Quantitative Assessment of Context for Mobile Services

A Case Study on the WebPark Service in the Swiss National Park

Master Thesis in Geography

Department of Geography
University of Zurich

Jonas Snozzi

Supervisors:

Dr. Tumasch Reichenbacher, University of Zurich
Prof. Dr. Dirk Burghardt, Dresden University of Technology
Prof. Dr. Robert Weibel, University of Zurich

Zurich 2009

Personal declaration:

I hereby declare that the submitted thesis is the result of my own, independent, work. All external sources are explicitly acknowldeged in the thesis.

Jonas Snozzi

Abstract

Over the last years, the number of mobile devices has exceeded the number of traditional PCs. Since the available information on the internet became unmanageable, a new problem in finding the most appropriate and relevant information has emerged. Additionally, mobile devices have a limited display resolution and size. Because of these developments, the idea to present and map only the most relevant and important information arose. Such a digital mobile guide for hiking has been developed in the Swiss National Park. This mobile service tries to answer the visitors' questions where and when they are the most relevant. The requests of the users are logged with the coordinates and time. It is possible to analyze these logfiles with respect to different context factors. In this thesis the context variables location (trail and picnic area), environmental variables (topography, vegetation, and weather condition), time (relative and absolute time), and user groups are analyzed. The user groups are derived from the duration of stay at three picnic areas on each trail. The relevance of each information group in specific contextual conditions can be derived from the relative distribution of the requests. Additionally the influence of each context variable and the sensitivity of each piece of information on context could be assessed. But the results are uncertain, because a major error in the calculation and the model must be assumed. One of the biggest sources of error is the GPS accuracy, which is coupled with the calculation of most of the context variables.

The most striking results are the high sensitivity of the data on the relative time and the meso space, which is in this thesis represented with three different trails. The picnic areas have also an influence on the request behavior of the users, because high peaks in terms of quantity are observable near the picnic areas.

Zusammenfassung

Die sich im Umlauf befindenden mobilen Geräte nahmen in den letzten Jahren stetig zu. Zum heutigen Zeitpunkt sind mehr solcher Geräte als herkömmliche PCs im Umlauf. Mit dem fast unendlichen Datenangebot des Internets kamen Probleme auf, die relevantesten Daten zu finden und zu repräsentieren. Zusätzlich verfügen mobile Geräte über ein limitiertes Potential detailreiche Karten auf dem meist kleinen Display darzustellen. Beide Probleme führten zur Idee, dem Nutzer nur die relevantesten und wichtigsten Informationen darzustellen. Im Schweizer Nationalpark ist seit einigen Jahren ein digitaler Wanderführer im Einsatz, welcher den Wanderern Informationen rund um den aktuellen Standort zur Verfügung stellt. Die Abfragen der Nutzer werden in einer Logdatei mit den jeweiligen Koordinaten und der jeweiligen Zeit gespeichert. Diese Logdateien bieten sich an, den Einfluss von verschiedenen Kontextfaktoren auf das Abfrageverhalten der Wanderer zu untersuchen. In dieser Arbeit wurde untersucht, wie sich der Standort (Wanderroute und Rastplatz), Umgebungsvariablen (Topographie, Vegetation und Wetter), Zeit (relative Wanderzeit und Tageszeit) und Nutzerkategorien auf das Abfrageverhalten der Wanderer auswirkt. Die Nutzerkategorien wurden anhand der Rastzeit an den Rastplätzen ermittelt. Aus der relativen Verteilung verschiedener abgefragter Informationen konnte die Relevanz einer jeder Information unter den gegebenen kontextuellen Umständen ermittelt werden. Die Stärke des Einflusses der verschiedenen Kontextfaktoren sowie die Einflussnahme auf verschiedene Informationsgruppen konnte abgeschätzt werden. Jedoch hat sich gezeigt, dass für die meisten Kontextfaktoren ein nicht zu vernachlässigender Rechenfehler und Modellfehler besteht, welcher nicht zuletzt von der Positionsgenaugkeit des Gerätes abhängt. Als einer der wichtigsten kontextuellen Faktoren konnte die relative Wanderzeit ermittelt werden. Weiter ist der Raum auf der Skala der Wanderwege wichtig. Für die quantitative Verteilung im Raum sind hauptsächlich die Rastplätze verantwortlich, an welchen viel mehr Abfragen als im restlichen Gebiet gemacht werden.

Acknowledgements

I want to thank my supervisors Tumasch Reichenbacher, Dirk Burghardt, and Robert Weibel for their support. They helped me to develop and realize this thesis. Then I want to thank Ruedi Haller and Antonia Eisenhut from the Swiss National Park. They gave me good inputs and ideas. I also want to thank David Mountain for his preliminary work on the data, which helped me realizing the preprocessing. Then I want to thank Livia Kurer, Simon Späth, Alex Nilson, Jonas Manser, and Daniel Rüegg for reviewing the thesis.

Table of Contents

List of Figures

List of Tables

1 INTRODUCTION

The number of mobile devices, such as personal digital assiastants (PDA) and mobile phones, has exceeded the number of traditional PCs in the last years. The usage of these devices has changed from toys to tools (Struss, 2004). Also multimedia phones are becoming more and more popular. On some of these platforms commercial map applications are available. By the majority they are car navigation systems (Sarjakoski & Nivala, 2005). Location-based services (LBS) become more and more integrated with the physical environment (Gellersen, 2003). The almost ubiquitous availability of the internet has brought web maps and web services to the mobile environment. Therefore not the general availability of the information, but other factors such as display size and computing power constrain the usage of such devices (Meng & Reichenbacher, 2005). Other factors such as the constantly changing environment, the volatile user emotions, or the time-critical user tasks constrain the efficiency of the delivered information (Reichenbacher, 2004).

So far the most utilized application in a mobile environment is a mobile map. Because of limitations of the display these map often contain only some points of interest (POI) on a more or less skeletonized background. The design of the map as well as its containing information must be calculated in real-time or pseudo real-time in order to ensure a high acceptance by the user (Meng & Reichenbacher, 2005). One of these LBS has been developed in the Swiss National Park (SNP) since 2001. The WebPark^{SNP} service has the goal to answer the questions of the park park visitors where and when they arise (Haller, Burghardt, & Weibel, 2005). Because the WebPark^{SNP} service was developed with respect to user needs it is accepted and liked by the majority of the users.

Log files have been produced as a product of the usage. This logging data has been analyzed by different research groups with different goals. Haller and Eisenhut (2008) analyzed how frequent the different information pieces were requested. A continuative study of Eisenhut et al. (2008) tried to clarify what impact the topography has on the request frequency of the WebPark^{snp} users. From a geographic perspective these analyses can be seen as a part from the field of geographical relevance and context research, with the general goal to find the most appropriate and relevant information in certain situations. Eisenhut et al. (2008) did not look into the distribution of the requests, but on the total amount of requested information under certain circumstances. Going a step further in analyzing these log files it can be determined *which* information is the most relevant in certain situations. In order to ensure that the presented information is relevant, ineterpretable and also readable, and due to the above mentioned constraints of a mobile device, only a limited amount on information can and should be presented to the user.

1.1 MOTIVATION

User-centered design has been a topic in a great amount of research papers for years. The general goal is *"to create usable design solutions that allow users to do the things they want to, not the things they have to"* (Meng & Reichenbacher, 2005, p. 6). The International Organization for Standardization (ISO) set the standard 13407 (1999) on user centered design process, which can be seen as a circle. The first step is to determine the *contexts of use,* which will lead to *requirements* of a system. These *requirements* can lead to a *design solution,* which can be *evaluated.* The *evaluation* can initiate new *context of use,* and the cyclic process starts again. If a specific *design solution* satisfies the current requirements, the cyclic process can (temporarily) be brought to an end. This process is illustrated in Figure 1. Officially the WebPark^{SNP} project finished in 2004, but the application and the system underlie a constant process of renewal and innovation. This study can be seen as part of the whole innovation process of the system. Analyzing the request for different context variables and their characteristics can be seen as a special for omf an evaluation task. The specific form motivation is to provide the SNP with new ideas in order to upgrade the system and make it even more efficient.

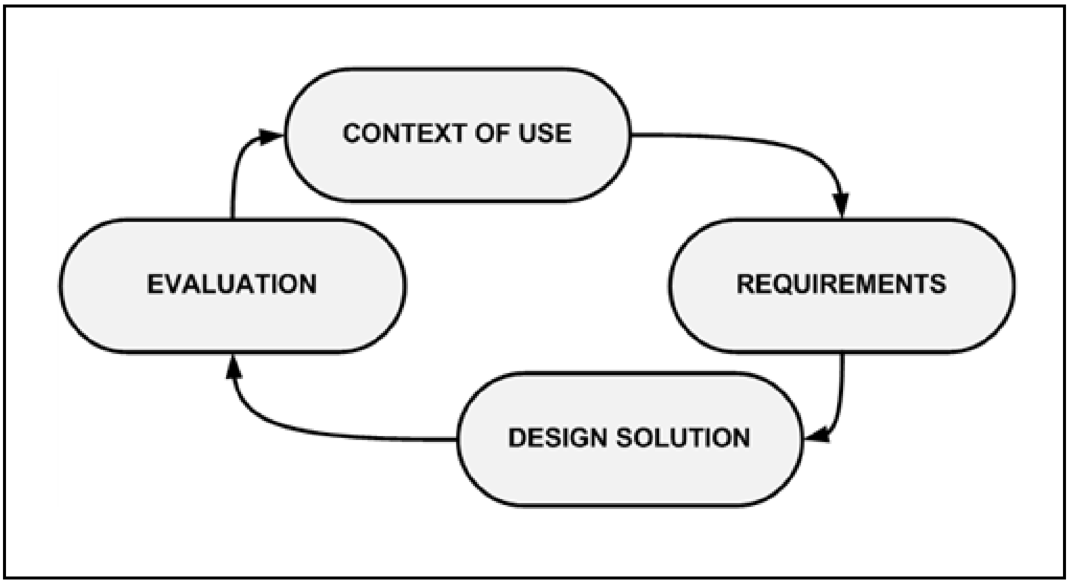

FIGURE 1 : USER-CENTRED DESIGN CIRCLE BASED ON ISO 13407.

The scientific motivation is to study both *context research/geographical relevance* in a mobile environment and *user classifications* with similar goals are in the scope of current GIS research. Relevance is especially important for mobile geovisualizations and maps. Due to several constraints, such as limited display resolution and size, not all information can be drawn to a map. Therefore only the most important / relevant ones should be selected. The importantce of a single piece of information has to be modeled and calculated accordingly.

1.2 RESEARCH OBJECTIVES

This thesis concerned with context research and user modeling in a mobile environment. A quantitative analysis of human-computer interaction based on requested information of the WebParkSNP service will be conducted in order to investigate the usage of context. It can be assumed that not always statistical prove statements can be formed, but since the data set is very large, the findings should be of some statistical significance. Based on the analysis of user requests qualitative statements shall then be formed. Based on this general objective, the following sub goals are:

- Proving that the request for certain pieces of information as well as the request frequency in general is depending on contextual factors.
- In which way is the behavior is dependent on contextual factors and reveal any potential patterns.
- The contextual factors shall be rated with respect to their influence. It shall be clarified which factors have more influence on the user's request patterns and which have less influence.
- How can the contextual variables be modeled coherently.

One of the contextual factors is a classification of the users into homogeneous groups. For this special case it shall be clarified whether it is possible to apply a simple model on the behavior of the users which leads to significant differences in the request behavior.

All models and calculations should be as simple as possible, because the long-term objective is a real time calculation of relevance scores performed by a mobile device. Because of the limited storage as well as the calculation capabilities the processing chain has to be as simple as possible, and therefore also the underlying model must be as simple as possible.

The context variables to be analyzed can be split into topographic variables, time variables, user classification and variables which describe the current location, be it the temperature or the vegetation.

1.3 RESEARCH QUESTIONS

Derived from the research objectives the following research questions can be phrased.

The leading research question is:

"Is the request behavior of the park visitors dependent on context?"

If this is the case, a follow up question arises:

- *"Can context factors be defined quantitatively?"*

If the request behavior and therefore the relevance of information are dependenton the context and the context can be defined quantitatively, a research question concerning relevance is:

- *"Is it possible to specify the relevance of information under different environmental conditions?"*

These three main questions lead to a serie of more specific questions:

- *"Which level of measurmenet is the most appropriate for the context and the relvance?"*
- *"Which variables can be modeled with a given dataset, and which variables are not assignable?"*
- *"Which context factor(s) has/have the greatest influence on the request patterns of the users?"*

The user groups will be a special kind of context variable in this study.

- *"Is it possible to classify the user by their behavior around the picnic areas, and do their requests pattern differ?"*

1.4 OUTLINE OF THE THESIS

This thesis is organized into 6 chapters. The first chapter describes the *research context* of this thesis with the themes relevance, context and the WebPark system. The resulting model development is documented in the second chapter. The third chapter, the *methods,* contains all preprocessing steps and also the processing steps in order to get all data prepared for the analysis. The results for each context variable will be discussed in the next chapter on the *results*. Both the quantity and the quality of the requested information under different environemental conditions will be presented and discussed in the fifth chapter. The thesis is going to end with a *conclusion* chapter, which will provide achievements and insights, limitations and open problems and an outlook.

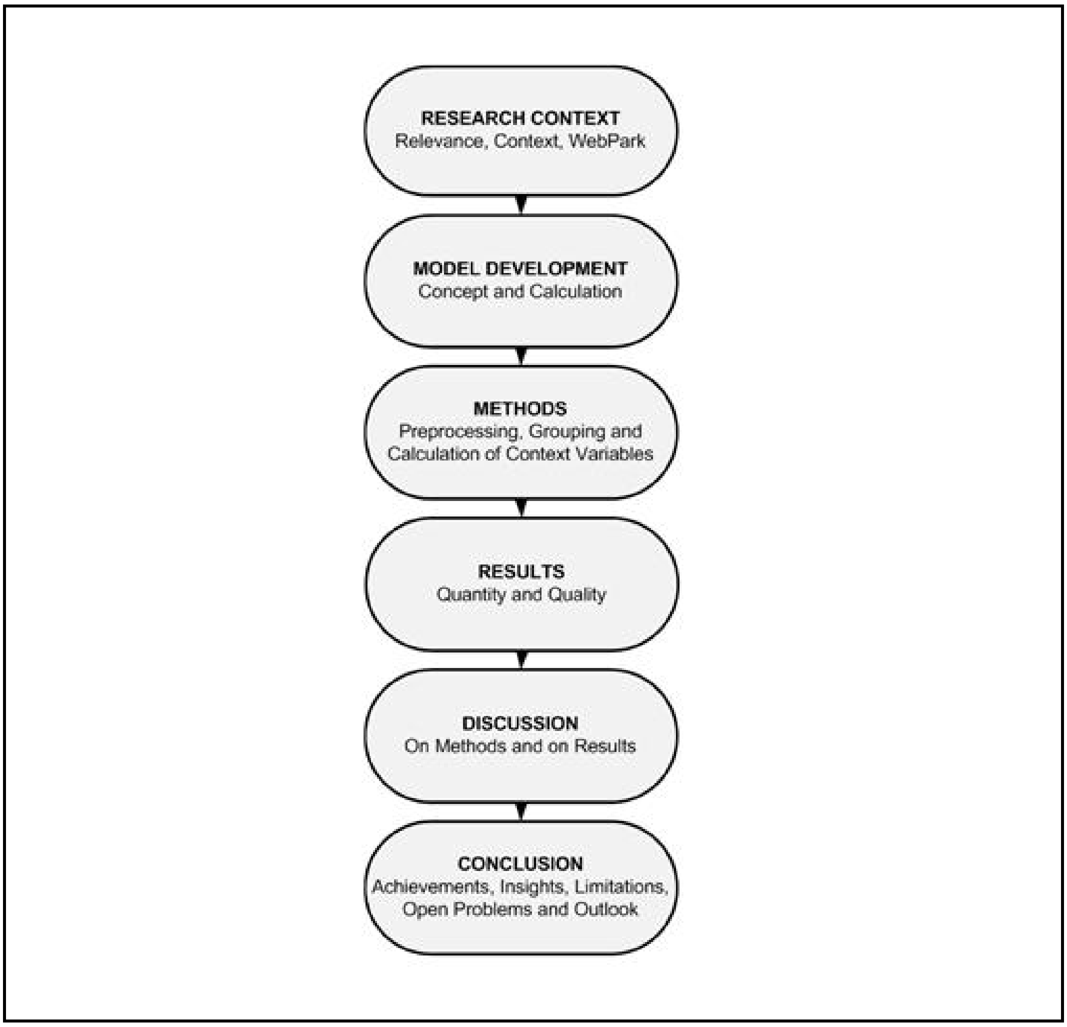

FIGURE 2 : OUTLINE OF THE THESIS.

2 RESEARCH CONTEXT

2.1 RELEVANCE

2.1.1 RELEVANCE MODELS IN OTHER DISCIPLINES

The term *relevance* does not only occur in the field of information science or even geographical information science. In the 1960s it was regarded as an "*evaluative tool for resolving problems associated with measuring the effectiveness of automated information systems*" (Greisdorf, 2000, p. 67). Other fields like philosophy and communication did also deal with the problem of categorizing information in terms of their relevance. From the relevance theory of Wilson & Sperber (2004) it can be deduced that what makes a person picking up a certain input is not only its relevance, but the relative relevance compared to an alternative input.
Saracevic (1996, p.206) proposes a list of general attributes of relevance such as *relation, intention, context, inference* and *interaction,* and defines relevance *"as a criterion reflecting the effectiveness of exchange of information between people".*

> *System or algorithmic relevance*: Every system has its own ways to represent and organize the files and text, as well to match the queries to these files.
> *Topical or subject relevance*: Describes the relation between the topics of the query and the retrieved text.
> *Cognitive relevance or pertinence*: Describes the relation between the level of knowledge and the retrieved text.
> *Situational relevance or utility*: Describes the usefulness of the retrieved text in the situation.
> *Motivational or affective relevance*: Relation between the goal or motivation of a user and the retrieved text.

Sperber and Wilson (1986) suggest three key points of relevance.

1. Relevance is associated with its context and the relation to the assumptions.
2. Relevance can be seen as a matter of degree.
3. Relevance can be utilized for comparative judgments, but also vague absolute judgments.

Relevance has also been used to rank information in order to be able to select the important information out of an unmanageable quantity of information. Internet search engines (e.g. Google) rank the relevance of information based on the scalar product of query vector and the document vector (Reichenbacher, 2007).
Different opinions about the measurability of relevance exist, ranging from immeasurable to scalable (Greisdorf, 2000).

2.1.2 GEOGRAPHICAL EXTEND

As presented by Raper et al. (2002) geographic information (GI) has a spatial and temporal extent. Derived from Tobler's (1970) first law of geography that "everything is related to everything else, but near things are more related than distant things", it can be assumed that information objects with a shorter distance in a given distance measurement have a higher relevance in a given context usage. The location itself plays an important role to evaluate an appropriate distance function, which includes not only the Euclidian space (Reichenbacher, 2009). Understanding the relevance for an individual is necessary to provide adapted and appropriate information on a location based service (Raper, 2007).

2.1.3 RELEVANCE IN MOBILE CARTOGRAPHY

The environment in a mobile situation influences also the information need of a user. Defining a task-driven geographical relevance in a mobile environment is complex and both task- and location-dependent (Raper, 2007). Reichenbacher (2001) tried to categorize the elements of relevance in a mobile system.
<u>High level tasks:</u>

- Locators
- Proximity
- Navigation
- Events

Therefore high level task, such as *locators, proximity, navigation* and *events* can be split up to several low level tasks. *Locator* tasks can be questions like *"where am I"* and *"where is…"*? *Proximity* tasks can be questions like *"where is the next …",* while a typical task for *navigation* is *routing. Events* can describe what conditions exist, or what's happening somewhere near. With concerns to the distraction of the user, the usage of visual solutions is depending on the context (Reichenbacher, 2001).

2.1.4 NEED FOR RELEVANCE CALCULATIONS

The assumptions following the first law of geography (Tobler, 1970) that things with a shorter distance to position are more relevant than things which are farther away, leads to the following thoughts (Reichenbacher, 2009). In a mobile environment the *distance* is not only defined as a Euclidian distance in space. Other possibilities might be the time distance or manhatten distances. In other words, an object is more *relevant* the closer it is in any relvance dimensions and a given context. In order to make these thoughs clearer an example will be discussed. In Figure 3 two users A and B are at the same location P. While user B is on foot and looking for a bookshop, user A is riding a bike and is looking for a bike shop. With LBS both users would only get the two objects 1 and 2 because they are inside the search radius of 250 m. But also the bike stores 3 ,4, and 5 are relevant to the user A and the book stores 6, 7, 8, and 9 are also relevant to the user B. Sure they are less relevant, because they are further away and might have some other attributes, which decrease their relevance to the user. Bike store 5 might just be closed and some of the book stores might not have the appropriate literature for the user, e.g. only selling books in a foreign language. Nevertheless they might be a valid alternative for the users. Another thing, which might be interesting, is that the relevance of bike store 4 and book store 8 are not equally relevant to their users, because user B is by foot, and the Euclidian or network distance to these storese plays a more important role than for the faster biker.

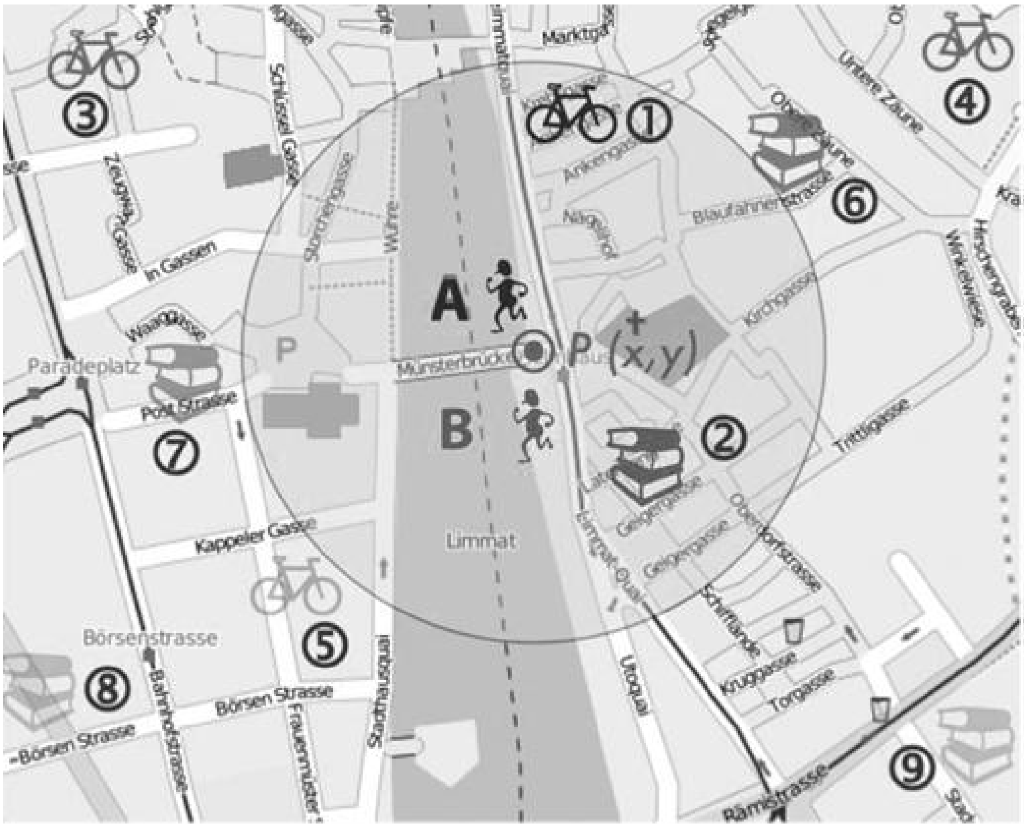

FIGURE 3 : DIFFERENT DISTANCES FOR DIFFERENT USERS (REICHENBACHER, 2009, P.2).

Both proximity and location are important factors in order to calculate the geographical relevance in a mobile environment (Reichenbacher, 2009). These calculations are not trivial. However, the use of relevance which inlcudes multiple factors and multiple concepts of distance enhance the quality of the presented information for the user. Therefore general concepts for contex modelling and relevance measurement are needed to enable an enhancement of LBS and mobile maps.

2.2 CONTEXT

Context can be defined in a holistic way as the sum of all circumstances and facts, which surround a certain activity (Dey, Salber, Abowd, & Futakawa, 1999). The term *context* is changing its meaning in the different domains of information science. In the field of artificial intelligence other general concepts are used as in the field of user interfaces (McCarthy & Sasa, 1997). In the field of mobile computing the differences between most of the definitions is the description of the context as well as the considered parameters (Dransch, 2005).
For computer systems, context can be defined as all factors that influence the calculation of a process beside the explicitly given arguments (Schmidt & Gellersen, 2001). Derived from these definitions Schnelle (2007) describes context as "the circumstances or events that form the environment within which something exist or takes place". One of the most cited publications of Dey and Abowd (2000, p. 3) contains the following definition of context:

> *„Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves. "*

Derived from these definitions it can be statet that context is much more than just location. It can vary a lot during the use and is therefore still difficult to measure or even identify (Kaasinen, 2002).

2.2.1 CONTEXT AWARENESS

From a general definition of context as the description of the surrounding area with its characteristics, additional information about the environment can be used to refine the system (Schmidt & Gellersen, 2001). Dey (2001) declares a system to be context-aware "if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task." Chen and Kotz (2000) distinguish between the context as a characteristic of the surrounding area, and the context that is relevant for mobile applications. Therefore they define context for mobile applications as *"the set of environmental states and settings that either determines an application's behavior or in which an application event occurs and is interesting to the user"* (Chen & Kotz, 2000, p. 3). From this definition they find a difference between *active context awareness* and *passive context awareness.* In the first case the application adapts automatically to the context, by changing its behavior, in the second case the context is presented for to user.
Because in some complicated navigation situations, the user might not be able to perform an active operation, Nivala and Sarjakoski (2003b) use a different approach for passive and active context awareness. They state that an active context-

awareness can be achieved by "*inputting the information for the application, e.g. by personalization"* (Nivala & Sarjakoski, 2003b, p. 46), while passive context-awareness is that the user can decide with a confirm-reject option, whether he wants context-aware information.

Van Setten et al. (2004) distinguish between context-awareness and recommender systems. For both types the goal is the same, namely to provide the user with the most appropriate content from an unmanageable amount of possible contents. The goal of context-aware systems is *"to provide a user with relevant information and/or services based on his current context"* (van Setten et al. , 2004, p. 236), while the goal of recommender systems is to provide the user with information of his interest. Both systems can be seen as a tool to provide the user with relevant information. In order to get a maximum accuracy for the provided information, both systems can be combined.

Becker and Nicklas (2004) proposed another classification for context-awareness. They distinguishe between:

- **Context-based selection:** The context information defines the information and services which are used by an application. Context factors can be the physical proximity or the user's preferences.
- **Context-based presentation:** The context is not responsible for *what* information is presented, but *how* it is presented.
- **Context-based action:** In the context-based presentation is the user directly involved in the interaction with the application. In contrast to that the context-based action automatically reacts to the context e.g. proximity.
- **Context-based tagging:** Tagging of information does not automatically lead to an immediate change of the behavior of the application.

For our purpose with the goal of dynamic mobile maps which enhance the usability of the device especially *context-based selection* and *context-based presentation* are important (Reichenbacher, 2007). *Context based selection* as well as *presentation* tries to supply more relevant information as illustrated in Figure 4.

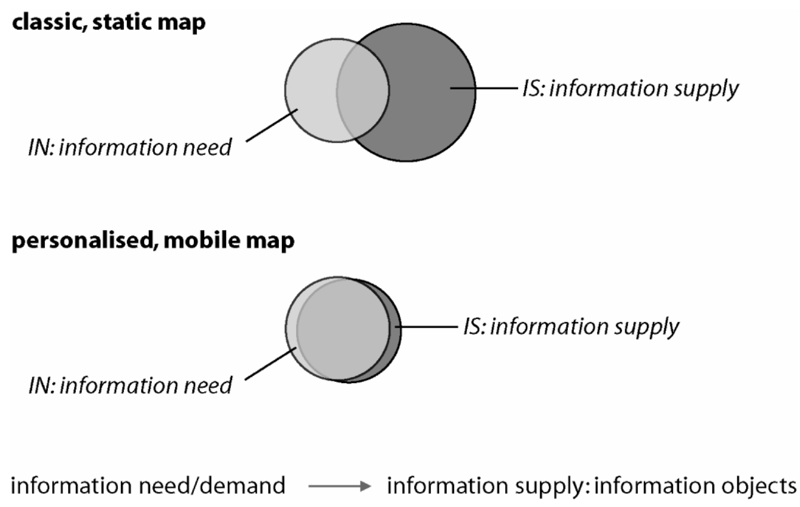

FIGURE 4 : GENERAL GOAL OF CONTEXT-AWARE APPLICATIONS (REICHENBACHER, 2007, P. 10).

2.2.2 CATEGORIZATION OF CONTEXT

For Schilit et al. (1994) the three most important aspects of context are the location (where are you?), the surrounding group of people (who are you with?) and what resources are nearby. Chen and Kotz (2000) find of this accumulation of single factors three general factors; the *computing context,* the *user context* and the *physical context.* They also add a fourth important factor, the *time context,* which indicates the season of the year, the time of day, week or month. They also distinguished between *high-level context,* and *low-level* context. *High-level* means the user's current activity and the complex social context, while *low-level* means time, location etc.
Other authors also tried to categorize the context into different factors. For our purpose the categorization of Sarjakoski and Nivala (2003) is the most appropriate. They defined the following eight main context variables, which influence the information needs in a mobile environment.

- *Location,*
- *System*
- *Purpose of use*
- *Time*
- *Physical surroundings*
- *Navigation history*
- *Orientation*
- *Cultural and social background*

These eight variables can be allocated to five general context categories which are illustrated in the following table. They are *computing, user, physical, time* and *history.*

TABLE 1 : CATEGORIZATION OF CONTEXT, (NIVALA ET AL., 2003, P. 26, BASED ON DAY, 2001 AND CHEN AND KOTZ, 2000)

General context categories	Context categories for mobile maps	Features
Computing	System	Size of display Type of the display (black – color screen) Input method (touch panels, buttons etc.) Network connectivity Communication cots and bandwidth Nearby resources (printers, displays)
User	Purpose of use Social Cultural	User's profile (experience, disabilities, etc.) People nearby Social situation
Physical	Location Physical surroundings Orientation	Lightning Temperature Surrounding landscape Weather conditions Noise levels
Time	Time	Time of day Week Month Season of the year
History	Navigation history	Previous locations Former requirements and points of interest

Because of several difficulties with other context variables, the location is the only context element in mobile map applications, which is currently exploited. Another reason, besides the difficulties to calculate the other context variables, is that the location is, concerning the map, the most appropriate variable. Other variables might just be irrelevant (Sarjakoski & Nivala, 2003).

2.3 RELEVANCE AND CONTEXT

Important factors for context are the conditions and surrounding influences which make a situation unique (Brézillon, 1999). Saracevic (1996) emphasized the role of context for the relevance of information. Reichenbacher (2007) stated that the most important context dimensions are location, time, user, activity, information, and the system. How relevance and context play together conceptually and computationally will be presented in the following subchapters.

2.3.1 EXISTING CONCEPTUAL MODELS

In general the context model *"separates applications from the process of sensor processing and context fusion"*. The context model also allows the application to *"share the gathered context"* (Becker & Nicklas, 2004, p. 3).

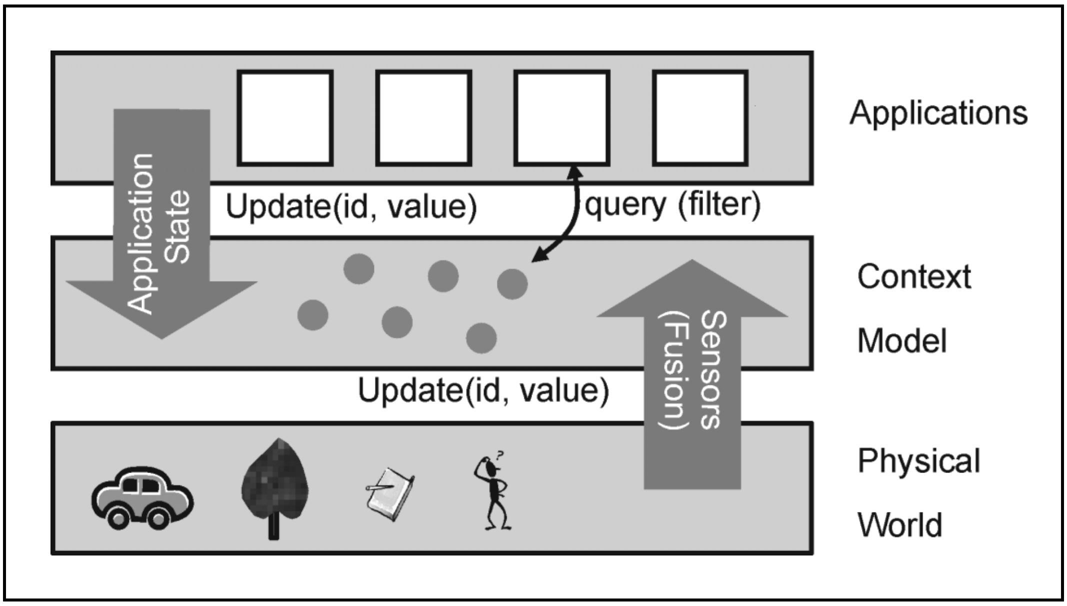

FIGURE 5 : GENERAL MODEL FOR THE RELATION BETWEEN THE PHYSICAL WORLD AND A CONTEXT-AWARE APPLICATION (BECKER & NICKLAS, 2004, P.2).

Dransch (2005) stated that the differences in the definitions of *context* mainly originate from which parameters the definitions consider. A very general model is presented by Jameson (2001). The relevant parts of the physical world are the situation, the current state of the user as well as the long-term properties of the user. The situation, mainly the location, can be read by context sensor such as GPS devices. Other sensors e.g. in the jewelry of the user could read the current state of the user, e.g. his emotional arousal. The long-term properties of the user could be evaluated through his interaction with the system, or his browsing history. All three main features of the physical world define whether the presented information is relevant to the user.

In order to achieve a maximum relevance or interestingness of the information, the system has to react on all three main features (depicted with dashed lines in Figure 6).

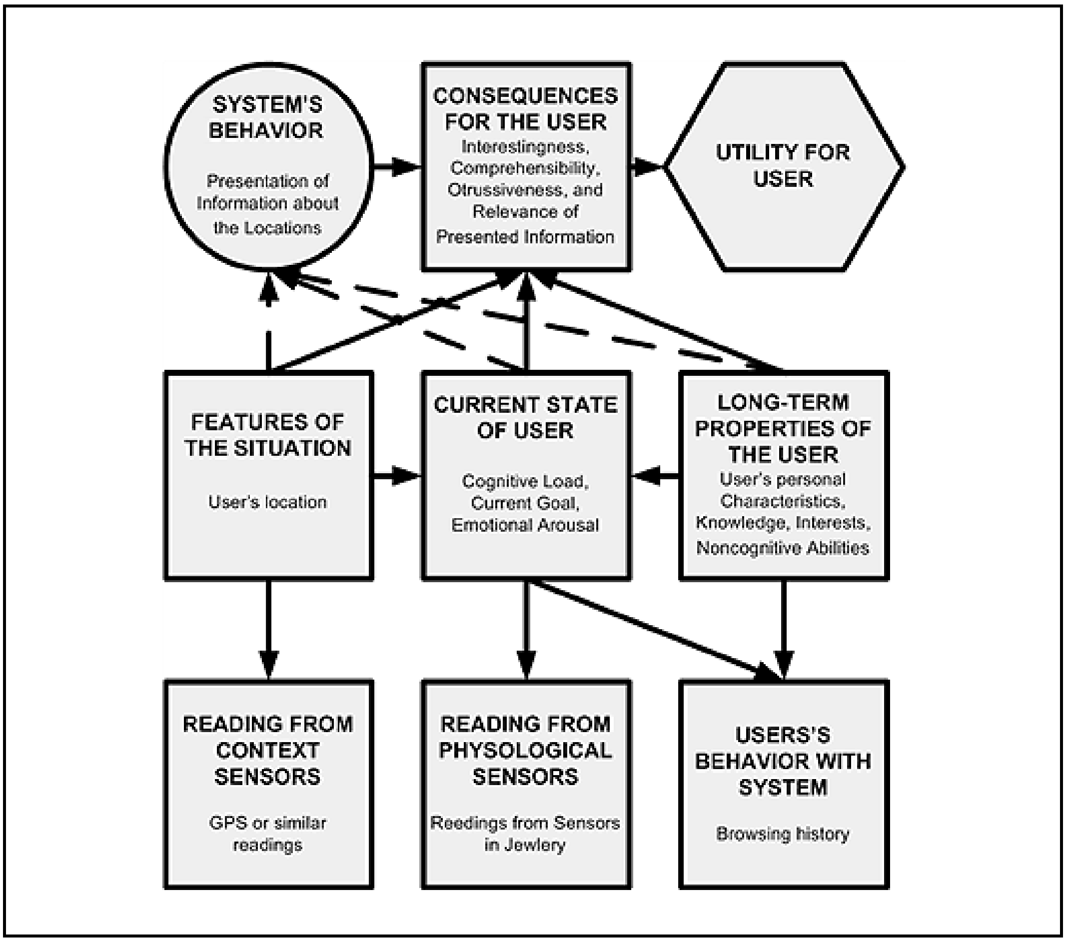

FIGURE 6 : CONTEXT, USER'S CURRENT STATE, USER'S BEHAVIOR, AND LONG-TERM PROPERTIES (RE-DRAWN FROM JAMESON (2001, P. 32)).

For Jameson the only feature of the situation is the location. For Reichenbacher (2009) location can be an index, a query parameter, an information attribute, a place, a mobile activity, a link to the neighborhood, or a predictor for future locations. Therefore location does not only contain an x, a y and maybe a z dimension, it can mean much more.

A similar but more concrete model based on a mobile situation is presented by Reichenbacher (2007). He categorizes relevance into *objective relevance* and *subjective relevance.* The *objective relevance* is devided into the *physical relevance* and the *system relevance.* While the *objective relevance* is independent from the user, the *subjective relevance* shows a clear dependency on the user. *Physical relevance,* which is devided into a temporal and spatial component, and the *system relevance* are part of the *objective relevance.* On the other is the *subjective relevance* based on *cognitions, activities etc.* of a person.

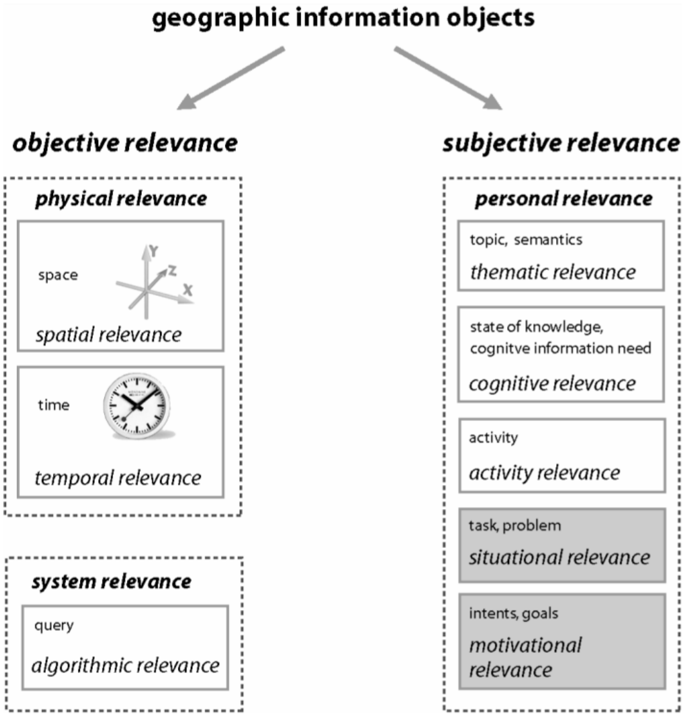

FIGURE 7 : RELEVANCE TYPES FOR MOBILE CARTOGRAPHIC APPLICATIONS (REICHENBACHER, 2007, P. 5, BASED ON SARACEVIC, 1996).

Because individuals are so different, it is almost impossible to model single persons. Defining user groups could be a reasonable way to model the *subjective relevance.* Social parameters such as age, sex, culture, or interests could be used to define user groups. But the most important factor for users of geoservices is their activity. *"You are what you do"* (Dransch, 2005, p. 39).

2.3.2 EXISTING CALCULATION MODELS

The basic concepts for calculating relevance scores are going to be introduced in this subchapter followed by examples in which some of these concepts were realized.
Keynotes
Reichenbacher (2007) proposed some general rules to assess relevancy. E.g. that near, visible and audible objects are more relevant than invisible, inaudible objects which are further away. Further he stated that if the object is "*needed for the successful completion of an activity*", or is "*linkable to users"* prevalent knowledge, it becomes more relevant, while contents with low information are less relevant. These functions are not easy to formalize. Some approaches how the difficult calculation problem could be handled will now be presented (Reichenbacher, 2007):

- **Utility functions:** Utility functions are used in economics to evaluate the relative satisfaction of goods and services. This concept can be adapted to our purpose. Every content or *point of interestest* (POI) would get a utility function, which depends on several different context variables. With a known context it would be possible to organize information by their utility (Reichenbacher, 2007). A utility function based on different context variables could look like the following (Bidgoli, 2004):

$$\boldsymbol{U}(\boldsymbol{x_1}, \dots, \boldsymbol{x_n}) = \sum w_i U_i(x_i)$$

  The Utility *U* is depending on n factors x. $U_i(x_i)$ are the attributes and w_i are the their specific weights.
- **Information retrieval functions:** In the field of information retrieval (IR) the documents get ranked by their similarity to the query, which represents the users's information need. The geographical extent or geographical information retrieval (GIR) could be measured by the *"geometric match between the query footprint and the document footprints*" (Jones & Purves, 2008).
- **Fuzzy set:** The memberships are not allocated binary but with a certain degree unlike in crisp set. Normally the relative membership is defined in [0,1] where 1 represents a full membership of the set. Both space and time can be modeled with a fuzzy function (Burrough & McDonnell, 2005).
- **Observation-based approaches:** The relevance of information with a geographical extent can be determined by the observation of the users and their activities. To every specific activity suitable features can be determined. Enhancing this model into predicting the next activity of the user is the next development step (Reichenbacher, 2007).
- **Geographical relevance assessment:** If an independency between the different relevance dimensions can be assumed a compound relevance factor outmatches isolated relevance factors. In general the relevance function can be written as (Jones, Alani, & Tudhope, 2001):

$$R(O_i) = \sum_{j=0}^{n} w_j * r_j$$

Where *r* is the relevance for every relevance dimension *j* and *w* is the weight for these dimensions. A single relevance score r_i can be modeled in different ways. The space can be modeld as a distance function etc. (Reichenbacher, 2007).

EXAMPLES

GIR

In terms of geographic information retrieval (GIR) Andrade and Silva (2006) propose a model, which considers *inclusion, proximity* and *siblings* into the calculation of a geographic similarity of a query. In this term *siblings* means their relatedness in the ontology graph, whereas *proximity* means the inverse of distance. All three factors are weighted similarly and summarized, what leads to a final value which lies in [0,1].

Also in the field of GIR Andogah and Bouma (2008) proposed a combination of both non-geographic relevance measure and geographic relevance measure. Special attention was paid to the scope, be it macro or micro based. They also discussed the possibilities of the combination of the geographic and non-geographic parts of relevance and proposed a *linear interpolated combination,* a *weighted harmonic mean combination* and a *extended harmonic mean combination.* The best result were achieved, when the geographic component was weighted low, compared to the non-geographic component.

As an addition to these factors Jones et al. (2001) calculated three different distance measurements, which can be seen as an analogue to the relevance concept. The authors calculated a Euclidian Distance Measure (ED), a Hierarchical Distance Measure (HD), which measures the hierarchical distance of an object in terms of spatial segmentation of a region, and the Thematic Distance (TD). Combining these three distance measurements the following formula for a normalized score was found:

$$Score = 100 - 100(w_t TD_n + (w_s (w_e ED_n + w_h HD_n)))$$

All *distances* have to be weighted with the weights w_t, w_s, w_e and w_h, and the subscript *n*'s indicate that the values were normalized.

FUZZY SET

Schmidt and Gellersen (2001) argue that the context validity is dependent on space and time. Their principle of *locality* and *temporality* defines a context being limited on a certain time and space situation. In order to model time and space a fuzzy function is used in dependency on the spatial and temporal origin of the context. The further away in time and space an element is from the origin of a context, the smaller the truth value of the membership to a set becomes. Because their theory is based on the fuzzy logic, the truth value is not a binary score, but a score in [0,1]. An element can therefore be part of more than one context. Other context factors beside distance and time are not discussed.

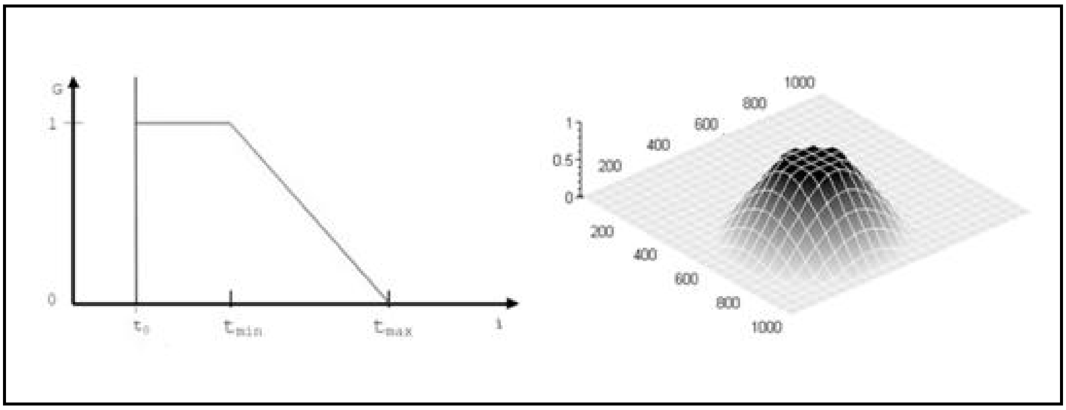

FIGURE 8 : FUZZY SET FOR TIME AND SPACE (SCHMIDT & GELLERSEN, 2001, P. 6).

OBSERVATION-BASED APPROACHES

An example of how the movement and therefore future positions can be predicted is presented by Mountain and MacFarlane (Mountain & MacFarlane, 2007). The lighter the raster cell in Figure 9 is, the more likely the location is to be visited in the next 30 minutes. Therefore functions about the surrounding area can be enhanced. Other variables such as *visited places* and *accessibility in a certain time* have also been analyzed by a questionary of the visitors of the SNP.

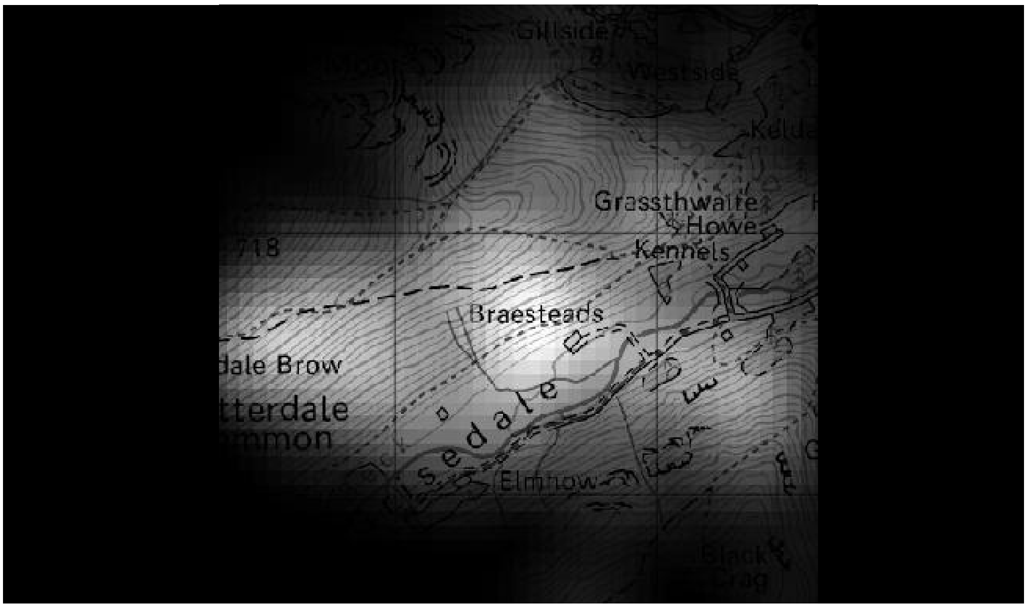

FIGURE 9 : LOCATIONS LIKELY TO BE VISITED IN NEXT 30 MIN (MOUNTAIN & MACFARLANE, 2007, P. 10)

Another example, which tries to predict the future location of a user is presented by Brimicombe and Li (2006). They try to model a mobile space-time envelope by the current position, the direction of travel and the velocity. This relative simple model of the user could supply the user with information about future locations.

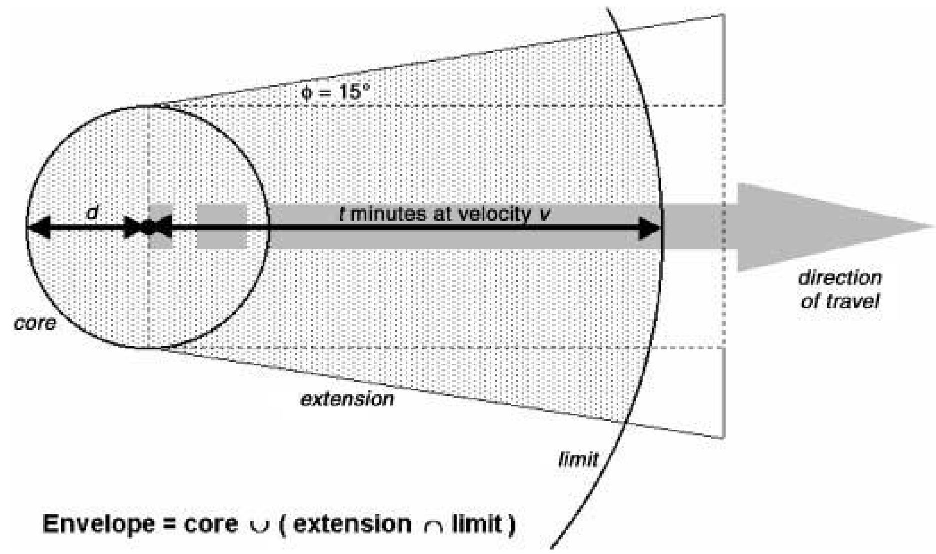

FIGURE 10 : SPACE-TIME ENVELOPE OF A USER OF LBS (BRIMICOMBE & LI, 2006, P. 15)

Geographical relevance assessment

In Reichenbacher (2004) the relevance values for space, time and thematic were calculated indivually. After that they were summed up to a total relevance factor. The relevance for an object *O* is a function of the distance from *O* to the current location.

$$(O, space) = f(\Delta space)$$

Similarly the relevance is also depending on the time

$$(O, time) = f(\Delta time)$$

and the distance between query and feature attribute

$$(O, thematic) = f(\Delta thematic)$$

All three relevance scores can be summed up to the total score, while the weights are depending on the context.

2.4 WEBPARK

The general aim for protected areas is to conserve the natural heritage and secondly to support the leisure/tourism industry. An additional goal for a majority of the protected areas is the environmental education of its visitors (Dias, 2004).
For this purpose the European Commission *Information Society Technology* (IST) enabled a program with the project number IST-2000-31041 with the name WebPark project. It ran for three years from 2001 to 2004. Its specific goals were to *"identify the geographic information needs of mobile users, to provide to such users geographically relevant personalized location-based services (LBS) and to create new G-commerce value-chains for recreation / protected area administrations and data integrators"* (WebPark, 2001, p.4). A new computational framework was created including aspects like the location, time, personal interests and activity of the visitors in order to leverage the data resources with a contextualized access and presentation. Within the WebPark framework several LBS were developed. These services can be built upon tailored collected data or other data such as environmental science research data or other existing information (Dias, Beinat, Rhin, Haller, & Scholten, 2004).

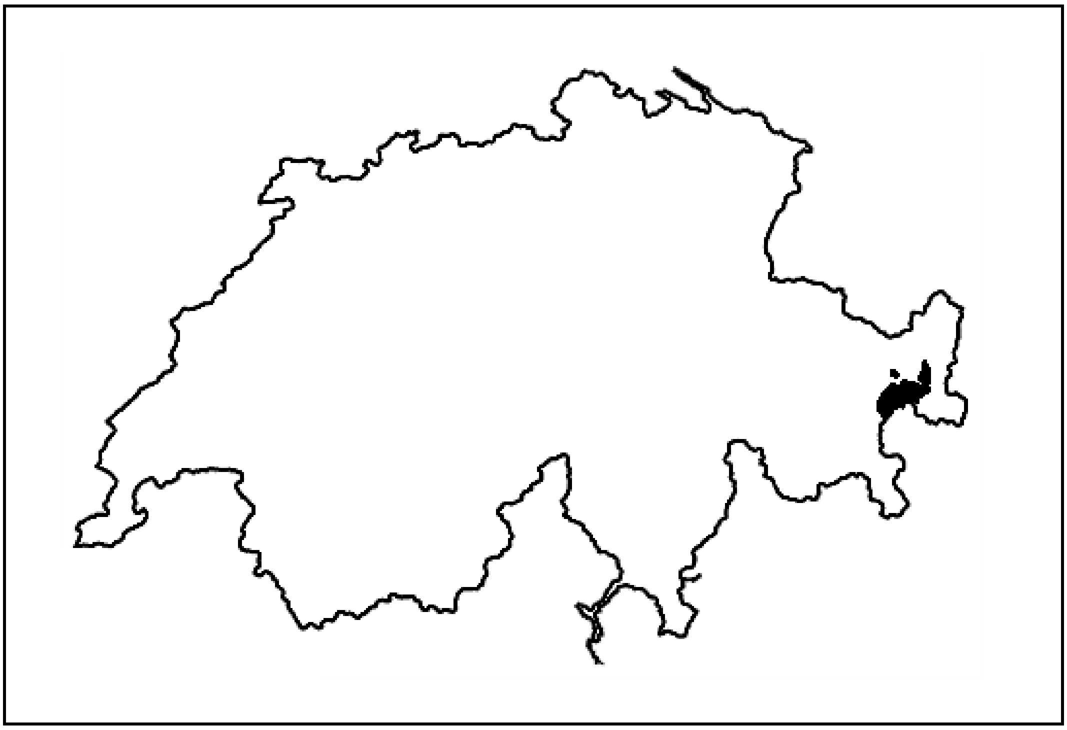

FIGURE 11 : LOCATION OF THE SWISS NATIONAL PARK.

As study areas the project focused on two sites; The Swiss National Park and the park in Texel (Holland). The Swiss National Park is located in the very south-eastern end of the canton of Grisons (see Figure 11). It was founded in 1914 and holds the strictest category of protection from the Union for Conservation of Nature and Natu-

ral Resources. The landscape is dominated by mountains and it supports 3 main types of habitat; high alpine, forest and alpine meadows, which leads to a wide variety of alpine flora and fauna. Nearly 150'000 people visit the park every year. Therefore strict regulations make sure that the disturbance of the animals and the plant life is minimized. The WebPark project can also be seen as one part to ensure the conservation of nature (Edwardes, 2007).
Four main areas sought to be innovated by the project (WebPark, 2001);
The visitors' question should be answered where and when they are the most relevant.

- The device should be available everywhere and anytime.
- The storage, handling, integration, and commodification of geographic content should be enhanced.
- The device should be aware of the past requests of the users, the future space and the users' activities.

The results of the analysis of the user needs by Krug et al. (2003), e.g. the importance of safety information and the location of animals and plants, leaded to the integration of the following services (Edwardes, 2007).

- **Mapping and Navigation** – Topographic maps, as well as a route vertical profile so that the users can locate themselves horizontally as well as vertically.
- **Geographic Bookmarking** – Allows the user to annotate special situations, locations. The resulting bookmarks could be used by the same users, as well as by other users.
- **Point of Interest Search** – Interesting locations, e.g. locations that are near by the user's current position can be looked up.
- **Flora and Fauna Search** – The users can access a variety of different information on the flora and the fauna of the Swiss National Park.

3 MODEL DEVELOPMENT

3.1 CONCEPTUAL MODEL

Derived from the context definition of Chen and Kotz (2000) a suitable context definition for our purpose can be formed. Therefore only variables that can be measured are declared as context. The only direct measurements, which can be seen as context sensors after Jameson (2001), are GPS coordinates, which are logged in the WebPark^{SNP} system. After Dransch (2005) not every user by itself, but user groups are modeled. Inspired by Sarjakoski and Nivala (2005) the categorization of context variables is visible in the following figure:

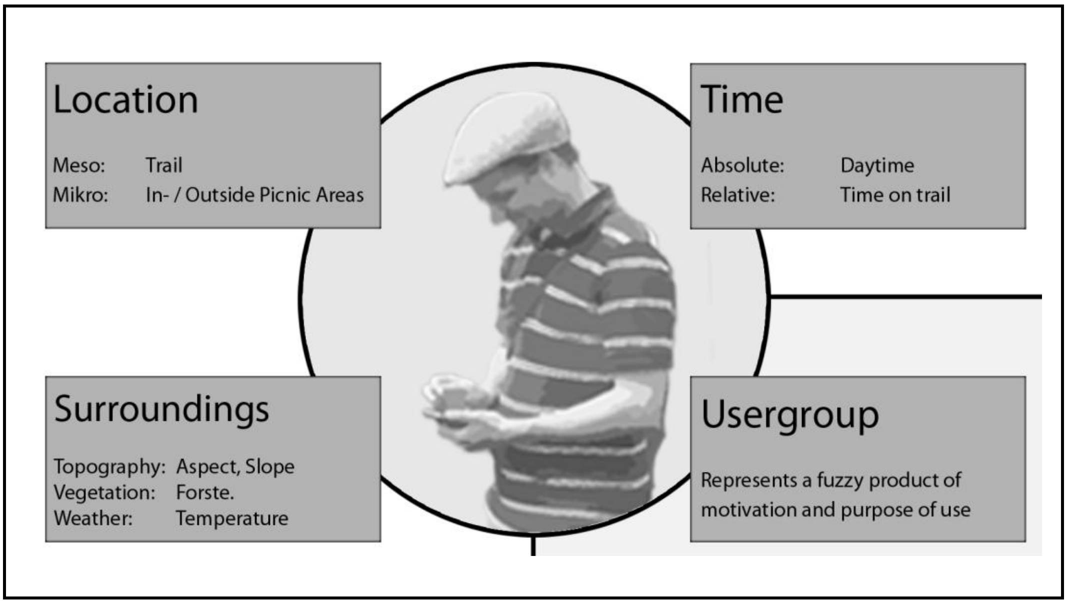

FIGURE 12 : CATEGORISATION OF CONTEXT VARIABLES.

Location in our sense is determined as a pair of coordinates. Therefore location has to be understood as a position on a specific trail or near a specific picnic area. But every location has also its surroudings.

TRAILS

The presented model states that location influences the user's request patterns. Location has several different scales. Compared to other situations, be it the internet, which gathers information from all over the planet, the dimension of the SNP is small. Therefore the trails cannot be regarded as a *macro* space, but more as a *meso* space.

USER GROUPS

The available data doesn't contain any direct information on the user's characteristics. There is no information about age, sex or preferences of users. Log files contain only the user's queries and their tracks during their hikes in the national park. A possibility to derive user groups would be to analyze the querying behavior and search for patterns. But the main goal of this thesis is to model this very request behavior. Dias et al. (2008) presented in their work techniques that allow comparisons of different tracks with the analysis of the users' speed on given tracks. One of the goals of that study was to describe aggregated patterns of group behavior. They used linear referencing and aggregation as main methods. Linear referencing was used to associate every GPS point with a single track. Due to aggregation of points, the track was divided into five meter segments. If it took longer than fifteen seconds to pass through a segment, it was considered as a stop or slowdown. This method could be adapted and applied to suit the purpose of defining user groups through their walking speed in the SNP. However, in order to reduce the complexity and calculation effort, only the duration of stay at the picnic areas is going to be analyzed.

NEAR PICNIC AREAS

The results of the user groups might be influenced by other factors such as whether the hikers are resting or hiking. Dransch (2005) stated that the most important factor is the activity. Therefore it can be derived that the information need changes when the activity changes from hiking to resting.

But some assumptions have to be made, because in the SNP it is not allowed to rest outside marked picnic areas: It can be assumed that the hikers do mostly respect these rules and only rest inside the picnic areas. Yet, some hikers might rest on other places, but this cannot be taken into account due the great effort that would have to be made to process such a behavior. Therefore the hikers are regarded in a resting state while they are inside the defined picnic areas, which are also used to classify the users.

TIME VARIABLES

In previous studies it could be shown that time has an influence on the renting frequency as well as the usage of a device. Eisenhut et al. (2008) declared that the longer hiker is already on its way, the less information is requested. According to the findings of Eisenhut et al. (2008) it can be assumed that the relative time and absolute time have an influence on the requesting patterns.

TOPOGRAPHY

Eisenhut et al. (2008) showed that there is a general difference in the quantity of information a user needs in different slopes and also in different aspects. They showed that information requests at southern expositions are more frequent than on northern expositions. Also the users are less interested in the WebPark^{snp} service in moderate slopes, while they use it more often in steep slope ore nearly flat areas. From these observations it can be derived that the aspect and slope might also have an influence on the requesting pattern.

VEGETATION

Eisenhut et al. (2008) present a short study on how the vegetation cover influences the request behavior of hikers in the SNP. Therefore the vegetation might also have an influence on the requesting patterns.

WEATHER

Temperature is seen as an important context factor. It can be seen as part of the physical surrounding (Nivala & Sarjakoski, 2003a). It is imaginable that the temperature has an influence on the handling of mobile devices. The usability for instance might decrease in a very cold surrounding. Also high temperatures and light reflections might have an effect on the use of the device. Another factor, which might influence the user's behavior, is the precipitation. While it is raining there might be less requests, and the requests might be more related to find the next shelter.

3.1.1 COMPARISON OF CATEGORIZATIONS

If we look at Table 2 several parallels between the categorization of different models can be found, even though a distinct allocation of a certain variable in one model to a context variable in another model is not always possible.

TABLE 2 : PARALLELS BETWEEN THE DIFFERENT CONTEXT CATEGORIZATIONS.

Context variables	Classes in Reichenbacher (2007)	Classes in Sarjakoski & Nivala (2005)	Classes in Hirakawa & Hewagamage (2001)
• **Trails**	• Objective relevance (OR) / space	• Location	• Location
• **User classes**	• Kind of subjective relevance	• Kind of purpose of use	• Contextual factors
• **Picnic Areas**	• OR / space	• Location	• Location / Activities
• **Relative Time** • **Absolute Time**	• OR /time	• Time	• Time
• **Slope** • **Aspect**	• OR / space	• Physical surroundings	• Location
• **Vegetation**	• OR / space	• Physical surroundings	• Contextual factors
• **Weather**	• OR / space	• Physical surroundings	• Contextual factors

For the conceptual model some assumptions have to be made. The model of Jameson (2001) includes the *current state of the user* the *system's behavior* and the *utility for the user.* No data is available for these three parts of the model. Therefore the model must be adapted and simplified.

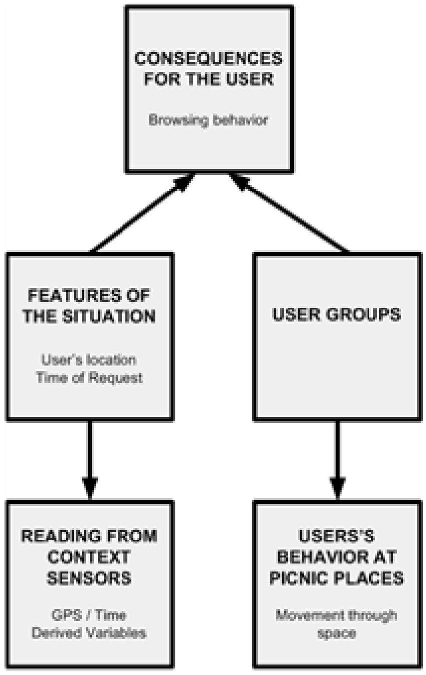

FIGURE 13 : SIMPLIFACTION OF JAMESON'S (2001) MODEL.

The utility of a single request is for all users constant, because measuring a single content is not possible. However, if the relative distribution of the points in a certain context is regarded, it can be said that if information is more often requested, it was also more useful. In our model the features of the situation are defined by the user's location and time for a certain request. The location has certain environment specifications which can be derived from other data sets. Similar to the *long-term properties of a user* the users can be categorized into user groups. Key factor for this categorization is the behavior at the picnic areas, namely the duration of stay. The impact or the consequences for the user can be measured in terms of the distribution as well as the quantity of requested information in certain contextual conditions. Because no data is available on the *system behavior* and the *utility for the user* they have to be assumed to be constant, even though not all requested content is equally utile for the user and the system architecture has an impact on the request behavior. Now that we have a concept of which context variables are going to be taken into account and how they generally are considered to interact with the users, a calculation model of how the context variables influence the users' request behavior can be developed.

3.2 CALCULATION MODEL

The calculation model is split into two values. The first value expresses how much information should be provided; the second expresses what kind of information should be provided. Both will be discussed individually. The finally provided information emanates from the combination of how much information in general is required, and the relative distribution of the different requested information.

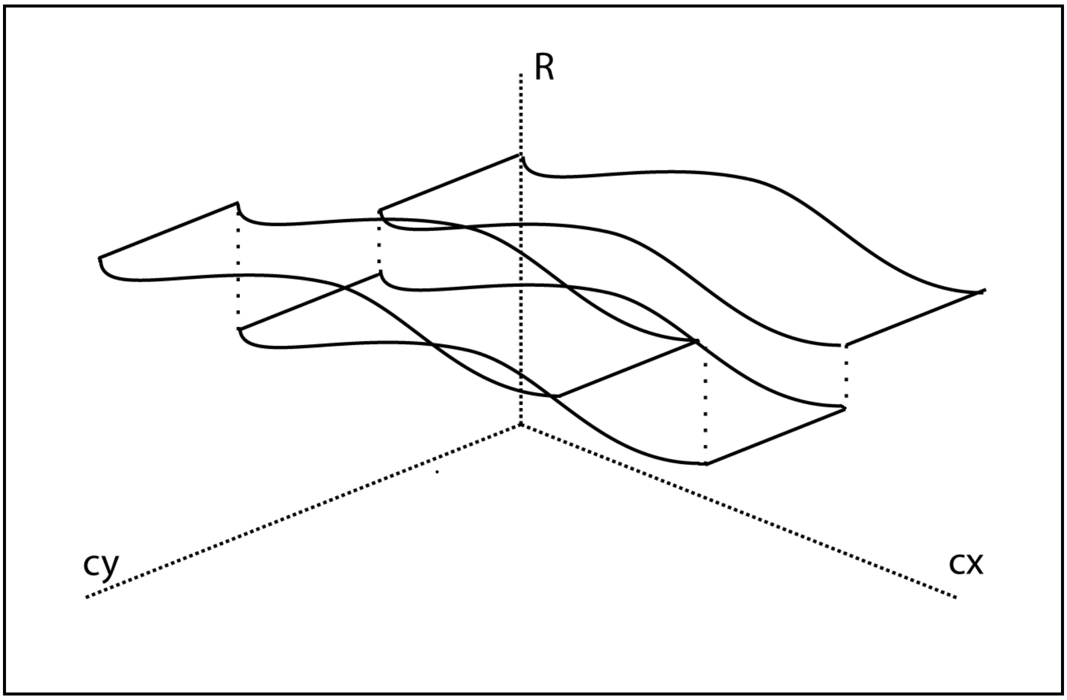

FIGURE 14 : GENERAL IDEA OF COMBINATION OF FACTORS.

The general concept treats the total relevance as a combination of several factors. Both continous (context factor *cx* in Figure 14) as well as step functions (context factor *cy*) are applicable. Every step can for instance represent a class of manifestations of a certain context factor. The combination of the context factors is a simple multiplication. Therefore, a number of n context factors in an n-dimensional field define how relevant certain information is. The context can be seen as a manifestation of a *utlity function* or a *geographical relevance assessment.* How the calculation was realized will be discussed in the following.

FORMULA

First, the set of different contents has to be defined as:

$$C = \{C_1, C_2, \dots, C_n\}$$

where C is the total set of contents, and n is the number of different contents. After that, a set of different context factors and situations have to be defined. The context factors shall be

$$F = \{F_1, F_2, \dots, F_m\}$$

F is the set of all context factors F_m. Every context factor F_m shall have k characteristics.

$$F_m = \{F_{m\,1}, F_{m\,2}, \dots, F_{m\,k}\}$$

If a categorization of a context factor is not possible and the context factor can be seen as a continous function, then the continous function can eighter be modeled by an interpolation of a categorization or the number of k characteristics can be assumed to be infinite.

With these three definitions it should be possible to model the context and also the relevance of particular information. The relative importance or relevance of information can be expressed with an index:

$$(1) \quad \frac{Importance_{C_n}}{Importance_{C_{std}}} = \frac{Req_{C_n}}{Req_{C_{std}}}$$

This index indicates how important certain content is compared to a standard content. The standard content C_{std} is just the total content C divided by the number of different contents n.

The intrinsic relevance of information can be assumed to be just one context factor F_m. Therefore it can be formed as the following

$$(2) \quad Rel_{C_n}^{F_{int}} = \frac{Req_{C_n}}{Req_{C_{std}}} = \frac{Req_{C_n * n}}{Req_C}$$

The intrinsic relevance score $\mathrm{Rel}_{\mathrm{C_n}}^{\mathrm{F_{int}}}$ is, as every other relevance score of a different context variable, defined between 0 and ∞. In order to be able to speak of an intrinsic value of an information item, the number of requests has to be great enough. If this is the case, the trends of the different context instances cancel each other out.

Now, as the intrinsic value of content is defined, the value of such information in certain contexts can be modeled. For a context factor F_m with k characteristics, the relevance of the content C_n in a situation k can be defined as a fraction of the importance under the given characteristics of the factor compared to a normal condition.

$$(3) \quad Rel_{C_n}^{F_{m\,k}} = \frac{Imp_{C_n}^{F_{m\,k}}}{Imp_{C_n}^{F_{m\,std}}} = \frac{Req_{C_n}^{F_{m\,k}} * Req_C}{Req_{C_n} * Req_C^{F_{m\,k}}}$$

The total relevance of a specific content C_n under the condition k can be calculated through the multiplication of all relative relevance scores of all relevance factors F_m.

A special case is the intrinsic relevance, because only one condition of k is possible, and Req_C is equal $Req_C^{F_{m\,k}}$ and Req_{C_n} is $\frac{Req_C}{n}$. The characteristic k of each relevance factor F_i can change. All combinations of characteristics are possible.

$$(4) \quad Rel_{C_n} = \prod_{i=1}^{m} Rel_{C_n}^{F_{i\,k}}$$

$$= \frac{Req_{C_n * n}}{Req_C} * \frac{Req_{C_n}^{F_{2\,k(j)}} * Req_C}{Req_{C_n} * Req_C^{F_{2\,k(j)}}} * \frac{Req_{C_n}^{F_{3\,k(j)}} * Req_C}{Req_{C_n} * Req_C^{F_{3\,k(j)}}}$$

$$* \ldots * \frac{Req_{C_n}^{F_{m-1\,k(j)}} * Req_C}{Req_{C_n} * Req_C^{F_{m-1\,k(j)}}} * \frac{Req_{C_n}^{F_{m\,k(j)}} * Req_C}{Req_{C_n} * Req_C^{F_{m\,k(j)}}}$$

$$= \frac{{Req_C}^{m-1} * n}{{Req_{C_n}}^{m-1}} * \prod_{i=1}^{m} \frac{Req_{C_n}^{F_{i\,k(j)}}}{Req_C^{F_{i\,k(j)}}}$$

The relevance of information C_j can be compared with the relevance of information C_l and set into a relation. Therefore all content C can be ranked. After this ranking it can be stated whether the information or content C_j is more or less relevant than the content C_l. But in order to get the relevant amount of information another factors has to be calculated; the quantity Q.

A simple example will illustrate the relevance of information under certain cirumstances in order to make the approach more comprehensible.

Two different contents C_1 and C_2 have been requested a total of 2000 times. The content C_1 has been requested 1200 times, while C_2 has been requested 800 times. If both contents had an equal intrinsic relevance, and because the number of different contents n is 2, they would be requested 1000 times each. Therefore the intrinsic relevance of C_1 is 1.2 and the intrinsic relevance of C_2 is 0.8 derived from formula 2. Under normal conditions 1200 of 2000 or 60% of the requests are *C_1's*. But under the condition m_2 only 210 of 500 or 42% are *C_1's*. In our example $Req_{C_n}^{F_{m\,k}}$ is 210, $Req_C^{F_{m\,k}}$ is 500, Req_{C_n} is 1200 and Req_C is 2000. According to formula 3, the index for the given circumstances is therefore:

$$Rel_{C_1}^{F_{m\,2}} = \frac{Req_{C_1}^{F_{m\,2}} * Req_C}{Req_{C_1} * Req_C^{F_{m\,2}}} = \frac{210 * 2000}{1200 * 500} = 0.7$$

The three other indexes are calculated similarly.

TABLE 3 : SIMPLE CROSSTAB WITH TWO CONTEXT CHARACTERISTICS AND TWO DIFFERENT INFORMATION PIECES.

		m_1	m_2	Total
Content C_1	Count	990	210	**1200**
	Expected	900	300	
	Index	**1.10**	**0.7**	
Content C_2	Count	510	290	**800**
	Expected	600	200	
	Index	**0.85**	**1.45**	
Total	**Count**	**1500**	**500**	**2000**

The intrinsic relevance has to be multiplied with the indexes from the cells in order to calculate the toal relevance of the contents C_1 and C_2 under the different conditions m_1 and m_2. The intrinsic value of C_1 is 1.2 and bigger than the intrinsic value of C_2, which is 0.8. But under the condition m_2 the total relevance of C_2 becomes bigger. With the values 1.16 for C_2 and 0.84 for C_1 it can be seen that the relation between the two contents has been reversed compared to the total distribution. On the other hand the relation under the condition m_2 has become more distinct.

The average requested quantity of information can be assumed as the total requests per total time:

$$(5) \quad Q_{avg} = \frac{Req_C}{T_{total}}$$

The amount of information for a specific context factor can be calculated analog to the calculation of the relevance of information. But the quantity is a description for the total content C and not only a specific part of it. For a single factor F_m with it's characteristic k the relative quantity ratio can be expressed with:

$$(6) \quad QR_{F_{m\,k}} = \frac{Q_{F_{m\,k}}}{Q_{std}} = \frac{Req_{F_{m\,k}} * T_{total}}{Req_C * T_{F_{m\,k}}}$$

An alternative to the total time could be the dimension of space, if a relative time could not be derived. The problem with the space dimension is, that it must be assumed that the movement through space is homogenous, which is in reality clearly not the case.

The quotient which includes all context variables and their characteristics can be expressed through a product of the average quantity Q_{avg} and all elements of F.

$$(7) \qquad QR_{k(j)} = \prod_{i=1}^{m} QR_{C_n}^{F_{i_{k(j)}}} = \prod_{i=1}^{m} \frac{Req_{F_{i_{k(j)}}} * T_{total}}{Req_C * T_{F_{i_{k(j)}}}}$$

$$= \frac{T_{total}{}^{m}}{Req_C{}^{m}} \prod_{i=1}^{m} \frac{Req_{F_{i_{k(j)}}}}{T_{F_{i_{k(j)}}}}$$

Each characteristic *k(a)* of one context factor *a* can be combined with every characteristic *k(b)* of the context factor *b*. It is therefore possible that the value *k* changes with the context factor C_i.

4 METHODS

4.1 PREPROCESSING

4.1.1 PREPROCESSING CHAIN

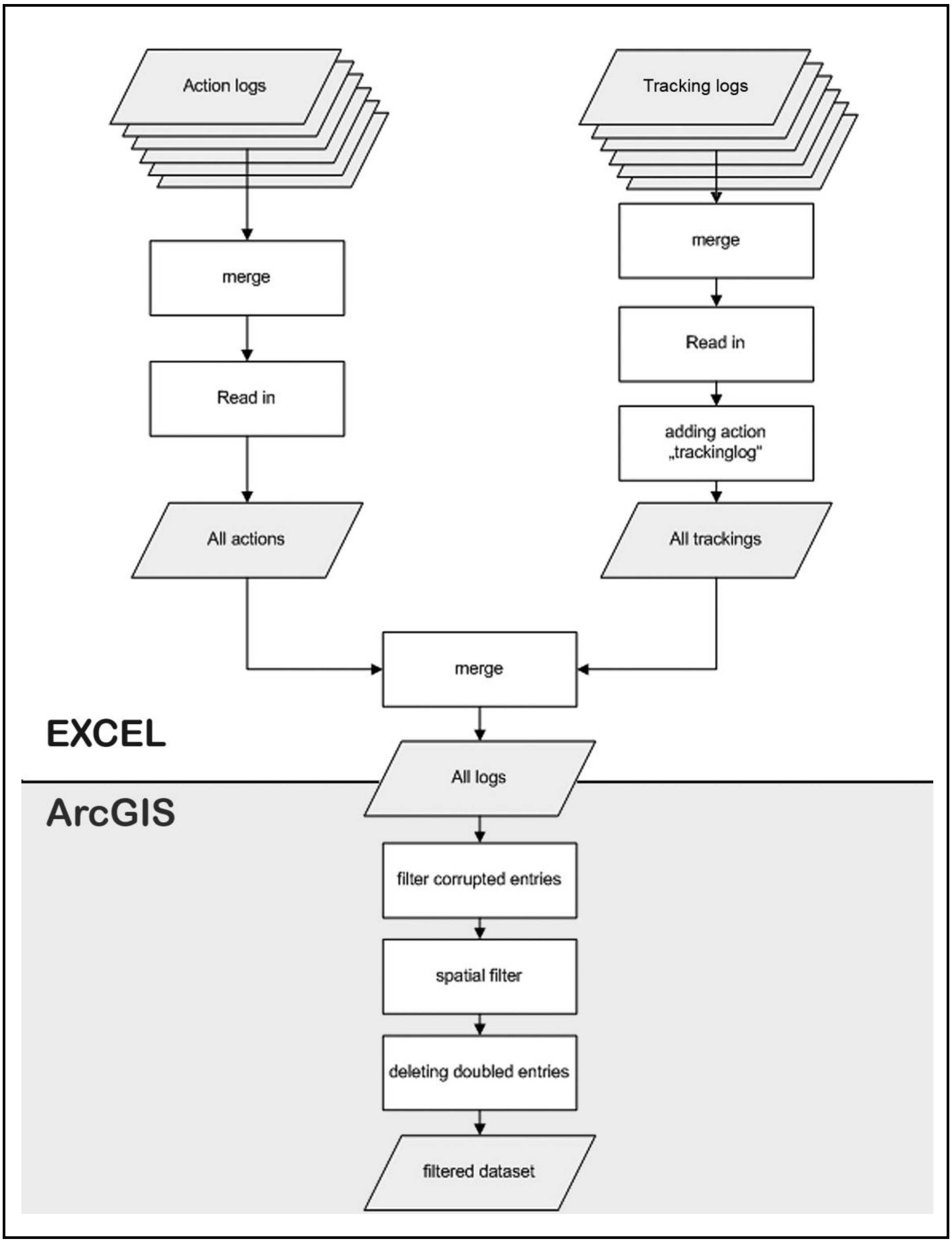

FIGURE 15 : PREPROCESSING CHAIN.

The general preprocessing workflow is illustrated in Figure 15. There are several different processing steps that need to be realized in order to be able to calculate the characteristics of the context variables. These preprocessing steps will be explained and discussed in the following sections.

4.1.2 AVAILABLE RAW DATA

The main data source is the WebPark log files from 2007. Every day and device hasits own log files and the user actions and the tracking data are stored seperately. These actions are stored similarly to a web browser history. Every page or information piece is stored in ascending order in a log file and a new log file is created every time the device was shut down or crashed. There are a great number of log files because the device shut down or crashed many times. Therefore, all files need to be merged first, which can be accomplished with any text editor. Merging all files of each device leads to 24 files: one action file and one tracking file for every device. All these files contain information with the following structure (Figure 16):

```
DATE;TIME;LONGITUDE;LATITUDE;ALTITUDE;ACCURACY;ACTION;ID

270707;101111.45;808374.94;170330.73;1897.0;22.0;/WebPark/ge
tTuto?h=f2147483604

270707;101127.49;808971.75;169854.17;1844.0;9.0;/WebPark/get
ButtonPage?PAGE_ID=MainPage

270707;101143.48;808971.75;169854.17;1844.0;6.9;/WebPark/get
ButtonPage?PAGE_ID=Bookmarks

270707;101143.48;808971.75;169854.17;1844.0;6.9;/WebPark/sea
rchBookmarks

...
```

FIGURE 16 : EXAMPLE OF A STRUCTURE OF A LOG FILE.

Every line contains one entry and the variables are separated by a semicolon. Due to this structure, the data could be imported to Microsoft Excel and stored as a dBase file which can either be read by SPSS or even by ArcGIS. The dBase format was always used when the data needed to be transferred from one program to the other.

4.1.3 READING IN THE DATA

The first line in a file was interpreted as the variable names during the import of the files toEXCEL. All other variables except the variable *"action"* were interpreted as numeric value. After this first step, a new variable called *"device"* was created, and every device received its ID ranging from one to twelve. The device identifications were necessary in order to be able to separate the different users in a later process. Almost all variables except the *"action"* variable were interpreted as numeric value. Different solutions were found for the variables *"time"* and the *"date".* In order to provide simple arithmetic functions, the variable "time" was split into *"hours", "minutes"* and *"seconds"* to finally calculate a single value *"timesecond"*, which indicates the time of day in seconds. The "date" was interpreted as a *calendar date class* in the dBase format.
In order to merge the tracking file and the action file together, in the action file, a new variable called *"action"* was created, and the values for all entries were set to *"trackinglog".* Both files were merged to one file, to provide an easy transfer to other programs such as SPSS and ArcGIS.

4.1.4 FILTERING THE DATA

So far, the raw log files were just imported and converted to a common format. In order to allow further processing and analysis, the data had to be be filtered in several different steps.

FILTERING CORRUPTED ENTRIES

Some of the entries were corrupted. This means crucial entries such as the date, time, longitude, latitude or the action were *null* or unreadable. The cause for this corruption is not known. It is possible that the device was shut down or crashed during the reading process or the GPS coordinates could not be determined. Since only a few cases were corrupt, there was no huge impact on the data set. The corrupted entries were removed, as they could not be used in the analysis anyway.

SPATIAL FILTERING

The dataset with over 130'000 entries was imported into ArcGIS. With the function *"make XY event layer"* dimensionless entries recieved an x and a y coordinate. It was possible to apply a python script to set a unique user identification (Code 1). With so-called *cursors* it is possible to access and to iterate through a set of rows in a table. With the *update cursor* it is possible to modify a row, which will be needed later. It is also possible to sort the entries in ascending order. The *UpdateCorsers* just needed to be extended with an additional variable (Tucker, 2004).

```
import sys, os, string, arcgisscripting
gp = arcgisscripting.create()
USERID = 0
TEMPDATE = 0
TEMPDEVIDE = 0
curs = gp.UpdateCursor("$path\\$file", "", "", "", "DATE A;
DEVICE A")
cur = curs.Next()
while cur:
      if (cur.DATE == TEMPDATE) and (cur.DEVICE == TEMPDE-
VICE):
            cur.USRID = USERID
            curs.UpdateRow(cur)
      else:
            TEMPDATE = cur.DATE
            TEMPDEVICE = cur.DEVICE
            USERID = USERID +1
            cur.USRID = USERID
            curs.UpdateRow(cur)
      cur = curs.Next()
```

CODE 1 : ITERATING THROUGH ALL ENTRIES, IN ORDER TO SET UNIQUE USER IDENTIFICATIONS.

The results of the iteration were 485 unique users of the WebPark device of 2007. These users and points had to be filtered in order to fulfill several criteria.

Users need to hike on specific trails: This study confines itself to analyze the four most frequently used trails according to Eisenhut et al. (2008). Therefore, only hikers using those four trails will be considered. The entries of the other hikers and trails were removed.

Only distinct trails are considered: Even if a user was hiking on one of the four specific trails it is possible that his/her entries are invalid, because in some cases certain segments of different trails are congruent. E.g. if someone hikes on the trail Fuorcla Val dal Botsch, he/she usually also hikes on the trail Margunet, which is one of the considered trails. But the entries of this hiker are not comparable to the ones from the users on the trail Margunet, because the trails have, besides the segments that they share, different characteristics like the average time for completing the tour. The aim is to derive a data set that only contains entries from hikers who only walk on the four most frequent used trails Margunet, Chamanna Cluozza, Murter and Val Trupchun.

No data from the arrival: Looking at the data from the filtered users, there are still a lot of points that are not wanted. Most of the hikers test their device in the visitor's centre in Zernez or do not turn it off in the bus. These points are unwanted, because they may distort the findings from the trail Chamanna Cluozza. These points also must be removed.

After filtering the data set with the explained methods, 201 unique users with over 80'000 GPS points remained. But the dataset had still not the desired quality due to a data storage error.

DELETING DOUBLED ENTRIES

For an unknown reason, there are some entries which have the exact same values, whether coordinates and time or the *"action"*; all entries are the same and can lead to misinterpretations. The dublicated entries have to be removed by using the following python script (Code 2). This script sorts all variables in ascending order and then determines whether the next entry has the exact same entries in the crucial variables. If this is the case, the row is removed; if not, new temporary values are set.

```
import sys, os, string, arcgisscripting
gp = arcgisscripting.create()
rows = gp.UpdateCursor("$path\\$file", "", "", "", "USRID A;
TIMESECOND A; LATITUDE A; LONGITUDE A; ACTION A")
row = rows.Next()
tempid = 0
temptime = 0
templon = 0
templat = 0
tempqry = ""
while row:
      if    (tempid == row.USRID) and
            (temptime ==row.TIMESECOND) and
            (templat == row.LATITUDE) and
            (templon == row.LONGITUDE) and
            (tempqry == row.Action)
                  rows.DeleteRow(row)
                  row = rows.Next()

      else:
            tempid = row.USRID
            temptime = row.TIMESECOND
            templat = row.LATINT
            templon = row.LONINT
            tempqry = row.QRY_REGROU
            row = rows.Next()
```

CODE 2 : PYTHON CODE TO REMOVE DUBLICATED ENTRIES.

This method eliminates about 10'000 entries from the originally 80'000 points. About half of the remaining points are tracking logs and half of them are requested information by the users. Deleting the dublicated points is a crucial action in the preprocessing chain, because some entries have been stored more than one thousand times, what can lead to misinterpretation and bias of the statistical analysis.

4.2 Grouping Information

4.2.1 Distinguishable information

The log files, and therefore also the variable "action" in the imported data files contain the file path, which is sometimes cryptic. An example is provided in the box below (Figure 17):

```
/WebPark/getMapIFoi?id=bookmark-
15210_1185542529181&TS=1185543416937

/WebPark/getNativeContent?id=/DE/by/by_de/butterfly_start.ht
m&mimetype=text/html+topMe

/WebPark/getNativeContent?id=/DE/bird_de/3_small.htm&mimetyp
e=text/html+topMenu
```

FIGURE 17 : EXAMPLE OF FILE PATHS IN THE LOG FILES.

If these file paths are assigned unambiguously to groups that are kept as small as possible, 39 groups of information can be identified. Nine of these groups, e.g. taking pictures or information on the "*privacy policy*", have less than 100 entries, while other pieces of information like returning to the "*main page*" have more than 4000 entries. The *"route information",* the *"get around"* function and the "*IFOI list*" are also very popular. Some of the entries like the "*battery low status*", or getting *"more"* information about any kind of thematic, can be defined as junk, or *"internal navigation"* meaning that these clicks are not directly related to information or content, but are used to gain access to those. Leaving these junk/internal navigation functions aside, the remaining groups need to be reclassified in order to get enough entries in every class. Examples are several different entries on *"bookmarks"* which were reclassified to *"bookmark".* It is also possible to merge groups in case the topics of the entries are very similar or answer the same purpose.

4.2.2 REGROUPING INFORMATION

Almost 23,000 action entries remain after the truncation of the data. They can be grouped into six main groups of information (Figure 18). The most frequently asked information is information about the surrounding area (*"Info Around")*, which can be accessed through the *"get around"* function or any kind of FOI/IFOI functions. The second most asked group of information is information on the trail. This group contains information on the route itself, about its vertical profile and virtual trails. The third group is the "*content*" group, which contains information about animals and plants etc. In the fourth group every kind of aid for orientation like *"where am I?"* or any kind of map are aggregated. The remaining requests of information are aggregated to the group about information on the device, like tutorials and key applications, and special functions, which contains any bookmark or search function.

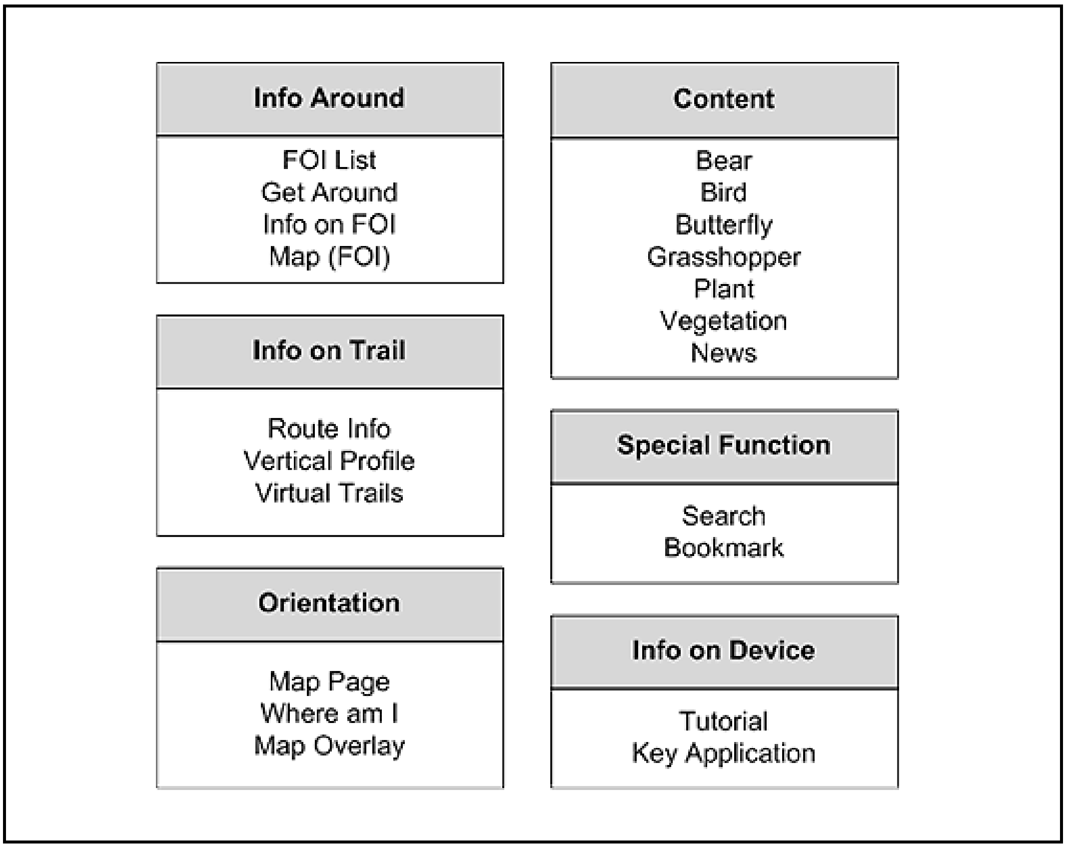

FIGURE 18 : GROUPS OF INFORMATION AND THEIR SUBGROUPS.

4.3 Trails

The request points can easily be allocated to one of the three trails due to the distinct spatial differentiation of the trails (Figure 19).

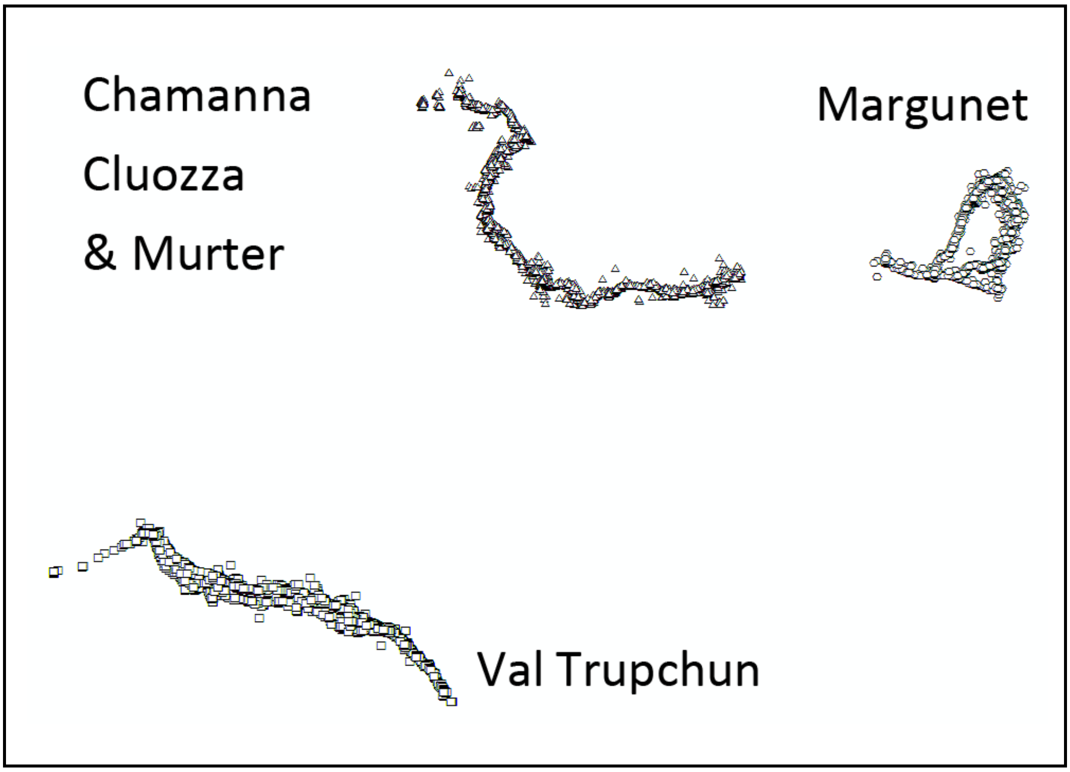

FIGURE 19 : SPATIAL DIFFERENTIATION OF THE TRAILS.

4.4 User groups

Model definition

The duration of stay at a single picnic area needs to be categorized first. A first group can be defined as those who do not have a single log entry near a picnic area. These users either do not hike in this area or they have the device turned off when they pass by a picnic area. All other user groups have valid points inside the picnic areas.
Derived from the idea of Dias et al. (2008), a differentiation between those users who traverse the picnic area without slowing down and those who stay for a certain time at the area can be made. Further differentiation of those who stay might be possible later. These thoughts lead to the following model (Figure 20):

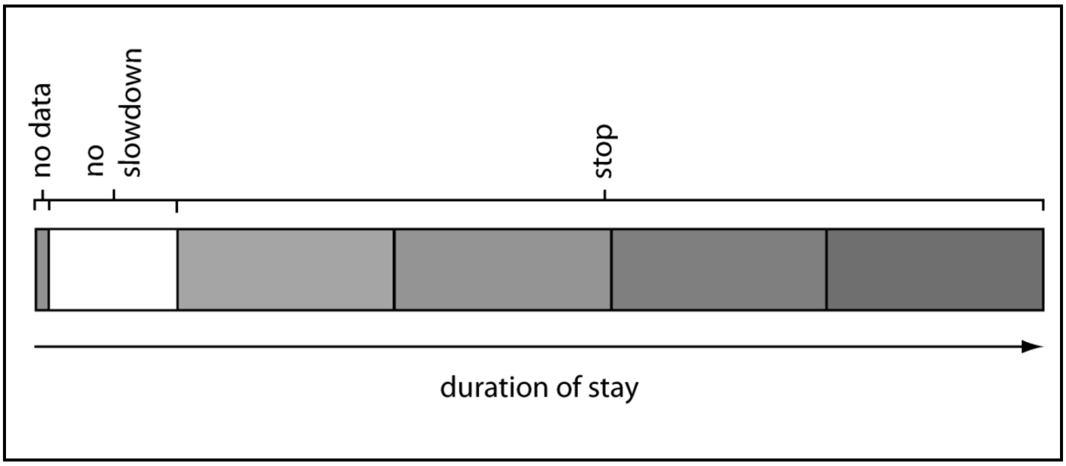

FIGURE 20 : ILLUSTRATION OF USER GROUPS AT A SINGLE PICNIC AREA. THE DURATION OF STAY DEFINES IN WHICH CATEGORY THEY FALL.

What is considered as a stop and what is considered as no slowdown has to be read from data, because the user's behavior cannot be anticipated.

After grouping the users at every picnic area, the problem becomes multidimensional for grouping them over all picnic areas. If N is the number of classes at a single picnic area and k is the number of regarded picnic areas, the number of classes in a k-dimensional perspective becomes N^k. This number can be reduced, if the behavior in a specific picnic area is not important and all picnic areas are rated the same. This theoretical quantity of classes can be reclassified with the knowledge of the practical appearance of classes. Maybe some abstract behavior is not observed in the data. Another factor that needs to be considered is the number of users in every class. Every class needs to contain a sufficient number of users. It would not make sense to specify the behavior only by a hand full of users.

SELECTION OF PICNIC AREAS

Several factors influence the selection of the picnic areas under study. The possible resting places need to be official ones, because only in these areas hikers are allowed to rest. It is not allowed to rest somewhere else it can be assumed that some hikers do it anyway. But these rests cannot be included due to their heterogeneity. For the official picnic areas other factors have to be considered, because not all picnic areas might be valid candidates. If the picnic area is not visited often, not many users were classified as *"no data"* class, which does not make much sense. A bias is produced if not all picnic areas are taken into the analysis. But this bias is probably not very strong, because most of the hikers probably use one of the most popular picnic areas. A third and last factor that influences the selection of the analyzed picnic areas is the comparability of the different trails. As emphasized in the last section, the users will be aggregated to user groups in a multidimensional space of the time of duration in every picnic area. If a consistent comparison wants to be achieved, the the same quantity of picnic areas needs to be considered for every trail.

For Margunet five or six picnic areas could be chosen. Whereas the picnic area near the street at the bottom left of Figure 21 on page 42 shows only a surprisingly low point density the three picnic areas on the right (east) show a high point density. From these three the middle one shows the lowest point density. But they are all valid candidates to investigate the hiker's behavior. The remaining picnic area in the middle is not a valid candidate, because the point density is too low.

The trails Chamanna Cluozza & Murter have four possible picnic area candidates. The one rightmost in Figure 21 does not suite our purpose, because the point density compared to the other three is too low. The peak near Zernez is also an invalid candidate, because it is not a real picnic area. When the users receive the mobile devices at the visitor's centre in Zernez, they might show an additional interest and activity, because they want to try it out, which would explain the high point density in Zernez.

Only three picnic areas are possible on the trail Val Trupchun. Two of these picnic areas show a high point density, whereas the last one shows a low point density compared to the other two. In order to provide all three trails with three picnic areas all three picnic areas will be selected, however:

A total of nine picnic areas will be further processed. In order to do so, the locations of the picnic areas have first been shifted and then the dimension of the picnic area has to be determined.

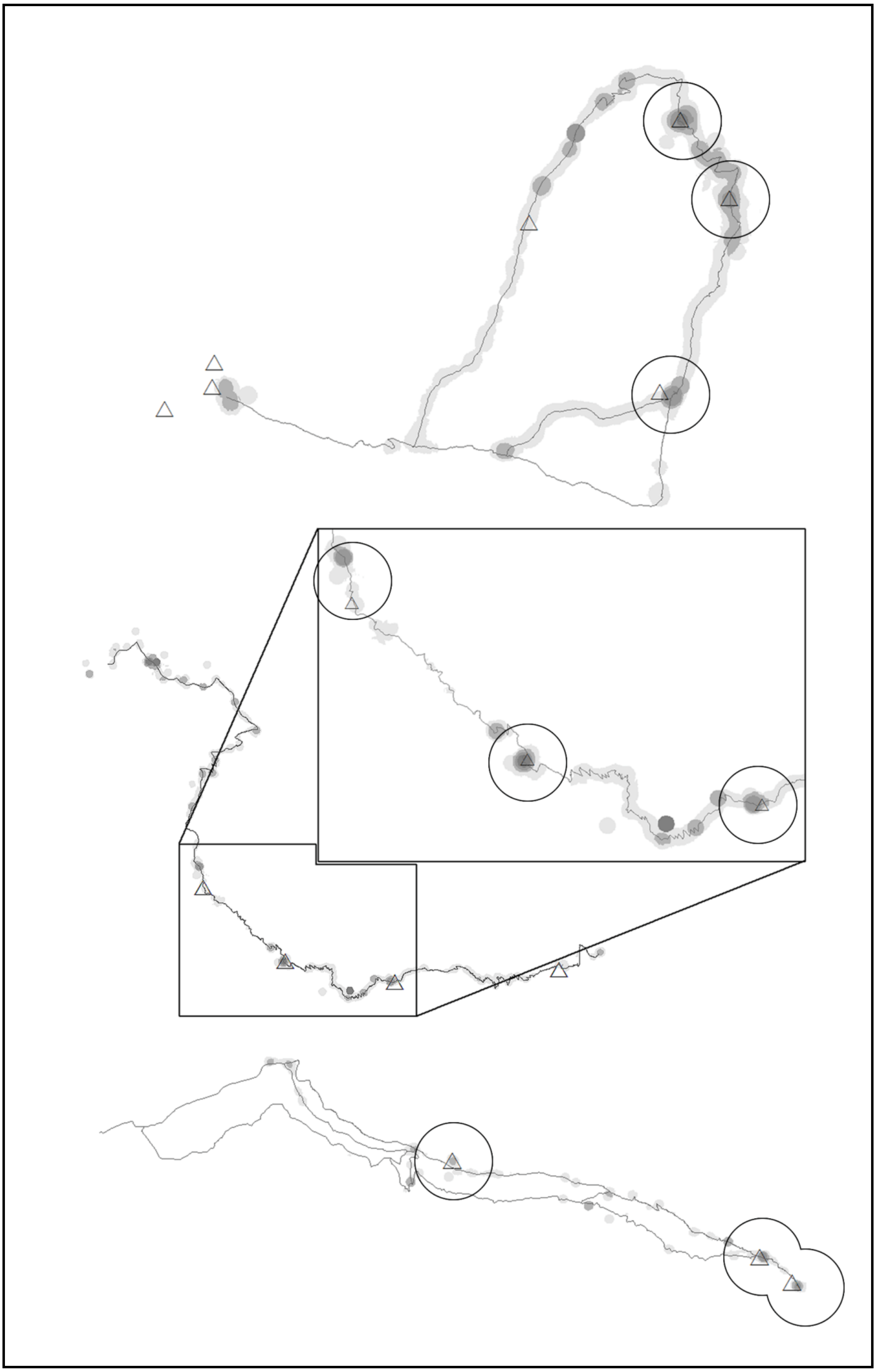

FIGURE 21 : THE TRAILS MARGUNET (TOP) CHAMANNA CLUOZZA & MURTER (MIDDLE) VAL TRUPCHUN (BOTTOM). POSSIBLE RESTING AREAS SYMBOLIZED WITH TRIANGLES, SELECTED ARE CIRCLED.

What seems striking in Figure 22 is the fact that the resting areas are not located at the very centre of the point density. Small displacements can be explained by the geometry or dimensionality of the resting areas. The data set contains only points. But the resting areas must be considered as fuzzy areas, which are not necessarily round or homogenous. E.g. the view from one particular side could be much better than on the other side of the picnic area. This would lead to a heterogeneous usage of the picnic area.

Major shifts can only be explained by errors in the data. Ruedi Haller[1] from the SNP ensured upon inquiry that there are some errors in the data. The very first stop on the trail Chamanna Cluozza is not even an official picnic area. But because of the very nice view that this spot has it is tolerated to rest there. Therefore a new dataset has to be created, where the picnic areas are located at the centers of the point density. In Figure 22 the original points are illustrated in white, whereas the new points are the black dots. Circles with distances 25, 50, 75, and 100 meters are drawn around these black dots in order to visualize the offsets.

[1] E-mail of March 13th 2009.

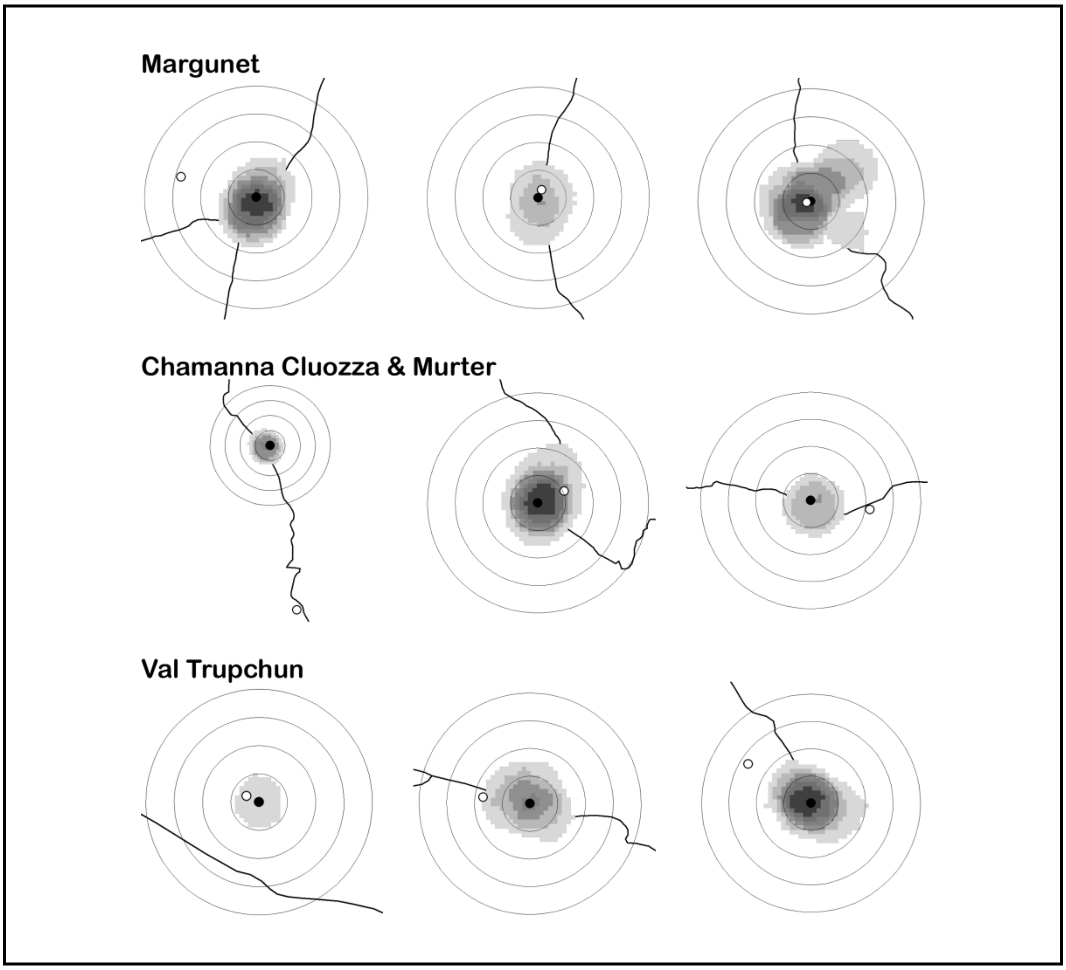

FIGURE 22 : NEW POINTS (BLACK DOTS), AND OLD POINTS (WHITE DOTS). CIRCLES WITH 25-100M DISTANCE TO VISUALISE THE SHIFTS.

The shift of the data is small in most instances. But on the trail Chamanna Cluozza & Murter, the shift is larger than 200 meters, which can be explained by the error in the dataset. What can also be seen is that the shapes of the point densities indicate that areas around the picnic areas are not used homogeneously. E.g. the third picnic area on the trail Margunet has a strongly anisotropic shape.

DIMENSION OF A PICNIC AREA

So far the picnic areas were defined as points with no dimension. In order to relate the picnic areas to the GPS points and also to allocate those points to the picnic areas, some kind of distance is needed. It has already been mentioned that the shape of the picnic areas cannot be assumed to be round, because of the heterogeneous use of the space around the centre of highest density. The points could be selected manually, but then the calculation process would be hard to reproduce. But the model has to be simplified in order to provide comparability between the trails. Therefore a circle around the centre of highest density is assumed to be the border of the picnic area and the radius of the circle has to be determined empirically.

Computationally this means the distance to the centre points has to be determined usingthe *near* function of ArcGIS. The radius can be determined empirically by looking at the frequencies in dependency on the distance to the center of highest density (Figure 23). It seems that in most of the cases the activity decreases with increasing distance, and after 50 meters the activity reaches a level of low activity. Even though there are some differences between picnic areas, one value for all nine picnic areas must be set for better comparability. Fifty meters seems to be a reasonable value for our purpose, so all points inside of a fifty meter buffer will be considered as points inside a picnic area.

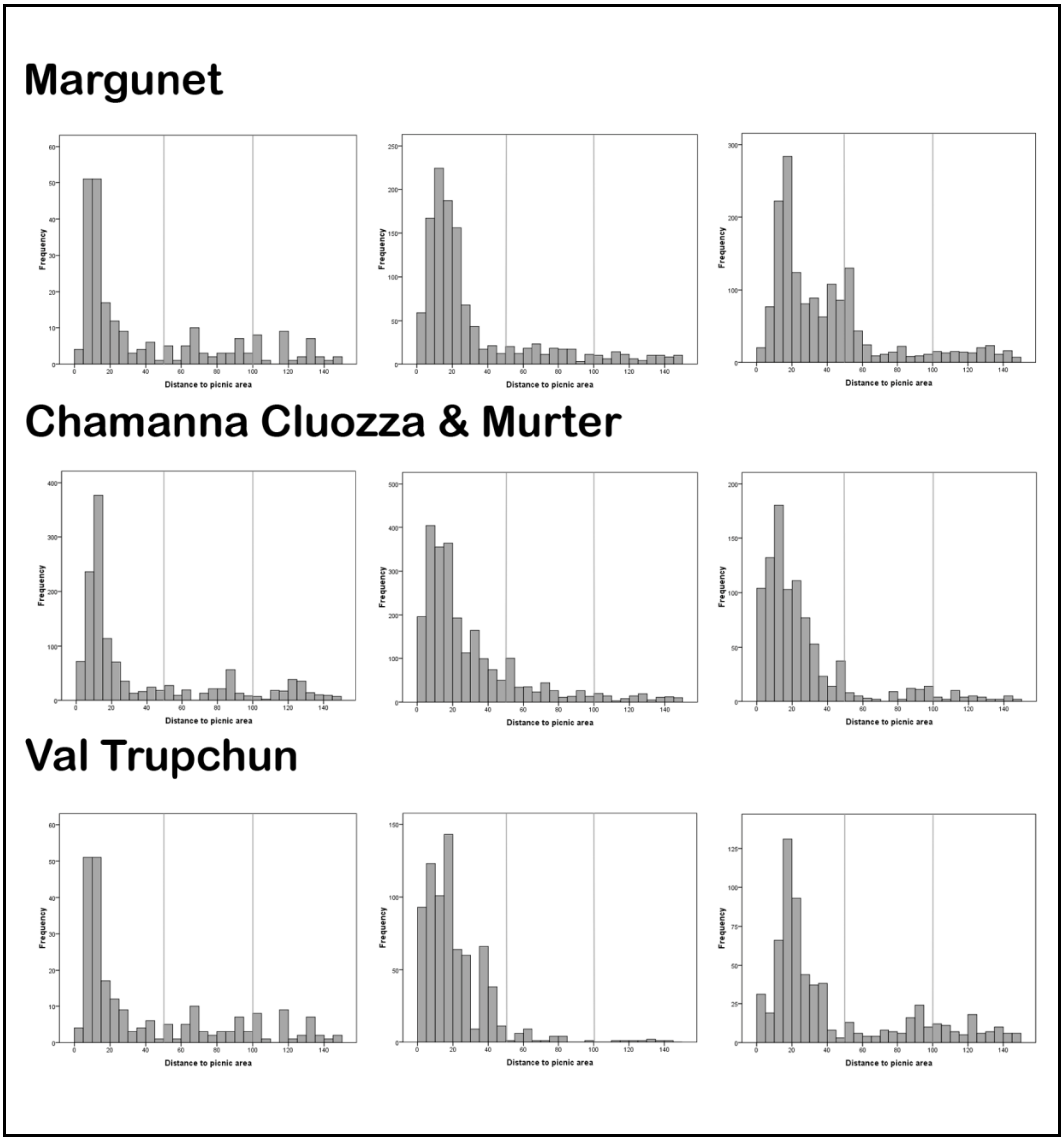

FIGURE 23 : HISTOGRAMS OF POINTS. X-AXIS IS DISTANCE TO THE CENTRE RANGING FROM 0 – 150 METERS (5 METER AGGREGATION), Y-AXIS IS THE FREQUENCY.

After setting the distance and determining what spatial dimension the picnic areas have, all variables are set for the calculation of the duration of stay. The basic function or the pseudo code to process the duration of stay is very simple and only needs a definition of some basic functions as minimum and maximum functions (Code 3).

```
For every user

      getMinTimeAtArea

      getMaxTimeAtArea

      calcDuration

      ApplyDurationToUserId
```

CODE 3 : PSEUDOCODE TO PROCESS THE DURATION OF STAY FOR A CERTAIN USER.

Each day and device has its own user identification (userid), which means that a user can utilize the mobile device only for one day. On the second day, every user would get a new *userid* and is therefore considered as a new user. Setting these *ids* can be realized with a while statement in python (Code 4).

```
gp = arcgisscripting.create()
users = gp.UpdateCursor("$path\\$shapefile ",
                         "", "", "", "USRID A")

TempID=0
user = users.Next()
while user:

       if user.USRID == TEMPID:
       user = users.Next()

 else:
       TEMPID = user.USRID
       User = users.Next ()
```

CODE 4 : SIMPLE ITERATION THROUGH THE USER IDENTIFICATIONS WITH ARCGIS AND PYTHON.

The *UpdateCursor* can be sorted by a certain variable. It is possible to exploit this function to easily get the maximum or minimum from a certain group. If just the first value of a sorted variable is read out the maximum or minimum, depending on whether the cases are sorted in ascending or descending order, can be read out in a simple way (Code 5).

```
rmax = gp.UpdateCursor( "$path\\$shapefile",
                        "USRID = " + str(MOMUSERID)+
                        "and dVT1 <= 50", "", "",
                        "TIMESECOND D")

rmma = rmax.Next()
localmax = rmma.TIMESECOND
```

CODE 5 : DEFINITION OF THE MAXIMUM. (CODE BASED ON ARCGIS 9.2 DESKTOP HELP).

Constraints can be defined for every *UpdateCursor* function. In the example of Code 5 the constraints are that only one *userid* or user is regarded at the same time and the distance to the set centre of the picnic area *dVT1* is smaller than 50 meters. The time (here *timesecond*) is sorted in descending order (*"timesecond D")*, which means that the first line contains the maximum value for this variable. This variable then can be written out in a variable *localmax.* A similar function can be defined for the minimum (*localmin)*. The only difference would be to sort the variables in ascending order (*"timesecond A")*.

The next step in our pseudo code is to calculate the duration of stay. This can easily be achieved by subtracting the *localmin* from the *localmax.* In order to apply the duration of stay to every point the following code has to be applied (Code 6):

```
inserts = gp.UpdateCursor("$path\\$shapefile",
                        "USRID = " +str(MOMUSERID),
                        "", "", "USRID A")
insert = inserts.Next()
while insert:
                        insert.dtVT1 = duration
                        inserts.UpdateRow(insert)
                        insert = inserts.Next()
```

CODE 6 : APPLYING THE VALUES TO ALL THE POINTS.

Similar to the other steps an *UpdateCursor* can be exploited to apply the calculated values to all the points with similar *USRID*. For this purpose the same constraint as in the previous code has to be set, namely the *USRID* must be the same as in the iteration step. Then the value of the duration must be set for the variable (here called *dtVT1*) and the row has to be updated.

Merging all these code parts together, the following code calculates the duration of stay for every user for a single picnic area (Code 7).

```
gp = arcgisscripting.create()
MOMUSERID = 0
#For every user

bounderies = gp.UpdateCursor("$path\\$shapefile ", "distance
<= 50", "", "", "USRID A")
boundery = bounderies.Next()
while boundery:
      #If usrid already applied, line can be skipped.

      if boundery.USRID == MOMUSERID:
            boundery = bounderies.Next()
      else:
            MOMUSERID = boundery.USRID


            #Calculating local maximum

            rmax    =    gp.UpdateCursor("$path\\$shapefile",
            "USRID  =  "  +str(MOMUSERID)+"  and  distance  <=
            50", "", "", "TIMESECOND D")
            rmma = rmax.Next()
            localmax = rmma.TIMESECOND


            #Calculating local minimum

            rmin    =    gp.UpdateCursor("$path\\$shapefile",
            "USRID = " +str(MOMUSERID)+" and distance <=50",
            "", "", "TIMESECOND A")
            rmmi = rmin.Next()
            localmin = rmmi.TIMESECOND


            #Calculating duration of stay

            duration = localmax - localmin

            #Applying values

            inserts  =  gp.UpdateCursor("$path\\$shapefile",
            "USRID = " +str(MOMUSERID), "", "", "USRID A")
            insert = inserts.Next()
            while insert:
                        insert.dtMu2_50 = duration
                        inserts.UpdateRow(insert)
                        insert = inserts.Next()
            boundery = bounderies.Next()
```

CODE 7 : ENTIRE CODE TO CALCULATE THE DURATION OF STAY

Hikers who pass the same place twice still pose a problem. E.g. it is possible to hike on the trail Margunet towards the north, rest at a picnic area and return the same way and pass by other picnic areas twice. Because this behavior is too complex to correct it with a script, has to be corrected manually.

4.4.1 User classification

One-dimensional classification

The group with no points can be classified distinctly. This means that either the hikers did not pass a certain picnic area or did not have the device turned on. In both cases the values for the duration of stay is 0, and it is not possible to distinguish between the two possible behaviors. This means both possible behaviors are labeled with the same group identification.

The next step is to determine the duration of a passing event. The question is, how long is the maximum duration such that it is not considered as a proper stay, but as a pass through event? Two perspectives are considered in order to answer this question: first the deductive perspective. This means it has to be determined deductively how long a normal hiker takes to pass a distance of about 100 meters including e.g. a short stop to watch the panorama and even query some information on the device. Assuming as a rough estimation that a hiker walks at a speed of roughly 3 km/h, which is an estimated speed average given in Robin (2009), he/she would need only about two minutes to walk these 100 meters. If there is some interesting information for the specific user, he/she could spend some minutes standing still and watching the environment. Five minutes can therefore be considered a reasonable estimate.

The second perspective that needs to be considered is certainly the data perspective. How long did the users rest at a certain place? Is it possible to distinguish between the two suggested groups? In Figure 24 we can see the distribution of the frequency for one picnic area split up for the three trails. In all three cases a peak can be seen in the first few minutes. After that the frequency decreases to a level and is more or less constant for the rest of the time. The rest of the distribution does not follow a clear pattern and is therefore not interpretable.

A result from the data perspective is that the frequencies on two of the three trails decrease significantly after around five minutes. These 5 minutes can be interpreted as a pass through event, if it is assumed that on average the hikers do not stop at one of the three possible picnic areas, but only pass through.

The frequencies on all three trails level off after about 70 minutes. If the long stoppers are assumed to be some kind of outlier it can be argued that after that time the stop can be regarded as a long stop. In between 5 and 70 minutes some peaks are visible. After 12.5 minutes at Chamanna Cluozza & Murter an increasing frequency is noticeable. On the other trails this differentiation is not possible. And due to the limited number of users, this local minimum cannot be associated with another user group.

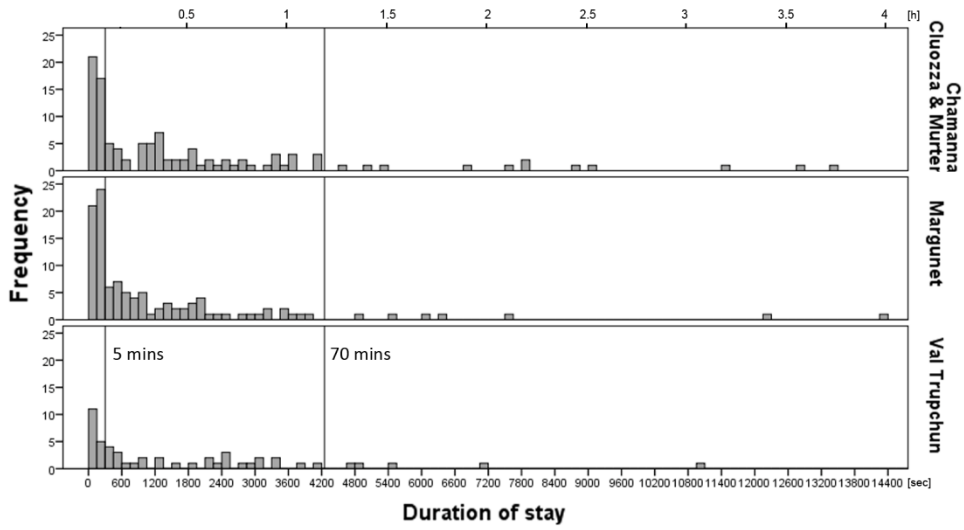

FIGURE 24 : HISTOGRAMS OF THE DURATION OF STAY AT EVERY TRAIL. ONLY ENTRIES THAT ARE BIGGER THAN ZERO ARE ACCOUNTED FOR A BETTER ILLUSTRATION.

These observations of the class limits lead to a classification into four classes with the following frequencies (Table 4).

TABLE 4 : USER CLASSIFICATION WITH RESPECT TO DURATINO OF STAY, WITH FREUENCIES. EVERY USER HAS THREE ENTRIES, AS HIS / HER BEHAVIOR IS LOGGED IN THREE PICNIC AREAS.

Class limits	Class name	Frequency	Percentage
0	No data / no show	331	55%
1-300 sec.	Pass through	99	16%
301-4200sec	Normal stop	149	25%
> 4200 sec	Long stop	24	4%

In 55% of the cases, no data is available at the picnic areas. This means that with three possible stops per trail a hiker on average at least passes through one and a half picnic areas, or on average has no data at 1.65 of 3 resting places, and a hiker on average uses about 30% of the selected picnic for longer than 5 minutes. But only 4 % of the users have entries at a picnic area, which can be considered as a long stop. The average time at a resting place is 12.5 minutes, which would be classified as a "normal stop". The average time at a resting place is about 28 minutes, if only the classes "pass trough", "normal stop" and "long stop" and therefore only users with entries at the picnic areas are considered.

A general problem is that presumably there are some cases in which a hiker stops at a picnic area but his device is not turned on. This has to be considered and a differentiation has to be made, whether there is some information on the hiker's behavior or not. Another general premise for the classification is that the classification has to be very coarse. At Val Trupchun there are only 49 valid entries of hikers, which means that if the users would be classified into four classes with equal sizes, there would only be 12 entries per class, which is not enough for a statistically firm analysis. Under these conditions there are several different options to classify the users over all three picnic areas with the classification of the users in one dimension. One possibility could be to calculate an average duration of stay over the three picnic areas which can be classified similarly as in the one dimensional classification. However, this classification method might cause the problem that an average of three low values, which all fall into the class of passing through, can have the same average as a hiker with one stop and two zeroes. The behavior of the two hikers is clearly not the same. One is passing through every picnic area, the other one selects a specific picnic area to relax at a picnic area. This approach is therefore not suitable for our purpose. An alternative would be to just classify the longest duration of stay. But this would have a similar disadvantage, namely that a hiker who stops at two or three stops is classified just like a hiker that only stops at a single picnic area. Because of the disadvantages of these two classification approaches a decision tree based classification will be applied.

If a hiker has no logging data at any picnic area he/she can clearly be assigned to a *no stop* category. 35 of 201 users fall into this category, which means that about one fifth of the users does not have any data at the selected picnic areas. If there is some information available the user clearly falls into another class. 38 of all users just march through one or multiple picnic areas. To provide a coarse classification it does not matter whether the hiker marches through one or multiple picnic areas. Otherwise the number of users per group would be too small. 38 users correspond to about 19 percent of all users.

Classifying the remaining 128 users who stop at least at one picnic area is not trivial. There are 16 possibilities based on the four classes in one dimension. Almost half of the remaining 128, namely 60 users, stop at a single resting place and do not have data on the other resting places. 25 also stop at a single picnic area, but have a passing through signature at the other picnic areas. To provide a coarse classification, these two groups can be considered as one, because the users from both groups stop only at one picnic area. The remaining 43 users show a very different behavior. Some of them stop at one single picnic area for a very long time. Others rest at more than one picnic area. They are assigned to one single group that can be labeled as *long stoppers.*

The suggested classification is not far away from the one dimensional classification. The two classifications differ mainly when hikers stop shortly at more than one picnic area. Then they are shifted up to the *long stoppers* category.

As already mentioned a hiker can pass through a picnic area twice. This means our values need to be corrected. Some seem unexpected high. For example one user has a resting time of almost four hours at the first picnic area at the trail Margunet. Such high values need to be checked and possibly the class affiliation needs to be adapted manually. The final classes with the corrected frequencies are shown in the following table:

TABLE 5 : SUGGESTED CLASSIFICATION FOR ALL THREE PICNIC AREAS.

Class name	Frequency	Percentage
No data / no show	35	18%
Pass through	38	19%
Normal stoppers	85	42%
Long stoppers	43	21%

At the trail Val Trupchun only one user with the user identification 233 had to be moved from the *long stopper* to the *pass through* class. One user with user identification 57 had to be moved from the *long stopper* to the *normal stopper* class and one user with user identification 121 had to be moved from the *long stopper* class to the *pass through* class at Margunet. There were two cases where some miscalculations took place but did not affect the classification. Both users remained in the long stopper class after the correction. No noticable cases were found on the trails Chamanna Cluozza & Murter. It cannot be ruled out that some cases were missed. But in general the classification seems stable and only a small percentage of users needed to be corrected manually, as only a few hikers came back to a particular picnic place.

4.5 PICNIC AREAS

The picnic areas have already been defined in chapter 4.4. All points within a distance of 50 meters around the center of highest point density are seen as *inside* the picnic areas. Therefore no further processing is necessary.

4.6 TIME VARIABLES

A relative time, namely the duration of stay at the picnic areas, was already introduced in chapter 4.4. The calculation of the relative time for the total time as the trails does, in principle, not differ from calculating relative times at the picnic areas. The code 7 has to be adjusted by removing the constraint of the distances to the picnic areas. All values for a user have to be included. Additionally the values of the maximum and the minimum time of every user's hike is stored in two variables. Other values, such as the relative time since the start, can be calculated by subtracting the absolute minimum from the current time. Also the relative time with respect to the duration of the total journey has to be stored. This can be achieved by dividing the relative time since the start by the total time of the journey, which leads to values from 0 to 1.

4.7 TOPOGRAPHY

All request points should feature the topographic values of the nearest point on the trail. The GPS inaccuracy and also the users breaking the rules of the SNP by not walking on the trails lead on average to a displacement of several meters. The accuracy of the data can be assumed to be about 5-10 m, which will be discussed in chapter 6. A 25 meter *digital elevation model* (DEM) provided from SwissTopo was used as elevation model, even though a model of 4 meters was available for the SNP. The elevation model with a lower resolution was selected due to three arguments: First, the lower resolution provides the whole area, while in the high resolution model some small parts on the trail Val Trupchun are missing. Second, the estimated GPS error is about 5-10 meters, which is up to two times more than the resolution of the high resolution DEM, which would not make much sense. Third, it can be assumed that changes in small areas do not matter for aspect and slope, because the general trend is much more important.

Next, the chosen DEM needed to be processed in several steps as shown in Figure 25 to gain the slope and aspect values on the trails. The first step is to convert the polylines of the trails to raster format with the same parameters as the DEM (cell size, grid spacing) in order to get the values on the trail, and not the general steepness of the flank of the hill. All cells containing trail points are set to 1, the remaining cells to 0. After that, both raster files can easily be multiplied in order to obtain the height values only on the trails, which then can be used to calculate slope and aspect.

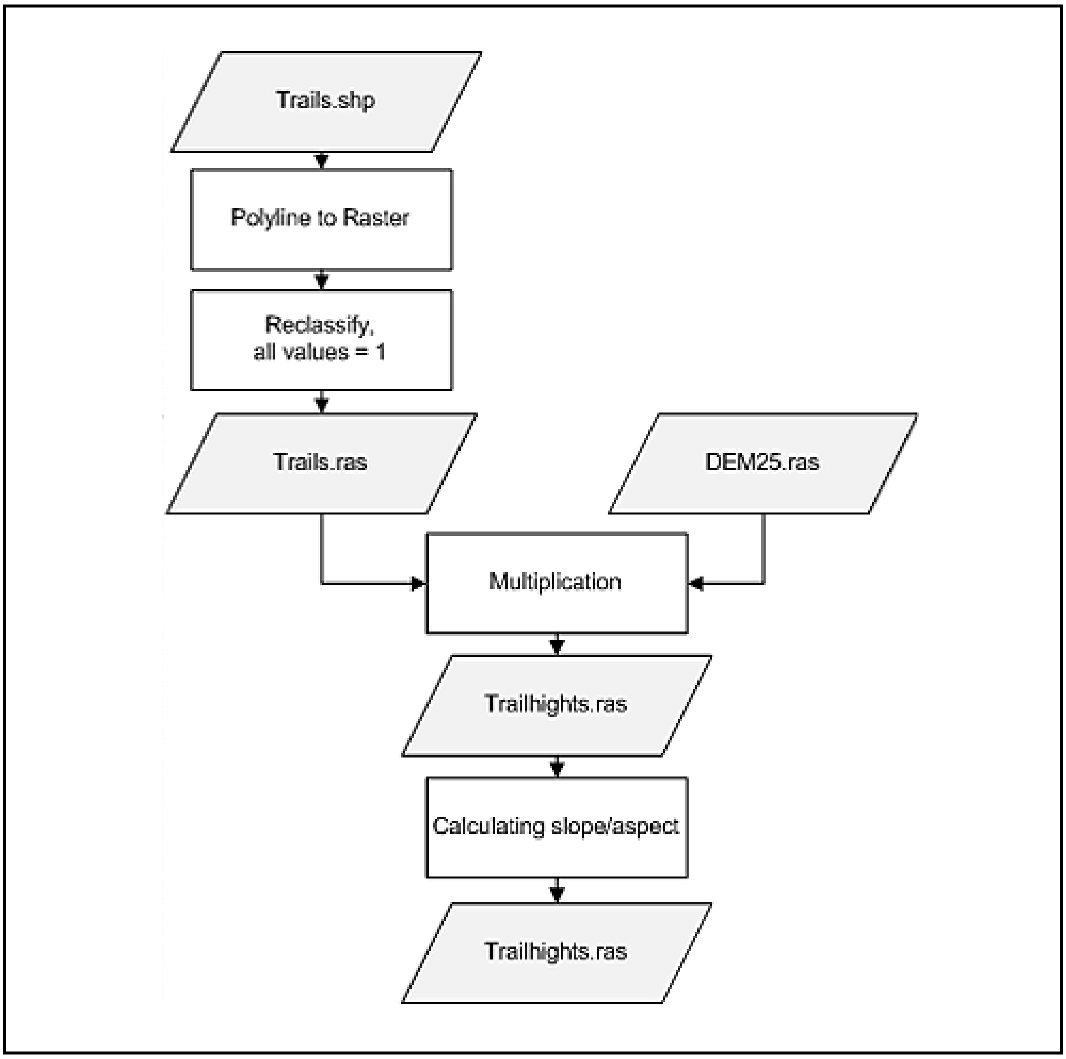

FIGURE 25 : WORKFLOW TO PROCESS SLOPE / ASPECT ON THE TRAILS.

ALGORITHMS

The following definitions are based on Zhou and Liu (2008). The drops in E-W and N-S direction are crucial for the aspect and the slope.

$$S = arctan\sqrt{p^2 + q^2}$$

$$A = 180° - arctan\left(\frac{q}{p}\right) + 90°\left(\frac{p}{|p|}\right)$$

S is the slope, *A* the aspect, *p* the gradient in W-E direction, and *q* the gradient in N-S direction.

$$p = f_x = \frac{\delta f}{\delta x}$$

$$q = f_x = \frac{\delta f}{\delta y}$$

To improve the simplicity of the function, the following window is defined for every centre cell Z.

A	**D**	**F**
B	**Z**	**G**
C	**E**	**H**

The difference between all surrounding cells and the centre cell will be calculated with:

$$d_i = i - Z$$

i stands for *A, B,C, D, E, F, G,* or *H* in this equation. But all corner cells A, C, F, and H have to be normalized by √2 because of the additional diagonal distance to the centre cell. The calculation therefore is a combination of the third-order finite difference weighted by the reciprocal of square distance (Horn, 1981) and the finite difference weighted by the reciprocal of distance (Unwin, 1981). In our calculation, both aspects, the additional distance of a corner cell, as well as the equal weighting of every surrounding cell, are included. The formulas for *p* and *q* are therefore:

$$p_{i,g} = \frac{1}{i * g}(\frac{1}{4}(dA + 2 * dB + dC) - \frac{1}{4}(dF + 2 * dG + dH)$$

$$q_{i,g} = \frac{1}{i * g}(\frac{1}{4}(dA + 2 * dD + dF) - \frac{1}{4}(dC + 2 * dE + dH)$$

In these equations *g* is the grid spacing and *i* either 1 or 2, depending on the missing values. The default value is 2. All distances of the corner cells are divided by the factor √2. The challenge is to be able to deal with missing values, because the DEM contains rasterized polylines with a lot of missing values in a 3x3 window. E.g. if no values are available in W-E direction, only the N-S direction is taken into account for the calculation. If also *dE* is missing, only the left part of the equation is considered and *i* is set to 1.
No special script is needed for the calculation of aspect, because ArcGIS 9.x handles missing values by replacing them with the value of the centre cell. Therefore, the steepest drop is calculated correctly. The slope script can be seen in the annex.

In order to obtain the slope and aspect values for all GPS points another technique needs to be applied. All raster points of the trails with their slope and aspect values were converted to points with the *raster to point* function of ArcGIS. The areas, in which the corresponding point on the trail is the nearest one, were created with the ArcGIS function *create Thiessen polygons*. Finally, the aspect and slope values could be added to the GPS points with the function *"spatial join"* with the option *"is_within".*

4.8 VEGETATION

A vegetation layer with several different vegetation classes is provided by the SNP. The dataset was collected during the HABITALP campaign (Lotz, 2006). In order to get those attributes on the GPS points the layer was spatially joined to the point dataset. Because the HABITALP layer does only cover 22,036 of the 22,986 points, 950 points were left aside and the neutral relevance of 1 was set as a default.

4.9 WEATHER

The source for our weather data is MeteoSwiss. They provide the University of Zurich with a JAVA application called Climap 7.0. With this application it was possible to extract the temperature and precipitation from the weather station Buffalora, which is located on the *Ofenpass.*

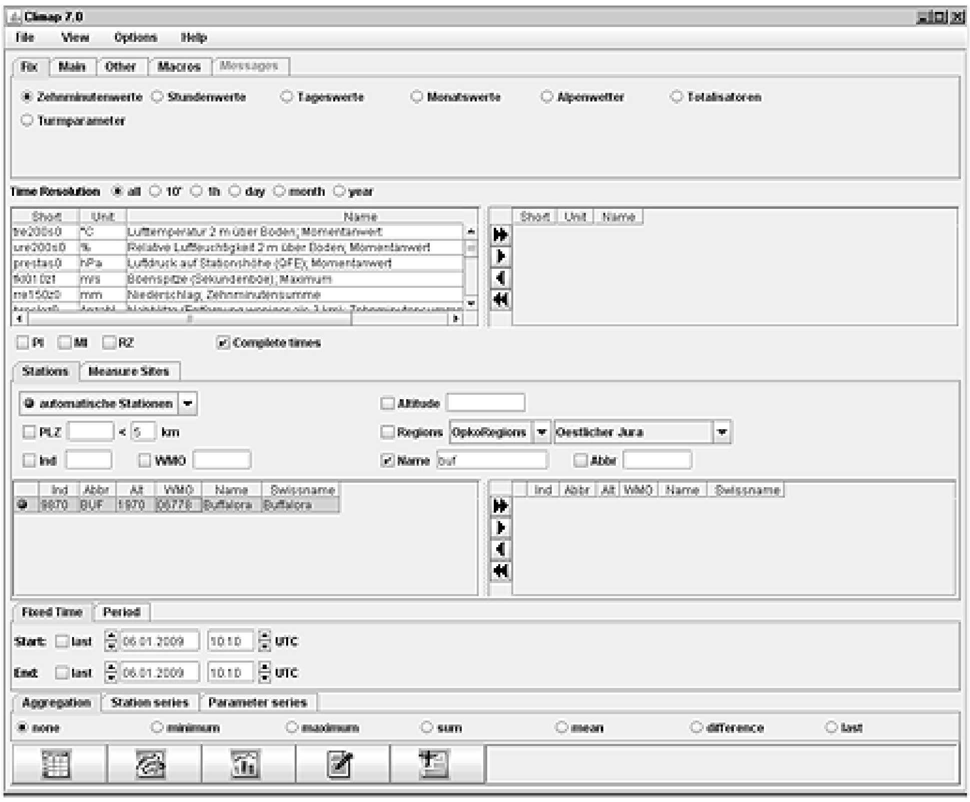

FIGURE 26 : SCREENSHOT OF THE CLIMAP 7.0 APPLICATION.

The variables in the text file from the Climap application are tab separated and can therefore be imported by EXCEL or SPSS.

stn	time	tre200h0	rre150h0
969	200706010000	5.4	0.0

FIGURE 27 : EXAMPLE OF A STRUCTURE OF THE WEATHER FILE.

After reading the data in, the temperature and precipitation information could be joined on the time information of the requests. The dataset contains only hourly averages. But they were joined at the GPS points which had a temporal resolution of seconds. This problem was solved by just keeping the coarse temporal resolution of the weather data and joining them at the corresponding hours of the GPS points.

5 RESULTS

5.1 INTRINSIC RELEVANCE

The intrinsic relevance of an information piece was introduced in chapter 2.3.2. It is calculated by dividing the counted requests of an information (sub)group by the standard request, which is the total number divided by the number of (sub)groups or just the average (sub)group size.

TABLE 6 : FREQUENCIES OF THE REQUESTED INFORMATION.

	Frequency	Intrinsic relevance
Info around	9115	2.38
Info on trail	5433	1.42
Content	3426	0.90
Orientation	2669	0.70
Special function	1338	0.35
Info on device	1005	0.26
Average group size	3831	1.00

Information about the surrounding area is the most frequently requested information. Information on the trail is also more important than the average. All the other requested information groups are less important than the average.

The distribution of the information subgroups can be seen in Table 7. The most frequently requested information subgroups are the *"route info"*, *"the FOI list"* and the *"get around"* function, whereas information on *"grasshoppers"*, *"bears"* and the *"map overlay"* function were requested the least. Seven information subgroups were requested above average and fourteen subgroups were requested below average, which means that some information subgroups are much more frequently requested than the average, while a lot of information subgroups are not requested very often. If the intrinsic relevance of each information subgroup is plotted in descending order, they form an inverse function on their rank. The distribution can be modeled with the function 3.358*exp(rank*-0.141) and has and R^2 of 0.98, which is very high.

TABLE 7 : INTRINSIC RELEVANCES OF SUBGROUPS.

	Frequency	Intrinsic relevance
Route Info	3646	3.33
FOI List	3354	3.06
Get Around	3296	3.01
Info on FOI	1877	1.71
Map Page	1645	1.5
Vertical Profile	1511	1.38
Vegetation	1245	1.14
Average group size	1095	1.00
Bookmarks	878	0.8
Where am I	802	0.73
Key Applications	659	0.6
Map(FOI)	588	0.54
Bird	538	0.49
Plant	481	0.44
Butterfly	464	0.42
Search	460	0.42
Tutorial	346	0.32
News	334	0.31
Virtual Trails	276	0.25
Map Overlay	222	0.2
Bear	196	0.18
Grasshopper	168	0.15

5.2 TRAILS

5.2.1 HOW MUCH INFORMATION WAS REQUESTED?

The first statement that can be derived from Table 8, is that there seem to be differences in the average request rate per trail. This means that the hikers on the trail Margunet on average request about 40 percent more information than hikers on the trail Val Trupchun. But the standard deviation is high on all three trails. Therefore this connection must be tested first with the Levene test (Levene, 1960) on the homogeneity of variances and afterwards with another test on the average value, for instance Duncan's test (Duncan, 1955). Homogeneity can therefore be assumed with a level of significance of 0.05. But it must be assumed that the average values do not differ from each other and the quantity of information that is needed on every trail must be assumed to be equal.

TABLE 8 : QUERIES PER USER OF THE THREE USER GROUPS, QRY/USR ROUNDED TO INTEGERS, THE NORMALIZED VALUES ARE ROUNDED TO TWO DECIMAL PLACES.

	Users	Queries	QRY / USR	Norm,	Stdev.
Cha. Cluozza & Mu.	92	10550	115	1.03	91
Margunet	64	7949	128	1.15	93
Val Trupchun	45	4497	92	0.82	72
Total	**201**	**22986**	**112**	**1**	**88**

5.2.2 WHAT KIND OF INFORMATION WAS REQUESTED?

In all six information groups, significant differences, which are indicated by a standardized residual that is greater than 2 or smaller than -2, can be observed (Table 9). On the trail Chamanna Cluozza & Murter less information about the surrounding area is requested than on the trail Margunet. On the other hand, much more information about the trail is asked on Chamanna Cluozza & Murter. The biggest difference between two trails can be found in the information class "*orientation*". On Chamanna Cluozza & Murter almost twice as much "*content*" was requested as on the trail Val Trupchun. But far more information that is for orientation purpose is requested on the trail Val Trupchun, meaning that map and other orientation functions are used much more. Special functions and information on the device were the least requested on the trail Val Trupchun. The standardized residuals are greater than 2 in almost every cell but in the cells *"info on device"* and *"content"* on the trail Margunet. This indicates that the observed values differ from the expected values.
With the Chi^2 test and a level of significance of 0.05 it can be determined that the mobile information need is depending on the trails. Also with a Chi^2 test on the information groups, it can be tested whether the subgroups within the information

groups differ from each other. In all six cases an independency cannot be assumed and therefore the subgroups must also be analyzed.

TABLE 9 : CROSSTABULATION OF THE INFORMATION GROUPS AGAINST THE TRAILS.

		Ch. Clu. & Murter	Margunet	Val Trup-chun	Total
Info around	Count	3228	3887	2000	**9115**
	Expected	4173.6	3152.1	1783.3	
	Rel. Score	**0.77**	**1.23**	**1.12**	
Info on trail	Count	2944	1409	1080	**5433**
	Expected	2491.2	1878.8	1062.9	
	Rel. Score	**1.18**	**0.75**	**1.02**	
Content	Count	1857	1134	435	**3426**
	Expected	1571.0	1184.8	670.3	
	Rel. Score	**1.18**	**0.96**	**0.65**	
Orientation	Count	1350	676	643	**2669**
	Expected	1223.8	923.0	522.2	
	Rel. Score	**1.10**	**0.73**	**1.23**	
Special function	Count	635	515	188	**1338**
	Expected	613.5	462.7	261.8	
	Rel. Score	**1.04**	**1.11**	**0.72**	
Info on device	Count	526	328	151	**1005**
	Expected	460.8	347.5	196.6	
	Rel. Score	**1.14**	**0.94**	**0.77**	
Total	**Count**	**10540**	**7949**	**4497**	**22986**

Differences in the group "Info around"

The general trend of the information group points towards a higher relevance of *"information around"* on the Margunet trail, and a lower relevance on the trails Chamanna Cluozza & Murter (Table 10). Almost all absolute values of the standardized residuals are greater than two, which indicates that there is a significant and also strong dependency between the subgroups in the information class *"info around"* and the trails themselves. All four subgroups have a value >1 on the trail Margunet, while all values are comparatively small on the trails Chamanna Cluozza & Murter.

In detail, as illustrated in Table 10, there is a substantial difference with regards to the FOI list. This list was, in relation to the trails Chamanna Cluozza & Murter, requested more than twice as much on the trail Val Trupchun.

This can be explained by the user need to locate the FOI, because they might not be as obvious as on the other trails. Another possibility is the additional orientation purpose of the map, which corresponds to the generally higher need of orientation information on that specific trail. If the *"FOI List"* and the *"get around"* function are combined, because they have a similar purpose, the index for both trails, Val Trupchun and Margunet would even out on a level of about 1.21. The general difference between these two trails originates therefore in the function *"info on trail"*, which was requested more on the trail Margunet.

A possible interpretation can be the distribution of FOI in the area. Because there might be more of them on the trail Margunet, they might be requested more. Another factor could be the visibility of these features near the trail.

TABLE 10 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO AROUND" AGAINST THE TRAILS.

		Ch. Clu. & Murter	Margunet	Val Trupchun	Total
FOI List	Count	1051	1297	1006	**3354**
	Expected	1537.94	1159.88	656.18	
	Rel. Score	**0.68**	**1.12**	**1.53**	
Get Around	Count	1231	1492	573	**3296**
	Expected	1511.35	1139.82	644.83	
	Rel. Score	**0.81**	**1.31**	**0.89**	
Info on FOI	Count	725	782	370	**1877**
	Expected	860.68	649.10	367.22	
	Rel. Score	**0.84**	**1.20**	**1.01**	
Map (FOI)	Count	221	221	221	**588**
	Expected	269.62	203.34	115.04	
	Rel. Score	**0.82**	**1.09**	**1.92**	
Total	**Count**	**3228**	**3887**	**2000**	**9115**

DIFFERENCES IN THE GROUP "INFO ON TRAIL"

Generally the *"information on trail"* is more important on the trails Chamanna Cluozza & Murter than on the other two trails (Table 11). This trend is visible in both subgroups *"route info"* and *"vertical profile",* while the function *"virtual trails"* follows a different trend. But only one standardized residual for the *"virtual trails"* is higher than 2. Therefore it can be assumed that they are requested more on the trail Margunet, while they were requested constantly on the other two trails. A possible interpretation of the distribution of the *"vertical profile"* might be the specific vertical profiles of the trails. Val Trupchun has a very homogenous profile, which rises towards the south-east, while the vertical profile for the trail Margunet is also clearly defined, because 75% of the trail is oriented towards the south and 20% is oriented towards west. Because the geometric shape of the trail is basically a circle and the distribution of the aspect is mainly towards north, the result is a positive gradient towards north and a negative gradient towards south. The distribution of the aspect as well as the vertical profile on the trails Chamanna Cluozza & Murter are more complicated, and therefore information to predict the hike such as the *"route vertical profile"* is more important. The influence of aspect will be further discussed in chapter 5.6. A plausible interpretation for the distribution for the information subgroups *"route info"* and *"virtual trails"* could not be found.

TABLE 11 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO ON TRAIL" AGAINST THE TRAILS.

		Ch. Clu. & Murter	Margunet	Val Trupchun	Total
Route info	Count	1877	974	795	**3646**
	Expected	1671.84	1260.86	713.31	
	Rel. Score	**1.12**	**0.77**	**1.11**	
Vertical profile	Count	959	308	244	**1511**
	Expected	692.85	522.53	295.61	
	Rel. Score	**1.38**	**0.59**	**0.83**	
Virtual trails	Count	108	127	41	**276**
	Expected	126.56	95.45	54.00	
	Rel. Score	**0.85**	**1.33**	**0.76**	
Total	**Count**	**2944**	**1409**	**1080**	**5433**

DIFFERENCES IN THE GROUP "CONTENT"

The trend is that hikers on the trail Val Trupchun show a lower interest in the category *"content"* than their equivalents on the other trails. Looking at Table 12 it becomes clear that there is indeed a trend that is visible in almost every information subgroup but the *"news"* and the *"bear"* category. The *"news"* category shows an equal relevance on all three trails, while the *"bear"* class is the only class that shows the least importance on the trail Margunet. In all other classes the lowest relevance values were achieved on the trail Val Trupchun. Those five classes can qualitatively be aggregated to two virtual groups: static and moving objects. The static objects such as plants and vegetation show smaller differences between the trails, but the general trend is still visible, whereas the moving group with butterflies, grasshoppers and birds, which are also small animals, show huge differences comparing the trails Val Trupchun and Chamanna Cluozza & Murter. On the trail Val Trupchun information on butterflies and grasshoppers was requested 5 times less frequently than on the trail Chamanna Cluozza & Murter, and information on birds was requested over twice as many times. What has been stated qualitatively can also be observed with the standardized residuals. A striking observation is that only the moving animals have large standardized residuals on the trail Margunet. All other values seem not to differ from the average.

Because animals have widespread habitats, a possible interpretation approach concerns the spatial appearance of the flora and fauna. Margunet for instance is known for its birds (bearded vulture). But interestingly, these "*bird"* requests were the most requested on Chamanna Cluozza & Murter. On the other hand this very trail is known for mixed forests (Robin, 2009), which might manifest themselves also in the requests. But other factors certainly also play a role. For instance the expected values are relatively small and therefore the scores might be random to a certain degree, because also the number of users requesting such information is relatively small. Also, the interesting information on the trails might already be implemented in FOI's, which are requested in the information group *"info around".* Therefore, the differences in flora and fauna can only partly explain the distribution.

TABLE 12 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "CONTENT" AGAINST THE TRAILS.

		Ch. Clu. & Murter	Margunet	Val Trupchun	Total
Bear	Count	105	56	35	**196**
	Expected	89.87	67.78	38.35	
	Rel. Score	**1.17**	**0.83**	**0.91**	
Bird	Count	287	203	48	**538**
	Expected	246.69	186.05	105.25	
	Rel. Score	**1.16**	**1.09**	**0.46**	
Butterfly	Count	320	117	27	**464**
	Expected	212.76	160.46	90.78	
	Rel. Score	**1.50**	**0.73**	**0.30**	
Grasshopper	Count	118	40	10	**168**
	Expected	77.03	58.10	32.87	
	Rel. Score	**1.53**	**0.69**	**0.30**	
Plant	Count	246	173	62	**481**
	Expected	220.56	166.34	94.10	
	Rel. Score	**1.12**	**1.04**	**0.66**	
Vegetation	Count	622	436	187	**1245**
	Expected	570.88	430.54	243.57	
	Rel. Score	**1.09**	**1.01**	**0.77**	
News	Count	159	109	66	**334**
	Expected	153.15	115.50	65.34	
	Rel. Score	**1.04**	**0.94**	**1.01**	
Total	**Count**	**1857**	**1134**	**435**	**3426**

Generally more information is needed on *"orientation"* on the trail Val Trupchun. In Table 13 it can be seen that this trend is only visible in the subgroup *"map overlay".* But the magnitude of the difference to the average is that big that it is also visible in the superior group *"orientation".* The other two subgroups have a similar distribution. The absolute values of the standardized residuals are greater than two in almost every cell. The only exception is the *"where am I"* function on the trail Val Trupchun, which can be assumed to be equally relevant as on average.

Because the trails are geometrically very different, variations in terms of *"orientation"* can occur. The trail Val Trupchun is branched and it might therefore be more important to know where to go to. On the other hand the trail Margunet has a round geometric shape but not many crossroads. Therefore, orientation might be less difficult.

TABLE 13 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "ORIENTATION" AGAINST THE TRAILS.

		Ch. Clu. & Murter	Margunet	Val Trupchun	Total
Map Page	Count	843	436	366	**1645**
	Expected	754.30	568.87	321.83	
	Rel. Score	**1.12**	**0.77**	**1.14**	
Where am I?	Count	452	198	152	**802**
	Expected	367.75	277.35	156.90	
	Rel. Score	**1.23**	**0.71**	**0.97**	
Map Overlay	Count	55	42	125	**222**
	Expected	101.80	76.77	43.43	
	Rel. Score	**0.54**	**0.55**	**2.88**	
Total	**Count**	**1350**	**676**	**643**	**2669**

The special function has no distinct trend and the standardized residuals are comparatively small (Table 14). It has two subgroups, the *"search"* and the *"bookmark"* function. On the trail Chamanna Cluozza & Murter both functions were requested just on average, where on the other two trails some differences occurred. It is possible to read out of Table 14 that the users on both trails showed complementary interests in those two functions. While the *"search"* function was requested more on the trail Val Trupchun, the *"bookmark"* function was more popular on the trail Margunet.
But the standardized residuals for the *"search"* function are very small, and it can be assumed that it has been requested equally on all three trails. The *"bookmark"* function has standardized residuals which are greater than two. Therefore the differences between the trails Margunet and Val Trupchun can be seen as statistically firm. A possible interpretation approach could be that on the trail Val Trupchun the requested information is only interesting in one location on the trail. Then the *"bookmark"* function would be less interesting, because no second request of the same information would be necessary. But this interpretation approach must be seen as just an idea, because it is rather far-fetched.

TABLE 14 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "SPECIAL FUNCTION" AGAINST THE TRAILS.

		Ch. Clu. & Murter	Margunet	Val Trup-chun	Total
Search	Count	223	148	89	**460**
	Expected	210.92	159.07	89.99	
	Rel. Score	**1.04**	**0.83**	**1.21**	
Bookmarks	Count	412	367	99	**878**
	Expected	402.60	303.62	171.77	
	Rel. Score	**1.01**	**1.26**	**0.52**	
Total	**Count**	**635**	**515**	**188**	**1338**

DIFFERENCES IN THE GROUP "INFO ON DEVICE"

In general, the information group *"info on device"* is distributed very similarly over the three different trails (Table 15). Looking deeper into the subgroups, there is no major difference to the general trend. The biggest difference is observable in the subgroup *"tutorials"* for which the users on the trail Chamanna Cluozza & Murter showed the highest interests, while on the trail Margunet the function seems less interesting. Tutorials might be functions that are depending on the user specific needs. The same can be stated for the *"key applications".* If the group *"information on device"* and *"special function"* are regarded as one big group of functions that do not only provide content information but some kind of interaction, a general trend towards a lower relevance on the trail Val Trupchun can be found.

TABLE 15 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO ON DEVICE" AGAINST THE TRAILS.

		Ch. Clu. & Murter	Margunet	Val Trupchun	Total
Tutorial	Count	195	91	60	**346**
	Expected	158.65	119.65	67.69	
	Rel. Score	**1.23**	**0.76**	**0.89**	
Key Applications	Count	331	237	91	**659**
	Expected	302.18	227.89	128.93	
	Rel. Score	**1.10**	**1.04**	**0.71**	
Total	**Count**	**526**	**328**	**151**	**1005**

5.3 User groups

5.3.1 How much information is needed?

The first question is not what the users need, but how much information the users need. This can be simply answered by focusing on the frequencies of the queries. The variances of the different users can later be tested, whether the variances are equal and whether the groups are varying in order to approve the aggregation of the users to different user classes.

TABLE 16 : QUERIES PER USER OF THE FOUR USER GROUPS. QRY/USR ROUNDED TO INTEGERS, THE NORMALISED VALUES ARE ROUNDED TO TWO DECIMAL PLACES.

	Users	Queries	QRY / USR	Normalized	Stdev.
No data / no show	35	2489	71	0.62	58
Pass through	38	3562	94	0.82	77
Normal stopper	85	10142	119	1.04	89
Long stopper	43	6810	158	1.39	94
Total	201	23003	114	1	88

Table 16 shows the distribution of the frequencies. On average a user requests 112 pieces of information. The hikers who have no logs at the picnic areas only request 60 percent of the average user. Also, the hikers who only pass through the picnic areas request less than average, while normal stoppers request only slightly more information than the average. But what strikes as important is that the long stoppers really have a greater information need. They need 36 percent more information than the average and more than 220 percent more information than hikers who do not have information at the picnic areas. The standard deviation is in all cases comparatively large compared to the average query per user. But the Levene's test of homogeneity shows with a significance of 0.066 that with a level of significance of 0.05 the null hypothesis cannot be rejected and it has to be assumed that the variances of the four different groups are equal. But the *analysis of variance* (ANOVA) of Kruskal-Wallis shows a high level of significance. Therefore, it can be assumed that the user groups do not have the same needs concerning the amount of information they request. With regards to the test of Duncan, it could be shown that the *"long stoppers"* are not in the same group as the rest. On the other hand the class of *"pass through"* and the *"normal stoppers"* or the *"pass through"* and the *"no data"* class could be aggregated, because they do not differ at a level of significance of 0.05.

5.3.2 WHAT INFORMATION IS NEEDED?

In Table 17 all observed and expected values for the information groups and the user groups are presented. As a first observation it can be stated that there are differences and dependencies in all six groups of information. Especially striking are the high differences in the group *"special function"* and the low dependencies in the group *"info on device".* While the *"special functions"* were used much less frequently by the non stopping users, the *"info on device"* is almost distributed homogenously over all four user groups.
Information about the surrounding area is much more requested by users that stop at picnic areas. On the other hand "*information on the trail"*, be it the vertical profile or just general information, is more important for the non stopping hikers. Regarding the regular "*content"*, this differentiation between stopping and non stopping hikers cannot be made. Normal stoppers and the hikers who do not have information at the picnic area seem to have lower interest in static content as the other groups. The long stoppers are less interested in information on "*orientation"*, whereas the other groups especially the pass through group, has a high interest in such information. These observations can be substantiated with the standardized residuals. In the group *"info on device"* the values barely reach a level where a dependency on the user groups can be assumed. The highest value is reached in the information group *"info on trail"* where the no data class has a value of 6.9.
In general there is in fact a difference between the groups that can even be proven with statistical tests like the Chi^2 test on crosstabs.
Summing up the observations, hikers who do stop at picnic areas are less interested in information on the trail, but show a higher interest in information on the surrounding area. Information on the surrounding area is therefore more important if a hiker stops, what corresponds to the assumption that hikers who are focusing on the trail and may hike uninterruptedly notice the environment less than a slow hiker that interrupts his/her trail several times. This also corresponds with the very low interest in the special function of the device. The long stoppers on the other hand have more time to get deeper information on several different things. Therefore information that helps them orientate becomes less important.
If the subgroups are tested with the Chi^2 test of Pearson it appears that the subgroups of the group *"info on device"* are with a significance of 0.553 independent concerning the user groups. Therefore, this particular group does not have to be analyzed further.

TABLE 17 : CROSSTABULATION OF THE INFORMATION GROUPS AGAINST THE USER GROUPS.

		No data	Pass through	Normal stopper	Long stopper	Total
Info around	Count	920	1140	4258	2797	**9115**
	Expected	980.3	1412.5	4021.8	2700.5	
	Rel. Score	**0.94**	**0.81**	**1.06**	**1.04**	
Info on trail	Count	751	990	2226	1466	**5433**
	Expected	584.3	841.9	2397.2	1609.6	
	Rel. Score	**1.29**	**1.18**	**0.93**	**0.91**	
Content	Count	324	653	1308	1141	**3426**
	Expected	368.4	530.9	1511.6	1015.0	
	Rel. Score	**0.88**	**1.23**	**0.87**	**1.12**	
Orientation	Count	287	465	1259	658	**2669**
	Expected	287.0	413.6	1177.6	790.7	
	Rel. Score	**1.00**	**1.16**	**1.07**	**0.83**	
Special function	Count	83	130	700	425	**1338**
	Expected	143.9	207.3	590.4	396.4	
	Rel. Score	**0.58**	**0.63**	**1.19**	**1.07**	
Info on de-vice	Count	107	184	391	323	**1005**
	Expected	108.1	155.7	443.4	297.7	
	Rel. Score	**0.99**	**1.18**	**0.88**	**1.08**	
Total	**Count**	**2472**	**3562**	**10142**	**6810**	**22986**

DIFFERENCES IN THE GROUP "INFO AROUND"

The overall trend in the information group *"info around"* goes towards a higher relevance for stopping hikers than for non-stopping hikers. Looking at the scores of the subgroups in Table 18 two values are particularly striking. First, the groups of hikers who only pass through the picnic areas do have a low interest in *"FOI list"*. On the other hand it seems that the group of hikers with no data at the picnic areas has a high interest in this particular information. Comparing the counts in this subgroup with the subgroup *"get around"*, a complementary distribution is observable. It could be argued that because the subgroups have similar functions only the aggregated group should be analyzed. After aggregating the scores as well as the standardized residuals shrink and it can be stated that only the user groups *pass through* and *normal stop* have remaining trends. While the *pass through* group shows a lower interest in these functions on the surrounding area, it is more relevant for the *normal stopping* group.

The independency of the group *"info around"* on the user groups could not be rejected, because in the other two groups *"info on FOI"* and *"Map (FOI)",* the standardized residuals are small and it seems that there is no dependency on the user groups.
A possible explanation can be that hikers who are stopping at the resting places generally walk a little more slowly and pay more attention to what is around them. But the trends are comparatively small with respect to the similarities of the function *"info around"* and *"FOI list"*, and in some cases the dependency on the user groups cannot always be assumed to be clearly determined.

TABLE 18 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO AROUND" AGAINST THE USER GROUPS.

		No data	Pass through	Normal stopper	Long stopper	Total
FOI list	Count	456	272	1403	1223	**3354**
	Expected	360.70	519.70	1479.90	993.70	
	Rel. Score	**1.26**	**0.52**	**0.95**	**1.23**	
Get Around	Count	213	542	1709	832	**3296**
	Expected	354.50	510.80	1454.30	976.50	
	Rel. Score	**0.60**	**1.06**	**1.18**	**0.85**	
Info on FOI	Count	201	249	891	536	**1877**
	Expected	201.90	290.90	828.20	556.10	
	Rel. Score	**1.00**	**0.86**	**1.08**	**0.96**	
Map(FOI)	Count	50	77	255	206	**588**
	Expected	63.20	91.10	259.40	174.20	
	Rel. Score	**0.79**	**0.85**	**0.98**	**1.18**	
Total	**Count**	**2472**	**3562**	**10142**	**6810**	**22986**

DIFFERENCES IN THE GROUP "INFO ON TRAIL"

In general, the *"information on the trails"* becomes less relevant the longer the users stay at a resting place (Table 19). Looking at the subgroups, in two of three cases, namely in the classes *"route information"* and *"vertical profile"*, this trend is also strongly visible, while for the *"virtual trails"*, a distribution which resembles the distribution of the *"special functions"* can be observed. Looking at the standardized residuals all values for the sub class *"virtual trails"* are smaller than 2, which could mean that the specific information need for this subgroup is not depending on the user group. And if there is a connection, it is not that strong. For the two other subgroups, the standardized residuals are always greater than 2, which indicates that there is indeed a connection between those subgroups and the user groups.

The described distribution can be explained by a possible higher interest in hiking purpose and hiking-relevant information for those who do not stop very long at a picnic area, while the hikers who stop at a picnic area might have an interest in getting more information about the flora and fauna and so forth.

TABLE 19 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO ON TRAIL" AGAINST THE USER GROUPS.

		No data	Pass through	Normal stopper	Long stopper	Total
Route Info	Count	528	657	1513	948	**3646**
	Expected	392.10	565.00	1608.70	1080.20	
	Rel. Score	**1.35**	**1.16**	**0.94**	**0.88**	
Vertical Profile	Count	195	299	595	422	**1511**
	Expected	162.50	234.20	666.70	447.70	
	Rel. Score	**1.20**	**1.28**	**0.89**	**0.94**	
Virtual Trails	Count	28	34	118	96	**276**
	Expected	29.70	42.80	121.80	81.80	
	Rel. Score	**0.94**	**0.79**	**0.97**	**1.17**	
Total	**Count**	**751**	**990**	**2226**	**1466**	**5433**

DIFFERENCES IN THE GROUP "CONTENT"

The general distribution shows no clear trend that could, for example, differ between stopping hikers and those who do not stop. Looking at Table 20 it strikes as important that the *"vegetation"* and *"plants"* have a similar distribution with high indexes and standardized residuals. For the user group of *passing through* hikers all other subgroups of information have very low standardized residuals and follow the general distribution. In the information subgroups *"bird", "grasshopper"* and *"butterfly"* the highest values and variations are in the user group of *long stopping* hikers. Because the standardized residuals are very small, it can be stated that the *"bear"* and *"news"* classes do not show any dependency on the user groups at all.

A possible explanation could be that hikers that only pass through the picnic areas have a generally higher pace and therefore they might have less time and interest in mobile objects. This argument can be supported by the fact that they show less interest in all mobile subclasses of animals except the "*bear*" class. On the other hand, the static *"vegetation"* and *"plants"* have an unexpected distribution and do not follow any trend. Therefore, it cannot be explained.

TABLE 20 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "CONTENT" AGAINST THE USER GROUPS.

		No data	Pass through	Normal stopper	Long stopper	Total
Bear	Count	20	34	70	72	**196**
	Expected	21.10	30.40	86.50	58.10	
	Rel. Score	**0.95**	**1.12**	**0.81**	**1.24**	
Bird	Count	32	72	207	227	**538**
	Expected	57.90	83.40	237.40	159.40	
	Rel. Score	**0.55**	**0.86**	**0.87**	**1.42**	
Butterfly	Count	31	71	278	84	**464**
	Expected	49.90	71.90	204.70	137.50	
	Rel. Score	**0.62**	**0.99**	**1.36**	**0.61**	
Grasshop.	Count	9	20	67	72	**168**
	Expected	18.10	26.00	74.10	49.80	
	Rel. Score	**0.50**	**0.77**	**0.90**	**1.45**	
Plant	Count	53	111	143	174	**481**
	Expected	51.70	74.50	212.20	142.50	
	Rel. Score	**1.03**	**1.49**	**0.67**	**1.22**	
Vegetation	Count	138	296	412	399	**1245**
	Expected	133.90	192.90	549.30	368.90	
	Rel. Score	**1.03**	**1.53**	**0.75**	**1.08**	
News	Count	41	49	131	113	**334**
	Expected	35.90	51.80	147.40	99.00	
	Rel. Score	**1.14**	**0.95**	**0.89**	**1.14**	
Total	**Count**	**324**	**653**	**1308**	**1141**	**3426**

The general trend of the group *"orientation"* is that hikers who stop at any picnic area have a lower need in orientation than the other users (Table 21). This is also visible in the sub classes *"map page"* and *"where am I"*. It seems as if this trend gets disturbed by the subgroup *"map overlay"*, which could also be interpreted as a special function. If the *"map overlay"* function is regarded as such a function, then the observed trend becomes clearer. On this aggregated group, it can be commented that the *long stopping* hikers clearly need less information on pure orientation purpose, where the *pass through* class has the opposite need. The *normal stopping* hikers and the hikers with *no data* at the picnic areas behave similarly with respect to pure orientation functions. This whole distribution could be explainable by the assumption that *passing through* hikers walk for exercise, while on the *no data* no such general statement can be formulated, because they just do not use the device, which is no direct indicator that they hike for exercise. For users who make comparatively long stops other functions become more important than those with an orientation purpose. But the fact that they have an equal interest in the "*map overlay"* function does not fit that line of argumentation. An uncertain assumption could be that the "*map overlay"* function needs more technical skills and the *long stoppers* might not have those, and because it can be assumed that older hikers and children might stay longer at the picnic areas, it could fit the distribution. But this hypothesis is highly debatable and cannot be proven with the given data.

TABLE 21 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "ORIENTATION" AGAINST THE USER GROUPS

		No data	Pass through	Normal stopper	Long stopper	Total
Map Page	Count	180	312	721	432	**1645**
	Expected	176.90	254.90	725.80	487.40	
	Rel. Score	**1.02**	**1.22**	**0.99**	**0.89**	
Where am I?	Count	97	139	373	193	**802**
	Expected	86.30	124.30	353.90	237.60	
	Rel. Score	**1.12**	**1.12**	**1.05**	**0.81**	
Map Overlay	Count	10	14	165	33	**222**
	Expected	23.90	34.40	98.00	65.80	
	Rel. Score	**0.42**	**0.41**	**1.68**	**0.50**	
Total	**Count**	**287**	**465**	**1259**	**658**	**2669**

DIFFERENCES IN THE GROUP "SPECIAL FUNCTION"

As presented in the previous groups, functions like *"virtual trails"* or the *"map overlay"* function are, in general, more popular for users that stay at picnic areas. Looking at the information groups *"special function"* in Table 22 the same distribution is visible in the subgroup *"bookmark"* as well as partially in the requests of a *"search"* function. A clear statement on the *"bookmark"* subgroup can be formulated, because the differences are very high and even the expected values for the non stopping hikers are low, the standardized residuals are still high. Therefore it seems clear that the stopping hikers do use the *"bookmark"* function far more frequently than the non stopping hikers. Especially the hikers with one *"normal stop"* have shown a high interest in this function. But also the *"long stoppers"* have, compared to the *non stopping* hikers, a high interest in this function. This clear statement cannot be transferred to the subgroup *"search",* because there is no general trend which distinguishes the *stopping* from the *non stopping* hiker. At least it can be said that users who have *no data* at the picnic areas are the least interested in search functions. Why the *"long stoppers"* show much more interest in the *"search"* function than the *"normal stoppers"* cannot be solved by argument. Maybe this behavior corresponds to the already proposed level of technical skill of those users, and because their technical knowledge might be below average they use more search functions. Another could be that they just show more interest in features that are neither orientation nor information about the surrounding area, and just like to search information. But these possible explanations are highly debatable and other random factors could be the cause of this distribution.

TABLE 22 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "SPECIAL FUNCTION" AGAINST THE USER GROUPS

		No data	Pass through	Normal stopper	Long stopper	Total
Search	Count	31	82	171	176	**460**
	Expected	49.50	71.30	203.00	136.30	
	Index	**0.63**	**1.15**	**0.84**	**1.29**	
Bookmarks	Count	52	48	529	249	**878**
	Expected	94.40	136.10	387.40	260.10	
	Index	**0.55**	**0.35**	**1.37**	**0.96**	
Total	**Count**	**83**	**130**	**700**	**425**	**1338**

5.4 Picnic Areas

5.4.1 How much information is needed?

The total length of all three considered trails is over 37 km, while only about 800 meters of the total length can be regarded as inside a picnic area, which is only about 2% of the total trail length (Table 23). But a normal hiker stays up to 30% of the time at the national park at a picnic area with an average of 12%. Looking at the queries per distance, at the picnic areas about 135 times more queries were performed than outside of them. The distances have to be regarded with caution, because the hikers do not necessarily hike the total trail length. But even if they just use about a third of the total trail length, the differencebetween in- and outside the picnic areas is still very big. Using the duration of stay as the normalization factor, the hikers request only about 2.4 times more information while being within picnic areas. Therefore both, the length normalization as well as the probably more appropriate time normalization show a clear pattern. More information is requested inside the picnic areas.

TABLE 23 : DESCRIPTIVES OF THE TRAILS AND PICNIC AREAS. QUERIES NORMALIZED PER TIME AND PER LENGTH.

	Length [km]	Queries	$\frac{\text{Queries}}{\text{km}}$	$\frac{\text{Queries}}{\text{h}}$	Norm.
Inside picnic areas	0,8	5823	7553	10039	2.10
Outside picnic areas	36,6	17163	56	4067	0.85
Total	**37,4**	**22986**	**615**	**4775**	**1**

5.4.2 What information is needed?

At first sight it seems that the differences between the request behavior in- and outside the picnic area are more or less equal (Table 24). The indexes from the information group *"info around"* and *"info on trail"* are all very close to 1. The scores beneath rise, but do not reach a very high level. The biggest deviation from 1 is visible in the information group *"special function".*

The standardized residuals for the information groups *"info around", "info on trail"* and as well as for the group *"info on device"* are very small. Therefore it can be assumed that they are not depending on the location in terms of in- or outside the picnic areas. In the information groups *"content"* and *"special function"* high positive standardized residuals inside the picnic areas are observable. Therefore it can be assumed that those two functions are especially interesting inside the picnic areas.

It seems that *"info around"* is equally important inside and outside the picnic areas. It can be argued that while the hikers are resting their interest does not rise compared to the interest while they are hiking. The same can be said for *"info on trail". "Orientation"* functions on the other hand, are the only information group that is more important outside the picnic areas. This seems logical, because orientation functions are more important when someone is moving. Both *"content"* and *"special function"* might be more interesting inside the picnic areas, because they need a higher degree of attention, and while the users are hiking they might just seek their information from the function *"info around".*
Interestingly, only the subgroups of the information group *"info on device"* show no difference at a level of significance of 0.05. Therefore all other groups must be analyzed further. Even though it seems that the information groups *"info around"* and *"info on trail"* show no dependency at all.

TABLE 24 : CROSSTABULATION OF THE INFORMATION GROUPS AGAINST THE PICNIC AREAS.

		Outside	Inside	Total
Info around	Count	6786	2329	**9115**
	Expected	6805.91	2309.09	
	Rel. Score	**1.00**	**1.01**	
Info on trail	Count	6786	2329	**5433**
	Expected	6805.91	2309.09	
	Rel. Score	**1.00**	**1.01**	
Content	Count	2411	1015	**3426**
	Expected	2558.10	867.90	
	Rel. Score	**0.94**	**1.17**	
Orientation	Count	2095	574	**2669**
	Expected	1992.87	676.13	
	Rel. Score	**1.05**	**0.85**	
Special function	Count	879	459	**1338**
	Expected	999.05	338.95	
	Rel. Score	**0.88**	**1.35**	
Info on device	Count	732	273	**1005**
	Expected	750.41	254.59	
	Rel. Score	**0.98**	**1.07**	
Total	**Count**	**17163**	**5823**	**22986**

DIFFERENCES IN THE GROUP "INFO AROUND"

It has already been shown that generally, there is no difference between the behavior of the users inside and outside the picnic area with respect to the information group *"info around".* Looking at the subgroups, the trend seems to be visible in almost every subgroup (Table 25). The *"FOI list"* and the *"Map (FOI)"* function are equally requested inside and outside the picnic areas. The *"get around"* function shows a small trend towards a smaller relevance inside the picnic areas, while the *"info on FOI"* function shows the opposite distribution.

The same trend as has already been described is also visible in the standardized residuals. Interestingly the expected and counted values for the *"FOI list"* show almost perfect equivalence. The biggest difference in terms of standardized residuals is visible in the subgroup *"info on FOI",* which has two values that are greater than 2.

It seems that even though the *"FOI list"* was requested equally, the information on the FOIs was requested more frequently inside the picnic areas. This might be because the users have more time for more information requests, whereas while they are hiking they are maybe more interested where the features are to look at in the real world.

TABLE 25 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO AROUND" AGAINST THE PICNIC AREAS.

		Outside	Inside	Total
FOI list	Count	2504	850	**3354**
	Expected	2504.30	849.70	
	Rel. Score	**1.00**	**1.00**	
Get Around	Count	2531	765	**3296**
	Expected	2461.00	835.00	
	Rel. Score	**1.03**	**0.92**	
Info on FOI	Count	1322	555	**1877**
	Expected	1401.50	475.50	
	Rel. Score	**0.94**	**1.17**	
Map(FOI)	Count	429	159	**588**
	Expected	439.00	149.00	
	Rel. Score	**0.98**	**1.07**	
Total	**Count**	**6786**	**2329**	**9115**

Looking at the distribution in the group *"info on trail"* it seems as if there is almost no difference between inside and outside the picnic areas. But the subgroups as shown in Table 26 seem to neutralize their trends. The *"virtual trails"* are far more used inside the picnic areas, which corresponds to the additional usage of *"special functions"*, whereas the *"route information"* seems to be more popular outside the trail, which can be explained by the assumed need of such information at the beginning of a journey. All analyzed picnic areas are not at the beginning of the trail, which could explain that behavior. Even though there might be some underlying variable, the trend depending on the picnic areas can be confirmed when the standardized residuals are analyzed. They are in both cases greater than 2. The difference in the subgroup *"virtual trails"* is not as big as the relevance index would indicate, because the standardized residuals are much smaller compared with the subgroup *"route information"*. In the subgroup *"vertical profile"* no difference between inside and outside the picnic areas can be ascertained. Both standardized residuals as well as the relevance index show low values.

A possible interpretation for this distribution again is the additional time and attention the hikers spend on the device while they are resting. *"Virtual trails"* need maybe a higher attention and are therefore less requested outside the picnic areas. On the other hand the other *"info on trail"* is more interesting while the users are hiking.

TABLE 26 CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO ON TRAIL" AGAINST THE PICNIC AREAS.

		Outside	Inside	Total
Route Info	Count	2936	710	**3646**
	Expected	2722.40	923.60	
	Rel. Score	**1.08**	**0.77**	
Vertical Profile	Count	1137	374	**1511**
	Expected	1128.20	382.80	
	Rel. Score	**1.01**	**0.98**	
Virtual Trails	Count	187	89	**276**
	Expected	206.10	69.90	
	Rel. Score	**0.91**	**1.27**	
Total	**Count**	**6786**	**2329**	**9115**

DIFFERENCES IN THE GROUP "CONTENT"

In general, more *"content"* information was requested inside the picnic areas than outside (Table 27). But this trend is not very strong. Looking at the subgroups *"vegetation"* and *"plants"* show no visible differences at all. The greatest relevance score is visible for the *"grasshoppers".* The requests on *"birds"* and *"bears"* are similarly distributed and requested more inside the picnic areas.

In terms of standardized residuals the same pattern is observable. The flora has very small residuals, while the highest value is reached in the subgroup *"bird".*

While the users are resting inside the picnic areas, they might be more aware of moving and small animals. Therefore they might request more of this information. The fauna on the other hand might be an equal eye catcher outside the picnic areas, while the users are walking. Therefore no real difference is observable. Interestingly the *"news"* function is also more important inside the picnic areas. The users might just have more time for such things. On the trails they are more occupied by other things and while they are resting they can inform themselves more about the information group *"content"*. Outstanding is the distribution for the *"butterflies"*. They were requested more outside the picnic areas. No plausible interpretation for this distribution could be found.

TABLE 27 CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "CONTENT" AGAINST THE PICNIC AREAS.

		Outside	Inside	Total
Bear	Count	120	76	**196**
	Expected	146.30	49.70	
	Rel. Score	**0.82**	**1.53**	
Bird	Count	314	224	**538**
	Expected	401.70	136.30	
	Rel. Score	**0.78**	**1.64**	
Butterfly	Count	358	106	**464**
	Expected	346.50	117.50	
	Rel. Score	**1.03**	**0.90**	
Grassh.	Count	97	71	**168**
	Expected	125.40	42.60	
	Rel. Score	**0.77**	**1.67**	
Plant	Count	358	123	**481**
	Expected	359.10	121.90	
	Rel. Score	**1.00**	**1.01**	
Vegetation	Count	939	306	**1245**
	Expected	929.60	315.40	
	Rel. Score	**1.01**	**0.97**	
News	Count	225	109	**334**
	Expected	249.40	84.60	
	Rel. Score	**0.90**	**1.29**	
Total	**Count**	**2411**	**1015**	**3426**

In the group *"orientation"* the overall trend shows that orientation functions are more important outside the picnic areas than inside of them. The subgroups show a slightly different distribution as visible in Table 28 for the function *"map page"* as well as for the function *"where am I"*. The opposite trend is visible in the function *"map overlay"*, which corresponds to the additional request of special functions inside the picnic areas.

The biggest difference between outside and inside the picnic areas in terms of standardized residuals is discoverable in the subgroup *"map page".* Interestingly for the other two subgroups the standardized residuals are small and it is not impossible that they are not even depending on the variable picnic area.

A suitable interpretation of the hiker's behavior could be that outside the picnic areas the orientation and especially the map function plays an important role, because the map is needed to continue the trail, while inside the picnic area during the rest other functions such as information on flora and fauna might be more important. Also, special functions such as the *"map overlay"* function are used more often, even though the standardized residuals are comparatively small, because the function is generally not used that often.

TABLE 28 CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "ORIENTATION" AGAINST THE PICNIC AREAS.

		Outside	Inside	Total
Map Page	Count	1330	315	**1645**
	Expected	1228.30	416.70	
	Rel. Score	**1.08**	**0.76**	
Where am I	Count	615	187	**802**
	Expected	598.80	203.20	
	Rel. Score	**1.03**	**0.92**	
Map Overlay	Count	150	72	**222**
	Expected	165.80	56.20	
	Rel. Score	**0.90**	**1.28**	
Total	**Count**	**2095**	**574**	**2669**

In general, far more *"special functions"* are requested inside the picnic areas than outside the picnic areas. This trend is visible in both subgroups, the *"search"* function as well as for the *"bookmark"* function (Table 29).

In both cases all standardized residuals, inside as well as outside the picnic areas, are greater than 2, indicating that there is indeed a dependency on this variable. The highest value 5.3 is observable in the cell *"search"* inside the picnic areas.

A possible interpretation could be that inside the picnic areas the users have more time to perform more complex functions. More simple functions are interesting while the users are hiking and have to concentrate on other things.

TABLE 29 CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "SPECIAL FUNCTION" AGAINST THE PICNIC AREAS.

		Outside	Inside	Total
Search	Count	286	174	**460**
	Exp.	343.50	116.50	
	Rel. Score	**0.83**	**1.49**	
Bookmarks	Count	593	285	**878**
	Exp.	655.60	222.40	
	Rel. Score	**0.90**	**1.28**	
Total	**Count**	**879**	**459**	**1338**

5.5 TIME

5.5.1 HOW MUCH INFORMATION IS NEEDED?

Eisenhut et al. (2008) showed that the distribution of the total dataset of all requests follow a normal distribution. In this thesis only a subset of almost 23,000 selected entries have been analyzed. Figure 30 illustrates the distribution of all requests during a day. Even though the distribution looks like it might be normally distributed, the Kolmogorov-Smirnov test indicates that at a level of significance of 0.05 it is not normally distributed. Some peaks around noon might differ too much from a normal distribution. An imaginable explanation for the differences might be the distribution of the resting time. But the peaks originate from outside the picnic areas, because the distribution of the requests inside the resting areas (dark bar) shows no such peaks. Therefore they originate from the behavior outside the picnic areas. The mean, which is before noon at 11.19, is the same as in the analysis of nearly 80,000 points by Eisenhut et al. (2008). Additionally it can be stated that the distribution has a slightly positive skewness (0.75), which indicates that more information is requested in the morning.

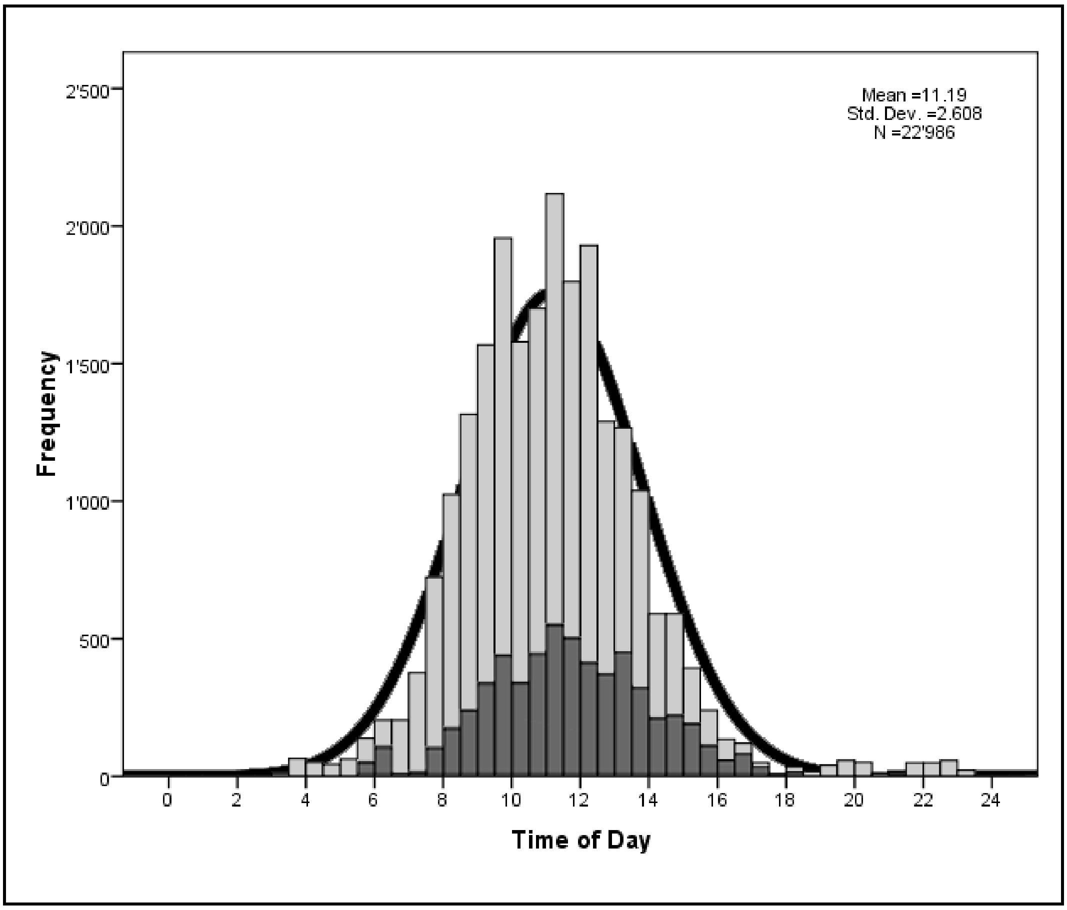

FIGURE 28 : DISTRIBUTINO OF REQUESTS DURING THE DAY, MEAN 11.19, INTERVAL WIDTH 30 MINUTES.

It could be proven that the requests are not normally distributed. An explanation of the distribution is visible in Figure 29. There is also a strong correlation (correlation coefficient 0.935) between the active users and the requests. It can be said that the distribution of the requests in the absolute time scale does not make much sense, unless it is normalized by the number of hikers.

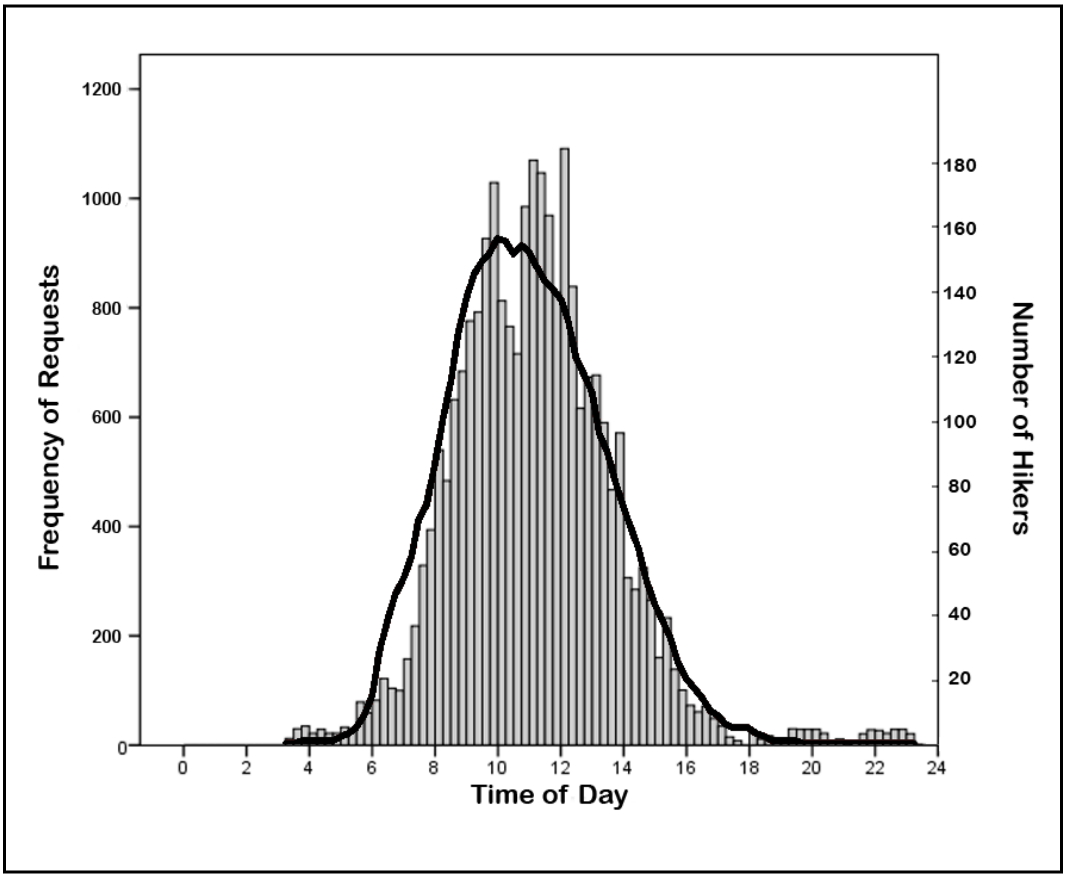

FIGURE 29 : REQUESTS AND USERS ON THE TRAIL, TIMESTEPS 15 MINUTES.

Looking at Figure 30 the total distribution (upper right corner) shows a different trend than expected in Figure 29. There seem to be much more requests per user and 15 minutes very early and very late in the day, as during the most popular hiking periods. But this also has to be put into perspective. Since only a few hikers were on the trails at that time, the number of requests during those 15 minute time periods is strongly depending on only a few hikers. From 6 AM to 5 PM there were always at least 10 hikers on the trails. Looking at this distribution another trend can be observed. It looks like more information is requested around noon than in the morning or in the afternoon. This distribution can be modeled with a quadratic function with an R^2 of 0.45 (Figure 30). Even though the residuals follow a normal distribution and show no signs of heteroscedasticity, the model might not be satisfying, because the function reaches values below zero, which is impossible. But other functions such as a constant function (R^2 of 0.05) and a normal distribution with a mean of 11.36 and a standard deviation of 3.08 are not satisfactory either. Therefore the quadratic function is the best choice of a model.

What seems clear is that there is a difference between early in the morning/afternoon and the maximum at noon. The maximum of the function is at 11:20, which is very close to the underlying frequency function. Although the difference is only a few requests per 15 minutes, the relative difference of 6 AM. to the maximum just before noon is about 30 percent.

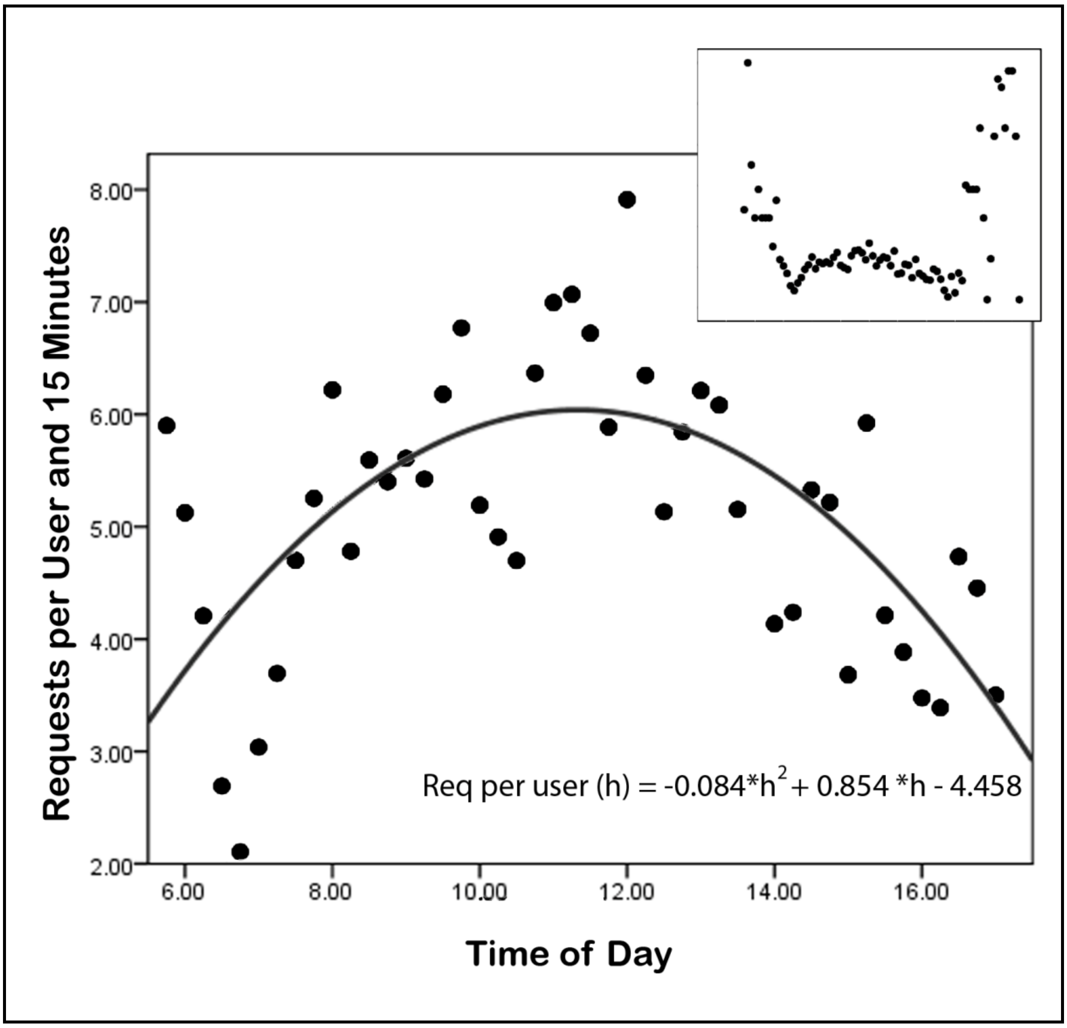

FIGURE 30 : NORMALIZED REQUESTS PER USER AND 15 MINUTES.

Secondly, the distribution of the relative time will be analyzed. Eisenhut et al. (2008) stated that the use of the device becomes less attractive until the end of the journey. Looking at Figure 31 the histogram shows a clear downward trend. But if the number of active users for such a long time is taken into account, the statement has to be put into perspective. Over 150 hikers used the device for at least 4 hours, while only one user used the device for 14 hours. Even if all hikers would behave similarly and constantly over time, the distribution would look similar to the shown histogram. The only outstanding feature is at the beginning where, compared to the number of users, much more information was requested.

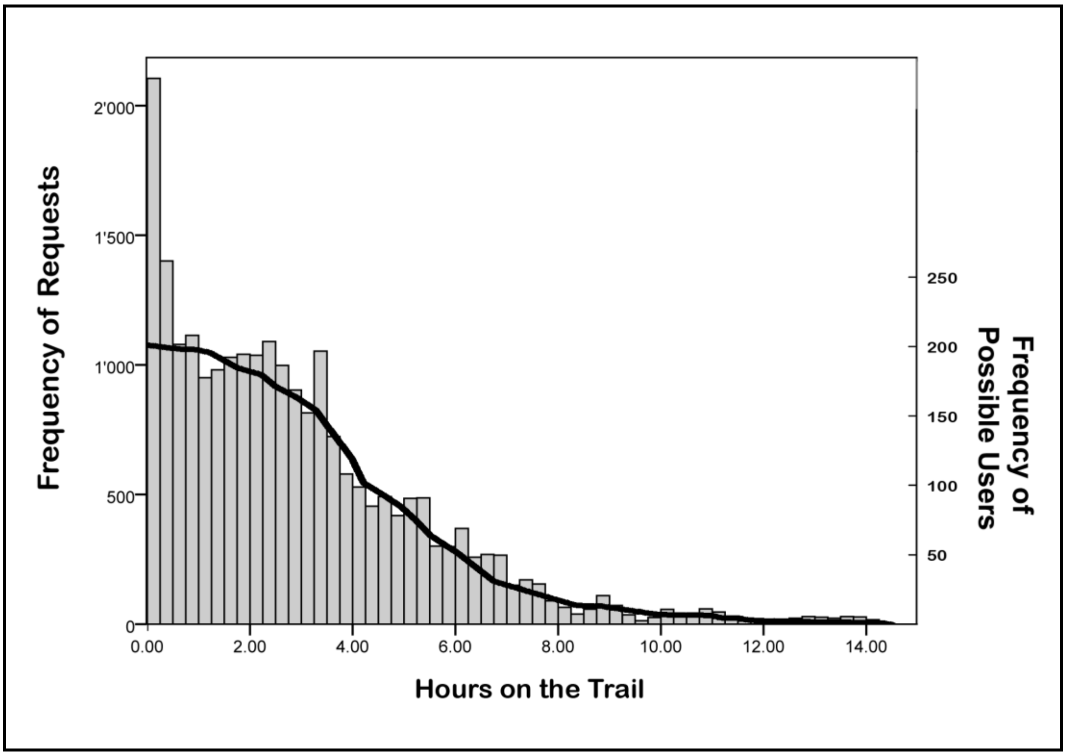

FIGURE 31 : FREQUENCY OF REQUESTS AS HISTOGRAM, AGGREGATION 15 MINUTES, RED CURVE: FREQUENCY OF USERS WHO ARE ON THE TRAILS FOR AT LEAST THAT LONG.

If the hours on the trail are normalized by the total duration of the hike the distribution becomes more meaningful. In Figure 32 the distribution of all points as well as the points from inside the picnic areas are displayed. Some features, such as the peak between 0.5 and 0.6, can be explained by the additional requests inside the picnic areas. In general, the distribution outside the picnic areas is rather homogenous. If a different aggregation scale is used, the values even become more homogenous. While the only outstanding value of outside the picnic areas is at the very beginning of the journey, inside the picnic areas some more features are visible, and the general trend shows that on average the users rest more in the second half of the travelled time.

Comparing these results to those of Eisenhut et al. (2008), a different outcome of the analysis is observed. Eisenhut et al. (2008) used the distance as an indicator of the duration on the trail, which is only to a certain degree plausible. In this study, 201 users on three different trails are analyzed, while Eisenhut et al. used only 30 users on one trail. A clear statement against the hypothesis of Eisenhut et al. (2008) can be made, because there is clearly no sign of decreasing interest in the later phase of a walking-tour. Only at the beginning there are significantly more requests recognizable.

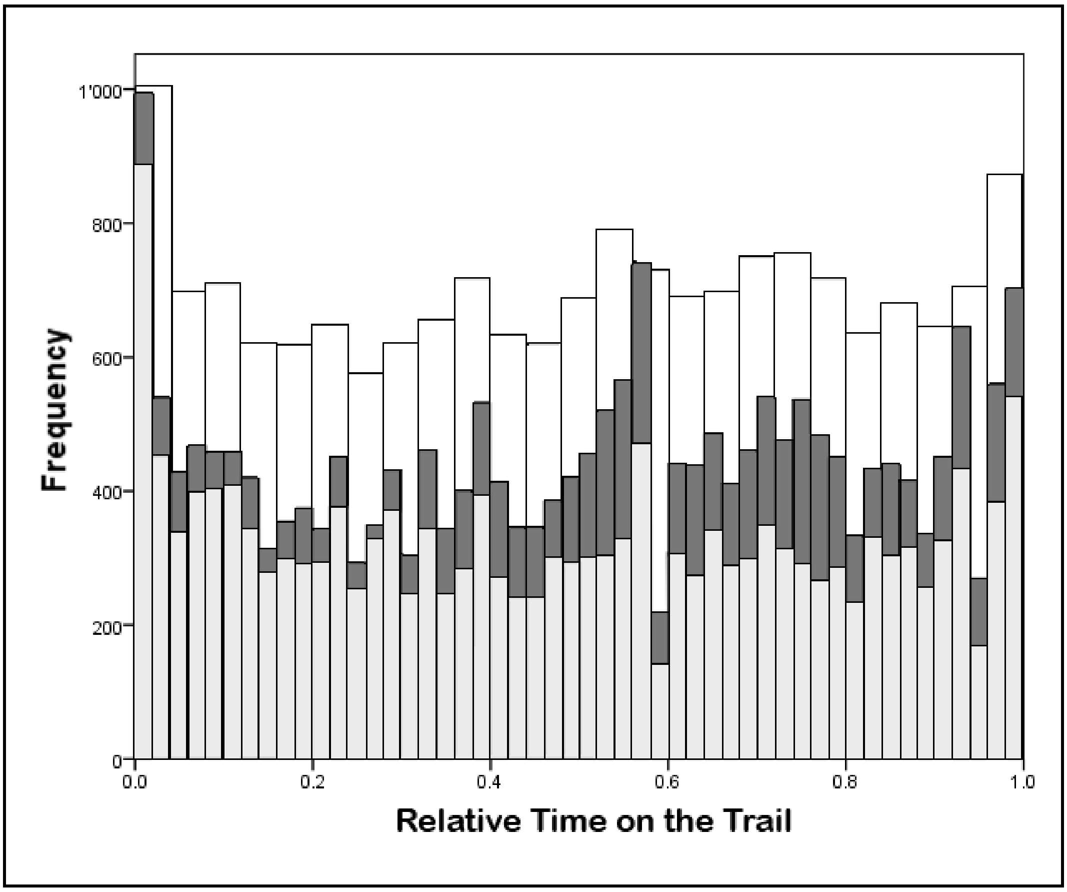

FIGURE 32 : FREQUENCIES OF REQUESTS, X-AXIS IS THE RELATIVE TIME ON THE TRAIL, DARK GREY ARE REQUESTS INSIDE THE PICNIC AREAS, LIGHT GREY ARE OUTSIDE THE PICNIC AREAS. WHITE ARE THE FREQUENCIES IN A DIFFERENT AGGREGATION.

Both the relative, as well as the absolute time have been analyzed in terms of quantity of requests. Therefore, also both will be analyzed separately in terms of quality of information.

5.5.2 WHAT INFORMATION IS NEEDED IN TERMS OF RELATIVE TIME?

We have seen that at the very beginning of a hiking journey more information is requested. These 2% of the relative time will also be regarded separately.
In the first 2% of a hike, the hikers might want to get used to the device. Therefore a number of requests in the group *"info on device"* could be expected. Looking at Figure 32 there is indeed a high interest in this group, but also the *"information on the trails"* is requested more frequently than during the average of the hike, while the *"information around"* shows the exact opposite. It might just not be interesting yet to have additional information about the surrounding area. In the other three groups, the information need is not special compared to the average, when looking at the expected values. But if the percentage of the total requested information of the several groups are concerned, all groups show higher frequencies than the expected 2% of the relative time interval.

TABLE 30 : COMPARISON OF THE FIRST TWO PERCENT OF A HIKE AND THE TOTAL DISTRIBUTION.

		First two percent	Total
Info around	Count	389.00	**9115**
	Expected	514.72	
	Std. Res.	**-5.54**	**4.3**
Info on trail	Count	400.00	**5433**
	Expected	306.80	
	Std. Res.	**5.32**	**7.4**
Content	Count	184.00	**3426**
	Expected	193.46	
	Std. Res.	**-0.68**	**5.4**
Orientation	Count	151.00	**2669**
	Expected	150.72	
	Std. Res.	**0.02**	**5.7**
Special function	Count	65.00	**1338**
	Expected	75.56	
	Std. Res.	**-1.21**	**4.9**
Info on device	Count	109.00	**1005**
	Expected	56.75	
	Std. Res.	**6.94**	**10.9**
Total	**Count**	**1298**	**22986**

The relative time variable is defined between 0 and 1. In order to have reasonable groups concerning Figure 32, five groups with equal interval width are defined, including the first interval from 0 to 0.2. This interval should indicate the information need at the beginning (called "start"). The second interval from 0.2 to 0.4 is an intermediate state (called "interim"), the third and fourth groups are groups of possible and reasonable times for a rest (called "rest1" and "rest2") and the fifth group from 0.8 to 1 is the complementary interval to the start, namely the ending period (called "end").

Looking at the distribution of the groups of information, it looks like values of the first 20 percent of the trail generally follow the distribution of the first two percent that have already been discussed. More information on the device and the trails are important, while less information about the surrounding area and the special functions were requested. The last time period shows an inverse distribution. Certainly the *"information on the trails"* becomes less important because the hikers have already chosen on what trail they want to hike, while they might be more interested in *"special functions"* and toying around before they have to return the device. During the main resting time periods, the distribution is clearly biased by the users need for rest. Therefore, the distribution might be falsified by the high percentage of requests inside picnic areas during that time, and generally shows the same trends.

Every information group has in at least one cell a value of standardized residuals which is bigger than 2. This indicates that every information group is requested significantly more or significantly less at least once during a hike. The highest value is in the information group *"info on trail"* at the start. Especially the *"info around*" and *"info on trail"* function show a high dependency on the relative time. The other information groups generally have low values. The *"orientation"* group has only one standardized residual which is smaller than -2. All other variables are closer to 0. Therefore, it can be assumed that functions with an orientation purpose are not that much time dependent as other functions like *"info around".*

The interpretation seems clear. At the beginning of a hike more information about the trail is requested, because the hikers want to know where they are going or should go. Also information about the device such as tutorials is very important at the beginning, because the users have to get used to the functions. Towards the end of a journey, special functions become more important, because the user had time to get used to the device and can therefore play a bit with other functions beside the basic ones. In between the start and the end no clear trend in any information group is observable. Periods 3 and 4 might be influenced largely by the behavior in the picnic areas, because a high percentage of the requests are performed inside those.

Looking at the subgroups of information an interesting observation can be made. With a level of significance of 0.05 the dependency on the time variables cannot be assumed inside the groups *"information on trail"* and *"information on device".* In both cases the information subgroups seem to be distributed homogenously and therefore they do not need to be analyzed further. This means that all three subgroups of the information group *"info on trail"* have statistically the same distribution as the main group.

TABLE 31 : CROSSTABULATION OF THE INFORMATION GROUPS AGAINST THE RELATIVE TIME.

		Start	Interim	Rest1	Rest2	End	Total
Info around	Count	1642	1374	1912	2074	2113	**9115**
	Expected	1975.59	1550.49	1764.23	1848.30	1976.38	
	Rel. Score	**0.83**	**0.89**	**1.08**	**1.12**	**1.07**	
Info on trail	Count	1453	1069	948	931	1032	**5433**
	Expected	1177.55	924.17	1051.57	1101.68	1178.02	
	Rel. Score	**1.23**	**1.16**	**0.90**	**0.85**	**0.88**	
Content	Count	748	617	735	606	720	**3426**
	Expected	742.55	582.77	663.11	694.71	742.85	
	Rel. Score	**1.01**	**1.06**	**1.11**	**0.87**	**0.97**	
Orientation	Count	620	467	462	578	542	**2669**
	Expected	578.48	454.01	516.59	541.21	578.71	
	Rel. Score	**1.07**	**1.03**	**0.89**	**1.07**	**0.94**	
Special function	Count	246	217	209	309	357	**1338**
	Expected	290.00	227.60	258.97	271.31	290.12	
	Rel. Score	**0.85**	**0.95**	**0.81**	**1.14**	**1.23**	
Info on device	Count	273	166	183	163	220	**1005**
	Expected	217.82	170.95	194.52	203.79	217.91	
	Rel. Score	**1.25**	**0.97**	**0.94**	**0.80**	**1.01**	
Total	**Count**	**4982**	**3910**	**4449**	**4661**	**4984**	**22986**

DIFFERENCES IN THE GROUP "INFO AROUND"

In general, the information group *"info around"* is requested more towards the end of a journey (Table 32). This general trend is based on overlapping trends of the subgroups. Some are additive, others are subtractive. The strongest trend is visible in the subgroup *"get around",* in which most of the information is requested during the main resting periods. Significantly less of this information is requested at the beginning of a journey, as well as in the end. The clear trend that differs between main resting periods and the other three periods might indicate that this variable is depending on another variable. But if the distribution in- and outside the picnic area is regarded, no such trend is visible. Therefore this strong trend might be due to relative time factors. A similar trend, even if it is not as strong as in the subgroup *"get around",* is visible in the subgroup *"FOI list",* where high values are also recognizable during the main resting periods. The opposite is visible looking at the subgroup *"info on FOI",* where high values are at the beginning and at the end of a journey. Also, this distribution does not reflect the distribution of in- and outside the picnic areas, and therefore some other reasons, especially the relative time itself, might cause this distribution. Also the *"info on FOI"* group shows a distribution that might not be expected. During the main resting periods, less of this information was requested, while at the beginning and the end of a journey, more information was requested. Looking at the function *"Map (FOI)",* no such strong trend as in the other subgroups is observable.

High standardized residuals exist at all time periods. Especially in the resting period 2, highly negative values are observable. The biggest amplitude is observable in the function *"get around".* With a value of -8 at the beginning and 8 in the third period a clear trend with a maximum in the middle is observable. In the subgroup *"Map (FOI)"* only one standardized residual at the start stands out. This function was requested much less at the start then during the rest of the journey. Looking at the group *"special function"*, which might also be a good fit for this subgroup, a similar distribution is observable.

Interestingly, major differences are observable between the subgroups. It is interesting that in the main resting periods less information on FOIs was requested. This is the opposite of what could have been expected from the distribution of in- and outside the picnic areas. It seems like this function is, in general, less popular in the middle of a hiking journey. For the other functions, more or less coherent interpretations are possibly. The *"FOI list"* and *"get around"* functions show a general trend towards a higher relevance from the middle until the end of a journey, which can be explained by a possible higher interest of the users in the surrounding area and the better knowledge of the device. On the other hand the distribution of the *"Map (FOI)"* function can be explained by its characteristic similar to a special function. It is less relevant at the beginning, and more relevant at the end of a hike, because the users are more used to the device.

TABLE 32 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO AROUND" AGAINST THE RELATIVE TIME.

		Start	Interim	Rest1	Rest2	End	Total
FOI List	Count	620	504	611	830	789	**3354**
	Expected	726.9	570.5	649.2	680.1	727.2	
	Rel. Score	**0.85**	**0.88**	**0.94**	**1.22**	**1.08**	
Get Around	Count	487	421	840	838	710	**3296**
	Expected	714.4	560.7	637.9	668.3	714.7	
	Rel. Score	**0.68**	**0.75**	**1.32**	**1.25**	**0.99**	
Info on FOI	Count	450	347	326	291	463	**1877**
	Expected	406.8	319.3	363.3	380.6	407	
	Rel. Score	**1.11**	**1.09**	**0.90**	**0.76**	**1.14**	
Map(FOI)	Count	85	102	135	115	151	**588**
	Expected	127.4	100	113.8	119.2	127.5	
	Rel. Score	**0.67**	**1.02**	**1.19**	**0.96**	**1.18**	
Total	**Count**	**1642**	**1374**	**1912**	**2074**	**2113**	**9115**

DIFFERENCES IN THE GROUP "CONTENT"

In general, the information class *"content"* reaches its maximum in the middle of the travelled time (Table 23). Additionally, this group is less relevant towards the end. As recognizable in other variables, the content on plants and vegetation is requested differently than the content on animals. The distribution of the requests on plants and vegetation is homogenous. In the subgroup *"bird"*, especially the values during the main resting periods are interesting. Information about birds was requested more than twice as often during the resting period 1 than during period 2. The difference between the resting period 1 and the resting period 2 is even larger concerning the subgroup *"butterflies"*, and does still exist in the subgroup *"grasshoppers"*. But in the last-mentioned subgroup another difference stands out: the difference between the start and the intermediate period.

In every subgroup of information at least one cell contains a value of a standardized residual which is greater than two, indicating that all subgroups are depending on the relative time. However, all standardized residuals for the flora are comparatively small. Therefore, it can be assumed that the content on *"vegetation"* and *"plants"* is not depending strongly on the relative time: also for the subgroups *"bear"* and *"news"*, generally small standardized residuals are observable, while in the subgroups *"bird"*, *"butterfly"*, and *"grasshopper"* some more pronounced differences are noticeable.

All these differences have to be analyzed with caution, because the number of requests is very small. None of the expected values is smaller than 5, and therefore a Chi^2 test on the independency of the variables is valid, but the differences might just be an effect of other variables, or might even be random. However, some general trends, such as the difference between the animals and the fauna seem more significant.

TABLE 33 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "CONTENT" AGAINST THE RELATIVE TIME.

		Start	Interim	Rest1	Rest2	End	Total
Bear	Count	38	41	31	38	48	**196**
	Expected	42.5	33.3	37.9	39.7	42.5	
	Rel. Score	**0.89**	**1.23**	**0.82**	**0.96**	**1.13**	
Bird	Count	112	100	127	66	133	**538**
	Expected	116.6	91.5	104.1	109.1	116.7	
	Rel. Score	**0.96**	**1.09**	**1.22**	**0.60**	**1.14**	
Butterfly	Count	85	68	169	48	94	**464**
	Expected	100.6	78.9	89.8	94.1	100.6	
	Rel. Score	**0.84**	**0.86**	**1.88**	**0.51**	**0.93**	
Grasshop.	Count	22	40	38	26	42	**168**
	Expected	36.4	28.6	32.5	34.1	36.4	
	Rel. Score	**0.60**	**1.40**	**1.17**	**0.76**	**1.15**	
Plant	Count	124	78	79	105	95	**481**
	Expected	104.3	81.8	93.1	97.5	104.3	
	Rel. Score	**1.19**	**0.95**	**0.85**	**1.08**	**0.91**	
Vegetation	Count	283	223	235	261	243	**1245**
	Expected	269.8	211.8	241	252.5	270	
	Rel. Score	**1.05**	**1.05**	**0.98**	**1.03**	**0.90**	
News	Count	84	67	56	62	65	**334**
	Expected	72.4	56.8	64.6	67.7	72.4	
	Rel. Score	**1.16**	**1.18**	**0.87**	**0.92**	**0.90**	
Total	**Count**	**748**	**617**	**735**	**606**	**720**	**3426**

The information class *“orientation”* shows no clear trend, which is perhaps caused by overlapping subgroups (Table 34). Looking at those, the *“where am I”* function for instance shows a clear pattern. At the beginning of a journey more information is requested, while in the resting phases less information is requested. On the other hand the *“map overlay”* function is used much less at the beginning. The trend for the first two groups is weakened a bit when the subgroup *“map page”* and *“where am I”* get aggregated.

The highest standardized residuals are observable for the *“map overlay”* function. Therefore, it can be assumed that this function is highly dependent on the relative time. On the other hand, both functions *“map overlay”* and *“where am I”* do not reach such high values, but show more values over 2.

Both distributions correspond to the distribution of in- and outside the picnic areas. Therefore, it might just be depending on the resting behavior, more than the relative time that the hikers spend on the trails. The high relevance of the *“where am I”* function at the beginning is explainable through the higher need for a general location information of a user. During the resting phases, especially during the first resting phase, these functions are less important, because the users know their current location.

TABLE 34 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP “ORIENTATION” AGAINST THE RELATIVE TIME.

		Start	Interim	Rest1	Rest2	End	Total
Map Page	Count	373	291	268	399	314	**1645**
	Expected	356.5	279.8	318.4	333.6	356.7	
	Rel. Score	**1.05**	**1.04**	**0.84**	**1.20**	**0.88**	
Where Am I?	Count	212	164	124	130	172	**802**
	Expected	173.8	136.4	155.2	162.6	173.9	
	Rel. Score	**1.22**	**1.20**	**0.80**	**0.80**	**0.99**	
Map Overlay	Count	35	12	70	49	56	**222**
	Expected	48.1	37.8	43	45	48.1	
	Rel. Score	**0.73**	**0.32**	**1.63**	**1.09**	**1.16**	
Total	**Count**	**620**	**467**	**462**	**578**	**542**	2669

"Special functions" are, in general, requested more at the end of a journey (Table 35). This can be explained by the user's intention to toy around with the device and finally use all functions. Looking at the distribution of the subgroups, one value especially stands out. While the *"search"* function is pretty much used uniformly during the whole journey, the *"bookmark"* function is clearly used just towards the end of a journey. The standardized residuals are very high during the last two periods of relative time, while they are low at the beginning. The only outstanding value in the *"search"* subgroup is in the interim phase between the start and a possible first rest. This might be due to the fact that the user first needs to get used to the device, and after that he/she tries to find information that is important to him/her.

The same explanation as for the information group *"special function"* can also be used for the subgroup *"bookmarks".* It seems like the users get more used to the device and can therefore use more demanding functions. They can also "review" the device with its information with *"bookmarks".* Therefore it is logical that the function is used more often towards the end of a hike.

TABLE 35 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "SPECIAL FUNCTION" AGAINST THE RELATIVE TIME.

		Start	Interim	Rest1	Rest2	End	Total
Search	Count	93	100	77	92	98	**460**
	Expected	99.7	78.2	89	93.3	99.7	
	Rel. Score	**0.93**	**1.28**	**0.87**	**0.99**	**0.98**	
Bookmarks	Count	153	117	132	217	259	**878**
	Expected	190.3	149.4	169.9	178	190.4	
	Rel. Score	**0.80**	**0.78**	**0.78**	**1.22**	**1.36**	
Total	**Count**	**246**	**217**	**209**	**309**	**357**	**1338**

5.5.3 WHAT INFORMATION IS NEEDED IN TERMS OF ABSOLUTE TIME?

After analyzing the relative time, also the absolute time must be analyzed. Therefore the time of day must be split into different periods. As seen in chapter 5.5.1, the information need reaches its maximum at 11:20 AM. With respect to that and the outliers, three different classes can be formed: one as a center class around the maximum, one before and one after this class. With respect to the distribution of the requests 3 percentiles have therefore the boundaries 10 AM and 12 AM. The groups are labeled with *"morning", "noon"* and *"afternoon",* even though it might not always be correct to speak of afternoon, because some of the requests were performed at 10 o'clock at night.

At first sight, the distribution of the values of the groups of information seems to be not that distinct (Table 36). But the standardized residuals show that there are indeed differences between the absolute time periods. It seems like the *"info around"* function was used more or less homogenously, while *"information on the trail"* was requested far more often in the morning, which corresponds to the distribution of the values in terms of relative time. At the beginning information on the trails is more important, because the users have to decide where they want to hike and where to start, or in what direction they want to go. *"Content"* information was requested more in the afternoon, which does not corresponds directly to the relative time. Maybe some animals are more active at that time, or the hikers might have a better sight of them in the afternoon. On the other hand information on *"orientation"* is requested less in the afternoon, which would correspond to the explanation of the distribution in the information group "*info on trail". "Special functions"* are again more frequently requested in the afternoon, which again corresponds to the relative time, while requests on *"information on the device"* seem to be more or less distributed homogenously over the time periods. Also, the standardized residuals are not as big as in the other groups. Interestingly, all standardized residuals of the information group *"info on device"* are very small. The highest value, on the other hand, is reached in the afternoon in the group *"special function".* In general, higher values are observable in the afternoon, while the values for noon are, in general, slightly smaller. Therefore, the afternoon can be seen as time of day with the highest fluctuation in terms of relevance, which is interesting because the long time period compared to the two hours at noon might have an obliterating effect on the general trends of the afternoon.

Testing the groups separately on independency of the variables, the group *"info on device"* can, with a level of significance of 0.05, be assumed to be independent on the variables. The variables inside the group "*information on trails"* are equally distributed with the same test and therefore do not have be further analyzed. All other groups of information show a dependency on their subgroups and must therefore be further analyzed.

TABLE 36 : CROSSTABULATION OF THE INFORMATION GROUPS AGAINST THE ABSOLUT TIME.

		Morning	Noon	Afternoon	Total
Info around	Count	2908	2985	3222	**9115**
	Expected	3070.85	2853.54	3190.61	
	Rel. Score	**0.95**	**1.05**	**1.01**	
Info on trail	Count	2078	1581	1774	**5433**
	Expected	1830.38	1700.86	1901.76	
	Rel. Score	**1.14**	**0.93**	**0.93**	
Content	Count	995	1081	1348	**3426**
	Expected	1154.22	1072.54	1199.23	
	Rel. Score	**0.86**	**1.01**	**1.12**	
Orientation	Count	985	937	747	**2669**
	Expected	899.19	835.56	934.25	
	Rel. Score	**1.10**	**1.12**	**0.80**	
Special function	Count	405	317	616	**1338**
	Expected	450.77	418.87	468.35	
	Rel. Score	**0.90**	**0.76**	**1.32**	
Info on device	Count	371	295	339	**1005**
	Expected	338.59	314.63	351.79	
	Rel. Score	**1.10**	**0.94**	**0.96**	
Total	**Count**	**7744**	**7196**	**8046**	**22986**

No strong trend is observable for the group *"info around".* Also, the standardized residuals are comparatively small. Looking at Table 37, the subgroups *"FOI List"* and *"get around"* show a similar distribution. The highest request rate is at noon in both cases, while in the beginning significantly less information about the surrounding area is requested. The function *"map (FOI)"* is, with respect to the standardized residuals, relatively homogenously distributed. Also the *"information on FOI"* seems to be homogenously distributed. More information on the FOI was requested in the morning, and far less is requested at noon. Some similarities to the distribution of the relative time are recognizable. But the differences are smaller, and no general trends can be observed.

In general, the standardized residuals are small. The only high values are reached for the subgroup *"FOI list"* with a maximum of 4.7. All values for the subgroup *"Map (FOI)"* are much smaller than 2 and therefore it can be assumed that this subgroup is not depending on the absolute time of day.

Due to the generally small standardized residuals it can be said that these subgroups are not highly dependent on the absolute time. A possible interpretation is similar to the interpretation of the relative time. The *"FOI list",* the *"Map (FOI)"* function as well as the *"get around"* function are more requested in the middle time periods of a journey, which is also manifested at noon.

TABLE 37 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO AROUND" AGAINST THE ABSOLUTE TIME.

		Morning	Noon	Afternoon	Total
FOI List	Count	1004	1202	1148	**3354**
	Expected	1129.97	1050.00	1174.03	
	Rel. Score	**0.89**	**1.14**	**0.98**	
Get Around	Count	1035	1060	1201	**3296**
	Expected	1110.42	1031.85	1153.73	
	Rel. Score	**0.93**	**1.03**	**1.04**	
Info on FOI	Count	696	530	651	**1877**
	Expected	632.36	587.61	657.02	
	Rel. Score	**1.10**	**0.90**	**0.99**	
Map(FOI)	Count	173	193	222	**588**
	Expected	198.10	184.08	205.82	
	Rel. Score	**0.87**	**1.05**	**1.08**	
Total	**Count**	**2908**	**2985**	**3222**	**9115**

The general trend of the group *"content"* shows more requests in the afternoon than in the morning (Table 38). This trend is also observable in the subgroup *"bear", "bird", "butterfly"* and *"grasshopper".* The fauna is, once again, distributed differently. Both subgroups *"vegetation"* and *"plants"* show high values at noon. The *"news"* function, on the other hand, is requested more frequently in the morning.

But if we look at the standardized residuals they show a different image. The fauna, the *"news"* function, and the subgroup *"bear"* have no standardized residuals of 2 or greater. Therefore, an independency of these subgroups on the absolute time is not impossible. The highest value of -4.5 is observable for the subgroup "*bird"* in the morning.

Possible interpretations are that the *"bear"* is most likely not visible at anytime and therefore no dependency on the absolute time can be assumed. The plants might be influenced by the sun, but do not change their location. Therefore, they are visible all day long. Interestingly, the *"news"* function shows a similar distribution as in the relative time steps. It is requested more frequently at the start and in the morning. The distribution of the requests on *"birds", "butterflies",* and *"grasshoppers"* might be influenced by their general appearance during the day. They might all be more active in the afternoon and therefore seen more often by the hikers.

TABLE 38 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "CONTENT" AGAINST THE ABSOLUTE TIME.

		Morning	Noon	Afternoon	Total
Bear	Count	61	53	82	**196**
	Expected	66.03	61.36	68.61	
	Rel. Score	**0.92**	**0.86**	**1.20**	
Bird	Count	121	187	230	**538**
	Expected	181.25	168.43	188.32	
	Rel. Score	**0.67**	**1.11**	**1.22**	
Butterfly	Count	127	126	211	**464**
	Expected	156.32	145.26	162.42	
	Rel. Score	**0.81**	**0.87**	**1.30**	
Grasshopper	Count	35	40	93	**168**
	Expected	56.60	52.59	58.81	
	Rel. Score	**0.62**	**0.76**	**1.58**	
Plant	Count	154	158	169	**481**
	Expected	162.05	150.58	168.37	
	Rel. Score	**0.95**	**1.05**	**1.00**	
Vegetation	Count	378	408	459	**1245**
	Expected	419.44	389.76	435.80	
	Rel. Score	**0.90**	**1.05**	**1.05**	
News	Count	121	109	104	**334**
	Expected	112.52	104.56	116.91	
	Rel. Score	**1.08**	**1.04**	**0.89**	
Total	**Count**	**995**	**1081**	**1348**	**3426**

The general trend in the information group *"orientation"* shows a decreasing relevance during a day (Table 39), which can be explained by the additional need of the user for orientation at the beginning of a journey, because they have to know where to go etc. Looking at the detailed distribution of the subgroups, especially in the subgroup *"map page"* this trend is indeed observable, while the *"map overlay"* function was used totally differently. This function is requested especially often at noon.

In all subgroups at least one standardized residual is greater than 2. Therefore, all information subgroups are depending on the time of day. But there are differences in the magnitude. For the *"map page"*, especially the morning and the afternoon are outstanding. There is a clear decreasing trend towards the afternoon. On the other hand, the *"where am I"* function has comparatively small residuals. Only the afternoon value is outstanding. For the subgroup *"map overlay"* all values are comparatively large. The highest value is reached at noon, which is also the global maximum of this information group. Therefore, it can be assumed that the *"map overlay"* function is mostly depending on absolute time.

A higher relevance is observable for orientation functions before 10 AM. This can be explained by the relative time as well as some visibility issues. Most of the users start their hike in the morning, where it might be darker and therefore a map might be more needed.

TABLE 39 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "ORIENTATION" AGAINST THE ABSOLUTE TIME.

		Morning	Noon	Afternoon	Total
Map Page	Count	655	524	466	**1645**
	Expected	554.20	514.98	575.81	
	Rel. Score	**1.18**	**1.02**	**0.81**	
Where am I	Count	287	286	229	**802**
	Expected	270.19	251.07	280.73	
	Rel. Score	**1.06**	**1.14**	**0.82**	
Map Overlay	Count	43	127	52	**222**
	Expected	74.79	69.50	77.71	
	Rel. Score	**0.57**	**1.83**	**0.67**	
Total	**Count**	**985**	**937**	**747**	**2669**

Differences in the group "Special Function"

In general, there is a trend for more requests of *"special functions"* in the afternoon. Concerning Table 40, this trend is dominated by the *"bookmark"* function. The subgroup *"search"* shows no high standardized residuals. Therefore, it can be assumed that during the whole day the *"search"* function is equally important. On the other hand the hikers requested the *"bookmark"* function much more often in the afternoon, which corresponds to the distribution of the relative time. Bookmarks are important after some other information was requested, which finally leads to the use of bookmarks. The general trend of *"special functions"*, also concerning the *"map overlay"* function, cannot be explained by the time of day. All three subgroups show a different distribution. Therefore, the relative time is a better indicator for the distribution of these requests during a journey.

TABLE 40 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "SPECIAL FUNCTION" AGAINSTE THE ABSOLUTE TIME.

		Morning	Noon	Afternoon	Total
Search	Count	151	159	150	**460**
	Expected	154.97	144.01	161.02	
	Rel. Score	**0.97**	**1.10**	**0.93**	
Bookmarks	Count	254	158	466	**878**
	Expected	295.80	274.87	307.33	
	Rel. Score	**0.86**	**0.57**	**1.52**	
Total	**Count**	**405**	**317**	**616**	**1338**

5.6 TOPOGRAPHY

5.6.1 HOW MUCH INFORMATION IS NEEDED?

The expected values of requests can be calculated with the distribution of slope on the trails as visible in Figure 33. The slope values in Figure 33 are ranging from 0 to 31° with a mean of 7.2° and quartile breaks at 2.80, 5.87 and 10.84°, which will be used later. Additionally, it can be seen that the trails are clearly located in lower slopes than the general area. They might be built perpendicular to the steepest slope i.e. following elevation contours and also might be positioned in lower slopes.

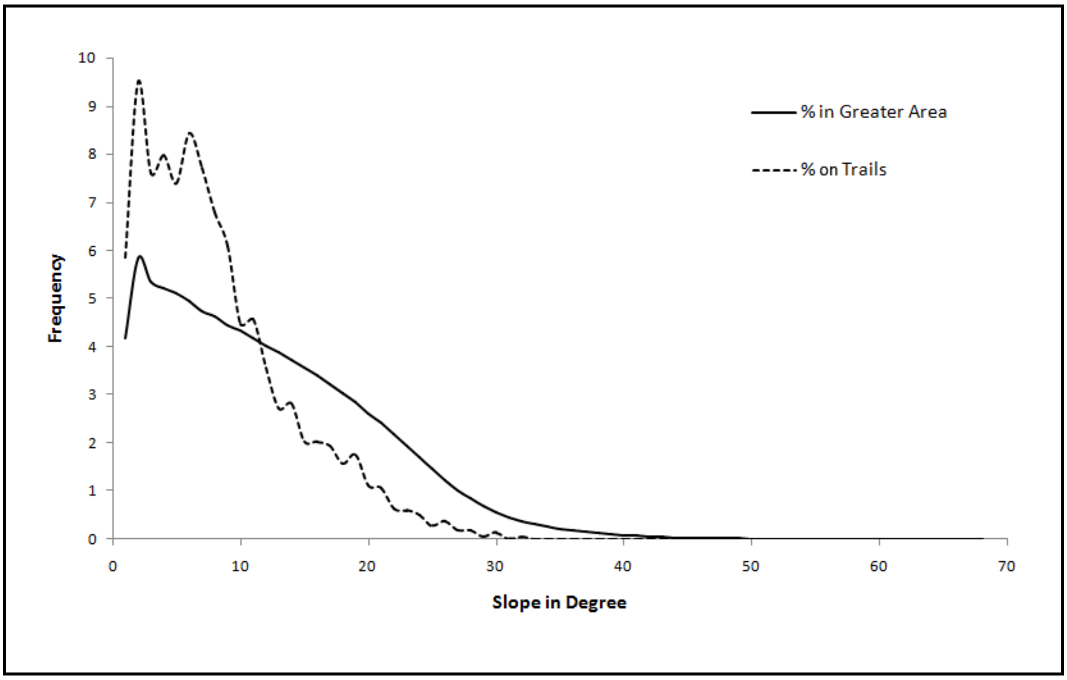

FIGURE 33 : DISTRIBUTION OF SLOPE ON THE TRAILS (DASHED LINE) AND IN THE REGION (SOLID LINE).

With the distribution of the slope values on the trails standardized residuals from the observed values can be calculated. These standardized residuals are visualized in Figure 34. Additionally, the percentages of requests in- and outside the picnic areas are represented in grey and white. It can be seen that most of the standardized residuals are negative. In only seven slopes the standardized residuals are positive, meaning that more requests were performed than expected. In four of these seven cases the frequencies might be strongly influenced by the additional requests inside the picnic areas, because the percentage of requests that were performed inside the picnic areas is high.

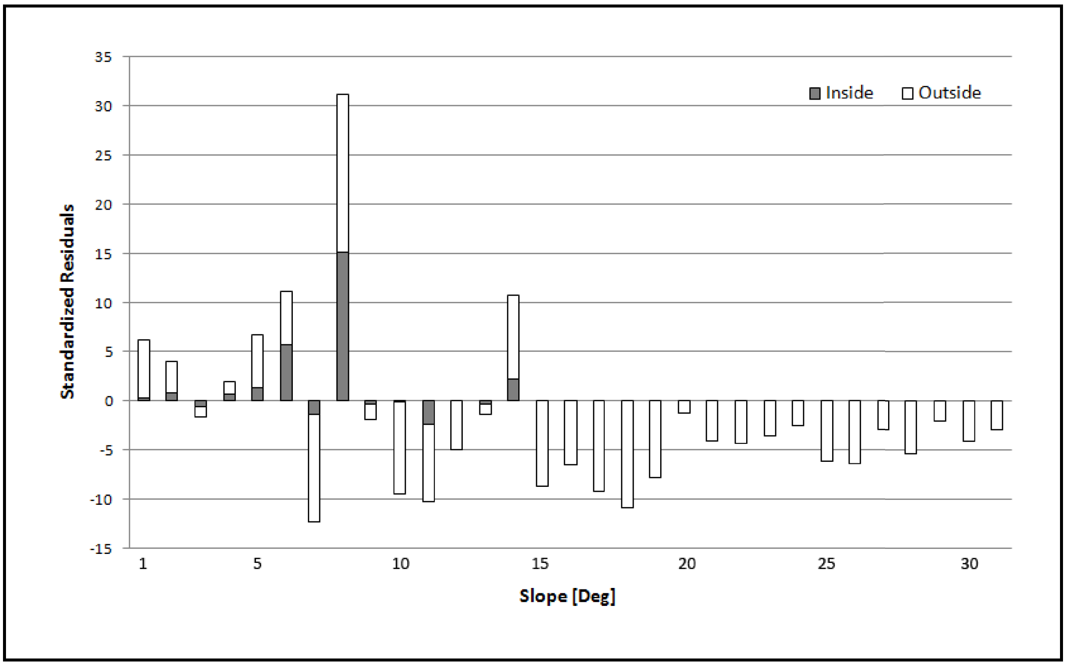

FIGURE 34 : STANDARDIZED RESIDUALS DEPENDING ON THE SLOPE. THE GREY AND WHITE BARS REPRESENT THE PERCENTAGE OF THE REQUESTS INSIDE AND OUTSIDE THE PICNIC AREAS.

In chapter 5.2 it could be shown that inside the picnic areas almost 2.5 times more requests are performed than outside. This effect can be corrected to receive purer information about the distribution of the requests depending on the slope. In Figure 35 the corrected standardized residuals (white bars) as well as the correction (grey bars) is visualized.

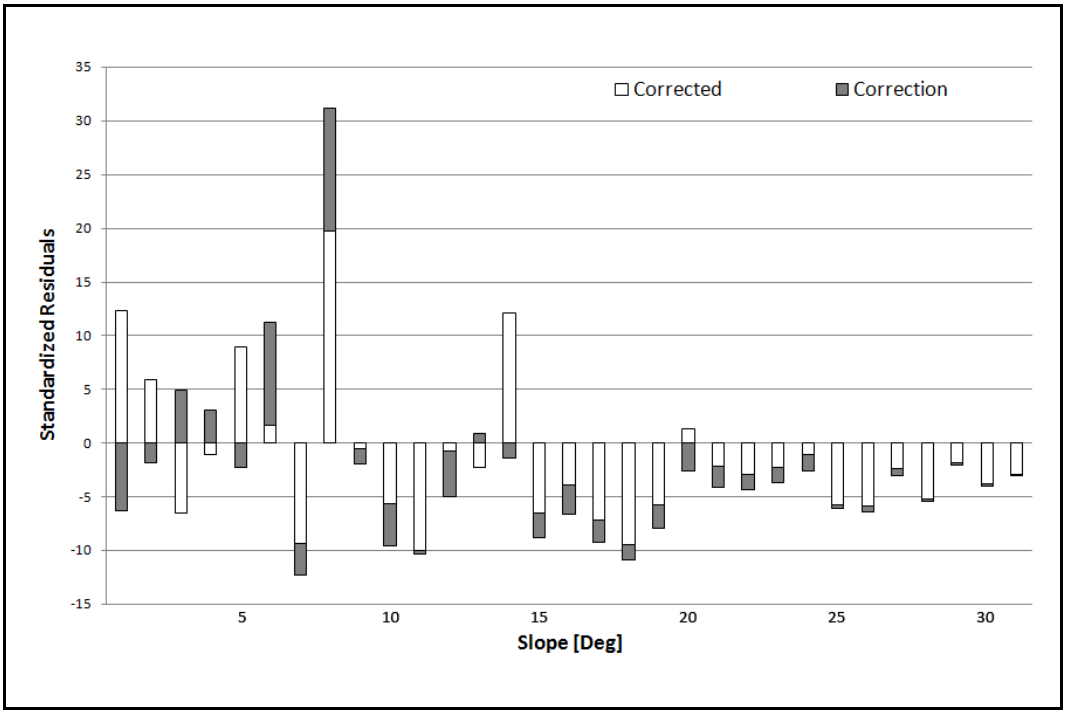

FIGURE 35 : CORRECTION OF THE VALUES WITH THE DISTRIBUTION OF THE PICNIC AREAS.

In the majority of the cases the absolute value of the standardized residuals isreduced by the correction of the picnic areas. Especially the value at 8° and 14° becomes reduced. The general reduction of residuals is about 18%.
Looking at the quantile classification in Figure 36 with class boundaries at 2.80, 5.87 and 10.84°, it becomes clear that the general trend points towards lower information need in steeper slope, which is greater than 10°, while there is an almost constant higher need in minor slopes with a smaller angle than 10°.

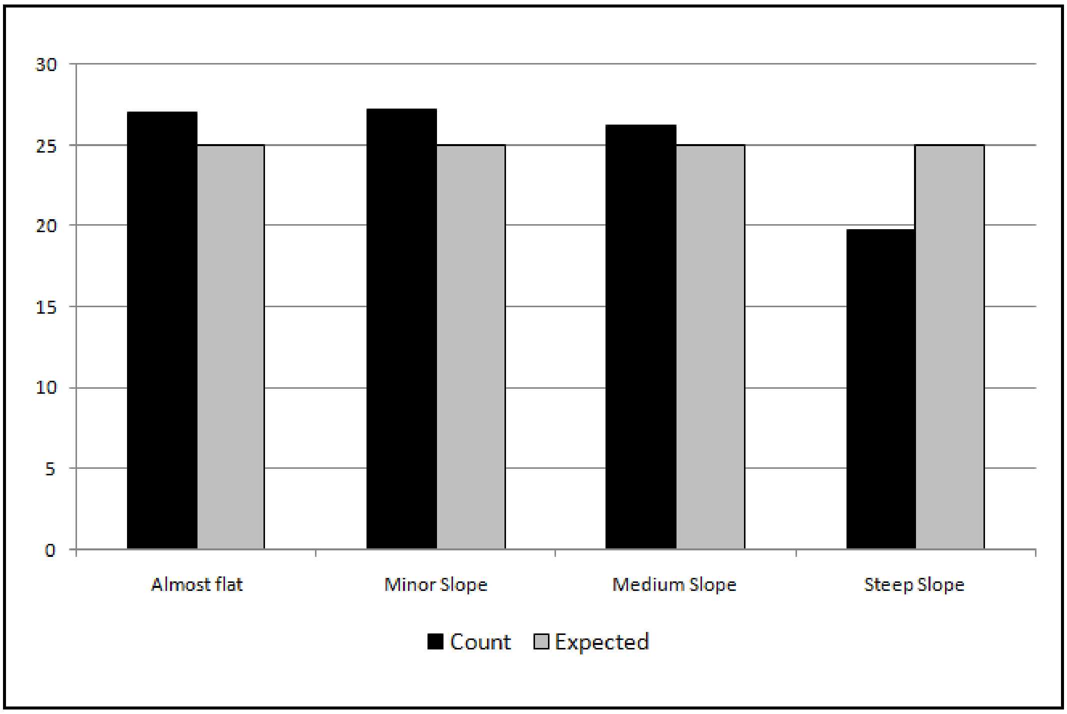

FIGURE 36 : EXPECTED AND COUNTED REQUESTS IN A QUANTILE CLASSIFICATION.

Comparing the results with those of Eisenhut et al. (2008) and their classes, a different result can be observed. In the recent analysis a general difference of the trend is obvious. Eisenhut et al. (2008) presented that in moderate slope the need for information is lower, while it increases in steeper slopes. In Figure 37 it can be seen that this trend cannot be seen in the recent analysis. Rather, an opposite trend is visible, because relatively speaking most requests were made in moderate slopes.

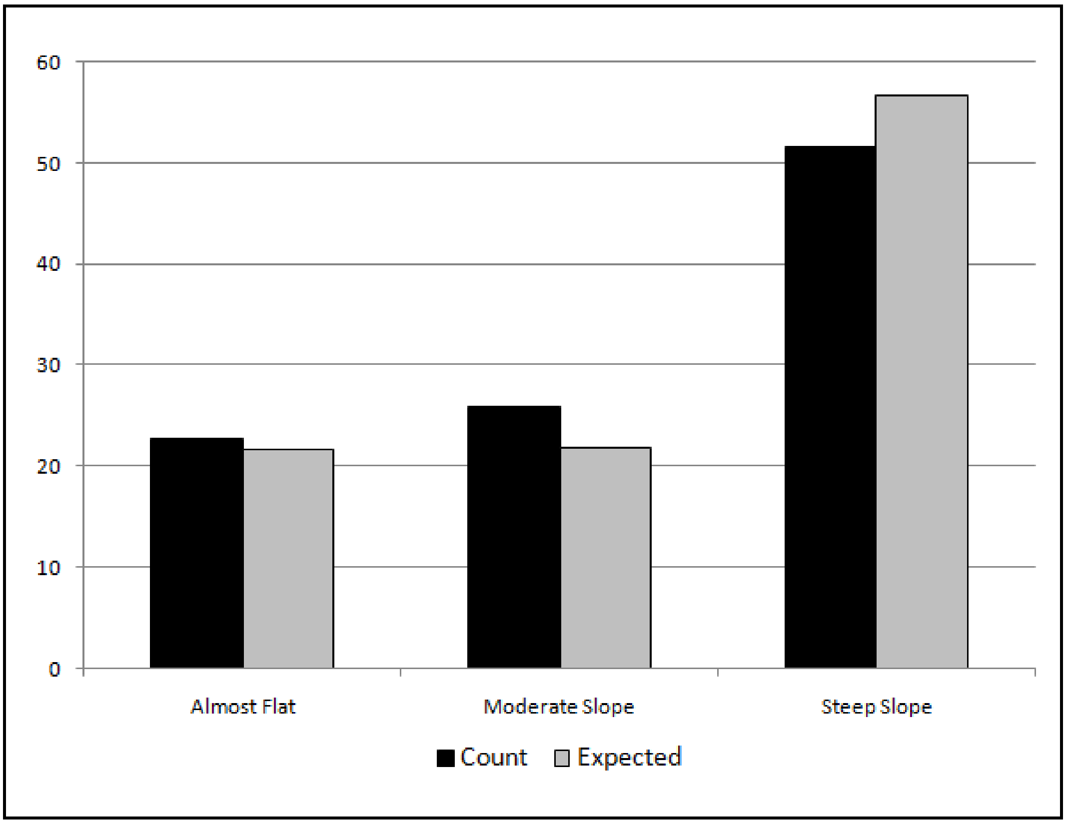

FIGURE 37 : EXPECTED AND OBSERVED VALUES WITH THE SLOPE CLASSES OF EISENHUT ET AL. (2008). BREAKS AT 2.22° AND 5.12°.

Summing up the answers to the question of *"How much information is neededrelative to slope*, it can be stated that only one general trend is observable: in slopes greater than 8° information becomes less important, while the distribution in slopes smaller than 8° shows an inconsistent form, which might be influenced by other factors.

Looking now at the request frequencies in different orientations, some other general trends can be stated. The standardized residuals of all orientations are visualized in Figure 38. The aggregation level is 5° and the expected values were calculated with the distribution of the aspect on the trails. In the north to north-east orientation of the diagram almost no positive standardized residuals (dark grey) are observable, while the majority of the values in the south to south-east part are positive (light grey). The maximum is reached at 100° azimuth with more than 6 times more requests than expected.

Before and after correction similar to the one presented for slope, the general trends of the relative needs for information remain the same. More information is requested in south-easterly aspects than in northern aspectswhich is consistent with the findings of Eisenhut et al. (2008).

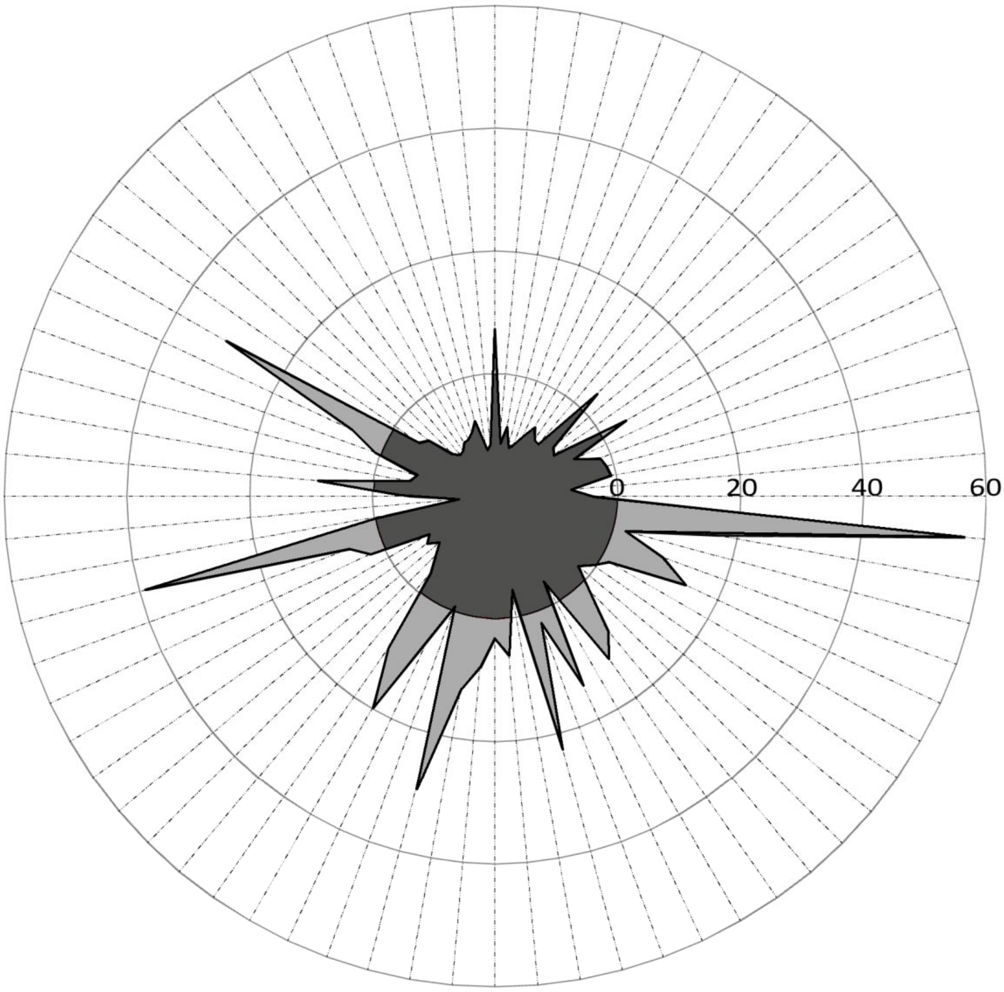

FIGURE 38 : STANDARDIZED RESIDUALS DEPENDING ON ASPECT. DARK GREY SPIKES ARE NEGATIVE RESIDUALS, LIGHT GREY SPIKES ARE POSITIVE RESIDUALS. MAJOR GRIDLINES ARE RANGING FROM -20 TO 60 STANDARDIZED RESIDUALS.

5.6.2 WHAT INFORMATION IS REQUESTED, DEPENDING ON THE SLOPE?

Slope can be classified in different ways. In Figure 33 on page 105 it can be seen that it is difficult to find natural breaks to classify the data. Therefore, quantiles are applied with the class breaks at 2.80, 5.87 and 10.84°.

Looking deeper into the distribution of the index values, it seems as if the *"information around"* function was requested more often in almost flat regions and less in minor to medium slopes. The opposite is observable in the information group *"content",* where more information was requested in medium slope and far less in almost flat regions and for the information group on *"orientation",* where no clear pattern can be found. However, it seems that in steep slopes that information becomes more important, while *"special functions"* become less important. This can be explained by a proposed focuson basic needs such as *"orientation"* and less on more demanding applications such as *"special functions".*

The first analysis of the information group in Table 41 shows that there are surprisingly few differences between the slopes. Especially the group *"information on trail"* shows no difference at all between the slopes. All standardized residuals but one are smaller than 1, which clearly indicates that there is no dependency on the variables. Also the groups *"special function"* and *"info on device"* have almost no big standardized residuals, while the group *"content", "info around"* and *"orientation"* have at least two values that are bigger than 3.

While the Chi^2 test on independency on all groups can be rejected with a level of significance of 0.05, some of the distributions within these groups show different results. Within the groups *"information on trail", "orientation",* and *"information on device"* an independency on the variables cannot be rejected and therefore these groups do not need to be analyzed further.

TABLE 41 : CROSSTABULATION OF THE INFORMATION GROUPS AGAINST SLOPE.

		Flat	Minor	Medium	Steep	Total
Info around	Count	2675	2322	2314	1804	**9115**
	Expected	2455.81	2476.82	2381.26	1801.11	
	Rel. Score	**1.09**	**0.94**	**0.97**	**1.00**	
Info on trail	Count	1431	1468	1425	1109	**5433**
	Expected	1463.79	1476.31	1419.35	1073.55	
	Rel. Score	**0.98**	**0.99**	**1.00**	**1.03**	
Content	Count	733	1061	996	636	**3426**
	Expected	923.05	930.95	895.03	676.97	
	Rel. Score	**0.79**	**1.14**	**1.11**	**0.94**	
Orientation	Count	682	744	617	626	**2669**
	Expected	719.09	725.25	697.27	527.39	
	Rel. Score	**0.95**	**1.03**	**0.88**	**1.19**	
Special function	Count	397	376	373	192	**1338**
	Expected	360.49	363.58	349.55	264.39	
	Rel. Score	**1.10**	**1.03**	**1.07**	**0.73**	
Info on device	Count	275	275	280	175	**1005**
	Expected	270.77	273.09	262.55	198.59	
	Rel. Score	**1.02**	**1.01**	**1.07**	**0.88**	
Total	**Count**	**6193**	**6246**	**6005**	**4542**	**22986**

DIFFERENCES IN THE GROUP "INFO AROUND"

Generally, more requests on the group *"information around"* were made in flat areas, while in medium slope less information was requested (Table 42). Looking at the subgroups, some general trends are observable. Just like the general trend of the group the subgroups *"get around"* and *"Map (FOI)"* were requested more often in flat areas, while the function *"FOI list"* becomes more important the steeper the slope is. Another general trend is that *"information on FOI"* is less important in steeper slopes and more important in almost flat areas.

In general, it can be stated that the biggest differences in terms of standardized residuals are in the *"get around"* subgroup, and the average residuals are not as big as when the information depend on other variables. But the interpretation of this feature is not easy. It might just be an artifact of other variables. On the other hand, the request of explicit information and more demanding functions such as *"Map (FOI)"* are less popular in steeper slope, because hikers might have less energy to put into the handling of the device.

TABLE 42 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO AROUND" AGAINST SLOPE.

		Flat	Minor	Medium	Steep	Total
FOI List	Count	856	875	875	748	**3354**
	Expected	903.65	911.38	876.22	662.75	
	Rel. Score	**0.95**	**0.96**	**1.00**	**1.13**	
Get Around	Count	1098	704	820	674	**3296**
	Expected	888.02	895.62	861.07	651.28	
	Rel. Score	**1.24**	**0.79**	**0.95**	**1.03**	
Info on FOI	Count	517	594	476	290	**1877**
	Expected	505.71	510.04	490.36	370.89	
	Rel. Score	**1.02**	**1.16**	**0.97**	**0.78**	
Map(FOI)	Count	204	149	143	92	**588**
	Expected	158.42	159.78	153.61	116.19	
	Rel. Score	**1.29**	**0.93**	**0.93**	**0.79**	
Total	**Count**	**2675**	**2322**	**2314**	**1804**	**9115**

DIFFERENCES IN THE GROUP "CONTENT"

The general distribution shows that there is less interest in this kind of information in flat areas, and more in minor slopes. Looking at the data, only in the subgroups *"bird"*, *"grasshopper"*, and *"vegetation"* residuals greater than 2 are observable. Requests on *"birds"* and *"grasshoppers"* were therefore performed much more often in minor slopes, while requests on the *"vegetation"* were more numerous in medium slopes. All other subgroups remain more or less constant in different slopes.
A possible explanation might be the different habitats of the plants and animals. But because the expected values are very small the distribution might just be depending on other variables.

TABLE 43 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "CONTENT" AGAINST SLOPE.

		Flat	Minor	Medium	Steep	Total
Bear	Count	48	60	52	36	**196**
	Expected	52.81	53.26	51.20	38.73	
	Rel. Score	**0.91**	**1.13**	**1.02**	**0.93**	
Bird	Count	89	238	130	81	**538**
	Expected	144.95	146.19	140.55	106.31	
	Rel. Score	**0.61**	**1.63**	**0.92**	**0.76**	
Butterfly	Count	139	111	136	78	**464**
	Expected	125.01	126.08	121.22	91.69	
	Rel. Score	**1.11**	**0.88**	**1.12**	**0.85**	
Grasshop.	Count	20	71	49	28	**168**
	Expected	45.26	45.65	43.89	33.20	
	Rel. Score	**0.44**	**1.56**	**1.12**	**0.84**	
Plant	Count	114	139	140	88	**481**
	Expected	129.59	130.70	125.66	95.04	
	Rel. Score	**0.88**	**1.06**	**1.11**	**0.93**	
Vegetation	Count	235	331	411	268	**1245**
	Expected	335.43	338.30	325.25	246.01	
	Rel. Score	**0.70**	**0.98**	**1.26**	**1.09**	
News	Count	88	111	78	57	**334**
	Expected	89.99	90.76	87.26	66.00	
	Rel. Score	**0.98**	**1.22**	**0.89**	**0.86**	
Total	**Count**	**733**	**1061**	**996**	**636**	**3426**

Differences in the group "Special Function"

The general trend of less need of *"special functions"* is clearly dominated by the distribution in the sub group *"bookmarks"* (Table 44). The subgroup *"search"* also shows a slight trend in the same direction, but all standardized residuals are too small to make a statistically firm statement. In the subgroup *"bookmarks"* major differences between steep slopes and flat areas are observable, corresponding to the *"Map (FOI)"* function in the group *"info around".*

It seems as if there is a lower need for demanding operations like making and requesting bookmarks in steep slopes, which again can be explained by the increased need on needed physical capacities of the hikers in steeper slopes.

TABLE 44 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "SPECIAL FUNCTION" AGAINST SLOPE.

		Flat	Minor	Medium	Steep	Total
Search	Count	120	149	115	76	**460**
	Expected	123.94	125.00	120.17	90.90	
	Rel. Score	**0.97**	**1.19**	**0.96**	**0.84**	
Bookmarks	Count	277	227	258	116	**878**
	Expected	236.56	238.58	229.37	173.49	
	Rel. Score	**1.17**	**0.95**	**1.12**	**0.67**	
Total	**Count**	**397**	**376**	**373**	**192**	**1338**

5.6.3 WHAT INFORMATION IS REQUESTED, DEPENDING ON THE ASPECT?

In order to be able to perform a statistically firm analysis, only the cardinal directions are regarded, leading to expected values that are large enough.
Looking at Table 45 generally four different cases are observable. While the information group *"content"* shows no differences between the cardinal directions all other information groups at least differ in two of the four orientations. The information groups *"information on trail"* and *"orientation"* have the same distribution with an observable difference between the eastern and southern sector, while the information groups *"special function"* and *"info on device"* clearly have the biggest difference between the northern and southern sector. Looking at the *"information around"* group, both trends are combined in inverse order. It seems that the *"information around"* function was more requested on the north-south axis than on the west-east axis. Both information groups *"information on trail"* and *"orientation"* are less important in the southern sector, while it seems that they are more important in the eastern sector. The information groups *"special functions"* as well as the *"information on device"* were requested more often in the western sector and less often in the northern sector, which is quite the opposite of the groups that have been mentioned before.
Even though some of the trends seem to be clear, no satisfactory interpretation of this distribution can be offered. Looking at the distribution within the information groups, the only group for which an independency cannot be discarded is the group *"info on device".* The Chi^2 test showed in this case that the variables are independent. In all other cases the test value is smaller than the level of significance of 0.05 and a dependency can be assumed.

TABLE 45 : CROSSTABULATION OF THE INFORMATION GROUPS AGAINST ASPECT.

		North	East	South	West	Total
Info around	Count	1252	1812	3440	2611	**9115**
	Expected	1171.40	2201.62	2952.28	2789.70	
	Rel. Score	**1.07**	**0.82**	**1.17**	**0.94**	
Info on trail	Count	710	1539	1440	1744	**5433**
	Expected	698.21	1312.28	1759.71	1662.80	
	Rel. Score	**1.02**	**1.17**	**0.82**	**1.05**	
Content	Count	454	855	1115	1002	**3426**
	Expected	440.29	827.51	1109.66	1048.55	
	Rel. Score	**1.03**	**1.03**	**1.00**	**0.96**	
Orientation	Count	326	835	710	798	**2669**
	Expected	343.00	644.67	864.47	816.86	
	Rel. Score	**0.95**	**1.30**	**0.82**	**0.98**	
Special function	Count	111	289	418	520	**1338**
	Expected	171.95	323.18	433.37	409.50	
	Rel. Score	**0.65**	**0.89**	**0.96**	**1.27**	
Info on device	Count	101	222	322	360	**1005**
	Expected	129.16	242.75	325.51	307.59	
	Rel. Score	**0.78**	**0.91**	**0.99**	**1.17**	
Total	**Count**	**2954**	**5552**	**7445**	**7035**	**22986**

As discussed before, the general distribution depending on the orientation shows an axis pattern, which differs between the east-west and the north-south axis. The highest value in terms of index as well as standardized residual can be found in the information subgroup *"get around"* in the northern sector. This subgroup is 1.44 times more important than on average. On the other hand the *"Map (FOI)"* function is much less important in the northern sector, whereas it is much more important in the southern sector. The lower relevance in the northern sector correlates with distribution of the *"special function"* information class.

Almost all cells have a higher standardized residual than 2, indicating that there is indeed a relation between these information subgroups and the cardinal directions. The lowest values are in the subgroup *"FOI list".* In this group only the eastern and southern values show a marked difference from the average. In the group *"get around"* in the northern sector the highest value is reached.

No clear interpretation for the presented distribution can be offered. Therefore, the values might just be random effect or have a spatial correlation with other factors.

TABLE 46 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO AROUND" AGAINST ASPECT.

		North	East	South	West	Total
FOI List	Count	436	710	1231	977	**3354**
	Expected	431.03	810.12	1086.34	1026.51	
	Rel. Score	**1.01**	**0.88**	**1.13**	**0.95**	
Get Around	Count	611	601	1248	836	**3296**
	Expected	423.58	796.11	1067.55	1008.76	
	Rel. Score	**1.44**	**0.75**	**1.17**	**0.83**	
Info on FOI	Count	176	392	711	598	**1877**
	Expected	241.22	453.37	607.95	574.47	
	Rel. Score	**0.73**	**0.86**	**1.17**	**1.04**	
Map(FOI)	Count	29	109	250	200	**588**
	Expected	75.57	142.02	190.45	179.96	
	Rel. Score	**0.38**	**0.77**	**1.31**	**1.11**	
Total	**Count**	**1252**	**1812**	**3440**	**2611**	**9115**

DIFFERENCES IN THE GROUP "INFO ON TRAIL"

The general trend of the results of the information group *"info on trail"* is that this kind of information is more important in theeastern sector, while being less important in the southern sector. Looking at the subgroups of the group *"info on trail"* it becomes clear that the general trend is strongly influenced by the distribution of the *"route info"* as well as the route *"vertical profile".* Especially the *"vertical profile"* shows a strong variation as described above. The *"virtual trail"* function shows a complementary distribution, because it is less important in north-eastern aspects. However, because of the low total count of requests of the *"virtual trail"* function this complementary trend does not have a big influence on the total distribution of this group.

If the standardized residuals are analyzed more closely an interesting aspect can be observed. All standardized residuals that are greater than 2 are at eastern and southern aspects, while all northern and western aspects show small standardized residuals. As could already be shown, the information on the trails is more important on the trail Chamanna Cluozza & Murter, on which the north-eastern aspect is most common. Therefore this might just be an effect of the aspect variations of the trails.

TABLE 47 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO ON TRAIL" AND THE ASPECT.

		North	East	South	West	Total
Route Info	Count	486	978	1007	1175	**3646**
	Expected	468.56	880.65	1180.91	1115.88	
	Rel. Score	**1.04**	**1.11**	**0.85**	**1.05**	
Vertical Profile	Count	200	515	324	472	**1511**
	Expected	194.18	364.96	489.40	462.45	
	Rel. Score	**1.03**	**1.41**	**0.66**	**1.02**	
Virtual Trails	Count	24	46	109	97	**276**
	Expected	35.47	66.66	89.39	84.47	
	Rel. Score	**0.68**	**0.69**	**1.22**	**1.15**	
Total	**Count**	**710**	**1539**	**1440**	**1744**	**5433**

No clear trend is observable in the information group *"content".* But if we look at the distribution inside this group, major differences occur between subgroups. Generally, all trends cancel each other out, and therefore no trend in the superior group can be observed.
The first observation that can be made is that in the northern sector great differences can be observed between subgroups. Information on *"butterflies"* and *"plants"* are requested more frequently, while almost all other groups show a lower request rate. Big differences are also visible in the subgroups *"birds"* and *"butterflies".* Both show a high dependency on aspect. Also, the *"vegetation"* and *"plants"* show differences between northern and eastern sectors.
Only for the requests on *"birds"* and *"butterflies"* the standardized residuals reach amplitudes where it can be clearly stated that there is a dependency. In all other cases the values are, in general, too low. For all other subgroups only one or two values out of four possible reach a level which is barely above the limit of 2. With respect to random effects it cannot be said that these subgroups depend on aspect.
The higher dependency of *"birds"* and *"butterflies"* on the exposition might be explained by a higher sensitivity than other animals on thermal column, which is depending on the exposition. And also the plants could be sensitive to the exposition, while the *"news"* subgroup clearly shows no direct connection to the exposition.

TABLE 48 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "CONTENT" AGAINST ASPECT.

		North	East	South	West	Total
Bear	Count	14	54	62	66	**196**
	Expected	25.19	47.34	63.48	59.99	
	Rel. Score	**0.56**	**1.14**	**0.98**	**1.10**	
Bird	Count	24	161	249	104	**538**
	Expected	69.14	129.95	174.25	164.66	
	Rel. Score	**0.35**	**1.24**	**1.43**	**0.63**	
Butterfly	Count	112	157	99	96	**464**
	Expected	59.63	112.07	150.29	142.01	
	Rel. Score	**1.88**	**1.40**	**0.66**	**0.68**	
Grasshop.	Count	10	50	43	65	**168**
	Expected	21.59	40.58	54.41	51.42	
	Rel. Score	**0.46**	**1.23**	**0.79**	**1.26**	
Plant	Count	81	93	148	159	**481**
	Expected	61.81	116.18	155.79	147.21	
	Rel. Score	**1.31**	**0.80**	**0.95**	**1.08**	
Vegetation	Count	185	252	396	412	**1245**
	Expected	160.00	300.72	403.25	381.04	
	Rel. Score	**1.16**	**0.84**	**0.98**	**1.08**	
News	Count	28	88	118	100	**334**
	Expected	42.92	80.67	108.18	102.22	
	Rel. Score	**0.65**	**1.09**	**1.09**	**0.98**	
Total	**Count**	**454**	**855**	**1115**	**1002**	**3426**

Generally, the information group *"orientation"* has the biggest differences between the eastern and the southern sector (Table 49). The differences between western and northern sector are minor.

Looking at the subgroups, the same general trend is noticeable in the *"route information"* and in the *"vertical profile"* subgroup, while the *"virtual trail"* shows the biggest difference on the east-west axis. In all three subgroups, the information seems less relevant in a southern aspect. The high relevance of the *"virtual trail"* in a western aspect corresponds to the high relevance of the *"special functions"* in this environment.

The standardized residuals reach a level which is higher than 2 in all cells of eastern aspect, and on two of the three southern cells. The only real exception of this general trend is the high relevance of the *"virtual trails"* in the western exposition.

Interpretations of this distribution are not easy. The high relevance of the route *"vertical profile"* might correspond to the direction of the hiking trails, and the high relevance of the *"virtual trails"* might just be a coincidence with the higher appearance of western aspect at the end of a journey, where this information was requested significantly more often.

TABLE 49 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "ORIENTATION" AGAINST ASPECT.

		North	East	South	West	Total
Route Info	Count	204	535	443	463	**1645**
	Expected	211.40	397.33	532.80	503.46	
	Rel. Score	**0.96**	**1.35**	**0.83**	**0.92**	
Vertical Profile	Count	96	264	212	230	**802**
	Expected	103.07	193.71	259.76	245.46	
	Rel. Score	**0.93**	**1.36**	**0.82**	**0.94**	
Virtual Trails	Count	26	36	55	105	**222**
	Expected	28.53	53.62	71.90	67.94	
	Rel. Score	**0.91**	**0.67**	**0.76**	**1.55**	
Total	**Count**	**326**	**835**	**710**	**798**	**2669**

DIFFERENCES IN THE GROUP "SPECIAL FUNCTION"

The general trend of the group *"special function"* is strongly dominated by the information subgroup *"bookmarks",* because the *"search"* function shows no dependency on aspect. It can be seen that the trend of the *"special function"* subgroup is reduced by the *"search"* function; thus, the standardized residuals in the group *"special function"* in an eastern exposition become too small to be significant. In the sub group *"bookmarks"* northern and western aspects seem to influence the relevance in the opposite direction. Again, the interpretation of these phenomena is not easy. It might just be a coincidence or spatial autocorrelation.

TABLE 50 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "SPECIAL FUNCTION" AGAINST ASPECT.

		North	East	South	West	Total
Search	Count	54	112	140	154	**460**
	Expected	59.12	111.11	148.99	140.79	
	Rel. Score	**0.91**	**1.01**	**0.94**	**1.09**	
Bookmarks	Count	57	177	278	366	**878**
	Expected	112.83	212.07	284.38	268.72	
	Rel. Score	**0.51**	**0.83**	**0.98**	**1.36**	
Total	**Count**	**111**	**289**	**418**	**520**	**1338**

5.7 VEGETATION

5.7.1 HOW MUCH INFORMATION IS NEEDED?

Eisenhut et al. (2008) discussed in their paper how the vegetation cover influences the users' request behavior. They stated that in *pastures* and *grassland* more requests were made than in the *forest.* Looking at Figure 39, the same distribution is also visible for the subset of points from this study. Because the studies are based on the same data set, it can be assumed that most of the differences are due to different processing. Most of the difference is observable in terms of amplitude. The trends are generally the same. In the vegetation classes *pastures* and *grassland, forest* and *residency and traffic* the same general trends are observable, namely that more than expected was requested in pasture/grassland and residency/traffic, while less than expected was requested in the forest. The trend in the vegetation class *"other",* which might correspond to the class *"extreme sites"* in Eisenhut et al. (2008), differs from the previous study. This might be explainable by a different regrouping of the HABITALP classes.

In the other three classes, which show the same general trend, the magnitude differs significantly. While both results in forest seem about equal, especially the relative distribution in *pasture and grassland* does differ a lot. In the study of Eisenhut et al. (2008) twice as many requests compared to the expected value were made in grassland, while in this study the difference is only about 16% higher than expected.

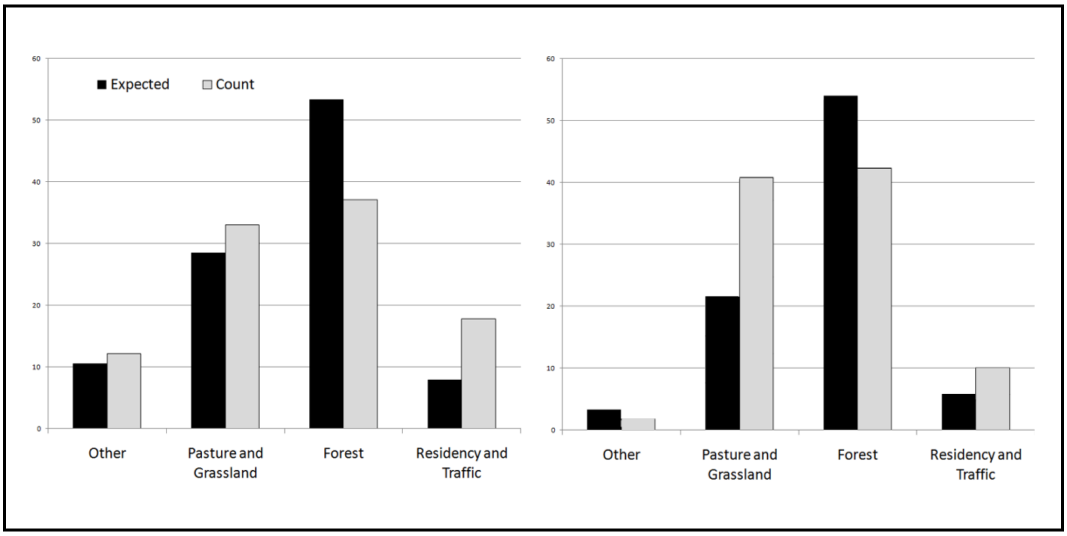

FIGURE 39 : COMPARISON WITH EISENHUT ET AL. (2008) ON THE RIGHT SIDE. BLACK ARE EXPECTED VALUES ON THE TRAILS, GREY VALUES ARE COUNTED FREQUENCIES.

A possible explanation for this distribution might be the limited intervisibility of objects in the forest, which could lead to fewer requests.

5.7.2 WHAT INFORMATION IS REQUESTED?

Looking at the cell values of the different groups, it seems that the biggest differences occur in grassland (Table 51). A major difference between the use of *"special functions"* and requests on *"information around"* is visible. On the other hand the differences in forest seem to be relatively small, because almost all values are high. Only *"information around"* was requested less. The increased number of requests in grassland was compensated by fewer requests in forest. For the *"info on trail"* the opposite is observable. The content function only shows differences between *grassland* and the *other* class. Exactly the opposite is visible for the class *"orientation". "Special functions"* are especially less requested in grassland. The other values seem less significant. The *"information on device"* function does not seem to depend on the vegetation, because the values are close to 1.

Looking at the standardized residuals the already mentioned trend is confirmed. But unlike the distribution of the relevance scores the influence of the different vegetation cover changes. Again, the biggest influence is observable in the grassland. On second place comes the forest, because the standardized residual for the information group *"info around"* is the one with the highest amplitude with -6.4. Therefore it can be stated that grassland and forest have the biggest influence on the request behavior, while the other two groups, which are underrepresented, have only a minor influence on the request behavior of the users.

An explanation for that distribution could be that it might be more difficult to locate oneself in the forest, because the view is limited. The same argument can be applied for the additional relevance of *"information around"* in *grassland.* Because of the better view, more objects can be seen in *grassland* and therefore more requests on that purpose were performed.

The distribution within the groups of information was tested with the Chi^2 test on independency. While the groups *"info around", "info on trail", "content"* and *"orientation"* show some differences in the subgroups, the groups *"special function"* and *"info on device"* did not. They therefore need not be analyzed further.

TABLE 51 : CROSSTABULATION OF THE INFORMATION GROUPS AGAINST VEGETATION.

		Other	Grassland	Forest	Residency	Total
Info around	Count	1012	3251	2909	1661	**8833**
	Expected	1073.86	2912.53	3277.30	1569.30	
	Rel. Score	**0.94**	**1.12**	**0.89**	**1.06**	
Info on trail	Count	640	1539	2140	853	**5172**
	Expected	628.78	1705.38	1918.96	918.88	
	Rel. Score	**1.02**	**0.90**	**1.12**	**0.93**	
Content	Count	439	970	1238	576	**3223**
	Expected	391.83	1062.73	1195.83	572.61	
	Rel. Score	**1.12**	**0.91**	**1.04**	**1.01**	
Orientation	Count	281	884	1013	355	**2533**
	Expected	307.95	835.21	939.82	450.02	
	Rel. Score	**0.91**	**1.06**	**1.08**	**0.79**	
Special function	Count	191	330	513	274	**1308**
	Expected	159.02	431.29	485.31	232.38	
	Rel. Score	**1.20**	**0.77**	**1.06**	**1.18**	
Info on device	Count	116	292	363	196	**967**
	Expected	117.56	318.85	358.79	171.80	
	Rel. Score	**0.99**	**0.92**	**1.01**	**1.14**	
Total	**Count**	**2679**	**7266**	**8176**	**3915**	**22036**

In grassland more *"info around"* was requested than in *forest*. In the other two categories the amplitude of the impact of the vegetation was not as big as for *forest* and *grassland.* The group *"info around"* seems to be homogenous within its subgroups (Table 51). All four subgroups show a higher relevance in *grassland,* while having a lower relevance in *forest.* The *"get around"* function is the only function which has a high value in the *residency* class. All values in the group *other* are close to 1 and the impact seems to be small.
The biggest differences in terms of standardized residuals are in the sub group *"get around",* and the subgroup *"Map (FOI)"* has no big standardized residuals at all. Therefore, it can be assumed that the function *"Map (FOI)"* is not depending on the vegetation cover. The biggest value is observable for the *"get around"* function in forest. There, the function was requested much less than on average while it has been requested much more in *grassland.*
A possible interpretation for this distribution might be that the hikers have a broader sight and therefore might be interested in what is around them, while the sight is limited in the forest, and therefore the hiker might be less stimulated for information on the surrounding area.

TABLE 52 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO AROUND" AGAINST VEGETATION.

		Other	Grassland	Forest	Residency	Total
FOI List	Count	387	1133	1127	548	**3195**
	Expected	388.43	1053.50	1185.44	567.64	
	Rel. Score	**1.00**	**1.08**	**0.95**	**0.97**	
Get Around	Count	357	1237	957	686	**3237**
	Expected	393.53	1067.35	1201.02	575.10	
	Rel. Score	**0.91**	**1.16**	**0.80**	**1.19**	
Info on FOI	Count	207	664	617	335	**1823**
	Expected	221.63	601.10	676.39	323.88	
	Rel. Score	**0.93**	**1.10**	**0.91**	**1.03**	
Map(FOI)	Count	61	217	208	92	**578**
	Expected	70.27	190.59	214.45	102.69	
	Rel. Score	**0.87**	**1.14**	**0.97**	**0.90**	
Total	**Count**	**1012**	**3251**	**2909**	**1661**	**8833**

The *"info on trail"* group exhibits a general trend, which is complementary to the distribution of the group *"info around"* (Table 53). Therefore the function is more relevant in the *forests* than in *grassland.* All subgroups show a higher relevance in forest than in grassland. The distribution of the subgroup *"vertical profile"* and *"virtual"* trail are very similar because the values for residency and the vegetation class *other* are also similar. The subgroup *"route info"* has a high value in the class *other,* while it has a low value in residency. The highest amplitude is observable in *forests,* while the lowest value is observable in the *residency* class.

All standardized residuals in the subgroup *"route info"* are greater than two, which indicates a clear dependency on the vegetation cover. A similar statement for the subgroup *"virtual trail"* cannot be made, because the standardized residuals indicate that the distribution might just be random.

Some trends might be just correlation effects with other variables such as time. At the end of a hike for instance, the *"virtual trails"* and other special functions become more important. At this time, the hikers are often in *residency* areas. But the *"route info"* which might also have an orientation component is requested more in *forests,* which would correspond to the distribution in the information group *"orientation".*

TABLE 53 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO ON TRAIL" AGAINST VEGETATION.

		Other	Grassland	Forest	Residency	Total
Route Info	Count	458	1061	1392	521	**3432**
	Expected	417.24	1131.64	1273.37	609.74	
	Rel. Score	**1.10**	**0.94**	**1.09**	**0.85**	
Vertical Profile	Count	159	411	631	266	**1467**
	Expected	178.35	483.72	544.30	260.63	
	Rel. Score	**0.89**	**0.85**	**1.16**	**1.02**	
Virtual Trail	Count	23	67	117	66	**273**
	Expected	33.19	90.02	101.29	48.50	
	Rel. Score	**0.69**	**0.74**	**1.16**	**1.36**	
Total	**Count**	**640**	**1539**	**2140**	**853**	**5172**

In general, the group *"content"* is influenced mainly by the vegetation class grassland (Table 54). The other three classes have a minor influence. Compared to other information groups only a minor influence is observable, which is in contrast to what could have been expected, because at least some subgroups like *"vegetation"* and *"plants"* have a proposed connection to the vegetation cover.

Looking deeper in the subclasses some complementary effects cancel each other out and result in a nonsignificant distribution in the superior class. In the subclass *"butterfly",* a major difference between *forest* and *pasture / grassland* is visible. Interestingly the *"grasshopper"* subgroup seems to be more relevant in the *residency* areas. Because of the low expected value the standardized residual is not as big as the index might suggest. Another interesting fact is the apparent independency of the subclass *"plant"* on the vegetation cover. The given standardized residuals are all greater than 1, but smaller than 2. It can therefore not be assumed that the differences are significant. Also, in the subgroup *"vegetation"* it seems that there is no difference between *forest* and *grassland,* but a difference between the vegetation classes *other* and *residency* occurs. In general, it can be stated that the biggest differences occur for the animals and surprisingly not for the vegetation, which could be assumed to be highly dependent on the vegetation cover.

No suitable explanation can be found for the distribution of the fauna. In contrast to that, it is imaginable that the vegetation has an influence on the occurrence of animals like "*butterflies".* Derived from the request pattern their habitats are in the forest, which is not unlikely after Weidmann (1995). But in order to reach a firm statement more research on this specific topic is necessary.

TABLE 54 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "CONTENT" AGAINST VEGETATION.

		Other	Grassland	Forest	Residency	Total
Bear	Count	18	46	76	50	**190**
	Expected	23.10	62.65	70.50	33.76	
	Rel. Score	**0.78**	**0.73**	**1.08**	**1.48**	
Bird	Count	63	179	179	97	**518**
	Expected	62.98	170.80	192.19	92.03	
	Rel. Score	**1.00**	**1.05**	**0.93**	**1.05**	
Butterfly	Count	64	98	218	78	**458**
	Expected	55.68	151.02	169.93	81.37	
	Rel. Score	**1.15**	**0.65**	**1.28**	**0.96**	
Grasshop.	Count	16	50	52	49	**167**
	Expected	20.30	55.07	61.96	29.67	
	Rel. Score	**0.79**	**0.91**	**0.84**	**1.65**	
Plant	Count	65	128	180	69	**442**
	Expected	53.74	145.74	163.99	78.53	
	Rel. Score	**1.21**	**0.88**	**1.10**	**0.88**	
Vegetation	Count	176	374	413	164	**1127**
	Expected	137.01	371.61	418.15	200.23	
	Rel. Score	**1.28**	**1.01**	**0.99**	**0.82**	
News	Count	37	95	120	69	**321**
	Expected	39.03	105.84	119.10	57.03	
	Rel. Score	**0.95**	**0.90**	**1.01**	**1.21**	
Total	**Count**	**439**	**970**	**1238**	**576**	**3223**

In general more information on *"orientation"* purposes was requested in *grassland* and *forest* (Table 55). Looking at the subgroups, this general trend is only conceivable in the subgroup *"map page".* The size of this subclass leads to the distribution of the superior group *"orientation".* The *"where am I"* function shows no significant dependency on the vegetation class, and the subclass *"map overlay"* shows no clear distribution.

The standardized residuals are not very large in all cells. The biggest magnitude of -4.8 is observable in the subgroup *"map page"* in residency. The *"where am I"* function only has values which are close to 0 and therefore independency on the vegetation cover must be assumed. Dependency on the vegetation covers grassland and forest can be assumed for the *"map overlay"* function, because only those cells have largeenough standardized residuals.

A possible explanation for distribution of the requests for the *"map page"* function might be the additional information which helps the users to orientate properly. Such help can take the form of streets and street signs. On the other hand, orientation is more difficult in the forest, where the *"map page"* function was requested more often. No suitable explanation for the additional requests of the *"map overlay"* function in grassland can be found.

TABLE 55 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "ORIENTATION" AGAINST VEGETATION.

		Other	Grassland	Forest	Residency	Total
Map Page	Count	167	550	642	194	**1553**
	Expected	188.80	512.08	576.21	275.91	
	Rel. Score	**0.88**	**1.07**	**1.11**	**0.70**	
Where am I?	Count	80	238	312	134	**764**
	Expected	92.88	251.92	283.47	135.74	
	Rel. Score	**0.86**	**0.94**	**1.10**	**0.99**	
Map Overlay	Count	34	96	59	27	**216**
	Expected	26.26	71.22	80.14	38.38	
	Rel. Score	**1.29**	**1.35**	**0.74**	**0.70**	
Total	**Count**	**281**	**884**	**1013**	**355**	**2533**

5.8 WEATHER

Unfortunately, the precipitation could not be analyzed, because only 18 of 201 users have performed 747 actions in a period where the precipitation was >0, which means that 10% of the users requested only 3% of the total amount of information in rainy weather. If the data was to be analyzed, it would be biased and no statistically firm statements could be made. A statement that can be made is that the device is in general less used while it is raining, be it due to the general usage of the device in the summer, or be it also due to fluctuations during a shorter time period. Unlike the precipitation, however, the temperature can be analyzed.

5.8.1 HOW MUCH INFORMATION IS NEEDED?

The first question that needs to be answered is the question of how much information is requested.

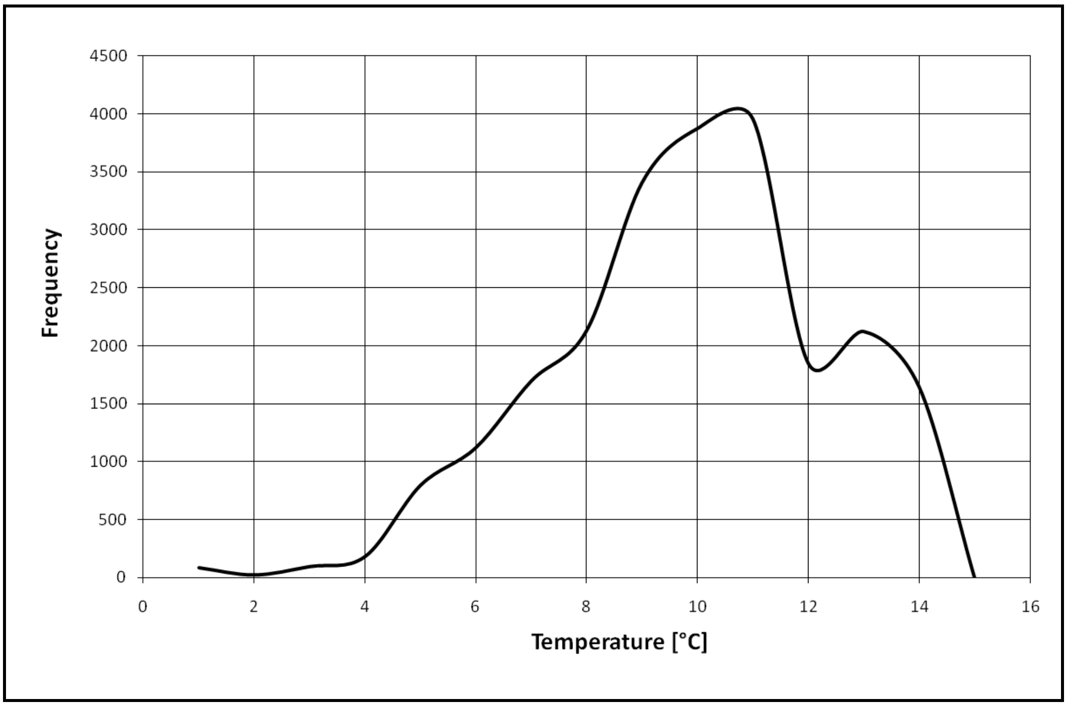

FIGURE 40 : INTERPOLATED FREQUENCIES PER °C.

The request frequency per temperature is shown in Figure 40. The maximum is reached in moderate temperature and the curve generally has a negative skew. Therefore, the device is in general more used in medium temperatures, and less in very warm and very cold temperatures. But this curve has to be normalized by the number of users, which even were active during such environmental setups. The following table 56 could be calculated with the temperature quantiles of [-3.5, 4.5], (4.5, 8.7], (8.7, 13.2] and (13.2, 22.6]. In a homogenous space in all classes the same amount of information would be requested.

TABLE 56 : USERS PER TEMPERATURE CLASS AND THEIR REQUESTS.

Temperature class	Number of Users	Number of Requests	Requests per User
Very cold (<4.5°C)	27	1199	44.41
Cold (4.5-8.7°C)	67	2866	42.78
Warm (8.7-13.2°C)	122	7226	59.23
Hot (>13.2°C)	132	11695	88.60
Mean	87	5746.5	66.05

Out of a total of 201 users, 132, or 65%, of the users were hiking in a period where at least once during their hike the temperature was comparatively warm, while only 27% of the hikers experienced temperatures less than 4.5°C. Looking at the requests that they made, it becomes clear that during the two coldest periods the same amount of information was requested, while the frequency risincreases to almost twice that value in the hot period. With this normalization it can be stated that the users are more active and request more information in warmer periods.

5.8.2 WHAT INFORMATION IS REQUESTED?

The first main observation of Table 57 is that there are big differences between the temperature classes. While *"information on the trail"* was requested much more in a cold environment, *"content"* and "*special functions"* seem to be much less interesting in such a situation. *"Special functions"* seem to be more popular in a comparatively warm environment and *"content"* and *"orientation"* functions seem to be the most desired in a hot period. The information group *"information on device"* shows almost no dependency on temperature.

In terms of standardized residuals all four temperature categories seem to influence the request pattern in a similar magnitude. The requests are the least influenced in a cold environment. Only two standardized residuals have a magnitude greater than 2, namely those of the *"info around"* and *"orientation".* Looking specifically at the distribution of the values of the information subgroups, the requests on *"content"* are the most sensitive to the temperature, while the *"info on device"* seems not to be dependent on the temperature at all. The standardized residuals for that group are all smaller than 1. By contrast to that a very high standardized residual is observable in a *very cold* environment in the information group *"info on trail".* With a value of almost 10 much more information on that specific content is requested.

A possible explanation for the distribution could be that in the morning when it is-cold *"information on the trail"* is more important, while animals and plants might be influenced by their "daily routine".

The groups of information were tested, whether their subgroups show different behavior. With a level of significance of 0.05 it can be said that in all six groups, the subgroups show a dependency on the variables, and must therefore be analyzed further.

TABLE 57 : CROSSTABULATION OF THE INFORMATION GROUPS AGAINST TEMPERATURE.

		Very cold	Cold	Warm	Hot	Total
Info around	Count	438	1238	3205	4234	**9115**
	Expected	475.46	1136.50	2865.44	4637.60	
	Rel. Score	**0.92**	**1.09**	**1.12**	**0.91**	
Info on trail	Count	450	700	1665	2618	**5433**
	Expected	283.40	677.41	1707.95	2764.24	
	Rel. Score	**1.59**	**1.03**	**0.97**	**0.95**	
Content	Count	69	430	825	2102	**3426**
	Expected	178.71	427.17	1077.02	1743.11	
	Rel. Score	**0.39**	**1.01**	**0.77**	**1.21**	
Orientation	Count	145	215	702	1607	**2669**
	Expected	139.22	332.78	839.04	1357.96	
	Rel. Score	**1.04**	**0.65**	**0.84**	**1.18**	
Special function	Count	43	146	516	633	**1338**
	Expected	69.79	166.83	420.62	680.76	
	Rel. Score	**0.62**	**0.88**	**1.23**	**0.93**	
Info on device	Count	54	137	313	501	**1005**
	Expected	52.42	125.31	315.94	511.33	
	Rel. Score	**1.03**	**1.09**	**0.99**	**0.98**	
Total	**Count**	**1199**	**2866**	**7226**	**11695**	**22986**

Differences in the group "Info around"

The general trend of the information group *"info around"* is towards a slightly higher relevance during moderate temperatures, while during very high or very low temperatures, it is less important (Table 58). By contrast to this distribution the subgroup *"FOI List"* shows an exactly inverse distribution. But the residuals are too small, and therefore it must be assumed that this function does not depend on the temperature. On the other hand the *"get around"* function has even an increased magnitude of the distribution of the superior group. Very high values are observable in a cold period. The *"info on FOI"* function in contrast shows a very indifferent distribution. High values in a very cold period, low values in a cold period, again high values in a warm period and low values again in a hot period, cannot be interpreted. Again a distinct and clear distribution of the sub group *"Map (FOI)"* shows a clear distribution towards high values in a moderate warm temperature environment.
A possible interpretation of the general trend might be that information on the surrounding area has a special importance in a moderate temperate environment, because if it is too hot or too cold, the hiker might be more interested in getting warm or resting. On the other hand the "*Map (FOI)"* function shows a clear distribution towards a moderate warm temperature profile, due to low interest in such advanced functions in a cold environment.

TABLE 58 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO AROUND" AGAINST TEMPERATURE.

		Very cold	Cold	Warm	Hot	Total
FOI List	Count	180	415	1045	1714	**3354**
	Expected	174.95	418.19	1054.38	1706.47	
	Rel. Score	**1.03**	**0.99**	**0.99**	**1.00**	
Get Around	Count	107	597	1238	1354	**3296**
	Expected	171.93	410.96	1036.15	1676.97	
	Rel. Score	**0.62**	**1.45**	**1.19**	**0.81**	
Info on FOI	Count	139	181	705	852	**1877**
	Expected	97.91	234.03	590.06	954.99	
	Rel. Score	**1.42**	**0.77**	**1.19**	**0.89**	
Map(FOI)	Count	12	45	217	314	**588**
	Expected	30.67	73.31	184.85	299.17	
	Rel. Score	**0.39**	**0.61**	**1.17**	**1.05**	
Total	**Count**	**438**	**1238**	**3205**	**4234**	**9115**

The trend of the information group *"info on trail"* strongly points towards a higher relevance in a cold environment, while this information is less relevant in a warm environment (Table 59). This trend is visible in the subgroups *"route information"* and *"vertical profile"*, while the *"virtual trails"* show a similar distribution as the "*Map (FOI)"*. This distribution is a bit indifferent, while the distributions on the other two information subgroups are more distinct and clear.
The standardized residuals are especially striking in a very cold environment. For both *"route info"* and *"vertical profile"* the residuals are greater than 7, which indicates that clearly more information is requested in such temperatures. All other values are comparatively small. Therefore the only real trend is that more *"information on the trail"* is requested when the temperatures are very low.
The distribution might be explained due to the higher importance of these functions at the beginning of a hike when the temperature is also low. It could have been argued that the *"vertical profile"* might be more important in a hot environment, because of the additional energy needed to go about a slope. But the opposite is observable. The function route *"vertical profile"* is much more often used in a cold environment.

TABLE 59 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO ON TRAIL" AND THE TEMPERATURE.

		Very cold	Cold	Warm	Hot	Total
Route Info	Count	281	472	1132	1761	3646
	Expected	190.18	454.60	1146.18	1855.04	
	Rel. Score	**1.48**	**1.04**	**0.99**	**0.95**	
Vertical Profile	Count	149	199	414	749	1511
	Expected	78.82	188.40	475.01	768.78	
	Rel. Score	**1.89**	**1.06**	**0.87**	**0.97**	
Virtual Trail	Count	20	29	119	108	276
	Expected	14.40	34.41	86.76	140.43	
	Rel. Score	**1.39**	**0.84**	**1.37**	**0.77**	
Total	**Count**	**450**	**700**	**1665**	**2618**	**5433**

Content is generally much less requested in a very cold environment, while it has been requested often in warmer conditions Looking at Table 60 the first general statement that has to be made concerns the frequencies in cells. The expected values never drop below 5. But the observed values are sometimes very small, therefore the values have to be evaluated with caution.

The values are low in a very cold environment for all information subgroups while they tend to be high in a hot environment. Only the subgroup *"news"* reaches its maximum in a warm environment. Interestingly, also the subgroup *"bird"* reaches its strong distinctive maximum in a cold environment. The magnitudes of the scores are extreme and reach extreme values such as 0.17 and 1.81. These values must taken with caution, because the expected values are very small and fall under 10. Therefore, the analysis of the standardized residuals is far more important.

It cannot be assumed that the subgroup *"bear"* depends on the temperature, because the standardized residuals are very small. Also, the two subgroups *"grasshopper"* and *"plant"* might not depend on temperature, even though the magnitudes are slightly bigger. The most extreme values are observable in the subgroup *"vegetation"* where in a very cold environment the value is -6.1 and in a hot environment 7. Interestingly, the value in a warm environment is also strongly negative (-6.8). Therefore the distribution has a discontinuous character. The absolute maximum in terms of standardized residuals is reached in the subgroup *"bird"* in a cold environment, where twice as many requests were made than on average.

The content is much more important in higher temperatures, while being less important in cold weather. A possible explanation can be the additional activity of some animals in warm temperatures. But because animals can also be more active at night, this would have to be investigated first. Another factor might be the visibility of the animals. In the morning, when it is usually also colder than at noon, they might be seen less often. For the plants a similar interpretation can be made, because plants, especially flowers are better visible when their blossoms are open.

Because at least the expected values in a very cold environment are very small all interpretations have to be handled with care. Random factors as well as correlation between variables might have a big influence on these values.

TABLE 60 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "CONTENT" AGAINST TEMPERATURE.

		Very cold	Cold	Warm	Hot	Total
Bear	Count	8	23	65	100	**196**
	Expected	10.22	24.44	61.62	99.72	
	Rel. Score	**0.78**	**0.94**	**1.05**	**1.00**	
Bird	Count	6	128	103	301	**538**
	Expected	28.06	67.08	169.13	273.73	
	Rel. Score	**0.21**	**1.91**	**0.61**	**1.10**	
Butterfly	Count	4	13	101	346	**464**
	Expected	24.20	57.85	145.87	236.08	
	Rel. Score	**0.17**	**0.22**	**0.69**	**1.47**	
Grasshop.	Count	8	20	31	109	**168**
	Expected	8.76	20.95	52.81	85.48	
	Rel. Score	**0.91**	**0.95**	**0.59**	**1.28**	
Plant	Count	12	63	124	282	**481**
	Expected	25.09	59.97	151.21	244.73	
	Rel. Score	**0.48**	**1.05**	**0.82**	**1.15**	
Vegetation	Count	15	150	265	815	**1245**
	Expected	64.94	155.23	391.38	633.44	
	Rel. Score	**0.23**	**0.97**	**0.68**	**1.29**	
News	Count	16	33	136	149	**334**
	Expected	17.42	41.64	105.00	169.94	
	Rel. Score	**0.92**	**0.79**	**1.30**	**0.88**	
Total	**Count**	**69**	**430**	**825**	**2102**	**3426**

DIFFERENCES IN THE GROUP "ORIENTATION"

Orientation functions are in general more important in a warm environment, which is also visible in two of three subgroups (Table 61). The only exception with a high relevance in a very cold environment is the *"where am I"* function. Nevertheless, it can be said that the trend points towards a higher importance of this information in a warmer environment. Interestingly also the *"map overlay"* function, with a total of only 222 requests follows this trend.

Looking at the standardized residuals the trend of the *"where am I"* function has to be put into perspective, because the only significant difference is visible in a very cold environment. And even there the amplitude is not strong enough to be completely sure. In all other cells small standardized residuals are observable. Far greater residuals are observable for the subgroup *"map page"*, which is complementary to the *"where am I*" function. Three of four cells have very high standardized residuals and both the absolute maximum and minimum for this information group can be found in the subgroup *"map page"*. Therefore it can be assumed that displaying the map is highly depending on the temperature. Also the *"Map Overlay"* function depends significantly on temperature and only in a hot environment more requests were made than on average. If it is slightly colder this function becomes less relevant.

The distribution of these values cannot be explained through the distribution depending on the different times of a day, because it is not affected by that. It just might be a result of the frequencies of other functions, such as the *"info on trail"*. An explanation for the distribution of the subgroup *"map page"* could be that hikers are interested in finding their way more efficiently, because they are sweating more.

TABLE 61 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "ORIENTATION" AGAINST TEMPERATURE.

		Very cold	Cold	Warm	Hot	Total
Map Page	Count	75	118	406	1046	**1645**
	Expected	85.81	205.11	517.13	836.96	
	Rel. Score	**0.87**	**0.58**	**0.79**	**1.25**	
Where am I?	Count	61	85	251	405	**802**
	Expected	41.83	100.00	252.12	408.05	
	Rel. Score	**1.46**	**0.85**	**1.00**	**0.99**	
Map Overlay	Count	9	12	45	156	**222**
	Expected	11.58	27.68	69.79	112.95	
	Rel. Score	**0.78**	**0.43**	**0.64**	**1.38**	
Total	**Count**	**145**	**215**	**702**	**1607**	**2669**

"Special functions" overallare mostly used in an environment of medium warmth. Looking at the subgroups the trends are complementary and not very distinct. It seems like the *"search"* function is not highly dependent on temperature. Only a small trend towards a higher relevance in a very cold environment is observable. The *"bookmark"* functions on the other hand show a higher relevance in warm environment and are less relevant under "extreme" weather conditions.

The standardized residuals for the *"search"* function are generally very small. The only exception is the small relevance in a cold environment. In contrast to that observation a more distinct distribution is noticeable for the *"bookmark"* functions. The differences between very cold and warm temperatures are significant, and the amplitude of both maximum and minimum at -4.7 and 5.4 is already distinct in terms of standardized residuals.

A possible explanation for these distributions might be that the users have to concentrate on other things in a very cold or hot environment. Either they are sweating or they are cold. In a warm environment the users can concentrate more on *"special functions"* which might be more demanding.

TABLE 62 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "SPECIAL FUNCTION" AGAINST TEMPERATURE.

		Very cold	Cold	Warm	Hot	Total
Search	Count	29	36	151	244	**460**
	Expected	23.99	57.35	144.61	234.04	
	Rel. Score	**1.21**	**0.63**	**1.04**	**1.04**	
Bookmarks	Count	14	110	365	389	**878**
	Expected	45.80	109.47	276.01	446.72	
	Rel. Score	**0.31**	**1.00**	**1.32**	**0.87**	
Total	**Count**	**43**	**146**	**516**	**633**	**1338**

Differences in the group "Info on Device"

The information group *"info on device"* does not follow a general trend (Table 63). It is interesting that there is a significant difference between the subgroups, meaning that their trends cancel each other out. Looking at the distributions of both subgroups, it becomes obvious that both functions have complementary trends. The *"tutorials"* are used more frequently in a very cold or cold environment, while the *"key applications"* were used more often in a hot environment.

But the standardized residuals show something different. The only significant difference between two cells is in the cells of a very cold and a hot environment for the *"tutorial"* function. Therefore it can be assumed that the *"key application"* function does not depend on the temperature, while the dependency of the *"tutorial"* function is not very distinct.

A possible explanation for this distribution are the small expected values, which increases possibly random effects, and that the tutorial is mostly requested at the beginning of a hiking journey, which is more likely in the morning when the temperatures are low.

TABLE 63 : CROSSTABULATION OF THE SUBGROUPS OF THE INFORMATION GROUP "INFO ON DEVICE" AGAINST TEMPERATURE.

		Very cold	Cold	Warm	Hot	Total
Tutorial	Count	31	56	121	138	**346**
	Expected	18.05	43.14	108.77	176.04	
	Rel. Score	**1.72**	**1.30**	**1.11**	**0.78**	
Key Applications	Count	23	81	192	363	**659**
	Expected	34.37	82.17	207.17	335.29	
	Rel. Score	**0.67**	**0.99**	**0.93**	**1.08**	
Total	**Count**	**54**	**137**	**313**	**501**	**1005**

5.9 AMPLITUDE OF INFLUENCE

The first step of the analysis is to become aware of the power the different variables have. In order to assess this power the standard deviation of the relevance scores of each context factor are calculated.

TABLE 64 : STANDARD DEVIATION OF THE RELEVANCE FACTORS.

Relevance Factors	Minimum	Maximum	Stddev.
Intrinsic relevance	0.15	3.33	1.14
Trails	0.30	2.88	0.30
Temperature	0.17	1.91	0.22
Aspect	0.35	1.88	0.21
User group	0.35	1.68	0.20
Relative time	0.32	1.88	0.19
Absolute time	0.57	1.83	0.15
Slope	0.44	1.63	0.14
Vegetation	0.65	1.65	0.13
Picnic Area	0.76	1.67	0.13
Mean	0.41	2.03	0.28

In Table 64 it can be seen that the intrinsic relevance has the biggest influence on the total relevance score. The standard deviation is almost four times bigger than the average. It can be explained with the heterogeneity of the information groups. The *"get around"* function, for instance, combines many different contents, while the content on *"butterflies"* is very specific. Therefore the assumed and observed values differ a lot.

For the other factors the influence of the larges scale, the trails, shows the biggest difference, while a feature, which manifests in a small scale, the distance to the picnic areas, has a comparatively small standard deviation. Also the vegetation, the slope and the absolute time have a relatively small influence.

After the evaluation of the magnitude of the influence of the different relevance factors, the correlation between the variance values needs to be calculated. Because all nine factors can, after testing them with the Kolmogorow-Smirnoff test, be assumed to not be normally distributed, Spearman's rank correlation coefficient can be applied.

Eight of the 45 possibilities have a correlation coefficient greater than 0.1, but only four do correlate significantly with a level of significance of 0.05. The factors slope and trails, temperature and picnic areas, intrinsic relevance and slope, as well as the factors intrinsic relevance and relative time correlate significantly. But in all cases the magnitude of the correlation coefficient is very small (-.003 - -.013). Therefore none of the factors correlates with another factor. All factors can be assumed to have their own variance in the data. A principle components analysis would therefore not make much sense.

5.10 Influence on the Information Subgroups

The standard deviation of the total relevance specified for each subgroup can help to estimate which information subgroup depends mostly on the context. But the standard deviation has to be normalized because the frequency of each subgroup has an influence on the relevance scores and, therefore, also their standard deviations. A possibility is to multiply the standard deviation with the *intrinsic relevance* of each subgroup. In Table 65 these scores can be seen sorted in descending order.
The information subgroups *"get around"*, *"butterfly"*, and *"map overlay"* show the highest dependency on context, while the subgroups "news", *"key applications"*, and the information on *"bears"* seem to be less sensitive to context.

TABLE 65 : INFORMATION SUBGROUPS DEPENDING ON THE CONTEXT.

	Intrinsic relevance	Standard deviation	Corrected Stdev.
Get Around	3.01	1.08	3.25
Butterfly	0.42	6.24	2.62
Map Overlay	0.2	11.57	2.31
Route Info	3.33	0.64	2.13
FOI List	3.06	0.59	1.81
Bird	0.49	3.00	1.47
Bookmarks	0.80	1.75	1.40
Vertical Profile	1.38	0.88	1.21
Map Page	1.50	0.78	1.17
Info on FOI	1.71	0.54	0.92
Vegetation	1.14	0.64	0.73
Grasshopper	0.15	3.87	0.58
Where am I	0.73	0.63	0.46
Map(FOI)	0.54	0.68	0.38
Tutorial	0.32	1.09	0.35
Search	0.42	0.69	0.29
Virtual Trails	0.25	1.00	0.25
Plant	0.44	0.56	0.25
Bear	0.18	1.16	0.21
Key Applications	0.60	0.30	0.18
News	0.31	0.53	0.16

5.11 ADDITIONAL QUESTIONS

Relevance scores can be applied to a mobile environment with raster cells as Raper (2007) suggested. Such raster files can be calculated with an ordinary krigin function (Burrough & McDonnell, 2005). As an example, the result of such a function of the information subgroups *"get around"* and *"FOI list"* can be seen in Figure 41. Because two different groups were considered at the same time, it had to be asured, that the intrinsic relevance does not influence the result. Therefore, only the relevance values of the nine extrinsic context factors were multiplied and rasterized in a 50m raster.
The relevance on the trail Chamanna Cluozza & Murter is lower than on the other two trails as it could be seen in chapter 5.2. Some regional differences within the trails are visible, e.g. the probability is higher on the west side than on the east side of the trail Val Trupchun. This result differs from a probability function based on the point density, because not only spatial variables were included in the calculation. Therefore the result claims to represent the relevance of the information better than the simpler point density calculation.
The presented probability is just an example of how the relevance scores could be applied in a mobile environment. How, and also which krigin functions are appropriate to suit the purpose of a mobile map, would have to be discussed.

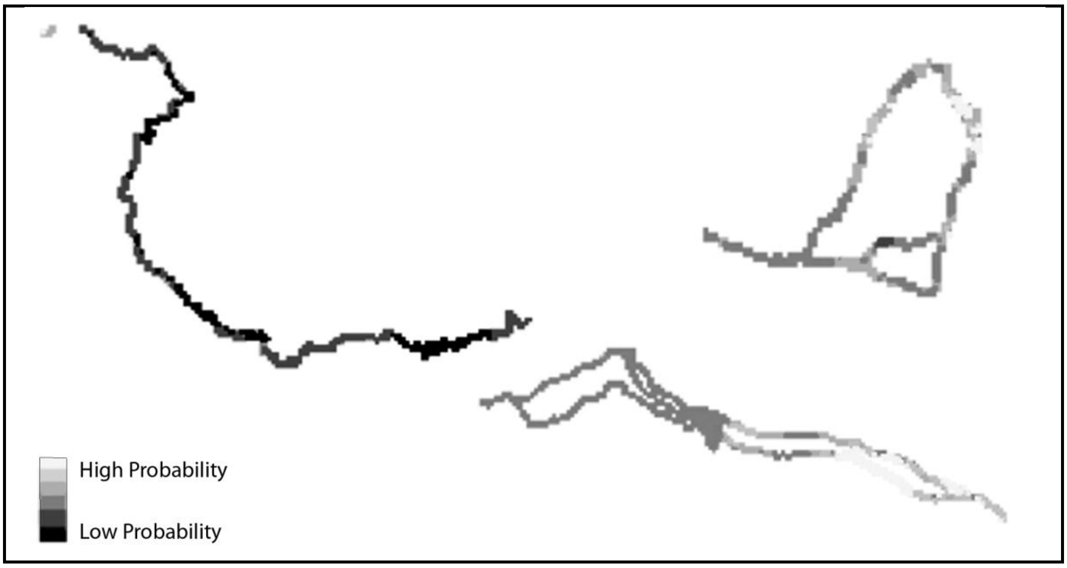

FIGURE 41 : EXAMPLE OF A PROBABILITY MAP FOR THE INFORMATION SUBGROUPS "GET AROUND" AND "FOI LIST".

6 DISCUSSION

6.1 DISCUSSION OF MODEL

The conceptual model, especially the choice of the context factors which are considered, is based on other models described in the literature. Because the choice of context factors was reasoned by their proposed measurability for the given datasets, the list of factors does not claim to be complete. Other variables might also be part of the total context and the list of contextual variables could therefore be expanded. The model of Jameson (2001) had to be simplified, because of limitations, which come mostly from the measurability of certain model parts. Even if the model of Jameson could be seen as complete, its simplified version with no system context and equal utility of all requests is certainly not. The simplifications made lead to results which could be biased. The system architecture and especially the software layout might influence the information requests significantly. In order to answer the research questions the model is sufficient, because significant results are observable in the end.

The computation concept has also some limitations. First it is highly depending on how the context is categorized into its characteristics. Every variable could be classified in a different way and other results might come out. Another problem is that the concept is dependent on high frequencies in every information subgroup, which is certainly not the case. Especially the requests of the subgroups in the information group *"content"* are not frequent enough. If the number of possible requests enlarges, and the relevance scores shall still be significant, then also the frequency of the total requests has to become more numerous.

6.2 DISCUSSION OF METHODS

The quality of a result of a quantitative model is dependent on three factors (Burrough & McDonnell, 2005):

- Data quality
- Model quality
- Interaction of model and data

All these three points will be discussed in order to be able to quantify the accuracy of the model.

6.2.1 DATA QUALITY

Based on Aalders (1996), data quality can be grouped into

- Position accuracy
- Temporal accuracy
- Thematic accuracy

There is no global accuracy for GPS data according to Wing et al. (2005) and DeCesare et al. (2005). Every receiver and region must individually be analyzed in the field. The GPS accuracy may depend on the topographic conditions (D'Eon, 2003), on the season, and the vegetation layer (Dussault, Courtois, & Huot, 1999). Furthermore, the available number of satellites, which is also dependent on weather, is an important factor (Hulbert, I. A.R.; French, J., 2001). No survey points are available for the GPS points from the WebPark^{SNP} dataset of 2007. However, if it is assumed that the hikers should stay on the given trails, then the distance from the points to the trails can be used to estimate the GPS accuracy. 81.5% of the points lie in between a 25 meters buffer, which is the cell size of the DEM. Still 65% lie in this 12.5% buffer, which is half the cell size. The average distance is about 20 meters. Yet, this value might be strongly influenced by outliers, e.g. in Zernez, where no absolute path is defined. The distance is also large in other regions, such as at the eastern end of the trail Chamanna Cluozza & Murter and in the north-east of the trail Margunet, and it is spatially autocorrelated because of these regional dependencies. But also the accuracy of the GIS layer of the trails is important for the estimation of the GPS accuracy. Including all these parameters, the empirical observation of the distances with their spatial autocorrelation, and a possible inaccuracy of the GIS layer of the trails, it can be assumed that the average GPS accuracy is about 10 meters.

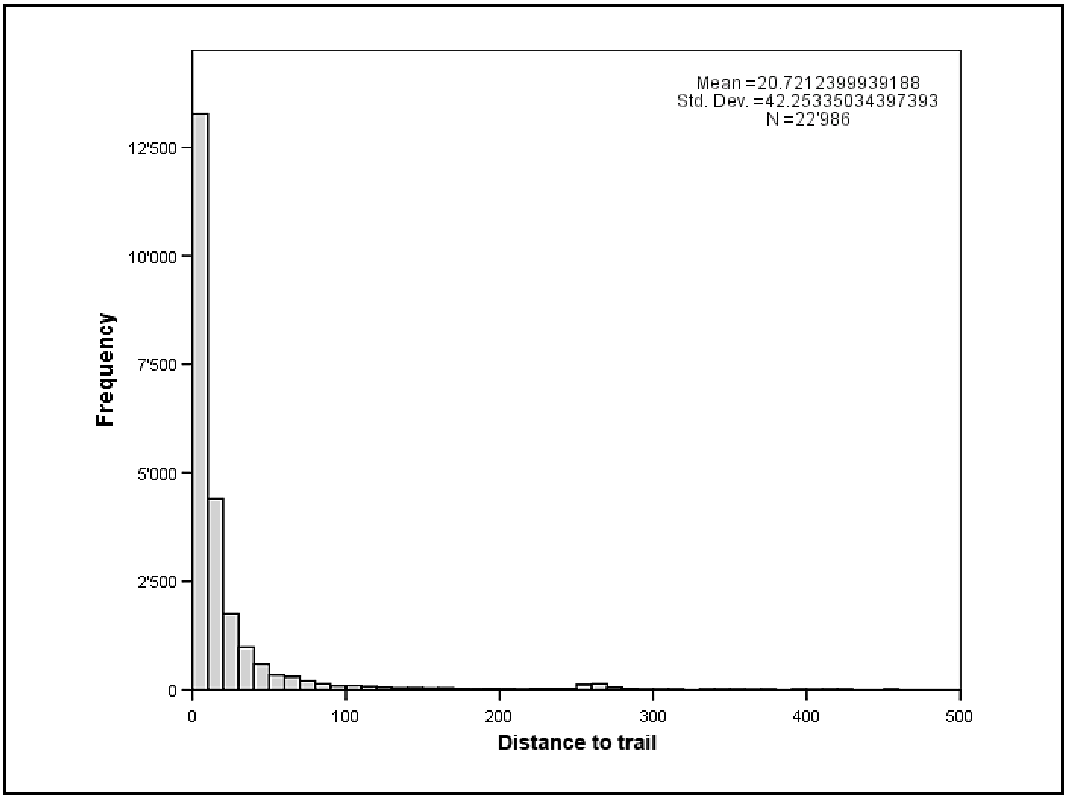

FIGURE 42 : DISTANCE OF THE GPS POINTS TO THE TRAILS.

DEM ACCURACY

The DEM with a resolution of 25 meters is based on the official Swiss topographic map with a scale of 1:25'000, which is produced by the Federal Office of Topography (swisstopo). The accuracy in the Alps is claimed to be 3 meters and was determined with control points on photogrammetric exposures. The error is specified with 4.1 and 5 meters in the region of Zernez and the Ofenpass (Swisstopo, 2005). For the same region Haller and Imfeld (2007) calculated an error of 2.54 and 0.77 meters in two test sites, which is less than the declaration from swisstopo. The data providers are overestimating the error, because they use survey points which are on ridges and exposed places leading to an underestimation in the DEM. But the authors state that global accuracy measurements are not always suitable, because the error is spatially dependent, even though the autocorrelation was lower than expected.
The data is based on measurements in the year 1997, i.e. is ten years before the collection of the GPS log files (Swisstopo, 2005). In this period some changes in the elevation model might have happened. But still the measurements in this area can be assumed to be suitable for our purpose.

VEGETATION LAYER

The vegetation classification is based on an orthoimage with an average resolution of 0.6 meters (Thompson, 2004) and an overall accuracy of the vegetation classification of 85% (Bley & Haller, 2006). The position error in a study by Bauch and Seitling (2006) on a similar dataset raised up to 5 meters, because of the fuzziness of the class boundaries. Therefore the spatial accuracy of the vegetation boundaries is more exact than the GPS accuracy.
The HABITALP project was finished in 2006, a year before the GPS log files were collected. Therefore natural changes in the vegetation could influence the accuracy as well. But the resulting error is rated smaller than the general classification error. The classes fit our purpose thematically well, because the classes of the HABITAP project are standardized, and the results can be reproduced.

METEOROLOGICAL DATA

The temperature can be assumed to be measured accurately, because of long time experiences on this field. Compared to the granularity of the time intervals for the GPS data, the temporal accuracy cannot be rated very high. The thematic accuracy is high, because the temperature is measured directly and no derivations have to be made.

TRAILS AND PICNIC AREAS

It was already mentioned in chapter 4.4, that the position accuracy of the original point dataset is not high. The corrected point dataset shows a much higher accuracy. However, some error must still be assumed, because the centers were determined with the aid of a density raster file, which had a resolution of 4 meters. Therefore, an accuracy of 4 meters can be assumed. A similar quality can be assumed for the trails, because the highest point densities are always near the trails.

The metadata provide information about the temporal accuracy. The trail dataset is from 2007, and because the new picnic areas were derived from the GPS points, they can be seen as being from 2007. Therefore both data sets have a good temporal fit for our purpose.
Thematically both datasets fit our purpose well. In the first place, only official resting areas were of interest. This quality could not be determined from the GPS points, and is therefore not given. But with the definition of the centre of highest density this problem becomes insignificant.

6.2.2 MODEL QUALITY

PREPROCESSING

Selecting only the four considered trails can be questioned with the findings of Eisenhut et al. (2008). High enough frequencies of requests and users might also occur on other trails. The advantage of selecting only these four trails is that they can distinctly be separated in space. Other trails, which provide a high number of hikers and requested information, may not be separated from other trails. For these reasons, the highest number of requested information and hikers and a clear spatial separation from other trails have been chosen as criteria aiming at reducing the complexity of this thesis.
The raw data had to be filtered after the import, in order to attain homogeneous data. Corrupted entries, entries of users who did not hike on the four main trails, and entries from users who were not logged during their main hiking period had to be removed. This processing step is probably the most controversial, because the rules to obtain the final data set are fuzzy. E.g. one rule was that the hikers need to be on one of the main trails for a long enough period of time. But what is *"long enough"*? Because this selection was done manually, another person might have selected different users and points.

GROUPING INFORMATION

In order to provide high enough frequencies for statistical analysis, the requests needed to be grouped. The downside of regrouping the data can be that some correlation or possible features of specific requests get lost.
Some of the information subgroups could also be aggregated with other subgroups. The function *"virtual trail"* for instance was aggregated with other functions and contents on the trails. But it could also be argued that it is a "*special function".* Each subgroup was analyzed separately to check, if the subgroups of each information group differed significantly. Therefore the aggregation problem is solved.
Interestingly some information classes, such as the *"privacy policy"* and information on *"wildfire",* do not appear in this selection of the total data, even though they are in the system. The same information might be accessed through other functions, such as the *"what's around me"* function.

TRAILS

The points can unambiguously be allocated to the trails. But the allocation is depending on the selection of the points in the preprocessing steps.

USER GROUPS

Grouping the users is depending on several factors in the processing chain, e.g. the selection of the picnic areas, setting the dimension of the picnic areas etc. The resulting classes are not distinct. The users could also be grouped differently. If the time for a *pass through* was be reset by only a few minutes, then the result of the classification would be different. If not 5 but 10 minutes were taken, then about 10 users would fall into the *pass through* group. This could influence the result significantly, because the *pass through* group contains only 38 users. And if other picnic areas were selected, up to 10 users would not be classified as *no data*. It could be shown that the amount of requested information by each user does vary a lot in each user group. The classification is not stable, and both criteria lead to the conclusion that the users might also be classified differently. But the requests still differ significantly from each user group, which approves that the classification is at least plausible.
If the classification criteria are correct, it can be assumed that most of the users were classified correctly and due to the calculation error only about 5-10% should have been in a different class.

PICNIC AREAS

There are four factors which mainly influence the result of the picnic areas. One of them, the GPS accuracy, has already been discussed. The selection of the picnic areas, the centre of highest density, and the dimension of the picnic areas influence the result as well. All three criteria could be defined in another way and other results would be the consequence. Looking back at Figure 23 on page 45 it can be estimated how many points could have been falsely classified. About 25% of the requests were made inside the picnic areas. Therefore it can be assumed that a lot of points are definitely not inside the picnic area and the maximum error is about 5%.

RELATIVE TIME

The relative time is mostly influenced by the preprocessing of the points, because in this processing step it is defined when the users start their hikes on the trails. Filtering out points can influence the relative time. This influence is always negative, i.e. the relative time gets reduced. The only possibility to enlarge the relative time, would be if the device was lent twice a day, and the second user starts at the start time of the first user. But this can be ruled out with a high certainty, because the movement of all users is continuously and they do not walk the same trail twice.
Beside the influence of the preprocessing steps, when and where the device gets started first is crucial. For instance, if a user starts his device the first time at the second picnic area after he has already hiked for hours, the result becomes biased. Most of the users have their first entries somewhere near the start of a trail, e.g. at the parking places in the south of the trail Margunet. Only about 5% of the users

have no entries near the start of the trail. Hikers, who do not have entries near the start of a trail, use the device on average less frequent. Therefore the percentage of the points, which are biased, is comparatively small. The average error on the data can roughly be estimated with 5%. Because the time classes are very coarse and one dimension above the estimated error, it can be assumed that at least 80% of the points were classified correctly. Only for 20% a wrong classification could even be possible.

ABSOLUTE TIME

Because the GPS logging system is very exact in terms of time, the absolute time is also very exact. Therefore the error on this variable can be estimated with 0%.

SLOPE / ASPECT

Three different error sources influence the error of slope and aspect:

- Algorithm errors
- DEM data errors
- DEM spatial resolution

The spatial resolution in our dataset is 25 meters with an accuracy of 2.5 meters. Because the raster cells of the dataset are comparatively big, the algorithmic error and the data error have in general less influence on the total error. The data error only influences the steepest drop algorithm in flat areas, because the aspect algorithm is influenced only in such areas, while the slope algorithm gets influenced everywhere (Zhou & Liu, 2008). However, the calculation of the slope in our case is special, because it is derived for linear objects. The effect of the algorithm therefore increases because several processing steps have to be applied. Already the first processing step can lead to a big inaccuracy. The hypothetical trail on the left side of Figure 43 would result in the raster structure on the right. If then the slope of the central pixel was calculated, all grey cells would be incorporated. A more appropriate way of calculating the slope in this special case would be if only the three vertically aligned cells in the middle were taken into account for the calculation. The selection of raster cells could only be considered as *exact* as in a normal environment if the trail was straight and had an angle of 0°, 45° or 90°. Based on geometric thoughts, an error of about 5-10% for straight lines can be assumed. Yet, the polylines in our dataset are normally not straight and therefore this error increases. As shown in Figure 43 the error is unpredictable if the slope is calculated for random curves.

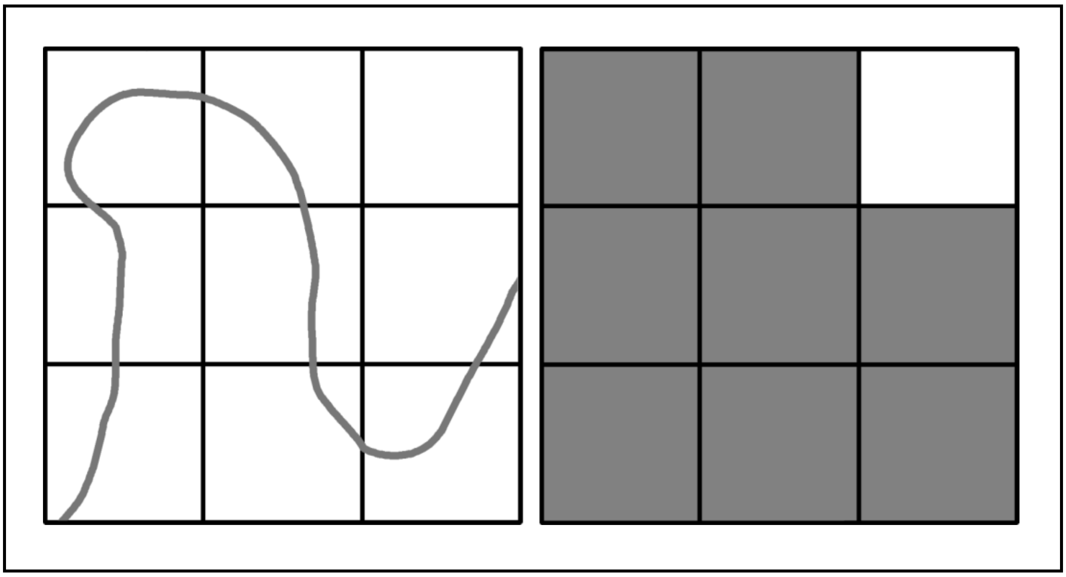

FIGURE 43 : PROBLEMS WITH LINE TO RASTER CONVERSION.

Thiessen polygons were calculated for the centre points of the cells after calculating the slope. These centre points are normally not on the trails and the points can be displaced by almost 18 meters. This error can even get larger, dependent on the distance between trails and GPS points. However, the assumption that tGPS points belong to the slope cell which is located nearest, which is accomplished by the Thiessen polygons, is not always true. Due to the displacement of GPS points, a point could also belong to a raster cell, which is further away.

All these presented factors influencing the association of a GPS point with a slope or aspect point have consequences for the final analysis of the frequencies. The setting of the class borders for the slope is argueable. However, the class-affiliation does not change abruptly in space, because only four characteristics for both context variables exist. Other factors come into play, because the classes for the aspect are given by the cardinal directions. Fisher et al. (2004) discussed the dependency of morphometric phenomena on the scale in which they are determined. They stated that a location can belong to multiple morphometric classes depending on the scale and a fuzzy membership could be a solution for that problem. This idea could be adapted for our purpose and the class membership for the different aspects, but also for the slopes could be defined by a truth value.

An estimation of the error of the slope classification is very difficult, because it is depending on many factors. Based on the estimated slope error, theoretically all GPS points could be falsely classified. Nonetheless, since the trails are most of the time straight, an error of 20% can be assumed, which is not an overestimation. The steepest drop algorithm is mainly affected in flat areas. Because only about 25% of the slopes are smaller than 5°, only these values could be affected. Therefore, about 5-10% of the values might be wrong. This number might even be smaller, due to the fact that only cardinal directions are considered.

VEGETATION

The error of the vegetation classification is depending on the GPS accuracy as well as the spatial accuracy of the HABITALP vegetation layer. The vegetation classification was provided by the HABITALP project and could not be manipulated. If a buffer of 15 m (5 m error of vegetation classification and 10 m of the average GPS error) is calculated around all points, the following possibilities for a false classification can be calculated:

TABLE 66 : POSSIBLY WRONG CLASSIFIED POINTS.

Class	Total Points	Possible wrong classified	Percentage of possible error
Other	2723	2420	89
Grassland	7235	3739	52
Forrest	8212	3036	37
Residency	3873	3032	78
Total	**22043**	**12227**	**55**

A maximum of 55% of the points might be classified into the wrong vegetation class. Possible errors for the classes *other* and *residency* are very big, which can be explained by the thin shape of these objects in the data layer. If it can be assumed that both position errors, which come from the GPS and vegetation datasets, are isotropic, then the percentage of false classification decrease and it can be assumed that more than 80% of the points were classified into the right class. Yet, the classes *other* and *residency* might still be more influenced by misclassifications than the other two classes.

TEMPERATURE

Assigning the weather information of just one station on the Ofenpass to all GPS points yields a big error. The altitudes alone, ranging from 1500 to 2300, could influence the temperature severely. It can be argued that because the climate in mountainous areas is very complex, it is very difficult or even impossible to model it (Yang & Xiao, 2008). Some models do exist, but they cannot predict the winds, which are important factors for the climate in the Alps. Therefore, it can be contended, that it is more reasonable to see the temperature just as a relative indicator for how high the temperature is compared to other days, than to model it and obtain uncertain values. The error can be estimated with the precondition that only the relative temperature is considered. The classes are on average 13°C wide and per degree Celsius on average almost 700 requests were made. An average error of some degrees Celsius can be assumed, because of the hourly aggregation of the temperature measurement. If this offset, the three class limits, and the requests per temperature are multiplied a maximum of about 20% of the points could have been classified into the wrong class.

6.2.3 INTERACTION OF MODEL AND DATA

The interaction between model and data is mainly depending on the complexity of the processing. These interactions could be estimated for each context variable, which can be seen in Table 67.

TABLE 67 : INTERACTINO OF DATA AND MODEL.

Context factor	Processing complexity	Interaction with Data	Estimated mean ζ of error in %	Estimated σ^2 of error
Trails	Low	Low	0	1
User groups	High	High	10	10
Picnic Areas	Middle	Middle	5	2.5
Relative Time	Middle	Middle	5	2.5
Absolute Time	Low	Low	0	1
Slope	High	High	20	50
Aspect	Middle	High	5	10
Vegetation	Low	Middle	15	5
Temperature	Low	Low	20	5
Average	Middle	Middle	10	14

The estimated errors were discussed in the previous chapters. It is even more difficult to estimate the standard deviation of the Gauss distribution. For the context factors *picnic areas* and *relative time* this estimated error is small, because the complexity of the model and the interaction is only rated middle. Therefore the estimated error is seen as stable. On the other hand the error on the slope can vary a lot, because all values could be classified into the wrong category.

6.2.4 ERROR PROPAGATION

The error in point datasets propagate without a spatial componentand and also the autocorrelation or crosscorrelation does not influence the final result (Heuvelink, 1998). Each Gaussian distribution with the estimated mean errors and estimated standard deviations given was summed up in order to obtain the final distribution of the error. The cumulative probability of the errors and the relative probability for each percent step on a false classification can be seen in Figure 44. The mean probability is at about 8%, while the value of explanation of 0.95 is breached at about 30%. The hightes probabilities for the estimated error are at 3%, while two other peaks are visible at 8 and 18%. With respect to the development and the curvature of both curves an error of less than 25 percent can be assumed, because the growth curve flattens out significantly after the 25% mark.

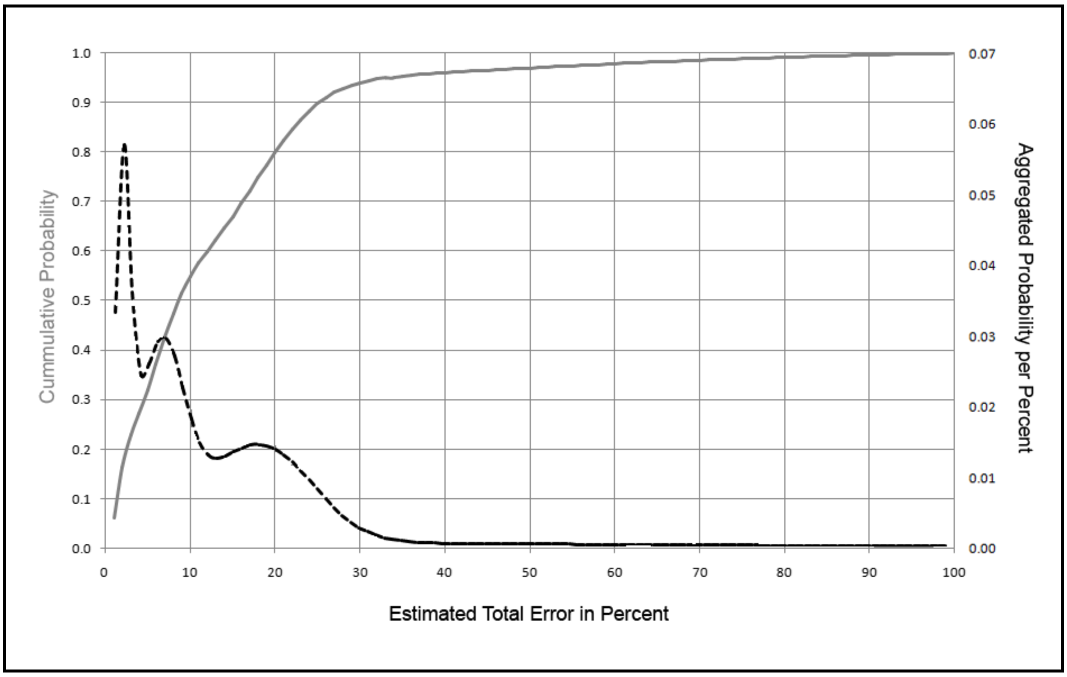

FIGURE 44 : CUMULATIVE PROBABILITY AND AGGREGATED PROBABILITY FOR EACH PERCENT STEP OF THE TOTAL ERROR ON THE CLASSIFICATION.

Yet, how strong is the influence of the classification on the relevance scores? If the classes have the same total quantity and the error is equal in each class, then it has no consequences at all. But it cannot be assumed that the error is equal for each class. The bigger the difference of the frequencies within the classes and the differences of the error within the different classes are, the bigger is the influence on the relevance scores. The more frequent the expected values are, the bigger are the errors on the standardized residuals. E.g. the standardized residuals with an assumed false classification of 25% become greater than 2 with an expected value of 70. The relevance scores are not depending on the size of the expected values. Relevance scores, which are smaller than 0.75 or greater than 1.25 are out of the range of the influence of the estimated calculation errors. If the distribution of the relevance scores in Table 68 is considered, it can be seen that 29% of the scores might be ones, but due to the calculation process they get different relevance scores.

TABLE 68 : DISTRIBUTION OF RELEVANCE SCORES.

Scores	Frequency	Percentage
< 0.75	5500	24
0.75 – 1.25	6646	29
1.25 >	10840	47

6.2.5 Random groups

One goal of this work was to prove a statistical dependency of the request patterns on the selected context factors. In all cases this dependency was significant at a level of significance of 0.05. In order to ensure that a random categorization of the variables could not cause a similar distribution, a new variable with a uniform distribution with four values was generated. Then a Chi^2 test on the independency of the variables was conducted. It could be shown that the information groups and subgroups are independent at a level of significant of 0.05. Only one cell value of the information subgroups reached a standardized residual of over 2, which indicates a general independency of the variables.

The average, empirically evaluated standardized residual for a random distribution with four classes is about 0.64, which is only a bit smaller than the standard deviations of some context factors. On the other hand the influence of the trails is about 3.5 times bigger than a random effect.

The rankings of the influences of every context variable could be shifted because of random effects. This is illustrated in Figure 45. Random effects influence in general all context variables in the same way. If this is assumed, then the ranking does not change.

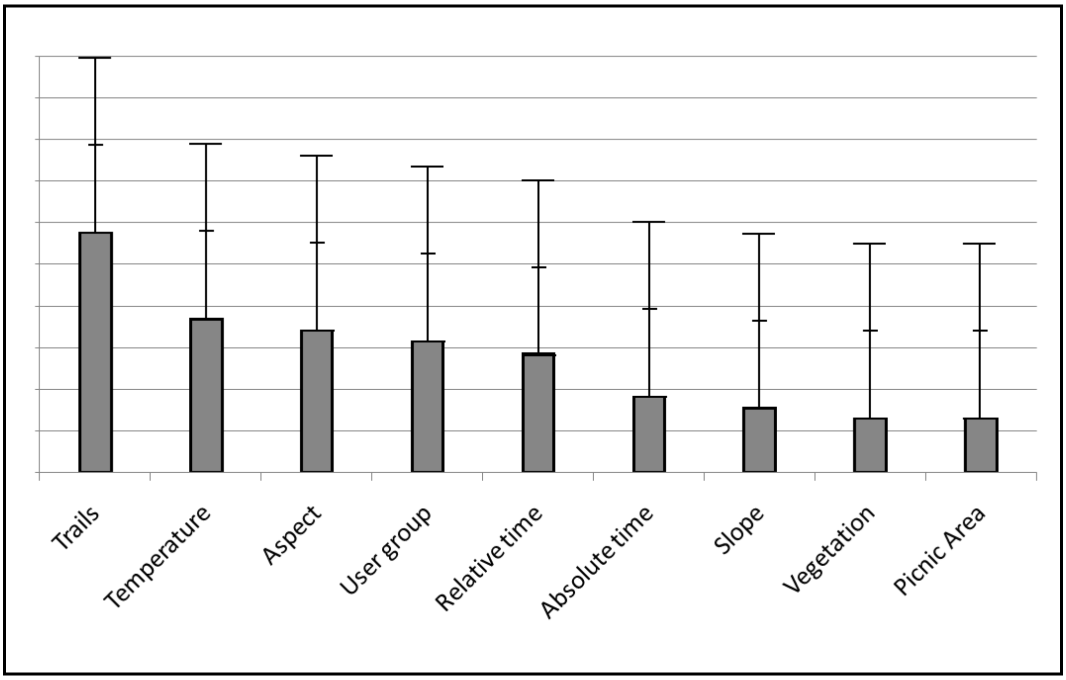

FIGURE 45 : ILLUSTRATION OF THE INFLUENCE OF EVERY CONTEXT VARIABLE WITH RESPECT TO THE UNCERTAINTIES DUE TO RANDOM EFFECTS OF THE CLASSIFICATION.

6.3 DISCUSSION OF RESULTS

Because of the uncertainties, which have been discussed in the previous chapter, only relevance scores and standardizes residuals with high amplitudes can be seen as being the result of influence of the context variables. The values with small amplidues could be a product of calculation errors and random effects.

6.3.1 SPATIAL AUTOCORRELATION

All groups and subgroups of information showed a statistical dependency on all context variables at a level of significance of 0.05. But all relevance scores also also spatially autocorrelate according to Moran's index (Moran, 1950). This is not surprising for those variables, which have a spatial component, such as the *trails* and the *picnic areas*. However, also the relevance scores of the *temperature, user groups, absolute,* and *relative time* show a spatial autocorrelation.

The spatial patterns of the scores, which were extrapolated to the area with an ordinary krigging function, are illustrated in Figure 46. Even though the influence of the *user groups* on the distribution of the trails could be ruled out statistically, they still show patterns on the different trails. A difference between east and west is for instance observable on the trail Val Trupchun. No clear patterns are recognizable for the *time* variables, even though some dark and bright spots are also visible. The trend for the temperature is on the other hand more distinct and the patterns are visible on a larger scale.

A possible interpretation for the distribution of the *user groups* might be that some user types just prefer specific trails, and some trails might be more interesting in specific *temperatures*. The patterns for *relative* and *absolute time* can be explained by the same starting and ending points of different users and the appearance at the same time at specific locations.

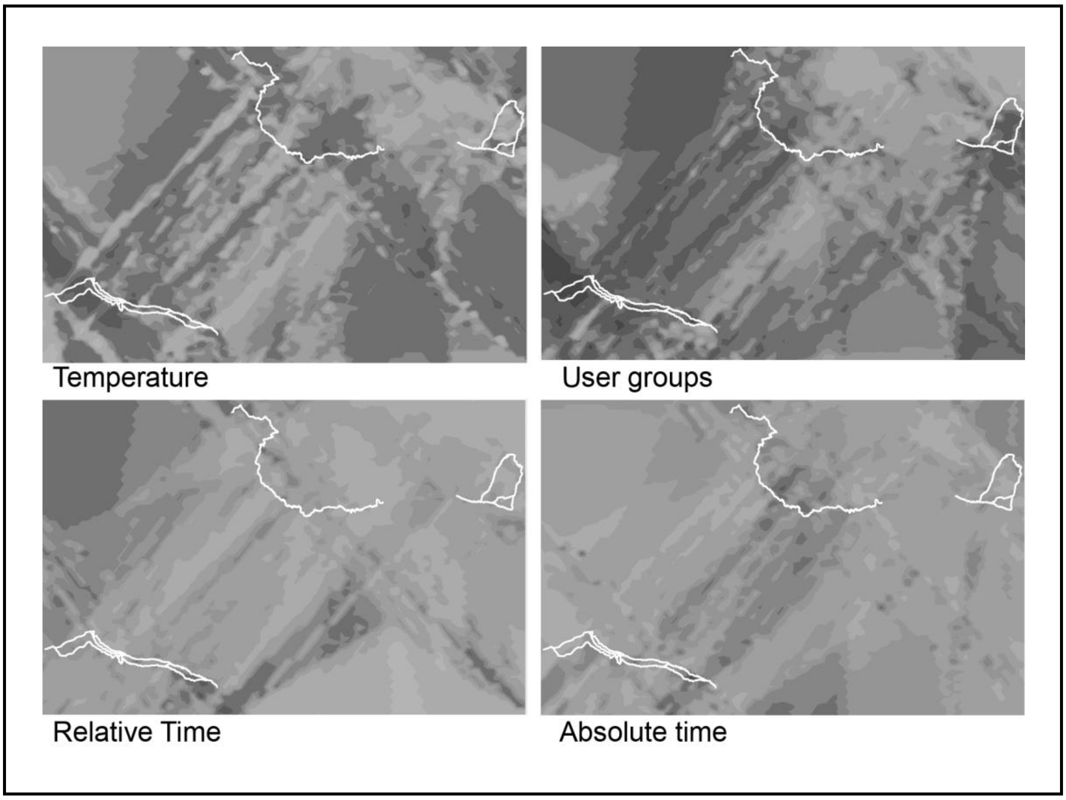

FIGURE 46 : SPATIAL PATTERNS OF NON-SPATIAL CONTEXT VARIABLES. DARK ARE HIGH RELEVANCE VALUES, BRIGTH ARE LOW RELEVANCE VALUES.

6.3.2 TRAILS

The influence of the context variable trail is very strong compared to the other context variables. Trails are defined as meso space. This means that there are regional differences between the trails, but they have also similarities. Location is seen as the most important context variable in the literature. If the spatial differences become greater, then also the requests should differ more. However, besides the differences of requests of content information, which are related to the characteristics of the different trails, especially requests on the surrounding area differ significantly from trail to trail. The information subgroups *"FOI List", "vertical profile"* and *"map overlay"* are strongly influenced by the trails, while the *"news"* and *"search"* functions did not show a dependency at all. 16 of 21 information subgroups have values, for which it can be assumed that they are not produced randomly because of their high absolute value. Processing effects are in general negligible. The preprocessing steps have the highest influences on the result. Yet, the discarded users mostly had only a few requests and therefore might just have an influence on the small subgroups, while the differences in the subgroups showing requests of high frequency would not have changed strongly. The smallest group, which shows high dependencies on the trails, is the *"map overlay"* function. This function was only requested 222 times in total and on the trail Margunet it was only 42 times. Small changes in the filtering process could have a strong influence on this result. The differences of the requests on the surrounding area might be caused by the system architecture, or maybe more by the available FOI. If more of these points are available, then the user has also more material to request. The high differences of the *"route info"* stand out after ignoring the subgroups, which might be caused by other factors than system context. This is not surprising, because this specific information is directly related to the trails. Another function, which is also strongly related to the trails, is the *"vertical profile".* This relation is also not surprising, because the trails show differences in their vertical profile as it is illustrated in Figure 47.

It can be seen that some of the trails have a supposably more demanding vertical profile than other trails, which might also influence the request behavior. But the figure is deceptive, because the natures of the trails are very different and hardly comparable. A hiker on the trail Val Trupchun for instance has to march back after reaching the highest point of the trail. In this case the vertical profile would have to be mirrored in horizontal direction.

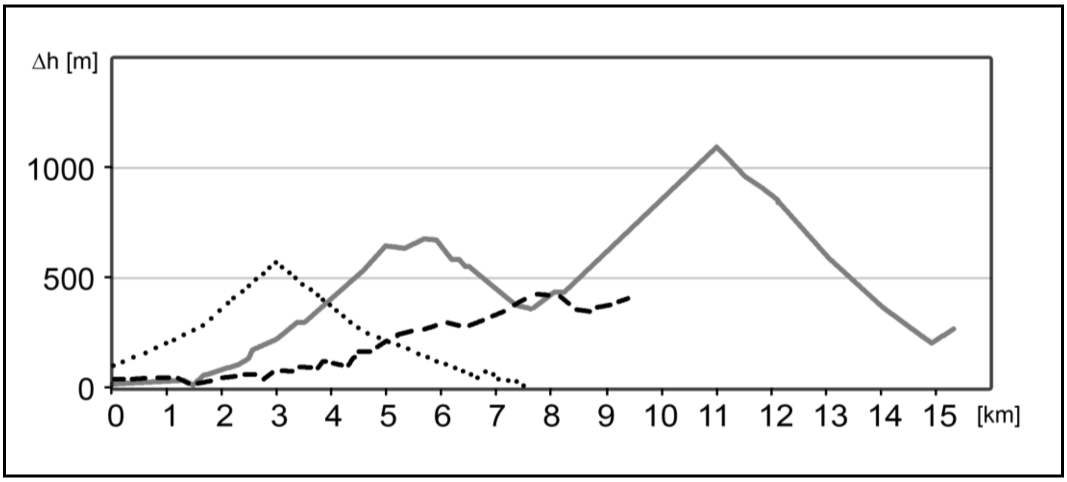

FIGURE 47 : VERTICAL PROFILES OF THE TRAILS MARGUNET (DOTTED), CHAMANNA CLUOZZA & MURTER (SOLID GREY) AND VAL TRUPCHUN (DASHED) BASED ON ROBIN (2009), P. 20, 44, 47, AND 88.

6.3.3 USER GROUPS

The assumed error in the classification of users is very large compared to the other context variable. Therefore the results have to be considered with caution. Errors in the classification could have a special impact on the two non-stopping user groups, because they have in general a smaller request rate. The *"map overlay"* function for instance has in the *no data* class been requested 1.7 times more than on average. This function could have been strongly influenced by only one falsely classified user, because the estimated value is only 25. If the expected values are small, e.g. in the *"content"* information subgroups, then a single user could have had a great impact on the result. Considering these thoughts, only the information subgroups *"FOI list"*, *"get around"* and *"route info"* can be seen as mainly defined by the user groups. The *"FOI list"* shows an unexpected trend, because the *long stoppers* and the *no data* group show a high interest in this particular information. The exact opposite is observable

for the *"get around"* function. If both functions were aggregated, then no significant differences between the user groups can be observed anymore. Therefore the only outstanding information subgroup is the *"route information",* which is far more requested by the *no data* group. These hikers could in general be more interested in information, which is directly related to hiking, while other contents are less interesting.
In contrast to the results of the quality of requested information, there are statistical firm differences between the quantities of requested information between the user groups. The longer a hiker stops at a picnic area, the more information is requested, which is not surprising, considering that at picnic areas much more information is requested than outside picnic areas.
After incorporating a great estimated classification error, no difference between the user groups, which are originated by the user groups, can be assumed. Therefore the classification cannot be seen as a success.

6.3.4 Picnic Areas

For the picnic areas only a few information subgroups have outstanding values. The *"route information",* the *"search"* function, the *"map page"* and the *"bookmark"* function show a high dependency on this context variable. Random effects and calculation errors could be the cause for the differences in all other information subgroups. While the functions *"search"* and *"bookmark"* have a higher relevance inside the picnic areas, the *"route information"* and the *"map page"* are more relevant outside the picnic areas. If the information groups are considered, because some of the subgroups are just too small for firm analysis, then it becomes clear that more *demanding* functions are more requested inside the picnic areas, while information on the hike is important, while the hiker is in motion.

6.3.5 Topography

Because of the high possible classification error only two trends can be recognized for the slope, and not even those are certain. Regarding the quantity of requests it can be stated that the steeper the slope, the fewer information is requested. But this trend becomes visible only in a very steep terrain. Regarding the quality of requests it can be stated, that in steep slopes, users have a higher interest in information, which is directly related to hiking, while more demanding functions such as the *"bookmark"* function were requested less.
With respect to aspect it can be stated that at south-eastern expositions more requests were made and especially the *"map page"* and the *"get around"* function are depending on the aspect. Yet, the result might be dependent on other factors, because the relevance scores of the aspect show a high spatial autocorrelation, and also the scale in which the aspect is calculated has an influence on the classification and therefore also on the results.

6.3.6 Time Variables

The error for the absolute time is insignificant and therefore the values can be seen as absolute. But the categorization into *morning, noon* and *afternoon* can be discussed. Maybe some other class limits would produce clearer results. Because most of the trends in the absolute time can be explained by the distribution in the relative time, it can be assumed that also other class limits would not have a higher explanatory value than the relative time. On the other hand it could be shown that the absolute time has an influence on the quantity of requests. But the additional quantity of requests at noon may also be explained by the additional requests at the picnic areas. Therefore the absolute time is not suitable for our purpose, and only the results of the relative time are considered. The relative time on the other hand shows an almost constant general request rate and the most distinct request patterns in terms of quality. Only at the start and significantly more information is requested. Slightly more information is also requested at the very end of a hike. There is a general trend that *"info around"* is more requested at the end of a hike, similar to the requests of *"special functions".* In contrast to that more *"info on the device"* and *"info on the trails"* are requested at the start of a hike, while the requests on content remain constant. The additional requests of *"info on the device"* can be explained by the need for tutorial functions at the start, while the use of *"special functions"* becomes more likely towards the end, because of the additional knowledge of users about the device. Also the *"info on the trail"* is more relevant at the start, which can be explained, by the need of the users to inform themselves about the trail on which they are going to hike.

6.3.7 Vegetation

For the context variable vegetation only the vegetation classes *grassland* and *forest* are considered, because of the high possible error in the vegetation classes *residency* and *other.* For these two vegetation classes only two information subgroups are remarkable: the *"get around"* function and the content on *"butterflies".* The get around function is more relevant in *"grassland",* while the opposite is the case for the requests on *"butterflies".* These requests could be explained by a sensitivity of these insects on the vegetation, while the *"get around"* function could be because the hikers have a broader sight and therefore might be interested in what is around them. In terms of quantity, more information was requested in the *grassland.* The significant differences can be explained by the intervisibility of the objects. Because of that, the vegetation might not have a direct influence and the differences might be originated by the intervisibility.

6.3.8 WEATHER

While it was raining no quantitative and qualitative analysis of the data is possible, because at these few times almost no requests were made. Therefore it can be said that the device is never used in *bad* weather, or at least that no users of the WebPark^{SNP} system are hiking during those times.

The temperature is difficult to handle. First it is almost impossible to calculate expected values for the assessment of the quantity of information. But it is more likely that the users requested more information in warmer than in colder weather. Second the temperature was not modeled, which can lead to great errors, because in mountainous areas the temperature can change significantly in comparable areas (Yang & Xiao, 2008). Therefore the temperature can only be seen as a trend of the weather condition. The classification is highly depending on the class limits, and falsely classified points can lead to big distortions, because there are great differences between the total frequencies of the temperature classes. The coldest temperature class is almost not usable, because the expected values for some information subgroups are smaller than 10. The most significant difference between the temperatures is the higher relevance of the *"map page"* in a hot environment. All other distributions could be caused by a wrong classification, by the non-existing temperature-model, or by random effects. Some of the other information subgroups have high standardized residuals and/or relevance scores, but do not follow a comprehensible pattern, and therefore unknown effects are possible causes.

7 Conclusions

7.1 Achievements

At the beginning of this thesis three general research objectives were defined, which will now be evaluated. The request frequency, in general, is mainly dependent on whether the users are in- or outside the picnic areas. Normalized by the passed time almost 2.5 times more requests were made inside the picnic areas than outside. The space on a larger scale, namely on the level of the trails, has no statistical influence on how many requests were made. In general, all three trails had, with respect to the individuals, the same request rate. Interestingly, the time variables had only a small influence on the general request rate. In the distribution of the relative time only the peaks at the beginning and at the end stand out and the request rate seems to be constant during the rest of the time.

Looking at the distributions of the specific requests, it could be shown that they are statistically depending on all presented context variables, be it on the level of information groups or information subgroups. It could also be shown in which way the several request patterns are depending on the context variables. With the introduced model it is possible to quantitatively specify the relevance of a certain variable under given contextual characteristics. But even though the introduced model allows for calculating a relevance score in a ratio scale, it might be more stable to measure the information pieces in an ordinal scale, in order to deal with the uncertainties of the calculation in a better way.

The influence of the introduced context variables is not constant. Some of them have a greater influence on the request patterns of the users' requests. It was possible to rank the context variables according to their influence. It was also possible to assess which information subgroups are influenced more, and which are influenced less by the context.

7.2 Insights

Eisenhut et al. (2008) stated that the request rate of the users decrease with their ongoing hike on the trails. Their analysis was based on a sub-sample of 30 users and a method which analyzed the frequency of requests depending on the distance to the start. In this study another method was chosen on a sub-sample of 201 users, which showed that only at the very beginning of a hiking tour significantly more information is requested and no trend towards fewer requests at the end of a hiking tour can be observed. Eisenhut et al. (2008, p.2) also claimed that the *"most popular time of use was the hour before noon",* which is true if only the absolute values are considered. However, this trend is weakened if the requests are normalized by the users. The hour before noon is the most popular time for hiking on the trails, and it is also predestinated to hold a picnic. Because of the combination of these two factors the time before noon becomes the period with the highest request frequency.

With the knowledge about the difficulties of processing certain context variables, it can be concluded that location on a bigger scale and time are the most influential parameters. The *trails,* which in our case are rated as *meso space,* are by far the most influential factor. *Relative time* is in second place, because the other variables, which are rated higher in terms of influence, show a greater dependency on the correct calculation of the values. Based on this observation it can be argued that the easier the calculation and the model, the more unambiguous the yielded results will be. If the calculation is uncertain, all results can get biased.

The context characteristics in this thesis were classified with crisp set methods. This is reasonable for the *trails,* because they can be spatially differentiated. But for all other context variables a fuzzy set classification could be a better choice. The *user groups* for instance could be classified with this method in order to avoid problems with the crisp class limits in the time axis. No changes on the general context model would have to take place, because the model can deal with fuzzy methods.

7.3 LIMITATIONS

The results in this thesis are masking out the *system context* completely. Also, other preconditions than the constant utility of the requested information have to be assumed, because no information is available on this matter. Therefore, the introduced model and the subsequent results are limited. It cannot be declared which information is more important than other information. But some trends are observable under the given system parameters of the WebPark^{SNP} system of 2007.

Mainly, three factors limit the accurateness of the results. The accuracy of the measurement and the dataset is the first. For all context factors with a spatial component especially the GPS accuracy reduced the preciseness of the results, because the attachment of other data layers was influenced.

The second factor is the processing of the values. The more complex the processing is, the less precise the results become. Especially the calculation of the *slope* is influenced by the calculation process. It could be shown that theoretically all points could have been classified into the wrong class. Even though this extreme outcome of the processing is unlikely, it shows that a great effort would have to be made in order to obtain more precise results on the dependency of the request patterns on topographic context variables. Another variable, which is depending highly on processing, is *temperature.* This variable was not modeled at all, because the temperature on the Ofenpass was considered as a trend of how *warm* or *cold* the weather is at the time of the hike. Maybe approaches which include a temperature model would result in more distinct trends and more certain results. The spatial autocorrelation of each variable indicates that some of the results might just be depending on unknown variables and not the presented contextual variables.

The crisp classification of the characteristics of each context is the third limitation. A crisp classification for all context variables but the *trails* is not suitable. The crisp method reaches its functional limit especially for the context variable *user groups.* Small changes in the class limits in the time axis can result in great changes in the classification. Therefore, a fuzzy set method could be more appropriate.

The posed research question on the possibility of classifying the users by their behavior at the picnic areas cannot be answered with a *yes* or a *no.* Even though the results between the requests of the different user groups are statistically firm with the Chi^2 test on the independency of the variables, they are also uncertain. Thus it was possible to classify the users into different groups, which also had different request patterns, but the results are also uncertain, because the groups contain only a few individuals and the differences in the groups are big. A similar answer can be given to most of the context variables. Yes, it is possible to model the context with a statistical significant dependency on the different characteristic of the context variable. But in most of the cases it has to be admitted that the results are too uncertain and are depending on too many other factors than only the context variable.

The requests for some information subgroups have been too few in order to make statistically firm statements, because classification errors and random effects could have a great influence. And even though almost 23'000 request and 201 users seem to be enough, they still limit the possibilities for a quantitative assessment and it would be better if a multiple of these numbers could be analyzed. Considering these limitations only the *trails* and maybe the *relative time* could have been modeled and rated.

The limitations can be summed up and are illustrated in Figure 48. As a first step the physical world has to be categorized into context variables. But is the selection of the variables suitable? These context variables have to be measured with sensors. But are these sensors suitable to measure the specific context variable? After that the raw data has to be processed (e.g. the slope function). Are the processing methods correct? After that the characteristics of each context variable have to be categorized (e.g. with time limits for the user groups). Are these categorizations appropriate? If the number of requests in general, but also for each subgroup, is frequent enough, the model can be considered suitable in order to calculate the relevance scores. But if this is not the case it must be asked if different approaches would not be more appropriate.

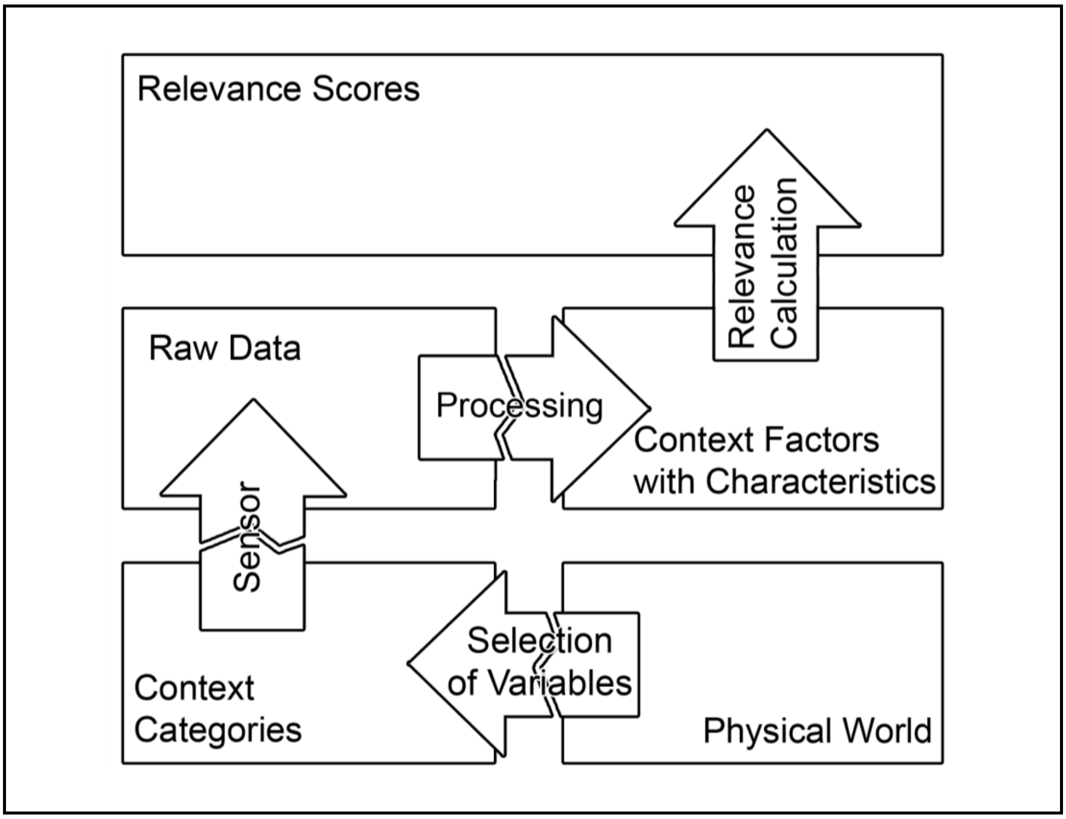

FIGURE 48 : UNCERTAINTIES IN THE PROCESSING OF THE RELEVANCE SCORES BASED ON BECKER & NICKLAS (2004, P.2).

7.4 OPEN PROBLEMS

Our model was suitable for some of the considered context variables. Because the context is changing continuously (Kaasinen, 2002) it might be more suitable to classify the context characteristics according to the fuzzy set theory. Some problems, which originate in the classification, could by reduced by doing so. But still problems on how to process the context variables would remain. What is the best way to define the slope and the aspect for an environment such as the WebPark^{SNP}? How could the temperature be modeled in a mountainous area such as the Swiss Alps? Is it even possible with the given quality of the data to assess all the introduced context variables quantitatively, because the calculation of the characteristics of the context variables is highly depending on the accuracy of the involved data layers? Even though the environment in the SNP had some limitations, it also offered some advantages, such as the limited number of hiking routes. Therefore, the question how the presented concept could be adapted and introduced to other environments such as city guide is unanswered. In such an environment it could be difficult to define similar situations, because far more daily routines are possible for a city tourist. And there are still open problems for the device in the SNP. How can the findings of this study be implemented? How can the device be adapted in terms of *context awareness?* And because the device already had implemented "*location" awareness*, how did this affect the request patterns? Could the request patterns of the *"FOI List"* and the *"get around"* function just be depending on how many *features of interest* are in the area? The distinct spatial autocorrelation of the relevance scores might lead to such a conclusion.

7.5 OUTLOOK

The findings of this study could help to update the system in the SNP. Possibilities might be an adaption of the relevance of the information pieces depending on which trail they are located. This is already the case to some degree, because FOI functions already exist, and the features are presented according to their location. Another simple implementation could be the incorporation of relative times. For instance it could be shown that the bookmark functions are more used at the end of a hiking tour. If a starting time could be set automatically or by user imputes, then the system could be adapted. The strongest trend in terms of total amount of information was observable for the picnic areas. Therefore, an adaption which deals with the proximity to the picnic areas can be suggested.

Because the user classification was not successful, a recommender system similar to the introduced system of Ricci and Nguyen (2007) could be used to model the context of the user. Analyzing the context variables based on the fuzzy set theory could help to get more significant results. Fuzzy set theory would pay more respect to the continuous changing context as proposed by Kaasinen (2002). If the relevance could be calculated for every information piece in every situation then it is possible to extrapolate the relevance scored to a probability map, which was introduced by Rapers (2007). Such maps could help to adapt the system to more complex context variables.

After the implementation of scores, it would be interesting to go on with research on the presentation of information according to their relevance. A variable, which represents the quantity of requested information was introduced. If not only the relation between the information pieces based on the relevance scores but also the quantity of information is included, then it would be possible to adapt the display not only with respect to usability criteria, but also with respect to how many different information pieces are needed.

A quantitative assessment of context is very difficult and this research faces many problems. But with the introduced model it could be possible to enhance mobile devices such as the WebPark^{SNP} system. The first step could be to use the introduced model on a different dataset and maybe already with a fuzzy classification of the context characteristics.

8 BIBLIOGRAPHY

Aalders, H. G. (1996). Quality metric for GIS. In M. J. Kraak, & M. Molenaar (Ed.), *Proc. 7th International Symposium on Spatial Data Handling*, (pp. 5B1-5b10). Delft.

Andogah, G., & Bouma, G. (2008). Relevance Measures Using Geographic Scopes and Types. In *Advances in Multilingual and Multimodal Information Retrieval* (pp. 794-801). Berlin / Heidelberg: Springer.

Andrade, L., & Silva, M. (2006). Relevance Ranking for Geographic IR. *Workshop on Geographical Information Retrieval, SIGIR'06.*

Bauch, K., & Seitlinger, G. (2006). Luftbildinterpretation über das Gebiet des Nationalparks Hohe Tauern - im Rahmen des EU-Projektes HABITALP. *Raumplanung Aktuell* , 47-49.

Becker, C., & Nicklas, D. (2004). Where do spatial context-models end and where do ontologies start? A proposal of a combined approach. *Workshop on Advanced Context Modelling, Reasoning and Management under the 6th Int. Conf. on Ubiquitous Computing (UbiComp 2004)* (pp. 1-6). Nottingham: Reasoning and Management.

Bidgoli, H. (2004). *The Internet Encyclopedia* (Vol. Volume 3). New Jersey: Wiley & Sons.

Bley, D., & Haller, R. (2006). Checking the spatial accuracy of class boundaries with a varying. *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences* (pp. 249-257). Lisboa: Instituto Geográfico Português.

Brézillon, P. (1999). Context in problem solving: A survey. *The Knowledge Engineering Review , 14(1)*, 1-24.

Brimicombe, A., & Li, Y. (2006). Mobile Space-Time Envelopes for Location-Based. *Transactions in GIS , 10*, 5-23.

Burrough, P. A., & McDonnell, R. A. (2005). *Principles of Geographical Information Systems.* Oxford: Oxford University Press.

Chen, G., & Kotz, D. A. (2000). *Survey of Context-Aware Mobile Computing Research, Technical Report, Dept. of Computer Science.* Darmouth College.

DeCesare, N. J.; Squires, J. R.; Kolbe, J. A. (2005). Effect of forest canopy on GPS-based. *Wildlife Society Bulletin , 33(3)*, 935-941.

D'Eon, R. G. (2003). Effects of a stationary GPS fix-rate bias on habitat selection analyses. *Journal of Wildlife Management , Vol. 67* (No. 4), 858-863.

Dey, A. K. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing, 5(1)* , 4-7.

Dey, A. K., & Abowd, G. D. (2000). Towards a better understanding of context and. *Proceedings of the What, Who, Where, When, and How of Context-AwarenessWorkshop.* New York: ACM.

Dey, A. K., Salber, D., Abowd, G. D., & Futakawa, M. (1999). *An Architecture to Support Context-Aware Applications.* Atlanta: Graphics, Visualization, and Usability Center, Georgia Tech.

Dias, E. (2004). *WebPark user Tests: WZ Case study.* WebPark Consortium.

Dias, E., Beinat, E., Rhin, C., Haller, R., & Scholten, H. (2004). Adding Value and Improving Processes Using Location-Based Services in Protected Areas. *Research on Computing Science , 11 (special edition on e-Environment)*, 291–302.

Dias, E., Edwardes, A. J., & Purves, R. S. (2008). Analysing and aggregating visitor tracks in a protected area. In S. A. al, *Quality Aspects in Spatial Data Mining* (pp. 1-29). Oxfordshire.

Dransch, D. (2005). Activity and Context - A Conceptual Framework for Mobile Geoservices. In L. Meng, A. Zipf, & T. Reichenbacher, *Map-based Mobile Services* (pp. 31-42). Berlin Heidelberg: Springer.

Duncan, D. B. (1955). Multiple range and multiple F tests. *Biometrics , 11*, 1-42.

Dussault, C. R., Courtois, J. P., & Huot, J. (1999). Evaluation of GPS telemetry collar performance for habitat. *Wildlife Society Bulletin , 27*, 965-972.

Edwardes, A. J. (2007). Re-placing Location: Geographic Perspectives in Location Based Services. *Ph.D Thesis* . University of Zurich.

Eisenhut, A., Haller, R., & Raper, J. (2008). *How does topography influence the use of the mobile guide WebParkSNP in the Swiss National Park?* Montecatini: The Fourth International Conference on Monitoring and Management of Visitor Flows in Recreational and Protected Areas.

Fisher, P., Wood, J., & Cheng, T. (2004). Where is Helvellyn? Fuzziness of Multiscale. *Transactions Institute of British Geographers , 29(1)*, 106-128.

Gellersen, H.-W. (2003). Embedded interactive systems: toward everyday environments as the interface. In G. Szwillus, & Z. Ziegler, *Mensch & Computer 2003 – Interaktion in Bewegung, Berichte des German Chapter of the ACM* (Vol. Band 57, pp. 25-28).

Greisdorf, H. (2000). Relevance: An Interdisciplinary and Information Science Perspective. *Informing Science , Volume 3* (No. 2), 67-71.

Haller, R. M., & Imfeld, S. (2007). Aassesment of Height Accuracy of DEM for Species Habitat Analysis and Modelling. *Spatial Data Quality.* Enschede: ITC.

Haller, R., & Eisenhut, A. (2008). Was fragen Wanderer den digitalen Wanderfürher im Webpark? In R. Eder, & A. Arnberger, *Auf den Pfaden von Natur und Kultur - Wodurch werden Lehrpfade, Themen- und Erlebniswege zu attraktiven Destinationen?* Wien: Institut für Landschaftsentwicklung, Erholungs- und Naturschutzplanung.

Haller, R., Burghardt, D., & Weibel, R. (2005). WebPark – neue Wege mit mobilen Lösungen in Tourismusgebieten. *Géomatique Suisse , 5*, 242 - 245.

Heuvelink, G. (1998). *Error Propagation in Environmental Modelling.* Padstow, UK: T.J. International Ltd.

Hirakawa, M., & Hewagamage, K. P. (2001). Situated Computing: A paradigm for the Mobile User_interaction with Multimedia Sources. *Annals of Software Engineering(12), no.1* , 213-239.

Horn, B. P. (1981). Hill shading and the reflectance map. *Proceedings of IEEEE , (69)1*, 14-47.

Hulbert, I. A.R.; French, J. (2001). The accuracy of GPS for wildlife telemetry and habitat mapping. *Journal of Applied Ecology , 38*, 869–878.

ISO. (1999). ISO 13407, Human-Centered Design for Interactive Systems. Geneva, Switzerland: International Organization for Standardization.

Jameson, A. (2001). Modelling both the Context and the User. *Personal and Ubiquitous Computing , 5*, 29–33.

Jones, C. B., & Purves, R. S. (2008). Geographical Information Retrieval. *International Journal of Geographical Information Science , 22* (3), 219-228.

Jones, C. B., Alani, H., & Tudhope, D. (2001). Geographical Information Retrieval with Ontologies of Place. In D. R. Montello, *Spatial Information Theory: Foundation of*

Geographic Information Science. International Conference, COSIT 2001. Morro Bay, CA, USA (pp. 322-335). New York: Springer.

Kaasinen, E. (2002). User needs for Location-Aware Mobile Services. *Third Wireless World Conference: The Social Shaping of Mobile Futures.* Guildford: Surrey University.
Krug, K., Abderhalden, W., & Haller, R. (2003). User needs for location-based services in protected areas; case study Swiss National Park. *Information Technology & Tourism , 5(4)*, 235–242.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, & H. Hotelling, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (pp. 278-292). Stanford University Press.

Lotz, A. (2006). *Alpine Habitat Diversity HABITALP.* EU Community Initiative INTERREG III B Alpine Space Programme.

McCarthy, J., & Sasa, B. (1997). Formalizing Context (Expanded Notes). *Computing Natural Language* , 13-50.

Meng, L., & Reichenbacher, T. (2005). Map-based mobile services. In L. Meng, T. Reichenbacher, & A. Zipf, *Map-based mobile services* (pp. 1-10). Berlin, Heidelberg, New York: Springer.

Moran, P. A. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika , 37*, 17-23.

Mountain, D., & MacFarlane, A. (2007). Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals. *Journal of Information Science* , 515-530.

Nivala, A.-M., & Sarjakoski, L. T. (2003b). An Approach to Intelligent Maps: Context Awareness. *Proceedings of the Workshop W1 "HCI in Mobile Guides 2003". In conjunction with Mobile HCI'03*, (pp. 45-50). Udine, Italy.

Nivala, A.-M., & Sarjakoski, L. T. (2003a). Need for context-aware topographic maps in mobile devices. *Proceedings of ScanGIS 2003*, (pp. 15-29). Espoo.

Nivala, A.-M., Sarjakoski, L. T., Jakobsson, A., & Kaasinen, E. (2003). Usability Evaluation of Topographic Maps in Mobile Devices. *Proceedings of the 21st Internationl Cartographic Conference (ICC)* (pp. 1903-1913). Durban, South Africa: Cartographic Renaissance.

Raper, J. F., Dykes, J. D., Wood, J., Mountain, D., Krause, A., & Rhind, D. (2002). A framework for evaluating geographical information. *Journal of Information Science, Vol. 28 No. 1* , 39-50.

Raper, J. (2007). Geographic relevance. *Journal of Documentation , Vol. 63* (No. 6), 836-852.

Reichenbacher, T. (2009). Geographic relevance in mobile services. *Proceedings of the 2nd International Workshop on Location and the Web. Vol. 370*, pp. 1-4. Boston: ACM.

Reichenbacher, T. (2004). Mobile Cartography - Adaptive Visualisation of Geographic Information on Mobile Devices. München: Verlag Dr. Hut.

Reichenbacher, T. (2007). The concept of relevance in mobile maps. In G. Gartner, W. Cartwright, & M. P. Peterson, *Location Based Services and TeleCartography.* Berlin, Heidelberg: Springer .

Reichenbacher, T. (2001). The World in Your Pocket - Towards a Mobile Cartography. *Proceedings of the 20th Internation Cartographic Conferecence*, (pp. 2514-2521). Beijing, China.

Ricci, F., & Nguyen, Q. N. (2007). Acquiring and Revising Preferences in a Critique-Based Mobile Recommender System. *Intelligent Systems, IEEE , 3* (22), 22-29.

Robin, K. (2009). *Wanderführer durch den Schweizerischen Nationalpark* (3. Auflage ed.). Basel: LAC AG.

Saracevic, T. (1996, October 14-17). Relevance reconsidered. pp. 201-218.

Sarjakoski, L. T., & Nivala, A.-M. (2005). Adaptation to Context - A way to Improve the Usability of Mobile Maps. In L. Meng, A. Zipf, & T. Reichenbacher, *Map-based Mobile Services, Theories, Methods and Implementations* (pp. 107-123). Berlin Heidelberg New: Springer.

Sarjakoski, L. T., & Nivala, A.-M. (2003). Context-aware maps in mobile devices. In A. Salovaara, A. Kuoppala, & M. Nieminen, *Perspectives on intelligent user interfaces* (pp. 112-133). Espoo: Helsinki University of Technology Software Businees and Engineering Institute, Technical Report 1.

Schilit, B., Adams, N., & Want, R. (1994, December). Context-aware computing applications. *Proceesings of IEEE Workshop on MObile Computing Systems and Applications* , pp. 85-90.

Schmidt, A., & Gellersen, H.-W. (2001). *Modell, Architektur und Plattform für Informationssysteme mit Kontextbezug.* Berlin / Heidelberg: Springer.

Schnelle, D. (2007). Context Aware Voice User Interfaces for Workflow Support. Technischen Universität Darmstadt.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition.* Cambridge: Harvard University Press.

Struss, H. (2004). Mobilität wird Alltag. *InformationWeek , 09/10*.

Swisstopo. (2005). DHM25 - Das digitale Höhenmodell der Schweiz. Wabern: Bundesamt für Landestopografie.

Thompson, M. K. (2004). *Quality evaluation procedure for the quantitative horizontal positional accuracy.* Technical Report.

Tobler, W. (1970). A computer movie simulating urban growth in the. *Economic Geography , 46* (2), 234-240.

Tucker, C. (2004). *Writing Geoprocessing Scripts With ArcGIS.* Redlands, California: ESRI.

Unwin, D. (1981). *Introductory Spatial Analysis.* London and New York: Methuen.

van Setten, M., Pokraev, S., & Koolwaaij, J. (2004). Context-Aware Recommendations in the Mobile Tourist Application COMPASS. In W. Nejdl, & P. De Bra, *Adaptive Hypermedia and Adaptive Web-Based Systems* (pp. 235-244). Eindhoven: Springer.

WebPark. (2001). WebPark: Geographically relevant information for mobile users in protected areas - Annex 1 Description of Work. *Proposal IST-2000 31041* . WebPark Consortium.

Weidmann, H.-J. (1995). *Tagfalter: beobachten, bestimmen.* Augsburg : Naturbuch-Verlag.

Wilson, D., & Sperber, D. (2004). Relevance Theory. In G. Ward, & L. Horn, *Handbook of Pragmatics* (pp. 607-632). Oxford: Blackwell.

Wing, M. G., Eklund, A., & Kellogg, L. D. (2005). Consumer-Grade Global Positioning System (GPS) Accuracy and Reliability. *Journal of Forestry , 103* (4), 169-173.

Yang, X., & Xiao, C. (2008). Terrain-based Revision of an Air Temperature Model in Mountain Areas. In Q. Zhou, B. Lees, & G. Tang, *Advances in Digital Terrain Analysis* (pp. 425-442). Berlin, Heidelberg: Springer.

Zhou, Q., & Liu, X. (2008). Assessing Uncertainties in Derived Slope and Aspect from a Grid DEM. In Q. Zhou, B. Lees, & G. Tang, *Advances in Digital Terrain Analysis* (pp. 279-306). Berlin, Heidelberg: Springer.

9 ANNEX

Slopescript

```
matA = -9999*ones(2335,1908);

slopemat = -9999*ones(2335,1908);

for x=1:2333,

        for y=1:1906,

            matA(x+1,y+1) = dhm25(x,y);

        end

end

for x=2:2334,

        for y=2:1907,

            Z = matA(x,y);

            A = matA(x-1,y-1);

            B = matA(x-1,y);

            C = matA(x-1,y+1);

            D = matA(x,y-1);

            E = matA(x,y+1);

            F = matA(x+1,y-1);

            G = matA(x+1,y);

            H = matA(x+1,y+1);

            if A ~= -9999

                cA = 1;

                dA = (A - Z)/sqrt(2);

            else

                cA = 0;
```

```
        dA = 0;
    end
    if B ~= -9999
        cB = 2;
        dB = B - Z;
    else
        cB = 0;
        dB = 0;
    end
    if C ~= -9999
        cC = 1;
        dC = (C - Z)/sqrt(2);
    else
        cC = 0;
        dC = 0;
    end
    if D ~= -9999
        cD = 2;
        dD = D - Z;
    else
        cD = 0;
        dD = 0;
    end
    if E ~= -9999
        cE = 2;
        dE = E - Z;
```

```
        else
            cE = 0;
            dE = 0;
        end
        if F ~= -9999
            cF = 1;
            dF = (F - Z)/sqrt(2);
        else
            cF = 0;
            dF = 0;
        end
        if G ~= -9999
            cG = 2;
            dG = G - Z;
        else
            cG = 0;
            dG = 0;
        end
        if H ~= -9999
            cH = 1;
            dH = (H - Z)/sqrt(2);
        else
            cH = 0;
            dH = 0;
        end
        qT = (cF+cG+cH);
```

```
avgT = (dF+2*dG+dH)/qT;

%AVG-Buttom

qB = (cA+cB+cC);

avgB = (dA+2*dB+dC)/qB;

%AVG-Left

qL =(cA+cD+cF);

avgL = (dA+2*dD+dF)/qL;

%AVG-Right

qR = (cC+cE+cH);

avgR = (dC+2*dE+dH)/qR;

if (qT == 0) && (qB == 0)

    s_vertical = -9999;

elseif (qT == 0) && (qB > 0)

    s_vertical = atan(avgB/25);

elseif (qB == 0) && (qT >0)

    s_vertical = atan(avgT/25);

else

    s_vertical = atan((avgT-avgB)/50);

end


if (qL == 0) && (qR == 0)

    s_horizontal = -9999;

elseif (qL == 0) && (qR > 0)

    s_horizontal = atan(avgR/25);

elseif (qR == 0) && (qL >0)

    s_horizontal = atan(avgL/25);
```

```
            else
                s_horizontal = atan((avgL-avgR)/50);
            end
      if (qL == 0) && (qR == 0) && (qB == 0) && (qT == 0)
            slopemat(x,y) = -9999;
      elseif Z == -9999
            slopemat(x,y) = -9999;
      elseif (s_horizontal == -9999) && (Z ~= -9999)
            slopemat(x,y) = sqrt((s_vertical)^2)/pi*180;
      elseif (s_vertical == -9999) && (Z ~= -9999)
            slopemat(x,y) = sqrt((s_horizontal)^2)/pi*180;
      else
            slopemat(x,y) =
         sqrt(((s_horizontal+s_vertical)/2)^2)/pi*180;
            end
end
end
output = ones(2333,1906);
for x=1:2333,
    for y=1:1906,
        output(x,y) = slopemat(x+1,y+1);
    end
```