Trend Filtering

A Modern Statistical Tool for Time-Domain Astronomy and Astronomical Spectroscopy

Collin A. Politsch

Carnegie Mellon University

Joint Statistical Meetings

Astrostatistics Student Paper Competition

August 5, 2020

Overview of paper submission

MNRAS **000**, 1–28 (2019) Preprint 16 December 2019 Compiled using MNRAS LaTeX style file v3.0

Trend Filtering: A Modern Statistical Tool for Time-Domain Astronomy and Astronomical Spectroscopy

Collin A. Politsch,[1,2,3★] Jessi Cisewski-Kehe,[4] Rupert A. C. Croft,[3,5,6] and Larry Wasserman[1,2,3]

[1] *Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA 15213*
[2] *Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213*
[3] *McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213*
[4] *Department of Statistics and Data Science, Yale University, New Haven, CT 06520*
[5] *Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213*
[6] *School of Physics, University of Melbourne, VIC 3010, Australia*

Accepted XXX. Received YYY; in original form 2019 August 20

ABSTRACT

The problem of denoising a one-dimensional signal possessing varying degrees of smoothness is ubiquitous in time-domain astronomy and astronomical spectroscopy. For example, in the time domain, an astronomical object may exhibit a smoothly varying intensity that is occasionally interrupted by abrupt dips or spikes. Likewise, in the spectroscopic setting, a noiseless spectrum typically contains intervals of relative smoothness mixed with localized higher frequency components such as emission peaks and absorption lines. In this work, we present trend filtering, a modern nonparametric statistical tool that yields significant improvements in this broad problem space of denoising *spatially heterogeneous* signals. When the underlying signal is spatially heterogeneous, trend filtering is superior to any statistical estimator that is a linear combination of the observed data—including kernels, LOESS, smoothing splines, Gaussian process regression, and many other popular methods. In the spirit of illustrating the broad utility of trend filtering, we discuss its relevance to a diverse set of spectroscopic and time-domain studies. The observations we discuss are (1) the Lyman-α forest of quasar spectra; (2) more general spectroscopy of quasars, galaxies, and stars; (3) stellar light curves with transiting exoplanet(s); (4) eclipsing binary light curves; and (5) supernova light curves. We study the Lyman-α forest in the greatest detail—using trend filtering to map the large-scale structure of the intergalactic medium along quasar-observer sightlines. The remaining studies broadly center around the themes of using trend filtering to estimate observable parameters and generate spectral/light-curve templates.

Key words: Methods: statistical, techniques: spectroscopic, cosmology: observations, stars: planetary systems, stars: binaries: eclipsing, supernovae: general

1 INTRODUCTION

Many astronomical observations produce one-dimensional data with unknown or varying degrees of smoothness. These include data from time-domain astronomy, where transient events such as supernovae can show light-curve variations on timescales ranging from seconds to years (e.g., Dimitriadis et al. 2017; Tolstov et al. 2019). Similarly, in astronomical spectroscopy, with wavelength (or frequency) as the independent variable, sharp absorption or emission-line features can be present alongside smoothly varying black-body or other continuum radiation (see, e.g., Tennyson 2019). In each of these general settings, we observe a signal plus noise and would like to denoise the signal as accurately as possible. Indeed the set of statistical tools available for addressing this general problem is quite vast. Commonly used nonparametric regression methods include kernel smoothers (e.g., Hall et al. 2002; Croft et al. 2002), local polynomial regression (LOESS; e.g., Maron & Howes 2003; Persson et al. 2004), splines (e.g., Peiris & Verde 2010;

★ E-mail: capolitsch@cmu.edu

Overview of paper submission

MNRAS **000**, 1–28 (2019) Preprint 16 December 2019 Compiled using MNRAS LaTeX style file v3.0

Trend Filtering: A Modern Statistical Tool for Time-Domain Astronomy and Astronomical Spectroscopy

Collin A. Politsch,[1,2,3]★ Jessi Cisewski-Kehe,[4] Rupert A. C. Croft,[3,5,6] and Larry Wasserman[1,2,3]

[1] *Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA 15213*
[2] *Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213*
[3] *McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213*
[4] *Department of Statistics and Data Science, Yale University, New Haven, CT 06520*
[5] *Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213*
[6] *School of Physics, University of Melbourne, VIC 3010, Australia*

Accepted XXX. Received YYY; in original form 2019 August 20

ABSTRACT

The problem of denoising a one-dimensional signal possessing varying degrees of smoothness is ubiquitous in time-domain astronomy and astronomical spectroscopy. For example, in the time domain, an astronomical object may exhibit a smoothly varying intensity that is occasionally interrupted by abrupt dips or spikes. Likewise, in the spectroscopic setting, a noiseless spectrum typically contains intervals of relative smoothness mixed with localized higher frequency components such as emission peaks and absorption lines. In this work, we present trend filtering, a modern nonparametric statistical tool that yields significant improvements in this broad problem space of denoising *spatially heterogeneous* signals. When the underlying signal is spatially heterogeneous, trend filtering is superior to any statistical estimator that is a linear combination of the observed data—including kernels, LOESS, smoothing splines, Gaussian process regression, and many other popular methods. In the spirit of illustrating the broad utility of trend filtering, we discuss its relevance to a diverse set of spectroscopic and time-domain studies. The observations we discuss are (1) the Lyman-α forest of quasar spectra; (2) more general spectroscopy of quasars, galaxies, and stars; (3) stellar light curves with transiting exoplanet(s); (4) eclipsing binary light curves; and (5) supernova light curves. We study the Lyman-α forest in the greatest detail—using trend filtering to map the large-scale structure of the intergalactic medium along quasar-observer sightlines. The remaining studies broadly center around the themes of using trend filtering to estimate observable parameters and generate spectral/light-curve templates.

Key words: Methods: statistical, techniques: spectroscopic, cosmology: observations, stars: planetary systems, stars: binaries: eclipsing, supernovae: general

1 INTRODUCTION

Many astronomical observations produce one-dimensional data with unknown or varying degrees of smoothness. These include data from time-domain astronomy, where transient events such as supernovae can show light-curve variations on timescales ranging from seconds to years (e.g., Dimitriadis et al. 2017; Tolstov et al. 2019). Similarly, in astronomical spectroscopy, with wavelength (or frequency) as the independent variable, sharp absorption or emission-line features can be present alongside smoothly varying black-body or other continuum radiation (see, e.g., Tennyson 2019). In each of these general settings, we observe a signal plus noise and would like to denoise the signal as accurately as possible. Indeed the set of statistical tools available for addressing this general problem is quite vast. Commonly used nonparametric regression methods include kernel smoothers (e.g., Hall et al. 2002; Croft et al. 2002), local polynomial regression (LOESS; e.g., Maron & Howes 2003; Persson et al. 2004), splines (e.g., Peiris & Verde 2010;

★ E-mail: capolitsch@cmu.edu

- Trend filtering (*Tibshirani 2014, Ann. Stat.*) is a modern statistical tool for 1D nonparametric regression

Overview of paper submission

MNRAS **000**, 1–28 (2019) Preprint 16 December 2019 Compiled using MNRAS LaTeX style file v3.0

Trend Filtering: A Modern Statistical Tool for Time-Domain Astronomy and Astronomical Spectroscopy

Collin A. Politsch,[1,2,3]★ Jessi Cisewski-Kehe,[4] Rupert A. C. Croft,[3,5,6] and Larry Wasserman[1,2,3]

[1] *Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA 15213*
[2] *Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213*
[3] *McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213*
[4] *Department of Statistics and Data Science, Yale University, New Haven, CT 06520*
[5] *Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213*
[6] *School of Physics, University of Melbourne, VIC 3010, Australia*

Accepted XXX. Received YYY; in original form 2019 August 20

ABSTRACT

The problem of denoising a one-dimensional signal possessing varying degrees of smoothness is ubiquitous in time-domain astronomy and astronomical spectroscopy. For example, in the time domain, an astronomical object may exhibit a smoothly varying intensity that is occasionally interrupted by abrupt dips or spikes. Likewise, in the spectroscopic setting, a noiseless spectrum typically contains intervals of relative smoothness mixed with localized higher frequency components such as emission peaks and absorption lines. In this work, we present trend filtering, a modern nonparametric statistical tool that yields significant improvements in this broad problem space of denoising *spatially heterogeneous* signals. When the underlying signal is spatially heterogeneous, trend filtering is superior to any statistical estimator that is a linear combination of the observed data—including kernels, LOESS, smoothing splines, Gaussian process regression, and many other popular methods. In the spirit of illustrating the broad utility of trend filtering, we discuss its relevance to a diverse set of spectroscopic and time-domain studies. The observations we discuss are (1) the Lyman-α forest of quasar spectra; (2) more general spectroscopy of quasars, galaxies, and stars; (3) stellar light curves with transiting exoplanet(s); (4) eclipsing binary light curves; and (5) supernova light curves. We study the Lyman-α forest in the greatest detail—using trend filtering to map the large-scale structure of the intergalactic medium along quasar-observer sightlines. The remaining studies broadly center around the themes of using trend filtering to estimate observable parameters and generate spectral/light-curve templates.

Key words: Methods: statistical, techniques: spectroscopic, cosmology: observations, stars: planetary systems, stars: binaries: eclipsing, supernovae: general

1 INTRODUCTION

Many astronomical observations produce one-dimensional data with unknown or varying degrees of smoothness. These include data from time-domain astronomy, where transient events such as supernovae can show light-curve variations on timescales ranging from seconds to years (e.g., Dimitriadis et al. 2017; Tolstov et al. 2019). Similarly, in astronomical spectroscopy, with wavelength (or frequency) as the independent variable, sharp absorption or emission-line features can be present alongside smoothly varying black-body or other continuum radiation (see, e.g., Tennyson 2019). In each of these general settings, we observe a signal plus noise and would like to denoise the signal as accurately as possible. Indeed the set of statistical tools available for addressing this general problem is quite vast. Commonly used nonparametric regression methods include kernel smoothers (e.g., Hall et al. 2002; Croft et al. 2002), local polynomial regression (LOESS; e.g., Maron & Howes 2003; Persson et al. 2004), splines (e.g., Peiris & Verde 2010;

★ E-mail: capolitsch@cmu.edu

- Trend filtering (*Tibshirani 2014, Ann. Stat.*) is a modern statistical tool for 1D nonparametric regression
- When the signal exhibits varying degrees of smoothness, trend filtering is superior to classical methods

Overview of paper submission

MNRAS 000, 1–28 (2019) Preprint 16 December 2019 Compiled using MNRAS LaTeX style file v3.0

Trend Filtering: A Modern Statistical Tool for Time-Domain Astronomy and Astronomical Spectroscopy

Collin A. Politsch,[1,2,3]★ Jessi Cisewski-Kehe,[4] Rupert A. C. Croft,[3,5,6] and Larry Wasserman[1,2,3]

[1] *Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA 15213*
[2] *Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213*
[3] *McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213*
[4] *Department of Statistics and Data Science, Yale University, New Haven, CT 06520*
[5] *Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213*
[6] *School of Physics, University of Melbourne, VIC 3010, Australia*

Accepted XXX. Received YYY; in original form 2019 August 20

ABSTRACT

The problem of denoising a one-dimensional signal possessing varying degrees of smoothness is ubiquitous in time-domain astronomy and astronomical spectroscopy. For example, in the time domain, an astronomical object may exhibit a smoothly varying intensity that is occasionally interrupted by abrupt dips or spikes. Likewise, in the spectroscopic setting, a noiseless spectrum typically contains intervals of relative smoothness mixed with localized higher frequency components such as emission peaks and absorption lines. In this work, we present trend filtering, a modern nonparametric statistical tool that yields significant improvements in this broad problem space of denoising *spatially heterogeneous* signals. When the underlying signal is spatially heterogeneous, trend filtering is superior to any statistical estimator that is a linear combination of the observed data—including kernels, LOESS, smoothing splines, Gaussian process regression, and many other popular methods. In the spirit of illustrating the broad utility of trend filtering, we discuss its relevance to a diverse set of spectroscopic and time-domain studies. The observations we discuss are (1) the Lyman-α forest of quasar spectra; (2) more general spectroscopy of quasars, galaxies, and stars; (3) stellar light curves with transiting exoplanet(s); (4) eclipsing binary light curves; and (5) supernova light curves. We study the Lyman-α forest in the greatest detail—using trend filtering to map the large-scale structure of the intergalactic medium along quasar-observer sightlines. The remaining studies broadly center around the themes of using trend filtering to estimate observable parameters and generate spectral/light-curve templates.

Key words: Methods: statistical, techniques: spectroscopic, cosmology: observations, stars: planetary systems, stars: binaries: eclipsing, supernovae: general

1 INTRODUCTION

Many astronomical observations produce one-dimensional data with unknown or varying degrees of smoothness. These include data from time-domain astronomy, where transient events such as supernovae can show light-curve variations on timescales ranging from seconds to years (e.g., Dimitriadis et al. 2017; Tolstov et al. 2019). Similarly, in astronomical spectroscopy, with wavelength (or frequency) as the independent variable, sharp absorption or emission-line features can be present alongside smoothly varying black-body or other continuum radiation (see, e.g., Tennyson 2019). In each of these general settings, we observe a signal plus noise and would like to denoise the signal as accurately as possible. Indeed the set of statistical tools available for addressing this general problem is quite vast. Commonly used nonparametric regression methods include kernel smoothers (e.g., Hall et al. 2002; Croft et al. 2002), local polynomial regression (LOESS; e.g., Maron & Howes 2003; Persson et al. 2004), splines (e.g., Peiris & Verde 2010;

★ E-mail: capolitsch@cmu.edu

- Trend filtering (*Tibshirani 2014, Ann. Stat.*) is a modern statistical tool for 1D nonparametric regression
- When the signal exhibits varying degrees of smoothness, trend filtering is superior to classical methods
- Trend filtering can have a significant impact on a wide variety of problems in astronomical data analysis

Overview of paper submission

MNRAS 000, 1–28 (2019) Preprint 16 December 2019 Compiled using MNRAS LaTeX style file v3.0

Trend Filtering: A Modern Statistical Tool for Time-Domain Astronomy and Astronomical Spectroscopy

Collin A. Politsch,[1,2,3]★ Jessi Cisewski-Kehe,[4] Rupert A. C. Croft,[3,5,6] and Larry Wasserman[1,2,3]

[1] *Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA 15213*
[2] *Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213*
[3] *McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213*
[4] *Department of Statistics and Data Science, Yale University, New Haven, CT 06520*
[5] *Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213*
[6] *School of Physics, University of Melbourne, VIC 3010, Australia*

Accepted XXX. Received YYY; in original form 2019 August 20

ABSTRACT

The problem of denoising a one-dimensional signal possessing varying degrees of smoothness is ubiquitous in time-domain astronomy and astronomical spectroscopy. For example, in the time domain, an astronomical object may exhibit a smoothly varying intensity that is occasionally interrupted by abrupt dips or spikes. Likewise, in the spectroscopic setting, a noiseless spectrum typically contains intervals of relative smoothness mixed with localized higher frequency components such as emission peaks and absorption lines. In this work, we present trend filtering, a modern nonparametric statistical tool that yields significant improvements in this broad problem space of denoising *spatially heterogeneous* signals. When the underlying signal is spatially heterogeneous, trend filtering is superior to any statistical estimator that is a linear combination of the observed data—including kernels, LOESS, smoothing splines, Gaussian process regression, and many other popular methods. In the spirit of illustrating the broad utility of trend filtering, we discuss its relevance to a diverse set of spectroscopic and time-domain studies. The observations we discuss are (1) the Lyman-α forest of quasar spectra; (2) more general spectroscopy of quasars, galaxies, and stars; (3) stellar light curves with transiting exoplanet(s); (4) eclipsing binary light curves; and (5) supernova light curves. We study the Lyman-α forest in the greatest detail—using trend filtering to map the large-scale structure of the intergalactic medium along quasar-observer sightlines. The remaining studies broadly center around the themes of using trend filtering to estimate observable parameters and generate spectral/light-curve templates.

Key words: Methods: statistical, techniques: spectroscopic, cosmology: observations, stars: planetary systems, stars: binaries: eclipsing, supernovae: general

1 INTRODUCTION

Many astronomical observations produce one-dimensional data with unknown or varying degrees of smoothness. These include data from time-domain astronomy, where transient events such as supernovae can show light-curve variations on timescales ranging from seconds to years (e.g., Dimitriadis et al. 2017; Tolstov et al. 2019). Similarly, in astronomical spectroscopy, with wavelength (or frequency) as the independent variable, sharp absorption or emission-line features can be present alongside smoothly varying black-body or other continuum radiation (see, e.g., Tennyson 2019). In each of these general settings, we observe a signal plus noise and would like to denoise the signal as accurately as possible. Indeed the set of statistical tools available for addressing this general problem is quite vast. Commonly used nonparametric regression methods include kernel smoothers (e.g., Hall et al. 2002; Croft et al. 2002), local polynomial regression (LOESS; e.g., Maron & Howes 2003; Persson et al. 2004), splines (e.g., Peiris & Verde 2010;

★ E-mail: capolitsch@cmu.edu

- Trend filtering (*Tibshirani 2014, Ann. Stat.*) is a modern statistical tool for 1D nonparametric regression
- When the signal exhibits varying degrees of smoothness, trend filtering is superior to classical methods
- Trend filtering can have a significant impact on a wide variety of problems in astronomical data analysis

Statistical tools used for 1D nonparametric regression in Astronomy & Astrophysics

Trend Filtering

Published as a two-part series in *MNRAS*

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **492,** 4005–4018 (2020) doi:10.1093/mnras/staa106
Advance Access publication 2020 January 14

Trend filtering – I. A modern statistical tool for time-domain astronomy and astronomical spectroscopy

Collin A. Politsch [1,2,3]⋆ Jessi Cisewski-Kehe,[4] Rupert A. C. Croft[3,5,6]⋆ and Larry Wasserman[1,2,3]

[1]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2]Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[3]McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[4]Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA
[5]Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[6]School of Physics, University of Melbourne, Parkville, VIC 3010, Australia

Accepted 2020 January 9. Received 2020 January 7; in original form 2019 August 21

ABSTRACT

The problem of denoising a 1D signal possessing varying degrees of smoothness is ubiquitous in time-domain astronomy and astronomical spectroscopy. For example, in the time domain, an astronomical object may exhibit a smoothly varying intensity that is occasionally interrupted by abrupt dips or spikes. Likewise, in the spectroscopic setting, a noiseless spectrum typically contains intervals of relative smoothness mixed with localized higher frequency components such as emission peaks and absorption lines. In this work, we present trend filtering, a modern non-parametric statistical tool that yields significant improvements in this broad problem space of denoising *spatially heterogeneous* signals. When the underlying signal is spatially heterogeneous, trend filtering is superior to any statistical estimator that is a linear combination of the observed data – including kernel smoothers, LOESS, smoothing splines, Gaussian process regression, and many other popular methods. Furthermore, the trend filtering estimate can be computed with practical and scalable efficiency via a specialized convex optimization algorithm, e.g. handling sample sizes of $n \gtrsim 10^7$ within a few minutes. In a companion paper, we explicitly demonstrate the broad utility of trend filtering to observational astronomy by carrying out a diverse set of spectroscopic and time-domain analyses.

Key words: methods: statistical – techniques: photometric – techniques: spectroscopic.

1 INTRODUCTION

Many astronomical observations produce 1D data with varying (or unknown) degrees of smoothness. These include data from time-domain astronomy, where transient events such as supernovae can show light-curve variations on time-scales ranging from seconds to years (e.g. Dimitriadis et al. 2017; Tolstov et al. 2019). Similarly, in astronomical spectroscopy, with wavelength (or frequency) as the input variable, sharp absorption or emission-line features can be present alongside smoothly varying blackbody or other continuum radiation (see e.g. Tennyson 2019). In each of these general settings, we observe a signal plus noise and would like to denoise the signal as accurately as possible. Indeed the set of statistical tools available for addressing this general problem is quite vast. Commonly used non-parametric regression methods include kernel smoothers (e.g. Hall et al. 2002; Croft et al. 2002), local polynomial regression (LOESS; e.g. Maron & Howes 2003; Persson et al. 2004), splines (e.g. Peiris & Verde 2010; Contreras et al. 2010; Dhawan et al. 2015), Gaussian process regression (e.g. Gibson et al. 2012; Aigrain, Parviainen & Pope 2016; Gómez-Valent & Amendola 2018), and wavelet decompositions (e.g. Fligge & Solanki 1997; Theuns & Zaroubi 2000; Golkhou & Butler 2014). A rich and elegant statistical literature exists on the theoretical and practical achievements of these methods (see e.g. Györfi et al. 2002; Wasserman 2006; Hastie, Tibshirani & Friedman 2009 for general references). However, when the underlying signal is *spatially heterogeneous*, i.e. exhibits varying degrees of smoothness, the power of classical statistical literature is quite limited. Kernels, LOESS, smoothing splines, and Gaussian process regression belong to a broad family of non-parametric methods called *linear smoothers*, which has been shown to be uniformly suboptimal for estimating spatially heterogeneous signals (Nemirovskii, Polyak & Tsybakov 1985; Nemirovskii 1985; Donoho & Johnstone 1998). The common limitation of these methods is that they are not locally adaptive; i.e. by construction, they do not adapt to local degrees of smoothness in a signal. In particular,

⋆ E-mail: capolitsch@cmu.edu (CAP); rcroft@cmu.edu (RAC)

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **492,** 4019–4032 (2020) doi:10.1093/mnras/staa110
Advance Access publication 2020 January 14

Trend filtering – II. Denoising astronomical signals with varying degrees of smoothness

Collin A. Politsch [1,2,3]⋆ Jessi Cisewski-Kehe,[4] Rupert A. C. Croft[3,5,6]⋆ and Larry Wasserman[1,2,3]

[1]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2]Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[3]McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[4]Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA
[5]Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[6]School of Physics, University of Melbourne, VIC 3010, Australia

Accepted 2020 January 9. in original form 2020 January 7

ABSTRACT

Trend filtering – first introduced into the astronomical literature in Paper I of this series – is a state-of-the-art statistical tool for denoising 1D signals that possess varying degrees of smoothness. In this work, we demonstrate the broad utility of trend filtering to observational astronomy by discussing how it can contribute to a variety of spectroscopic and time-domain studies. The observations we discuss are (1) the Lyman-α (Lyα) forest of quasar spectra; (2) more general spectroscopy of quasars, galaxies, and stars; (3) stellar light curves with planetary transits; (4) eclipsing binary light curves; and (5) supernova light curves. We study the Lyα forest in the greatest detail – using trend filtering to map the large-scale structure of the intergalactic medium along quasar-observer lines of sight. The remaining studies share broad themes of: (1) estimating observable parameters of light curves and spectra; and (2) constructing observational spectral/light-curve templates. We also briefly discuss the utility of trend filtering as a tool for 1D data reduction and compression.

Key words: methods: statistical – binaries: eclipsing – planetary systems – supernovae: general – cosmology: observations.

1 INTRODUCTION

Many astronomical analyses can be described by the following problem setup. Suppose we collect noisy measurements of an observable quantity (e.g. flux, magnitude, or photon counts) according to the data generating process

$$f(t_i) = f_0(t_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where $t_1, \ldots, t_n$ is an arbitrarily spaced grid of 1D inputs (e.g. times or wavelengths), ϵ_i is mean zero noise, and f_0 is a signal that may exhibit varying degrees of smoothness across the input domain (e.g. a smooth signal with abrupt dips/spikes). Given the observed data sample, we then attempt to estimate (or denoise) the underlying signal f_0 from the observations by applying appropriate statistical methods. In Politsch et al. (2020) – hereafter referred to as Paper I – we introduced trend filtering (Tibshirani & Taylor 2011; Tibshirani 2014) into the astronomical literature. When the underlying signal is spatially heterogeneous, i.e. possesses varying degrees of smoothness, trend filtering is superior to popular statistical methods such as Gaussian process regression, smoothing splines, kernels, LOESS, and many others (Nemirovskii, Polyak & Tsybakov 1985; Nemirovskii 1985; Tibshirani 2014). Furthermore, the trend filtering estimate can be computed via a highly efficient and scalable convex optimization algorithm (Ramdas & Tibshirani 2016) and only requires data-driven selection of a single scalar hyperparameter. In this paper, we directly demonstrate the broad utility of trend filtering to observational astronomy by using it to carry out a diverse set of spectroscopic and time-domain analyses.

The outline of this paper is as follows. In Section 2, we use trend filtering to study the Lyman-α (Lyα) forest of quasar spectra – a series of absorption features that can be used as a tracer of the matter density distribution along quasar-observer lines of sight. We choose to study this application in depth and then illustrate the breadth of trend filtering's utility through our discussions in Section 3. The applications we discuss in Section 3 can be grouped into two broad (and often intertwined) categories: (1) deriving estimates of observable parameters from trend filtered observations; and (2) using trend filtering to construct spectral/light-curve templates of astronomical objects/events. In Section 3.1, we discuss constructing

⋆ E-mail: capolitsch@cmu.edu (CAP); rcroft@cmu.edu (RAC)

Outline

1. *Trend Filtering – I. A Modern Statistical Tool for Time-Domain Astronomy and Astronomical Spectroscopy, MNRAS*
    - Introduced trend filtering (Tibshirani 2014, *Annals of Statistics*) into the astronomical literature
    - Punchline: Trend filtering is superior to classical methods for analyzing noisy signals with varying degrees of smoothness

Outline

1. *Trend Filtering – I. A Modern Statistical Tool for Time-Domain Astronomy and Astronomical Spectroscopy, MNRAS*
    - Introduced trend filtering (Tibshirani 2014, *Annals of Statistics*) into the astronomical literature
    - Punchline: Trend filtering is superior to classical methods for analyzing noisy signals with varying degrees of smoothness

2. *Trend Filtering – II. Denoising Astronomical Signals with Varying Degrees of Smoothness, MNRAS*
    - **Extrasolar planets**
        - Characterization via the transit method
    - **Eclipsing binary stars**
        - Nonparametric modeling of phase-folded light curves
    - **Supernovae**
        - Light-curve template generation
        - Nonparametric estimation of observable parameters
    - **Stars/galaxies/quasars**
        - Spectral template generation
        - Nonparametric estimation of emission line parameters
    - **Intergalactic medium (via the Lyman-α forest)**
        - One-dimensional reconstruction of the intergalactic medium
    - **Data reduction and compression**
        - Efficient compression of 1D data sets

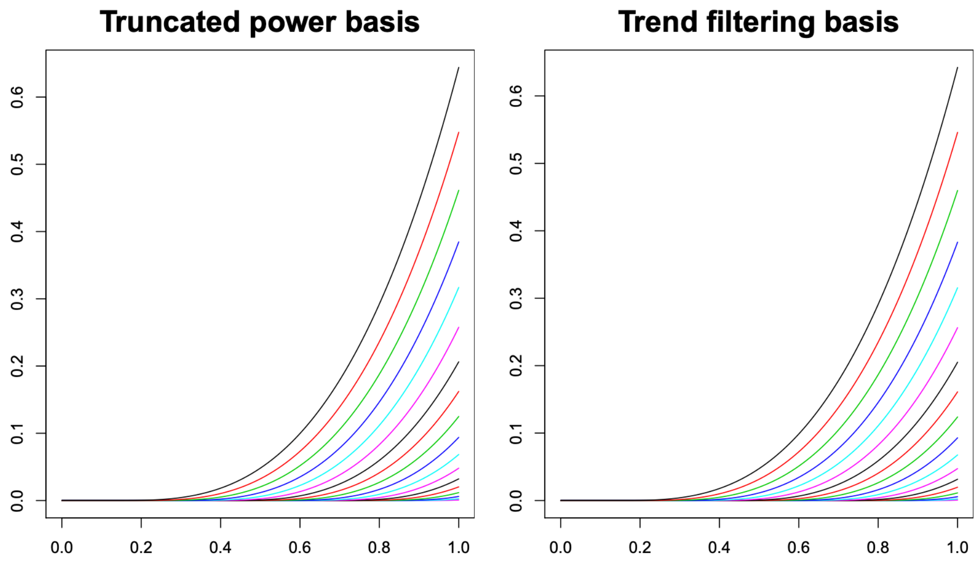
Overview of trend filtering

$$f(t_i) = f_0(t_i) + \epsilon_i \qquad\qquad i = 1, \ldots, n$$

Overview of trend filtering

$$f(t_i) = f_0(t_i) + \epsilon_i \qquad\qquad i = 1, \ldots, n$$

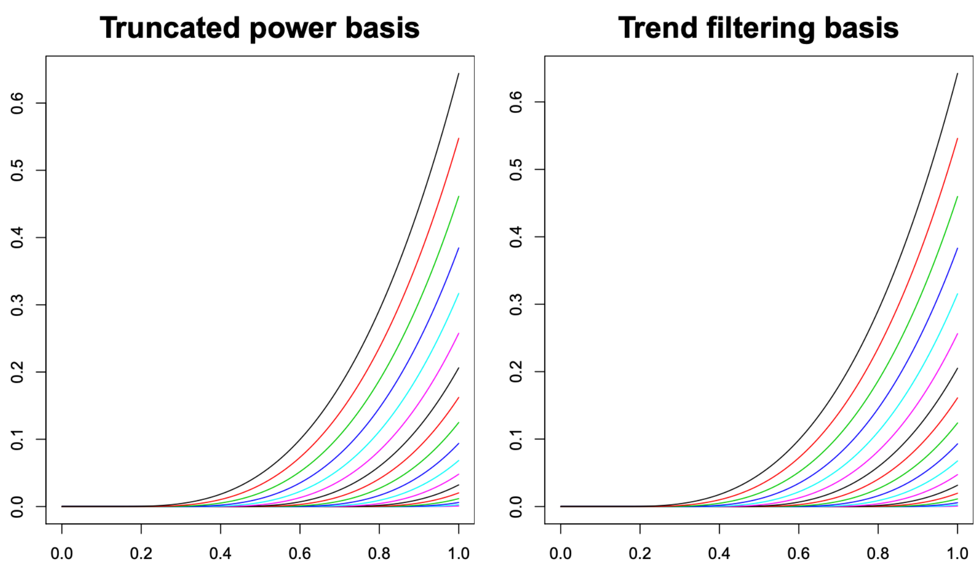
$$f_0(t) = \sum_j \beta_j h_j(t)$$



Overview of trend filtering

$$f(t_i) = f_0(t_i) + \epsilon_i \qquad\qquad i = 1, \ldots, n$$

$$f_0(t) = \sum_j \beta_j h_j(t)$$

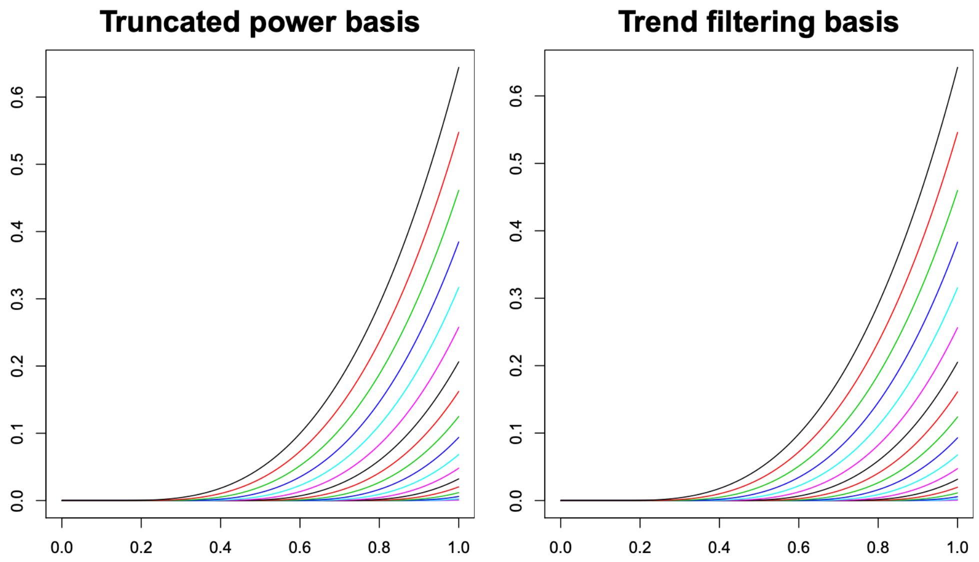


$$\min_{\{\beta_j\}} \quad \sum_{i=1}^{n} \left(f(t_i) - \sum_{j=1}^{n} \beta_j h_j(t_i) \right)^2 + \gamma \cdot k! \cdot \Delta t^k \sum_{j=k+2}^{n} |\beta_j|$$

Overview of trend filtering

$$f(t_i) = f_0(t_i) + \epsilon_i \qquad\qquad i = 1, \ldots, n$$

$$f_0(t) = \sum_j \beta_j h_j(t)$$



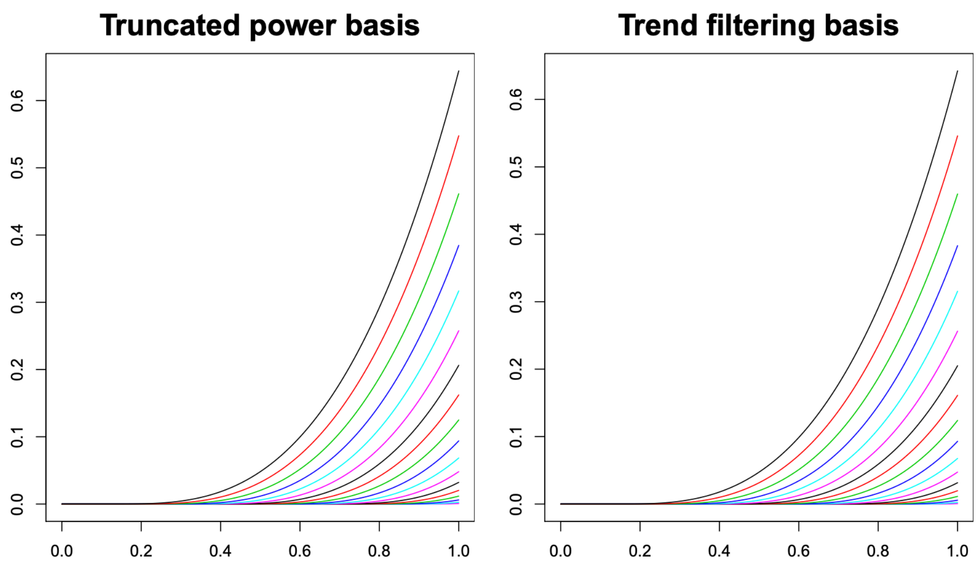
$$\min_{\{\beta_j\}} \quad \sum_{i=1}^{n} \left(f(t_i) - \sum_{j=1}^{n} \beta_j h_j(t_i) \right)^2 + \gamma \cdot k! \cdot \Delta t^k \sum_{j=k+2}^{n} |\beta_j|$$

$$\widehat{f}_0(t;\gamma) = \sum_{j=1}^{n} \widehat{\beta}_j h_j(t)$$

Overview of trend filtering

$$f(t_i) = f_0(t_i) + \epsilon_i \qquad i = 1, \ldots, n$$

$$f_0(t) = \sum_j \beta_j h_j(t)$$



$$\min_{\{\beta_j\}} \sum_{i=1}^{n} \left(f(t_i) - \sum_{j=1}^{n} \beta_j h_j(t_i) \right)^2 + \gamma \cdot k! \cdot \Delta t^k \sum_{j=k+2}^{n} |\beta_j|$$

$$\widehat{f}_0(t;\gamma) = \sum_{j=1}^{n} \widehat{\beta}_j h_j(t)$$

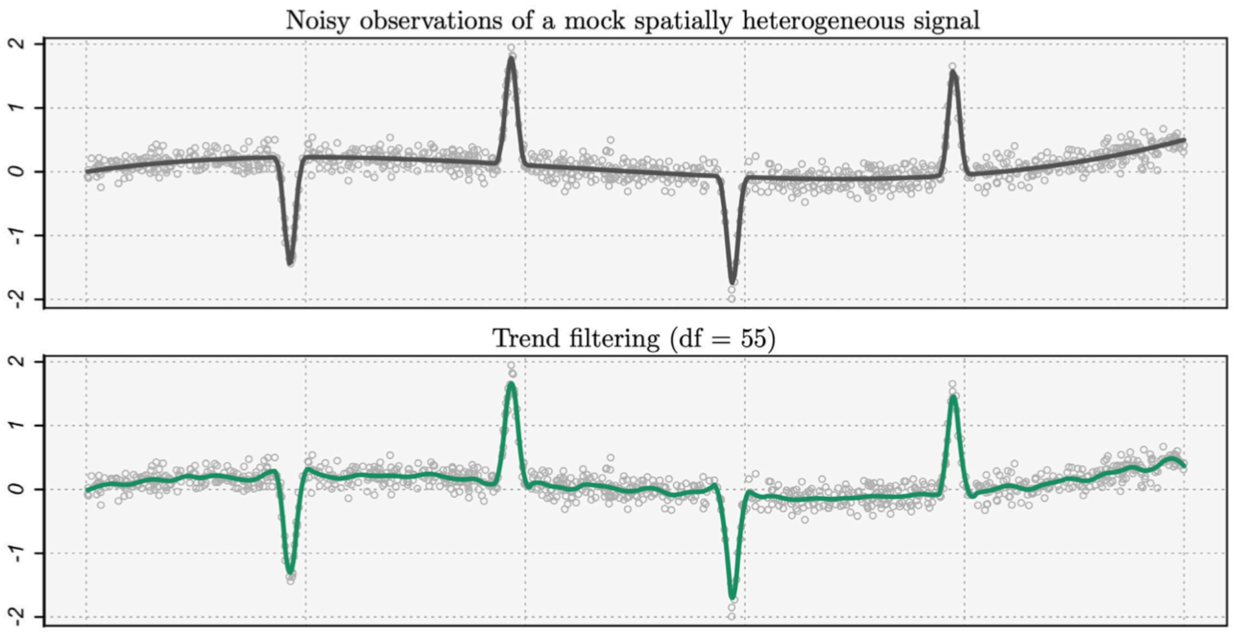
ℓ_p-penalized linear models

	High-dimensional	1D splines*
$p = 0$	**Subset selection** $\min_{\{\beta_j\}} \sum_{i=1}^{n} \Big(y_i - \sum_j \beta_j x_{ji} \Big)^2$ s.t. $\sum_{j=1}^{d} \mathbb{1}\{\beta_j \neq 0\} = s$	**Variable-knot regression spline** $\min_{\{\beta_j\}} \sum_{i=1}^{n} \Big(f(t_i) - \sum_j \beta_j \eta_j(t_i) \Big)^2$ s.t. $\sum_{j \geq k+2} \mathbb{1}\{\beta_j \neq 0\} = s$
$p = 1$	**Lasso** $\min_{\{\beta_j\}} \sum_{i=1}^{n} \Big(y_i - \sum_j \beta_j x_{ji} \Big)^2$ s.t. $\sum_{j=1}^{d} \lvert\beta_j\rvert \leq \tau$	
$p = 2$	**Ridge regression** $\min_{\{\beta_j\}} \sum_{i=1}^{n} \Big(y_i - \sum_j \beta_j x_{ji} \Big)^2$ s.t. $\sum_{j=1}^{d} \beta_j^2 \leq \tau$	**Smoothing spline** $\min_{\{\beta_j\}} \sum_{i=1}^{n} \Big(f(t_i) - \sum_j \beta_j \eta_j(t_i) \Big)^2$ s.t. $\sum_{j,k=1}^{n} \beta_j \beta_k \omega_{jk} \leq \tau$

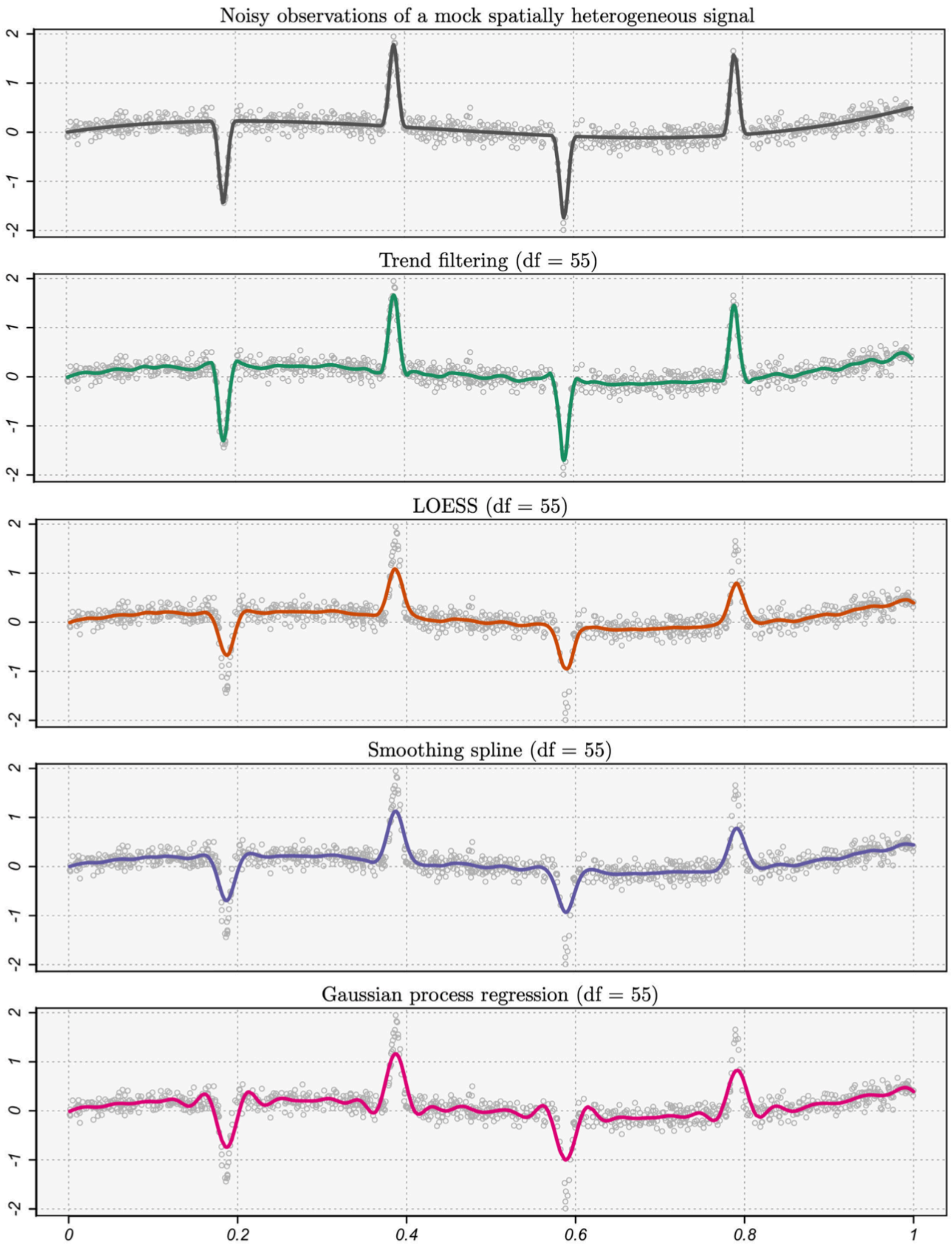
ℓ_p-penalized linear models

	High-dimensional	1D splines*
$p = 0$	**Subset selection** $$\min_{\{\beta_j\}} \sum_{i=1}^n \Big(y_i - \sum_j \beta_j x_{ji}\Big)^2 \quad \text{s.t.} \quad \sum_{j=1}^d \mathbb{1}\{\beta_j \neq 0\} = s$$	**Variable-knot regression spline** $$\min_{\{\beta_j\}} \sum_{i=1}^n \Big(f(t_i) - \sum_j \beta_j \eta_j(t_i)\Big)^2 \quad \text{s.t.} \quad \sum_{j \geq k+2} \mathbb{1}\{\beta_j \neq 0\} = s$$
$p = 1$	**Lasso** $$\min_{\{\beta_j\}} \sum_{i=1}^n \Big(y_i - \sum_j \beta_j x_{ji}\Big)^2 \quad \text{s.t.} \quad \sum_{j=1}^d \lvert\beta_j\rvert \leq \tau$$	**Trend filtering** $$\min_{\{\beta_j\}} \sum_{i=1}^n \Big(f(t_i) - \sum_j \beta_j h_j(t_i)\Big)^2 \quad \text{s.t.} \quad \sum_{j \geq k+2} \lvert\beta_j\rvert \leq \tau$$
$p = 2$	**Ridge regression** $$\min_{\{\beta_j\}} \sum_{i=1}^n \Big(y_i - \sum_j \beta_j x_{ji}\Big)^2 \quad \text{s.t.} \quad \sum_{j=1}^d \beta_j^2 \leq \tau$$	**Smoothing spline** $$\min_{\{\beta_j\}} \sum_{i=1}^n \Big(f(t_i) - \sum_j \beta_j \eta_j(t_i)\Big)^2 \quad \text{s.t.} \quad \sum_{j,k=1}^n \beta_j \beta_k \omega_{jk} \leq \tau$$

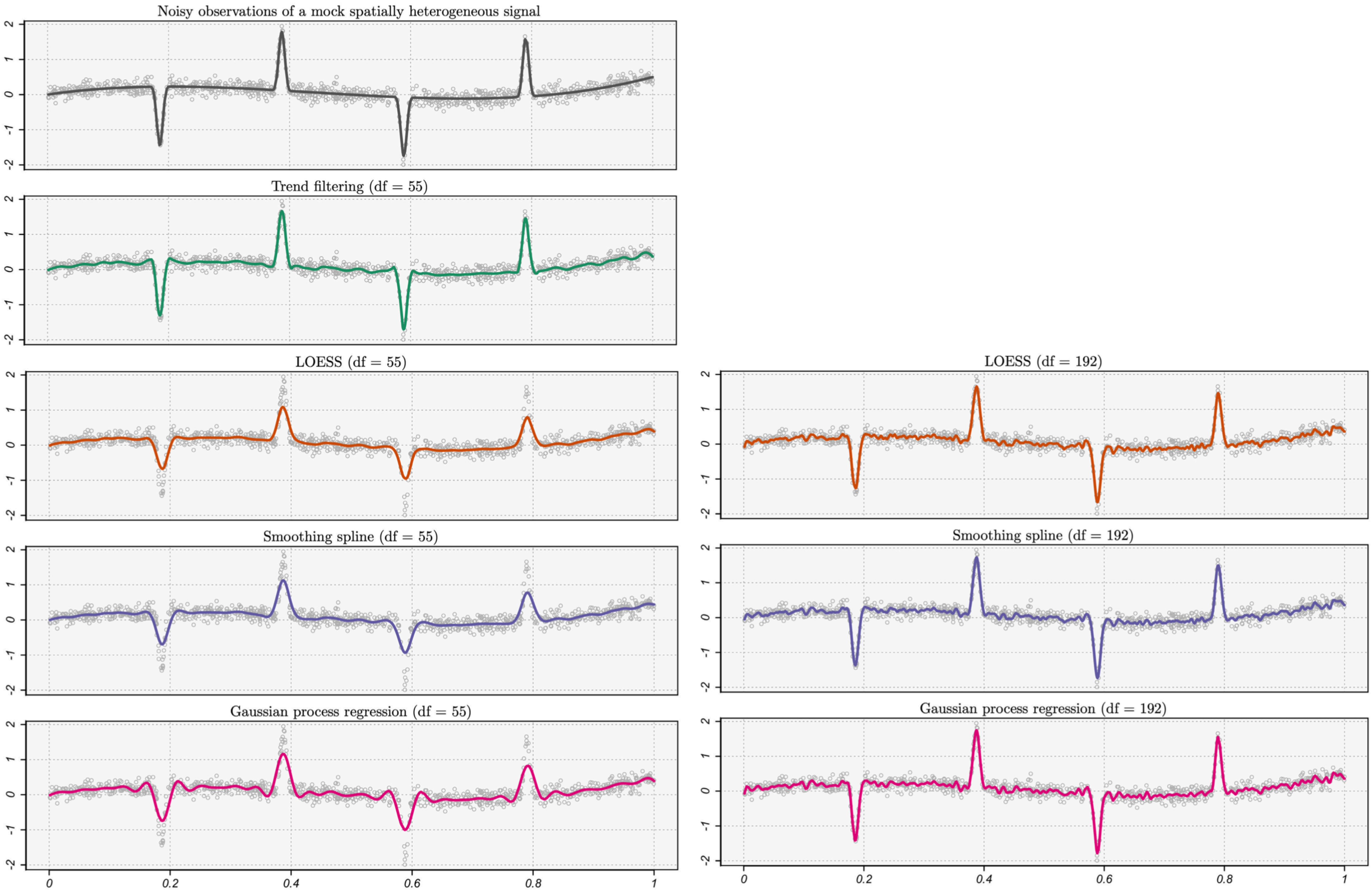
Sample comparison to other methods



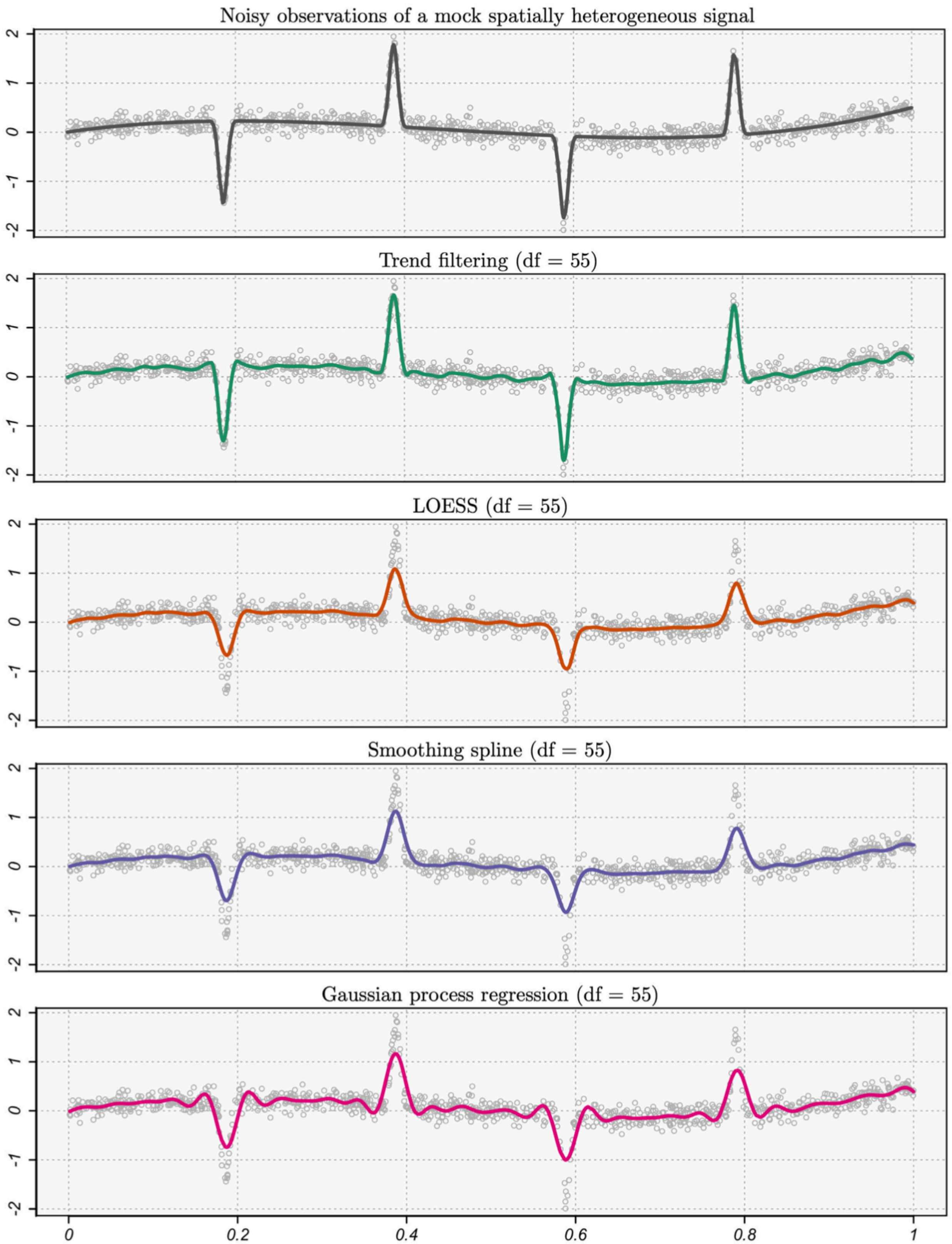
Sample comparison to other methods

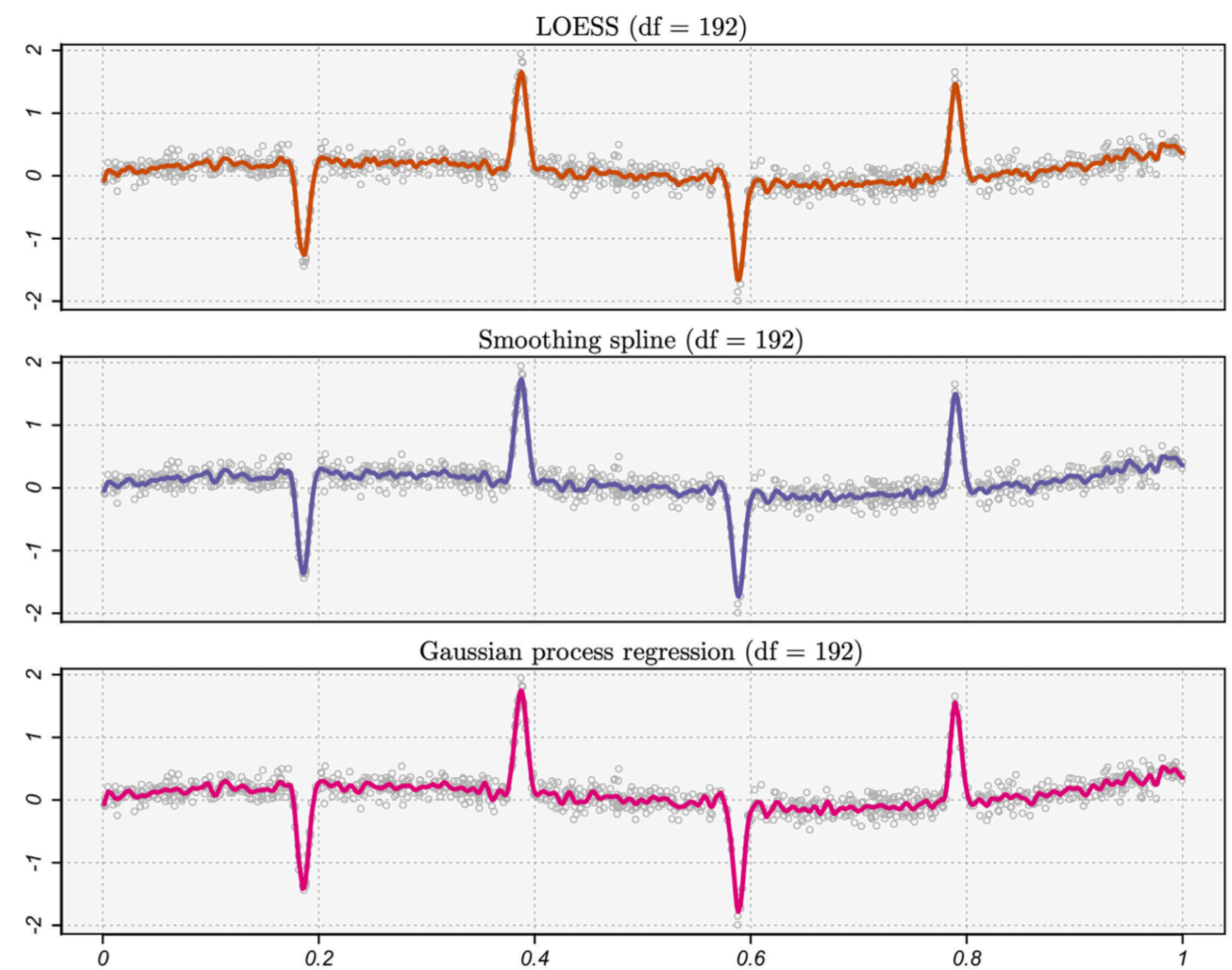


Sample comparison to other methods

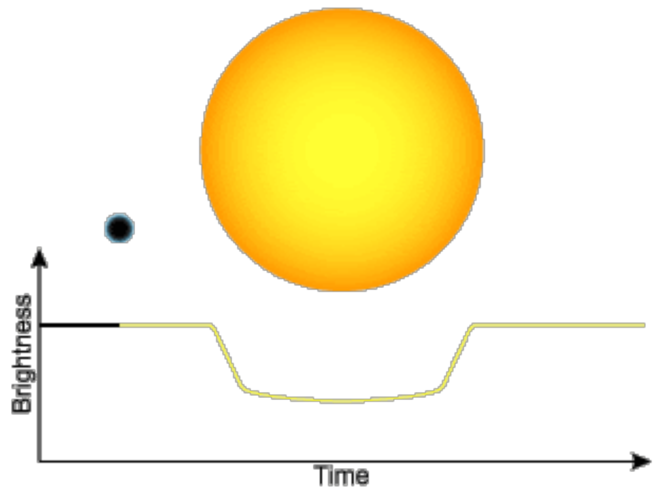


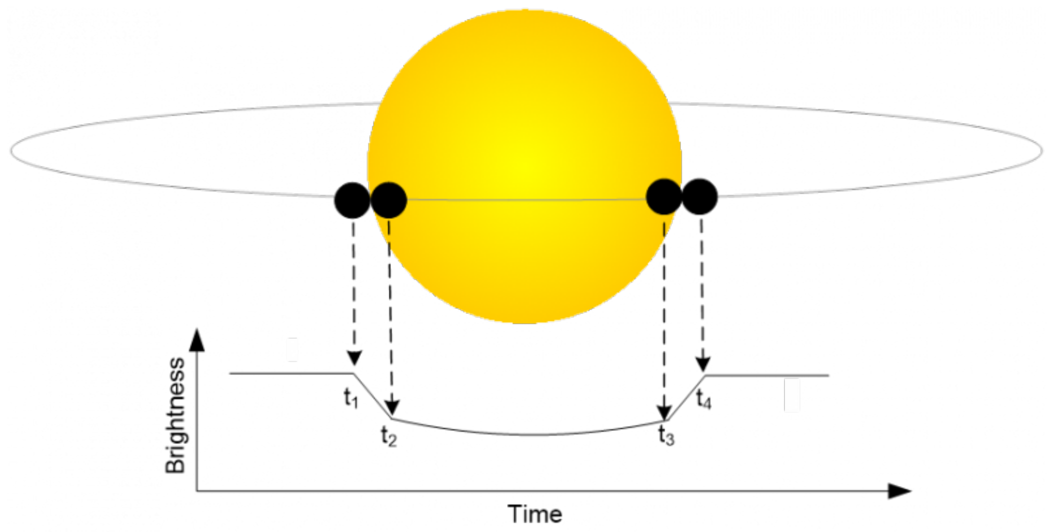
Sample comparison to other methods



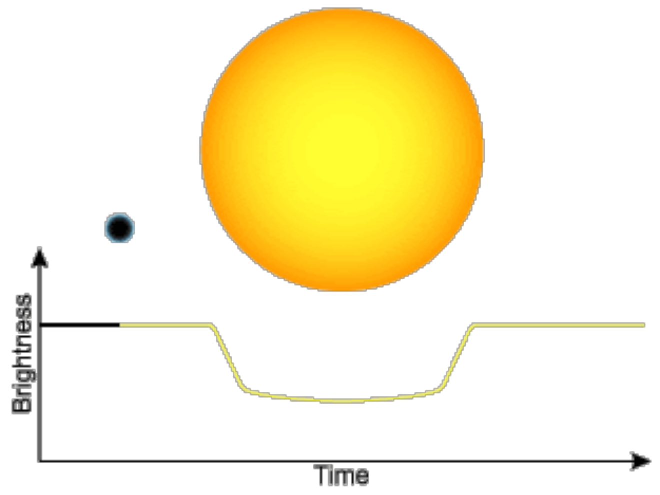


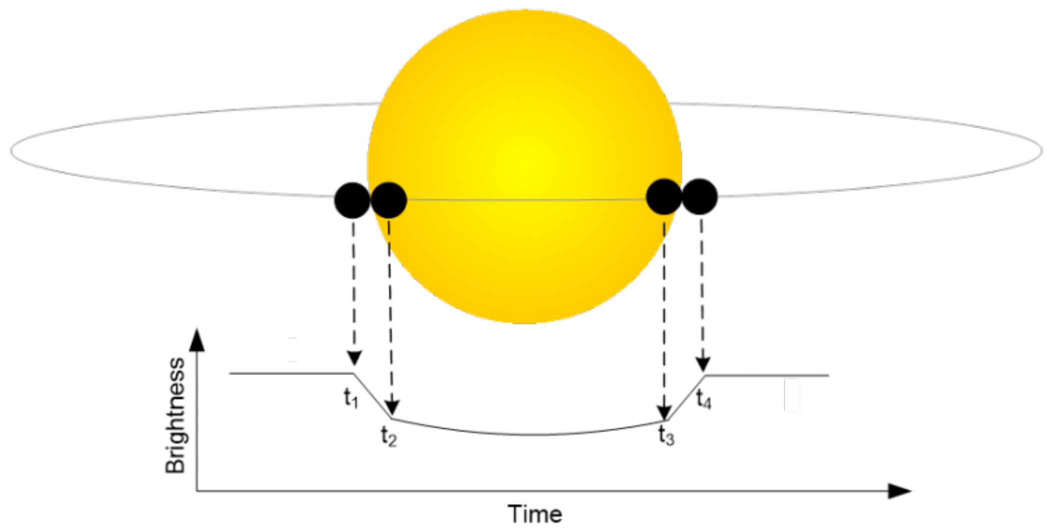
Exoplanet transit light curves





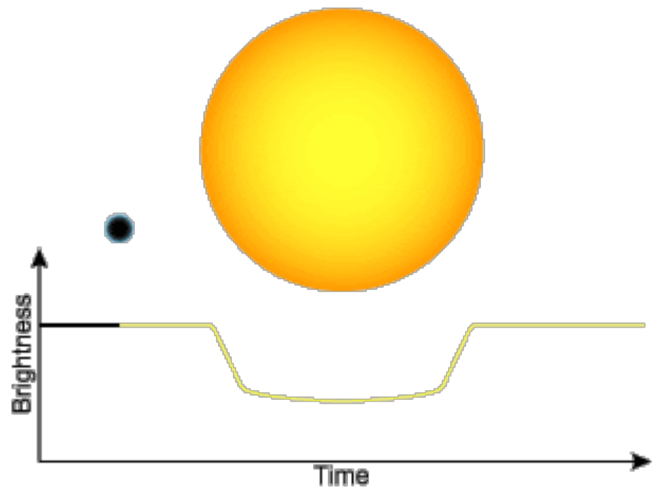
Exoplanet transit light curves

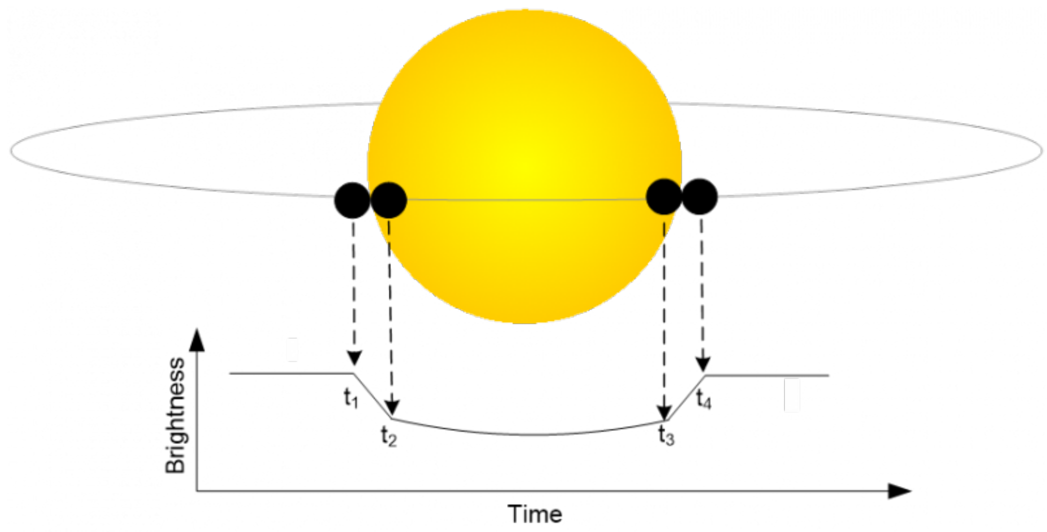




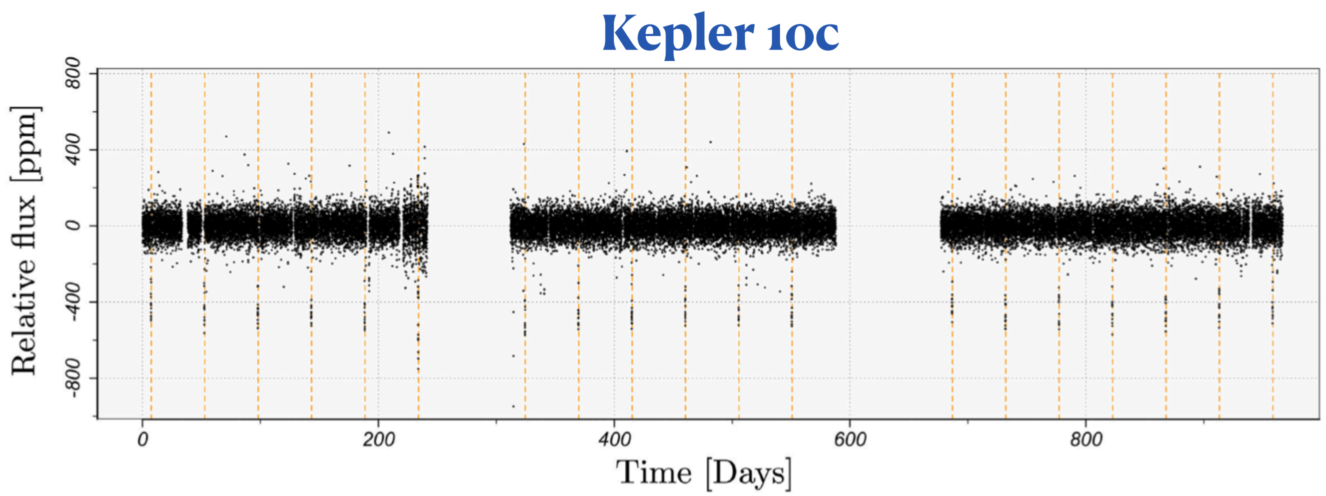
- **Relaxed trend filtering** can be used to efficiently estimate the transit duration and depth and provide a full uncertainty distribution

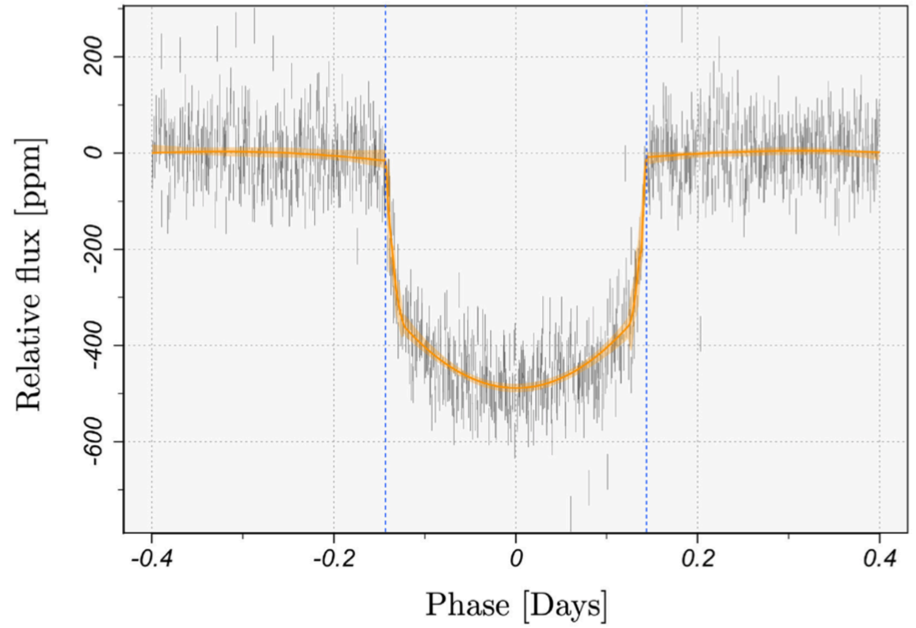
Exoplanet transit light curves

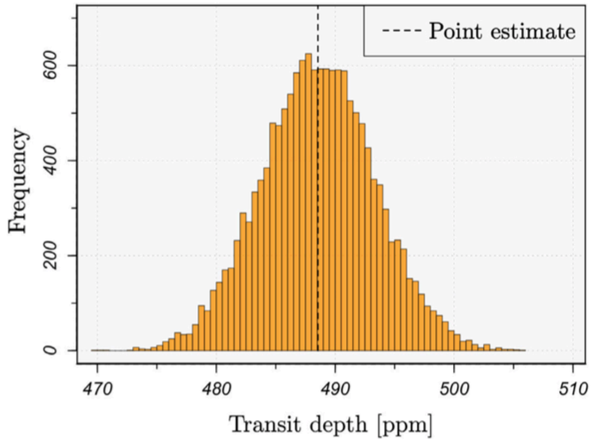


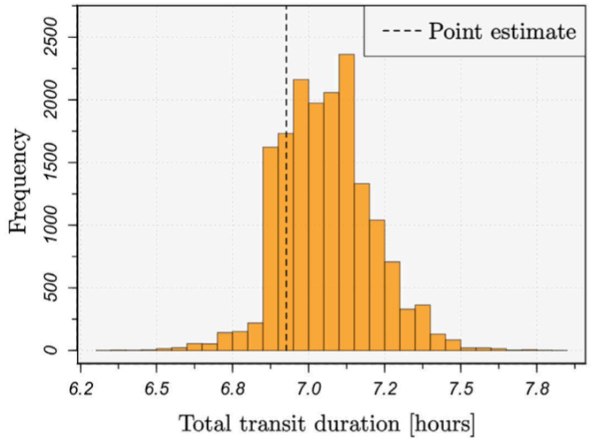


- **Relaxed trend filtering** can be used to efficiently estimate the transit duration and depth and provide a full uncertainty distribution

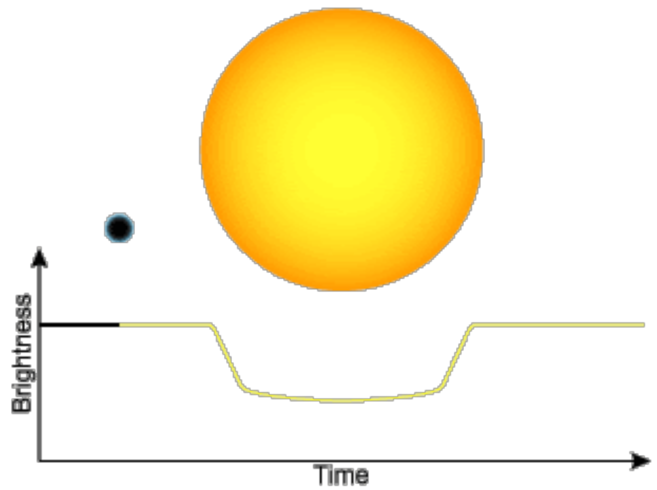


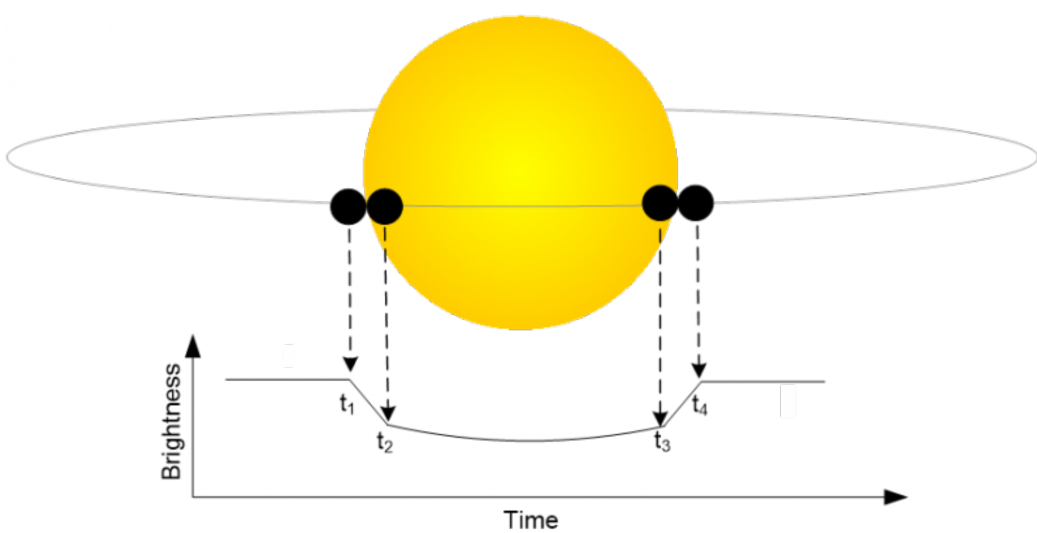


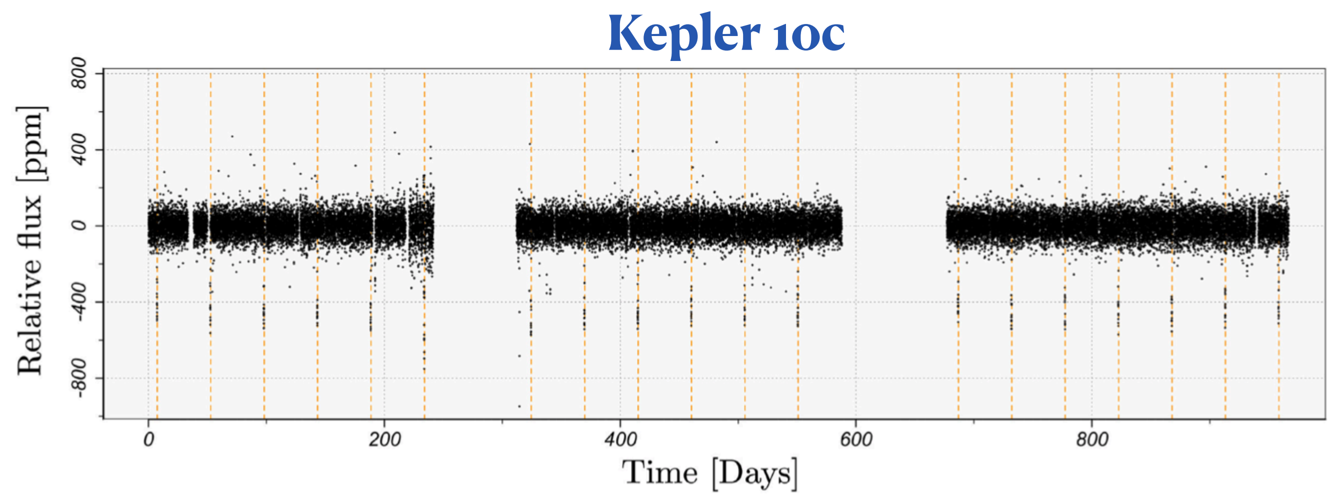


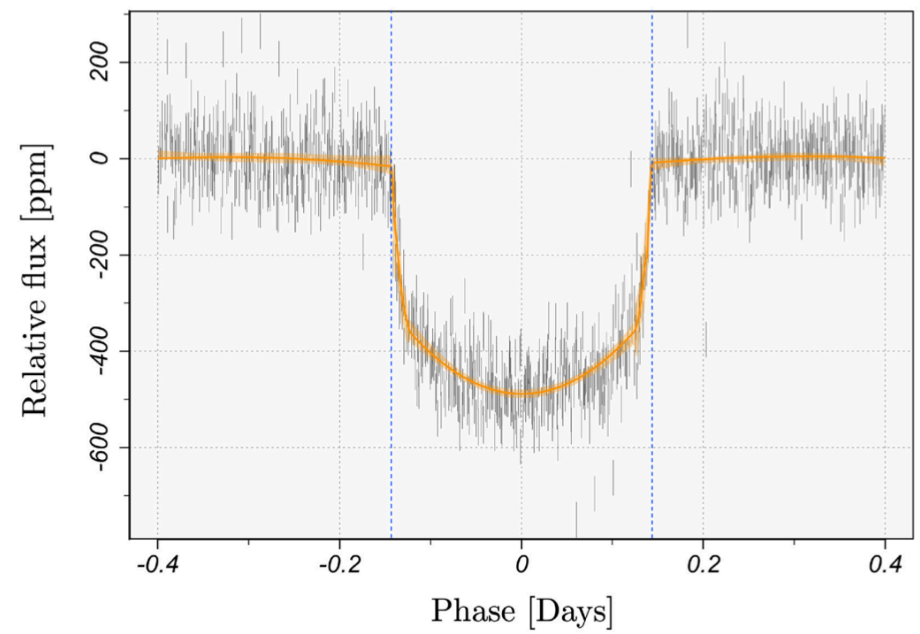


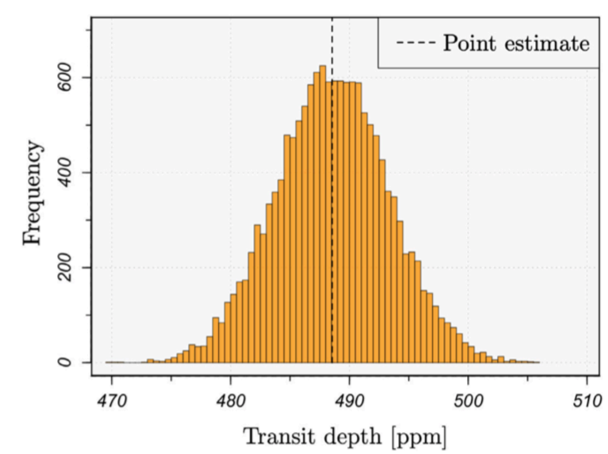
Exoplanet transit light curves

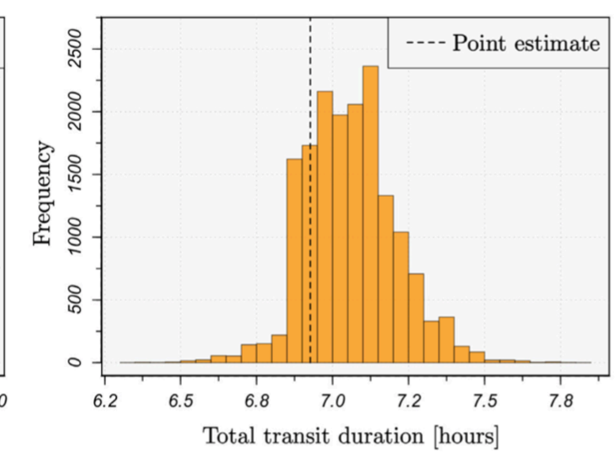






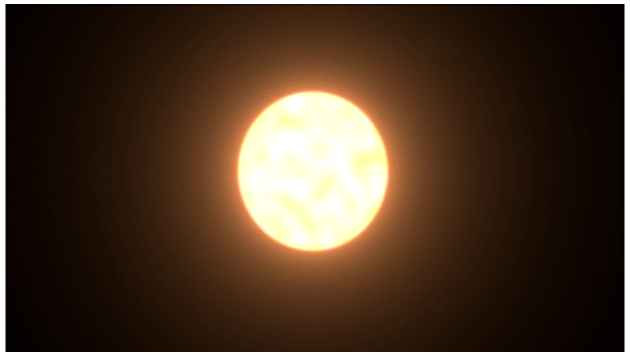


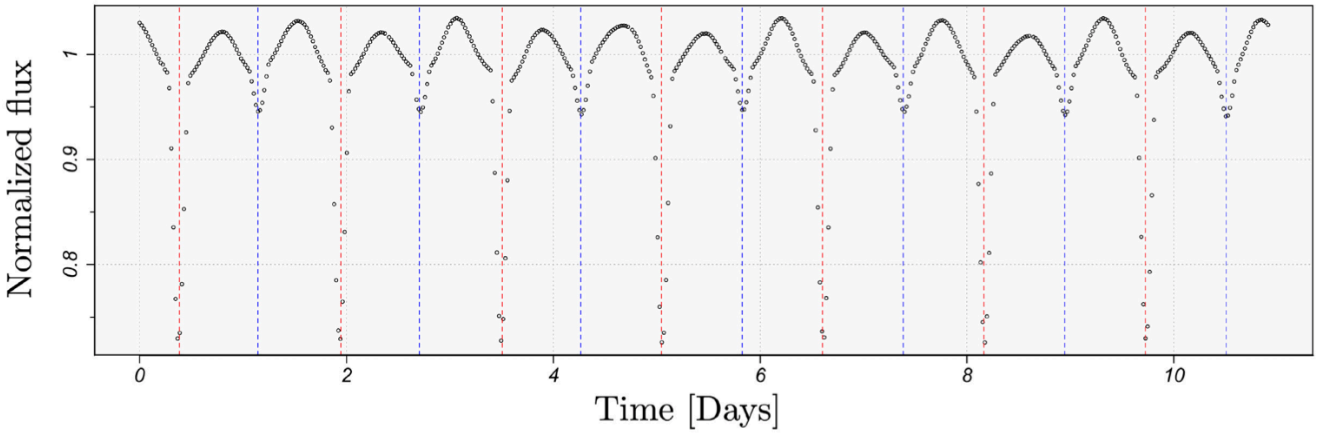




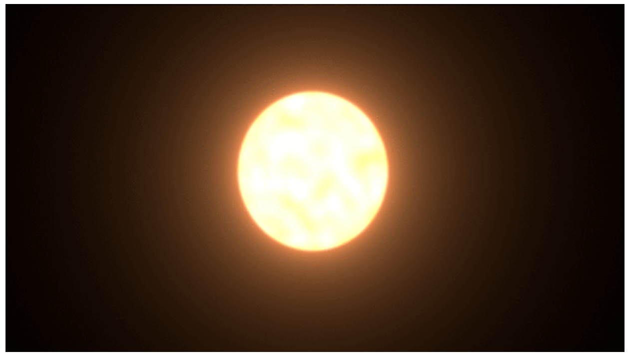
- **Relaxed trend filtering** can be used to efficiently estimate the transit duration and depth and provide a full uncertainty distribution
- Nonparametric + Fast alternative/supplement to Mandel & Agol analytic LC models

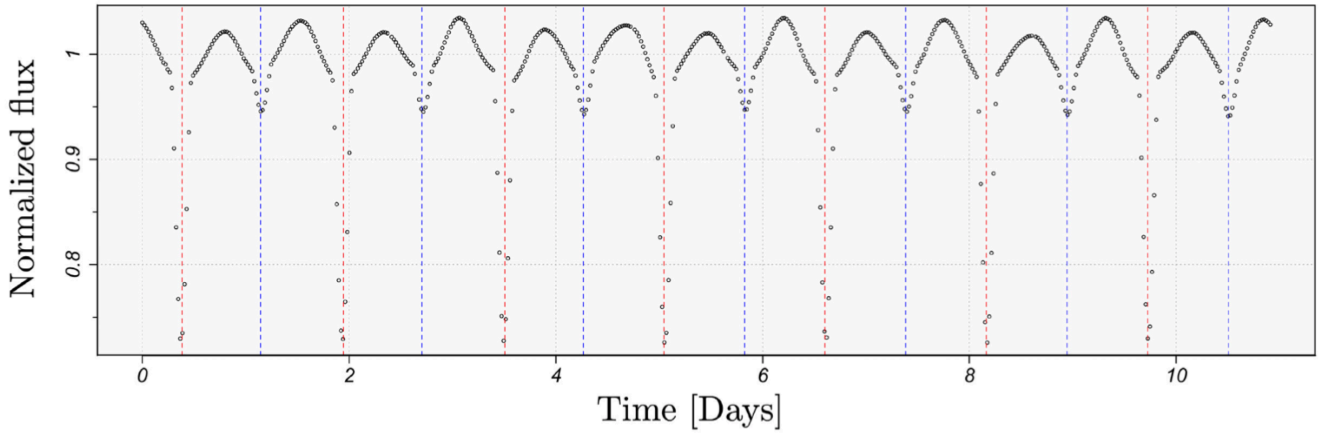
Eclipsing binary light curves

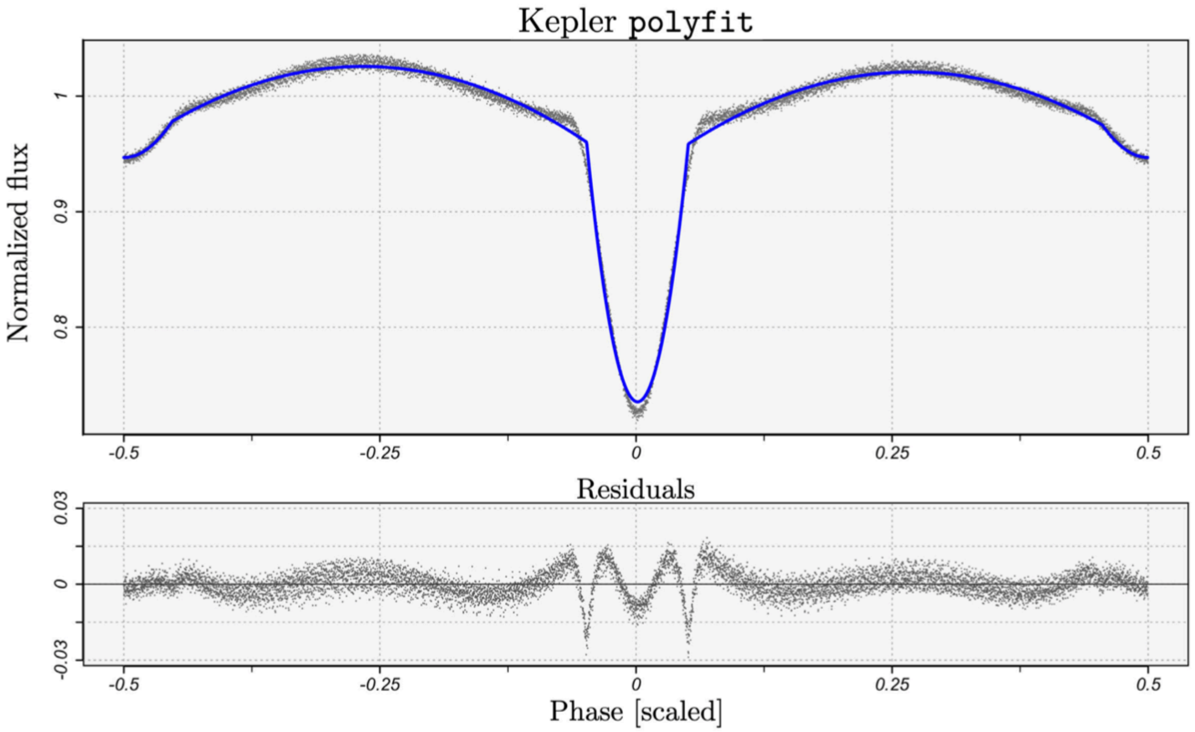




Eclipsing binary light curves



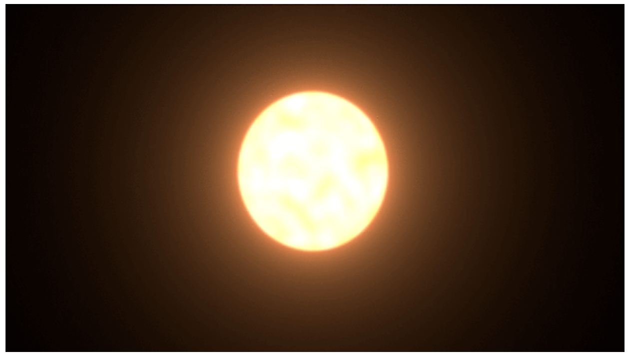


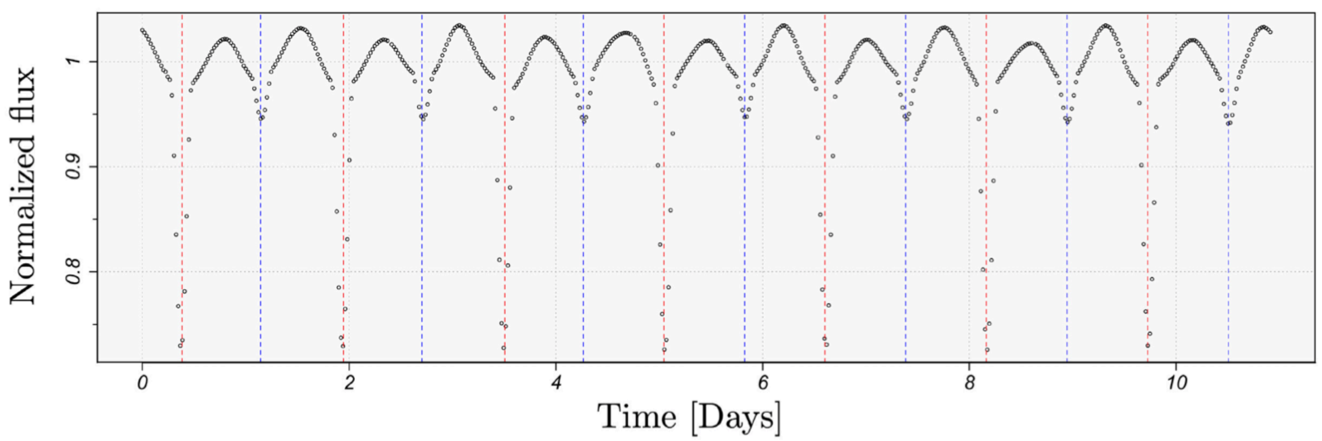


Kepler Eclipsing Binaries via Artificial Intelligence (EBAI) pipeline

Phase-folded LC → **Polyfit** → Denoised LC → **ANN** → EB physical parameters

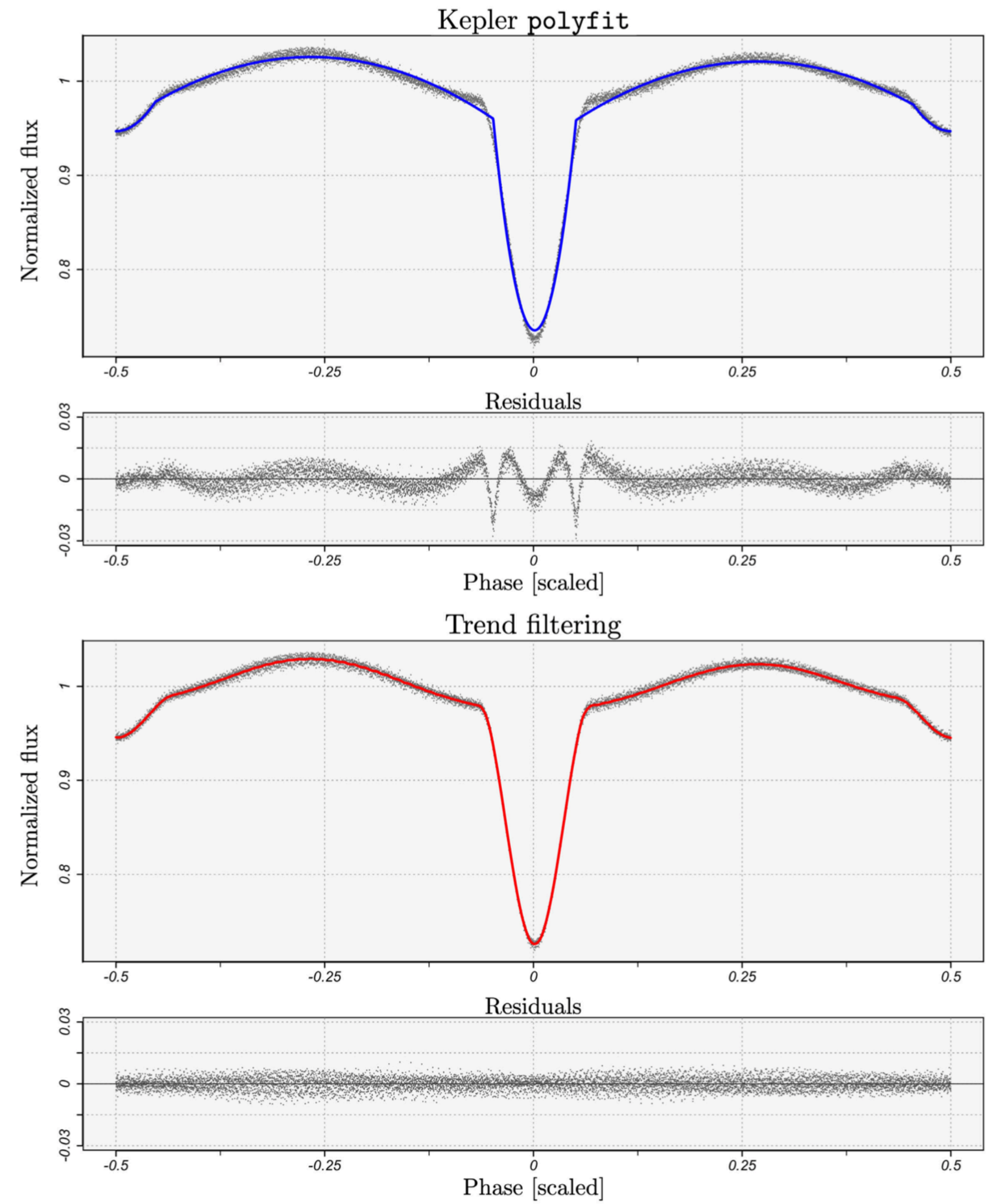
Eclipsing binary light curves



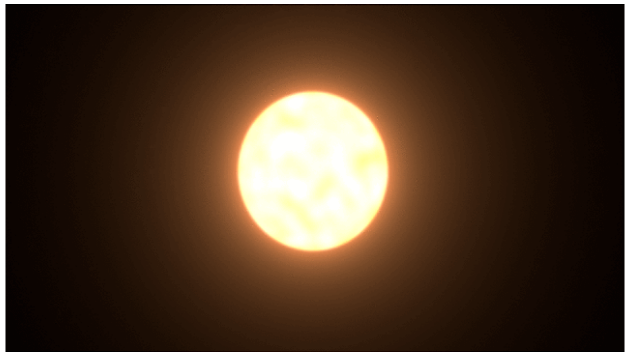


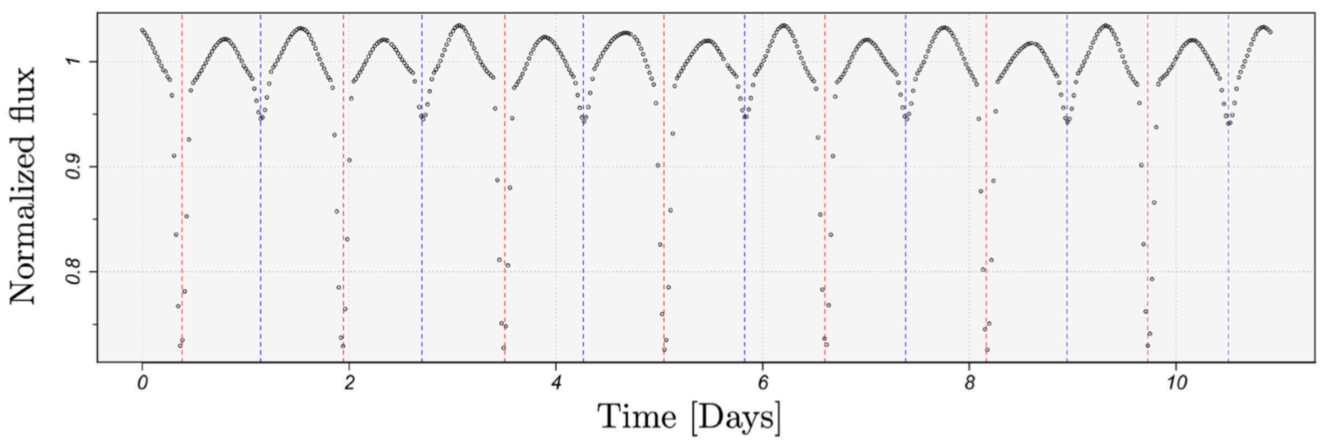
Kepler Eclipsing Binaries via Artificial Intelligence (EBAI) pipeline

Phase-folded LC → **Polyfit** → Denoised LC → **ANN** → EB physical parameters

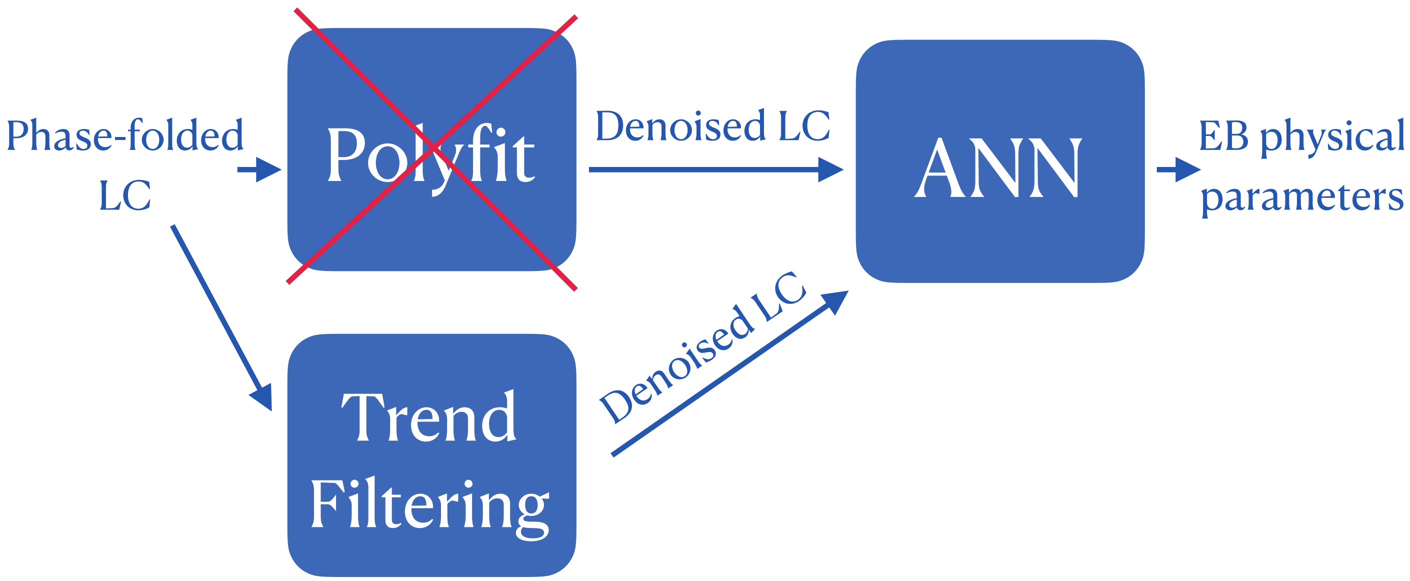


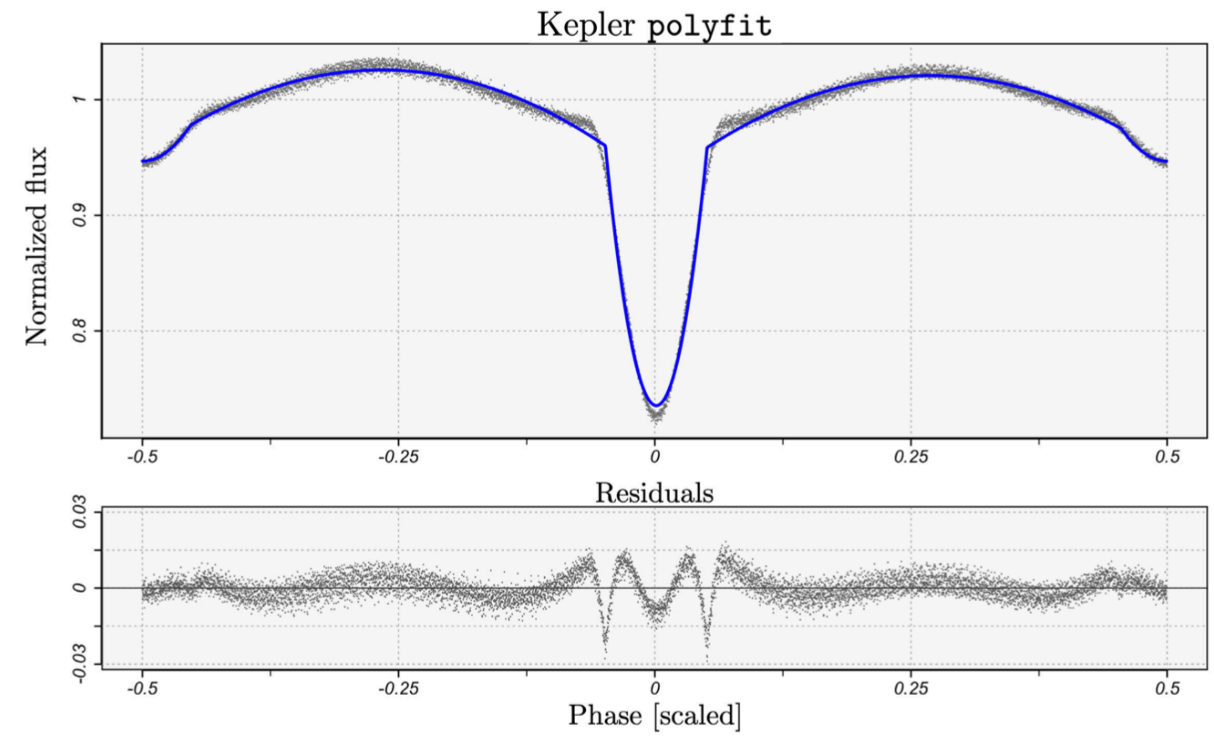
Eclipsing binary light curves

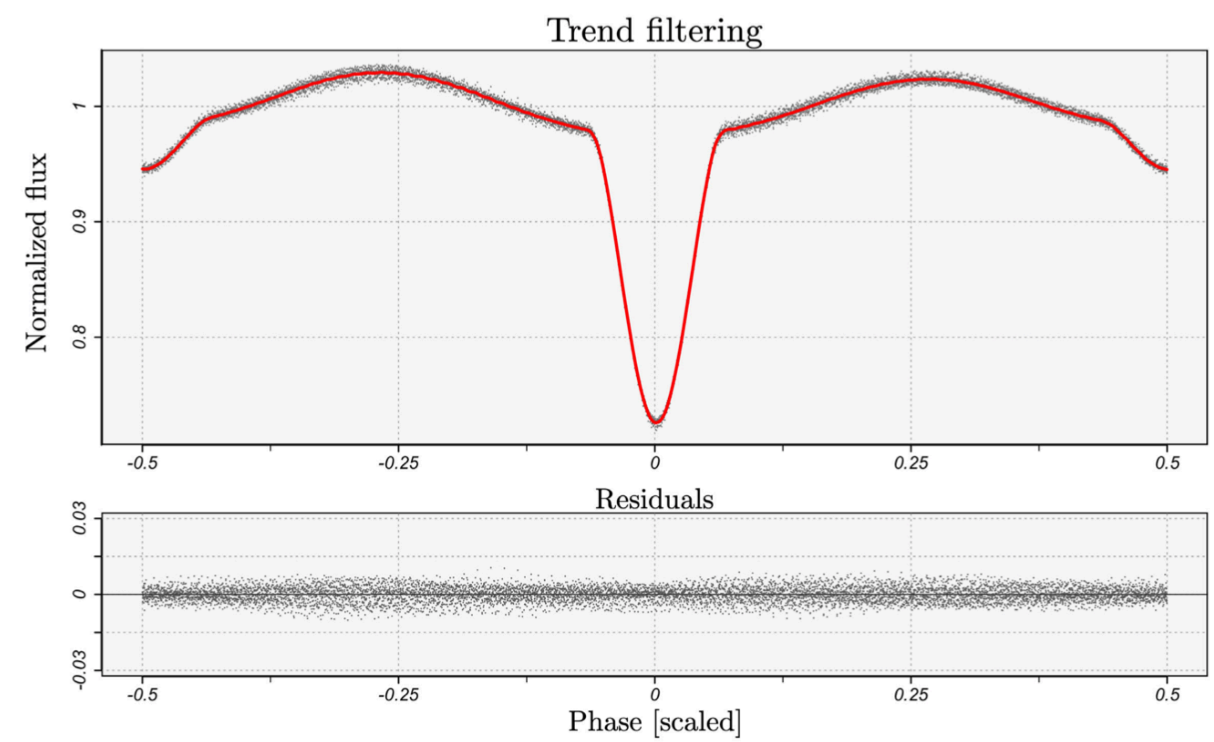




Kepler Eclipsing Binaries via Artificial Intelligence (EBAI) pipeline

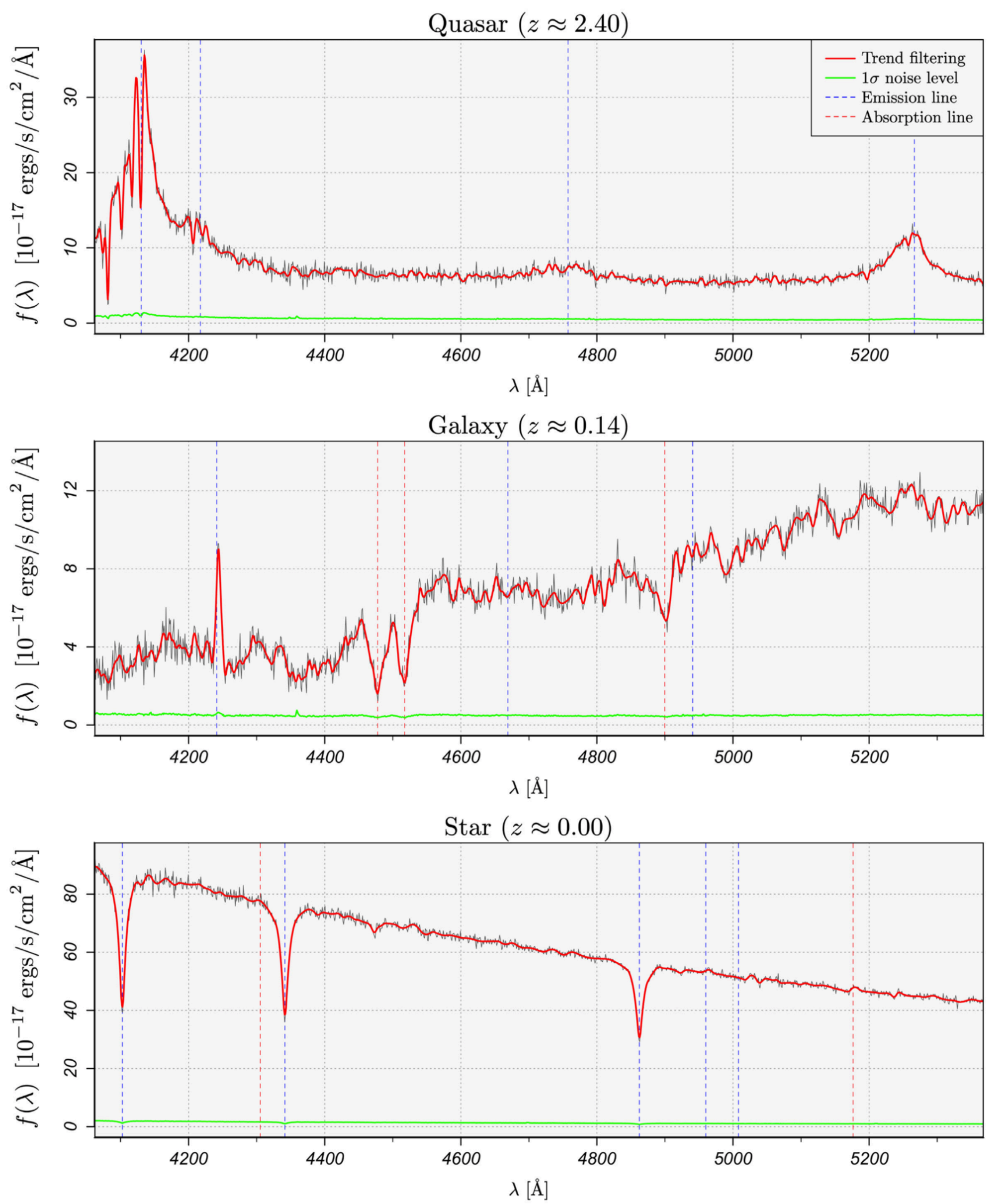






Spectral template generation

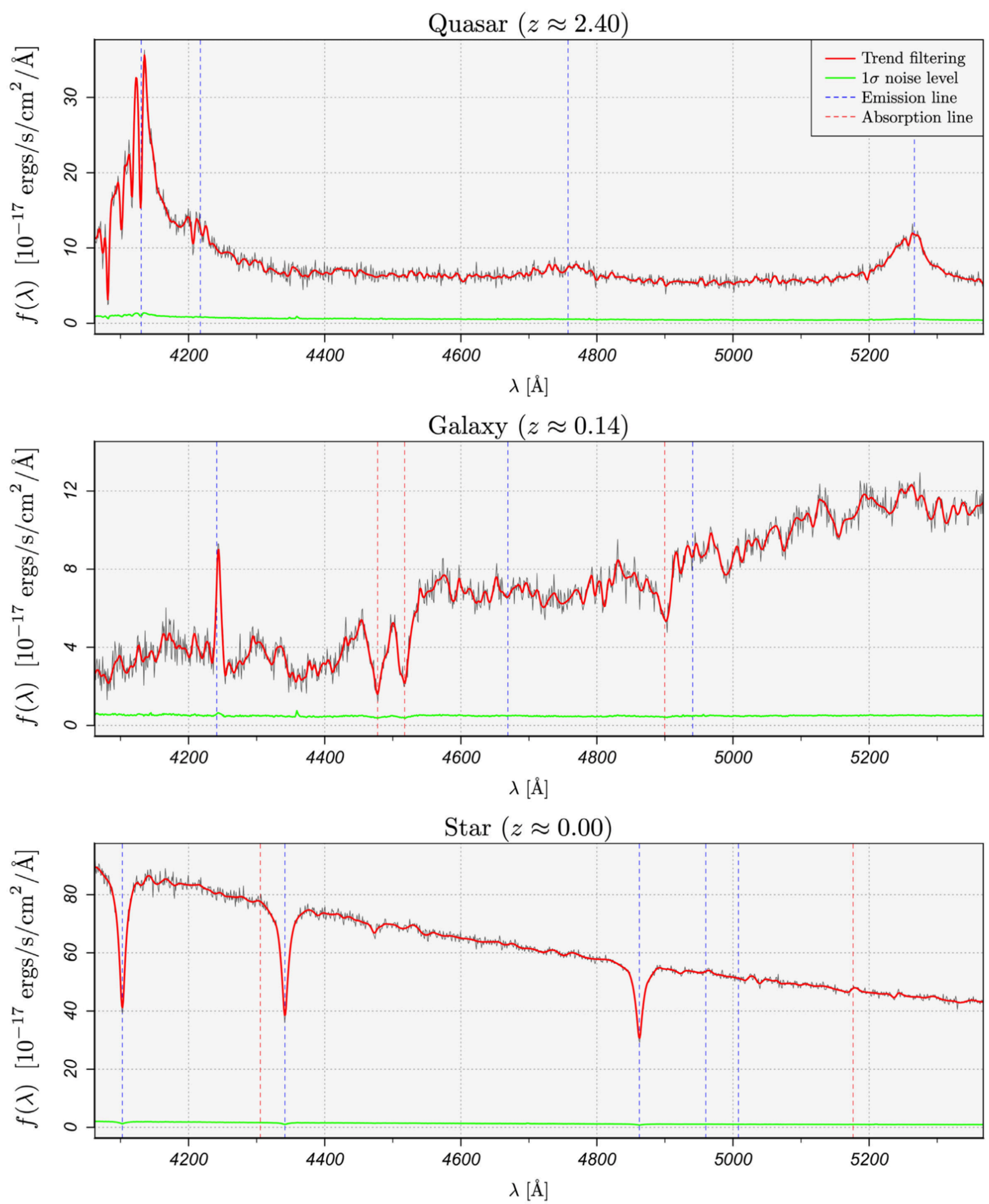
Automated spectral classification and redshift estimation algorithms rely heavily on template libraries that span the full physical diversity of observed objects



Spectral template generation

Automated spectral classification and redshift estimation algorithms rely heavily on template libraries that span the full physical diversity of observed objects

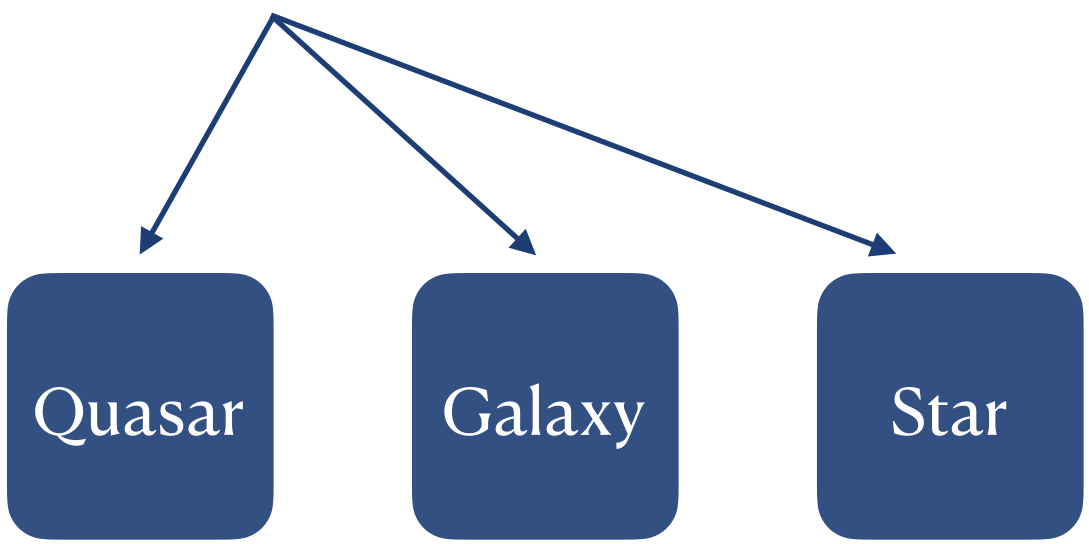
$$b(\lambda_{\text{rest}}) = \widehat{f}_0(\lambda/(z_0+1))$$



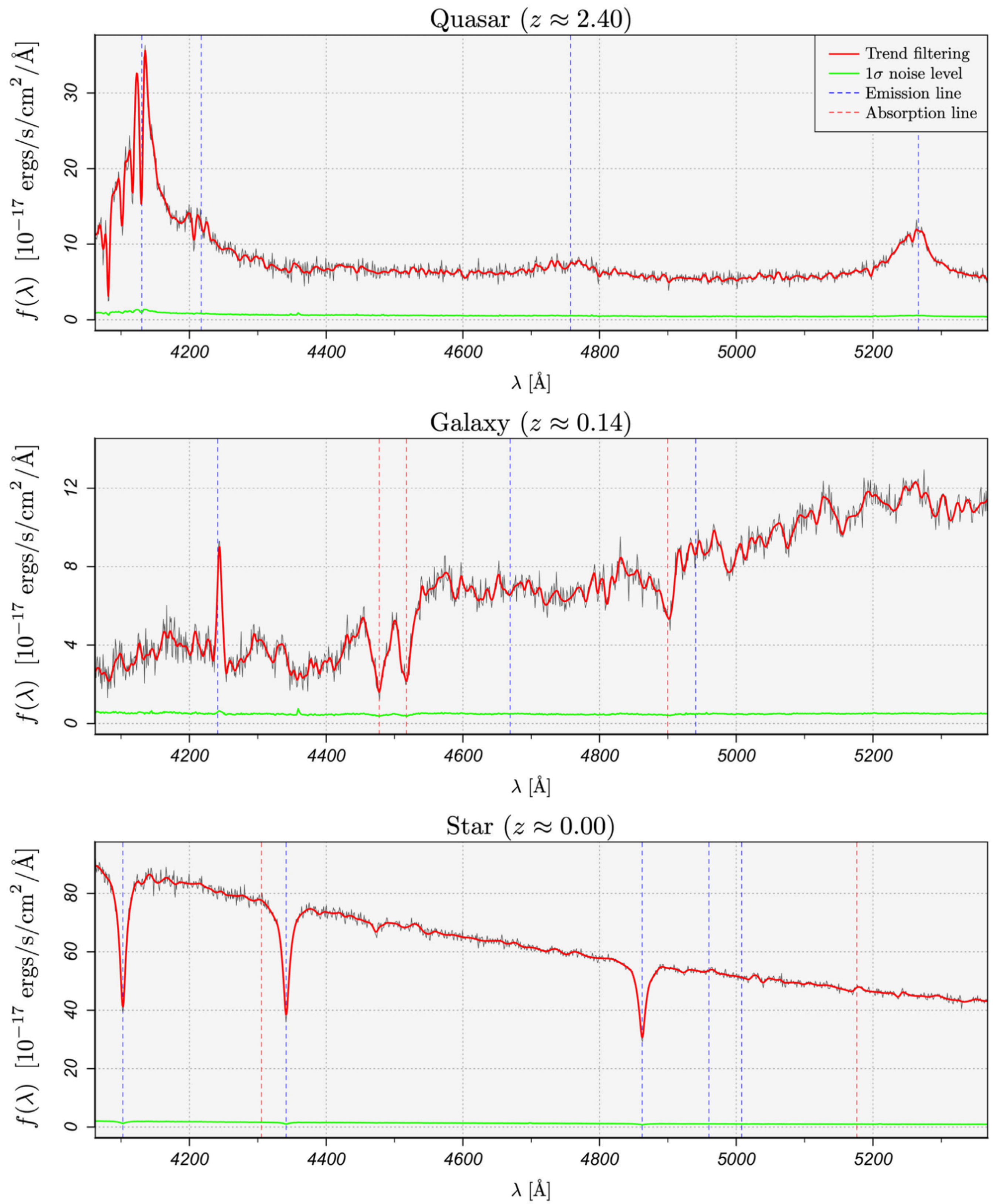
Spectral template generation

Automated spectral classification and redshift estimation algorithms rely heavily on template libraries that span the full physical diversity of observed objects

$$b(\lambda_{\text{rest}}) = \widehat{f}_0(\lambda/(z_0 + 1))$$

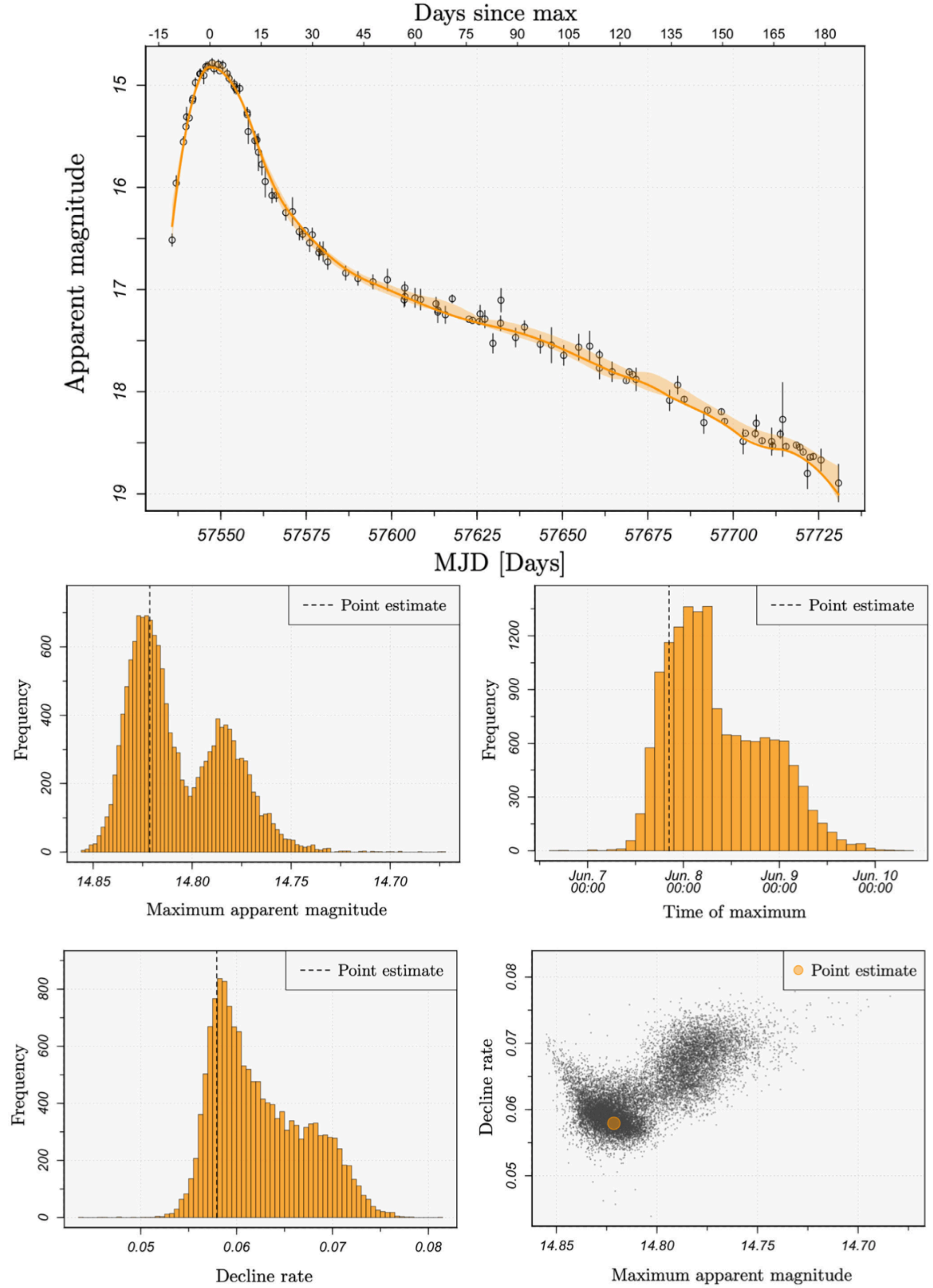


Spectral template libraries



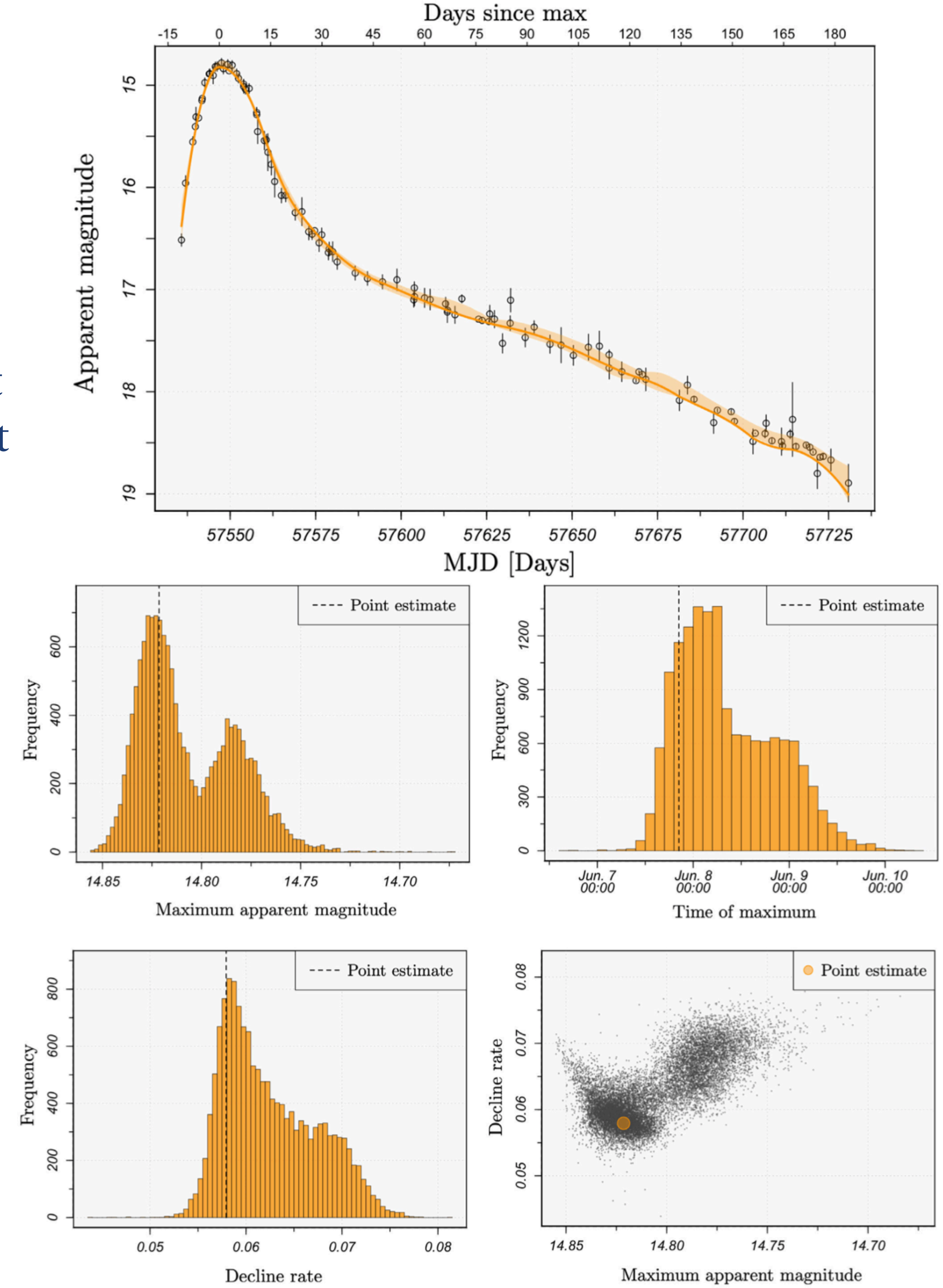
Supernova light curves

- Classification of SNe relies on light-curve/spectral template libraries that span the full physical diversity of observed objects



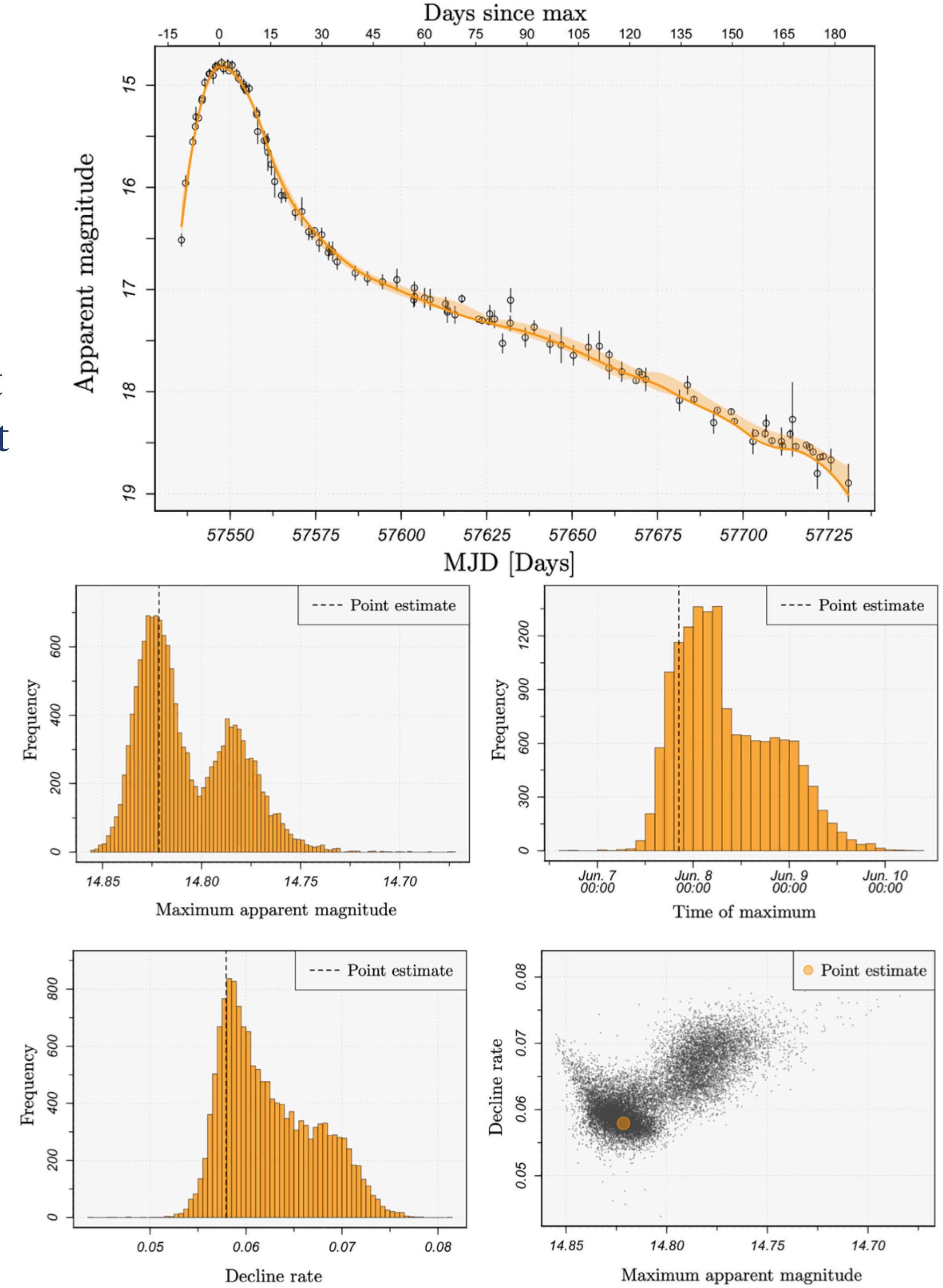
Supernova light curves

- Classification of SNe relies on light-curve/spectral template libraries that span the full physical diversity of observed objects
- Trend filtering improves upon the smoothing spline template generation of the Carnegie Supernova Project when the peak luminosity is high and decline rate is fast (e.g. Type Ia SNe)

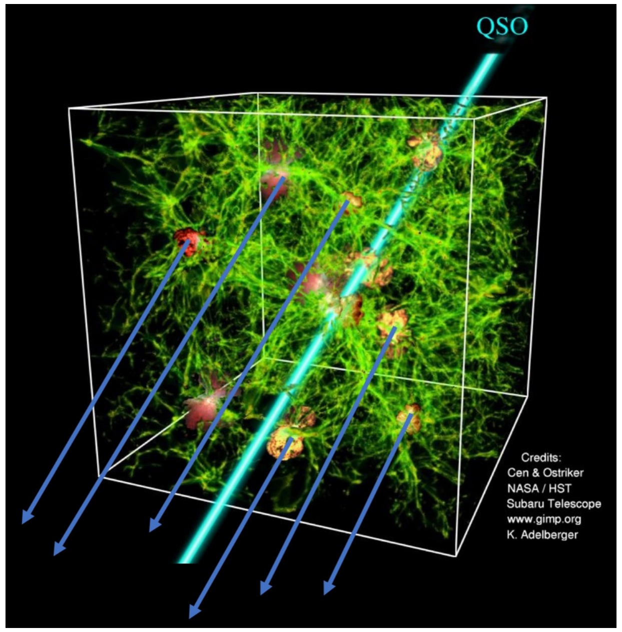


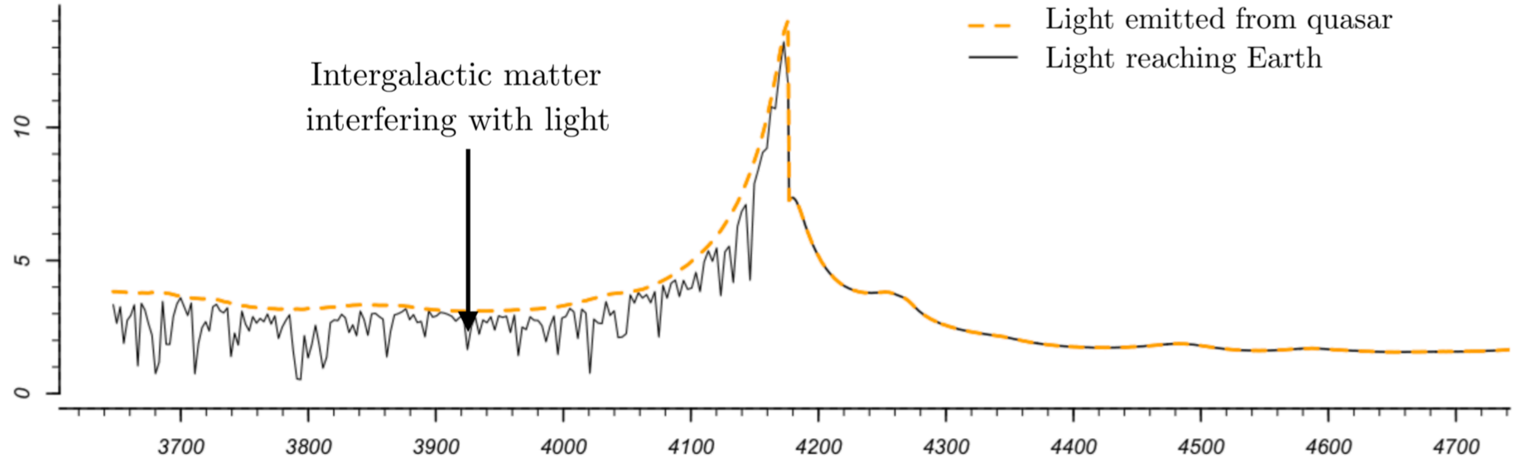
Supernova light curves

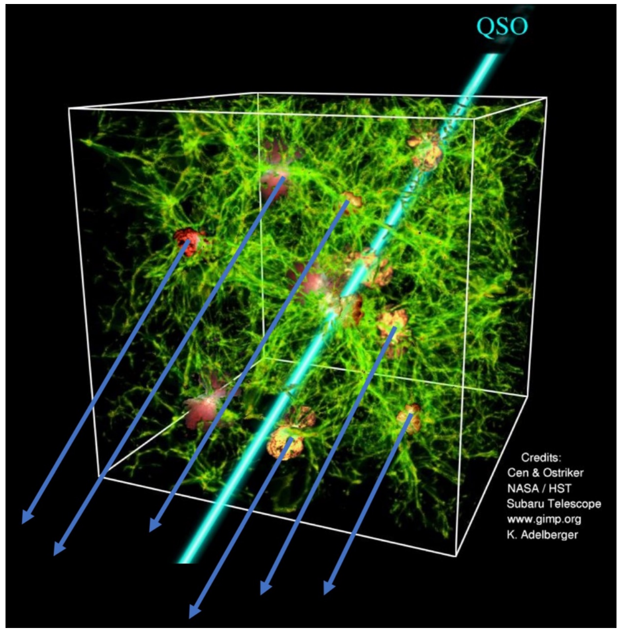
- Classification of SNe relies on light-curve/spectral template libraries that span the full physical diversity of observed objects
- Trend filtering improves upon the smoothing spline template generation of the Carnegie Supernova Project when the peak luminosity is high and decline rate is fast (e.g. Type Ia SNe)
- Trend filtering can produce efficient nonparametric estimates and uncertainty distributions for the maximum brightness, the time of maximum, and the decline rate

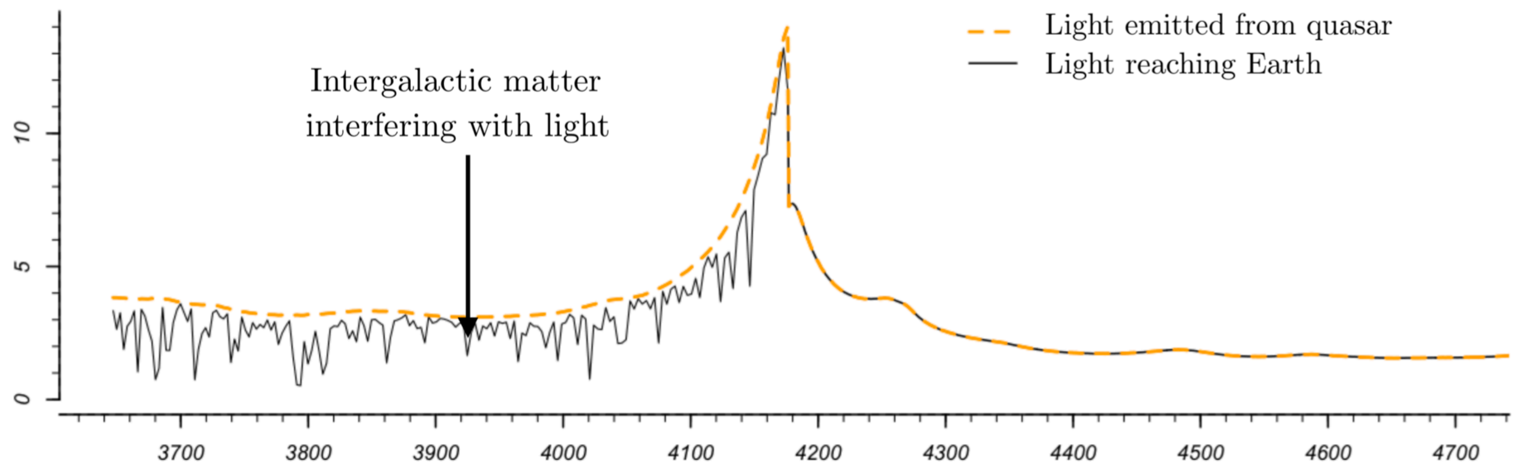


Mapping the intergalactic medium via the Lyman-alpha forest





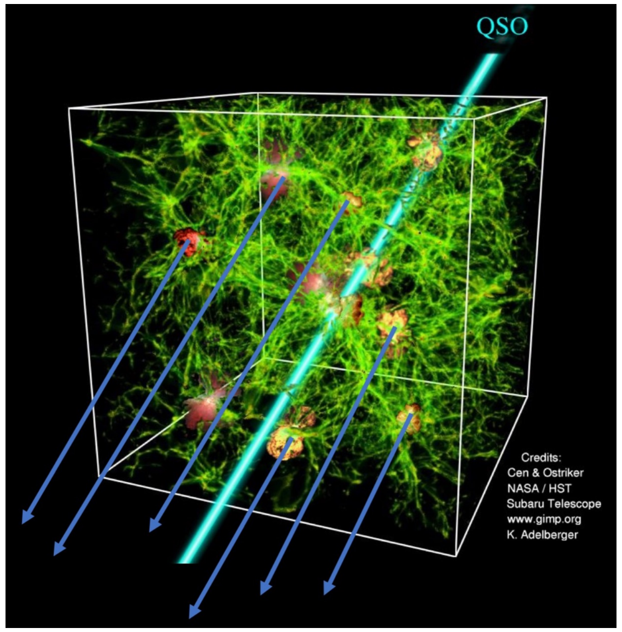


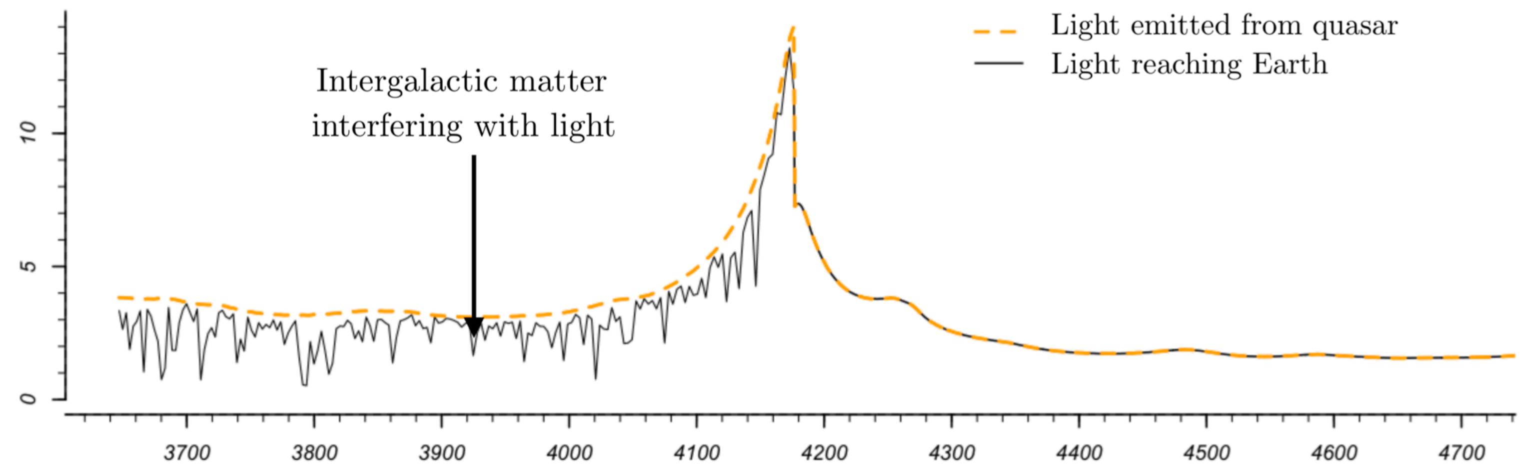


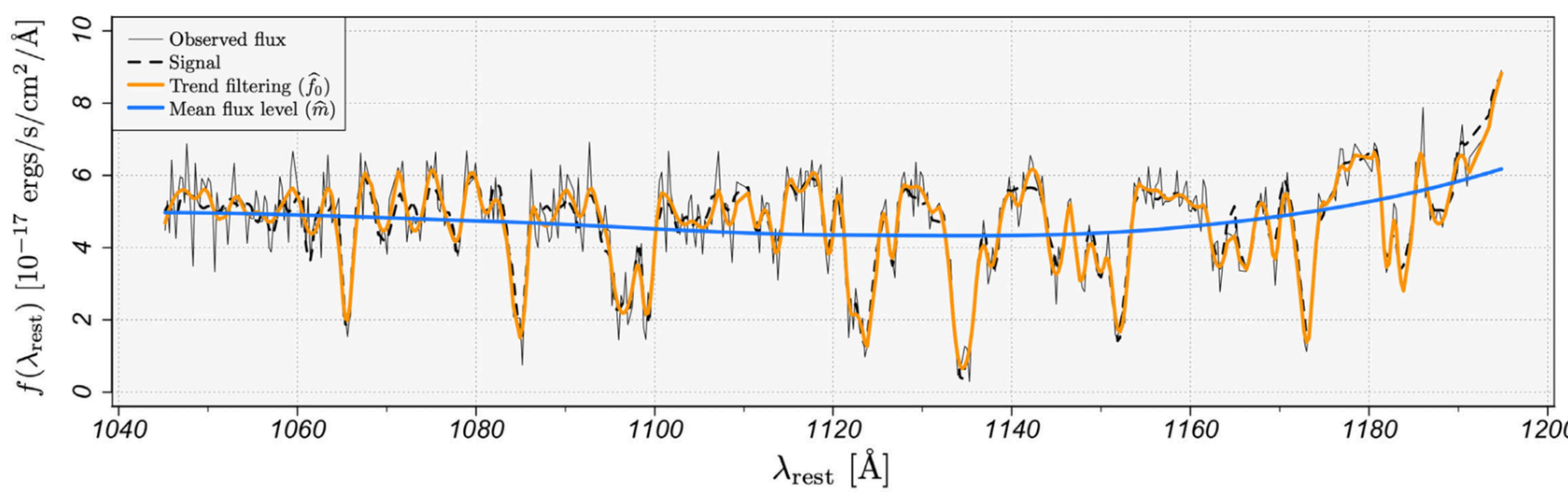
Lyman-α flux contrast:

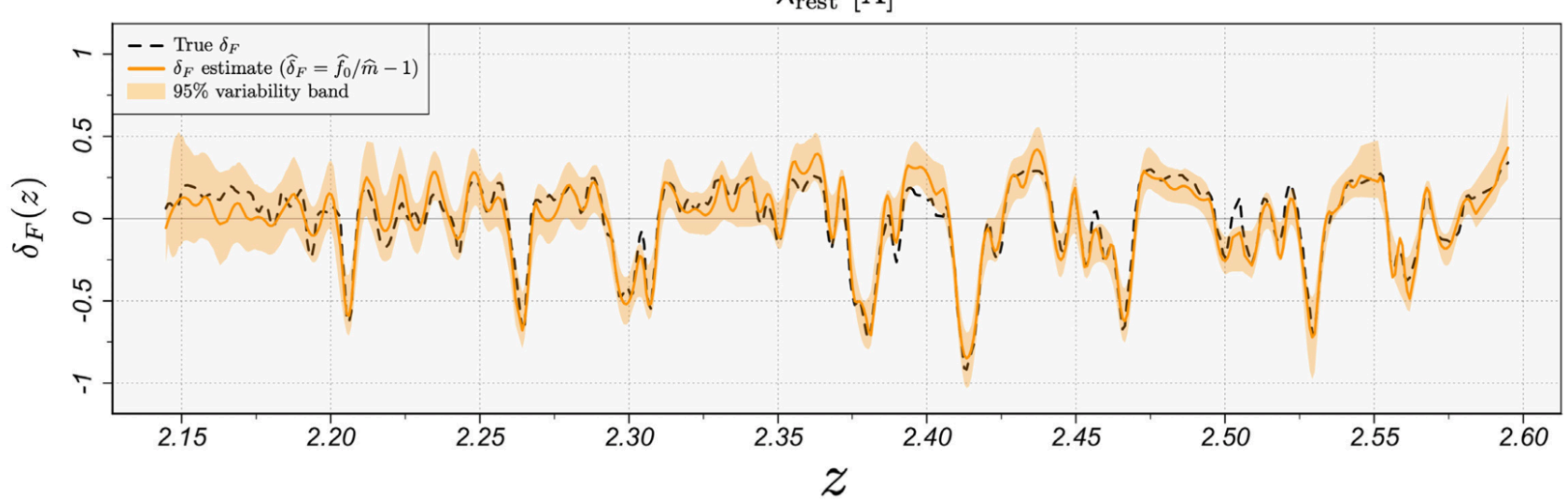
$$\delta_F(z) = \frac{F(z) - \overline{F}(z)}{\overline{F}(z)}$$
$$\propto \left(\frac{\rho(z) - \overline{\rho}(z)}{\overline{\rho}(z)}\right)^{-1}$$

Mapping the intergalactic medium via the Lyman-alpha forest







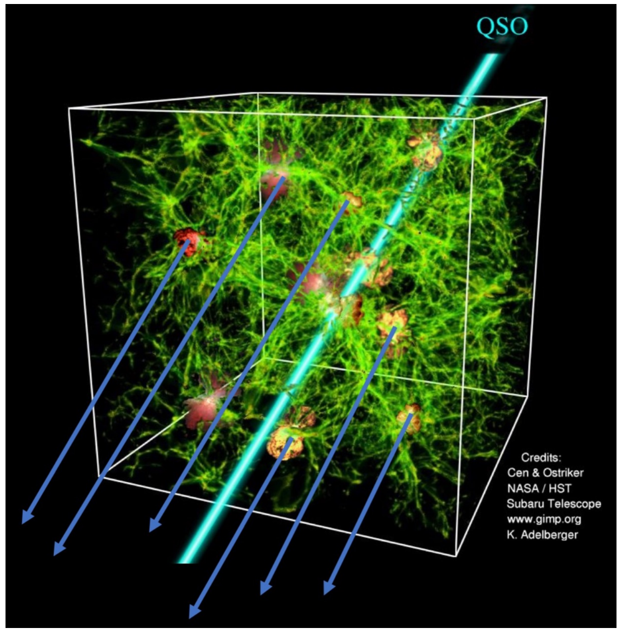


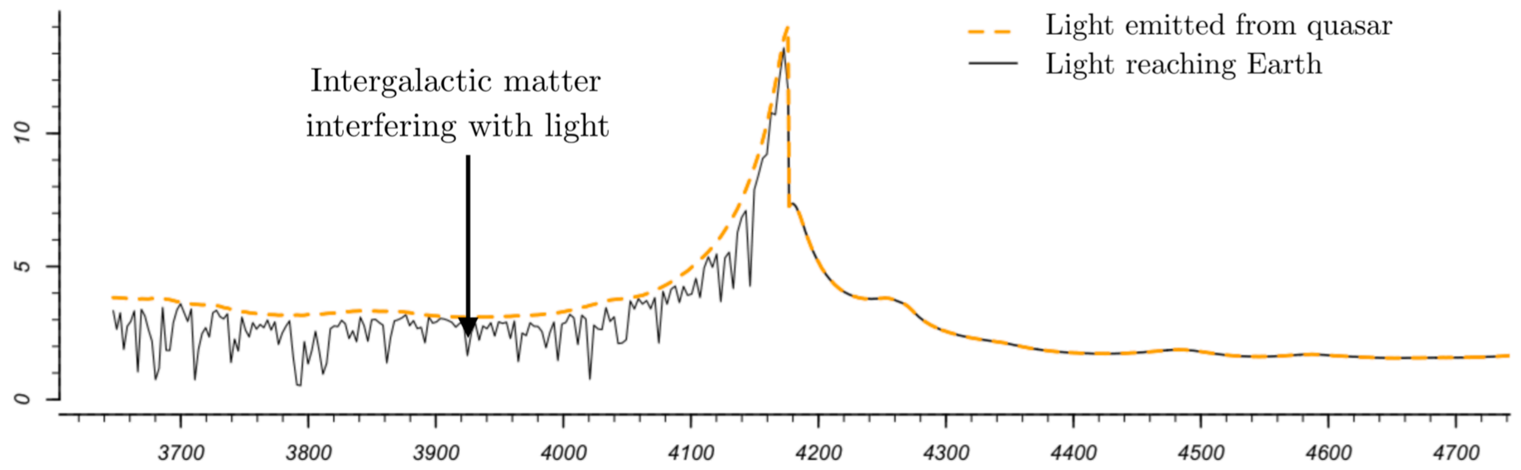
Lyman-α flux contrast:

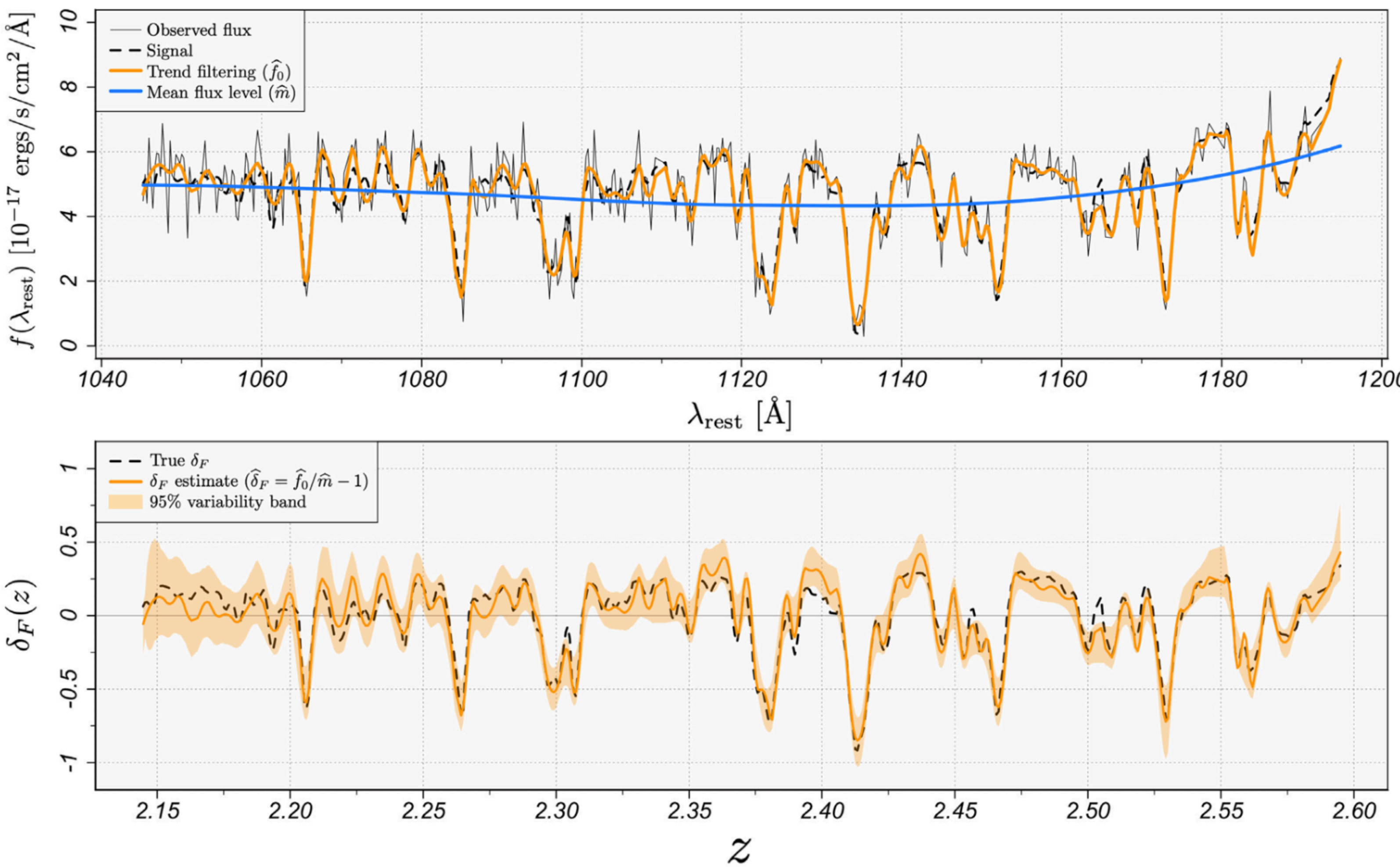
$$\delta_F(z) = \frac{F(z) - \overline{F}(z)}{\overline{F}(z)}$$

$$\propto \left(\frac{\rho(z) - \overline{\rho}(z)}{\overline{\rho}(z)}\right)^{-1}$$

Mapping the intergalactic medium via the Lyman-alpha forest







Lyman-α flux contrast:

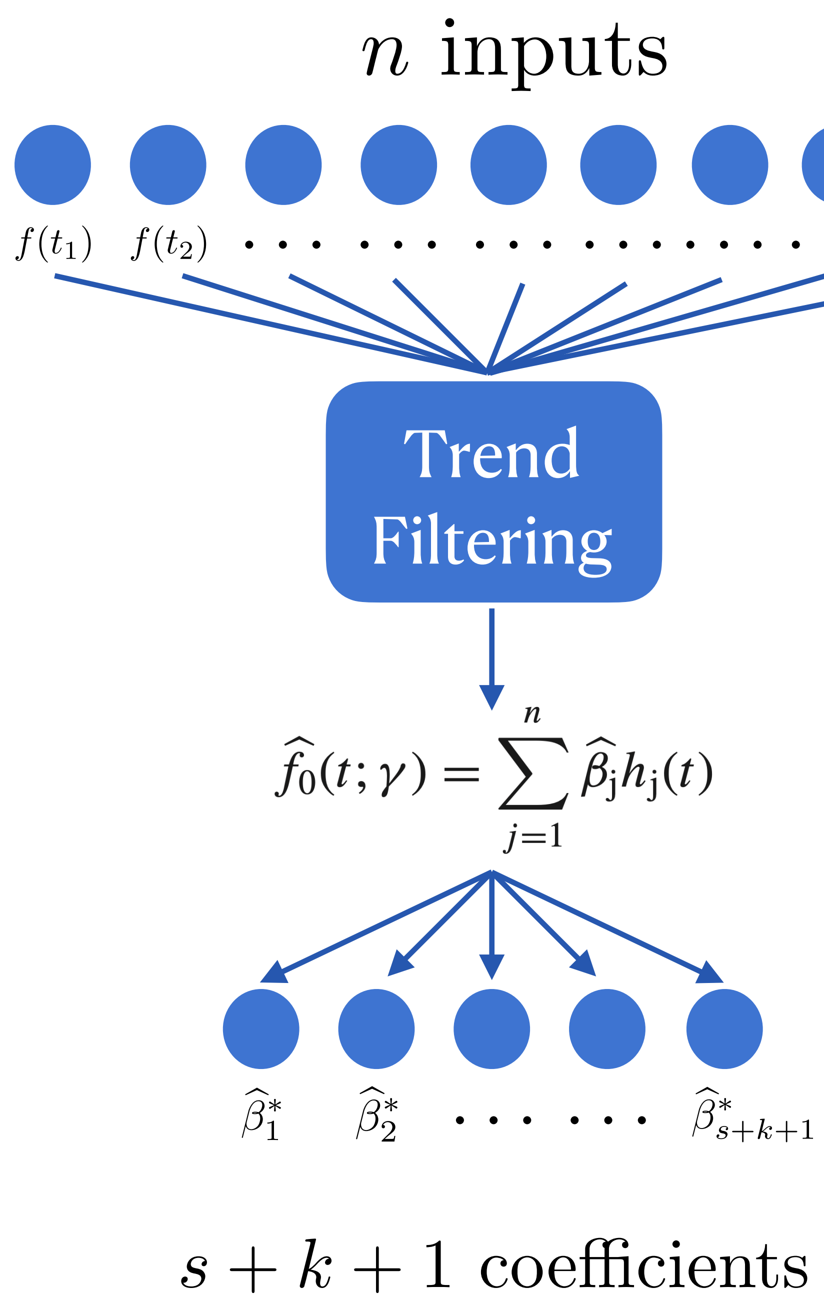
$$\delta_F(z) = \frac{F(z) - \overline{F}(z)}{\overline{F}(z)}$$

$$\propto \left(\frac{\rho(z) - \overline{\rho}(z)}{\overline{\rho}(z)}\right)^{-1}$$

Estimator:

$$\widehat{\delta}_F(z) = \frac{\widehat{f}_0(z) - \widehat{m}(z)}{\widehat{m}(z)}$$

Compression of one-dimensional data sets



References

C. A. Politsch et al. Trend Filtering - I. A Modern Statistical Tool for Time-Domain Astronomy and Astronomical Spectroscopy. Monthly Notices of the Royal Astronomical Society, 492(3), p. 4005-4018, 2020.
https://academic.oup.com/mnras/article/492/3/4005/5704413

C. A. Politsch et al. Trend Filtering - II. Denoising Astronomical Signals with Varying Degrees of Smoothness. Monthly Notices of the Royal Astronomical Society, 492(3), p. 4019-4032, 2020.
https://academic.oup.com/mnras/article/492/3/4019/5704414

R. J. Tibshirani & J. Taylor. The solution path of the generalized lasso. Ann. Statist. 39 (2011), no. 3, 1335-1371.
https://projecteuclid.org/euclid.aos/1304514656

R. J. Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. The Annals of Statistics. 42 (2014), no. 1, 285-323. https://projecteuclid.org/euclid.aos/1395234979

A. Ramdas & R. J. Tibshirani. Fast and Flexible ADMM Algorithms for Trend Filtering. Journal of Computational and Graphical Statistics, 25:3 (2016), 839-858.
https://amstat.tandfonline.com/doi/abs/10.1080/10618600.2015.1054033#.XfJpNpNKju0

T. B. Arnold, V. Sadhanala, and R. J. Tibshirani. Fast algorithms for generalized lasso problems.
https://github.com/glmgen. Version 0.0.3 (2014)

R. J. Tibshirani. Divided Differences, Falling Factorials, and Discrete Splines: Another Look at Trend Filtering and Related Problems. *arXiv:2003.03886* (2020).

Y. Wang, A. Smola, and R. J. Tibshirani. The falling factorial basis and its statistical properties. In International Conference on Machine Learning, vol. 31, p. 9. 2014.
http://proceedings.mlr.press/v32/wange14.html

Thank you!