

# Inspecting numerical methods for determining the kinematics of coronal mass ejections and coronal waves

J. P. Byrne<sup>1</sup>, D. M. Long<sup>2</sup> et al.

Address(es) of author(s) should be given

Received ?; accepted ?

## ABSTRACT

*Context.* Why are we doing this work?

*Aims.* In this paper we show that traditional techniques for the determination of CME and “EIT wave” kinematics, as currently applied, do not return accurate estimates of the true kinematics of the feature. We highlight the errors inherent in these approaches and illustrate a recipe for accurate estimates of the kinematics using a residual resampling bootstrapping approach to determine the confidence interval associated with the model used to measure them.

*Methods.* We discuss the errors inherent in the use of numerical differentiation techniques when applied to small data-sets. We present a residual resampling bootstrapping approach as a statistically rigorous technique for the determination of accurate kinematic estimates.

*Results.* It is shown that accurate feature kinematics can only be estimated by applying a pre-determined model to the position measurements. The validity of this model must be based on the physical properties of the feature that are to be measured, and the accuracy of applying that model to the data can be examined using a bootstrapping approach to determine the confidence interval associated with the estimated model parameters.

*Conclusions.* What are our conclusions?

**Key words.** Sun: activity – Sun: corona – Sun: coronal mass ejections (CMEs)

## 1. Introduction

The most defining feature of a transient solar phenomenon such as a Coronal Mass Ejection (CME) or a coronal wave (commonly called an “EIT Wave”) is its motion. These generally short-lived features, resulting from a solar eruption, are observed to propagate across the solar corona (i.e., coronal waves) or in the case of CMEs, outward from the Sun into the heliosphere. Observational catalogues of both phenomena have been compiled over more than  $\sim$ 20 years of observations (e.g., Illing & Hundhausen, 1985; Yashiro et al., 2004; Thompson & Myers, 2009), with the physical properties of both phenomena very well characterised (see the recent reviews by e.g., Gallagher & Long, 2011; Patsourakos & Vourlidas, 2012; Howard, 2011; Webb & Howard, 2012).

As transient phenomena, the kinematics of both sets of features continue to be one of the most important characteristics used to classify them. The motion of both phenomena is traditionally identified using difference images, where a preceding image is subtracted from a leading image to highlight motion, allowing the feature to be identified “by eye”. However, this approach highlights *relative* rather than *actual* motion, and is prone to undefined user-dependent bias. More recent work has used single image processing techniques such as wavelet transforms (Byrne et al., 2009; Morgan et al., 2012) and automated approaches (e.g., Long et al., 2011; Byrne et al., 2012; Podladchikova et al., 2012) to minimise user-error and reveal the true physical characteristics of the feature. By tracking the position of the feature with time it is possible to determine its kinemat-

ics, allowing an insight into the physical properties of the phenomenon.

The kinematics of these features are important for a variety of reasons. The true physical nature of coronal waves is not fully understood, with two main competing theories; that they are waves (e.g., Shen & Liu, 2012; Veronig et al., 2010) or signatures of magnetic field restructuring during a CME eruption (e.g., Schrijver et al., 2011; Chen & Wu, 2011). The kinematics of the feature have been proposed as one of the main discriminators between these competing theories, with the relatively high velocities measured thus far for this phenomenon suggesting a wave interpretation may be appropriate (cf. Warmuth & Mann, 2011; Zheng et al., 2012). Similarly, CME kinematics are vitally important from a space weather point of view as they allow increased accuracy in the predicted arrival time of the feature at Earth. The kinematic behaviour of the CME very low down in the solar corona may also be used to discriminate between eruption mechanisms (cf. Lin et al., 2010). However Wen et al. (2007) demonstrate that the errors in CME acceleration values can be of the same order as the accelerations typically measured.

A variety of different mathematical techniques exist for deriving the kinematics of transient features, ranging from the fitting of polynomial functions to the distance-time measurements to the numerical differentiation of the measurements. While these techniques may be mathematically sound, some of them are not necessarily applicable to the derivation of kinematics for these features and can produce spurious results. A recap of standard numerical methods and error propagation are presented in

Section 2. Simulations of the drawbacks of a standard numerical derivative are presented in Section 3. In Section 4 we outline a more appropriate method for inspecting the kinematics of CMEs and waves, as applied to model data. In Section 5 some real-data cases are revisited with these methods, and motivate the proposed treatment of data from the new coronal image processing CME catalogue (CORIMP; Morgan et al., 2012; Byrne et al., 2012), and coronal pulse identification and tracking algorithm catalogue (CorPITA; Long et al., 2011). A final discussion and conclusions are presented in Section 6.

## 2. Numerical Differentiation & Error Propagation (recap)

When presented with a moving object through a sequence of image frames such that it is possible to measure its position at each time step, the technique of numerical differentiation is often used to derive the velocity and acceleration of the object. In the standard 2-point approach, it should be possible to derive the time evolution of a system at time step  $t + \delta t$  according to the system values at time step  $t$ . This may be applied through the technique of forward, reverse or centre differencing, resulting in an estimate of the speed of the object at a specific time step given its positional information. More commonly, a 3-point Lagrangian interpolation is applied to better approximate the kinematics of a moving object by solving for the Lagrange polynomials that best fit across three given datapoints (e.g. DERIV.PRO in IDL). Each of these schemes is based upon the Taylor series expansion of a real function  $f(t)$ :

$$f(t_0 + \delta t) = f(t_0) + f'(t_0)\delta t + \frac{f''(t_0)}{2!}(\delta t)^2 + \dots \quad (1)$$

but due to the approximation of an infinite series with a finite number of terms and iterations, an error must be associated with the result, based on its deviation from the true solution. Generally the Euler method is employed, using the formula:

$$y_{n+1} = y_n + h f(t_n, y_n) \quad (2)$$

to solve the initial value problem  $y' = f(t, y)$  given  $y(t_0) = y_0$ , where  $h$  is the stepsize such that  $t_n = t_0 + nh$ . The convergence of such an approximation to the actual solution is prone to two sources of error; truncation error (the difference between the true solution and the approximation) and round-off error (the limited precision of the approximation). Added to this is the fact that the data measurements themselves are subject to uncertainties in both the positional and temporal information, and the ability of the numerical differentiation techniques to derive kinematics becomes highly jeopardised, as shall be shown.

Given a function  $x = f(u, v)$ , the error propagation equation (based on the standard deviations  $\sigma$  of the variables) is written:

$$\sigma_x^2 = \sigma_u^2 \left( \frac{\partial x}{\partial u} \right)^2 + \sigma_v^2 \left( \frac{\partial x}{\partial v} \right)^2 + 2\sigma_{uv}^2 \left( \frac{\partial x}{\partial u} \right) \left( \frac{\partial x}{\partial v} \right) + \dots \quad (3)$$

Specifically in the case of kinematic analyses, this is used to propagate the errors on the distance-time data  $r(t)$  into the velocity  $v(t)$  and acceleration  $a(t)$  profiles to determine the associated uncertainties. In the case of distance-time

data the covariance terms are zero because the quantities are uncorrelated.

When presented with relatively low sampling of the data, as in the case of coronagraph observations of CMEs and disk observations of waves, it is generally found that the simplest differentiation techniques are not applicable. The forward and/or reverse differencing techniques act to shift the kinematic profiles by one time-step, which is substantial enough to be of concern here (i.e., they derive a result at the current time-step, based on the pro-/preceding time-step). Centre differencing employs the two neighbouring data-points of the point under examination, and so is a better indication of the result at that time-step, but it fails at both endpoints. In any case these should not be employed when the spacing of the data-points is unequal, i.e., when the cadence  $\delta t$  is not constant, and so the 3-point Lagrangian interpolation technique is used (which gives the same result as the centre-difference otherwise, but includes the endpoints and has an associated error propagation formulation). The Lagrangian interpolation polynomial is given by:

$$L(x) = \sum_{j=0}^2 y_j l_j(x) \quad (4)$$

$$\text{where } l_j(x) = \prod_{i=0, i \neq j}^2 \frac{x - x_i}{x_j - x_i} \quad (5)$$

The derivative at point  $x$  is given by  $L' = \partial_x L(x)$ . So for the case of height-time data being used to derive velocity (and similarly acceleration) the associated error is given by:

$$\sigma_{v_1}^2 = \frac{\sigma_{r(t_2)}^2 + \sigma_{r(t_0)}^2}{(t_2 - t_0)^2} + v^2 \left( \frac{\sigma_{t_2}^2 + \sigma_{t_0}^2}{(t_2 - t_0)^2} \right) \quad (6)$$

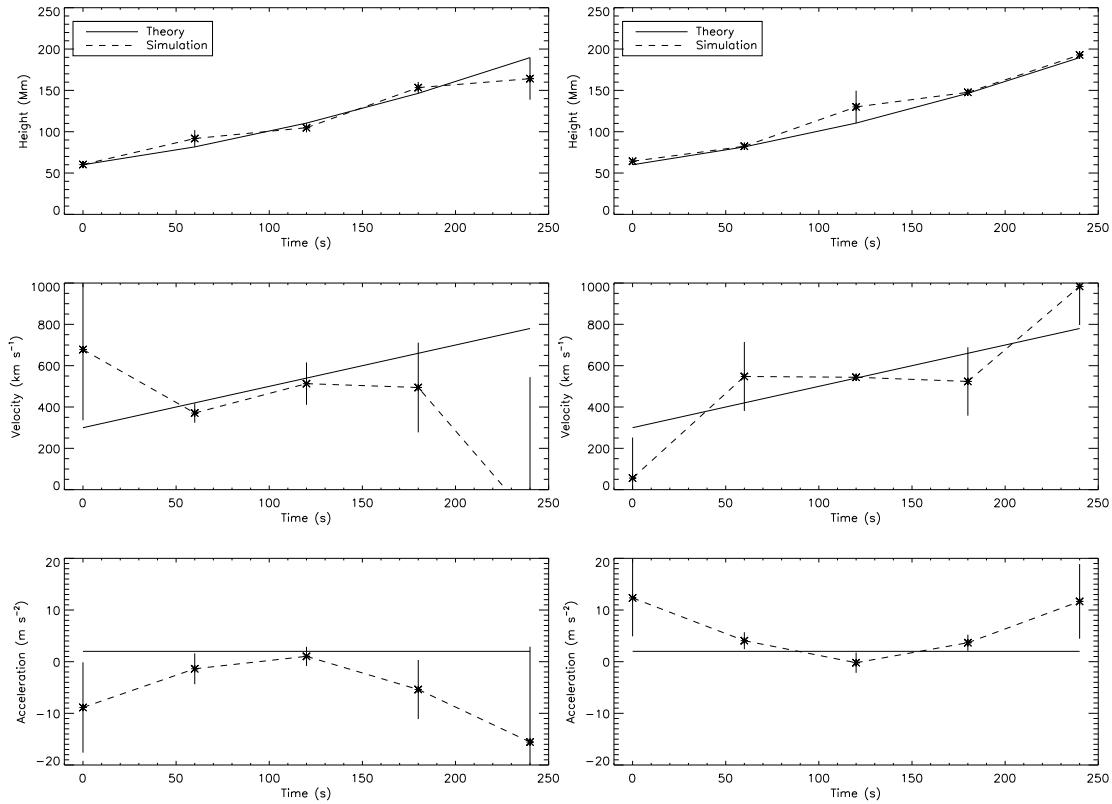
as in DERIVSIG.PRO in IDL. The endpoint errors are derived from a weighting of the pro/preceding two data-points, that is therefore larger to reflect the unknown nature of the trend beyond the sample points.

Although the 3-point Lagrangian is mathematically sound, its application to solar eruptive event kinematics proves problematic. The main drawbacks are two-fold:

1. The noise level, especially across low-cadence sampling, can scatter the measurements to such a degree that the numerical derivatives become untrustworthy and even misleading compared to the actual trends of the kinematic data.
2. The error-propagation formulation results in a misleading uncertainty interval on the velocity and acceleration profiles. For example, it counter-intuitively increases for increasing cadence measurements.

## 3. Simulations

In the case of CMEs and waves, there is great motivation to resolve the dynamics of their propagation as precisely as possible in order to study the forces at play. For example, CMEs in general may be undergoing continued driving (internal) forces, or positive or negative drag (external) forces, or most assuredly some interplay of both. Similarly wave propagation may be affected by changes to the low-coronal environment, e.g., low-density coronal hole regions, or strong magnetic field active regions. Thus any changes to event acceleration that result from different phases of dominating force, and where or why this can occur, are of great interest. But the true kinematics of such events remains



**Fig. 1.** A theoretical model for a CME with constant acceleration  $2 \text{ m s}^{-2}$  and initial velocity  $300 \text{ km s}^{-1}$ , and two simulations of how the resulting profiles for a noisy sample of data-points behave using 3-point Lagrangian interpolation.

somewhat elusive given the inherent limitations of the data and the numerical methods employed. In our treatment of both phenomena here, we use cases of simulated CME and wave data interchangeably, to demonstrate both constant and non-constant acceleration profiles with varying scatter and cadence.

### 3.1. Effect of noisy scatter on deriving kinematics

As an example of the effect of scatter due to noisy data, we first simulate a simple height-time profile of a CME that propagates according to the quadratic equation:

$$r(t) = r_0 + v_0 t + \frac{1}{2} a t^2 \quad (7)$$

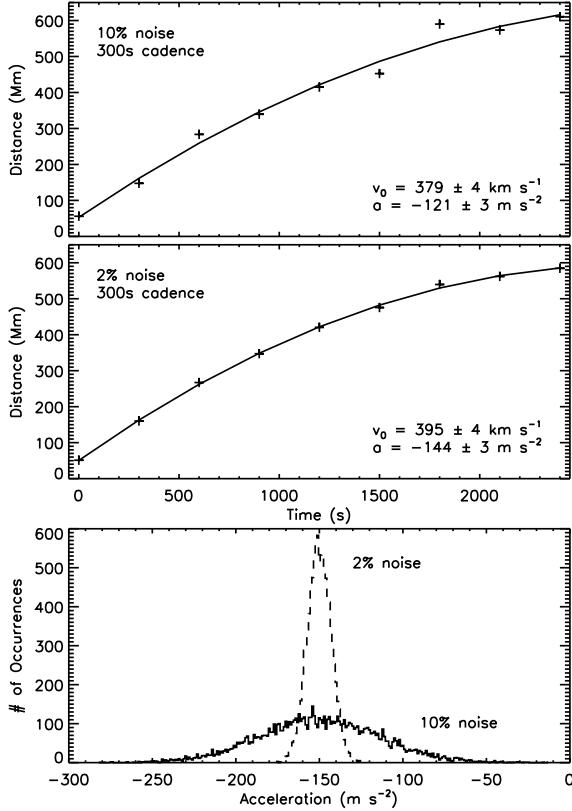
where  $r_0 = 60 \text{ Mm}$  is the initial height,  $v_0 = 300 \text{ km s}^{-1}$  is the initial velocity, and  $a = 2 \text{ m s}^{-2}$  is the acceleration of the CME. Varying levels of noise are randomly added up to a maximum of 20 %. An ‘extreme case’ errorbar on each datapoint is determined by its distance from the true height-time profile, to represent a scenario wherein all measurement uncertainties just manage to overlap the true profile. Various instances of randomized datapoint scatters result in erroneous trends in the velocity and acceleration profiles, even with the proper error treatment ascribed by the 3-point Lagrangian interpolation technique, and even in this simplest case of constant acceleration. Two examples are shown in the left and right of Figure 1 where completely opposing acceleration trends are determined for different samplings of the same dataset, indicating that the nature of the scatter in the samples is not satisfactorily reflected in the derived kinematics and their associated errors. At

the very least the datapoints should be expected to overlap the truth in each plot so that it remains a valid solution. A possible assurance, even in the case of low-cadence samplings (or very fast events), is that instead of trusting the endpoints they simply be removed. Figure 1 would then show three datapoints for velocity and one datapoint for acceleration, that would go somewhat toward removing the biased trends and imply a constant acceleration close to the true value. However, when dealing with low-number samples especially, it would be better not to have to remove datapoints.

The effects of varying scatter due to noise on the data of a coronal wave were also examined, demonstrated here for the case of a constant acceleration event. The wave motion is modeled by Equation 7, where  $r_0 = 50 \text{ Mm}$  is the initial distance of the wave from the source,  $v_0 = 400 \text{ km s}^{-1}$  is the initial velocity of the wave, and  $a = -150 \text{ m s}^{-2}$  is the acceleration of the wave. Figure 2 shows the derived kinematics for the simulated dataset with random noise added, shown here for  $3\sigma$  limits of 10 % (top panel) and 2 % (middle panel). A quadratic is then fit to each dataset to test how the noise level affects the precision of the derived kinematics, even in this idealized case of knowing the expected form of the data. The increased noise level acts to smooth out the true kinematics, as demonstrated by the different distributions of derived accelerations (bottom panel).

### 3.2. Effect of sampling cadence on deriving kinematics

As an example of the effect of cadence, we first simulate again the constant-acceleration profile of a coronal wave (as in Equation 7 and Figure 2). The data is sample at cadences



**Fig. 2.** Simulated datapoints for noise distributions with  $3\sigma$  widths of  $\pm 10\%$  (top), and  $\pm 2\%$  (middle) of the model value, at a fixed cadence of 300s, and the resulting quadratic fit and  $v_0$  and  $a$  parameters. The reduced noise level increases the precision in obtaining the true kinematics, as demonstrated by the different distributions of derived accelerations (bottom).

akin to the instruments EIT (12 mins) and AIA (12 s), at a fixed noise level of 10 %. Figure 3 shows examples of these, with a quadratic fit to the sample data to test the effect on the derived kinematics. It is clear that the higher-cadence data best resolves the true kinematic profile, providing an accurate estimation of the wave velocity and acceleration. These results are consistent with the observations made by both Long et al. (2008) and Ma et al. (2009) and show that the effects of image cadence must be accounted for when trying to derive the true kinematics of a coronal wave.

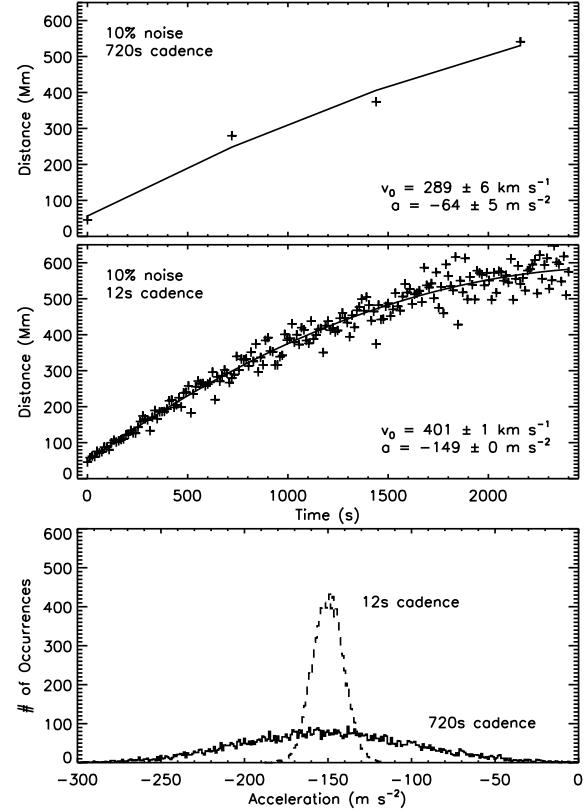
We next simulate a non-constant acceleration profile for a CME via the following equations (where the constant  $x$  is just a scaling factor):

$$h(t) = \sqrt{2x} t \tan^{-1} \left( \frac{e^{t/2x}}{\sqrt{2x}} \right) \quad (8)$$

$$v(t) = \sqrt{2x} \tan^{-1} \left( \frac{e^{t/2x}}{\sqrt{2x}} \right) + \frac{e^{t/x+2x}}{e^{t/x}+2x} \quad (9)$$

$$a(t) = \frac{e^{t/2x}(2x(t+4x)-e^{t/x}(t-4x))}{2x(e^{t/x}+2x)^2} \quad (10)$$

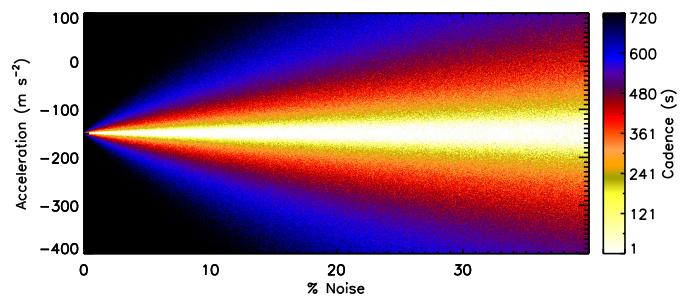
The acceleration profile exhibits an initial peak followed by a deceleration and then leveling to zero. This is akin to a general impulsive CME that undergoes an initial high-acceleration eruptive phase, and then decelerates to match the solar wind speed during its propagation phase. Thus a



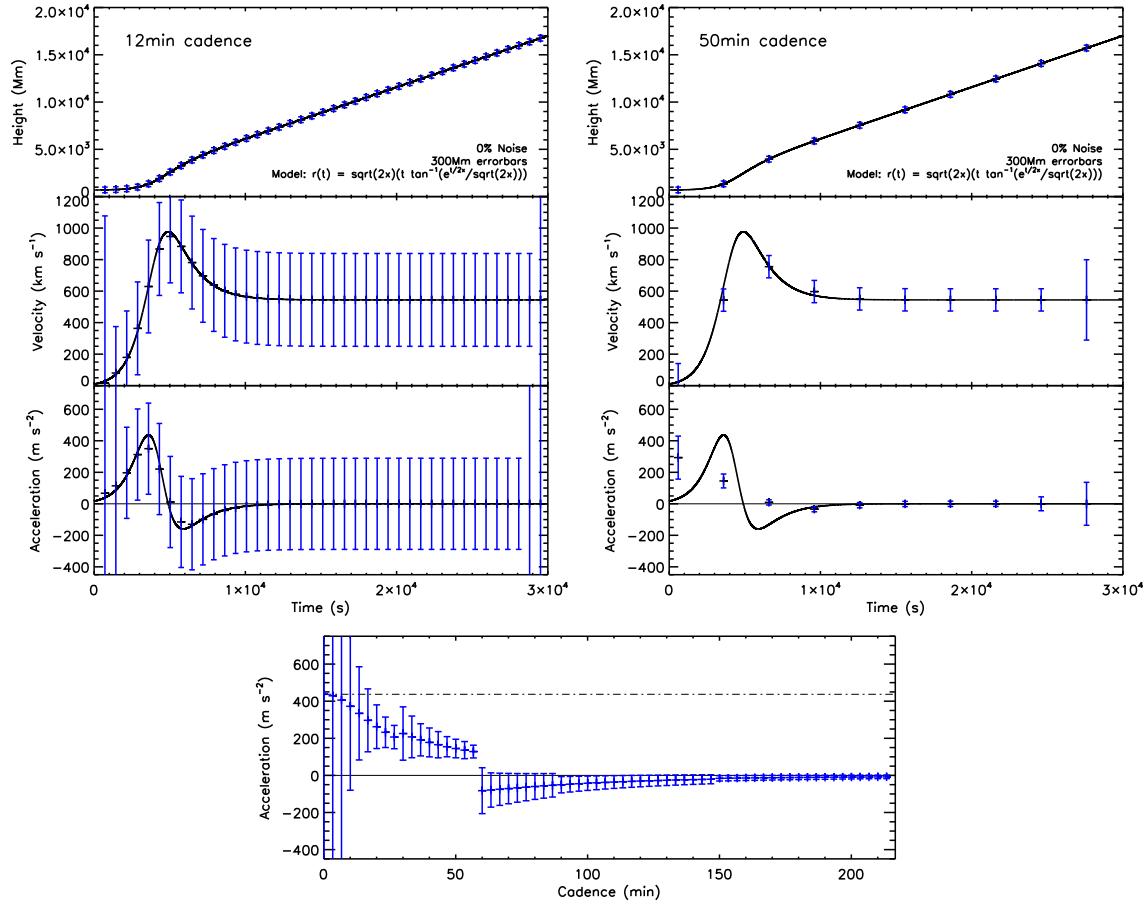
**Fig. 3.** Simulated datapoints for sampling cadences of 720 s (EIT; top) and 12 s (AIA; middle), at a fixed noise level of  $\pm 10\%$ , and the resulting quadratic fit and  $v_0$  and  $a$  parameters. The increased cadence offers better precision in obtaining the true kinematics, as demonstrated by the different distributions of derived accelerations (bottom).

model CME height-time profile is generated enabling synthetic observation samples to be taken at different cadences (Figure 5).

We investigate the effect of the cadence of the observations on the derivation of the kinematics and associated errorbars using the standard 3-point Lagrangian interpola-



**Fig. 4.** A simulation of derived acceleration values, output from the model fit of Equation 7 (with examples shown in Figures 2 and 3), for varying noise levels of 0–40% and varying cadence times of 1–720 s. As shown, both decreasing noise level and decreasing cadence time improves the chances of obtaining the correct acceleration value, being  $-150 \text{ m s}^{-2}$  in this model coronal wave case.

**Fig. 5.**

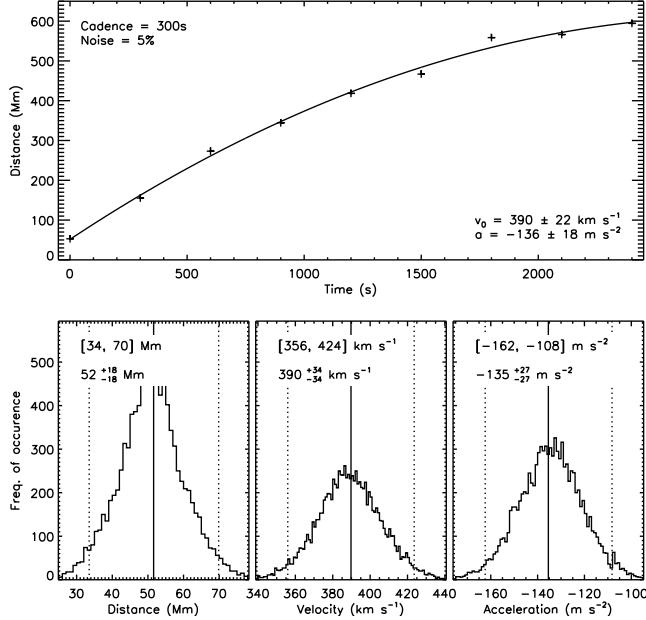
tion. In the first instance fixed  $3\sigma$  errorbars of  $\pm 300$  Mm are applied to the height-time points, without any noise added. This is useful to simply test the effects of the cadence on the derived velocity and acceleration profiles and their associated errors. The top left and right plots of Figure 5 show the model height-time, velocity and acceleration profiles sampled at cadences of 12 and 50 mins respectively. As the cadence is reduced, i.e., the cadence time is increased, the errorbars become smaller due to the inverse dependence of the Lagrangian error terms on the time between the datapoints  $\Delta t^{-2}$  (see Equation 6). However, reducing the cadence reduces the resolution with which the acceleration peak is detectable, and so the acceleration profile is smoothed out. Conversely, the errorbars would become erroneously large for very high-cadence measurements, even though the measurements would better reveal the true trends of the kinematic profiles. This fundamentally implies that the errorbars do not truly reflect the uncertainty on the data at a given cadence, and are in fact completely redundant for these cases.

It is clear therefore, that the variation in both noise and imaging cadence can strongly influence the derived kinematics of a CME or coronal wave, and the 3-point Lagrangian technique does not return useful estimates of the associated uncertainty. However, it is possible to use other techniques along with a bootstrapping approach to overcome these issues, and produce a more statistically sound method for dealing with small datasets such as these.

#### 4. Bootstrapping

When trying to determine an estimator for a particular parameter of interest and subsequently evaluate the accuracy of that estimator, a small sample-size is immediately limiting. So techniques based on resampling methods have been developed, in order to approximate the behavior of the true distribution by resampling the data enough times to generate a maximum likelihood estimator of the distribution. Bootstrapping, first introduced by Efron (1979) and more recently described in e.g. Efron & Tibshirani (1994) and Chernick (1999), is one such technique, that may be formally defined as follows: Given a random sample of  $n$  independent identically distributed vectors  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  from an unknown probability distribution function  $\mathbf{F}$  and a real-valued estimator  $\hat{\theta} = s(\mathbf{x})$ , a procedure (the bootstrap) to assess the accuracy of  $\hat{\theta}$  is defined in terms of the empirical distribution function  $\hat{\mathbf{F}}$ , which is the maximum likelihood estimator of the distribution for the observations when no parametric assumptions are made. Otherwise stated, a bootstrap sample  $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$  is generated by randomly sampling with replacement from the population of  $n$  objects, giving a resampled version of  $\mathbf{x}$ . The bootstrap algorithm then works by drawing many independent bootstrap samples, to estimate the standard error of  $\hat{\theta}$  from the observed data  $\mathbf{x}$ .

Thus in the cases of CME and coronal wave observations, bootstrap techniques can prove very useful for determining the accuracy of the derived form of their kinemat-



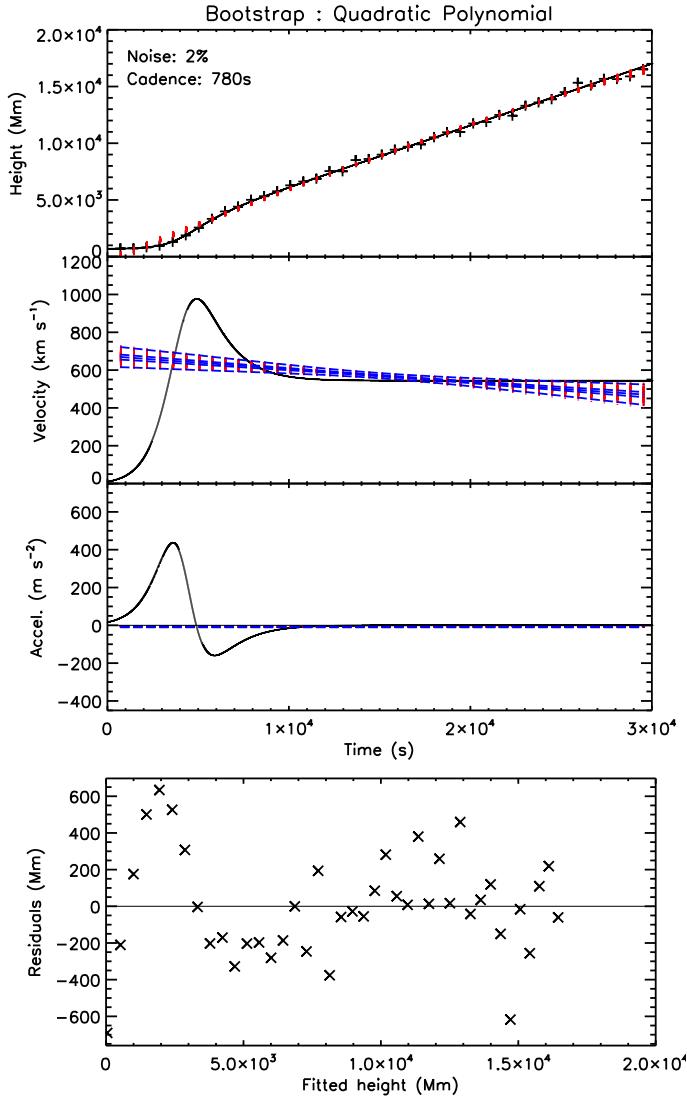
**Fig. 6.** Top panel shows the initial fit of Equation 7 to simulated coronal wave measurements with 5% noise and 300 s cadence. The fit parameters are quoted with 1-sigma uncertainties. The bottom panels show histograms of the initial distance, initial velocity, and acceleration values derived using the bootstrapping technique. The mean and 95% confidence interval are indicated by the solid and dotted lines respectively, and quoted on each panel in two ways!

ics. The implementation of the bootstrapping technique is as follows:

1. An initial fit to the data  $y$  is obtained, yielding the model fit  $\hat{y}$  with parameters  $p$ .
2. The residuals of the fit are calculated:  $\epsilon = y - \hat{y}$ .
3. The residuals are randomly resampled with replacement to give  $\epsilon^*$ .
4. The model is then fit to a new data vector  $y^* = y + \epsilon^*$  and the parameters  $p^*$  stored.
5. Steps 3–4 are repeated many times (e.g. 10,000).
6. Confidence intervals on the parameters are determined from the resulting distributions.

This technique was used to fit a quadratic model, first to the simulated coronal wave moving with constant acceleration (Equation 7), and second to the simulated CME moving with non-constant acceleration (Equation 10). In the case of the wave, the initial fit to the data and the bootstrapped distributions of initial height, initial velocity, and acceleration values are shown in Figure 6. Bootstrapping in this manner allows us to derive a confidence interval on the fit parameters. This is taken from the  $100\alpha$ th and  $100(1 - \alpha)$ th percentiles of the distribution (giving a 95% confidence interval when  $\alpha = .025$ ). Since in this case the distributions are close to symmetric about their means, we can see that the bootstrapped confidence interval is more precise than the corresponding uncertainty intervals on the single fit parameters. For example, the 1-sigma uncertainty on the acceleration given by the single fit in Figure 6 is  $\pm 17 \text{ m s}^{-2}$ , which is greater than half the 95% confidence interval range of  $\pm 27 \text{ m s}^{-2}$  (i.e., likening the 95% range to a 2-sigma uncertainty).

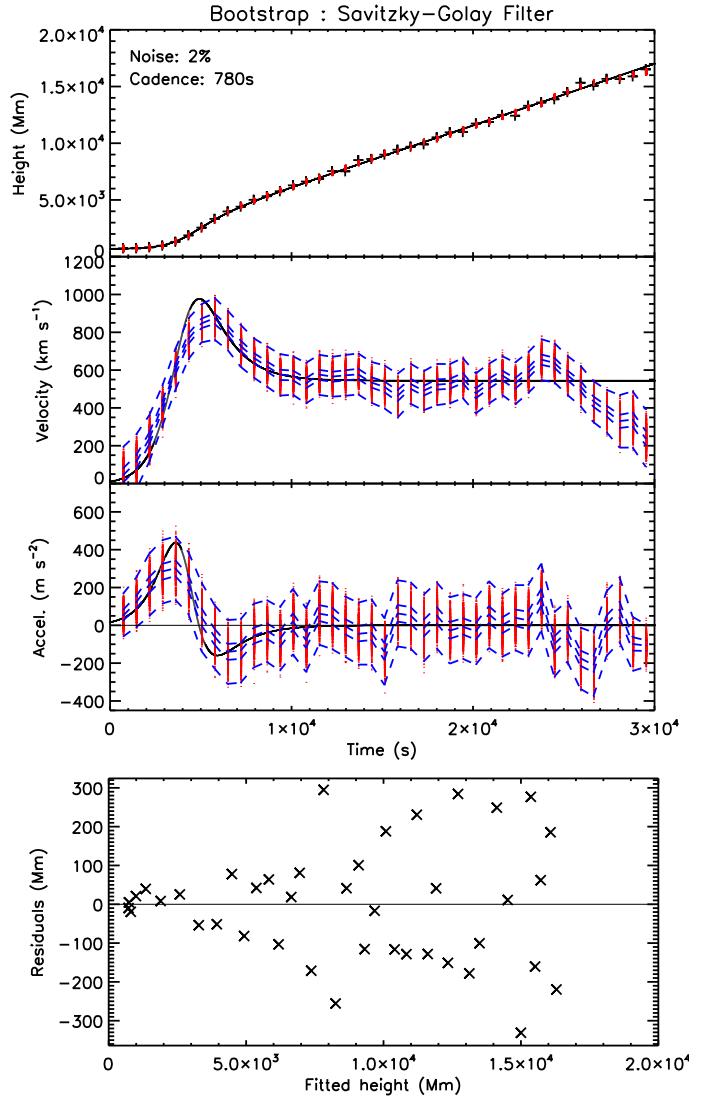
However, bootstrapping of a model cannot be applied blindly, as we will demonstrate for a simple case of fitting the same constant acceleration form of Equation 7 to the non-constant acceleration of Equation 10. Figure 7 shows the associated model CME profile, sampled at 780 s cadence with 2% noise. The red points show the resulting distributions of resampled points after the residuals are resampled with replacement. Thus a distribution of velocity and acceleration values are derived, with the corresponding median, interquartile range, and upper and lower fences overlaid in blue. However, the true non-constant acceleration profile is being washed-out by this second-order fit, as revealed by the trend in the residuals of the initial fit (bottom panel of Figure 7). So for any cases where possible non-constant acceleration profiles are to be revealed, the method for deriving the kinematics must be applied at an appropriate scale, and such that the residuals scatter randomly. In this specific case, the early acceleration peak is affecting the trend of the fit over the whole interval, when ideally a piecewise function should be used to characterize the different phases of motion. This is inspected in great detail by Schrijver et al. (2008) in an effort to model the early acceleration phase of erupting filaments involved in CMEs. They show that one functional form alone cannot describe the entire phase as well as a number of different functions can. Since we do not know the functional form that CME or coronal wave kinematics will have, we must rely on some form of numerical derivative for revealing possible trends. Since the issues with the 3-point Lagrangian have been highlighted in Section 3, we shall opt instead to implement the Savitzky-Golay filter (Savitzky & Golay, 1964). This is a form of local polynomial regression of chosen degree of smoothing polynomial, and of chosen order to produce smoothed first order, second order, etc., derivatives of the signal. The number of data points either side of the case point to be included in the filter is also specified. So for treating the distance-time data of CMEs and coronal waves, the Savitzky-Golay filter is a better method for smoothing small-scale noise while still revealing the true kinematic profiles. This is illustrated in Figure 8 for the simulated non-constant acceleration profile, with the residuals resampled with replacement as per the bootstrapping technique described above and demonstrated in Figures 6 and 7. Since it is a form of “moving window averaging”, slight biases may be introduced at local maxima and minima where the function value can be reduced, but its implementation still proves more robust than the standard 3-point Lagrangian at smoothing the scatter while still revealing the overall trend. Note that the residuals can be seen to scatter somewhat randomly, as desired (bottom panel of Figure 8). Also, for this case-study it is important to note how the scatter at the endpoints gives the impression of a decreasing velocity which the Savitzky-Golay filter faithfully reproduces even though we know it’s not how the model is behaving. This is akin to how the intensity of a CME or coronal wave often becomes too close to the background intensity at such distances, that it is lost to the noise and the measured profile falls off. This alludes to considerations that must be made when dealing with automated systems of kinematic determination, whereby the algorithmic limitations can introduce systematic biases not accounted for in the derived kinematics and associated uncertainties. This is discussed further in the next section.



**Fig. 7.** The bootstrapped second order polynomial (quadratic) fit to simulated CME height-time data, with 2% noise and a sampling cadence of 780 s. The panels from top to bottom show the height, velocity, and acceleration plots, and the residuals of the initial fitted height. The red points show the resampled residuals with replacement, and the blue lines are the median, interquartiles range, and upper and lower fences on the bootstrapped fit. The quadratic form tends to smooth out the non-constant acceleration profile, as revealed by the trend in the residuals that indicates the fit is not appropriate for the data.

## 5. Case Studies

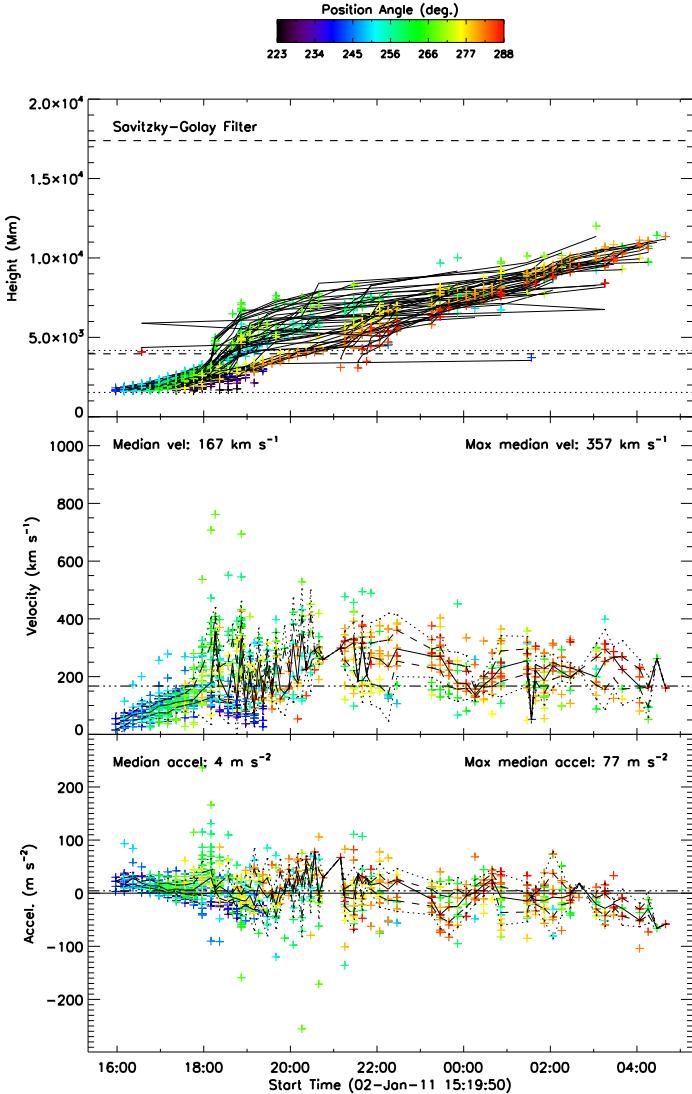
In this section some examples of real data are presented as case studies for deriving the CME and coronal wave kinematics in light of the discussion of the previous sections. The newly developed CORIMP and CorPITA catalogues are used to detect and measure the height-time profiles of CMEs and coronal waves in such a manner as to allow a determination of the ranges of speeds across the expanding phenomena as they propagate in time. They are inspected separately as follows.



**Fig. 8.** The bootstrapped Savitzky-Golay filter method applied to the simulated CME height-time data as in Figure 7. This manner of piecewise fit smooths the data by fitting a polynomial to the three neighboring points either side of each datapoint, and is successful in revealing the non-constant acceleration profile. The randomly scattered residuals also indicate its appropriateness.

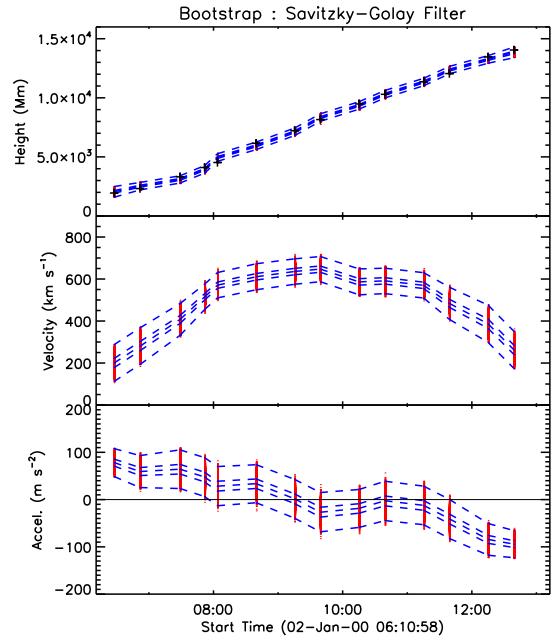
### 5.1. CORIMP

A CME that occurred on 2011 January 2, (ideally shown events as shown in Figure ?? of Byrne et al. (2012)), is presented here as a case-study from the CORIMP CME catalogue. The height-time points are measured at 1 degree intervals across the angular span of the event as it moves through the corona in time (see top of Figure 9). The Savitzky-Golay filer is then applied to each of these angular slices across the CME, at a chosen window-size of 6 data points (3 neighboring points either side of the datapoint in question) and the corresponding first and second order derivatives of the fit are determined (to reveal the spread of velocity and acceleration points in the middle and bottom plots of Figure 9). Since a spread of measurements across the CME can be determined in these cases, that spread is treated as the range of variation in the kine-



**Fig. 9.** The Savitzky-Golay filter method applied to a real CME data case from CORIMP.

matics without necessarily requiring bootstrap methods to be employed. The implication here is that any variation in CME speeds that may result from the expansion of the ejecta across the plane-of-sky, should provide a distribution of kinematics that is centered on the motion of the bulk of the CME (as opposed to the relatively slower moving flanks of the CME). In essence, this provides a solution space of CME kinematics that is based purely on the distribution of CME height-time measurements. Additional bootstrapping is not necessary at this point since a specific model is not being tested, nor would any algorithmic bias that may have occurred be overcome, e.g. if the CME detection is lost to the background or a false detection contaminates the measurements. However if a user of the catalogue has a specific model that they wish to test, then a bootstrapping procedure and residual analysis would be of use, as demonstrated in Section 4. For the spread of CME kinematic measurements, as in Figure 9, the median, interquartile range, and upper and lower fences, are determined in order to characterize the significance of the range of velocity and acceleration points. This allows the underlying trend of the data



**Fig. 10.** The bootstrapped Savitzky-Golay filter method applied to a real CME data case (not from CORIMP but revisited from [Byrne et al. 2009](#))...

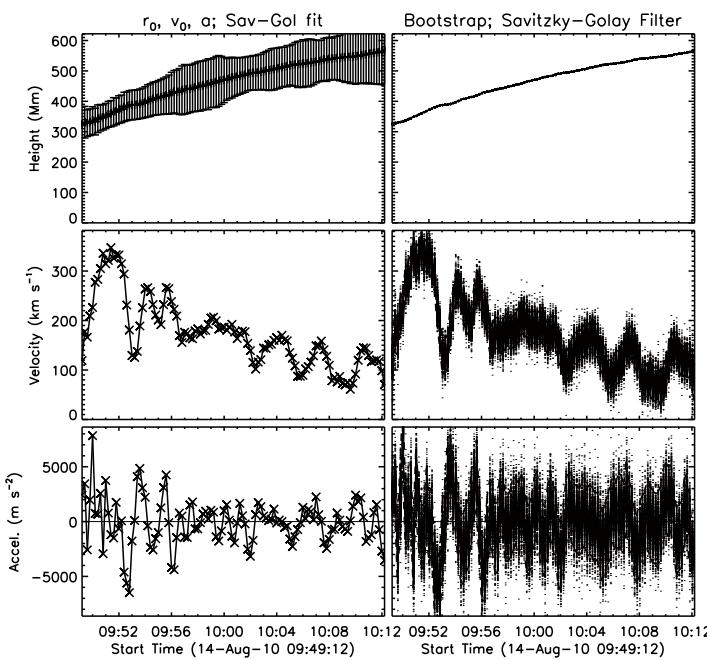
to be inspected, revealing an initial acceleration of approximately  $20 \text{ m s}^{-2}$  up to a constant velocity of approximately  $250 \text{ km s}^{-1}$ . In order to obtain quotable single-values for the velocity and acceleration of the event, the median value of all of the velocity and acceleration points is determined; being  $167 \text{ km s}^{-1}$  and  $4 \text{ m s}^{-2}$  respectively. The angular width of the event is quoted as  $65^\circ$ , centered on position angle  $255^\circ$  (measured counter-clockwise from solar north). Thus these values may be compared directly with the other catalogues, as listed in Table 1. However, the output produced from the more appropriate treatment of the kinematics as per the motivation and discussion of this paper, greatly improves the ability to study CME dynamics.

## 5.2. CorPITA

## 6. Conclusions

- 3 point Lagrangian unreliable.
- Cadence issues of great consequence.
- Unknown bias (notably automated detection scatter) cannot be measured.
- Errorbars redundant beyond testing single models.
- Scatter about fit is the only legitimate concern.
- What that fit should be is ‘*a priori*’ knowledge.

*Acknowledgements.* This work is supported by SHINE grant 0962716 and NASA grant NNX08AJ07G to the Institute for Astronomy. The SOHO/LASCO data used here are produced by a consortium of the Naval Research Laboratory (USA), Max-Planck-Institut fuer Aeronomie (Germany), Laboratoire d’Astronomie (France), and the University of Birmingham (UK). SOHO is a project of international cooperation between ESA and NASA. The STEREO/SECCHI project is an international consortium of the Naval Research Laboratory (USA), Lockheed Martin Solar and Astrophysics Lab



**Fig. 11.** The bootstrapped Savitzky-Golay filter method applied to a real wave data case. Note the errorbars need work!!!

(USA), NASA Goddard Space Flight Center (USA), Rutherford Appleton Laboratory (UK), University of Birmingham (UK), Max-Planck-Institut für Sonnen-systemforschung (Germany), Centre Spatial de Liege (Belgium), Institut d'Optique Théorique et Appliquée (France), and Institut d'Astrophysique Spatiale (France).

## References

- Byrne, J. P., Gallagher, P. T., McAteer, R. T. J., & Young, C. A. 2009, *A&A*, 495, 325
- Byrne, J. P., Morgan, H., Habbal, S. R., & Gallagher, P. T. 2012, *ApJ*, 752, 145
- Chen, P. F., & Wu, Y. 2011, *ApJ*, 732, L20
- Chernick, M. 1999, *Bootstrap methods: A practitioner's guide* (New York: Wiley)
- Efron, B. 1979, *Ann. Stat.*
- Efron, B., & Tibshirani, R. J. 1994, *An Introduction to the Bootstrap* (Chapman & Hall/CRC)
- Gallagher, P. T., & Long, D. M. 2011, *Space Sci. Rev.*, 158, 365
- Howard, T., ed. 2011, *Astrophysics and Space Science Library*, Vol. 376, Coronal Mass Ejections%
- Illing, R. M. E., & Hundhausen, A. J. 1985, *J. Geophys. Res.*, 90, 275
- Lin, C., Gallagher, P. T., & Raftery, C. L. 2010, *A&A*, 516, A44+
- Long, D. M., Gallagher, P. T., McAteer, R. T. J., & Bloomfield, D. S. 2008, *ApJ*, 680, L81
- . 2011, *A&A*, 531, A42
- Ma, S., Wills-Davey, M. J., Lin, J., et al. 2009, *ApJ*, 707, 503
- Morgan, H., Byrne, J. P., & Habbal, S. R. 2012, *ApJ*, 752, 144
- Patsourakos, S., & Vourlidas, A. 2012, *Sol. Phys.*, 93
- Podladchikova, O., Vuets, A., Leontiev, P., & van der Linden, R. A. M. 2012, *Sol. Phys.*, 276, 479
- Savitzky, A., & Golay, M. 1964, *Analytical Chemistry*, 36, 1627
- Schrijver, C. J., Aulanier, G., Title, A. M., Pariat, E., & Delannée, C. 2011, *ApJ*, 738, 167
- Schrijver, C. J., Elmore, C., Kliem, B., Török, T., & Title, A. M. 2008, *ApJ*, 674, 586
- Shen, Y., & Liu, Y. 2012, *ApJ*, 754, 7
- Thompson, B. J., & Myers, D. C. 2009, *ApJS*, 183, 225
- Veronig, A. M., Muhr, N., Kienreich, I. W., Temmer, M., & Vršnak, B. 2010, *ApJ*, 716, L57
- Warmuth, A., & Mann, G. 2011, *A&A*, 532, A151
- Webb, D. F., & Howard, T. A. 2012, *Living Reviews in Solar Physics*, 9, 3
- Wen, Y., Maia, D. J. F., & Wang, J. 2007, *ApJ*, 657, 1117
- Yashiro, S., Gopalswamy, N., Michalek, G., et al. 2004, *Journal of Geophysical Research (Space Physics)*, 109, 7105
- Zheng, R., Jiang, Y., Yang, J., et al. 2012, *ApJ*, 753, 112