Numerical differentiation analysis - Draft

David Long

Contents

List of Tables

List of Figures

This report deals with the statistical analysis of the kinematics of seven disturbance events observed using the *STEREO* spacecraft. The aim of this study was to test the different numerical differencing techniques available for determining the kinematics of these disturbances. Should these numerical differencing techniques fail, an accurate method must be found for determining their kinematics to a high confidence level. This should indicate if the events are freely-propagating MHD waves, or alternatively if they are propagating fronts of magnetic reconnection.

Rather than testing real data, it was decided to use simulated data to test each numerical differencing technique. This allowed the different parameters to be known and enabled the simultaneous testing of the different techniques.

Section 1 presents the different numerical differencing techniques that were examined in the course of this work. These are presented in detail, and the truncation errors determined mathematically in each case. Next, Section 2 discusses the kinematical model that was chosen for the simulation work. The noise added to the simulated data is then discussed in Section 3. The simulation results are presented and interpreted in Section 4, before some conclusions are drawn in Section 5.

1 Numerical differencing techniques

Several different numerical differencing techniques were studied for this report. The Taylor expansion method was used to derive the forward, reverse and centre differencing techniques, while the built-in IDL numerical differencing technique (DERIV) uses a three-point Lagrangian technique.

A Taylor series is the expansion of the real function $f(t)$ about the point $t = a$ and generally takes the form

$$f(t) = f(a) + f'(a)(t-a) + \frac{f''(a)}{2!}(t-a)^2 + \frac{f'''(a)}{3!}(t-a)^3 + ... \tag{1}$$

The difference between points $(t-a)$ is defined as Δt. This expansion can be used to determine the numerical derivative of a given point using a variety of different methods, each of which is outlined in the following sections.

1.1 Forward differencing

The forward differencing technique involves the computation of the derivative at the point $t + \Delta t$ by extrapolating forward from the point t. This uses the Taylor series:

$$r(t+\Delta t) = r(t) + r'(t)\Delta t + \frac{r''(t)}{2!}(\Delta t)^2 + \frac{r'''(t)}{3!}(\Delta t)^3 + ... \tag{2}$$

This equation can be re-arranged to give

$$r'(t)\Delta t = r(t+\Delta t) - r(t) - \frac{r''(t)}{2!}(\Delta t)^2 - \frac{r'''(t)}{3!}(\Delta t)^3 + ... \tag{3}$$

which then gives

$$r'(t) = \frac{r(t+\Delta t) - r(t)}{\Delta t} - \frac{r''(t)}{2!}(\Delta t) - \frac{r'''(t)}{3!}(\Delta t)^2 + ... \tag{4}$$

This is usually written as

$$r'(t) = v = \frac{r(t+\Delta t) - r(t)}{\Delta t} + O(\Delta t) \tag{5}$$

where $O(\Delta t)$ is the truncation error term, determined by the distance between neighbouring points (Δt). This technique assumes a straight line gradient between points.

This estimate of the velocity is dependent on the initial units used for the distance. In the case of the simulation work done here, the units of distance are mega-metres (1 Mm = 10^6 m). This produces an estimate of velocity in units of Mm s^{-1}. To convert this to acceptable units of km s^{-1} requires multiplying the estimated velocity values by 10^3. This has been done in all plots showing the numerically derived velocity. Similarly, converting the acceleration from units of Mm s^{-2} requires multiplying the estimated acceleration values by 10^6.

It is possible to derive the value of the truncation error term in terms of the original $r(t)$ values. The truncation error is given as

$$O(\Delta t) = \frac{r''(t)}{2!}(\Delta t) \tag{6}$$

The $r''(t)$ term may be decomposed using the original forward-difference definition:

$$r''(t) = \frac{r'(t+\Delta t) - r'(t)}{\Delta t} \tag{7}$$

Rewriting each term using the original functional forms produces

$$O(\Delta t) = \frac{r(t+2\Delta t) - 2r(t+\Delta t) + r(t)}{2!\Delta t} \tag{8}$$

The error term associated with the velocity estimate using the forward-difference technique is therefore dependent on the value of the function $r(t)$ at the points t, $t+\Delta t$ and $t+2\Delta t$.

This equation must be modified when dealing with the error associated with the acceleration estimate. In this case, the truncation error term would be given as:

$$O(\Delta t) = \frac{v(t+2\Delta t) - 2v(t+\Delta t) + v(t)}{2!\Delta t} \tag{9}$$

Here, the velocity function $v(t)$ is treated as the base function, rather than the distance function $r(t)$ as above.

The forward-difference technique is a very simplistic technique that produces spiky plots, with large variation between points. This is a result of the inherent assumption made by the forward-difference technique that there is a straight-line gradient between points. In addition, the forward-difference technique removes a point from the end of the data-set with each differentiation due to the way it calculates the derivative.

1.2 Reverse differencing

A similar method to the forward difference method is the reverse differencing technique. In this case, the derivative at the point $t - \Delta t$ is estimated by extrapolating backwards from the point t, rather than forwards (hence the name). This method again uses the Taylor series:

$$r(t-\Delta t) = r(t) - r'(t)\Delta t + \frac{r''(t)}{2!}(\Delta t)^2 - \frac{r'''(t)}{3!}(\Delta t)^3 + ... \tag{10}$$

This equation can be re-arranged to give

$$r'(t)\Delta t = r(t) - r(t-\Delta t) + \frac{r''(t)}{2!}(\Delta t)^2 - \frac{r'''(t)}{3!}(\Delta t)^3 + ... \tag{11}$$

which then gives

$$r'(t) = \frac{r(t) - r(t-\Delta t)}{\Delta t} + \frac{r''(t)}{2!}(\Delta t) - \frac{r'''(t)}{3!}(\Delta t)^2 + ... \tag{12}$$

This is usually written as

$$r'(t) = v = \frac{r(t) - r(t-\Delta t)}{\Delta t} + O(\Delta t) \tag{13}$$

where $O(\Delta t)$ is the truncation error term which, as with the forward difference method, is determined by the distance between neighbouring points (Δt), assuming a straight line gradient between points.

The units of the velocity estimate produced using this method once again depend on the units used for the distance. The velocity estimate produced using this method must again be multiplied by 10^3 to give an estimate of velocity in units of km s^{-1}, while the acceleration estimate must be multiplied by 10^6 to give an estimate in units of m s^{-2}. As with the forward-difference technique, this has been done for all plots showing the velocity derived numerically using the reverse-difference technique.

Once again, the truncation error can be estimated in terms of the original distance function $r(t)$. The truncation error is defined as:

$$O(\Delta t) = \frac{r''(t)}{2!}(\Delta t) \tag{14}$$

Using the definition of the reverse-difference technique, the $r''(t)$ term may be defined as

$$r''(t) = \frac{r'(t) - r'(t - \Delta t)}{\Delta t} \tag{15}$$

It is possible to re-write each term using the the original functional form and the definition of the reverse-difference technique. When this is done, and the values substituted into the equation for the truncation error term, this produces

$$O(\Delta t) = \frac{r(t) - 2r(t - \Delta t) + r(t - 2\Delta t)}{2!\Delta t} \tag{16}$$

This shows a similar form to the truncation error estimate for the forward-difference technique. Both techniques are quite similar and it is not unexpected that they should show a similar form for the truncation error estimate. However, despite the similar form, the two estimates are not necessarily equal.

As before, this equation can be modified to find the truncation error associated with the acceleration term by taking the velocity function $v(t)$ as the base function rather than the distance function $r(t)$. When this is done, this produces a truncation error of:

$$O(\Delta t) = \frac{v(t) - 2v(t - \Delta t) + v(t - 2\Delta t)}{2!\Delta t} \tag{17}$$

Both the forward and reverse difference techniques are very similar and use two adjacent points (t & $t + \Delta t$ and t & $t - \Delta t$ respectively) to find the derivative at a chosen point. They are both very simplistic techniques and are only applicable in very simplistic cases. The reverse-difference technique is also observed to remove a point from the beginning of the data-set with each numerical derivative due to the way the technique is calculated.

1.3 Centre differencing

A more accurate technique using the Taylor expansion method is the centre difference technique. This uses the points either side of the point under examination t (i.e..: $t - \Delta t$ and $t+\Delta t$), smoothing the data and producing a more accurate estimate of the numerical derivative at that point. From above, the Taylor series expansion is defined as

$$r(t - \Delta t) = r(t) - r'(t)\Delta t + \frac{r''(t)}{2!}(\Delta t)^2 - \frac{r'''(t)}{3!}(\Delta t)^3 + ... \tag{18}$$

at the point $t - \Delta t$ and

$$r(t + \Delta t) = r(t) + r'(t)\Delta t + \frac{r''(t)}{2!}(\Delta t)^2 + \frac{r'''(t)}{3!}(\Delta t)^3 + ... \tag{19}$$

at the point $t + \Delta t$. The centre differencing technique subtracts the following point from the leading point:

$$r(t + \Delta t) - r(t - \Delta t) \tag{20}$$

This produces the equation

$$r(t + \Delta t) - r(t - \Delta t) = 2r'(t)\Delta t + 2\frac{r'''(t)}{3!}(\Delta t)^3 + ... \tag{21}$$

This equation can then be re-arranged to give

$$r'(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} - \frac{r'''(t)}{3!}(\Delta t)^2 + ... \tag{22}$$

Following the notation from before, this is usually written

$$r'(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} + O(\Delta t^2) \tag{23}$$

where $O(\Delta t^2)$ is the truncation error term. The truncation error term in this case is determined by the square of the distance between neighbouring points, producing a more accurate estimate than the forward and reverse difference methods.

As with the forward and reverse-difference techniques, the velocity estimate produced by the centre-difference method must be multiplied by 10^3 to give the physically realistic units of km s^{-1}, with the acceleration estimate multiplied by 10^6 to give the physically realistic units of m s^{-2}. This has been done for all the simulated data-sets that have been processed using the centre-difference technique.

The centre-difference method effectively smoothes the data set while differentiating it by using the two points either side of the point under examination. The truncation error can once again be given in terms of the original function. From above the truncation error is defined as:

$$O(\Delta t^2) = \frac{r'''(t)}{3!}(\Delta t)^2 \tag{24}$$

In this case, the truncation error is defined by the third order derivative of the original function (unlike the second-order derivative for the forward and reverse-difference methods). As before, the third-order derivative can be defined as

$$r'''(t) = \frac{r''(t + \Delta t) - r''(t - \Delta t)}{2\Delta t} \tag{25}$$

using the definition of the centre-difference technique. Once all of the terms have been re-written in terms of the original equation, the truncation error term, may be written as

$$O(\Delta t^2) = \frac{r(t+3\Delta t) - 3r(t+\Delta t) + 3r(t-\Delta t) - r(t-3\Delta t)}{(3!)(8)\Delta t} \tag{26}$$

This produces a mean truncation error estimate that is much smaller than the mean truncation error calculated using the forward and reverse-difference techniques.

As with the forward and reverse-difference techniques, the truncation error term can be estimated for the acceleration term by using the velocity function $v(t)$ as the base function rather than the distance function $r(t)$. When this is done, it produces the estimate

$$O(\Delta t^2) = \frac{v(t+3\Delta t) - 3v(t+\Delta t) + 3v(t-\Delta t) - v(t-3\Delta t)}{(3!)(8)\Delta t} \tag{27}$$

for the truncation error of the acceleration data.

The centre-difference technique produces a mean truncation error estimate and a scatter in the data that are much smaller than the forward and reverse-difference techniques. As a result, the centre-difference technique produces a better estimate of the velocity and acceleration, and their associated truncation errors than either the forward or reverse-difference techniques.

1.4 Lagrangian differentiation

A more advanced numerical differentiation method is the three-point Lagrangian method used by the DERIV routine in IDL. This method uses three adjacent points ($t-\Delta t$, t and $t+\Delta t$) to fit a Lagrange polynomial function of the form

$$\begin{aligned} P(x) = {} & \frac{(x-t)(x-(t+\Delta t))}{((t-\Delta t)-t)((t-\Delta t)-(t+\Delta t))} y_1 \\ & + \frac{(x-(t-\Delta t))(x-(t+\Delta t))}{(t-(t-\Delta t))(t-(t+\Delta t))} y_2 \\ & + \frac{(x-(t-\Delta t))(x-t)}{((t+\Delta t)-(t-\Delta t))((t+\Delta t)-t)} y_3 \end{aligned} \tag{28}$$

to the three points, and hence find the derived value at a point numerically.

This method gives a more accurate indication of the numerically differentiated data as it does not assume a straight line in between points. In addition, the routine compensates for edge points, increasing the errors of the points. It also uses three points to find the numerical derivative, rather than two (c.f. centre, forward and reverse-difference techniques). The Lagrangian technique also smoothes the data-set, removing the spiky appearance that arises from use of the different Taylor series techniques.

Table 1. Numerical differentiation errors

Method	Kinematics	Error term
Forward diff.	Velocity	$O(\Delta t) = \frac{r(t+2\Delta t)-2r(t+\Delta t)+r(t)}{2!\Delta t}$
	Acceleration	$O(\Delta t) = \frac{v(t+2\Delta t)-2v(t+\Delta t)+v(t)}{2!\Delta t}$
Reverse diff.	Velocity	$O(\Delta t) = \frac{r(t)-2r(t-\Delta t)+r(t-2\Delta t)}{2!\Delta t}$
	Acceleration	$O(\Delta t) = \frac{v(t)-2v(t-\Delta t)+v(t-2\Delta t)}{2!\Delta t}$
Centre diff.	Velocity	$O(\Delta t^2) = \frac{r(t+3\Delta t)-3r(t+\Delta t)+3r(t-\Delta t)-r(t-3\Delta t)}{(3!)(8)\Delta t}$
	Acceleration	$O(\Delta t^2) = \frac{v(t+3\Delta t)-3v(t+\Delta t)+3v(t-\Delta t)-v(t-3\Delta t)}{(3!)(8)\Delta t}$
Lagrangian	Velocity	σ
	Acceleration	σ

The problems encountered in this work arise from the small data-sets available in the case of these disturbances. The small data-sets mean that there are only four or five points available of which three are needed to calculate the numerical derivative of each point. This makes the Lagrangian technique unsuitable for estimating the kinematics of these disturbances.

The three-point Lagrangian interpolation method used by IDL uses the standard deviation of the derivative calculated by the Lagrangian polynomial as the truncation error of the numerical derivative. This produces a consistent estimate of the mean truncation error that only varies with the end-points due to the increased uncertainty of these points.

Table 1 shows the different estimates for the truncation error term for each numerical differencing technique. These truncation error estimates are given in terms of the original base function, the exception being for the Lagrangian technique, which takes the standard deviation of the Lagrange derivative at a point as the truncation error estimate at that point.

2 Kinematic model

To test the different numerical difference techniques, a simple linear equation model was chosen as a basis for the simulation work. This model took the form

$$r(t) = r_0 + v_0 t \tag{29}$$

This is a simple zero acceleration/constant velocity model with two free parameters (the initial offset distance and the initial velocity). This model assumes that once the disturbance is freely-propagating, it moves with a constant velocity and zero acceleration. The different parameters were chosen as:

$$r_0 = 150\,\text{Mm} \tag{30}$$

$$v_0 = 400\,\text{km/s} \tag{31}$$

resulting in an equation of the form

$$r(t) = 150 + 0.4t \tag{32}$$

This model was used as a basis for both an ideal data-set and a simulated data-set. The ideal data-set followed the model exactly, and was used to show the expected kinematics. The simulated data-set used the model as a basis, but added Gaussian noise of varying strength (see Section 3 for a detailed discussion of the noise added to the data). The simulated data-set was therefore of the form

$$r(t) = r_0 + v_0 t + \delta r \tag{33}$$

with the δr term in this case corresponding to the added noise.

A fit was applied to the distance-time plot in the case of the noisy data to examine the effect of varying the percentage noise on the initial offset distance and the initial velocity. This could then be compared to the known kinematics, allowing the limiting noise level to be found. The derived velocity and acceleration data were also plotted to show the effect of the different numerical differencing techniques. This is discussed in more detail in Section 4.

A total time range of 2000s was chosen for this analysis as this is comparable to real data. within this total time range, the cadence was varied from 10s to 240s. This range of cadence values encompasses cadences allowed by both the *SDO* and *STEREO* missions, and produces data-sets ranging from 200 to 8 points. Any fewer points than this and the different numerical differencing methods would start to fail due to the small number of data-points.

3 Noise

To ensure that the different techniques under review operated as expected and were accurate, it proved necessary to add a quantified amount of noise to the data-sets. This was necessary as if the noise added was not quantifiable, then the results of the simulation would be meaningless. In this case, it was decided to add Gaussian noise obtained using the RANDOMN routine from IDL. This noise was added to the ideal data-set, producing a simulated data-set that was kept separate from the ideal data-set.

The RANDOMN routine from IDL was chosen as it produces an array of randomly chosen values of a pre-determined size that have a mean of $\mu = 0$ and a standard deviation of $\sigma = 1$. In this case, the noise would be varying about the original equation, so the mean of $\mu = 0$ was preferable. The standard deviation of this random array could then be varied by multiplying the array by any number x, producing an array of mean $\mu = 0$, with the standard deviation given by that number (i.e.: $\sigma = x$).

The noise was quantified at source by varying the standard deviation of the Gaussian noise added to the ideal data-set. This noise was added (as shown in Section 2) by adding the noise term to the original distance function:

$$r(t) = r_0 + v_0 t + \delta r \tag{34}$$

Due to the random nature of the noise, the resulting percentage noise was measured by taking the minimum, mean and maximum values of the percentage difference between the simulated data and the ideal data at each data-point. The standard deviation was varied between 0 (no noise) and 300, producing a variation in mean percentage noise of 0 to ∼60 %.

The effect of varying the standard deviation of the Gaussian distribution used to produce the noise is shown in Figure 1. This shows the ideal and simulated distance-time data for a standard deviation of 0 Mm (top plot), 150 Mm (middle plot) and 300 Mm (bottom plot). This may be seen to result in a mean percentage scatter of 0 %, ∼30 % and ∼60 % respectively.

For the purposes of the first simulation, the standard deviation of the Gaussian distribution used to obtain the noise was varied from 0 to 300 Mm in increments of 10 Mm. While the level of noise was being examined, the cadence of the data was kept constant at 100s. This cadence was chosen as it allowed sufficient data-points to examine the effect of the numerical differentiation methods while simultaneously being a small enough data-set for the results to be comparable with real data. This resulted in a total data-set of 20 data-points for the simulation.

In each case, the noisy distance-time data was fitted with a best-fit equation to try and identify the kinematics. This fit was observed to vary with increasing noise, with both of the coefficients varying according to the data. The numerically-derived velocity and accelera-

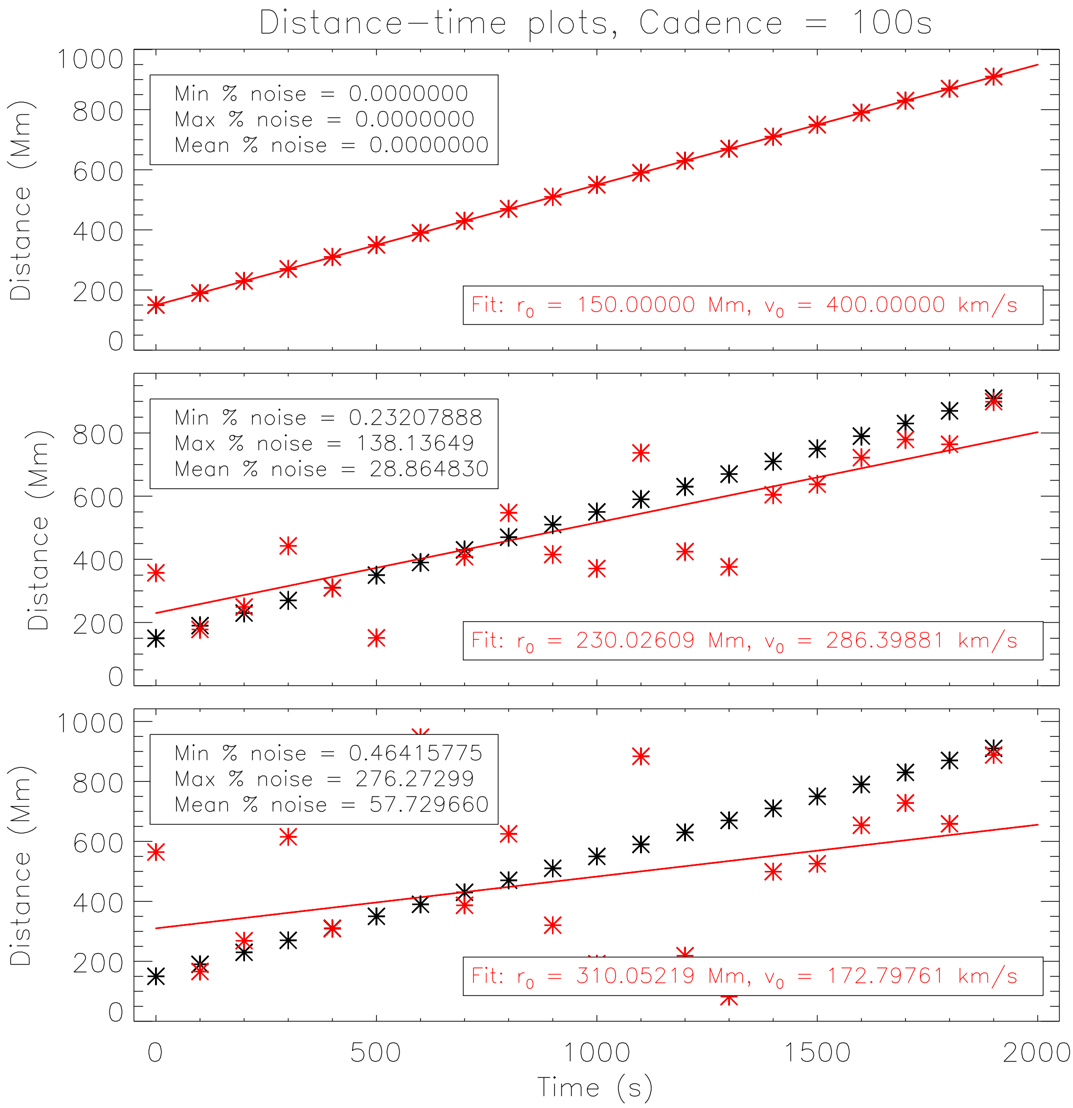

Figure 1 – Each plot shows the same distance-time model, with the noise level for the simulated data different in each case. *Top*: Standard deviation = 0 Mm. The ideal and simulated data match perfectly in this case. *Middle*: Standard deviation = 150 Mm. The simulated data shows some scatter. *Bottom*: Standard deviation = 300 Mm. The simulated data shows large scatter in this case.

tions were also fitted in each case, to try and return the best-fit kinematics in each case, for comparison with the original kinematics.

The effects of varying the cadence of the data were also examined, keeping the noise-level of the data constant. This was done by using a constant standard deviation of 100 Mm. Both this simulation and the simulation examining the effects of the varying noise level are discussed in more detail in Section 4.

4 Simulations and Analysis

Several simulations were carried out in the course of this work to try and determine the effects of varying both the noise-level and cadence of the noisy data on the abilities of the different numerical techniques to find the kinematics of the original data.

4.1 Varying noise-level

The first simulation carried out involved the examination of the effects of varying noise-level on the accuracy of the different numerical differencing techniques. To perform this simulation, each numerical technique was applied to the same noisy data-array and noise-level in turn, with the cadence kept constant at 100 s. The standard deviation of the noise array was varied between 0 and 300 Mm in steps of 10 Mm (as discussed in Section 3). The distance-time, derived velocity-time and derived acceleration-time data was then plotted for each standard deviation for each numerical differencing technique. The resulting HTML movies may be found online[1].

Each of the numerical differencing techniques returned the ideal velocity and acceleration when the standard deviation of the noise applied to the simulated data array was set to zero. This indicates that in the absence of noise, the numerical differencing techniques work as expected. Once noise was added to the data, the derived velocity and acceleration showed varying degrees of scatter and varying truncation error estimates, depending on the technique being analysed.

The plots shown in the online movies have had the scales set to vary with the data being plotted. The data being plotted therefore appears stationary, but the scaling changes rapidly. This should be kept in mind when comparing the data from different movies.

A plot showing the distance, derived velocity and derived acceleration with time as obtained using the forward-difference technique for a standard deviation of 100 Mm is shown in Figure 2. This plot indicates that for a standard deviation of 100 Mm, the percentage noise on the distance data varies between 3 and 98 %, with a mean value of 22 %. Note that when comparing the different techniques and their effectiveness, the mean value of the percentage noise will be taken as this provides a more consistent estimate of the percentage noise.

This plot shows the wide scatter of the derived velocity and acceleration data due to the noisy distance data. In addition, the estimated truncation error in each case is quite large.

Figure 3 shows the distance, derived velocity and derived acceleration data obtained using the reverse-difference technique. This figure is quite similar to the forward-difference plot, and shows comparable scatter and truncation error estimates.

The forward and reverse-difference techniques can be observed to remove points from

[1] http://www.maths.tcd.ie/~dlong/num_diff/simulations

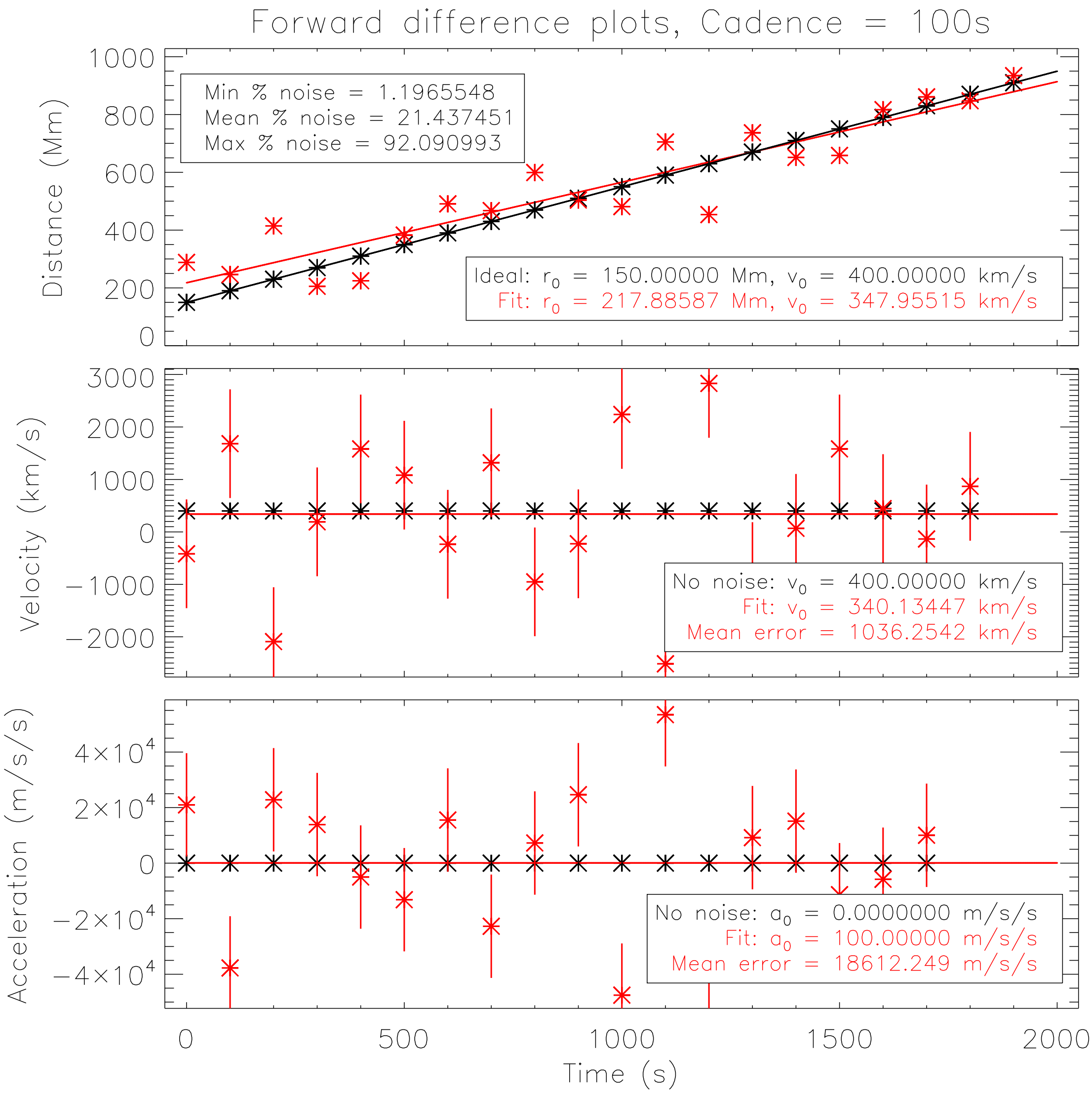

Figure 2 – The kinematics plots for the forward-difference technique, with Std. Dev = 100 Mm, giving mean percentage noise $\sim$ 21%. *Top*: Ideal (black) and noisy (red) distance-time data. *Middle*: Velocity-time data derived using the forward-difference method. *Bottom*: Acceleration time data derived using the forward-difference method.

the start and end of the data-set respectively. This does not greatly affect the results using the current simulation, but for a small data-set (of <10 data points), this can have a huge affect on the ability of the methods to return an accurate estimate of the kinematics.

Section 1.3 suggested that the centre-difference technique should provide a more accurate estimate of the derived velocity and acceleration than the forward and reverse-difference techniques. This is evident in Figure 4, which shows a much smaller scatter and reduced truncation error estimates for the centre-difference technique.

Despite the ability of the centre-difference method to produce smaller truncation error estimates and smaller scatter in the derived data-points, the technique does remove a point from both the start and end of the data-set with every differentiation. As with the forward

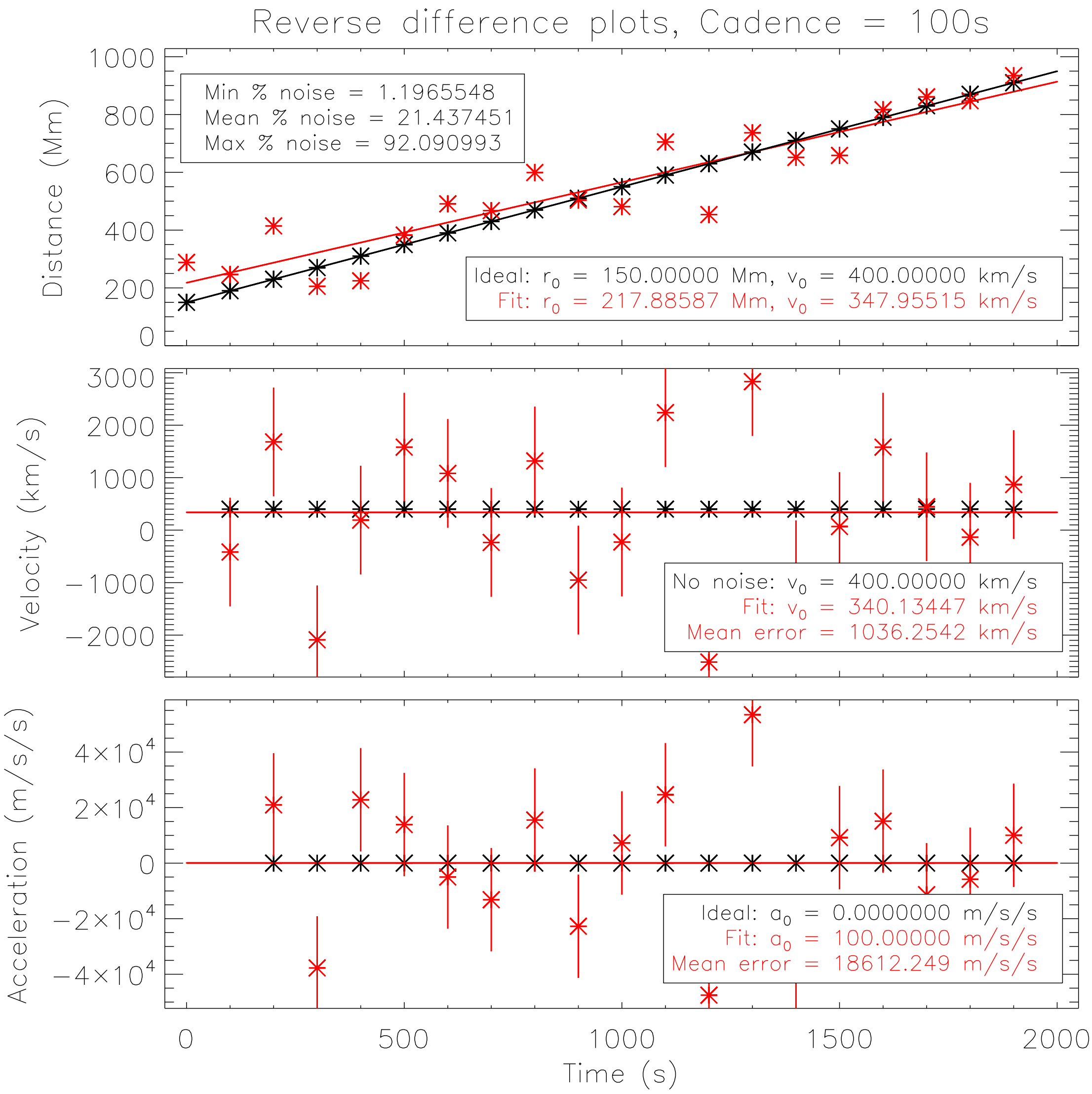

Figure 3 – The kinematics plots for the reverse-difference technique, with Std. Dev = 100 Mm, giving mean percentage noise $\sim$ 21%. *Top*: Ideal (black) and noisy (red) distance-time data. *Middle*: Velocity-time data derived using the reverse-difference method. *Bottom*: Acceleration time data derived using the reverse-difference method.

and reverse-difference methods, this does not have much effect here, but would have a huge effect on small data-sets of the type produced by globally-propagating coronal disturbances.

These results suggest that the centre-difference technique is the best technique to use of those derived using the Taylor expansion series method. A more advanced numerical differencing technique is that which uses Lagrange polynomials to estimate the numerical derivative of a function at a given point. Figure 5 shows this technique used on the same noisy data-set as the three different Taylor methods.

The velocity and acceleration data derived using the Lagrangian method show a similar scatter to that of the centre-difference technique. However, the end points have not been removed in this case. As a result, the Lagrangian method would be expected to return a

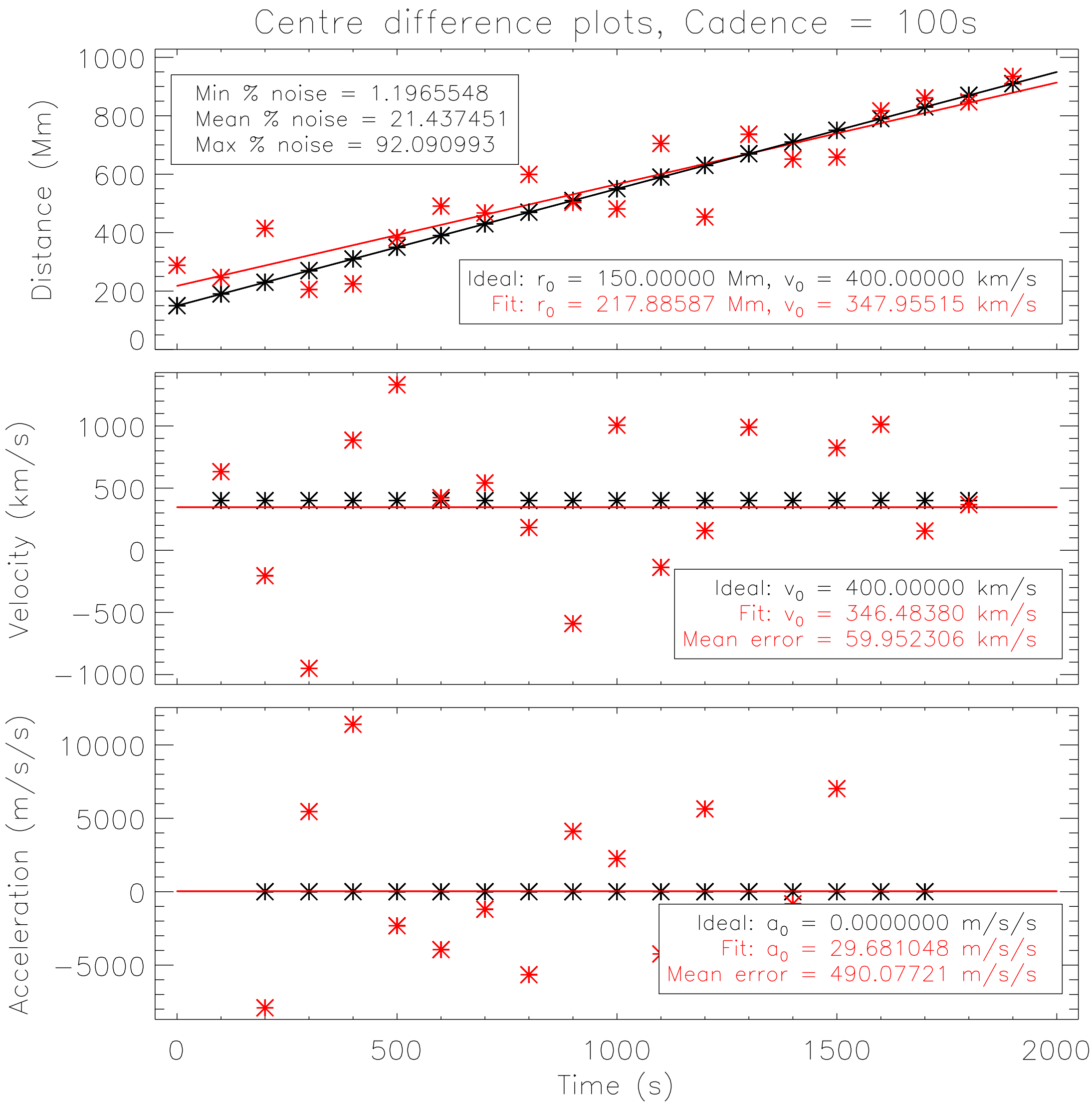

Figure 4 – The kinematics plots for the centre-difference technique, with Std. Dev = 100 Mm, giving mean percentage noise $\sim$ 21%. *Top*: Ideal (black) and noisy (red) distance-time data. *Middle*: Velocity-time data derived using the centre-difference method. *Bottom*: Acceleration time data derived using the centre-difference method.

better estimate of the numerical derivative of a small data-set. However, the end-points in the case of the Lagrangian method are skewed away from the rest of the data-set. This is reflected in the larger error bars associated with these points.

As discussed in Section 1.4, the truncation errors of the Lagrangian technique at a point are estimated using the standard deviation of the Lagrangian derivative at that point. As a result, the truncation error is more consistent across the data-points, unlike the truncation error estimate for the Taylor methods, which can vary substantially from point-to-point. As a result of this variation, the global truncation error (found by taking the mean of the absolute values of the local truncation error estimates) is plotted for each of the different Taylor method plots. In the case of the Lagrangian method, the local truncation error is

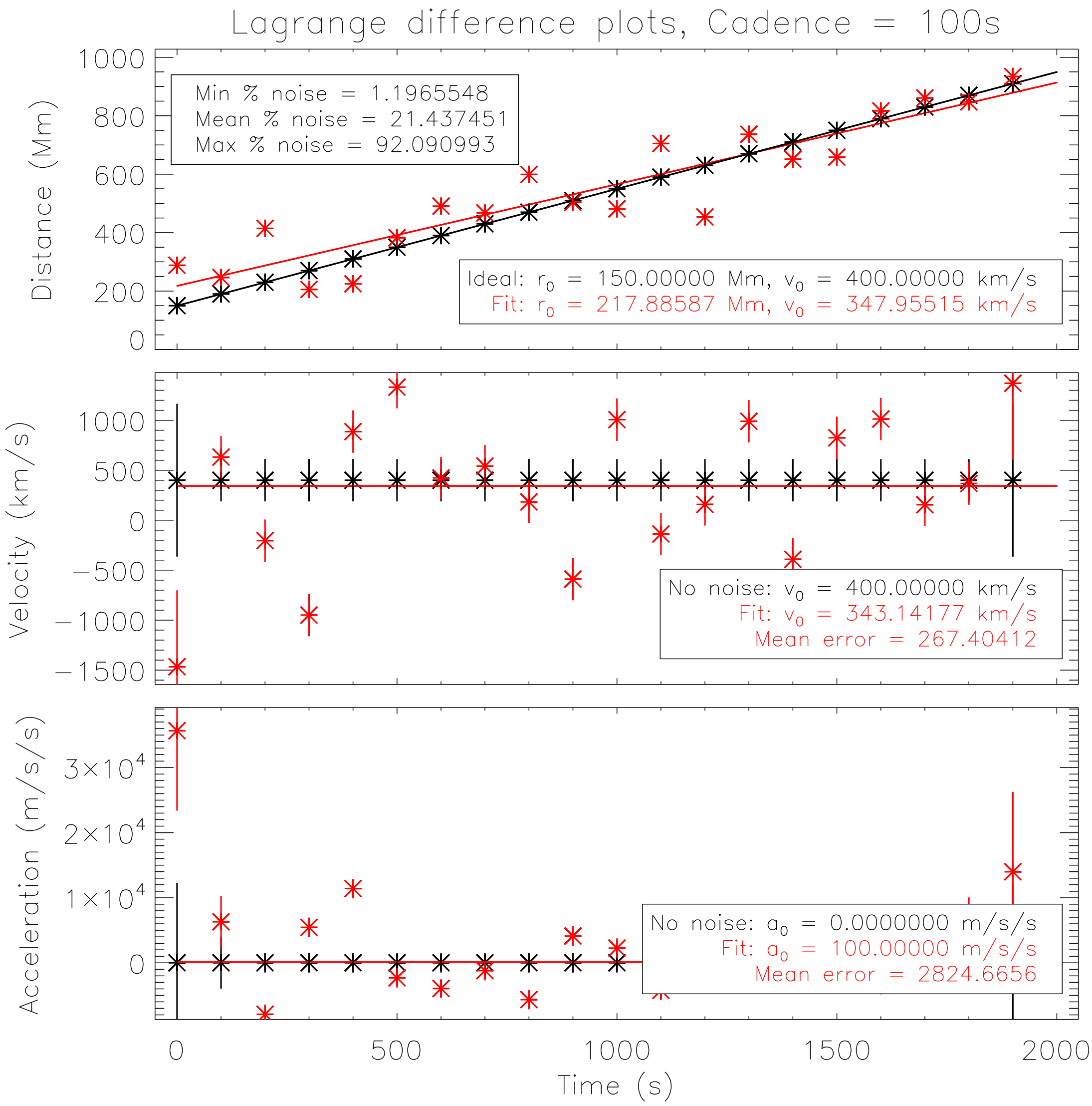

Figure 5 – The kinematics plots for the lagrangian-difference technique, with Std. Dev = 100 Mm, giving mean percentage noise $\sim$ 21%. *Top*: Ideal (black) and noisy (red) distance-time data. *Middle*: Velocity-time data derived using the Lagrangian-difference method. *Bottom*: Acceleration time data derived using the Lagrangian-difference method.

plotted, with the global truncation error quoted on the plots.

The large data-set in this case was chosen to show the general effects of each numerical differencing techniques when confronted with a data-set that has a varying noise level. The next step is to examine the effects of varying the cadence of the data-set on the numerical-differencing techniques. This is discussed in Section 4.2. Once this has been done, it will be possible to examine the effects of these techniques on a realistic data-set of a given noise level and cadence. This will be discussed in Section 4.3.

The effect of varying the mean percentage noise on the truncation error estimate of the velocity and acceleration for each technique is shown in the left-hand plot in Figure 10. The truncation error for both velocity and acceleration increases linearly for both the for-

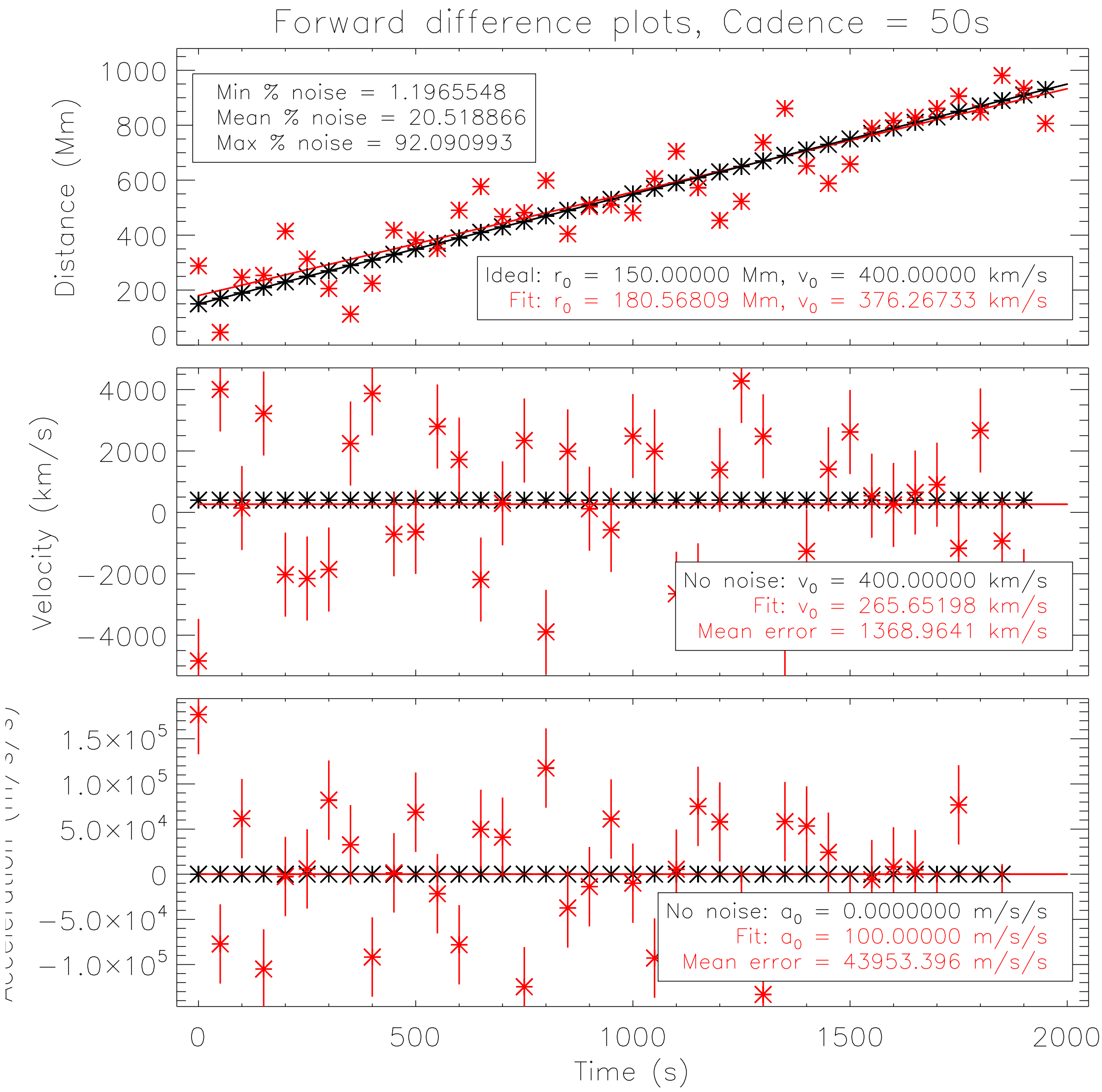

Figure 6 – The kinematics plots for the forward-difference technique, with cadence = 100 s, giving mean percentage noise $\sim$ 21%. *Top*: Ideal (black) and noisy (red) distance-time data. *Middle*: Velocity-time data derived using the forward-difference method. *Bottom*: Acceleration time data derived using the forward-difference method.

ward and reverse-difference techniques. The Lagrange-difference method shows a constant truncation error, while the centre-difference error is quite small, but increasing with mean percentage noise.

4.2 Varying cadence

The simulation outlined in Section 4.1 was repeated with the noise-level constant at a standard deviation of 100 Mm, but with the cadence of the data-set varying between 10 and 240 s in steps of 10 s. This enabled the ability of each numerical differencing technique to cope with a varying image cadence to be examined.

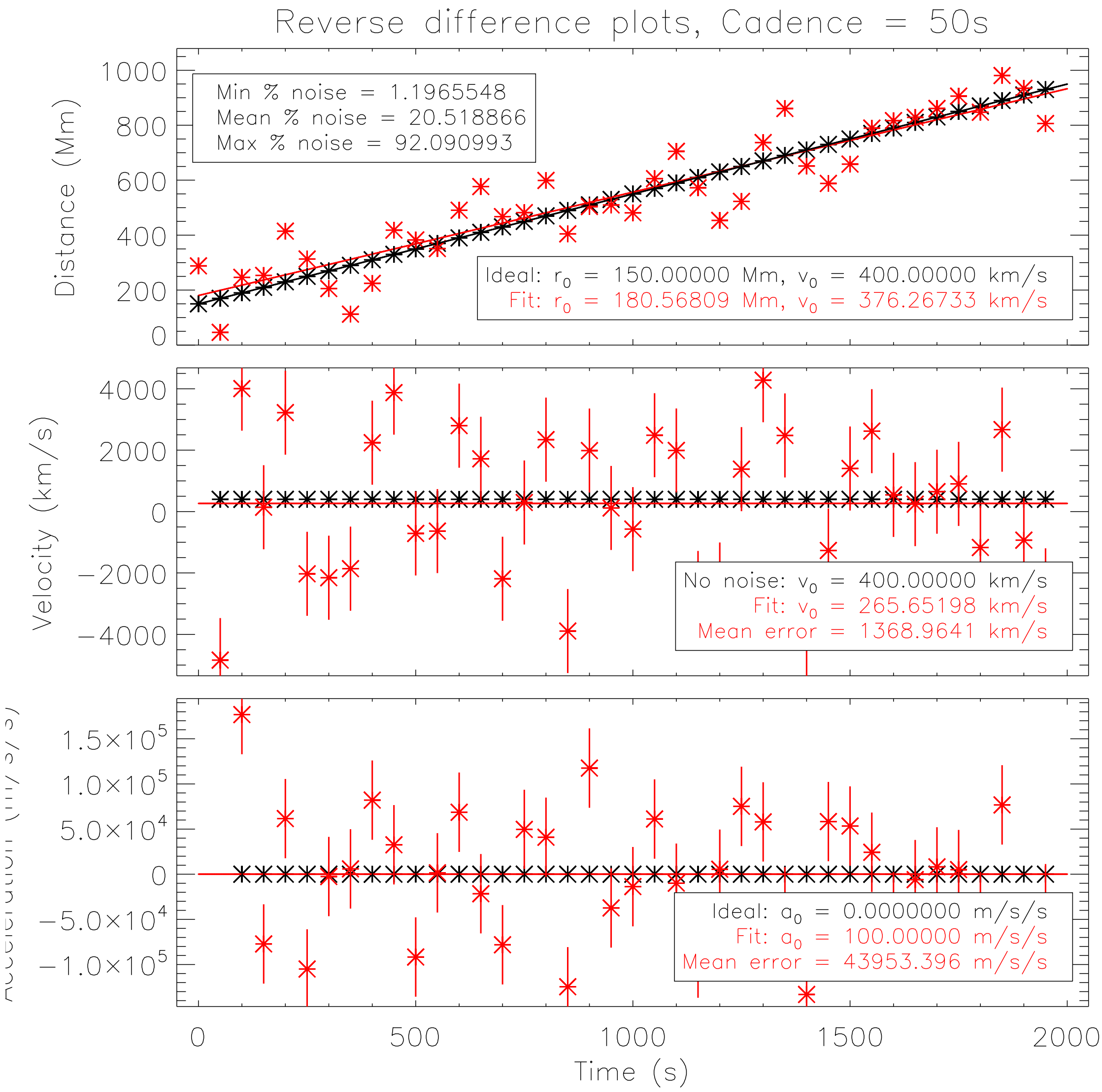

Figure 7 – The kinematics plots for the reverse-difference technique, with cadence = 100 s, giving mean percentage noise ∼ 21%. *Top*: Ideal (black) and noisy (red) distance-time data. *Middle*: Velocity-time data derived using the reverse-difference method. *Bottom*: Acceleration time data derived using the reverse-difference method.

As before, the distance-time, derived velocity-time and derived acceleration-time were then plotted for each cadence, and combined into a HTML movie (with one movie for each numerical differencing technique). The movies allow the variation with cadence to be viewed easily, and may be viewed online[2].

Figure 6 shows the distance-time, derived velocity-time and derived acceleration-time plots for a data-set with a standard deviation of 100 Mm and a cadence of 50 s using the forward-difference technique. This produces a percentage noise array with a minimum value of ∼ 1%, a mean value of ∼ 21% and a maximum value of ∼ 92%. As with the simulation examining varying noise-level, there is a large scatter in the derived velocity and

[2] http://www.maths.tcd.ie/~dlong/num_diff/simulations

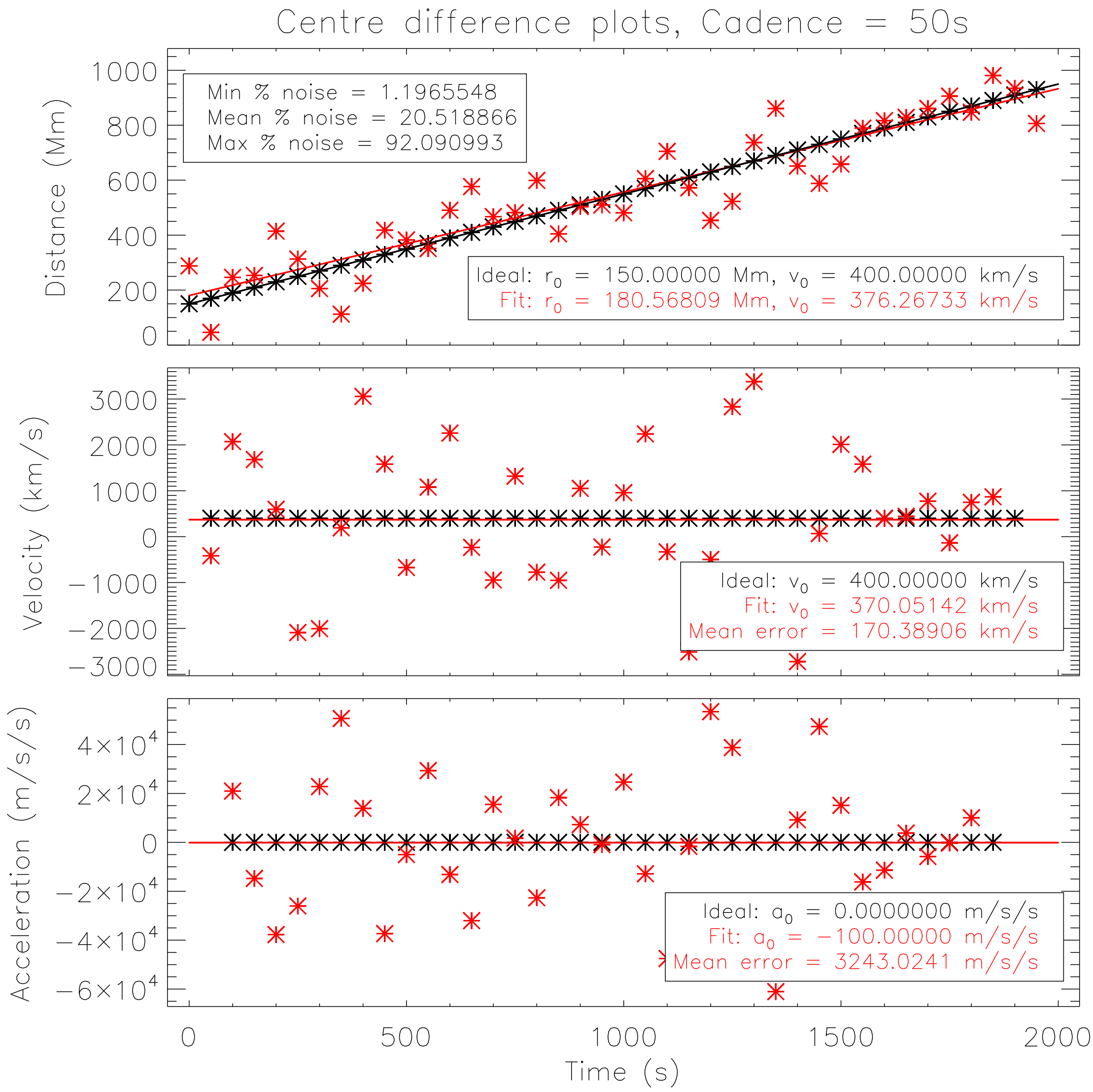

Figure 8 – The kinematics plots for the centre-difference technique, with cadence = 100 s, giving mean percentage noise $\sim$ 21%. *Top*: Ideal (black) and noisy (red) distance-time data. *Middle*: Velocity-time data derived using the centre-difference method. *Bottom*: Acceleration time data derived using the centre-difference method.

acceleration data. The estimated truncation error is also quite large.

The corresponding plot for the reverse-difference technique is shown in Figure 7. The derived velocity and acceleration data again shows a large scatter, with the points dropping from the front of the data-set as previously shown. The truncation error estimate is once again of the same order as the forward-difference technique.

Figure 8 shows the derived kinematics plots for the centre-difference method. These plots show a much reduced scatter in the data, in addition to a smaller estimated truncation error. Once again, the plots show that the centre-difference technique removes points from either end of the data-set with each derivative. This is the reason for the upper limit of the cadence simulation of 240 s; the centre-difference technique requires seven unique points to

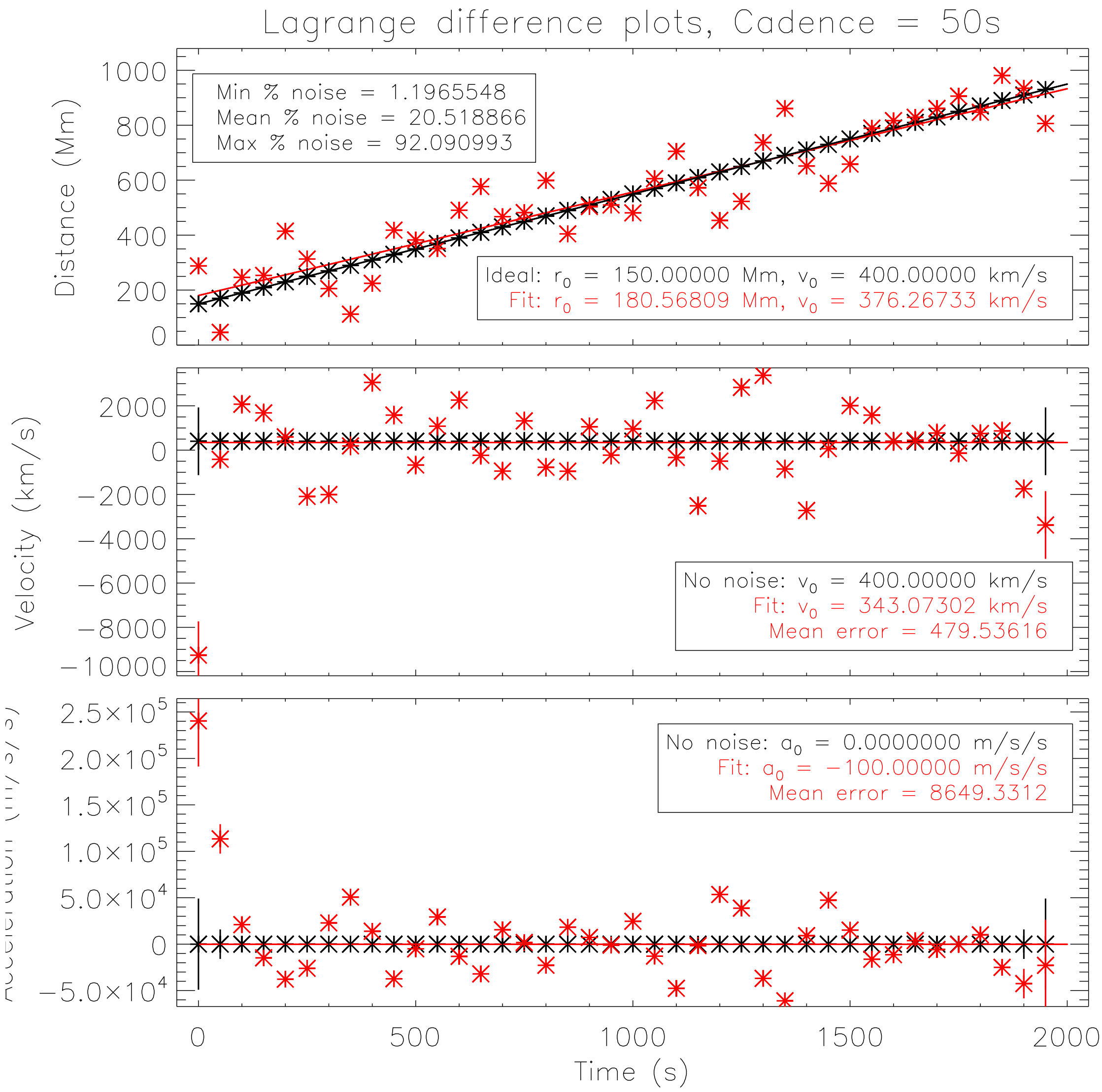

Figure 9 – The kinematics plots for the Lagrange-difference technique, with cadence = 100 s, giving mean percentage noise $\sim$ 21%. *Top*: Ideal (black) and noisy (red) distance-time data. *Middle*: Velocity-time data derived using the Lagrange-difference method. *Bottom*: Acceleration time data derived using the Lagrange-difference method.

determine the truncation error associated with each individual point in the acceleration data. This places a lower limit on the number of points required for the centre-difference technique to operate with even some semblance of accuracy. If there are fewer than seven points for the acceleration data, the truncation error cannot be estimated properly, and the technique is effectively useless. This implies that a minimum of seven data points are required in the original data-set are required before the centre-difference technique can return an estimate of the truncation error associated with the second-derivative of the given function.

The Lagrangian technique is shown in Figure 9 using the same parameters as in the previous plots. In this case, the unique abilities of the Lagrangian technique are clear; it is not necessary to remove points from either end of the data-set in order to calculate the numerical derivative. This means that the Lagrangian technique may be used with data-sets that have as few as three data-points, making it ideal for the small data-sets that will be encountered when dealing with globally-propagating coronal disturbances. However, the errors associated with this technique are still quite large, reducing the ability of this numerical differencing technique to distinguish between different possible kinematic theories.

The variation in the mean truncation error with cadence is shown in the right-hand plot in Figure 10. All four techniques show a drop in mean truncation error with increasing cadence, as would be expected due to the larger time step value appearing on the bottom of the truncation error term in each case (see Table 1 for the different error terms). As a result, the larger time-step has the effect of reducing the associated truncation error estimate. This is somewhat counter-intuitive, but is clear once the mathematics are examined.

The forward and reverse-difference techniques show the largest truncation error terms, followed (somewhat surprisingly) by the Lagrangian technique. The centre-difference technique has the lowest truncation error estimate. However, as noted above, the centre-difference technique requires seven data-points to return a truncation error estimate, while the Lagrange technique requires only three.

With the results of the simulations on both varying noise-level and varying cadence in mind, we shall now examine the ability of the different techniques to return accurate kinematics estimates for a realistic data-set.

4.3 Realistic Simulation

The realistic simulation requires the use of a small data-set with a cadence of 150 s (for examining 171 Å-like data) and 600 s (for examining 195 Å-like data). In addition, the data should be constantly increasing, This requires the use of a noise-level that does not vary too excessively. This is necessary as the observed disturbances do not contract, instead showing only expansion. With this in mind, a standard deviation of 20 Mm was chosen, producing a percentage noise level that ranged from $\sim 0.3\%$ to $\sim 18\%$ with a mean of $\sim 4\%$. This noise-level produces a range of distance-time values that are constantly increasing as required.

It was decided to examine the 171 Å-like data first by setting the cadence of the data-set to 150 s. Assuming a reasonable estimate of eleven data-points, this meant a total time array of 1650 s. Figures 11 and 12 show the results of applying the different numerical differencing techniques to the noisy data-set at this cadence. In Figure 11 the forward and reverse-difference techniques produce similar results as expected. In both cases, the fit to the derived velocity and acceleration data produces a good estimate of the actual velocity and acceleration within error. However, the size of the error in each case makes this result inevitable. The derived data itself shows no similarity to the actual value due to the nature of the numerical differentiation methods.

Figure 12 shows the results of the centre-difference and Lagrange-difference methods applied to the noisy data-set. The centre-difference technique shows a large scatter, but a small truncation error estimate. In comparison, the Lagrange-difference technique shows a similar scatter, but a much larger truncation error estimate. In both cases, the derived data-sets show no similarity to the actual values, with the fits in all cases different from the actual values (although approximately similar within error).

To examine a typical 195 Å-like data-set, the cadence was increased to 600 s. At this cadence, real data-sets consist of a maximum of four data-points, resulting in a total time-array of 2400 s. The resulting plots may be seen in Figures 13 and 14.

The problems associated with the Taylor expansion techniques are immediately apparent from the given plots. The forward and reverse-difference techniques remove one data-point with each derivative. Although this is not a problem when the data-sets are quite large, with a data-set of four points, this leaves three data-points to estimate the derived velocity, and two data-points to estimate the derived acceleration in both cases. This cannot be expected to return a reasonable estimate for either the velocity or acceleration, and although the best-fit estimate is reasonable within error, this is merely due to the size of the truncation error estimate.

Figure 14 shows the centre and Lagrangian-difference techniques. The centre-difference technique removes two data-points with each derivative, leaving two data-points to estimate the velocity and no data-points for the acceleration. The centre-difference technique is clearly useless when confronted with such a small data-set. By comparison, the Lagrangian-difference technique retains all the data-points, but shows consistent variation in both the derived velocity and acceleration plots which can easily be misinterpreted. This is shown by the apparent dip in velocity, and constantly increasing acceleration, both of which are artifacts of the numerical differentiation method, and have no basis in the real data.

5 Discussion and Conclusions

The results of the analysis with regards to realistic data-sets suggest that all the numerical differentiation techniques are ineffective at returning a reasonable estimate of the kinematics of globally-propagating disturbances. In particular, the different techniques fail when confronted with small data-sets that are typical of this work. This is seen in the removal of data-points from already small data-sets, and also the tendency of techniques to produce plots that show trends that are lacking in the actual kinematics.

It must be concluded that the different numerical differentiation techniques examined here are useless for this work. This research requires the accurate determination of kinematics of globally-propagating disturbances from data-sets of less than ten data-points. While the different techniques may have some success at returning accurate kinematics from large data-sets, the small data-sets that are frequently encountered in this work cause the different techniques to return inaccurate results (assuming that results can be returned at all).

As a result of this analysis, the suggestion would be to discard the use of numerical differentiation techniques and to concentrate instead on dealing with the actual distance-time data itself. Through the use of models fitted to the physical data, it should be possible to derive estimates of the kinematics of these disturbances within error.

The size of the data-sets under examination is the main downfall of the numerical differentiation techniques, and is also a cause for concern when fitting a model to the data. This could be overcome through the use of bootstrapping techniques, in particular the use of a residual-resampling bootstrapping analysis.

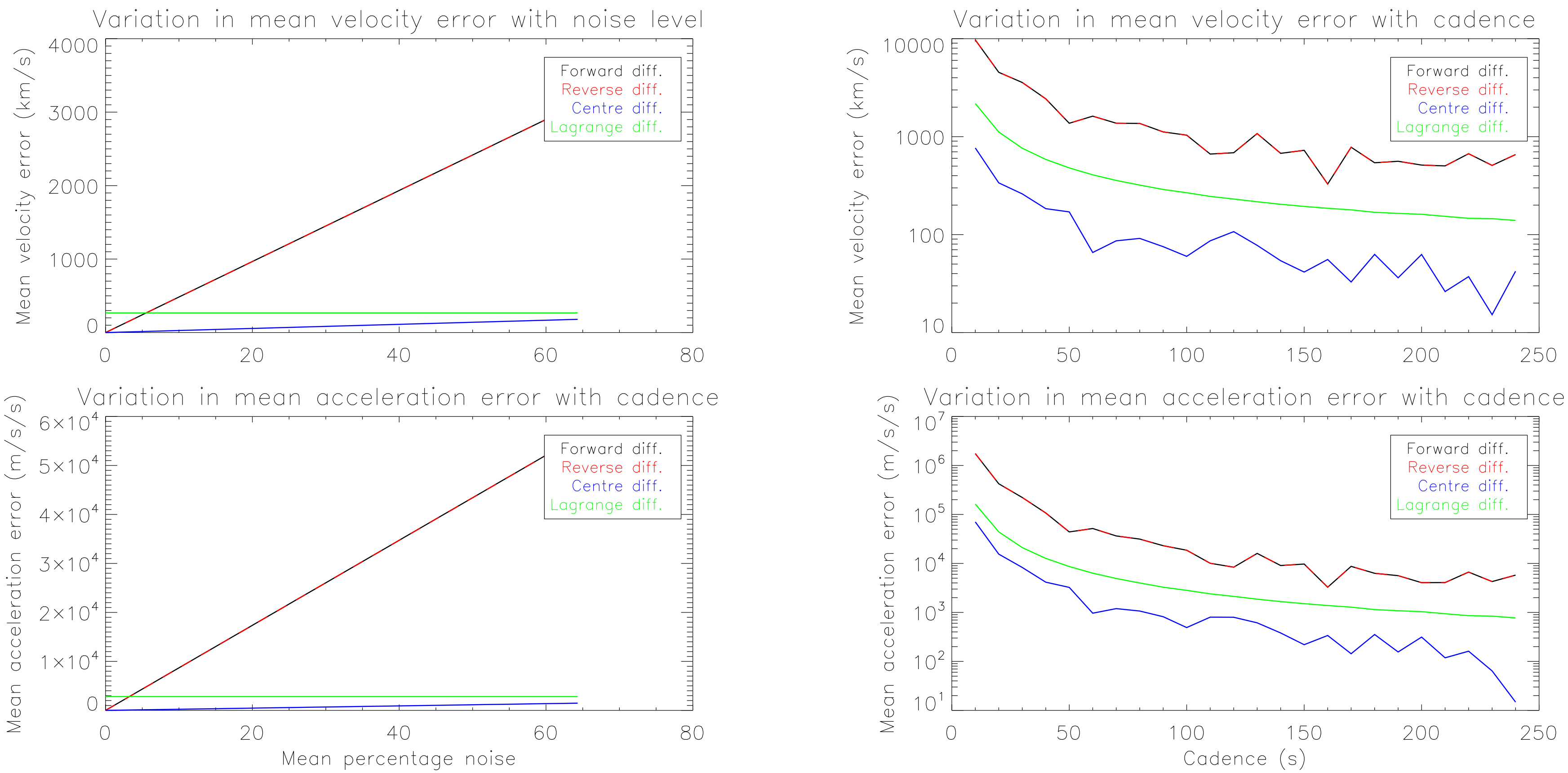

Figure 10 – *Left*: Variation in mean truncation error for velocity (top) and acceleration (bottom) with variation in mean percentage noise at a cadence of 100 s. *Right*: Variation in mean truncation error for velocity (top) and acceleration (bottom) with variation in cadence at a mean percentage noise of ∼21%.

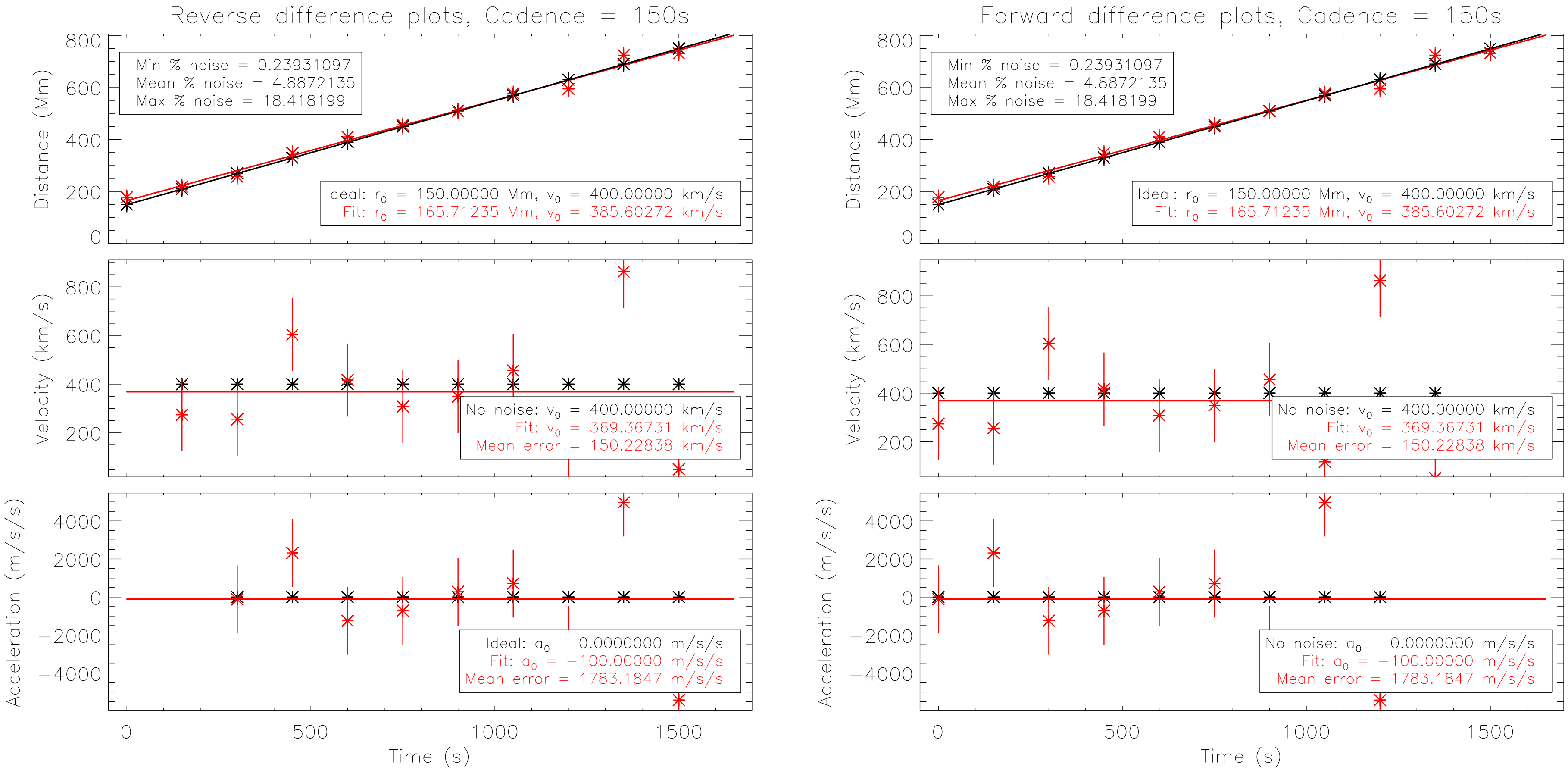

Figure 11 – *Left*: Realistic 171 Å-like data treated using the forward-difference method, showing the distance-time (top), derived velocity-time (middle) and derived acceleration-time (bottom). *Right*: Realistic 171 Å-like data treated using the reverse-difference method, showing the distance-time (top), derived velocity-time (middle) and derived acceleration-time (bottom).

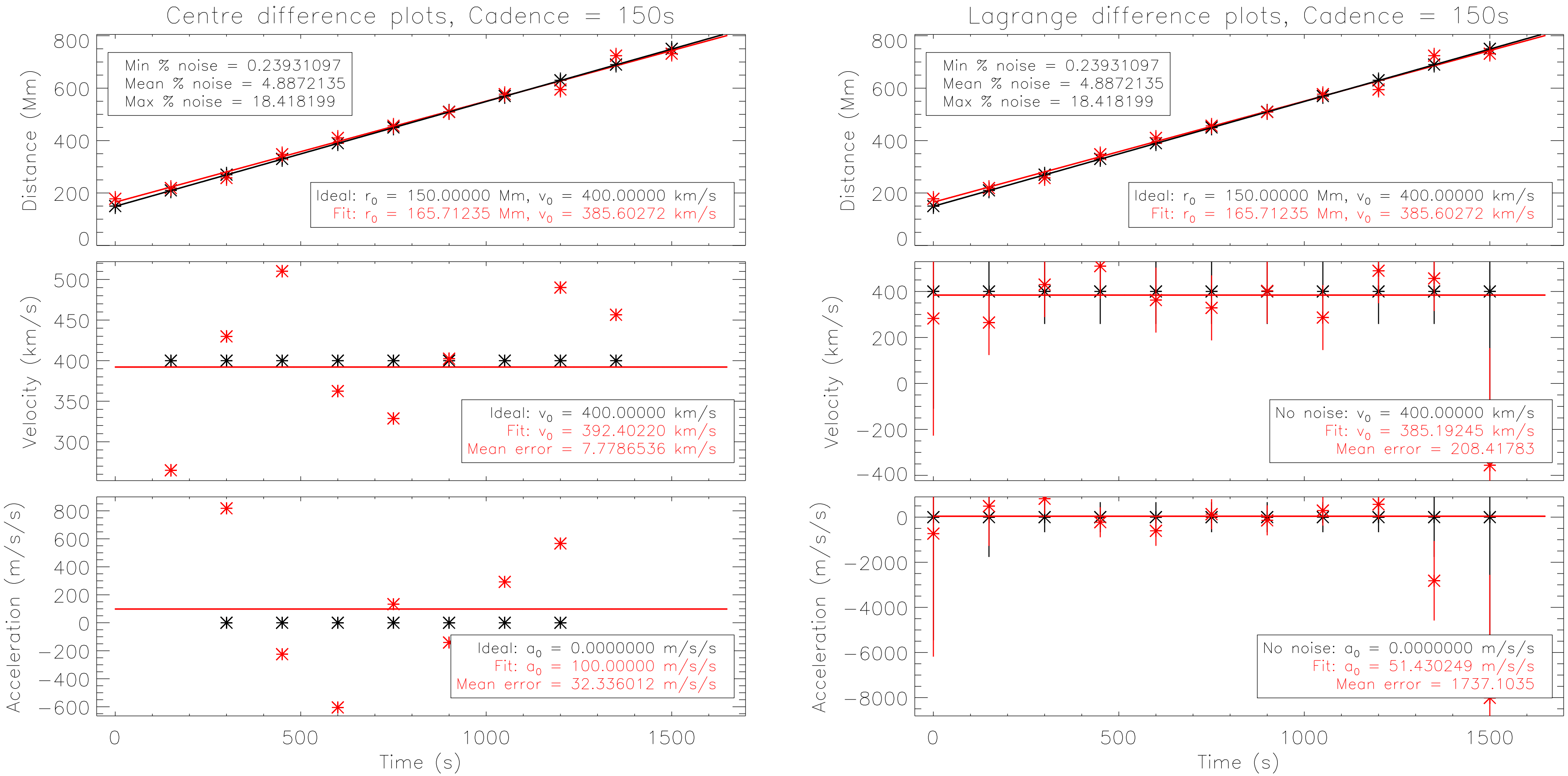

Figure 12 – *Left*: Realistic 171 Å-like data treated using the centre-difference method, showing the distance-time (top), derived velocity-time (middle) and derived acceleration-time (bottom). *Right*: Realistic 171 Å-like data treated using the Lagrange-difference method, showing the distance-time (top), derived velocity-time (middle) and derived acceleration-time (bottom).

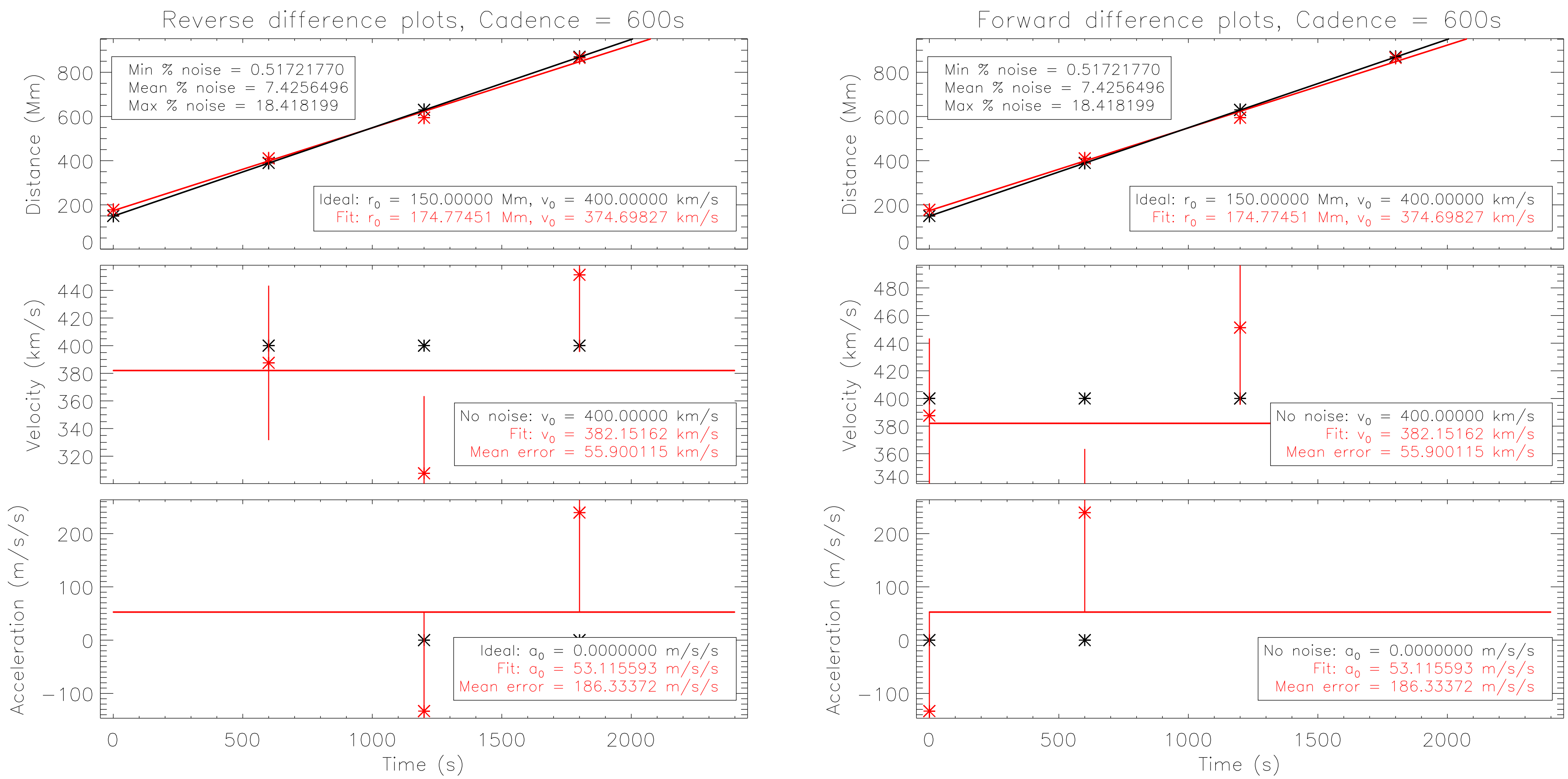

Figure 13 – *Left*: Realistic 195 Å-like data treated using the forward-difference method, showing the distance-time (top), derived velocity-time (middle) and derived acceleration-time (bottom). *Right*: Realistic 195 Å-like data treated using the reverse-difference method, showing the distance-time (top), derived velocity-time (middle) and derived acceleration-time (bottom).

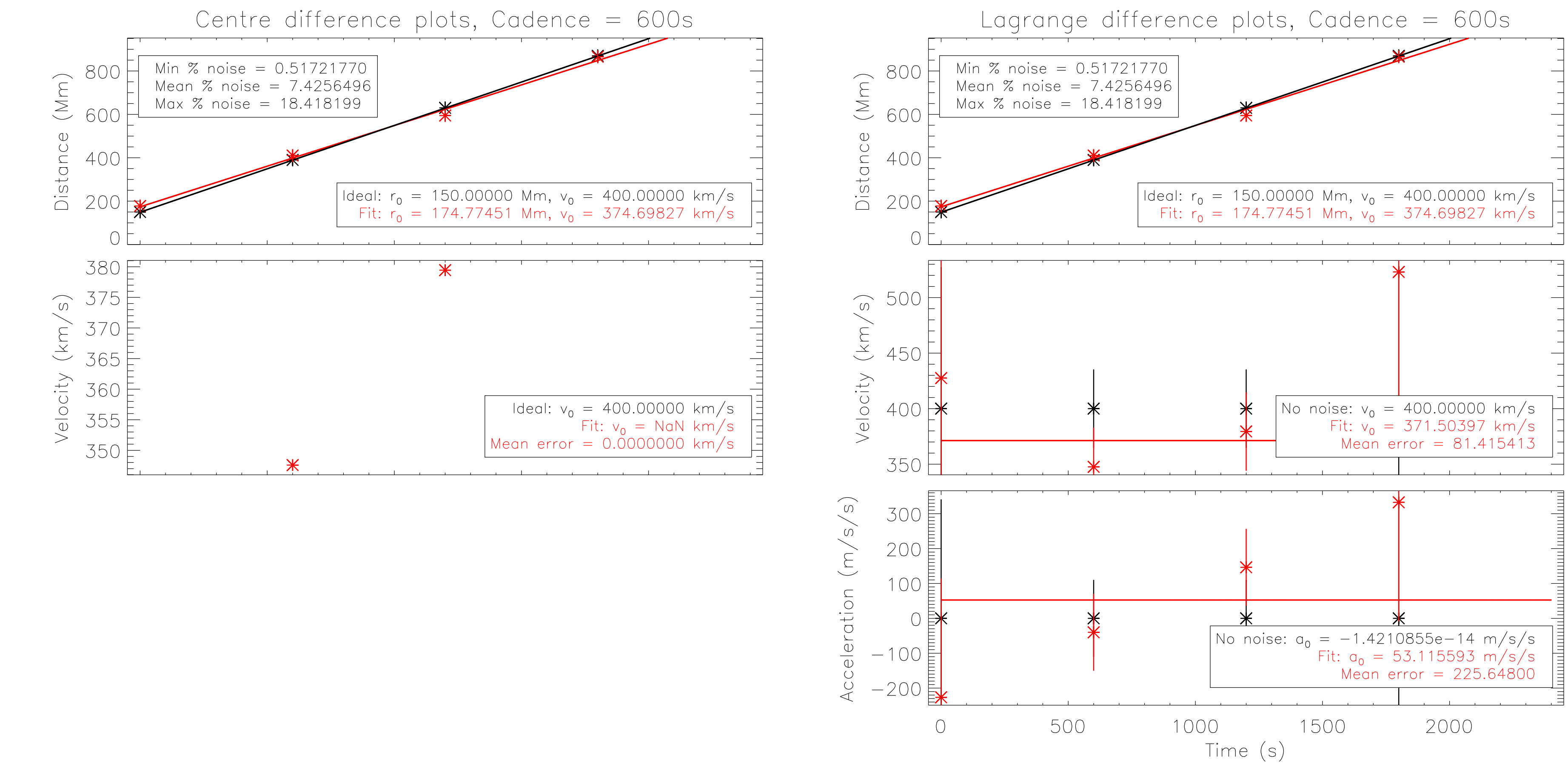

Figure 14 – *Left*: Realistic 195 Å-like data treated using the centre-difference method, showing the distance-time (top), derived velocity-time (middle) and derived acceleration-time (bottom). *Right*: Realistic 195 Å-like data treated using the Lagrange-difference method, showing the distance-time (top), derived velocity-time (middle) and derived acceleration-time (bottom).