

**Large-scale biological transportation
networks**
cargo ship traffic and bird migration

Von der Fakultät für Mathematik und Naturwissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels
eines Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

von Andrea Kölzsch,
geboren am 15.07.1980 in Wismar

Gutachter
Zweitgutachter
Tag der Disputation

Prof. Dr. Bernd Blasius
Prof. Dr. Franz Bairlein
11.09.2009

Contents

1	General Introduction	5
1.1	Transportation in a globalised world	5
1.2	Bioinvasion and epidemics spread	6
1.3	Random walk theory and movement analysis	8
1.4	Complex networks	9
1.5	The importance of global cargo ship traffic	11
1.6	Issues of bird migration	12
1.7	Outline of the included papers	13
2	Paper I.	
	Regularity and randomness in the global network of cargo ship movements	17
2.1	Introduction	17
2.2	Data	18
2.3	The global network of cargo ships	20
2.4	The network layers of different ship types	22
2.5	Network trajectories	24
2.6	Discussion	26
2.7	Supplementary Information	29
3	Paper II.	
	Indications of marine bioinvasion from network theory	37
3.1	Introduction	37
3.2	Methods	39
3.3	Results	44
3.4	Discussion	50
4	Paper III.	
	Theoretical approaches to bird migration	55
4.1	Introduction	55
4.2	Data for identifying bird migration routes	57
4.3	Movement analysis for the white stork	59
4.4	Quantitative modelling of bird migration	69
4.5	Conclusions	74
5	Paper IV.	
	Lévy flights in bird movement after all?	77

6 Paper V.	
A periodic Markov model of bird migration on a network	81
6.1 Introduction	81
6.2 Model	83
6.3 Data	85
6.4 Methods	86
6.5 Results	90
6.6 Discussion	94
7 General Discussion	99
7.1 Transportation networks in comparison	100
7.2 Superdiffusion of bird migration movement	103
7.3 Future perspectives of biological transportation	104
8 Summary	107
9 Zusammenfassung	111
10 Bibliography	115
11 Acknowledgements	129
12 Curriculum vitae	131
13 Personal contributions	135
Erklärung	137

1 General Introduction

1.1 Transportation in a globalised world

Life is, in a large part, formed by active and passive movement of organisms. Nearly all animals locomote from place to place for foraging, to look for mates and shelter and to avoid predators (Begon *et al.*, 2006). Such motion is mostly small-scale, many animals reside in a certain local territory or habitat wherein they move about. However, there are some species that conduct movement over much longer ranges, often during certain times of their life. Such movement is for example seasonal migration or the dismigration of juveniles. The former are return migrations between more or less distant areas and are conducted to exploit exceptionally good food and weather conditions of some strongly seasonal governed regions for e.g. breeding, but avoid their harsh winter conditions. The latter, dismigration, is a once in a lifetime long-distance displacement of young animals to find an appropriate breeding territory (Berthold, 2001a). These two processes determine the species' dispersion. Also plants, that are usually immobile, can be dispersed over long distances by seeds being transported by the wind, water currents or animals (Nathan, 2006). While moving, many animals transport smaller organisms, pathogens, seeds and else, purposefully or accidentally.

Migratory animals, especially birds, regularly perform long-range movement and are thus very likely to spread organisms attached to their plumage. Migratory birds have travelled their migration routes for a long time and ecosystems have over the years adapted to the influences of the birds passing through, feeding and interacting with other species. Humans, however, extremely increased their movement over the last centuries. We have dispersed to almost all regions of the earth and move around at ever increasing rates for the exchange of goods and knowledge, finding jobs, tourism and other reasons. Often small organisms or seeds are taken along deliberately or as stowaways. In former times the transport of goods and information has been conducted comparably slow by foot, on horseback or with small boats. Today transportation has by orders of magnitude increased in velocity and quantity, due to the usage of huge ships, trains, trucks and airplanes. Most such means of transportation are organised in networks spanning large parts of the world. Geographical barriers that in the past complicated transportation and the spread of organisms do not apply any more in today's times of globalisation.

On the one hand, due to the transport process itself the environment is polluted by vehicles' exhaust fumes and natural habitat destroyed for building roads and ports. On the other hand, global transport leads to uncontrolled global dispersal of alien species, i.e. bioinvasion. These processes lead to several ecological complications, not the least important being the modification of ecosystems' structures and functioning (Crowl *et al.*, 2008). Regarding bioinvasion this means that the introduction of large numbers of alien species at high rates may lead to increased competition for resources with native species. Such can entail the extinction of some

native species, the modification of species interaction structures and therefore change provided ecosystem services. One important motivation for the here presented work is the accelerated global spread of bioinvasive organisms and epidemics. It is, surpassed only by habitat destruction, the second most important threat to global biodiversity and human health and livelihood (Mack *et al.*, 2000).

1.2 Bioinvasion and epidemics spread

Biological invasions are geographical expansions of a species into an area that was not previously occupied by it. This quite natural process has become problematic for biodiversity and ecosystem functioning worldwide recently, as it has been greatly intensified due to deliberate and accidental human transportation (Elton, 1958). Bioinvasion is a process of three stages (Vermeij, 1996). After an invasive species that has arrived (stage 1) to a novel region has also become established (stage 2) it may proliferate (stage 3) and cause devastating changes in ecosystems. The adverse effects of successful biological invasions vary tremendously (Mack *et al.*, 2000). Invasive species may have very little impact, but in the extreme, they can drive native species to extinction as well as extremely impact the economy, e.g. agriculture and fishing. Such cases are mostly not predictable, therefore bioinvasion research is of great general relevance.

One example of a detrimental invasion is the introduction of the comb jelly *Mnemiopsis leidyi*, possibly by trading ships, into the Black Sea in the early 1980's (Vinogradov *et al.*, 1989). There it caused an extreme decrease in fish populations, especially the commercially important European anchovy (*Engraulis encrasicolus*), by predation and competition for food. Since 2006 the comb jelly is also recorded in the Baltic Sea (Javidpour *et al.*, 2006) which is recently of great concern.

Mnemiopsis leidyi is only one of extremely many invasive species that negatively impact the global biodiversity, ecosystem functions and human life. Several data bases that list invasive species and their histories have lately become available on the web. The most prominent one is the Global Invasive Species Database (<http://www.issg.org/database/>) that was developed by the Global Invasive Species Programme (GISP) and is managed by the Invasive Species Specialist Group (ISSG) of the Species Survival Commission of the IUCN-World Conservation Union. It lists invasive species of all taxa. A highly invasive plant is gorse (*Ulex europaeus*), a shrub that is very competitive and alters the soil conditions and fire regime in wide ranges of the Americas, Australia and several Pacific islands. As an example of the large number of invasive insects we want to name the brown house-ant (*Pheidole megacephala*). It is a pest to agriculture, destroys electrical wiring and has displaced several native invertebrates throughout all temperate and tropical regions of the world. Examples of aquatic invasive species are the marine algae *Caulerpa taxifolia* that by its dense growth excludes almost all marine life in some regions of the Mediterranean, and the Nile perch (*Lates niloticus*). The introduction of this large freshwater fish into Lake Victoria led to the mass extinction of more than 200 endemic fish. One of the oldest invasive species is the ship rat (*Rattus rattus*) that has spread throughout the world by ships. It has caused the extinction of a large number of often endemic species on islands and transmits the plague (*Yersinia pestis*) in certain parts of the world.

A large number of studies on biological invasions have examined species and habitat characteristics favourable for the establishment and proliferation of aliens (Mack *et al.*, 2000; Kolar & Lodge, 2001). In particular, often high impact invasion events were analysed and control measures suggested. It is surprising that comparably few studies focus on the first stage of bioinvasion, the transport and introduction. At this stage it is most likely that an invasion event can be prevented and detrimental impacts be avoided. Therefore, in this work we focus on characterising patterns of movement and transportation, identifying routes of possible introductions and spread of bioinvasive organisms.

Also diseases are today spreading further and more quickly than ever before, thus forming epidemics that quickly disperse on earth, transported by human travel. One disease of recent concern has been the severe acute respiratory syndrome, SARS, which has almost become a pandemic in 2003 when it rapidly spread from China to more than 30 countries around the world (Smith, 2006). How important the global aviation network of passenger airplanes was for the extremely fast spread of SARS is apparent from the study of Hufnagel *et al.* (2004). They developed and analysed the aviation network and simulated an epidemics spread, mimicking SARS, on it. Results coincide surprisingly well with the patterns of the real expansion of SARS throughout the world in 2003. A very recent example of the propagation of human infectious diseases is the rapid spread of influenza A, H1N1, in the year 2009. Despite all undertaken measures spread over the globe could not be prevented and the WHO declared it a pandemic (www.who.int).

The dynamics of epidemics spread are in line with bioinvasion and can be considered a special case of it. The major effects of epidemics, however, are directed on human and animal/plant health rather than ecosystem functioning. This difference influences establishment probabilities, because e.g. human diseases find suitable conditions in hosts almost anywhere on earth, whereas bioinvasive organisms have to adapt to novel habitats. Lately also the spread of avian diseases by migratory birds, e.g. avian influenza H5N1 (Olsen *et al.*, 2006) and the West Nile virus (Blitvich, 2008), has become of great concern (see below).

Most studies of bioinvasion and disease spread assume the movement of the transport vectors or individuals as random and independent of time and space (Turchin, 1998). It has been shown in several studies that this is usually not the case (Turchin, 1998; Viswanathan *et al.*, 2008). The motion of many organisms is reminiscent of a Lévy flight, i.e. its displacement distances follow a power law $P(d) = d^{-\beta}$ (see also below). These facts should be accounted for in any study of dispersal and spread, especially because long-range displacements strongly dominate proliferation patterns and accelerate the spread of infectious diseases and bioinvasion extremely. Consequently, bioinvasion risk and epidemics spread can be of variable prominence depending on the frequency of long-distance displacement events.

Furthermore, bioinvasion success and the outbreak of infectious diseases often depend on the environmental conditions during the transport and in the recipient areas. Not only survival, but also the patterns of movement and spread of many species are impacted by the environment (Begon *et al.*, 2006). Environmental characteristics prominently change with season and can in this sense drive spread dynamics. One example is the global spread of cholera and its sometimes seasonal outbreaks. Cholera is one of the most feared infectious diseases worldwide. It is usually spread by contaminated food or water, but lately even transported in ships' ballast tanks (Lee, 2001). High prevalence of cholera has been associated with El Niño events, high temperatures

and humidity. Recently, it was discovered that toxigenic cholera bacteria can survive for long times in association with zooplankton (Colwell & Huq, 1994) the population dynamics of which follow seasonal algal blooms. Other transportation vectors may as well be seasonally driven, e.g. tourist transportation in temperate regions, trade with seasonal fruits and seasonal migrations of mammals and birds. In this respect, it is often important to consider issues of seasonality and environmental conditions when studying long-range movement and spread.

1.3 Random walk theory and movement analysis

Human and animals' movements are, as other processes in nature, very complex phenomena. For unravelling its properties and dynamics they are often compared with random movement. Random processes have been studied in physics for more than a century. A starting point was the description of complex, somewhat erratic processes like Brownian motion with the simple model of a random walk (Hughes, 1995). It is defined as stepwise movement of equidistant increments which develop into an independent, identically distributed random direction. This basic random walk concept can be modified in different ways. Important ones are to introduce variability of the step lengths, a preference of the movement direction, i.e. drift, and correlations of successive directions. Recently, the importance of studying the complex movement of animals and explicitly including it in population dynamics models has been pointed out (see above; Turchin, 1998). Usually, movement has been assumed random in such models, a fact which can have a strong effect on the model outcome. Several studies have shown that animals' movement cannot be described as purely random, but rather resembles correlated random walks (Taylor, 1922; Kareiva & Shigesada, 1983), Lévy flights (Shlesinger *et al.*, 1982; Viswanathan *et al.*, 1996) or other complex patterns.

For analysing movement trajectories there are two tools commonly applied. First, one is interested in how far the object under consideration moves per time. Calculating the mean squared displacement (MSD) and relating it to time provides a measure of the distribution of displacement distances and of the directionality of the movement. In case the MSD increases proportional to the square root of time $MSD \sim \sqrt{t}$ the movement is diffusive. A direct proportionality of MSD to time $MSD \sim t$ indicates ballistic, directed movement (Ben-Avraham & Havlin, 2000). A second tool for the analysis of movement is the turning angle distribution of directional changes during short time intervals. Its shape allows for indications if the movement is homogeneous or composed of different modes, and how they are characterised.

A recently widely applied model of movement is the Lévy flight that has first been mentioned by Taylor (1922). It is a special type of superdiffusive movement. The stepwise motion is comprised of randomly directed displacements of lengths d that are drawn from a power law distribution $P(d) = d^{-\beta}$. Its exponent $1 < \beta < 3$ determines the character of the Lévy flight, how strongly superdiffusive it is. The extreme cases mark for $\beta = 3$ Brownian movement, i.e. random movement with identical step lengths, and for $\beta = 1$ indefinite ballistic movement. Special properties are that the standard deviation of the distribution is not defined for $\beta < 3$ and that for $\beta < 2$ even the mean is infinite. A generalisation of the Lévy flight is the so called Lévy walk. Its displacement lengths are also drawn from a power law distribution and turning angles are independent, identically distributed. Time is, however, not considered in discrete

steps, but as continuously flowing proportional to the distance covered while moving.

The movement of several animal species as well as humans has been shown to resemble Lévy flights (e.g. Viswanathan *et al.*, 1996; Marell *et al.*, 2002; Ramos-Fernandez *et al.*, 2004; Brockmann *et al.*, 2006; Sims *et al.*, 2008; Viswanathan *et al.*, 2008; González *et al.*, 2008). Different methods for studying these patterns have been proposed and controversially discussed. The simplest is to examine the fit of a line to the doubly logarithmic distribution plot. However, this may be strongly biased, influenced by binning intervals and allows for spurious results (Newman, 2005; Clauset *et al.*, 2009). Therefore, it was proposed to analyse the cumulative distributions and use model selection with Akaike weights (Burnham & Anderson, 1998; Edwards *et al.*, 2007).

The issue of biological Lévy flights has lately been raised to question. Edwards *et al.* (2007) proposed that previous studies used inadequate methods and Lévy flights do not exist in nature. However, recently Sims *et al.* (2008) studied the vertical displacement of several fish species using rigorous statistics like model selection. They clearly reveal Lévy-like motion patterns. We would like to note the importance of not only examining the existence of Lévy flights from data, but to determine the mechanisms that bring about such movement patterns and if they are optimal in any respect. However, to discern such mechanisms is very complicated and would have to include extensive empirical and theoretical studies. One mechanism that has been proposed is optimal foraging (Viswanathan *et al.*, 1999), others may be heterogeneities in the environment, food distribution and more complex behavioural aspects.

Another concept for analysing movement and flow through some kind of environment is percolation theory (Stauffer, 1994). It is concerned with the properties of the environment and how different transition probabilities between spatially distinct regions influence the flow of an agent through it (Ben-Avraham & Havlin, 2000; Bollobás & Riordan, 2006). Of special concern is the transition probability that allows for global spread of the agent, called the critical threshold. Such can be quantified and readily be used for a preliminary study of bioinvasive and epidemics spread on discretised environments like e.g. complex networks (see below).

1.4 Complex networks

A network consists of a number of nodes that represent certain well defined entities, e.g. people, places or substances that are pairwise connected by links whenever a specified relation exists between them (Newman, 2003b). These relations can be symmetric, leading to the development of an undirected network, or asymmetric in a directed network. Furthermore, network links can be weighted by a measure of distance or interaction strength or unweighted. As many complex systems of our world are comprised of specific discrete items that interact, the concept of networks can be applied in a wide range of fields. So, networks and network theory have been developed not only in mathematics and theoretical computer science, but in sociology, physics and biology. A large number of measures characterising the structure and dynamics of networks have become available, readily being used for examining practical issues of the studied systems (Costa *et al.*, 2007).

One example of complex networks are metabolic networks, the nodes of which represent metabolic substrates and products and directed links are drawn if a substrate can be converted

into a given product by a metabolic reaction. Jeong *et al.* (2000), for example, have characterised the metabolic networks of several species and found notions of robustness and so called scale-free behaviour. This means that the longest path between any pair of nodes does not change much when deleting single nodes, and that the distribution of the number of links a node has, i.e. the degree k , is distributed as a power law $P(k) = k^{-\gamma}$. Thus, a small number of substrates/products are exceptionally important for the metabolism while a large number of others are involved in only very few reactions. The striking property is that this relation holds on any scale of degrees. Other network measures characterise small-scale as well as large-scale topological structures and spread properties. They are local and global efficiencies and network cost (Latora & Marchiori, 2001), the kind and strength of assortativity, symmetry and the motif distribution (Milo *et al.*, 2002). Furthermore, network nodes can be specified by centrality and other functional roles and grouped into exceptionally well connected subgraphs (Newman, 2003b). Several measures of the robustness of networks and spread through networks, using e.g. percolation theory (Ben-Avraham & Havlin, 2000), have been proposed.

Much research has been concerned with examining the topology *of* networks and processes *on* networks (Gross & Blasius, 2008). Mechanistic models have been developed to explain the topologies *of* different kinds of networks. Some examples are the random graph of Erdős & Rényi (1959), the small-world model of Watts & Strogatz (1998) and the scale-free networks of Barabási & Albert (1999). These models describe mechanisms of possible network generation, trying to mimic and explain natural processes of network construction. One example is preferential attachment (Barabási & Albert, 1999), the main idea of which is that new nodes prefer to link to existing nodes that already possess a large neighbourhood, i.e. many nodes linked directly to them.

A statistical model to describe processes *on* a network is the Markov chain model that is used to describe simple transition processes between discrete nodes. It is characterised by a matrix of transition probabilities that are usually constant in time, i.e. homogeneous, and do not depend on process states in the past. This model is thus memoryless and transition events independent of the time and from each other (Norris, 1997). Markov processes are characterised by time interval lengths needed for returning to a certain node (recurrence and transience times) and eventually converge to a stable state of density distribution on the nodes. One application of a Markov model in biology is the SIS model of epidemics (Bailey, 1975). In this model susceptible individuals get infected from infected individuals with a constant infection rate and infected ones recover with a constant recovery rate. In the SIS model, after a transient time the proportions of infected and susceptible individuals reach an equilibrium.

An extension to these simple, homogeneous Markov models are non-homogeneous ones. Then transition probabilities are not constant but dependent on time or other parameters that change during the process. Non-homogeneous Markov models have first been developed for modelling manpower systems (Young & Almond, 1961), but were not much studied theoretically (Vassiliou, 1998). They have recently been applied in ecology for modelling forest succession (Usher, 1979) and seasonal population dynamics of a zoobenthos community (Patoucheas & Stamou, 1993). In those studies transition probabilities between different successional community stages depend on the season.

The special group of networks that we are concerned with in this study are transportation networks. They describe structures that convey certain entities from one point to the other.

Examples are roads, railways, powerlines and conveyor belts, but also river networks, mammalian circulatory systems and plant leaves. A main objective of transportation networks is that they should somehow be able to optimise the flow of goods or other entities that they facilitate. The recent availability of a large amount of data of transportation networks made it possible to analyse their network characteristics. One optimal structure that has been observed in several transportation networks is scale-free behaviour (see above), e.g. in the worldwide aviation network (Guimera *et al.*, 2005). Thus, such transportation networks have a large number of nodes with small degree, but also a few rather well connected nodes, the hubs. These structures enable a quick flow, i.e. a small shortest path, between most pairs of nodes. Further characteristics typical for transportation networks are high efficiencies and low cost, also called small world behaviour, and robustness to random node deletions.

1.5 The importance of global cargo ship traffic

As noted above, levels of intensity and the rate of global travel and transportation of goods are at a height never seen before, and there are immense networks of transportation that provide this. Of the many different modes of goods transport, conveyance by large, oceangoing cargo ships emerges as the most energy efficient one for large quantities of goods (Rodrigue *et al.*, 2006). It has been estimated that a striking proportion of about 90% of the world trade (in terms of tonnes and distance) is hauled by ships, and it is increasing still, at a rate even faster than global economy (IMO, 2006). Therefore, global cargo ship traffic is already very interesting itself, projecting the patterns of world trade.

Already in the 13th and 14th century ships have travelled long distances and spread foreign species, in the wooden ships it were often wood-boring shipworms (Carlton, 1996a). However, since Columbus discovered America in 1492, ships have become larger and larger and increasingly safe for travelling the open oceans. So, as they went along they travelled all parts of the world. Then, marine bioinvasive organisms spread at a highly increased rate more or less accidentally into pristine habitats and native ecosystems were altered.

Today, the spread of bioinvasive organisms by cargo ships is of great importance. The major vectors for marine bioinvasion are discharged water from ships' ballast tanks (Carlton, 1996a; Ruiz *et al.*, 2000) and hull fouling (Drake & Lodge, 2007), but also terrestrial species are inadvertently transported in shipping containers (Lounibos, 2002). Depending on the number of invasive organisms that survive the transport and environmental conditions in the recipient port, they may establish into new habitats, proliferate and impact ecosystem functioning.

Cargo ship traffic has been quantified before using gravity models (Drake & Lodge, 2004; Haynes & Fotheringham, 1984). This means that, given only incoming and outgoing amounts of commodity for most ports worldwide, fluxes between the ports have to be estimated. Usually, the transition rate between pairs of ports is described as a function of distance. Thus, a randomisation of global trade is assumed. As trade is, however, strongly structured because of political, cultural and other reasons, this assumption introduces faults in the proposed ship traffic network.

A newly established source of data of actual ship movement trajectories has become available recently. To avoid collisions and improve port security in 2001 ships and ports have begun

installing Automatic Identification System (AIS) equipment. Whenever a ship arrives to or departs from a port its AIS transmitter automatically reports it to the port authorities. Such data of port arrival and departure are made commercially available by Lloyd's Register Fairplay in its sea-web database (www.sea-web.com). They provide lists of ports that are consecutively called at by single ships. As these "ship trajectories" are available for the majority of all ships worldwide they map the global ship traffic as it is. Interesting questions are now what general properties of global cargo traffic are, if and in what respect ocean trade structures are optimised and how biological invasions can be propagated by it.

1.6 Issues of bird migration

Bird migration, defined as seasonal return migrations of various lengths and patterns, has fascinated man for a very long time (Berthold, 2001a). It has been examined by numerous studies and in various aspects, e.g. its physiological mechanisms, the performance of navigation and the different manners of flight (Hedenström & Ålerstam, 1997; Ålerstam & Hedenström, 1998; Bairlein, 2003; Wiltschko & Wiltschko, 2003). The timing and routes of bird migration have evolved over hundreds and thousands of years. Due to climate change and habitat modification arrival times and migration routes are recently changing more or less rapidly, depending on the adaptability of each migratory species (Jenni & Kéry, 2003). Furthermore, with the danger of a H5N1 influenza pandemic the possibility of disease transmission by migratory birds came into public and scientific focus (Olsen *et al.*, 2006). In the light of these issues it is clear how strongly intertwined bird migration is with human life. Therefore, it is especially important to not only understand single mechanisms of bird migration flight, but also examine its general spatiotemporal patterns and variability under changing environmental conditions.

For over 100 years birds have been ringed and recaptures collected in several ringing centres in Europe and worldwide (Bairlein, 2003). Using this vast amount of data, so called migration atlases presenting recapture positions were compiled for many different species. Thus, one can deduce the whereabouts of migratory species. However, the time aspect is very crude, especially because of the greatly differing recapture rates in different regions of the world (Fiedler *et al.*, 2004). Ring-recapture data are, therefore, well suited for studies when the time aspect does not need to be accounted for explicitly, like e.g. survival and life history studies.

For a few years satellite telemetry and GPS data have become available (Berthold *et al.*, 1997). They are much more homogeneous in time and space and provide many localisations per bird, thus allowing for usage in quantitative analyses. To obtain and apply satellite telemetry transmitters on wild birds is a very demanding and expensive matter. Therefore, usually only a few individuals are tagged. This is in contrast to ring-recapture data that contain information of few positions for a very large number of individuals. This accounts for individual variability in the population rather than providing a sufficient homogeneity of localisations in space and time, as is the case for satellite telemetry and GPS data. The finely resolved satellite telemetry data allow very nicely for mapping migratory routes of a subset of a population's individuals and specifying it in terms of time progression. It is, however, important to account for the sometimes large inaccuracies of satellite telemetry localisations (Kaatz, 1999).

It has been discovered that several migratory bird species perform some kind of stepping

stone migration (Berthold *et al.*, 1992; Kjellén *et al.*, 1997; Beekman *et al.*, 2002; Eichhorn *et al.*, 2006). They journey from staging area to staging area, replenishing their food resources on the way to reach their wintering or breeding regions, respectively. This is in line with the fact that suitable habitat often appears in patches (e.g. Bos *et al.*, 2005). The general dispersion of such species during migration is therefore a discretised one, each stepping stone region being important for the species' viability. Furthermore, at particularly favourable stepping stones or wintering regions high densities of different avian species accumulate. Thus, in these sites probabilities of infection and disease transmission are increased (Olsen *et al.*, 2006), and consideration of them is therefore especially important for the epidemiology of avian diseases.

It is more or less controversially debated if wild migratory birds are able to and spread diseases that are potentially dangerous for humans, like the avian influenza H5N1 (Olsen *et al.*, 2006). Up to now it is only doubtlessly confirmed that H5N1 is globally transmitted and distributed by poultry trade. However, there are other avian diseases of concern, like the West Nile virus (WNV). Possibly transported by a cargo ship WNV has been introduced to Northern America in 1999 where it caused a dramatic outbreak of human encephalitis (Blitvich, 2008). This dreadful incident has been promoted by the fact that the mosquito transmitted WNV has been a novelty to Northern America, thus truly bioinvasive, and native species were not adapted to cope with it. Furthermore, the mosquito species that transmit WNV feed on birds as well as horses and humans. Because it is a hazard to human health the disease is especially problematic. Birds are the major reservoir hosts of WNV, i.e. the virus propagates basically in birds, and viremic migratory birds are widely regarded to be the main vectors of long-distance dispersal (Peterson *et al.*, 2003). The West Nile virus has by now spread through almost all Northern America, and the migratory American crow (*Corvus brachyrhynchos*) is considered one important vector (Komar, 2003). There are more diseases that are spread by long-distance migrating birds (Hubálek, 2004), and it has even been proposed that birds play a role in distributing small organisms, like snails, in their plumage (Gittenberger *et al.*, 2006).

1.7 Outline of the included papers

In this work, two different kinds of transportation systems have been examined and quantified. On the one hand, we were concerned with the global cargo ship traffic, its structure and how it takes part in promoting marine bioinvasion. On the other hand, we looked at avian migration as a more naturally shaped movement structure. From trajectories of birds' motion we wanted to deduce general statements about their quantitative movement patterns. We, thus, aimed to develop a basis for the examination of the dynamical properties of bird migration and eventually evaluate its importance for disease spread. In the different parts of this work, we have quantified both processes and developed network representations of them. Finally, we determined network properties that give first indications of the networks' topologies and dynamics in the light of transportation.

1.7.1 Cargo ship traffic quantified

The first part of this work is intended to quantify and characterise global cargo ship traffic and gain implications for marine bioinvasion. In **paper I** we aimed at developing the global cargo

ship network (GCSN) from real ship movement trajectories. This has not been attempted before (but see Hu & Zhu, 2009). We used detailed data on the whereabouts of a large amount of ships globally that we obtained from Lloyd's Register Fairplay. They are a company that collects and provides a wide range of information for the shipping industry including online freight tracking and ship movement. Using the subset of movements of cargo ships of sizes larger than 10,000 GT (gross tonnage) in 2007 the GCSN was generated. We were interested in analysing the large-scale global movement patterns of cargo ships, its statistical properties and community structure. Additionally, the similarities and differences of the movements of distinct ship types were examined. In relation to a previous study on the quantification of global ship traffic (Drake & Lodge, 2004) we wanted to compare our network with predictions that can be made using the previously developed one. Results were expected to differ considerably, because global trade is not as randomisable as had been assumed.

Using the results from **paper I**, in the following **paper II** the analysis of the global cargo ship network was extended. In that part of this study, we wanted to focus on possible implications of the transportation network characteristics for bioinvasion. A main objective was to apply the concept of networks and state of the art network characteristics, like e.g. the different centrality measures, for bioinvasion research. Such may provide a sufficient first approximation of invasion risk for species with yet poorly understood ecology, especially in the light of the ease of calculating network properties. So, we aimed at explaining possible meanings of network measures for bioinvasion, e.g. that the closeness centrality, i.e. the average shortest path of one node to all others, is an indicator for the invasiveness of the species of a certain region and characterises the GCSN accordingly. Furthermore, we examined the network's robustness to node deletions and estimated a critical per ship transmission probability of global spread. Results were compared to previous estimates of actual bioinvasion transmission.

1.7.2 Bird migration theoretically examined

The second part of this work, **papers III, IV and V**, deals with a theoretical analysis and modelling of bird migration. We were mainly motivated by presently observed effects of climate change on migration routes and timing and the issue of epidemics spread by migratory birds. As a first theoretical, quantitative analysis of bird migration patterns, in **paper III** we analyse different kinds of data with methods of random walk theory. The different available data sources and new devices under development for observing bird movement and displacement are described in detail. The most prominent ones are ring-recapture and satellite telemetry, of which we examined reasonable data sets for the white stork (*Ciconia ciconia*).

When quantifying patterns and rates of spread the statistical properties of the movement of the dispersal vectors are of great importance. Epidemics spread, for example, is much accelerated if movement is not random, but superdiffusive. Therefore, in **paper IV** we examine short time displacement distributions in more detail. We analyse large amounts of ring-recapture data of several migratory and non-migratory bird species, and assess if their movement may be reminiscent of a Lévy flight, as has been proposed for other species. Furthermore, mechanisms and possible sources of bias are discussed, as well as the results' implications for the spread of avian diseases.

As we found two modes of movement in the data of the migratory white stork (**paper III**),

long-range migration and short-range foraging, we consider it reasonable to develop a network model of the process of such migration, disregarding short-range movement. Apart from that, it is very important to incorporate seasonality into the model, as it is the main driver of bird migration. Therefore, network links were weighted by seasonally changing transition rates, the network topology thus changing with time of the year. At the end of **paper III** the idea of such a seasonally driven network model of bird migration is shortly described, the details of which are then presented in **paper V**. Furthermore, we derive how one can parameterise this network model with real data, and perform such a parameterisation (**paper V**) using satellite telemetry data for the white stork and GPS localisations of the greater white-fronted goose (*Anser albifrons*). These two species conduct stepping stone like migration. Therefore, it is suitable to describe their migration as a network process. Once more using network measures we analysed the cumulative and seasonal migration networks of the two considered bird species in the light of transportation. This is sensible, because migratory birds are regarded as vectors for the spread of infectious diseases and other small organisms.

1 General Introduction

2 Paper I.

Regularity and randomness in the global network of cargo ship movements

Pablo Kaluza, Andrea Kölzsch, Michael T. Gastner and Bernd Blasius;
submitted to Proceedings of the National Academy of Sciences of the United States of America

Abstract

Transportation networks play a crucial role in human mobility, the exchange of goods, and the spread of invasive species. With 90% of world trade carried by sea, the global network of merchant ships provides one of the most important modes of transportation. Here we use information about the itineraries of 16,363 cargo ships during the year 2007 to construct a network of links between ports. We show that the network has several features which set it apart from other transportation networks. In particular, most ships can be classified in three categories: bulk dry carriers, container ships, and oil tankers. These three categories do not only differ in the ships' physical characteristics, but also in their mobility patterns and networks. Container ships follow regularly repeating paths whereas bulk dry carriers and oil tankers move apparently randomly between ports. The network of all ship movements possesses a heavy-tailed distribution for the connectivity of ports and for the loads transported on the links with systematic differences between ship types. The data analyzed in this paper improve current assumptions based on gravity models of ship movements, an important step towards understanding patterns of global trade and bioinvasion.

2.1 Introduction

The ability to travel, trade commodities, and share information around the world with unprecedented efficiency is a defining feature of the modern globalized economy. Among the different means of transport, ocean shipping stands out as the most energy efficient mode of long-distance transport for large quantities of goods (Rodrigue *et al.*, 2006). According to estimates, as much as 90% of world trade is hauled by ships (IMO, 2006). In 2006, 7.4 billion tons of goods were loaded at the world's ports. The trade volume currently exceeds 30 trillion ton-miles and is growing at a rate faster than the global economy (UN, 2007).

The worldwide maritime network also plays a crucial role in today's spread of invasive species. Two major pathways for marine bioinvasion are discharged water from ships' ballast tanks (Ruiz *et al.*, 2000) and hull fouling (Drake & Lodge, 2007). Even terrestrial species such as insects are sometimes inadvertently transported in shipping containers (Lounibos, 2002). In several parts

of the world, invasive species have caused dramatic levels of species extinction and landscape alteration, thus damaging ecosystems and creating hazards for human livelihoods, health, and local economies (Mack *et al.*, 2000). The financial loss due to bioinvasion is estimated to be \$120 billion per year in the United States alone (Pimentel *et al.*, 2005).

Despite affecting everybody's daily lives, the shipping industry is far less in the public eye than other sectors of the global transport infrastructure. Accordingly, it has also received little attention in the recent literature on complex networks (Wei *et al.*, 2007; Hu & Zhu, 2009). This neglect is surprising considering the current interest in networks (Albert & Barabási, 2002; Newman, 2003b; Gross & Blasius, 2008), especially airport (Barrat *et al.*, 2004; Hufnagel *et al.*, 2004; Guimera *et al.*, 2005), road (Buhl *et al.*, 2006; Barthelemy & Flammini, 2008) and train networks (Latora & Marchiori, 2002; Sen *et al.*, 2003). In the spirit of current network research, we take here a "holistic" perspective on the global cargo ship network (GCSN) as a complex system defined as the network of ports that are connected by links if ship traffic passes between them.

Similar research in the past had to make strong assumptions about flows on hypothetical networks with connections between all pairs of ports in order to approximate ship movements (Drake & Lodge, 2004; Tatem *et al.*, 2006). By contrast, our analysis is based on comprehensive data of real ship journeys allowing us to construct the actual network. We show that it has a small-world topology where the combined cargo capacity of ships calling at a given port follows a heavy-tailed distribution. This capacity scales superlinearly with the number of directly connected ports. We identify the most central ports in the network and find several groups of highly interconnected ports showing the importance of regional geopolitical and trading blocks.

A high-level description of the complete network, however, does not yet fully capture the network's complexity. Unlike previously studied transportation networks, the GCSN has a multi-layered structure. There are, broadly speaking, three classes of cargo ships – container ships, bulk dry carriers, and oil tankers – that span distinct subnetworks. Ships in different categories tend to call at different ports and travel in distinct patterns. We analyze the trajectories of individual ships in the GCSN and develop techniques to extract quantitative information about characteristic movement types. With these methods we can quantify that container ships sail along more predictable, frequently repeating routes than oil tankers or bulk dry carriers. Understanding the movement patterns in the network can guide future international policy decisions concerning the stability of worldwide trade and reducing the risks of bioinvasion.

2.2 Data

An analysis of global ship movements requires detailed knowledge of ships' arrival and departure times at their ports of call. Such data has become available in recent years. Starting in 2001, ships and ports have begun installing Automatic Identification System (AIS) equipment. AIS transmitters on board of the ships automatically report the arrival and departure times to the port authorities. This technology is primarily used to avoid collisions and increase port security, but arrival and departure records are also made available by Lloyd's Register Fairplay for commercial purposes as part of its Sea-web data base (www.sea-web.com). AIS devices

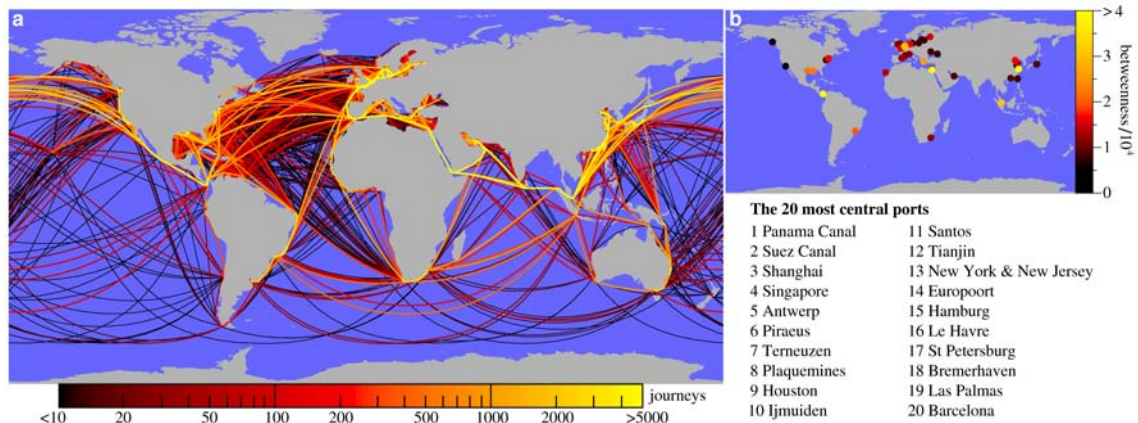


Figure 2.1: Routes, ports, and betweenness centralities in the global cargo ship network (GCSN). (a) The trajectories of all cargo ships bigger than 10,000 GT during 2007. The color scale indicates the number of journeys along each route. Ships are assumed to travel along the shortest (geodesic) paths on water. (b) A map of the 50 ports of highest betweenness centrality and a ranked list of the 20 most central ports.

have not been installed in all ships and ports yet, and therefore there are some gaps in the data. Still, all major ports and the largest ships are included, thus the data base represents the majority of cargo transported on ships.

Our study is based on Sea-web's arrival and departure records in the calendar year 2007 as well as Sea-Web's comprehensive data on the ships' physical characteristics. We restrict our study to cargo ships bigger than 10,000 GT (gross tonnage) which make up 93% of the world's total capacity for ship cargo transport. From these we select all 16,363 ships for which AIS data are available, taken as representative of the global traffic and long-distance trade between the 951 ports equipped with AIS receivers. For each ship we obtain a trajectory from the data base, i.e., a list of ports visited by the ship sorted by date. In 2007, there were 490,517 nonstop journeys linking 36,351 distinct pairs of arrival and departure ports. The complete set of trajectories, each path representing the shortest route at sea and colored by the number of journeys passing through it, is shown in Fig. 2.1 a.

Each trajectory can be interpreted as a small directed network where the nodes are ports linked together if the ship traveled directly between the ports. Larger networks can be defined by merging trajectories of different ships. In this article we aggregate trajectories in four different ways: the combined network of all available trajectories, and the subnetworks of container ships (3,100 ships), bulk dry carriers (5,498) and oil tankers (2,628). These three subnetworks combined cover 74% of the GCSN's total gross tonnage. In all four networks, we assign a weight w_{ij} to the link from port i to j equal to the sum of the available space on all ships that have traveled on the link during 2007 measured in GT. If a ship made the journey from i to j more than once, its capacity contributes multiple times to w_{ij} .

2.3 The global network of cargo ships

The directed network of the entire cargo fleet is noticeably asymmetric, with 59% of all linked pairs of ports being connected only in one direction. Still, the vast majority of ports (935 out of 951) belongs to one single strongly connected component, i.e., for any two ports in this component, there are routes in both directions, though possibly visiting different intermediate ports. The routes are intriguingly short: only few steps in the network are needed to get from one port to another. The shortest path length l between two ports is the minimum number of nonstop connections one must take to travel between origin and destination. In the GCSN, the average over all pairs of ports is extremely small, $\langle l \rangle = 2.5$. Even the maximum shortest path between any two ports (e.g., from Skagway, Alaska, to the small Italian island of Lampedusa), is only of length $l_{\max} = 8$. In fact, the majority of all possible origin-destination pairs (52%) can already be connected by two steps or less.

Comparing these findings to those reported for the worldwide airport network (WAN) shows interesting differences and similarities. In the WAN, the average and maximum shortest path lengths are $\langle l \rangle = 4.4$ and $l_{\max} = 15$ respectively (Guimera *et al.*, 2005), i.e., about twice as long as in the GCSN. Similar to the WAN, the GCSN is highly clustered: if a port X is linked to ports Y and Z , there is a high probability that there is also a connection from Y to Z . We calculated a clustering coefficient C (Watts & Strogatz, 1998) for directed networks and find $C = 0.49$ whereas random networks with the same number of nodes and links only yield $C = 0.04$ on average. Therefore, the GCSN – like the WAN – can be regarded as a small-world network possessing short path lengths despite substantial clustering (Watts & Strogatz, 1998). However, the average degree of the GCSN, i.e., the average number of links arriving at and departing from a given port, $\langle k \rangle = 76.5$, is notably higher than in the WAN where $\langle k \rangle = 19.4$ (Barrat *et al.*, 2004).

The degree distribution $P(k)$ shows that most ports have few connections, but there are some ports linked to hundreds of other ports (Fig. 2.2 a). Similar right-skewed degree distributions have been observed in many real-world networks (Barabási & Albert, 1999). While the GCSN's degree distribution is not exactly scale-free, the distribution of link weights, $P(w)$, follows approximately a power law $P(w) \propto w^{-\mu}$ with $\mu = 1.71 \pm 0.14$ ($R^2 = 0.986$, Fig. 2.2 b). By averaging the sums of the link weights arriving at and departing from port i , we obtain the node strength s_i (Barrat *et al.*, 2004). The strength distribution is also approximated by a power law $P(s) \propto s^{-\eta}$ with $\eta = 1.02 \pm 0.17$ ($R^2 = 0.945$), meaning that a small number of ports handle huge amounts of cargo (Fig. 2.2 c). Model selection by Akaike weights (Burnham & Anderson, 1998) confirm that power law is a better fit than exponential for $P(w)$ and $P(s)$, but not $P(k)$ (see Supplementary Information). Strengths and degrees of the ports are related according to the scaling relation $\langle s(k) \rangle \propto k^{1.46 \pm 0.1}$ (95% CI for SMA regression; Warton *et al.*, 2006). Hence, the strength of a port grows generally faster than its degree (Fig. 2.2 d). In other words, highly connected ports not only have many links, but their links also have a higher than average weight. This observation agrees with the fact that busy ports are better equipped to handle large ships with large amounts of cargo. A similar result, $\langle s(k) \rangle \propto k^{1.5 \pm 0.1}$, was found for airports (Barrat *et al.*, 2004), which may hint at a general pattern in transportation networks.

A further indication of the importance of a node is its betweenness centrality (Freeman,

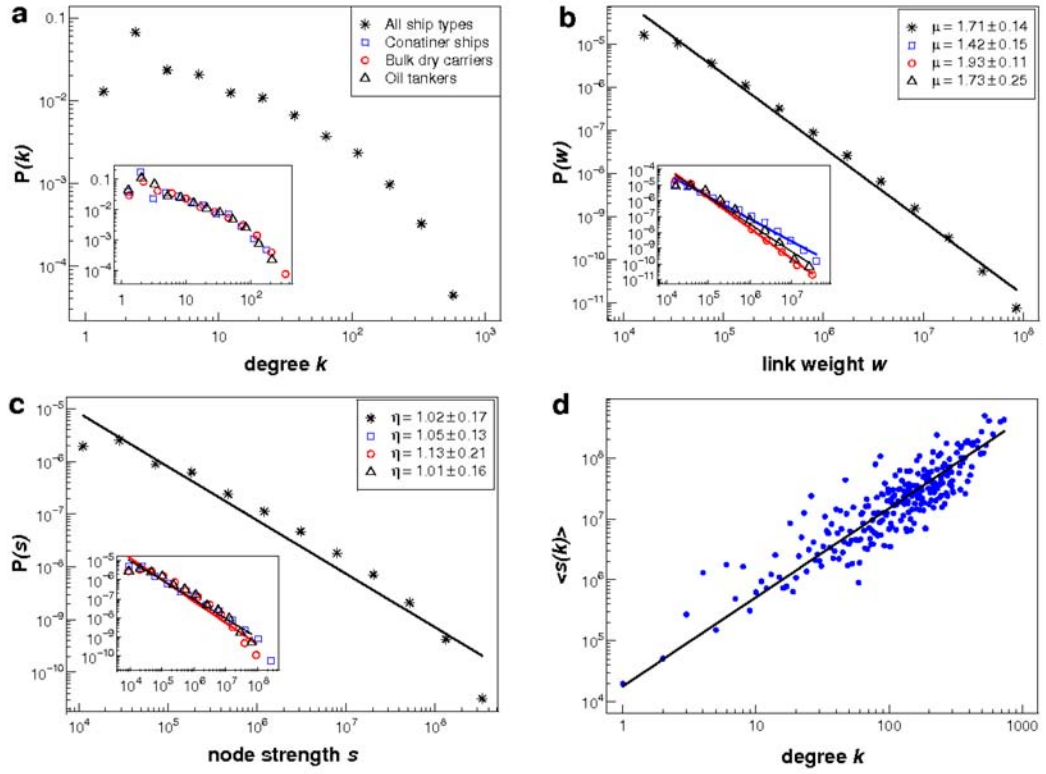


Figure 2.2: Degrees and weights in the global cargo ship network * (insets: subnetworks for container ships \square , bulk dry carriers \circ , and oil tankers \triangle). (a) The degree distributions $P(k)$ are right-skewed, but not power laws, neither for the GCSN nor its subnetworks. The degree k is defined here as the sum of in- and out-degree, thus $k = 1$ is rather rare. (b) The link weight distributions $P(w)$ reveal clear power law relationships for the GCSN and the three subnetworks, with exponents μ characteristic for the movement patterns of the different ship types. (c) The node strength distributions $P(s)$ are also heavy-tailed, showing power law relationships. The stated exponents are calculated by linear regression with 95% confidence intervals (similar results are obtained with maximum likelihood estimates, see Supplementary Information). (d) The average strength of a node $\langle s(k) \rangle$ scales superlinearly with its degree, $\langle s(k) \rangle \propto k^{1.46 \pm 0.1}$, indicating that highly connected ports have, on average, links of higher weight.

1979; Newman, 2004). The betweenness of a port is the number of topologically shortest directed paths in the network that pass through this port. In Fig. 2.1 b we plot and list the most central ports. There are different reasons why these nodes play a special role in the network. The Panama and Suez Canal are shortcuts to avoid long passages around South America and Africa. Other ports have a high centrality because they are visited by a large number of ships (e.g., Shanghai) whereas others gain their status primarily by being connected to many different ports (e.g., Antwerp).

2.4 The network layers of different ship types

To compare the movements of cargo ships of different types, separate networks were generated for each of the three main ship types: container ships, bulk dry carriers, and oil tankers. Applying the network parameters introduced in the previous section to these three subnetworks reveals some broad-scale differences (see Table 2.1). The network of container ships is densely clustered, $C = 0.52$, has a rather low mean degree, $\langle k \rangle = 32.44$, and a large mean number of journeys (i.e., number of times any ship passes) per link, $\langle J \rangle = 24.26$. The bulk dry carrier network, on the other hand, is less clustered, has a higher mean degree, and fewer journeys per link ($C = 0.43$, $\langle k \rangle = 44.61$, $\langle J \rangle = 4.65$). For the oil tankers, we find intermediate values ($C = 0.44$, $\langle k \rangle = 33.32$, $\langle J \rangle = 5.07$). Note that the mean degrees $\langle k \rangle$ of the subnetworks are substantially smaller than that of the full GCSN, indicating that different ship types use essentially the same ports but different connections.

A similar tendency appears in the scaling of the link weight distributions (Fig. 2.2 b). $P(w)$ can be approximated as power laws for each network, but with different exponents μ . The container ships have the smallest exponent ($\mu = 1.42$) and bulk dry carriers the largest ($\mu = 1.93$) with oil tankers in between ($\mu = 1.73$). In contrast the exponents for the distribution of node strength $P(s)$ are nearly identical in all three subnetworks, $\eta = 1.05$, $\eta = 1.13$ and $\eta = 1.01$, respectively.

These numbers give a first indication that different ship types move in distinctive patterns. Container ships typically follow set schedules visiting several ports in a fixed sequence along their way, thus providing regular services. Bulk dry carriers, by contrast, appear rather random as they frequently change their routes on short notice depending on the current supply and demand of the goods they carry. The larger variety of origins and destinations in the bulk dry carrier network ($n = 616$ ports, compared to $n = 378$ for container ships) explains the higher average degree and the smaller number of journeys for a given link. Oil tankers also follow short-term market trends, but, because they can only load oil and oil products, the number of possible destinations ($n = 505$) is more limited than for bulk dry carriers.

The betweenness centralities for the three network layers also underline their differences (see Supplementary Information). While some ports rank highly in all categories (e.g. Suez Canal, Shanghai), others are specialized on certain ship types. For example, the German port of Wilhelmshaven ranks tenth in terms of its world-wide betweenness for oil tankers, but is only 241st for bulk dry carriers and 324th for container ships.

We can gain further insight into the roles of the ports by examining their community structure. Communities are groups of ports with many links within the groups but few links between different groups. We calculated these communities for the three subnetworks with a modularity optimization method for directed networks (Leicht & Newman, 2008). The network of container trade shows 12 communities (Fig. 2.3 a). The largest ones are located (1) on the Arabian, Asian, and South African coasts, (2) on the North American east coast and in the Caribbean, (3) in the Mediterranean, the Black Sea, and on the European west coast, (4) in Northern Europe, and (5) in the Far East and on the American west coast. The transport of bulk dry goods reveals 7 groups (Fig. 2.3 b). Some can be interpreted as geographic entities (e.g., North American east coast, trans-Pacific trade) while others are dispersed on multiple continents. Especially interesting is the community structure of the oil transportation network

Ship class	ships	MGT	n	$\langle k \rangle$	C	$\langle l \rangle$	$\langle J \rangle$
Whole fleet	16363	664.7	951	76.4	0.49	2.5	13.57
Container ships	3100	116.8	378	32.4	0.52	2.76	24.25
Bulk dry carriers	5498	196.8	616	44.6	0.43	2.57	4.65
Oil tankers	2628	178.4	505	33.3	0.44	2.74	5.07

Ship class	μ	η	$\langle N \rangle$	$\langle L \rangle$	$\langle S \rangle$	$\langle p \rangle$
Whole fleet	1.71	1.02	10.4	15.6	31.8	0.63
Container ships	1.42	1.05	11.2	21.2	48.9	1.84
Bulk dry carriers	1.93	1.13	8.9	10.4	12.2	0.03
Oil tankers	1.73	1.01	9.2	12.9	17.7	0.19

Number of ships, total gross tonnage [10^6 GT] and number of ports n in each subnetwork; together with network characteristics: mean degree $\langle k \rangle$, clustering coefficient C , mean shortest path length $\langle l \rangle$, mean journeys per link $\langle J \rangle$, power-law exponents μ and η ; and trajectory properties: average number of distinct ports $\langle N \rangle$, links $\langle L \rangle$, port calls $\langle S \rangle$ per ship and regularity index $\langle p \rangle$.

Table 2.1: Characterization of different subnetworks.

which shows 6 groups (Fig. 2.3 c): (1) the European, north and west African market (2) a large community comprising Asia, South Africa and Australia, (3) three groups for the Atlantic market with trade between Venezuela, the Gulf of Mexico, the American east coast and Northern Europe, and (4) the American Pacific Coast. It should be noted that the network includes the transport of crude oil as well as commerce with already refined oil products so that oil producing regions do not appear as separate communities.

Despite the differences between the three main cargo fleets, there is one unifying feature: their motif distribution (Milo *et al.*, 2002). Like most previous studies, we focus here on the occurrence of three-node motifs and present their normalized Z score, a measure for their abundance in a network (Fig. 2.4). Strikingly, the three fleets have practically the same motif distribution. In fact, the Z scores closely resemble those found in the World Wide Web and different social networks which were conjectured to form a superfamily of networks (Milo *et al.*, 2004). This superfamily displays many transitive triplet interactions (i.e., if $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$); for example, the overrepresented motif 13 in Fig. 2.4, has six such interactions. Intransitive motifs, like motif 6, are comparably infrequent. The abundance of transitive interactions in the ship networks indicates that cargo can be transported both directly between ports as well as via several intermediate ports. It remains to be seen whether the transitivity in the cargo ship network is caused by mechanisms similar to those at work in other networks in this superfamily.

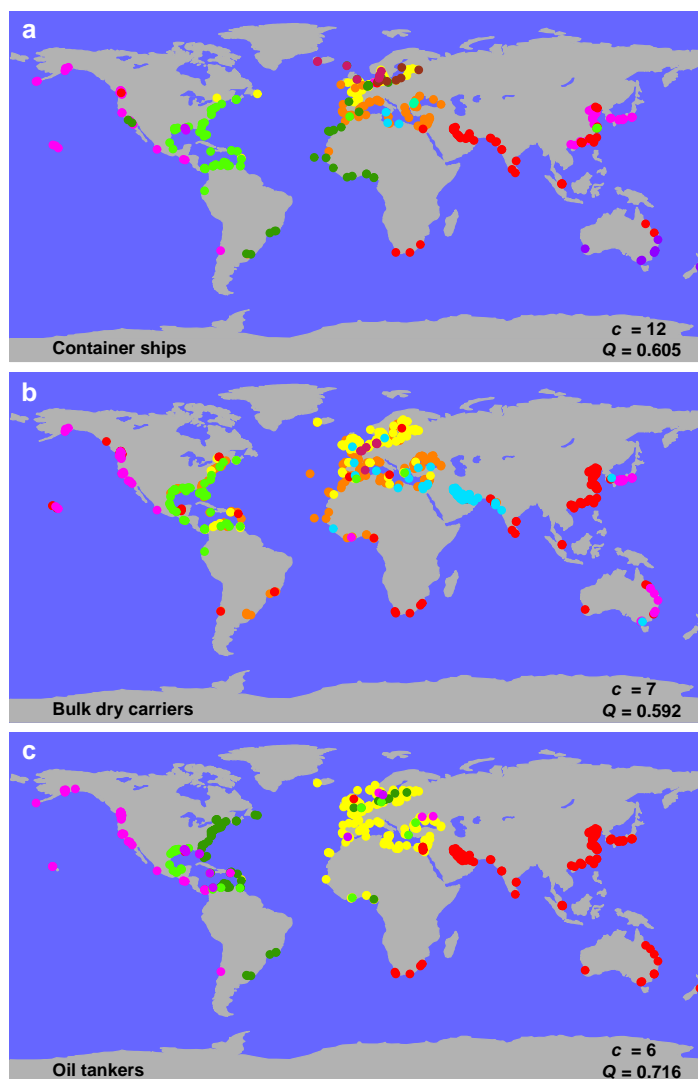


Figure 2.3: Communities of ports in three cargo ship subnetworks. The communities are groups of ports that maximize the number of links within the groups, as opposed to between the groups, in terms of the modularity Q (Leicht & Newman, 2008). In each map, the colors represent the c distinct trading communities for the goods transported by (a) container ships, (b) bulk dry carriers, and (c) oil tankers. The optimal values for c and Q are stated in the lower right corners.

2.5 Network trajectories

Going beyond the network perspective, the data base also provides information about the movement characteristics per individual ship (Table 2.1). The average number of distinct ports per ship $\langle N \rangle$ do not differ much between different ship classes, but container ships call much more frequently at ports than bulk dry carriers and oil tankers. This difference is

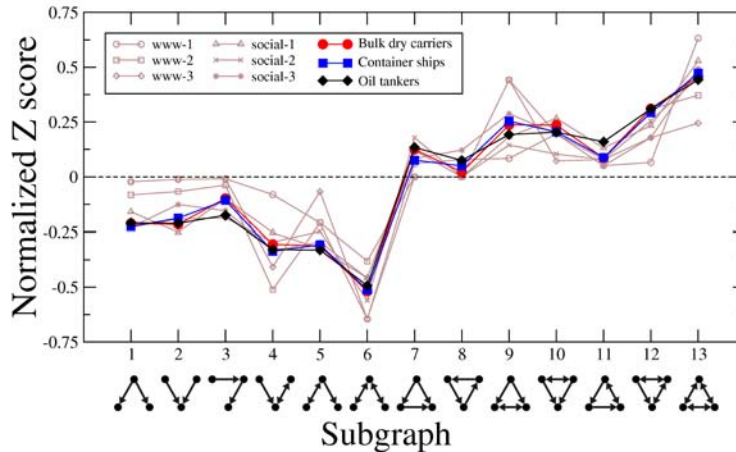


Figure 2.4: Motif distributions of the three main cargo fleets. A positive (negative) normalized Z score indicates that a motif is more (less) frequent in the real network than in random networks with the same degree sequence. For comparison, we overlay the Z scores of the World Wide Web and social networks. The agreement suggests that the ship networks fall in the same superfamily of networks (Milo *et al.*, 2004). The motif distributions of the fleets are maintained even when 25%, 50% and 75% of the weakest connections are removed.

explained by the characteristics and operational mode of these ships. Normally, container ships are fast (between 20 and 25 knots) and spend less time (1.9 days on average in our data) in the port for cargo operations. By contrast, bulk dry carriers and oil tankers move more slowly (between 13 and 17 knots) and stay longer in the ports (on average 5.6 days for bulk dry carriers, 4.6 days for oil tankers).

The speed at sea and of cargo handling, however, is not the only operational difference. The topology of the trajectories also differs substantially. Characteristic sample trajectories for each ship type are presented in Fig. 2.5 a-c. The container ship (Fig. 2.5 a) travels on some of the links several times during the study period whereas the bulk dry carrier (Fig. 2.5 b) passes almost every link exactly once. The oil tanker (Fig. 2.5 c) commutes a few times between some ports, but by and large also serves most links only once.

We can express these trends in terms of a “regularity index” p that quantifies how much the frequency with which each link is used deviates from a random network. Consider the trajectory of a ship calling S times at N distinct ports and travelling on L distinct links. We compare the mean number of journeys per link $f_{real} = S/L$ to the average link usage f_{ran} in an ensemble of randomized trajectories with the same number of nodes N and port calls S . To quantify the difference between real and random trajectories we calculate the Z score $p = (f_{real} - f_{ran})/\sigma$ (where σ is the standard deviation of f in the random ensemble). If $p = 0$, the real trajectory is indistinguishable from a random walk. The larger p , the more regular is the movement of the ship. Figs. 2.5 d-f present the distributions of the regularity index p for the different fleets. For container ships it is distributed broadly over strongly positive values of p , thus supporting our earlier observation that most container ships provide regular services

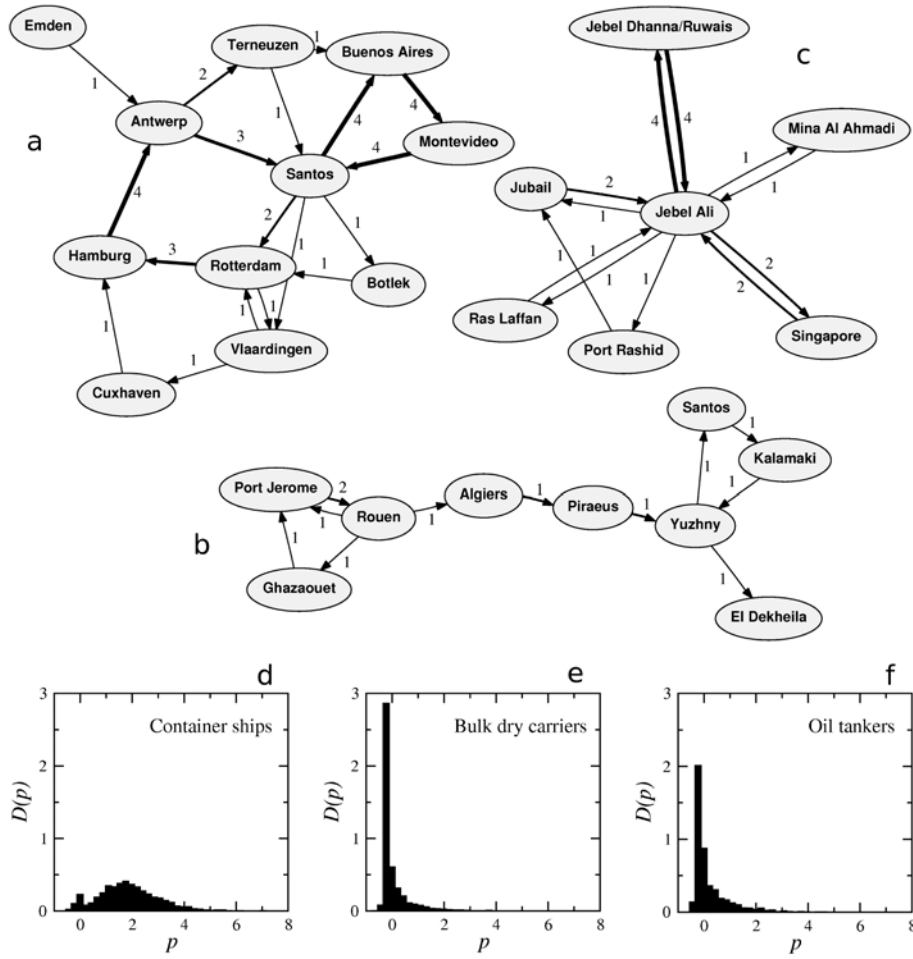


Figure 2.5: Sample trajectories of (a) a container ship with a regularity index $p = 2.09$, (b) a bulk dry carrier, $p = 0.098$, (c) an oil tanker, $p = 1.027$. In the three trajectories, numbers and thickness of drawn links indicate the frequency of journeys on each link. (d)-(f) Distribution of p for the three main fleets.

between ports along their way. Trajectories of bulk dry carriers and oil tankers, on the other hand, appear essentially random with the vast majority of ships near $p = 0$.

2.6 Discussion

In this article, we view global ship movements as a network based on detailed arrival and departure records. Until recently, surveys of seaborne trade had to rely on far less data: only the total number of arrivals at some major ports were publicly accessible, but not the ships' actual paths (Zachcial & Heideloff, 2001). Missing information about the frequency

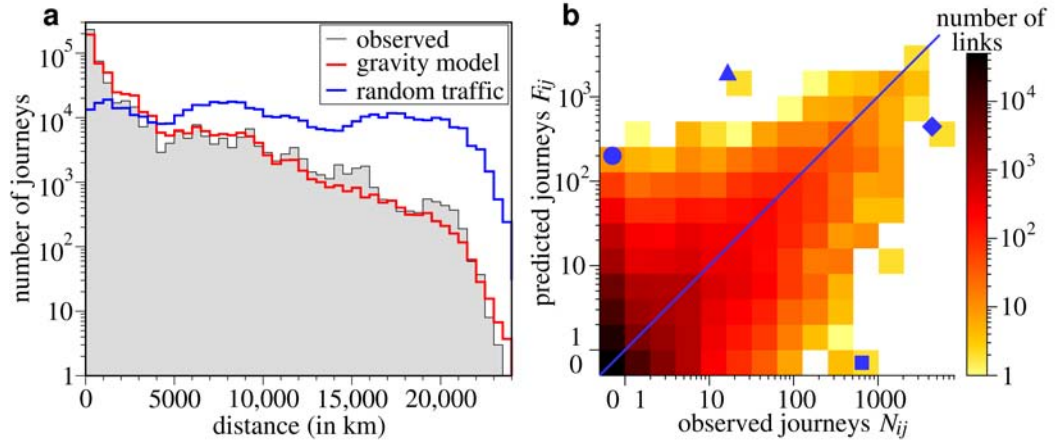


Figure 2.6: (a) Histogram of port-to-port distances travelled in the GCSN (navigable distances around continents as indicated in Fig. 2.1). We overlay the predictions of two different models. The gravity model (red), based on information about distances between ports and total port calls, gives a much better fit than a simpler model (blue) which only fixes the total number of journeys. (b) Count of port pairs with N_{ij} observed and F_{ij} predicted journeys. The flows F_{ij} were calculated with the gravity model (rounded to the nearest integer). Some of the worst outliers are highlighted in blue. \circ : Antwerp to Calais ($N_{ij} = 0$ vs. $F_{ij} = 200$). \triangle : Hook of Holland to Europoort (16 vs. 1895). \diamond : Calais to Dover (4392 vs. 443). \square : Harwich to Hook of Holland (644 vs. 0).

of journeys, thus, had to be replaced by plausible assumptions, the gravity model being the most popular choice. It posits that trips are, in general, more likely between nearby ports than between ports far apart. If d_{ij} is the distance between ports i and j , the decline in mutual interaction is expressed in terms of a distance deterrence function $f(d_{ij})$. The number of journeys from i to j then takes the form $F_{ij} = a_i b_j O_i I_j f(d_{ij})$, where O_i is the total number of departures from port i and I_j the number of arrivals at j (Haynes & Fotheringham, 1984). The coefficients a_i and b_j are needed to ensure $\sum_j F_{ij} = O_i$ and $\sum_i F_{ij} = I_j$.

How well can the gravity model approximate real ship traffic? We choose a truncated power law for the deterrence function, $f(d_{ij}) = d_{ij}^{-\beta} \exp(-d_{ij}/\kappa)$. The strongest correlation between model and data is obtained for $\beta = 0.59$ and $\kappa = 4900$ km (see Supplementary Information). At first sight, the agreement between data and model appears indeed impressive. The predicted distribution of travelled distances (Fig. 2.6 a) fits the data far better than a simpler non-spatial model that preserves the total number of journeys, but assumes completely random origins and destinations.

A closer look at the gravity model, however, reveals its limitations. In Fig. 2.6 b we count how often links with an observed number of journeys N_{ij} are predicted to be passed F_{ij} times. Ideally all data points would align along the diagonal $F_{ij} = N_{ij}$, but we find that the data are substantially scattered. Although the parameters β and κ were chosen to minimize the scatter, the correlation between data and model is only moderate (Kendall's $\tau = 0.433$). In some cases, the prediction is off by several thousand journeys per year.

In summary, the gravity model captures some broad trends of global cargo trade, but for

many applications its results are too crude. Recent studies have used the gravity model to pinpoint the ports and routes central to the spread of invasive species (Drake & Lodge, 2004; Tatem *et al.*, 2006). However, the actual movements of ships are more complex than anticipated. Future strategies to curb biological invasions will have to take these details into account. The network structure presented in this article can be used as a first step in this direction.

Acknowledgements

We thank B. Volk, K.H. Holocher, A. Wilkinson, J.M. Drake and H. Rosenthal for stimulating discussions and helpful suggestions. We also thank Lloyd's Register Fairplay for providing their shipping data base. This work was supported by German VW-Stiftung.

2.7 Supplementary Information

2.7.1 Network analysis

Network construction

Construction of the GCSN is based on the Sea-web data base (www.sea-web.com), containing ship arrival and departure records, as well as data on ships' physical characteristics, in the calendar year 2007. From the 58,056 ships in the data base we select the 24,375 ships bigger than 10,000 GT, which comprises 42% of all ships in the data base and more than 90% of the total cargo capacity (measure in dead weight tonnage DWT). From these we further select all 16,363 ships for which AIS are available in 2007 (comprising 28% of all ships in the data base and about 58% of the world's total cargo capacity).

Each ship trajectory consists of a list of visited locations, which can be ports, port terminals or anchorage areas. Thus, a specific port will usually be listed with different references, since a ship can visit several of the ports' terminals. As we are interested in the interaction between ports, different terminals of the same port are uniquely referred to by the port name (ports have been adjusted using the portguide database www.portguide.com) so that each port is represented by a single node in the network (technically speaking, we performed vertex contractions among the terminals until we obtain only one node). Anchorage points are removed from the list of visited places as they do not provide new information about the interaction of ships between ports.

After this preprocessing of the data base, each trajectory of a ship consists of a list of S port calls during the calendar year 2007 sorted by date. For each such trajectory we generate an associated small network with N nodes and L links, where each visited port is represented by a node and a directed link connects two visited ports according to the ship movement. As a ship, in general, can travel the same link many times during the considered period of time, the number of distinct links can be smaller than the number of port calls, $L \leq S$. The links in a trajectory are directed and can be weighted either by the number J of times the ship sailed between the same ports during the study period, or by the cumulative gross tonnage (GT) that crossed the link in the considered time window, i.e. J multiplied by the ship's gross tonnage.

Networks, corresponding to larger ensembles of vessels, are generated by merging all single ship networks for a particular class of fleet. To merge an ensemble of subnetworks we generate a new network, comprising all ports and links in the subnetworks. If a link is present in more than one of the subnetworks, its weight in the merged network is the sum of all weights of this link in the subnetworks.

Clustering coefficient

To determine a clustering coefficient in a directed network we calculate the clustering coefficient c_i of a node i as

$$c_i = \frac{E_i}{k_i(k_i - 1)}, \quad (2.1)$$

where k_i is the number of neighbors of the node i , and E_i is the number of directed connections that exist between the k_i neighbors. Two nodes are considered to be neighbors if there exists

at least one directed connection between them (Kaluza, 2007). The clustering coefficient C for the whole network is obtained by averaging c_i over all nodes of the network

$$C = \frac{1}{n} \sum_i c_i. \quad (2.2)$$

Motif structure

To determine motif distributions of networks, the software package MFINDER has been used (www.weizmann.ac.il/mcb/UriAlon).

Regularity index

We define a “regularity index” p to characterize the regularity or mode of operation of a ship’s trajectory, i.e. to measure to which extent the ship is operating periodically on some prescribed service route rather than moving randomly between the ports. Consider a ship calling S times at N distinct ports and traveling on L distinct links. The regularity index is based on the average frequency f_{real} with which the ship crosses the links during the considered period of time

$$f_{real} = \frac{S}{L} = \frac{1}{L} \sum_{i=1}^L f_i, \quad (2.3)$$

where f_i is the number of journeys across link i . Given a fixed number of port calls S , high values of f_{real} arise if a ship is crossing the same links many times, hinting at a regular service route. In contrast, a small value means that the ship crosses many links only a few times - an indication of a rather random mode of operation.

The measure f_{real} is still ambiguous as it is affected by the number of steps that the ship has taken. Therefore we compare the mean number of journeys of the real trajectory, f_{real} , to the link usage f_{ran} of an ensemble of randomized trajectories with the same number of nodes N and port calls S . We construct a random trajectory by starting at an arbitrary port and choosing randomly between the other possible ports in order to make a step. This process is repeated S times, creating a contiguous trajectory (each new link starts where the previous one ended). In this way we create an ensemble with M randomized trajectories and define f_{ran} as the ensemble average

$$f_{ran} = \frac{1}{M} \sum_{j=1}^M \frac{1}{L_j} \sum_{i=1}^{L_j} f_{ji}, \quad (2.4)$$

where L_j is the number of distinct links used by the randomized trajectory j (note that in general L_j will be different from L) and f_{ji} describes the number of journeys across link i in trajectory j . For normalization one can use the standard variation σ_{ran} which indicates the variation of the frequencies f_{ji} .

To quantify the difference between real and random trajectories, we calculate the Z score, i.e., the difference between the two mean frequencies, taking their distance in units of standard

Port	Mbc	ships	port calls	k	s (mill. GT)
1 Panama Canal	74.55	3224	12719	733	422.4
2 Suez Canal	45.27	3516	7632	686	406.1
3 Shanghai	39.46	4156	12882	564	404.8
4 Singapore	31.06	3190	11875	524	489.5
5 Antwerp	30.71	2268	7447	603	245.6
6 Piraeus	25.59	916	2535	358	83.7
7 Terneuzen	23.14	1877	5218	534	161.3
8 Plaquemines	22.81	2166	8449	496	269.7
9 Houston	21.87	1629	3504	487	103.7
10 Ijmuiden	20.48	1057	3879	487	129.7
11 Santos	19.90	1490	3826	513	119.2
12 Tianjin	16.73	2450	5152	415	169.4
13 New York & New Jersey	16.68	1510	4495	462	189.7
14 Europoort	16.44	1148	4080	464	147.3
15 Hamburg	15.79	1287	4015	430	176.6
16 Le Havre	15.07	1102	3865	422	163.1
17 St Petersburg	14.88	472	1506	343	30.0
18 Bremerhaven	14.02	1025	4004	353	156.4
19 Las Palmas	13.42	764	2008	408	54.2
20 Barcelona	12.71	1015	5043	376	177.5

Table 2.2: The 20 most central ports in the GCSN, their betweenness centrality in units of 1000 (Mbc), the number of different ships that pass this port per year, number of times this port is called by all ships, degree (k), and strength (s).

deviations

$$p = \frac{f_{real} - f_{ran}}{\sigma_{ran}}. \quad (2.5)$$

The index p is a measure for the regularity of the trajectory. If p is close to zero the trajectory cannot be distinguished from a random walk among the ports, whereas with larger values of p the movement of the ship is increasingly regular.

2.7.2 Betweenness centralities in comparison

Characterization of the full GCSN

In the main text we have listed the most important ports of the GCSN according to their betweenness centralities. In Table 2.2 we additionally list their betweenness values and related measures of port size and importance. The table shows that ports that handle many ships or have high degree, in general also have a high betweenness centrality. On the other hand, a port may have a rather small node strength s and still be very central (e.g. Piraeus). One possible explanation is that such a port has many weak links that function as local hubs.

rank	Container ships		Bulk dry carriers		Oil tankers	
	Port	Mbc	Port	Mbc	Port	Mbc
1	Shanghai	11.76	Panama Canal	40.00	Singapore	30.09
2	Vlaardingen	11.76	Shanghai	29.71	Panama Canal	20.33
3	Panama Canal	10.95	Suez Canal	23.83	Europoort	18.41
4	Singapore	9.94	Plaquemines	23.51	New York/New Jersey	13.98
5	Jebel Ali	8.72	Santos	18.30	Suez Canal	13.85
6	Algeciras	6.97	Tianjin	12.82	Houston	11.31
7	Le Havre	6.63	Durban	11.51	Shanghai	11.18
8	Barcelona	6.13	Terneuzen	11.38	Ijmuiden	10.18
9	Bremerhaven	5.83	St Petersburg	11.19	New Orleans	8.80
10	Hamburg	5.72	Lianyungang	9.47	Wilhelmshaven	8.63
11	Tacoma	5.44	Qingdao	9.36	Fos	8.22
12	Malaga	5.29	Houston	9.25	Ventspils	7.00
13	Antwerp	4.83	Las Palmas	9.21	Maasvlakte	6.81
14	Suez Canal	4.73	Antwerp	8.77	Corpus Christi	6.11
15	Santos	4.16	Guangzhou	8.37	Jebel Ali	5.91
16	Rotterdam	3.99	Riga	7.03	Chiba	5.67
17	Piraeus	3.99	Yuzhny	6.90	Benicia	5.65
18	Busan	3.64	Novorossiysk	6.88	Jebel Dhanna/Ruwais	5.41
19	Felixstowe	3.59	Kaohsiung	6.63	Antwerp	5.38
20	Colombo	3.22	Ijmuiden	6.61	Long Beach	4.50

Table 2.3: The 20 most central ports in the different network layers of different ship types and their betweenness centrality in units of 1000 (Mbc).

Different ship types

Other interesting aspects regarding the betweenness centralities become evident by comparing these values for the three ship type specific networks (see Table 2.3). This allows for example to deduce which of the ports are more important for oil trade, which for bulk dry and which for container goods. Interestingly the importance of the two main canals, the Suez Canal and Panama Canal, shows much variation between the ship types. This is plausible because, for example for container ships, there are, in comparison, less direct links between the Suez Canal and other ports, as they stop at several, usually coinciding, ports on their route.

2.7.3 Distances traveled in the global network of cargo ships

Calculating ship distances

Calculating distances between two ports is not a trivial matter. The spherical shape of the earth complicates some of the equations: the shortest path between two points on a sphere is not a line, but a “geodesic”, a segment of a great-circle arc. To make matters worse, ships cannot always travel along the geodesic if it is blocked by land (i.e., a continent or an island) or ice. To calculate effective distances, one has to consider all obstacles and find the shortest

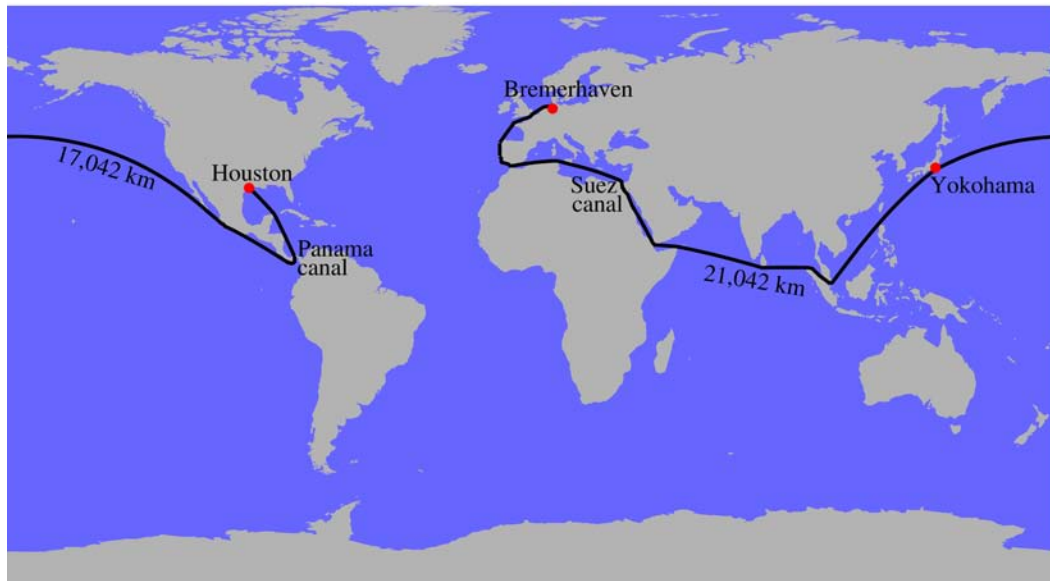


Figure 2.7: Examples of distances between ports (Bremerhaven \leftrightarrow Yokohama \leftrightarrow Houston). Ships are assumed to move around the continents, with the Suez and Panama canals as the only exceptions.

way around them.

As a first step in this calculation, we rasterized the world's coast lines. The coordinates along the coasts were treated as vertices of a so-called visibility graph: pairs of vertices were connected by edges if and only if the geodesic between the vertices is entirely on water. (We assume for simplicity's sake that the northernmost latitude free of sea ice is 69° N and the southernmost latitude 62° S.) To determine the edges, we parametrized the geodesic, stepped along the curve in small discrete steps, and determined at each step if the corresponding point on the geodesic is on water. Each edge was assigned the great-circle distance, $R \arccos(\cos \phi_1 \cos \phi_2 \cos(\lambda_1 - \lambda_2) + \sin \phi_1 \sin \phi_2)$, where R is the earth's radius (6,371 km), and $\lambda_{1(2)}$ and $\phi_{1(2)}$ are the longitude and latitude of the vertices at the end points. Then the shortest paths between ports were determined with Dijkstra's algorithm. Undoubtedly, more elegant algorithms exist (Mitchell, 1991), but this technique was sufficiently reliable for our purpose.

Examples of calculated ship routes are shown in Fig. 2.7. Note that we permit transit through the Panama and Suez canals, a reasonable assumption for all but the very largest ships. A list of pairwise port distances is available upon request from the authors.

Characterization by gravity models

An important question in transportation forecasting is how distances influence the number of trips between ports. Several classic studies in economics have proposed "gravity models" where the frequency of trips (Zipf, 1946) or the volume of trade (Reilly, 1931; Isard, 1954) decays as a function of distance. Following along these lines, we use the doubly constrained

gravity model (Haynes & Fotheringham, 1984) to fit the observed distribution of ship traffic. The number of ships per year from port i to j is treated as a flow F_{ij} on the network of all possible port connections and is assumed to have the form

$$F_{ij} = a_i b_j O_i I_j f(d_{ij}) \quad (2.6)$$

where O_i is the total flow out of port i and I_j the total flow into port j . The distance deterrence function $f(d_{ij})$ describes the level of interaction if the distance between the ports is d_{ij} . The coefficients a_i and b_j have to be chosen such that the calculated flows F_{ij} are consistent with the actually observed total in- and out-flows, thus

$$a_i = \left(\sum_j b_j I_j f(d_{ij}) \right)^{-1}, \quad (2.7)$$

$$b_j = \left(\sum_i a_i O_i f(d_{ij}) \right)^{-1}. \quad (2.8)$$

This system of non-linear equations can be solved by iteratively putting an approximate solution for all a_i into Eq. 2.8 and the new solution for all b_j back into Eq. 2.7 until all coefficients converge.

The deterrence function is assumed to follow a truncated power law

$$f(d_{ij}) = d_{ij}^{-\beta} \exp(-d_{ij}/\kappa). \quad (2.9)$$

The quality of the model for a given exponent β and cutoff distance κ is assessed in terms of Kendall's rank correlation between the calculated flow F_{ij} and the actual number of ships between i and j in 2007. F_{ij} was rounded to the nearest integer; all pairs of calculated and observed flows were cross-tabulated in exponentially increasing bins (cf. Fig. 6b in the main text). The strongest correlation is obtained for $\beta = 0.59$ and $\kappa = 4900$ km (Fig. 2.8).

A histogram of the number of journeys binned by distance in 500 km intervals shows indeed striking similarities between real data and the gravity model for these parameters (Fig. 2.6 a in the main text). In particular, the gravity model performs considerably better than a null model that preserves the total number of journeys, but with completely random movements between ports. However, if journeys over similar distances are not aggregated, even the gravity model possesses only limited predictive power. In Fig. 2.6 b in the main text, we compare the observed and predicted flow on each link. While ideally all data points would align along the diagonal, there is substantial scatter on both sides of it. For example, on the link from Antwerp (Belgium) to Calais (France) the gravity law predicts 200 journeys whereas none are observed. Conversely, zero journeys are predicted between Harwich (UK) and Hook of Holland (Netherlands), but 644 were actually recorded. The rank correlation is, accordingly, even at its optimum value only moderate ($\tau = 0.433$). Therefore, gravity-model based proposals to limit bioinvasion mediated by ships should be interpreted with caution (Drake & Lodge, 2004; Tatem *et al.*, 2006).

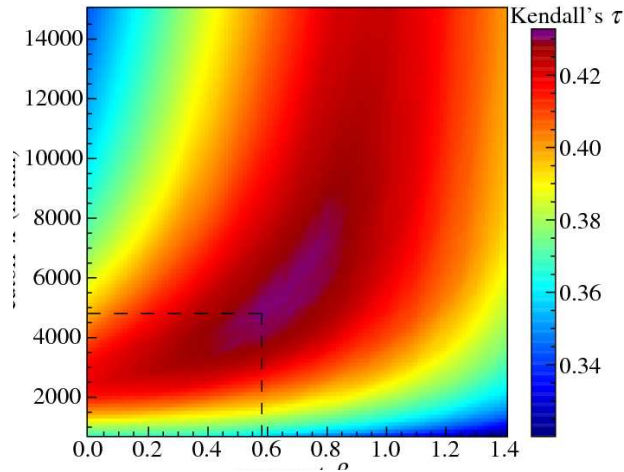


Figure 2.8: Kendall's τ as a function of the exponent β and cutoff distance κ in the distance deterrence function. The dashed lines indicate the parameters at maximum τ .

2.7.4 Model selection for degree and weight distributions

One central aim for the characterization of broad scale distributions of e.g., networks degrees is to quantify whether such distributions are consistent with a power law or whether different distributions (such as exponential or log-normal) are more appropriate. The classic approach to identify which distribution best fits a given set of data is to simply assess the quality of a linear regression in a log-log plot of the logarithmically binned data. If it is a sufficiently close fit, it is often concluded that the distribution coincides with a power law $P(x) = x^{-\gamma}$, and that the slope of the regression curve is an appropriate estimate of the exponent γ . This approach has, however, recently been shown to often wrongly favour power laws and an alternative methodology has been developed and introduced to power law fitting (Burnham & Anderson, 1998; Edwards *et al.*, 2007): model selection by Akaike weights and nonparametric estimation of the model parameters by maximum likelihood.

To strengthen our results concerning the forms of the distributions of the degree $P(k)$, link weight $P(w)$ and node strength $P(s)$ (see Fig. 2.2 in the main text) we calculated the Akaike weights w_i for these distributions in the GCSN and each of the three ship type specific networks. We investigated the strength of evidence for the power law distribution and two alternatives: the exponential $P(x) = \lambda \exp(-\lambda x)$ and log-normal distribution $P(x) = \frac{\exp(-(\log x - \mu)^2 / (2\sigma^2))}{x\sigma\sqrt{2\pi}}$.

Akaike weights were always very distinctive, one of the distribution types revealing $0.98 \leq w_i \leq 1$. In Table 2.4 we specify the preferred distribution for each network and measure. If it is a power law, we additionally provide the maximum likelihood estimate of the exponent $\hat{\mu}$ or $\hat{\eta}$, respectively, with its standard deviation.

Agreeing with our previous results, the Akaike weights for the degree distributions favor the exponential. Thus, in contrast to other real-world networks, the degree distribution of the cargo ship network does not conform with strict scale-free behavior. The results for the link weight and node strength distributions as well strengthen our previous results. The weight

network	degree k		link weight w		node strength s	
	$P(k)$	$\hat{\gamma}(SD)$	$P(w)$	$\hat{\mu}(SD)$	$P(s)$	$\hat{\eta}(SD)$
GCSN	exponential	-	power law	1.49 (0.00)	power law	1.17 (0.01)
container ships	exponential	-	power law	1.36 (0.00)	power law	1.18 (0.01)
bulk dry carrier	exponential	-	power law	1.64 (0.01)	log-normal	-
oil tanker	exponential	-	log-normal	-	power law	1.21 (0.01)

Table 2.4: Best fit functions for the degree, link weight and node strength distributions selected by Akaike weights. For power law distributions also the maximum likelihood estimates of the exponent are given with standard deviation (SD).

distributions for the GCSN, the container ship and bulk dry carrier networks show clear power law behaviour with exponents $\hat{\mu}$ a little smaller, but of similar order as the regression estimates μ . For the oil tanker network, $P(w)$ is, however, best described by a log-normal distribution. This may be related to the large confidence interval of the according regression slope. A similar picture becomes apparent for the strength distributions. All but the network of bulk dry carriers reveal power law distributions of their node strengths, the exception again coinciding with a large confidence interval of its regression estimate. The maximum likelihood exponents $\hat{\eta}$ are slightly larger than the regression estimates, still they are very small, indicating a rather big number of very strong nodes or very busy ports, respectively.

3 Paper II.

Indications of marine bioinvasion from network theory

An analysis of the global cargo ship network

Andrea Kölzsch and Bernd Blasius;
to be submitted

Abstract

The transport of huge amounts of small aquatic organisms in the ballast tanks and at the hull of large cargo ships leads to ever increasing rates of marine bioinvasion. In this study, we theoretically examined this first stage of bioinvasion, the introduction, because then one can still intervene with regulating measures. We applied a selection of network properties and analysed the structure and spread characteristics of the directed and weighted global cargo ship network (GCSN). The GCSN is highly efficient, shows small world characteristics and is positive assortative. This indicates a strong community structure and that quick spread of invasive organisms between the ports is likely. We revealed that the GCSN contains two large communities, the Atlantic and Pacific trading groups, and several small, local ones. Ports that appear as connector hubs between the communities are also most important regarding spread specific measures of centrality. They are the Suez and Panama Canal, Singapore and Shanghai. Furthermore, from our calculations of the robustness of the cargo ship network and determining its percolation behaviour, differences of onboard and in port ballast water treatment were evaluated and set into context with previous studies. For effective bioinvasion management one would have to outfit around 80% of the largest ports with in port treatment or apply onboard treatment of an efficiency above 80% to each ship. Because of the presently unclear efficiency of different ballast treatment devices, we propose that it is most sensible to combine the two.

3.1 Introduction

Biological invasions are geographical expansions of species into areas not previously occupied by it. This is a natural process. However, lately bioinvasion events occur at extremely accelerated rates due to human actions, sometimes by deliberate introductions, but often accidentally (Elton, 1958; Vermeij, 1996; Mack *et al.*, 2000; Kolar & Lodge, 2001). Such increased rates of bioinvasion greatly threaten biodiversity and ecosystem functioning worldwide (Lodge, 1993;

Sala *et al.*, 2000). Additionally, they cause damages of human facilities, impact the economy, and pose unpredictable hazards to our health and livelihood (Pimentel *et al.*, 2005; Ruiz *et al.*, 2000; Tatem *et al.*, 2006). Bioinvasion is a three stage process consisting of the introduction, establishment and proliferation of the invasive organisms (Elton, 1958; Williamson, 1996). Introduction is the stage where man can still intervene and the prevention of bioinvasion is possible. However, this phase of bioinvasion is least studied up to now (Puth & Post, 2005). Thus, we want to concentrate on it here.

Bioinvasion of marine organisms is leading to an ever increasing level of homogenisation of the world oceans' ecosystems (Carlton, 1996b; Ruiz *et al.*, 1997). In some cases it has caused habitat destruction and ecosystem degeneration (e.g. *Caulerpa taxifolia* in the Mediterranean; Meinesz *et al.*, 1993) and pronounced fishing declines (e.g. *Mnemiopsis leydii* in the Black Sea; GESAMP, 1997). There are different vectors by which marine bioinvasive organisms can be spread. By cargo ships a large number of small organisms are transported between the ports of the world. Plankton, larval molluscs and small fishes travel within the huge amounts of ships' ballast water (Carlton, 1985). Plankton samples from several cargo ships from Japanese ports e.g. contained at least 367 different taxa (Carlton & Geller, 1993; see also Gollasch *et al.*, 2000). A large number of invasive organisms can also be attached to the ships' hulls (Drake & Lodge, 2007). For bioinvasion success one has to consider that most species are very sensitive to oxygen depletion in the tanks and high salinity of open seas water (Williamson & Fitter, 1996; Gollasch *et al.*, 2000; Mack *et al.*, 2000). Approximately 90% of the present world trade is being transported by ships and trade volumes increase greatly every year (IMO, 2006; UN, 2007). Therefore, a great number of large cargo ships travel the world's oceans at ever increasing rates, thereby advancing levels of marine bioinvasion. For containment thereof the applicability of several ballast water management and anti-fouling options are presently discussed more or less controversially (Gollasch *et al.*, 2007; Hopkins & Forrest, 2008). A third, important way of marine bioinvasion is introduction by aquaculture. However, here we want to focus on and examine ocean shipping as the main vector of transport of marine bioinvasive organisms.

For the quantification of bioinvasive spread by ships and the identification of important donor and recipient regions it appears straightforward to apply network theory to a global network of ports. Network theory has become a widely applied and rather diverse field of study in physics, social sciences, transportation, ecology and epidemics spread (Newman, 2003b, 2006; Urban *et al.*, 2009). In several studies habitat patches were modelled as nodes that are linked by dispersal events. The spread of SARS during 2002/2003, for example, has been reasonably well reproduced simulating propagation dynamics on the worldwide airport network (Hufnagel *et al.*, 2004). Many different measures have been developed to characterise real world networks and applied to a variety of systems (Costa *et al.*, 2007). Often unexpected properties were revealed and similarities and differences to other networks became obvious. So, for example, in various large and very complex networks (i) pairs of nodes are connected by paths of only few consecutive links and (ii) nodes are locally densely clustered. These two properties have been summarised as "small world behaviour" (Watts & Strogatz, 1998), indicating that any pair of nodes is surprisingly well connected.

Drake & Lodge (2004) had developed a global network of ship traffic from information on the inflow and outflow of goods in a number of ports worldwide, assuming spatially homogeneous

trade flows (i.e. gravity modelling; Haynes & Fotheringham, 1984). We regard this network rather inaccurate, since world trade and ship traffic are very heterogeneous, much influenced by cultural and political issues. However, it has been used in combination with the worldwide airport network (Guimera *et al.*, 2005) and climate information for studying disease dispersal by global traffic (Tatem *et al.*, 2006). For improvement, lately we and others (Kaluza *et al.*, 2009) developed the global cargo ship network (GCSN) from real cargo ship trajectories. When comparing our network with the one of Drake & Lodge (2004), we find that already the number and list of ports differs greatly. There are many important ports that have not been included previously, since the list of ship trade (Zachcial & Heideloff, 2001) that was used is not complete. On the other hand there are a number of ports that are not present in the GCSN, mainly African ports. However, when considering the small amount of trade that is transacted through these ports, they may not be very important for large scale consideration of bioinvasion by ships. Thus, we propose that the GCSN is more appropriate for quantification and analysis of marine bioinvasion by ballast water transport and hull fouling.

In the here presented study we examine structural and spread characteristics of the GCSN by applying a selection of network measures. They are characteristics like small world properties, network efficiency and assortativity, but mainly measures that indicate community structure, centralities of single ports and network robustness. We explain how they allow assertions about invasion spread and extract preliminary indications of the importance of different ports and trade structures for bioinvasion. Furthermore, we quantify how node deletions, i.e. application of ballast water treatment in ports, can affect bioinvasion spread on the GCSN and which transmission rates are of concern. Applying network measures for bioinvasion predictions and evaluation has the advantage that no complicated models have to be developed and no costly and time consuming field studies be conducted. Already by simply examining transportation vector networks preliminary indications of spread characteristics can be derived for decision making authorities.

3.2 Methods

Because of the strong sensitivity of invasion success to travel conditions and propagule pressure it is very important to consider travel frequency and the duration of transport. Therefore, we consider it crucial to analyse the *weighted* cargo ship network for indications of the structural importance and dynamical properties of marine bioinvasive spread. In some cases we additionally analysed the unweighted and even the geographically embedded network for comparison. Many links in the GCSN are unidirectional, i.e. the network is highly asymmetric (Kaluza *et al.*, 2009). As this may be important for the patterns of spread, we retain the directionality of the GCSN. Not many network measures have been developed for such directed, weighted networks (but see e.g. Miguéns & Mendes, 2008). Thus, we provide a summary of several characteristics (see Tab. 1) that are useful for spread analyses.

3.2.1 Small world characteristics

One of the most important network properties for transportation is its connectivity. A quantitative concept of such characterisation was introduced by Latora & Marchiori (2001). It is

Network measure	Definition	
	Topological network	Geographically embedded network
Global efficiency	$E_{glob} = E(G) = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{ \sigma_{ij} }$	$E_{glob}^{geo} = E^{geo}(G) = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{ \sigma_{ij}^{geo} }$
Local efficiency	$E_{loc} = \frac{1}{n} \sum_{i \in G} E(G_i)$	$E_{loc}^{geo} = \frac{1}{n} \sum_{i \in G} E^{geo}(G_i)$
Network cost	$c = \frac{\sum_{i \neq j \in G} a_{ij}}{n(n-1)}$	$c^{geo} = \frac{\sum_{i \neq j \in G} a_{ij} d_{ij}}{\sum_{i \neq j \in G} d_{ij}}$

Table 3.1: Global and local efficiencies and network cost for the topological and geographically embedded GCSN. In each formula n is the number of nodes in the network G . G_i is the subnetwork of neighbours of i and σ_{ij} are the shortest paths between nodes i and j . Different measures of pair distance are the length of the shortest topological path $|\sigma_{ij}|$, the length of the shortest geographically embedded path $|\sigma_{ij}^{geo}|$ and the waterway shortest distance d_{ij} . a_{ij} are the elements of the adjacency matrix, $a_{ij} = 1$ if there exists a link from i to j , $a_{ij} = 0$ else.

based on directed topological or geographical distances between nodes and includes calculations of the local and global efficiencies, E_{loc} and E_{glob} , and the network cost c (Tab. 3.1). High local and global efficiencies in combination with low cost indicate small world behaviour. This definition differs from the classical concept of small worlds (see above; Watts & Strogatz, 1998). For the topological as well as the geographically embedded GCSN, i.e. each node is characterised by its geographical position, efficiencies and cost were calculated, indicating how well connected ports are and how long travels take. Geographical distances between ports were estimated as shortest sea travel routes (for details see Kaluza *et al.*, 2009).

3.2.2 Assortativity

Another characteristic of network structure and possible spread is the assortativity $k_{nn,i}$ of node linkages. It determines the correlation between a node's number of ingoing links (indegree) and the average of its neighbours' mean number of outgoing links (outdegrees) (Pastor-Satorras *et al.*, 2001). We calculated this relationship for the unweighted as well as weighted GCSN. For the weighted assortativity, $k_{nn,i}^w$, links' weights were integrated (de Montis *et al.*, 2007, see Tab. 3.2 a). A positive relationship, i.e. positive assortativity, indicates that nodes are mostly connected to neighbours with a similar number of links. For the large scale structure of the network this implies clustering into several strongly intertwined groups of nodes (Newman, 2002, 2003a). On the contrary, negative assortativity, also called disassortativity, indicates that strongly connected ports are connected to a very large number of less well-connected nodes, and the network has a relatively homogeneous, star-like structure (Redner, 2008). Furthermore, the comparison between the unweighted and weighted nearest neighbours' degrees shows if nodes of high or low strength are clustered together and how degrees and strengths of neighbours correlate (Miguéns & Mendes, 2008). This indicates if bioinvasive spread concentrates on groups of strongly connected ports only or if organisms can quickly spread through the whole

network.

3.2.3 Network community structure

A detailed analysis of the community structure (Newman, 2006) of the GCSN provides further insight into its transportation characteristics. The optimal network communities are obtained by minimising the network's modularity. This is defined as the proportion of links that fall within compartments of the considered network compared to that expected in an equivalent random network (see Table 3.2 c). Leicht & Newman (2008) developed an algorithm that incorporates the directionality of links of directed networks in the partitioning method. This is appropriate and applied here for calculating the communities of the weighted GCSN. Additionally, we characterised the members of each community according to its role based on its intra- and intercommunity connections. On that account we calculated the within-community degree's z-score of each node and its participation coefficient (Tab 3.2 b; Guimera & Amaral, 2005b; Guimera *et al.*, 2005). The participation coefficient describes how many of a node's links are connected only within its community or also to nodes of other communities. The phase plot of the within-community degrees and the participation coefficients reveals how nodes can easily be distinguished into provincial hubs, connector hubs or non-hub connectors (Guimera & Amaral, 2005a). In the light of bioinvasion one can conclude from these measures between which ports there is regular, strong exchange of organisms, which ports function as hubs globally or within their community and which ports are the most important for spread of invasives outside of their community.

3.2.4 Network node centralities

Small world networks usually contain a giant component that is a strongly connected (i.e. there is a directed path between any pair of nodes; Newman, 2007) subnetwork of the majority of all ports. This is the case for the GCSN. It has 951 nodes, 935 of which comprise the giant component (Kaluza *et al.*, 2009). In the following section we consider properties of the nodes of the giant component only.

Centrality of a network addresses the issue of which nodes are most important for the structure and dynamical characteristics of the network (Newman, 2007). According to the question of what is considered important there are a number of different centrality measures; the three most widely used ones are presented here (see Tab. 3.2 c). Additionally, we develop a new measure of centrality that is especially suited to characterise the general possibilities of spread on a network. For each of the four centralities we provide a distribution of the by centrality weighted proportion of ports situated in each of the six continents. Finally, we examined the correlations of the four centrality measures, using rank correlation by Kendall.

One centrality that has been widely used for characterising spread is the eigenvector centrality x (Bonacich, 1972; Canright & Engo-Monsen, 2006). It is based on the very simple measures of node degree and strength (Newman, 2003b) that count the number of links of each node or sum its weights, respectively. The eigenvector centrality additionally considers that not all links of a node are equally important for spread. It incorporates the recursive dependency of a node's centrality on the average centrality of its neighbours (see Tab. 3.2 c). In that respect,

Network measure	Definition
a	
Nearest neighbours' degree (unweighted)	$k_{nn,i} = \frac{1}{k_i^{in}} \sum_{j \in N_i} k_j^{out}$
Nearest neighbours' degree (weighted)	$k_{nn,i}^w = \frac{1}{s_i^{in}} \sum_{j \in N_i} w_{ij} s_j^{out}$
b	
Modularity	$Q = \frac{1}{\sum_{i \in G} s_i^{in}} \sum_{i,j \in G} \left(w_{ij} - \frac{s_i^{out} s_j^{in}}{\sum_{i \in G} s_i^{in}} \right) \delta_{c_i, c_j}$
Within-community degree z-score	$z_i = \frac{\varsigma_i - \bar{\varsigma}_{c_i}}{SD(\varsigma_{c_i})}$
Participation coefficient	$P_i = 1 - \sum_{s=1}^{N_c} \left(\frac{\varsigma_{ic}}{s_i} \right)^2$
c	
Eigenvector centrality	$x_i = \frac{1}{\lambda} \sum_{j \in G} w_{ij} x_j$
Closeness centrality	$cc_i = \sum_{j \in G} \sigma_{ij}^w $
Betweenness centrality	$bc_i = \sum_{j,k \in G} \frac{\#\{\sigma_{jik}^w\}}{\#\{\sigma_{jk}^w\}}$
R_0 centrality	$rc_i = \alpha s_i^{in} s_i^{out} + (1 - \alpha) s_i^{in} (s_i^{out} - \bar{w}_i)$
d	
Cluster size growth	$R_0 = r \frac{\alpha \sum_{i \in G} s_i^{in} s_i^{out} + (1 - \alpha) \sum_{i \in G} s_i^{in} (s_i^{out} - \bar{w}_i)}{\sum_{i \in G} s_i^{in}}$

Table 3.2: Network measures for the characterisation of the directed, weighted GCSN. (a) Formulae for the calculation of the nearest neighbours' degrees for the examination of assortativity in the unweighted and weighted network. w_{ij} is the weight of the link from i to j , k_i^{in} and k_i^{out} are the number of in- and outgoing links (in/outdegrees) of node i , and s_i^{in} and s_i^{out} the cumulative weights of in- and outgoing links (in/outstrength). **(b)** For modularity calculations m is the number of links in the network G , δ_{ij} the Kronecker delta and c_i the label of the community to which node i belongs. ς_i is the sum of the weights of all links that connect node i to nodes in its community c_i . $\bar{\varsigma}_{c_i}$ is the mean ς over all nodes belonging to community c_i and $SD(\varsigma_{c_i})$ is the corresponding standard deviation. ς_{ic} represents the summed weights of the links of node i to nodes in any community c . **(c)** Formulae for centrality measures further contain the largest eigenvalue λ of the weight matrix and shortest weighted paths σ_{jk}^w between j and k (link distance is set to $1/w_{ij}$). The shortest path length is $|\sigma_{jk}^w|$ and $\{\sigma_{jk}^w\}$ is the number of paths of shortest length between j and k in general and $\{\sigma_{jik}^w\}$ the number of paths passing node i on their way from j to k . α is the proportion of unidirectional connections between pairs of ports and \bar{w}_i the mean weight of outgoing links from i . **(d)** Variables for calculating cluster size growth R_0 are the transition rates r per year and link or unit weight (here per 16,232 GT, the mean value of all included ships), α , in- and outdegrees and -strengths, and the mean outweigh from i , \bar{w}_i . For the critical transition to global spread, $R_0 = 1$, r is the percolation threshold.

ports with many well connected neighbouring ports are more strongly involved in bioinvasion.

Two other measures of centrality that are widely used in social studies are closeness centrality and betweenness centrality. Both are based on the set of shortest directed paths between all pairs of ports. For the weighted network we use shortest paths in regard to the link distances of $1/w_{ij}$, with w_{ij} being the cumulated cargo capacity of ships travelling through. Shortest paths then minimise the sum of these distances. Thus, links that are heavily frequented are related to a “shorter” distance, and bioinvasion along those routes is more likely. Closeness centrality cc (Wassermann & Faust, 1994) for each port i is the reciprocal of the average of shortest path lengths from i to all ports $j \neq i$. It points out from and to which ports invasive organisms can spread quickly. Betweenness centrality bc (Freeman, 1979) is the number of shortest paths between all pairs of nodes that pass through node i (Tab. 3.2 c). Ports that are most “between” are especially important for guaranteeing short paths and high connectedness. If they are deleted, shortest path lengths will increase much and bioinvasion spread will slow down.

We developed a new centrality measure of spread from percolation theory (Ben-Avraham & Havlin, 2000) that we want to call R_0 centrality rc . It is derived from the cluster size growth R_0 (Tab. 3.2 c, d), being proportional to each nodes' contribution to it. The R_0 centrality is negatively related to the percolation threshold r^* (see below; Callaway *et al.*, 2000) that is the transmission rate above which spread through the whole network is likely within a certain time interval. In contrast to the other centralities the R_0 centrality is a very local one, depending on a node's instrength and outstrength only. It is a general measure of how much each node and its neighbours contribute to the possibility of global spread of e.g. bioinvasive organisms through the network.

3.2.5 Network robustness and percolation

In the following, we examine how different ballast water management and anti-fouling options affect bioinvasion using network theory. First, we analyse error and attack tolerances (Albert *et al.*, 2000) of the GCSN. This means that we delete nodes and all their links one by one and examine the size of the giant component of the remaining network. In analogy to error tolerance we select the nodes to be deleted in random order, for attack tolerance analyses we consecutively delete nodes of maximum degree or strength. If the networks are more sensitive to selective deletions than to random ones, bioinvasion could be slowed down by equipping a certain proportion of the largest ports with ballast water management devices.

Inspired by results of Drake & Lodge (2004) we examined how invasive spread is determined by per ship transmission intensity and how this is affected by node deletions. For sensible quantitative predictions and comparability, in this analysis we adjusted link strengths of the GCSN multiplying them with a factor of 1.72, because only ships comprising 58% of the global cargo capacity were available for network development (Kaluza *et al.*, 2009). Average ship numbers for each link were calculated by dividing the link's weight in GT by the mean of all ships' GTs. For the full GCSN and a set of different per year transmission probabilities r we calculated the average fraction of ports infected from one random, initially infected port (1000 simulations each). Furthermore, we determined the critical transmission probability, the percolation threshold r^* , for which the cluster size growth $R_0 = 1$. For transmission

probabilities above this value, $r > r^*$, the number of invaded ports grows exponentially as ships travel the oceans, and soon all will be invaded. However, if $R_0 < 1$ and thus transition rates $r < r^*$, bioinvasion spread will decrease and diminish during the considered year. R_0 is also called the basic reproductive number in epidemics spread theory (Anderson & May, 1991; Newman, 2007).

To put the issues of transmission intensity decrease and node deletion into perspective we calculated r^* for the GCSN under random and selective node deletion. This provides indications of the magnitudes of the per year transition probability decrease and the number of ports with ballast water treatment required for controlling bioinvasion. Results were compared to a specific per ship transmission probability that has been estimated from empirical data (Drake & Lodge, 2004).

3.3 Results

3.3.1 Small world properties

Global and local efficiencies and network cost for the plainly topological, but directed GCSN are $E_{glob} = 0.43$, $E_{loc} = 0.75$ and $c = 0.04$. This reveals that network topology very cost efficiently connects the ports of the world (Latora & Marchiori, 2001). The high value of E_{loc} points out that the ship network is very fault tolerant, i.e. failure of one node will not much affect the efficiency of its neighbouring nodes. E_{glob} indicates that the topological GCSN is approximately half as efficient as a fully connected network of the same size at a cost of only 4%.

Efficiencies and cost for the geographically embedded GCSN are somewhat different, $E_{glob} = 0.99$, $E_{loc} = 0.70$ and $c = 0.22$. It becomes obvious that, when considering the geographical locations of each port, the GCSN is only 1% less efficient than an ideal ship traffic network with a direct connection between each pair of ports. This is very remarkable. The local efficiency E_{loc} does not change a lot in comparison to the topological network, error tolerance is still given. The cost, in accordance with the increased E_{glob} , increased somewhat, but is still far from the 100% for the fully connected network. The geographically embedded GCSN is practically as efficient as the fully connected one at a cost of only 22%. Concluding, one might state that the GCSN is a small world network.

3.3.2 Assortativity

The relationship between the indegree k_i^{in} of each port i and the average degree of its neighbours $k_{nn,i}$ and $k_{nn,i}^w$ (see Fig. 3.1) reveals that the global cargo ship network is positively assortative. Especially up to port sizes of $k_i^{in} = 70$ there is a strong increase of $k_{nn,i}$ and $k_{nn,i}^w$. Above that average neighbours' degrees of the unweighted GCSN seem to asymptotically approach a limit at around $k_{nn} = 70$. This means that at this point larger ports cannot obtain more links to other ports. Then, as can be observed from the further growth of the weighted k_{nn}^w with k_i^{in} , ports with extremely many connections are preferably linked with larger ports. The positive assortativity, and this special kind of it in particular, indicates

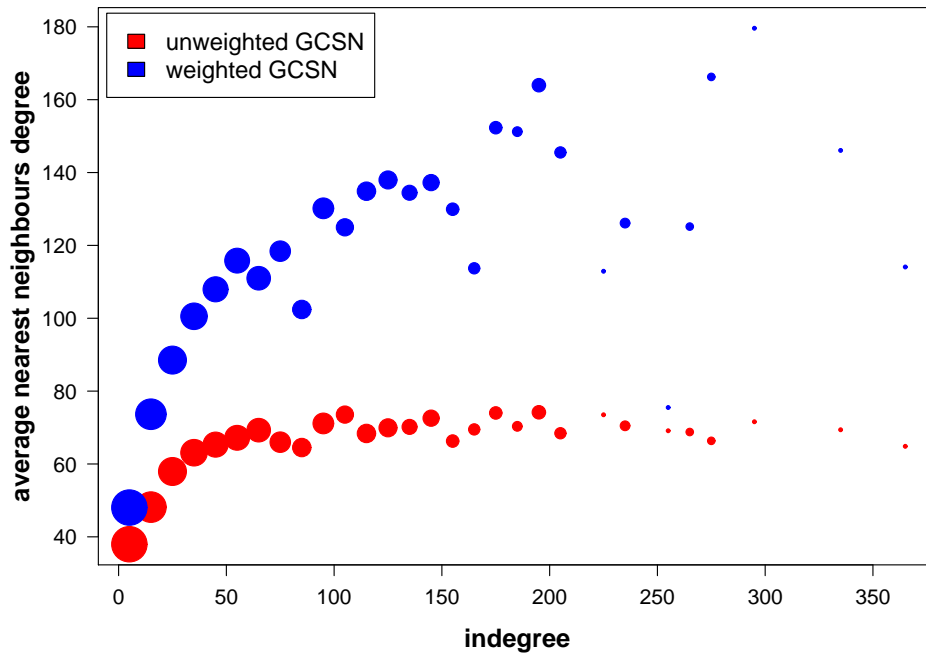


Figure 3.1: Degree correlations as signs of network assortativity. Average nearest neighbours' degrees of each node of the unweighted (red dots) and weighted (blue dots) GCSN (Tab. 3.2 a) averaged over bins of size 10 for k^{in} . Radii of the dots indicate the sample size, i.e. number of different ports within the respective bin.

that the global cargo ship network contains several communities, i.e. groups of strongly connected ports. The fact that k_{nn}^w is always above k_{nn} underlines that links with large weights are directed towards neighbours with large degrees. This is emphasised by their averages, $\langle k_{nn,i} \rangle = 52.21 < \langle k_{nn,i}^w \rangle = 81.68$. If k_{nn}^w approaches a limit is not clear, because sample sizes for ports with large indegrees are rather small. In the dynamics of both measures (Fig. 3.1) one can observe slight periodic behaviour, k_{nn} and k_{nn}^w showing a number of equidistant, local minima. This may be another indication of the network structure and sizes of its compartments.

3.3.3 Network community structure

The community structure of the weighted GCSN is presented in Fig 3.2 a. There are nine smaller groups and two large trading communities, the Atlantic group including European and American countries and the Pacific group of Asia and Australia. Small communities are specialised, often local trade routes (e.g. West Africa – Argentina/Brazil) and ferry connections (e.g. Dover – Calais). The phase plot of participation coefficients indicates which ports are especially important for global and local network connectivity (Fig. 3.2 b). Oil ports at the coast of Louisiana as well as some ferry ports in Europe fall into the group of provincial hubs. Many large, well known and globally significant ports are connector hubs, e.g. the Panama and

Suez Canal, Shanghai, New York & New Jersey, Singapore and Antwerp. It may be notable that the Panama Canal has a very high within-module degree whereas the Suez Canal has an extremely high participation coefficient, thus being very important for network connectedness. Some ports, like Santos and Le Havre are nonhub connectors. They are not of a very high within module degree, but important to keep the GCSN globally connected. Note that there are no kinless hubs or nonhubs, i.e. ports the links of which are homogeneously distributed among all communities.

3.3.4 Network node centralities

The rankings of the ports of the weighted GCSN according to each of the different centrality measures are presented in Fig. 3.3. We highlighted the top 100 of the by each of the four measures most central ports and provide pie charts of the weighted proportions of central ports in each continent. The Panama and Suez Canal frequently appear in the top 10 central ports (see Tab. 3). This underlines their high importance in topologically holding together the network. Furthermore, Shanghai and Singapore are always among the first ranks for whichever type of centrality. Singapore is even first for three of the four kinds of centrality. One can observe that there are always very few ports of Africa, Australia or South America among the most central.

There are several differences between the most important ports in respect to the four kinds of centrality. Ports of highest eigenvector centrality that are important for spread in the long term, are mostly situated at the Gulf of Mexico. Closeness centrality, a more short term indicator for invasive spread, reveals a more homogeneous distribution of most central ports. Among them are many European, North American and Asian ports, with a slight preference of ports in the Suez region. The high closeness of this region to any other port underlines its importance for keeping the two major components of the weighted GCSN connected (see previous section). In terms of betweenness centrality again Singapore, Shanghai and the two big canals are most important besides a large number of ports in northern Europe. The R_0 centrality provides somewhat intermediate results. Ports of highest R_0 centrality are Shanghai, Singapore, the Suez and Panama Canal, some ports at the Gulf of Mexico and several ones in northern Europe (Tab. 3).

Rank correlations between all pairs of the four centralities are positive $0.43 \leq \tau \leq 0.82$ (all $p < 0.001$). The two largest ones are (i) between cc and rc and (ii) x and rc , emphasising that the R_0 centrality is a measure that incorporates several different notions of spread. Correlations that involved bc are lowest, the mainly global character of this measure demarcating it from the others.

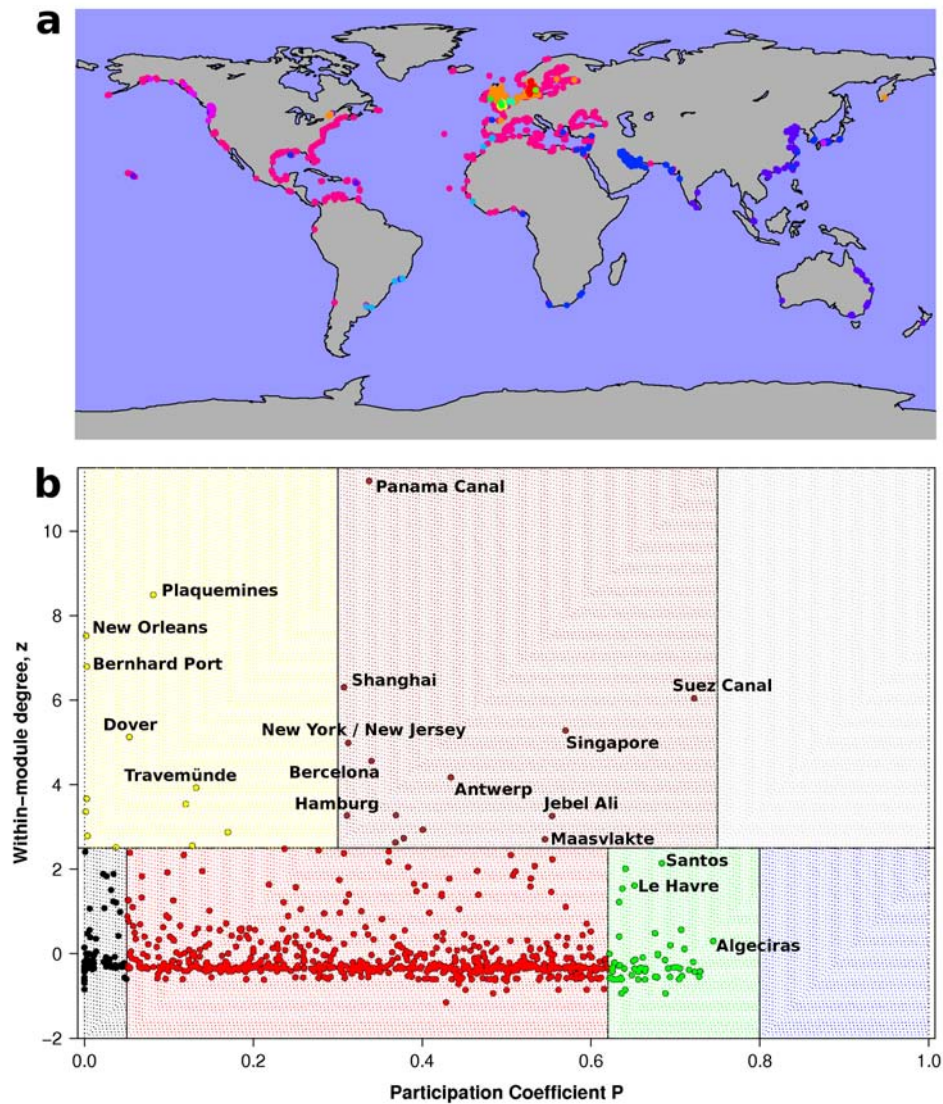


Figure 3.2: Community structure and representation of the roles of each port. (a) Colour coded, the 11 communities of the weighted GCSN. It becomes apparent that trading preferences and geographical distances basically determine these communities. (b) Each port in the phase-space of the z-score of the within-module degree vs. the participation coefficient. The positions of the ports indicate their different roles for network structure and connectivity. Hubs always have a high within-module degree and the larger the participation coefficient of a port the greater its influence on connecting the different communities of the networks. Regions in this phase space are coloured according to Guimera *et al.* (2005) into provincial hubs (yellow), connector hubs (brown), kinless hubs (grey), ultraperipheral (black), peripheral (red), nonhub connectors (green) and kinless nonhubs (blue). The names of the largest hubs and some connecting ports are given.

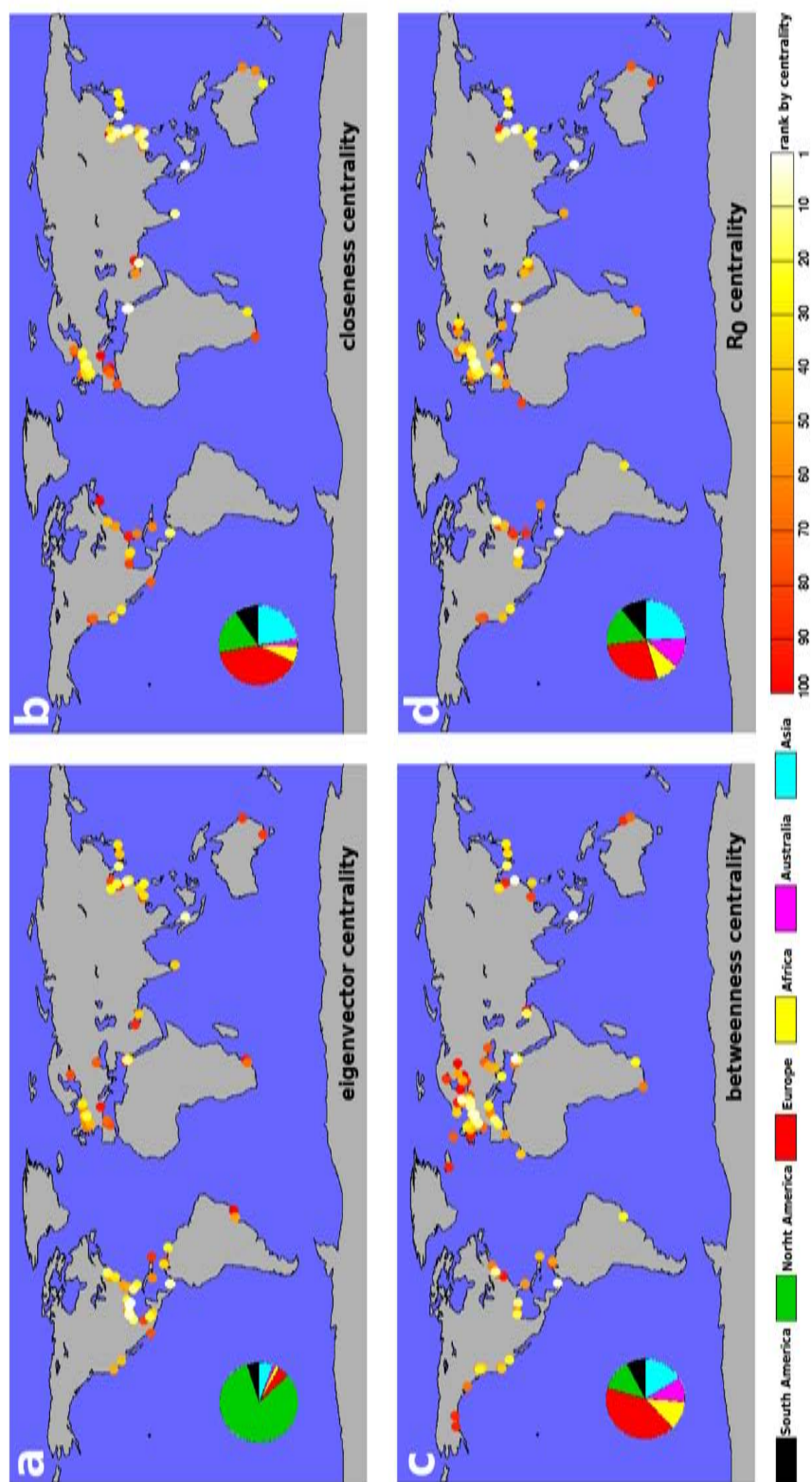


Figure 3.3: Top ports of different centrality measures of the weighted global cargo ship network. Each map shows the positions and colour coded the ranks of the 100 highest ranked ports according to different centrality measures. They are (a) eigenvector centrality, (b) closeness centrality, (c) betweenness centrality and (d) R_0 centrality. The most central ports are coloured light yellow and white. Insets depict by the respective centrality measure weighted distributions of ports in the six continents.

rank	Eigenvector centrality	Closeness centr.	Betweenness centr.	R_0 centrality
1	St Bernhard Port	Singapore	Singapore	Singapore
2	New Orleans	Suez Canal	Suez Canal	Panama Canal
3	Plaquemines	Shanghai	Shanghai	Shanghai
4	Port of South Louisiana	Port Said	Panama Canal	Suez Canal
5	Mississippi River	Suez	Maasvlakte	Plaquemines
6	Houston	Jebel Ali	Gothenburg	Antwerp
7	Panama Canal	Busan	Antwerp	Busan
8	Barbours Cut	East Port Said	Zeebrugge	New Orleans
9	Singapore	Kaohsiung	Le Havre	St Bernhard Port
10	Shanghai	Qingdao	Bremerhaven	Maasvlakte

Table 3.3: Top ten most central ports in terms of eigenvector, closeness and betweenness centrality and the novel R_0 centrality.

3.3.5 Network robustness and percolation

The global cargo ship network is very robust to random node deletions (Fig. 3.4 a). Up to a fraction of 80 – 90% of ports remaining, the size of the largest component decreases almost linearly in steps of size one. Thus, if randomly deleting any proportion less than 80% of all nodes the leftover ports are still strongly connected to each other, and spread of bioinvasive organisms through the remaining network prevails. Against selective, ordered deletion of preferred high degree (blue lines) and high strength (red lines) nodes, respectively, the GCSN is less robust. Interestingly, the effect on network connectivity of the deletion of nodes selected by degree or strength is very similar. In Fig. 3.4 a it is shown that connectivity of the giant component is only retained up to a deletion of 35 – 45% of the most strongly connected ports; else the whole network loses its strong connectivity. Additionally, the slopes of the red and blue lines are relatively steep. This means that the network becomes disconnected quickly, losing its strong connectivity already after removing a very small proportion of the best connected nodes.

Results of network percolation, i.e. global spread, with different per ship transmission rates are depicted in Fig. 3.4 b. With increasing transmission probability per ship and year the fraction of infected ports grows from 0 to 0.99, steeply increasing at transmission probabilities slightly above the percolation threshold $r^* = 9.14 \cdot 10^{-5}$. The proportion of infected ports never reaches 1, because due to edge effects of data sampling a subset of 10 ports has no ingoing links and cannot be infected. They are Papenburg, Leer, Stralsund and Wolgast in Germany, Marugame and Hashihama in Japan, Haikou and Jinshan in China, Cagliari in Italy and Cekisan in Turkey. The value of the percolation threshold is very small, even below the transmission probability estimated from empirical data (Drake & Lodge, 2004; see also Carlton & Geller, 1993), $r^{DL} = 4.4 \cdot 10^{-4}$. The cluster size growth for r^{DL} in our network is $R_0 = 4.81$. To push r^{DL} below r^* for the full GCSN it would have to be reduced by at least $1 - r^*/r^{DL} = 79.2\%$. When additionally deleting nodes one by one the percolation threshold r^* increases, in the beginning a little faster for by strength selective node deletion (Fig. 3.4 c). The increase is, however, very slow, and above deleting 40% of all nodes there is no difference in r^* between random and selective node removal. This is a sign of high clustering and local efficiency

throughout the GCSN. To push the threshold value r^* above the estimated transmission rate r^{DL} by port deletions only, about 70% of all ports would have to be deleted, even if by then any strongly connected component of the network has fallen apart (see above). In several ways optimal combinations of node deletion and transmission probability decrease can be read off Fig. 3.4 c. For example, when deleting 40% of the strongest connected nodes r^* would have to be reduced by more than 53.3% to avert global spread.

3.4 Discussion

In the here presented work we extracted novel information about ship traffic structures and its indications for marine bioinvasion calculating simple network characteristics of the global cargo ship network. Our results point out that cargo ship traffic is globally very efficient and ports closely connected. This agrees well with other transportation networks (e.g. Latora & Marchiori, 2002; Barrat *et al.*, 2004) and previous results concerning the GCSN (high clustering $C = 0.49$, small topologically shortest paths $\langle l \rangle = 4.4$; Kaluza *et al.*, 2009). Concerning bioinvasion this means that the structure of the GCSN strongly supports quick spread of invasive organisms, so that if species can establish in new habitats sooner or later a global homogenisation of marine ecosystems is likely. Such is even accelerating as global trade of goods is increasing at unprecedented rates (IMO, 2006). However, due to the global finance crisis, presently ship traffic seems to decline.

The here applied efficiency and network cost measures for the quantification of network connectivity differ from the classical concept of small world determination (Watts & Strogatz, 1998), but results nicely coincide (Kaluza *et al.*, 2009). The efficiency-cost concept allows for clear physical interpretation, how well are pairs of nodes connected. They do not regard the network links' weights, but are calculated for the purely topological network and the geographically embedded one. For many invasion events not propagule pressure, but conditions in the recipient regions determine invasion success. Therefore, the examination of the simple network structure may reveal new insights of general bioinvasion. Especially the optimised transport efficiency on the geographical map of ports is impressive, as the GCSN seems to be almost as efficient as a fully connected network of the same size. The positive assortativity of global shipping underlines that large and well connected ports are strongly linked to other large ones between which bioinvasive organisms quickly spread. So, port size is a good indicator for invasibility and invasiveness. Positive assortativity also indicates a distinct community structure of the GCSN, resulting in more strongly pronounced spread within certain groups of ports.

Community structure, as calculated by modularity minimisation, underlines differences in the ports' spread behaviour. It is strongly influenced by intensive local ship movement, as can be seen by the identified small, local port communities. Often they are highly frequented by ferries. Apart from that the GCSN contains only two very large communities, an Atlantic group of ports and a Pacific one. Thus, global connectivity is very high and invasion spread likely to quickly advance. Furthermore, the importance of different types of hubs is pointed out. Provincial hubs that are mainly European ferry ports and US oil ports at the Gulf of Mexico keep local passenger traffic going and receive oil from off-shore. Bioinvasive organisms that originate

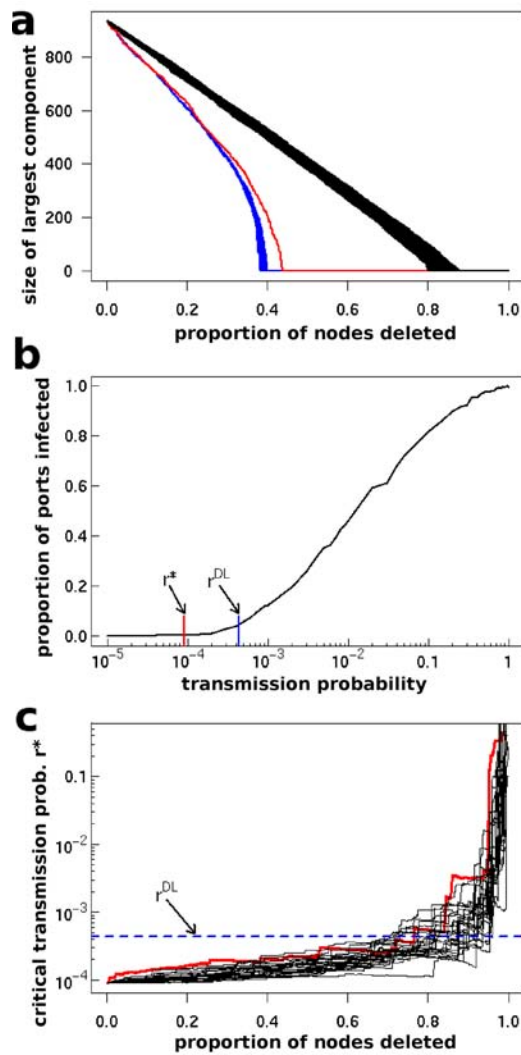


Figure 3.4: Robustness to node removal and critical transmission probabilities. (a) Decrease of network connectivity with random and selective node removal. The black lines show the size of the largest component after random removal of the respective proportions of nodes. The blue lines represent the largest component size if nodes were selected by their degree, i.e. in each step the node of highest degree was deleted. Red lines are results of node deletion preferred in terms of the highest strength. For each scenario 100 samples were calculated and depicted. (b) For the GCSN simulated mean proportion of infected ports for different per mean ship, i.e. 16,232 *GT*, transmission probabilities r . For each r we averaged over 1000 samples, each time infection starting in one random port. The critical transmission probability, the percolation threshold r^* , (red line) was calculated for $R_0 = 1$ and the by 1.72 adjusted node strengths (see Methods and Tab. 3.2d). r^{DL} (blue line) is a by ship transmission probability estimated from empirical data (Drake & Lodge, 2004). (c) Critical transmission probabilities r^* for the cargo ship network with successively deleted nodes. Black lines resemble random node deletion, the red one by strength selected deletion. We marked the by data estimated transmission probability r^{DL} with a blue dashed line.

from them can be quickly spread throughout their local community. Global spread over large distances is basically propagated by connector hubs. They are disproportionately important for bioinvasion spread and coincide with the highly connected ports of Shanghai, Singapore, Antwerp, the Panama Canal and Suez Canal. The importance of hubs for bioinvasion spread has been proposed already by Floerl *et al.* (2009).

The fact that the GCSN contains a giant component (Kaluza *et al.*, 2009) underlines that almost all ports are connected with each other and thus prone to bioinvasion. The more specific centrality measures focus on various aspects of spread through the network. They coincide nicely with the results of the community structure analysis, but additionally reveal that some ports are only important for certain issues of transport, whereas others seem to be central in any. The latter ones coincide well with the connector hubs. Singapore, appearing first in the centrality rankings three times, seems to be of special importance. This conforms well with the fact that currently Singapore is the world's busiest port in terms of total shipping tonnage and container and crude oil transshipment (Heng, 2007). Considering its position at the Strait of Malacca it also provides the functions of a canal, connecting the Pacific with the Indian Ocean. The eigenvalue centrality ranking, revealing a concentration of most central ports around the Gulf of Mexico, deviates from the general findings. The accumulation of highly connected ports at the South coast of Northern America may be surprising, but can be explained by the intensive oil transportation and trade between close ports within this region. The large proportion of highly "between" and "close" ports in Northern Europe indicates that they are very well connected to a wide range of distant ports worldwide, and that European ports are important as intermediate transshipment ports on shortest routes. The intermediate results by the novel R_0 centrality point out that it includes contributions of short and long term global spread characteristics as well as connectivity of each port. Thus, it provides a sufficient general measure for spread in transportation networks. Consequently, the very importance and prospective effectiveness of installing ballast water and antifouling management facilities in the named most central ports are obvious.

As a note of caution one has to add that here no information on habitat characteristics in the ports and geographical port distances are considered. On the one hand, between very close ports many species may be exchanged, but are often also native in the recipient port. On the other hand, if conditions in very distant ports are extremely different, establishment of alien species may be difficult. Thus, local homogenisation of species assemblies and global bioinvasion are treated similarly here. In this study, only the first stage of invasion, the introduction, but not establishment and proliferation were evaluated. As noted above, at this first stage it is most likely that one is able to hinder bioinvasive spread and exert effective control measures. It may be surprising that Rotterdam, the largest port in Europe, is not listed in Tab. 3. By developing the GCSN Rotterdam's different port terminals were not merged and therefore count as own, single ports. Otherwise, it would be among the most central ports.

The discussion about how to most effectively manage marine bioinvasion has been clarified somewhat. Which of the several techniques and treatments to clean ballast water and the large selection of antifouling-paints is most effective to slow down bioinvasion? Drake & Lodge (2004) have directly compared if the exclusion of ports, in-port treatment, or the reduction of transition rates per ship, onboard treatment, more effectively decreased numbers of infected ports in their approximate ship network. Their study indicated that a decrease in

the transmission rates of organisms by ships has a stronger effect than the deletion of single ports from the network. The analysis was rather case study like, only comparing three specific spread scenarios. We find that this is not sufficient for drawing general conclusions, however fits in the general framework that we propose.

We applied distinct, general and quantitative methods for the examination of effects of node deletion and the modification of the transmission probability on network connectivity and bioinvasive proliferation properties. Our results of network robustness to node deletion are not port specific, but quantitative and allow for general conclusions. We found that the global cargo ship network cannot be controlled by the application of treatment techniques in a random set of ports. To this treatment it is highly robust. However, similar to many other scale-free networks (Albert *et al.*, 2000) the GCSN is very vulnerable to the deletion of the most strongly connected nodes. In line with this result one could claim that the “deletion” of at least 40% of the largest ports could substantially decrease spread of bioinvasive organisms. However, this only means that most ports cannot be reached from all other ones (strong connectedness), but spread is still possible. This was revealed by our study of threshold percolation transmission rates.

We showed that there is a steep increase of spread intensity from transmission rates that are insignificant to those that facilitate global spread (percolation). The percolation threshold is quite low, but less than an order of magnitude below the value estimated by Drake & Lodge (2004). Therefore, in their study a decrease of r had a relatively pronounced effect on spread. Here, we systematically quantified how much decrease of the transmission probability per ship is necessary for the inhibition of global spread. At least 80% of all invasive organisms in ballast tanks have to be extinguished to push r^{DL} below r^* . If onboard ballast treatment is that efficient remains to be seen. Therefore, it may be useful to couple in-port treatment and onboard treatment of bioinvasive stowaways. Our simulations of this situation show that a surprisingly high number of ports have to be treated for the percolation threshold to considerably decrease. Selective treatment of strongest connected ports has only a slight advantage, above 40% deletion being not better than some random cases. Thus, it is not sufficient to only apply ballast water and hull fouling treatment in the 40% biggest ports, but a combination of in port and onboard treatment most promising.

Concluding, we provide and apply a number of measures for the characterisation of directed and weighted networks in the light of spread. We specified them for examining the efficiency, structure and robustness of transportation vectors, for the evaluation and prediction of bioinvasion and, possibly, the spread of infectious diseases. In the case of marine bioinvasion by cargo ships we are able to point out which ports are most vulnerable to and promoting invasive spread. Furthermore, we quantified the robustness of network connectivity and spread. However, these results provide only first indications for bioinvasion risk management. In the light of the obtained insights one should now be able to develop general population dynamics models including species' establishment and proliferation in the ports. Additionally, one can examine the effects of ballast water treatment in more detail and integrate specific environmental conditions into modelling. The global cargo ship network provides a sensible framework for such an examination of bioinvasion risk for management and decision making.

Acknowledgements

We want to thank H. Rosenthal and H. Seebens for helpful discussion. This work was supported by the German VW-Stiftung and the BMBF.

4 Paper III.

Theoretical approaches to bird migration

The white stork as a case study

Andrea Kölzsch and Bernd Blasius;
The European Physical Journal Special Topics **157**: 191-208 (2008)

Abstract

Birds are often considered to be one of the best studied groups of organisms. However, only a few investigations have been devoted to a theoretical analysis of avian migration patterns in time and space. This paper is meant to be a first step into this direction. We start by presenting different types of observational data sets that are available and discuss their advantages and disadvantages for use in quantitative analysis of bird movement and migration. Based on ring recovery and satellite telemetry data we perform a statistical analysis of the migratory patterns of the white stork *Ciconia ciconia*. We find that standard methods from random walk theory can be applied, but have to be carefully interpreted and possibly modified to analyse migration movement data which are dominated by seasonal drift. Our analysis reveals two different modes of movement - fast, directed migration and slow, undirected resting. Furthermore, we present a conceptual network model of avian migration. In our model a number of discrete breeding, resting and wintering habitats are linked by migration in the form of seasonally driven transition probabilities that are described by a unimodal circular function of time. Our study emphasises the need for more rigorous quantitative data analysis and mathematical modelling to gain a better understanding of the dynamic processes of avian migration.

4.1 Introduction

Bird migration has fascinated man for a very long time and it has been studied in many aspects and from various perspectives. As a large number of bird species from all parts of the world venture periodic, seasonally driven return migrations of various lengths, one is interested in their routes and the mechanisms driving this phenomenon (Berthold, 2001a). There are different forms of bird migration. Some species fly thousands of kilometers between their wintering and breeding areas (long-distance migrants, e.g. the arctic tern *Sterna paradisaea*), whereas others just move a few hundred kilometers from their breeding grounds to areas of more favourable feeding conditions in the winter. Each migratory bird species has different physiological features that determine its movement behaviour and migration routes. For example, large birds need

thermalling to fly long distances (Schüz *et al.*, 1971) which restricts their routes mainly to over-land.

Many species have certain areas on their way that they visit each year for moulting, resting and feeding up. Thus, their migration can be described as a jumping process on a network of more or less discrete habitat patches. Complex networks have been widely used to model spatial processes in the real world (Albert & Barabási, 2002; Newman, 2003b), not least because the discrete space dimension used by this method makes it relatively easy to examine very complex processes and patterns. In spatial ecology network approaches have, for example, been popular for describing metapopulation dynamics on a network of habitat patches or the spread of infectious diseases (Tilman & Kareiva, 1997; Pastor-Satorras & Vespignani, 2001; Gross *et al.*, 2006; Hufnagel *et al.*, 2004). While the statistics underlying transport processes on complex networks has been intensively studied (Albert & Barabási, 2002; Newman, 2003b), seasonal bird migration implies jumping rates which are not constant, but periodic functions of time. This will change the network and transport characteristics in a fundamental way.

The process of migratory bird movement is not only varying within years, but is exposed to individual variation, between year variability and gradual directional changes during the last years. Due to human mediated habitat destruction and climate change, some birds are altering their migratory routes and timing (Walther *et al.*, 2002; Jenni & Kéry, 2003). This affects species with different strategies in various ways and may become of concern for their persistence. Another topic that is presently of concern is that the spread of diseases and pathogenic microorganisms has been associated with bird migration (Hubálek, 2004; Kilpatrick *et al.*, 2006). From an epidemiological point of view it is extremely important to predict the impact of such pathogen dispersal on human populations along migratory routes. These scenarios further emphasise our motivation to characterise migration patterns and to quantitatively describe the process of bird migration.

Up to now only few studies have been devoted to investigate bird migration from a theoretical point of view (Bairlein, 2003), and if then mostly the physiological mechanisms of optimal flight strategies (Erni *et al.*, 2002; Hedenström & Ålerstam, 1997) and navigation mechanisms (Wiltschko & Wiltschko, 2003; Mouritsen, 1998) were examined in experimental and simulation studies. These aspects of avian migration are thus well under study, even if widely discussed lately (Ålerstam, 2006). To our knowledge, nobody has yet systematically explored and reproduced the movement of migratory birds in time and space. One noteworthy exception is a seminal work by (Viswanathan *et al.*, 1996). They investigated the flight distances of albatrosses in the South Atlantic and showed that bird movements can be described as Lévy flights. However, this study regarded only foraging movement, while migratory movements which proceed on much larger spatial scales were not included in their analysis.

In a fast changing world it becomes more and more important to understand general ecological processes in detail and determine its underlying courses. By mathematical modelling many phenomena in nature have been systematically examined and one was able to characterise important driving mechanisms (e.g. in epidemiology or ecology (Tilman & Kareiva, 1997; Pastor-Satorras & Vespignani, 2001; Hufnagel *et al.*, 2004; Hethcote, 2000; Blasius *et al.*, 1999)). We want to point out that also in bird migration time has come for more abstract and rigorous modelling and data evaluation. There are several challenges that will have to be met, because of the wide ranges of migration routes, their variability between the years and

individuals, data deficiencies and the several time dependencies of movement parameters and population characteristics. However, if they are met in a convenient way we may be rewarded with a thorough understanding of avian migratory movement patterns which can be applied in a wide range of disciplines, like habitat management, climate change, optimal flight and epidemiology. This study is meant to be a first step into this direction and our aim is to present some first approaches for a theoretical description of the migratory movement of birds. Another major motivation is to propose this topic to the physics community as an intriguing problem and interesting case study.

We begin with a short review of different kinds of observation techniques that were developed and are applied for qualitatively and quantitatively exploring bird migration in time and space. Special emphasis is given on possible sources of bias and data uncertainty. As an example, we present some preliminary results for the statistical analysis of the movement patterns of the white stork *Ciconia ciconia*. First, we study the mean square displacement using ring recovery and satellite telemetry data. We find that the migratory dislocation for small time steps is dominated by ballistic movement. For longer time spans, however, the average travel distances are periodic functions of time, which underlines the importance of large-scale seasonal drift and recurrence in the migratory behaviour of white stork. Second, we investigate distributions of turning angles and are able to identify two different flight patterns - a migratory mode characterised by fast, directed movement with small turning angles during the time of migration and a resting mode characterised by slow, undirected movement with strong changes in flight directions especially during the breeding season. Data accuracy is, however, rather poor on small scales, indicating that great care should be taken of any consideration of turning angles. Based on these results we propose a conceptual quantitative model for global avian migration. The model describes a number of discrete breeding, resting and wintering habitats which are connected by fast, directed migratory flight routes. These seasonal transitions of individual birds between different habitat patches are described as unimodal circular functions of time. We are able to parametrise the transition probabilities with data from the white stork and find that here satellite telemetry data are relatively well suited. Finally, we discuss applications and possible extensions of our model.

4.2 Data for identifying bird migration routes

4.2.1 Ring recovery

The beginning of the systematic exploration of the routes of migratory birds has been marked by H. C. Mortensen in 1890 putting small metal rings on blackbirds *Sturnus vulgaris* and other species and receiving recoveries from distant locations. In the following century the ringing of birds spread and has become a highly important tool in exploring the behaviour and ecology of migratory birds all over the world (Bairlein, 2001). Many ringing centres were founded that provide standardised rings and ringing techniques and gather recovery data from a diligent crowd of volunteer bird watchers. Before that time bird migration was a rather mysterious phenomenon and theories trying to explain the absence of birds in a given area during a certain time of the year were rather speculative and controversial. The combination of positions of ring recoveries into so called "Atlases" gave a first picture of the spatial distribution of the

migration routes of different bird species (Berthold, 2001a). Techniques to mark individuals of different species of birds have become more and more refined during the years, e.g. putting large neck or wing bands with numbers or colour rings to be read off from a distance. Thus, birds did not need to be captured for identification any more. As the number of recoveries increased for many species it was also possible to get a crude idea of the temporal course of migration.

However, as recovery data is mainly obtained from amateur bird watchers and the public, the disposition to report observations of ringed birds and also the sampling efforts are not constant (Bairlein, 2001; Fiedler *et al.*, 2004). According to when and in what regions bird watching activity and awareness is high reporting rates and, thus, recovery probabilities differ greatly. In most areas in Africa, for example, reporting rates are suggested to be rather low because of low population densities and rather little awareness of the public of the subject matter of ringed birds and its importance. Therefore, ring data are typically strongly biased, a fact that makes it extremely difficult to quantify the process of migration from such data. Concluding, one can say that ring data has been successfully used in life history and behavioural studies and usually provides a good first idea about migration routes. However, the problems associated with sampling biases may pose serious difficulties when aiming for a quantitatively explicit description of migration patterns.

4.2.2 Satellite telemetry

For many years radio telemetry tags have been used to tackle questions of the behaviour and movement of birds (Priede & Swift, 1992). Individual birds are equipped with a radio transmitter which is sending a signal that can be directly tracked by an observer using a receiver. The transmitters are rather small and obtain positions sufficiently accurate. However, their tracking range is, depending on the vegetation, rather short (< 30 km), and the process of following a bird is considered very labour intensive.

When the ARGOS system (www.argos-system.com) was established in the 1970s for monitoring environmental processes with the help of satellites that orbit our earth, a new tool was available to explore the routes and characteristics of bird migration (Berthold *et al.*, 1997). Soon, battery powered transmitters were developed. They can be carried by individual birds anywhere on earth, for several months or even years sending signals to a satellite which is calculating the position of the bird and gives back longitude and latitude values to the ARGOS data base. From there researchers can retrieve information about the whereabouts of their studied animals. The technique of satellite telemetry is very powerful, having a wide range, not being very labour intensive and, especially since using solar batteries, even long lived. It has been widely applied in animal movement, orientation and migration studies (Jouventin & Weimerskirch, 1990; Berthold, 2001b; Thorup *et al.*, 2003; van den Bossche *et al.*, 2002), and its potential to e.g. resolve the discussion regarding flight navigation has been emphasised (Alerstam, 2006).

However, there are some drawbacks and problems that one has to consider when analysing satellite telemetry data. The equipment is very costly and the fee for using the satellites high. The transmitters are relatively heavy (> 10 g, i.e. only applicable to large bird species) and the localisations show large measurement error (especially for smaller transmitters) that is mainly

due to inconsistency in the frequency of the signal sent to the satellite (Fukuda *et al.*, 2004; Kaatz, 2004). Anyway, there is a huge and fast growing amount of satellite telemetry data around, its accuracy is appropriate for analysing long distance movements of migratory animals, and they are highly valuable for determining migration routes of several bird species, especially when compared with the accuracy and time resolution of ring-recapture data. Its accuracy may even be improved by selecting only high quality localisations analysing the additional information that is sent by the satellite (Kaatz, 1999).

4.2.3 New technologies

Recently, a number of new technologies have become available for animal tracking. Logging devices that receive geographical positions via GPS, for example, are very accurate in positioning, but still rather heavy, and the tags have to be retrieved from the animal to get hold of the stored data (Leick, 1995; Weimerskirch *et al.*, 2002). Another approach, geolocation via GLS is used to determine the geographical coordinates by measuring ambient light intensity (Hill, 1994). Compared to GPS the accuracy of GLS is rather poor, being in the range between 30 and 150 km, and the tags also store their data internally (Wilson, 2001). However, GLS tags can be very small (down to 1.5 g) so that even small bird species can be studied using them (www.antarctica.ac.uk).

Radar technology has been very promising for studying the behaviour of migrating birds (like e.g. swarming) in a refined area (Leshem & YomTov, 1996). However, it does not apply to fully explore the migratory routes of long-distance migrants, because immobile radar installations can only track individuals up to a distance of about 300 km. With active radar (e.g. using the cross-band transponder technology, www.earthspan.org) one tried to overcome these limitations and revolutionise radar technology. However, the earthspan project, for example, still faces large problems with detection and battery power. One of the most recent projects in the area of animal tracking is that proposed by Wikelski *et al.* (2007). They plan to launch a new satellite that can locate low power signals from very small (< 1 g) radio transmitters. If this technique renders feasible localisations obtained from such a satellite will be of extraordinary spatial precision.

To sum up, one can say that there is a wide range of different technologies under development which may revolutionise bird migration mapping in the immediate future. While currently, each method is still facing some problems it is important to be aware of these approaches for future research.

4.3 Movement analysis for the white stork

As discussed above, up to now a large amount of data has been accumulated on the whereabouts of individual birds and is ready for analysis. This is basically the case for ring and satellite telemetry data. In the following, we consider these data types and present a statistical analysis examining the movement of one long-distance migratory bird species, pointing out special features of the processes that govern its movement patterns.

4.3.1 Data sets

As an example of a long distance migrant we have chosen the white stork, because it is one of the best studied bird species in Europe. Some reasons for that are its prominent size and association with civilisation that make observation and captures relatively easy. Much ring recovery data is available for this species and to attach satellite telemetry transponders to several individuals was possible. The white stork is breeding in middle and eastern Europe and migrates either to southern Spain and west Africa or via Israel to eastern and southern Africa. Most individuals migrate each year, juveniles usually only partly, but there are also a few nonmigratory white stork individuals that do not leave their breeding grounds in Europe (Bauer *et al.*, 2005). When migrating the white stork flies only during the day, because it uses thermalling to fly long distances. It does not deposit fat prior to its migration and therefore, in some areas where the white stork rests overnight it may even stay to feed for a few days. Thus, its migration can be described as a stepping-stone process (Berthold *et al.*, 2006).

For the analysis of the movement patterns of the white stork we were able to obtain two different data sets. The first one is based on ring recoveries. The ringing centre “Vogelwarte Hiddensee” has systematically administered bird ringing and compiled recoveries of over 250 bird species in eastern Germany since 1964 (Köppen & Scheil, 2004). They provided us with ring recovery data of white stork that were ringed and recovered in the eastern part of Germany or that were ringed by different ringing centres and recovered in eastern Germany. Most of the recoveries concentrate on Germany, but several birds have also been observed in their African wintering areas. In total, the data set contains information about 7043 ringed white stork individuals that were at least recaptured one time. Because we are interested in the temporally explicit movement of the birds, we selected only the part of the data set of which the date of the ring recovery had been evaluated to be “correctly known”. In Germany some white stork are held in captivity or had been raised by humans and do not migrate. 35 such individuals had to be excluded from the data set, as we wanted to concentrate on strictly migratory birds. The new data set still held recoveries of 5,306 individuals; 35.8% of them were recovered three or more times, one was even recorded 53 times. Of the recovered white stork individuals only 3.8% have been observed in Africa. A spatial representation of the whereabouts of migratory white storks in the four seasons, as it is obtained from the selected ring data, is shown in Fig. 4.1. Juveniles (1-3 years old) are included in this data set, because they join the migration of the adults. However, in their first and maybe second summer, these birds possibly stay in the African habitats or just incompletely migrate north, what can be seen in Fig. 4.1 by the number of recoveries in Africa in June-August.

The second data set that we worked with is of satellite telemetry localisations. During the last 15 years Berthold (2001b) have attached satellite telemetry transmitters on a large number of individual white storks breeding in Germany and Poland. They were able to obtain several years of movement trajectories for many white stork individuals. One individual has been tracked for more than ten years, revealing much information about its life history and migration ecology (Berthold *et al.*, 2004). In the light of data quality evaluation, white stork conservation and habitat utilisation the data set has been elaborately analysed (Kaatz, 2004; van den Bossche *et al.*, 2002; Gerkmann, 2007). For our analysis we obtained 3 years (2001-2004) of data from three individual white storks, “Annamarie”, “Jonas” and “Prinzesschen”.

4.3 Movement analysis for the white stork

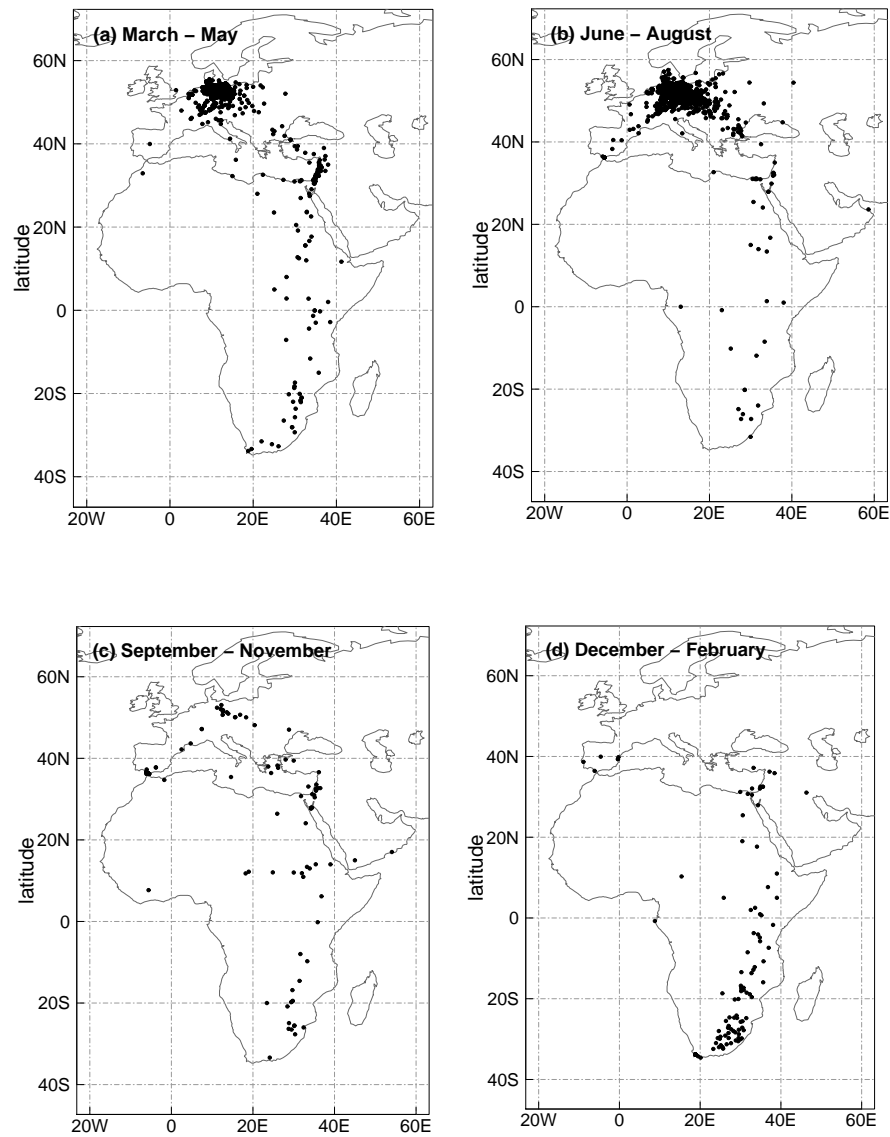


Figure 4.1: Migration routes of the white stork, obtained from a subset of ring recovery data of migratory individuals for 1964-2006. Positions of rings recovered in (a) March - May, (b) June - August, (c) September - November and (d) December - February are depicted. Juvenile birds are included here, some of them stay in the African habitats during the breeding period.

For “Annamarie” it contained 4,771, for “Jonas” 1,968 and for “Prinzesschen” 5,273 relocations. These localisations were spaced relatively evenly over the years, with the exception of poor sampling in May, June and July in the breeding areas. The data sets include individual

ring numbers, time, longitude and latitude of each recorded position and several technical and quality measures.

The satellites provide positions of the bird that is equipped with a transmitter quite regularly and rather independently of habitat characteristics and infrastructure. Thus, reporting rates can be assumed constant in time and space. Inaccuracy of the localisations can be detected using frequency stability and other measures provided by ARGOS with each transmitted position (Kaatz, 1999) and then out-selected. There is naturally a trade off between spatial accuracy and sample size which one has to consider carefully before selecting data for an analysis. Time of recovery is a very accurate measure, and sometimes, especially in more southerly areas, many localisations of the individual bird are available per day. Aware of the problem of poor localisation precision we considered frequency stability measures and chose only data which were evaluated to be accurate to an average distance of about 15 km (Kaatz, 1999). From this set we selected the best position per each day if available for each of the three individuals. Even though the data set was so strongly reduced, we chose to proceed in such a way in order to avoid major problems of spatial inaccuracy on small scales. However we have checked that our main results do not change if we keep all positions per day for analysis. The preprocessed data set consisted of 1538 localisations altogether that we worked with. Fig. 4.2 gives a visualisation of the movements of the three birds. Note, that even though we have data of only three individuals, they comprise three years each what means that in total our data cover 9 full spring and autumn migration paths. As white stork show individual as well as between year variability in migration routes (differential migration) (Berthold *et al.*, 2004) and because of the selection of three quite differentially migrating individuals we assume that a large amount of migration route variability of the white stork population is captured by our data set. This is supported by a comparison with Fig. 4.1, where one can see that the migratory routes from ringing and satellite telemetry data agree quite well.

4.3.2 Mean square displacement

One fundamental quantity in random walk theory to characterise the motion of a moving individual is the mean squared displacement. It can be calculated in the same way for both data sets, i.e. the ring recoveries and the satellite telemetry, since both data types yield longitude and latitude coordinates $(x(t_i), y(t_i))$ for the geographic position of individual birds at different time instances t_i . Using these coordinates we analysed the great circle distances

$$r(x_1, x_2, y_1, y_2) = \begin{cases} R \arccos(\text{sign}(s)) & : \text{ if } |s| > 1 \\ R \arccos(s) & : \text{ else} \end{cases} \quad (4.1)$$

where $R = 6378.388$ km is the earth radius and s is given by

$$s = \cos(y_1) \cos(x_1) \cos(y_2) \cos(x_2) + \cos(y_1) \sin(x_1) \cos(y_2) \sin(x_2) + \sin(y_1) \sin(y_2) \quad (4.2)$$

with $x_i = x(t_i)$ and $y_i = y(t_i)$ being the longitude and latitude coordinates in radians of each pair of positions of the same individual. The average distance $d(\Delta t)$ that is travelled in a time span $\Delta t = t_2 - t_1$ is then calculated by the mean square displacement as

$$d(\Delta t) = \sqrt{\langle r^2(x_1, x_2, y_1, y_2) \rangle}; \text{ for } t_2 - t_1 = \Delta t \quad (4.3)$$

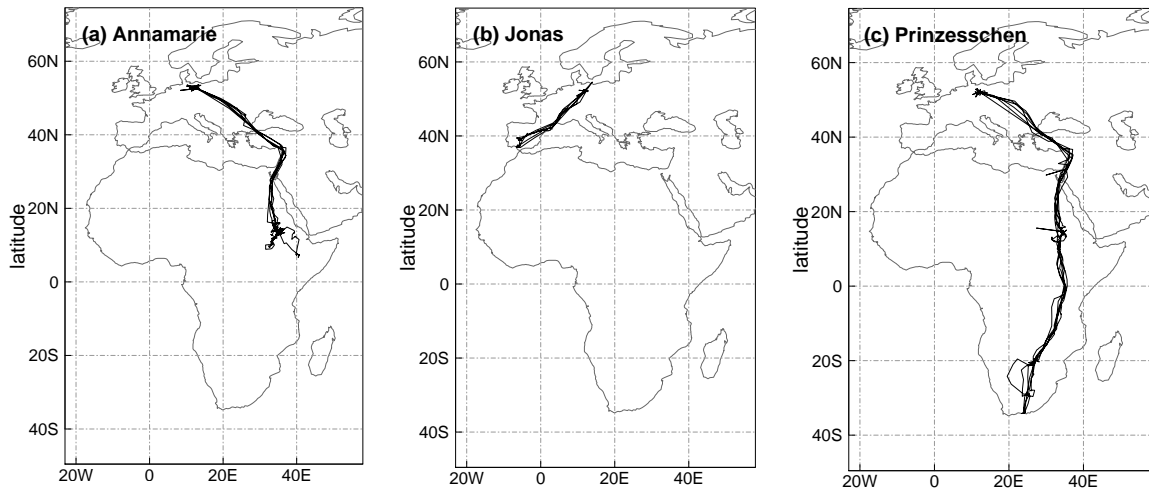


Figure 4.2: Migratory routes of the white stork, obtained from satellite telemetry data of three individuals from 1 August 2001 - 31 July 2004. The trajectories represent the migration flight routes of the two east migrating individuals (a) “Annamarie” and (c) “Prinzesschen” and the west migrant (b) “Jonas”.

where the brackets $\langle \rangle$ indicate averaging over all initial positions, t_1 , and all bird individuals in the data set.

For constructing graphical presentations of $d(\Delta t)$ we binned our data into equal 5 day (satellite telemetry data, Figs. 4.3, 4.4 a) or 30 day (ring recapture data, Fig. 4.5 a) time span intervals. We chose these as the minimum interval lengths so that each bin contained a reasonable amount of data (only 1% of the bins with less than 50 (for telemetry) or 10 (for rings) values of displacements). As can be expected, the number of displacement values calculated from satellite telemetry trajectories decreases for increasing Δt , ranging from 4,328 for $\Delta t \in (5, 10]$ days to 14 for $\Delta t \in (1,090, 1,095]$ days. The number of values for different Δt in the ring recapture data rather varies depending on the season, 3,531 to 207 per month in the breeding period and 89 to 6 values per month during wintering. This variation may somewhat decrease power in the statistics, but sample sizes are still sufficient to allow for interpretation. Furthermore, when varying the bin size between 2 and 50 days for the satellite telemetry data the main characteristics of the plots do not change.

Usually in diffusion problems the average travel distance is an increasing function of the time lag and scales as a power law $d \sim (\Delta t)^\alpha$. This relationship can be conveniently depicted in a log-log-plot of $d(\Delta t)$ against all values of Δt , where the exponent α is obtained as the slope of a least-square fit in the scaling region. This is shown in Fig. 4.3 for the satellite telemetry data of the three white stork individuals. For small time differences of $\Delta t < 6$ months the average travel distance roughly scales with an exponent of $\alpha \approx 0.7$ and the examined motion clearly is superdiffusive. More detailed, for very small time distances of $\Delta t < 30$ days the motion is nearly ballistic ($\alpha \approx 1$), while in the intermediate range of $30 \text{ days} < \Delta t < 6 \text{ months}$ it may be

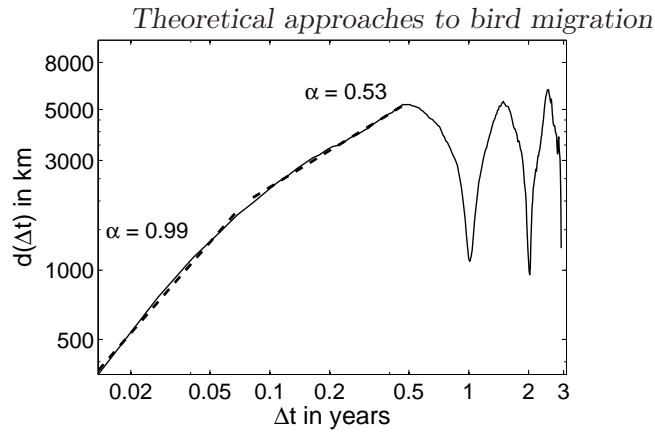


Figure 4.3: Average travel distance $d(\Delta t)$ of the white stork on their migratory path obtained from satellite telemetry data, binned into equal 5 day time intervals. This visualisation as a log-log plot clearly indicates the superdiffusive character of the migratory movement for small time differences. It may even be split into one part of ballistic ($\alpha \approx 1$) and another part of diffusive ($\alpha \approx 0.5$) motion.

best described as diffusive ($\alpha \approx 0.5$). Similar transitions from ballistic to diffusive behaviour are frequently observed in random walks with short term correlations.

Additionally, as can be expected from the seasonally driven character of migration, for larger time scales of $\Delta t > 6$ months $d(\Delta t)$ becomes a periodic function of time (Fig. 4.4 a). On average each bird returns to its previous position after one year. When aggregating the displacements according to normalised time differences in years, $\Delta\tau = \Delta t \bmod 1$, we can observe this seasonality in form of a hump shaped curve (Figs. 4.4 b). Its general form shows that on their migratory path the birds' positions are on average maximally separated by a distance of about 6000 km at a time difference of about 6 months. This is basically a result of the dominating long distances between the birds' breeding and wintering areas. The slopes of the curve are determined by the speed of migration in spring and autumn, respectively, and the width of the peak's plateau indicates how long the birds stay in their wintering and breeding areas. It is interesting to note the changes of scaling detected in Fig. 4.3 also in Fig. 4.4 b at $\Delta t \approx 1.5$ months and, because of symmetry, at $\Delta t \approx 10.5$ months.

Similar dynamics as for the satellite telemetry data can be deduced from ring recapture data, where $d(\Delta t)$ is obtained from rather sparse trajectories of a very large number of birds over many years (Fig. 4.5). As we have mentioned above, the ring recovery data base covers a time interval of more than 50 years. This data set contains many recoveries of individual birds with a time interval Δt of more than 20 years. However, the frequency of such entries decays for larger time intervals, which results in poor statistics. Therefore, in Fig. 4.5 we have plotted $d(\Delta t)$ only for time lags up to 10 years. Nevertheless, ring recovery data reflect much larger time spans than the three years of satellite telemetry data that we obtained. In this sense the two data sets can be regarded to be "orthogonal". While with satellite telemetry, by averaging over many time instances of only three individual birds, $d(\Delta t)$ is basically calculated as a time average, in the case of the ring recoveries, by averaging over a large number of individuals, it rather corresponds to a sample average.

Even though the ring data fits well into the patterns obtained from satellite telemetry data

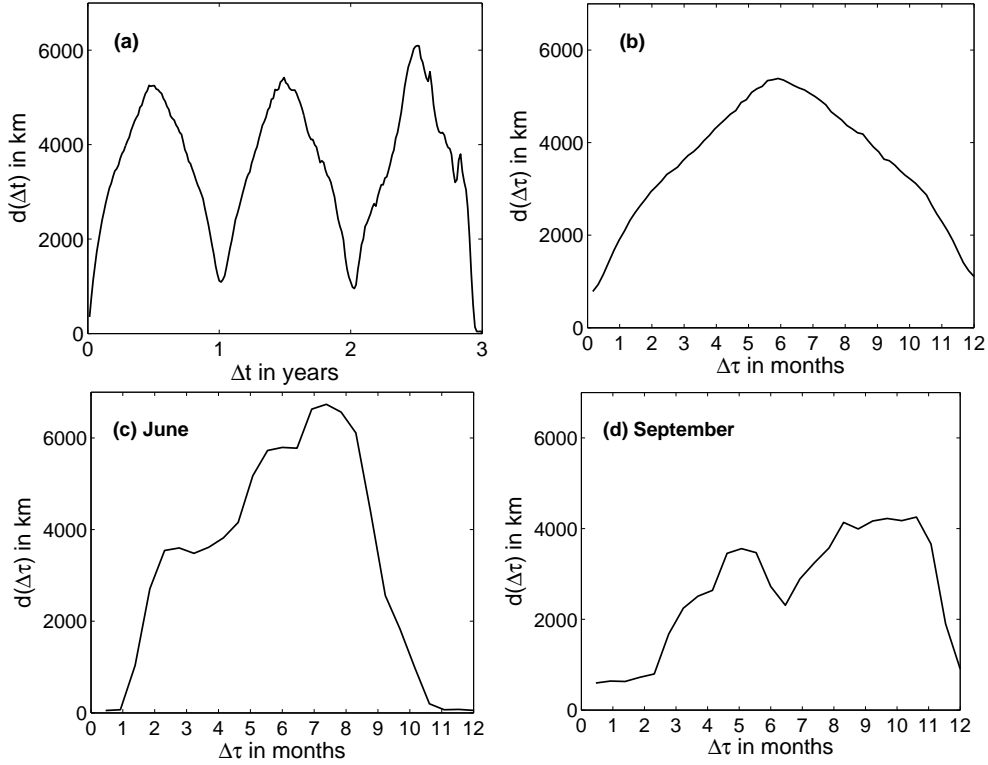


Figure 4.4: Average travel distance d , from Eq. (4.3), of the white stork based on satellite telemetry data as a function of (a) the time differences $\Delta t = t_2 - t_1$ and (b) the time differences modulo one year $\Delta\tau$. Similar to Fig. 4.3 note the change of scaling at $\Delta\tau = 1.5$ months and $\Delta\tau = 10.5$ months in (b). For further investigation the data are pooled by the displacement starting times τ_1 : (c) the middle of the breeding period (18 June – 1 July) and (d) the autumn migration period (10 - 23 September). Thus, (c) represents the consecutive displacement of the storks from their breeding area to wintering places and (d) the displacement from some resting site on their autumn migratory route to the wintering grounds, then back, close by and to the breeding areas. The data was binned into equal 5 day time intervals.

and the overall form of the average travel distances $d(\Delta t)$ agree in both data sets, there are some marked differences in the shape of the curves. While the rise of $d(\Delta\tau)$ is nearly linear for the satellite telemetry data (Fig. 4.4 b), in the case of the ring recovery data we find abrupt changes at $\Delta t = 5$ months and at $\Delta t = 8$ months (Fig. 4.5 b). We can attribute these discrepancies to the fact that the ring recovery data are not uniformly sampled through all seasons (see also Fig. 4.1). Individual birds which travel the full distance from Northern Europe to South Africa are most strongly sampled in the wintering and breeding areas at the extreme ends of the migratory path. This results in a sudden peak of $d(\Delta\tau)$ when $\Delta\tau \approx 6$ month. This edge in the curve of typical travel distances is absent in the satellite telemetry data which are uniformly sampled. Thus we believe that in this case the data from satellite telemetry reveal a better picture of the white stork movement patterns.

Interestingly, in both data sets the oscillations of $d(\Delta t)$ are not damped for long time intervals. Instead, the average travel distance always returns to a small value close to zero

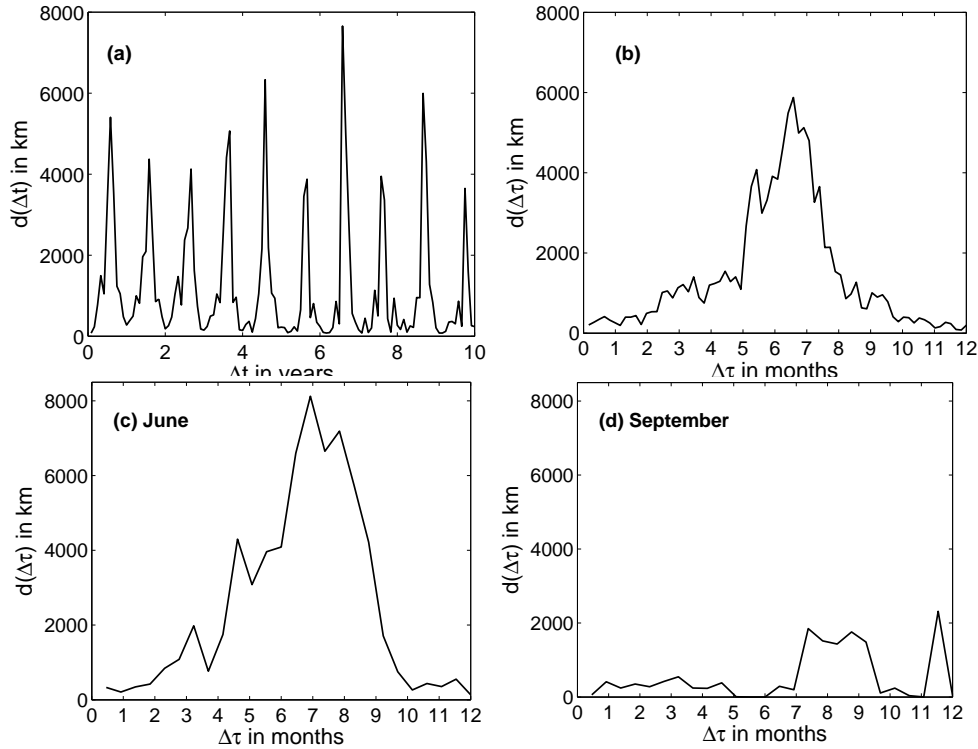


Figure 4.5: Average travel distance d of the white stork as in Fig.4.4 but based on ring recovery data. The displacements were binned into equal 30 day time intervals for (a), but resolved more finely for the aggregated plots: 5 day intervals for (b) and 14 day intervals for (c) and (d).

when Δt is a multiple of one year and these minima do not increase with the number of years between two measurement points. This observation confirms the legendary spatial memory and faithfulness to its nesting place of the white stork (Creutz, 1985). In the ring recovery data several birds are located at exactly the same nest, breeding, for many years. The satellite telemetry data underline this notion. The minimum displacements $d_{min}(\Delta t)$ for Δt being multiples of around one year, are of the orders of quality uncertainty of this data set (“Annamarie”: $\langle d_{min}(\Delta t \approx k) \rangle = 272 \text{ m}$, “Jonas”: $\langle d_{min}(\Delta t \approx k) \rangle = 121 \text{ m}$, “Prinzesschen”: $\langle d_{min}(\Delta t \approx k) \rangle = 303 \text{ m}$, for $k = 1, 2, 3$ years). Thus, individual birds come very close to their previous resting, and especially breeding, positions. As shown in Fig. 4.5 this memory lasts at least 10 years, i.e., the length of our time series. However, also in the full time series including many ring recoveries with time intervals well over 20 years (not plotted) this tendency prevails. Indeed, the longest time interval of recovery in the data base was obtained for a bird that was ringed in Germany as a hatchling and recovered 29 years later only 17 km from its birth place. This suggests that the spatial memory of white stork individuals extends over the full life span of the birds.

Our analysis has shown that many characteristics of bird migration can be obtained from and are reflected by an inspection of the mean square displacement. However, it also becomes

clear that in a seasonally driven process like migration such standard techniques from random walk theory should not be applied naively. It is advisable to modify them somewhat and interpret results carefully. A convenient way to obtain a visual impression of the seasonal effect on displacement distances is to split the data according to the starting month of each displacement $d(\tau_1, \Delta\tau)$ (as examples see subplots c and d in Figs. 4.4, 4.5). Here τ_1 is the season of the first location of the bird displacement, $\tau_1 = t_1 \bmod 1$. In such seasonally pooled presentations the times of breeding and migration can clearly be distinguished, which in this way gives a better indication of how the averaged quantity $d(\Delta\tau)$ is assembled from very different processes during the different stages of bird migration. In Figs. 4.4 c/4.5 c and 4.4 d/4.5d respectively, we contrast such pooled displacement plots for trajectories starting in the breeding area (c) and sometime in the autumn migration (d). In the former ones one sees the difference in migration velocity between the autumn migration (shallowly increasing slope) and the spring migration (steeply decreasing slope), whereas the latter figures more clearly indicate the back-and-forth character of the migratory movement with their bimodal shape.

By comparison of the results for both the ring recapture and satellite telemetry data one obtains insight into the accuracy and quality of the two data types. Additionally, our investigation reveals the advantages to combine such two “orthogonal” data sets for analysis. While one data set has a very fine resolution over a short time span, the other covers larger time distances at the cost of poor resolution and biased sampling. By combining both data sources we obtain a more complete picture of the migratory movement from two different points of view.

4.3.3 Turning angles

Another important statistics to consider in animal movement analysis is the distribution of turning angles. It indicates the character of the movement, i.e. how directed or curved it is. For migratory birds one would expect that during migration flight is very directed and turning angles are relatively small and that during resting, breeding and wintering turning angles are larger and more or less random in the process of foraging. Analysis of turning angles has been carried out to examine the motion of several aquatic and terrestrial organisms, e.g. daphnia (Komin *et al.*, 2004), barnacle larvae (Pasternak *et al.*, 2004), fish (Gutenkunst *et al.*, 2007), marine mammals (Bailey & Thompson, 2006), ungulates (Morales *et al.*, 2004), birds (Nolet & Mooij, 2002) and others (Turchin, 1998), usually describing their foraging behaviour or dispersal. However to our knowledge, the distribution of turning angles has not yet been investigated for the year round travel and motion of migratory birds.

Any reasonable measurement of turning angles requires that the time differences between successive positions are small. In our case this is only valid for satellite telemetry data, and therefore we have disregarded ring recovery data for the investigation of turning angles. For the analysis, we selected one best position of each individual per each day that data was available for. From this reduced time series we then selected the triples of positions, whose consecutive time differences were not larger than one week, and calculated a movement direction

$$\varphi_i = \arctan \left(\frac{\Delta y_i}{\Delta x_i} \right), \quad (4.4)$$

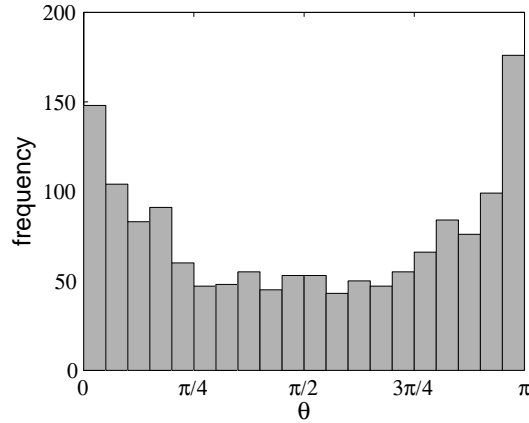


Figure 4.6: Histogram $P(\theta)$ of turning angles θ along the migratory path of the white stork accumulated over the whole year, based on satellite telemetry data.

where $\Delta y_i = |y_{i+1} - y_i|$ and $\Delta x_i = |x_{i+1} - x_i|$ describe the differences of geographic latitude and longitude coordinates. Turning angles θ_i are then obtained as the differences of two consecutive movement directions

$$\theta_i = \varphi_{i+1} - \varphi_i. \quad (4.5)$$

The (non-normalised) distribution $P(\theta)$ of turning angles for the migratory movement of the white stork is shown in Fig. 4.6. It clearly shows a maximum for small turning angles $\theta \approx 0$ and is decaying for larger angles. However unexpectedly, there is a second maximum for large turning angles of $\theta \approx \pi$. Thus, one can suggest that the flight behaviour of migratory birds splits into two characteristic modes. The first mode describes directed flight where the bird changes its direction only little between the days. In contrast, the second mode is characterised by large turning angles of about 180° corresponding to ongoing reversions of the flight direction.

To obtain more insight into this phenomenon we have resorted to the same approach as in the previous section and have split the histogram according to the starting month τ_1 . Inspection of the pooled distributions $P(\tau_1, \theta)$ in Fig. 4.7 reveals that the small turning angles correspond to the typical times of migration (March, August) and that the large turning angles occur mainly during the breeding and wintering seasons. Thus, the cluster of small turning angles results as expected from the directed migration flights in spring and autumn. Similarly, the cluster of large angles can be interpreted in the way that in their breeding and wintering grounds individuals have a central nest or sleeping tree, from where they fly back and forth to a small number of foraging areas. To examine this matter in more detail, in Fig. 4.8 we present a number of two-monthly trajectories for a single white stork and year. Visual inspection of the flight trajectories confirms the existence of these two different migratory flight modes.

We note though that an additional contribution to the cluster of large turning angles might possibly arise from measurement error since the small scale movement in the resting areas contains some localisations of very low accuracy, whose incorporation into the trajectories can produce large spurious turning angles. Indeed, a closer examination of the localisation errors in the breeding and wintering areas (see Figs. 4.8 b and 4.8 d) shows that the positions at the

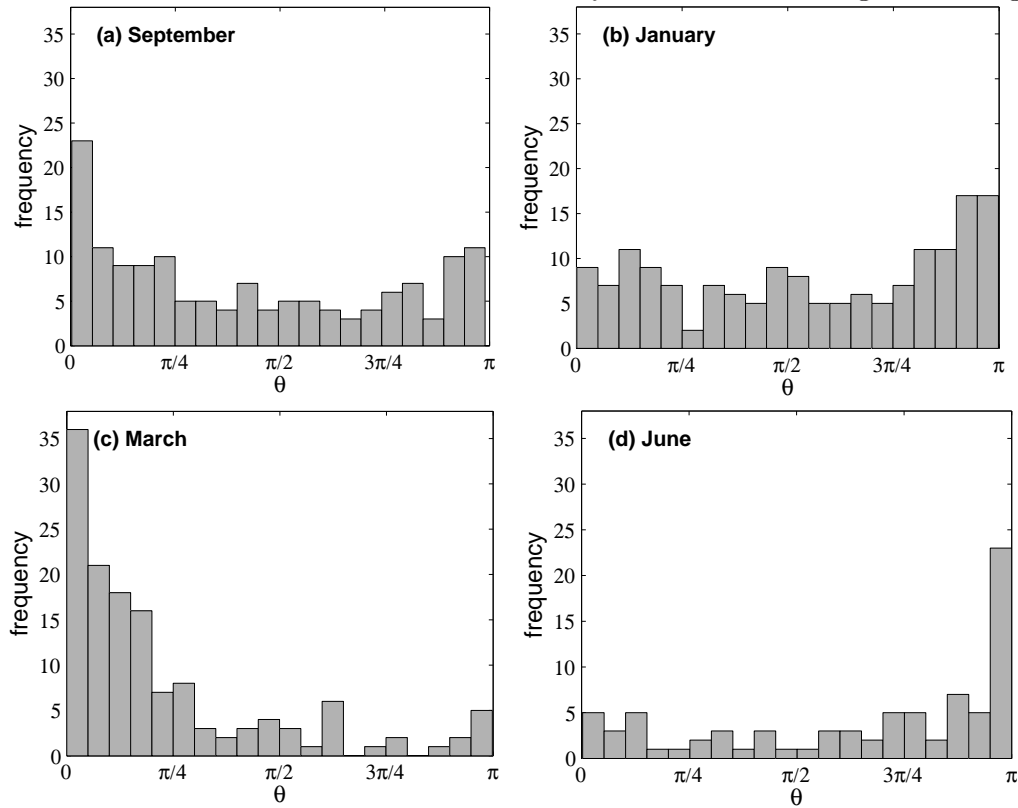


Figure 4.7: Pooled distributions of turning angles for the white stork. In contrast to Fig. 4.6 here the histograms $P(\theta)$ have been split according to the time of the year τ , for (a) September - autumn migration, (b) January - wintering, (c) March - spring migration and (d) June - breeding.

end of the “spikes” that determine the large turning angles are of rather low accuracy of on average ± 15 km, opposed to only ± 3 km for central positions. However, the typical distance between the two extreme ends of a back-and-forth step is larger than 15 km, which makes it unlikely that imprecise measurement can be the sole explanation for the peak around $\theta \approx \pi$. Another phenomenon where large turning angles may occur even during migration is reverse migration, where individual birds reverse their migration direction for a few stages due to some cause that has not yet fully been understood (Berthold, 2001a; Schüz *et al.*, 1971).

4.4 Quantitative modelling of bird migration

As the result of our movement analysis, a typical path of a migrating bird can be described as an alternating sequence of directed (moving) and undirected (resting) stages. This is consistent with observations in other studies, where the migration of birds has been described as a succession of flying and stopover (Hedenström & Ålerstam, 1997). In the following we will use these ideas for setting up a conceptual model to describe the seasonal process of migration.

The main idea of our model is to replace the spatially continuous trajectory of a migrating

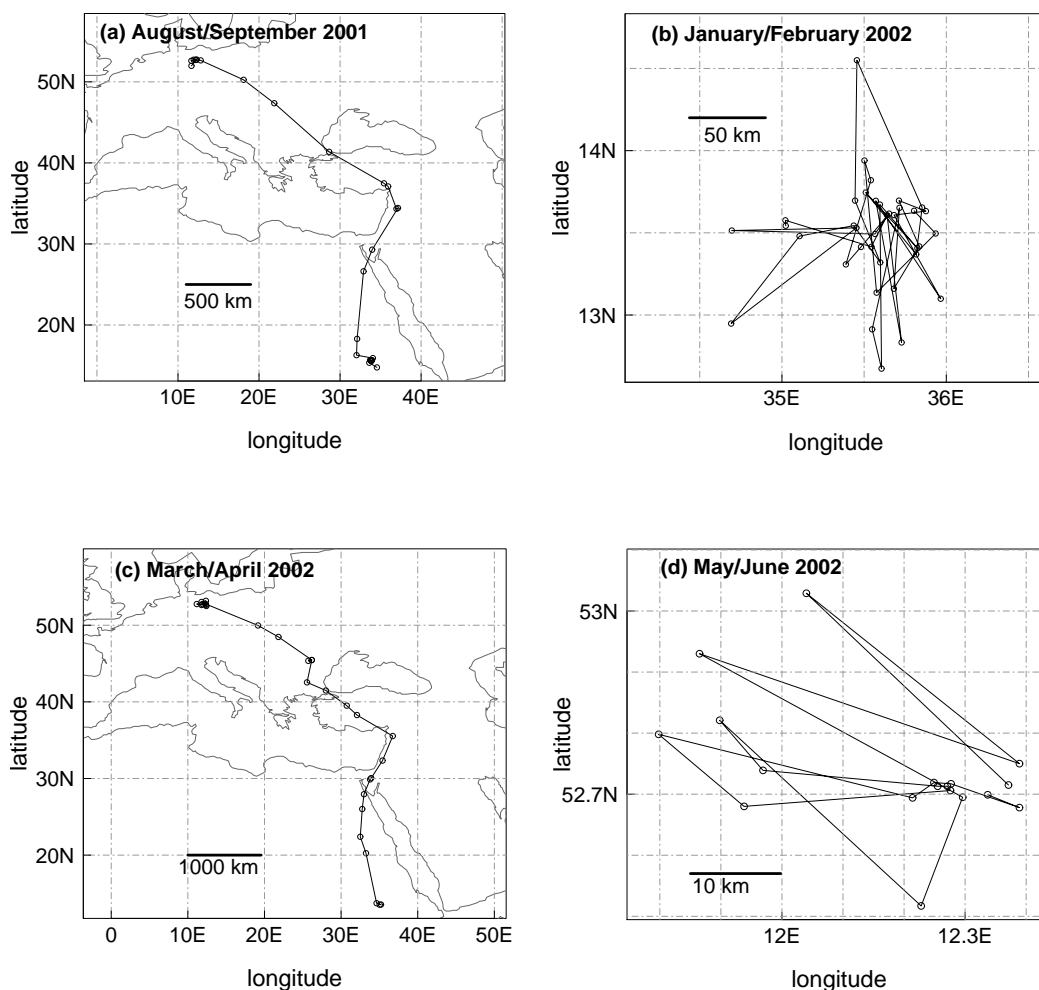


Figure 4.8: Selection of typical trajectories of two months along the migratory path of the white stork “Annamarie” for (a) August/September 2001 (autumn migration), (b) January/February 2002 (wintering), (c) March/April 2002 (spring migration), (d) May/June 2002 (breeding).

bird by a jumping process on a network of a number of discrete habitat patches. Thus, the model describes the seasonally driven long-distance migration of birds on a discrete network of m species specific breeding, moulting, resting and wintering places. These local regions, the nodes of the network, are fixed in time and are connected to one another by directed flight routes. Accounting for differential migration we assume that avian migration does not take place only on a linear chain of patches, but instead permit branching on the network of possible flight connections.

Consider the time varying density $n_i(t)$ of N birds that are present in area $i = 1, \dots, m$ at

time t . The densities are normalised as

$$\sum_{i=1}^m n_i(t) = N. \quad (4.6)$$

Thus, we assume that the total number of birds, N , stays constant and we neglect population growth during the time scale of our analysis. The dynamics of the migration is described as a Markovian jumping process

$$n_i(t+1) = \sum_{j:j \neq i} p_{ij}(\tau) n_j(t) + p_{ii}(\tau) n_i(t), \quad (4.7)$$

where we assume that the birds jump from patch j to patch i according to certain transition probabilities p_{ij} which do not depend on previous jumping events or bird densities. So, in our model individual birds have no memory and choose the jumping routes every year anew from the same probability distribution. Furthermore, each bird in a given area will continue its migratory path independently of the behaviour of the other birds, thus, we neglect swarming effects.

The novel aspect of our model is that the transition probabilities p_{ij} are not constant within the year, but instead are periodic functions of the season $p_{ij} = p_{ij}(\tau)$ (however, they are assumed to be constant between the years). Thus, the transition probabilities are described by a circular function $g(\tau)$ of the season τ that quantifies the intensity of birds flying between different patches

$$p_{ij}(\tau) = g(h_{ij}, \phi_{ij}, \sigma_{ij}, \tau). \quad (4.8)$$

Each jumping probability $p_{ij}(\tau)$ is characterised by the mean season when jumping takes place (the phase) ϕ_{ij} , the maximum jumping intensity (the amplitude) h_{ij} and a typical length of the time interval during which jumping occurs (the width) σ_{ij} . An additional assumption that simplifies the migration jumping process in a reasonable manner is to set the jumping durations equal for all connections of the network $\sigma_{ij} = \sigma = \text{const}$ (see Fig. 4.9). Then each link is characterised by a complex number $z = h e^{i\phi}$, which describes the amplitude h and the phase ϕ of the transition.

Because it is a canonical distribution to describe unimodal, circular data, we used a generalised von-Mises distribution function as a first approximation of the functional form of g

$$g(h_{ij}, \phi_{ij}, \sigma, \tau) = h_{ij} \exp \left[\frac{\cos(\tau - \phi_{ij}) - 1}{\sigma^2} \right]. \quad (4.9)$$

For small values of σ this form is nearly a Gaussian with a standard deviation of σ , while in the limit of large σ this approaches a uniform distribution $g(\tau) = h_{ij}$. Furthermore, by using the von-Mises distribution we also assume that the jumping probabilities are symmetrical in time, which may not be the case in real bird migration.

A systematic presentation of the resulting model dynamics will be presented elsewhere. Here we are concerned with the problem of finding an optimal parametrisation in accord with real observational data, i.e. we are interested to estimate the transition rates $p_{ij}(\tau)$ for the migratory movement of the white stork. Data ideal for fitting the described model need to

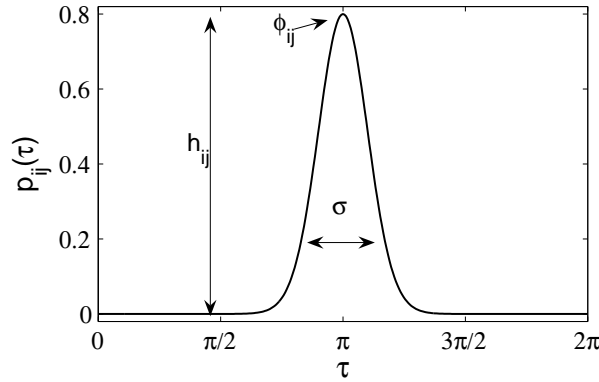


Figure 4.9: Typical form of the circular jumping probability function $g(\tau)$, here as a modified form of the von-Mises distribution function, Eq. (4.9), with parameters $h_{ij} = 0.8$, $\phi_{ij} = \pi$ and $\sigma = 0.3$.

be explicit, accurate and consistent in time and space. As explained above satellite telemetry data, rather than ring recovery data, fulfil these demands relatively well. Using telemetry data one can follow each individual bird's trajectory on its migratory path to estimate the transition times between predetermined resting, breeding or wintering places for each year.

In an exemplary way we have calculated a rather crude network from the satellite telemetry data of the three white stork individuals (Fig. 4.10). As the nodes of the network we defined the four main breeding and wintering areas of the species. There are surely additional resting areas used by the white stork (Gerkmann, 2007; Berthold *et al.*, 2006), however, here we want to keep the presentation as simple as possible. For each of the realised directed transitions we recorded the dates of leaving the one area and arriving in the other. The jumping times were defined as the mean time between the two events. By histogramming these data we were able to estimate the time-dependent flows $J_{ij}(t) = p_{ij}(\tau)n_j(t)$, from the model Eq. (4.7), between each pair of patches. After normalisation with the bird densities $n_j(t)$ this yields the desired transition probabilities $p_{ij}(\tau)$ (see Fig. 4.10, grey bars). We then, by hand, fitted a von-Mises functional form to these records, thus, having parametrised our model. Because of the low sample size we consider it not reasonable to use any more advanced fitting techniques, as e.g. maximum likelihood. They should be applied when larger data sets become available.

The transition probabilities in Fig. 4.10 provide valuable information about the spatio-temporal dynamics of the migration process. Despite the shortness of transition data all phases ϕ_{ij} have reasonably been reconstructed. Note that the phases ϕ_{ij} are ordered along the migratory path, in concord to the naive expectation. Further, the transitions are rather sharp and can be assumed to have the same width for all transitions. The estimated width equals $\sigma = 0.12$ which corresponds to about 7 days. The small sample size does not allow a sensible estimation of the transition amplitudes, yet. Nevertheless, the crude fit in Fig. 4.10 appears already to be quite reasonable.

Interestingly, the transition amplitudes of the north going birds are all more or less equally of a large value. This reflects a basic north-south symmetry breaking in the migratory route of the white stork: In our data base the birds always return to their breeding area in the north (area 2), but in the southgoing direction their routes split into different resting and wintering areas.

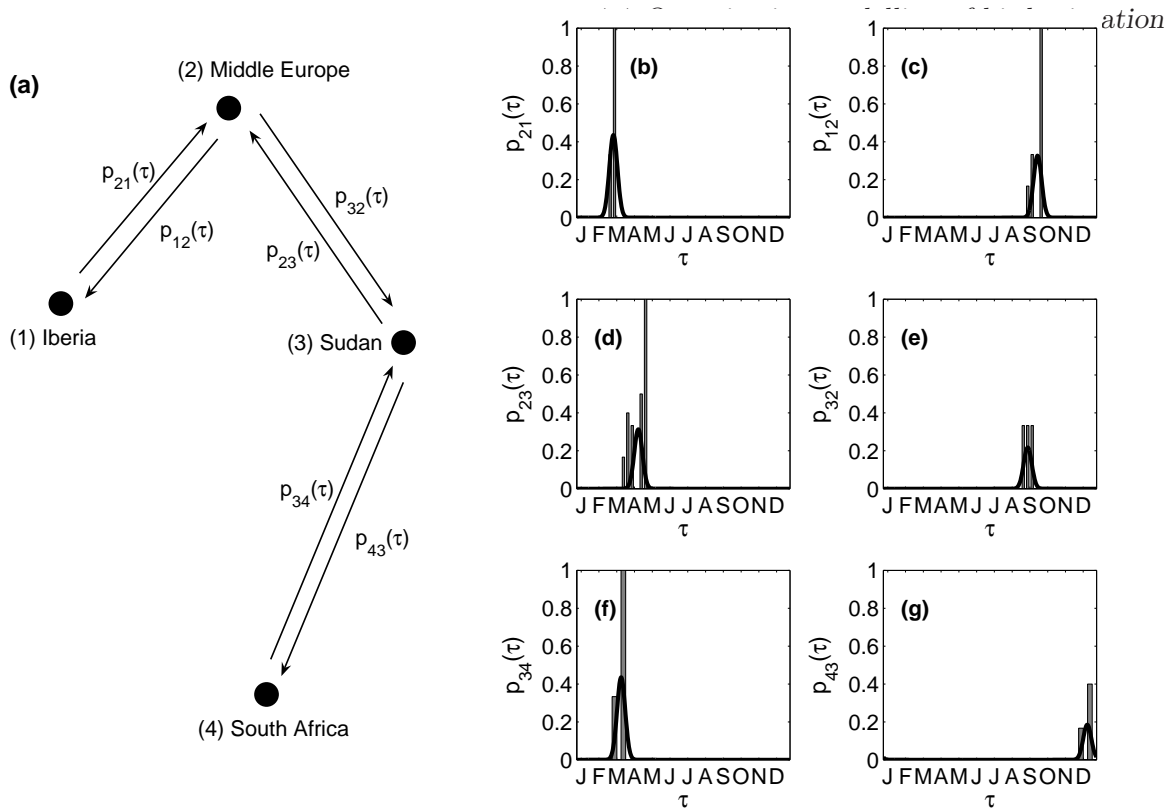


Figure 4.10: Network model for white stork migration. (a) The breeding area (2) Middle Europe is connected to the main wintering areas (1) the Iberian Penninsular, (3) the Sudan area and (4) South Africa by seasonally driven jumping probabilities $p_{ij}(\tau)$. (b) - (g) show the transition probabilities $p_{ij}(\tau)$ (represented as grey bars) that are fit by the von-Mises functional form, Eq. (4.9), (black solid lines) for north going (left column) and south going (right column) birds obtained from the satellite telemetry data.

Thus, all north going transitions have a large amplitude ($h_{21} \approx h_{23} \approx h_{34}$) which insures that all individuals jump within a short time interval. In contrast, when going south the birds have to decide on the western or eastern route, i.e. either to overwinter on the Iberian Penninsular (with probability p_{12}) or to fly to the Sudan (with probability p_{32}). Note that the transition to the Sudan area is slightly advanced $\phi_{32} < \phi_{12}$. Therefore, the transition probability p_{32} has to be small, to ensure that some birds stay to use the western route, and p_{12} must be large, so that no birds remain in Middle Europe over the winter. In this way we can explain the observed relation of the transition amplitudes $h_{32} < h_{12}$. A second branching arises in the Sudan region (area 3) where the birds have to decide whether to continue their migratory path to South Africa (area 4) or to stay. Consequently, the transition amplitude h_{43} is rather small. This simple example already shows that it is not possible to read off the branching ratios simply from the transition parameters, if the transitions are not synchronised in season. Therefore, we expect that the seasonally driven model has nontrivial properties, especially for more complex network topologies.

As one can also observe for the white stork many migratory bird species exhibit differential

migration (Gauthreaux, 1982). This means that different populations, age groups or sexes of a species migrate on different routes, by varying time schedules and maybe even to different areas. Depending on how pronounced this effect is in the considered species a large amount of data is needed to obtain a sufficient picture of the species' migration patterns. The sample size here is rather small but well selected, so that we are confident that our model gives a good first approximation of the general migratory pattern of the white stork.

In principle one could also use ring data for parametrisation of the presented network migration model. However, as has already been pointed out, the major shortcoming of ring data is the difference in observation effort and reporting. For quantitative estimation (that we attempt here) the reporting rates would need to be quantified. We expect them to correlate with human population density or social indices such as alphabetisation or economic income, but any such relationship may also be illusory. Furthermore, ring data provide only the number of birds that have visited certain places during the year, but not exact time instances of the transitions between consecutive resting places. Therefore, jumping probabilities would have to be estimated from the observed densities by optimisation techniques like e.g. linear programming or simulated annealing. Hence, the use of ring data for estimating the transition probabilities of our bird migration model in time and space is possible, but poses many problems and it may render more successful to give other kinds of data the preference.

4.5 Conclusions

We strongly agree with and want to promote the notion of Bairlein (2003) that the analysis of recovery data and modelling of migration patterns in time and space need increased emphasis in ecological research. On the one hand large amounts of bird observational data are collected on an ever increasing rate and quality. To gain insight into patterns and characteristics of such data they have to be analysed in much detail using statistical methods. On the other hand several research groups specialise on only few species and step by step try to disentangle its behaviour and ecology. As this is undoubtedly important we think it, however, crucial not to lose the more general view on ecology and e.g. bird migration mechanisms. Here the development of time and space explicit models may be of high value. With sensibly constructed models one could examine the influence of parameters as for example the peak time or interval length of leaving the breeding areas on the migration process. One could thus make predictions about the consequences of global factors such as climate change and its effects on the migration behaviour of several species and by simulations and model exploration obtain general insight into the underlying mechanisms. In the long run it may even be possible to make comparisons with the movement of other organism types.

The above described network model of bird migration makes many simplifying assumptions. In particular, it disregards any variability of the migration process between years. This means that environmental factors are assumed to have no influence on survival, timing and the course of migration. We are certainly aware that this is not the case in reality (e.g. Zink & Bairlein, 1995; Kölzsch *et al.*, 2007), especially in the light of the recent climate change. In this sense the presented ideas represent only a first step in the direction of setting up a full-fledged, conceptual bird migration model. Once the general properties of such a model are fully

understood, other environmental covariates can be incorporated.

In this article we have summarised the kind of data that are available for quantitative studies of bird migration. Our aim is to encourage the systematic, theoretical study of this process, but also to call for caution regarding the use of such data. Of course, any kind of data needs to be carefully considered before it can be incorporated into a model. In bird migration we face the situation that much data has already been collected, and the advantages and disadvantages of them have to be taken into account before any decision can be made regarding its analysis. In theoretical modelling it is always possible to just explore a model and its properties per se, but only by parametrisation with sensible data one will be able to conclude sensible information about the ecological processes acting in avian migration.

Acknowledgements

We are grateful to the ringing centre Vogelwarte Hiddensee in Stralsund and the Vogelwarte Radolfzell at the Max-Planck-Institute for Ornithology for providing us with ring recovery and satellite telemetry data that was collected by numerous professional and amateur ornithologists for many years. We want to thank Ulrich Köppen and Birgit Gerkmann for discussing the data analysis and results. Special thanks to Michael Kaatz for elaborate discussions about data quality and Ralf Tönjes for help with model formulation. This work was funded by the VW-Stiftung and the BMBF.

5 Paper IV.

Lévy flights in bird movement after all?

Andrea Kölzsch, Ulrich Köppen, Wolfgang Fiedler, Franz Bairlein and Bernd Blasius;
submitted to Nature

Arising from: A.M. Edwards, R.A. Phillips, N.W. Watkins, M.P. Freeman, E.J. Murphy, V. Afanasyev, S.V. Buldyrev, M.G.E. da Luz, E.P. Pappas, H.E. Stanley and G.M. Viswanathan. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*. **449**, 1004-1048 (2007).

Following the seminal study on wandering albatross *Diomedea exulans* (Viswanathan *et al.*, 1996), during the last decade movement data of many different animal species (Viswanathan *et al.*, 1999; Marell *et al.*, 2002; Ramos-Fernandez *et al.*, 2004; Sims *et al.*, 2008; Viswanathan *et al.*, 2008) and humans (Brockmann *et al.*, 2006; González *et al.*, 2008) were analysed and indications of Lévy flight behaviour were found. However, recently Edwards *et al.* (2007) applied rigorous statistical methods and came to different conclusions. Not only did they clearly rule out Lévy flight behaviour in the foraging movement of the wandering albatross and other reanalysed data sets, but they question the strength of the empirical evidence for biological Lévy flights in general. Here we study short-term individual bird displacements based on ring-recapture data using model selection statistical methods as suggested by Edwards *et al.* (2007) We find strong indications of Lévy flights for all analysed bird species, however, concerning their large scale migration movement rather than foraging.

Ring-recapture data have been collected for more than 100 years (Berthold, 2001a), providing a large and variable sample of individual displacement events (Kölzsch & Blasius, 2008). For this analysis we chose five ecologically rather different species with data of reasonable sample sizes (core data consists of 96,835 ring recaptures): the long-distance migrating eastern population of the white stork (*Ciconia ciconia*), the barn swallow (*Hirundo rustica*), the medium-distance migratory chiffchaff (*Phylloscopus collybita*), the in central Europe almost completely sedentary mute swan (*Cygnus olor*) and the mallard (*Anas platyrhynchos*) that has short-distance migratory as well as sedentary populations in Europe.

For all tested bird species the density of displacement distances decays with a fat tail, reminiscent of a power law $P(d) \sim d^{-\beta}$ for $d_{min} \leq d \leq d_{max}$ in intervals of 2-3 orders of magnitude (see Fig. 5.1). We found good agreement for the power law exponents fitted by linear regression (b) and maximum likelihood (β). The measured exponents reflect each species' migratory behaviour: Largest exponents are found for the sedentary mute swan ($\beta = 2.42$), whereas small exponents are found for the migratory chiffchaff and barn swallow ($1.25 \leq \beta \leq 1.32$), and the white stork ($\beta = 1.48$). The mallard, being partially migratory, assumes an

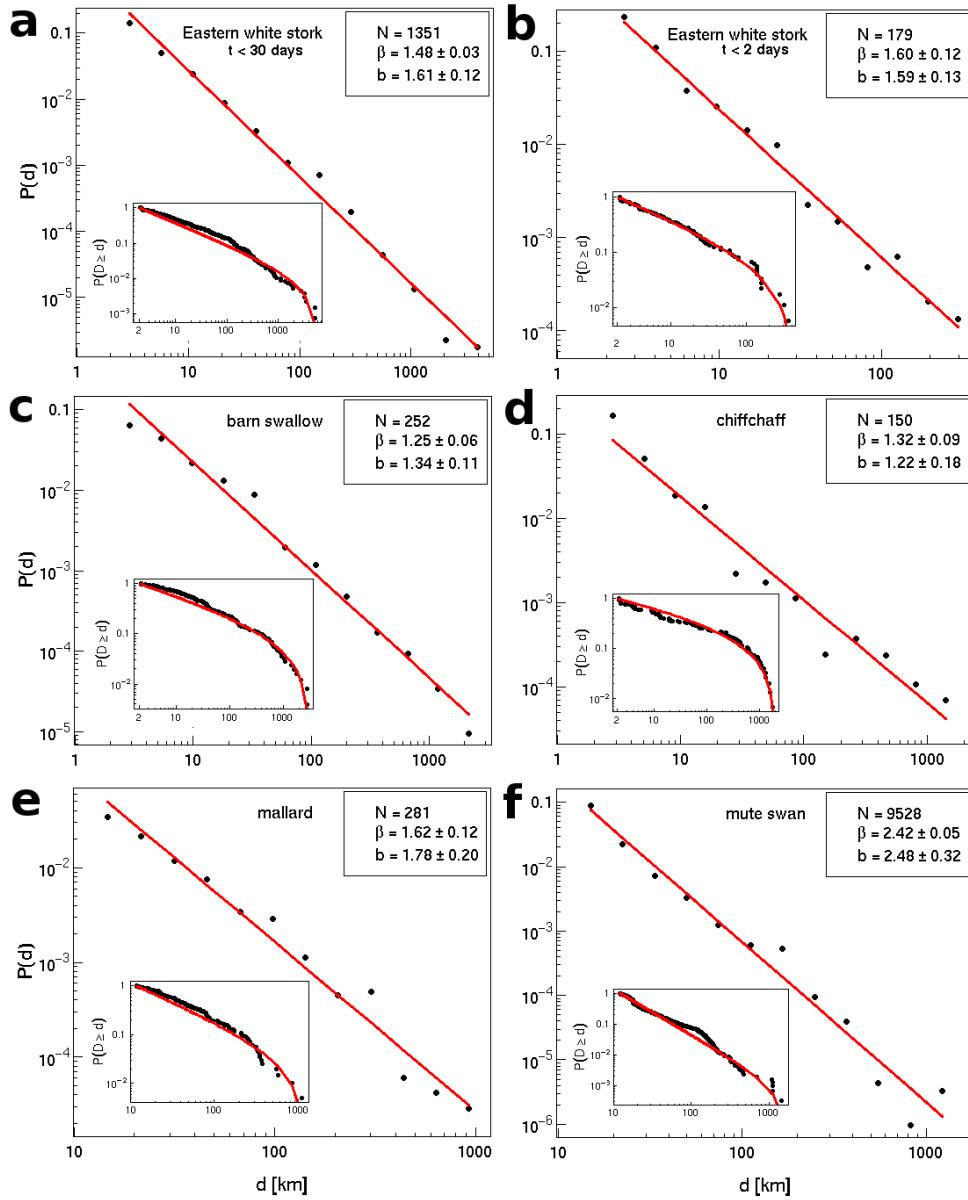


Figure 5.1: Short - term dispersal kernels of bird migration. Measured distribution $P(d)$ of traversing a distance d in less than τ days using logarithmic binning for (a), white stork ($\tau = 30$); (b), white stork ($\tau = 2$); (c), barn swallow; (d), chiffchaff; (e), mallard; and (f), mute swan (all $\tau = 30$). Further indicated is the sample size N , a maximum likelihood estimate of the power law exponent β and the exponent b from linear regression (shown as a solid line) with 95% confidence limits. Coefficients of determination for all regressions are $r^2 > 0.96$. Insets compare the cumulative distribution $P(D > d)$, i.e. the probability for a displacement D larger than d , for the bird data (black points) with the analytic curve $P(D > d) = (d^{1-\beta} - d_{max}^{1-\beta}) / (d_{min}^{1-\beta} - d_{max}^{1-\beta})$ for a truncated power law with exponent β and upper cut-off d_{max} (red lines). d_{max} was estimated as 110% of the maximal displacement for each species. .

intermediate position ($\beta = 1.48$). The small exponents for migratory birds, $\beta < 2$, indicate strongly directional motion. For sedentary birds scaling is observed only for $d \geq 12\text{km}$, reflecting the different mechanism of their long distance displacements, dismigration (Berthold, 2001a).

The cumulative distributions, corresponding to rank frequency plots (Newman, 2003b), do not show scaling, but agree remarkably well with a power law with upper cut-off d_{max} (insets in Fig. 5.1). For bird motion such a maximal displacement distance naturally arises from physiological as well as geographical constraints.

Following Edwards *et al.* (2007), using Akaike weights w_i we tested for unbounded and truncated power law distributions, considering three alternative distributions (exponential, truncated exponential and log-normal distribution). For the two sedentary bird species we find $w_1 \approx 1$ for the power-law distribution and $w_i < 0.01$ for the other distributions ($i = 2, 3, 4, 5$). For the three migratory species truncated power laws gave best results $w_2 \approx 1$ ($w_i < 0.01$ for $i = 1, 3, 4, 5$). Thus, our data are better described by an unbounded or truncated power law, whereas exponential and log-normal distributions can be ruled out.

Lévy flight movement patterns can be explained by several possible means, such as distributions of landmasses, climate conditions or vegetation types, as well as many ecological and behavioural mechanisms that can shape movement patterns (Viswanathan *et al.*, 1999, 2008; Benhamou, 2007). Possible artefacts from sampling bias in data acquisition, ring reporting rates differ according to region (Fiedler *et al.*, 2004), can be ruled out because very similar scaling is observed irrespective whether displacements are examined in a time window of $\tau = 2$ days or $\tau = 30$ days (see e.g. Fig. 5.1 a, b). This insensitivity to time window length (i.e. scale invariance) further supports notions of an underlying Lévy process.

Our investigation points to the fact that Lévy flights might in fact be much more frequent in nature than ever thought before. This is notable, since Lévy flight patterns reversely have strong influence on many ecological processes, not least concerning the spread of infectious diseases by migratory birds.

Methods

Bird positions were obtained from ring-recapture data available from the three German ringing centres. We calculated the geographical distances d along loxodromes (Wiltschko & Wiltschko, 1995) connecting all pairs of capture and recapture locations of each individual (similar to data format in Brockmann *et al.* (2006)). The analysis is restricted to 11,562 displacements of short time intervals, $t < \tau$, and a minimum distance, $d \geq 2$ km. Furthermore, we excluded dead-recaptures, data from juveniles and birds raised by humans, held in captivity or relocated by man otherwise.

We used maximum likelihood estimates and calculated the Akaike weights (Edwards *et al.*, 2007) for five model distributions: unbounded and truncated power law, unbounded and truncated exponential and the log-normal distribution, with likelihoods:

$$\begin{aligned}
L_{PL}(\beta|d) &= \prod_{j=1}^n \frac{\beta - 1}{d_{min}^{1-\beta}} d_j^{-\beta}, \\
L_{trPL}(\beta|d) &= \prod_{j=1}^n \frac{\beta - 1}{d_{min}^{1-\beta} - d_{max}^{1-\beta}} d_j^{-\beta}, \\
L_{Exp}(\lambda|d) &= \prod_{j=1}^n \frac{1}{e^{-\lambda d_{min}}} \lambda e^{-\lambda d_j}, \\
L_{trExp}(\lambda|d) &= \prod_{j=1}^n \frac{1}{e^{-\lambda d_{min}} - e^{-\lambda d_{max}}} \lambda e^{-\lambda d_j} \quad \text{and} \\
L_{LogN}(\mu, \sigma|d) &= \prod_{j=1}^n \frac{1}{(d_j - d_{min})\sigma\sqrt{2\pi}} e^{\frac{-(\log(d_j - d_{min}) - \mu)^2}{2\sigma^2}}, \text{ respectively.}
\end{aligned}$$

6 Paper V.

A periodic Markov model of bird migration on a network

Theory and data analysis for the white stork and greater white-fronted goose

Andrea Kölzsch, Helmut Kruckenberg, Michael Kaatz and Bernd Blasius;
to be submitted

Abstract

Bird migration has fascinated man for a long time and its wide interconnectivity with human life emphasises the importance to advance our knowledge of general issues of migration patterns. In this study, we describe bird migration as a Markov process with seasonally periodic, time-dependent transition rates on a directed stepping stone network. Each breeding, resting and wintering site is occupied by some fraction of birds during a certain time of the year and these nodes are connected by edges that represent the seasonally changing transition intensities between the habitats. The transition rates are considered to be periodic, unimodal functions of the time of the year. For the construction of the model and parameter estimation we used satellite telemetry and GPS data of the white stork (*Ciconia ciconia*) and the greater white-fronted goose (*Anser albifrons*). Finally, with methods of network analysis we examined the dynamical properties of the generated cumulative networks and time specific sub-networks of the two species. The results allow for a characterisation of the complex migration routes, also giving indications about the species' vulnerability to the loss of certain resting or breeding areas and changes in migration timing. The model is suitable for devising conservation strategies and studying the spread of pathogens by migratory birds.

6.1 Introduction

Movement is a basic property of life and has fascinated ecologists for many decades (Turchin, 1998; Begon *et al.*, 2006). Almost all organisms locomote around the earth at different timely and spatial scales. They undertake dispersal to exploit new habitats and resources, and some species migrate more or less regularly to avoid unfavourable environmental conditions. Most obvious are mass movements of e.g. buffalo herds, swarms of locusts, upstream motion of fish for spawning, but also the single flight of bees, birds and bats (Begon *et al.*, 2006). Up to now, the movement of insects appears to be best studied. Paths of butterflies, ants and beetles

have been observed, and in a seminal study by Kareiva & Shigesada (1983) the flight of two butterfly species was described by a mechanistic model.

So far, in many population dynamics models movement has not been explicitly included because of a lack of sufficiently accurate and well resolved data (Turchin, 1998). With the recent wide availability of tagging technologies, like GPS and satellite telemetry, this has changed. Now a flood of data has become available on the moving about of individuals of many taxa, e.g. fish, birds, terrestrial and aquatic mammals (Berthold *et al.*, 1992; Marell *et al.*, 2002; Kruckenberg *et al.*, 2007; Sims *et al.*, 2008), and more is expected as the technologies are improving. This gives rise to two challenges: (i) powerful analysing techniques will be needed for these data, and (ii) by theoretically describing the data we want to develop quantitative models that are able to capture the various types of movement.

A prominent example of animal movement is the seasonal migration of birds. Numerous bird species migrate long distances to avoid barren winter conditions in their breeding areas and winter in regions that provide food. Hence, bird migration is a periodic movement process that is driven by seasonally changing temperatures and weather conditions (Bairlein, 2003). Many studies have examined a wide range of aspects of bird migration (Berthold, 2001a; Berthold & Terrill, 1991; Bairlein, 2003). Up to now, most work has been devoted to understanding the physiological and ecological mechanisms of the migration process, metabolism, navigation and flight (Hedenström & Ålerstam, 1997; Wiltschko & Wiltschko, 2003). Incidental positions of marked migratory birds have been catalogued in migration atlases (Berthold, 2001a). However, explicit spatio-temporal trajectories from e.g. satellite telemetry and GPS are known for a few species only. And even for those it is unclear how variable the migration process is and how changes in environmental conditions may affect it.

Bird migration is widely intertwined with human life and activities. Recently, this has been stressed in the light of disease transmission by migratory birds (Peterson *et al.*, 2003; Hubálek, 2004; Kilpatrick *et al.*, 2006) and the influence of human induced climate change on migration routes and timing (Jenni & Kéry, 2003). Thus, it is of special importance to advance our knowledge about the structure of migration patterns and mechanisms behind their forming. Theoretical predictions about the spread of avian infectious diseases by migratory birds require a quantitative description of the seasonal movement of the birds. However such a description is evasive.

For several bird species migration has been described as a stepping stone process by eco-physiological field studies (Berthold & Terrill, 1991; Hedenström & Ålerstam, 1997; Bos *et al.*, 2005). Species for which seasonal hyperphagia and fattening for migration (Berthold, 2001a) is not practicable need to take intermediate rests for feeding during their highly energy consuming migration flights. Such resting regions are especially important for successful migration, but also facilitate disease transmission across different populations and species. From this structure, it comes natural to model such migration as a stepping stone process on a network of regions where birds concentrate at some time of the year, i.e. breeding, resting and wintering areas. It has also been shown previously (Kölzsch & Blasius, 2008) that the motion of such a stepping stone migratory bird, the white stork (*Ciconia ciconia*), is composed of two modes of movement: slow resting and fast migration. Here we present a novel, quantitative network model of bird migration and validate it with high quality temporal and spatial data of two bird species.

Theoretical and applied modelling has been performed for many ecological issues and even migration physiology, flight, orientation and the evolution of bird migration (Alerstam & Hedenström, 1998; Houston, 1998; Mouritsen, 1998; Erni *et al.*, 2002). However, explicit spatio-temporal models of the process of bird migration have, to the best of our knowledge, not been developed yet (Bairlein, 2003).

Our model describes bird migration on a complex network. A large number of processes in nature and human life, transportation as well as others have been modelled as networks and their examination revealed many interesting characteristics and patterns (Newman, 2003b; Barrat *et al.*, 2004; Urban *et al.*, 2009; Kaluza *et al.*, 2009). Global transportation, e.g. within the aviation network, makes human movement non-diffusive by providing efficient mixing through long range displacements. This is crucial for understanding the rapid spread of human diseases on earth (Hufnagel *et al.*, 2004). In studies of landscape ecology and metapopulation theory the concept of networks has also been successfully applied, but rarely (Urban *et al.*, 2009).

6.2 Model

Bird migration is modelled as a Markovian jumping-process on a geographically embedded network. The model describes the time dependent densities of birds in region $i = 1, \dots, n$ at time t . Links represent main transition events – the flights between the different areas (see also Kölzsch & Blasius, 2008). The changing density of birds in each node i is then given by the difference of outgoing and ingoing flow $J_{ji}(\tau)$ and $J_{ij}(\tau)$ in season τ . One can formulate this space discretised process as a time continuous Markov process using the master equations

$$\dot{N}_i(t) = \sum_{j \neq i} J_{ij}(\tau) - \sum_{j \neq i} J_{ji}(\tau). \quad (6.1)$$

$J_{ij}(\tau) = r_{ij}(\tau)N_j(t)$ is the flow from node j to i at season τ and $r_{ij}(\tau)$ the corresponding transition rate with $\tau = t \bmod T$. The period $T = 2\pi$ corresponds to one year. For a schematic representation of an example of the process see Fig. 6.1.

The time-dependent, periodic transition rates $r_{ij}(\tau) = r_{ij}(\tau + T)$ between any regions j and i are described by a three parameter von Mises functional form (see Fig. 6.1 b)

$$r_{ij}(\tau) = \omega_{ij} \exp\left(\frac{\cos(\tau - \varphi_{ij}) - 1}{\sigma_{ij}^2}\right) \quad (6.2)$$

Here, φ_{ij} is the main transition season, σ_{ij} characterises the width of the transition period and ω_{ij} is the maximum transition intensity that is typically realised during $\tau = \varphi_{ij}$. This functional form of $r_{ij}(\tau)$ allows to interpolate smoothly between the δ -function ($\sigma_{ij} \rightarrow 0$), when a fraction of ω_{ij} birds jumps exactly at season τ , and the uniform distribution ($\sigma_{ij} \rightarrow \infty$) where transitions happen independently of the season.

By introducing this model we have made several simplifying assumptions. First, the Markovian property that a transition probability depends only on the previous, but not earlier densities implies that birds use no memory of their former whereabouts for the decision where to go

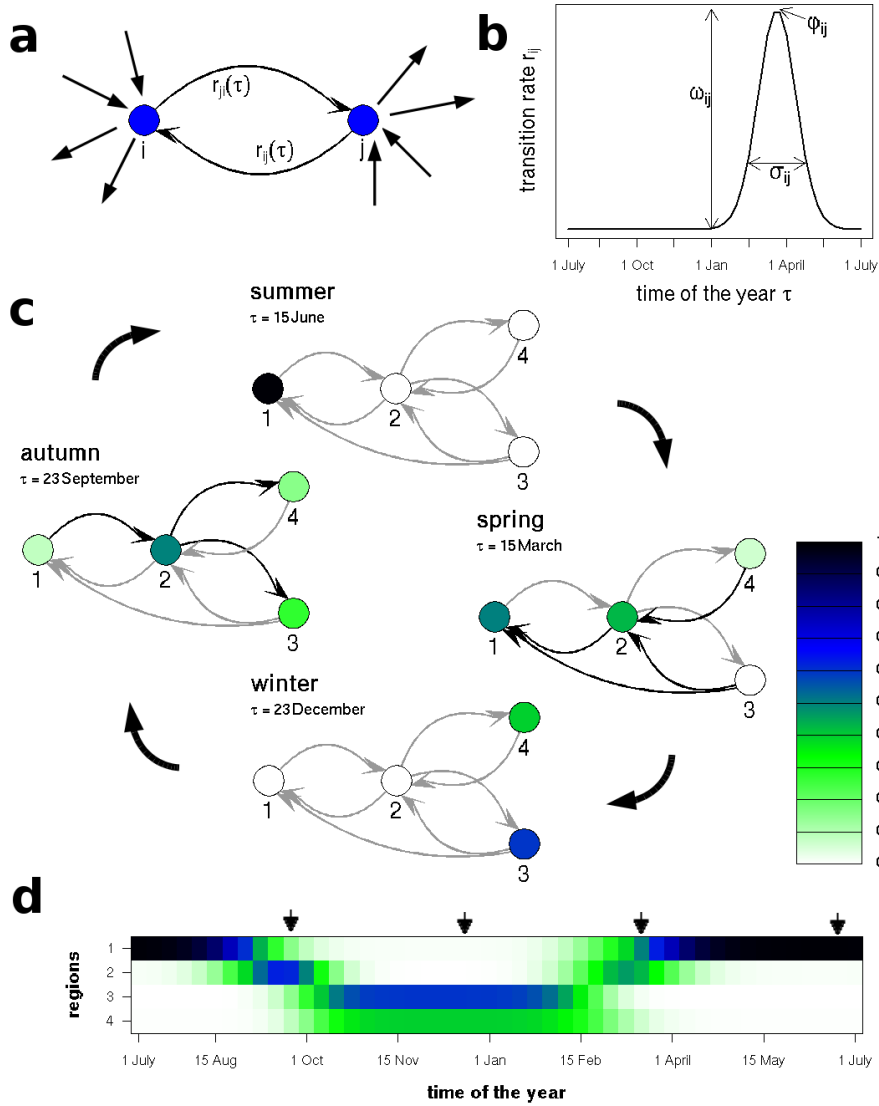


Figure 6.1: Concept of the Migration network model. (a) One isolated pair of nodes i, j that is connected by directed links weighted by seasonally changing transition rates $r_{ij}(\tau)$ and $r_{ji}(\tau)$. (b) Illustration of a transition rate as function of the season. It is of a von Mises functional form (eq. 6.2), the arrows indicate how the three function parameters determine its form. (c) and (d) present a model example on a simple topology of four nodes and seven transition links. Parameters are: (i) phases (ii) amplitudes and (iii) widths. In (c) for each season one typical density pattern of a time interval of one week is given. One can clearly observe breeding in node 1 in summer, spring and autumn migration via node 2 and wintering in nodes 3 and 4. The densities of birds in each node are colour coded (see legend). Links are drawn in black whenever transition probabilities were larger than 0.05 in the respective time interval, grey else. (d) depicts year-round network flow, densities are colour coded as above, each row representing density changes in the respective node over time of the year. The small arrows indicate for which time intervals network plots are shown in (c).

next or if to stay. This depends on the season and their present state only. Second, the model obeys the conservation law, thus the population size is considered constant. Third, the transition rates are independent of the present densities in the nodes. Therefore, we disregard any swarming or density dependence effects. Finally, it has not been explicitly accounted for the fact that the birds' decision for migratory departure from breeding, wintering and stopover sites depends on environmental variables like the photoperiod and food availability and uptake (Berthold & Terrill, 1991; Gwinner *et al.*, 1988). However, this issue is implicitly included in the time dependence of the model's transition probabilities.

6.3 Data

We have parameterised the model, using high precision movement data (satellite telemetry and GPS localisations) for the white stork (*Ciconia ciconia*) and the greater white-fronted goose (*Anser albifrons*). These two species perform stepping stone migration, and are thus suitable for being described by our network model. Both species are relatively large long-distance migrants with low maximum fuel deposition rates and need a number of stopover sites on their migration route for refuelling (Beekman *et al.*, 2002). However, they differ in their migratory paths and flight behaviour.

White stork: The white stork is one of the best studied bird species that breed in Europe. Its life history is well understood and migration routes have been worked out (Kaatz, 2004; Berthold *et al.*, 2006). There is an eastern population that migrates east around the Mediterranean, passes Turkey and Israel to winter or rest in different areas in the Sudan and Chad. Some individuals continue to southern Africa for wintering. A western population used to migrate west around the Mediterranean and wintered in certain places in western Africa. However, many birds nowadays stay in Spain during the winter. Depending on availability white storks feed on large insects, amphibians and small reptiles and mammals. To fly long distances they use thermalling. Therefore, for migration storks rely on favourable weather conditions, their routes are restricted to over-land and they fly at daytime only. Since 1992 more than 100 individual storks have been equipped with a satellite telemetry transmitter by the MPI for Ornithology – Vogelwarte Radolfzell (Berthold *et al.*, 1992). The transmitters regularly send signals which are received and localised by satellites using the principles of the Doppler effect. The location data has been analysed in various ways by previous studies (e.g. Berthold *et al.*, 2002; van den Bossche *et al.*, 2002; Kaatz, 2004). We obtained a data set of 106,220 localisations of 107 individual white storks.

Greater white-fronted goose: The greater white-fronted goose winters in Europe from the United Kingdom to Kazakhstan and migrates to the arctic tundra for breeding and moulting (Kruckenberg *et al.*, 2008). Its wintering behaviour has been extensively studied (Mooij *et al.*, 1999; Kruckenberg, 2002), but there is much unknown about its ecology, breeding biology and migration behaviour. Greater white-fronted geese use flapwing flight and often move up to 2000 km non-stop within one or two days between migration stepping stone regions. There they have to refill their fat deposits by grazing on young grassy vegetation. In contrast to other goose species white-fronted geese do not fly along the coastal lines of the Baltic and Barents-Sea, rather they migrate more inland through the European part of Russia. Their

breeding range expands from Kanin Peninsula in the west to Taimyr Peninsula in the east. Nonbreeders gather in large flocks of more than 100,000 birds on traditional moulting sites at Taimyr or Novaya Zemlya (Mooij *et al.*, 1999). Since 2006 a number of adult birds (males) have been fitted with combined GPS and Doppler location satellite telemetry transmitters at their wintering sites in the Netherlands by the European greater white-fronted goose scientific programme (Kruckenberg *et al.*, 2007). We obtained flight trajectories of 32 individuals that consisted of 17,803 GPS localisations and 19,181 Doppler location positions.

Quality selection: Before one can analyse the described data sets their quality deficiencies, i.e. localisation inaccuracies (Kaatz, 1999) have to be accounted for. For the white stork data we have used an algorithm previously developed for this data set (Kaatz, 1999) to estimate how much a localisation deviates from the real position of the bird. We selected stork localisations of mean deviations less than 20 km for our analysis. The GPS data of the greater white-fronted goose are correct up to 15 m, thus we can use it as is. However, there are often gaps in the data sets due to poor light conditions that do not allow for the solar batteries to recharge sufficiently for GPS functionality. For times like this only satellite telemetry data based on Doppler location are available. Because different kinds of transmitters have been used, the quality algorithm by Kaatz (1999) cannot be applied here. Therefore, we have used a filtering algorithm developed by Austin & McMillan (2003) to select only localisations that conform to realistic flight velocities and travel directionality of the moving birds.

For further use, only trajectories of adult birds that encompass at least one full year of movement data were chosen. That is necessary, because only adult birds migrate regularly and our model requires data that are evenly distributed over a full calendar year. For the white stork 32 such individual year trajectories of 12 different individual birds were extracted, whereas the geese data encompassed 8 individual year trajectories of 7 adult geese.

We selected the best possible resting localisation per day for each of the individual year trajectories. For the storks localisations with least spatial deviation and the ones that were taken at times of the day with lowest sunset were preferred. This is sensible, because the need for thermalling during migration restricts their flight to daytime and low sunset data provide positions of resting on the ground. For geese we selected the GPS localisation, if available, that was taken latest during the day or the best quality satellite telemetry localisation. The positions later in the day were preferred, because the solar batteries have probably more power then. A discrimination of geese localisations by sunset is not possible; they fly during the day as well as at night. Finally, sets of continuous movement trajectories $x_i(t)$ of one localisation per day were available.

6.4 Methods

6.4.1 Construction of a migration network

The breeding, wintering and resting areas of the respective bird species, i.e. the network nodes, were determined based on the velocity with which the birds were moving. For each position we calculated the approaching and departing velocities $v_{i,1} = (x(t) - x(t - \Delta t))/\Delta t$ and $v_{i,2} = (x(t + \Delta t) - x(t))/\Delta t$ if previous and subsequent positions were taken less than one week before or after the present position. Locations were selected as resting localisations (see

Fig. 6.2) when the approaching and departing velocities were less than 50 km/day (Gerkmann & Riede, 2005) for the white stork and 20 km/day for the greater white-fronted geese. The so specified localisations were then clustered hierarchically by a combination of their pairwise Loxodrome distances (Wiltschko & Wiltschko, 1995) and circular time differences (Batschelet, 1981). Loxodromes were selected, because birds have been proposed to fly along these routes of constant bearing rather than along the orthodromes of shortest distance (Berthold, 2001a). The cut off values for clusters were selected according to maximum daily average flight distances and dendrogram patterns, 600 km/day for the white stork and 350 km/day for the white-fronted goose. All clusters that contained more than 3 resting localisations were finally selected to represent the set of typical breeding, resting and wintering regions of the considered population of each species.

6.4.2 Calculation of the links and their transition rates

The proposed migration model assumes that the continuous time transition rates $r_{ij}(\tau)$ are of a von Mises functional form (eq. 6.2). To parameterise this function one first has to approximate a discrete set of transition fluxes $J_{ij}(\tau, \tau + \Delta\tau)$ and bird densities $N_j(\tau)$ from the data for small, discrete time steps Δt . We selected Δt to be around 1 week ($\Delta t = 0.13$ if 1 year = 2π), thus there are 48 different seasonal time intervals $(\tau, \tau + \Delta t)$. $J_{ij}(\tau, \tau + \Delta t)$ between all pairs of nodes i, j were counted for each time interval and converted into transition rates

$$r_{ij}(\tau) = \frac{J_{ij}(\tau, \tau + \Delta t)}{N_j(\tau)\Delta t}. \quad (6.3)$$

Bird densities as extracted from the data are included here, but are not general solutions of our model. The determination of smooth transition rates allows for the assignment of generalised population densities, and not just assertions about the small set of individual birds the data of which is available.

For each realised transition link the von Mises parameters φ_{ij} , σ_{ij} and ω_{ij} were estimated. As this estimation is often based on very few data it is sufficient to calculate parameters with circular statistics (Batschelet, 1981) as follows:

$$\varphi_{ij} = \arctan \left(\frac{\sum_{\tau} r_{ij}(\tau) \sin(\tau)}{\sum_{\tau} r_{ij}(\tau) \cos(\tau)} \right) \quad (6.4)$$

$$\sigma_{ij} = \sqrt{\frac{1}{\sum_{\tau} r_{ij}(\tau)} \sum_{\tau} r_{ij}(\tau) (|\tau - \varphi_{ij}| \bmod (2\pi))} \quad (6.5)$$

$$\omega_{ij} = \max_{\tau} (r_{ij}(\tau)). \quad (6.6)$$

Here, φ_{ij} is the circular mean of the transition times from j to i with respect to the transition rates, σ_{ij} is the circular standard deviation of the transition times, and ω_{ij} the maximum transition intensity (see also Fig. 6.1). Using these parameters one can calculate a smooth, estimated transition rate $\hat{r}_{ij}(\tau)$ for any season τ using eq. (6.2).

6.4.3 Estimation of the transition probabilities for the non-homogeneous Markov chain

In this work, we use the discrete Markov chain that corresponds to the continuous model. Reasons were the relatively sparse data that are obtainable for parameterisation and the readily available methods for the analysis of Markov chains (Norris, 1997). The density of birds $N_i(t)$ thus develops as

$$N_i(t + \Delta t) = \sum_{j \neq i} p_{ij}(\tau; \Delta t) N_j(t) + p_{ii}(\tau; \Delta t) N_i(t), \quad (6.7)$$

$p_{ij}(\tau; \Delta t)$ being the transition probability at season τ for the time interval length Δt . The two descriptions of the transition process can be transformed into one another by

$$P(\tau; \Delta t) = e^{R(\tau) \Delta t}, \quad (6.8)$$

using the Chapman-Kolmogorov differential equations (Logofet & Lesnaya, 2000), $P(\tau; \Delta t)$ is the matrix $(p_{ij}(\tau; \Delta t))_{ij}$ and $R(\tau)$ corresponds to $(r_{ij}(\tau))_{ij}$. Because of the monotony of the exponential the functional forms of the transition probabilities $p_{ij}(\tau; \Delta t)$ also resemble the von Mises function.

To solve the model, probability matrices $P(\tau; \Delta t)$ were calculated from the transition rate matrices $R(\tau)$ (eq. 6.8) for $\Delta t = 0.13$ (one week). Then, the stable distributions $N_i^*(\tau)$ can be calculated by using the Markov chain equations (eq. 6.7) and calculating the Poincaré map from one year to the next through each time step τ by matrix multiplication

$$S(\tau) = \prod_{s=0}^{T-\Delta t} P(\tau + s; \Delta t). \quad (6.9)$$

Each matrix $S(\tau)$ is a regular stochastic matrix and has a dominant eigenvalue $\lambda(\tau) = 1$ (Perron-Frobenius Theorem). The corresponding eigenvector is the stable distribution $N_i^*(\tau)$, i.e. $S(\tau) = S(\tau) N^*(\tau)$.

6.4.4 Characterisation of the migration network and Markov chain

Season specific migration network topologies were derived such that a pair of nodes is connected by a directed link from j to i whenever $p_{ij}(\tau; \Delta t) > 0.05$. For all species, additionally, a cumulative network was constructed by merging the time specific networks. First, we want to examine which important network characteristics the cumulative migration networks have in common with other transportation networks (e.g. the internet, metabolic networks and the worldwide airport network). We calculated global and local efficiencies and network costs (Latora & Marchiori, 2001), as well as average shortest topological path distances and clustering coefficients (Watts & Strogatz, 1998) to characterise if the two migration networks are similar to small worlds (Watts, 1999). This is a sign of networks optimised for quick and efficient transport of entities between the nodes. Furthermore, we analysed network robustness to random or selective deletion of strongly connected nodes (Albert *et al.*, 2000). High robustness

can be another indication for small world behaviour, but also suggests how strong flow on networks depends on each single node.

For further specification of migration network properties in the light of optimisation for transport we examined motif distributions (Milo *et al.*, 2002), i.e. the “network building blocks”, and the networks’ symmetry. The latter shows which proportion of the migration routes is used in both directions, i.e. during spring as well as autumn migration.

For detailed characterisation of each resting area’s (node’s) importance for migration flow we determined degrees k , betweenness centralities bc (Freeman, 1979), closeness centrality cc (Wassermann & Faust, 1994), clustering coefficients C , cumulative densities, mean staying times following the highest density $\langle t_s \rangle$ and mean first passage times $\langle t_{fp} \rangle$ (Norris, 1997). Betweenness centrality is the number of topologically shortest paths in the network that pass a node, closeness centrality is a measure of the shortest distance of one to all other nodes. These measures characterise how much each node contributes to network connectivity. The cumulative densities, staying times and mean first passage times indicate which nodes are frequently reached and used often and by a large number of birds. First passage times are studied in more detail, because they seem especially interesting in the light of bird migration. How long does it take to move from one resting area to the other and when is it likely to return to a node? First passage times vary with the time of the year one starts off a region. Because we are interested in the behaviour of the majority of individuals we have selected the first passage time starting in the season of highest $N_i^*(\tau)$ for each transition from i to j and return time to i .

The topologies of the bird migration networks change with season. Therefore, to just examine the cumulative network characteristics is not sufficient. Additionally, we present simple network measures for each of the season-specific networks. The changes of the mean degree as well as the size of the giant component (i.e. largest connected subnetwork) with season are indications of network changes.

6.4.5 Model and parameterisation verification and sensitivity analysis

For the verification of how good node positions coincide with regions that are valid for breeding, resting and wintering of the white stork (Gerkmann & Riede, 2005) or greater white-fronted goose, respectively, we determined a dominant land cover type in the area of a radius of 200 km around each node centre. They were compared with the habitat preferences of the specific bird species. Land cover data were obtained from the USGS Land Cover Project (<http://edc2.usgs.gov/g1cc/>).

To validate the appropriateness of our model for mapping bird migration, we explored the quality of rate fitting and how well the solutions $N_i^*(\tau)$ represent counted nonzero numbers of birds in the nodes. Differences and similarities of $r_{ij}(\tau)$ and $\hat{r}_{ij}(\tau)$ were analysed by a two-sample paired Wilcoxon test. To assess the fit of bird numbers by the stationary densities $N_i^*(\tau)$, G-Test (Sokal & Rohlf, 1995) results were evaluated. Furthermore, we examined the sensitivity of the model outcome $N_i^*(\tau)$ to changes of the phases φ_{ij} , widths σ_{ij} and amplitudes ω_{ij} of single transition rates, again by G-Tests.

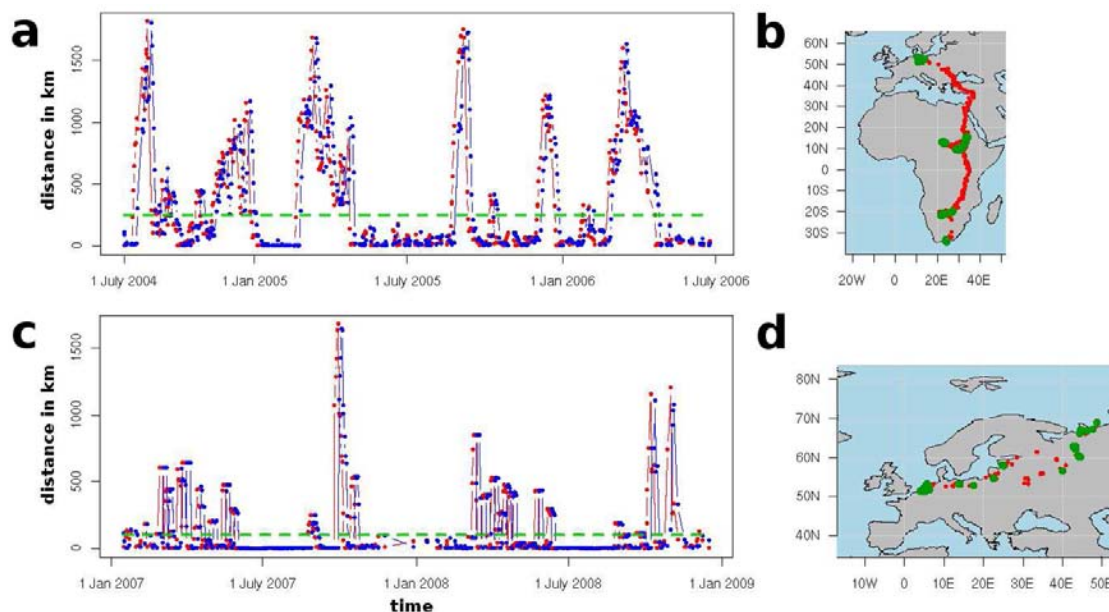


Figure 6.2: Determining resting localisations by flight velocity. (a) for a white stork individual and (c) a greater white-fronted goose approaching (red dots) and descending velocities (blue dots) along a two-year trajectory. Empirically given thresholds for velocities of resting and migration movement are shown as green dotted lines. For the stork (b) and the goose individual (d) selected breeding, resting and wintering localisations are mapped as green dots; red dots depict localisations of fast migration movement. Resting areas for the white stork are located in northern Germany, the Sudan area and southern Africa. The goose individual winters in the Netherland area, migrating via a number of stopovers in wetland areas in northern Poland, the Baltic coast and Russia up to the island of Kolgueiv.

6.5 Results

6.5.1 Constructed migration network and Markov chain

For two years of migration of a white stork and a greater white-fronted goose individual approach and departure velocities are presented in Fig. 6.2. It is clearly observable that they consist of two modes, slow resting movement and faster migration movement. Presented positions of slow movement give an indication of where breeding, resting and wintering regions are located. For the full set of calculated breeding, wintering and resting regions, in the form of nodes of a directed network, see Fig. 6.3 a, c. The network of white stork migration consists of $n = 24$ nodes and $l = 96$ directed links, whereas the greater white-fronted goose network has $n = 17$ nodes and $l = 53$ links. When examining the localisations that have been clustered into each node, it becomes clear that a few are rather scattered around the midpoints of the nodes, whereas the others nicely cumulate within.

The cyclic, seasonal Markov model flow of birds between the nodes of each network is presented in Fig. 6.3 b, d. It becomes clear that there is not just a linear flow of birds along a number of nodes, but that there is branching, especially in the stork network. Furthermore,

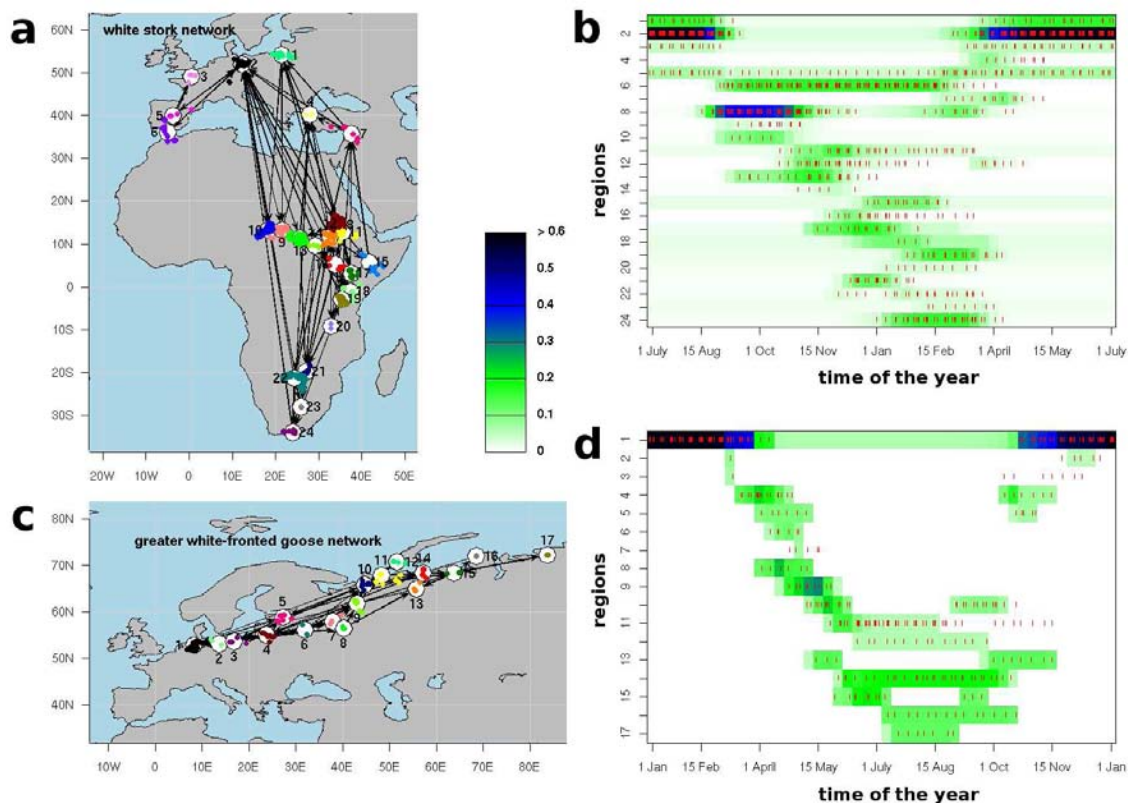


Figure 6.3: Migration networks and flows. For (a) the white stork and (c) the greater white-fronted goose we provide cumulative networks with resting localisation clusters, i.e. network nodes. Locations grouped into the same node are presented in the same colour. Bird densities of the non-homogeneous Markov chains on the network of the white stork (b) and the greater white-fronted goose (d) are indicated by colour for each node for a number of $n = 48$ time intervals during the year. The small red lines indicate times when birds of the respective data set were located in the specific region.

the nodes seem not to be equally involved and frequently used as bird species specific seasonal habitats. Some are frequented regularly by many birds or for a long time, whereas others seem to be simply short-time resting stations. The added small lines in Fig. 6.3 c, d that indicate when a bird in the data set was located in the respective node region reveal that the models by and large represent the data sets well. To get an even better impression of the models' performance, in Fig. 6.4 six season-specific sub-networks of the white stork are presented. According to the phase of migration, the link intensities change, thus nicely mapping the real migration process.

6.5.2 Characterisation of the migration network and Markov chain

There are several indications that the determined topological migration networks are efficient transportation networks. Local and global efficiencies as well as network costs for the white

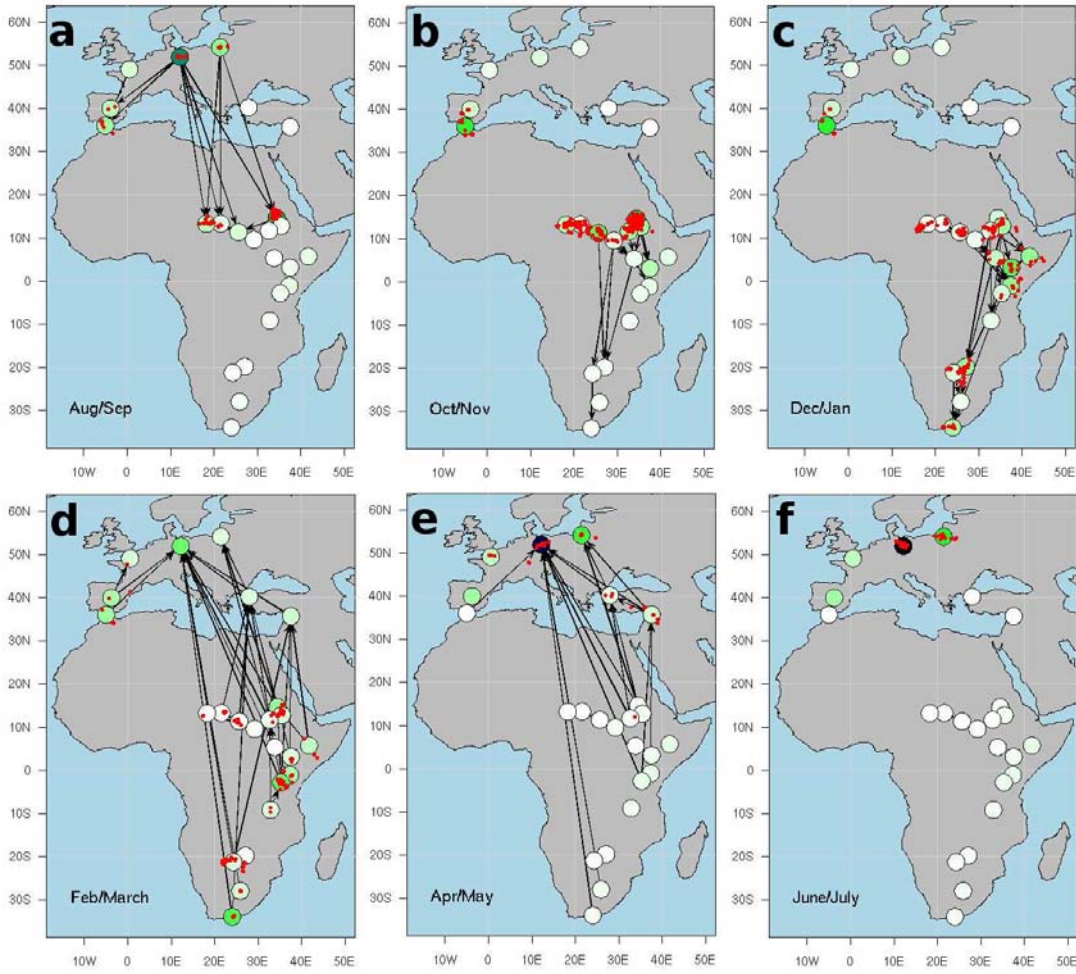


Figure 6.4: Season-specific, partially cumulative networks of the white stork. Colouring of the nodes indicates mean densities of birds there during the respective two months time window. Bird densities are colour coded corresponding to the label in Fig. 6.3. A link is drawn here, whenever transition probabilities were larger than 0.05 for any one week period within the considered time interval. The change in the connectivity is an indication of the migration phase. Red dots show real data positions of white storks within this period of time.

stork are $E_{glob,s} = 0.49$, $E_{loc,s} = 0.61$, $c_s = 0.17$, and for the greater white-fronted goose $E_{glob,g} = 0.51$, $E_{loc,g} = 0.64$, $c_g = 0.19$. They reveal that these networks warrant fast transportation at relatively low cost (few links) and high robustness to failures. Also mean clustering coefficients $\langle C \rangle_s = 0.31$, $\langle C \rangle_g = 0.35$ and average shortest path lengths $\langle d \rangle_s = 2.47$, $\langle d \rangle_g = 2.46$ support this notion. The small $\langle d \rangle$ coincide with the fact that the individual white storks in our data use only 2 – 9 nodes and the greater white-fronted geese 6 - 9 different nodes per year. Furthermore, if one examines robustness of the networks to node deletion, it becomes apparent that the connectedness of the giant component is less affected by random node removal than selective removal for highly connected nodes (figures not presented here).

These are properties that the migration networks share with several small world networks.

The transportation optimised character of our migration networks is additionally supported by the motif distributions of their 3-node-subgraphs. Subgraphs that represent transit movement (sub-networks 3, 4) and patterns of branching character (subgraphs 7, 9 and 10) are overrepresented in comparison to random graphs of similar structure. It can be observed that the motif distribution (Fig. 6.5 b) falls into the second superfamily (Milo *et al.*, 2004) which includes the signal-transduction and neurons networks. Another property of migration networks that shows that they are special kinds of transportation networks is their high degree of asymmetry. Only $s_s = 16\%$ and $s_g = 20\%$ of the network links are bidirectional. This points out that most resting areas are not used both during the autumn and spring migrations, at least not in a similar ordering.

Several centrality measures (see Tab. 6.1) indicate which of the network nodes are especially important for migration flow. For the white stork network nodes 2 (Middle European breeding region) and 8 (Sudan wintering and staging area) are especially important, whereas nodes 3 (French breeding area), 20 (staging area close to Lake Malawi) and 23 (staging region at the Vaal and Orange Rivers) are not central at all. For the greater white-fronted goose network nodes 1, 4 and 15 are more important and nodes 2 and 12 less so. Measures of clustering, cumulative densities, staying times and mean first passage times generally emphasize these findings. Regions that are frequented for long times during breeding and wintering have large densities that often coincide with longer staying times (Tab. 6.1). Maximum density first passage times (Fig. 6.5 c, d) emphasize that nodes 2 and 8 for the white stork network and nodes 1 and 4 for the greater white-fronted goose are very important for migration and should lead to large scale evasion or collapse if destroyed.

Simple measures like the mean degree $\langle k \rangle$ and the size of the giant component for each time specific network illustrate nicely times of migration movement and times when birds are moving slowly, thus rest (see Fig. 6.5 a). For any more sophisticated analyses the single season-specific networks are too small.

6.5.3 Model and parameterisation verification

Dominant land cover types for the nodes of each network (Tab. 6.1) reveal a consistent picture with habitat preferences of the respective species. Major land cover types of the white stork network nodes are cropland in Europe and grassland, savannas and open shrublands in Africa. Network nodes' dominating land cover types for the greater white-fronted goose indicate that in the western wintering and stopover sites the geese stay in regions that are dominated by cropland. The further east they migrate and in their breeding areas land cover is mainly edges of forest, barren or sparsely vegetated shrublands. Resting regions are often situated by and in deltas of large rivers that flow into the sea.

From the data estimated nonzero $r_{ij}(\tau)$ and fitted rates $\hat{r}_{ij}(\tau)$ differ slightly but not significantly (two-sample paired Wilcoxon test: $W = 9040$, $p = 0.23$ (stork); $W = 1870$, $p = 0.29$ (goose)). Furthermore, estimated densities $N_i^*(\tau)$ from the model are very similar to the numbers of counted birds in each region during the different seasons. The numbers of birds in the regions differ on average by 0.65 (stork)/ 0.67 (goose) individuals (of $N = 28$ and $N = 8$, respectively). G – Tests of the goodness of fit for each distribution of the individuals

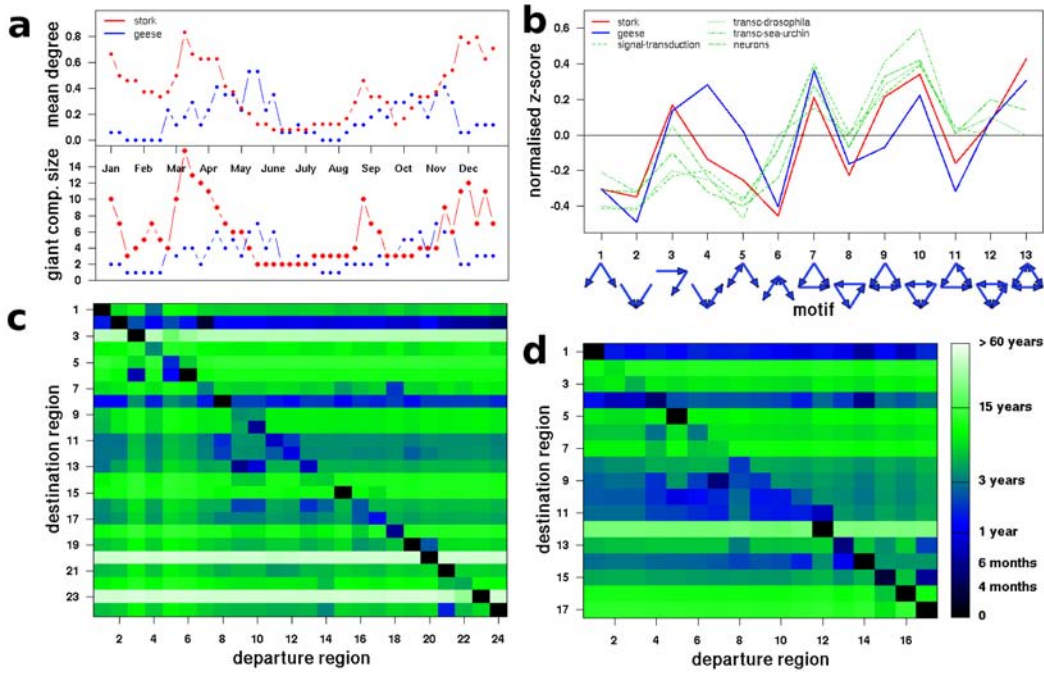


Figure 6.5: Characteristics of the white stork and greater white-fronted goose migration networks. (a) Presented are the mean degree and giant component size of the weekly season-specific networks. One can observe how they change with times of migration, breeding, resting and wintering for the stork (red) and geese (blue). (b) shows the motif distributions of the two cumulative migration networks. They coincide nicely with motif distributions of other real world networks of the second superfamily (Milo *et al.*, 2004) that are also depicted here. The first passage times, each starting from the season of highest density in the outgoing node (departure region), characterise the migration process of the stork (c) and geese (d). They indicate from which regions long transition times can be expected whereas others are reached and departed from quickly.

during the seasons confirm this likeness. Mean test results are $\langle G \rangle_s = 15.77$ ($p = 0.76$) and $\langle G \rangle_g = 11.01$ ($p = 0.96$), and 87.5% and 100% of the stork and goose models, respectively, show p-values of $p > 0.1$. Thus, we propose that our model is sensibly parameterised for the migration of the white stork and the greater white-fronted goose. This is remarkable especially because of the small sample sizes. Model outcome is not very sensitive to small changes in most parameters, e.g. transition phases φ_{ij} can be changed by up to one month without notable effect. However, one exception is the decrease of transition rate widths σ_{ij} that may often lead to $\sigma_{ij} \approx 0$ which strongly changes model results. Furthermore, the model is more sensitive to changes in transition rates from and to more important nodes (as by centrality), but not largely.

6.6 Discussion

From a principal and more or less simplified understanding of bird migration we propose for this dynamic process a Markov model with periodic, time-dependent transition rates on a network.

Such has also been called a cyclostationary (Gardner *et al.*, 2006), cyclic non-homogeneous (Patoucheas & Stamou, 1993; Vassiliou, 1998) or periodically inhomogeneous (Gee *et al.*, 2006) Markov process. We parameterised the model using real, high quality bird movement data. One may say that it convincingly mimics the migration process in time and space for two different bird species. The fact that the model is rather simple and discrete in time and space makes it especially suitable for examining a number of issues on bird migration dynamics by simulations.

Model validation shows that the model is not very sensitive to small shifts in single parameters. One qualitative difference between data and model dynamics is that transition rates that were fitted by the von Mises function are smooth in time. Considering the sparse number of transition events in the data this stands in contrast to the discrete transition events. Therefore, transition may start earlier in the model than in the data and peak transition times can be shifted. Such problems should disappear when using a larger and more randomised set of data, which may be available soon, as satellite telemetry and bird tagging becomes more advanced and widely applied.

Node <i>i</i>	Network node characteristics							Dominant land cover
	<i>k</i>	<i>bc</i>	<i>cc</i>	<i>C</i>	$\langle N_i \rangle$	$\langle t_s \rangle$	$\langle t_{fp} \rangle$	
a stork								
1	3.5	16.81	0.46	0.27	3.74	1.76	6.30	croplands
2	9.0	231.69	0.66	0.15	12.66	2.23	0.98	croplands/natural veget.
3	<i>1.0</i>	<i>0.00</i>	<i>0.30</i>	<i>1.00</i>	1.49	7.81	<i>34.20</i>	croplands
4	4.5	13.56	0.52	0.28	0.42	0.25	8.47	croplands/natural veget.
5	2.5	22.00	0.42	0.50	2.65	4.56	11.44	croplands
6	2.5	22.00	0.43	0.50	3.87	3.65	8.84	croplands
7	3.5	13.02	0.38	0.12	0.61	0.24	6.69	open shrublands
8	7.5	121.84	0.56	0.22	4.78	1.50	1.76	savannas
9	4.5	23.90	0.47	0.31	0.28	0.16	7.58	savannas
10	3.5	9.90	0.47	0.35	0.97	0.64	7.36	savannas
11	4.5	10.89	0.38	0.33	1.66	0.66	3.42	savannas
12	6.5	46.21	0.41	0.23	0.91	0.41	3.43	savannas
13	6.0	78.00	0.49	0.27	1.40	0.51	4.35	savannas
14	4.5	30.54	0.35	0.25	0.38	0.26	9.69	savannas
15	2.5	8.82	0.40	0.25	1.63	3.27	9.71	closed shrublands
16	5.0	23.10	0.42	0.27	0.51	0.36	4.51	woody savannas
17	5.0	27.89	0.43	0.22	2.00	1.18	4.09	open shrublands
18	2.5	10.18	0.38	0.25	1.80	3.46	9.75	grassland
19	4.0	34.73	0.35	0.25	1.72	1.67	5.74	savannas
20	<i>2.0</i>	<i>0.96</i>	<i>0.30</i>	<i>0.50</i>	0.63	6.23	<i>42.86</i>	savannas
21	4.5	38.62	0.44	0.21	0.80	0.79	5.55	savannas
22	3.0	12.57	0.34	0.20	0.56	1.46	8.24	savannas
23	<i>1.5</i>	<i>5.67</i>	<i>0.33</i>	0.33	0.64	5.98	<i>42.76</i>	savannas
24	2.5	10.10	0.35	0.25	1.88	1.16	5.47	grassland

i	k	bc	cc	C	$\langle N_i \rangle$	$\langle t_s \rangle$	$\langle t_{fp} \rangle$	Dominant land cover
b geese								
1	6.5	73.84	0.62	0.17	12.66	2.50	1.14	croplands/natural veget.
2	2.5	<i>0.00</i>	0.43	<i>1.00</i>	1.47	3.78	10.61	croplands/natural veget.
3	3.5	10.08	0.48	0.55	0.89	2.27	7.36	croplands
4	6.0	63.84	0.44	0.19	2.48	1.31	1.84	croplands
5	3.0	17.93	0.47	0.58	1.87	3.29	7.59	croplands/natural veget.
6	2.0	2.07	0.33	0.42	1.69	2.98	6.24	croplands
7	2.5	6.49	0.35	0.40	0.42	0.78	8.47	croplands/natural veget.
8	2.0	12.95	0.41	0.33	2.80	2.90	4.11	croplands
9	3.5	30.22	0.40	0.26	3.20	2.43	2.93	mixes forest
10	5.5	47.50	0.43	0.15	2.22	1.06	2.47	mixes forest
11	4.0	26.86	0.37	0.21	1.67	1.09	2.52	mixes forest
12	1.0	<i>0.00</i>	0.31	<i>1.00</i>	1.46	7.23	<i>21.08</i>	open shrublands
13	2.0	23.99	0.43	0.25	2.58	2.74	4.80	mixed forest
14	3.0	19.13	0.42	0.30	4.03	2.45	2.54	open shrublands
15	4.0	51.50	0.50	0.17	2.75	1.32	3.88	open shrublands
16	1.0	10.60	0.36	0.00	3.11	7.73	9.10	open shrublands
17	1.0	<i>0.00</i>	0.34	0.00	2.43	5.98	10.80	open shrublands

Table 6.1: Centrality and clustering characteristics, staying and passage times and land cover specifications for each specified migration network node. For each node of the (a) white stork network and the (b) greater white-fronted goose network we present values of the degree k , betweenness centrality bc , closeness centrality cc and the clustering coefficient C . Furthermore, the cumulative density $\langle N_i \rangle = \sum_{\tau} N_i(\tau)$ indicates how much each node is frequented during the year. Staying times $\langle t_s \rangle$ starting at the season of maximum density in the node (they vary with starting time) in units of months reveal how long a bird would stay on average in the respective region. Mean first passage times $\langle t_{fp} \rangle = 1/n \sum_j t_{fp,ij}$ in units of years are average times it takes to reach a node from any other one. All these measures indicate each node's importance for connectedness and network flow and dynamics. Values highlighted in bold mark nodes of high connectedness and importance, italic numbers indicate the opposite. Dominant land cover specifications provide a crude means for quality assessment of the resting habitat determination.

As good as our model mimics real bird migration, it is rather conceptual and has not been derived from first principles of bird migration like navigation, conditions at resting sites, responses to weather and food availability, etc. This may be strongly criticised by many empirical researchers. However, we have here intended to derive a null model as simple as possible that imitates bird migration movement for use in further analyses of climate change effects on migration routes and timing and epidemics spread by migratory birds. Naturally, the model can and should be refined by including environmental data, e.g. making transition rates dependent on not only time, but weather conditions, food availability, density and others.

The network model of bird migration applies only to species that migrate in a stepping-stone-like manner, i.e. that interrupt migration at some points for longer than just one or two days for resting and feeding up. Such is known for many species and suspected for others (Hedenström & Ålerstam, 1997). The two species that were selected, the white stork and the greater white-fronted goose, are of this type, but differ in their flight modes (flapping flight

vs. thermalling), feeding and migration routes.

With the presented methods of velocity analysis and clustering the determination of resting, breeding and wintering sites has been automated. Already such a method may be of value for the analysis of the respective bird specie's habitat preferences and, thus, for making habitat conservation decisions. The two case studies show that the determined resting regions are rather large and crudely determined, but the validations with land cover data emphasise its quality. Habitat types that are preferred by white stork are cropland, grassland and open shrubs, they avoid forested regions and deserts (Gerkmann, 2007). Dominant land cover types of network nodes match these preferences and are compatible with white storks basically feeding on amphibians and small reptiles that live in wetlands (Europe) as well as dry grasslands (Africa). Greater white-fronted geese favour habitat that is free of dense bushes: bogs, grassland and fields, cropland, marsh and coasts, but seem to be rather opportunistic in their habitat selection. If good grazing is available they also stay at the edge of forests for some days during migration (Kruckenberg *et al.*, 2008). Our findings of dominant land cover types of the resting node areas match with these preferences and conform to their feeding requirements, young grass during migration and insects rich of protein in the breeding areas.

Furthermore, the network node regions for the white stork coincide nicely with results by Gerkmann (2007) and Berthold *et al.* (2006), even if we did not use habitat characteristics like them. Increased data quality, density and availability will improve habitat detection and decrease the spatial scale of breeding, wintering and resting regions. Additionally, we rate the different regions by a number of measures of connectivity and intensity of usage, which gives a more detailed picture for conservation importance than only a list of resting areas.

The two here derived migration networks are very similar in structure and dynamic properties. They show characteristics of small worlds and are similar to other transportation systems. The novel issue of our model is that network structure changes periodically with time of the year, which makes processes on this interchanging network very complex. In general, one may propose that wintering and breeding nodes are more important for network connectivity, which agrees nicely with biological knowledge of population dynamics. For the goose network results, we would ask for some caution, because sample size here is very small and most individuals were caught and equipped with a transmitter in the same area. Such may bias results towards the importance of just this region (the western European part of the geese's wintering range).

Concluding, we developed a model of bird migration that is novel in several ways. It provides a representation of migration in time and space and introduces the notion of network theory to bird migration theory. We use very recent, high quality movement data for parameterisation: satellite telemetry with GPS and Doppler localisations. This kind of data is well suited for such a regular, dynamic process. Furthermore, the proposed Markov process with seasonally periodic, time-dependent transition rates of a von Mises functional form is a relatively new concept in modelling (but see Gardner *et al.*, 2006; Gee *et al.*, 2006), especially in ecological modelling (Patoucheas & Stamou, 1993). Thus, we not only present a new statistical analysis of migration movement data, but present a relatively new type of model with highly complex dynamics for ecological modelling. In an ecological context this null model of migration can be extended and used for the examination of population dynamics of migratory species, disease spread by migratory species and simulations of how different scenarios of habitat loss and changes of environmental conditions may influence migration dynamics in the future.

Acknowledgements

We are grateful to the Max-Planck-Institute for Ornithology – Vogelwarte Radolfzell for providing us with satellite telemetry data of white storks. Many thanks to the Vogelschutz-Komitee e.V. (VsK, Hamburg), the Alterra Institute (Wageningen) and the Dutch Society of Goose-catchers for financial and technical support for geese catching as well as tracking. Furthermore, we want to thank D. Syga and R. Tönjes for helpful discussion on the modelling approach. This work was funded by the VW-Stiftung and the BMBF.

7 General Discussion

Networks have become widely used for the study of complex, real world systems. Structural properties and characteristics of processes on networks have provided important insights into systems like the world wide web, contact networks, food webs, gene regulation systems and others (Newman, 2003b). Especially the examination of biological systems networks is proposed more and more important (Jeong *et al.*, 2000; Farkas *et al.*, 2003; Jordano *et al.*, 2003; Urban *et al.*, 2009). In this work we have developed two types of important biological networks, examined their complexity and discerned several general properties with methods of network theory.

Often, it is clear that a certain system has a network structure, as for example for the internet, contact networks and ship traffic between ports. However, there are also systems that seem not to be of network-like structure at first glance. So, for example, the movement of migratory birds. With methods from random walk theory we have statistically examined migratory movement data of the white stork (**paper III**). It emerged that its movement is composed of two different modes, migration and resting. Thus, it is a stepping stone process that can be described by a simplified, discrete process of migration on a network and continuous diffusion models do not have to be applied (Turchin, 1998).

For the quantification of large-scale biological and ecological systems only recently appropriate data have become available. By international cooperation of local data sampling, data of long-distance movement and transportation processes can be compiled. However, by new techniques like satellite telemetry and other technical systems of globally interacting devices one can recently obtain such data in a less labour intensive and error-prone way. In **paper I** and **III** we presented different kinds of such newly available data, ship positions by the Automatic Identification System (AIS) and localisations of migratory birds by GPS and satellite telemetry.

To analyse the movement structure of systems like the cargo ship trade and bird migration one first needs to quantify their flows reasonably well. We have achieved this by developing sensibly weighted networks of the two mentioned systems from high quality data (**papers I, V**). A comprehensive framework of methods for the development and structural analysis of such transportation networks has so far not been available. Therefore, considering a wide selection of different techniques from network theory we provide a methodology adapted for examining issues of transportation on large-scale biological networks (**paper I, II, V**). Techniques for network assessment provide small- and large-scale structural characteristics, node centralities and spread properties. The given methods are applicable for the development and characterisation of further transportation networks.

We have not only structurally described the mentioned transportation networks but also discussed them with focus on the global spread of organisms. In this context we compared them with other real world transportation networks and set our findings into context with bioinvasion, epidemics spread and other issues of globalisation. We realised and want to promote that

network theory can be used in bioinvasion studies and become an important tool for preliminary risk analyses (**paper II**). The same can be stated for bird migration studies and the spread of avian diseases. However, here one has to carefully consider seasonality, as has been confirmed by random walk analyses (**paper III**). Consequently, for the description of bird migration simple, homogeneous transportation networks are not sufficient. Therefore, we introduced non-homogeneous Markov processes of transportation on a network to migration ecology (**paper V**). In the wake of these main analyses, we characterised bird migration movement patterns in detail and found structures comparable to previous studies of optimal foraging (**paper IV**; Viswanathan *et al.*, 1999), i.e. Lévy flights. The consequent superdiffusivity of migration movement should be considered in further modelling, because it may have strong impact on spread rates of avian diseases. The weighted and geographically embedded structure of the migration network considers this property.

7.1 Transportation networks in comparison

7.1.1 The global cargo ship network

We have developed the global cargo ship network (GCSN) from real ships' trajectories of consecutively called at ports. Therefore, it sensibly maps global trade conducted via ships. Results from **paper I** show that link weight distributions of the GCSN agree reasonably well with predictions from the gravity model developed by Drake & Lodge (2004, see above). However, direct comparisons between single observed and expected link weights differ much. Therefore, we propose that the gravity model approximation is not sufficiently realistic, real ship traffic is more complex. Furthermore, when comparing the list of ports of the ship network by Drake & Lodge (2004) (DLNW) and the GCSN, we find great deviations. Many ports of great strength and importance for global trade are not included in the DLNW. They are for example Port Said, Jebel Ali in the United Arab Emirates, Xiamen in China and Kawasaki in Japan. Also in the GCSN some ports are missing, but they are mainly small ports situated in Africa and Southern America and usually contribute very little to the global flow of goods. One exception is the port of Rotterdam. It is not explicitly included in the GCSN, but its different parts appear as single ports, like e.g. Maasvlakte and Europort. As they are very central ports in the GCSN, we propose that their inclusion in that way is sufficient. Concluding, we find it more acceptable to disregard the ports missing in the GCSN than the ones missing in the DLNW. In a study by Tatem *et al.* (2006) the DLNW has been used for quantifying bioinvasion risk. Because of the shortcomings of it we recommend to use the GCSN for such studies instead.

In line with the work of Drake & Lodge (2004) we conducted a preliminary comparative analysis if it is better to isolate specific, strongly connected ports or if bioinvasion is prevented more readily by decreasing general per ship transition rates (**paper II**). Our results confirm their findings that a decrease of the transmission rates by a factor of ten has more impact than the deletion of a single port. Additionally, we put these results into a quantitative framework. It provides indications of how a combined approach of ballast water treatment in the ports and onboard the ships can best decrease bioinvasion rates. This is a novel analysis and can be useful to support management decisions.

The global cargo ship network (GCSN) shows several properties that are similar as well as some that are different to other transportation networks. The network most sensible to compare the GCSN with is the worldwide airport network (WAN). It is formed by a composition of airports connected by flights, but in contrast to the GCSN mainly passengers are transported (Barrat *et al.*, 2004; Guimera *et al.*, 2005). One important structural difference is that the WAN is organised in single flights forth and back between pairs of airports, whereas the GCSN incorporates many circular routes of container ships that are travelled through in only one direction. Thus, the GCSN is highly asymmetric and should be mapped as a directed and by ship numbers weighted network. It should not be symmetrised like the WAN. Ships' ports also seem to be more strongly connected, because the mean degree $\langle k \rangle$ of the GCSN by far surpasses that of the WAN. Most other simple structural properties of the two transportation networks are rather similar. Clustering is very strong in both systems, and weighted clustering coefficients exceed the unweighted ones. This indicates that interconnected clusters are mainly formed by very frequently and strongly travelled links. An interesting property is the scaling relation between the strengths and degrees of the ports or airports, respectively. Both show a relationship of the form $\langle s(k) \rangle \propto k^\alpha$ with α being approximately 1.5. This indicates that the such defined superlinear increase of node strength with degree may be some general property of social transportation networks.

Cargo ship traffic is dominated by three types of ships: oil tankers, container ships and bulk carriers. The three respective ship type specific cargo networks mainly reveal differing properties. This is caused by their varying patterns of travel. Container ships follow scheduled, circular routes, but oil tankers and bulk dry carriers operate more irregularly between pairs of ports. The GCSN and its subnetworks are not scale-free in their degree distributions. This is different in comparison to other real world networks, e.g. the world wide web (Barabási & Albert, 1999). The link weight distributions as well as the node strength distributions, however, are reminiscent of a power law, thus indicating scale-free behaviour. The slopes of these relationships differ according to the ship types. Container ships, for example, reveal a less steep decrease of abundance with distance. Surprisingly, when we examined motif distributions the GCSN and its subnetworks showed very similar patterns that additionally even coincided with motif distributions of the world wide web and several other social networks. The patterns of higher frequency include mainly closed triangles which coincides with the high clustering of the GCSN. This property seems to be another unifying feature of socially optimised transportation networks.

7.1.2 Seasonal bird migration networks

In **paper III** we confirm that migration patterns of the white stork vary with season and show two different modes: slow, undirected resting and fast, directed migration. Migration movement of the white stork is reminiscent of ballistic movement for time differences of less than one month and merely random for longer times. For time differences of half a year up to three years we obtain a recurrent behaviour, as was expected for such periodic movement as bird migration. Turning angles show a peak at zero (ballistic movement), are rather evenly distributed for intermediate angles and show another peak at 180° . This was at first surprising, but can be explained by the fact that during breeding and wintering storks have a constant

night place to which they return each night.

Following these results, we find it reasonable to model this migration process on a network of breeding, resting, moulting and wintering places. This agrees well with the concept of stepping stone migration that several bird species conduct (Berthold *et al.*, 2006). Movement characteristics of the white stork differ between the seasons as is intuitively clear from the concept of seasonal return migrations. Bird migration is a very complex process, and the selection of simplifying assumptions determines the usefulness of the resulting movement model. Therefore, we developed a seasonally changing, non-homogeneous Markov model on a network of bird migration, i.e. the transition rates between the different staging regions depend on the season. In contrast to unweighted networks and networks, like the GCSN, the links of which are weighted by (in time) constant intensities of travel, the bird migration network is more complex. Transition rates are seasonally driven and links thus associated with functions. We decided to use a modified von Mises function that is the circular equivalent to the Gaussian distribution. It is characterised by three parameters that are easily interpreted for the movement process: the main transition time (phase), the maximum transition intensity (amplitude) and the length of the transition interval.

Such a seasonally driven network is, to our knowing, novel in network theory and dynamics can become extremely complex (but see Gardner *et al.*, 2006). A similar approach has been conducted by Lämmer *et al.* (2006) who modelled traffic flows using phase oscillators as nodes representing red lights that are linked by streets' traffic flows. In their study each node was given a phase and amplitude, which is sensible for the system they explored. However, this is not the case for bird migration, because resting regions are sometimes used during spring and autumn migration and would thus have two times of peak densities (phases). Instead associating network links with phases and amplitudes is more straight forward in this context. We recommend that it is advisable to consider both possibilities for modelling seasonal or else periodically driven processes on a network.

The use of networks and graph theory for ecological studies has been proposed earlier (Urban *et al.*, 2009), but the only study using a non-homogeneous Markov model is by Patoucheas & Stamou (1993). They not only carefully defined the model, derived properties of it and pointed out possible fields of applications in ecology, but also described seasonal changes of species richness and population density of zoobenthos in Thermaikos Gulf in Greece. From comparing transition probabilities and population densities in the four different seasons the authors conclude that density reductions from winter to summer are correlated with oxygen depletion. Thus, this study was also initiated by concern about anthropogenically induced habitat changes. In contrast to our Markov model of bird migration their study was much simpler. The time steps were larger, only four different times of the year were differentiated, contrary to our 48 time intervals per year. This may be a consequence of their poorer resolved data and differences in the question of study. The authors considered not movement, but dynamics of ecosystem succession. Beyond the methodology of that study we suggest a functional form of the transition rates as well as network properties for the characterisation of the non-homogeneous Markov chain (**paper V**). Concluding, we have taken up the suggestions of Patoucheas & Stamou (1993) and extended the proposed uses and methods for examining seasonal ecological systems with network theory and non-homogeneous Markov models.

In **paper V** the cyclic network model of bird migration has been parameterised for two

different bird species, the white stork and the greater white-fronted goose. Both species conduct a stepping stone like migration and well resolved movement data were available. The migration routes and timing are different for storks and geese so that similarities of the derived networks and transition processes could not be expected. However, properties of the time cumulative migration networks of both species are very similar. They have small shortest path lengths, are highly efficient, asymmetric and flow patterns indicate similar forth and back movements with some degree of branching. The motif distributions coincide nicely with each other and with additional networks of biological transport (Milo *et al.*, 2004), a signal transduction network, neurons networks and the transcription network of *Drosophila melanogaster*. Thus, the two migration networks have similar structural characteristics, some of which they share with other biologically optimised transportation networks, even if they are small- rather than large-scale ones.

7.1.3 Differences and similarities of ships and birds

In the previous paragraphs it has become obvious that several characteristics of the global cargo ship network coincide with other transportation networks. Also the migration networks reveal characteristics of transportation structures. Some are similar to those of the GCSN, others differ, but many cannot be compared, because it is not sensible to determine highly complex measures of the rather small migration networks. In detail, both the GCSN and the two derived migration networks show signs of small worlds, they are highly efficient at low costs, have small average shortest paths and high clustering. Furthermore, they are highly asymmetric and degree distributions are heavy tailed. Strikingly different are their motif distributions, they can be associated with different superfamilies (Milo *et al.*, 2004). The GCSN motifs were attributed to a number of social networks, whereas the migration networks' motif distributions are similar to some biological networks. Thus, one may conclude that the GCSN is optimised according to social aspects, whereas the migration networks reveal biologically optimised patterns. These differences agree well with common knowledge and understanding. By calculating centralities for both network types another common property emerges. There is a small fraction of nodes that are extremely important for the connectivity of the networks. Consequently, a relatively quick spread of certain agents through the networks is likely but can also be effectively obstructed by the deletion of highly central nodes. For the propagation of bioinvasion or disease spread on the developed networks such properties are of importance. So, there are several similarities of the global movement of ships and that of some migratory birds.

7.2 Superdiffusion of bird migration movement

In **paper IV** it has been clearly shown that the movement of several migratory but also non-migratory bird species is superdiffusive and resembles Lévy flights. Striking was the scale-free behaviour of displacement distances in time and space over several decades for species that are very different in their behaviour, ecology and migration routes. Our study contributes to the controversial debate about the general existence of biological Lévy flights (Edwards *et al.*, 2007; Sims *et al.*, 2008; Viswanathan *et al.*, 2008). We applied rigorous model selection

statistics and show that unbounded and truncated power laws fit the displacement data of all five examined species best. Thus, Lévy flights might in fact be more frequent in nature than previously thought.

Mechanisms that may be accountable for forming the observed movement patterns have been discussed in **paper IV**. Landmass distributions, climate conditions, heterogeneous habitat characters and other behavioural mechanisms were named, but data acquisition bias ruled out. For several other animal species as well as humans Lévy flight movement properties have been revealed (see above; Viswanathan *et al.*, 2008). However, mostly they comprised only data analyses over small spatial scales, and mechanisms of the emergence of the superdiffusive movement patterns were not examined. Exceptions were the studies by Viswanathan *et al.* (1999) and Benhamou (2007) that propose optimal foraging or a random movement of different modes (composite random walk) to form the Lévy-flight-like patterns.

The movement of all examined species is dominated by Lévy flight characteristics. From our results of **paper III** the composition of different random walks seems a reasonable mechanism to bring about the observed patterns for the white stork. Random walk measures showed that the movement of this species consists of two modes, migration and resting. In the dispersal kernels of the white stork and also the mute swan (see Fig. 5.1a,b,f) a small bend of the slope at a distance of about 100 km also indicates that the movement may consist of different overlaid modes. They are the seasonal migration and resting of the white stork and dismigration and foraging movement of the mute swan. However, in the data of the other four bird species such patterns did not emerge, so mode composition seems not to be a unifying property. Therefore, the network model (**paper III, V**) is only appropriate for species the movement of which shows this property (see also above).

An important aspect of migratory movement that should be kept in mind is the seasonality of habitat and climate conditions. We have shown that there are profound differences in walk characteristics of the white stork between the seasons (**paper III**). Thus, one should be careful interpreting results of data cumulated over the times of the year. We have included a seasonal dependency of large-scale movement in the network model, but how the analysis of Lévy flight patterns and possible mechanisms should be modified for this fact is left to further studies. One important issue regarding seasonality is the relationship between movement patterns and spatiotemporal habitat heterogeneities and changes. Seasonal changes in the climate and food availability are the driving factors of migration itself. Therefore, it may be optimal foraging on large spatial and temporal scales that in turn can explain the observed Lévy flight movement patterns of short-range as well as long-range migration.

7.3 Future perspectives of biological transportation

Whenever humans and animals move around the earth transportation of goods and smaller organisms is performed. Due to the current globalisation human movement has increased in intensity, range and velocity, and will continue to do so. Also animal movement, especially over long ranges, has been modified due to the rapid human induced climate change. However, in that case it is not clear to what extent and in which direction the changes happen. Some migratory bird species, for example, adapt their timing to advanced maximal food availability

in the breeding ranges. Others are not able to arrive earlier in their breeding areas, and even have to delay their spring migration because of disadvantageous environmental circumstances on the way (Jenni & Kéry, 2003). Thus, their time of breeding becomes decoupled from the peak food density and population densities may decrease. Not only migration timing, but also migration routes have changed because of habitat modification and climate change. It is not clear how these developments influence patterns of transportation, the spread of diseases and a possible homogenisation of the global biota. At that point it is sensible to develop models of such transportation systems and examine the effects of possible changes on the structure and spread dynamics of model outcomes. Naturally, high quality data are necessary to evaluate how well the models map reality and which mechanistic modules of the model apply. In a world of even more accelerated transportation spread characteristics will change drastically and much research is needed to understand its implications for the future functioning of ecosystems and human welfare.

In the here presented work we have taken a first step in this direction. We have developed simplified mappings of complex movement processes and conducted data analyses that point out important features thereof. The approach of network modelling and Markov processes incorporates several simplifications, not only the discretisations of space, but also the neglect of population dynamics, habitat characteristics, behavioural interactions and others. Some of these assumptions can be relaxed at further stages of model development and refinement. However, in the beginning of such studies it is advantageous that simple relationships can be uncovered from such clearly defined, simplified system.

As soon as more high quality bird movement data or transportation system quantifications become available, an important extension of our simple network models can be conducted. Not only will one be able to develop transportation networks of additional migratory species and human transportation systems, but transportation networks of e.g. different bird species should be merged. The spread characteristics of the resulting networks will give novel insights in the interaction properties of different transport vectors on bioinvasion and the global spread of diseases. They would furthermore allow for the development of global spread scenarios.

Models as simple as the presented networks are very useful for the prediction and the exploration of the general dynamics of spatiotemporal systems, but it does not extend our understanding of the exact processes that form the observed patterns. To obtain such insight one would have to develop mechanistic models and compare its outcomes with field data and experimental results. This could be individual based models as well as systems of differential equations (Turchin, 1998). In line with such analyses one may even discover which properties the observed movement and transportation systems are optimised for. As noted above, human induced transportation and social networks seem to differ from biologically optimised transportation networks. Costs and optimality may differ between these systems. Human business is often optimised by cost in terms of money and time, whereas bird migration movement may rather be driven by energy expenditure and synchronised with seasonally driven habitat characteristics.

Last but not least, a very natural extension of the presented studies is to relax the assumptions of the simple network approach and to incorporate, for example, population growth and environmental conditions. So can transition rates in the migration network depend not only on the season, but also on wind conditions or food availability in the network node regions. If

7 General Discussion

one is interested in the spread of bioinvasive organisms and diseases on networks population dynamics should be incorporated in the model, i.e. a population growth model of the invasive organism applied on each network node. Thus, the phases of establishment and proliferation of bioinvasion can be included in the modelling approach. The eradication of problematic invasives that have already established in a certain region is usually difficult and unpredictable (Mack *et al.*, 2000), therefore predictions and modelling outcomes should be regarded with caution and tested with real data.

Concluding, we provide methodologies for the quantification of movement in different systems and the examination of general transportation characteristics. Transportation networks were fully quantified and baseline movement statistics provided for indications of global spread of bioinvasive organisms and avian diseases. In the following we are very excited to see how these findings will be used for the development of more detailed transportation and spread models, as well as how model extensions can increase their predictive quality.

8 Summary

Movement is a fundamental property of life and almost all living creatures locomote around the earth in complex patterns at differing spatial and temporal scales. Often, movement incorporates the spread of e.g. small organisms, pathogens or seeds attached to the moving animals or other transportation vectors. Despite the fact that movement is natural and necessary, the extremely intensive human travel and cargo transportation presently lead to problems like global bioinvasion and epidemics spread.

Large-scale movement and transportation systems often possess an inherent network structure and can easily be modelled as commotion events between discrete habitat patches. In this work we developed network representations of the global cargo ship traffic as well as the migratory movement of some bird species. We have quantified movement patterns on these networks with the use of highly resolved AIS-trajectories, satellite telemetry and GPS data, and characterised it with different network measures. Furthermore, large-scale displacement patterns of several bird species were statistically described from ring-recapture data.

As humans travel the earth at unprecedented rates and intensities and migration routes of e.g. birds change due to climate change, spread patterns vary and are more complex than ever before. Thus, to obtain a sufficient understanding of the movement dynamics of ships and migratory birds is crucial for the management of further expansions of bioinvasive organisms that are detrimental to ecosystems functioning and infectious diseases that threaten human health. Based on the results of this thesis we discuss problems and possibilities to develop such management strategies.

In the first part of this work we have generated the global cargo ship network (GCSN) from real ship trajectories. It is a strongly directed network of all ports worldwide the links between which are weighted by yearly cargo capacities. We analysed its general structure and revealed several similarities with other transportation networks. The weighted GCSN shows high clustering, small average shortest paths and its link weight and node strength distributions are scale-free. This promotes quick spread and robustness to random traffic disruptions. Cargo ship traffic is dominated by ships of the three most important ship types: oil tankers, container ships and bulk carriers. The three according subnetworks reveal differences in their movement patterns. Container ships serve a regular schedule, whereas the other two perform irregular journeys according to demand. Additionally, key ports and community structures differ considerably between the subnetworks. Finally, the GCSN was compared with a simple gravity model of ship transportation, and clear differences of transmission intensities showed. Therefore, we conclude that it is more sensible to use the GCSN for studies of bioinvasion and other spread by ocean shipping.

In a subsequent analysis we take up the issue of marine bioinvasion by transport in ballast water and hull fouling of cargo ships. We discuss its present extents and implications for

the global biodiversity, ecosystem functioning and the global economy. By examining the structure of the GCSN in detail and with regard to spread we obtained new insights into the first stage of marine bioinvasion, the introduction of nonindigenous, aquatic species to novel regions. We confirm that the GCSN is a “small world”, and show that the global efficiency of the geographically embedded network is remarkably high. Furthermore, it is positively assortative and thus distinctly clustered into highly interconnected groups of ports. With different measures of centrality we discern which ports are most important for spread on the GCSN. They are basically ports that preserve network connectivity, the Suez and Panama Canal, Singapore and Shanghai. For the support of ballast water management decisions we provide a framework of the effects of node deletions and transmission rate decrease on spread. We propose to select a most practicable combination of in-port and onboard ballast water treatment to control further marine bioinvasion.

The second part of this thesis is concerned with the exploration, characterisation and network modelling of the large-scale process of bird migration. In a preliminary study we analyse satellite telemetry and ring-recapture data of the white stork. From mean squared displacement and turning angle distributions we find that its movement is composed of two different modes, (i) fast and directed migration, and (ii) slow and undirected resting and foraging. This coincides with the concept of stepping stone migration that has been empirically described for several bird species.

In a more detailed study of displacement distances, we compared the movement of five very different bird species, the white stork, barn swallow, chiffchaff, mallard and mute swan. For all of them unbounded and truncated power laws fit the displacement distributions better than exponential or lognormal distributions. This indicates that long-time bird movements resemble Lévy flights. Consequently, they seem to be much more frequent in nature than thought before, especially since recently the mere existence of biological Lévy flights has been questioned.

The Lévy flight characteristics dominate the movement patterns of all the examined species. However, for the white stork one may observe two slopes of the displacement histogramme for different intervals. This coincides with the previous findings of two different modes of movement. Therefore, we considered it sensible to disregard small-scale movement and develop a network model of the migration routes. Nodes and active links were derived from satellite telemetry and GPS trajectories of the white stork and the greater white-fronted goose. Localisations of birds with small flight velocity were clustered to breeding, resting and wintering regions. The vegetation characteristics of each of these regions were determined from satellite images, and they coincide nicely with habitat preferences of the stork and goose, respectively.

For modelling bird migration, a simple network structure is not sufficient. Migration is a rather complex process that is mainly driven by seasonality. Therefore, we incorporated a seasonal dependence into the migration flow model, mathematically defining it as a non-homogeneous Markov chain. The transition probabilities associated with the network links were weighted by a circular, unimodal function of the time of the year and fitted by timing found in the movement trajectories of the white stork and the greater white-fronted goose. So, we have developed this model as initial model of bird migration on a network.

Finally, we analysed the structure and spread characteristics of the cumulative and time specific migration networks and determined regions most important for migration flow. Both

networks have very small average shortest paths and are highly asymmetric. This confirms that they migrate on average with few stopovers and that the spring and autumn migration routes differ, especially regarding the usage of consecutive resting areas. Habitat areas that are most important to keep the network connected and that are frequently used by the birds of our data sets are (i) the breeding region in northern Europe and a resting and wintering area in central Sudan for the stork and (ii) the wintering region in the Netherlands and a few breeding places on the Siberian coast for the geese. Such indications may serve as support for habitat conservation decisions and risk assessment of the propagation of avian borne diseases.

Concluding, this thesis provides a special combination of the development and characterisation of different transportation networks. We developed networks of the global movement of cargo ships and migratory birds and parameterised the movement flow on them from high quality data. Detailed network analytical examinations revealed typical properties of transportation networks, but also novel insights into the movement patterns of ships and birds. We showed similarities of the two developed network types, contrary to the distinct differences of the processes they describe. The models shall now be extended by case studies with real population dynamics and environmental conditions and bioinvasion/disease spread dynamics. The presented results provide an important basis for that. However, the described, fundamental spread characteristics of the examined systems already reveal different approaches to cope with the negative impacts of modified movement patterns, like bioinvasion and the global spread of infectious diseases.

9 Zusammenfassung

Bewegung ist eine grundlegende und notwendige Eigenschaft des Lebens. Fast alle Lebewesen bewegen sich in komplexen Mustern und auf verschiedenen räumlichen und zeitlichen Skalen über die Erde. Dabei werden durch Tiere und menschliche Transportvektoren sekundäre Ausbreitungen von z.B. kleinen Organismen, Krankheitserregern und Samen stark begünstigt. Menschliches Reisen und Gütertransport nie dagewesener Häufigkeit und Geschwindigkeit sowie die Änderung von Migrationsrouten verschiedener Tierarten aufgrund des Klimawandels führen derzeit zu Problemen wie globaler Bioinvasion und der Ausbreitung von Epidemien.

Großskalige Tierbewegungen und Transportsysteme besitzen oft eine inherente Netzwerkstruktur und können leicht als Bewegungsereignisse zwischen diskreten Regionen modelliert werden. Das Ziel dieser Arbeit besteht in der Entwicklung und Beschreibung von Transportnetzwerken, die zum Einen den globalen Güterschiffsverkehr oder zum Anderen die Zugbewegungen zweier verschiedener Vogelarten abbilden. Bewegungsmuster auf den Netzwerken wurden mit hochaufgelösten Bewegungsdaten, AIS-Trajektorien, Satellitentelemetrie und GPS, quantifiziert und mit verschiedenen Netzwerkmaßen charakterisiert. Des Weiteren wurden großskalige Ausbreitungsmuster verschiedener Vogelarten anhand von Beringungsdaten statistisch beschrieben. Auf der Grundlage dieser neu gewonnenen Ergebnisse konnten Probleme und Möglichkeiten zur Bekämpfung der fortschreitender globalen Ausbreitung von Organismen diskutiert werden. Dies bezog sich insbesondere auf bioinvasiver Arten, die Ökosysteme schädigen und Infektionskrankheiten, welche die menschliche Gesundheit gefährden.

Im ersten Teil dieser Arbeit wurde das Globale CargoSchiffsNetzwerk (GCSN) aus realen Schiffstrajektorien zwischen den Häfen generiert. Das GCSN ist ein gerichtetes Netzwerk, dessen Links mit der Summe der Güterkapazitäten der verkehrenden Schiffe gewichtet wurden. Eine Analyse der Netzwerkstruktur zeigte zahlreiche Gemeinsamkeiten mit anderen Transportnetzwerken auf. Das gewichtete GCSN ist stark geclustert, hat im Mittel sehr kleine kürzeste Pfade und seine Kanten- und Knotengewichte folgen skalenfreien Verteilungen. Dies begünstigt schnelle Ausbreitung auf dem Netzwerk und Robustheit gegenüber zufälligen Störungen im Schiffsverkehr. Der Verkehr von Handelsschiffen ist dominiert von Schiffen der drei wichtigsten Typen: Öltankern, Containerschiffen und Massengutfrachtern. Die drei zugehörigen Teilnetzwerke zeigen deutliche Unterschiede in ihren Bewegungsmustern. Containerschiffe fahren nach einem regulären Fahrplan, wohingegen die anderen beiden Schiffstypen abhängig von Angebot und Nachfrage unregelmäßig verkehren. Des Weiteren unterscheidet es sich für die jeweiligen Teilnetzwerke, welche Häfen besonders zentral sind und wie die Strukturen der Handelsbeziehungen organisiert sind. Der Vergleich des GCSN mit einem Gravitationsmodell des Schiffshandels zeigte deutliche Unterschiede der Flussraten. Daraus kann geschlossen werden, dass es für Untersuchungen von Bioinvasion und anderen Ausbreitungen per Schiff realistisch ist das GCSN zu verwenden.

In darauf folgenden Analysen wurde das Thema der marinen Bioinvasion aufgegriffen, insbesondere in Bezug auf den Transport von Organismen im Ballastwasser und Bewuchs des Schiffsrumpfes. Die derzeitigen Ausmaße von mariner Bioinvasion wurden aufgezeigt und ihre ungünstigen Auswirkungen auf die globale Biodiversität, die Funktion von Ökosystemen und die weltweite Ökonomie diskutiert. Strukturbeschreibungen und Untersuchungen der Ausbreitungsmöglichkeiten von Organismen auf dem GCSN liefern neue Einsichten zur marinen Bioinvasion. Die Ergebnisse beziehen sich hauptsächlich auf die erste Phase mariner Bioinvasion, nämlich der Einführung nicht heimischer, aquatischer Arten in neue Regionen. Es konnte bestätigt werden, dass das GCSN zahlreiche Eigenschaften einer "kleinen Welt" (small world) besitzt. Außerdem wurde gezeigt, dass die globale Effizienz des geografisch eingebetteten Netzwerks bemerkenswert hoch ist. Das Schiffsnetzwerk ist positiv assortativ und somit in ausgeprägte, stark zusammenhängende Gruppen von Häfen aufgeteilt. Mit verschiedenen Zentralitätsmaßen wurden Häfen bestimmt, die wichtig für den Transport und die Konnektivität des GCSN sind, nämlich der Suez- und Panama-Kanal, Singapore und Shanghai. Schließlich wurden die Auswirkungen der Isolation bestimmter Gebiete und der Reduzierung der Übergangsraten zwischen Häfen auf die Ausbreitungsgeschwindigkeit berechnet. Dies beschreibt die verschiedenen Arten von Ballastwasserreinigung. Der Zusammenhang von der Reinigung von Ballastwasser in den Häfen und an Bord der Schiffe und ihr Effekt auf Bioinvasion wurde dargestellt. Anhand dieser Ergebnisse wird empfohlen, eine in die Praxis umsetzbare Kombination der beiden Reinigungsarten auszuwählen, um Strategien zur Bekämpfung mariner Bioinvasion zu entwickeln.

Der zweite Teil dieser Dissertation befasst sich damit den großskaligen Prozesses des Vogelzugs zu untersuchen, statistisch zu beschreiben und als Migrationsnetzwerk zu modellieren. In einer Vorabstudie wurden Satellitentelemetrie- und Beringungsdaten des Weißstorchs bezüglich seiner Bewegungsmuster ausgewertet. Aus Resultaten zur mittleren quadratischen Abweichung und der Drehwinkelverteilung ist zu schließen, dass sich die Bewegung aus zwei verschiedenen Modi zusammensetzt, (i) schnellem, gerichtetem Zug und (ii) langsamem, ungerichtetem Rasten und Futtersuchen. Dies bestätigt das Trittstein-Vogelzug-Konzept, welches bereits für verschiedene Vogelarten empirisch beschrieben wurde.

In einer detaillierteren Studie der Verbreitungsdistanzen von Vögeln wurden die Bewegungsmuster von fünf sehr unterschiedlichen Vogelarten verglichen: dem Weißstorch, dem Zilpzalp, der Rauchschnalbe, der Stockente und dem Höckerschwan. Verbreitungsverteilungen für all diese Arten können besser durch unbegrenzte und "gestutzte" (truncated) Potenzgesetze als Exponential- oder Lognormalverteilungen beschrieben werden. Dies deutet darauf hin, dass Langzeitvogelbewegungen Lévyflügen ähneln. Kürzlich wurde die Existenz von biologischen Lévyflügen in Frage gestellt. Die hier präsentierte Arbeit zeigt jedoch, dass sie in der Natur viel häufiger vorzukommen als bisher vermutet.

Die Lévyflugeigenschaften dominieren die Bewegungsmuster aller untersuchten Arten. Es können jedoch für den Weißstorch zwei verschiedene Steigungen der Verbreitungsverteilungen beobachtet werden. Das stimmt mit den vorigen Resultaten, die zwei Bewegungsmodi zeigen, überein. Deshalb konnten die kleinskaligen Bewegungen des Weißstorchs und anderer Trittstein-Zug-Arten vernachlässigt werden, um ein Netzwerkmodell der Zugrouten zu entwickeln. Knoten und Kanten wurden durch Analysen von Satellitentelemetrie- und GPS-

Trajektorien des Weißstorches und der Blässgans bestimmt. Insbesondere wurden Lokalisationen von Vögeln mit geringer Fluggeschwindigkeit in Netzwerkknoten zusammengefasst. Die Vegetationseigenschaften dieser Regionen wurden aus Satellitenbildern abgeleitet und stimmen gut mit den Habitatansprüchen der Störche bzw. Gänse überein.

Das entwickelte, einfache Netzwerk aus Brut-, Rast- und Überwinterungsregionen kann den dynamischen Prozess des Vogelzugs nicht vollständig modellieren. Migration ist sehr komplex und wird zu einem Großteil von der Saisonalität in verschiedenen Regionen bestimmt. Deshalb wurde eine Abhängigkeit von den Jahreszeiten in unser Vogelzugmodell eingebunden. Mathematisch wurde es als nicht-homogene Markovkette definiert. Die Übertragungswahrscheinlichkeiten, welche mit den Netzwerkkanten assoziiert sind, wurden mit einer zirkulären, unimodalen Funktion der Jahreszeit gewichtet und an die Bewegungszeiten in den Weißstorch- und Blässgansdaten angepasst. Somit wurde ein erstes quantitatives Vogelzugnetzwerkmodell entwickelt.

Darüber hinaus wurden die Struktur und Ausbreitungseigenschaften der kumulativen und zeitspezifischen Vogelzugnetzwerke analysiert und Gebiete bestimmt, die besonders wichtig für die Erhaltung durchgehender Zugbewegungen sind. Beide Netzwerke haben sehr kleine gemittelte kürzeste Pfade zwischen den Knoten und sind asymmetrisch. Dies bestätigt, dass die Vögel während des Zugs im Mittel wenige Zwischenstopps einlegen und die Routen des Herbstzuges besonders bezüglich der Nutzung der Rastgebiete von denen des Frühjahrzuges abweichen. Von den Vögeln häufig benutzte und für die Konnektivität des Netzwerks wichtige Regionen sind: (i) das Brutgebiet in Nordeuropa und ein Rastgebiet im zentralen Sudan für die Störche und (ii) das Überwinterungsgebiet in den Niederlanden und einige Brutgebiete an der Sibirischen Küste für die Gänse. Solche Ergebnisse können als vorläufige Hinweise für Umweltschutzentscheidungen und Risikoabschätzungen zur Ausbreitung von Infektionskrankheiten durch Zugvögel dienen.

Die vorliegende Arbeit stellt eine besondere Kombination der Entwicklung und Charakterisierung verschiedener Transportnetzwerke dar. Es wurden Netzwerke der globalen Bewegungsmuster von Handelsschiffen und Zugvögeln entwickelt und der Bewegungsfluss darauf mit hochqualitativen Daten parametrisiert. Detaillierte Netzwerkanalysen zeigen typische Eigenschaften anderer Transportnetzwerke. Des Weiteren wurden Ähnlichkeiten der beiden entwickelten Netzwerke hervorgehoben, obwohl die von ihnen beschriebenen Prozesse sehr unterschiedlich sind. Ausbreitungsmodelle auf den Netzwerken sollen nun mit speziellen Fallstudien unter Einbezug von Populationswachstum, Umwelteinflüssen und Bioinvasion bzw. Epidemieausbreitung entwickelt werden. Die vorliegenden Resultate bieten eine wichtige Grundlage dafür. Die dargestellten, grundlegenden Ausbreitungseigenschaften der untersuchten Systeme zeigen verschiedene Ansätze zur Bekämpfung der negativen Auswirkungen veränderter globaler Bewegungsmuster, wie mariner Bioinvasion und der globalen Verbreitung von Infektionskrankheiten auf.

10 Bibliography

- Albert, R., & Barabási, A.-L. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics*, **74**(1), 47–97.
- Albert, R., Jeong, H., & Barabási, A.-L. 2000. Error and attack tolerance of complex networks. *Nature*, **406**, 378–382.
- Alerstam, T. 2006. Conflicting evidence about long-distance animal navigation. *Science*, **313**(5788), 791–794.
- Alerstam, T., & Hedenström, A. 1998. The development of bird migration theory. *Journal of Avian Biology*, **29**, 343–369.
- Anderson, R. M., & May, R. M. 1991. *Infectious diseases of humans - dynamics and control*. Oxford University Press, New York.
- Austin, D., & McMillan, J. I. 2003. A three-stage algorithm for filtering erroneous Argos satellite locations. *Marine Mammals Science*, **19**, 371–383.
- Bailey, H., & Thompson, P. 2006. Quantitative analysis of bottlenose dolphin movement patterns and their relationship with foraging. *Journal of Animal Ecology*, **75**(2), 456–465.
- Bailey, N. T. J. 1975. *The mathematical theory of infectious diseases*. Griffin, London.
- Bairlein, F. 2001. Results of bird ringing in the study of migration routes. *Ardea*, **89**(1), 7–19.
- Bairlein, F. 2003. The study of bird migrations - some future perspectives. *Bird Study*, **50**, 243–253.
- Barabási, A.-L., & Albert, R. 1999. Emergence of scaling in random networks. *Science*, **286**, 509–512.
- Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(11), 3747–3752.
- Barthelemy, M., & Flammini, A. 2008. Modeling urban street patterns. *Physical Review Letters*, **100**(13), 138702.
- Batschelet, E. 1981. *Circular statistics in biology*. Academic Press, London.
- Bauer, H.-G., Bezzel, E., & Fiedler, W. 2005. *Das Kompendium der Vögel Mitteleuropas*. Aula, Wiebelsheim. (in German).

- Beekman, J. H., Nolet, B. A., & Klaassen, M. 2002. Skipping swans: fuelling rates and wind conditions determine differential use of migratory stopover sites of Bewick's Swans *Cygnus Bewickii*. *Ardea*, **90**, 437–460.
- Begon, M. E., Townsend, C. R., & Harper, J. L. 2006. *Ecology: from individuals to ecosystems*. Blackwell, Oxford.
- Ben-Avraham, D., & Havlin, S. 2000. *Diffusion and reactions in fractals and disordered systems*. Cambridge University Press, Cambridge.
- Benhamou, S. 2007. How many animals really do the Lévy walk? *Ecology*, **88**(8), 1962–1969.
- Berthold, P. 2001a. *Bird migration. A general survey*. Oxford University Press, Oxford.
- Berthold, P. 2001b. Bird migration: a novel theory for the evolution, the control and the adaptability of bird migration. *Journal for Ornithology*, **142**, 148–159.
- Berthold, P., & Terrill, S. B. 1991. Recent advances in studies of bird migration. *Annual Reviews of Ecological Systems*, **22**, 357–378.
- Berthold, P., Nowak, E., & Querner, U. 1992. Tracking white storks on migration through Europe to Africa. *ARGOS Newsletter*, **43**, 1–3.
- Berthold, P., Nowak, E., & Querner, U. 1997. Eine neue Dimension der Vogelforschung: Die Satelliten-Telemetrie. *Falke*, **44**, 134–140. (in German).
- Berthold, P., van den Bossche, W., Jakubiec, Z. and Kaatz, C., Kaatz, M., & Querner, U. 2002. Long-term satellite tracking sheds light upon variable migration strategies of white storks (*Ciconia ciconia*). *Journal of Ornithology*, **143**, 489–495.
- Berthold, P., Kaatz, M., & Querner, U. 2004. Long-term satellite tracking of white stork (*Ciconia ciconia*) migration: constancy versus variability. *Journal of Ornithology*, **145**(4), 356–359.
- Berthold, P., van den Bossche, W., Kaatz, M., & Querner, U. 2006. Conservation measures based on migration research in white storks (*Ciconia ciconia*, *Ciconia boyciana*). *Acta Zoologica Sinica*, **52**, S 211–214.
- Blasius, B., Huppert, A., & Stone, L. 1999. Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature*, **399**(6734), 354–359.
- Blitvich, B. J. 2008. Transmission dynamics and changing epidemiology of West Nile virus. *Animal Health Research Reviews*, **9**, 71–86.
- Bollobás, B., & Riordan, O. 2006. *Percolation*. Cambridge University Press, Cambridge.
- Bonacich, P. 1972. Factoring and weighting approaches to status scores and clique identification. *Journal of the Mathematical Society*, **2**, 113–120.

- Bos, D., Drent, R. H., Rubinigg, M., & Stahl, J. 2005. The relative importance of food quantity and quality for patch and habitat choice in Brent Geese. *Ardea*, **93**, 5–16.
- Brockmann, D., Hufnagel, L., & Geisel, T. 2006. The scaling laws of human travel. *Nature*, **439**, 462–465.
- Buhl, J., Gautrais, J., Reeves, N., Sole, R. V., Valverde, S., Kuntz, P., & Theraulaz, G. 2006. Topological patterns in street networks of self-organized urban settlements. *European Physical Journal B*, **49**(4), 513–522.
- Burnham, K. P., & Anderson, D. R. 1998. *Model selection and multimodel inference. A practical information-theoretic approach*. Springer, New York.
- Callaway, D. S., Newman, M. E. J., Strogatz, S. H., & Watts, D. J. 2000. Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, **85**, 5468–5471.
- Canright, G. S., & Engo-Monsen, K. 2006. Spreading on networks: a topographic view. *Complexus*, **3**, 131–146.
- Carlton, J. T. 1985. Trans-oceanic and interoceanic dispersal of coastal marine organisms - the biology of ballast water. *Oceanography and Marine Biology*, **23**, 313–371.
- Carlton, J. T. 1996a. Marine bioinvasions: the alteration of marine ecosystems by nonindigenous species. *Oceanography*, **9**, 36–43.
- Carlton, J. T. 1996b. Pattern, process, and prediction in marine invasion ecology. *Biological Conservation*, **78**, 97–106.
- Carlton, J. T., & Geller, J. B. 1993. Ecological roulette: the global transport of nonindigenous marine organisms. *Science*, **261**, 78–82.
- Clauset, A., Shalizi, C. R., & Newman, M. E. J. 2009. Power-law distributions in empirical data. *SIAM Review*, forthcoming. arXiv:0706.1062v1.
- Colwell, R. R., & Huq, A. 1994. Environmental reservoir of *Vibrio cholerae*. The causative agent of cholera. *Annals of the New York Academy Society*, **740**, 44–54.
- Costa, L. da F., Rodrigues, F. A., Travieso, G., & Villas Boas, P. R. 2007. Characterization of complex networks: a survey of measurements. *Advances in Physics*, **56**, 167–242.
- Creutz, G. 1985. *Der Weißstorch. Ciconia ciconia*. Ziemsen Verlag, Wittenberg Lutherstadt. (in German).
- Crowl, T. A., Crist, T. O., Parmenter, R. R., Belovsky, G., & Lugo, A. E. 2008. The spread of invasive species and infectious diseases as drivers of ecosystem change. *Frontiers in Ecology and the Environment*, **6**, 238–246.
- de Montis, A., Barthelemy, M. and Chessa, A., & Vespignani, A. 2007. The structure of inter-urban traffic: a weighted network analysis. *Environmental Planning B: Planning and Design*, **34**, 905–924.

- Drake, J. M., & Lodge, D. M. 2004. Global hot spots of biological invasions: evaluating options for ballast-water management. *Proceedings of the Royal Society of London B*, **271**, 575–580.
- Drake, J. M., & Lodge, D. M. 2007. Hull fouling is a risk factor for intercontinental species exchange in aquatic ecosystems. *Aquatic Invasions*, **2**, 121–131.
- Edwards, A. M., Phillips, R. A., Watkins, N. W., Freeman, M. P., Murphy, E. J., Afanasyev, V., Buldyrev, S. V., da Luz, M. G. E., Raposo, E. P., Stanley, H. E., & Vishwanathan, G. M. 2007. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, **449**, 1044–1049.
- Eichhorn, G., Afanasyev, V., Drent, R. H., & van der Jeugd, H. P. 2006. Spring stopover routines in Russian Barnacle Geese *Branta leucopsis* tracked by resightings and geolocation. *Ardea*, **94**, 667–678.
- Elton, C. S. 1958. *The ecology of invasions by animals and plants*. Chapman and Hall, London.
- Erdős, P., & Rényi, A. 1959. On random graphs. *Publicationes Mathematicae*, **6**, 290–297.
- Erni, B., Liechti, F., & Bruderer, B. 2002. Stopover strategies in passerine bird migration: A simulation study. *Journal of Theoretical Biology*, **219**(4), 479–493.
- Farkas, I. J., Jeong, H., Vicsek, T., Barabási, A.-L., & Oltvai, Z. N. 2003. The topology of the transcription regulatory network in yeast, *Saccharomyces cerevisiae*. *Physica A*, **381**, 601 – 612.
- Fiedler, W., Bairlein, F., & Köppen, U. 2004. Using large-scale data from ringed birds from the investigation of effects of climate change on migrating birds: pitfalls and prospects. *Pages 49–67 of: Caswell, H. (ed), Advances in ecological research. Birds and climate change*, vol. 35. Elsevier, Amsterdam.
- Floerl, O., Inglis, G. J., Dey, K., & Smith, A. 2009. The importance of transport hubs in stepping-stone invasions. *Journal of Applied Ecology*, **46**, 37–45.
- Freeman, L. C. 1979. Centrality in social networks. Conceptual clarification. *Social Networks*, **1**(3), 215–239.
- Fukuda, A., Katsuji, M., Hirano, E., Suzuki, M., Higuchi, H., Morishita, E., Anderson, D. J., Waugh, S., & Philips, R. 2004. BGD-LII - a GPS data logger for birds. *Memoirs of National Institute of Polar Research*, **58**, 234–245.
- Gardner, W. A., Napolitano, A., & Paura, L. 2006. Cyclosationarity: Half a century of research. *Signal Processing*, **86**, 639–697.
- Gauthreaux, S. A. 1982. The ecology and evolution of avian migration systems. *Pages 93–167 of: Farner, D. S., & King, J. R. (eds), Avian Biology*. Academic Press, New York.

- Gee, H., Jiang, D.-Q., & Qian, M. 2006. A simple discrete model of Brownian motors: time-periodic Markov chains. *Journal of Statistical Physics*, **123**, 831–859.
- Gerkmann, B. 2007. *Nutzung von Telemetrie- und Satellitendaten zur Identifizierung wichtiger Habitats wandernder Vogelarten*. Ph.D. thesis, University of Bonn.
- Gerkmann, B., & Riede, K. 2005. Use of satellite telemetry and remote sensing data to identify important habitats of migratory birds (*Ciconia ciconia*). Pages 261–269 of: Huber, B. A., Sinclair, B. J., & Lampe, K.-H. (eds), *African Biodiversity: Molecules, Organisms, Ecosystems*. Springer Verlag, Bonn.
- GESAMP. 1997. Opportunistic settlers and the problem of the ctenophore *Mnemiopsis leidyi* invasion in the Black Sea. *GESAMP Report Studies*, **58**, 1–84.
- Gittenberger, E., Groenenberg, D. S. J., Kokshoorn, B., & Preece, R. C. 2006. Molecular trails from hitch-hiking snails. *Nature*, **439**, 409.
- Gollasch, S., Lenz, J., Dammer, M., & Andres, H.-G. 2000. Survival of tropical ballast water organisms during a cruise from the Indian Ocean to the North Sea. *Journal of Plankton Research*, **22**, 923–937.
- Gollasch, S., David, M., Voigt, M., Dragsund, E., Hewitt, C., & Fukuyo, Y. 2007. Critical review of the IMO international convention on the management of ships' ballast water and sediments. *Harmful Algae*, **4**, 585–600.
- González, M. C., Hidalgo, C. A., & Barabási, A.-L. 2008. Understanding individual human mobility patterns. *Nature*, **453**, 779–782.
- Gross, T., & Blasius, B. 2008. Adaptive coevolutionary networks: a review. *Journal of the Royal Society Interface*, **5**(20), 259–271.
- Gross, T., Dommar, C. J., & Blasius, B. 2006. Epidemic dynamics on an adaptive network. *Physical Review Letters*, **96**(20), 208701.
- Guimera, R., & Amaral, L. A. N. 2005a. Cartography of complex networks: modules and universal roles. *Journal of Statistical Mechanics and Theoretical Experiments*, P02001.
- Guimera, R., & Amaral, L. A. N. 2005b. Functional cartography of complex metabolic networks. *Nature*, **433**, 895–900.
- Guimera, R., Mossa, S., Turtleschi, A., & Amaral, L. A. N. 2005. The worldwide air transportation network - Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, **102**, 7794–7799.
- Gutenkunst, R., Newlands, N., Lutcavage, M., & Edelstein-Keshet, L. 2007. Inferring resource distributions from Atlantic bluefin tuna movements: An analysis based on net displacement and length of track. *Journal of Theoretical Biology*, **245**(2), 243–257.

10 Bibliography

- Gwinner, E., Schwabl, H., & Schwabl-Bensinger, I. 1988. Effects of food deprivation on migratory restlessness and diurnal activity in the garden warbler *Sylvia borin*. *Oecologia*, **77**, 321–326.
- Haynes, K. E., & Fotheringham, A. S. 1984. *Gravity and spatial interaction models*. Sage, Beverly Hills.
- Hedenström, A., & Ålerstam, T. 1997. Optimum fuel loads in migratory birds: distinguishing between time and energy minimization. *Journal of Theoretical Biology*, **189**, 227–234.
- Heng, D. 2007. Continuities and changes: Singapore as a port-city over 700 years. *Biblioasia*, **1**, 12–16.
- Hethcote, H. W. 2000. The mathematics of infectious diseases. *SIAM Review*, **42**(4), 599–653.
- Hill, R. D. 1994. Theory of geolocation by light levels. *Pages 227–236 of: Le Boeuf, B. J., & Laws, R. M. (eds), Elephant seals: population, ecology, behaviour and physiology*. University of California Press, Berkeley.
- Hopkins, G. A., & Forrest, B. M. 2008. Management options for vessel hull fouling. An overview of risks posed by in-water cleaning. *ICES Journal of Marine Science*, **65**, 811–815.
- Houston, A. I. 1998. Model of optimal avian migration: state, time and predation. *Journal of Avian Biology*, **29**, 395–404.
- Hu, Y., & Zhu, D. 2009. Empirical analysis of the worldwide maritime transportation network. *Physica A*, **388**, 2061–2071.
- Hubálek, Z. 2004. An annotated checklist of pathogenic microorganisms associated with migratory birds. *Journal Of Wildlife Diseases*, **40**, 639–659.
- Hufnagel, L., Brockmann, D., & Geisel, T. 2004. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences*, **101**, 15124–15129.
- Hughes, B. D. 1995. *Random walks and random environments*. Oxford University Press, New York.
- IMO. 2006. *International shipping and world trade. Facts and figures*. International Maritime Organization. (<http://www.imo.org>).
- Isard, W. 1954. Location Theory and Trade Theory: Short-Run Analysis. *Quarterly Journal of Economics*, **68**, 305–320.
- Javidpour, J., Sommer, U., & Shiganova, T. 2006. First record of *Mnemiopsis leidyi* A. Agassiz 1865 in the Baltic Sea. *Aquatic Invasions*, **1**, 299–302.
- Jenni, L., & Kéry, M. 2003. Timing of autumn bird migration under climate change: advances in long-distance migrants, delay in short-distance migrants. *Proceedings of the Royal Society of London B*, **270**, 1467–1471.

- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A.-L. 2000. The large-scale organization of metabolic networks. *Nature*, **407**, 651 – 654.
- Jordano, P., Bascompte, J., & Olesen, J. M. 2003. Invariant properties in coevolutionary networks of plant-animal interactions. *Ecology Letters*, **6**, 69 – 81.
- Jouventin, P., & Weimerskirch, H. 1990. Satellite tracking of wandering albatrosses. *Nature*, **343**(6260), 746–748.
- Kaatz, M. 1999. *Studie zur Neubewertung von ARGOS-Satellitenkoordinaten*.
- Kaatz, M. 2004. *Der Zug des Weißstorchs *Ciconia ciconia* auf der Europäischen Ostroute über den Nahen Osten nach Afrika*. Ph.D. thesis, University of Halle/Saale. (unpublished).
- Kaluza, P. 2007. *Evolutionary engineering of complex functional networks*. Ph.D. thesis, Technical University of Berlin.
- Kaluza, P., Kölzsch, A., Gastner, M. T., & Blasius, B. 2009. Regularity and randomness in the global network of cargo ship movements. (*submitted*).
- Kareiva, P. M., & Shigesada, N. 1983. Analyzing insect movement as a correlated random walk. *Oecologia*, **56**, 1432–1939.
- Kilpatrick, A. M., Chmura, A. A., Gibbons, D. W., Fleischer, R. C., Marra, P. P., & Daszak, P. 2006. Predicting the global spread of H5N1 avian influenza. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(51), 19368–19373.
- Kjellén, N., Hake, M., & Alerstam, T. 1997. Migration strategies of two Ospreys *Pandion haliaetus* between Sweden and tropical Africa revealed by satellite telemetry. *Journal of Avian Biology*, **28**, 15–23.
- Kolar, C. S., & Lodge, D. M. 2001. Progress in invasion biology: predicting invaders. *Trends in Ecology and Evolution*, **16**, 199–204.
- Komar, N. 2003. West Nile virus: epidemiology and ecology in North America. *Advances in Virus Research*, **61**, 185–234.
- Komin, N., Erdmann, U., & Schimansky-Geier, L. 2004. Random walk theory applied to *Daphnia* motion. *Fluctuation and Noise Letters*, **4**(1), L151–L159.
- Kruckenberger, H. 2002. *Patterns of spatial use of colour-marked Whitefronted geese *Anser albifrons* in Western and Middle Europe with contribution to social aspects*. Ph.D. thesis, University of Osnabrück.
- Kruckenberger, H., Müskens, G., & Ebbinge, B. S. 2007. Satellitentelemetrie von Blessgänsen *Anser albifrons albifrons* auf dem Frühjahrszug 2006 und 2007. *Vogelwarte*, **45**, 330–331. (in German).

- Kruckenberg, H., Kondratyev, A., Mooij, J., Zöckler, C., & Zaigudinova, E. 2008. White-fronted goose flyway population status. - Interim Report of a preliminary study in 2006. *Angewandte Feldbiologie*, **Bd. II**, 1–63.
- Kölzsch, A., & Blasius, B. 2008. Theoretical approaches to bird migration - The white stork as a case study. *European Physical Journal Special Topics*, **157**, 191–208.
- Kölzsch, A., Saether, S. A., Gustafsson, H., Fiske, P., Höglund, J., & Kalas, J. A. 2007. Population fluctuations and regulation in great snipe: a time-series analysis. *Journal Of Animal Ecology*, **76**(4), 740–749.
- Köppen, U., & Scheil, S. 2004. Bericht der Beringungszentrale Hiddensee für die Jahre 2001 und 2002. *Apus*, **12**, 5–36. (in German).
- Latora, V., & Marchiori, M. 2001. Efficient behaviour of small-world networks. *Physical Review Letters*, **87**, 198701.
- Latora, V., & Marchiori, M. 2002. Is the Boston subway a small-world network? *Physica A*, **314**(1-4), 109–113.
- Lee, K. 2001. The global dimensions of cholera. *Global change and human health*, **2**, 6–17.
- Leicht, E. A., & Newman, M. E. J. 2008. Community structure in directed networks. *Physical Review Letters*, **100**(11), 118703.
- Leick, A. 1995. *GPS Satellite Surveying*. Wiley, New York.
- Leshem, Y., & YomTov, Y. 1996. The magnitude and timing of migration by soaring raptors, pelicans and storks over Israel. *Ibis*, **138**(2), 188–203.
- Lodge, D. M. 1993. Biological invasions: lessons for ecology. *Trends in Ecology and Evolution*, **8**, 133–137.
- Logofet, D. O., & Lesnaya, E. V. 2000. The mathematics of Markov models: what Markov chains can really predict in forest succession. *Ecological Modelling*, **126**, 285–298.
- Lounibos, L. P. 2002. Invasions by insect vectors of human disease. *Annual Review of Entomology*, **47**, 233–266.
- Lämmer, S., Kori, H., Peters, K., & Helbing, D. 2006. Decentralised control of material or traffic flows in networks using phase-synchronisation. *Physica A*, **363**, 39–47.
- Mack, R. N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M., & Bazzaz, F. A. 2000. Biotic invasions - causes, epidemiology, global consequences, and control. *Ecological Applications*, **10**, 689–710.
- Marell, A., Ball, J. P., & Hofgaard, A. 2002. Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Levy flights. *Canadian Journal of Zoology-Revue Canadienne de Zoologie*, **80**(5), 854–865.

- Meinesz, A., de Vaugelas, J., Hesse, B., & Mari, X. 1993. Spread of the introduced tropical green alga *Caulerpa taxifolia* in the northern Mediterranean waters. *Journal of Applied Phycology*, **5**, 141–147.
- Miguéns, J. I. L., & Mendes, J. F. F. 2008. Weighted and directed network on traveling patterns. *Lecture Notes in Computer Science*, **5151**, 145–154.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. 2002. Network motifs: Simple building blocks of complex networks. *Science*, **298**(5594), 824–827.
- Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., & Alon, U. 2004. Superfamilies of evolved and designed networks. *Science*, **303**(5663), 1538–1542.
- Mitchell, J. S. B. 1991. A new algorithm for shortest paths among obstacles in the plane. *Annals of Mathematics and Artificial Intelligence*, **3**, 83–106.
- Mooij, J. H., Faragó, S., & Kirby, J. S. 1999. White-fronted Goose *Anser albifrons albifrons*. Pages 94–129 of: Madsen, J. G., Cracknell, G., & Fox, A. D. (eds), *Goose Population of the Western Palearctic*. Wetlands International Publications, vol. 48.
- Morales, J. M., Haydon, D. T., Frair, J., Holsiner, K. E., & Fryxell, J. M. 2004. Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology*, **85**(9), 2436–2445.
- Mouritsen, H. 1998. Modelling migration: the clock-and-compass model can explain the distribution of ringing recoveries. *Animal Behaviour*, **56**, 899–907.
- Nathan, R. 2006. Long-distance dispersal of plants. *Science*, **313**, 786–788.
- Newman, M. E. J. 2002. Assortative mixing in networks. *Physical Review Letters*, **89**, 208701.
- Newman, M. E. J. 2003a. Mixing patterns in networks.
- Newman, M. E. J. 2003b. The structure and function of complex networks. *SIAM Review*, **45**, 167–256.
- Newman, M. E. J. 2004. Who is the best connected scientist? A study of scientific coauthorship networks. Pages 337–370 of: Ben-Naim, E., Frauenfelder, H., & Toroczkai, Z. (eds), *Complex networks*. Springer, Berlin.
- Newman, M. E. J. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, **46**(5), 323–351.
- Newman, M. E. J. 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 8577–8582.
- Newman, M. E. J. 2007. The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, 1–12.

- Nolet, B. A., & Mooij, W. M. 2002. Search paths of swans foraging on spatially autocorrelated tubers. *Journal of Animal Ecology*, **71**(3), 451–462.
- Norris, J. R. 1997. *Markov Chains*. Cambridge University Press, Cambridge.
- Olsen, B., Munster, V. J., Wallensten, A., Waldenström, J., Osterhaus, A. D. M. E., & Fouchier, R. A. M. 2006. Global patterns of Influenza A virus in wild birds. *Science*, **312**, 384–388.
- Pasternak, Z., Blasius, B., & Abelson, A. 2004. Host location by larvae of a parasitic barnacle: larval chemotaxis and plume tracking in flow. *Journal of Plankton Research*, **26**(4), 487–493.
- Pastor-Satorras, R., & Vespignani, A. 2001. Epidemic spreading in scale-free networks. *Physical Review Letters*, **86**(14), 3200–3203.
- Pastor-Satorras, R., Vazquez, A., & Vespignani, A. 2001. Dynamical and correlation properties of the internet. *Physical Review Letters*, **87**(25), 258701.
- Patoucheas, D. P., & Stamou, G. 1993. Non homogeneous Markovian models in ecological modelling: A study of zoobenthos dynamics in Thermaikos Gulf, Greece. *Ecological Modelling*, **66**, 197–215.
- Peterson, A. T., Vieglais, D. A., & Andreasen, J. K. 2003. Migratory birds modeled as critical transport agents for West Nile virus in North America. *Vector-Borne and Zoonotic Diseases*, **3**, 27–37.
- Pimentel, D., Zuniga, R., & Morrison, D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, **52**(3), 273–288.
- Priede, I. G., & Swift, S. M. (eds). 1992. *Wildlife telemetry: Remote monitoring and tracking of animals*. Ellis Horwood Ltd, New York.
- Puth, L. M., & Post, D. M. 2005. Studying invasion: have we missed the boat? *Ecology Letters*, **8**, 715–721.
- Ramos-Fernandez, G., Mateos, J. L., Miramontes, O., Cocho, G., Larralde, H., & Ayala-Orozco, B. 2004. Levy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*). *Behavioral Ecology and Sociobiology*, **55**(3), 223–230.
- Redner, S. 2008. Networks: teasing out the missing links. *Nature*, **453**, 47–48.
- Reilly, W. J. 1931. *The law of retail gravitation*. Reilly, New York.
- Rodrigue, J.-P., Comtois, C., & Slack, B. 2006. *The geography of transport systems*. Routledge, London.
- Ruiz, G. M., Carlton, J. T., Grosholz, E. D., & Hines, A. H. 1997. Global invasions of marine and estuarine habitats by non-indigenous species: mechanisms, extent, and consequences. *American Zoologist*, **37**, 621–632.

- Ruiz, G. M., Rawlings, T. K., Dobbs, F. C., Drake, L. A., Mullady, T., Huq, A., & Colwell, R. R. 2000. Global spread of microorganisms by ships. *Nature*, **408**, 49–50.
- Sala, O. E., Chapin, F. S. III., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L. F., Jackson, R. B., Kinzig, A., Leemans, R., Lodge, D. M., Mooney, H. A., Oesterheld, M., Poff, N. L., Sykes, M. T., Walker, B. H., Walker, M., & Wall, D. H. 2000. Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.
- Schüz, E., Berthold, P., Gwinner, E., & Oelke, H. 1971. *Grundriss der Vogelzugkunde*. Parey, Berlin. (in German).
- Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G., & Manna, S. S. 2003. Small-world properties of the Indian railway network. *Physical Review E*, **67**(3), 036106.
- Shlesinger, M. F., Klafter, J., & Wong, Y. M. 1982. Random walks with infinite spatial and temporal moments. *Journal of Statistical Physics*, **27**, 499–512.
- Sims, D. W., Southall, E. J., Humphries, N. E., Hays, G. C., Bradshaw, C. J. A., Pitchford, J. W., James, A., Ahmed, M. Z., Brierley, A. S., Hindell, M. A., Morritt, D., Musyl, M. K., Righton, D., Shepard, E. L. C., Wearmouth, V. J., Wilson, R. P., Witt, M. J., & Metcalfe, J. D. 2008. Scaling laws of marine predator search behaviour. *Nature*, **451**, 1098–1103.
- Smith, R. D. 2006. Responding to global infectious disease outbreaks: Lessons from SARS on the role of risk perception, communication and management. *Journal of Social Science and Medicine*, **63**, 3113–3123.
- Sokal, R. R., & Rohlf, F. J. 1995. *Biometry: The principles and practice of statistics in biological research*. Freeman and Company, New York.
- Stauffer, D. 1994. *Introduction to Percolation Theory*. CRC Press, London.
- Tatem, A. J., Hay, S. I., & Rogers, D. J. 2006. Global traffic and disease vector dispersal. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(16), 6242–6247.
- Taylor, G. I. 1922. Diffusion by continuous movements. *Proceedings of the London Mathematical Society*, **20**, 196–212.
- Thorup, K., Alerstam, T., Hake, M., & Kjellen, N. 2003. Can vector summation describe the orientation system of juvenile ospreys and honey buzzards? - An analysis of ring recoveries and satellite tracking. *Oikos*, **103**(2), 350–359.
- Tilman, D., & Kareiva, P. 1997. *Spatial Ecology*. Princeton University Press, Princeton.
- Turchin, P. 1998. *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Sinauer Associates, Sunderland, Massachusetts.
- UN. 2007. *Review of maritime transport*. United Nations conference on trade and development. (http://www.unctad.org/en/docs/rmt2007_en.pdf).

10 Bibliography

- Urban, D. L., Minor, E. S., Treml, E. A., & Schick, R. S. 2009. Graph models of habitat mosaics. *Ecology Letters*, **12**, 260–273.
- Usher, M. B. 1979. Markovian approaches to ecological succession. *Journal of Animal Ecology*, **48**, 413–426.
- van den Bossche, W., Berthold, P., Kaatz, M., Nowak, E., & Querner, U. 2002. Eastern European white stork populations: Migration studies and elaboration of conservation measures. *BfN-Skripten*, **66**, 1–197.
- Vassiliou, P.-C. G. 1998. The evolution of the theory of non-homogeneous Markov systems. *Applied Stochastic Models and Data Analysis*, **13**, 159–176.
- Vermeij, G. J. 1996. An agenda for invasion biology. *Biological Conservation*, **78**, 3–9.
- Vinogradov, M. E., Shushkina, E. A., Musayeva, E. I., & Sorokin, P. Y. 1989. A new exotic species in the Black Sea: the ctenophore *Mnemiopsis leidyi* (Ctenophora: Lobata). *Oceanology*, **29**, 220–224.
- Viswanathan, G. M., Afanasyev, V., Buldyrev, S. V., Murphy, E. J., Prince, P. A., & Stanley, H. E. 1996. Lévy flight search patterns of wandering albatrosses. *Nature*, **381**, 413–415.
- Viswanathan, G. M., Buldyrev, S. V., Havlin, S., da Luz, M. G. E., Raposo, E. P., & Stanley, H. E. 1999. Optimizing the success of random searches. *Nature*, **401**(6756), 911–914.
- Viswanathan, G. M., Raposo, E. P., & da Luz, M. G. E. 2008. Lévy flights and superdiffusion in the context of biological encounters and random searches. *Physics of Life Reviews*, **5**, 133–150.
- Walther, G. R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J. M., Hoegh-Guldberg, O., & Bairlein, F. 2002. Ecological responses to recent climate change. *Nature*, **416**(6879), 389–395.
- Warton, D. I., Wright, I. J., Falster, D. S., & Westoby, M. 2006. Bivariate line-fitting methods for allometry. *Biological Reviews*, **81**(2), 259–291.
- Wassermann, S., & Faust, K. 1994. *Social Network Analysis*. Cambridge University Press, Cambridge.
- Watts, D. J. 1999. Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, **105**, 493–592.
- Watts, D. J., & Strogatz, S. H. 1998. Collective dynamics of small-world networks. *Nature*, **393**, 440–442.
- Wei, T., Deng, G., & Wu, P. 2007. Analysis of network effect in port and shipping system characterized by scale-free network. In: *Proceedings of the 2007 International Conference on Intelligent System and Knowledge Engineering*. Atlantic Press, Amsterdam.

- Weimerskirch, H., Bonadonna, F., Bailleul, F., Mabile, G., Dell'Omo, G., & Lipp, H. P. 2002. GPS tracking of foraging albatrosses. *Science*, **295**(5558), 1259–1259.
- Wikelski, M., Kays, R. W., Kasdin, N. J., Thorup, K., Smith, J. A., & Swenson, G. W. 2007. Going wild: what a global small-animal tracking system could do for experimental biologists. *Journal Of Experimental Biology*, **210**(2), 181–186.
- Williamson, M. H. 1996. *Biological Invasions*. Chapman and Hall, London.
- Williamson, M. H., & Fitter, A. 1996. The characters of successful invaders. *Biological Conservation*, **78**, 163–170.
- Wilson, R. P. 2001. Beyond rings on birds for determination of movements: Wither the archival tag? *Ardea*, **89**(1), 231–240.
- Wiltschko, R., & Wiltschko, W. 1995. *Magnetic Orientation in Animals*. Springer, Berlin.
- Wiltschko, R., & Wiltschko, W. 2003. Avian navigation: from historical to modern concepts. *Animal Behaviour*, **65**, 257–272.
- Young, A., & Almond, G. 1961. Predicting distributions of staff. *The Computer Journal*, **3**, 246–250.
- Zachcial, M., & Heideloff, C. 2001. *ISL shipping statistics yearbook 2001*. Institute of Shipping Economics and Logistics, Bremen.
- Zink, G., & Bairlein, F. 1995. *Der Zug europäischer Singvögel. Ein Atlas der Wiederfunde beringter Vögel*. Aula, Wiesbaden. (in German).
- Zipf, G. K. 1946. The P1 P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review*, **11**, 677–686.

10 Bibliography

11 Acknowledgements

During my work on biological transportation networks a large social network developed around me, of people to guide and support me.

First of all many thanks to my supervisor Bernd Blasius, who gave me the opportunity to work in his interdisciplinary group of theoretical ecologists, always encouraged me to follow own ideas, explained new ones and directed me through the world of scientific research.

As a theoretical ecologist one usually depends on empirical researchers to provide data for model development and evaluation. I am very grateful to Helmut Kruckenberg, Michael Kaatz, Ulrich Köppen, Wolfgang Fiedler and Franz Bairlein for providing me with ring-recapture, satellite telemetry and GPS data of several bird species. Thank you also to the large crowd of diligent bird watchers and expert bird catchers without the help of who the large data sets would not exist. Furthermore, I would like to thank Alex Wilkinson from Lloyd's Register Fairplay for the support in obtaining access to their unique data base of ship movement. I am also grateful to Avigdor Abelson, Eva Strothotte and Harald Rosenthal for data and helpful discussion about marine bioinvasion.

Sensible working is only possible in a supportive environment, which I was very happy to experience in the group of Theoretical Ecology in Potsdam. Many thanks to Nina for the pleasant company in our office and help with intractable computers, Ralf for endless discussions about statistical physics and Lars for fruitful exchanges about the great software R (Ihaka & Gentleman, 1996) and an inexhaustible, encouraging cheerfulness. Furthermore, I want to thank Carlos and Stefan for one or the other happy discussion. Last but not least, I am very grateful to Thilo for expertise advice and motivation to pursue theoretical research. The time in Potsdam was one of the greatest I have had, thanks to those six people that I not only shared working places with.

After moving to Oldenburg our new group of Mathematical Modelling was forming slowly. I am very grateful to Cora for company in the beginning and always having time to help with maths and computers. Special thanks to the group of Theoretical Physics and Complex Systems for innumerable cheerful coffee breaks and helpful discussions about physics and mathematics. Later, our group expanded and with Pablo, Hanno and Michael there were three postdocs supporting me with work on ships and bioinvasion, thank you. Furthermore, I would like to acknowledge Katharina, Daniel, Levin and Johannes. For providing the link back to applied ecology I am grateful to the group of Landscape Ecology. Thanks to Martin and Julia for helpful discussions about ornithology and GIS. However, my two years in Oldenburg would not have been the same if it was not for Mira who always encouraged and built me up, when my work would not proceed as desired.

On purpose I did not list Alex in the two paragraphs above. He started working with Bernd and arrived in Potsdam the same day as I did and we shared the office ever since, in Potsdam as well as Oldenburg. Thank you so much for being one of the most uncomplicated and helpful

11 Acknowledgements

office mates and colleagues. More than that I happily recall times of cooking and socialising with your family.

Special thanks to Elke, Gudrun and Elisa in Oldenburg and Birgit in Potsdam for support with administrative matters. I am also very grateful to our system administrators Nina and Tessi in Potsdam and Klemens in Oldenburg. Thank you for being so patient with me.

During my Phd-studies I had the possibility to attend many conferences and workshops. Two of them I recall as especially encouraging, the Lem II workshop in the Ukraine and the Yomo workshop in Gülpe. Thanks to all people around there.

Several people proofread and helped me writing this thesis, for this I am very grateful to first of all Mira, but also Bernd, Jana, Hanno, Ralf, Lars, Nina, Ulf and my mother.

I would like to give my special thanks to two very good friends, Gerlind and Janusch that constantly motivated me during the work on this thesis and never got tired of listening to one or the other worry.

During the final year of work on this thesis Ulf has become a very important person to me. Thank you so much for holding my head up, bearing with my strange moods and keeping me the greatest company.

My family strongly supported and encouraged me during my years of Phd-work, for this I am especially grateful to my Mama, Papa, Sebastian, Oma, Opa and Oma.

12 Curriculum vitae

Personal details

Name	Andrea Kölzsch
Born	15 July 1980 in Wismar, Germany
Nationality	German
Contact details	Carl-von-Ossietzky-Str. 5, 26129 Oldenburg andrea_jule@web.de

Education

April 2007 - present	Carl von Ossietzky Universität Oldenburg, Germany Institute for Chemistry and Biology of the Marine Environment (ICBM) PhD programme in Mathematical Modelling (supervised by Prof. Dr. B. Blasius) Topic: <i>Modelling of Migration Movement and Bioinvasion on Networks</i>
January 2006 - March 2007	Universität Potsdam, Germany Institute of Physics, Nonlinear Dynamics Group PhD programme in Theoretical Ecology (supervised by Prof. Dr. B. Blasius) Topic: <i>Modelling of Animal Movement and Bioinvasion on Networks</i>
December 2005	Diplom (MSc) in Biomathematics, Ernst Moritz Arndt Universität Greifswald, Germany Thesis: <i>Population Dynamics and Statistical Analysis of longtime mark-recapture data in Great Snipe (Gallinago media)</i> Grade: 1.2 ("very good")
February 2003 - June 2003	Aberdeen University, Scotland Courses within zoology, Level 4
August 2002 - January 2003	Uppsala University, Sweden Courses withing mathematics and biology

12 Curriculum vitae

- | | |
|-----------|---|
| July 2000 | Abitur (A-level), Große Stadtschule - Geschwister Scholl
Gymnasium, Wismar, Germany
Grade: 1.0 ("very good") |
| June 1998 | US High School Diploma , LaSalle High School, Olla,
Louisiana
GPA: 3.73 ("very good") |

Practical Experience

- | | |
|-----------------------------------|--|
| June 2008, May 2006,
June 2005 | Norwegian Institute for Nature Research (NINA), Trondheim,
Norway
Field work on great snipe |
| February 2004 | Institute for Zoology, Ernst Moritz Arndt Universität Greifswald,
Germany
Analysis of biological data: <i>Differentiating three ant species by
morphometrical data</i> |
| August 2003 | Field Station Hiddensee, Ernst Moritz Arndt Universität Greifswald,
Germany
Ecology Field Course |
| June 2003 | Cromarty Field Station, Aberdeen University, Scotland
Biodiversity Field Course |

Teaching Experience

- Lecturer and tutor for 1-day Master course *Statistical Analysis of spatial data*, Carl von Ossietzky Universität Oldenburg, 2008
- Supervisor of a Bachelor student for 9 week traineeships in *Mathematical modelling*, Carl von Ossietzky Universität Oldenburg, 2008
- Tutor for *Mathematical Modelling I & II*, Carl von Ossietzky Universität Oldenburg, 2007/2008
- Tutor for *Statistics I & II*, Ernst Moritz Arndt Universität Greifswald, 2004/2005

Conferences and Presentations

- Attended the
141. Jahresversammlung der DO-G, Bremen, Germany, Oktober 2008
- Talk: *Bird migration as seasonally driven movement on a network*
Workshop "Pattern formation in biological networks", Berlin, Germany, September 2008
- Talk: *Bird migration as seasonally driven movement on a network*
EURECO - GFOE 2008, "Biodiversity in an Ecosystem Context", Leipzig, Germany,
September 2008

- Poster: *Bird migration as seasonally driven movement on a network*
European Conference on Mathematical and Theoretical Biology, Edinburgh, Scotland, June-July 2008
- Talk: *Modelling of migration movement and bioinvasion on networks*
Discussion Group: *Animal dispersal and Lévy distributions*
Third YoMo Workshop, Gülpe, Germany, April 2008
- Talk: *A network model of bird migration.*
The 6th European Conference on Ecological Modelling, Trieste, Italy, November 2007
- Talk: *Bird migration from a theoretical point of view: Movement patterns, data analysis and a network model.*
Seminar at the Max Planck Institute for Ornithology, Vogelwarte Radolfzell, Radolfzell, Germany, August 2007
- Talk: *The diversity-invasibility hypothesis - models, field studies and the importance of scale.*
Workshop "Stanislaw Lem II. Theories of species richness - an entangled route to Santa Rosalia", Theoretical Ecology Potsdam/Oldenburg, Yasinya, Ukraine, February 2007
- Talk: *Modelling bird migration: stochastic movement on a network with seasonally driven jumping rates.*
Workshop "Active Motion and Swarming", Sfb555, Berlin, Germany, December 2006
- Attended the
4th European Conference of the working group NEOBIOTA on Biological Invasions, Vienna, Austria, September 2006
- Attended the
International Conference on Macroecological Tools for Global Change Research, Virtual Institute Macroecology, Potsdam, Germany, August 2006
- Talk: *Population Dynamics Analysis in Great Snipe (Gallinago media).*
Seminar at the Mathematical Statistics Group, Uppsala University, Sweden, November 2005

Publications

- Diplom Thesis: Andrea Kölzsch *Population Dynamics and Statistical Analysis of long-time mark-recapture data in Great Snipe (Gallinago media)*, Greifswald (2005).
- Andrea Kölzsch, Stein Are Sæther, Henrik Gustafsson, Peder Fiske, Jacob Höglund and John Atle Kålås. Population fluctuations and regulation in great snipe: a time-series analysis. *Journal of Animal Ecology* **76**: 740-749 (2007).

12 Curriculum vitae

- Andrea Kölzsch and Bernd Blasius. Theoretical approaches to bird migration: the white stork as a case study. *The European Physical Journal Special Topics* **157**: 191-208 (2008).
- Andrea Kölzsch, Ulrich Köppen, Wolfgang Fiedler, Franz Bairlein and Bernd Blasius. Lévy flights in bird movement after all? (submitted).
- Pablo Kaluza, Andrea Kölzsch, Michael T. Gastner and Bernd Blasius. Regularity and randomness in the global network of cargo ship movement. (submitted).

13 Personal contributions

In the following, the contributions of Andrea Kölzsch (AK) and the coauthors of the different papers are listed.

Paper I.

Regularity and randomness in the global network of cargo ship movements

PK and AK developed the network from real movement trajectory data. PK implemented the network derivation, examined single ship trajectories, developed the regularity index and analyzed the network's motif distribution. AK calculated the scaling properties of the GCSN and its subnetworks, analyzed betweenness centralities and determined the community structures. MTG developed the graphical presentation of the ship network and conducted the gravity model analyses. BB coordinated the work. Writing was a joint task of all coauthors.

Paper II.

Indications of marine bioinvasion from network theory

AK developed this study and performed the network analyses. BB suggested the R_0 centrality, provided critical advice for the percolation analysis and discussion of the results. The writing was done by AK.

Paper III.

Theoretical approaches to bird migration

AK conducted the literature compilation of bird migration data, communicated with the ringing centres and analysed the obtained data. BB provided advice on random walk theory and the interpretation of the obtained results. AK and BB jointly developed the migration network model, which was implemented and parameterised by AK. The writing was done by AK with revisions by BB.

Paper IV.

Lévy flights in bird movement after all?

This study was initiated by BB after discussion of preliminary results of data analysis by AK. The ring-recapture data were provided by UK, WF and FB acting in the interest of the three German ringing centres. UK and FB advised the selection of species. The data were analysed by AK and she conducted the statistical tests. AK wrote the paper with revisions by BB.

Paper V.

A periodic Markov model of bird migration on a network

AK and BB jointly developed the migration network model. It was implemented, parameterised with data and characterised by AK. HK compiled the greater white-fronted goose data, and MK provided data of the white stork and a quality selection algorithm. The writing was done by AK, with revisions by BB (introduction, model) and HK (geese ecology).

Erklärung

gemäß § 10 Abs. 2 der Promotionsordnung der Fakultät für Mathematik und Naturwissenschaften der Carl von Ossietzky Universität Oldenburg vom 11.12.2003 und den Änderung der Promotionsordnung der Fakultät für Mathematik und Naturwissenschaften vom 16.12.2005.

Hiermit erkläre ich, dass ich eine Promotion zum Dr. rer. nat. anstrebe. Es soll der Grad eines Doktors verliehen werden.

Des Weiteren erkläre ich, die vorliegende Arbeit in allen Teilen selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt zu haben.

Außerdem erkläre ich, dass die Dissertation weder in ihrer Gesamtheit noch in Teilen einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorliegt oder vorgelegen hat.

Paper III dieser Arbeit wurde bereits veröffentlicht:

Andrea Kölzsch and Bernd Blasius. Theoretical approaches to bird migration: the white stork as a case study. *The European Physical Journal Special Topics* **157**: 191-208 (2008).

Oldenburg, 14.09.2009

Andrea Kölzsch