Numerical methods for determining the kinematics of CMEs and "EIT Waves"

J.P. Byrne[1], D.M. Long[2], P.T. Gallagher[3], S.A. Maloney?[4], S.D. Bloomfield?[3], H. Morgan[5, 1], and S.R. Habbal[1]

[1] Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA.
e-mail: `jbyrne@ifa.hawaii.edu`
[2] UCL–Mullard Space Science Laboratory, Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK.
[3] School of Physics, Trinity College Dublin, College Green, Dublin 2, Ireland.
[4] Skytek, 51/52 Fitzwilliam Square West, Dublin 2, Ireland
[5] Institute of Mathematics and Physics, Aberystwyth University, Ceredigion, SY23 3BZ, UK.

Received ?; accepted ?

ABSTRACT

Context. Why are we doing this work?
Aims. The aims of the paper
In this paper we show that traditional techniques for the determination of CME and "EIT waves", 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. The methods used in the paper
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. The results of this analysis
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

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 $f(t)$:

$$f(t_0 + \delta t) \; = \; f(t_0) + f'(t_0)\delta t + \frac{f''(t_0)}{2!}(\delta t)^2 + ... \qquad (1)$$

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:

$$y_{n+1} \; = \; y_n + hf(t_n, y_n) \qquad (2)$$

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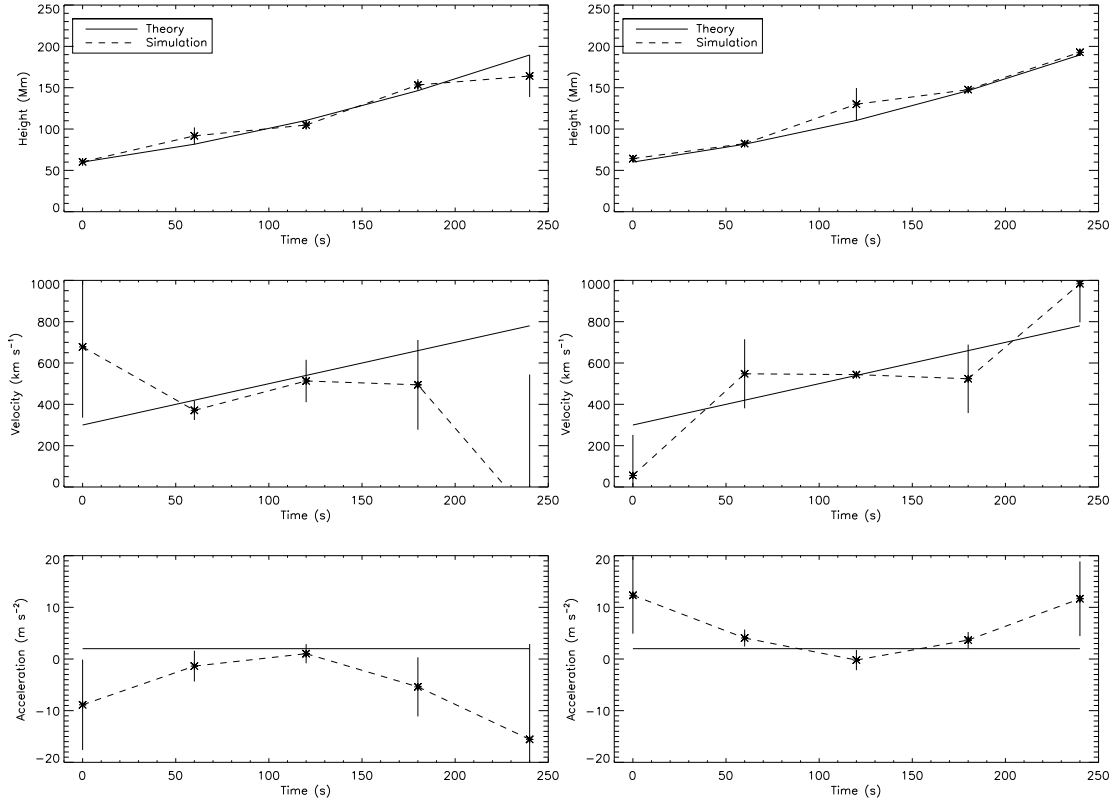
Fig. 1. A theoretical model for a CME with constant acceleration 2 m s^{-2} and initial velocity 300 km s^{-1}, and two simulations of how the resulting profiles for a noisy sample of data–points behave using 3-point Lagrangian interpolation.

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

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

Specifically in the case of kinematic analyses, this is used to propagate the errors on the height-time data $r(t)$ into the velocity $v(t)$ and acceleration $a(t)$ profiles to determine the associated uncertainties. In the case of height-time data the covariance terms are zero because the quantities are uncorrelated.

In the case of relatively low sampling of the data, as in the case of coronagraph observations of CMEs, 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 timestep, 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 timestep, 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 δt is not constant, and so the 3-point Lagrangian interpolation technique is used for such cases (and gives the same result as the centre-difference otherwise, but includes the endpoints). The Lagrangian interpolation polynomial is given by:

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

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

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 for acceleration) the associated error is given by:

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

with the endpoint errors derived from a weighting of the pro-/preceding two data–points, that is therefore larger to reflect the uncertainty of the trend beyond the sample points.

Although the 3-point Lagrangian is mathematically sound, its application to CME data has proved problematic due to the limitations of the data. The main drawback is that the noise level across the few, low-cadence measurements, often belies the actual trends of the data such that the numerical derivations of velocity and acceleration become untrustworthy and even misleading. As an example of this, we simulate a height-time profile of a CME that propagates with a constant acceleration of 2 m s^{-2} and initial velocity 300 km s^{-1}, and add random noise of various levels up to a maximum of $\sim$20%. An errorbar on each datapoint is determined by its distance from the true height-time profile, to represent the worst best-case scenario wherein all measurements lie on the very edge of the known error. Various instances of 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. Two such cases are shown in Figure 1 where completely opposing acceleration trends are revealed, indicating that the nature of the scatter in the samples is not satisfactorily reflected in the

derived kinematics and associated errors. The worry here is that for CMEs undergoing non-constant acceleration, the resulting trends in the derived profiles may be completely intractable via these methods. One possible assurance has been that instead of trusting the endpoints, even given their larger errorbars, they be removed. Figure 1 would then show three datapoints for CME velocity and one for acceleration, that would in the left case go somewhat toward removing the biased trend and implying a constant acceleration, while in the right case it would remain inconclusive. So qualitatively this might achieve a more accurate result and avoid assigning erroneous trends, especially to fast CMEs for which there are few data–points.

Beyond the interest certainly in the initial acceleration from the forces that drive a CME to erupt, there is a real physical motivation to resolve the subsequent low or zero acceleration as precisely as possible. This is important for determining whether the predominant forces acting are such that the CME is, in general, undergoing continued driving (internal) forces, or positive or negative drag (external) forces, or most assuredly some interplay of both. Any changes to its acceleration that result from different phases of dominating force, and where or why this occurs, are of great interest. Yet this all remains somewhat elusive given the limitations of current CME data, especially as Wen et al. (2007) demonstrate that the errors in the acceleration values can be of the same order as the accelerations typically measured.

We investigate the output of the recently developed CORIMP catalogue with new techniques (and stringent error analyses) for deriving CME kinematics.

3. Simulations

3.1. Kinematic models

A theoretical CME with non-constant acceleration is modelled by the following kinematic profile equations (as used in Byrne et al. 2012, based on a variation of the acceleration function chosen by Gallagher et al. 2003):

$$h(t) = \sqrt{2x}\, t \tan^{-1}\left(\frac{e^{t/2x}}{\sqrt{2x}}\right) \qquad (7)$$

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

$$a(t) = \frac{e^{t/2x}\left(2x(t+4x) - e^{t/x}(t-4x)\right)}{2x\left(e^{t/x} + 2x\right)^2} \qquad (9)$$

where x is just a scaling factor, set to be constant. The resulting acceleration profile exhibits an initial peak followed by a deceleration and then levelling 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 model CME height-time profile is generated enabling synthetic observation samples to be taken at different cadences from $1 - 47$ mins.

Simulating the kinematics of the on–disk "EIT Wave" poses a slightly different problem as the theoretical interpretation of this phenomenon is not yet fully codified. As a result, two distinct kinematic forms were simulated here: a quadratic model and a power–law model. Both of these models are typically used to determine the kinematics of
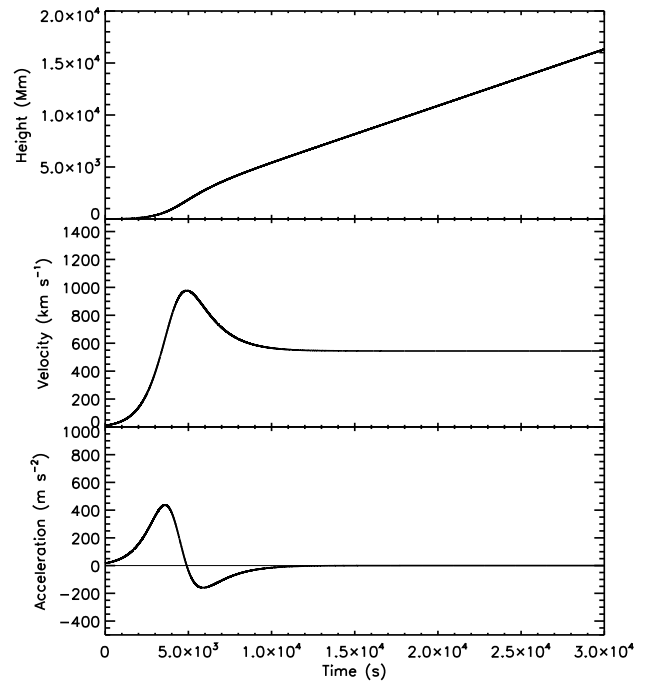

Fig. 2. The height, velocity and acceleration profiles for a theoretical event with non-constant acceleration peaking at 437 m s^{-2}.

these disturbances and have been used extensively in previous work (e.g., Long et al. 2008; Muhr et al. 2011; Veronig et al. 2008; Warmuth et al. 2004). The quadratic model takes the form,

$$r(t) = r_0 + v_0 t + \frac{1}{2} a_0 t^2 \qquad (10)$$

where r_0 is the initial distance of the pulse from the source, v_0 is the initial velocity of the pulse and a_0 is the acceleration of the pulse. This may then be used to identify the situation where the pulse has constant acceleration (either positive or negative) or zero acceleration (where the error associated with the acceleration term passes through zero). The power–law fit may be used to identify variable acceleration and is of the form,

$$r(t) = c(t - t_0)^\delta \qquad (11)$$

where c is a constant, t_0 is the starting time given by the fit and δ is the exponent.

3.2. Testing for non-constant acceleration with the 3-point Lagrangian

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 interpolation. In the first instance fixed 1σ errorbars of approximately ± 25 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. Figure 3 shows a sample output of the test. Figures 3(a) and (b) show the model CME acceleration profile sampled at cadences of 3 and 10 mins respectively, relative to the initial point at 0 mins (blue) held fixed

while the varying cadence changes the locations of the other sample points along the profile. Note that the errorbars of the two points at both the start and end of the interval are larger than the rest to account for not having enough points to interpolate across them (the endpoints of velocity having already undergone the same treatment). 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 treatment on the time between points (Δt^{-2}). However, reducing the cadence reduces the resolution with which it is able to detect the acceleration peak, and so the peak is smoothed out until the interval is eventually approximated as being one of constant zero acceleration, as demonstrated in Figure 3(d) (where the non-smooth transition and errorbar sizes are as a result of stepping down in sample size such that the fixed point at 9 mins eventually becomes the start-point of the interval). Figure 3(c) shows the trend of derived acceleration on the fixed point at 0 mins for increasing cadence time, corresponding directly to the start-point (blue) of the top two plots (a) and (b). In this case, as the cadence time is increased the derived acceleration value at the fixed 0 mins point approaches the true acceleration peak as the points average across the general peaked trend, before it is smoothed out and the interval is again approximated as having constant zero acceleration. Figures 3(c) and (d) clearly show the effect of the Δt^{-2} dependence of the errorbars derived through the 3-point Lagrangian interpolation, wherein the errors would become infinitely large for infinitely high cadence, and imply erroneously high precision for very low cadence sampling of the data. The importance of where along the profile the sample points are taken is also highlighted here.

It is also noted that the trend of the acceleration profile at high-cadence samplings is also revealed by the errorbars' span across the interval, as is shown in Figure 3(a) for example. This fundamentally implies that the errors do not truly reflect a 1 σ uncertainty on the data in this case, but recall here that no noise has been added to the dataset. The addition of noise, even to only a small degree, creates large problems for derivatives.

3.3. Bootstrapping

Bootstrapping was first proposed by Efron (1979) as a general interpretation of the jackknife, allowing statistical quantities to be estimated from a limited sample to a high degree of accuracy. In principle this allows a statistical quantity (such as the mean, variance etc.) of a distribution to be estimated using a small sample of measurements from that distribution. The resulting estimates are statistically significant, with statistically defined associated error estimates.

While there are a number of different bootstrapping techniques, each of which is useful for a different purpose, in this case we will be using the residual resampling technique. Whereas the majority of bootstrapping techniques involve removing one point at random and recalculating the fit to the data, this is inappropriate in this case due to the small sample set available. The residual resampling technique however, does not require data-point removal, making it more appropriate for small data-sets such as those available here.

The residual resampling technique works as follows. The distance-time measurements (y_i; $i = 1, 2, \ldots, n$) are first
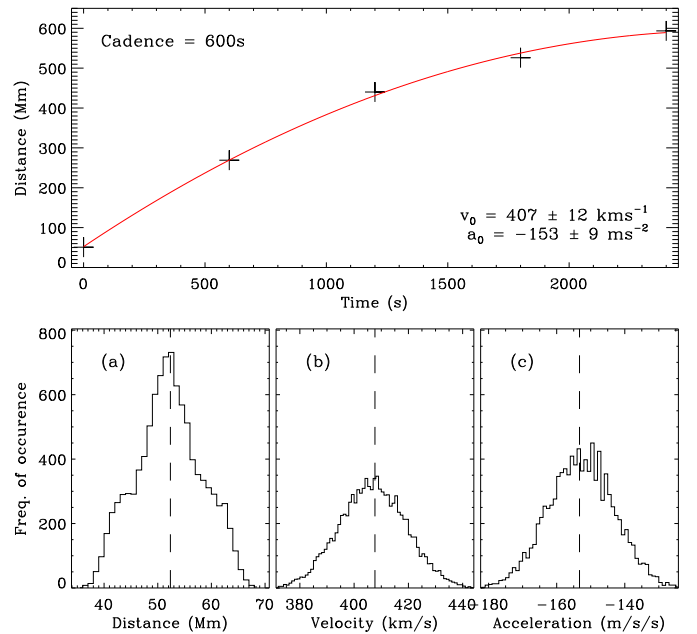

Fig. 4. *Top*: Simulated data (asterisks) with bootstrapped fit shown in red and parameters in bottom right. *Bottom*: Histograms showing the distance (a), velocity (b) and acceleration (c) derived using the bootstrapping technique.

fitted using a quadratic equation of the form

$$r(t) = r_0 + v_0 t + \frac{1}{2}at^2, \tag{12}$$

where r_0 is the offset distance, v_0 is the initial velocity, a is the constant acceleration of the pulse, and t is the time elapsed since the first observation of the disturbance. This yields the fitted values ($\hat{y}_i$) and residuals ($\hat{\epsilon} = y_i - \hat{y}_i$) for each point. The residuals are then randomly ordered and randomly assigned a sign (1 or -1), before being added to the original fit values to produce a new data set. These new data are then fitted using the same model, with the resulting re-fitted parameters recorded. The process is then repeated a large number of times ($\sim$10 000).

This has the effect of repeating the fit to slightly varying data a large number of times, allowing a distribution of fit values to be obtained. The mean and standard deviation of this distribution then provide the estimated value and associated error of the parameter. This technique is statistically rigorous and produces a more accurate result than a simple model fit to the given data.

This technique was applied to the simulated data–set used in Section **??** to allow a comparison of the effectiveness of the bootstrapping approach with the numerical differencing techniques. The results of this analysis are shown in Figure 4. The upper panel here shows the simulated data (indicated by the asterisks) with the bootstrapped fit to the data shown by the red line. The kinematics obtained by the bootstrapping approach are given in the bottom right. The bottom panels (a), (b) and (c) show the histograms for the distance, velocity and acceleration parameters respectively. These histograms returned by the bootstrapping approach for each parameter allow a statistical determination of the fit parameters, a fact reflected in the error term for the derived kinematics.
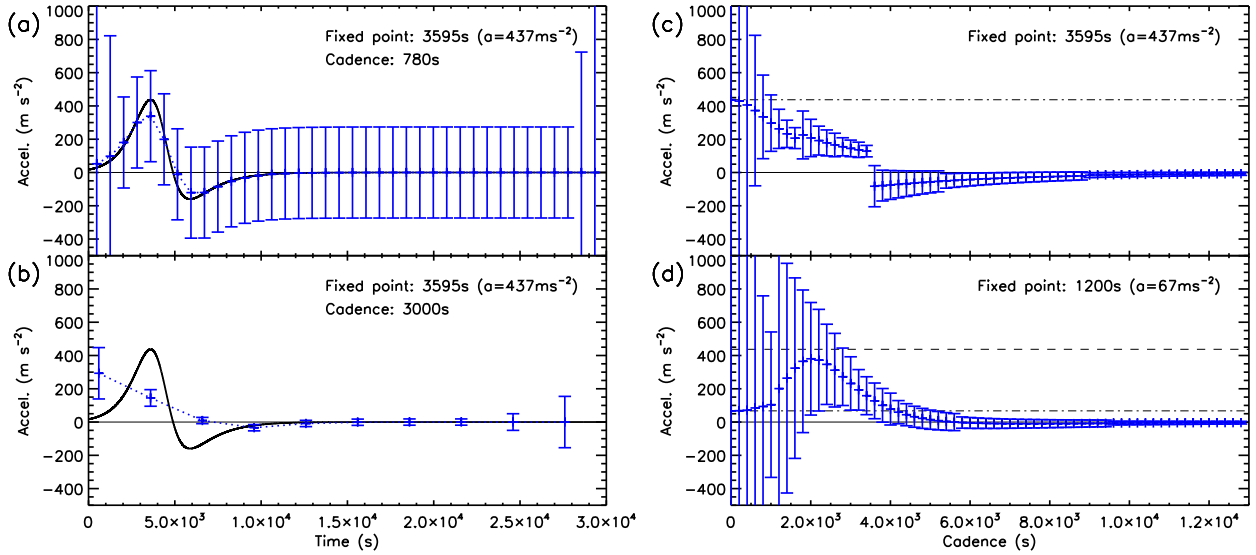
Fig. 3. A demonstration of the effects of cadence on the error propagation according to the 3-point Lagrangian interpolation. A theoretical model for a CME with non-constant acceleration peaking at 437 m s^{-2} is tested for varying cadences of 1 to 47 mins. Plots (a) and (b) show the cases when the cadence is 3 mins and 10 mins respectively, relative to a fixed point at the 0 mins frame (start-point). Plots (c) and (d) show the derived acceleration values and associated errorbars as a function of cadence, relative to fixed points at 0 mins and 9 mins (peak acceleration point). The dotted line represents the peak acceleration value, and the dashed line represents the true acceleration expected in each case (which is the peak acceleration anyway for the case of the fixed point at 9 mins).

The kinematics returned by the bootstrapping approach match the model kinematics within one standard deviation, while also producing an accurate estimate with well–defined errors, unlike those of the numerical differencing techniques. It should also be noted that none of the data–points have been removed, unlike the majority of the numerical differencing techniques tested.

As a result of the simulations carried out to test the different approaches to determining the kinematics of a CBF pulse, the residual resampling bootstrapping technique was chosen as the best method. This fits a model directly to the distance–time measurements, allowing the kinematics of the CBF pulse to be determined to a high degree of accuracy. The kinematics given for each future event studied here therefore reflects the mean value of the bootstrapped distribution, with the associated error given by the standard deviation.

3.4. Inversion Techniques?

4. Sources of error?

4.1. Cadence effects

In the absence of a numerical technique to determine the kinematics of a CBF pulse, the cadence of the observing instrument and the degree of uncertainty in the data must also be considered. Here, the influence of both cadence and uncertainty are examined using the simulated data discussed in Section **??**. The effects of varying image cadence were the first to be examined; this is shown in Figures 5 and 6.

The same data–set was used for each of the panels shown in Figure 5, with the cadence varied to best reflect data from *SOHO*/EIT (top panel at 720 s), *STEREO*/EUVI (middle panel at 300 s) and *SDO*/AIA (bottom panel at 12 s). The
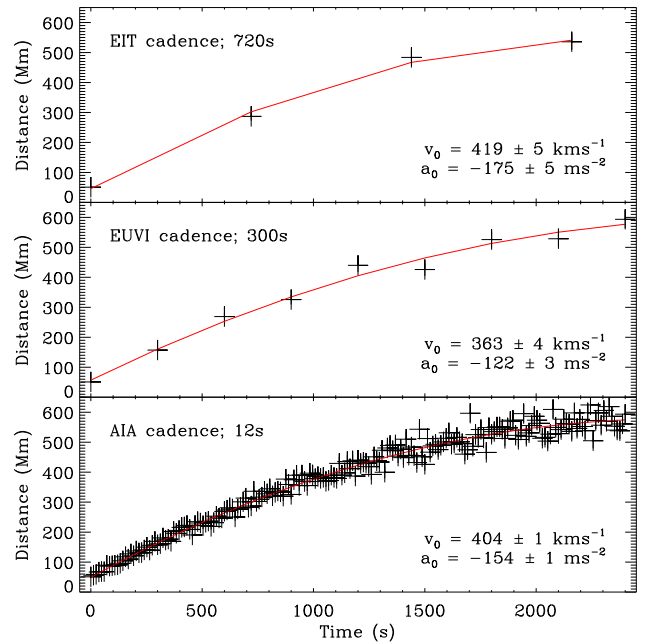

Fig. 5. Simulated data (crosses) for cadences of 720 s (*top*; comparable to EIT), 300 s (*middle*; comparable to EUVI) and 12 s (*bottom*; comparable to AIA). In each case, the fit to the data is shown in red with derived kinematics in the bottom right of each panel.

derived kinematics are given in the bottom right of each panel with the fit to the data shown by the red line. It is immediately apparent that the higher cadence data allows a better fit to the data despite the uncertainty in the data (which was kept at ±5% of the model for each data–set).

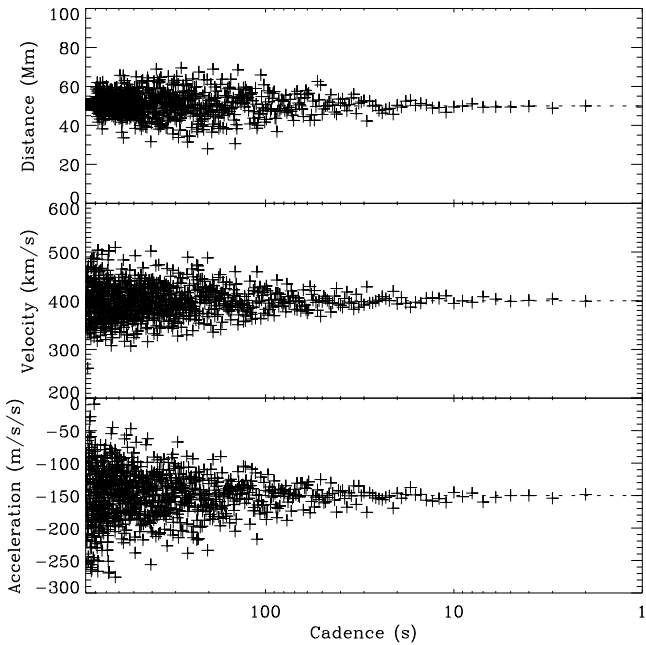**Fig. 6.** Derived kinematics for varying image cadence with ±5% uncertainty. Distance, velocity and acceleration are shown in the top, middle and bottom panels respectively, with the x–axis shown using logarithmic scaling to highlight the effects of varying the image cadence.

Fig. 7. Simulated data (crosses) for noise distributions with widths of ±20% (*top*), ±10% (*middle*) and ±5% (*bottom*) of the data value. In each case, the fit to the data is shown in red with derived kinematics in the bottom right of each panel.

This implies that higher cadence data is required to derive the true kinematics of a CBF pulse.

The variation in derived kinematics with cadence is shown in Figure 6 (again for ±5% uncertainty). As the cadence decreases, the derived velocity and acceleration approach the model values, with the scatter showing a dramatic reduction below ∼50 s cadence. 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 CBF pulse.

The effects of varying uncertainty in the data were also examined for constant image cadence with the results of this analysis shown in Figures 7 and 8. Figure 7 shows the derived kinematics for the simulated data–set with the top panel showing uncertainties of ±20%, the middle panel showing uncertainties of ±10% and the bottom panel showing uncertainties of ±5%. The variation in uncertainty has a noticeable effect on the derived kinematics in each case, with the acceleration in particular exhibiting significant variation with uncertainty.

The variation in the derived kinematics with uncertainty is shown in Figure 8, with the variation in offset distance (r_0), initial velocity (v_0) and acceleration (a_0) shown in the top, middle and bottom panels respectively. Some variation is apparent in each case, with the acceleration again showing the strongest variation. The derived kinematics are seen to approach the model kinematics as the noise reduces to 0% as expected.

It is clear that the variation in both noise and imaging cadence can strongly influence the derived kinematics of a CBF pulse, while the different numerical differencing techniques do not return accurate estimates of the pulse kinematics. However, it is possible to use a bootstrapping ap-
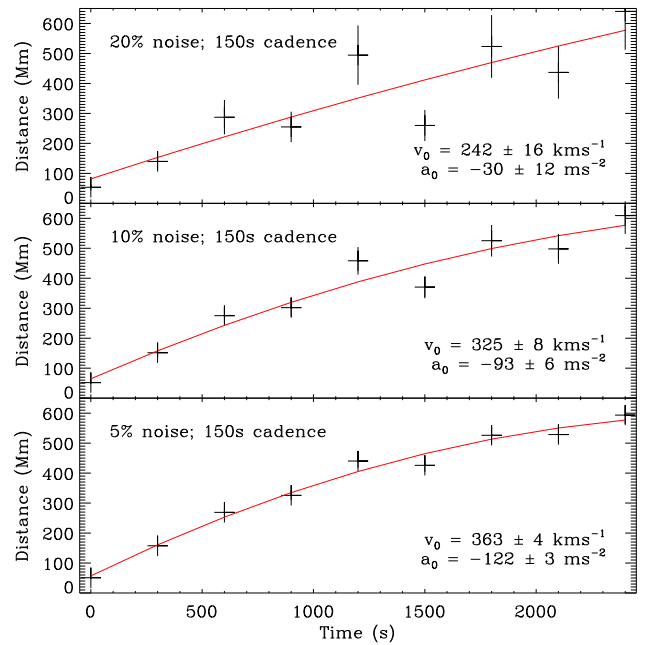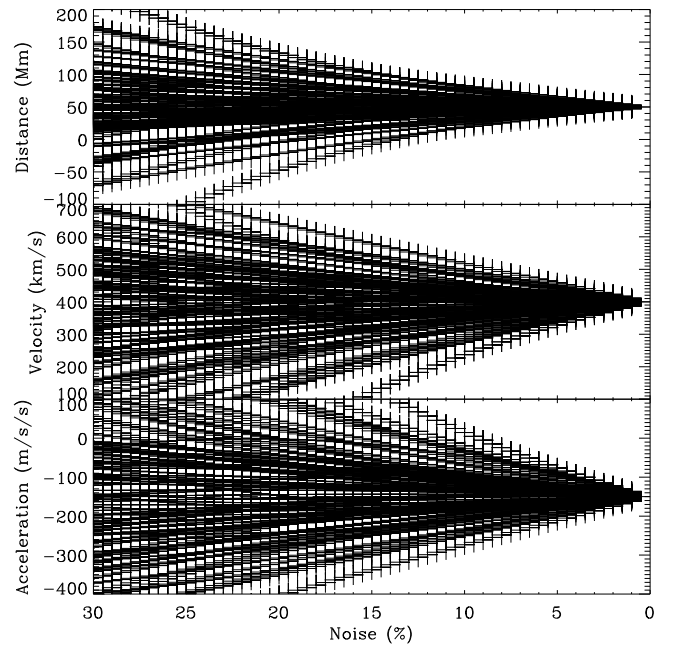

Fig. 8. Derived kinematics for varying uncertainty at 150 s image cadence. Distance, velocity and acceleration are shown in the top, middle and bottom panels respectively.

proach to overcome these issues as this is a statistically significant technique that has been optimised for small data-sets such as those typically obtained when studying CBF pulses.

4.2. Positional uncertainty?

5. Case Studies

5.1. CORIMP

5.2. CorPITA

6. Conclusions

References

Byrne, J. P., Morgan, H., Habbal, S. R., & Gallagher, P. T. 2012, ApJ, 752, 145

Efron, B. 1979, Ann. Stat., 7, 1

Gallagher, P. T., Lawrence, G. R., & Dennis, B. R. 2003, ApJ, 588, L53

Long, D. M., Gallagher, P. T., McAteer, R. T. J., & Bloomfield, D. S. 2008, ApJ, 680, L81

Ma, S., Wills-Davey, M. J., Lin, J., et al. 2009, ApJ, 707, 503

Muhr, N., Veronig, A. M., Kienreich, I. W., Temmer, M., & Vršnak, B. 2011, ApJ, 739, 89

Veronig, A. M., Temmer, M., & Vršnak, B. 2008, ApJ, 681, L113

Wen, Y., Maia, D. J. F., & Wang, J. 2007, ApJ, 657, 1117

Warmuth, A., Vršnak, B., Magdalenić, J., Hanslmeier, A., & Otruba, W. 2004, A&A, 418, 1101