MASTERARBEIT

Probabilistic Fusion of Ku and C Band Scatterometer Data for Determining the Freeze/Thaw State

ausgeführt am

Institut für Photogrammetrie & Fernerkundung

der Technischen Universität Wien

unter der Anleitung von

Prof. Dr. Wolfgang Wagner

sowie

Dr. Annett Bartsch

Dr. Thomas Melzer

als mitwirkenden Universitätsassistenten

durch

Simon Zwieback

Weißes Kreuzgasse 96
2340 Mödling

Wien, 2011-03-17

Abstract

The transition of the landscape from frozen to non-frozen conditions has far-reaching consequences on numerous geo- and biophysical processes such as plant growth and the hydrologic cycle. Microwave remote sensing has been shown to be an apt tool for monitoring the landscape freeze/thaw state. As the measured signal σ^0 is sensitive to different factors at different radar frequencies, the combination of distinct data sources can potentially lead to improved results.

In light of this observation, a novel sensor fusion algorithm is proposed – it estimates the F/T state based on scatterometer data: SeaWinds on QuikScat in K_u band and ASCAT on MetOp in C band. In addition, a widely used backscatter model for snow packs is extended, whose purpose is twofold: firstly, it can give insight into the dependence of σ^0 on various factors and secondly, it facilitates the parameterization of the aforementioned sensor fusion model.

The sensor fusion approach is based on a probabilistic model, an adaptation of the well-known Hidden Markov model (HMM). The F/T state is assumed to be a Markov chain, whose value is not directly observable. At each epoch, however, its current state influences the measurements: σ^0 at both frequency bands. The simple structure assures that inference can be done efficiently, e.g. the calculation of the probability of the state on a given day. The algorithm does not use training data; the parameters are estimated for each time series in an unsupervised fashion. This is achieved by maximizing the marginal likelihood in the framework of the Expectation Maximization algorithm.

The algorithm is analyzed and tested in a study area in Russia and northern China, which encompasses the region of $120 - 130\,E$ and $50 - 75\,N$. The time series of the probability of the state are validated with in-situ snow and temperature data as well as global climate models. In general, the accuracy exceeds 90%, but the algorithm can fail in agriculturally used land (fields, pastures) and bare rock outcrops in mountainous regions. On a more qualitative level, the study affirms the importance of using two distinct frequencies, as particularly dry snow, vegetation and the freezing of the soil water manifest themselves differently at K_u and C band.

Zusammenfassung

Der Frier/Tau Zustand (Freeze/Thaw state) der Landschaft kann mit Methoden der Mikro-wellenfernerkundung beobachtet und analysiert werden. Da die Bedeutung diverser Ein-flussfaktoren auf das gemessene Signal σ^0 eine Funktion der Radarfrequenz ist, erscheint die Kombination unterschiedlicher Datenquellen eine vielversprechende Möglichkeit zur Verbesserung der Ergebnisse.

Angesichts dieser Erkenntnis wird ein neuartiger Sensor-Fusionsalgorithmus vorgestellt, welcher den Frier/Tau (F/T) Zustand anhand von Scatterometerdaten schätzt: QuikScat im K_u-Band und ASCAT im C-Band. Darüber hinaus wird ein weit verbreitetes Backscatter Modell adaptiert. Dieses vereinfacht einerseits der Parametrisierung des Sensor-Fusions-modells und ermöglicht andrerseits ein besseres Verständnis der Abhängigkeit von σ^0 von verschiedenen Einflussfaktoren.

Der Sensor-Fusionsalgorithmus basiert auf einem wahrscheinlichkeitstheoretischen Mo-dell, einer Adaption des Hidden Markov Models (HMM). Der F/T Zustand, dessen Wert nicht direkt beobachtbar ist, wird durch eine Markov-Kette beschrieben. Allerdings bein-flusst der F/T Zustand die gemessenen Signale: σ^0 in beiden Frequenzen. Die einfache Struktur des Modells garantiert, dass beispielsweise die Wahrscheinlichkeit des Zustandes an einem einzelnen Tag effizient berechnet werden kann. Der Algorithmus benötigt keine Trainingsdaten; die Parameter werden für jede Zeitreihe geschätzt. Dies wird durch Max-imierung der marginalen Likelihood auf Basis des Expectation Maximization Algorithmus erreicht.

Das Verfahren wird anhand eines Testgebiets in Russland und Nordchina ($120 - 130\,E$, $50 - 75\,N$) analysiert und validiert. Die Zeitreihen der Wahrscheinlichkeit des F/T Zus-tandes werden mit In-Situ Schnee- und Temperaturmessungen sowie globalen Klimamod-ellen verglichen. Im Allgemeinen werden Genauigkeiten von mehr als 90% erzielt, jedoch kann der Algorithmus in landwirtschaftlich genutzten Gebieten sowie über blankem Fels im Gebirge versagen. Darüber hinaus bestätigt diese Arbeit die Bedeutung der Kombina-tion zweier verschiedener Frequenzen, da inbesondere trockener Schnee, Vegetation und das Gefrieren des Bodenwassers unterschiedliche Auswirkungen auf σ^0 im K_u- und C-Band haben.

Acknowledgements

Contents

Contents

1. Introduction

1.1. Remote sensing

Remote sensing aims at retrieving relevant information by measurement devices that are not in physical contact with the object of interest. More specifically, the scope of the discipline is usually restricted to those approaches that employ electromagnetic radiation to collect data [84]. The radiative energy can either be naturally occurring (e.g. thermal emission) or generated as part of the measurement process – the former methods are referred to as passive, the latter as active.

Besides platforms mounted on airplanes, helicopters, balloons etc., space-borne instruments are at the heart of the practice of remote sensing. They are particularly suitable for mapping global, large-scale processes which would be impossible to monitor otherwise. This point is particularly important when considering remote regions such as the oceans or the high-latitude landmasses, where extensive in-situ data collection is not feasible.

Many different processes and factors determine the signal detected by the sensor. Consequently, it cannot be expected that unambiguous inferences are possible: the problem of retrieving parameters of the surveyed scene is generally ill-posed [57]. It is thus unavoidable to make simplifying assumptions or rely on simplistic empirical methods altogether.

Apart from the interaction of the electromagnetic waves with the object of interest, another important factor to consider is the way the radiation propagates between the sensor and the target. If one is primarily interested in the monitoring of the Earth's surface, only those regions of the spectrum where a significant amount of radiation is transmitted through the atmosphere may be used. Such parts of the spectrum are called atmospheric windows; the frequency of the analyzed radiation is one of the properties the most common remote sensing methods for studying the Earth's surface can be classified by:

Passive Visible to Short-Wave Infrared methods employ electromagnetic radiation in the visible part as well as the near and short-wave infrared, chiefly reflected sun light.

LiDAR - Light Detection and Ranging is an active method concerned with the analysis of backscattered pulses of electromagnetic radiation – typically in the visible and near infrared.

Passive Thermal Infrared remote sensing deals with thermal infrared radiation. It makes use of the fact that the thermal radiation emitted by any object depends on its temperature and material properties.

Passive Microwave sensors detect thermal radiation in the microwave region.

Active Microwave measurements employ artificially generated microwave radiation to study the target of interest.

Active microwave remote sensing techniques are particularly suitable for regular, global monitoring because they do not depend on the illumination by the sun and because they

are insensitive to the state of the atmosphere (e.g. cloud cover). The way microwave radiation is scattered by a target chiefly depends on two parameters: firstly, the structure and roughness of the surface and secondly, its dielectric properties.

Such data are generally difficult to interpret because the factors influencing the backscattered signal are different from those which determine our visual perception of the very same target [106]. For certain types of instruments, the imaging geometry can also lead to counterintuitive results.

Nevertheless, the usefulness of active microwave data has led to the development of numerous instruments [106]. Scatterometers are radar instruments used to accurately measure the backscattered energy, whereas radar altimeters are designed to determine the range with high precision. Side-looking radars focus on recording a continuous map of the illuminated region. Wave diffraction dictates that in order to provide high spatial resolution, a large antenna is needed. Synthetic aperture radars circumvent this limitation by mimicking a huge antenna by coherent processing of the data. By comparing the phases of the received signals, additional kinds of information such as the topography or surface deformation can be retrieved; this is called radar interferometry [3]. Another property of electromagnetic waves that can be used to infer properties of the target is the polarization. The study of the dependence of the backscatter on polarization is referred to as polarimetry [21].

Scatterometers are particularly suitable for global earth observation purposes because they offer excellent temporal resolution and spatial coverage, albeit at the cost of a rather poor spatial resolution.

1.2. Sensor fusion and data assimilation

The idea of sensor fusion, although ubiquitous in the Earth sciences, is very hard to pinpoint. Different terms, such as fusion, integration, merging and combination, can take on a myriad of meanings depending on the context and background. Wald [103] provides an overview of various definitions and interpretations with the emphasis on remote sensing.

In this thesis, the term 'sensor fusion' is employed to describe the process of combining information from different sources in a systematic fashion, the aim being better results. The phrasing is deliberately vague. The focus, however, is clearly on the organized nature of the fusion, whereas quite often it is simply considered to be a collection of tools and algorithms, e.g. [79].

The benefits generally attributed to this process derive from the varying properties of the distinct data sources. In the present context, primarily the dependency of the backscattering mechanisms on frequency are of interest, as are differences in the temporal coverage.

The combination of the various kinds of data is often deemed to take place at different levels of abstraction [103]. In the context of this thesis, fusion at the data level consists of merging the data before inferring the freeze/thaw state. An alternative would be to derive independent estimates of the freeze/thaw states and combine these.

In this work, we fuse the distinct measurements directly and the framework we employ is that of probability theory. The concept of probability, as summarized by the axioms of Kolmogorov, allows us to describe non-deterministic phenomena such as instrument noise in a clearly specified way. After having made certain kinds of observations, inferences

about the unknown quantities can be made. Alternative frameworks for problems of this kind are, for example, fuzzy methods and the belief theory of Dempster and Shafer [111].

In contrast to the terms listed in the first paragraph, 'data assimilation' usually has a narrower scope [103]: it describes the incorporation of measurements into numerical models. The most well-known example is that of weather forecasting, e.g. by the ECMWF [78].

A popular tool for data assimilation is the Kalman Filter and its variants, which is an efficient algorithm to make certain kinds of inferences in a particular probabilistic state space model. The Hidden Markov Model employed in this work belongs to the same family of models; it is applicable when the state – the freeze/thaw state in our case – is described by a discrete variable.

The approach taken in this thesis is very similar to the data assimilation schemes described above. Due to the lack of input data, the models describing the physical phenomena as well as the observation process are empirical in nature, rather than having a sound physical basis. These different kinds of measurements can then be used, depending on availability, to infer the freeze/thaw state. The framework employed is sufficiently flexible to handle missing data and varying noise levels.

1.3. Freeze/thaw state

The transition of the landscape from frozen to non-frozen conditions has far-reaching consequences on numerous geo- and biophysical processes. The snow cover is another factor of crucial importance with similar implications for those very same processes. These two physical phenomena affect the backscatter which thus enables us to study them with active microwave remote sensing techniques. It should be mentioned that in the remote sensing literature, often no distinction is made between snow cover and the state of the soil water [10] when referring to the freeze/thaw state at continental scale.

One of the processes most affected by the freeze/thaw state is the hydrologic cycle. The liquid water generated when snow melts infiltrates the soil or leads to overland run-off [16]. The former process, however, is severely inhibited in many soils in high-latitude regions because the soil is frozen or close to saturation.

The annual duration of frozen conditions limits the extent of the growing season of plants in high latitudes. Low temperatures inhibit protein synthesis and activity and consequently the metabolism [54]. Frozen conditions also lead to reduced water and nutrient availability. The lack of light in winter, together with the screening effect of snow, has to be considered as well. Plants thus enter a state of dormancy characterized by significantly reduced metabolic activity [95].

The transition to non-frozen conditions also triggers significant changes in the carbon balance [44]: carbon sequestration by the vegetation increases, whereas the accelerated decomposition of organic soil matter releases CO_2. The carbon cycle at high latitudes is still not very well understood and opposing views regarding the feedback mechanisms with climate are commonly held; Davidson and Jannsens [23] provide a balanced review. Another important gas in the carbon cycle is methane CH_4 [20], which is produced by an anaerobic form of respiration called methanogenesis. It has recently received considerable attention due to it being a greenhouse gas with significant global warming potential [53].

1.4. Previous work

As the freeze/thaw states has significant implications to ecological and geochemical processes, numerous studies about its derivation from remotely sensed data have been conducted. At this point, several seminal papers will be briefly introduced; the physical background of many of these approaches will be described later.

The excellent temporal resolution and spatial coverage of scatterometers and radiometers – in conjunction with the sensitivity of microwave signals to snow and water – make them particularly suitable for deriving large-scale products. The complexity of the topography and the land cover as well as the low spatial resolution renders the estimation of parameters by model inversion difficult; instead, change detection approaches are preferred.

Numerous approaches for monitoring the freeze/thaw state of the landscape using active microwave remote sensing have been developed and applied [10][69]:

Seasonal threshold The key idea of these techniques is to contrast the observation at a given time with typical values for the frozen and thawed state. The freeze/thaw state can consequently be estimated by comparing it with a threshold derived from the reference values. The determination of the parameters and the cut-off varies. This approach is particularly suitable for sensors with poor temporal sampling [69]. It has thus been applied to SAR data [85], but also to scatterometers [14][63].

Moving average With these approaches, it is attempted to track changes in the signal and derive salient events. These can then be related to the phenomena of interest. Commonly, the measurement is compared to a moving average of the preceding days. The difference or some related measure is subsequently compared to various thresholds. Studies relying on such an approach include [62][38].

Diurnal difference Kidd et al. [61] proposed to detect thawing by analyzing the diurnal changes of the signal measured by the SeaWinds scatterometer, which offers excellent temporal coverage at high latitudes. A day is flagged by means of a statistical test. The variance of the test statistic incorporates both the noise level of the cell and the number of observations. The significant events, called thaw indicators, are subsequently combined to determine the periods of thawing.

Several change detection approaches applied to scatterometer data have also been used to determine the freeze/thaw state based on radiometer measurements [69]. There have also been numerous studies based on a static approach originally proposed by Zuerndorfer et al. [112], whereby the brightness temperature at one band and the difference between two bands is used to determine the freeze/thaw state of bare soil based on a simple rule-based approach.

In addition, many algorithms for determining the snow cover and snow water equivalent have been developed, e.g. [19][81]. They are usually based on the inversion of radiative transfer models or semi-empirical simplifications thereof. The most serious limitation is due to the unknown structure of the snow, particularly the size of the grains. In order to take these effects into account, assimilation of in-situ data [80] and temporal modelling of the snow metamorphosis [59] have been suggested.

In the visible and near-infrared domain, the freezing of the soil water barely influences the reflection properties. Snow cover, on the other hand, can be retrieved rather easily; this is, for example, routinely done for MODIS data [45].

1.5. Scope and structure

The purpose of this thesis is essentially twofold. The first – more specific – point is the development and validation of a probabilistic sensor fusion model for estimating the freeze/thaw state based on different sensors: SeaWinds QuikScat Scatterometer operating at K_u-band and MetOp ASCAT Scatterometer operating at C-band. The second – more general – objective concerns the study of the general characteristics of the backscatter for frozen and non-frozen conditions in these two bands. This analysis is based on the aforementioned data sources and the probabilistic model as well as a simple model capable of simulating the backscatter of several layers of snow as well as a bare soil surface.

The insight gained from this analysis helps explain the benefits associated with the sensor fusion: increased robustness and flexibility and more accurate results, when compared to ad hoc empirical methods. In addition, it might be useful for designing algorithms, e.g. for the proposed Dual Frequency Scatterometer [56] operating at K_u- and C-band (its launch onboard the GCOM-W2 satellite is scheduled for 2016).

Information about the freeze/thaw state can also be put to good use in other remote sensing applications. In the course of the retrieval of soil moisture from active microwave remote sensing data, frozen conditions have to be detected as they might otherwise lead to gross errors [89].

Chapter 2 is concerned with the geographical and geophysical aspects of the freeze/thaw state, covering topics such as the dependence of the freezing point of water on the soil matrix and the amount of solutes, permafrost and snow metamorphosis. In addition, a brief overview of the test area is provided. In chapter 3, the physical principles of remote sensing and the backscattering behaviour of the landscape at high latitudes are discussed. It introduces concepts such as Mie scattering, complex dielectric constants and the equation of radiative transfer. Chapter 4 contains a concise summary of the theory of Hidden Markov Models: their key properties, how inferences can be made efficiently and how parameters can be estimated. Among the ideas presented in that chapter are Belief Networks, the Forward-Backward Algorithm and the Expectation Maximization Algorithm.

Chapter 5 provides an overview of the data sources used in the context of this thesis. The detailed description of the probabilistic model for sensor fusion can be found in chapter 6. The backscatter model used for validation purposes is also presented. The discussion of the data and the model forms the backbone of chapter 7: it contains a general study of the properties of the backscatter time series and an analysis of the most important contributing factors. The results of the sensor fusion are compared to external data in chapter 8, which also includes an investigation of the assumptions and limitations of the method. Chapter 9 wraps up this thesis by providing an outlook into possible extensions and improvements as well as summing up the crucial aspects.

1. Introduction

2. Geography

In this chapter, the two crucial geophysical aspects that determine the freeze/thaw state will be discussed: the freezing and thawing of the soil water as well as snow. The description is focussed on the most important properties and interactions with processes such as the hydrologic cycle.

There is also a second part to this chapter: a brief overview of the test area.

2.1. Freezing and thawing of the landscape

2.1.1. Freezing and thawing of soil

Physical description

The phase transition of water between its liquid and solid form[1] results in abrupt changes of most material properties such as density or heat conductivity. At the phase transition boundary, both phases can co-exist: they are in equilibrium. There is a state function (analogous to a potential) which is thus the same for both phases. At constant temperature T and pressure p, this is the Gibbs free energy G (2.1) and its partial derivative with respect to the number of moles is the chemical potential μ [43]:

$$G(p,T) = U + pV - TS \tag{2.1}$$

$$\mu_i = \left(\frac{\partial G}{\partial N_i} \right)_{T,V,N_{j \neq i}} \tag{2.2}$$

where U is the internal energy, S the entropy, V the volume and N_i the number of particles of species/phase i. The Gibbs free energy tends towards a minimum, which implies equilibrium. At the transition, in particular, the chemical potentials have to be equal: $\mu_{solid} = \mu_{liquid}$.

The Gibbs free energy G considers both energetic and entropic aspects. The solid phase is more structured and in order to move from the solid to the liquid state, heat has to be added to overcome particle interactions. It is called the latent heat of fusion or enthalpy of fusion ΔH_{fusion} and for water it is given by [66]

$$\Delta H_{fusion} = 333.55 \, kJ/kg \tag{2.3}$$

The importance of the enthalpy of fusion for the heat balance is illustrated by the following example. We will assume a temperature-independent specific heat capacity of liquid

---

[1]there are actually many different solid phases, the most recently discovered being ice XV [88]. The only form commonly encountered on Earth is ice l_h.

water (at constant pressure) of $4.22\,^{kJ}/_{kg\,K}$ [109]. The heat energy needed to melt one kilogram of ice is sufficient to raise the temperature of $1\,kg$ of water by almost $80°\,C$. Imagine you had $1\,kg$ of frozen and $1\,kg$ of liquid water at $0°\,C$ in separate containers and suppose you wanted to raise their temperature to $10°\,C$. In this case, you would need almost nine times as much energy for the ice block than for the water.

The freezing point of water depends on the pressure; at standard atmospheric pressure, it is at $0°\,C$. Two very important factors can also change the freezing point [49]:

Solutes Dissolved constituents in the liquid water such as salts lead to an increase of the entropy, thus lowering μ_{liquid} and the freezing point.

Surface Tension Interactions at the interface of the water with e.g. the soil matrix influence the energy. Adsorption to the soil matrix results in a lower μ_{liquid} and this change depends on the pore size. Consequently, as the freezing point depression grows as the pore size decreases, the water in the larger pores is first converted to ice, whereas the tightly bound water only freezes at lower temperatures.

When simulating the freezing of a soil, the energy and mass balance have to be satisfied and the changes in the material properties with the phase transition considered. The Stefan equation is a commonly used simplification [110]: the phase transition is assumed to occur at a fixed temperature, the fluid movements are not considered and the heat capacity of the frozen soil is neglected. It describes the slow propagation of the freezing front.

Consequences

The freezing and thawing of the soil have important implications for its structural stability; a particularly famous example being frost heave. At a constant temperature below the freezing point, there is ice present in the larger pores and liquid water in the smaller ones. Due to the surface tension, there is a pressure difference between the two phases and it depends on the pore size. As the temperature drops, additional pores freeze over and the pressure difference has to increase to maintain equilibrium [49]. Liquid water is thus sucked towards the ice and under the right conditions, ice lenses form.

These processes have a great influence on the metabolic activity of the organisms in the soil and they also regulate the hydraulic conductivity and thus water and nutrient availability, cf. section 1.3.

Permafrost

Permafrost can be defined to be 'perennially frozen ground – rock, sediment or any other earth material with a temperature that remains below $0°\,C$ for two or more years' [86].

The ice contained in permafrost is of great importance for the stability of the ground. The top layer, which thaws in summer and re-freezes in autumn, is called the active layer. Permafrost has recently received considerable attention due to the discussion about global change. It is expected that, in general, the growth in depth of the active layer will continue and that overall permafrost will warm or thaw [86]. Apart from the impacts on the hydrological cycle and the stability of the ground, the interconnection with the carbon cycle is a focus of research. As the permafrost thaws, the increased decomposition of the organic matter stored therein leads to the emission of carbon dioxide and methane, cf. 1.3.

2.1.2. Snow

Formation

Snow is 'frozen precipitation in the form of ice crystals that fall in soft, white flakes' [6].

Snow flakes are crystalline ice particles that form due to the accumulation of H_2O molecules. Pure liquid water can be supercooled to about $-40°C$ without the presence of condensation nuclei [65], which enable nucleation to set in at higher temperatures. The reason why this does not already happen at the freezing point is surface tension: this is referred to as the Gibbs-Thomson effect. The Gibbs free energy G (2.1) thus has to be adapted to account for the surface tension [36]. In a way similar to the influence of the soil matrix on the bound water, the complex interaction between the nucleus and the water molecules affects the surface tension and can thus facilitate nucleation.

As the snowflake grows, the new molecules have to be incorporated into the crystal, additional molecules transported to the site by diffusion and heat removed by heat diffusion. Libbrecht [65] provides an overview of these mechanisms and how they lead to the remarkable variability in shapes and sizes.

Melting and sublimation

When the snow melts, the ice is converted to liquid water. This process sets in at the snow-air boundary once the temperature has risen to the melting point. The liquid water percolates downwards where it re-freezes, thus increasing the temperature of the snow pack due to its ΔH_{fusion} (2.3) [98]. There is thus a wetting front which propagates from the surface. This process is associated with metamorphism (see below).

Sublimation is the direct transition from the solid state to the vapour phase; it is thus somewhat equivalent to evaporation. The process is governed by the difference between the actual and the saturation vapour pressure [74]. It is important for the mass balance of the snow pack as well as the heat balance due to the enthalpy associated with the sublimation.

Metamorphism

A snow pack is not static but evolves dynamically – it metamorphoses. According to the classification of Colbeck [22], two kinds have to be distinguished: dry-snow and wet-snow metamorphism.

Dry-snow metamorphism is, as the name suggests, prevalent in dry snow; it is a rather slow process towards bigger, less intricate snowflakes. Two regimes are commonly considered, depending on the temperature profile of the snow [83]. In case of only a small T gradient, the differences in the radii of curvature are the driving force. Due to surface tension, water molecules tend to sublimate from the convex parts and re-condense at concave interfaces. This very slow process leads to more rounded grains.

When the temperature at the air-snow interface is much less than the one close to the ground, the differences in vapour pressure lead to a flux towards the surface [83]. Along the way, the vapour molecules can attach to grains, thus enlarging them. This process is

much quicker than the previous one and is the driving force in the formation of depth hoar (large crystals that are detrimental to the structural stability of the snow pack).

Wet-snow metamorphism   occurs in snow containing liquid water. When the snow is close to saturation, metamorphism proceeds very quickly, the crucial limit being the heat transport [22]. As the melting point of small grains can be very low, they preferentially melt while the bigger ones might re-freeze. This leads to very large grains. When there is only a small amount of liquid water present, the crystals tend to join by forming interfaces.

A particular aspect of wet-snow metamorphism occurs in the presence of freeze/thaw cycles; it can result in multicrystalline grains and a very stratified appearance (especially when it rains) [83].

Consequences

Snow cover greatly influences the heat balance in the following ways [6]:

- due to its high albedo in the visible and near infrared, the incoming solar radiation cannot efficiently increase the temperature
- the significant enthalpy of fusion leads to a delayed warming in spring; it acts as a heat sink
- it is a formidable insulator and thus keeps the soil temperature from falling precipitously in winter

Barry et al. [6] provide a concise summary of the most important ecological impacts of snow. It greatly influences many aspects of the hydrological cycle such as run-off and thus human activities, e.g. agriculture, power generation and drinking water consumption. It is also of great importance to many animals by limiting access to food, regulating the temperature changes or providing shelter. Snow limits the growing season of plants and greatly affects their growth due to its impacts on the heat and water balance.

2.2. Study area

The test area is between $120°E$ and $130°E$ north of $50°$ latitude. A small part is in Chinese territory, but most of it belongs to Russia, cf. figure 2.1.

2.2.1. Topography

Figure 2.2 provides an overview of the most important topographic features and land cover classes.

The lowlands are dominated by the Lena River, whose source is close to Lake Baikal and which flows in a northeastern direction until about Yakutsk ($62°N$, $129°E$). From there it flows northwest before continuing northwards until it reaches the delta.

East of the Lena River, the Verkhoyansk Range runs in a meridional direction. Its highest peaks in the test area rise above 2000 metres.
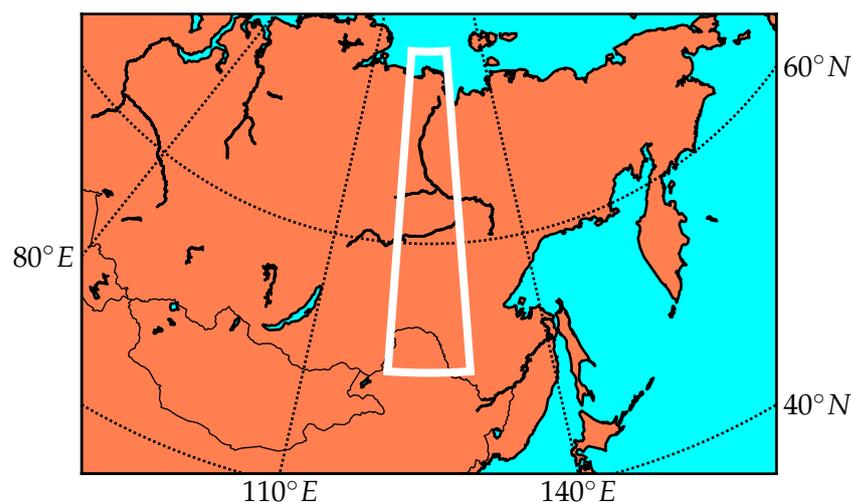
Figure 2.1.: Location of the test area. Albers Equal Area projection.

The Stanovoy Range and nearby highlands stretch in an east-west direction at around $57°N$. Towards the south they give way to the foothills of the Great Khingan. The river Amur forms the boundary between China and Russia; its watershed dominates the very South-East of the study area.

2.2.2. Land cover

Tundra vegetation prevails in the northernmost part of the test area, such as the wetlands in the Lena delta. The grasses and shrubs gradually give way to larches towards the south [94]. Stretches of the Lena River and its tributaries are fringed with wetlands and pastures. The Verkhoyansk range is mostly barren except for the valleys, cf. fig 2.2b.

Towards the south-west of the study area, the taiga gives way to the steppe, with some fields close to the rivers. Pastures and crops also characterize the areas close to the Amur river and its tributaries in the south-east of the test area.

2.2.3. Climate

The climate of the study area is continental; it is in fact quite well-known for the huge differences between winter and summer temperatures. Jakutsk, for example, has an average minimum temperature of $-46°$ in winter and an average maximum temperature of $25°C$ in summer [48].

In winter, the area is at the heart of the Siberian Anticyclone – a stable high-pressure field of very cold and dry air.

The study area is characterized by rather arid conditions. The annual precipitation is around $200 - 300mm$ in the north and $400 - 600mm$ in the south [48]. There are of course topographically induced variations. Most precipitation falls between June and October.
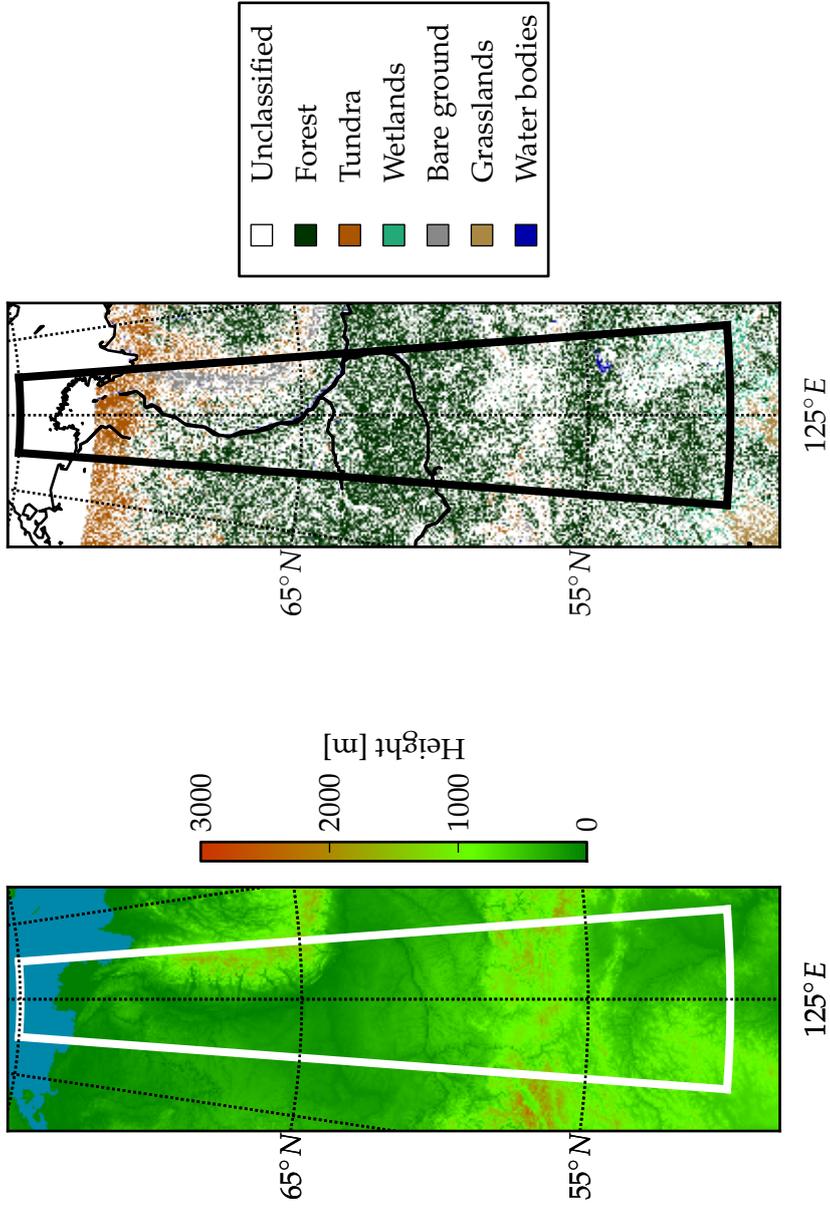
(a) Topographic map based on SRTM30 data [29]

(b) Land cover map based on [7]

Figure 2.2.: Topographic and land cover maps of the study area.

3. Remote Sensing of the Freeze/Thaw State

Many relevant geophysical and biophysical parameters can be estimated on the basis of space- and air-borne measurements. Different sensors and techniques have been shown to be sensitive to different properties of snow, liquid water and ice, which are the features we are primarily interested in.

In the first part of this chapter, the physical principles of electromagnetic waves and their interaction with matter are described. This knowledge is then applied to characterize the backscattering behaviour of the targets of interest in more detail.

3.1. Physical principles

3.1.1. Electromagnetic waves and their interactions with matter

Electromagnetic waves are at the heart of microwave remote sensing and their propagation and interaction with the targets of interest make the inference of quantities of interest possible. The following concise summary generally follows [51].

Maxwell's equations

Maxwell's equations are central to the theory of electromagnetism. In their differential form and separating free and bound charges and currents, they are given by [21]:

$$\nabla \cdot \mathbf{D} = \rho_f \tag{3.1}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{3.2}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{3.3}$$

$$\nabla \times \mathbf{H} = \mathbf{j}_f + \frac{\partial \mathbf{D}}{\partial t} \tag{3.4}$$

where ρ_f is the volumetric density of free charges and $\mathbf{j}_f$ is the current density of free charges.

The $\mathbf{D}$ and $\mathbf{E}$ fields are related by the properties of the material. For a linear, isotropic and homogeneous material, the constitutive relations for the static case are [51]:

$$\mathbf{D} = \varepsilon \mathbf{E} = \varepsilon_r \varepsilon_0 \mathbf{E} \tag{3.5}$$
$$= \varepsilon_0 \mathbf{E} + \mathbf{P}$$

Vacuum has a relative permittivity ε_r of 1 and thus the polarization density $\mathbf{P}$ is 0. The latter expresses the volumetric density of dipole moments. In a dielectric medium, an external $\mathbf{E}$ field leads to a re-arrangement of the internal charges. Even though the net

charge does not change, the formation and re-orientation of electric dipoles results in a non-zero polarization density.

In the case of time-varying fields, the constitutive relations become more complicated. If the temporal evolution of the external fields is sufficiently fast, the polarization cannot follow instantaneously. It is usually assumed that **P** and **E** are related by a causal linear time-invariant (LTI) system. As the complex exponentials are eigenfunctions of any LTI, the constitutive relations reduce to simple proportionality in the frequency domain. The permittivity $\varepsilon(f)$ as a function of frequency is a complex number and commonly decomposed as:

$$\varepsilon(f) = \varepsilon'(f) - i\varepsilon''(f) \tag{3.6}$$

The relation between **H** and **B** is essentially analogous (at least for dia- and paramagnetic materials), the proportionality constant being the relative permeability μ [21]:

$$\mathbf{B} = \mu\mathbf{H} = \mu_r\mu_0\mathbf{H} \tag{3.7}$$
$$= \mu_0(\mathbf{H} + \mathbf{M})$$

For most materials of interest in remote sensing, $\mu_r \approx 1$ to a very good approximation, so that it is usually assumed to be one [106].

Electromagnetic waves

Light and other forms of electromagnetic waves in vacuum can be described mathematically by the wave equation [51]:

$$\nabla^2\mathbf{E} = \mu_0\varepsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2} \tag{3.8}$$

This equation can be derived from Maxwell's equations assuming no currents and no free charges by combining the curl of Faraday's law (3.3) and Ampere's law (3.4). Simultaneously, **H** has to fulfill its own wave equation:

$$\nabla^2\mathbf{H} = \mu_0\varepsilon_0\frac{\partial^2\mathbf{H}}{\partial t^2} \tag{3.9}$$

The plane harmonic wave given by [21]

$$\mathbf{E} = E_0 e^{i(\mathbf{k}^T\mathbf{x}-\omega t)}\mathbf{e}_E \tag{3.10}$$

$$\mathbf{H} = H_0 e^{i(\mathbf{k}^T\mathbf{x}-\omega t)}\mathbf{e}_H = \frac{1}{\eta}E_0 e^{i(\mathbf{k}^T\mathbf{x}-\omega t)}\mathbf{e}_H \tag{3.11}$$

can be shown to satisfy both (3.8) and (3.9). It is the real part that is of physical significance.

The quantity **k** is called the wave vector – its direction determines the direction of propagation and its magnitude satisfies the dispersion relation $|\mathbf{k}| = k = \omega/c$, where $c = \sqrt{1/\varepsilon_0\mu_0}$ is the velocity of light in vacuum. The ratio of E_0 and H_0 is called the impedance of free space and is given by $\eta = \mu_0/\varepsilon_0$. As it is real, the two fields are in phase.

The three vectors $\mathbf{e}_E$, $\mathbf{e}_H$ and **k** are mutually orthogonal. For a fixed direction of propagation, there is still an inherent symmetry, which gives rise to the phenomenon of polarization

[21]. A planar wave as given by (3.8) and (3.9) is called linearly polarized because both field vectors oscillate in a plane. In remote sensing, the polarization is usually referred to with respect to a vertical plane: vertical and horizontal.

In lossless media – ε_r being real – there are analogous relations between the properties of the waves. The most important difference is the change in phase velocity [21]. In media with losses, most of these quantities are complex, leading to exponential decay of the amplitudes and a phase difference between the two fields.

It can be shown that the power transported by an EM wave is proportional to $|\mathbf{E}|^2$ [51]. More specifically, it is given by the Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$, which points in the direction of energy transport.

Fresnel equations

When a plane wave impinges on a planar, infinitely extending interface between two media, one part of the energy is reflected and another transmitted [106]. The geometry is illustrated in figure 3.1. Two different polarization states have to be distinguished:

TM $\mathbf{H}$ perpendicular to the plane of incidence

TE $\mathbf{E}$ perpendicular to the plane of incidence

It can be shown that a stationary solution to Maxwell's equations for the appropriate boundary conditions is given by the Fresnel equations. They characterize the magnitude of the reflected and transmitted $\mathbf{E}$ field relative to the incident one by the reflection coefficient Γ and the transmission coefficient T, respectively. Assuming perfect (non-conducting) dielectrics with $\mu_r = 1$, the Fresnel equations are given by [51]:

Law of Reflection

$$\theta_r = \theta_i \tag{3.12}$$

Snell's Law

$$\frac{\sin \theta_i}{\sin \theta_t} = \sqrt{\frac{\varepsilon_2}{\varepsilon_1}} = \frac{n_2}{n_1} \tag{3.13}$$

TM polarization

$$\Gamma_{TM}(\theta_i) = \frac{\sqrt{\varepsilon_2} \cos \theta_i - \sqrt{\varepsilon_1} \cos \theta_t}{\sqrt{\varepsilon_2} \cos \theta_i + \sqrt{\varepsilon_1} \cos \theta_t} \tag{3.14}$$

$$T_{TM}(\theta_i) = \frac{2\sqrt{\varepsilon_1} \cos \theta_i}{\sqrt{\varepsilon_2} \cos \theta_i + \sqrt{\varepsilon_1} \cos \theta_t} \tag{3.15}$$

TE polarization

$$\Gamma_{TE}(\theta_i) = \frac{\sqrt{\varepsilon_1} \cos \theta_i - \sqrt{\varepsilon_2} \cos \theta_t}{\sqrt{\varepsilon_1} \cos \theta_i + \sqrt{\varepsilon_2} \cos \theta_t} \tag{3.16}$$

$$T_{TE}(\theta_i) = \frac{2\sqrt{\varepsilon_1} \cos \theta_i}{\sqrt{\varepsilon_1} \cos \theta_i + \sqrt{\varepsilon_2} \cos \theta_t} \tag{3.17}$$
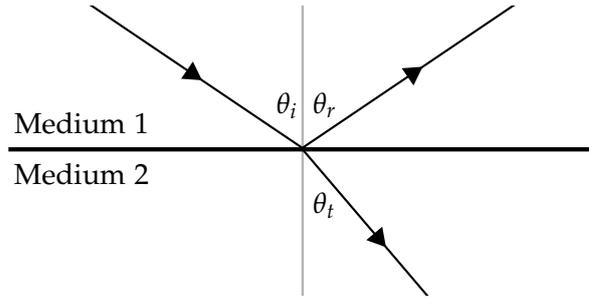
Figure 3.1.: Geometry of the interface and the plane waves

Scattering of EM waves by a rough surface

The reflection implied by the Fresnel equations is perfectly specular. However, when the interface is rough compared to the wavelength, the scattering behaviour becomes more complex. When dealing with natural surfaces, the minute variations in height are usually neither known nor of interest, so that a statistical description seems reasonable. The RMS height s is a measure of the deviation from the mean height. Additionally, it is important to know how rapidly the height changes in the horizontal direction; this can be described by an autocorrelation function, which can often be summarized by a characteristic length l.

In order to quantify the roughness with respect to a specific measurement, it has to be related to the geometry and the frequency. The Rayleigh criterion states that a surface can be considered as smooth if [106]

$$s \leq \frac{\lambda}{8\cos(\theta)} \tag{3.18}$$

In general it holds that the rougher a surface, the more diffuse the backscatter. Thus for the measurement geometries commonly experienced with scatterometers, an increase in surface roughness leads to an increase in the backscatter.

The considerations above only apply for surfaces whose height varies in a random fashion; periodic undulations are excluded. They can lead to constructive interference of the scattered components depending on the direction. If the variations in height are small compared to the wavelength, the scattering regime is referred to as Bragg scattering and described by the Small Perturbation model [84].

Another widely used approach is the Kirchhoff approximation: the key assumption is that the radius of curvature of the surface is small compared to the wavelength. Thus it is usually more appropriate at lower frequencies [84].

As these two models cater to different regimes regarding the patterns of the surface, many more methods that are applicable in different circumstances have been proposed, among them the widely used Integral Equation model by Fung [39]. The majority of these physically based models try to capture the structure of the surface by making use of complex parameterizations. In practice, these descriptions are often too intricate to be measured in situ or to be estimated by inverting such models [102].

Due to the small range of validity and the issues concerning the characterization of the roughness, semi-empirical models enjoy considerable popularity. These are based on extensive experiments and are usually parameterized by the permittivity of the soil and some measure of the roughness. Among them are the models by Oh [77] and Dubois [26].

Scattering of EM waves by a particle

Electromagnetic waves interact with particles; the redirection of energy is called scattering. The proportional amount of scattering, for fixed frequency and particle shape and size, depends on the direction of both incoming and outgoing wave. The total cross section σ_s summarizes how much energy is scattered, the directionality is described by a phase function. These terms are introduced more formally in section 3.1.2.

The most important observation is the strong dependence of σ_s on the size of the particle with respect to the wavelength. Due to the complexity of solving Maxwell's equations for an arbitrary scattering, the discussion is restricted to a qualitative description of the arguably simplest geometric shape: the sphere.

The analytical solution for the scattering of a plane EM wave by a dielectric sphere is provided by Mie theory. Figure 3.2 illustrates the importance of the particle size and the frequency.
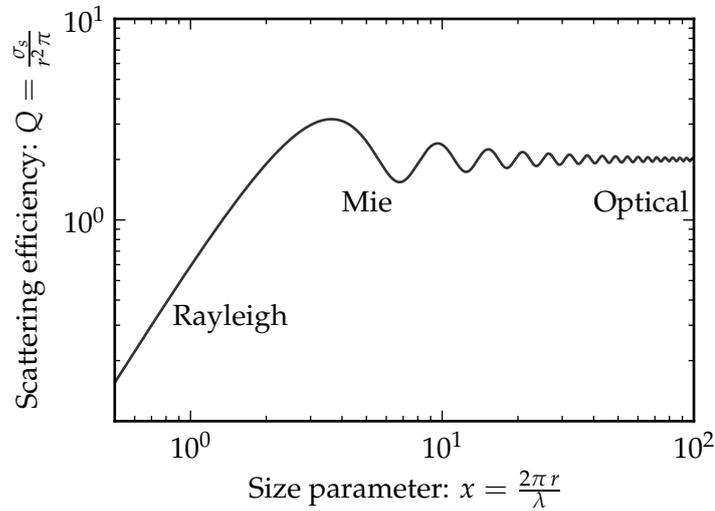
Figure 3.2.: Ratio Q of σ_s to the geometric cross section as predicted by Mie theory (in the Walstra approximation [104]) for $\varepsilon_r = 1.25$. The scattering behaviour of a perfectly conducting sphere shows similar patterns.

In the optical limit – the particle being much bigger than the wavelength – σ_s becomes proportional to the geometric cross-section. Strong fluctuations are characteristic for the transition region – the Mie regime.

The Rayleigh approximation is valid if the particle is smaller than the wavelength. In the microwave region, due to the long wavelengths, this regime is thus applicable for a wide range of scatterers – snow being the most important example in the context of this thesis. The total cross-section is given by [84]:

$$\sigma_s = \frac{128\,\pi^5}{3c^4}\, r^6\, f^4 \left(\frac{\varepsilon - 1}{\varepsilon + 2}\right)^2 \tag{3.19}$$

where r is the radius of the sphere, ε its permittivity and c the velocity of light. Note the strong dependence on the frequency f for fixed particle size.

For symmetry reasons, the directional behaviour is completely described by the angle between the incoming and outgoing wave, θ. If we do not distinguish different polarizations, the phase function is proportional to $1 + \cos(\theta)^2$. The scatterer can be thought of as a small dipole antenna, resulting in the two terms for unpolarized light. The dependence of the scattering behaviour on the polarization is more complicated for arbitrary scatters and beyond the scope of this thesis.

3.1.2. Radiative transfer

When describing the propagation of electromagnetic radiation through a medium that consists of many small particles (e.g. snow), it turns out to be convenient to consider the changes of intensity along the path. This intensity can be expressed mathematically as follows:

Definition *(Brightness)* The brightness (or radiance) L is defined by:

$$L = \frac{d^2\Phi}{dA \cos(\theta) d\Omega} \tag{3.20}$$

where Φ is the power, A the area and Ω the solid angle. The spectral brightness L_f is defined as the brightness per unit frequency.

Other radiometric quantities of interest can be obtained by integration; most importantly the power density S by integrating with respect to Ω. The power density is directly related to the power detected by the sensor, cf. section 3.1.3.

The radiative transfer equation (RTE) expresses the conservation of energy by balancing the change in power with the external losses and gains. By only considering the intensity rather than the **E**-field, the effects of interference due to coherent superposition cannot be modelled and are thus neglected. In the stationary case, the RTE can be written as

$$\frac{dL_f(s, \mathbf{\Omega})}{ds} = -\beta_f(s) L_f(s, \mathbf{\Omega}) + j_f + \frac{1}{4\pi} \alpha_f(s) \iint_{4\pi} L_f(s, \mathbf{\Omega}') p_f(\mathbf{\Omega}' \to \mathbf{\Omega}) d\mathbf{\Omega}' \tag{3.21}$$

where α_f, $\beta_f = \alpha_f + \gamma_f$ and γ_f are referred to as the spectral scattering, extinction and absorption coefficients, respectively. Often, $\omega = \alpha_f / \beta_f$ is called the single scattering albedo. j_f accounts for the thermal emission, but it is negligible compared to the other effects when dealing with active radar measurements. The phase function $p_f(\mathbf{\Omega}' \to \mathbf{\Omega})$ describes the scattering of radiation from direction $\mathbf{\Omega}'$ into $\mathbf{\Omega}$. This last term turns (3.21) into an integro-differential equation, which is extremely hard to solve in general.

For sparse media, it is commonly assumed that the scattering coefficient simply scales with the number of particles per unit volume N [30]:

$$\alpha = N \sigma_s \tag{3.22}$$

Copious methods for approximating solutions to the RTE have been proposed in the literature [37], many of which aim at replacing the integral with a sum by transforming, discretizing or parameterizing the solution and thus reducing the task to the solution of a

system of linear ordinary differential equations (ODEs). A very simple approach is the so-called two-stream approximation, whereby only two streams $F^\downarrow$, $F^\uparrow$ are considered. They satisfy a system of ODEs [70]:

$$\frac{d}{ds} \begin{pmatrix} F^\downarrow \\ F^\uparrow \end{pmatrix} = \begin{pmatrix} \gamma_1 & -\gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} \begin{pmatrix} F^\downarrow \\ F^\uparrow \end{pmatrix} \tag{3.23}$$

subject to appropriate boundary conditions[1]. Many methods lead to this system; the meaning of the $F^\downarrow$, $F^\uparrow$ and the coefficients depends on the specific approach [70]. Commonly, $L_f(\mathbf{\Omega})$ is assumed to be of a certain nature, and averaging with respect to the upper and lower hemisphere leads to the diffuse fluxes $F^\downarrow$, $F^\uparrow$.

Alternatively, if only forward and backward scattering are accounted for, $L_f(\mathbf{\Omega})$ can be thought of as a pair of delta functions: $F^\downarrow = S_1 \delta(\mathbf{\Omega})$ and similarly for $F^\uparrow$. This is usually referred to as the δ two-stream approximation. A further simplification can be achieved if the backward scattering is only considered if it points towards the sensor, corresponding to

$$\frac{d}{ds} \begin{pmatrix} F^\downarrow \\ F^\uparrow \end{pmatrix} = \begin{pmatrix} \gamma_1 & 0 \\ \gamma_2 & -\gamma_1 \end{pmatrix} \begin{pmatrix} F^\downarrow \\ F^\uparrow \end{pmatrix} \tag{3.24}$$

At the lower boundary surface at depth d, the upwelling radiation $F^\uparrow(d) = \gamma^0_{surf} F^\downarrow(d) = \sigma^0_{surf}/\cos(\theta) \, F^\downarrow(d)$. After reparameterization with respect to depth ($z = \cos(\theta) \, s$), (3.24) can be solved easily (e.g. by an eigendecomposition):

$$F^\uparrow(0) = \frac{\gamma_2}{2\gamma_1} \left(1 - e^{-2\gamma_1 d/\cos(\theta)} \right) F^\downarrow(0) + \gamma^0_{surf} e^{-2\gamma_1 d/\cos(\theta)} F^\downarrow(0)$$

$$\sigma^0 = \frac{\cos(\theta)\gamma_2}{2\gamma_1} \left(1 - e^{-2\gamma_1 d/\cos(\theta)} \right) + \sigma^0_{surf} e^{-2\gamma_1 d/\cos(\theta)} \tag{3.25}$$

The first term corresponds to the volume scattering, the second to the attenuated reflection. This interpretation forms the basis of the famous cloud model for vegetation [2] and also to the most widely used simple model for the scattering of snow, cf. section 3.2.2.

Several assumptions are inherent to the basic theory of radiative transfer as treated above, the most important of which are [90]:

- the scatterers are independent and do not interact
- the fields add incoherently (e.g. the scattering and the reflection from the boundary)
- the scattering is elastic so that the radiation at each frequency can be considered independently

The RTE in the form of (3.21) does not account for polarization; this can be circumvented by adopting the Stokes formalism [25].

3.1.3. Monostatic radar equation

The radar equation relates the target with the geometry and instrumental properties of the measurement. It is expressed in the quantities of radiometry described in table 3.1.

---

[1]The different signs of the coefficients are due to the fact that s is now a global parameter: compared to (3.21), the sign is flipped for $F^\uparrow$.

Table 3.1.: Quantities in the radar equation

Symbol	Unit	Description
P_e	W	power emitted by the antenna
P_r	W	power received by the antenna
S_i	Wm^{-2}	Irradiance incident at the target
S_s	Wm^{-2}	Irradiance scattered by the target
R	m	distance between the antenna and the target
σ	m^2	radar cross section
A_r	m^2	(effective) area of the antenna
G	1	gain w.r.t. an isotropic antenna with the same P_e

For an antenna emitting a power P_e in a direction where its gain is G, the irradiance at the target is given by $S_i = {}^{P_e G}/_{4\pi R^2}$. The target is replaced by a hypothetical perfectly scattering and isotropic scatterer of size σ in such a way that the irradiance at the receiving antenna is the same as the one actually measured [51]:

$$\sigma = \lim_{R \to \infty} 4\pi R^2 \frac{S_s}{S_i}$$

The receiving (isotropic) antenna thus measures a power of $P_r = S_s A_r$. Combining these results, we arrive at the monostatic radar equation for lossless, isotropic antennas:

$$P_r = S_s A_r = \frac{S_i \sigma}{4\pi R^2} A_r$$
$$= \frac{G P_e \sigma}{(4\pi)^2 R^4} A_r$$

The gain pattern of the antenna is described by G, which itself is related to the effective area and the wavelength [51]. The received power can thus be expressed by [106]

$$P_r = \frac{P_e \sigma G^2 \lambda^2}{(4\pi)^3 R^4} \tag{3.26}$$

Any losses in the atmosphere or due to the measuring device have to be accounted for separately, as does the polarization of the EM waves. Furthermore, both the gains and the radar cross section vary with direction.

Extended targets are conveniently described by a radar cross section per unit area σ^0, also referred to as normalized radar cross section and backscattering coefficient. σ^0 is a dimensionless quantity that is usually reported in decibels.

3.2. Backscatter behaviour

3.2.1. Permittivity of liquid and frozen water

Water

Water molecules have a permanent dipole moment. An external electric field **E** thus exerts a torque on them, causing them to preferentially align in the direction of **E** [34]. This

tendency is counteracted by the internal, thermal energy. This alignment and thus the macroscopic polarization is not instantaneous, but happens on a characteristic timescale τ_D, the relaxation time. The permittivity is thus a function of frequency and its dependence is often considered to follow the Debye model [64]:

$$\varepsilon(f) = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 - i\dfrac{f}{f_D}} \tag{3.27}$$

where $f_D = \tau_D^{-1}$ is the relaxation frequency and ε_s and ε_∞ the permittivity in the static and the high-frequency case.
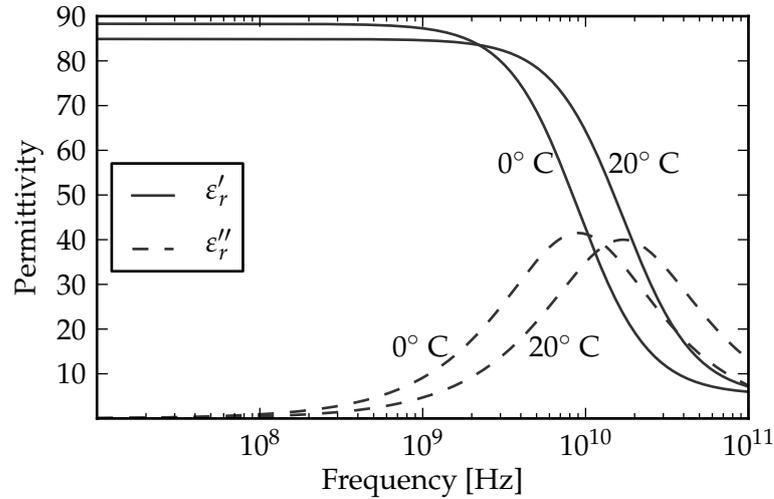
Figure 3.3.: Spectra for liquid water at $T = 0°\,C$ and $T = 20°\,C$ [71].

Figure 3.3 illustrates a Debye spectrum for water at $T = 0°\,C$ and $T = 20°\,C$. The imaginary part of the dielectric constant has a peak at f_D. The relaxation time decreases with increasing temperature.

Ice

In the solid phase, the structure of the lattice gives rise to a very high static permittivity [47]. It even turns out that the frequency dependence can also be approximated with the Debye model, albeit at a much smaller relaxation frequency. Consequently the dielectric constant is rather small at C- and K_u-band; even more so the absorption – cf. figure 3.4.

3.2.2. Land cover types

Bare ground

The backscatter of bare soil is mainly influenced by two features: roughness and soil moisture [106]. The effect of surface roughness is discussed in section 3.1.1.
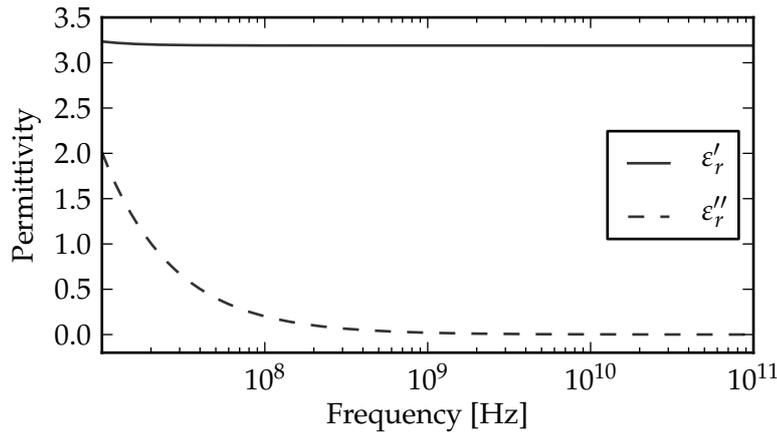
Figure 3.4.: Spectrum for pure ice at $T = -0.1°\,C$ [71].

The water content in the top layers of a soil is referred to as soil moisture. As already mentioned in section 3.2.1, the permittivity of water is remarkably high. The Fresnel equations as well as all models describing scattering from rough surfaces are a function of the dielectric constant; thus wet soils can be recognized by their high σ^0. The water molecules in the soil experience various kinds of interactions with the soil matrix which reduce their polarizability; consequently free and bound water should be distinguished [24]. The dielectric constant of soils as a function of soil moisture can be expressed mathematically by so-called mixing formulas, e.g. [12][24].

As the water in the soil freezes, the dielectric properties in the microwave range change: both ε'_r and ε''_r decrease and due to their comparatively small values, frozen soil shows striking similarities to very dry soil [105]. Tikhonov [96] provides a model as well as experimental data.

Water bodies

The huge dielectric contrast between air and water results in strong reflection – mainly in the specular direction unless the surface is sufficiently roughened by wind or otherwise. For the viewing geometries commonly employed for scatterometers and SARs, specular reflection leads to very low backscatter.

Periodic undulations that are of similar size as the wavelength are particularly important as they result in coherent Bragg scattering. The azimuthal dependence of this backscatter is used to infer the wind direction with scatterometers [40].

Vertically oriented features such as tree trunks can result in the so-called double bounce effect, whereby the surface acts like a corner reflector: a significant amount of the energy is reflected specularly by both the water surface and the vertical object [28].

Vegetation

Vegetation canopies consist of an intricate arrangement of scatterers of different sizes, shapes and orientations. In addition, the contribution of the underlying soil has to be con-

sidered. If the surface is sufficiently smooth, the vertical features of the plants can induce a so-called double bounce, cf. section 3.2.2. Due to the complexity of vegetation canopies, numerous studies on the backscattering characteristics have been conducted, some of them of entirely empirical nature, others based on modelling considerations. The observed dependence of σ^0 on frequency, polarization and incidence angle are of particular interest.

The radar cross section of a scattering object is very sensitive to the observation frequency. Consequently, small wavelengths such as K_u-band are subject to strong volume scattering from leaves and twigs, whereas longer wavelengths are less affected. The backscatter at P-band, for example, is dominated by large elements such as trunks and the underlying ground [106].

Vegetation canopies give rise to distinctive polarimetric characteristics of the backscatter. The cross-polarized backscatter σ^0_{hv} is usually very low for bare soil surfaces. Multiple scattering as encountered in dense canopies, however, leads to elevated σ^0_{hv}. The co-polarized backscatter is sensitive to the orientation of the canopy elements: the **E**-field of vertically polarized waves induces currents in vertical structures much more efficiently than in horizontally oriented ones (this is also the principle behind linear polarizers based on parallel wires).

Another key feature of vegetation canopies is the dependence of σ^0 on the incidence angle θ. The larger θ, the longer the optical path in the canopy and consequently the contribution of volume scattering compared to the surface. Very dense canopies can be assumed to be perfect volume scatterers at K_u- and C-band with good accuracy.

Water is the main constituent of most kinds of plant tissues. The permittivity of dried plants is in general rather small: around 2 for the real and less than 0.1 for the imaginary part. The dependence of the dielectric constant on the volumetric water content was studied by Ulaby and El-Rayes [99]. They developed a simple model, which distinguishes between bound and free water (this is somewhat similar to soil water). In the former, the interactions with surrounding molecules inhibit the rotational alignment: it is modelled as a Cole dielectric with a relaxation frequency $f = 0.18\,GHz$. The latter is assumed to behave like pure liquid water. Based on empirical data, the mixture data permits calculation of the permittivity as a function of the gravimetric moisture content; an example plot is shown in figure 3.5.

The freezing of water is accompanied by a large drop of the permittivity at the frequencies of interest. Depending on the moisture content in autumn, this trend also manifests itself in plant tissue to a certain extent [62]. From a qualitative point of view, the backscatter change due to freezing is thus expected to be quite similar in plants and soils.

A simple way to describe the backscatter from vegetation is the Cloud model by Attema and Ulaby [2]. It is equivalent to the simplified solution (3.25) to the Radiative Transfer Equation. More elaborate models such as MIMICS [100] have been proposed to obtain a better description of the scattering processes. The drawback, however, is that they tend to rely on a rather intricate parametrization of the canopy.

Snow

The backscatter from snow is often considered to consist of three parts: scattering from the top surface and the underlying surface as well as volume scattering from within the
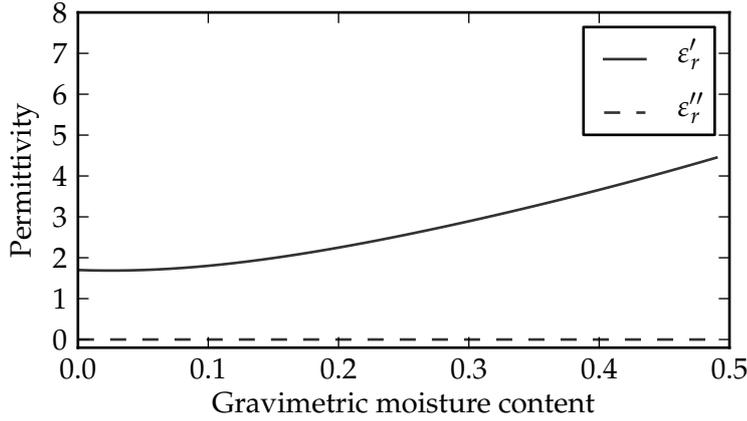
Figure 3.5.: Relative dielectric constant of plant tissue as a function of the gravimetric moisture content [99]. The difference between *C*- and K_u-band is neglible, so only K_u-band data are shown.

snow pack. A commonly used model is the extension of the simple RTE model (3.25) with a surface scattering term:

$$\sigma^0 = \frac{\cos(\theta)\gamma_2}{2\gamma_1} \left(1 - e^{-2\gamma_1 d/\cos(\theta)}\right) + \sigma^0_{ground} e^{-2\gamma_1 d/\cos(\theta)} + \sigma^0_{snow\_surface} \qquad (3.28)$$

where the notation of the preceding sections is employed. The coefficients γ_1 and γ_2 have to be related to the shape, size and density of the snow grains.

Due to its simplicity, this model or variants thereof are commonly applied, e.g. [1][72]. Its most important limitations are the neglect of multiple scattering, the simplistic description of the particles (spherical, uniform size) and that the scatterers are treated independently. The interactions, however, can become significant for dense snow packs. In order to alleviate these problems, several approaches have been suggested, e.g. [33].

A snow pack consists of particles of a variety of sizes and shapes. Consequently, the radius *r* in (3.19), as well as other parameters in different models, refer to an effective quantity which is chosen as to mimic the overall backscattering behaviour well. Snow grains act as Rayleigh scatterers in the microwave region, thus leading to a strong dependence on frequency and particle size.

Dry snow exhibits very little absorption; the extinction coefficient is dominated by the scattering coefficient. This changes drastically with the introduction of liquid water, which is characterized by a large ε_r''. The permittivity as a function of liquid water content can be described by mixing formulas, e.g. [68]. The simple model in (3.28) correctly predicts a strong diminishment of the backscatter in the presence of liquid water in the snow. This is caused by the steep increase in absorption.

The remarkable dependence of σ^0 with respect to frequency is illustrated in figure 3.6. The behaviour is generated using the model described in section 6.2, where more details about the implementation can be found.
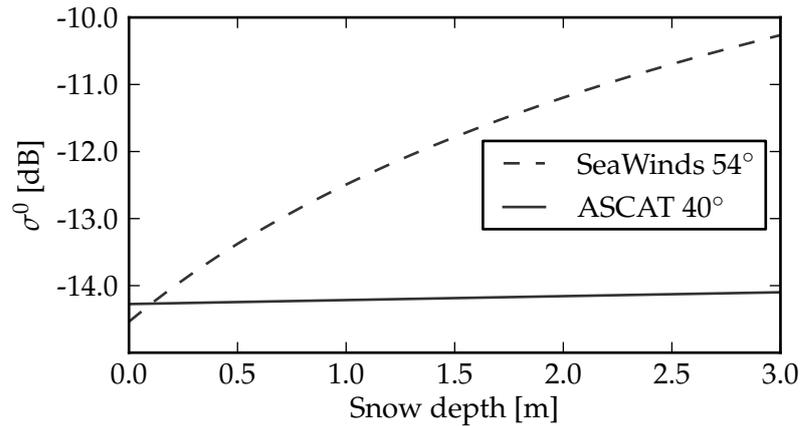
Figure 3.6.: Simulation of σ^0 of dry snow for the frequencies and incidence angles of QuikScat and ASCAT; cf. section 6.2 for a detailed description of the model and 7.1.2 for the parameters used.

Freshwater ice

Many different parameters govern the backscatter characteristics of ice, so that it is challenging to infer geophysical variables using radar data. The most important factors influencing σ^0 are:

Surface Roughness Differences in the texture of the ice surface can be brought about by e.g. the movement of water masses and pressure due to the loading with snow [55]. Increased roughness leads to a higher backscatter coefficient in general.

Grounded vs. floating ice For the frequencies of interest, the difference in the permittivity between water and either ice or soil is much bigger than between the latter. The reflectivity is thus greatly enhanced in the presence of a liquid water layer between the ice and the ground. Floating ice appears brighter in radar images than grounded ice, all other factors being equal.

Bubbles and inclusions The number, sizes and orientations of bubbles depend on many factors during ice formation and metamorphosis [55]. They influence the volume scattering behaviour and thus the σ^0 of the ice. Vertically oriented, tubular bubbles have been reported to increase backscatter in floating ice due to a double bounce mechanism [27].

Snow cover and layered structure Due to the coherent nature of radar systems, layers of different electromagnetic properties can lead to interference effects – similar to a Fabry-Pérot interferometer, but more complex because of the geometry, surface roughness and volume scattering. Snow-covered ice is very common and the increase in its liquid water content leads to a sharp decrease in backscatter during melting.

3. Remote Sensing of the Freeze/Thaw State

4. Hidden Markov Models

A Hidden Markov model (HMM) is a probabilistic model of a sequence of observations. These observations are assumed to be generated by an unobserved process. HMMs have been successfully applied in a host of scientific applications [18]. Their use is particularly widespread in the analysis of DNA and protein sequences, as well as in text and speech recognition.

Ghahramani [41] gives a brief overview of Hidden Markov models and some of their generalizations, whereas Elliott et al. [31] provide an in-depth, self-contained treatment.

This chapter contains an elementary presentation of Belief networks, which are graphical representations of probability distributions. They are introduced because Hidden Markov models can be conveniently expressed as Belief networks. Subsequently, the probabilistic structure of HMMs, as well as the important tasks of inference and the learning of parameters are described.

4.1. Belief networks

Belief networks (BN), also called Bayesian networks, are a widespread form of graphical model of multivariate probability distributions. In a graphical model, each random variable is represented as a vertex in a graph. The edges of the graph encode relations between the variables. There are different kinds of graphical models, but Bayesian networks are among the most common and are particularly amenable for the description of Hidden Markov models. The following brief survey mainly follows Barber [4] and Bishop [13].

Random variables are denoted by capital letters, e.g. X. In the discrete case, they can can take on any of the states $x_1 \ldots x_n$ in their domain Ω_X and the probability of the event $X = x_i$ is described by a probability mass function $P(X = x_i)$. For simplicity, in the continuous case, the probability density function will also be denoted in the same way $P(X = x)$. When describing graphical models, it is commonplace to abbreviate $P(X = x)$ as $P(x)$, thus making the random variable and the associated distribution function implicit. When marginalizing over a variable, $\sum_x P(x, y)$ is shorthand for $\sum_{s \in \Omega_X} P(X = s, Y = y)$.

Belief networks are motivated by the fact that any probability distribution over the random variables $X_1, X_2 \ldots X_N$ can be written in cascade form [4]:

$$P(x_1, x_2 \ldots x_N) = P(x_1)P(x_2|x_1) \ldots P(x_N|x_1, x_2 \ldots x_{N-1})$$

More specific distributions might allow for factorizations where some of the dependencies in can be dropped, e.g.

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_2) \tag{4.1a}$$
$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1) \tag{4.1b}$$
$$P(x_1, x_2, x_3, x_4) = P(x_1)P(x_2)P(x_3|x_1, x_2)P(x_4|x_3) \tag{4.1c}$$

Belief networks are graphical representations of such factorizations. Each node (vertex) of the graph encodes a variable. A directed link is drawn from node i to j if j is conditioned on i in the factorization. Figure 4.1 illustrates these principles by depicting the BN corresponding to equation 4.1.

(a) Markov Chain   (b) 3-node BN   (c) 4-node BN
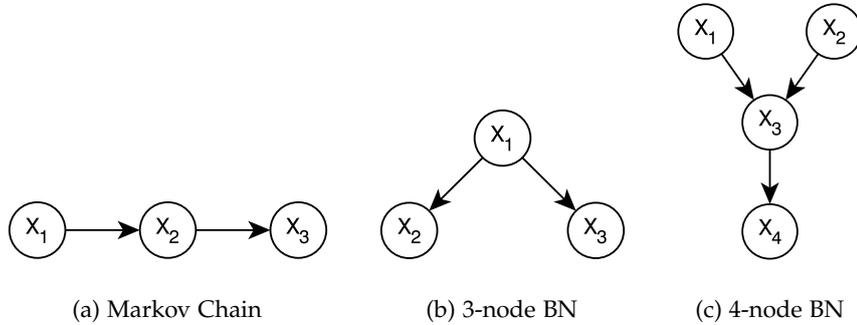
Figure 4.1.: Three simple Belief networks, cf. equation 4.1 for the corresponding factorizations

For the subsequent sections some terminology has to be introduced. We follow the notation of [4]. If there is a directed arc from node i to j, j is said to be a child of i and i a parent of j. A path is a sequence of vertices and two consecutive vertices are connected by an edge. Note that, unless explicitly stated otherwise, we do not consider the direction of the arcs in the definition of a path, i.e. we neglect the arrows. A vertex i is called a collider relative to a path, if the arcs connecting it to both its neighbours point to i.

These concepts are now illustrated by referring to figure 4.1. In figure 4.1a, X_2 is child of X_1 and parent of X_3 In figure 4.1b, X_2 and X_3 are both children of X_1. The fact that a collider is defined relative to a path can be appreciated in figure 4.1c, where X_3 is a collider on the path from X_1 to X_2, but not from X_1 to X_4.

4.1.1. Conditional independence

The links in the graph, or rather the lack thereof, give important information about the conditional independence relations of the probability distribution. Variable X is said to be conditionally independent of Y given Z if for all x, y, z $P(X = x, Y = y | Z = z) = P(X = x | Z = z)P(Y = y | Z = z)$ or equivalently if $P(X = x | Y = y, Z = z) = P(X = x | Z = z)$.

The concept of d-separation allows one to read off the conditional independence structures of a distribution from its graph. The following definition is taken almost verbatim from [4].

Definition *(d-separation)* Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be disjoint sets of vertices of a graph G. $\mathcal{X}$ and $\mathcal{Y}$ are d-separated iff all paths[1] between any $x \in \mathcal{X}$ and any $y \in \mathcal{Y}$ are blocked.

A path U is said to be blocked if there is a node w on U such that either

- w is a collider and neither w nor any of its descendants is in $\mathcal{Z}$

---

[1] the direction of the arcs is not taken into account

- w is not a collider on U and w is in $\mathcal{Z}$

The sets of variables $\mathcal{X}$ and $\mathcal{Y}$ are conditionally independent given $\mathcal{Z}$ in any probability distribution that can be represented by G, if they are d-separated by $\mathcal{Z}$.

4.2. Probabilistic structure of Hidden Markov models

An HMM is a probability distribution of two sequences of random variables: $X_{1:T}$[2] and $Y_{1:T}$. X_i and Y_i are called the state and observation random variable at time i, respectively.

HMMs get their name from the fact that the sequence $x_{1:T}$ is not known. At a given epoch i, however, the current value of the state x_i influences the probability distribution of the variable Y_i, whose observed realization is y_i.

Definition *(Hidden Markov Model)* Let $X_{1:T}$ and $Y_{1:T}$ be sequences of random variables. Each X_i is a discrete random variable over the states $s_1 \ldots s_n$. Each observation y_i can be multi-dimensional and either continuous or discrete. The joint probability of a Hidden Markov Model is given by [41]:

$$P(x_{1:T}, y_{1:T}) = P(x_1)P(y_1|x_1)\prod_{t=2}^{T} P(x_t|x_{t-1})P(y_t|x_t) \tag{4.2}$$

In this notation, x_i is a realization of the random variable X_i and thus represents an arbitrary fixed element of the state space $\Omega_{X_i} = \{s_1 \ldots s_n\}$. The Hidden Markov Model can thus be represented by the Bayesian network shown in figure 4.2.
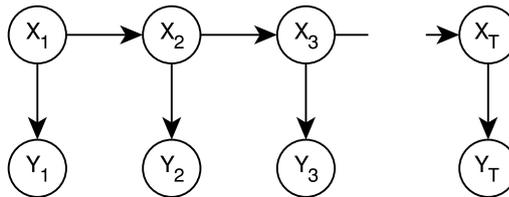
Figure 4.2.: Belief network of the Hidden Markov model.

The Belief network in figure 4.2 encodes many different conditional independence assumptions. The following two are the most important ones and can be easily checked using the concept of d-separation:

- Two observations Y_i and Y_j are independent given any state X_k if $i \leq k \leq j$. Without conditioning on any state, they are dependent.

- Given the current state X_t, any future state X_u is independent from any past state X_v, where $u > t$ and $v < t$. They are said to fulfill the Markov Property [42].

---

[2]$X_{1:T}$ is shorthand for $X_1, \ldots X_T$

Probability distributions that can be described by the BN of figure 4.2 are often referred to as state-space models, of which the HMM is an important example. Characteristics of this BN such as the Markov Property also hold when the state variable is continuous. An important example of this class is the linear Gaussian state-space model, which is also often called Kalman Filter. This nomenclature, however, is slightly confusing because it does not distinguish between the probabilistic model and an algorithm to estimate the state [5].

4.3. Inference

Inference describes the determination of posterior probabilities given some observed variables [13]. Such computations can become intractable very quickly for general probability distributions, but due to the structure of the Hidden Markov Model, several types of inference can be done efficiently. In this thesis, we require two different kinds.

Firstly, we are interested in the probability distribution of the state variable X_i at a fixed epoch i, having observed $Y_{1:T} = y_{1:T}$. That is, we want to compute $P(X_i = x_i | Y_{1:T} = y_{1:T})$. This can be achieved in a fast way with the Forward-Backward algorithm.

Secondly, the most probable sequence of states $X_{1:T}^\star$ is of interest and can be determined by the Viterbi algorithm.

4.3.1. Forward-Backward algorithm

The objective of the Forward-Backward algorithm is to compute $P(x_i | y_{1:T})$ for a given i. It turns out that it is hardly more time-consuming to compute the posterior probabilities for all epochs i.

The quantity of interest can be expressed in a more amenable form:

$$
\begin{aligned}
P(x_i | y_{1:T}) = \frac{P(x_i, y_{1:T})}{P(y_{1:T})} &\propto P(x_i, y_{1:T}) \\
&= \sum_{x_{1:i-1}, x_{i+1:T}} P(x_{1:i}, x_{i+1:T}, y_{1:i}, y_{i+1:T}) \\
&= \underbrace{\sum_{x_{1:i-1}} P(x_{1:i}, y_{1:i})}_{P(x_i, y_{1:i}) \equiv \alpha(x_i)} \underbrace{\sum_{x_{i+1:T}} P(x_{i+1:T}, y_{i+1:T} | x_i)}_{P(y_{i+1:T} | x_i) \equiv \beta(x_i)}
\end{aligned}
$$

The simplification in the last step is allowed due to the distributive law and the conditional independence relations. $\alpha(x_i)$ can be expanded as

$$
\begin{aligned}
\alpha(x_i) &= P(x_i, y_{1:i}) \\
&= \sum_{x_{i-1}} P(x_i, x_{i-1}, y_{1:i-1}, y_i) \\
&= \sum_{x_{i-1}} P(y_i | x_i, x_{i-1}, y_{1:i-1}) P(x_i | x_{i-1}, y_{1:i-1}) P(x_{i-1}, y_{1:i-1}) \\
&= \sum_{x_{i-1}} P(y_i | x_i) P(x_i | x_{i-1}) \underbrace{P(x_{i-1}, y_{1:i-1})}_{\alpha(x_{i-1})}
\end{aligned}
$$

Algorithm 1 Alpha messages

Base case:

$$\alpha(x_1) = P(x_1)P(y_1|x_1) \tag{4.3}$$

Recursion step:

$$\alpha(x_i) = P(y_i|x_i) \sum_{x_{i-1}} P(x_i|x_{i-1})\alpha(x_{i-1}) \tag{4.4}$$

Algorithm 2 Beta messages

Base case:

$$\beta(x_T) = 1 \tag{4.5}$$

Recursion step:

$$\beta(x_i) = \sum_{x_{i+1}} P(y_{i+1}|x_{i+1})P(x_{i+1}|x_i)\beta(x_{i+1}) \tag{4.6}$$

This is the key idea of the recursive scheme outlined in algorithm 1, and similarly for the β-messages in algorithm 2.

These two types of messages can be combined to compute the required posterior distribution $P(x_i|y_{1:T})$. The numerator ensures the normalization of the posterior. Because of this, the messages can also be normalized at intermediate steps to alleviate numerical errors [4]. It can also be helpful to work with the logarithms of the messages [13]. Provided all the α and β-messages are stored, the posterior can be evaluated very efficiently at any time step.

$$P(x_i|y_{1:T}) = \frac{\alpha(x_i)\beta(x_i)}{\sum_{x_i}\alpha(x_i)\beta(x_i)} \tag{4.7}$$

Note that this approach can also be applied to more general graphs, namely those without undirected loops. These are called singly-connected graphs and the message-passing is referred to as the Sum-Product algorithm [13].

The computation of the marginal probabilities is only possible by clever exploitation of the structure of the joint distribution. The number of possible time-series $x_{1:T}$ grows as $\mathcal{O}(N^T)$ where N is the number of states. Evaluating the posterior distribution $P(x_i|y_{1:T})$ corresponds to taking weighted averages of this exponentially growing number of time-series of the state, which would be infeasible even for small T. The message passing framework ensures that this can be done efficiently, namely in time linear in the length, $\mathcal{O}(TN^2)$.

In case of a continuous state, the idea of the recursive message passing scheme can still be applied. The sums, however, have to be replaced by integrals and these integrals cannot be evaluated in closed form in general [5]. The particle filter approximates the probability

density functions by sampling and the integrals thus reduce to sums. The linear Gaussian state-space model, on the other hand, has the property that all probability distributions involved are Gaussians [4]; the message passing can thus be reduced to sequential computation of the mean and covariance. This is done by the Kalman Filter, which corresponds to the α-messages.

4.3.2. Viterbi algorithm

The Viterbi algorithm can be employed to find the most probable sequence of states $X_{1:T}^{\star}$ conditioned on all the observations. For fixed $y_{1:T}$, this is equivalent to maximizing the joint probability $P(x_{1:T}, y_{1:T})$:

$$X_{1:T}^{\star} = \arg\max_{x_1 \dots x_T} P(x_{1:T}, y_{1:T}) \tag{4.8}$$

As probabilities are non-negative, the operation of maximization is distributive over multiplication, so that message-passing can be employed to make the computations very efficient. The computations are almost the same as in the Forward-Backward algorithm. On more general singly-connected graphs, the sum product algorithm is replaced with the max product algorithm, of which the Viterbi algorithm is a special case. The following formulae (with a slightly different notation) are taken from [4].

Algorithm 3 Viterbi algorithm

Forward Pass

    Base case:

$$\mu(x_T) = 1 \tag{4.9}$$

    Recursion step:

$$\mu(x_{i-1}) = \max_{x_i} P(y_i|x_i)P(x_i|x_{i-1})\mu(x_i) \tag{4.10}$$

Backward Pass

    Base case:

$$X_1^{\star} = \arg\max_{x_1} P(x_1)P(y_1|x_1)\mu(x_1) \tag{4.11}$$

    Recursion step:

$$X_i^{\star} = \arg\max_{x_i} P(x_i|X_{i-1}^{\star})P(y_1|x_1)\mu(x_i) \tag{4.12}$$

The latter pair of equations in algorithm 3 is also referred to as backtracking. Because the logarithm is a monotonous function, it is again possible to work with logarithms [13].

4.4. Learning

In Machine Learning and Pattern Recognition, the expression 'Learning' refers to the estimation of parameters. In this section, the paradigms of Maximum Likelihood and Maximum A Posteriori estimation are presented. In presence of hidden variables, the likelihood is usually very difficult to evaluate and manipulate, so that more sophisticated tools are called for. One of these is the Expectation Maximization algorithm, which can be used to estimate parameters in Hidden Markov models.

4.4.1. Maximum likelihood and maximum a posteriori

Every sufficiently complex model of a real-world phenomenon has parameters, which are not or only approximately known. The objective of learning is the estimation of these parameters.

According to the Bayesian philosophy, parameters are treated as random variables. Having observed some data $\mathcal{D}$, the posterior distribution of the parameters θ is given by Bayes' rule [67]:

$$P(\theta|\mathcal{D}) = \frac{P(\mathcal{D}|\theta)P(\theta)}{P(\mathcal{D})} \tag{4.13}$$

where $P(\mathcal{D}|\theta)$ is called the likelihood of θ and $P(\theta)$ is referred to as the prior. It encodes the a priori assumptions of the modeller. Predictions and inferences are performed by averaging with respect to all the parameters. These computations, although straight-forward, are very challenging from a computational point of view [67]. Also for this reason, it is common to only consider point estimates, i.e. to determine one vector of parameters.

(4.13) suggests to take the maximum of either $P(\mathcal{D}|\theta)$ or $P(\mathcal{D}|\theta)P(\theta)$ as point estimate. These approaches are called the Maximum Likelihood and the Maximum A Posteriori estimates, respectively. These, particularly the former, also have a rich history in their own right. As Maximum Likelihood estimators can be derived in a systematic fashion and can be shown to possess very pleasing properties as the number of observations grows, they are very popular.

It is common to work with the logarithm of the likelihood (log likelihood), because it tends to be more suited for computations.

4.4.2. Learning in presence of hidden variables

The log likelihood $l(\theta)$ of the parameters θ often becomes difficult to evaluate and handle in presence of latent variables. More specifically for the HMM, where $y_{1:T}$ is one sequence of observations:

$$l(\theta) = \log\left(P(y_{1:T}|\theta)\right) = \log\left(\sum_{x_{1:T}} P(x_{1:T}, y_{1:T}|\theta)\right) \tag{4.14}$$

Note that $P(Y|\theta)$ simply means that the parameters of the distribution take the value θ. The function on the right-hand side of (4.14) is also referred to as the marginal log likelihood. It only provides reasonable information about the parameters if the hidden variables are 'missing at random' [4], which, however, is the case for the given problem.

The marginal log likelihood is rather cumbersome to deal with because it is the logarithm of a sum. The EM algorithm provides an elegant framework for estimating the parameters without having to work with $l(\theta)$ directly.

Before describing the algorithm, the Kullback-Leibler divergence has to be introduced.

Definition *(Kullback-Leibler (KL) divergence, relative entropy)* The KL divergence between two probability distributions $P(X)$, $Q(X)$ is given by [67]:

$$D_{KL}(P||Q) = \left\langle \log \frac{P(x)}{Q(x)} \right\rangle_{P(x)} \tag{4.15}$$

where $\langle\rangle_{P(x)}$ denotes averaging with respect to $P(x)$. $D_{KL}(P||Q)$ can be shown to be non-negative, with equality if and only if the distributions P and Q are identical.

Expectation Maximization algorithm

There are different ways to motivate and interpret the EM algorithm. This concise survey adopts the approach taken by Neal[73]. In the following, Y and X refer to all the observed and hidden variables, respectively.

Instead of maximizing the log likelihood $l(\theta)$, the objective function is taken to be $F(\theta, Q(x))$, which is analogous to the 'variational free energy' in statistical physics [67]. $Q(x)$ is a probability distribution over the hidden variables. $F(\theta, Q(x))$ is given by the following equivalent expressions [73]:

$$F(\theta, Q(x)) = \langle \log P(x, y|\theta) \rangle_{Q(x)} - \langle \log Q(x) \rangle_{Q(x)} \tag{4.16}$$
$$= l(\theta) - D_{KL}(Q(x)||P(x|y, \theta)) \tag{4.17}$$

(4.16) puts emphasis on the decomposition into an energy term and an entropy term. The former, $\langle \log P(x, y|\theta) \rangle_{Q(x)}$, has a much more pleasing structure than the log likelihood because it is the sum of logarithms rather than the other way round. It is this fact that makes the optimization problem easier. (4.17), on the other hand, shows that the variational free energy is a lower bound on $l(\theta)$, which the EM algorithm maximizes by performing the two steps in algorithm 4 repeatedly [73].

Algorithm 4 Expectation Maximization

E-Step Set $Q(x)$ to $P(x|y, \theta)$

M-Step Choose θ so that $F(\theta, Q(x))$ or equivalently $\langle \log P(x, y|\theta) \rangle_{Q(x)}$ is maximized

---

Due to the non-negativity of the KL-Divergence and (4.17), the E-Step also corresponds to a maximization of $F(\theta, Q(x))$ w.r.t. $Q(x)$. The EM algorithm can thus be interpreted as a coordinate descent scheme on the variational free energy.

Determining $Q(x)$ in the E-Step corresponds to performing inference in the probability distribution. In the given case, the Forward-Backward algorithm ensures that this can be done efficiently.

It can be shown that the likelihood increases after each step, which, for bounded $l(\theta)$, entails convergence (of the likelihood)[108]. There is no guarantee that the EM algorithm converges to the global optimum, so that it is advisable to use different initial values [4].

5. Data Sources

5.1. SeaWinds on QuikScat

5.1.1. Instrument

SeaWinds was a K_u band scatterometer instrument onboard the QuikScat satellite, which was launched on June 19th, 1999 [97]. It was the first 'pencil-beam' scatterometer in space – the spatial coverage was achieved by a spinning dish. This rotating motion enabled the instrument to sample the swath with its two beams: inner and outer. Table 5.1 contains a summary of the two modes.

Table 5.1.: Key parameters of the two modes. Source: [97].

Parameter	Inner Beam	Outer Beam
Polarization	HH	VV
Frequency [GHz]	13.4	13.4
Incidence Angle [°]	47	55
Slant Range [km]	1100	1245

SeaWinds employed pulse compression using a linear chirp to achieve range discrimination inside the footprint [107]. The range-resolved measurements are referred to as 'slices' [93], whereas the measurements encompassing the entire footprint are called 'eggs'. As the noise of the σ^0 measurements is considered to be of prime importance in most applications, the slices are averaged to composites to achieve the required balance between spatial and radiometric resolution. The technical specifications of the data products can be found in [58].

On November 23rd, 2009, NASA [17] reported that the mechanical system responsible for spinning the antenna ceased operating due to wear and tear. The life-span of the instrument was much longer than the two-year period originally envisaged.

5.1.2. Pre-processing

The data are taken from the archive at the Institute of Photogrammetry and Remote Sensing, TU Vienna, a detailed description of which can be found in [60].

The raw data were processed at the Jet Propulsion Laboratory (JPL), and published in different formats. The L2A product consists of slice-composite backscatter measurements and they are spatially grouped into 'wind vector cells', which make up a preliminary grid for each orbit [58].

In order to analyze the data in a systematic fashion and conduct time-series studies, these σ_0 measurements including relevant meta-data have to be re-gridded.

In the processing chain described by [60], the same ellipsoid (GRS 80) as the one used by JPL is adopted. The surface of this ellipsoid is tesselated by a grid with a fixed spacing of $10km$. These grid points are further grouped into cells to facilitate processing tasks.

All σ^0 measurements extracted from the L2A product whose centre of the footprint is within 12.5km of the centre of the grid point are allocated to the respective grid point.

5.2. ASCAT on MetOp

5.2.1. Instrument

ASCAT is the *C* band scatterometer onboard the MetOp-A satellite, which was launched into its sun-synchronous orbit in October 2006 [32]. Like its predecessor SCAT onboard the ERS satellites, it is a fan beam instrument, but it contains six antennae, three on each side [40]. The two swaths guarantee an improved temporal coverage compared to the previous fan beam instruments [35]. The main characteristics of the scatterometer are summarized in table 5.2.

Table 5.2.: Characteristics of the ASCAT instrument. Adapted from [40]

Parameter	Value
Frequency	5.255GHz
Polarization	VV
Swath Width	550km

5.2.2. Pre-processing

In order to derive a σ^0 value, normalized to an incidence angle of $40°$, the following pre-processing steps have to be applied [9]:

Resampling The backscatter coefficient σ^0 and additional parameters are resampled to a grid, which is similar to the one used for archiving the QuikScat data but based on the GEM6 Ellipsoid [8].

Azimuthal normalization An azimuthal correction, which also depends on the incidence angle, is applied. The value of the correction is estimated based on previously acquired data.

Incidence angle normalization The backscatter coefficient σ^0 is extrapolated to an incidence angle of $\theta = 40°$, based on a second-order polynomial. As the dependence of σ^0 on θ is very sensitive to the vegetation cover, the parameters of the polynomial have to be determined dynamically; cf. [9] for a detailed description.

5.3. ERA-Interim

The ERA Interim data set, produced by the European Centre for Medium Range Weather Forecasts (ECMWF), encompasses a host of meteorological and oceanographic parameters

[11]. It is a data assimilation scheme that incorporates in-situ and satellite observations into a physical model of the atmosphere, oceans and land surface [92]. The spatial resolution of the model and the data is about 80km (which is less than the one of the ECMWF Forecasts). The set of parameters includes the air temperature at two meter height and the temperature of the topmost soil layer.

5.4. IMS

The Interactive Multisensor Snow and Ice Mapping System (IMS) is a global snow and sea/lake ice data set produced by the National Oceanic and Atmospheric Administration's National Environmental Satellite Data and Information Service (NOAA/NESDIS) [76]. The reference resolution is $4\,km$ but a product with $24\,km$ resolution is disseminated as well; this is the data used in this work. The areal extent of the snow and ice cover are derived manually by analysts on the basis of the following data sources [46]:

Visible and IR instruments on polar satellites e.g. AVHRR, MODIS

Visible and IR instruments on geostationary satellites e.g. Meteosat-5

Microwave radiometers e.g. AMSR-E

Active microwave instruments e.g. ASCAT, ASAR

External products meteorological data, snow models (e.g. SNODAS)

According to Helfrich et al. [46], the analysts prefer to work with visible and IR data but have to rely on microwave and external data when the former are unavailable or inconclusive; this can be due to the viewing geometry at high latitudes, insufficient illumination or cloud cover.

5.5. In-situ data

5.5.1. Snow measurements

The data set DSI-9808 by the Russian National Climate Data Centre [87] encompasses around 400 stations in Russia. At these locations, several properties of the snow cover are monitored every 10 days (or more frequently during the melt periods). Among the variables recorded in the database are: snow depth, snow structure and ice crust thickness.

5.5.2. WMO CISL RDA ds512.0

This data set, a global collection of in-situ meteorological observations, is maintained by the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR) [101]. Subsequently, it will often be referred to as the 'WMO' data set. Among the variables stored in the archive are the maximum and minimum daily temperature and precipitation. In addition, observations of the snow height are available at some of the stations.

5. Data Sources

6. Implementation

In this chapter, we describe the processing chain that yields the estimate of the Freeze/Thaw state, thereby drawing heavily on the theory presented in the preceding chapters. Subsequently, the backscatter model will be presented. It allows us to better understand the signals we observe and intend to combine.

6.1. Sensor fusion model

The flowchart in figure 6.1 illustrates the main steps:

Preprocessing At this stage, the input data are read and the probabilistic model is set up.

Parameter Estimation After having retrieved initial values, the parameters are estimated with the help of the EM Algorithm.

Inference and Output The quantities of interest such as the marginal posterior distributions are inferred. Subsequently, these data are summarized to facilitate their interpretation.
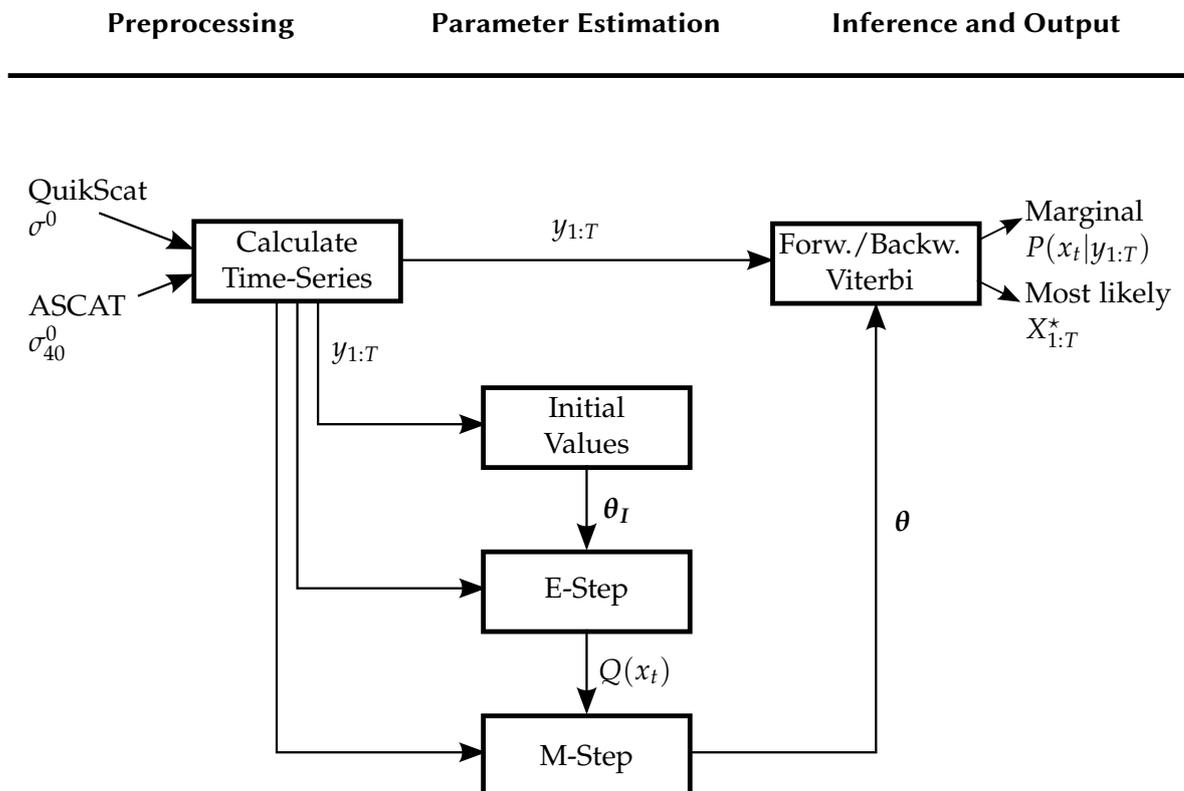
Figure 6.1.: Illustration of the data flow.

6.1.1. Preprocessing

Probabilistic Model

The probability distribution of a Hidden Markov Model is given by equation 4.2

$$P(x_{1:T}, y_{1:T}) = P(x_1)P(y_1|x_1) \prod_{t=2}^{T} P(x_t|x_{t-1})P(y_t|x_t) \tag{6.1}$$

X_t and Y_t being the random variables at epoch t of the states and observations, respectively.

The key properties of this model were discussed in chapter 4. In the context of this thesis, we will associate the latent variables X_t with the Freeze/Thaw state and the observable variables Y_t with the input data.

The idea of the probabilistic modelling can be summarized as follows: the observed quantities and the unknown state are assumed to follow the distribution (6.1). The probabilistic framework allows us to calculate probabilities of the Freeze/Thaw state based on these observations.

Input data

This model is very flexible; adaptations to incorporate different sensors as well as weather forecasts or in situ data are easily possible. In this thesis, however, we concentrate on the data described in chapter 5: SeaWinds outer beam and the ASCAT measurements reduced to a fixed incidence angle of $\theta = 40°$.

Depending on several factors, there can be varying numbers of measurements by each instrument. To illustrate: SeaWinds can acquire more than 25 recordings for each beam at high latitudes but significantly fewer closer to the Equator. This is why we summarize the data by extracting the following three features:

SeaWinds Difference: *swd* The difference of the mean QuikScat σ^0 values (in *dB*) acquired in the morning and the evening[1]. This is in analogy to the diurnal difference approach discussed in section 1.4.

SeaWinds Mean: *swm* The daily mean of the SeaWinds σ^0 values.

ASCAT Mean: *am* The daily mean of the ASCAT σ^0 values. The grid point closest to QuikScat point is considered.

The reason for there being no ASCAT diurnal difference feature is the comparably small number of observations by this instrument even at high latitudes.

These three features are assumed to be independent conditioned on the state X_t. The Belief network implied by this assumption is sketched in figure 6.2. For each state and each feature, we assume that the conditional probability distribution (emission probability) can be reasonably well approximated by a Gaussian distribution.

---

[1]Morning: 01:00 - 05:00; Evening: 17:00 - 21:00. Measured in mean local solar time.
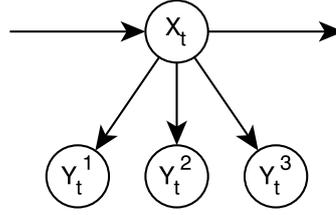
Figure 6.2.: Belief network for a HMM with three independent observational random variables, indicated by the superscripts.

States

The states X_t are discrete random variables; in our case they correspond to the freeze/thaw state we want to infer. As briefly described in section 1.4, the existing approaches usually consider either two or three states. The former just distinguish between frozen and non-frozen, the latter also allow for a 'thawing' state, characterized by diurnal thawing and freezing.

Experiments have shown, however, that due to the different properties of K_u and C band, a fourth state is called for: the soil being frozen but without snow cover. During these times, C band backscatter is low (like it usually is in winter), but so is K_u band due to the lack of snow. Note that, if we only dealt with C band data, we would also restrict ourselves to three states. The sample space of the states X_t is thus $S_4 = \{0, 1, 2, 3\}$:

0 Frozen, with snow cover

1 Non-Frozen

2 Thawing

3 Frozen, without snow cover

6.1.2. Estimation of parameters

Based on the assumptions described in the previous section, the probabilistic model of (6.1) can be expanded as

$$
\begin{aligned}
P(x_{1:T}, y_{1:T}) =\, & P(x_1; \boldsymbol{\theta_1}) P(y_1^{swd} | x_1; \mu_{swd}, \sigma_{swd}^2) \\
& \cdot P(y_1^{swm} | x_1; \mu_{swm}, \sigma_{swm}^2) P(y_1^{am} | x_1; \mu_{am}, \sigma_{am}^2) \\
& \cdot \prod_{t=2}^{T} P(x_t | x_{t-1}; \mathbf{A}) P(y_t^{swd} | x_t; \mu_{swd}, \sigma_{swd}^2) \\
& \cdot P(y_t^{swm} | x_t; \mu_{swm}, \sigma_{swm}^2) P(y_t^{am} | x_t; \mu_{am}, \sigma_{am}^2)
\end{aligned}
$$ (6.2)

where the symbols after the semicolons represent the parameters of the respective term. Due to the varying number of observations, the variances of the emission probability terms are not constant:

SeaWinds diurnal difference *swd* $\quad \sigma_i^2[t] = \sigma_{swd,i}^2 {}^{(N_{am}[t] + N_{pm}[t])}/{N_{am}[t] N_{pm}[t]}$

SeaWinds mean *swm* $\quad \sigma_i^2[t] = \sigma_{swm,i}^2/N[t]$

ASCAT mean *am* $\sigma_i^2[t] = \sigma_{am,i}^2/N[t]$

where N is the total number of observations for the given sensor, and N_{am} (N_{pm}) the number of those in the morning (evening).

Table 6.1 contains all these parameters of the probabilistic model that have to be estimated or that are held constant based on the backscatter model and empirical data.

We set the parameters based on Maximum Likelihood learning; this approach is often taken because it provides a convenient framework that fits in nicely with graphical models. Besides, it is comparatively easy to compute the estimates; in the presence of latent variables, the EM Algorithm provides an iterative scheme for maximizing the marginal likelihood, which would be hard to optimize or even evaluate otherwise.

Being an iterative approach, the EM algorithm needs initial values of the parameters. In addition, several parameters will be constrained based on empirical observations and theoretical considerations.

Initial values and constraints

The probability of the initial state X_1, $\boldsymbol{\theta_1}$, is a parameter of the distribution that is specified beforehand. Similarly, the transition probabilities, as described by the stochastic matrix $\mathbf{A}$, are held constant. As they are expected to be roughly the same everywhere, only one set of values is used; these can be found in the appendix.

In contrast to the transition probabilities, the parameters of the emission probabilities $P(y_t|x_t)$ depend strongly on the spatial location of the grid point considered. Consequently, they have to be estimated for each time series.

Extensive data analysis has revealed that the means of the emission probability densities for the thawing parameter 2 cannot be estimated reliably because they occur so rarely. They are thus set empirically to the values shown in table 6.1. Similarly, the variances are chosen to be a multiple of those of state 0 and the means for states 2 and 3 are expressed as functions of those of states 0 and 1. The means of the diurnal differences are set to zero for all but the thawing state. The reasoning behind these values will be explained in chapter 7, based on the backscatter model as well as real data.

The initial values for the means of the SeaWinds and ASCAT σ^0, these are extracted from the medians of the measurements in January and July – due to the restrictions described above, initial values are only needed for states 0 and 1.

EM algorithm

The EM algorithm consists of two steps – the E-Step and the M-Step –, that are applied repeatedly in an alternating fashion. In order to avoid overfitting (and because it works well in practice) the EM algorithm is only performed once. This is similar to the regularization induced by limiting the number of steps in the backpropagation algorithm when estimating the weights in feedforward neural networks [67]. In our case, the likelihood is a very complicated function exhibiting a possibly large number of local maxima. If we did not fix the transition probabilities, there would be an inherent symmetry: the state 'frozen', for example, could map equally well to 1 instead of 0. Such issues are often encountered

in unsupervised learning [4]. In contrast to typical machine learning situations, we employ models to fix and constrain certain parameters; the nature of these constraints was described above.

In the M-Step, the energy term $\langle \log P(x, y|\theta)\rangle_{Q(x)}$ has to be maximized. Substituting (6.2) into this expression, we notice that we only need to know the marginals $Q(x_t)$ for estimating the parameters of the emission probabilities. These can be found by the Forward-Backward algorithm (this is the E-Step).

For constant variances of the emission probabilities, the M-Step for the Gaussian HMM can be found in e.g. [4]. Due to our parameterization, the derivation is a bit more complicated. The resulting equations are summarized in algorithm 5 in the appendix.

6.1.3. Inference

Inference is the task of deriving probabilistic information from the graphical model. In terms of the Hidden Markov Model, two quantities are important:

Marginal probability $P(x_t|y_{1:T})$ can be determined with the Forward/Backward Algorithm – algorithms 1 and 2, equation (4.7).

Most probable time-series $X^\star_{1:T}$ is easily computable by the Viterbi Algorithm – algorithm 3.

The state space $S_4 = \{0, 1, 2, 3\}$ is not very convenient to deal with, particularly because not only frozen landscapes without snow cover but also thawing landscapes are often classified as state 3. Thus we map S_4 to $S_3 = \{0, 1, 2\}$: state 3 is converted to 2 from February to July and to 0 otherwise. When dealing with marginal distributions $P(x_t)$, the probabilities simply have to be added (as they correspond to disjoint events). This representation is more pleasing intuitively and facilitates visualization and comparison with other data sources.

These results can be summarized in numerous ways; an example is the extraction of key dates, such as the date of freeze-up. As will be shown in section 7.2, such very simple products lend themselves to visualization and human interpretation.

6.2. Backscatter model

The previous discussions have shown that the dynamic evolution of snow as well as the freezing and thawing of soils manifest themselves in σ_0 in a way which is strongly dependent on the observation frequency.

Having primarily the tundra biome in mind, we focus on the contributions by the ground and snow, which can possibly be layered. The model we adopt is very similar to a family of approaches, which have been widely used [1][72]; the differences mainly affect smaller details. The total backscatter is modelled as the incoherent sum of the following terms:

$$\sigma^0 = \kappa \sigma^0_{soil} + \sigma^0_{snow} \tag{6.3}$$

where κ describes the attenuation of the soil component due to the overlying snow, σ^0_{soil} the backscatter of the soil and σ^0_{snow} the total backscatter of the snow. The last term contains the contribution due to volume scattering and any interface, i.e. the surface of the snow.

6.2.1. Soil

The discussion in section 3.2.2 has shown that σ^0 is primarily dependent on two factors: the dielectric constant and the texture.

Variations of ε_r of a given soil are mainly caused by changes in soil moisture. Commonly, mixing models are employed to describe this dependence. Even though quite sophisticated models which distinguish between bound and free water – which, for a given soil water content depends on the composition of the soil, e.g. the amount of clay – have been developed, we adopt the very simple model by Birchak [12]. This model can also be used to predict ε_r of the soil when it is frozen.

The second factor is the surface roughness. In order to keep the model and the parametrization simple, we adopt the empirical model by Oh et al. [77]: the backscatter (for a fixed polarization and incidence angle) is only a function of the RMS height s and the bulk dielectric constant ε_r. The former encompasses equations for co-polarized (horizontal and vertical) and cross-polarized backscatter:

$$q = \frac{\sigma_{hv}^0}{\sigma_{vv}^0} = 0.23\Gamma_{TE}(0)\left(1 - e^{-ks}\right) \tag{6.4}$$

$$\sqrt{p(\theta_i)} = \sqrt{\frac{\sigma_{hh}^0}{\sigma_{vv}^0}} = 1 - \left(\frac{2\theta_i}{\pi}\right)^{\left(3\Gamma_{TE}^2(0)\right)^{-1}} e^{-ks} \tag{6.5}$$

$$\sigma_{vv}^0(\theta_i) = 0.7\left(1 - e^{-0.65(ks)^{1.8}}\right)\frac{cos^3(\theta_i)}{\sqrt{p}}\left(\Gamma_{TE}^2(\theta_i) + \Gamma_{TM}^2(\theta_i)\right) \tag{6.6}$$

where the reflectivities $\Gamma(\theta_i)$ are given by (3.14) and (3.16) and ks is the product of the norm of the wave vector and the RMS height. According to the authors, the model shows good agreement with empirical data for $0.1 \leq kh \leq 6$ and $2.5 \leq kl \leq 20$.

6.2.2. Snow

The most widespread simple model for describing the backscatter of snow has already been introduced in section 3.2.2. The formula reads

$$\sigma_{snow}^0 = \frac{\cos(\theta)\gamma_2}{2\gamma_1}\left(1 - e^{-2\gamma_1 d/\cos(\theta)}\right) + \sigma_{ground}^0 e^{-2\gamma_1 d/\cos(\theta)} + \sigma_{snow\_surface}^0 \tag{6.7}$$

The coefficients γ_1 and γ_2 have to be related to the physical properties of the snow pack. By assuming that the medium is not too dense and that due to the long wavelength, the individual snow grains can be considered as Rayleigh scatterers, α_f can be obtained by substituting (3.22) into (3.19).

$$\alpha_f = N\frac{128\,\pi^5}{3c^4}r^6 f^4\left(\frac{\varepsilon_r - 1}{\varepsilon_r + 2}\right)^2 \tag{6.8}$$

$$\tag{6.9}$$

.

The absorption coefficient can be obtained by analogy with a homogeneous medium: it is chosen in such a way that the absorption corresponds to the reduction in power encountered by a planar wave in a medium with complex dielectric constant ε_r [1]:

$$\gamma_f = 2|\mathbf{k}|\Im(\sqrt{\epsilon_r}) \tag{6.10}$$

In order to apply the simplified treatment of radiative transfer, we choose the δ two-stream approximation according to Meador & Weaver [70]. For the phase function of Rayleigh scatterers, the coefficients can be shown to be:

$$\gamma_1 = \beta_f - \frac{\alpha_f}{4} \tag{6.11}$$

$$\gamma_2 = \frac{3}{4}\alpha_f \tag{6.12}$$

In this approximation, the coefficient κ of (6.3) corresponds to $e^{-2\gamma_1 d/\cos(\theta)}$; the surface scattering can be incorporated by setting

$$\kappa = T_\uparrow^2 T_\downarrow^2 \, e^{-2\gamma_1 d/\cos(\theta)} \tag{6.13}$$

where $T_\uparrow^2$ and $T_\downarrow^2$ are the appropriate Fresnel power transmission coefficient as defined in section 3.1.1.

The dielectric constant and thus the absorption coefficient is strongly dependent on the liquid water content. We use the dielectric mixing formula for snow proposed by Maetzler et al. [68].

Besides the radiation scattered due to the heterogeneous nature of a layer of snow, such a layer also acts as a bulk object; this leads to surface scattering. The dielectric contrast is very low for dry snow, as is thus the surface scattering term. As with the soil contribution, we employ the empirical model by Oh to describe the scattering off the rough surface, even though this model was derived on the basis of measurements of bare soil.

Multiple layers

This simple model, where multiple scattering effects are neglected, can be extended to several planar scattering layers in a straight forward way.

The easiest way to generalize this model is to implement it recursively: the term $\sigma^0_{soil\_surface}$ can then be thought of as the total contribution of the snow and soil beneath the given layer.

6.2.3. Discussion

The model presented above is semi-empirical in nature; many similar approaches have been proposed. There are three important points where many of these models differ:

1. Computation of the transmission κ

2. Determination of the incidence angle beneath a scattering layer

3. Choice of the dielectric constant in the rough surface scattering model

We consider the surface scattering contribution in the computation of κ by the Fresnel power transmission coefficient. This approach, which has also been taken by Shi & Dozier [91], suffers from the fact that it is not consistent with the rough surface scattering model, thus violating the principle of energy conservation. For the situations commonly encountered when modelling snow, the choice of κ is not very critical. In case of dry snow, the bulk dielectric constant and thus the surface scattering is very small. The absorption is extremely high and the contribution of the soil negligible if, on the other hand, the snow is wet.

The second point concerns the incidence angle on the soil, which we consider to be equal to the incidence angle at the snow surface. An alternative would be to employ Snell's law; this approach is for example taken in [91]. For dry snow, the difference is negligible. For wet snow, on the other hand, the backscattering from the soil is small due to the absorption.

A tacit assumption of the rough surface scattering model is that the relative dielectric constant of the upper medium is 1, corresponding to air or vacuum. For loose, dry snow, this approximation is permissible; it breaks down with increasing liquid water content. Shi and Dozier [91] replace the permittivity of the soil by its dielectric contrast $\varepsilon_{new} = \frac{\varepsilon_{soil}}{\varepsilon_{snow}}$. There is, however, no physical justification for this assumption; the Fresnel equations call for replacing the index of refraction: $\sqrt{\varepsilon_{new}} = \sqrt{\frac{\varepsilon_{soil}}{\varepsilon_{snow}}}$. This is the approach taken in our model. Note that – similar to the varying incidence angle – the contribution of the soil to σ^0 becomes negligible quickly with increasing liquid water content.

6.2.4. Vegetation

The main purpose of the model is to provide a better understanding of the influence of the snow height and its composition. It is not intended to estimate parameters from measurements. This is the reason why vegetation is not considered; in the following, we will briefly discuss the expected consequences.

The cloud model is the simplest backscatter model of vegetation. As it is of the same form as our snow model, it could be easily included as an additional layer. The parameters, however, are usually estimated from data because they do not have a clear physical interpretation. At *C*-band, the dynamic effects of the vegetation cover at high latitudes are presumably rather small. At K_u-band, the scattering from canopy elements is stronger. Nghiem and Tsai [75], however, report that even in the taiga, the vegetation cover is often so sparse that σ^0 at K_u-band is dominated by the effects of snow and liquid water.

Table 6.1.: Parameters used in the probabilistic model.

Symbol	Description	Estimation	Initial value
$\boldsymbol{\theta_1}$	Probability mass vector of X_1	–	$(0.01, 0.97, 0.01, 0.01)^2$
A	Transition probability matrix	–	see Appendix
$\mu_{swm,0}$	Mean of Y_t^{swm} for the frozen state 0	EM Algorithm	median of window in winter
$\mu_{swm,1}$	Mean of Y_t^{swm} for the thawed state 1	EM Algorithm	median of window in summer
$\mu_{swm,2}$	Mean of Y_t^{swm} for the thawing state 2	set to $\mu_{swm,1} - 2dB$	$\mu_{swm,1} - 2dB$
$\mu_{swm,3}$	Mean of Y_t^{swm} for state 3	set to $\mu_{swm,1}$	$\mu_{swm,1}$
$\mu_{am,0}$	Mean of Y_t^{am} for the frozen state 0	EM Algorithm	median of window in winter
$\mu_{am,1}$	Mean of Y_t^{am} for the thawed state 1	EM Algorithm	median of window in summer
$\mu_{am,2}$	Mean of Y_t^{am} for the thawing state 2	set to $\mu_{am,1} - 2dB$	$\mu_{am,1} - 2dB$
$\mu_{am,3}$	Mean of Y_t^{am} for state 3	set to $\mu_{am,0}$	$\mu_{am,0}$
$\mu_{swd,0}$	Mean of Y_t^{swd} for the frozen state 0	set to 0	0
$\mu_{swd,1}$	Mean of Y_t^{swd} for the thawed state 1	set to 0	0
$\mu_{swd,2}$	Mean of Y_t^{swd} for the thawing state 2	set to 3dB	3dB
$\mu_{swd,3}$	Mean of Y_t^{swd} for state 3	set to 0	0
σ^2_{swm}	Variance of Y_t^{swm} for states 0-3	EM algorithm	0.5
σ^2_{am}	Variance of Y_t^{am} for states 0-3	EM algorithm	0.2

6. Implementation

7. Results

This chapter focusses on the results obtained with the previously discussed algorithm. The main characteristics of the time-series are analyzed by means of the backscatter model. Subsequently, several examples of how the probabilistic information can be exploited to derive higher level products, such as maps, are given.

The results described in this chapter will be discussed in more detail in the next chapter, where the emphasis is put on validating the outcomes of the sensor fusion algorithm.

7.1. Freeze/thaw time series

According to the analysis of chapters 2 and 3, there are several key factors that have a significant impact on the dynamics of σ^0 and this influence possibly strongly depends on the radar frequency:

- Frozen and thawed soil
- Snow depth
- Snow metamorphosis (varying grain sizes)
- Liquid water content of snow
- Vegetation

The relative impact of these influences depends on the biome and land cover; the results of the sensor fusion model applied to real data and the predictions of the backscatter model can elucidate the connection.

7.1.1. Representative time series

Tundra wetlands

A summary of the key results for a grid point in the tundra ($72\,N, 121\,E$) is shown in figure 7.1a. As there is hardly any difference between the time series of the most probable state X_t^m and the results of the Viterbi algorithm $X_t^\star$, only the former is displayed.

The freeze-up in late September is sudden and very evident at C band, less so at K_u band. The corresponding transition in the estimated state, as described by $P(x_t)$, is thus also clearly defined.

The slow growth of *swm* suggests a gradual build-up of snow and accompanying structural changes, which is not noticeable at C band. A melting event in May leads to a precipitous fall of σ^0 at both bands and simultaneously to a significant diurnal difference (*swd*). As not all the snow melts, the remaining snow undergoes metamorphosis: an increase in grain size that results in very large backscatter, even at C band. Subsequently, there is another melting period.

Larch taiga

A typical time series $(63\,N, 127\,E)$ for the taiga biome can be found in figure 7.1b.

The transition from frozen to non-frozen conditions is not as sudden and pronounced as in figure 7.1a. The slow accumulation of snow and its metamorphosis is particularly evident at K_u band. The melting sets in at the end of April; it leads to the expected changes in σ^0 and the diurnal difference *swd*. Overall, these transitions are well reflected in X_t^m.

A comparison with the tundra time series, where vegetation presumably plays a comparatively minor role, reveals that the influence of the canopy on σ^0 is secondary to the one caused by snow and the freeze/thaw dynamics of the soil. This could partly be explained by the fact that the larch is a deciduous tree.

Agricultural fields

The data shown in figure 7.2a are the measurements of agricultural fields close to Kasan $(57\,N, 48\,E)$.

During non-frozen condition, the backscatter at C band (and to a much smaller extent at K_u band) exhibits an interesting behaviour: a sudden increase is followed by a slow decay. Although the validation with external data is only treated in chapter 8, it should be mentioned that, as revealed by ERA Interim data, this is caused by rain fall events, or rather the corresponding evolution of the soil moisture.

Compared to the other time series, the backscatter in winter is not dominated by snow. Its influence, however, is particularly striking in spring: there are complicated melt/refreeze patterns. This is noticeable in both bands as well as in the diurnal difference. The results of the analysis reflect these intricate dynamics.

Steppe

Figure 7.2b contains the results of a representative area $(51\,N, 119\,E)$ of grasslands typical for the steppe biome.

The behaviour in summer is similar to the one seen in figure 7.2a; in particular the sudden increases in soil moisture and thus σ^0. Unlike the other time series discussed here, this one only gradually falls off to its average value in winter, both in K_u and C band. There is a rather small increase at K_u band, which suggests an accumulation in snow and continuous metamorphosis.

As opposed to the other time series, this one lacks a clear melting or thawing signature. Bearing the latitude in mind, it is obvious that due to the aforementioned reasons the probabilistic analysis is not working very well. In fact, such difficulties are also encountered in the test area. A more in-depth in analysis (in particular, comparisons with external data) can be found in chapter 8.
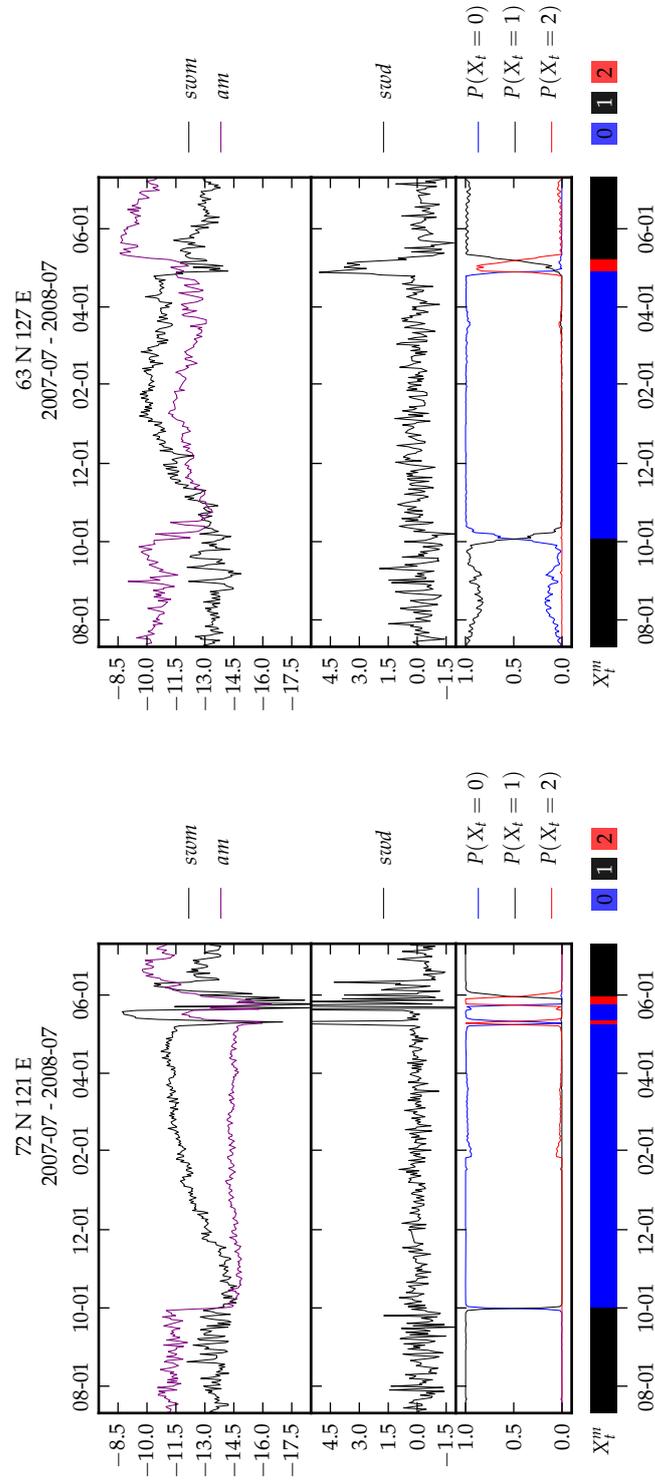
(a) A grid point in the tundra.

(b) A grid point in the taiga.

Figure 7.1.: Results for two grid points in the tundra and taiga.
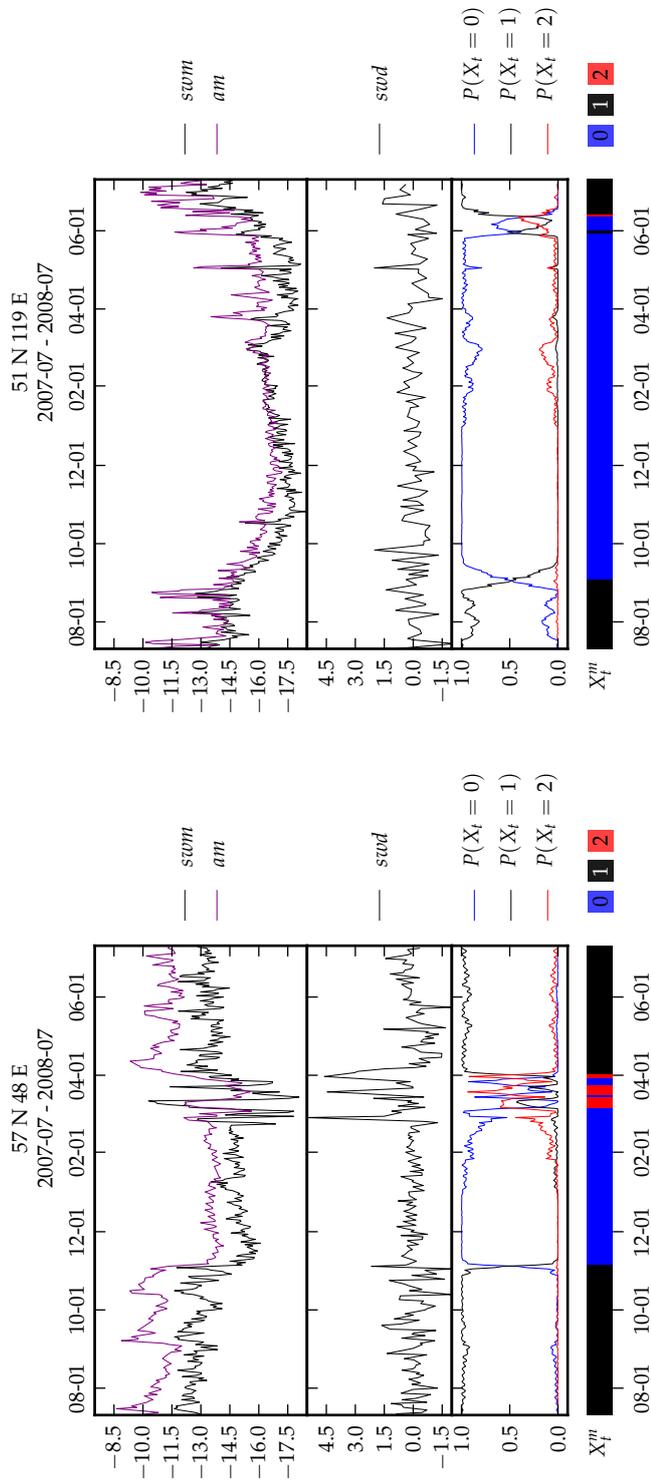
(a) A grid point in agriculturally used land.

(b) A grid point in the steppe.

Figure 7.2.: Results for two grid points in the steppe and in fields.

7.1.2. Model predictions

Snow depth

In light of the theoretical discussion in chapter 3 and the previous examples, it is evident that the depth of the snow has a significant impact on σ^0, particularly at K_u band.

The dependence on the frequency is depicted in figure 7.3, which is actually the same plot as the one shown in chapter 3. The following parameter settings have been used:

Snow structure Following [15], the number density $N = 1.3 \cdot 10^8 \, m^{-3}$ and the grain radius $r = 4.5 \cdot 10^{-4} \, m$. The RMS height of the surface is $1 \, cm$.

Snow permittivity The liquid water content is assumed to be 0.

Soil $\varepsilon_r = 2.3$ for dry soil, which is a reasonable value [12]. The frozen ground has a volumetric ice content of 0.2 and its RMS height is taken to be $5 \, cm$.
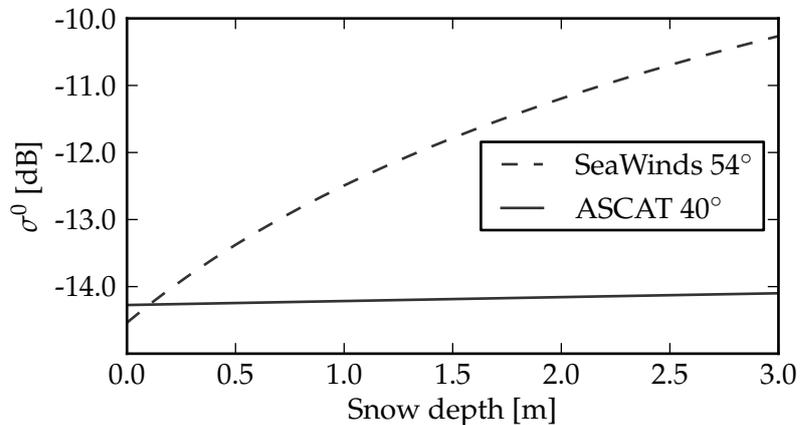
Figure 7.3.: Simulation of σ^0 of dry snow for the frequencies and incidence angles of QuikScat and ASCAT

Snow metamorphism

Snow metamorphism is the process of compaction of a snow pack or changes in the grain size distribution with time; a discussion of the phenomenon and its causes can be found in section 2.1.2.

Due to the pronounced influence of the size of an object on its scattering cross-section in the Rayleigh regime, snow metamorphosis manifests itself in σ^0. Figure 7.4 depicts simulation results of the backscatter model, where all other parameters are set to the same values as before.

The refreezing of a melting snow pack results in characteristically large snow grains; this is particularly evident in the time series in figure 7.1.
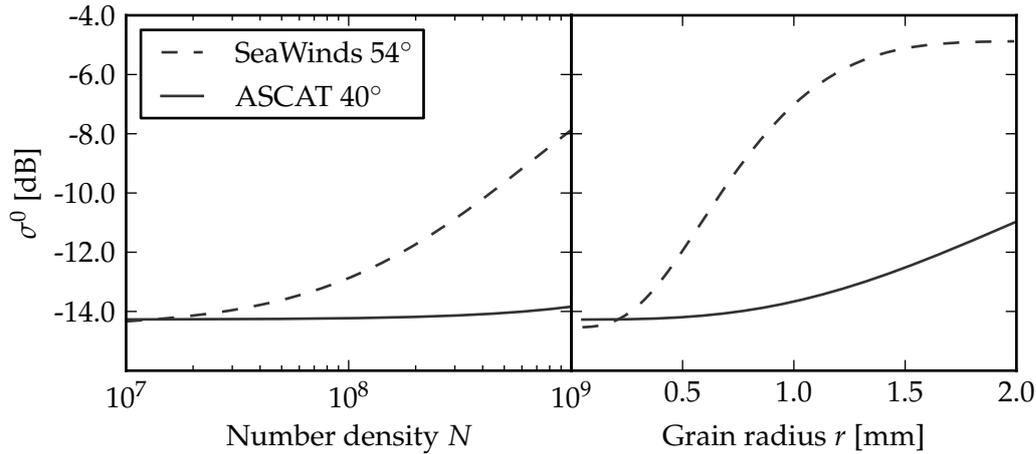
Figure 7.4.: Simulation of σ^0 of a snow pack of $1\,m$ height: in the left subplot, the number density varies with the grain size being constant; the grain size is changed at constant mass density in the right subplot. The other parameters are the same as in figure 7.3.

Snow melt

There are two striking features dominating the time series during snow melt events: diurnal cycling and a precipitous drop in σ^0. The former is caused by the fact that the backscatter rises during the night as the liquid water refreezes and the snow undergoes metamorphism.

The latter is due to the steep increase in the absorption; ε''_r of water is much bigger than that of ice. The following simulation can make these assertions more quantitative: a two-layered snow pack, whose top layer has a varying liquid water content. The lower layer is half a metre in height, the other properties of this layer and the soil are the same as before. The parameters of the wet snow layer: $h = 0.1\,m$, $N = 7 \cdot 10^7\,m^{-3}$, $r = 9 \cdot 10^{-4}\,m$.

Note that the backscatter is increasingly dominated by the contribution of the snow surface; it is thus very dependent on the roughness.

Frozen and thawed soil

The freeze-up is particularly evident in the time series in figures 7.1a and 7.2a: there is a precipitous drop of σ^0 at C band and, to a lesser extent, at K_u band. Figure 7.6 illustrates the dependence of σ^0 on the volumetric ice/water content, the parameters of the soil being the same as previously. For simplicity, only completely frozen or non-frozen conditions are considered and all the liquid water is treated as unbound (as always in the Birchak model for ε_r).

7.2. Secondary products

The output of the probabilistic sensor fusion algorithm can be summarized to facilitate analysis and visualization. The following three examples can only offer a glimpse at the
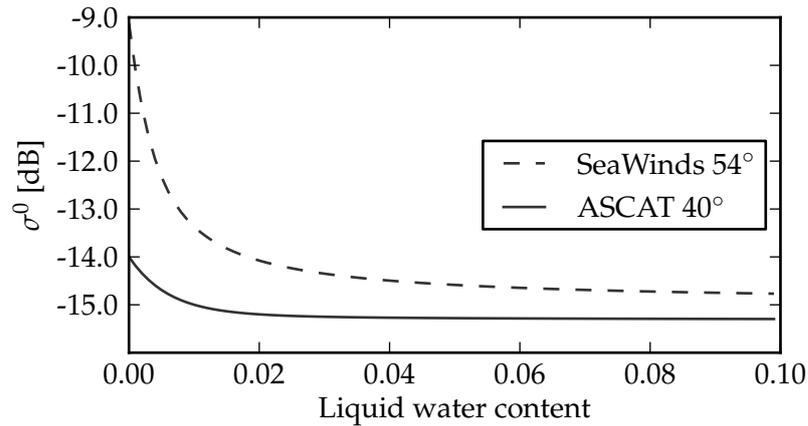
Figure 7.5.: Simulation of σ^0 of a layered snow pack; the liquid water content of the topmost layer varies.
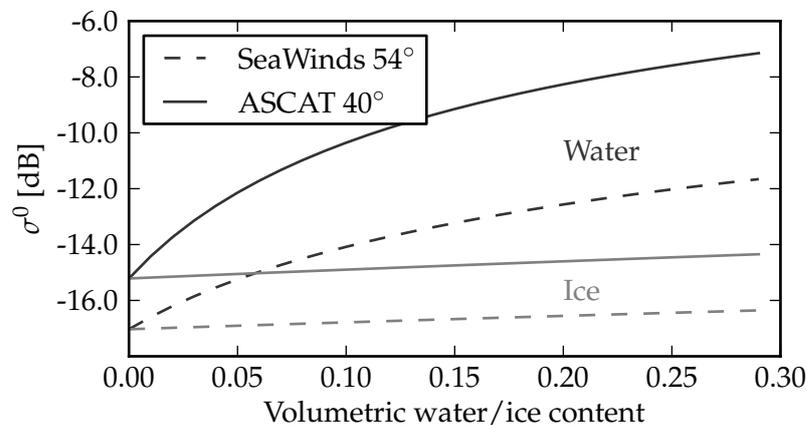
Figure 7.6.: Simulation of σ^0 of a bare soil as a function of frozen/liquid water content.

wealth of information contained in the time series.

Time of freeze and length of frozen season

The day of freeze-up (figure 7.7a) has been derived from the marginal probabilities by a very simple method: the first day in autumn when 0 (frozen) is the most probable state. The number of days between the freeze-up and the beginning of thaw can give insight into the limitations imposed on many biophysical processes, e.g. metabolic activity; it is depicted in figure 7.7b. The map clearly exhibits a north-south trend. Furthermore, there is a striking correlation with topography – in particular the Verkhoyansk Range and Stanovoy Range, cf. figure 2.2a.

There are two outliers in the Verkhoyansk Range and more importantly severe problems south of 53°N. These will be discussed in more detail in chapter 8.

Probability of thaw

The state 2, which corresponds to thawing, is characterized by wet snow during the day and consequently very low overall backscatter and a significant diurnal difference. Figure 7.7c shows the marginal probability of thawing on 2008-05-16, once again emphasizing the effects of altitude and latitude.

(c) Probability of thawing on 2008-05-16.
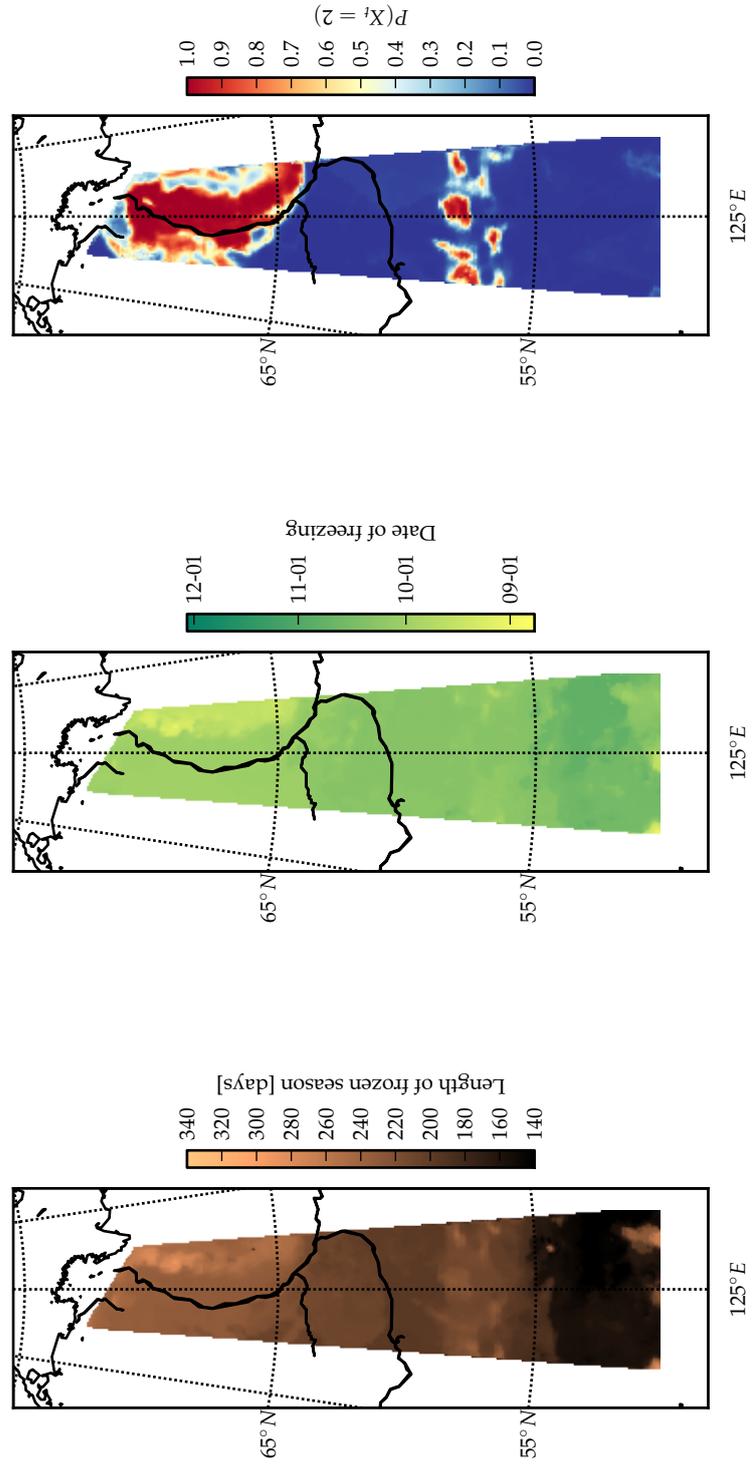
(b) Day of freezing.

(a) Length of the frozen season.

Figure 7.7.: Three exemplary maps derived from the probability information.

8. Validation & Discussion

A discussion of the assumptions and limitations inherent in the probabilistic model forms the first part of this chapter. The validation of any remote sensing product is of paramount importance for determining its usefulness: the second section contains a comparison with in-situ observations and global data products. The third part of this chapter consists of an in-depth analysis of situations when the approach does not work satisfactorily.

8.1. Discussion of the assumptions

8.1.1. States

The notion of having a small number of discrete states that encode all information about the physical reality is a serious limitation. Apart from the scale issue inherent to most remote sensing practices, it would also represent a severe simplification for a perfectly homogeneous terrain.

The theoretical considerations and the analysis of the data have revealed the following points:

Snow height/Snow Water Equivalent influence the backscatter gradually, whereas the assumption of having a frozen state 0 calls for a step-like dependence (which would be a better assumption in the visible spectrum).

Snow metamorphism can lead to changes in σ^0; it thus stands to reason that such information should be included in the state variables.

Snow melting is the most important indicator for the thawing state 2. The liquid water content of the snow, however, is not explicitly represented.

Soil freezing is assumed to be a binary decision: frozen or non-frozen. According to the discussion in section 2.1.1, the larger pores freeze over before the smaller ones. In addition, the depth profile has to be considered: when the active layer in permafrost regions begins to freeze, a thawed layer of non-frozen soil can be sandwiched between two frozen ones. The data, however, often show a distinct, precipitous drop in σ^0, indicating that with respect to microwave remote sensing, these limitations are not that severe. Note that similar observations hold with regards to the freezing vegetation.

Lake ice (although barely considered in this thesis) is not a binary decision either, but depends on many factors; cf. section 3.2.2 for a short discussion regarding its microwave backscatter characteristics.

Inundation is not explicitly considered either. In the study area, it often occurs in conjunction with the snow melt due to the limited infiltration capacity of the soil.

In light of these simplifying assumptions, the choice of the number of states appears rather pragmatic; it is in particular dependent on the sensors/features available. This is

also the reason for adding the state $3 \in S_4$, which enables the model to distinguish frozen conditions with no or very little snow from conditions with significant snow cover. This distinction would barely be possible without the simultaneous availability of K_u and C band observations.

The advantage of adopting a discrete state is chiefly due to the fact that it allows a rather robust inference (few possibilities) and, more importantly, the use of a simple model for its evolution.

8.1.2. Conditional independence assumptions

There are several conditional independence relations inherent in the probabilistic model defined generally in section 4 and specifically in section 6.

Firstly, the time series of the states is assumed to obey the Markov condition: given the state at time t, states $s < t$ and $u > t$ are independent. This is presumably not the biggest limitation, even though the continuous variables that are not explicitly considered could reveal more information. As an example, consider t to be the onset of snow melt. The knowledge that there was snow cover at $s = t - 1$ would barely influence the prediction at $u = t + 7$, but information about the snow height at s probably would. If this variable was part of the state, however, it would also be available at t and the Markov assumption would still be reasonable.

Secondly, the visible variables Y at different epochs are deemed to be independent given the state. Figure 7.1, for example, suggests that this is not the case. In this example, it is the soil moisture which makes the backscatter at subsequent days dependent (even if the states are known). Similar correlations are also induced by the snow height and grain size. The obvious way to mitigate this issue would be to include such variables in the latent state.

Thirdly, the features Y_t^{swm}, Y_t^{swd}, Y_t^{am} are assumed to be independent given X_t. This approximation can break down for essentially the same reasons: snow metamorphism, for example, can greatly increase the backscatter at both K_u and C band.

8.1.3. Parametrization

The decision which parameters to estimate freely and which to constrain was made based on an analysis of the data and preliminary results. The rather small number of free parameters is also due to the fact that there is only a rather small overlap of the times of operation of the two sensors. The values of the constrained parameters are based on extensive data analysis and modelling considerations. Similarly, the emission probability distributions are assumed to be Gaussians mainly for simplicity. More extensive sensitivity studies would be needed to determine the validity and limits of these assumptions, i.e. in which conditions they can give reasonable results.

The initial values of the parameters were determined similarly. As the expected values of Y_t^{swm} and Y_t^{am} vary significantly, the corresponding parameters have to be estimated for each grid point. The way this is done (taking the median of a representative period) is quite arbitrary and could certainly be improved upon. It appears to work quite well in general (and in the high-latitude regions in particular) but can break down when the time span that forms the basis of the estimation is not representative. This could, for example, be due to inundation in summer or a snow melt in winter.

As the model is probabilistic in nature, choosing a method for estimating the parameters that is based on probability theory only seems natural. The (marginal) likelihood has the key advantage of being comparatively easy to work with and maximum likelihood does not need any additional assumptions either (as a fully Bayesian approach would). In addition, the model and the parametrization are sufficiently complex to render an ad-hoc, empirical approach impossible.

The step-wise nature of the EM Algorithm for maximizing the marginal likelihood may seem unattractive at first (although there are no known closed-form solutions), but it offers the ability of regularization by only performing a limited number of steps. The modelling decisions are included in the likelihood via the constraints and the initial values. In fact, the 'meaning' of the states is nowhere made explicit. The EM Algorithm, by only considering the likelihood, has a tendency to run off and choose parameters which are not consistent with the physical interpretation. In practice, the decision to only make one iteration works well and in general a second step does not lead to a significant improvement. More detailed studies about the estimation process could offer additional insight.

8.2. Comparison with external data sources

8.2.1. In-situ observations

In-situ observations provide the opportunity to validate the results of remote sensing. They are, however, sparse and they refer to a single point. It should be borne in mind that remote sensing in general and scatterometry in particular pertain to much larger areas. This is particularly true for the snow cover, which is one of the two in-situ parameters available; the other is the air temperature.

In general, the transition from frozen to non-frozen conditions and vice versa is marked by a period when the maximum temperature is above and the minimum temperature below $0°C$. The dependence of the Freeze/Thaw state on air temperature, although quite obvious at first sight, is not that straightforward. Due to the reasons discussed above and in chapter 2, it should only be regarded as a proxy variable. A description of the data set can be found in section 5.5.2.

As described in section 5.5.1, the observations of the snow cover were carried out in steps of either two weeks or four weeks; they begin soon after the first snow fall and stop before the snow cover disappears. The time of snow fall and melt can thus only be guessed at, as it is not explicitly recorded. The data set of meteorological observations (section 5.5.2) contains records of the snow heights at some of the station[1]; however, the storage of these parameters was discontinued in December 2007. Figure 8.1 shows the distribution of the observation stations.

In light of these reasons and the scarcity of the ground stations, the following analysis will focus on the qualitative aspects.

Freeze-up can be detected quite well in the high latitudes (cf. figure 8.4b); this is mainly due to the fact that the transition period is very short. The accumulation of snow is rather

---

[1]Although the snow height is given in centimeters according to the metadata, the values appear unrealistic, cf. figure 8.4
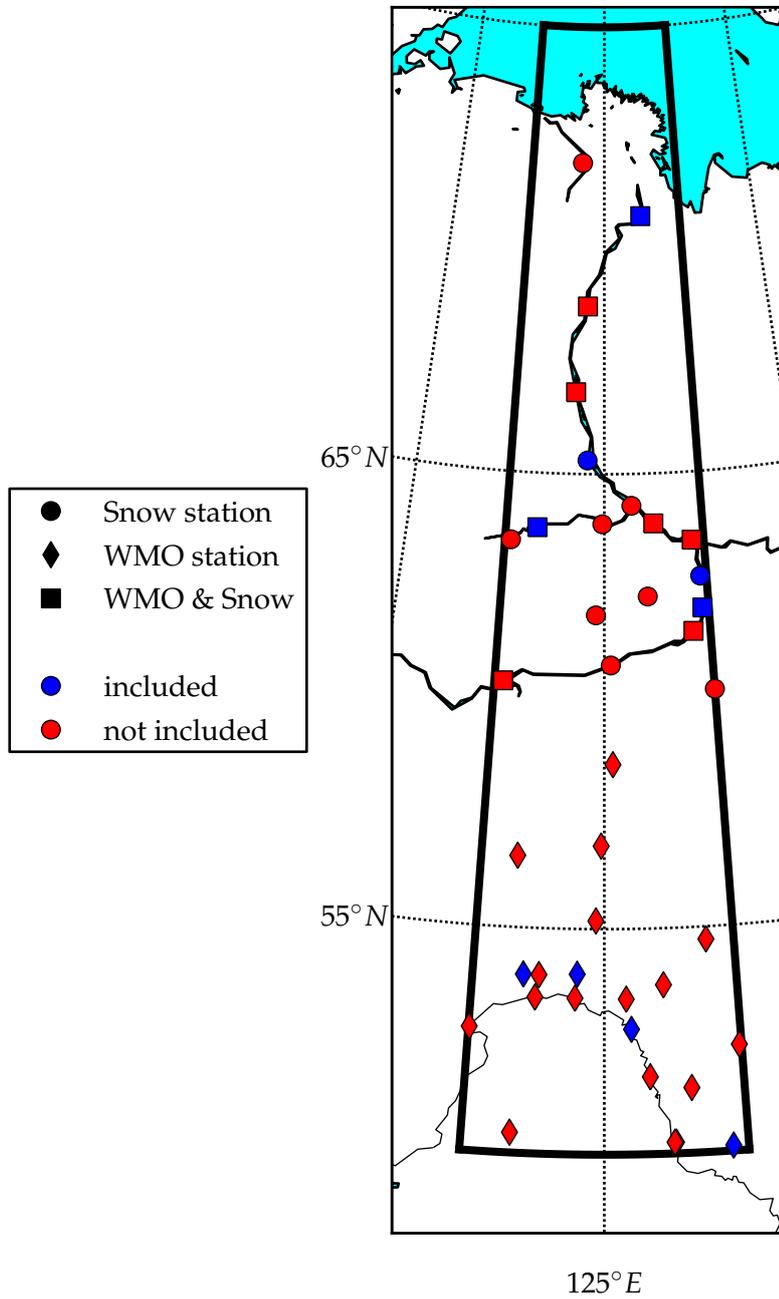
Figure 8.1.: Overview of the distribution of the observation stations. The blue symbols ('included') represent stations that are analyzed in detail.

slow, which is made evident when comparing the K_u and the C band backscatter. Towards the south, e.g. figure 8.4a, the transition period is longer but the drop in C band is often very pronounced. The detected freeze date corresponds well with the first (significant) snow fall in figures 8.2a and 8.2b.

The point at $50°N$ in figure 8.3a exhibits a very gradual change from non-frozen to frozen conditions; this can be seen in both the remote sensing and the in situ meteorological data. The description of the freeze/thaw state with a discrete variable thus appears questionable.

The time series of figure 8.3b reveals a temporary snow cover of a few centimetres during thawed soil conditions. It is not evident in the backscatter data and thus not reflected in the results.

Midwinter conditions   are stable in the study area, which is reflected in the temperature records. The backscatter can change due to gradual snow metamorphism, which will subsequently be analyzed in more detail.

A striking feature of the marginal probabilities is the kink in $P(X_t = 0)$ in February, e.g. in figure 8.3b. This is caused by the mapping from S_4 to S_3. The data show that the current implementation leads to a rather unnatural evolution of $P(x_t)$ and should thus be improved upon.

Thawing   is characterized by melting snow: very low backscatter and positive diurnal differences. The former, however, is also indicative of inundation. Figures 8.3a and 8.4b show time series with pronounced thawing periods, whereas figures 8.2a, 8.2b and 8.4a exhibit rapid transitions. These figures show that the period of thawing can be underestimated at lower latitudes. In figure 8.6b, for example, the transition of $P(x_t)$ happens when the backscatter at K_u band reaches its summer level, even though the one at C band is still below it; this is corroborated by the snow measurements.

The period when the temperature drops below $0°C$ during the night but rises above it during the day, can be rather long; particularly at lower latitudes. The remote sensing signal tends to make a quicker transition, particularly when there is comparatively little snow.

Metamorphism   can result in greatly elevated σ^0; it is thus possible that the C band measurements barely differ in summer and winter. The influence of gradual snow metamorphism is evident in figures 8.4a and 8.4b as opposed to figures 7.1a and 8.2b – one of each corresponding to the tundra and taiga biome.

In figures 8.2a and 8.2b a single warming event in mid-November, which led to metamorphism but – judging from the σ^0 observations – not to melting, can be observed. It increased the backscatter in both bands. Figure 8.4a reveals a similar event in early spring.

(a) WMO station 30683
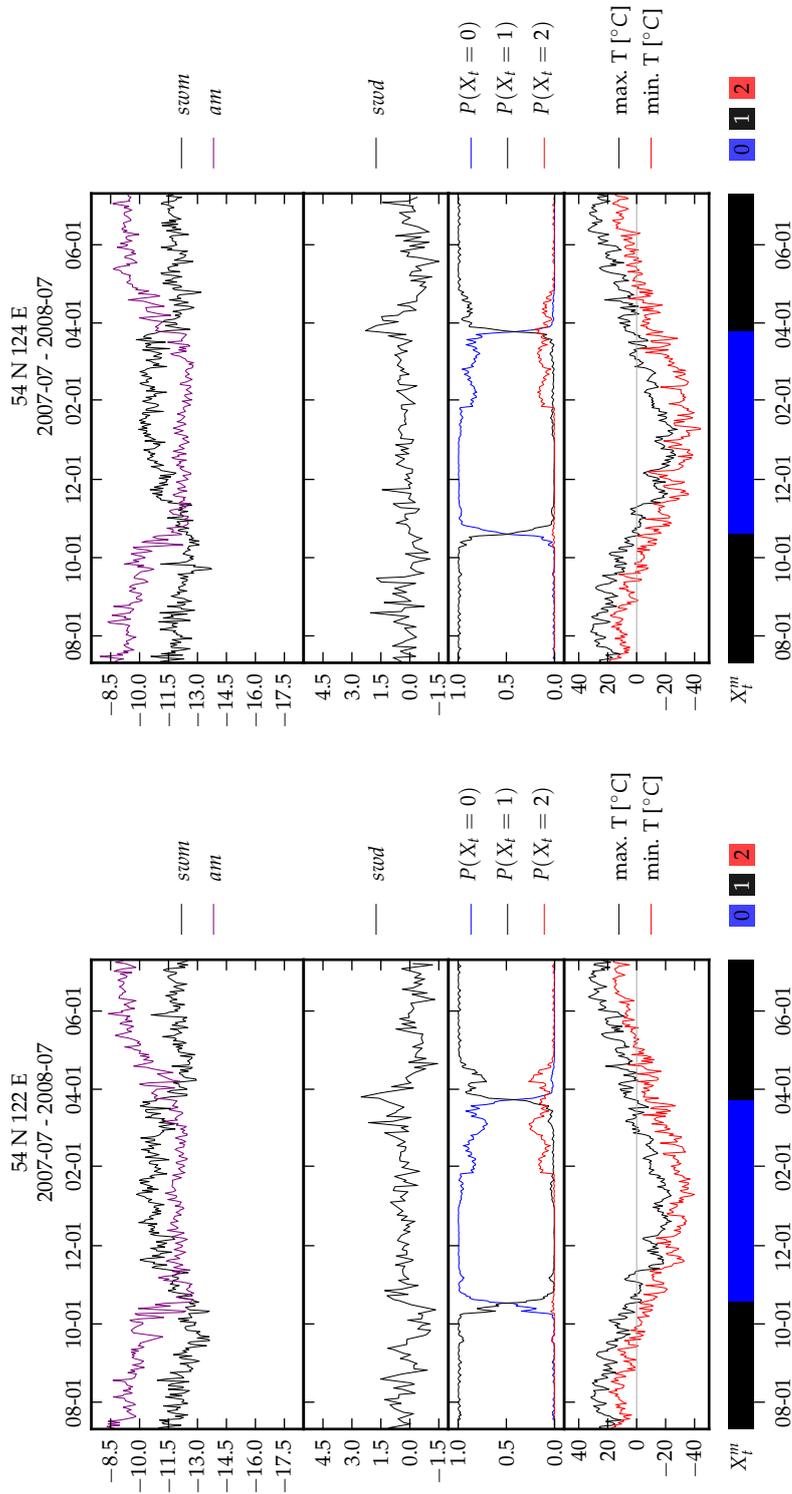
(b) WMO station 30692

Figure 8.2.: Results for two different WMO observation stations.
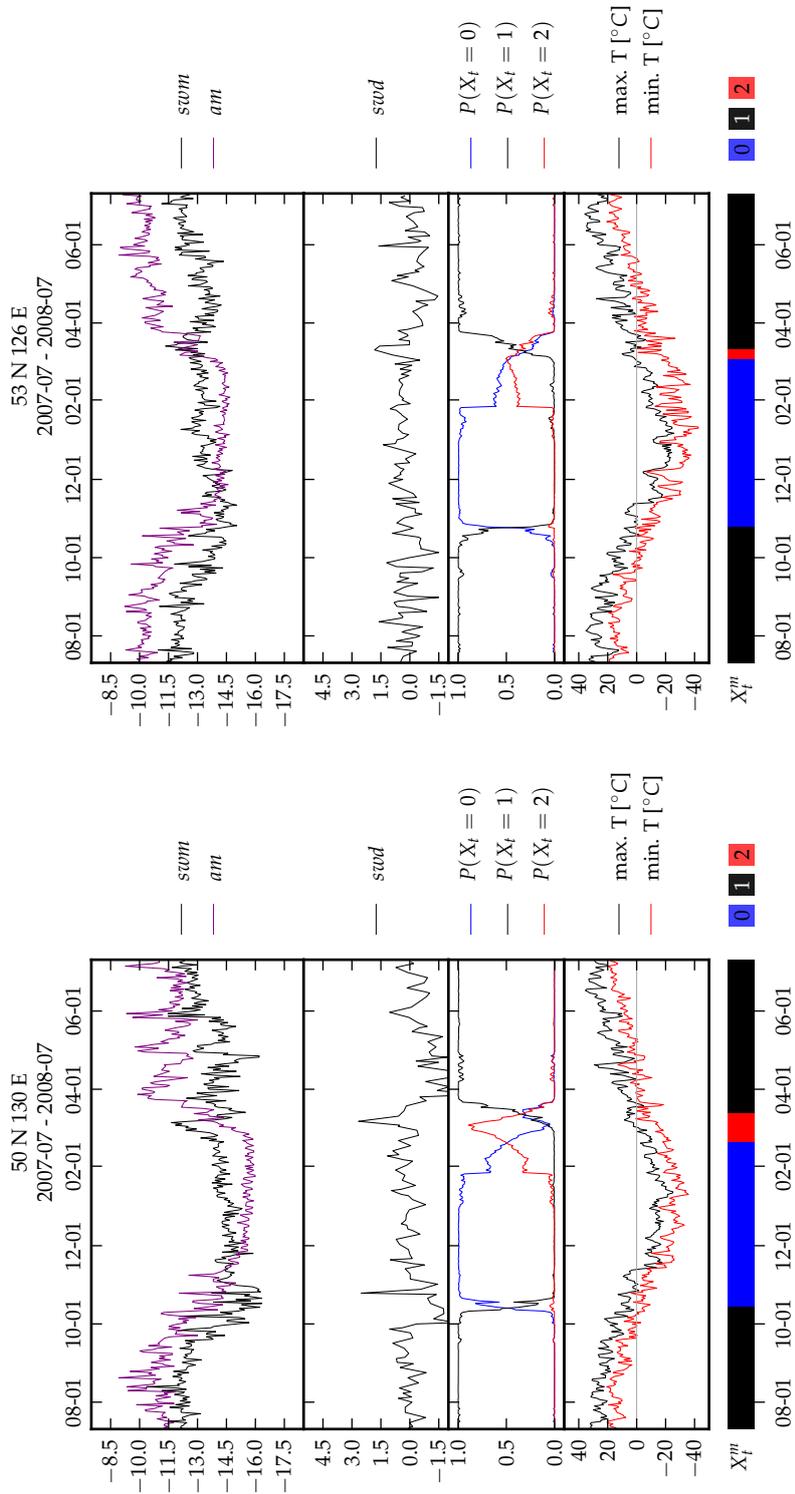
(a) *WMO station 31527*

(b) *WMO station 31371*

Figure 8.3.: Results for two different WMO observation stations.
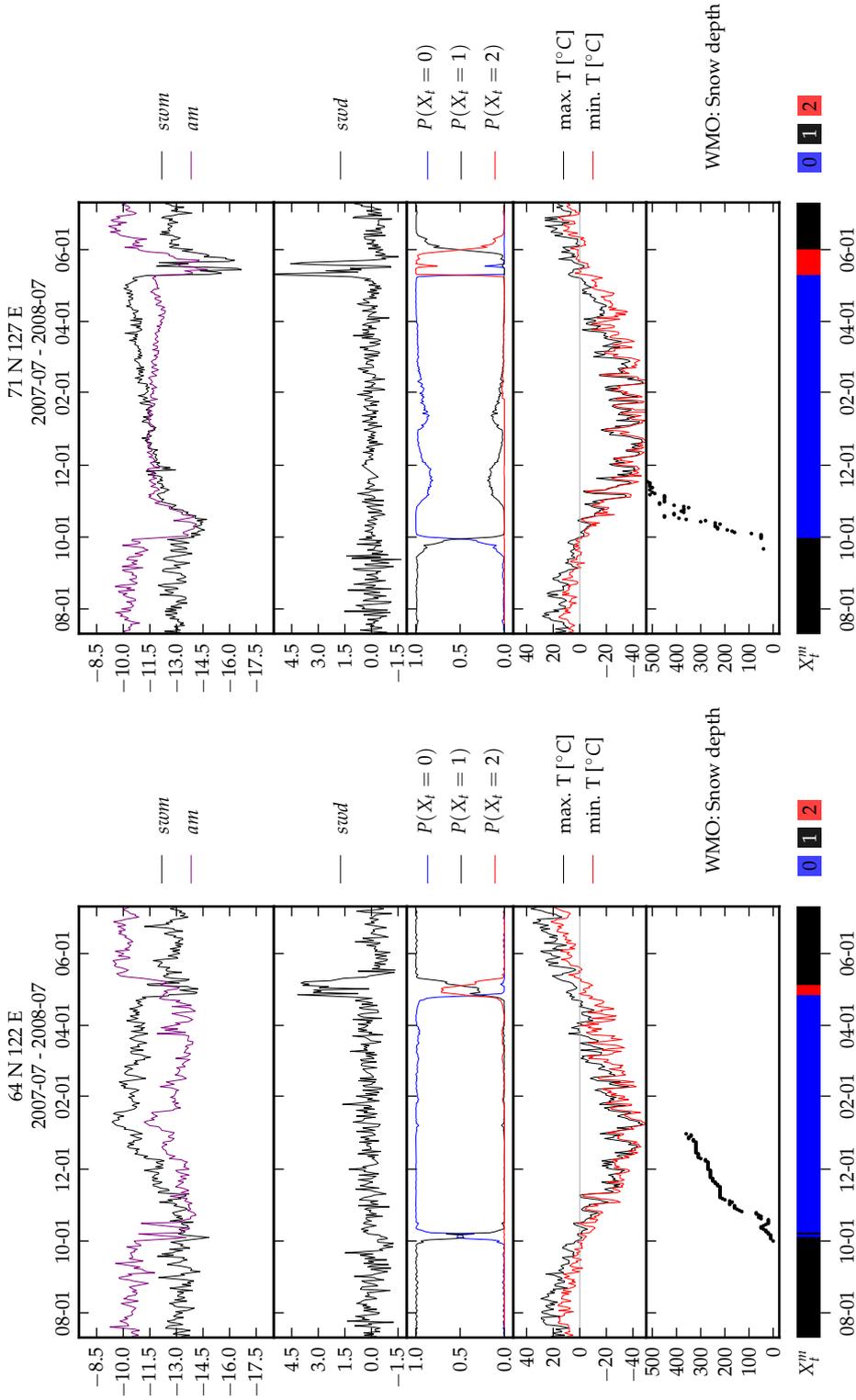
(a) *WMO station 24641*

(b) *WMO station 21921*

Figure 8.4.: Results for two different WMO observation stations, where the snow height was also recorded.

(a) WMO station 24449
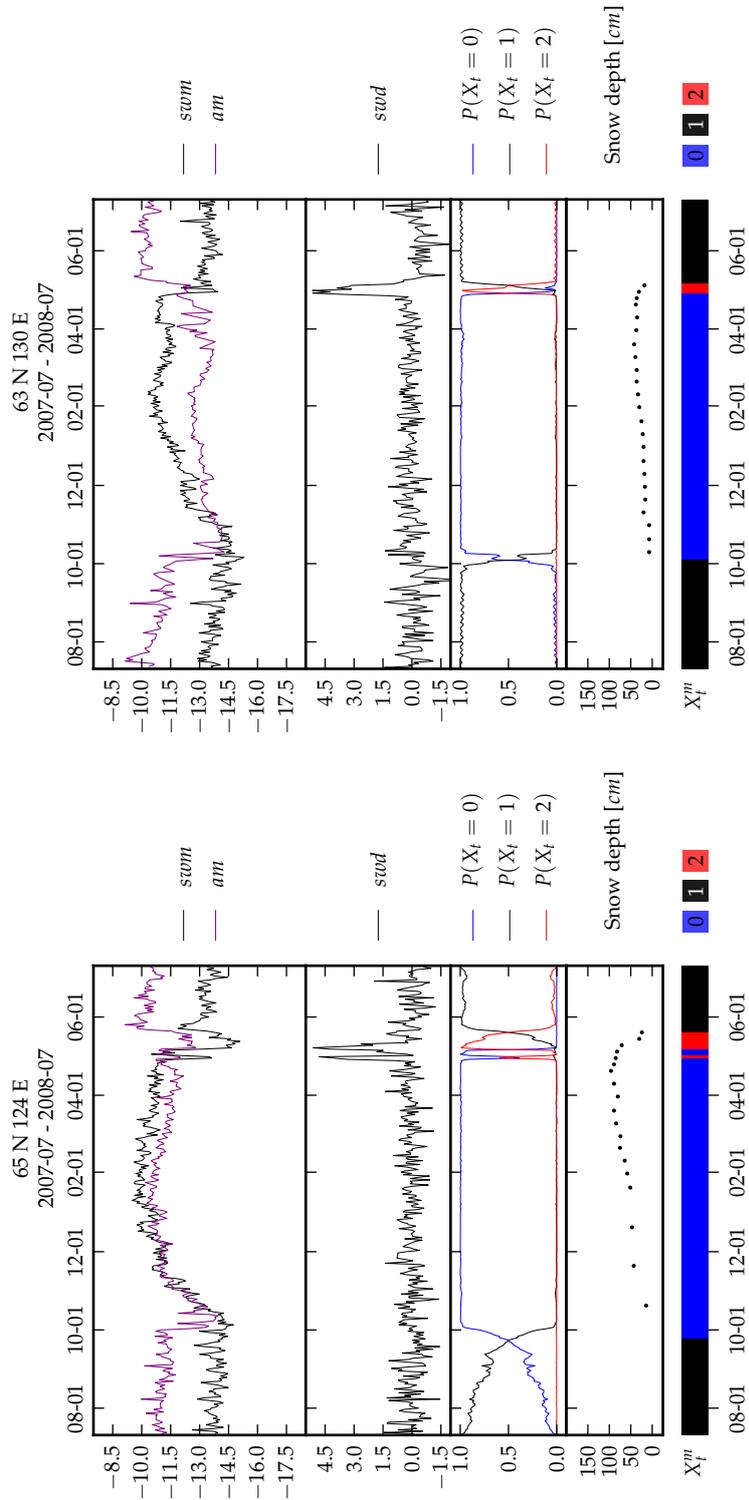
(b) WMO station 24753

Figure 8.5.: Snow observations at two stations.
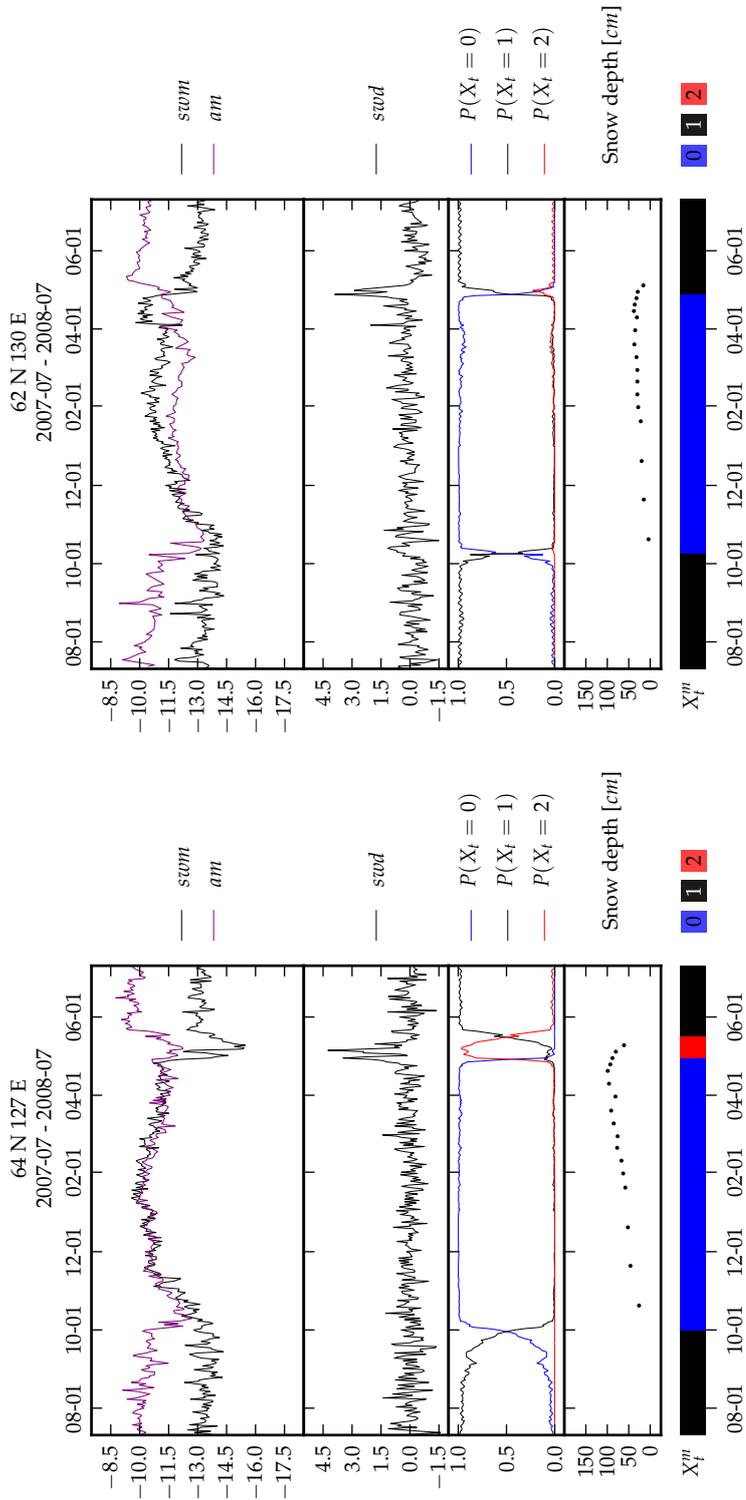
(a) WMO station 24557

(b) WMO station 24959

Figure 8.6.:: Snow observations at two stations.

8.2.2. Global data products

The global snow data set described in section 5.4 provides binary information – snow covered (0) or snow free (1) – on a daily basis. Although the data are derived in a semi-automatic way, there appear to be artifacts in the product.

The ERA Interim product (section 5.3) contains the air and the soil temperature of the top layer, both of which can be used as proxies for the freeze/thaw state. In order to facilitate comparison with the results of the sensor fusion model, the four observations per day are mapped to a discrete Freeze/Thaw state:

Frozen: 0 maximum temperature less than $-3°C$

Non-Frozen: 1 minimum temperature above $3°C$

Thawing: 2 otherwise

Based on the confusion matrix defined below, a measure of the accuracy a can be derived:

		Reference data		
		0	1	2
Results	0	c_{00}	c_{10}	c_{20}
	1	c_{01}	c_{11}	c_{21}
	2	c_{02}	c_{12}	c_{22}

$$a_T = \frac{c_{00} + c_{11}}{c_{00} + c_{01} + c_{02} + c_{10} + c_{11} + c_{12}}$$
$$a_S = \frac{c_{00} + c_{11}}{c_{00} + c_{01} + c_{10} + c_{11}}$$

where the former is applicable to temperature data and the latter to the snow cover data set. This measure of accuracy does not consider cases when the reference data indicate thawing as the transition periods predicted by temperature data alone can be very long.

The complete confusion matrix contains a lot of information and thus allows for a very thorough analysis (bearing in mind, of course, that the standard error of certain entries can be quite large). The advantage of using a single measure of accuracy lies in the fact that in can be conveniently mapped and compared. It is thus easily possible to find out in which situations the algorithm performs unsatisfactorily.

The results for the three data sets are mapped in figure 8.7. Considering the limitations of the approaches, they give a very clear and uniform picture of the overall quality.
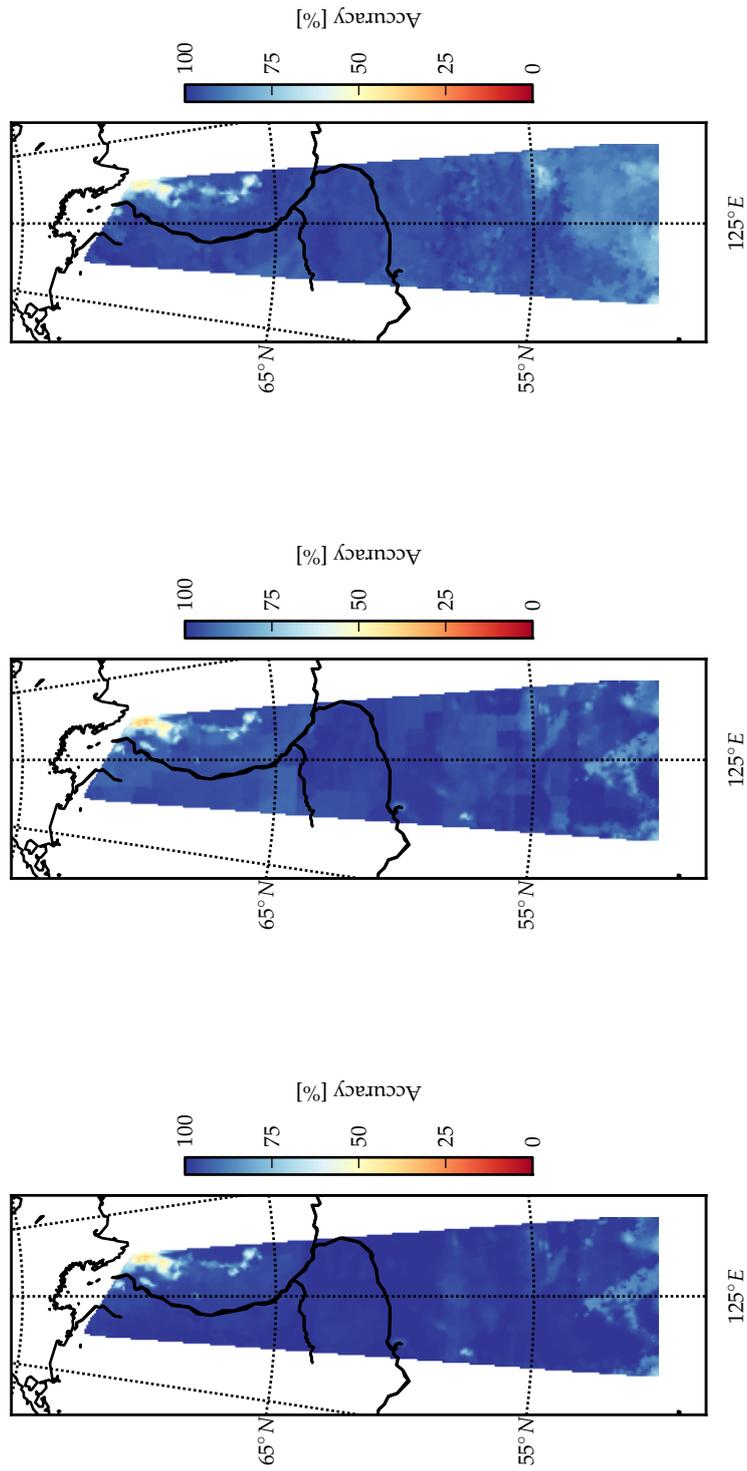
(a) ERA Interim Air Temperature.

(b) ERA Interim Soil Temperature.

(c) IMS Snow Cover.

Figure 8.7.: Accuracy a (as defined above) of the results with respect to the three reference data sets.

The algorithm works well in the higher latitudes, except for the Verkhoyansk Range, where serious problems are encountered. Even though the land cover and also the climate are quite different, there is very little overall difference in the performance between the tundra and the taiga.

The Stanovoy Range and nearby highlands (at around $57°N$) can be clearly discerned in these maps. In particular, there are two small areas where the approach seems to work insufficiently: one is barren, mountainous land similar to the Verkhoyansk range, the other is the Zeyskoye Reservoir.

In the very south of the study area, along the Amur and its tributaries as well as in the south-west corner, all reference data indicate unsatisfactory results. These areas are predominantly pastures, fields and grasslands.

Even though the reference data sets should be interpreted with some caution as they describe related but distinct geophysical parameters and are also subject to errors, they can give important insight. This is particularly applicable in high-latitudes, where in-situ observations are very sparse – there is, for example, no station in the Verkhoyansk Range. Both temperature data sets as well as the snow data reveal similar patters regarding the overall performance of the algorithm. It works well in vast stretches of the study area, but breaks down in the south and in mountain ranges. These failure cases are analysed in more detail in section 8.3.

8.3. Failure cases

The two failure cases identified previously are barren mountainous slopes and agriculturally used areas.

Mountains The time series in figure 8.8a corresponds to a grid point in the Verkhoyansk Range. The ground is barren and mainly consists of rock outcrops, characterized by excessive drainage [94]. The overall backscatter is high, which is probably due to roughness effects. Consequently, the accumulation of dry snow is barely noticeable at K_u band: the σ^0 values in winter and summer are almost the same. At C band, the freezing of the ground diminishes the backscatter by about $1dB$, which is a comparatively small drop and caused by the lack of water due to the rapid infiltration into the ground. The most important contribution is thus wet snow, as it acts as an absorbing layer and thus strongly decreases σ^0.

The detection of the spring melt works well enough, but the freeze-up is not correctly identified. This is presumably due to two reasons: firstly, the small difference between winter and summer backscatter and secondly, the parameter estimation. The initial values for the mean of *swm* and *am* for non-frozen conditions, which are used in the EM algorithm, are given by the median of the features in mid-July. In the present case, these are particularly low, which is probably caused by remaining wet snow.

Similar effects detriment the performance of the algorithm at the grid point of figure 8.8b: the difference between σ^0 in summer and winter is small, caused by the lack of soil, which makes the freeze-up only barely visible at C band. The problems are mainly encountered in winter, when $P(X_t = 0) \approx P(X_t = 1)$. The sudden change in the probabilities is due to

the mapping from S_4 to S_3 and could thus be improved by a more sophisticated mapping or additional data.

Agricultural land   The performance of the algorithm varies widely in fields and pastures. An example where it does not work well can be found in figure 8.9a. As the temperatures drop in autumn, the soil becomes gradually drier (this is corroborated by ERA Interim data); the freeze-up is thus not easily identifiable. The thawing signal is not very pronounced either, particularly when compared to the soil moisture dynamics caused by rain. In addition, there appears to be a backscatter maximum in mid-summer, which is possibly related to the rather complex vegetation phenology and human interactions such as tillage.

The same patterns are discernible in figure 8.9b, an agricultural area close to the Amur River. The rather dry conditions in autumn and presumably the plant phenology make the freeze-up difficult to pick out. Overall, the differences associated with wet and dry soil are much bigger than those associated with thawed/frozen soil. Particularly at K_u band, the backscatter within a season varies more than in between seasons and thus cannot be interpreted as easily as in the higher latitudes.

This example illustrates particularly well that the modelling choice of having one state for non-frozen condition (and not taking soil moisture into consideration) is a severe simplification. A possible work-around, that does not change the discrete state, could consist a modification of the transition probabilities so that non-frozen conditions are strongly favoured in summer.
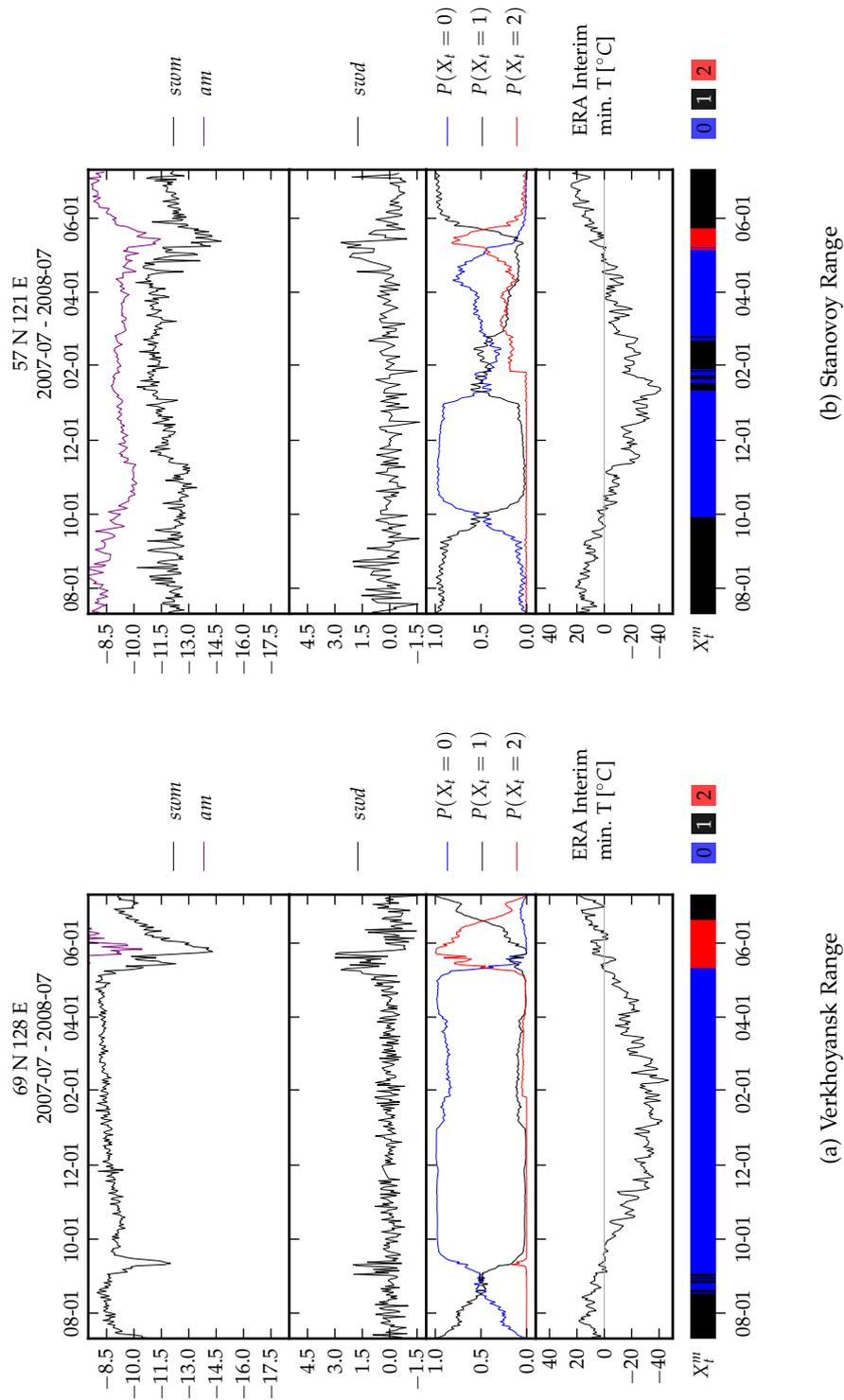
Figure 8.8.: Failure cases in the mountains.
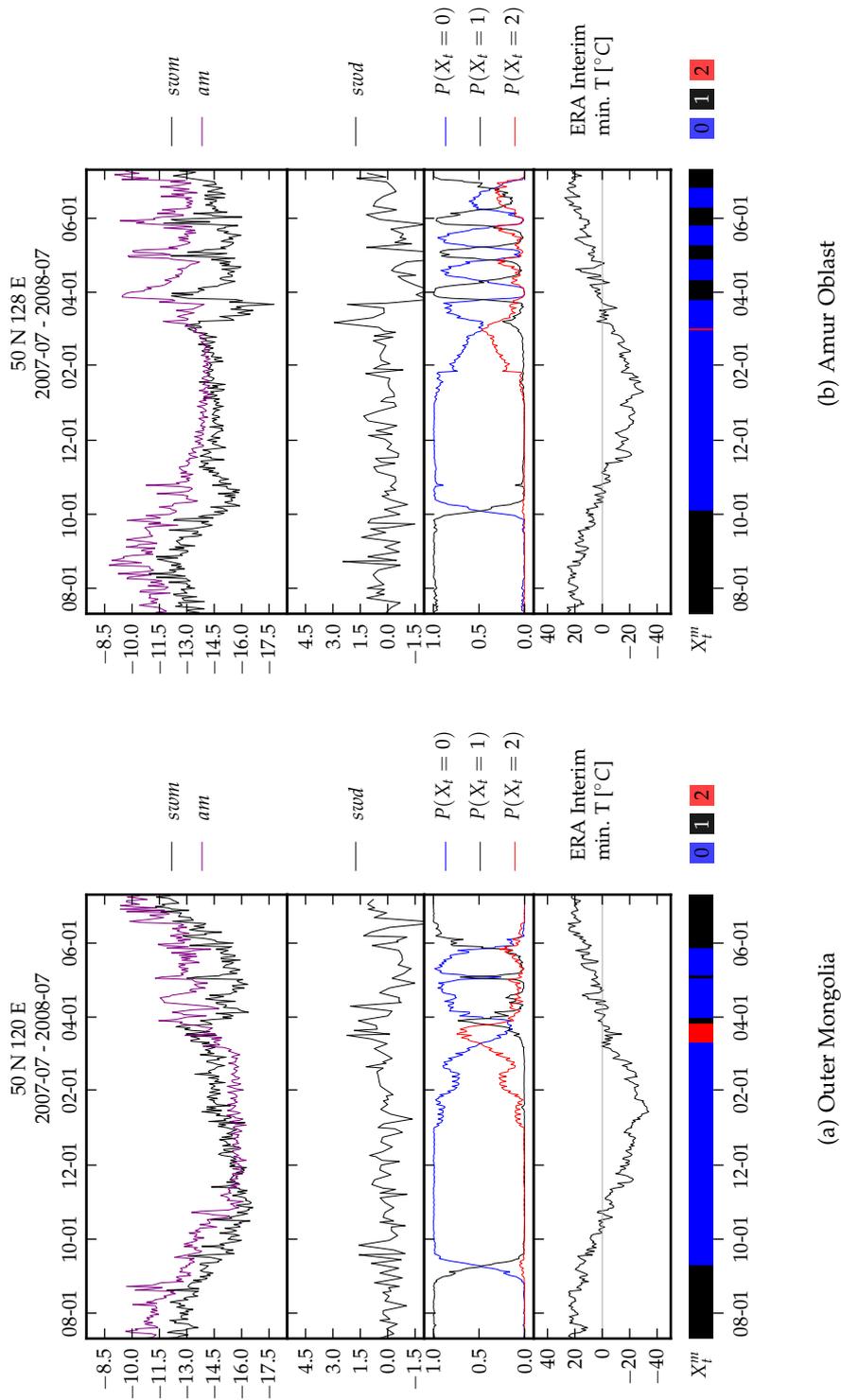
Figure 8.9.: Failure cases in agricultural land.

9. Conclusion

9.1. Outlook

The probabilistic sensor fusion model introduced in this thesis is a variant of a Hidden Markov model. Such approaches have recently enjoyed a surge in popularity, particularly when combined with pre-existing models of the 'state', in which case the term 'data assimilation' is often used. Specific applications include weather forecasting (e.g. by the ECMWF [78]) or ecosystem modelling [82].

Contrary to the aforementioned examples, in this thesis the state is a discrete variable. This concept of sensor fusion based on Hidden Markov models could be more widely applicable in remote sensing, e.g. for mapping inundated or burnt areas.

The analysis of the preceding chapters as well as considerations as to possible applications have raised several points; these concern the input data, the processing chain and the validation of the results.

9.1.1. Different Sensors

As the SeaWinds instrument onboard the QuikScat satellite has ceased operations, the question about which sensor could be used instead naturally arises. The most obvious replacement is the K_u band scatterometer onboard Oceansat-2 [52], although the combined C and K_u band scatterometers could be even more interesting: the Chinese satellite FengYun-3E [50] and the American-Japanese satellite GCOM-W2 [56].

In addition, the use of different polarizations and frequencies could be worthwhile. The algorithm could also be extended to work with data of a completely different origin, such as:

Microwave radiometers offer similar spatial and temporal resolution and coverage. The physics of the measurement process are different and could thus potentially be exploited to gain a better picture.

Visible/NIR instruments can offer data with sufficient temporal resolution, but suffer from issues related to insulation and cloud cover. The signal is influenced by different factors; for example, it would make the detection of very shallow snow possible.

9.1.2. Additional input data

Apart from remote sensing observations, global data sets derived from models could potentially enrich the sensor fusion model. Meteorological models such as the forecasts by the ECMWF are an interesting example.

Such data can be incorporated into the model in several different ways. Firstly, they can be additional visible variables Y_t; this might, for example, work well with soil temperature.

Secondly, they can be used to parameterize the emission probabilities, which might be a reasonable idea when working with snow height. Thirdly, they can be used as input variables, thus essentially modifying the transition probabilities between successive states. The air temperature lends itself to such an approach.

9.1.3. Spatial independence

In this thesis, each grid point (and thus each time series) is processed independently. In general, the results between neighbouring grid points correspond quite well; i.e. the results do not exhibit excessive noise of high spatial frequency.

Nevertheless, there are ways to extend the approach by introducing dependencies between the time series; the result would be a probability distribution of all epochs and grid points. Apart from significant difficulties with finding an appropriate parametrization, the handling would be extremely difficult and the probabilistic inference could only be done approximately.

9.1.4. NRT processing

Remote sensing data such as raw ASCAT observations are disseminated in near real time (NRT). The processing pipeline can easily be adapted to also provide results in near real time.

The inference of the marginal probability $P(x_t)$ for data available at epoch t relies exclusively on the α-messages. Due to their recursive nature, the update of the α-messages is very cheap and fast: $\mathcal{O}(N^2)$. The β-messages are not necessary for estimating the current state, but only for estimating past states. In theory, they have to be re-computed every time new data become available. Depending on the application, simple heuristics (e.g. only updating them for one week) might work very well in practice. An alternative algorithm suitable for this task is the correction smoothing by γ-messages, which is often the preferred choice in such online environments.

The Viterbi algorithm for determining $X_t^\star$ can be adapted for NRT processing as well. Compared to the implementation presented in section 4.3, the direction of the message passing has to be changed. Backtracking is not necessary if only the current state is of interest. This approach yields the state at epoch t of the most probable sequence.

The estimation of the parameters requires a sufficient number of past observations. Reasonable ways of updating them can certainly be found, e.g. re-computing them once per year.

9.1.5. Open questions

Due to the limited scope of the thesis and the paucity of external data, several important questions have been left unanswered.

The analysis of the results in chapter 8 is limited to an area with very continental climate. The next step in validating the sensor fusion algorithm should thus focus on additional biomes and climate zones.

The backscatter model and the analysis of the data mostly focus on snow as well as the freezing and thawing of soil. However, the influence of vegetation, standing water and freshwater ice is of paramount importance in many parts of the world and these factors merit additional attention.

The distinct backscatter characteristics of bare rock in the mountains and agricultural areas often lead to unsatisfactory results. The discussion in section 8 contains several possible reasons, each of which might be a starting point for possible improvements of the model:

Parameter estimation and constraints The determination of the initial values and the constraints, which are based on the backscatter model and data analysis, are somewhat arbitrary. They thus warrant additional investigations, as does the number of iterations of the EM algorithm.

State variables The notion of having a discrete state variable presents serious limitations as to the applicability of the problem. Improvements could include the inclusion of additional states (e.g. one for dry and one for wet soil) or the adoption of a more intricate description using continuous states (e.g. snow height).

State transitions As the modification of the representation of the state would necessitate significant adaptations of the algorithm, a better model of the state transitions appears very attractive. It could be based on, for example, long-term climate observations or weather forecast/re-analysis data.

Additional data Apart from such meteorological information, different remote sensing products can be potentially put to good use.

9.2. Contributions

The main contributions of this thesis can be summarized as follows:

Sensor fusion model A novel approach for sensor fusion has been introduced. The implementation presented in this work relies on K_u and C band σ^0 data. The model is rooted in probability theory; the inference and estimation of parameters can thus be conducted in a systematic way. The approach can potentially be extended to a much wider range of sensors and applications.

Backscatter model Due to the lack of literature and experiments concerning K_u and C band data on such a large scale – in particular in relation to the freeze/thaw state – a backscatter model applicable to bare ground with multiple layers of snow has been developed.

Comparison of K_u and C band In combination with extensive data analysis, the backscatter model provides insight into the frequency dependence of σ^0. This knowledge is of vital importance for the sensor fusion; in particular for the parametrization of the model. The thesis has also illustrated the significance of the metamorphism of snow, the freezing of the soil water and snow height.

9.3. Summary

The landscape freeze/thaw state is of paramount importance for all geo- and biophysical processes at high latitudes, such as the hydrological cycle and vegetation phenology. Consequently knowledge about its dynamics can be very useful for a host of applications.

For these reasons, there has been considerable interest in inferring the freeze/thaw state from remote sensing observations. Microwave remote sensing has proven to be an adequate tool and scatterometers, in particular, can offer global coverage at a reasonable temporal resolution. As the backscatter at different frequencies is influenced by different physical phenomena, the combination of several frequencies can be potentially put to good use.

This thesis focusses on two distinct points: firstly, the development and application of a novel sensor fusion model for determining the freeze thaw state – it incorporates SeaWinds K_u and ASCAT C band scatterometer measurements. Secondly, it investigates the most important factors influencing the backscatter. This analysis is based on empirical data analysis as well as a model capable of describing bare soil covered by multiple layers of snow.

The sensor fusion model is based on a probabilistic model (the so-called Hidden Markov model); the tasks of inference and parameter estimation can be conducted in the very same framework of probability. The most important output of the model is the probability of the landscape being frozen, non-frozen or thawing.

The validation of the results with external data is very promising in general; issues are, however, encountered in agricultural land and barren rock outcrops in the mountains. An in-depth analysis of the results, and the shortcomings in particular, is provided: it focusses on the validity of the modelling assumptions and points out several possible improvements.

Even though the sensor fusion model is empirical in nature, it greatly benefits from a good understanding of the signal characteristics. To this end, a widely used backscatter model for a single layer of snow on top of barren ground has been adapted; it can simulate σ^0 for multiple layers of snow with different snow grain sizes and liquid water content. It is also capable of accounting for frozen and non-frozen soil water.

The analysis of the data, the results of the sensor fusion and the simulated backscatter have reaffirmed the importance of the following factors of influence:

Snow height SeaWinds K_u band measurements are particularly sensitive to this parameter; as opposed to ASCAT C band

Frozen/thawed soil The freezing of the soil water is particularly evident in ASCAT C band. The drop in σ^0 is strongly dependent on the soil moisture

Dry snow metamorphism The backscatter is very sensitive to changes in grain size – a point which has often been neglected in global freeze/thaw state studies at C-band.

Wet snow metamorphism The absorption in the snow pack grows with increasing liquid water content; the backscatter thus drops precipitously. It also leads, however, to a rapid increase in the grain size, which can result in high σ^0 once the liquid water re-freezes.

Overall, the thesis has revealed that probabilistic sensor fusion for determining the landscape freeze/thaw state is a very promising remote sensing application. It well warrants additional studies in order to make it more robust and more widely applicable.

A. Addenda

A.1. EM updates

The energy term of the variational free energy (4.17) for the full probabilistic model in (6.2) can be written as (considering only one feature in order to minimize the notational burden):

$$\langle \log P(x,y) \rangle_{Q(x)} = \left\langle \log \left(P(x_1)P(y_1|x_1) \prod_{t=2}^{T} P(x_t|x_{t-1})P(y_t|x_t) \right) \right\rangle_{Q(x)}$$

$$= \left\langle \log \left(P(x_1) \frac{1}{\sqrt{2\pi\sigma_i^2[t]}} e^{-\frac{1}{2}\frac{(y_t-\mu_i)^2}{\sigma_i^2[t]}} \right. \right.$$

$$\left. \left. \prod_{t=2}^{T} P(x_t|x_{t-1}) \sqrt{2\pi\sigma_i^2[t]} \, e^{-\frac{1}{2}\frac{(y_t-\mu_i)^2}{\sigma_i^2[t]}} \right) \right\rangle_{Q(x)}$$

where the variances depend on the number of observations at that epoch, cf. 6.1.2

$$= \left\langle \sum_{t=1}^{T} -\frac{1}{2} \log \left(\sigma_i^2[t] \right) - \frac{1}{2} \frac{(y_t-\mu_i)^2}{\sigma_i^2[t]} \right\rangle_{Q(x)} + c$$

where c is a constant independent of the parameters. Due to the structure of the distribution, this can be simplified to

$$= \sum_{t=1}^{T} \left\langle -\frac{1}{2} \log \left(\sigma_i^2[t] \right) - \frac{1}{2} \frac{(y_t-\mu_i)^2}{\sigma_i^2[t]} \right\rangle_{Q(x_t)} + c$$

In the M-Step, this function has to be maximized with respect to the parameters. The formulae corresponding to a simple coordinate ascent scheme are summarized in algorithm 5.

Algorithm 5 M-Step for the Gaussian Hidden Markov model (6.2) with varying variances

Means $\mu_i \leftarrow \frac{\sum_{t=1}^{T} y_t Q(X_t=i) 1/\sigma_i^2[t]}{\sum_{t=1}^{T} Q(x_t=i)/\sigma_i^2[t]}$

Variances $\sigma_i^2 \leftarrow \frac{\sum_{t=1}^{T} (y_t-\mu_n)^2 Q(X_t=i) 1/\sigma_i^2[t]}{\sum_{t=1}^{T} Q(X_t=i)}$

A.2. State transitions

The state transition matrix **A**, which has to be a stochastic matrix, was chosen manually:

$$\begin{pmatrix} 0.95 & 0.05 & 0.05 & 0.12 \\ 0.01 & 0.85 & 0.05 & 0.01 \\ 0.01 & 0.05 & 0.85 & 0.02 \\ 0.02 & 0.05 & 0.05 & 0.85 \end{pmatrix}$$

for the time period between 09-15 and 12-31. During the rest of the year it was set to:

$$\begin{pmatrix} 0.85 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.85 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.85 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.85 \end{pmatrix}$$

B. Bibliography

[1] I. Ashcraft and D. Long, *Comparison of methods for melt detection over greenland using active and passive microwave measurements*, International Journal of Remote Sensing, 27(12-14) (2006), pp. 2469–2488.

[2] E. Attema and F. Ulaby, *Vegetation modeled as a water cloud*, Radio Science, 13(2) (1978), pp. 357–364.

[3] R. Bamler and P. Hartl, *Synthetic aperture radar interferometry*, Inverse Problems, 14 (1998), p. R1.

[4] D. Barber, *Bayesian Reasoning and Machine Learning*, Cambridge University Press, 2011.

[5] D. Barber and C. Taylan, *Graphical models for time series*, IEEE Signal Processing Magazine, 27(6) (2010), pp. 18–28.

[6] R. Barry, R. Armstrong, T. Callaghan, J. Cherry, S. Gearheard, A. Nolin, D. Russell, and C. Zöckler, *Global Outlook for Ice and Snow: Snow*, United Nations Environment Programme, 2007, ch. 4, pp. 39–62.

[7] S. Bartalev, A. Belward, D. Erchov, A. Isaev, E. BartholomÃl', V. Gond, P. Vogt, F. Achard, A. Zubkov, D. Mollicone, I. Savin, S. Fritz, G. Repina, and A. Hartley, *The land cover map for northern eurasia for the year 2000*. `http://www-gem.jrc.it/glc2000`, 2003.

[8] Z. Bartalis, R. Kidd, and K. Scipal, *Development and implementation of a discrete global grid system for soil moisture retrieval using the metop ascat scatterometer*, in Proocedings of the 1st EPS/MetOp RAO Workshop, 2006.

[9] Z. Bartalis, V. Naeimi, S. Hasenauer, and W. Wagner, *Ascat soil moisture product handbook*, ASCAT Soil Moisture Report Series 15, Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Austria, 2008.

[10] A. Bartsch, *10 years of seawinds on quikscat for snow applications*, Remote Sensing, 2(4) (2010), pp. 1142–1156.

[11] P. Berrisford, D. Dee, K. Fielding, M. Fuentes, P. Kallberg, S. Kobayashi, and S. Uppala, *The era-interim archive*, era report series, ECMWF, 2009.

[12] J. Birchak, C. Gardner, J. Hipp, and J. Victor, *High dielectric constant microwave probes for sensing soil moisture*, in Proceeding of the IEEE, vol. 62(1), 1974, pp. 93–98.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 1st ed. 2006. corr. 2nd printing ed., October 2007.

[14] K. Boehnke and V. Wismann, *Thawing of soil in siberia observed by the ers-1 scatterometer between 1992 and 1995*, in Proceedings of the IGARSS '96, 1996.

[15] R. Brown, M. Edens, M. Barber, and A. Sato, *Equitemperature metamorphism of snow*, vol. Snow engineering: recent advances, Balkema, 1997, ch. 1.1, pp. 41–49.

[16] W. Brutsaert, *Hydrology*, Cambridge University Press, 2005.

[17] A. Buis, *Nasa assessing new roles for ailing quikscat satellite. Press release: November 23, 2009*, tech. report, NASA - Jet Propulsion Laboratory, 2009.

[18] O. Cappé, *Ten years of hmms.* `http://www.tsi.enst.fr/~cappe/docs/hmmbib.html`, 2001.

[19] A. Chang, *Remote sensing of snowpack properties by microwave radiometry*, in Hydrologic Applications of Space Technology. Proceedings of the Cocoa Beach Workshop, Florida, August 1985, 1985.

[20] T. Christensen, *Methane emission from arctic tundra*, Biogeochemistry, 21 (1993), pp. 117–139.

[21] S. Cloude, *Polarisation*, Oxford University Press, 2010.

[22] J. Colbeck, *Thermodynamics of snow metamorphism due to variations in curvature*, Journal of Glaciology, 26(94) (1980), pp. 291–301.

[23] E. Davidson and I. Janssens, *Temperature sensitivity of soil carbon decomposition and feedbacks to climate change*, Nature, 440 (2006), pp. 165–173.

[24] M. Dobson, F. Ulaby, M. Hallikainen, and M. El-Rayes, *Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric Mixing Models*, in IEEE Transactions on Geoscience and Remote Sensing, vol. 23, 1985, pp. 35–46.

[25] M. Drusch and S. Crewell, *Principles of Radiative Transfer*, Encyclopedia of Hydrological Sciences, Wiley, 2006, ch. 46, pp. 659–671.

[26] P. Dubois, J. van Zyl, and T. Engman, *Measuring soil moisture with imaging radars*, in IEEE Transactions on Geoscience and Remote Sensing, vol. 33(4), 1995, pp. 915–926.

[27] C. Duguay and P. Lafleur, *Determining depth and ice thickness of shallow sub-arctic lakes using space-borne optical and sar data*, International Journal of Remote Sensing, 24(3) (2003), pp. 475–489.

[28] S. Durden, J. van Zyl, and H. Zebker, *Modeling and observation of the radar polarization signature of forested areas*, in IEEE Transactions on Geosciences and Remote Sensing, vol. 27(3), 1989, pp. 209–301.

[29] Earth Resources Observation and Science Center, *Srtm data*.

[30] C. Elachi and J. van Zyl, *Introduction to the physics and techniques of remote sensing*, Wiley, 2006.

[31] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models - Estimation and Control*, vol. 29 of Applications of Mathematics, Springer, 1995.

[32] European Space Agency, *Metop.* `http://www.esa.int/esaLP/LPmetop.html`, 2010.

[33] H. Ewe, H. Chuah, and A. Fung, *A backscatter model for a dense discrete medium: Analysis and numerical results*, Remote Sensing of Environment, 65 (1998), pp. 195 – 203.

[34] R. Feynman, R. Leighton, and M. Sands, *Feynman Lectures on Physics*, Addison Wesley Publishing Company, 2005.

[35] J. Figa-Saldaña, J. Wilson, E. Attema, R. Gelsthorpe, M. Drinkwater, and A. Stoffelen, *The advanced scatterometer (ascat) on the meteorological operational (metop) platform: A follow on for european wind scatterometers*, Canadian Journal of Remote Sensing, 28(3) (2002), pp. 404–412.

[36] N. Fletcher, *The nanoscience of clouds, rain and rainmaking*, Australian Physics, 43(4) (1992), pp. 116–122.

[37] M. Frank, *Approximate models for radiative transfer*, Bulletin of the Institute of Mathematics, Academia Sinica, 2 (2007), pp. 409–432.

[38] S. Frolking, K. McDonald, J. Kimball, J. Way, R. Zimmermann, and S. Running, *Using the space-borne nasa scatterometer (nscat) to determine the frozen and thawed seasons of a boreal landscape*, Journal of Geophysical Research, 104(D22) (1999), pp. 27895–27907.

[39] A. Fung and K. Cheng, *Microwave Scattering and Emission Models for Users*, Artech House, 2010.

[40] R. Gelsthorpe, E. Schied, and J. Wilson, *Ascat - metop's advanced scatterometer*, ESA bulletin, 102 (2000), pp. 19–27.

[41] Z. Ghahramani, *An introduction to hidden markov models and bayesian networks*, Journal of Pattern Recognition and Artificial Intelligence, 15(1) (2001), pp. 9–42.

[42] B. V. Gnedenko, *Theory of probability*, Gordon and Breach Science Publisher, 1997.

[43] H. Gould and J. Tobochnik, *Statistical and Thermal Physics*, Princeton University Press, 2010.

[44] M. L. Goulden, S. C. Wofsy, J. W. Harden, S. E. Trumbore, P. M. Crill, S. T. Gower, T. Fries, B. C. Daube, S.-M. Fan, D. J. Sutton, A. Bazzaz, and J. W. Munger, *Sensitivity of Boreal Forest Carbon Balance to Soil Thaw*, Science, 279 (1998), pp. 214–217.

[45] D. K. Hall, G. A. Riggs, and S. V. V, *Algorithm theoretical basis document (atbd) for the modis snow and sea ice mapping algorithms*, tech. report, MODIS Science Team, 2001.

[46] S. Helfrich, D. McNamara, B. Ramsay, T. Baldwin, and T. Kasheta, *Enhancements to, and forthcoming developments in the interactive multisensor snow and ice mapping system (ims)*, Hydrological Processes, 21 (2007), pp. 1576–1586.

[47] A. Hippel, *The dielectric relaxation spectra of water, ice, and aqueous solutions, and their interpretation. iii. proton organization and proton transfer in ice*, in IEEE Transactions on Electrical Insulation, 1988.

[48] R. Hoare, *World climate.* http://www.worldclimate.com/, 2010.

[49] M. Hohmann, *Soil freezing - the concept of soil water potential. state of the art*, Cold Regions Science and Technology, 25 (1997), pp. 101–110.

[50] Y. Honggang and D. Xialong, *Predesign of the wind field measuring radar for fengyun-3e meteorological satellite*, in International Conference on Microwave and Millimeter Wave Technology, 2010, pp. 1228–1230.

[51] N. Ida, *Engineering electromagnetics*, Springer, 2003.

[52] Indian Space Research Organisation, *Oceansat-2.* http://www.isro.org, 2008.

[53] IPCC, *Climate Change 2001: The Scientific Basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change.*, Cambridge University Press, 2001.

[54] P. Jarvis and S. Linder, *Constraints to growth of boreal forests*, Nature, 405 (2000), pp. 904–905.

[55] M. Jeffries, K. Morris, and N. Kozlenko, *Remote Sensing in Northern Hydrology: Measuring Environmental Change. Geophysical Monograph 163*, American Geophysical Union, 2005, ch. Ice Characteristics and Processes, and Remote Sensing of Frozen Rivers and Lakes, pp. 63–90.

[56] Z. Jelanek, P. Chang, J. Sienkiewicz, M. Brennan, R. Knabb, B. Ward, R. Edson, G. Hufford, C. Scott, J. Lorens, S. Stripling, B. Sjoberg, K. Schrab, and H. Kite-Powell, *Dual frequency scatterometer (dfs) on gcom-w2 satellite: Instrument design, expected performance and user impact. draft version*, tech. report, NESDIS, 2009.

[57] J. Jensen, *Remote sensing of the environment: an earth resource perspective*, Prentice Hall, 2000.

[58] Jet Propulsion Laboratory, *Quikscat science data product. user's manual. overview & geophysical data products.*, tech. report, NASA, 2006.

[59] R. Kelly, A. Chang, L. Tsang, and J. Foster, *A prototype amsr-e global snow area and snow depth algorithm*, IEEE Transaction on Geoscience and Remote Sensing, 41(2) (2003), pp. 230–242.

[60] R. Kidd, M. Trommler, and W. Wagner, *The development of a processing environment for time-series analyses of seawinds scatterometer data*, in Proceedings of IEEE 2003 International Geoscience and Remote Sensing Symposium. Toulouse. France., 2003, pp. 4110–4112.

[61] ——, *Monitoring freeze-thaw events in siberia using the seawinds ku-band scatterometer: First results*, in Geoscience and Remote Sensing Symposium, 2003. IGARSS '03, 2003.

[62] J. Kimball, K. McDonald, S. Frolking, and S. Running, *Radar remote sensing of the spring thaw transition across a boreal landscape*, Remote Sensing of Environment, 89 (2004), pp. 163–175.

[63] J. Kimball, K. McDonald, A. Keyser, S. Frolking, and S. Running, *Application of the nasa scatterometer (nscat) for determining the daily frozen and nonfrozen landscape of alaska*, Remote Sensing of Environment, 75 (2001), pp. 113–126.

[64] F. Kremer and A. Schönhals, *Theory of Dielectric Relaxation*, Broadband dielectric spectroscopy, Springer, 2003, ch. 1, pp. 1–34.

[65] K. Libbrecht, *The physics of snow crystals*, Reports on Progress in Physics, 68 (2005), pp. 855–895.

[66] D. Lide, *CRC Handbook of Chemistry and Physics*, Taylor & Francis, 2010.

[67] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2005.

[68] C. Mätzler, H. Aebischer, and E. Schanda, *Microwave dielectric properties of surface snow*, IEEE Journal of Oceanic Engineering, 9(5) (1984), pp. 366–373.

[69] K. McDonald and J. Kimball, *Estimation of surface freeze-thaw states using microwave sensors*, in Encyclopedia of Hydrological Sciences, M. Anderson, ed., Wiley, 2006, ch. 53.

[70] M. Meador and W. Weaver, *Two-stream approximations to radiative transfer in planetary atmospheres: A unified description of existing methods and a new improvement*, Journal of the Atmospheric Sciences, 37 (1980), pp. 630–643.

[71] R. Morey, A. Kovacs, and G. Cox, *Electromagnetic properties of sea ice*, Cold Regions Science and Technology, 9(1) (1984), pp. 53–75.

[72] T. Nagler and H. Rott, *Retrieval of wet snow by means of multitemporal sar data*, in IEEE Transaction on Geoscience and Remote Sensing, vol. 38(2), 2000.

[73] R. M. Neal and G. E. Hinton, *A view of the em algorithm that justifies incremental, sparse, and other variants*, in Learning in Graphical Models, M. I. Jordan, ed., Kluwer Academic Publishers, 1998, pp. 355–368.

[74] T. Neumann, M. Albert, C. Engel, Z. Courville, and D. Perron, *Sublimation rate and the mass-transfer coefficient for snow sublimation*, International Journal of Heat and Mass Transfer, 52 (2009), pp. 309–315.

[75] S. Nghiem and W. Tsai, *Global snow cover monitoring with spaceborne ku-band scatterometer*, in IEEE Transaction on Geoscience and Remote Sensing, vol. 29(10), 2001, pp. 2118–2136.

[76] NOAA/NESDIS/OSDPD/SSD, *Ims daily northern hemisphere snow and ice analysis at 4 km and 24 km resolution.* Digital media, 2004.

[77] Y. Oh, K. Sarabandi, and F. Ulaby, *An empirical model of an inversion technique for radar scattering from bare soil surfaces*, in IEEE Transaction on Geoscience and Remote Sensing, vol. 30(2), 1992, pp. 370–372.

[78] A. Persson and F. Grazzini, *User guide to ecmwf forecast products*, tech. report, European Centre for Medium-Range Weather Forecasts, 2007.

[79] C. Pohl and J. van Genderen, *Review article: multisensor image fusion. in remote sensing: concepts, methods and applications*, International Journal of Remote Sensing, 19(5) (1998), pp. 823–854.

[80] J. Pulliainen, *Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations*, Remote Sensing of Environment, 101 (2006), pp. 257 – 269.

[81] J. Pulliainen and M. Hallikainen, *Retrieval of regional snow water equivalent from space-borne passive microwave observations*, Remote Sensing of Environment, 75 (2001), pp. 76–85.

[82] T. Quaife, p. Lewis, M. de Kauwe, M. Williams, B. Law, M. Disney, and P. Bowyer, *Assimilating canopy reflectance data into an ecosystem model with an ensemble kalman filter*, Remote Sensing of Environment, 112 (2008), pp. 1347–1364.

[83] S. Rasmus, *Snow pack structure characteristics in Finland - measurements and modelling*, PhD thesis, University of Helsinki, 2005.

[84] G. Rees, *Physical principles of remote sensing*, Cambridge University Press, 2001.

[85] E. Rignot and J. Way, *Monitoring freeze-thaw cycles along north-south alaskan transects using ers-1 sar*, Remote Sensing of Environment, 49 (1994), pp. 131–137.

[86] V. Romanovsky, S. Gruber, A. Instanes, J. Jin, S. Marchenko, S. Smith, D. Trombotto, and K. Walter, *Global Outlook for Ice and Snow: Frozen Ground*, United Nations Environment Programme, 2007, ch. 7, pp. 181–201.

[87] Russian National Climate Data Centre (NCDL), *Dsi-9808*, 2010.

[88] C. Salzmann, P. Radaelli, E. Mayer, and J. Finney, *Ice xv: A new thermodynamically stable phase of ice*, Physical Review Letters, 103(10) (2009), p. 105701.

[89] K. Scipal, *Global Soil Moisture Retrieval from ERS Scatterometer Data*, PhD thesis, Technical University Vienna, 2002.

[90] E. Sharkov, *Passive Microwave Remote Sensing of the Earth*, Springer, 2003.

[91] J. Shi and J. Dozier, *Estimation of snow water equivalence using sir-c/x-sar, part i: Inferring snow density and subsurface properties*, in IEEE Transaction on Geoscience and Remote Sensing, vol. 28(6), 2000, pp. 2465–2474.

[92] A. Simmons, S. Uppala, D. Dee, and S. Kobayashi, *Era-interim: New ecmwf reanalysis product from 1989 onwards*, ECMWF Newsletter, 10 (2006), pp. 25–35.

[93] M. Spencer, C. Wu, and D. Long, *Improved resolution backscatter measurements with the seawinds pencil-beam scatterometer*, IEEE Transactions on Geoscience and Remote Sensing, 38(1) (2000), pp. 89–104.

[94] V. Stolbovoi and I. McCallum, *Land resources of russia*. CD-ROM, 2002.

[95] J. Svoboda, *Tundra, Arctic and Antarctica*, Encyclopedia of Environmental Sciences, Kluwer Academic Publishers, 1999, pp. 630–635.

[96] V. Tikhonov, *Model of complex dielectric constant of wet and frozen soil in the 1-40 ghz frequency range*, in IGARSS. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation, vol. 3, 1944, pp. 1576–1578.

[97] W.-Y. Tsai, M. Spencer, C. Wu, C. Winn, and K. Kellogg, *Seawinds on quikscat: Sensor description and mission overview*, in IGARSS 2000 - Proceedings of IEEE 2000 National Geoscience and Remote Sensing Symposium, Honolulu. 24-28 July 2000., 2000, pp. 1021–1023.

[98] P.-H. Tseng, T. Illangasekare, and M. Meier, *Modeling of snow melting and uniform wetting front migration in a layered subfreezing snowpack*, Water Resources Research, 30(8) (1994), pp. 2363–2376.

[99] F. Ulaby and M. El-Rayes, *Microwave dielectric spectrum of vegetation - part ii: Dual-dispersion model*, in IEEE Transaction on Geoscience and Remote Sensing,, vol. 25(5), 1987.

[100] F. Ulaby, K. Sarabandi, K. McDonald, M. Whitt, and C. Dobson, *Michigan microwave canopy scattering model*, International Journal of Remote Sensing, 11(7) (1990), pp. 1223–1253.

[101] US Climate Prediction Center, *ds512.0: Cpc global summary of day/month observations*. `http://dss.ucar.edu/datasets/ds512.0/`.

[102] N. Verhoest, H. Lievens, W. Wagner, J. Alvarez-Mozos, M. Moran, and F. Mattia, *On the soil roughness parameterization problem in soil moisture retrieval of bare surfaces from synthetic aperture radar*, Sensors, 8(7) (2008), pp. 4213–4248.

[103] L. Wald, *Some terms of reference in data fusion*, IEEE Transactions on Geoscience and Remote Sensing, 37(3) (1999), pp. 1190–1193.

[104] P. Walstra, *Approximation formulae for the light scattering coefficient of dielectric spheres*, British Journal of Applied Physics, 15 (1964), pp. 1545–1552.

[105] U. Wegmüller, *The effect of freezing and thawing on the microwave signatures of bare soil*, Remote Sensing of Environment, 33 (1990), p. 123:135.

[106] I. H. Woodhouse, *Introduction to Microwave Remote Sensing*, Taylor & Francis, 2006.

[107] C. Wu, J. Graf, M. Spencer, and C. Winn, *The seawinds scatterometer instrument using a pulse compression radar*, in Proceedings of SPIE: Sensors, Systems, and Next-Generation Satellites, 1997, pp. 340–345.

[108] J. Wu, *On the convergence properties of the em algorithm*, The Annals of Statistics, 11(1) (1983), pp. 95–103.

[109] XYdatasource. `http://www.xydatasource.com`, 2010. last access: 2010-12-06.

[110] E. Yershov, *General Geocryology*, Cambridge University Press, 1998.

[111] Y. Zeng, J. Zhang, and J. van Genderen, *Comparison and analysis of remote sensing data fusion techniques at feature and decision levels*, in From Pixels to Processes. ISPRS Mid-term Symposium, 2006.

[112] B. Zuerndorfer, A. England, C. Dobson, and F. T. Ulaby, *Mapping freeze/thaw boundaries with smmr data*, Agricultural and Forest Meteorology, 52(1-2) (1990), pp. 199–225.