

GAP method: a gap-based technique for outlier detection and its application to wildlife GPS data

Hirofumi Hirakawa (✉ hiroh215@gmail.com)

Independent

Daisuke Muramatsu

Wildlife Research Center of Kyoto University

Marcelo Gordo

Laboratório Biologia da Conservação DB/ICB, Universidade Federal do Amazonas

Akiko Takii

Shinshu University

Shigeyuki Izumiyama

Shinshu University

Article

Keywords:

Posted Date: April 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1567713/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Outliers in datasets are a concern for analysts as noise or signals. To detect them, we must quantify how much each data point lies away from others and then set a threshold to separate outliers from the others. However, there is scant research on the latter. We developed a method to address this issue. The method uses value gaps between adjacent data pairs in a univariate dataset, where data are in ascending order of values. Its core process is to find the largest gap in the upper range of the dataset and remove the data above the gap as outliers, and its supplementary process is to repeat the core process with the same dataset after removal. Analysts must decide when to stop this repetition; however, this technique lessens subjectivity by leaving only a few logical options. We can apply this method to any dataset, such as a time series or multivariate dataset, by quantifying data to expose anomalies as outliers in a univariate dataset. We demonstrate how to implement the technique using wildlife GPS data. We expect that this method has a much broader application because, to our knowledge, there are no other techniques to play the role.

Introduction

Outliers in datasets are a concern for analysts; they may be noise that obstructs identifying patterns in data or signals for unusual but meaningful events. In either case, the outliers deserve attention. Many studies have been on outlier detection^{1,2,3,4,5,6,7,8}. However, most of these focused on quantifying how much each data point lies away from others and implicitly left the rest to analysts or assumed a preset number or proportion of outliers^{9,10}. Although some techniques separate outliers in the output, they have 'tuning' parameters for analysts to set a threshold¹¹, or require separate, outlier-labelled or outlier-free datasets for training in advance¹². Quantification differs from dichotomisation, and for the latter, setting a threshold is imperative.

Let us consider a univariate unimodal dataset. There are two classical approaches: parametric and non-parametric. The former uses means and standard deviations of statistical probability distributions to quantify and expose outliers, while the latter quantiles. These classics even suggest a 'standard' threshold for dichotomisation. For example, when a normal distribution is assumed, data points with more than three standard deviations away from the mean are often treated as outliers¹³. For quantiles, the often proposed threshold is 1.5 times the interquartile range width away from the upper or lower quartile¹⁴. However, these are only conventional 'rules of thumb,' and their reasoning is unclear. 'There is no rigid mathematical definition of what constitutes an outlier; determining whether or not an observation is an outlier is ultimately a subjective exercise' (<https://en.wikipedia.org/wiki/Outlier>).

In this study, we propose a dichotomisation technique for outlier detection. We named it the 'GAP method' because it uses value gaps among data. Its repetitive process removes outliers stepwise. This technique requires analysts' decision on when to stop the repetition (to establish a final threshold) but lessens subjectivity by leaving only a few logical options. We developed this technique for wildlife GPS data, which often have a wide range of data values due to unstable measurement precision. In the following, we describe the principle and procedures of this technique and show applications using wildlife GPS data. We discuss the novelty and usefulness of the technique, tips for its wider application, and the implication of this study for anomaly/outlier detection in general.

Gap Method

a) Principle

The GAP method uses the gap index, GAP , to search for a threshold of outliers. We define GAP for each adjacent data pair in a dataset, where the data are in ascending order of values. Let the values of a pair be A and B and assume $0 \leq A \leq B$. Then, $GAP = B / A$ ($GAP \geq 1$ by definition). The core idea of the GAP method is to regard the gap of a data pair with the maximum GAP in the upper range of the dataset as the threshold for outliers. The complementary idea is to repeat the removal of outliers based on the core idea until the maximum GAP falls below a certain level. The 'maxGAP' denotes the maximum GAP hereafter.

b) The nature of GAP

To understand why this is effective, we will elucidate the nature of GAP .

Let us assume a dataset with ten positive integers in ascending order: (1, 2, 3, 4, 5, 6, 7, 8, 9, 10). This dataset has nine GAP s: (2.00, 1.50, 1.33, 1.25, 1.20, 1.17, 1.14, 1.13, 1.11). Of these, maxGAP is 2.00, which lies between 1 and 2 in the original dataset. This dataset is an arithmetic sequence in which the differences between adjacent data are the same. In an arithmetic sequence with

positive numbers, maxGAP lies between the first and second minimum values. In contrast, all *GAPs* in a geometric sequence, for example, (1, 2, 4, 8, 16, 32, 64, 128, 256, 512), are the same by definition. Now, let us assume a dataset with a wide range of positive real numbers in ascending order. Now, we can imagine that, in its bottom range, even a small value difference in an adjacent data pair can result in a large *GAP*, however, to have a large *GAP* in the top range, the difference must be relatively large.

The nature of *GAP* has three implications. First, *GAP* is useful only in the upper range of a dataset as a tool to search for a threshold for outliers. Second, we can expect a large *GAP* in the top range only when the range is sparsely populated, which is the situation where outliers occur. As the range becomes lower and more densely populated, *GAPs* decrease and become less significant as an indicator of the threshold for dichotomisation. Third, because *GAP* is a ratio between positive values, the dataset must be a ratio scale ¹⁵, where 0 means the ultimate minimum, not a partition between negative and positive values as seen in the interval scale.

c) Basic procedure

The basic procedure of the *GAP* method can be described in steps as follows:

- 1) Sort the data in a dataset in ascending order of values.
- 2) Find a data pair with maxGAP in the upper range of the dataset and regard all data above the pair's gap as outliers.
- 3) Remove the outliers from the dataset.
- 4) Return to step 2) with the dataset after removal.

Analysts decide when to stop the repetition; however, maxGAP, which quickly decreases with repetition, provides a clue for this decision. Our experience suggests that if maxGAP falls below 1.10, it is time to consider stopping the repetition (see the Application and Discussion). In implementing the procedure, we set 5% of the data size as the upper range of the datasets to search for maxGAP. This range was wide enough for our data because the procedure was invariably ended up with a much smaller percentage of removal (see the Application).

d) Generalised procedure

The basic procedure works only for a sequence-irrelevant univariate dataset. However, there are many cases that this procedure cannot handle. For example, when the dataset is a time series, a question of the temporal sequence of values may arise instead of one relating to the individual values of data points; when the dataset is multivariate, the question may not concern the individual variate values, but the multivariate composite values (i.e. the position of data points in multidimensional data space). In such a case, before applying the basic procedure, we must quantify data points to expose anomalies in question and create a univariate ratio-scale dataset. Let us call the quantified values 'anomaly-scores'. There is more than one way of quantification. It must be devised according to the nature of the question (see Application Example 1). The generalised procedure that adapts to such a case is as follows (the differences from the basic procedure are in italics):

0) Create an anomaly-score dataset from the original dataset.

- 1) Sort the data in *the anomaly-score* dataset in ascending order of values.
- 2) Find a data pair with maxGAP in the upper range of the *anomaly-score* dataset and regard all the data above the pair's gap as outliers.
- 3) Remove *the data from the original dataset that corresponds to the outliers in the anomaly-score dataset.*
- 4) Return to step *0)* with the original dataset after removal.

The difference from the basic procedure is the addition of step 0) and the inclusion of steps 0) and 1) for repetition. The creation of the anomaly-score dataset in step 0) is necessary at every repetition because the removal in step 3) causes a change in the data structure in the original dataset. In modified step 3), outliers exposed in the anomaly-score dataset in step 2) indicate which data in the original dataset to remove. Step 3) is simple only when a one-to-one correspondence between the original and the anomaly-score

datasets holds. When this is not the case, step 3) is difficult to generalise. We will show a specific example of step 3) for this case in Application Example 2.

Application to wildlife GPS data

The detection of outliers in wildlife GPS data is challenging because the data often have an extraordinarily wide range of values. Highly unstable measurement precision, caused by multiple factors of a highly variable and unpredictable nature, is partly responsible. Each GPS fix is usually given with the dilution of precision (DOP), which is the precision computable from the number of GPS satellites used and their constellation¹⁶. However, other factors affect the precision, such as the multipath of signals from satellites to receivers and the movement speed of animals. The multipath effect is inelible under thick forest cover^{17, 18, 19}, and fast movements naturally can make precise measurements of location difficult. However, their effects under changing conditions in the wild are difficult to measure and not given. Overall, the DOP is merely a computable part of precision. In addition, low precision does not necessarily imply low accuracy. The DOP-dependent data removal is both insufficient and misleading.

We are now developing an Excel system for analysing wildlife GPS data, which first checks the data to create suitable datasets for subsequent analysis. We developed the GAP method for this purpose. The GPS data are multivariate of time, height, and a pair of latitudes and longitudes. We applied the GAP method to each of these separately to combine the results later. We will describe three ways to implement the method using data obtained from two terrestrial mammalian species: the sika deer (*Cervus nippon*) and the pale-throated three-toed sloth (*Bradypus tridactylus*).

Example 1: The removal of height outliers

We used two procedures of the GAP method to remove height outliers: the basic procedure for sequence-independent outliers (extreme heights) and the generalised procedure for sequence-dependent outliers (height spikes). The latter procedure mostly suffices for our data, but the former ensures the removal of extreme values in a row. The default lower bounds of maxGAP in the former and latter procedures were set at 1.5 and 1.1, respectively. The bound for the former is higher because we expect it only to handle extreme values. The two procedures are combined and run sequentially. We used the following two datasets.

Original dataset(time-series of heights)

$(H_1, H_2, H_3, \dots, H_{i-2}, H_{i-1}, H_i, H_{i+1}, H_{i+2}, \dots)$

Anomaly-score dataset(sequence-dependent quantification for height spikes)

$(S_1, S_2, S_3, \dots, S_{i-2}, S_{i-1}, S_i, S_{i+1}, S_{i+2}, \dots)$

The former procedure uses only the original dataset. Although heights are an interval scale, we treated them as a ratio scale here for convenience (see Discussion). For the anomaly-score dataset in the latter procedure, we devised three variations of quantification for analysts to choose: (A) $S_i = (H_i - H_{i-1}) \times (H_i - H_{i+1})$; (B) $S_i = (2H_i - H_{i-1} - H_{i-2}) \times (2H_i - H_{i+1} - H_{i+2})$; (C) $S_i =$ the maximum of the following three products $((H_i - H_{i-1}) \times (H_i - H_{i+1}), (H_i - H_{i-2}) \times (H_i - H_{i+1}), (H_i - H_{i-1}) \times (H_i - H_{i+2}))$. In all of these variations, a one-to-one correspondence holds between the original and anomaly-score datasets. Some of the resulting values in these variations become negative; however, these only indicate the absence of height spikes and do not affect the procedure.

We used the GPS data obtained from a female sika deer living in a mountainous region in central Japan as an example (Supplementary Excel-DataFile 1). The original data spans 606 days, but for ease of demonstration, we selected data covering one year – from August 2012 to July 2013 at 2-hour intervals. The initial data size was 4254. The former and latter procedures resulted in 1 and 4 repetitions, 4.765 and 1.159 for the last maxGAPs, and 7 (0.16%) and 15 (0.35%) removals, respectively (Fig. 1; Table 1; Supplementary Video 1). Removal by the latter procedure included three downward spikes. We used the quantification variation (B) in this example. Naturally, the other variations had partly different results.

Example 2: The removal of horizontal position outliers

We used the generalised procedure to remove horizontal position outliers. The original dataset is a time series of latitude-longitude pairs. We created the interim dataset of XY coordinate pairs from the original using gnomonic projection and used it as a proxy for

distance calculation and mapping. We created the anomaly-score dataset using Euclidean distances between sequential data pairs in the original dataset.

Original dataset(time-series of latitude-longitude pairs)

$((L_{t_1}, L_{n_1}), (L_{t_2}, L_{n_2}), (L_{t_3}, L_{n_3}), (L_{t_4}, L_{n_4}), (L_{t_5}, L_{n_5}), \dots)$

Interim dataset(time-series of XY coordinates pairs)

$((X_1, Y_1), (X_2, Y_2), (X_3, Y_3), (X_4, Y_4), (X_5, Y_5), \dots)$

Anomaly-score dataset (Euclidean distances between sequential data pairs in original dataset)

$(D_{12}, D_{23}, D_{34}, D_{45}, D_{56}, \dots)$

Note that a one-to-one correspondence does not hold. The idea behind this quantification is to use distance outliers in the anomaly-score dataset to separate data points in the original dataset. Of those separated, let us call single points and groups of a small number of points 'minors'. Minors are likely to be spatial outliers but not always (see G5 in Phase 1 and G7 and G8 in Phase 4 in Fig. 2). For this reason, in implementing the procedure, we let analysts check which minors are spatial outliers. As a result, step 3) in the generalised procedure is described in the following sub-steps:

3 - 1) Separate the data points in the original dataset using distance outliers in the anomaly-score dataset.

3 - 2) Choose single points and groups of a small number of points as candidates for removal

3-3) Show these candidates and others on the map distinctively

3-4) Let analysts choose and remove only spatial outliers from the candidates

As another example, we used the GPS data obtained from a pale-throated three-toed sloth living in a thick tropical forest in Manaus, Brazil. The data covered 243 days from October 2019 to June 2020 at 15-minute intervals (Supplementary Excel-File 2). The initial data size was 22618. We, as an analyst, stopped the procedure after four repetitions with 1.088 for the last maxGAP and 14 (0.06%) removals (Fig. 2; Tables 2 & 3; Supplementary Video 2). Note that maxGAP and the spatial scale quickly decrease with repetition.

Discussion

The method proposed in this paper addresses the issue mostly neglected in outlier detection studies: dichotomisation. Its core idea of using significant value gaps for the purpose is simple and intuitive; hence, it is highly probable that analysts have long used it subjectively and unconsciously when separating outliers from others^{1,20}. However, we are unaware that the idea is elaborated and formalised as the technique proposed here.

The most appealing feature of the GAP method is that the quickly decreasing maxGAP lessons guesswork of analysts by providing only a few logical options. Analysts must make the final decision because they have broad background information, including the aim of the analysis. However, our experience suggested that analysts need to consider stopping the repetition of the procedure when the maxGAP falls below 1.10 (seen in Tables 1 and 3). We believe that this guideline is robust for a wide range of data sizes; however, it is only empirical and requires further elaboration. Developing 'gap statistics' may help in this regard.

The procedure of the GAP method is repetitive, and the removal of outliers is stepwise. This feature allows it to handle a wide range of data values in a single procedure (see Application Example 2). Extreme data values, often caused by device or system malfunctions, are easy for analysts to identify; however, not having to handle them separately from other outliers reduces the analysis burden. In addition, the stepwise feature of this method enables the stratification of outliers. This stratification is useful when analysts are interested in outliers, not as noise, but as a signal – for example, to analyse the causes of anomalies.

For the generalised procedure of the GAP method, quantification to create an anomaly-score dataset from the original dataset is necessary (Fig. 3). Although prior studies have proposed various quantification techniques (Fig. 4), the GAP method simplifies the task. First, separate (outlier-labelled or outlier-free) datasets for training (Fig. 4) may serve quantification but is unnecessary for

dichotomisation because the GAP method handles it. Second, the measurement unit (Fig. 4) is irrelevant because the GAP method uses ratios. For example, to quantify anomalies in a univariate and unimodal dataset, taking the absolute value difference between each data point and its central point is adequate; there is no need to consider the standard deviation or interquartile range (Fig. 4). Third, the central point as the origin of measurement can be mean, median, mode, or any other convenient point because the GAP method is robust to this choice. To expose extreme heights in Application Example 1, we regarded interval-scale heights as a ratio scale and applied the basic procedure of the GAP method. This expedient is equivalent to regarding height 0 as the central point. Overall, the 'robust statistics'^{3,21}, which tries to lessen the influence of outliers in quantification, is less needed for the GAP method.

Some anomalies appear as outliers as given, but others do not. In either case, quantification to expose them as outliers in a univariate dataset is usually necessary for their detection (Fig. 3). Here are two points. First, this quantification defines 'anomaly'. This aspect of quantification needs due attention besides the technical one. Second, anomaly/outlier 'detection' in the strict sense comes after quantification. In this regard, using the term 'anomaly/outlier detection' in studies that only deal with quantification techniques is misleading. Studies for 'anomaly/outlier detection' must come with a dichotomisation technique such as the GAP method (Fig. 3).

After quantification, the envisaged question is: 'how large is large enough for the value to make data points outliers?' Instead, the GAP method asks, 'how large is large enough for a value gap to separate outliers from others?' This modification is a significant shift of focus, and this question well represents humans' natural cognition of outliers²⁰. Hawkins (1980)¹ states:

A sample containing outliers would show up such characteristics as large gaps between 'outlying' and 'inlying' observations and the deviation between outliers and the group of inliers, as measured on some suitably standardized scale.

This statement exactly matches the idea of the GAP method. There have long been indecisive arguments as to what is outliers^{2,5}. The GAP method addresses the issue from a long neglected perspective.

Declarations

Acknowledgements

The GPS survey of the sloth was supported by JST/JICA, SATREPS, JSPS Core-to-Core Program, A. Advanced Research Networks (Wildlife Research Center of Kyoto University), and JSPS KAKENHI Grant Number JP18K06414 .

Ethics declaration

1) Sloth GPS study

Harnesses and collars were made of 10-mm-wide, 1.3-mm-thin PP webbings which were wide and soft enough to prevent damage to the animals. Every time we captured or observed the animals, we checked the condition of the animals, especially on the skin around the webbings. Marking and capturing were authorised by the biodiversity authorisation and information system (SISBIO-60116-1) and by the Ethics Committee of the Federal University of Amazonas (006/2018-CEUA/UFAM).

2) Deer GPS study

An academic capture permission for the deer was granted from Nagano Prefecture. We adhered to the animal capture and handling guidelines of the Mammalogical Society of Japan (2009). The GPS collars were equipped with a radio-and-timer-controlled system for drop-off within 2 years.

*The Mammalogical Society of Japan. The guidelines for the treatment of animal samples. *Honyurui Kagaku* **49**, 303–319 (2009) [in Japanese].

Author contributions

H.H. developed this technique in collaboration with the other authors. D.M. and M.G. conducted the sloth GPS surveys. A.T. and S.I. conducted the deer GPS surveys. H.H. wrote the manuscript, and all the other authors reviewed it.

Data availability statement

The data used in the present study are available in Excel files provided as supplementary information of this paper.

Competing interests

The authors declare no competing interests.

References

1. Hawkins, D. *Identification of Outliers*. (Chapman and Hall, 1980).
2. Beckman, R. J. & Cook, R. D. Outlier ... s, *Technometrics* **25**, 119–149 (1983).
3. Rousseeuw, P. & Leroy, A. *Robust Regression and Outlier Detection*. (John Wiley & Sons, 1987).
4. Hodge, V. J., & Austin, J. A survey of outlier detection methodologies. *Artif. Intell. Rev.* **22**, 85–126 (2004).
5. Chandola, V., Banerjee, A., & Kumar, V. Outlier detection: a survey. *ACM Comput. Surv.* **41**, 1–83 (2009).
6. Aggarwal, C. C. *Outlier Analysis*. (Springer, 2017).
7. Zimek, A. & Filtzmoser, P. There and back again: outlier detection between statistical reasoning and data mining algorithms. *WIREs Data Min. Knowl. Discov.* **8**:e1280 (2018).
8. Blázquez-García, A., Conde, A., Mori, U. & Lozano, J. A. A review on outlier/anomaly detection in time series data. *ACM Comput. Surv.* **54**, 1–33 (2022).
9. Jin, W., Tung, A. K. H. & Han, J. Mining Top-n Local Outliers in Large Databases. *Proc. 7th ACM SIGKDD* 293–298 (2001).
10. Lu, H., Liu, Y., Fei, Z., & Guan, C. An outlier detection algorithm based on cross-correlation analysis for time series dataset. *IEEE Access* **6**, 53593–53610 (2018).
11. Ester, M., Kriegel, H. P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *kdd* **96**, 226–231 (1996).
12. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
13. Grubbs, F. E. Procedures for detecting outlying observations in samples. *Technometrics* **11**, 1–21 (1969).
14. Tukey, J. *Exploratory Data Analysis*. (Addison-Wesley, 1977).
15. Stevens, S. S. On the theory of scales of measurement. *Science* **103**, 677–680 (1946).
16. Langley, R. B. Dilution of precision. *GPS World* **38**, 52–59 (1999).
17. Frair, J. L. *et al.* Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Phil. Trans. R. Soc. B* **365**, 2187–2200 (2010).
18. Pirti, A. Using GPS near the forest and quality control. *Surv. Rev.* **38**, 286–298 (2013).
19. Forin-Wiart, M.-A., Hubert, P., Sirguey, P. & Poulle M.-L. Performance and accuracy of lightweight and low-cost GPS data loggers according to antenna positions, fix intervals, habitats and animal movements. *PLoS One* **10**, e0129271 (2015).
20. Collett, D. & Lewis, T. The subjective nature of outlier rejection procedures. *J. Roy. Stat. Soc. C-App.* **25**, 228–237 (1976).
21. Rousseeuw, P. J. & Hubert, M. Robust statistics for outlier detection. *WIREs Data Min. Knowl. Discov.* **1**, 73–79 (2011).

Tables

Table 1. Repetition phases in the removal of height outliers in Fig. 1.

The initial data size was 4254; the number of removals was 22 (0.52%), and three were downward peaks. Values in the threshold column represent the lower values of the data pairs with maxGAP in each phase. The upper values of the data pairs is a product of the maxGAP and threshold values.

Height Anomaly Scores				Removal Candidates		Removed	
Sequence	Repetition	maxGAP	Threshold	Up. Peaks	Up+Downs	Up. Peaks	Up+Downs
Independent	Phase 1	4.765	/ 1788	7	7	7	7
	stopped	1.098	/ 1609	2	2		
Dependent	Phase 1	3.178	/ 442133	1	1	1	1
	Phase 2	1.484	/ 297920	1	1	1	1
	Phase 3	1.411	/ 211214	1	1	1	1
	Phase 4	1.159	/ 109984	9	12	9	12
	stopped	1.068	/ 42261	16	46		
Total						19	22

Table 2. Separated point groups in each phase in the removal of horizontal position outliers in Fig. 2.

Of the groups separated in each phase, minors represent single points or groups of a small number of points. In each phase, analysts decided which minors were spatial outliers to remove. See also Table 3.

Phase 1			Phase 2			Phase 3			Phase 4		
GroupID	Size	Minors									
G1	16620		G1	16286		G1	11533		G1	8148	
S2	1	1	G2	4	4	S2	1	1	G2	2	2
G3	5993		G3	6326		G3	6214		G3	4465	
S4	1	1	Total	22616	4	G4	2	2	S4	1	1
G5	3	3				G5	4862		G5	5937	
Total	22618	5				Total	22612	3	S6	1	1
									G7	78	78
									G8	69	69
									S9	1	1
									G10	3907	
									Total	22609	152
Removed		S2,S4	Removed		G2	Removed		S2,G4	Removed		G2,S4,S6,S9
Not Removed		G5	Not Removed			Not Removed			Not Removed		

Table 3. Repetition phases in the removal of horizontal position outliers in Fig. 2.

The initial data size was 22618; there were 14 (0.06%) removals. See also Table 2. Note that maxGAP and the spatial scale quickly decrease with repetition.

Repetition	Distance (m)		Separated	Removal Candidates		Removed	
Phase	maxGAP	Threshold	#Groups	#Group	#Points	#Group	#Points
1	978.8	/ 11243	5	3	5	2	2
2	14.55	/ 763	3	1	4	1	4
3	1.423	/ 337	5	2	3	2	3
4	1.088	/ 265	10	6	152	4	5
stopped	1.034	/ 234	12	2	2		
Total						9	14

Figures

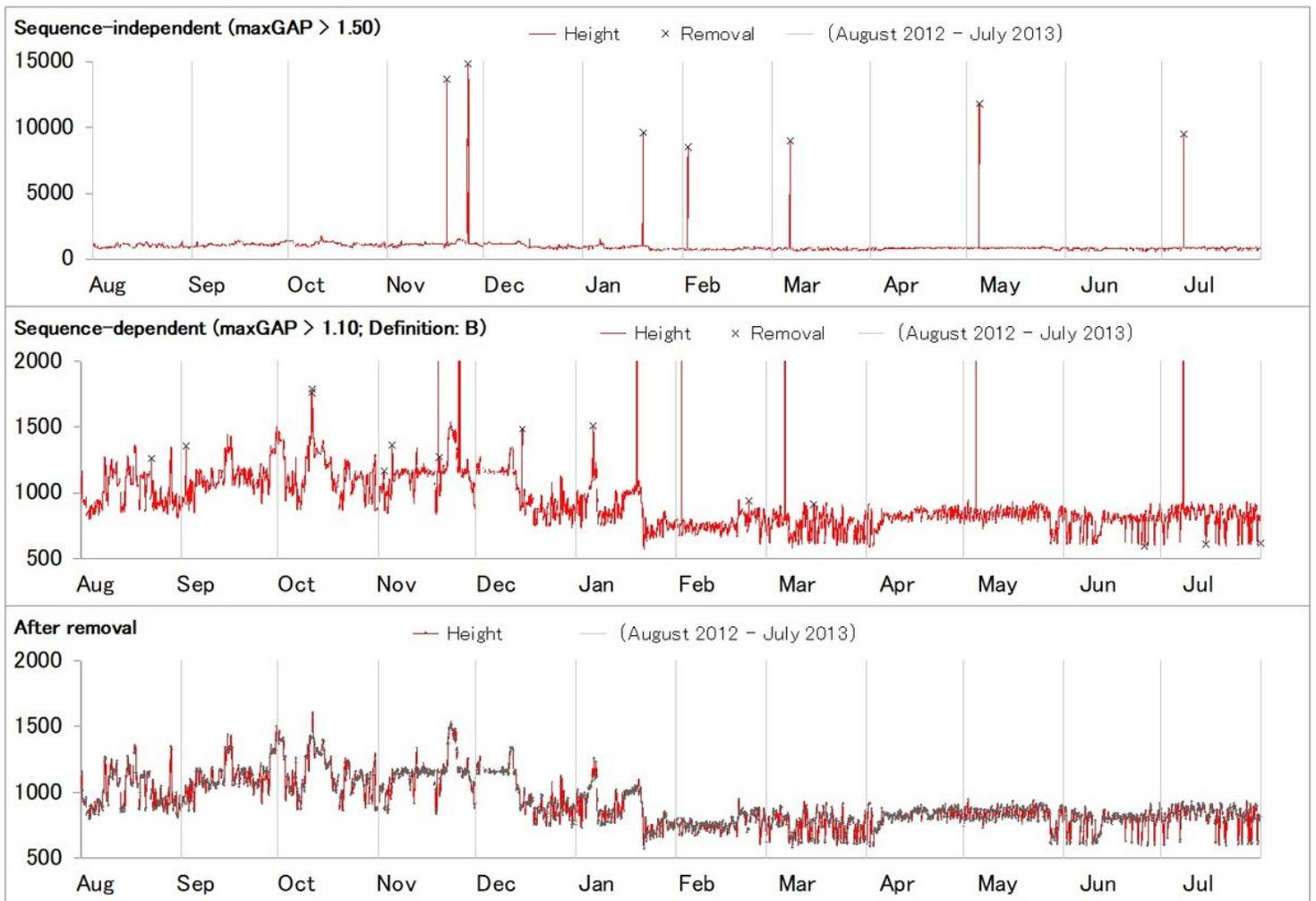


Figure 1

Removal of height outliers in the Sika deer data

We obtained the data from a female sika deer (*Cervus nippon*) living in a mountainous region in central Japan, spanning one year from August 2012 to July 2013 at 2-hour intervals. The initial data size was 4254. Red lines represent the sequence of height values; crosses indicate the heights to remove. Top panel: Removal of sequence-independent outliers (maxGAP > 1.5); middle panel: removal of sequence-dependent outliers (maxGAP > 1.1); bottom panel: after removal. See Table 1.

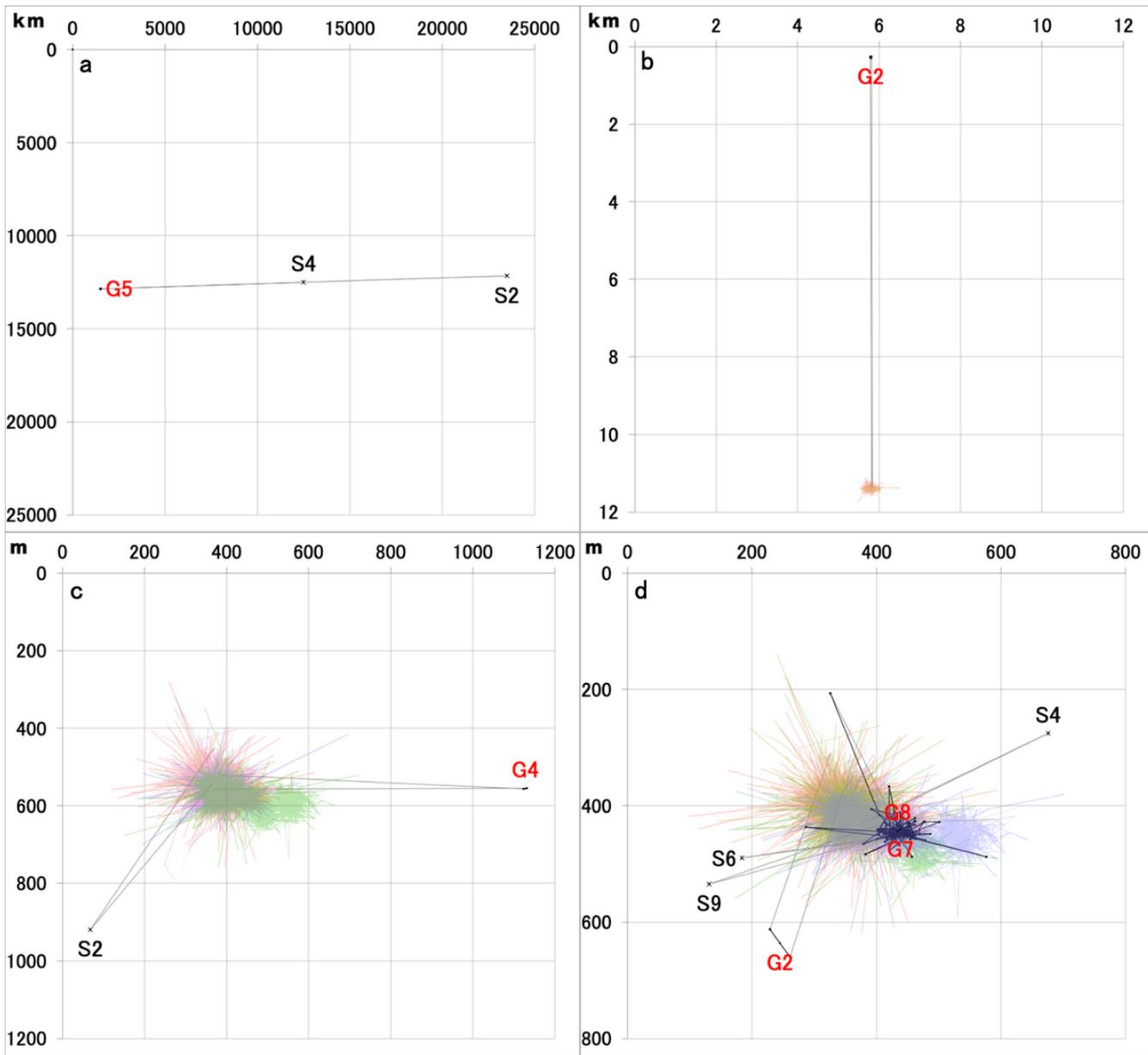


Figure 2

Phases in the removal of horizontal position outliers in the three-toed sloth data.

We obtained the data from a pale-throated three-toed sloth (*Bradypus tridactylus*) living in a tropical forest in Manaus, Brazil. The data span 243 days from October 2019 to June 2020 at 15-minute intervals. The panels showing each removal phase are arranged from top to bottom and left to right. See Table 2 and 3.

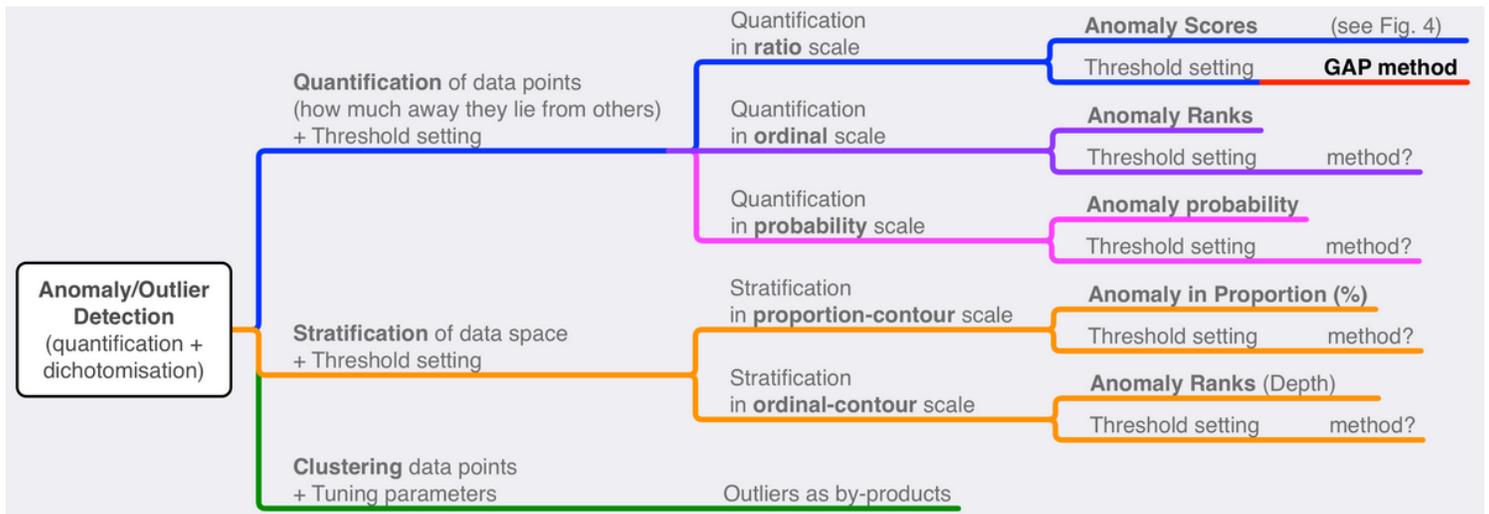


Figure 3

A quantification-based classification of outlier detection methods and the position of the GAP method.

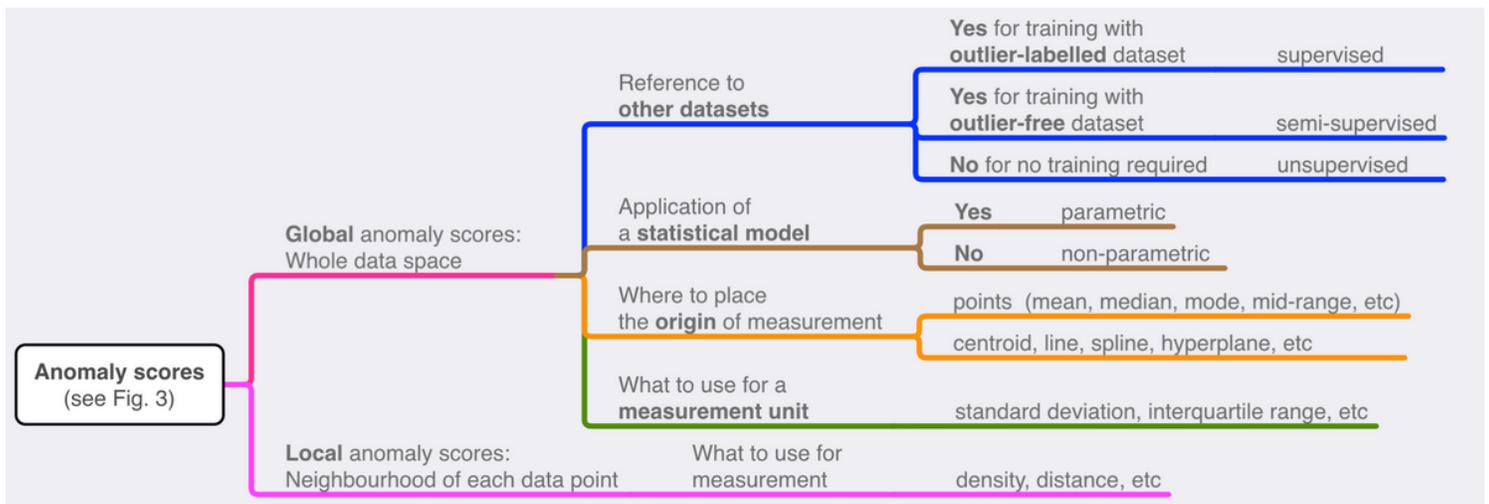


Figure 4

Variations of the techniques to measure anomaly scores.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [DSBT45619102006E.xlsx](#)
- [DSRina12081307E.xlsx](#)
- [HeightCheck.mov](#)
- [RemoteCheck.mov](#)