The Series With Elastic Extents Problem (SWEEP) and "Gerrymandering" Urban Time Series

Samuel Stehle*

National Centre for Geocomputation, Maynooth University

March 5, 2019

Summary
The SWEEP influences temporal aggregation via scale, segmentation, and boundary effects. This paper uses the analogy of gerrymandering, which is the purposeful segmentation of space such that the underlying aggregations prove a specific point, to demonstrate such effects on time series. This paper evaluates sound monitoring and communication in Dublin, Ireland with respect to local regulations. Regularities in urban human social patterns and other time series properties generate predictable, and gerrymanderable, effects on the measurement of ambient sound with respect to established standards of aural health.

KEYWORDS: Modifiable Aerial Unit Problem, spatio-temporal analysis, time series, sound monitoring

1. Introduction

Analogous to the Modifiable Aerial Unit Problem is the Series With Elastic Extents Problem, or SWEEP. SWEEP is also known as the Modifiable Temporal Unit Problem (MTUP) (Çöltekin, et al., 2011), where aggregations of fine temporal scale observations summarize all observations within a bounded temporal interval. This causes a *scale effect*, where the local observations are summarised in some mathematical way into a smaller number of groups, each with a longer duration than local observations. The choices of when to draw the boundaries between units introduces *zoning effects*, each of which have the potential to alter the agreement between local and aggregate measurement.

The ability to change the statistics representative of an area by defining aggregation properties can be a purposeful activity with vague ethics. Gerrymandering is the manipulation of spatial boundaries to create administrative areas with more or less homogeneity among the individual members of those areas. It is possible to gerrymander a spatial aggregation with the knowledge of the local composition of the area – such as the voting tendencies of individuals in political administrative areas – or of other predictable patterns which lend themselves to alternative aggregate arrangements.

2. Series With Elastic Extents Problem

Çöltekin et al (Çöltekin, et al., 2011) describe temporal parallels to scalar and zoning problems in the spatial example. Temporal data contains some key differences which influence the ways that the SWEEP is conceived and measured. Common temporal patterns, such as periodicity, strongly effect the choices of scale of temporal aggregation (deJong & deBruin, 2012; Cheng & Adepeju, 2014). In

* sam.stehle@mu.ie

both the spatial and temporal implementations of aggregation effects, autocorrelated data impacts certain types of summary statistics. Arithmetic mean is unaffected by the autocorrelation of data comprising an aggregation, but the trend, for example, of a set of temporal points, is highly dependent on how the zones of aggregation are defined (deJong & deBruin, 2012).

3. Measuring Sound

Fifteen stations in Dublin collect decibel measurements every five minutes and make them available through a public API (http://dublincitynoise.sonitussystems.com/applications/api/api-doc.html). The data is used for public reference in real-time, hourly, and daily averages with a historical collection of multiple years. I capture two years of this record, between October 14, 2016 and October 14, 2018. Following guidelines and recommendations from the WHO, Dublin established thresholds for sound exposure in its Noise Action Plan. Dublin defines three value limits of sound for day and night: a ‘desirable low’ threshold, an ‘undesirable’ high threshold, and a lower threshold used to define ‘quiet areas’ – an special designation for places of urban solitude. The decibel value limits are in **Table 1.**

Table 1 Decibel value limits corresponding to levels of noise defined in Dublin’s Noise Action Plan

Threshold	Day-Time (07:00 – 19:00)	Night-time (23:00 – 07:00)
Quiet Area	< 55	< 45
Desirable Low	< 55	< 50
Undesirable High	> 70	> 55

4. Demonstrating the SWEEP

4.1 Scalar Effect

Timing influences the significance of sound measurement and communication. Peaks of relatively high noise cause temporary disruption to wellbeing and health, but sustained high decibel levels are damaging to quality of life. Thus, urban sound measurement is shared for public consumption at aggregated timeframes – particularly hourly and daily levels in Dublin – rather than the five-minute observations measured at sound sensor sites. I summarise the production of harmful sound with the proportion of time over the two years of the study period in which each of the 15 monitoring sites experienced sound levels greater than Dublin’s Undesirable High value limit.

The proportion of all five-minute observations over the two years which surpassed the Undesirable High noise limit level is compared to the proportion of all hourly means and daily means. With no scalar SWEEP effects, there would be no variation between these aggregations. **Figure 1**, showing the percent of total measurements at each temporal scale which are over the nighttime Undesirable High, shows the vast difference in the amount of time recording unhealthy sound, dependent on the level of aggregation. Moreover, that difference is predictable, as stations with undesirably high sound measurements at the 5-minute level produce even greater frequency of undesirably high means at greater aggregations. The observed five-minute proportions over the Undesirable High are highly correlated with the change in proportions at both the hourly and daily levels via Pearson’s r (significance level .01).

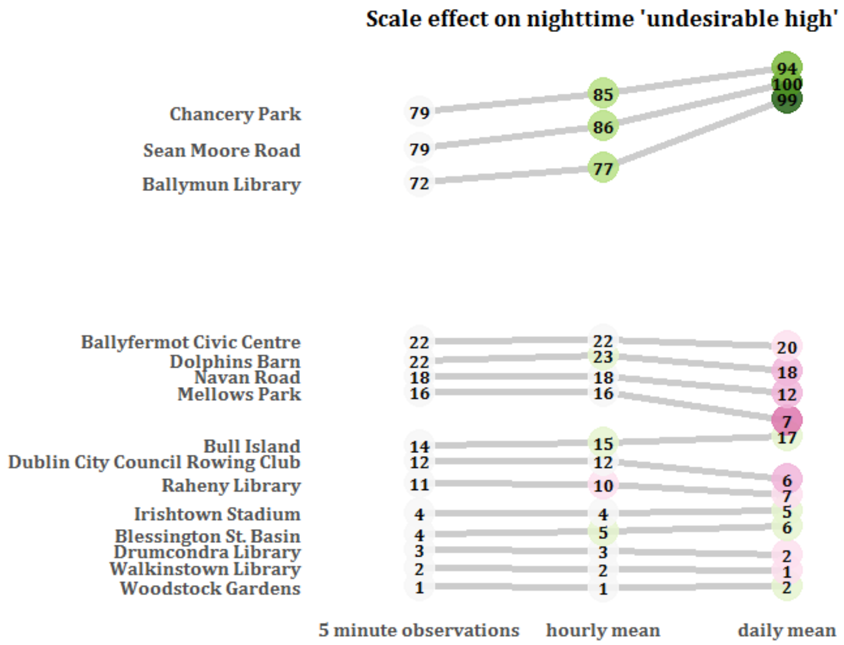

Figure 1 Proportion of time in which measurements at each monitoring station surpassed Undesirable High value limit, across three temporal aggregations.

4.2 Segmentation Effect

Segmentation effects stem from the choice of where to define the boundaries between aggregations. For example, the hourly aggregations described above consist of the 12 five-minute observations beginning on the transition to a new hour, or from 12:00 through 12:55. Temporal data is dividable into segments which carry environmental and logistical meaning via clock time. However, segmenting time in the same ways may reproduce specific inherencies in temporal data, and sometimes such inherencies do not follow natural rhythms. I quantify the effect of alternative segmentations of hourly aggregations and show that temporal properties of the data cause such visible patterns.

Testing the effect of hourly segmentation is akin to using a temporal moving average with an order of 12 (five-minute observations per hour) and stringing every 12th value from the result into a separate new time series. Thus, each new time series, or segmentation, is a decibel mean of the 12 closest observations in time. **Figure 2** shows the process diagrammatically.

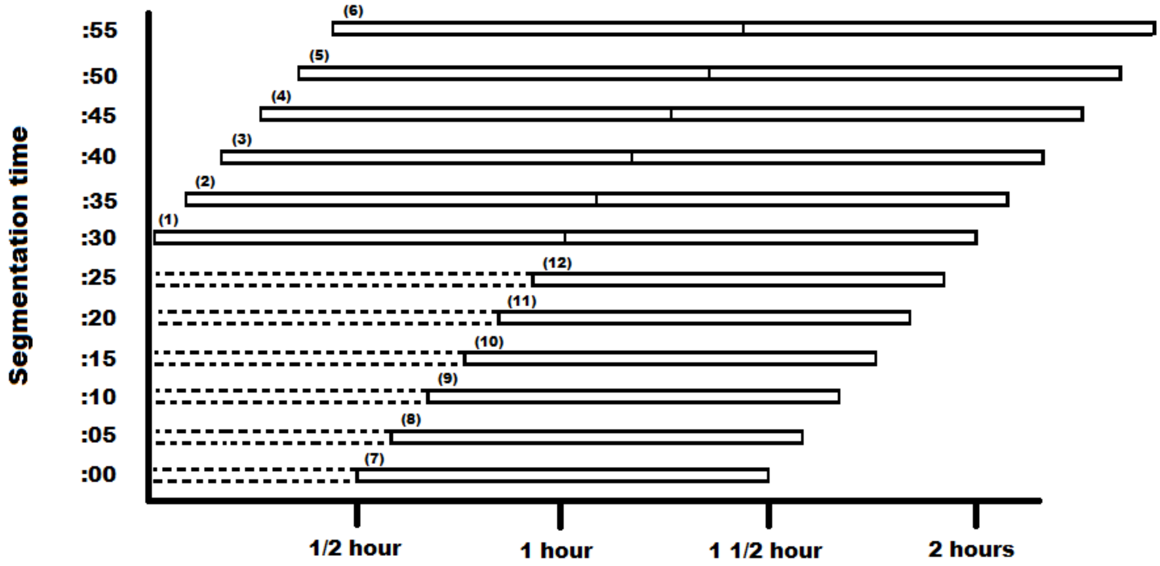

Figure 2 Hourly aggregations in 5-minute segmentations. Moving average of order 12 corresponds to the numbered segments (1) through (12).

I performed the same calculation of proportion of each hourly mean decibel measurement within each 5-minute segmentation which was greater than the value limit for desirable low and undesirable high levels in the day and nighttime. Consistently, each station produced a similar result; the proportion of total time spent above the given value limits increase as segmentations centre around times later in the hour. **Figure 3** shows two of the stations with the most pronounced version of this pattern.

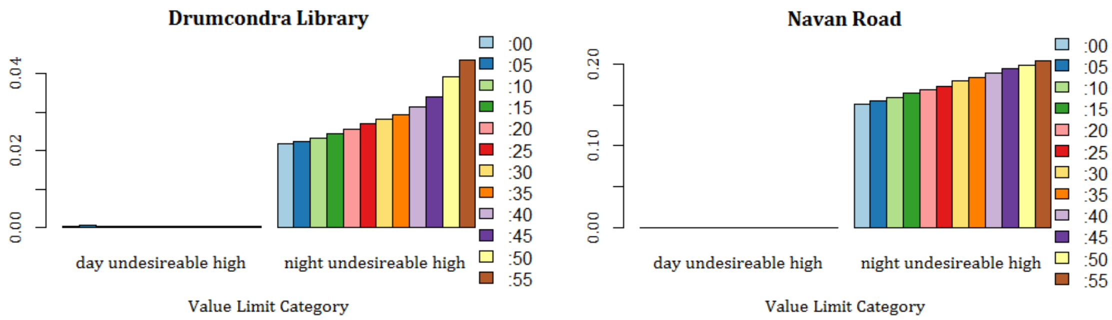

Figure 3 Proportion of aggregated mean decibel level greater than the given value limit.

Figure 2 shows that the only difference between two adjacent segmentations (for example, a segmentation with mean centred around :35 past the hour vs. :40 past the hour) is a later shift of a single five-minute observation. Most of the mean values within each segmentation consist of the same observations. The primary difference is simply that later segmentations consist of a greater number of observations from later in each day.

A general trend increasing at scales greater than the aggregation level, such as the day or the study period, explain the pattern visible in Figure 3. Across all stations, 89.1 percent of all days have a generally positive slope, according to Sen's Slope estimator (Sen, 1968). Thus, segmentations which consist of a greater number of measurements later in the day will have larger mean decibel readings, and are more likely to breach value limits.

5. Discussion

This paper seeks to show that, not only does the SWEEP exist, but has predictable impacts on aggregate statistics. The consequences of gerrymandering spatial data can be large, most notably election rigging. Knowledge of the composition of observations within spatial aggregations are necessary for "effective" gerrymandering, but I argue that less prior knowledge is necessary for doing the same with temporal data, and that care must be taken in using temporal groupings of official data, such as urban sound monitoring.

Time series properties correlate with demonstrable effects in the measurement of unhealthy noise. Aggregating information which tends to be higher in the observations' range yields even larger values, and aggregating lesser values generate even lesser statistics, as demonstrated in Section 4.1. As such, aggregations are also dependent on the long and short term trends of time series data. Section 4.2 showed that an increasing trend at a larger temporal scale than the aggregation level can influence aggregate statistics simply by making small modifications to the boundaries between aggregations.

The ability to define which values get aggregated means that there is control over what those statistics communicate. Specific decisions have been made on how aggregations would be defined and presented as functions of temporal scale and segmentations. Because those decisions are known to effect statistical outcomes in ways that could potentially alter the narrative on the instance of urban sound in predictable ways, the SWEEP is not an ignorable factor in temporal analysis.

6. Acknowledgements

The author would like to acknowledge the support of Science Foundation Ireland (Award No. 15/IA/3090) and the contribution from our data partners and providers.

7. References

Cheng, T. & Adepeju, M., 2014. Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PloS one,* Volume 9

Çöltekin, A. et al., 2011. Modifiable temporal unit problem. In *ISPRS/ICA workshop "Persistent Problems in geographic visualisation" ((ICC 2001)*. Paris, France (Vol 2)

deJong, R. & deBruin, S., 2012. Linear trends in seasonal vegetation time series and the modifiable temporal unit problem. *Biogeosciences,* 1, Volume 9, pp. 71-77.

Biography

Sam Stehle is a postdoctoral researcher on the Building City Dashboards project at Maynooth University. His work involves analysing, visualising, and deriving official statistics from real-time urban data. He is interested in machine learning, visual analytics, and applying these techniques to spatio-temporal and urban issues such as transportation and the Internet of Things.