Integrated Assessment
2002, Vol. 3, No. 1, pp. 63–77

1389-5176/02/0301-063$16.00
© Swets & Zeitlinger

Agent Based Modelling for Integrated Assessment

SCOTT MOSS

Centre for Policy Modelling, The Business School, Manchester Metropolitan University, Manchester, United Kingdom

ABSTRACT

The argument of this paper is that, in the present state of knowledge, it is not good science to base integrated assessment models on any social theory and it is not appropriate to evaluate integrated assessments in relation to conventional positivist precepts. Instead, means should be sought to produce well validated descriptions of social and natural systems. A powerful means of validation turns on the application of agent based models. The agents are self-contained computer programs that interact with one another and can be designed and implemented to describe the behaviour and modes of interaction of observed social entities (individuals or groups). Reported models and simulation experiments demonstrate a procedure for the qualitative and statistical validation of agent based integrated assessment models.

Keywords: validation, verification, statistical signature, self-organised criticality, water.

1. THE ISSUES

In the course of developing agent based social simulation techniques for application to integrated assessment, a number of issues have become foci of attention both within the European Fifth Framework RTD project on Freshwater Integrated Resource Management with Agents (FIRMA) and the the European special interest group on Agent Based Social Simulation (the ABSS SIG) which is a part of the European Network of Excellence in Agent Based Computing (AgentLink).[1] These issues are

- the nature of good science;
- what makes an analysis scientifically respectable;
- the relationship for scientific method between theory and observation;
- the importance of embedding new developments in the existing literature;
- the desirable degree of simplicity for each model.

To attempt to deal with all of these issues in general is obviously well beyond the scope of a single paper. Consequently, they will be addressed specifically in relation to integrated assessment of issues relating to climate change and related environmental issues and the value of agent based modeling in support of that assessment. This in itself is too substantial a task but it is at least possible to identify some key issues, to enunciate some principles for discussion by the integrated assessment community and to present some examples of a promising agent based approach.

Climate change issues and the importance of anthropogenesis in climate change are of extreme scientific interest (as well as social importance) because together they require an analysis of physical, biological and social systems that are each too complex to support any forecasting or prediction together with the complex interactions among those systems. No one would argue that the course of climate change with or without anthropogenic influence can be predicted over the time horizons, from 25 years to centuries ahead, that are of interest to the integrated assessment community. While there is enough good physical science to identify potential threats from human activity to the natural system on a global scale, the complexities of the relevant systems and the interactions among them prevent any convincing or widely agreed prediction of the distribution of those effects in terms of either trends or the magnitude, frequency and timing of extreme natural events such as droughts and floods, heat waves and spells of extreme cold, wind conditions including the incidence of hurricanes and typhoons and even the effects of global warming on earthquake and volcanic activity.

The threat is clearly immense and yet there is no established natural or social science to guide a usefully detailed analysis of the impacts of global climate change or the mitigation of those impacts. Indeed, the purpose of integrated assessment research is to develop just such a useful science. In order to consider the effectiveness and

Address correspondence to: Scott Moss, Centre for Policy Modelling, The Business School, Manchester Metropolitan University, Manchester, United Kingdom. Tel.: +44 (0)161 247 3886; Fax: +44 (0)161 247 6802; E-mail: S.moss@mmu.ac.uk

[1]Their respective web sites are at http://www.cpm.mmu.ac.uk/firma and http://www.cpm.mmu.ac.uk/abss

appropriate characteristics of an agent based approach to the development of such science, I start by offering a number of *obiter dicta* relating to the issues raised in the opening paragraph above.

The nature of good science. The longstanding encomium of the logical positivists that good science yields accurate predictions is clearly not relevant to issues of climate change where prediction is not remotely possible over a useful timeframe and in any conceivable state of forecasting technology. Of course, prediction relates to observation and if future observation cannot provide a measure of the goodness of a science, then we must make do with the use of current and historical observation as the basis for the formation of an understanding of some target system. There can be many purposes in seeking to understand a system. For the integrated assessment community the core purpose is to mitigate the extent and negative impacts of anthropogenic climate change and to support sustainable development of human social systems. In relation to integrated assessment, therefore, the measure of good science is the extent to which the analysis supports specific policy and strategic objectives. It follows that any analytical tools, in particular any approaches to modeling, are to be judged strictly on the basis of their usefulness in guiding social policy formation with respect to climate change and sustainable development issues.

What makes an analysis scientifically respectable? There is a widespread presumption that theory, no matter how remote from reality, is an essential aspect of respectable (as distinct from good) science. The arguments do not usually appear in published work so I simply report that the defense of use of unvalidated and unvalidatable theory in model specification turns on the role of theory in communicating with other scientists. The problem with this view is that it gives pride of place to theory over observation and policy analysis. It is by no means obvious that the adoption and possible adaptation of any particular conceptual framework such as economic theory or psychological or social psychological theory has any merit in the development of policy formation processes and policy alternatives. The scientific respectability of such theories should derive from their demonstrated capacity to support plausible accounts of observed social processes and phenomena. Where no such theories meet that criterion, scientific respectability should derive from the implementation of descriptions of specific observed social and physical processes. It may be that over the course of time commonly useful descriptive techniques are developed that not only support description but also form the basis of a comparison of different processes or processes that emerge in different conditions of social and natural systems.

The relationship for scientific method between theory and observation. In giving pride of place to observation over theory in the specification of respectable scientific method, we are also giving pride of place to validation over verification. That is, how well an analysis captures relevant aspects of the target system is more important than issues of formal consistency, soundness or decidability. Whatever the analytic approach, validation will be rendered far more robust if there is support for direct comparisons between observable entities and the analytical descriptions of those entities.

The importance of embedding new developments in the existing literature. To build on the literature of a progressive research programme that repeatedly produces replicable and well validated propositions that describe observable and observed phenomena is obviously sensible. It is by no means clear that this is an accurate description of the literature on integrated assessment. The IPCC and similar (e.g., UKCIP) approaches to scenario analysis are a good case in point. These scenarios are explicitly not intended to be forecasts. Nor do they describe any validated social process. Typically, they describe fixed behavioural patterns such as consumerism and social phenomena such as globalisation and perhaps some assumptions about available future technologies. These fixed assumptions then underlie speculations (involving some arbitrary quantifications) about future events.[2] Essentially this whole process is the specification of future outcomes based on no evidence and no description of any existing social process. The problem here is not the lack of validation of the *scenarios* but rather the lack of validation of the specification of individual behaviour and social interaction. These scenario approaches offer no understanding of how behaviour and system properties might change under the pressure of global change. And, yet, it is hard to imagine that such behavioural, social and technological changes will not be important consequences of the global changes. Certainly there is no historical precedent for such monumental individual and system stasis in the face of significant environmental change. The mainstream economics literature is similarly unhelpful. Models based on aggregate production functions, in particular the DICE [2] model and its derivatives, are untenable even on grounds of coherence within the framework of economic theory [3]. The reason is that these models rest crucially on the assumption that a single unrealistically specified commodity is produced in the economy and used both for consumption and as an input to production and that the welfare propositions of the models rest on the assumption that the economy is always and precisely in equilibrium with never the slightest deviation therefrom. In addition, the representation of technological change depends entirely on the time pattern of income distribution as between profits and wages and not in any way on technology. Appeals to game theory are no better since game theoretic models either restrict interactions among agents to two or, occasionally, three agents in reaching any decision or they offer propositions about systems in equilibrium with no propositions about processes of change [4]. Effectively, if economic models are well verified they

[2]See the IPCC Special Report on Emissions Scenarios, especially Section 4.3 online at http://www.grida.no/climate/ipcc/emission/093.htm#1 or [1].

cannot be validated. If they are not verified, it is hard to see any advantage in using such unrealistic descriptions of the social and physical systems to inform policy.

If analysis is to be embedded in an existing literature, there should be some independent reason to believe that such a literature supports the validation of the analysis. Neither the literature on scenarios nor the literature on cultural stereotypes nor the literature on economics provides such support. Unless there is some other branch of science that does provide a basis to validate analysis of the impacts and beneficial responses to global change, the integrated assessment research programme will have to develop its own core literature without appeals to any existing social science literature.

The desirable degree of simplicity. Frequently, particularly among economists (cf. [5]), assumptions are made allegedly for simplicity but in practice to restate the nature or facts of a problem in order to render it susceptible to a previously chosen analytical technique. Such assumptions for the sake of simplicity made it impossible to validate an analysis since the system being described is not the system being observed or experienced.

A different and respectable claim is that simple models are easier to understand than complicated models and there is no advantage to devising a model that is not transparently clear since the target system is itself already not transparently clear. This is not a claim to be dismissed and there is always a danger of sophistry in contesting such an obvious and attractive truth. However, one must set this advantage of simplicity against the loss of content that simplicity entails.

Clearly there is a trade-off between simplicity and content and, since neither has an unambiguous metric, there is no unambiguously optimal point on the simplicity-content trade-off. However, it may be that there are some discontinuities in the trade-off such that greater simplicity entails an unacceptable loss of content. Such is arguably the case where social systems are concerned. If the behaviour of socially embedded individuals is different from the behaviour of individuals who are not socially embedded, as argued by Granovettor [6] and demonstrated in simulation models by Edmonds [7], then any reduction in the content of a model which misspecifies or ignores social embeddedness invalidates the model with respect to systems in which individuals are in fact socially embedded.

Consequently, if we make the case that social embeddedness is an essential aspect of systems generating observed phenomena, then the desirable degree of simplicity will be not less than is required to capture social embeddedness. Within that constraint, it remains true that simpler models are easier to analyse and therefore more effective than more complex models. While the extent of social embeddedness will be seen below to be an important issue, there may well be other such issues. Consequently, it is important to develop a body of knowledge concerning appropriate degrees of simplicity for particular target systems (or types thereof) and particular purposes of analysis.

Organisation of the paper. The argument of the paper starts with the systems targeted by the integrated assessment community, indicates why simulation modeling rather than analytical or informal approaches are indicated, then deals with the importance of agent interaction and social embeddedness and finally the necessity for stakeholder participation in the modeling process.

2. DESCRIBING TARGET SYSTEMS: THE NECESSITY OF SIMULATION MODELING

The target systems of integrated assessment are natural and social systems. Data series generated by both systems are marked by occasional and unpredictable clusters of extreme events. This pattern is well documented for physical systems – especially with regard to earthquakes, avalanches, sunspot activity, quasar light intensities and many others. Time patterns of speciation and extinction have the same characteristic. In social systems, it has long been known that time series of financial asset prices in organised exchanges entail unpredictable clusters of volatility.

2.1. Leptokurtosis

One effect of such clustered volatility is that the frequency distribution of first differences is *leptokurtic* (thin-peaked) and fat tailed where both terms are relative to a normal distribution of the same mean and standard deviation. An example of such a distribution, shown in Figure 1, is the distribution of mean January temperatures over nearly two centuries in southeast England. The distribution of the actual temperatures (given by the histogram) is clearly more skewed and with a higher and thinner peak than the reference normal distribution. Non-parametric tests of the actual distribution for normality (Table 1) indicate a significant

Table 1. Tests of normality of average monthly temperatures in southeast England, 1816–1998.

	Kolmogorov–Smirnov statistic	*df*	Sig.	Shapiro–Wilk statistic	*df*	Sig.
Jan	.078	182	.008	.978	182	.006
Feb	.077	182	.010	.967	182	.000
Mar	.062	182	.087	.995	182	.760
Apr	.046	182	.200	.995	182	.810
May	.066	182	.054	.990	182	.253
Jun	.047	182	.200	.991	182	.304
Jul	.061	182	.094	.988	182	.134
Aug	.080	182	.007	.980	182	.010
Sep	.045	182	.200	.994	182	.646
Oct	.085	182	.003	.986	182	.078
Nov	.038	182	.200	.995	182	.855
Dec	.066	182	.051	.980	182	.009

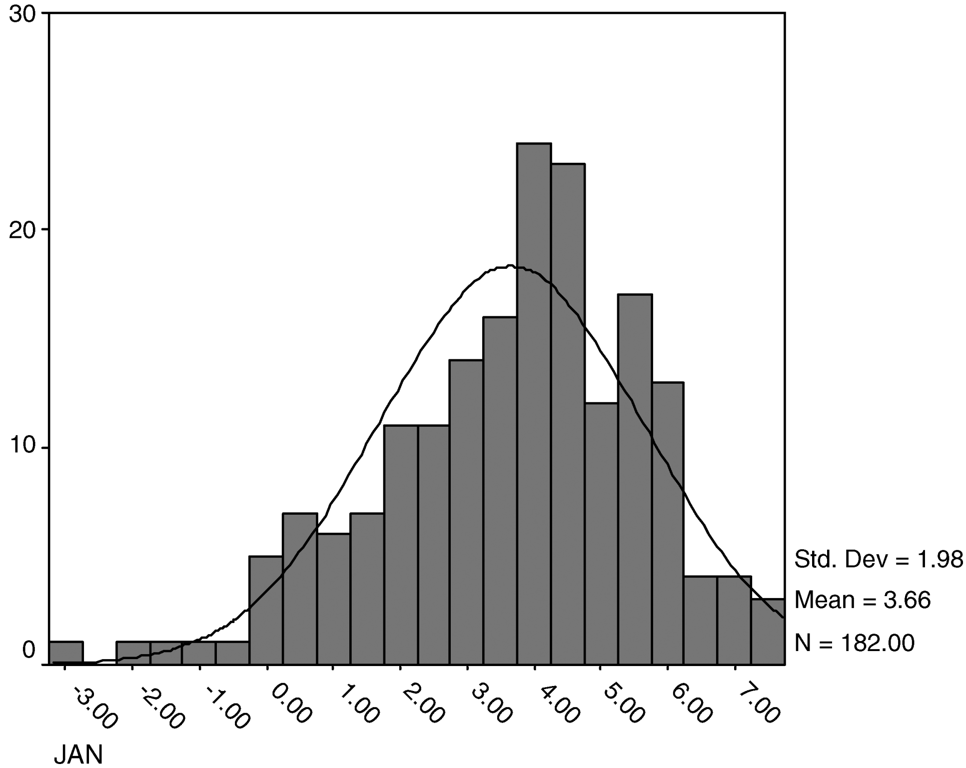

Fig. 1. Frequency histogram of average January temperatures in southeast England, 1816–1998 with normal distribution for same mean and standard deviation.

difference at the 0.008 level of confidence (Kolmogorov–Smirnov statistic) or at the 0.006 level of confidence (Shapiro–Wilk statistic).

The columns headed "Sig.," give the confidence interval for the actual distribution being insignificantly different from normal. Using the Kalmogorov–Smirnov statistic for January, the actual distribution is normal at the 0.008 confidence level (i.e. with less than 1% confidence). The months yielding the distributions closest to normal are evidently April and November, but the confidence intervals even for those months are hardly compelling (0.2 using the Kolmogorov–Smirnov statistic but a more respectable 0.81 and 0.855, respectively using the Shapiro–Wilk statistic).

One explanation of the non-normality of these frequency distributions is that the physical relationships are non-linear and lie within the chaotic range.[3] Another explanation is that the physical processes determining temperature levels are characterised by self organised criticality. Yet another, suggested by an anonymous reviewer, is that the data for winter months might be skewed by phase transitions associated with freezing. Whatever the cause of the leptokurtosis, the consequences for forecasting are clear. The consequences for modeling social elements in the system are discussed below.

[3]This explanation was suggested by Michael Bishell now in the University of Cambridge, Department of Geography.

2.2. Alternative Explanations of Leptokurtosis

If a leptokurtic distribution is stable, then the only known theoretical distribution that captures the characteristics of the such observed distributions is the stable paretian distribution [8]. The characteristic function of the stable paretian distribution takes the logarithmic form

$$\log \int_{-\infty}^{\infty} \exp(iuz)\, d\Pr(U < u) = i\delta z - \gamma |z|^{\alpha} \left[1 + i\beta \left(\frac{z}{|z|}\right) \tan\left(\frac{\alpha\pi}{2}\right)\right]$$

where α in the interval (0, 2] is a "peakedness parameter" and β in the interval $[-1, 1]$ determines skewness. Together with values of δ and γ, these parameters determine the first four moments of the distribution.

The value of α is of the most interest here. When $\alpha = 2$, this characteristic function reduces to that of the normal distribution. For all values of $\alpha < 2$, the variance of the distribution is infinite. Moreover, for values of $\alpha < 1$, the distribution mean is undefined. This means that, for all values of $\alpha < 2$, the law of large numbers does not apply to the variance and for all values of $\alpha < 1$ the law of large numbers does not apply to the mean. No laws or theorems of classical statistics or econometrics are applicable in these circumstances.

It is, however, important to note that the central limit theorem will typically apply to data with a stable paretian distribution. Since aggregating time series data – say daily into weekly data points – is effectively to calculate the mean of daily data over seven data points and then multiply by seven, samples of these sample means will be approximately normally distributed even if the underlying daily data is not. Consequently, it is possible to generate data that appears to be normally distributed simply by taking data of sufficiently low frequency. However, the variance and possibly the mean of the distribution of sample distributions will not converge to stable values.

The consequences for forecasting are enormous since, in all cases where the distribution of time series data is stable paretian but not gaussian, parametric statistical forecasting techniques are wholly otiose [9].

Economists and econometricians have responded to this problem by assuming that there is a normal distribution but that the parameters of the distribution vary over time and the values of these parameters can themselves be forecast.[4] Since there are no obviously successful forecasts of the values of these time varying parameters, there seems little reason to credit this approach.

A more promising approach is that of the statistical physicists who have been concerned to explain the observation that an extraordinarily wide range of physical phenomena that are power law distributed. The power law distribution, is:

$$N(s) \sim s^{-\tau}$$

where N is the number of observation at scale s and τ is a parameter. Mandelbrot [8] pointed out that the power law distribution is a characteristic of the stable Paretian distribution.

The question of concern to statistical physicists starting with Per Bak and his colleagues [14] was to find a process that is both very general and yields power law distributed time series. The canonical model they developed was an idealisation of a sandpile with grains of sand being continually added. The sandpile model is closely related to a cellular automaton model in that it is located on a grid with non-periodic boundaries with grains of sand added to cells at each time step. Whenever the number of grains of sand in a cell reaches some specified critical level – say 4 – there is a ‘‘toppling’’ of the sand in that cell. This toppling takes the form of a redistribution of the grains of sand in the critical cell to other (not necessarily adjacent) cells in the grid. Not all of the grains are redistributed but the number of grains in the critical cell is nonetheless reduced to 0. That some grains are lost from the system in this way makes it dissipative.

Of course adding toppled sand to the grains at other cells increases the numbers in those cells until some of them become critical and topple and so increase the number of grains in yet other cells, and so on. The consequence is that, once the system reaches a critical state, there will be a sequence of topplings involving different numbers of cells in the grid. The time series of these topplings is power law distributed.

There is a growing family of such models that yield power law distributed time series and cross sectional data. The key feature of these models is that they do not require fine tuning of the parameters of interest in order to produce data with this statistical signature. In this sense, the models self-organise into the critical state and remain in that state thereby to produce such power law distributed data with clusters of extreme events.

Jensen [15] has summarised the conditions in which self-organised criticality (SOC) emerges as those where:

- Model components (cells, agents, etc.) are metastable in the sense that they do not change their behaviour until some level of stimulus has been reached.
- Interaction among the model components is a dominant feature of the model dynamics.
- The model is a dissipative system.
- The system is slowly driven so that most components are below their threshold (or critical) states most of the time.

Whether these conditions apply to the relevant physical systems determining precipitation, temperature and the like is a matter for discussion by physicists. Arguably, however, these conditions are good descriptors of social systems.

2.3. Self Organised Criticality in Social Systems

In social terms, agents and the individuals they represent are metastable if they do not respond to every minute stimulus they face. They would not, for example, reconfigure their desired shopping basket as a result of a penny rise in the price of a tin of tuna. A particular implication of metastability is that the behaviour of individuals cannot be represented by utility maximising software agents. The dominance of interaction among the agents amounts to what Granovettor [6] and Edmonds [15] have called social embeddedness: the behaviour of individuals cannot be explained except in terms of their interaction with other individuals. Dissipation in a social system, analogous to the dissipation of grains of sand in the sandpile model, equates to individuals being influenced by other individuals without slavishly imitating them.

In the social science literature, Moss [17] has produced evidence that leptokurtic frequency distributions apply widely to sales volumes and values for fast moving consumer goods such as alcoholic beverages, biscuits, tea, shampoo and shaving products. This evidence depends on the highest frequency, best quality data ever available as a result of electronic point of sale data capture. In the same paper, Moss demonstrated that a general model of inter-

[4] Bollerslev [10] identifies the core econometric processes of relevance here to be the ARCH process [11], the GMM process [12] and GARCH [13].

mediated exchange in which agent cognition and behaviour is specified on the basis of well validated concepts in cognitive science and social psychology (and is therefore metastable) with intensive local interaction among agents and word of mouth communication yields the statistical signatures of data from real markets for fast moving consumer goods. In particular, sales volumes by intermediaries were leptokurtically distributed with clusters of volatility while market shares were power law distributed.

The implication of these results is that leptokurtosis and clustered extreme events are found in competitive intermediated markets generally and are not restricted to the financial markets where speculation is an important and sometimes dominant feature. That is, the statistical signature of the financial markets is due to particular instances of much more general phenomena: individual metastability, social interaction and influence by but not imitation of friends and acquaintances.

If we consider the natural and social systems as two data generating mechanisms, then we observe that the data emerging from natural systems typically have stable statistical properties while data emerging from social systems typically do not. For example, earthquake magnitudes are power law distributed (the Gutenberg–Richter law) and the observed power law distribution does not change over time. Indeed, if it did it would not be observed as a power law distribution. So while it is not possible to predict the occurrence of specific earthquakes or earthquakes of specific magnitudes, it is possible to describe with considerable accuracy the distribution of earthquake magnitudes. However, in social systems experience of extreme events leads to a search for means of reducing their incidence if that is possible or their impact if it is not. The point is to change the observed distribution by, in effect, changing the data generating mechanism. To the extent that such social engineering is successful, the magnitude of the exponent of the power law distribution will be reduced. The goodness of fit of a power law distribution will be reduced although that will not eliminate the leptokurtosis of the observed data.

An important feature of SOC models for purposes of the present discussion is that there are a few, special analytical results concerning their properties but there are no general results. Our understanding of self organised criticality depends almost entirely on simulation experiments. Moreover, the social simulation models of Moss described above were produced initially in total ignorance of the literature on self organised criticality. The agents were implemented as metastable entities because that is a consequence of representations of cognition based on computational and experimental cognitive science and on the implementations designed to enable the modeller to capture criteria used by real actors in devising courses of action. The patterns of interaction among agents were developed in light of evidence concerning the importance of word of mouth communication. That individuals influence but do not imitate one another is again a property of the model that emerged naturally but with which the modellers were (and remain) comfortable. That these models reproduced the statistical signatures – time-series leptokurtosis and cross-sectional power law distributions – obtained from retail sales data is, we now but did not initially know, due to the emergence of self organised criticality in the models.

The models themselves were designed to describe specific social processes. The reasons for their statistical signatures are explained by the SOC literature which also indicates that, in the present state of mathematical understanding, the same processes cannot be captured by analytical models. Simulation is the only tool we have that yields models validated by their generation of leptokurtic time series and power law distributed cross sectional data. Because the process yielding these results is explicit, its plausibility as a representation of real social processes can be assessed by domain experts.

3. SOCIAL EMBEDDEDNESS AND METASTABILITY: AN EXAMPLE

If an individual is socially embedded then that individual's behaviour cannot be explained except in relation to interaction with some other individuals.

In this section, results are reported from a model of domestic water demand and consumption. The model, developed in ignorance of SOC and the implications of leptokurtosis, was designed to analyse the effect of policy pronouncements and exhortation on water demand in various circumstances. In particular, water companies and regulatory authorities participating in two projects headed up by the Environmental Change Institute in the University of Oxford expressed the clear and common view that domestic water consumption in the UK has historically been influenced during dry periods by requests from water suppliers and government agencies. The purpose of the model reported here is to investigate the social process generating that result.

The model incorporated both a hydrological model to determine soil water content on the basis of historical temperature and precipitation data and a social model to capture the effects of soil water content on policy agencies and their pronouncements as well as the effects of those pronouncements on household consumption. These component models are described in turn.

3.1. The Hydrological Model

The purpose of the hydrological model is to simulate the occurrence of drought conditions on the basis of real precipitation and temperature data. The policy agent determines there to be a drought when soil water is less than 85 per cent of the soil water retention capacity for two or more consecutive months.

The soil water content for any month is determined by adding to the previous month's soil water content the total precipitation during the month in question less the evapotranspiration. If the result is less than the water retention capacity of the soil, then that is the current soil water content. If the result is greater than the capacity, the difference is runoff.

Potential evapotranspiration (PET) in the model is determined by temperature, mean day length during the month (calculated for the relevant latitude for each day relative to the winter solstice) and a monthly correction factor taken from the following table:

Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
0.8	0.8	0.8	0.9	1	1.1	1.1	1.1	1.05	0.85	0.8	0.8

The uncorrected value of PET is calculated from the modified Thornthwaite algorithm from temperature and hours of daylight per day. The value of the unadjusted PET is calculated as:

PET	Temperature (T) range
$-415.8547+32.2441T-0.4325T^2$	$26.5 \leq T$
$16.5\ (9\ T/H)^a$	$0 = T \leq 26.5$
0	$T < 0$

where H is heat defined as

$$H \equiv \left(\frac{T}{0.7}\right)^{1.514}$$

and the exponent a is

$$a = 6.75e - 7H^3 - 7.71e - 5H^2 + 0.01792H + 0.49239$$

The resulting unadjusted value of PET is adjusted to allow for day length so that

$$aPET = uPET * (days/30) * (mdl/12)$$

where $uPET$ is the unadjusted value of PET determined as above, *days* is the number of days in the month and *mdl* is the mean day length in the month.[5]

3.2. The Social Model

Because nobody knows everybody but everybody knows somebody, the social model was constructed to capture the fact that every individual has a limited by usually non-empty set of acquaintances. This was achieved by situating agents on a toroidal grid and enabling each agent to know every agent within a set number of cells in each of the four cardinal directions. In all of the simulation runs reported here, the grid size was 16^2 ($=256$) cells, there were either 80 or 100 agents distributed at random on the grid with no more than one agent to a cell and each agent had a horizon of 4 cells in the sense that it knew all agents in the four cells to the right, to the left, up and down.

Two abstract types of water consumption event were defined: private consumption events that are not visible to any but the consuming agent and public consumption events that are publicly visible. Each agent is able to observe the public consumption events of all agents known to it. Each agent decides how much water to use during each event and the frequency of use events measured by the number of events per month. This is also the information viewed by the agent's neighbours.

In addition, each agent can read the exhortatory policy pronouncements of the policy agent. These pronouncements take the form of recommended frequencies of each activity (public and private) as well as volume consumed per use event. As indicated, such pronouncements are made only during periods of drought defined as months in which soil water is below 85 per cent of capacity and has been below that level for at least two consecutive months. The frequency and volume per event recommended by the policy agent are both determined by the previous month's use and the current shortfall of soil water below 85 per cent of capacity. So if, for all domestic households, the previous average frequency of a water using activity were f and the volume used per event were u and the current soil water magnitude is 75 (capacity $= 100$), then the policy agent would recommend that each household engage in the consumption activity on $0.75 \times f$ occasions (rounded down to the nearest integer) and that the volume used per event be $0.75 \times u$. In every drought month when the soil water diminishes, the frequency and volume demanded by the policy authority is reduced according to the above formulae. The policy agent does *not* change the demanded frequency or use per event when drought conditions persist but the soil water increases. When the drought ends – because soil water exceeds 85 per cent of capacity – the policy agent simply ceases to offer any exhortation.

This behaviour by the policy agent captures at coarse grain the behaviour actually observed with regard to the authorities' public pronouncements and statutory restrictions in the UK during periods of drought.

3.3. Agent Cognition

The cognitive processes of the agents were specified to capture observed patterns of demand for goods and services by representing social interactions that have been well documented by observation and experiment by social psychologists in a manner that also reflects the fact that individuals change their behaviour in response to a significant weight of evidence and social pressure but not

[5]This model was provided by Tom Downing of the Environmental Change Unit, University of Oxford.

in response to small changes in price, incomes or any other variable. The representation of the determinants of behaviour are described first and then the specific representations developed to capture the role of exhortation and social interaction.

The representation of the decision-making process is the *endorsements* mechanism devised by Paul Cohen [18] and used to good effect in several social simulation studies (e.g., [19]) to capture observed behaviour. The endorsements mechanism entails the use of rules to attach tokens to objects. Two types of object were endorsed by household agents in the Thames prototype model: actions and other household agents. The actions available to the household agents are water-use actions consisting of use frequencies and consumption per use-event. The other household agents are, of course, the other household agents that they can see.

There is an endorsement scheme for each type of object to be endorsed. So in the Thames model there is one endorsement scheme for water-use actions and one for visible household agents. For all households, the tokens with which actions and agents, respectively, can be endorsed are identical.

The endorsements on actions are *globallySourced*, *neighbourhoodSourced*, *selfSourced* or *bestEndorsedNeighbourSourced*. The first three are straightforward. Globally sourced actions are actions that the policy agent is recommending publicly. Neighbourhood sourced actions are actions that are observed to be taken by visible neighbours while self sourced actions are actions preferred by the individual independently of any external influences. The fourth endorsement, *bestEndorsedNeighbourSourced*, is accorded to the action observed by the neighbour that is most highly valued among all of the visible neighbours of the agent. That agent is the neighbour that is most like the endorsing agent and is determined by the agents endorsements scheme.

The endorsements on other agents are *closestActivityConsumption*, *closestActivityFrequency* or *closestActivityFrequencyAndConsumption*. The idea here is that other agents can be valued because their consumption per use-event is closer to one's own than the consumption per use-event of any other agent or because the frequency with which an agent engages in an activity is closest to one's own frequency of engaging in that activity. There is an added appreciation of any agent whose consumption is most like one's own in both frequency of events and consumption per event. The motivation for these endorsements is the common finding by social psychologists that similarity of attitudes and mutual (not necessarily romantic) attractiveness are highly correlated. That is, individuals tend to share the attitudes of those they like and to like those whose attitudes they share. The attitudes that lead to similar consumption patterns are taken here to induce personal attraction which in turn reinforce the similarity of attitudes. In this model, attitudes towards authority and towards social norms determine patterns of water consumption.

It is, of course, essential to be able to evaluate which actions and which neighbours are the most attractive. This requires some means of comparing different collections of endorsement tokens. Cohen's original endorsements approach allocated endorsement tokens to classes of importance. The action chosen would be that which had the most endorsements of the highest class or, if several had the same number in the highest class, the action that was tied in the highest class but had the most endorsements in the second highest class. If there was another tie at the second highest class of endorsements, the third or if necessary the fourth or lower class would be used to break the tie.

A more general approach, and that used here, is to define a number base b and evaluate each endorsed object according to the formula

$$V = \sum_{e_i \geq 0} b^{e_i} - \sum_{e_i < 0} b^{|e_i|}$$

where e_i is a (usually integer) value associated with the ith endorsement token. Negative values of endorsement tokens indicate naturally enough that they are undesirable. The higher the value associated with an endorsement token, the higher the class of tokens containing that particular token. The value of b is the importance of an endorsement token relative to the value of a token in the class below. If the base is 2, then an endorsement of class three contributes 8 to the endorsement value of an object while an endorsement of class two contributes only 4. For values of b larger than the number of tokens in any class used to endorse any object, the results from this evaluation scheme are the same as from Cohen's evaluation scheme. For smaller values of b it is possible for a large number of lesser endorsements to outweigh a small number of endorsements of greater value.

For each agent, the value of each endorsement token was set at random as was the numerical base (the value of b) of the endorsement scheme. The desired consequence of this approach was to allow for heterogeneity among agents as well as different agent populations in each simulation run. Some agents would be most influenced by external authority, some by their neighbours and some would be mainly self-directed. On average, a third of the agents would fall into each category.

A useful set of experiments would specify the proportions of agents in each category with results that could be validated by surveys or focus groups to determine the actual proportions in different cultural contexts and the critical proportions that make exhortation effective. Clearly, there must be some minimum proportion of agents who respond to public exhortation if the exhortation is to be noticed, much less effective. Equally clearly, some agents must be influenced by their neighbours if the behaviour of the agents influenced by exhortation is to proliferate through the rest of the community. Moreover, the greater the proportion of agents who are self-directed, the less likely are there to be agents who respond to exhortation and the less will be density of agents influenced by their neighbours.

The point of further experimentation with this model will be in part to verify that there are critical proportions of agents influenced by authority, by neighbours or not at all. While it is conceivable that the critical proportions found in the simulation experiments are reflected in reality, a more robust finding would be statistical patterns corresponding to the presence or absence of the proportions of each type of agent and their concentrations in the grid. Such statistical indicators have been found in other simulation studies (e.g., [20, 21]).

3.4. Simulation Results

Earlier versions of the model resulted in a reduction in water consumption by agents representing domestic households with the effect taking hold as the length of time over which the policy exhortation persisted was longer. However, water demand returned to more normal levels immediately the exhortation ended. This result was implausible to stakeholders (water companies and regulatory agencies). Consequently, the representation of cognition was changed so that water demand was more influenced by individuals' norms of water consumption in different uses and also by allowing for most of the endorsements attaching to each object to be forgotten over time. The consumption norms were generated at random for each agent at the time of its creation. The norms for frequency and use-per-event were endorsed as such by each agent and were remembered throughout the simulations. All other endorsements would be remembered with a probability that is directly related to the importance of the endorsement token (the magnitude of the token's value) and inversely related to the lapse of time since the endorsement was attached to its object. This mechanism, suggested by Anderson's work on memory and recall [22], sets the probability of remembering an endorsement as vt^{-d} where v is the value of the endorsement token, t is the number of months since the token was attached to the action or agent and $d > 0$ is the factor determining the rate of decay of the probability of remembering the endorsement. The value of d was determined at random with a maximum set by the model operator for each experiment.

These changes in model specification had the intended effect as confirmed by the consumption and duration of soil dryness reported in Figure 2 where we see that consumption recovered over a period of months after the drought spikes in the chart.

The relationship of consumption to exhortation in the simulation experiments is reported in Figure 3 covering two periods of drought: a relative short period in the summer of 1989 and the longer drought during much of 1990 and into 1991. In both cases, simulated domestic users reduced consumption more or less in line with exhortation. An interesting result to be confirmed by data and domain experts

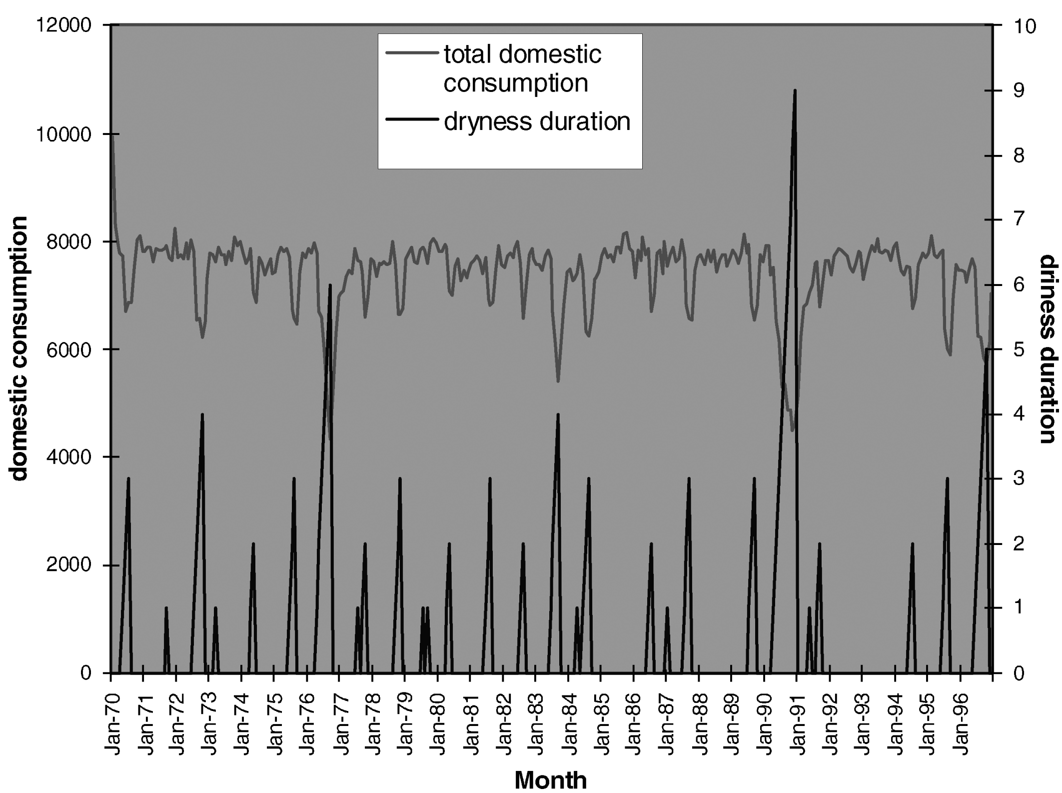

Fig. 2. Simulated consumption and calculated drought, 1970–1996.

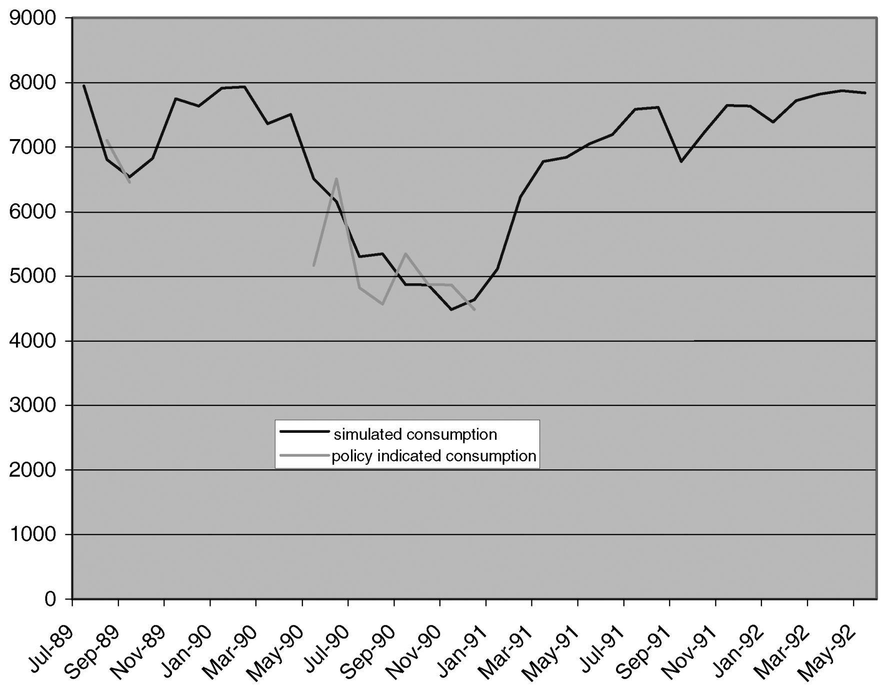

Fig. 3. Exhortation and domestic use (simulated).

is the reduction in consumption was longer lasting and the recovery more gradual as a result of the longer period of exhortation and deeper cuts in water user in 1990–1991 in comparison with the relatively brief drought and period of exhortation in 1989.

The specification of individual behaviour certainly entails metastability and some relatively local interaction. The enhanced importance of individual consumption norms implies that individuals are less likely and less completely to imitate their neighbours so that the social system is likely to be more dissipative than in the first versions of the model. It is of course important to realise that the dissipativeness follows from stakeholder descriptions of household behaviour and not from any attempt to produce a self organised critical model. The specifications of agent behaviour stemmed from the validation process made possible by stakeholder participation. In that sense and to that extent, validating the descriptive accuracy of the model specification is part and parcel of the process of model development.

It remains to demonstrate that the statistical signature of the model is the same as the statistical signature of the target system. The simulated water consumption reported in Figure 2 is certainly leptokurtically distributed.

The frequency distributions of consumption and the first difference in consumption are seen in Figure 4 to be very far from normal and certainly leptokurtic. Both the

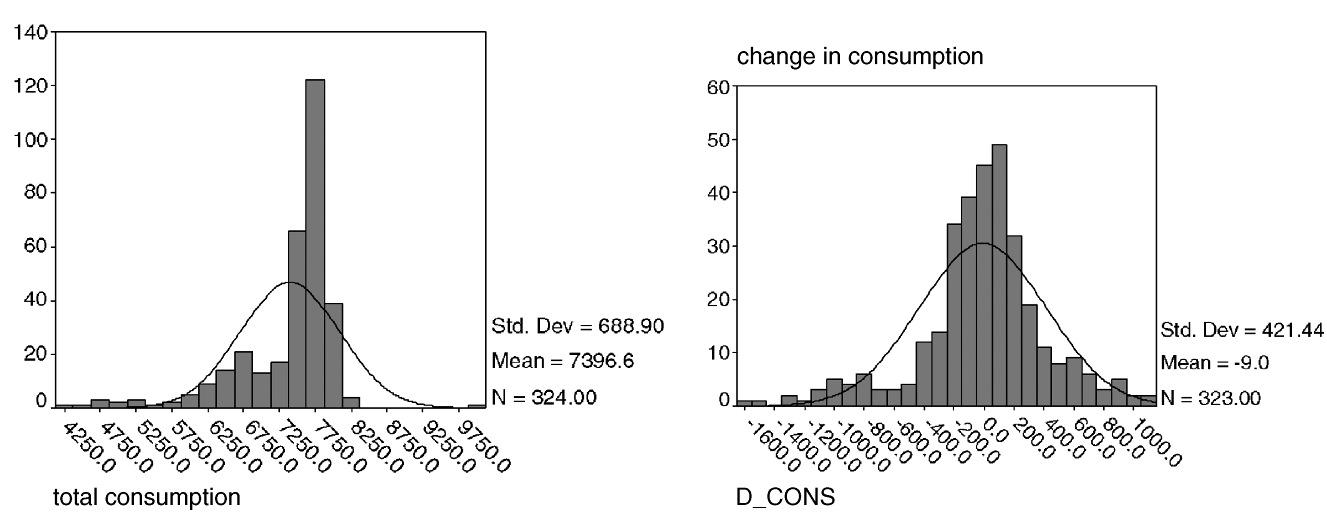

Fig. 4. Frequency distribution of total simulated consumption (left) and change in simulated consumption (right).

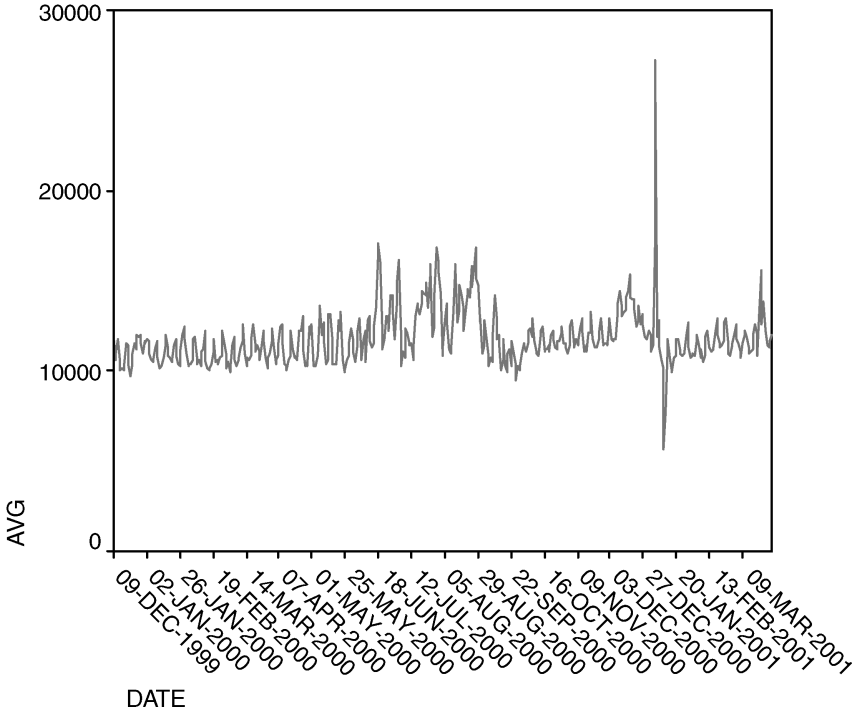

Fig. 5. Average household consumption (metered) over 18 months for a neighbourhood in the south of England.

Kolmogorov–Smirnov and Shapiro–Wilk statistics show that the confidence that the observed distributions are normal is zero to at least three significant digits. These results together with the observed clustering of episodes of volatility are classic indicators of self organised criticality.

While there is no reason to expect the scale of any actual consumption series to match the simulated series, we do require evidence of self organised criticality of the target system. This evidence is provided in Figure 5 for the time series of average daily household water consumption in the south of England and Figure 6 for the comparison of the actual and normal distributions.

The distribution of both the average daily consumption and the change in average daily consumption are significantly different from normal to any measurable confidence level.

It is clear that both the water demand model and the target system when viewed as data generating processes yield the statistical signatures of self organised critical systems. The skewness of the simulated consumption frequency distribution is different from that of the observed actual distribution although the skewness of both actual and simulated consumption first differences are virtually non-existent.

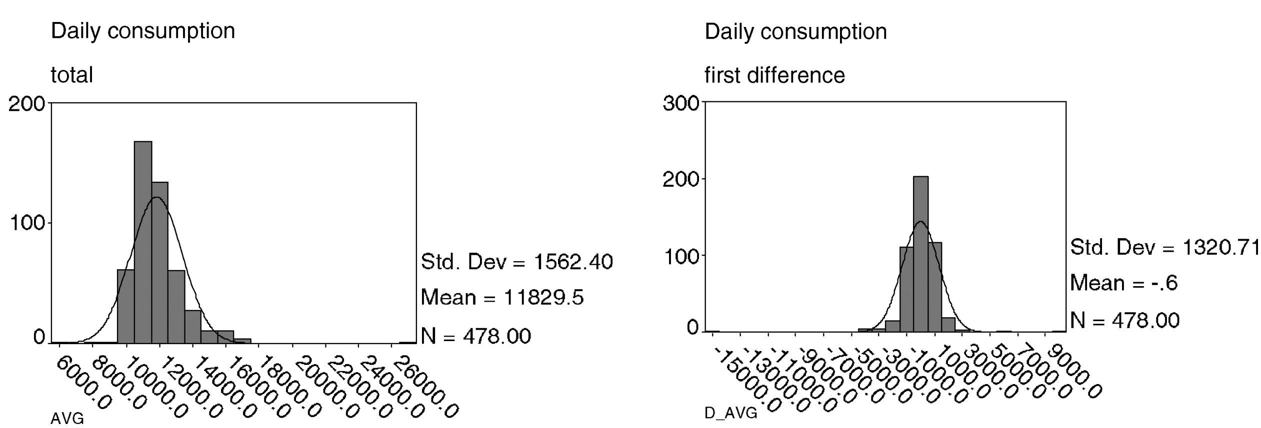

Fig. 6. Frequency distribution of average actual consumption (left) and change in actual consumption (right).

4. EXTENDING THE MODEL

The model reported in the previous section accounted for the frequency with which various abstract water activities were undertaken and volume of water used on each occasion. However, the UK water companies, regulatory agencies and academic students of water issues are also concerned with the ownership of appliances necessary to use water. This is the ownership-frequency-volume (OFV) model. A particular concern of the regulatory agency for water quality and security of supply, the UK Environment Agency (EA) has been the growing use of power showers among UK households. A more general concern has been the development of an understanding of the microcomponents of water demand for use in discussions with the regulated water companies. Since there are natural conflicts of interest between regulated companies in any industry and their regulators, it is important for the companies and agencies to have a basis of discussion of their different perceptions of such key issues as domestic water demand. Insight into the determination and evolution of the different microcomponents is clearly essential in planning for future water supply capacity and quality issues.

In order to assess whether an agent based social simulation approach would be a suitable basis for both the exploration of the microcomponents and a platform for a comparison of perspectives on water issues more generally, the water demand model reported in the previous section was modified to take account of the requirement for appliances in order to undertake water consumption activities. The simulations with the revised model were limited to representations of demand for water for purposes of personal washing.

4.1. The Extended Model Specification

The only changes made to the previous model were to capture the ownership of appliances and also to include in the model agents who tend to be early adopters of new technology.

In the setting up of each simulation run, data on existing appliances, the ownership penetration, the average frequency of use and the volume of water per use for each of these appliances was used to allocate the appliances to households and to specify the frequency and volume for each household. Thus the proportion of agents representing households that owned each type of appliance was an accurate reflection of the penetration of that appliance in an actual population.

To allocate the preferred frequency and volume for each household agent, it was assumed that these values would be power law distributed. This assumption is certainly open to independent validation *and* it conforms to what we know about cross section data in self organised critical systems and their target systems. For example, Moss [4] has shown both that market shares in competitive intermediated markets and simulated market shares in an SOC model of intermediated markets are power law distributed. Data provided by the EA included mean frequencies and volumes associated with each type of appliance. Given the number of household agents and the mean frequency (or volume), it is straightforward to calculate the appropriate exponent of the power law distribution.

Taking the rank of the household from the smallest frequency (or volume) to the highest, the value of the frequency (or volume) of the household at rank i will be

$$n_i = r_i^{\alpha}$$

where n_i is the frequency (or volume) associated with the household at rank r_i. Transforming to logs and taking the geometric mean of frequencies (or volumes),

$$\mu = \alpha \frac{1}{N} \sum_{i=1}^{N} \ln r_i$$

where N is the number of households. Consequently,

$$\alpha = \frac{N\mu}{\sum_{i=1}^{N} \ln r_i}$$

The households were ranked at random for frequency and volume associated with each appliance and the frequency was rounded for obvious reasons.

The model operator was offered the option of selecting an appliance to be "introduced" during the simulation run. This feature made it possible in the simulation runs reported below to capture the introduction of power showers during 1990. No ownership of power showers was taken from the data. The other appliances represented in the model were ordinary showers, wash hand basins and baths. In the reported runs, showers, power showers and baths were assumed to be public activities while wash hand basin activities were assumed to be private. This distinction was meant to capture the fact (if fact it be) that people talk about installing new appliances or having baths or showers but tend not to discuss washing their hands and faces. In developing such a model, it will be essential to identify whether the private-public distinction is valid and, if so, which activities are public or private in the sense used here.

In early runs, power showers were adopted by the whole simulated population within a year of their introduction, no matter how small were the number of agents with a propensity for early adoption. On the suggestion of the EA informants, the later versions allowed for the relatively infrequent (once every 5 years) refurbishment of a bathroom and this was taken to be a necessary condition for the installation of a power shower. With experimentation on the frequency of refurbishment, any length of time to achieve high penetration of a new appliance can be generated – in particular a length of time corresponding to that found in reality.

4.2. Simulation Results

Once again, the statistical signature of the simulated consumption series entailed leptokurtosis and clustered volatility. The histograms and data are not very different in shape from those reported for the earlier version of the model and, so, are not reproduced here.

It is, of course, important to emphasise that this model and its predecessors are constructed on the basis of a large number of simple assumptions. We require a process of model development that replaces the arbitrary (though plausible) assumptions about the distribution of frequency and volume (e.g., whether and how the two are related) for individual appliances and across appliances. Much of this data will have to come from surveys and similar means rather than by trying to infer microcomponents of demand from total consumption and demographic data alone. Claudia Pahl-Wostl and her colleagues at EAWAG [23] have already made substantial progress in developing the means of eliciting this information from individuals. The key point here is that all of these assumptions are open to independent validation.

5. CONCLUSION

To give a general perspective on the modeling approach and specific developments discussed above, it is useful to return to the issues raised in the opening section.

5.1. Is Participatory Agent Based Social Simulation Good Science?

The objective of a science of integrated assessment is to inform and support specific policy and strategic objectives. The three versions of the model of water demand reported in Sections 3 and 4 above were developed precisely to support policy discussions among water supply and treatment companies and regulatory agencies. The future course of water demand and use as well as the means of influencing or at needs restricting that demand are crucial policy issues. The means of developing these models by interaction between modeller and stakeholders has had two benefits in this regard. One is that the models reflect the expertise and opinions of both the regulated companies and the regulating agencies. Specific examples of the influence of this expertise are the relatively major modification of the first version of the water demand model made in order to capture the understanding of stakeholders from both sides of the regulatory divide regarding the post-drought time pattern of water usage recovery. The changes in the description of the determinants of household behaviour were checked for plausibility with stakeholders. This practice not only amounts to useful validation of an important aspect of the model, but it also engages the stakeholders in model construction in a way that can give them some understanding of the key model assumptions and also some confidence that the model provides a plausible description of a complex social process interacting with a complex physical (hydrological) process.

5.2. Is the Water Demand Model Scientifically Respectable?

The claim that the model is respectable rests on the extent to which it conforms to observation rather than the extent to which it draws on any grand theory. There are well verified concepts, these are used at a low level. The key example here is the consistent experimental and observational finding of social psychologists that individuals tend to like those who agree with them and agree with those they like. This led to the inclusion in the agents' endorsement schemes of the endorsements on known agents for similarity of water consumption behaviour. It is extremely important to note that this behaviour led to the influence of agents on one another without full blown imitation which amounts to dissipative interaction among the agents – an apparently necessary condition for self organised criticality of the whole system. A further necessary condition, agent metastability, was ensured by the adoption of the endorsements mechanism itself since any change in behaviour is then the consequence of a growing weight of evidence and stimulus rather than the possibly minute adjustments to optimality required by economic theory.

The scientific respectability of the particular models described here is arguably enhanced by the combination of the provenance of low level specifications of cognition and social interaction that are well validated by scientists working on entirely different research programmes and the consequent conformity of the statistical signature of the whole system with that of the target system as evidenced by the leptokurtosis and clustered volatility of the simulated and actual water consumption data.

5.3. The Relationship Between Theory and Observation

The models described above were as well validated as possible with available evidence and this, of course, is the importance of observation in relation to theory. Because software agents are implemented to capture the behaviour of observable social entities – in this case households – it is possible further to validate or change the model to yield improved descriptions based on more detailed data. It is certainly possible using standard interview and survey techniques to determine the distribution of intervals of time between bathroom refurbishments in different households or even to determine any correlations between the bathroom refurbishment interval and demographic characteristics. It is equally possible to identify patterns of social interaction – who is influenced by whom in the determination of water consumption patterns and which water consumption activities and hardware are discussed or displayed socially and which are not. In short, by virtue of the implementation of *agent based* social simulation models, the detail of validation can be undertaken to any desired extent.

5.4. The Models in Relation to Existing Literature

The only literature on which the models described above rest derives from three research programmes: artificial intelligence (including expert systems), social psychology and cognitive science. The idea of the endorsements mechanism was developed to enable expert systems to give reasons for decisions that would be clear and comprehensible to the expert system users. The advantage of this feature for purposes of validation is that the reasons generated by the software agents can be compared with reasons given by the stakeholders represented by those agents. The literature on social psychology and cognitive science led to the assumptions entailing dissipative social interaction and metastable behaviour, respectively. The validation of the model by comparison of statistical signatures clearly draws heavily on the statistical physics and self organised criticality literature. That the literature informing the model specification should lead to results that are explicable in relation to a literature that is distinct to the point that the link has not previously been made should be a source of confidence in the coherence of the approach. It is simply noted here that each of these branches of the scientific literature have been selected or used because they support validation of the models insofar as that is possible. Validation by detailed prediction is not indicated with these models. But then successful validation by detailed prediction is not a feature of any of the social science literature on which integrated assessment research has rested heretofore.

5.5. Modeling and Simplicity

The simplicity of the water consumption models is due to the coarseness of the grain of the specifications. For example, in the ownership-frequency-volume (OFV) formulation, the distribution of frequencies and volumes per use of each appliance is a power law distribution with each agent given a ranking chosen at random in each case. Experience with self organised critical systems suggests that cross sectional distributions are power law distributed so that, in the OFV demand model, demand by households would be characterised by a small number of consumers of large amounts of water and a large number consuming much smaller amounts. If it were to turn out that consumption by household were power law distributed and a model of household consumption at finer grain generated that distribution of consumption, then the assumption of that distribution used to determine normal consumption patterns in the OFV model would be validated against independently generated data and verified with respect to a model at finer grain. In general, coarseness of grain is typically required to retain model simplicity within the constraints implied by (for example) social embeddedness, but finer grain is necessary to validate the behaviour taken for granted in the more coarse grain models. Provided that fine and coarse grain models are mutually consistent, as in the example suggested (but not yet implemented) here, simplicity and detailed validation can both be obtained without compromise.

5.6. A Final Assessment

While participatory agent based social simulation has no basis in the existing integrated assessment literature or, more generally, in the literature of the social sciences apart from social psychology, it provides a useful basis for the engagement of stakeholders in the analysis of social policy and strategies. The models described here are already influenced by the domain expertise of important and knowledgeable stakeholders. UK regulatory agencies at least are interested in using developments of the OFV model to explore with regulated companies different assumptions underlying their respective scenarios of future water demand. The model has been developed so that the only exogenous data is temperature and precipitation for the hydrological submodel and the names, average frequencies and volumes associated with different appliances. Consequently, even with the present demonstrator model extended to include other important classes of appliance, it will be possible to explore the consequences of climate change (or its effect on temperature and precipitation) for the microcomponents of water demand. Unlike the IPCC and UKCIP scenarios, the scenarios based on participatory agent based social simulation models such as those described here produce emergent system behaviour based on validated specifications of individual behaviour and social interaction. Moreover, the specification and implementation of agent cognition has been developed to provide extensive information about the reasoning that led the individual agents to evolve the simulated behavioural patterns. This and more standard data captured from the models provides ample information for stakeholders to evaluate and agree or debate the plausibility of the scenarios and the sorts of opportunities and crises that might emerge from the enormously complex and inherently unpredictable natural and social systems that dominate the outcomes of their decisions.

ACKNOWLEDGEMENTS

The arguments of this paper are the result of discussions with my longstanding colleague Bruce Edmonds and collaborators on the FIRMA project, especially Jan Rotmans, Anne van der Veen, Tom Downing and Claudia Pahl Wostl. I am very grateful to an anonymous reviewer and to Claudia Pahl Wostl, both of whom clearly took a great deal of time with an earlier draft. None of the above bear any responsibility for the resulting paper and none necessarily agree with my arguments.

REFERENCES

1. IPCC-TGSCIA: *Special Report on Emissions Scenarios*. Cambridge University Press, Cambridge, 2001.
2. Nordhaus, W.: *Managing the Global Commons: The Economics of Climate Change*. MIT Press, Cambridge, MA, 1994.
3. Moss, S., Pahl, C.W. and Downing, T.E.: Agent Based Integrated Assessment Modelling: The example of Climate Change. *Integrated Assessment* 2(1) (2001), pp. 17–30.
4. Moss, S.: Game Theory: Limitations and an Alternative. *Journal of Artificial Societies and Social Simulation* 4(2) (2001) ⟨http://jasss.soc.surrey.ac.uk/4/2/2.html⟩
5. Moss, S.: *Relevance, Realism and Rigour: A Third Way for Social and Economic Research*. Centre for Policy Modelling, Manchester Metropolitan University, Manchester, 1999.
6. Granovetter, M.: Economic-Action and Social-Structure – The Problem of Embeddedness. *American Journal of Sociology* 91(3) (1985), pp. 481–510.
7. Edmonds, B.: Capturing Social Embeddedness: A Constructivist Approach. *Adaptive Behaviour* 7 (1999), pp. 323–348.
8. Mandelbrot, B.: The Variation of Certain Speculative Prices. *Journal of Business* 36(4) (1963), pp. 394–419.
9. Mandelbrot, B.: *Fractales, Hasard et Finance*. Flammarion, Paris, 1997.
10. Bollerslev, T.: Financial Econometrics: Past Developments and Future Challenges. *Journal of Econometrics* 100(1) (2001), pp. 41–51.
11. Engle, R.F.: Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50 (1982), pp. 987–1007.
12. Hansen, L.P.: Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50 (1982), pp. 1029–1054.
13. Bollerslev, T.: Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31 (1986), pp. 307–327.
14. Bak, P., Tang, C. and Weisenfeld, K.: Self Organized Criticality: An Explanation of 1/*f* Noise. *Physical Review Letters* 59(4) (1987), pp. 381–384.
15. Jensen, H.: *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*. Cambridge University Press, Cambridge, 1998.
16. Edmonds, B.: *Social Embeddedness and Agent Development*. Manchester, Centre for Policy Modelling, Manchester Metropolitan University, 1998.
17. Moss, S.: Policy Analysis from First Principles. *Proceedings of the US National Academy of Sciences* 99 (suppl 3), pp. 7267–7274.
18. Cohen, P.R.: *Heuristic Reasoning: An Artificial Intelligence Approach*. Pitman Advanced Publishing Program, Boston, 1985.
19. Moss, S. and Sent, E.-M.: Boundedly Versus Procedurally Rational Expectations. In: A.H. Hallett and P. McAdam (eds.): *Analyses in Macro Modelling*. Kluwer Academic Publishers, Boston, 1999, pp. 115–146.
20. Moss, S.: Critical Incident Management: An Empirically Derived Computational Model. *Journal of Artificial Societies and Social Simulation* 1(4) (1998).
21. Moss, S.: Canonical Tasks, Environments and Models for Social Simulation. *Computational and Mathematical Organization Theory* 6(3) (2000), pp. 249–275.
22. Anderson, J.R.: *Rules of the Mind*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.
23. Pahl-Wostl, C., et al.: Models at the interface between science and society: Impacts and Options. *Integrated Assessment* 1 (2000), pp. 267–280.