

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

J. P. Byrne<sup>1</sup>, D. M. Long<sup>2</sup>, P. T. Gallagher<sup>3</sup>, D. S. Bloomfield<sup>3</sup>, S. A. Maloney<sup>4</sup>, R. T. J. McAteer<sup>5</sup>,  
H. Morgan<sup>6,1,7</sup>, and S. R. Habbal<sup>1</sup>

<sup>1</sup> Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA  
e-mail: jbyrne@ifa.hawaii.edu

<sup>2</sup> UCL–Mullard Space Science Laboratory, Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK

<sup>3</sup> School of Physics, Trinity College Dublin, College Green, Dublin 2, Ireland

<sup>4</sup> Skytek, 51/52 Fitzwilliam Square West, Dublin 2, Ireland

<sup>5</sup> Department of Astronomy, New Mexico State University, Las Cruces, NM 88003-8001, USA

<sup>6</sup> Sefydliad Mathemateg a Ffiseg, Prifysgol Aberystwyth, Ceredigion, Cymru, SY23 3BZ, UK

<sup>7</sup> Coleg Cymraeg Cenedlaethol, Y Llwfan, Ffordd y Coleg, Caerfyrddin, Cymru, SA31 3EQ, UK

Received 1 February 2013 / Accepted 26 July 2013

## ABSTRACT

**Context.** The study of solar eruptive events and associated phenomena is of great importance in the context of solar and heliophysics. Coronal mass ejections (CMEs) and coronal waves are energetic manifestations of the restructuring of the solar magnetic field and mass motion of the plasma. Characterising this motion is vital for deriving the dynamics of these events and thus understanding the physics driving their initiation and propagation. The development and use of appropriate methods for measuring event kinematics is therefore imperative.

**Aims.** Traditional approaches to the study of CME and coronal wave kinematics do not return wholly accurate nor robust estimates of the true event kinematics and associated uncertainties. We highlight the drawbacks of these approaches, and demonstrate improved methods for accurate and reliable determination of the kinematics.

**Methods.** The Savitzky-Golay filter is demonstrated as a more appropriate fitting technique for CME and coronal wave studies, and a residual resampling bootstrap technique is demonstrated as a statistically rigorous method for the determination of kinematic error estimates and goodness-of-fit tests.

**Results.** It is shown that the scatter on distance-time measurements of small sample size can significantly limit the ability to derive accurate and reliable kinematics. This may be overcome by (i) increasing measurement precision and sampling cadence; and (ii) applying robust methods for deriving the kinematics and reliably determining their associated uncertainties. If a priori knowledge exists and a pre-determined model form for the kinematics is available (or indeed any justified fitting-form to be tested against the data), then its precision can be examined using a bootstrapping technique to determine the confidence interval associated with the model/fitting parameters.

**Conclusions.** Improved methods for determining the kinematics of CMEs and coronal waves are demonstrated to great effect, overcoming many issues highlighted in traditional numerical differencing and error propagation techniques.

**Key words.** Sun: activity – Sun: corona – Sun: coronal mass ejections (CMEs) – methods: data analysis – methods: numerical – methods: statistical

## 1. Introduction

Coronal mass ejections (CMEs) and coronal waves (commonly known as “EIT waves”) are large-scale manifestations of solar activity that indicate a restructuring of the global solar magnetic field. These phenomena involve the mass motion of plasma through the solar corona, with energies on the order of  $10^{25}$  J for CMEs (Emslie et al. 2004), and upwards of  $10^{18}$  J for coronal waves (Ballai et al. 2005). CME speeds range from about 20 to  $>2500$  km s<sup>-1</sup> (Yashiro et al. 2004), most typically moving at speeds similar to those of coronal waves, which range from 50 to  $>700$  km s<sup>-1</sup> (Thompson & Myers 2009). Observational catalogues of these events have been compiled from over ~20 years of observations, with the aim of characterising their physical properties in order to better understand the dynamics of their initiation and propagation (see some recent reviews by Howard 2011; Gallagher & Long 2011; Patsourakos & Vourlidas 2012; Webb & Howard 2012), and form an integral part of such efforts as the Solar Dynamics Observatory Feature

Finding Team (SDO FFT; Martens et al. 2012), the Heliophysics Event Knowledgebase (HEK; Hurlburt et al. 2012), and the HELiophysics Integrated Observatory (HELIO; Bentley et al. 2011). CME dynamics, in particular, are of great interest in a space weather context (e.g., Howard & Tappin 2005; Schrijver & Siscoe 2010; Riley 2012). Therefore, methods for determining their kinematics with improved accuracy are extremely important if scientists are to become skilled at predicting their behaviour and arrival times (see, for example, efforts by Prangé et al. 2004; Schwenn et al. 2005; Colaninno & Vourlidas 2006; Byrne et al. 2010). Temmer et al. (2012) also highlight the importance of characterising event kinematics as accurately as possible, in order to understand the effects of drag and CME-CME interactions and the subsequent implications for the underlying magnetic field structures involved.

In order to observe and characterise the motion of the bulk plasma of these events, a method of image-differencing is traditionally used, whereby a preceding image is subtracted from a leading image to highlight moving features. However, this

approach enhances relative rather than actual motion and is prone to spatiotemporal crosstalk and user-dependent bias. More recent work has used single-image processing techniques such as multiscale filters (Young & Gallagher 2008; Byrne et al. 2009; Gallagher et al. 2011) and robust automated approaches (e.g., Long et al. 2011; Podladchikova et al. 2012; Byrne et al. 2012) to overcome these issues and reveal the true physical characteristics of the events. Thus, by accurately tracking the position of a feature over time, it is possible to determine the kinematics of the event.

The true physical nature of coronal waves is not fully understood, with two main competing theories: that they are indeed waves (e.g., Veronig et al. 2010; Shen & Liu 2012), or that they are signatures of magnetic field restructuring during a CME eruption (e.g., Schrijver et al. 2011; Chen & Wu 2011). Their kinematic behaviour has been proposed as one of the main discriminators between these competing theories, with the relatively high velocities measured thus far suggesting a wave interpretation to be appropriate (Warmuth & Mann 2011; Zheng et al. 2012). Similarly, the low-coronal kinematics of CMEs may be used to discriminate between eruption mechanisms (see, for example, Lin et al. 2010, and the CME models discussed therein). However, Wen et al. (2007) demonstrate that the errors in CME acceleration values can be of the same order as the accelerations typically measured, making this task difficult. This has led to many statistical studies employing large numbers of events to try and determine a general form for typical CME motion (e.g., Gopalswamy et al. 2000; dal Lago et al. 2003; Zhang & Dere 2006). However, individual events need to be studied with rigour in order to satisfactorily derive the kinematics and gain insight to the physics at play.

A variety of different mathematical techniques exist for deriving the kinematics of transient features, most being based upon some form of numerical differentiation of the distance-time measurements and/or the fitting of a pre-assumed model function. While such techniques may appear mathematically sound, some of them are not necessarily applicable to the derivation of kinematics for these specific forms of events and can produce spurious results. Temmer et al. (2010) demonstrated the applicability of an inversion technique (developed by Kontar & MacKinnon 2005) to overcome some of these issues. They highlight its effectiveness at reducing the errors on three CME/flare studies compared to a standard spline-fit approach, and so the technique shows promise if it can be applied robustly across a variety of events.

Here, we further explore improved methods for overcoming the issues outlined above to robustly determine the kinematics of CMEs and coronal waves with improved accuracy and reliability. Simulations of the drawbacks of a standard numerical derivative are presented in Sect. 2. In Sect. 3, we outline a more appropriate method for inspecting the kinematics of CMEs and coronal waves, as applied to model data. In Sect. 4, some real-data cases are inspected via these methods, motivated by 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). The main conclusions and future work are discussed in Sect. 5.

## 2. Simulated data

When presented with relatively low sampling of the data, it is generally found that the simplest differentiation techniques are not applicable. The forward and 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 change from the preceding or proceeding 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 the endpoints. In any case, these techniques should not be employed when the spacing of the data points is unequal (i.e., when the cadence,  $\delta t$ , is not constant). Therefore, the 3-point Lagrangian interpolation technique is often used (as in DERIV.PRO in IDL), which includes the endpoints and has an associated error propagation formulation, which for the velocity is given by,

$$\sigma_{v_1}^2 = \frac{\sigma_{r(t_2)}^2 + \sigma_{r(t_0)}^2}{(t_2 - t_0)^2} + v^2 \frac{\sigma_{t_2}^2 + \sigma_{t_0}^2}{(t_2 - t_0)^2}, \quad (1)$$

and similarly for acceleration (as in DERIVSIG.PRO in IDL, from Bevington & Robinson 2003). The endpoint errors are derived from a weighting of the preceding or proceeding two data points and are therefore larger, reflecting the unknown nature of the trend beyond these points. Although the 3-point Lagrangian is mathematically sound, its application to CME and coronal wave kinematics proves problematic. The main drawbacks, which will be discussed in detail in the following subsections, are two-fold:

1. The scatter in the measurements, especially across low-cadence sampling, can cause the numerical derivatives to become untrustworthy and even misleading compared to the actual trends of the kinematic data.
2. The error-propagation formulation results in a misleading uncertainty on the velocity and acceleration profiles, whereby the errorbars counter-intuitively increase in size when the cadence itself is increased (i.e., when more frequent observations are made).

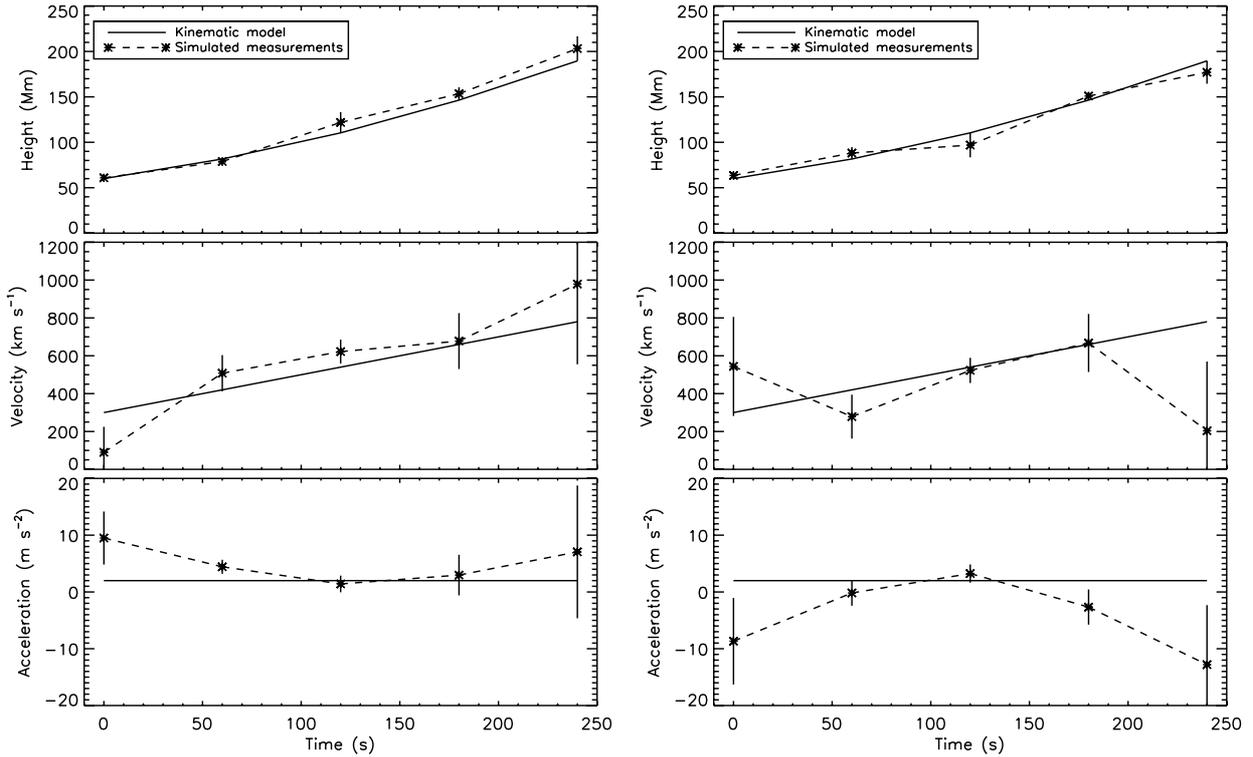
In our treatment of both phenomena here, we use cases of simulated CME and coronal wave measurements interchangeably to demonstrate both constant and non-constant acceleration profiles with varying measurement scatter and cadence.

### 2.1. Effect of measurement scatter on deriving kinematics

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

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

where  $r_0 = 60$  Mm is the initial height,  $v_0 = 300$  km s<sup>-1</sup> is the initial velocity, and  $a = 2$  m s<sup>-2</sup> is the acceleration of the CME. Measurement scatter is applied to the height-time points via a normally-distributed random number generator with a standard deviation of 10% of the model height at each time step. This is to simply convey the fact that the events that will be studied are transient phenomena that become increasingly difficult to discern in the images as their signal-to-noise ratios decrease and their structure often becomes disjoint. Each measurement is assigned an errorbar sufficient to overlap the true height-time profile. Even in this simple case of constant acceleration, various instances of randomised measurement scatter result in erroneous trends in the velocity and acceleration profiles, despite using the proper error treatment prescribed by the 3-point Lagrangian interpolation technique (Eq. (1)). The endpoint errors are derived



**Fig. 1.** Kinematic model for a CME with constant acceleration  $2 \text{ m s}^{-2}$  and initial velocity  $300 \text{ km s}^{-1}$ , and two data simulations of how the resulting profiles for different scatters on the height-time measurements (*top panels*) behave when 3-point Lagrangian interpolation is used to derive the velocities (*middle panels*) and accelerations (*bottom panels*). Both simulated cases were produced by adding scatter to the height measurements, via a normally-distributed random number generator with a standard deviation of 10% of the model height at each time-step. The different instances of scatter shown here produce completely opposing trends in the accelerations, with the errorbars failing to appropriately overlap the model, therefore belying the true trend.

from a weighting of the preceding or proceeding two data points and are therefore larger, reflecting the unknown nature of the trend beyond these points.

Two examples are shown in the left and right panels of Fig. 1, where completely opposing acceleration trends are determined for different samplings of the same height-time data set. This indicates that the nature of the scatter in the samples is not satisfactorily reflected in the derived kinematics and their associated errorbars. At the very least, the kinematic uncertainties should be expected to overlap the truth in each plot so that it remains a valid solution. A possible workaround is that instead of trusting the endpoints they simply be removed. Figure 1 would then show three data points for velocity and one data point for acceleration, reducing the biased trends. However, when dealing with low-number samples it would be better not to have to remove data points.

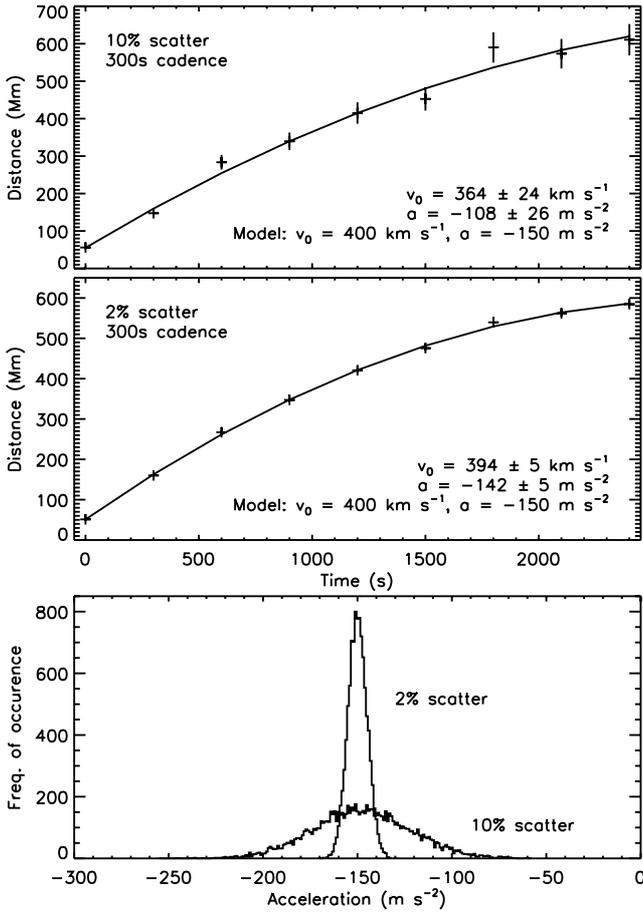
The effects of scatter on the distance-time measurements of a coronal wave were also examined, demonstrated here for the case of a constant acceleration event. The wave motion is modelled by Eq. (2), where  $r_0 = 50 \text{ Mm}$  is the initial distance of the wave from the source,  $v_0 = 400 \text{ km s}^{-1}$  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 distance-time measurements with random scatter added, shown here for  $3\sigma$  limits of 10% (top panel) and 2% (middle panel). Errorbars are applied with magnitude equal to the measurement scatter percentage. A second-order polynomial (quadratic) is then fit to each data set to test how the scatter affects the precision of the derived kinematics, even in this idealised case of knowing the underlying form of the data. The increased scatter

acts to smooth out the true kinematics, as demonstrated by the different distributions of derived accelerations from 10 000 runs of the simulation (bottom panel).

## 2.2. 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 Eq. (2) and Fig. 2). The simulated distances were sampled at cadences similar to the Solar & Heliospheric Observatory (SOHO; Domingo et al. 1995) Extreme-ultraviolet Imaging Telescope (EIT; 12 min) and SDO, (Pesnell et al. 2012) Atmospheric Imaging Assembly (AIA; 12 s), with a  $3\sigma$  scatter of 10% of the model distance. Errorbars are again applied with magnitude equal to the measurement scatter percentage. Figure 3 shows examples of these, with a quadratic fit to the sample measurements 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 determining the kinematics of a coronal wave.

In Fig. 4 the combined effects of measurement scatter and sampling cadence are simulated for the model coronal wave case. The image shows the result of plotting the distribution of acceleration fit parameters against all variations of scatter from 0–40%, at all variations of cadence from 1–720 s. Essentially the plots of the bottom panels of Figs. 2 and 3 represent slices through the corresponding locations of Fig. 4. This demonstrates



**Fig. 2.** Simulated distance-time measurements with  $3\sigma$  scatters of  $\pm 10\%$  (*top*), and  $\pm 2\%$  (*middle*) of the model value, at a fixed cadence of 300 s, and the resulting quadratic fit and  $v_0$  and  $a$  parameters. The reduced scatter increases the precision for obtaining the true kinematics, as demonstrated by the different distributions of derived acceleration fit parameters from 10 000 runs of the simulation (*bottom*).

that the reduction of measurement scatter and increase of sampling cadence together are required to improve the accuracy of the derived kinematics. However, measurement scatter is less problematic for higher cadence data sets.

We next simulate a non-constant acceleration profile for a typical fast CME via the following equations,

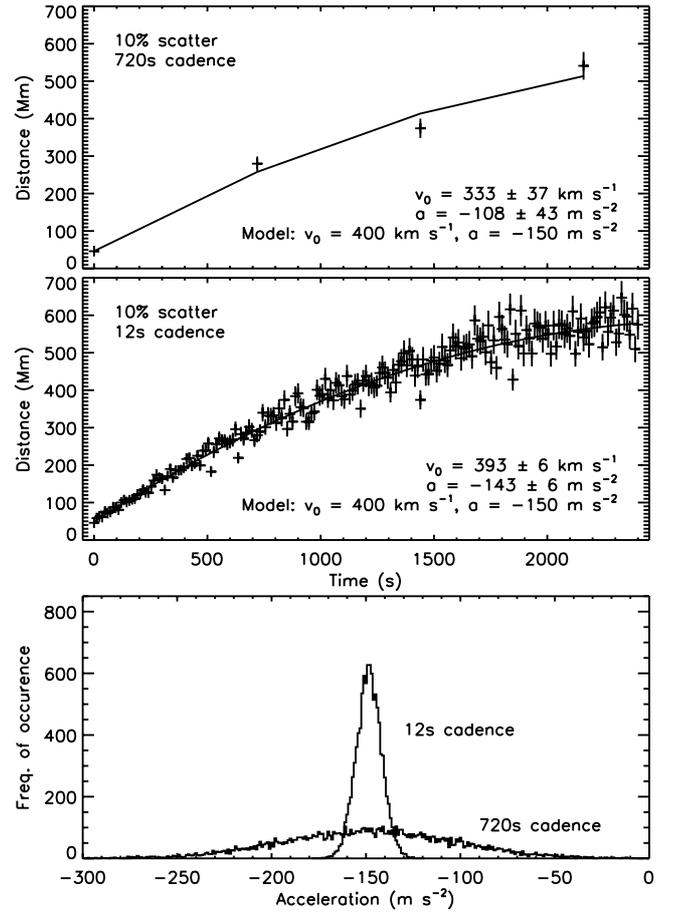
$$h(t) = \sqrt{2s} t \tan^{-1} \left( \frac{e^{t/2s}}{\sqrt{2s}} \right), \quad (3)$$

$$v(t) = \sqrt{2s} \tan^{-1} \left( \frac{e^{t/2s}}{\sqrt{2s}} \right) + \frac{e^{t/2s} t}{e^{t/s} + 2s}, \quad (4)$$

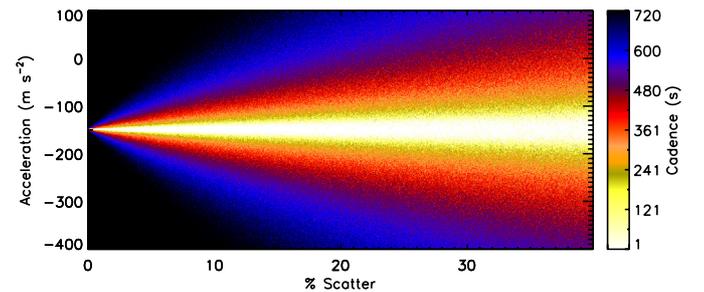
$$a(t) = \frac{e^{t/2s} (2s(t+4s) - e^{t/s}(t-4s))}{2s(e^{t/s} + 2s)^2}, \quad (5)$$

where  $s$  is a constant scaling factor. The acceleration profile exhibits an initial peak followed by a deceleration and then levels to zero. This is similar to a general impulsive CME that undergoes an initial high-acceleration eruptive phase before it decelerates to match the solar wind speed during its propagation phase. A model CME height-time profile is generated, enabling synthetic observation samples to be taken at different cadences.

We investigate the effect of observational cadence on the derivation of the kinematics and associated errorbars using the standard 3-point Lagrangian interpolation. In the first instance,

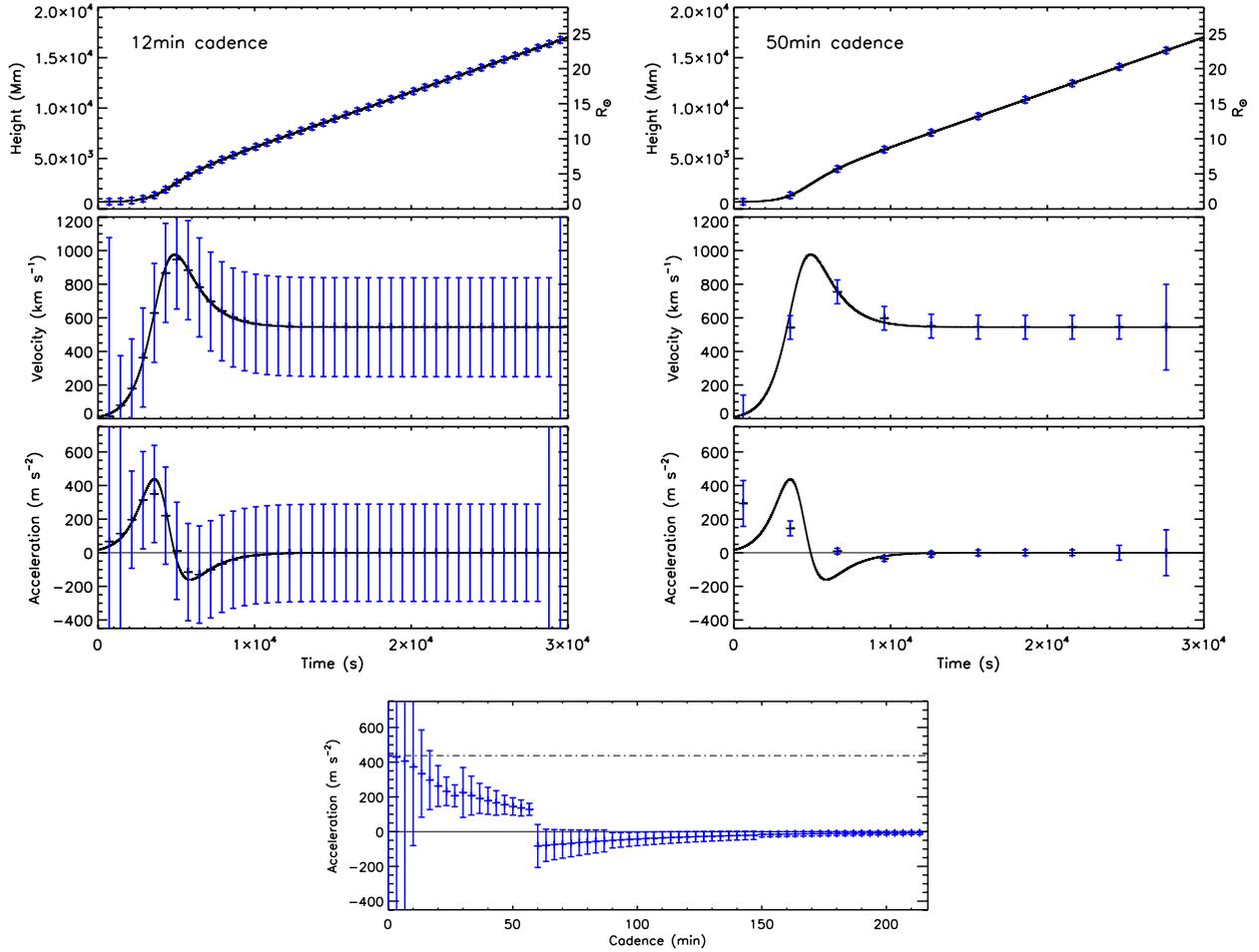


**Fig. 3.** Simulated distance-time measurements for sampling cadences of 720 s (EIT; *top*) and 12 s (AIA; *middle*), at a fixed scatter of  $\pm 10\%$ , and the resulting quadratic fit and  $v_0$  and  $a$  parameters. The increased cadence offers better precision for obtaining the true kinematics, as demonstrated by the different distributions of derived acceleration fit parameters from 10 000 runs of the simulation (*bottom*).



**Fig. 4.** Simulation of the derived accelerations from the model fits to Eq. (2) (with examples shown in Figs. 2 and 3), for scatters of 0–40% and cadences of 1–720 s. As shown, a decrease in both the data scatter and the cadence time improves the chances of obtaining the correct acceleration value, being  $-150 \text{ m s}^{-2}$  in this model coronal wave case.

fixed errorbars of  $\pm 300 \text{ Mm}$  are applied to the height-time measurements without any scatter. This is useful to simply test the effects of the cadence on the derived velocity and acceleration profiles and their associated errors. The left and right plots of Fig. 5 show the model height, velocity, and acceleration profiles sampled at cadences of 12 and 50 min, respectively. As the cadence is reduced (i.e., the time interval between observations is increased), the errorbars become smaller due to the inverse dependence of the Lagrangian error terms on the time between the data points  $\Delta t^{-2}$  (see Eq. (1)). However, reducing



**Fig. 5.** Demonstration of the effects of cadence on the error propagation according to the 3-point Lagrangian interpolation. A kinematic model for a CME with non-constant acceleration peaking at  $437 \text{ m s}^{-2}$  is tested for varying cadences, without any scatter. The *top left plots* show the height, velocity, and acceleration profiles for data sampled at 12 min cadence. The *top right plots* show the same data sampled at 50 min cadence. Note how the errorbars of the higher cadence measurements are counter-intuitively larger than the lower cadence measurements, even though the higher sampling rate better reveals the true kinematic trend. The *bottom plot* shows the derived peak acceleration against cadence, where the dot-dashed line indicates the true value of  $437 \text{ m s}^{-2}$ . The errorbars are shown to reduce in magnitude (implying greater precision) even though the derived acceleration at lower cadence is less accurate.

the cadence reduces the resolution at which the acceleration peak is detectable, and so the acceleration profile is less well characterised. Conversely, the errorbars become unrealistically large for very high-cadence observations even though the measurements better reveal the true trends of the kinematic profiles, as demonstrated in the bottom plot of Fig. 5. This fundamentally implies that the errorbars do not truly reflect the uncertainty on the data at a given cadence, and are in fact redundant for these cases.

It is clear that the variation in both scatter and imaging cadence can strongly influence the derived kinematics of a CME or coronal wave, and the commonly used 3-point Lagrangian technique does not return useful estimates of the associated uncertainty. Furthermore, the quantification of uncertainty on the physical measurements themselves is extremely non-trivial. Uncertainties exist that are attributable to the unknown physical mechanisms operating in the phenomena being studied, as well as the uncertainties involved in the techniques used for the analysis, making a robust error estimate practically impossible. This is especially true for automated routines designed to detect faint and transient phenomena such as CMEs and coronal waves. However, it is possible to use other techniques in an effort

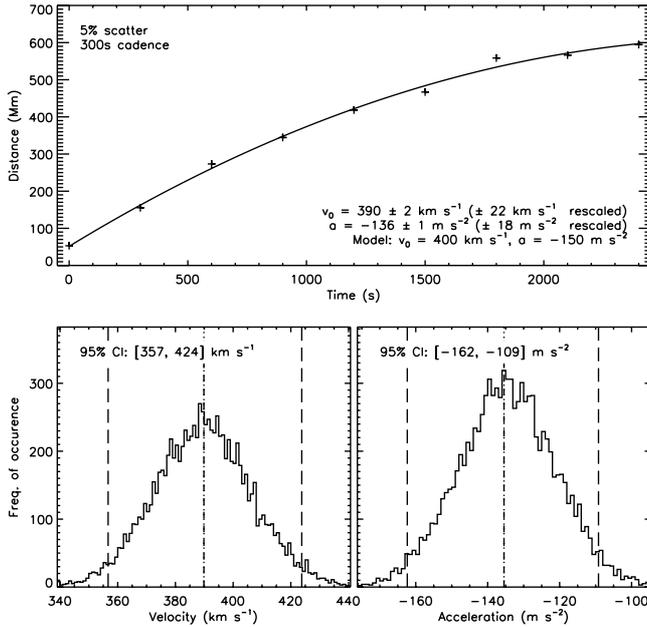
to overcome these issues and produce a more statistically sound method for dealing with these types of measurements.

### 3. Bootstrapping: a resampling method

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. Therefore, in order to approximate the behaviour of the true distribution, techniques based on resampling methods have been developed. These work by resampling the data enough times to generate a maximum likelihood estimator of the distribution. Bootstrapping is one such technique, first introduced by Efron (1979) and more recently described in, e.g., Efron & Tibshirani (1994) and Chernick (1999).

In the cases of CME and coronal wave observations, bootstrapping techniques can prove very useful for determining the uncertainty on the derived kinematic model parameters. The implementation of the residual resampling bootstrapping scheme is as follows:

1. An initial fit to the data  $y$  is obtained, yielding the model fit  $\hat{y}$  with parameters  $p$ .



**Fig. 6.** *Top panel:* initial fit of Eq. (2) to simulated coronal wave distance-time measurements, with 5% scatter and 300 s cadence. Errorbars of  $\pm 1.1$  Mm are assigned to the measurements, based on the 1.5 arcsec resolution of the imager. The fit parameters are quoted with  $1\sigma$  uncertainties, with the rescaled uncertainties shown in brackets. *Bottom panels:* histograms of the initial velocity and acceleration values derived using the bootstrapping technique. The mean and 95% confidence interval are indicated by the dot-dashed and dashed lines respectively. Bootstrapping provides a distribution of fitting parameters that is unattainable via a standard single-fit to data when unknown sources of uncertainty exist.

2. The residuals of the fit are calculated as  $\epsilon = y - \hat{y}$ .
3. The residuals are randomly resampled with replacement to give  $\epsilon^*$ .
4. The model is then fit to a new data vector  $y^* = \hat{y} + \epsilon^*$  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 to the simulated coronal wave moving with constant acceleration (Eq. (2)) and to the simulated CME moving with non-constant acceleration (Eq. (5)). In the case of the constant acceleration wave, the initial fit to the measurements and the bootstrapped distributions of initial velocity and acceleration values are shown in Fig. 6. Bootstrapping in this manner allows the determination of confidence intervals 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  $\alpha = 0.025$ ). Since unknown sources of error can exist that affect the measurement scatter, the only uncertainty that can be confidently attributed to the measurements is that which is due to the resolution limit of the data and similar quantifiable sources of uncertainty. Therefore, in an effort to avoid assigning incorrect uncertainties that will undermine the interpretation of any subsequent model-fit, the scatter of the data instead may be investigated via resampling methods.

For the simulated measurements in Fig. 6, the data points were assigned errorbars of  $\pm 1.1$  Mm to represent the lower limit of quantifiable uncertainty, based on the 1.5 arcsec resolution of AIA. If, as in this simulated coronal wave case, it may be assumed that the model should give an appropriate fit to the data,

and the measurement uncertainties are known to be too small (or alternatively too large), then the output uncertainties on the fit parameters can be rescaled accordingly, as per the following equation,

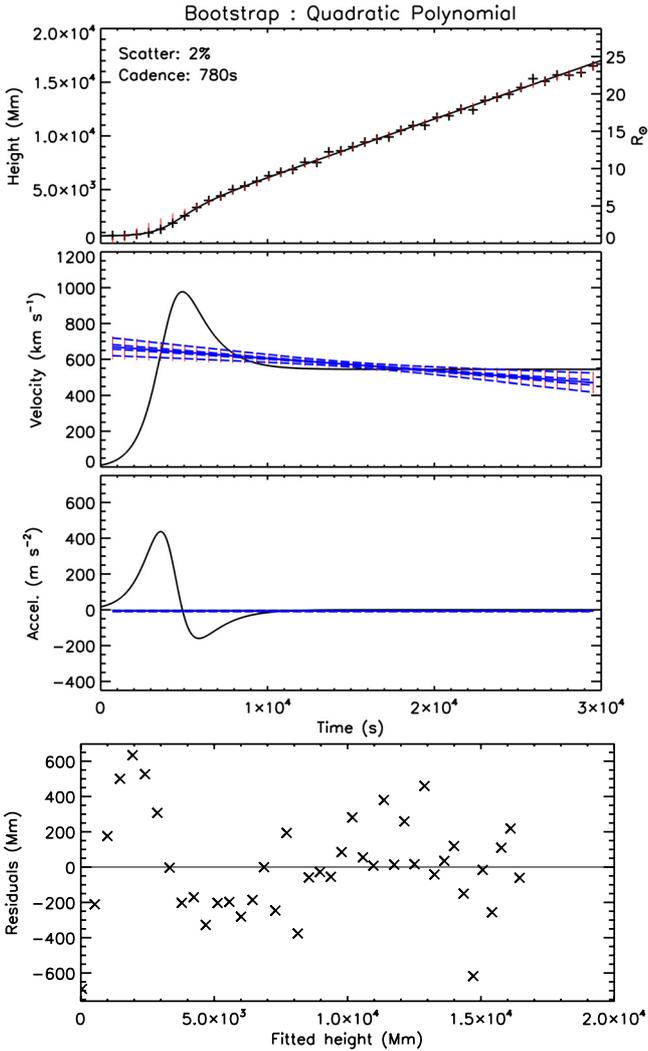
$$\sigma'^2 = \sigma^2 \frac{\chi^2}{\nu}, \quad (6)$$

where  $\chi^2$  is the standard measure of goodness-of-fit and  $\nu$  is the number of degrees of freedom in the fit (see, for example, [Bevington & Robinson 2003](#)). The rescaled uncertainties for this case are quoted in the top panel of Fig. 6. However, since in truth we do not know the exact kinematic form a CME or coronal wave should take nor the true uncertainty due to possible unknown sources of error, we cannot make such assumptions in our treatment of the real-data measurements. Therefore, the power of a bootstrapping technique is clear, by allowing an appropriate confidence interval to be assigned to the fit parameters. For the simulated measurements in Fig. 6, it is seen that the bootstrapped distributions for the initial velocity and acceleration do contain the true model values and have a range similar to the rescaled uncertainties of the single fit (i.e., likening the 95% confidence intervals to  $2\sigma$  uncertainty ranges, which are approximately double the rescaled  $1\sigma$  uncertainty ranges of the single fit).

Bootstrapping of a model cannot be applied blindly, as we demonstrate for a simple case of fitting the same constant acceleration form of Eq. (2) to the non-constant acceleration of Eq. (5). Figure 7 shows a model non-constant acceleration CME profile sampled at 780 s cadence with 2% scatter. The red points show the resulting distributions of points after the residuals are resampled with replacement. Thus a distribution of velocity and acceleration values are derived at each data point, with the corresponding median, interquartile range, and upper and lower fences overlaid in blue<sup>1</sup>. The second-order model is not appropriate to the true non-constant acceleration profile, as revealed by the trend in the residuals of the initial fit (bottom panel of Fig. 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 such that the residuals scatter randomly. Ideally, a piecewise function should be used to characterise 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. We must rely on some form of numerical derivative for revealing possible trends because we do not know the functional form that CME or coronal wave kinematics will have.

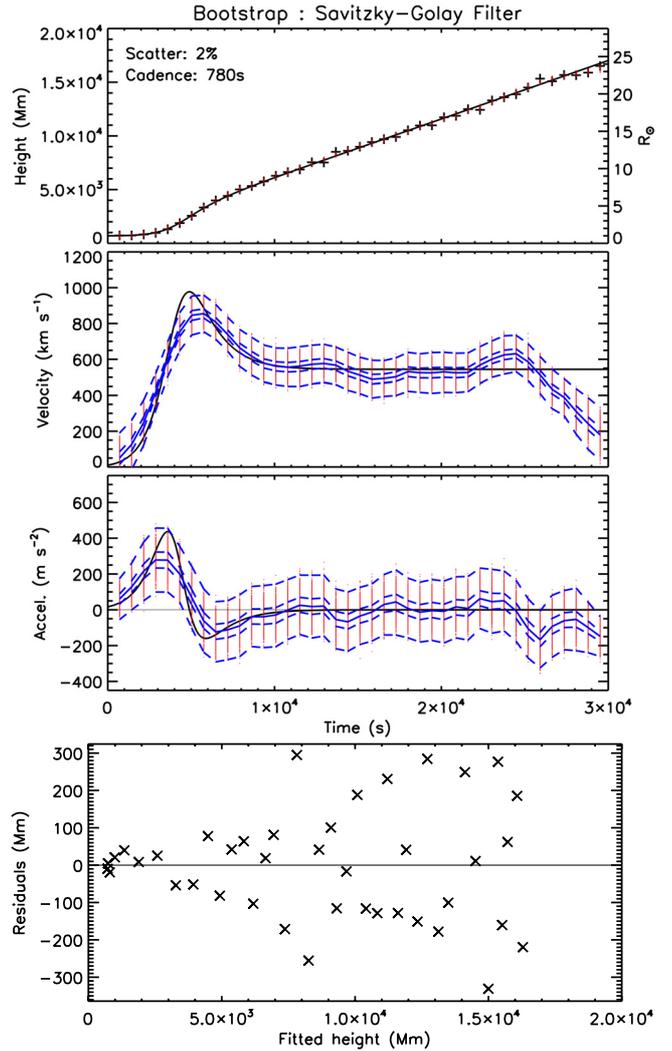
Since the issues with the 3-point Lagrangian have been highlighted in Sect. 2, we shall opt instead to implement the Savitzky-Golay filter ([Savitzky & Golay 1964](#)). This is a form of local polynomial regression of chosen degree of smoothing polynomial, and of chosen order to produce smoothed first 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. Therefore, in the case of CME and coronal wave distance-time measurements, the Savitzky-Golay filter is

<sup>1</sup> The interquartile range is the difference between the upper and lower quartiles of a dataset ( $IQR = Q_3 - Q_1$ ), determined by the 25th and 75th percentiles, with the second quartile  $Q_2$  being the 50th percentile, which is the statistical median of the dataset. Outliers are determined by the upper and lower fences of the dataset:  $lower\ fence = Q_1 - 1.5(IQR)$ ;  $upper\ fence = Q_3 + 1.5(IQR)$ .



**Fig. 7.** Bootstrapped second order polynomial fit to simulated CME height-time measurements, with 2% scatter and 780 s cadence. 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 dashed 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, indicating that the fit is not appropriate for the measurements.

a better method for smoothing small-scale scatter while still revealing the true kinematic profiles. This is illustrated in Fig. 8 for the same simulated non-constant acceleration CME profile used in Fig. 7, with the residuals resampled with replacement as per the bootstrapping technique described above. For this case, the neighbouring 3 points to the left and right of each data point were considered, and the filter applied with a chosen order of 2. 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. Note that the residuals in the bottom panel of Fig. 8 can be seen to scatter somewhat randomly, as desired. Also, for this realisation of the model 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 is not how the model is behaving. This is similar to how the intensity of a CME



**Fig. 8.** Bootstrapped Savitzky-Golay filter method applied to the simulated CME height-time measurements as in Fig. 7. This manner of piecewise fit smooths the measurements by fitting a polynomial to the 3 neighbouring points either side of each data point, and is successful in revealing the non-constant acceleration profile. The randomly scattered residuals also indicate its appropriateness.

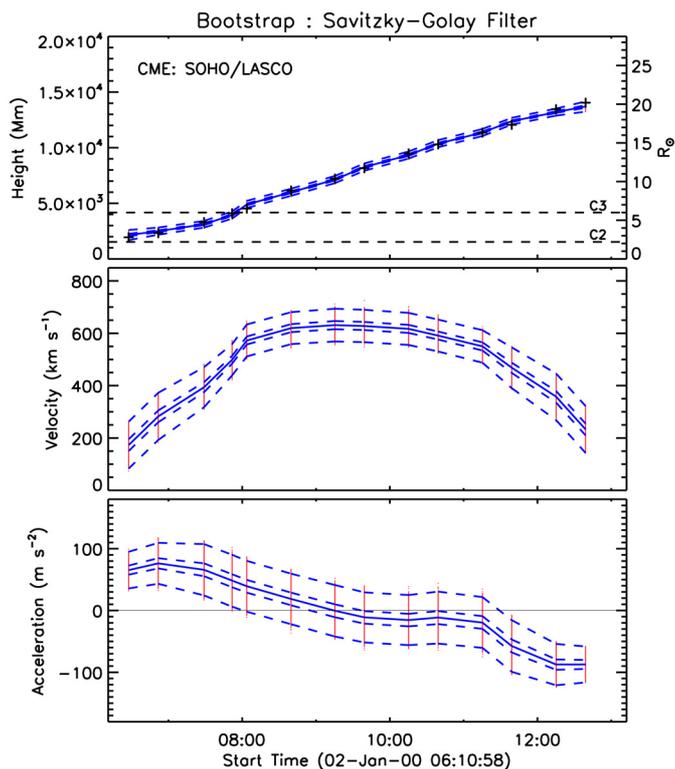
or coronal wave often lies too close to the background intensity at such distances, being lost to the noise and causing the measured profile to drop off. This alludes to considerations that must be made when dealing with automated systems of kinematic determination, whereby the algorithmic limitations can introduce systematic biases not accounted for in the derived kinematics and associated uncertainties. This is discussed further in the next section.

## 4. 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.

### 4.1. CME kinematics

First, we shall revisit a CME studied by Byrne et al. (2009) that was observed by SOHO/LASCO on 2000 January 2. In that study, the CME front edge was detected via multiscale

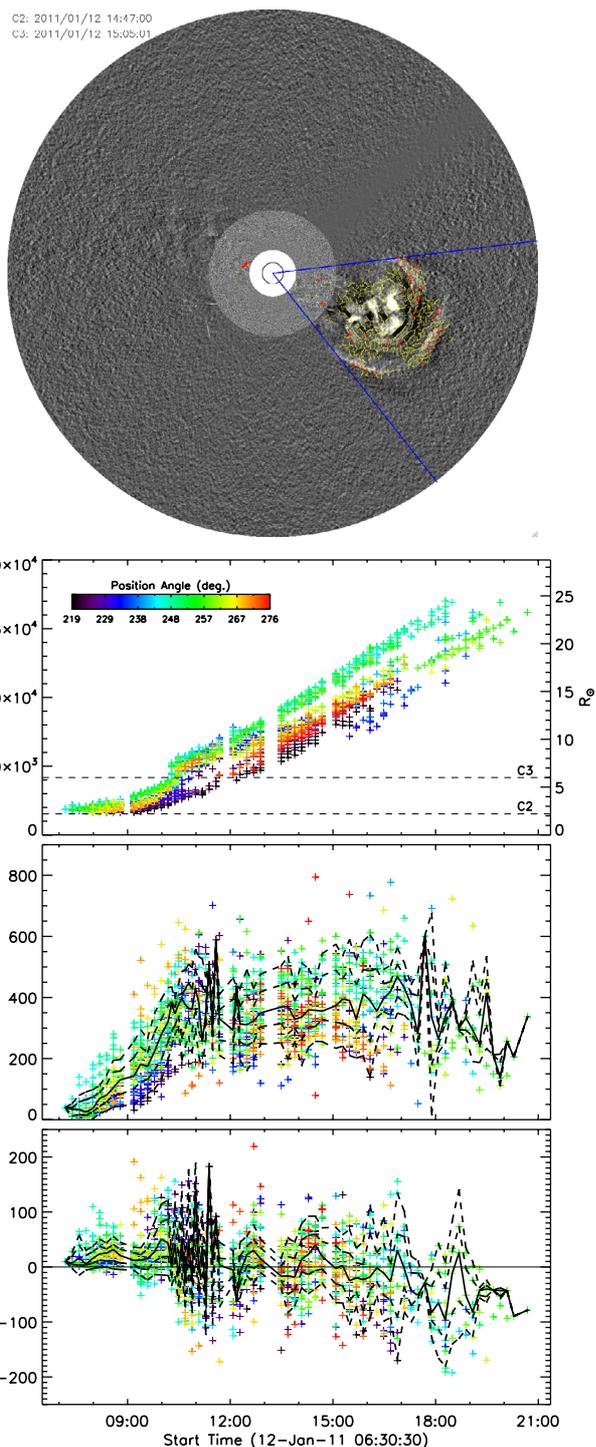


**Fig. 9.** Bootstrapped Savitzky-Golay filter method applied to a CME event observed by SOHO/LASCO on 2000 January 2, revisited from Byrne et al. (2009). The *top plot* shows the height-time measurements (plus symbols), the resampled residuals after applying the Savitzky-Golay filter (in red), and corresponding median, interquartile range, and upper and lower fences on the bootstrapped data (blue lines). The *middle and bottom plots* show the derived velocity and acceleration profiles.

methods and characterised with an ellipse-fit that was used to track changes to the CME front over time. The apex of the fit (furthest distance from Sun-centre) was measured in each frame and a height-time profile produced. This allowed an investigation into the kinematics of the event, derived using the 3-point Lagrangian method and associated error propagation formulation. However, as has been outlined in the preceding sections, this formulation is somewhat redundant for such a small sample size, and so a new method is applied to test the validity of the analysis.

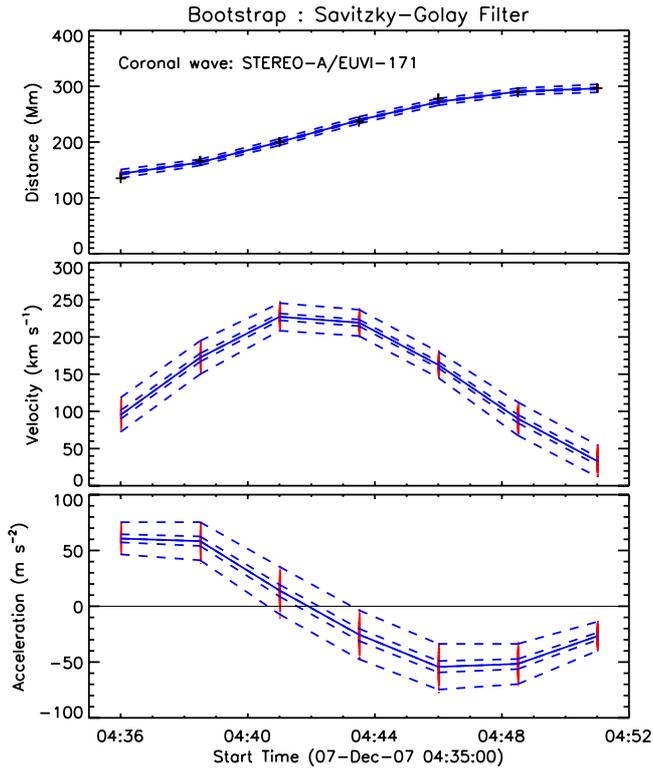
The method chosen, and deemed most appropriate from our investigations in this paper, is that of the Savitzky-Golay filter and bootstrapping technique. The filter width is set at 7 neighbouring data points (i.e., 3 points either side of the point in question). The top plot of Fig. 9 shows the height-time plot of the CME (plus symbols), the resampled residuals after applying the Savitzky-Golay filter (in red), and corresponding median, interquartile range, and upper and lower fences on the bootstrapped data (blue lines). The middle and bottom plots show the derived velocity and acceleration profiles. The variation of the velocity profile is consistent with that of Byrne et al. (2009), however the improvement offered by the approach here is the ability to investigate the acceleration profile, which was not possible to perform in Byrne et al. (2009) due to the unrealistically large errorbars resulting from the numerical differencing technique.

An important point to note is that a priori knowledge about such events and the manner in which they are tracked must



**Fig. 10.** Savitzky-Golay filter applied to the automated CORIMP CME detection and tracking of an event observed by SOHO/LASCO on 2011 January 12 (*top image*). The detected CME structure is highlighted in yellow, and the outermost height measurements indicated in red. The *top plot of the lower panels* shows the height-time measurements across the angular range of the CME (indicated by the colourbar). The *middle and bottom plots* show the derived velocity and acceleration profiles, with the median (solid line), interquartile range (inner dashed lines) and upper and lower fences (outer dashed lines) over-plotted.

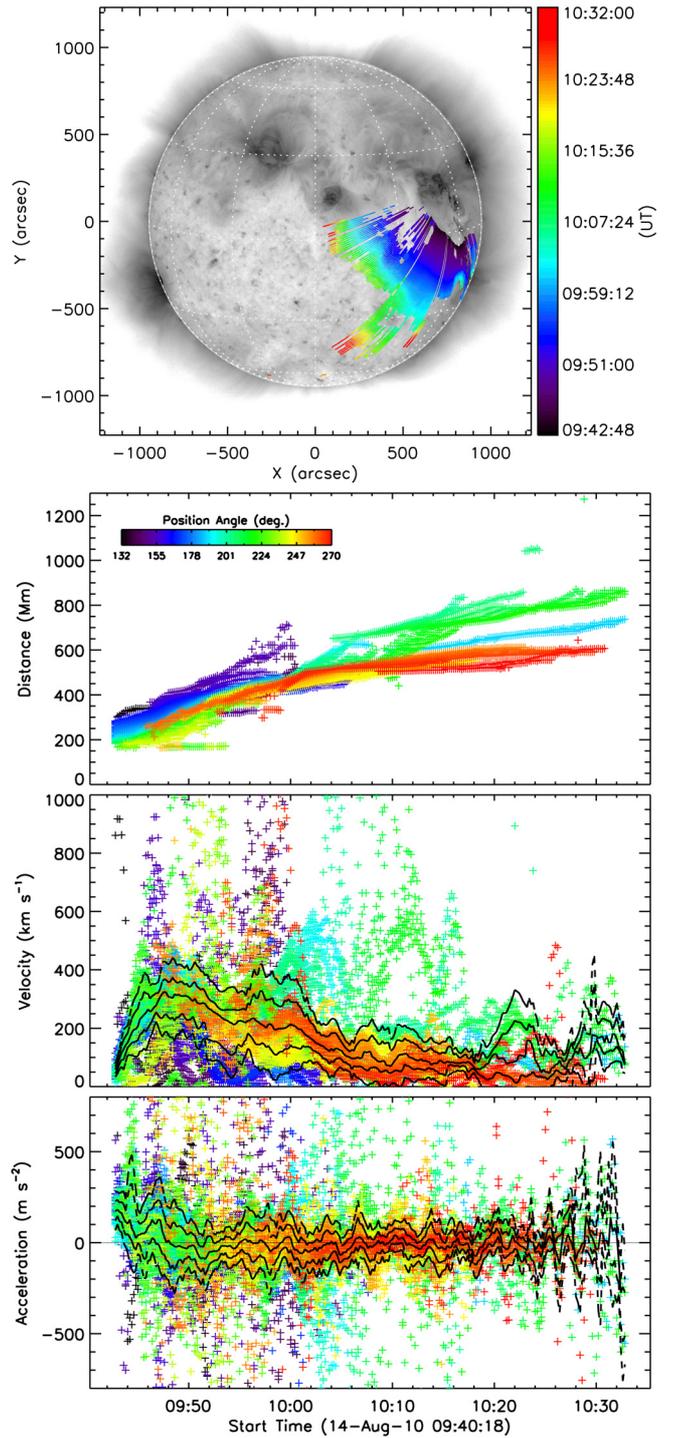
be called upon when interpreting the derived kinematic profiles. Edge effects in the derivation of the kinematics can be problematic at both ends of a time-series, though often less so at the start for these types of observations when the accuracy



**Fig. 11.** Bootstrapped Savitzky-Golay filter method applied to a coronal wave event observed by STEREO-Ahead/EUVI 171 Å on 2007 December 7, revisited from Long et al. (2011). The *top plot* shows the distance-time measurements (plus symbols), the resampled residuals (red points), and the median (solid line), interquartile range (inner dashed lines), and upper and lower fences (outer dashed lines). The *middle and bottom plots* show the corresponding velocity and acceleration profiles.

and precision are relatively high before the structure becomes too faint and/or disjoint. A caveat on any method for determining the kinematics is that the coverage of the observations may not be sufficient, either spatially or temporally, to confidently reveal a true trend in the profiles rather than succumbing to edge effects. For CMEs, the overlap between coronagraphs can cause discrepancies in the derived kinematics, and the outer edge of the field-of-view is inherently less reliable due to increased noise as the CME intensity falls-off. For this example, there appears to be a decrease in velocity (and negative acceleration) at the outer edge of C3, but by inspecting the resampled height-time measurements this may be deemed an artefact of the smoothing by the Savitzky-Golay filter when dealing with the scatter on the endpoints – an ongoing issue for small data sets. This strongly indicates the need for improved CME observations through increased imager cadences and greater field-of-view coverage, but otherwise a spread of measurements across the full span of a CME should be considered in current observations.

To this end, we next investigate a CME that occurred on 2011 January 2 as a case-study from the newly developed CORIMP CME catalogue (Morgan et al. 2012; Byrne et al. 2012). Height-time profiles are measured at 1 degree intervals across the angular span of the event as it moves through the corona in time (see top plot of Fig. 10). The Savitzky-Golay filter is then applied to each of these slices along the CME path, and the corresponding first and second order derivatives of the fit are determined. This reveals the spread of velocity and acceleration points as plotted in Fig. 10, with the median, interquartile range,



**Fig. 12.** Savitzky-Golay filter applied to the automated CorPITA detection and tracking of a coronal wave observed by SDO/AIA on 2010 August 14. The *top image* shows a percentage base differenced frame during the event, with an overlay of the detected wave motion in time (indicated by the top right colourbar). The *top plot of the lower panels* shows the distance-time measurements across the angular range of the coronal wave (indicated by the inset colourbar). The *middle and bottom plots* show the derived velocity and acceleration profiles, with the median (solid line), interquartile range (inner dashed lines) and upper and lower fences (outer dashed lines) over-plotted.

and upper and lower fences over-plotted in order to characterise the significance of the kinematic range. This allows the underlying trend of the data to be inspected, revealing an initial

acceleration of approximately  $20 \text{ m s}^{-2}$  up to a constant velocity of approximately  $400 \text{ km s}^{-1}$ . Large scatter can inevitably occur at the crossover of the two fields-of-view and at the outer edge of C3.

Since a spread of measurements across the CME can be determined in these cases, the sample size is therefore larger and the spread can be treated as the range of variation in the kinematics 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 centred 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. For the automated catalogue, avoiding the need for bootstrapping in this manner also saves on computing time. However, if a user of the catalogue has a specific model that they wish to test against the output, then a bootstrapping procedure and residual analysis would be of use, as demonstrated in Sect. 3. Any position angle may be chosen for such analyses, such as the central position angle that may correspond to the maximum speed of the CME. For example, the position angles around  $250^\circ$  in Fig. 10 indicate speeds up to approximately  $500 \text{ km s}^{-1}$ , so for these specific height-time measurements a model may be fit and a bootstrapping technique used to provide a confidence interval for testing the goodness-of-fit. Thus, the output produced from this treatment of the kinematics greatly improves the ability to study CME dynamics.

#### 4.2. Coronal wave kinematics

The treatment of coronal wave kinematics is very similar to that of CMEs, in that a small sample size of distance measurements is obtained, being prone to the effects of scatter from observational and algorithmic biases. We first revisit a coronal wave event studied by Long et al. (2011), that was observed by the Solar TERrestrial RELations Observatory (STEREO; Kaiser et al. 2008) Extreme-ultraviolet Imager (EUVI 171Å) on 2007 December 7. In that study, the coronal wave was detected via an intensity thresholding technique on percentage base-difference images. Distance-time measurements were produced and studied with a quadratic model of the form of Eq. (2). Here, we use the Savitzky-Golay filter and bootstrapping technique, as in the CME case study of Fig. 9. The filter width was reduced to operate on 5 neighbouring data points (i.e., 2 either side) since the sample size itself is so small, consisting of only 7 measurements. The top panel of Fig. 11 shows the distance-time plot of the coronal wave (plus symbols), the resampled residuals after applying the Savitzky-Golay filter (in red), and corresponding median, interquartile range, and upper and lower fences on the bootstrapped data (blue lines). The middle and bottom panels show the derived velocity and acceleration profiles. The kinematic trend strongly implies an initial acceleration of approximately  $60 \text{ m s}^{-2}$  followed by deceleration of  $-50 \text{ m s}^{-2}$  as it then approaches zero velocity. It is interesting that the velocity profile of Fig. 11 differs dramatically from that expected of a constant-acceleration kinematic profile, as was fit in Long et al. (2011). Care must still be taken when interpreting this trend from such a small data set, and such trends would be better justified with increased imager cadences, as is now possible with SDO/AIA.

The benefit of a higher cadence was demonstrated in Fig. 3 for the 12 second cadence of AIA and, since the resolution of the images is also significantly higher than previous missions, the

determination of the true kinematics should be greatly improved as a result (cf. Fig. 4). A coronal wave event observed by AIA, and tracked via the automated CorPITA algorithm, is shown in Fig. 12. The top image and upper plot show the distance-time measurements, and middle and bottom plots show the derived velocities and accelerations, respectively. The Savitzky-Golay filter was again deemed optimal for determining the kinematics of the event, at a filter width of 11 neighbouring data points (i.e., 5 either side). Similar to the previous CME case, this provides a spread of kinematics that is centred on the motion of the bulk of the event, with the median, interquartile ranges and upper/lower fences over-plotted. The trend indicates a possible initial acceleration of  $100\text{--}150 \text{ m s}^{-2}$ , attaining speeds of approximately  $300 \text{ km s}^{-1}$  before slowing down. The general limitation to be highlighted here is that the interpretation of the kinematics along any single position angle can result in different profiles that would imply different wave dynamics. A spread of measurements must be obtained in order to best characterise the eruption as a whole, noting that the scatter provides an insight into the variability of the results. Any position angle may be chosen for specific investigation of whether the detection technique has introduced erroneous trends, or is in fact revealing the different physics underlying the propagation of any given event.

## 5. Conclusions

We have demonstrated the unreliability of some common approaches to studying the kinematics of large scale dynamic features in the solar corona, such as CMEs and coronal waves. The dynamics of these phenomena are of great consequence to understanding the physics governing their initiation and propagation, and so the drawbacks of traditional numerical techniques for deriving their velocities and accelerations have been highlighted and efforts to overcome them discussed. Of particular note is the counter-intuitive behaviour of the error propagation in the standard 3-point Lagrangian interpolation, due to its inverse dependence on the observing cadence. It is also shown how strongly affected it can be by noise levels similar to those found in the data sets of these events. Therefore, we outline a two-fold approach.

Firstly, we suggest the use of bootstrapping techniques when dealing with the small sample sizes of these events. We demonstrate both their ease of implementation and usefulness in estimating the confidence intervals on fit parameters. It is an effective method for overcoming the limitation of having only a single set of observations of a phenomenon, which is the case for the dynamic events studied here. As a class of resampling methods, bootstrapping allows the construction of a number of resamples of the observed data from which a distribution of fitting parameters may be generated. Thus, the most likely parameter values and associated uncertainties may be determined. Uncertainty estimates calculated via a defined and appropriate procedure are crucial for determining the appropriateness of any chosen theoretical model seeking to characterise the physics of their motion. A robust estimate of the goodness-of-fit can be obtained when used in conjunction with residual analysis, as demonstrated in Figs. 7 and 8. This is because bootstrapping itself cannot be applied blindly, which leads us to the next point.

Secondly, we suggest that the events may be treated with a priori knowledge since there is a physical basis for their behaviour that can be confidently assumed to a certain extent. For example, the plasma is known to start from rest, and thus undergo an early acceleration to reach its maximum speed. What remain to be tested are the exact form of that initial acceleration

phase, whether or not unexpected trends or changes occur in its motion, and whether additional effects come into play (such as from drag and/or interactions with other features of the surrounding environs). The interpretation of the measurements must also allow for unknown biases that can occur, either due to observational uncertainties (e.g., low intensity profiles embedded in background noise) or algorithmic uncertainties (e.g., erroneous scatter derived at the boundary between instrument fields-of-view). Thus, while the uncertainty of a measurement might be mathematically sound, the measurement itself could be offset in a manner the user must aim to interpret appropriately in the derived kinematics, as demonstrated in Fig. 9.

Many of the issues of traditional numerical differencing techniques and standard error propagation can be overcome by the use of more appropriate methods. Thus, we may improve the determination of CME and coronal wave kinematics, and indeed this extends to any similar scenario of a dynamic event. We note that the significance attributed to the kinematic profiles must still be interpreted with care, relying on a priori knowledge so that none of these techniques be blindly applied or interpreted. This is essentially the idea behind Bayesian inference, wherein the probability of a hypothesis is updated as additional evidence is learned. Techniques derived from Bayesian statistics may prove useful for studying the kinematics of solar dynamic phenomena as observations and catalogues of their properties are built-up over time. For example, Butala et al. (2010) employ a Kalman filter in their development of a dynamic model for three-dimensional tomography of the solar corona. This particular type of algorithm operates recursively (over two steps of prediction and update) to build an estimate of the state of a system and associated uncertainties, having prior knowledge of the state of the system. Such approaches may hold promise if they can be adapted for the treatment of solar event kinematics.

*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. SDO data supplied courtesy of the NASA/SDO consortia. The STEREO/SECCHI project is an international consortium of the Naval Research Laboratory (USA), Lockheed Martin Solar and Astrophysics Lab (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 Liège (Belgium), Institut d'Optique Théorique et Appliquée (France), and Institut d'Astrophysique Spatiale (France). The authors would like to acknowledge the help and guidance from the SDO Feature Finding Team. DSB was supported by a Marie Curie Intra-European Fellowship under the European Community's Seventh Framework Programme (FP7) and the ESA Prodex program. D.L. would like to thank the Institute for Astronomy at the University of Hawaii for their generous hospitality during his visit. The authors thank the anonymous referee for their valuable comments.

## References

Ballai, I., Erdélyi, R., & Pintér, B. 2005, *ApJ*, 633, L145  
 Bentley, R. D., Csillaghy, A., Abouadarham, J., et al. 2011, *Adv. Space Res.*, 47, 2235

Bevington, P. R., & Robinson, D. K. 2003, *Data reduction and error analysis for the physical sciences* (McGraw-Hill Higher Education)  
 Butala, M. D., Hewett, R. J., Frazin, R. A., & Kamalabadi, F. 2010, *Sol. Phys.*, 262, 495  
 Byrne, J. P., Gallagher, P. T., McAteer, R. T. J., & Young, C. A. 2009, *A&A*, 495, 325  
 Byrne, J. P., Maloney, S. A., McAteer, R. T. J., Refojo, J. M., & Gallagher, P. T. 2010, *Nat. Comm.*, 1, 74  
 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)  
 Colaninno, R. C., & Vourlidas, A. 2006, *ApJ*, 652, 1747  
 dal Lago, A., Schwenn, R., & Gonzalez, W. D. 2003, *Adv. Space Res.*, 32, 2637  
 Domingo, V., Fleck, B., & Poland, A. I. 1995, *Sol. Phys.*, 162, 1  
 Efron, B. 1979, *Ann. Stat.*, 7, 1  
 Efron, B., & Tibshirani, R. J. 1994, *An Introduction to the Bootstrap* (Chapman & Hall/CRC)  
 Emslie, A. G., Kucharek, H., Dennis, B. R., et al. 2004, *J. Geophys. Res. (Space Physics)*, 109, 10104  
 Gallagher, P. T., & Long, D. M. 2011, *Space Sci. Rev.*, 158, 365  
 Gallagher, P. T., Young, C. A., Byrne, J. P., & McAteer, R. T. J. 2011, *Adv. Space Res.*, 47, 2118  
 Gopalswamy, N., Lara, A., Lepping, R. P., et al. 2000, *Geophys. Res. Lett.*, 27, 145  
 Howard, T. 2011, *Coronal Mass Ejections*, *Astrophys. Space Sci. Lib.*, 376  
 Howard, T. A., & Tappin, S. J. 2005, *A&A*, 440, 373  
 Hurlburt, N., Cheung, M., Schrijver, C., et al. 2012, *Sol. Phys.*, 275, 67  
 Kaiser, M. L., Kucera, T. A., Davila, J. M., et al. 2008, *Space Sci. Rev.*, 136, 5  
 Kontar, E. P., & MacKinnon, A. L. 2005, *Sol. Phys.*, 227, 299  
 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  
 Long, D. M., Gallagher, P. T., McAteer, R. T. J., & Bloomfield, D. S. 2011, *A&A*, 531, A42  
 Ma, S., Wills-Davey, M. J., Lin, J., et al. 2009, *ApJ*, 707, 503  
 Martens, P. C. H., Attrill, G. D. R., Davey, A. R., et al. 2012, *Sol. Phys.*, 275, 79  
 Morgan, H., Byrne, J. P., & Habbal, S. R. 2012, *ApJ*, 752, 144  
 Patsourakos, S., & Vourlidas, A. 2012, *Sol. Phys.*, 93  
 Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, *Sol. Phys.*, 275, 3  
 Podladchikova, O., Vuets, A., Leontiev, P., & van der Linden, R. A. M. 2012, *Sol. Phys.*, 276, 479  
 Prangé, R., Pallier, L., Hansen, K. C., et al. 2004, *Nature*, 432, 78  
 Riley, P. 2012, *Space Weather*, 10, S02012  
 Savitzky, A., & Golay, M. 1964, *Anal. Chem.*, 36, 1627  
 Schrijver, C. J., & Siscoe, G. L. 2010, *Heliophysics: Space Storms and Radiation: Causes and Effects* (Cambridge, UK: Cambridge University Press)  
 Schrijver, C. J., Elmore, C., Kliem, B., Török, T., & Title, A. M. 2008, *ApJ*, 674, 586  
 Schrijver, C. J., Aulanier, G., Title, A. M., Pariat, E., & Delannée, C. 2011, *ApJ*, 738, 167  
 Schwenn, R., dal Lago, A., Huttunen, E., & Gonzalez, W. D. 2005, *Ann. Geophys.*, 23, 1033  
 Shen, Y., & Liu, Y. 2012, *ApJ*, 754, 7  
 Temmer, M., Veronig, A. M., Kontar, E. P., Krucker, S., & Vršnak, B. 2010, *ApJ*, 712, 1410  
 Temmer, M., Vršnak, B., Rollett, T., et al. 2012, *ApJ*, 749, 57  
 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, *Liv. Rev. Sol. Phys.*, 9, 3  
 Wen, Y., Maia, D. J. F., & Wang, J. 2007, *ApJ*, 657, 1117  
 Yashiro, S., Gopalswamy, N., Michalek, G., et al. 2004, *J. Geophys. Res. (Space Physics)*, 109, 7105  
 Young, C. A., & Gallagher, P. T. 2008, *Sol. Phys.*, 248, 457  
 Zhang, J., & Dere, K. P. 2006, *ApJ*, 649, 1100  
 Zheng, R., Jiang, Y., Yang, J., et al. 2012, *ApJ*, 753, 112