# Trend Damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment?

Nigel Harvey[1] and Stian Reimers[1,2]

[1]University College London, UK

[2]City University London, UK

Nigel Harvey

Department of Cognitive, Perceptual and Brain Sciences

University College London

Gower Street

London WC1E 6BT

UK

Tel: +44 207 679 5387

Fax: +44 207 436 4276

n.harvey@ucl.ac.uk

Running head: Trend damping

Abstract

People's forecasts from time series underestimate future values for upward trends and overestimate them for downward ones. This trend damping may occur because 1) people anchor on the last data point and make insufficient adjustment to take the trend into account, 2) they adjust towards the average of the trends they have encountered within the experiment, or 3) they are adapted to an environment in which natural trends tend to be damped. Two experiments eliminated the first account: for series that are negatively accelerated or have shallow slopes, people showed anti-damping (the opposite of damping), a phenomenon that cannot be interpreted in terms of under-adjustment. These experiments also produced results consistent with the second account: forecasts for a given function clearly depended on the other functions that were forecast within the same experiment. However, this second account was itself eliminated by a third experiment demonstrating both damping and, to a lesser degree, anti-damping when people forecast from a single series. We conclude that people have adapted to degrees of growth and decay that are representative of their environment: damping occurs when trends in presented series are steeper than this and anti-damping occurs when they are shallower.

Keywords: adaptation, under-adjustment, forecasting, trend damping, context effects

How do people predict the future and how does that help them to control it? In dynamic judgment tasks, people are presented with time series information and use their judgment to make forecasts (Lawrence, Goodwin, O'Connor & Önkal, 2006), to gain control over the series (Osman, 2010), or to detect a change in the way it is being produced (Brown & Steyvers, 2009). To perform these tasks well, they have to take some account of how information is patterned over time. This feature distinguishes them from static tasks such as multiple cue probability learning (Cooksey, 1996), function learning (Kalish, Lewandowsky & Kruschke, 2004), and causal learning (Shanks, Holyoak & Medin, 1996). However, though static and dynamic tasks differ in terms of the demands they make on temporal pattern perception, they are similar in other ways. For example, both require systematic relations to be separated from random noise and the way that this separation is achieved may be similar for the two classes of task.

Here we focus on one particular dynamic task, judgmental forecasting. However, the issue that concerns us is relevant to other dynamic and, indeed, many static tasks. Specifically, we address the issue of whether people's performance depends on use of very simple judgment heuristics or on more complex pattern extraction strategies.

People's forecasts from time series are subject to a number of systematic errors: they add random noise to their forecasts (Harvey, 1995), their forecasts show less regression from the last data point towards the mean of the series than they should do (Reimers & Harvey, 2011), and they damp trends in data so that forecasts lie below upward trend lines but above downward ones (Lawrence & Makridakis, 1989). These errors have been taken as evidence that forecasting performance relies on simple heuristics, such as representativeness and anchoring-and-adjustment (Tversky & Kahneman, 1974).

Here we focus on trend damping. We first outline the accepted heuristic account of it and then consider two other explanations. Then we report three experiments designed to select between these alternatives.

*Trend damping as under-adjustment*

Hogarth and Makridakis (1981) argued that judgmental forecasters use the same cognitive heuristics that can produce judgmental biases in other domains (Tversky & Kahneman, 1974). When they forecast from time series, people would use the anchoring-and-adjustment heuristic (Eggleton, 1982). They could treat long-term mean of the series as a mental anchor and adjust away from it to take the trend into account. Alternatively, they could use the last data point in the series, which, on average, will lie on the trend line, as their anchor and adjust away from that to allow for the series trend. In either case, the under-adjustment that is associated with use of this heuristic (Tversky & Kahneman, 1974) would produce the observed damping effect.

*Trend damping as a context effect*

There have been many studies of trend damping. Typically, people have forecast from a number of series, some of which have upward trends and some of which have downward ones (e.g., Bolger & Harvey, 1993; Harvey & Bolger, 1996; Lawrence & Makridakis, 1989; Mackinnon & Wearing, 1991; O'Connor, Remus, & Griggs, 1997). If people in those studies regressed their forecasts towards the mean trend they encountered within their respective experiments, the observed damping effects would have occurred. This contextual explanation attributes damping to a contraction bias (Poulton, 1989), a widespread phenomenon that is also known variously as the central tendency of judgment (Hollingsworth, 1910), regression (Stevens & Greenbaum, 1966), and assimilation (Warren, 1985).

Although Poulton (1989) referred to this type of context effect as a bias, it may reflect a reasonable response on the part of participants to their uncertainty about the steepness and/or acceleration of the trend on each trial. If only the mean trend across trials can be estimated, expected forecast error is minimized by forecasting according to that estimated mean trend. If, however, the trend on a particular trial can be estimated subject to uncertainty, expected forecast error is minimized by forecasting according to a trend produced by taking a weighted average of the estimated trend on that trial and the estimated mean trend.  This sensible strategy would have produced trend damping in the studies cited above. However, accounting for the effect in these terms implies that it should be regarded as an artifact of the mixture of trends included in the experimental designs used in these studies.

*Trend damping as adaptation*

This final explanation is an extension of the context-effect account, but assumes that we are adapted not just to the immediate context of the experiment, but to the wider context of the environment more generally. If trends in real series that do not currently show negative acceleration tend to develop such acceleration in the future, then people able to take that into account are likely to damp current trends when making their forecasts.

Why would the future points in real series show negative acceleration? In the natural world, growth initially tends to accelerate positively because sufficient resources are available to allow it to continue in an unconstrained manner. It may, for example, approximate an exponential or power function. Eventually, however, the demands of this type of growth outpace the resources available. As a result, the original pattern of growth becomes damped. Thus, series that initially show positive acceleration become sigmoidal. For example, adding a damping term to the differential equation for exponential growth produces logistic growth. Such logistic growth initially accelerates in a way that is indistinguishable from exponential growth but later the effects of resource constraints produce an inflection in the curve. It decelerates and eventually levels off at a value known as the carrying capacity of the environment.

Many studies have shown that sigmoidal growth is characteristic of a very wide variety of natural time series (Tsoularis & Wallace, 2002). These include population growth of animals and plants (Law, Murrell

& Dieckmann, 2003; Watson, 1984), growth of tumors (Foryś & Marciniak-Czochra, 2003; Ledzewicz, Munden & Schättler, 2009), growth of linguistic variants in speech communities (Altmann, Buttlar, Rott, & Strauß, 1983; Niyogi & Berwick, 1995), growth of charitable donations after natural disasters (Schweitzer & Mach, 2008), diffusion of technologies (Kucharavy & De Guio, 2011), changes in energy consumption (Bodger & Tay, 1987), growth of human populations within individual countries (Meade, 1988), the socio-economic growth of countries (Herman & Montroll, 1972), and many more. Although there is broad agreement that growth curves are typically sigmoidal, it is recognized that exceptions occur when growth of one element influences that of another, as in the predator-prey relations described by Lotka-Volterra equations. In these cases, both elements typically show cycles of growth and decay.

Whether growth levels off or is followed by decay, it is clear that, from experience of their environment, people can reasonably expect that growth that is positively accelerating will, in the future, decelerate and that growth that has already started to decelerate will continue to do so. This means that, in the words of Lawrence and Makridakis (1989), trend damping "appears to reflect a good deal of common sense".

If it is reasonable to expect that trends in natural series will decelerate (or accelerate less) in the future, forecasts that are based on trends extracted from data collected up to the present time should be damped to increase their accuracy. So is there any evidence that forecasts based purely on statistical analysis of the available data can be improved by damping? Yes, there is. Gardner and McKenzie (1985) compared linear trend and damped trend versions of exponential smoothing forecasts for the 1001 real time series originally collected for the M1 forecasting competition (Makridakis, Anderson, Carbone, Fildes, Hibon, Lewandowski, Newton, Parzen, & Winkler, 1982). They found that damping the trend in series improved forecast quality for longer forecast horizons. In the later M3 forecasting competition (Makridakis & Hibon, 2000), a similar comparison for 3003 real series confirmed this finding. Furthermore, Collopy and Armstrong (1992) and Adya, Armstrong, Collopy and Kennedy (2000) found that including rules in their expert system that damped trends improved the way it produced forecasts from real series. All these results demonstrate that, on average, trends observed when data are taken from a particular sampling period are steeper than they would have been if that sampling period had been extended. In other words, natural time series are, on average, less accelerated than limited samples of them indicate.

Gardner and McKenzie's (1985) analysis also led them to conclude damping should be greater when series are noisier or when trends are erratic. Furthermore, Collopy & Armstrong's (1992) found that forecasting by their expert system was improved if they included rules specifying that damping should be greater when series are noisier and trends are less clear. Thus, the greater damping with noisier series observed in judgmental forecasting experiments (Eggleton, 1982; Harvey & Bolger, 1996) can be explained within an ecological framework. These results suggest that, when uncertainty in the data is higher, greater weight should be given to ecological knowledge

The growth rate in sigmoidal functions can vary. Increasing their growth rate parameters causes carrying capacity to be reached more quickly. Presumably, there is some growth rate that is broadly representative of the environment. As there will be deceleration whatever the growth rate, we can expect some damping of positively accelerated series. However, it is harder to predict the degree of damping without knowing what growth rate is representative of the environment.

What should we expect for negatively accelerated series? These series correspond to that part of the growth period during which carrying capacity is being reached. If it is being reached quite slowly so that the series is failing to become asymptotic when people would expect it to do so from their experience of their natural environment, the adaptation account predicts that trend damping will occur and that this damping will be greater when series are noisier. On the other hand, if the asymptote is being reached more quickly than people expect from their experience of their natural environment, then this account predicts that the opposite of damping will occur. In other words, when series are sufficiently negatively accelerated, people will make forecasts that are above upward sloping trend lines and below downward sloping ones. For convenience, we shall term such a phenomenon 'anti-damping'. To date, it has not been demonstrated. If it exists, we expect that, like damping (and for the same reasons), it will be greater when series are noisier.

## Experiment 1

This experiment was designed to examine two key issues. First, we ask whether anti-damping can be observed for decelerating curves, a finding that would challenge the under-adjustment account of trend damping, and whether noise increases damping and anti-damping. The second issue concerns whether forecasts for a given function are affected by the acceleration of the other functions presented in the experiment. A finding that they are would provide support for the two context-sensitive models.

The under-adjustment account would not be able to explain any anti-damping that we obtain. Any such effect would correspond to over-adjustment: it would arise because the adjustments away from the last data point (or away from the long-term mean of the series) that people make to take the series trend into account are too large.

The context effect and adaptation models also predict that both damping and anti-damping will be greater when series are noisier. This is because, with higher uncertainty, more weight would be put on context, either from within the experiment (contextual account) or from the environment more generally (adaptation account). Thus, to test this prediction, we included high and low noise series among those from which participants made their forecasts.

The context effect account predicts that degree of damping for any given trend will depend on the average steepness of trends that people encounter within an experimental session. The adaptation account also makes this prediction. We have argued that ecological knowledge provides people with an *a priori* expectation of the degree of acceleration in data series from which they make forecasts. This knowledge must be derived from data series that they have experienced in their natural environment. However, given that environments change, it is likely to be mutable. If we assume that people sample their environment in a naïve (non-selective) way, they are likely to be as sensitive to the laboratory environment as to any other. Hence, by providing people with contexts comprising series that have either high or low acceleration, we should be able to change the characteristics of the ecological knowledge that people use when making their forecasts.

This means that both the context effect account and the adaptation account predict that, for a given series, damping should be less after a context of high acceleration series than after a context of low acceleration series. In contrast, the under-adjustment account of the phenomenon would require further elaboration to explain any effect of context. Thus to examine whether such any effect occurs, we studied damping of two target series under two conditions. In one condition, all the series (including the target series) were positively accelerated. In the other, only the target series were positively accelerated; the other series were linear and negatively accelerated.

*Method*

Participants completed a set of 20 forecasting trials, with distribution of function accelerations manipulated between participants. The experiment was run online.

*Participants.* A total of 793 sets of data were submitted. Of these, 451 participants indicated they were female, 321 indicated they were male, and 21 did not give their gender. Participants reached the experiment through web searches, clicking links from our main testing site, or from a list of web-based psychology studies hosted by Hanover College http://psych.hanover.edu/Research/exponnet.html.

*Design.* We constructed time series of 50 observations using power-law functions, of the general form:

$$y = 50 + 300 \times \left(\frac{x}{50}\right)^k \qquad \text{Equation 1}$$

where $y$ represents number of pixels above the x-axis, $x$ has the dimension time, and $k$ controls the acceleration of the function. For $k < 1$, the function has negative acceleration, for $k = 1$, the function is linear, and for $k > 1$ the function is positively accelerated. For all values of $k > 0$, functions have positive slope for all x. Constants ensured that the value of the function at the final observation, $x = 50$, was always at the same point on the screen – 350 pixels above the x-axis – irrespective of $k$, and that all pre-noise observations were between 50 and 350 pixels above the x-axis.

Using different values of k, we created a total of eight functions. However, a given participant would only see a subset of these functions, depending on the between-participants variable of context, which had two levels: low and high. The low-context and high-context conditions each contained five different functions, two of which were found in both conditions. In the low context condition, we used Equation 1 with k = 0.2, 0.4, 1.0, 1.5, 2.0. In the high context condition, we had k = 1.25, 1.5, 1.75, 2.0, 2.25. Thus, low- and high- context conditions both contained two identical functions, with k = 1.5 and k = 2.0. However, in the low context condition, all of the other functions participants saw were less accelerated; in the high context condition, some of the other functions participants saw were more accelerated and others less accelerated.

There was one within-participants variable of noise (low: Gaussian noise with M = 0, SD = 3 pixels; high: Gaussian noise with M = 0, SD = 10 pixels). There were 20 trials in total: two repetitions of each of the five functions crossed with the two levels of noise. Trial order was randomized for each participant.

Figure 1 about here

*Procedure.* The experiment was coded in Adobe Flash (see Reimers & Stewart, 2007, for an introduction to using Flash in experiments). Participants received instructions about the experiment, including an animation showing how the selections should be made. Introductory instructions for the task were:

> "In this experiment, you play the part of an advisor in a corporation. You'll see some sets of sales figures, and will have to make predictions about numbers of sales in the future, based on the trends you see. The closer you get to the actual outcome, the higher the score you get. Overall you'll make 20 sets of predictions, and the whole test should last less than 10 minutes - you'll get feedback on your performance at the end."

As part of a short animation showing a sample trend and how to respond, the following instructions were given:

> You'll see a series of graphs like this one. Here you'll see how many sales the company has made in the past 50 sessions. You have to estimate how many sales will be made in the next 8 sales sessions, each of which is indicated with a vertical line. You'll select how many sales you think will be made for the 8 sessions by clicking with the mouse. You'll then get feedback on how close you were to the actual outcome. For each of the predictions you make, you'll get a score between 20 (if you're spot on) and 0 (if you're quite a way out). These 8 scores will be added up and will give you a total score for that trial. Your job is to try to get as high a score as possible. But note that sometimes it will be relatively easy; sometimes it'll be quite hard. At the end of 20 trials you'll get an overall final score and some feedback on your performance. You can track how far you've got by looking at the progress bar that will be at the bottom of the screen.

After reading the information, participants pressed a 'NEXT' button to continue. They then completed questions about their age, gender, and highest level of education. After this, they saw the first trend, presented in a similar way to Figure 1. Participants indicated their predictions by clicking somewhere on each of the eight vertical lines that followed the last data point. A red 'x' appeared where they clicked, and this could be moved by clicking elsewhere on the line. This was the 8-level 'time horizon' variable: The further from the last observation a forecast is, the more distant the time horizon.

After making a prediction on each of the eight lines, participants pressed a ‘SUBMIT’ button. If participants pressed the ‘SUBMIT’ button before making all eight predictions, they received a warning message asking them ensure they had made all the required predictions. Once the predictions were submitted, participants received feedback. Actual values from the noisy trend function were displayed on the vertical lines on which participants had made their prediction, along with a score for each of the eight predictions, ranging linearly from 20 (if prediction and actual outcome were identical) to zero (if prediction and actual outcome were 20 or more pixels apart). A total score for that trial – the sum of the scores for each of the eight predictions – was then displayed in the center of the screen, and added to a running total that remained on the screen throughout the experiment. Participants then clicked on a ‘NEXT’ button to begin the next trial. A progress bar informed participants how many of the 20 trials remained. At the end of the experiment, data were transmitted to the server and feedback was given.

*Results*

We were conservative in our inclusion of datasets, because it was possible that some participants misunderstood the task – and they were not able to ask questions – or that they were responding randomly or mischievously. We first removed 20 participants who indicated they had completed the experiment already, leaving 773. Next, we took the average of each participant’s forecasts across the two exemplars of each trend line at each noise level, and ran an iterative procedure for each of the 160 cells in the design (8 time horizons x 5 function accelerations x 2 noise levels x 2 contexts), to identify any outlying forecasts that were more than two inter-quartile ranges above the upper quartile or below the lower quartile for a given cell. We excluded data from participants who made any outlying forecasts from subsequent analysis. This left 662 participants’ data in the analysis[1]. Average forecasts are given in Figure 2 (low context) and Figure 3 (high context).

---

Tables 1 & 2 and Figures 2 & 3 about here

---

*Trend damping and anti-damping.* It is clear from the results that participants' forecasts deviate systematically from the trend functions. Our first aim was to investigate in which of these conditions significant trend damping or anti-damping occurred. There are two distinct ways in which forecasts could deviate from a trend function. The first is in elevation, that is, participants' predictions are consistently above or below the trend line, but that the difference is the same at all 8 levels of the time horizon. We argue that elevation errors may be interesting phenomena – with their own mechanisms – but are not the same as trend damping, which is what we are concerned with here.

We therefore focus on the second, and to us more interesting, deviation: trend damping. This occurs when participants' forecasts deviate further from a trend line with increasing time horizon[2]. (In a linear function, this would be equivalent to underestimating the gradient of a positive trend line.) Thus, we calculated the signed deviation from the trend line of each point that was forecast by a participant. For each function we ran a repeated-measures ANOVA with the dependent variable of signed deviation from the trend line, and independent variable of time horizon (eight levels). Significant effects, viewed in conjunction with Figures 2 and 3, indicate higher deviation for more distant levels of time horizon than for earlier levels, in other words, significant trend damping. In this analysis and in all of those that follow, we correct degrees of freedom for sphericity violations using the Huynh-Feldt method. Results are shown in Table 1.

These analyses show that significant anti-damping was found for the two shallowest functions in the low context condition, and for the low noise / shallowest function of the high context condition. Traditional trend damping was found in most of the steeper functions in both contexts.

*Effects of noise on forecasting.* Having established the conditions under which trend damping was observed, we investigated the effect of noise. Using the same dependent variable – signed deviation from the trend line – as before, we ran a two-factor ANOVA (noise, time horizon) for each function. Results of these analyses are shown in Table 2. All interactions between noise and forecast horizon are significant. They show that damping and anti-damping effects were always greater when series contained more noise.

---

Figure 4 and Table 3 about here

---

*Effects of context on forecasts.* We examined whether the other functions that participants had recently seen affected their judgment by comparing forecasts for the same trend line in the two between-trials contexts: low (in which the accelerations of the other functions in the experiment were comparatively low) and high (in which the accelerations of the other functions in the experiment were comparatively high). The results can be seen in Figure 4. Forecasts are clearly numerically lower in the low context than in the high context. To test the significance of this effect, we examined the absolute vertical position of forecasts – rather than deviations from trend lines before – as a function of context and time horizon. We ran separate two-factor (context, time horizon) ANOVAs for each function and level of noise (Table 3). The results indicate that forecasts for both functions at both levels of noise were different in the two contexts, and that differences increased with increasing time horizon.

*Discussion*

We obtained anti-damping effects for the two negatively accelerated functions under both levels of noise in the low context condition and for the function with the lowest level of acceleration under low noise in the high context condition. Anti-damping has not previously been shown in forecasts from noisy series. It was predicted by the context effect and adaptation models but cannot be explained in terms of under-adjustment. According to the latter approach, damping should decrease with a decrease in the slope of the presented series but it should not become negative.

As Figures 2 and 3 indicate, both damping and anti-damping were greater when displayed series were noisier. In the past, this has been shown for damping (e.g., Eggleton, 1982; Harvey & Bolger, 1996) but not for anti-damping. As we have seen, these effects of noise are fully consistent with both the context effect and adaptation accounts.

We examined whether the other functions that participants had seen recently affected their judgment by comparing forecasts for two target trend lines in two different contexts: low (in which the level of acceleration in the other functions that participants saw was comparatively low) and high (in which the level of acceleration of the other functions that they saw was comparatively high). We found that forecasts were significantly lower in the low context than in the high context across the two function types and two levels of noise (Figure 4). As we have seen, this fits well with predictions from both the context effect model and the adaptation model of damping. However, it is not immediately explicable in terms of under-adjustment. This is because anchor values and degrees of adjustment are determined within series rather than across series.

## Experiment 2

Having demonstrated anti-damping, noise effects, and trial-to-trial context effects using accelerated functions, we now examine generalization of our findings to a different function shape, specifically, noisy linear functions. As far as we are aware, the effects of this manipulation on the damping of point forecasts have not been investigated before. Indeed Thomson, Önkal-Atay, Pollock and Macaulay (2003, p. 242) point out that ‘despite the considerable attention devoted to trend as an efficacious time-series component, the potential influence of *trend-strength* on forecasting performance has been virtually ignored’.

In our natural environment, linear series do not continue indefinitely. Either they correspond to the central sections of sigmoidal growth functions or they are parts of long-term cycles between peaks and troughs. In either case, there is likely to be some slope that is representative of the environment. When people forecast from noisy linear series that are steeper than this, we should expect the damping that has been demonstrated in a number of studies in the past (e.g., Eggleton, 1982; Harvey & Bolger, 1996; Lawrence & Makridakis, 1989). However, when they forecast from noisy linear series that are less steep than the slope that is representative of linear sections of growth curves in the environment, we should expect anti-damping. Thus, the adaptation account predicts that damping should change to anti-damping as the slope of a linear series is reduced. The account of damping in terms of a context effect makes the same prediction because judgments are assimilated towards the mean trend in the whole set of presented series. However, the account of the effect in terms of under-adjustment merely predicts less damping as the slope of linearly trended series is reduced.

We made one other change in this study. Participants in Experiment 1 were not given financial incentives for good performance. Economists argue that, without such incentives, participants do not put sufficient cognitive effort into experimental tasks (Smith & Walker, 1993). If they are correct, it could be argued that the damping and context effects that we obtained in Experiment 1 arose because our participants followed a 'cognitively lazy' strategy of averaging previous trend lines. In other words, low incentives produced performance that was consistent with the context effect account. We therefore paid participants in Experiment 2, and offered additional payment for accurate performance.

*Method*

*Participants*. A total of 299 British participants were recruited using the ipoints scheme (www.ipoints.co.uk), which rewards participation with virtual 'ipoints', which are exchangeable for consumer goods, flights, and shopping vouchers. (The scheme has since been rebranded as maximiles:

www.maximiles.co.uk) Participants were each paid 100 ipoints (approx USD1) for their time. Participants were informed that the most accurate 10% of participants would receive an additional 100 ipoints, and the participant whose performance was best overall would receive a bonus of 1000 ipoints.

*Design.* A total of eight linear functions were generated, with gradients of 0.6, 1.4, 2.2, 3.0, 3.8, 4.6, 5.4, and 6.2 pixels per inter-observation interval. Functions were constrained such that the noiseless $50^{th}$ observation was the same. There was one between-participants variable, function set (low: lines of gradient 0.6, 1.4, 2.2, 3.0, 3.8; high: lines of gradient 3.0, 3.8, 4.6, 5.4, 6.2). The within-participants variable was noise (low: Gaussian noise with M = 0, SD = 3; high: Gaussian noise with M = 0, SD = 10, as before). There were 20 trials in total: two repetitions of each of the five functions crossed with the two levels of noise.

*Procedure.* Participants received an email sent by ipoints.co.uk inviting them to participate. The email described the experiment in a sentence, informed participants how many ipoints they would receive for participating, and included a link to the experiment, hosted on our servers. At the start of the experiment, participants entered their email address in order to allow payment and bonuses to be allocated. The procedure was in all other respects the same as in Experiment 1.

*Results*

As before, we were conservative in our exclusion of participants. We eliminated any participant whose average prediction (across the two exemplars for each cell) in any of the 10 within-participant cells at any of the eight levels of time horizon was more than two inter-quartile ranges above the upper quartile, or more than two inter-quartile ranges below the lower quartile. A total of 243 participants' data (81%) remained in the analysis[3]. Mean participant responses, collapsing across noise, are given in Figure 5, for the low function set, and Figure 6, for the high function set. It should be noted that the two steepest trend lines of Figure 5 are the same as the two shallowest trend lines of Figure 6.

Our analytical approach for Experiment 2 differed slightly from that used in Experiment 1. As the functions were linear and participants' responses were, on average, approximately linear, rather than using a repeated measures ANOVA design across the eight levels of time horizon, we instead took the simpler approach of fitting a linear regression function to each participant's data, with time horizon (1 – 8) as the x variable and participant prediction (in pixels) as the y variable. We then used slope of the regression lines as our dependent measure. An average slope across participants that was significantly shallower than the trend line would provide evidence of trend damping. (In none of the conditions was there any significant evidence from the regression intercepts that participants' predictions showed an absolute shift in y value. We therefore do not discuss the intercept analysis here.)

---

Figures 5, 6 & 7 and Tables 4 & 5 about here

---

We used planned one-sample t-tests to compare participants' prediction gradients with actual trend line gradients. In the low noise condition (Table 4), participants in the low function set significantly overestimated trend gradients for Functions 1 and 2 (indicating anti-damping), and significantly underestimated trend gradients for Functions 4 and 5 (indicating damping). Participants in the high function set significantly underestimated trend gradients for Functions 6, 7, and 8 (indicating damping). Most importantly, prediction gradients for Functions 4 and 5 were significantly higher in the high function set than in the low function set, showing again that trend damping can be manipulated by context (Figure 7).

Similar results are found in the high noise conditions (Table 5). Participants in the low function set significantly overestimated trend gradients for Functions 1 and 2 (indicating anti-damping), and significantly underestimated trend gradients for Function 5 (indicating damping). Participants in the high function set significantly overestimated trend gradients for Function 4 (indicating anti-damping), and underestimated trend gradients for Functions 7 and 8 (indicating damping). Again, prediction gradients for Functions 4 and 5 were significantly higher in the high function set than in the low function set (Figure 7). Comparing the data in Tables 4 and 5, to examine the effect of noise, we found that in the low function context, participants overestimated the gradient of Function 1 significantly more in the high noise condition than in the low noise condition. Similarly, in the high function context, participants overestimated the gradient of Function 4 significantly more in the high noise condition than in the low noise condition and underestimated the gradient of Functions 7 and 8 significantly more in the high noise condition than in the low noise conditions. In other words, where noise had a significant effect it pushed participants' estimates closer to the average gradient of the function set.

*Discussion*

Despite the fact that participants in this experiment were provided with financial incentives for good performance, damping and anti-damping effects were broadly similar in magnitude to those obtained in Experiment 1 and, when present, the effect of noise was the same as it was before. These results are again consistent with the context effect and adaptation accounts of damping but not consistent with the explanation in terms of under-adjustment from an anchor. They are also important because an increase in damping with an increase in the steepness of the trend of a linear series has not been previously demonstrated.

Participants' forecasts for the two target trend lines were higher when the other lines that participants saw were of steeper gradient than when the other lines participants saw were of shallower gradient. Again, this fits well both with the notion that damping represents a regression to the average trend presented in the experiment and with the idea that it is an adaptation to the forecaster's environment, if that environment is taken to include the local context of the experiment as well as the environment outside the laboratory.

**Experiment 3**

The results of Experiments 1 and 2 are consistent with the contextual account of damping. According to this explanation, both the presence of anti-damping and the effects of other series on damping in the target series arise because of the influence of series from which participants have previously made forecasts within the experiment. Regression towards the average trend seen during the session would produce damping for the steeper trends and anti-damping for the shallower ones. A change in the average trend would shift forecasts in the direction of the change.

As far as we are aware, no experiments have been reported in which participants made predictions for just a single trended series. From the perspective of the context effect model, no damping should occur in a single-shot experiment. From the perspective of the adaptation model, it should occur and the degree of damping should reflect participants' experience of the world prior to the experiment. In other words, it should allow us to assess their expectations of how series are trended within their natural ecology. Thus we designed Experiment 3 as a single-shot version of Experiment 1: series acceleration and variability were varied as before but between rather than within participants.

Damping has been found to be greater for downward than for upward trended series. This asymmetry can be explained directly in terms of the adaptation model of damping. Thus, Harvey and Bolger (1996, p. 130) suggested that it may occur because 'people more frequently experience data series that are increasing than data series that are decreasing. As a result they develop expectations about how series typically change and these expectations influence their forecasts'. Alternatively, people may condition their expectations on series type. Most (but not all) experiments on trend damping have required people to forecast quantities for which higher values are better than lower ones (e.g., profits, sales). This could have led to greater damping for downward trended series for two reasons. First, people's expectations for such quantities may have been subject to an optimism bias (Weinstein, 1989). Second, they may have expected actions to be taken to reverse downward trends but not to reverse upward ones (O'Connor et al., 1997).

These accounts of the damping asymmetry are not mutually exclusive: people may be simultaneously influenced by a greater expectation of upward series, by optimism, and by a higher expectation of actions to reverse downward trends. However, predictions of the first of these three accounts for the effects of whether the task is framed as one of forecasting gains or losses are different from the predictions of the second two. If, as the first account assumes, expectations are not conditioned on series type, then the asymmetry should not be affected by this manipulation: all that matters is the relative proportion of upward and downward trended series that participants have encountered before the experiment. However, if expectations are conditioned on series type in some way (via optimism or via anticipation of the effects of selective attempts to reverse the trend), the damping asymmetry should reverse when people forecast losses rather than gains. Hence, in Experiment 3, we also examined the effects of trend direction (up, down) and series type (gains, losses) to find out whether this reversal occurred.

*Method*

*Participants.* Participants were recruited using the ipoints scheme, as in Experiment 2 and were each paid 25 ipoints (approx USD 0.25) for their time. Participants were informed that the 50 most accurate respondents would receive an additional 75 ipoints, and the participant whose performance was the best overall would receive a bonus of 1000 ipoints.

*Design.* Each participant saw a single time series comprising 50 points (Figure 8). Three function shapes were examined and, for each function shape, three factors were varied across participants: slope (positive, negative), noise (low, high) and framing (profit, loss). Within each function type, participants were randomly allocated to one of the cells when the experiment initialized. Function shape was generated using Equation 1, with $k = 0.2$, 1.0 and 1.5. Noise was generated as in Experiments 1 and 2. Negative slope was generated by reflecting a positive-sloped function in the horizontal line passing through the horizontal mid-point of the grey area that can be seen in Figure 8.[4]

Figure 8 about here

*Procedure.* Potential participants received an email from ipoints inviting them to take part by clicking on a link in the email. On clicking, participants opened a URL which hosted an Adobe Flash movie used to run the experiment. There were two versions of instructions. In the profit framing condition, participants received the following instructions:

> In this ultra-quick one-shot experiment, you'll take the role of an advisor to a successful company. Your job is to make predictions about how much profit the company will make in the next eight periods of trading, using trends from the previous 50 periods. The company is concerned to identify levels of profits over these periods because they will determine its future acquisitions policy.
>
> You'll see a graph of how much profit the company has made in the last 50 periods, and all you do is click on the graph to indicate how much profit it will make in each of the next eight periods. By using your judgement based on the existing trend, you should be able to make a fairly accurate prediction. There are no tricks here - the trend you see is based on real trends seen in business forecasting.

In the loss framing condition, participants saw the following text:

> In this ultra-quick one-shot experiment, you'll take the role of an advisor to a loss-making company. Your job is to make predictions about how much money the company will lose in the next eight periods of trading, using trends from the previous 50 periods. The company is concerned to identify levels of losses over these periods because they will determine whether an attempt is made to sell the company.
>
> You'll see a graph of the losses the company has made in the last 50 periods, and all you do is click on the graph to indicate how much loss it will make in each of the next eight periods…

The experimental procedure was in other respects the same as that described for previous experiments.

*Results*

A total of 450 participants made forecasts for decelerating trends, 402 made forecasts for linear trends, and 413 made forecasts for accelerating trends. The data were, unsurprisingly, noisier than the previous within-participants experiments. We were therefore conservative in our exclusion of participants. For each of the 8 cells for each function type, we excluded participants who in any of their eight predictions made an estimate that was more than two interquartile ranges below the lower quartile or more than two inter-quartile ranges above the upper quartile. As before, we used an iterative procedure that updated interquartile ranges after each run and checked again whether any outliers remained. This left 339 participants in the decelerating condition, 309 in the linear condition, and 317 in the accelerating condition[5]. Cells contained between 24 and 56 observations.

Our first aim was to determine whether trend damping occurred in this single-shot experiment and whether it varied with type of series in a manner similar to that observed for Experiment 1. Thus, the dependent variable we used was deviation of forecasts towards the horizontal from the trend line. This means that for both positive and negative slopes, a positive value indicates trend damping. The analysis we conducted was similar to that for Experiment 1: For each function type (decelerating, linear, accelerating), we constructed a general linear model with repeated measure of time horizon (eight levels), and between-participants factors of slope (positive, negative), noise (low, high), and series label (profit, loss). Results of the analysis were as follows[6]. (For clarity, we discuss the influence of series label later.)

*Decelerating function.* Data are shown in the top panel of Figure 9. Anti-damping was evident as a main effect of time horizon on deviation away from the horizontal, $F(4.0, 1316) = 3.60$, $p = .006$. There was also an interaction between time horizon and noise, $F(4.0, 1316) = 2.50$, $p = .04$, suggesting that anti-damping was larger in the high noise condition.

*Linear function.* Data are shown in the middle panel of Figure 9. Trend damping is indicated by a significant effect of time horizon on deviation towards the horizontal $F(3.1, 946) = 120.2$, $p < .001$. There was also an effect of slope $F(1, 301) = 6.93$, $p = .009$, suggesting that trend damping was larger for positive slopes than for negative slopes, a finding which runs contrary to existing research.

*Accelerating function.* Trend damping was found again, as shown by a significant effect of time horizon on deviation towards the horizontal $F(2.1, 656) = 391.7$, $p < .001$. There was a main effect of noise $F(1, 309) = 11.2$, $p = .001$, and an interaction between time horizon and noise $F(2.1, 656) = 6.22$, $p = .002$, which indicates that there was significantly more trend damping in the high noise condition than in the low noise condition.

In summary, the analysis showed both trend damping and anti-damping to be present in our single-shot task. This effect is clear for the linear and accelerating series. Although somewhat weaker, evidence for anti-damping in the decelerating series was still clear and significant.

Figures 9 & 10 about here

The secondary aim of the experiment was to examine the effect of profit-loss series labeling on participants' forecasts. To reveal the general effect of series labels independent of slope, we used absolute prediction as our dependent variable. (Deviation from the trend line towards the horizontal, the variable we used in the above analyses, would have canceled out framing effects when collapsed across slope). Our independent variables were time horizon (1-8) and series label (profit, loss). We also included slope and noise, but as we have reported their effects above, we do not discuss them here – the variables are only included to capture the variance that their manipulations introduced into predictions. Overall effects of series labels can be seen in Figure 10. Average predictions for trends framed as profits were higher (estimated marginal mean deviation = +0.36, SEM = 0.57) than those framed as losses (estimated marginal mean deviation = -2.10, SEM = 0.59), $F(1, 941) = 8.95$, $p = .003$. There was also an interaction between series label and time horizon, $F(2.9, 2705) = 3.53$, $p = .02$, as can be seen in Figure 10, suggesting that the effect of series label gets larger for predictions further into the future.

*Discussion*

Trend damping effects were obtained despite participants completing only a single trial. This provides evidence that the phenomenon is not an experimental artifact related to the mixture of trends seen by participants. Although formal comparisons between the size of the damping effects here and in Experiment 1 are precluded because experiments differed in ways other than single versus multiple trials per participant (e.g., incentive conditions), informal comparison suggests that the damping effect is at least as large, if not larger, in the present study for linear and accelerating functions. Conversely, we also observed significant anti-damping for the decelerating function, suggesting that anti-damping is not merely an artifact of a multiple-trial experimental design.

We also observed a clear labeling bias: When the axes were labeled as ‘profit’, forecasts were significantly higher than when they were labeled as ‘losses’. The finding suggests either the presence of an optimism bias (Weinstein, 1989), or that people expect actions to be taken to reverse upward trends of undesirable quantities and downward trends of desirable ones (O’Connor et al., 1997). Our experiment was not designed to distinguish between these two possibilities. However, it would be simple to do so by repeating the experiment using desirable and undesirable quantities that are recognized as uncontrollable (e.g., volcanic eruptions). Preservation of the asymmetry would favor an explanation in terms of optimism whereas its disappearance would favor one in terms of expectations of selective action to reverse undesirable trends. Given that we have also demonstrated similarly elevated forecasts for ‘profit’-labeled axes relative to ‘loss’-labeled axes in forecasts for untrended but autocorrelated series (Reimers & Harvey, 2011), we argue that the optimism-bias explanation currently has slightly more support.

## General Discussion

Although trend damping is a well-established phenomenon, reasons for its occurrence have been a matter of some debate. It can be considered, at a process level, as a bias that reflects the under-adjustment that characterizes use of the anchor-and-adjustment heuristic. Alternatively, it may be the product of a within-experiment context effect. Finally, at a functional level it can be regarded as an adaptation that ensures that people's forecasts are well-adjusted to the environment in which they are made.

We obtained anti-damping for series with negative acceleration (Experiments 1 and 3) and for linear series with shallow slopes (Experiment 2). This phenomenon cannot be interpreted in terms of under-adjustment either from the long-term mean of the series or from the last data point. It would have to be seen as over-adjustment. However, an anchor-and-adjustment model that permits both under-adjustment and over-adjustment would be of no predictive use without additional assumptions allowing us to specify the conditions under which each type of bias occurs.

Experiment 3 demonstrated damping with positively accelerated series and anti-damping with negatively accelerated series in a single-shot between-participants experiment. This shows that the damping and anti-damping phenomena that we observed in the earlier experiments cannot be attributed solely to the type of within-experiment context effects demonstrated in those studies. Damping and anti-damping may therefore arise from long-term adaptation to the natural environment (Experiment 3), a process that takes into account the immediate context of the experiment as well as longer-term representations of trends (Experiments 1 and 2). As such, our explanation of context effects and context-free judgments using a single, memory-based mechanism is similar to other recent models of judgment and decision making that emphasize the role of sampling both from the immediate experimental context and from long-term memory (Stewart, Chater and Brown, 2006).

The particular type of experience that produces trend damping and anti-damping is experience of trends. We have argued that the primary examples of trends that people come across in their natural environment are those of growth and decay and that it is primarily the sigmoidal and cyclical characteristics of growth and decay curves that explain why damping occurs and why it improves the quality of forecasts (Collopy & Armstrong, 1992; Gardner & McKenzie, 1985).

Our suggestion that judgment is influenced by the fact that certain types of functions are more representative of our environment than others is not new. In cue learning tasks, people complete a series of trials in which they are presented with one or more cues (e.g., employee age), estimate a criterion value (e.g. employee salary), and receive information about the correct criterion value. Learning is faster when the relation between the cue(s) and criterion is positively linear than when it is negatively linear and faster in both these cases than when it is U-shaped. Brehmer (1974) argued that this was because that ordering reflected the relative likelihood of those functional forms in natural environment and people used that ranking to determine the order in which they would attempt to fit those functional forms to the information they received when performing their task. Karelaia and Hogarth (2008) argue from data collected since Brehmer (1974) made his claim that environmental relations between cues and criterion are indeed well fitted by linear models.

Cue learning tasks are static: the environmental relations discussed by Brehmer (1974) were not time-dependent. Have there been comparable suggestions for dynamic tasks? For example, is there evidence, other than that documented above, that people make use of their ecological knowledge of the way information is patterned over time when making forecasts? It is known that forecasts for an untrended series of independent points lie between the last data point and the mean of the series rather than on the mean (Bolger & Harvey, 1993; Eggleton, 1982). Reimers and Harvey's (2011) experiments indicated that this failure to regress to the mean (Kahneman & Tversky, 1973) arose because people made use of their ecological knowledge that natural time series tend to be positively autocorrelated.

*Remaining questions*

First, how does integration of ecological knowledge with information in presented series occur? People may first extract pattern information from the data they are given. Because the series are noisy, this information is uncertain. As a result, they may represent it as a range of possibilities that hold at some level of confidence. Ecological knowledge is then used in some way to select from this range. An alternative possibility is that people use their ecological knowledge as *a priori* hypotheses for the temporal pattern characteristics that are most likely to be present in any series that they encounter. When they do encounter a series, they use the evidence in it to modify those *a priori* hypotheses in some Bayesian manner. Because series are limited in length and noisy, their *a posteriori* hypotheses about the temporal characteristics in the series still show a residual influence of their ecological knowledge[7].

Another issue concerns whether integration of newly encountered temporal patterns into ecological knowledge depends on the format in which those patterns are received. We argued above that the context effects in Experiments 1 and 2 occurred because previously encountered temporal patterns presented as graphs were integrated into ecological knowledge about temporal patterns and that this change in ecological knowledge affected how forecasts were made from later graphs. But would the effects have been the same if the temporal patterns providing the context had been presented as tables of numbers or as sequences of events experienced in real time?

Keren's (1983) study of cultural differences in trend damping suggests that they would. He asked participants in Canada and in Israel to make forecasts for food prices on the basis of a tabular display of prices for each product over the previous four years. The simulated prices in this display increased exponentially. Trend damping was observed for both groups of participants but was much less for Israeli participants. Keren (1983) argued that it arose because they had experienced higher levels of food-price inflation than their Canadian counterparts. In other words, experience of a temporal pattern in a sequence of events in real time was integrated into ecological knowledge that then later influenced how forecasts were made from time series presented in a tabular format. Thus, Keren's (1983) study provides evidence that ecological knowledge is not format-dependent.

A final issue concerns whether the ecological knowledge that people retrieve to use in their forecasting depends solely on the broad pattern in the presented data (e.g., upward versus downward trend) or whether it also depends on the content domain from which those data are drawn. One way of resolving the issue would be to ask one group of people to make forecasts from context series containing steep trends and another group would make forecasts from context series containing shallow trends. Afterwards, both groups would make forecasts for a target series containing a medium trend. For half the participants in each group, context and target series would refer to the same domain; for the other half, they would refer to different domains. A finding that context effects are larger when context and target series refer to the same content domain than when they refer to different domains would imply that ecological knowledge is domain dependent.

*Implications*

Forecasting practitioners often make predictions from a sequence of separate time series. For example, sales forecasters make forecasts for a succession of consumer products on the basis of the sales history of each one. Our findings suggest that the forecasts they make for a given product will be influenced by trends in the sales history of products that they have previously forecast.

*Summary*

When people use their judgment to make forecasts from time series data, they do not merely extrapolate from the pattern present in that series. Instead, they deviate systematically from it. In the past, these deviations have been characterized as biases that are diagnostic of the heuristics that are used to make the forecasts. However, the research reported here demonstrates that this approach is inadequate: the new findings cannot be explained in terms of the heuristics that have been used to account for the earlier ones.

To account for the whole data set, we advocate an alternative approach. People's forecasts deviate systematically from the pattern in the presented series because they take account of the fact that that pattern is just a part of a larger pattern that is present in their environment. The nature of the deviations of their forecasts from an extrapolation of the pattern in the presented series can be predicted from the way the pattern in the presented series differs from the larger ecologically representative pattern of which it forms a part.

Trend damping has been found when people forecast from artificial simulated series in which trends in the presented series persist. However, in real series, trends that are initially present in a series do not persist: they mutate into different ones.  Growth may be initially exponential but later we see that this initial trend is merely the first part of a logistic growth curve. People's forecasts take this into account. Furthermore, taking into account the fact that current trends in real data will become damped as we move into the future is a sensible strategy. We know this because the accuracy of statistical forecasts produced from currently available data is increased by introducing *ad hoc* damping terms (Gardner & McKenzie, 1985).

References

Adya, M., Armstrong, J.S., Collopy, F. & Kennedy, M. (2000). An application of rule-based forecasting to a situation lacking domain knowledge. *International Journal of Forecasting*, *16*, 477-484.

Altmann, G., Buttlar, H.v., Rott, W., & Strauß, U. (1983). A law of change in language. In B. Brainerd (Ed*.), Historical linguistics* (pp. 104-115). Bochum: Studienverlag Dr. N. Brockmeyer.

Bodger, P.S. & Tay, H.S. (1987). Logistic and energy substitution models for electricity forecasting: A comparison using New Zealand consumption data. *Technological Forecasting and Social Change*, *31*, 27-48.

Bolger, F. & Harvey, N. (1993). Context-sensitive heuristics in statistical reasoning. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 46,* 779-811.

Brehmer, B (1974). Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior and Human Performance*, *11,* 1-27.

Brown, S. D. & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology, 58*, 49-57.

Collopy, F. & Armstrong, J. S. (1992). Rule-based forecasting - development and validation of an expert systems-approach to combining time-series extrapolations. *Management Science, 38,* 1394-1414.

Cooksey, R. (1996). *Judgment Analysis: Theory, Methods and Applications*. San Diego: Academic Press.

Eggleton, I. R. C. (1982). Intuitive time-series extrapolation. *Journal of Accounting Research, 20,* 68-102.

Foryś, U. & Marciniak-Czochra A. (2003). Logistic equations in tumour growth modelling. *International Journal of Applied Mathematics and Computer Science*, *13*, 313-325.

Gardner, E. S., Jr. & McKenzie, E. (1985). Forecasting trends in time series. *Management Science, 31,* 1237-1246.

Harvey, N. (1995). Why are judgments less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes, 63,* 247-263.

Harvey, N. (2011). Anchoring and adjustment: A Bayesian heuristic? In W. Brun, G. Keren, G. Kirkebøen, & H. Montgomery (Eds.), *Perspectives on Thinking, Judging, and Decision Making* (pp 98-108). Oslo: Universitetsforlaget.

Harvey, N. & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting, 12,* 119-137.

Herman, R. & Montroll, E.W. (1972). A manner of characterizing the development of countries. *Proceedings of the National Academy of Sciences, USA, 69*, 3019-3023.

Hogarth, R. M. & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management Science, 27,* 115-138.

Hollingsworth, H. L. (1910). The central tendency in judgment. *Journal of Philosophy, Psychology, and Scientific Methods, 7,* 461-469.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Kalish, M.L., Lewandowsky, S. & Kruschke, J.K. (2004). Population of linear experts: Knowledge partitioning and function learning. *Psychological Review, 111*, 1072-1099.

Karelaia, N. & Hogarth, R.M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*, 404-426.

Keren, G. (1983). Cultural differences in the misperception of exponential growth. *Perception & Psychophysics, 34,* 289-293.

Kucharavy, D. & De Guio, R. (2011). Logistic substitution model and technological forecasting. *Procedia Engineering*, *9*, 402-416

Law, R., Murrell, D.J. & Dieckmann, U. (2003). Population growth in space and time: Spatial logistic equations. *Ecology, 84*, 252-262.

Lawrence, M. & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes, 43,* 172-187.

Lawrence, M. Goodwin, P., O'Connor, M & Önkal, D (2006) Judgemental Forecasting: A review of progress over the last 25 Years. *International Journal of Forecasting*, *22*, 493-518.

Ledzewicz, U., Munden, J. & Schättler, H. (2009). Scheduling of angiogenic inhibitors for Gompertzian and logistic tumor growth models. *Discrete and Continuous Dynamical Systems, Series B*, *12*, 415-438.

Mackinnon, A. J. & Wearing, A. J. (1991). Feedback and the forecasting of exponential change. *Acta Psychologica, 76,* 177-191.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time-series) methods: Results of a forecasting competition. *Journal of Forecasting, 1,* 111-153.

Makridakis, S. & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting, 16,* 451-476.

Meade, N. (1988). A modified logistic model applied to human populations. *The Journal of the Royal Statistical Society, Series A, 151*, 1103-1115.

Niyogi, P. & Berwick, R.C. (1995). The logical problems of language change. AI-Memo 1516, Artificial Intelligence Lab, MIT. (ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-1516.pdf)

O'Connor, M., Remus, W., & Griggs, K. (1997). Going up - going down: How good are people at forecasting trends and changes in trends? *Journal of Forecasting, 16,* 165-176.

Osman, M. (2010). Controlling uncertainty: A review of human behavior in complex dynamic environments. *Psychological Bulletin, 136*, 65-86.

Poulton, E. C. (1989). *Bias in Quantifying Judgments.* Hillsdale, NJ: Erlbaum.

Reimers, S. & Harvey, N. (2011). Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting, 27*, 1196-1214.

Reimers, S., & Stewart, N. (2007). Adobe Flash as a medium for online experimentation: A test of RT measurement capabilities. *Behavior Research Methods, 39,* 365-370..

Schweitzer, F. & Mach, R. (2008) The Epidemics of Donations: Logistic Growth and Power-Laws. *PLoS ONE,* 3(1): e1458, doi:10.1371/journal.pone.0001458

Shanks, D.R., Holyoak, K.J. & Medin, D.L. (Eds.), (1996). *The Psychology of Learning and Motivation (Vol. 34): Causal Learning*. San Diego, CA: Academic Press.

Smith, V. L. & Walker, J. M. (1993). Rewards, experience, and decision costs in first price auctions. *Economic Inquiry, 31,* 237-244.

Stevens, S. S. & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception and Psychophysics, 1*, 439-446.

Stewart, N., Chater, N. & Brown, G.D. A. (2006). Decision by sampling. *Cognitive Psychology, 53*, 1-26.

Thomson, M. E., Önkal-Atay, D., Pollock, A. C. & Macauley, A. (2003). The influence of trend strength on directional probabilistic currency predictions. *International Journal of Forecasting, 19,* 241-256.

Tsoularis, A. & Wallace, J. (2002). Analysis of logistic growth models. *Mathematical Biosciences, 179*, 21-55.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Warren, R. M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review, 92,* 574-584.

Watson, W.A. (1984). Developmental constraints: Effect on population growth and patterns of resource allocation in a clonal plant. *The American Naturalist, 123*, 411-426.

Weinstein, N. D. (1989). Optimistic biases about personal risks. *Science, 246,* 1232-1233.

Footnotes

[1] The exact cut-off point does not affect results a great deal. Specifically, in an analysis that does not exclude outliers, all significant effects and interactions in Tables 1, 2 and 3 remain significant.

[2] To be precise, trend damping clearly also requires that average forecasts all lie on the same side of the trend function, to exclude forecasts which intersect the trend line at one of the time horizons. We therefore tested this, and found, as is clear from Figures 2 and 3, that mean forecasts all lay on the appropriate side of the trend line for trend damping

[3] To check the effect of outlier removal, we ran the same analyses on the full dataset. All significant comparisons in Tables 4 and 5 remain significant with the exception of difference between contexts for line 4 in the low noise condition, where in this analysis $p = .06$.

[4] There was inadvertently also a small vertical shift of 10 pixels to the whole function in the flipping process.

[5] Here, outlier removal is more important than in Experiments 1 and 2, as some participants made very extreme predictions. For the decelerating trends, keeping all outliers in the analysis ($n = 450$) leads to no significant antidamping. Using the same procedure as described in the main text to remove only very extreme outliers (4 IQR above upper or below lower quartile, $n = 406$) gives a trend towards antidamping, but not a significant effect ($p = .12$). Excluding only extreme outliers (3 IQR above upper or below lower quartile, $n = 386$) gives a significant antidamping effect ($p = .048$). For the linear and accelerating functions, a main effect of horizon remains when outliers are not removed, suggesting damping is robust across outlier removal strategies.

[6] We completed an alternative analysis by fitting a linear regression to each participant's data and taking from it a slope and intercept. Findings were very similar: main effects in the GLM were largely the same as effects found in the analysis of intercepts, and interactions between time horizon and other variables in the GLM were largely the same as the effects found in the analysis of slopes. We prefer the GLM in this case because it does not assume a linear relationship between time horizon and prediction. Given that two

of the three functions in Experiment 3 were non-linear, this appears more appropriate, even though we found little evidence that participants made non-linear predictions at the group level.

[7] In this second case, ecological knowledge can be regarded as a mental anchor. Adjustment away from it is made on the basis of newly presented evidence. However, because this evidence is subject to uncertainty, adjustment is (and should be) only partial. Thus, in this case, under-adjustment is a sensible strategy. Harvey (2011) has considered anchoring-and-adjustment as an automatically implemented but broadly adaptive strategy that nevertheless produces characteristic errors when over-generalized.

Note

The authors contributed equally to the work reported here. Their research was supported by ESRC grant RES – 000-22-2007. SR was also supported by a fellowship from the ESRC Centre for Economic Learning and Social Evolution.

Table 1. Experiment 1: Damping effect assessed in terms of the main effect of time horizon on signed deviation from the trend line, given for each function type in each condition.

| Condition | Exponent (k) | Noise level | Type of damping effect | |
|---|---|---|---|---|
| Low context | 0.2 | Low | – | $F(2.8, 911) = 68.3$, $p < .01$ |
| | | High | – | $F(5.4, 1787) = 133.1$, $p < .001$ |
| | 0.4 | Low | – | $F(2.1, 679) = 3.22$, $p = .04$ |
| | | High | – | $F(4.2, 1389) = 88.5$, $p < .001$ |
| | 1.0 | Low | 0 | $F(1.9, 643) = 3.01$, $p > .05$ |
| | | High | | $F(3.9, 1297) = 4.52$, $p = .001$* |
| | 1.5 | Low | + | $F(1.4, 473) = 114.5$, $p < .001$ |
| | | High | + | $F(3.2, 1073) = 211.5$, $p < .001$ |
| | 2.0 | Low | + | $F(1.2, 406) = 245.7$, $p < .001$ |
| | | High | + | $F(2.0, 652) = 692.9$, $p < .001$ |
| High context | 1.25 | Low | – | $F(1.7, 546) = 13.9$, $p < .001$ |
| | | High | 0 | $F(3.4, 1111) = 2.33$, $p = .06$ |
| | 1.5 | Low | + | $F(1.5, 495) = 10.2$, $p < .001$ |
| | | High | + | $F(3.1, 1013) = 30.5$, $p < .001$ |
| | 1.75 | Low | + | $F(1.5, 487) = 70.3$, $p < .001$ |
| | | High | + | $F(2.6, 851) = 177.9$, $p < .001$ |
| | 2.0 | Low | + | $F(1.3, 441) = 163.2$, $p < .001$ |
| | | High | + | $F(2.2, 721) = 301.6$, $p < .001$ |
| | 2.25 | Low | + | $F(1.2, 407) = 281.0$, $p < .001$ |
| | | High | + | $F(2.3, 758) = 641.8$, $p < .001$ |

Negative damping effects refer to anti-damping.
*In this case, damping did not meet the criterion that all mean forecasts should lie on the same side of the trend line.

Table 2. Experiment 1: Results of two-way ANOVAs on signed error for each function type in each condition.

| Condition | Exponent (k) | Noise | Horizon | Interaction |
|---|---|---|---|---|
| Low context | 0.2 | $F(1, 331) = 104.1$<br>$p < .001$ | $F(4.3, 1410) = 176.0$<br>$p < .001$ | $F(5.5, 1805) = 49.5$<br>$p < .001$ |
| | 0.4 | $F(1, 331) = 100.0$<br>$p < .001$ | $F(2.9, 968) = 71.2$<br>$p < .001$ | $F(4.6, 1518) = 66.2$<br>$p < .001$ |
| | 1.0 | $F(1, 331) = 5.16$<br>$p = .02$ | $F(3.1, 1013) = 3.73$<br>$p = .01$ | $F(3.6, 1183) = 4.66$<br>$p = .002$ |
| | 1.5 | $F(1, 331) = 40.2$<br>$p < .001$ | $F(2.2, 714) = 282.4$<br>$p < .001$ | $F(2.8, 921) = 35.6$<br>$p < .001$ |
| | 2.0 | $F(1, 331) = 119.5$<br>$p < .001$ | $F(1.4, 472) = 702.3$<br>$p < .001$ | $F(2.0, 646) = 126.5$<br>$p < .001$ |
| High context | 1.25 | $F(1, 329) = 0.25$<br>NS | $F(2.6, 842) = 5.97$<br>$p = .001$ | $F(3.2, 1040) = 4.39$<br>$p = .004$ |
| | 1.5 | $F(1, 329) = 8.32$<br>$p = .004$ | $F(2.2, 728) = 34.4$<br>$p < .001$ | $F(2.8, 914) = 10.1$<br>$p < .001$ |
| | 1.75 | $F(1, 329) = 6.50$<br>$p < .001$ | $F(2.0, 667) = 3.73$<br>$p < .001$ | $F(2.4, 789) = 52.4$<br>$p < .001$ |
| | 2.0 | $F(1, 329) = 119.6$<br>$p < .001$ | $F(1.7, 561) = 366.8$<br>$p < .001$ | $F(2.2, 731) = 66.3$<br>$p < .001$ |
| | 2.25 | $F(1, 329) = 155.0$<br>$p < .001$ | $F(1.6, 515) = 684.3$<br>$p < .001$ | $F(2.2, 728) = 108.3$<br>$p < .001$ |

Table 3. Experiment 1: Results of two-way ANOVAs on vertical positions of forecasts showing effects of context for two target functions in low and high noise conditions.

| Noise | Target | Context | Horizon | Interaction |
|---|---|---|---|---|
| Low | k = 1.5 | F(1, 660) = 36.1<br>p < .001 | F(1.5, 967) = 11985<br>p < .001 | F(1.5, 967) = 31.3<br>p < .001 |
| | k = 2.0 | F(1, 660) = 23.8<br>p < .001 | F(1.3, 841) = 10233<br>p < .001 | F(1.3, 841) = 13.6<br>p < .001 |
| High | k = 1.5 | F(1, 660) = 41.0<br>p <.001 | F(3.2, 2110) = 5005<br>p < .001 | F(3.2, 2110) = 31.3<br>p < .001 |
| | k = 2.0 | F(1, 660) = 13.3<br>p < .001 | F(2.1, 1402) = 4540<br>p < .001 | F(2.1, 1402) = 13.7<br>p < .001 |

Table 4. Experiment 2: Trend line gradients and mean gradients of regression lines fitted to individual participant predictions. Low noise condition.

| Trend line number | Trend line gradient | Low context prediction gradient | High context prediction gradient | Difference between contexts |
|---|---|---|---|---|
| 1 | 0.6 | 0.89** | | |
| 2 | 1.4 | 1.77** | | |
| 3 | 2.2 | 2.32 | | |
| 4 | 3.0 | 2.80* | 3.03 | 0.23§ |
| 5 | 3.8 | 3.34** | 3.68 | 0.34§ |
| 6 | 4.6 | | 4.15** | |
| 7 | 5.4 | | 4.95** | |
| 8 | 6.2 | | 5.42** | |

* Significantly different from trend line gradient, $p<.01$

** Significantly different from trend line gradient, $p<.001$

§ Significant difference between low and high context prediction gradients, $p<.05$

Table 5. Experiment 2: Trend line gradients and mean gradients of regression lines fitted to individual participant predictions. High noise condition.

| Trend line number | Trend line gradient | Low context prediction gradient | High context prediction gradient | Difference between contexts |
|---|---|---|---|---|
| 1 | 0.6 | 1.20**† | | |
| 2 | 1.4 | 1.87** | | |
| 3 | 2.2 | 2.28 | | |
| 4 | 3.0 | 2.78 | 3.51**† | 0.73§ |
| 5 | 3.8 | 3.15** | 3.90 | 0.75§ |
| 6 | 4.6 | | 4.39 | |
| 7 | 5.4 | | 4.53**† | |
| 8 | 6.2 | | 4.89**† | |

** Significantly different from line trend gradient, $p < .001$

§ Significant difference between low and high context prediction gradients, $p < .001$

† Significantly different from within-participants low-noise gradient (see Table 4), $p < .01$

Figure Captions

*Figure 1*. Screenshot showing basic experimental setup for Experiment 1 and 2. Participants made their predictions by clicking on each of the eight closely spaced vertical lines, which are indicated to the right of the data series. In the experiment, participants' predictions appeared as red crosses whereas the data series was displayed using black crosses..

*Figure 2.* Trend lines and actual participant estimates for Experiment 1, low function context condition. For clarity, only some of the trend line is shown. The panels, going from left to right and then top to bottom, show power-law functions with exponents of 0.2, 0.4, 1.0, 1.5, 2.0.

*Figure 3.* Trend lines and actual participant estimates for Experiment 1, high function context condition. For clarity, only some of the trend line is shown. The panels, going from left to right and then top to bottom, show power-law functions with exponents of 1.25, 1.5, 1.75, 2.0, 2.25.

*Figure 4.* Comparison of forecasts for the two trend lines that appeared in both low and high contexts. The upper two panels show the function with an exponent of 1.5; the lower two panels show the function with an exponent of 2.0. Left panels show the low noise condition; right panels the high noise condition.

*Figure 5.* Trend lines and actual participant estimates for Experiment 2, low function context condition. For clarity, only some of the trend line is shown. The panels, going from left to right and then top to bottom, show functions with gradients of 0.6, 1.4, 2.2, 3.0, and 3.8 pixels per sales period. Note that the last two panels have the same trend lines as the first two panels of Figure 6.

*Figure 6.* Trend lines and actual participant estimates for Experiment 2, high function context condition. For clarity, only some of the trend line is shown. The panels, going from left to right and then top to bottom, show functions with gradients of 3.0, 3.8, 4.6, 5.4, and 6.2 pixels per sales period. Note that the first two panels have the same trend lines as the last two panels of Figure 5.

*Figure 7.* Comparison of forecasts for the two trend lines that appeared in both low and high contexts. The upper two panels show the function with a gradient of 3.0; the lower two panels show the function with a gradient of 3.8. Left panels show the low noise condition; right panels the high noise condition.

*Figure 8.* Screenshot from Experiment 3. In this case, axes were labeled ‘profits’ or ‘losses’ on both left and right side, to maximize salience. Note the absence of feedback.

*Figure 9*. Trend lines and mean predictions for decelerating (top panel), linear (middle panel), and accelerating (bottom panel) trends, collapsed across label and slope.

*Figure 10.* Series label effects in Experiment 3 as a whole. On average, participants’ predictions for trends framed as ‘profits’ were higher than for trends framed as ‘losses’.

Figure 1

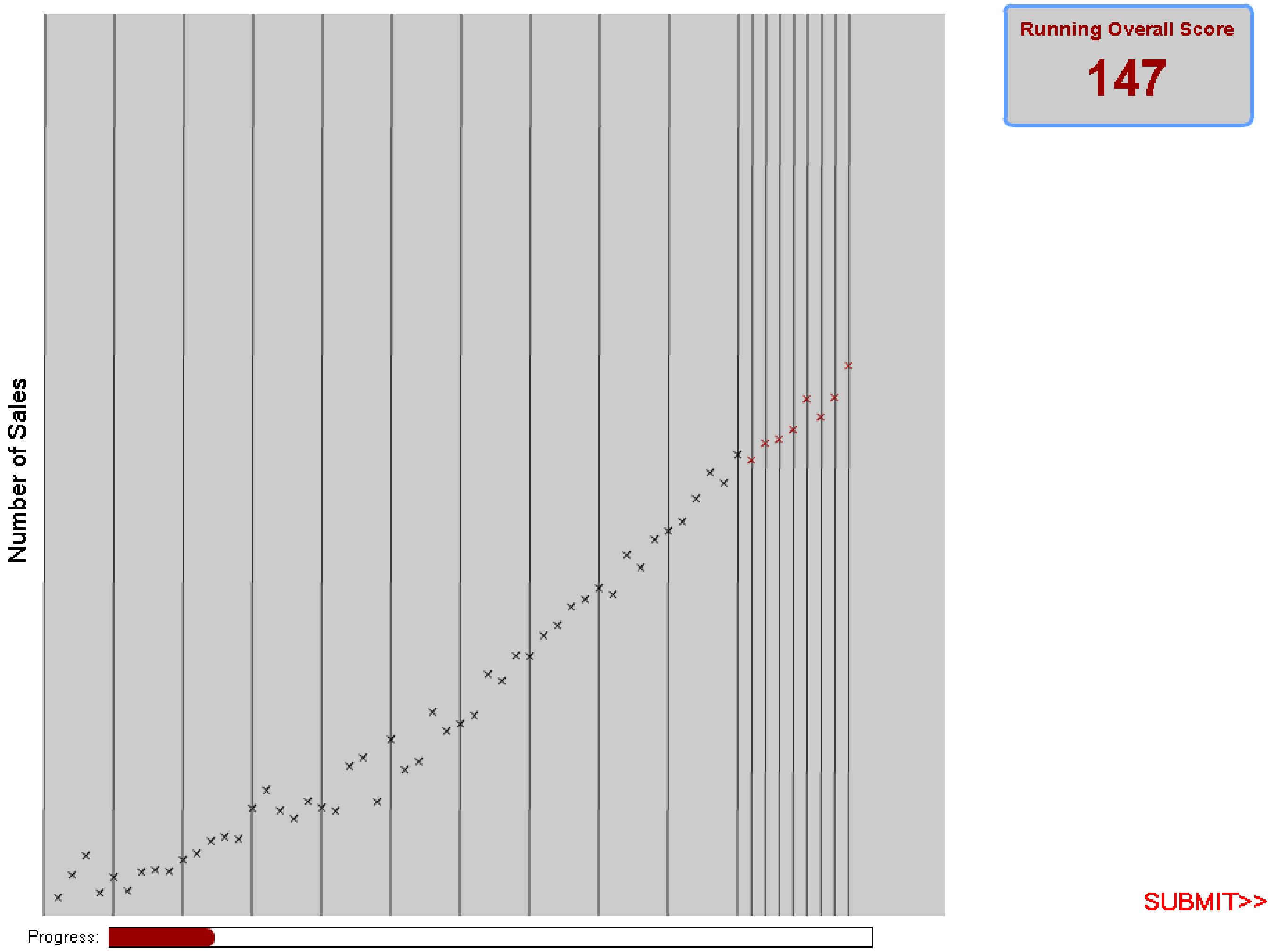

Figure 2

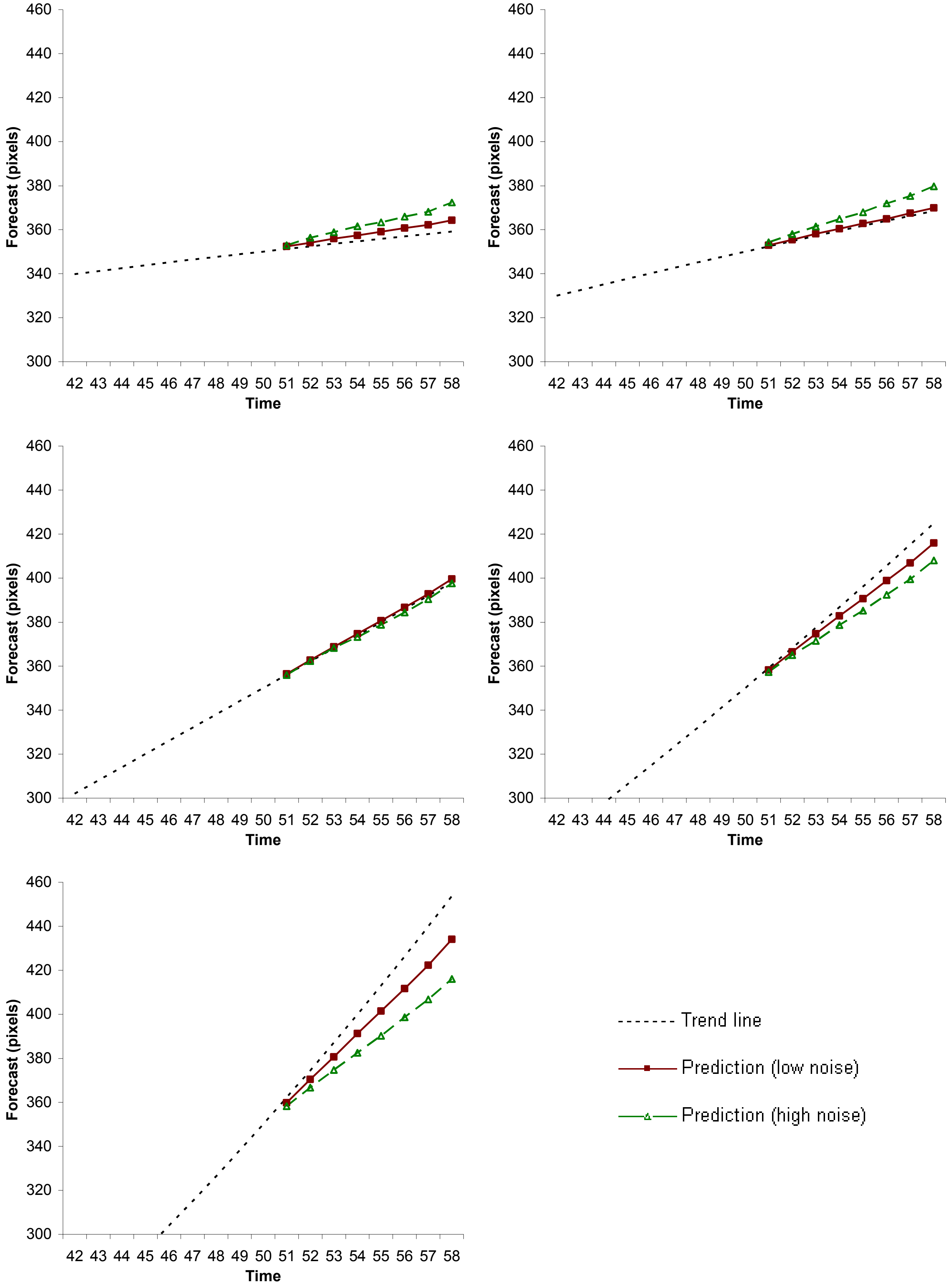
Forecast (pixels)
460
440
420
400
380
360
340
320
300
42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58
Time
Trend line
Prediction (low noise)
Prediction (high noise)

Figure 3

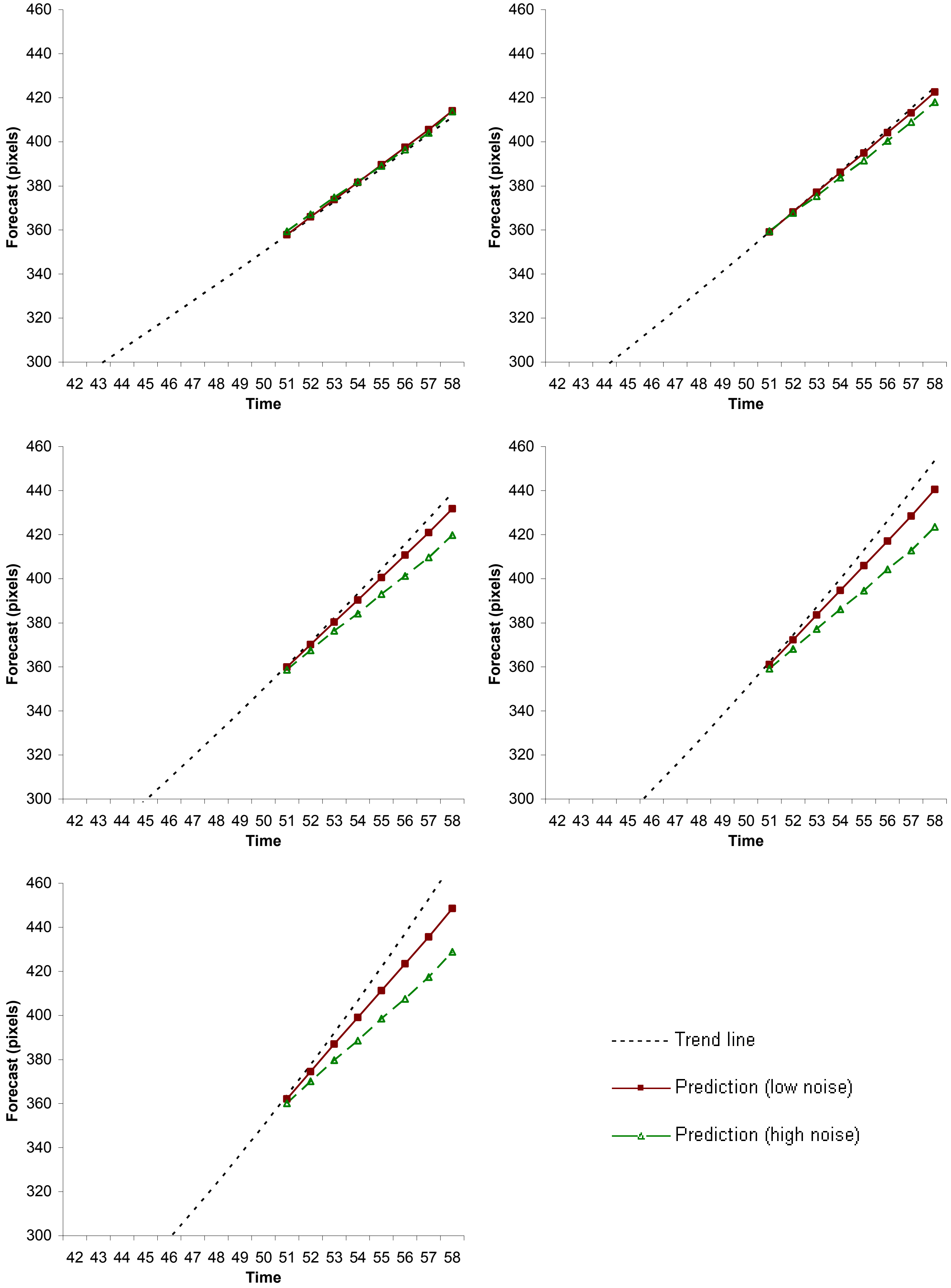
Forecast (pixels)
Time
Trend line
Prediction (low noise)
Prediction (high noise)

Figure 4

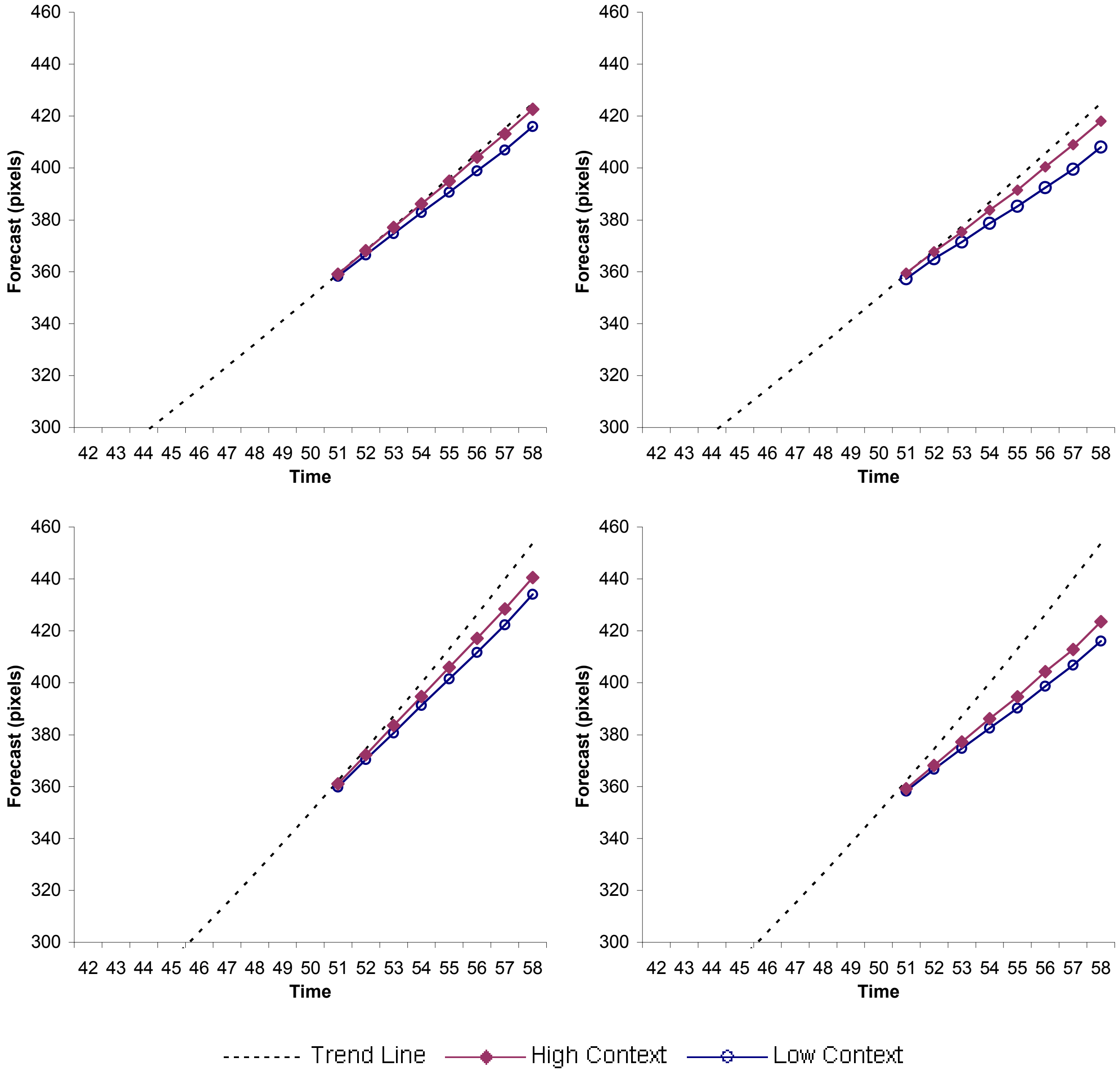
460
440
420
400
380
360
340
320
300
Forecast (pixels)
42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58
Time
Trend Line
High Context
Low Context

Figure 5

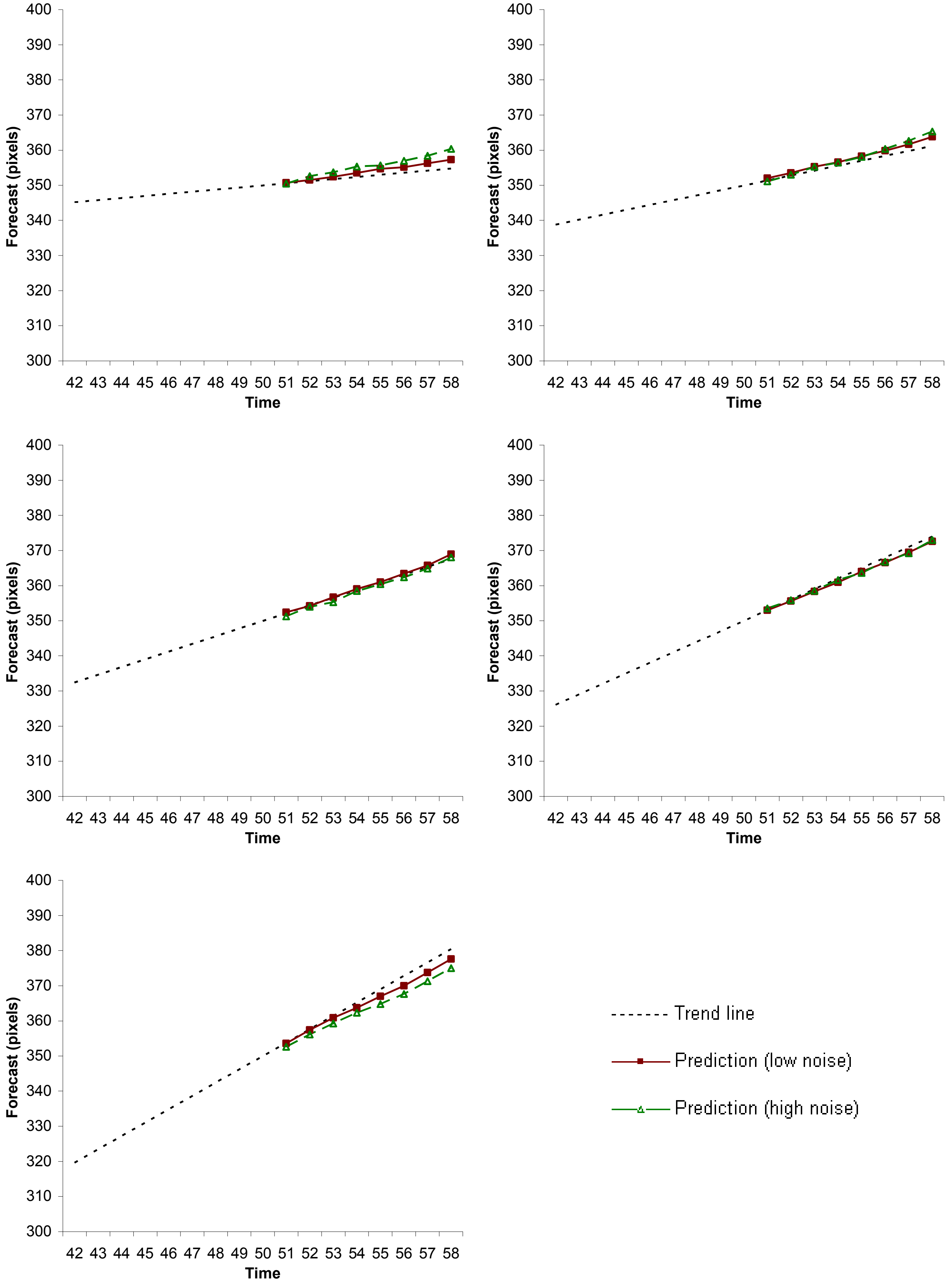
Forecast (pixels)
Time
300
310
320
330
340
350
360
370
380
390
400
42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58
Trend line
Prediction (low noise)
Prediction (high noise)

Figure 6

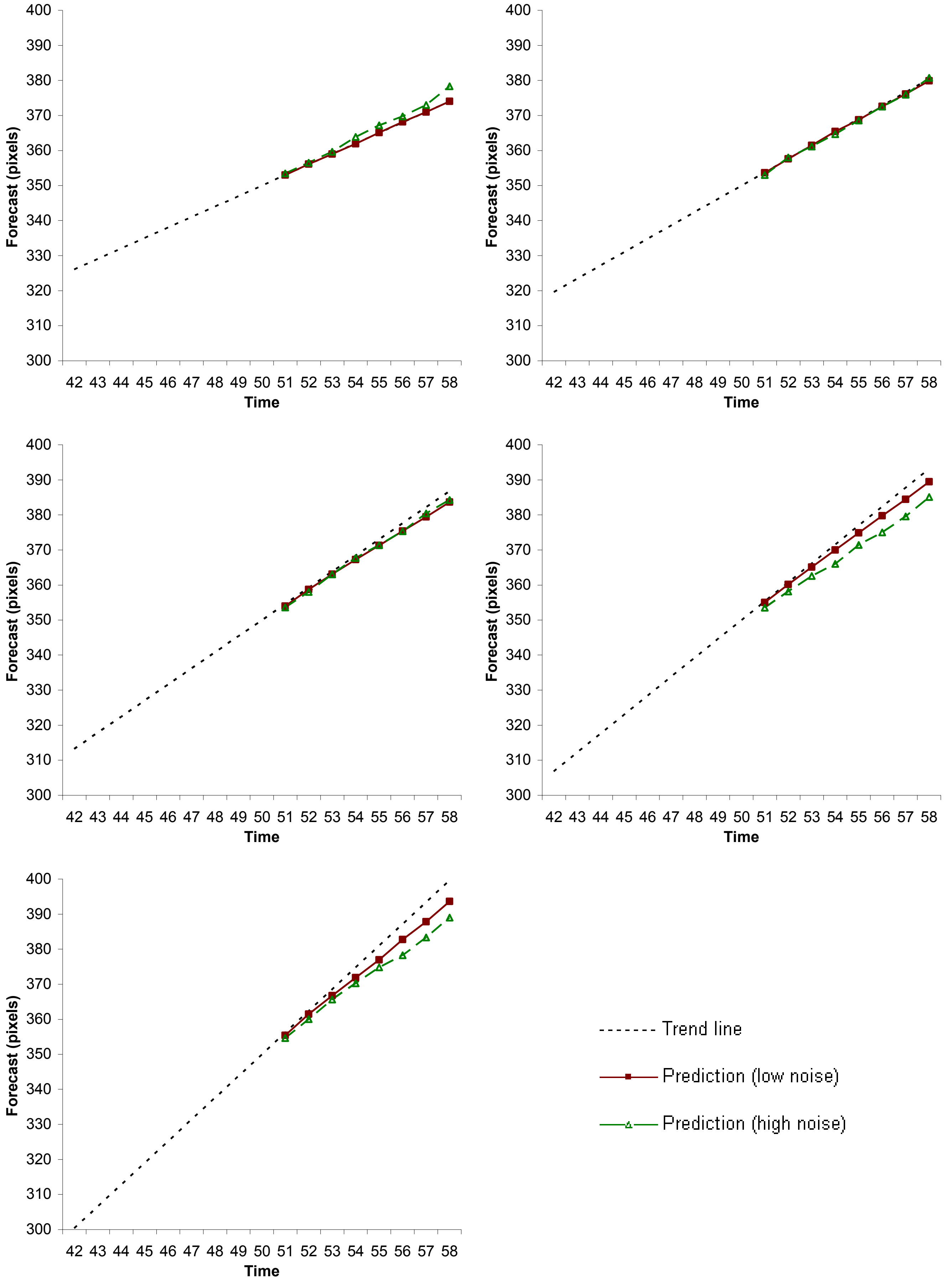
Forecast (pixels)
Time
400
390
380
370
360
350
340
330
320
310
300
42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58
Trend line
Prediction (low noise)
Prediction (high noise)

Figure 7

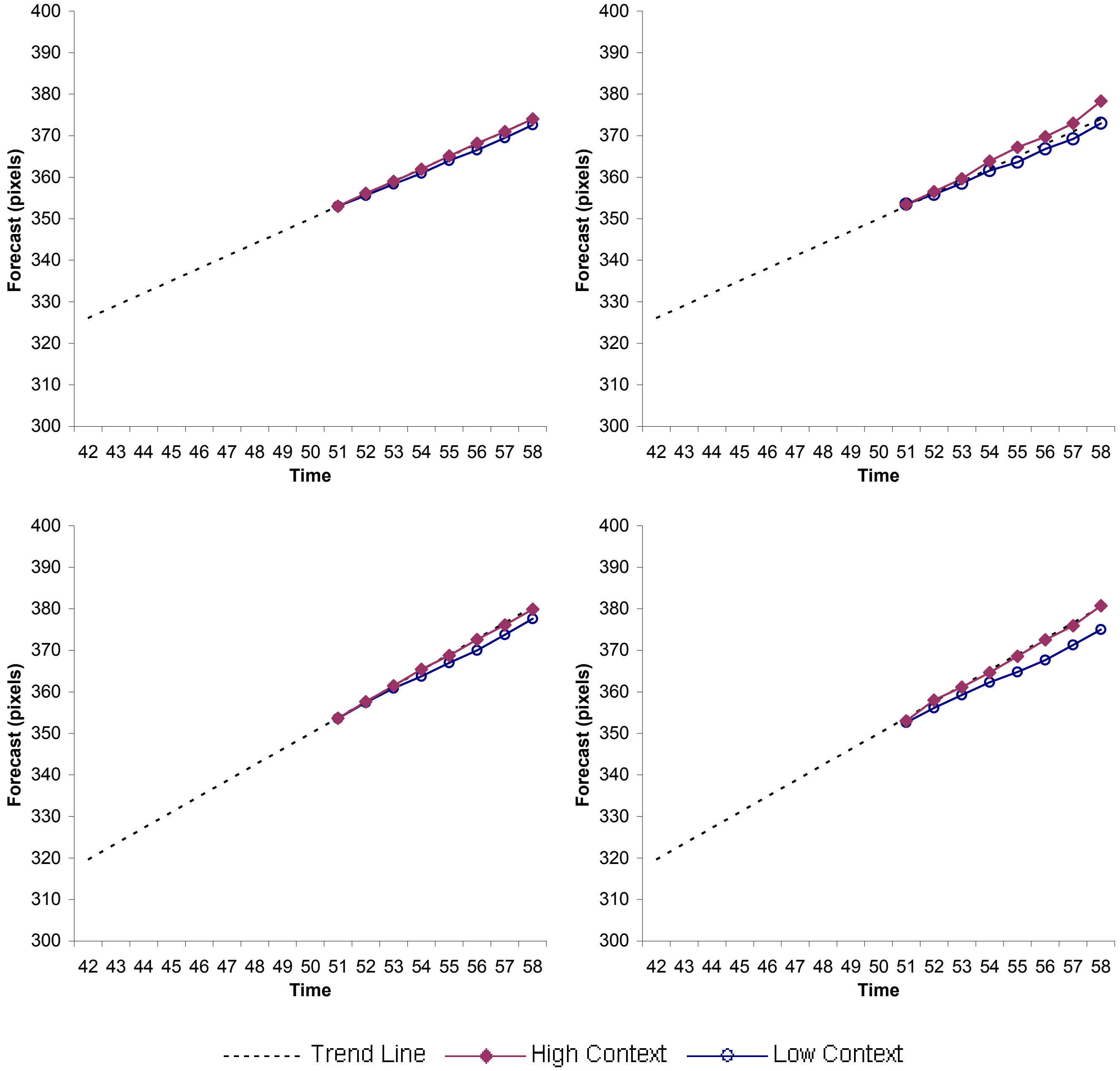

Figure 8

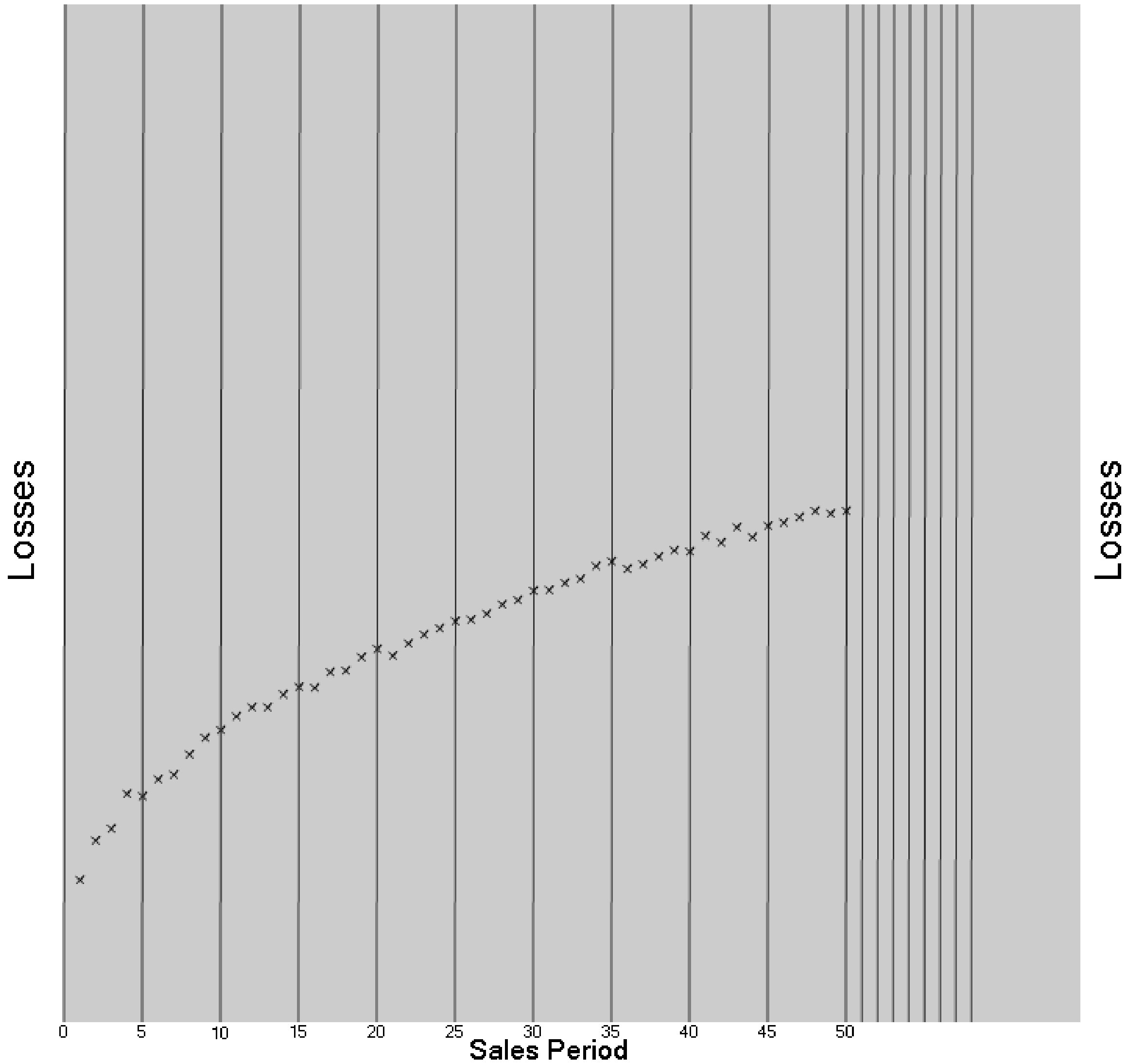

Figure 9

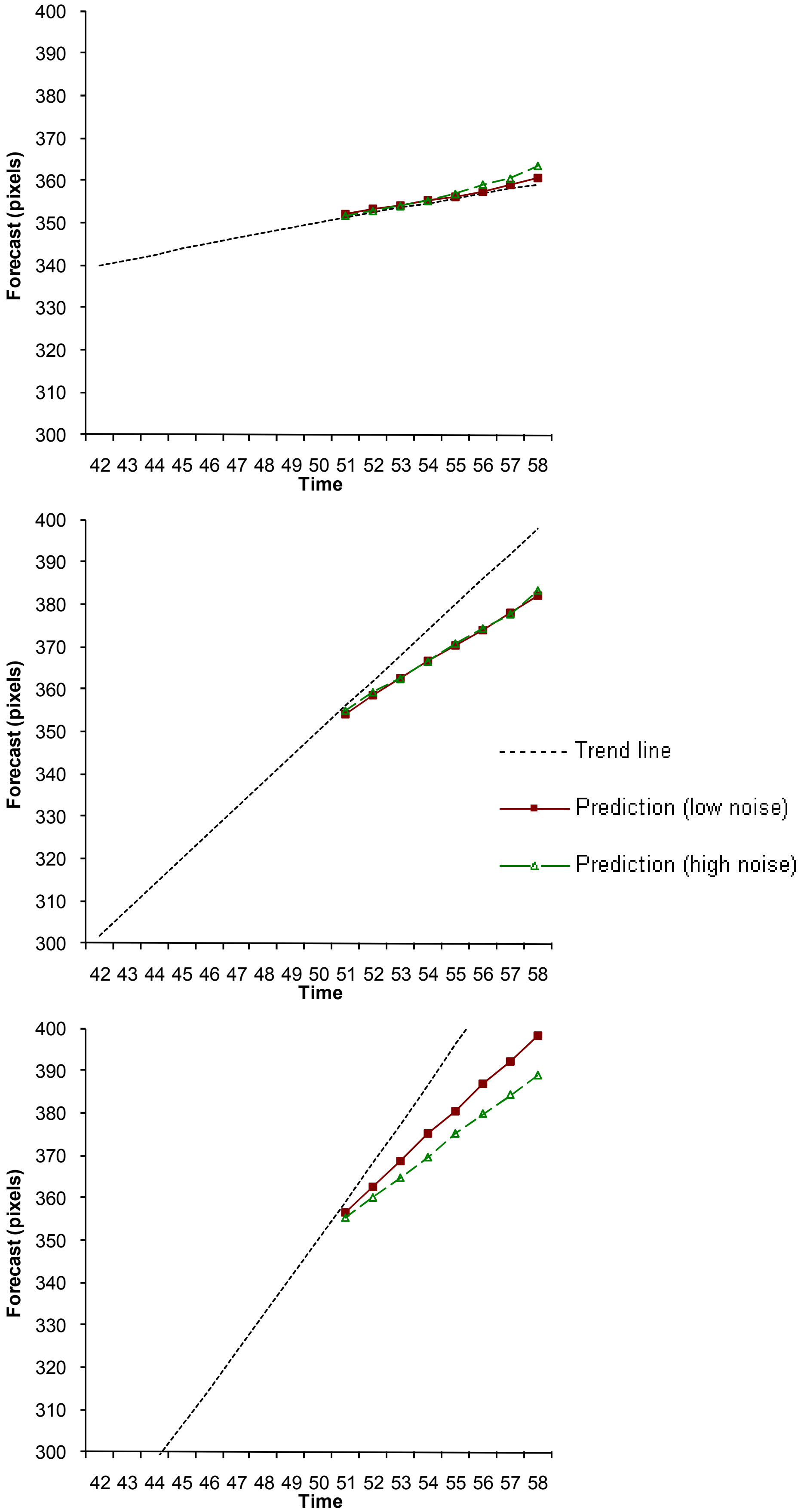
400
390
380
370
360
350
340
330
320
310
300
Forecast (pixels)
42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58
Time
Trend line
Prediction (low noise)
Prediction (high noise)

Figure 10