# ANALYSING AND AGGREGATING VISITOR TRACKS IN A PROTECTED AREA

E. S. Dias [a, *], A. J. Edwardes [b], R. S. Purves [b]

[a] Vrije Universiteit (FEWEB, RE, GIS), De Boelelaan 1105, 1081 HV Amsterdam, Netherlands - edias@feweb.vu.nl
[b] Dept. of Geography, University of Zurich, Winterthurerstr. 190 CH-8057 Zurich, Switzerland - (aje,rsp)@geo.unizh.ch

**KEY WORDS:** Location-based information, spatio-temporal analysis, geographic behaviour, geo-data-mining, positioning uncertainty

**ABSTRACT:**

The advent of location-based technologies deployed in protected areas provides both visitors and managers of such areas with new opportunities. In this paper we investigate the potential for mining individual tracks of visitors' geospatial lifelines to both extract information describing aggregated patterns of group behaviour and characterise individual actions. Methods to spatio-temporally cluster individual behaviour and identify potential locations for specific actions (e.g. do visitors stop here to look at wildlife), whilst handling uncertainty in location, are described and applied to test the hypotheses that firstly, visitor behaviour is altered by the provision of information, and secondly whether the mode of information provision (e.g. in the form of a paper map or though an location-based service) influences visitor behaviour. The results of experiments with 140 visitors to a nature trail on the island of Texel in the Netherlands show statistically significant differences in time spent at locations where information was "pushed" to the visitors.

## 1. INTRODUCTION

### 1.1 Motivation

In recent decades, recreational use of natural areas has grown rapidly from low intensive and relatively passive use to a situation where tourism is the dominant force driving change in many rural areas and their associated communities (Butler et al., 1998). However, excessive use of natural areas can have significant direct and indirect negative impacts. These include both environmental degradation (Farrell and Marion, 2001) and a diminishing of the quality of visitors' recreational experience (Lynn and Brown, 2003). Mobile Information Services have been suggested as one means of supplying park managers with the possibility to monitor and manage visitor distribution within parks and, concurrently, help visitors achieve a fuller awareness of the richness of natural and cultural resources they visit. In this paper we analyse data collected using the prototype of such an information tool and assess its usefulness in monitoring and influencing the whereabouts of the visitors.

### 1.2 Research context

Location-Based Services (LBS) allow access to information for which the content is filtered and tailored based on the location of the user. We tend to spend the majority of our time in known or familiar environments, where we either do not require information or know where to obtain it. LBS may therefore be particularly useful in tourism and leisure sectors where visitors are both eager for information and unfamiliar with a locale (Dias et al, 2004). LBS can provide a wide variety of useful information, for example, answering questions such as:

- What birds of prey can be found here? (presence)
- Where can Sea Holly be found? (distribution)
- Can orchids be found in these dunes? (confirmation)
- Are these Elderberries? (identification)
- Are these lichens always found on southerly dune slopes? (association)

(Edwardes et al, 2003).

In the context of this work, previous research from three different domains is relevant: that exploring how users behave and impact upon natural spaces; techniques to analyse GPS tracks from individual users and methods to visualise, explore and analyse large volumes of so-called moving point objects.

Previous research addressing issues of visitors' spatial distribution and behaviour within natural areas has been carried out from the context of crowding, visitor density and visitor simulation modelling (Elands and van Marwijk, 2005; Manning 2005). Such research is typically centred within the field of recreation management, and aims, for example, to model the carrying capacities of natural areas.

As technologies allowing tracking of individual paths have developed, researchers have started to apply conceptual research concerned with the analysis of space and time (e.g. the space-time aquarium suggested by Hägerstrand (1970)). However, as real, high volume data describing geo-spatial lifelines (Mark, 1998) have become available the inadequacies of techniques such as the space-time aquarium as more than a simple visualisation tool for a limited number of paths have also become apparent (Kwan, 2000).

These limitations have in turn led to the emergence of so-called Geographic Knowledge Discovery Techniques (for a full review see Laube et al., 2006) which seek to allow both the qualitative and quantitative exploration of motion tracks. Laube et al. (2005) introduced a set of methods for analysing relative motion in groups of objects, while Mountain and MacFarlane (in press) discuss methods for predicting an object's likely

* Corresponding author.

position based on previous fixes and describe examples uses such as the filtering of queries to a Geographic Information Retrieval system.

One of the key limitations identified by Laube et al. (2006) is the lack of availability of real data with multiple geo-spatial lifelines for analysis. For this work, we collected data specifically to allow exploration of the behaviour of visitors to a natural area, thus overcoming this problem. By contrast to previous work, park users were constrained to the same path, with few chances to leave the network, thus vastly simplifying the role of space in our work, and allowing us to focus on users' behaviour along this constrained track. We developed a set of techniques aimed at investigating how the spatial behaviour of visitors to a protected area changes in response to information being supplied to them in differing forms. This problem is framed within the following research questions:

- How can tracks of multiple visitors to a park be used to explore visitor behaviour?
- Is the geographic behaviour of visitors altered by the provision of information?
- Do different forms of information media alter the geographic behaviour of visitors?

## 2. METHODOLOGY

### 2.1 Experimental design

A controlled experiment was designed to measure the influence that location-based information had on the behaviour of visitors to natural areas. In the experiment all subjects were issued with GPSs which recorded their positions regularly and divided into control and test groups. The test groups were each issued with different forms of information, ranging from location-based services to traditional paper-based information. The control group were provided with no additional information. The tests were carried out between August 22 and September 9, 2005.

**2.1.1 Study area:** The National Park "Dunes of Texel" located on an island in the north of the Netherlands served as the testing ground for this work. Part of the dune park is only accessible via the EcoMare museum and visitor centre, which is visited by a large number of tourists during the summer period. EcoMare together with Camineo Systems and Geodan b.v. developed a location-based service to serve the visitors to the dune park. This system has two main components:

1) A cross indicating the exact location of the visitor while walking in the dune park on a map.
2) Information content is pushed to the visitor when they are at specific locations. A soft cuckoo-song-sound is emitted by the device at these locations and the relevant content page is automatically shown.

Random visitors to the EcoMare museum were approached and asked if they would be interested in participating in this research. In order to test four different information media, the test subjects were divided among four groups: *No information*, *Paper booklet*, *Digital information* and *LBS*. The *No information* group (the control group) were given no information during their visit, while for the other three groups (the test groups), all subjects had access to the same information, but delivered using different media. The group "Paper booklet" carried the information during the visit in a booklet. The group "Digital info" accessed the information via a Personal Digital Assistant (PDA). The "LBS" group also had access to the information in the PDA, but enhanced with the location sensitivity explained above.

The composition of the groups was controlled to ensure their profiles were as similar as possible. In addition, all subjects set out to follow the same route, in similar weather conditions. A GPS receiver was given to every participant irrespective of the group they were in. GPS tracks were recorded at a rate of one position fix every *five* seconds in order to analyse the subjects' spatial behaviour.

**2.1.2 Information content:** The information provided to the test groups subjects comprised of a map of the route with the locations of a number of Points-of-Interest (POI) displayed (see Figure 1). Detailed information about each of these was supplied in the subsequent information. This content consisted of a prominent title, a photo of the feature and a text description. The POIs were classified into four categories: "Directions" (indicating the path the subject should follow); "Plants" (information about a particular plant visible from the path); "Animals" (information about animals relevant at a particular point of the path) and "Landscape" (information about landscape features visible from a certain location).


Figure 1. Map of the trail given to visitors.

### 2.2 Analysis techniques

The passage of each visitor traversing the dune park was recorded by a unique GPS track. Whilst analysis of these tracks independently could yield valuable information about individual movements, the purpose of the analysis here was to investigate whether significantly different behaviours occurred across groups as a result of the introduction of information in different forms.

As such, our first task was to develop a method to aggregate the data. As shown in Figure 2, GPS tracks vary both as function of the precision of the device and differences in subject behaviour. The main types of variability include:

- Uncertainty introduced by imprecision in the GPS coordinates recorded;
- The visitor leaving the prescribed path;
- Missing GPS data for periods of traversal;
- Individual differences in walking pace; and
- Differences in the period of time spent stopping at particular locations.

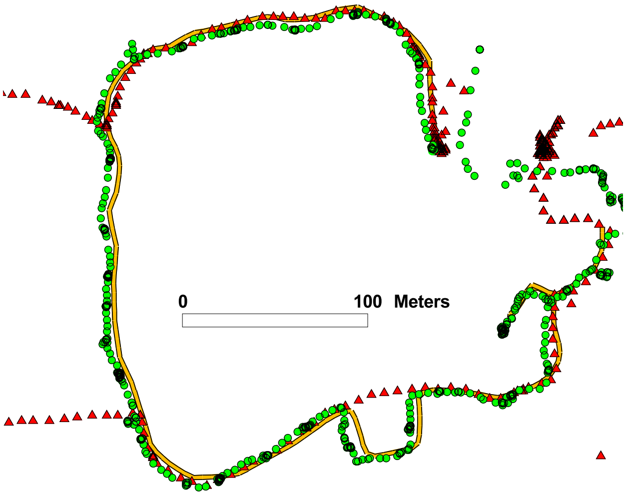


Figure 2. Example of GPS Tracks for two visitors superimposed on the digitised path

In order to allow data analysis two main methods were employed: linear referencing and aggregation. The purpose of linear referencing was to associate all individual GPS fixes with a single common baseline. In our case, the path provided the obvious reference to perform this function. It was therefore extracted as a linear geometry using a 1:10,000 topographic base map (the TOP10 vector dataset of the Dutch National Mapping Agency). GPS fixes were referenced by projecting them onto their closest path position. Aggregation involved the definition of a sampling frame segmenting the path, into which the referenced positions could be aggregated. To achieve this, the path was indexed at five meter intervals and the number of fixes occurring in each interval recorded. The size of the interval was chosen because it reflected the approximate precision of the GPS receivers.

A number of issues were encountered in performing these tasks. During aggregation, situations were found where the GPS fixes were not representative of the visitor's movement along the path, with for example, fixes occurring a considerable distance from the path. To handle these situations a filter was employed to reject fixes that were projected over a distance of more than ten meters. This value represented twice the theoretical GPS precision and was validated by visual inspection of the tracks.
A second problem was that at one point the path forked taking visitors up to a viewpoint, indicated by the POI labelled 34 in Figure 1. This presented a difficulty in defining a single linear reference. To handle this, the stretch of path leading to the viewpoint was duplicated within the linear reference, once for each direction. The closest fix to the viewpoint, measured along the path, was then used to discriminate which of the duplicated path segments should be referenced. Fixes within the segment that occurred before the closest position were assigned to the first segment and those thereafter to the second.

Two additional aggregations were also performed to consider sources of error that might influence the data quality. To investigate the errors arising from the two different GPS receivers used, the dispersion of fixes allocated to each interval was recorded. This involved computing the centroid of the fixes assigned to a particular interval and the mean distance of the points to this centroid. To consider errors in the digitisation of the path, the average projection distance to an interval for every segment was also calculated. This value was signed according to the side of the path that the fixes fell on.

After indexing each valid fix to its corresponding path interval, fix frequencies were calculated for each interval. Using these results, the tracks were graphically visualised and statistically analysed. One issue emerged from this analysis: for a particular track, an interval could have zero recorded fixes. This situation could be indicative of one of two possibilities, either the visitor had moved rapidly through the five meter interval and there were truly no fixes, or there was no data available for the segment due to receiver issues. Since it was relatively unlikely that a visitor could move fast enough that there were no fixes over more than two segments(since the frequency of fixes was 5 seconds, this would represent a speed of more than 7km/hr), consecutive intervals with no fixes were selected and their values set to null. The average number of fixes on each interval for each visitor was calculated and used as a measure of time spent at an interval. Aggregated values for each information medium were also calculated and allowed for inter-group comparisons.

## 3. TRACK ANALYSIS

### 3.1 General observations

The main goal of this research was to uncover differences in the spatial behaviour caused by the provision of different information media to visitors of protected areas. The characterisation of behaviour was simplified into the variables time and place. This simplification was implemented by linearising the space, dividing it into consecutive five meter segments and calculating for each segment the time the visitors spent there. When the visitors spent 15 seconds or more in a segment, then it was considered that they either stopped or significantly slowed down there.

Table 1 and Table 2 summarise the overall influence that the different information media have on the behaviour. Table 1 shows the average time each group spends per interval. This value is indicative of the overall time spent in the park, therefore we can conclude that the technology has some effect since it is visible that visitors who had access to information via the PDA (the digital and the LBS groups) spent on average more time (around 45%) than the other groups (the no info and paper groups). The maximum amount of time that a visitor has spent on a certain segment is also displayed in the same table - for all groups, visitors can be found that have spent long amounts of time in a segment (more then 10 minutes for a visitor with the digital info and more then 20 minutes for visitors in all the other groups). These values are indicative of activities such as picnicking or reading.

| | Mean (sec.) | SD. (sec.) | Min. (sec.) | Max. (min.) | N (#segs) |
|---|---|---|---|---|---|
| No info | 7.3 | 27.5 | 0 | 23 | 4999 |
| Paper | 8.7 | 22.2 | 0 | 23 | 6684 |
| Digital | 11.9 | 24.7 | 0 | 12 | 6896 |
| LBS | 11.3 | 21.6 | 0 | 20.8 | 12228 |

Table 1. Time statistics regarding the time the user spends per segment.

Table 2 indicates the number of stops each visitor made during their visit, averaged over the group. A stop was defined as when a visitor spends 15 consecutive seconds (or more) in the same segment. The visitors without information, the control group, stop on average in 16.6 places. For the visitors with paper information, the average number of stops increases to 26.6. For the visitors with access to digital information, the average number of stops increases to 39.2. Finally, the visitors receiving location-based information stop on average 48.6 times.

| | Mean | SD. | Min. | Max. | N |
|---|---|---|---|---|---|
| No info | 16.6 | 10.5 | 0 | 42 | 38 |
| Paper | 26.6 | 17.7 | 3 | 82 | 49 |
| Digital | 39.2 | 15.0 | 15 | 69 | 46 |
| LBS | 48.6 | 14.6 | 16 | 85 | 75 |

Table 2– Average number of stops (15 seconds or more in a certain place) per visitor per group

These results suggest that the number of stops increases based on the increasing complexity of the information delivery mechanism.

### 3.2 Visual analysis of results

The previous results demonstrate the influence of information in the number of stops, but we also wanted to analyse where the stops occur and if these stops are correlated in space. Figure 3 shows the information on spatial behaviour for all the segments and for all the visitors grouped by information medium. POIs are shown at the top of the figure, indicating places where visitors were provided with information. Information categories are shown at the bottom of the figure using the same pictograms as in Figure 1. In order to simplify the visual analysis, segments were classified according to the time spent at the segment into four classes: rest locations (more than 2 minutes at location; long stops (between 30 seconds and 2 minutes at location); short stops (15 – 30 seconds at location) and walking. The segments for which there is no data collected (due to either extreme inaccuracy of the GPS receiver or to the visitor taking a shortcut) were given a nod data value. This method of presenting the data drew on the technique for identifying relative motion patterns suggest by Laube et al. (2005). The visualisation reveals the stops that are spatially autocorrelated among the visitors, these are indicated by the darker vertical bars. The smeared areas (where the darker cells are not aligned along vertical structures) are indicative of low autocorrelations. This figure is also helpful in revealing shortcuts where the visitors did not take the correct path. Two areas of common shortcuts are clearly visible in the second half of the path, indicated by continuous missing data for about 13 segments. Scattered missing values that are not correlated in space (not vertically aligned) are due to GPS inaccuracy if they occur singly, or if temporally autocorrelated (i.e. horizontal bands of null values) indicate individual users leaving the path. Figure 3 also indicates "natural" stopping places where all groups stop irrespective of the information medium. An interesting observation is the fact that the group with location-sensitive digital information appears to display more correlated stopping places (clearly defined red bars).

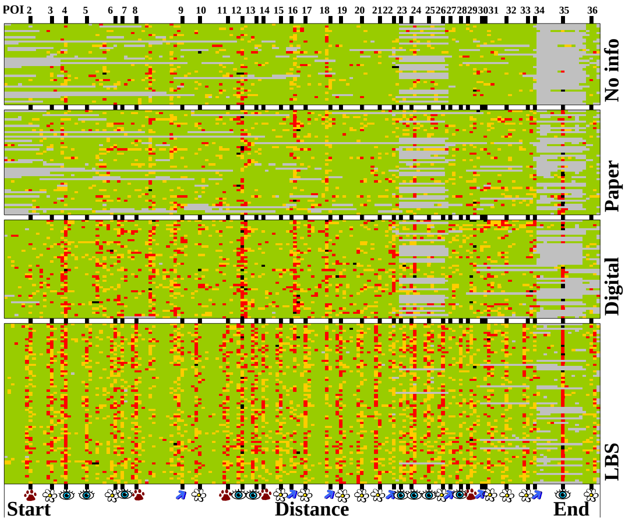


Figure 3. Visualisation of the frequency of fixes per interval of path for every track grouped by information type.

Legend:
- Walk - until 15 seconds in the segment (3 or less fixes)
- Short stop - 15 to 30 seconds in the segment (3 to 6 fixes)
- Long stops – 30 sec to 2 min in the segment (6 to 24 fixes)
- Rest - more than 2 min in the segment (more than 24 fixes)
- No data

These data were then averaged according to information media and then plotted along the path in order to visualise the coordinated stops in space (Figure 4).


Figure 4. Average number of fixes per interval shown along the path for each information medium: a) No info; b) Paper booklet; c) Digital info and d) LBS.

Figure 4-a) shows that for the visitors with no access to information, there are, nevertheless, places that were common stopping points. This is indicative that the control group does not move at a constant pace along the entire route. It is also

noticeable that most of the stops defined by the control group are also to be found in the other groups. A visual analysis of the aggregated tracks shows little difference between the control group (Figure 4-a) and the paper booklet group (Figure 4-b). Although the digital info and the LBS groups show some similarities the LBS group in particular has more stopping points and these stopping points are more uniformly scattered along the path.

### 3.3 Analysis of errors

As introduced in the methodology, the collected data (GPS fixes for moving visitors) had different possible sources of errors and uncertainty, primarily related to GPS positional error through canyoning effects and multipath reception, and the representation of the base path (on to which the fixes were being projected). In order to visualise these errors and identify biases or systematic errors in the data, Figure 5 was produced. It presents for all the visitors' tracks (grouped by information medium) and for all segments, the average distance of the fixes to the base path. This distance was classified as positive for the fixes measured on the left side of the path and as negative for the fixes measured on the right side of the path. Systematic error or GPS biases can be identified in the figure as the spatially autocorrelated bands of colour (the same colour vertically aligned), meaning that on those specific segments, all points for all tracks were being measured either on one side of the path or on the other.

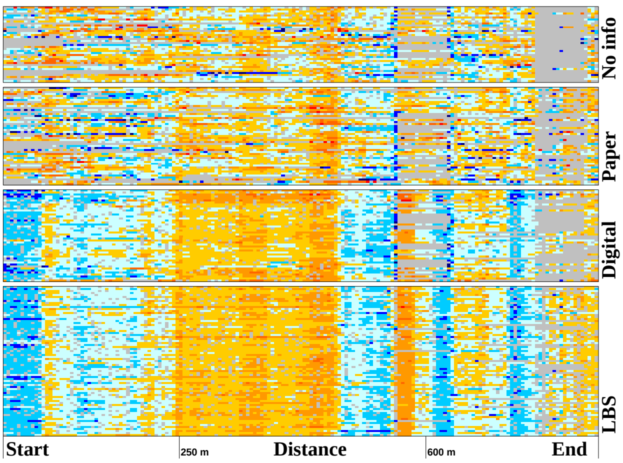


Figure 5. Visualisation of the average distance to the path for all fixes within a single interval for each track, grouped by information type.

Legend: Left side of the path: Right side of the path:

| Left side of the path | Right side of the path |
|---|---|
| < 3 meters | < 3 meters |
| 3 to 6 meters | 3 to 6 meters |
| 6 to 9 meters | 6 to 9 meters |
| > 9 meters | > 9 meters |
| | No data |

Figure 5 also enables the identification of differences in the degree of uncertainty between the two types of GPS receivers used. The positional information for the non-Tech groups (no info and paper booklet groups) was collected using a handheld Garmin12 GPS unit and for the Tech groups (the digital info and the LBS groups) positional measurements were made using a Bluetooth Globalsat receiver. The visitors from the non-tech groups show less autocorrelation than the tech groups, suggesting that the uncertainty related to the Garmin12 receiver is greater than for the Globalsat receivers. The spatial autocorrelation, for the information collected with the Globalsat receiver, is also much more apparent (vertical alignment of the same colour patches). Figure 6 displays the distance data averaged and aggregated to path segments for each receiver. The average variance of the location data, represented by the delimiting lines on both sides of the path is also shown. The variance was calculated as the mean radius of fixes per segment interval. To compute this, the mean position (centroid) of all fixes falling in a given interval was first calculated. The resulting point was therefore independent of the geometry of the interval itself. The variance was then given by the mean of the distances between each fix and this centroid.

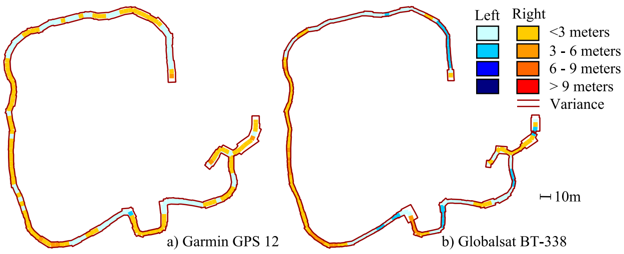


Figure 6. Average distance of fixes to the path with, outline showing mean variance amongst fixes allocated to each interval. Results are aggregated by GPS receiver a) Garmin GPS 12. b) Globalsat BT-338

It can be observed in Figure 6 that this variance is generally consistent in width along all the segments of the path for each receiver taken independently. The exceptions (segments where the variance is much greater) can all be explained by shortcuts (places where the visitors took a different way and therefore distanced themselves from the path increasing the variance level). It can also be observed that the variance is higher overall for the Garmin GPS 12 receiver, compared to the Globalsat BT receiver. This is a reflection of differences in the positional error between the devices. Overall, Figure 6b shows a source of errors that is accountable to digitisation (the path is shifted) rather than uncertainty in the GPS fixes. This is indicated by the fact that the distance values, which also consider the side of the path fixes fall on, contain autocorrelation. However, since the variance of the GPS error is constant along the path, we can conclude that this autocorrelation must be due to a mismatch between the path on the ground and the digitised path. This divergence is less apparent for the Garmin receivers because the positional error of the fixes there are in a similar range to that of the positional error of the path digitisation (Figure 6a). The uncertainty analysis (variance and distance to the path) also allows validation of the method used in projecting points to segments. Figure 7 shows a histogram of the frequency of distances measured from the path in assigning individual fixes, which shows a normal distribution centred on the path itself.

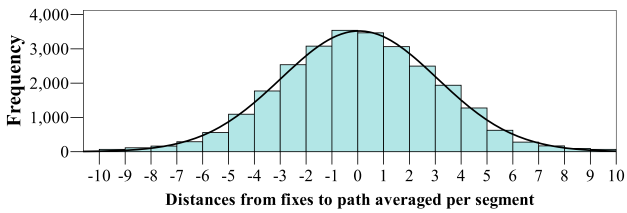


Figure 7. Histogram of average fix-path distances for the intervals with normal distribution curve. Mean = 0.05; Std Dev. = 3.025; Nvalid = 26702; Missing = 8450; Valid Percent = 76%)

Such results give confidence in the choice of both buffer size (10m) and segment length (5m) and indicate that the potential positional and digitising errors did not significantly affect the location counts and the resulting classifications.

### 3.4 Statistical analysis

In this section we set out to quantify the influence that information and its delivery mode has on movement behaviour of visitors. In an attempt to create "artificial" stopping places, information was provided to the three test groups (paper booklet, digital info and LBS), this information was relevant to the locations along the path indicated in Figure 1.

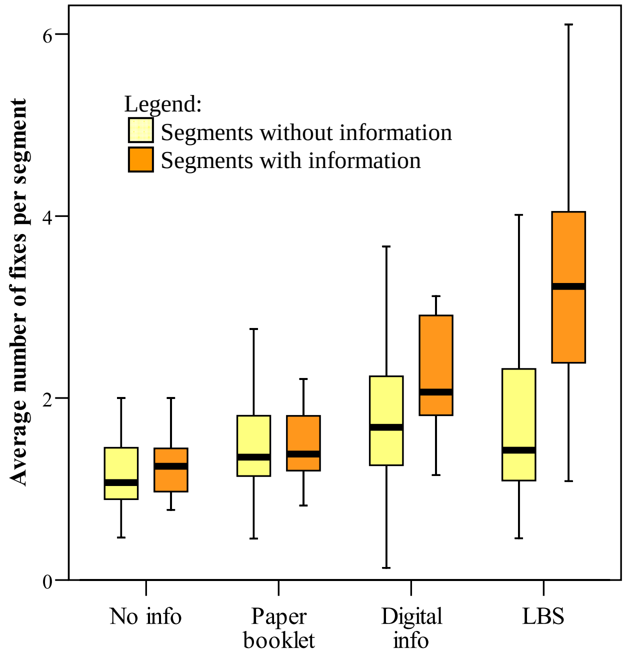


Figure 8 Box plot of average number of fixes per path segment grouped by information medium and whether the interval was related to a POI location or not.

Figure 8 illustrates the average number of stops per segment for each information type, classified according to whether locations were POIs or not. Both the No info and the Paper booklet groups spent roughly the same amount of time at all segments on the path. This finding was expected for the No info group because these visitors do not have knowledge of the information at certain segments, but is more surprising for the Paper booklet group where it was expected that the visitors would spend more time at the POIs exploring these places and the information. By contrast, the group issued with Digital info show a significant difference in their behaviour at POIs, even though the only difference between them and the Paper booklet group was in the method of information provision. Finally, the LBS group displayed similar behaviour to the Digital info group, once again spending significantly more time at POIs. These results suggested that the method of providing information had an influence on visitors' behaviour. In a second step, we wished to examine whether the type of information also influenced behaviour. As explained in section 2.1.2, the information available could be classified into four categories (POIs related to Navigation, Animals, Plants and Landscape).

Table 3 presents the results of four binary logistic regressions between stops (defined as more than 15 seconds in a segment) and four information types that originated four different spatial behaviours. In the first column, below the information type, are the overall model statistics. $\chi^2$ and *M.Sig* are the chi-square statistic and its significance. They result from the Omnibus Tests of Model Coefficients and measure how well the model performs. Only the model for the LBS group has a high performance, meaning that the stops and the information provision places are correlated for this group. For the other groups, a correlation could not be found. *N* is the number of valid segments included in the regression and the *Nagelkerke* $R^2$ is an approximation of the proportion of the variation in the response that is explained by the model (comparable to the $R^2$ in linear regressions). As expected, the LBS information provision explains a bigger proportion of the stops than any of the other groups. Also presented in Table 3 are the specific results for the variables performance within the models. *Exp(B)* is the predicted change in odds for a unit increase in the predictor. The *Wald* and *Variable Sig.* columns provide the Wald chi-square value and 2-tailed p-value used in testing the null hypothesis. Coefficients that have *V. Sig.* (p-values) less than alpha=0.01 are statistically significant at 1% level.

| Spatial behaviour | POI category | Exp(B) | Wald | V.Sig. |
|---|---|---|---|---|
| **No info**<br>$\chi^2$ = 9.029<br>M.Sig = 0.060<br>Nagelkerke $R^2$ = 0.154<br>N = 166 | Navigation | 0.000 | 0.000 | 0.999 |
| | Animals | 0.000 | 0.000 | 0.999 |
| | Plants | 0.000 | 0.000 | 0.999 |
| | * Landscape | 8.929 | 7.364 | 0.007 |
| **Paper booklet**<br>$\chi^2$ = 5.328<br>Sig = 0.255<br>Nagelkerke $R^2$ = 0.086<br>N = 169 | Navigation | 0.000 | 0.000 | 0.999 |
| | Animals | 0.000 | 0.000 | 0.999 |
| | Plants | 0.000 | 0.000 | 0.999 |
| | Landscape | 3.938 | 2.478 | 0.115 |
| **Digital info**<br>$\chi^2$=5.026<br>Sig = 0.285<br>Nagelkerke $R^2$ = 0.049<br>N = 169 | Navigation | 0.897 | 0.010 | 0.922 |
| | Animals | 0.000 | 0.000 | 0.999 |
| | Plants | 0.978 | 0.001 | 0.978 |
| | Landscape | 3.587 | 3.449 | 0.063 |
| **LBS ***<br>$\chi^2$=33.688<br>Sig = 0.000<br>Nagelkerke $R^2$ = 0.268<br>N = 169 | Navigation | 0.000 | 0.000 | 0.999 |
| | * Animals | 19.304 | 6.728 | 0.009 |
| | * Plants | 5.630 | 8.250 | 0.004 |
| | * Landscape | 19.304 | 12.935 | 0.000 |

* significant at the 1% level

Table 3. Logistic Regression results for the influence of POI push positions in the spatial behaviour, represented by stops (longer than 15 seconds, freq $> = 3$).

For the control group, who were given no information, there is none the less a significant correlation with the Landscape POIs – this suggests that these POIs are in locations where park users might naturally stop. For both groups who were provided with information passively, no significant correlations were found. Finally, the group who were pushed information show significant correlations with all POIs except for the navigation information. It is suggested that this is because when pushed information, users stop to read it. However, at navigation points given the simplicity of the route the users were on, it was not necessary to travel significantly slower.

# 4. DISCUSSION

## 4.1 Handling large volumes of spatio-temporal data

In order to obtain knowledge of the spatial behaviour of visitors, it is necessary to capture fine-grained spatio-temporal data, but the collection of this high resolution data leads to an problem in itself: Individual tracks contain too much variation (in terms of data quality *and* actual movement) to allow direct inter-track comparisons of spatial behaviour between them. To deal with this issue, several techniques were applied to extract useful information and identify trends. The first step was to define when to accept or reject data as a valid measurement. To do this a distance-based filter was applied, such that only the points close enough (within 10m) to the path were considered. The choice of tolerance was validated by analysis of the data (see Figure 7). The second step involved the aggregation of the data to a common baseline, i.e. valid GPS tracks were *warped* on to the path. This technique allowed the high variability of the tracks to be handled by referencing them all to a common baseline. In addition, because often the data sets were not complete (due to inaccuracies of the GPS receivers or to visitors' shortcuts), the analysis was not performed over the full tracks (which would require complete datasets). Instead, the data were analysed by averaging them over single path intervals which allowed null values to be ignored. It was still necessary to characterise such errors in order to contextualise the effects of them on the results and analysis. To achieve this various visualisation methods were employed (Figure 5 and Figure 6).

## 4.2 Observations on spatial behaviour

Providing visitors with information was expected to have an influence on their spatial behaviour. Comparing only the no information and paper information groups there is some evidence to support this hypothesis though it is far from compelling. The average number of stops >15 seconds, shown by Table 2, is significantly higher (T-test p>0.001). However, the visual difference in the patterns shown in Figure 3 and Figure 4 is negligible. More importantly, the interpretation of box plot Figure 8 indicates little difference in behaviour, both between the groups and between the segments with and without information for the paper group. Likewise, the Logistic Regression shown in Table 3 was unable to find evidence that the positions of POIs were influencing stopping behaviour for this group.

A difference in behaviour between the digital info and the paper groups was not expected since the information content was identical, but was found. The visitors with the digital info not only stopped more (see Table 2) overall, but the places they stopped at were correlated along points of the path not investigated by the paper group. This can be seen in Figure 4. However, interpretation of the box-plot in Figure 8 would suggest this difference should not be stressed too strongly. Indeed the Logistic regression shown in Table 3 was unable to correlate the places that visitors stopped in with the POI information for the digital information group. Two reasons can be hypothesised to explain these finding: 1) the visitors from this group needed to interact more when handling the device, causing them to stop more and 2) the technology had a "novelty effect", i.e. the visitors were more motivated to explore the information because it was presented in a media that was unfamiliar to them.

It is important to consider the potential impact of granularity – for example the sensitivity of the results to the chosen length of stopping time (15 seconds) – and further work is required to explore this issue. Equally the chosen segmentation length (5m) and GPS sample rate (5s), although to some extent validated by the experiments on GPS uncertainty, is another example of variable granularity whose influence on the results should be explored. Previous work from Laube and Purves (2006) has shown that seemingly significant results can be artefacts produced as a function of granularity.

In terms of the overall results, it was possible to observe a clear difference between the non-tech (the no info and the paper booklet) groups and the tech groups (the groups that accessed the information via a PDA. One can assume that this difference indicates that the technologies have an intrusive effect on the behaviour of visitors. Although both tech groups spent more or less the same amount of time on the route (see Table 1), two main differences were observable. The visitors with LBS information stopped more (see Table 2). Visual inspection of the data presented in Figure 3 and Figure 4 clearly shows more frequent autocorrelated stops for the LBS group when compared with the other groups. In addition, Figure 8 indicates that there is a significant difference in behaviour around path segments where the POIs were positioned and those without information and the Logistic Regression of Table 3 is able to detect that this behaviour is significantly influenced by the animal, plant and landscape POIs. These findings indicate that location-sensitive information provision can alter the spatial behaviour of the visitors. In terms effectiveness in behaviour altering, of the type of POIs, plants and animal information seem to cause "un-natural" stops, since the Landscape information POIs appeared to be natural stopping points anyway, as shown by the correlation with the stopping points of the No info group (see Table 3). Therefore, information about plants and animals can lead people to explore the park in a different way. Information about plants *at the right place*, for example, can lead people to direct experiences of nature, stopping to see plants about which they are receiving information.

## 4.3 Influencing behaviour towards sustainability

The collection of anonymous-aggregated movement data allowed two additional qualitative behaviour analyses: 1) do visitors leave the trail and trample the protected dunes and 2) if visitors accept the park management advice to visit particular places. Regarding the latter, the information provided to the three information groups intended to help the visitors fully explore and become more aware of the park's natural richness (e.g. it recommended the visitors to walk through a south loop [POIs 23-26] and to see a breathtaking park (over)view by climbing to a dune top [POI 35]). The spatial data shows that for the *Paper booklet* group, 43% did not walk through the loop, 39% did not see the viewpoint and 31% went off-path in one or more places. The results were even more alarming for the *Digital info* group, where 46% took the shortcut, 59% did not visit the viewpoint and 22% were off-path at least once. Significantly different results were obtained for the LBS group, where only 4% took the shortcut, 20% did not visit the viewpoint and only 7% were found off-path. These results indicate that delivering location-based information is a more efficient channel for the park managers to communicate and influence visitors' behaviour towards eco-friendliness.

## 5. CONCLUSIONS

The results described in this work underscore the value of spatio-temporal data for assessing the impact of mobile information technologies. This is particularly important because it provides a geographical basis for evaluating such technologies that extends and complements more commonly used approaches grounded in psychology and usability.

The main issue for the development of methods in this regard was how to handle the uncertainty associated with the variability of high-resolution track data. This uncertainty arises from errors in positioning, incomplete information, and the general variability in individual movements. To cope with these issues a number of techniques were described in this work. In terms of data handling; filtering, linear-reference and aggregation techniques were described that brought the data into forms that allowed comparison amongst tracks to be made and the influence of different variables to be explored. In terms of analysis, a number of visualisation techniques were described that identified patterns of autocorrelation within the data that could be explored and the effect of systematic errors to be considered. The patterns suggested using these techniques were then validated using statistical methods. The combination of these methods proved successful in allowing inferences about spatial behaviour to be made. In particular, it could be shown that location-sensitive provision of information significantly effects how visitors behave whilst other media for delivery have little effect. Additionally, it was found that there are places where visitors tend to stop irrespective of the mode of delivery and information content. To some extent these could be related to features in the landscape, however in general a more comprehensive answer was lacking. This suggests future avenues of work that might attempt to complement track data with participant observation and interviews, as well as analysis of track data as a function of the environment. It was also noted that the "novelty effect" of technology was not one that had been adequately controlled for. In particular this made explaining the results for the visitor group using non-location-based information presented on a PDA difficult to account for. It will be important in future work to better control for this effect and determine whether it is undesirable, transient, or useful in terms of encouraging visitors to explore natural environments.

Whilst aggregation was useful to smooth out local variations amongst the singular tracks and so explore the more general trends of the data, it also caused much potential about interesting information about individual behaviour to be lost. Future work will aim to look at the data more when disaggregated. The work undertaken here will be able to complement this by providing baseline models of movement against which individual tracks can be considered.

## 6. REFERENCES

Butler, R., Hall, M. and Jenkins, J., 1998. *Tourism and Recreation in Rural Areas*. Wiley, Europe, pp. 1-274.

Dias, E., Beinat, E., Rhin, C. and Scholten, H., 2004. Location Aware ICT in Addressing Protected Areas' Goals. In: *Research on Computing Science vol. 11 (special edition on e-Environment)*, Prastacos, P. and Murillo, M. (eds). Centre for Computing Research at IPN, Mexico City. pp. 273-289.

Edwardes A., Burghardt D. and Weibel R., 2003. Webpark location based services for species search in recreation area. In: *Proc. 21st Intl. Cartographic Conference (ACI/ICA)*, Durban, South Africa, pp. 1012–1021.

Elands, B. and R. van Marwijk, 2005. Expressing recreation quality through simulation models: useful management tool or wishful thinking? In: *Proc. 11th International Symposium on Society and Natural Resource Management*, June 16-19, 2005. Östersund , Sweden.

Farrell, T.A. and Marion, J.L., 2001. Identifying and assessing ecotourism visitor impacts at eight protected areas in Costa Rica and Belize. *Environmental Conservation*, 28(3), pp. 215-225.

Hägerstrand, T., 1970. What about people in regional science. *Papers of the Regional Science Association*, 24, pp. 7–21.

Kwan, M.-P., 2000. Interactive geovisualization of activity-travel patterns using three dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C*, 8(1-6), pp. 185-203.

Lynn, N.A. and Brown, R.D., 2003. Effects of recreational use impacts on hiking experiences in natural areas. *Landscape and Urban Planning*, 64(1), pp. 77-87.

Laube, P., Imfeld, S., and Weibel, R., 2005. Discovering relative motion patterns in groups of moving point objects. *Int. J. Geographical Information Science*, 19(6), pp. 639-668.

Laube, P., and Purves, R. S., 2006. An approach to evaluating motion pattern detection techniques in spatio-temporal data. *Computers, Environment and Urban Systems*, 30(3), pp. 347-374.

Laube, P., Purves, R. S., Imfeld, S., and Weibel, R., 2006. Analysing Point Motion with Geographic Knowledge Discovery Techniques. In: *Dynamic and Mobile GIS: Investigating Change in Space and Time*, Drummond, J., Billen, R., Forrest, D., and João, E., (eds.). Taylor and Francis, London, pp. 263-286.

Manning, R.E., 2002. How much is too much? Carrying Capacity of National Parks and Protected Areas In: *Proc.s of Monitoring and Management of Visitor Flows in Recreational and Protected Areas*. Bodenkultur University Vienna, Austria, January 30 - February 2, 2002. pp. 306-313

Mark, D. M., 1998. Geospatial lifelines, *Integrating Spatial and Temporal Databases*, *Dagstuhl Seminars*, no. 98471.

Mountain, D.M. and MacFarlane, A., in press. Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals. *Journal of Information Science*.

## 7. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the study's participants; EcoMare, Camineo and Geodan for allowing the use of the LBS system in this research; Patrick Laube for his insightful comments. The first author would like to thank the support from the Portuguese National Science Foundation (FCT/MCT) under the PhD grant SFRH/BD/12758/2003. Additionally, the authors gratefully acknowledge the European Commission through their funding of the IST FP6 project TRIPOD (045335) which supported parts of this work.