

RELATING MOVEMENT TO GEOGRAPHIC CONTEXT – EFFECTS  
OF PREPROCESSING, RELATION METHODS AND SCALE

---

**Dissertation**

zur

Erlangung der naturwissenschaftlichen Doktorwürde  
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

CHRISTIAN GSCHWEND

von

Altstätten SG

**Promotionskomitee**

Prof. Dr. Robert Weibel (Vorsitz)  
Prof. Dr. Ross Purves  
Prof. Dr. Nico Van de Weghe  
PD Dr. Patrick Laube (Leitung der Dissertation)

Zürich, 2015



# Summary

Movement (in space and time) is studied by different research fields, such as GIScience, movement ecology, or urban planning in order to gain insights into behavior of animals or people. Recent technological advancements allow collecting huge volumes of movement data in the form of GPS records. Movement analysis aims at developing methods with a strong focus on detecting movement patterns in relation to movement behavior. In this context, data mining as the key process of knowledge discovery in databases (KDD) is often applied on movement data in order to achieve a better understanding of movement patterns in terms of the moving objects' behavior.

Within spatial data mining, movement patterns today are mostly quantified on the basis of geometric properties and the arrangement of the GPS points. Although information about the geographic environment surrounding the movement patterns could provide semantics, from a methodological point of view, thus far the relation of movement to its surrounding geographic context has largely been neglected. Therefore, in this thesis, the relation of movement to the geographic context is investigated with regard to effects of preprocessing, its sensitivity to different methods for relating the movement to its embedding geographic context, and scale issues.

Preprocessing primarily aims at removing erroneous and filtering irrelevant data. In the KDD process, the focus lies on data mining, and the therewith development of algorithms. Besides data mining, preprocessing is also well established in recent research on movement analysis, as for example with the detection of stops and moves, or the identification of inaccurate GPS measurements. However, preprocessing is specific to almost each study, depending on various factors, such as, for example, the species under investigation or the problem scenario. Nevertheless, a common ground for preprocessing in movement analysis can be formalized on a more general level. In this work, a preprocessing workflow based on averaged speed values is presented. This preprocessing workflow proved to enable the detection of stops and moves at different temporal scales, applicable within different research contexts. Further, effects of preprocessing on the computation of movement parameters and their relation to geographic context were investigated in different case studies. The various experiments on quantitative effects of preprocessing procedures produced a complex picture. While in some cases the effect of preprocessing was significant (e.g., computing speed values on map-matched movement data), in others it was negligible (e.g., filtering stops in movement data with low temporal sampling rate). The experiments illustrated that preprocessing of movement data can cause expected (e.g., low speed values at stops), but also unexpected effects (e.g., unrealistic speed values due to map matching). All in all, on a more specific level, it is not possible to establish generic guidelines; instead,

besides the presented preprocessing workflow, individual considerations are required on a case-by-case basis.

GPS data are usually captured as a temporal sequence of point locations. Although a variety of movement models from different disciplines (e.g., GIScience, movement ecology) exist, in GIScience, movement most often is modeled in the form of points. As a consequence, movement-context relation methods are commonly based on the point geometry of the movement data with the use of some form of buffer analysis. In this work, we present a movement-context relation matrix, where different movement-context relation methods are categorized on the basis of models for movement (x-axis) and context (y-axis). The matrix was validated with a case study relating the movement of animals to factors of their geographic context. Different movement models and the therewith application of different movement-context relation methods resulted in significant differences in the space-use distributions generated with regard to the geographic context. Therefore, the proposed movement-context relation matrix is a reasonable means for choosing a suitable movement-context relation method.

In the relation of movement to geographic context different types of scale, such as the temporal scale of the movement, as well as the spatial and thematic scale of the context come into play. Recent studies in movement analysis most often focused on effects of varying the temporal scale. When even different sorts of scale are considered, interdependencies between different scales have rarely been investigated thus far. Therefore, sensitivity experiments were conducted with regard to different types of scale (temporal scale of movement, and spatial and thematic scale of geographic context). Besides effects of spatial and thematic scale of the geographic context on the result of relating movement to geographic context, interdependencies between all the types of scale under investigation in this thesis, were revealed. As a consequence, different types of scale cannot be regarded in isolation with regard to the relation of movement and geographic context. This thesis confirmed that taking into account different types of scale is crucial in order to understand movement in relation to its surrounding geographic environment. Coefficient of variation analysis allowed gaining the insight that within the relation of movement and geographic context, effects of scale and variations caused by different relation methods can be in the same order of magnitude.

In a first case study – relating animal movement (and its properties) to land cover and terrain using different relation methods – revealed manifold and not easy-to-predict dependencies between problem scenario, data granularity, chosen relation method and relation method parameterizations. Similarly, in a second case study – relating the movement of outdoor sportsmen and urban shoppers to their movement context – also revealed that in some constellations, the choice of relation method and the method for deriving context variables has significant impact on the strength of relations found. For instance, it was shown that correlation analysis is highly sensitive to different quantitative methods for modeling slope. This study illustrated the substantial difficulty to distinguish between what is really a new insight in terms of movement behavior and what appears to be an artifact of the methodology.

In the course of this work, different factors that play an important role in the relation of movement to geographic context were revealed. On the one hand, the quantification of aspects of movement and the geographic context proved to be crucial when relating them. On the other hand, preprocessing, movement-context relation methods, and scale showed having a significant impact on quantifying the relation of movement to geographic context. The insights gained in this thesis can support future studies with methods for developing more specific tools for context-aware movement analysis. This thesis presented means for bridging the semantic gap between movement patterns and movement behavior through the relation of movement to geographic context. Revealing causal relationships between movement and geographic context, however, still remains a major challenge of future research. In relation to statements of DODGE *et al.* (2008, p.245, according to BLYTHE *et al.*, 1996) and ORELLANA & RENSO (2010, p.67), this dissertation allows the following conclusion: “Movement is behavior, but patterns are not yet.”



# Zusammenfassung

Bewegung (in Raum und Zeit) ist Gegenstand verschiedener Forschungsfelder, wie etwa der Geographischen Informationswissenschaft, der Bewegungsökologie oder der Stadtplanung. Bewegungsanalyse vermag Erkenntnisse über das Verhalten von Tieren und Menschen aufzudecken. Heute verfügbare Technologien erlauben das Sammeln grosser Mengen von Bewegungsdaten in Form von GPS-Aufnahmen. Die methodische Disziplin der Bewegungsanalyse beabsichtigt das Auffinden von Bewegungsmustern hinsichtlich eines besseren Verständnisses von Bewegungsverhalten. In diesem Zusammenhang wird häufig Data-Mining als Teilprozess des Knowledge Discovery in Databases (KDD) auf Bewegungsdaten angewendet.

Innerhalb des räumlichen Data-Mining werden Bewegungsmuster heutzutage meistens auf der Basis geometrischer Eigenschaften und der Anordnung von GPS-Punkten quantifiziert. Obwohl Informationen über die geographische Umgebung zur semantischen Anreicherung von Bewegungsmustern dienen könnten, wurden bisher Methoden zur expliziten Quantifizierung des Zusammenhangs zwischen der Bewegung und dem geographischen Kontext weitgehend vernachlässigt. Deshalb wird in dieser Arbeit die Verknüpfung der Bewegung und dem geographischem Kontext in Bezug auf Effekte der Vorverarbeitung, dessen Sensitivität gegenüber Methoden zur Verknüpfung der Bewegung mit geographischem Kontext und Skalenaspekte untersucht.

Die Vorverarbeitung dient primär dem Entfernen fehlerhafter und dem Filtern irrelevanter Daten. Im KDD-Prozess liegt der Fokus auf dem Data-Mining und der damit verbundenen Entwicklung von Algorithmen. Nebst Data-Mining ist die Vorverarbeitung ein anerkannter Bestandteil des KDD-Prozesses innerhalb kürzlich erschienener Forschungsarbeiten der Bewegungsanalyse, wie zum Beispiel beim Detektieren von Stopps und Bewegungsphasen oder der Identifikation ungenauer GPS-Messungen. Es zeigt sich allerdings, dass die Vorverarbeitung für beinahe jede Studie in Abhängigkeit spezifischer Rahmenbedingungen (z.B. Art des Bewegungsobjekts oder Problemszenario) unterschiedlich ausgestaltet werden muss. Dennoch kann eine gemeinsame Basis für unterschiedliche Studien hinsichtlich der Vorverarbeitung in der Bewegungsanalyse auf einem allgemeineren Level formuliert werden. In dieser Arbeit wird ein Vorverarbeitungs-Arbeitsablauf, welcher auf gemittelten Geschwindigkeitswerten basiert, präsentiert. Dieser Vorverarbeitungsprozess hat sich für die Detektion von Stopps und Bewegungsphasen verschiedener zeitlicher Skalen innerhalb verschiedener Forschungskontexte als geeignet erwiesen. Des Weiteren wurden in verschiedenen Fallstudien die Sensitivität der Berechnung von Bewegungsparametern und deren Zusammenhang zu geographischem Kontext untersucht. Die verschiedenen Experimente bezüglich quantitativer Effekte der Vorverarbeitungs-Prozeduren erzeugten ein komplexes Bild. Die Vorverarbeitungs-Effekte waren in einigen Fällen

signifikant (z.B. Berechnung der Geschwindigkeit auf mit Map-Matching korrigierte Bewegungsdaten), in einigen anderen vernachlässigbar (z.B. Filtern von Stopps in Bewegungsdaten mit tiefer zeitlicher Auflösung). Die Experimente illustrierten, dass die Vorverarbeitung der Bewegungsdaten erwartete (z.B. tiefe Geschwindigkeiten bei Stoppphasen), aber auch unerwartete Effekte (z.B. unrealistische Geschwindigkeitswerte aufgrund des Map-Matching) zur Folge haben kann. Alles in allem gibt es diesbezüglich keine generellen Richtlinien; stattdessen benötigt man nebst einem allgemeineren Vorverarbeitungs-Arbeitsablauf auch spezifische Annahmen von Fall zu Fall.

GPS-Daten werden normalerweise als Beobachtungspunkte in zeitlicher Abfolge erfasst. Obwohl eine grosse Vielfalt an Bewegungsmodellen aus verschiedenen Disziplinen (z.B. Geographische Informationswissenschaft, Bewegungsökologie) existieren, wird die Bewegung in der Geographischen Informationswissenschaft oft in Form von Punkten modelliert. Demzufolge basieren auch Bewegung-Kontext-Verknüpfungsmethoden normalerweise auf Punktgeometrien der Bewegungsdaten, wie beispielsweise mit der Anwendung einer Pufferanalyse. In dieser Arbeit präsentieren wir eine Bewegung-Kontext-Verknüpfungsmatrix, bei welcher Bewegung-Kontext-Verknüpfungsmethoden auf der Basis von Bewegungs- (x-Achse) und Kontextmodellen (y-Achse) strukturiert werden. Die Matrix wurde mit einer Fallstudie validiert, indem Tierbewegungen in Bezug zu Faktoren des umgebenden geographischen Kontexts gesetzt wurde. Verschiedene Varianten zur Modellierung der Bewegung und die damit verbundene Anwendung verschiedener Bewegung-Kontext-Verknüpfungsmethoden resultierten in signifikanten Differenzen in den Verteilungen der Raumnutzung generiert in Bezug auf den geographischen Kontext. Deshalb stellt die vorgeschlagene Bewegung-Kontext-Verknüpfungsmatrix ein sinnvolles Mittel zur Auswahl einer geeigneten Bewegung-Kontext-Verknüpfungsmethode dar.

Bei der Verknüpfung der Bewegung und des geographischen Kontexts kommen verschiedene Skalenarten, wie die zeitliche Skala der Bewegung und die räumliche und thematische Skala des Kontexts ins Spiel. Neuste Studien der Bewegungsanalyse tragen häufig nur der zeitlichen Skala der Bewegung Rechnung. Falls sogar verschiedene Skalensorten berücksichtigt werden, werden jedoch Abhängigkeiten zwischen den Skalen bisher selten untersucht. Deshalb wurden in dieser Dissertation Sensitivitätsexperimente hinsichtlich verschiedener Skalentypen (zeitliche Skala der Bewegung, und räumliche und thematische Skala des geographischen Kontexts) durchgeführt. Nebst Effekten räumlicher und thematischer Skala des geographischen Kontexts auf das Resultat der Verknüpfung der Bewegung und geographischem Kontext, wurden Abhängigkeiten zwischen allen in dieser Arbeit untersuchten Skalentypen aufgezeigt. Deshalb können verschiedene Skalenarten in Bezug auf die Verknüpfung von Bewegung und geographischem Kontext nicht isoliert betrachtet werden. Diese Arbeit bestätigte, dass die Berücksichtigung verschiedener Instanzen verschiedener Skalenarten für das Verständnis der Bewegung in Bezug auf dessen geographische Umwelt wesentlich ist. Die Analyse mittels Variationskoeffizienten hat die Erkenntnis erlaubt, dass Effekte in der Bewegung-Kontext-Verknüpfung aufgrund der Skalenvariation



und unterschiedlicher Verknüpfungsmethoden in der gleichen Grössenordnung liegen können.

In einer ersten Fallstudie hat die Verknüpfung von Tierbewegungen (und deren Eigenschaften) zu Bodenbedeckung und Gelände mit verschiedenen Verknüpfungsmethoden, vielfältige und nicht einfach vorherzusagende Abhängigkeiten zwischen Problemszenario, Datengranularität, gewählter Verknüpfungsmethode und Methodenparametrisierung aufgedeckt. Ebenso zeigte die Verknüpfung der Bewegung von Outdoor-Sportlern und städtischen Einkäufern zu deren Bewegungskontext in einer zweiten Fallstudie, dass in einigen Konstellationen die Wahl der Verknüpfungsmethode und die Methode zum Ableiten von Kontextvariablen einen signifikanten Einfluss auf die gefundene Zusammenhangsstärke haben. Es wurde zum Beispiel gezeigt, dass die Korrelationsanalyse hoch sensitiv gegenüber verschiedener quantitativer Methoden für die Modellierung der Hangneigung ist. Diese Studie illustrierte die beachtliche Schwierigkeit der Unterscheidung zwischen neuen Erkenntnissen in Bezug auf Bewegungsverhalten und Artefakten der Methodik.

Im Zuge dieser Arbeit wurde aufgedeckt, dass verschiedene Faktoren in der Verknüpfung der Bewegung und des geographischen Kontexts eine wichtige Rolle spielen. Einerseits hat sich die Methodik zur Quantifizierung von Aspekten der Bewegung und des geographischen Kontexts bei deren Verknüpfung als entscheidend erwiesen. Andererseits hat diese Arbeit aufgezeigt, dass die Vorverarbeitung, Bewegung-Kontext-Verknüpfungsmethoden und Skala, massgebliche Einflussfaktoren auf die quantitative Ausprägung des Zusammenhangs zwischen der Bewegung und dessen geographischen Kontext sind. Die Erkenntnisse, welche in dieser Dissertation gewonnen werden konnten, können in zukünftigen Studien zur Entwicklung spezifischer Werkzeuge für die kontext-sensitive Bewegungsanalyse beitragen. In dieser Arbeit wurden Mittel zur Überbrückung der semantischen Lücke zwischen Bewegungsmustern und Bewegungsverhalten durch die Verknüpfung der Bewegung und geographischem Kontext präsentiert. Das Enthüllen kausaler Zusammenhänge zwischen Bewegung und geographischem Kontext bleibt dennoch eine Hauptherausforderung zukünftiger Forschung. Diese Erkenntnisse erlauben die folgende Schlussfolgerung in Anlehnung an die Aussagen von DODGE *et al.* (2008, p.245, according to BLYTHE *et al.*, 1996) und ORELLANA & RENSO (2010, p.67): “Bewegung ist Verhalten, aber Muster sind dies noch nicht.”



# Acknowledgements

This thesis was granted by the Swiss National Science Foundation within the projects 200021-129963 and 200020-149705, and carried out at the Department of Geography, University of Zurich. I want to thank both institutions for supporting my work.

This is the time and place to express my sincere gratitude to all the people, who contributed in many different ways to this thesis. First and foremost, I am indebted to my main advisor PD Dr. Patrick Laube for guiding me through the ups and downs of my time as a PhD student. Patrick found the right words for motivating me in a honest and positive way, and always showed patience and understanding for my problems. The exchange of ideas with Patrick was always very inspiring. Without him, it would have been impossible to bring this thesis to an end.

I owe debts to Prof. Dr. Rob Weibel for allowing me to develop my own research interests within his research group. A special thank goes to Prof. Dr. Ross Purves, University of Zurich, and Prof. Dr. Nico Van de Weghe, Ghent University, for their insightful comments and discussions in the course of the annual committee meetings.

GIScience mostly means analyzing data, making me dependent on others, collecting these data. Thus, this thesis would not have been possible without many people providing me with movement, as well as context data. I want to acknowledge the Swiss National Park for movement data of ungulates and detailed data about different aspects of the surrounding environment. I owe a debt of gratitude to Prof. Dr. Stefan van der Spek, Delft University of Technology, for different data sets of moving pedestrians, and Prof. Dr. Reto Rupf, Zurich University of Applied Sciences, for movement data of hikers and bikers collected within the “mafreina” project.

I want to thank Annica Mandola for her administrative support and the numerous interesting discussions about things outside of research.

All my colleagues, including PhD students, as well as students, technical staff, postdocs, lecturers and professors made my life at University of Zurich very delightful; I will never forget the time with you. I want to separately mention Dr. Curdin Derungs. I had the pleasure to share the office with him, and could not imagine a better office mate; it was fun. Curdin knew how to motivate and inspire me in my daily work with numerous discussions about research and “God and the world”.

I owe my deepest gratitude to my parents, who always believed in me and allowed me to explore the world from many different perspectives. My grandparents showed me the meaning of modesty and happiness, enabling me to do a PhD.

A dissertation seems to be like climbing a mountain. I sometimes did not believe that I can make it to the “peak”, and I sometimes lost track in the “fog of research”. My wife and my daughter were always there, and helped me to reach the “mountain top” with much more than supporting and motivating me.



# Contents

<b>Summary</b>	<b>i</b>
<b>Zusammenfassung</b>	<b>v</b>
<b>Acknowledgements</b>	<b>ix</b>
<b>List of Figures</b>	<b>xv</b>
<b>List of Tables</b>	<b>xix</b>
<b>List of Abbreviations</b>	<b>xxi</b>
<b>1. Introduction</b>	<b>1</b>
1.1. Motivation . . . . .	1
1.1.1. Why is movement studied? . . . . .	1
1.1.2. Putting movement into context . . . . .	2
1.2. Research gaps and research questions . . . . .	3
1.3. Structure of the thesis . . . . .	5
<b>2. Background</b>	<b>7</b>
2.1. Movement data . . . . .	7
2.1.1. Capturing movement data . . . . .	7
2.1.2. Characteristics of movement data . . . . .	8
2.2. About spaces and places . . . . .	9
2.2.1. Movement spaces and geographic context . . . . .	9
2.2.2. Definitions of context . . . . .	11
2.2.3. Modeling context . . . . .	12
2.3. From movement towards a better understanding of behavior . . . . .	16
2.3.1. Preprocessing and semantic enrichment . . . . .	17
2.3.2. Movement in its embedding geographic context . . . . .	20
2.3.3. Scale in movement analysis . . . . .	25
2.4. Research gaps and research questions . . . . .	27
<b>3. Preprocessing and semantic enrichment of movement data</b>	<b>29</b>
3.1. Preprocessing: segmenting stops and moves . . . . .	29
3.2. From preprocessing to semantic enrichment: a preprocessing workflow	31
3.3. Stops and moves: application of the preprocessing workflow . . . . .	33

3.4.	Exploring effects of preprocessing . . . . .	36
3.4.1.	Quantifying effects of preprocessing . . . . .	36
3.4.2.	Data and experiments . . . . .	36
3.4.3.	Results and discussion . . . . .	39
3.5.	Main findings and contributions . . . . .	41
<b>4.</b>	<b>Concepts for relating movement to geographic context</b>	<b>43</b>
4.1.	A movement-context relation matrix . . . . .	43
4.1.1.	The movement axis . . . . .	43
4.1.2.	The context axis . . . . .	45
4.1.3.	Movement-context combinations . . . . .	45
4.2.	Vector-field movement-context relation methods . . . . .	48
4.3.	Move-field movement-context relation methods . . . . .	49
4.4.	Segment-field movement-context relation methods . . . . .	50
4.5.	Polygon-field and field-field movement-context relation methods . . . . .	51
<b>5.</b>	<b>Relating ungulates' movement to geographic context</b>	<b>53</b>
5.1.	Study area and data . . . . .	53
5.2.	Experiments . . . . .	54
5.2.1.	Validity experiment: Is the movement at all related to the terrain? . . . . .	54
5.2.2.	Sensitivity experiment #1: Differences between relation methods . . . . .	54
5.2.3.	Sensitivity experiment #2: Differences due to varying sector parameters . . . . .	56
5.3.	Results . . . . .	57
5.3.1.	Validity experiment: Is the movement at all related to the terrain? . . . . .	57
5.3.2.	Sensitivity experiment #1: Differences between relation methods . . . . .	59
5.3.3.	Sensitivity experiment #2: Differences due to varying sector parameters . . . . .	60
5.4.	Main findings and contribution . . . . .	60
<b>6.</b>	<b>Relating people's movement to geographic context</b>	<b>63</b>
6.1.	Study area and data . . . . .	63
6.2.	Experiments . . . . .	64
6.2.1.	Relating speed and slope . . . . .	64
6.2.2.	Relating movement to user-generated content . . . . .	65
6.3.	Results . . . . .	67
6.3.1.	Relating speed and slope . . . . .	67
6.3.2.	Relating movement to user-generated content . . . . .	70
6.4.	Main findings and contribution . . . . .	73
<b>7.</b>	<b>Scale effects in context-aware movement analysis</b>	<b>75</b>
7.1.	Study area and data . . . . .	75
7.2.	Experiments . . . . .	76

7.3. Results . . . . .	78
7.3.1. Differences due to scales and methods . . . . .	79
7.3.2. Interdependencies between scales . . . . .	80
7.4. Main findings and contribution . . . . .	81
<b>8. Discussion</b>	<b>89</b>
8.1. Effects of preprocessing on context-aware movement analysis . . . . .	89
8.2. Modeling movement and its embedding geographic context . . . . .	91
8.3. Methods for relating movement to geographic context . . . . .	94
8.4. Scale effects in context-aware movement analysis . . . . .	97
<b>9. Conclusions</b>	<b>99</b>
9.1. Achievements . . . . .	99
9.2. Insights . . . . .	101
9.3. Outlook and challenges of future research . . . . .	102
<b>Bibliography</b>	<b>105</b>
<b>A. Curriculum vitae</b>	<b>117</b>
<b>B. Complete publication list</b>	<b>119</b>





# List of Figures

2.1. Concept of space-time cube. Left: grouping of individuals; right: space-time prism, grey: potential path area based on a maximum speed threshold (Source: author’s own graphs according to HÄGERSTRAND, 1970). . . . .	10
2.2. Landform classification with an automated procedure into 8 classes for the Yatsugatake volcano in Japan (IWAHASHI & PIKE, 2007). . . . .	14
2.3. KDD process (according to MILLER, 2008, extended version) in relation to semantic enrichment process (BAGLIONI <i>et al.</i> , 2009). . . . .	17
2.4. Movement ecology paradigm (NATHAN <i>et al.</i> , 2008, p.19054). . . . .	23
3.1. Process chain for the segmentation of stops and moves (author’s own graph). . . . .	30
3.2. Illustrative example of the process chain for detecting stops and moves (author’s own graph). . . . .	31
3.3. Self-tracking in the city of St. Gallen, Switzerland (author’s own data, 2011). Orange: actual movement path, red: GPS measurements (author’s own data, 2011), map: “Landeskarte 1:25’000” (“Bundesamt für Landestopographie”, 2011). . . . .	32
3.4. Space-time cube and local scenario in “Neltje Jans” with stops and moves according to the preprocessing workflow. Grey: short stop/“-2”, white: short move/“-1”, red: stop/“0”, any other color: move/“>1”. Data source: TU Delft, S. van der Spek, 2011. . . . .	34
3.5. Another local scenario with stops and moves according to the preprocessing workflow. Grey: short stop/“-2”, white: short move/“-1”, red: stop/“0”, any other color: move/“>1”. Data source: TU Delft, S. van der Spek, 2011. . . . .	35
3.6. Example trajectory section for a pedestrian in Delft, without (green) and with naïve map matching (red), fix indices at sampling rate of 2 seconds (Data source: VAN DER SPEK <i>et al.</i> , 2013). . . . .	38
3.7. Example trajectory of chamois with habitat context in Swiss National Park. Stationary fixes (white), moves in various colors (purple, dark blue, dark red, orange, red), and time of day (hh:mm:ss). Data sources: movement data, Swiss National Park (2002-2010); context data, www.habitalp.de (2002-2006). . . . .	39

List of Figures

3.8. (a) Case study #1, speed for pedestrians; GPS vs. computed; raw data (r), filtered (f) and map-matched (mm). (b) Case study #2, speed vs. habitat for chamois; raw data (r) vs. filtered (f). . . . .	40
4.1. <i>mc</i> -relation model (“relation matrix”, author’s own graph). . . . .	46
4.2. Concepts for relating the vector-based movement model to the embedding (field-based) geographic context ( $m_{VCF}$ , author’s own graph based on established GIS concepts). . . . .	48
4.3. Relating (a) move- and (b) segment-based models of movement to geographic context (author’s own graph based on established GIS concepts). . . . .	50
4.4. Concepts for movement models based on representations as polygons or fields and the relation to geographic context: (a) Convex hull, (b) KDE, (c) BBMM and its relation to (d) geographic context (author’s own graph based on established GIS concepts). . . . .	52
5.1. CRW vs. GPS data of a female ibex (no. 469) in relation to geomorphometric landforms by IWAHASHI & PIKE (2007) with a hillshading in the background in Swiss National Park (data sources: GPS data, Swiss National Park, 2006 – 2007; DEM “dtm4”, Swiss National Park, 2010). . . . .	55
5.2. Dissolved MCPs per species in Swiss National Park with a hillshading (data source: DEM “dtm4”, Swiss National Park, 2010). . . . .	56
5.3. Representations of the movement path of a female chamois (no. 47) in Swiss National Park. (a) Spatial models with KDE and MCP, and (b) spatio-temporal approach with BBMM (data source: GPS data, Swiss National Park, 2002 – 2004). . . . .	57
5.4. Histograms representing results of experiments with regard to VALIDITY, RELATION METHODS and VARIATION. . . . .	58
6.1. Different approaches for computing slope and speed (distance). Adapted from <a href="http://www.innovativegis.com/basis/present/gita_denver05/default.htm">http://www.innovativegis.com/basis/present/gita_denver05/default.htm</a> , accessed on 29/12/2014. $\alpha$ & $\beta$ : gradients from DEM, $\gamma$ : slope of the beeline, $d_{2D}$ & $d_{3D}$ : 2D and 3D distances between consecutive GPS fixes. . . . .	65
6.2. Methodology for the relation of movement of pedestrians in Rotterdam and context in form of Foursquare categories based on KDE (author’s own graph). . . . .	67
6.3. Correlation of speed (logarithmic scale) and slope using a linear fit model for bikers and hikers around the Fuorn pass considering two different quantitative models of slope. . . . .	69

6.4.	Correlation of speed (logarithmic scale) and slope using a linear fit model of single down- and uphill trails around the Fuorn pass considering the slope of the beeline. (a) Single downhill trail of a biker, (b) Single uphill trail of a biker, (c) Single downhill trail of a biker for slopes between $-40^\circ$ and $-10^\circ$ , (d) Single downhill trail of a biker for slopes between $-10^\circ$ and $-5^\circ$ . . . . .	70
6.5.	Correlation of speed (logarithmic scale) and slope using a linear fit model for bikers and hikers around the Fuorn pass considering two different quantitative models of speed. . . . .	71
6.6.	Resulting KDEs for the movement (all data), a Foursquare category (“shop & service”) and the pairwise multiplication of the two with a small extent of the resulting classification and the original point locations of the “shop & service” Foursquare category. . . . .	72
6.7.	Resulting classification of the relation of movement of pedestrians in Rotterdam and context in form of Foursquare categories based on KDE. . . . .	73
7.1.	Aspect represented in 5 classes on different spatial scales (4, 20 and 100 <i>meters</i> ), and a chamois’ (no. 22) movement modeled with the Brownian Bridge Movement Model (BBMM) (increasing probability density from black to white) on different temporal scales (10 <i>min</i> / 4 <i>h</i> , 30 <i>min</i> / 12 <i>h</i> and 1 <i>h</i> / 1 <i>day</i> ) around the Fuorn pass within the Swiss National Park (purple). . . . .	76
7.2.	Aspect classes at different thematic scales (without class Flat, author’s own graph). . . . .	77
7.3.	Relative distributions (0-1) of the geographic context (terrain aspect) in relation to the chamois’ movement. . . . .	84
7.4.	Coefficients of variation representing within-class variation due to variation of temporal (a) and spatial (b) scale across different thematic scales. . . . .	85
7.5.	Relative distribution of aspect related to the chamois’ movement (with map pin) for whole year (all), winter and summer (@10 <i>min</i> /4 <i>h</i> , @4 <i>m</i> , @9 categories). . . . .	86
7.6.	Overview of scale effects and interdependencies in relating movement to geographic context. Green: effects/interdependencies detected, red: no effects/interdependencies detected (author’s own graph). . . . .	87



# List of Tables

3.1. Overview of the pedestrians' and chamois' movement data sets (data sources: pedestrians, VAN DER SPEK <i>et al.</i> , 2013 / chamois, Swiss National Park, 2002 – 2010). . . . .	37
4.1. Overview of $m_{VC_F}$ -relation methods for vector-based movement model (linked to Figure 4.2). . . . .	48
5.1. Overview of the ungulates' movement data sets. . . . .	53
5.2. Mean and standard deviation and coefficient of variation (in percentage) of relative frequencies per landform class and species (mean $\pm$ std and std/mean). . . . .	60



# List of Abbreviations

<b>BBMM</b>	Brownian Bridge Movement Model
<b>CAMA</b>	Context-Aware Movement Analysis
<b>CRW</b>	Correlated Random Walk
<b>DEM</b>	Digital Elevation Model
<b>DOP</b>	Dilution Of Precision
<b>GIR</b>	Geographic Information Retrieval
<b>GIS</b>	Geographic Information System
<b>GIScience</b>	Geographic Information Science
<b>GKD</b>	Geographic Knowledge Discovery
<b>GME</b>	Geospatial Modelling Environment
<b>GPS</b>	Global Positioning System
<b>HDOP</b>	Horizontal Dilution Of Precision
<b>HMM</b>	Hidden Markov Model
<b>ICT</b>	Information and Communication Technology
<b>IDW</b>	Inverse Distance Weighting
<b>IR</b>	Information Retrieval
<b>KDD</b>	Knowledge Discovery in Databases
<b>KDE</b>	Kernel Density Estimation
<b>KDE-DT</b>	Kernel Density Estimation-Delaunay Triangulation
<b>LBS</b>	Location-Based Service
<b>LBSN</b>	Location-Based Social Network
<b>LISA</b>	Local Index of Spatial Association
<b>MAUP</b>	Modifiable Areal Unit Problem

*List of Tables*

<b>MCP</b>	Minimum Convex Polygon
<b>MOD</b>	Moving Object Database
<b>MPO</b>	Moving Point Object
<b>PDOP</b>	Position Dilution Of Precision
<b>POI</b>	Point Of Interest
<b>RDF</b>	Radial Distance Function
<b>REMO</b>	RElative MOtion
<b>RFID</b>	Radio-Frequency IDentification
<b>RSF</b>	Resource Selection Function
<b>SMoT</b>	Stops and Moves of Trajectories
<b>TGDE</b>	Time-Geographic Density Estimation
<b>TIN</b>	Triangulated Irregular Network
<b>T-OPTICS</b>	Trajectory-Ordering Points To Identify the Clustering Structure
<b>UD</b>	Utilization Distribution
<b>UGC</b>	User Generated Content
<b>VGI</b>	Volunteered Geographic Information



# Chapter 1.

## Introduction

### 1.1. Motivation

#### 1.1.1. Why is movement studied?

“Movement is behavior.” (DODGE *et al.*, 2008, p.245; according to BLYTHE *et al.*, 1996)

Movement can be seen as traces of behavior, which is mirrored in the above quote as the key motivation of studying movement. In light of the assumption that individuals' behavior is reflected in their movement, different research fields – including Geographic Information Science (GIScience), behavioral ecology, urban planning (including studies of urban mobility) and computer science, for example – all contribute to the analysis of movement. Accordingly, movement analysis aims to achieve a better understanding of individuals' movement in terms of behavior. GIScience and computer science are mainly focused on the development of methodologies that allow the application fields (e.g., behavioral ecology and urban planning) to reveal new insights into movement behavior for addressing problems of animal conservation or urban mobility. In the following paragraphs, motivations specific to the mentioned research fields – and recognized as relevant within the framework of this thesis – are discussed in more detail.

**GIScience.** GIScience has strong expertise in developing methods for capturing, managing, analyzing, and presenting spatial data (e.g., CLARKE, 1995). Besides the spatial component, inherently, movement also has a temporal one; in other words, movement is spatio-temporal and, therewith, is represented with spatio-temporal data. With the strong methodological background, GIScience strengths are also identified in the field of movement analysis. For instance, IMFELD (2000) illustrates (as one of the first) that the expertise of GIScience in analyzing spatial data can be applied and expanded to the spatio-temporal analyses of movement data. Generally, GIScience contributes to movement analysis predominantly based on a methodological motivation.

**Behavioral ecology.** As the name of the field suggests, this research area is primarily interested in better understanding the impact of evolutionary and environmental

factors on animals' behavior. Behavioral ecology has a long tradition in observing animals and their behavior in the field (e.g., HEBBLEWHITE & HAYDON, 2010). With technological progress allowing for an almost continuous capturing of animals' movement, new possibilities of studying animals' behavior have been seen to emerge (e.g., TOMKIEWICZ *et al.*, 2010), where movement is supposed to reflect evolutionary and environmental effects on animals' behavior.

**Urban planning.** Urban planners aim to achieve a better understanding of people's spatio-temporal behavior (e.g., VAN DER SPEK, 2008), in addition to how people use the physical environment (e.g., VAN SCHAICK, 2010) since they seek to design urban areas as efficiently and attractively as possible with regard to residents' and visitors' needs. In order to achieve these goals, the study of people's mobility in urban areas is an established means of garnering greater insights into people's spatio-temporal behavior – and, finally, in regard to the possible directions of urban planning (e.g., VAN DER SPEK *et al.*, 2009).

**Computer science.** Knowledge Discovery in Databases (KDD) is a process concerned with generating knowledge from data (FAYYAD *et al.*, 1996). This procedure originated in the context of database research in the area of computer science, and is better known as data mining; importantly, however, data mining is only one aspect of the entire KDD process. KDD and data mining are not only applicable to the use of database structures, but also are applicable to a number of different research fields. Since KDD is used and adapted in mind of addressing geographic questions (within Geographic Knowledge Discovery (GKD), MILLER, 2008), data mining is also recognized as relevant with regard to spatio-temporal problems within the analysis of movement (e.g., GUDMUNDSSON *et al.*, 2008).

This thesis is embedded within the aforementioned research areas, with a strong focus directed towards methodological issues (GIScience). The main methodologically motivated analysis of movement is realized within the application areas, namely behavioral ecology and urban planning, predominantly on the basis of the KDD process, i.e., computer science.

### 1.1.2. Putting movement into context

Technical progress within Information and Communication Technology (ICT) allows an almost continuous tracking of moving objects in space and time, and accordingly yields a significant potential in terms of assessing movement behavior. Although large volumes of movement data are available, the development of methods for movement analysis with regard to the problems of the application areas is lagging behind, since methodologies differ significantly for different research purposes and application areas. Data mining mainly aims at the detection of movement patterns as a type of aggregation of the movement data in a way that allows humans to better understand the data in terms of movement behavior. Most of the methods put forward thus

far in terms of identifying patterns in movement data, are focused on geometric characteristics and arrangement patterns in movement paths (e.g., LAUBE *et al.*, 2005). The semantic meaning of such geometric properties has to be approached in order to develop better understanding of movement patterns as movement behavior. Consequently, movement has to be put into the correct context (PURVES *et al.*, 2014; LAUBE, 2014).

The study of geometric properties of movement paths in particular facilitated the investigation of interaction between different moving objects (e.g., leadership, ANDERSSON *et al.*, 2008). ANDRIENKO *et al.* (2011b) propose an event-based conceptual model for Context-Aware Movement Analysis (CAMA), and validate this model in terms of the real movement data of roe deer, including relations amongst movers (roe deer and lynxes), as well as the relations of the roe deer to spatial locations (e.g., geographic context in form of open areas). However, from a methodological perspective, the movement as a response to other moving entities has been the focus of much more in-depth research than the impact of the physical surrounding environment (geographic context) on actual movement. Thus, this thesis aims at relating movement to its embedding geographic context, thus enabling a better understanding of movement patterns in terms of movement behavior.

## 1.2. Research gaps and research questions

Besides the well-known data mining, the KDD process also includes equally important issues of background knowledge, preprocessing and knowledge construction (FAYYAD *et al.*, 1996, terminology according to MILLER, 2008). Background knowledge and knowledge construction are mainly the expertise of the application areas (e.g. behavioral ecology or urban planning). Since GIScience rather focuses on methodological research, the step of preprocessing is particularly relevant for this thesis. Many studies in movement analysis illustrate that, in any case, work based on real movement data always requires preprocessing (e.g., LAUBE & PURVES, 2011). However, although preprocessing is very time-consuming, few details relating to preprocessing steps and their impact on the actual results have been studied in most research of movement analysis. This issue is approached through the application of Research Question 1 (RQ1).

Studies within GIScience often model movement as time-stamped point locations (a triple consisting of spatial coordinates and time as  $(x, y, t)$ ), where consecutive Global Positioning System (GPS) records are connected by straight lines, which are usually referred to as trajectories. In other fields, such as behavioral ecology, movement is, amongst others, represented as fields (e.g., BBMM, HORNE *et al.*, 2007) or aggregated into polygons (e.g., Minimum Convex Polygon (MCP), POWELL, 2000). The selection of a specific movement model over another is very task-dependent. However, the different research fields have in common that they often do not explicitly motivate their choice for representing the movement. As an exception, research within home range analysis leads the way as, in this regard, different estimators are compared

(e.g., Kernel Density Estimation (KDE) vs. MCP, WARTMANN *et al.*, 2010). Still, more than one or two models for the representation of movement are rarely discussed and considered in studies of movement analysis. This research gap is addressed with Research Question 2 (RQ2).

With different movement models, different methods for relating this movement (in the form of specific data structures) to its embedding geographic context are required. Moreover, in many studies of movement analysis within GIScience, context in relation to movement is only taken into account in the point locations of the GPS fixes, although the surrounding environment also has an influence on the movement (e.g., DODGE *et al.*, 2012). Similar to the line of argumentation in the course of Research Question 2 (RQ2), within studies of movement analysis, few methods for relating movement to its surrounding geographic context are considered in the actual relation of the two. Research Question 3 (RQ3) addresses this gap by proposing and comparing various quantitative approaches for relating movement to its surrounding geographic context.

Scale issues have been considered in many studies in both GIScience and behavioral ecology (e.g., LEVIN, 1992). In the main, the effects of temporal scale (sampling rate of the movement data) (e.g., LAUBE & PURVES, 2011) and movement on different spatio-temporal scales (meters to kilometers and minutes to years) (e.g., FRYXELL *et al.*, 2008) have been investigated in research with regard to movement analysis. Throughout the procedure of relating movement to geographic context, different types of scale of both movement and geographic context come into play. Overall, on different scales, different patterns can be observed, with patterns (e.g., movement patterns) seen to emerge on completely different scales than the processes (e.g., geographic context) that produce them (LEVIN, 1992). Different sorts of scale as temporal scale of movement, as well as spatial and thematic scale of the context data, have been taken into account – mainly in investigations of behavioral ecology (e.g., BÖRGER *et al.*, 2006). Nevertheless, different types of scale have been considered very rarely in one single study, and interdependencies between the scales due to the relation of movement to geographic context have been neglected thus far. Accordingly, sensitivity to the relation of movement and geographic context to different types of scale and interdependency between different kinds of scales are addressed in the course of Research Question 4 (RQ4).

In mind of the above, we propose to address the following research questions throughout the completion this thesis:

- **RQ1.** What is the influence of commonly used preprocessing steps (e.g., segmentation, map matching) on context-aware movement analysis, and how can it be quantitatively evaluated?
- **RQ2.** How should movement and the geographic context embedding movement be modeled so as to allow for a quantitative relation between the two?

- **RQ3.** How can movement – and, explicitly, its spatio-temporality (not only its mere spatial footprint) – and the geographic context embedding this movement be quantitatively interrelated?
- **RQ4.** How sensitive is the computation of a quantitative relation between movement and its embedding context to a systematic variation of the temporal, spatial and thematic scale of analysis, and can interdependencies between the different scale dimensions be identified and quantified?

Within the framework of this thesis, movement is understood to be the change in the spatial location of the entire body of an individual in time (NATHAN *et al.*, 2008). Therefore, movement is not intended to be the movement of only parts of an individual's body: for example, shaking the head without changing spatial location over time is not considered movement within this research of movement analysis.

In this thesis, the methods for the quantitative analyses of relations between movement and geographic context do not intend to assess the relevance of context factors with regard to the movement of specific moving objects; this research rather aims to achieve the quantification of relations between movement and geographic context in an effort to quantitatively confirm educated guesses from within the application areas for the existence of movement-context relations.

The relation of movement and geographic context is assessed in light of the different aspects of the KDD process and methodological issues, such as preprocessing, different models for representing movement and geographic context, and different movement-context relation methods. Furthermore, scale effects with regard to different types of scales of movement and its embedding geographic context are addressed.

Within this thesis, the applied, extended and proposed methods are mainly based on the theoretical framework of the KDD process and data mining. Specifically, the focus lies on spatial, temporal and spatio-temporal methodologies on the basis of concepts from GIScience.

### 1.3. Structure of the thesis

In the following chapter, a literature review is provided, in an effort to provide a link of relevant theory and to motivate the research questions in greater depth. Chapter 3 presents methods and a case study addressing the various effects of preprocessing on the relation of movement characteristics and geographic context, within which a preprocessing workflow is suggested in mind of the identification of stops and moves. After presenting a movement-context relation framework based on models for movement and geographic context (Chapter 4), this framework is applied on the movement data of ungulates (Chapter 5). Further, in Chapter 6, movement is related to geographic context, utilizing the movement data of pedestrians and outdoor sportsmen in urban and mountainous areas, respectively, with a focus on two examples for modeling geographic context. In Chapter 7, scale effects in the procedure of relating movement to geographic context are assessed. Finally, Chapter

## *Chapter 1. Introduction*

8 provides a discussion of the research questions on a more general level, with the conclusions of this thesis presented in the last chapter (Chapter 9).

# Chapter 2.

## Background

### 2.1. Movement data

#### 2.1.1. Capturing movement data

Different methods can be applied to capture the positional data of moving organisms. Progress in ICT brings about a wealth of advantages and new possibilities; at the same time, however, these have been critically reviewed mainly by biologists (HEBBLEWHITE & HAYDON, 2010; TOMKIEWICZ *et al.*, 2010; URBANO *et al.*, 2010). Traditionally, behavioral ecologists observe animals and their interactions with one another and to with their environment, directly in the field, and manually register the positions of individuals. Furthermore, HEBBLEWHITE & HAYDON (2010) argue that, due to the technical advances in localization techniques, behavioral ecologists spend less time in the field, causing the biological “feel” to be lost in terms of understanding animals in their native environment. On the other hand, however, biologists – specifically those studying organisms in their habitat from single “hotspots” – cannot follow a movement path over a longer period of time owing to, for example, topography, which avoids access to regions where animals still can live, where wide-ranging species as birds were impossible to follow over huge distances – even crossing the sea – without the inventions of ICT approaches. In the case of field observations, behavioral ecologists potentially disturb animals by entering their habitat. The longer term impacts on the behavior of observed animals are reduced by tagging them with GPS collars. From biologists’ perspectives, the main advantage of new localization technologies is predominantly concerned with establishing where animals move, rather than the increasingly finer granularity of GPS-recorded movement data (HEBBLEWHITE & HAYDON, 2010).

Capturing movement data with localization techniques, including GPS and Radio-Frequency IDentification (RFID), leads to growing movement data repositories for different types of moving object, such as various species of animal, pedestrian, bicycle, or car. The shadowing effects caused by mountains in rural regions or buildings in urban areas block the GPS signal, which can have an influence on the quality of data, with indoor measures (mostly in urban areas) in particular recognized as rather inaccurate. The RFID technology also enables the localization of moving objects, and especially is recognized as suitable for indoor environments. However, in comparison with GPS covering localization on the entire globe (TOMKIEWICZ *et al.*, 2010), the

range of RFID is in the magnitude of meters. The reason for this is that the RFID technique is based on remotely retrieving and storing data using tags and readers, whilst spatial information is derived from the signal strength (ZHOU & SHI, 2009). On the other hand, in the case of GPS, the time delay of signals between the satellite and receiver is considered in order to localize moving objects. Nowadays, in many studies taking place in outdoor environments, movement data is recorded by GPS receivers, such as in the domain of wildlife ecology, studying animal behavior (TOMKIEWICZ *et al.*, 2010), or studies in urban environments (VAN SCHAICK & VAN DER SPEK, 2008), for example.

Essentially, there are two perspectives of perceiving movement, the *Eulerian* and the *Lagrangian* view (TURCHIN, 1998); the former approach investigates movement at fix points in space (e.g., classic field observation of behavioral ecologists), whilst the latter perspective aims at analyzing changes of movement (e.g., speed) along the movement path. In this research project, the focus will be directed towards working with the Lagrangian approach as the modeling of movement paths with Moving Point Objects (MPOs) as trajectories is most suited to the Lagrangian view.

### 2.1.2. Characteristics of movement data

In the research context of movement analysis, movement is defined as the change of spatial location over time. The change of objects' geometric form is not in the focus of this research area: for example, moving the upper part of the body without changing spatial location is not understood as movement (LAUBE, 2009); hence, movement can be modeled as a temporal sequence of geographic point locations, referred to as MPOs. A set of moving points sorted by time is defined as a trajectory. Under the assumption that a MPO moves in a straight direction and at constant speed between two consecutive moving points, it is common to connect consecutive moving points through the use of a straight line (DODGE *et al.*, 2008).

Naturally, space and time are continuous, whereas movement, on the other hand, cannot be captured and stored on computers as continuous paths (PEUQUET, 1994). Technically, there are several methods for recording locations as a discrete subset of the real movement path. Discretization arises in regard to the question of scale, where the choice of a specific temporal scale of movement data depends on the purpose of the movement analysis (LEVIN, 1992). Most GPS devices register point locations at constant temporal intervals, ranging from fractions of seconds up to hours. This sampling rate, at constant time steps, is understood as the temporal granularity of movement data. Consequently, movement data is also most often modeled as temporal sequence of point locations, comprising a unique identifier, point coordinates of spatial location (where?), time (when?) and a number of additional properties of the MPO itself or of the surroundings (what?), such as in relation to the geographic context embedding the MPO ( $id, x, y, t, <additional\ attributes>$ ). In the triad model, PEUQUET (1994) distinguishes three fundamental sets for movement: space (where?), time (when?) and objects (what?). The various elements of each set – or the combinations of them (e.g., speed is a combination of elements of space and



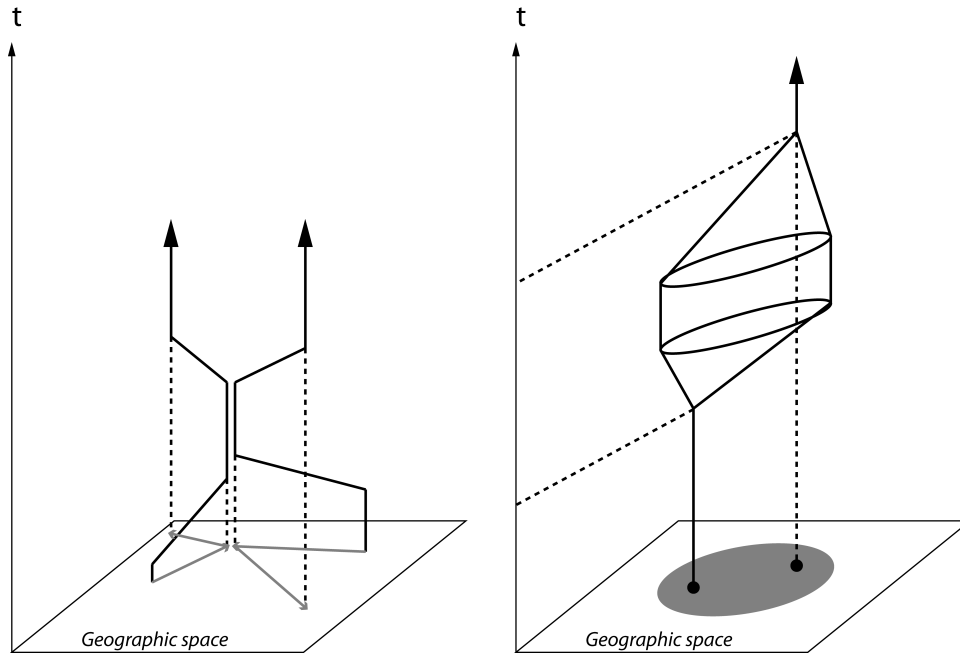
time) – can be represented as attributes building a movement dataset (PEUQUET, 1994; ANDRIENKO *et al.*, 2011a).

## 2.2. About spaces and places

### 2.2.1. Movement spaces and geographic context

Methods for analyzing movement data towards achieving a better understanding of movement behavior cannot cope with the increasing amounts of movement data facilitated by the technical advances with regard to ICT during the last decade. Methods for movement analysis are sensitive to different movement models (representations of movement); in turn these can take different forms, depending on how the movement embedding space is modeled (LAUBE, 2009). To this end, a variety of different movement spaces – such as Euclidean homogeneous space, constrained Euclidean space, space-time aquarium, heterogeneous field space, irregular tessellation and network space – require specific representations of movement, as well as movement analysis methods (LAUBE, 2014). Through this research project, the focus lies on animals moving in Euclidean (homogeneous) space and on people moving in an urban environment on street networks (network space). A prominent example amongst different models of spaces originates from time geography using the space-time cube. HÄGERSTRAND (1970) introduced the concept of the space-time cube in the analysis of movement in the context of regional science for policy and planning. The author presents how groupings of several individuals can be visually identified in a kind of tube. Furthermore, the paper establishes the space-time prism based on the potential path area constrained by a maximum speed, depending on the type of MPO (Figure 2.1). Further, irregular tessellation is often applied with movement derived from mobile phone data, where antenna connections are logged over time.

Within GIScience, movement analysis research thus far has been predominantly concerned with algorithmically detecting shape, arrangement, or interaction patterns based on geometric criteria (LAUBE, 2014). Movement is driven by different factors, including navigation capacity, motion capacity, internal state and external factors (NATHAN *et al.*, 2008). Therefore, movement ecology aims at achieving a better understanding of movement as a response to biotic and abiotic external factors, amongst others, whereas biotic external factors can also be referred to the interaction between individuals (*second order effects*), whilst abiotic external factors can be seen as environmental factors (*first order effects*) influencing the movement. Relating movement to its embedding geographic context allows for a better understanding of where, when and particularly why individuals move as they do. As a consequence, from a data mining perspective, in movement pattern analysis, the relation of movement patterns to the embedding geographic environment is a key challenge (GUDMUNDSSON *et al.*, 2008; PURVES *et al.*, 2014). The GIScience community has only rarely considered geographic context in the domain of movement analysis (BITTERLICH *et al.*, 2009; GUDMUNDSSON *et al.*, 2012). Earlier, IMFELD (2000) recognized the need for extending the classic Geographic Information System (GIS) toolbox by methodologies allowing



**Figure 2.1.:** Concept of space-time cube. Left: grouping of individuals; right: space-time prism, grey: potential path area based on a maximum speed threshold (Source: author's own graphs according to HÄGERSTRAND, 1970).

for taking into consideration geographic surroundings for a better understanding of individuals' movement; therefore, Radial Distance Function (RDF) was developed in mind of quantitatively capturing the geographic surroundings of an MPO.

In movement ecology, movement is linked to environmental factors for gaining a better understanding of processes that have an influence on movement. Movement data can be enriched with environmental data from many sources and many types using remotely sensed data, for example (DEMŠAR *et al.*, 2015). DODGE *et al.* (2013) implemented an automated system for annotating GPS tracks with environmental variables from global remote sensing, weather, and ecosystem products. With the use of this annotation system in another study, the relation of movement of turkey vultures to environmental conditions revealed correlations explaining how far vultures need to move to find food and how fast they can move depending on thermals and temperature (DODGE *et al.*, 2014). Moreover, (MORELLET *et al.*, 2013) discovered an interesting interplay between home range size and seasonality, weather and climate, using different sources of environmental data (e.g., remote sensing data). These studies illustrate that in recent decades movement ecology underlies remarkable progress in analyzing movement related to environmental factors using ICT technologies; from a methodological point of view, however, there is potential for improvement with regard to the quantitative relation of movement and geographic context, as IMFELD (2000) introduced the concept of RDFs in the context of movement analysis already some years ago.

Although geographic context thus far not has been commonly considered in movement analysis, the quantitative formulation of movement patterns in any case depends on the movement space; therefore, it is important to distinguish geographic context from movement spaces. Movement spaces are models of space, in which movement is embedded, such as in the case of a road network in network space, for example. On the other hand, geographic context represents characteristics of the movement space, whereas an edge in a network representing a road can, for instance, have the property one-way. Different movement spaces allow the addressing of different types of question, whilst also requiring different types of method. For example, mobile phone data most often represents information relating to antenna connections, and are then modeled as irregular tessellation. Movement space modeled as irregular tessellation allows the assessment of movement on larger scales (e.g., commuter patterns); however, the analysis of instantaneous changes in movement on fine temporal scales is hardly possible.

### 2.2.2. Definitions of context

Context has a wealth of different meanings depending on the research area and application. Even within the research area of GIScience, there lacks clear and detailed consensus in relation to what context should encompass, since most often the conceptualization of context adapts to the purpose for which it is used. Some definitions of context can be found in the following list, which is not intended to cover all aspects of context, nor is exhaustive, but rather is a suitable set of definitions for the purpose of CAMA:

- ORELLANA & RENSO (2010, p.69): “[...] context is formed by a **partial set of parameters and values that is never complete, precise nor objective**, but is useful for movement representation and for reasoning on it.”
- ORELLANA & RENSO (2010, p.69): “[...] context is **usually linked to geographic environment**.”
- EDWARDES (2009, p.1): “[...] context is people and their direct engagement with the world [...]”
- EDWARDES (2009, p.2): “[...] location can be seen to describe the geographic embedding that binds position and different forms of information together.”
- DEY (2001, p.5): “Context is **any information that can be used to characterise 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.”
- DEY (2001, p.5): “A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user’s task.”

- SCHMIDT *et al.* (1999, p.893): “[...] **surrounding facts that add meaning.**”
- SCHLIT *et al.* (1994, p.85): “[...] **collection of nearby people, hosts, and accessible devices**, as well as to changes to such things over time.”

The above detailed descriptions illustrate that it is difficult to define context in a single sentence. In this vein, it can be seen that most of these definitions are quite general and not unique; however, in this thesis, context is understood to be a mix of the above highlighted statements. More concrete, the focus in this project can be identified on geographic context (*first order effects*), not taking into account the interaction between individuals (*second order effects*).

### 2.2.3. Modeling context

To facilitate reasoning surrounding movement in terms of geographic context, it is crucial to establish efficient management of context information and feasible context representation (BECKER & NICKLAS, 2004). Classically, GIScience provides representations of phenomena either as objects or fields. The literature review shows that, most often, geographic context in urban regions comprises discrete objects, such as buildings or Points Of Interest (POIs). These phenomena are modeled with geometries as points, lines and polygons, utilizing the concept of vectors. On the other side, animals are most likely influenced by both discrete but also continuous geographic external factors, where the latter are more likely to model as fields applying a raster representation.

Locations span beyond their representation as coordinate pairs; otherwise stated, making sense of space involves more than measuring distances, for example. First, movement is influenced by phenomena occurring in the surroundings of a certain location. Second, movement is not purely determined by parameters based on the physical properties of the surrounding space, but also, individuals’ perceptions of space are suggested as an important component in mind of understanding MPOs’ movement as a response to the geographic environment. Therefore, the notion of place seems to be important in terms of addressing geographic context, where place can be understood as ways in which people make sense of space (EDWARDES, 2009). Consequently, after modeling context with classical data sources as POIs, land use maps or Digital Elevation Models (DEMs), geographic context should also be addressed using, for example, User Generated Content (UGC) in mind of exploring how space is perceived by people.

Currently, there are no general mechanisms geared towards assessing context beyond location (SCHMIDT *et al.*, 1999). This research project aims at investigating how context could be modeled enabling CAMA. However, there are a number of different disciplines delivering countless approaches for making sense of data in terms of geographic context. Within the framework of this thesis, two examples of modeling geographic context on the level of generating semantic meaning from data are considered relevant and therefore are presented in this subsection: primarily, terrain is quantitatively captured based on DEMs since parameters as slope often

have an impact – particularly on animals’ movements – in rural areas; and secondarily, in the context of network-bound movement, semantics of urban places are assessed using UGC.

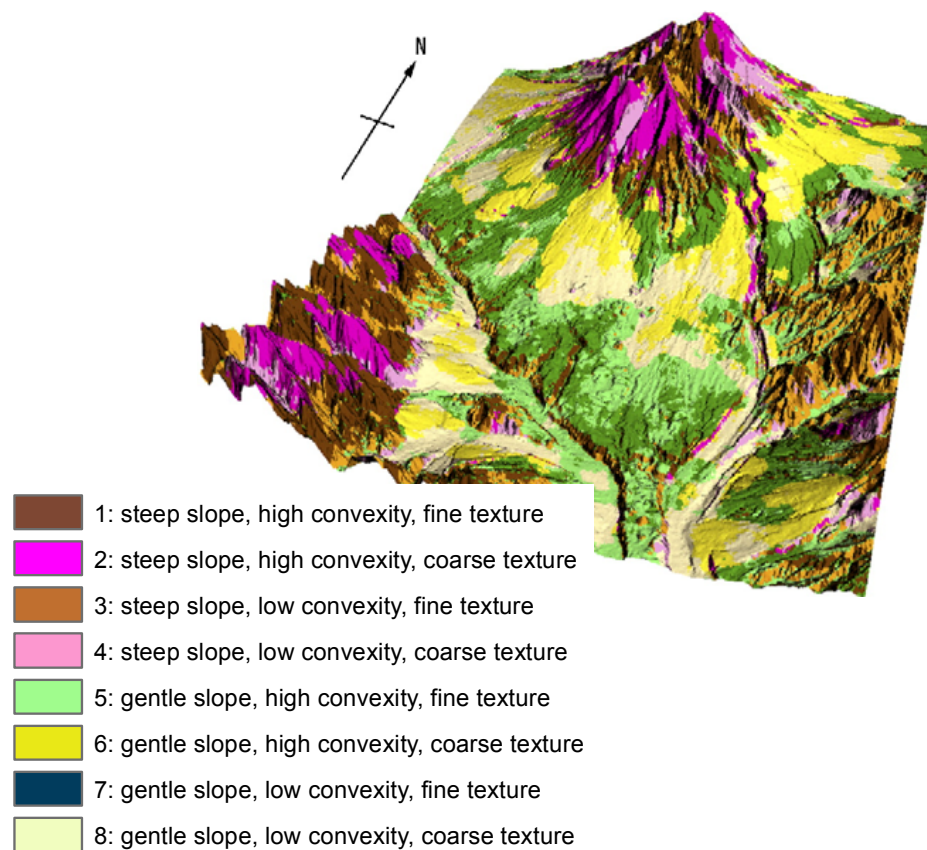
**Geomorphometry: capturing the terrain.** Geomorphometry aims at capturing the terrain in the form of quantitative descriptions and analyses of geometric and topologic characteristics of the landscape (RASEMANN *et al.*, 2004). Originally, geomorphometry has been seen to emerge from physical geography – or, in particular, from geomorphology – which explores landforms and the processes shaping them. Initially, geomorphometric investigations approached the topography by measuring the altitudes of mountain ridges and peaks, thus allowing for the relative comparison of different mountain systems. Moreover, linear structures, such as coast lines and rivers, were surveyed. Along with the availability of computers, quantitative parameters (e.g., slope) covering some terrain characteristics were formalized (EVANS, 1972; MARK, 1975*b*), and data structures for a digital representation of the topography in the form of DEMs were developed (MARK, 1975*a*; PIKE *et al.*, 2009). Besides terrain parameters – where, for instance, slope mostly represents the continuous phenomena of the land surface – geomorphometry also addresses discrete terrain features, largely on the basis of landform classifications (EVANS, 1972; PIKE *et al.*, 2009).

Although rasters representing data in rectangular data arrays have large storage requirements – and also considering the data structure does not account for the morphology of the study area – most algorithms in geomorphometry are conceptualized for raster data structures because a square-grid DEM significantly simplifies the implementation of geomorphometric methods (PIKE *et al.*, 2009). Most often, geomorphometric algorithms consider some neighborhoods (i.e. filter window) in relation to a specific pixel. Therefore, basically, a filter window is moved across the raster in order to achieve a quantitative measure in terms of a certain terrain characteristic for the current central pixel. In the case of slope, by default, a 3-by-3 window is moved across the raster, whereas North-South and East-West neighborhood differences (finite differences) are considered for their computation (e.g., HORN, 1981). With similar approaches, terrain parameters, such as curvature, aspect or roughness, can also be quantified.

The quantitative detection of landforms is an essential aspect of geomorphometry. The aim is basically centered on the identification of landscape features, such as mountain peaks (FISHER *et al.*, 2004) or valleys (STRAUMANN & PURVES, 2008). FISHER *et al.* (2004) propose a multi-scale approach for defining fuzzy set membership of geomorphometric classes, such as peak, which allows the quantifying of peakness. Moreover, STRAUMANN & PURVES (2008) present an object-based, top-down approach using thalwegs, where pixels are added to the valley region depending on a gradient threshold value (region growing). WOOD (1996) defines landscape features, including peak, ridge, pass, plane, channel and pit, with the terrain parameters slope and different types of curvature (supervised) in consideration of multiple scales. More recent work still formalizes landforms using the geometric signatures of the terrain:

for example, IWAHASHI & PIKE (2007) introduced an unsupervised nested-means algorithm taking into account slope, convexity and texture.

The iterative procedure presented by IWAHASHI & PIKE (2007) aims at classifying topography represented as raster-based DEMs into categories of surface form. Based on the various thresholds in form of the mean values of the three terrain parameters (slope, local convexity and surface texture), a classification into either 8, 12 or 16 landform classes is realized. Slope is calculated following the approach proposed by HORN (1981). Local convexity is measured with the application of a Laplacian filter, which is applied in image processing for edge detection. The number of peaks in a certain radius is used as a proxy for surface texture, whereas peaks are identified based on the difference between original and median-filtered DEM. Figure 2.2 illustrates the result of the automated classification of landforms on the basis of the geometric signature by IWAHASHI & PIKE (2007), encompassing 8 resulting classes for the Yatsugatake volcano in Japan.



**Figure 2.2.:** Landform classification with an automated procedure into 8 classes for the Yatsugatake volcano in Japan (IWAHASHI & PIKE, 2007).

As the above examples of capturing the terrain illustrate, the area of geomorphometry allows the quantitative assessment of geographic context with regard to gaining

the semantic insights of the surrounding topography of moving objects based on DEMs. Accordingly, in the case studies of this thesis, amongst others, geomorphometric approaches are applied for modeling geographic context. More specifically, the unsupervised algorithm for the detection of landform features by IWAHASHI & PIKE (2007) is considered in further investigations of moving animals.

**Semantics from user-generated content.** Nowadays, Web 2.0 allows users to generate (geographic) content accessible to many other users. Various platforms, such as OpenStreetMap, enables people to create geographic information available for almost everyone. More and more, people's engagement with the world in everyday life is reflected in content on the internet. In this context, citizens can be seen as sensors collecting and sharing information about their perspectives on the surrounding geographic setting (GOODCHILD, 2007). GOODCHILD (2007) refers to this kind of data as Volunteered Geographic Information (VGI) – a special case of UGC. As an example of VGI, beside photos, the photosharing platform Flickr contains photo metadata, which, in many cases, includes geographic coordinates and user-generated tags, most often describing the content of the photographs (e.g., landscape).

More generally, naïve geography aims at capturing and reflecting the knowledge people have relating to the surrounding geographic world (EGENHOFER & MARK, 1995). VGI, as it is accessible for instance on Flickr, covers naïve geography to some degree. People's perspectives on the surrounding geographic environment are crucial in terms of accessing geographic context that is relevant for moving individuals in urban areas, for example. From a theoretical point of view, semantics from UGC can support a better understanding of the moving objects' surrounding geographic context, assessing locations beyond mere geographic coordinates from a more semantic and social perspective, which relates to the concept of place.

Practically, various studies have shown that UGC – more specifically VGI – in form of Flickr tags can be considered a useful data source for extracting place semantics (e.g., EDWARDES & PURVES, 2007; RATTENBURY & NAAMAN, 2009). Another popular UGC source is the Location-Based Social Network (LBSN) Foursquare, which allows users to explore places of interest in close proximity to their current location. Furthermore, users can inform others about their current location through check-ins and also by composing reviews of certain venues. Check-ins are a suitable source of geographic information in terms of providing user profiles or, more specifically, activity patterns (VASCONCELOS *et al.*, 2012; NOULAS *et al.*, 2011). LONG *et al.* (2012) show, in their paper, that, with Foursquare check-ins, local geographic topics can be explored. Consequently, the geographic information relating to venues in Foursquare can be seen as a proxy of geographic context with certain social relevance.

GPS tracking is becoming more and more widely applied in the fields of urban design and planning, and, in addition to traditional methodologies of urban studies, it allows for gaining insights into pedestrians' actual spatio-temporal behavior (VAN DER SPEK, 2008). Understanding ways in which people use the physical environment is of major importance in urban design and planning VAN SCHAICK (2010). KRÜGER *et al.* (2014)

employed Foursquare as an information source of up-to-date POIs for contextually enriching movement data in a system of visual analytics. However, movement data has not been quantitatively related beyond overlaying some information layers; therefore, in order to achieve a better understanding of people's space use with regard to the physical environment in urban areas, it is critical that movement is related to the surrounding environment, whilst UGC suggests a potential representation of the urban environment.

### 2.3. From movement towards a better understanding of behavior

Generally, movement research deals with the two key problems: collecting movement data and deriving mobility knowledge from such data (GIANNOTTI & PEDRESCHI, 2008). GIScience has shown special expertise in developing methods for extracting high-level knowledge from large volumes of low-level spatial data; however, GISs remain weak in terms of handling the temporal dimension of movement data (MARK, 2003). This underpins the fact that methods generating new knowledge from movement data are not able to keep up with the development of ICT techniques for capturing this data.

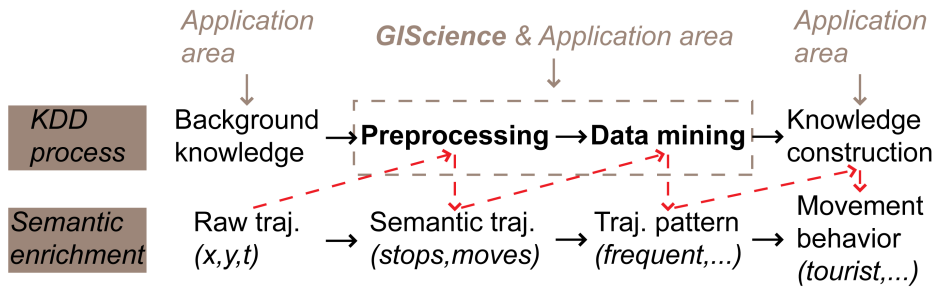
The significant challenge in analyzing movement data – specifically for dynamic collectives – is identified in terms of establishing possibilities to bridge the gulf between low-level data and high-level schemes through which we undertake interpretation (GALTON, 2005). The process of revealing human understandable structure (patterns) from huge datasets is established from KDD. The KDD process – and data mining in particular – seeks to bridge this gap between data and knowledge. Data mining is the core of the KDD process and is the most intensely discussed KDD step in the literature, generally including techniques such as classification, pattern discovery, regression and cluster analysis (FAYYAD *et al.*, 1996).

KDD aims at deriving knowledge from data in several steps, whereas this knowledge is intended to be valid, unexpected, relevant and understandable (FAYYAD *et al.*, 1996). MILLER (2008) identifies four major steps in the KDD process: background knowledge, preprocessing, data mining and knowledge construction (Figure 2.3). The KDD process is not intended to be strictly sequential: different steps can be revisited and repeated several times. In movement analysis, GIScience mainly contributes to the methodological part of the KDD process, namely preprocessing and data mining. In addition to the development of techniques for preprocessing and data mining, the application area specifically supports the KDD process in the steps of background knowledge and knowledge construction.

Corresponding to KDD, the GKD introduces this process, specifically with methods for spatial phenomena as movement. Therefore, the GKD process with movement data predominantly is based on trajectory reconstruction (preprocessing), knowledge extraction (geographic/mobility data mining) and knowledge discovery (reasoning) (MILLER, 2008). The most prominent techniques in mobility data mining are classification,



### 2.3. From movement towards a better understanding of behavior



**Figure 2.3.:** KDD process (according to MILLER, 2008, extended version) in relation to semantic enrichment process (BAGLIONI *et al.*, 2009).

frequent patterns and clustering (GIANNOTTI & PEDRESCHI, 2008). Therefore, the general framework of KDD can be applied to movement analysis, adopting methods that take into account the spatial and temporal dimension of the movement.

As these paragraphs illustrate, in an effort to enable the interpretation of movement data as traces of behavior, several steps need to be considered; therefore, in this thesis, three main dimensions are considered, namely preprocessing, methods for relating movement to geographic context and different types of scale in CAMA. In the following subsections, the literature available with regard to these three topics is reviewed.

#### 2.3.1. Preprocessing and semantic enrichment

The preprocessing of movement data is needed for data reduction (e.g., semantic trajectory compression, RICHTER *et al.*, 2012) and filtering purposes (e.g., filtering stops, LAUBE & PURVES, 2011). Data reduction techniques aim at decreasing the data size due to storage costs, whilst maintaining the therewith introduced error as small as possible. GPS measurement errors are reduced with trajectory filtering and smoothing methods, thus allowing for a more accurate estimation of high-level properties, such as e.g., speed or turning angle (LEE & KRUMM, 2011). On the one hand, data filtering deals with systematic errors, whereas on the other hand, data smoothing addresses random errors. Sensor noise and other technical and environmental factors cause errors in the GPS signal. Systematic errors often occur with an insufficient number of satellites (less than four), or satellites that are not ideally positioned in view. Therefore, systematic errors can be quantitatively filtered with the number of satellites and Position Dilution Of Precision (PDOP) values (SCHUESSLER & AXHAUSEN, 2009). Random errors are mainly caused by sensor noise, satellite issues, disturbances in the atmosphere and ionosphere, multi-path signal reflection or signal blocking. Multi-path signal reflection typically occurs in urban areas at the walls of buildings, resulting in jumps in GPS positions.

Mean and median filters smooth noise by taking into account measured values in a certain temporal neighbourhood (time window) of a specific GPS measurement.

These simple filters can only estimate what was directly measured, whereas more sophisticated filtering techniques, such as the Kalman filter, allow the estimation of other variables (e.g., speed), using models for the measurement noise as well as the dynamics of the trajectory (LEE & KRUMM, 2011).

Similarly, LAUBE & PURVES (2011) apply a mean filter on distances within a certain time window in order to separate stops and moves in trajectories using a distance threshold. This segmentation can be seen as preprocessing; in this case, this enables a cross-scale analysis of cow movements. A great deal of research has been performed in segmenting trajectories into stops, and moves in the context of semantic enrichment of trajectories. BAGLIONI *et al.* (2009) introduce the semantic enrichment process in the form of a model for deductive reasoning of trajectory patterns, allowing for interpretations of revealed movement patterns in terms of movement behavior (Figure 2.3). The semantic enrichment process is strongly related to the more generally formulated KDD process. Both approaches have been developed in the context of database research. The semantic enrichment process (Figure 2.3) is specific to movement analysis, and is focused on the results from raw data to movement behavior (BAGLIONI *et al.*, 2009), whereas the KDD process, on the other hand, illustrates, on a more general level, the processes needed in order to get from data to knowledge. Accordingly, the steps in the concept of KDD can be seen as processes concerned with getting from raw data to movement behavior in the semantic enrichment process (Figure 2.3, red dashed arrows).

SPACCAPIETRA *et al.* (2008) conceptually define stops and moves as semantic units of a trajectory, whereas a trajectory is seen as a sequence of stops and moves. In urban planning, researchers aim at understanding where people – and, specifically, tourists – stop and spend their time. In this context, semantic enrichment mostly is realized based on the segmentation of stops and moves, linking particular stops to domain-specific semantics, such as touristic areas of interest in an urban environment (e.g., ALVARES *et al.*, 2007). More concrete, ALVARES *et al.* (2007) developed an algorithm, known as Stops and Moves of Trajectories (SMoT), for extracting stops and moves from movement data based on candidate regions for stops (e.g., areas of attractions for tourists) and a minimal stay duration within a certain candidate region. From a preprocessing perspective, LAUBE & PURVES (2011) removed stops from movement data because, otherwise, scale effects on rather slow movement of cows would be masked by relatively long sequences of stops. Generally, this means that segmenting stops and moves can be seen as preprocessing, on the one hand, whereas, on the other hand, this may be viewed as the first step from raw data to semantic trajectories within the semantic enrichment process, where detecting stops and moves is more than preprocessing, and can be viewed as part of the data mining process.

The conceptual view on trajectories, including the formal definitions of stops and moves by SPACCAPIETRA *et al.* (2008), serves as a key foundation for many other studies, mainly in movement analysis, but not only limited to database structures. For instance, a conceptual data model for trajectory data mining is proposed by BOGORNY *et al.* (2010) for modeling trajectory patterns. Due to errors in the recording

### 2.3. From movement towards a better understanding of behavior

of GPS locations, the identification of stops in trajectories is not straightforward. Since stops are particularly prone to errors, stops often contain artifacts in form of movements actually not taking place (pseudo-movements) (ZIMMERMANN *et al.*, 2009). Therewith, different research areas with different backgrounds and research questions, using movement data with different properties, develop rather specific approaches for identifying stops in trajectories. ZIMMERMANN *et al.* (2009) illustrates that a density-based clustering approach incorporating spatial as well as temporal aspects of the movement (Trajectory-Ordering Points To Identify the Clustering Structure (T-OPTICS)) allows the identification of stops in trajectories comprising artifacts and movements at different speeds (different transportation modes).

In eye-tracking experiments, eye-movements on the screens of computer users are tracked, where the identification of fixations is of particular interest. SALVUCCI & GOLDBERG (2000) present a taxonomy of approaches for extracting fixations in eye-movement data, which is also valid for the data of moving individuals. Essentially, this taxonomy distinguishes different algorithms based on spatial and temporal criteria. Spatial characteristics allow the separation of velocity-based, dispersion-based and area-based algorithms. With temporal characteristics, two types of algorithm are identified: duration sensitive and locally adaptive. In this regard, BUCHIN *et al.* (2013) segment the movement of migrating geese into movement states (stop, flight) through the use of spatio-temporal criteria, where geese are supposed to stop when staying in a 30-kilometer radius for more than at least 48 hours, with the movement behavior of flying quantified with the variation in heading and a minimum speed over a certain time interval. This example, as well as the previously mentioned taxonomy by SALVUCCI & GOLDBERG (2000), together mirrors the limitations of most of the approaches in identifying stops in movement data. Often, a-priori knowledge is needed. More concrete, in many methods for detecting stops, spatial and/or temporal thresholds or predefined potential areas of suspension are required. As a consequence of these limitations, ORELLANA *et al.* (2010) developed an exploratory statistical approach able to detect stops in movement data based on the Local Index of Spatial Association (LISA). This methodology, complete with the LISA, proved to be robust in terms of identifying patterns of suspension for different type of movement without the use of any thresholds or predefined areas of potential stops. However, threshold-based approaches still enable assessing stops and moves on different temporal scales (different durations). To the author's knowledge, stops and moves of different duration have not often been addressed in movement analysis.

On a more general level, identifying stops and moves in movement data is related to segmentation, semantic trajectories and travel mode classification. In other words, the semantic enrichment and the segmentation of trajectories go beyond the detection of stops and moves including a whole range of concepts and methodologies (e.g., for travel mode classification), improving the understanding of why and how people and animals move (PARENT *et al.*, 2013). The segmentation of trajectories can have different aims, as for example revealing typical, as well as unexpected behavior (e.g., SESTER *et al.*, 2012 applying graph theory), the extraction of significant places (e.g., BHATTACHARYA *et al.*, 2012 based on movement parameters, or UMAIR *et al.*, 2014

considering data point density and stay duration), or the detection of travel modes (e.g., ZHENG *et al.*, 2010 on the basis of supervised learning).

Literature from different research areas, such as GIScience, movement ecology or urban planning analyzing movement data of different moving entities (various animals, or people with different transportation modes), illustrates that, for knowledge discovery in movement data, preprocessing is well established but, most often, is specific with regard to the movement data and the species under investigation. The identification of stops and moves in trajectories has been the focus of particular attention. In preprocessing, the detection of stops is used to filter error-prone GPS fixes. Preprocessing and semantic enrichment are closely related, with both originating in database research. As some selected studies illustrated, semantic enrichment does not exclusively focus on the segmentation of stops and moves. Semantic enrichment is a process that can have different purposes. On a more general level, semantic enrichment aims at adding semantic meaning to the raw movement trajectories, facilitating a better understanding of movement in terms of behavior. The identification of stops and moves can be seen as part of semantic enrichment or as part of preprocessing in movement analysis, whereas preprocessing is a step within the process from data to knowledge. Most remarkably for this thesis, the detection of stops and moves as part of preprocessing, and its effects on the relation of movement to its surrounding environment, has only rarely been considered in movement research thus far.

### 2.3.2. Movement in its embedding geographic context

Related work from three different areas is relevant specifically to CAMA: recently emerging work on including context in GIScience, habitat and home range analysis in movement ecology, as well as semantic enrichment of trajectories in moving object databases.

**GIScience.** In GIScience, ample research has been performed in the field of detecting arrangement patterns according to geometric properties. Such geometric patterns are supposed to reflect specific movement behavior, such as leadership (ANDERSSON *et al.*, 2008) or flock (BENKERT *et al.*, 2008). Aiming at explaining why patterns emerge, some authors claim that movement patterns evolve due to the variability in the embedding geographic context (BITTERLICH *et al.*, 2009; LAUBE, 2009). Consequently, initial research on CAMA emerge, complementing arrangement and shape studies.

For example, ANDRIENKO *et al.* (2011b) introduced an event-based conceptual model for CAMA allowing for the identification and analysis of relations between moving objects and elements of the spatio-temporal context. Special emphasis is placed on movement events (that is, part of trajectories) and their spatial and/or temporal proximity relations to spatial context features – ultimately aiming at visualizing the relations and thereby supporting a human analyst in inferring relations between movement and context. Similarly, BUCHIN *et al.* (2012) also presented a model for geographic context that allows the integration of environmental factors into algorithmic movement analysis; in their case, a context-aware trajectory similarity measure. Their

### 2.3. From movement towards a better understanding of behavior

model explicitly includes vector, graph and field representations of the surrounding environment. In yet another study, animal movement is linked to spatio-temporal dynamic variables (e.g., wind speed) through interpolation in space and time (DODGE *et al.*, 2012). Lastly, in an early piece, IMFELD presented already in 2000 the concept of RDFs for capturing the surrounding environment around fixes on multiple scales. Most research in movement analysis has examined the movement of individual MPOs (e.g., DODGE *et al.*, 2008), typically collected by GPS receivers. The most atomic primitive is a fix – a time-stamped point primitive  $(x, y, t)$ . A more holistic overview of alternative concepts for the modeling of movement is presented in Chapter 4.

GIScience has a strong background in modeling spatial phenomena. Spatial data and the corresponding geographic context is modeled for accurately representing geographic reality. Classic concepts for representations of space within GISs are objects or fields, which simplify geographic phenomena to different abstractions, representing specific aspects of the surrounding environment. Data modeling has a direct impact on the users' point of view of the geographic world, since data models – and models overall – are limited representations of the real world (GOODCHILD, 1992). The modeling of geographic context addresses the debate of the object-field dichotomy in the field of GIScience. Otherwise stated, the representation of geographic space can be approached from two different perspectives of reality – either from the object or from the field view. For instance, the dissertation of IMFELD (2000) and the paper of BUCHIN *et al.* (2012) together show that, in order to cover different sorts and perspectives of spatial phenomena, geographic context is modeled predominantly as point, line, polygon or field.

Furthermore, GIScience has a long-standing tradition in relating different sources of geographic information. The classic GIS techniques for combining different spatial layers are generically captured within the name of *map overlay* (O'SULLIVAN & UNWIN, 2010). Further, buffering is a useful transformation in a GIS for identifying objects and areas within a certain distance of a specific object (LONGLEY *et al.*, 2010). Since CAMA is not only based on spatial data but also on spatio-temporal data representing movement, a challenge is recognized in terms of preserving the characteristics of movement in the relation of movement and geographic context. Both static (e.g., land cover) as well as dynamic (e.g., temperature) geographic context have a spatial component, thus meaning that a spatial intersection is always part of the relation of movement and geographic context. Therefore, movement-context relations basically build on classic GIS methodologies as *map overlay* and buffering. Actual implementations of the interrelation across the fields of GIScience, movement ecology and KDD are based on the concepts and methods of the GIScience toolbox, including *overlay* and *spatial intersection* (“What is the value of a certain attribute at a specific location?”), notions of *proximity* (“What is the distance to the next object of a certain type?”), as well as *spatial statistics* (“What is the quantity of a certain parameter within an area around a MPO?”) (BURROUGH & McDONNELL, 1998; IMFELD, 2000). RDFs involve nested circular buffers with variable radii for the different context scales (small rings for local neighbourhood, larger rings for regional neighbourhoods), where the relative distribution of specific context values (e.g., forest

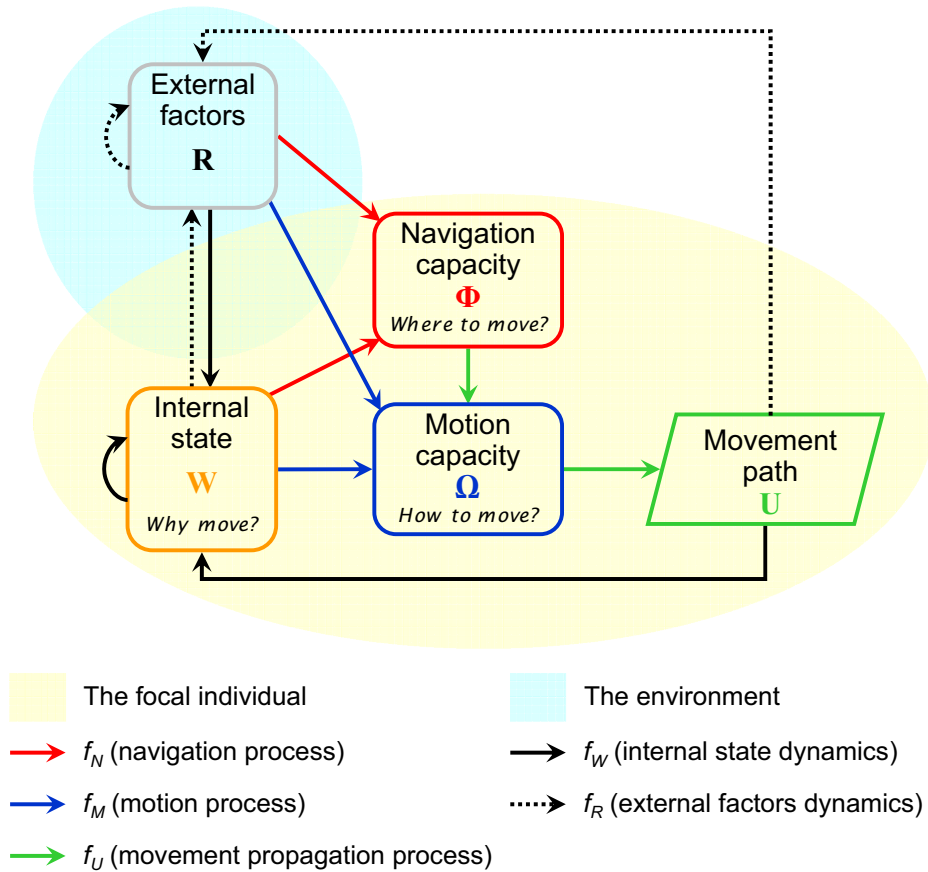
vs. meadow for land cover) is quantified for variable radius buffers. Different spatial scales using the RDF approach accounts for the fact that the radius parameter for the vector-based point buffers is often chosen rather arbitrarily.

Work addressing CAMA in GIScience thus far has mostly focused on conceptual characterization, and the modeling and capturing of spatio-temporal context. In contrast, an integral perspective not only considering context models but equally directing emphasis to alternatives of conceptually modeling the movement itself and the options for quantifying the emerging movement-context combinations is still lacking.

**Movement ecology.** NATHAN *et al.* (2008) present a movement ecology paradigm for unifying movement research, identifying external factors, internal state, navigation capacity and motion capacity as crucial components influencing behavior and, consequently, an individual's movement path (Figure 2.4). Movement ecology has a long tradition in habitat and home range analysis, investigating animals' space use in relation to various environmental factors. For instance, ecological-niche factor analysis is applied in an effort to describe an organism's habitat selection in terms of external environmental factors and to explore realized niches within the available space (BASILLE *et al.*, 2008). Moreover, research investigating habitat use, typically tests for seasonal differences or otherwise examines relations to food abundance or habitat availability (home range size) (AEBISCHER *et al.*, 1993). In general, spatial range methods aim at delineating areas of the movement's occurrence, thus allowing investigation of mobility and space use (LONG & NELSON, 2013). More concrete, a home range is characterized by an interplay between the environment and an animal's understanding of that environment (POWELL & MITCHELL, 2012; DEMŠAR *et al.*, 2015). From a conceptual point of view, spatial ranges are most commonly modeled with polygons or fields, and consequently are some form of aggregation of the movement data, with the purpose of delineating home ranges, e.g., using MCPs (POWELL, 2000). Such polygon-based models of movement, besides MCP, include approaches as harmonic mean, Voronoi polygons, characteristic hull and polygon contours derived from density fields (LONG & NELSON, 2013). Different variations of polygon-based models are discussed in the literature. For example, GETZ & WILMERS (2004) extend the MCP approach by incorporating the local nearest-neighbours for estimating home range areas, and in this regard for constructing Utilization Distributions (UDs). In this context, UD is a formal way to represent home ranges by capturing the probability of encountering an animal in a given location (DEMŠAR *et al.*, 2015). DOWNS & HORNER (2009) present a characteristic-hull approach for home range estimation, based on Delaunay triangulation, whereas resulting home ranges can have concave edges and unoccupied space within their interiors, notably – in contrast to the MCP method.

Moreover, polygon contours can be derived by various field-based movement models, whereas the field-models of movement overall are based on spatial or spatio-temporal density estimations representing individuals' home range with a UD. WARTMANN *et al.*

### 2.3. From movement towards a better understanding of behavior



**Figure 2.4.:** Movement ecology paradigm (NATHAN *et al.*, 2008, p.19054).

(2010) quantify spatial needs and space use patterns using home range calculations with KDE. As an exemplary outcome of such a habitat study, they could show that context, in the form of food abundance (here fruit availability), has a significant effect on the per-day path length. DOWNS & HORNER (2012) introduce KDE based on distances on a network computed with Delaunay triangulation (KDE-DT), with such a network representing typical or average pathways. Kernel Density Estimation-Delaunay Triangulation (KDE-DT) suggests the production of more accurate home range estimates than traditional KDE.

Pure KDE has the shortcoming that resulting fields only reflect the spatial distribution or footprint of fixes, but not the temporal characteristics of movement; otherwise stated, when an individual revisits a certain location, high densities appear in the resulting density surface, irrespective of the sequence of the visits. Therefore, classic KDE is limited in terms of modeling movement, since not only spatial but also spatio-temporal proximity is required in order to adequately represent the sequential aspects of movement. Furthermore, importantly, the kernel bandwidth for computing

KDE has an influence on the resulting density surface. A Time-Geographic Density Estimation (TGDE) approach (DOWNS, 2010) extends on traditional KDE, taking into account a maximum speed value for modeling potential path area, within which intensity values are computed. Applying this maximum speed parameter appears less controversial than the difficult choice of a kernel bandwidth with classic KDE. The time-geographic concept, within home range estimation, proves to be a suitable approach for analyzing movement data with unsystematic and irregular sampling intervals; the resulting home range estimates, however, are significantly influenced by the maximum speed parameter (DOWNS *et al.*, 2011). Although TGDE accounts for the amount of time seen to elapse between MPOs, in essence, the representation continues to be based on spatial proximity of points. Brownian bridges describe movement incorporating both spatial and temporal aspects. More specifically, the BBMM is a stochastic model of movement, where the probability density value in a particular area is conditioned on the temporal order of visited locations, as well as the time gaps between visits and the location error of the GPS device (HORNE *et al.*, 2007). The relation of polygon- and field-based movement models to geographic context is realized, in the main, with a simple *overlay*, and the distribution of context values within these areas of space use is investigated.

Resource selection aims at understanding animals' behavior in terms of habitat use by relating areas of presence and absence – or even UD – to resource attributes with mathematical models, which are also referred to as Resource Selection Functions (RSFs). Essentially, RSFs are developed in an effort to mathematically describe the relation between environmental factors and mostly spatial footprints of movement. For instance, MILLSPAUGH *et al.* (2006) relate kernel-based UD estimates to environmental factors through the use of multiple regression. As an exception, HUNTER's work 2007 has to be mentioned since he links speed, as a spatio-temporal characteristic of movement, to the activities of bears (e.g., foraging and locomotion), accordingly applying a resource selection model using logistic regression in order to characterize environmental variables for the activities of foraging and locomotion. In comparison, our research does not explicitly contribute to mathematical descriptions of relations between movement and geographic context, but rather assesses different opportunities of relating movement and geographic context on a methodological and more GIS-related level.

Whilst movement ecology has developed a strong theory for analyzing space use and exploring ecological niches, investigating space use in relation to external factors is most often static, ignoring temporal aspects of movement, meaning that only the spatial footprints of movement are considered. Movement is largely modeled as sample points, enclosing polygons (e.g., home range defined by MCP) or fields (e.g., home range defined by raster as a result of KDE), whereas fields are often simplified to contour polygons, thus ignoring the UD in the relation process. Although a variety of concepts is used for modeling movement, much less emphasis is placed upon alternative methodologies for relating movement to external factors.



**Semantic enrichment in moving object databases.** KDD provides a means for deriving behavior knowledge from raw trajectory data. Researchers working with Moving Object Databases (MODs) argue that movement adopts a particular meaning in a geographic environment (BAGLIONI *et al.*, 2009), and emphasize that movement patterns should not be identified with consideration to only geometric properties of a trajectory (ALVARES *et al.*, 2007): “Movement is behaviour, but patterns are not” (ORELLANA & RENSO, 2010, p.67). KDD and data mining techniques in particular allow the identification of movement patterns, but linking specific movement behavior to such patterns remains problematic at best (BAGLIONI *et al.*, 2009). Consequently, aiming at the interpretation of movement patterns and hence deriving knowledge about movement behavior relevant to a specific application domain, requires that trajectory data is semantically enriched (BAGLIONI *et al.*, 2009). With focus on relating movement to its embedding geographic environment within semantic enrichment, points and segments can be identified as the basic, underlying concepts for modeling movement, whereas such models representing movement are related to points of interest as part of the geographic context in the form of points or polygons. Accordingly, movement-context relations are most commonly quantified using approaches as “distance to” or “point in polygon”.

Semantic enrichment and our research in CAMA both aim to put movement or movement patterns into context, aiming towards understanding individuals’ movement as behavior. In the semantic enrichment process, however, methods for relating movement to its underlying context are most often based on movement modeled as sample points or segments. In contrast, throughout the course of this thesis, the argument is posited that alternative models incorporating the spatio-temporal characteristics of movement and respective relation methods can be used in order to garner a more dynamic perspective on CAMA.

#### 2.3.3. Scale in movement analysis

With regard to scale in movement analysis, work is relevant from the research fields of GIScience and movement ecology, with both fields recognized as contributing to the theory of scale effects within spatio-temporal analyses. Here, the focus lies on scale effects when relating movement to its geographic context in the form of the surrounding environment.

**GIScience.** Scale is a quintessential concept of every geographic analysis, whether spatial, temporal or thematic scale (MONTELLO, 2001). Although movement simultaneously is a spatial and temporal phenomenon, most often, scale effects are neglected in the context of movement analysis. Since modeling movement at multiple granularities is crucial for processing large amounts of data, HORNSBY & EGENHOFER (2002) present a framework for shifting across different scales of a moving objects’ lifeline. In their paper, the authors illustrate that temporal and spatial scales of movement are directly linked; in other words, a change in temporal scale of a movement trajectory has a direct impact on the spatial scale of the movement. LAUBE & PURVES (2011)

conducted a sensitivity analysis with regard to the temporal scale, demonstrating that the temporal scale has a significant influence on the computation of movement parameters based on the trajectory, such as speed or turning angle. SOLEYMANI *et al.* (2014) introduced a methodology for the characterization and classification of behavior based on movement data, considering spatial, as well as temporal scale.

CAMA aims at relating movement to its surrounding context in an effort to better understand movement patterns as responses of the moving individuals to its surrounding context. In particular, if geographic context is represented in categories (e.g., land cover as meadow, forest, etc.), on the context side, spatial, as well as thematic, scale then becomes relevant for CAMA. It is common for buffer analysis to be applied in order to relate movement trajectories to their surrounding geographic context, whereas spatial scale, on the other hand, is assessed by varying the buffer size. For example, IMFELD (2000) introduces RDFs in an attempt to capture the geographic context in relation to the movement on different spatial scales using simple point buffers with varying radii.

To conclude, the theory clearly shows that, when relating movement to its surrounding geographic context, spatial, temporal and thematic scale effects matter. However, to the authors' knowledge, the three types of scale – and, more specifically, their interplay – have not often been investigated, thus far, in CAMA within GIScience.

**Movement ecology.** Animals' behavior and their movement are determined by internal and external factors (NATHAN *et al.*, 2008), meaning that movement patterns emerge due to internal and external processes. LEVIN (1992) identifies a main challenge in ecology as understanding patterns observed at one level of detail in terms of processes operating on other scales. More concrete, CAMA faces the problem of understanding movement patterns at a specific scale emerging due to processes on other scales, where such processes can be referred to context. Otherwise stated, movement ecologists are essentially interested in understanding movement patterns as a response to processes, such as internal and external factors. Consequently, in the relation of movement to its surrounding geographic environment, scale plays a crucial role.

In different research projects within movement ecology, the sinuosity of movement trajectories has been assessed using fractal dimensions, reflecting movement as a response to the environment on different spatial scales (FRITZ *et al.*, 2003; NAMS, 2005; WEBB *et al.*, 2009). FRITZ *et al.* (2003) identify three nested scale-dependent domains, where seabirds adjust sinuosity to different environmental conditions. WEBB *et al.* (2009) show in their paper that fractal dimensions facilitate better understanding of deer movement paths affected by environmental (e.g., monthly rainfall) and behavioral (e.g., sex) factors. FRYXELL *et al.* (2008) demonstrate that, depending on the spatio-temporal scale at which movement is studied, different movement modes can be discovered.

BÖRGER *et al.* (2006) apply a generalized mixed-effects model that allows for explaining variance in home range size with variation in spatial, temporal and individual-level

factors. More specifically, the authors demonstrate that the temporal scale at which movement data is collected (biweeks, seasons and half-years) determines the patterns in the time series of home range sizes that can be discovered. Furthermore, home range size has been found to significantly vary with different dominating land cover within such home ranges. The variation of the classification scheme (thematic scale) and different home range definitions resulted in similar outcomes.

The literature review focusing on movement ecology and scale illustrates that spatial, temporal and even thematic scale effects are considered in numerous studies. Again, the majority of studies investigate sensitivities to spatial or temporal scale, although their interdependency has been very rarely examined.

## 2.4. Research gaps and research questions

In this section, the research gaps identified in the literature review are summarized, with the research questions resulting from the identification of such gaps presented.

### Preprocessing.

- Preprocessing is well established in various research fields, but most often is specific to movement data and the species under examination. A common ground on a more general level, as a workflow for the preprocessing of movement data, is not realized for the most part.
- The effects of preprocessing on the actual analysis of movement, and the movement-context relation in particular, has been only rarely considered so far.

→ **Research Question 1 (RQ1):** What is the influence of commonly used preprocessing steps (e.g., segmentation, map matching) on context-aware movement analysis, and how can it be quantitatively evaluated?

### Movement models.

- Alternatives of conceptually modeling the movement have not been a topic of focus so much in GIScience, with the conceptual characterization of purely spatial phenomena garnering most attention.
- Various concepts for modeling movement are applied in movement ecology (e.g., polygon or raster); often, however, only the spatial footprint of the movement is considered.
- In semantic enrichment, other models of movement are taken into account than in movement ecology. Nevertheless, the spatio-temporal characteristics of movement are not integrated within such conceptual models.
- It is rare for more than one or two models of movement to be used in studies of movement analysis.

→ **Research Question 2 (RQ2)**: How should movement and the geographic context embedding movement be modeled allowing for a quantitative relation between the two?

**Movement-context relation methods.**

- Different models for representing movement (refers to RQ2) suggest different options for quantifying the movement-context relations. Various methodologies for relating movement and geographic context are most often not taken into account.
- Different research fields tend to use different models for representing movement. Therefore, an entire range of potential movement models (also with some shortcomings, see paragraph movement models) is already applied, although various methods for relating movement and geographic context are rarely taken into account.

→ **Research Question 3 (RQ3)**: How can movement and explicitly its spatio-temporality (not only its mere spatial footprint) and the geographic context embedding this movement be quantitatively interrelated?

**Scale.**

- In the procedure of relating movement and geographic context, different types of scales (spatial, temporal and thematic) and their interdependencies have only rarely been specifically investigated thus far.

→ **Research Question 4 (RQ4)**: How sensitive is the computation of a quantitative relation between movement and its embedding context to a systematic variation of the temporal, spatial and thematic scale of analysis, and can interdependencies between the different scale dimensions be identified and quantified?

## Chapter 3.

# Preprocessing and semantic enrichment of movement data

In this chapter, preprocessing as a step of the KDD process is tackled in view of movement analysis. Firstly, a novel preprocessing method for segmenting a trajectory into stops and moves is proposed. Furthermore, a preprocessing workflow is presented with the inclusion of the segmentation of stops and moves. For validation purposes, the proposed preprocessing workflow is then applied on a data set of people moving in a metropolitan region. In Section 3.4, a method is presented that allows for quantifying the effects of preprocessing on movement analysis. Finally, the effects of preprocessing (in the form of filtering stops) on the computation of movement parameters (e.g., speed) and the relation of movement characteristics to its surrounding geographic environment are explored, using data of people moving in an urban setting, as well as the movement data of ungulates in a protected area.

### 3.1. Preprocessing: segmenting stops and moves

This section presents a novel preprocessing method concerned with segmenting a trajectory into stops and moves. Owing to the fact that the identification of stops and moves in GPS data is a prominent example in preprocessing and semantic enrichment, the main focus lies on the methods for segmenting a trajectory into stops and moves. A segmentation approach for identifying stops and moves is proposed in the subsequent paragraphs. Moreover, a more holistic preprocessing workflow, grounding on the mentioned segmentation method, is introduced in the next section and accordingly tested in Section 3.3.

A segmentation method for the detection of stops and moves is introduced, allowing different types of stops and moves to be defined with a threshold-based approach (Figure 3.1, *segmentation*). The Figure 3.2 shows an illustrative example of the segmentation procedure, including the quantitative definition of stops and moves, as well as the handling of too short stops and moves. Essentially, an MPO is classified as a stop or move based on mean speed, turning angle and duration. Mean speed is an average value of speeds on different temporal scales (different time windows around a GPS fix), and conceptually follows the approach introduced by LAUBE & PURVES (2011), where a mean distance method was presented within a similar research context.

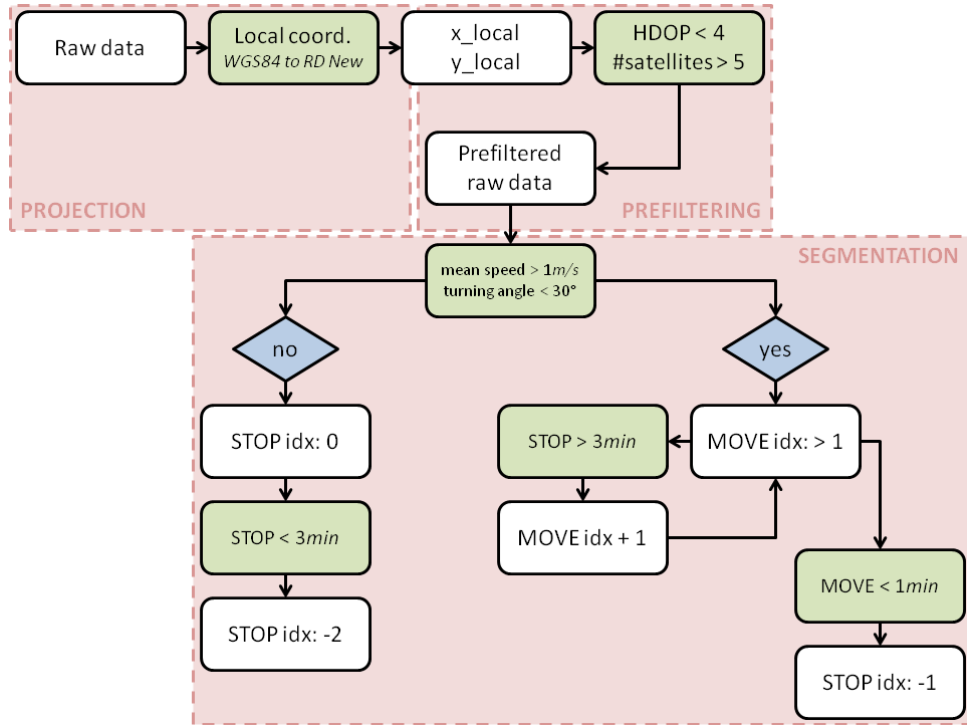


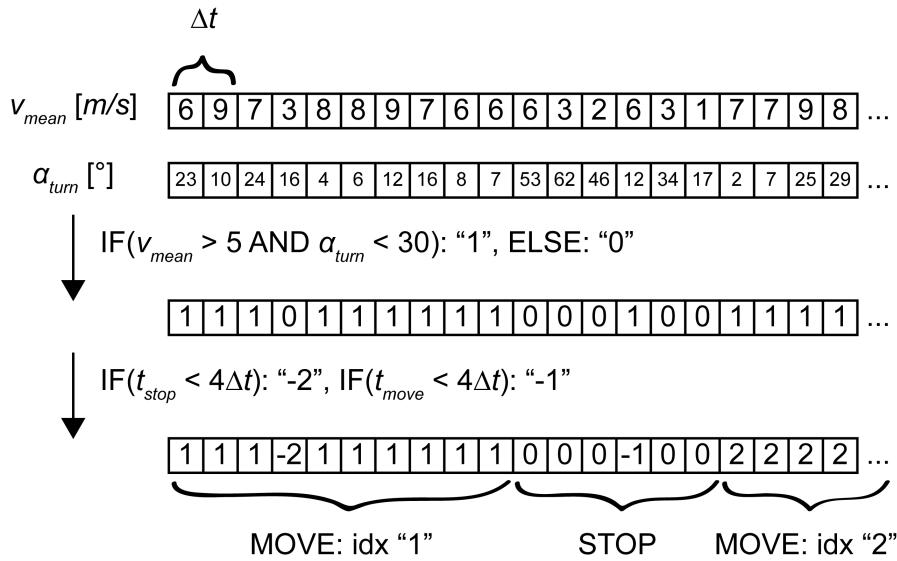
Figure 3.1.: Process chain for the segmentation of stops and moves (author’s own graph).

Practically, the procedure for detecting stops and moves from the prefiltered raw data consists of three iterations (Figure 3.1, *segmentation*). In the first step, every MPO, which meets the demands of a move based on mean speed and turning angle, is assigned the index “1”; otherwise, an MPO is supposed to be a stop and is labeled with “0”. In a second iteration, durations of stops (sequences of “0”) and moves (sequences of “1”) are determined. Finally, relatively short sequences of stops and moves are identified, using two different thresholds to characterize the minimal duration of a stop and a move. Sequences of “0” and “1” that are too short (smaller than the stop or move threshold) are labeled with “-2” and “-1”, respectively. Sequences of “-2” refer to stops that are too short to split a move, and consequently are considered part of the current move. Likewise, sequences of “-1” are not allowed to disconnect stops. Long enough stops split moves into separate ones, whereas a new move gets a new index (increased by one, new index = old index + 1). The different stop-move indices can be summarized as follows:

- Index -2: sequence of MPOs, where each MPO meets the demands of a stop, but the sequence is too short to split a move → not considered as stop, part of a MOVE

### 3.2. From preprocessing to semantic enrichment: a preprocessing workflow

- Index -1: sequence of MPOs, where each MPO does not meet the demands of a stop, consequently is a move, but the sequence is too short to split a stop  $\rightarrow$  not considered as move, part of a STOP
- Index 0: sequence of MPOs, where each MPO meets the demands of a stop and the sequence is long enough to be considered as a stop  $\rightarrow$  STOP
- Index  $> 0$ : sequence of MPOs, where each MPO does not meet the demands of a stop, consequently is a move and the sequence is long enough to be considered as a move  $\rightarrow$  MOVE

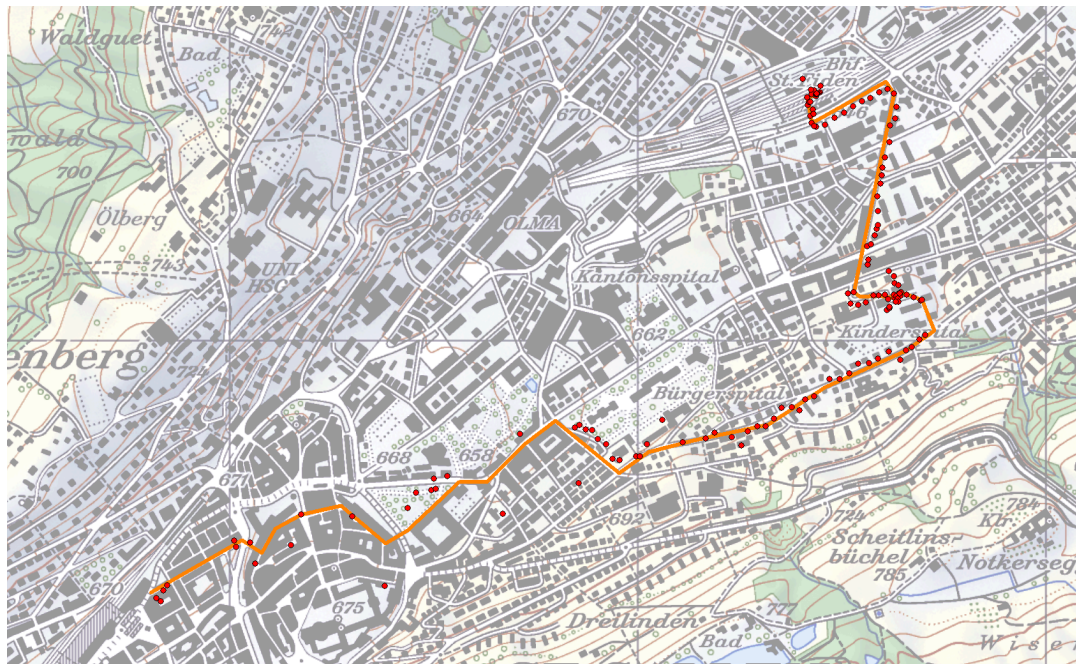


**Figure 3.2.:** Illustrative example of the process chain for detecting stops and moves (author's own graph).

### 3.2. From preprocessing to semantic enrichment: a preprocessing workflow

In case the MPO stops (theoretically, speed equals zero), the GPS signal is prone to relatively large scattering, which results in artifacts (pseudo-movement), since in practice speed does not equal zero. Thus, stops can be seen as locations or areas, which are susceptible to errors. Figure 3.3 illustrates the spreading in stopovers, with the accuracy of GPS measurements depending on building density, amongst others. The detection of stops and moves can serve for the filtering of noise in the movement data (LAUBE & PURVES, 2011); however, as the literature review of database research with semantic enrichment demonstrates, finding stops can span beyond the

identification of errors (e.g., ALVARES *et al.*, 2007), namely adding semantic meaning to the trajectory that facilitates a better understanding of the movement in relation to its surrounding environment (e.g., tourists checking in at a hotel). This means that, in the context of semantic enrichment, the detection of stops and moves can be more than a preprocessing step in the sense of trajectory segmentation. As within the related work in the Subsection 2.3.2 states, the segmentation of trajectories aims at partitioning movement paths into subtrajectories that are uniform with regard to various movement characteristics (e.g., speed, turning angle). Therefore, in summary, segmentation adds semantic meaning to subtrajectories towards achieving a better understanding of movement patterns as behavior. In case of the process of semantic enrichment, segments of stops in particular allow insights into reasons for resting, where in urban areas, for instance, behaviors of tourists can be assessed.



**Figure 3.3.:** Self-tracking in the city of St. Gallen, Switzerland (author’s own data, 2011). Orange: actual movement path, red: GPS measurements (author’s own data, 2011), map: “Landeskarte 1:25’000” (“Bundesamt für Landestopographie”, 2011).

In order to take this idea of semantic enrichment a step further, a workflow for preprocessing and semantically enriching movement data – based on the integration of preprocessing steps discussed in the previous section – is introduced. The segmentation part of the procedure can easily be replaced by any other segmentation approach. This procedure aims at establishing a concrete means for preprocessing movement data and accordingly illustrates the smooth transition from preprocessing to semantic enrichment (Figure 3.1). Furthermore, the above introduced segmentation



### 3.3. Stops and moves: application of the preprocessing workflow

methodology aims at contributing to a better understanding of stops and moves on different temporal scales.

**Projection.** As a first preprocessing step, the coordinates within the movement data are projected to a local coordinate system since it is common for GPS data to be delivered in longitude/latitude coordinate pairs in a geographic coordinate system (WGS84). Geographic coordinate systems are based on the globe – not projected to a local coordinate system – and therefore are based on spherical geometry, most often in degrees. As soon as distances between GPS fixes are required – such as for computing speed values, for example – it is simpler to choose a planar local coordinate system, where distances are measured in meters. Local coordinate systems are not feasible for areas of investigation on a global scale, but are appropriate for relatively small study areas (e.g.,  $100km \times 100km$ ) because, for larger areas, any kind of projection is not accurate enough.

**Prefiltering.** Dilution Of Precision (DOP) values as Horizontal Dilution Of Precision (HDOP) and PDOP, as well as the number of satellites, often are co-registered with the GPS data and are common parameters for assessing potentially inaccurate GPS fixes, thus serving as an initial filtering of the movement data. A PDOP or HDOP smaller than four mirrors relatively low spreading in GPS measurements. Moreover, the more satellites that contribute to the measurement of the position of the GPS receiver, the more accurate the GPS position is supposed to be. The minimum amount of satellites in view for GPS localization is four. As a consequence, the number of satellites chosen was greater than five in our preprocessing procedure.

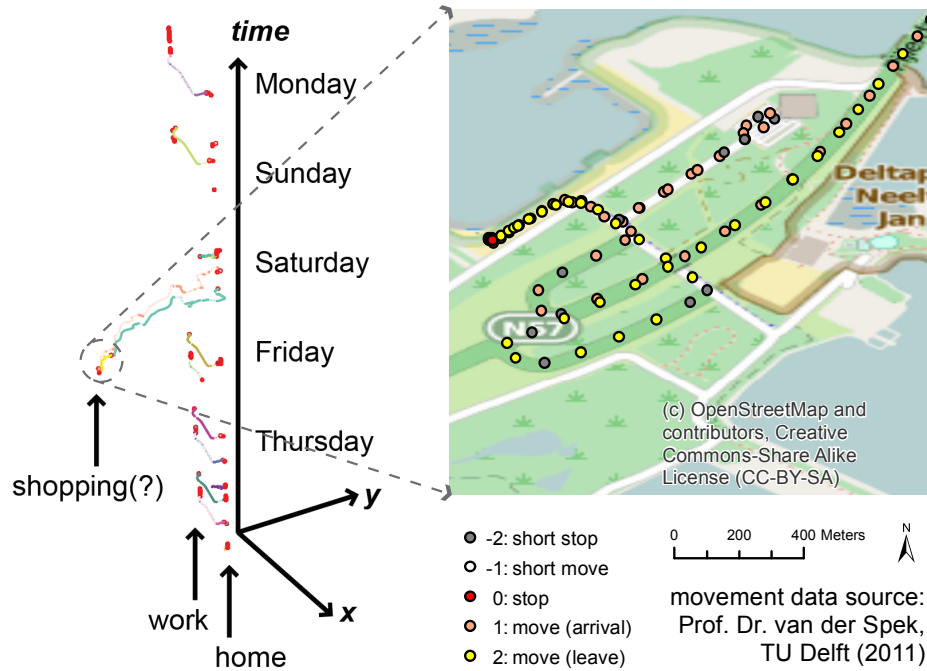
The two steps of **projection** and **prefiltering** of the raw movement data are followed by the **segmentation** approach. In Section 3.3, this preprocessing workflow is applied on the movement data of people moving mostly on street or railway networks in and around the city of Rotterdam.

### 3.3. Stops and moves: application of the preprocessing workflow

Movement data provided by the research group around Prof. Dr. Stefan van der Spek, from the faculty of architecture and the built environment at Delft University of Technology, is used in order to test the proposed preprocessing procedure in Section 3.2. The movement data was collected from end of April until the beginning of May, 2011. Approximately 50 people were asked to track their daily activities with the use of a GPS receiver for a period of approximately one week. This data aims at supporting household research, detailing participants recording their activity patterns and when they leave their home, etc. (e.g., home to work or weekend activities as shopping). The data sets were sampled at five *second* intervals.

The projection of the geographic coordinates (WGS84) to the local coordinate system (RD New) is realized with FME (Safe Software, version Desktop 2011). The

thresholds for mean speed, turning angle and minimal duration for stops and moves used in this case study correspond to the values in the workflow (Figure 3.1,  $v_{mean} > 1m/s$ ,  $\alpha_{turn} < 30^\circ$ ,  $t_{min,stop} = 3min$ ,  $t_{min,move} = 1min$ ). Besides using FME for the projection of the coordinate system, JAVA (version 1.6.0) was used to implement the rest of the procedure. Since no ground truth for evaluating the results of stops and moves is available, exemplary outcomes are shown visually. According to RYKIEL JR. (1996), *visualization techniques* are used to compare the system to the actual model output (*model validation*). Figure 3.4 illustrates the resulting stop-move classification for a single participant.

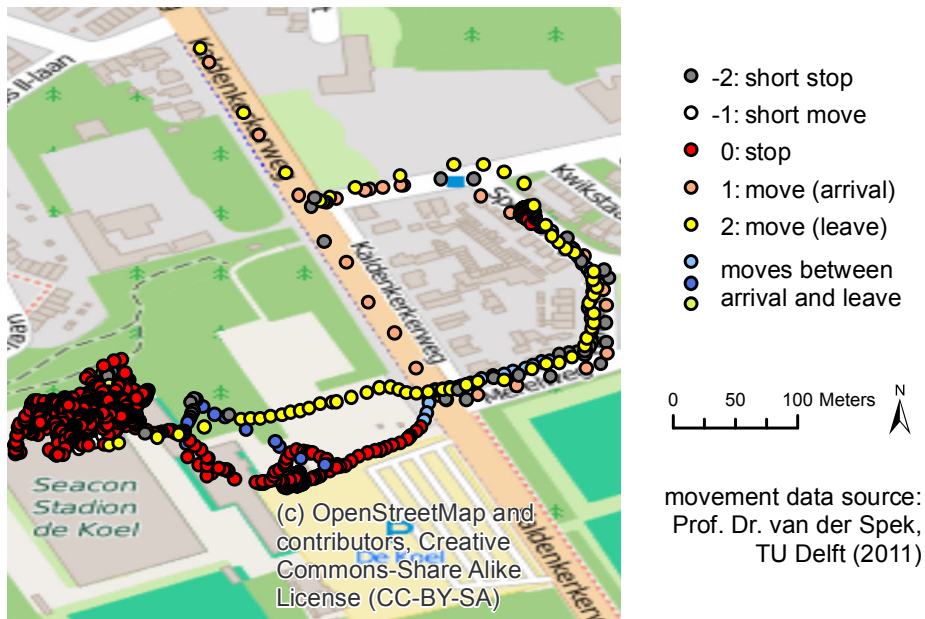


**Figure 3.4.:** Space-time cube and local scenario in “Neltje Jans” with stops and moves according to the preprocessing workflow. Grey: short stop/“-2”, white: short move/“-1”, red: stop/“0”, any other color: move/“>1”. Data source: TU Delft, S. van der Spek, 2011.

Along with the visualization of a participant’s movement data in the space-time cube, a specific situation, where a stop divides a move, is presented in Figure 3.4. This illustration facilitates the visual inspection of patterns on different spatio-temporal scales. On a larger scale, the stops in red are clearly visible. The moves detected during working days most likely correspond to commuting between home and the workplace. The moves on Saturday show different patterns compared to the others. On Sunday, again the person is supposed to move from home to work. On a smaller scale, the differing shape of the movement path on Saturday seems to reflect a trip to “Neltje Jans” – a small island with a kind of theme park. The light red GPS fixes

### 3.3. Stops and moves: application of the preprocessing workflow

illustrate the incoming movement track, whilst the yellow GPS points represent the way home. Again, the red points show the actual stop, where, in this case, the car was parked. The grey dots also represent stops in terms of mean speed and turning angle, but are assumed to be too short to divide the current move. Most often, the grey short stops are close to crossings or to locations, where the moving person turned its car. Altogether, the visual inspection in the case of this GPS track, and the stop-move classification with different types of stops and moves, allows a rough evaluation of the stop-move detection. In this example, we argue that it is reasonable to split a move at the location, where the car was parked, but not at places of short stops, such as in the case of crossings, for example.



**Figure 3.5.:** Another local scenario with stops and moves according to the preprocessing workflow. Grey: short stop/“-2”, white: short move/“-1”, red: stop/“0”, any other color: move/“>1”. Data source: TU Delft, S. van der Spek, 2011.

The incoming movement – in this case, the arrival at the soccer stadium – again is colored in light red, whilst the part of the track that corresponds to the movement leaving the current location after the soccer match, is shown in yellow. Also in this scenario, too short stops in grey mostly occur at crossings, and the first longer stop (red) can be observed some hundred meters away from the stadium, where most likely the car was parked. This example underpins that movements on different scales can be detected. For example, movements in walking distance are identified as separate moves with the presented algorithm, as the examples of short movements from the car to the cash point, or thereafter, from the cash point to the stadium, illustrate. However, as soon as the pedestrian starts moving very slowly in a crowded area in front of or within the stadium, or as soon as the person queued for a ticket – a phenomenon that would be visually identified as move – is classified as stop. The

yellow track shows that leaving the stadium area happens more fluently because no longer stops are detected just after the soccer match. Although the presented preprocessing workflow allows the detection of different types of stops and moves on various spatio-temporal scales, the application of the fix thresholds remains a limitation in terms of identifying slow movements as moves and not as stops.

## 3.4. Exploring effects of preprocessing

### 3.4.1. Quantifying effects of preprocessing

Preprocessing is considered necessary for assessing erroneous movement data. Often, preprocessing has a very specific purpose and therefore might have an impact on further calculations with the movement data, which are not explicitly obvious or expected. So, the purpose of the experiments in this section is to reveal effects of preprocessing on movement analysis and CAMA in particular. The quantification of preprocessing effects is realized by comparing distributions of trajectory-based movement parameters (e.g., speed), since the computation of movement parameters is a crucial task in many analyses of movement data. Accordingly, our assumption is that the effects of preprocessing are reflected in the distributions of movement parameters. Furthermore, the impact of preprocessing on movement parameters is investigated with regard to the various aspects of the surrounding environment in an effort to reveal potential links between preprocessing effects and the surrounding geographic context. Throughout the course of this section, effects of preprocessing on movement analysis are studied by applying established methods for map matching (BERNSTEIN & KORNHAUSER, 1998) and segmentation (detection of stops and moves, LAUBE & PURVES, 2011) as preprocessing steps.

### 3.4.2. Data and experiments

On the movement side, GPS localization was permitted, which allows for the quasi-continuous tracking of moving individuals in space-time (VAN DER SPEK *et al.*, 2009). GPS trajectories allow derivation of fine-grained descriptive movement parameters, such as speed, sinuosity or turning angle. The geographic context, enabling and constraining movement, is clearly application-dependent. For wild animals, relevant context might be habitat type or terrain, whereas for shoppers, on the other hand, it might include the spatio-temporal properties of the urban transit network and personal points of interest (e.g., home, work, gym). It should be noted that the goal here is not the identification of what context factors are important for a given movement process, but rather the quantification of the movement-context interrelation when making the assumption that we have access to expertise, capable of identifying relevant context (i.e. habitat type for a foraging animal).

Case studies were selected from urban mobility and behavioral ecology, featuring data with differing properties in terms of temporal resolution and movement space (Table 3.1). First, there was the analysis of the movement properties of the finely

### 3.4. Exploring effects of preprocessing

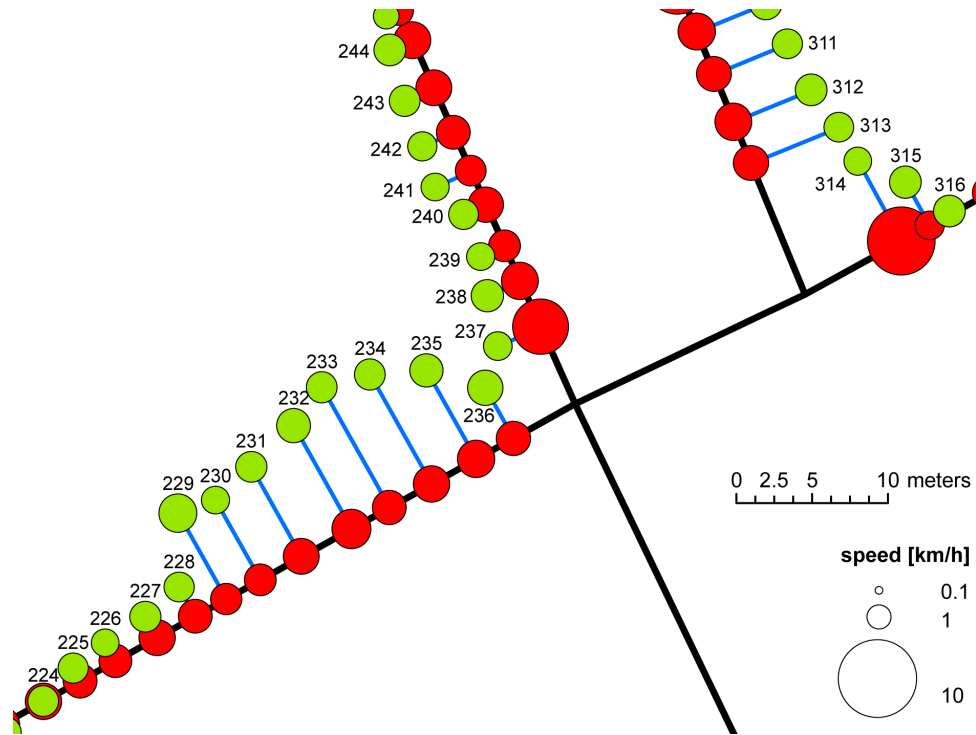
sampled trajectories of pedestrians moving in the urban network space of the city of Delft (Netherlands); people leaving a parking deck in the centre of Delft were given a GPS device and their trips through the city were recorded, comprising both raw GPS data, as well as preprocessed trip data, where stationary phases were manually removed by the data provider (VAN DER SPEK *et al.*, 2013). Second, the movement data of chamois foraging in the Swiss National Park (2002 – 2010) were used to perform an experiment relating speed to the underlying habitat type. This data set reflects typical data from monitoring studies in behavioral ecology, where technical constraints may dictate rather coarse sampling rates.

	<b>Pedestrians</b> <i>(Delft)</i>	<b>Chamois</b> <i>(Swiss National Park)</i>
Data collection	11/2009	12/2002 – 04/2010
Temporal resolution	2 <i>sec</i>	10 <i>min</i>
Space	Network, OpenStreetMap	Euclidean <i>(“unconstrained”)</i>
Moving objects	Pedestrians <i>(Homo sapiens s.)</i>	Chamois <i>(Rupicapra rupicapra)</i>
Context	Shopping and leisure POIs <i>(points)</i>	Habitat types <i>(polygons)</i>
No. of GPS points	2’300	29’571
Source	TU Delft, Stefan van der Spek	Swiss National Park

**Table 3.1.:** Overview of the pedestrians’ and chamois’ movement data sets (data sources: pedestrians, VAN DER SPEK *et al.*, 2013 / chamois, Swiss National Park, 2002 – 2010).

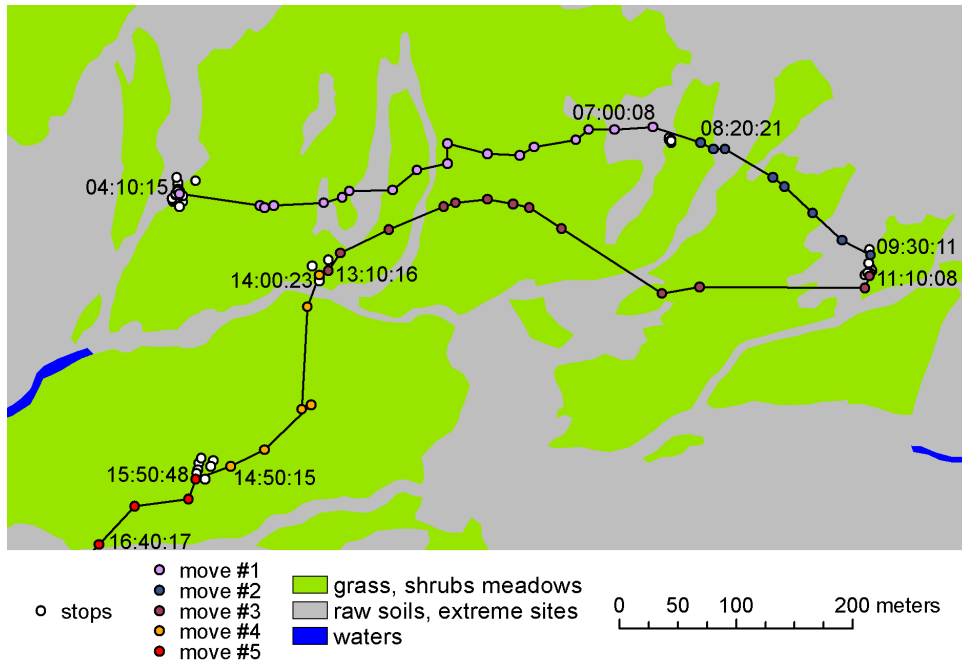
With regard to preprocessing, the problem of map matching and its effects on deriving movement parameters (in this case speed) from the trajectories were addressed with the network-bound data. For the other data set of chamois moving in an “unconstrained” Euclidean space, the effects of filtering stops on the relation of movement characteristics (again speed) to geographic context (habitat type) were studied.

Accordingly, the first case study examined the effects of preprocessing movement data in an urban context. More specifically, speed values provided by the GPS device were compared with different ways of computing speed from location fixes – both for raw GPS data and manually filtered trip data. First, speed was calculated from the distance moved within consecutive fixes (sampling rate of 2 *seconds*, few longer intervals). Second, speed was computed after the application of a naïve map matching (c.f. BERNSTEIN & KORNHAUSER, 1998; WHITE *et al.*, 2000) technique. With map matching, GPS fixes are matched on the underlying street network. For the naïve map matching, fixes were snapped to the closest network edge, with a maximal snapping threshold of 15 *meters* (Figure 3.6).



**Figure 3.6.:** Example trajectory section for a pedestrian in Delft, without (green) and with naïve map matching (red), fix indices at sampling rate of 2 *seconds* (Data source: VAN DER SPEK *et al.*, 2013).

The second case study aimed to relate speed to the underlying habitat type embedding the movement of seven GPS-tracked chamois in the Swiss National Park (Figure 3.7). In this case study, the focus was specifically put on the habitat type, as shown in Figure 3.7, because it is expected to have an influence on the chamois' movement. The chamois data set consists of MPOs at different temporal granularities with a sampling rate of 10 *minutes* every second Wednesday and 4 *hours* for the rest. In this case study, only the segments of the movement data at a temporal resolution of 10 *minutes* were considered. The data set at a temporal resolution of 10 *minutes* was selected in an effort to investigate whether or not movement data with such a coarse temporal granularity could be used to relate movement and context. Again, speed was calculated assuming constant speed between two consecutive fixes. Here, raw GPS data was segmented into stops (removed) and moves utilizing a simple algorithmic approach, which considers a fix-based quantification of mean distance (LAUBE & PURVES, 2011). Raw and filtered movement data were then related (“point in polygon”) to three habitat types aggregated from a detailed habitat data set ([www.habitatlp.de](http://www.habitatlp.de), 2002-2006). The habitatlp project dealt with the diversity of alpine habitats. The project aimed at monitoring habitats and their environmental changes during four years with the use of areal photographs.



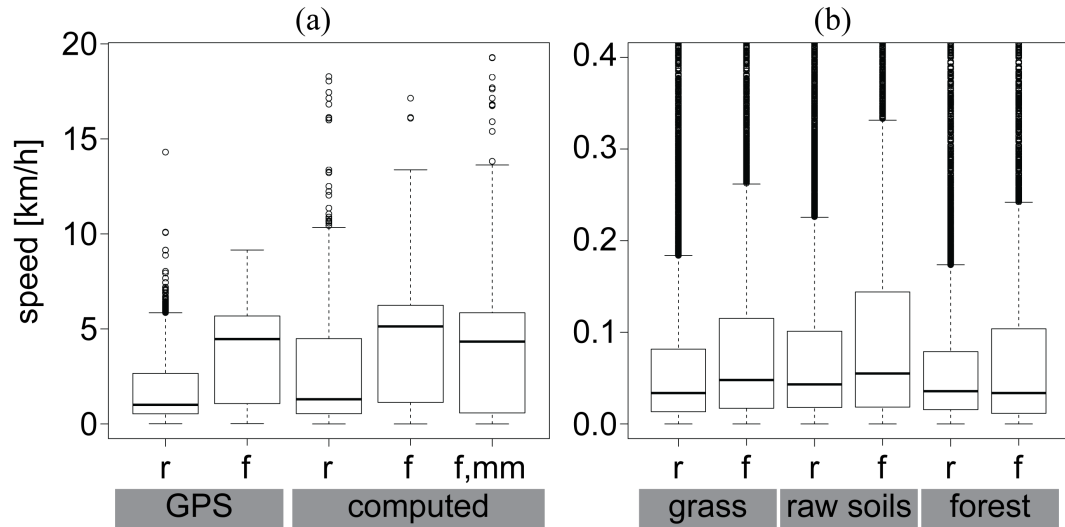
**Figure 3.7.:** Example trajectory of chamois with habitat context in Swiss National Park. Stationary fixes (white), moves in various colors (purple, dark blue, dark red, orange, red), and time of day (hh:mm:ss). Data sources: movement data, Swiss National Park (2002-2010); context data, [www.habitalp.de](http://www.habitalp.de) (2002-2006).

### 3.4.3. Results and discussion

For both case studies, speed values were binned, with each bin resulting in an item on the ordinate of the box whisker plots (Figure 3.8). The box whisker plots show medians (horizontal bar), 25<sup>th</sup> and 75<sup>th</sup> percentiles enclosing the middle 50% of the data (boxes, also interquartile range), minimum and maximum values (whiskers), and outliers (data points more than 1.5 times the interquartile range from either end of the box).

Figure 3.8a shows the results for the Delft pedestrians. The first two items describe speed values measured by the GPS device itself – first for raw (r) and second for filtered data (f). Computed speeds for raw (r), filtered (f), and both filtered and map-matched data (f,mm) were then followed. The “computed” speeds were calculated based on the trajectory, assuming a straight line connection between consecutive GPS fixes. Figure 3.8b illustrates variable speed values over three different habitat types (grass, raw soils, forest). Here, for every habitat type, raw GPS trajectories are compared to segmented and filtered data (stops removed).

Figure 3.8a first illustrates that separating moves from stops (raw: r vs. filtered: f or f,mm) has an important influence on computed speeds (median 1km/h vs. 5km/h). Second, the median of all three filtered speed categories (GPS, filtered, filtered and map-matched) is in the same order of magnitude. Third, the median



**Figure 3.8.:** (a) Case study #1, speed for pedestrians; GPS vs. computed; raw data (r), filtered (f) and map-matched (mm). (b) Case study #2, speed vs. habitat for chamois; raw data (r) vs. filtered (f). Note the smaller order of magnitude of the speed values in (b) than in (a) (but both with same unit:  $km/h$ ).

for “map-matched” is slightly below the uncorrected signal. As to this latter point it is argued that, for this result, two effects must be considered (Figure 3.6): (i) Map matching introduces error at intersections, where unrealistically large (overestimated) speed values result from the distorted geometry (fixes 236 to 237 or 313 to 314); and (ii) Changing strength of shadow effects in a 3D urban setting at building transitions results in positive speed artifacts owing to the positional inaccuracy of the GPS signal (e.g., fixes 228 to 229), where an error can even be removed through map matching. Further, we argue that, in this case, the latter effect (building transitions) outnumbers the first (intersections), explaining the lower median for filtered and map-matched (f,mm) than for only filtered (f) data. In this example, a purely distance-based map matching algorithm was applied. ZEILER (2012) showed that a more sophisticated map matching approach, including distance, neighboring GPS fixes and the topology of the road network, causes remarkable variation in trajectory-derived movement parameters, whereas these differences in the resulting movement parameters depend on the temporal scale as well.

Figure 3.8b shows no significant difference in speed depending on the embedding habitat types. For grass and raw soils, filtering out stops again results in higher speeds. The signal for forests is more complex, with a lower median but a larger range. One reason for this mixed signal could be that, in forests, animals generally move slower; however, averaged values of speed over time intervals of 10 *minutes* are, in general, very low. We argue that such low, average speed values do not represent actual instantaneous speed of moving animals. For instance, in Figure 3.7, the first segment between 04:10:15 and 07:00:18 shows several transitions between habitat



type, where the granularity of the trajectory does not often allow for a conclusive link between speed and habitat type.

### 3.5. Main findings and contributions

The workflow presented in Section 3.2 allows the detection of stops and moves using mean speed and turning angle, as well as time duration. With regard to the concept, BUCHIN *et al.* (2013) implement a similar threshold-based approach for segmenting trajectories into movement states (flight and stopovers of migrating geese) on the basis of a framework for segmentation with spatio-temporal criteria (BUCHIN *et al.*, 2011). Our preprocessing workflow has the advantage that the detection of stops and moves with one single set of thresholds still provides stops on different temporal scales, which can be assessed through the different indices. In the context of people moving in urban areas and commuting between larger cities, for example, stops at crossings/traffic lights are detected (index “-2”); at the same time, however, they are not handled as “real” stops on a larger scale (e.g., staying in a shopping center). However, with this threshold-based preprocessing workflow, movements at low speeds are prone to be classified as stops. ORELLANA *et al.* (2010) detect patterns of movement suspension based on a statistical approach, using the LISA, where neither spatial and temporal thresholds, nor knowledge about the moving object are required. Along this line of argumentation, our preprocessing workflow contributes to a better understanding of stops at different temporal scales for the same moving object. The presented approach is particularly suitable for movement studies in urban areas, where stops of different temporal duration are relevant for a better understanding of a moving object’s behavior (e.g., stop at a traffic light vs. visiting a shopping mall).

The experiments, linking movement parameters to the embedding geographic context in section 3.4 allow summarizing the main findings in the following list of lessons learned:

- Removing (filtering) stops is a paramount preprocessing step, as artifacts (pseudo-movement) introduced by inaccurate GPS fixes of stationary objects (spreading of GPS fixes at low speeds) swamp any signal with respect to speed values of moves.
- For network-bound movement, we argue that there is an unavoidable catch-22 between computing derived movement parameters from unmatched fixes (which may not lie on network edges and therefore are erroneous) or from map-matched fixes (which have an altered geometry and hence cannot represent the “true” movement in terms of speed).
- When the temporal granularity of movement data is so coarse that the interval between two consecutive fixes includes several stops and moves, computing instantaneous speed is not suitable; hence, establishing a link between such derived speed properties and movement context is not suitable either.

The results from the case study concerning effects of map matching with movement data in an urban setting suggest computing movement parameters (e.g., speed) based on the original geometry of the trajectory. Accordingly, we propose the use of the original geometry of the trajectory, as long as the geometry is relevant. Since the map matching algorithm implemented significantly changes the geometry, the computation of trajectory-derived movement parameters is not accurate enough with map-matched movement data. This outcome corresponds with the findings from ZEILER (2012). Although ZEILER (2012) applied a more sophisticated map matching algorithm than the naïve map matching used in this experiment, she also found remarkable variation of movement parameters as a result of their computation on the original and the map-matched GPS data. Moreover, map matching could also be used for other purposes besides correcting the actual GPS data. In the case of movement in network space, map matching could be applied as a method for relating movement to geographic context in the form of edge properties (e.g., one-way street). Furthermore, the relation of speed and land cover illustrate that the scale and modelling of movement characteristics play an important role within CAMA, and therewith supports the more detailed investigation of scale effects in Chapter 7 and the modeling of movement and geographic context (Chapter 4 and 5).

## Chapter 4.

# Concepts for relating movement to geographic context

### 4.1. A movement-context relation matrix

In this section, the *mc*-relation model is presented. This model comprises of a cross-tabulation matrix relating conceptual models for representing movement (*m*) in the horizontal axis with conceptual models for representing the embedding geographic context (*c*) in the vertical axis (Figure 4.1). This model is the first attempt to provide a structured theory for the interrelation of movement with its embedding geographic context. The movement-context relation matrix is not intended to be an exhaustive model capturing all kinds of movement-context relations. In this thesis, the *mc*-relation model is developed specifically for static geographic context. In future work, this framework could be extended for more dynamic context or other kinds of context (e.g., in form of other MPOs). The aim of the *mc*-relation model is exploring the characteristics of quantitative relations between the conceptual models for movement and its embedding geographic context for various interrelation forms.

#### 4.1.1. The movement axis

The movement axis aligns the basic conceptual models for representing MPOs. The movement axis first features the entity-based primitives vector ( $m_V$ ), move ( $m_M$ ), segment ( $m_S$ ) and polygon ( $m_P$ ), complemented with the field-based representation ( $m_F$ ). From left to right, the conceptual models gain in complexity, as their temporal extent increases from *instantaneous* (vector), over *interval* (again vector and move), *episodal* (segment) to *global* (polygon and field) (LAUBE *et al.*, 2007).

**Vector ( $m_V$ ).** Most research in movement analysis studies the movement of individual MPOs (e.g., DODGE *et al.*, 2008). The most atomic primitive then is a fix – a time-stamped point primitive ( $x,y,t$ ), typically collected through the use of GPS receivers. For our *mc*-relation model, however, we argue that the primitive element of movement is more than merely point-geometry: a fix has a velocity, a movement direction (azimuth) and a speed. This is important because a directed point – or a *vector* – offers different context interrelation methods than a mere point. It is important to note that, for our model, we adhere to the notion of movement

parameters that can be measured at infinitesimal short intervals – even though their computation may often require at least two consecutive fixes or even longer intervals.

**Move ( $m_M$ ).** Our second movement primitive *move* follows the definition of TURCHIN (1998). In his early conceptual framework for studying the movement of animals, TURCHIN defines a move as the displacement between two consecutive stopping points. In our *mc*-relation model, we adhere to this notion of a move as a straight line connector between two consecutive fixes. Whilst a move is clearly equally directed as a vector, it additionally has a spatial and temporal extent. Moreover, a move could also connect two vectors expressing different azimuths. Depending on the temporal granularity, moves can represent the displacement of a butterfly moving from one flower to another, as in TURCHIN’s original text, or residential displacement in geospatial lifelines (SINHA & MARK, 2005) – and even origin-destination pairs (THÉRIAULT *et al.*, 1999) or flows in flow maps (VERSICHELE *et al.*, 2012).

**Segment ( $m_S$ ).** A *segment* represents the movement of an individual over a longer temporal interval. This comprises a sequence of moves. As a geometric primitive, it is a polyline that can self-intersect. Most often segments emerge from a segmentation process, whereas trajectories are divided into meaningful subtrajectories, according to some specific quantitative rules (e.g., DODGE, 2011). A basic example is the segmentation of a trajectory into moves separated by stops (Subsection 2.3.1), whereas a stop could be defined using a speed threshold or even more complex procedures (e.g., ALVARES *et al.*, 2007; ORELLANA, 2012). Another example is the diurnal (days, days vs. nights) or seasonal (e.g., winter, summer) aggregation. A further, more complex example would be the modeling of movement behavior with different movement parameters (e.g., speed, turning angle, sinuosity) (e.g., DODGE, 2011). In short, segmentation aims at aggregating several consecutive vectors or moves into segments, which are homogeneous with respect to some spatio-temporal characteristics, or even in terms of some properties of contextual information.

**Polygon ( $m_P$ ).** When the temporal extent is widened even further, conceptual models for capturing the aggregated spatio-temporal extent of a movement process come into focus. For instance, in movement ecology, there is a large body of literature concerned with analyzing the movement of animals with respect to their space use (e.g., BASILLE *et al.*, 2008). For this purpose, home ranges are defined with polygons, aggregating GPS fixes to polygons, e.g., using MCPs POWELL (2000).

**Field ( $m_F$ ).** In a second areal conceptual model for movement, movement can be modeled with fields, representing point densities and probabilities of presence based on purely spatial or even spatio-temporal aspects of movement. WARTMANN *et al.* (2010) use KDE for defining animals’ home ranges. Pure KDE has the shortcoming that resulting fields only reflect the spatial distribution or footprint of fixes, but not the temporal characteristics of movement. More details relating to field-representations

of movement, including spatio-temporal characteristics of movement, are discussed in the following sections, where selected *mc*-relation methods are presented.

#### 4.1.2. The context axis

The description of the context axis can be kept shorter as it follows the basic conceptual models for representing spatial information – objects and fields. The context axis features the object-based primitives *points* ( $c_{Pt}$ ), *lines* ( $c_L$ ) and *polygons* ( $c_P$ ), as well as *fields* ( $c_F$ ), which typically are modeled as rasters or Triangulated Irregular Networks (TINs). The *mc*-relation model foresees relating any feature of the natural or built environment to movement data, with the exact nature of the context then being dependent on the application area.

Points could be POIs, checkpoints, work or home locations, or animals’ nest sites. One might want to relate animal movement to line-shaped features, such as rivers or coastlines, or human mobility to the course of motorways or inner urban edges of a street network. Polygon-shaped context may take the form of land cover classes or habitat types. Finally, field-based context typically refers to the continuous physical properties of the natural environment, such as temperature, magnetism or wind fields.

#### 4.1.3. Movement-context combinations

Cross-relating the movement axis with the context axis results in 20 different relation types for which characteristic profiles are described in the following and depicted in the signature in Figure 4.1. Each combination can be specified through consideration to its combination of the axes codes: for example  $m_{SCP_t}$  refers to methods relating segments to point-shaped context.

The *mc*-relation model does not aim at an exhaustive description of every possible technique per combination, but rather aims at the generic characterization of the methods required for a specific combination in the form of a profile. Characterization can be based on typical questions arising from a specific interrelation combination: for example, methods of the  $m_{SCP_t}$  type, in essence, require a notion of proximity, evaluating whether a moving object passes by or visits a point-shaped spatial feature. The following discussion of relation method profiles is structured according to the columns of the *mc*-relation model, starting with combinations for vectors, then moves, segments, polygons and, finally, fields.

**Vector.** Typical questions posed to movement vectors can be summarized as “Where am I? What is around me? Where am I heading to?” The first two questions require concepts of proximity and neighbourhood. Since a vector has a velocity, specific methods are required for the third question, assessing whether an MPO moves towards a point ( $m_{VCP_t}$ ) or a line ( $m_{VCL}$ ). When the context has an areal character ( $m_{VCP}$  and  $m_{VCF}$ ), spatial relation concepts allow for the semantic enrichment (Subsection 2.3.1) of the vector location with the underlying spatial variable (e.g., land cover type for  $m_{VCP}$  or elevation for  $m_{VCF}$ ).

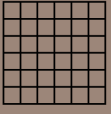
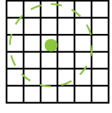
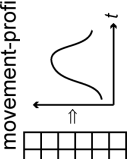
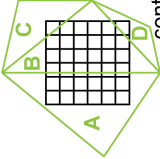
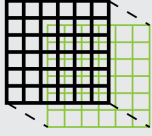
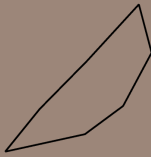
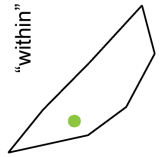

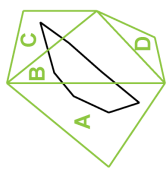


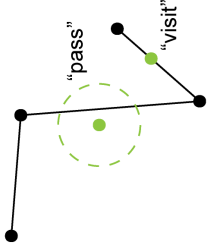
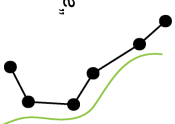
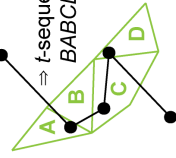
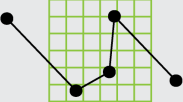


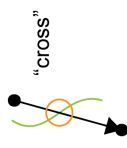
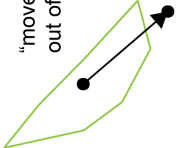
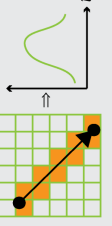


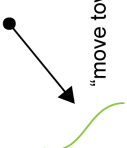
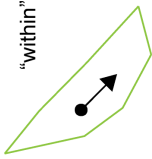
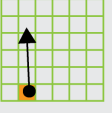


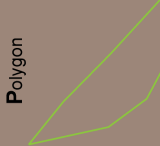
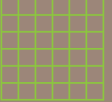
		$m_F$				
	Field					
			"around"			context weighted by movement: zonal statistics
	Polygon					
			"within"	"intersect"	A: 30% B: 50% C: 20% D: 0%	
	Segment					
			"pass" "visit"	"aligned"	$\Rightarrow$ t-sequence: BABCCD	
	Move					
			"nearest"	"cross"	"move out of"	context-profile
	Vector					
			"around"	"move towards"	"within"	semantic enrichment (map pin)
<b>MOVEMENT</b>	<b>CONTEXT</b>					
	Point	$c_{Pt}$		Line	$c_L$	
	Polygon	$c_P$		Field	$c_F$	

Figure 4.1.: mc-relation model ("relation matrix", author's own graph).

**Move.** Moves can be interpreted in two different ways, with suitable context relation methods varying accordingly. First, the edge between the fixes can be interpreted as a likely approximation of the actual path travelled in between the known fixes. This notion allows for questioning whether the move passed a point within distance (question of proximity: “near”) ( $m_{MCpt}$ ) or crossed a line ( $m_{MCL}$ ). For areal context ( $m_{MCP}$  and  $m_{MCF}$ ) relation methods aim at quantifying the distribution of the spatial variable covered by the move. When the connector does not represent a physical travel path but rather an indication of sequence and direction (as in origin-destination pairs), questions then can be asked about a move approaching an MPO to a point ( $m_{MCpt}$ ) or a move being aligned with a linear context feature ( $m_{MCL}$ ). Additionally, for areal context, topological relations such as stays within, moves into and moves out of an area, are all possible.

**Segment.** One key feature of segments in terms of movement is sequence. Relation methods therefore can establish whether context points are visited or passed in a sequence ( $m_{SCPt}$ ), lines intersected in a sequence ( $m_{SCL}$ ), and polygons again visited in a sequence ( $m_{SCP}$ ). Focusing on the line property of a sequence allows for concepts of alignment or orientation ( $m_{SCL}$ ). Similar to moves, the areal context covered by segments can also be quantitatively analyzed ( $m_{SCP}$  and  $m_{SCF}$ ).

**Polygon.** As outlined in Subsection 4.1.1, the polygon representation of movement adheres to an aggregation perspective. As a result, typical questions for this group of relation methods circle around topological concepts, such as inclusion, intersection and overlay. Context points can be included ( $m_{PCPt}$ ) and lines intersected ( $m_{PCL}$ ). The two areal context forms lend themselves to the quantitative analysis of the coverage distribution ( $m_{PCP}$  and  $m_{PCF}$ ).

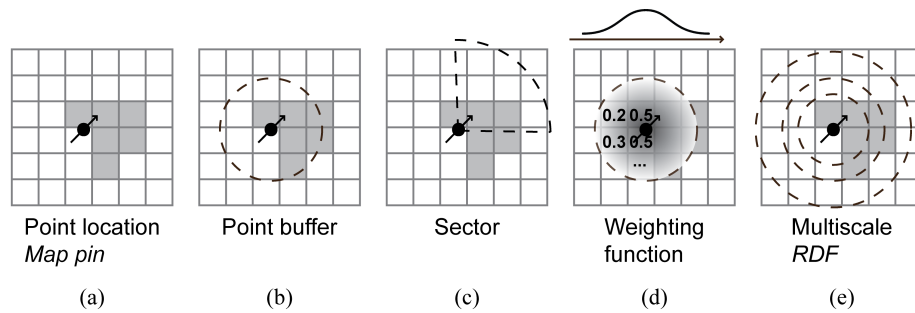
**Field.** Relation methods for field-based representations of movement are, in many aspects, similar to the ones for polygons, with both adhering to an aggregative perspective. In contrast to the previously discussed methods for polygons, inclusion, intersection and coverage will be used for an interrelation with a probability of occurrence as opposed to a crisp presence-absence.

A systematic and exhaustive development of implementations for the entire *mc*-relation matrix goes beyond the scope of this thesis; however, we argue that the usefulness and expressiveness of the matrix becomes evident when systematically exploring implementation techniques for only one row of the *mc*-relation matrix. The next sections present the exemplary development for field-based context  $c_F$  and its relation to different models of movement (Figure 4.1, light gray row). The field-based context has been chosen for this illustration because the subsequent experiments (Chapter 5 aim at relating the movement of alpine animals to the diversity of the embedding alpine topography, as an example of quintessential geographic context enabling and constraining animal movement. In the following sections, first, a family of vector-based relation techniques is developed, and second, a novel relation technique

based on Brownian bridges (HORNE *et al.*, 2007) is proposed for relating field-based movement with field-based context.

## 4.2. Vector-field movement-context relation methods

The development of vector-based relation techniques is initiated with the basic movement feature *vector*; in essence, a point object with velocity. Clearly, a very basic way of acquiring a context value for a vector is simply by querying the context attribute at the vector's point location (*map pin*, Figure 4.2a, Table 4.1). This approach is appropriate for context variables, where we assume a very local influence of the context on the movement, e.g., with wind speed for a cyclist. However, in other cases, the surroundings of a moving individual will be just as relevant for the actual movement as the context value at the very vector location. For example, topographic features as hills or land cover types as forest have an influence on the movement path across longer distances. For this reason, a series of increasingly complex approaches is suggested in mind of capturing the surrounding environment, exploiting the vector's additional direction and speed information.



**Figure 4.2.:** Concepts for relating the vector-based movement model to the embedding (field-based) geographic context ( $m_{VCF}$ , author's own graph based on established GIS concepts).

Relation technique	Fig. 4.2	Variables	Characteristic	Weighting
Map pin	a	–	Punctual	–
Point buffer	b	$r$	Areal	Homogeneous
Sector	c	$r, \alpha$	Areal	Homogeneous
Weighting function	d	$r, w(d)$	Areal	IDW
RDF	e	$r_{s1}, r_{s2}, r_{s3}, \dots$	Areal	Homogeneous

**Table 4.1.:** Overview of  $m_{VCF}$ -relation methods for vector-based movement model (linked to Figure 4.2).



The techniques described in Figure 4.2 and Table 4.1 allow varying the shape and also the weighting of the context information with the distance to the vector location. The simplest inclusion of the surrounding environment comes in the form of circular point buffer around the vector location (*point buffer*, Figure 4.2b, Table 4.1). Since most of the individuals have several abilities to perceive the surroundings, different buffer shapes allow further specifying the inclusion of the neighboring context: for example, visual perception is limited to a certain field of view (e.g., around  $180^\circ$  for human beings). Subsequently, a point buffer restricted to a certain area could be the more appropriate way of capturing relevant surroundings to the movement. This area, where context information is considered relevant, can be defined with different approaches, where the point buffer is limited to a specific angle of view as a *sector* (Figure 4.2c, Table 4.1) or in a more sophisticated way, only visible parts of the surrounding environment are taken into account (*viewshed*).

Once an area is specified for the inclusion of the geographic context, different approaches can be used for its quantitative spatial analysis. In the simplest case, the vector can be assigned that context type, expressing the largest percentage within the area (*dominance* principle). Alternatively, the relative fractions of all context types are captured for each vector of the trajectory, and then can be summarized along the trajectory (*proportion* principle).

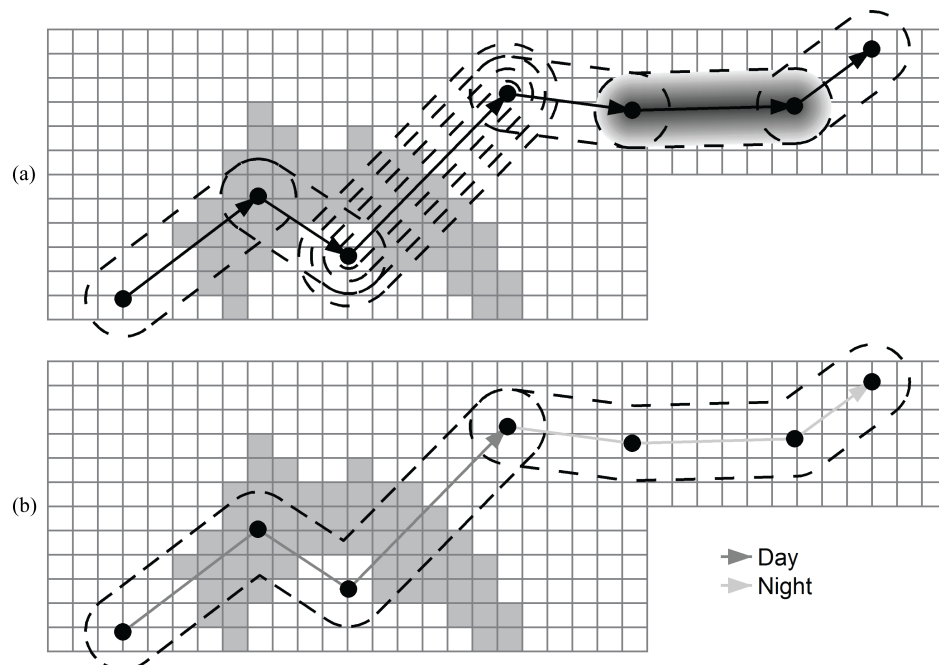
Since with spatial distance the importance of surrounding context can vary, we include *distance-weighting functions* (e.g., Inverse Distance Weighting (IDW) using a linear function, or a normal distribution around sample points, etc., Figure 4.2d, Table 4.1). Finally, variable scales of context can be evaluated, using differently shaped buffer rings for the respective scales, as introduced in Imfeld's RDFs (IMFELD, 2000; Subsections 2.3.2 & 2.3.3) (Figure 4.2e, Table 4.1). Table 4.1 gives an overview of characteristics and crucial parameters shaping the form of the different vector-based relation methods. In the case study in Chapter 5, the map pin, the point buffer and the sector approach (Figure 4.2a-c) are applied.

### 4.3. Move-field movement-context relation methods

As was shown in Subsection 4.1.3, a move can be defined in different ways, such as in terms of start- and end-point, for example, where the movement is directed from the start- to the end-point, or as a line connecting the start- and end-point. The introduction of the concept of moves, as defined by TURCHIN (1998), in comparison to vectors is reasonable owing to the fact that, besides the direction, moves include a step length from the start- to the end-point, and a line approximating the movement from the start- to the end-point allows considering context information in relation to movement in between the GPS fixes.

For movement interpreted as moves, some of the basic concepts remain similar to the case of the movement modeled as vector. For instance, a line profile along a move can be thought of as the corresponding relation method to the map pin approach, and instead of using point buffers, line buffers can be applied, again changing in shape

and weighting depending on the distance to the line. Further, methods comparable to the vector-based RDF approach can be used to assess the geographic context around and, more specifically, between GPS fixes (Figure 4.3a). In the experimental part (Chapter 5), where some of these techniques are tested, the line buffer approach is implemented.



**Figure 4.3.:** Relating (a) move- and (b) segment-based models of movement to geographic context (author’s own graph based on established GIS concepts).

#### 4.4. Segment-field movement-context relation methods

A segment can be a whole trajectory or a subset of a trajectory, where segments are distinguished with regard to spatial, temporal or spatio-temporal characteristics of the movement. Figure 4.3b schematically illustrates two segments separated according to daytime. By definition, segments most often comprise a sequence of moves and, therefore, conceptually, are, to some extent, analogue to the movement modeled as moves. Accordingly, again, relation methods discussed in connection with move-based approaches (Figure 4.3a) can be transferred to movement as segments.

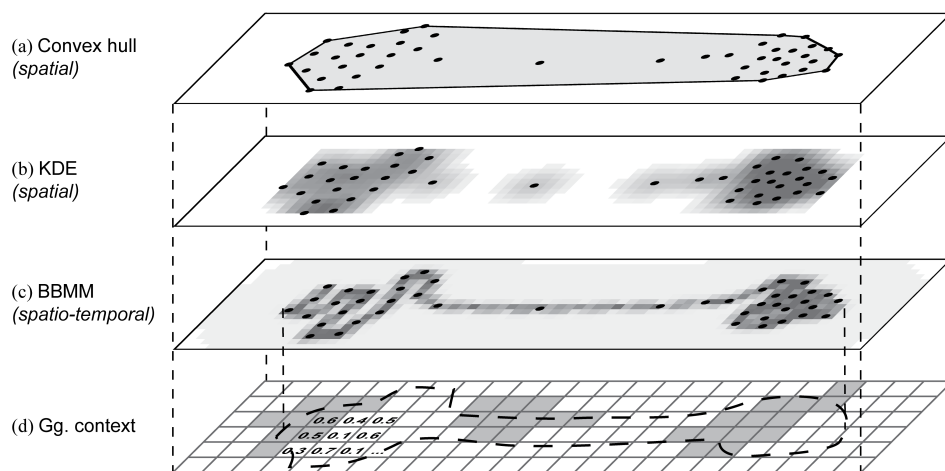
Furthermore, segments allow the quantification of movement in relation to the surrounding geographic context with regard to changes in the environment, such as by counting the number of context transitions per segment, for example. Different methods of counting the context transitions are possible: for instance, considering context in the segment’s GPS points or changes in context along the segment line resulting in a temporal sequence of context values can be used in order to establish the

number of context transitions. In Figure 4.3b, the two approaches (in points vs. along line) lead to different results in terms of context transitions. The segment representing the “day” part of the trajectory results in the temporal sequences of context values “white-grey-white-white” with two transitions or “white-grey-white-grey-white” with four transitions considering context in the GPS points or along the segment line, respectively. In the empirical investigation (Chapter 5), a line buffer is used in order to relate the segments to the underlying geographic context, as movement modeled as segments and its relation with line buffers to the embedding geographic context remains comparable to other approaches used to relate other models of movement and geographic context.

## 4.5. Polygon-field and field-field movement-context relation methods

In increasing complexity, Figure 4.4a-c schematically displays some of the areal models of movement, as discussed in the literature review (Subsection 2.3.2). Figure 4.4a illustrates the MCP as a simple polygon-based movement model, delineating individuals’ used space, whereas context values within the polygon are intended to be relevant. In the relation process of the MCP to the underlying geographic context, context values are weighted homogeneously across the polygon area – irrespective of the spatial and temporal characteristics of movement (e.g., duration of stay, revisit of locations). Further, the delineation of the used space is only based on the geometry of the trajectory – again, not on spatio-temporal characteristics of movement. The MCP approach and its relation to the underlying geographic context is applied in the case study in Chapter 5.

Figure 4.4b illustrates a probability density surface resulting from the KDE, taking into account only the spatial arrangement of the MPO, but regardless of the temporal order of MPOs, for instance. Figure 4.4c nicely illustrates that BBMM manages to maintain a certain notion of a traveled path in between known fixes, whereas KDE produces mere sequence-insensitive hot spots. Consequently, BBMM (HORNE *et al.*, 2007) best suits our needs as a field-based representation of movement for the corresponding case study in Chapter 5. It is this BBMM-based probability density value depending on the spatio-temporal characteristics of movement (Figure 4.4c) that we exploit for our  $m_{FCF}$ -relation, utilizing the probability density surface for a weighted relation with the context field (Figure 4.4d). We argue that the higher the probability density value for an individual in a certain area, the more relevant the underlying context can be assumed to be. The effective relation is then realized using map algebra (TOMLIN, 1983) for field-based context as a model of continuous phenomena (e.g., temperature, terrain) or zonal statistics for geographic context modeled with polygons representing categories of different land cover types for example, where for each category or zone (e.g., land cover type), statistics of movement characteristics are computed.



**Figure 4.4.:** Concepts for movement models based on representations as polygons or fields and the relation to geographic context: (a) Convex hull, (b) KDE, (c) BBMM and its relation to (d) geographic context (author's own graph based on established GIS concepts).

# Chapter 5.

## Relating ungulates' movement to geographic context

### 5.1. Study area and data

The family of *mc*-relation methods introduced in Chapter 4 is now illustrated in a case study relating animal movement to the embedding terrain. GPS-tracked movement data of chamois, deer and ibex from the Swiss National Park are related to the underlying and surrounding landforms, whereas the landforms are assumed to have an influence on animals' movement. The Swiss National Park is situated in Southeast Switzerland in an area of around  $170 \text{ km}^2$ . Therewith, it is the largest protected area in the country. The ungulates were tracked with GPS loggers at a temporal sampling rate of 30 *minutes* (almost all deer and some ibex), or between 10 *minutes* and 4 *hours* (10 *minutes* every second Wednesday for most of the chamois and some ibex), depending on the species. The following Table 5.1 provides more detailed insights into the ungulates' movement data.

	<b>Chamois</b>	<b>Deer</b>	<b>ibex</b>
Data collection	12/2002 – 04/2010	05/1999 – 05/2010	06/2003 – 06/2009
Mean time span per animal	1.4 <i>years</i>	1.5 <i>years</i>	1.2 <i>years</i>
Temporal resolution	10 <i>min</i> – 4 <i>h</i>	30 <i>min</i>	30 <i>min</i> 10 <i>min</i> – 4 <i>h</i>
No. of animals	7 (6 f, 1 m)	10 (10 f)	8 (4 f, 4 m)
No. of GPS points	29'571	19'084	25'509
Source	Swiss National Park	Swiss National Park	Swiss National Park

**Table 5.1.:** Overview of the ungulates' movement data sets (source: Swiss National Park, 1999 – 2010).

The Swiss National Park has a pronounced topography, with elevations ranging from 1'400 *meters* up to more than 3'000 *meters* above sea level. For our case study, context is used in the form of a landform classification. To this end, we use a field-based DEM (“dtm4”, Swiss National Park, 2010) with a spatial resolution of 4 *meters*. The “dtm4”

originates from different sources with various cell sizes (1, 4, 10 and 20 *meter(s)*). Landforms are characterized by topographic parameters, such as slope, convexity and texture according to the unsupervised landform classification by IWAHASHI & PIKE (2007) (Figure 5.1). The background of Figure 5.1 illustrates an example section of the terrain with the superimposed landform classification. Figure 5.2 provides an overview of the space use of the studied chamois, deer and ibex in the Swiss National Park in the form of the dissolved MCPs. This figure shows that the species express rather dissimilar space use patterns, and overall stay away from the populated area in the valley floors.

## 5.2. Experiments

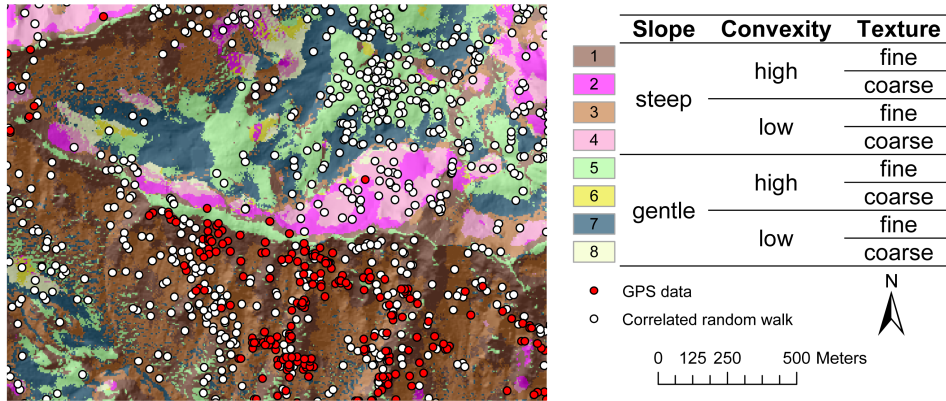
The experimental set-up is threefold. First, we assess whether we can expect landforms to influence ungulate movement in the area. To this end, we compare the landform use of real tracking data with Correlated Random Walk (CRW) data. Second, we then test the outlined family of  $m_{FCF}$ -relation techniques (Section 4.5), investigating whether or not the use of different relation methods results in significant differences in landform use, now only for the real data. Third, we choose the vector-based sector relation method (Figure 4.2c) in order to analyze the sensitivity of the relation method when varying the sector parameters radius and angle of view.

### 5.2.1. Validity experiment: Is the movement at all related to the terrain?

The rationale of the validity experiment is as follows. We assume that the landforms have an influence on the ungulate movement when the real tracking data shows a significantly different landform use than random walk data expressing similar movement properties as the observed data (Figure 5.1). We ensure comparability of observed and simulated data when using a CRW approach and by constraining random walks to the MCPs of the spatial extent of the corresponding species' GPS fixes distribution (Figure 5.2). The CRWs generate synthetic trajectories based on relative distributions of step length and turning angle, as observed in the respective species' real data. We generated 40, 60 and 50 trajectories for chamois, deer and ibex, respectively, with 1'000 points for each trajectory based on the relative distribution of step length and turning angle from the real movement data. For this baseline experiment, the basic map pin approach was implemented in order to relate the movement to its underlying topography (Figure 4.2a). The CRW generator was applied, using the Geospatial Modelling Environment (BEYER, 2012).

### 5.2.2. Sensitivity experiment #1: Differences between relation methods

The relative landform use for each species was quantified by applying the following family of  $c_F$ -relation techniques: map pin, point buffer, sector, line buffer, MCP, and BBMM. As parameters, we used a point buffer radius of  $r = 250$  *meters* and a front angle for the sector of  $\alpha = 60^\circ$ . The radius  $r$  for the sector was fixed to a value of

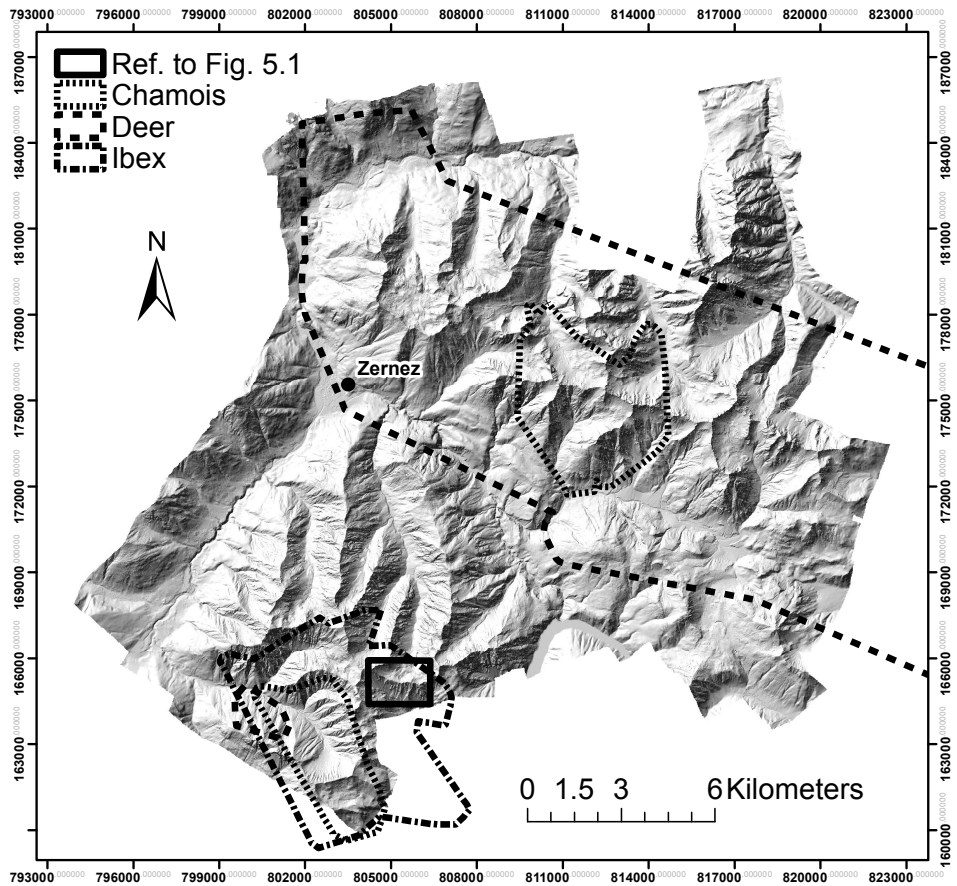


**Figure 5.1.:** CRW vs. GPS data of a female ibex (no. 469) in relation to geomorphometric landforms by IWAHASHI & PIKE (2007) with a hillshading in the background in Swiss National Park (data sources: GPS data, Swiss National Park, 2006 – 2007; DEM “dtm4”, Swiss National Park, 2010). Note: only spatial distribution of movement data is relevant here; temporal sequence in form of trajectory is not illustrated.

600 *meters* since the sector then covers approximately the same area as the buffer with a radius of 250 *meters*, assuming the animals perceive an area of a comparable size. Subsequently, line buffers were used, also with a radius of 250 *meters*. The radii for the different buffers were chosen with relatively large values (250 *meters* and 600 *meters*) since salient topographic features are supposed to have an impact on animals’ movement across relatively large distances. For point buffer, sector and line buffer, applied the *dominance* principle was applied (Section 4.2). This experiment was mostly realized using a JAVA implementation (version 1.6.0).

Completing the series of  $c_F$ -relation techniques, we also included the basic  $m_P$ -movement model MCP in our experimental set-up. Each individual’s movement is aggregated within its MCP and accordingly dissolved to a single polygon per species. Within these resulting dissolved polygons, the landform class distribution per species is analysed (Figure 5.2).

Finally, the field-based representation of movement with BBMM is calculated using the R package BBMM 3.0 (NIELSON *et al.*, 2011), which is based on the stochastic model presented by HORNE *et al.* (2007). The BBMM probabilities of occurrence then serve as weighting values for assessing the relative distribution of the landform classes. The BBMMs are computed with a location error of 10 *meters*, a resulting cell size of 20 *meters* and only GPS fixes with a time gap below 20’000 *seconds* (around 5.5 *hours*) are taken into account for the computation. In the current BBMM calculations, the probability density function is integrated with a time step of 10 *seconds*. The BBMMs are calculated separately for each individual, and then used for the relation to the landform classes, where each raster in the BBMM is related to the underlying nearest landform class raster. Similarly to the MCP methodology, the resulting context distributions (per individual) are aggregated per species. Figure



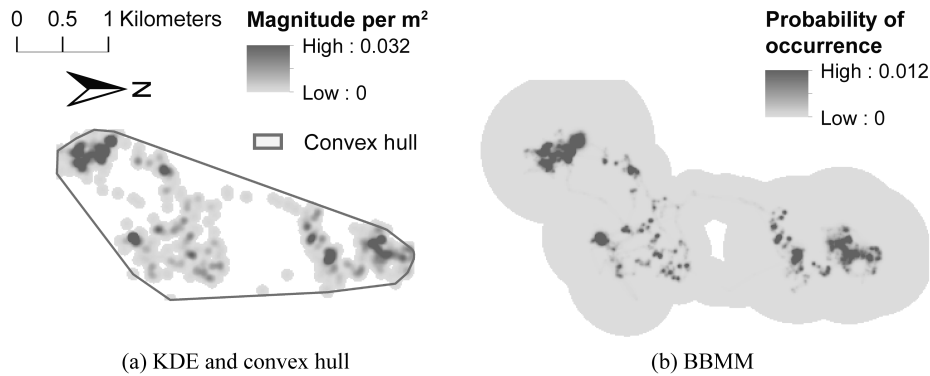
**Figure 5.2.:** Dissolved MCPs per species in Swiss National Park with a hillshading (data source: DEM “dtm4”, Swiss National Park, 2010). Note the cut-out of Figure 5.1.

5.3b illustrates the BBMM for a single chamois in comparison to the KDE and MCP approach (Figure 5.3a).

### 5.2.3. Sensitivity experiment #2: Differences due to varying sector parameters

In a second sensitivity experiment, the differences emerging when systematically varying the parameterization of one single relation method were analyzed. To this end, we chose the  $m_{VC_F}$ -relation method *sector* introduced in Figure 4.2c and Table 4.1, varying the sector radius and opening angle. Sectors were calculated with radii of 120 *meters*, 250 *meters* and 600 *meters* for a fixed angle of 60°. Subsequently, the sector radius is fixed to 120 *meters* and the angle is varied (30°, 60° and 180°). The emerging variance is accordingly compared to the variation between relation methods (previous Section 5.2.2), since effects of method parameterization are prominent in





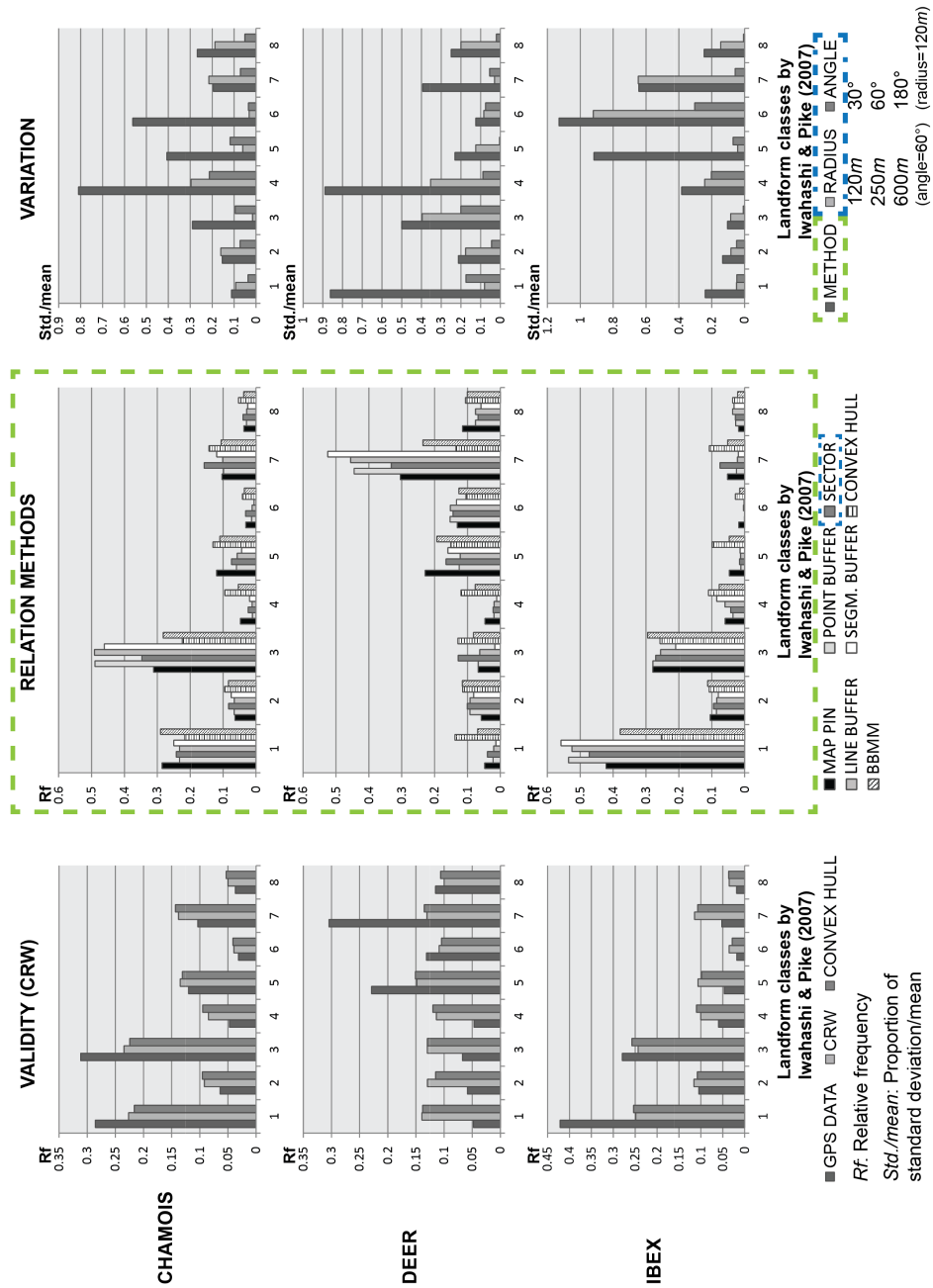
**Figure 5.3.:** Representations of the movement path of a female chamois (no. 47) in Swiss National Park. (a) Spatial models with KDE and MCP, and (b) spatio-temporal approach with BBMM (data source: GPS data, Swiss National Park, 2002 – 2004).

multi-scale analyses, such as by varying buffer radii, for example. This experiment aims at quantifying whether the effect of different relation methods is at a similar order of magnitude as the variation due to the better investigated variation of parameters.

## 5.3. Results

### 5.3.1. Validity experiment: Is the movement at all related to the terrain?

The geographic context covered by the actual movement is compared to CRW, restricted to the dissolved MCPs per species. Figure 5.1 illustrates in an exemplary way of how terrain constrains animals' movement. In this particular case, the mountain ridge builds a natural border for the female ibex (no. 469) since that individual rarely moves north of the ridge. The overall picture is presented in Figure 5.4 in the first column (VALIDITY). This shows a comparison of the distribution of landform classes for both observed animal movement and its corresponding CRW. This figure shows that, for the landforms by IWAHASHI & PIKE (2007), animals' actual movement produces a significantly different landform use pattern than the CRW (Wilcoxon rank sum test:  $p < 0.05$ ). This result indicates that ungulates' movement is, to some extent, influenced by the landform classes.



**Figure 5.4.:** Histograms representing results of experiments with regard to VALIDITY, RELATION METHODS and VARIATION. VALIDITY: relative distributions of CRWs vs. GPS data. RELATION METHODS: relative distributions for different methods relating movement and landform classification. VARIATION: coefficient of variation (standard deviation/mean) due to different relation methods (green dashed box) and sector parameters (radius and angle, blue dashed box).

The real movement of chamois and ibex captured with GPS loggers has the largest difference to the CRW in Landform Class 1, which represents landforms with steep slope, high convexity and fine texture. More specifically, around 60-70% of the chamois and ibex observations are found in areas with steep slope and fine texture, which corresponds to Landform Classes 1 and 3. For the deer, Landform Classes 5 and 7 (gentle slope and fine texture) are overrepresented in comparison to the CRW. Finally, the landform use of the CRW trajectories were also compared with the landform distribution in the MCPs (second and third bars in bar chart). We observe that both landform class distributions are similar. This confirms that the CRW landform class distribution not only depends on the parameters step length and turning angle, but is also heavily influenced by the landforms available within the MCP.

### 5.3.2. Sensitivity experiment #1: Differences between relation methods

The second column in Figure 5.4 (RELATION METHODS) shows the histograms for different relation methods across the landform classes for chamois, deer and ibex. Table 5.2 illustrates absolute variations across different relation methods using standard deviation. It is important to note that the standard deviation represents an averaged value for the variation; thus, single pairs of relation methods even show absolute deviation of up to 30% (e.g., deer, Landform Class 7, line buffer vs. MCP). The chamois show large absolute variation across different relation methods for Landform Class 3 (steep slope, low convexity, fine texture), for the deer Landform Class 7 (gentle slope, low convexity, fine texture) comprises most of absolute variation over the different relation methods and for the ibex the highest absolute variability between relation methods occur for Landform Class 1 (steep slope, high convexity, fine texture). In general, the highest absolute variation occurs in the landform class, where the ungulates were recorded the most.

Furthermore, the variability of the relative frequencies per landform class is assessed with a normalized measure for variation, the coefficient of variation (or synonymously relative standard deviation), since this measure allows different results of variation to be compared. This statistical value (in percentage) is defined as the quotient of standard deviation and mean multiplied by 100. A numeric example for the chamois and Landform Class 3 illustrates the computation of the coefficient of variation. With the mean (0.36) and standard deviation (0.11) of the relative frequencies for the chamois and Landform Class 3 (map pin: 0.31, point buffer: 0.49, sector: 0.35, line buffer: 0.49, convex hull: 0.22, BBMM: 0.28) the coefficient of variation ( $0.11/0.36 * 100 = 31\%$ , Table 5.2) is computed, accordingly. These relative variation values show notable variability across all landform classes for all ungulates (Table 5.2 and third column of Figure 5.4, VARIATION; note the green dashed box, indicating the methods variation).

Comparing individual methods, we observe that point and line buffers manifest similar landform class distributions since the large radii, when compared to the step length of the actual movement (step length  $< 100$  meters for around 80% of the moves), lead to a large overlap of the buffers; thus, point and line buffers cover

Species	1	2	3	4	5	6	7	8
Chamois	0.25±0.03 (12%)	0.08±0.01 (17%)	0.36±0.11 (31%)	0.04±0.03 (79%)	0.09±0.03 (35%)	0.03±0.01 (44%)	0.12±0.03 (21%)	0.04±0.01 (24%)
Deer	0.06±0.04 (78%)	0.10±0.02 (22%)	0.09±0.03 (35%)	0.05±0.04 (80%)	0.16±0.04 (25%)	0.14±0.02 (14%)	0.32±0.12 (39%)	0.09±0.02 (22%)
Ibex	0.43±0.11 (25%)	0.10±0.01 (12%)	0.27±0.02 (6%)	0.06±0.03 (42%)	0.04±0.03 (87%)	0.01±0.01 (103%)	0.06±0.03 (58%)	0.03±0.01 (27%)

**Table 5.2.:** Mean and standard deviation and coefficient of variation (in percentage) of relative frequencies per landform class and species (mean ± std and std/mean).

approximately the same area. Moreover, the point and line buffer approaches show the largest absolute differences to the other relation methods. The related other vector-based approaches (map pin and sector methods) are more similar to the BBMM and MCP approaches than to the point and line buffer approaches. Therefore, the sector approach is recognized as comparable to the map pin approach but quite different from the distributions emerging from the point and line buffer relation methods – even though, for the sector approach, a relatively large radius was selected. Although map pin and sector approaches in comparison to the Brownian bridge-based relation method are technically substantially different, the resulting distributions are surprisingly similar – especially when comparing with the MCP approach, which, in general, differs much more from BBMM – particularly in the case of the deer.

### 5.3.3. Sensitivity experiment #2: Differences due to varying sector parameters

Relative variations in the landform class distributions due to different relation methods (green dashed box) and sector parameters (radius and angle, within-method variation, blue dashed box) are illustrated in Figure 5.4 in the third column (VARIATION). The result shows that the largest relative variation occurs with the different relation methods in most of the cases. The variation of the radius (120 *meters*, 250 *meters* and 600 *meters*) of a 60°-sector causes quite a significant variation in the landform class distribution, whereas the different sector angles (30°, 60° and 180° at a radius of 120 *meters*) have the least effect on the result.

## 5.4. Main findings and contribution

The significant differences for the spatio-temporal landform use between the CRW and the observed animals indeed indicate a link between animals' movement and landform classes. The distribution established for all three species matched our expectations, given the habitat of the respective species. Chamois and ibex, on the other hand, use steep areas with fine textures, whereas deer prefer gentler slopes in fine-textured terrain. More specifically, ecologists confirm that agile and deft chamois (TOÏGO, 1999) and ibex (NESTI *et al.*, 2010) tend to prefer very steep slopes. In contrast, we are not surprised to find heavier deer in gentler terrain. All in all, the results of the

validity experiment indicate a strong relation between mountain ungulate movement and geomorphometric landform classes in the first place.

From the sensitivity experiment #1 (difference between vector-based relation techniques), we are able to single out three key findings. First, point and line buffer relations produce similar results when the sampling rate of the GPS device is high or animals do not move over large distances, consequently causing the buffers to overlap. Map pin and sector relations produce similar distributions that are, in turn, rather different from point and line buffer relations. The difference between the sector and buffer relation clearly indicates the potential importance of directionality when relating movement and context. Third, as expected, the MCP shows substantial deviations from all other relation methods (e.g., Landform Class 7 for the deer). Deer cover much larger distances than the smaller chamois or ibex, resulting in a large MCP that, in most cases, is only a very poor representation of the species space use (Figure 5.2). In contrast, BBMM aiming at a more realistic areal representation of the movement diverges from the MCP results and is closer to vector-, move-, and segment-based relation methods. Sensitivity experiment #1 underlines the complex interplay of *mc*-relation methods, their parameterization and the data characteristics.

Sensitivity experiment #2 (parameterization of *mc*-relation methods) revealed – at least in some cases – a substantial influence of parameterization. In many cases, the differences due to the parameterization are much smaller than from using different methods (e.g., ibex, Landform Class 5), whereas in other cases, the variation due to parameterization is clearly in the same order of magnitude (e.g., ibex, Landform Class 6 and 7, method vs. radius).

Whilst the overall landform class distribution can be similar for *mc*-relations, we expect variable effective landform class allocation on the level of individual vectors or moves depending on the used *mc*-relation method. However, the focus of our experiments was not directed towards investigating the actual and spatially explicit allocation of landform class along our trajectories, but rather on garnering insight into quantifiable differences resulting from the different conceptualizations of the movement-context relation.

Most other work on CAMA is application-driven, typically proposing one specific solution for one specific problem. In contrast, we opted to first detail a theoretical foundation with our *mc*-relation model (Chapter 4), which then underwent partial implementation and experimental tests in order to assess its contribution in this chapter. The *mc*-relation model clearly has an illustrative and structural purpose, aiming at pointing at the methodological diversity of movement-context relations. We argue that our systematic development of relation methods for field-based context produced a remarkable methodological diversity indeed leading to quantifiable differences in the movement-context relation.



## Chapter 6.

# Relating people's movement to geographic context

This chapter presents two experiments relating people's movement to different types of geographic context, with focus directed towards the modeling of geographic context rather than on different relation methods presented in the previous chapter. Therefore, this chapter does not directly relate to the *mc*-relation matrix introduced in Chapter 4. In the first experiment, the speed of bikers and hikers moving in a mountainous area is related to the terrain slope through the use of different quantitative models for computing slope as well as speed. The second experiment is realized with the adoption of Foursquare check-in data in order to illustrate, with a simple example, how UGC is able to provide the geographic context, thus supporting a better understanding of pedestrians' movement in an urban context.

### 6.1. Study area and data

For the first experiment, movement data is provided from the research project "Management-Toolkit Freizeit & Natur – mafreina", which is led by the Institute for Natural Resource Sciences at Zurich University of Applied Sciences. Visitors of the "Biosfera Val Müstair" were asked to log their trips with a GPS logger. The data set used in this study comprises 20 biking and 30 hiking GPS tracks, encompassing a temporal sampling rate of usually 5 *seconds* in the area of the Fuorn pass ("Biosfera Val Müstair") in Switzerland. The GPS tracking was conducted in 2009 and 2010, resulting in around 65'000 and 100'000 GPS points for the bikers and hikers, respectively. For calculating slope, a DEM from the Swisstopo – more precisely the swissALTI3D (version 2013 with a pixel resolution of 2 *meters*) from the "Bundesamt für Landestopographie", is used.

The movement data for the second experiment was collected by Prof. Dr. van der Spek and colleagues from TU Delft in a campaign with 547 participants in the city of Rotterdam, Netherlands. People arriving with their cars in one of three different parking decks were asked to record their city visit with a GPS receiver between May 3, 2012 and May 6, 2012. Therefore, the single trips, with a sampling rate of typically 5 *seconds*, are usually rather short. Foursquare venues were collected with the python-based Foursquare API "foursquare" (version 2), where around 16'000

unique locations within the area of Rotterdam covering over 350 different categories were received (date of collection: 13/06/2012).

## 6.2. Experiments

### 6.2.1. Relating speed and slope

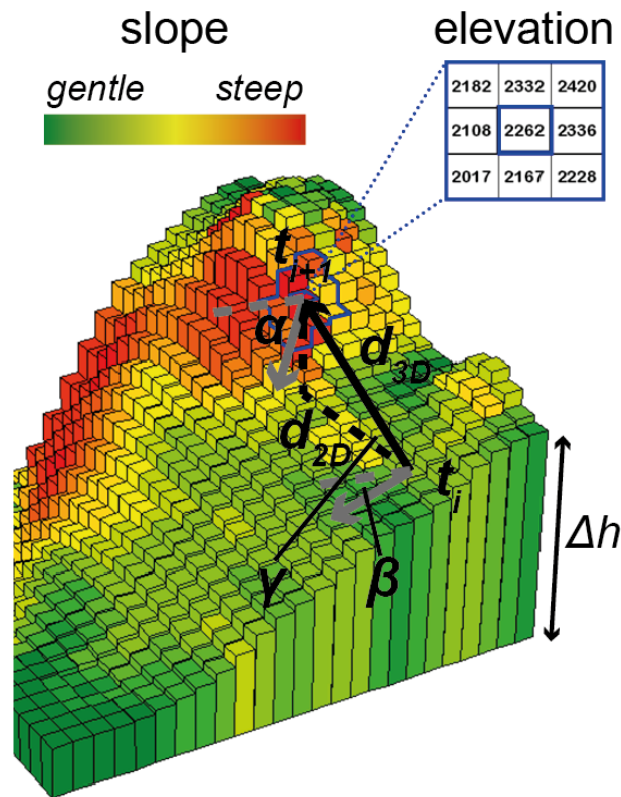
This subsection provides the experiment where the speed of bikers and hikers is related to terrain slope. These two parameters representing specific aspects of movement and the surrounding geographic context are chosen, owing to the speed of outdoor sports activities in the mountains being expected to be highly related to slope. Correlation analyses are performed, whereas different quantitative methods are applied in mind of capturing speed and slope (Figure 6.1). This empirical study aims at garnering insight into the effect of different quantitative methods on the correlation of certain characteristics of movement and geographic context.

On the one hand, slope is calculated based on the classic approach by HORN (1981) using the DEM (Figure 6.1, finite differences, angles  $\alpha$  and  $\beta$ ), whilst on the other hand, slope is modeled as the decline of the beeline between two consecutive GPS fixes (Figure 6.1, angle  $\gamma$ ), since the paths of bikers and hikers usually do not follow the gradient (steepest slope, e.g.,  $\alpha$ ) of the terrain. Further, speed is computed taking into account different measures of distance between two consecutive GPS points, where firstly distance is calculated in 2D Euclidean space (Figure 6.1, distance  $d_{2D}$ ) and secondly, elevation is considered for computing the straight-line distance in 3D (Figure 6.1, distance  $d_{3D}$ ). More specifically, in terrain with steep slopes, differences between 2D and 3D speed are expected.

The preprocessing and statistical analysis is realized with the adoption of R statistics. Potential outliers are filtered with a speed (computed in 2D between consecutive GPS points) threshold of 40 *kilometers per hour* for bikers and 15 *kilometers per hour* in the case of hikers. Notably, movements at low speeds are prone to be related to extremely high slopes, in the case of the slope of the beeline. Firstly, since the spreading of the GPS data is larger at low speeds, in steep terrain, the difference between the related and the real slope increases. Secondly, at low speeds, the planar distance traveled within 5 *seconds* is fairly small (less than 1 *meter*); at the same time, however, in steep terrain, the difference in elevation from one pixel to another can be large in proportion to the planar distance, which leads to unrealistic slope values. For these two reasons, the movement data is filtered with regard to the slope of the beeline at a threshold of  $\pm 45$  degrees. First, different regression models were applied, and the scaling (e.g., linear, or logarithmic) of the dependent and independent variables was varied as well. Then, the best performing regression model and variable scales were chosen. Consequently, for the correlation analyses with slope as the independent and speed as the dependent variable, a linear fit model, within which the logarithm of speed is considered, is applied. Separate linear fit models are computed for downhill and uphill, since different relationships between speed and slope are expected for



different sorts of terrain in terms of slope. Downhill and uphill slopes are mirrored in negative and positive values, respectively.



**Figure 6.1.:** Different approaches for computing slope and speed (distance). Adapted from [http://www.innovativegis.com/basis/present/gita\\_denver05/default.htm](http://www.innovativegis.com/basis/present/gita_denver05/default.htm), accessed on 29/12/2014.  $\alpha$  &  $\beta$ : gradients from DEM,  $\gamma$ : slope of the beeline,  $d_{2D}$  &  $d_{3D}$ : 2D and 3D distances between consecutive GPS fixes.

### 6.2.2. Relating movement to user-generated content

An empirical study with movement data of pedestrians in Rotterdam and Fourquare venues representing geographic context is conducted. From the movement data the information of where and when people spend their time is gathered, whereas the geographic context in form of the Foursquare venues should allow for gaining insights about the reasons of the people's stay in a certain area of the city.

Firstly, the initial categories are reclassified to the 9 top level Foursquare categories. The Foursquare records with the category "Residence" are removed owing to the fact that these entries are not valid and do not make sense in terms of the covered locations with an unrealistic regular pattern across the city. For the whole movement data set, as well as each of the 8 remaining Foursquare categories, a KDE with a

radius of 100 *meters* and a resulting cell size of 10 *meters* is performed with the aim of achieving representations of the spatial distribution of the movement and the single Foursquare categories (Chapter 4, *mc*-relation matrix,  $m_{FCF}$ ). The resulting KDEs are normalized with the maximal cell value. This means that, for every KDE (movement as well as Foursquare categories), every raster cell is divided with the maximal cell value of the whole raster. Therefore, for each KDE, the maximal value gets the value “1”. This normalization takes into account the fact that the categories are not equally distributed with regard to the amount of instances, but, for example, “shop/service” should not be considered more relevant than “arts/entertainment” simply because of the larger number of POIs with shops and services available in Rotterdam than locations with arts and entertainment.

Figure 6.2 illustrates the methodology of the relation of the movement of pedestrians and the Foursquare categories on the basis of the computed KDEs. The relation of movement and context is realized with a pairwise multiplication of the movement KDE (KDE of whole data set, stops or moves), with each of the Foursquare category KDEs. This means that, for every Foursquare category, a new raster is computed, which depends on the movement and the corresponding context category, based on map algebra (Section 4.5 and Figure 4.1,  $m_{FCF}$ , according to TOMLIN, 1983). The maximal possible value of the multiplied KDEs is “1”, in the case of overlaying maximal cell values in the movement and the Foursquare KDE raster. Most often, multiplied KDEs do not comprise the maximal value “1” simply because maximal values of movement and context rasters rarely overlap.

Let us consider the lower left corner in Figure 6.2 in order to make an illustrative example of the relation methodology. The lower left corner comprises the following values for the movement, and the brown and the gray Foursquare category: 0.5, 0.1 and 0.0, respectively. Therefore, the pairwise multiplication of the movement and the Foursquare category raster results in the values of 0.05 ( $= 0.5 * 0.1$ ) for the brown category and 0.0 ( $= 0.5 * 0.0$ ) for the gray category. Further, the category with the maximal value in the multiplied rasters (brown category in the example of the lower left corner) is considered the most valid reason for the pedestrians to be in the actual region of the city, and therewith assigned to the corresponding cell in a new raster (brown raster cell at lower left). The reasons for single persons remaining in certain areas of the city obviously can be completely different from the categories given in the Foursquare data. Therefore, the final classification of the city into Foursquare categories should only be understood as plausible hints for reasons of people’s stay. The resulting classification of the city of Rotterdam into the Foursquare categories, as based on KDE representations of movement and context (Foursquare categories), is shown in Figure 6.7.

The same methodology is applied to subsets of the movement data, namely on KDEs of stops and moves, where stops and moves are separated with a speed threshold of 2.5 *kilometers per hour*. The aim of this further investigation is centered on establishing whether people stop for different reasons than they move on at certain locations in the city of Rotterdam.

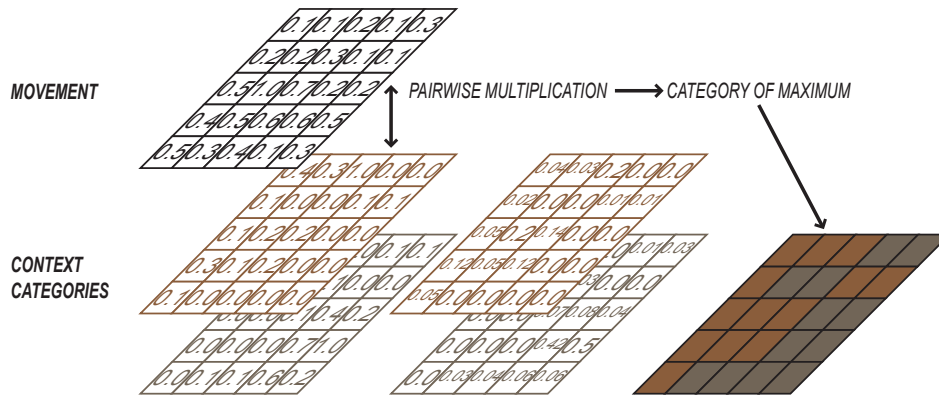


Figure 6.2.: Methodology for the relation of movement of pedestrians in Rotterdam and context in form of Foursquare categories based on KDE (author’s own graph).

## 6.3. Results

### 6.3.1. Relating speed and slope

Figure 6.3 shows the results of the correlation analysis for the different quantitative models of slope. The results of the regression models are represented in blue and green for downhill and uphill movements, respectively. In the case of slope derived from the DEM (Figure 6.1,  $\alpha$  or  $\beta$ ), no correlation to speed is found for bikers or for hikers, but a significant correlation with an  $R^2$  of around 0.1 for bikers (0.1 for downhill and 0.14 for uphill) and hikers (0.05 for downhill and 0.07 for uphill) results for the relation of the slope as the decline of the beeline (Figure 6.1,  $\gamma$ ) to speed. Although the results of the regression are significant at a significance level of 0.05, the correlation coefficients are very low. Higher correlation coefficient values are revealed with the slope as the decline of the beeline ( $\gamma$ ), most probably because slope, as based on the DEM ( $\alpha$  or  $\beta$ ) mirrors the steepest slopes on the terrain’s surface, whereas the slope as the decline of the beeline ( $\gamma$ ) rather reflects the actual slope of the movement path. Another reason for this result could be that in mountainous steep terrain, the DEM slope is more sensitive to the positional inaccuracy of the GPS data than the slope of the beeline. However, the interplay of steep slopes and inaccuracies in GPS positions also lead to low correlations for the slope as the decline of the beeline.

The results with different slope models even show different patterns. For the slope as the decline of the beeline, people tend to move more slowly in steeper terrain – equally for downhill as for uphill movements of bikers and hikers. In case of the DEM slope, no clear trend is detected. Although the results are not statistically meaningful, the difference as a result of the different quantitative methods for computing slope is remarkable with regard to Tobler’s hiking function (TOBLER, 1993). For the hiking data, the result empirically mirrors the hiking function presented by TOBLER (1993), where hiking speed is formalized as an exponential function of the slope of the beeline ( $\gamma$ ), with a maximum speed (around 6 kilometers per hour) at a slightly negative

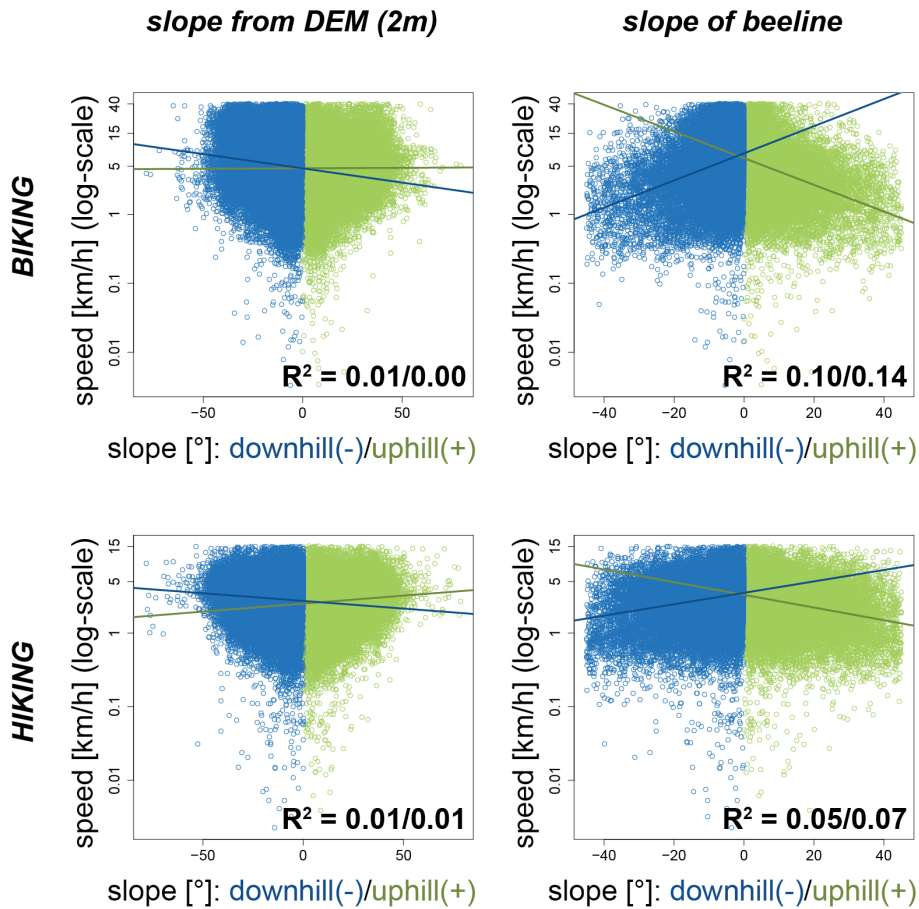
slope (gentle downhill slope). Slope models are not a strong focus of this thesis, but rather should illustrate the remarkable impact of different models for representing the same aspect of the geographic context. Therefore, for more details about Tobler's hiking function, TOBLER (1993) can be considered.

For bikers we did not expect this result since downhill movement is supposed to be faster in steep slopes. This expectation rather can be illustrated with single down- and uphill trails of bikers (Figure 6.4a-d) where, for slopes between -10 and -5 *degrees*, higher speeds for steeper slopes are revealed (Figure 6.4d). Accordingly, specific parts of the movement paths confirm expectations for bikers, because in some parts of the terrain, the sensitivity of slope to positional inaccuracy of the GPS points is less pronounced. Moreover, with the aggregation of the data of all bikers, the expected result is masked.

In order to support the line of argumentation with regard to the sensitivity of slope to positional inaccuracies in GPS data, the correlation analyses were also computed with the "DHM25" ("Bundesamt für Landestopographie", version 2005), which offers a coarser resolution of 25 *meters* (instead of 2 *meters* for the swissALTI3D). The results with the use of the "DHM25" show the same patterns in terms of different slope and speed models, and therewith, are not graphically presented in detail here. However in case of the slope of the beeline, significantly higher  $R^2$  values of around 0.4 and 0.2 resulted for bikers and hikers, respectively, because the DEM with the coarser resolution ("DHM25") is smoother than the swissALTI3D, and therefore, is less sensitive with regard to the positional inaccuracy of the GPS data. However, the results in connection with the swissALTI3D are presented here, since a finer resolution of the DEM is more realistic in mountainous regions with steep terrain. This additional experiment shows that changing strength of the sensitivity effect at different spatial scales of the geographic context is crucial for the relation of movement to its embedding geographic context, and further provides evidence for the necessity of the scale studies in Chapter 7.

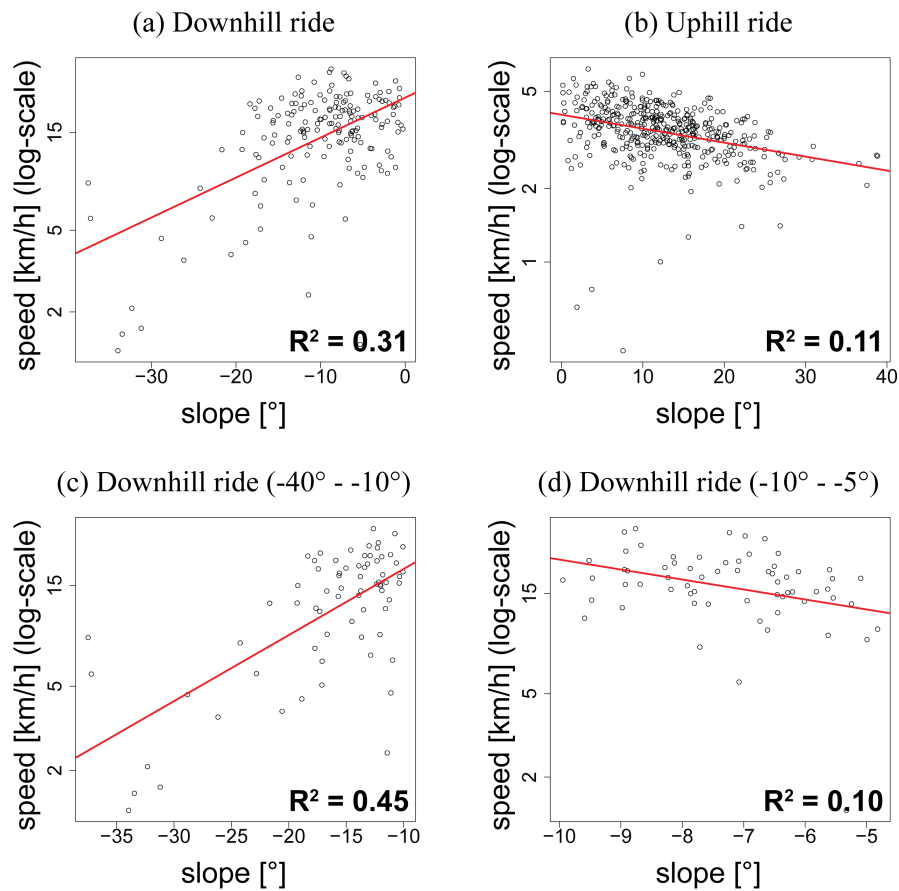
Further, the complexity of the trails and the capabilities of bikers and hikers might be explanations for low correlation coefficients in general, as well as for the unexpected result for downhill biking, which is surprisingly close to the outcome for hikers in particular. A better correlation is found for bikers than for hikers with the slope as the decline of the beeline ( $\gamma$ ), since the sensitivity of slope to positional inaccuracy of GPS data is lower at higher speeds for bikers than for hikers. Moreover, bikers and hikers are able to move at different speeds in a terrain with constant slope. Although the spreading of speed values at fix terrain slopes decreases towards steeper terrain, a perfect correlation cannot be expected between speed and slope.

The sensitivity of the correlation analysis, with regard to different measures of speed, is provided in Figure 6.5. Taking into account distance in 3D (Figure 6.1,  $d_{3D}$ ) for computing speed based on consecutive GPS points is intended to be closer to reality than merely calculating speed values in 2D Euclidean space (considering distance in 2D, Figure 6.1,  $d_{2D}$ ). Although we would expect higher correlation values for speed considering elevation (or distance in 3D,  $d_{3D}$ ), correlations for speed based on  $d_{2D}$  with slope as the decline of the beeline ( $\gamma$ ) show higher correlation coefficients



**Figure 6.3.:** Correlation of speed (logarithmic scale) and slope using a linear fit model for bikers and hikers around the Fuorn pass considering two different quantitative models of slope. Note the cut off at 40 and 15 kilometers per hour to remove outliers.

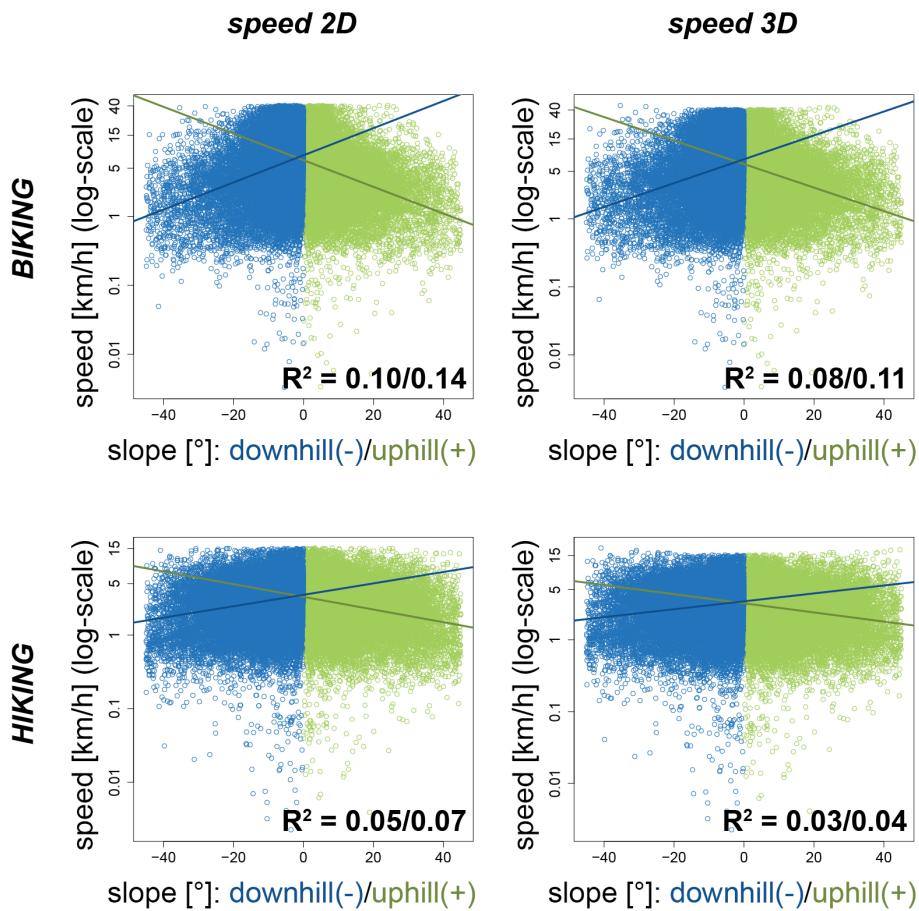
for biking as well as for hiking. I would argue that introducing another dimension from a different data source (DEM) also implies more uncertainty for the resulting distances in 3D ( $d_{3D}$ ). This could lead to an overestimation of speed values, whereas, in the case of speed values considering  $d_{2D}$ , this effect is compensated, to some degree, with a slight underestimation. Underestimating speed values, as based on distances in 2D Euclidean space ( $d_{2D}$ ) seems to be less significant than the overestimation in case of considering elevation (or distance in 3D,  $d_{3D}$ ) and its uncertainty for computing speed.



**Figure 6.4.:** Correlation of speed (logarithmic scale) and slope using a linear fit model of single down- and uphill trails around the Fuorn pass considering the slope of the beeline. (a) Single downhill trail of a biker, (b) Single uphill trail of a biker, (c) Single downhill trail of a biker for slopes between  $-40^\circ$  and  $-10^\circ$ , (d) Single downhill trail of a biker for slopes between  $-10^\circ$  and  $-5^\circ$ .

### 6.3.2. Relating movement to user-generated content

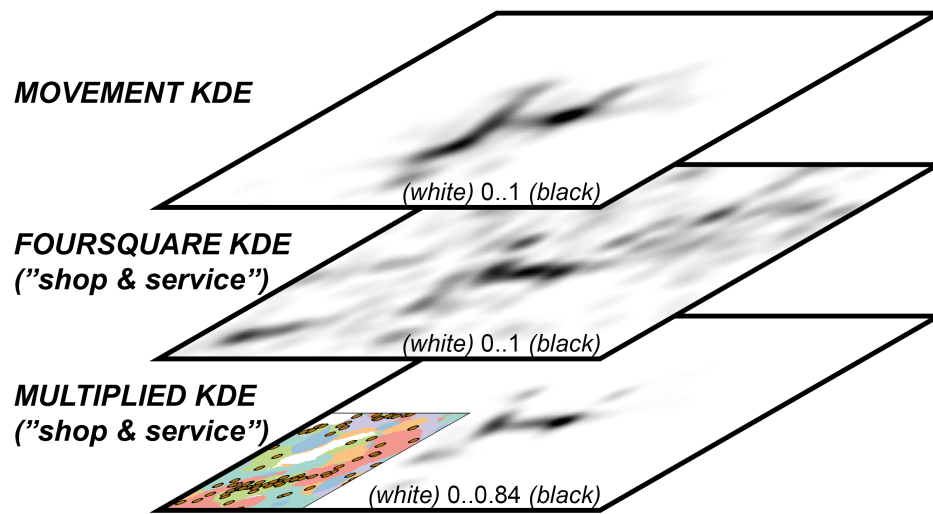
Figure 6.6 shows the resulting movement KDE for the entire data set, a Foursquare KDE for the “shop & service” category and the pairwise multiplication of the two. The movement KDE as well as the Foursquare KDE are normalized, meaning that the maximal values for the rasters are “1”, as explained earlier. Since the high values in the movement KDE overlap with the high values in the Foursquare layer, the pairwise multiplied KDE shows a relatively high maximal value of “0.84” (close to “1”). This spatial correlation between the movement and the context raster explains, to some degree, that the resulting classification in Figure 6.7 in the center of Rotterdam (green arrow number (1)) mainly comprises the Foursquare category “shop & service”. The



**Figure 6.5.:** Correlation of speed (logarithmic scale) and slope using a linear fit model for bikers and hikers around the Fuorn pass considering two different quantitative models of speed. Note the cut off at 40 and 15 kilometers per hour to remove outliers.

pairwise multiplied KDE includes an extent of the resulting Foursquare classification and the original point locations of the “shop & service” category at the lower left. This cut-out illustrates that also smaller connected areas in the resulting Foursquare classification usually originating of more than only one point location.

Each of the resulting categories mostly comprises relatively large connected areas, thus meaning that the spatial autocorrelation is relatively high, and the categories are not distributed randomly across the city of Rotterdam. Since a ground truth for the evaluation of our result does not exist, some specific aspects are discussed. In Figure 6.7, green and red arrows point out the illustrative examples for discussion, distinguishing between results that rather make sense (green, no. 1-3) and findings that are not explicitly comprehensible (red, no. 4-6). Therefore, the model is validated

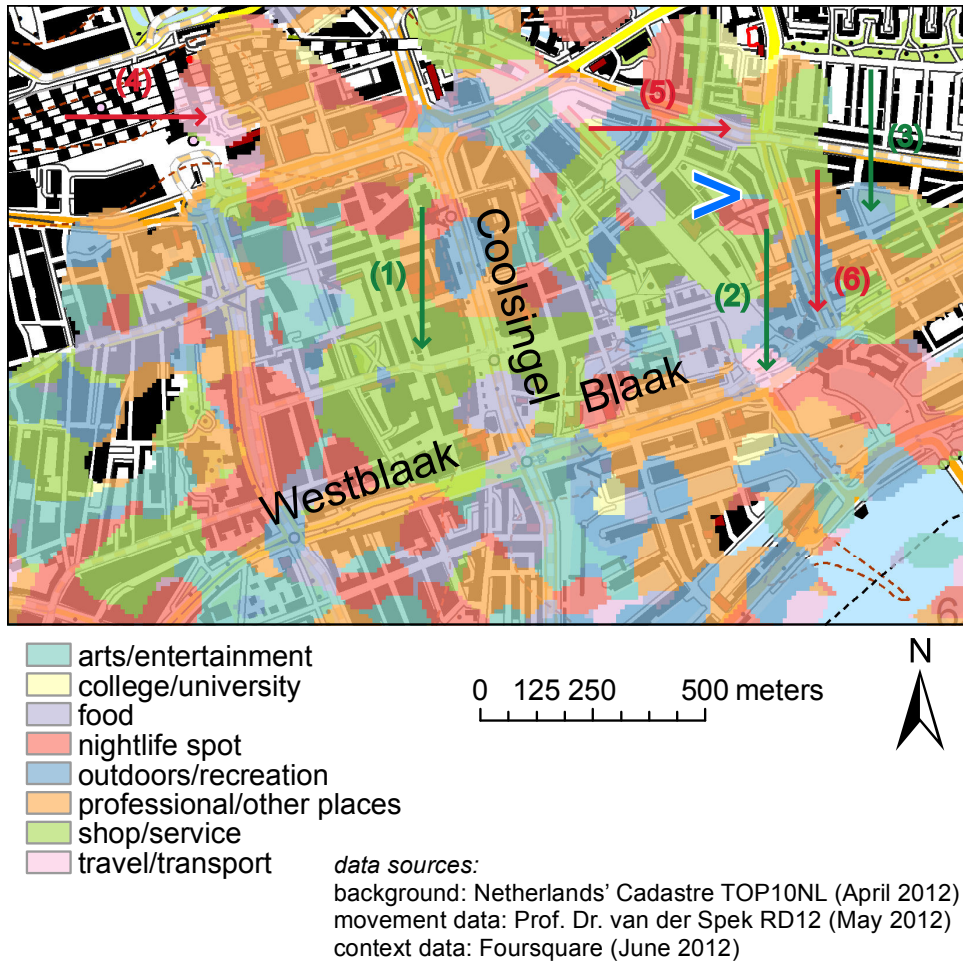


**Figure 6.6.:** Resulting KDEs for the movement (all data), a Foursquare category (“shop & service”) and the pairwise multiplication of the two with a small extent of the resulting classification and the original point locations of the “shop & service” Foursquare category.

using *face validity* (knowledge of local people) and *visualization techniques* (overlay with a basemap), according to RYKIEL JR. (1996). Result no. 1 of the category “shop/service” covers the well-known shopping area in the city core of Rotterdam quite well. Finding no. 2 (“travel/transport”) refers to the Rotterdam Blaak station. The area no. 3 is classified with the Foursquare category “outdoors/recreation” and covers a relatively large meadow or park, which could be used for recreation. The red arrow no. 4 points out the main station of Rotterdam. In and around the main station, I would expect a larger area of the category “transport/travel”. For the red arrows with the numbers 5 and 6, the reasons for the actual classifications are not directly obvious, and with the current knowledge these findings do not make sense.

Unfortunately, the resulting classifications of the city of Rotterdam into Foursquare categories, based on the KDEs of stops or moves separately, and the categories themselves, did not differ from the presented result using the whole data set. This finding indicates that there are no particular regions, at which people stop more often; they stop and move equally in all the regions they have visited. Therefore, the stops are spatially distributed in a similar way as the moves and as the whole data set. Therefore, the classifications are also the same for the subsets of the movement data. Moreover, KDE might overly aggregate and smooth the differences between stops and moves in order to be mirrored in the relation of movement and geographic context.





*Figure 6.7.:* Resulting classification of the relation of movement of pedestrians in Rotterdam and context in form of Foursquare categories based on KDE.

## 6.4. Main findings and contribution

The speed-vs-slope experiment was very simple; nevertheless, it made a very important contribution. Using out-of-the-box tools (just because they are easily accessible) has the ability to trigger an inappropriate movement-context relation (speed at a point vs slope of terrain as the gradient using a DEM). For an investigation centered on whether or not terrain slope influences movement speed, it is not the slope of the terrain from a DEM that is relevant, but rather the slope of the actual movement path (approximated with the slope of the beeline between two consecutive GPS fixes). Although this seems trivial, the present study quantifies this effect. Further, the first investigation in the form of the correlation analyses illustrates that, besides many different factors (e.g., data structure, scale), the methods for quantifying the specific phenomena of the surrounding environment play an important role in modeling

geographic context. The same applies to the quantification of movement characteristics (e.g., speed).

The UGC experiment really only scratched the surface in terms of possibilities with UGC as a source of context. Nonetheless, it also showed the notorious problems with UGC. First, the lack of validation opportunities due to lacking ground truth: even though it appears very tempting at first sight, the use of UGC as a context source is problematic. Moreover, although the methodology of the second experiment could be improved in many aspects, the findings demonstrate that geographic context in the form of UGC has the potential to garner greater insight into the reasons behind the actual movement.

# Chapter 7.

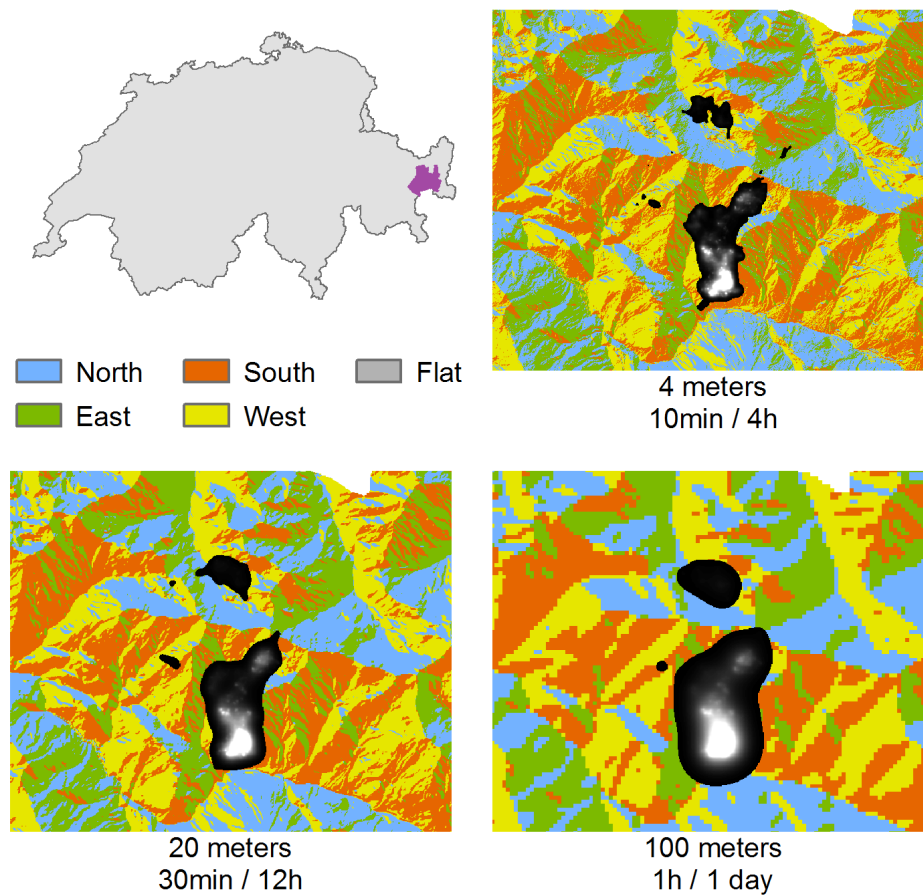
## Scale effects in context-aware movement analysis

### 7.1. Study area and data

In this case study, GPS movement data of ungulates (seven chamois, ten deer and eight ibexes) from the Swiss National Park is related to the aspect of the terrain. The experiments assess in what aspect classes the animal move, whilst systematically varying the temporal scale of the movement data and the spatial and thematic scales of the aspect. Exactly the same movement data and DEM was already used in the case study concerning relation methods in Chapter 5. Figure 5.2 (p.56) again illustrates the space use of the different species of ungulates. An overview of the study area in the South-East of Switzerland and illustrative examples of aspect on different spatial scales and the BBMM on various temporal scales is given in Figure 7.1.

The aspect is derived from a DEM with a resolution of 4 *meters* using ArcGIS 10.0, whereas the aspect for the coarser resolutions is computed based on aggregated versions of the 4-*meters*-DEM. For the aggregation to coarser resolutions of 20 and 100 *meters*, averaged cell values of the 4-*meters*-DEM are calculated. Figure 7.1 illustrates the aspect in 5 classes for the three spatial scales. In addition, thematic scales of 9 and 17 classes are considered in the quantitative experiment, as well in order to assess the influence of the different thematic scales on the aspect's distribution and its interplay with spatial and temporal scales. Specific characteristics of the movement data are summarized in Table 5.1 (p.53) for the original form of the movement data, corresponding to the finest temporal scale within this investigation, where a smaller sampling rate occurs on every second Wednesday for most of the chamois and some of the ibex.

The coarser temporal scales of the movement data are generated considering every third (e.g., for chamois, sampling rate: 30 *min* every second Wednesday / 12 *h* for the rest of the data) and every sixth (sampling rate: 1 *h* / 1 *day*) GPS fix, respectively. As a consequence, the number of GPS fixes is reduced by the factor of three and six. In Figure 7.1, movement on different temporal scales is presented with the raster-based BBMM (HORNE *et al.*, 2007), whereas coarser temporal scales appear with less detail and smoother. Moreover, the area covered by the 99% volume contours is getting larger, with increasing temporal sampling rate since, with less GPS fixes, more uncertainty is introduced.

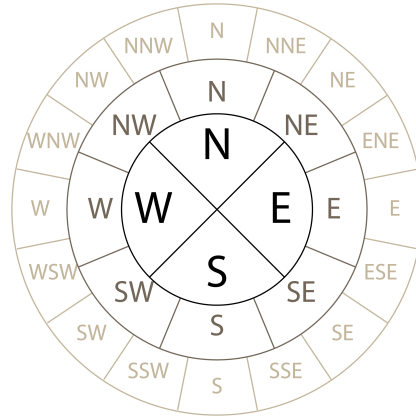


**Figure 7.1.:** Aspect represented in 5 classes on different spatial scales (4, 20 and 100 *meters*), and a chamois' (no. 22) movement modeled with the BBMM (increasing probability density from black to white) on different temporal scales (10 *min* / 4 *h*, 30 *min* / 12 *h* and 1 *h* / 1 *day*) around the Fuorn pass within the Swiss National Park (purple).

## 7.2. Experiments

The experiments relating movement to aspect are carried out at three spatial scales (4, 20 and 100 *meters* raster resolution), three thematic scales of the context (5, 9 and 17 classes), and three temporal scales of the movement (10 *min* / 4 *h*, 30 *min* / 12 *h* and 1 *h* / 1 *day*). Aspect is selected as a context factor because it is simple, meaning that it does not introduce much more complexity into the analysis. Furthermore, this terrain parameter is intended to be relevant for the movement of ungulates. Spatial scales are selected between 4 and 100 *meters* since this covers a reasonable range of different levels of detail at which terrain is typically modeled. Moreover, for the thematic scales, it is straightforward to classify aspect in the specified number of classes. With the temporal sampling rates of the GPS fixes, various scales within the order of a day

are chosen. The terrain parameter aspect shows a particular characteristic in terms of the thematic scale, since classes at finer thematic scales are not necessarily contained by classes at coarser thematic scales (Figure 7.2). Figure 7.2 illustrates that, for instance, North (N) at a thematic scale of 17 classes is contained by the North classes of the coarser thematic scales, but e.g., North-East (NE) at the thematic scales of 9 and 17 classes obviously overlap with the classes North and East (E) at a thematic scale of 5 classes.



**Figure 7.2.:** Aspect classes at different thematic scales (without class Flat, author’s own graph).

Movement is modeled in two different ways (similarly to the methodology in Chapter 5) in order to discover whether the effect of different scales is dependent on the chosen conceptual movement model. First, movement is represented as the mere GPS points. As a second model, the BBMM is realized, representing movement in the form of a probability density surface as a raster (HORNE *et al.*, 2007). This means that movement and the terrain’s aspect are related based on the GPS fixes and the BBMM within the 99% volume contours. The 99% volume contours are calculated using the Geospatial Modelling Environment (GME) (BEYER, 2012). The 99% volume contours were chosen in order to delineate the BBMM since more than only the core area and therewith, as well as locations with low probability density, shall be covered. In this experiment, the BBMM is computed with the “BBMM”-library in R statistics (NIELSON *et al.*, 2011), with a resulting cell size of 20 *meters*, a location error of 10 *meters* and the maximal allowed temporal gap between consecutive GPS points is set to 5.6 *hours* (20’000 *seconds*), 16.7 *hours* (60’000 *seconds*) or 33.3 *hours* (120’000 *seconds*) corresponding to the different temporal scales in increasing order. Most of the GPS fixes are regularly sampled with the specified temporal scales; in practice, however, due to technical and environmental reasons (e.g., terrain-related heterogeneity of the GPS signal), the effective sampling rate differs from these values. For this reason, the thresholds for the maximal temporal gap allow that the sampling rate can be up to 40% larger than the largest expected temporal scale in order to be taken into account for the computation of the BBMM.

The point-based movement model is related to the terrain parameter through considering the aspect values in the exact location of the GPS fixes, which relate to the map pin approach presented in Section 4.2. In the case of the BBMM, each cell of the BBMM-raster is related to the nearest neighbor in the raster representing the terrain parameter, and the probability density value is used to weight the aspect. Consequently, the more frequently a location is visited and the longer an ungulate stays at a certain place, the more importance is assigned to the corresponding aspect of the terrain. The relation procedure is implemented with the programming language JAVA.

Relative distributions for different combinations of spatial, temporal and thematic scales are statistically analyzed by quantitatively assessing the differences using the coefficient of variation. The coefficient of variation is a normalized measure of dispersion and is represented by the ratio of the standard deviation to the mean (equally to Chapter 5). Thus, this statistical measure allows the quantitative comparison of variation within the context categories for various spatial, temporal and thematic scales. Barplots and coefficients of variation are computed with the R statistics software.

### 7.3. Results

Chamois, deer and ibexes express different terrain aspect preferences. Throughout the course of this experiment, focus has been directed towards the differences between the resulting context distributions caused by different kinds of scale or relation methods; the shapes and patterns of the actual distributions (e.g., general preference of southerly aspects) are of minor interest. Since the general scale effects are very similar for the different species, we only present and discuss the results for the chamois here.

The statistical results are presented in Figure 7.3, which consists of a “time”, “space”, “method” and “coefficient of variation” column. The different rows represent the variation of the thematic scale of the geographic context (5, 9 and 17 categories). In the first three columns (“time”, “space” and “method”), the bars in each category illustrate the relative distribution of movement context for that category when varying temporal scales, spatial scales, and relation methods. The last column (“coefficient of variation”) shows the coefficients of variation, where the bars reflect the within-class variation of the first three columns in each category.

- The “time” column illustrates for three different context scales the resulting context distribution for three systematically varied temporal scales ( $10min/4h$ ,  $30min/12h$ ,  $1h/1d$ ) when the spatial scale is kept constant at 4 meters (@4m).
- Accordingly, the “space” column demonstrates the effects of systematically varied spatial scales of the geographic context (4m, 20m, 100m), for the three thematic scales at a fixed temporal scale of the original dataset of  $10min/4h$  (@10min/4h).

- The “method” column illustrates the sensitivity to the chosen relation method, again for three thematic scales. Here, the resulting context distributions are plotted relating to a point-based (map pin) or a raster-based (BBMM) movement model.
- The “coefficient of variation” column assesses the variation within single categories (e.g., South aspect) for each of the columns “time”, “space” and “method”. The coefficient of variation is calculated as the ratio between the standard deviation and the mean of the relative values per category. In other words, the coefficients of variation in the last column of Figure 7.3 is based on the relative distribution values established in the corresponding columns “time”, “space” and “method”. In the color version, the colors in the last column correspond to the colors of the three first columns: orange for “time”, green for “space” and blue for “method”. Study, for example, the dashed box in Figure 7.3 (17 categories, South): for aspect South, the bars in the “space” and “method” columns vary more than the bars in the “time” column. This is mirrored in the corresponding coefficients of variation: high values for “space” (0.23) and “method” (0.16), a low value for “time” (0.004).

Since the characteristics of the results with regard to the within-class variations and the interrelation between the different sorts of scale are similar for the spatial scales of 20 and 100 *meters* related to the “time” column, for the temporal scales of 30min/12h and 1h/1d related to the “space” column, and also for the BBMM-based relation method, only the described specific results are presented. Nevertheless, there are various specific cases of scale interdependencies remarkably differing from these patterns, which are addressed with the results illustrated in Figure 7.4.

### 7.3.1. Differences due to scales and methods

The resulting relative distributions in Figure 7.3 show roughly the same pattern across different temporal, spatial and thematic scales, and also when varying the method used for relating movement to the underlying geographic context. For chamois, the context factor aspect shows roughly a normal distribution around South. This means that chamois predominantly move on South exposed slopes. However, remarkable variations can be found within categories when varying scales and relation methods. Figure 7.3 indicates the following findings:

- The variation in the context distributions due to different spatial scales (coefficient of variation around 0.2) is larger than the variation caused by the various temporal scales (coefficient of variation around 0.02) (Figure 7.3, “coefficient of variation”). Spatial scale effects are reflected in differences of up to more than 13% in the relative distribution of the context factor aspect for the chamois (Figure 7.3, “space”, 17 categories, South; e.g., 13.2% for 4m vs. 100m).
- Varying the temporal scale of the movement data has little effect on the resulting context distributions, the differences are negligible (Figure 7.3, “time”).

- The different relation methods (map pin vs. BBMM-based) have a notable effect on the distribution of the context factor, but the general pattern of aspect preference remains the same (normally distributed) (Figure 7.3, “method”).
- The variation due to the different relation methods is in the same order of magnitude of the variation caused by the spatial scale (Figure 7.3, “coefficient of variation”).
- The coefficients of variation express some negative correlation to the relative distributions; in other words, categories with higher relative values of distribution (e.g., South, 17 categories, dashed box) show in general smaller coefficients of variation than categories with a smaller share (e.g., North, 17 categories).

### 7.3.2. Interdependencies between scales

The findings referring to the interdependencies between scales are illustrated with Figure 7.4. Figure 7.4 shows only coefficients of variation. These coefficients again are used to quantify the within-category variation that results when systematically varying the temporal and spatial scale. The coefficients are computed in analogy to the procedure applied for the variation values in Figure fig:chamoisStatistics (“coefficient of variation” column). However, in Figure fig:chamoisStatistics, temporal scale effects (“time” column) are presented for a fixed spatial scale of 4 *meters* (@4*m*). Correspondingly, the spatial scale effects (“space” column) are illustrated for a fixed temporal scale of 10*min/4h* (@10*min/4h*). By contrast, Figure 7.4a shows the coefficients of variation resulting from a systematic variation of the temporal scale (10*min/4h*, 30*min/12h*, 1*h/1d*, “temporal scale effects”) on all the spatial scales of the geographic context (4*m*, 20*m* and 100*m*). Accordingly, Figure 7.4b shows variations caused by systematically varying spatial scale (4*m*, 20*m*, 100*m*, “spatial scale-effects”) for all the temporal scales of movement (10*min/4h*, 30*min/12h* and 1*h/1d*). For example, the dashed box in Figure 7.4a (17 categories, “North”) shows the coefficient of variation values, representing variation due to different temporal scales (10*min/4h*, 30*min/12h*, 1*h/1d*) ranging between 0.02 and 0.09 for different spatial scales. Hence, the coefficients of variation in Figure 7.4a reveal whether or not temporal scale effects depend on the spatial scale of the geographic context. Likewise, in Figure 7.4b the potential dependencies of spatial scale effects on different temporal scales of the movement can be studied. Figure 7.4 indicates the following findings:

- It is clear, the more categories of aspect that are introduced, the smaller the relative values per category, and the higher the potential coefficients of variation grow (ranging between 0 and max. 0.2 for 5 categories, and between 0 and max. 0.6 for 17 categories).
- Although the differences due to temporal scale are not prominent for all the different spatial scales, these differences become more pronounced with coarser spatial scale (Figure 7.4a). Therefore, variations caused by different temporal scales, to some degree, depend on the spatial scale of the geographic context.



- Contrarily, spatial scale effects are more stable with regard to different temporal scales (Figure 7.4b). This means, in this specific example, that effects triggered by various spatial scales are independent of the temporal scale at which the movement data is sampled.
- The coefficients of variation furthermore reveal a tendency of higher in-class-variation for increasing thematic scale irrespective of the spatial and temporal scale (Figure 7.4).

## 7.4. Main findings and contribution

When varying the spatial scale of the geographic context, we found differences for individual context categories of up to two-digit percentages. This finding accords with our expectations, as well as for the general GIScience and movement ecology literature for that matter. Less expected was that the chosen temporal scales, ranging from hours to the order of magnitude of a day, are much less relevant. This finding suggests that chamois do not have specific preferences with regard to the terrain's aspect, varying on hourly or daily basis; however, this does not mean that temporal scale has no effect at all. Figure 7.5 illustrates that, on a temporal scale of a half-year (e.g., winter vs. summer), for instance, seasonal differences, such as in terms of terrain aspect, can be discovered, whereas chamois prefer South slopes more (around +20%) in winter than in summer.

Furthermore, the spatial scale effects are stable with regard to different temporal scales (Figure 7.4b) owing to the chosen temporal scales being irrelevant in relation to the terrain's aspect. For a seasonal temporal scale of the movement, spatial scale effects are expected to be less stable. As such, this result does not allow the conclusion to be drawn that spatial scale effects generally are independent of the temporal scale of the movement; rather, this reveals a complex interplay of all three scales (temporal, spatial and thematic) with the chosen analysis method, which hardly allows for universally valid statements.

Our explanation for the negative correlation of the relative distributions of the geographic context and the coefficients of variation is the law of large numbers. Low relative values for a specific class of the geographic context emerge owing to a smaller number of samples in that class. For the relative distributions that are compared, only a single factor is varied (e.g., temporal scale of movement); as a consequence of the law of large number, however, classes with low sample sizes are less balanced with regard to variation. Therefore, resulting coefficients of variation are prone to being higher for categories of the geographic context with low values in the relative distribution. Using the same line of argumentation, the tendency of higher coefficients of variation with finer thematic scale can be explained when considering that class sizes become smaller with the addition of more classes. However, variations due to different temporal scales of movement are relatively small in comparison to the variations caused by different spatial scales of the geographic context across all

categories. We argue that the effect of negative correlation is not so strong that it masks the actual scale effects. Independent of the reasons for the higher coefficients of variation with increasing number of context categories, thematic scale can have an impact on the variation established owing to different temporal and spatial scales.

The methodology proposed in this case study is transferable to many other sources of movement data and their respective geographic context factors. In this experiment, many different factors (temporal, spatial and thematic scale, relation method, and species) were varied. Accordingly, our cross-scale sensitivity study was limited to only three scale levels for the temporal, spatial and thematic scale, and to only two relation methods, in order to ensure the complexity of the experiments was not increased too much (combinatorial explosion). However, we argue that the results are representative.

In GIScience, it is common for only a single type of scale to be investigated – most prominently the spatial scale, but also temporal scale of movement and its impact on the computation of trajectory-based parameters, such as speed and turning angle, for example (LAUBE & PURVES, 2011). Within movement ecology, for instance, BÖRGER *et al.* (2006) investigated the variance of home range size depending on temporal, spatial and thematic scale effects; however, interdependencies between the different sorts of scales were not considered by the authors. In this case study, we contribute to achieving a better understanding of the role of different types of scale – and, more specifically, their interdependencies within the relation of movement to the surrounding geographic context.

Finally, the variation due to the different relation methods confirms the findings in Chapter 5, where we empirically illustrated that different relation methods can significantly influence the resulting relative distributions of the geographic context, legitimizing the structured overview of relation methods based on different conceptual models for movement and geographic context (Chapter 4).

In these experiments, the complex interplay between different types of analysis scales was illustrated when relating movement to its embedding geographic context with an empirical study. In the case study, the resulting scale effects and the interdependencies of different scales showed similar patterns independent of the species (chamois, deer and ibex) and the relation method (map pin vs. BBMM-based).

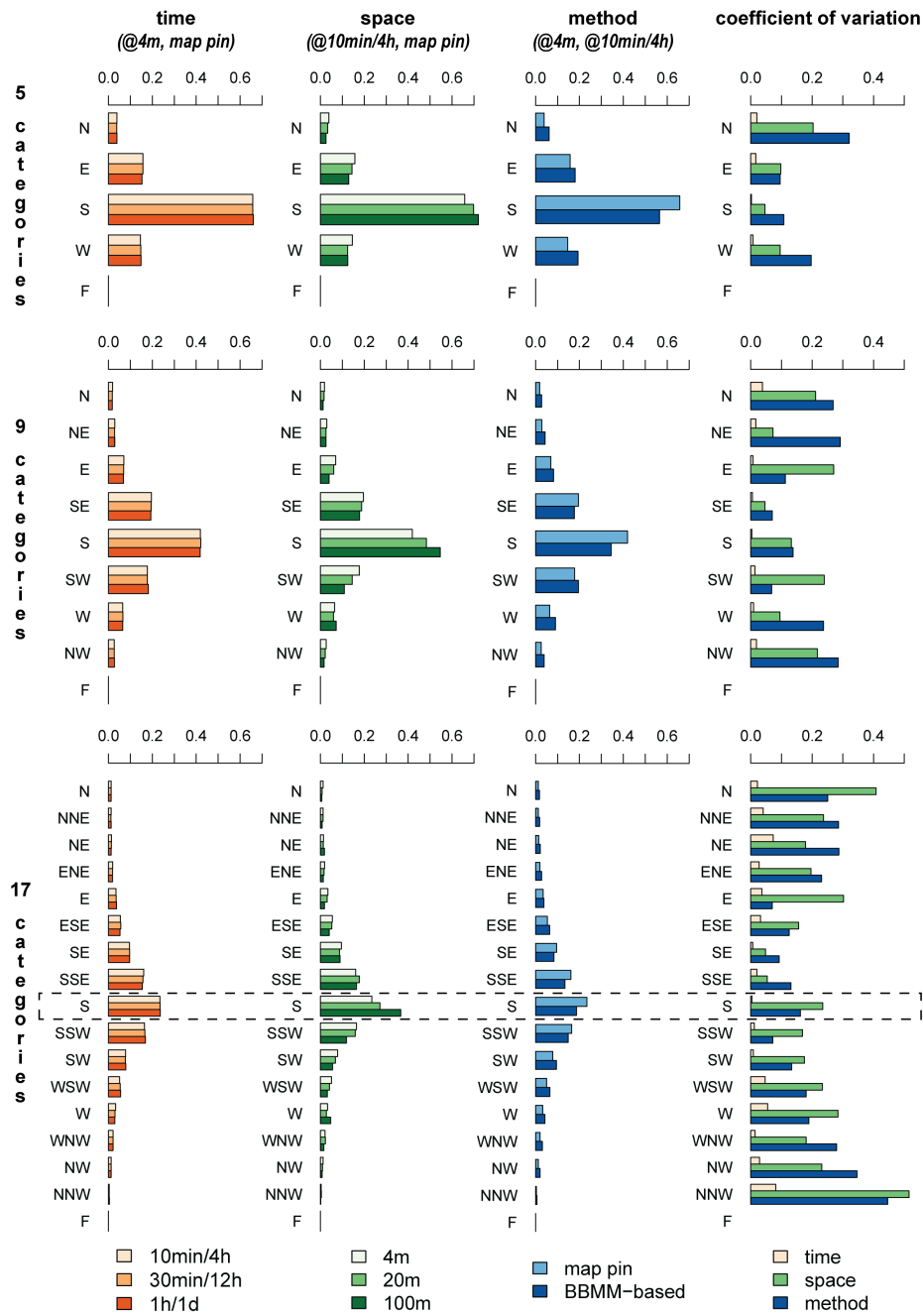
Figure 7.6 summarizes the main findings garnered by this case study. The effects of different factors (scales and relation methods) and their interplay with regard to relating movement and geographic context are illustrated in green in case there was found some remarkable impact, otherwise in red. If no notable effect was identified, as visualized in red, this would not mean that there are no scale effects and interdependencies at all, especially in the case of the temporal scale. This rather illustrates that, for the chosen movement and geographic context, as well as for the selected instances of scales and relation methods, no mentionable variation was identified.

At a first level of insight, our study illustrates that, for a given case study (in our case, ungulate movement related to terrain aspect), the movement-context relation may be sensitive to some scales (spatial scale, thematic scale), but not so much to others (temporal scale, Figure 7.6). The behavior of the species determines the

sensitivity of the results to the various scales; for example, ungulates do not have specific preferences with regard to the terrain's aspect on a daily basis, but seasonally notable differences in their preferences can be revealed.

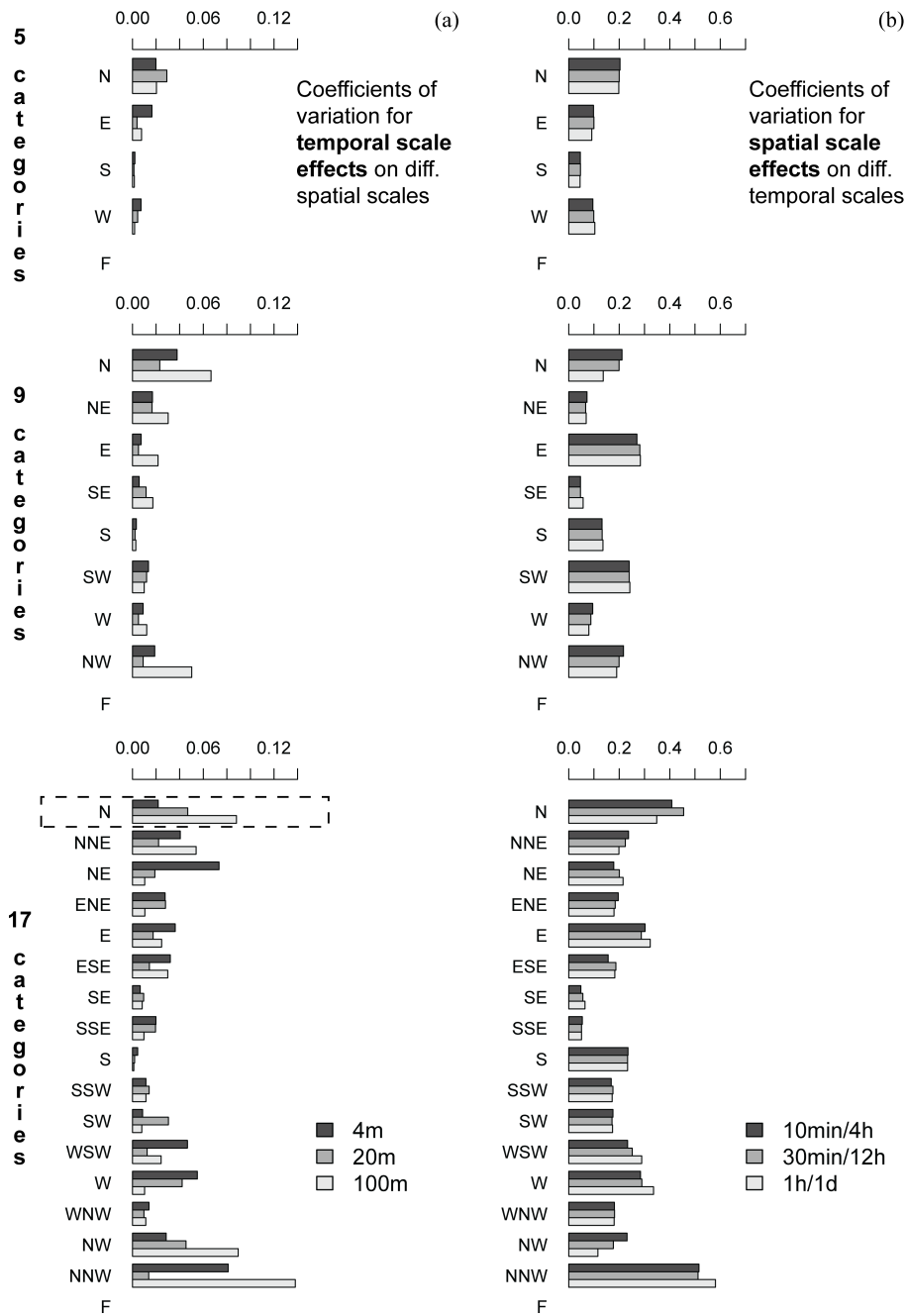
More importantly, on a second level of insight, our systematic cross-scale sensitivity study provided quantitative evidence for the existence of complex interdependencies between all involved scales (Figure 7.6). The sampling rate of the movement may matter more or less depending on the spatial and thematic scale of the context, and the chosen interrelation method. Similarly, the spatial granularity of the embedding context matters more or less depending on the sampling rate of the movement, and again the interrelation method. Similar interdependencies may be found for the influence of the used thematic scale and interrelation method. We are happy to acknowledge that, in most applied movement analysis studies, extensive cross-scale sensitivity analyses are not possible or indeed even necessary. We also understand that, in the data-driven field of movement analysis, the analysis scale is often ruled by cost or technology constraints, and hence beyond what the analyst can influence. However, we argue that it should be good practice in movement analysis to perform basic scale sensitivity tests. We see the key contribution of this work in raising the awareness for a problem, comparable to the awareness every good spatial data analyst has for the Modifiable Areal Unit Problem (MAUP).

In future work, other context variables could be considered in order to establish whether or not scale effects and scale interdependencies are specific to the geographic context variable. Further, in order to assess in greater detail the sensitivity of the distributions of the geographic context in relation to the movement with regard to the different scales and relation methods, more instances of temporal, spatial and thematic scales can be included.

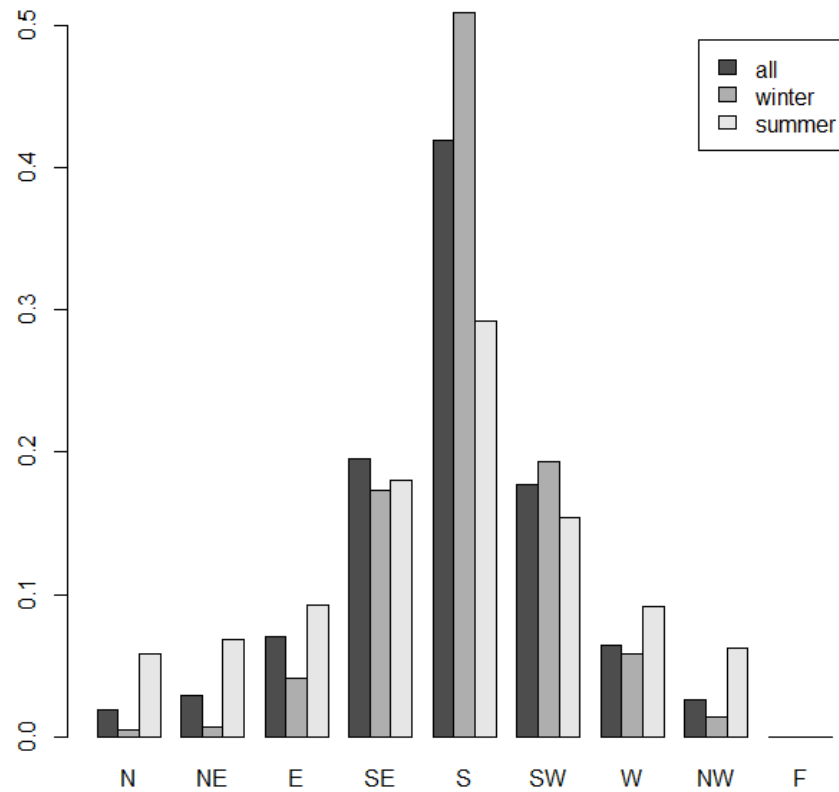


**Figure 7.3.:** Relative distributions (0-1) of the geographic context (terrain aspect) in relation to the chamois' movement. The sensitivity experiments systematically varied the temporal scale of the movement (“time” column), the spatial scale of the context (“space” column), and thematic scale of the context (y-axis), as well as the relation method (“method” column). The “coefficient of variation” (0-?) column presents the within-category variation. For example, the coefficient of variation for S, 17 categories (dashed box) is very low for “time”, much higher for “method”, and highest for “space”.

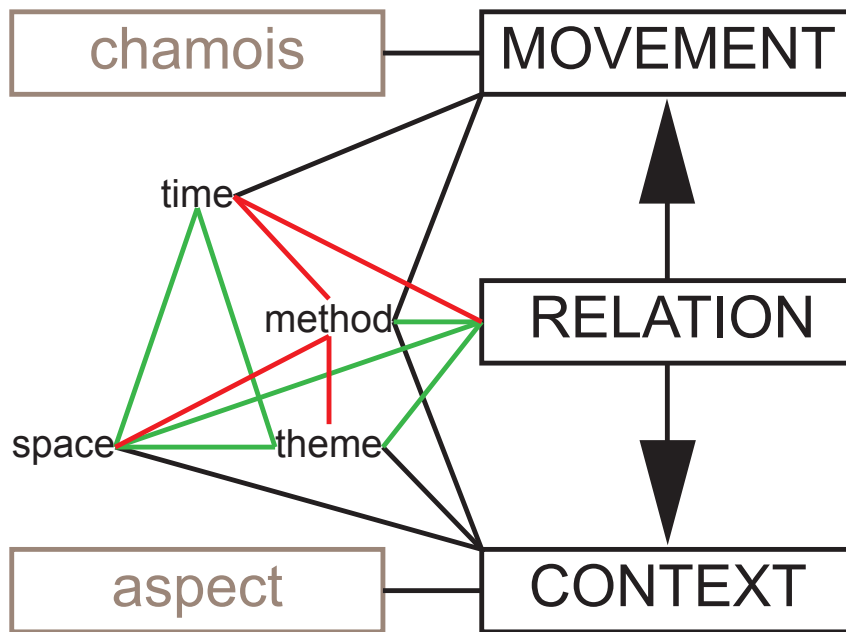
#### 7.4. Main findings and contribution



**Figure 7.4.:** Coefficients of variation representing within-class variation due to variation of temporal (a) and spatial (b) scale across different thematic scales. (a) Variation due to temporal scale effects depending on different spatial scales of geographic context. (b) Variation due to spatial scale effects depending on different temporal scales of geographic context. For example, in column (a), N, 17 categories (dashed box), the variation resulting from varying the temporal scale (10min/4h, 30min/12h, 1h/1d) is much smaller at a spatial scale of 4 meters than it is on 100 meters.



**Figure 7.5.:** Relative distribution of aspect related to the chamois' movement (with map pin) for whole year (all), winter and summer (@10min/4h, @4m, @9 categories).



*Figure 7.6.:* Overview of scale effects and interdependencies in relating movement to geographic context. Green: effects/interdependencies detected, red: no effects/interdependencies detected (author's own graph).





# Chapter 8.

## Discussion

In this chapter, the outcomes regarding the research questions of this thesis are reflected in the light of recent literature.

### 8.1. Effects of preprocessing on context-aware movement analysis

- *RQ1. What is the influence of commonly used preprocessing steps (e.g., segmentation, map matching) on context-aware movement analysis, and how can it be quantitatively evaluated?*

Besides background knowledge, data mining and knowledge construction, the KDD process includes preprocessing as a major stage to progress from data to knowledge (FAYYAD *et al.*, 1996; MILLER, 2008). Since the KDD process is also applied in the context of movement analysis, preprocessing is well established in this research area; however, although researchers within GIScience, movement ecology and urban planning acknowledge preprocessing as a crucial and fundamental step in the analysis of movement data, applying some kind of preprocessing in their studies, explanations of the detailed preprocessing remain sparse (e.g., in case of filtering outliers). The main reason for this issue might be the fact that results from data mining are much more interesting to present since these are more closely related to the construction of new knowledge within a certain application area. Within computer science, for example, procedures of preprocessing are even main topics of research (e.g., map matching); in the context of computer science, however, researchers have different research objectives. Consequently, computer scientists have introduced detailed methodologies, but do not aim at testing the impact of a preprocessing technique on the actual movement analysis. Knowledge construction should still be the major and final aim of studies with regard to movement analysis; nevertheless, preprocessing remains an integral step within movement analysis, and therefore should not be neglected. Exactly this issue describes where the contribution of this thesis lies in terms of preprocessing of movement data. In Section 3.4, the effects of preprocessing techniques, such as filtering stops (segmentation of stops and moves) and map matching, were investigated in regard to (context-aware) movement analysis.

In the first experiment in Section 3.4, we confirmed the findings of LAUBE & PURVES (2011), who stated that removing stops is an important issue in movement

analysis for deriving speed. Further, the detection of stops, as presented in this thesis, supports distinguishing “real” movement at low speeds from pseudo-movement. Further, in Section 3.4, it was illustrated that map matching can have a crucial impact on the computation of speed owing to the fact that the geometry of the trajectory is significantly changed – particularly at crossings. Since roads often meet at a 90-degree angle, the change of geometry for the trajectory is even larger at crossings. Map matching allows for assigning movement to the properties of network edges (geographic context), but simultaneously, even with more sophisticated map matching algorithms, geometry and therewith physics, inherently being part of movement data, are manipulated to a considerable extent. For linking movement to properties of streets in network space, map matching is a suitable means; however, as soon as the geometry or the physics of movement, respectively, is at the focus of investigation, the original trajectory geometry should be considered.

In a further experiment (Section 3.4), speed was related to habitat type, firstly for the unfiltered movement data, and secondly for the movement data, where stops were removed. Removing stops revealed different changes in the speed signal, depending on the habitat type. Therefore, preprocessing takes on a particular meaning for CAMA. This experiment further illustrates that the computation of movement parameters, based on the trajectory, such as speed, does not make sense on a too-coarse temporal granularity, which adheres to several moves and stops in-between two consecutive GPS points. Moreover, this outcome also links to Research Question 4 (RQ4). In connection with Research Question 4 (RQ4), however, scale effects with regard to CAMA (Chapter 7 and Section 8.4) were addressed rather than the computation of movement parameters.

Preprocessing of movement data, as well as the preprocessing of context data, requires significant effort in terms of time. This thesis mainly focused on the preprocessing of movement data since some research on the preprocessing of static (spatial) context data has already been done across different fields. In our opinion, the preprocessing of movement data should be better incorporated into the methodological procedure, besides the data mining, and the effects of preprocessing on results of data mining and CAMA should be taken into account and viewed through a more careful lens. As a consequence, a threshold-based preprocessing workflow was developed (Section 3.2) that is easy to use, but which still allows for the detection of stops at different temporal scales based on mean speed and turning angle (extended from mean distance in LAUBE & PURVES, 2011). There are many other threshold-based methodologies concerned with the identification of stops and moves (e.g., BUCHIN *et al.*, 2013); however, such approaches most often do not deliver stops at different temporal scales. As the case studies in connection with this workflow have illustrated, short stops at crossings can even be distinguished from longer stays, and the knowledge relating to the temporal context (duration of stop) helps to better understand possible reasons for the stop, simply by putting the stops (and moves) on a map.

A weakness of the approach lies in the arbitrariness of setting the different thresholds. Accordingly, some background knowledge is needed in order to derive stops on different temporal scales. However, as shown in the experiment with the household data in

Section 3.3, selecting the “correct” thresholds can be rather intuitive, and the outcome still can be unexpected in terms of locations, where people stop or reveal new insights into the process that triggered the person to stop. However, this effect is taken into account by testing the sensitivity of the approach to different threshold values. ORELLANA *et al.* (2010) present an exploratory statistical approach concerned with identifying movement patterns related to stops, where no a-priori knowledge about the moving entity and the surrounding context, and no thresholds, are required. These authors even showed that their approach works for different moving objects with different temporal resolutions of stop durations. Nevertheless, stops of different magnitudes of duration for the same moving entity cannot be distinguished easily, which we would argue is the main strength of our preprocessing workflow.

Owing to the fact that often we cannot ask people about their purposes and reasons for stopping at certain locations and moving on at others, such a preprocessing workflow and the relation of movement to its surrounding environment (Chapter 6) is a step towards a better understanding of people’s intention of movement.

## 8.2. Modeling movement and its embedding geographic context

- *RQ2. How should movement and the geographic context embedding movement be captured and modeled allowing for a quantitative relation between the two?*

First of all, it is important to emphasize that modeling spatio-temporal phenomena, such as movement and geographic context, involves several steps. Conceptual models (object vs. field for spatial phenomena) are represented with the use of certain data structures (e.g., raster can be a realization of a field), and methods allow for the quantification of specific spatio-temporal phenomena; for instance, a field-based representation of movement can be realized with the BBMM (statistical method by HORNE *et al.*, 2007), which is applied on a raster data structure. Equally, a method introduced by HORN (1981) can be applied on a raster data structure, in order to represent hill slope as a specific characteristic of the geographic context in form of a field.

Chapter 4 suggests conceptual models and data structures with the purpose of representing movement, as well as geographic context. Concepts for modeling movement and geographic context are based on the fundamental primitives in GIScience for the representation of geographic phenomena: objects and fields (e.g., GOODCHILD, 1992). When striving to capture characteristics of the geographic context, the classic data structures are used (point, line, polygon, and raster); similarly, data structures were presented for representing movement – in particular, incorporating the temporal dimension of movement, including vector (directed point), move (according to TURCHIN, 1998), segment (subtrajectory, trajectory), polygon (e.g., MCP) and raster (Chapter 4). Raster representations of movement most often implicitly take into consideration the time dimension of movement, based on some theory (e.g., time geography) or

statistical model (e.g., BBMM: continuous-time stochastic process, HORNE *et al.*, 2007) that is used.

A review of the literature from movement ecology and GIScience (Chapter 4) showed the tendency of separate preferences between these research fields with regard to conceptual models and data structures for the representation of movement. Movement ecologists rather prefer areal aggregation movement models as polygons (object, e.g., MCP, POWELL, 2000, or characteristic hull, DOWNS & HORNER, 2009) and rasters (field, e.g., BBMM, HORNE *et al.*, 2007, dynamic BBMM, KRANSTAUBER *et al.*, 2012, or bivariate Gaussian bridges, KRANSTAUBER *et al.*, 2014) in order to access the resource preferences of different animals. In GIScience often vectors, moves or segments (trajectories) are favored (e.g., DODGE *et al.*, 2008), which are mainly object-based representations of movement. It is clear that the purpose of a study determines the movement models deemed suitable for answering specific research questions; nevertheless, the proposed data structures for representing movement should allow for a more structured assessment of the possibilities at hand, providing researchers from different fields with more guidance in terms of more carefully choosing a suitable concept for modeling movement.

From a technical point of view, vector (corresponds to a point geometry) is arguably the most intuitive way of representing movement, when movement is captured in form of GPS data as time-stamped points, in the first place. Moreover, moves and segments can be seen as direct extensions of vectors; in addition, however, the temporal sequence is taken into account by connecting consecutive GPS points with straight lines, where segments usually even comprehend a semantic homogeneity in terms of movement parameters (e.g., speed, DODGE, 2011; BUCHIN *et al.*, 2011, 2013). Finally, as GIScience mostly relates to technical disciplines as computer science, this might go some way to explaining the popularity of the concepts of vectors, moves and segments. In regard to the conceptual models (object-field dichotomy) and data structures (vector vs. raster) for representing spatial phenomena as geographic context, much related work has already been carried out (e.g., GOODCHILD, 1992). Therefore, this issue with reference to geographic context, is not further discussed within this thesis.

Furthermore, with regard to Research Question 2 (*RQ2*), this work contributes more so to a structured overview of conceptual models and data structures for movement and geographic context in the form of a matrix than to a presentation of new conceptual models and data structures. Therefore, we propose a matrix as a structuring element with movement models on the x-axis and models for representing context on the y-axis in order to support a systematic access of different methods for relating movement to geographic context. Along this line of argumentation, methods allowing for a quantitative relation between movement and geographic context are captured in Section 8.3 in connection with Research Question 3 (*RQ3*) in more detail. In terms of the structured overview in the form of the matrix, Research Question 2 (*RQ2*) addressed the axis of the matrix building the fundament for the systematic assessment of movement-context relation methods accessed with Research Question 3 (*RQ3*). Nevertheless, the results in reference to Research Question 3 (*RQ3*) support quantitative evidence for using the proposed axes of the movement-context relation

matrix as a means of facilitating a broader, conceptual access to methods for relating movement and geographic context.

In order to understand the purpose of this movement-context relation matrix (*mc*-relation model) – and, therewith, its limitations – it is important to note two aspects (explicit validation of the matrix with RQ3, Section 8.3); firstly, this matrix is not supposed to be an exhaustive framework in the sense of a formal ontology; and secondly, emphasis should again be directed towards the fact that data structures (e.g., raster) should clearly be distinguished from quantitative methods (e.g., BBMM, HORNE *et al.*, 2007) – with importance also recognized in the fact that it is even more crucial not to confuse conceptual movement models (on the x-axis of the matrix) with movement spaces (according to LAUBE, 2009, e.g., space-time cube from time geography). The same data structure (e.g., segment) can be applied in different movement spaces (e.g., Euclidean space vs. space-time cube). To our understanding, the space-time cube is one way of representing the data structure of segments, what however, requires other methods for analyzing the movement, compared to the case of segments represented only in the spatial dimensions of Euclidean space. In other words, in different movement spaces, the same data structure can have different meanings, and requires different methods for establishing a link between movement (e.g., modeled as segment) and geographic context.

The experiments in Chapter 6 were conducted with the objective to demonstrate two examples of quantitatively modeling geographic context or, more precisely, specific properties of the surrounding environment, and to assess the effect of different quantitative models on the relation of people’s movement to geographic context. In the first case study in Chapter 6, efforts were centered on relating speed of bikers and hikers to the terrain’s slope. This example showed that not only the conceptual model and the data structure, but also the way in which parameters, such as slope, are modeled, can be of major importance in terms of finding correlations between movement and geographic context. For example, when slope was modeled as the decline of the beeline, much better correlations to speed were revealed than for slope derived from a DEM ( $R^2$  around 0.14 for biking and 0.07 for hiking instead of values close to 0.0). Our case study particularly illustrates that we should be careful in concluding that slope has no effect on the speed of bikers and hikers based on correlation analysis because, in this case, it obviously would be wrong, as the results of correlation analyses of single selected downhill and uphill tracks have shown. On a general level, on the other hand, correlation does not mean causation; therefore, even if we would identify a correlation between two variables as slope and speed, we cannot imply that slope has an effect on speed. It is rather an indication for a cause-effect relationship (also relates to RQ3, Section 8.3). In other words, finding no correlation might be seen as a hint that slope has no effect on speed, but in our case study, other reasons for the low correlations were an inappropriate modeling of the reality (in this case slope), and particularly the sensitivity of the movement’s underlying slope in steep terrain to positional errors in the GPS data. In only few studies, more than one quantitative method is considered for capturing a specific characteristic of the surrounding environment; for example, SAFI *et al.* (2013) investigated the impact of

different aspects of wind (wind speed, windsupport, crosswind, airspeed) on birds' flight speed.

The purpose of the Foursquare example (second case study in Chapter 6) was to show the potential of UGC as a source of geographic context in the analysis of movement data. The manual evaluation, based on some examples using the information of a cadastre map, emphasized the limitations and opportunities of modeling context with UGC. In many aspects, the use of UGC seems to lead towards gaining insights into people's stay in a city; for instance, the shopping area in Rotterdam as a motivation for people to visit the city core was a factor successfully identified. Also, small spots identified as recreation areas, such as in the form of a park, are rather accurately identified.

Importantly, however, this Foursquare study could be improved in some aspects. Firstly, within the modeling of movement with classic KDE, the implicit integration of the temporal component of movement (e.g., with BBMM) was not realized. Moreover, the BBMM approach would need a slightly different methodology since this movement model considers the temporal sequence of the GPS fixes, and therewith, a spatio-temporal probability density surface including all the movement data is not straight forward. Accordingly, adopting BBMM as opposed to KDE would not simply mean a replacement of KDE with BBMM. Secondly, the typically uneven distribution of the number of contributions per user (most of the data is generated by very few users: long-tail distribution) in UGC was not considered in our investigation. Regardless, however, this investigation allows a semantic enrichment of the movement data towards a better understanding of possible reasons about the stay in certain areas of a city.

KRÜGER *et al.* (2014) demonstrated that Foursquare data can be used for the contextual enrichment within the visual analysis of movement behavior. In their study, they support the user with a highly interactive tool in the visual investigation of POIs defined through a density-based clustering of the movement data in terms of foursquare locations (e.g., "shops/services"). In our case study, we integrated Foursquare data in the actual analysis of movement data in an effort to derive functional regions of Rotterdam, representing the most probable reason for the stay of people in this regions of the city. The second case study illustrated that the use of UGC for capturing semantics of the people's surrounding environment in urban areas, has great potential in CAMA.

### 8.3. Methods for relating movement to geographic context

- **RQ3.** *How can movement and explicitly its spatio-temporality (not only its mere spatial footprint) and the geographic context embedding this movement be quantitatively interrelated?*

This research question (*RQ3*) was addressed with the movement-context relation matrix. This matrix is intended to mirror different possibilities of relating movement to geographic context on the basis of the concepts for representing movement and geographic context in connection with Research Question 2 (*RQ2*, Section 8.2). In the

### 8.3. Methods for relating movement to geographic context

Sections 4.2 - 4.5, some specific movement-context relation methods were presented for geographic context represented as field. Further, Chapter 5 includes a case study, where some of these relation methods have been tested on movement data of ungulates. The experiments with the movement data of people and its relation to geographic context, with focus on modeling geographic context in Chapter 6, were already discussed in Section 8.2.

In the case study in Chapter 5, a simple movement analysis task related to home range estimation was applied on the movement data of ungulates. Other more complex movement analyses often have different purposes than merely assessing context factors' distribution based on their relation to the actual movement. The objectives of a study also define the movement and context models and therewith, the movement-context relations suitable for a specific CAMA. As such, often, only parts of the range of available movement-context relation methods, as presented with the *mc*-relation matrix, are considered appropriate for particular CAMA tasks. However, even if only some of the movement-context relation methods can be used for certain movement analysis tasks, the proposed matrix structure still supports analysts in thinking about different approaches for assessing CAMA tasks. Therefore, the proposed *mc*-relation matrix also allows for a systematic sensitivity analysis with regard to movement-context relation methods, as it is already done for other factors as for instance scale (Chapter 7).

ANDRIENKO *et al.* (2011b) presented an event-based conceptual model for CAMA that allows for analyzing relations of moving objects to spatial locations, other moving objects and events. In terms of this event-based conceptual model, our structured overview mainly focused only on the relations of moving objects to its surrounding geographic context (spatial locations). Our movement-context relation matrix instead specifically assesses the relations of moving objects to almost static geographic context (more or less stable over time, e.g., topography), with a whole range of specific methods for this kind of relation presented in more detail.

As a consequence of the finding that different disciplines with different aims for studying movement tend to use different movement models (discussed in Section 8.2), different methods for relating movement to geographic context are applied. In GIScience, point-based (map pin approach) or buffer overlay (e.g., IMFELD, 2000) are prominent techniques for relating movement to geographic context. On the other hand, within movement ecology in the field of home range estimation, the technical focus on the movement side lies on areal geometries (polygon, raster) for linking movement to its surrounding environment. Therefore, in this context, mainly MCP or KDE (e.g., WARTMANN *et al.*, 2010) are applied in order to assess the space use of animals with regard to external factors. Therefore, in this thesis, we attempt to link movement and geographic context on a whole range of different relation methodologies.

In the case study in Chapter 5, different methods for relating movement to geographic context with regard to the estimation of space use in terms of environmental factors were compared. This experiment (Chapter 5) showed that the variation due to different *mc*-relation methods is considerable – and even in the order of magnitude of the well acknowledged scale effects in geographic analyses. The different relation

methods with regard to the movement data of ungulates showed remarkable differences in the actual movement-dependent distributions of the context factors of up to almost 0.4 (in relative frequencies), in the example of segment buffer vs. MCP for movement of chamois and the context factor represented by landform classes by IWAHASHI & PIKE (2007). Thus, the *mc*-relation matrix seems to be reasonable with regard to relation methods, since notable differences due to relation methods are mirrored on the basis of the given matrix structure. Therefore, we claim that researchers across different fields within CAMA should carefully consider different methods for quantitatively linking the movement and geographic factors of the surrounding environment. Moreover, we have practically demonstrated that our *mc*-relation model suggests a reasonable means for supporting a more exhaustive basis for choosing an appropriate *mc*-relation method in the example of a task related to home range estimation. In most of the studies reviewed in the state of the art (Chapter 2), only one single method for relating movement and parameters of the surrounding environment was applied. More importantly, a detailed discussion of the used *mc*-relation methods within other work most often is lacking. In other words, most often there is no reasoning about the chosen *mc*-relation model.

However, IMFELD (2000) proposed RDFs as an alternative to the simple map pin approach, but still based on the conceptual movement model of vector (point-based). RDFs allow for capturing the surrounding environment on different spatial scales. The concept of RDFs is not only restricted to vectors, and could also be applied on moves or segments. From a GIScience perspective, in this thesis, with the suggested *mc*-relation matrix and the specific *mc*-relation methods for a whole range of conceptual movement models, we sought to reach beyond a point-based interpretation of movement with concepts as move, segment, polygon (e.g., MCP) and raster (e.g., BBMM).

Concerning the relation of movement to geographic context in general, DODGE *et al.* (2012) developed an automated procedure allowing for enriching animal tracking data from the open MoveBank portal with environmental variables. In comparison to this annotation tool, our *mc*-relation model supports a more exhaustive range of possibilities for relating movement to environmental factors, but at the same time is more conceptual and less directly applicable as the annotation tool. The annotation tool on MoveBank is highly applicable for efficiently putting animals' movements into context (mostly variables of the geographic environment). From a conceptual perspective with regard to our *mc*-relation model, this annotation service only incorporates a small subset of possible *mc*-relation methods. Since most of the contextual information provided by the annotation system are dynamic (spatio-temporally highly fluctuating) variables (e.g., temperature or wind speed), spatial and temporal interpolation builds the core of the annotation procedure, but various concepts of methods for relating movement to the actual context variable are not considered. Further, most of the environmental factors are supposed to be modeled as rasters, which is not necessarily the optimal solution for other context factors of the geographic environment (e.g., land cover). The annotation service might improve, considering a more holistic approach in terms of conceptual models for movement and geographic context, and *mc*-relation methods with regard to our *mc*-relation model.



## 8.4. Scale effects in context-aware movement analysis

- *RQ4.* How sensitive is the computation of a quantitative relation between movement and its embedding context to a systematic variation of the temporal, spatial and thematic scale of analysis, and can interdependencies between the different scale dimensions be identified and quantified?

The empirical study in Chapter 7, during which we related movement to geographic context with various sorts of analysis scales (temporal scale of movement, spatial and thematic scale of geographic context), showed that CAMA is sensitive to the spatial (up to around 10% difference in relative frequencies) and thematic scale of the geographic context. In this case, varying the temporal scale of movement had no effect on the results of the CAMA, since the variation of the temporal scale was not significant enough (less than 1% in relative frequencies) in order to reflect changes in the animals' habitat use in terms of the investigated parameter of the geographic environment (here: aspect). This does not mean that temporal scale has no impact on movement-context relation. In the case of aspect, on a seasonal temporal scale, notable variances in space use are expected. The quantitative assessment of the sensitivity of CAMA to different analysis scales and *mc*-relation methods with the coefficient of variation facilitated the cross-comparison of effects of different analysis scales – and even the comparison of the scale effects with the impact of different *mc*-relation methods on CAMA.

The investigation of scale effects was performed with regard to a clearly defined task of relating movement to geographic context, rather than introducing more complexity with a specific algorithm for CAMA. A simple and not-too-specific movement analysis task was applied, and we did not consider more complex data mining algorithms from CAMA for the following reasons, similar to the approach that was chosen to investigate effects due to different relation methods (Chapter 5); firstly, an approach primarily incorporating the movement-context relation was selected; secondly, this task is relevant for several research areas, mostly referring to research from home range estimation and resource selection in movement ecology, but also to work from GIScience in annotating movement data with information relating to the surrounding geographic environment; and thirdly, with the presented approach, sensitivities caused by scales are still comparable to effects of different relation methods using this task.

Our case study in connection with scale effects (Chapter 7), interdependencies between the different scales of analysis were revealed. Although different *mc*-relation methods caused a remarkable variation in the results of linking movement and geographic context, scale effects were shown to be independent of the relation method. With different relation methods, similar scale effects were discovered.

BÖRGER *et al.* (2006) did not reveal any impact of the thematic scale (classification scale) on the analysis of home ranges. In contrast to their study, we identified various effects of thematic scale on relating movement to its embedding geographic context. Moreover, BÖRGER *et al.* (2006) illustrated variation of home range sizes for different temporal scales, linking these variances to external sources. This finding underpins

the assumption that, in our case study, larger variation of the temporal scale (e.g., seasonal scale) would have an impact on linking movement to geographic context as well. Whether sensitivity to scale is significant or not depends on many factors, including how strongly scale is varied, the movement analysis task, and the type of moving object and geographic context under examination. In comparison to other studies in movement analysis with regard to different types of scale, besides the effects caused by the variation of a single sort of scale (e.g., temporal scale in BÖRGER *et al.*, 2006; LAUBE & PURVES, 2011), our study revealed a complex interplay between the temporal, spatial and thematic scales.

The interdependencies between the different types of scale were quantified by computing the within-context-class variation using the coefficient of variation. The within-context-class variations caused by different kinds of scales were mirrored in the coefficient of variation values successfully. Therefore, the coefficient of variation showed to be an adequate means, allowing for the quantitative assessment of the interdependencies between the different sorts of scales.

# Chapter 9.

## Conclusions

Researchers from different fields, such as GIScience, behavioral ecology and urban mobility, for example, study different aspects of animals' or people's movement in an effort to learn more about their behavior. KDD offers a procedure to get from data to knowledge – or, in the context of movement analysis, to get from movement data to movement behavior. With regard to the KDD procedure, GIScience mainly has its expertise in preprocessing and data mining, and the application areas rather contribute to bridging the semantic gap in terms of constructing knowledge based on quantitative data mining results (e.g., movement patterns). In most of the research conducted in movement analysis, solely the geometric characteristics of the movement trajectory are analyzed in an effort to achieve a better understanding of movement behavior. Although there is a volume of research that integrating information about the context, wherein the actual movement is embedded, concepts and methods that allow for investigating movement in the context of its surrounding environment are not yet well-developed. In this thesis, we focused mainly on conceptual and methodological contributions to CAMA, referring to preprocessing, methods for relating movement and geographic context, and scale effects.

### 9.1. Achievements

**Preprocessing and its effects.** Research in this methodological project carefully investigated effects of different preprocessing techniques, preparing movement and context data for subsequent quantitative interrelation (RQ1). Specifically, the project produced several filtering and segmentation procedures, starting with crucial separation of stops and moves (e.g., for animals moving through their habitat). Furthermore, map matching and its impact on the computation of movement parameters, as a consequence of changed trajectory geometry, were analyzed (e.g., for pedestrians moving through the street network). The proposed preprocessing methods were finally integrated into a standardized preprocessing workflow allowing for the segmentation of raw trajectories, considering movement properties at different granularities. With the presented preprocessing methods and experiments, we confirmed that preprocessing is a crucial step in applying the KDD process with GIScience approaches in an effort to resolve questions of behavioral ecology and urban mobility on the basis of movement data and its relation to geographic context. Moreover, the experiments showed that preprocessing can have a significant impact on further CAMA steps.

**Models for movement and geographic context.** This thesis then introduced a conceptual framework for structuring different movement-context relation methods in matrix form according to models of movement (x-axis) and geographic context (y-axis) (RQ2). The axes of the *mc*-relation model built up the matrix, allowing for a systematic approach of methods allowing movement to be related to geographic context (refers to RQ3, see following paragraph). A classification into five conceptual models for the representation of movement was proposed, allowing for the assessment of different approaches for relating movement and geographic context. These concepts for modeling movement include vector, move, segment, polygon and raster, building on the classic concepts for representing purely spatial phenomena, but additionally also considering time on a conceptual level (e.g., vector as point geometry with velocity). Conceptual models for geographic context are realized with entities and field conceptual data models. In the course of comparing different movement-context relation methods with the use of the proposed matrix structure, the entire range of this movement model classification was validated. The suggested movement model classification makes sense as it allows identifying differences in resulting relations of movement and geographic context. The systematic approach of the relation framework not only identified gaps in the analytical toolbox, but also crucially helped defining the requirements and prerequisites for the missing methods (RQ2).

Moreover, different ways of the quantitative assessment of parameters capturing characteristics of the geographic environment were presented for the case of animal movement, as well as for people moving on networks of urban areas. The case study with the animal movement data, combined with geographic context, modeled with an approach from geomorphometry, illustrated that quantitative methods for modeling the context remarkably influence the results of CAMA. When assessing the geographic context of urban areas, UGC was found to be suitable in the context of CAMA and urban mobility; notably, however, the usefulness of UGC depends on the research goals of the actual study, and therefore might not be appropriate within other application domains.

**Methods for relating movement and geographic context.** The proposed matrix structure (refers to RQ2, see previous paragraph) with regard to the relation of movement and geographic context was validated by applying a whole range of movement-context relation methods on real movement data (RQ3). Therefore, a wide range of relation methods were developed, adapted and tested. A first focus was placed on relating vector-, move- and segment-based movement to field context, including a range of buffer-related concepts and RDFs. The second focus turned to relating field movement models, based on kernel density and Brownian bridge concepts, to again field context (e.g., terrain or land cover). Relation methods were put to the test and validated with a set of case studies from animal movement ecology, urban planning and nature reserve management. The coefficient of variation allowed for assessing the methodological differences mirrored in the resulting distributions of geographic context after relating movement and geographic context. This approach of revealing

methodological effects within linking movement and geographic context enabled us to discover significant differences due to different movement-context relation methods.

**Scale effects.** This thesis suggested an approach, similar to the methodology used with regard to methodological effects (RQ3) for quantitatively assessing scale effects caused by different sorts of analysis scales in CAMA. The sensitivity of relating movement to geographic context to varying temporal scales of movement as well as spatial and thematic scales of geographic context was investigated. Moreover, a methodology for the quantitative assessment of interdependencies between these different analysis scales in CAMA was proposed and applied. The scale sensitivity of relating movement to geographic context was compared to the effects of different *mc*-relation methods in CAMA (RQ3). Accordingly, the impact of different *mc*-relation methods was studied against the background of the well acknowledged scale effects inherent to geographic investigations. Again, the coefficient of variation allowed for identifying the scale effects of different types of analysis scales and interdependencies between them in the procedure of relating movement and geographic context. Further, the methodological effects (RQ3) was shown to be in the same order of magnitude as the effects caused by spatial and thematic scale of the geographic context.

## 9.2. Insights

The key contribution of this project lies in substantiating quantifiable effects of method design, choice, and parameterization when relating movement to its embedding geographic context - effects that were mostly neglected so far in the literature.

This thesis aimed at investigating the procedure of relating movement to its embedding geographic context in order to facilitate a better understanding of external factors influencing the actual movement and to enable gaining more detailed insights about movement as behavior. This research did not explicitly show that the integration of geographic context leads to better results with the KDD process (data mining), but rather focused on the effects of particular factors, which are part of the KDD process explicitly (preprocessing), as well as implicitly (models for movement and geographic context, movement-context relation methods, scale), and which have been mostly neglected in some of its aspects so far. In other words, the focus of this work was directed towards the parts of the KDD process that, thus far, have not been considered in movement analysis as prominently as, for example, data mining.

This research has shown that different aspects play an important role in relating movement to geographic context; this work confirmed that CAMA is very complex with regard to various interdependent factors, along with the well-researched techniques of data mining. This thesis provided quantitative evidence that for a better understanding of movement processes, it is at least equally important to carefully think about preprocessing, appropriate representations and quantitative methods for modeling movement and geographic context, suitable movement-context relation methods, and

scale effects in CAMA, than it is to conceptualize, formalize and detect meaningful movement patterns within later data mining phases.

Relating animal movement (and its properties) to land cover and terrain, using different relation methods revealed manifold, and not easy to predict dependencies between problem scenario, data granularity, chosen relation method and method parameterizations. Similarly, relating the movement of outdoor sportsmen and urban shoppers to their movement context also revealed that in some constellations, the choice of relation method and the method for deriving context variables has significant impact on found relation strengths. For instance, correlation analysis was found to be highly sensitive to different quantitative methods for modeling slope. This study illustrated the substantiated difficulty to distinguish between what is really a new insight in terms of movement behavior, and what appears to be an artifact of the methodology. However, an adequate modeling of the geographic context facilitated an improved result in terms of a better understanding of the movement-context relation.

The same problem in terms of sensitivity applies to preprocessing, movement-context relation methods, and different types of scale and their interdependencies. Preprocessing was confirmed as a crucial step of the KDD process for detecting inaccurate GPS measurements. Further, the various experiments on quantitative effects of preprocessing procedures produced a complex picture. This thesis illustrated that the preprocessing of movement data can cause expected (e.g., low speed values at stops), but also unexpected effects (e.g., unrealistic speed values due to map matching) on different movement analysis tasks; therefore, a standardized preprocessing workflow – such as that presented in this work, for example – would allow a better comparison of movement analysis results since these results would not be masked by different preprocessing procedures. However, on a more specific level, there can be no generic guidelines, instead, besides the presented preprocessing workflow, individual considerations on a case-to-case basis are required.

This thesis showed that it is necessary to consider different movement models and relation methods for establishing a feasible quantification of the movement-context relation. Thus, the movement-context relation matrix is a reasonable means of assessing different types of model for representing movement, and different types of movement-context relation method. Moreover, in the procedure of relating movement to geographic context, different forms of scale play an important role. This thesis went a step further and demonstrated that the meaning of different scales cannot be comprehended by considering the temporal scale of the movement and the spatial and thematic scales of the geographic context separated from each other. Instead, when relating movement and geographic context, scale effects go beyond an isolated view on the different kinds of scales, because of interdependencies between the scales.

### **9.3. Outlook and challenges of future research**

Relating movement to geographic context in movement analysis is a crucial aspect when striving to gain a better understanding of movement patterns in terms of

### 9.3. Outlook and challenges of future research

movement behavior, and generating insights about factors that have an influence on the actual movement of different individuals. Practically, we know a lot about different concepts for modeling static phenomena. Furthermore, the technical requirements for capturing context and movement data, as dynamic phenomena with regard to time, are available. However, we know less about concepts for modeling movement and about methods for relating movement and geographic context. The proposed matrix model as a structured overview of potential methods for relating movement and geographic context can be seen as a starting point for guiding the development of methods for a quantitative embedding of movement in its surrounding geographic context.

In this thesis, the movement-context relation matrix was validated with regard to geographic context represented in the form of a raster. The remaining rows of the matrix in terms of other representations of context (e.g., polygon representation of context) could be evaluated in future research. Although the movement-context relation matrix is also applicable for more dynamic spatial phenomena (e.g., wind) than for instance the topography, the matrix and the presented movement-context relation methods mainly have been conceptualized for static geographic context. Therefore, the movement-context relation matrix could be revisited to accommodate also geographic context that changes more quickly in time.

Further, more specific movement analysis tasks – for example, algorithms defining a concrete context-aware movement pattern (e.g., leadership) – could be applied to evaluate the proposed movement-context matrix in more detail. Moreover, new algorithms implicitly incorporating geographic context into the analysis of movement may be developed in a next step, or including geographic context into existing context-unaware algorithms. This would mean that the added value of integrating context information into movement analysis could be assessed by comparing classic algorithms, which define movement patterns on purely geometric properties of the movement path, to context extensions of basically the same algorithm.

The evaluation of the semantic meaning of detected movement patterns is a major challenge in movement analysis for different reasons. Movement data most often is not annotated with additional information about the moving object, as for example field-observed behavior or special conditions under which the movement took place. The annotation of movement data would enable a more detailed evaluation of resulting movement patterns. Although movement data is collected in huge amounts on a wide range of different temporal granularities, its annotation with some additional information about the moving object most often is sparse. Moreover, animals, for instance, cannot be interviewed, in order to get insights into their internal state, for example.

“Movement is behaviour, but patterns are not.” (ORELLANA *et al.*, 2010, p.67)

The above quote nicely mirrors the fact that the identification of movement patterns’ semantic meaning in terms of movement behavior is not straightforward. The relation

## *Chapter 9. Conclusions*

of movement to geographic context is an attempt at bringing movement patterns closer to an interpretation of movement data as movement behavior. Therefore, it may be valuable to slightly rephrase the quote by ORELLANA *et al.* (2010, p.67) as “movement is behavior, but patterns are not yet”. Bridging the semantic gap remains a challenge of future research in the area of movement analysis, and CAMA in particular.



# Bibliography

- AEBISCHER, N., ROBERTSON, P. & KENWARD, R. (1993). Compositional analysis of habitat use from animal radio-tracking data. *Ecology*, vol. 74(5): pp. 1313–1325.
- ALVARES, L., BOGORNY, V., KUIJPERS, B., DE MACEDO, J., MOELANS, B. & VAISMAN, A. (2007). A Model for Enriching Trajectories with Semantic Geographical Information. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, ACM, pp. 22–29. ACM-GIS 2007, November 7-9, 2007, Seattle, WA, USA.
- ANDERSSON, M., GUDMUNDSSON, J., LAUBE, P. & WOLLE, T. (2008). Reporting Leaders and Followers among Trajectories of Moving Point Objects. *Geoinformatica*, vol. 12(4): pp. 497–528.
- ANDRIENKO, G., ANDRIENKO, N., BAK, P., KEIM, D., KISILEVICH, S. & WROBEL, S. (2011a). A conceptual framework and taxonomy of techniques for analyzing movement. *Journal of Visual Languages & Computing*, vol. 22: pp. 213–232.
- ANDRIENKO, G., ANDRIENKO, N. & HEURICH, M. (2011b). An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, vol. iFirst: pp. 1–24.
- BAGLIONI, M., DE MACÊDO, J., RENSO, C., TRASARTI, R. & WACHOWICZ, M. (2009). Towards Semantic Interpretation of Movement Behavior. In *Advances in GIScience, Proceedings of 12th AGILE Conference*, (eds.) M. Sester, L. Bernhard & V. Paelke, Springer, Berlin, DE, Lecture Notes in Geoinformation and Cartography, pp. 271–288. AGILE 2009, June 2-5, 2009, Hannover, DE.
- BASILLE, M., CALENGE, C., MARBOUTIN, E., ANDERSEN, R. & GAILLARD, J. (2008). Assessing habitat selection using multivariate statistics: Some refinements of the ecological-niche factor analysis. *Ecological Modelling*, vol. 211(1-2): pp. 233–240.
- BECKER, C. & NICKLAS, D. (2004). Where do spatial context-models end and where do ontologies start? A proposal of a combined approach. In *Proceedings of the First International Workshop on Advanced Context Modelling, Reasoning and Management*, Citeseer, vol. 4. UbiComp 2004, September 7-10, 2004, Nottingham, UK.
- BENKERT, M., GUDMUNDSSON, J., HÜBNER, F. & WOLLE, T. (2008). Reporting flock patterns. *Computational Geometry*, vol. 41(3): pp. 111–125.

## Bibliography

- BERNSTEIN, D. & KORNHAUSER, A. (1998). An introduction to map matching for personal navigation assistants. Tech. rep., TIDE Center, New Jersey Institute of Technology, Newark, NJ, USA.
- BEYER, H. (2012). Geospatial Modelling Environment (Version 0.7.1.0). (Software). URL: <http://www.spatial ecology.com/gme>.
- BHATTACHARYA, T., KULIK, L. & BAILEY, J. (2012). Extracting significant places from mobile user GPS trajectories: a bearing change based approach. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, ACM, pp. 398–401. ACM SIGSPATIAL GIS 2012, November 6–9, 2012, Redondo Beach, CA, USA.
- BITTERLICH, W., SACK, J., SESTER, M. & WEIBEL, R. (2009). Summary Report—Representation, Analysis and Visualization of Moving Objects. In *Dagstuhl Seminar Proceedings 08451*. Dagstuhl Seminar 2008, November 2–7, 2008, Dagstuhl, DE.
- BLYTHE, P., MILLER, G. & TODD, P. (1996). Human simulation of adaptive behavior: Interactive studies of pursuit, evasion, courtship, fighting, and play. In *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, (eds.) P. Maes, S. Wilson & M. Mataric, pp. 13–22. 4th International Conference on Simulation of Adaptive Behavior, September 9–13, 1996, Cape Cod, MA, USA.
- BOGORNY, V., HEUSER, C. & ALVARES, L. (2010). A Conceptual Data Model for Trajectory Data Mining. In *Geographic Information Science*, (eds.) S. Fabrikant, T. Reichenbacher, M. van Kreeveld & C. Schlieder, Springer, Berlin, DE, *Lecture Notes in Computer Science*, vol. 6292, pp. 1–15. 6th International Conference, GIScience 2010, September 14–17, 2010, Zurich, CH.
- BÖRGER, L., FRANCONI, N., FERRETTI, F., MESCHI, F., DE MICHELE, G., GANTZ, A. & COULSON, T. (2006). An integrated approach to identify spatiotemporal and individual-level determinants of animal home range size. *The American Naturalist*, vol. 168(4): pp. 471–485.
- BUCHIN, M., DODGE, S. & SPECKMANN, B. (2012). Context-Aware Similarity of Trajectories. In *Geographic Information Science*, (eds.) N. Xiao, M.P. Kwan, M. Goodchild & S. Shekhar, Springer, Berlin, DE, *Lecture Notes in Computer Science*, vol. 7478, pp. 43–56. 7th International Conference, GIScience 2012, September 18–21, 2012, Columbus, OH, USA.
- BUCHIN, M., DRIEMEL, A., VAN KREVELD, M. & SACRISTÁN, V. (2011). Segmenting trajectories: A framework and algorithms using spatiotemporal criteria. *Journal of Spatial Information Science*, vol. 3: pp. 33–63.
- BUCHIN, M., KRUCKENBERG, H. & KÖLZSCH, A. (2013). Segmenting Trajectories by Movement States. In *Advances in Spatial Data Handling*, (eds.) S. Timpf & P. Laube, Springer, Berlin, DE, pp. 15–25.

- BURROUGH, P. & McDONNELL, R. (1998). *Principles of Geographical Information Systems*. Oxford University Press, Oxford, UK.
- CLARKE, K. (1995). *Analytical and Computer Cartography*. Prentice-Hall, Englewood Cliffs, NJ, 2nd edn.
- DEMŠAR, U., BUCHIN, K., CAGNACCI, F., SAFI, K., SPECKMANN, B., VAN DE WEGHE, N., WEISKOPF, D. & WEIBEL, R. (2015). Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology*, vol. 3(5).
- DEY, A. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing*, vol. 5(1): pp. 4–7.
- DODGE, S. (2011). *Exploring Movement Using Similarity Analysis*. Ph.D. thesis, GIS Unit, Department of Geography, University of Zurich.
- DODGE, S., BOHRER, G., BILDSTEIN, K., DAVIDSON, S., WEINZIERL, R., BECHARD, M., BARBER, D., KAYS, R., BRANDES, D., HAN, J. & WIKELSKI, M. (2014). Environmental drivers of variability in the movement ecology of turkey vultures (*Cathartes aura*) in North and South America. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 369(1643): p. 20130195.
- DODGE, S., BOHRER, G. & WEINZIERL, R. (2012). MoveBank Track Annotation Project: Linking Animal Movement Data with the Environment to Discover the Impact of Environmental Change in Animal migration. In *Proceedings of the Workshop on GIScience in the Big Data Age*, (eds.) K. Janowicz, C. Keßler, T. Kauppinen, D. Kolas & S. Scheider, pp. 35–41. Workshop on GIScience in the Big Data Age, September 18, 2012, Columbus, OH, USA.
- DODGE, S., BOHRER, G., WEINZIERL, R., DAVIDSON, S., KAYS, R., DOUGLAS, D., CRUZ, S., HAN, J., BRANDES, D. & WIKELSKI, M. (2013). The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data. *Movement Ecology*, vol. 1(3): pp. 1–14.
- DODGE, S., WEIBEL, R. & LAUTENSCHÜTZ, A. (2008). Towards a taxonomy of movement patterns. *Information Visualization*, vol. 7(3): pp. 240–252.
- DOWNS, J. (2010). Time-Geographic Density Estimation for Moving Point Objects. In *Geographic Information Science*, (eds.) S. Fabrikant, T. Reichenbacher, M. van Kreeveld & C. Schlieder, Springer, Berlin, DE, *Lecture Notes in Computer Science*, vol. 6292, pp. 16–26. 6th International Conference, GIScience 2010, September 14–17, 2010, Zurich, CH.
- DOWNS, J. & HORNER, M. (2009). A Characteristic-Hull Based Method for Home Range Estimation. *Transactions in GIS*, vol. 13(5–6): pp. 527–537.
- DOWNS, J. & HORNER, M. (2012). Analysing infrequently sampled animal tracking data by incorporating generalized movement trajectories with kernel density estimation. *Computers, Environment and Urban Systems*, vol. 36(4): pp. 302–310.

## Bibliography

- DOWNES, J., HORNER, M. & TUCKER, A. (2011). Time-geographic density estimation for home range analysis. *Annals of GIS*, vol. 17(3): pp. 163–171.
- EDUARDES, A. (2009). Geographical Perspectives on Location for Location Based Services. In *Proceedings of the 2nd International Workshop on Location and the Web*, ACM, New York, NY, USA, pp. 1–4. LocWeb 2009, April 4, 2009, Boston, MA, USA.
- EDUARDES, A. & PURVES, R. (2007). A Theoretical Grounding for Semantic Descriptions of Place. In *Web and Wireless Geographical Information Systems*, (eds.) J. Ware & G. Taylor, Springer, Berlin, DE, *Lecture Notes in Computer Science*, vol. 4857, pp. 91–106. Proceedings of the 7th International Symposium, W2GIS 2007, November 28-29, Cardiff, UK.
- EGENHOFER, M. & MARK, D. (1995). Naive Geography. In *Spatial Information Theory A Theoretical Basis for GIS*, (eds.) G. Goos, J. Hartmanis & J. Van Leeuwen, Springer, Berlin, DE, *Lecture Notes in Computer Science*, vol. 988, pp. 1–15. Proceedings of the International Conference COSIT'95, September 21-23, Semmering, Austria.
- EVANS, I. (1972). General geomorphometry, derivatives of altitude, and descriptive statistics. In *Spatial Analysis in Geomorphology*, (ed.) R. Chorley, Harper & Row, pp. 17–90.
- FAYYAD, U., PIATETSKY-SHAPIO, G. & SMYTH, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI magazine*, vol. 17(3): pp. 37–54.
- FISHER, P., WOOD, J. & CHENG, T. (2004). Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers*, vol. 29(1): pp. 106–128.
- FRITZ, H., SAID, S. & WEIMERSKIRCH, H. (2003). Scale-dependent hierarchical adjustments of movement patterns in a long-range foraging seabird. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 270(1520): pp. 1143–1148.
- FRYXELL, J.M., HAZELL, M., BÖRGER, L., DALZIEL, B.D., HAYDON, D.T., MORALES, J.M., MCINTOSH, T. & ROSATTE, R.C. (2008). Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proceedings of the National Academy of Sciences*, vol. 105(49): pp. 19114–19119.
- GALTON, A. (2005). Dynamic Collectives and Their Collective Dynamics. In *Spatial Information Theory*, (eds.) A. Cohn & D. Mark, Springer, Berlin, DE, *Lecture Notes in Computer Science*, vol. 3693, pp. 300–315. COSIT 2005, September 14-18, 2005, Ellicottville, NY, USA.
- GETZ, W. & WILMERS, C. (2004). A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography*, vol. 27(4): pp. 489–505.

- GIANNOTTI, F. & PEDRESCHI, D. (2008). Mobility, Data Mining and Privacy: A Vision of Convergence. In *Mobility, Data Mining and Privacy – Geographic Knowledge Discovery*, Springer, Berlin, DE, chap. 1, pp. 1–14.
- GOODCHILD, M. (1992). Geographical data modeling. *Computers & Geosciences*, vol. 18(4): pp. 401–408.
- GOODCHILD, M. (2007). Citizens as Sensors: The World of Volunteered Geography. *GeoJournal*, vol. 69(4): pp. 211–221.
- GUDMUNDSSON, J., LAUBE, P. & WOLLE, T. (2008). Movement Patterns in Spatio-Temporal Data. In *Encyclopedia of GIS*, (eds.) S. Shekar & H. Xiong, Springer, Berlin, DE, pp. 726–732.
- GUDMUNDSSON, J., LAUBE, P. & WOLLE, T. (2012). Computational Movement Analysis. In *Springer Handbook of Geographic Information*, (eds.) W. Kresse & D. Danko, Springer, Berlin, DE, pp. 423–438.
- HÄGERSTRAND, T. (1970). What about people in regional science? *Papers in Regional Science*, vol. 24(1): pp. 6–21.
- HEBBLEWHITE, M. & HAYDON, D. (2010). Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365(1550): pp. 2303–2312.
- HORN, B. (1981). Hill Shading and the Reflectance Map. *Proceedings of the IEEE*, vol. 69(1): pp. 14–47.
- HORNE, J.S., GARTON, E.O., KRONE, S.M. & LEWIS, J.S. (2007). Analyzing animal movements using Brownian bridges. *Ecology*, vol. 88(9): pp. 2354–2363.
- HORNSBY, K. & EGENHOFER, M. (2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, vol. 36(1): pp. 177–194.
- HUNTER, A. (2007). *Sensor-based animal tracking*. Ph.D. thesis, Department of Geomatics Engineering, University of Calgary.
- IMFELD, S. (2000). *Time, Points and Space - towards a better Analysis of Wildlife Data in GIS*. Ph.D. thesis, GIS Unit, Department of Geography, University of Zurich.
- IWAHASHI, J. & PIKE, R.J. (2007). Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology*, vol. 86(3): pp. 409–440.
- KRANSTAUBER, B., KAYS, R., LAPOINT, S., WIKELSKI, M. & SAFI, K. (2012). A dynamic Brownian bridge movement model to estimate utilization distributions

## Bibliography

- for heterogeneous animal movement. *Journal of Animal Ecology*, vol. 81(4): pp. 738–746.
- KRANSTAUBER, B., SAFI, K. & BARTUMEUS, F. (2014). Bivariate Gaussian bridges: directional factorization of diffusion in Brownian bridge models. *Movement Ecology*, vol. 2(1): p. 5.
- KRÜGER, R., THOM, D. & ERTL, T. (2014). Visual analysis of movement behavior using web data for context enrichment. In *Proceedings of the IEEE Pacific Visualization Symposium 2014*, IEEE, pp. 193–200. 7th International IEEE Pacific Visualization Symposium 2014, PacificVis 2014, March 4-7, 2014, Yokohama, Kanagawa, JP.
- LAUBE, P. (2009). Progress in Movement Pattern Analysis. In *Behaviour Monitoring and Interpretation – BMI*, (eds.) B. Gottfried & H. Aghajan, IOS Press, Amsterdam, NL, *Ambient Intelligence and Smart Environments*, vol. 3, pp. 43–71.
- LAUBE, P. (2014). *Computational Movement Analysis*. Springer, Berlin, DE.
- LAUBE, P., DENNIS, T., FORER, P. & WALKER, M. (2007). Movement beyond the snapshot-Dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems*, vol. 31(5): pp. 481–501.
- LAUBE, P., KREVELD, M. & IMFELD, S. (2005). Finding REMO – Detecting Relative Motion Patterns in Geospatial Lifelines. In *Developments in Spatial Data Handling*, (ed.) P. Fisher, Springer, Berlin, DE, pp. 201–215.
- LAUBE, P. & PURVES, R. (2011). How fast is a cow? Cross-Scale Analysis of Movement Data. *Transactions in GIS*, vol. 15(3): pp. 401–418.
- LEE, W.C. & KRUMM, J. (2011). Trajectory preprocessing. In *Computing with spatial trajectories*, (eds.) J. Zheng & X. Zhou, Springer, Berlin, DE, chap. 1, pp. 3–33.
- LEVIN, S. (1992). The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, vol. 73(6): pp. 1943–1967.
- LONG, J. & NELSON, T. (2013). A review of quantitative methods for movement data. *International Journal of Geographical Information Science*, vol. iFirst: pp. 1–27.
- LONG, X., JIN, L. & JOSHI, J. (2012). Exploring trajectory-driven local geographic topics in foursquare. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ACM, New York, NY, USA, pp. 927–934. 14th ACM Conference on Ubiquitous Computing, UbiComp 2012, September 5-8, 2012, Pittsburgh, PA, USA.
- LONGLEY, P., GOODCHILD, M., MAGUIRE, D. & RHIND, D. (2010). *Geographic Information Systems and Science*. John Wiley & Sons Ltd, Chichester, West Sussex, England.

- MARK, D. (1975a). Computer Analysis of Topography: A Comparison of Terrain Storage Methods. *Geografiska Annaler*, vol. 57(3/4): pp. 179–188.
- MARK, D. (1975b). Geomorphometric Parameters: A Review and Evaluation. *Geografiska Annaler*, vol. 57(3/4): pp. 165–177.
- MARK, D. (2003). Geographic Information Science: Defining the Field. In *Foundations of Geographic Information Science*, (eds.) M. Worboys, M. Duckham & M. Goodchild, CRC Press, chap. 1, pp. 3–18.
- MILLER, H. (2008). Geographic Data Mining and Knowledge Discovery. In *The Handbook of Geographic Information Science*, (eds.) J. Wilson & A. Fotheringham, Wiley Online Library, chap. 19, pp. 352–366.
- MILLSPAUGH, J., NIELSON, R., McDONALD, L., MARZLUFF, J., GITZEN, R., RITTENHOUSE, C., HUBBARD, M. & SHERIFF, S. (2006). Analysis of resource selection using utilization distributions. *Journal of Wildlife Management*, vol. 70(2): pp. 384–395.
- MONTELLO, D. (2001). Scale in geography. In *International Encyclopedia of the Social and Behavioral Sciences*, (eds.) J. Neil & B. Paul, Elsevier, pp. 13501–13504.
- MORELLET, N., BONENFANT, C., BÖRGER, L., OSSI, F., CAGNACCI, F., HEURICH, M., KJELLANDER, P., LINNELL, J., NICOLOSO, S., SUSTR, P., URBANO, F. & MYSTERUD, A. (2013). Seasonality, weather and climate affect home range size in roe deer across a wide latitudinal gradient within Europe. *Journal of Animal Ecology*, vol. 82(6): pp. 1326–1339.
- NAMS, V. (2005). Using animal movement paths to measure response to spatial scale. *Oecologia*, vol. 143(2): pp. 179–188. Springer, Berlin, DE.
- NATHAN, R., GETZ, W., REVILLA, E., HOLYOAK, M., KADMON, R., SALTZ, D. & SMOUSE, P. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, vol. 105(49): pp. 19052–19059.
- NESTI, I., POSILLICO, M. & LOVARI, S. (2010). Ranging behaviour and habitat selection of Alpine chamois. *Ethology Ecology & Evolution*, vol. 22(3): pp. 215–231.
- NIELSON, R., SAWYER, H. & McDONALD, T. (2011). R package BBMM, Brownian bridge movement model.
- NOULAS, A., SCELLATO, S., MASCOLO, C. & PONTIL, M. (2011). An Empirical Study of Geographic User Activity Patterns in Foursquare. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, AAAI Press, Menlo Park, CA, USA, pp. 570–573. 5th International AAAI Conference on Weblogs and Social Media, ICWSM 2011, July 17–21, 2011, Barcelona, ES.

## Bibliography

- ORELLANA, D. (2012). *Exploring Pedestrian Movement Patterns*. Ph.D. thesis, Wageningen University.
- ORELLANA, D. & RENSO, C. (2010). Developing an interactions ontology for characterizing pedestrian movement behavior. In *Movement-Aware Applications for Sustainable Mobility: Technologies and Approaches*, (ed.) M. Wachowicz, IGI Global Publishing, New York, NY, USA, chap. 5, pp. 62–86.
- ORELLANA, D., WACHOWICZ, M., DE KNEGT, H., LIGTENBERG, A. & BREGT, A. (2010). Uncovering Patterns of Suspension of Movement. In *GIScience 2010: Extended Abstracts*, (eds.) R. Purves & R. Weibel. 6th International Conference, GIScience 2010, September 14-17, 2010, Zurich, CH.
- O’SULLIVAN, D. & UNWIN, D. (2010). *Geographic Information Analysis*. John Wiley & Sons Inc, Hoboken, NJ, USA.
- PARENT, C., SPACCAPIETRA, S., RENSO, C., ANDRIENKO, G., ANDRIENKO, N., BOGORNY, V., DAMIANI, M., GKOUALALAS-DIVANIS, A., MACEDO, J., PELEKIS, N., THEODORIDIS, Y. & YAN, Z. (2013). Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR)*, vol. 45(4): p. 42.
- PEUQUET, D. (1994). It’s About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. *Annals of the Association of American Geographers*, vol. 84(3): pp. 441–461.
- PIKE, R., EVANS, I. & HENGL, T. (2009). Geomorphometry: A Brief Guide. In *Geomorphometry: Concepts, Software, Applications*, (eds.) T. Hengl & H. Reuter, Elsevier, chap. 1, pp. 3–30.
- POWELL, R. (2000). Animal Home Ranges and Territories and Home Range Estimators. In *Research Techniques in Animal Ecology: Controversies and Consequences*, (eds.) L. Boitani & T. Fuller, Columbia University Press, New York, New York, USA, chap. 3, pp. 65–110.
- POWELL, R. & MITCHELL, M. (2012). What is a home range? *Journal of Mammalogy*, vol. 93(4): pp. 948–958.
- PURVES, R., LAUBE, P., BUCHIN, M. & SPECKMANN, B. (2014). Moving beyond the point: An agenda for research in movement analysis with real data. *Computers, Environment and Urban Systems*, vol. 47: pp. 1–4.
- RASEMANN, S., SCHMIDT, J., SCHROTT, L. & DIKAU, R. (2004). Geomorphometry in Mountain Terrain. In *Geographic Information Science in Mountain Geomorphology*, (eds.) M. Bishop & J. Shroder Jr, Springer, Berlin, DE, chap. 4, pp. 101–146.
- RATTENBURY, T. & NAAMAN, M. (2009). Methods for Extracting Place Semantics from Flickr Tags. *ACM Transactions on the Web*, vol. 3(1): pp. 1–30.



- RICHTER, K.F., SCHMID, F. & LAUBE, P. (2012). Semantic trajectory compression: Representing urban movement in a nutshell. *Journal of Spatial Information Science*, vol. 4: pp. 3–30.
- RYKIEL JR., E. (1996). Testing ecological models: the meaning of validation. *Ecological modelling*, vol. 90(3): pp. 229–244.
- SAFI, K., KRANSTAUBER, B., WEINZIERL, R., GRIFFIN, L., REES, E., CABOT, D., CRUZ, S., PROAÑO, C., TAKEKAWA, J., NEWMAN, S., WALDENSTRÖM, J., BENGTSSON, D., KAYS, R., WIKELSKI, M. & BOHRER, G. (2013). Flying with the wind: scale dependency of speed and direction measurements in the modelling of wind support in avian flight. *Movement Ecology*, vol. 1(4).
- SALVUCCI, D. & GOLDBERG, J. (2000). Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the Eye tracking Research & Applications Symposium*, (ed.) A. Duchowski, ACM, pp. 71–78. Eye Tracking Research & Applications Symposium, ETRA 2000, November 6-8, 2000, Palm Beach Gardens, Florida, USA.
- VAN SCHAIK, J. (2010). Future scenarios for the relation between advanced tracking research and urban design and planning. *Journal of Location Based Services*, vol. 4(2): pp. 70–92.
- VAN SCHAIK, J. & VAN DER SPEK, S. (2008). *Urbanism on track: application of tracking technologies in urbanism*, vol. 1. IOS Press Amsterdam.
- SCHILIT, B., ADAMS, N. & WANT, R. (1994). Context-aware computing applications. In *Mobile Computing Systems and Applications*, IEEE, pp. 85–90. 1st Workshop on Mobile Computing Systems and Applications 1994, December 8-9, 1994, Santa Cruz, CA, USA.
- SCHMIDT, A., BEIGL, M. & GELLERSEN, H. (1999). There is more to context than location. *Computers & Graphics*, vol. 23(6): pp. 893–901.
- SCHUESSLER, N. & AXHAUSEN, K. (2009). Processing raw data from global positioning systems without additional information. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2105(1): pp. 28–36.
- SESTER, M., FEUERHAKE, U., KUNTZSCH, C. & ZHANG, L. (2012). Revealing underlying structure and behaviour from movement data. *KI-Künstliche Intelligenz*, vol. 26(3): pp. 223–231.
- SINHA, G. & MARK, D. (2005). Measuring similarity between geospatial lifelines in studies of environmental health. *Journal of Geographical Systems*, vol. 7(1): pp. 115–136.
- SOLEYMANI, A., CACHAT, J., ROBINSON, K., DODGE, S., KALUEFF, A. & WEIBEL, R. (2014). Integrating cross-scale analysis in the spatial and temporal domains for

## Bibliography

- classification of behavioral movement. *Journal of Spatial Information Science*, (8): pp. 1–25.
- SPACCAPIETRA, S., PARENT, C., DAMIANI, M., DE MACEDO, J., PORTO, F. & VANGENOT, C. (2008). A conceptual view on trajectories. *Data & knowledge engineering*, vol. 65(1): pp. 126–146.
- VAN DER SPEK, S., VAN SCHAIK, J., DE BOIS, P. & DE HAAN, R. (2009). Sensing human activity: GPS tracking. *Sensors*, vol. 9(4): pp. 3033–3055.
- VAN DER SPEK, S. (2008). Tracking technologies: an overview. In *Urbanism on track: application of tracking technologies in urbanism*, (eds.) J. van Schaick & S. van der Spek, IOS Press Amsterdam, vol. 1, chap. 3, pp. 25–32.
- VAN DER SPEK, S., VAN LANGELAAR, C. & KICKERT, C. (2013). Evidence-based design: satellite positioning studies of city centre user groups. *Proceedings of the ICE-Urban Design and Planning*, vol. 166(4): pp. 206–216.
- STRAUMANN, R. & PURVES, R. (2008). Delineation of valleys and valley floors. In *Geographic Information Science*, (eds.) T. Cova, H. Miller, K. Beard, A. Frank & M. Goodchild, Springer, Berlin, DE, *Lecture Notes in Computer Science*, vol. 5266, pp. 320–336. 5th International Conference on Geographic Information Science, GIScience 2008, September 23–26, 2008, Park City, UT, USA.
- THÉRIAULT, M., VANDERSMISSEN, M.H., LEE-GOSSELIN, M. & LEROUX, D. (1999). Modelling commuter trip length and duration within GIS: Application to an OD survey. *Journal for Geographic Information and Decision Analysis*, vol. 3(1): pp. 41–55.
- TOBLER, W. (1993). Three presentations on geographical analysis and modeling. Tech. rep., National Center for Geographic Information and Analysis (NCGIA).
- TOÏGO, C. (1999). Vigilance behavior in lactating female Alpine ibex. *Canadian Journal of Zoology*, vol. 77(7): pp. 1060–1063.
- TOMKIEWICZ, S., FULLER, M., KIE, J. & BATES, K. (2010). Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365(1550): pp. 2163–2176.
- TOMLIN, C. (1983). A map algebra. In *Proceedings of the Harvard Computer Graphics Conference 1983*, Laboratory for Computer Graphics and Spatial Analysis, Harvard. Harvard Computer Graphics Conference, July 31–August 4, 1983, Cambridge, Massachusetts, USA.
- TURCHIN, P. (1998). *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Sinauer Associates Sunderland, Massachusetts.

- UMAIR, M., KIM, W., CHOI, B. & JUNG, S. (2014). Discovering personal places from location traces. In *Proceedings of the 16th International Conference on Advanced Communication Technology (ICACT 2014)*, IEEE, pp. 709–713. ICACT 2014, February 16-19, 2014, Phoenix Park, South Korea.
- URBANO, F., CAGNACCI, F., CALENGE, C., DETTKI, H., CAMERON, A. & NETELER, M. (2010). Wildlife tracking data management: a new vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365(1550): pp. 2177–2185.
- VASCONCELOS, M., RICCI, S., ALMEIDA, J., BENEVENUTO, F. & ALMEIDA, V. (2012). Tips, dones and todos: uncovering user profiles in foursquare. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*, ACM, New York, NY, USA, pp. 653–662. 5th ACM International Conference on Web Search and Data Mining, WSDM'12, February 8-12, 2012, Seattle, WA, USA.
- VERSICHELE, M., NEUTENS, T., DELAFONTAINE, M. & VAN DE WEGHE, N. (2012). The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography*, vol. 32(2): pp. 208–220.
- WARTMANN, F., PURVES, R. & VAN SCHAIK, C. (2010). Modelling ranging behaviour of female orang-utans: a case study in Tuanan, Central Kalimantan, Indonesia. *Primates*, vol. 51(2): pp. 119–130.
- WEBB, S., RIFFELL, S., GEE, K. & DEMARAIS, S. (2009). Using fractal analyses to characterize movement paths of white-tailed deer and response to spatial scale. *Journal of Mammalogy*, vol. 90(5): pp. 1210–1217.
- WHITE, C., BERNSTEIN, D. & KORNHAUSER, A. (2000). Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies*, vol. 8(1–6): pp. 91–108.
- WOOD, J. (1996). *The Geomorphological Characterisation of Digital Elevation Models*. Ph.D. thesis, University of Leicester, UK. <http://www.soi.city.ac.uk/~jwo/phd/>, Zugriff: 08.07.2009.
- ZEILER, A. (2012). *Zeitliche Multiskalen-Analyse von Bewegungsdaten*. Master's thesis, GIS Unit, Department of Geography, University of Zurich.
- ZHENG, Y., CHEN, Y., LI, Q., XIE, X. & MA, W.Y. (2010). Understanding transportation modes based on GPS data for web applications. *ACM Transactions on the Web (TWEB)*, vol. 4(1): p. 1.
- ZHOU, J. & SHI, J. (2009). RFID localization algorithms and applications — a review. *Journal of Intelligent Manufacturing*, vol. 20(6): pp. 695–707.
- ZIMMERMANN, M., KIRSTE, T. & SPILIOPOULOU, M. (2009). Finding Stops in Error-Prone Trajectories of Moving Objects with Time-Based Clustering. In *Intelligent*

## Bibliography

*Interactive Assistance and Mobile Multimedia Computing*, (eds.) D. Tavangarian, T. Kirste, D. Timmermann, U. Lucke & D. Versick, Springer, Berlin, DE, *Communications in Computer and Information Science*, vol. 53, pp. 275–286. International Conference, IMC 2009, November 9-11, 2009, Rostock-Warnemünde, DE.

# Appendix A.

## Curriculum vitae

GSCHWEND Christian

Born August 28, 1981 in St. Gallen, Switzerland

Citizen of Altstätten SG, Switzerland

### Education

- 1996 – 2000 High School “Matura” type “C”, Kantonsschule am Burggraben, St. Gallen, Switzerland.
- 2001 – 2003 Physics, Diploma studies (1st year), Department of Physics, ETH Zurich, Switzerland. (*without qualification*)
- 2003 – 2005 Computer Science, BSc studies (1st year), Department of Computer Science ETH Zurich, Switzerland. (*without qualification*)
- 2005 – 2010 Geography, BSc and MSc studies, Department of Geography, University of Zurich, Switzerland. *Specialization: GIScience. Minor: Mathematics.*
- 2010 *MSc thesis: “Erfassung umgangssprachlicher Geographie im Zusammenhang mit Geomorphometrie”,* advised by Prof. Dr. Ross Purves and Prof. Dr. Robert Weibel.
- 2011 – 2015 Dissertation, PhD studies, Department of Geography, University of Zurich, Switzerland.  
*PhD thesis: “Relating Movement to Geographic Context – Effects of Preprocessing, Relation Methods and Scale”,* advised by PD Dr. Patrick Laube, Prof. Dr. Robert Weibel, Prof. Dr. Ross Purves and Prof. Dr. Nico Van de Weghe.



## Appendix B.

### Complete publication list

The research published during the time of my PhD studies is mentioned in the following list. Publications in connection with my MSc thesis are marked with a star (\*).

- (\*)GSCHWEND, C. & PURVES, R.S. (2011). Comparing Flickr tags to a geomorphometric classification. In *Proceedings of the 19th GIS Research UK Annual Conference*, (eds.) C.E. Jones, A. Pearson, G. Hart & N. Groome, University of Portsmouth with Ordnance Survey, pp. 174-178. GISRUK 2011, April 27-29, 2011, Portsmouth, UK.
- GSCHWEND, C. & LAUBE, P. (2012). Challenges of Context-Aware Movement Analysis - Lessons learned about Crucial Data Requirements and Pre-processing. In *Proceedings of the 20th GIS Research UK Annual Conference*, (eds.) D. Whyatt & B. Rowlingson, Lancaster University, pp. 241-246. GISRUK 2012, April 11-13, 2012, Lancaster, UK.
- (\*)GSCHWEND, C. & PURVES, R.S. (2012). Exploring Geomorphometry through User Generated Content: Comparing an Unsupervised Geomorphometric Classification with Terms Attached to Georeferenced Images in Great Britain. *Transactions in GIS*, vol. 16(4): pp. 499-522.
- GSCHWEND, C. & LAUBE, P. (2014). Scale Effects in Relating Movement to Geographic Context. In *Proceedings of the 8th International Conference on Geographic Information Science*, (eds.) M. Duckham, E. Pebesma, K. Stewart & A.U. Frank, TU Vienna, pp. 51-56. GIScience 2014, September 23-26, 2014, Vienna, AT.