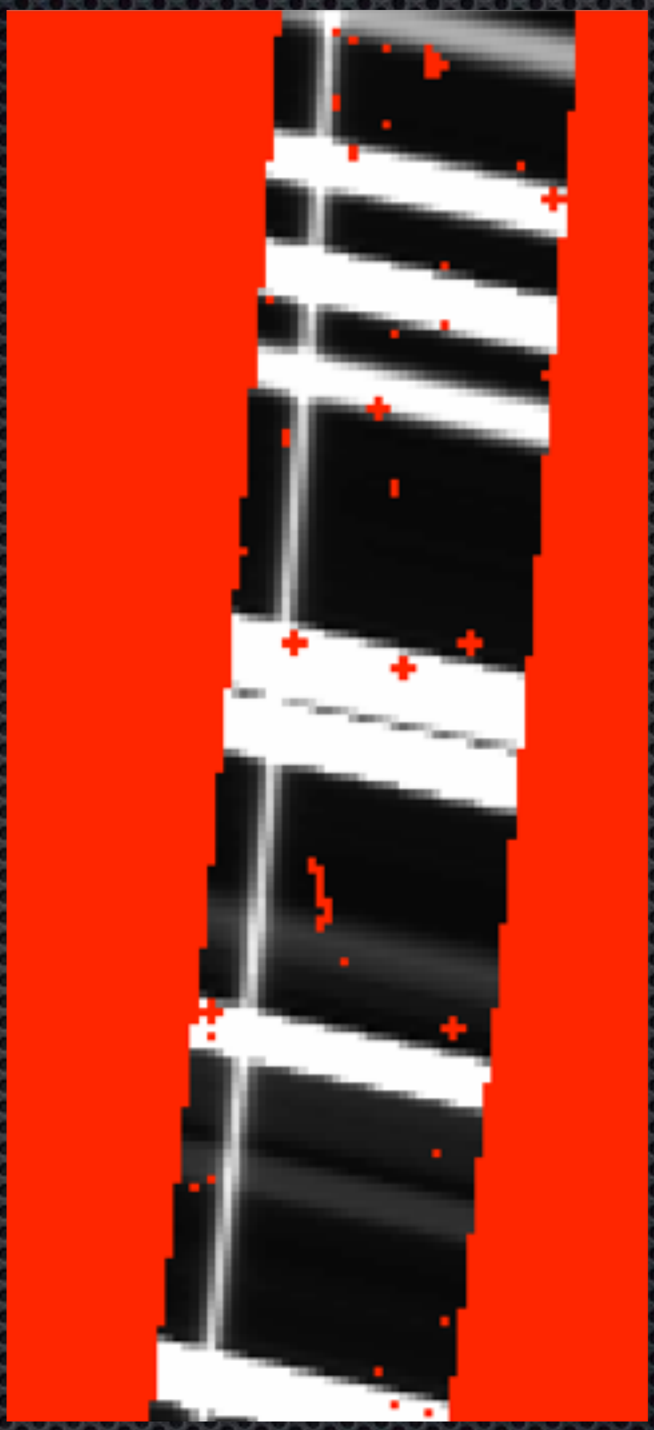
For this example n= 1000 * 100 in just rows
c will be one c_sky and one c_trace amplitude
for each iso-wavelength contour

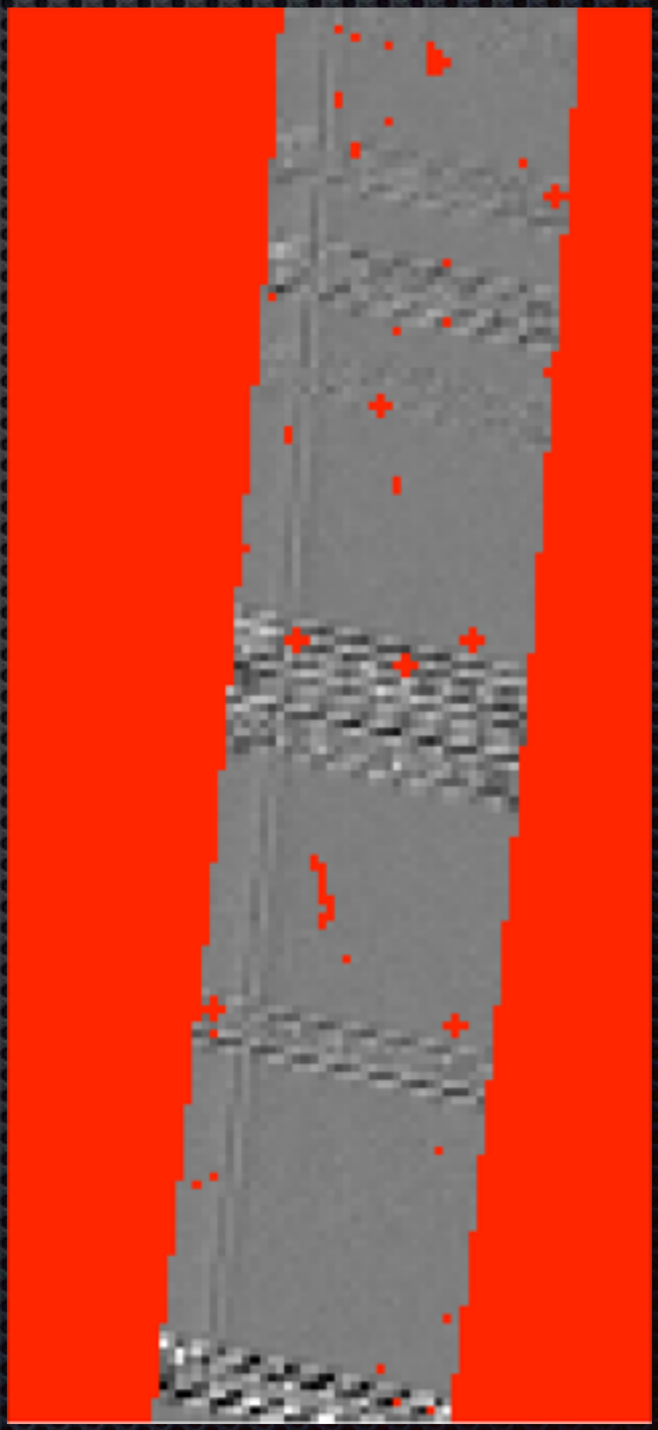
Big matrices don't cry if they are sparse



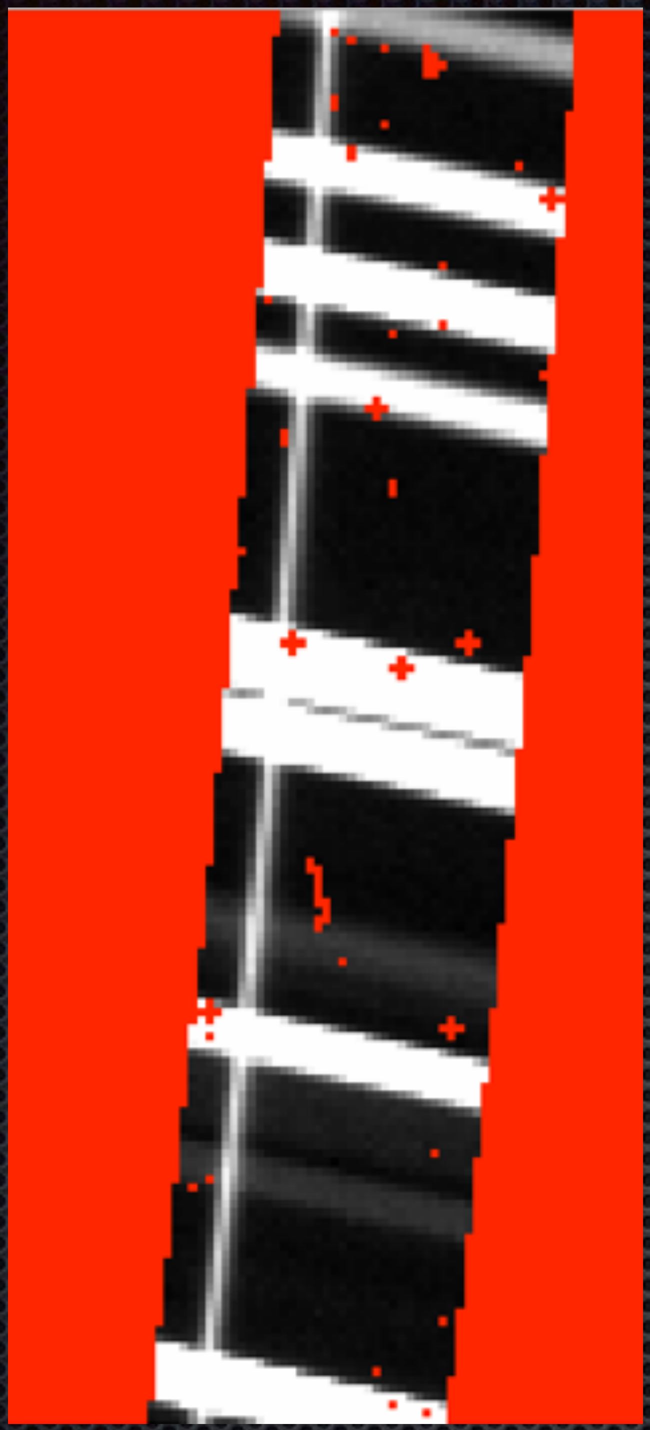
The results

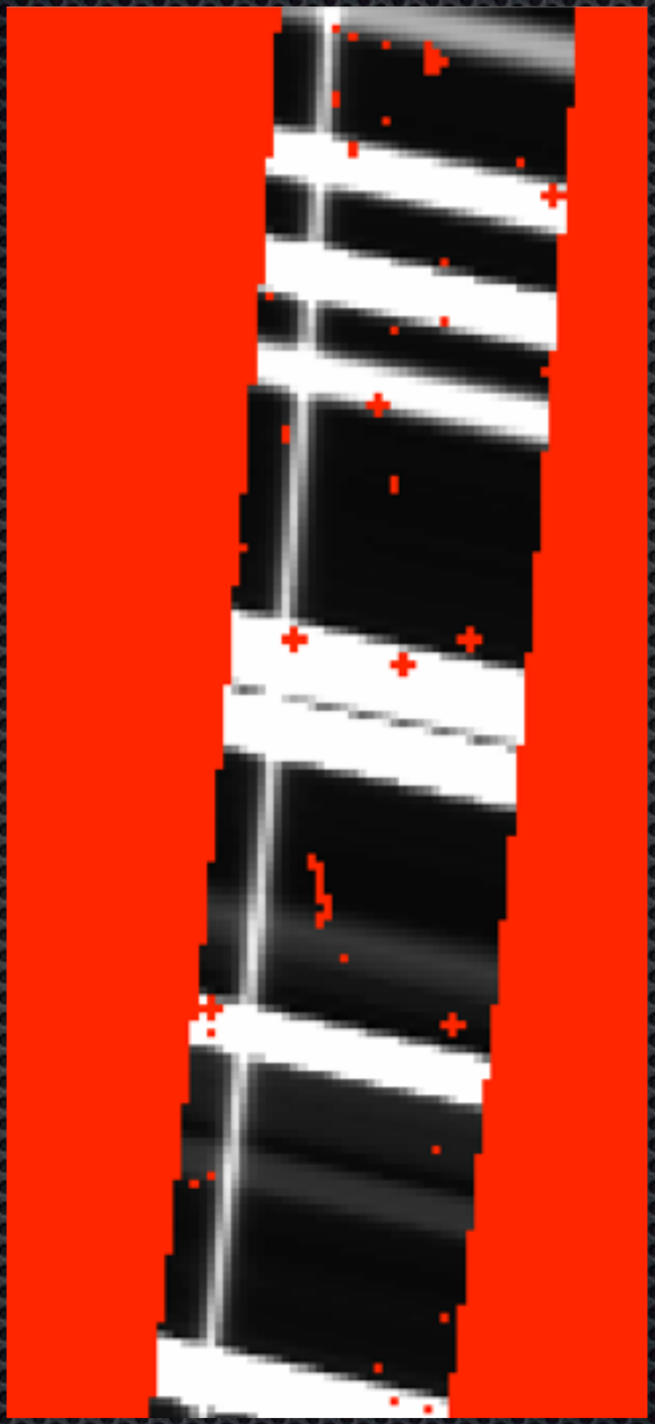




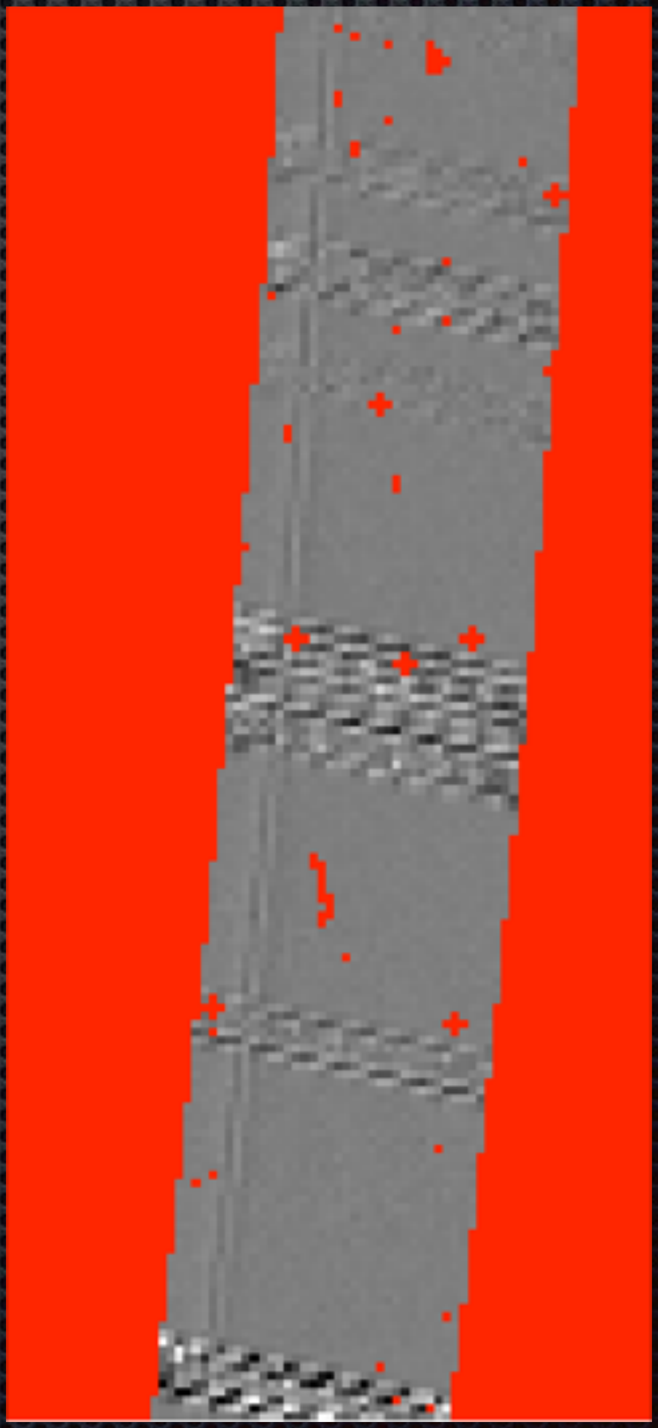


Residual

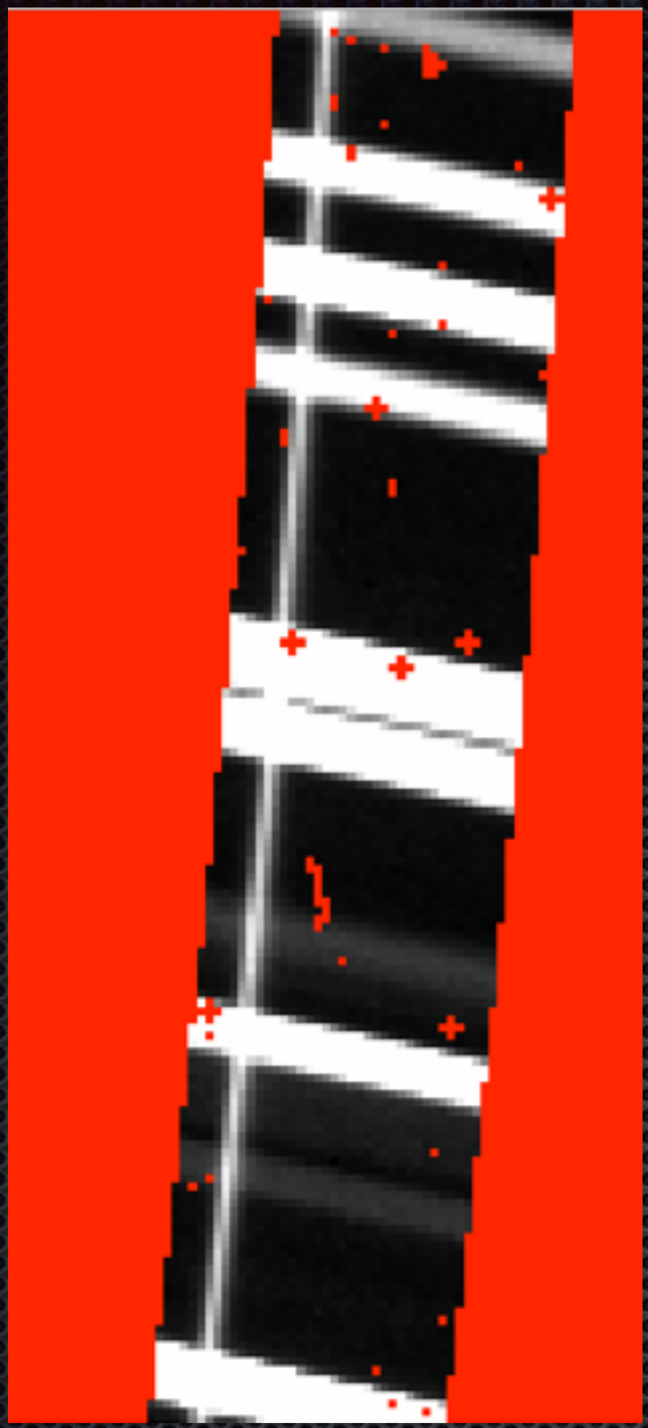




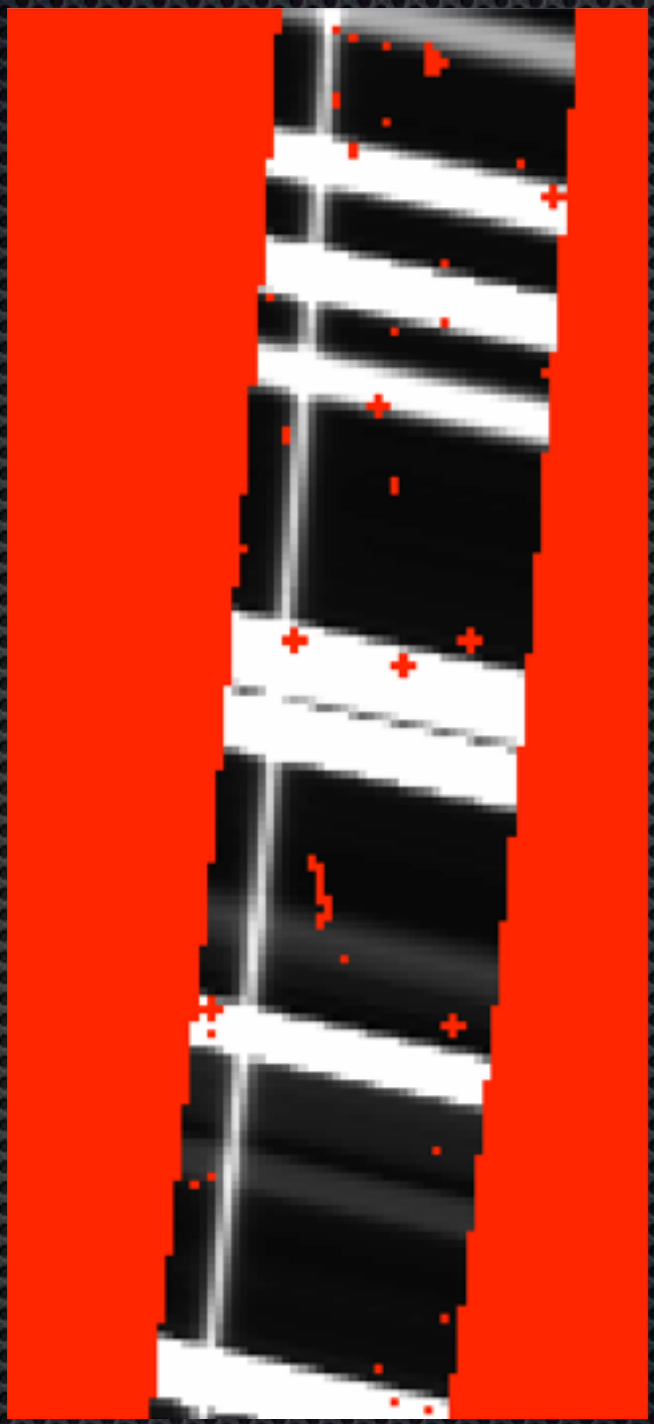
Model



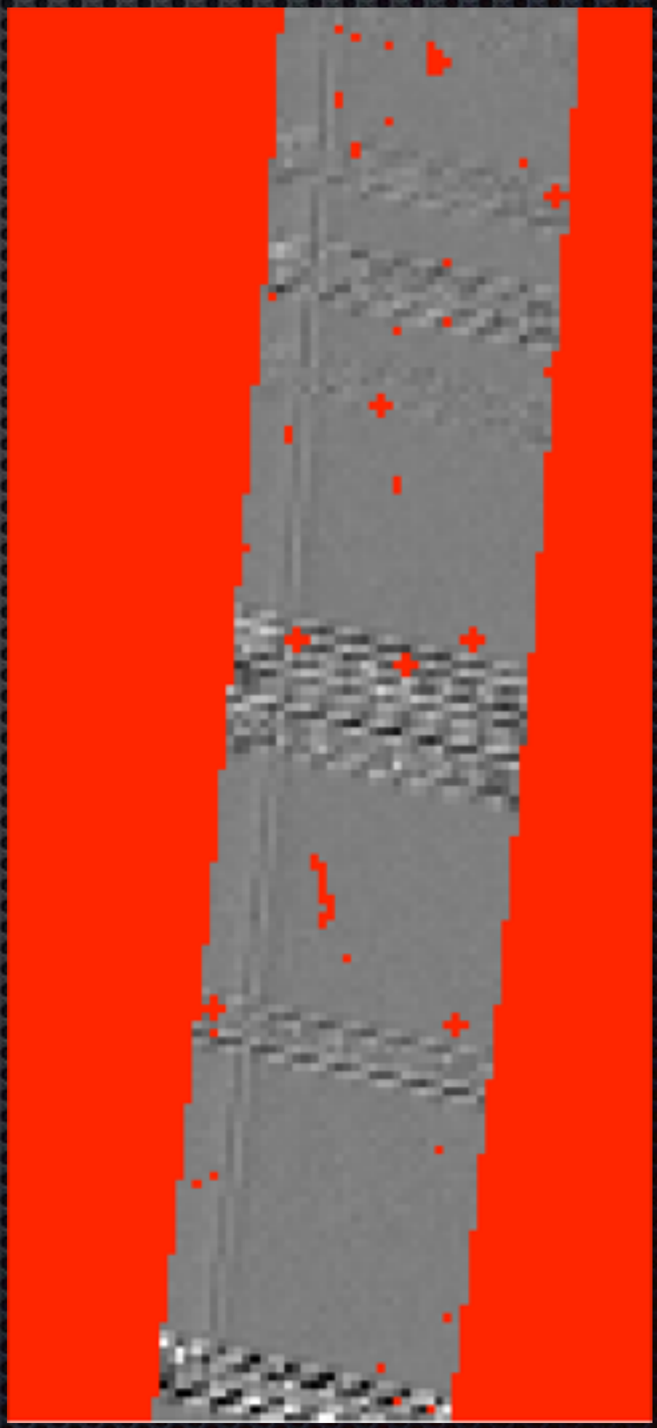
Residual



Raw Observations



Model



Residual

XTool - a prototype implementation

XTool - a prototype implementation

- uses pre-processed raw frames from the pipelines

XTool - a prototype implementation

- uses pre-processed raw frames from the pipelines
- extracts point-source spectra

XTool - a prototype implementation

- uses pre-processed raw frames from the pipelines
- extracts point-source spectra
- models the background for each spectral channel

XTool - a prototype implementation

- uses pre-processed raw frames from the pipelines
- extracts point-source spectra
- models the background for each spectral channel
- takes 1 s for each evaluation - ca. 15 minutes for one fit

XTool - a prototype implementation

- uses pre-processed raw frames from the pipelines
- extracts point-source spectra
- models the background for each spectral channel
- takes 1 s for each evaluation - ca. 15 minutes for one fit
- open collaboration @ github.com/eso/xtool

XTool - a prototype implementation

- uses pre-processed raw frames from the pipelines
- extracts point-source spectra
- models the background for each spectral channel
- takes 1 s for each evaluation - ca. 15 minutes for one fit
- open collaboration @ github.com/eso/xtool
    - see xtool.readthedocs.org for documentation

Conclusions

Conclusions

- Data simulators useful for fitting and data reduction as well

Conclusions

- Data simulators useful for fitting and data reduction as well
- Obtain science data, instrument health and robust removal of comics

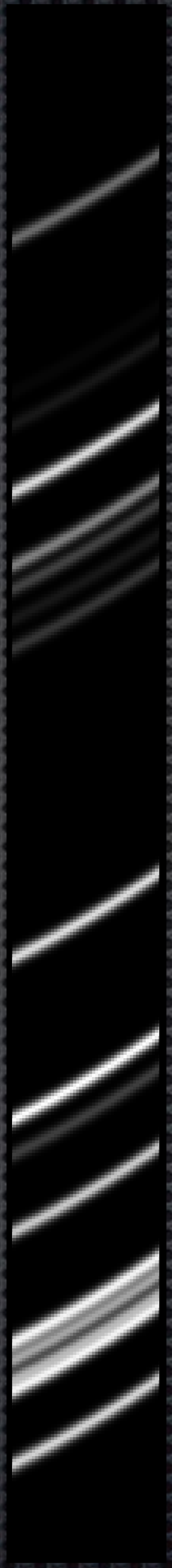
Conclusions

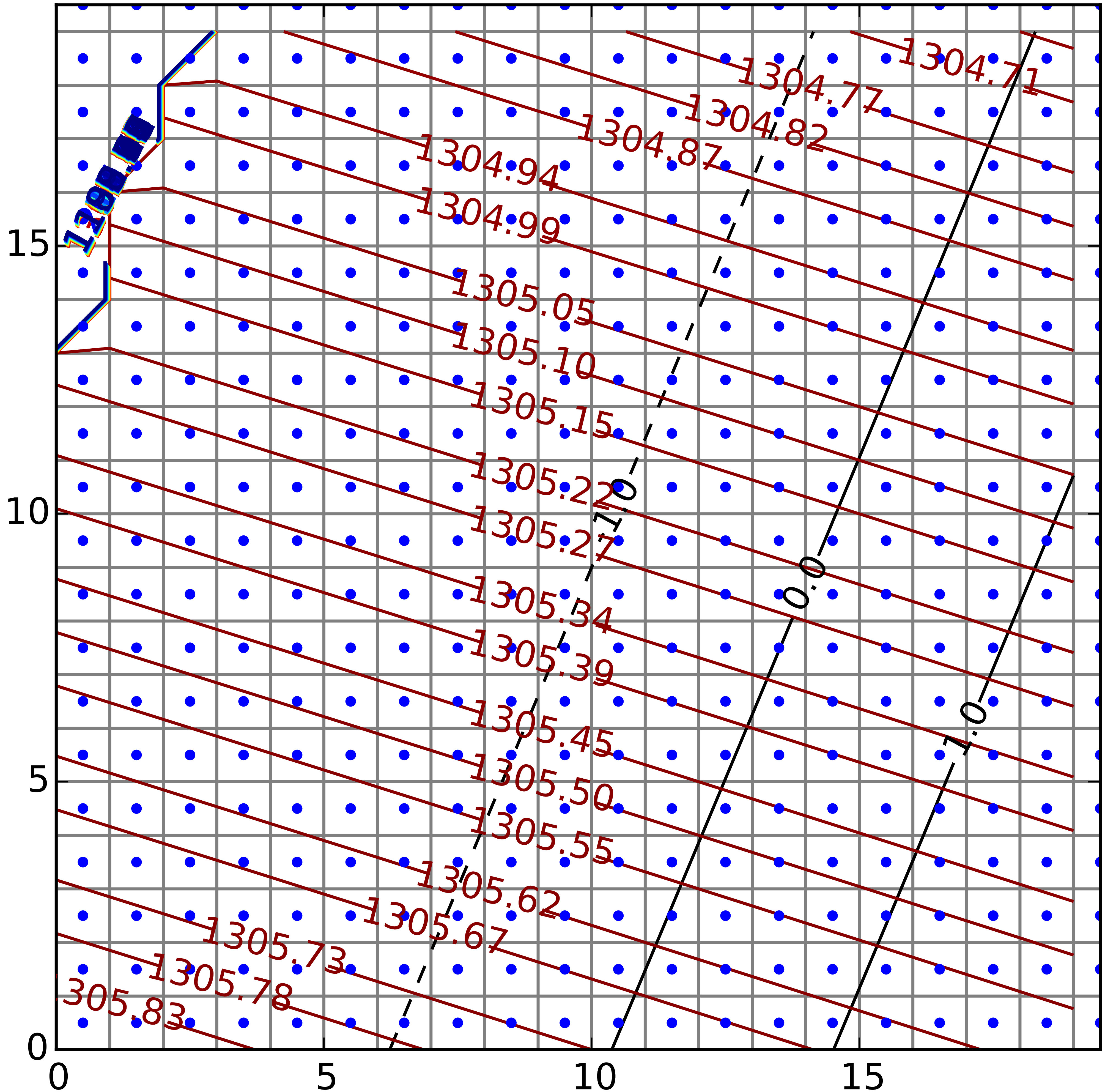
- Data simulators useful for fitting and data reduction as well
- Obtain science data, instrument health and robust removal of comics
- XTool - a prototype that will be expanded

Thank you!

Background

Background + Sky

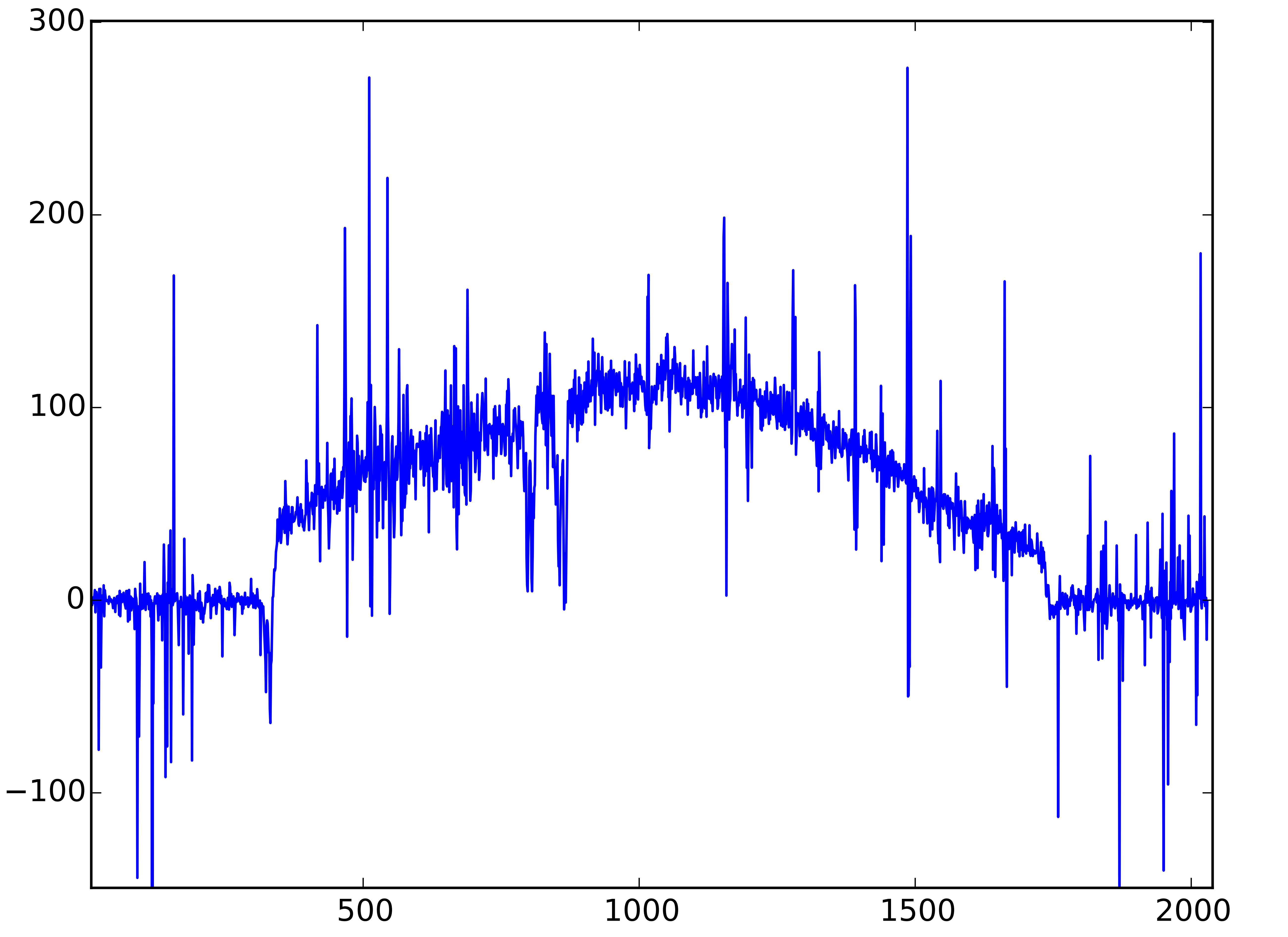




1304.71
1304.77
1304.82
1304.87
1304.94
1304.99
1305.05
1305.10
1305.15
1305.22
1305.27
1305.34
1305.39
1305.45
1305.50
1305.55
1305.62
1305.67
1305.73
1305.78
305.83
-1.0
0.0
1.0
0
5
10
15
0
5
10
15

I’m convinced

Now show me the money!


300
200
100
0
−100
500
1000
1500
2000

The short term future

- Do the uncertainties
- Make a script that does it -> Mogdiliani compare
- adjust the instrument model
- build a framework to allow users to use this
- coming to an open source repository near you in 2024

The long term future

- Build a fittable model for sky and instrument
- Fit calibration frames (i.e. Arcs) to get priors
- Build a bayesian posterior for instruments
- …. profit!

Questions?