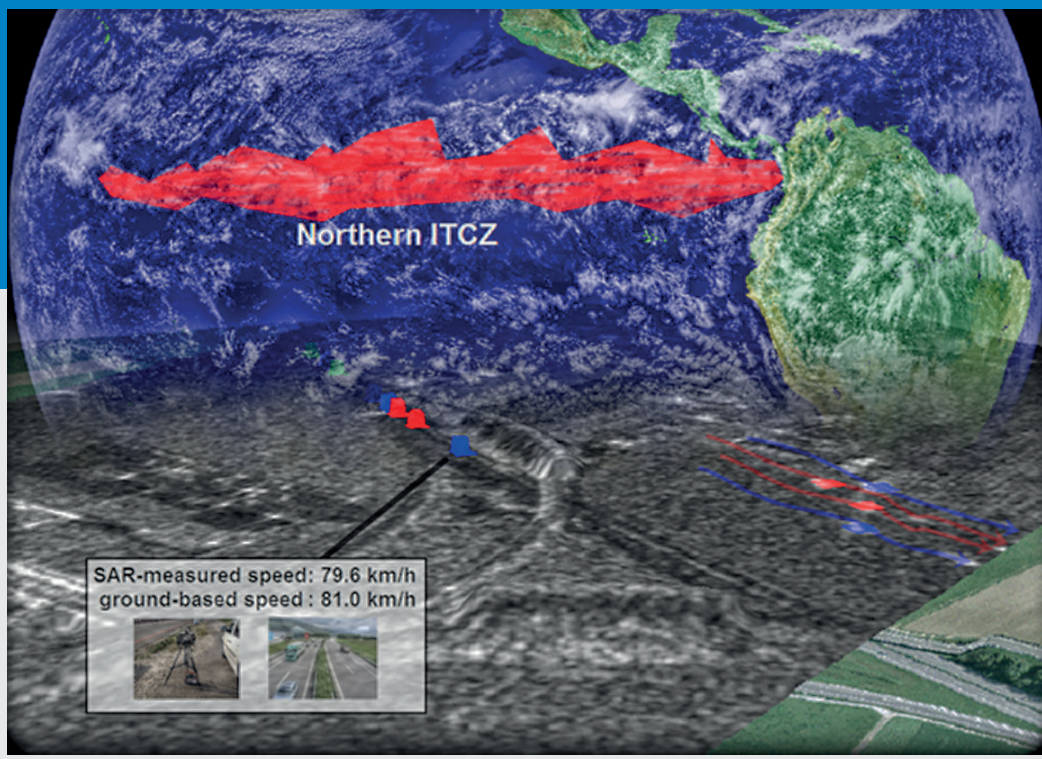
Remote Sensing Series 64

Daniel Henke

Object tracking in remotely sensed data



Remote Sensing Laboratories
Department of Geography
University of Zurich, 2014

Remote Sensing Series 64

Daniel Henke

Object tracking in remotely sensed data

Remote Sensing Laboratories
Department of Geography
University of Zurich, 2014

Front page: image collage of the two main projects within this PhD thesis. Top: the tracking of the double ITCZ in optical satellite imagery. Bottom: tracking of moving vehicles on a highway using SAR data.

Henke, Daniel

Object tracking in remotely sensed data.

Remote Sensing Series, Vol. 64
Remote Sensing Laboratories, Department of Geography, University of Zurich
Switzerland, 2014

ISBN: 978-3-03703-030-1

This work was approved as a PhD thesis by the Faculty of Science of the University of Zurich in the spring semester 2014. Doctorate committee: Prof. Dr. Michael E. Schaepman (chair), Dr. Erich Meier (dissertation supervisor), Dr. David Small, Prof. Dr. Renato Pajarola. External examiner: Prof. Dr. Konrad Schindler (ETH Zürich, Switzerland).

SUMMARY

Remote sensing is one of the most valuable techniques available today for acquiring consistent, repeated high resolution observations of large-scaled phenomena and processes. However, the amount of data is extremely high due to the continuing increase in coverage and repetition rates in recent years. Manual interpretation of the raw data by human experts is therefore in many cases impracticable. The need for automatic procedures to extract useful semantic information is obvious and also crucial for meeting the increasing demands of the latest generations of remote sensing instruments and associated applications.

Automatic methods offer a wide gamut of potential benefits, particularly when considering the temporal dimension, and are the focus of many research fields. Potential applications range from “global change” studies, e.g. changes in vegetation and climate, ice melt, etc. to the mapping of natural disasters such as earthquakes, tsunamis, landslides or avalanches, to tracking of moving objects, e.g. for traffic monitoring or surveillance. Due to the complexity of these phenomena a fully automatic processing is often not possible yet. However, reliable, fast and objective processing of the recorded data in all these cases is of great importance and requires advanced methods. A class of methods used in recent years for analyzing temporal processes in other research fields are "hidden state space" models: they deliver promising results and have been established as some of the most advanced available.

This dissertation investigates the potential of this specific class of methods for remote sensing applications demonstrated for two *prima facie* very different phenomena: One application deals with tracking the state of the Intertropical Convergence Zone (ITCZ) in optical satellite imagery, and the other with traffic monitoring using synthetic aperture radar (SAR) data. Apart from the differing sensors and the resulting different image processing methodologies, the diversity of applications concerned with temporal-dynamic behavior (temporal interval of one day vs. a few fractions of a second) in their respective state space (discrete vs. continuous) is of interest. This thesis demonstrates that with sophisticated image processing methods and adapted, refined "hidden state space" models, significant improvements in the tracking of objects in remotely sensed data can be achieved despite readily apparent differences between the presented remote sensing applications. In the ITCZ project, the formation states of the convection zone of a five-year test period could be determined with an accuracy on par with human experts. In a traffic monitoring application for a particular highway segment, 13 out of 14 trucks were detected correctly and their velocities and positions were estimated with an average accuracy of 1.2 m/s and 10m respectively. A discussion of the main findings highlights advantages and limitations of the proposed techniques and suggests possible improvements as well as future perspectives with respect to both the presented and potential further application areas.

ZUSAMMENFASSUNG

Die Fernerkundung ist heutzutage eine der wertvollsten Methoden, um verschiedenste grossräumige Phänomene und Prozesse konsistent, repetitiv und hochauflösend zu erfassen. Allerdings sind die so anfallenden Datenmengen aufgrund der hohen räumlichen und zeitlichen Dimensionen extrem gross und über die letzten Jahre stetig angewachsen. Eine direkte Interpretation der Bilder durch das menschliche Auge ist daher in vielen Fällen kaum möglich. Deshalb sind automatische Verfahren zur semantischen Informationsextraktion entscheidend, um den immer höheren Anforderungen der neusten Generationen von Fernerkundungssensoren und -anwendungen auch zukünftig gerecht zu werden.

Dabei ist der potentielle Nutzen solcher automatischer Verfahren besonders unter Berücksichtigung der zeitlichen Dimension breitgefächert und im Fokus vieler Forschungsgebiete. Mögliche Anwendungsfelder reichen von „Global Change“ - Studien, z.B. Vegetationsveränderungen, Klimawandel, Eisschmelze, usw. über das Kartieren von Naturkatastrophen wie etwa Erdbeben, Tsunamis, Erdrutsche oder Lawinen bis hin zur Erfassung bewegter Objekte wie beispielwiese zur Verkehrskontrolle oder Überwachung. Aufgrund der Komplexität dieser Phänomene ist bisher eine vollautomatische Verarbeitung oft noch nicht möglich. Für all diese Anwendungen ist eine zuverlässige, schnelle und objektive Verarbeitung der aufgenommenen Daten von grosser Bedeutung und erfordert modernste Verfahren. Eine Klasse von Verfahren zur Verarbeitung zeitlicher Prozesse, die in den letzten Jahren in anderen Forschungsgebieten vielversprechende Ergebnisse lieferte und sich als eine der fortschrittlichsten etabliert hat, sind die „Hidden State Space“ Modelle.

Diese Dissertation untersucht das Potential dieser speziellen Klasse von Verfahren für Fernerkundungsanwendungen anhand zweier auf den ersten Blick sehr unterschiedlicher Prozesse: Eine Anwendung beschäftigt sich mit der Verfolgung des Zustands der Intertropischen Konvergenzzone (ITCZ) in optischen Satellitenbilddaten und die andere mit der Verkehrsüberwachung mit Hilfe von Daten eines abbildenden Radarsystems (Synthetic Aperture Radar, SAR). Neben den unterschiedlichen Sensoren und der daraus folgenden unterschiedlichen Bildverarbeitung ist besonders die Verschiedenheit der Anwendungen im zeitlich dynamischen Verhalten (Zeitreiheninterval von einem Tag vs. wenige Sekundenbruchteile) im jeweiligen Zustandsraum (diskrete vs. kontinuierlich) von Interesse. Im Rahmen dieser Arbeit wird gezeigt, dass trotz dieser augenscheinlichen Unterschiede der präsentierten Fernerkundungsanwendungen mit angepassten Bildverarbeitungsmethoden und verfeinerten „Hidden State Space“ Modellen deutliche Verbesserungen bei der Verfolgung von Objekten in Fernerkundungsdaten erreicht werden können. Im ITCZ Projekt konnten die Formationszustände der Konvektionszone von einer fünfjährigen Testperiode mit einer von Experten visuell erzielten, vergleichbaren Genauigkeit bestimmt werden. Bei der Verkehrsüberwachung konnten auf einem Autobahnabschnitt 13 von 14 Fahrzeugen richtig detektiert werden und ihre Geschwindigkeit und Position mit einer durchschnittlichen Genauigkeit von 1.2 m/s bzw. 10m geschätzt werden. In der anschliessenden Diskussion der Forschungsergebnisse werden die Vor- und Nachteile der entwickelten Ansätze besprochen und auf mögliche Verbesserungen hingewiesen. Abschliessend wird ein Ausblick für weiterführende Arbeiten und Anwendungsgebiete gegeben.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Remote sensing applications

Understanding, planning and monitoring earth systems and processes of regional to global scale are currently of increasing importance (Kump et al., 2010; Cornell et al. 2012; Biermann et al., 2012). To tackle these issues, remote sensing technologies can play a key role to support decision makers in their work (De Leeuw et al., 2010). However, to extract useful information from the large volume of remotely sensed data sets – utilizing the spatial, temporal and spectral dimensions – enabling experts in the respective field to make meaningful data interpretations, state-of-the-art methods have to be applied. One can differentiate between "static" problems where the necessary information can be extracted using sophisticating image processing and interpretation algorithms based on a single data acquisition, and "dynamic" problems where the temporal component in successive data acquisitions must be additionally considered.

The variety of remote sensing applications is immense, having grown steadily in recent years and gaining access to more and more research fields. In the broad field of environmental science (Barrett and Crutis, 1999), remote sensing is extensively used for monitoring the Earth's biosphere, hydrosphere, lithosphere and atmosphere. A broad overview of biosphere remote sensing of vegetation and ecology can be found in Thenkabail et al. (2012) and Horning et al. (2010). An overview of hydrosphere applications of remote sensing is given for snow in Dietz et al. (2012), for ice in Quincey and Luckman (2009) and for water e.g. in Odermatt et al. (2012). Soil and terrain mapping is reviewed by Mulder at el. (2011), cloud and precipitation retrieval using atmospheric remote sensing is reviewed by Stephens and Kummerow (2007); a specific application can be found in Henke et al. (2012a). Further fields of substantial remote sensing services can be found in mapping natural hazards including earthquakes, flooding and landslides (Joyce et al., 2009; Tralli et al. 2005; Metternicht, 2005), land use (Rogan and Chen, 2004; Schindler 2012), urban planning (Tooke and Coops, 2013), the generation of Digital Elevation Models (DEMs) (SRTM: e.g. Rabus et al., 2003; TanDEM-X: e.g. Gruber et al., 2012) with ground penetrating SAR systems experimentally for glacier beds (Henke and Meier, 2010), surveillance and traffic monitoring (Henke et al., 2013).

Not only the fields of applications are manifold, but also the acquisition technologies consisting of passive sensors, mainly optical with a large number of spectral bands (Schaepman et al., 2009), and active sensors (Jin et al., 2013), mainly radar (Synthetic Aperture Radar, SAR) e.g. in Reigber et al. (2013) and laser (Light Detection And Ranging, LiDAR) e.g. in (Mallet and Bretar, 2009), on spaceborne or airborne platforms. The study areas vary from global to local regions (Foody and Curran, 1994) and from long-term phenomena protracting over several years (e.g. climatology, phenology) to fast processes evolving within the time scale of minutes or even seconds (e.g. traffic, some natural hazards). Especially analyzing the temporal behavior of phenomena are complex and challenging tasks that need sophisticated algorithms to automatically extract meaningful information and are emphasized in this thesis.

In the past, many studies have been conducted: long-term studies deal mainly with phenological or climatological trends in optical data, and are based on clustering (Viovy, 2000), linear regression (Tottrup and Rasmussen, 2004), least-square methods (Jönsson and Eklundh, 2004), wavelets (Sakamoto et al., 2005; Galford et al., 2008), curve fitting using harmonic and spline functions (Bradley et al., 2007), intercomparing and ensemble-estimating using a spectrum of trend analysis methods (White et al., 2009), seasonal-trend decomposition algorithms (De Jong et al, 2012), hypothesis tests on general image statistics (Chen et al., 2008) or spatio-temporal Markov random fields (Bain et al., 2011). Mid- to short-term studies in the order of weeks to minutes rely often on active sensors like SAR; examples are the tracking of icebergs using semi-automatic iceberg segmentation and classification with subsequent matching of size and similarity over time (Silva and Bigg, 2005), the estimation of glacier flows using correla-

tion- and wavelet-based image matching (Schubert at al., 2013) or traffic monitoring applications using Displaced Phased Center Array (DPCA) solutions (Gierull and Sikaneta, 2003), Along Track Interferometry (ATI) (Budillon et al., 2008) or Space-Time Adaptive Processing (STAP) (Cerutti-Maori et al., 2012). However, only a few studies rely on new developments in the field of discrete-time hidden state space models already used frequently in state-of-the-art tracking algorithms in different fields of computer science for a robust analysis of a process' temporal behavior. Thus, one major motivation of this thesis is to demonstrate the benefits of these models in the context of remote sensing applications.

1.2 Hidden state space models

Hidden state space models exist for continuous state spaces, first introduced by Kalman (1960) and discrete state spaces with the hidden Markov model (HMM), first indroduced by Rabiner (1989).

The concept is common for all approaches: there are two time-discrete sequences of random variables, a "*hidden*" one and an observed one. The structure in the system is provided by assuming that the hidden sequence is a Markov chain (in general a first order Markov chain, i.e. the state x_t at time t depends only on the state x_{t-1} at the previous time step $t-1$). One also assumes that only the hidden variable at a specific time t has a *direct* influence on the observed variable at the same time t. A basic graphical schema of the model is illustrated in Fig. 1.1.

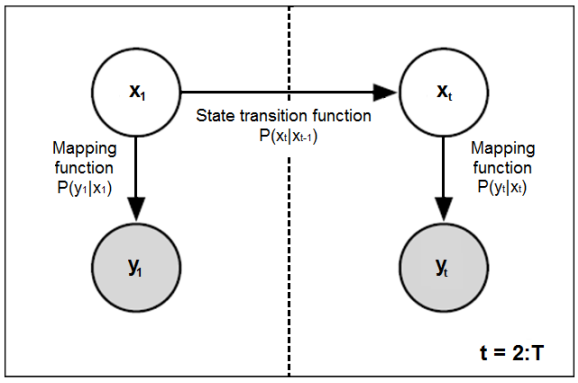


Fig. 1.1 – Graphical representation of a basic hidden state space model. (Similar e.g. in Smyth, 1997)

Thus, both approaches estimate the true values of states recursively over time using incoming measurements and a mathematical process model. The major difference between a Kalman filter model and an HMM is the continuous and discrete state space, respectively: a hidden state space model where the state can take an arbitrary value in the state space is denoted as a Kalman filter, a hidden state space model where the state belongs to a pre-defined, discrete state set is called an HMM. The difference in the state space implies some minor differences in the mapping and state transition functions as well as the learning of the model parameters and the estimation of the optimal state sequence for an unknown set of observations. The two models are juxtaposed in Table 1.1.

TABLE 1.1
HIDDEN STATE SPACE MODEL: HIDDEN MARKOV MODEL VS. KALMAN FILTER

	Hidden Markov Model	Kalman Filter
State space	The state space is discrete.	The state space is continuous. A corresponding error covariance matrix gives information about the certainty of the estimated state.
Mapping of the state space to the observation space	Emission probability B: b: state space S→ observation space V	Transformation function H: h: state space X→ observation space Z
State transition	Discrete state-transition matrix A.	Continuous state progression function A. Generally, an underlying physical (dynamic) model
Optimal state sequence estimation	Maximization of the probability of a state sequence given the observation sequence and the model parameters using numerical solutions (common: Viterbi-algorithm)	Iterative updating and predicting of the state determined by the innovation and the underlying model, respectively.
Learning	Supervised and unsupervised approaches are common	In general, supervised, as the dynamics of the underlying model and the transformation function are commonly known in advance.

Starting from the standard Kalman filter and HMM, many extensions and modifications were developed to achieve robust tracking capability for complex scenarios. Important advancements to the standard Kalman filter comprise the capability of tracking non-linear temporal behavior (unscented Kalman filter reported in Julier and Uhlmann, 2004), the ensemble Kalman filter (e.g. Evensen, 2009) and the multi-target tracking capability (e.g. Khan et al., 2005; Särkkä et al., 2007; Frank et al., 2012). Significant enhancements to standard HMMs include the modeling of an unknown, potentially infinite state set (Teh et al., 2006), the replacement of Gaussian emission probabilities by other estimators (e.g. Platt, 1999), the modeling of semi-Markov properties (e.g. Yu, 2010) and supervised and semi-supervised learning (e.g. Zhong, 2005). Examples of fields of applications for Kalman filters are robotics (Chen, 2012), sensor fusion with an emphasis on navigation data (Mohamed and Schwarz, 1999), and tracking of moving objects like human motions (e.g. eye tracking by Zhang and Zhang, 2010), animals (e.g. ants by Khan et al., 2005) or vehicles (e.g. Henke et al., 2013). HMMs were traditionally used in speech recognition (Bourlard and Morgan, 1994) but have gained access to fields like medicine (Krogh et al., 2001; Rosen et al., 2002), seismology (Ebel et al., 2007; Chambers 2012) and atmospheric science (Greene et al., 2011; Henke et al., 2012a).

The integration of these developments for improving remote sensing applications is discussed in the following section.

1.3 Hidden state space models for improved remote sensing applications

1.3.1 Overview

In the remote sensing context, hidden state space models have not yet gotten the necessary attention for analyzing temporal phenomena with only few approaches being applied to solely instantaneous remotely sensed input data (Storm tracking by Scharenbroich et al., 2009; crop recognition in remote sensing image sequences by Leite et al., 2011). Most approaches rely rather on reanalysis or composed data that by design include numerical model assumptions and intermediate remote-sensing-aided products as auxiliary inputs in combination with other data sources (e.g. ground- or model-based). An overview of assimilating remote sensing data using hidden state space models (most common: ensemble Kalman filter) in earth science can be found in Reichle, 2008. Specific examples include soil moisture estimations (e.g. Sabater et al., 2007), crop yield forecasts (as in de Wit and van Deipen, 2007) or atmospheric (Houtekamer et al., 2005; Torn and Hakim, 2008), storm (Kieu et al., 2012) and rainfall predictions (French et al., 1994; Ushio et al., 2009). However, these approaches potentially lead to shortcomings

caused by an undervaluing of the interdependencies of instantaneous remotely sensed data on the one hand and the extracted high-level information on the other hand. For improved remote sensing applications, a successful adaption of hidden state space models often requires an application-driven intermeshing of several tasks based on instantaneous raw data: (a) the selection of an appropriate sensor system, (b) an elaborated extraction method of the observation set from the acquired raw data at specific discrete times utilizing sophisticated image and signal processing methods, (c) a meaningful definition of the hidden state space corresponding to the high-level information under investigation that helps experts in the respective field analyze the data, and (d) the incorporation of state-of-the-art model extensions for a highly accurate, robust estimation of the state sequence of the phenomenon.

Thus, in this thesis two studies where modern hidden state space models are used to track the state of the investigated system in instantaneous remotely sensed data are developed, presented and compared: (1) analyzing the temporal behavior of the double ITCZ in optical data (Henke et al., 2012a) and (2) tracking vehicles in SAR images (Henke et al., 2012b; Henke et al., 2013). Although, the two studies differ considerably in the fields of application, their temporal and spatial scales, and in the sensor technology used, it is demonstrated that, even though application specific adaption to the models are necessary, the general processing methods are similar and the achieved results can help to significantly improve existing remote sensing applications. In (1), a large-scale, long-term climate study was conducted using spaceborne optical data. The double ITCZ (dITCZ) is a phenomenon occurring in the boreal spring in the eastern Pacific. Long-term Geostationary Operational Environmental Satellite (GOES) visible (VS) and infrared (IR) data were used to spatially identify and segment the convection zones on both sides of the equator and to track the temporal variability of the ITCZ, specifically to identify four discrete ITCZ cases: dITCZs, northern ITCZ, southern ITCZ, or non-presence events. The proposed method relies on image processing techniques and a combination of a hidden semi Markov model with support vector machine (SVM) methods. In (2), short-term traffic monitoring was conducted using airborne SAR data. The focus was placed on the tracking of vehicles on motorways. Different methods for extracting features from the SAR data were designed depending on the available data channels (single- and multi-channel extraction methods). These features were then tracked using a multi-target Kalman filtering approach.

In the later chapters of this thesis, remote sensing specific problems are addressed and refinements to the models discussed in detail, helping to promote hidden state space models for remote sensing applications.

1.3.2 General processing chain

First, a short overview of the processing chain shows how to extract state sequences from remotely sensed data as illustrated in Fig. 1.2. An appropriate sensor observes the real world with the phenomenon of interest. The dimension of possible scenarios is infinite ("∞"). The acquired data which are disturbed by noise effects (e.g. thermal noise) are digitized to obtain a remotely sensed image that can be displayed on a screen for human perception. However, the contained information is low-level and often hardly interpretable even by experts, in part due to the enormous amount of data ("P"). Therefore, appropriate methods have to be developed to automatically generate semantic information from the raw data.

The first step, which will just be touched on, is the image-preprocessing. This includes for remote sensing applications: noise reduction, geo-referencing, geo-coding and technology-dependent focusing, especially relevant for SAR data. The focus of this thesis is mainly placed on the two successive steps: the appropriate feature extraction, and the tracking of the hidden state of the system. Both have the goal of reducing the dimensionality of the problem's complexity and extracting high-level information ("N" & "S"). The feature extraction is in general based only on single acquisitions at this stage ignoring the dynamic behavior of the process investigated. Important to notice is also that the features extracted for the later tracking do not

necessarily have to be easily interpretable for humans. These features then make up the observation set directly influenced by the hidden state with the high-level, semantic information content. To estimate these hidden states an appropriate model has to be set up to accommodate the temporal behavior of the phenomenon investigated. This is done by adapting appropriate hidden state space models to the specific remote sensing task.

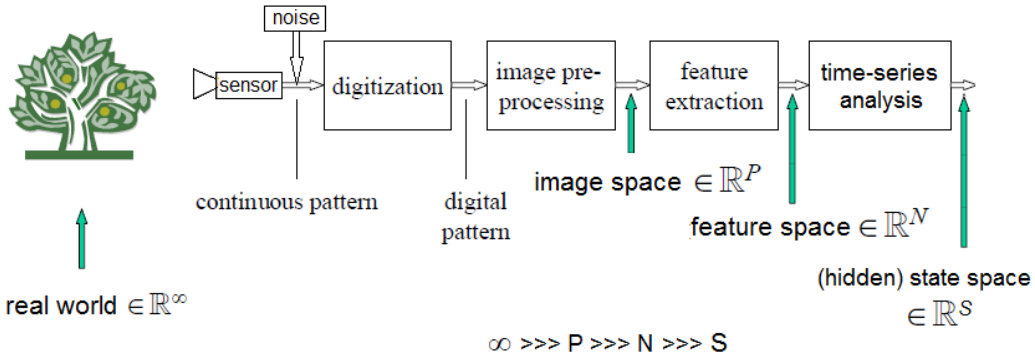


Fig. 1.2 – Overview of processing to extract a system's state of a phenomenon from remotely sensed data utilizing hidden state space models.

1.3.3 Preliminary considerations for the two specific case studies

In remote sensing applications, hidden state space models could potentially improve a large number of temporal analysis methods. As a first step, in this dissertation the realization on two specific case studies is conducted demonstrating the benefits when dealing with "dynamic" remote sensing applications: a climatological investigation of dITCZ phenomena, and traffic management study of vehicles on motorways. For a successful realization in remote sensing studies, the choice of an appropriate experimental setup (sensor & test site), a sophisticated observation set, a meaningful definition of the hidden state space and the incorporation of model extensions are discussed.

1.3.3.1 Sensor & test site

For the dITCZ application, the requirements of the sensor are: (a) long-term, consistent, and frequently acquired images from the area of interest; (b) imaging of the cloud convection and cloud top temperature with a stable geometry (i.e. geostationary optical satellite sensors); and (c) optional: further improved temporal resolution (i.e. IR data can be used to provide nighttime acquisitions). High spatial resolution (~ 8km) is not crucial for observing the large-scale ITCZ phenomenon.

Considering this requirement list, the sensors of choice are geostationary satellite sensors with VS and IR channels. The GOES GridSat dataset (Knapp et al., 2011) with data available for the years 1980 to present, with one VS image per day (21:00 GMT, daylight) and IR images at 3 hour intervals (8 per day) meets the requests.

The test site was determined by the location and time of occurrence of the dITCZ phenomenon: the area of interest is restricted to longitudinal and latitudinal ranges of 180° W to 90° W and 20° N to 20°S (Eastern Pacific) and the period of potential occurrence between February and May of each year (the active dITCZ season).

For traffic monitoring and surveillance applications, the sensor should fulfill the following requirements: (a) provide an (active) illumintation to guarantee applicability independent of sun-light and weather (i.e. work also at night, or when cloudy); (b) high spatial resolution (~ 1m) to be able to correctly detect vehicles; (c) for vehicle tracking in unknown areas (surveil-

lance): imaging capability to map the area while simultaneously tracking the vehicles; (d) flexibility regarding the flight path of the platform; and (e) optional: multi-channel data for more robust moving object extraction.

Given this list, the sensors of choice are Synthetic Aperture Radars mounted on a small airplane or unmanned aerial vehicle (UAV) with good spatial resolution. For the experiments described later, the sensors used are the EADS single-channel UAV-capable ultra-lightweight Miniature Synthetic Aperture Radar (MiSAR) system operating in Ka-band (35 GHz) with a spatial resolution of ~0.5m (Edrich and Weiss, 2008) and DLR's multi-channel F-SAR sensor operating in X-band (9.6 GHz) with a range resolution of about 1.5m (Reigber et al., 2011). The MiSAR experiment was flown in a simple straight geometry, the F-SAR experiment in a more complex, circular flight path.

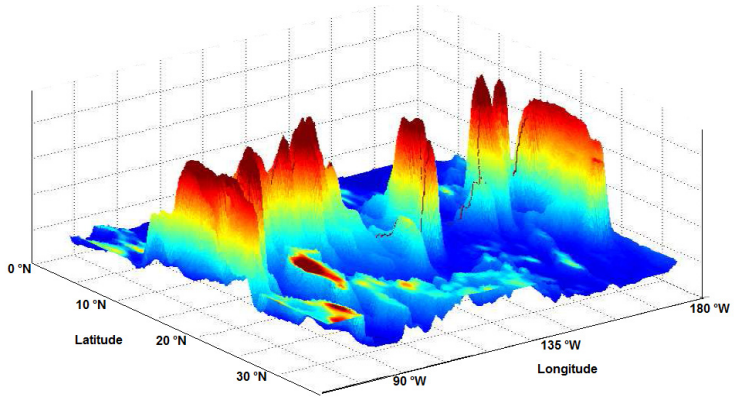
The test sites for the traffic monitoring experiments were selected at motorway intersections as preferred locations: the MiSAR experiment took place in South Germany (three-leg interchange "*Neu-Ulm*"); the F-SAR experiment in central Switzerland (Exit "Oensingen" and interchange "*Härkingen*").

1.3.3.2 The observation sets

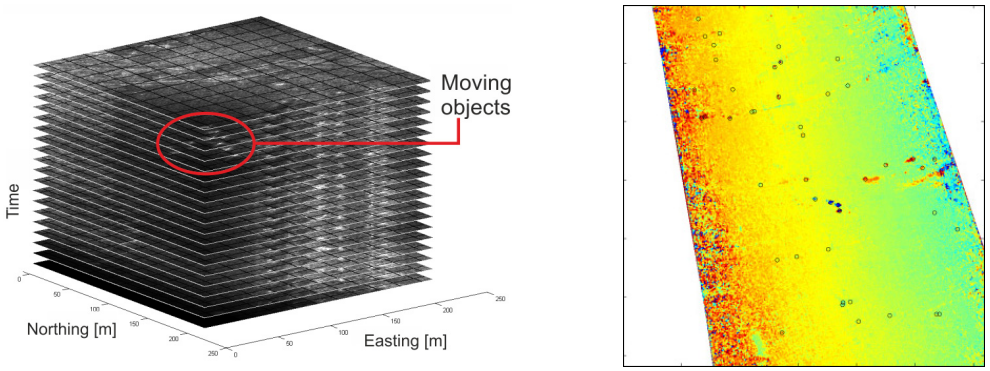
The extraction of a meaningful and reliable observation set is a crucial factor in remote sensing applications. Decisive is a sophisticated approach to separate the signal relevant to the investigated phenomena from signals caused by other irrelevant processes (background clutter).

For the dITCZ study, the major task was the separation of the signal of clouds belonging to the ITCZ convectional band from other cloud structures. The Earth's surface reflectance over the eastern Pacific is extremely weak and can thus be neglected. The decisive factor here is an optimal path (later called: backbone path) that finds the location of the ITCZ band and a subsequent region growing algorithm for extracting the spatial extent of the ITCZ structure, while simultaneously suppressing other cloud formations. Fig. 1.3(a) shows the 3D-plot of the extracted ITCZ structure in the VS image overlaid by the color-coded intensity GOES image containing all cloud structures. The location of the backbone path and the spatial extend of the ITCZ serve then as most important basis for calculating features for the observation set of the ITCZ phenomena.

For the vehicle tracking project, the major task in the feature extraction step was the separation of moving target candidates from the static objects. The decisive factor here was to make use of the temporal component inherited from the SAR acquisition generally left disregarded in standard SAR focusing products (one image for the entire aperture). In contrast to standard outputs of SAR processers, for the approach proposed in this thesis the high number of acquired consecutive radar echo pulses is processed in many highly overlapping images with reduced bandwidth, generating a video-like SAR image stack. The extraction of the observation set can be achieved by two approaches: in the single-channel approach, the SAR image stack is interpreted as a video, and pixels with anomalies compared to a long-term mean background are marked as moving object candidates. The robustness of the single-channel approach increases with the time a pixel is illuminated and can be achieved by either a sensor with a large azimuth antenna beam or spotlight acquisitions (i.e. circular flight tracks). In the multi-channel approach, the phase shift induced by a short spatial separation of the antennas of different channels can be extracted containing non-zero values for moving object candidates and zeros for static objects. Afterwards, both extraction methods can be combined. The observation set is calculated by the center of mass of the potential moving objects. Fig. 1.3(b) shows intermediate results for both the single- and multi-channel extraction methods.



(a) Extraction of ITCZ band (example: northern hemisphere) and suppression of other cloud structures. 3D-plot: ITCZ structure; overlaid image: color-coded intensity GOES image containing all cloud structures with backbone path indicated. Strong non-ITCZ cloud structures in the north east of the region of interest are successfully suppressed.



(b) Extraction of moving object candidates. Left: single-channel extraction based on video-like image stack interpretation. Abnormal signals of moving objects are marked by the red circle. Right: multi-channel extraction based on the motion-induced phase shift. Objects with strong phase shifts compared to the background are extracted.

Fig. 1.3 – Important extraction steps for the observation sets of the two case studies.

1.3.3.3 The semantic interpretation of the hidden state space

Before choosing the hidden state space model, high-level information for the interpreter of the remote sensing data needs to be elaborated in interdisciplinary collaborations.

For the dITCZ study, the state space consists of four mutually exclusive states: double ITCZ, northern ITCZ, southern ITCZ and the non-presence case. The distribution of the ITCZ states in different years and months provides insights to atmospheric scientists helping them understand the dependencies between different global circulation mechanisms (e.g. El Niño and the Southern Oscillation) and improve global climate models. Remote sensing is the only reliable, large-area and frequently available data source in the sparsely observed Eastern Pacific. The need for a robust extraction of the hidden state sequence of the dITCZ phenomena from satellite data is therefore obvious.

For the vehicle tracking study, the state space consists of the continuous space in cartographic coordinates. After SAR-specific corrections (e.g. for Doppler shifts), the hidden state sequence contains the positions and velocities of vehicles in cartographic coordinates and can be used by experts in the field of surveillance, traffic management or traffic police. For surveillance tasks, object movements in uncharted regions can be displayed in a geographic context

giving the extracted trajectories together with SAR imagery to help decision makers. For traffic management, statistics about traffic flows such as traffic volume, average driving speed and predominant driving directions on individual roads or the entire observed road network can be derived to provide useful information for traffic planning. The speed of individual vehicles can be calculated from the extracted trajectories and support speedometer measurements. With remote sensing technology, a synoptic view of traffic can be realized in a cost- and time-efficient way. Hidden state space models provide significant improvements for a robust and reliable estimation of moving objects from SAR data.

1.3.3.4 Model extensions

A further absolutely crucial factor when applying hidden state space models to remote sensing applications is the choice of the model and extensions needed for the specific problem. An understanding of the entire processing chain is required for a successful adaptation of the model.

For the dITCZ phenomena, the discrete state set demands an HMM. However, several extensions are necessary to improve on standard HMM for the dITCZ study:

- The state duration inherited from standard HMMs is not consistent with the typical state durations of dITCZ states (as found in previous publications of the dITCZ). Therefore, the state duration is modeled explicitly (Hidden semi-Markov Model (HsMM));
- The natural Gaussian clustering as in standard HMMs is not valid for the dITCZ phenomena. Thus, the Gaussian emission probabilities of the states are replaced by other classifiers such as logistic regression, Artificial Neural Networks (ANN) or Support Vector Machines (SVM). The most promising approach in the latter studies was a HsMM/SVM hybrid approach;
- Compared to the unsupervised learning approach of standard HMMs, the classification accuracy can be clearly improved by supervised and semi-supervised learning.

For the traffic monitoring project, the state space is continuous, suggesting a Kalman filter based approach. Also here, several extensions are necessary for an appropriate tracking model:

- SAR data experiments imply non-linearity due to SAR-specific effects (such as Doppler), non-linear flight geometry, non-linear phase history or potential multi-static acquisitions. Unscented Kalman filters handle these conditions with high accuracy;
- The presence of multiple moving objects of unknown number demands multi-target tracking capability. This is realized by an extended Markov Chain Monte Carlo (MCMC) sampling technique, known as Rao-Blackwellized Monte Carlo data association. It subdivides the problem into the tracking of a single object using an appropriate Kalman filter and the association problem, i.e. finding the best possible association with a tracking history at a specific time step given several observations and several active trajectories.

These preliminary considerations give a broad overview of the required steps when applying hidden state space models in the remote sensing context. Details on these four topics and their practical implementation are part of the publications presented later in the dissertation. Before that, the main objective, associated research questions and the structure of the thesis are summarized in the next subsections.

1.4 Objective and research questions

The present dissertation promotes the use of hidden state space models to improve existing remote sensing applications. An interdisciplinary approach is needed as the understanding and realization of the processing chain is critical for the improvement of the temporal analysis of remotely sensed data and is required for correct interpretation by experts in the respective field. In this context, the dissertation should answer the following research questions:

- Can hidden state space models help to improve the analysis of the temporal behavior of objects/phenomena in remotely sensed data?
- What prerequisites have to be met to integrate remotely sensed data into the hidden state space model framework?
- What extensions to standard models can be applied and how are they suited to improve the performance?
- Can a certain generality of the investigated methods be confirmed on different applications, scales (temporal & spatial) and data sets?

1.5 Structure of the dissertation

Chapter 1 provides the framework and the definitions required for understanding the peer-reviewed contributions. It familiarizes the reader with the problem of dynamic processes in remote sensing applications and gives an introduction to hidden state space models used to cope with large data volumes required for an in-depth analysis of temporal phenomena. Preliminary considerations for the two case studies are summarized. Research questions and an outline of the dissertation are presented.

Chapter 2 is based on a first-authored peer-reviewed journal article (Henke et al., 2012a) published in Remote Sensing of Environment. The temporal analysis of the dITCZ phenomena over the Eastern Pacific is presented using extended HMM methods. The publication is self-contained in terms of structure and content.

Chapter 3 is based on a first-authored peer-reviewed journal article (Henke et al., 2012b) published in IEEE Transactions on Geoscience and Remote Sensing. It demonstrates how vehicles on a motorway can be tracked in single-channel SAR data using a multi-target Kalman filter approach. The publication is self-contained in terms of structure and content.

Chapter 4 is based on a first-authored peer-reviewed journal article (Henke et al., 2013) submitted to IEEE Transactions on Geoscience and Remote Sensing. It builds upon the publication of Chapter 3 and extends the method for multi-channel data acquired in a non-linear flight geometry. Furthermore, it incorporates data from an along-track interferometer, and demonstrates the correction of SAR-specific position distortions. The publication is self-contained in terms of structure and content.

Chapter 5 addresses the individual research questions, summarizes the main findings from the publications presented in chapters 2 to 4, and provides concluding remarks including an outlook of potential future research directions.

2 AUTOMATED ANALYSIS OF THE TEMPORAL BEHAVIOR OF THE DOUBLE INTERTROPICAL CONVERGENCE ZONE OVER THE EAST PACIFIC

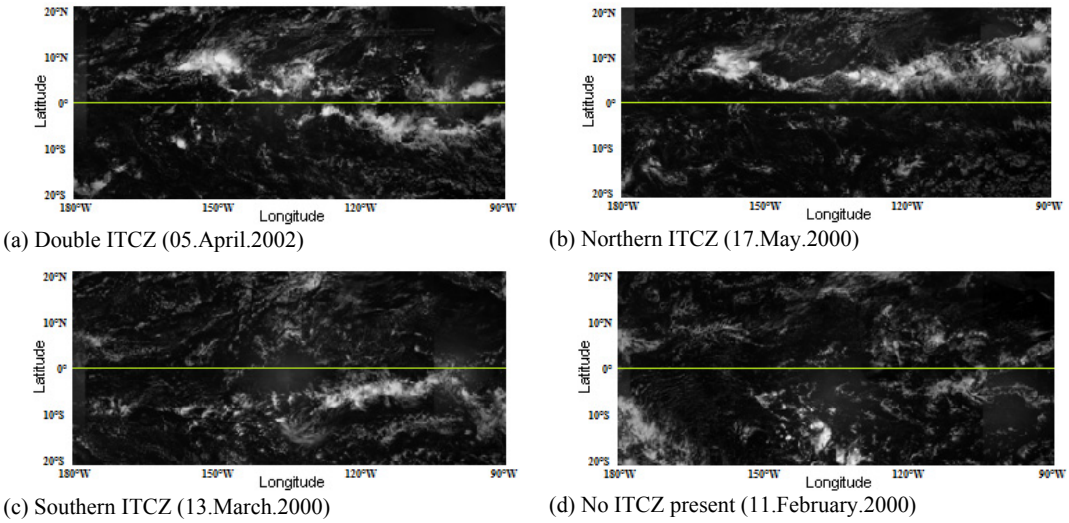
This chapter has been published as: Henke D., Smyth P., Haffke C., & Magnusdottir G. (2012). Automated analysis of the temporal behavior of the double Intertropical Convergence Zone over the east Pacific. *Remote Sensing of Environment, 123*: 418-433.

Abstract

This paper presents new methods for an automated analysis of the double InterTropical Convergence Zone (dITCZ) phenomena on a daily time scale over the east Pacific. Long-term Geostationary Operational Environmental Satellite (GOES) visible and infrared data are used to spatially identify and segment the convection zones over the east Pacific basin on both sides of the equator and to track the temporal variability of the ITCZ, specifically to identify cases of dITCZs, northern or southern ITCZ, or non-presence events. For the segmentation approach, image processing techniques are developed to extract information about the spatial features of the ITCZ in both hemispheres for each satellite image. These features serve as input to a temporal classification algorithm that is based on a combination of hidden semi Markov model (HsMM) and support vector machine (SVM) methods. The performance of the proposed method is competitive with human experts and the methodology can thus be used to conduct an in-depth analysis of the dITCZ. Such an analysis could provide precise information for refining existing weather and climate models over the sparsely observed east Pacific where the dITCZ is greatly over-represented in most models.

2.1 Introduction

The InterTropical Convergence Zone (ITCZ) is a zonally elongated, rather narrow zone of convection over the tropical oceans, playing a key role in the general circulation of the atmosphere. Understanding its variability is essential for improving global climate models. The ITCZ is frequently observed north of the equator in the east Pacific (hereafter referred to as the northern ITCZ (nITCZ)) especially during the boreal summer half-year. Even in winter, the ITCZ may be detected north of the equator as seen in cloud fields. Only in the boreal spring (especially March to April) is the ITCZ also observed south of the equator, which we will refer to as the southern ITCZ (sITCZ). At times this leads to two convergence zones, strong enough to be revealed in cloud fields, simultaneously existing on both sides of the equator in the tropical east Pacific. This phenomenon has been named the double ITCZ (dITCZ) (Hubert et al., 1969). Fig. 2.1 illustrates representative images from the visible channel on geostationary satellites for all four possible combinations of east Pacific ITCZ occurrences in the boreal spring, i.e. (a) dITCZ, (b) nITCZ, (c) sITCZ and (d) with no ITCZ signal present on either side of the equator. In this paper we reserve the symbol dITCZ to describe the occurrence of simultaneous convection zones on both sides of the equator in the east Pacific in instantaneous data. Most previous studies have examined time-averaged and/or area-averaged fields and upon detecting convection both north and south of the equator refer to the phenomenon as a double ITCZ — this is different from our definition of dITCZ above.



(a) Double ITCZ (05.April.2002) (b) Northern ITCZ (17.May.2000)

(c) Southern ITCZ (13.March.2000) (d) No ITCZ present (11.February.2000)

Fig. 2.1 – Examples of the four possible ITCZ states in GOES VS images. The green lines indicate the equator.

Most current global climate models grossly over-represent the dITCZ during the boreal spring (active season). Some of the models even show a dITCZ outside of its active season. This leads to serious model misrepresentations of surface energy fluxes and radiative heating in the east Pacific with global consequences. Previous studies of the double ITCZ have focused on monthly (e.g. Liu and Xie, 2002 and Zhang, 2001) or at most weekly time scales (e.g. Gu et al., 2005 and Lietzke et al., 2001). Magnusdottir and Wang (2008) examined double ITCZ occurrence in daily data but used reanalysis data that by design include numerical model assumptions that may not be valid in the tropics. Masunaga and L'Ecuyer (2010) examined seasonal evolution of the southern branch of the ITCZ in eight years of daily satellite data (2000–2007) and the associated seasonal evolution of surface energy fluxes and the ocean mixed layer.

We are interested in characterizing the occurrence of each of the configurations (or states) of ITCZ on both sides of the equator in the east Pacific on the daily time scale. This requires an

approach that relies only on instantaneous meteorological satellite images, avoiding potential corruption from meteorological analyses that may misrepresent the state of the tropical atmosphere in data sparse regions, such as the tropical east Pacific. In addition, the approach should work in an automated manner to the extent possible in order to handle large data sets for long-term analyses that are impractical for manual annotation.

The algorithm we propose for this task consists of two major parts: first, we spatially localize the ITCZ in both hemispheres in single images (detection); and second, we use time-series analysis to identify different ITCZ states for later statistical analyses. For the first task, we rely on finding a backbone path to indicate the elongated ITCZ cloud band and use a subsequent region-growing algorithm to extract the connected ITCZ cloud field. An alternative approach to this unsupervised methodology would be to use spatio-temporal Markov random fields as described in Bain et al. (2011) — however, this requires time-consuming labeling of images at the pixel level, whereas in the approach proposed here no labeling is required for the spatial segmentation.

For the second task, features extracted from the spatial segmentation are used to classify each image into one of four discrete states (dITCZ, nITCZ, sITCZ and no presence). We make use of an extended hidden semi Markov model/support vector machine (HsMM/SVM) approach, leveraging HMM modeling techniques to integrate information over time (Rabiner, 1989), adapting hybrid HMM methods that traditionally have been applied to speech recognition (Bourlard and Morgan, 1994).

Our results in this paper demonstrate that the proposed method can extract extensive information about the ITCZ phenomenon – about its current state as well as characteristics such as extent and location – at a high temporal resolution and in a fully automatic manner. The method is competitive in accuracy when compared to human experts and it significantly reduces the amount of human effort involved in annotation since only an initial set of training images need to be labeled and not the whole data set.

In the following, we first briefly describe the satellite data that we use and the ground truth acquisition by human experts. Then the method is described in detail. Beginning with localization of the ITCZ, we will focus on the temporal classification of the dITCZ states. We evaluate the performance of the proposed algorithm and present first basic climate statistics based on a five-season sample case. Finally, we conclude with potential future research directions.

2.2 Methods

2.2.1 Dataset and ground truth

The visible (VS) and infrared (IR) data of the GridSat data set from Knapp et al. (2011) are used in this paper[1]. In GridSat, raw data from multiple geostationary satellites are remapped to a standard map projection, recalibrated to optimize temporal homogeneity (especially optimized for the IR data), and stitched together from different satellite instruments to provide a standardized data set for the years 1980 to present. The spatial resolution of the data is approximately 8 km which allows accurate identification of large-scale ITCZ phenomena. VS images are available daily at 21:00 GMT (daylight) while IR images are available every 3 hours since they are independent of the position of the sun. High reflectivity values in VS images correspond to optically thick cloud cover and low IR temperature corresponds to cold cloud tops or cloud tops high in the troposphere. These characteristics are useful indicators for the presence of the ITCZ which, in general, is associated with deep as well as shallow convection (see also Bain et al.,

[1] The files stored in GRISAT-B1 beta netCDF-4 format were provided by Dr. Ken Knapp and are precisely described in the referenced paper. They were generated for the boreal spring (i.e. February-May) of the years 2000-2004 and the specific region of interest (i.e. east Pacific).

2011). For detecting the dITCZ, nITCZ, and sITCZ states, we restrict the area of interest to longitudinal and latitudinal ranges of 180° W to 90° W and 20° N to 20° S and the period of potential occurrence to February until May of each year. In subsequent analysis of the images we rescaled the [0…1] reflectivity values of the VS data and the temperature-normalized IR data to 256-integer images.

Human experts (atmospheric scientists familiar with the ITCZ process) manually generated image labels using a graphical user interface. To assist the human experts three different satellite fields (VS, IR and Total Precipitable Water (TPW)) were displayed in the interface. TPW was obtained from Remote Sensing Systems and is a composite from various microwave sensors. Two atmospheric scientists (Experts I and II) each independently labeled two seasons of data (2000 and 2002), and two more experts (Expert III and IV) each labeled a sequence for March 2000. The sequence labeled by Expert I was used for training and testing of the model; the accuracy of Expert I was measured using the labels from Experts II, III and IV. The scientists assigned each day to one of four stage tags, where the mutually exclusive stage tags were (a) dITCZ, (b) nITCZ, (c) sITCZ, or (d) a non-presence event. The purpose of collecting labels in this manner was to provide ground truth data for training pattern recognition algorithms, as well as for systematically evaluating the quality of such algorithms by comparing algorithm and human accuracy on held-out data.

2.2.2 Locating the ITCZ — the backbone algorithm

The first challenge for automatic tracking of the dITCZ phenomena is to locate the ITCZ position (if present) and its spatial extent, on both sides of the equator. As a preprocessing step, the data are filtered by a standard Gaussian spatial filter. A latitude-dependent weighting (a mixture of two smooth Gaussian functions with maxima at 8° N and 8° S, see Appendix A) is used to add prior knowledge about the most likely ITCZ location in order to bias the path-finding algorithm towards these regions. Thereafter, we estimate the location of the ITCZ by finding a path (or "backbone") that is optimal in the sense that it maximizes the intensity value of the summed pixels from the west to the east. Dynamic programming is used to compute the optimal path as outlined below.

Let I be a two-dimensional N x M array of intensity values corresponding to the satellite image with N referring to the number of columns and M to the number of rows. Let P be the matrix storing the estimated backbone path from left to right and P^A an auxiliary variable containing the path positions:

Initiation: $P(1,m) = I(1,m) \qquad \forall m: 1 \le m \le M$

Recursion:

$$P(n,m) = I(n,m) + \max_{m-1 \le k \le m+1} P(n-1,k) \quad \forall n: 2 \le n \le N;\ \forall m: 1 \le m \le M \tag{2.1.1}$$

$$P^A(n,m) = argmax_{m-1 \le k \le m+1} P(n-1,k) \quad \forall n: 2 \le n \le N;\ \forall m: 1 \le m \le M$$

To estimate the backbone path positions P^O the maximum of P in column N is chosen and backtracked using the auxiliary variable P^A.

Initiation: $P^O(N) = argmax_{1 \le k \le M} P(N,k)$

Recursion: $P^O(n) = P^A\big(n+1, P^O(n+1)\big) \qquad with\ n = N-1, N-2, \dots, 1$ (2.1.2)

We apply this algorithm to both sides of the equator, i.e. 20°N to 0° for the northern ITCZ and 0° to 20°S for the southern ITCZ. Fig. 2.2 shows a typical result from this algorithm, illus-

trating the estimated backbone paths (and corresponding ITCZ locations) obtained[2] from VS data.

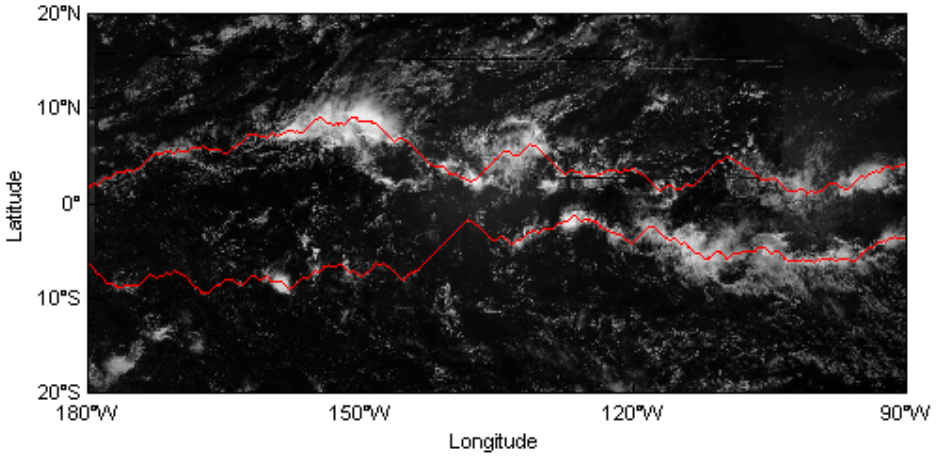


Fig. 2.2 – The red line indicates the location of the estimated backbone path (in both the northern and southern hemispheres), and correspondingly, the estimate of the dITCZ location. From VS data.

Given the estimated path we extract the spatial extent of the ITCZ as follows. Morphological reconstruction by dilation is used to find a spatial segmentation. To avoid the computationally-intensive process of iterative geodesic dilation we use a downhill filter which relies on a random access queue and is initialized via seed points defined by the backbone path – this approach was suggested by Robinson and Whelan (2004) in the context of medical image analysis. The spatial segmentation algorithm is as follows:

Initialize:

The mask image C is defined by the original satellite image

$$I(p) = \begin{cases} C(p) & if\ p\ is\ a\ pixel\ of\ the\ backbone\ path \\ 0 & otherwise \end{cases} \tag{2.2.1}$$

The access list L is defined as: $\forall p: I(p) \neq 0:$ Append p *to* $L(I(p))$

The auxiliary variable m is set to the maximum value of I.

Loop over all $n = m \ldots 1$:

While $L(n) \neq \{\}$ do

Take first element p out of $L(n)$ and do $\forall q \in$ Neighborhood(p) and q not already visited: (2.2.2)

$I(q) = \min(n, C(q))$ and append q to $L(I(q))$

After the loop is completed the pixels in I belonging to the ITCZ signal are preserved while non-ITCZ pixels are suppressed (see Appendix A).

This type of region-growing algorithm tends to preserve the main cloud band of the ITCZ while simultaneously suppressing cloud formations which are not connected to the ITCZ complex. The resulting image I is thresholded to produce a mask – this mask corresponds to our estimate of the spatial location and extent of the ITCZ (see Appendix A). The method is not par-

[2] When we display or refer to IR images in what follows we use inverted IR images (i.e., bright pixels correspond to lower intensity values) since this improves readability and understanding.

ticularly sensitive to the choice of threshold as long as it is above a minimum value - a value of 60 (using a scaled VS image with a reflectivity range [0…255]) was used in the experiments. Finally, we use morphologic open and close operations to connect nearby segments and to filter out small regions. For the close operation we use an elliptically shaped mask with an east-west orientation of the principal axis to favor elongated east-west oriented cloud bands (see Appendix A for details). The method is applicable to both the VS and inverted IR images. Fig. 2.3 illustrates the estimated mask for the example from Fig. 2.2.

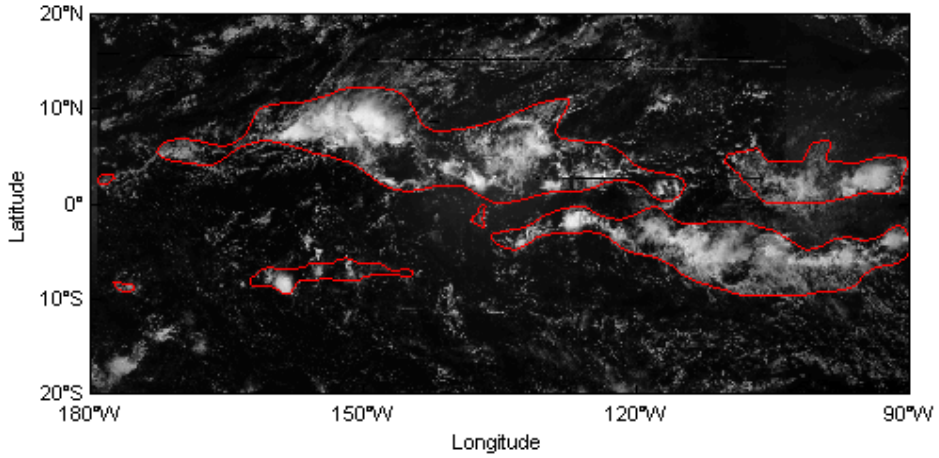


Fig. 2.3 – Result from automatic spatial segmentation of the ITCZ corresponding to the image in Fig. 2.2.

2.2.3 Time series analysis of the dITCZ phenomena

The second primary challenge in analyzing the dITCZ phenomena is to track and classify ITCZ presence over time in both the northern and southern hemisphere. Our approach to this temporal classification problem is to first estimate features from the satellite images, and to then use these features as observations for a hidden state-space model, from which we can estimate the most probable state sequence, i.e. presence or non-presence over time of ITCZ north and south of the equator. As mentioned earlier, and as illustrated in Fig. 2.1, we distinguish between four categorical and mutually-exclusive states:

- The dITCZ state where an ITCZ signal is visible (consisting of at least a roughly 90° longitudinal stretch of a connected cloud band[3]) on both sides of the equator;
- The northern state, where only one ITCZ is formed in the northern hemisphere (nITCZ);
- The southern state, where only one ITCZ is formed in the southern hemisphere (sITCZ);
- The state of non-presence, i.e. no significant ITCZ signal in the area of interest.

In preliminary experiments (not described here) we found that the 4-state model described in this paper outperformed two independent 2-state models that analyze the northern and southern ITCZ separately. This is suggestive of correlations in the dITCZ phenomenon which we will comment on again later in the paper. In what follows below, we will first describe the feature extraction and then give a detailed overview of the temporal classification methods.

[3] Note: A 90° longitudinal connected cloud band is our working definition of the ITCZ being present but the interpretation of a "connected cloud band" is subjective. This subjectivity naturally leads to variations among the labels assigned by the atmospheric scientists.

2.2.3.1 Feature extraction

We use a large number of different image-based features in our temporal classification approach (34 for VS and 37 for IR) - here we present only an overview - the complete list of features is defined in Appendix B. The features can be sub-divided into three different classes: those based on the backbone path, those based on the segmentation mask, and those based on general image statistics. Features are calculated separately for the northern part (20° N to 0°), the southern part (0° to 20° S), or in some cases for the complete image (20° N to 20° S). The features are calculated in the same manner for the VS and IR images, the only difference being in the threshold used for the segmentation mask.

Features based on the location of the backbone path

We extract features from the backbone path such as the slope or spatial variance of the path with respect to a linear best-fit regression. We find that this variance is typically lower when the ITCZ is present due to the typical elongated linear form of an ITCZ. Additionally, we derive statistics from the pixel intensities along the backbone path. For example, the mean of the intensity values as well as the distance of local minima from each other tend to be higher when the ITCZ is present since the ITCZ tends to be associated with elongated high-intensity cloud cover.

Shape-based features

Shape-based features are derived from the binary mask that is calculated by morphological reconstruction described in subsection 2.2.2. We use border tracing as described in Klette and Rosenfeld (2004) to extract all connected objects. Subsequent features that can be computed include for example the width of the object that has the largest east-west extent. This width is often quite large when the ITCZ is present relative to images without the ITCZ. For the object with the widest east-west extent additional characteristics such as area, perimeter, center of mass and compactness are also computed and used as features.

Shape-based features are also derived from Fourier descriptors (Arbter et al., 1990) computed on the border. The Fourier coefficient c_0 defines the balance point of the border line. Using the Fourier coefficient c_1, and the corresponding negative coefficient c_{-1}, an ellipse is fitted to the extracted object, as described in Appendix B. The major axis (which for the ITCZ is related to the maximal east-west width), the minor axis, and its ratio and orientation all provide useful features. The higher order Fourier coefficients are indicative of the smoothness of the object's border and thus the energy spectrum of the Fourier coefficients also serves as an additional feature.

Features based on general image statistics

As basic features we calculate the mean intensity of the complete image, for both the northern and southern parts of the image separately, as well as for the areas 12.5° N to 2.5°N and 2.5° S to 12.5° S (as defined for the double ITCZ phenomenon over the western Pacific in Chen et al., 2008). Central and invariant moments of the intensity distribution can also provide useful information and a number of such moments are used as features.

2.2.3.2 Temporal classification of the dITCZ phenomena with hidden state space models

Our aim is to classify each time snapshot into one of four states, conditioned on the observed features. Since the features are somewhat noisy, we use a temporal classifier to model the persistence of the dITCZ states. Specifically, we use a hidden Markov model (HMM) to incorporate knowledge about the temporal behavior of the dITCZ phenomena. In an HMM the probability distribution over the hidden state of the system at a particular time depends only on (a) the previous state of the system (the Markov property) and (b) the current observation.

As described in Rabiner (1989) an HMM is formally defined as

(1) a set of N discrete states $\{q1, q2, \dots, qN\}$

(2) a state transition matrix $\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix}$ with $a_{ij} = P(q_t = j | q_{t-1} = i)$ being the probability that state j follows state i under the conditions $a_{ij} \geq 0$ and $\forall i: \sum_{j=1}^{n} a_{ij} = 1$.

(3) a set of emission probability distributions or densities: $B=\{b_1,\dots,b_n\}$; where $b_i(\boldsymbol{x}) = P(\boldsymbol{o_t} = \boldsymbol{x} | q_t = i)$ is the probability of generating observation $\boldsymbol{x}$ in state i.

(4) a feature or observation space O which is the domain of the b_i's.

(5) an initial probability distribution $\boldsymbol{\pi}$ where $\pi(i)$ is the probability of starting in state i.

For modeling the dITCZ phenomena the states consist of {dITCZ, nITCZ, sITCZ, non-presence} and the observations $\boldsymbol{O}$ correspond to the set of features described earlier.

In a typical application of the HMM approach, the model parameters $\boldsymbol{\theta}=(\boldsymbol{A},B,\boldsymbol{\pi})$ are learned in an unsupervised manner by maximizing the probability $P(\boldsymbol{O}|\boldsymbol{\theta})$ for a given realization $\boldsymbol{O}=\{\boldsymbol{o_1},\dots,\boldsymbol{o_T}\}$ (the training data) via the Baum-Welch algorithm, which is a specific case of the more general expectation maximization (EM) procedure (e.g., see Rabiner, 1989). For a new set of images whose states are unknown, given the features of the image sequence $\boldsymbol{O}$ and the trained model parameters $\boldsymbol{\theta}$, the computation of the optimal state sequence $\boldsymbol{Q}=\{q_1,q_2,\dots,q_T\}$ – here, the temporal tracking of the dITCZ phenomena – is carried out using the Viterbi-algorithm (based on dynamic programming similar to the optimal path algorithm in Eq. (2.1) of subsection 2.2.2) by maximizing the probability $P(\boldsymbol{Q}|\boldsymbol{O},\boldsymbol{\theta})$.

To improve the performance of the model for the purpose of dITCZ tracking we make three changes to the standard HMM framework. These modifications have been investigated independently in the past and proven to be useful in general but to our knowledge have never been used in combination:

We begin by relaxing the strict Markov property for the transition matrix $\boldsymbol{A}$ by introducing a time-dependent self-transition function to the transition matrix, also known as a hidden semi-Markov model (HsMM) (e.g. Murphy, 2002). This modification allows for explicit modeling of the distribution of state durations.

Second, instead of modeling the emission probabilities B via a conditional (generative) model of the probability of the *observations given the states* (e.g., via a Gaussian distribution) we instead model the conditional distribution of the *states given the observations* using support vector machines (SVMs). The latter approach is generally less sensitive to parametric distributional assumptions than the former, particularly in high-dimensional observation spaces. The general approach is discussed for example in Bourlard and Morgan (1994) – here we specifically use SVMs and logistic regression (as in Platt, 1999) to model the conditional distribution of states given observations, as described later in this paper.

Finally, since training on known expert labels can improve the model's performance significantly (compared to unsupervised learning without any labeled data), we replace unsupervised parameter estimation with both supervised (all labels known) or semi-supervised (partly known and partly unknown labels) learning, as described for example by Zhong (2005).

Relaxing the strict Markov property: the HsMM

The standard HMM implies a geometric distribution for the state transition probability due to the constant self-transition coefficients (the a_{ii}'s in the transition matrix $\boldsymbol{A}$). The state duration probability density $pi(d)$ for staying in state i for exactly d time-steps is

$$p_i(d) = (a_{ii})^{d-1} * (1 - a_{ii}) \tag{2.3}$$

This geometric distribution is inconsistent with prior knowledge about ITCZ persistence. For example, in Wang and Magnusdottir (2006), where the durations of northern ITCZs in the summer months were manually identified for the years 1999-2003, typical durations were found to be 9 days in length and non-geometric in terms of distribution. Furthermore, based on the human-generated labels used in this present study, we found that the state duration for the ITCZ in either hemisphere tends to have a mode that is centered between 5 and 15 days, rather than at 0 days as required by a geometric distribution. For these reasons we model the state duration explicitly by introducing a time-dependent component in the state transition matrix $\boldsymbol{A}$.

Let L be a contiguous (in time) labeled training sequence obtained from a human labeler. We define $d_k(i)$ as the duration of the k^{th} connected occurrence of state i in L. The state duration distribution $D_i(t)$ over time for each state can be calculated as

$$\begin{aligned} &D_i(t) = \sum_k 1\,(d_k(i) \le t); \\ &with\ 1(d_k(i) \le t) = \begin{cases} 1 & if\ d_k(i) \le t \\ 0 & else \end{cases} \quad \forall t\colon 1 \le t \le \max\,(d_k(i)) \end{aligned} \tag{2.4.1}$$

We can approximate the state duration probability by fitting a logistic regression model of the following form:

$$D_i(t) \approx \frac{e^{\beta_0+\beta_1 t}}{1 + e^{\beta_0+\beta_1 t}} \tag{2.4.2}$$

where the β_k's are estimated by maximum likelihood estimation (MLE). Since the training data for learning the state duration is relatively small, outliers can have a severe impact on the fitting of the β_k coefficients. To overcome this we use a scaled, interpolated and smoothed version of $D_i(t)$ as a look-up table for finite amounts of training data, recovering the logistic regression form for $t \gg \max\,(d_k(i))$. Additional details, including an example of such fitted duration models, are provided in Appendix C.

The constant a_{ii}'s in the transition matrix $\boldsymbol{A}$ can now be replaced by the time-dependent $a_{ii}(t)$'s which are defined as the scaled, interpolated and smoothed version of the $D_i(t)$ values described above. To satisfy the criteria $\forall i\colon \sum_{j=1}^{n} a_{ij} = 1$ for each row of $A(t)$, we normalize the a_{ij} with $i \neq j$.

As mentioned earlier, the training data for learning the state duration is often relatively small since there are a limited number of ITCZ transition events per season. Thus, we found it useful to use a transition matrix that is a combination of purely dynamic and static transition matrices, i.e., $\boldsymbol{A}^{new}(t) = \lambda * \boldsymbol{A}^{dynamic}(t) + (1-\lambda) * \boldsymbol{A}^{static}$ where λ is a weighting factor. $\boldsymbol{A}^{static}$ is more robust since it depends on fewer parameters than $\boldsymbol{A}^{dynamic}(t)$. As the number of training samples increases, $\boldsymbol{A}^{dynamic}(t)$ contributes more information, and the weighting factor λ can be increased. In our dITCZ experiments, for the supervised VS model with 1 image per day, and relatively little training data, we rely mainly on the static matrix with λ=0.1. For the semi-supervised IR model with 8 images per day (thus 8 times as much training data) we use a purely dynamic transition matrix (λ=1). Furthermore, as discussed below, λ=1 is also necessary for robust semi-supervised learning. Finally, we use modified forms of the alpha-beta and Viterbi-algorithms by introducing an auxiliary variable that keeps track of the time spent in the current state making it possible to update the self-transition probability of $\boldsymbol{A}^{dynamic}(t)$ at a given time step t in the classification step.

Modeling the state-observation probabilities by SVMs: the HsMM/SVM hybrid

In the typical application of HMMs the bi's are defined as $b_i(\boldsymbol{x}) = P(\boldsymbol{o}_t = \boldsymbol{x} \mid q_t = i)$, the probability of generating observation $\boldsymbol{x}$ in state i. However, common parametric forms for the bi's (such as the multivariate Gaussian distribution) are often not well-matched to real data distribu-

tions. As discussed earlier, the alternative parametrization we pursue here is to model $P(q_t = i \mid \boldsymbol{o_t} = \boldsymbol{x})$ by an SVM or by logistic regression. The necessary conditional probabilities for learning and classification for this parametrization can be derived using Bayes' Rule:

$$P(\boldsymbol{x}|q_t = i) = \frac{P(q_t = i|\boldsymbol{x}) * P(\boldsymbol{x})}{P(q_t = i)} \tag{2.5}$$

$P(q_t = i)$ can be estimated via the labeled training set and $P(\boldsymbol{x})$ is a scaling factor since it is independent of the state s. It can be shown that the parameters of the H(s)MM are invariant with respect to scaling and thus one can convert any classifier which satisfies the maximum a posteriori (MAP) criterion (i.e., it estimates $P(q_t = i \mid \boldsymbol{x})$) into an estimator of the emission densities B of the HsMM (e.g. see Bourlard & Morgan, 1994).

When using SVMs some additional adjustments need to be made to implement this parametrization. Firstly, standard binary SVMs can only discriminate between two classes. Since we deal with four different classes (dITCZ, northern, southern, not present), we use "one-versus-all" SVMs for all classes, i.e., we use four binary classifiers, one per state. Given our labeled state sequence L, we set up the l training data pairs $(\boldsymbol{x_t}, y_{i,t})$ for the SVM as follows:

$$y_{i,t} = \begin{cases} 1 & if\ the\ state\ at\ time\ t\ is\ i \\ -1 & otherwise \end{cases} \quad \text{with a corresponding feature vector } \boldsymbol{x_t}. \tag{2.6.1}$$

With these l training data pairs the support vectors are learned by maximizing:

$$L(\alpha) = \sum_{k=1}^{l} \alpha_k - \frac{1}{2}\sum_{k=1}^{l}\sum_{m=1}^{l} y_k y_m \alpha_k \alpha_m \langle \boldsymbol{x_k}, \boldsymbol{x_m} \rangle \tag{2.6.2}$$
under the constraints $0 \le \alpha_k \le C$ and $\sum_{k=1}^{l} y_k \alpha_k = 0$

The cost parameter, C, controls the trade-off between allowing training errors and forcing rigid margins, and was set to a fixed default value $C = 1$ in all of our experiments.

Given the α's, the hyperplane parameters of the SVM are defined as:

$$\boldsymbol{w} = \sum_{k=1}^{l} \alpha_k y_k \boldsymbol{x_k} \text{ and } u = -\frac{1}{2}(\max_{k,y_k=-1}(\langle \boldsymbol{w}, \boldsymbol{x_k} \rangle) + \min_{k,y_k=+1}(\langle \boldsymbol{w}, \boldsymbol{x_k} \rangle)) \tag{2.6.3}$$

A second issue is that SVMs only calculate the distance of a new sample $\boldsymbol{x_{new}}$ to the optimal hyperplane $< \boldsymbol{w_i}, \boldsymbol{x_{new}} > + u_i$ but do not directly compute the probability $P(q = i \mid \boldsymbol{x})$. However, the distance to the optimal hyperplane implies a measurement of certainty about the classification decision (small distances imply low certainty, large distances imply high certainty). One can approximate the probability of $P(y_{i,t} = 1 \mid \boldsymbol{x_t})$ by a sigmoid function (Platt, 1999):

$$P(y_{i,t} = 1|\boldsymbol{x_t}) \approx \frac{1}{1 + e^{-\varepsilon_i * (<\boldsymbol{w_i}, x_t> + u_i) + \gamma_i}} \tag{2.6.4}$$

where $\boldsymbol{w_i}$ and u_i are the parameters learned by the SVM and ε_i and γ_i are estimated iteratively with respect to the training data. To fit the SVM models in our experiments we used a non-linear least square SVM (LS-SVM) with a radial basis function (RBF) kernel, as implemented in the LS-SVM MATLAB toolbox by De Brabanter et al. (2010). The width parameter of the RBF was set to 110 based on preliminary experiments with the dITCZ training data.

Supervised and semi-supervised learning approaches

For learning the model parameters $\boldsymbol{\theta}$ we use both supervised and semi-supervised approaches depending on the particular data set used for the classification. Both data sets, VS and IR, have their advantages and disadvantages. The VS imagery (which is also the main source for the atmospheric scientists to visually detect the ITCZ signal) achieves the better results for classifying the dITCZ phenomena. One of the main reasons for this is that the ITCZ phenomena in the east Pacific in spring (Feb-May) can be weak, and thus, there are numerous shallow ITCZ events

which are only detectable in VS data but are not seen in IR images. Including IR features in the VS model downgrades the model's accuracy. These facts in general favor VS data over IR data. On the other hand, IR data are available at a higher temporal resolution (8 per day vs. only 1 per day for VS), and more importantly, the time series for analyzing the long-term behavior is more reliable for the IR data since VS data are less temporally homogeneous. Furthermore, the VS data prior to 1995 can be corrupt or non-existent as discussed in previous investigations (e.g. Bain et al., 2011).

These differences between IR and VS motivated our investigation of two different learning approaches. For the VS model with only daily images, we have labels for each image and thus, the model parameters $\boldsymbol{\theta}$ can be learned in a completely supervised manner, i.e. we have an expert labeled sequence $\boldsymbol{L}$ (which yields by definition a most probable (with probability 1) state sequence $\boldsymbol{Q}$) and an associated observation sequence $\boldsymbol{O}$. $\boldsymbol{A}$ and $\boldsymbol{\pi}$ can be estimated directly from $\boldsymbol{L}$, and B can be trained by the SVMs using the ($\boldsymbol{O}$,$\boldsymbol{L}$)- tuples.

In contrast, the IR images are available every 3 hours, i.e. 8 per day. In principle, these additional images (7 of which are unlabeled) provide the opportunity to improve the model's performance via a semi-supervised learning approach as follows. We make the reasonable assumption that the state duration for a specific ITCZ phase spans at least 1 day. We enforce this by making the self-transition duration of Eq. (2.4.1) be 1 for the first 8 time steps (8 IR images per day): $\forall i, 0 < t < 8: D_i(t) = 1$ and we use a purely dynamic transition matrix, i.e. λ=1. Given an observation sequence $\boldsymbol{O} = \{\boldsymbol{o_1},\boldsymbol{o_2}...,\boldsymbol{o_N}\}$ and the corresponding labels $\boldsymbol{L^{known}} = \{l_{8*k+m}\}$ with $0 \leq 8 * k \leq N - m, 0 < m < 8$, we wish to maximize $P(\boldsymbol{O}|\boldsymbol{\theta})$ under the constraint that the state sequence $\boldsymbol{Q} = \{q_1, q_2, \ldots, q_N\}$ at times $8*k+m$ equals exactly the supervised labels $\boldsymbol{L^{known}}$, i.e. $\forall k: l_{8*k+m} = q_{8*k+m}$. We first initialize our model parameters $\boldsymbol{\theta} = (\boldsymbol{A},\boldsymbol{\pi},\boldsymbol{w},\boldsymbol{u})$ as follows:

- Train parameters $\boldsymbol{w_i}$ and u_i of the SVMs by using the labels $\boldsymbol{L^{known}}$ and the corresponding observation sub-set $\boldsymbol{O} = \{\boldsymbol{o_{8*k+m}}\}$ with $0 \leq 8 * k \leq N - m, 0 < m < 8$ as for the VS model (Eq. (2.6)).
- Define new labels $\boldsymbol{L^{r=\theta}}$ for the missing labels via the following simple heuristic: between two labels l_{8*k+m} and $l_{8*(k+1)+m}$, where $l_{8*k+m} == l_{8*(k+1)+m}$, set all labels between, i.e. $l_{8*k+m+1}$, $l_{8*k+m+2}, \ldots, l_{8*k+m+7}$, equal to l_{8*k+m}. If a state transition exists between two labels, i.e. $l_{8*k+m} \neq l_{8*(k+1)+m}$, we set all values $l_{8*k+m+1}, \ldots, l_{8*k+m+4}$ equal to l_{8*k+m} and all labels $l_{8*k+m+5}, \ldots, l_{8*k+m+7}$ equal to $l_{8*(k+1)+m}$:

$$\begin{aligned} &For\ j = \{1,2\ 3,4\}: l_{8*k+m+j} = l_{8*k+m} \\ &For\ j = \{5,6,7\}: \quad l_{8*k+m+j} = l_{8*(k+1)+m} \end{aligned} \tag{2.7}$$

- Finally, the dynamic transition matrix $\boldsymbol{A}$, derived from $\boldsymbol{L^{\theta}}$ and Eq. (2.4), as well as the prior distribution $\boldsymbol{\pi}$, are learned in a manner equivalent to the VS model case, except for forcing the minimum self-transition time to be at least 8 steps.

We use a quasi-EM algorithm for semi-supervised learning of HsMM/SVM models on the IR data. The aim is to obtain the optimal model parameters $\boldsymbol{\theta}^*$ by maximizing the marginal likelihood function $\mathcal{L}(\boldsymbol{\theta};\boldsymbol{O}) = P(\boldsymbol{O}|\boldsymbol{\theta}) = \sum_{\forall \boldsymbol{Q}} P(\boldsymbol{O},\boldsymbol{Q}|\boldsymbol{\theta})$ given the observation sequence $\boldsymbol{O}$: $\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} P(\boldsymbol{O}|\boldsymbol{\theta})$. For our semi-supervised approach we include the additional constraint that the state sequence $\boldsymbol{Q}$ has to traverse the given labels $\boldsymbol{L^{known}}$. Since there is no closed analytic solution to the problem, we solve it iteratively. In the estimation (E) step of the EM algorithm, the expected value of the log likelihood function under the current estimate of the parameters $\boldsymbol{\theta^r}$ is calculated with respect to the conditional distribution $\boldsymbol{Q}$ given $\boldsymbol{O}$ using the following auxiliary function: $Aux(\boldsymbol{\theta}|\boldsymbol{\theta^r}) = E_{\boldsymbol{Q}}[\log P(\boldsymbol{O},\boldsymbol{Q}|\boldsymbol{\theta})|\boldsymbol{O},\boldsymbol{\theta^r}]$. For our model this is simply equivalent to computing the most probable state sequence $\boldsymbol{Q^r}$ given $\boldsymbol{O}$ and $\boldsymbol{\theta^r}$ under the constraint that $\boldsymbol{Q^r}$ traverses $\boldsymbol{L^{known}}$. In the maximization (M) step we seek the parameters maximizing the auxiliary function: $\boldsymbol{\theta^{r+1}} = \arg\max_{\boldsymbol{\theta}} Aux(\boldsymbol{\theta}|\boldsymbol{\theta^r})$. Thus, the estimate of the parameters $\boldsymbol{\theta^{r+1}}$ is updated according to the state sequence $\boldsymbol{Q^r}$ of the E-step and the observation sequence $\boldsymbol{O}$.

A more detailed description of the algorithm can be found in appendix D.

We repeat the EM-algorithm until no changes between the labels at r and $r+1$ occur, i.e. $\boldsymbol{L}^r == \boldsymbol{L}^{r+1}$, and thus, convergence is reached. It is possible (although it did not happen in any of our experiments) that the algorithm might not converge – to avoid this one can define a maximum number of iterations and then select the maximum over all parameter estimates $\boldsymbol{\Theta} = \{\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^R\}$, thus, guaranteeing that the semi-supervised case has at least as high a likelihood as the supervised IR model: $P\left(\boldsymbol{O}\middle|\boldsymbol{\theta}^{\boldsymbol{semi-supervised}}\right) \geq P\left(\boldsymbol{O}\middle|\boldsymbol{\theta}^{\boldsymbol{supervised}}\right)$.

Since we use a semi-supervised SVM/HsMM hybrid approach the strict EM-criterion, that $P\left(\boldsymbol{O}\middle|\boldsymbol{\theta}^{\boldsymbol{r+1}}\right) \geq P(\boldsymbol{O}|\boldsymbol{\theta}^{\boldsymbol{r}})$, is not necessarily guaranteed - but the method appears to work well empirically as illustrated below.

2.3 Experimental Results

We applied the methods described in section 2.2 to the VS and IR satellite data and discuss below the results in terms of model performance relative to human labeling. We also discuss a number of insights gained about the temporal behavior of the dITCZ phenomena.

2.3.1 Performance of the models

As described in section 2.2.1, to train and evaluate the models, four atmospheric scientists (Experts I-IV) independently labeled a sequence of satellite images using VS, IR and TPW images of the area of interest for each day. Two of the scientists labeled two seasons (Feb-May of 2000 and 2002, 241 days total). The other two scientists labeled 31 days in March 2000. The labels of Expert I were used for training and validating the model. The three additional expert labels provided information about typical subjective variability across human experts for the ITCZ-labeling task.

Table 2.1 shows the distribution of the fraction of days per label for the four disjunctive dITCZ states, using the labels of Expert I. We see that the nITCZ state dominates (55.6% of days).

TABLE 2.1
ITCZ STATE DISTRIBUTION OF EXPERT I

	dITCZ	nITCZ	sITCZ	not present
Expert I (241 days labeled)	27.8% (67 days)	55.6% (134 days)	11.6% (28 days)	5.0% (12 days)

Table 2.2 shows the classification accuracy of the labels of Experts II, III, and IV, relative to Expert I, i.e., treating Expert I's labels as ground truth. The accuracy ranges from 74% to 81%, indicating that there is a relatively high degree of subjectivity for this task. Measuring the inter-labeler accuracy is useful because it allows us to calibrate an automated algorithm's performance relative to human performance on the same task.

TABLE 2.2
CLASSIFICATION ACCURACY OF HUMAN EXPERTS

	Expert II	Expert III	Expert IV
Expert I	77.2% (241 common days)	80.7% (31 common days)	74.2% (31 common days)

To evaluate the performance of our models, we divided the images (and labels from Expert 1) into a training and test set of equal size, such that the distribution of state occurrences in the training set and test set is approximately the same. The training set used for the statistics in the paper comprises 121 days (dITCZ: 37 (30.6%), nITCZ: 60 (49.6%), sITCZ: 17 (14.0%), no ITCZ: 7 (5.8%)) the test set 120 days (dITCZ: 30 (25.0%), nITCZ: 74 (61.7%), sITCZ: 11 (9.2%), no ITCZ: 5 (4.2%)).[4]

The performance of the VS model on the test data is shown in Table 2.3. The model achieved an overall classification accuracy of 84.2%, broken down into classification errors between specific pairs of states via the confusion matrix. The majority of errors (10 days) occurred when the model predicted an nITCZ and the expert labeled it as a dITCZ. The model's classification accuracy of 84.2% is comparable with that of human experts among each other (Table 2.2). That the model accuracy is slightly higher than that of the best expert (the best human accuracy was 80.7%) could be explained by the fact that the model was trained on labels from Expert I, and thus, the model more closely mimics the labeling behavior of Expert I compared to the other experts who may be using slightly different heuristics for image labeling.

TABLE 2.3
CONFUSION MATRIX OF THE VS MODEL ON THE TEST SET

----		Expert I			
		Double	Northern	Southern	None
Model	Double	20	2	3	2
	Northern	10	71	0	1
	Southern	0	0	8	0
	None	0	1	0	2
Classification accuracy		101 of 120 (84.2%)			

It is also informative to compare the full VS model (using the HsMM/SVM approach, with all features) to simpler variants. Table 2.4Table 2.4 shows the classification accuracies for various models trained on Expert I's labels, using the same training and test setup as in Table 2.3. The closest-performing model is an SVM-only model, which classifies each day separately without any temporal (HsMM) component – it is 5% less accurate overall than the HsMM/SVM approach, indicating that including temporal information in the classification approach leads to higher accuracy. Two approaches that each only use a single feature achieve accuracies of 70% and 73%, indicating that the full set of features provide an improvement of 11 to 14% in accuracy. Finally, the standard HMM approach with a multivariate Gaussian for modeling observations given states (using a full covariance matrix) is about 12% less accurate than the HsMM/SVM method.

[4] The partitioning of the training and test set is not straightforward, partially caused by the relatively low amount of labeled data. To robustly learn the parameters of the proposed model, the training set has to be temporally contiguous (to train the transition matrix of the HMM) and have enough days for each state to have sufficient training samples for the SVMs of the four different ITCZ states. Only few of such train/test splits exist, and these are shifted only by a few days from each other and consequently highly correlated. These different train/test set partitions with similar state distributions have little influence on the model's performance (varying not more than +/-1.5%).

TABLE 2.4
CLASSIFICATION ACCURACY OF DIFFERENT MODELS

HsMM/SVM as in paper	84.2%
SVM only	79.2%
HsMM/SVM only with the east-west extent feature for the longest object in the northern and southern image (see appendix)	73.3%
Standard HMM	72.5%
HsMM/SVM only with the average of 12.5°N to 2°N and 2°S to 12.5° S feature (see appendix)	70.0%

We focus next on the performance of IR-based models, using the same training and test data from Expert I as used for the VS model. It is important to keep in mind that Expert I considered VS and TPW images as well as IR images when determining the ITCZ state but the IR model only considers IR data. This means that the expert can identify regions of shallow convection while the model cannot. Therefore we expect a decrease in accuracy when the IR model is compared to Expert I's ground truth, particularly when there are weak ITCZ cases present that are visible to the human eye in VS and TPW, but that are not visible in IR imagery alone. The best performance was obtained with a semi-supervised IR model - its performance is 75.8% in terms of classification accuracy on the daily time-scale (Table 2.5) - this is 8.4% lower than the accuracy of 84.2% obtained by VS model. It is reasonable to assume that the IR model would be more 'accurate' if it were compared to a human expert who identified ITCZ states based only on IR images (see also Table 2.8). Despite this lower accuracy, given that the IR data records are available for a much longer time-span than VS data, and have fewer data quality issues, the use of the IR model is worth exploring. Furthermore, results from the IR model provide times when a deep convective ITCZ is present, adding additional useful information to the analysis.

TABLE 2.5
CONFUSION MATRIX OF THE IR MODEL ON THE TEST SET

----		Expert I			
		Double	Northern	Southern	None
Model	Double	9	0	1	1
	Northern	14	71	1	1
	Southern	2	0	8	0
	None	5	3	1	3
Classification accuracy		91 of 120 (75.8%)			

Table 2.6 shows how the semi-supervised HsMM/SVM IR model performed relative to other variants. As with the VS model, these other approaches have lower accuracy. In particular, the supervised HsMM/SVM approach (which does not take advantage of 7 of the 8 IR images available each day) is 5% less accurate than the semi-supervised approach, indicating that the addition of the unlabeled IR images leads to a more accurate classifier. We conjecture that the reason the SVM only model is more accurate than the version with an HMM is that the HMM propagates some of the incorrect label decisions of the SVM to neighboring states (especially for rare non-present case where experts labeled an ITCZ presence but IR only features indicate non-presence due to shallow ITCZ occurrences), increasing the error rather than reducing it.

TABLE 2.6
CLASSIFICATION ACCURACY OF DIFFERENT MODELS

Semi-supervised HsMM/SVM as in paper	75.8%
SVM only	72.5%
Supervised HsMM/SVM	70.8%
Standard HMM	65.8%

A useful feature of the semi-supervised approach is its ability to incorporate partially labeled data at prediction time (on the test data). For example, the IR model could be used as part of a semi-automated tool whereby a human user labels every k^{th} image and the model infers the rest of the labels. To illustrate this concept, we trained both a semi-supervised HsMM/SVM model and an SVM-only model on fully labeled data (at the daily scale), and then made predictions with each on partially labeled data (again at the daily scale). At prediction time the models were provided with the labels on a subset of images, and then predicted the labels for the rest. The predictions on the rest of the images were then compared to the actual human expert labels from Expert I for these images. The resulting accuracies on the predicted test images are shown in Table 2.7. The addition of partially labeled data significantly improves the accuracy of the HsMM/SVM approach, from 75.8% with fully unlabeled test data, to 80% with every 2nd day labeled – this increase is because the HsMM component of the model can "propagate" information from the labeled to the unlabeled images at prediction time. In contrast, the supervised SVM-only method cannot leverage these temporal dependencies, and its prediction accuracies remain largely the same (the variation across the columns for SVMs in Table 2.7 is due only to sampling noise in the training data, i.e., due to predictions being made on different subsets of unlabeled images in each case).

TABLE 2.7
PERFORMANCE IMPROVEMENT BY INCLUDING EXPERT LABELS FOR CLASSIFICATION
(ACCURACY ON UNLABELED DAYS)

----	No expert labels	Expert labels every 8th day	Expert labels every 2nd day
Semi-supervised HsMM/SVM	75.8%	76.0%	80.0%
SVM only	72.5%	71.2%	73.3%

As a final test of our methodology, we applied the supervised VS model and the semi-supervised IR model, both trained on Expert I's labels, to a completely unlabeled image sequence from 2001. We then provided the predictions from each model to Expert I for interpretation. The expert looked at the images using all three image fields (VS, IR, and TPW) and identified which classification decisions from each model were correct or incorrect in the context of the expert's visual inspection. The expert then repeated this analysis, but this time only using the IR images for visual reference, and identified which classifications from the IR model were correct or incorrect relative to visual inspection of the IR images. (We did not ask the expert to examine the VS model prediction relative to IR-only images). Table 2.8 shows the results. On this test, each of the models (VS and IR) were able to achieve 86.7% accuracy relative to Expert

I's inspection of the same data (VS or IR). These accuracies may be slightly optimistically biased since the human labeler was shown both the model predictions and the raw images at the same time (whereas previously the labeling was done independently of any model predictions), allowing for the possibility that the models' predictions could bias the human's decisions. Nonetheless the general trend is clear, namely that both VS and IR models can achieve relatively high accuracy (86.7%) when compared to a human looking at the same data that the algorithm has available. When comparing the expert's labels that include the additional information of VS and TPW data with the IR model (IR-only information) less ITCZ cases (11.7%) are detected because of shallow ITCZ cases that show no ITCZ signal in IR images.

TABLE 2.8
MODEL CLASSIFICATION ACCURACY FOR SEASON 2001 AS DETERMINED BY EXPERT I

------	VS model	IR model
Expert I; decision based on all data (VS, IR, TPW)	86.7%	75.0%
Expert I; decision based on IR only	------	86.7%

2.3.2 Using the model to study dITCZ climatology over five active seasons

For the purposes of this discussion we focus on the classification labels produced by the VS model over a relatively short illustrative period of 5 seasons (2000-2004). The statistics calculated here are mainly intended to demonstrate the potential of the model rather than an in-depth climatological analysis.

Fig. 2.4 shows the distribution of the occurrences of the ITCZ states from the model's predictions during the 2000-2004 time period. Red indicates dITCZ days, blue indicates nITCZ days, green indicates sITCZ days, and yellow indicates days with no ITCZ signal. Seasonal totals are shown in the last column and monthly totals (over all 5 years) are shown in the bottom row. Overall, the nITCZ is dominant during this period, accounting for 64% of the days. This is followed by dITCZ (27%), sITCZ (6%) and non-present days (3%).

The bottom row in Fig. 2.4 provides a general idea of the seasonal evolution of the east Pacific ITCZ through the boreal spring. In February, the dITCZ is present approximately 26% of the time, but this month is dominated by the nITCZ. By March, the occurrence of a dITCZ has increased to 49% and it is now more dominant than the nITCZ. The sITCZ is more present in March than February. The number of dITCZs decreases in April and, by May the ITCZ is located almost exclusively in the north (nITCZ). Based on these five seasons, the peak in dITCZ occurrence happens in March. When considered individually, all years follow this pattern except 2003. In this year the highest number of dITCZs occurs in February and this was the only year where a dITCZ was detected in May. In future studies it would be interesting to see if this general picture holds true or if there are more seasons similar to 2003. It would also be of interest to extend the analysis to January to see if any cases of dITCZ are detected earlier than February.

The last column in Fig. 2.4 gives an indication of interannual variability. The 2000 season is an outlier compared to the other years. While the number of dITCZ is similar to that in other years (especially 2001 and 2004) the number of nITCZs is greatly reduced and the number of sITCZs is greatly increased. This is the only year in which there are more dITCZs than nITCZs. The 2000 season also has the largest number of days when the ITCZ was labeled 'not present'. The 2003 season has the fewest dITCZs and sITCZs out of all five seasons. It is possible that the large number of dITCZs and sITCZs seen in 2000 (and maybe even 2001) are due to a

somewhat strong La Nina during the 1999/2000 winter. A longer time series analysis could better clarify this hypothesis and indicate the effects of El Nino - Southern Oscillation (ENSO) on the different states of the east Pacific ITCZ.

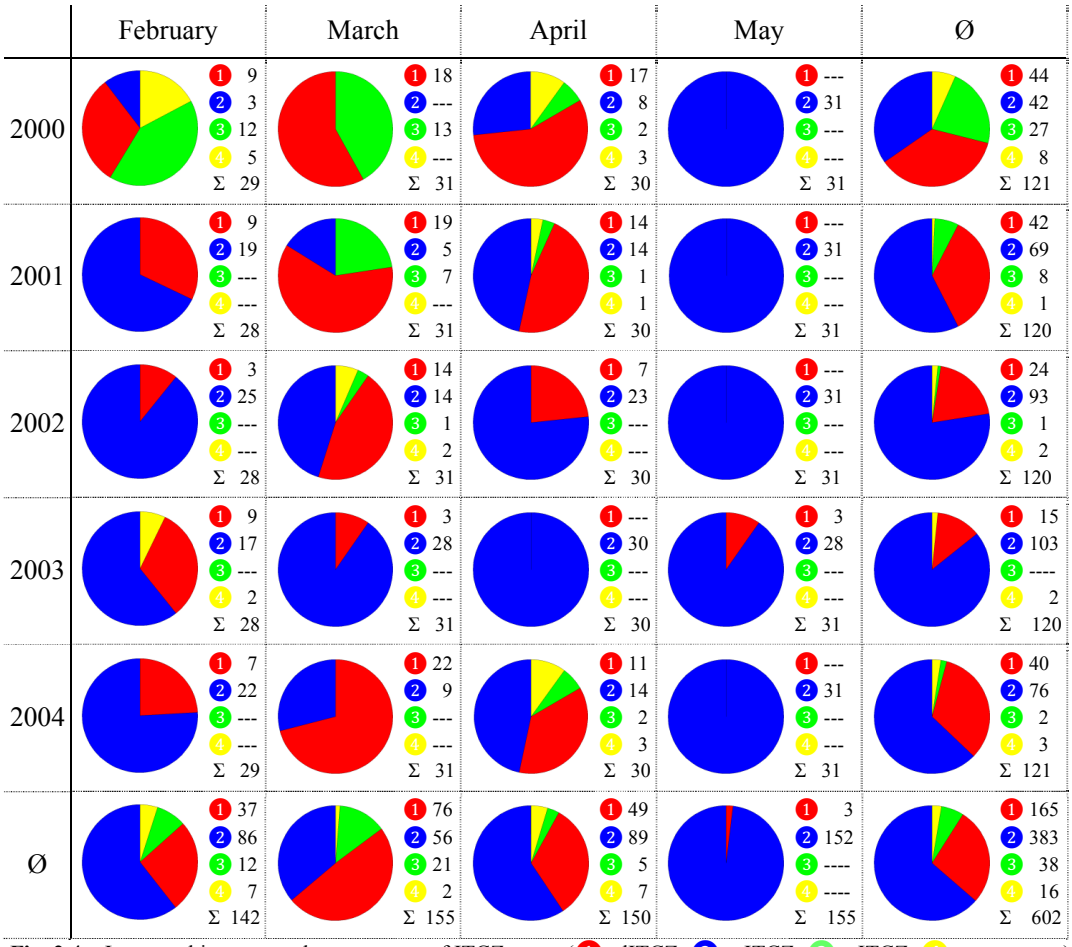


Fig. 2.4 – Intra- and inter-annual occurrences of ITCZ states (1: dITCZ; 2: nITCZ; 3: sITCZ; 4: not present). Values on the right of each chart indicate number of days for each ITCZ state.

Fig. 2.5 shows composite VS satellite images for each of the four states, showing the average of the VS intensity. The top left plot shows the dITCZ with two zonally elongated cloud bands on each side of the equator. From this composite plot one can observe a longitudinal offset between and northern and southern parts of the dITCZ such that the northern part appears to be located west of the southern part. The top right panel shows the typical nITCZ. The bottom left panel shows the sITCZ. It is interesting to note that the composite sITCZ shown here looks spatially different from the southern part of the dITCZ in the top left panel. This introduces the question of whether there is asymmetry in ITCZ structure in one hemisphere depending on presence or absence of a simultaneous ITCZ structure in the other hemisphere. Finally the bottom right panel shows the average VS field when no ITCZ is present. The output images are generally consistent with the limited prior climatological knowledge we have for each state, providing additional validation of the model.

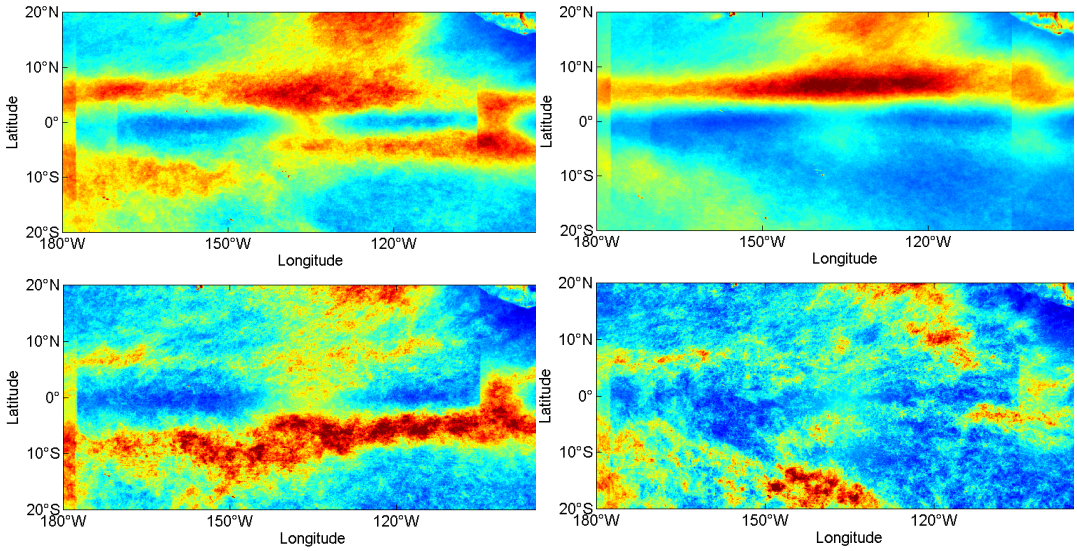


Fig. 2.5 – Average VS images for each dITCZ state: double, top left; north, top right; south, bottom left; no presence, bottom right

In Fig. 2.6 the average VS images after the morphologic reconstruction step are displayed, filtering out non-ITCZ clouds and emphasizing the ITCZ signal. In both Fig. 2.5 and Fig. 2.6 some vertical artifacts are visible at the location where measurements from two different GOES satellites are combined. This type of artifact has been cleaned up in the more consistent IR output in the GridSat dataset (Knapp et al., 2011). Suppressing such artifacts is not of direct relevance to the results presented; for the interested reader, additional information on this topic can e.g. be found in Minnis (1989) and Govaerts et al. (2008).

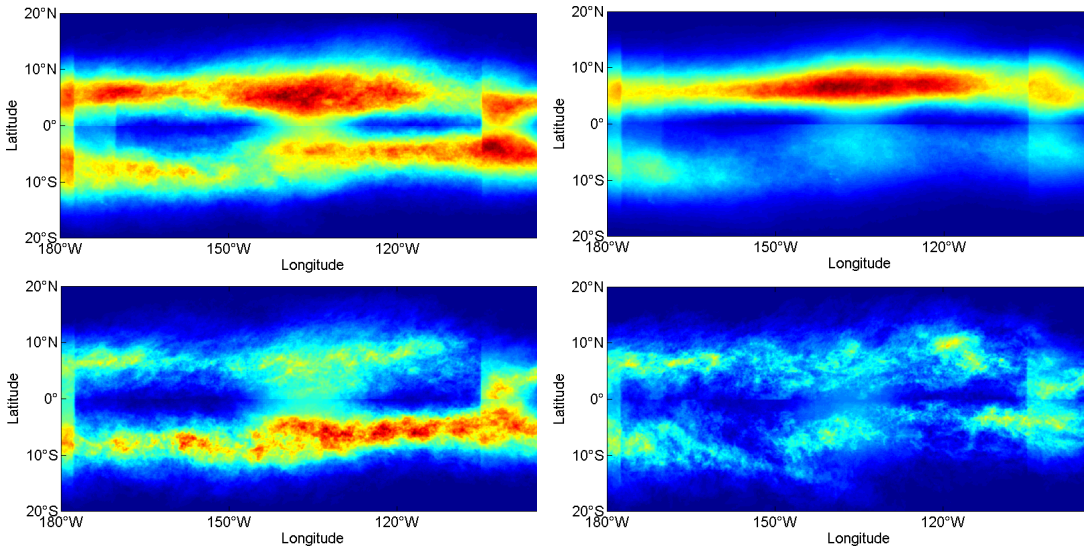


Fig. 2.6 – Average VS morphologic reconstructed images for each dITCZ state: double, top left; north, top right; south, bottom left; no presence, bottom right

Furthermore, we briefly demonstrate the benefits of the individual features discussed in section 2.2 by examining the geometry of the dITCZ, using the northern and southern centers of mass of dITCZ structures. Specifically, given the ITCZ states as classified by the model, we investigate if the shift between the northern and southern dITCZ, human experts observed while labeling the data (and rudimentary visible in the composite images of Fig. 2.5 & 2.Fig. 2.6, top

left), also appears in the feature statistics. Out of the 165 dITCZ days during the five seasons, in 109 cases (66%) the center of mass of the southern dITCZ was located to the east of the northern dITCZ (we define this as a 'positive offset') and the two centers were, on average, 23.4° apart in longitude. The other 56 cases (34%) have the southern part of dITCZ located to the west of the northern part with a 18.9° mean longitudinal difference in their centers of mass ('negative offset'). Fig. 2.7 illustrates the mean center of mass in each hemisphere for all of the dITCZ cases and the positive and negative offset cases. Note that the mean location of the northern dITCZ does not change drastically between the positive and negative offset cases (change in longitude: 10°). It is the southern dITCZ location that shifts widely between positive and negative offset cases (change in longitude: 32°).

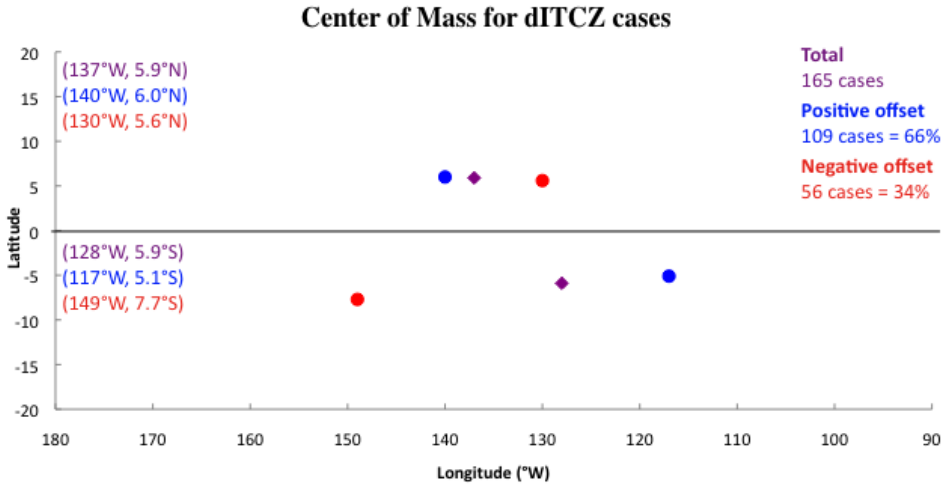


Fig. 2.7 – dITCZ cases for Feb-May, 2000-2004. Mean location of the centers of mass of the north part and south part of the dITCZ is dipicted for all cases (purple), for positive offset cases (blue, when north part is west of south part) and for negative offset cases (red, when north part is east of south part).

2.4 Conclusion

We have introduced an algorithm to extract the spatial location and expansion of the dITCZ phenomenon and to track its temporal behavior. For spatial segmentation, we developed a fully automatic unsupervised method based on the backbone path method. The extracted ITCZ regions are consistent and robust, and form – with additional derived features – the observation set for the temporal segmentation. The HsMM/SVM hybrid approach performed best in our experiments for this temporal task. We showed that the classification labels from this model are comparable in reliability to those from human experts. A VS model was developed that identifies ITCZ states even when there is only shallow convection present. A corresponding IR model performs the same task but only identifies the ITCZ states when deep convection is present. Both models proved very accurate (well within the range of human labeling variability) when compared to expert opinion using the appropriate field (either VS or IR).

The model's output for the period of 2000-2004 shows clear evidence that this method would be useful in classifying the east Pacific ITCZ states. The results suggest that it is worth conducting an in-depth long-term analysis of dITCZ phenomena, using both of the models developed here with VS and IR data available (with some gaps) since 1980. Recent studies dealing with the temporal behavior of the dITCZ phenomenon used either (a) a larger temporal time scale (Gu et al., 2005), (b) rather general image statistics (Chen et al., 2008) neglecting the ITCZ signal itself and the temporal dependencies, or (c) manual identification (Wang and Magnusdottir 2006). In contrast, with the model proposed in this paper, it is now feasible to au-

tomatically detect the dITCZ phenomenon and track its signal on a daily time scale, in turn computing meaningful statistics about location, extent and temporal distribution of the dITCZ phenomenon. Furthermore, this type of analysis can help us better understand the mechanisms causing the dITCZ phenomena and thus to improve existing weather and climate models for the tropical circulation system.

Future work may include fine tuning of the model in terms of feature selection, improved methods for classification of rare cases (such as 'no ITCZ'), as well as a detailed comparison of the unsupervised backbone method developed in this work with the Markov random field in Bain et al. (2011). The results achieved with the proposed HMM-based method for detecting and characterizing the ITCZ signals on both hemispheres encourage an in-depth analysis to investigate if initial hypotheses derived from the five season analysis can be proven on long-term data. Initial hypotheses include the accumulation of sITCZ events in La Nina years and the asymmetry in spatial structure of the dITCZ depending on relative longitudinal offset across the equator. With a longer time series we will be able to study the effects of ENSO in general on the four different classes of double ITCZ variability. We will examine the effects of propagation of the Madden Julian Oscillation (MJO) on the double ITCZ variability. Furthermore, with the temporally finer resolution IR channel we can examine whether the results of Bain et al (2010) for the diurnal cycle of the boreal summer ITCZ hold up for the weaker boreal spring convergence zones.

Acknowledgments

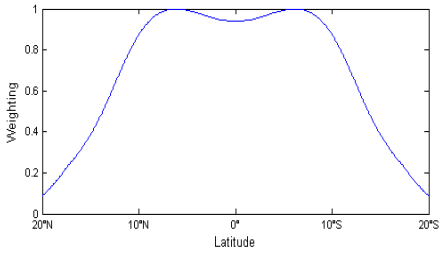
We thank Ashley Payne and Yi-Hui Wang for help with labeling satellite images. We thank Dr. Ken Knapp of NOAA National Climate Data Center for providing the GridSat data. We thank two anonymous reviewers for comments on the manuscript. The project was originally supported by NSF grant ATM-0530926.

2.5 Appendix

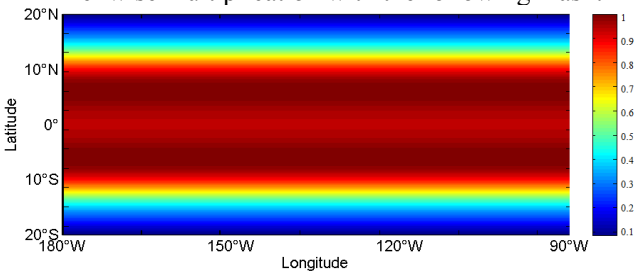
Appendix A: Image Segmentation

Pre-processing

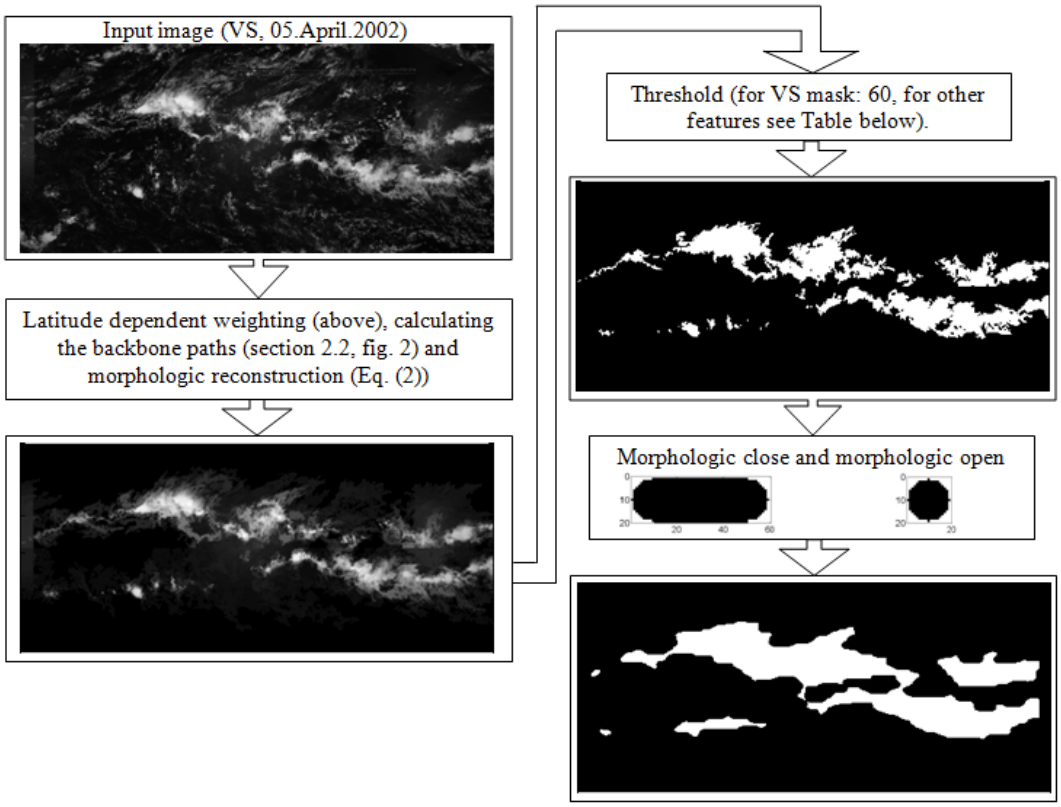
Latitude dependent weighting:



Pixel wise multiplication with the following mask:



Processing chain for the feature extraction:



Appendix B: Feature Calculation

Details on the calculation of selected features:

- *Border tracing (Klette & Rosenfeld, 2004)*

Let *(p,r)* be the first edge where a 0/1 switch occurs:

$$B_t(p) = 0 \rightarrow B_t(r) = 1 \tag{2.5.1.1}$$

Let $\xi(p)$ be the 4-neigborhood of p in circular order:

$$\xi(p) = \{\rho_0, \rho_1, \rho_2, \rho_3 = r\} \tag{2.5.1.2}$$

With the 4-neigborhood being defined as follows:

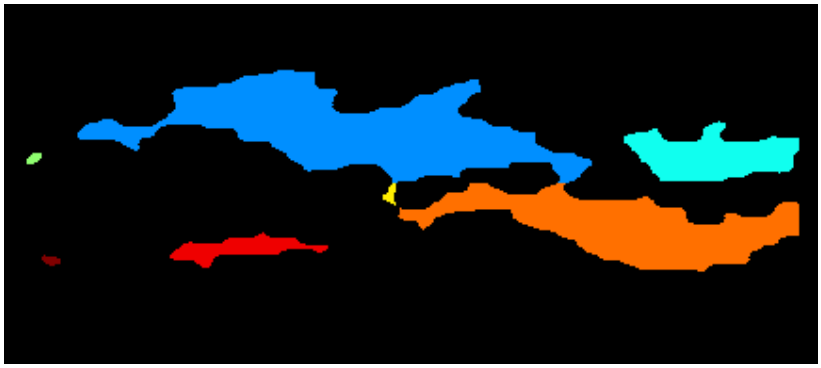
$$\begin{matrix} * & \rho_0 & * \\ \rho_3 & p & \rho_1 \\ * & \rho_2 & * \end{matrix}$$

Find the first $r^* := \rho_k$ in $\xi(p)$ where $\rho_k = 1$ holds: $\min_{0 \le k \le 3} \rho_k = 1$
and add r^* to the list of border points. (2.5.1.3)

If r^* is equal to the initial point of Eq. (2.5.1.1) the complete object border was traced and the next 0/1 switch in the image needs to be found to extract the next object starting from Eq. (2.5.1.1). Otherwise set $r = p$ and $p = r^*$ and repeat steps (2.5.1.2) and (2.5.1.3) until r^* equals the initial point of (2.5.1.1)

$$r^* == initial\ p \quad ? \quad \begin{cases} yes & \rightarrow algorithm\ finished, object\ border\ traced \\ no & \rightarrow r := p, p := r^*\ and\ proceed\ with\ Eq.\ (2.5.1.2) \end{cases} \tag{2.5.1.4}$$

The image is scanned from left-top to right-bottom until all objects are traced.
Result of border tracing:



- *Fourier descriptors*

The extracted object borders are interpreted as closed-form contours of periodic functions $x(t)$ in the complex plane. For a continuous line in the complex space the Fourier coefficients c_n are defined as:

$$c_n = \frac{1}{T} \int_0^T x(t) e^{-2i\pi nt/T} dt \tag{2.5.2.1}$$

with T being the perimeter of the continuous line

Due to the discrete character of the border line, we obtain a set of nodes $\{x_1, x_2, \ldots, x_N = x_0\}$ in the complex plane connected by section-wise straight lines instead. Thus, an explicit calculation

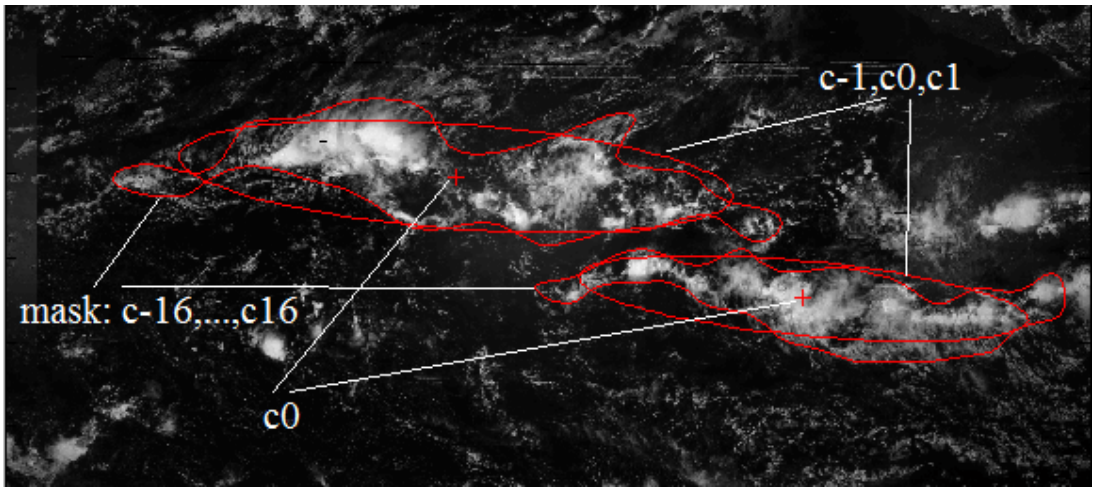
of the section-wise straight lines has to be conducted and it can be shown (e.g. Arbter et al., 1990) that the Fourier coefficients c_n can be efficiently computed by:

$$c_0 = \frac{1}{2T}\sum_{k=0}^{N-1}(x_k + x_{k+1})|x_{k+1} - x_k|$$

$$c_n = \frac{T}{(2\pi n)^2}\sum_{k=0}^{N-1}\left(\frac{x_k - x_{k-1}}{|x_k - x_{k-1}|} - \frac{x_{k+1} - x_k}{|x_{k+1} - x_k|}\right) e^{-2i\pi n t_k/T} \qquad (2.5.2.2)$$

with $t_k = \sum_{l=0}^{k-1}|x_{k+1} - x_k|$ for k>0, $t_0 = 0$

Fourier descriptors:



- *Moments*

Central and invariant moments can provide useful information. The first order moment corresponds to the centre of mass of the image *I*:

$$\bar{x} = \frac{\sum_x \sum_y x\, I(x,y)}{\sum_x \sum_y I(x,y)} \text{ and } \bar{y} = \frac{\sum_x \sum_y y\, I(x,y)}{\sum_x \sum_y I(x,y)} \qquad (2.5.3.1)$$

Higher order moments are less easy to interpret geometrically but nevertheless contain valuable information about the current state of the ITCZ.

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p\, (y - \bar{y})^q I(x,y) \qquad (2.5.3.2)$$

To achieve invariance to scale and translation the central moments are scaled:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{1+(p+q)/2}} \qquad (2.5.3.3)$$

Furthermore by combining the scale invariant moments a set of rotationally invariant moments – the Hu set of moments (Hu, 1962) – can be computed. A subset of these moments is used for classification in this paper.

List of features used for the models:

The table below defines the lists of features used for each of the VS and IR models used in the paper. The final set of features was selected from an initial set of 177 features based on the following simple stepwise algorithm.

- Build a classifier with all of the features
- For each feature
  - Deselect the feature
  - If the classification on the training set improves exclude the feature
  - Otherwise include the feature in the list again
- Proceed until the last feature in the list is reached

This greedy feature selection method is certainly not optimal, but provided a simple method for feature selection for our problem. More sophisticated feature selection methods could likely be used to further improve the performance of the algorithms described in this paper.

Definition: Complete image: 20°N to 20°S; northern image: 20°N to 0°; southern image: 0° to 20°S

VS model feature list:

ID	Feature description	Trend north vs. not-north	Trend south vs. not-south
(General image statistics)			
1	Average pixel value of 12.5°N to 2°N	↑	↓
2	Average pixel value of 2°S to 12.5°S	↓	↑
3	Latitude of the center of mass of the complete image	↓	↑
4	η_{03} of Eq. (2.5.3.3) for the complete image	↑	↓
5	Latitude of the center of mass of the northern image	↑	↑
6	η_{02} for the southern image	↑	↓
7	η_{20} for the southern image	↑	↓
(based on the backbone path)			
8	Regression line parameters of the optimal path (complete image): $linreg(P^O)$ with P^O result of Eq. (2.1.2)	↓	↑
9	Mean of pixel intensities along the optimal path (northern image)	↑	↓
10	Variance of pixel intensities along the optimal path (northern image)	↑	↓
11	Mean of pixel intensities along the optimal path (southern image)	↑	↓
(based on the extracted binary mask, with two different thresholds: 60 and 100 on the scaled reconstructed image with values in the interval [0 255]) *Definition of "longest object": The object with the largest east-west extent of all extracted objects*			
	Features 12-17: Northern image with threshold 100		
12	East-west extent of the longest object	↑	↓
13	Area (in pixels) of the longest object	↑	↓
14	Perimeter (in pixels along the border line) of the longest object	↑	↓
15	Number of direction changes in the border line of the longest object	↑	↓
16	Length of the major axis of the ellipse of the longest object derived by the first Fourier descriptors c_0, c_{-1}, c_1	↑	↓
17	Sum of the east-west extent of all objects with an area larger than 500 pixels	↑	↓
	Features 18-26: Southern image with threshold 100		

18	Compactness of the longest object	↑	↓
19	Perimeter of the longest object	↓	↑
20	Number of direction changes in the border line of the longest object	↓	↑
21	Length of the major axis of the ellipse of the longest object derived by the first Fourier descriptors c_0, c_{-1}, c_1	↓	↑
22	Ratio of major/minor axis of the ellipse of the longest object (Fourier descriptor)	↓	↑
23	Major axis' inclination of the ellipse of the longest object (Fourier descriptor)	↓	↑
24	Number of objects with an area larger than 500 pixels	↑	↓
25	Sum of the perimeter of all objects with an area larger than 500 pixels	↓	↑
26	Sum of the east-west extent of all objects with an area larger than 500 pixels	↓	↑
	Features 27-32: Northern image with threshold 60		
27	Area of the longest object	↑	↓
28	Compactness of the longest object	↓	↑
29	Perimeter of the longest object	↑	↓
30	Sum of the area of all objects with an area larger than 500 pixels	↑	↓
31	Sum of the perimeter of all objects with an area larger than 500 pixels	↑	↓
32	Sum of the east-west extent of all objects with an area larger than 500 pixels	↑	↓
	Features 33-34: Southern image with threshold 60		
33	Sum of the area of all objects with an area larger than 500 pixels	↓	↑
34	Sum of the east-west extent of all objects with an area larger than 500 pixels	↓	↑

IR model feature list:

ID	Feature description	Trend north vs. not-north	Trend south vs. not-south
(General image statistics)			
1	Average pixel value of 2°S to 12.5°S	↓	↑
2	Latitude of the center of mass of the complete image	↓	↑
3	η_{02} of Eq. (2.5.3.3) for the complete image	↓	↑
4	Latitude of the center of mass of the northern image	↑	↓
5	η_{12} for the northern image	↓	↑
6	η_{03} for the northern image	↓	↑
7	Hu moment $J_1 = \eta_{02} + \eta_{20}$ (northern image)	↓	↑
8	Hu moment $J_2 = (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2$ (northern image)	↓	↑
9	Hu moment $J_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$ (northern image)	↑	↓
10	Latitude of the center of mass of the southern image	↑	↓
11	η_{02} for the southern image	↑	↓
12	η_{20} for the southern image	↑	↓
13	Hu moment $J_1 = \eta_{02} + \eta_{20}$ (southern image)	↑	↓
(based on the backbone path)			
14	Mean of pixel intensities along the optimal path (northern image)	↑	↓
15	Variance of pixel intensities along the optimal path (northern image)	↑	↓
(based on the extracted binary mask, with two different thresholds: 30 and 18 on the scaled re-			

constructed inverted IR image with values in the interval [0 255]) *Definition of "longest object": The object with the largest east-west extent of all extracted objects*			
	Features 16-20: Northern image with threshold 30		
16	East-west extent of the longest object	↑	↓
17	Area (in pixels) of the longest object	↑	↓
18	Perimeter (in pixels along the border line) of the longest object	↑	↓
19	Major axis' inclination of the ellipse of the longest object (Fourier descriptor)	↑	↓
20	Sum of the area of all objects with an area larger than 500 pixels	↑	↓
	Features 21-28: Southern image with threshold 30		
21	Compactness of the longest object	↑	↓
22	Perimeter of the longest object	↓	↑
23	Number of direction changes in the border line of the longest object	↓	↑
24	Length of the major axis of the ellipse of the longest object derived by the first Fourier descriptors c_0, c_{-1}, c_1	↓	↑
25	Major axis' inclination of the ellipse of the longest object (Fourier descriptor)	↓	↓
26	Sum of the area of all objects with an area larger than 500 pixels	↓	↑
27	Sum of the perimeter of all objects with an area larger than 500 pixels	↓	↑
28	Sum of the east-west extent of all objects with an area larger than 500 pixels	↓	↑
	Features 29-34: Northern image with threshold 18		
29	Perimeter of the longest object	↑	↓
30	Length of the major axis of the ellipse of the longest object derived by the first Fourier descriptors c_0, c_{-1}, c_1	↑	↓
31	Major axis' inclination of the ellipse of the longest object (Fourier descriptor)	↑	↓
32	Sum of the perimeter of all objects with an area larger than 500 pixels	↑	↓
33	Sum of the east-west extent of all objects with an area larger than 500 pixels	↑	↓
34	Ratio of the sum of the east-west to the north-south extent of all objects with an area larger than 500 pixels	↑	↓
	Features 35-37: Southern image with threshold 18		
35	East-west extent of the longest object	↓	↑
36	Sum of the area of all objects with an area larger than 500 pixels	↓	↑
37	Sum of the east-west extent of all objects with an area larger than 500 pixels	↓	↑

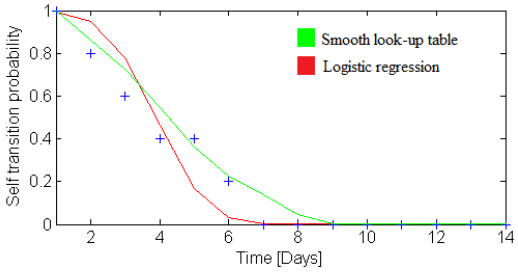
Appendix C: HsMM Details

Description of the Models:

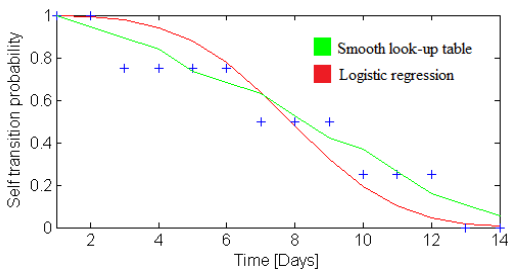
Self-transition duration estimate (simple example, 2 states only, 46 labels):
Top: training sequence *L*, Bottom: $d_k(i)$'s

1 1 1 1 1	2 2	1 1 1	2 2 2 2 2 2 2 2 2	1 1	2 2 2 2 2 2	1	2 2 2 2 2 2 2 2 2 2 2 2	1 1 1 1 1 1
$d_1(1)=5$	$d_1(2)=2$	$d_2(1)=3$	$d_2(2)=9$	$d_3(1)=2$	$d_3(2)=6$	$d_4(1)=1$	$d_4(2)=12$	$d_5(1)=6$

$\Rightarrow D_1(t) = [5,4,3,2,2,1,0]$



$\Rightarrow D_2(t) = [4,4,3,3,3,3,2,2,2,1,1,1,0]$



For the example above the differences between the fitted logistic regression and the look-up table are rather small since there are no outliers in the example - for the real data with few training examples, the look-up table provides robustness against outliers.

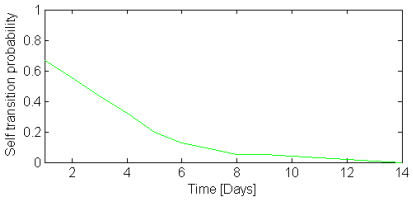
VS model parameters used in the experiments:

- $\pi = [0.3058 \quad 0.4959 \quad 0.1405 \quad 0.0579]$
- $A_{static} = \begin{bmatrix} 0.6389 & 0.0833 & 0.2222 & 0.0556 \\ 0.0833 & 0.9000 & 0.0000 & 0.0167 \\ 0.4706 & 0.0000 & 0.4118 & 0.1176 \\ 0.1429 & 0.4286 & 0.1429 & 0.2857 \end{bmatrix}; \quad \lambda = 0.1$
- Emission probabilities by SVMs (parameters for each state are learned by the toolbox by De Brabanter et al., 2010)
- Logistic regression parameters of Eq. (2.6.4):

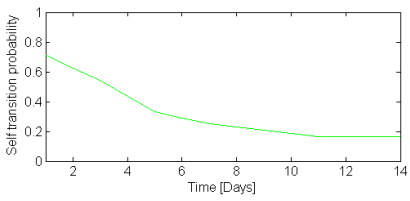
  $\varepsilon_{DITCZ} = 10.48 \quad \varepsilon_{north} = 4.84 \quad \varepsilon_{south} = 9.85 \quad \varepsilon_{none} = 1.49$

  $\gamma_{DITCZ} = -1.31 \quad \gamma_{north} = -0.04 \quad \gamma_{south} = -1.66 \quad \gamma_{none} = 1.51$
- Self-transition look-up tables:

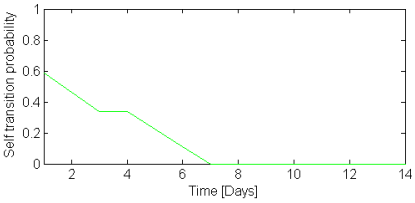
dITCZ



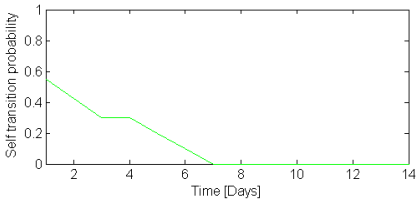
nITCZ



sITCZ



non-present

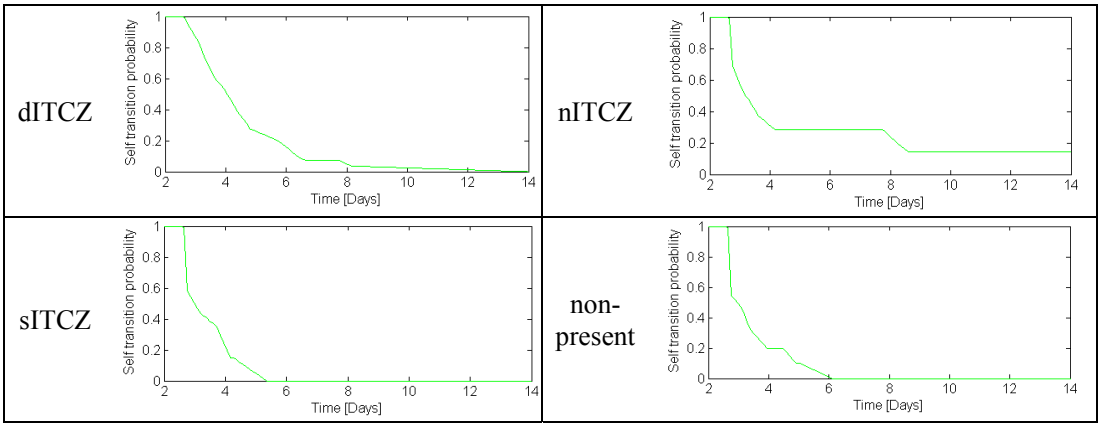


IR model parameters used in the experiments:

- $\pi = [0.3337 \quad 0.4898 \quad 0.1163 \quad 0.0602]$
- $A_{static} = \begin{bmatrix} 0.9572 & 0.0122 & 0.0245 & 0.0061 \\ 0.0104 & 0.9875 & 0.0000 & 0.0021 \\ 0.0702 & 0.0000 & 0.9123 & 0.0175 \\ 0.0169 & 0.0508 & 0.0169 & 0.9153 \end{bmatrix}; \quad \lambda = 1$
- Emission probabilities by SVMs (parameters for each state are learned by the toolbox by De Brabanter et al., 2010)
- Logistic regression parameters of Eq. (2.6.4):

$$\varepsilon_{DITCZ} = 5.92 \quad \varepsilon_{north} = 6.45 \quad \varepsilon_{south} = 4.92 \quad \varepsilon_{none} = 1.03$$
$$\gamma_{DITCZ} = -0.22 \quad \gamma_{north} = -0.46 \quad \gamma_{south} = -0.56 \quad \gamma_{none} = 1.96$$

- Self-transition look-up tables:



Appendix D: Detailed Description of the semi-supervised EM algorithm

E-Step (Find $\boldsymbol{Q}^r$ given $\boldsymbol{O}$ and $\boldsymbol{\theta}^r$ under the constraint: $\boldsymbol{Q}^r$ traverses $\boldsymbol{L}^{known}$)

- Calculate the observation probabilities $b_i(\boldsymbol{x}) = P(\boldsymbol{o}_t = \boldsymbol{x} \mid q = i)$ for all time steps t and all states i using the SVMs.
- Force the state sequence Q to traverse the supervised labels $\boldsymbol{L}^{known}$, by setting the transition probabilities at given labeling times to 1 for the transition into the labeled state and to 0 otherwise:

$$\forall\, j,k\text{:}\ \begin{matrix} a_{j,l_{8*k+m}} = P(q_{8*k+m} = l_{8*k+m} \mid q_{8*k+m-1} = j) = 1 \\ a_{j,other} = P(q_{8*k+m} \neq l_{8*k+m} \mid q_{8*k+m-1} = j) = 0 \end{matrix} \qquad (2.5.4.1)$$

- By doing so, the probabilities $P(\boldsymbol{O},\boldsymbol{Q}|\boldsymbol{\theta})$ are 0 for all sequences $\boldsymbol{Q}$ not fulfilling the supervised label constraint. Another interpretation is that for our likelihood maximization problem we actually marginalize only over the possible state sequences $\boldsymbol{Q}^{supervised}$, i.e.:

$$\forall \boldsymbol{Q} \neq \boldsymbol{Q}^{supervised} : P(\boldsymbol{O},\boldsymbol{Q}|\boldsymbol{\theta}) = 0 \Rightarrow \mathcal{L}(\boldsymbol{\theta};\boldsymbol{O}) = \sum_{\forall \boldsymbol{Q}} P(\boldsymbol{O},\boldsymbol{Q}|\boldsymbol{\theta}) = \sum_{\forall \boldsymbol{Q}^{supervised}} P(\boldsymbol{O},\boldsymbol{Q}|\boldsymbol{\theta}) \qquad (2.5.4.2)$$

- Now either the alpha-beta or the Viterbi algorithm can be used to estimate the most probable state sequence $\boldsymbol{Q}$ which satisfies the condition of traversing all supervised labels $\boldsymbol{L}^{known}$. Since for the Viterbi algorithm the constraint that a minimum of 8 consecutive steps in the same state can be guaranteed, in our application we use it instead of the alpha-beta algorithm (which does not enforce this constraint) and we can obtain directly from the state sequence $\boldsymbol{Q}$ a new training sequence $\boldsymbol{L}^{r+1} = \boldsymbol{Q}$ for the M-Step.

M-Step:

- Update the model parameters $\boldsymbol{\theta}^{r+1}$ as follows:
  - Train the SVM parameters $\boldsymbol{w_i}$ and u_i by using the labels $\boldsymbol{L}^{r+1}$ from the E-Step and the corresponding complete observation set $\boldsymbol{O}$. Set up the training pairs and learn the SVMs as for the VS case (Eq. (2.6)).
  - Learn $\boldsymbol{A}$ and $\boldsymbol{\pi}$ using the labels $\boldsymbol{L}^{r+1}$ equivalent to the supervised VS case (Eq.(2.4)).
  - Force the minimum self-transition time to be at least 8 time steps: $\forall i, 0 < t < 8$: $D_i(t) = 1$

3 MOVING TARGET TRACKING IN SINGLE-CHANNEL, WIDE-BEAM SAR

This chapter has been published as: Henke D., Magnard C., Frioud M., Small D., Meier E. & Schaepman M.E. (2012). Moving target tracking in single-channel, wide-beam SAR. *IEEE Transactions on Geoscience and Remote Sensing, 50*(11): 4735-4747.

Abstract

A novel method for moving target tracking using single-channel Synthetic Aperture Radar (SAR) with a large antenna beam width is introduced and evaluated using a field experiment and real SAR data. The presented approach is based on sub-aperture SAR processing, image statistics and multitarget unscented Kalman filtering. The method is capable of robustly detecting and tracking moving objects over time, providing information not only about the existence of moving targets but additionally about their trajectories in the image space while illuminated by the radar beam. We have successfully applied the method on an experimental data set using MiSAR to accurately characterize the movement of vehicles on a highway section in the radar image space.

3.1 Introduction

Tracking moving objects in SAR data is a challenging task demanding state-of-the-art processing. Traditional moving target indication (MTI) in SAR data relies on space-time adaptive processing (Corbell et al., 2007; Zhang et al. 2009) along-track interferometry (Budillon et al., 2008; Chapin and Chen, 2008), displaced phase centre array (Genyuan et al., 2004), detection by focusing with different estimated velocity vectors (Pettersson, 2007), monopulse processing (Rüegg et al., 2007), exploiting patch differences in autofocus results (Fienup, 2001) or a combination of the above (Suchandt et al., 2010; Gierull, 2006). An alternative approach for MTI solutions is the development of new sensor technologies, e.g. a special MTI mode as the scan-MTI mode of PAMIR (Brenner and Ender, 2006; Cerutti-Maori et al. 2008; Cerutti-Maori et al.; 2010). However, for some of these approaches the restrictions on the movements of the object, e.g. only linear motion, and on the transmitter flight track limit the potential applications. Furthermore, in most methodologies, multiple antenna channels are required to achieve satisfactory detection probability and the presence of moving objects is only indicated with limited information about the speed and direction of detected objects. However, when dealing with ultra-lightweight sensors such as MiSAR (Edrich and Weiss, 2008; Edrich, 2006) – commonly required in unmanned aerial vehicles (UAVs) – generation of multi-channel data may not be feasible. Therefore, to overcome these problems, we make use of an approach based on image statistics of geocoded SAR images, unscented Kalman filtering and multi-target hypotheses to extract the movement of objects illuminated by the beam of a suitable single-channel, wide-beam SAR sensor. We propose a combination of image processing and advanced multi-target tracking techniques with hidden state space representation as a promising approach for solving the problem of multiple targets with movements that contain a certain degree of non-uniformity and non-linearity, to overcome the above listed limitations. Multi-target tracking for radar applications is presented in Bar-Shalom et al. (2009). Special emphasis has been placed on improving existing tracking methods in recent years within the computer science community, with applications to a variety of problems (Khan et al., 2005; Pei et al., 2008; Scharenbroich et al., 2009). These promising approaches are introduced and adapted to the tracking of moving objects in SAR data (Henke et al., 2010) where a general framework is demonstrated.

Based on above progress, an algorithm to track moving objects in wide-beam, single-channel systems was developed and evaluated. The proposed algorithm which relies on the processing of a large number of highly "Overlapping Apertures with reduced Bandwidth" (OAwrB) and subsequent tracking of anomalies in the temporal image statistics provides a flexible and promising solution for the tracking of moving objects even under conditions where most other methods available for single-channel SAR systems fail. We demonstrate that by successfully testing the method on a data set recorded with the wide-beam, single-channel MiSAR system in southern Germany, where we were able to accurately extract the trajectories of most of the cars and trucks on a highway and its associate on-ramps in the *image* space. However, these trajectories are the movements of the targets in the *image* space only and not the real movements in *3D* space, which are distorted due to the Doppler shift.

In the following, we briefly introduce the MiSAR sensor and its specifications as well as the test area. Then the method is described in detail. Beginning with the SPECAN-processed OAwrB images, we will focus on the extraction of potential moving objects in each OAwrB and subsequently use these observations to integrate the findings in a general tracking framework to obtain the trajectories of the moving objects over time. We describe the results obtained on the MiSAR data set and conclude with potential future research and applications.

3.2 Methods

3.2.1 Test site and sensor

The study area is located south of the city of Ulm in southern Germany. In Fig. 3.1 the focused geocoded single-channel SAR image is overlaid on a topographical map. The strip is ~2450 m long and ~850 m wide. Two highways, one with along-track and one with cross-track orientation, are present: they serve as the main areas for testing the tracking method.

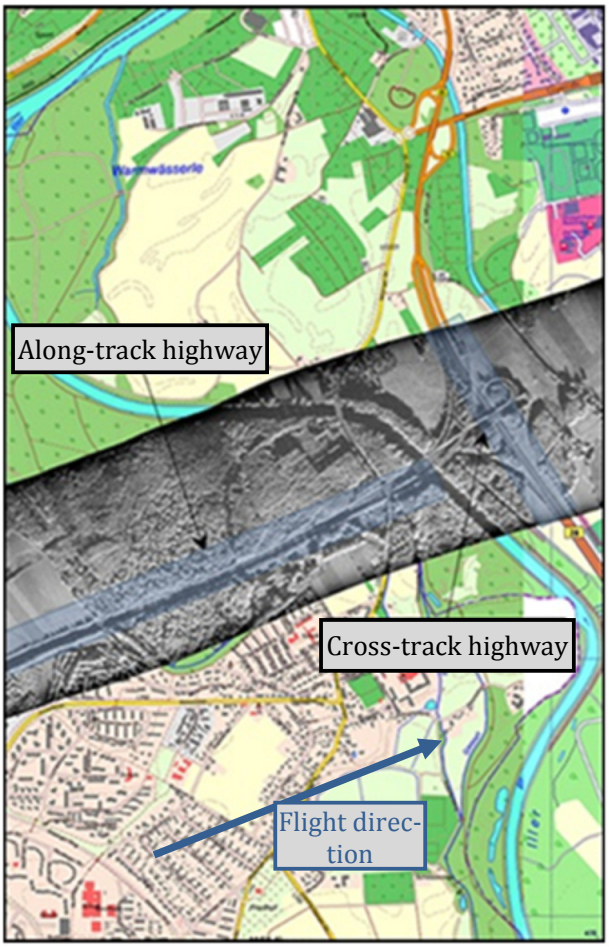


Fig. 3.1 – Test site in South Germany: Topographic map overlaid with the focused SAR image with the two main highways investigated for moving object tracking.

The sensor system used for the experiments is the frequency-modulated continuous-wave (FMCW) Miniature Synthetic Aperture radar (MiSAR) system, developed by EADS, Germany (Edrich and Weiss, 2008; Edrich, 2006). It is a small lightweight system that operates at 35 GHz in Ka-band and can be mounted on Unmanned Aerial Vehicles (UAVs) or small manned aircrafts. The sensor's azimuth beam width for the experiment is ~14° and the pulse repetition frequency (PRF) 2000 Hz, with resolution ~0.5 m in range and azimuth. The MiSAR's slant range is approximately three times the cruise altitude. In the experiment discussed, the slant range was ~1500 m.

3.2.2 Overview of the processing chain

The proposed method for tracking moving targets in SAR data with a wide beam sensor can be subdivided into three major parts. The first task is to process and geocode the SAR data to OAwrB images. The tracking algorithm works independently of the focusing algorithm. We use the SPECtral ANalysis (SPECAN) algorithm (Cumming and Wong, 2005; Curlander and McDonough, 1991) to focus the OAwrB images, as it is an ideal burst processing method for low bandwidth images. Having obtained a sequence of images where static targets stay at the same position while the moving targets are displaced over time, we collect image statistics and use image processing methods to extract potential moving objects in each image. The candidate objects are based only on single images, with the temporal behavior ignored at this stage. In the third processing step, we use a multi-target unscented Kalman filter approach incorporating a dynamic model to associate the extracted time-independent observation of step two to generate maximum probability moving target trajectories. An overview of the processing chain can be found in Fig. 3.2.

3.2.3 Processing and geocoding the raw data

In preparation for the tracking algorithm, the OAwrB images are first processed. They have an azimuth extent of 250 m - 300 m, determined by the azimuth beam width and slant range. An implicit tradeoff is made between a low bandwidth with a higher energy concentration of moving vs. static targets vs. a high bandwidth with a better image resolution. We chose short bursts of only 0.25 s, corresponding to ~8 m synthetic aperture length, i.e. around 2% of the full synthetic aperture, resulting in a good energy conservation of moving targets while also offering a reasonable spatial resolution (~1 m). A compromise has to be made between (a) a high overlap implying a high temporal resolution and thus an improved tracking reliability, and (b) computation time. We chose an overlap between the OAwrB images of ~90% (step width: 0.0195 s).

Preprocessing steps such as Kalman filter based GPS/INS integration, track linearization including first order motion compensation, range cell migration correction (performed in the range-Doppler domain) and side-lobe suppression are performed before the SPECAN algorithm can be applied. SPECAN, a burst processing method ideal for OAwrB images, is a very computationally efficient method. It directly delivers a focused image for nearly the entire illuminated footprint, also in the azimuth dimension, using a short azimuth segment (or burst), where the azimuth extent of the processed image is much larger than the burst length. SPECAN is significantly less computationally demanding than other stripmap processing methods such as range-Doppler, Ω-k or chirp scaling.

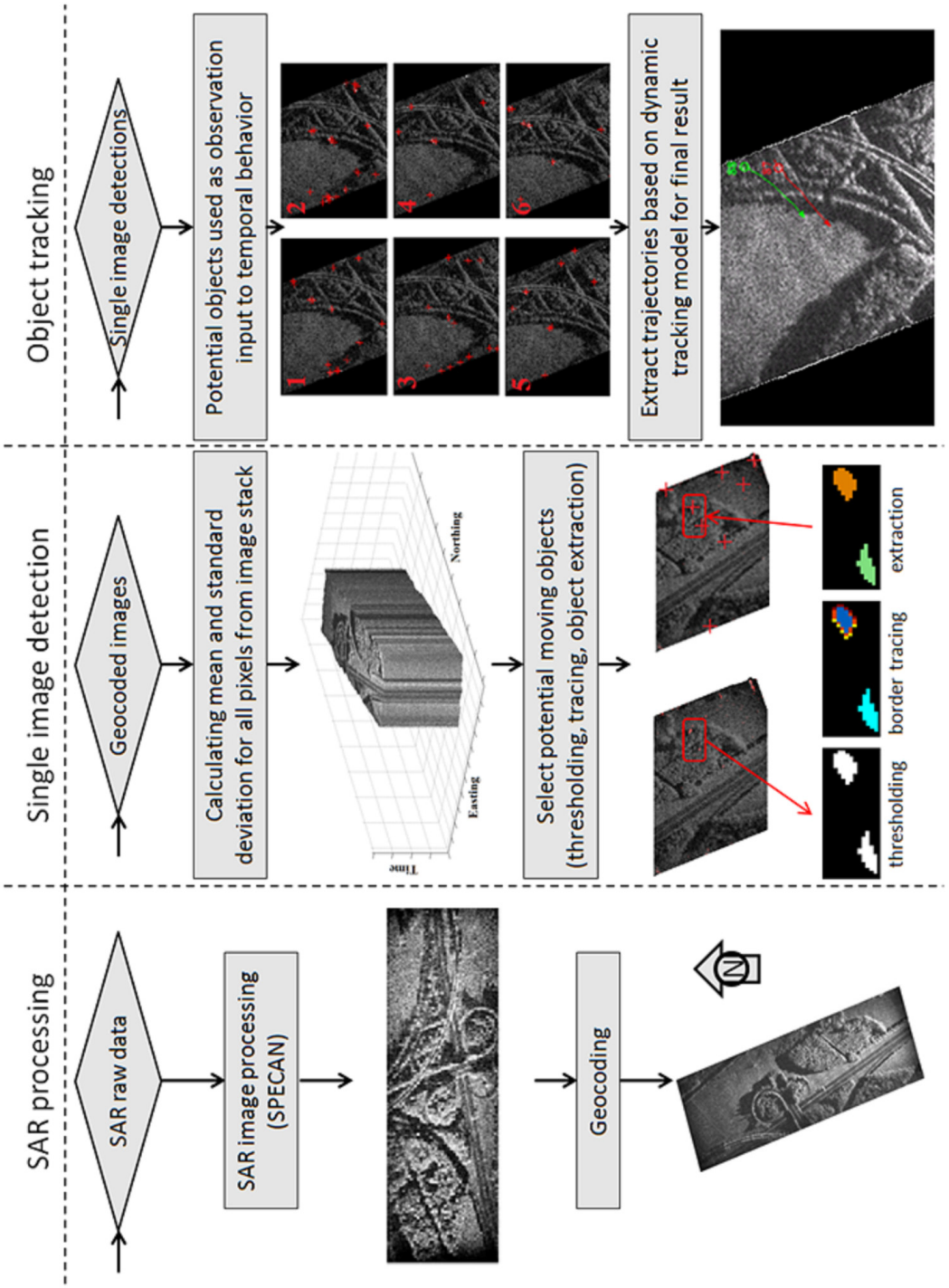


Fig. 3.2 – Flow-chart of the processing chain.

The use of SPECAN in the past (Cumming and Wong, 2005; Curlander and McDonough, 1991) has often been restricted to narrow antenna beam-widths and small squint angles (i.e. satellite sensors). We implemented an algorithm described in Magnard et al. (2010) for wide-beam airborne sensors, briefly summarized here. The most important assumption made in SPECAN is that the Doppler rate should be constant for any target during the whole of its integration time and that this rate is constant for all targets at the same slant range. While the Doppler centroid is deduced from the navigational data, the Doppler rate is taken as its instantaneous value for a

target at a given slant range and a given squint angle associated with the Doppler centroid relative to the linearized track. With the previously mentioned constant Doppler rate assumption, all targets at the same slant range have similar Doppler histories. Plotting the Doppler centroid versus azimuth time, these histories are approximated by straight lines with the slope of the Doppler rate (ramps). The principle of SPECAN is based on the observation that the position of any target (or equivalently its azimuth time) is uniquely determined by its frequency value at some reference time. This equivalence makes it possible to reduce the required processing to only two steps, namely de-ramping and spectral analysis. The de-ramping transforms the Doppler history into constant frequency lines, and the spectral analysis maps these frequency components to the azimuth time. The spectral analysis is performed through a Scaled Fourier Transform implemented via fast-convolution using the Chirp-Z Transform as described in Lanari (1998). Compared to the basic variant of SPECAN using a simple Fourier Transform, this has the advantage of producing images with an azimuth pixel spacing independent of range, therefore avoiding the need to interpolate the data at the final stage. In addition, the azimuth pixel spacing can be adjusted using a scale factor.

Finally, tracking the target requires that the stack of images share a common geometry. The OAwrB images are geocoded once they have been focused with the SPECAN algorithm to provide the position and velocity of the moving targets in a geodetic coordinate system. For the geocoding, first a set of points around the perimeter of the focused image is projected into a map geometry using a forward geolocation algorithm. Then a region of interest containing the coordinates of these points plus a margin is defined and a regular grid in local map coordinates is created in this region of interest. Finally, the image samples are extracted from the input data using a standard backward geocoding algorithm. Both forward and backward geocoding processes are based on a standard range-Doppler approach (Meier et al., 1993) using a Digital Elevation Model (DEM) and all necessary geodetic and cartographic transformations and parameters. A detailed description of this algorithm can be found in Magnard et al. (2010).

The geocoded images have a pixel-spacing of 0.5 m and were evaluated to analyze the internal accuracy of the geocoding procedure. Since no reflectors were deployed in the campaign, three bright static targets on a dark concrete area were used to estimate the geolocation accuracy, relevant for the later detection and tracking algorithm. The standard deviation of these static targets varies between 0.51 m and 1.37 m in northing and between 0.94 m and 1.86 m in easting over the whole sequence of images where the static targets were in the antenna beam. The image to image shift of the geolocation is on average 0.06m in northing and 0.18m in easting. The accuracy of the geocoded images is most reliable in the center of the images and degrades to towards the edges. Causes of error are inaccurate navigation data, motion compensation of small flight track segments and also SPECAN algorithm uncertainties at the image edges (mainly azimuth dimension). Nevertheless, the variations in the relative image-to-image positions are sufficiently accurate for our detection and tracking purposes.

3.2.4 Extraction of moving target candidates

After the raw data focusing to geocoded OAwrB images, one obtains an image stack *S(m,n,t)* over time for which the following statement holds:

Let $I_t(m,n)$ be the OAwrB geocoded amplitude image at time t.

$$S(m,n,t) = \begin{cases} I_t(m,n) & \begin{array}{l} if\ the\ pixel\ at \\ position\ (m,n)\ is\ a \\ valid\ pixel\ in\ image\ I_t \end{array} \\ undefined & otherwise \end{cases} \tag{3.1}$$

From Eq. (3.1) one can derive the mean and standard deviation of each pixel's intensity over time:

$\forall m, n$:

$$\bar{S}(m,n) = \frac{\sum_{\forall t:S(m,n,t)\neq undef} S(m,n,t)}{\sum_{\forall t:S(m,n,t)\neq undef} 1} \tag{3.2.1}$$

$$\sigma(m,n) = \sqrt{\frac{\sum_{\forall t:S(m,n,t)\neq undef} (S(m,n,t) - \bar{S}(m,n))^2}{\sum_{\forall t:S(m,n,t)\neq undef} 1}} \tag{3.2.2}$$

As we approximate the mean and standard deviation based on pixel samples, it is clear that with increasing numbers of samples the reliability of the estimated statistics and thus the detection probability of potential moving objects improves. This increase in the number of pixel samples is dependent on the pixel's beam dwell time and is consequently best achieved by using a wide beam sensor where our proposed method works independently of the acquisition mode. With minor adaptations the method could also be applied to sensors with a narrower azimuth beam width using spotlight mode to achieve longer dwell times.

From the statistics calculated in Eq. (3.2), we can derive the outlier pixels with high positive variance that indicate the presence of non-static pixels at specific OAwrB instances:

$B(m,n,t)$ is defined as a Boolean matrix with value 1 for potential moving pixels and 0 otherwise.

$$B(m,n,t) = \begin{cases} 1 & if\ S(m,n,t) > \bar{S}(m,n) + \alpha\sigma(m,n) \\ 0 & otherwise \end{cases} \tag{3.3}$$

The value of α specifies the threshold that determines which pixels are treated as candidate moving objects. The parameter α is a sensor-specific property and was chosen on an experimental basis. In theory, the parameter could be derived from an extensive calibration data take, overflying the same restricted control area twice – once with and once without moving objects – to estimate the optimal value for α. However, experiments have shown that the choice of α is not critical providing that it stays within a certain range. We set $\alpha=4.5$ in our MiSAR experiment, providing an appropriate tradeoff between false positive detections and missing candidates.

Until this point, we have been dealing with a pixel-based matrix, but to obtain a set of potential moving object candidates at each time step, we next make use of segmentation algorithms to group single pixels into connected object areas. First of all, morphological open and close operations are conducted to reduce noise in the Boolean matrix B. A morphological reconstruction based on the new matrix B and S is then performed to connect neighboring objects that have a high positive variance on a connecting path. An efficient morphologic reconstruction is described in Robinson and Whelan (2004). For each image at time t, B_t acts as a seed point mask and S_t as a marker image to obtain S_t^{new} which is used to calculate B_t^{new} equivalent to Eq. (3.3), using a slightly lower threshold α_2 (in our MiSAR experiment: $\alpha_2=3.5$). Fig. 3.3 illustrates the output of the algorithm for a particular time.

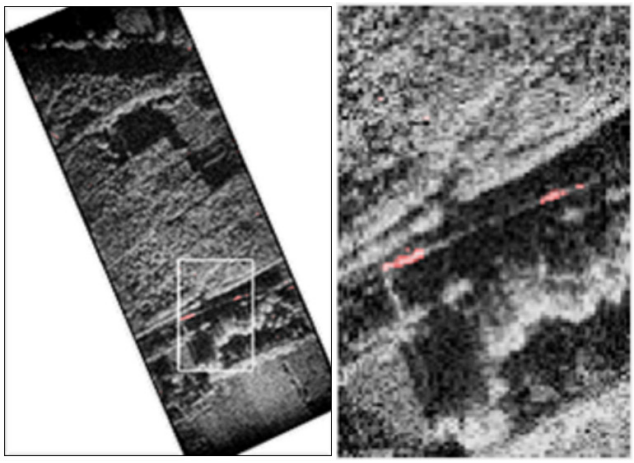


Fig. 3.3 – Pixel areas marked as potential moving objects after morphological reconstruction.

Finally, one has to extract connected objects from the Boolean pixel mask. We use an extraction algorithm based on border tracing as described in Klette and Rosenfeld (2004):

Let *(p,r)* be the first spatial edge where a 0/1 switch occurs:
$B_t(p,n) = 0 \rightarrow B_t(r,n) = 1$ (3.4.1)

Let $\xi(p)$ be the 4-neigborhood of p in circular order:
$\xi(p) = \{\rho_0, \rho_1, \rho_2, \rho_3 = r\}$ with the 4-neigborhood being defined as follows:
$$\begin{bmatrix} * & \rho_0 & * \\ \rho_3 & p & \rho_1 \\ * & \rho_2 & * \end{bmatrix} \quad (3.4.2)$$

Find the first $r^* := \rho_k$ in $\xi(p)$ where $\rho_k = 1$ holds: $\min_{0 \le k \le 3} \rho_k = 1$
and add r^* to the list of border points. (3.4.3)

If r^* is equal to the initial point of Eq. (3.4.1) the complete object border was traced and the next 0/1 switch in the image needs to be found to extract the next object starting from Eq. (3.4.1). Otherwise, set $r = p$ and $p = r^*$, and repeat steps (3.4.2) and (3.4.3) until r^* equals the initial point of (3.4.1):

$$r^* == initial\ p\ ? \begin{cases} yes & \rightarrow algorithm\ finished,\ object\ border\ traced \\ no & \rightarrow r := p, p := r^*\ and\ proceed\ with\ Eq.(4.2) \end{cases} \quad (3.4.4)$$

Thus, one obtains a set of objects for each binary image B_t indicating potential moving objects. This set of moving object candidates includes a considerably higher number of (false) detections than the final detection set after the tracking model is applied. The reasons are (a) the relatively low values for the threshold α and the processed bandwidth of the OAwrB which enable detection of objects with low Radar Cross Section (RCS)[5] at the cost of a higher number of false detections in single OAwrB images and (b) speckle noise effects. Fortunately, the dynamic modeling of the temporal tracking successfully suppresses most of these misdetections given that non-moving detections (e.g. speckle noise, static above-threshold detections) appear randomly or completely static in the image space thus lacking any dynamic behavior over time and therefore getting identified as clutter by the tracking model. For each object, characteristics like

[5] To calculate the ranges for objects that are detectable is impracticable as it depends on factors like its RCS, orientation, moving direction, speed (i.e. energy concentration), background clutter, location within the antenna beam etc.

the center of mass and the area are calculated. The area can be used to eliminate small objects most likely caused by noisy artifacts. Additional object characteristics might be useful for the tracking algorithm but were not investigated in this work. The center of mass illustrated in Fig. 3.4 is used as the "raw" observation for the Kalman tracking approach discussed in the next section when considering temporal behavior.

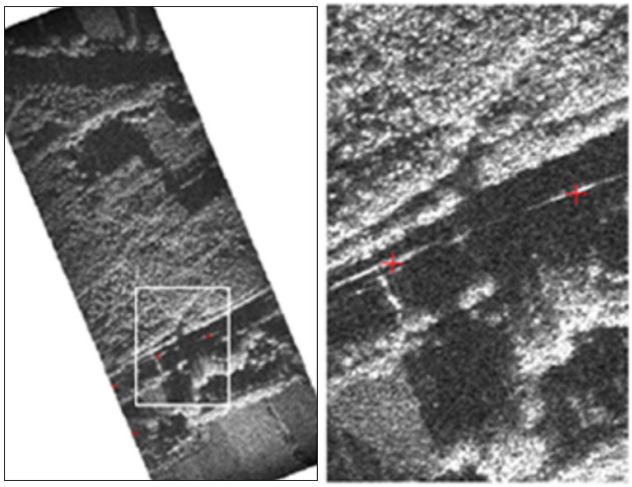


Fig. 3.4 – Output of the center of mass points serving as observations for the multitarget Kalman filter approach. Time step is the same as in Fig. 3.3.

3.2.5 Tracking of the moving objects

A broad overview of multi-target tracking algorithms is provided in Bar-Shalom et al. (2011). The advanced multi-target tracking algorithm we investigate in this paper is based on the theory of unscented Kalman filtering (e.g. Wan and Van Der Merwe, 2000; Khairnar et al., 2007) and sequential Monte Carlo methods (e.g. Doucet et al., 2001) in general, and on a combination of both for multi-target tracking specifically on a work by Särkkä et al. (2007). The tracking approach developed by Särkkä et al. (2007) outperforms several multi-target tracking methods in accuracy and computational efficiency. For a detailed comparison with classical multi-target tracking approaches with a fixed number of targets, see Särkkä et al. (2004). The basic framework for SAR moving target tracking using multi-target unscented Kalman filters was previously introduced and demonstrated for a simplified simulated data set in Henke et al. (2010). For the real data experiments, two major changes have been made compared to the previous work: first, while in Henke et al. (2010) we developed a tracking method for simulated multi-static data, we deal here with a monostatic real data experiment, implying some presumptions concerning the movement of the objects: we assume the moving objects to be at ground level and have only moderate non-linear motion and acceleration events. Thus, we can estimate the moving target's position based only on monostatic acquisitions. Second, the extraction of potential moving targets tuned for the MiSAR sensor (see III.B) is now integrated into the tracking framework: several adaptations to the Kalman filter model were necessary.

In the following, we summarize the computational chain, explaining the parameters used within the tracking algorithm.

3.2.5.1 Filtering models and the Unscented Kalman Filter (UKF)

Kalman filters are time discrete filters modeled with Markov chain properties and additional Gaussian noise to estimate the hidden state of a dynamic system from noisy observations. An iterative estimation-maximization approach is used to find the optimal hidden state estimate at each time step. Thereby, the unscented Kalman filter is an extension of the standard Kalman

filter allowing non-linear state transitions and transformation functions. This is realized by the so-called unscented transformation (Julier, 2002; Julier and Uhlmann, 2004).

For the unscented transformation, we consider a random variable ψ (of dimension L) with mean $\bar{\psi}$ and covariance C_ψ propagated through a nonlinear function $y = g(\psi)$. To calculate the transformed random variable y, we first select $2L+1$ sigma points ψ_i with corresponding weights w_i as defined in the following equations (3.5.1) and (3.5.2):

First, form the sigma point matrix $\Psi = [\psi_o^T \dots \psi_{2L}^T]^T$

$$\begin{aligned} \psi_0 &= \bar{\psi} \\ \psi_i &= \bar{\psi} + \left(\sqrt{(L+\lambda)C_\psi}\right)_i \qquad i = 1, \dots, L \\ \psi_i &= \bar{\psi} - \left(\sqrt{(L+\lambda)C_\psi}\right)_{i-L} \quad i = L+1, \dots, 2L \end{aligned} \tag{3.5.1}$$

Form the mean $W^{(m)} = \left[w_0^{(m)T} \dots w_{2l}^{(m)T}\right]^T$ and the covariance $W^{(c)} = \left[w_0^{(c)T} \dots w_{2l}^{(c)T}\right]^T$ weights:

$$\begin{aligned} w_0^{(m)} &= \frac{\lambda}{L+\lambda} \\ w_0^{(c)} &= \frac{\lambda}{L+\lambda} + (1 - \alpha^2 + \beta) \\ w_i^{(m)} &= w_i^{(c)} = \frac{1}{2(L+\lambda)} \qquad i = 1, \dots, 2L \end{aligned} \tag{3.5.2}$$

$\lambda = \alpha^2(L + \kappa) - L$ is a scaling parameter, α determines the spread of the sigma points around $\bar{\psi}$ (usually a small positive value, e.g. 0.001), κ is a secondary scaling parameter (set to 0) and β is used to incorporate prior knowledge of the distribution of ψ (for a Gaussian: $\beta = 2$). The lower index *i* indicates the *i*th row of a matrix. These sigma points are then interpreted using the non-linear function *g*: $\mathcal{Y}_i = g(\psi_i)$. The transformed $\mathcal{Y}_i$'s are used to approximate the *a posteriori* distribution of the random variable *y*:

Weighted sample mean

$$\bar{y} \approx \sum_{i=0}^{2L} w_i^{(m)} \mathcal{Y}_i \tag{3.5.3}$$

Weighted sample covariance

$$C_y \approx \sum_{i=0}^{2L} w_i^{(c)} (\mathcal{Y}_i - \bar{y})(\mathcal{Y}_i - \bar{y})^T \tag{3.5.4}$$

This approximation preserves at relatively low computational cost accuracy to the third order for Gaussian inputs for all nonlinearities and thus outperforms other non-linear Kalman filter approaches as e.g. the extended Kalman filter. By using the unscented transformation we can use arbitrary non-linear state transition and transformation functions for the Kalman filter to propagate the state and covariance matrix forward in time. The state transition function

$$x_k^- = f(x_{k-1}, q_k) \text{ with } q_k \sim N(0, Q_k) \tag{3.6.1}$$

projects the estimated hidden state forward in time. q_k models the process noise; x_{k-1} and x_k^- are the states at the previous time step *k-1 after* the observation update and the current time step *k before* the observation update, respectively. This part of the iterative procedure is called the *prediction* step. In the *update* step, a new observation is made at time step *k* and the current state x_k^- is updated according to the transformation function

$$z_k = h(x_k^-, v_k) \text{ with } v_k \sim N(O, R_k) \tag{3.6.2}$$

which maps the state x_k^- *before* the observation into the observation space Z with v_k modeling the uncertainty of the observations. This function is used to calculate the Kalman gain K_k and update the state x_k to obtain the most probable state estimate considering the observation at time step *k*.

In our experiments, the state *x* is defined as:

$$x = \begin{pmatrix} pos_{east} \\ pos_{north} \\ V_{abs} \\ \varphi \end{pmatrix} \tag{3.6.3}$$

with pos_{east} and pos_{north} being the map coordinates of the object [m], V_{abs} the absolute velocity [m/s] and φ the direction of movement [°]. The transformation function mapping from the state into the observation space is in our case linear:

$$h(x) = Hx \tag{3.6.4}$$

with *x* being the state and $H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$ the transformation matrix. The extracted centers of mass in the observation space correspond to the coordinates pos_{east} and pos_{north} of the state.

The non-linear state transition function feeding forward the state *x* in time *t* is defined in our experiment as follows:

$$f(x) = \begin{pmatrix} pos_{east} + t * cos(\varphi) * V_{abs} \\ pos_{north} + t * sin(\varphi) * V_{abs} \\ V_{abs} \\ \varphi \end{pmatrix} \tag{3.6.5}$$

Eq. (3.6.3)-(3.6.5) define the fundamental parameters of the Kalman filters. For the complete definition, the initial values for the state x_o and its associated covariance matrix C_0 as well as the observation noise R and process noise Q must also be set appropriately. The positions pos_{east} and pos_{north} are initialized to the coordinates of the center of the geocoded SAR image. For the absolute velocity V_{abs}, a value of 30 m/s (108 km/h) is chosen as a best initial estimate of the velocity of vehicles on the highway. The initial direction φ can have an arbitrary value, as no prior knowledge about the main driving direction of the vehicles is available. We account for the arbitrary direction value by setting the initial covariance matrix C_0 with a quasi uniformly distributed direction, equivalent to a very high value for the diagonal value of C_0 associated with the direction. Also for the diagonal values of C_0 affecting the positions we choose a quasi uniform distribution over the complete valid area of the SAR image given that we have only limited knowledge of the "birth" position of moving objects. The diagonal value of C_0 affecting the absolute velocity is set to 100, equivalent to a standard deviation of 10 m/s implying an initial guess that ~68% of the vehicles move with a speed between 20 – 40 m/s (72 – 144 km/h) and ~95% between 10 – 50 m/s (36 – 180 km/h). Non-diagonal values of C_0 are set to 0. For the observation noise R, the diagonal values are set to 25, implying a standard deviation of 5m in northing and easting of the observations, respectively, reflecting a realistic uncertainty

after visual inspection of the results of the method in III.B. Finally, the process noise matrix Q has to be defined. The diagonal values in Q associated with the position values of the state are set to low values of 0.01m. The state transition function of the model implies uniform and linear dynamic behavior of the moving objects, but we want to be able to track non-uniform, non-linear targets, so we allow as a noise component a standard deviation of 0.5° for the value associated with the movement direction (this favors linear movement as highway vehicles are assumed to be moving mainly straight but allows a degree of non-linearity) and a standard deviation of 1.5 m/s for the value associated with the absolute velocity V_{abs}, to realistically model vehicular acceleration and slowdown events. We use a model that assumes uniform, linear movement only to keep the number of parameters for the state low and thus more robust. However, by allowing some noise component for the moving direction and a relatively high noise level in the velocity, non-linearity as well as acceleration events are handled well within realistic boundaries.

3.2.5.2 Tracking multiple moving objects

The Kalman filter described above handles an accurate tracking of one vehicle. An open problem remains the tracking of multiple objects in the presence of clutter, equivalent to associating each observation to the correct tracking history or identifying it as clutter. This is done based on Rao-Blackwellized particle filtering (Särkkä et al., 2007), relying on an extended Markov Chain Monte Carlo (MCMC) sampling technique to best approximate the posterior distribution for the data association to the set of Kalman filters. In the following, we briefly describe the main steps of the processing chain to understand the parameters used for our test data set. The problem can be subdivided in to two tasks. The first task is to find the correct association of a new observation under the assumption of a known number of targets. This is, generating the optimal importance distribution for all particles according to the Rao-Blackwell theorem:

$$p(d_k \mid z_{1:k}, d_{1:k-1}^{(i)}) \quad \forall particles\ i \tag{3.7.1}$$

which is the probability of the association d_k given the association history $d_{1:k-1}^{(i)}$ and all observations $z_{1:k}$ up to and including time step k. Using Bayes rule, we can derive:

$$p(d_k \mid z_{1:k}, d_{1:k-1}^{(i)}) \propto p(z_k \mid d_k, z_{1:k-1}, d_{1:k-1}^{(i)}) \times p(d_k \mid d_{k-m:k-1}^{(i)}) \tag{3.7.2}$$

where $p(z_k \mid d_k, z_{1:k-1}, d_{1:k-1}^{(i)})$ is given by the clutter density function $p_{clutter} \coloneqq p(z_k \mid clutter)$ for the case of d_k being clutter and by the Kalman filter measurement likelihood evaluation of the Kalman filter associated to d_k otherwise. According to Eq. (3.7.2), we can use sampling techniques to approximate the posterior distribution for the data association.

In the second task, we deal with an unknown number of targets, which are "born", exist for a while and then "die" again. It is realized by allowing a very high number of targets in the model (in theory infinite; in practice the maximal conceivable number). The probability of a target's birth is derived from the clutter probability and the detection probability p_{detect}, and its life time t_d is modeled by a gamma distribution Γ indicating the time a target is active (i.e. the typical time of a target existence).

For the clutter density function $p_{clutter}$, a uniform distribution in the measurement space is chosen and the detection probability p_{detect} is set to 0.6. For the lifetime t_d we use a gamma distribution with a peak at a duration of 1.5s and a smooth gradient allowing a quite wide range of target lifetimes, accounting for the fact that targets moving in the cross-track direction remain in the images for longer periods than targets that move along-track.

The setting of the parameters is done based on prior knowledge and observations, relying also on human expertise and testing to tune the method for optimal performance. The choice of

parameters is optimized for the specific MiSAR experiment and has to be adapted for other sensors. Note that the parameters are set very generically to allow a very general tracking framework. Additional prior knowledge can improve performance further at the cost of generality. This could e.g. be done by a reduction of potential "birth" locations by excluding large forested or other roadless areas.

3.3 Results of the MiSAR experiment

We applied the above method to the MiSAR data set described in section 3.2. Two factors were unfavorable for our experiments. First, no ground truth was acquired during the campaign, implying that only rough estimates of the real speed of a vehicle can be used for validation. Second, the grazing angle of the sensor is low (~18°). The forested areas on both sides of the highway therefore cause shadow and layover effects, significantly impairing the detection of moving targets in these regions. However, the MiSAR sensor with its wide antenna beam and the test area are well suited to the task of tracking moving objects in a sequence of OAwrB images.

One hundred and fourteen objects were detected; the tracks are indicated in Fig. 3.5.

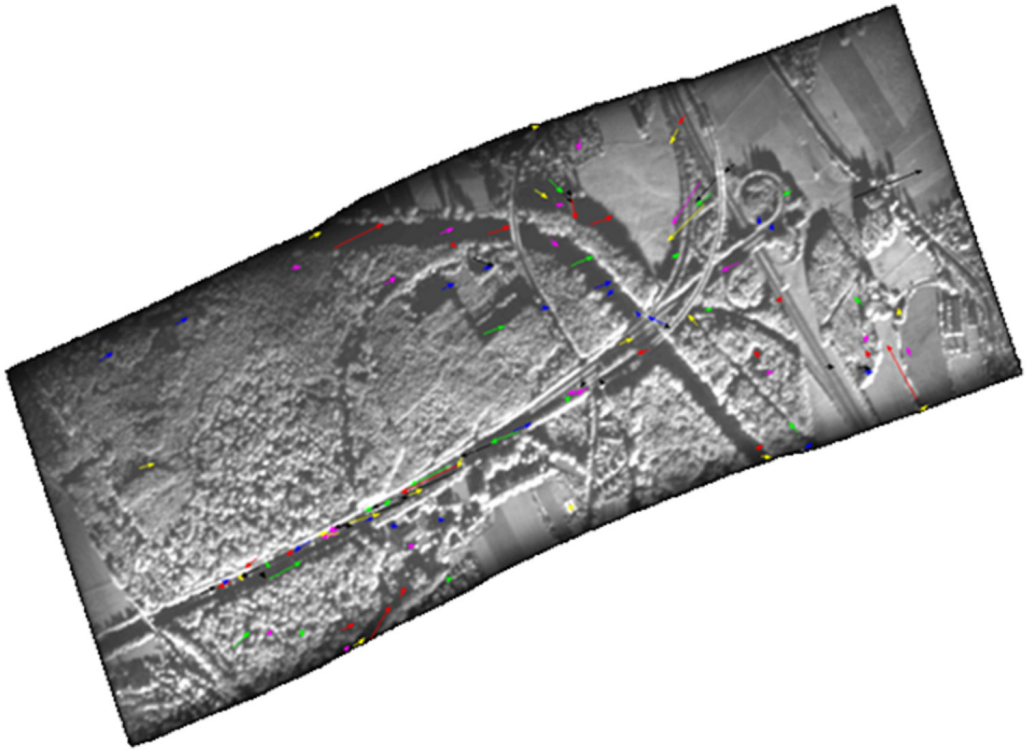


Fig. 3.5 – The output of the 114 detected objects and the associated movements for the complete scene. The arrows indicate the beginning and end of the movements; the usage of different colors helps keep targets separable. An agglomeration of objects along the highways is clear.

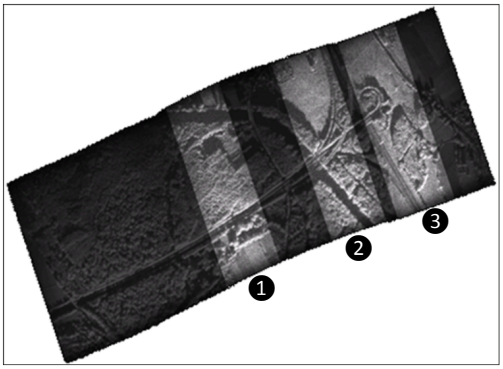
Additional post-processing not discussed in this paper might be able to exclude some artifacts caused mainly by noisy observation at the edges of the images (lower signal-to-noise ratio) and processing artifacts in the OAwrB images. In the following we use the estimated velocities of the vehicles in the *image* space as rough approximations of the true *3D* velocities to evaluate our method without applying any *image to 3D velocity* corrections. This allows meaningful comparisons between image and 3D velocities only for cross-track moving objects. We comment on possible ways to transform the image space velocities distorted by the Doppler effect to

3D velocities later in the paper. The Doppler corrections are not part of this paper, as (a) we do not have ground truth velocities to calculate reliable error bars even after performing a Doppler shift correction, (b) the accuracy of the navigation data is insufficient (no differential GPS (dGPS) data available) and (c) the MiSAR sensor is a high-frequency sensor, thus inducing significant additional uncertainty to *image to 3D space* transformations. The sensor has a wavelength of 8.69 mm and a PRF of 2000 Hz, the Doppler ambiguity equates to a radial velocity difference of only 8.69 m/s. With a slant range of roughly 1500 m, a platform speed of ~35 m/s and an inaccuracy of 1 m/s in radial velocity, the Doppler shift is given by Rüegg et al., (2007):

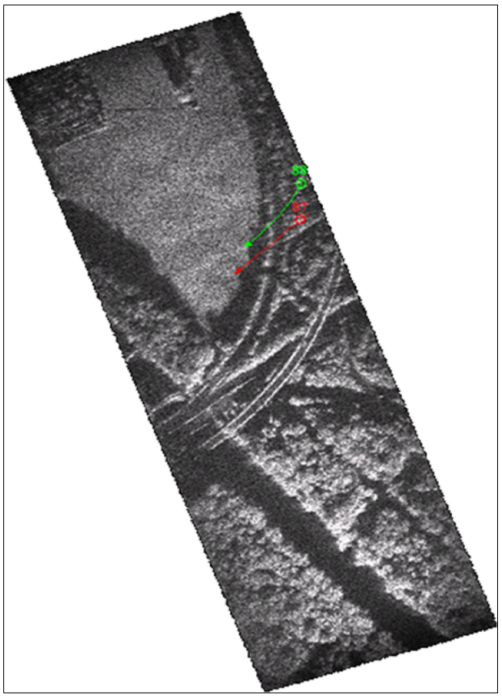
$$\Delta azimuth = -\frac{V_{radial}}{V_{platform}} * \left|\overrightarrow{Range}\right| \approx -43m \tag{3.8.1}$$

The above relation implies that especially for targets moving mainly in the along-track direction, the velocity estimates are distorted, as even a small Doppler shift in the azimuth direction has a large influence on the velocity as the direction of the shift is parallel to the object's direction of movement. For cross-track moving objects, the shift in the absolute position is much larger as the radial velocity of cross-track moving objects is larger. However, its distortion of the velocity between image and 3D space is considerably smaller, as the direction of movement and the Doppler shift are perpendicular and the relative change of the radial velocity (and Doppler shift, respectively) between the first and last detection of the object (in the first and last OAwrB/burst) is smaller than for objects moving opposite to the sensor's movement in along-track direction.

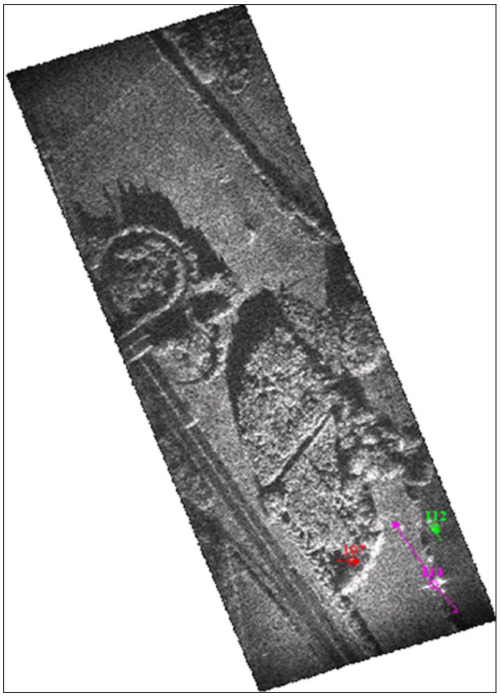
In the following, we discuss three exemplary cases (shown in Fig. 3.6(a)) to demonstrate the algorithm's output in the image space. Fig. 3.6(b) illustrates the trajectories of three moving vehicles (with identification number). Target 114 is moving in the cross-track direction (implying large positional and small velocity distortions introduced between image and 3D space); the estimated image velocity of 28.87 m/s (~103.9 km/h) is a realistic 3D velocity. This object has the longest dwell time (4.70 s). The real world position of object 114 is most probably the highway about 143m west of the detected position in the image space. Fig. 3.6(c) shows two vehicles (ID: 87 and 88) moving in a diagonal direction to the cross-/along-track. They are tracked for 2.40 s and 2.93 s, respectively; their real positions are probably on the on-ramp southeast of their image positions. As they have a non-negligible along-track direction component in their movement, the estimated target velocities are distorted with values around 40 – 45 m/s (144 – 162 km/h). Their image velocities are surely overestimating the 3D velocities for this road section. We obtain even higher values of over 70 m/s (252 km/h) for the vehicles number 47 and 49 in Fig. 3.6(d) moving predominantly in the along-track direction. These values are distorted by a Doppler shift (which is small in absolute values but with larger relative changes between first and last detection and parallel to the moving direction thus directly influencing its velocity estimate). The image positions of the vehicles are only slightly shifted, as the radial velocity component for targets moving along-track is considerably smaller; this is why they are mapped on the highway, probably shifted only slightly to the south.



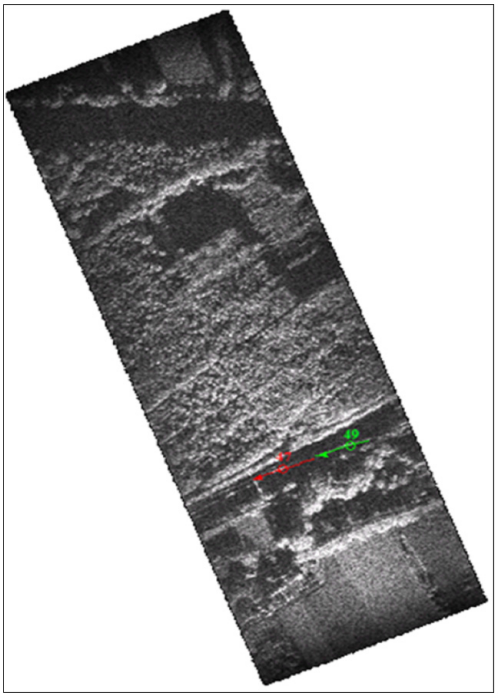
(a) Locations of the three exemplary cases.



(c) Scene 2 of Fig. 3.6(a): Two moving targets with a cross- and along-track direction of the movement.



(b) Scene 3 of Fig. 3.6(a): Three moving targets; the longest movement of object 114 is mainly in the cross-track direction.



(d) Scene 1 of Fig. 3.6(a) with two targets moving in the along-track direction on the highway.

Fig. 3.6 – Three selected cases for the output of the tracking algorithm. The circles indicate the positions at the current time step and the arrows the movement for the time the objects are detected in the beam of the radar sensor.

The results from the test data set seem promising and accurate in the image space; unfortunately, evaluation of the results based on 3D space velocities is difficult due to the sensitivity of

the Doppler shift to the radial velocity because of the high frequency sensor and imprecise navigation data and due to the lack of ground velocity measurements. To give the reader an impression of the accuracy, we summarize the algorithm's output in a visual evaluation (Table 3.1, note: velocities are only given for cross-track targets as they are better approximations of their 3D velocities; misdetections are almost exclusively caused by processing artifacts, e. g. glint points, and noise effects at the edges of the OAwrB).

TABLE 3.1
VISUAL EVALUATION OF THE TRACKING ALGORITHM

TID	TT	MD	MV	R?	TID	TT	MD	MV	R?	TID	TT	MD	MV	R?
1	0.84	21.88	n/a	no	39	0.62	5.07	n/a	no	77	1.07	117.9	n/a	yes
2	0.86	10.01	n/a	no	40	0.45	-2.83	n/a	?	78	0.53	8.64	n/a	no
3	0.9	-2.56	n/a	no	41	0.66	-9	n/a	yes	79	0.7	4.75	n/a	yes
4	0.49	33.48	n/a	no	42	1.64	3.16	n/a	yes	80	3.06	138.12	n/a	yes
5	1.38	7.73	n/a	?	43	1.58	3.24	n/a	yes	81	1.03	130	n/a	yes
6	0.43	2.3	n/a	yes	44	1.31	0.1	n/a	yes	82	0.8	55.07	n/a	yes
7	0.47	48.31	n/a	no	45	0.8	0.89	n/a	yes	83	0.47	118.16	n/a	yes
8	0.45	1.25	n/a	?	46	0.59	6.04	n/a	no	84	1.29	84.27	36.22	yes
9	0.45	96.89	26.44	?	47	0.8	0.38	n/a	yes	85	0.43	-11.99	n/a	yes
10	0.47	34.98	n/a	no	48	2.38	2.73	n/a	no	86	1.19	102.53	7.92	?
11	0.66	4.15	n/a	yes	49	0.74	-1.31	n/a	yes	87	2.4	20.96	n/a	yes
12	0.7	13.14	n/a	yes	50	0.64	2.33	n/a	no	88	2.93	33.16	n/a	yes
13	0.59	100.51	21.09	?	51	0.8	50.49	n/a	yes	89	0.84	18.62	n/a	yes
14	0.7	8.31	n/a	no	52	0.66	1.11	n/a	no	90	0.57	40.39	n/a	yes
15	0.92	-6.94	n/a	yes	53	1.77	143.05	n/a	?	91	1.31	13.34	n/a	yes
16	0.45	-12.91	n/a	yes	54	0.43	128.52	n/a	?	92	0.53	59.44	n/a	?
17	0.53	108.21	n/a	?	55	1.27	-2.34	n/a	no	93	0.88	42.36	n/a	yes
18	0.62	-2.07	n/a	yes	56	0.66	-3.61	n/a	no	94	0.7	3.39	n/a	yes
19	1.99	5.63	n/a	yes	57	0.6	82.23	7.24	no	95	1.48	22.91	n/a	yes
20	0.82	13.36	n/a	yes	58	1.11	-4.82	n/a	yes	96	0.49	133.24	n/a	yes
21	0.84	1.87	n/a	yes	59	0.45	99.87	14.35	yes	97	0.43	0.87	n/a	yes
22	0.47	2.63	n/a	yes	60	1.25	-2.56	n/a	no	98	1.17	68.48	n/a	yes
23	1.7	-1.3	n/a	yes	61	0.62	81.49	14.89	yes	99	0.64	157.15	n/a	no
24	0.9	5.39	n/a	no	62	0.55	-4.12	n/a	?	100	0.51	-13.87	n/a	yes
25	0.47	46.86	n/a	yes	63	0.9	0.63	n/a	no	101	0.55	155.8	n/a	yes
26	0.6	158.05	n/a	?	64	0.43	-2.76	n/a	no	102	0.49	99.73	33.05	yes
27	0.9	13.4	n/a	no	65	0.49	38.91	n/a	yes	103	1.48	118.12	n/a	yes
28	0.43	27.85	n/a	yes	66	0.47	-4.62	n/a	yes	104	0.55	10.06	n/a	?
29	0.92	-16.6	n/a	yes	67	1.42	-0.74	n/a	no	105	0.6	5.32	n/a	?
30	2.2	36.67	n/a	no	68	1.09	-2.02	n/a	no	106	0.7	51.65	n/a	yes
31	0.49	6.77	n/a	yes	69	0.92	-1.25	n/a	no	107	0.74	155.7	n/a	yes
32	0.62	13.52	n/a	?	70	0.66	3.81	n/a	yes	108	0.43	-0.74	n/a	?
33	2.48	-4.95	n/a	yes	71	0.49	87.88	13.65	no	109	1.13	123.92	n/a	?
34	0.64	-8.91	n/a	no	72	1.68	-4.26	n/a	no	110	1.27	-17.27	n/a	?
35	2.16	1.1	n/a	yes	73	1.17	119.75	n/a	yes	111	0.53	81.07	26.76	yes
36	0.6	40.48	n/a	no	74	0.51	25.39	n/a	yes	112	0.99	97.78	19.72	yes
37	0.62	22.27	n/a	yes	75	0.66	116.75	n/a	?	113	4.09	-0.84	n/a	no
38	0.49	151.34	n/a	?	76	0.76	116.62	n/a	yes	114	4.7	102.4	28.87	yes

* TID: Target Identification; TT: Tracked Time [s]; MD: Moving direction [°] with respect to azimuth direction; MV: velocity [m/s] for targets moving within 15° of cross-track direction; R?: visually identified as real vehicle.

We additionally tried to visually identify objects not detected by the algorithm, finding 25. The derived statistics are summarized in Tables 3.2 and 3.3.

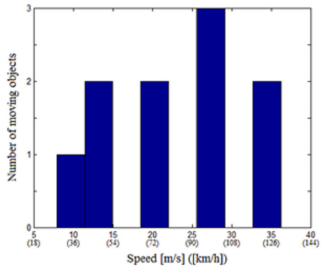
TABLE 3.2
PERFORMANCE OF THE ALGORITHM

		(VISUAL) GROUND TRUTH		
		Moving Target (MT)	Possibly (?)	False detection (FD)
ALGO-RITHM	Detected	63	20	31
	Not detected	25	-----	-----

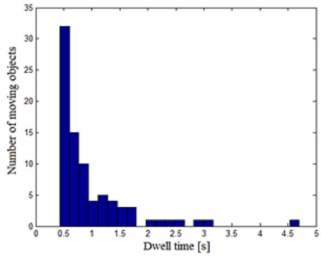
TABLE 3.3
CLASSIFICATION AND FALSE ALARM RATE OF THE ALGORITHM

	(MT + ?) vs. FD	MT vs. (FD + ?)	MT vs. FD
Classification Accuracy	76.9%	71.6%	71.6%
False Alarm Rate	27.2%	44.7%	33.0%

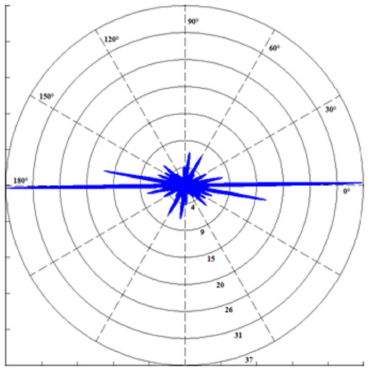
Although this is not an absolute validation of the method, these results with single channel SAR data demonstrate good performance of the algorithm as far as possible in the absence of ground truth data (classification probability of ~72% with a false alarm rate of only ~33%). Furthermore, in Fig 3.7. characteristics of the detected moving objects are illustrated. The distribution of the speeds of cross-track moving objects (varying between 10 m/s and 35 m/s, Fig. 3.7(a)) and the distribution of the direction of movement which is dominated by the coincident highway direction (Fig. 3.7(c)) give additional indications of successful velocity estimations in the image space.



(a) Distribution of speeds for moving objects with cross-range moving direction (comparable with real 3D velocity) from Table 3.1.



(b) Distribution of the dwell time for moving objects from Table 3.1.



(c) Distribution of the direction of movement for moving objects from Table 3.1. Note the large along-track accumulation coincident with the highway.

Fig. 3.7 – Distributions of moving objects' characteristics derived from Table 3.1.

A further way of evaluating the results is by using road models to overcome the problem of the sensitivity of the Doppler shift. We did this for the trajectory of target 114 by projecting the movement of the vehicle onto the south-north running highway. By inverting Eq. (3.8.1), we obtain an average radial velocity of 30.89m/s (varying between 30.83 and 31.02) assuming the most probable Doppler ambiguity. That is almost identical to the averaged values for platform speed, azimuth shift and range:

$$V_{radial} = -\frac{\Delta azimuth * V_{platform}}{|\overrightarrow{Range}|} \approx -\frac{143m * 34.43m/s}{1272m} \approx -3.87m/s. \tag{3.8.2}$$

Due to the Doppler ambiguity, it holds that

$$V_{radial}^{DopplerAmbiguity} = V_{radial} + n * 8.69m/s \tag{3.8.3}$$

and thus the most probable radial velocity of the vehicle on the highway is 30.89 m/s (with n=4). Below, we compare the cross-range component of the estimated speed

$$V_{cross-range}^{estimated} = \sin(102.4°) * 28.87m/s \approx 28.20m/s \text{ (for values see Table 3.1)} \tag{3.8.4}$$

with the cross-range component of the calculated radial velocity:

$$V_{cross-range}^{Doppler\ shift} = \sin(71.35°) * 30.89m/s \approx 29.27m/s \quad (3.8.5)$$

(with an average incidence angle of 71.35°)

obtaining a difference of only 1.07 m/s. The consistency between model and observation is highly encouraging.

3.4 Conclusion & Outlook

We have introduced a tracking method based on multitarget unscented Kalman filters to not only identify moving targets but also estimate their trajectories in radar image space. The proposed algorithm can be practically applied to any single-channel, wide-beam SAR and shows promising results. For cross-range velocities, it was possible to manually validate the velocity with the help of a road model: the speed estimation was within an error margin of ~1 m/s. A more complete validation is impaired by the lack of ground truth data. A further flight campaign with a different sensor is in progress, where ground truth measurements *are* available, which will be used in a more complete validation process. In the newer campaign, along-track interferometric (ATI) data were recorded in parallel, which will provide the opportunity to compare the method developed here with ATI-based approaches. The OAwrB image processing and tracking is computationally intensive, but with high-performance computers and a trade-off in the ratio of the overlap of OAwrB images and minimum detection probability, a quasi real time processing of the data is achievable. A remaining task not addressed in this paper is the translation of an object's position and velocity in the image space to its corresponding real world position/velocity. Existing works in this field (e.g. Kirscht, 2003) can be incorporated to our algorithm in the future. However, as the sensor operates at high frequencies and the navigation data are not based on dGPS measurements, these corrections induce additional uncertainties and thus might not be desirable for some applications. Generally, the image to 3D correction can be done either by including the Doppler shift in the transformation function (Eq. (3.6.4.)), by post-processing, or by introducing additional prior information. The first solution implies additional parameters in the tracking model and consequently a less robust detection capability. The second approach for the image to 3D corrections has the advantage that it has no negative influence on the detection accuracy, but its estimation of the 3D position and velocity is less accurate and, in some cases, needs manual intervention to resolve Doppler ambiguities. The most precise method to transform image space estimates of moving targets into 3D space is by using additional prior information, most suitably road models. This allows a robust solution for the corrections even with high frequency sensors. As the main output of the paper, we were able to demonstrate with real data a novel method for tracking moving objects in single-channel wide-beam SAR systems with high accuracy in the image space and flexibility concerning the sensor and object moving directions (cross-track as well as along-track) and speeds.

Acknowledgments

The authors would like to thank the MiSAR team at EADS for their cooperation and for providing the raw data for the experiments, the anonymous reviewers for their valuable comments that improved this paper and armasuisse W+T for funding the work. This work was conducted at the Remote Sensing Laboratories, University of Zurich, Zurich, Switzerland.

4 MOVING TARGET TRACKING IN SINGLE- AND MULTICHANNEL SAR

This chapter has been submitted as: Henke D., Mendez Dominguez E., Small D., Schaepman M.E., & Meier E. (2013). Moving target tracking in single- and multichannel SAR. *IEEE Transactions on Geoscience and Remote Sensing*.

Abstract

Methods for moving target tracking in single- and multi-channel Synthetic Aperture Radar (SAR) are presented and evaluated using data from the F-SAR system. An approach based on temporal and spatial overlapping SAR images, image statistics and multitarget unscented Kalman filtering showed promising results for a single-channel linear SAR experiment. Here we extend this approach to integrate along-track multi-channel data, combine single- and multi-channel extraction methods and verify our results on non-linear test data on all the different extraction methods. The results are validated and compared with ground-based measurements on a highway section, demonstrating the accuracy of the presented method.

4.1 Introduction

Ground moving target indication (GMTI) in SAR data addresses the task of extracting information about moving objects (Brenner & Ender, 2006; Lipps et al., 2004). Compared to other GMTI systems, SAR sensors can indicate moving objects *while* simultaneously imaging the area of interest. Furthermore, SAR data can be acquired almost independently of weather conditions and daylight. These SAR-specific properties enable their deployment over a wide range of target tracking applications from traffic monitoring (Baumgartner and Krieger, 2012; Cerutti-Maori et al., 2012) to surveillance (Yan et al., 2013) in uncharted areas.

However, since SAR was originally developed for the use in static scenes, the problem of extracting information of moving objects is a challenging task which has evolved over the last decade and demands state-of-the-art processing techniques. Most of the traditional GMTI approaches in SAR rely on multi-channel data e.g. space-time adaptive processing (STAP) (Cerutti-Maori et al., 2012; Ender, 1999), along-track interferometry (ATI) (Chapin and Chen, 2009), monopulse processing (Rüegg et al., 2007), displaced phase centre array (Gierull and Sikaneta, 2003) or a combination (Gierull et al., 2013; Suchandt et al., 2010). However, many of these approaches impose restrictions on the acquisition, such as linear SAR sensor flight tracks. In Henke at el. (2012b), we presented an approach which – compared to traditional SAR GMTI – is capable of not only indicating the presence of moving targets, but also the extraction of an object's movement while being observed by the SAR sensor by making use of temporal processing methods. The approach relies on image statistics, unscented Kalman filtering and multi-target hypotheses. In contrast to the aforementioned traditional GMTI methods, it needs only *single-channel* SAR data as input. This can be especially interesting for low-cost, light-weight SAR systems.

However, given the availability of multi-channel data, additional information content can be used to harness the advantages of both, traditional multi-channel GMTI methods and the tracking approach employed in Henke et al. (2012b) to improve GMTI performance. Therefore, in this paper we demonstrate two major extensions to the single-channel algorithm (Henke et al., 2012b): special emphasis is placed on the multi-channel extension, i.e. integrating traditional multi-channel methods, here namely ATI, into the multi-target tracking method (see also Fig. 4.2, green box). Second, we demonstrate the generality of the tracking method by testing it using a highly non-linear data acquisition mode (by using a circular flight track) and both single- and multi-channel data.

The data set was recorded using the airborne F-SAR sensor (Horn et al., 2009) in ATI mode based on a circular acquisition geometry. We processed the data with many highly "overlapping apertures with reduced bandwidth" (OAwrB) using a time-domain back-projection (TDBP) algorithm (Frey et al., 2009) implemented efficiently on a graphics processing unit (GPU). Subsequently, moving object candidates were extracted at each time step and tracked over time by analyzing their dynamic behavior. For the candidate-selection step, we rely on the extraction method for single-channel data as described in Henke et al. (2012b), and on a multi-channel ATI extraction method based on a generalized likelihood ratio test (GLRT) commonly used with ATI data (Budillon et al., 2012; Kohlleppel and Gierull, 2008). In addition to the purely single- and multi-channel extractions, there are several conceivable methods to combine the results of both extractions to form a new set of moving target candidates: a few representative approaches are described. Afterwards, the tracking algorithm is applied to the extracted potential moving objects (observations) and, based on an underlying physical model, one obtains an estimate of the objects' most probable movements. For this task, a multi-target unscented Kalman filter proposed by Särkkä et al. (2007) is applied. In addition to providing trajectory information for moving objects, it significantly suppresses noisy observations (in ATI as well as in single-channel observations) by analyzing the dynamic behavior of moving object candidates. Finally, the trajectories are mapped from the SAR image geometry into real world or map coor-

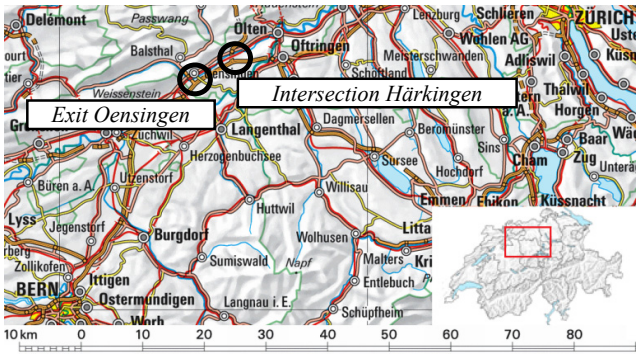
dinates, respectively. The resulting objects' trajectories are validated by comparing them with ground-truth measurements on a section of the highway. We found that with a combined single-/multi-channel observation set, a detection rate of 93% with a false alarm rate of only 14% can be achieved for large vehicles. Depending on the applied "SAR to real world"-projection, average position accuracies with the best mapping approach were off by 10 m, with an average speed uncertainty of less than 1.2 m/s.

In sections 4.2.1 & 4.2.2, we briefly introduce the F-SAR sensor, its specifications and the test sites, and give an overview of the processing chain. Then the method is described in detail. Starting with the TDBP processed OAwrB images (4.2.3), we focus on the extraction of the potential moving objects in section 4.2.4. These potential moving objects serve then as observations for the multi-target tracking approach to obtain the trajectories of moving objects over time in the SAR image space (4.2.5) and transform them into map coordinates (4.2.6). The validation methods are introduced in 4.2.7. In section 4.3.1 & 4.3.2, the extracted observation sets are used to test the method in highly non-linear flight configurations. Results are intercompared and validated using ground measurements acquired during the overflight. In section 4.4, we conclude with a detailed analysis of the results and a discussion of potential future improvements and applications.

4.2 Methods

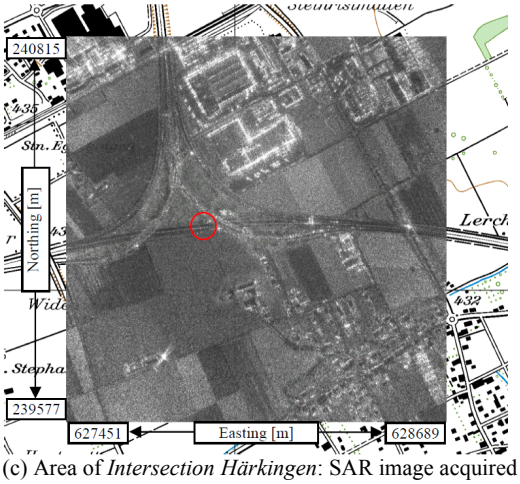
4.2.1 Test site and sensor

The SAR sensor utilized in the experiments is the F-SAR system developed by DLR (Horn et al., 2009; Reigber et al., 2011). The sensor is mounted on a Dornier Do-228 aircraft, providing different operation modes (multi-frequency, fully-polarimetric, interferometric, etc.): we make use of the X-band ATI mode at 9.6 GHz. This mode comprises four equidistant receiving antennas (channels) and one transmitting antenna operating with a pulse repetition frequency of 2016 Hz at HH polarization. The receiver antennas are separated by approximately 20 cm, with an azimuth beam width of 7.3° each. For a single OAwrB image processed with 700 pulses, the theoretical resolution is 1.5 m in range, and 2.2 m in azimuth. Fig. 4.1 shows the focused geocoded single-channel SAR images (area of interest ~1.25 x 1.25 km^2 located in the center of the circles) overlaid on topographical maps of the two test sites *Oensingen* and *Härkingen.* The data was focused using time-domain processing with a square grid-pixel spacing of 40 cm. The study areas for the traffic monitoring were a highway intersection in *Härkingen* and a highway exit in *Oensingen*, located on the Swiss plateau. The circular flight paths have an approximate diameter of 3.5 km, a mean altitude above ground of 2.7 km, and an average platform velocity of 76 m/s.



(a) Overview of the test sites.

(b) Area of *Exit Oensingen*: SAR image acquired with a circular flight track overlaid on topographic map. The red circle indicates the location of ground truth measurements (camcorder on the bridge)



(c) Area of *Intersection Härkingen*: SAR image acquired with a circular flight track overlaid on topographic map. The red circle indicates the location of ground truth measurements by METAS (radar and infrared speedometer)

Fig. 4.1 – Test sites.

4.2.2 Overview of the processing chain

The proposed method for tracking moving targets in SAR data can be subdivided into four major parts. An overview of the processing chain can be found in Fig. 4.2. The first task is to process and geocode the SAR data, producing a set of OAwrB images. The tracking algorithm works independently of the focusing algorithm. We use a TDBP algorithm to focus the OAwrB images, as it is one of the most precise methods and can handle non-linear flight tracks at high quality. Having obtained a sequence of images, we collect image statistics and use image processing methods to discern potential moving object candidates in each image; their temporal behavior is ignored at this stage. For the candidate-selection, we rely on two different approaches: one based on single-channel data as in Henke et al. (2012b) and the other based on multi-channel ATI data. In the third processing step, we use a multi-target unscented Kalman filter approach (Särkkä et al., 2007) incorporating a dynamic model to associate the extracted time-independent observation of step two to generate moving target trajectories of maximum probability. Finally, the projection of the trajectories from the SAR geometry into 3D real world coordinates is performed for two cases of *a priori* information: with and without the existence of a road network model.

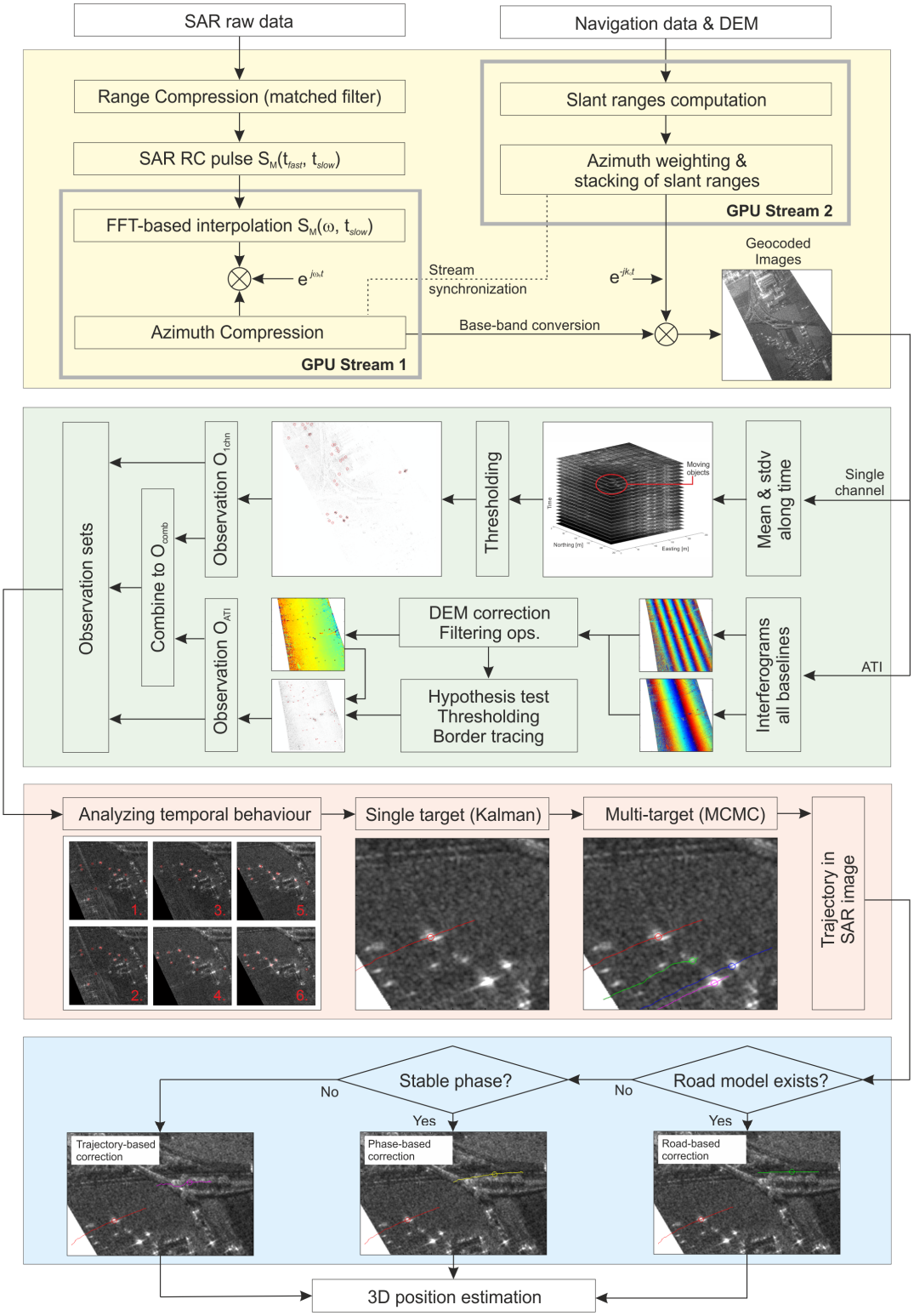


Fig. 4.2 – Overview of the processing chain in four major steps: (1) focusing geocoded images (yellow), (2) calculating features (green), (3) estimating trajectories in SAR image space (red) and (4) projecting the trajectories from the SAR image space to real 3D position (blue).

4.2.3 Processing the raw data

4.2.3.1 TDBP-based image focusing

The aim of the first processing step is to generate a stack of OAwrB images with high accuracy in both, focusing quality and geolocation precision. Although the exact inversion of the SAR equation is not guaranteed for non-linear SAR geometries (Ulander et al., 2003), the TDBP algorithm offers high quality focusing at the expense of a relatively high computational burden compared to approaches that use frequency-domain algorithms. Without accounting for the radiometric terms and considering base-band signals, TDBP calculates the geocoded image $OAwrB(E_o,N_o)$ as follows (Eq. 4.1):

$$OAwrB(E_0,N_0) = \left[\sum_{t_{slow}=1}^{L} S_M\left(\frac{\sqrt{(X(t_{slow})-x_0)^2+(Y(t_{slow})-y_0)^2+(Z(t_{slow})-z_0)^2}}{c}, t_{slow}\right) e^{-i2\pi f_c t_{fast}}\right] e^{-i2k_c R(x_0,y_0,z_0)} \quad (4.1)$$

where (E_0,N_0) denotes the map coordinates of a backscattering element, which are transformed to the equivalent global Cartesian coordinates (x_0,y_0,z_0) using a Digital Elevation Model (DEM). $(X(t_{slow}),Y(t_{slow}),Z(t_{slow}))$ is the sensor position in global Cartesian coordinates at slow time instant t_{slow}, k_c is the demodulated frequency wavenumber at carrier frequency f_c, c is the speed of the light, L is the aperture length, t_{fast} is the corresponding fast time and S_M is the base-band range compressed pulse at slow time instant t_{slow}. TDBP requires an interpolation of the pulse currently being projected to properly reconstruct the SAR image (Soumekh, 1999): we make use of the classical zero padding FFT interpolation method, but different interpolation kernels could also be utilized.

4.2.3.2 GPU-based computational procedure

A Central Processor Unit (CPU) does not exploit the inherent parallelism of the algorithm due to its typically serial operation. From Eq. (4.1), one observes that the azimuth compression (projection or aperture integration) of the range bins is independent for each pixel of the DEM, and can be executed in parallel using a GPU. Employing the NVIDIA-CUDA C toolbox (Sanders and Kandrot, 2011), the total computation time of the TDBP is considerably reduced depending on the number of GPU-cores. Moreover, to provide a maximum speed-up factor, we exploit an asynchronous execution between the CPU and the GPU using a two-streamed GPU TDBP SAR processor as shown in Fig. 4.2 (top, yellow box).

In our approach we process L-length OAwrB images with K overlapped pulses among consecutive OAwrBs (L=700, K=500). Considering the linearity of the back-projection, and in order to remove the dependency on the overlapping factor for the total computation time, we use intermediate matrices with L-K projected pulses, such that after combination they form the current L-length OAwrB image. Following this approach, the total computation time is comparable to a standard single look processing of the full length of the (circular) SAR image.

The geometric accuracy of the data was validated utilizing four bright static point-like targets. The obtained relative geo-location errors vary with a standard deviation of 12 cm, whereas the mean geo-location error between consecutive OAwrBs was observed to be ~4 cm.

4.2.4 Extraction of moving target candidates

After focusing the raw data to OAwrB images, the next task is to extract potential moving objects, ignoring their temporal behavior for the moment. One way of doing so is to use only single-channel data as in Henke et al. (2012b) which for completeness we summarize in the following paragraph briefly, as it is used in the later experiments. The second possibility is based on ATI moving target indication, described later in more detail. Later, we suggest methods for combining the two extraction methods to take advantage of both approaches.

4.2.4.1 *Single-channel extraction*

For the single-channel data, we build on the single-channel extraction method presented in Henke at el. (2012b). Departing from conventional linear data acquisitions as in Henke at el. (2012b), this paper treats data acquired in a circular flight geometry, leading to a longer dwell time for backscattering elements and therefore, an improvement of the estimated statistics and probability of detection for moving object candidates. The longer dwell time compensates for the narrower beam width of the F-SAR sensor in comparison to the MiSAR sensor treated in Henke at el. (2012b). The processing steps for the single-channel extraction from the generated OAwrB image stack over time consist of a Gaussian smoothing and a subsequent calculation of the mean and standard deviation of each pixel's intensity through time (along the z-axis). The final detection mask is generated by appropriate thresholding and morphologic filtering (for more details on these processing steps, see Henke at el., 2012b). The intermediate results of the processing chain of the single-channel extraction are illustrated in Fig. 4.3.

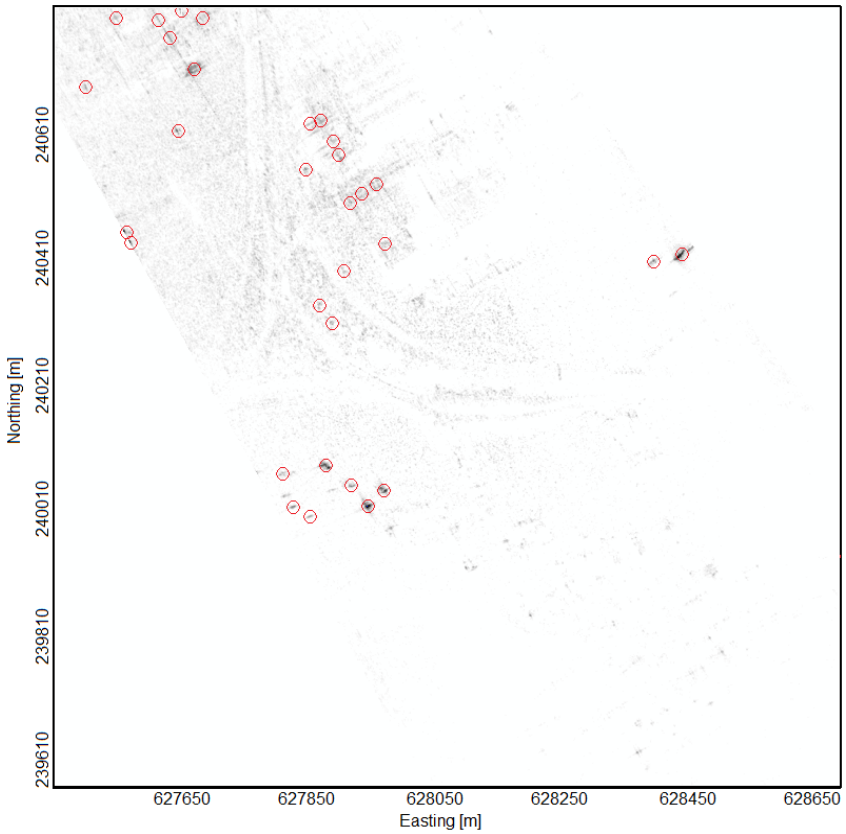


Fig. 4.3 – Single-channel detection mask before thresholding. The red circles mark the extracted observation set O_{1chn} at the corresponding time step.

4.2.4.2 *ATI extraction*

In the presence of ATI multi-channel data, the extraction of moving object candidates can be based on the fact that *moving* objects induce a phase shift between the acquisitions of two consecutive channels while *static* objects do not. The phase shift $\Delta\varphi$ is given by

$$\Delta\varphi = -\frac{4\pi}{\lambda} \cdot V_m \cdot \tau \mod 2\pi \tag{4.2}$$

where λ is the wavelength of the sensor, τ is the time lag between the two acquisitions and V_m is the radial velocity of the moving object. For a constant platform velocity on a linear flight track, the time lag τ is directly proportional to the spatial distance of the two antennas. It is obvious from Eq. (4.2) that for static objects ($V_m = 0$), no phase shift is measured.

Methods to detect moving objects in multi-channel data were developed in recent years and rely mainly on either STAP (Cerutti-Maori et al., 2012; Ender, 1999) or ATI (Chapin and Chen, 2009; Budillon et al., 2012). For slowly moving targets, the performance of the two approaches was compared in Oriot et al. (2008) and found to be similar. We use an ATI-based method; however, with minor modifications, also a STAP approach could be integrated for the detection step instead.

In Eq. (4.3), the phase shift $\Delta\varphi$ can be calculated via

$$\Delta\varphi = atan2(Im(I), Re(I)) \tag{4.3}$$

i.e. the four-quadrant inverse tangent (arc tangent) of the imaginary and real part of an interferogram I.

Having four channels enables calculation of pair-wise interferograms using all four geocoded SLCs of the OAwrB images (Eq. 4.4):

$$I\,(rx_a, rx_b) = geoSLC(rx_a) * geoSLC(rx_b)^* \tag{4.4}$$

with rx_a and rx_b representing the receiving antennas and $I(rx_a, rx_b)$ the interferogram of the antenna pair calculated by a pixel-wise multiplication of the geocoded SLC of the first antenna "*a*" with the complex-conjugated of the geocoded SLC of the second antenna "*b*". Depending on the spatial distance between the chosen pair's along-track baseline, the time lag τ is smaller or larger implying either fewer phase ambiguities at the cost of lower velocity resolution (given a short baseline), or higher velocity resolution at the cost of more phase ambiguities (given longer baselines). Fig. 4.4(a) & 4.4(b) compare the interferogram from a short baseline with one from a long baseline. The interferometric phase is unwrapped using

$$\Delta\varphi^{unwrapped}_{A_x,A_y} = \left\lfloor \frac{1}{2\pi}\left(\frac{d_{A_x,A_y}}{d_{A_x,A_z}} * \Delta\varphi^{unwrapped}_{A_x,A_z} - \Delta\varphi_{A_x,A_y}\right)\right\rfloor \cdot 2\pi + \Delta\varphi_{A_x,A_y} \tag{4.5}$$

with $d_{Ax,Ay} > d_{Ax,Az}$

where $d_{A\ldots,A\ldots}$ is the distance between an antenna pair, $\Delta\varphi_{A_\ldots,A_\ldots}$ is the phase difference of the interferogram of the antenna pair, and the superscript "*unwrapped*" indicates the phase difference after unwrapping. Thus, it combines the advantages of the shorter and longer baselines. Multi-channel ATI has two main advantages in comparison to dual-channel ATI: (a) a better velocity resolution with fewer phase ambiguities, and (b) use of all antenna pairs reduces phase noise effects, improving the detection rate. Eq. (4.5) is applied recursively starting with the least sensitive shortest baseline (i.e. smallest antenna distance d_{A_x,A_z}) and ending with the longest baseline d_{A_x,A_y}, initialized with $\Delta\varphi^{unwrapped}_{A_x,A_z} := \Delta\varphi_{A_x,A_z}$. Using this formula, many phase jumps that give rise to ambiguities in data acquired with longer baselines can be resolved. Systematic phase jumps existing at the shortest baseline are suppressed using range information derived from the flight track positions and the DEM (Fig. 4.4(c)). Remaining noise effects were handled by considering spatial dependencies in the interferograms of all antenna pairs. This was done by utilizing the likelihood ratio test commonly used to identify if a moving target is present in the ATI data: hypothesis H_0, "no moving target is present" is tested against hypothesis H_1, "a moving target is present" as in Eq. (4.6).

$$\begin{aligned} &H_0: Z = C + N && \text{No moving target present} \\ &H_1: Z = C + N + S && \text{Moving target present} \end{aligned} \tag{4.6}$$

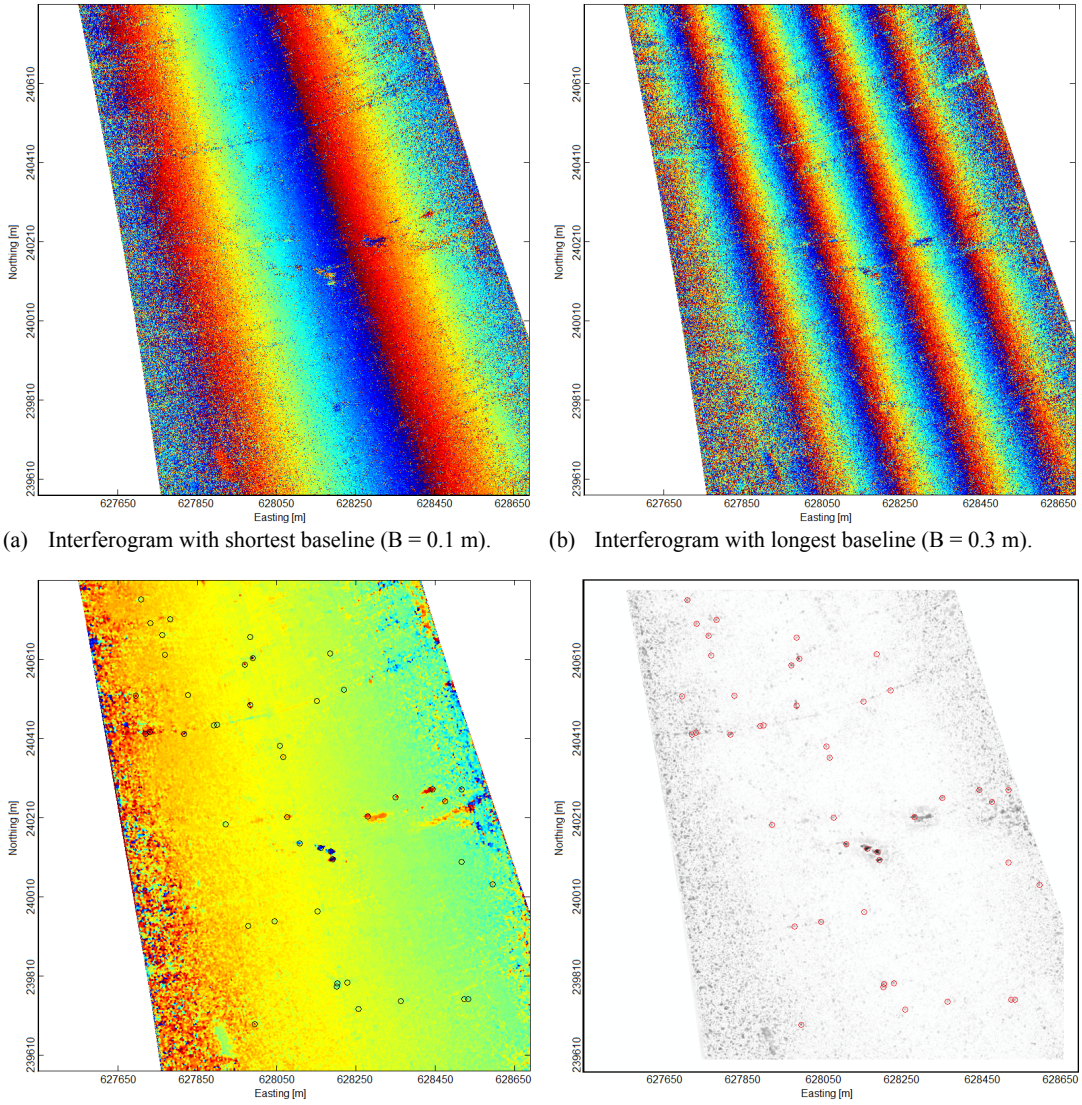
where Z is a vector containing for each specific pixel the complex corresponding values of the interferograms at this pixel. Thereby, Z is modeled with a clutter component C (caused e.g. by scattering behavior), a noise component N (e.g. caused by thermal noise) and a potential signal component S contributed by the moving target. Subsequently, the ratio $\Lambda(Z)$ of the likelihood of hypothesis H_1 and the likelihood of hypothesis H_0 for a given vector Z is calculated via Eq. (4.7):

$$\Lambda(Z) = \frac{L(H_1|Z)}{L(H_0|Z)} \text{ with } L(H|Z) \text{ being the likelihood function} \tag{4.7}$$

If $\Lambda(Z)$ is larger than a certain threshold k_Λ, then H_1 is accepted, otherwise rejected.

C and N are modeled as having a Gaussian distribution with zero mean $\mu_C = \mu_N = 0$, and standard deviations σ_C and σ_N respectively. While C is spatially correlated, N is uncorrelated noise. In contrast, S is distributed with a non-zero mean value $\mu_S \neq 0$. The parameters of the probability distributions can be estimated from the data: given an accurate DEM, smooth topography and reliable data calibration, the disrupting effects ($C + N$) can be summed up in one normal distribution $\mathcal{N}_{R=C+N}(\mu_{C+N} \approx 0, \sigma_{C+N})$ in an area with adequate spatial dimension. This means one can estimate the parameters μ_{C+N} and σ_{C+N} using samples in an appropriate area given a sufficient number of neighboring pixels and data from all channels. It can be efficiently realized by two linear filtering operations with appropriate sample sizes, the first to derive the mean μ_{C+N} and the second the standard deviation σ_{C+N}. The selected sample area should neither be too large nor too small; we used an area of 250 x 250 pixels (~100m x 100m) in our experiments. When the areas are too large, $\mathcal{N}_{R=C+N}$ may lose spatial correlation in certain parts. When sample areas are too small, not enough samples are available for a robust estimate: unavoidably, moving target signal components will also be included in the estimate since these are not known in advance. But for a sufficiently large sampling area, the realistic assumption holds that the number of moving targets is significantly smaller compared to the static background, and therefore can approximately be regarded as negligible. The probability distribution of the moving target signal component can be derived from manually identified moving target samples. Having parameterized both probability distributions, the optimal threshold value k_Λ can be derived by minimizing the false alarm probability (Gaussian classifier). In practice. In practice, we use a slightly lower semi-empiric threshold value k_Λ^{new}, i.e. $k_\Lambda^{new} < k_\Lambda$, artificially increasing the number of pixels indicated as moving target candidates. This is done because k_Λ is optimized for single detections at one specific time instance. However, in the latter, we use Kalman filtering techniques to analyze the dynamic behavior of the moving target candidates: false detections in the single-time detection step are discarded if they do not show a realistic dynamic behavior over time. Consequently, the aim is to guarantee a high detection rate at the single-time detection step and thus, downgrade the importance of a low false alarm rate at this point in the processing chain – we account for this by using a slightly lower threshold k_Λ^{new}.

Subsequently, objects and their properties (like the center of mass) are extracted from this binary ATI mask using morphological open and close operations as well as border tracing as described for the binary single-channel mask (details in Henke et al. 2012b). The resulting centers of the objects of the multi-channel detection are plotted in Fig. 4.4(d).



(a) Interferogram with shortest baseline (B = 0.1 m).

(b) Interferogram with longest baseline (B = 0.3 m).

(c) After terrain-correction, filtering and phase-unwrapping.

(d) ATI detection mask before thresholding. The red circles mark the extracted observation set O_{ATI} at the corresponding time step.

Fig. 4.4 – ATI-based extraction of moving target candidates.

4.2.4.3 Combining single- and multi-channel detections

Having proposed both single- and multi-channel detection methods, the individual observation sets O_{1chn} and O_{ATI} contain the centers of mass extracted with their respective approach. The combination of both methods gives additional sets of observations fostering advantages of both approaches. Two straightforward ways of combining single- and multi-channel data are introduced and compared to single-channel and multi-channel only detections: one is based on a distance metric and the other on an adjustment of the thresholds α and k_A^{new}.

The metric-based approach takes the two observation sets calculated from the single- and multi-channel data as input, and clusters the data points according to the distance between the centers of mass: the observation pair with the shortest distance is linked together and deleted from the list iteratively until a distance-threshold d_{max} is exceeded; we chose d_{max} = 10 m in our experiments as a realistic approximation of typical maximum vehicle dimension. If the distance between a single- and a multi-channel observation is larger than d_{max}, we assume that the single-

and multi-channel signal originate from two different objects. As a result of this procedure, one obtains a list of observations that can be found in both observation sets, and assembles two observation lists – one from the single-channel and one from the multi-channel observations, where no observation was found based on the other extraction method within a distance smaller than d_{max}. Thus, two new combined observation sets can be formed: (a) the union set of both extractions $O_{ATI \cup 1chn}$ containing observations found in at least one of the two extraction methods, and (b) the intersection set of both extractions $O_{ATI \cap 1chn}$ containing only observations found in both extraction methods.

The other combination procedure involves an adjustment of the two thresholds α and k_{Λ}^{new}. While for the binary masks of single- and multi-channel observations the values are set to 1 if a value exceeds the threshold and 0 otherwise. For the combined case we demand that one value must exceed the associate threshold and the other must exceed at least a pre-defined fraction ε (Eq. 4.8):

$$B_{combined}(m,n,t) = \begin{cases} 1 & \{[\Lambda(Z(m,n,t)) > k_{\Lambda}^{new}] \wedge [\theta(m,n,t) > \alpha * \varepsilon]\} \\ & \vee \\ & \{[\Lambda(Z(m,n,t)) > k_{\Lambda}^{new} * \varepsilon] \wedge [\theta(m,n,t) > \alpha]\};\ 0 \le \varepsilon \le 1 \\ 0 & otherwise \end{cases} \tag{4.8}$$

Thereafter, objects and their properties (like the center of mass) are extracted from the combined mask $B_{combined}$, using morphological open and close operations as well as border tracing as described in Henke et al. (2012b) for the single-channel mask. Depending on the choice of ε, one obtains the observation set $O_{ATI+1chn}^{\varepsilon}$. Given $\varepsilon = 0$, the observation set is similar to the union set from above $O_{ATI+1chn}^{\varepsilon=0} \approx O_{ATI \cup 1chn}$. Given $\varepsilon = 1$, the observation set is similar to the intersection set from above $O_{ATI+1chn}^{\varepsilon=1} \approx O_{ATI \cap 1chn}$.

Up to this point, the detection scheme has ignored the temporal behavior of potential moving objects. To include the dynamic behavior of the objects over time, the observation sets O_{1chn}, O_{ATI}, $O_{ATI \cap 1chn}$, $O_{ATI \cup 1chn}$ and $O_{ATI+1chn}^{\varepsilon}$ are passed to the tracking algorithm described in the next subsection to extract realistic moving target trajectories.

4.2.5 Multi-target tracking

In this processing step, two major benefits are realized by modeling the temporal behavior: first, false detections (caused primarily by SAR noise effects in the extraction method at single time steps) can be significantly reduced as in general they do not follow a realistic movement pattern over time. Second, by tracking the dynamic behavior over time, one not only detects moving objects but also obtains their trajectories in the "SAR image space"[6].

A broad overview of multi-target tracking algorithms is provided in Bar-Shalom et al. (2011). In this paper, we rely on an advanced multi-target tracking algorithm based on the theory of unscented Kalman filtering (see e.g. Khairnar et al., 2007) and sequential Monte Carlo methods (e.g. Doucet et al., 2001) in general, and on a combination of the two for multi-target tracking (Särkkä et al., 2007). It outperforms several classical multi-target tracking methods in

[6] When using the term "SAR image space" we refer to *geocoded SAR images* in map geometry for which the *retrieved trajectories* of the moving objects *are initially different* from their corresponding *real world* trajectories due to the SAR focusing process and the induced Doppler-shift. Note that the *range-azimuth SAR images* commonly referred to as "SAR image space" in SAR literature (e.g. Kropatsch and Strobl, 1990; Van Zyl et al., 1993) are not referred to in this context.

accuracy and computational efficiency (Särkkä et al., 2004). The adaptation to SAR multi-target tracking is described in detail in Henke et al. (2012b). It consists of two major parts: (a) tracking a single object using unscented Kalman filtering and (b) associating observations from multiple targets to the "correct" tracking history using Rao-Blackwellized particle filtering (Särkkä et al., 2007).

For (a), the state $x = (pos_{east}, pos_{north}, V_{abs}, \varphi)^T$ of an object, consisting of the SAR map coordinates pos_{east}, pos_{north} of the object [m], the scalar velocity V_{abs} [m/s] and the direction of movement φ [°] is updated by incoming observations and propagated forward in time based on an underlying dynamic model. To minimize the number of the degrees of freedom in the model, the object's motion contains neither acceleration nor rotation events; they are account for via variability in the process noise. More details and equations of the definition of the dynamic model and the input parameters can be found in Henke et al. (2012b). Notable differences in the input parameters are the velocity V_{abs} initially set to 22 m/s (79.2 km/h) with a standard deviation of 10 m/s (36 km/h) due to a slower expected average velocity in the area of interest and the observation noise, where we approximate a standard deviation of 3m in both directions (easting and northing) from the data due to higher absolute geo-location accuracy.

For (b), the problem of associating a new observation to the correct tracking history in the presence of multiple targets, clutter, and misdetection is solved based on Rao-Blackwellized particle filtering (Särkkä et al., 2007) to best approximate the *a posteriori* distribution for the data association to the tracking history. Again, details of the equations and input parameters can be found in Henke et al. (2012b). Note that the values set for the probability of clutter $p_{clutter}$ and of detection p_{detect} vary depending on the specific observation set used, i.e. for the union set $O_{ATI \cup 1chn}$, $p_{clutter}$ and p_{detect} are highest, and for the intersection set $O_{ATI \cap 1chn}$ are lowest of all.

4.2.6 From SAR image measurements to real world measurements

The trajectories are extracted in SAR image geometry. This means, compared to the real world, the objects' positions and thus also their velocities are distorted due to radar-specific effects, e.g. Doppler shift. To transform the trajectories from SAR geometry to real world coordinates, several solutions exist. In this work, we correct the trajectory (a) according to the estimated radial velocity derived from both the single-channel information and the phase from the multi-channel data, and (b) by integrating road models, i.e. projecting the vehicle's positions along iso-range lines on the respective road (i.e. intersection with the road network). For (a), the correction is possible without any prior knowledge. The positions are corrected on the iso-range lines according to the Doppler induced azimuth shift as in Eq. (4.9):

$$\Delta azimuth = -\frac{V_{radial}}{V_{platform}} * R \tag{4.9}$$

where the platform velocity $V_{platform}$ is known from the navigation data, and the slant range R is derived from the vehicle's position in the SAR image and the platform position. V_{radial} has to be approximated either from the single-channel track or from the phase shift in the ATI data. The former has the advantage that the estimation is more robust and available for all vehicles' trajectories; the drawback is a lower precision compared to the phase measurements. The phase derived velocity, however, can be corrupted by noise and clutter effects and is therefore not as reliable. The non-linearity of the platform's flight path and the antenna calibration (e.g. baseline between antenna phase centers, antenna effects, vibration, …) add instability to the experimental SAR system, increasing the uncertainty in the velocity estimation from the phase information. The most accurate results are thus achieved integrating information from a road model. The moving objects are projected along the iso-range lines onto the road network, and the road where the corresponding velocity has the highest probability is chosen. While assigning the most probable road to a moving object, its velocity and position can be corrected in an accurate

manner. Obviously, the tracking algorithm itself could also be improved with such prior knowledge of the underlying road network for the specific scene; however, we want to demonstrate the functionality even without a road model. The projection on the road model therefore serves here as a ground truth validation.

4.2.7 Validation method

For evaluation purposes, two ground truth measurements were acquired during the overflights of both test sites. For the test site "*Oensingen*", a video was captured from a bridge over the highway with a conventional camcorder and an associated GPS clock; for the test site "*Härkingen*", a ground-based radar speedometer with a digital picture from an optical camera was deployed by the Swiss Federal Institute of Metrology (METAS), producing for each passing vehicle an image with speed information and associated time stamp. The "*Härkingen*" data provide the main basis for the evaluation, as it can be used for both the evaluation of the tracking reliability and the later evaluation of the derived vehicle velocities.

The ground-truth camcorder of the test site "*Oensingen*" was for logistical reasons several hundred meters outside the center of the circular flight track (see Fig. 4.1(b)) and thus only partially covers the SAR data. During this time of about 20 seconds, five trucks were recorded with the camcorder, four traveling in the direction Zurich-Bern, and one in the direction Bern-Zurich.

For the test site "*Härkingen*", the ground measurement station was placed in the center of the circular flight track (see Fig. 4.1(c)) and covered the entire SAR illumination time of 130 seconds. Unfortunately, during 20 seconds of the flight track the ground-based measurements were interrupted. Only the traveling direction Zurich-Bern was recorded by the METAS speedometer device. During this time, information on nine trucks was collected; 29 cars detected by the ground-based radar speedometer during the overflight were not considered due to the limited spatial resolution of the SAR system. The output of the METAS system "Robot Multiradar CD" has a 1-second precision and a high velocity accuracy estimating ranges of [8.3m/s, 55.6m/s] ([30km/h, 200km/h]) within an interval of [-0.35m/s, +0.18m/s] ([-1.27km/h, +0.63km/h]) and 95% confidence based on extensive tests by METAS.

4.3 Results

4.3.1 Performance of the tracking algorithm

The computational performance considering the OAwrB image size of 3100×3100 pixels and current Commercial Off-The-Shelf GPUs with 400 cores was improved by speed-up factors that vary between 50 and 100 with respect to a CPU-based TDBP processing.

For the evaluation of the accuracy of the tracking algorithm, the most relevant factors are the detection and false alarm rate. One restriction on potentially detectable vehicles is imposed by the sensor's resolution: the moving objects must be large to produce reliable backscatter responses, i.e. with the given wavelength and bandwidth, the objects of interest are trucks: the signatures of normal passenger cars were too weak and small.

To demonstrate the generality of the approach, we briefly summarize results achieved at the bridge of the ground truth measurements for "*Oensingen*". Only the combined observation set $O_{ATI \cup 1chn}$ was tested. All five objects were detected in the SAR data, four were tracked correctly over the entire illumination time, one was picked up for 3.4 seconds, then lost for 0.9 seconds and picked up again for the remaining 15.7 seconds (with a different internal vehicle identification tag). No false alarms occurred during the test period on the video-recorded highway section (see Fig. 4.5).

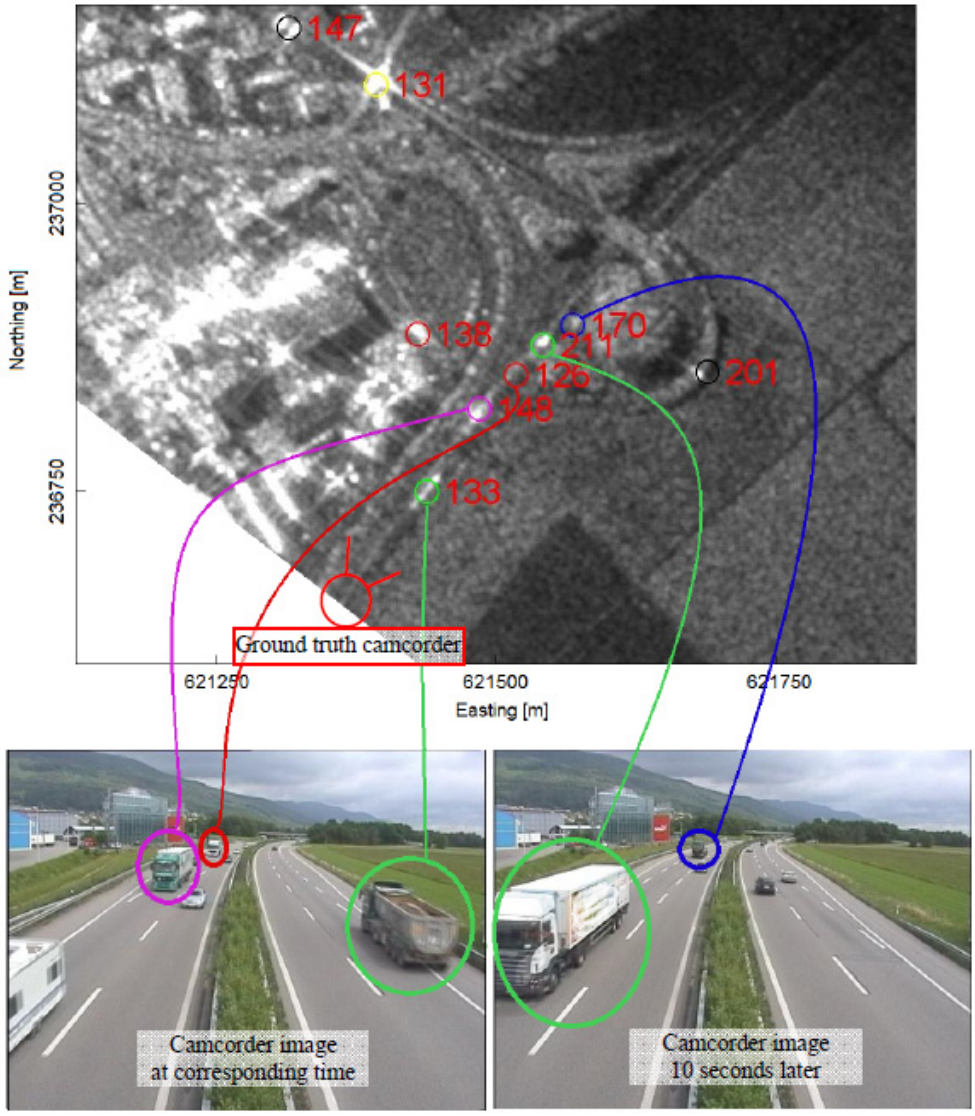


Fig. 4.5 – Ground-truth measurements using a camcorder at the *Oensingen* test site.

A more detailed evaluation was conducted for the test site "*Härkingen*". The SAR measurements were projected on the "ground truth"-observed Zurich-Bern highway using a road model and then propagated from their SAR-observed positions to the ground truth speedometer location using the vehicles' estimated speeds. Tables 4.1 & 4.2 illustrate the detection and false alarm rates for these vehicles with respect to the SAR measurements in four different observation sets including the SAR-estimated and ground-measured speeds and time stamps discussed later.

TABLE 4.1
GROUND TRUTH AND SAR MEASUREMENTS

	Ground	81 km/h	12:29:44	Δv = -1.4 km/h
	SAR	79.6 km/h	12:29:43.65	Δt = -0.35s
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{1chn}, O_{ATI}, $O_{ATI \cap 1chn}$			
	Ground	no catch by speedometer		Δv = n/a
	SAR	86.4 km/h	12:29:47.60	Δt = n/a
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{1chn}, O_{ATI}, $O_{ATI \cap 1chn}$			
	Ground	89 km/h	12:29:49	Δv = -4.4 km/h
	SAR	84.6 km/h	12:29:48.51	Δt = -0.49s
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{1chn}, O_{ATI}, $O_{ATI \cap 1chn}$			
20 seconds missing ground truth: 5 objects detected in SAR data				
	Ground	84 km/h	12:30:12	Δv = 4.6 km/h
	SAR	88.6 km/h	12:30:11.67	Δt = -0.33s
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{1chn}, O_{ATI}, $O_{ATI \cap 1chn}$			
	Ground	80 km/h	12:30:15	Δv = n/a
	SAR	not detected in SAR data		Δt = n/a
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{1chn}, O_{ATI}, $O_{ATI \cap 1chn}$			
	Ground	84 km/h	12:30:30	Δv = 0.1 km/h
	SAR	84.1 km/h	12:30:30.52	Δt = 0.52s
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{1chn}			
	Ground	87 km/h	12:30:53	Δv = -4.5 km/h
	SAR	82.5 km/h	12:30:56.77	Δt = 3.77s
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{ATI}			
 too small target?	Ground	106 km/h	12:30:04	Δv = -3.8 km/h
	SAR	102.2 km/h	12:31:02.50	Δt = 1.50s
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{ATI}			
NO PHOTO (FALSE ALARM?)	SAR	74.8 km/h	12:31:19.78	n/a
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{1chn}, O_{ATI}, $O_{ATI \cap 1chn}$			
	Ground	79 km/h	12:31:24	Δv = 9.2 km/h
	SAR	88.2 km/h	12:31:23.33	Δt = -0.67s
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{1chn}, O_{ATI}, $O_{ATI \cap 1chn}$			
	Ground	85 km/h	12:31:49	Δv = 5.7 km/h
	SAR	90.7 km/h	12:31:48.70	Δt = -0.30s
	DETECTED IN: $O_{ATI \cup 1chn}$, O_{1chn}			

TABLE 4.2

PERFORMANCE OF THE TRACKING ALGORITHM

		OBSERVATION SET			
		$O_{ATI \cup 1chn}$	O_{1chn}	O_{ATI}	$O_{ATI \cap 1chn}$
TRUCKS IN GROUND TRUTH	Detected	8	7	6	5
	Not detected	1	2	3	4
	False alarm	2	1	2	1
Detection rate		89%	78%	67%	56%
False alarm rate		22%	11%	22%	11%

Even though the number of ground truth measurements is rather small, some important information can be extracted. Especially when considering the video-recorded ground truth of "*Oensingen*", a union of the observation sets of ATI and single-channel data was found to lead to best results. The false alarm rate remained low, while the detection rate improved from the intersection to the union set of observations. Use of both single-channel and ATI observations combined outperformed the accuracies achieved using each observation set individually. This assumption was confirmed when the tracking duration of the objects for the united vs. the individual observation sets were compared: for the individual observation sets, the average tracking durations were 4.7s and 6.4s for ATI and single-channel respectively. For the united observation set, the average tracking time was raised to 9.2s, implying more robust estimates of the moving objects. A closer analysis of the tracking sequences in ATI and single-channel data revealed that ATI is more powerful in the center of the antenna beam with rising noise level towards the edges (see also Fig. 4.4). In contrast, the single-channel observation set showed its strength at the edges of the antenna beam, where the static background is muted, and relative weakness in the center. This explains the longer tracking durations of the union set: objects traveling through the antenna beam were supported mainly by single-channel extractions towards the edges of the antenna beam and taken over mainly by ATI detections in the center of the antenna beam. For the 15 detections in $O_{ATI \cup 1chn}$, the tracks of two objects were lost for 0.8s and 1.2s, respectively and then assigned a different identification number. These short breaks in the track occurred at the transition, where the ATI extractions vanished and the single-channel extractions were not yet strong enough. However, due to close temporal and spatial proximity, the tracks can be easily linked later in a post-processing step. In the future, more elegant solutions to these potential breaks in the tracks on an ATI-/single-channel observation transition could be realized by intensified parameter tuning with more training data examples (detection probability, object lifetime,…) or by additional features (e.g. object size and intensity). Causes for the two false alarms and one missed object in the united observation set can be manifold. We comment on this topic later in section IV.A (discussion).

Nevertheless, including all available ground truth measurements, the algorithm achieved a detection rate of 93% and false alarm rate of only 14% with the union set of observations $O_{ATI \cup 1chn}$. Given this encouraging performance at detecting trucks, the evaluation of their speed and position estimates is evaluated in the next section.

4.3.2 Evaluation of speed and position accuracy

First, the evaluation of the position and velocity of the SAR measurements incorporating a road model is addressed. When comparing with the ground truth measurements, it is important to keep in mind that in general the estimation of speed and velocity takes place at a different location in the SAR data than in the ground measurements, i.e. the location of detection is propagated using the estimated speed assuming a locally constant vehicle velocity. The average absolute difference between the ground truth and SAR speeds was then 1.19 m/s (4.27 km/h) $\pm$ 0.82 m/s (2.95 km/h). Concerning the position accuracy calculated from the time difference, the absolute disparity averaged to 21.82 m $\pm$ 30.69 m. One outlier with a temporal disagreement of

3.77s (≙ 91.1m) was conspicuous and could have been caused by a sudden acceleration or breaking event (e.g. passing) between the SAR and ground measurement locations. For the remaining six trucks, the differences between ground and SAR time were always less than one second (~25m). Given that the ground clock had an accuracy of only 1 second and the location was propagated with a constant estimate of the speed the average disagreement of 10.27m ± 3.13m (0.44s ± 0.14s) verifies the preciseness of the SAR derived measurements.

Projecting all detected vehicles on the roads, 128 objects were linked to one of the nine main roads of the road network (see Fig. 4.6). In this way, over three quarters of the detected objects were associated to a road, i.e. projected with a realistic speed interval of 20-120 km/h covering the possible speed range of trucks on non-motorway roads, on- and off-ramps and the motorway. Less than a quarter of the objects could not be associated with one of the main roads, therefore, containing either objects moving on a road not included in the road network or being a potential false-positive. The speed distribution ranged from 73.5 km/h to 85.6 km/h on the motorway, from 51.1 km/h to 64.0 km/h on the on- and off-ramps, and from 34.4 km/h to 52.5 km/h on regular roads in an industrial area and out of town, respectively.

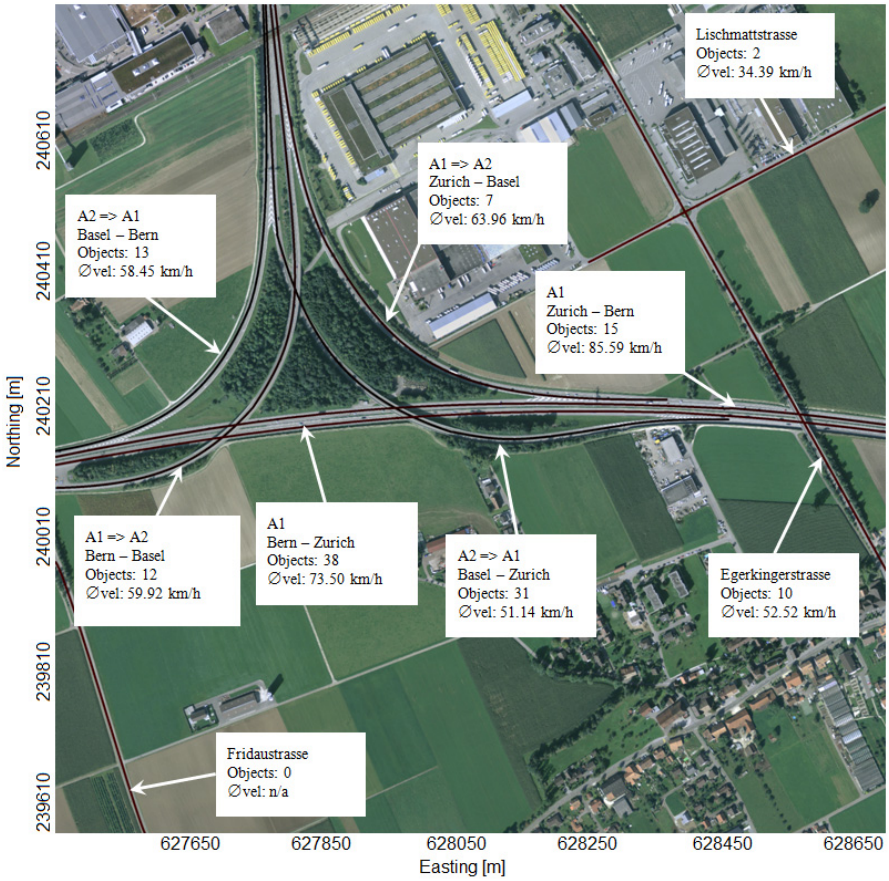


Fig. 4.6 – Moving objects projected onto the road network, indicating for each road the number of vehicles and the average velocity.

For the generic case where no road model was available, the position and speed was corrected relying only on the phase shift or the trajectories in SAR geometry. Fig. 4.7 shows vehicle 1 of table 4.1 as a specific example; overall accuracies using the phase-shift- and trajectory-based approach are illustrated in Fig 4.8(a) and 4.8(b). For 48 of the 128 investigated objects, the phase was stable enough for a robust phase-based estimation, resulting in an average position

difference between road model and phase-based correction of 64.1m, with two thirds of the objects having an error of less than 52m. With the trajectory-based correction, a decreased position accuracy of 85.5m and a two-third confidence interval of 92m were obtained for all 128 objects.

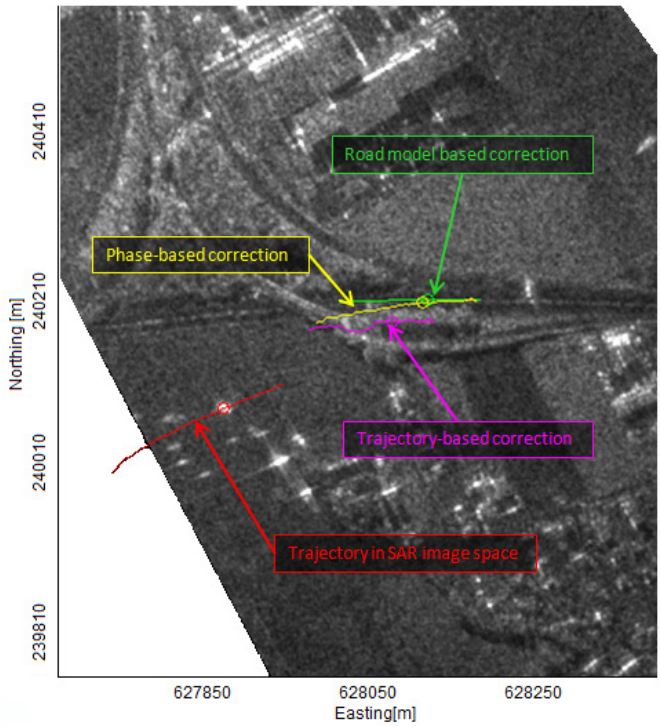
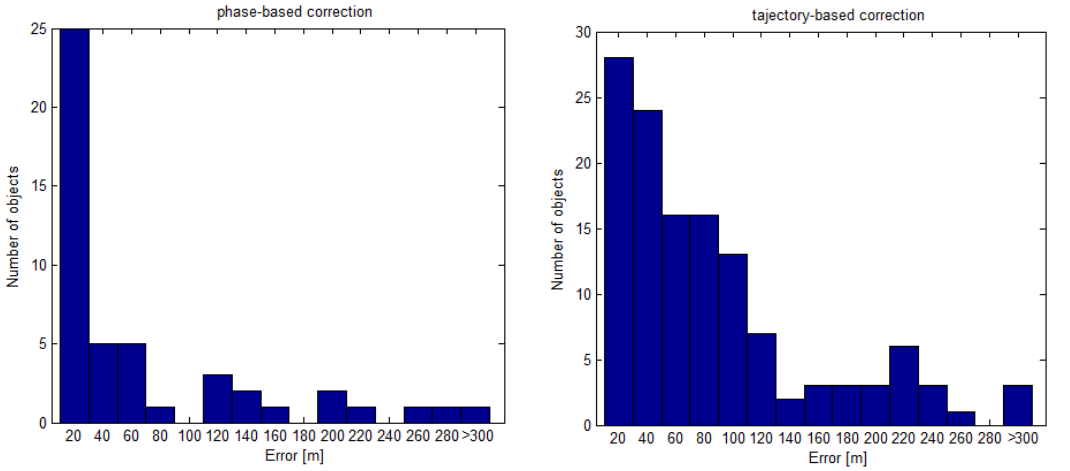


Fig. 4.7 – Example of position corrections (vehicle 1 of table 4.1): (a) using road models (green), (b) using phase information (ATI, yellow), (c) using trajectory information (1chn, magenta). The uncorrected original trajectory in SAR image is illustrated in red.



(a) Error distribution for the position correction using phase information (ATI).

(b) Error distribution for the position correction using trajectory information (1chn).

Fig. 4.8 – Evaluation of different position correction methods to project Doppler-corrupted SAR trajectories into real world 3D positions.

For the objects' velocities, the difference between velocity estimated from the SAR geometry and the road model solution correlated highly with the relative velocity between the sensor and vehicle in the SAR image space. Fig. 4.9 illustrates the scatter plot with the relative velocity

between the moving object in SAR geometry and the sensor as x-axis and the speed error as y-axis. Considering only the 15 vehicles of the ground measured highway (Zurich-Bern), the coefficient of determination of the linear regression was $R^2 = 0.91$, considering all 128 vehicles (including possible false alarms) it remained high at $R^2 = 0.80$, suggesting a reliable speed estimation even without the use of a road network. However, the absolute error had a rather high standard deviation of 5.5 m/s (19.8 km/h). Further investigations aimed at reducing this error are planned for future improvements.

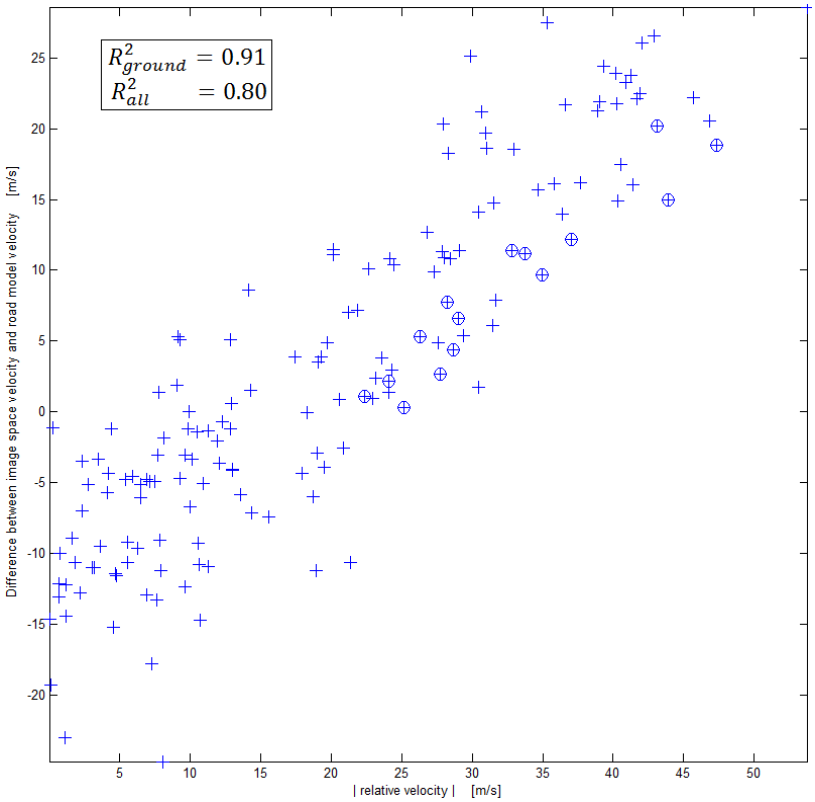


Fig. 4.9 – Scatter graph of the absolute relative velocity between sensor and SAR trajectory plotted against the difference between SAR image space velocity and road-model derived velocity. Crosses: All moving objects. Circles: Moving objects listed in table 4.1.

4.4 Conclusion & Outlook

4.4.1 Discussion

An extended tracking method based on multitarget Kalman filtering working on single-, multi-channel and combined data extraction was introduced and evaluated on highly non-linear flight tracks. Comparing the different extracted observation sets, best results with detection rates of 93% at low false alarm rates of 14% were attained when using combined ATI / single-channel extraction methods. We discuss some possible explanations for the two false alarms and one missed object in the united observation set found in the validation of the "*Härkingen*" test site. The false-positive detections could be caused either by inaccuracies in the processing chain or by uncertainties in the ground truth validation. Potential errors in the processing chain include: (a) artifacts in the focusing step (especially bright scatterers in urban areas with significant side lobes or reduced quality due to strong attitude variations), (b) coherent scattering artifacts along elongated metal structures parallel to the flight direction leading to ghost targets (e.g. caused by power lines, railways, guardrails, roofs of large industrial buildings, etc.),

(c) phase noise (e.g. on the edges of the antenna beam, forested areas, low SNR targets, etc.) or (d) other noise effects with incidentally dynamic behavior (less likely). Note that the circular flight geometry sets higher requirements on the focusing processor, but leads to fewer ghost targets caused by the largely linear character of most metal structures. Uncertainties in the ground truth validation can be caused by: vehicles being detected in the SAR measurements on a highway section not necessarily coincident with the ground truth location (the object is then propagated with a constant speed to the ground measurement station). If detected on a highway section before the motorway exit, the vehicle can leave the highway even though it is assumed to be propagated to the ground measurement location. In fact, that is a realistic assumption for the vehicle in row 9 of table 4.1 (local time: 12:31:19), as it was detected in the SAR data on a section before a motorway off-ramp. Another potential source of false alarms is the fact that the ground truth measurements can miss observations if two consecutive vehicles are separated by only a short time interval. This was the case for vehicles 2 & 3 of table 4.1: two consecutive trucks of which only one was measured, the other only visible on the edge of the optical image. The same can happen for a car and a truck leading to a missed ground truth measurement at the specific time. Furthermore, the relatively high speed of the false-positive vehicle in row 8 of table 4.1 (102.2 km/h) could suggest the detection of a large car (instead of a truck). At the time of the false alarm, a minivan was observed by the speed camera with a ground-measured speed of 106 km/h. Several such similar-sized cars passed during the SAR acquisition and were not detected in general due to the system's resolution. However, one could speculate that even for smaller vehicles, a fortunate illumination geometry may occasionally generate a backscatter response strong enough to be detected by the algorithm. A potential explanation for the undetected vehicle is rather complicated; theoretically, a vehicle that is passing at that time could be projected onto the wrong road using the road network, thus leading to a misdetection. However, that does not seem to be the case here. Even after visual inspection of the data, the only two potential objects strongly suggest being on the motorway exit rather than staying on the highway and passing the ground truth station. Therefore, the other explanation is that for some reason the backscatter response of the truck was not strong enough to generate a significant signal, perhaps caused by an unfavorable illumination geometry, occlusion or shadowing, and strong background clutter in the area of detection.

For the SAR image space to real world transformations, we found that the vehicles' positions and velocities were estimated with an accuracy of 10 m and 1.2 m/s when using road models. Without prior road knowledge, position accuracies of 64 - 86 m and a standard deviation of 5.5 m/s in speed were derived from phase information and SAR image space trajectories.

Compared to traditional SAR-based GMTI algorithms, the proposed method has several advantages: through the dynamic tracking filter, SAR-specific noise effects can be suppressed and moving targets are not only identified but also tracked during the illumination time. Furthermore, the presented method can be practically applied to any SAR system that has a sufficiently wide antenna beam. In contrast to commonly used GMTI algorithms, it provides useful results using a single-channel approach, even though detection rates generally improve when using additional multi-channel information. By using the TDBP on a GPU, non-linear flight geometries can be focused time-efficiently with high quality, providing a high algorithm flexibility concerning acquisition geometry.

4.4.2 Outlook & Summary

Further enhancements are conceivable for the ATI processing with improved antenna calibrations, raising both the detection rate as well as the positional accuracy. In addition, the ATI extractions could be replaced by other frequently used multi-channel methods (e.g. displaced phase center antenna techniques or STAP). The interrelationships between the velocities in the SAR image space and the real 3D space will be investigated in more detail in the future, especially important in areas without existing road network information. Extensions are also planned

for the way of combining information derived from single- and multi-channel data, adding to the “center of mass” - feature other object characteristics such as object size, intensity or phase. The detection rate and the dimension of detectable objects will be investigated with regard to illumination angle, transmit power, wavelength, bandwidth and other factors.

This includes specifically investigations of whether or not limitations on detecting normal passenger cars using the current sensor configuration can be overcome using higher bandwidths. For this purpose, a study with data from a new F-SAR campaign with a bandwidth of 384MHz (instead of 100MHz as in the current experiment) is in progress. Moreover, a further tuning of the model parameters could potentially improve the algorithm’s performance.

The achieved results from the proposed algorithm have been demonstrated to be robust and are highly encouraging for further studies. High detection rates with low false alarm rates under flexible configuration environments (flight geometry, number of channels, independent of the objects’ moving direction, i.e. cross- and along-track, etc.) were demonstrated, clearly indicating the generality of the proposed method for SAR vehicle tracking. The transformation of SAR image positions and velocities to real 3D coordinates varies in accuracy, depending on the inclusion of prior road models or not. The overall performance of the presented method has the potential to support on-ground measurements and monitor large-scale traffic flows in a precise, time-efficient manner at relatively low cost.

5 SYNOPSIS

5.1 Main results

The main achievements of this dissertation are structured according to the research questions presented in chapter 1.5.

5.1.1 Can hidden state space models help to improve the analysis of the temporal behavior of objects/phenomena in remotely sensed data?

In both the dITCZ and traffic monitoring experiments, the results support the interpretation of the temporal behavior being investigated considerably.

In the dITCZ project, the spatial location and expansion of the dITCZ phenomenon could be identified precisely. Extended hidden state space models showed classification performances comparable in reliability to those from human experts for both VS data and IR data. Due to shallow ITCZ events which are only recognizable in VS images, VS data are generally favored. However, the higher temporal resolution and reliability of the IR data legitimates their use in long-term studies especially for the years before 1994 (from 1980-1994, VS data was often corrupted). With the models proposed in this thesis, it is now feasible to automatically detect the dITCZ phenomenon and track its signal on a daily time scale with an accuracy in the range of human expert variability. It clearly outperforms previous studies which either used a larger temporal time scale (Gu et al., 2005) and general image statistics (Chen et al., 2008), both neglecting the ITCZ signal itself and the temporal dependencies, or labeled ITCZ occurrences in a time-consuming, manual identification procedure (Wang and Magnusdottir, 2006). The reliable results of the model already showed on a 5 year test period its potential to support atmospheric scientists in an in-depth temporal analysis of dITCZ phenomena. Initial hypotheses presented in chapter 2 include the asymmetry in spatial structure of the dITCZ depending on relative longitudinal offset across the equator and the accumulation of sITCZ events in La Niña years.

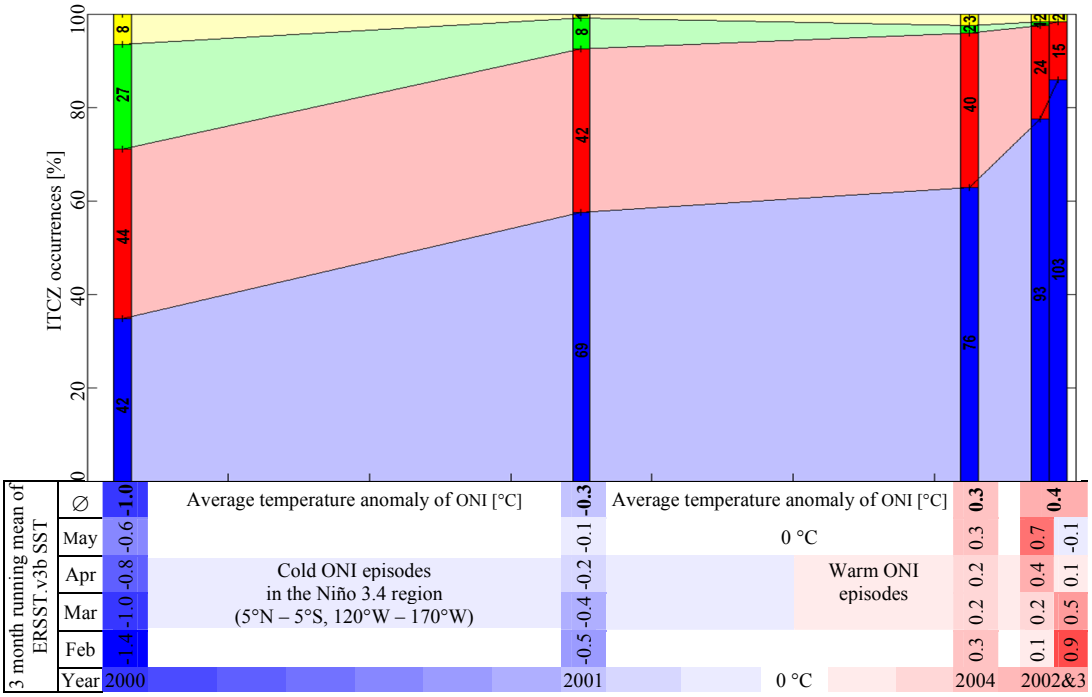


Fig. 5.1 – Dependencies between dITCZ variability derived from the model and the Oceanic Niño Index (ONI) based on 1971-2000 climatology data as provided by the National Oceanic and Atmospheric Administration (NOAA).

Fig 5.1 fortifies this hypothesis of dependencies between ENSO and the four different classes of double ITCZ variability by comparing the model's output and the Oceanic Niño Index (ONI) for the active season of the test period. Warmer average ONI values over the active season correlate with more nITCZ cases occuring in the model output. Colder ONI values over the active season of a year increase the likelihood of more dITCZ and sITCZ cases being registered by the model. A warm ONI characterizes El Niño periods, while a cold ONI indicates La Niña periods.

The model had already been applied to a long-term period spanning the years 1994-2012. The evaluation of the results is ongoing by atmospheric scientists to confirm the benefits of the proposed model.

In the traffic monitoring project, it becomes possible to not only identify vehicles in SAR data as in most traditional SAR/GMTI algorithms but also estimate their trajectories with fine granularity in time and without *a priori* knowledge. Furthermore, the proposed hidden state space model can cope with single-channel only data. This is a significant improvement compared to previous approaches. First methods applied at times when only single-channel SAR systems existed rely on SAR-specific effects which are manually detected by human operators (Raney, 1971; Freeman, 1984). Examples of visually identifiable SAR-specific effects are a defocusing/smearing of the target and a displacement caused by the Doppler shift in focused SAR images. Furthermore, radar shadows can be visually detected at the place of the moving target. Fig. 5.2 illustrates these effects using a SAR image acquired with a modern high-resolution SAR system (COBRA sensor operated by Fraunhofer FHR in Ka-Band developed by Essen et al., 2008).

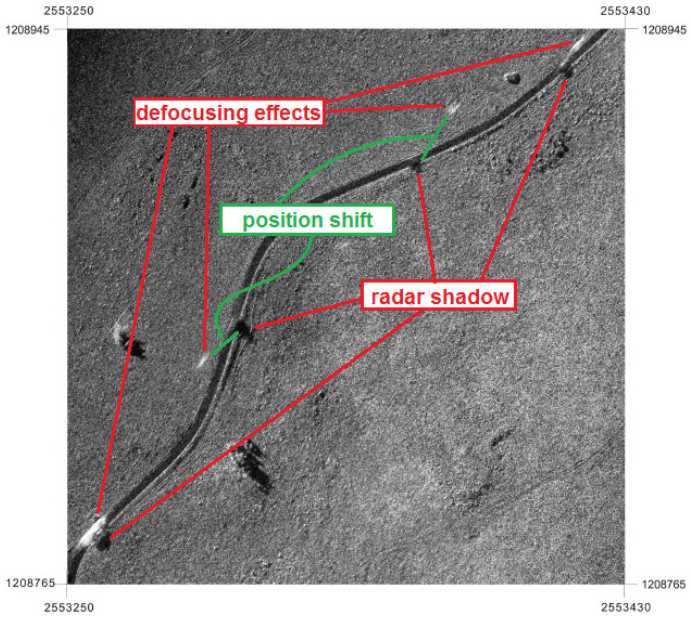


Fig 5.2 – Examples of SAR-specific effects in single-channel data used for visual moving object identification in the past using a 180m x 180m COBRA SAR image of Mt. Racine, Jura, Switzerland as illustrated in Frioud et al., 2012.

More recent developments (Vu et al., 2010; Pettersson, 2007) suggest a brute-force exhaustive search algorithm to detect moving objects by iteratively testing different velocity hypotheses of moving objects. However, the computational burden is insurmountable if a small region in the dimension of the moving object itself cannot be isolated beforehand. Kirscht (2003) suggested a sub-aperture processing schema which is to a certain extent comparable with the OAwrB image processing used for the extraction of the single-channel observation set in chap-

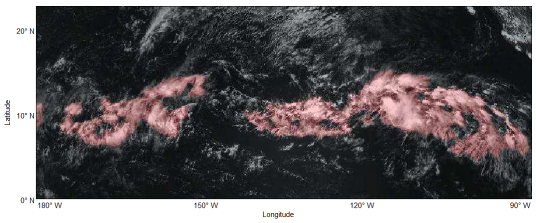
ter 3 and 4 of this dissertation. However, the decisive factor of using the dynamic behavior of the objects was neglected: the method was only applied to semi-real data and simulated moving objects. Consequently, the proposed hidden state space model presented here clearly outperforms previous approaches in detecting and tracking vehicles in single-channel SAR data. For multi-channel data, more robust solutions for vehicle tracking are established (DPCA, ATI, STAP and others). However, many restrictions on flight geometry, system properties and direction of the object's motion exist. The method proposed in this thesis allows high flexibility and achieves high detection rates (93%) while maintaining low false alarm rates (14%) under highly non-linear acquisition geometries (circular flight track), not yet achieved using any other method. The hidden state space model approach can therefore be applied also in low-cost UAV-based single-channel systems as well as multi-channel systems when available. As shown in this thesis, in the presence of multi-channel data, single-channel / ATI combined observation sets contribute to an improved performance compared to single-channel only systems. Furthermore, the experimental results summarized in table 4.1 demonstrate that ground-based speedometer measurements of trucks on motorways can be supported by the proposed method as it shows good positional accuracies and useful velocity accuracies, especially if a road network is available (for traffic applications generally the case). The big advantage of remote sensing applications in the traffic monitoring context is that large regions of several square kilometers can be covered at low cost with low effort while in general ground-based speedometers deployed in situ deliver only very few sample points. Thus, the combination of an extended hidden state space model and remote sensing technology demonstrates the benefits of the proposed approach for traffic monitoring applications.

5.1.2 What prerequisites have to be met to integrate remotely sensed data into the hidden state space model framework?

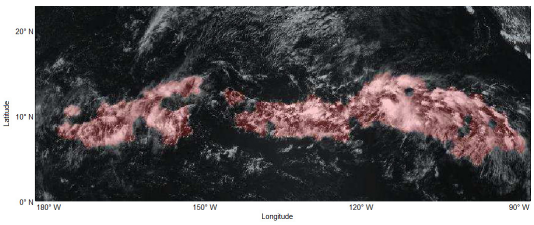
Hidden state space models can be applied to almost every problem in remote sensing where a dynamic process is analyzed that has a direct influence on the data observed using an appropriate remote sensing sensor system. Having a well-defined state set, a robust preprocessing of the data, a thoughtful extraction of the observation sets and a reliable training setup for the case of (semi-)supervised models is fundamental for their successful integration into the hidden state space model framework.

For the climatological study, the definition of the states of the dITCZ phenomena and the observation set were the most challenging tasks for integrating the information extracted from the satellite data into the tracking framework. Even among atmospheric scientists, the clear definition of an ITCZ structure being present or absent is subject to individual fluctuations (see table 2.1). A 90° longitudinal connected cloud band is our working definition of the ITCZ being present but the interpretation of a "connected cloud band" is subjective. This subjectivity naturally leads to variations among the labels assigned by atmospheric scientists and therefore induces some uncertainty in the definition of the states. However, with a sufficiently large number of expert labels, the variation averages out to a certain degree, leading to more reliable training data. For the test run presented in chapter 2, the quantity of available labels was adequate even though we found weaknesses in classifying the rare case of noITCZ states. When extending the model to longer time series, robustness can be enhanced with additional labels. For an extended study period (1994-2012), four additional active seasons were already labeled and used for training the model, resulting in more robust model parameters especially for rare states (noITCZ). In addition, the extracted observation set is a crucial prerequisite for the integration of the remotely sensed data into the dITCZ framework. The most promising observations rely on the backbone path and a subsequent region growing, as well as general image statistics. While the general image statistics are hardly interpretable by humans (despite their importance to the model), the backbone and region growing based features have the beneficial side effect that their values provide useful information for atmospheric scientist when interpreting the dITCZ phenomena. Examples are given in Fig. 2.5-2.7. More such relationships between the

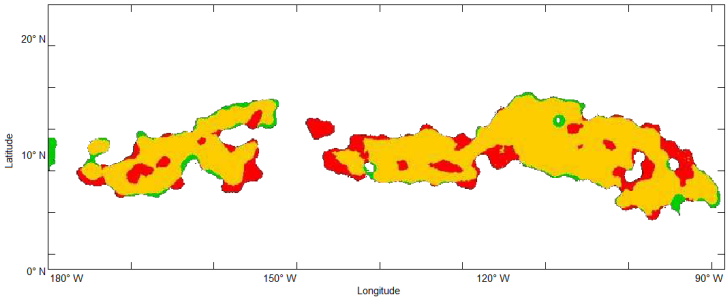
extracted observations and the classified state are being currently investigated for a longer study period. The reason for the meaningful information in the backbone and region-growing based features is rooted in the fact that the region-growing derived mask of the ITCZ cloud structure highly correlates with the ITCZ structure identified by atmospheric scientists. The comparison of the automatic extraction of the ITCZ structure and the expert labeled ITCZ structure was already conducted in a previous study based on a one-month test period (Henke et al., 2007), finding an overlap factor between automatic and human labeled mask in the order of 90% for the test period. An exemplary case is shown in Fig. 5.3 where 92% of the automatically predicted ITCZ structure is part of the expert labeled ITCZ and 85% of the expert labeled ITCZ is covered by the automatically generated mask.



(a) Automatic region-growing mask of ITCZ structures in the northern hemisphere.



(b) Expert labeled mask of ITCZ structures in the northern hemisphere.



(c) Overlaid masks: Yellow are ITCZ pixels marked in both masks, red are only marked by the expert and green only by the automatic approach

Fig. 5.3 – Comparison of automatically extracted and expert labeled ITCZ structures in GOES imagery of August 22, 2000.

An appropriate training setup and a meaningful observation set of the dITCZ (similar to expert labeled features) are thus the key prerequisites for a successful integration in the hidden state space approach. The necessary preprocessing of the raw data (projection, calibration, geocoding and mosaicing) is available as a standard product (Knapp et al., 2011).

For the traffic monitoring, the decisive preprocessing steps for a successful integration into the hidden state space model are the raw data focusing and the observation extraction since the state space is set in well-defined cartographic coordinates, the parameters of the Kalman filter are defined by known physical motion dynamics initialized with typical vehicle motion behavior, and the multi-target parameters (clutter density, object's lifetime,…) are derived from the data using some inexpensive supervised labeling. Important is the raw data processing, as the temporal component inherited from the SAR acquisition and needed for the tracking is not extractable from SAR images focused by a standard SAR processor. Instead many highly overlapping images with reduced bandwidth are demanded to generate an image stack from which the observations can be derived. To process this video-like SAR image sequence, potentially even under non-linear acquisition geometries, a SAR algorithm with high focusing quality, geometric accuracy and computational efficiency is required. Having generated this image stack, different observation sets can be extracted depending on the available channels of the SAR sensor. The

approaches developed in this dissertation use the image statistics of single-channel data and interferometric processing of along-track multi-channel data as well as combined approaches. A further potential way of separating moving targets from the static background planned for an upcoming SAR flight campaign is relying additionally on an exo-clutter sensor configuration using high pulse repetition frequencies. Furthermore, the integration of polarimetric data is contemplated. It is common for all different SAR acquisition modes (single-channel, ATI, exo-clutter, polarimetric) that the combination of the video-like raw data processing and the subsequent methods to extract moving object observations is the key component for a robust and flexible moving object tracking application.

As demonstrated on two exemplary projects in this thesis, a sophisticated integration of remotely sensed data into the hidden state space model is feasible. It requires fundamental knowledge of the imaging system, the specific application and the complete processing chain as a successful realization demands application-driven solutions that need to be developed elaborately for each project.

5.1.3 What extensions to standard models can be applied and how are they suited to improve the performance?

In addition to the prerequisites for the integration, the adaptations and extensions to the standard models have to be carefully optimized for each individual application as the investigated phenomena are subject to different dynamic processes that have to be accommodated by the extended model.

In the dITCZ case, it means that certain assumptions of a standard HMM, i.e. a natural clustering in the observation space and a geometric distribution of the state duration, do not hold for the dITCZ formation process. Thus, to better approximate the non-Gaussian distributions of the emission probabilities, the unsupervised fitting by normal distributions is replaced by supervised HMM hybrid approaches, most successfully using SVMs. Moreover, the typical state duration distributions of the dITCZ phenomena are not in accordance with the geometric distribution of standard HMMs, and are therefore replaced using explicit modeling of the state duration (known as HsMMs). A further extension allows a semi-supervised learning in the model. This is justified by the fact that the labeling of data by atmospheric scientist is very time-consuming. Labeling VS data for a couple of active seasons is doable within reasonable times, however, for IR data with 8 images per day an 8 times higher workload can be avoided if using a semi-supervised learning approach estimating the 7 unlabeled intermediate images.

The tracking of vehicles in SAR data also requires several extensions to a standard Kalman filter. One adaptation is needed to cope with non-linearity existing in dynamic SAR processes. It is realized by the unscented Kalman filter preserving at relatively low computational cost high accuracies. More importantly, in traffic monitoring applications, a large number of moving objects are present at the same time and thus, a collection of Kalman filters is adapted to track multiple objects simultaneously, i.e. each object with an individual filter. The major task of the model is therefore to associate multiple observations at a specific time step with the correct tracking history of a specific filter. Noise effects can cause false observations or missed observations which have to be intercepted by the model to extract one continuous trajectory for each individual vehicle. This is realized by an extended MCMC particle filtering approach allowing the modeling of objects' lifetimes to accommodate for the fact that moving objects enter and leave the area illuminated by the beam of the SAR sensor.

For a successful realization of a specific application it is essential that adaptations and extensions to the hidden state space model are in coherence with the temporal behavior of the underlying physical process.

5.1.4 Can a certain generality of the investigated methods be confirmed on different applications, scales (temporal & spatial) and data sets?

The two at first glance very different applications confirm that hidden state space models can be successfully applied to different remote sensing studies supposing it is possible to combine an appropriate sensor system, a robust data preprocessing, a reliable observation extraction and model extensions reflecting the temporal behavior of the underlying dynamic process in a sophisticated way. Comparing the two case studies, it is obvious that they differ considerably in many factors: the sensor system, the temporal and spatial scales & resolutions, and the research fields. One of the most decisive factors is the temporal component: the different dITCZ states occur in cycles with durations of several days to weeks and it is observed only once per day in the VS channel and at most 8 times per day in the IR channel. In contrast, the tracking of vehicles on motorways is a fast process covering typically a duration of several seconds to minutes with a sampling rate of milliseconds.

A further important difference is the dynamic process itself. While the trajectories of the vehicles contain the SAR-distorted motion as the extracted high-level information, the dITCZ state sequence keeps track of a more abstract atmospheric circulation system status. Nevertheless, the general processing chain and the hidden state space models are analog for both studies outperforming existing methods in their respective, although different, research fields.

With the necessary experience and knowledge of hidden state space models and an understanding of the associated application-specific processing chain, a certain universality of the method on remote sensing applications could be demonstrated in this dissertation, showing the potential of this class of models for remote sensing application in general.

5.2 Conclusion & Outlook

An efficient implementation of hidden state space models was demonstrated for two different remote sensing applications. In both cases, the results are encourage promoting this class of models specifically for remote sensing applications in order to improve existing methods. Especially in complex fields like remote sensing, the gap between the technology and hardware developers on the one side and the final policy makers on the other side is large; the importance of methods for narrowing this gap is obvious. Both experiments illustrate that with an interdisciplinary approach and an elaborated processing chain, the linkage of remote sensing raw data and high-level information retrieval can be achieved. Important steps for a successful high-level extraction from the raw data involve an appropriate choice and/or development of adequate sensor technology, high-quality image (pre-)processing and focusing algorithms, sophisticated feature extraction and selection methods, adaptation and refinement of hidden state space models. Experts and decision makers need to communicate their needs and assist in the model tuning e.g. by providing feedback or labeling training sets. Collaborations within this broad spectrum of experts in different science fields is challenging and can be difficult, but it provides the opportunity to enhance the understanding of environmental and social earth systems using modern remote sensing technology. In one experiment, climatological statistics of the dITCZ phenomena could be derived from instantaneous meteorological satellite images. In the other study, traffic flow statistics could be derived from SAR raw data with high-end sensors. Both studies outperform existing methods in accuracy and time-efficiency and demonstrate the potential of sophisticated feature extraction and adapted hidden state space models for improving remote sensing applications helping to close the gap between hardly interpretable low-level raw data and the high-level information.

Further research is ongoing for the two presented applications, involving enhancements of the sensors, optimizing features and their selection, tuning model parameters and conducting in-depth studies with the retrieved high-level information.

Specifically, for the dITCZ project, the model is being extended to a longer test period (1994-2012) with additional expert labels especially for rare noITCZ cases. The analysis of the model output and the derived features is in progress and will potentially provide additional insights into the performance of the features and the model, and the understanding of the dITCZ phenomenon and its interactions with the global circulation system (e.g. ENSO). Depending on the in-depth analysis, some tuning of the feature selection and the model itself might be necessary. The preliminary analysis of the long-term output is promising except for some inconsistencies in the year 1998, an extraordinarily strong El Niño year. In that season, the northern ITCZ sometimes shifted very far to the south, at extreme times visible on both sides of the equator, i.e. one ITCZ structure spanning both hemispheres. This extreme ITCZ state is not incorporated in the model and therefore leads to a misclassification as dITCZ. To find a solution, further investigations and potentially additional training data for these extreme cases or even the definition of a new state may be needed.

For the vehicle tracking model, the inclusion of further observations (e.g. phase, size and intensity of a moving object) is ongoing. Furthermore, a new F-SAR experiment is planned that will offer the possibility of testing exo-clutter methods based on high pulse repetition frequencies. For this new campaign, an potential improvement in detectability and a reduction in misdetections for the single image observation extraction can be expected, while opening up the possibility of validating the implemented methods on test sites with generally very different traffic flows (from traffic jams to thin traffic density) using different non-linear flight configurations (circular, L-shaped curve). Moreover, in the new flight campaign the sensor's bandwidth will increase, leading to better spatial resolution, potentially allowing the tracking of smaller vehicles. An in-depth analysis of the detection rate as a function of factors like vehicle size, illumination direction and time, incidence angle, sensor resolution and others is pending. Additionally, a later recursive focusing of moving targets given precise velocity estimates is currently in a first test phase. Mid-term goals include the analysis of polarimetric data for tracking purposes, space-time adaptive processing based observation extractions and the extension of the model for satellite based sensors.

The results presented for the two case studies encourage the application of similar methods to other remote sensing applications. For example, the development of models for temporal biodiversity analyses or vegetation cycles similar in structure to the climatologic study, potentially improving the understanding of interactions between changes in climate and vegetation or for the estimation of the state of vegetation parameters from the inversion of radiative transfer models. Further potential applications include the temporal analysis of deforestation events or melting and freezing cycles in alpine and (sub-)polar regions. The experiments described in this thesis provide a reference framework for high-level temporal analysis based on remote sensing raw data using hidden state space models.

REFERENCES

Arbter K., Snyder W.E., Burkhardt H., & Hirzinger G. (1990). Application of affine-invariant Fourier descriptors to recognition of 3-D objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*, 640-647.

Bain C.L., Magnusdottir G., Smyth P., & Stern H. (2010). Diurnal cycle of the intertropical convergence zone in the east Pacific. *Journal of Geophysical Research*, *115*, 1-10.

Bain C.L., De Paz J., Kramer J., Magnusdottir G., Smyth P., Stern H., & Wang C.-C. (2011). Detecting the ITCZ in Instantaneous Satellite Data using Spatiotemporal Statistical Modeling: ITCZ Climatology in the East Pacific. *Journal of Climate, 24*, 216-230.

Barrett E.C., & Curtis L.F. (1999). *Introduction to Environmental Remote Sensing.* New York, NY: Routledge.

Bar-Shalom Y., Daum F., & Huang J. (2009). The probabilistic data association filter. *IEEE Control Systems, 29*, 82-100.

Bar-Shalom Y., Willett P.K., & Tian X. (2011). *Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT: YBS Publishing.

Baumgartner S. V., & Krieger G. (2012). Fast GMTI Algorithm For Traffic Monitoring Based On A Priori Knowledge. *IEEE Transactions on Geoscience and Remote Sensing, 50*(11), 4626-4641.

Biermann F., Abbott K., Andresen S., Bäckstrand K., Bernstein S., Betsill M.M., Bulkeley H., Cashore B., Clapp J., Folke C., Gupta A., Gupta J., Haas P.M., Jordan A., Kanie N., Kluvánková-Oravská T., Lebel L., Liverman D., Meadowcroft J., Mitchell R.B., Newell P., Oberthür S., Olsson L., Pattberg P., Sánchez-Rodríguez R., Schroeder H., Underdal A., Camargo Vieira S., Vogel C., Young O.R., Brock A., & Zondervan, R. (2012). Navigating the Anthropocene: improving earth system governance. *Science*, *335*(6074), 1306-1307.

Bourlard H.A., & Morgan N. (1994). *Connectionist speech recognition: a hybrid approach*. Kluwer Academic Publishers.

Bradley B.A., Jacob R.W., Hermance J.F., & Mustard J.F. (2007). A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. *Remote Sensing of Environment*, *106*(2), 137-145.

Brenner A.R., & Ender J.H.G. (2006). Demonstration of advanced reconnaissance techniques with the airborne SAR/GMTI sensor PAMIR. *IEE Proceedings on Radar, Sonar and Navigation, 153*, 152-162.

Budillon A., Pascazio V., & Schirinzi G. (2008). Estimation of Radial Velocity of Moving Targets by Along-Track Interferometric SAR Systems. *IEEE Geoscience and Remote Sensing Letters, 5*, 349-353.

Budillon A., Evangelista A., & Schirinzi G. (2012). GLRT Detection of Moving Targets via Multibaseline Along-Track Interferometric SAR Systems. *IEEE Geoscience and Remote Sensing Letters, 9*(3), 348-352.

Cerutti-Maori D., Klare J., Brenner A.R., & Ender J.H.G. (2008). Wide-Area Traffic Monitoring With the SAR/GMTI System PAMIR. *IEEE Transactions on Geoscience and Remote Sensing, 46*, 3019-3030.

Cerutti-Maori D., Gierull C.H., & Ender J.H.G. (2010). Experimental Verification of SAR-GMTI Improvement Through Antenna Switching. *IEEE Transactions on Geoscience and Remote Sensing, 48*, 2066-2075.

Cerutti-Maori D., Sikaneta I., & Gierull C. H. (2012). Optimum SAR/GMTI Processing and Its Application to the Radar Satellite RADARSAT-2 for Traffic Monitoring. *IEEE Transactions on Geoscience and Remote Sensing, 50*(10), 3868-3881.

Chambers D.W., Baglivo J.A., Ebel J.E., & Kafka A.L. (2012). Earthquake Forecasting Using Hidden Markov Models. *Pure and Applied Geophysics*, *169*(4), 625-639.

Chapin E., & Chen C. W. (2009). Airborne along-track interferometry for GMTI. *IEEE Aerospace and Electronic Systems Magazine, 24*(5), 13-18.

Chen B., Lin X., & Bacmeister J.T. (2008). Frequency distribution of daily ITCZ patterns over the Western-Central Pacific. *Journal of Climate, 21*, 4207-4222.

Chen S.Y. (2012). Kalman filter for robot vision: a survey. *IEEE Transactions on Industrial Electronics, 59*(11), 4409-4420.

Corbell P.M., Perez J.J.. & Rangaswamy M. (2007). Enhancing GMTI Performance in Non-Stationary Clutter Using 3D STAP. In *IEEE Radar Conference 2007*, 647-652.

Cornell S.E., Prentice I.C., House J.I., & Downy C.J. (2012). *Understanding the earth system: global change science for application*. Cambridge University Press.

Cumming I.G., & Wong F.H. (2005). *Digital Processing of Synthetic Aperture Radar Data, Algorithms and Implementation*. Artech House, Boston.

Curlander J.C., & McDonough R.N. (1991). *Synthetic Aperture Radar - Systems and Signal Processing*. John Wiley & Sons, New York, 440-443.

De Brabanter K., Karsmakers P., Ojeda F., Alzate C., De Brabanter J., Pelckmans K., De Moor B., Vandewalle J., & Suykens J.A.K. (2010). LS-SVMlab Toolbox User's Guide version 1.7. Internal Report 10-146, ESAT-SISTA, K.U.Leuven, Belgium. http://www.esat.kuleuven.be/sista/lssvmlab/ (04.02.2011)

De Jong R., Verbesselt J., Schaepman M.E., & Bruin S. (2012). Trend changes in global greening and browning: contribution of short-term trends to longer-term change. *Global Change Biology*, *18*(2), 642-655.

De Leeuw, J., Georgiadou, Y., Kerle, N., De Gier, A., Inoue, Y., Ferwerda, J., ... & Narantuya, D. (2010). The function of remote sensing in support of environmental policy. *Remote Sensing*, *2*(7), 1731-1750.

De Wit A. D., & Van Diepen C. A. (2007). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, *146*(1), 38-56.

Dietz A.J., Kuenzer C., Gessner U., & Dech S. (2012). Remote sensing of snow–a review of available methods. *International Journal of Remote Sensing*, *33*(13), 4094-4134.

Doucet A., de Freitas N., & Gordon N. (2001). *Sequential Monte Carlo Methods in Practice*. New York, NY: Springer.

Ebel J.E., Chambers D.W., Kafka A.L., & Baglivo J.A. (2007). Non-Poissonian earthquake clustering and the hidden Markov model as bases for earthquake forecasting in California. *Seismological Research Letters*, *78*(1), 57-65.

Edrich M. (2006). Ultra-lightweight synthetic aperture radar based on a 35 GHz FMCW sensor concept and online raw data transmission. *IEE Proceedings on Radar, Sonar and Navigation, 153*, 129-134.

Edrich M., & Weiss G. (2008). Second-generation Ka-Band UAV SAR System. In *European Radar Conference, EuRAD 2008*, 479-482.

Ender J.H.G. (1999). Space-time processing for multichannel synthetic aperture radar. *Electronics & Communication Engineering Journal, 11*(1), 29-38.

Essen H., Hagelen M., Johannes W., Sommer R., Wahlen A., Schlechtweg M., & Tessmann A. (2008). High resolution millimetre wave measurement radars for ground based SAR and ISAR imaging. In *IEEE Radar Conference, 2008*, 1-5.

Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter*. Springer.

Fienup J.R. (2001). Detecting moving targets in SAR imagery by focusing. *IEEE Transactions on Aerospace and Electronic Systems, 37*, 794-809.

Foody G.M., & Curran P.J. (1994). *Environmental remote sensing from regional to global scales*. John Wiley & Sons Ltd.

Frank A., Smyth P., & Ihler A. (2012). A graphical model representation of the track-oriented multiple hypothesis tracker. In *IEEE Statistical Signal Processing Workshop (SSP) 2012,* 768-771.

Freeman, A. (1984). Simple MTI using synthetic aperture radar. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 1984,* 65-70.

French M.N., Andrieu H., & Krajewski W.F. (1994). A model for real-time quantitative rainfall forecasting using remote sensing: 2. Case studies. *Water Resources Research*, *30*(4), 1085-1097.

Frey O., Magnard C., Ruegg M., & Meier E. (2009). Focusing of Airborne Synthetic Aperture Radar Data From Highly Nonlinear Flight Tracks. *IEEE Transactions on Geoscience and Remote Sensing, 47*(6), 1844-1858.

Frioud M., Wahlen A., Essen H., & Meier, E. (2012). Processing of COBRA FMCW SAR Data. In *Progress in Electromagnetics Research Symposium (PIERS) 2012*, 322-327.

Galford G.L., Mustard J.F., Melillo J., Gendrin A., Cerri C.C., & Cerri C.E. (2008). Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sensing of Environment*, *112*(2), 576-587.

Genyuan W., Xiang-Gen X., Chen V.C., & Fielder R.L. (2004). Detection, location, and imaging of fast moving targets using multifrequency antenna array SAR. *IEEE Transactions on Aerospace and Electronic Systems, 40*, 345-355.

Gierull C.H., & Sikaneta I.C. (2003). Raw data based two-aperture SAR ground moving target indication. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2003, 2*, 1032-1034.

Gierull C.H. (2006). Ground moving target parameter estimation for two-channel SAR. *IEE Proceedings on Radar, Sonar and Navigation, 153*, 224-233.

Gierull C.H., Sikaneta I., & Cerutti-Maori D. (2013). Two-Step Detector for RADARSAT-2's Experimental GMTI Mode. *IEEE Transactions on Geoscience and Remote Sensing, 51*(1), 436-454.

Govaerts Y.M., Lattanzio A., Taberner M., & Pinty B. (2008). Generating global surface albedo products from multiple geostationary satellites. *Remote Sensing of Environment, 112*(6), 2804-2816

Greene, A. M., Robertson, A. W., Smyth, P., & Triglia, S. (2011). Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. *Quarterly Journal of the Royal Meteorological Society*, *137*(655), 347-359.

Gruber A., Wessel B., Huber M., & Roth A. (2012). Operational TanDEM-X DEM calibration and first validation results. *ISPRS Journal of Photogrammetry and Remote Sensing, 73,* 39-49.

Gu G., Adler R.F., & Sobel A.H. (2005). The eastern Pacific ITCZ during the boreal spring. *Journal of Atmospheric Sciences, 62*, 1157-1174.

Henke D., Smyth P., & Burkhardt H. (2007). *Pattern recognition algorithms for detecting and classifying ITCZ phenomena over time*. Diploma Thesis. Computer Science Department, Albert-Ludwigs University Freiburg.

Henke D., & Meier E. (2010). Preliminary results of a low-frequency 3D-SAR approach for glacier volume mapping. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2010,* 2027-2030.

Henke D., Smyth P., Haffke C., & Magnusdottir G. (2012a). Automated analysis of the temporal behavior of the double Intertropical Convergence Zone over the east Pacific. *Remote Sensing of Environment, 123*: 418-433.

Henke D., Magnard C., Frioud M., Small D., Meier E., & Schaepman M. E. (2012b). Moving-Target Tracking in Single-Channel Wide-Beam SAR. *IEEE Transactions on Geoscience and Remote Sensing, 50*(11), 4735 - 4747.

Henke D., Mendez Dominguez E., Small D., Schaepman M.E., & Meier E. (2013). Moving target tracking in single- and multichannel SAR. *IEEE Transactions on Geoscience and Remote Sensing*, submitted.

Horn R., Nottensteiner A., Reigber A., Fischer J., & Scheiber R. (2009). F-SAR – DLR's new multifrequency polarimetric airborne SAR. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2009, 2*, 902- 905.

Horning N., Robinson J., Sterling E., Turner W., & Spector S. (2010). *Remote Sensing for Ecology and Conservation*. Oxford University Press.

Houtekamer P.L., Mitchell H.L., Pellerin G., Buehner M., Charron M., Spacek L., & Hansen B. (2005). Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review*, *133*(3), 604-620.

Hu M.K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory, 8*, 179-187.

Hubert L.F., Krueger A.F., & Winston J.S. (1969). The Double Intertropical Convergence Zone-Fact or Fiction?. *Journal of Atmospheric Sciences, 26*, 771-773.

Jin, S., van Dam, T., & Wdowinski, S. (2013). Observing and understanding the Earth system variations from space geodesy. *Journal of Geodynamics*, in press.

Jönsson, P., & Eklundh, L. (2004). TIMESAT—A program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, *30*(8), 833-845.

Joyce K.E., Belliss S.E., Samsonov S.V., McNeill S.J., & Glassey P.J. (2009). A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. *Progress in Physical Geography*, *33*(2), 183-207.

Julier S.J. (2002). The scaled unscented transformation. In *Proceedings of the American Control Conference 2002, 6*, 4555-4559.

Julier S.J., & Uhlmann J.K. (2004). Unscented filtering and nonlinear estimation. In *Proceedings of the IEEE, 92*, 401-422.

Kalman R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, *82*(1), 35-45.

Khairnar D.G., Merchant S.N., & Desai U.B. (2007). Nonlinear Target Identification and Tracking Using UKF. In *Proceedings of IEEE International Conference on Granular Computing*.

Khan Z., Balch T., & Dellaert F. (2005). MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Transactions on Pattern Analysis and Machine Intelligence, , 27*(11), 1805-1819.

Kieu C.Q., Truong N.M., Mai H.T., & Ngo-Duc T. (2012). Sensitivity of the Track and Intensity Forecasts of Typhoon Megi (2010) to Satellite-Derived Atmospheric Motion Vectors with the Ensemble Kalman Filter. *Journal of Atmospheric and Oceanic Technology*, *29*(12), 1794-1810.

Kirscht M. (2003). Detection and imaging of arbitrarily moving targets with single-channel SAR. *IEE Proceedings on Radar, Sonar and Navigation, 150*, 7-11.

Klette R., & Rosenfeld A. (2004). *Digital geometry: geometric methods for digital picture analysis*. Morgan Kaufmann.

Knapp K.R., Ansari S., Bain C.L., Bourassa M.A., Dickinson M.J., Funk C., Helms C.N., Hennon C.C., Holmes C.D., Huffman G.J., Kossin J.P., Lee H.-T., Loew A., &

Magnusdottir G. (2011). Globally Gridded Satellite Observations for Climate Studies. *Bulletin of the American Meteorological Society, 92*, 893-907

Kohlleppel R., & Gierull C. H. (2008). Enhancement of Along-Track Interferometry for Ground Moving Target Indication. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2008, 2*, 229-232.

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, *305*(3), 567-580.

Kropatsch W.G., & Strobl D. (1990). The generation of SAR layover and shadow maps from digital elevation models. *IEEE Transactions on Geoscience and Remote Sensing*, 28(1), 98-107.

Kump L.R., Kasting J.F., & Crane R.G. (2010). *The earth system*. Prentice Hall.

Lanari R., Hensley S., & Rosen P. (1998). Modified SPECAN algorithm for ScanSAR data processing. In *IEEE International Symposium on Geoscience and Remote Sensing (IGARSS) 1998, 2*, 636-638.

Leite P.B.C., Feitosa R.Q., Formaggio A.R., da Costa G.A.O.P., Pakzad K., & Sanches I.D.A. (2011). Hidden Markov Models for crop recognition in remote sensing image sequences. *Pattern Recognition Letters*, *32*(1), 19-26.

Lietzke C.E., Deser C., & Vonder Haar T.H. (2001). Evolutionary structure of the eastern Pacific double ITCZ based on satellite moisture profile retrievals. *Journal of Climate, 14*, 743-751.

Lipps R., Chen V., & Bottoms M. (2004). Advanced SAR GMTI techniques. In *Proceedings of the IEEE Radar Conference, 2004,* 105-110.

Liu W.T., & Xie X. (2002). Double intertropical convergence zones a new look using scatterometer. *Geophysical Research Letters, 29*, 2072.

Magnard C., Frioud M., & Meier E. (2010). Processing, geocoding and mosaicking of MiSAR data. In *Proceedings of ISPRS 3-Ländertagung DGPF - OVG – SGPBF*, 524-531.

Magnusdottir G., & Wang C.-C. (2008). Intertropical convergence zones during the active season in daily data. *Journal of Atmospheric Sciences, 65*, 2425-2436.

Mallet C., & Bretar F. (2009). Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(1), 1-16.

Masunaga H., & L'Ecuyer T.S. (2010). The southeast Pacific warm band and double ITCZ. *Journal of Climate, 23*, 1189-1208.

Meier E., Frei U., & Nüesch D. (1993). Precise Terrain Corrected Geocoded Images In Schreier G. (ed.). *SAR Geocoding: Data and System*. Wichmann, 173-186.

Metternicht G., Hurni L., & Gogu R. (2005). Remote sensing of landslides: An analysis of the potential contribution to geo-spatial systems for hazard assessment in mountainous environments. *Remote Sensing of Environment*, *98*(2), 284-303.

Minnis P. (1989). Viewing zenith angle dependence of cloudiness determined from coincident GOES East and GOES West data. *Journal of Geophysical Research, 94*, 2303-2320.

Mohamed A.H., & Schwarz K.P. (1999). Adaptive Kalman filtering for INS/GPS. *Journal of Geodesy*, *73*(4), 193-203.

Mulder V. L., De Bruin S., Schaepman M. E., & Mayr T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, *162*(1), 1-19.

Murphy K.P. (2002). Hidden semi-Markov models (HSMMs). http://www.ai.mit.edu/murphyk (04.02.2011)

Odermatt D., Gitelson A., Brando V.E., & Schaepman M.E. (2012). Review of constituent retrieval in optically deep and complex waters from satellite imagery. *Remote Sensing of Environment*, *118*, 116-126.

Oriot H., Cantalloube H., & Vaizan B. (2008). STAP or Along-Track Interferometry: two different perspectives. In *7th European Conference on Synthetic Aperture Radar (EUSAR) 2008*, 1-4.

Pei F., Cui P., & Chen Y. (2008). Adaptive MCMC Particle Filter for Nonlinear and Non-Gaussian State Estimation. In *Proceedings of the 2008 3rd International Conference on Innovative Computing Information and Control*.

Quincey D.J., & Luckman A. (2009). Progress in satellite remote sensing of ice sheets. *Progress in Physical Geography*, *33*(4), 547-567.

Pettersson M.I. (2007). Optimum relative speed discretisation for detection of moving objects in wide band SAR. *IET Radar, Sonar and Navigation, 1*, 213-220.

Platt J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.). *Advances in Large Margin Classifiers*, MIT Press.

Rabiner L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*, 257-286.

Rabus B., Eineder M., Roth A., & Bamler R. (2003). The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*(4), 241-262.

Raney R.K. (1971). Synthetic Aperture Imaging Radar and Moving Targets. *IEEE Transactions on Aerospace and Electronic Systems, 7*(3), 499-505.

Reichle R.H. (2008). Data assimilation methods in the Earth sciences. *Advances in Water Resources*, *31*(11), 1411-1418.

Reigber A., Jager M., Fischer J., Horn R., Scheiber R., Prats P., & Nottensteiner A. (2011). System status and calibration of the F-SAR airborne SAR instrument. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2011,* 1520-1523.

Reigber A., Scheiber R., Jager M., Prats-Iraola P., Hajnsek I., Jagdhuber T., Papathanassiou K.P., Nannini M., Aguilera E., Baumgartner S., Horn R., Nottensteiner A.,& Moreira A.

(2013). Very-High-Resolution Airborne Synthetic Aperture Radar Imaging: Signal Processing and Applications. *Proceedings of the IEEE, 101*(3), 759-783.

Robinson K., & Whelan P.F. (2004). Efficient morphological reconstruction: a downhill filter. *Pattern Recognition Letters, 25*, 1759-1767.

Rogan J., & Chen D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning, 61*(4), 301-325.

Rosen, J., Solazzo, M., Hannaford, B., & Sinanan, M. (2002). Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. *Computer Aided Surgery*, *7*(1), 49-61.

Rüegg M., Meier E., & Nüesch D. (2007). Capabilities of Dual-Frequency Millimeter Wave SAR with Monopulse Processing for Ground Moving Target Indication. *IEEE Transactions on Geoscience and Remote Sensing, 45*, 539-553.

Sabater J.M., Jarlan L., Calvet J.C., Bouyssel F., & De Rosnay P. (2007). From near-surface to root-zone soil moisture using different assimilation techniques. *Journal of hydrometeorology*, *8*(2), 194-206.

Sakamoto T., Yokozawa M., Toritani H., Shibayama M., Ishitsuka N., & Ohno H. (2005). A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment*, *96*(3), 366-374.

Sanders J., & Kandrot E. (2011). *CUDA by example: an introduction to general-purpose GPU programming*. Upper Saddle River, NJ: Addison-Wesley.

Särkkä S., Vehtari A., & Lampinen J. (2004). Rao-Blackwellized Monte Carlo data association for multiple target tracking. In *Proceedings of the International Conference on Information Fusion*, 583-590.

Särkkä S., Vehtari A., & Lampinen J. (2007). Rao-Blackwellized particle filter for multiple target tracking. *Information Fusion, 8*, 2-15.

Schaepman M.E., Ustin S.L., Plaza A.J., Painter T.H., Verrelst J., & Liang S. (2009). Earth system science related imaging spectroscopy—An assessment. *Remote Sensing of Environment*, *113*, 123-137.

Scharenbroich L., Magnusdottir G., Smyth P., Stern H., & Wang C.-C. (2009). A Bayesian framework for storm tracking using a hidden-state representation. *Monthly Weather Review, 138*, 2132–2148.

Schindler, K. (2012). An Overview and Comparison of Smooth Labeling Methods for Land-Cover Classification. *IEEE Transactions on Geosciences and Remote Sensing, 11*(50), 4534-4545.

Schubert A., Faes A., Kääb A., & Meier E. (2013). Glacier surface velocity estimation using repeat TerraSAR-X images: Wavelet- vs. correlation-based image matching. *ISPRS Journal of Photogrammetry and Remote Sensing*, *82*, 49-62.

Silva T.A., & Bigg G.R. (2005). Computer-based identification and tracking of Antarctic icebergs in SAR images. *Remote Sensing of Environment*, *94*(3), 287-297.

Smyth P. (1997). Belief networks, hidden Markov models, and Markov random fields: a unifying view. *Pattern recognition letters*, *18*(11), 1261-1268.

Soumekh M. (1999). *Synthetic Aperture Radar Signal Processing: With MATLAB Algorithms*. Hoboken, NJ: Wiley.

Stephens G.L., & Kummerow C.D. (2007). The remote sensing of clouds and precipitation from space: A review. *Journal of the Atmospheric Sciences*, *64*(11), 3742-3765.

Suchandt S., Runge H., Breit H., Steinbrecher U., Kotenkov A., & Balss U. (2010). Automatic Extraction of Traffic Flows Using TerraSAR-X Along-Track Interferometry. *IEEE Transactions on Geoscience and Remote Sensing, 48*, 807-819.

Teh Y.W., Jordan M.I., Beal M.J., & Blei D.M. (2006). Hierarchical dirichlet processes. *Journal of the American Statistical Association*, *101*(476).

Thenkabail P.S., Lyon J.G., & Huete A. (2012). *Hyperspectral remote sensing of vegetation*. CRC Press.

Tooke T.R., & Coops N.C. (2013). A review of remote sensing for urban energy system management and planning. In *IEEE Joint Urban Remote Sensing Event (JURSE) 2013,* 167-170.

Torn R.D., & Hakim G.J. (2008). Performance characteristics of a pseudo-operational ensemble Kalman filter. *Monthly Weather Review*, *136*(10), 3947-3963.

Tottrup C., & Rasmussen M.S. (2004). Mapping long-term changes in savannah crop productivity in Senegal through trend analysis of time series of remote sensing data. *Agriculture, Ecosystems & Environment*, *103*(3), 545-560.

Tralli D.M., Blom R.G., Zlotnicki V., Donnellan A., & Evans D.L. (2005). Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. *ISPRS Journal of Photogrammetry and Remote Sensing*, *59*(4), 185-198.

Ulander L.M.H., Hellsten H., & Stenstrom G. (2003). Synthetic-aperture radar processing using fast factorized back-projection. *IEEE Transactions on Aerospace and Electronic Systems, 39*(3), 760-776.

Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K. I., Aonashi, K., Inoue T., Takahashi N., Iguchi T., Kachi M., Oki R., Morimoto T., & Kawasaki, Z. I. (2009). A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data. *Journal of the Meteorological Society of Japan, 87A*, 137-151.

Van Zyl J.J., Chapman B.D., Dubois P., & Shi J. (1993). The effect of topography on SAR calibration. *IEEE Transactions on Geoscience and Remote Sensing, 31*(5), 1036-1043.

Viovy N. (2000). Automatic classification of time series (ACTS): a new clustering method for remote sensing time series. *International Journal of Remote Sensing*, *21*(6-7), 1537-1560.

Vu V.T., Sjogren T.K., Pettersson M.I., Gustavsson A., & Ulander L.M. (2010). Detection of moving targets by focusing in UWB SAR—Theory and experimental results. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(10), 3799-3815.

Wan E.A., & Van Der Merwe R. (2000). The unscented Kalman filter for nonlinear estimation. In *Proceedings of IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium, AS-SPCC 2000*, 153-158.

Wang C.-C., & Magnusdottir G. (2006). The ITCZ in the central and eastern Pacific on synoptic time scales. *Monthly Weather Review, 134*, 1405-1421.

White M.A., Beurs D., Kirsten M., Didan K., Inouye D.W., Richardson A.D., Jensen O.P., O'Keefe J., Zhang G., Nemani R.R., Van Leeuwen W.J.D., Brown J.F., De Wit A., Schaepamn M.E., Lin X., Dettinger M., Bailey A.S., Kimball J., Schwartz M.D., Baldocchi D.D., Lee J.T., & Lauenroth, W.K. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Global Change Biology*, *15*(10), 2335-2359.

Yan H., Wang R., Li F., Deng Y., & Liu Y. (2013). Ground Moving Target Extraction in a Multichannel Wide-Area Surveillance SAR/GMTI System via the Relaxed PCP. *IEEE Geoscience and Remote Sensing Letters, 10*(3), 617-621.

Yu, S. Z. (2010). Hidden semi-Markov models. *Artificial Intelligence*, *174*(2), 215-243.

Zhang C. (2001). Double ITCZs. *Journal of Geophysical Research, 106*, 11785-11792.

Zhang Z., Xie W., Hu W., & Yu W. (2009). Local Degrees of Freedom of Airborne Array Radar Clutter for STAP. *IEEE Geoscience and Remote Sensing Letters, 6*, 97-101.

Zhang, Z., & Zhang, J. (2010). A new real-time eye tracking based on nonlinear unscented Kalman filter for monitoring driver fatigue. *Journal of Control Theory and Applications*, *8*(2), 181-188.

Zhong, S. (2005). Semi-supervised sequence classification with HMMs. *International Journal of Pattern Recognition and Artificial Intelligence, 19*, 165-182.

CURRICULUM VITAE

Education

2009－2013	**Ph.D.** University of Zurich (Switzerland), Department of Geography, Remote Sensing Laboratories (RSL). Thesis: Object tracking in remotely sensed data.
2001－2007	**Diploma studies** in Computer Science, University of Freiburg (Germany).
2006	**Diploma thesis** at the University of California, Irvine (USA) *Title*: Pattern recognition algorithms for detecting and classifying ITCZ phenomena over time
2004	**Semester abroad** at the University of New South Wales, Sydney (Australia)
1991－2000	**Abitur,** *Staffelsee Gymnasium*, Murnau (Germany)

Professional experience

2008－2013	**Research assistant,** University of Zurich, Department of Geography, Remote Sensing Laboratories (RSL), Zurich, Switzerland.
2006 & 2011	**Visiting Scientist**, Computer Science Department, Donald Bren School of Information and Computer Science & Earth System Science Department, School of Physics, University of California, Irvine (USA).
2000－2001	**Internship & civilian service,** daycare facility for disabled children (*Kinder, Jugend und Erwachsenen e.V.*) , Garmisch-Partenkirchen/Farchant (Germany)

Graduate courses and professional training

- Graduate Seminar I & II
- Graduate Retreat I & II
- Scientific writing
- Project management
- International Workshop on Compressed Sensing applied to Radar (CoSeRa), 2012, Bonn, Germany
- Lecture Series on "Radar and SAR Systems for airborne and space-based surveillance and reconnaissance", 2013, Roma, Italy

Oral contributions:

- IEEE International Geoscience and Remote Sensing Symposium 2009, Cape Town, South Africa.
- IEEE International Geoscience and Remote Sensing Symposium 2010, Honolulu, HI, USA.
- IEEE International Geoscience and Remote Sensing Symposium 2012, Munich, Germany
- Swiss GeoScience meeting, 2011, Zurich, Switzerland

Teaching:

- GEO 443, Lectures SAR I & II (2010 – 2013)
- GEO 803, Lecture in MatLAB Course (2009 & 2013)

Publications

Peer-reviewed publications

Henke D., Mendez Dominguez E., Small D., Schaepman M.E., & Meier E. (2013). Moving target tracking in single- and multichannel SAR. *IEEE Transactions on Geoscience and Remote Sensing*, submitted.

Henke D., Magnard C., Frioud M., Small D., Meier E., & Schaepman M. E. (2012). Moving-Target Tracking in Single-Channel Wide-Beam SAR. *IEEE Transactions on Geoscience and Remote Sensing, 50*(11), 4735 - 4747.

Henke D., Smyth P., Haffke C., & Magnusdottir G. (2012). Automated analysis of the temporal behavior of the double Intertropical Convergence Zone over the east Pacific. *Remote Sensing of Environment, 123*: 418-433.

Other scientific publications

Conference contributions:

Frey O., Werner C., Wegmüller U., Wiesmann A., **Henke D.,** & Magnard C. (2013). A car-borne SAR and InSAR experiment. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2013,* in press.

Henke D., & Meier E. (2012). Moving target tracking in single-channel SAR. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2012,* 6813-6816.

Henke D., & Meier E. (2011). Initial results of a low-frequency 3D-SAR approach for mapping glacier volumes. In *Swiss Geoscience Meeting, 2011*, 382-383.

Henke D., & Meier E. (2010). Preliminary results of a low-frequency 3D-SAR approach for glacier volume mapping. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2010,* 2027-2030.

Henke D., Barmettler A., & Meier E. (2009). Bistatic experiment with the UWB-CARABAS sensor—First results and prospects of future applications. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2009, 2*: 234-237.

Reports:

Henke D., Meier E., & Schaepman M.E. (2013). *Studien auf dem Gebiet von Ground Moving Target Indication GMTI,* Schlussbericht zum Vertragszeitraum 2012-2013, Department of Geography, University of Zurich, p. 71.

Henke D., Meier E., & Schaepman M.E. (2012). *Studien auf dem Gebiet von Ground Moving Target Indication und GMTI,* Schlussbericht zum Vertragszeitraum 2011-2012. Zurich (CH), Department of Geography, University of Zurich, p. 56.

Henke D., Meier E., & Schaepman M.E. (2011). *Studien auf dem Gebiet von tieffrequenten abbildenden Breitbandradar-Systemen,* Schlussbericht zum Vertragszeitraum 2010-2011. Zurich (CH), Department of Geography, University of Zurich, p. 64.

Barmettler A., **Henke D.,** Meier E., & Schaepman M.E. (2010). *Studien auf dem Gebiet von tieffrequenten abbildenden Breitbandradar-Systemen,* Schlussbericht zum Vertragszeitraum 2009-2010. Zurich (CH), Department of Geography, University of Zurich, p. 118.

Barmettler A., **Henke D.,** Meier E., & Itten K.I. (2009). *Studien auf dem Gebiet von tieffrequenten abbildenden Breitbandradar-Systemen,* Schlussbericht zum Vertragszeitraum 2008-2009. Zurich (CH), Department of Geography, University of Zurich, p. 84.

ACKNOWLEDGMENTS

I would like to gratefully acknowledge the input of people that fundamentally contributed to the success of my PhD thesis.
First, and most importantly, I would like to say a very special thank you to my supervisor, Dr. Erich Meier, for giving me the opportunity to conduct my PhD at the SAR Lab of the University of Zurich under his guidance. With his broad knowledge in SAR processing and SAR applications, it was easy to familiarize with SAR related topics. I very much appreciate his efforts, his inspiration, his confidence in me, his motivation and his valuable feedback, the fruitful discussions and ideas as well as the freedom he gave me to realize my research ideas. Without this, my PhD would not have been possible. I am very grateful for an excellent time and for the very pleasant working atmosphere which will always make me remember my PhD thesis as a great time.
I also want to particularly thank Prof. Michael Schaepman for his suggestions and advises which have been very helpful to me. His door was always open to discuss about different research topics revealing opportunities that greatly contributed to the success of my work. His support and his fundamental knowledge in all remote sensing related topics contributed significantly to this work.
A special thank for his important contributions to this work also goes to Dr. David Small. With his broad experience in all SAR related topics he was supporting me to work out the details of an idea. I am grateful for the effort he put into systematically and thoroughly reviewing and improving draft versions of manuscripts to elaborate a completed scientific document.
Furthermore, I would like to thank the other members of the PhD committee, Prof. Konrad Schindler (Institute of Photogrammetry and Remote Sensing, ETH Zurich) and Prof. Renato Pajarola (Visualization and Multimedia Lab, University of Zurich), for their valuable inputs and suggestions.
I would like to thank all current and former members of the SAR Lab. A special thanks to my office colleagues, Christophe, Max and Elias. They not only contributed by providing high-quality SAR products as input to the algorithms developed within this PhD thesis but also by their friendly and helpful character in and outside of University. Adrian and Lukas were helping me out several times; especially Adrian by prove-reading and correcting several manuscripts. I also want to thank the former SAR members Michl and Othi who made my start in Zurich very easy and enjoyable.
Furthermore, I would like to thank all current and former members of RSL. It is a great pleasure to work in such an international and manifold working environment. I want to name specifically Mathias as leader of the SpectroLab, Damien and Bruno for their technical support, Rita and Sandra for their help in all administrative work. Furthermore, I especially appreciate the help when organizing my PhD Apero (amongst other things): many thanks to Maitane(!), Daniela, Irene, Qingyang, Sanne and Marta. Then, Andy with who I spent great times on conferences and field campaigns as well as field work and/or sport activities with Parviz, Hossein, Reik, Jörg, Philip, Feilx, Lucie, Lucia, Valerie, Dani, Zbynek, and others. It has been a great pleasure to work with all of you at RSL and GIUZ.
Another group of people having contributed significantly to this work are from the University of California, Irvine. I am very grateful that I had the chance to work with Prof. Padhraic Smyth (DataLab) and continue some great collaboration in California. His knowledge in probabilistic modeling and his enthusiasm about this topic in the context of remote sensing made me start a PhD in this area. I very much appreciate his very valuable inputs and advises and had a great time at UCI. My thanks goes not only to Padhraic but also to Prof. Gudrun Magnusdottir (Earth System Science Department) who always gave me great support and confidence. I really appreciate the enthusiasm about the ITCZ project and the inspiration and knowledge about atmospheric science. Also, Colene Haffke contributed to the work during my stay at UCI. Thanks for the fruitful discussions and the time-consuming manual labeling (together with

Ashley and Yi-Hui).
Other institutions contributing to this thesis with their financial, technical, logistical and scientific support are: the Procurement and Technology Center "*armasuisse*" of the Swiss Federal Department of Defense (Dr. Peter Wellig), the German Aerospace Center DLR (Dr. Andreas Reigber and colleagues), Airbus Defence and Space (Dr. Klaus Hoffmann) and the Fraunhofer Institute for High Frequency Physics and Radar Techniques FHR (Prof. Nils Pohl and colleagues).
Finally, I want to extend my thanks and great appreciation to my family and my friends in Bavaria, Switzerland and the rest of the world. Without them I would not be where I am today.

ISBN Nr. 978-3-03703-030-1