Nonlinear Processes in Time-Ordered Observations: Self-Organized Criticality in Daily High School Attendance

MATTHIJS KOOPMANS
Mercy College (USA)

In the United States, high school attendance and drop-out are important policy concerns receiving extensive coverage in the research literature. Traditionally, the focus in this work is on the summary of dropout rates and mean attendance rates in specific schools, regions or socio-economic groups. However, the question how stable those attendance rates are over time has received scant attention. Since instability in attendance may affect how long individual students stay in school, the issue deserves attention. The school districts that have begun to keep record of daily attendance rates in their schools over multi-year periods, such as those in New York City, have created an opportunity to investigate the temporal dimension of daily attendance, and thereby explore its stability. This paper will focus on its long-term characteristics, specifically the following: self-similarity, meta-stability or pink noise, and the impact of sudden departures from the central tendency of the series. Such departures can be used to estimate the impact of exogenous influences on the behavior of the system. The findings illustrate the importance of describing the dynamical patterns underlying attendance that remain concealed in traditional summary measures.

Few educators would dispute that attending school is an important contributor to academic success. The question how indicators such as these are in turn predicated on other variables is subject to more debate. Among the likely factors affecting attendance and achievement are socio-economic status, support from parents and/or caretakers, quality of instruction and effective school building leadership. In our analyses of educational effectiveness, daily

attendance rates are not often analyzed in their own right, and if attendance information is reported at all, it is usually done in terms of weekly, monthly or yearly averages. While these measures have their value, they conceal potentially valuable information about how daily attendance fluctuates over time, and what these fluctuations tell us about attendance behavior. The description and interpretation of such fluctuations are the focus of this paper.

The inadequacy of aggregated summary measures to capture the dynamical aspects of daily attendance rates has been noted before (Koopmans, 2011; 2015). Several urban school districts including those in New York City have begun to build data repositories of daily attendance rates for all of their schools, thus creating an opportunity to investigate how attendance behaves across the time spectrum and what type of patterns we might expect to encounter. We can test specific models of change and instability that have been forwarded in the complex dynamical systems literature and examine their applicability to the field of education, as well as contributing to that literature by providing prototypical cases of dynamical processes.

In addition to its theoretical relevance, a description of the time-related dependencies in daily attendance rates also serves a practical purpose because they can help enhance practitioners' knowledge about cyclical patterns of attendance in their school buildings, and while school personnel might be well able to detect weekly or monthly cycles of fluctuation in daily attendance, it is much harder to observe the irregular cycles that capture complexity, and therefore, some kind of statistical modeling would be required to detect them. This paper focuses on three of those less readily observable dynamical patterns, namely self-similarity, self-organized criticality, and meta-stability or pink noise. I also propose an approach to the modeling of the exogenous influences on daily attendance patterns, which is generally known in the time series literature as pulse analysis, or interrupted time series analysis.

This paper is planned as follows. First, I will review some of the basics of traditional time series analysis, the statistical approach to the analysis of temporal variability. Subsequently, the three aforementioned dynamical scenarios will be illustrated using daily attendance data from two schools in New York City. I will also discuss the utility of pulse analyses in that context to model influences exogenous to the system of interest, and reflect on how these approaches are helping us obtain a better understanding of temporal processes in education, daily attendance rates in particular.

Time Series Analysis – The Basic Approach

The basic idea behind time series analysis is that if one has a long string of observations whose order reflects the passage of time, one can predict the value of future observations based on the shape of that trajectory. For example, if there is an upward trend from time 1 to time 2 and time 3, one might predict that at time 4, this upward trend will continue, barring any additional information about the behavior of this time series. Likewise, if a time series oscillates during a given period, one might expect this pattern to continue going forward, again assuming that no new information becomes available about the series. The more observations are used when making predictions of this kind, the greater the accuracy of the prediction, and the narrower the confidence interval will be around the predictions.

In situations where the behavior of individuals is measured repeatedly, it is not safe to assume that there is no relationship between the measures that are repeated, as conventional statistics would have it, because those measures belong to the same analytic unit (individual or other system). One of the primary concerns in a time series analysis is therefore to model these dependencies, and remove any bias they might produce in the estimates of time effects on a behavioral outcome such as attendance. These dependencies sometimes involve a correlation between one observation and an immediately neighboring one or other observations that are close by. Correlations of observations with neighboring observations are usually referred to as *autocorrelations*, and in time series analysis, estimating their impact on the series enhances the reliability of the model predictions. This estimation is one of the central aims of a time analysis.

Sometimes modeling the dependency between observations is simply done to make the statistical corrections to enhance the accuracy of predictions. There may also be instances where autocorrelative patterns in time series that are of substantive interest. For instance, it could be that the data show cyclical patterns, such that observations on a given day of the week correlated with those made on the same day in other weeks. Likewise, there could be hourly or monthly patterns in the data. Time series analysis can model those cycles, and incorporate them into the prediction of future occurrences. These modeling features have always been part and parcel of the time series approach (e.g., Box, Jenkins, & Reinsel, 2008). In its contemporary form, time series analysis is also capable of modeling more complex processes such as unpredictability, self-similarities and self-organized criticality, processes that typically reveal themselves over the longer range of the series (Beran, 1994). The ability to capture these unrulier parts of the system's behavior is important because they address the concern that in conventional approaches to research, systems of interest are assumed to be stable, an assumption that been subjected to challenge quite some time ago (Goldstein, 1988).

Another scenario that requires due attention is the possibility that the series is impacted suddenly due to shocks from sources that are external to the system. Such shocks are typically not predictable based on the previous behavior of the time series, but their impact on this series can nonetheless be estimated. Modeling such shocks is typically done using an approach called interrupted time series analysis, intervention analysis or pulse analysis. An example of such a shock to a time series consisting of daily attendance rates would be inclement weather, a situation that would temporarily disrupt the attendance pattern. Pulse analysis models such disruptions, and it also estimates to what extent the time series recovers its original properties, and if so, how quickly.

The ongoing pattern in the series, and the sudden perturbations thereof, characterize different aspects of the behavior of a system. As far as daily attendance is concerned, both are important, as the examples provided below will be meant to illustrate. The former can be used to characterize the baseline settings of the system (i.e., the *endogenous process*), which describes how the system behaves in the course of its usual adaptive cycle, while the latter models a sudden impact of external events on the system (the *exogenous process*), and the resilience of the system toward such events.

White Noise and Pink Noise – An Illustration

When describing or predicting the future behavior of a system, the first question that time series analysis must answer is of course is whether there is indeed a pattern that may replicate. Figure 1 illustrates two different aspects of this issue in a simulated example. The panel on the left illustrates random fluctuation, also known as white noise. In this situation, knowing the past behavior of the series will not help predict its future behavior, except to say that observations are expected to continue to fluctuation randomly. The panel in the top right of the figure shows pink noise. The pattern shown in these two panels is similar, except that there is considerable fluctuation around the central tendency of the series in the case of pink noise, except for the fact that this central tendency shows a nonlinear pattern over time, giving the trajectory this undulating appearance, as well as showing irregularity in the clustering in the observations.
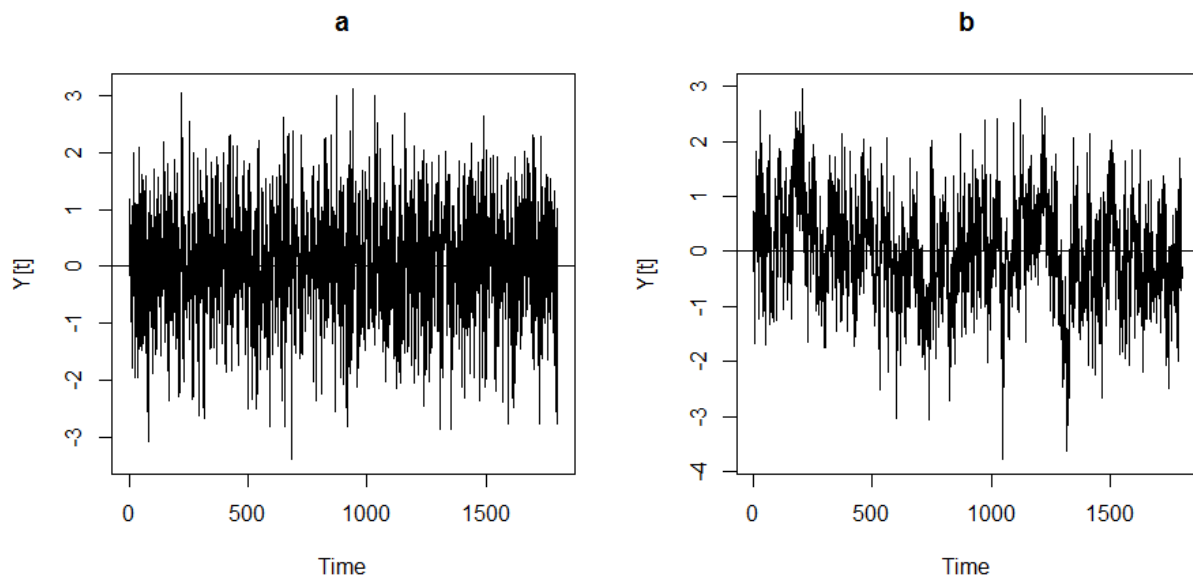

Figure 1. A simulation with a string of $N = 1,800$ data points Y_t, illustrating two scenarios. **(a)** Random variability (white noise) and **(b)** Long range dependency (pink noise). Both distributions have a mean of $M = 0$.

Preliminary Data Description

The analyses presented here concern on two New York City schools whose daily attendance rates were analyzed as part of a larger study that distinguishes small high schools (total enrollment of 500 students or less) to large ones whose enrollment is in the thousands of students. Both schools described in this paper are small ones. Periods in which schools were not in session (e.g., summer vacation, winter and spring breaks) were removed and median values were imputed when observations were missing. The series were then connected such that one long string of observations describes the attendance rates in a given school over a multi-year period, and that multi-year pattern was subjected to analysis. In the present study, ten years' worth of attendance information is used. The mean attendance rate in School 1 is 76.4%, and 91.4% in School 2. Both schools serve a predominantly Hispanic student population (52.0% in

School 1 and 78.7% in School 2), with a significant representation of African American students as well (37.3% in School 1 and 11.4% in School 2).
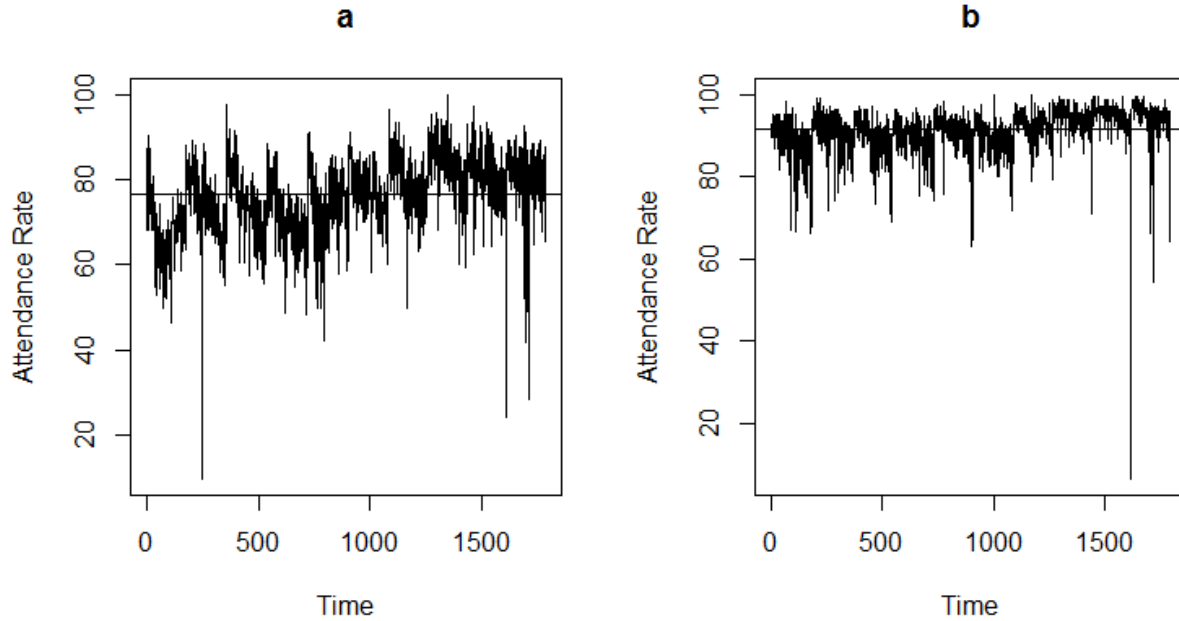

Figure 2. Variability in Attendance Rates at Two schools over a 10-year Period (Sept. 2004 – June 2014). **(a)** School 1 ($N = 1,785$, $M = 76.4$); **(b)** School 2 ($N = 1,795$, $M = 91.4$)

Daily Attendance as a Time Dependent Dynamical Process

Figure 2 shows how the dependencies described above actually play out in two schools in New York City (referred to here as School 1 and school 2). There appears to be a clustering of observations as well as rising and falling fluctuation patterns over the broader range of the time scale that is characteristic of pink noise or long range memory processes. The statistical modeling of these attendance patterns confirms this possibility. In a statistical comparison of models describing the data shown in Figure 2, those models that include a parameter for long range memory processes fare better than those that rely exclusively on short range processes. Koopmans (2017) describes the statistical modeling process in greater detail. While the presence of pink noise points to the possibility of self-organized criticality or self-similarity, researchers are generally careful to point at that statistically establishing pink noise does not confirm these patterns, but merely allows us to conclude that we cannot rule them out (Beran, 1994; Stadnitski, 2012).

While the comparison of the statistical models may not conclusively establish the presence of self-similarity, closer inspection of the series does suggest that it might indeed be the case. The plot in Figure 3 illustrates this point for School 2. The four panels in Figure 3 give an impression of a self-similar process in School 2, i.e., a pattern where within a time series pattern, the very same pattern replicates itself on a smaller time scale and then again at a yet smaller

time scale. This pattern is also sometimes referred to as scale invariance (Mandelbrot, 1997) in recognition of the fact that no matter how the scale is time set for the observations to be included in the plot, one would encounter this pattern. In other words, the cycle at which these patterns repeat themselves it not fixed, as it would be, for example, in weekly or monthly cycles. Instead, one expects to find the same process, repeated at time scales that are different. Figure 3 shows such self-similar patterns in School 2, where initially small variability in the observations increases, followed by a sudden drop in the attendance. In this figure, the patterns seen for the one school year (2004-05) are replicated on progressively smaller scales in Figures 3b, c and d. With respect to daily attendance, Koopmans (2015; 2016) suggests that these repeating patterns may reflect increasing tension associated with the compulsory nature of attendance followed by the occasional need for release, as shown by the significantly lower attendance rates on
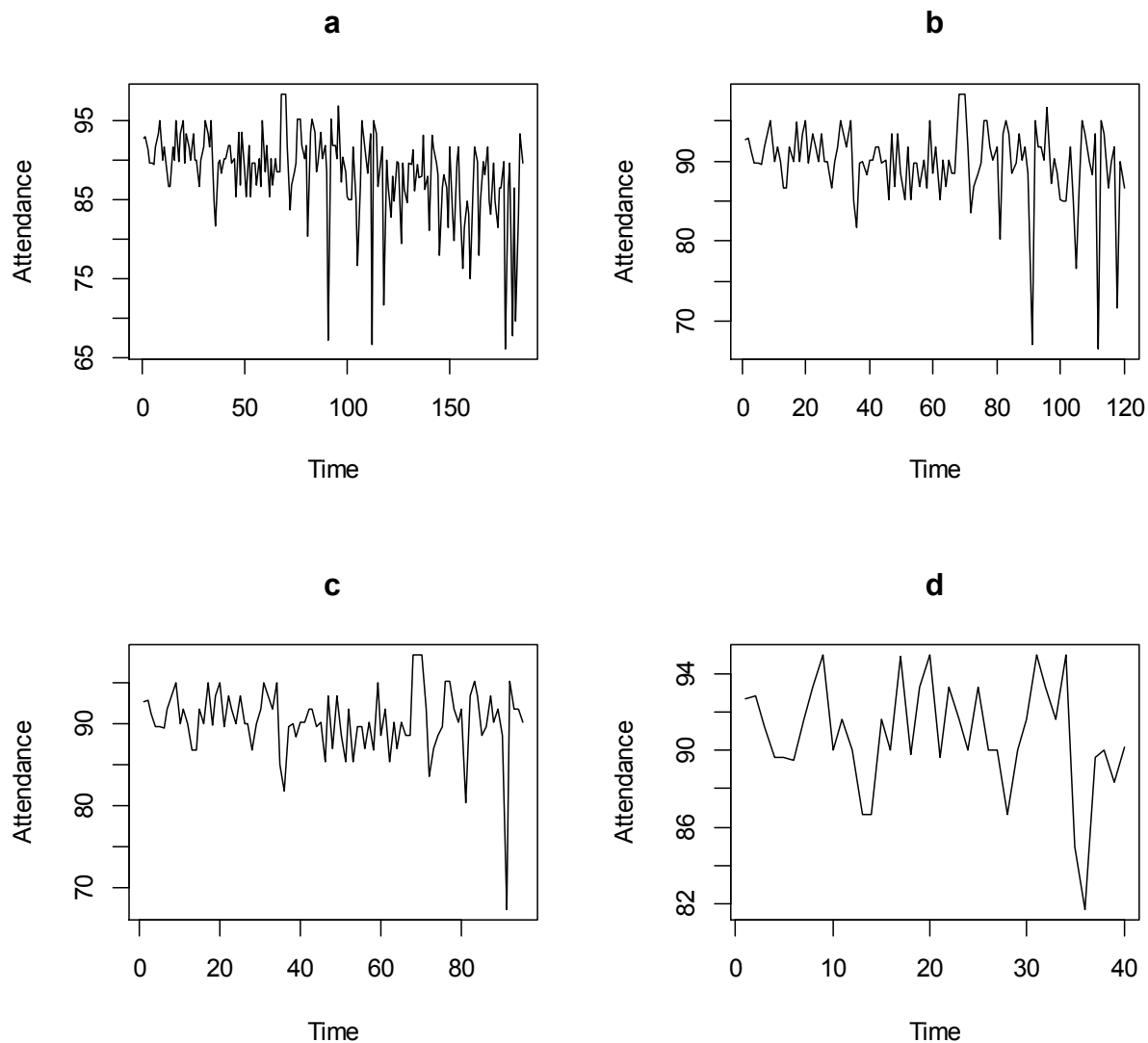

Figure 3. Self-similar patterns and scale invariance in daily attendance rates over time in School 2 for the trajectories starting on September 13, 2004. Attendance rates through **(a)** June 17, 2005 ($N = 185$); **(b)** March 11, 2005 ($N = 120$); **(c)** January 28, 2005 ($N = 95$); **(d)** November 5, 2004 ($N = 40$)

particular days. Such tension-release patterns are consistent with how self-organized criticality has been interpreted elsewhere in the dynamical literature (e.g., Beran, 1994).

The Impact of Extreme Values

One of the attractive features of time series analysis as a statistical approach is its capability of incorporating outliers into the analysis and to model their impact on the statistical properties of the series. The analysis presented here illustrates three of the most well- known scenarios. One is the occurrence of extreme values that do not impact the series beyond their own time point. These are called *additive outliers*. The second scenario is the occurrence of extreme values, after which the series gradually returns to its range of original values. In other words, after the pulse, there is a recovery period. Such observations are called *transient change* outliers. The third scenario, called *level shift* marks a change of a more permanent nature, which is to say that there is no recovery from the departure (Peña, 2001). These three scenarios are illustrated in Figures 4 and 5.

Figure 4 illustrates the occurrence of level shifts in the trajectories for School 1. Figure 4a covers the 2004-05 school year. The table shows a downward level shift toward the beginning of the series (t = 19, October 7), as well as a transient outlier at t=58 (December 1) from which there is a slow but decisive recovery. Figure 4b shows the daily attendance rates for the spring of 2009, with a quite sizable upward level shift at t = 816 (February 2).
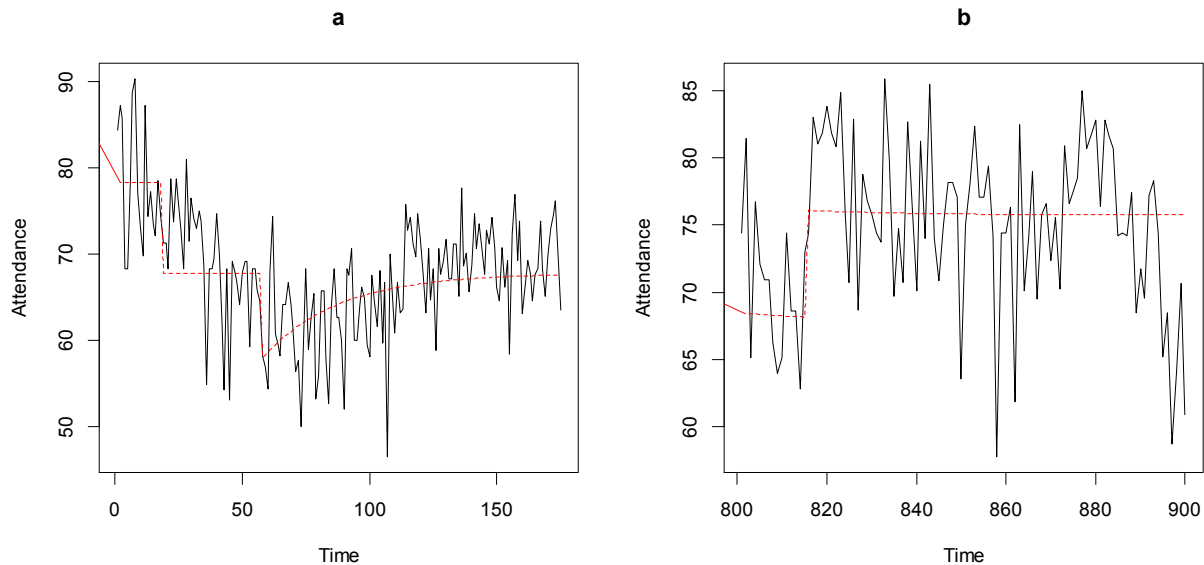

Figure 4. Modeling sudden departures in time series trajectories (School 1). **(a)** Daily attendance rates from September 13, 2004 through June 10, 2005 with a Level Shift at t = 19 (October 7) and a Transient Change outlier at t = 58 (December 1); **(b)** January 5, 2009 through June 12, 2009 with a Level Shift at t = 816 (February 2)

Figure 5 shows a level shift and some additive outliers in School 2. Figure 5a shows the trajectory in this school from January 2 through June 13, 2008. At t = 666 (March 10, 2008), the model observed a sudden drop in the series followed by a relatively quick recovery toward its original central tendency. A similar drop can be seen at t = 721 (June 2, 2008). The substantive question is what specific event (policy or otherwise) triggered these sudden drop, and whether this recovery, in conjunction with recovery rates elsewhere in the system offer a description of the resilience of this system to perturbation.

Figure 5b shows the attendance rate trajectory in School 2 from September 9, 2013 through June 13, 2014. As can be seen in the figure, four additive outliers were observed during this period in the series, marking sudden changes from which the series soon recovers. Comparison of the two periods depicted in these two figures in relation to these two types of outliers is informative as well. In addition to the swift recovery from extreme observations, the series depicted in the panel toward the right looks more stable in one important other regard, which is that variability between observations is much lower than it is on the series shown on the left, in which high degrees of variability are observed and clearly are not captured by the predictions made by the pulse models.
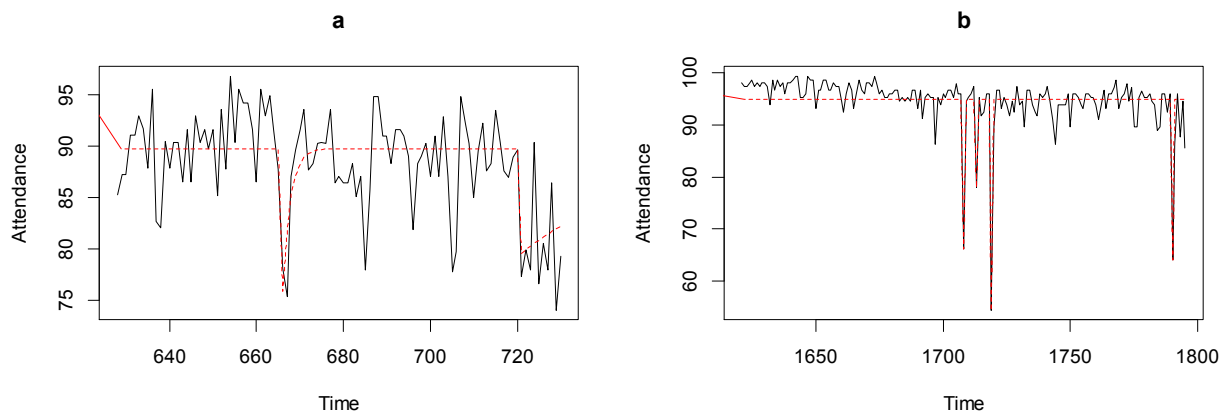

Figure 5. Modeling sudden departures in time series trajectories (School 2). **(a)** January 2 through June 13, 2008 with Transient Outliers at t=666 (March 10, 2008) and t=721 (June 2, 2008); **(b)** September 9, 2013 through June 13, 2014 with Additive Outliers at t=1708 (January 22, 2014), 1713 (February 5, 2014), 1719 (February 13, 2014) and 1790 (June 6, 2014).

Discussion

The analyses presented above indicate that critical information is lost if daily high school attendance is reported in terms of weekly, monthly or yearly averages instead of a time dependent phenomenon. The trajectories presented here show turbulence, patterns that repeat in unpredictable ways, and unexpected departures of individual observations from the central tendency of the series. While teachers and school building administrators may be familiar enough with the predictable patterns such as weekly cycles and similarities between attendance rates on subsequent days, the clustered variability patterns that constitute the pink noise in the series may not be familiar, nor easy to predict. Therefore, the statistical confirmation of such patterns adds to our knowledge and understanding of how attendance behaves over time.

The modeling of sudden departures is an analytical approach that is particularly suitable to address the impact of external factors (e.g., inclement weather) on attendance behavior in a school building. More importantly, such modeling efforts can be used to address the impact of specific policy events on the attendance trajectory. Such policy initiatives may or may not be directly related to attendance, but the most important thing is that their impact on attendance rates can be directly measured. There are two aspects of this impact. One is the departure of individual observation from the central tendency of the series, and the second is the degree and extent to which the original statistical properties of the trajectory are fully recovered afterward. Examples of this recovery process are shown in Figure 4a for School 1, and in Figure 5a for School 2. The estimation of this recovery has policy implications because it tells us how quickly schools go 'back to normal' after sudden perturbations. Also important, is the possibility that lasting change is produced in a time series (for better or worse). In such instances, the departure of an individual observation from the series marks the beginning of the shift of the series to a new level, and in policy research, the question is pertinent what the size is of the change, as is the question whether these shifts constitute a change predicted by policy events. With respect to sudden departures, it should be remembered that they do not necessarily imply a response to an external stimulus to the system, but may also be due to the internal dynamics. An exploration of such departures across the full time spectrum leaves that question in the middle, while analyses that pinpoint specific policy events on the timeline to study their hypothesized consequences to the system specifically seek to explain the systems' behavior in terms of external pressures.

When interpreting these results, it is important to remember that they concern school level information, without discriminating individual students or grade levels, and thus, it is a very crude measure. We also need approaches such as those described here, approaches that were originally developed to handle linear processes, and extend them to the modeling of nonlinear ones. The examples presented here illustrate some of the nonlinearity that can occur in strings of repeated observations, as well as suggesting underlying dynamical processes. Particularly the tension-release patterns that replicate unpredictably across the time spectrum are of importance here because they provide an important qualifier to what we know about the daily attendance of our high school students and the processes that might impact it. Triangulation of findings such as these with the analyses of external factors and events is important, because it would enable us to better understand how we can have a favorable impact on attendance behavior and, by extension, their life and career trajectories.

References

Beran, J. (1994). Statistics for long-memory processes. New York: Chapman-Hall.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). Time series analysis: Forecasting and control. 4th Edition. Hoboken, NJ: Wiley.

Goldstein, J. (1988). A far-from-equilibrium systems approach to resistance to change. Organizational Dynamics, 17(1), 16-26.

Koopmans, M. (2011, April). Time series in education: The analysis of daily attendance in two high schools. Presented at the annual meeting of the American Educational Research Association. New Orleans, LA. http://files.eric.ed.gov/fulltext/ED546476.pdf

Koopmans, M. (2015). A dynamical view of high school attendance: An assessment of short-term and long-term dependencies in five urban schools. Nonlinear Dynamics, Psychology and Life Sciences, 19(1), 65-80.

Koopmans, M. (2016). Investigating the long memory process in daily high school attendance data. In M. Koopmans, & D. Stamovlasis (Eds.) Complex dynamical systems in education: Concepts, methods and applications (pp. 299-321). New York: Springer.

Koopmans, M. (2017, April). Estimating perturbation and meta-stability in the daily attendance rates of six small high schools. Paper presented at the annual meeting of the American Educational Research Association, San Antonio, TX.

Mandelbrot, B. B. (1997). Fractals and scaling in finance: Discontinuity, concentration, risk. New York: Springer.

Peña, D. (2001). Outliers, influential observations, and missing data. In D. Peña, G. C. Tiao, & R. S. Tsay (Eds.) A course in time series analysis (pp. 136-170). New York: Wiley & Sons, Inc.

Stadnitski, T. (2012). Some critical aspects of fractality research. Nonlinear Dynamics, Psychology, and Life Sciences, 16(2), 137-158.

About the Author

Matthijs Koopmans is an Associate Professor at Mercy College. His current research interests include the application of complex dynamical systems approaches to education, cause and effect relationships, and nonlinear time series. Correspondence: mkoopmans@mercy.edu