

# The cluster NDNS<sup>+</sup>

## AIMS & SCOPES

### 1 Nonlinear Dynamics of Natural Systems

#### 1.1 Research plan

Processes in the natural environment or in living organisms differ intrinsically from those studied in the ‘hard sciences’. Classically, a physical or chemical process is reduced into one or more idealized building blocks which are subsequently studied in isolation. Modern developments in earth and life sciences have exposed the limitations of this approach. Collective cell behavior or the behavior of a neuro-physiological or a geophysical system is inherently governed by the simultaneous interaction and competition between various mechanisms. Since the coupling between these mechanisms is typically nonlinear, the system cannot be understood as a combination of one or more of its subsystems.

The mathematics of nonlinear systems is firmly rooted in the classical hard sciences, especially since the reduced models in, for example, hydrodynamics and classical mechanics are highly nonlinear. However, at present, the challenges posed by processes in the brain, in the cell, or in the atmosphere, to the theory of dynamical systems surpass those coming from more classical fields in richness, in magnitude and in difficulty. This leads to a deep cross-fertilization, because the mathematics of nonlinear systems itself is at a critical phase in its development as well. Especially in idealized settings, many fundamental aspects of the theory are well-understood. It has only very recently become possible to delve beneath the surface of more complex phenomena by combining these aspects and subsequently extending these into new theoretical insights. Hence, mathematics is ready, and well-prepared, to meet the challenges posed by, for instance, biological networks and atmospherically driven oceanic convection. Since the interaction between mathematics and the earth and life sciences has traditionally not been very intense, the mathematical approach has not yet been exploited to its full strength and depth within this context.

#### 1.2 Nonlinear life sciences

##### 1.2.1 The challenge

The process of evolution has generated an enormous diversity of behavioral and physiological interaction in biological systems. Given the known genetic information in a DNA sequence, what processes produce an observed function in a macro-molecule? How are macro-molecules organized and controlled to make a living cell? How are cells organized at higher levels such as tissues and organisms? These are difficult questions to answer. Cells never form thermodynamically equilibrated structures and can have a complex feedback interaction with their environment. Cells can, for example, modify their surroundings by secreting chemicals. On the other hand, the cell environment can cause changes in cell properties, for example, by changing the levels of gene expression within a cell. These (dynamical) systems with feedback are very common in biology [17]. Mathematically, it is known that feedback decreases the uncertainty due to noise and other disturbances

On a totally different scale, the emerging global health care problems associated with the spread of infectious diseases and the antibiotic resistance of micro-organisms, induce similar fundamental questions. This is a dynamical process, driven by the spread of micro-organisms and resistance genes, and often fueled by the very use of antibiotics. Data on prevalence and incidence of resistant micro-organisms are abundant. But how to base sound conclusions, e.g. concerning the effectiveness of various potential control measures, on these data?

A true predictive understanding of these questions requires the discovery of simple mechanisms that are not species-specific. Multidisciplinary research teams from mathematics, computer science, physics,

chemistry and biology are needed in order to reveal such basic mechanisms.

### 1.2.2 Research themes

Many models of biological phenomena use partial differential equations to combine elements of random diffusive motion with biologically motivated rules that generate more ordered motion. These models, however, do typically not include nonlocal interactions, nor do they take into account the discrete nature of cells and their non-trivial geometry and orientation. To model biological phenomena more realistically one needs to incorporate nonlocal terms, for example, using structured populations models [34], integro-partial differential equations [35] or cellular automata [14].

Cellular automata consist of discrete agents, which occupy sites of a lattice. The agents can have internal state variables and a set of rules describing the evolution of their state and position. The internal state variables and the rules can be either discrete or continuous, deterministic or probabilistic. At one extreme, the rules may approximate partial differential equations, at the other end, the rules may resemble discrete logical interactions of Boolean computers or interacting ensembles of particles [33].

Lattice differential equations form an intermediate class. Here, the internal state variables are continuous functions of time. Such differential equations modeled on a lattice play an important role in modeling a variety of applications [26]. The study of patterns and synchronization in lattice equations naturally leads to systems of functional differential equations of mixed type. The theory for such equations is not yet completely developed, but the development of an effective perturbation theory and of numerical methods is well underway [32]. The ability to incorporate nonlocal interactions into a model together with the relatively rich solution structure of lattice differential equations, presents a strong motivation for studying such systems (see [43] and the references given there).

The interaction between continuum and discrete models is fundamental and should be exploited. For example, in the case of embryonic segmentation of *Drosophila melanogaster*, it is shown that a simpler Boolean network model [1] may approximate a fairly complex continuum model while retaining the essential characteristics.

The challenging research problem is to establish firm mathematical relationships between at one hand the component and interaction network structure and, at the other hand, the extent to which the internal dynamics of a component or the dynamics of a part of the interaction system may be approximated by a simpler model, while essentially preserving the global behavior of the full system.

The development of efficient algorithms, such as Monte Carlo simulations, makes it possible to simulate large coupled dynamical systems. In turn, these simulations lead to fundamental questions concerning the qualitative behavior of large coupled dynamical systems, such as the stability of configurations, interaction between time scales, pattern formation and traveling waves, which can only be solved in close interaction with numerical analysis and probability theory.

There is a large body of mathematical expertise that can be exploited to attack these questions, for example, regularization by randomness, methods from perturbation theory, inertial manifolds, and the theory of dissipative systems (see [44] and the references given there).

## 1.3 Nonlinear earth sciences

### 1.3.1 The challenge

Our climate is a nonlinear, highly complex, dynamical system. It is driven by the interaction of fast processes in the atmosphere (the weather) and slow processes in the oceans. The climate is not stationary or ‘stable’, it evolves. Although its dynamics is very slow, the climate may also undergo ‘sudden’ changes on relatively short time scales, as is known from various kinds of data. This behavior is typical for environmental processes. Of course, the impact on society of any kind of structural change in the climate is enormous. On a smaller scale, the same is true for most processes in the earth sciences. For instance, the evolution of sandbanks in rivers and the coastal zone has direct economical and environmental consequences. The Dutch government spends large sums of money and energy on trying to control the changes, which requires the understanding of the underlying mechanisms.

A typical model in the earth sciences has a core that is based on well-known laws of physics. However, almost without exception, the key mechanisms involved have different spatial and temporal scales. As a consequence, the full model, in general, includes a large number of equations. Moreover, one often has to rely on empirically motivated assumptions, especially to model the interactions between the different mechanisms, or to reduce the number of unknowns and thus to simplify the model.

For instance, the evolution of sand banks is for a large part driven by hydrodynamics. However, the mechanism by which sand is transported through the water is highly complicated, and not fully understood. The transport takes place at micro-scales compared to the scales of the objects of study, the sand banks. Furthermore, the natural time scale at which sand banks travel and grow (years) is much larger than that of the driving fluid flow (hours/minutes).

Although the models of earth sciences are in general composed of many strongly coupled highly nonlinear partial differential equations that act on various spatial and temporal scales, the dynamics generated by these models are often surprisingly simple, especially on finite, but long, intervals of time. Moreover, the complex phenomena exhibited by these models can most often be seen as the products of interaction between relatively simple ‘building blocks’. In the terminology of dynamical systems, this complex behavior might thus be seen as a bifurcation scenario that is initiated by the destabilization of a fundamental ‘pattern’.

The most urgent challenge posed by the earth sciences to the field of nonlinear systems is to develop methods by which this low-dimensional behavior of the infinite-dimensional complex multi-scale models can be derived, understood and predicted.

### 1.3.2 Research themes

There is at present a large body of mathematical knowledge on which the response to this challenge can be based. The concept of a low-dimensional attractor embedded in a finite-dimensional attracting invariant manifold of an infinite dimensional flow lies at the core of the dynamical systems approach to partial differential equations [40, 44]. Bifurcations in partial differential equations can often be reduced to a finite-dimensional setting by methods that have been introduced in the context of low-dimensional systems [27]. In partial differential equations destabilizations might be induced by essential spectrum and can therefore not be directly studied by the methods of low-dimensional systems. Nevertheless, there is a mathematical approach by which these bifurcations can be successfully studied [11]. The mathematical backbones of this approach, the weakly nonlinear stability theory or the theory of modulation equations, has only very recently been established rigorously [15]. In this situation, the model cannot be reduced directly to a finite-dimensional system, however, near the bifurcation, the dynamics are captured by a ‘normalized’ nonlinear modulation equation of partial differential type. Studying the full model has thus been reduced to studying this modulation equation. The Ginzburg-Landau equation is the most well-known example of such modulation equations and, although it generates highly non-trivial dynamics that are certainly not fully understood, there is a fundamental mathematical understanding of various aspects of the dynamics [15].

In realistic models, the system cannot be assumed to be (mathematically) close to a bifurcation, i.e., it is in general ‘far from equilibrium’. The mathematical ‘tool-box’ includes various methods by which these far-from-equilibrium-systems can be studied. Various of these methods exploit the multiple-scale structure of the model. The theories of (singular) perturbations, averaging, (re)normalization, etc. are well-established ‘tools’ that have been applied successfully in infinite-dimensional settings to deal with strongly separated spatial and/or temporal scales [15, 36]. In fact, they are partly developed in this context. Especially in the earth sciences, the truncation techniques of projecting an infinite dimensional system on a finite set of ‘high energy modes’ are widely applied. Of course this is partly due to the seminal work by Lorenz [30], which can be seen as the benchmark example of how the fields of earth science and nonlinear mathematics inspire and challenge each other. Since the work of Lorenz, these techniques have evolved into the sophisticated methods of proper and/or empirical orthogonal decompositions in which the statistical characteristics of the model play a central role [21]. Nevertheless, the low-dimensional truncations produced by these techniques may still include up to 50 – 100 components, which is beyond the realm of ‘standard’ finite-dimensional dynamical systems theory. Thus, although truncated, these models inspire and challenge the mathematical theory [6].

## 1.4 Coherence

### 1.4.1 Earth and live sciences from a dynamical viewpoint

First of all the unifying feature of this cluster lies in the theory of dynamical systems and the associated qualitative methods and qualitative tools. But there is more. From the viewpoint of dynamical systems, the study of a nonlinear process in the life or in the earth sciences is remarkably similar. A mathematician who intends to contribute to the life or the earth sciences necessarily has to be in close contact with

specialists in, for instance, cell systems or oceanography. Moreover, he or she has to be actively involved in the modeling process. This involvement is essential for such a joint research project to be successful. The derivation of reduced models, i.e., models that are suitable for further study, is an intrinsic part of the modeling process. It may incorporate mathematical techniques and/or empirical arguments. Most likely, the (reduced) model is obtained by a creative mixture of various ingredients and approaches.

The resulting models, irrespective of whether they describe biotic or a-biotic natural phenomena, are almost without exception too complex to be studied in full detail. Nevertheless, like the modeled phenomenon, the dynamics generated by the model equations will most often have a low-dimensional character. For example, a model of a large neural network may exhibit a periodically ‘spiking’ output; a model of a wind-driven ocean produces large vortices (‘eddies’). The central issues to be addressed by the nonlinear scientist are: ‘What is the mechanism behind this low-dimensional behavior? Can it be understood? Can it be predicted? Can it be reduced or transformed to another mathematical representation and as a result be studied in more detail?’ Of course, this is the most simple setting, in general there will be several interacting mechanisms. For instance, the neural network will only produce transient ‘bursts’ of spiking activities, eddies in the ocean will eventually be annihilated. Nevertheless, this is the level at which the cross-fertilization between the mathematician and the cell biologist, or meteorologist etc., takes place. The questions and experimental observations of the cell-biologist can now be translated into mathematical concepts, mathematical theory can be interpreted in terms of behavior of the atmosphere. At this level, there might be a stunning similarity between the eddies of the oceanographer and the spikes of the neuro-scientist, for instance since they can both be seen as the product of a Hopf bifurcation [22, 46].

Often, however, the mathematics behind a natural phenomenon cannot be identified right away. The questions of the cell-biologist initiate a novel way of thinking about the dynamical system, and for instance stimulate the further development of the bifurcation theory for systems with delay. Or, the behavior of a projected finite-dimensional model of Lorenz-type may be realistic from the meteorological point of view, but there is (yet) no mathematical theory available by which its persistence can be established, i.e., it cannot be concluded that the underlying full infinite-dimensional PDE model generates qualitatively the same behavior.

Numerical analysis is an essential ingredient of the interaction between mathematics and earth and life sciences. State of the art numerical methods are necessary to simulate the models accurately, reliably and fast. The three major aspects of the application of mathematics to the earth and life sciences, modeling, analysis and simulations, are strongly intertwined. For instance, the numerical approach to systems with separated spatial and/or temporal scales is intimately linked to mathematical analysis [23]. Projection or truncation methods arise both in numerical analysis and in dynamical system theory. The application of dynamical system theory to explicit applications has been boosted by numerical continuation methods. Although programs such as AUTO have originally been developed in the context of low-dimensional models, they are these days crucial for a dynamical system approach to realistic models in earth and life sciences.

For the mathematician, simulations may play the role experiments play for the physicist and the combination of analytical studies of prototypes and numerical studies of complex models incorporating quantitative details gives more insight than each of them separately.

### 1.4.2 Building a national research team

Models in earth and life sciences appeal to significant parts of the fields of dynamical systems theory, of scientific computing and of partial, functional and stochastic differential equations. The composition of the research team is such that the cluster covers in essence all the theoretical ingredients that are necessary for the interaction with the earth and the life sciences. Within the cluster, there is an extensive experience in joint projects with scientists in the earth and life sciences (see section 1.4), but also with physicists, engineers, chemists and economists. Therefore the cluster is indeed an optimal foundation on which a national research team in the field of ‘natural dynamics’ can be built.

Traditionally, there is a stimulating exchange of ideas between Dutch nonlinear mathematicians at the theoretical level. However, the contacts between these mathematicians and, for example, biologists or meteorologists, have mostly been restricted to individual, or local, cooperation projects. Building on the experience of these successful cooperation projects (section 1.4), the cluster intends to make the next step and to establish a strong team in which the research and education in nonlinear dynamics of natural systems is stimulated and coordinated at a national level.

The aim of the cluster is to foster the exchange of ideas, views, insights, problem formulation and methods amongst scientists interested in mathematics and earth and life sciences. The members of the cluster are among the active participants, but there are many more. An important objective is to attract young talented researchers to the field, to motivate them and to help them getting an overview.

### 1.4.3 Central themes

The mathematical theory of dynamical systems has reached the stage in which it can serve as a ‘tool-box’ for the development and understanding of the complex models of life and earth sciences. Conversely, the phenomena generated by these models strongly stimulate the further development of nonlinear mathematics. Here, we mention some of the central themes associated to this interaction.

**Bifurcations and chaos.** A dynamical system most often depends on external parameters. A central theme which underlies many other sub-themes, concerns bifurcations in the dynamical behavior upon variation of such parameters. A sound characterization of the phenomenon of chaos, as well as the transitions leading to chaos has, up to now, only been possible in low-dimensional systems [25, 37]. One specific aim is to extend this characterization to higher dimensions. Identifying finite-dimensional phenomena that can explain the actual behavior of an infinite dimensional system lies at the core of the theme. Part of the higher and infinite dimensional connections with finite-dimensional dynamics is in the celebrated Ruelle-Takens program [40], see also [4, 5]. The reduced models in earth and life sciences often have additional structures and/or symmetries which strongly influence the dynamical behavior of the system [12, 21], and therefore special attention will be given to the further development of bifurcation theory for structured systems. There is a strong cross-fertilization between fundamental and applied elements of dynamical systems. The analysis of applications will also lead to the new developments in the theory of global bifurcations (both homoclinic and quasi-periodic) and to the characterization of chaos in intermediate dimensions [4, 37].

**Networks and delays.** Cell activity involves a huge regulatory network of protein interactions driven by external physical and chemical signals in which the physical processes themselves influence the way in which the information coded in the genes is used. This leads to complicated dynamical systems with feedback loops [17]. It is very natural that there are time delays in the feedback loop and this leads to an infinite dimensional dynamical system [10, 20]. Although at a different scale, similar infinite dimensional dynamical systems arise to describe physiologically structured populations. These last models are developed to tackle challenging questions, like whether one can predict catastrophic shifts in ecosystems from observable spatial patterns. A difficult problem in modeling the behavior of networks is that they often consist of huge numbers of ‘nodes’ but cannot be described by continuous models. This leads to models based on cellular automata, coupled map lattices and lattice differential equations [26, 43]. The analysis of such models leads to fundamental questions in the theory of functional analysis and dynamical systems.

**Scientific computing.** At present, a detailed bifurcation analysis of a realistic model is beyond the scope of existing numerical techniques. Novel numerical insights are necessary, especially since numerical methods in systems with strongly differing scales are typically unstable and errors may grow in an uncontrollable fashion. Like dynamical system theory itself, numerical analysis may have to adopt ideas from differential geometry to develop new methods. Numerical procedures that preserve conserved structures and symmetries are promising new approaches [19]. Although software packages such as AUTO and CONTENT (see url: [www.enm.bris.ac.uk/staff/hinke/dss/continuation/](http://www.enm.bris.ac.uk/staff/hinke/dss/continuation/)) have evolved into widely applicable packages, they are not yet capable of analyzing high-dimensional complex systems in full detail. Active research on how these tools can be improved has a high priority.

**Transient dynamics and multiple scales.** The low-dimensional ‘coherent structures’ exhibited by realistic models in meteorology, oceanography, neuro-physiology etc., are often not stable in the classical sense (i.e., for all time). Coherent structures may appear, exist for some (long) time, and eventually disappear. This continuous transient process of emergence, slow evolution followed by annihilation is remarkably subtle, but several aspects can nowadays be studied mathematically in reduced settings. It is often directly related to the multiple-scale nature of the model, but the slow and/or localized dynamics can also be induced by other intrinsic dynamical properties of the model. This sub-area is especially promising, since there is an intimate relationship between experimental observations, computer simulations, modeling, numerical analysis and dynamical systems theory [15, 21, 23]. Another feature of

multi-scale problems is that there can be hierarchies of length and time scales that should be eliminated subsequently, leading to a hierarchy of reduced models. Examples of such a hierarchy are moving fronts and interface approximations. The development of these model reductions is essential for deeper analytical understanding as well as for efficient numerical simulations.

**Patterns and waves.** From the mathematical perspective, pattern formation is the discipline in which the evolution of solutions to high and/or infinite dimensional systems is studied. At present, the mathematical techniques can only cover the most idealized behavior of complex systems. Many patterns formed in nature are induced by interacting instability mechanisms or competing localized structures [12, 21]. The existing theory strongly exploits on the underlying linear structures, but some patterns cannot be traced back to linear mechanisms [45]. The development of novel mathematical procedures by which these patterns can be studied is essential to the successful application of dynamical systems theory to the earth and life sciences. Such procedures will involve ideas from spatial dynamics, spectral theory, (global) bifurcation theory, etc. [4, 15, 37, 41].

## 2 Life Sciences & Stochastics

The Life Sciences have already been named the “science of the 21th century”. The first few years of this revolution have been dominated by “genomics”, triggered by the complete sequencing of the genomes of humans and other organisms, and by the advent of new high-throughput experimental technologies that produce massive amounts of data (e.g. microarrays, mass spectrometry). Because probability models explicitly allow for indeterminism, they are highly suitable to model complicated processes in living cells, as well as the behaviour of organisms and populations that are their aggregates. At present the application of such models in the life sciences is a wide-open area, which we are eager to explore more. In addition to modelling there are many opportunities and needs for statisticians helping to analyse the data arising from the new experimental platforms.

In the past years many steps have been taken in setting up new cooperations with biologists, medical scientists, geneticists and biological psychologists<sup>1</sup>. For instance, we are involved in the *Centre for Medical Systems Biology* (CMSB, UL/LUMC-VUA/VUMC-EMC-TNO), the *Institute for Integrative Biology* (IBIVU), and the *Centre for Neurogenomics and Cognitive Research* (CNCR-VU). We wish the cluster to become the main expertise centre for stochastics in the life sciences in the Netherlands.

Statistical techniques for investigating correlations between diseases or treatments and patient variables have long been an important motivation for the research of members in the cluster (e.g. survival analysis, semiparametric modelling, causality). Other topics in the life sciences that have been studied for several years in our group are carcinogenesis and ion channel kinetics, and many other projects were started up more recently. A large variety of mathematical methods is involved: hidden Markov models, Bayesian techniques and MCMC algorithms, stereological techniques, branching processes, point processes, Lévy processes, complexity regularization techniques, combinatorics, Mellin analysis of generating functions, asymptotic algebra, system and control theory in combination with analysis and polynomial algebra, and Bayesian networks, to mention a few. In several projects a systematic approach based on probabilistic modelling and statistical analysis is completely new, in others the current modelling or statistical approach does not suffice for the complex biological situation or for the massive data set at hand. The cluster’s life science research will be concentrated in a couple of sub-themes which we list below.

### 2.1 Statistical genetics

The major focus here will be statistics for genetic epidemiology aimed at connecting disease status and other trait variables to genetic factors by linkage and association analysis for humans. We shall in particular be involved in twin research in cooperation with the Biological Psychology group of D.I. Boomsma at the VUA. We will also collaborate with P. Heutink’s Clinical Genetics and Human Genetics group at the VUMC, and the Medical Statistics group of the LUMC headed by J.C. van Houwelingen, and with the mathematicians B. Basrak (Mathematics, University of Zagreb), S. Hernandez-Alonso (Mathematics, Universidad Rey Juan Carlos, Madrid), and D.O. Siegmund (Statistics, Stanford University).

On the methodological side we shall investigate and extend the use of MCMC methods for genetic analysis of twin data, of which the first examples have only recently been published in the genetics journals and which allow a greater flexibility in modelling; we shall introduce semiparametric models, e.g.

---

<sup>1</sup>See e.g. <http://www.math.vu.nl/~degunst/statsforlife.html> .

copula models, for greater realism than the usual normality assumptions, which are known to fail in many situations; and we shall introduce models from survival analysis. Much attention will go to the modelling and analysis of the genetic effects of multiple genes and their interactions on complex quantitative traits. Although this is inherently difficult, modern statistical methods, such as high-dimensional and semiparametric regression methods combined with methods of model selection through penalization or Bayesian analysis, will help to link trait values to e.g. IBD status at multiple loci. For traits that develop over time, such as migraine, we shall embed genetic models in models for survival analysis. The development of suitable copula models describing the dependence of the onset disease in twins or otherwise related persons will be an essential ingredient. The methods used to fit these models to the available data need to be able to cope with all the usual modes of censoring: e.g. right censoring, truncation and current status censoring. Combination of data for different countries have raised questions regarding the assumption of linkage equilibrium, which is basic to many genetic studies, and requires fundamental rethinking of the standard assumptions and methods of analysis, in particular regarding methods for combined linkage and association analysis.

An essential element of the investigations will be the practical implementation of the new models. It is an unfortunate fact that many great statistical ideas (e.g. model search) are hard to apply to the available data, given the current limited (!) computational possibilities. In some cases we shall cooperate with computer scientists specializing in parallel and distributed computing (specifically the group of H.E. Bal at the Vrije Universiteit) to alleviate these problems.

## 2.2 Epidemiology

This includes survival analysis and causal inference, research subjects that are well established for our group, and that are strongly connected to fundamental research in semiparametric statistical modelling. Our current research is joint with J. Robins (Harvard School of Public Health), M. van der Laan (Biostatistics, University of California, Berkeley), and N. Keiding (Biostatistics, University of Copenhagen), and concerns conditions under which causal conclusions can be drawn from observational data, and the development of the necessary statistical techniques, and the inclusion of very high-dimensional covariate data. We recently also started collaboration with the epidemiology group of Prof. dr. D.E. Grobbee, Clinical Epidemiology, Julius Centre, UMC Utrecht, on the effect of policy decisions on mortality due to tobacco smoking, and with Prof dr. L. van de Gaag, Computer Science, on missing data in expert systems based on graphical models.

## 2.3 Techniques for high-dimensional data

High-throughput genomics, proteomics, metabolomics and physiology generate massive amounts of data, thought to contain “all the information about life”. Extraction of signal from noise for this type of data sets requires new statistical techniques and computational paradigms. Under this sub-theme fall the statistical analysis of comparative genome hybridization (CGH) and gene expression microarray data, and proteomics analysis. We currently collaborate with the groups of A. Frigessi (Mathematics, University of Oslo), J.C. van Houwelingen (Medical Statistics, VUMC), G.A. Meier (Pathology, VUMC), A.B. Smit (Molecular and Cellular Neurobiology, VUA), and B. Ylstra (Microarray Facility, VUMC), but there is a definite demand for further input from statisticians to answer the many questions that arise through the application of continuously changing data-gathering platforms. For instance, multiple tests for significance of genes, latent variable models, reconstruction of CGH-profiles from data with overlapping or nonoverlapping clones, classification of disease status from SELDI-TOF mass/charge ratio profiles, gene and protein selection through feature selection methods, clustering of genes, proteins, or patient groups, linking survival to gene expression or CGH profiles, or design of CGH-chips. At a deeper level we focus on the development of techniques for the simultaneous analysis of different types of data, in the context of systems biology, e.g. as described below under network modelling. In this area our cluster will cooperate with computer scientists involved in data-mining and machine learning techniques: J. Heringa and E. Marchiori, both Computer Science, VUA. The research is strongly connected to fundamental research in high-dimensional statistical modelling (e.g. penalized minimum contrast methods, nonparametric and empirical Bayes methods, semiparametric modelling, classification methods), and links up to computational methods, including evolutionary computing and kernel learning methods (e.g. support vector machines).

## 2.4 Network modelling

This sub-theme concludes both genetic networks, and activity patterns in neuronal and other cellular networks. One project, with A.B. Smit (Molecular and Cellular Neurobiology, VUA) and J. Verhaagen (Neurogeneration, Netherlands Institute for Brain Research), aims at development of statistical and computational methods directed at building a model of the initial genomic response that underlies out-growth and elongation of neurons, and at finding regulatory cis-acting elements in genes involved in the coordinated growth response. The goals will be pursued by a combination of statistical approaches, such as Bayesian network analysis for genomic networks and hidden Markov modelling for genomic sequence searching, and be based on gene expression data from the rat dorsal root ganglion, as well as on information from existing databases. Some of the issues that will need to be solved are the incorporation of noise and temporal aspects, model selection, and the issue of causality.

Jointly with N. Lalam (EURANDOM) we are starting a collaboration in research concerning developmental gene networks in *Drosophila* with J. Kaandorp (Computer Science, UvA) and J. Reinitz (Applied Mathematics and Statistics/Developmental Genetics, Stony Brook University). The goal is to develop realistic models, based on data concerning gene products, for the gene networks that rule the embryonic development of several species, and, still more ambitiously, models for the living cell.

In cooperation with the groups of A.B. Brussaard (Experimental Neurophysiology, VUA) and J. van Pelt (Neurons and Networks, Netherlands Institute for Brain Research), we will deal with the statistical analysis of spatiotemporal patterns of electrical activity in neuronal networks. Recently introduced experimental techniques allow the monitoring of these patterns in great detail by simultaneous recording of neuronal activity from a large number of locations in the network (e.g. in cortical brain slices and cultured neuronal networks). Ultimately, we want to be able to analyze the changes in spatiotemporal neuronal activity patterns of mice that are the result of random mutagenesis, in order to determine whether a gene mutation has taken place that affects network activity and which synaptic parameter has most likely been affected.

With the same biologists we investigate the firing patterns of axons. The fire frequency of axons is determined by the firing processes at the synapses and a threshold value  $q$ : if the cumulative signals from the synapses exceed  $q$ , then the axon “fires”. The effect of synapses nearby is clearly different from the effect of synapses far away. Using entropy considerations to quantify this we are studying the contribution of individual synapses, or individual segments of synapses, to the firing patterns of an axon.

Finally, the ion channel research mentioned above (with B. van Duijn, Center for Phytotechnology UL/TNO, and J.G. Schouten, CBS) which aims at development and assessment of new Markov Chain Monte Carlo (MCMC) techniques for the statistical analysis of ion channel kinetics, will move from studying single channel kinetics to networks of ion channels. In the carcinogenesis research (with E.G. Luebeck, Fred Hutchinson Cancer Research Centre, Seattle) the focus has also changed, namely from understanding the influence of certain chemical compounds on the different stages of the carcinogenesis process to investigating gene regulation networks for carcinogenesis.

## 2.5 Modelling of cell processes

An ultimate goal of integrative bio-informatics, or systems biology, is to build a model that can simulate the functioning of a cell. It would be natural to think here of probabilistic models, which guide molecules to certain interactions, of the same type as described under “Physics” below. For example, a small pilot project directed at investigating the feasibility of the application of Markov models to understand the influence of macromolecular crowding on the flux through the bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) has been performed and showed promising results.

Another approach is being taken via control and system theory of biochemical reaction networks. The aim here is to gain insight in the functioning of the cell in relation to rational drug design. System theory is used to formulate methods for system reduction and for realization of rational positive systems. Control theory will be used to formulate methods for rational drug design. The innovative aspects concern the control and system theory for positive systems. This area is largely underdeveloped while it is urgently needed by biology and biotechnology. The long term applications are drug design and control of biochemical processes as in bioreactors and in waste water treatment plants.

The research in this area will be performed in collaboration with J.-L. Gouzé (INRIA, Sophia Antipolis, France) and with the group of H.V. Westerhoff (Molecular Cell Physiology, VUA). In the direction of systems biology there are contacts as well with R. van Driel (BioCentrum Amsterdam, UvA).

## 2.6 Modelling of population dynamics

One of the goals of population genetics has been to explore the genetic diversity in a population in order to understand DNA evolution. Mathematically, the diversity can be modelled as a partition of a sample into distinct types or categories. A central model of this type is a random partition described by Ewens' sampling formula (ESF). ESF naturally appears under the assumption of allelic non-interference, also called "neutrality". Recently, considerable progress has been achieved by members in the cluster in constructing a very large family of partition structures using the idea of regeneration, which may be seen as one-sided neutrality, already explored in some form in statistics and the coalescence theory. Current research is aimed at developing the related ancestral models of fragmentation and coalescence, and is joint with A.D. Barbour (Mathematics, University of Zürich), J. Pitman (Statistics, University of California, Berkeley), and M. Yor (Laboratoire de Probabilités et Modèles aléatoires, Université Pierre et Marie Curie, Paris).

On a different level of population dynamics we study the spread of infectious diseases, like swine fever, among animals. We are interested in modelling and estimation of parameters. We approach the problem using branching processes. Our interest is both theoretical (maximum likelihood estimators, martingale methods) as well as practical (computational methods, approximation methods). We work with unusual datasets in which we only observe those farms that are no longer infective. This is different from the usual statistical procedures in branching processes, in which one observes the full generations. The other innovative aspect is that we allow for a changing environment, that is, we allow the parameters to change in time and/or place. We cooperate with the groups of O. Diekmann (Mathematics, UU) and M.C.M. de Jong (Veterinary Medicine, UU/Animal Sciences, WUR).

## 2.7 Image analysis

Probabilistic models (e.g. point processes) are at the basis of many techniques for recognizing, reconstructing, improving, tracking, or comparing 2D or 3D pictures, or 4D videos of which numerous applications can be found in the life sciences. Our cluster is currently involved in a number of multidisciplinary projects in image analysis. For instance, germ-grain and marked point processes are useful for modelling the semantic content of images, and we develop MCMC algorithms for fitting such models. There are cross connections with quantum tomography, as discussed under Section 1.3.

In cooperation with the PET centre of the VUA medical centre (A.A. Lammertsma) we work on the analysis of images acquired by Positron Emission Tomography (PET), which gives indirect measurements of radioactivity in living organisms (e.g. the human brain). Because of the Poisson nature of radiation, the best reconstruction and comparison techniques are statistical in nature. We are a.o. developing likelihood based techniques for comparing images of brain activity of a single person under multiple conditions. The EM algorithm based on a Poisson statistical model of radioactive counts is the state of the art reconstruction technique to back project measured radioactivity projected along lines (Radon transform) to intensity in the brain. Our research aims at developing a similar likelihood based technique for comparing images of brain activity of a single person under multiple conditions. Furthermore, we focus on making similar methods available for the analysis of ligand studies, where there is the additional complication of having to determine the change point in the observed intensity curves, corresponding to the changed stimulus environment.

# 3 Het uitgebreide NDNS<sup>+</sup> cluster

## 3.1 Inleiding

Levenswetenschappers hebben behoefte aan wiskundigen die met hen meedenken over het opstellen van wiskundige modellen, het doorrekenen van kwalitatieve en kwantitatieve voorspellingen uit die modellen, en het aanpassen van die modellen aan empirische data. Voor een deel gaat het om standaard wiskunde, waarvan het bestaan en het gebruik echter vaak alleen bij wiskundigen bekend is, en voor een deel gaat het om nieuwe vraagstellingen die nader wiskundig onderzoek noodzakelijk maken.

De meest betrokken onderdelen van de wiskunde zijn de analyse en de stochastiek. Voor de levenswetenschappers zelf is het onderscheid tussen de diverse typen wiskundigen niet relevant. Ook voor de wiskundigen zelf is het nuttig om over de schotten tussen de deelgebieden heen te kijken, met name waar het doel multi-disciplinair onderzoek is. Bezien vanuit het bevorderen van het thema levenswetenschappen is het

zinnig om in gezamenlijkheid als wiskunde met een breed initiatief naar buiten te treden in de richting van de levenswetenschappen.

Het NDNS<sup>+</sup> voorstel is gericht op dit doel. Het voorstel beoogt stochastiek voor levenswetenschappen onder te brengen binnen het cluster ‘Nonlinear Dynamics of Natural Systems’ (NDNS). Inhoudelijk is dit stochastiek deel beschreven in sectie 2.

Dit voorstel richt zich in essentie op de interactie tussen de wiskunde en de levenswetenschappen. Opgemerkt dient te worden, dat dezelfde thematiek – de raakvlakken tussen de analyse en de stochastiek – een vergelijkbaar belangrijke rol speelt binnen de interactie tussen de wiskunde en de aardwetenschappen. De hier voorgestelde uitbreiding van NDNS sluit daarom goed aan bij de binnen NDNS gekozen focus op de aard- en levenswetenschappen.

### 3.2 Research plan

Voor het modelleren van processen in de levenswetenschappen ligt het toevoegen van stochastische storingstermen aan analytische modellen voor de hand, of andersom, het afleiden van een deterministisch systeem voor een evenwichtstoestand, of als eerste benadering, van een stochastisch systeem. Analyse en stochastiek zijn op deze wijze te onderscheiden gezichtspunten, maar grijpen duidelijk in elkaar, en hebben een groot raakvlak in het modelleren van niet-lineaire processen. Voor het aanpassen van modellen aan empirische gegevens zijn zowel numerieke methoden als de statistiek essentieel. Dit zijn onderdelen van respectievelijk de analyse en stochastiek die een sterke overlap hebben, en vormen tegelijkertijd een voorbeeld van de interactie tussen deze vakgebieden.

Gezien de betrokken wetenschappers, de ontwikkelingen binnen beide vakgebieden (de levenswetenschappen en de wiskunde) richt de uitbreiding van NDNS<sup>+</sup> ten opzichte van NDNS zich op de volgende 4 samenwerkingsgebieden:

- Cell Biology:
  - Membranes & Macro-Molecules
  - Systems Biology
- Neur(on)al & Gene Networks
- Epidemiology & Infectious Diseases

Dit zijn onderzoeksgebieden die in het brandpunt van de aandacht staan binnen de levenswetenschappen en komen mede daarom zowel in sectie 1 als sectie 2 aan de orde. Voor meer inhoudelijke details wordt daarom naar deze secties verwezen.

## References

- [1] R. Albert, H.G. Othmer [2003], The topology of the regulatory interactions predict the expression pattern of the segment polarity genes in *Drosophila melanogaster*, *J. Theor. Biol.* **233** 1–18.
- [2] M. Arrayas, U. Ebert, W. Hundsdoerfer [2002], Spontaneous branching of anode-directed streamers between planar electrodes, *Phys. Rev. Lett.* **88** 174502.
- [3] J.G. Blom, M.A. Peletier [2004], A continuum model of lipid bilayers, to appear *Euro. J. Appl. Math.*.
- [4] H.W. Broer, G.B. Huitema, M.B. Sevryuk [1996], ‘*Quasi-periodic tori in families of dynamical systems: order amidst chaos*’, LNM **1645**, Springer.
- [5] H.W. Broer, B. Krauskopf, G. Vegter (eds.) [2001], ‘*Global Analysis of Dynamical Systems*’, IOP.
- [6] H.W. Broer, C. Simó, R. Vitolo [2002], Bifurcations and strange attractors in the Lorenz-84 climate model with seasonal forcing, *Nonlinearity* **15**(4) 1205–1267.
- [7] S. Camazin, J.-L. Deneubourg, N.R. Franks, J. Sneyd, G. Theraulaz, E. Bonabeau [2001], ‘*Self-organization in Biological Systems*’, Princeton University Press.

- [8] C. Cotter, J. Frank and S. Reich [2004], Hamiltonian particle-mesh method for two-layer shallow-water equations subject to the rigid-lid approximation, *SIAM J. Appl. Dyn. Syst.* **3** 69–83.
- [9] D. Crommelin [2002], Homoclinic dynamics: a scenario for atmospheric ultra-low frequency variability, *J. Atmos. Sci.* **59** 1533–1549.
- [10] O. Diekmann, S.A. van Gils, S.M. Verduyn Lunel and H.O. Walther [1995], ‘*Delay Equations: Functional-, Complex-, and Nonlinear Analysis*’, Springer-Verlag.
- [11] H.A. Dijkstra [2000], ‘*Nonlinear Ocean Dynamics*’, Kluwer.
- [12] A. Doelman, A. van Harten (eds.) [1995], ‘*Nonlinear Dynamics and Pattern Formation in the Natural Environment*’, Longman.
- [13] A. Ehrenfeucht, T. Harju, I. Petre, D.M. Presscott, G. Rozenberg [2004], ‘*Computing in Living Cells*’, Springer.
- [14] G. Ermentrout, L. Edelstein-Keshet [1993], Cellular automata approach in biological modelling, *J. Theor. Biol.* **160** 97–133.
- [15] B. Fiedler (ed.) [2002], ‘*Handbook of Dynamical Systems, volume II*’, Elsevier.
- [16] Ch. Francke, P.W. Postma, H.V. Westerhoff, J.G. Blom, M.A. Peletier [2003], Why the phosphotransferase system of *Escherichia Coli* escapes the diffusion limitation of signal transduction, transport and metabolism that confronts mammalian cells, *Biophys. J.* **85** 612–622.
- [17] M. Freeman [2000], Feedback control of intercellular signalling in development, *Nature* **408** 313–319.
- [18] A. Gerisch, J.G. Verwer [2002], Operator splitting and approximate factorization for taxis-diffusion-reaction models, *Appl. Numer. Math.* **42** 159–176.
- [19] E. Hairer, C. Lubich, G. Wanner [2002], ‘*Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*’, Springer.
- [20] J.K. Hale, S.M. Verduyn Lunel [1993], ‘*Introduction to Functional Differential Equations*’, Springer.
- [21] P. Holmes, J.L. Lumley, G. Berkooz [1996], ‘*Turbulence, Coherent Structures, Dynamical Systems and Symmetry*’, Cambridge.
- [22] F.C. Hoppensteadt, E.M. Izhikevich [1997], ‘*Weakly Connected Neural Networks*’, Springer.
- [23] W. Hundsdorfer, J.G. Verwer [2003], ‘*Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*’, Springer.
- [24] J. Huisman, M. Arrays, U. Ebert, B. Sommeijer [2002], How do sinking phytoplankton species manage to persist? *The American Naturalist* **159** 245–254.
- [25] A. Katok, B. Hasselblatt [1995], ‘*Introduction to the Modern Theory of Dynamical Systems*’, Cambridge.
- [26] J.P. Keener [1987], Propagation and its failure in coupled systems of discrete excitable cells, *SIAM J. Appl. Math.* **47** 556–572.
- [27] H. Kielhöfer [2004], ‘*Bifurcation Theory. An Introduction with Applications to PDEs*’, Springer.
- [28] J.K. Krottje, On the dynamics of a mixed parabolic-gradient system [2003], *Comm. Pure App. Anal.* **2**(4), 521–537.
- [29] Yu.A. Kuznetsov [1998], ‘*Elements of Applied Bifurcation Theory*’, Springer.
- [30] E. Lorenz [1963], Deterministic non-periodic flow, *J. Atmos. Sci.* **20** 130–141.
- [31] L.R.M. Maas, A. Doelman [2002], ‘Chaotic tides’ *J. Phys. Oceanogr.* **32**(3) 870–890.
- [32] J. Mallet-Paret, S.M. Verduyn Lunel [2004], Exponential dichotomies and Wiener-Hopf Factorization for mixed-type functional differential equations, to appear *J. Differential Eq.*

- [33] A. Marée, P. Hogeweg [2001], How amoeboids self-organize into a fruiting body: Multicellular coordination in *Dictyostelium discoideum*, *Proc. Natl. Acad. Sci USA* **98** 3879–3883.
- [34] J.A.J. Metz, O. Diekmann [1986], ‘*The Dynamics of Physiologically Structured Populations*’, Springer.
- [35] A. Mogilner, L. Edelstein-Keshet [1996], Spatio-angular order in populations of self-aligning objects: formation of ordered patches, *Physica D* **89** 346–367.
- [36] Y. Nishiura [2002], ‘*Far-from-Equilibrium Dynamics*’, Americal Mathematical Society.
- [37] J. Palis, F. Takens [1993], ‘*Hyperbolicity & Sensitive Chaotic Dynamics at Homoclinic Bifurcations*’, Cambridge.
- [38] Gh. Păun [2002], ‘*Membrane Computing, an Introduction*’, Springer.
- [39] N.N. Pham Thi, J. Huisman, B.P. Sommeijer [2004], Simulation of three-dimensional phytoplankton dynamics: competition in light-limited environments, to appear *J. Comput. Appl. Math.*
- [40] D. Ruelle, F. Takens [1971], On the nature of turbulence, *Comm. Math. Phys.* **20** 167–192; **23** 343–344.
- [41] B. Sandstede, A. Scheel [2004], Defects in oscillatory media: toward a classification, *SIAM J. Appl. Dyn. Syst.* **3** 1–68.
- [42] V. Schwammle, H. J. Herrmann [2003], Geomorphology: Solitary wave behaviour of sand dunes, *Nature* **426**(6967), 619-620.
- [43] S.J. van Strien and S.M. Verduyn Lunel [1996], ‘*Stochastic and Spatial Structures of Dynamical Systems*’, Royal Netherlands Academy of Arts and Sciences, Amsterdam: North-Holland.
- [44] R. Temam [1997], ‘*Infinite Dimensional Dynamical Systems in Mechanics and Physics*’, Springer.
- [45] L.N. Trefethen [1997], Pseudospectra of linear operators, *SIAM Rev.* **39**(3) 383–406.
- [46] P.C.F. van der Vaart, H.M. Schuttelaars, D. Calvete, H.A. Dijkstra [2002], Instability of time-dependent wind-driven ocean gyres, *Phys. Fluids* **14**(10) 3601–3615.
- [47] L. van Veen, J.D. Opsteegh, F. Verhulst [2001], Active and passive regimes a in low-order climate model, *Tellus* **53A** 616–628.