

Published as: Qiang, Y., Chavoshi, S.H., Logghe, S., De Maeyer, P., Van de Weghe, N. (2014), Multi-scale analysis of linear data in a two-dimensional space. *Information Visualization*, vol. 13 (3): 248-265.

# MULTI-SCALE ANALYSIS OF LINEAR DATA IN A TWO-DIMENSIONAL SPACE

## 1 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<sup>1-3</sup>. 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 at a certain scale cannot be detected at other scales. Even at the same scale, different partitions of intervals may result in different patterns being revealed. On the other hand, a question can be answered at different scales. For example, the answers to the question when there are a lot of traffic jams in Belgium may include ‘between 7:00am and 9:00am’, ‘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 at 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 at different scales can also be important in certain analytical tasks. 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.

Visualisation has been proven to be an effective analytical approach for time series data <sup>4, 5</sup>. An explicit visualization can effectively combine the insight of humans and processing ability of computers <sup>6, 7</sup> to tackle analysis tasks. While a number of approaches have been developed to visualise time series <sup>8-11</sup>, the line chart remains the most frequently used. In a line chart, the horizontal dimension indicates positions in the time line, and the vertical dimension indicates the values at the positions. 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 at a certain scale. Displaying time series at 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 at all different scales.

The Continuous Triangular Model (CTM) provides an alternative approach to represent time series and overcomes the difficulty of traditional approaches in visualising time series at multiple scales. The CTM is based on a diagrammatic representation of time intervals initially proposed by Kulpa <sup>12, 13</sup>. Later, Van de Weghe named it the Triangular Model (TM) and applied it to archaeological use cases <sup>14</sup>. More recently, Qiang investigated its use in reasoning imperfect intervals <sup>15</sup> and visual analytics <sup>16, 17</sup>. 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 etc. On the one hand, the CTM can provide an overview of linear time series at 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) are now open to be employed to manipulate and analyse the CTM data <sup>18, 19</sup>. 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 runoff along a river, or objects with a linear structure, such as voltage along power lines and DNA sequences.

An idea similar to the CTM is the Growth Matrix introduced by Keim <sup>20</sup>, which visualises time series of stock prices 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 work 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). Moreover, we show how the methods of map algebra and cartographic modelling<sup>21</sup> can be applied to the CTM to solve multi-criteria decision-making problems based on time series.

In the remainder of this paper, we first review the basic concept of the TM. Next, the concept of the CTM is introduced. We then demonstrate how the CTM can be applied to visualise 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, conclusions are drawn and directions of future work are proposed.

## 2 BASIC CONCEPTS

### 2.1 Triangular Model

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  $I^-$ ;  $L_2$  through  $I^+$  (Figure 1).  $\alpha_1$  is the angle between  $L_1$  and the horizontal axis, while  $\alpha_2$  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  $I$ , i.e.,  $mid(I) = (I^- + I^+)/2$ , while the vertical position indicates the duration of  $I$ , i.e.,  $dur(I) = \tan \alpha \cdot (I^+ - I^-)/2$ . 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 Triangular Model (TM). The angle  $\alpha$  is a pre-defined constant that is identical to the construction of all interval points. Here, we set  $\alpha = 45^\circ$  to be consistent with previous work<sup>12, 13, 15</sup>, though  $\alpha$  can be set to any value between  $0^\circ$  and  $90^\circ$  for specific purposes. In the TM, every time interval can be represented as a unique point in the 2D space. The 2D space where interval points are located in is called the Interval Space ( $I\mathbb{R}$ ).

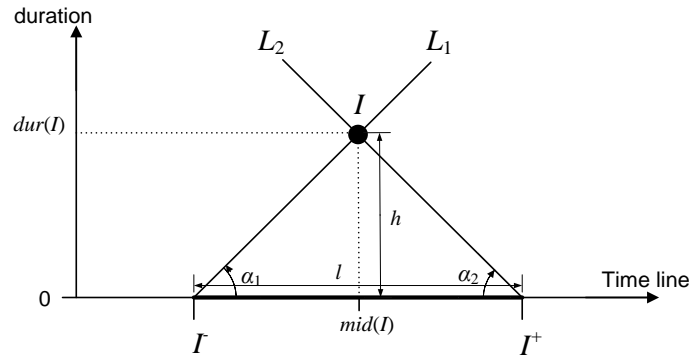


Figure 1. The configuration of the Triangular Model (TM).

According to the interval algebra introduced by Allen<sup>22</sup>, there exist thirteen atomic relations between two time intervals. In the TM, the intervals in a certain temporal relation to a reference interval  $I$  are located in a specific zone in  $I\mathbb{R}^{12,13}$ . Figure 2 illustrates the zones of the thirteen Allen relations. In Figure 2, it is assumed that there exists a triangular study area that contains all intervals, and the referenced interval  $I$  is in the centre of the study area. The black zones contain the sets of intervals in certain temporal relations to the referenced interval. For instance, the black zone in the left corner of the study area represents the set of intervals that are *before*  $I$ , which is denoted as *before*( $I$ ). Temporal constraints based on Allen 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 one of the several constraints is located in the union of the zones. For a more detailed description of the relational zones and their compositions please refer to 15.

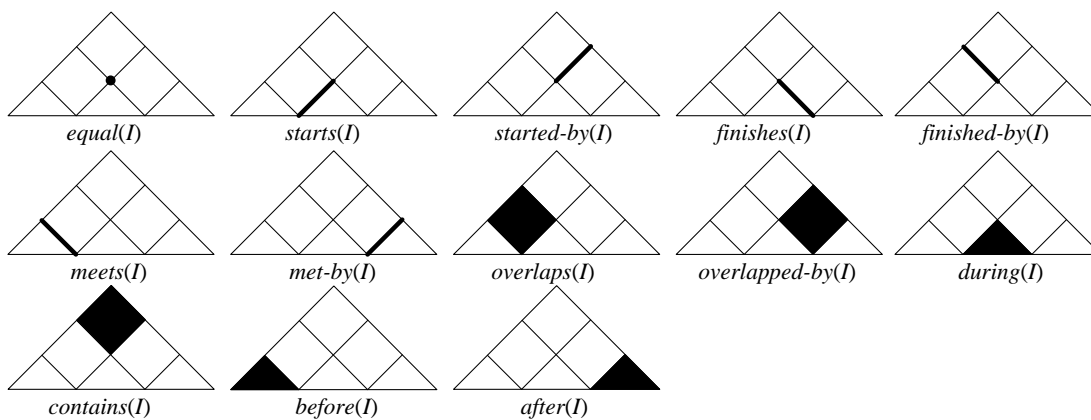


Figure 2: The representation of the thirteen Allen relations in the TM.

## 2.2 Continuous Triangular Model

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 2). 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 sub-interval  $I_n$  of the linear data. If every point is assigned a certain value, i.e.  $f(I_n)$ , 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 3 illustrates how the CTM is 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 assigned a number that is calculated from the numbers within the subinterval. Here the granularity of the CTM is consistent to that of the linear data sequence. Finer granularity can be obtained through interpolation. Figure 4 gives an example of the implementation of the CTM in a raster space. Through colour-coding, the CTM can be displayed as an image.

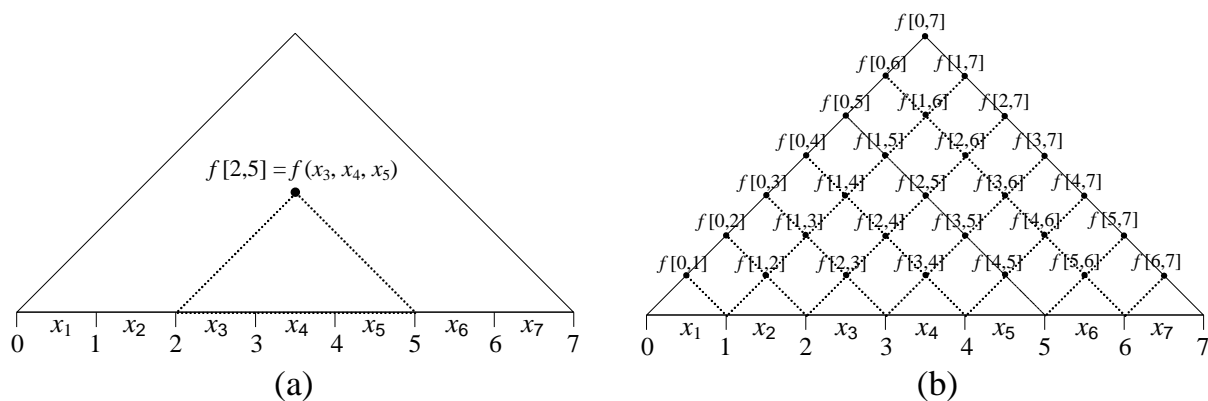


Figure 3: Representing a linear data sequence with seven numbers in the CTM. (a): A point is assigned a number calculated from the numbers within a subinterval. (b): Every point in the triangular space is assigned a number of a specific subinterval.

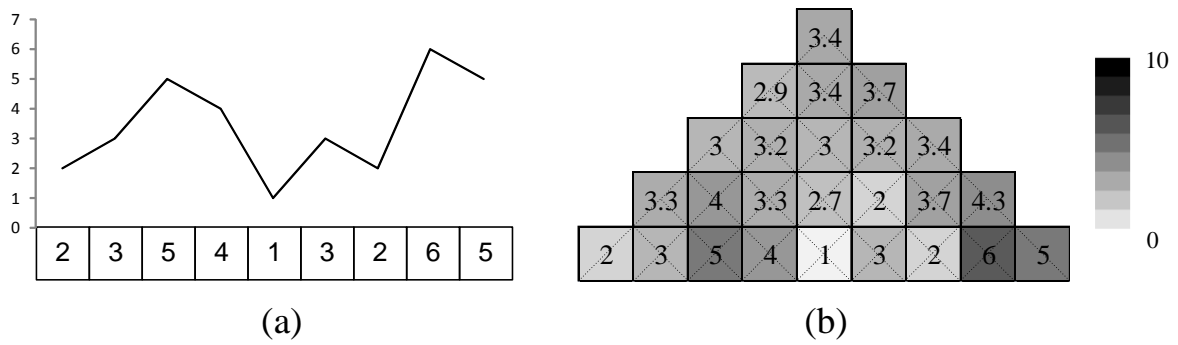


Figure 4: The implementation of the CTM in a raster space. (a): A linear data sequence and its representation in a line chart. (b): The CTM representation of the linear data sequence in (a), with the average formula applied.

### 3 VISUALIZATION OF LINEAR DATA

#### 3.1 Visualizing Time Series

This sub-section 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 5 players. The time series is a player’s speed in every second (i.e. meter/second or m/s) during a study interval of 800 seconds. As indoor soccer is rather intensive and fast-paced, the line chart (i.e. Figure 5 (a)) exhibits dramatic changes of speed from second to second. However, variations in longer intervals (e.g. one minute or two minutes) are hard to observe. In Figure 5 (b), the time series of the player’s speed is represented by the CTM, in which  $f(I_n)$  is the average of the player’s 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 at all different scales, in which one can observe the relationship between the short-term variations and long-term variations. From this diagram, one can identify intervals of sprint from the red areas on the bottom of the CTM, e.g.  $I_1$ ,  $I_2$ , and  $I_3$ . At 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 minutes (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 <sup>20</sup>, in which  $I^-$  and  $I^+$  are respectively coordinated along the vertical and horizontal axes, 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.

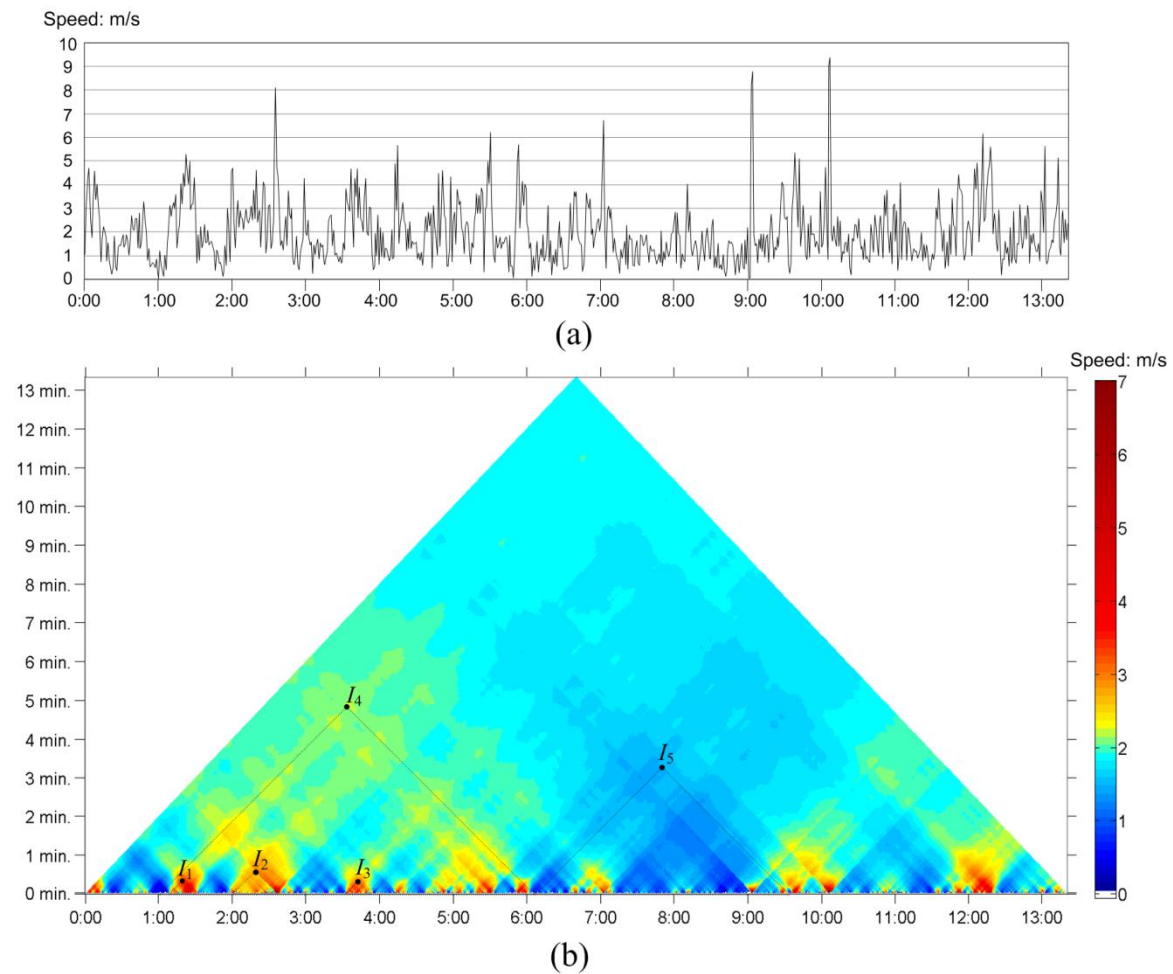


Figure 5: The time series of a soccer player’s speed in a line diagram (a) and the CTM (b).

### 3.2 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 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 Figure 6, 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 is always less than 1 kilometre, a single tick is used to mark both the exit to and entrance from the same place. Lines that are in  $\alpha$  and  $-\alpha$  to the horizontal axis are drawn from these ticks. The average

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 at 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 6, one can observe that the average speed from Erpe-Mere to Ternat at 7:20 has fallen below 20 km/hour (the 24 hour clock is applied in this paper), 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/hour. 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/hour. At 7:40 and 8:00, although some short segments with low traffic speed can be observed (e.g., at 8:00 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/hour. At 8:20, 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 7 display the travel time of the same motorway at the same time stamps as that in Figure 6. 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-minute 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, it takes 10 minutes to travel from Merelbeke (Ghent) to Erpe-Mere. However, for a similar distance, it takes 30 minutes to travel from Erpe-Mere to Affligem. Such a difference is less obvious at 7:40, which implies that the traffic jam reduced to some degree and is distributed more evenly. At



8:00 and 8:20, 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.

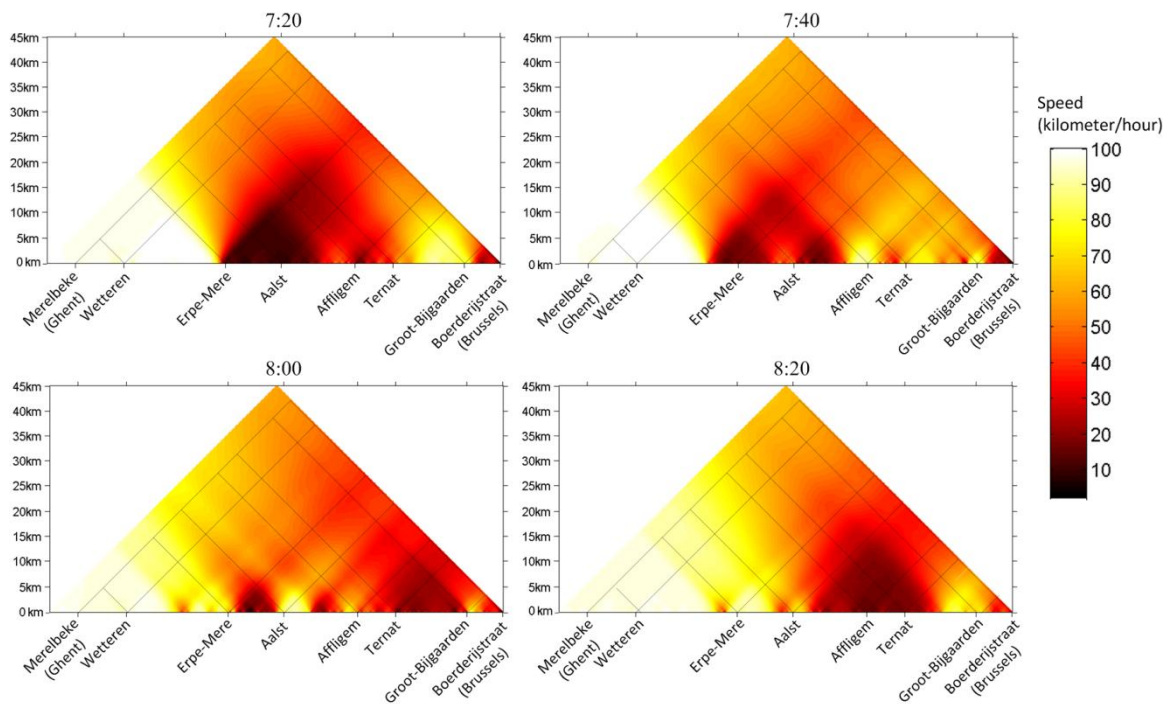


Figure 6: CTM diagrams of the average speed on the E40 motorway from Ghent to Brussels.

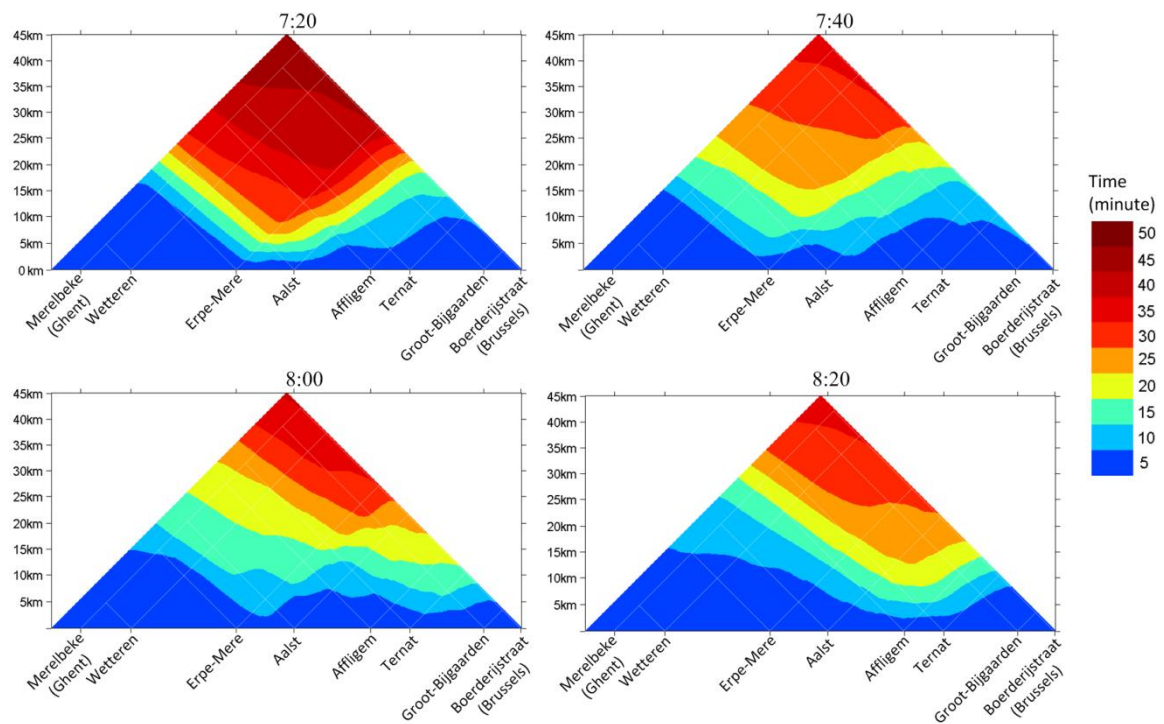


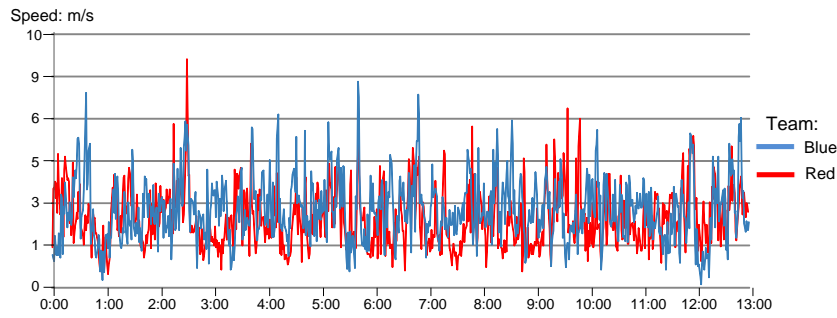
Figure 7: CTM diagrams of travel time on the E40 motorway from Ghent to Brussels.

## 4 ANALYSING MULTIPLE TIME SERIES

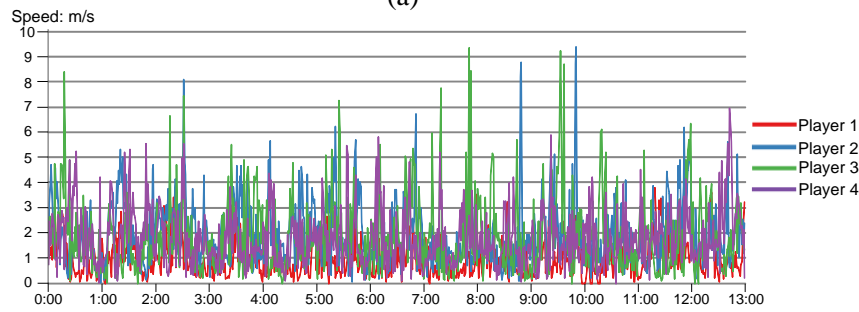
As the CTM is based on a 2D coordinate interval spaces, many spatial analysis techniques in GIScience can be employed to analyse CTM diagrams. This section demonstrates how the methods of map algebra and cartographic modelling are used to analyse multiple time series modelled by the CTM.

### 4.1 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 8, one can only compare the speed between the indoor soccer teams or players at granularity of second. The speed at 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 result 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 9 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: 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 4<sup>th</sup> minute and from the 9<sup>th</sup> to the 13<sup>th</sup> minute. In Figure 9(b), the darkness of the colours indicates the degree of difference, which gives a better sense of the actual difference. From Figure 9(b), one can see that only during some very short intervals (less than 2 minutes), the red team is apparently more active than the blue team.



(a)



(b)

Figure 8: Line charts of multiple time series. (a): the average running speed of two competing soccer teams. (b): the running speed of individual players of the red team.

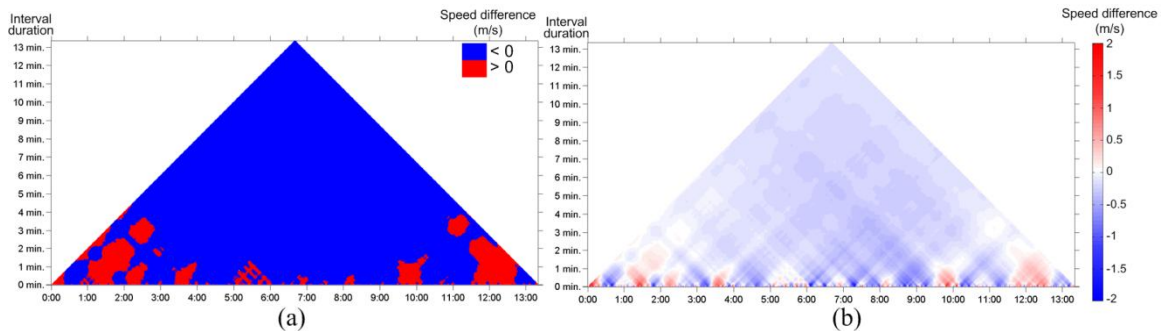


Figure 9: The output of subtracting the CTM of the blue team from that of the red team. (a): Only colour hues are used to indicate the team with higher speed. (b): Colour darkness is applied to indicate the speed difference.

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 greatest running speed is selected. We define this player the dominant player of the interval. If the dominant players of every interval is displayed in a specific colour hue, the output becomes a nominal diagram with four

zones (i.e. Figure 10 (a)), each of which represents a set of intervals during which a certain player is dominant. In Figure 10 (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 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 10 (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 first diagram of Figure 11, every colour indicates a player that 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 12 uses two examples to illustrate this algebra. 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 Player 2 and Player 3 are much more active than the others during two successive 5-minute 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 responds to a slider setting the dominance level.

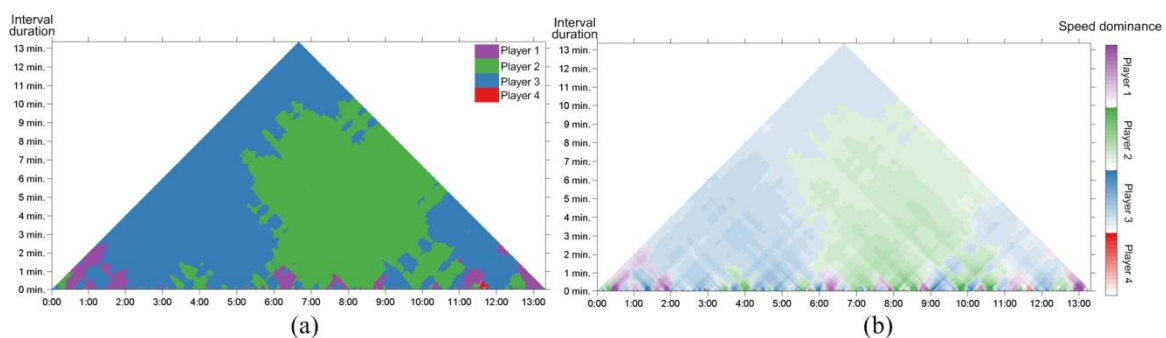


Figure 10: Comparison of multiple players. (a): Dominant players is represented by discrete colour hues. (b): Dominance degree is represented by darkness of colours.

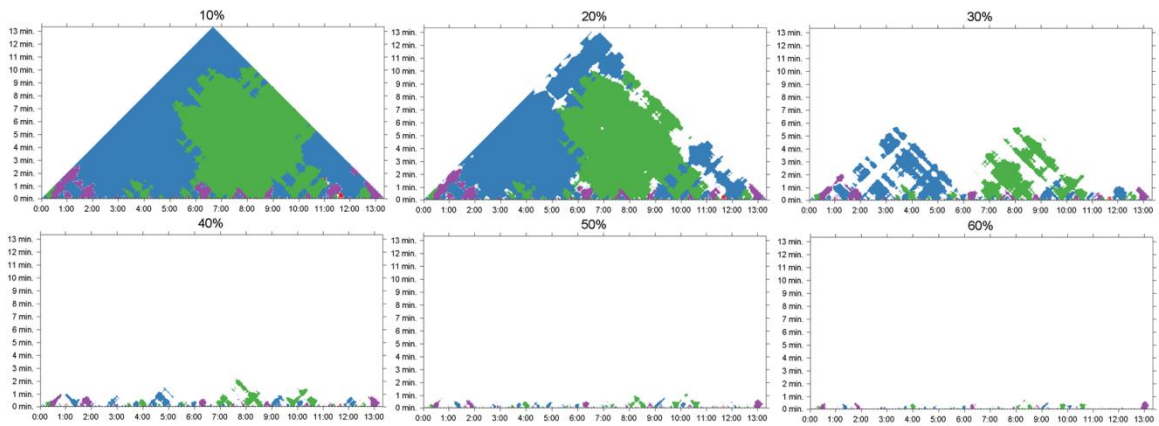


Figure 11: Players dominant at different degrees are represented in multiple diagrams. The meaning of colour hues is same as that in Figure 10.

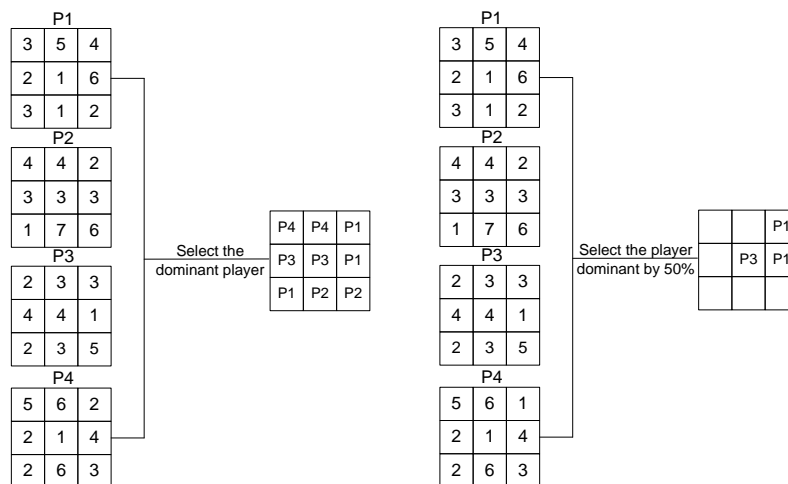


Figure 12: Select the dominant player at every point in a CTM diagram. P means player.

## 4.2 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 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) <sup>23-25</sup>. 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 <sup>26, 27</sup>. 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. Considering the surfers' requirements and preferences in Table 1, they need to select one of these four sites and also decide when and for how long 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.

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.		The more, the better.
Sea temperature	Average sea temperature in every month.		The higher, the better.
Avoidance		Avoiding the international tournament in August.	

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 13 illustrates the specific procedure of the first step. The monthly weather conditions are rated according to the three criteria, i.e., percentage of days with ridable waves<sup>1</sup>, percentage of days with ground waves<sup>2</sup> 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 of their importance (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 at 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.

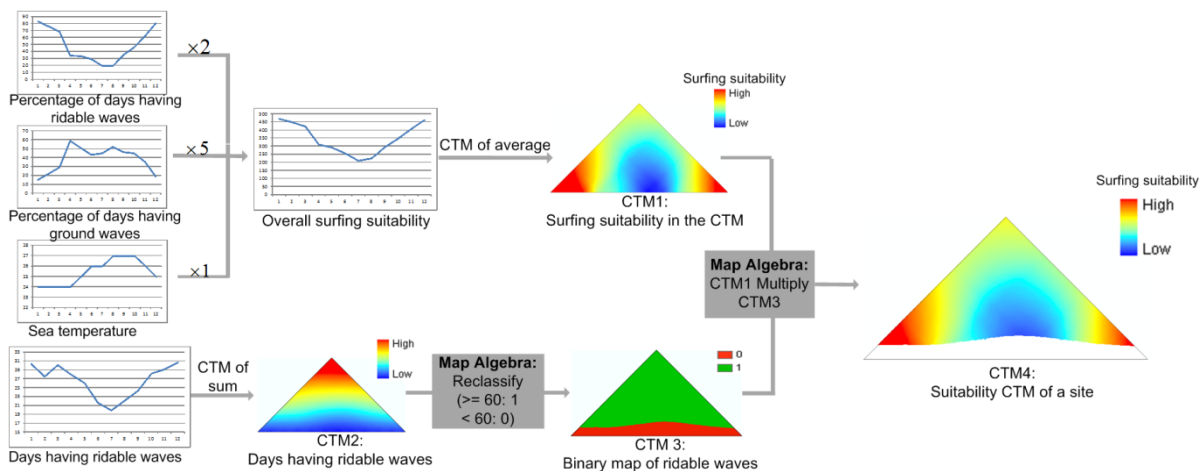


Figure 13: The procedure of generating the suitability CTM diagram for a site.

Figure 14 illustrates the specific procedure of the second step. The suitability diagrams (CTM4s) of all sites are combined together using the dominance algebra. At every position,

<sup>1</sup> Ridable wave: waves last for 7 seconds period or more. (<http://magicseaweed.com/>)

<sup>2</sup> Ground wave: waves last for 10 seconds period or more and over 3ft. (<http://magicseaweed.com/>)



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 in the four candidate sites. 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 Section 2.1, 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, i.e., the rightmost CTM diagram in Figure 14.

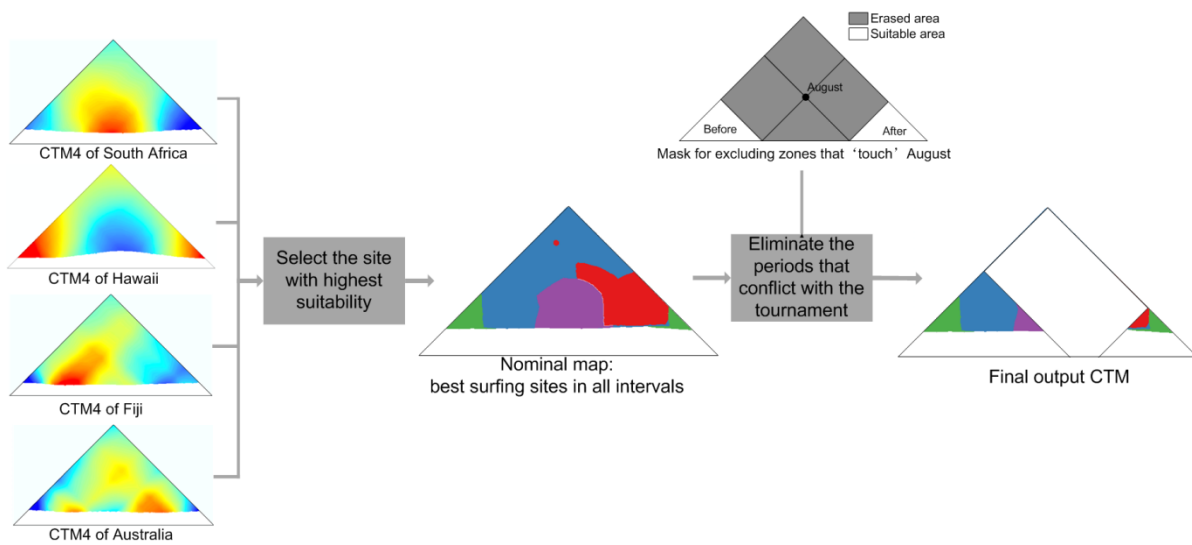


Figure 14: The composition of the suitability CTM diagrams of all sites.

The final output is enlarged in Figure 15. From this diagram, one may have an idea which site is the best candidate during which intervals. Instead of providing 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 2.2), the result in Figure 15 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 15), 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 rideable waves, Hawaii is the very 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 rideable 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 4.1, in this case, the dominance is the percentage that the highest score is over the second highest score. From Figure 16 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 to 3 months.

Compared with the traditional approaches, the CTM-based cartographical modelling has two merits. First, the CTM is not limited to a certain time scale 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 types of visual 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.

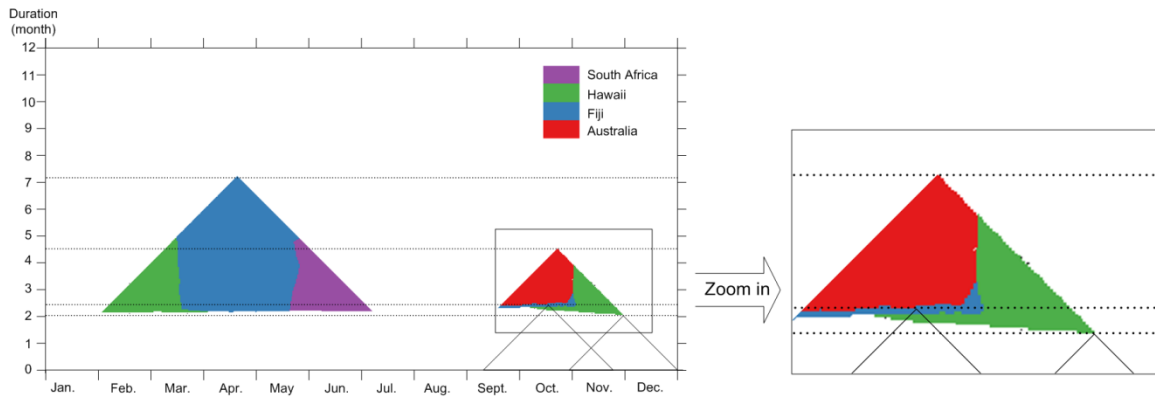


Figure 15: The enlarged result CTM diagram. (Left: the overall diagram. Right: zoom into the right part of the overall diagram).

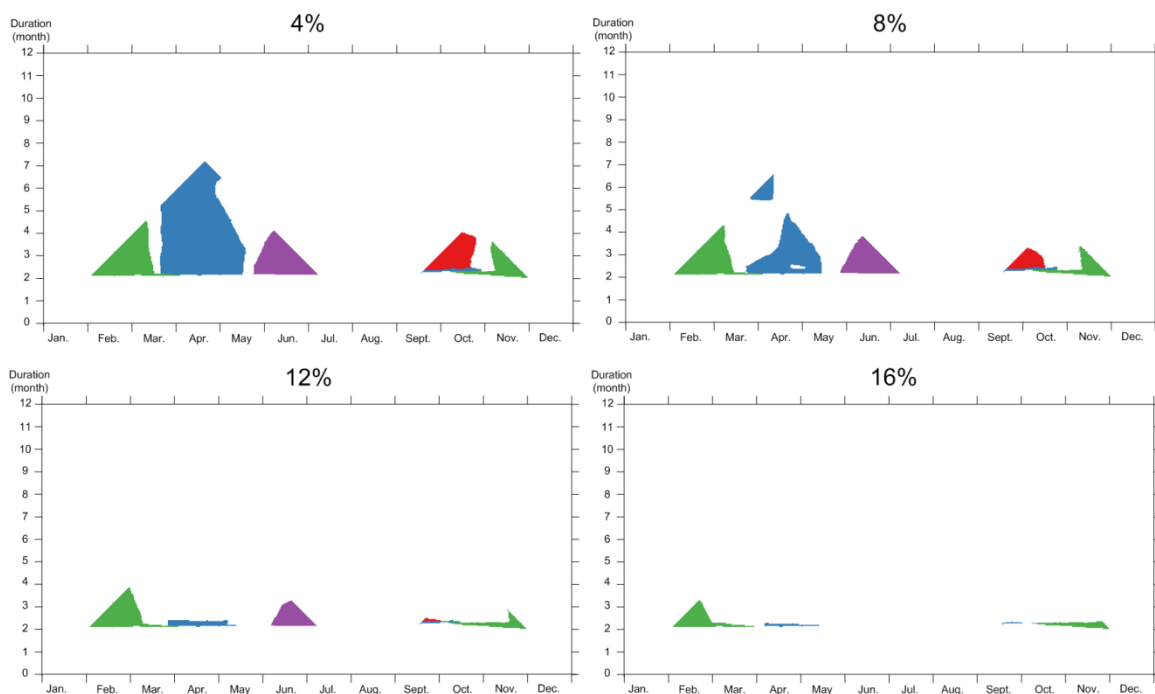


Figure 16: The result CTM based on different dominance threshold. The meaning of colours is identical to that of Figure 15.

## 5 CONCLUSION AND FUTURE WORK

This paper 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 visualisation of linear data at 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 at different scales. The hierarchy and relationships between short-interval and long-interval variations can be directly observed. Second, the CTM is based on a 2D coordinate space, which is very similar to prevalent geospatial datasets. Thus, a vast set of existing, high-level techniques in GIScience can be used to manipulate and analyse CTM diagrams. For instance, by applying map algebra to the CTM, multiple time series can be compared at different scales. Using special visualization techniques, one can identify time series dominant during different time intervals. Based on map algebra, a similar idea of cartographical modelling can be applied to 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 length that meet their criteria and constraints. This analysis is difficult with the traditional representations where analysis is 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 paper gave several examples to show possible applications of the CTM. In future work, there are a lot of 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 minute time frame. The visualization in the CTM may support the coach to make strategies or adjustments accordingly in all different time intervals. The idea of cartographic modelling in the CTM can be also applied to agriculture and farming, which heavily rely on the analysis of time series of weather statistics. The analysis based on the CTM can potentially provide advice for the irrigation strategies or fish egg incubation control. Moreover, implementing the CTM into an interactive system can improve the analytical usability of the CTM. In such a system, the operation and manipulation of the CTM diagrams can be performed more conveniently and automatically. 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 reference. From a general perspective, the CTM can be considered as a conceptual model of generalisation. 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 work.

## References

1. Enders W. *Applied Econometric Time Series (Second Edition)*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2008, p.480.
2. Hamilton JD. *Time Series Analysis*. Princeton University Press, 1994, p.799.
3. Liao TW. Clustering of time series data: a survey. *Pattern Recognition*. 2005; 38(11): 1857-1874.
4. Aigner W, Miksch S, Muller W, Schumann H and Tominski C. Visualizing time-oriented data - A systematic view. *Computers & Graphics-UK*. 2007; 31(3): 401-409.
5. Muller W and Schumann H. Visualization methods for time-dependent data - An overview. *Proceedings of the 2003 Winter Simulation Conference*. New Orleans, U.S.A.: 2003, pp. 737-745.
6. Andrienko N and Andrienko G. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Heidelberg: Springer-Verlag Berlin, 2006.
7. Keim DA, Mansmann F, Schneidewind J and Ziegler H. Challenges in Visual Data Analysis. *Proceedings of the conference on Information Visualization*. IEEE Computer Society, 2006.
8. Havre S, Hetzler B and Nowell L. ThemeRiver: Visualizing theme changes over time. *Proceedings of the IEEE Symposium on Information Visualization 2000*. Salt Lake City, Utah: IEEE Computer Society, 2000, pp. 115-123.
9. Hochheiser H and Shneiderman B. Dynamic query tools for time series data sets: timebox widgets for interactive exploration. *Information Visualization*. 2004; 3(1): 1-18.
10. Lin J, Keogh E, Lonardi S, Lankford JP and Nystrom DM. Visually mining and monitoring massive time series. *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining*. Seattle, WA, USA: ACM, 2004, pp. 460-469.
11. Weber M, Alexa M and Muller W. Visualizing time-series on spirals. *Proceedings of IEEE Symposium on Information Visualization 2001*. Washington, DC, U.S.A.: 2001, pp. 7-13.
12. Kulpa Z. A Diagrammatic Approach to Investigate Interval Relations. *Journal of Visual Languages and Computing*. 2006; 17(5): 466-502.
13. Kulpa Z. Diagrammatic Representation for a Space of Intervals. *Machine Graphics & Vision*. 1997; 6(5-24).
14. Van de Weghe N, Docter R, De Maeyer P, Bechtold B and Ryckbosch K. The Triangular Model as an Instrument for Visualising and Analysing Residuality. *Journal of Archaeological Science*. 2007; 34(4): 649-655.
15. Qiang Y, Delafontaine M, Asmussen K, et al. Modelling Imperfect Time Intervals in a Two-Dimensional Space. *Control and Cybernetics*. 2010; 39(4): 983-1010.

16. Qiang Y, Delafontaine M, Versichele M, De Maeyer P and Van de Weghe N. Interactive Analysis of Time Interval in a Two-Dimensional Space. *Information Visualization*. in press.
17. Qiang Y, Delafontaine M, Neutens T, et al. Analysing Imperfect Temporal Information in GIS using the Triangular Model. *The Cartographic Journal*. in press.
18. 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.
19. Smith MJd, Goodchild MF and Longley PA. *Geospatial Analysis: a Comprehensive Guide to Principles, Techniques and Software Tools*. Leicester, UK: Troubador Publishing Limited, 2007.
20. Keim DA, Nietzsche T, Schelwies N, Schneidewind J, Schreck T and Ziegler H. A Spectral Visualization System for Analyzing Financial Time Series Data. In: Ertl T, Joy K and Santos B, (eds.). *Proceedings of the Eurographics/IEEE-VGTC Symposium on Visualization (EuroVis 2006)*. Lisbon, Portugal: 2006, pp. 195-200.
21. Tomlin CD. *Geographic information systems and cartographic modeling*. New Jersey: Prentice Hall, 1990.
22. Allen JF. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*. 1983; 26(11): 832-843.
23. Collins MG, Steiner FR and Rushman MJ. Land-Use Suitability Analysis in the United States: Historical Development and Promising Technological Achievements. *Environmental Management*. 2001; 28(5): 611-621.
24. Smith L. Site Selection for Establishment & Management of Vineyards. Paper presented in *The 14th Annual Colloquium of the Spatial Information Research Centre University of Otago*. Dunedin, New Zealand: 2002.
25. Walsh SJ, Butler DR, Brown DG and Bian L. Cartographic Modelling of Snow Avalanche Path Location within Glacier-National-Park, Montana. *Photogramm Eng Remote Sens*. 1990; 56(5): 615-621.
26. 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.
27. Jankowski P. Integrating geographical information systems and multiple criteria decision-making methods. *Int J Geogr Inf Syst*. 1995; 9(3): 251-273.