

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

Collin A. Politsch<sup>1</sup>, Jessi Cisewski-Kehe<sup>2</sup>, Rupert A.C. Croft<sup>1</sup>, Larry Wasserman<sup>1</sup>

<sup>1</sup> Carnegie Mellon University, Pittsburgh, PA 15213

<sup>2</sup> Yale University, New Haven, CT 06520

**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- $\alpha$  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- $\alpha$  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.** Trend filtering, Lyman- $\alpha$  forest, spectroscopy, light curves, exoplanet transits, eclipsing binaries

**1. Introduction.** The contents of this paper were published as a two-part series in [101] and [102], and were also submitted as part of the first author’s doctoral dissertation [100]. 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., 29, 126]. 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., 117]. 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., 55, 21], local polynomial regression [LOESS; e.g., 77, 97], splines [e.g., 96, 20, 26], Gaussian process regression [e.g., 48, 1, 52], and wavelet decompositions [e.g., 46, 119, 51]. A rich and elegant statistical literature exists on the theoretical and practical achievements of these methods [see, e.g., 54, 133, 56] 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 nonparametric methods called *linear smoothers*, which has been shown to be uniformly suboptimal for estimating spatially heterogeneous signals [85, 84, 31]. 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, continuing with the example of a smoothly varying signal with occasional sharp features, a linear smoother will tend to oversmooth the sharp features and/or overfit the smooth regions in its effort to optimally balance statistical bias and variance. Considerable effort has been made to address this problem by locally varying the hyperparameter(s) of a linear smoother. For example, locally varying the kernel bandwidth [e.g., 82, 40, 41, 72, 49] irregularly varying spline knot locations [e.g., 24, 62, 28], and constructing non-stationary covariance functions for Gaussian process regression [e.g., 111, 88, 89]. However, since hyperparameters typically need to be estimated from the data, such exponential increases in the hyperparameter complexity severely limit the practicality of choosing the hyperparameters in a fully data-driven, generalizable, and computationally efficient fashion. Wavelet decompositions offer an elegant solution to the problem of estimating spatially heterogeneous signals, providing both statistical optimality [e.g., 30, 31] and only requiring data-driven tuning of a single (scalar) hyperparameter. Wavelets, however, possess the practical limitation of requiring a stringent analysis setting, e.g. equally-spaced inputs and sample size equal to a power of two, among other provisions; and when these conditions are violated, the optimality guarantees are void. So, seemingly at an impasse, the motivating question for this work is *can we have the best of both worlds?* More precisely, is there a statistical tool that simultaneously possesses the following properties:

- P1.** Statistical optimality for estimating spatially heterogeneous signals
- P2.** Practical analysis assumptions; for example, not limited to equally-spaced inputs
- P3.** Practical computational speed
- P4.** A one-dimensional hyperparameter space, with automatic data-driven methods for selection

In this paper we introduce trend filtering [114, 63, 122], a statistical method that is new to the astronomical literature and provides a strong affirmative answer to this question.

The layout of the paper is as follows. In Section 2 we provide both theoretical and empirical evidence of the superiority of trend filtering for estimating spatially heterogeneous signals compared to classical statistical methods. In Section 3 we introduce trend filtering, including a general overview of the estimator’s machinery, its connection to spline methods, automatic methods for choosing the hyperparameter, uncertainty quantification, generalizations, and recommended software implementations in various programming languages. In Section 4 we use trend filtering to rigorously study the Lyman- $\alpha$  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. Finally, in Section 5 we illustrate the utility of trend filtering to a much wider variety of one-dimensional problems across astronomical spectroscopy and time-domain astronomy (necessarily discussing each application less comprehensively than our Lyman- $\alpha$  forest study). In particular, we discuss how trend filtering can be used to (1) generate galaxy, quasar, and stellar spectral templates and produce robust estimates of emission-line parameters; (2) model exoplanet transit events in phase-folded stellar light curves, providing transit depth and duration measurements; (3) improve upon the `polyfit` algorithm of [106] for denoising phase-folded eclipsing binary light curves; (4) generate supernova light-curve templates and produce nonparametric estimates of the observable parameters—e.g., the maximum apparent magnitude, the time of maximum, and the decline rate. Beyond its uses for astronomical data analysis, we also briefly discuss its relevance to large-scale one-dimensional data reduction and compression.

**2. Classical statistical methods and their limitations.** We begin this section by providing background and motivation for the nonparametric approach to estimating (or denoising) signals. We then discuss statistical optimality for estimating spatially heterogeneous signals, with an emphasis on providing evidence for the claim that trend filtering is broadly superior to classical statistical methods in this setting. Finally, we end this section by illustrating this superiority with a direct empirical comparison of trend filtering and several popular classical methods on simulated observations of a spatially heterogeneous signal.

2.1. *Nonparametric regression.* Suppose we observe noisy measurements of a response variable of interest (e.g., flux, magnitude, photon counts) according to the data generating process (DGP)

$$(1) \quad f(t_i) = f_0(t_i) + \epsilon_i, i = 1, \dots, n$$

where  $f_0(t_i)$  is the signal at input  $t_i$  (e.g., a time or wavelength) and  $\epsilon_i$  is the noise at  $t_i$  that contaminates the signal, giving rise to the observation  $f(t_i)$ . Let  $t_1, \dots, t_n \in (a, b)$  denote the observed input interval and  $\mathbb{E}[\epsilon_i] = 0$  (where we use  $\mathbb{E}[\cdot]$  to denote mathematical expectation). Here, the general statistical problem is to estimate (or denoise) the underlying signal  $f_0$  from the observations as accurately as possible. In the nonparametric setting, we refrain from making strong *a priori* assumptions about  $f_0$  that could lead to significant modeling bias, e.g. assuming a power law or a light-curve/spectral template fit. Mathematically, a nonparametric approach is defined through the deliberately weak assumption  $f_0 \in \mathcal{F}$  (i.e. the signal belongs to the function class  $\mathcal{F}$ ) where  $\mathcal{F}$  is *infinite-dimensional*. In other words, the assumed class of all possible signals  $\mathcal{F}$  cannot be spanned by a finite number of parameters. Contrast this to the assumption that the signal follows a  $p$ th degree power law, i.e.  $f_0 \in \mathcal{F}_{\text{PL}}$  where

$$(2) \quad \mathcal{F}_{\text{PL}} = \left\{ f_0 : f_0(t) = \beta_0 + \sum_{j=1}^p \beta_j t^j \right\},$$

a class that is spanned by  $p + 1$  parameters. Similarly, given a set of  $p$  spectral/light-curve templates  $b_1(t), \dots, b_p(t)$ , the usual template-fitting assumption is that  $f_0 \in \mathcal{F}_{\text{TEMP}}$  where

$$(3) \quad \mathcal{F}_{\text{TEMP}} = \left\{ f_0 : f_0(t) = \beta_0 + \sum_{j=1}^p \beta_j b_j((t - s)/v) \right\}$$

and  $s$  and  $v$  are horizontal shift and stretch hyperparameters, respectively. Both (2) and (3) represent very stringent assumptions about the underlying signal  $f_0$ . If the signal is anything other than exactly a power law in  $t$ —a highly unlikely occurrence—nontrivial statistical bias will arise by modeling it as such. Likewise, if a class of signals has a rich physical diversity (e.g., Type Ia supernova light curves; [135]) that is not sufficiently spanned by the library of templates used in modeling, then statistical biases will arise. Depending on the size of the imbalance between class diversity and the completeness of the template basis, the biases could be significant. Moreover, these biases are rarely tracked by uncertainty quantification. To be clear, this is not a uniform criticism of template-fitting. For example, templates are exceptionally powerful tools for object classification and redshift estimation (e.g., [59], [12]). Furthermore, much of our discussion in Section 5 centers around utilizing the flexible nonparametric nature of trend filtering to construct more complete spectral/light-curve template libraries for various observational objects and transient events.

Let  $\hat{f}_0$  be any statistical estimator for the signal  $f_0$ , derived from the noisy observations in (1). Further, let  $p_t(t)$  denote the probability density function (pdf) that specifies the sampling distribution of the inputs on the interval  $(a, b)$ , and let  $\sigma^2(t) = \text{Var}(\epsilon(t))$  denote the noise

level at input  $t$ . In order to assess the accuracy of the estimator it is common to consider the mean-squared prediction error (MSPE):

$$(4) \quad R(\hat{f}_0) = \mathbb{E}[(\hat{f}_0 - f)^2]$$

$$(5) \quad = \mathbb{E}[(\hat{f}_0 - f_0)^2] + \bar{\sigma}^2$$

$$(6) \quad = \int_a^b \left( \text{Bias}^2(\hat{f}_0(t)) + \text{Var}(\hat{f}_0(t)) \right) \cdot p_t(t) dt + \bar{\sigma}^2,$$

where

$$(7) \quad \text{Bias}(\hat{f}_0(t)) = \mathbb{E}[\hat{f}_0(t)] - f_0(t)$$

$$(8) \quad \text{Var}(\hat{f}_0(t)) = \mathbb{E} \left( \hat{f}_0(t) - \mathbb{E}[\hat{f}_0(t)] \right)^2$$

$$(9) \quad \bar{\sigma}^2 = \int_a^b \sigma^2(t) \cdot p_t(t) dt.$$

The equality in (6) is commonly referred to as the bias-variance decomposition. The first term is the squared bias of the estimator  $\hat{f}_0$  (integrated over the input interval) and intuitively measures how appropriate the chosen statistical estimator is for modeling the observed phenomenon. The second term is the variance of the estimator that measures how stable or sensitive the estimator is to the observed data. And the third term is the irreducible error—the minimum prediction error we cannot hope to improve upon. The bias-variance decomposition therefore illustrates that an optimal estimator is one that combines appropriate modeling assumptions (low bias) with high stability (low variance).

2.1.1. *Statistical optimality (minimax theory).* In this section, we briefly discuss a mathematical framework for evaluating the performance of statistical methods over nonparametric signal classes in order to demonstrate that the superiority of trend filtering is a highly general result. Ignoring the irreducible error, the problem of minimizing the MSPE of a statistical estimator can be equivalently stated as a minimization of the first term in (5)—the mean-squared estimation error (MSEE). In practice, low bias is attained by only making very weak assumptions about what the underlying signal may look like, e.g.  $f_0$  has  $k$  continuous derivatives. An ideal statistical estimator for estimating signals in such a class (call it  $\mathcal{F}$ ) may then be defined as

$$(10) \quad \inf_{\hat{f}_0} \left( \sup_{f_0 \in \mathcal{F}} \mathbb{E}[(\hat{f}_0 - f_0)^2] \right).$$

That is, we would like our statistical estimator to be the minimizer (infimum) of the worst-case (supremum) MSEE over the signal class  $\mathcal{F}$ . This is rarely a mathematically tractable problem for any practical signal class  $\mathcal{F}$ . A more tractable approach is to consider how the worst-case MSEE behaves as a function of the sample size  $n$ . A reasonable baseline metric for a statistical estimator is to require that it satisfies

$$(11) \quad \sup_{f_0 \in \mathcal{F}} \mathbb{E}[(\hat{f}_0 - f_0)^2] \rightarrow 0$$

as  $n \rightarrow \infty$ . That is, for any signal  $f_0 \in \mathcal{F}$ , when a large amount of data is available,  $\hat{f}_0$  gets arbitrarily close to the true signal. In any practical situation, this is not true for parametric models because the bias component of the bias-variance decomposition never vanishes. This, however, is a widely-held—perhaps, defining—property of nonparametric methods. Therefore, in order to distinguish optimality among nonparametric estimators, we require a stronger

metric. In particular, we study *how quickly* the worst-case error goes to zero as more data is observed. This is the core idea of a rich area of statistical literature called *minimax theory* [see, e.g., 128, 133, 127]. For many infinite-dimensional classes of signals, theoretical lower-bounds exist on the rate at which the MSEE of *any* statistical estimator can approach zero. Therefore, if a statistical estimator is shown to achieve that rate, it can be considered optimal for estimating that class of signals. Formally, letting  $g(n)$  be the rate at which the MSEE of the theoretically optimal estimator (10) goes to zero (a monotonically decreasing function in  $n$ ), we would like our estimator  $\hat{f}_0$  to satisfy

$$(12) \quad \sup_{f_0 \in \mathcal{F}} \mathbb{E}[(\hat{f}_0 - f_0)^2] = \mathcal{O}(g(n)),$$

where we use  $\mathcal{O}(\cdot)$  to denote big O notation. If this is shown to be true, we say the estimator *achieves the minimax rate over the signal class  $\mathcal{F}$* . Loosely speaking, we are stating that a *minimax optimal* estimator is an estimator that learns the signal from the data just as quickly as the theoretical gold standard estimator (10).

**2.1.2. Spatially heterogeneous signals.** Thus far we have only specified that the signal underlying most one-dimensional astronomical observations should be assumed to belong to a class  $\mathcal{F}$  that is infinite-dimensional (i.e. nonparametric). Further, in Section 2.1.1 we introduced the standard metric used to measure the performance of a statistical estimator over an infinite-dimensional class of signals. Recalling the discussion in the abstract and Section 1, trend filtering provides significant advances for estimating signals that exhibit varying degrees of smoothness across the input domain. We restate this definition below.

**Definition.** A *spatially heterogeneous* signal is a signal that exhibits varying degrees of smoothness in different regions of its input domain.

**Example.** A smooth light curve with abrupt transient events.

**Example.** An electromagnetic spectrum with smooth continuum radiation and sharp absorption/emission-line features.

To complement the above definition we may also loosely define a *spatially homogeneous signal* as a signal that is *either* smooth *or* wiggly<sup>1</sup> across its input domain, but not both. As “smoothness” can be quantified in various ways these definitions are intentionally mathematically imprecise. A class that is commonly considered in the statistical literature is the  $L_2$  Sobolev class:

$$(13) \quad \mathcal{F}_{2,k}(C_1) := \left\{ f_0 : \int_a^b f_0^{(k)}(t)^2 dt < C_1 \right\}, C_1 > 0, k \in \mathbb{N}.$$

That is, an  $L_2$  Sobolev class is a class of all signals such that the integral of the square (the “ $L_2$  norm”) of the  $k$ th derivative of each signal is less than some constant  $C_1$ . Statistical optimality in the sense of Section 2.1.1 for estimating signals in these classes (and some other closely related ones) is widely held among statistical methods in the classical toolkit; for example, kernel smoothers [60, 116], LOESS [38, 39], and smoothing splines [86]. However, a seminal result by [85] and [84] showed that a statistical estimator can be minimax optimal over signal

<sup>1</sup>This is, in fact, a technical term used in the statistical literature.

classes of the form (13) and still perform quite poorly on other signals. In particular, the authors showed that, when considering the broader  $L_1$  Sobolev class

$$(14) \quad \mathcal{F}_{1,k}(C_2) := \left\{ f_0 : \int_a^b |f_0^{(k)}(t)| dt < C_2 \right\}, C_2 > 0, k \in \mathbb{N},$$

all linear smoothers<sup>2</sup>—including kernels, LOESS, smoothing splines, Gaussian process regression, and many other methods—are strictly suboptimal. The key difference between these two types of classes is that  $L_2$  Sobolev classes are rich in spatially homogeneous signals but not spatially heterogeneous signals, while  $L_1$  Sobolev classes<sup>3</sup> are rich in both [see, e.g., 31].

The intuition of this result is that linear smoothers cannot optimally recover signals that exhibit varying degrees of smoothness across their input domain because they operate as if the signal possesses a fixed degree of smoothness. For example, this intuition is perhaps most clear when considering a kernel smoother with a fixed bandwidth. The result of [85] and [84] therefore implies that, in order to achieve statistical optimality for estimating spatially heterogeneous signals, a statistical estimator must be nonlinear (more specifically, it must be locally adaptive). [122] showed that trend filtering is minimax optimal for estimating signals in  $L_1$  Sobolev classes. Since  $L_2$  Sobolev classes are contained within  $L_1$  Sobolev classes, this result also guarantees that trend filtering is also minimax optimal for estimating signals in  $L_2$  Sobolev classes. Wavelets share this property, but require restrictive assumptions on the sampling of the data [30].

*How large is this performance gap?* The collective results of [85], [84], and [122] reveal that the performance gap between trend filtering and linear smoothers when estimating spatially heterogeneous signals is significant. For example, when  $k = 0$ , the minimax rate over  $L_1$  Sobolev classes (which trend filtering achieves) is  $n^{-2/3}$ , but linear smoothers cannot achieve better than  $n^{-1/2}$ . To put this in perspective, this result says that the trend filtering estimator, training on  $n$  data points, learns these signals with varying smoothness as quickly as a linear smoother training on  $n^{4/3}$  data points. As we demonstrate in the next section, this gap in theoretical optimality has clear practical consequences.

In order to minimize the pervasion of technical statistical jargon throughout the paper, henceforth we simply refer to a statistical estimator that achieves the minimax rate over  $L_2$  Sobolev classes as *statistically optimal for estimating spatially homogeneous signals*, and we refer to a statistical estimator that achieves the minimax rate over  $L_1$  Sobolev classes as *statistically optimal for estimating spatially heterogeneous signals*. As previously mentioned, the latter implies the former, but not vice versa.

**2.2. Empirical comparison.** In this section we analyze noisy observations of a simulated spatially heterogeneous signal in order to compare the empirical performance of trend filtering and several classical statistical methods—namely, LOESS, smoothing splines, and Gaussian process regression. The mock observations are simulated on an unequally-spaced grid  $t_1, \dots, t_n \sim \text{Unif}(0, 1)$  according to the data generating process

$$(15) \quad f(t_i) = f_0(t_i) + \epsilon_i$$

---

<sup>2</sup>A *linear smoother* is a statistical estimator that is a linear combination of the observed data. Many popular statistical estimators, although often motivated from seemingly disparate premises, can be shown to fall under this definition. See, e.g., [133] for more details.

<sup>3</sup>The  $L_1$  Sobolev class is often generalized to a nearly equivalent but slightly larger class—namely, signals with derivatives of bounded total variation. See [122] for the generalized definition.

with

$$(16) \quad f_0(t_i) = 6 \sum_{k=1}^3 (t_i - 0.5)^k + 2.5 \sum_{j=1}^4 (-1)^j \phi_j(t_i),$$

where  $\phi_j(t)$ ,  $j = 1, \dots, 4$  are compactly-supported radial basis functions distributed throughout the input space and  $\epsilon_i \sim N(0, 0.125^2)$ . We therefore construct the signal  $f_0$  to have a smoothly varying global trend with four sharp localized features—two dips and two spikes. The signal and noisy observations are shown in the top panel of Figure 1.

In order to facilitate the comparison of methods we utilize a metric for the total statistical complexity (i.e. total wiggleness) of an estimator known as the *effective degrees of freedom* [see, e.g., 123]. Formally, the effective degrees of freedom of an estimator  $\hat{f}_0$  is defined as

$$(17) \quad \text{df}(\hat{f}_0) = \bar{\sigma}^{-2} \sum_{i=1}^n \text{Cov}(\hat{f}_0(t_i), f(t_i))$$

where  $\bar{\sigma}^2$  is defined in (9). In Figure 1 we fix all estimators to have 55 effective degrees of freedom. This exercise provides insight into how each estimator relatively distributes its complexity across the input domain. In the second panel we see that the trend filtering estimate has sufficiently recovered the underlying signal, including both the smoothness of the global trend and the abruptness of the localized features. All three of the linear smoothers, on the other hand, severely oversmooth the localized peaks and dips. Gaussian process regression also exhibits some undesirable oscillatory features that do not correspond to any real trend in the signal. In order to better recover the localized features the linear smoothers require a more complex fit, i.e. smaller LOESS kernel bandwidth, smaller smoothing spline penalization, and smaller Gaussian process noise-signal variance. In Figure 2 we show the same comparison, but we grant the linear smoothers more complexity. Specifically, in order to recover the sharp features comparably with the trend filtering estimator with 55 effective degrees of freedom, the linear smoothers require 192 effective degrees of freedom—approximately 3.5 times the complexity. As a result, although they now adequately recover the peaks and dips, each linear smoother severely overfits the data in the other regions of the input domain, resulting in many spurious fluctuations.

As discussed in Section 2.1.2, the suboptimality of LOESS, smoothing splines, and Gaussian process regression illustrated in this example is an inherent limitation of the broad *linear smoother* family of statistical estimators. Linear smoothers are adequate tools for estimating signals that exhibit approximately the same degree of smoothness throughout their input domain. For example, in Section 4.3 we use LOESS to nonparametrically reconstruct the smooth background continuum radiation in quasar spectra. However, when a signal is expected to exhibit varying degrees of smoothness across its domain, a locally-adaptive statistical estimator is needed.

**3. Trend filtering.** Trend filtering, in its original form, was independently proposed in the computer vision literature [114] and the applied mathematics literature [63], and has recently been further developed in the statistical and machine learning literature, most notably with [124, 122, 130, 107]. This work is in no way related to the work of [67], which goes by a similar name. At a high level, trend filtering is closely related to two familiar nonparametric regression methods: variable-knot regression splines and smoothing splines. We elaborate on these relationships below.

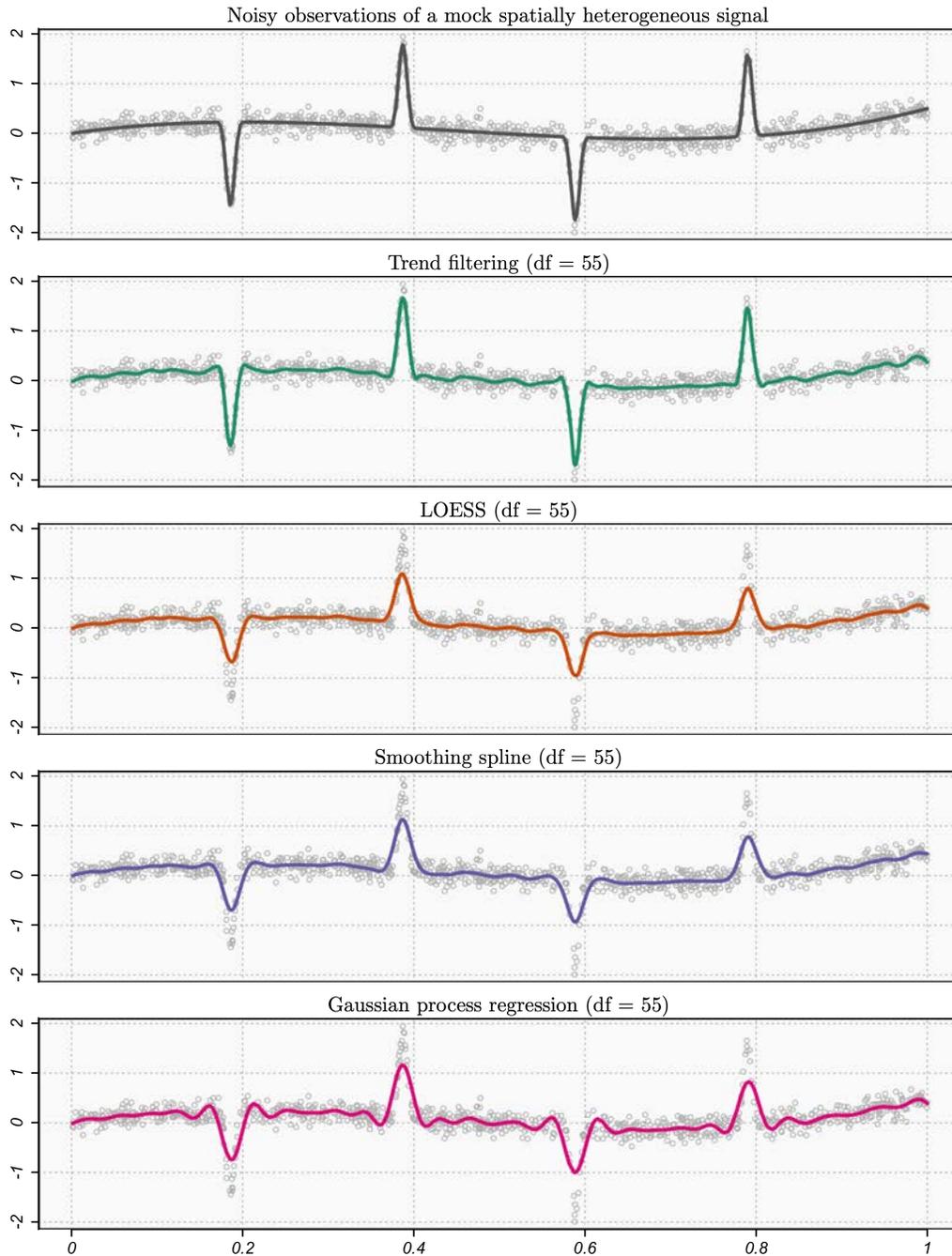


FIG 1. Comparison of statistical methods on data simulated from a spatially heterogeneous signal. Each statistical estimator is fixed to have 55 effective degrees of freedom in order to facilitate a direct comparison. The trend filtering estimator is able to sufficiently distribute its effective degrees of freedom such that it simultaneously recovers the smoothness of the global trend, as well as the abrupt localized features. The LOESS, smoothing spline, and Gaussian process regression each estimates the smooth global trend reasonably well here, but significantly oversmooths the sharp peaks and dips. Here, we utilize quadratic trend filtering (see Section 3.2).

## Trend Filtering in Astronomy

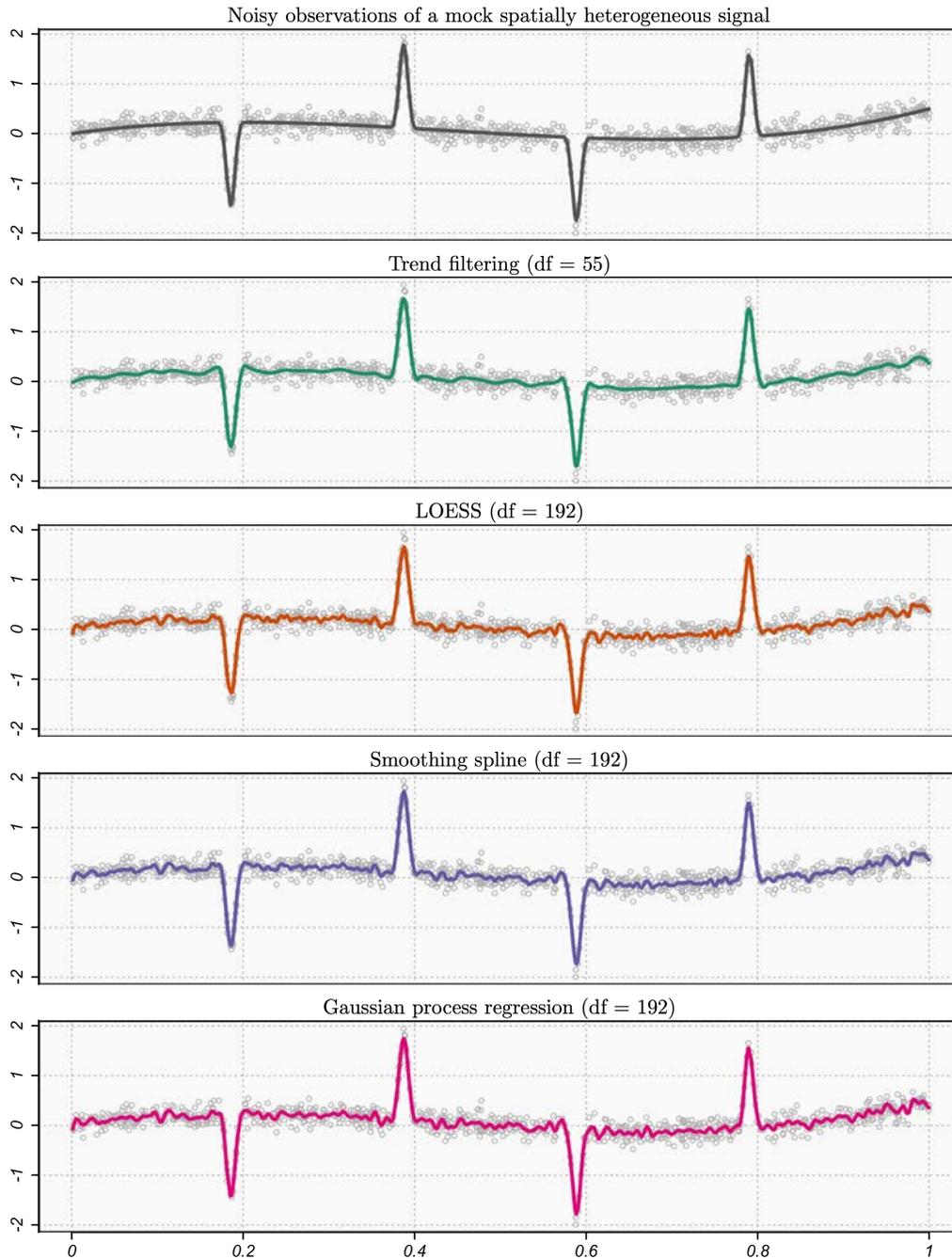


FIG 2. (Continued): Comparison of statistical methods on data simulated from a spatially heterogeneous signal. Here, each of the linear smoothers (i.e. the LOESS, smoothing spline, and Gaussian process regression) is fixed at 192 effective degrees of freedom—the complexity necessary for each estimator to recover the sharp localized features approximately as well as the trend filtering estimator with 55 effective degrees of freedom. While the linear smoothers now estimate the four abrupt features well, each severely overfits the data in the other regions of the input domain.

3.1. *Closely-related methods.* Splines have long played a central role in estimating complex signals (see, e.g., [25] and [129] for general references). Formally, a  $k$ th order spline is a piecewise polynomial (i.e. piecewise power law) of degree  $k$  that is continuous and has  $k - 1$  continuous derivatives at the knots. As their names suggest, variable-knot regression splines and smoothing splines center around fitting splines to observational data. Recall from (1) the observational data generating process (DGP)

$$(18) \quad f(t_i) = f_0(t_i) + \epsilon_i, t_1, \dots, t_n \in (a, b),$$

where  $f(t_i)$  is a noisy measurement of the signal  $f_0(t_i)$ , and  $\mathbb{E}[\epsilon_i] = 0$ . Given a set of knots  $\kappa_1, \dots, \kappa_p \in (a, b)$ , the space of all  $k$ th order splines on the interval  $(a, b)$  with knots at  $\kappa_1, \dots, \kappa_p$  can be parametrized via a basis representation

$$(19) \quad m(t) = \sum_j \beta_j \eta_j(t),$$

where  $\{\eta_j\}$  is typically the truncated power basis or B-spline basis. A suitable estimator for the signal  $f_0$  may then be

$$(20) \quad \hat{f}_0(t) = \sum_j \hat{\beta}_j \eta_j(t),$$

where the  $\hat{\beta}_j$  are the ordinary least-squares (OLS) estimates of the basis coefficients. This is called a *regression spline*. The question of course remains where to place the knots.

3.1.1. *Variable-knot regression splines.* The variable-knot (or free-knot) regression spline approach is to consider all regression spline estimators with knots at a subset of the observed inputs, i.e.  $\{\kappa_1, \dots, \kappa_p\} \subset \{t_1, \dots, t_n\}$  for all possible  $p$ . Formally, the variable-knot regression spline estimator is the solution to the following constrained least-squares minimization problem:

$$(21) \quad \begin{aligned} \min_{\{\beta_j\}} \quad & \sum_{i=1}^n \left( f(t_i) - \sum_j \beta_j \eta_j(t_i) \right)^2 \\ \text{s.t.} \quad & \sum_{j \geq k+2} \mathbb{1}\{\beta_j \neq 0\} = p \\ & p \geq 0 \end{aligned}$$

where  $p \geq 0$  is the number of knots in the spline and  $\mathbb{1}(\cdot)$  is the indicator function satisfying

$$(22) \quad \mathbb{1}\{\beta_j \neq 0\} = \begin{cases} 1 & \beta_j \neq 0, \\ 0 & \beta_j = 0. \end{cases}$$

Furthermore, note that the equality constraint on the basis coefficients excludes those of the “first”  $k + 1$  basis functions that span the space of global polynomials and only counts the number of active basis functions that produce knots. The variable-knot regression spline optimization is therefore a problem of finding the *best subset* of knots for the regression spline estimator. Due to the sparsity of the coefficient constraint, the variable-knot regression spline estimator allows for highly locally-adaptive behavior for estimating signals that exhibit varying degrees of smoothness. However, the problem itself cannot be solved in polynomial time, requiring an exhaustive combinatorial search over all  $\sim 2^n$  feasible models. It is common to utilize stepwise procedures based on iterative addition and deletion of knots in the active set,

but these partial searches over the feasible set inherently provide no guarantee of finding the optimal global solution to (21).

In order to make the connection to trend filtering more explicit it is helpful to reformulate the constrained minimization (21) into the following penalized unconstrained minimization problem:

$$(23) \quad \min_{\{\beta_j\}} \sum_{i=1}^n \left( f(t_i) - \sum_j \beta_j \eta_j(t_i) \right)^2 + \gamma \sum_{j \geq k+2} \mathbb{1}(\beta_j \neq 0),$$

where  $\gamma > 0$  is a hyperparameter that determines the number of knots in the spline and the sum of indicator functions serves as a smoothness “penalty” on the ordinary least-squares minimization. Penalized regression is a popular area of statistical methodology [see, e.g., 56], in which the cost functional (i.e. the quantity to be minimized) quantifies a tradeoff between the training error of the estimator (here, the squared residuals) and the statistical complexity of the estimator (here, the number of knots in the spline). In particular, (23) is known as an  $\ell_0$ -penalized least-squares regression because of the penalty’s connection to the mathematical  $\ell_0$  vector quasi-norm.

3.1.2. *Smoothing splines.* Smoothing splines counteract the computational issue faced by variable-knot regression splines by simply placing knots at all of the observed inputs  $t_1, \dots, t_n$  and regularizing the smoothness of the fitted spline. For example, letting  $\mathcal{G}$  be the space of all cubic natural splines with knots at  $t_1, \dots, t_n$ , the cubic smoothing spline is the solution to the optimization problem

$$(24) \quad \min_{m \in \mathcal{G}} \sum_{i=1}^n (f(t_i) - m(t_i))^2 + \gamma \int_a^b (m''(t))^2 dt,$$

where  $m''$  is the second derivative of  $m$  and  $\gamma > 0$  tunes the amount of regularization. Letting  $\eta_1, \dots, \eta_n$  be a basis for cubic natural splines with knots at the observed inputs, (24) can be equivalently stated as a minimization over the basis coefficients:

$$(25) \quad \min_{\{\beta_j\}} \sum_{i=1}^n \left( f(t_i) - \sum_j \beta_j \eta_j(t_i) \right)^2 + \gamma \sum_{j,k=1}^n \beta_j \beta_k \omega_{jk}$$

where

$$(26) \quad \omega_{jk} = \int_a^b \eta_j''(t) \eta_k''(t) dt.$$

The cost functional (25) is differentiable and leads to a linear system with a special sparse structure (i.e. bandedness), which yields a solution that can both be found in closed-form and computed very quickly—in  $O(n)$  elementary operations. This particular choice of cost functional, however, produces an estimator that is a linear combination of the observations—a *linear smoother*. Therefore, as discussed and demonstrated in Section 2, smoothing splines are suboptimal for estimating spatially heterogeneous signals. Equation (25) is known as an  $\ell_2$ -penalized least-squares regression because of the penalty’s connection to the mathematical  $\ell_2$  vector-norm.

3.2. *Definition.* Trend filtering can be viewed as a blending of the strengths of variable-knot regression splines (local adaptivity and interpretability) and the strengths of smoothing splines (simplicity and speed). Mathematically, this is achieved by choosing an appropriate

set of basis functions and penalizing the least-squares problem with an  $\ell_1$  norm on the basis coefficients (sum of absolute values), instead of the  $\ell_0$  norm of variable-knot regression splines (sum of indicator functions) or the  $\ell_2$  norm of smoothing splines (sum of squares).

This section is primarily summarized from [122] and [131]. Let the inputs be ordered with respect to the index, i.e.  $t_1 < \dots < t_n$ . For the sake of simplicity, we consider the case when the inputs  $t_1, \dots, t_n \in (a, b)$  are equally spaced with  $\Delta t = t_{i+1} - t_i$ . See the aforementioned papers for the generalized definition of trend filtering to unequally spaced inputs.

For any given integer  $k \geq 0$ , the  $k$ th order trend filtering estimate is a piecewise polynomial of degree  $k$  with knots *automatically selected* at a sparse subset of the observed inputs  $t_1, \dots, t_n$ . In Figure 3, we provide an example of a trend-filtered data set for orders  $k = 0, 1, 2$ , and 3. Specifically, the panels of the figure respectively display piecewise constant, piecewise linear, piecewise quadratic, and piecewise cubic fits to the data with the automatically-selected knots indicated by the tick marks on the horizontal axes. Constant trend filtering is equivalent to total variation denoising [109], as well as special forms of the fused lasso of [121] and the variable fusion estimator of [70]. Linear trend filtering was independently proposed by [114] and [63]. Higher-order polynomial trend filtering ( $k \geq 2$ ) was developed by [124] and [122]. In the Figure 3 example, the quadratic and cubic trend filtering estimates are nearly visually indistinguishable, and this is true in general. Although, as we see here, trend filtering estimates of different orders typically select different sets of knots.

Like the spline methods discussed in Section 3.1, for any order  $k \geq 0$ , the trend filtering estimator has a basis representation

$$(27) \quad m(t) = \sum_{j=1}^n \beta_j h_j(t),$$

but, here,  $\{h_1, \dots, h_n\}$  is the *falling factorial* basis

$$(28) \quad h_j(t) = \begin{cases} \prod_{i=1}^{j-1} (t - t_i) & j \leq k + 1, \\ \prod_{i=1}^{j-1} (t - t_{j-k-1+i}) \cdot \mathbb{1}\{t \geq t_{j-1}\} & j \geq k + 2. \end{cases}$$

Like the truncated power basis, the first  $k + 1$  basis functions span the space of global  $k$ th order polynomials and the rest of the basis adds the piecewise polynomial structure. However, the knot-producing basis functions of the falling factorial basis  $h_j, j \geq k + 2$  have small discontinuities in their  $j$ th order derivatives at the knots for all  $j = 1, \dots, k - 1$ , and therefore for orders  $k \geq 2$  the trend filtering estimate is *close to*, but not quite a spline. The discontinuities are small enough, however, that the trend filtering estimate defined through the falling factorial basis representation is visually indistinguishable from the analogous spline produced by the truncated power basis (see [122] and [131]). The advantage of utilizing the falling factorial basis in this context instead of the truncated power basis (or the B-spline basis) comes in the form of significant computational speedups, as we detail below.

Analogous to the continuous smoothing spline problem (24), we let  $\mathcal{H}_k$  be the space of all functions spanned by the  $k$ th order falling factorial basis, and pose the trend filtering problem as a least-squares minimization with a derivative-based penalty on the fitted function. In particular, the  $k$ th order trend filtering estimator is the solution to the problem

$$(29) \quad \min_{m \in \mathcal{H}_k} \sum_{i=1}^n (f(t_i) - m(t_i))^2 + \gamma \cdot \text{TV}(m^{(k)}),$$

where  $m^{(k)}$  is the  $k$ th derivative of  $m$ ,  $\text{TV}(m^{(k)})$  is the *total variation* of  $m^{(k)}$ , and  $\gamma > 0$  is the model hyperparameter that controls the smoothness of the fit. When  $m^{(k)}$  is differentiable everywhere in its domain, the penalty term simplifies to

$$(30) \quad \text{TV}(m^{(k)}) = \int_a^b |m^{(k+1)}(t)| dt.$$

## Trend Filtering in Astronomy

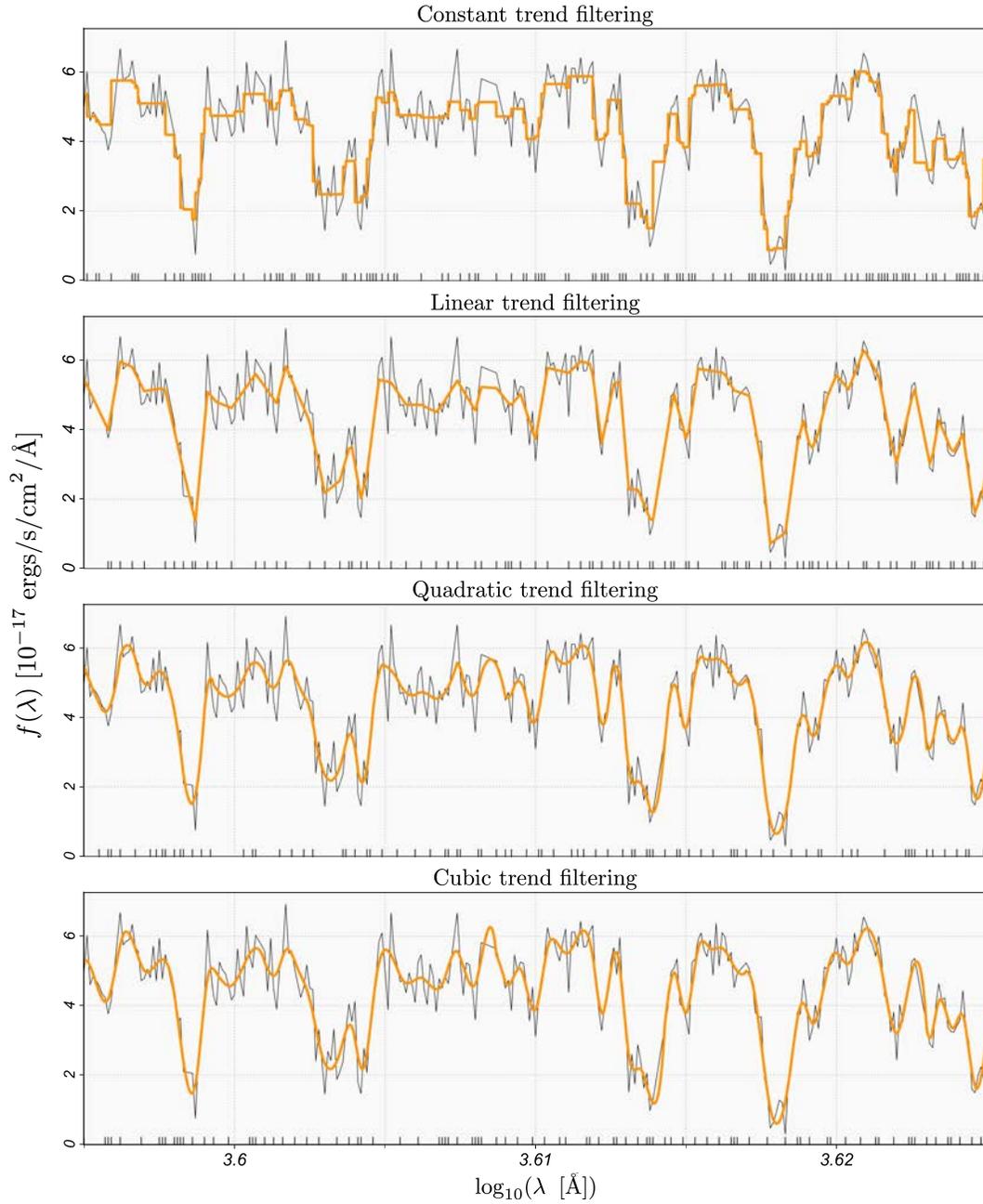


FIG 3. Piecewise polynomials with adaptively-chosen knots produced by trend filtering. From top to bottom, we show trend filtering estimates of orders  $k = 0, 1, 2$  and  $3$ , which take the form of piecewise constant, piecewise linear, piecewise quadratic, and piecewise cubic polynomials, respectively. The adaptively-chosen knots of each piecewise polynomial are indicated by the tick marks along the horizontal axes. The constant trend filtering estimate is discontinuous at the knots, but we interpolate here for visual purposes. The data set is taken from the Lyman- $\alpha$  forest of a mock quasar spectrum ([8]), sampled in logarithmic-angstrom space. We study this phenomenon in detail in Section 4.

Avoiding the technical generalized definition of total variation [see, e.g., 122], we can simply think of  $\text{TV}(\cdot)$  as a generalized  $L_1$  norm<sup>4</sup> for our piecewise polynomials that possess small discontinuities in the derivatives. Again referring back to the smoothing spline problem (24), definitions (29) and (30) reveal that trend filtering can be thought of as an  $L_1$  analog of the ( $L_2$ -penalized) smoothing spline problem. Moreover, note that unlike smoothing splines, trend filtering can produce piecewise polynomials of all orders  $k \geq 0$ .

Replacing  $m$  with its basis representation, i.e.  $m(t) = \sum_j \beta_j h_j(t)$ , yields the equivalent finite-dimensional minimization problem<sup>5</sup>:

$$(31) \quad \min_{\{\beta_j\}} \sum_{i=1}^n \left( f(t_i) - \sum_{j=1}^n \beta_j h_j(t_i) \right)^2 + \gamma \cdot k! \cdot \Delta t^k \sum_{j=k+2}^n |\beta_j|.$$

The terms  $k!$  and  $\Delta t^k$  are constants and can therefore be ignored by absorbing them into the hyperparameter  $\gamma$ . Visual inspection of (31) reveals that trend filtering is also analogous to the variable-knot regression spline problem (21)—namely, by replacing the  $\ell_0$  norm on the basis coefficients with an  $\ell_1$  norm. The advantage here is that the problem is now strictly convex and can be efficiently solved by various convex optimization algorithms. Furthermore, the  $\ell_1$  penalty still yields a sparse solution (i.e. many  $\beta_j = 0$ ), which provides the automatic knot-selection property. Letting  $\hat{\beta}_1, \dots, \hat{\beta}_n$  denote the solution to (31) for a particular choice of  $\gamma > 0$ , the trend filtering estimate is then given by

$$(32) \quad \hat{f}_0(t; \gamma) = \sum_{j=1}^n \hat{\beta}_j h_j(t),$$

with the automatically-selected knots corresponding to the basis functions with  $\hat{\beta}_j \neq 0$ ,  $j \geq k + 1$ .

The advantage of utilizing the falling factorial basis is found by reparametrizing the problem (31) into an optimization over the fitted values  $m(t_1), \dots, m(t_n)$ . The problem then reduces to

$$(33) \quad \min_{\{m(t_i)\}} \sum_{i=1}^n (f(t_i) - m(t_i))^2 + \gamma \sum_{i=1}^{n-k-1} |\Delta^{(k+1)} m(t_i)| \cdot \Delta t$$

where  $\Delta^{(k+1)} m(t_i)$  can be viewed as a discrete approximation of the  $(k + 1)$ st derivative of  $m$  at  $t_i$ . For  $k = 0$  the discrete derivatives are

$$(34) \quad \Delta^{(1)} m(t_i) = \frac{m(t_{i+1}) - m(t_i)}{\Delta t},$$

and then can be defined recursively for  $k \geq 1$ :

$$(35) \quad \Delta^{(k+1)} m(t_i) = \frac{\Delta^{(k)} m(t_{i+1}) - \Delta^{(k)} m(t_i)}{\Delta t}.$$

The penalty term in (33) can be viewed as a Riemann-like discrete approximation of the integral in (30). Because of the choice of basis, the problem has reduced to a simple generalized lasso problem [124, 6] with an identity predictor matrix and a banded<sup>6</sup> penalty matrix. This

<sup>4</sup>We use the upper-case notation  $L_p$ ,  $p = 1, 2$  for the  $p$ -norm of a continuous function, and  $\ell_p$ ,  $p = 0, 1, 2$  for the  $p$ -norm of a vector.

<sup>5</sup>This may be recognized as a lasso regression [120], with the features being the falling factorial basis functions.

<sup>6</sup>A banded matrix only contains nonzero elements in the main diagonal and zero or more diagonals on either side.

## Trend Filtering in Astronomy

	Method	Computational Complexity	Hyperparameters to estimate
Locally-adaptive	Wavelets	$\mathcal{O}(n)$	1
	<b>Trend filtering</b>	$\mathcal{O}(n^{1.5})$	<b>1</b>
	Variable-knot regression splines	$\mathcal{O}(n \cdot \binom{n}{p})$	1
Non-adaptive	Uniform-knot regression splines	$\mathcal{O}(n)$	1
	Smoothing splines	$\mathcal{O}(n)$	1
	Kernel smoothers	$\mathcal{O}(n^2)$	1
	LOESS	$\mathcal{O}(n^2)$	1
	Gaussian process regression	$\mathcal{O}(n^3)$	3+

TABLE 1

Comparison of computational costs associated with popular one-dimensional nonparametric regression methods. The computational complexity column states the dependence on the sample size  $n$  of the number of elementary operations necessary to obtain the fitted values of each estimator (i.e. the estimator evaluated at the observed inputs). For trend filtering, the  $\mathcal{O}(n^{1.5})$  complexity represents the worst-case complexity of the [107] convex optimization algorithm. In most practical settings the actual complexity of this algorithm is close to  $\mathcal{O}(n)$ . Variable-knot regression splines require a (nonconvex) exhaustive combinatorial search over the set of possible knots and the complexity therefore includes a binomial coefficient term  $\binom{n}{p} = n!/(n!(n-p)!)$ , where  $p$  is the number of knots in the spline. The remaining methods are explicitly solvable and the stated complexity represents the cost of an exact calculation. The  $\mathcal{O}(n)$  complexity of wavelets relies on restrictive sampling assumptions (e.g., equally-spaced inputs, sample size equal to a power of two). The stated computational complexity of all methods represents the cost of a single model fit and does not include the cost of hyperparameter tuning. Gaussian process regression suffers from the most additional overhead in this regard because of the (often) large number of hyperparameters used to parametrize the covariance function (e.g., shape, range, marginal variance, noise variance). Each of the non-adaptive methods (linear smoothers) can be made to be locally adaptive (e.g., by locally varying the hyperparameters of the model), but at the expense of greatly increasing the dimensionality of the hyperparameter space to be searched over.

special structure allows the solution to be found in nearly linear time, i.e.  $\mathcal{O}(n)$  elementary operations. In this work we utilize the specialized alternating direction method of multipliers (ADMM) algorithm of [107]. This algorithm has a linear complexity per iteration, so the overall complexity is  $\mathcal{O}(nr)$  where  $r$  is the number of iterations necessary to converge to the solution. In the worst case scenario  $r \sim n^{1/2}$ , so the worst-case overall complexity is  $\mathcal{O}(n^{1.5})$ . The practical computational speed further illustrates the value of trend filtering to astronomy, as it is readily compatible with the large-scale analysis of one-dimensional data sets that has become increasingly ubiquitous in large sky surveys. We show a comparison in Table 1 of the computational costs associated with trend filtering and other popular one-dimensional nonparametric methods.

Given the trend filtering fitted values obtained by the optimization (33) the full continuous-time representation of the trend filtering estimate follows by inverting the parametrization back to the basis function coefficients and plugging them into the basis representation (32).

3.3. *Extension to heteroskedastic weighting.* Thus far we have considered the simple case where the observations are treated as equally-weighted in the cost functional (33). Recall from (18) the observational data generating process and define  $\sigma_i^2 = \text{Var}(\epsilon_i)$  to be the noise level—the (typically heteroskedastic) uncertainty in the measurements that arises from instrumental errors and removal of systematic effects. When estimates for  $\sigma_i^2$ ,  $i = 1, \dots, n$  accompany the observations, as they often do, they can be used to weight the observations to yield a more efficient statistical estimator (i.e. smaller mean-squared error). The error-

Language	Recommended implementation
R	<a href="https://github.com/glmgen">github.com/glmgen</a>
C	<a href="https://github.com/glmgen">github.com/glmgen</a>
Python	<a href="https://cvxpy.org">cvxpy.org</a>
Matlab	<a href="http://stanford.edu/~boyd/l1_tf">http://stanford.edu/~boyd/l1_tf</a>
Julia	<a href="https://github.com/JuliaStats/Lasso.jl">github.com/JuliaStats/Lasso.jl</a>

TABLE 2

Recommended implementations for trend filtering in various programming languages. See Section 3.4 for details. We provide additional code at [github.com/capolitsch/trend-filtering](https://github.com/capolitsch/trend-filtering) for selecting the hyperparameter via minimization of Stein’s unbiased risk estimate (see Section 3.5) and various bootstrap methods for uncertainty quantification (see Section 3.6). Our implementations are built on top of the `glmgen` R package of [5].

weighted trend filtering estimator is the solution to the following minimization problem:

$$(36) \quad \min_{\{m(t_i)\}} \sum_{i=1}^n (f(t_i) - m(t_i))^2 w_i + \gamma \sum_{i=1}^{n-k-1} |\Delta^{(k+1)} m(t_i)| \cdot \Delta t,$$

where the optimal choice of weights is  $w_i = \sigma_i^{-2}$ ,  $i = 1, \dots, n$ . Much of the publically available software for trend filtering allows for a heteroskedastic weighting scheme (see Section 3.4).

3.4. *Software.* Trend filtering software is available online across various platforms. For the ADMM implementation of [107] that we utilize in this work, implementations are available in R and C [5], as well as Julia [66]. Matlab and Python implementations are available for the primal-dual interior point method of [63], but only for equally-weighted linear trend filtering [65, 27]. We provide links to our recommended implementations in Table 2. Note that in all software implementations the trend filtering hyperparameter is called  $\lambda$  instead of  $\gamma$ , which we use here to avoid ambiguity with the notation for wavelength in our spectroscopic examples.

3.5. *Choosing the hyperparameter.* The choice of the piecewise polynomial order  $k$  generally has minimal effect on the performance of the trend filtering estimator in terms of mean-squared error and therefore can be treated as an *a priori* aesthetic choice based on how much smoothness is desired or believed to be present. For example, we use  $k = 2$  (quadratic trend filtering) throughout Sections 4 and 5 so the fitted curves are smooth, i.e. differentiable everywhere.

Given the choice of  $k$ , the hyperparameter  $\gamma > 0$  is used to tune the complexity (i.e. the wiggleness) of the trend filtering estimate by weighting the tradeoff between the complexity of the estimate and the size of the squared residuals. Obtaining an accurate estimate is therefore intrinsically tied to finding an optimal choice of  $\gamma$ . The selection of  $\gamma$  is typically done by minimizing an estimate of the mean-squared prediction error (MSPE) of the trend filtering estimator. Here, there are two different notions of error to consider, namely, *fixed-input* error and *random-input* error. As the names suggest, the distinction between which type of error to consider is made based on how the inputs are sampled. As a general rule-of-thumb, we recommend optimizing with respect to fixed-input error when the inputs are regularly-sampled and optimizing with respect to random-input error on irregularly-sampled data.

Recall the DGP stated in (18) and let it be denoted by  $Q$  so that  $\mathbb{E}_Q[\cdot]$  is the mathematical expectation with respect to the randomness of the DGP. Further, let  $\sigma_i^2 = \text{Var}(\epsilon_i)$ . The fixed-

input MSPE is given by

$$(37) \quad R(\gamma) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_Q \left[ (f(t_i) - \hat{f}_0(t_i; \gamma))^2 \mid t_1, \dots, t_n \right]$$

$$(38) \quad = \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E}_Q \left[ (f_0(t_i) - \hat{f}_0(t_i; \gamma))^2 \mid t_1, \dots, t_n \right] + \sigma_i^2 \right)$$

and the random-input MSPE is given by

$$(39) \quad \tilde{R}(\gamma) = \mathbb{E}_Q \left[ (f(t) - \hat{f}_0(t; \gamma))^2 \right],$$

where, in the latter,  $t$  is considered to be a random component of the DGP with a marginal probability density  $p_t(t)$  supported on the observed input interval. In each case, the theoretically optimal choice of  $\gamma$  is defined as the minimizer of the respective choice of error. Empirically, we estimate the theoretically optimal choice of  $\gamma$  by minimizing an estimate of (37) or (39). For fixed-input error we recommend Stein's unbiased risk estimate (SURE; [115, 33]) and for random-input error we recommend  $K$ -fold cross validation with  $K = 10$ . We elaborate on SURE here and refer the reader to [132] for  $K$ -fold cross validation.

The SURE formula provides an unbiased estimate of the fixed-input MSPE of a statistical estimator:

$$(40) \quad \hat{R}_0(\gamma) = \frac{1}{n} \sum_{i=1}^n (f(t_i) - \hat{f}_0(t_i; \gamma))^2 + \frac{2\bar{\sigma}^2 \text{df}(\hat{f}_0)}{n},$$

where  $\bar{\sigma}^2 = n^{-1} \sum_{i=1}^n \sigma_i^2$  and  $\text{df}(\hat{f}_0)$  is defined in (17). A formula for the effective degrees of freedom of the trend filtering estimator is available via the generalized lasso results of [125]; namely,

$$(41) \quad \text{df}(\hat{f}_0) = \mathbb{E}[\text{number of knots in } \hat{f}_0] + k + 1.$$

We then obtain our hyperparameter estimate  $\hat{\gamma}$  by minimizing the following plug-in estimate for (40):

$$(42) \quad \hat{R}(\gamma) = \frac{1}{n} \sum_{i=1}^n (f(t_i) - \hat{f}_0(t_i; \gamma))^2 + \frac{2\hat{\sigma}^2 \hat{\text{df}}(\hat{f}_0)}{n},$$

where  $\hat{\text{df}}$  is the estimate for the effective degrees of freedom that is obtained by replacing the expectation in (41) with the observed number of knots, and  $\hat{\sigma}^2$  is an estimate of  $\bar{\sigma}^2$ . If a reliable estimate of  $\bar{\sigma}^2$  is not available *a priori*, a data-driven estimate can be constructed (see, e.g., [133]). We provide R code on the corresponding author's GitHub page<sup>7</sup> for implementing SURE with trend filtering. The code is built on top of the `glmgen` R package of [5], which already includes an implementation of  $K$ -fold cross validation.

Because of the existence of the degrees of freedom expression (41), trend filtering is also compatible with reduced chi-squared model assessment and comparison procedures under a Gaussian noise assumption [94, 19].

### 3.6. Uncertainty quantification.

<sup>7</sup><https://github.com/capolitsch/trend-filtering>

3.6.1. *Frequentist.* Frequentist uncertainty quantification for trend filtering follows by studying the sampling distribution of the estimator that arises from the randomness of the observational data generating process (DGP). In particular, most of the uncertainty in the estimates is captured by studying the variability of the estimator with respect to the DGP. We advise three different bootstrap methods [32] for estimating the variability of the trend filtering estimator, with each method corresponding to a distinct analysis setting. Here, we emphasize the terminology *variability*—as opposed to the variance of the trend filtering estimator—since, by construction, as a nonlinear function of the observed data, the trend filtering estimator has a non-Gaussian sampling distribution even when the observational noise is Gaussian. For that reason, each of our recommended bootstrap approaches is based on computing sample quantiles (instead of pairing standard errors with Gaussian quantiles).

We restate the assumed DGP here for clarity:

$$(43) \quad f(t_i) = f_0(t_i) + \epsilon_i, t_1, \dots, t_n \in (a, b)$$

where  $\mathbb{E}[\epsilon_i] = 0$ . We make the further assumption that the errors  $\epsilon_1, \dots, \epsilon_n$  are independent<sup>8</sup>. The three distinct settings we consider are:

- S1.** The inputs are irregularly sampled
- S2.** The inputs are regularly sampled and the noise distribution is known
- S3.** The inputs are regularly sampled and the noise distribution is unknown

The corresponding bootstrap methods are detailed in Algorithm 1 [nonparametric bootstrap; 32], Algorithm 2 [parametric bootstrap; 34], and Algorithm 3 [wild bootstrap; 136, 73, 75], respectively. We provide code on our GitHub page for implementing each of these methods with trend filtering.

---

**Algorithm 1** Nonparametric bootstrap for random-input uncertainty quantification

---

**Require:** Training Data  $(t_1, f(t_1)), \dots, (t_n, f(t_n))$ , hyperparameters  $\gamma$  and  $k$ , prediction input grid  $t'_1, \dots, t'_m$

- 1: **for all**  $b$  in  $1 : B$  **do**
- 2: Define a bootstrap sample of size  $n$  by resampling the observed pairs with replacement:

$$(t_1^*, f_b^*(t_1^*)), \dots, (t_n^*, f_b^*(t_n^*))$$

- 3: Let  $\hat{f}_b^*(t'_1), \dots, \hat{f}_b^*(t'_m)$  denote the trend filtering estimate fit on the bootstrap sample and evaluated on the prediction grid  $t'_1, \dots, t'_m$
- 4: **end for**

**Output:** The full trend filtering bootstrap ensemble

$$\{\hat{f}_b^*(t'_i)\}_{i=1, \dots, m}^{b=1, \dots, B}$$


---

Given the full trend filtering bootstrap ensemble provided by the relevant bootstrap algorithm, for any  $\alpha \in (0, 1)$ , a  $(1 - \alpha) \cdot 100\%$  quantile-based pointwise variability band is given by

$$(44) \quad V_{1-\alpha}(t'_i) = \left( \hat{f}_{\alpha/2}^*(t'_i), \hat{f}_{1-\alpha/2}^*(t'_i) \right), i = 1, \dots, m$$

where

$$(45) \quad \hat{f}_\beta^*(t'_i) = \inf_g \left\{ g : \frac{1}{B} \sum_{b=1}^B \mathbb{1}\{\hat{f}_b^*(t'_i) \leq g\} \geq \beta \right\}, \beta \in (0, 1).$$

---

<sup>8</sup>If nontrivial autocorrelation exists in the noise then a block bootstrap [69] will yield a better approximation of the trend filtering variability than the bootstrap implementations we discuss.

---

**Algorithm 2** Parametric bootstrap for fixed-input uncertainty quantification (when noise distribution  $\epsilon_i \sim Q_i$  is known *a priori*)

---

**Require:** Training Data  $(t_1, f(t_1)), \dots, (t_n, f(t_n))$ , hyperparameters  $\gamma$  and  $k$ , assumed noise distribution  $\epsilon_i \sim Q_i$ , prediction input grid  $t'_1, \dots, t'_m$

1: Compute the trend filtering point estimate at the observed inputs:

$$(t_1, \hat{f}_0(t_1)), \dots, (t_n, \hat{f}_0(t_n))$$

2: **for all**  $b$  in  $1 : B$  **do**

3: Define a bootstrap sample by sampling from the assumed noise distribution:

$$f_b^*(t_i) = \hat{f}_0(t_i) + \epsilon_i^* \quad \text{where } \epsilon_i^* \sim Q_i, \quad i = 1, \dots, n$$

4: Let  $f_b^*(t'_1), \dots, f_b^*(t'_m)$  denote the trend filtering estimate fit on the bootstrap sample and evaluated on the prediction grid  $t'_1, \dots, t'_m$

5: **end for**

**Output:** The full trend filtering bootstrap ensemble

$$\{f_b^*(t'_i)\}_{i=1, \dots, m} \\ b=1, \dots, B$$


---

**Algorithm 3** Wild bootstrap for fixed-input uncertainty quantification (when noise distribution is not known *a priori*)

---

**Require:** Training Data  $(t_1, f(t_1)), \dots, (t_n, f(t_n))$ , hyperparameters  $\gamma$  and  $k$ , prediction input grid  $t'_1, \dots, t'_m$

1: Compute the trend filtering point estimate at the observed inputs:

$$(t_1, \hat{f}_0(t_1)), \dots, (t_n, \hat{f}_0(t_n))$$

2: Let  $\hat{\epsilon}_i = f(t_i) - \hat{f}_0(t_i)$ ,  $i = 1, \dots, n$  denote the residuals

3: **for all**  $i$  **do**

4: Define a bootstrap sample by sampling from the following distribution:

$$f_b^*(t_i) = \hat{f}_0(t_i) + u_i^* \quad i = 1, \dots, n$$

where

$$u_i^* = \begin{cases} \hat{\epsilon}_i(1 + \sqrt{5})/2 & \text{with probability } (1 + \sqrt{5})/(2\sqrt{5}) \\ \hat{\epsilon}_i(1 - \sqrt{5})/2 & \text{with probability } (\sqrt{5} - 1)/(2\sqrt{5}) \end{cases}$$

5: Let  $f_b^*(t'_1), \dots, f_b^*(t'_m)$  denote the trend filtering estimate fit on the bootstrap sample and evaluated on the prediction grid  $t'_1, \dots, t'_m$

6: **end for**

**Output:** The full trend filtering bootstrap ensemble

$$\{f_b^*(t'_i)\}_{i=1, \dots, m} \\ b=1, \dots, B$$


---

Analogously, bootstrap sampling distributions and variability intervals for observable parameters of the signal may be studied by deriving a bootstrap parameter estimate from each trend filtering estimate within the bootstrap ensemble. For example, in Sections 5.2 and 5.4 we examine the bootstrap sampling distributions of several observable parameters of exoplanet transits and supernovae, respectively.

3.6.2. *Bayesian.* There is a well-studied connection between  $\ell_1$ -penalized least-squares regression and a Bayesian framework [see, e.g., 120, 45, 93]. A discussion specific to trend filtering can be found in [44].

3.7. *Relaxed trend filtering.* We are indebted to Ryan Tibshirani for a private conversation that motivated the discussion in this section. Trend filtering can be generalized to allow for greater flexibility through a technique that we call *relaxed trend filtering*<sup>9</sup>. Although the traditional trend filtering estimator is already highly flexible, there are certain settings in which the relaxed trend filtering estimator provides nontrivial improvements. In our experience, these typically correspond to settings where the optimally-tuned trend filtering estimator selects very few knots. For example, we use relaxed trend filtering in Section 5.2 to model the detrended, phase-folded light curve of a Kepler star with a planetary transit event.

The relaxed trend filtering estimate is defined through a two-stage sequential procedure in which the first stage amounts to computing the traditional trend filtering estimate discussed in Section 3.2. Recall the trend filtering minimization problem (31). For any given order  $k \in \{0, 1, 2, \dots\}$  and hyperparameter  $\gamma > 0$ , let us amend our notation so that

$$(46) \quad \hat{f}_0^{TF}(t) = \sum_{j=1}^n \hat{\beta}_j^{TF} h_j(t)$$

denotes the basis representation of the traditional trend filtering estimate. Further, define the index set

$$(47) \quad \mathcal{K}_\gamma = \left\{ 1 \leq j \leq n \mid \hat{\beta}_j^{TF} \neq 0 \right\}$$

that includes the indices of the non-zero falling factorial basis coefficients for the given choice of  $\gamma$ . Now let  $\hat{\beta}_j^{OLS}$ ,  $j \in \mathcal{K}_\gamma$ , denote the solution to the ordinary least-squares (OLS) minimization problem

$$(48) \quad \min_{\{\beta_j\}} \sum_{i=1}^n \left( f(t_i) - \sum_{j \in \mathcal{K}_\gamma} \beta_j h_j(t_i) \right)^2,$$

and define the corresponding OLS estimate as

$$(49) \quad \hat{f}_0^{OLS}(t) = \sum_{j \in \mathcal{K}_\gamma} \hat{\beta}_j^{OLS} h_j(t).$$

That is, the OLS estimate (49) uses trend filtering to find the knots in the piecewise polynomial, but then uses ordinary least-squares to estimate the reduced set of basis coefficients. The relaxed trend filtering estimate is then defined as a weighted average of the traditional trend filtering estimate and the corresponding OLS estimate:

$$(50) \quad \hat{f}_0^{RTF}(t) = \phi \hat{f}_0^{TF}(t) + (1 - \phi) \hat{f}_0^{OLS}(t),$$

for some choice of relaxation hyperparameter  $\phi \in [0, 1]$ .

In principle, it is preferable to jointly optimize the trend filtering hyperparameter  $\gamma$  and the relaxation hyperparameter  $\phi$ , e.g. via cross validation. However, it often suffices to choose  $\gamma$  and  $\phi$  sequentially, which in turn adds minimal computational cost on top of the traditional trend filtering procedure. Because of the trivial proximity of the falling factorial basis to the truncated power basis (established in [122] and [131]), it is sufficient to let  $\hat{f}_0^{OLS}$  be the  $k$ th order regression spline with knots at the input locations selected by the trend filtering estimator. In heteroskedastic settings, as discussed in Section 3.3, a piecewise polynomial or regression spline fit by weighted least-squares should be used in place of the OLS estimate (49).

---

<sup>9</sup>We choose this term because the generalization of trend filtering to relaxed trend filtering is analogous to the generalization of the lasso [120] to the relaxed lasso [81].

**4. MAIN APPLICATION: QUASAR LYMAN- $\alpha$  FOREST.** In this section, we use trend filtering to study the Lyman- $\alpha$  forest of quasar spectra. We choose to study this application in depth and then illustrate the breadth of trend filtering’s utility through our discussions in Section 5.

The Lyman- $\alpha$  ( $\text{Ly}\alpha$ ) forest is the name given to the absorption features seen in quasar spectra which are caused by neutral hydrogen ( $\text{H I}$ ) in the intergalactic medium between a quasar and an observer. When emitted from an accretion disk close to the central black hole, the light from the quasar has a relatively smooth spectrum—a continuum—caused by the summed black-body emission of gas with different temperatures at different disk radii [87]. Emission lines are also seen, and their intensities and line ratios supply information on the physical conditions in the line emitting gas. At least twenty broad emission lines, broadened by high velocities and temperatures can be measured in a single AGN, along with a similar number of narrow lines from colder gas [78]. The emitted spectrum therefore already consists of a superposition of components with varying degrees of smoothness. The  $\text{Ly}\alpha$  forest arises when this spectrum is further processed with the addition of absorption lines. Light moving towards the observer is redshifted into resonance with the  $\text{Ly}\alpha$  transition of  $\text{H I}$ , and the strength of absorption features is dictated by the densities of intergalactic material along the line of sight [108]. The smoothness of the absorption lines varies depending on the gas pressure, and thermal doppler broadening [50, 95]. Sharper absorption features, metal lines, are also caused by other intergalactic species, such as  $\text{C IV}$ ,  $\text{O VI}$  and  $\text{Mg II}$  [57, 99]. The usefulness of the  $\text{Ly}\alpha$  forest as a cosmological probe [e.g., 90] stems from its relationship to the matter density field in the Universe, effectively mapping out structure along each quasar-observer line of sight [e.g., 22, 71, 103]. In order to extract this information from noisy spectra and separate it from other components, it is useful to have a method that can deal with the complexities outlined above, i.e. one that can naturally adapt to varying degrees of smoothness without extensive tuning.

The relative fluctuations in the  $\text{Ly}\alpha$  forest transmitted flux fraction are of primary interest since they possess a monotonic relationship with the relative distribution of the absorbing  $\text{H I}$ . We utilize trend filtering to first denoise the spatially heterogeneous flux signal in an observed  $\text{Ly}\alpha$  forest. Estimates for the fluctuations in transmitted flux due to absorbing  $\text{H I}$  are then typically produced by coupling the denoised  $\text{Ly}\alpha$  forest with estimates for the quasar continuum and the cosmic mean transmitted flux in the  $\text{Ly}\alpha$  forest. We take an alternative approach: directly estimating the mean flux level—defined as the product of the continuum and cosmic mean transmitted flux—as in [21]. The mean flux level is a very smooth, spatially homogeneous function within the truncated  $\text{Ly}\alpha$  forest restframe. It is therefore appropriate to use a linear smoother (see Section 2) for this stage of estimation. Specifically, we use local polynomial regression (LOESS; [18]; [42]; [74]). In this section, we illustrate these methods on a mock quasar  $\text{Ly}\alpha$  forest from [8] and a real quasar  $\text{Ly}\alpha$  forest from the Baryon Oscillation Spectroscopic Survey Data Release 12 (BOSS DR12; [4]) of the Sloan Digital Sky Survey III [SDSS-III; 35, 2].

Historically,  $\text{Ly}\alpha$  forest analyses have typically utilized kernel smoothers [e.g., 21, 64], wavelets [e.g., 119], or Gaussian processes [e.g., 91].

4.1. *Notation.* Suppose we observe a quasar located at redshift  $z = z_0$ . Ignoring systematic effects such as sky contamination and interstellar extinction for the moment, the observational DGP of the  $\text{Ly}\alpha$  forest can be assumed to follow the model

$$(51) \quad f(\lambda) = f_0(\lambda) + \epsilon(\lambda), \quad \lambda \in \Lambda(z_0),$$

$$(52) \quad = \bar{F}(\lambda) \cdot C(\lambda) \cdot (1 + \delta_F(\lambda)) + \epsilon(\lambda),$$

Input	Definition	Range (quasar at $z = z_0$ )
$\lambda$	Observed wavelength	$\bar{\Lambda}(z_0)$
$\nu$	Rest wavelength	$\bar{\Lambda}_{\text{rest}}(z_0) = \bar{\Lambda}(z_0)/(1 + z_0)$
$z$	Redshift	$\Pi(z_0) = \bar{\Lambda}(z_0)/\lambda_{\text{Ly}\alpha} - 1$
$\zeta$	Log-wavelength (scaled)	$Z(z_0) = 10^4 \cdot \log_{10}(\bar{\Lambda}(z_0))$

TABLE 3

Various input spaces utilized for the Ly $\alpha$  forest analysis. Notation of functions is held constant, e.g.  $\delta_F(\cdot)$ , and an alteration of the input variable implicitly indicates a change of input spaces. Logarithmic wavelengths are scaled for numerical stability of the trend filtering optimization algorithm.

where  $f(\lambda)$  is the observed flux at wavelength  $\lambda$ ,  $f_0(\lambda)$  is the flux signal,  $\epsilon(\lambda)$  is zero mean white Gaussian noise,  $\Lambda(z_0) = (\lambda_{\text{Ly}\beta}, \lambda_{\text{Ly}\alpha}) \cdot (1 + z_0)$  is the redshifted Ly $\alpha$  forest,  $C(\lambda)$  is the flux of the unabsorbed quasar continuum,  $F(\lambda) = f_0(\lambda)/C(\lambda)$  is the transmitted flux fraction,  $\bar{F}(\lambda) = \mathbb{E}[F(\lambda)]$  is the mean transmitted flux fraction (over the sky) in the Ly $\alpha$  forest at redshift  $z = \lambda/\lambda_{\text{Ly}\alpha} - 1$ , and

$$(53) \quad \delta_F(\lambda) = F(\lambda)/\bar{F}(\lambda) - 1$$

is the fluctuation about the mean Ly $\alpha$  transmitted flux at redshift  $z = \lambda/\lambda_{\text{Ly}\alpha} - 1$ . Here,  $\delta_F$  is the quantity we are primarily interested in estimating since  $\delta_F \propto \delta_{\text{HI}}^{-1}$  at each fixed redshift, where  $\delta_{\text{HI}}$  is the density of H I. The estimation of the flux signal  $f_0$  is viewed as an ancillary step.

Although, in principle, it is preferable to study the full spectral range  $\Lambda(z_0)$  we have found that, in the nonparametric setting, estimating the quasar continuum near the localized Ly $\alpha$  and Ly $\beta$  emission peaks at the boundaries of the Ly $\alpha$  forest reduces the estimation accuracy in the interior of  $\Lambda(z_0)$ . Therefore, in this work we limit our analysis to the truncated Ly $\alpha$  forest range

$$(54) \quad \bar{\Lambda}(z_0) = (1045 \text{ \AA}, 1195 \text{ \AA}) \cdot (1 + z_0).$$

We simplify notation in this work by changing the input space of the functions introduced above by merely altering the input variable. For example, with respect to  $\delta_F$ , we maintain the notation  $\delta_F(\cdot)$  for all inputs  $\lambda, \nu, z, \zeta$ , while it is understood that a proper change of input spaces has taken place. The various input spaces are defined in Table 3.

*4.2. Trend filtering the observed flux.* We use quadratic trend filtering to estimate the flux signal  $f_0$  of the observational model (51). In both BOSS DR12 and the [8] mock catalog, the quasar spectra are sampled on equally-spaced grids in logarithmic wavelength space with  $\Delta \log_{10}(\lambda_i) = 10^{-4}$  dex (in logarithmic angstroms). Furthermore, flux measurement variances are provided by the BOSS pipeline [12], accounting for the statistical uncertainty introduced by photon noise, CCD read noise, and sky-subtraction error. We correct the BOSS spectrum for interstellar extinction with the [17] extinction law and the [110] dust map.

We fit the trend filtering estimator on the equally-spaced logarithmic grid and tune the complexity by minimizing Stein's unbiased risk estimate (SURE) of the fixed-input mean-squared error (42). More precisely, we fit the trend filtering estimator in the input space  $Z(z_0) = 10^4 \cdot \log_{10}(\bar{\Lambda}(z_0))$ , as defined in Table 3, where we add the scaling to unit spacing for numerical stability of the trend filtering convex optimization.

*4.3. Nonparametric continuum estimation.* We utilize a modified [21] approach to propagate the trend filtering estimate for the flux signal  $f_0$  from Section 4.2 into an estimate for the fluctuation field  $\delta_F$  along a line of sight to an observed quasar. Namely, given the trend

filtering estimate  $\hat{f}_0$ , we directly estimate the smooth mean flux level defined as the product  $m = \bar{F} \cdot C$  and then define the  $\delta_F$  estimates via the transformation  $\hat{\delta}_F := \hat{f}_0 / \hat{m} - 1$ . We carry out the estimation of  $m$  via a wide-kernel LOESS smooth of the trend filtering estimate, with the specific bandwidth of the kernel selected by optimizing over a large sample of mock spectra (detailed in Section 4.4). We find that regressing on the fitted values of the trend filtering estimate—instead of the observational DGP (51)—significantly improves the accuracy and robustness of the  $\delta_F$  estimates. We carry out the LOESS estimation in the Ly $\alpha$  restframe  $\bar{\Lambda}_{\text{rest}}(z_0)$  (see Table 3) in order to remove the effect of redshifting on the smoothness of  $m$ . The LOESS estimation of  $m$  is a fully nonparametric procedure and provides a reduction in bias over popular parametric approaches such as low-order power laws and principal components analyses (PCA). The sole assumption of our LOESS approach is that, in the restframe, the mean flux level  $m$  always has a fixed degree of smoothness, defined by an optimal fixed kernel bandwidth. There is of course a bias-variance tradeoff here; namely, the decreased bias comes with a modest increase in variance compared to a parametric power law or low-dimensional PCA model. Our decision in favor of the LOESS approach directly reflects our stance that low bias is preferable to low variance in this context since statistical uncertainty due to estimator variability is tracked by our uncertainty quantification (Section 4.5), while uncertainty due to modeling bias is not easily quantifiable. Therefore, we can be more confident that significant fluctuations in the estimated  $\delta_F$  field are in fact real, and not due to statistical bias in the quasar continuum estimate.

To be explicit, the LOESS estimator for  $m$  is a regression on the data set

$$(55) \quad \{(\nu_i, \hat{f}_0(\nu_i; \hat{\gamma}))\}_{i=1}^n, \nu_i \in \bar{\Lambda}_{\text{rest}}(z_0),$$

which can be viewed as arising from the DGP

$$(56) \quad \hat{f}_0(\nu_i; \hat{\gamma}) = m(\nu_i) + \rho_i,$$

where  $\hat{f}_0$  is the trend filtering estimate fixed at the minimum SURE hyperparameter  $\hat{\gamma}$ ,  $e_i = \hat{f}_0(\nu_i; \hat{\gamma}) - f_0(\nu_i)$  are the errors of the trend filtering estimate, and  $\rho_i = m(\nu_i) \cdot \delta_F(\nu_i) + e_i$  are autocorrelated fluctuations about zero. The LOESS estimator is the natural extension of kernel regression [83, 134] to higher-order local polynomials. Given a kernel function  $K(\cdot)$  with bandwidth  $h > 0$ , for each  $i = 1, \dots, n$ , the LOESS estimator is obtained by minimizing

$$(57) \quad \sum_{j=1}^n \left( \hat{f}_0(\nu_j; \hat{\gamma}) - \phi_{\nu_i}(\nu_j; \beta_0, \dots, \beta_d) \right)^2 K\left(\frac{|\nu_j - \nu_i|}{h}\right),$$

and letting  $\hat{m}(\nu_i) = \hat{\beta}_0$ , where  $\phi_{\nu_i}(\cdot; \beta_0, \dots, \beta_d)$  is a  $d$ th order polynomial centered at  $\nu_i$ . Specifically, we utilize the local linear regression estimator (LLR;  $d = 1$ ) and the Epanechnikov kernel [37]

$$(58) \quad K(t) = \frac{3}{4}(1 - t^2) \mathbb{1}\{|t| < 1\}.$$

The LLR estimator is described in full detail by Algorithm 4. Given the trend filtering estimate  $\hat{f}_0$  and the LLR estimate  $\hat{m}$ , the  $\delta_F$  estimates are then defined as

$$(59) \quad \hat{\delta}_F(z_i; \hat{\gamma}) = \frac{\hat{f}_0(z_i; \hat{\gamma})}{\hat{m}(z_i)} - 1, z_1, \dots, z_n \in \Pi(z_0),$$

where we deliberately express  $\hat{m}$  as “hyperparameter-less” since  $\gamma$  has already been fixed at the minimum SURE value  $\hat{\gamma}$  and we provide the optimal LOESS bandwidth value  $h_0 = 74 \text{ \AA}$ —optimized over the large mock sample. Here, we have also done a change of variables

---

**Algorithm 4** LOESS (local linear) estimator for mean flux level
 

---

**Require:** Training Data  $\{(\nu_i, \hat{f}_0(\nu_i; \hat{\gamma}))\}_{i=1}^n$ , Bandwidth  $h_0 = 74 \text{ \AA}$

1: **for all**  $i$  **do**

2: Let  $\hat{\beta}_0, \hat{\beta}_1$  minimize the locally weighted sum of squares

$$\sum_{j=1}^n \left( \hat{f}_0(\nu_j; \hat{\gamma}) - \beta_0 - \beta_1(\nu_j - \nu_i) \right)^2 \cdot K \left( \frac{|\nu_j - \nu_i|}{h_0} \right).$$

3: Let  $\hat{m}(\nu_i) = \hat{\beta}_0(\nu_i)$ .

4: **end for**

**Output:**  $\{\hat{m}(\nu_i)\}_{i=1}^n$

---

to redshift space—our desired input domain for studying the H I density fluctuations in the IGM.

Although  $h_0$  is chosen to directly optimize  $\hat{\delta}_F$  accuracy, an estimate for the quasar continuum  $C(\cdot)$  arises intrinsically:

$$(60) \quad \hat{C}(\nu_i) = \bar{F}(\nu_i)^{-1} \cdot \hat{m}(\nu_i), \nu_i \in \bar{\Lambda}_{\text{rest}}(z_0),$$

where precise estimates of  $\bar{F}$  follow from a rich literature [e.g., 11, 80, 43, 23, 92, 9]. The  $\delta_F$  estimates could then be equivalently restated as

$$(61) \quad \hat{\delta}_F(z_i; \hat{\gamma}) = \frac{\hat{F}(z_i; \hat{\gamma})}{\bar{F}(z_i)} - 1, z_1, \dots, z_n \in \Pi(z_0),$$

where

$$(62) \quad \hat{F}(z_i; \hat{\gamma}) = \hat{f}_0(z_i; \hat{\gamma}) / \hat{C}(z_i).$$

4.4. *Calibrating continuum smoothness.* We utilize a sample of 124,709 mock quasar spectra from the [8] catalog to optimize the bandwidth of the LOESS estimator for the mean flux level that intrinsically removes the effect of the quasar continuum. Our mock data reduction is detailed in the appendix and the redshift distribution of the quasars is shown in the top panel of Figure 4.

For each mock quasar Ly $\alpha$  forest with DGP  $Q_j$ ,  $j = 1, \dots, 124,709$ , we first compute the trend filtering hyperparameter value that minimizes the *true* fixed-input mean-squared prediction error

$$(63) \quad \gamma_0^j = \arg \min_{\gamma > 0} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{Q_j} \left[ (f(\zeta_i) - \hat{f}_0(\zeta_i; \gamma))^2 \mid \zeta_1, \dots, \zeta_n \right].$$

Then, given the trend filtering restframe fitted values  $\{(\nu_i, \hat{f}_0(\nu_i; \gamma_0^j))\}_{i=1}^n$ , we fit a LOESS estimator with bandwidth  $h$  and define the error (as a function of  $h$ ) of the resulting  $\delta_F$  estimator as the fixed-input mean absolute deviation (MAD) error

$$(64) \quad R_j(h) = \frac{1}{n} \sum_i \mathbb{E}_{Q_j} \left[ |\delta_F(\nu_i) - \hat{\delta}_F(\nu_i; \gamma_0^j, h)| \mid \nu_1, \dots, \nu_n \right],$$

where  $\mathbb{E}_{Q_j}$  denotes the mathematical expectation over the randomness arising from the observational DGP  $Q_j$ . Because we can repeatedly sample from each mock quasar DGP, the expectations in (63) and (64) can be computed to an arbitrary precision. Here, we utilize 300 realizations of each DGP to approximate the mathematical expectations.

## Trend Filtering in Astronomy

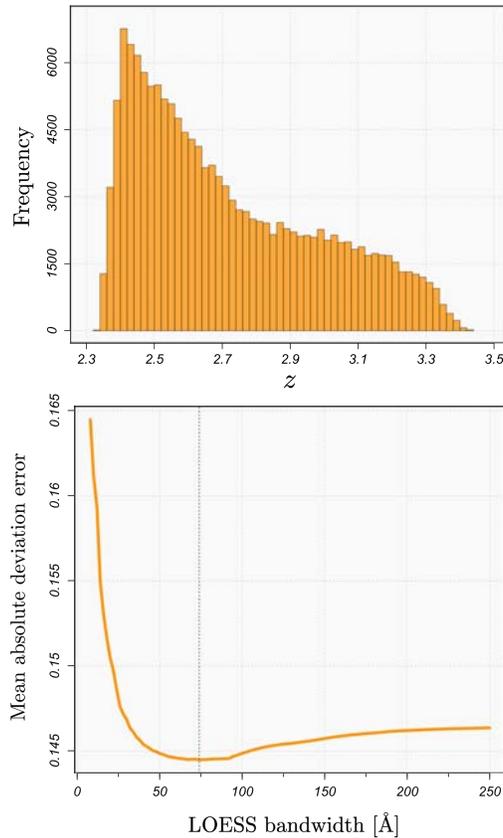


FIG 4. **Top:** Distribution of mock quasar redshifts (data reduction detailed in the appendix). We utilize this sample of 124,709 quasars to calibrate the optimal nonparametric continuum smoothness. **Bottom:** Mean absolute deviation error curve for selecting the optimal kernel bandwidth for the LOESS (local linear) estimator of the mean flux level, averaged over the 124,709 spectra in the mock sample. The optimal choice of bandwidth is  $h_0 = 74 \text{ \AA}$ .

We then define the optimal choice of  $h$  as the minimizer of the summed error over the full sample of  $m = 124,709$  mock quasar spectra:

$$(65) \quad h_0 = \arg \min_{h>0} \sum_{j=1}^m R_j(h).$$

The aggregate error curve is shown in the bottom panel of Figure 4, yielding an optimal value of  $h_0 = 74 \text{ \AA}$ . We find that defining  $R(h)$  as the conditional MAD error—instead of the conditional MSE—provides an essential boost in robustness that keeps the error from being dominated by a very small proportion of worst-case estimates. More complex choices of bandwidth, e.g. variable bandwidths that depend on the S/N ratio of the trend filtering estimator, the restframe pixel spacing, and/or the redshift of the quasar, do not significantly improve upon the  $h_0 = 74 \text{ \AA}$  restframe fixed bandwidth.

**4.5. Uncertainty quantification.** Given the assumed noise model  $\epsilon_i \sim N(0, \sigma_i^2)$ ,  $i = 1, \dots, n$  provided by the BOSS pipeline [12], we can construct a pointwise variability band for  $\hat{\delta}_F$  via an augmentation of the parametric bootstrap outlined in Algorithm 2. Specifically, given the parametric bootstrap ensemble of trend filtering estimates provided by Algorithm 2, we fit the mean flux level of each with the LOESS estimator detailed in Algorithm 4, and then

define the  $\widehat{\delta}_F$  bootstrap ensemble

$$(66) \quad \widehat{\delta}_{F,b}^*(z_i) = \frac{\widehat{f}_b^*(z_i)}{\widehat{m}_b^*(z_i)} - 1, i = 1, \dots, n, b = 1, \dots, B,$$

where, for each  $b = 1, \dots, B$ ,  $\widehat{m}_b^*$  is the LOESS estimate fit to the data set  $\{(\nu_i, \widehat{f}_b^*(\nu_i))\}_{i=1}^n$ . Note that refitting the LOESS estimator on each bootstrap realization allows us to track the extra variability introduced into the  $\delta_F$  estimates by the uncertainty in the continuum estimate. Analogous to (44), a  $(1 - \alpha) \cdot 100\%$  quantile-based pointwise variability band for  $\widehat{\delta}_F$  is then given by

$$(67) \quad V_{1-\alpha}(z_i) = \left( \widehat{\delta}_{F,\alpha/2}^*(z_i), \widehat{\delta}_{F,1-\alpha/2}^*(z_i) \right), i = 1, \dots, n.$$

where

$$(68) \quad \widehat{\delta}_{F,\beta}^*(z_i) = \inf_g \left\{ g : \frac{1}{B} \sum_{b=1}^B \mathbb{1} \{ \widehat{\delta}_{F,b}^*(z_i) \leq g \} \geq \beta \right\},$$

for any  $\beta \in (0, 1)$ .

**4.6. Results.** A mock quasar spectrum from [8] and a real quasar spectrum from the Baryon Oscillation Spectroscopic Survey (BOSS; [4]) are displayed in Figure 5, with the results of our Ly $\alpha$  forest analysis overlaid. Starting with the top panel, the quadratic trend filtering estimate is fit on the equally-spaced observations in logarithmic wavelength space and then transformed to the restframe wavelength space, where the LOESS estimate for the mean flux level—the product of the continuum and cosmic mean Ly $\alpha$  absorption—is then fit to the trend filtering fitted values. The  $\delta_F$  estimates are then computed according to (59) and displayed in redshift space in the second panel, where they closely track the true  $\delta_F$  defined by (53). Furthermore, the 95% bootstrap variability band defined in (67) can be seen to almost fully cover the true  $\delta_F$  signal. The estimated  $\delta_F$  can be interpreted as an inversely proportional proxy for the deviations from the mean H I density in the intervening intergalactic medium at each redshift—with negative values of  $\delta_F$  corresponding to (epoch-relative) over-densities of H I and positive values corresponding to under-densities.

The third and fourth panels are the analogous plots for a BOSS quasar spectrum Ly $\alpha$  forest (Data Release 12, Plate = 6487, MJD = 56362, Fiber = 647). The quasar is located in the northern galactic cap at (RA, Dec,  $z$ )  $\approx$  (196.680°, 31.078°, 2.560).

**5. Further Applications.** The analysis of signals possessing varying degrees of smoothness permeates many areas of astronomy. In this section, we discuss a variety of further applications for which trend filtering may find use. For the sake of brevity, we discuss these applications in less detail than the Ly $\alpha$  forest analysis in Section 4. Naturally, we expect trend filtering may find many uses in astronomy beyond those we explicitly discuss.

The applications we discuss in this section 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 5.1, we discuss constructing spectral template libraries for astronomical objects by trend filtering coadded spectroscopic observations. We illustrate our approach on quasar, galaxy, and stellar spectra from the Sloan Digital Sky Survey. Emission-line parameters can also be robustly estimated by fitting radial basis functions (e.g., Gaussians) to trend-filtered estimates near emission lines. In Section 5.2, we use relaxed trend filtering to model the detrended, phase-folded light curve of a Kepler stellar system with

## Trend Filtering in Astronomy

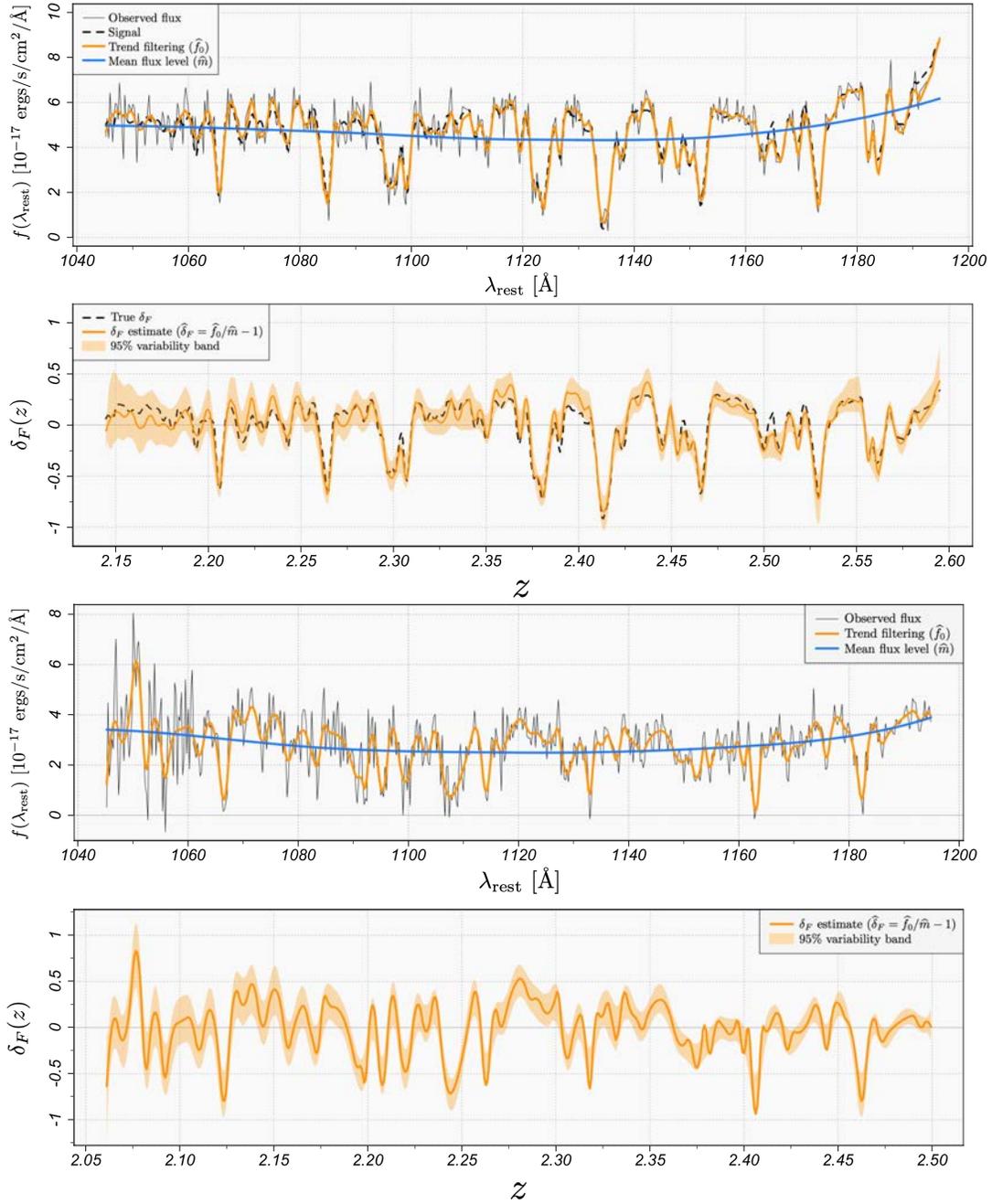


FIG 5. Results of Ly $\alpha$  forest analysis. **Top panel:** Ly $\alpha$  forest of a mock quasar spectrum [8] in the restframe, with the quadratic trend filtering estimate shown in orange and the LOESS (local linear) estimate for the mean flux level shown in blue. **Second panel:** The redshift-space fluctuations in the Ly $\alpha$  transmitted flux fraction, with our estimate superposed. The fluctuations inversely trace the relative under- and over-densities of H I in the intervening intergalactic medium between Earth and the quasar. **Third and Fourth panels:** Analogous plots for a real quasar Ly $\alpha$  forest from the twelfth data release of the Baryon Oscillation Spectroscopic Survey ([4]; Plate = 6487, MJD = 56362, Fiber = 647). The quasar is located at (RA, Dec,  $z$ )  $\approx$  (196.680 $^\circ$ , 31.078 $^\circ$ , 2.560).

a transiting exoplanet. We derive estimates and full uncertainty distributions for the transit depth and total transit duration. In Section 5.3, we use trend filtering to denoise a detrended, phase-folded Kepler light curve of an eclipsing binary (EB) system. We illustrate that trend filtering provides significant improvements upon the popular `polyfit` method of [106] that is used to model Kepler EB light curves and derive observable parameters. In Section 5.4, we discuss using trend filtering to construct light-curve templates of supernova (SN) explosions. We illustrate this approach on a SN light curve obtained from the Open Supernova Catalog [53]. Furthermore, we derive estimates and full uncertainty distributions for a set of observable parameters—namely, the maximum apparent magnitude, the time of maximum, and the decline rate. Finally, in Section 5.5, we briefly discuss a different, non-data-analysis application of trend filtering. Specifically, we discuss the use of trend filtering as a tool for fast and flexible one-dimensional data reduction and compression. The flexibility of the trend filtering estimator, paired with its efficient speed and storage capabilities, make it a potentially powerful tool to include in large-scale (one-dimensional) astronomical data reduction and storage pipelines.

5.1. *Spectral template generation and estimation of emission-line parameters.* Automated spectral classification and redshift estimation pipelines require rich template libraries that span the full space of physical objects in the targeted set in order to achieve high statistical accuracy. In this section, we discuss using trend filtering to construct spectral templates from observational spectra. We describe the procedure here for generating a single spectral template from a well-sampled observed spectrum. Suppose we observe coadded flux measurements of a targeted source at wavelengths  $\lambda_1, \dots, \lambda_n \in \Lambda$  according to the data generating process

$$(69) \quad f(\lambda_i) = f_0(\lambda_i) + \epsilon_i,$$

where the observations have been corrected for systematic effects (e.g., sky subtraction, interstellar extinction, etc.) and the  $\epsilon_i$  are mean-zero errors that arise from instrumental uncertainty and systematic miscalibrations. After removal of potentially problematic pixels (e.g., near bright sky lines), let  $\hat{f}_0(\lambda)$ ,  $\lambda \in \Lambda$  denote the trend filtering estimate fit to the observations. Given a confident object classification and redshift estimate  $z_0$  (e.g., determined by visual inspection), we then define the restframe spectral template

$$(70) \quad b(\lambda_{\text{rest}}) = \hat{f}_0(\lambda/(z_0 + 1)), \lambda_{\text{rest}} \in \Lambda/(z_0 + 1),$$

and store it in the respective object template library.

In Figure 6, we show three optical coadded spectra from the twelfth data release of the Baryon Oscillation Spectroscopic Survey (BOSS DR12; [4]) of the Sloan Digital Sky Survey III (SDSS-III; [35, 2]). The figure is zoomed to a subinterval of the optical range for visual clarity and the spectra are easily identifiable as a quasar, a galaxy, and a star, respectively. We fit an error-weighted quadratic trend filtering estimate to each spectrum in the logarithmic-angstrom wavelength space in which the BOSS spectra are gridded, and tune the hyperparameter to minimize Stein’s unbiased risk estimate (see Section 3.5). The trend filtering estimates give good results, adequately adapting to even the strongest emission and absorption features in each spectrum.

As in the BOSS spectroscopic pipeline [12], after a spectral classification and redshift measurement have been precisely determined, emission-line parameter estimates can then be obtained by fitting Gaussian radial basis functions to the emission lines of the spectrum “best fit”—nonlinearly optimizing the amplitudes, centroids, and widths. We propose that the trend filtering estimate of a spectrum be used as the “best fit” for this type of procedure, e.g. instead of the low-dimensional principal components models currently used by the BOSS pipeline. The relative magnitudes of the emission-line parameter estimates can then be used to determine object subclassifications.

## Trend Filtering in Astronomy

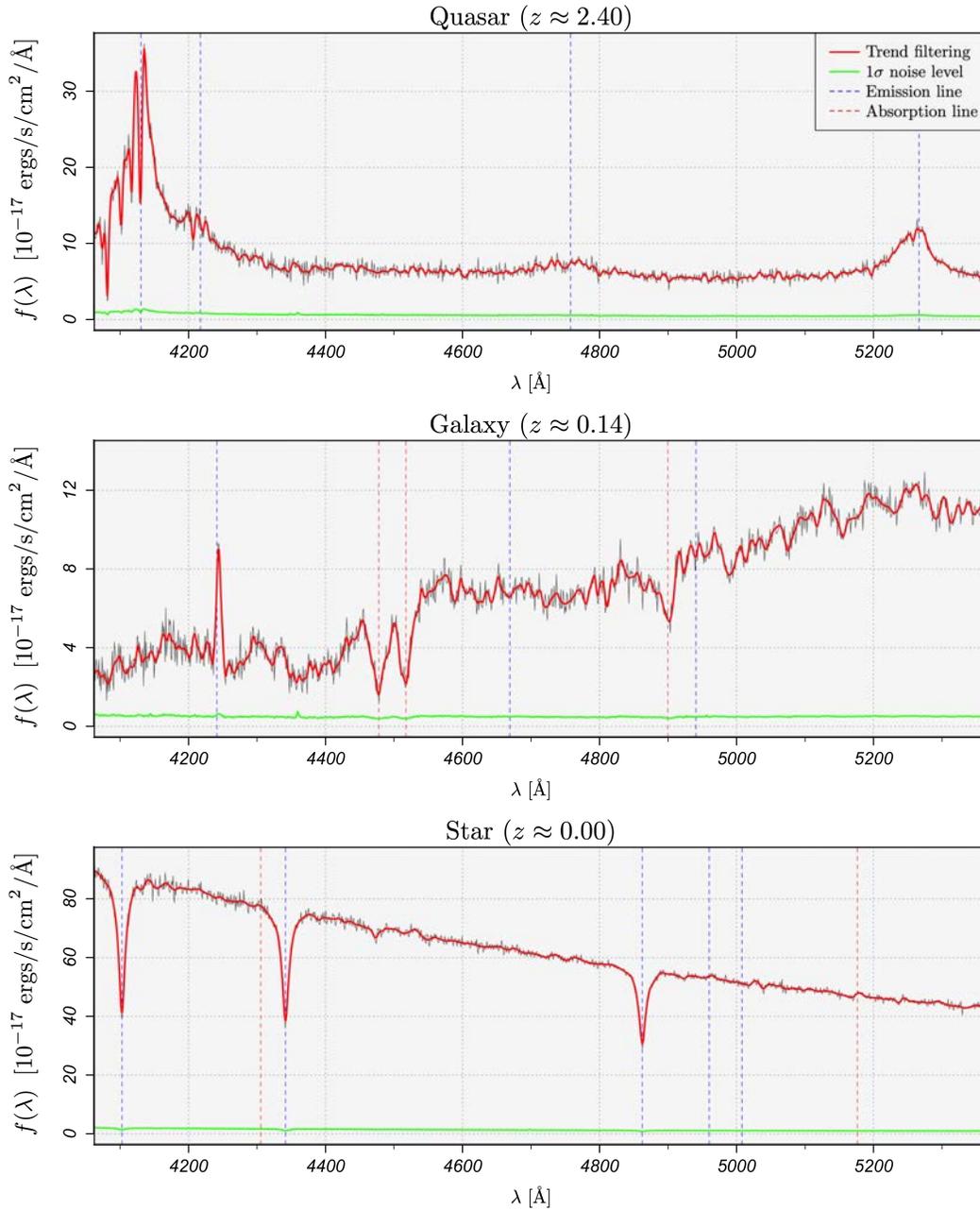


FIG 6. Optical coadded spectra collected by the Baryon Oscillation Spectroscopic Survey [4] of the Sloan Digital Sky Survey III [35, 2]. From top to bottom, a quasar (DR12, Plate = 7140, MJD = 56569, Fiber = 58) located at (RA, Dec,  $z$ )  $\approx$  (349.737°, 33.414°, 2.399), a galaxy (DR12, Plate = 7140, MJD = 56569, Fiber = 68) located at (RA, Dec,  $z$ )  $\approx$  (349.374°, 33.617°, 0.138), and a star (DR12, Plate = 4055, MJD = 55359, Fiber = 84) located at (RA, Dec,  $z$ )  $\approx$  (236.834°, 0.680°, 0.000). We fit a quadratic trend filtering estimate to each spectrum in the logarithmic wavelength space in which the observations are equally spaced, and optimize the hyperparameter by minimizing Stein's unbiased risk estimate. Given confidently determined redshifts (e.g., determined by visual inspection), the trend filtering estimate for each object can be scaled to the restframe and stored as a spectral template. Furthermore, emission-line parameter estimates for a spectrum can be obtained by fitting Gaussian radial basis functions to the emission lines of the trend filtering estimate.

5.2. *Exoplanet transit modeling.* In this section we discuss the use of trend filtering for modeling the photometric time series of an extrasolar planet transiting its host star. Given a phase-folded stellar light curve, corrected for stellar variability and spacecraft motion systematics, trend filtering can be used to automatically produce fully nonparametric estimates and uncertainties for the transit depth and total transit duration.

We demonstrate our approach on long-cadence photometric observations of Kepler-10 (KOI-72, KIC 11904151; [13]). Kepler-10 is confirmed to host at least two exoplanets: Kepler-10b (KOI-72 b, KIC 11904151 b; [7]) and Kepler-10c (KOI-72 c, KIC 11904151 c; [47]). Each planet was first detected via the transit method—a measurable periodic dimming in the photometry caused by the planet crossing in front of the host star. Here, we use trend filtering to estimate a nonparametric transit model for Kepler-10c and derive depth and duration measurements. The results of our analysis are displayed in Figure 7. The top panel displays a sample of the Kepler-10 long-cadence (30-min. increment), quarter-stitched, median detrended, relative flux light curve processed by the Kepler pipeline [61], which we accessed from the NASA Exoplanet Archive [3]. The observed transit events of Kepler-10c are indicated by the vertical dashed lines. The middle panel displays the light curve (with  $1\sigma$  error bars) after phase-folding with respect to the  $\sim 45.29$  day orbital period of Kepler-10c and zooming in on the transit event. We fit a relaxed quadratic trend filtering estimate (see Section 3.7) to the phased data, weighted by the measurement variances provided by the Kepler pipeline and tuned by sequential  $K$ -fold cross validation. The optimal relaxation hyperparameter is  $\hat{\phi} = 0.14$ , indicating that relaxation significantly improves the traditional trend filtering estimate in this setting. In particular, we find that relaxation allows the estimate to faithfully capture the sharp transitions corresponding to the beginning of the ingress phase and the end of the egress phase. The relaxed trend filtering estimate is overlaid on the phase-folded light curve, along with 95% variability bands. Estimates for the transit depth and total transit duration follow immediately from the relaxed trend filtering estimate, as detailed below. The estimated inception and termination of the transit event are indicated by the vertical dashed lines in the middle panel. The nonparametric bootstrap sampling distributions (see Algorithm 1) of the transit depth and total transit duration measurements are displayed in the bottom panel. Our point estimates for the transit depth and total transit duration for Kepler-10c are  $\hat{\delta} = 488.292$  ppm and  $\hat{T} = 6.927$  hours, respectively.

The knot-selection property of trend filtering is particularly appealing in this setting because it provides interpretation as to where the transit event begins and ends. Specifically, we define our estimate of the inception of the transit event  $\hat{T}_0$  as the leftmost knot selected by the trend filtering estimator and we define our estimate of the termination of the transit event  $\hat{T}_1$  as the rightmost knot<sup>10</sup>. The total transit duration estimate then follows as  $\hat{T} = \hat{T}_1 - \hat{T}_0$ . Naturally, we define our transit depth estimate  $\hat{\delta}$  to be the minimum of the relaxed trend filtering estimate.

It is thus far unclear to us whether trend filtering can also reliably detect the end of the ingress and beginning of the egress phases—e.g. via knot selection or examining changes in the estimated derivatives—and therefore also provide nonparametric ingress/egress duration measurements. We recommend pairing trend filtering with the traditional analytical transit model search [e.g., 76] for these particular measurements. That is, given the transit depth and total transit duration measurements provided by trend filtering, an analytical planet model can be fit over a parameter space that is constrained by the trend filtering parameter measurements. Coupling the two methods in this way therefore also provides significant computational speedups over a traditional single-stage analytical model search since it greatly reduces

<sup>10</sup>Note that the boundary points of the input space are not considered knots.

## Trend Filtering in Astronomy

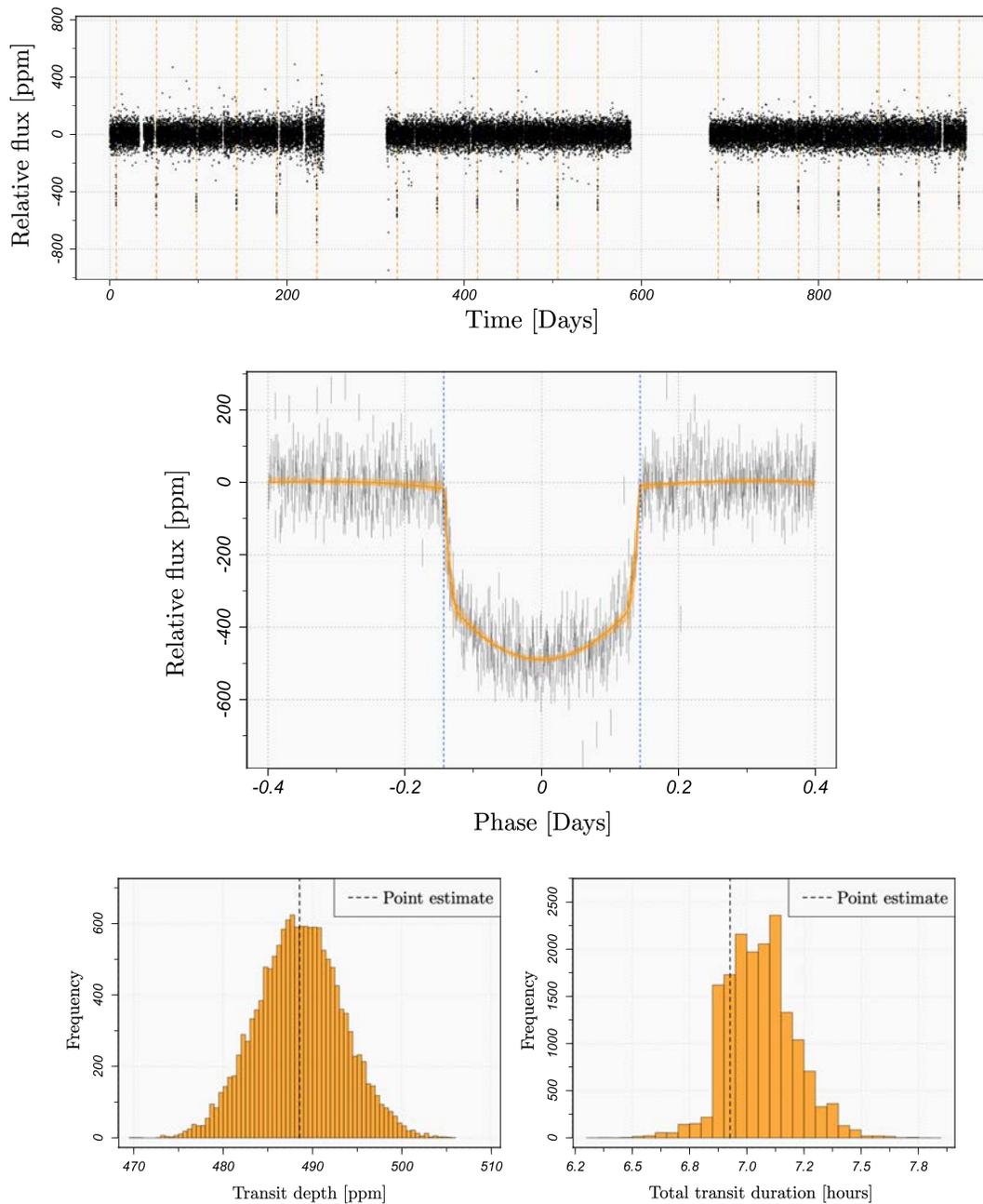


FIG 7. *Kepler-10c* transit light curve analysis. **Top:** Long-cadence (30-min.), quarter-stitched, median detrended, relative flux light curve ( $LC\_DETRENDED$ ) of the confirmed exoplanet host *Kepler-10* (KOI-072, KIC 11904151; [7]), processed by the Kepler pipeline [61] and obtained from the NASA Exoplanet Archive [3]. Vertical lines indicate the observed transit events of the system's second confirmed planet *Kepler-10c* (KOI-072 c, KIC 11904151 c; [47]). **Middle:** Phase-folded transit light curve for *Kepler-10c* ( $\sim 45.29$  day orbital period) with  $1\sigma$  error bars. The error-weighted relaxed trend filtering estimate, optimized by sequential  $K$ -fold cross validation, is superposed with 95% variability bands. The estimated inception and termination of the transit event are indicated by the vertical dashed lines. The estimated transit depth and total transit duration are  $\hat{\delta} = 488.292$  ppm and  $\hat{T} = 6.927$  hours, respectively. **Bottom:** Bootstrap sampling distributions of the transit depth and transit duration estimates.

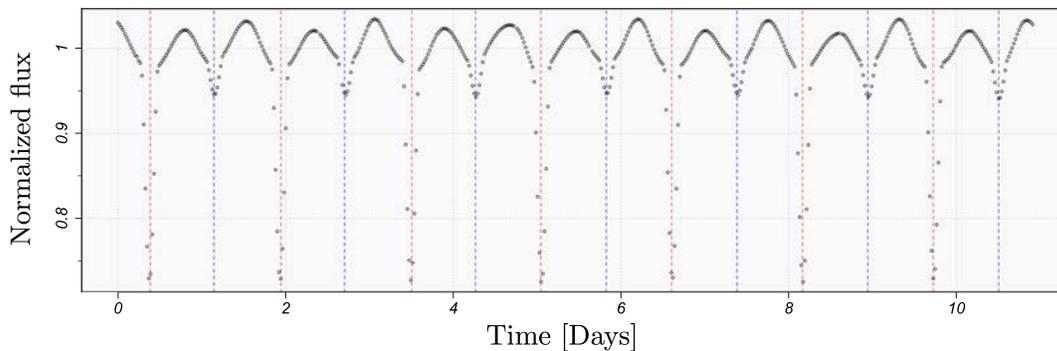


FIG 8. Long-cadence (30-min. increment), detrended, median-normalized light curve of a Kepler eclipsing binary system (KIC 6048106; [13, 105]). The vertical red lines mark the primary eclipses (the eclipses of the hotter star) and the vertical blue lines mark the secondary eclipses (the eclipses of the cooler star). KIC 6048106 has an orbital period of  $\sim 1.559$  days.

the dimensionality of the parameter space to be searched over. Furthermore, the relaxed trend filtering estimate may provide a benchmark  $\chi^2$  statistic for the constrained analytical model comparison.

A dedicated paper on this application of trend filtering is forthcoming.

**5.3. Eclipsing binary modeling.** In this section we discuss how trend filtering can further improve the work of the Kepler Eclipsing Binary working group, specifically in regard to the Eclipsing Binaries via Artificial Intelligence (EBAI) pipeline used to characterize Kepler eclipsing binary stars via their phase-folded light curves [106, 105, 113, 79]. The EBAI pipeline utilizes an artificial neural network (ANN) to estimate a set of physical parameters for each binary pair (the temperature ratio, sum of fractional radii, photometric mass ratio, radial and tangential components of the eccentricity, fillout factor, and inclination) from the observables of the phase-folded light curve (e.g., the eclipse widths, depths, and separations). [106] outline the EBAI light curve pre-processing algorithm, which they call `polyfit`, that provides the crucial step of taking a noisy, irregularly-spaced, phase-folded light curve (detrended for spacecraft motion and normalized by the median flux) and outputting a denoised and gridded phase-folded light curve, which is then fed to the ANN. We propose that trend filtering be used for this pre-processing step instead of the `polyfit` algorithm for the reasons detailed below.

An eclipsing binary (EB) light curve is characterized by periodic dips in the observed brightness that correspond to eclipse events along the line of sight to an observer. In particular, there are two eclipses per orbital period—a primary and a secondary eclipse. The primary eclipse occurs when the hotter star is eclipsed by the cooler star and produces a comparatively deep dip in observed brightness. Analogously, the secondary eclipse occurs when the cooler star is eclipsed by the hotter star and produces a comparatively shallow dip in the observed brightness. Depending on the effective temperature ratio and orbital period of the EB, the dips may range from very narrow and abrupt to very wide and smooth. In Figure 8, we display a detrended, median-normalized, long-cadence (30 min. increment) light curve of a Kepler EB (KIC 6048106; [13, 105]), with the primary eclipses and secondary eclipses designated by dashed red and blue lines, respectively. The orbital period of KIC 6048106 is  $\sim 1.559$  days.

After phase-folding with respect to the estimated EB orbital period and centering the primary eclipse at `Phase = 0`, the purpose of the EBAI light-curve pre-processing step is to faithfully extract the signal of the phase-folded light curve and evaluate it on a regular grid so

that it can then be input into the EBAI ANN. We show a comparison of the `polyfit` algorithm of [106] and our trend filtering approach in Figure 9 on the phase-folded KIC 6048106 light curve. We choose a relatively high S/N light curve here in order to elucidate the significant statistical bias that underlies `polyfit`. The `polyfit` algorithm fits a piecewise quadratic polynomial by weighted least-squares with four knots selected by a randomized computational search over the phase space. The piecewise quadratic polynomial is forced to be continuous, but no smoothness constraints are placed on the derivatives of the estimate at the knots. Recalling our discussion in Section 2.1, this overly-stringent modeling assumption leads to significant statistical bias in the light-curve estimate. The bias is particularly apparent by examining the residuals of the `polyfit` estimate, which we display below the light curve. Moreover, recalling our discussion of variable-knot regression splines in Section 3.1.1, the randomized partial search over the space of feasible knots inherently provides no guarantee of finding a global solution—leaving the algorithm susceptible to extreme failure scenarios. We display a quadratic trend filtering estimate of the KIC 6048106 phase-folded light curve in the bottom panel of Figure 9, with the hyperparameter chosen by  $K$ -fold cross validation. The trend filtering estimate accurately recovers the signal of the light curve (clearly apparent here by examining a high S/N light curve) and produces a desirable random residual scatter about zero. Since the statistical bias introduced by the `polyfit` pre-processing stage propagates through to the EBAI ANN as systematic bias in the input data, we are confident that the use of trend filtering will in turn improve the error rate of the EBAI ANN output-parameter estimates.

#### 5.4. *Supernova light-curve template generation and estimation of observable parameters.*

In this section we demonstrate the use of trend filtering for generating light-curve templates of supernova (SN) events and estimating observable parameters. We illustrate our approach on SN 2016coi (ASASSN-16fp; [58]) by constructing a B-band light-curve template from the well-sampled observations of [14] and [104] and deriving nonparametric estimates and full uncertainty distributions for the maximum apparent brightness, the time of maximum, and the decline rate parameter  $\Delta m_{15}(B)$  introduced by [98]. The improvement yielded by trend filtering as a tool for SN light-curve template generation, compared to, for example, the `SN00PY` cubic smoothing spline approach of the Carnegie Supernova Project [CSP; 20, 16, 15] primarily corresponds to light curves with an especially bright peak magnitude and fast decline rate. In such cases, trend filtering is better able to recover the abruptness of the peak, the initial sharp decline, and the subsequent slow decay. This behavior is particularly characteristic of Type Ia SNe [e.g., 135]. In cases where the peak is not particularly prominent, trend filtering and cubic smoothing splines produce nearly identical estimates. Our procedure for generating the SN light-curve templates requires reasonably well-sampled observations (in particular, with the initial observation occurring before maximum light). The resulting template libraries can then be used to classify SNe with partially-sampled light curves and derive parameter estimates [e.g., 10].

The SN light-curve template generation procedure is analogous to the spectral template generation procedure discussed in Section 5.1, so we discuss it in less formal detail here. Naturally, the same procedure can also be implemented for generating fixed-time spectral templates of SN events. Given a well-sampled light curve, corrected for systematic effects (e.g.,  $K$ -corrections and interstellar reddening), we propose the use of quadratic trend filtering to generate a “best fit” to the observations. Given a confident type classification, the trend filtering estimate can then be stored as a light-curve template in the respective library.

We show a B-band light curve for SN 2016coi (ASASSN-16fp; [58]) in the top panel of Figure 10. The light curve is an aggregation of observations collected by [14] and [104], which we accessed from the Open Supernova Catalog ([53]). SN 2016coi was discovered

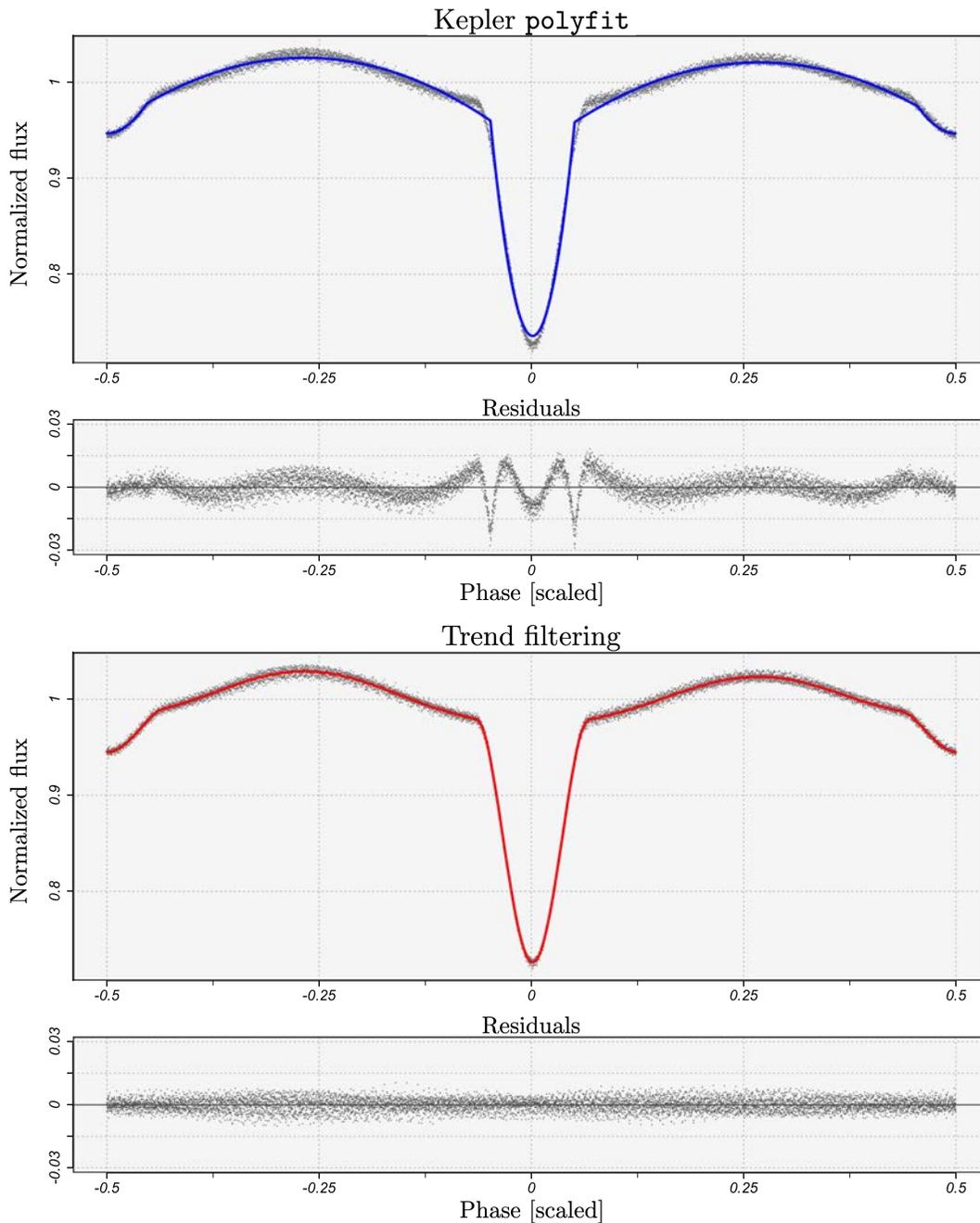


FIG 9. Comparison of the `polyfit` algorithm of [106] and our trend filtering approach for denoising phase-folded eclipsing binary light curves. The light curve shown in this example comes from the Kepler eclipsing binary system KIC 6048106 [13, 105]. **Top:** The `polyfit` algorithm fits a piecewise quadratic polynomial by weighted least-squares with four knots selected by a randomized search over the phase space. The estimate is constrained to be continuous but no constraints are enforced on the derivatives at the knots. The overly-stringent assumed model leads to significant statistical bias, which is readily apparent by examining the autocorrelation in the residuals. **Bottom:** Trend filtering is sufficiently flexible to accurately denoise the diverse set of signals observed in phase-folded eclipsing binary light curves. Here, the goodness-of-fit is clear by the random, mean-zero residual scatter.

## Trend Filtering in Astronomy

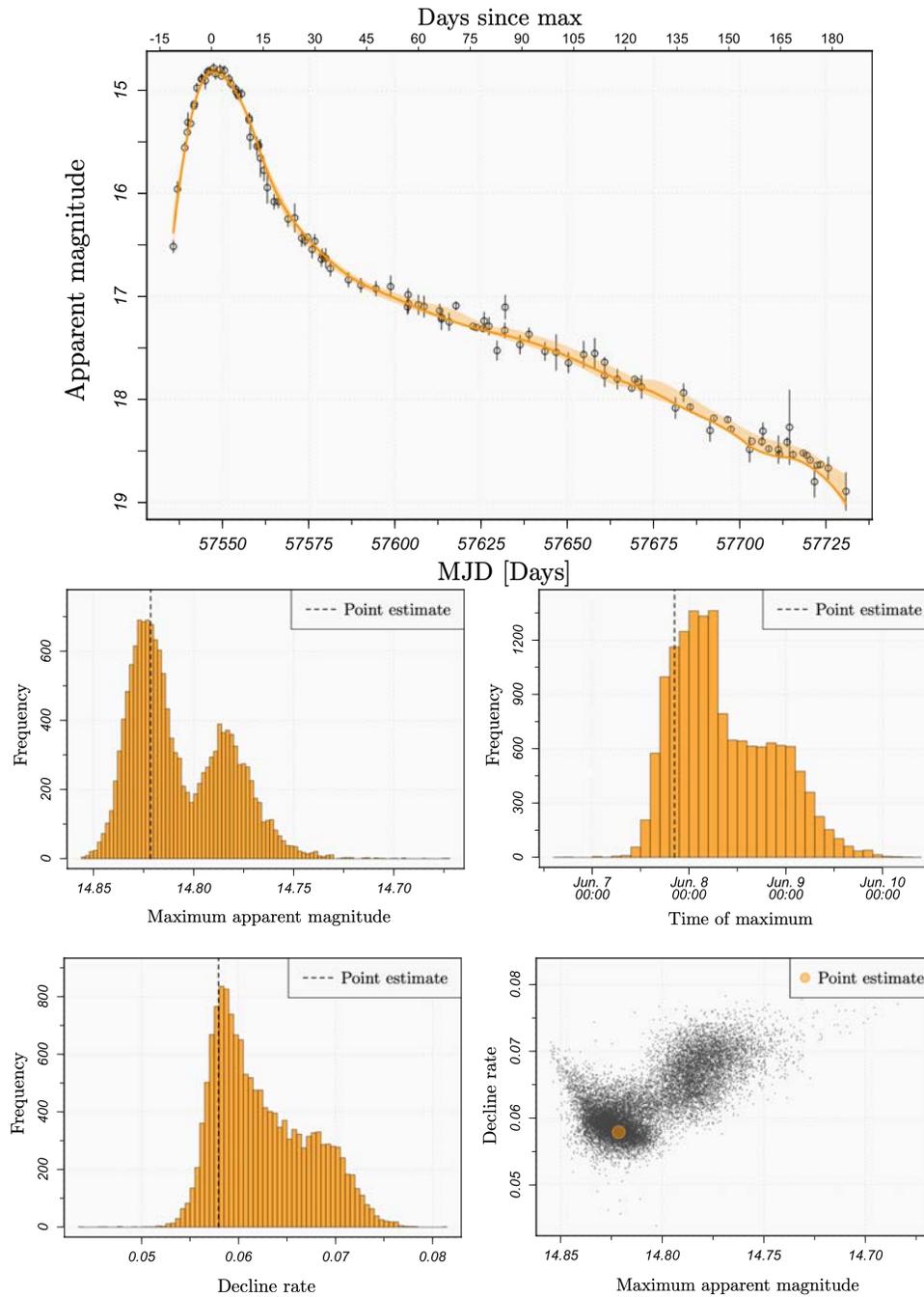


FIG 10. SN light-curve analysis (SN 2016coi). **Top:** B-band photometry of the supernova SN 2016coi (ASASSN-16 fp; [58]) discovered on May 27, 2016 by the All Sky Automated Survey for SuperNovae (ASAS-SN; [112]) in the galaxy UGC 11868 at redshift  $z \approx 0.0036$ . We fit a quadratic trend filtering estimate, tuned by  $K$ -fold cross validation, and overlay 95% nonparametric bootstrap variability bands. **Bottom:** Univariate/bivariate nonparametric bootstrap sampling distributions of the observable parameter estimates derived from the trend filtering light-curve estimate. The bimodality in the bootstrap parameter distributions arises from systematic discrepancies between the observations of the two separate observers [14, 104].

on May 27, 2016 (UT 2016-05-27.55) by the All Sky Automated Survey for SuperNovae (ASAS-SN; [112]) in the galaxy UGC 11868 at redshift  $z \approx 0.0036$ . The classification of SN 2016coi currently remains intermediate between Type Ib and Type Ic [36, 68, 104, 118]. We fit a quadratic trend filtering estimate to the observations, weighted by measurement uncertainties and with the hyperparameter selected by  $K$ -fold cross validation. The trend filtering estimate, along with 95% nonparametric bootstrap variability bands, is overlaid in the top panel of the figure. In the bottom panels we show the univariate nonparametric bootstrap sampling distributions for the estimates of the maximum apparent magnitude, the time of maximum, and the decline rate—defined as the relative change in the B-magnitude light curve from maximum light to the magnitude 15 days after the maximum [98]. We also show a bivariate bootstrap sampling distribution for maximum apparent magnitude versus decline rate. The bimodality in the bootstrap sampling distributions arises from systematic discrepancies between the two separate sets of B-band observations that form the light curve [14, 104].

*5.5. Data reduction and compression.* Although the primary focus of this paper is the use of trend filtering as a tool for astronomical data analysis, it also possesses potential utility for large-scale reduction and compression of one-dimensional data sets. This owes to several factors: its speed, flexibility, and representation as a sum of basis functions with a sparse coefficient vector.

Given a one-dimensional set of  $n$  observations  $(t_1, f(t_1)), \dots, (t_n, f(t_n)) \in (a, b) \times \mathbb{R}$ , trend filtering can quickly provide a flexible, lower-dimensional approximation of the data where the dimensionality is controlled by the choice of the hyperparameter  $\gamma$ . In this context  $\gamma$  is a subjective choice that specifies the amount of (lossy) compression desired—unrelated to the discussion in Section 3.5. For any given choice of  $\gamma$ , let  $p$  be the number of knots selected by the trend filtering estimator. The corresponding continuous-time representation of this lower-dimensional approximation is fully encoded by the bounds of the observed input interval  $(a, b)$  and the sparse basis vector with  $p + k + 1$  nonzero entries, which can be stored efficiently. The falling factorial basis then serves as the “dictionary” from which the continuous-time representation can be losslessly recovered. Gridded uncertainty measurements for the reduced observations can also be computed and stored, though not in a sparse format, via the methods discussed in Section 3.6.

**6. Concluding remarks.** The analysis of one-dimensional data arising from signals possessing varying degrees of smoothness is central to a wide variety of problems in time-domain astronomy and astronomical spectroscopy. Trend filtering is a modern statistical tool that provides a unique combination of (1) statistical optimality for estimating spatially heterogeneous signals; (2) natural flexibility for handling general sampling designs, heteroskedastic noise distributions, etc.; (3) practical computational speed, i.e. nearly linear computational complexity in the sample size; and (4) a single model hyperparameter that can be chosen via automatic data-driven methods.

In order to illustrate the broad utility of trend filtering to astronomy, we demonstrated its promising performance on a wide variety of problems across time-domain astronomy and astronomical spectroscopy. We studied the Lyman- $\alpha$  forest of quasar spectra with the most depth—using trend filtering to map the relative distribution of neutral hydrogen in the high redshift intergalactic medium along quasar-observer lines of sight. Furthermore, we discussed how trend filtering can be used to (1) generate galaxy, quasar, and stellar spectral templates and estimate emission-line parameters; (2) produce nonparametric models for exoplanet transit events in phase-folded stellar light curves, providing estimates for the transit depth and total duration; (3) improve upon the `polyfit` algorithm utilized by the Kepler Eclipsing Binary via Artificial Intelligence (EBAI) pipeline for denoising phase-folded eclipsing binary light curves (as a preliminary step to estimating the physical parameters); (4) generate

supernova light-curve templates and produce nonparametric estimates of the maximum apparent magnitude, the time of maximum, and the decline rate; (5) quickly and efficiently compress large collections of one-dimensional data sets. Naturally, we expect trend filtering will find uses in astronomy beyond those that we explicitly discussed.

Software for trend filtering is freely available online across various platforms and we provide links to our recommendations in Table 3.4. Additionally, we make supplementary R code available on the corresponding author's GitHub page<sup>11</sup> for: (1) selecting the trend filtering hyperparameter by minimizing Stein's unbiased risk estimate (see Section 3.5); and (2) various bootstrap methods for trend filtering uncertainty quantification (see Section 3.6).

## REFERENCES

- [1] AIGRAIN, S., PARVIAINEN, H. and POPE, B. J. S. (2016). K2SC: flexible systematics correction and detrending of K2 light curves using Gaussian process regression. *Monthly Notices of the Royal Astronomical Society* **459** 2408-2419.
- [2] AIHARA, H., PRIETO, C. A., AN, D., ANDERSON, S. F., AUBOURG, É., BALBINOT, E., BEERS, T. C., BERLIND, A. A., BICKERTON, S. J., BIZYAEV, D., BLANTON, M. R., BOCHANSKI, J. J., BOLTON, A. S., BOVY, J., BRANDT, W. N., BRINKMANN, J., BROWN, P. J., BROWNSTEIN, J. R., BUSCA, N. G., CAMPBELL, H., CARR, M. A., CHEN, Y., CHIAPPINI, C., COMPARAT, J., CONNOLLY, N., CORTES, M., CROFT, R. A. C., CUESTA, A. J., DA COSTA, L. N., DAVENPORT, J. R. A., DAWSON, K., DHITAL, S., EALET, A., EBELKE, G. L., EDMONDSON, E. M., EISENSTEIN, D. J., ESCOFFIER, S., ESPOSITO, M., EVANS, M. L., FAN, X., CASTELLÁ, B. F., FONT-RIBERA, A., FRINCHABOY, P. M., GE, J., GILLESPIE, B. A., GILMORE, G., HERNÁNDEZ, J. I. G., GOTT, J. R., GOULD, A., GREBEL, E. K., GUNN, J. E., HAMILTON, J.-C., HARDING, P., HARRIS, D. W., HAWLEY, S. L., HEARTY, F. R., HO, S., HOGG, D. W., HOLTZMAN, J. A., HONSCHIED, K., INADA, N., IVANS, I. I., JIANG, L., JOHNSON, J. A., JORDAN, C., JORDAN, W. P., KAZIN, E. A., KIRKBY, D., KLAENE, M. A., KNAPP, G. R., KNEIB, J.-P., KOCHANÉK, C. S., KOESTERKE, L., KOLLMEIER, J. A., KRON, R. G., LAMPEITL, H., LANG, D., GOFF, J.-M. L., LEE, Y. S., LIN, Y.-T., LONG, D. C., LOOMIS, C. P., LUCATELLO, S., LUNDGREN, B., LUPTON, R. H., MA, Z., MACDONALD, N., MAHADEVAN, S., MAIA, M. A. G., MAKLER, M., MALANUSHENKO, E., MALANUSHENKO, V., MANDELBAUM, R., MARASTON, C., MARGALA, D., MASTERS, K. L., MCBRIDE, C. K., MCGEHEE, P. M., MCGREER, I. D., MÉNARD, B., MIRALDA-ESCUDE, J., MORRISON, H. L., MULLALLY, F., MUNA, D., MUNN, J. A., MURAYAMA, H., MYERS, A. D., NAUGLE, T., NETO, A. F., NGUYEN, D. C., NICHOL, R. C., O'CONNELL, R. W., OGANDO, R. L. C., OLMSTEAD, M. D., ORAVETZ, D. J., PADMANABHAN, N., PALANQUE-DELABROUILLE, N., PAN, K., PANDEY, P., PÁRIS, I., PERCIVAL, W. J., PETITJEAN, P., PFAFFENBERGER, R., PFORR, J., PHLEPS, S., PICHON, C., PIERI, M. M., PRADA, F., PRICE-WHELAN, A. M., RADDICK, M. J., RAMOS, B. H. F., REYLÉ, C., RICH, J., RICHARDS, G. T., RIX, H.-W., ROBIN, A. C., ROCHA-PINTO, H. J., ROCKOSI, C. M., ROE, N. A., ROLLINDE, E., ROSS, A. J., ROSS, N. P., ROSSETTO, B. M., SÁNCHEZ, A. G., SAYRES, C., SCHLEGEL, D. J., SCHLESINGER, K. J., SCHMIDT, S. J., SCHNEIDER, D. P., SHELDON, E., SHU, Y., SIMMERER, J., SIMMONS, A. E., SIVARANI, T., SNEDDEN, S. A., SOBECK, J. S., STEINMETZ, M., STRAUSS, M. A., SZALAY, A. S., TANAKA, M., THAKAR, A. R., THOMAS, D., TINKER, J. L., TOFFLEMIRE, B. M., TOJEIRO, R., TREMONTI, C. A., VANDENBERG, J., MAGAÑA, M. V., VERDE, L., VOGT, N. P., WAKE, D. A., WANG, J., WEAVER, B. A., WEINBERG, D. H., WHITE, M., WHITE, S. D. M., YANNY, B., YASUDA, N., YECHE, C. and ZEHAVI, I. (2011). THE EIGHTH DATA RELEASE OF THE SLOAN DIGITAL SKY SURVEY: FIRST DATA FROM SDSS-III. *The Astrophysical Journal Supplement Series* **193** 29.
- [3] AKESON, R. L., CHEN, X., CIARDI, D., CRANE, M., GOOD, J., HARBUT, M., JACKSON, E., KANE, S. R., LAITY, A. C., LEIFER, S. and ET AL. (2013). The NASA Exoplanet Archive: Data and Tools for Exoplanet Research. *Publications of the Astronomical Society of the Pacific* **125** 989-999.
- [4] ALAM, S. et al. (2015). The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. *The Astrophysical Journal Supplement Series* **219**.
- [5] ARNOLD, T. B., SADHANALA, V. and TIBSHIRANI, R. J. (2014). Fast algorithms for generalized lasso problems.

<sup>11</sup><https://github.com/capolitsch/trend-filtering>

- [6] ARNOLD, T. B. and TIBSHIRANI, R. J. (2016). Efficient implementations of the generalized lasso dual path algorithm. *Journal of Computational and Graphical Statistics* **25** 1-27.
- [7] BATALHA, N. M., BORUCKI, W. J., BRYSON, S. T., BUCHHAVE, L. A., CALDWELL, D. A., CHRISTENSEN-DALSGAARD, J., CIARDI, D., DUNHAM, E. W., FRESSIN, F., GAUTIER, T. N., GILLILAND, R. L., HAAS, M. R., HOWELL, S. B., JENKINS, J. M., KJELDSSEN, H., KOCH, D. G., LATHAM, D. W., LISSAUER, J. J., MARCY, G. W., ROWE, J. F., SASSELOV, D. D., SEAGER, S., STEFFEN, J. H., TORRES, G., BASRI, G. S., BROWN, T. M., CHARBONNEAU, D., CHRISTIANSEN, J., CLARKE, B., COCHRAN, W. D., DUPREE, A., FABRYCKY, D. C., FISCHER, D., FORD, E. B., FORTNEY, J., GIROUARD, F. R., HOLMAN, M. J., JOHNSON, J., ISAACSON, H., KLAUS, T. C., MACHALEK, P., MOOREHEAD, A. V., MOREHEAD, R. C., RAGOZZINE, D., TENENBAUM, P., TWICKEN, J., QUINN, S., VANCLEVE, J., WALKOWICZ, L. M., WELSH, W. F., DEVORE, E. and GOULD, A. (2011). KEPLER'S FIRST ROCKY PLANET: KEPLER-10b. *The Astrophysical Journal* **729** 27.
- [8] BAUTISTA, J. E. et al. (2015). Mock Quasar-Lyman- $\alpha$  Forest Data-sets for the SDSS-III Baryon Oscillation Spectroscopic Survey. *Journal of Cosmology and Astroparticle Physics* **1505**.
- [9] BECKER, G. D., HEWETT, P. C., WORSECK, G. and PROCHASKA, J. X. (2013). A Refined Measurement of the Mean Transmitted Flux in the Ly-alpha Forest over  $2 < z < 5$  Using Composite Quasar Spectra. *Monthly Notices of the Royal Astronomical Society* **430** 2067-2081.
- [10] BELOKUROV, V., EVANS, N. W. and LE DU, Y. (2004). Light-curve classification in massive variability surveys – II. Transients towards the Large Magellanic Cloud. *Monthly Notices of the Royal Astronomical Society* **352** 233-242.
- [11] BERNARDI, M., SHETH, R. K., SUBBARAO, M., RICHARDS, G. T., BURLES, S., CONNOLLY, A. J., FRIEMAN, J., NICHOL, R., SCHAYE, J., SCHNEIDER, D. P., BERK, D. E. V., YORK, D. G., BRINKMANN, J. and LAMB, D. Q. (2003). A Feature at  $z \sim 3.2$  in the Evolution of the Ly $\alpha$  Forest Optical Depth. *The Astronomical Journal* **125** 32–52.
- [12] BOLTON, A. S., SCHLEGEL, D. J., AUBOURG, É., BAILEY, S., BHARDWAJ, V., BROWNSTEIN, J. R., BURLES, S., CHEN, Y.-M., DAWSON, K., EISENSTEIN, D. J., GUNN, J. E., KNAPP, G. R., LOOMIS, C. P., LUPTON, R. H., MARASTON, C., MUNA, D., MYERS, A. D., OLMSTEAD, M. D., PADMANABHAN, N., PÀRIS, I., PERCIVAL, W. J., PETITJEAN, P., ROCKOSI, C. M., ROSS, N. P., SCHNEIDER, D. P., SHU, Y., STRAUSS, M. A., THOMAS, D., TREMONTI, C. A., WAKE, D. A., WEAVER, B. A. and WOOD-VASEY, W. M. (2012). SPECTRAL CLASSIFICATION AND REDSHIFT MEASUREMENT FOR THE SDSS-III BARYON OSCILLATION SPECTROSCOPIC SURVEY. *The Astronomical Journal* **144** 144.
- [13] BORUCKI, W. J. et al. (2010). Kepler Planet-Detection Mission: Introduction and First Results. *Science* **327** 977-980.
- [14] BROWN, P. J., BREEVELD, A. A., HOLLAND, S., KUIN, P. and PRITCHARD, T. (2014). SOUSA: the Swift Optical/Ultraviolet Supernova Archive. *Astrophysics and Space Science* **354** 89-96.
- [15] BURNS, C. R., STRITZINGER, M., PHILLIPS, M. M., HSIAO, E. Y., CONTRERAS, C., PERSSON, S. E., FOLATELLI, G., BOLDT, L., CAMPILLAY, A., CASTELLÓN, S., FREEDMAN, W. L., MADORE, B. F., MORRELL, N., SALGADO, F. and SUNTZEFF, N. B. (2014). The Carnegie Supernova Project: Intrinsic Colors of Type Ia Supernovae. *The Astrophysical Journal* **789** 32.
- [16] BURNS, C. R., STRITZINGER, M., PHILLIPS, M. M., KATTNER, S., PERSSON, S. E., MADORE, B. F., FREEDMAN, W. L., BOLDT, L., CAMPILLAY, A., CONTRERAS, C., FOLATELLI, G., GONZÁLEZ, S., KRZEMINSKI, W., MORRELL, N., SALGADO, F. and SUNTZEFF, N. B. (2010). THE CARNEGIE SUPERNOVA PROJECT: LIGHT-CURVE FITTING WITH SNooPy. *The Astronomical Journal* **141** 19.
- [17] CARDELLI, J., CLAYTON, G. and MATHIS, J. (1989). The Relationship Between Infrared, Optical, and Ultraviolet Extinction. *The Astrophysical Journal* **345** 245-256.
- [18] CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74** 829-836.
- [19] COCHRAN, W. G. (1952). The  $\chi^2$  Test of Goodness of Fit. *The Annals of Mathematical Statistics* **23** 315–345.
- [20] CONTRERAS, C., HAMUY, M., PHILLIPS, M. M., FOLATELLI, G., SUNTZEFF, N. B., PERSSON, S. E., STRITZINGER, M., BOLDT, L., GONZÁLEZ, S., KRZEMINSKI, W., MORRELL, N., ROTH, M., SALGADO, F., MAUREIRA, M. J., BURNS, C. R., FREEDMAN, W. L., MADORE, B. F., MURPHY, D., WYATT, P., LI, W. and FILIPPENKO, A. V. (2010). THE CARNEGIE SUPERNOVA PROJECT: FIRST PHOTOMETRY DATA RELEASE OF LOW-REDSHIFT TYPE Ia SUPERNOVAE. *The Astronomical Journal* **139** 519–539.
- [21] CROFT, R. A. C. et al. (2002). TOWARDS A PRECISE MEASUREMENT OF MATTER CLUSTERING: LYMAN-ALPHA FOREST DATA AT REDSHIFTS 2-4. *The Astrophysical Journal* **581**.

- [22] CROFT, R. A. C., WEINBERG, D. H., KATZ, N. and HERNQUIST, L. (1997). Intergalactic Helium Absorption in Cold Dark Matter Models. *The Astrophysical Journal* **488** 532-549.
- [23] DALL'AGLIO, A., WISOTZKI, L. and WORSECK, G. (2009). The UV background photoionization rate at  $2.3 < z < 4.6$  as measured from the Sloan Digital Sky Survey.
- [24] DE BOOR, C. (1974). Good approximation by splines with variable knots. II. In *Conference on the numerical solution of differential equations* 12–20. Springer.
- [25] DE BOOR, C. (1978). A Practical Guide to Splines. In *Applied Mathematical Sciences*.
- [26] DHAWAN, S., LEIBUNDGUT, B., SPYROMILIO, J. and MAGUIRE, K. (2015). Near-infrared light curves of Type Ia supernovae: studying properties of the second maximum. *Monthly Notices of the Royal Astronomical Society* **448** 1345-1359.
- [27] DIAMOND, S. and BOYD, S. (2016). CVXPY: A Python-Embedded Modeling Language for Convex Optimization. *Journal of Machine Learning Research* **17** 1–5.
- [28] DIMATTEO, I., GENOVESE, C. R. and KASS, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88** 1055-1071.
- [29] DIMITRIADIS, G., SULLIVAN, M., KERZENDORF, W., RUITER, A. J., SEITENZAHL, I. R., TAUBENBERGER, S., DORAN, G. B., GAL-YAM, A., LAHER, R. R., MAGUIRE, K., NUGENT, P., OFEK, E. O. and SURACE, J. (2017). The late-time light curve of the Type Ia supernova SN 2011fe. *Monthly Notices of the Royal Astronomical Society* **468** 3798-3812.
- [30] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Minimax risk over  $\ell_1$ -balls for  $\ell_2$ -error. *Probability Theory and Related Fields* **99** 277-303.
- [31] DONOHO, D. L. and JOHNSTONE, I. M. (1998). Minimax estimation via wavelet shrinkage. *The Annals of Statistics* **26** 879-921.
- [32] EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7** 1–26.
- [33] EFRON, B. (1986). How Biased is the Apparent Error Rate of a Prediction Rule? *Journal of the American Statistical Association* **81** 461-470.
- [34] EFRON, B. and TIBSHIRANI, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* **1** 54-75.
- [35] EISENSTEIN, D. J., WEINBERG, D. H., AGOL, E., AIHARA, H., PRIETO, C. A., ANDERSON, S. F., ARNS, J. A., AUBOURG, É., BAILEY, S., BALBINOT, E., BARKHOUSER, R., BEERS, T. C., BERLIND, A. A., BICKERTON, S. J., BIZYAIEV, D., BLANTON, M. R., BOCHANSKI, J. J., BOLTON, A. S., BOSMAN, C. T., BOVY, J., BRANDT, W. N., BRESLAUER, B., BREWINGTON, H. J., BRINKMANN, J., BROWN, P. J., BROWNSTEIN, J. R., BURGER, D., BUSCA, N. G., CAMPBELL, H., CARGILE, P. A., CARITHERS, W. C., CARLBERG, J. K., CARR, M. A., CHANG, L., CHEN, Y., CHIAPPINI, C., COMPARAT, J., CONNOLLY, N., CORTES, M., CROFT, R. A. C., CUNHA, K., DA COSTA, L. N., DAVENPORT, J. R. A., DAWSON, K., LEE, N. D., DE MELLO, G. F. P., DE SIMONI, F., DEAN, J., DHITAL, S., EALET, A., EBELKE, G. L., EDMONDSON, E. M., EITING, J. M., ESCOFFIER, S., ESPOSITO, M., EVANS, M. L., FAN, X., CASTELLÁ, B. F., FERREIRA, L. D., FITZGERALD, G., FLEMING, S. W., FONT-RIBERA, A., FORD, E. B., FRINCHABOY, P. M., PÉREZ, A. E. G., GAUDI, B. S., GE, J., GHEZZI, L., GILLESPIE, B. A., GILMORE, G., GIRARDI, L., GOTT, J. R., GOULD, A., GREBEL, E. K., GUNN, J. E., HAMILTON, J.-C., HARDING, P., HARRIS, D. W., HAWLEY, S. L., HEARTY, F. R., HENNAWI, J. F., HERNÁNDEZ, J. I. G., HO, S., HOGG, D. W., HOLTZMAN, J. A., HONSCHEID, K., INADA, N., IVANS, I. I., JIANG, L., JIANG, P., JOHNSON, J. A., JORDAN, C., JORDAN, W. P., KAUFFMANN, G., KAZIN, E., KIRKBY, D., KLAENE, M. A., KNAPP, G. R., KNEIB, J.-P., KOCHANÉK, C. S., KOESTERKE, L., KOLLMEIER, J. A., KRON, R. G., LAMPEITL, H., LANG, D., LAWLER, J. E., GOFF, J.-M. L., LEE, B. L., LEE, Y. S., LEISENRING, J. M., LIN, Y.-T., LIU, J., LONG, D. C., LOOMIS, C. P., LUCATELLO, S., LUNDGREN, B., LUP-TON, R. H., MA, B., MA, Z., MACDONALD, N., MACK, C., MAHADEVAN, S., MAIA, M. A. G., MAJEWSKI, S. R., MAKLER, M., MALANUSHENKO, E., MALANUSHENKO, V., MANDELBAUM, R., MARASTON, C., MARGALA, D., MASEMAN, P., MASTERS, K. L., MCBRIDE, C. K., MCDONALD, P., MCGREER, I. D., MCMAHON, R. G., REQUEJO, O. M., MÉNARD, B., MIRALDA-ESCUDE, J., MORRISON, H. L., MULLALLY, F., MUNA, D., MURAYAMA, H., MYERS, A. D., NAUGLE, T., NETO, A. F., NGUYEN, D. C., NICHOL, R. C., NIDEVER, D. L., O'CONNELL, R. W., OGANDO, R. L. C., OLMSTEAD, M. D., ORAVETZ, D. J., PADMANABHAN, N., PAEGERT, M., PALANQUE-DELABROUILLE, N., PAN, K., PANDEY, P., PAREJKO, J. K., PÁRIS, I., PELLEGRINI, P., PEPPER, J., PERCIVAL, W. J., PETITJEAN, P., PFAFFENBERGER, R., PFORR, J., PHLEPS, S., PICHON, C., PIERI, M. M., PRADA, F., PRICE-WHELAN, A. M., RADDICK, M. J., RAMOS, B. H. F., REID, I. N., REYLE, C., RICH, J., RICHARDS, G. T., RIEKE, G. H., RIEKE, M. J., RIX, H.-W., ROBIN, A. C., ROCHA-PINTO, H. J., ROCKOSI, C. M.,

- ROE, N. A., ROLLINDE, E., ROSS, A. J., ROSS, N. P., ROSSETTO, B., SÁNCHEZ, A. G., SANTIAGO, B., SAYRES, C., SCHIAVON, R., SCHLEGEL, D. J., SCHLESINGER, K. J., SCHMIDT, S. J., SCHNEIDER, D. P., SELLGREN, K., SHELDON, A., SHELDON, E., SHETRONE, M., SHU, Y., SILVERMAN, J. D., SIMMERER, J., SIMMONS, A. E., SIVARANI, T., SKRUTSKIE, M. F., SLOSAR, A., SMEE, S., SMITH, V. V., SNEDDEN, S. A., STASSUN, K. G., STEELE, O., STEINMETZ, M., STOCKETT, M. H., STOLLBERG, T., STRAUSS, M. A., SZALAY, A. S., TANAKA, M., THAKAR, A. R., THOMAS, D., TINKER, J. L., TOFFLEMIRE, B. M., TOJEIRO, R., TREMONTI, C. A., MAGAÑA, M. V., VERDE, L., VOGT, N. P., WAKE, D. A., WAN, X., WANG, J., WEAVER, B. A., WHITE, M., WHITE, S. D. M., WILSON, J. C., WISNIEWSKI, J. P., WOODVASEY, W. M., YANNY, B., YASUDA, N., YÈCHE, C., YORK, D. G., YOUNG, E., ZASOWSKI, G., ZEHAVI, I. and ZHAO, B. (2011). SDSS-III: MASSIVE SPECTROSCOPIC SURVEYS OF THE DISTANT UNIVERSE, THE MILKY WAY, AND EXTRA-SOLAR PLANETARY SYSTEMS. *The Astronomical Journal* **142** 72.
- [36] ELIAS-ROSA, N., MATTILA, S., LUNDQVIST, P., STRITZINGER, M., KUNCARAYAKTI, H., HARMANNEN, J., PASTORELLO, A., BENETTI, S., CAPPELLARO, E., BLAGORODNOVA, N., DAVIS, S., DONG, S., FRASER, M., GALL, C., HARRISON, D., HODGKIN, S., HSIAO, E. Y., JONKER, P., KANGAS, T., KANKARE, E., KOSTRZEWA-RUTKOWSKA, Z., NIELSEN, M., OCHNER, P., PRIETO, J. L., REYNOLDS, T., ROMERO-CANIZALES, C., TADDIA, F., TARTAGLIA, L., TERRERAN, G., TOMASELLA, L. and WYRZYKOWSKI, L. (2016). Spectroscopic classification of ASASSN-16fp with the Nordic Optical Telescope. *The Astronomer's Telegram* **9090** 1.
- [37] EPANECHNIKOV, V. A. (1967). Non-Parametric Estimation of a Multivariate Probability Density. *Theory of Probability & its Applications* **14** 153-158.
- [38] FAN, J. (1993). Local Linear Regression Smoothers and Their Minimax Efficiencies. *The Annals of Statistics* **21** 196–216.
- [39] FAN, J., GASSER, T., GIJBELS, I., BROCKMANN, M. and ENGEL, J. (1997). Local Polynomial Regression: Optimal Kernels and Asymptotic Minimax Efficiency. *Annals of the Institute of Statistical Mathematics* **49** 79-99.
- [40] FAN, J. and GIJBELS, I. (1992). Variable bandwidth and local linear regression smoother. *The Annals of Statistics* **20** 2008-2036.
- [41] FAN, J. and GIJBELS, I. (1995). Data-Driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation. *Journal of the Royal Statistical Society. Series B (Methodological)* **57** 371–394.
- [42] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman and Hall.
- [43] FAUCHER-GIGUÈRE, C.-A., LIDZ, A., HERNQUIST, L. and ZALDARRIAGA, M. (2008). Evolution of the Intergalactic Opacity: Implications for the Ionizing Background, Cosmic Star Formation, and Quasar Activity. *The Astrophysical Journal* **688** 85–107.
- [44] FAULKNER, J. R. and MININ, V. N. (2018). Locally Adaptive Smoothing with Markov Random Fields and Shrinkage Priors. *Bayesian Analysis* **13** 225–252.
- [45] FIGUEIREDO, M. A. T. (2003). Adaptive Sparseness for Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25** 1150-1159.
- [46] FLIGGE, M. and SOLANKI, S. K. (1997). Noise reduction in astronomical spectra using wavelet packets. *A&AS* **124** 579-587.
- [47] FRESSIN, F., TORRES, G., DÉSSERT, J.-M., CHARBONNEAU, D., BATALHA, N. M., FORTNEY, J. J., ROWE, J. F., ALLEN, C., BORUCKI, W. J., BROWN, T. M., BRYSON, S. T., CIARDI, D. R., COCHRAN, W. D., DEMING, D., DUNHAM, E. W., FABRYCKY, D. C., III, T. N. G., GILLILAND, R. L., HENZE, C. E., HOLMAN, M. J., HOWELL, S. B., JENKINS, J. M., KINEMUCHI, K., KNUTSON, H., KOCH, D. G., LATHAM, D. W., LISSAUER, J. J., MARCY, G. W., RAGOZZINE, D., SASSELOV, D. D., STILL, M., TENENBAUM, P. and UDDIN, K. (2011). KEPLER-10 c: A 2.2 EARTH RADIUS TRANSITING PLANET IN A MULTIPLE SYSTEM. *The Astrophysical Journal Supplement Series* **197** 5.
- [48] GIBSON, N. P., AIGRAIN, S., ROBERTS, S., EVANS, T. M., OSBORNE, M. and PONT, F. (2012). A Gaussian process framework for modelling instrumental systematics: application to transmission spectroscopy. *Monthly Notices of the Royal Astronomical Society* **419** 2683-2694.
- [49] GIJBELS, I. and MAMMEN, E. (1998). Local Adaptivity of Kernel Estimates with Plug-in Local Bandwidth Selectors. *Scandinavian Journal of Statistics* **25** 503–520.
- [50] GNEDIN, N. Y. and HUI, L. (1998). Probing the Universe with the Ly $\alpha$  forest - I. Hydrodynamics of the low-density intergalactic medium. *Monthly Notices of the Royal Astronomical Society* **296** 44-55.
- [51] GOLKHOV, V. Z. and BUTLER, N. R. (2014). Uncovering the Intrinsic Variability of Gamma-Ray Bursts. *The Astrophysical Journal* **787** 90.

- [52] GÓMEZ-VALENT, A. and AMENDOLA, L. (2018). H0 from cosmic chronometers and Type Ia supernovae, with Gaussian Processes and the novel Weighted Polynomial Regression method. *Journal of Cosmology and Astroparticle Physics* **2018** 051–051.
- [53] GUILLOCHON, J., PARRENT, J., KELLEY, L. Z. and MARGUTTI, R. (2017). An Open Catalog for Supernova Data. *The Astrophysical Journal* **835** 64.
- [54] GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). A Distribution-Free Theory of Non-parametric Regression. In *Springer Series in Statistics*.
- [55] HALL, P. B., F. ANDERSON, S., STRAUSS, M., YORK, D., T. RICHARDS, G., FAN, X., R. KNAPP, G., P. SCHNEIDER, D., E. VANDEN BERK, D., GEBALLE, T., BAUER, A., H. BECKER, R., DAVIS, M., RIX, H.-W., C. NICHOL, R., BAHCALL, N., BRINKMANN, J., BRUNNER, R., J. CONNOLLY, A. and ZHENG, W. (2002). Unusual Broad Absorption Line Quasars from the Sloan Digital Sky Survey. *The Astrophysical Journal Supplement Series* **141**.
- [56] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2 ed. Springer.
- [57] HELLSTEN, U., HERNQUIST, L., KATZ, N. and WEINBERG, D. H. (1998). The Observability of Metal Lines Associated with the Ly $\alpha$  Forest. *The Astrophysical Journal* **499** 172–180.
- [58] HOLOIEN, T. W. S., STANEK, K. Z., BROWN, J. S., KOCHANEK, C. S., GODOY-RIVERA, D., BASU, U., SHAPPEE, B. J., PRIETO, J. L., BERSIER, D., DONG, S., CHEN, P. and BRIMACOMBE, J. (2016). ASASSN-16fp: Discovery of A Probable Supernova in UGC 11868. *The Astronomer's Telegram* **9086**.
- [59] HOWELL, D. A., SULLIVAN, M., PERRETT, K., BRONDER, T. J., HOOK, I. M., ASTIER, P., AUBOURG, E., BALAM, D., BASA, S., CARLBERG, R. G., FABBRO, S., FOCHEZ, D., GUY, J., LAFoux, H., NEILL, J. D., PAIN, R., PALANQUE-DELABROUILLE, N., PRITCHET, C. J., REGNAULT, N., RICH, J., TAILLET, R., KNOP, R., MCMAHON, R. G., PERLMUTTER, S. and WALTON, N. A. (2005). Gemini Spectroscopy of Supernovae from the Supernova Legacy Survey: Improving High-Redshift Supernova Selection and Classification. *The Astrophysical Journal* **634** 1190–1201.
- [60] IBRAGIMOV, I. A. and HASMINISKII, R. Z. (1980). Asymptotic efficiency bounds for non-parametric estimation of a regression function in  $L_p$ . *Zapiski Nauchnykh Seminarov LOMI (in Russian)* **97** 88–101.
- [61] JENKINS, J. M., CALDWELL, D. A., CHANDRASEKARAN, H., TWICKEN, J. D., BRYSON, S. T., QUINTANA, E. V., CLARKE, B. D., LI, J., ALLEN, C., TENENBAUM, P., WU, H., KLAUS, T. C., MIDDOUR, C. K., COTE, M. T., MCCAULIFF, S., GIROUARD, F. R., GUNTER, J. P., WOHLER, B., SOMMERS, J., HALL, J. R., UDDIN, A. K., WU, M. S., BHAVSAR, P. A., CLEVE, J. V., PLETCHER, D. L., DOTSON, J. A., HAAS, M. R., GILLILAND, R. L., KOCH, D. G. and BORUCKI, W. J. (2010). OVERVIEW OF THE KEPLER SCIENCE PROCESSING PIPELINE. *The Astrophysical Journal Letters* **713** L87–L91.
- [62] JUPP, D. L. (1978). Approximation to data by splines with free knots. *SIAM Journal on Numerical Analysis* **15** 328–343.
- [63] KIM, S.-J. et al. (2009).  $\ell_1$  Trend Filtering. *SIAM Review* **51** 339–360.
- [64] KIM, T. S., VIEL, M., HAEHNELT, M. G., CARSWELL, R. F. and CRISTIANI, S. (2004). The power spectrum of the flux distribution in the Lyman $\alpha$  forest of a large sample of UVES QSO absorption spectra (LUQAS). *Monthly Notices of the Royal Astronomical Society* **347** 355–366.
- [65] KOH, K., KIM, S.-J. and BOYD, S. (2008).  $\ell_1$ - $\ell_2$ : Software for  $\ell_1$  Trend Filtering.
- [66] KORNBLITH, S. (2014). Lasso/Elastic Net linear and generalized linear models.
- [67] KOVÁCS, G., BAKOS, G. and NOYES, R. W. (2005). A trend filtering algorithm for wide-field variability surveys. *Monthly Notices of the Royal Astronomical Society* **356** 557–567.
- [68] KUMAR, B., SINGH, A., SRIVASTAV, S., SAHU, D. K. and ANUPAMA, G. C. (2018). ASASSN-16fp (SN 2016coi): a transitional supernova between Type Ic and broad-lined Ic. *Monthly Notices of the Royal Astronomical Society* **473** 3776–3788.
- [69] KUNSCH, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics* **17** 1217–1241.
- [70] LAND, S. and FRIEDMAN, J. (1996). Variable fusion: a new method of adaptive signal regression Technical Report, Department of Statistics, Stanford University.
- [71] LEE, K.-G., HENNAWI, J. F., STARK, C., PROCHASKA, J. X., WHITE, M., SCHLEGEL, D. J., EILERS, A.-C., ARINYO-I-PRATS, A., SUZUKI, N., CROFT, R. A. C., CAPUTI, K. I., CASSATA, P., ILBERT, O., GARILLI, B., KOEKEMOER, A. M., LE BRUN, V., LE FÈVRE, O., MACCAGNI, D., NUGENT, P., TANIGUCHI, Y., TASCIA, L. A. M., TRESSE, L., ZAMORANI, G. and ZUCCA, E. (2014). Ly $\alpha$  Forest Tomography from Background Galaxies: The First Megaparsec-resolution Large-scale Structure Map at  $z \approx 2$ . *The Astrophysical Journal Letters* **795** L12.

- [72] LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics* **25** 929–947.
- [73] LIU, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *The Annals of Statistics* **16** 1696–1708.
- [74] LOADER, C. (1999). *Local Regression and Likelihood*. Springer-Verlag.
- [75] MAMMEN, E. (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *The Annals of Statistics* **21** 255–285.
- [76] MANDEL, K. and AGOL, E. (2002). Analytic Light Curves for Planetary Transit Searches. *The Astrophysical Journal* **580** L171–L175.
- [77] MARON, J. L. and HOWES, G. G. (2003). Gradient Particle Magnetohydrodynamics: A Lagrangian Particle Code for Astrophysical Magnetohydrodynamics. *The Astrophysical Journal* **595** 564–572.
- [78] MARZIANI, P., SULENTIC, J. W., DULTZIN-HACYAN, D., CALVANI, M. and MOLES, M. (1996). Comparative Analysis of the High- and Low-Ionization Lines in the Broad-Line Region of Active Galactic Nuclei. *The Astrophysical Journal Supplement Series* **104** 37.
- [79] MATIJEVIĆ, G., PRŠA, A., OROSZ, J. A., WELSH, W. F., BLOEMEN, S. and BARCLAY, T. (2012). KEPLERRECLIPSING BINARY STARS. III. CLASSIFICATION OF KEPLERRECLIPSING BINARY LIGHT CURVES WITH LOCALLY LINEAR EMBEDDING. *The Astronomical Journal* **143** 123.
- [80] McDONALD, P. et al. (2005). The Linear theory power spectrum from the Lyman-alpha forest in the Sloan Digital Sky Survey. *The Astrophysical Journal* **635** 761–783.
- [81] MEINSHAUSEN, N. (2007). Relaxed Lasso. *Computational Statistics & Data Analysis* **52** 374 - 393.
- [82] MULLER, H.-G. and STADTMULLER, U. (1987). Estimation of Heteroscedasticity in Regression Analysis. *The Annals of Statistics* **15** 610–625.
- [83] NADARAYA, E. A. (1964). On estimating regression. *Theory of Probability & its Applications* **9** 141–142.
- [84] NEMIROVSKII, A. (1985). Nonparametric estimation of smooth regression function. *Izv. Akad. Nauk. SSSR Tekhn. Kibernet. (in Russian)* **3** 50–60.
- [85] NEMIROVSKII, A., POLYAK, B. and TSYBAKOV, A. (1985). Rate of convergence of non-parametric estimates of maximum-likelihood type. *Problems of Information Transmission* **21**.
- [86] NUSSBAUM, M. (1985). Spline Smoothing in Regression Models and Asymptotic Efficiency in L<sub>2</sub>. *The Annals of Statistics* **13** 984–997.
- [87] OSTERBROCK, D. E. and FERLAND, G. J. (2006). *Astrophysics of gaseous nebulae and active galactic nuclei*.
- [88] PACIOREK, C. J. and SCHERVISH, M. J. (2004). Nonstationary covariance functions for Gaussian process regression. In *Advances in neural information processing systems* 273–280.
- [89] PACIOREK, C. J. and SCHERVISH, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17** 483–506.
- [90] PALANQUE-DELABROUILLE, N., YÈCHE, C., BAUR, J., MAGNEVILLE, C., ROSSI, G., LESGOURGUES, J., BORDE, A., BURTIN, E., LEGOFF, J.-M., RICH, J., VIEL, M. and WEINBERG, D. (2015). Neutrino masses and cosmology with Lyman-alpha forest power spectrum. *Journal of Cosmology and Astroparticle Physics* **2015** 011.
- [91] PALANQUE-DELABROUILLE, N., YÈCHE, CHRISTOPHE, BORDE, ARNAUD, LE GOFF, JEAN-MARC, ROSSI, GRAZIANO, VIEL, MATTEO, AUBOURG, ÉRIC, BAILEY, STEPHEN, BAUTISTA, JULIAN, BLOMQVIST, MICHAEL, BOLTON, ADAM, BOLTON, JAMES S. , BUSCA, NICOLÁS G. , CARITHERS, BILL, CROFT, RUPERT A. C. , DAWSON, KYLE S. , DELUBAC, TIMOTHÉE, FONT-RIBERA, ANDREU, HO, SHIRLEY, KIRKBY, DAVID, LEE, KHEE-GAN, MARGALA, DANIEL, MIRALDA-ESCUDE, JORDI, MUNA, DEMITRI, MYERS, ADAM D. , NOTERDAEME, PASQUIER, PÂRIS, ISABELLE, PETITJEAN, PATRICK, PIERI, MATTHEW M. , RICH, JAMES, ROLLINDE, EMANUEL, ROSS, NICHOLAS P. , SCHLEGEL, DAVID J. , SCHNEIDER, DONALD P. , SLOSAR, ANZE and WEINBERG, DAVID H. (2013). The one-dimensional Lyest power spectrum from BOSS. *A&A* **559** A85.
- [92] PÂRIS, I., PETITJEAN, P., ROLLINDE, E., AUBOURG, E., BUSCA, N. G., CHARLASSIER, R., DELUBAC, T., JEAN-CHRISTOPHE, H., M LE GOFF, J., PALANQUE-DELABROUILLE, N., PEIRANI, S., PICHON, C., RICH, J., MAGANA, M. and YÈCHE, C. (2011). A Principal Component Analysis of quasar UV spectra at z 3. *A&A* **530**.
- [93] PARK, T. and CASELLA, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* **103** 681–686.
- [94] PEARSON, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **50** 157–175.

- [95] PEEPLES, M. S., WEINBERG, D. H., DAVÉ, R., FARDAL, M. A. and KATZ, N. (2010). Pressure support versus thermal broadening in the Lyman  $\alpha$  forest - I. Effects of the equation of state on longitudinal structure. *Monthly Notices of the Royal Astronomical Society* **404** 1281-1294.
- [96] PEIRIS, H. V. and VERDE, L. (2010). The Shape of the Primordial Power Spectrum: A Last Stand Before Planck. *Phys. Rev.* **D81** 021302.
- [97] PERSSON, S. E., MADORE, B. F., KRZEMIŃSKI, W., FREEDMAN, W. L., ROTH, M. and MURPHY, D. C. (2004). New Cepheid Period-Luminosity Relations for the Large Magellanic Cloud: 92 Near-Infrared Light Curves. *The Astronomical Journal* **128** 2239-2264.
- [98] PHILLIPS, M. M. (1993). The Absolute Magnitudes of Type IA Supernovae. *The Astrophysical Journal Letters* **413** L105.
- [99] PIERI, M. M., FRANK, S., WEINBERG, D. H., MATHUR, S. and YORK, D. G. (2010). The Composite Spectrum of Strong Ly $\alpha$  Forest Absorbers. *The Astrophysical Journal Letters* **724** L69-L73.
- [100] POLITSCH, C. (2020). Statistical Astrophysics: From Extrasolar Planets to the Large-scale Structure of the Universe.
- [101] POLITSCH, C. A., CISEWSKI-KEHE, J., CROFT, R. A. C. and WASSERMAN, L. (2020). Trend filtering – I. A modern statistical tool for time-domain astronomy and astronomical spectroscopy. *Monthly Notices of the Royal Astronomical Society* **492** 4005-4018.
- [102] POLITSCH, C. A., CISEWSKI-KEHE, J., CROFT, R. A. C. and WASSERMAN, L. (2020). Trend filtering – II. Denoising astronomical signals with varying degrees of smoothness. *Monthly Notices of the Royal Astronomical Society* **492** 4019-4032.
- [103] POLITSCH, C. A. and CROFT, R. A. C. (2019). Mapping the Large-Scale Universe through Intergalactic Silhouettes. *CHANCE* **32** 14-19.
- [104] PRENTICE, S. J., ASHALL, C., MAZZALI, P. A., ZHANG, J. J., JAMES, P. A., WANG, X. F., VINKÓ, J., PERCIVAL, S., SHORT, L., PIASCIK, A., HUANG, F., MO, J., RUI, L. M., WANG, J. G., XIANG, D. F., XIN, Y. X., YI, W. M., YU, X. G., ZHAI, Q., ZHANG, T. M., HOSSEINZADEH, G., HOWELL, D. A., MCCULLY, C., VALENTI, S., CSEH, B., HANYECZ, O., KRISKOVICS, L., PÁL, A., SÁRNECZKY, K., SÓDOR, Á., SZAKÁTS, R., SZÉKELY, P., VARGA-VEREBÉLYI, E., VIDA, K., BRADAC, M., REICHAERT, D. E., SAND, D. and TARTAGLIA, L. (2018). SN 2016coi/ASASSN-16fp: an example of residual helium in a typeIc supernova? *Monthly Notices of the Royal Astronomical Society* **478** 4162-4192.
- [105] PRŠA, A., BATALHA, N., SLAWSON, R. W., DOYLE, L. R., WELSH, W. F., OROSZ, J. A., SEAGER, S., RUCKER, M., MJASETH, K., ENGLE, S. G., CONROY, K., JENKINS, J., CALDWELL, D., KOCH, D. and BORUCKI, W. (2011). KEPLERECLIPSING BINARY STARS. I. CATALOG AND PRINCIPAL CHARACTERIZATION OF 1879 ECLIPSING BINARIES IN THE FIRST DATA RELEASE. *The Astronomical Journal* **141** 83.
- [106] PRŠA, A., GUINAN, E. F., DEVINNEY, E. J., DEGEORGE, M., BRADSTREET, D. H., GIAMMARCO, J. M., ALCOCK, C. R. and ENGLE, S. G. (2008). Artificial Intelligence Approach to the Determination of Physical Properties of Eclipsing Binaries. I. The EBAI Project. *The Astrophysical Journal* **687** 542–565.
- [107] RAMDAS, A. and TIBSHIRANI, R. J. (2016). Fast and Flexible ADMM Algorithms for Trend Filtering. *Journal of Computational and Graphical Statistics* **25** 839-858.
- [108] RAUCH, M. (1998). THE LYMAN ALPHA FOREST IN THE SPECTRA OF QUASISTELLAR OBJECTS. *ARA&A* **36** 267-316.
- [109] RUDIN, L. I., OSHER, S. and FATERNI, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* **60** 259-268.
- [110] SCHLAFLY, E. F. and FINKBEINER, D. P. (2011). Measuring Reddening with Sloan Digital Sky Survey Stellar Spectra and Recalibrating SFD. *The Astrophysical Journal* **737** 103.
- [111] SCHMIDT, A. M. and O’HAGAN, A. (2003). Bayesian Inference for Non-Stationary Spatial Covariance Structure via Spatial Deformations. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **65** 743–758.
- [112] SHAPPEE, B. J., PRIETO, J. L., GRUPE, D., KOCHANÉK, C. S., STANEK, K. Z., ROSA, G. D., MATHUR, S., ZU, Y., PETERSON, B. M., POGGE, R. W., KOMOSSA, S., IM, M., JENKINSON, J., HOLOIEN, T. W.-S., BASU, U., BEACOM, J. F., SZCZYGIEL, D. M., BRIMACOMBE, J., ADAMS, S., CAMPILLAY, A., CHOI, C., CONTRERAS, C., DIETRICH, M., DUBBERLEY, M., ELPHICK, M., FOALE, S., GIUSTINI, M., GONZALEZ, C., HAWKINS, E., HOWELL, D. A., HSIAO, E. Y., KOSS, M., LEIGHLY, K. M., MORRELL, N., MUDD, D., MULLINS, D., NUGENT, J. M., PARRENT, J., PHILLIPS, M. M., POJMANSKI, G., ROSING, W., ROSS, R., SAND, D., TERNDRUP, D. M., VALENTI, S., WALKER, Z. and YOON, Y. (2014). THE MAN BEHIND THE CURTAIN: X-RAYS DRIVE THE UV THROUGH NIR VARIABILITY IN THE 2013 ACTIVE GALACTIC NUCLEUS OUTBURST IN NGC 2617. *The Astrophysical Journal* **788** 48.

- [113] SLAWSON, R. W., PRŠA, A., WELSH, W. F., OROSZ, J. A., RUCKER, M., BATALHA, N., DOYLE, L. R., ENGLE, S. G., CONROY, K., COUGHLIN, J. and ET AL. (2011). KEPLERECLIPSING BINARY STARS. II. 2165 ECLIPSING BINARIES IN THE SECOND DATA RELEASE. *The Astronomical Journal* **142** 160.
- [114] STEIDL, G., DIDAS, S. and NEUMANN, J. (2006). Splines in Higher Order TV Regularization. *International Journal of Computer Vision* **70** 241-255.
- [115] STEIN, C. M. (1981). Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics* **9** 1135–1151.
- [116] STONE, C. J. (1982). Optimal Global Rates of Convergence for Nonparametric Regression. *The Annals of Statistics* **10** 1040–1053.
- [117] TENNYSON, J. (2019). *Astronomical Spectroscopy: an Introduction to the Atomic and Molecular Physics of Astronomical Spectroscopy*.
- [118] TERRERAN, G., MARGUTTI, R., BERSIER, D., BRIMACOMBE, J., CAPRIOLI, D., CHALLIS, P., CHORNOCK, R., COPPEJANS, D. L., DONG, S., GUIDORZI, C., HURLEY, K., KIRSHNER, R., MIGLIORI, G., MILISAVLJEVIC, D., PALMER, D. M., PRIETO, J. L., TOMASELLA, L., MARCHANT, P., PASTORELLO, A., SHAPPEE, B. J., STANEK, K. Z., STRITZINGER, M. D., BENETTI, S., CHEN, P., DEMARCHI, L., ELIAS-ROSA, N., GALL, C., HARMANEN, J. and MATILTA, S. (2019). SN 2016coi (ASASSN-16fp): An Energetic H-stripped Core-collapse Supernova from a Massive Stellar Progenitor with Large Mass Loss. *The Astrophysical Journal* **883** 147.
- [119] THEUNS, T. and ZAROUBI, S. (2000). A wavelet analysis of the spectra of quasi-stellar objects. *Monthly Notices of the Royal Astronomical Society* **317** 989-995.
- [120] TIBSHIRANI, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58** 267–288.
- [121] TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B* **67** 91-108.
- [122] TIBSHIRANI, R. J. (2014). Adaptive Piecewise Polynomial Estimation via Trend Filtering. *The Annals of Statistics* **42** 285-323.
- [123] TIBSHIRANI, R. J. (2015). Degrees of freedom and model search. *Statistica Sinica* 1265–1296.
- [124] TIBSHIRANI, R. J. and TAYLOR, J. (2011). The solution path of the generalized lasso. *The Annals of Statistics* **39** 1335-1371.
- [125] TIBSHIRANI, R. J. and TAYLOR, J. (2012). Degrees of freedom in lasso problems. *The Annals of Statistics* **40** 1198-1232.
- [126] TOLSTOV, A., NOMOTO, K., SOROKINA, E., BLINNIKOV, S., TOMINAGA, N. and TANIGUCHI, Y. (2019). Light-curve Modeling of Fast-evolving Supernova KSN 2015K: Explosion in Circumstellar Matter of a Super-AGB Progenitor. *The Astrophysical Journal* **881** 35.
- [127] TSYBAKOV, A. B. (2008). *Introduction to Nonparametric Estimation*, 1st ed. Springer Publishing Company, Incorporated.
- [128] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.
- [129] WAHBA, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics.
- [130] WANG, Y.-X. et al. (2016). Trend Filtering on Graphs. *Journal of Machine Learning Research* **17** 1-41.
- [131] WANG, Y.-X., SMOLA, A. and TIBSHIRANI, R. (2014). The Falling Factorial Basis and Its Statistical Applications. In *Proceedings of the 31st International Conference on Machine Learning* (E. P. XING and T. JEBARA, eds.). *Proceedings of Machine Learning Research* **32** 730–738. PMLR, Beijing, China.
- [132] WASSERMAN, L. (2003). *All of Statistics: A Concise Course in Statistical Inference*. Springer Publishing Company, Incorporated.
- [133] WASSERMAN, L. (2006). *All of Nonparametric Statistics*. Springer Texts in Statistics.
- [134] WATSON, G. S. (1964). Smooth Regression Analysis. *Sankhyā: The Indian Journal of Statistics, Series A* **26** 359-372.
- [135] WOOSLEY, S. E., KASEN, D., BLINNIKOV, S. and SOROKINA, E. (2007). Type Ia Supernova Light Curves. *The Astrophysical Journal* **662** 487–503.
- [136] WU, C. F. J. (1986). Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion). *The Annals of Statistics* **14** 1261-1295.