Integrating physical and social modelling - the example of climate change

Scott Moss Centre for Policy Modelling Manchester Metropolitan University

Claudia Pahl-Wostl Swiss Federal Institute of Environmental Science and Technology (EAWAG), Dübendorf

1. Methodological issues

Climate change brings into sharp relief the most difficult problems in addressing the interaction between science and policy. Instead of trying to solve an existing environmental problem one tries to prevent a problem that scientists expect to come. Results from model simulations have an unprecedented role in shaping the public debate. Forecasts of climate change and potential impacts are basic for any risk assessment. Economic models that compare costs and benefits of different measures provide important arguments for discussing response strategies. Modelling approaches in climate and economics draw their legitimisation largely from their being based on a body of theory that is well accepted in the scientific community. However, the information available for policy makers is plagued by large inherent uncertainties. Reality is complex and the prediction of future developments is beyond our capacities now and for the foreseeable future. In situations where decision stakes are high and uncertainty looms large the role of scientific arguments in general and of model simulations in particular may have to change (Pahl-Wostl et al 1998, Funtowicz and Ravetz 1996). We note little effort in the community of economists to take complexity and its implications into account. We argue therefore that another tradition of modelling, social simulation, is more appropriate in dealing with the complex environmental problems we face today.

Climate modelling shares a number of the key difficulties with social simulation. Each is based on a long history of observation and each can appeal to a number of well verified theories on which to base model components. For example, GCMs, the scientifically most highly valued class of climate models, are based on fundamental physical laws, the so-called "first principles" (e.g. the laws of thermodynamics). However, reality is complex and due to a lack of resolution in space, time and process, parameterisation generates a lot of uncertainty. Parameterisation implies that spatially heterogeneous processes have to be represented by an aggregated parameter scheme (e.g. evapotranspiration or cloud formation). Opinions diverge regarding the question how much uncertainty is inherent in the complexity of the climate system itself and how much predictability can be enhanced by advances in scientific knowledge and computational capacities. It is well understood that the chaotic behaviour of the atmosphere puts an intrinsic limit to weather forecasting which is about 10 to 14 days. Whether the chaotic behaviour of the oceans imposes a similar limit on climate forecasting, *albeit* on a longer time scale, is still a matter of dispute (Pahl-Wostl et al 1998). In social simulation modelling, representations of cognition are based on intelligent planning and diagnostic formalisms (typically KD45 logics) or on implementations of theories of cognition which are themselves well verified in psychological experiments (Moss *et al.*, 1997; Moss, 1998). The range of widely accepted and well verified relations available for use in social simulation is much smaller and perhaps rather less convincing than the relations used in physical modelling. Nonetheless, when the physical models are applied to complex environments and used to simulate phenomena occurring over centuries of historical time, the interactions taking place within the models and the phenomena that emerge are not in any obvious sense harder or more reliable than the phenomena that emerge in social simulation models.

We take it that the function of physical models of climate change is to inform discussions of the possible consequences of intended or inadvertent influences of humanity on the natural environment. This is also the function of social and socio-economic (but not economists') simulations. Sociologists, anthropologists, computational organisation theorists develop simulations to test the consistency and plausibility of relationships postulated by theorists and practitioners. Sometimes, these relationships are of very long standing indeed. The purpose of the simulations is better to understand or to give additional credence to postulated relationships. In addressing specific policy issues, the point is not to predict the outcomes of alternative government measures in a quantitative fashion but to heighten awareness of possible consequences or, indeed, the difficulty of anticipating consequences with any precision. The function of policy simulations in particular is to help policy analysts formulate expectations and to consider the procedures to be adopted when the unanticipated happens. Both in social simulation and in physical simulation modelling of complex, long-term interactions, the aim should be to integrate well validated, independently verified relationships into a framework which informs policy analysis and debate but which does not presume to predict.

2. The integrated assessment experience and the interface between physical and socio-economic modelling

The interface between physical and the social sciences in general and between physical and socio-economic modelling in particular is shaped by the dominance of the prevailing economist's view on how the decision problem should be framed. In this view dealing with climate change is a cost-benefit problem. The costs of measures for preventing climate change to happen have to be compared to the benefits of preventing potential damages from climate change. Physical models provide climate scenarios, economic models serve to quantify the costs and benefits for reducing greenhouse gas emissions (GHGs).

In the field of integrated assessment one attempts to integrate physical and economic aspects within one modelling framework to be able to provide more meaningful information to decision makers. Integrated assessment models range from highly aggregated models such as the DICE model of Nordhaus (Nordhaus, 1994) to process based models where processes from climate to ecosystem change to the response of humankind are addressed in a detailed fashion (Rotmans and Dowlatabadi, 1998). The DICE model is a dynamic optimisation model for estimating the optimal path of reductions in GHGs. An aggregated global welfare function is optimised where choice is limited to consuming goods and services, to investing in productive capital or to slowing climate change. Such approaches are not always met with unanimous approval. Among economists there is for example a dispute over the appropriate level of the discount rate and to what extent cost-benefit considerations are useful for informing political debate. Some people try to include agent based models in political response - e.g. to model an adaptive climate in order to determine the optimal carbon tax. However, there is a notable absence of alternative approaches which try to tackle the problem by starting from the realisation that we simply do not understand the relationships involved or their consequences.

In essence the interface between physical and social modelling has so far rested on something like a damage function which either entered a cost benefit analysis or served as a target to measure the effectiveness of response strategies. Current approaches, following standard economic modelling practice, imply greater predictiveness than experience shows to be warranted - even (or particularly) with economic models on their own. These models do not allow for new behavioural patterns and social processes to emerge. The modelling technology which supports such emergence is not easily reconciled with the equilibrium models of economists.

One could approach the problem from a different perspective and model collective learning processes. This approach has already been used by the Manchester Centre for Policy Modelling in models of transition economies (Moss and Kuznetsova, 1995; Moss, 1999), organisational change (Moss, 1998) and consumer behaviour (Moss and Edmonds, 1997; Edmonds and Moss, 1998). It has also been investigated by EAWAG using focus groups and in emphasising business opportunities. Before going into further details of an alternative modelling approach we devote some critical thoughts to current approaches in economic modelling.

3. Does specification matter?

Since economists typically apply a cost-benefit analysis to climate change, it seems reasonable to apply a cost-benefit analysis to their own approach. Our particular concern is whether is the conditions in which it is rational to apply standard economic (or indeed any other) models to analyse climate change in particular. The argument is taken from Moss (1992).

It is by no means inconsistent with conventional economic reasoning to assume that a model should be used for policy analysis in a manner which maximises the policy analyst's subjective expectation of policy benefit net of all costs associated with the analysis and policy implementation. An economist who satisfies his own definition of rationality will maximise the benefit of the policy actions less any costs of implementing the policy or any costs of identifying whether or not the conditions of application are satisfied. Thus we shall say that a particular policy is implied by a model whenever the model is the best available and, at least, is no worse than any previous or current policy model.

To consider the determination of this subjective expectation, we define a policy as a set of individual actions *P*. We suppose that any model used to generate policy recommendations has a set of conditions of application *C*. Since *C* could be the null set, this supposition includes the possibility that no conditions of application have been specified. There will be *n* such conditions C_i where n is a non-negative integer and $C_i \in [true, false]$.

Let *B* be the image of the mapping $[(P \mid C) \rightarrow R]$, the value of the benefits expected from the set of policy actions in *P* given that a set of conditions *C* are satisfied.

The "observation tag" for the *i*th condition is $\phi_i \in [true, false]$ which takes the value true if it is intended to observe the *i*th condition and *false* otherwise. The intention of the policy analyst to observe conditions of application is captured by the set $\Phi = \{\phi_i \mid (i = 1 \ldots n)\} \cap true$.

In addition, we denote by $C(\Phi)$ the cost of observing all of the conditions $\phi_i \in \Phi$.

To complete our notation we require some means of representing degrees of prior belief in the satisfaction of the conditions of application which it is intended to observe. The standard representation is in terms of subjective probabilities. For this reason, we adopt the mapping $\Psi(\Phi) \rightarrow [0,1]$ which we interpret as the subjective probability that all conditions of application in Φ will be found to be satisfied.

By hypothesis, if all of the conditions of application of the theory are true, then the acts in *P* will imply some expected benefit E(*B* | *C*). Otherwise, some different benefit E(*B* | ¬*C*) will result. Since the benefit will be net of the cost of ascertaining whether conditions of application are satisfied, we define the benefit as $B = \mathrm{B}(\Phi)$.

Evidently, the prior expected benefit of *P* when the set of conditions to be observed is empty is

(1) $\mathrm{E}(B \mid \Phi = 0) = \mathrm{E}(B \mid C) \cdot \mathrm{E}(C) + \mathrm{E}(B \mid \neg C) \cdot (1 - \mathrm{E}(C))$

More generally, the expected benefit given any arbitrary set of conditions of application to be observed will be

(2) $$\begin{aligned}\mathrm{E}(B \mid \Phi) = \Psi(\Phi) \cdot \{&\mathrm{E}(C \mid \Psi(\Phi)) \cdot \mathrm{E}(B \mid C) + \\ &[1 - \mathrm{E}(C \mid \Psi(\Phi))] \cdot \mathrm{E}(C) \cdot \mathrm{E}(B \mid \neg C) \cdot [1 - \mathrm{E}(C)] - \mathrm{c}(\Phi)\} \\ &-[1 - \Psi(\Phi)] \cdot \mathrm{c}(\Phi)\end{aligned}$$

In equation (2), c(Φ) is the cost of observing the conditions in Φ. The expression $\mathrm{E}(C \mid \Psi(\Phi))$ is the expectation that all of the conditions of application are satisfied given that the individual conditions in the set Φ are known to be satisfied. Expanding and simplifying equation (2), we get

(3) $\mathrm{E}(B \mid \Phi) = \mathrm{E}(C \mid \Psi(\Phi)) \cdot \Psi(\Phi) \cdot \mathrm{E}(B \mid C) + [1 - \mathrm{E}(C \mid \Psi(\Phi))] \cdot \mathrm{E}(B \mid \neg C) - \mathrm{c}(\Phi)$

Since, from the definition of C, $\mathrm{E}(\Psi(\Phi) \mid C) = 1$

$$\mathrm{E}(C \mid \Psi(\Phi)) \cdot \Psi(\Phi) = \mathrm{E}(\Psi(\Phi) \mid C) \cdot C = C$$

where the first equality is Bayes' Law. In consequence, equation (3) can be written

(4) $\mathrm{E}(B \mid \Phi) = \mathrm{E}(B \mid C) \cdot \mathrm{E}(C) + (1 - \mathrm{E}(C)) \cdot \mathrm{E}(B \mid \neg C) - (1 - \Psi(\Phi)) \cdot \mathrm{E}(B \mid \neg C) - \mathrm{c}(\Phi)$

Substituting into equation (4) from equation (1),

(5) $\mathrm{E}(B \mid \Phi) = \mathrm{E}(B \mid \Phi = 0) - (1 - \Psi(\Phi)) \cdot \mathrm{E}(B \mid \neg C) - \mathrm{c}(\Phi)$

The interpretation of equation 5 is that, taking the case where no conditions of application are verified as the base case, the expected benefit is reduced by the expectation of benefit when the conditions of application are not satisfied and by the cost of ascertaining whether those conditions are satisfied. If this is true for every possible combination of conditions of application, then it is rational never to test for model applicability.

Evidently, the rational modeller who accepts the implications of economic theory for rational behaviour and cost-benefit analysis will investigate the conditions of application of the model only when

(6) $(1 - \Psi(\Phi)) \cdot \mathrm{E}(B \mid \neg C) \geq \mathrm{c}(\Phi)$

indicating that the expected benefit of the policy implied by an inapplicable model is negative to an extent which is greater in magnitude the costs of determining whether the model is applicable. This result accords with economic reasoning.

In short, if the costs of adopting policies based on an inapplicable model is less than the cost of determining applicability, do not determine whether the conditions of application hold. If the cost of the inapplicable policy exceeds the cost of assessing the conditions of application, then that assessment is indicated.

There might well be many cases in which condition (6) is not satisfied and, so, the conditions of application would be too costly to evaluate in light of the expected benefit of the evaluation. It is hard to judge just how widespread this phenomenon might be in economics since conditions of application are not in practice taken into account. When it comes to climate modelling, however, the assessments of the costs and benefits of different policies vary with the model used. This is an issue which we will pick up presently in some detail. We first demonstrate that some conditions of application of economic theory in general and the climate models in particular have in important cases been deduced from the internal properties of the theoretical core of economics.

Following Sanstad and Greening (1998), we note that three types of economic modelling approaches are applied to climate modelling:

- neo-classical general equilibrium theory
- neo-classical growth theory based on aggregate production functions
- large-scale energy-sector models

All of these approaches also rely on some measure of social welfare represented as a social welfare function which takes as inputs the utility functions of individual agents in the economy. Both general equilibrium and neo-classical growth theory have properties proved in the 1960s to be empirically untenable. In addition, social welfare judgements were shown to be impossible to make unless the economic system is in a full-blown general equilibrium. We consider these in turn.

3.1 General equilibrium and the Radner theorem

The general equilibrium model which is most fundamental is the Arrow-Debreu (1954) model which is in some ways like a set of organised commodity markets in the sense that contracts are bought and sold for the purchases and sales of specified amounts of specified commodities at specified dates. These contracts are contingent on certain specified states of the world prevailing on the contract dates. Arrow and Debreu proved that there are a set of sufficient conditions for the existence of an equilibrium in which supplies of every one of these contingent contracts equal demands for them at non-negative prices. The sufficient conditions included, for example, the convexity of all utility and production functions and that all trades for every date are concluded at one time. Production function convexity precludes increasing returns to the scale of production in any industry which is itself empirically untenable. Even if increasing output scales will eventually entail reductions in outputs per unit of every input, it is not always true at realised production scales.

Much of the literature following from the Arrow-Debreu model was dedicated to finding alternative specifications of the sufficient conditions for the existence of equilibrium which could be interpreted in a more realistic fashion. A good, though by no means exhaustive, example of this work is Arrow and Hahn (1972). None of this work, with the exception of Radner's was concerned with necessary conditions for the existence of equilibrium.

Radner (1968) considered the effects of assuming that new transactions in these contingent contracts could take place at each date. He demonstrated that the amount of information required for each agent to calculate its optimal supplies and demands in relation to each of these contracts would increase over time as agents had more information about other agents' preferences and production functions. The rate of increase was exponential with time. In consequence, the computational capacity of every agent would have to be sufficient to calculate the initial general equilibrium and then grow exponentially without limit over time. This is the only known necessary condition we have been able to find for the existence of Arrow-Debreu (1954) general equilibrium.

There is a special case in which the Radner theory is not apposite. That is the general equilibrium model in which all agents are identical (or sometimes all households are identical to one another and, separately, all firms are identical to one another). Consequently, since each agent knows its own preferences it knows everyone's preferences and there is nothing to learn about them. This assumption is usually justified by the suggestion that the agent is somehow "representative" of all agents but its analytical purpose is specifically to get round the problem that, without that assumption, general equilibrium cannot exist in the absence of unlimited computational capacity for all agents.

3.2 Social welfare and the Lipsey-Lancaster theorem

Welfare judgements in a general equilibrium framework are based on the assumption of a social welfare function which takes as inputs the values of the utility functions of every agent in the economy. General equilibrium is itself known to be Pareto efficient which means that it is not possible to increase the utility of any agent in the economy without reducing the utility of at least one other agent. Indeed, general equilibrium is necessary and sufficient for Pareto. If there is any disequilibrium in the system (because supplies and demands are not equal for all contracts), then it is possible to increase the utility of at least one agent without reducing the utility of any other agent. The problem is that Lipsey and Lancaster (1956) showed that the effect on utilities of any policy measure in the form of taxes or subsidies will have an unknown effect on social welfare. Such a measure might improve welfare or reduce it but whichever transpires cannot be determined from general equilibrium theory. The only exception is the lump-sum tax which is welfare-neutral.

If there are more than two commodities and two traders in the economy, then there will be more than one general equilibrium. It is possible in principle to compare the values of a social welfare function for each general equilibrium configuration of the economy. However, the existence of general equilibria are always proved by demonstrated that a set of conditions support the application of a fixed-point theorem - usually the Kakutani fixed-point theorem. This is the essence of demonstrating the sufficiency of the conditions for general equilibrium to exist. Of course, fixed-point theorems can be used to identify the properties of a point on a topological surface but they cannot be used to identify the properties of any process. In general, there is no reason to believe that general equilibria are either locally or globally stable. Indeed, we know nothing about processes in a system satisfying any of the sufficient conditions for general equilibrium. What Lipsey and Lancaster showed was that among the things we do not know is the effect of such processes on the value of a social welfare function.

3.3 Neo-classical growth theory and the aggregate production function

Neo-classical growth theory is an alternative to general equilibrium theory and social welfare functions in economic analyses of climate policy issues. This theory rests on the assumption that there is a function relating aggregate inputs of capital and labour to some aggregate output. Shifts in that function are a consequence of technological change and the education and training of labour. This approach was first suggested by Robert Solow (1957) though it was immediately pointed out by Hogan (1958) that the econometric estimation of production functions actually estimates the distribution of income between property and employment income (the functional distribution of income). Sheikh (1974) demonstrated by simulation and algebraically that the estimation of the production function typically used in these endogenous growth studies (the Cobb-Douglas function) gives a better fit the more constant is the functional distribution of income. Indeed, it fits perfectly if and only if the functional distribution of income is constant. The measures of the effects of technological change and education and training are the residuals from the estimating equation. Consequently, the representation of technological change in these models is actually determined by the deviation of the actual income distribution from constancy. It has no direct relationship to technological change except to the extent that such change influences the distribution of income.

Solow's (1974) defence against Shaikh's argument was this:

> Mr. Shaikh's article is based on misconception pure and simple. The factor-share device of my 1957 article is in no sense a *test* of aggregate production functions or of anything else. It merely shows how one goes about interpreting given time series if one starts by *assuming* that they were generated from a production function and that the competitive marginal-product relations apply. Therefore, it is not only not surprising but it is exactly the point that if the observe factor shares were exactly constant the method would yield an exact Cobb-Douglas and tuck everything else into the shift factor. That is what one would *want* such a method to do.

Remember that the device being discussed here is used in most economic models of the impacts of climate change. What Solow asserted is that, assuming some production function and the associated theoretical propositions of economic theory to have generated the observed data, then the Cobb-Douglas function will be the particular function which generated the data. Otherwise, the goodness of fit of the Cobb-Douglas function is no more than a measure of the constancy of factor shares over the data period.

So a rational economic analyst would consider whether there are any conditions in which the Cobb-Douglas function - or perhaps some other function of its type - could generate that data.

The conditions of application of the Cobb-Douglas and similar functions was drawn from the theory itself during the course of an extensive controversy in the 1960s chronicled in detail by Harcourt (1972). Apart from certain curious examples of steady-state equilibrium, a function of the Cobb-Douglas type would describe the economic processes involved only if labour unassisted by any capital equipment were to use some malleable, homogenous substance to produce more of the same substance. Now the statistical properties of the technical change measures estimated by endogenous growth theorists preclude the existence of a steady-growth equilibrium. Consequently, the conditions of application of this economic theory are restricted to the case where a single malleable, homogenous good is the only produced input to its own production. We doubt that even the most rabidly orthodox economist would care to argue that that condition is ever likely to be satisfied. The cost of determining that it is not satisfied is small since it requires nothing more than counting to 2.

We conclude that the theoretical foundations underpinning current approaches to economic modelling of climate change are inappropriate for the type of questions that being addressed. Since these economic models draw their legitimacy largely from being based on established theory, this is a serious problem. The theoretical foundations of climate models seem to be more solid. Who would argue against the validity of the thermodynamic laws stating the conservation of energy and/or the increase of entropy? However, it has been questioned whether a model's being grounded in theoretical foundations is sufficient for establishing a high status for certain modelling approaches (Shackley et al in Climate Change 1998, Oreskes et al in Science 1994). Empirical terms (e.g. parameterisation) and corrections (e.g. flux adjustments) are essential for any theoretically derived climate model to produce plausible results. If such models are applied outside of their range of validation the confidence in model simulations is reduced. Sometimes one hears a claim for simpler models with a complex dynamics present only to a limited extent in current GCMs (e.g. chaotic fluctuations, positive feedback and abrupt transitions). Such claims are based on the argument that models should provide insights rather than detailed predictions. The possibility of certain changes (e.g. the increase in extreme events) should be highlighted and discussed. Interestingly, this is a major difference from economic models applied to integrated assessment which are simple and highly aggregated with extremely simple dynamics.

4. Implications for the methodology of integrated assessment of climate change

Conventional economic modelling is based on a top-down modelling style. General equilibrium models impose a structure on the modelled system in which there is a market for every good and there is a price in every market that equates the supply of the good to the demand for it. In the aggregated models, there is a production function relating inputs to the output with the assumption that input prices are equal to the values of their marginal products (*i.e.* the additional output valued at market price from an additional unit of input with all other inputs held constant). The first specification of this model was by John Bates Clark who argued in his 1893 *Distribution of Wealth* that because the wage rate would be equal to the value of output contributed by the marginal worker, there was no exploitation of labour. Whatever the reason, economics rests on prior characterisations of the properties of economic states. Among mainstream economists, there is no concern for the processes which might give rise to such states. When a process is described in relation to the general equilibrium model, it is the existence of an auctioneer to receive notice of the supplies and demands for each commodity and by some iterative process finds a set of prices that will clear all markets simultaneously. Recently, Binmore *et. al.* (1997) justified an evolutionary process by appeal to the auctioneer of general equilibrium as an analogy for an unmodelled process by means of which supplies and demands eventually become equal.

While the process is typically ignored, that is not the universal case. Nelson and Winter (1982), for example, argued for the existence of a "technological trajectory" on which there was a slow start during which inventors or innovators learned how to use a new technology and then, as they came to understand it better, there would be a period of rapid innovation and technological improvement followed by a flattening of the trajectory as the promise of the technology was realised and there was increasingly little productivity improvement left to be gained. This, too, of course, is imposing the structure of what is to be represented on the model.

All of the economic models for climate policy advice surveyed by Sanstad and Greening (1998) use aggregate production functions which, we saw in section 3, have been known for more than 30 years to be empirically untenable. They all assume that consumers maximise utility which means that they know all of the goods (if more than one) which will be available for purchase over the indefinite future. The utility function used is usually Cobb-Douglas which implies that the share of expenditure on different products (for example, energy and food) is constant - an assumption which has been known to be false for 150 years and more. Several of the models assume rational expectations which means that all individuals have the *correct* model of the economic system. In other spheres of mainstream economics (the classic work being Sargent, 1993), agents are allowed to learn by generating and testing models. This is represented by genetic algorithms which specify learning as a global search over *all possible relations* in order to focus increasingly on the best relations as guides to spending and production decisions. In all of this work, however, the test of the success of a model is how closely it converges to a rational expectations equilibrium. Practitioners in the field are well aware of the fact that the models' inability to deal with the issue of induced technological change is a major draw back (Sanstad and Greening, Goulder and Schneider and further references). However, there seems to be little effort to take that conclusion as the point of departure for an entirely new modelling effort.

The problem here seems to be quite general: economic models are constructed on the basis of a specification of individual behaviour that conforms to the imposed, usually equilibrium, structure. As a result the model of the agent is very simple. Agents are social atoms without the ability to learn from their own set of experiences. All of their knowledge is explicit and their procedural knowledge is unchanging. Communication within such models is always limited to a broadcasting of prices or quantities of outputs. We know of no economic models where agents communicate in a social network in a meaningful fashion. At the current leading edge of economics, agents are modelled as participating in round-robin tournaments wherein the play a prisoners' dilemma game at each stage.

This simplifying approach allowed economics to link microfoundations to macrobehaviour. It is largely the physicists approach chosen in particle physics and statistical thermodynamics. Whenever one has a large number of uniform particles it is possible to use the law of large numbers. One of the major achievements in theoretical physics is the derivation of the macroscopic laws of thermodynamics from their microscopic base. The macroscopic laws were a well established empirical fact prior to the work on the microfoundations. Economics is not in an analogous position in that, apart from a few apparently trendless time series, there are no well verified, invariant macroscopic relations. Consequently, the approaches of equilibrium thermodynamics and Newtonian physics are inappropriate for economics. It is amazing that much of economic theory that was derived around the turn of the century survived all the empirical challenges from within the discipline and the challenges from complex system studies. It may not be by chance that metaphors for innovative approaches make a closer link to biology and ecology than to physics. What is required is an understanding of evolution, function, communication and diversity, all concepts of major importance in biological systems.

In order to achieve such an understanding that will be convincing to the integrated assessment community at large, we require to develop an analysis of social behaviour, including its economic aspects, that has some empirical verification and qualitative plausibility. This is a pragmatic issue. If, instead of assuming that unobserved relationships are always correct, we look to externally verified elements for our models and then assess the outputs from our models in terms of goodness of correspondence to both available statistical records and the qualitative historical record as assessed by domain experts, we can then use those externally verified elements of the models to define the models' conditions of application.

Prediction of the course of climate change and social relations and behaviour is, we have suggested, beyond our capabilities for dealing with the complex and poorly understood processes underlying these phenomena. We have also focussed on the models and modelling techniques used and developed by Nordhaus as a representative of the view of economists such as Solow that theories are not tested by their assumptions. We have shown that results obtained by economists concerning economic theories and models imply a number of clear features of agent cognition (Radner, 1968), the effects of economic policies on social welfare (Lancaster and Lipsey) and the nature of production (chronicled by Harcourt, 1972). None of these results are obscure and their implications are well understood. We therefore conclude that the descriptions of the world implied by these results are not taken to be conditions of application of the respective theoretical constructs. Indeed, we see no interpretation of Nordhaus' implicit approach to modelling the effects of climate change other than that inequality (6) always holds when no conditions of application are investigated. Either this is because the benefits of the policies implied by the economists' contribution to integrated assessment are assumed always to be beneficial or because the approach is assumed to be applicable in all circumstances.

We take a more modest approach to conditions of application. We agree that anything that might reasonably be called a condition of application for economists' approaches to integrated assessment is sufficiently implausible and remote from the circumstances we observe that these approaches are unlikely to be useful in the suggested methodological regime. Other approaches taken by computational organisation theorists such as Carley and her colleagues (Ye and Carley, 1995; Carley and Prietula, 1998) or by the Centre for Policy Modelling (Moss, et. al., 1997; Moss, et. al, 1997; Moss, 1998) or at EAWAG (Kottonau and Pahl-Wostl, in prep; Tilman et al in press) demonstrate that simulation models can usefully entail representations of agent cognition or institutional developments that have clear empirical referents. Such components of simulation models are naturally taken to be conditions of application.

Physical modellers are concerned with the conditions of application of their models. Conditions of application refer often to the empirical relationships that have to be included. Confidence is generally enhanced by a model's ability to reproduce empirical data. As in economics, physical modellers use part of a data series for fitting model parameters and another part of the data series to test model predictions. That is not without drawbacks. Often models have a higher resolution than the available data. It is a saying that with enough parameters one is able to fit an elephant. The problem of overparameterisation emerges. More recently in particular with the increasing importance of models as tools for producing policy advice new criteria become important. Do models produce results that are useful for discussing a topic in public debate? Are assumptions and uncertainties communicated in sufficient detail?

But what then about prediction? We do not propose to predict outcomes. Our modelling rationale is the development of tools for counterfactual experiments and for what-if analyses to inform and help focus discussions about policy measures and also to help identify social and perhaps physical processes that analysts might not otherwise have considered.

We certainly insist that our models have clear, empirical conditions of application *and* yield outputs that can correspond to relevant observations. So we start our analyses from a plausible account of the important relationships in social responses to climate change. By plausible, we mean that observed qualitative conditions and statistical descriptions put into the models yield outputs which also correspond to observed qualitative features and statistical descriptions of the phenomena of concern. An important feature of such models is that they capture representations of changing social relations. Such changing relations would include institutional changes in exchange, changing organisational structures, the development of new mental models by agents and how these affect policy assessments.

4.1 A new approach to integrating social and physical modelling

A new modelling approach should account for the different ways of the interaction of agents with their social and natural environment. In current approaches the interface between social and physical modelling is mainly given by the price mechanism. Potential damage of climate change serves as input into economic growth models. In processes of social learning environmental awareness and the formation of values are important. For strategic planning one has to take into account the formation of expectations that may be informed by results from climate forecasts and expected policy measures. This implies that one has to account for the flow of information other than market prices. Rather than building large fully integrated physical-social models we will investigate the effects of different types of scientific information. We will represent the influence of information, the attitude towards and the perception of risks, different levels of individual and social values, the expectations of a positive future etc. How can one build trust in models that cannot be validated by experience? What is the importance of uncertainties for the processes of policy formation?

Our approach to social modelling is pragmatically reductionist or, equivalently, pragmatically holist (Edmonds, 1998). We intend to be sufficiently reductionist that we specify individual agents whose behaviour can be represented in a manner that is either independently verifiable or that is validated relative to some well verified theory that is itself independent of our own models. The correspondence between these representations and their empirical referents constitute conditions of application for our models. Representations of some agents will be based in established, verified theories of cognition. The agents represented in this way are those engaged in strategic behaviour including planning, generating and modifying social policies, guiding the process of technological and institutional innovation, determining the scale and direction of investment, *etc.* Other agents can be represented more simply without loss of descriptive accuracy. We have consumers in mind here particularly. We will put a major emphasis on the social embedding of individual action. We consider changes at the level of procedural knowledge (rules) as the major driving force for the evolutionary socio-economic change required for sustainable development.

We cannot say in advance what the necessary degree of reductionism must be - that is, how fine grained must be our representations of agents. Initially, we represent enterprises as engaging in activities determined by actors represented as problem space architectures, who learn by generating, testing and evolving models of their environments and other agents, converting these models into rules of behaviour so that declarative knowledge (the models) become procedural knowledge (the rules) in accordance with Newell's unified theory of cognition (implemented as Soar) or Anderson's theory of memory (implemented as Act or more recently ACT-R). Our representation draws from both Soar and ACT-R where the elements of each are compatible and experimentally well verified. Examples of the implementation of these representations is to be found in recent papers by Moss (1998, 1999). We also, initially, represent consumers as genetic programming algorithms which build up consumption patterns taking account of the influences of other consumers, new products and educational programmes. Several models in this vein have been published by Edmonds (1998a, 1998b). Another approach will involve the development of models for consumer behaviour based on goal networks and means-end chain theory that is empirically well grounded in marketing research (Kottonau and Pahl-Wostl, in prep). This approach focuses in particular on the mutual relationship between collective and individual learning.

We will investigate means of aggregating over these agents in order to reduce computational expense without loss of verifiability and, correspondingly, accuracy of our representations with respect to relevant observations. In addition, we will represent markets as emergent trading relationships and practices. We will not assume market structures, degrees of competition or the effects of competition on the abilities or inclinations of individual agents to set prices or determine sales volumes. In this, we follow Marshall (1893 [1961, pp.323-330), Kaldor (1939 [1961]) and Moss (1980).

The theoretical analysis of major innovations turns on the relationship between the technology of exchange and the institutions that support exchange. For example, ships are bought and sold under a completely different set of arrangements than are chocolate bars. The reasons are that ships are generally not well standardised and they are expensive to store. Consequently, building ships for stock to sell to any customer that comes by would imply a long time lag between production and sale with a rapidly increasing price to cover storage and financing changes. There is less risk to the shipyards and less expense to the purchasers if ships are built to order. Because of the high costs of maintaining shipyards, the yards maintain order books so that their work is planned out for several years ahead. The order book enables the shipyards to be fully utilised in the face of fluctuating demands for ships. Chocolate bars, on the other hand, are cheap to store and they are (partly as a result of branding) highly standardised and they are quite durable. As a result, they can be sold to passing customers whose identity is not important and they can cheaply be held in stock so that the inventory fluctuations. The basis of this difference is the technologies involved in storage, transportation and communication.

Technological changes in these activities can radically change the nature of exchange. In the last century, refrigeration replaced the prevailing system of selling live animals to local butchers with the creation of huge slaughtering factories and the transport of chilled meat to the local butchers. This and many similar examples have been reported by Porter and Livesay (1971) The effects of the internet and electronic trading could -- indeed should -- be analysed in these terms.

It is plausible that an important factor in the scale and pattern of energy use will involve changes in the practices of exchange. In order to investigate the importance of this possibility, we propose to model the relationships between trading patterns and practices on the one hand and the technologies of storage, transportation and communication (which collectively comprise the technology of exchange) on the other. A particular issue to be investigated at an early stage in the project will be whether an effect of a carbon tax or a tax (say) on diesel fuel would make transportation of some goods so expensive as to encourage local production and markets with consequent effects on economies of scale in production. Can electronic control and robotics reduce minimum efficient scales in production thereby to make it possible to increase transportation costs by some kind of tax and, as a result, encourage local production without loss of scale economies? If the degree of standardisation of products were unaffected (because electronic control programs can be shared via the Internet), would there be a savings in the combined costs of exchange and production without adversely affecting consumers? What kinds of market arrangements might we expect to emerge from such taxes? Would it be possible by public provision of infrastructure to facilitate such changes? And, of course, how important might be the influence of such changes on the course of climate change?

In order to bring these representations together within a coherent framework that also supports well validated and verified representations of environmental relations and determinants of climate, we intend to build a number of models of small scale phenomena such as the introduction of community central heating in Copenhagen, the management of critical incidents and crises in environmentally sensitive organisations, the effects of particular technological changes in urban water management on resource utilisation and adaptability in times of increased uncertainty due to climate change, the diffusion of ecological innovations and consumer attitudes in Zürich.

4.2 Future developments

We can imagine a range of different approaches which would meet the criteria of our methodological regime in application to climate change.

The issue of scale in both time and space does not raise questions only about the use of physical models but also about how to combine a physical climate model and a social model . Decision making is local and short term. Climate change is in the end a global phenomenon and long term. Bringing these two scales together to achieve meaningful results requires much conceptual work. We offer a few suggestions for an initial set of models integrating representations of learning processes with established models of climate. These would provide both substantive output and a focus for developing the required modelling technology. Examples of applications are:

- How the effects of adaptability and emergent technology on a regional scale diffuse globally.
- The consequences of different speeds of response associated with different dynamics of the physical/ecological systems and how these might influence technological change related to changing (or not) patterns of energy use under different policy regimes.
- Analysis of the principle of robust action where one should choose short term decisions such that long term degrees of freedom are maintained.
- Modelling the process of knowledge generation about the climate system, the way this knowledge affects people's belief about climate and climate change. We believe such models must take into account that uncertainties about both the physical and the social processes make it possible (even likely) that individuals will hold contradictory beliefs.
- Models which allow that qualitative aspects are important for consumer choice and where shared environmental awareness influences decision making. How might expectations of future developments determine investment and life-style decisions which would support functional responses to the threat of unfavourable climate change?

In general terms, we are proposing a modelling methodology and technology to support the exploration of different global scenarios where one has different response strategies and different scenarios of change in the climate system. Our approach is based on a view of socio-economic systems where the notion of an equilibrium state does not make sense.

The legitimisation of a modelling approach for policy advice derives not only from internal criteria within science, in particular not from being based on theoretical foundations that may be subject to dispute. A model must produce plausible results regarding empirical relationships. And it must be plausible for the non-expert audience - extended peer community and an embedding of the whole modelling process into a social process with a dialogue with non-scientific experts and/or citizens being concerned as consumers and decision-makers. It is essential in this process to included local domain knowledge and the subjective assessments of the people concerned (Pahl-Wostl et al 1998). The modelling approach presented in this paper provides an excellent framework for taking such requirements into account.

FOOTNOTE:

It is helpful to distinguish between validation and verification as used in the context of this paper. A well validated model is one which is demonstrated to be sound and consistent relative to some formal framework. We are requiring our models to be validated relative to a formal framework which is itself verified in the sense that it is well grounded in and supported by empirical data including the qualitative assessments of domain experts.

REFERENCES

Anderson, J. R. (1993), *Rules of the Mind* (Hillside NJ: Lawrence Erlbaum Associates).

Arrow, K.J. and G. Debreu (1954), "Existence of an Equilibrium for a Competitive Economy", *Econometrica*, pp. 265-90.

Arrow, K.J. and F.H. Hahn (1972), *General Competitive Analysis* (Edinburgh: Oliver and Boyd).

Clark, John Bates (1893) *The Distribution of Wealth* (New York: Macmillan).

Harcourt, G.C. (1972), *Some Cambridge Controversies in the Theory of Capital* (Cambridge: The University Press).

Hogan, Warren P. (1958), "Technical Progress and Production Functions", *The Review of Economics and Statistics*, 40:4, pp. 407-411.

Kaldor, Nicholas (1939 [1960]), "Speculation and Economic Stability", *Review of Economic Studies*, [reprinted in Kaldor, *Essays on Economic Stability and Growth* (London: Gerald Duckworth & Co., pp. 17-58.

Lipsey, R.G. and K. J. Lancaster (1956), "The General Theory of the Second Best", *Review of Economic Studies* 24: 11-32.

Marshall (1893[1961]), *Principles of Economics*, (London: Macmillan); [8th (variorum edition) published in 1961].

Moss, Scott(1980), *An Economic Theory of Business Strategy* (Oxford: Basil Blackwell).

Moss, Scott, Helen Gaylard, Steve Wallis and Bruce Edmonds (1998), "SDML: A Multi-Agent Language for Organisational Modelling, *Computational and Mathematical Organization Theory*, 4:1, pp. 43-69.

Moss, Scott (1992) "The Economics of Positive Methodology" in R. Blackwell, J. Chatha and E. Nell (eds), *Economics as Worldly Philosophy* (New York: Macmillan).

Moss, Scott (1998), "Critical Incident Management: An Empirically Derived Computational Model", *Journal of Artificial Societies and Social Simulation*, 1:4, http://www.soc.surrey.ac.uk/JASSS/1/4/1.html.

Moss, Scott (1999), "A Cognitively Rich Methodology for Modelling Emergent Socioeconomic Phenomena" in Thomas Brenner (ed.), *Computational Techniques for Modelling Learning in Economics* (Amsterdam: Kluwer Academic Publishers).

Newell, Alan (1990), *Unified Theories of Cognition* (Cambridge MA: Harvard University Press)

Nordhaus, W., 1994, Managing the Global Commons. The Economics of Climate Change, (Cambridge MA: MIT Press).

Oreskes, Shrader -Frecheette, Belitz, 1994. Verification. Validation and Conformation of Numerical Models in the Earth Sciences. Science, 263: 641-646

Pahl-Wostl, C.: 1996, "Integrated Assessment of Regional Climate Change and a new Role for Computer Models at the Interface between Science and Society", in Sors, A., Liberatore, A., Funtowicz, S., Hourcade, J.C., Fellous, J.L., Prospects for Integrated Environmental Assessment: Lessons learnt from the case of climate change, European Commission, DG XII, Toulouse, pp. 156-160.

Pahl-Wostl, C., et al.: 1998, "Integrated Assessment of Climate Change and the Problem of Indeterminacy", in Cebon, P., Dahinden, U., Davies, H.C., Imboden, D.M., and Jaeger, C.C., Views from the Alps: Regional Perspectives on Climate Change, The MIT Press, Cambridge, MA, pp. 435-498.

Porter, Glenn and Harold Livesay (1971), *Merchants and Manufacturers* (Baltimore MD: Johns Hopkins Press).

Radner, Roy (1968), "Competitive Equilibrium under Uncertainty", *Econometrica* pp. 31-59.

Rotmans, J. and Dowlatabadi, H.: 1998, "Integrated Assessment of Climate Change: Evaluation of methods and strategies", in Rayner, S. and Malone. L. (eds.), Human Choice and Climate Change, Battelle Press, Columbus Ohio.

Sanstad, A.H., and Greening, L.A., 1998. Economic models for climate policy analysis: A critical discussion. *Environmental Modelling and Assessment* 3: 3-18.

Shackley, Young, Parkinson and Wynne, 1998. Uncertainty, complexity and concepts of good science in climate change modelling: are GCMs the best tools? *Climate Change*, 38: 159-205

Shaikh, Anwar (1974), "Laws of Production and Laws of Distribution: The Humbug Production Function", *The Review of Economics and Statistics* 56:1, pp. 115-120.

Solow, Robert (1957), "Technical Change and the Aggregate Production Function" *The Review of Economics and Statistics* 39, pp. 312-320.

Solow, Robert (1974), "Laws of Production and Laws of Distribution: The Humbug Production Function: A Comment", *The Review of Economics and Statistics* 56:1, p. 121.