Selecting the Content of Textual Descriptions of Geographically Located Events in Spatio-Temporal Weather Data

Ross Turner,Somayajulu Sripada,Ehud Reiter
Dept of Computing Science, University of Aberdeen,
Aberdeen, UK

Ian P Davy
Aerospace and Marine International,
Banchory, UK

Abstract

In several domains spatio-temporal data consisting of references to both space and time are collected in large volumes. Textual summaries of spatio-temporal data will complement the map displays used in Geographical Information Systems (GIS) to present data to decision makers. In the RoadSafe project we are working on developing Natural Language Generation (NLG) techniques to generate textual summaries of spatio-temporal numerical weather prediction data. Our approach exploits existing video processing techniques to analyse spatio-temporal weather prediction data and uses Qualitative Spatial Reasoning(QSR) techniques to reason with geographical data in order to compute the required content (information) for generating descriptions of geographically located events. Our evaluation shows that our approach extracts information similar to human experts.

1 Introduction

There has been increasing interest in applying NLG technology to develop systems that generate textual summaries of numerical data. Some recent examples include Babytalk [Portet2007], a system currently being developed for generating textual summaries of Neonatal intensive care data. SumTime [Sripada2003a], has been applied to produce summaries of time series data in the weather, gas turbine and medical domains. While StockReporter [Dale2003], has built upon previous work for generating summaries of stock market statistics, and Trend [Boyd1998] was developed for summarising archived time series weather data. As current applications (such as the above examples) have concentrated exclusively on time series, the increasing use and availability of low cost Geographical Information Systems (GIS) has made availability of spatial data commonplace in many scientific areas. In many domains, data varies across both spatial and temporal dimensions. Thus there is a need to develop NLG techniques to summarize spatio-temporal data.

In the RoadSafe project our long term goal is to generate textual weather forecasts for winter road maintenance application. Modern weather forecasting is largely guided by numerical weather prediction (NWP) data generated by computer simulations of weather models. In the RoadSafe project NWP data generated by Aerospace and Marine International's (AMI) GRIP[1] model contains predictions of several weather parameters (such as road surface temperature and wind speed) for several thousand geographical locations in a council area and also for several tens of time points in a day. In other words, RoadSafe data sets are large spatio-temporal data. An example data set showing the NWP data for a council area at a specific time point is shown in Figure 1. The actual input to RoadSafe consists of several such weather prediction snapshots corresponding to the different time points.

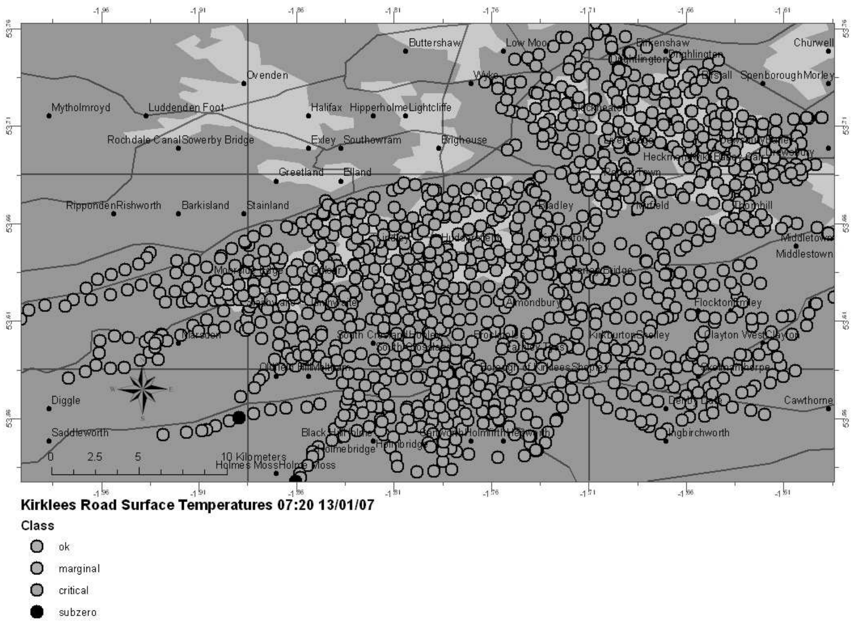

Figure 1: GRIP model data visualised in a GIS

From an analysis of human written weather forecasts we observed that summarising such spatio-temporal weather data mainly involves textually describing two types of events (weather phenomena):

- Type A event - a weather condition developing at a specific geographic location and time e.g. the surface temperature of a road falling below zero in a specific area
- Type B event - a weather condition moving across a geographic region over a period of time - e.g. a band of rain moving over a geographic area

[1] Geographical Road Ice Predictor

GRIP	Latitude	Longitude	RST 0720	RST 0740	RST 0800	RST 0820	RST 0840
UKKL1020001440	53.5293	-1.8568	0	0	0	-0.1	-0.1
UKKL1020001869	53.556	-1.8369	0.5	0.4	0.4	0.4	0.4
UKKL1020001926	53.56	-1.8333	0.5	0.5	0.5	0.4	0.4
UKKL1020001992	53.5631	-1.8319	0.3	0.3	0.3	0.3	0.3
UKKL1020002076	53.5691	-1.8308	0.3	0.3	0.3	0.3	0.3

Figure 2: Spatio-Temporal input data

Road surface temperature will become subzero [E] on high ground in the south west [L] by early morning [T]

Figure 3: A spatial description of an event

In this paper we focus on the subtask of generating textual descriptions of Type A events, and in particular, how to compute the required content from raw data for this task. Figure 2 shows an excerpt from a RoadSafe input data set and figure 3 shows a textual description of a type A event in that data. We define a description of an event in this context as an unordered set of phrases {E,L,T} where [E] describes an event in the data, [L] describes the location of the event and [T] is a phrase denoting the time that the event occurred.

The description in Figure 3 is an example of an event description generated by the RoadSafe system, a current data-to-text application we describe in Section 2. Our approach to this task and our evaluation is described in Sections 3 and 4. Section 5 provides discussion of current issues and future work while Section 6 presents our conclusions.

2 Background

The weather domain has proved a particularly popular area of application for data-to-text systems. Many notable examples exist: SumTime-Mousam [Sripada2003a], MultiMeteo [Coch1998] and Fog [Goldberg1994] have all been used commercially; however, all have concentrated on non spatial aspects of weather data or produced single site forecasts. RoadSafe is a data-to-text application currently being developed for producing winter road maintenance reports for local councils; these help to ensure correct application of salt and grit to roads and efficient routing of gritting vehicles. We are working with an industrial collaborator, AMI who are responsible for providing these reports as part of their Network Forecasting service[2].

Traditionally road weather reports are based upon data sampled at a small number of outstations at varied locations throughout a local authority or even in some cases a single site. In contrast, AMI have developed a computer simulation model called GRIP that generates detailed NWP data for thousands of points along a road network and provides the prediction data on which their road

[2]Network Forecasting is the name of the technique used by AMI to deliver their winter road maintenance forecasts.

weather forecasts are based. In GRIP, each point is sampled at 20 minute intervals throughout the day, and 9 meteorological parameters such as road surface temperature are measured every interval. The resultant output of this model is a huge amount of spatio-temporal data; for larger local authorities the model can produce output files up to 33mb in size. This poses two main problems: how can a meteorologist make full use of the available information in a limited time period, and what is the best way to present this information to users of the reports. To solve these problems we have been developing RoadSafe to automatically generate the weather reports from the GRIP model output, an example is shown in Figure 4(N.B. the textual parts in this example have been written by a human forecaster).

Issued by Aerospace & Marine International:- Valid from noon on 13/01/2007

24 Hour Forecast for Kirklees

Routes	Min RST	Time <= 0c	Ice	Frost	Snow	Fog	MaxGusts	Rain	TS
Worst/Be	-0.1 /0.9	07:20 /NA	Yes /No	Heavy	Light /No	No/No	58/52	Heavy/Heavy	No
Wind (mph)	Wind will be SW fresh to strong but touching gale force over the moors. They will very soon veer WNW but continue at similar strengths easing only slowly by the end of the night moderate but still fresh to stong over the moors.								
Weather	There will be cloud and some rain at first this evening but it will soon clear and a much colder air will arrive on the WNW winds. It will be dry with good clear spells for the rest of the night and into tomorrow morning. Roads will dip away sharply at first in the night but winds should keep most of them just above zero although almost all except urban ones will become critical and a few higher routes will dip below zero. All road should dry quickly in the wind after the rain clears.(JB)								
Route	All routes summary worst/best								
1	0.4/1.4	NA/NA	No/No	No/No	No/No	No/No	54/49	Heavy/Heavy	No
2	0.5/1.6	NA/NA	No/No	No/No	No/No	No/No	54/43	Heavy/Heavy	No
3	0.3/1.4	NA/NA	No/No	No/No	No/No	No/No	54/41	Heavy/Heavy	No
4	0.3/1.4	NA/NA	No/No	No/No	No/No	No/No	54/49	Heavy/Heavy	No
5	0.5/1.6	NA/NA	No/No	No/No	No/No	No/No	53/41	Heavy/Heavy	No
6	0.5/1.7	NA/NA	No/No	No/No	No/No	No/No	54/48	Heavy/Heavy	No
7	0.6/1.5	NA/NA	No/No	No/No	No/No	No/No	53/41	Heavy/Heavy	No
8	0.5/1.7	NA/NA	No/No	No/No	No/No	No/No	53/33	Heavy/Heavy	No
9	0.9/1.9	NA/NA	No/No	No/No	No/No	No/No	54/41	Heavy/Heavy	No
10	0.5/1.6	NA/NA	No/No	No/No	No/No	No/No	52/33	Heavy/Heavy	No
11	0.3/1.2	NA/NA	No/No	No/No	No/No	No/No	54/43	Heavy/Heavy	No
12	0.1/1.3	NA/NA	No/No	No/No	Flurries /No	No/No	58/49	Heavy/Heavy	No
13	0.2/1.4	NA/NA	No/No	No/No	No/No	No/No	54/41	Heavy/Heavy	No

Figure 4: RoadSafe system output

The reports produced by the system are structured into two distinct parts. The tabular data serves as more of an 'alert system' in order to facilitate quick decision making for users, this is achieved through colour coding of extreme values and use of traffic light colour coding to indicate the treatment requirements of a route; green for ok, orange for caution and red for critical.The wind and weather forecast texts are designed to provide a more general overview of weather conditions and highlight more complex spatio-temporal relationships in the data, such as 'some snow or sleet over high ground at first'.

3 Approach

Our Knowledge Acquisition (KA) activities in RoadSafe thus far have consisted of working with domain experts at AMI and building a parallel data-text corpus from road weather forecasts issued by them, and the associated NWP data they are based upon. Section 3.1 describes this corpus collection and analysis process while Sections 3.2, 3.3 and 3.4 describe the approach taken to abstracting content from spatio-temporal data and generating spatial descriptions.

3.1 Corpus Collection and Analysis

Meteorologists and Councils currently have access to a commercial website where they can view meteorological information such as visualisations of NWP data and weather forecasts for their respective area. At present Meteorologists at AMI use the RoadSafe system to generate the tabular part of the weather report and manually enter the textual wind and weather forecasts. The forecast parts of the report are then extracted and combined with the system input data to form a parallel data-text corpus. This forms the key element of the KA approach in RoadSafe, which is based upon a novel technique developed in [Reiter2003] and [Sripada2003c] for analysing word meaning based on aligning words and phrases to data in a parallel corpus.

The corpus currently consists of around 326 data-text pairs, the texts have been semantically annotated and our initial studies have concentrated upon the extraction and alignment of spatial phrases. This was carried out by parsing the corpus to extract individual spatial phrases along with other semantic information corresponding to the overall description of the event, such as the spatial frame of reference, time period and parameter being described. The date and county values connected with each phrase were also extracted to cross reference it with the input data file. A total of 648 phrases have been extracted and aligned with the input data so far. Table 1 shows an example of two aligned spatial phrases; the first refers to an description of the temperature of a road surface decreasing to zero, the second refers to the wind increasing in strength.

Frame of Ref.	Direction	Geofeature Combination
Spatial Phrase	'in the southwest'	'on the coast and over the downs'
County	Kirklees	Hampshire
Date	2007-01-18	2006-11-11
Time period	aftermidnight	morning
Parameters	road surface temp	wind
Event	decrease to 0	increase to fresh

Table 1: Aligned Spatial Phrases

After analysing the spatial phrases extracted from our corpus a broad classification of the spatial frames of reference experts used to refer to the locations in our domain was possible. These can be represented computationally by thematic layers in a spatial database. The relevant spatial frames of reference identified by the analysis including an example of a description of an event using each reference frame are:

1. Altitude(distinctions between areas of high and low ground) - e.g. ‘possible gale force gusts on higher ground’
2. Direction(absolute and motion) - e.g. ‘minimum temperatures around 5-6 degrees in the more Northern Routes’
3. Population(distinctions between urban and rural areas) - e.g. ‘many urban routes will drop to be critical but remain above zero’
4. Coastal Proximity - e.g. ‘a few showers at first mainly along the coast’
5. Geofeature Direction Combination(combination of direction with any of 1,3 or 4) - e.g. ‘cooler in the south east on higher ground’
6. Geofeature Combination(any combination of 1,3 and 4) - e.g. ‘Most higher level and rural roads will drop below zero later in the night’

The analysis found that the most commonly used frames of reference in this domain were altitude and direction as shown in Figure 5. Preferences for the use of certain frames of reference in descriptions of events in certain parameters were also observed, for example variation in wind speed was most often described using altitude or where the spatial domain incorporated a coastline, using a combination of coastal proximity and altitude. A similar effect was observed for population, which was mainly used to describe variation in Road Surface Temperature. It was also observed that experts use combinations of spatial frames of reference in their descriptions similar to the common GIS map overlay operation.

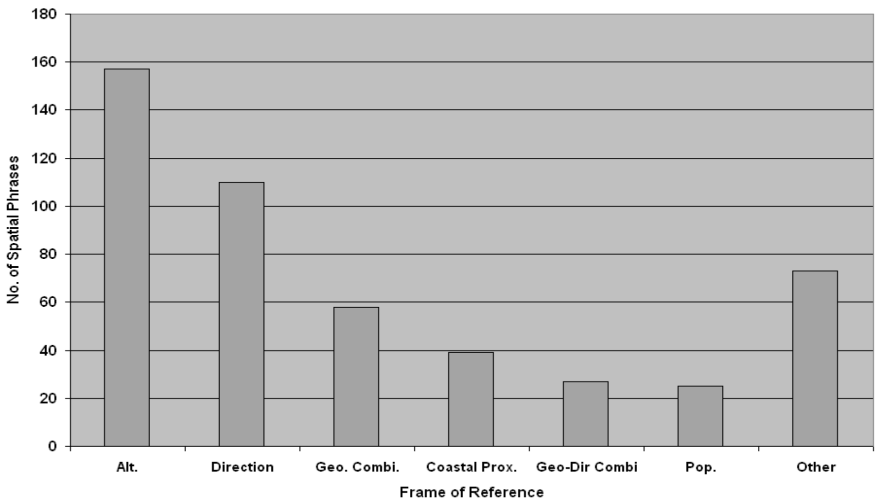

Figure 5: Distribution of Spatial Frames of Reference in the RoadSafe Corpus

3.2 Architecture

Figure 6 shows the architecture of the RoadSafe system. The modules enclosed inside the dotted rectangle are the standard NLG modules as suggested by [Reiter2000]. The two additional data analysis and spatial reasoner modules have been introduced into our work for reasons described next. While most NLG systems work with symbolic inputs with known semantics, in our case the input is numerical whose high level meaning needs to be computed by our system. The data analysis module performs this function. For example, the data analysis module computes that wind speed is 'rising' and road surface temperature is 'falling'. In other words, the data analysis module grounds concepts such as 'rising' and 'falling' in the NWP data.

It is also important that our system not only has the knowledge of weather events (rises and falls) but also has the ability to compute their location. QSR provides the ability to represent spatial information abstractly and a mechanism for reasoning with that information (cf. [Cohn2001] for an overview). This can be beneficial for interacting with GIS and representing the semantics of spatial prepositions [Cohn2001]; therefore, we incorporate a spatial reasoning module into the RoadSafe Architecture as shown in Figure 6. The spatial reasoning module is responsible for reasoning over geographic data stored in the spatial database. It's purpose is to provide functionality for other modules to compute the spatial relationship between objects in the study region and also pose spatial queries such as retrieve all the forecast points above 500m. For example, this module computes that wind speed 'rises on high ground' and road surface temperature 'falls in the southwest of Kirklees'. In the next two sections we describe these two modules in detail.

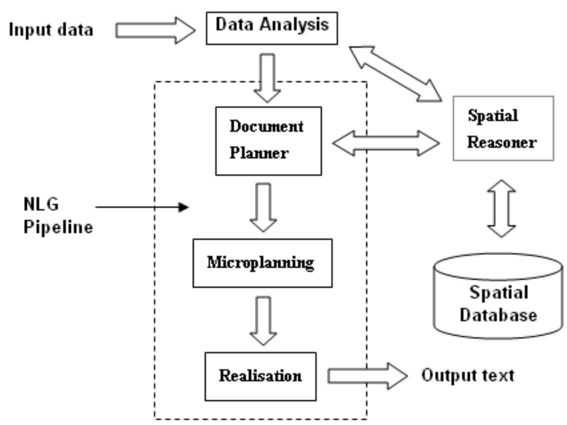

Figure 6: RoadSafe architecture

3.3 Spatio-Temporal Data Analysis for Text Generation

Spatio-Temporal data is typically represented in GIS by a series of static snapshots, analogous to a sequence of video frames. Therefore we base our approach upon structural and semantic analysis used in video processing, c.f. [Calic2005] for an overview. Our approach involves 3 key stages:

1. Low level feature description.
2. Event detection and indexing.
3. Keyframe extraction.

Low level feature description consists of applying spatial segmentation[3] (see, for example [Miller2001]) to each individual snapshot for each parameter in the data set. This is a technique which has been successfully applied in a previous prototype system developed by the authors [Turner2006]. Spatial segmentation is a two stage process, the first stage classifies points using simple thresholding based on their non-spatial attributes. We have identified significant thresholds for each parameter through our KA studies with domain experts, example thresholds for the Road Surface Temperature parameter are:

1. Ok - road surface temperature value > 5
2. Marginal - road surface temperature value > 2 and ≤ 5
3. Critical - road surface temperature value $> 0 \leq 2$
4. Subzero - road surface temperature value ≤ 0

After all points have been assigned to classes the second stage takes local density estimates of the points in each class to look for signs of clustering. Density estimation is a common spatial analysis technique in point pattern analysis [O'Sullivan2003] that measures the proportion of events at every location in the study region. In RoadSafe the sampling locations are based upon the frames of reference identified by the corpus analysis in Section 3.1. More formally, if x is a set of classified points (for example all subzero points) at time t_x, and y is a set of related reference objects within that frame of reference(e.g. all altitude contours within the study region at 100m) density is given by:

$$Density(x, y) = \frac{no.points(x \cap y)}{\sum_{i=1}^{n} no.points(i \in y)}$$

A part of a density estimation output for the example forecast in Figure 4 is shown in Table 2. After each snapshot has been spatially segmented, our method indexes sequences of snapshots that describe higher level semantic events in the data; such as, when road surface temperature is increasing or when a band of rain appears. This involves applying temporal segmentation to the series of non spatial attributes (min,mean,max) for each parameter. This helps to summarise the sequence into a smaller number of important time intervals

[3]Spatial segmentation in this sense is defined as including both clustering and classification

Frame of Reference		Proportion of subzero points				
		07:20	0740	08:00	08:20	08:40
Altitude	0m:	0.0	0.0	0.0	0.0	0.0
	100m:	0.0	0.0	0.0	0.0	0.0
	200m:	0.0	0.0	0.0	0.0	0.0
	300m:	0.0	0.0	0.0	0.0	0.0
	400m:	0.041	0.041	0.12	0.125	0.166
	500m:	0.5	1.0	1.0	1.0	1.0
Direction	Central:	0.0	0.0	0.0	0.0	0.0
	Northeast:	0.0	0.0	0.0	0.0	0.0
	Northwest:	0.0	0.0	0.0	0.0	0.0
	Southeast:	0.0	0.0	0.0	0.0	0.0
	Southwest:	0.014	0.021	0.035	0.0354	0.0426
Urban/Rural	Rural:	0.002	0.003	0.005	0.006	0.007
	Urban:	0.0	0.0	0.0	0.0	0.0

Table 2: Density Estimation Output Kirklees 13/01/2007

by approximating a time series of length n with k straight lines, where k is much smaller than n. We use the linear segmentation algorithm developed by [Keogh2001] and successfully applied in the SumTime system [Sripada2002a] for this purpose.

The final stage of data analysis is based upon the process of key-frame extraction typically used for semantic analysis in video processing. This involves extracting a set of representative timestamps describing important events in the data, such as extreme values or when significant thresholds are crossed. As an example, Figure 7 shows temporally segmented minimum road surface temperature values. The dashed lines represent values that cross class thresholds for spatial segmentation. In this case the minimum road surface temperature for the area becomes subzero at 0720 and this is shown in Figure 1, which is a visualisation of the corresponding key frame.

3.4 Generating Spatial Descriptions

The input to the NLG component of the RoadSafe system is a sequence of timestamped events for each parameter and their associated density estimations. The task of the Document Planning stage is to map from these events to an Abstract Event Description such as the one shown in Figure 8, which is the abstract representation for the event(RST $\leq$ 0) shown in Figure 7. The mapping process simply instantiates all the fields in the event description with the information from the basic event, with the exception of the Relation and Container fields. These fields are computed by performing a topological query to find the single smallest enclosing region in the spatial database.

This representation provides the input to the Microplanning module. Microplanning converts Abstract Event Descriptions into basic syntactic struc-

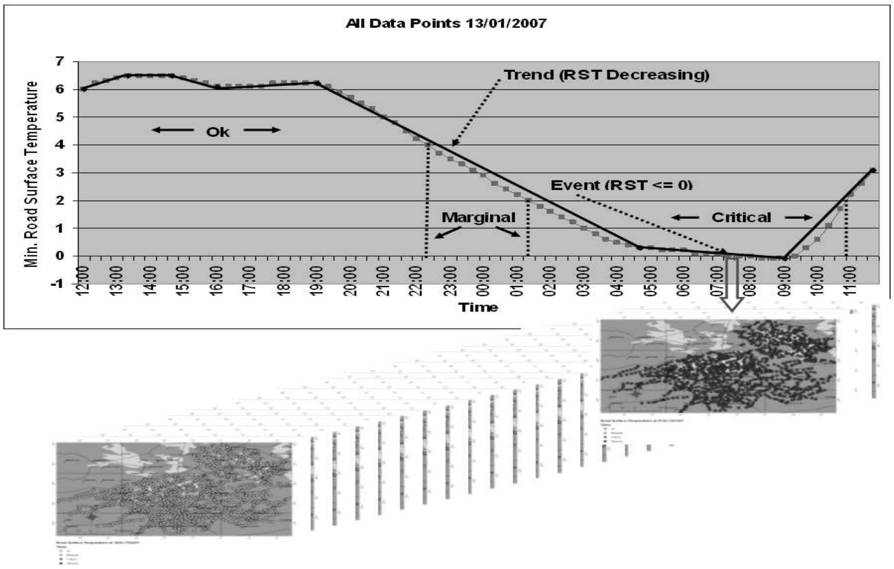

Figure 7: Key Frame Extraction

tures that are the input to the Realisation module. Realisation is responsible for converting the syntactic structures into the final output text. For example, converting the Abstract Event Description in Figure 8 would give object ‘road surface temperature’, verb phrase ‘will become subzero’,and prepositional phrases ‘on high ground’, ‘in the south west’ and ‘by early morning’. After Realisation this produces the final output text in Figure 3. While event and time phrases are simply translated by the Microplanner, spatial phrases are translated according to the corresponding density estimation. This means that the frame of reference with the highest density will be used, or as in this example, a combination will be used where no single frame of reference has a density higher than 0.8.

{Event:
    {Parameter: road surface temperature
    Type: decrease
    Threshold: subzero}
Location:
    { Density: 0.014
    Relation: inside
    Container: southwest}
Time: 07:20}

Figure 8: RST below subzero abstract event description

4 Evaluation

Our main focus in this current paper is on extracting the correct information for generating geographically located event descriptions. In an initial evaluation of our approach described in Section 3, we compared a subset of the event information extracted from the human written corpus (as described in Section 3.1) with the abstract event descriptions generated by the RoadSafe Document Planning module. This was done by taking 40 parsed phrases from the corpus and running the system on the corresponding data set. Only phrases that used a directional frame of reference, such as "‘in the north"’ were used.

The results were split in 3 categories; aligned, partially aligned and unaligned. Results were considered aligned when the Spatial Phrase,Time Period,Parameters and Event fields from the corpus data matched the abstract event description fields. Unaligned results were when the system did not generate any corresponding event description.

Aligned Corpus data	**Abstact Event Description**
Frame of Reference: Direction	Event:
Spatial Phrase: "some places in the south"	{Parameter: road surface temperature
County: Kirklees	Type: decrease
Date: 30-01-2007	Threshold: critical}
Time period: evening	Location:
Parameters: rst	{Pointratio: 0.01238390092879257
Event: decrease to critical	Relation: inside
	Container: southwest}
	Time: 17:40

Table 3: Spatial Phrase Alignment with System Results

Partial alignment occurred when there was only a minor difference observed between the location indicated by the parsed corpus phrase and the system event description. This was particularly apparent with phrases indicating north/south and east/west divides, forecasters appeared to use this distinction more flexibly than when using south west or north east for example. One expert explained that this could be due to the fact that when writing forecasts they avoid use of more specific spatial descriptions unless the pattern in the data is very clear cut. This is due to the inherent fuzziness of weather system boundaries and also to avoid ambiguity where a forecaster may not be aware of more provincial terminology used by road engineers and other users of the forecasts. This conscious effort by forecasters to avoid ambiguity when making spatial descriptions could provide an explanation for our observation in Section 3.1, where we found that altitude and direction were the main frames of reference used to make spatial descriptions in our corpus as opposed to specific place names. Table 3 shows an example of a parsed corpus phrase partially aligned with a system event description, the actual description of the event in the corpus and the description generated by the system are:

Corpus:	‘some places in the south will be critical tonight’
System generated:	‘road surface temperature will be critical in some southwestern places by early evening’

The results of the evaluation are shown in Table 4, one phrase had to be discarded due to the corresponding input data being corrupt. We found that the majority of directional phrases in the corpus did align either completely or partially align to the information being generated by the system. Differences in linguistic realisations of the event will be addressed in future versions of the system. This small evaluation provides indication that this method can be successfully applied to generate descriptions of geographically located events in raw spatio-temporal data. However, the study does indicate a need for the use of more flexible boundaries and other spatial relations.

Aligned	Partially Aligned	Unaligned
24 (62%)	13 (33%)	2 (5%)

Table 4: Evaluation Results

5 Discussion and Future Work

We have outlined a data analysis method for summarising spatio-temporal NWP data based on video processing, we have also described how we can map from the results of this method to textual descriptions of specific events in the data through interacting with a spatial database using QSR; we have focused upon generating descriptions of static events as this is a fundamental first step in reaching our goal of describing the movements of weather systems over time. For example, consider how to generate a descriptions such as ‘A weak secondary weather front will move East across the north of Kirklees tonight’ or ‘a cold front passes through the Kirklees area from the North’. In both examples we must be able to identify two basic events, where the weather system appears and disappears. We hope to extend the method described here to describe the movement of weather systems.

Another focus of extending the work presented here is the generation of more complex spatial descriptions. This issue is inherently complex and we have simplified the issue here greatly by only describing containment between 2 dimensional areas. As a consequence we have chosen to ignore more complex spatial relationships between objects such as overlap, and thus the mapping between the semantics of that relationship and spatial prepositions. There has been much work on spatial prepositions in the psycho-linguistic literature, e.g. [Vandeloise1991], [Herskovits1986]; in particular, [Coventry2004] highlight the fact that spatial description is not only influenced by an object’s geometric location in space, but also by the functions afforded by that object and its functional relations between other objects. This is of particular importance in our current domain as both the semantics of the spatial object and the parameter being described need to be considered. For example the following phrases exemplify this problem:

1. 'strong winds over high ground'
2. 'strong winds along the coast'
3. 'rain falling as snow on high ground'

Phrases 1 and 2 are an example where the semantics of the underlying spatial data type of the geographic feature in the phrase requires the use of a different preposition, 'along' for a line as opposed to 'over' for an elevated area. The use of 'over' in phrase 1 and 'on' in phrase 3 also exemplifies the same requirement due to the semantic properties of the parameter being described.

6 Conclusions

Developing NLG techniques to textually summarise spatio-temporal NWP data involves as a first step, developing techniques to generate descriptions of geographically located events. In this paper we described a method for computing the information required to generate such event descriptions by exploiting existing techniques from video processing and Qualitative Spatial Reasoning. Our evaluation which involved comparing the information computed by our method with that computed by humans (based on a parallel data-text corpus) showed that our method works well. We plan to extend the method to generate descriptions of events involving motion in spatio-temporal data.

References

[Boyd1998] S. Boyd. Trend: a system for generating intelligent descriptions of time-series data. In *IEEE International Conference on Intelligent Processing Systems (ICIPS1998)*, 1998.

[Calic2005] J. Calic, N. Campbell, S. Dasiopoulou, and Y. Kompatsiaris. An overview of multimodal video representation for semantic analysis. In *European Workshop on the Integration of Knowledge, Semantics and Digital Media Technologies (EWIMT 2005)*. IEE, December 2005.

[Coch1998] J. Coch. Multimeteo: multilingual production of weather forecasts. *ELRA Newsletter*, 3(2), 1998.

[Cohn2001] A G Cohn and S M Hazarika. Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, 46(1-2):1–29, 2001.

[Coventry2004] K. R. Coventry and S. C. Garrod. *Saying, Seeing and Acting: The Psychological Semantics of Spatial Prepositions.* Psychology Press, 2004.

[Dale2003] R. Dale. Stockreporter. http://www.ics.mq.edu.au/ ltgdemo/StockReporter/, November 2003.

[Goldberg1994] E. Goldberg, N. Driedger, and R. Kittredge. Using natural-language processing to produce weather forecasts. *IEEE Expert*, 9(2):45–53, 1994.

[Herskovits1986] A. Herskovits. *Language and Spatial Cognition: an interdisciplinary study of the prepositions in English.* Studies in Natural Language Processing. Cambridge University Press, London, 1986.

[Keogh2001] E. Keogh, S. Chu, D. Hart, and M. Pazzani. An online algorithm for segmenting time series. In *Proceedings of IEEE International Conference on Data Mining*, pages 189–196, 2001.

[Miller2001] H. J. Miller and J. Han. *Geographic Data Mining and Knowledge Discovery.* Taylor and Francis, 2001.

[O'Sullivan2003] D. O'Sullivan and D. J. Unwin. *Geographic Information Analysis.* John Wiley & Sons, 2003.

[Portet2007] F. Portet, E. Reiter, J. Hunter, and S. Sripada. Automatic generation of textual summaries from neonatal intensive care data. In *11th Conference on Artificial Intelligence in Medicine (AIME 07)*, pages 227–236, 2007.

[Reiter2000] E. Reiter and R. Dale. *Building Natural Language Generation Systems.* Cambridge University Press, 2000.

[Reiter2003] E. Reiter, S. Sripada, and R. Robertson. Acquiring correct knowledge for natural language generation. *Journal of Artificial Intelligence Research*, 18:491–516, 2003.

[Sripada2002a] S. Sripada, E. Reiter, J. Hunter, and J. Yu. Segmenting time series for weather forecasting. In *Applications and Innovations in Intelligent Systems X*, pages 105–118. Springer-Verlag, 2002.

[Sripada2003a] S. Sripada, E. Reiter, and I. Davy. Sumtime-mousam: Configurable marine weather forecast generator. *Expert Update*, 6:4–10, 2003.

[Sripada2003c] S. Sripada, E. Reiter, J. Hunter, and J. Yu. Exploiting a parallel text-data corpus. In *Proceedings of Corpus Linguistics*, pages 734–743, Lancaster, UK, 2003.

[Turner2006] R. Turner, S. Sripada, E. Reiter, and I. Davy. Generating spatio-temporal descriptions in pollen forecasts. *EACL06 Companion Volume*, pages 163–166, 2006.

[Vandeloise1991] C. Vandeloise. *Spatial Prepositions: : a Case Study from French.* University of Chicago Press, 1991.