Information Visualization

http://ivi.sagepub.com/

Multi-scale analysis of linear data in a two-dimensional space

Yi Qiang, Seyed H Chavoshi, Steven Logghe, Philippe De Maeyer and Nico Van de Weghe

Information Visualization published online 25 March 2013

DOI: 10.1177/1473871613477853

Info Vis

Special Issue Article

Information Visualization
0(0) 1–18

DOI: 10.1177/1473871613477853
ivi.sagepub.com
SAGE

Multi-scale analysis of linear data in a two-dimensional space

Yi Qiang[1], Seyed H Chavoshi[1], Steven Logghe[2], Philippe De Maeyer[1] and Nico Van de Weghe[1]

Abstract

Many disciplines are faced with the problem of handling time-series data. This study introduces an innovative visual representation for time series, namely the continuous triangular model. In the continuous triangular model, all subintervals of a time series can be represented in a two-dimensional continuous field, where every point represents a subinterval of the time series, and the value at the point is derived through a certain function (e.g. average or summation) of the time series within the subinterval. The continuous triangular model thus provides an explicit overview of time series at all different scales. In addition to time series, the continuous triangular model can be applied to a broader sense of linear data, such as traffic along a road. This study shows how the continuous triangular model can facilitate the visual analysis of different types of linear data. We also show how the coordinate interval space in the continuous triangular model can support the analysis of multiple time series through spatial analysis methods, including map algebra and cartographic modelling. Real-world datasets and scenarios are employed to demonstrate the usefulness of this approach.

Keywords

Time series, linear data, multi-scale analysis, information visualization, multi-criteria analysis

Introduction

Many disciplines are faced with the problem of handling time-series data, which lead to considerable efforts dedicated to the research of time series.[1–3] The temporal scale is one of the most important issues in time-series analysis. Analogous to the well-known modifiable areal unit problem in spatial analysis, the way of aggregating temporal data may also significantly affect analysis results. Sometimes, patterns or relationships detectible in a certain scale cannot be detected in other scales. Even in the same scale, different partitions of intervals may result in different patterns being revealed. On the other hand, a question can be answered in different scales. For example, the answers to the question when there are a lot of traffic jams in Belgium may include 'between 7:00 a.m. and 9:00 a.m.', 'during the days it snows' and 'in the months of school semester'. All these answers make sense because they may guide people to take actions in corresponding scales. Therefore, an appropriate choice of the temporal scale should take account of the characteristics of phenomena under study, the level of questions being asked, and the scale of actions to be taken. This choice is not easy, particularly in the phase of exploratory analysis, when there is not much known about the data and when the objective of the analysis is not accurately specified. In addition to specifying an appropriate scale for analysis, the hierarchy of phenomena in different scales can also be important

[1]Department of Geography, Ghent University, Ghent, Belgium
[2]Be-Mobile NV, Ghent, Belgium

Corresponding author:
Yi Qiang, Department of Geography, Ghent University, Krijgslaan 281, Ghent 9000, Belgium.
Email: yi_qiang@hotmail.com

in certain analytical tasks.[4] Analysts may be interested in how long-term patterns are composed or influenced by short-term patterns within them. As a result, multi-scale analysis is of critical importance for analysing temporal data. Due to the complexity of this issue, the solution requires considerable human intelligence to be involved.

Visualization has been proven to be an effective analytical approach for time-series data.[5] An explicit visualization can effectively combine the insight of humans and processing ability of computers[6,7] to tackle analysis tasks. While a number of approaches have been developed to visualize time series,[8–11] the line chart remains the most frequently used. In a line chart, the horizontal dimension indicates positions in the timeline, and the vertical dimension indicates the values at time stamps. The time series is represented as a curve, offering a direct view of the variation of time series along the linear space. Line diagrams usually only display time series in a certain scale. Displaying time series in different scales would require drawing more curves, which makes the data display matted. Manipulating a sliding bar to shift the scales to be displayed is an alternative approach. However, with the slider, one still cannot obtain an overall picture of time series in all different scales.

The continuous triangular model (CTM) introduced in this study provides an alternative approach to represent time series and overcomes the difficulty of traditional approaches in visualizing time series in multiple scales. The CTM is based on a diagrammatic representation of time intervals initially proposed by Kulpa.[12,13] Later, Van de Weghe et al.[14] named it the triangular model (TM) and applied it to archaeological use cases More recently, Qiang et al.[15–17] investigated its use in reasoning imperfect intervals and visual analytics. The basic idea of the TM is representing time intervals as points in a coordinated two-dimensional (2D) space. Evolved from the TM, the CTM adds the third dimension to the interval space of the TM and forms a continuous field, which can display time series in all different intervals. In the continuous field, every point represents a specific interval and is referenced to a certain value of the interval, such as the summation, average, or standard deviation. On the one hand, the CTM can provide an overview of linear time series in all different scales. On the other hand, as the CTM is based on a 2D coordinate space, the glossary of spatial analysis methods in geographical information science (GIScience) is now open to be employed to manipulate and analyse the CTM data.[18,19] In addition to time series, the CTM can also be applied to a broader sense of linear data, which refers to data sequences ordered in a one-dimensional (1D) space. Linear data can be derived from a linear geographical space, such as traffic speed along a road and run-off along a river, or objects with a linear structure, such as text and DNA sequences.

In the remainder of this article, we first review the representative approaches in temporal visualization. Next, the basic concept of the TM is introduced, followed by its extension to the CTM. We then demonstrate how the CTM can be applied to visualize time series of soccer players and traffic speed along a motorway. Afterwards, we show how map algebra and cartographic modelling can be applied to analyse time series represented in the CTM. Finally, we discuss the data and task coverage of the CTM at a conceptual level, followed by conclusions and future study.

Related study

As time is perceivable but invisible, the visual representation of time is a key issue in the analysis of temporal information. Due to the sophisticated nature of time, visualization approaches for temporal data appear to be highly diverse and customized according to particular analytical tasks. In the past, visualizations of temporal data were limited to static diagrams on paper. Later, the invention of computer has liberated people from manually chart drawing by automatic graphic techniques. The advanced processing ability of computers has allowed a wide variety of techniques to be applied to visualize temporal data. Time series is the data type that receives most attention. Many visualization tools extend the conventional line chart with interactively linked components. For instance, Van Wijk and van Selow[20] developed a tool that combines a line chart and a calendar, in order to display pattern clusters of daily time series. The line chart visualizes averages of different time-series clusters, while the colour of the dates in the calendar indicates which cluster the dates belong to. This system enables visualization of time series in two different timescales, that is, daily and hourly. Besides, Lin et al.[10] developed a visualization tool (VisTree) that combines a tree representation with line charts for visual detection of special patterns in time series. In the tree representation, every branch represents a specific subsequence of time series, and the thickness of the branch indicates the occurrence of identical subsequences. Using this approach, the frequently appeared patterns can be visually recognized from the thickness of branches.

In addition to the interactive line charts, advanced display and printing techniques allow the use of large amounts of colours and raster images, leading to a number of innovative visual metaphors for time series. Compared with traditional line charts, colour lines[21,22] are considered as a more efficient visualization for a large number of time series. This technique uses a straight line with a certain colour scheme to represent a time series so that numerous parallel time series can be compactly arranged within a rectangular region. With the support of interactive sorting tools, users are able to visually identify clusters of time series with similar patterns. The two-tone pseudo colouring technique[23] can be viewed as a combination between colour-coding and the line chart, which facilitates the overview of a large number of time series and also supports the precise observation of specific values. The other well-known approach is ThemeRiver,[8] which uses a continuous flow of colour bands to represent multiple time series. Each band in the ThemeRiver represents a specific theme, and the bandwidth indicates the number of its appearance in selected documents at a certain point of time. ThemeRiver combined the advantages of the line chart and the bar chart, enabling users to observe the variation of individual time series and also the difference among multiple series.

Some tools are specially designed to visualize the time series with cyclic characteristics. Most of them derive from the idea of the spiral graph, which portraits time as a spiral line flowing from the inside out. The examples of the spiral representation can be found in the visualization tools developed by Carlis and Konstan[24] and Weber et al.[11] In these spiral visualizations, the linear variation of time series can be observed along the spiral, while the periodical patterns can be detected in the straight line that extends from the centre of the spiral outwards. On the contrary, Li and Kraak[25] proposed an intermediate approach between the timeline and the spiral, which depicts the timeline as waves. A promising feature of this approach is that the time circles of more than one scale (e.g. yearly and monthly) can be displayed in one diagram. The other two intriguing approaches are the TimeWheel and the MutliCombo introduced by Tominski et al.[26]

More extensive reviews of time-series visualization can be found in the works of Aigner et al.,[5] Muller and Schumann,[27] and Aigner et al.[28] A common weakness of existing approaches is that time series can only be displayed in one or a few preset temporal scales, which may miss interesting patterns in other scales. This also prohibits the observation of the complete hierarchy of scales, where smaller phenomena are nested within larger phenomena. Moreover, in a certain scale, time series is usually visualized in equal-length time granules, for example, hour, day, month, or year. The patterns in intervals that partially overlap granules cannot be displayed. To the best of our knowledge, there is no approach that can break away the barriers between scales and visualize time series in all intervals within the considered time frame. This problem also exists in a broader sense of linear data. The CTM presented in this study can be considered as a solution to this problem. Also, the CTM can be transferred to a cone to visualize cyclic data. The cyclic transformation of the CTM is in our future research agenda.

An idea similar to the CTM is the growth matrix introduced by Keim et al.,[29] which visualizes stock price changes in a 2D space. In the growth matrix, the horizontal axis indicates the time when the fund is purchased, and the vertical axis indicates when the fund is sold. Every point in the matrix is referenced to the price difference between the purchasing and selling times. Beyond Keim's research, this study demonstrates how other formulas (i.e. average and summations) can be applied to calculate the values of intervals, and how this representation can be useful for analysing different types of linear data (e.g. traffic data along a road). Moreover, it shows the use of map algebra in comparing multiple time series, and how cartographic modelling[30] can be applied to the CTM to solve multi-criteria decision-making problems based on time series.

Basic concepts

The TM

In the classical linear representation, a time interval I is represented as a linear segment bounded by a start point I^- and end point I^+. The properties of an interval are expressed by the location and extent of the linear segment in a 1D space. The basic idea of the TM is mapping the linear segment in the 1D space into points in a 2D space. Given an arbitrary time interval I, two straight lines (L_1 and L_2) are projected from the two extremes (I^- and I^+), with L_1 passing through I^- and L_2 passing through I^+ (Figure 1). α_1 is the angle between L_1 and the horizontal axis, while α_2 is the angle between L_2 and the horizontal axis, where $\alpha_1 = \alpha_2 = \alpha$. The intersection point of L_1 and L_2 is called the interval point, which expresses the properties of the time interval I. The horizontal position indicates the midpoint of I, that is, $mid(I) = (I^- + I^+)/2$, while the vertical position indicates the duration of I, that is, $dur(I) = \tan\alpha(I^+ - I^-)/2$. The start of the interval I^-, the end of the interval I^+, and interval point I form an isosceles triangle. Therefore, this representation of time intervals is called the TM. The angle α is a predefined constant that is identical to the construction of all interval points. Here, we set $\alpha = 45°$ to be consistent with previous study,[13,15,31] though α can be set to any value between 0° and 90° for specific purposes. In the TM, every time interval can be represented as a unique point in the 2D space. The 2D space where

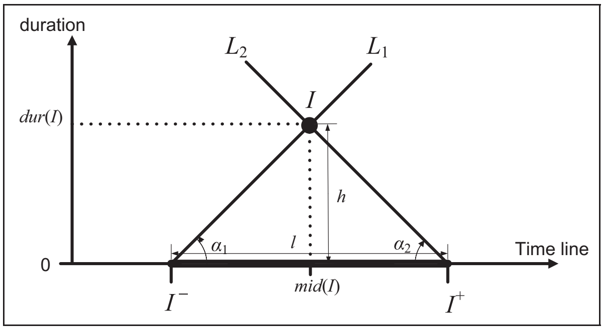

Figure 1. Configuration of the triangular model (TM).

interval points are located in is called the interval space ($I\mathbb{R}$).

According to the interval algebra introduced by Allen,[32] there exist 13 atomic relations between two time intervals. Since the TM represents time intervals as points in a 2D coordinate space, the relations between time intervals are expressed by spatial relations. Given a study interval I [0, 100], all examined intervals are located within the isosceles triangle formed by I^-, I^+, and the interval point of I. Let us consider a reference interval I_1 [33,66] and several intervals (I_2, I_3, and I_4) that are *before* the interval I_1 (Figure 2(a)). In TM, I_2, I_3, and I_4 are located in the zone in the left corner of the study interval (Figure 2(b)). Therefore, it is easy to deduce that this zone (i.e. the black zone in Figure 2(c)) encloses all intervals that are *before* the interval I_1. In this manner, all Allen's relations with respect to an interval can be represented by such zones in TM[13,31] (see Figure 3). In each diagram in Figure 3, it is assumed that there exists a triangular study area that contains all intervals, and the referenced interval I_1 is in the centre of the study area. Each black zone represents the set of intervals that are in a certain relation to I_1. Temporal constraints based on Allen's relations can thus be modelled as such zones. The composition of temporal constraints is based on the same principle as that of the Venn diagram. For instance, the set of intervals that satisfy several constraints are located in the intersection of the corresponding zones. The set of intervals that satisfy any of the constraints is located in the union of the zones. For a more detailed description of the relational zones and their compositions, please refer to our previous study.[17]

The CTM

In addition to discrete time intervals, the TM can be extended to represent continuous temporal data. Given a time interval I, all intervals *during* I are enclosed in a triangular zone below it (see Figure 3). In other words, every interval I_n *during* I corresponds to a specific point in this triangular zone. Let us consider a linear dataset arranged within I. Every point in the triangular zone represents a subinterval I_n of the linear data. If every point is assigned a certain value, that is, $f(I_n)$, of the interval it represents, then the triangular area can be filled and becomes a continuous field. $f(I_n)$ is a certain formula dependent on I_n, such as the average, summation, or standard deviation of the linear data in I_n. Figure 4 illustrates how the CTMis built from a linear data sequence consisting of seven numbers. It shows that every point in the triangular area represents a certain subinterval of the sequence and assigns a value that is calculated from the numbers within the subinterval. Here, the granularity of the CTM is consistent with that of the linear data sequence. Finer granularity can be obtained through interpolation. Figure 5 gives an example of the implementation of the CTM in a raster space. Through colour-coding, the CTM can be displayed as an image.

Visualization of linear data

Visualizing time series

This subsection demonstrates how the time series of the moving speed of a soccer player can be represented in the CTM. The movement of the soccer players is obtained through digitalization of the game video. Here, we study an indoor soccer game, in which each team has five players. The time series is a player's speed in every second (i.e. meter/second) during a study interval of 800 s. As indoor soccer is rather intensive and fast-paced, the line chart (i.e. Figure 6(a)) exhibits dramatic changes in speed from second to second. However, variations in longer intervals (e.g. 1 min or 2 min) are hard to observe. In Figure 6(b), the time series of the player's speed is represented by the CTM, where $f(I_n)$ is the average of the player's average speed during I_n. In the CTM, short-term fluctuations can be observed in lower levels, while the long-term patterns can be observed in higher levels. Moreover, it explicitly displays a hierarchy of the time series in all different scales, in which one can observe the relationship between the short-term variations and long-term variations. From Figure 6(b), one can identify intervals of sprint from the red areas at the bottom of the CTM, for example, I_1, I_2, and I_3. On a larger scale, it is clear that the player had a high average speed from 1:00 to 6:15 (i.e. I_4). However, during the next 3.5 min (i.e. I_5), he experienced a less active period, although there are still several sprints during it. Compared to the growth matrix of Keim et al.,[29] in

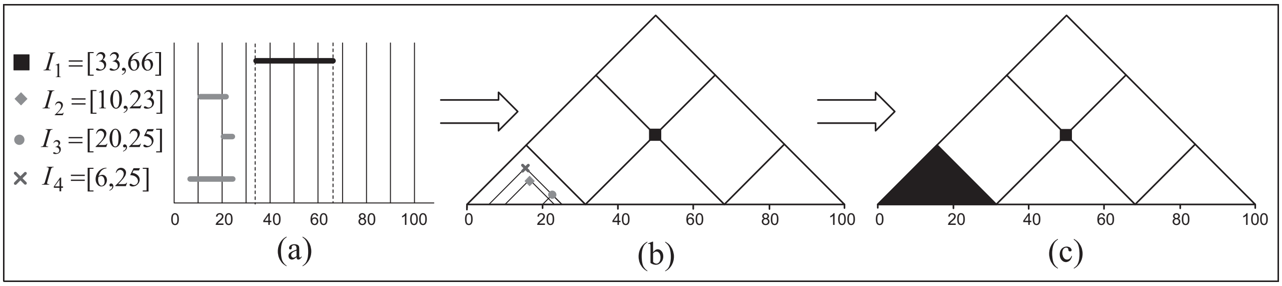

Figure 2. Temporal relations in the linear model and in the TM, taking *before* as an example.
TM: triangular model.

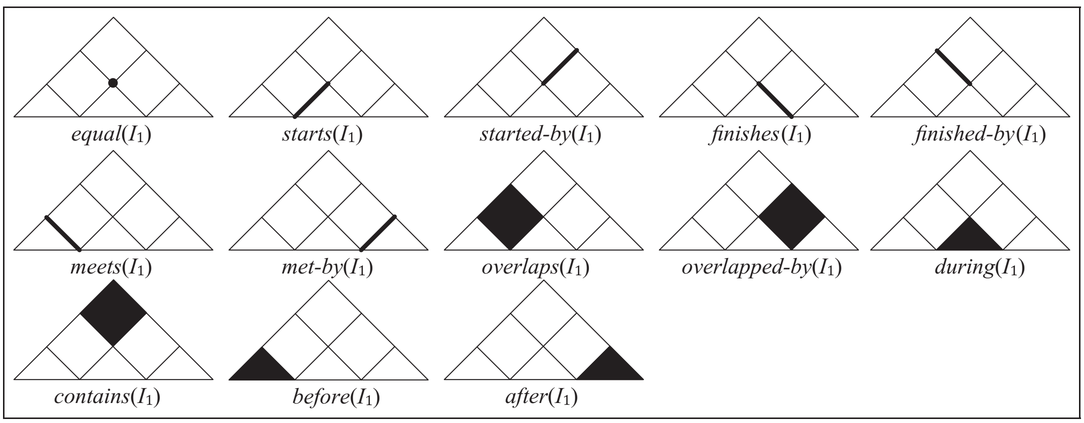

Figure 3. Representation of 13 Allen's relations in the TM.
TM: triangular model.

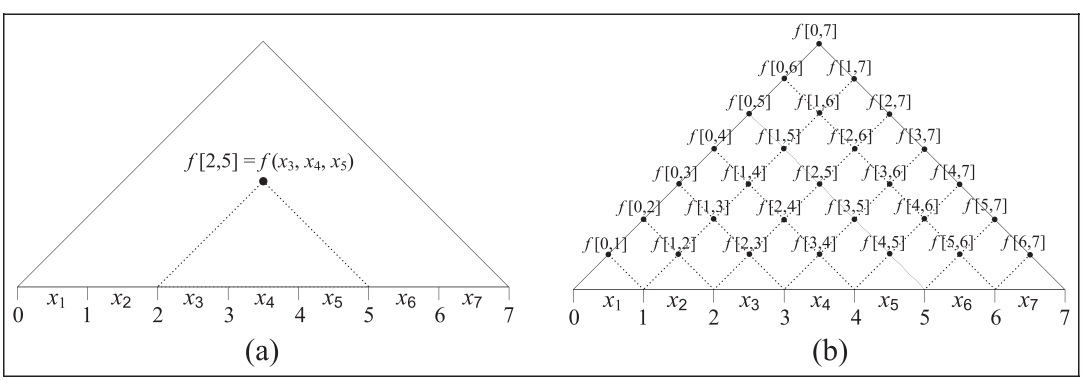

Figure 4. Representing a linear data sequence with seven numbers in the CTM: (a) a point is assigned a number calculated from the numbers within a subinterval and (b) every point in the triangular space is assigned a number of a specific subinterval.
CTM: continuous triangular model.

which I^- and I^+ are coordinated along the vertical and horizontal axes, respectively, the coordinate space of the CTM preserves the linear nature of time that flows from left to right. Mapping longer intervals in higher positions is also somehow more intuitive than the growth matrix.

Visualizing traffic speed

The CTM can also be applied to other types of linear data, such as traffic speed along a road, which is detected at a sequence of minimum road segments. In this case, every point in the CTM represents a specific road segment, and the colour at the point indicates the

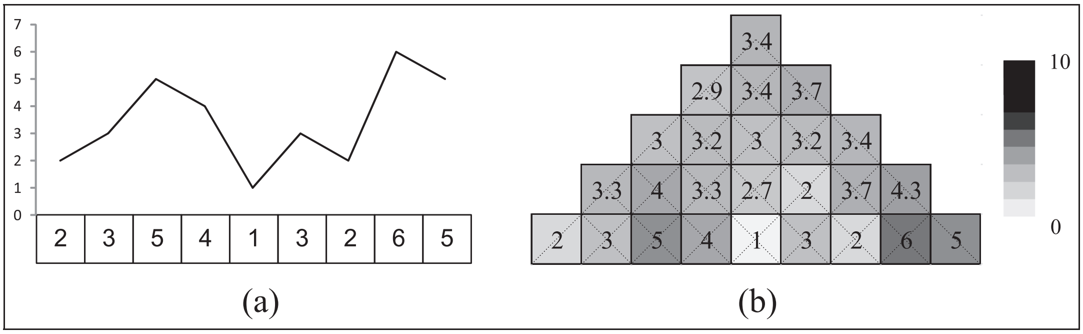

Figure 5. Implementation of the CTM in a raster space: (a) a linear data sequence and its representation in a line chart and (b) CTM representation of the linear data sequence in (a), with the average formula applied.
CTM: continuous triangular model.

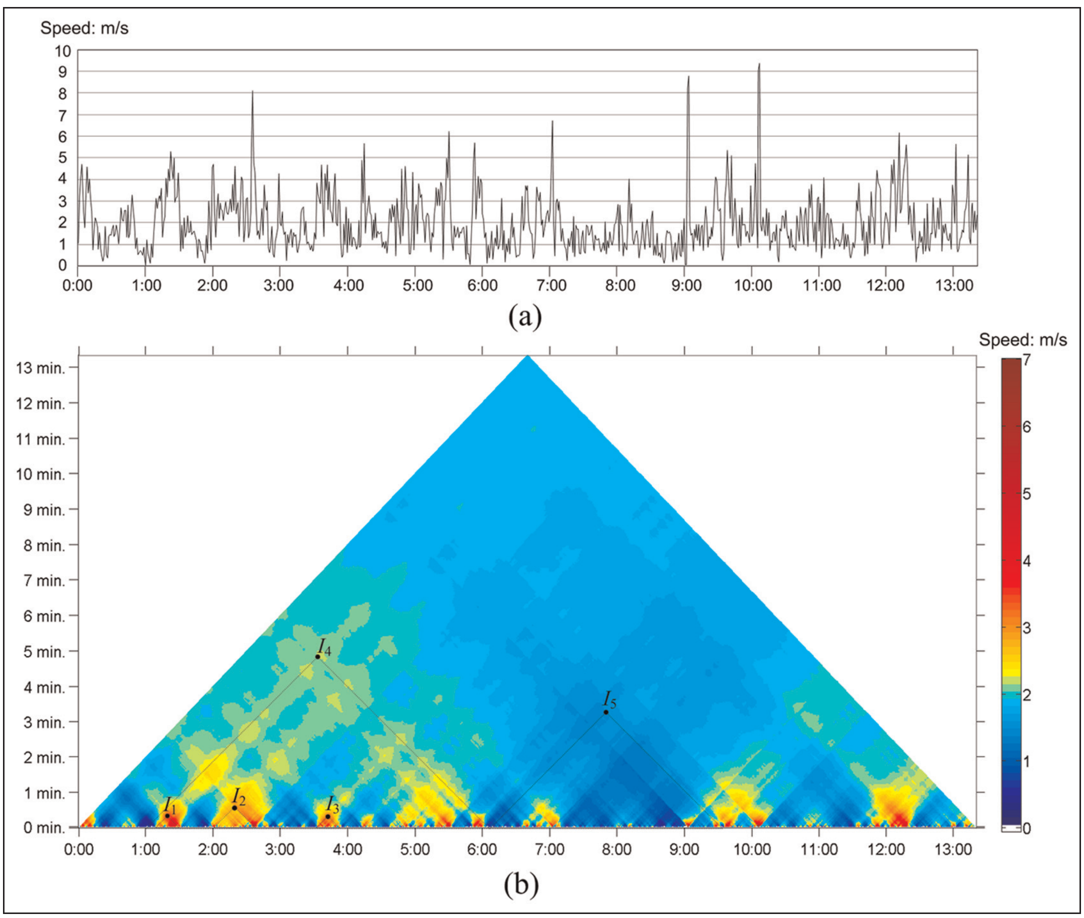

Figure 6. (a) Time series of a soccer player's speed in a line diagram and (b) the CTM.
CTM: continuous triangular model.

average speed of traffic in this segment. One CTM diagram can only represent the traffic in one direction, which is from left to right in this case. In Figure 7, the four diagrams represent the traffic speeds along the E40 motorway in Belgium from Merelbeke (near Ghent) to Boerderijstraat (in Brussels) at four different timestamps. The ticks on the horizontal axis indicate the exits and entrances along the road. As the exit and entrance for one place are always less than 1 km, a single tick is used to mark both the exit to and entrance from the same place. Lines that are in α and $-\alpha$ to the horizontal axis are drawn from these ticks. The average

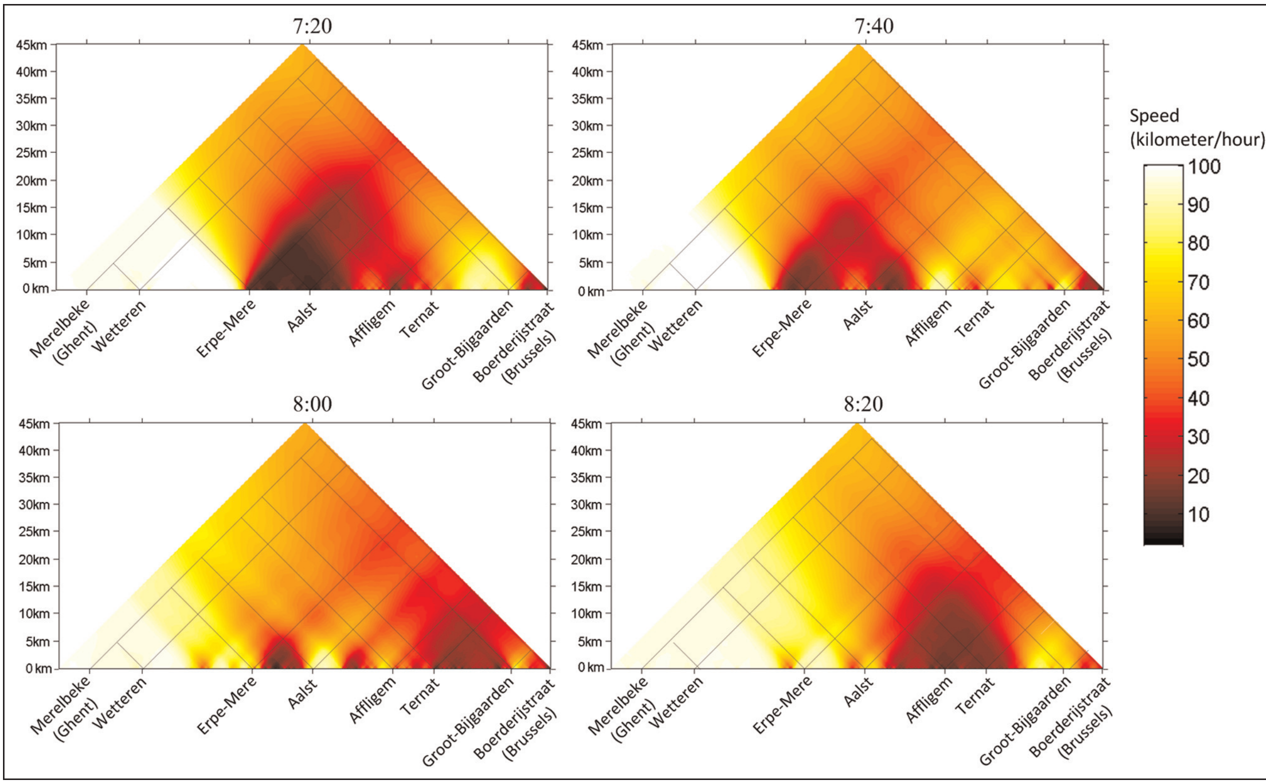

Figure 7. CTM diagrams of the average speed on the E40 motorway from Ghent to Brussels.
CTM: continuous triangular model.

speed from one place to the other can be read from the colour at the intersection point of two lines projected from the two places. In traditional representations, such as line charts and colour-coded polylines on maps, one can only read the traffic speed in road segments that are partitioned in a certain scale. The average speed across several partitions is obtained through mental estimation, which is not precise. In contrast, the CTM diagram provides an explicit overview of the average speed in all different segments of the road. One can observe the location of traffic jams along the road (i.e. low-speed road segments) in the bottom of the triangular field and also how much the traffic jams influence the average speed of longer distance from the higher levels. In Figure 7, one can observe that the average speed from Erpe-Mere to Ternat at 7:20 a.m. has fallen below 20 km/h (the 24-h clock is applied in this article), as the intersection point of the lines from these two places is in a very dark area. Thus, it is strongly not recommended that drivers take the motorway during this segment. In the higher level, one can identify that the average speed from Ghent to Brussels is approximately 60 km/h. Thus, taking this motorway to travel from Ghent to Brussels is still feasible at this moment, as the secondary road nearby is limited to 50 km/h. At 7:40 a.m. and 8:00 a.m., although some short segments with low traffic speed can be observed (e.g. at 8:00 a.m. from Erpe-Mere to Aalst and from Affligem to Groot-Bijgaarden), the average speed of most medium-distance and long-distance segments can reach 50 km/h. At 8:20 a.m., the traffic speed from Aalst to Ternat is below 30 km.

In addition, the time that drivers need to spend on the road can also be represented in the CTM. The linear data of travel time on the road can be calculated by dividing the length of the minimum road segments by the speed detected in these segments. In this case, summation is applied to $f(I_n)$, as the travel time of any road segment is the summation of the travel time of all road segments within it. The CTM diagrams in Figure 8 display the travel time of the same motorway at the same timestamps as that in Figure 7. The colour at every point indicates the time that travellers need to spend to travel from one place to another, according to the speed along the road detected at the time stamp. A discrete colour scale with 5-min increments is applied for the ease of visual identification. From these diagrams, one can observe the time that one needs to travel between any two places. At 7:20 a.m., it takes 10 min to travel from Merelbeke (Ghent) to Erpe-Mere.

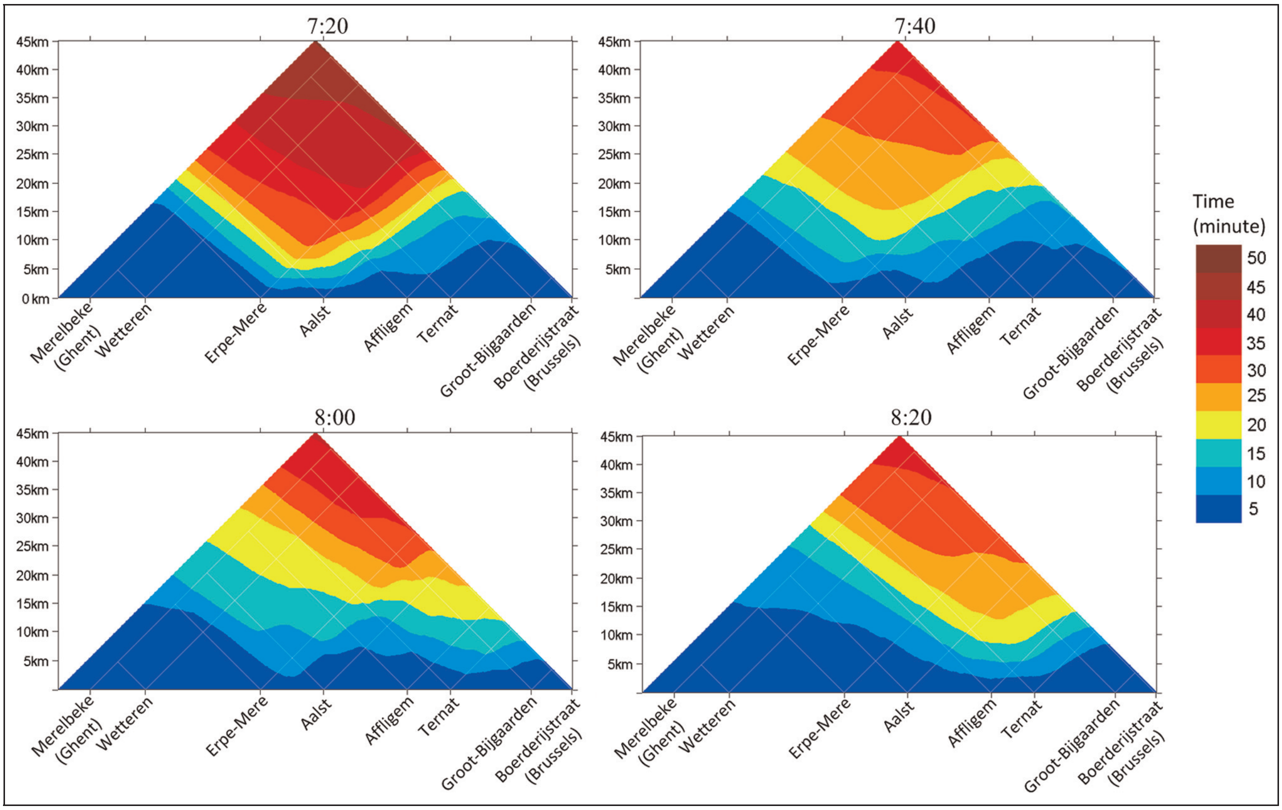

Figure 8. CTM diagrams of travel time on the E40 motorway from Ghent to Brussels.
CTM: continuous triangular model.

However, for a similar distance, it takes 30 min to travel from Erpe-Mere to Affligem. Such a difference is less obvious at 7:40 a.m., which implies that the traffic jam reduced to some degree and is distributed more evenly. At 8:00 a.m. and 8:20 a.m., it takes more time to travel through the second half than the first half of the motorway, as the red area sinks in the right part of the triangle.

Analysing multiple time series

As the CTM is based on a 2D coordinate interval spaces, many spatial analysis techniques in GIScience can be employed to analyse CTM diagrams. This section demonstrates how the methods of map algebra and cartographic modelling are used to analyse multiple time series modelled by the CTM.

Map algebra

With the traditional line chart, the comparison of multiple time series can only be made in a fixed temporal scale and partition. For example, in Figure 9, one can only compare the speed between the indoor soccer teams or players at granularity of second. The speed in other scales is hard to compare. Alternatively, using map algebra in the CTM, these time series can be compared over all different time intervals. The time series of average running speed of the two competing teams can be compared by applying the 'subtract' algebra to their CTM diagrams. In the resulting CTM diagram, every point corresponds to a specific time interval, and the value at that point is the difference of the running speed between the two teams. Figure 10 illustrates the result of subtracting the CTM of the blue team from that of the red team. Blue represents the positive value, meaning that the average speed of the blue team is greater than that of the red team. Red represents negative value, meaning that the average speed of the red team is greater. The result diagram can be interpreted as follows: in general, the blue team is more active (i.e. has greater running speed during long intervals); however, during some short time intervals, the red team has greater running speed, for instance, from the beginning to the 4th minute and from the 9th to the 13th minute. In Figure 10(b), the darkness of the colours indicates the degree of difference, which gives a better sense of the actual difference. From Figure 10(b), one can see that only during some very short intervals (less

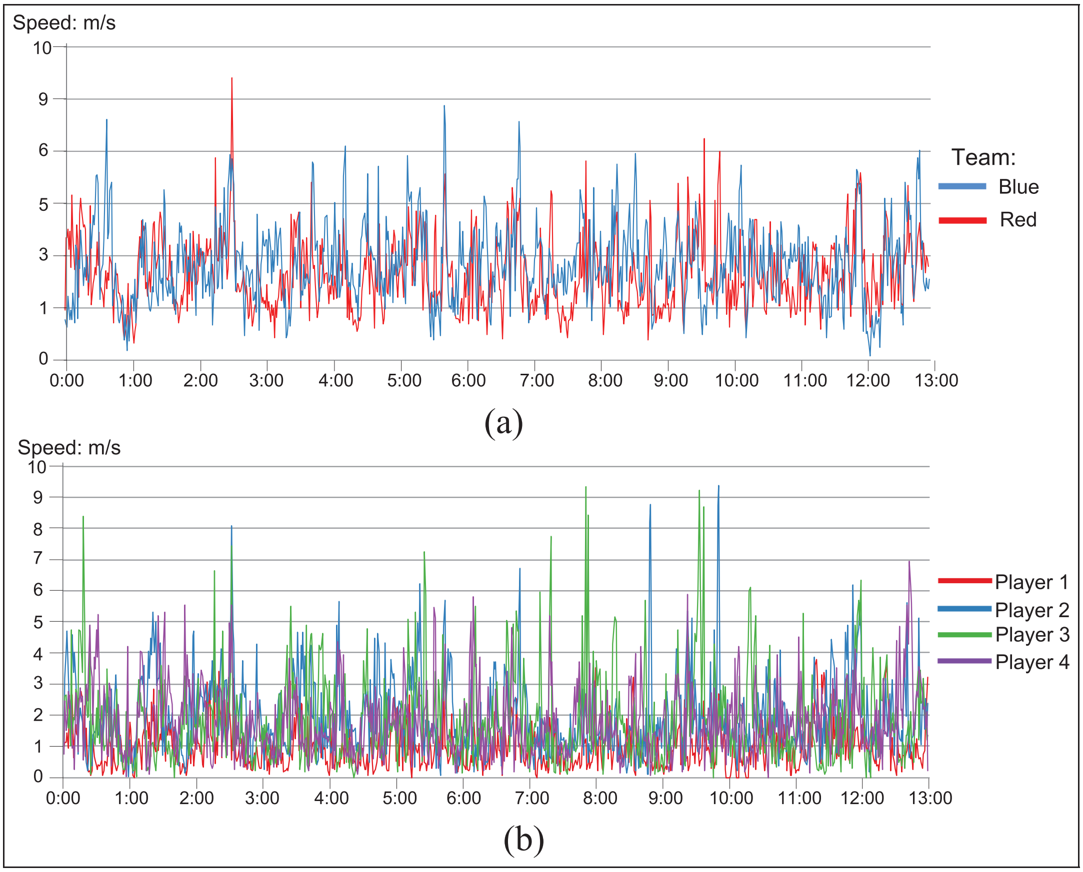

Figure 9. Line charts of multiple time series: (a) the average running speed of two competing soccer teams and (b) the running speed of individual players of the red team.

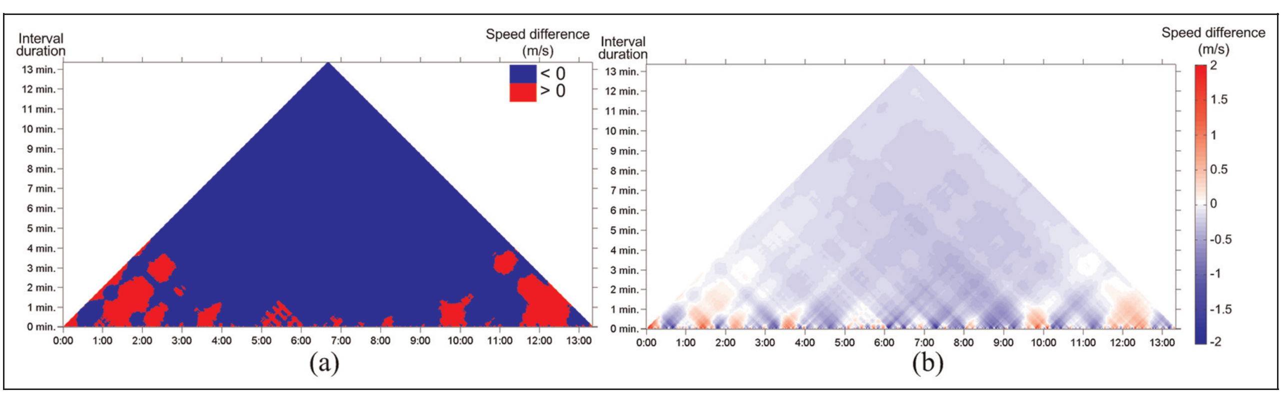

Figure 10. Output of subtracting the CTM of the blue team from that of the red team: (a) only colour hues are used to indicate the team with higher speed and (b) colour darkness is applied to indicate the speed difference.
CTM: continuous triangular model.

than 2 min), the red team is apparently more active than the blue team.

The CTM can also be used to compare more time series, for instance, the running speed of several soccer players. Here, we compare the running speed of four players in the red team, who have played through the entire study period. This can be done by combining the CTM diagrams of these four players into one CTM using the following map algebra: at every point (i.e. during every interval), the player having the

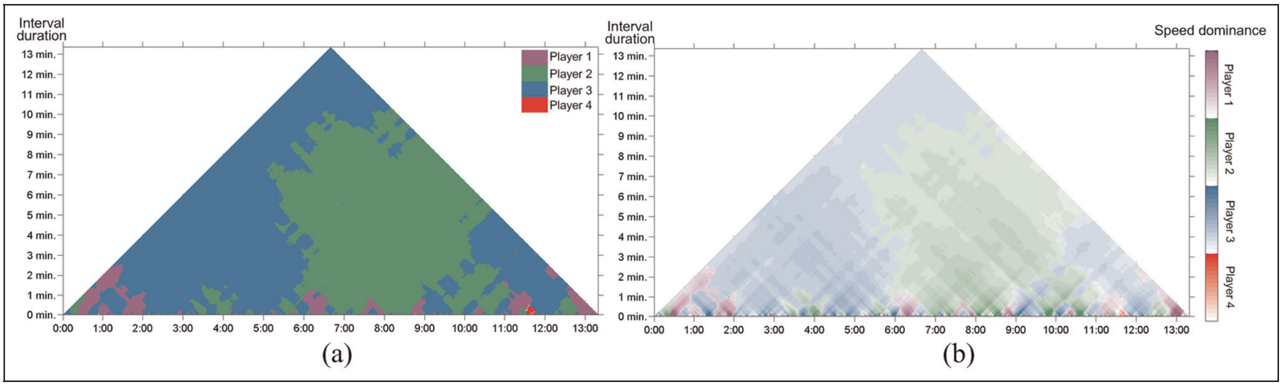

Figure 11. Comparison of multiple players: (a) dominant players are represented by discrete colour hues and (b) dominance degree is represented by darkness of colours.

greatest running speed is selected. We define this player the dominant player of the interval. If the dominant player of every interval is displayed in a specific colour hue, the output becomes a nominal diagram with four zones (i.e. Figure 11(a)). Each zone represents a set of intervals during which a certain player is dominant. In Figure 11(b), darkness of colours is used to indicate the degree of dominance. There are many ways to calculate the degree of dominance. In this case, we define the degree of dominance as the percentage that the dominant player's speed is greater than the average of the others. For example, if the speed of Player 1 is 5 and the average speed of Players 2–4 is 4, Player 1 is dominant over the others by 25%. Due to the many colour hues applied, it is a little hard to observe both the dominant players (represented by colour hues) and the degree of dominance (represented by darkness) in Figure 11(b).

This problem can be overcome by representing players dominant at certain levels in multiple CTM diagrams. In these diagrams, each colour hue represents the player dominant by a certain percentage. For example, in the top-left diagram of Figure 12, every colour indicates a player who is dominant over the others by at least 10%, which means the speed of this player is more than the average speed of the others by at least 10%. Figure 13 uses two examples to illustrate this algebra in a raster space. From these diagrams, one can see that with the increase of dominance threshold, the colour zones with low dominance gradually disappear, while the remaining zones represent the intervals during which a player is dominant above a certain level. With the dominance threshold of 30%, it becomes clear that Players 2 and 3 are much more active than the others during two successive 5-min intervals, which possibly reveals a strategy change or position shift during the game. The variation of these CTM diagrams can be better observed through a controlled animation, where the CTM diagram dynamically responses to a slider setting the difference between these CTM.

Cartographic modelling

Cartographic modelling addresses complex geographical problems by decomposing the problem into component criteria or constraints, which are usually modelled in different forms of geospatial datasets (e.g. raster, vector, or triangulated irregular network (TIN)). Through a logical sequence of operations on these geospatial datasets, the final result (normally a map) is generated, indicating the solution to the problem. One typical application of cartographic modelling is site selection (e.g. selecting the site for a vineyard or windfarm), which takes account of many geographical criteria and constraints (e.g. climate, soil, topography, and demography).[33–35] Through a series of operations, a suitability map is produced that indicates the suitable areas. This approach can also be referred to as multi-criteria decision-making analysis.[36,37] Using the same methods, cartographic modelling can also apply to CTM diagrams to solve multi-criteria decision-making problems based on time series. Analogous to geographical site selection, cartographic modelling on the CTM can help one to select time intervals that satisfy different criteria and constraints. In the CTM, these suitable intervals are represented as areas in the interval space. Next, we use a concrete scenario to explain how cartographic modelling can be applied to CTM diagrams.

Suppose several professional surfers want to select a training site for the next year. There are four candidate surfing sites, including South Africa, Hawaii, Fiji, and Australia. These four sites have different weather conditions throughout the year. Every site may be the best option during some specific periods of the year.

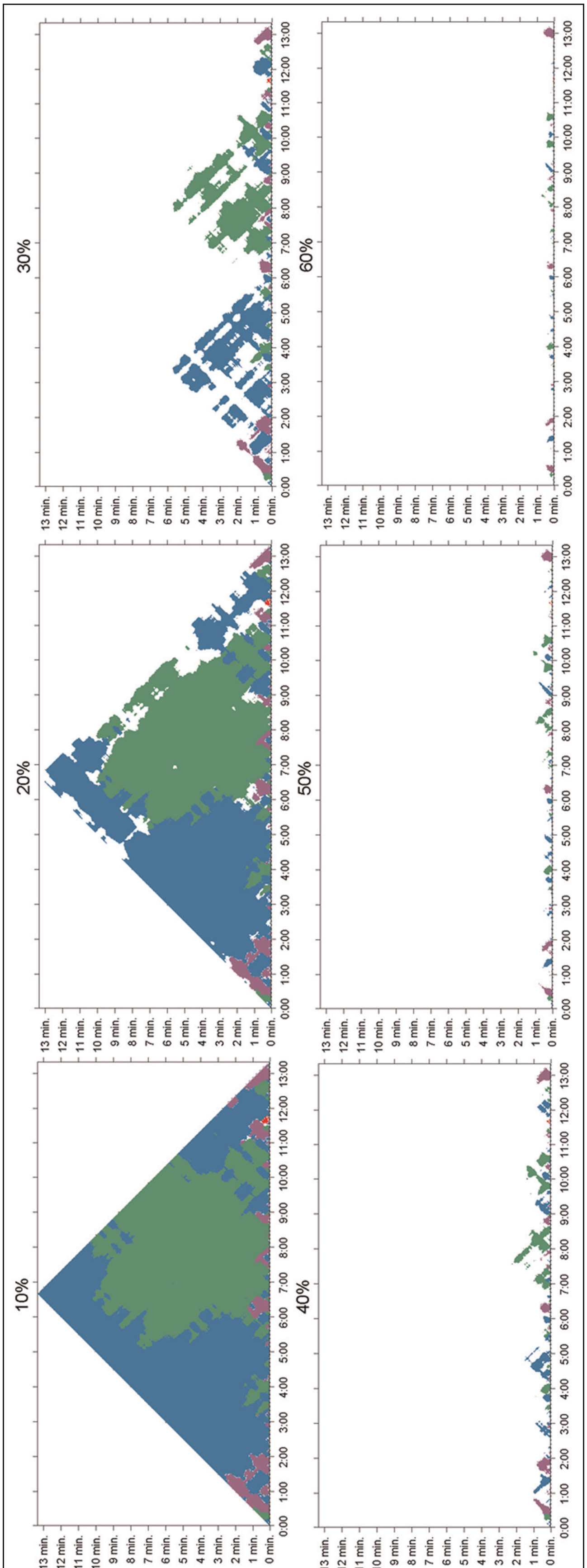

Figure 12. Players dominant at different degrees are represented in multiple diagrams. The meaning of colour hues is identical to that in Figure 11.

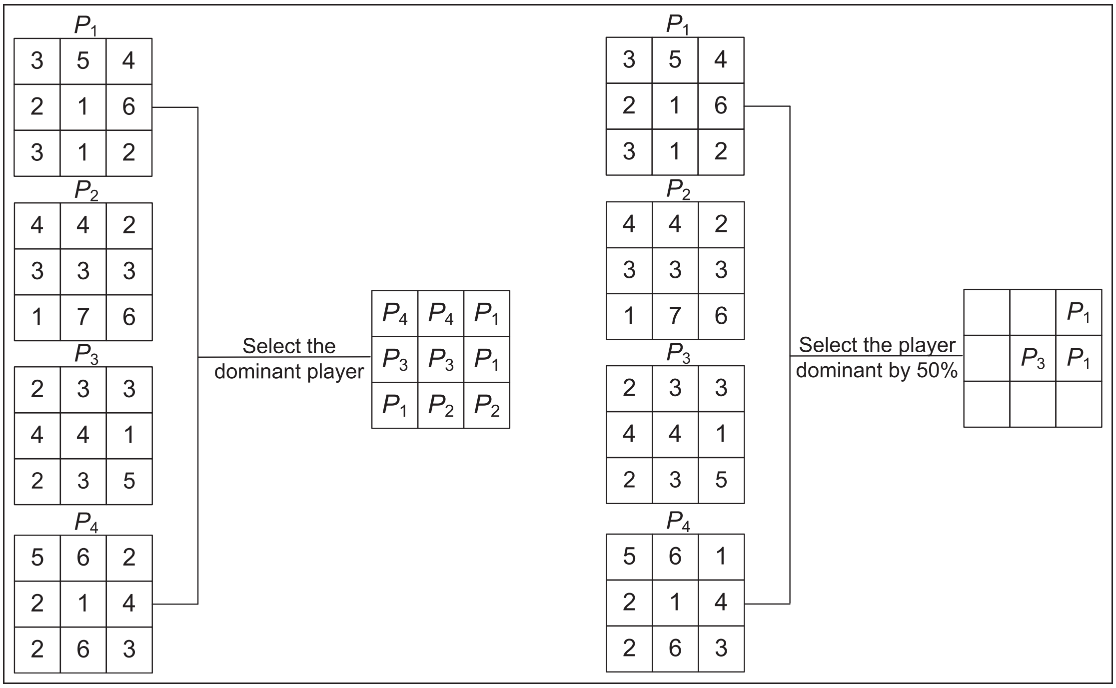

Figure 13. Select the dominant player at every point in a CTM diagram. *P* means player.
CTM: continuous triangular model.

Table 1. Preferences and constraints of surfers and available statistics of the surfing sites.

	Statistics of the site	Constraints of surfers	Preference of surfers
Ridable wave	Percentage of days that have ridable waves in every month	At least 60 days having ridable wave	The more, the better
Ground wave	Percentage of days that have ground waves in every month	N/A	The more, the better
Sea temperature	Average sea temperature in every month	N/A	The higher, the better
Avoidance	N/A	Avoiding the international tournament in August	N/A

Considering the surfers' requirements and preferences in Table 1, they need to select one of these four sites and also decide during which period they will arrange their training there. This question involves two sub-questions: which site and during which period? Given the annual weather statistics of these sites, including seasonal wave situations and sea temperature, this question is not easy to answer with traditional representations. However, by means of cartographic modelling, the CTM can give an explicit answer.

In general, there are two steps in the problem-solving procedure. First, the suitability diagram of each surfing site is created, taking account of constraints and criteria, in this case, surfers' requirements and preferences (Table 1). Second, suitability diagrams of the four sites are combined into one summary diagram to answer which site is the best during which period.

Figure 14 illustrates the specific procedure of the first step. The monthly weather conditions are rated according to the three criteria, that is, percentage of days with ridable waves (ridable wave: waves last for a period of 7 s or more (http://magicseaweed.com/)), percentage of days with ground waves (ground wave: waves last for a period of 10 s or more and over 3 ft (http://magicseaweed.com/)), and sea temperature. The ridable wave is the minimum condition on which the surfing training can be performed. The ground wave is more attractive for surfers because high-level skills can be practiced. Furthermore, warmer sea temperature increases surfing comfort. Different weights are given to these criteria based on their importance

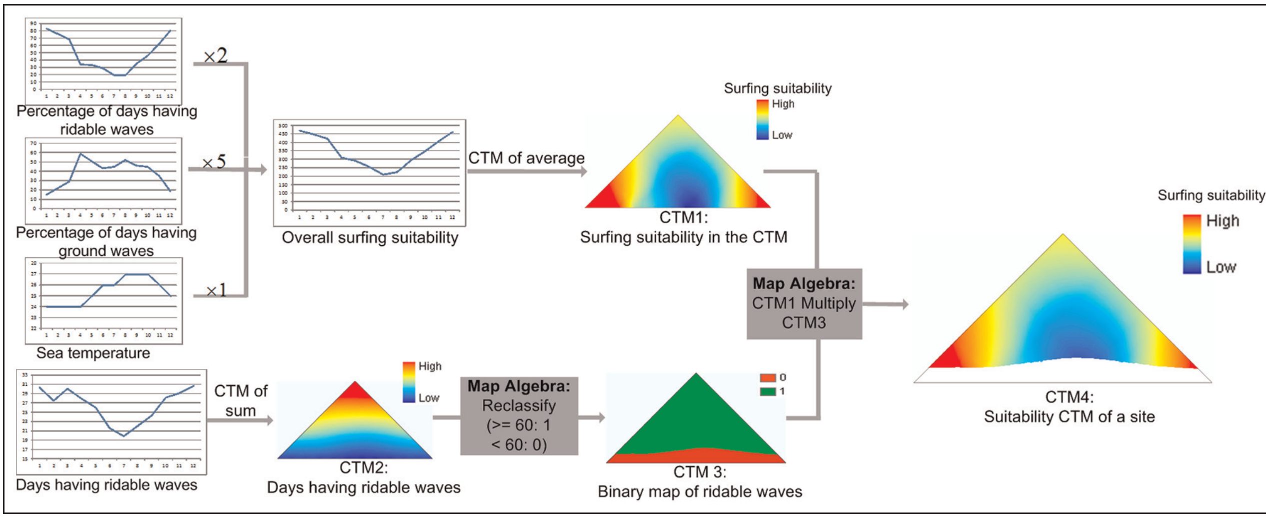

Figure 14. Procedure of generating the suitability CTM diagram for a site.
CTM: continuous triangular model.

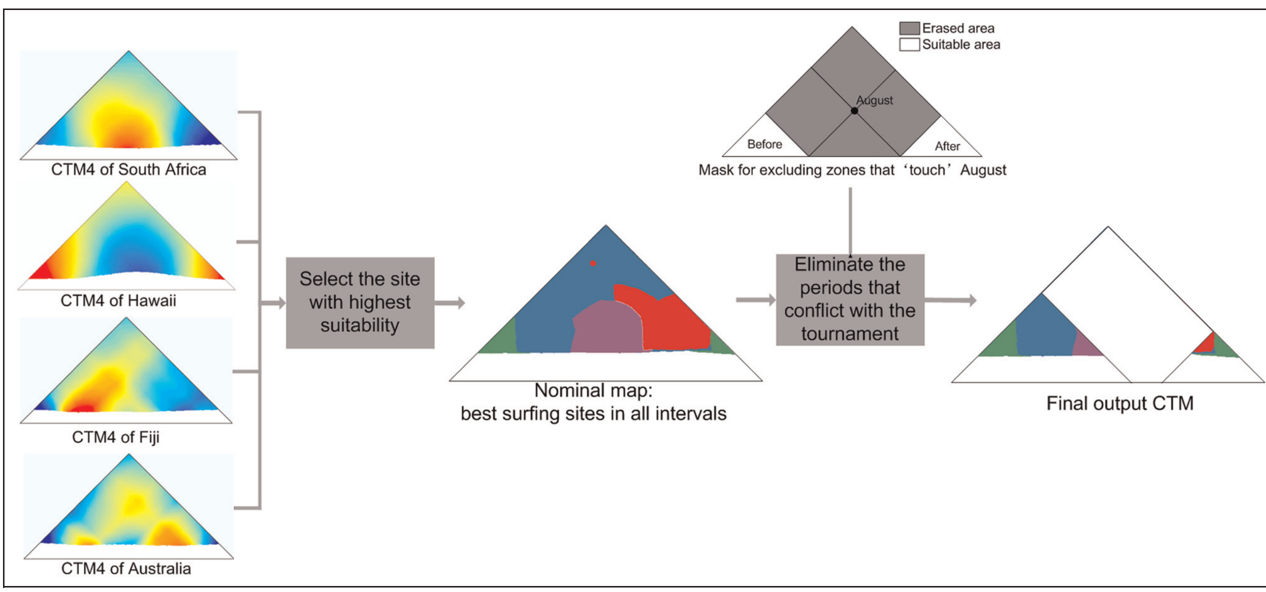

Figure 15. Composition of the suitability CTM diagrams of all sites.
CTM: continuous triangular model.

(i.e. ridable wave: 2; ground wave: 5; sea temperature: 2). Combining these weighted criteria produces a time series of suitability rates according to the general weather condition. The time series of suitability rates is represented by a CTM diagram (i.e. CTM1) with the average formula applied to $f(I_n)$. The number of days with ridable waves is represented by a CTM diagram (i.e. CTM2) with the summation formula applied to $f(I_n)$. In this case, all statistic data are recorded in a monthly scale. Therefore, the values of intervals across monthly partitions are obtained by interpolation. Considering the requirement of at least 60 days of ridable waves (Table 1), CTM2 is reclassified into a binary diagram CTM3, where values above 60 are set to one and the remaining are set to zero. Multiplying CTM1 by CTM3, we obtain the suitability diagram CTM4 of a surfing site, excluding intervals that do not have 60 days with ridable waves. Following the same procedure, the suitability diagrams of the other sites can be generated.

Figure 15 illustrates the specific procedure of the second step. The suitability diagrams (CTM4s) of all

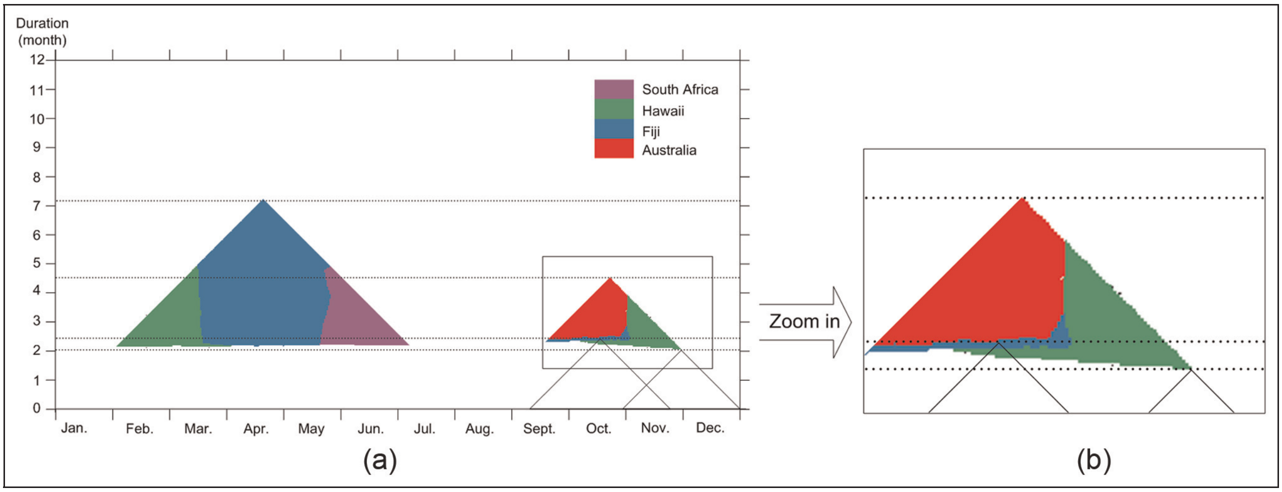

Figure 16. Enlarged result CTM diagram: (a) the overall diagram and (b) zoom into the right part of the overall diagram. CTM: continuous triangular model.

sites are combined together using the dominance algebra. At every position, the site with the highest score is selected, resulting in a nominal diagram with four zones. Each zone contains the intervals during which a certain site is the best of the four candidates. In other words, each candidate site has a set of intervals during which it is the best, which are in the corresponding zone in the CTM diagram. To avoid the international tournament in August, only intervals before and after August can be used for training, and thus, all other intervals have to be excluded. Referring to the relational zones of the TM in Figure 3, relational zones that 'touch' August have been erased. Only intervals in the *before* and *after* zones are suitable. After all operations, the final output is produced, that is, the rightmost CTM diagram in Figure 15.

The final output is enlarged in Figure 16. From this diagram, one may have an idea which site is the best candidate during which intervals. Instead of offering a fixed choice, this diagram presents all possible intervals through the entire year. This diagram is flexible enough for surfers to make rough plans and remains open for modifications caused by other constraints. The output CTM is a nominal diagram with four distinctive zones, which can visually answer the surfer's question.

According to the coordinate system of the CTM (described in section 'CTM'), the result in Figure 16 can be interpreted. At the beginning of the year, Hawaii is the best site for surfing, while Fiji and South Africa are the best in the spring and summer, respectively, before the tournament in August. In autumn and early winter (the right part of Figure 16(a)), the situation is more complex. There is a thin blue slice extending from September to November, which means that between September and November, Fiji is the best surfing site for approximately 2.5 months. If the period is longer or shorter than 2.5 months, either Australia (red) or Hawaii (green) is the best choice. Furthermore, if surfers would like to stay for the shortest period that guarantees 60 days with ridable waves, Hawaii is the best place, because the lowest position is in the green zone, at the right corner of the suitable area. In November and December, surfers only need to stay in Hawaii for just over 2 months to get 60 days with ridable waves. Moreover, if surfers would like to stay at one site as long as possible, Fiji is the best choice, because the highest point is in blue. This means that Fiji has the best average weather condition during the period from the beginning of the year to the tournament start. After August, Australia is the best site for a long stay, because it has the highest rate during the period between August and the end of the year.

Moreover, one can increase the dominance threshold to screen out the periods during which a surfing site is more suitable than the others to a certain degree. Different from the analysis of soccer players in section 'Map algebra', in this case, the dominance is the percentage that the highest score is over the second highest score. From Figure 17, one can see that with the increase of dominance threshold, long intervals gradually disappear. When the threshold has increased to 16%, the remaining intervals have a duration between 2 and 3 months. This means that if the surfer wants to select a surfing site that is more suitable than the others by at least 16%, his staying period should be 2–3 months.

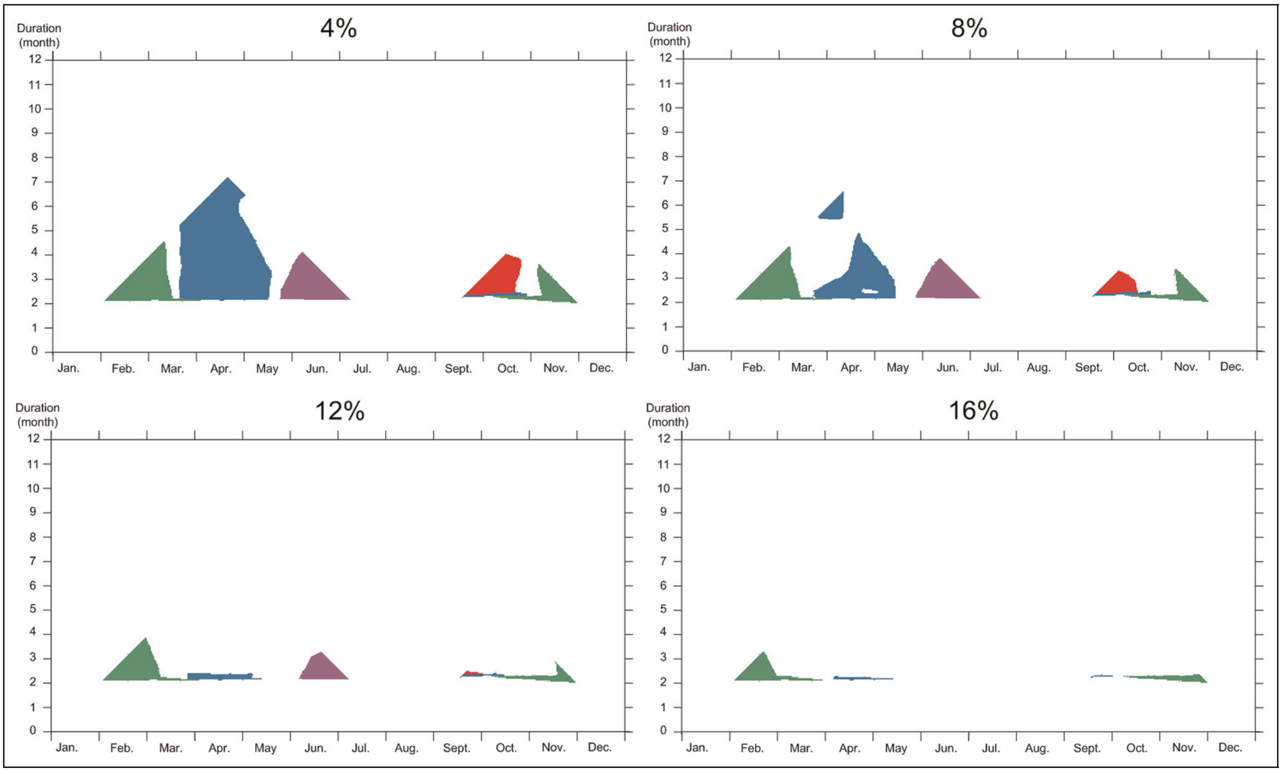

Figure 17. Result CTM based on different dominance thresholds. The meaning of colours is identical to that of Figure 16. CTM: continuous triangular model.

Compared with the traditional approaches, the CTM-based cartographical modelling has two merits. First, the CTM is not limited to a certain timescale or partition. It offers multi-scale suggestions for choosing suitable sites and periods. Second, the answer is presented in a structured diagram, which is more informative and perceivable than text, table, and other forms of representations. One can visually identify the suitable intervals and observe how they distribute in a 2D space. In this surfing site selection case, a simple set of criteria and constraints are modelled by the CTM. When it is applied to more complex cases, other time-related criteria can be added. Also, the weights of criteria can be adjusted according to specific requirements.

Discussion

Due to the increasing availability of tools for visual analysis, many scientists have been working on building a theoretical framework to generalize the data and tasks the different tools aim to tackle. To ground our contribution in the area of visual analytics, this section attempts to elaborate the tasks that our approach is able to solve, according to an established framework. There are some remarkable studies about such a theoretical framework,[38–40] each of which has strength in particular aspects. We choose the Andrienkos approach[6] due to its generalization, comprehensive coverage of updated techniques and sufficient consideration about temporal data. This approach distinguishes two components of data, that is, the referrers and attributes, which can also be called independent and dependent variables. For example, in geographical datasets, the geographical locations are referrers, while the measurements of a location (e.g. land cover and elevation) are attributes. Values of referrers are called references, and values of attributes are called characteristics. The CTM specially deals with data that have linearly arranged referrers. As previously mentioned, such referrers can be the timeline, a road, a river, or even a section of text. Although the attributes of data used in this study are quantitative (i.e. ratio and interval), the CTM is applicable to qualitative attributes as long as corresponding operators are available. Andrienko's approach separates elementary and synoptic tasks. The former addresses individual elements of data, while the latter deals with the whole reference set or its subsets. The specific explanation of the task categorization is documented in their study.[6] In general, the CTM is applicable to both levels of tasks. At the elementary level, the references (i.e.

intervals) and characteristics (attributes of intervals) of individual elements are expressed by positions and colours, respectively, so that one can easily look up, identify, and compare data elements in a CTM diagram. Using map algebra, several datasets sharing the same linear axis can be parallel compared, such as the comparison of movement of two soccer teams in Figure 10. Moreover, temporal constraints can be modelled as zones in the CTM, which can be used as masks to screen out intervals satisfying certain temporal relation. Conversely, intervals satisfying certain attribute constraints can be selected with the support of attribute queries, which is widely implemented in many geographic information system (GIS) tools. This kind of tasks can also be referred to as relation-seeking task.[6] As a CTM diagram presents all possible references within the study frame, the abovementioned tasks can cover references in all different granularities. On the other hand, all these tasks can also be performed at the synoptic level, which then become lookup, comparison, and relation-seeking of patterns. The patterns of characteristics and references are expressed by colour patterns and spatial patterns of points or zones in the CTM, respectively. The special feature of the CTM is that patterns can be identified, characterized, and compared in all different scales. Also, an explicit hierarchy of patterns can be displayed, from which one can perceive the interaction between small and large patterns. While some of these pattern analyses tasks may rely on image operations or manipulations (e.g. map algebra, animation, or small multiples), we believe that a visualization system based on the CTM will have comprehensive coverage of analysis tasks of linear data, which can be even enhanced by its multi-scale capability.

Conclusion and future study

This article introduced an innovative representation of linear data, namely the CTM. In the CTM, the linear data in different intervals are displayed in a two-dimensional space, constituting a basis for a multi-scale analysis of linear data. In general, the CTM has two major advantages. First, it provides an explicit and compact visualization of linear data in different scales. In the CTM, moving statistics (e.g. the average and summation) during intervals of different lengths can be displayed in one diagram, which offers an explicit overview of patterns in different scales. Compared with traditional multi-scale visualization approaches, which only display data in a few selected scales, the CTM can present the data in all scales. This feature is particularly useful for the exploration of unfamiliar datasets, in which interesting patterns may emerge in any scales. Also, the CTM offers an overview of the hierarchy of scales, which allows the observation of the interaction between large-scale patterns and small-scale patterns. Second, the CTM is based on a universal 2D coordinate space, which is very similar to prevalent geospatial datasets. This study demonstrates how existing techniques in GIScience can be used to manipulate and analyse CTM diagrams. By applying map algebra to the CTM, multiple time series can be compared at different scales. The CTM can use a single diagram to present the answer to the questions such as 'whether Brussels is warmer than New York' or 'whether Player 2 is more active than the other players' according to all possible intervals within the considered time frame. We contend that the CTM representation of these answers is more informative and perceivable than traditional approaches such as line chart or colour lines. With the support of special visualization techniques (e.g. setting a dominance threshold), it can have an extensive coverage of different analysis tasks. Furthermore, we demonstrated the application of cartographical modelling in the CTM for multi-criteria decision-making analysis. Temporal criteria and constraints can be modelled as areas in CTM diagrams, which can be interpreted as sets of intervals that are in a certain temporal relation or attribute range. These areas can be used as masks to screen the intervals that satisfy the constraints. Through a sequence of map algebra on CTM diagrams, one can obtain all intervals of different lengths that meet their criteria and constraints. Such all-scale analysis is difficult with the traditional representations where analysis is usually performed on intervals of the same length. Another interesting feature is that the whole procedure is based on operations of diagrams, which offers a visual impression of every step of the analysis process and intermediate outputs.

This article gave several examples to show possible applications of the CTM. In future study, there are many possible scenarios in which the CTM can be applied. For instance, we plan to extend the use of the CTM to standard soccer games to analyse the time series of the performance and physical conditions of soccer players in a 90-min time frame. The visualization in the CTM may support the coach to make strategies or adjustments accordingly in all different time intervals. Not only football games, the analysis of cyclists or Marathon runners during a competition may also advice the strategy making. Thus, we believe that sport analysis may constitute a potential market of the CTM. The idea of cartographic modelling can be used to support decision-making analysis based on different types of time-dependent data. For instance, it can be applied to agriculture and farming, which heavily rely on the analysis of time series of weather statistics. Similar to the demonstrated scenario of surfing site selection, the CTM-based analysis can potentially provide advices for the scheduling of agricultural activities such as irrigation or fish egg incubation. Besides these, there are many other areas, such as economic analysis, climate change research, and traffic analysis, where the CTM could be applied to.

In view of the bulking data volume faced by many disciplines, using the CTM to analyse a large number of linear datasets is another interesting direction. Instead of individually displaying every linear dataset, those with a certain degree of similarity can be clustered. Due to the special feature of the CTM, that is, a 2D representation built up from 1D data, special clustering technique needs to be developed to cluster CTM diagrams. Moreover, implementing the CTM into an interactive system can improve the usability of the CTM. On the one hand, interactive controls and dynamic linkage with other visualization components may assist users to comprehend and interpret the detected patterns. On the other hand, such a system can even magnify the analytical power of the CTM. Currently, a prototype tool with basic functionalities of loading/exporting data, animating multiple time series, setting colour ramp, and running map algebra has been developed for the experimental use. In future versions, operations of the CTM diagrams should be performed in minimal time, so that users can perceive the reaction of the CTM to the change of particular settings. For instance, dynamic controls for colour ramps, the formulas in the CTM, and area selection by attribute can facilitate visual exploration and analysis in the CTM. The sensitivity of the cartographic modelling approach to different parameters and utility functions can be better analysed within this system. Integrating the CTM to other information systems can also be interesting. For example, linking the CTM with a GIS can enhance the analysis of time series with geographical references. At the moment, available empirical evidence has pointed out the advantages of the TM in visualizing discrete time intervals.[41] Our next step is to quantitatively evaluate the usability of the CTM and its implementations before introducing them to a broader range of users. Not only limited to linear data, the transformation of the CTM can be used to represent cyclic data. From a broader perspective, the CTM can be considered as a conceptual model of generalization. Many types of data can be described at different granularities, for instance, the border of a country can be displayed at different granularities, and a system can be decomposed into different levels. These types of data can also be plotted into a CTM diagram, which could potentially benefit the analysis. These ideas will be further investigated in future studies.

Funding

This study was supported by the Research Foundation Flanders (FWO) of Belgium and the Special Research Fund (BOF) of Ghent University.

References

1. Enders W. *Applied econometric time series*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc, 2008, p. 480.
2. Hamilton JD. *Time series analysis*. Princeton, NJ: Princeton University Press, 1994, p. 799.
3. Liao TW. Clustering of time series data: a survey. *Pattern Recogn* 2005; 38(11): 1857–1874.
4. Andrienko G, Andrienko N, Demsar U, et al. Space, time and visual analytics. *Int J Geogr Inf Sci* 2010; 24(10): 1577–1600.
5. Aigner W, Miksch S, Muller W, et al. Visualizing time-oriented data – a systematic view. *Comput Graph-UK* 2007; 31(3): 401–409.
6. Andrienko N and Andrienko G. *Exploratory analysis of spatial and temporal data: a systematic approach*. Heidelberg: Springer-Verlag Berlin, 2006.
7. Keim D, Mansmann F, Schneidewind J, et al. Challenges in visual data analysis. In: *Proceedings of the 10th international conference on information visualization*, London, UK, 5–7 July 2006, pp. 9–16. IEEE Computer Society, Washington, DC, USA.
8. Havre S, Hetzler B and Nowell L. ThemeRiver: visualizing theme changes over time. In: *Proceedings of the IEEE symposium on information vizualization 2000*, Salt Lake City, UT, 9–10 October 2000, pp. 115–123. IEEE Computer Society, Washington, DC, USA.
9. Hochheiser H and Shneiderman B. Dynamic query tools for time series data sets: timebox widgets for interactive exploration. *Inform Visual* 2004; 3(1): 1–18.
10. Lin J, Keogh E, Lonardi S, et al. Visually mining and monitoring massive time series. In: *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining*, Seattle, WA, 22–25 August 2004, pp. 460–469. ACM, New York, USA.
11. Weber M, Alexa M and Muller W. Visualizing time-series on spirals. In: *Proceedings of IEEE symposium on information visualization 2001*, San Diego, CA, 22–23 October 2001, pp. 7–13. IEEE Computer Society Press, Washington, DC, USA.
12. Kulpa Z. Diagrammatic representation of interval space in proving theorems about interval relations. *Reliab Comput* 1997; 3(3): 209–217.
13. Kulpa Z. A diagrammatic approach to investigate interval relations. *J Visual Lang Comput* 2006; 17(5): 466–502.
14. Van de Weghe N, Docter R, de Maeyer P, et al. The triangular model as an instrument for visualising and analysing residuality. *J Archaeol Sci* 2007; 34(4): 649–655.
15. Qiang Y, Delafontaine M, Asmussen K, et al. Modelling imperfect time intervals in a two-dimensional space. *Control Cybern* 2010; 39(4): 983–1010.
16. Qiang Y, Delafontaine M, Neutens T, et al. Analysing imperfect temporal information in GIS using the triangular model. *Cartogr J* 2012; 49(3): 265–280.

17. Qiang Y, Delafontaine M, Versichele M, et al. Interactive analysis of time interval in a two-dimensional space. *Inform Visual* 2012; 11(4): 255–272.
18. De Smith MJ, Goodchild MF and Longley PA. *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Leicester: Troubador Publishing Limited, 2007.
19. Goodchild MF, Yuan M and Cova TJ. Towards a general theory of geographic representation in GIS. *Int J Geogr Inf Sci* 2007; 21(3): 239–260.
20. Van Wijk JJ and van Selow ER. Cluster and calendar-based visualization of time series data. In: *Proceedings of IEEE symposium of information visualization (InfoVis '99)* (eds G Wills and D Keim), San Francisco, CA, 24–29 October 1999, pp. 4–9. IEEE Computer Society, Washington, DC, USA.
21. Kincaid R and Lam H. Line graph explorer: scalable display of line graphs using focus + context. In: *Proceedings of the working conference on advanced visual interfaces*, Venezia, Italy, 23–26 May 2006, pp. 401–411. ACM, New York, USA.
22. Matkovic K, Gracanin D, Konyha Z, et al. Color lines view: an approach to visualization of families of function graphs. In: *11th international conference on information visualization, 2007 IV '07*, Zurich, Switzerland, 2–6 July 2007, pp. 59–64. IEEE Computer Society, Washington, DC, USA.
23. Saito T, Miyamura HN, Yamamoto M, et al. Two-tone pseudo coloring: compact visualization for one-dimensional data. In: *Proceedings of the 2005 IEEE symposium on information visualization*, Minneapolis, MN, 23–25 October 2005, pp. 173–180. IEEE Computer Society, Washington, DC, USA.
24. Carlis JV and Konstan JA. Interactive visualization of serial periodic data. In: *Proceedings of the 11th annual ACM symposium on user interface software and technology*, San Francisco, CA, 1–4 November 1998. ACM, Washington, DC, USA.
25. Li X and Kraak M-J. The time wave. A new method of visual exploration of geo-data in time–space. *Cartogr J* 2008; 45(3): 193–200.
26. Tominski C, Abello J and Schumann H. Axes-based visualizations with radial layouts. In: *Proceedings of the 2004 ACM symposium on applied computing*, Nicosia, Cyprus, 14–17 March 2004, pp. 1242–1247. ACM, New York, USA.
27. Muller W and Schumann H. Visualization methods for time-dependent data – an overview. In: *Proceedings of the 2003 winter simulation conference*, New Orleans, LA, 7–10 December 2003, pp. 737–745. ACM, New York, USA.
28. Aigner W, Miksch S, Schumann H, et al. *Visualization of time-oriented data*. London: Springer-Verlag, 2011.
29. Keim D, Nietzschmann T, Schelwies N, et al. A spectral visualization system for analyzing financial time series data. In: *Proceedings of the eurographics/IEEE-VGTC symposium on visualization (EuroVis 2006)* (eds T Ertl, K Joy and B Santos), Lisbon, Portugal, 8–10 May 2006, pp. 195–200. Eurographics Association, Geneva, Switzerland.
30. Tomlin CD. *Geographic information systems and cartographic modeling*. NJ: Prentice Hall, 1990.
31. Kulpa Z. Diagrammatic representation for a space of intervals. *Mach Graph Vis* 1997; 6: 5–24.
32. Allen JF. Maintaining knowledge about temporal intervals. *Commun ACM* 1983; 26(11): 832–843.
33. Collins MG, Steiner FR and Rushman MJ. Land-use suitability analysis in the United States: historical development and promising technological achievements. *Environ Manage* 2001; 28(5): 611–621.
34. Smith L. Site selection for establishment & management of vineyards. In: *The 14th annual colloquium of the spatial information research centre University of Otago*, Dunedin, New Zealand, 3–5 December 2002.
35. Walsh SJ, Butler DR, Brown DG, et al. Cartographic modelling of snow avalanche path location within Glacier-National-Park, Montana. *Photogramm Eng Rem S* 1990; 56(5): 615–621.
36. Jankowski P, Andrienko N and Andrienko G. Map-centred exploratory approach to multiple criteria spatial decision making. *Int J Geogr Inf Sci* 2001; 15(2): 101–127.
37. Jankowski P. Integrating geographical information systems and multiple criteria decision-making methods. *Int J Geogr Inf Syst* 1995; 9(3): 251–273.
38. Bertin J. *Semiology of graphics: diagrams, networks, maps*. Madison, WI: The University of Wisconsin Press, 1983.
39. Koussoulakou A and Kraak MJ. Spatia-temporal maps and cartographic communication. *Cartogr J* 1992; 29(2): 101–108.
40. Yi JS, Kang YA, Stasko JT, et al. Toward a deeper understanding of the role of interaction in information visualization. *IEEE T Vis Comput Gr* 2007; 13(6): 1224–1231.
41. Qiang Y, Valcke M, van de Weghe N, et al. Representing time intervals in a two-dimensional space: an empirical study. *Spat Cogn Comput* (in press).