

Published as: Qiang, Y., Delafontaine, M., Neutens, T., Stichelbaut, B., De Tré, G., De Maeyer, P., Van de Weghe, N. (2012). Analysing imperfect temporal information in GIS using the Triangular Model. *The Cartographic Journal*, vol. 49(3) (in press).

Analysing Imperfect Temporal Information in GIS Using the Triangular Model

1. Introduction

Every entity has an extent in time, such as the lifetime of an object or the duration of an event. These temporal extents are usually described by crisp time intervals bounded by a well-defined start and end. However, under some circumstances, the temporal extent of an entity is imperfect, and cannot be adequately modelled by a crisp time interval. On the one hand, some events may start or end gradually and therefore their start and end cannot be pinned to exact time stamps. For example, it is difficult to decide when the Industrial Revolution started and finished. Intervals of this kind of events are usually modelled by fuzzy sets (Zadeh, 1965) through the quantification of the graded truth of whether a time point is in the interval, bring the concept of fuzzy time interval. On the other hand, in some other cases, it is only known that the start and end of a crisp interval are within certain ranges, but no extra information or assumptions about the distribution of the start and the end is available. Modelling this kind of imperfect time intervals with fuzzy sets would induce extra overhead and unnecessary complexity. In these cases, the alternative approach, i.e. rough sets (Pawlak, 1982), can excellently suit the modelling and handling of time intervals. Currently, a lot of disciplines are faced with the issue of imperfect time intervals, which is reflected in many contributions in modelling intervals by fuzzy sets (De Caluwe et al., 1999, De Caluwe et al., 1997, Nagypál and Motik, 2003, Ohlbach, 2004, Schockaert et al., 2008, Garrido et al., 2009) and rough sets (Bittner, 2002, Bassiri et al., 2009). However, while most of this work focuses on modelling and reasoning about imperfect time intervals, techniques and tools for visualising and analysing imperfect time intervals are still lacking. This probably stems from the conventional representation of time intervals (e.g. Gantt chart, time table and historical timeline) which represents time intervals as linear segments along a one-dimensional axis. The second dimension is often exploited merely to differentiate intervals of entities and has no temporal meaning. In this representation, the visual distribution of intervals is variable, according to the application of different ordering rules in the second dimension, for example, ordering intervals from the shortest to the longest, or from the earliest started to the latest started. The characteristics of the distribution of intervals cannot be observed in one single display. This is not convenient for visual pattern detection of time intervals, let alone imperfect time intervals with more complex structure.

This issue also exists in geographical information science (GIScience), which considers time as one of the most important component of geographical information (Peuquet, 2002, Li and Kraak, 2008, Maceachren et al., 1999, Neutens et al., 2008). In the recent development of GIScience, considerable effort has been made in handling the temporal aspect of

geographical data (Andrienko et al., 2003, Smith et al., 2007, Neutens et al., 2007). Due to the limitations of data acquisition techniques, spatial data are often linked to imperfect temporal information. Dealing with imperfect temporal information becomes an increasingly significant issue in spatio-temporal data analysis, particularly in exploratory spatio-temporal data analysis (ESTDA) which greatly relies on graphical representations and visualisations. Since most prevalent techniques and tools of ESTDA represent time in the linear form, their ability in dealing with temporal imperfectness in spatio-temporal data is still not satisfactory.

To address these issues, attempts have been made to represent time in a two-dimensional (2D) space. For example, Keim et al. (2006) arranged time series of financial investments in a 2D representation called the Growth Matrix. This matrix is able to display the growth rates of investments in all possible sub-intervals in the time series. The TT-plot introduced by Imfeld (2000) applied a similar idea to analyse movement patterns. Besides these approaches, Kulpa proposed the ER Diagram that represents time intervals as 2D points, and investigated the use of this diagram for interval reasoning and arithmetic (Kulpa, 1997). Later, Van de Weghe *et al.* (2007) named this representation the Triangular Model (TM) and applied it in an archaeological context. Recently, Qiang *et al.* (Qiang et al., 2009, Qiang et al., 2010) have extended TM to represent time intervals modelled by rough sets and fuzzy sets. This work majorly focused on the use of TM in temporal reasoning. The practical value of TM in visualising and analysing imperfect time intervals is still yet to be exploited.

To fill this gap, this paper will investigate the use of TM in visualising and analysing imperfect time intervals. The focus is on imperfect time intervals that can be modelled by rough set theory. A probabilistic framework has been proposed to model the uncertainties in the temporal relations of such roughly-described intervals. In order to better demonstrate the use of TM, a prototype tool will be introduced, which implements TM in a geographical information system (GIS). This tool support analysis of geographical entities with imperfect intervals. We will show that the advantages of TM in visualising and querying roughly-described intervals can be better exploited through a computer application with a graphical user interface (GUI) and interactive functionalities. In addition, we will use a case study to illustrate how the unique insights gained by TM can assist a GIS for ESTDA involving imperfect time intervals.

In the remainder of the paper, we first introduce the basic concept of TM and how rough approximations of intervals are represented in TM. In Section 3, we apply probability theory to model uncertainties in temporal queries of imperfect intervals. Section 4 presents GeoTM, including its GUI, functionalities and supported data model. In section 5, the use of GeoTM is demonstrated in a concrete use case. The paper ends with a brief conclusion and an outline of avenues for future work. All notations and symbols used in this paper are summarised in the table in the appendix.

2. The Triangular Model

2.1. The Representation of Time Intervals

A time interval I is usually modelled as a convex set of real numbers, i.e. $[I^-, I^+]$ with $I^- < I^+$. I^- and I^+ respectively denote the start and end of I . In the traditional linear representation, a time interval is represented by a finite linear segment bounded by I^- and I^+ (see Figure 1a). This linear representation of time intervals is widely used in our daily life, for example time tables and historical time lines. The transformation from the linear

representation of a time interval to TM can be achieved by constructing two lines through the extremes of an interval (Figure 1b). For each interval I , two straight lines (L_1 and L_2) are constructed, with L_1 passing through I^- and L_2 passing through I^+ . α_1 is the angle between L_1 and the horizontal axis and α_2 is the angle between L_2 and the horizontal axis, with $\alpha_1 = -\alpha_2 = \alpha$. The intersection of L_1 and L_2 is called the interval point. The angle α is a predefined constant which is identical for all intervals to ensure that every interval is mapped to a unique point in the 2D space. Although α can be any value between 0 and 90°, we consider $\alpha = 45^\circ$ for consistency with earlier work (Kulpa, 1997, Van de Weghe et al., 2007, Qiang et al., 2010, Kulpa, 2006). In this way, TM represents all time intervals as points in a 2D space, which is called the interval space (Figure 1c). The interval space is denoted as $I\mathbb{R}$ (Kulpa, 2006). In $I\mathbb{R}$, given an interval point I , the horizontal position indicates its midpoint (e.g. $mid(I)$) and the vertical position indicates its duration (e.g. $dur(I)$).

2.2. The Representation of Temporal Relations

James F. Allen (1983) specified thirteen possible relations between two time intervals (see Table 1), which are referred to as Allen relations. In TM, every Allen relation can be represented as a specific zone (Kulpa, 1997). Given a study interval $I_S = [0, 100]$, in TM all examined intervals are located within the isosceles triangle formed by I_S^- , I_S^+ and I . Let us consider, for example, a reference interval $I_2 [33,66]$. Any intervals (e.g. I_{1a} , I_{1b} , I_{1c}) *before* I_2 (Figure 2a) are located in the triangular area in the left corner of the study area (Figure 2b). Therefore, it is easy to deduce that all intervals *before* I_2 must be located in the black zone in Figure 2c, which is called the *before* zone of I_2 . Likewise, all Allen relations with respect to an interval can be represented by zones in $I\mathbb{R}$ (Figure 3), which are called relational zones. For each relation in Figure 3, the reference interval I has been chosen in the centre of the study period in order to avoid visual bias. Each relational zone represents the set of intervals that are in a specific relation to the reference interval I , which are denoted as $Rel(I)$. For example, the *during* zone of I represents the set of intervals that are *during* I and is denoted as $during(I)$. On the other hand, $Rel(I_1, I_2)$ expresses the statement that I_1 is in the relation Rel to I_2 . For example, $overlaps(I_1, I_2)$ expresses that I_1 overlaps I_2 .

Table 1: Thirteen Allen Relations (Allen 1983)

I_1 equal I_2	if $I_1^- = I_2^- \wedge I_1^+ = I_2^+$
I_1 starts I_2	if $I_1^- = I_2^- \wedge I_1^+ < I_2^+$
I_1 started-by I_2	if $I_1^- = I_2^- \wedge I_2^+ < I_1^+$
I_1 finishes I_2	if $I_1^+ = I_2^+ \wedge I_1^- > I_2^-$
I_1 finished-by I_2	if $I_1^+ = I_2^+ \wedge I_2^- > I_1^-$
I_1 meets I_2	if $I_1^+ = I_2^-$
I_1 met-by I_2	if $I_2^+ = I_1^-$
I_1 overlaps I_2	if $I_2^- > I_1^- \wedge I_1^+ < I_2^+ \wedge I_1^+ > I_2^-$
I_1 overlapped-by I_2	if $I_1^- > I_2^- \wedge I_1^- < I_2^+ \wedge I_2^+ < I_1^+$
I_1 during I_2	if $I_1^- > I_2^- \wedge I_1^+ < I_2^+$
I_1 contains I_2	if $I_2^- > I_1^- \wedge I_2^+ < I_1^+$
I_1 before I_2	if $I_1^+ < I_2^-$
I_1 after I_2	if $I_2^+ < I_1^-$

2.3. The Representation of Rough Approximation

Incomplete information may result in uncertainties about the exact start and end of a time interval. This may happen in many observation activities in which data are acquired at discrete time stamps, such as images or photographs in remote sensing. From a sequence of images, for example, one can determine whether a feature exists at specific time stamps. However, the feature's status between two time stamps is unknown. With these discrete snapshots, the interval of the feature's existence is thus imperfect. If there is no prior knowledge about the distribution of the start and end, modelling this kind of imperfect time intervals with membership functions of fuzzy sets would induce extra overhead and unnecessary complexity. In these cases, rough sets can be considered as an appropriate and adequate solution (Bassiri et al., 2009, Bittner, 2002).

In the rough sets approach, an imperfect time interval I is described by an upper approximation \bar{I} and a lower approximation \underline{I} , with $\underline{I} \subseteq \bar{I}$ (Figure 4). We call such a pair of \underline{I} and \bar{I} the rough approximation of I , which is denoted as $R(I)$. Time points in \underline{I} are *definitely* in I , whereas all time points not in \bar{I} are *definitely not* in I . \bar{I} is bounded by the earliest possible start \bar{I}^- and the latest possible end \bar{I}^+ , while \underline{I} is bounded by the latest possible start \underline{I}^- and the earliest possible end \underline{I}^+ . In between \underline{I} and \bar{I} , there are two rough boundaries $R(I^-)$ and $R(I^+)$ gathering the time points that are *possibly* in I . Unlike fuzzy time intervals which define the extent to which the time points are possibly in I , the rough approximation of I classifies time points into *definitely in*, *definitely not in* and *possibly in* I .

In the linear representation, the rough approximation of an interval is usually represented as a tripartite linear segment (Figure 5a). However, in TM, a rough approximation of an interval is represented by a convex 2D geometry in $I\mathbb{R}$. To construct a rough approximation $R(I)$ in TM, two parallel lines are projected from \bar{I}^- and \underline{I}^- with angle α to the horizontal axis; and the other two lines are projected from \underline{I}^+ and \bar{I}^+ with angle $-\alpha$ to the horizontal axis. These four lines form a rectangle (Figure 5b) which indicates a zone where the exact interval I can be found. In other words, this zone represents the set of intervals that are possibly equal to I . The shape and location of the rectangle completely express the characteristics of $R(I)$. In this way, the rough approximation of a time interval can be represented by such a rectangle (Figure 5c). If the lower approximation is empty, the rough approximation becomes a triangle on the horizontal axis, e.g. I_2 in Figure 5.

3. Probabilities of Rough Approximations

Because time intervals described by rough approximations have an uncertain start and an uncertain end, temporal queries about such imperfect intervals may result in uncertain answers. To handle the uncertainties in temporal queries, a probabilistic framework is needed. Given a rough approximation $R(I)$, $R(I^-)$ and $R(I^+)$ define the ranges where the exact start (I^-) and the exact end (I^+) can be found (Eq. 1 and Eq. 2). As there is no further knowledge about the distribution of I^- and I^+ in $R(I^-)$ and $R(I^+)$, we assume that the probability distribution in $R(I^-)$ and $R(I^+)$ are uniform. That is to say, every time point in $R(I^-)$ has the same probability of being equal to I^- (i.e. Eq. 3). Therefore, the probability density of $x = I^-$ ($x \in R(I^-)$) is the quotient of 1 and the cardinality of $R(I^-)$ (Eq. 4). In the same way, every time point in $R(I^+)$ has the same probability of being equal to I^+ (Eq. 5), and therefore the probability density of $x = I^+$ ($x \in R(I^+)$) is the quotient of 1 and the cardinality of $R(I^+)$

(Eq. 6). Based on these prerequisites, we can deduce that every interval in $R(I)$ has the same probability of being equal to I (Eq. 7). Because $R(I)$ defines the interval set that includes I (Eq. 8), the probability density of $x = I$ ($x \in R(I)$) is the quotient of 1 and the cardinality of $R(I)$ (Eq. 9). Therefore, $R(I)$ can be considered as an interval set in which every interval has the same probability of being equal to I . Because TM maps time intervals into points in a 2D space, the cardinality of an interval set is proportional to its area, and therefore the probability density of $x = I$ ($x \in R(I)$) is inversely proportional to the area of $R(I)$ in TM (see Eq. 9).

According to Eq. 9, we can deduce that, given a set of intervals A and a rough approximation $R(I)$ (such as Figure 6), the probability that I is in A is the ratio between the cardinality of $A \cap R(I)$ and the cardinality of $R(I)$ (Eq. 10). In the TM representation, this probability is the ratio between the area of $A \cap R(I)$ and the area of $R(I)$ (Eq. 10). If $R(I)$ is totally contained in A (i.e. $R(I) \subseteq A$) the probability that I is in A is 1. If $R(I)$ and A do not intersect (i.e. $R(I) \cap A = \emptyset$), the probability that I is in A is 0. The interval set A can also be a relational zone of a certain interval, e.g. *before*(I'). In this case, the probability that I is in A can be interpreted as the probability that I is *before* I' . With respect to the principles of probability theory (Jaynes and Bretthorst, 2003), the probabilities of multiple rough approximations can be deduced. For example, given an interval set A and two independent rough approximations $R(I_1)$ and $R(I_2)$ (Figure 7), the probability that only one of I_1 and I_2 is in A can be obtained from Eq. 11, and the probability that both I_1 and I_2 are in A can be obtained by Eq. 12. Analogously, given an interval set A and n independent rough approximations i.e. $R(I_1), R(I_2) \dots R(I_n)$, the probability that m intervals ($0 \leq m \leq n$) are in A can be obtained by Eq. 13. Since interval sets and rough approximations of intervals are represented as 2D geometries in TM, the probability that an interval is in an interval set is expressed by the overlap ratio of their geometries. This feature of TM are similar to Venn diagrams (Venn, 1881) in which the areas represents the occurrences of a certain event and the intersection of areas expresses the coincidence of events. Compared with mathematic formulas, we contend that such graphic representation is more intuitive for human beings. Hence, TM can be considered as a promising platform for visual queries of intervals described by rough approximations.

$$\sum_{x \in R(I^-)} P(I^- = x) = 1 \quad \text{Eq. 1}$$

$$\sum_{x \in R(I^+)} P(I^+ = x) = 1 \quad \text{Eq. 2}$$

$$P(I^- = x_1) = P(I^- = x_2), \quad x_1, x_2 \in R(I^-) \quad \text{Eq. 3}$$

$$f_{I^-}(x) = P(I^- = x) = \frac{1}{|R(I^-)|}, \quad x \in R(I^-) \quad \text{Eq. 4}$$

$$P(I^+ = x_1) = P(I^+ = x_2), \quad x_1, x_2 \in R(I^+) \quad \text{Eq. 5}$$

$$f_{I^+}(x) = P(I^+ = x) = \frac{1}{|R(I^+)|}, \quad x \in R(I^+) \quad \text{Eq. 6}$$

$$P(I = I_1) = P(I = I_2), \quad I_1, I_2 \in R(I) \quad \text{Eq. 7}$$

$$\sum_{x \in R(I)} P(I = x) = 1 \quad \text{Eq. 8}$$

$$f_I(x) = P(I = x) = \frac{1}{|R(I)|} \propto \frac{1}{\text{Area}(R(I))}, \quad x \in R(I) \quad \text{Eq. 9}$$

$$P(I \in A) = \sum_{x \in A} \frac{1}{|R(I)|} = \frac{|A \cap R(I)|}{|R(I)|} = \frac{\text{Area}(A \cap R(I))}{\text{Area}(R(I))} \quad \text{Eq. 10}$$

$$\begin{aligned} P(\text{either } I_1 \text{ or } I_2 \text{ is in } A) &= P(I_1 \in A \wedge I_2 \notin A) + P(I_1 \notin A \wedge I_2 \in A) \\ &= P(I_1 \in A) \times P(I_2 \notin A) + P(I_1 \notin A) \times P(I_2 \in A) \\ &= \frac{|A \cap R(I_1)|}{|R(I_1)|} \times \frac{|A - (A \cap R(I_2))|}{|R(I_2)|} + \frac{|A \cap R(I_2)|}{|R(I_2)|} \times \frac{|A - (A \cap R(I_1))|}{|R(I_1)|} \\ &= \frac{\text{Area}(A \cap R(I_1)) \times \text{Area}(A - A \cap R(I_2)) + \text{Area}(A \cap R(I_2)) \times \text{Area}(A - A \cap R(I_1))}{[\text{Area}(R(I_1)) \times \text{Area}(R(I_2))]^2} \end{aligned} \quad \text{Eq. 11}$$

$$\begin{aligned} P(\text{both } I_1 \text{ and } I_2 \text{ are in } A) &= P(I_1 \in A) \times P(I_2 \in A) \\ &= \frac{|A \cap R(I_1)|}{|R(I_1)|} \times \frac{|A \cap R(I_2)|}{|R(I_2)|} \\ &= \frac{\text{Area}(A \cap R(I_1))}{\text{Area}(R(I_1))} \times \frac{\text{Area}(A \cap R(I_2))}{\text{Area}(R(I_2))} \end{aligned} \quad \text{Eq. 12}$$

$$P(m \text{ intervals of } I_1, I_2, \dots, I_n \text{ are in } A) = \sum_{Q \subseteq S} \left[\prod_{i \in Q} P(I_i \in A) \times \prod_{j \in (S-Q)} P(I_j \notin A) \right], \quad \text{Eq. 13}$$

$$S = (1, 2, 3 \dots n) \wedge Q \subseteq S \wedge |Q| = m$$

4. The Implementation

To demonstrate and exploit practical uses of TM, we have implemented TM into a prototype tool (i.e. GeoTM) which incorporates TM into a GIS. GeoTM allows visualising, querying and analysing spatio-temporal data. More specifically, it is able to handle discrete entities with spatial locations and temporal extents. The spatial locations are modelled as vector geometries (i.e. point, line and polygon), while the temporal extents are modelled as time intervals. In this paper, we will take special care of time intervals described by rough approximations. Besides spatial and temporal extents, these geographical entities may have other attributes. The spatial locations, temporal extents and other attributes are linked to geographical entities by unique identifiers (ID) of entities. GeoTM is built on top of ArcGISTM which is a desktop GIS produced by ESRI[®]. Within its object model (ArcObjects), developers can call on existing functions and components of ArcGISTM to develop customised applications. Consequently, GeoTM is compatible with the supported data formats in ArcGISTM, such as shapefile and dBASE files. In the representation of geographical entities, spatial objects and time interval points are stored in two shapefiles, while the attributes are stored in a database table (i.e. a dBASE file). Figure 8 gives an overview of how geographical entities are modelled and represented in GeoTM.

4.1. The Graphic User Interface of GeoTM

The user interface of GeoTM consists of a map view, a TM view and controls that can trigger specific functions or user forms (Figure 9). The map view is used to display spatial locations of geographical entities, which are modelled as points, lines or polygons. The TM view displays time intervals of entities, using the TM representation. Rough approximations of intervals are represented as polygons. In the TM view, there may exist overlaps of polygons. In order to display the pattern of overlapped polygons, gradual colours are assigned to areas

according to the numbers of overlaps: a darker colour is assigned to the areas with more overlaps. In Figure 9, a dataset of military features in the First World War (WWI) is displayed in GeoTM. The spatial locations of these features are displayed in the map view, while their lifetimes (represented by rough approximations of intervals) are displayed in the TM view. In this section, we will use this dataset to illustrate the functionalities of GeoTM. The entire scenario of this dataset will be described in Section 5. The temporal resolution of the dataset is one day. Thus, in this context, a time interval is the period between two specific dates formatted as Year/Month/Day.

Users can interactively browse the map view and TM view by zooming and panning. In the map view, objects can be selected by dragging a rectangle. In the TM view, special selection tools have been designed to select intervals according to specific temporal queries, which will be described in detail in the next sub-section. Other attributes of entities are stored in the attribute table that can pop up when the user clicks the corresponding button. ‘Linked brushing’ is supported among the map view, the TM view and the attribute table. This function allows selecting objects from any of these three views and dynamically updating the other two views to highlight the corresponding objects. With this function, geographical entities in GeoTM can be queried with spatial, temporal and attribute constraints. Additionally, many common functions of a conventional GIS are supported in GeoTM. For example, several datasets can be loaded into GeoTM as multiple layers.

4.2. Temporal Queries in GeoTM

As introduced in Section 2.2, intervals that satisfy a certain temporal relation are located within a relational zone. Thus, queries based on temporal relations can be modelled as specific zones in TM, i.e. query zones. A query zone can be a relational zone or combinations (e.g. intersection and union) of multiple relational zones. Whether an interval satisfies the temporal query depends on whether the interval point lies within the query zone. However, when intervals are described by rough approximations, in TM the rough approximation can be partially in the query zone. If this is the case, temporal queries can be answered with probability thresholds. For example, within a dataset of rough approximations of intervals, one can select all intervals that have more than 90% probability of being *during* an indicated interval. In TM, this query is expressed by selecting rough approximations that have more than 90% of its area in the *before* zone of the indicated interval. Therefore, in the TM view of GeoTM, querying tools are available to define temporal queries by creating query zones. On the one hand, users can indicate an interval by moving the cursor to a specific position in the TM view. When right-clicking on this position, a drop-down menu of Allen relations appears. Next, by clicking an Allen relation in the menu, all intervals in this relation are selected, according to a pre-defined probability threshold. Figure 10 shows an example of selecting intervals that are probably *before* [1916/06/25,1917/05/31]. In this example, the lower probability threshold is 0.6 and the upper probability threshold is 1. As a result, intervals with more than 60% probability *before* [1916/6/25,1917/5/31] are selected.

Besides temporal queries of Allen relations, some other queries can be made by dragging specific geometries in the TM view. For example, a convex set of intervals can be selected by dragging a rectangle whose sides are in α or $-\alpha$ angle to the horizontal axis (Figure 11a). A convex interval set is defined as the set of time intervals *in-between* two different time intervals (Kulpa, 2006). We developed this query tool because convex interval sets can be easily interpreted by Allen relations or combinations of Allen relations. $\ll I_1, I_2 \gg$ is used to

denote a convex interval set *in-between* I_1 and I_2 . The formal definition of $\ll I_1, I_2 \gg$ can be found in Eq. 14. Moreover, two parallel lines can be dragged in α or $-\alpha$ angle with the horizontal axis, in order to select intervals that *start-within* or *end-within* a certain interval (e.g. Figure 11b and c). *starts-within* (I_1) is defined as a set of intervals whose starts are in I_1 (Eq. 15). Analogously, *ends-within* (I_1) is defined as a set of intervals whose ends are in I_1 (Eq. 16). Though the *starts-within* and *ends-within* can be expressed by unions of Allen relations, we design such query tools for people that are more accustomed to the expressions of ‘starts within’ and ‘ends within’ than unions of Allen relations. Furthermore, by dragging a range along the vertical axis, users can select intervals whose durations are in a specific range (Figure 11d). In GeoTM, all queries are carried out with respect to pre-defined probability thresholds. In the examples in Figure 11, the lower threshold is set to 0.6 and the upper threshold is set to 1, meaning that the selected intervals have more than 60% probability of satisfying these queries.

$$\ll I_1, I_2 \gg \equiv \{I \mid \min(I_1^-, I_2^-) < I^- < \max(I_1^-, I_2^-) \wedge \min(I_1^+, I_2^+) < I^+ < \max(I_1^+, I_2^+)\} \quad \text{Eq. 14}$$

$$\text{starts-within}(I_1) \equiv \{I \mid I_1^- < I^- < I_1^+\} \quad \text{Eq. 15}$$

$$\text{ends-within}(I_1) \equiv \{I \mid I_1^- < I^+ < I_1^+\} \quad \text{Eq. 16}$$

Based on Eq. 13, when a temporal query is defined, users may obtain the probability that a given number of intervals satisfy this query. For example, GeoTM can return the probabilities that a certain number of intervals are *before* [1915/11/21, 1916/08/30]. In this case, 67 rough approximations have common parts with (intersect or within) the query zone, while 43 out of these 67 rough approximations are totally within the query zone. This means that there are minimum 43 intervals *before* [1915/11/21, 1916/08/30] and maximum 67 intervals *before* [1915/11/21, 1916/08/30]. The probability that n intervals ($43 < n < 67$) are *before* [1915/11/21, 1916/08/30] is between 0 and 1. GeoTM can automatically generate the probability that n intervals are *before* [1915/11/21, 1916/08/30], from $n=43$ to $n=67$ (Figure 12). In this way, one can know how many intervals satisfy the query, with respect to a confidence level. According to Figure 12, there are 59 intervals *before* [1915/11/21, 1916/08/30], at 0.9 confidence level. This function is also useful in distinguishing ‘real clusters’ and ‘fake clusters’ of intervals. Because the colours in TM represent the number of overlaps, which is actually the maximum number of intervals within the area, an area with a darker colour indicates a potential cluster of intervals. However, the probability that intervals are clustered in this area varies. By dragging a query zone over the dark area, one can be aware of the numbers of intervals clustered in this area with respect to specific probabilities.

5. The Case Study

Having introduced basic functionalities of GeoTM in the previous section, this section will illustrate the use of TM in supporting GIS for ESTDA by means of a case study.

5.1. The Dataset

During World War One (WWI), a large number of aerial photos in West-Flanders (Belgium) were taken at discrete time stamps by all combating nations as an intelligent tool to collect information on the enemy’s intentions. These aerial photographs are preserved in archives all over the world. The largest collections are held at the Belgian Royal Army Museum, the Imperial War Museum, the Australian War Memorial and the Bavarian Military archive. From these aerial photos, one can observe whether a military feature (e.g. a fire trench, gun

position or barrack) was not yet present, present, or destroyed (Stichelbaut and Bourgeois, 2009). Although the state of a feature is uncertain between two snapshots, we assume that it does not change between two snapshots which show similar states. Thus, the uncertainty only exists between two snapshots showing different states. Certainly, this assumption relies on our knowledge that snapshots are dense enough to capture most of features' changes. When more volatile entities are considered, an appropriate temporal resolution will be required. In this context, rough sets are excellently suited for temporal modelling, because no knowledge is available about the state of a feature between two snapshots. Thus, we can consider a period of snapshots showing similar states as the lower approximation, its neighbouring uncertain intervals as boundary regions, and all of them form the upper approximation (Figure 13). Thus, a feature's lifetime can be meaningfully represented by a rough approximation.

More specifically, four photos determine the lifetime of a feature (Figure 13): (1) the last photo in which the feature is not yet present, (2) the first photo in which the feature is present, (3) the last photo in which the feature is present, and (4) the first photo in which the feature is destroyed or abandoned. The interval between the dates of photo (2) and (3) is the lower approximation of the feature's lifetime, while the interval between the dates of photo (1) and photo (4) is the upper approximation. Intervals between the dates of photo (1) and (2), and intervals between the dates of photo (3) and (4) are the boundary regions, which indicate respectively the range of the feature's construction and destruction/abandonment dates. There are a few exceptions, where a feature was not yet present in one photo and already destroyed in the following photo. In these cases, photo (2) and (3) are missing, and therefore the rough approximation has an empty lower approximation. As described in Section 2.3, such rough approximations are represented as triangles on the horizontal axis. In this case study, a rectangular area (around 3 kilometre \times 3.5 kilometre) is selected near Ypres (Belgium) as the study area, containing 2466 military features (Figure 14). This study area near Ypres is one of the most important battlefields of WWI. From 1914 to 1918 there was constant artillery fighting going on. And more importantly, Ypres was the scene of several large offensive actions by both Germans and Allies. Therefore, it is an ideal area to test the potential of TM in exploratory analysis.

For analysis, we also consider spatio-temporal information of the frontlines during the war. There are 11 snapshots of the states of the German and Allied frontlines during WWI. From these snapshots we have observed that the frontlines were stable in I_2 [1915/5/25, 1916/6/14] (we format a date as year/month/date), but had significant shifts in three time intervals, i.e. I_1 [1915/5/23, 1915/5/25], I_3 [1916/6/14, 1917/10/1] and I_4 [1917/10/1, 1918/4/15] (Figure 15). Note that the frontline shifts in I_3 and I_4 might only take a few days. But from snapshots of frontlines, it is only known that the shifts happened in these three intervals and the exact dates of the shifts are unavailable. Figure 15 displays the locations and shifts of frontlines in these four intervals. Note that the German army is always on the east side of the Allied army. If the positions of frontlines were out of the study area, such as in I_3 and I_4 , we have used arrows to indicate the direction of the shifts.

5.2. Analysing Rough Time Intervals in GeoTM

First, the dataset of the military features is imported to GeoTM. The spatial locations of these features are displayed in the map view, while the rough approximations of their intervals are displayed in the TM view (Figure 16). To identify the intervals of frontline shifts, the interval points and *during* zones of I_1 , I_2 , I_3 and I_4 are added to the TM view (Figure 16). The zone with solid boundary is the *during* zone of the interval in which the frontlines were stable, i.e.

I_2 , while the zones with dashed boundaries are the *during* zones of the intervals in which the frontlines had significant shifts, i.e. I_1 , I_3 , and I_4 . As I_1 is very short, its *during* zone is invisible in this scale. Using the selection tool we find that 99% of features (2433 out of 2466) in the study area were built after I_1 . Considering the frontline shift in I_1 , we can infer that the major military activities in the study area started after I_1 when the frontline moved from the edge of the study area to the centre. Three dark areas that indicate potential clusters of intervals can be observed in TM (Figure 16). In the next three sub-sections these potential clusters will be analysed with respect to frontline states in I_2 , I_3 and I_4 . Note that in the following three sub-sections, we set the lower probability threshold of all temporal queries as 0.9 and the upper probability threshold as 1, in order to select intervals that satisfy the query with more 90% probability.

5.2.1. The stable period of frontlines

Most intervals in Cluster 1 and Cluster 2 *start-within* I_2 , when the frontlines were relatively stable. When selecting intervals probably *start-within* I_2 (more than 0.9 probability), we obtain 1531 intervals (62% of all features in the study area), which reflects that a majority of military features were constructed during this stable period. There are also some intervals that *end-within* I_2 , which are lifetimes of features that were destroyed within I_2 . By selecting intervals that *end-within* I_2 , we obtained 200 intervals (8% of all features) and the map view shows all these features were close to the frontlines (Figure 17). This is reasonable because military features close to the frontline are easier to be destroyed by artillery attacks or minor offensives. Therefore, it is natural to infer that during I_2 there were no significant military actions, and the armies at both sides were building military features in order to keep their positions. Only a small portion of features close to the frontlines were destroyed.

5.2.2. The shift of frontlines in I_3

During I_3 , the frontlines shifted from their positions to the east of the study area, towards the German side. In the TM view, it is observable that the intervals in Cluster 2 *end-within* I_3 . More specifically, the intervals in Cluster 2 *end-within* a quite short interval [1917/07/10, 1917/08/01], which results in a slim shape (see Cluster 2 in Figure 16). When selecting intervals that *end-within* [1917/07/10, 1917/08/01], we obtain 1050 features, which is 43% of all features (2466 features) in the study area. This finding reflects that a large number of features were intensively destroyed during [1917/07/10, 1917/08/01]. The map view shows that all these features were distributed in the eastern side of the old frontlines, which were earlier occupied by the German army (Figure 18). By checking the attribute table, we can see that 99% of these 1050 features were German features, which is 83% of all German features (1260 features) in the study area. With these findings, it becomes possible to infer, without consulting any historical documents, that an overwhelming and intensive destruction happened to German features in July 1917 in the study area. Probably due to this destruction, the German army lost its area and was pushed to the east. Consulting historical literature (Barton, 2005, Verbeke, 2006), these findings reflect the fact that during this period the Allies intensively destroyed German features using artillery fire, in a battle which became known as the Third Battle of Ypres. During the battle, Allied army took Passchendaele (Belgium), pushing the front line towards the east.

5.2.3. The shift of frontlines in I_4

Through observation in the TM view, most intervals in Cluster 1 and Cluster 3 *end-within* I_4 . When selecting intervals probably *end-within* I_4 (with more than 90% probability), 817 features are obtained and the attribute table shows that all these features are Allied. This means that 68% Allied features (817 out of 1206) in the study area were probably destroyed

within I_4 . Considering that within I_4 the frontlines have moved from east to west towards the Allied side, it is natural to infer that these features were destroyed due to the offensive of German army. Features in Cluster 1 were built earlier (i.e. during I_2), when the frontlines were relatively stable. By selecting features in Cluster 1, one can see these features are mostly located in the eastern part of the study area, which was occupied by the Allied army during I_2 . Features in Cluster 3 were built later (i.e. during I_4), and are evenly distributed over the entire study area (Figure 19). These findings suggest that after the Third Battle of Ypres, the Allied army controlled the whole study area and built military features over the area (i.e. features in Cluster 3). Later on, the German army attacked back and destroyed all Allied features (features in both Cluster 1 and Cluster 3) in the study area. After this military action, the frontlines shifted from the east of the study area to the west of the study area. Known from historical literatures (Howard, 2002), these findings probably reflect the Battle of Lys (a part of Spring Offensives of Germany) in April 1918, during which the German army attacked the Allied army and pushed the frontline back to the west.

6. Conclusions and Future Work

This paper has investigated the use of TM in visualising and analysing time intervals modelled by rough sets. Compared to the fuzzy set approach, the rough set approach is excellently suited to imperfect time intervals with no prior knowledge and assumptions about the distribution of their starts and ends, which broadly exist in discrete data acquisitions. In TM, rough approximations of intervals are represented as polygons in a 2D space. We contend that this representation is advantageous in visualising the distribution of imperfect time intervals, because the patterns (e.g. clusters) in the distribution can be explicitly displayed. If an interval is described by rough approximation, its relation with other intervals may have uncertainties. Due to these uncertainties, temporal queries may generate uncertain answers. Therefore, we have proposed a probabilistic framework to model uncertainties in temporal relations of roughly-described intervals. Since TM represents temporal relations as 2D geometries, the probability of temporal relations can be expressed by the overlap rate of geometries, which is analogous to Venn diagrams. Compared to mathematical expressions, such graphical representation is potentially more intuitive to human beings and can offer a promising basis for visual temporal queries.

In order to evaluate the capabilities of TM, we have implemented it into a prototype tool (i.e. GeoTM) which incorporates TM within a GIS. GeoTM shows that the advantages of TM in visualisation and querying can be better exploited when it is implemented in a computer program. Instead of providing a static TM visualisation, GeoTM supports interactive functionalities that enable flexible manipulations of the TM display. These functionalities have potential of assisting visual observation and pattern detections. Besides visualisation, GeoTM also offers possibilities to make temporal queries by creating 2D geometries. Whether intervals satisfy the query depends on the extent to which the intervals are part of the query zone. Moreover, query zones can be directly created on top of the visualisation of intervals so that users can select an observed cluster of intervals by designing a proper temporal query. In order to handle the uncertainties in temporal relations of roughly-described intervals, special functions based on the proposed probabilistic framework have been applied in GeoTM, allowing users to make temporal queries with consideration of probabilities. On the one hand, temporal queries can be defined with probability thresholds, which ensure that every retrieved interval satisfies the query in a certain probability. On the other hand, once a query is made, GeoTM can return the probability that a certain number of intervals satisfy the query. With this feature, users can be aware of how many intervals

satisfy the query, with respect to a certain confidence level. With these functionalities, GeoTM is able to support ESTDA of geographical entities with reference to roughly-described intervals. As shown in the case study of WWI features, GeoTM can be used to explore the military features from spatial, temporal and attribute aspects. From the detected patterns and clusters, users can discover interesting phenomena in the war without consulting historical literatures.

Not only spatio-temporal data, we believe that TM can be applied in other contexts that involve roughly-described intervals for purposes of information visualisation, exploratory analysis and data mining. In future work, the applicability of TM needs to be further assessed by more use cases, preferably covering different research contexts. Also, future extensions will improve the usability and interactivity of the implementation. Before the implementation is released to a broader community, its scalability needs to be systematically evaluated. Furthermore, we plan to investigate the possibility of representing and analysing fuzzy time interval by TM. The fuzzy set approach is more suitable for modelling imperfect time intervals when assumptions or knowledge of the distributions of starting points and end points are available. This extension may rely on the application of more advanced visualisation techniques such as 3D visualisation.

Acknowledgement

The research work of Yi Qiang, Matthias Delafontaine and Tijds Neutens is funded by the Research Foundation–Flanders (FWO).

Appendix: The summary of notations used in the manuscript

Symbols	Name of symbols	Examples
I	A crisp time interval	
I^-	The start of I	
I^+	The end of I	
$dur(I)$	The duration of I	
$mid(I)$	The midpoint of I	
$I\mathbb{R}$	The interval space, i.e. the universal set that contains all time intervals	
\bar{I}	The upper approximation of I	
\underline{I}	The lower approximation of I	
\bar{I}^-	The start of \bar{I}	
\bar{I}^+	The end of \bar{I}	
\underline{I}^-	The start of \underline{I}	
\underline{I}^+	The end of \underline{I}	
$R(I)$	The rough approximation of I	
$R(I^-)$	The earlier boundary region or the rough start of I	
$R(I^+)$	The later boundary region or the rough end of I	
$f_X(x)$	The probability density function of X	$f_{I^-}(x)$ is the probability density function of I^- .
$P(X)$	The probability that the statement	$P(I_1 = I)$ denotes the probability

	X is true	that $I_1 = I$ is true.
$ A $	The cardinality of the set A , i.e. the total number of elements in A . A can be a set of real numbers or a set of intervals.	$ R(I) $ denotes the cardinality of $R(I)$.
$Rel(I)$	A relational zone of I or the set of intervals that are in a certain relation to I .	$before(I)$ is the <i>before</i> zone of I and denotes the set of intervals that are <i>before</i> I .
$Rel(I_1, I_2)$	The statement that I_1 and I_2 are in a certain relation.	$overlaps(I_1, I_2)$ expresses the statement that I_1 <i>overlaps</i> I_2 .
$\ll I_1, I_2 \gg$	The set of intervals that are <i>in-between</i> I_1 and I_2 .	The formal definition of <i>in-between</i> can be found in Eq. 14

References

- Allen, J. F. (1983). Maintaining Knowledge About Temporal Intervals. *Communications of the ACM*, 26(11), pp.832-843.
- Andrienko, N., Andrienko, G. and Gatalsky, P. (2003). Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing*, 14(6), pp.503-541.
- Barton, P. (2005). *De Slagvelden van Wereldoorlog 1. Van de Eerste Slag om Ieper to Passendale*, Tielt, Lannoo.
- Bassiri, A., Malek, M., Alesheikh, A. and Amirian, P. (2009). Temporal Relationships between Rough Time Intervals. in *Computational Science and Its Applications – ICCSA 2009*, pp.543-552.
- Bittner, T. (2002). Approximate Qualitative Temporal Reasoning. *Annals of Mathematics and Artificial Intelligence*, 36(1), pp.39-80.
- De Caluwe, R., Van Der Cruyssen, B., De Tré, G., Devos, F. and Maesfranckx, P. (1999). The Semantics and Modelling of Flexible Time Indications. in *Computing with Words in Information/Intelligent Systems 1 (Foundations)*, pp.229-256, Springer/Physica-Verlag, Heidelberg, Germany.
- De Caluwe, R., Van Der Cruyssen, B., De Tré, G., Devos, Filip and Maesfranckx, P. (1997). Fuzzy Time Indications in Natural Language Interfaces. in *Flexible Query Answering Systems*, pp.163-185, Kluwer Academic Publishers.
- Garrido, C., Marín, N. and Pons, O. (2009). Fuzzy Intervals of Represent Fuzzy Valid Time In a Temporal Relational Database. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 17(1), pp.173-192.
- Howard, M. (2002). *The First World War*, Oxford, Oxford University Press.
- Imfeld, S. (2000). *Time, Points and Space: Towards a better analysis of wildlife data in GIS*. Ph.D., University of Zurich.
- Jaynes, E. T. and Bretthorst, G. L. (2003). *Probability Theory: The Logic of Science*, Cambridge, United Kingdom, Cambridge University Press.
- Keim, D. A., Nietzsche, T., Schelwies, N., Schneidewind, J., Schreck, T. and Ziegler, H. A Spectral Visualization System for Analyzing Financial Time Series Data. ed. by Ertl, T., Joy, K. and Santos, B., *Proceedings of the Eurographics/IEEE-VGTC Symposium on Visualization (EuroVis 2006)*, Lisbon, Portugal, pp.195-200.

- Kulpa, Z. (1997). Diagrammatic Representation for a Space of Intervals. *Machine Graphics & Vision*, 6, pp.5-24.
- Kulpa, Z. (2006). A Diagrammatic Approach to Investigate Interval Relations. *Journal of Visual Languages and Computing*, 17(5), pp.466-502.
- Li, X. and Kraak, M.-J. (2008). The Time Wave. A New Method of Visual Exploration of Geo-data in Time–space. *The Cartographic Journal*, 45(3), pp.193-200.
- Maceachren, A. M., Wachowicz, M., Edsall, R., Haug, D. and Masters, R. (1999). Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods. *International Journal of Geographical Information Science*, 13(4), pp.311-334.
- Nagypál, G. and Motik, B. A Fuzzy Model for Representing Uncertain, Subjective, and Vague Temporal Knowledge in Ontologies. *Proceedings of the International Conference on Ontologies, Databases and Applications of Semantics, (ODBASE)*, Catania, Italy, pp.906-923.
- Neutens, T., Van de Weghe, N., Witlox, F. and De Maeyer, P. (2008). A three-dimensional network-based space-time prism. *Journal of Geographical Systems*, 10(1), pp.89-107.
- Neutens, T., Witlox, F., Van de Weghe, N. and De Maeyer, P. (2007). Space-time opportunities for multiple agents: a constraint-based approach. *International Journal of Geographical Information Science*, 21(10), pp.1061-1076.
- Ohlbach, H. J. Relations between fuzzy time intervals. *Proceedings of 11th International Symposium on Temporal Representation and Reasoning (TIME 2004)*, Montréal, Québec, pp.44-51.
- Pawlak, Z. (1982). ROUGH SETS. *International Journal of Computer & Information Sciences*, 11(5), pp.341-356.
- Peuquet, D. J. (2002). *Representations of space and time*, New York, The Guilford Press.
- Qiang, Y., Asmussen, K., Delafontaine, M., De Tré, G., Stichelbaut, B., De Maeyer, P. and Van de Weghe, N. Visualising rough time intervals in a two-dimensional space. *Proceedings of 2009 IFSA World Congress / EUSFLAT Conference*, Lisbon, Portugal.
- Qiang, Y., Delafontaine, M., Asmussen, K., Stichelbaut, B., De Tré, G., De Maeyer, P. and Van de Weghe, N. (2010). Handling Imperfect Time Intervals in a Two-Dimensional Space. *Journal of Control and Cybernetics*, 39(4), pp.983-1010.
- Schockaert, S., De Cock, M. and Kerre, E. E. (2008). Fuzzifying Allen's Temporal Interval Relations. *IEEE Transactions on Fuzzy Systems*, 16(2), pp.517-533.
- Smith, M. J. d., Goodchild, M. F. and Longley, P. A. (2007). *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools* Leicester, UK, Troubador Publishing Ltd.
- Stichelbaut, B. and Bourgeois, J. (2009). The Aerial Imagery of World War One: A Unique Source for Conflict and Landscape Archaeology. *Photogrammetrie - Fernerkundung - Geoinformation*, 2009, pp.235-244.
- Van de Weghe, N., Docter, R., De Maeyer, P., Bechtold, B. and Ryckbosch, K. (2007). The Triangular Model as an Instrument for Visualising and Analysing Residuality. *Journal of Archaeological Science*, 34(4), pp.649-655.
- Venn, J. (1881). *Symbolic Logic*, London, Macmillan and Company.
- Verbeke, R. (2006). De Derde Slag bij Ieper-Passendale: de Honderddaagse slag in 1917. ed. by Chielens, P., Dendooven, D. and Decoodt, H., in *De Laatste Getuige. Het Landschap van Wereldoorlog 1 in Vlaanderen*, pp.63 - 66, Lannoo, Tielt.
- Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8(3), pp.338-353.

Figure captions:

Figure 1: The transformation from the linear representation to TM. (a): The linear representation of time intervals. (b): The construction of an interval point in TM. (c): The TM representation of time intervals.

Figure 2: Temporal relations in the linear model and TM, taking *before* as an example. (a): I_{1a} , I_{1b} , I_{1c} and I_2 in the linear representation. (b): I_{1a} , I_{1b} , I_{1c} and I_2 in TM. (c): The *before* zone of I_2 .

Figure 3: Relational zones representing sets of intervals in certain Allen relations to the reference interval I .

Figure 4: The linear representation of the rough approximation of I .

Figure 5: The transformation from the linear representation to the TM representation of rough approximations of intervals. (a): Rough approximations of intervals in the linear representation. Full lines denote \underline{I} , and dashed lines denote $R(I^-)$ and $R(I^+)$. The combination of full lines and dotted lines forms \bar{I} . (b): The construction of the rough approximation of an interval in TM. (c): Rough approximations of intervals in TM.

Figure 6: An arbitrary interval set A and an arbitrary rough approximation $R(I)$.

Figure 7: An arbitrary interval set A and two arbitrary rough approximations $R(I_1)$ and $R(I_2)$.

Figure 8: Representation of spatio-temporal data in GeoTM.

Figure 9: The user interface of GeoTM, consisting of a map view (left) and a TM view (right).

Figure 10: The selection of intervals probably *before* [1916/06/25, 1917/05/31], with more than 60% probability. (a): Moving the mouse cursor to the interval [1916/06/25, 1917/05/31], and right-clicking to trigger the menu of Allen relations, and then right-clicking the 'before' option. (b): Intervals that are probably *before* [1916/6/25,1917/5/31] are selected.

Figure 11: Making temporal queries by dragging geometries in the TM view. (a): Selecting intervals *in-between* [1915/10/01, 1918/04/20] and [1916/05/10, 1917/09/10]. (b): Selecting intervals that *start-within* [1915/04/01, 1915/10/20]. (c): Selecting intervals that *end-within* [1916/06/01,1917/04/01]. (d) Selecting intervals that are longer than 2 years and shorter than 3 years.

Figure 12: The line diagram of probabilities that at least n intervals are *before* [1915/11/21,1916/08/30].

Figure 13: The rough approximation of the lifetime of a military feature.

Figure 14: The location of the study area.

Figure 15: The states of German and Allied frontlines in WWI. The map in the solid box shows the period during which the frontlines were relatively stable, while the maps in the dashed boxes show periods during which the frontlines had significantly shifted.

Figure 16: The rough approximations of the feature lifetimes in the TM view, with the *during* zones of I_1 , I_2 , I_3 and I_4 .

Figure 17: The selection of intervals that *end-within* I_2 .

Figure 18: The selection of intervals that *end-within* [1917/7/10, 1917/8/1].

Figure 19: The selection of intervals in Cluster 3.