Uncertainty from model calibration: applying a new method to transport energy demand modelling

Bas van Ruijven[*1], Jeroen P. van der Sluijs [2], Detlef P. van Vuuren [1], Peter Janssen [1], Peter S.C. Heuberger [3], Bert de Vries [1 2]

[*] Corresponding Author

[1] *Netherlands Environment Assessment Agency (PBL)*
PO Box 1, 3720 BA Bilthoven, The Netherlands

[2] *Utrecht University, Copernicus Institute for Sustainable Development and Innovation, Department of Science, Technology and Society*
Heidelberglaan 2, 3584 CS Utrecht, The Netherlands

[3] *Delft University, Delft Center for Systems and Control*
Mekelweg 2, 2628 CD Delft, The Netherlands

NWS-E-2009-50

ISBN: 9789 086 720 491

Abstract:

Uncertainties in energy demand modelling originate from both limited understanding of the real-world system and a lack of data for model development, calibration and validation. These uncertainties allow for the development of different models (from different scientific paradigms), but also leave room for different calibrations of a single model. Here, an automated model calibration procedure was developed and tested for transport sector energy use modelling in the TIMER 2.0 global energy model. This model describes energy use on the basis of activity levels, structural change and autonomous and price induced energy efficiency improvements. We found that the model could reasonably reproduce historic data under different sets of parameter values, which project different future energy demand levels. Projected energy use for 2030 shows a range of 44-95% around the best-fit projection. Two different model interpretations of the past can generally be distinguished: 1) high useful energy intensity and major energy efficiency improvements or 2) low useful energy intensity and little efficiency improvement. Generally, the first lead to higher future energy demand levels than the second, but model and insights do not provide decisive arguments to attribute a higher likelihood to one of the alternatives.

Keywords

Model calibration, uncertainty, global energy model, transport energy use

Abbreviations

AEEI	Autonomous Energy Efficiency Improvement
GDP	Gross Domestic Product
GLUE	Generalised Likelihood Uncertainty Estimation
IEA	International Energy Agency
IMAGE	Integrated Model to Assess the Global Environment
IPCC	Intergovernmental Panel on Climate Change
NRMSE	Normalised Root Mean Square Error
OECD	Organisation for Economic Co-operation and Development
OECD-EO	OECD Environmental Outlook
PEST	Parameter Estimation
PIEEI	Price Induced Energy Efficiency Improvement
SRES	IPCC Special Report on Emission Scenarios
TIMER	The Image Energy Regional Model
UE	Useful Energy
UEI	Useful Energy Intensity
UNEP	United Nations Environmental Program
WDI	World Development Indicators

1 Introduction

Uncertainties play a key role in projecting future developments of the energy system. At least two factors contribute to this: 1) the energy system is determined by complex interactions of a wide range of drivers and 2) there is a lack of empirical data. Factors that influence future energy demand and supply include economic activity, developments in economic structure, lifestyle changes and technology development. Our understanding of the interaction of these factors is still limited (and they may range over a wide range of possible outcomes). On top of this, the lack of empirical data complicates the development and calibration of models, especially for developing regions.

Despite limitations in both theory and data availability, a wide range of models has been developed to explore trends at global, regional and national scales. These models are partly developed from different scientific paradigms, which may lead to different interpretations of the past and different expectations of the future [1, 2]. A classic example is the difference between models from a macro-economic tradition (top-down) and those from a technological tradition (bottom-up). These two traditions tend to interpret the present situation differently with respect to energy efficiency ('improvement of energy efficiency leads to higher costs' vis-à-vis 'major opportunities for improvement without substantial costs') and as a result also expect different mitigation costs in the future [3]. Even within one model, however, often different options exist on how to interpret the current and past situation. For instance, macro-economic demand functions often include both income-elasticity and price-elasticity, which are hard to identify unambiguously in historic data. A different interpretation of the past may lead to different calibrations of the model and uncertainty in future projections. So far, different methods have been used to explore uncertainty in global energy models [4-7], but relatively little attention has been given to the influence of model calibration on future projections.

The issue of multiple model calibration is closely related to the concept of equifinality, which focuses attention 'on the fact that there are many acceptable representations that cannot easily be rejected and should be considered in assessing the uncertainty associated with predictions' [8]. These 'acceptable representations' are called *behavioural*. The "acceptance criterion" can be defined strictly quantitative (e.g. above a threshold value of a likelihood measure) or more qualitative (e.g. reproduction of trends). At present, calibration of energy models is often done on the basis of the modeller's expert knowledge to identify a single set of plausible parameter values. However, if multiple sets of parameter values are tenable and model projections are sensitive to the parameter values chosen, this practice is questionable [9].

In this context, we have developed a method to automatically calibrate models and obtain sets of parameter values that perform reasonably against historic data. These calibrated sets are obtained by varying the main model parameters within a limited range, choosing an initial estimate in this range, and searching consecutively for a (local) optimum to minimise the error between observations and model results. Repeating this procedure many times, initialised at different locations in the parameter space, generates a series of (different) calibrated sets of parameter values. This method is related to both nonlinear regression methods like PEST [10] or UCODE [11] and (sequential) Monte Carlo based approaches like GLUE [12] or SimLab [13].

We apply this method to the energy demand module of the global energy model TIMER 2.0, a system dynamics model that simulates developments in global energy supply and demand [14, 15]. The TIMER 2.0 model is the energy sub-model of the Integrated Model to Assess the Global Environment, IMAGE 2.4, that describes the main aspects of global environmental change [16]. In recent years, this model has been used in several global scenario studies like the IPCC Special Report on Emission Scenarios [17], the Millennium Ecosystem Assessment [18], UNEP Global Environmental Outlook [19] and the OECD Environmental Outlook [20].

Since the development of the TIMER model several uncertainty studies have been performed [21-23]. These analyses accepted the model's initial calibration and focused on the spread in model outcomes based on variation in central input values. Moreover, all TIMER uncertainty studies (and for that matter the same applied to other global energy models) focused on the global level, neglecting interesting underlying trends in different regions. Recent analysis of TIMER found that uncertainty in energy demand trends – and thus the factors underlying these trends – is a major source of model uncertainty [6]. Therefore, we focus this analysis on the TIMER energy demand sub-model. Within energy demand modelling a further choice was made to focus on the transport sector, which is the sector with the fastest growth in energy demand. For the regional focus, 6 regions were selected: the USA, Western Europe, Brazil, Russia, India and China. These regions are among the largest regions in terms of energy use and, moreover, represent a wide spectrum of development levels.

In this paper, first in Section 2 we discuss the role of uncertainty in energy modelling and introduce a methodology to capture uncertainty in model calibration. In the second part of the article, we elaborate on the application of the method: Section 3 describes the structure of the TIMER 2.0 energy demand model and selects parameters that are useful for model calibration. Section 4 presents the results of the analysis, Section 5 evaluates the presented methodology and Section 6 discusses and concludes.

2 Uncertainty in model calibration

2.1 Uncertainty in energy models

Exploration of different futures on the basis of models is complicated by inherent uncertainties [24-31]. Uncertainty and associated terms (such as error, risk and ignorance) are defined and interpreted differently by different authors [for reviews see 29, 32, 33, 34]. These different definitions partly reflect the underlying traditions and their associated scientific philosophical way of thinking. In general, uncertainty may be identified of input parameters, model structure or even different theories at a more aggregated level. Part of these uncertainties are related to natural randomness (ontic). Other uncertainties results from limited knowledge (epidemistic). One phase of model development where uncertainties become apparent is during model calibration. Model calibration and validation are of critical importance. As Oreskes et al. [35] highlight, "In areas where public policy and public safety are at stake, the burden is on the modeller to demonstrate the degree of correspondence between the model and the material world it seeks to represent and to delineate the limits of that correspondence." However, given existing uncertainties in most cases historic trends and data can be interpreted in different ways. This is also emphasized by Beck [36] when he noted that almost all models suffer from a lack of identifiability, i.e. many combinations of values for the model's parameters may permit the model to fit the observed data more or less equally well.

The notion of ambiguity in model identification and calibration can be valued differently [37, 38]. In statistical modelling traditions, ambiguity in model calibration is typically interpreted as over-parameterisation of the model. Following Occam's razor, this could be solved with model reduction [39-42] or developing multiple specialised models [43] to strike a balance between model complexity and data-availability. In rule-based (system-dynamic) and engineering models[1] the model structure is based on (intuitive) causal relations and rules (either in physical or in monetary terms) that are calibrated to historic data [44, 45]. Such causal relations may be postulated, even in the absence of sufficient data for calibration. Beven [8] aims to extend traditional schemes with a more realistic account of uncertainty and rejects the idea that a single optimal model exists for any given case. Instead, models may not be unique in their accuracy of both reproduction of observations and prediction (i.e. unidentifiable or equifinal) and subject to only a conditional confirmation, due to e.g. errors in model structure, calibration of parameters and period of data used for evaluation.

In energy modelling literature, the most analysed sources of uncertainty are parameters and model structure in direct relation with future projections of model drivers. As a typical example, Tschang and Dowlatabadi [4] deal with input parameter uncertainty when performing an uncertainty analysis of the Edmonds-Reilly global energy model. They use Bayesian updating techniques to filter out model simulations that do not conform to outputs on energy consumption and carbon emissions and determine updated prior distributions for several core parameters. Van Vuuren et al. [6] use a slightly more complicated method, in which sampling of input parameters is made conditional upon different consistent descriptions of the future. With respect to model structure, a nice example is provided by Da Costa [5] who compares the results of two different energy models for Brazil. He concludes that although the aggregate results of these models are comparable, considerable differences exist when the results are broken down.

This study focuses on uncertainty that originates from the calibration of parameter values. We explore whether "acceptable sets of parameter values in model calibration (so-called 'behavioural sets') can be identified for the TIMER energy demand model and what these imply for the model's projection, inspired by Beven's work on equifinality.

It should be noted that the mismatch between model prediction and observation can stem from many different sources [8], including those related to measurement,

[1] Also, especially global energy models are highly policy relevant and are applied for multiple purposes (for instance looking into carbon emission, total energy use, structure of energy use or costs of mitigation measures). This implies that not all model-parameters influence the results of all outputs. Hence, these models are de-facto over-parameterised.

random error, but also the representation of reality by the model as a results of both parameter error and model structure. To keep our analysis manageable, here we assume that the parameter error is the dominant error component – and focus on the question whether our calibration procedure can indeed identify multiple, equally valid, calibrations of the energy demand model. Techniques exist to overcome this simplification and better deconstruct the mismatch between observation and prediction into the six constituting error terms of Beven [8] but this is beyond the scope of the present paper.

2.2 Methodology to identify calibrated sets of parameter values

We developed an automated parameter estimation procedure in order to explore the impact of different sets of parameter values on model outcomes. The aim of the developed parameter estimation methodology is two-fold. First, it is an automated model calibration procedure that minimises the error between model results and observations, generating a set of calibrated parameter values. In this sense it is related to nonlinear regression methods like PEST [10] or UCODE [11]. Second, by repeatedly applying the method it can be used to perform an uncertainty analysis on model calibration. This generates a series of calibrated sets of parameter values. This aspect is more related to (sequential) Monte Carlo based methods like GLUE [12] or SimLab [13]. The procedure closely follows the manual model calibration process that is normally applied to the TIMER model. This method involves several steps:

A. Determining useful parameters for model calibration and their associated ranges
B. Performing a series of model calibrations and identify sets of input parameters that perform well against historic data
C. Analysing the sets of calibrated parameter values
D. Analysing the impacts of calibration uncertainty on future projections.

2.2.1 Determining useful parameters for model calibration and their associated ranges

The first step of the method involves analysis of the model, to select useful parameters for the model calibration process. We also identify ranges for the calibration parameters, based on analysis of the model formulation, the values used in former calibrations, literature and expert judgement. This step is described in detail in Section 3 and related appendices. These ranges are used as boundaries in the parameter estimation process.

2.2.2 Performing a series of model calibrations and identify sets of input parameters that perform well against historic data

Criteria for calibration fit.

Several measures exist to evaluate the deviation between model results (predictions, P) and observed data (O), of which an overview can be found in Janssen and Heuberger [46]. We choose to use the normalised root mean square error (*NRMSE*), comparing individual time series of observations and predictions, and defined as:

$$NRMSE = \sqrt{\frac{\sum_{t=1}^{T}\left(\frac{P_t - O_t}{O_t}\right)^2}{T}} \qquad 1$$

In this, P_t and O_t indicate the predicted and observed value in year t and T is the number of years in the time series. This measure has values between zero (perfect fit) and infinite (random). Multiplied with 100, the *NRMSE* can be seen as the time averaged percentage deviation between the time series of model results and the time series of observations. A certain threshold level for the *NRMSE* can be defined, below which models are called *behavioural* with the data (e.g. a *NRMSE*<10%), but in Section 4 we show that it is hardly possible to find criteria for such general numeric threshold.

We use the *NRMSE* for several reasons. First, it expresses model error at the individual data level. The alternative, expressing model error on the average level, only provides a rough impression of the model-data-discrepancy and averages out the dynamic features [46], whereas with calibration one wants to simulate both trends and patters in the data. Second, the *NRMSE* can easily be normalised in each year to observed energy use to prevent that years with higher energy demand dominate the estimated overall error.

Series of model calibrations
As starting point for the parameter estimations, we use the initial dataset (*SI*) for *P* parameters and *N* parameter estimation attempts: $SI_{P,N}$ (i.e. for the parameter and ranges identified in the previous step). We use a combination of design of experiments (central composite design [47], to explore the extremes of the parameter space) accomplished with a series of random numbers. In the model calibrations, the input parameters are varied in order to minimize the *NRMSE*, starting at the locations in the parameter space defined in the dataset $SI_{P,N}$. We look for optimal parameter estimations by using a MATLAB build-in functionality for constrained nonlinear optimisation, using sequential quadratic programming [48]. This algorithm varies the parameter values until the derivative of the objective function (i.e. the *NRMSE*) reaches values between zero and a pre-defined threshold level. This results in a dataset with calibrated parameter values that have a good (or best obtainable) fit with observations of energy use for the period 1970-2003: $SC_{P,N}$. This can be best imagined as the collection of local optima in the objective function landscape spanned up by the explored parameter space.

2.2.3 Analysis of calibrated parameter values
We analyse the series of calibrated sets of parameter values in $SC_{P,N}$ in several ways. First, the distribution of the calibrated parameter values over their range is analysed. Second, we plot the calibrated parameter values against the *NRMSE* (see Figure 3, upper graphs). Relations between parameters and the impact of parameters on the *NRMSE* can be numerically expressed by the (linear) Pearson correlation coefficient between parameters. We use this as the simplest indicator to express a relation between two parameters, although it does not capture non-linearity or the existence of multimodal distributions.

Based on this, behavioural sets of parameter values can be selected. The most straightforward method is based on the *NRMSE* value, for instance, one can decide to call sets of parameter values with *NRMSE* < 10% behavioural. An alternative, but less reproducible criterion is based on visual inspection of the parameter values and the observed and simulated time series of energy demand. In our analysis, we decided not to remove any sets of parameter values based on non-behavioural outcomes. However, we use the *NRMSE* (hence, behavioural/non-behavioural) to weight future projections that are derived from the different sets of parameter values.

2.2.4 Analysing the impacts of calibration uncertainty on future projections

To analyse the impact of different parameter values on future projections of the model, we use the series of calibrated sets of parameter values in $SC_{P,N}$ to run the model forward for the period 2003-2030 using a similar scenario on the model drivers (see Section 4.2). This leads to a range of projected future energy use, based on the different sets of parameter values. We analyse this in a frequency diagram of energy use in 2030 and weigh the frequencies in the diagram relative to the *NRMSE* of the parameter set that obtained the best fit to historic data in $SC_{P,N}$ (implicitly assuming that sets of parameter values with a better fit to historic data lead to more plausible future projections). The weight (*W*) that the *N*'th calibrated parameter set gets in the prediction ensemble is defined as the normalisation of the relative weight (*R*) of the parameter set to the best performing parameter set[2]:

$$W_N = \frac{R_N}{\sum_N R_N} \text{ where } R_N = \frac{NRMSE_{best}}{NRMSE_N} \qquad 2$$

In the remainder of this article, we apply this method to the transport sector energy use model of TIMER 2.0.

3 The TIMER 2.0 Energy Demand Model: parameters and ranges

The global energy model TIMER includes both demand and supply of energy [14, 15, 21]. Because of the many feedbacks, interactions and sub-modules, the TIMER model as a whole would be far too large and too complex to analyse the uncertainty from calibration. Therefore, we here confine the analysis to the sub-model that simulates the demand for energy on the basis of economic activity and autonomous and price-induced efficiency improvements.

In the TIMER model, energy use is first modelled as the annual demand for useful energy[3] (*UE*, in GJ/year, see Figure 1), which is converted to secondary energy use, using specific efficiencies for different fuels. Useful energy demand is modelled as function of four dynamic factors: structural change, autonomous energy efficiency improvement (*AEEI*), price induced energy efficiency improvement (*PIEEI*) and price-based fuel substitution. Thus:

---

[2] This measure does not hold in the unlikely situation that the model exactly reproduces historic data and the best obtained fit becomes zero.

[3] With useful energy defined as the level of energy services or energy functions, for instance a heated room or cooled food; conversion efficiencies are taken from statistics

$$UE_{R,S,F} = POP_R(t) \cdot X_{R,S}(t) \cdot Y_{R,S,F}(t) \cdot AEEI_{R,S,F}(t) \cdot PIEEI_{R,S,F}(t) \qquad \text{(GJ/yr) 3}$$

in which *POP* is the population (in persons), *X* is the per capita economic activity of a sector (in purchasing power parity (PPP), constant 1995 international \$/capita/yr), useful energy intensity (*Y,* in GJ/\$/capita) captures intra-sectoral structural change and the *AEEI* and *PIEEI* (dimensionless) multipliers represent autonomous and price induced efficiency improvements. The indices *R*, *S* and *F* respectively indicate region, sector and energy form (heat or electricity).

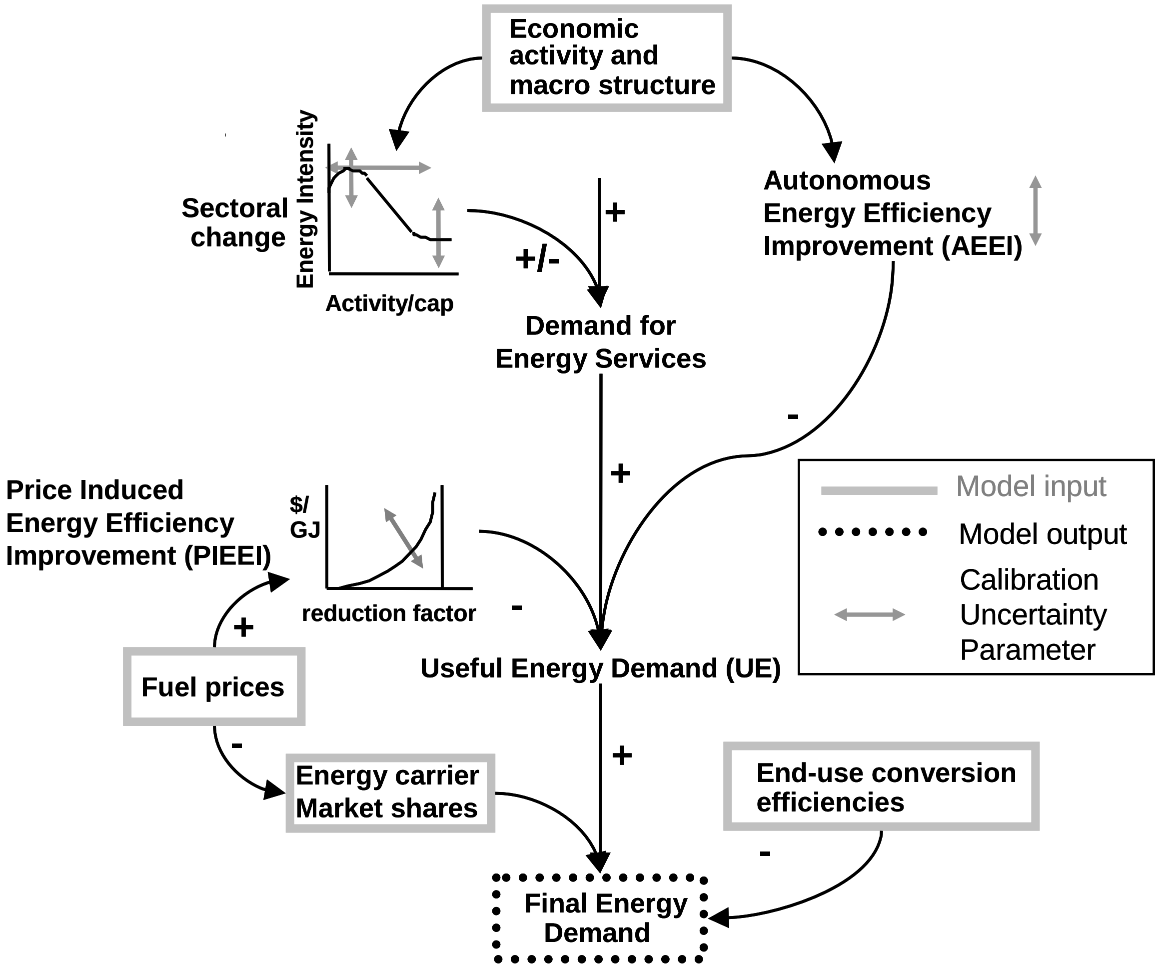

Figure 1: Overview of the TIMER energy demand model and identification of model inputs, output and parameters used to determine calibration uncertainty.

Statistical time series are available for two variables: economic activity and secondary energy use. Between these observable variables, the model tells a story of useful energy intensity (structural change) and autonomous and price induced efficiency improvements, aggregates that can hardly be measured in the real world. The multiplicative structure of this model leaves room for different behavioural sets of parameter values: for different implementations of the *UEI*-curve, *AEEI* and *PIEEI*, a similar result can be obtained for the observable value of final energy use.

The model distinguishes two forms of energy: electricity and fuels. In this analysis, we focus on the total demand for energy (i.e. the sum of all energy carriers); the fuel mix is assumed constant, calibrated to (historic) energy prices. We equate energy demand and energy use, as the statistical data are assumed to have satisfied demand in a state of economic equilibrium on an annual basis; hence, we do not consider the concept of latent (or unfulfilled) demand for energy (which is relevant for low-income regions).

3.1 Energy intensity curve

From energy analysis [for instance 49, 50, 51] it is known that:

1. there is a tendency for total energy use to increase with population and economic activity;
2. in many countries, energy intensity tends first to rise then decline; this takes place at the level of the whole economy but also at sector level. This pattern is often referred to as the Environmental Kuznets Curve [for discussions see 52, 53]. It is usually explained from a mix of saturation and dematerialization, i.e. change to more value-added per unit of energy input [for analyses see e.g. 54, 55-57]. The income level at which such a maximum in intensity is reached tends to decrease over time – interpreted as the collective dissemination of energy-innovations and of learning-by-doing [58, 59].

Assuming that this also holds for useful energy, these stylized facts are represented in the model equation for useful energy intensity ($Y_{(t)}$) in the form of a (asymmetric) bell-shaped function of the sector-specific per capita economic activity. For each region (R), sector (S) and energy form (F) at time t, this can be expressed as[4]:

$$Y_{(t)R,S,F} = Y_0 + \frac{1}{\beta \cdot X_{(t)} + \gamma \cdot X_{(t)}{}^{\delta}} \quad (4)$$

with $X_{(t)}$ the sectoral economic activity per capita and β, γ and δ parameters (of which δ is negative to maintain a bell-shaped form, see Figure 2). All parameters in this equation are defined per region, sector and energy form.

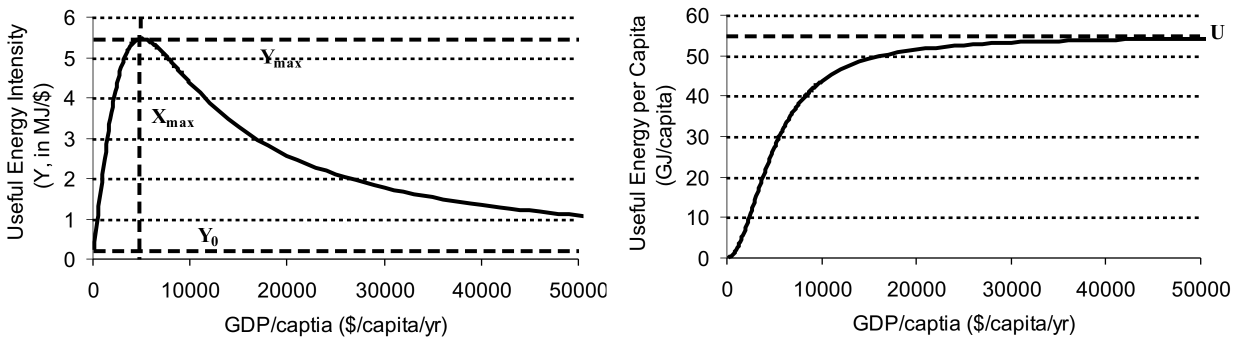

Figure 2: *UEI* curve (left) and useful energy use per capita (right) for hypothetical parameter values

The flexible formulation of this curve implies also a high sensitivity to parameter values. From an energy point-of-view, some reasonable constraints can be made to limit the potential parameter space to a relevant subspace and to shape the curve on the basis of understandable quantities:

[4] This bell-shaped curve can also be written in terms of elasticity with GDP/capita as is common for energy use, but for the transport sector it can also be done for vehicle ownership [60].

1. the activity level at which the maximum occurs, X_{max}, can be estimated from regional energy use data. This has the risk of cyclical reasoning, because one draws conclusions from the observed data which are to be explained; one should do it only for datasets (regions, periods) where presumably end-use conversion efficiency has hardly changed[5].
2. The second term of the curve may be related to the saturation level of useful energy per capita per year at high income levels (U, see Figure 2, right graph). This saturation level can be based on sector and region specific features such as climate or population density.
3. Y_0 can be interpreted as the ultimately lowest energy-intensity of sectoral activity (in $/GJ) in the both limits $X \to \infty$ and $X \to 0$.

Values and ranges for the parameters β, γ and δ can be derived from these constraints, in combination with the assumption that the curve should be forced through one observed reference point, defined as (X_{ref},Y_{ref}), which can be any year in the period 1971-2003[6] (see Appendix 3). Each implementation of the curve (as function of X_{max}, U and Y_0) can be characterised by its maximum energy intensity, i.e. the top of the curve (Y_{max}, see Figure 2), derived as:

$$Y_{max} = Y_0 + \frac{U}{X_{max}} \cdot \frac{\delta}{(\delta - 1)} \qquad 5$$

This allows a consistent set of parameter choices, for which these three key variables have to be investigated. We first establish suitable prior ranges for the variables X_{max}, U and Y_0 and translate these into values for the curve parameters β, γ and δ. The range for values of X_{max} and U in the parameter estimation process is defined as 10% broader than the maximum and minimum values applied in earlier (manual) calibrations of the TIMER model and is shown in the appendix (Table A 1). Conceptually, Y_0 is only limited by the value of Y_{ref}, because the ultimately lowest energy intensity cannot be higher than the observed historic energy intensity[7]. However, if Y_0 equals Y_{ref}, the second term of eqn. 4 (the 'curve' itself) would be irrelevant and the model would become linear. To force the model to explain the major part of energy intensity from the curve, we assume Y_0 to be lower than 20% of Y_{ref}.

3.2 Autonomous Energy Efficiency Improvement (*AEEI*)

The continuous decline of energy intensity due to technology change is represented in the TIMER model by the autonomous energy efficiency improvement (*AEEI)* multiplier. Marginal *AEEI* is defined as fraction of economic activity growth [62]:

$$AEEI_{marg}(t)_{R,S} = F_S \cdot \left(\frac{GDPpc_{(t)R}}{GDPpc_{(t-1)R}} - 1\right) \cdot 100 \qquad (\%/yr) \quad 6$$

---

[5] For instance, transport energy efficiency in the USA, where improved fuel efficiency is offset by vehicle mass [61]

[6] In our model implementation this is the year 2003, the latest year of the calibration period

[7] Since energy intensity is defined in energy use per (monetary) unit of GDP, there is no theoretical or thermodynamic limit to the value of Y_0.

with F_S a sectoral specific fraction of economic activity growth. The vintage structure modelling for energy using capital in TIMER determines that the current *AEEI* is the weighted average of the marginal *AEEI* over the capital life time [14]. This means that rapid economic growth leads to a faster decline in *AEEI*, due to both increased decline in the marginal *AEEI* and a larger share of the capital stock that is relatively new [21]. In case of economic decline, the marginal *AEEI* cannot become negative and is limited to zero.

The parameter that has to be estimated for *AEEI* is the fraction of GDP growth (F_S). For the percentage of annual *AEEI* a range of 0.2-1.5% per year is suggested by experts consulted by Van der Sluijs et al (2001). This is used to establish a range for F_S by using the average annual regional GDP per capita growth over the period 1970 to 2003 (see Table A 1 and A2). In the above formulation, *AEEI* is related to the general economic growth in a region and not to a sector specific activity indicator, such as value added. During economic structural change, some sectors will grow faster or slower than the average economic growth, but this can be accounted for by using a sector specific value for F_S. In the presentation of the results, *AEEI* is expressed as the average percentage of annual sectoral efficiency improvement, based on the average historic regional GDP per capita growth for the period 1971-2003.

3.3 Price Induced Energy Efficiency Improvement (*PIEEI*)

The *PIEEI* reflects that with increasing energy prices end-users take measures to use energy more efficiently. The description of *PIEEI* in TIMER is based on an assumed energy conservation supply-cost-curve. This curve describes the increasing marginal cost of energy conservation. By comparing the gains of efficiency improvement (annual saved energy times payback time and energy prices) to the cost of investments, an optimum can be found. As such, there are three main factors that determine the level of energy efficiency: first, the form of the supply-cost-curve; second, the value of the pay-back time and third, learning-by-doing of energy efficiency technology. In the TIMER model, the energy conservation supply-cost-curve can be compared to bottom-up technology data [14] but is modelled as an aggregated stylized function. The optimal level of energy efficiency (*E*, as fraction of total energy use) is defined as the point at which marginal energy conservation measures still yield net revenue:

$$E_{R,S,F} = M_{R,S,F} - \frac{1}{\sqrt{M_{R,S,F}^{-2} + \frac{C_{R,S,F} \cdot T_{R,S,F}}{S_{R,S,F} \cdot I_{R,S,F}}}} \qquad 7$$

in which *M* is the maximum potential price-induced efficiency improvement (as fraction of total frozen energy use), *C* the sectoral average costs of useful energy (in *$/GJ)*) and *T* the (apparent or desired) pay-back time (*in years*). *I* is the dimensionless factor with which the cost curve declines as a result of learning-by-doing. The scaling parameter S is used to scale the curve to the sector-specific costs of useful energy. The *PIEEI* on marginal capital investments, which is used in eqn 3, is a dimensionless multiplier defined as: $1-E_{R,S,F}$. Vintage modelling of energy demand capital delays the impact of the *PIEEI*, as the current *PIEEI* is the weighted average of the marginal *PIEEI* over the capital life time.

In the parameter estimation procedure we vary values of payback time (T) and the learning parameter (I)[8] using historic energy prices. From equation 7 it can be seen that both a higher payback time and a lower learning multiplier lead to more efficiency improvement. These two parameters are linearly interchangeable, but the overall *PIEEI* value provides more accessible information. Therefore we express these two *PIEEI* related parameters together as the cumulative efficiency improvement up to the year 2003. The ranges for the learning and payback time parameters in the experiment are shown in Table A 1.

4 Application to transport energy demand modelling

We tested our method to identify multiple behavioural sets of parameter values to the transport sector energy demand sub-model of TIMER. We performed 100 parameter estimation attempts per region (so N=100 in $SI_{P,N}$ and $SC_{P,N}$). First, Section 4.1 discusses the results of calibration to historic data (i.e. step B and C, explained in Section 2.2). Second, Section 4.2 explores the impact of the calibrated sets of parameter values on future projections (step D of the procedure).

4.1 Calibration to historic data

Energy consumption in the transport sector is rapidly increasing and might become the major final energy use in the near future. We tested our method to identify multiple behavioural sets of parameter values to the transport sector energy demand sub-model of TIMER. We performed 100 parameter estimation attempts per region (so N=100 in $SI_{P,N}$ and $SC_{P,N}$). If we only look at the *NRMSE*, an error of less than 10% between model results and observations is obtained for the regions USA, Europe and India; the results for Brazil, Russia and China are less good (Figure 10, Appendix 2).

Table 1: Linear correlation coefficient of calibrated parameter values

USA	*UEI* (Y_{max})	*AEEI*	*PIEEI*	**Europe**	*UEI* (Y_{max})	*AEEI*	*PIEEI*
AEEI	-0.17	-		*AEEI*	0.33	-	
PIEEI	0.77	-0.69	-	*PIEEI*	-0.54	-0.86	-
NRMSE	-0.71	-0.46	-0.26	*NRMSE*	-0.52	-0.57	0.85
India	*UEI* (Y_{max})	*AEEI*	*PIEEI*	**China**	*UEI* (Y_{max})	*AEEI*	*PIEEI*
AEEI	-0.52	-		*AEEI*	-0.35	-	
PIEEI	0.06	-0.23	-	*PIEEI*	0.69	-0.24	-
NRMSE	0.91	-0.75	0.17	*NRMSE*	0.91	-0.48	0.82
Brazil	*UEI* (Y_{max})	*AEEI*	*PIEEI*	**Russia**	*UEI* (Y_{max})	*AEEI*	*PIEEI*
AEEI	0.16	-		*AEEI*	0.54	-	
PIEEI	-0.15	-0.63	-	*PIEEI*	-0.26	-0.85	-
NRMSE	-0.04	0.37	-0.32	*NRMSE*	0.53	0.96	-0.76

[8] Alternative parameters to vary would be the maximum improvement level (M) or the steepness (S). However, M is based on a theoretical maximum efficiency improvement expressed in energy intensity terms. This is a useful parameter to explore, but has more impact on future projections than on historic calibration. The steepness parameter (S) is used to scale the *PIEEI* curve to the useful energy costs per sector and is therefore not useful to vary.

4.1.1 Europe and USA: equifinal sets of parameter values

Final energy use of the transport sector in both the USA and Europe shows an increasing trend, with temporary slower growth after 1980 due to oil-price increases. Generally, the model simulates transport energy use in Europe quite well with a best *NRMSE* of 2.8% (Figure 10). Also, the fluctuations during the 1980s are well-captured (Figure 3, lower graphs). The calibrated parameter values vary over a wide range and only *U*, *AEEI* and *PIEEI* have relations with the *NRMSE*, although X_{max} is generally high and Y_0 is low (Figure 3, upper graphs). About 5% of the sets of parameter values have an *NRMSE* higher than 10% and can be identified as outliers on the basis of the parameter values. Generally, the parameter values follow two model stories: the best-fitting sets of parameter values have high values for *AEEI* (>1 %/yr) and no *PIEEI*; a second group has low values for *AEEI* and high *PIEEI*. The high correlations between *AEEI*/*PIEEI* and *NRMSE* (Table 1) also indicate these different options for parameter values.

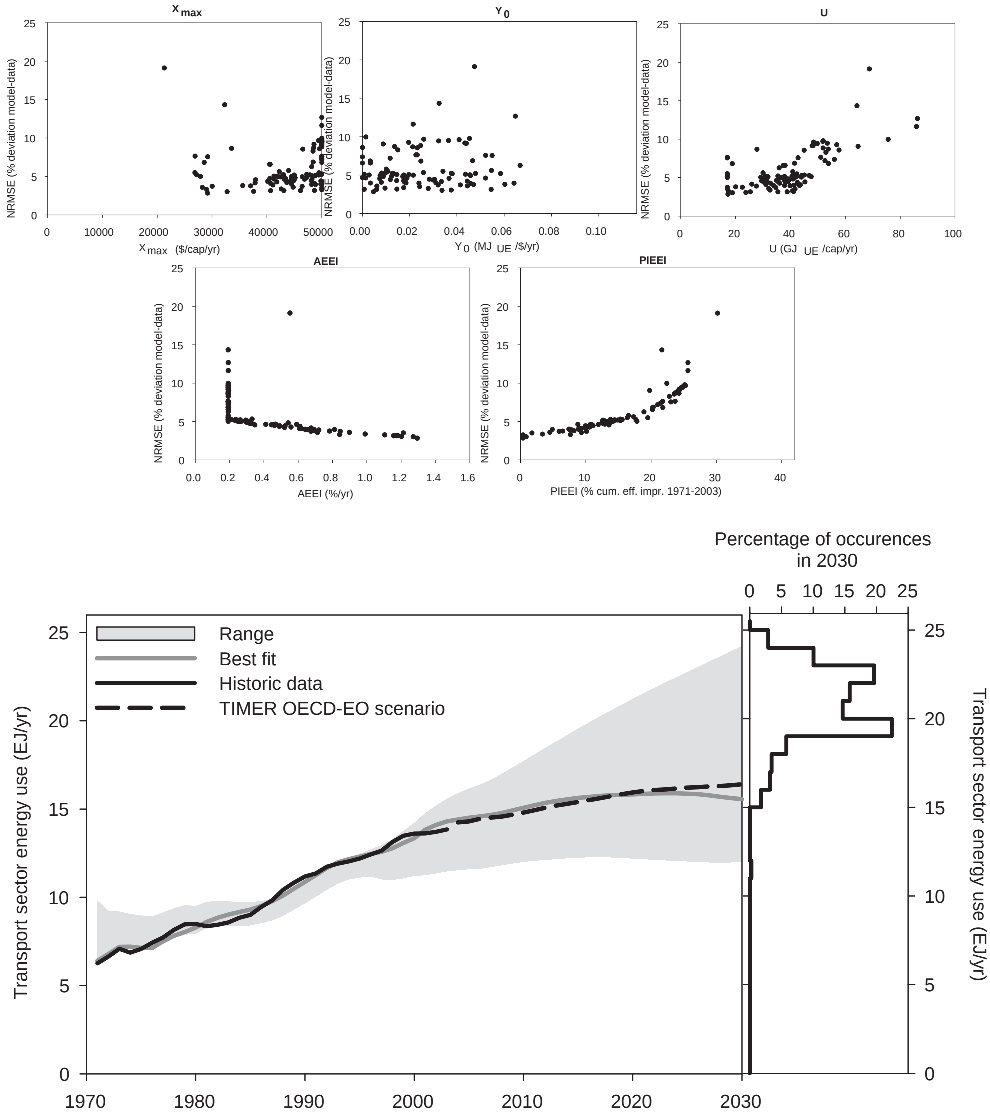

Figure 3: Upper graphs: plot of 100 calibrated sets of parameter values for transport sector energy use in Western Europe. Each dot represents a calibrated parameter value for the period 1971-2003. Lower graphs: historic and projected transport energy use for Western Europe up to 2030 (left graph) and histogram (right graph) of energy use in 2030 using the *NRMSE* as weighting factor. Projections based on OECD-EO scenario inputs and calibrated sets of parameter values

The best *NRMSE* values for the USA is 3.5% (Figure 10), and also here the model simulates both long-term and short-term trends (Figure 4). The calibrated parameter values show hardly any relation with the *NRMSE*: the distributions of Y_0, U, *AEEI* and *PIEEI* involve a wide range are rather flat to the *NRMSE* (with the exception of about 10% outliers). In general, however, we can distinguish two different groups of behavioural sets of parameter values as well. The relation between *AEEI*/*PIEEI* and *NRMSE* is opposite to that of Europe (Table 1): the best fitting sets of parameter values have a low *AEEI* and high *PIEEI*, a second group has high *AEEI* and low *PIEEI*. This implies that the USA is more sensitive to energy price changes than Western Europe. All *UEI*-curve solutions for the USA have a top at low income levels: useful energy intensity has been declining in the period 1971-2003. The negative correlation between Y_{max} and *NRMSE* (Table 1) shows that the more behavioural sets of parameter values have (historically) higher energy intensity.

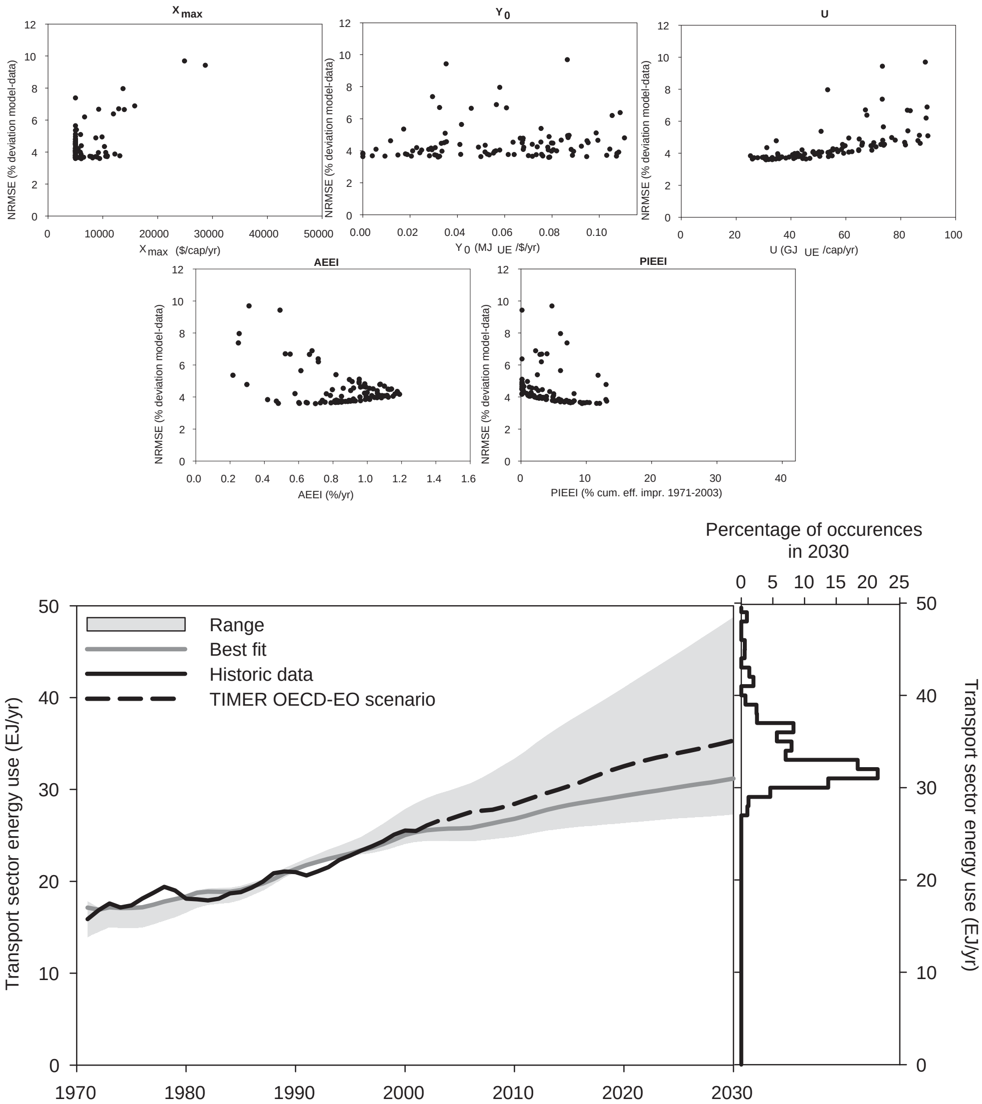

Figure 4: Upper graphs: plot of 100 calibrated sets of parameter values for transport sector energy use in the USA. Each dot represents a calibrated parameter value for the period 1971-2003. Lower graphs: historic and projected transport energy use for the USA up to 2030 (left graph) and histogram (right graph) of energy use in 2030 using the *NRMSE* as weighting factor. Projections based on OECD-EO scenario inputs and calibrated sets of parameter values

These results indicate that the model performs quite well in simulating energy use in the USA and Western Europe, regions that have been important during the model development phase. However, they also indicate that distinguishing between the drivers of energy efficiency improvement, technology (*AEEI*) vs. prices (*PIEEI*), is difficult and maybe even questionable; especially since the reaction of *PIEEI* on energy prices is slow, due to delays from capital turnover.

4.1.2 Brazil: fluctuation of economy and energy use

Brazilian GDP per capita and transport sector energy use have been fluctuating during the period 1971-2003. This complicates model calibration for this region, which can be seen in high *NRMSE* values: the best *NRMSE* is 10.6% (Figure 5). However, the simulation might be called behavioural in following the long-term trends. The calibrated values of all parameters are distributed rather evenly over the range and show hardly relations with the *NRMSE* (see also Table 1). Analysis of the parameter values shows that *AEEI* and *PIEEI* are negatively correlated and rather interchangeable.

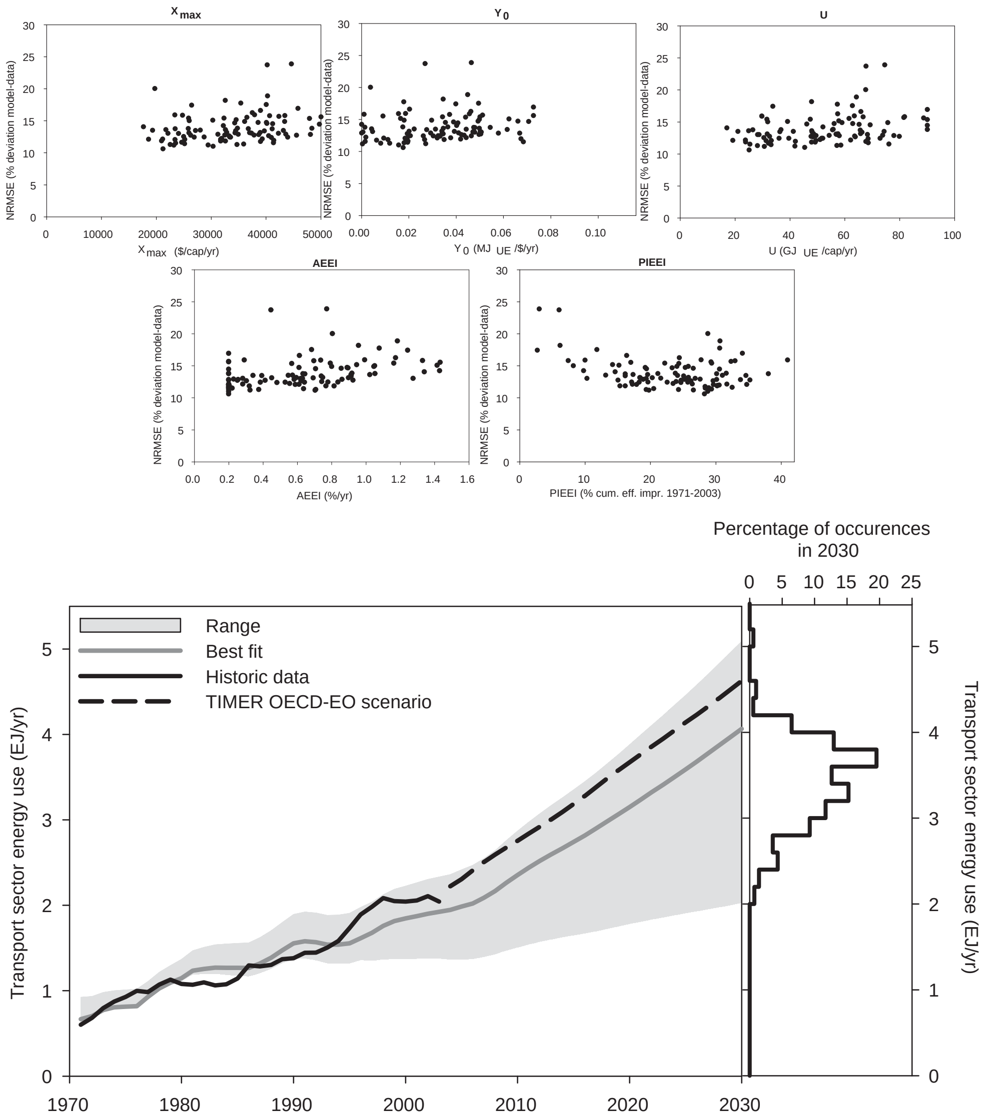

Figure 5: Upper graphs: plot of 100 calibrated sets of parameter values for transport sector energy use in Brazil. Each dot represents a calibrated parameter value for the period 1971-2003. Lower graphs: historic and projected transport energy use for Brazil up to 2030 (left graph) and histogram (right graph) of energy use in 2030 using the *NRMSE* as weighting factor. Projections based on OECD-EO scenario inputs and calibrated sets of parameter values

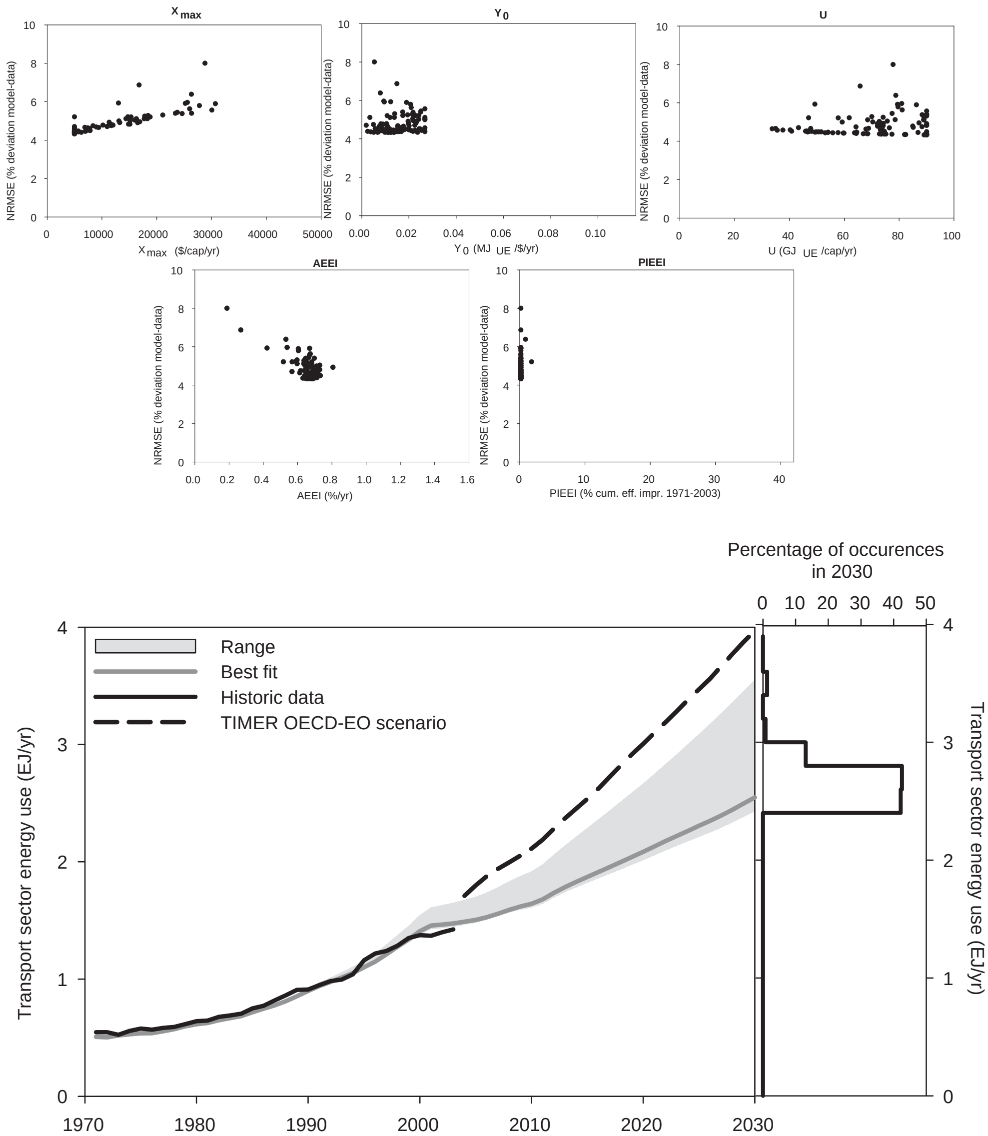

Figure 6: Upper graphs: plot of 100 calibrated sets of parameter values for transport sector energy use in India. Each dot represents a calibrated parameter value for the period 1971-2003. Lower graphs: historic and projected transport energy use for India up to 2030 (left graph) and histogram (right graph) of energy use in 2030 using the *NRMSE* as weighting factor. Projections based on OECD-EO scenario inputs and calibrated sets of parameter values

4.1.3 India and China: exponential growth

In the historic 1971-2003 period, energy use in the transport sectors of India and China has been growing exponentially. The Chinese data include some periods of decreasing energy use (1978-1980, 1990 and 1994), which makes the curve more difficult to simulate, especially before 1990. This shows up clearly in the *NRMSE* values: the best value for India is 4.3%, for China all calibrated parameter sets are between 17.18% (Figure 10). The main source of this high number is in mismatch in the period 1970-1990, but still, all sets of parameter values are generally behavioural in the sense that they simulate the exponentially increasing trend in the data. Both regions are simulated best with constant useful energy intensity (in the 1971-2003 GDP/capita range), *AEEI* of about 1 %/yr and no *PIEEI*. In relation to the *NRMSE*, both regions show a better fit with low values for X_{max}, high *U* and high *AEEI* (Figure 6 and Figure 7). There are no systematic relations between parameters (Table 1), except between maximum energy intensity (Y_{max}) and *NRMSE* (i.e. a lower Y_{max} leads to a better fit).

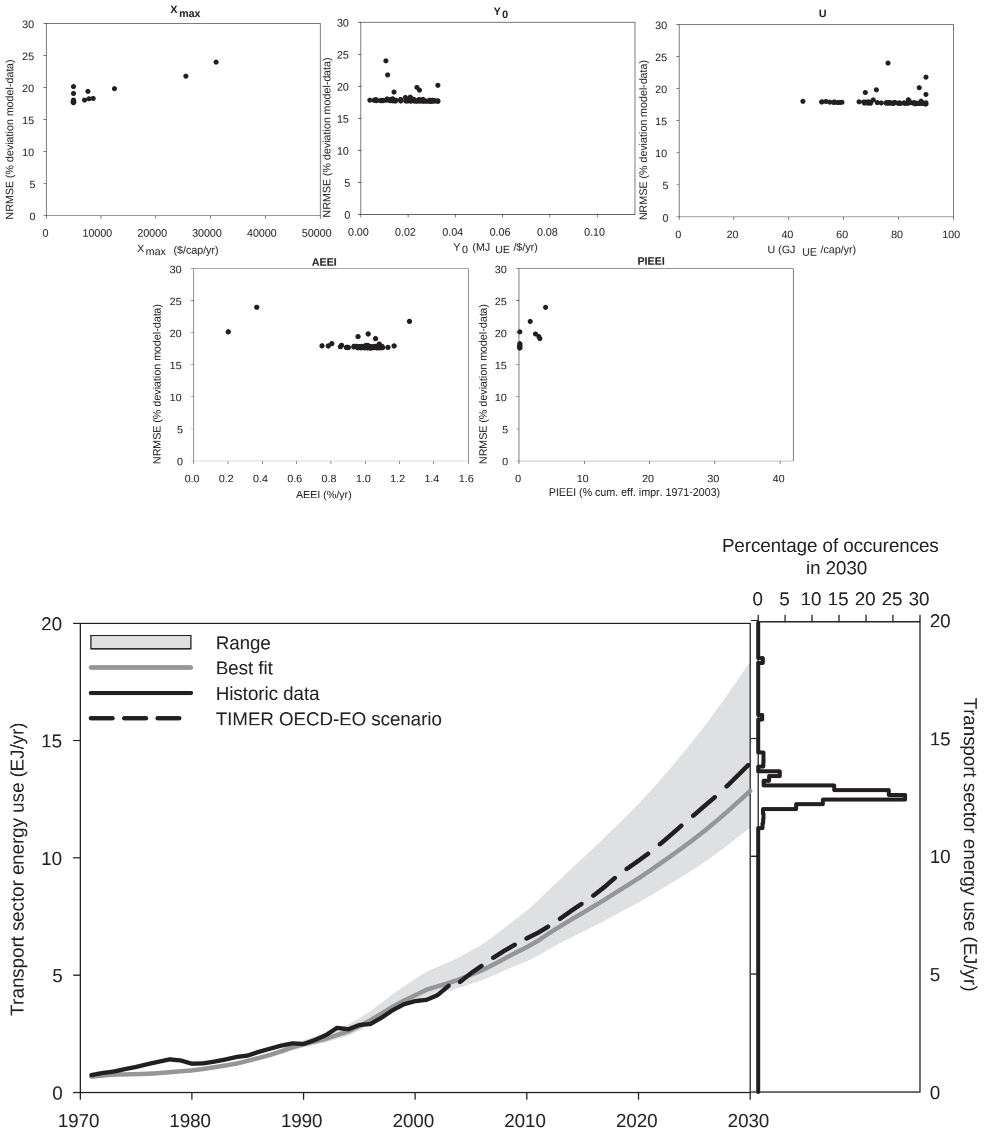

Figure 7: Upper graphs: plot of 100 calibrated sets of parameter values for transport sector energy use in China. Each dot represents a calibrated parameter value for the period 1971-2003. Lower graphs: historic and projected transport energy use for China up to 2030 (left graph) and histogram (right graph) of energy use in 2030 using the *NRMSE* as weighting factor. Projections based on OECD-EO scenario inputs and calibrated sets of parameter values

Several issues play a role in estimating the model parameters for the regions of India and China. With respect to the *UEI*-curve, these regions have rather narrow absolute GDP per capita ranges between 1971 and 2003 and they are forced to be below the top of the *UEI*-curve (the lower bound of X_{max} is 5000 \$/capita/yr). Historically, useful energy intensity might have been constant, but it can be questioned whether such implementation of the model is representative outside the range of historically observed economic activity. Another source for the model error in India and China (but also Brazil) might be that the TIMER model does not capture some important concepts that are relevant for developing countries (e.g. urban/rural divide and unequal income distribution [see 63]) and ignores the role of specific technologies (e.g. modal split).

4.1.4 Russia: dealing with (ir-)reversibility

The Russian combination of economic growth and decline within a range of 5000-9000 international $ per capita puts the model and its parameterisation to the test. Energy use in the Russian transport sector shows a sharp break of the increasing trend after the fall of the Soviet Union, with energy use decreasing from 4.3 EJ/yr in 1990 to 2.8 EJ/yr in 1997 (Figure 8). The model appears to be rather able to simulate historic Russian transport energy use with best *NRMSE* values of 11.6% (Figure 10). There are relations of Y_0 and *U* with the *NRMSE*, and the values of these parameter are scattered over the range (Figure 10). However, lower X_{max}, low *AEEI* and high *PIEEI* lead clearly to a better fit.

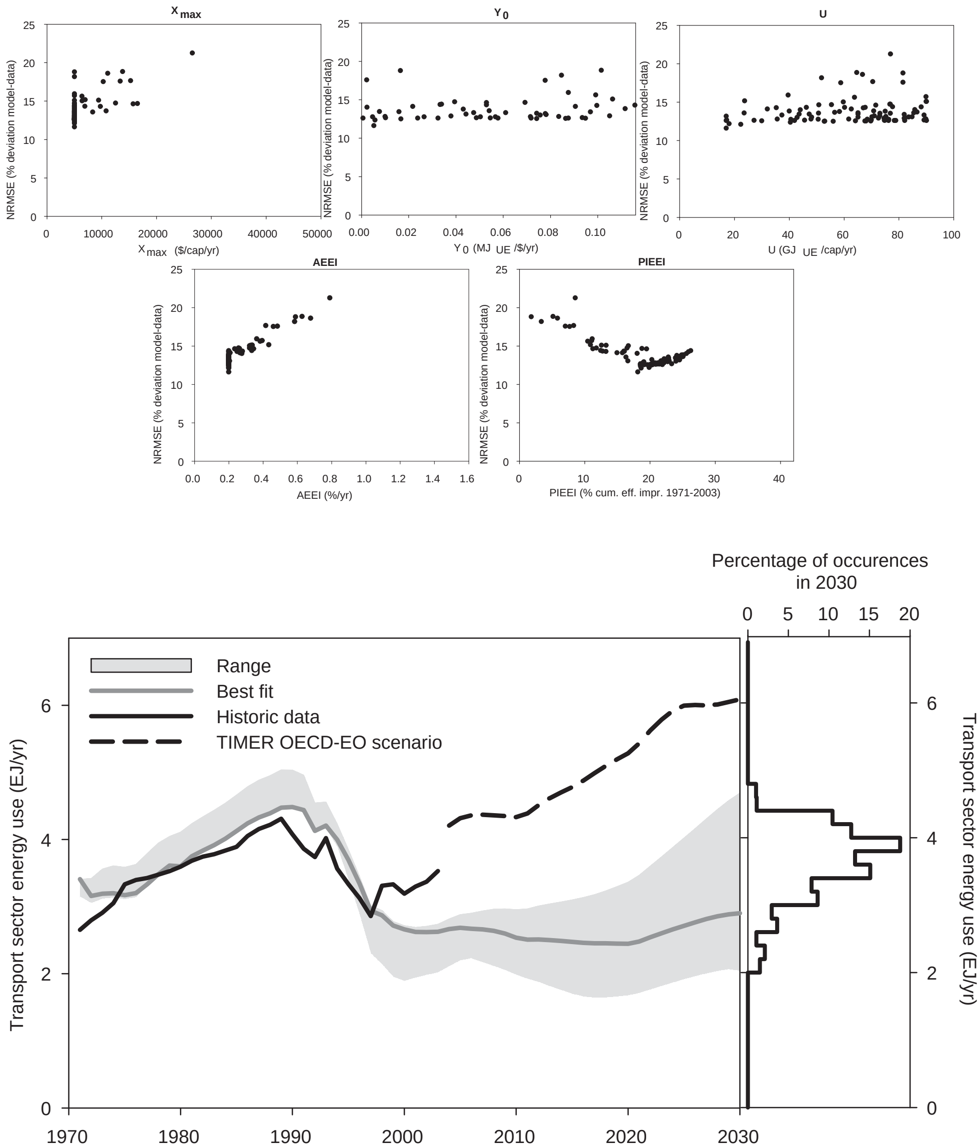

Figure 8: Upper graphs: plot of 100 calibrated sets of parameter values for transport sector energy use in Russia. Each dot represents a calibrated parameter value for the period 1971-2003. Lower graphs: historic and projected transport energy use for Russia up to 2030 (left graph) and histogram (right graph) of energy use in 2030 using the *NRMSE* as weighting factor. Projections based on OECD-EO scenario inputs and calibrated sets of parameter values

4.2 Impact on future projections

To determine the influence of the different sets of parameter values on future projections of the model we calculate the projected energy demand in 2030, using scenario inputs of the OECD environmental outlook scenario [OECD-EO, described in detail in 20, 64, 65].These scenario inputs include projections for GDP, sectoral value added and population. The OECD-EO is a baseline scenario without new policies on economy and environment, in which energy use is based on moderate projections of population and economy. In this analysis we use the same energy prices for all forward calculations; these prices correspond with the default implementation of this scenario[9].

The TIMER model was used in its original setting within the OECD-EO study to project development of the future energy system, including energy transport demand. These projections can be very different from the current as 1) TIMER modellers have focused in model calibration not only on the performance of a single region but aimed to have similar parameter settings for different regions and 2) have calibrated to the model projections also against the IEA World Energy Outlook.

The projections of future transport sector energy use in Western Europe in 2030, based on the calibrated sets of parameter values, show a slowly increasing energy use toward 15-25 EJ/yr. In 2030, these projections vary over a wide range (Figure 3); expressed as percentage around the 'best fit' in 2030, this range amounts 79% (Table 2). However, from the distribution of projections (weighted to the *NRMSE*-value, see Section 2.2.4) it can be seen that the lower bound is heavily influenced by a singly outlier. The most behavioural sets of parameter values and the OECD-EO scenario are on the lower bound of this range. However, most sets of parameter values (weighted to the *NRMSE*) project an energy use of 19-23 EJ/yr in 2030, higher than the best fitted parameter set and the OECD-EO scenario.

Table 2: Correlation coefficient between calibrated parameter values (for both branches of δ) and projected energy use in 2030 for the transport sector

	UEI (Y_{max})	*AEEI*	*PIEEI*	*Range in 2030*
USA	-0.68	-0.50	-0.27	69%
Europe	0.65	-0.32	-0.15	79%
India	0.65	-0.78	0.41	44%
China	0.26	-0.97	0.06	55%
Brazil	0.11	-0.49	-0.34	75%
Russia	-0.22	-0.67	0.41	91%

A second issue of interest is which parameters mainly influence the projected energy use. This is explored in Table 2, showing the correlation between the calibrated parameter values and projected energy use in 2030. However, for Europe (and most other regions) there are no strong correlations, but we can analyse the direction. Generally, it can be stated that higher energy intensity and lower *AEEI* lead to higher projections for European energy demand for transport.

[9] Normally energy prices for future projections are calculated endogenously in the model based on depletion and learning. In this way, different energy demand projections lead to different energy prices, causing different market shares of fuels and other values for end-use-efficiency and *PIEEI*

For the USA, model-projections based on the calibrated sets of parameter values lead to a wide range of energy use in 2030: 30-50 EJ/yr, or 69% around the 'best fit' (Figure 4). However, this upper limit of this range is mainly determined by outliers that have little weight in the histogram. The best fitting sets of parameter values, which account for more than 50% of the weighted occurrences, project energy use in the range of 30-35 EJ/yr. The OECD-EO scenario is slightly above this range. Also for the USA, there is hardly any correlation between parameter values and projected energy use in 2030. The correlation with Y_{max} is negative, because the top of the *UEI*-curve is calibrated at low income levels. Correlation with the *NRMSE* is strong, indicating that a better fit leads to lower energy use projections.

For Brazil, which currently has a transport sector energy use of 2 EJ/yr, the projections vary in the range of 2-5 EJ/yr, with a peak of occurrences at 3.5-4 EJ/yr (Figure 5). The 'best fit' and the OECD-EO scenario project a somewhat higher energy use. Correlations between calibrated parameter values and projected energy use are weak; the negative correlations with efficiency improvements are the strongest.

Forward calculations for India indicate an increasing transport sector energy use from 1.5 EJ/yr in 2003 to 2.5-3 EJ/yr in 2030 (Figure 6). Relative to the 'best fit', the range for India is narrow: only 44%. The OECD-EO scenario is clearly above the range of projections, leading to 4 EJ/yr in 2030. Projected energy use correlates strongest with *AEEI* and Y_{max}: higher *AEEI* (and thus, better fit) leads to lower projected energy use (Table 2).

Projections for energy use in the Chinese transport sector in 2030 vary over a range of 11-19 EJ/yr, but a clear peak exists at 12.5 EJ/yr (Figure 7). The upper limit of the range is (>14 EJ/yr) is determined by two outliers. The 'best fit' is located in the peak and the OECD-EO scenario projects a slightly higher energy use of 14 EJ/yr. *AEEI* is the most decisive parameter for future energy use, with negative correlation of 0.97 (Table 2).

Future projections for Russia show that, although the historic calibrations are reasonable, the deviation between model results and data after 1997 has a crucial impact on future projections (Figure 8). The OECD-EO scenario projections are more in line with the historically increasing trend, but it seems that historic calibration can hardly be used as a ground for future projections for this regions. A solution might be to manually calibrate the model, specifically looking for sets of parameter values with a better fit in later years. Another option is to redefine how the model structure copes with the process of historic economic decline.

In general, the variation in parameter values accounts for quite some uncertainty in future projections (see the ranges of 44-90% around the 'best fit'). However, it is hard to attribute this uncertainty to single parameters. In most cases the *AEEI* is the most important model parameter for future projections, followed by the intensity curve. *PIEEI* seems less influential, although this is also related to slowly increasing projected energy prices.

4.3 General results and trends

The general results that emerge from this analysis are summarised in Figure 9, showing transport sector final energy intensity (i.e. transport sector final energy use per unit of GDP) and annual transport sector energy use per capita. With respect to energy intensity, major differences exist between countries, both on absolute levels and direction of trends. It is complicated to distinguish a general pattern in the results: China and India show maxima at low income levels, and historically declining energy intensities. Brazilian energy intensity is higher and historically rather stable. European energy intensity shows a maximum at GDP levels of about 20000 $/capita/year; final energy intensity of the USA is significantly higher than all other regions, but rapidly decreasing. Russia is an exceptional case: historically high energy intensity and low future projection (though these seem not very likely, see Figure 8).

Differences in per capita use of energy for transport are also outspoken. The USA shows a rather stable pattern at 80-90 GJ/capita per year both historically and in future projections. Europe increased historically from 20 to 40 GJ/capita, but is projected to stabilise. India has the current and projected lowest energy use per capita of the three low-income regions, whereas Chinese per capita energy use for transport is projected to increase from 4 to 8-12 GJ/yr. Also here, Russia is an exceptional case with a rapid decline in energy use. For the developing regions, the European level of energy use seems more likely than the high level of the USA.

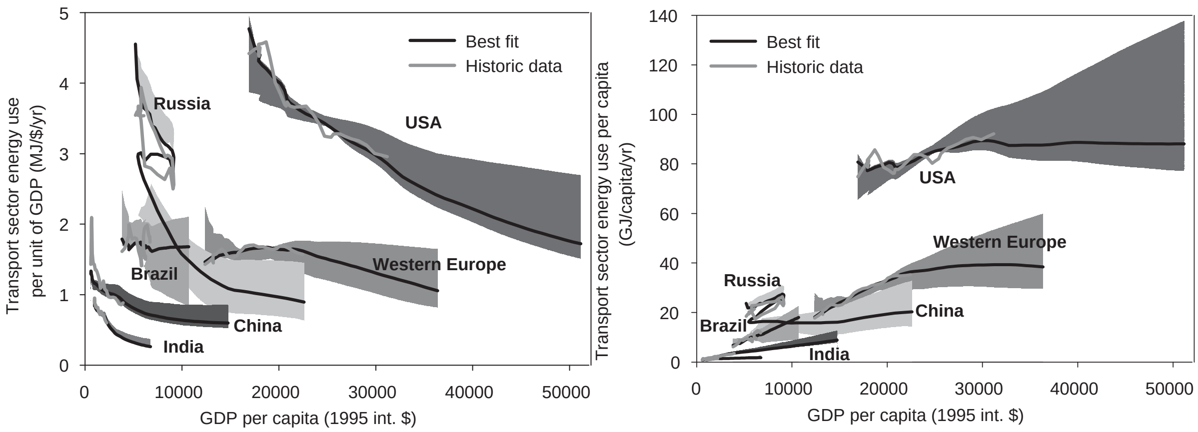

Figure 9: overview of transport sector final energy intensity (left graph, transport energy use per unit of GDP in PPP) and final energy use per capita (right graph) for all regions (except Russia) versus GDP/capita

The method applied in this paper calibrates the behaviour of each region individually to observed data. However, regions do not develop totally independently (e.g technology to improve efficiency is likely to be coupled between different regions) and especially the intensity curve (as described in Section 3.1) does originate from comparing different regions (cross regional data) [14, 57]. For instance, transport energy use in India and China is calibrated against a period in which car ownership was low and (motorised) two-wheelers were the major transport mode; a possible future rise in car ownership and air-transport cannot be foreseen in these data. Therefore, a further step in the analysis would be to restrict the allowed parameter space in different regions, as a function of the values chosen in other regions. The current method leads to somewhat low energy demand trends in Russia, India and

China compared to other projections (represented here by the trends for the OECD-EO).

5 Method evaluation

Several remarks can be made about the presented method to identify variation in model calibration parameters. Because the method applies an optimisation-algorithm to minimise the error between model results and data, it does not guarantee the identification of the total fit-landscape. Especially if the fit-landscape is flat this algorithm identifies the best-fitting (local) optimum, possibly ignoring other well-fitting sets of parameter values that have a slightly higher NRMSE. This indicates that the uncertainty from equifinality on forward projections might be larger than estimated in this study. A detailed Monte Carlo sampling analysis would guarantee that the whole fit-landscape is identified. However, we found in early stages of this analysis that equifinality sometimes takes place within very small ranges of the parameter values. Hence, the sampling has to be very detailed in order not to overlook the relevant parameter values, driving up calculation time. We used optimisation to efficiently scan the parameter space, and partly overcome this issue by initialising the parameter estimation process from many different locations in the parameter space (including 'design of experiments' to initialise at the corners of the parameter space). However, advanced adaptive sampling methods (see for instance Hendrix and Klepper [66]) might be better able to identify the full range of equifinality.

For this model we chose 100 different initialisations, balancing between calculation time and size of the database. Analysis of the results shows that for this model the shape of the distribution of the parameters and the *NRMSE* did not change significantly after 60 to 80 parameter estimation attempts. We expect this to be specific for each model. If this automated calibration procedure would be applied to another model, convergence of the *NRMSE* and the shape of the parameter distributions should be monitored to see whether enough initialisations have been chosen. It is clear that the method also identifies outliers, cases in which the optimisation-algorithm is terminated at relatively high *NRMSE* values. In the analysis that we performed, about 5-10% of the calibrated sets of parameter values could be identified as outliers. We conclude that the estimation technique performs well and most of the identified variation can be attributed to the model at hand.

In the error model that we use, we oversimplified by attributing the difference between modelled and observed values completely to the parameter error. One could extent the method towards more focus on measurement error in the observation, for instance by adding white noise to the calibration variable, or input and boundary condition error. In the specific case of TIMER, an error distribution on the reference energy intensity for the *UEI*-curve might deal with data-error and allow a broader range of sets of parameter values to be behavioural with the data. Another issue in the TIMER case is that parameter error and model structure error can hardly be separated, because the parameters related to the *UEI*-curve, can change the functional form of the model dramatically (e.g. from bell-shaped to linear).

The development of the described method is inspired by the concept of equifinality, developed by Beven based on his experiences with the GLUE methodology. The GLUE methodology has recently been subject of a scientific debate on its consistency with Bayesian statistics. A major criticism on GLUE was its application of ‘less formal likelihood’ measures; this may imply that it looses the learning properties of the Bayesian approach, leading to ‘flat’ parameter posterior densities and thus equifinality is build in the methodology [67, 68]. In response, it has been argued that if strong assumptions about the error model cannot be justified, GLUE provides a reasonable alternative [69]. The method applied here differs from both Bayesian updating and GLUE, because it does not apply sequential Monte Carlo analysis. Moreover, it also has elements of nonlinear regression methods like PEST and UCODE, in that its purpose is to identify ‘peaks’ in the fit-landscape. Therefore, we conclude that this discussion does not apply to this method.

6 Discussion, conclusion and implications

A method was developed to identify sets of parameter values that perform reasonably against historic data. Energy use modelling knows many scientific paradigms and traditions, which lead to different interpretations of past and present and to different expectations of the future. Even within one model, several options may exist on how to interpret the past and current situation. We developed a method to identify the range of sets of parameter values that perform reasonably against historic data and analyse the impact of these different calibrations on future projections. The essence of this method is that by varying several essential parameter values, we search to minimise the error between model results and observations. By repeating this parameter estimation procedure, starting from different locations in the parameter space, we were able to identify a range of local optima in the error-landscape within the parameter space. These co-existing different interpretations (i.e. values of essential parameters) that explain historic energy use comparably well are incorporated in the prediction ensemble.

In the energy demand modelling of the TIMER model, different parameters sets can be observed that all lead to reasonable calibration (equifinality). From the application of this method to the TIMER 2.0 energy demand model for the transport sector, we found that its model formulation, in combination with the aggregated character of energy statistics available for calibration, leaves room for multiple behavioural sets of parameter values. In the given model formulation, the different options for calibrated parameter values are related to the balance between useful energy intensity and energy efficiency improvement. Generally, high useful energy intensity combined with major efficiency improvements leads to similar results as low energy intensity and stagnant efficiency improvement.

Different model calibrations lead to different future projections. The range in outcomes is about 44-79% around the best-fit option. With respect to future projections, we found that different (behavioural) sets of parameter values can lead to a wide range of future projections. *AEEI* and useful energy intensity are the most decisive model aspects with respect to future energy levels.

Equifinality of the TIMER model can partly be improved by further model development. What does this analysis imply for the application and development of the TIMER model? Given the aggregate nature of both model and data some parameter ambiguity is inevitable and does not a priori disqualify the model. For the existing model, a workable situation can be created by using the ‘best-fit’ calibrated parameter values and communicating the calibration uncertainty range with the model results. More fundamentally, two options exist for model improvement. First, the data-based solution would be model reduction. However, because the model only involves three well-established concepts (energy intensity and autonomous and price induced efficiency improvement) model reduction implies econometric curve-fitting. A second option is to convert the model to a more bottom-up nature and use the increasingly available data and insights from the underlying physical activity (in this specific case: data on person or freight kilometres, or ownership of cars, trucks, planes etc.; and the concepts of time and money budgets). Such development would lead to two major improvements: first, it provides an extra model layer (of physical activity) that can be calibrated to data and second, such model enhances insight in the actual activity that is simulated and projected.

Appendix 1: Values and ranges for parameter estimation

Table A 1: Ranges for the estimated parameter values

Variable	Minimum	Maximum
***UEI*-curve**		
$\mathbf{X_{max}}$	5000	50000
$\mathbf{Y_0}$	0	0.116 (USA) 0.066 (EU) 0.072 (Brazil) 0.17 (Russia) 0.026 (India) 0.03 (China)
U	17	90
AEEI		
$\mathbf{F_S}$	0.09 (USA) 0.09 (EU) 0.09 (Brazil) 0.28 (Russia) 0.07 (India) 0.03 (China)	0.68 (USA) 0.70 (EU) 0.68 (Brazil) 2.11 (Russia) 1.0 (India) 1.0 (China)[10]
PIEEI		
Payback time	0.007	6.760
P-value learning curve	0.70	1.00

Table A 2: Average annual GDP per capita growth for the analysed regions

	USA	**WEU**	**Brazil**	**Russia**	**India**	**China**
Average annual GDP/cap growth 1971-2003	2.19%	2.13%	2.21%	0.71%	2.69%	6.69%

[10] Because of the high economic growth in India and China, it might be that historic *AEEI* has been higher as well; therefore we increases the upped bound of the range to the total economic growth: F_S=1.

Appendix 2: Analysis of calibrated parameter values and future projections

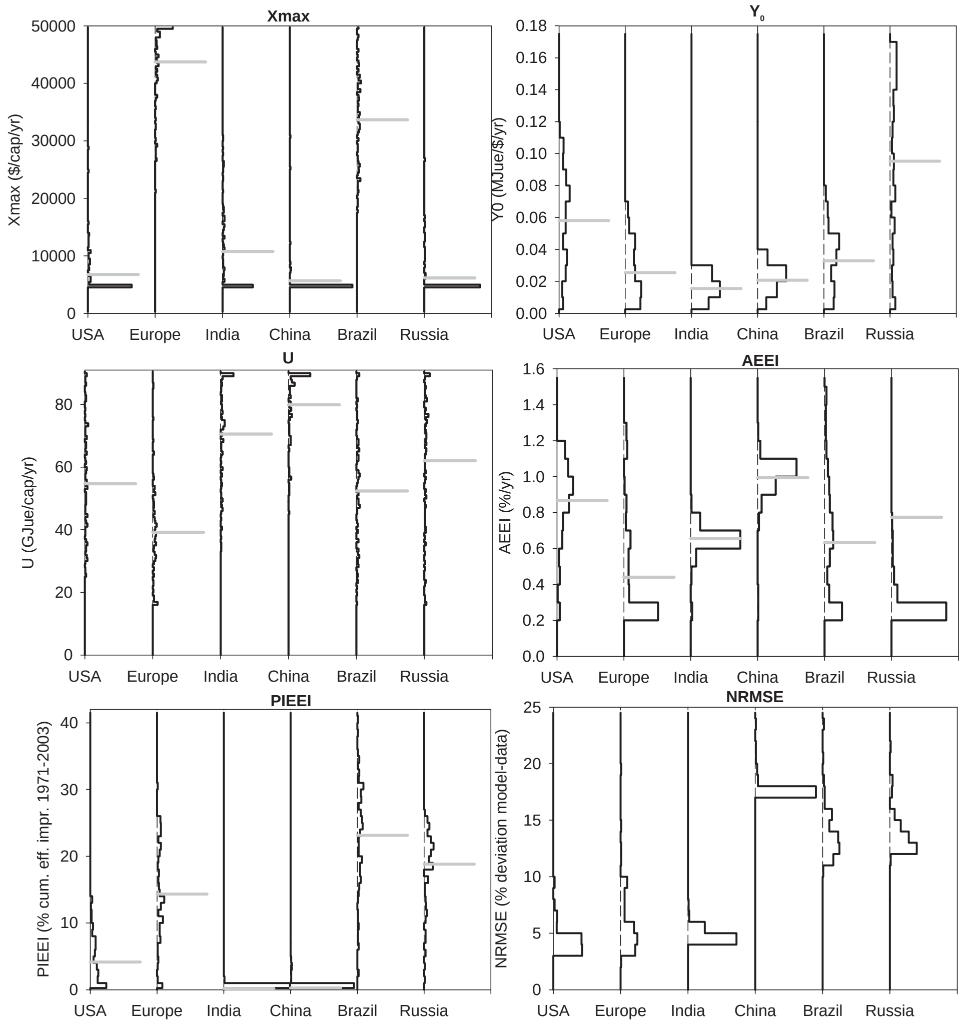

Figure 10: Calibrated parameter values for transport sector energy use in all regions. Distribution (black lines) and mean value (grey)

Appendix 3: Mathematical derivation of parameters for *UEI*-curve

The form of the *UEI* curve (Y) is given by:

$$Y = Y_0 + \frac{1}{\beta \cdot X + \gamma \cdot X^{\delta}} \qquad 8$$

where X denotes the per capita sectoral activity level. The condition for the maximum value of this curve is:

$$\beta + \gamma \cdot \delta \cdot X_{\max}{}^{\delta - 1} = 0 \qquad 9$$

which renders an explicit expression for γ (or likewise for β).

Relating the curve to the saturation level of useful energy per capita per year at high income levels (U), given that δ is negative and assuming that $Y_0=0$ in eqn. 8 (hence, focussing at the second term), means that:

$$U = X \cdot Y = \frac{X}{\beta \cdot X + \gamma \cdot X^{\delta}} \Rightarrow \lim_{X \to \infty} \frac{X}{\beta \cdot X + \gamma \cdot X^{\delta}} = \frac{1}{\beta} \text{ and thus } \beta = \frac{1}{U} \qquad 10$$

If the curve is forced through one observed reference point (X_{ref}, Y_{ref}), the combination of eqn. 9 and using $\tilde{X}_{ref} = \frac{X_{ref}}{X_{\max}}$, renders the following expression for δ:

$$\delta = -\frac{ProductLog\left[\dfrac{\ln(\tilde{X}_{ref}) \cdot X_{\max} \cdot (Y_{ref} - Y_0)}{U + X_{\max} \cdot \tilde{X}_{ref} \cdot (-Y_{ref} + Y_0)}\right]}{\ln(\tilde{X}_{ref})} \qquad 11$$

where ProductLog[z] gives the solution for w in $z = w \cdot e^{w}$; depending on the value of z, this function has multiple branches of solutions involving complex numbers. If $z > -e^{-1}$ the principle branch solution is a real number; if $-e^{-1} < z < 0$ the secondary branch solution is a real number as well. This means that, depending on the values of X_{max}, Y_0 and U, multiple values for δ might exist. In model terms, the primary branch solutions of δ are closest to zero and generally lead to a lower maximum in the curve (more linear) than the secondary branch solutions (more bell shaped). Based on these derivations, the *UEI* curve can be determined as function of the quantities X_{max}, Y_0 and U.

References:

1. Löschel, A., *Technological change in economic models of environmental policy: a survey.* Ecological Economics, 2002. **43**(2-3): p. 105-126.
2. Rotmans, J. and H.J.M. de Vries, *Perspectives on Global Change, the TARGETS approach*. 1997, Cambridge (UK): Cambridge University Press. 463.
3. Grubb, M., J. Edmonds, P.T. Brink, and M. Morrison, *The Costs of Limiting Fossil-Fuel CO2 Emissions: A Survey and Analysis.* Annual Review of Energy and the Environment, 1993. **18**(1): p. 397.
4. Tschang, F.T. and H. Dowlatabadi, *A Bayesian technique for refining the uncertainty in global energy model forecasts.* International Journal of Forecasting, 1995. **11**(1): p. 43-61.
5. da Costa, R.C., *Do model structures affect findings? Two energy consumption and CO2 emission scenarios for Brazil in 2010.* Energy Policy, 2001. **29**(10): p. 777-785.
6. van Vuuren, D.P., B. de Vries, A. Beusen, and P.S.C. Heuberger, *Conditional probabilistic estimates of 21st century greenhouse gas emissions based on the storylines of the IPCC-SRES scenarios.* Global Environmental Change, 2008. **18**(4): p. 635-654.
7. Kann, A. and J. Weyant, *Approaches for performing uncertainty analysis in large-scale energy/economic policy models.* Environmental Modeling and Assessment, 2000. **5**(1): p. 29-46.
8. Beven, K., *A manifesto for the equifinality thesis.* Journal of Hydrology, 2006. **320**(1-2): p. 18-36.
9. Draper, D., *Assessment and Propagation of Model Uncertainty.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 45-97.
10. Doherty, J., *PEST Model-Indipendent Parameter Estimation, User Manual: 5th edition*. 2004, Brisbane, Australia: Watermark Numerical Computing. 336. Available from: www.sspa.com/pest.
11. Poeter, E.P., M.C. Hill, E.R. Banta, S. Mehl, and S. Christensen, *UCODE_2005 and Six Other Computer Codes for Universal Sensitivity Analysis, Calibration, and Uncertainty Evaluation*. U.S. Geological Survey Techniques and Methods. 2005: U.S. Geological Survey. 283.
12. Beven, K. and A. Binley, *The future of distributed models: Model calibration and uncertainty prediction.* Hydrological Processes, 1992. **6**(3): p. 279-298.
13. Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity analysis in practice, a guide to assessing scientific models*. 2004, Chichester: John Wiley & Sons, Ltd.
14. de Vries, H.J.M., D.P. van Vuuren, M.G.J. den Elzen, and M.A. Janssen, *The TIMER IMage Energy Regional (TIMER) Model*. 2001, Bilthoven: National Institute for Public Health and the Environment (RIVM). 188. Available from: http://www.mnp.nl/bibliotheek/rapporten/461502024.pdf.
15. van Vuuren, D.P., B.J. van Ruijven, M.M. Hoogwijk, M. Isaac, and H.J.M. de Vries, *TIMER 2.0, Model description and application*, in *Integrated modelling of global environmental change. An overview of IMAGE 2.4*, A.F. Bouwman, M.P.M. Hartman, and C.G.M. Klein Goldewijk, Editors. 2006, Netherlands Environmental Assessment Agency (MNP): Bilthoven.
16. Bouwman, A.F., M.P.M. Hartman, and C.G.M. Klein Goldewijk, eds. *Integrated modelling of global environmental change. An overview of IMAGE*

2.4. 2006, Netherlands Environmental Assessment Agency: Bilthoven, The Netherlands.
17. IPCC, *Special Report on Emission Scenarios*. 2000, Cambridge: Intergovernmental Panel on Climate Change, Cambridge University Press.
18. MA, *Millenium Ecosystem Assessment: Ecosystems for Human Wellbeing*. 2005, Washington DC: Island Press.
19. UNEP, *Global Environment Outlook: Environment for Development*. 2007, Nairobi: United Nations Environment Program. Available from: http://www.unep.org/geo/geo4.
20. OECD, *OECD Environmental Outlook to 2030*. 2008, Paris: OECD. Available from: www.oecd.org/environment/outlookto2030.
21. van Vuuren, D.P., *Energy systems and climate policy*. Dept. of Science, Technology and Society, Faculty of Science. 2007, Utrecht: Utrecht University. 326.
22. van den Berg, H., *Calibration & Evaluation of a Global Energy Model (submodel of TARGETS)*. Centre for Energy and Environmental Studies (IVEM). 1994, Groningen: University of Groningen. 95.
23. van der Sluijs, J.P., J. Potting, J. Risbey, D. van Vuuren, B. de Vries, A. Beusen, et al., *Uncertainty assessment of the IMAGE-TIMER B1 CO2 emissions scenario, using the NUSAP method*. 2001: Dutch National Research Program on Climate Change. 225.
24. Risbey, J., J.P. Van der Sluijs, P. Kloprogge, J. Ravetz, S. Funtowicz, and S. Corral Quintana, *Application of a Checklist for Quality Assistance in Environmental Modelling to an Energy Model.* Environmental Modeling & Assessment, 2005. **10**(1): p. 63-79.
25. van der Sluijs, J.P., *Uncertainty, assumptions, and value commitments in the knowledge-base of complex environmental problems*, in *Interfaces between Science and Society*, Â.G. Pereira, S.G. Vaz, and S. Tognetti, Editors. 2006, Green Leaf Publishing. p. 67-84.
26. van der Sluijs, J.P., *A way out of the credibility crisis of models used in integrated environmental assessment.* Futures, 2002. **34**(2): p. 133-146.
27. van der Sluijs, J.P., *Uncertainty as a monster in the science policy interface: four coping strategies.* Water science and technology, 2005. **52**(6): p. 87–92.
28. van der Sluijs, J.P., *Uncertainty and precaution in environmental management: Insights from the UPEM conference.* Environmental Modelling & Software, 2007. **22**(5): p. 590-598.
29. Refsgaard, J.C., J.P. van der Sluijs, A.L. Hojberg, and P.A. Vanrolleghem, *Uncertainty in the environmental modelling process - A framework and guidance.* Environmental Modelling & Software, 2007. **22**(11): p. 1543-1556.
30. Refsgaard, J.C., J.P. van der Sluijs, J. Brown, and P. van der Keur, *A framework for dealing with uncertainty due to model structure error.* Advances in Water Resources, 2006. **29**: p. 1586–1597.
31. Filar, J.A., *Mathematical Models*, in *Knowledge for Sustainable Development - An Insight into the Encyclopedia of Life Support Systems*. 2002, UNESCO/EOLSS: Released at the World Summit on Sustainable Development, Johannesburg, South Africa. p. 339-354.
32. van der Sluijs, J.P., *Anchoring Amid Uncertainty, on the Management of Uncertainties in Risk Assessment of Anthropogenic Climate Change*. Department of Science, Technology and Society. Vol. PhD thesis. 1997, Utrecht: Utrecht University.

33. Walker, W.E., P. Harremoës, J. Rotmans, J.P. Van der Sluijs, M.B.A. Van Asselt, P. Janssen, et al., *Defining Uncertainty A Conceptual Basis for Uncertainty Management in Model-Based Decision Support.* Integrated Assessment 2003. **4**(1): p. 5-17.
34. Janssen, P.H.M., A.C. Petersen, J.P. Van der Sluijs, J. Risbey, and J.R. Ravetz, *A guidance for assessing and communicating uncertainties.* Water science and technology, 2005. **52**(6): p. 125–131.
35. Oreskes, N., K. Shrader-Frechette, and K. Belitz, *Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences.* Science, 1994. **263**: p. 641-646.
36. Beck, B., *Model evaluation and performance*, in *Encyclopedia of Environmetrics*, A. El-Shaarawi and W. Piegorsch, Editors. 2002, John Wiley & Sons, Ltd: Chichester. p. 1275–1279.
37. Edwards, P.N., *Global climate science, uncertainty and politics: Data-laden models, model-filtered data.* Science As Culture, 1999. **8**(4): p. 437-472.
38. Barlas, Y., *Multiple tests for validation of system dynamics type of simulation models.* European Journal of Operational Research, 1989. **42**(1): p. 59-87.
39. Young, P., *Data-based mechanistic modelling of environmental, ecological, economic and engineering systems.* Environmental Modelling and Software, 1998. **13**(2): p. 105-122.
40. Young, P.C., S.D. Parkinson, and M. Lees, *Simplicity out of complexity: Occam's razor revisited.* Journal of Applied Statistics, 1996. **23**: p. 165-210.
41. Crout, N.M.J., D. Tarsitano, and A.T. Wood, *Is my model too complex? Evaluating model formulation using model reduction.* Environmental Modelling & Software, 2009. **24**(1): p. 1-7.
42. Jakeman, A.J., R.A. Letcher, and J.P. Norton, *Ten iterative steps in development and evaluation of environmental models.* Environmental Modelling & Software, 2006. **21**(5): p. 602-614.
43. Beck, M.B., J.R. Ravetz, L.A. Mulkey, and T.O. Barnwell, *On the problem of model validation for predictive exposure assessments.* Stochastic Hydrology and Hydraulics, 1997. **11**(3): p. 229-254.
44. Dogan, G. *Confidence Interval Estimation in System Dynamics Models: Bootstrapping vs. Likelyhood Ration Method*. in *22nd International Conference of the System Dynamics Society*. 2004. Oxford, UK.
45. Oliva, R., *Model calibration as a testing strategy for system dynamics models.* European Journal of Operational Research, 2003. **151**(3): p. 552-568.
46. Janssen, P.H.M. and P.S.C. Heuberger, *Calibration of process-oriented models.* Ecological Modelling, 1995. **83**(1-2): p. 55-66.
47. NIST/SEMATECH. *e-Handbook of Statistical Methods*. 2006 [cited 2007 5 October]; Available from: http://www.itl.nist.gov/div898/handbook/.
48. Mathworks, *Optimization Toolbox, user's guide*. 2007, Natick, MA, USA. Available from: http://www.mathworks.com/access/helpdesk/help/pdf_doc/optim/optim_tb.pdf.
49. Focacci, A., *Emperical analysis of the environmental and energy policies in some developing countries using widely employed macroeconomic indicators: the cases of Brazil, China and India.* Energy Policy, 2005. **33**: p. 543-554.
50. Medlock III, K.B. and R. Soligo, *Economic Development and End-Use Energy Demand.* Energy Journal, 2001. **22**(2): p. 77.

51. Reddy, A.K.N. and J. Goldemberg, *Energy for the Developing World.* Scientific American, 1990. **263**(3): p. 111.
52. Stern, D.I., *The Rise and Fall of the Environmental Kuznets Curve.* World Development, 2004. **32**(8): p. 1419-1439.
53. van Ruijven, B., F. Urban, R.M.J. Benders, H.C. Moll, J.P. van der Sluijs, B. de Vries, et al., *Modeling Energy and Development: An Evaluation of Models and Concepts.* World Development, 2008. **36**(12): p. 2801-2821.
54. Kander, A. and L. Schon, *The energy-capital relation--Sweden 1870-2000.* Structural Change and Economic Dynamics, 2007. **18**(3): p. 291-305.
55. Gales, B., A. Kander, P. Malanima, and M. Rubio, *North versus South: energy transition and energy intensity in Europe over 200 years.* European Review of Economic History, 2007. **11**(2): p. 219-253.
56. Seppälä, T., T. Haukioja, and J. Kaivo-oja, *The EKC Hypothesis Does Not Hold for Direct Material Flows: Environmental Kuznets Curve Hypothesis Tests for Direct Material Flows in Five Industrial Countries.* Population & Environment, 2001. **23**(2): p. 217-238.
57. van Vuuren, D.P., B.J. Strengers, and H.J.M. De Vries, *Long-term perspectives on world metal use--a system-dynamics model.* Resources Policy, 1999. **25**(4): p. 239-255.
58. Bernardini, O. and R. Galli, *Dematerialization: Long-term trends in the intensity of use of materials and energy.* Futures, 1993. **25**(4): p. 431-448.
59. Groenenberg, H., *Development and Convergence, a bottom-up analysis for the differentiation of future commitments under the Climate Convention*. Faculty of Chemistry. Vol. PhD Thesis. 2002, Utrecht: Universiteit Utrecht. 230.
60. Dargay, J., D. Gately, and M. Sommer, *Vehicle Ownership and Income Growth, Worldwide: 1960-2030.* Energy Journal, 2007. **28**(4): p. 143-170.
61. IEA. *Perspectives énergétiques mondiales à long term – le cas spécifique des transports*. 2007 [cited 2007 December].
62. Manne, A., R. Richels, and J. Edmonds, *Market Exchange Rates Or Purchasing Power Parity: Does The Choice Make A Difference To The Climate Debate?* Climatic Change, 2005. **71**(1): p. 1-8.
63. van Ruijven, B.J., F. Urban, R.M.J. Benders, H.C. Moll, J.P. van der Sluijs, D.P. van Vuuren, et al., *Modeling Energy and Development: an Evaluation of Models and Concepts.* World Development, 2008. **36**(12).
64. Bakkes, J., P.R. Bosch, A.F. Bouwman, H.E. Eerens, M. den Elzen, M. Isaac, et al., *Background report to the OECD Environmental Outlook to 2030. Overviews, details, and methodology of model-based analysis*. 2008, Bilthoven: Netherlands Environmental Assessment Agency (MNP). 186. Available from: http://www.mnp.nl/bibliotheek/rapporten/500113001.pdf.
65. van Vuuren, D.P., M.G.J. Den Elzen, J. Van Vliet, T. Kram, P.L. Lucas, and M. Isaac, *Comparison of different climate regimes: the impact of broadening participation.* in review.
66. Hendrix, E.M.T. and O. Klepper, *On Uniform Covering, Adaptive Random Search and Raspberries.* Journal of Global Optimization, 2000. **18**(2): p. 143-163.
67. Mantovan, P. and E. Todini, *Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology.* Journal of Hydrology, 2006. **330**(1-2): p. 368-381.
68. Mantovan, P., E. Todini, and M.L.V. Martina, *Reply to comment by Keith Beven, Paul Smith and Jim Freer on "Hydrological forecasting uncertainty*

assessment: Incoherence of the GLUE methodology". Journal of Hydrology, 2007. **338**(3-4): p. 319-324.
69. Beven, K., P. Smith, and J. Freer, *Comment on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini.* Journal of Hydrology, 2007. **338**(3-4): p. 315-318.