UK Collaborative on Development Sciences Report: Complexity Science and International Development

Written by James Porter, Centre for Complexity Science, University of Warwick

Disclaimer

This review was requested by the UK Collaborative on Development Science membership; however the views presented in this paper are those of the author and do not necessarily represent the views of the organisations that authorised the document.

Complexity Science and International Development

Author: James Porter[1]

Overview

Research for international development draws on many different disciplines and involves many different objectives which may interrelate in complex ways. These objectives often relate to complex, partially understood, changing systems.

This report focuses on how to apply the methodologies of complexity science to international development. It introduces key complexity science concepts and examines several areas of complexity science in detail, showing how they are likely to be of potential utility for a range of international development scenarios.

The report first defines the idea of complexity science and surveys the various branches of the field. The direct application of complexity science methodologies to international development is examined next, with attention focused on three major methodologies: machine learning, computational modelling and network modelling.

The next section explores the connections between complexity science and complex policy questions, examining how the issues of uncertainty, the inherent complexity of development, multi-objective problems and multiple stakeholders might be illuminated by complexity science.

[1] University of Warwick, Centre for Complexity Science. Email address: j.a.porter@warwick.ac.uk
Thanks to Martine Barons, Andrée Carter, Sayantan Ghosal, Frances Griffiths, Thomas House, Matthew Keeling, Igor Khovanov, Charlie McClaren, Ben Ramalingam and Nicholas Watkins for their suggestions and/or feedback.

Executive Summary

Key Findings

- Ideas from complexity science are of clear relevance to international development.
- There are a wide range of methodologies to tackle complex problems in a scientific way which can be grouped together as complexity science.
- Approaches from machine learning of complex data, agent-based modelling and network modelling may be particularly useful in tackling a wide range of complex developmental issues.
- Complexity science provides some useful ways of framing and modelling issues such as multiple stakeholders and multiple objectives.

Background

Research for international development draws on many different disciplines and involves many different objectives which may interrelate in a complex way. These objectives relate to complex, partially understood, changing systems. Recognising this UK Collaborative on Development Sciences members expressed an interest in exploring whether complexity science might provide new understanding of international development. This report is an initial review to inform UKCDS members and other interested funders.

Objectives

The objectives of this report are:

- Provide a definition of complexity science.
- Give examples of actual and potential applications of complexity science to international development.
- Identify future possibilities and challenges in the application of complexity science to international development.

Complexity Science

This report focuses on how one might apply some of the methodologies of complexity science to international development. Complexity science is a difficult term to define. It is easier to give a list of features that are considered complex such as interconnected components, nonlinearity, self-organisation, adaptive agents, evolving and emergent properties. And to give examples of methods which are used to tackle systems featuring these properties such as nonlinear dynamics, stochastic processes, agent-based models and machine learning.

Complexity Science and its application to Development

Learning about processes with many interacting parts is computationally hard and will typically require large amounts of data. Often it is not clear what the relevant data for a particular problem is and the data which is available may be noisy. There are various approaches to deal with these and related problems, which will become more important for international development.

Agent-based models have been increasingly used for real world applications. They offer a flexible, tractable way of modelling a range of phenomena, especially those with some kind of purposeful agent. While the work in international development explicitly using such models has been limited, they have great potential for the future.

The mathematical model of a network (roughly speaking a series of points with connections to other points) has proved to be a very useful modelling technique for healthcare, epidemiology and for many other purposes. By focusing on individuals and their connections one can ignore much irrelevant detail and identify policies to most effectively ameliorate problems.

Complexity Science, Development and Policy

The question of how to model uncertainty is by its very nature a difficult one. While one can distinguish between uncertainty and risk, even the latter can be problematic to understand and model. Various mathematical and computational techniques introduce randomness in order to try and assess the robustness of conclusions or policy interventions.

Development is a complex process and in many cases averages, such as gross domestic product, may be of little use in assessing the effectiveness of a policy. More sophisticated measures are required.

There may be many objectives and many stakeholders. Such groups interacting and making decisions is often a complex task. While this is in theory difficult, in practice there may be benefits from such situations; the theoretical difficulties may not actually be particularly relevant to real situations.

Conclusions and Future Work

Complexity science offers a range of methods which at least attempt to tackle some of the many complex problems inherent in development.

The increasing availability of data will make many techniques from complexity sciences, which are designed for large data sets, increasingly important. Accelerating the availability and reliability of data relevant to development and how to best use this data are important on-going questions.

Techniques such as agent-based modelling and network modelling are beginning to be widely adopted across a range of areas, and supply chain analysis and crisis modelling are two areas of international development where such techniques could be particularly effective as they allow the explicit modelling of a dynamic, partially understood system.

When processes are complex we may be unable to predict or design precisely, but we may be able to achieve a high level of robustness. The kinds of techniques applied to complex climatic predictions or economic forecasting could be useful for a wider variety of applications.

Introduction

Relating the science of complexity to development is a broad, difficult and mostly unrealised task. Here some key complexity concepts are introduced along with several example areas of research with clear potential utility for international development. Examples of practice which draw explicitly or implicitly on complexity theory are introduced and connected to the areas of research.

In contrast to most existing work on complexity and development which looks at the important question of how to use the insights gained by viewing communities, societies, countries and development as complex systems, see for example Rihani, (2002) and Warner (2001), this report focuses on how one might apply some of the methodologies of complexity science to international development.

Several of these areas are examined in more detail; basic concepts are outlined, areas of obvious relevance are highlighted and real examples where these methods have been applied are studied. The rest of the report is split into three parts.

The first part sketches a fuller account of complexity science as defined for the purposes of this report. The second part examines some examples of those areas where complexity science is directly relevant to development. In this part issues surrounding complex data, the computational modelling of complex systems and modelling using networks are explored. Interspersed are applications which give an account of a hypothetical or actual use of Complexity Science methodologies, often drawing on supporting examples from parallel sectors where the ideas have already proved to be useful. In the third the scope of the investigation is widened to consider methods from complexity science which are important for policy matters. Finally the key conclusions of the report are summarised and areas for future work are outlined.

Complexity Science

Complexity Science is an imprecise term, covering an array of disciplines, methodologies and ideas. Defining it along the lines of *the scientific study of complex systems* still leaves one with an impossibly wide range of topics to cover. One can characterise a complex systems as one featuring some or all of the following properties:

- Interconnected components
- Nonlinearity
- Self-organisation/self-organised criticality
- Adaptive agents
- Evolution
- Emergent properties

A system is complex if it has many interacting components; so one cannot understand it as merely the sum of its parts and study each part in isolation – one must pay attention to the way in which different components interact. Furthermore these interactions may be nonlinear: given inputs to a system the output may not simply be the sum of some linear scaling of inputs, but some more sensitive process based on the input. But while this process may be more sensitive to input it may also be more robust, self-organising to classes or states; that is states with particular overall qualitative properties, without fine tuning of inputs to the process or initial conditions. But it may self-organise to a critical state; one where a small change in a particular input may have dramatic outcomes. Clear examples of this kind of phenomenon are avalanches or earthquakes, in both these cases the system slowly changes over time, but the overall state is apparently stable until a catastrophic event occurs, as the result of a small change in input.

There may be adaptive agents whose behaviour changes depending on their local environment and in doing so may regulate the overall behaviour of a system. This system may evolve over time, perhaps with co-evolution of individual components. And these various behaviours may result in the emergence of an outcome at a higher level that was not intended or explicitly aimed for at a lower level.

The term Complexity is often confused with 'chaos'. In a strictly mathematical sense they are quite distinct concepts. A chaotic system is entirely deterministic, that is for a particular set of inputs one will always obtain the same output, however that output may vary greatly with input[2]; hence the phrase "sensitive dependence on initial conditions". In contrast complex systems may actually result in qualitatively similar outputs for a wide range of initial conditions; and these systems themselves may be noisy (randomly perturbed) or stochastic (inherently random).

But given that a system of interest such as a (developing) country, a city, a piece of technology or a process has such features, the key question is what that actually entails. What methodologies, what "sciences", are of use in analysing, improving or building such systems?

[2] More precisely there is an exponential change in output given a change in input.

In general for systems with such properties exact mathematical solutions are impossible to obtain, so one must utilise other techniques. Complexity science draws on and combines a range of methodologies from disparate disciplines including physics, ecology, mathematics, computer science, economics and others. These include:

- Nonlinear dynamics
- Stochastic processes
- Statistical physics
- Agent-based modelling
- Evolutionary game theory
- Machine learning/statistical inference
- Decision theory

In calling a piece of work complexity science, in contrast to just pursuing say evolutionary game theory within a biological context, one is typically doing two things:

1. Emphasising the interdisciplinary aspects of the work. That is to say, applying work from one field to a seemingly unrelated field.
2. Emphasising the complex nature of the problem one is facing as work in each of the above fields could be split, albeit imprecisely, into “complex” and “non-complex”.

Both of these points are of importance for the application of scientific ideas to international development. The first almost by definition, and the second if one recognises international development scenarios as (often) having complex features.

Below several such methods are introduced. The list is of course incomplete but does cover most of the major techniques widely recognised as being part of complexity science.

Nonlinear dynamics is the mathematical study of systems which change in nonlinear ways. This can include chaotic systems. Results in this field typically try to characterise the final (asymptotic) state of the system, or to identify regularities in it such as fixed points[3] or orbits[4]. This analysis may reveal parameters for which a small quantitative change can qualitatively alter outcomes.

Stochastic processes is the study of probabilistic (random) systems. Results here often characterise the probabilistic distribution the system, for example a weather forecasting model which tells us there is a 70% chance of rain tomorrow in London.

Statistical physics shows how a difficult to model at the micro level system with many particles may in fact be very predicable at the macro level. Many physicists have tried to adapt results from this field to social or economic systems.

Agent-based modelling is a computational modelling technique in which the researcher explicitly models the behaviour of an agent in a system by a computer program. This program is then run in order to identify regularities in outcomes at the system level. In this way it is possible to identify

[3] A state of the system which doesn’t change.
[4] A state of the system which does change, but does so in a regular, predictable way; for example the moon’s orbit around the earth or a predator-prey population cycle which may arise in a Lotka-Volterra model.

unexpected features or to carry out virtual experiments where real experiments are impossible to carry out.

Evolutionary game theory takes ideas from game theory but rather than having a system where agents rationally consider all other agents and their payoffs one has agents whose actions are determined by a “genotype”. The sophistication of this “genotype” may vary from model to model; it could range from a number denoting a certain type of agent to a more sophisticated representation of the properties of the agent more closely aligned to the evolutionary biological inspiration for this approach. The population evolves with reproductive likelihood of a “genotype” dependent on the success of the actions the agents carry out, where the success is determined by the effectiveness of the “phenotype” arising from the “genotype”.

Networks are a widely used modelling tool in complexity science and have been applied to various social contexts, such as the spread of epidemics and friendships; along with many other processes such as protein interactions. They are an ideal tool for the consideration of structured interactions as they may be able to abstract away much unnecessary detail. For example when modelling the spread of a highly virulent disease, one thing that matters is how individuals are connected to other individuals; a network model may capture this detail, while ignoring other less important details. Furthermore, many aspects of networks are mathematically well understood.

Machine learning/statistical inference. The name for this will vary depending on the discipline, but here the general question is simply how to learn from data. In terms of complexity science one is typically thinking of problems where simple techniques such as linear regression do not work very well, where there are many interactions between variables and where there may be high levels of noise or uncertainty in our data. There are various established techniques for looking at such data, including the use of neural networks, interaction networks and nonlinear time series analysis which are recognised as being appropriate for complex problems.

Further Reading

For a fuller overview of many of the key concepts of complexity science and their relationship to international development the Overseas Development Institute’s recent working paper (Ramalingam, Jones, Reba, & Young, 2008) is an excellent starting point.

There are also a wide variety of general introductions to complexity science for a general audience, such as Strogatz (2003) and for a more technical audience Bar-Yam (1997), Érdi (2007) and Miller & Page (2007). A recent and clear introduction to the application of quantitative methods across many branches of the social sciences is Gelman & Cortina (2009).

Obviously each topic identified above has its own set of publications, conferences and so on as each is either an entire field of academic research or a category that is included within multiple fields.

Complexity Science and its application to Development

Complexity and Data

A classic example of the success of the automated use of data to produce a useful model is that of soybean disease classification (Chakrabarti, et al., 2009). Around 700 soybean plants were measured on thirty five different discrete attributes. The disease the plant was affected by was determined by an expert in the field (and thus each sample was given one of nineteen labels, each corresponding to a particular disease). From a portion of this data set (300 values) the expert came up with a set of rules to give a plant a disease label given the original thirty five measurements. The success rate (as measured on the remainder of the data) was around 72%. However, when this problem was attempted by a computer, the success rate was 97.5%; obviously a substantial improvement on the set of rules generated directly by the domain expert. But how does this kind of approach generalise to more complex problems and more complex models?

There is not yet a clear answer to this question but from, to take one example, the computational models utilised by investment banks to make trading decisions, we can see that it is both being attempted and has significant potential utility. The above problem used traditional (inference of rule based classification) machine learning techniques. In this section the key disciplines which seek to learn from data are introduced. Recent developments and prospects for the future are summarised and a brief overview of 'complex data' is given; focusing on the interrelated 'complex' nature of data with associated modelling techniques and the issues of volume of data with associated data mining techniques.

Depending on the broader discipline different names are used for the process of learning from data. For example, machine learning is the general term favoured by computer scientists and pattern recognition that by engineers. Definitions vary, but the following, taken from Alpaydin (2004) is representative, at least of the computer science view:

"Machine learning is programming computers to optimize a performance criterion using example data or past experience."

Across a wide range of phenomena (natural, social, technological) it is clear that changes are not simply random, that there is some underlying process which generates the data that one wishes to understand and act upon. Machine learning is a collection of techniques which attempt to do this.

While much traditional machine learning work focused on coming up with deterministic models of systems such as decision trees, much recent work has focused on statistical inference: that is learning from some data with noise or random variation. Bishop (2006), traces machine learning through three generations of approaches: the first is traditional machine learning, the second uses neural networks and the third, our current generation, is dominated by Bayesian approaches[5].

[5] Bayesian approaches explicitly models prior beliefs, which are then (repeatedly) updated when fresh data is available. These approaches make it possible to include in the modelling process the guidance of experts about how they believe a process works and then for this to be refined or corrected by actual data.

When high correlation is observed in data, and this is typically the case with social data, then traditional statistical techniques with their inherent assumptions of independence may be of limited value. One needs to model interactions between variables, something which is extremely (computationally) difficult. Considering even small numbers of combinations of variables requires both a lot of computation (to consider all such possible combinations) and a large amount of data to make meaningful conclusions. This poses a particular challenge for international development work (and socio-economic work in general) because it is often not possible to run repeated experiments or even to access reliable data on the processes which are naturally occurring.

One example of applying such ideas to highly correlated data was in looking at the treatment efficacy on various subtypes of a cancer. Gene expression microarrays can be used to get a (noisy) simultaneous measurement of the expression of thousands of genes in a sample[6]. By, for example, using neural networks the key genetic expression variables for certain types of diagnosis can be identified. This can allow treatment targeted to individuals and may be a substantial improvement on untargeted treatment. The survey by Greer & Khan (2004) introduces some of the key ideas and while in that form are only of indirect relevance to development; the techniques are useful in a much wider set of cases where sufficient data is available.

Data mining is the branch of machine learning which attempts to learn from very large quantities of data, much of which may not be relevant to the proper under investigation and much of which may have been collected or generated for other purposes. While most work in data mining takes place in information technology rich societies, it seems to be the case that a rapidly increasing amount of information is being generated by more and more people across the world. In fact one estimate[7] suggests we are now generating as much data every two days as we did from the dawn of civilisation to 2003.

For these problems the identification and pre-processing of data are often key issues, in contrast to many scientific investigations where the data is collected or even generated in a specific, controlled way (though it may be intrinsically noisy). One branch of data mining, text mining, focuses on the data obtained from large sources of textual content, such as the internet.

The statistical methods used must be computationally tractable for large amounts of data, so the ability to parallelise operations (work on individual parts separately) is desirable. While data mining has not been extensively used in the field of international development, given the increasing, often automated, capture of large quantities of information, it will become increasingly important.

Application

One strand of development economics has been the attempt to identify growth constraints on a country's economy. If one can identify constraints such as lack of education or specific types of infrastructure then one should obviously focus development efforts on these. But this kind of problem could be reframed from a complexity perspective: where there are many strongly interacting systems simply identifying one or two aspects of an economy which need work may not be the appropriate task. Is it possible instead to come up with a tractable, and realistic, way of

[6] Gene expression is how genetic information at the most fundamental level, results in functional products. Expression level data can be used in diagnose, for example of cancer tumour subtypes.
[7] From Eric Schmidt (Google CEO), August 2010, at Techonomy Conference.

thinking about how fundamental economic systems interact and in doing so formulate a more useful way of thinking about constraints on growth?

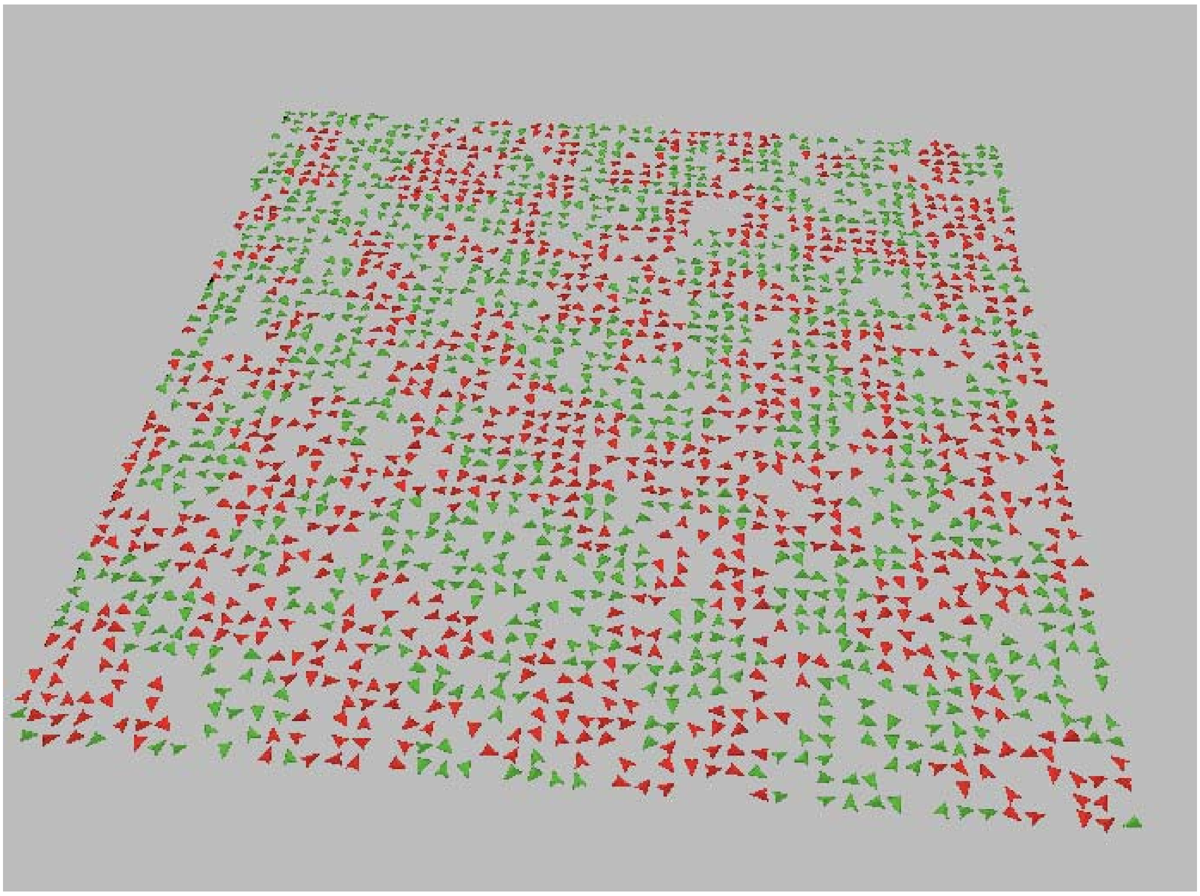
Ideally it would take the form of a computational model which could learn from the large amounts of data that is and will be available, and which could produce or guide policy recommendations which may be more effective than those produced by experts. Obviously this is towards the speculative end of the scale; a more modest, and attainable, goal might be the automated identification of infrastructure weaknesses. In the later section on Multiple Objective and Multiple Stakeholders a couple of examples of real world collection and aggregation of data are given; though from a machine learning perspective these are very straightforward as they involve little actual processing of machine learning from the data.

Complex Computational Modelling

Agent-Based Computational approaches have been widely recognised as a promising technique for general economic and organisational modelling, having been featured recently in high profile publications such as The Economist (Economist, 2010), the Proceedings of the National Academy of Sciences (Bonabeau, 2002) and Nature (Farmer & Foley, 2009). Typically the application area focused on is financial markets or western economies in a more general sense. However, it will be argued in this section that these techniques should be of use in international development, specifically in development economics, logistic modelling and organisational modelling.

A classic example of Agent-Based Modelling is the Schelling Segregation model (Schelling, 1971). In it there are two colours of agents. Each agent weakly prefers to live in a neighbourhood consisting mostly of his own colour. Agents are initially randomly scattered across a landscape. When agents are “unhappy” they move to a new location where they will be happy. Even though they would be “happy” to live in a mixed neighbourhood, they quickly segregate into blocks of their own colour. The figure below shows the final state of such a society. Notice how agents are segregated into distinct blocks of their own colour.

Perhaps most interestingly even if agents actually want to live in a mixed neighbourhood, then the global result of segregation is still the outcome. This kind of model which has very different micro and macro properties has been an active area of research right up to the present day. A recent paper (Kirman, 2010) explores various extensions to the original model, for example having both income and race, and how it may be related to observed instances of racial segregation in cities such as New York.



There have been some simple applications of this kind of model to racial and societal segregation within the international development literature. However, usually this kind of model is kept quite simple and the focus is on what qualitative insights one can gain from it, rather than attempting to model real scenarios and guide policy. In the remainder of this section agent-based modelling is introduced more generally and some of the future potential uses for international development are identified.

Agent-Based Modelling takes a bottom up approach. Instead of trying to build a single model incorporating the entire complex system, models of individual parts or agents are constructed, typically by explicitly specifying agent's behaviours in the form of a computer program. This stands in contrast to more traditional approaches adopted, particularly within Economics, which assume that agents maximise some function without specifying how this might actually happen.

Having specified our model as a computer program simulations are carried out under a variety of conditions. This allows one to see how high level effects (such as sustained growth in GDP) may arise from micro level interactions such as bidding, buying, selling or individual decision making. In Axelrod & Tesfatsion (2006) three goals for agent-based modelling of relevance to international development are identified:

1. *Empirical understanding*: direct understanding of existing process through computational modelling. For international development this might (ambitiously) mean directly and accurately simulating an on-going crisis in order to determine the best ways of responding.
2. *Normative understanding* or determining whether a proposed policy works or is better than others. Given that we only observe one outcome for a policy (whatever actually happened) it

can be hard to judge what policy is likely to be most effective for a particular scenario; if we can simulate many realisations we may be able to obtain a better answer (this is done in, for example, weather forecasting).

3. *Heuristic understanding*, one may gain insight from attempting to model a system in an explicit way.

Simple qualitative models can be built that investigate behaviour whilst not explicitly modelling the system at the micro level, such as those examined above. However, it is possible to build more realistic models of processes and behaviour. Investigating a particular field an attempt could be made to come up with a "medium range" model: one which includes realistic parameters but attempts to keep the model relatively straightforward and tractable. Alternatively a larger model may be built, for example, a central bank may attempt to model of an entire economy with many sectors and detailed features drawn from the real world.

There has been some work within the field of international development drawing on agent-based computational modelling, although there is much scope for more work. The existing use of agent-based models has been mostly restricted to simple qualitative models; usually examining issues relating to territorial conflict or segregation, one of the most prominent such models is by Epstein (2002). There is a limited amount of work which uses agent-based modelling as a way of producing models of conflicts to attempt the more difficult task of trying to model real world scenarios, though none has yet achieved an authoritative status[8], perhaps reflecting the difficulty of such a task.

The purported benefits of using agent-based models relate to their ability to model partially understood rapidly changing systems involving many heterogeneous elements. Traditional mathematical or economic models are often unable to provide solutions when modified to make them more realistic; agent-based models, subject to sufficient computational power and ability to properly specify a model, can.

Applications

Agent-based models have been applied to a wide range of practical problems from logistics to traffic modelling. The resource AgentLink, although a commercial promotional document, gives an attractive overview of many of the "success stories" and statistics related to real applications of agent-based modelling.

An important part of international development research has been the field of humanitarian logistics. The area of logistics has been one lauded application area for agent-based modelling so this seems like an obvious potential application. The ability to experiment with different supply chains and scenarios on computers, rather than either carrying out costly real world experiments or simply not experimenting at all, is an attractive possibility. While there has been substantial work done in supply chain analysis in international development, particularly in humanitarian relief, there would seem to have been little use of agent–based modelling.

A further use of agent-based modelling may be in simulating the results of policy changes on on-going crises. A proposal for a €1 Billion social modelling project, FuturICT[9], is suggestive. A key goal of this project proposal is the modelling of economic processes in response to an early warning crisis

[8] Judging by the metric of citation count.
[9] See http://www.futurict.ethz.ch/FuturIcT for details of the proposal.

system. A similar system might be of use, in say, predicting the effect of food price increases on entire societies.

Network Modelling

The use of a network as a modelling tool for societies and complex organisations is becoming increasingly common. In a network model there are a set of vertices, which could represent for example people, and a set of edges representing connections, for example social relations.

These models have proven particularly useful for work in healthcare. For example Christakis & Fowler (2007) looked at how obesity is linked to one's social relations. This longitudinal study looked at over 12000 individuals over thirty years and found that a person's chances of becoming obese were significantly increased if (s)he had a friend or partner who became obese, even when they corrected for other causes (as the data was for a heart disease study it included data for a large number of relevant factors). Many researchers have also argued that social networks are extremely important for understanding issues such as happiness, loneliness and cessation of smoking.

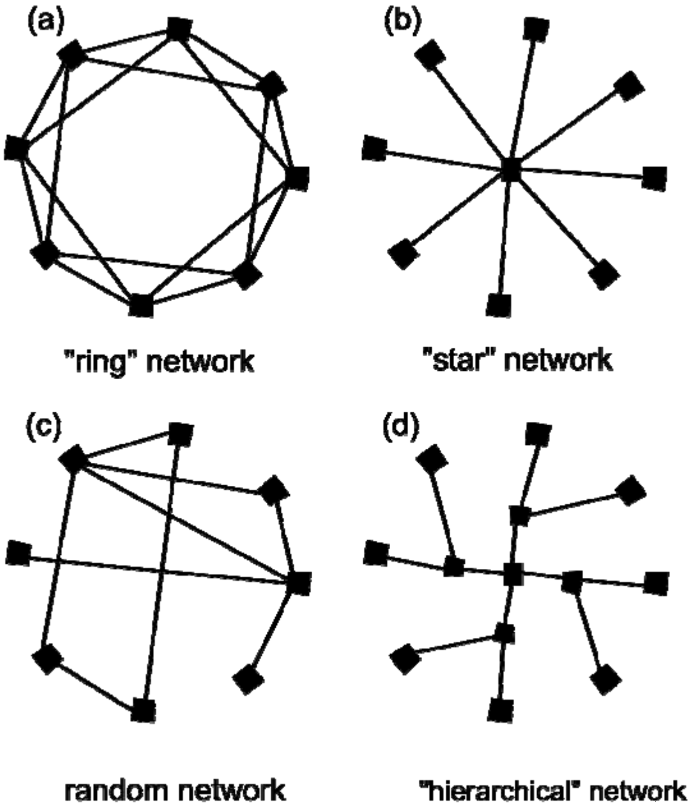
It is intuitively plausible, and has been widely argued, that social networks are significant for many issues. But how should this affect policy? Recent research has actually constructed social networks in order to investigate how structural features affect outcomes. For example Centola (2007) compares tightly clustered social networks with networks with 'longer range'[10] connections. He concludes that for certain behaviour spreads the repeated reinforcement of peers is necessary for that spread to propagate through the network. More recent work by Centola explores a social network set up for a fitness programme; each person signing up is assigned to a group with certain properties (such as consisting of very similar members to oneself). Through this the effectiveness of interventions and spread of beneficial behaviours can be examined, and attempts made to answer policy relevant questions about social network structure.

Unfortunately such studies, require the explicit construction of and control over a social network, which is extremely difficult to achieve. Furthermore, how applicable the results would be to the real world - with evolving social networks and many types of relations between individuals - is not clear. Realistically what one might obtain from these kinds of study are heuristic insights, such as the idea there may be behavioural patterns whose spread requires many individuals who have adopted the behaviour to be linked to an individual in order to spread it to that individual.

[10] Where individuals are connected to individuals their friends are unlikely to be connected also (i.e. people knowing many people perhaps none of their friends know).

Network Concepts

Networks[11] consist of a set of **vertices** or points and **edges** which connect them. These edges can be **directed** in which case they point from one node to another or **undirected** when there is just a connection, usually when modelling some kind of symmetric relationship. Below some simple networks are illustrated (all are undirected, usually directed networks are illustrated by the use of arrows from one node to another).



Networks provide a simple, computationally workable[12] way of modelling a wide range of phenomena, but one which allows modellers to keep the important details in the model. By explicitly capturing interactions between parts of a system, they are a natural tool to use for the modelling of complex systems.

There are various mathematical properties which may result in quite different macro level outcomes. The **degree** of a vertex is how many other vertices it is connected to. **Clustering** is a measure of how likely a neighbour of a vertex shares other neighbours, that is whether my friend is likely to be a friend of another one of my friends. If there is a **path** (some chain of edges) from each vertex to each other, then it is a **connected** network.

There are many sources of further information on network theory. Perhaps the obvious place to start for international development is Jackson (2010), a recent text which introduces the main mathematical theories and offers a comprehensive survey of existing work using networks.

[11] Networks are often referred to as graphs In scientific work, however here the term network is used for consistency and clarity.

[12] http://en.wikipedia.org/wiki/Social_network_analysis_software provides a comprehensive, though far from exhaustive, list of software packages for network modelling and analysis of social networks.

Social networks are one special case of network study and are by no means the only one relevant to international development. Promising avenues of research in networks include: looking at stability of network structures against random failures and directed attacks; epidemiological network models which may provide more accurate models of disease spread; educational models which look at how ideas spread through a network; and policy models which look at, for example, how individual agents minimising their journey time may result in poor overall traffic flow.

There are many examples of success stories of bottom-up development efforts; but, given a particular crisis, widespread risky or damaging behaviour or practice, how does one target interventions? Network analysis may offer one tool to address this. Networks have already demonstrated their utility in data rich healthcare work in the developed world. As more data becomes available in other areas, such as for developing societies, their usefulness for empirical and targeted interventions should increase. Furthermore theoretical insights into structural properties of networks can guide designs of processes even in environments with little available data.

Application

During the 2001 foot and mouth outbreak in the United Kingdom a team of scientists were assembled to analyse the data and to suggest the most effective policy interventions. The disease is a complex problem exemplified by listing just a few of its characteristics: it can infect multiple species including cows, pigs and sheep; it can be spread by various means including human movement; there is a vaccine but it may not be effective for all species; there are international flows (export and import) of meat; and restrictions on movement in areas may have debilitating effects on tourism and other industries. The challenge of identifying the most effective ways of mitigating the problem(s) was substantial, and the modelling techniques used then (Keeling, 2001) and for similar crises may be useful in tackling complex problems in international development when issues such as movement, disease spread and targeting of interventions are important.

Building complex models to look at diseases is an on-going area of research. Particular challenges are modelling the geographical spread of diseases (see Pareham et al. 2008), assessing the efficacy of vaccination and other interventions given uncertainty about disease spread (see Yang, et al., 2009) and looking at the interactions between disease and other systems, such as linking Cholera dynamics to El Niño-Southern Oscillation (Pascual et al. 2000).

The modelling of current disease spread, how it may interact with other systems and how policy can best tackle it is a challenge to which techniques from complexity science are, and should, increasingly prove useful. More generally complex models, and in particular network models, may be useful for modelling on-going crises and guiding policy.

Complexity Science, Development and Policy

Uncertainty and Stochasticity

Climate change has been widely recognised as a pertinent issue within the international development literature. Mitigation and adaptation may conflict or interrelate with development goals and climate change impacts must be dealt with as part of a comprehensive development framework. The former issue will be examined in the next section, but here the latter issue is focused on. Given that climate change will have dramatic but uncertain consequences for crop production, flooding, health care and more; what should be done? In this section uncertainty in climate change modelling, and the techniques used to deal with this, are examined as an example of attempting to deal with uncertainty in general.

Some of the key conclusions of the International Panel on Climate Change (IPCC) Fourth Assessment Report on the results and nature of climate change are:

1. That warming of the climate system is *"unequivocal"*.
2. World temperatures could rise by between 1.1 and 6.4 °C during the 21st century and sea levels will probably rise by 18 to 59 cm.
3. There is a high confidence level that there will be more frequent warm spells, heat waves, and heavy rainfall.

Notice the ranges (often substantial) in predicted event and the use of terms such as 'high confidence'. A recent UKCDS publication (Conway & Waage, 2010) provides an accessible introduction to the scientific techniques involved in the IPCC report and related publications, along with an overview of predictions and an account of changes.

In the early twentieth century Knight (1921) made the influential distinction between risk: where the result was random but the probability of particular results was known, and uncertainty: random results where the probabilities aren't/can't be known.

Large problems like climate change or macroeconomics typically involve aspects of both. Based on past observations or at least partial understanding of systems it is possible to at least estimate risk, while being uncertain as to how things that are not in the model or have been inaccurately included may affect outcomes.

Existing approaches from climate change research which are more generally applicable to other areas include looking at a large variety of parameters for a model (around estimated values) and try to test for robustness in conclusions. This could be done for either a traditional deterministic model (for a given input the same result always occurs) or a properly stochastic model (there is a probability distribution over results). Another approach is to take a model and randomly perturb it (change the state) and see if it will settle back to its original state or will remain in roughly the same state as before. While this approach is widely used in economics, it only makes sense if equilibria of the system (stable states) can be assumed to be the typical or important states of the system.

For any kind of large models both of these approaches are difficult. There may be computational limits to how many parameters can be explored and any limits in understanding i.e. uncertainty may fundamentally limit how much can be concluded. However, methodological improvements arising from complexity science or other research should be applicable to the complex modelling when applied to international development.

Complex development

India has seen substantial gross domestic product growth in recent decades, however there has been relatively little poverty reduction compared to China or Brazil (Ravallion & Datt, 2002). The Gini index (a measure of inequality) has increased and it would seem that the poor are largely unable to participate in the opportunities made available by the general economic growth. The issue of intra society inequality is important and can take a multitude of forms from the relatively straightforward inequality of income to issues such as the availability of opportunities for education and employment.

Much traditional work assessing the progress of development, indeed including that for the developed world, focused narrowly on measures such as GDP. While more attention has been given in recent decades to a more comprehensive, multi-objective view of development (most obviously through the Millennium Development Goals)[13]; the inherent complexity of development processes has commanded less attention. This section looks at two examples of approaches to the complexity of development, there are undoubtedly many more.

One attempt to address unequal development is via the construction of a measure of polarisation within a society via perceived ‘difference’ from other individuals such as that proposed by Duclos *et al.* (2004). One wishes to be able to distinguish between a society that is merely unequal in terms of end results[14] and one where the people are highly polarised and in which there is potential for unrest and violence. There may be polarisation between cultural, religious or class groupings which in extreme cases may result in conflict. Measuring and determining how to intervene in a way that reduces or minimises polarisation may be vitally important in avoiding future violence and for development in general.

Another concept to consider is that of aspiration, or what kind of career or form of living individuals consider possible and what they aim for. One strand of research in this area considers the cognitive window of individuals, or basically that set of other people whose jobs, education or lifestyle an individual considers attainable for herself or her children. There may be a failure of aspiration and an associated aspiration trap, whereby low aspirations among sections of society, such as rural poor, start low and as a result of the aspirations and experience, future aspirations remain low. The key issues here are how individuals determine their cognitive window and how aspirations might affect behaviour. The key complexity issue here is how individual decisions and the macroeconomic outcome are related; it may be beneficial for everyone for a greater proportion of the population to become well educated, but thanks to the way decisions are made individually, this collectively desirable outcome does not occur. Insights from psychology and behaviour economics should prove useful for addressing both questions.

[13] See the next section for an outline of work on multi-objective problems.
[14] Even this in itself is unlikely to be a particularly desirable outcome.

The above examples are both drawn from the recent development economics literature and are both fairly recent areas of study. Much work remains to be done on these and related approaches especially with respect to empirical implementation, but the development of models and measures which can comprehensively take account of important aspects of the complex process of development should lead to an improved understanding of development, and may suggest improved ways of achieving particular goals, or sets of goals.

Multiple Objectives and Multiple Stakeholders

In response to a disputed Kenyan election a group of ex-Kenyan residents quickly put together a website named *Ushahidi* [15] to aggregate information from witnesses on the ground; it collected reports of violence via email or text message, possibly including photographs and video along with written content, and placed these on an online map. These reports were then verified by a variety of known people and organisations. This enabled users of the service to obtain a spatial overview of the changing situation and to learn about events in particular areas[16]. In a retrospective examination of the situation in Kenya Rotich (2008) claims that: *"Ushahidi is revolutionary for human rights campaigns in the way that Wikipedia is revolutionary for encyclopaedias"* as it allows collaboration on a larger scale than previously possible.

In the aftermath of the 2010 earthquake in Haiti an open source street mapping organisation[17], drawing upon newly captured satellite imagery among other sources, *crowdsourced*[18] the creation of up to date street maps to volunteers across the world. These up to date maps were then able to be used by aid organisations for a variety of relief and longer term development purposes.

In the 5th chapter of Lathrop & Ruma (2010)[19] the notion of collaborative government is contrasted with deliberative democracy: instead of asking how to use modern technology and information to make decisions and improve policy, one asks how to divide up tasks and problems so they can be tackled by "the crowd"; instead of drawing on a diverse range of voices to ensure a lively and potentially improved policy debate occurs, the policy implementers instead attempt to draw on a diverse range of people and skills to solve problems. This kind of reframing, asking not what the right policy to reconcile various stakeholders' objectives but how to encourage stakeholder collaboration, may be particularly relevant to development.

These are examples of how modern technology has been used in crisis response and development efforts. However, from a machine learning point of view many of these efforts are relatively simple (at least in terms of the processing and collection of data) though they involve large groups related in complex ways. Perhaps improved methods can be developed to work with the kind of fragmented, multi-source and potentially unreliable data. But perhaps more interestingly from another point of view these examples are already "complex": they involve an unpredictable, bottom-

[15] Ushahidi is Swahili for "witness".

[16] The framework has since been improved and released as open source software and has been used in a wide variety of situations. See http://www.ushahidi.com/ for details.

[17] http://www.openstreetmap.org/

[18] It seems appropriate to use Wikipedia's definition: *crowdsourcing* is the act of outsourcing tasks, traditionally performed by an employee or contractor, to a large group of people or community through an open call.

[19] The first eight chapters including this one are available for free from http://oreilly.com/catalog/9780596804367

up process which allowed for the direct involvement of stakeholders in information generation and processing in crises. In both cases individuals not directly connected to the crisis, perhaps in a far off country were able to play an immediate role in crisis response and in longer term development efforts (via the establishment of higher expectations of security or the provision of improved mapping of these areas). The challenge here is how to better facilitate this kind of collaboration. Can we model and better understand these kinds of micro-level processes and design or facilitate better macro-level systems for complex collaboration processes? Models of networks, agent-based modelling and other complexity science approaches may be extremely useful.

The above analysis looked at the problem (and benefits) of multiple stakeholders and ways to address this. There is however the further widespread and complex issue of multiple objectives. As an example take the United Nations' Millennium Development Goals (MDGs). These are split into eight categories and many of these categories contain multiple targets. These various targets relate to each other and with respect to policy in ways which can be far from clear. In some cases an improvement in one area may lead to improvements in another, in other cases one may have to make a choice about which to focus one's attention on.

One of the problems is the lack of thinking about how to tackle multiple objectives. Most work, in particular within Economics but also within other disciplines, focuses on the optimisation or achievement of a target for one quantity as this makes the problem both easier to think about and make assessing the success of particular efforts clearer.

There are a few pieces of quite technical work examining the issue of multiple objective optimisation from economics, computer science and operational research (for a recent general introduction see Ehrgott, 2005). Perhaps some of the most useful thinking comes from Sen (2009). It argues on one had that many problems we face involve multiple incommensurable objectives but that in practice one often makes such decisions, so the lack of ability to give a single, precise answer should not and does not hold one back from making such decisions and should not force one to collapse everything into one flawed metric. Sen argues that the some of the classic paradoxes[20] in multi-person/objective decision making had the effect of diverting research attention away from this important area[21].

[20] The most famous of these is Arrow's impossibility theorem which states that, when voters have at least three options, no voting system can convert the rankings of individuals into a group ranking while also satisfying four apparently straightforward assumptions of an unrestricted domain, non-dictatorship, Pareto efficiency, and independence of irrelevant alternatives. A classic text which gives many of the fundamental results which were obtained is by Arrow and Raynaud (1986).

[21] See Sen (2009) for this argument.

Conclusions and Future Work

Complexity science offers a range of methods which at least attempt to tackle some of the many complex problems inherent in development. This report has summarised some of the key methods. Others such as statistical physics or nonlinear dynamics could have been examined, but the focus was on those areas with clearer applicability to existing problems in international development. That is not to say that other areas are not important, for example there is recent interest within economics on drawing on ideas from statistical physics to model complex economic markets.

The increasing availability of data will make many techniques from complexity sciences, which are designed for large data sets, increasingly important. Machine learning and data mining techniques have already proved useful in parallel sectors to international development. They are currently constrained in their application to international development by the lack in both volume and reliability of data. As has been argued above this should become less of a problem, but how to accelerate the availability and reliability of data relevant to development is an important on-going question.

Techniques such as agent-based modelling and network modelling are beginning to be widely adopted across a range of areas, and supply chain analysis and crisis modelling are two areas of international development where such techniques should be particularly effective as they allow the explicit modelling of a dynamic, partially understood system.

When processes are complex we may be unable to predict or design precisely, but we may be able to achieve a high level of robustness or stability against shocks. Methods which use random processes and modelling are one attempt at achieving this. The kinds of techniques applied to complex climatic predictions or economic forecasting are useful for a wider variety of applications than those to which they are already applied.

International development is fundamentally complex, while other works such as that by Ramalingam *et al.* (2008), have more to say about the inherent complexity, the technical approaches outlined above are promising avenues of research in thinking about many aspects of international development. However, perhaps the key and most immediate future challenge for international development and complexity science will be facilitating communication and interaction between international development practitioners and academic and theoretical scientists.

Supporting Materials

AgentLink. (n.d.). *50 Facts About Agent-Based Computing.* Retrieved October 6th, 2010, from http://www2.econ.iastate.edu/tesfatsi/AgentLink.50CommercialApplic.MLuck.pdf

Alpaydin, E. (2004). *Introductiono to Machine Learning.* MIT Press.

Arrow, K. J., & Raynaud, H. (1986). *Social Choice and Multicriterion Decision-Making.* MIT Press.

Axelrod, R., & Tesfatsion, L. (2006). A Guide for Newcomers to Agent-Based Modeling in the Social Sciences. *Handbook of Computational Economics, Volume II*, 1647-1659.

Bar-Yam, Y. (1997). *Dynamics of Complex Systems.* Westview Press.

Bishop, C. (2006). *Pattern Recognition and Machine Learning.* Springer.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 7280-7287.

Centola, D. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*.

Chakrabarti, S., Cox, E., Frank, E., Guting, R., Han, J., Jiang, X., et al. (2009). *Data Mining.* Elsevier.

Christakis, N. A., & Fowler, J. H. (2007). The Spread of Obesity in a Large Social Network over 32 Years. *The New England Journal of Medicine*, 370-379.

Conway, G., & Waage, J. (2010). *Science and Innovation for Development.* UKCDS.

Duclos, J.-Y., Esteban, J., & Ray, D. (2004). Polarization: Concepts, Measurement, Estimation. *Econometrica*, 1737-1772.

Economist. (2010, July 22). Agents of Change. *The Economist*.

Ehrgott, M. (2005). *Multicriteria Optimization.* Springer.

Érdi, P. (2007). *Complexity Explained.* Springer.

Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 685-686.

Gelman, A., & Cortina, J. (2009). *A Quantitative Tour of the Social Sciences.* Cambridge University Press.

Jackson, M. O. (2010). *Social and Economic Networks.* Princeton University Press.

Keeling, W. S.-T. (2001). Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. *Science*, 813-817.

Kirman, A. (2010). *Complex Economics: Individual and Collective Rationality.* Routledge.

Knight, F. H. (1921). *Risk, Uncertainty and Profit.*

Lathrop, D., & Ruma, L. R. (2010). *Open Government.* O'Reilly.

Miller, J. H., & Page, S. E. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life .* Princeton University Press.

Pascual, M., Rodó, X., Ellner, S. P., Colwell, R., & Bouma, M. J. (2000). Cholera Dynamics and El Niño-Southern Oscillation. *Science*, 1766 - 1769.

Pareham, P. E.., Brajendra, K. S., & Ferguson, N. M. (2008). Analytic approximation of spatial epidemic models of foot and mouth disease. *Theoretical Population Biology*, 349-368.

Ramalingam, B., Jones, H., Reba, T., & Young, J. (2008). Exploring the science of complexity: Ideas and implications for development and humanitarian efforts. *ODI Working Paper*.

Ravallion, M., & Datt, G. (2002). Is India's Economic Growth Leaving the Poor Behind? *World Bank Working Paper*.

Rihani, S. (2002). Implications of adopting a complexity framework for development. *Progress in Development Studies*, 133-143.

Rotich, J. G. (2008). Digitally Networked Technology in Kenya’s 2007–2008 Post-Election Crisis.

Schelling, T. C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 143-186.

Sen, A. (2009). *The Idea of Justice.* Allen Lane.

Strogatz, S. (2003). *Sync.* Penguin.

Warner, M. (2001). *Comlex problems... negotiated solutions: The Practical applications of chaos and complexity theory to community-based natural resource management.* Working Paper.

Yang, Y., Sugimoto, J. D., Halloran, M. E., Basta, N. E., Chao, D. L., Matrajt, L., et al. (2009). The Transmissibility and Control of Pandemic Influenza A (H1N1) Virus. *Science*, 729 - 733.