

ORIGINAL ARTICLE

Open Access



# Quantitative scenario design with Bayesian model averaging: constructing consistent scenarios for quantitative models exemplified for energy economics

Monika Culka 

## Abstract

**Background:** Scenario design is currently not a standardised process. The formulation of storylines representing different dimensions (for example economic or societal developments) demands an investigation of assumption compatibility, coherence, and consistency. Scenario techniques that use expert opinion as the sole information source are particularly appropriate for personal decisions. Contexts where scenarios serve as decision support on a societal level—for example in political decision-making—benefit from unbiased, fact-depicting, multi-dimensional information that is available in statistical data.

**Methods:** The presented approach uses the well-established method of Bayesian model averaging for the formulation of consistent, transparent, and intuitively understandable quantitative scenario assumptions. These assumptions are used in quantitative models to produce outlooks and forecasts. Illustrated by the example of quantitative energy models used to investigate developments of the energy system by scenario technique, the approach contrasts with other scenario methods. Bayesian model averaging (BMA) is a method that allows for an evaluation of both system relation stability in terms of observable co-evolution of phenomena in the past and of future system states of interest based on expert opinion where past evolutions serve as a point of reference.

**Results:** The results are scenarios assessable with respect to (1) the consistency of scenario assumptions in terms of statistical confirmation, (2) the suitability of a quantitative model to represent the scenario, and (3) the statistical uncertainty of the scenario for a given quantitative model. A transparent scenario construction process results in traceable assumption documentation (an exemplary communication is provided in the [Appendix](#)). Perhaps, the most important novelty of the approach is the possibility of communicating to decision-makers the associated uncertainty in easily understandable terms. The distinction between provable possible assumptions (based on statistical evidence) and hypothetical assumptions is a novelty and significantly improves the aptitude of scenario study recipients to evaluate scenarios on their part.

**Conclusions:** BMA provides the possibility for decision-makers (and all recipients of outlooks based on scenario technique) to trace back results to assumptions and provide an evaluation of these assumptions in terms of statistical confirmation. As such, the approach adds to the currently limited methodological diversity in scenario construction techniques.

**Keywords:** Scenario technique, Uncertainty modelling, Assumption consistency, Empirical adequacy

Correspondence: [energy.environment.engineer@gmail.com](mailto:energy.environment.engineer@gmail.com)  
NICA New Interdisciplinary Collaboration Association, 6330 Kufstein, Austria

## Background

Formulating scenarios is a relevant part of future research. This paper aims to contribute to the currently limited toolbox of scenario construction methodologies. The Bayesian model averaging (BMA) technique is a well-established methodology today and, as I will argue, is an appropriate conceptual setting for consistent scenario construction for application cases where (some) cause-effect relations are uncertain and the mathematical representation in models should account for that uncertainty. Exemplified for the case of energy modelling in this paper, the idea of the BMA scenario technique is that what is observed in the past (documented by statistical data) is a proven possible state of the world. A state of the world that observably (re-)occurred in the past is more probable and less uncertain than a state of the world that has not been observed before.

However, in many circumstances, investigating unprecedented situations is the very reason for creating scenarios! BMA offers a way that uses “knowledge of the past” about parts of the world—say, the number of unemployed people documented in statistical data—to formulate expectations about these parts of the world (unemployment rate) in different states of accompanying phenomena. It is important to understand that the statistical method tries to find *relations* of phenomena expressed as statistical data, based on similarities or differences in the changes these phenomena undergo. It is the expected impact of an assumption on other assumptions of the scenario, given the data record we consider. The technique is particularly suitable for scenarios that figure as assumptions for consequent processing in quantitative models, e.g. optimization models and simulation models.

In contrast to most scenario techniques, BMA does not primarily rely on expert judgement. I emphasise that judgement-based scenarios are suitable in different contexts. In the case of energy scenarios, the need for techniques improving known difficulties associated with judgement-based scenario design demands empirical evidence as a further source of information. In the following discussion, I will present the BMA method for consistent scenario construction from a mainly conceptual perspective. Constructing scenarios is generally best performed with a perspective on the specific application case and purpose of the scenario. This implies that the scenario construction process should rely on different methodologies, in particular, qualitative techniques, to avoid an overemphasis of statistics. At the same token, applying only qualitative judgement-based approaches risks neglecting evidence and promotes an apodictic expert opinion. Choosing the appropriate methodologies in a scenario construction process remains the main task of scenario designers.

The following discussion is a *conceptual* discussion. In contrast to a technical presentation of a method, the focus lies here on arguing for Bayesian model averaging in the context of quantitative models. That means, *for this paper, I explain how to make sense of BMA results in scenario construction, and not primarily how to derive a BMA analysis*. For an example of a BMA analysis, I would like to refer to [1]. A detailed *methodological* discussion of BMA in terms of mathematical formulation and computation options is given in [2–9], to name just a few. Fragoso et al. [10] present a taxonomy of BMA literature by means of a meta-analysis of published works. While earlier works of mine could be viewed as Fragoso et al.’s usage category “joint estimation”, the present discussion would match the “joint prediction” category, which is justifiably a separate category.<sup>1</sup> Though the exemplary application case is again energy-economic modelling, the inferences drawn from applying BMA for scenario design are novel.

The focus of the paper is to present BMA as a technique to compute scenarios. This is particularly suitable in contexts where mathematical relations are approximations and the modelled entities’ behaviour is uncertain. In contrast, the modelling of parts of the world that obey laws of nature is often straightforward in mathematical terms. A model of a dropped ball is a precise mathematical relation for example. The results computed with such a model depict empirically well-confirmed system states for known parameters (e.g. gravitational pull) and variables (e.g. weight of the ball). The computed model results are quite precise—expectation for, say, the ball’s position at time  $t$ . It is insensible to repeat a ball throw over and over to gather statistical data when we know a mathematical relation describing the relations precisely, a mathematical function. But when it comes to scenarios depicting (also) human decision(s), for example, scenarios in social sciences, then we lack a precise mathematical formulation as humans can decide differently at any time. The behaviour of humans and the consequent “behaviour” of relevant variables depicted in statistical data (e.g. GDP, trade balance, demand) can only be approximated. This is what BMA, as a statistical method, does. BMA approximates with a view on the statistically evident behaviour in the past. Judgement-based scenario techniques reflect the expert understanding of the phenomena’s behaviour in the past.

In the “[Energy \(economic\) modelling as particular context](#)” section, I investigate the context of energy scenarios. In the sections “[Quantitative and qualitative scenario construction](#)” and “[Scenario definition and uncertainty evaluation](#)”, I review multiple scenario construction techniques to contrast them with each other and to relate them to the requirements for scenario construction in the energy modelling context. A basic principle depicts scenario design in terms of phenomena and energy model boundaries. Using an example, my

discussion of consistent scenario construction extends to two cases: an existing energy model and the case of a scenario-adapted energy model in the “[Consistent scenario construction for a given energy model and an adapted energy model](#)” section. The BMA results are numerical assumptions necessary for quantitative energy models as an input as discussed in the section “[Results: consistent numerical value estimation](#)”. The following section consists of a critical discussion of the approach and its limits in detail. Based thereon, I draw conclusions and end with a brief summary. The [Appendix](#) is an exemplary communication of scenario assumptions.

The hope is that this paper helps to acknowledge that scenarios depicting a future world should also respect the world of the past and the present. Here, BMA would be one way to do so.

### Energy (economic) modelling as particular context

The energy system of a country is interrelated with different societal aspects forming “systems” on their part. Economic, social, environmental, and governmental policies influence the design and desired changes of an energy system. Stakeholders are present in all societal “categories”, for example, industry, public, government, or non-governmental organisations. An adequate design and adaptation of the energy system to the changing needs of a society are in the interest of all stakeholders. The priorities may, however, vary according to stakeholder objectives, planning, and societal duties.

Changes to an energy system cannot be experimentally tested, as compared to the design of a physical experiment. Implementing “new” policies is a delicate process that needs to balance economic feasibility, societal acceptance, industrial attractiveness, and political rigour. In addition, international agreements, such as the security of supply agreements or environmental protection agreements, demand for strategies that are respectful in regard to both the accepted duties and their practicability.

Investigating potential consequences of policy measures for different stakeholders of an energy system in terms of monetary, technical, environmental, and social burdens has become a major concern of quantitative energy modelling for policy advice. Assisting the impact evaluation for policy advice is a central role of quantitative energy modelling [11, 12]. To account for different possibilities, *scenarios* are developed representing a set of numerical assumptions interpreted in a narrative way, the so-called storyline. What the term “scenario” refers to is not clearly defined in the literature.

Van Notten discusses 11 definitions and application examples for scenarios [13]. Lindgren debates paradoxical situations and practical indications of the scenario technique [14]. Van Notten and also Lindgren accord to the scenario technique qualities as intuitiveness, creativity,

associational thinking, causal relation assumptions, and other possibly non-standardised characteristics. The main objective of scenarios is to create a set of assumptions representing a state of the world of interest, used for the evaluation of future developments [15]. Önköl et al. have addressed the difference between method-based statistical forecasting and the scenario technique. According to them, scenario technique reflects plausible futures based on the reasoning of the scenario designer [16].

In quantitative modelling, possible future states of the energy system are limited to some defined *input scenarios*, also called storylines or key assumptions, what implies a subjective and decisive pre-selection of futures scrutinised with an energy model. This is a delicate process that should involve expert knowledge, and rigorous attention must be paid to plausibility. Individually stipulated assumptions may, in concert with other individually plausible assumptions, amount to implausibility due to reciprocal assumption impact. An energy model, designed to represent an existing energy system, is typically applied to investigate potential consequences for the target system, given things were as assumed in a scenario. However, due to the interrelated nature of the target system, experimental confirmation of scenario assumptions is limited, if not impossible. Therefore, the assumptions figuring in a scenario cannot solely be derived from intuitive scenario methodologies, if the energy model results should represent a provable possible or even probable energy system state. I refer to *energy models* as quantitative descriptions of an existing energy system, e.g. [17–19]. The literature on existing energy models is given for example in [20–22], or [12], where reviews and evaluations are published.

The method proposed for scenario construction addresses the problem of scenario representation in energy models and evaluates the scenario assumptions for a given energy model in terms of their *empirical adequacy*. The empirical adequacy of an assumption is its propensity to represent possible states of the world as confirmed by statistical evidence.<sup>2</sup> In other words, I take *consistent* scenario construction to mean that numerical assumptions are consistent with statistically evident stable relations in the target system. The transparent documentation and communication of the assumptions’ statistical confirmation can help recipients find their own opinion of a scenario.

Energy model results are typically presented as energy scenario studies, for example [23]. A consistent scenario construction as an accompanying document is an uncertainty assessment, as presented in [24], as well as the consequent predictive density computations, the scenarios. The scenarios come thus automatically with an uncertainty estimation for the specific energy model and the specific scenarios computed with it.

In other words, consistent BMA scenarios estimate a quantitative (energy) model’s suitability to *represent* a

scenario. Consistent scenario construction can assess scenarios of a specific energy model in terms deemed relevant by Goodwin [25] p. 7: transparency (what are the relevant phenomena according to the data), the ease of judgement (how good are the relevant phenomena captured in the energy model—both quantitative via posterior model probability (PMP) and qualitative via posterior inclusion probability (PIP)), the versatility (the BMA can be applied to many quantitative models), the flexibility (provided statistical data are available, and different phenomena can be included in the statistical analysis), and theoretical correctness (the mathematical core of the BMA is set, applying BMA means exercising that theory on the data). I will return to these criteria by Goodwin in the conclusion section and discuss the BMA methods' aptitude as a "formal strategy evaluation process within the scenario planning".

The general characteristic of scenario construction that is specific to the energy modelling context is a tight connection of the scenario to the actual world. In other words, scenarios modelling potential energy futures are (partially) used as a replacement of experiments (which cannot be carried out) and serve as *concrete* guidance in decision support. This places requirements on the employed scenario technique in terms of empirical adequacy, as the purpose of energy scenario studies is an evaluation of actual, possible, and plausible future options, which decision-makers may have to consider.

### Methods: quantitative and qualitative scenario construction

For clarity, I would like to start with a clarification of the terminology used. I take a *phenomenon* to be either a physically observable or an invisible socially emerged constellation of parts of reality that are naturally interrelated. Physically observable phenomena are quantifiable via measurements and/or observation records. Social phenomena are observable and quantifiable via an interrogation and/or observation record. An *observation record*, also called *empirical evidence*, is in this case the statistical data. In fact, a phenomenon may exhibit different empirical evidence of itself. Statistical data have the advantage above personal observations that they are collected systematically, according to a method, and data observed the *same aspects* of a phenomenon over time. This methodological transparency of statistical data serves as common ground for different persons to speak about reality. However, one must not think that statistical data describe or capture a phenomenon exhaustively or even just appropriately. They are merely a basis allowing different people to speak about the same aspects of the target system.

In this discussion, an *energy model* is a mathematical representation of an energy system with the aim to depict a real energy system simplified and idealised, but

nonetheless empirically adequate. The term *energy system* refers to the part of reality that is (1) physically existent in the world used to generate and deliver energy (e.g. electricity, heat); (2) economically associated with the processes of generating, transporting, and consuming energy; (3) socially related to the effects induced by energy consumption and access (e.g. fuel poverty); (4) related to environmental phenomena (e.g. change of gaseous composition of the atmosphere due to energy sector CO<sub>2</sub> emissions); and (5) part of individual human reality, i.e. a human is aware that the energy model and the scenarios represent a part of her reality. The energy system (ESS) is part of the world (WSS), and the world is part of the universe.

In Fig. 1, a subset illustrates being "part of" the larger reality, to which I refer to as the *target system*. The energy system is not naturally demarcated from any other system of the world, and the world is not naturally separated from any other system of the universe. The energy model depicts parts of the energy system and all interrelated systems of the world *by stipulation*, expressed in the *energy model design and energy model boundaries*. Energy model boundaries are an artificial demarcation between naturally interrelated phenomena. The energy model design is basically the choice of input variables, parameters, and output variables representing the phenomena in WSS. *Statistical data* are quantitative, i.e. continuous or discrete or qualitative, i.e. categorical (nominal and ordinal) descriptions of phenomena in WSS according to the definition of the data collecting institution.<sup>3</sup>

I take a *true data generating process* to be a process that causally influences the numerical appearance of statistical data. A true data generating process can be artificial, for example, if the data are created to test statistical methods, or natural, if known or unknown phenomena in WSS cause the data recorded. For example, let our data be the coordinates of a ball at a time. If the ball is thrown, the true data generating process is the force influencing the ball, which "changes" the coordinates' numbers—the data we collect. The *relation* in WSS of the phenomena can, for the ball example, be described by a mathematical formulation using laws of classical mechanics. Depending on the phenomena described by a statistical data point, the number of true data generating processes can vary and their causal status too. For example, the statistical numerical description of the phenomenon GDP is caused by, or correlated with, different phenomena in WSS as for example consumer satisfaction, trade balance, tax burdens, unemployment, etc. Statistical methods typically aim at identifying what phenomena are influential. The proposed BMA method does so, too.

Scenario construction has so far not been considered as a scientific area itself, and to an extent it should have been [26]. Some independently developed



methodologies, techniques, and quality standards have recently emerged. Research addressing the crucial role of scenario construction and the difficulty in classifying the diverse techniques has been undertaken, for example, by [27–29]. Scenario construction is the systematic choice of numerical values for exogenous variables (input variables) and parameters of an energy model as assumptions. Scenarios constructed to derive recommendations for decision support necessitate a balance between confirmed possibility and hypothetical assumption creation.

Unfortunately, in energy-economic modelling, the current practice of scenario construction is often opaque and unsystematic [30]. Sometimes, scenarios are defined, that is, agreed upon by modellers and sponsors. The consequently stipulated numerical values for assumptions translate the storyline. With opaque scenario construction, a “result design” for sponsors is also possible, what is in my view scientific misconduct. Evaluating the *representation quality* of the agreed storyline in a specific energy model is rarely addressed. The proposed method does so by looking at the energy model input variables’ ability to represent the scenario. If the defined scenarios are not based on a systematic analysis of interrelations in the target system, *the truth and legitimacy of the claim that energy model results represent a response of the energy system to the scenario* cannot be evaluated. BMA scenarios for input variables allow for the construction and evaluation of consistent scenarios based on observable energy system relations in the past using statistical data.

Methodologies and techniques reported in the literature are presented in Table 1. The comparison of scenario techniques is based on the categorisation introduced by [31].

The following series of arguments addresses the question as to whether the proposed BMA technique is a sensible addition to the purely qualitative approaches of the scenario method. To do so, I highlight some aspects of opinion-based approaches which are implicit to them. I emphasise that these aspects are *not* erroneous in themselves. They may serve as an advantage in scenario design in some contexts. The basic claim I argue for is that scenario construction techniques which use opinion as a sole source of information fall short of empirical adequacy required *for decision support and policy advice*. Scenarios based on opinion

1. Lack a democratic perspective,
2. Lack the possibility to evaluate the scenario quality,
3. Suffer from detrimental psychological effects in the context of decision support,
4. Cannot reflect the target system’s complexity due to limitations of human reasoning capacity.

I recall that the particular context of energy modelling uses scenarios as a basis for decision support, replacing experiments. In other words, scenarios computed with energy models may lead to political decisions *influencing real people in the real world*. Scenario construction techniques which suffer extensively from the four points raised in this section bias the futures investigated, which could have far-reaching and society-relevant consequences.

Constructing scenarios for decision support with consequences surpassing the realm of personal experience of the scenario constructor places in a sense an obligation on the scenario to account for the interests of all people affected. If the information source of scenarios is an expert opinion, the constructed scenarios are necessarily and inevitably biased towards the personal situation of the expert(s). In a democracy, however, possible (probable, plausible, and consistent) futures presented to decision-makers ought to envision futures respectful of all stakeholders. Statistical data as an information source, in a sense, encode stakeholder choices and implicitly reflect different interests, for instance, technology acceptance, economic commitment to changes, ecological concerns, institutional (personal, societal) priorities, etc. Many statistical data are available from trustworthy sources that collect data non-discriminatory, regularly, reliably, methodologically sound, and freely available.

Moreover, statistical data connect conceivable future scenarios to relevant constraints. Relating the thinkable to the feasible, and showing where changes are necessary to render scenarios feasible, can be achieved in a straightforward manner based on statistical evidence. For example, the statistical information of average processing times for construction permits of electricity transportation facilities could constrain scenarios of generation capacity increase realistically. As a transparent assumption in a scenario, it is at the same time the action recommended to decision-makers; the *assumption* that processing times for construction permits are stipulated shorter (or equal, longer) than statistically evident in the scenario can be communicated in detail.<sup>4</sup>

The second and the third point concern the problem that expert elicitation as an information source for scenario construction cannot be evaluated in terms of quality. There are no standards as to who should be considered as an “expert” and no criteria for the status of being an expert. There are no requirements for group design and environmental design to prevent psychological effects reported in the literature [32–35]. Although standards have been proposed [36], it is not a common practice to accompany the judgement-based scenario construction with a methodological assessment, and more importantly, the quality of the standards is itself a question of opinion. In contrast, statistical data can be evaluated in several dimensions. Time, scope, collecting agency, post-processing, data arrangement, accessibility, and *financing* of the data

**Table 1** Comparison of scenario design methodologies and techniques adapted from [31]

Methodological approach	Technique	Keywords or a (very) brief description	Source of information
Judgement	Genius forecasting	"Think the unthinkable" [55]	Opinion
	Visualisation	Intuitive images are combined to scenarios that are in juxtaposition to analytical strategies, e.g. [56]	Opinion
	Role playing	A group judgement technique where individuals create a response to a hypothetical situation which is considered as a scenario. Playing the devil's advocate is an example of scenario forming with focus on unprecedented or highly unlikely scenarios [57].	(Group) opinion
	Coates and Jarratt	For a given time frame and domain, four to six scenario themes regarding the most significant kinds of potential future developments based on judgement are formulated, cf. [58]	Opinion
Baseline scenario: the expected future	Trend extrapolation	Measures existing trends and extrapolates effects into the future; both judgement and empirical analyses are possible. The Manoa technique defines three strong trends which are analysed w.r.t. their implications separately, and in conjunction using a cross-matrix, cf. [26]	Opinion or statistical data
Elaboration of fixed scenarios	Incasting	Based on an (extreme) state of the world participants judge potential impacts in various respects as politics, economics, etc. Qualitative or quantitative for example life cycle assessment technique [59]	Opinion/data
	SRI matrix (Stanford Research Institute matrix)	From a column-wise classification of fixed scenarios as for example worst case or expected future the dimensions (e.g. population, environment) are evaluated row-wise, cf. [60]	Opinion
Event sequences	Probability trees and scenario trees	Different future conditions constitute individual paths which are assigned probabilities. Probability trees are used in risk management. A related technique is scenario trees where a reduction of probable paths to relevant paths is carried out, cf. [61]	Opinion
	Intuitive scenario building	Probability trees are evaluated to identify characteristics that are common to several branches. Summarising these branches is considered as a way to create coherent scenarios, e.g. sociovision [62]	Opinion
	Divergence mapping	A set of events, derived from brainstorming, is aligned in different time horizons forming the storyline of a scenario. The relation of earlier events to the later events is seen to be a plausible sequence, cf. [63]	Opinion
Backcasting	Horizon mission methodology	Supposing a hypothetical situation (the scenario) was an actual situation, the ways and necessary components at present are analysed to achieve the scenario [31, 64]	Opinion, state-of-the-art technology data
	Impact of future technologies	Multiple future scenarios are the basis from which experts work backward and identify necessary (technological) breakthroughs, e.g. [65]	Opinion
	Future mapping	An expert elicitation technique where pre-defined events and pre-defined end-states are arranged to investigate interrelations and consequences, cf. [66]	Opinion
Dimensions of uncertainty	GBN (Global Business Network)	Based on two dimensions of uncertainty and polarities, four combinations are seen to constitute plausible futures, cf. [62]	Opinion
	Morphological analysis and field anomaly relaxation	Multiple dimensions of uncertainty captured in columns are related to alternative events in rows. A scenario is created by the alignment of alternatives of each column, cf. [67]	Opinion
Cross-impact analysis	Interactive future simulation IFS	Based on a set of variables (descriptors) an assessment of their mutual relevance based on expert judgement is carried out. Consistent scenarios are constructed in the sense that variable combinations are computed that have been judged to be compatible. There are probabilistic versions of cross-impact analyses [37]	Opinion arranged by mathematical method
Modelling	Trend impact analysis	Based on a business-as-usual trend assumption, the impact of a potential event on that baseline scenario is evaluated in distinguished impact sequences (first depart from trend continuation, maximum impact, and effect integration), cf. [68]	Opinion (and statistical data)

**Table 1** Comparison of scenario design methodologies and techniques adapted from [31] (Continued)

Methodological approach	Technique	Keywords or a (very) brief description	Source of information
	Sensitivity analyses	Given a model, exogenous variables or model parameters are varied. The changes of model results given varying input/parameter assumptions are evaluated as "sensitivity". Often the one-parameter-at-a-time technique is employed, a kind of <i>ceteris paribus</i> approach [69]	(Opinion and) statistical data
	Dynamic scenarios	From brainstormed scenario themes a system is mapped using causal models. The variables figuring in different causal models are combined in a meta-model mapping the whole domain. The meta-model is analysed for different uncertainties involved in the variables, cf. [70]	Opinion and statistical data
	Bayesian model averaging (BMA)	Scenarios are constructed from statistical data records of phenomena most relevant for an input variable of a consequent quantitative model. Uncertainty is evaluated based on the explanatory power of influencing phenomena most relevant in the historical record, cf. [24]	Opinion and statistical data

collection are transparent and often follow a methodologically rigorous process. It seems legitimate to consider that scenarios based on statistical data are less susceptible to personal interests, personal experience, and group dynamics. Psychological factors (as consensual attitudes or authority biases, in addition to cultural factors as pedigree, or gender prejudice, and environmental factors such as meeting facilities, meeting location, travel times and housing, as well as economic factors such as remuneration, funding, or nepotism) are not observable in statistical data analyses. The unclear quality of expert elicitation and the reported psychological phenomena involved in the construction of opinion-based scenarios gain a dramatic momentum if we recall that these scenarios are presented as possible (sometimes even consistent) futures to decision-makers.

Using the technique of cross-impact analysis as an example for a judgement-based scenario construction technique, I would like to discuss the fourth point. However, my criticisms apply to all techniques based on expert opinion. Consistent scenario construction is a relevant prerequisite for the legitimation of energy model results and one approach addressing this issue is cross-impact analysis [37, 38], reviewed by [39]. Briefly described, the method presented in [37] defines the so-called *descriptors* figuring as a representation of context assumptions for a scenario. Experts are elicited to stipulate the reciprocal influence of the descriptors and via an algorithm compatible context assumption combinations are derived. Although cross-impact analysis (CIB) is preferable to an unsystematic assumption choice and numerical value stipulation, the method has some drawbacks.

First, the number of so-called descriptors is limited due to both practicability, cf. p359 [37], and reliance on experts. In contrast, due to the Monte Carlo simulation, the number of potential *influences* (corresponding to descriptors) is not limited computationally using Bayesian model averaging. Expert knowledge is not required but can be included through prior choice.

Secondly, the CIB methodology defines *consistency* in a particular manner based on expert judgement. The descriptors are evaluated with respect to their *reciprocal influence*, one-on-one, as assumed by the expert interrogated. The basic principle describing the consistency is the *principle of compensation* that says "two opposing influences on one state are to be judged as equally strong if their effects can compensate each other. If it is to be estimated that one of the influences predominates during a confrontation, this one shall be judged higher, i.e. be given a higher number." p.340 [37].

Underlying the principle of compensation are three arguable assumptions (1) dominance is generally valid, (2) dominance can be extended, and (3) dominance is pairwise invariable if additional descriptors are simultaneously considered in a scenario.

Assumption 1 is a general statement for two descriptors that can be for example "+ 1" meaning "weakly promoting direct influence", "- 3" meaning "strongly restricting direct influence", or "0" meaning "no direct influence". The problem is that such an evaluation needs to be related explicitly to another context (i.e. descriptor) and contexts may vary in different scenarios, and for different experts. For example, if an expert generally judges a descriptor as weakly promoting another descriptor, she (implicitly) presupposes conditions where the statement is valid, what I call that a *state of the world*. But the *presupposed* conditions are exactly those that are varied in different scenarios. In contrast, if we use statistical data and the BMA method, we evaluate the reciprocal influences of descriptors in *many different states* of the world, to be precise, all those states of the world, our data record contains. This is in fact what statistical analyses do, regarding the rate of change of a variable (i.e. descriptor, influence) relative to all other variables. I call the property of a descriptor to be influential in many different states of the world *stability*. It is advantageous to formulate scenarios based on stable relations in the target system as these relations are most likely to hold in scenarios too.

Assumption 2 implies that dominance of descriptors can be extended to *hypothetical states* of the world. A hypothetical state of the world is a state with unprecedented conditions. In contrast, a provable possible state of the world is a combination of occurrences in the world observed in the past. If the scenario construction is based on intuitive or hypothetically possible relations in the target system, we are confronted with two kinds of uncertainty: (a) assumption uncertainty and (b) representation uncertainty. Now, (a) is a natural uncertainty for every future scenario and, in fact, it is the very reason why we construct scenarios. Assumption uncertainty arises because we do not know what will happen in the future. So, if we assume a numerical value we *stipulate an assumption for an input variable*, we face assumption uncertainty as the value might prove to be different in due course. In contrast, (b) representation uncertainty means that the relations in the target system are, at least partially, flexible and unknown. Intuitive cause-effect relations or expert opinion on reciprocal relations can be empirically adequate; however, the only way to *evaluate the adequacy is to compare the assumed relation with actual target system behaviour in the past*. This amounts to a statistical analysis. The proposed method merely circumvents the introduction of additional uncertainty due to poor system relation representation and straightforwardly uses the stable relations that are statistically confirmed for the observation period in the target system. BMA for consistent scenario construction extends to hypothetical states of the world too but allows specifying that a scenario *represents* a hypothetical state and allows for a clear communication as to why and to which extent the hypothetical scenario differs from the observations in the considered historical period.

Assumption 3 is, in a sense, a *ceteris paribus* assumption known to be an idealisation. It is related to assumption 1 but now I mean consistency within the same scenario. I would like to give a simple, intuitive counterexample to this assumption based on the interrelated nature of the energy system with social, environmental, and political systems. Consider the three descriptors: gross domestic product, world tensions, and oil price, taken from the example in [37]. The experts are supposed to judge the influence of the gross domestic product on the world tensions, and so forth, pairwise. But if an expert was asked to assess the relevance of the oil price on the world tensions *given* a high gross domestic product (a relaxed economic situation), it may be different than when this relation is assessed given a low gross domestic product (distressed economic situation). Whatever the expert's subjective reasoning behind the supposed impact of one descriptor on another would be, it must not necessarily hold true when a third, a fourth, etc. descriptor enters the picture (and is variable).

However, for a human being, also for an expert, it is difficult to assess the strength of relations between the descriptors when the number of descriptors exceeds two or three. I suppose the cause for this is that human's reason about correlation and reasoning is difficult when relations become multi-dimensional. In contrast, the statistical BMA model can (and does) take dozens, even hundreds, of potential combinations of descriptors (influences) into account and assesses their explanatory power with respect to all other descriptors in a model simultaneously. Thousands of such models are investigated in the MCMP sampler, which is not restricted by human capacity. And, perhaps most relevant, the explanatory power (equivalent to the "cross-impact judgements" of the experts) is *not based on human reasoning, but on statistical confirmation*. Consistency is hence defined as non-contradiction with empirically confirmed states of the system (i.e. with statistical fact) rather than an expert *opinion*.

In addition to being non-contradictory, the strength of evidential support in a relation (given the data used) can be analysed *in principle* in detail for any region, any time resolution, any historical period, and any type of statistical information by one person. In contrast, even if the group of *available* experts has remarkably diverse backgrounds (what brings about other problems, e.g. language issues, incompatible implicit worldviews), this is for judgement-based scenarios *in principle not possible*. All experts employ human reasoning. I would like to remark that this is a fundamental difference to all techniques presented in Table 1 with information source opinion. To be clear, I do not say that experts err in principle their assessment of relations, even if a multi-dimensional scenario is constructed from their two-dimensional assessment. But we cannot assess the quality of the "human black box" directly. Whatever the reasoning behind the expert's opinion of the assessed relation would be, it should be empirically adequate, and this adequacy needs to be evaluated. Using statistical data in the BMA method, we can obviate the risk of empirical inadequacy due to human reasoning naturally involving psychological effects and computational limitations.

Another drawback of a judgement-based scenario design with a semi-formal ordinal categorical formulation is exemplified here using cross-impact analysis. However, the criticism applies to all techniques where quantitative scales are used as a way to *represent* expert judgement as an information source, rather than using a quantitative scale *as* an *information source*. In contrast, the quantitative values of statistical data represent statistical evidence, i.e. observable empirical facts. Reciprocal effects and dominance in the semi-formal formulation are represented for example as "+1" or "-2", "medium" or "low", and it is unclear according to what procedure these formalisations are translated into values acceptable by an energy model. Even in case the assumed relations were right, the scenario formulation of



numerical assumptions seems to not be based on a methodologically stringent interpretation procedure. In other words, the method invites interpretation opacity and uncertainty in the numerical value stipulation necessary for an energy model.<sup>5</sup> Different modellers may interpret the impact of the descriptor oil price “quantified” as +2 on the descriptor gross domestic product in completely different numerical value stipulations “fed in” the energy model. The CIB method in fact provides “sets of consistent assumptions” for a number of descriptors forming the scenario itself. To me, it is unclear if the interpretation of the scenario B1 (p. 343 [37]) consisting of moderate world tensions, medium borrowing industrial countries, strong cohesion of OPEC, an oil price of 35–50\$, and 2–3% world GDP growth would be interpreted by different energy modellers numerically in the same way. If an oil price of 35\$ is equally well confirmed as an oil price of 50\$, the “rough” scenarios are possibly too general for energy models being highly sensitive to numerical assumption changes.

Secondly, how would “moderate world tensions” be numerically interpreted? Thirdly, if it was not included in the model numerically, how is the scenario-relevant context assumption (and its non-inclusion in the energy model) accounted for? The proposed BMA method for consistent scenario construction uses statistical data with explicit numerical assumptions for quantitative statistical data. For example, the descriptor GDP is not *interpreted* but simply is the GDP in units, e.g. billion EUR [40]. The derived statistical data, as for example GDP per capita, inherit their meaning from the basic statistics and the transparently published computation procedure by statistic agencies. Vague scenario formulations as “high GDP” or “moderate world tensions” reflect the relative nature of *every* judgement. In the case of CIB, the judgement expresses the expert’s opinion and the modeller’s interpretation stipulating the numerical value. In the case of BMA, the judgement expresses the statistical data used, e.g. the stipulated value lies in the third quartile of the data (in the highest 25%). If data are transformed [41] for BMA analyses, a systematic reversal to derive a numerical value for scenario assumptions is possible. In some contexts, categorical data or ordinate statistical data are available. They can be included in the scenario construction, but in contrast to semi-formal quantification, BMA is not limited to them. Their interpretation is straightforward from the statistical sources, for instance, defined indices<sup>6</sup> or systematic data treatment as, e.g. seasonal adjustments.

The empirical assessment’s importance is particular to energy modelling and policy advice scenarios in general. The severe consequences of experimental trial-and-error strategies in terms of economic impacts on society demands for a scenario construction method that takes into account assumption compatibility and strength of evidential confirmation. As energy modelling is the best

“experiment” of the energy system we have, it is meaningful to make sure that the constructed scenarios relate to what was possible in the target system and to estimate how much deviation, we could say “novelty”, a scenario introduces compared to the states of the energy system as documented in statistical data. In this way, we estimate and give credit to the efforts we take in our pursuit of hypothetical system states we dream of in scenarios.

Expert knowledge is part of this process; however, evaluating the *relations* of target system phenomena is *not* an adequate scope for expert judgement, as speculations about relations are more adequately scrutinised with statistical methods. If *in fact* one phenomenon in the target systems happens to systematically change in concert with other phenomena, statistical methods can point to that – *whatever the reasons for the simultaneous changes are*. Experts can interpret changes and guess reasons for correlations if the *true data generating process* is unknown. But a guess remains a guess, even if performed by an expert. Pretending expertise on something one does not understand is not professionalism but unscientific conduct. And as we cannot directly evaluate expert understanding and the “correctness” of reasoning as confirmed by evidence we might as well guess ourselves or directly consult evidence, as we do in statistical analyses.

If an expert *knows* a true data generating process, expert judgement is valuable for scenario design. For example, if an expert knows of signed contracts to build a pipeline, she knows a reason for a change. In a scenario the phenomenon “infrastructure capacity” can be adjusted according to the expert’s opinion in due time of the scenario.

### Scenario definition and uncertainty evaluation

The first step when creating BMA-scenarios is defining (i.e. choosing upon known) phenomena in the target system we want to make a scenario of for an energy model. I refer to such phenomena of interest as “scenario phenomenon”. Public invitations to tenders for modelling exercises often describe the scenario phenomena of interest in detail. For example “the impact of unconventional gas production on the electricity price of country X” could be a phenomenon of interest. But also phenomena, as for example “implications for the energy system of a legal threshold for CO<sub>2</sub> emissions in country X”, could be subject of a scenario. Let me replace “country X” with “Nicastan”, an in-existent country, to improve legibility and intuitive understanding coupled with the generality of the presentation.

What interesting phenomena for a scenario depend on the person who asks the question—the societal group concerned. For energy models, governmental stakeholders sometimes investigate consequences of debated political measures by scenarios. Industrial stakeholders might be interested in energy system developments given different investment decisions. Public stakeholders

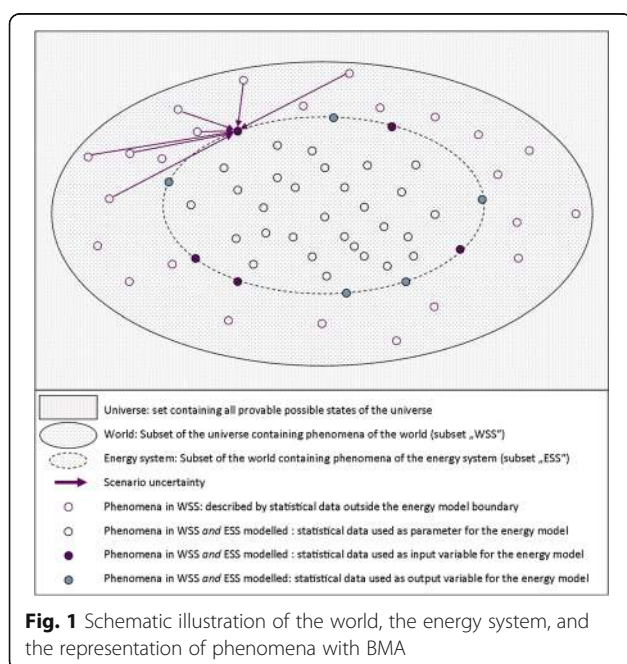
might be interested in tax burdens or security of supply scenarios, to name just a few. For a consistent scenario with BMA, the phenomena investigated are not limited in number and nature (social phenomena, environmental phenomena, economic phenomena) provided statistical data are available. I return to that limitation in the “Discussion and limits of the approach” section.

We discern between modelled phenomena and context phenomena defined by the model boundaries of a specific energy model. The consistent scenario construction is solely concerned with the aptitude of an energy model to represent the scenario given the energy model design. The relevant parts of an energy model for scenario representation are the *input variables* and sometimes *parameters*. For the presented method of consistent scenario construction, we do not need any judgement of the energy model’s quality, as we only look at an energy model’s capacity to represent the scenario, and construct numerical assumptions consistent with statistical data and expert opinion. In other words, the consistent scenario construction as presented here is independent of consequent processing by an energy model, its mathematical formulation, and internal error propagation, what makes the method applicable to a large number of quantitative modelling techniques. The “interface” of scenarios and energy models are typically input variables; therefore, the following example uses a typical input variable. Figure 1 depicts a schematic illustration of scenario phenomena, statistical data, energy model representation of the energy system (ESS), and the target system (WSS), i.e. the “real” energy system.

The *rectangle* should be considered extending beyond the graph as “the universe”; it is associated with all possible states of the universe, one at a time, according to the time increment we choose for the record. The *solid ellipse* depicts schematically “the world”; it is a subset of the universe associated with all possible states of the world. Dots in the WSS subset indicate existing phenomena. The *dotted ellipse* indicates “the energy system”; it is a subset of the world associated possible states of the energy system. The circles indicate *statistical data* depicting (a collection of) phenomena in WSS as follows. *Purple circles* schematically depict phenomena of the world subset (WSS), explicitly outside the energy model boundaries. *Blue circles* depict phenomena of WSS and the energy system subset (ESS); statistical data of these phenomena are accounted for in an energy model in the form of *parameters*. *Solid purple circles* depict phenomena of WSS and the energy system subset (ESS); statistical data of these phenomena are accounted for in an energy model in the form of *input variables*. *Solid blue circles* depict phenomena of WSS and the energy system subset (ESS); statistical data of these phenomena are accounted for in an energy model in the form of *output variables*. The arrows show the BMA description of phenomena, to be read as “explains” in the direction of the arrow tip (Fig. 1).

I created a fictitious scenario and elaborate the different steps for the example. The *storyline* of the scenario is translated to numerical assumptions for input variables. Suppose the government of Nicastan is interested in a low-cost scenario for natural gas because in the real world, recent developments in unconventional gas production indicate a lasting period of low natural gas prices. We want to investigate the effects of such a low-cost period on the energy system in Nicastan with the fictitious energy model My-model.

Assessing the empirical adequacy of My-model is our aim, expressed as a lower bound of uncertainty that My-model is apt to represent this scenario. Finding the numerical assumptions, value that best describes the scenario follows. Phenomena of interest in this example are different target system phenomena, including the unconventional gas production. The formulation of the scenario is captured in the input variable “natural gas price” of My-model. So, if the energy models’ assumption should reflect the bearing of unconventional gas production on natural gas prices in Nicastan, there should be statistical evidence for that. The strength of the supposed relation between the energy model input variable (the price for natural gas) and the scenario-phenomenon of the target system (unconventional gas production) is assessed by statistical evidence.



It is, in a straightforward sense, evaluating whether the considered scenario *can* be described by the energy model at all (and how well), based on the energy model design. The input variable of My-model represents many phenomena; with BMA, we can find out how relevant the one phenomenon we are interested in (the unconventional gas production) is. The uncertainty assessment estimates how *relevant* the relation of *scenario* phenomena and energy model input variables is, based on observations. For statistical data of phenomena outside the energy model boundary, it is a quantitative assessment of *scenario uncertainty* as defined by Walker et al. [42]. This is a fundamental assessment because it concerns *the grounds of justification for any energy model result claiming to illustrate the scenario*.

I would like to emphasise that *hypothetical* energy system state scenarios (i.e. *not* designed to represent energy system states observed in the past) *are not in conflict* with using statistical data for scenario construction. Both energy models and scenarios are designed to describe *possible* energy system states. Statistical data are evidence for a system state *to be* possible. Equivalent to energy model calibration, consistent scenario construction uses the evidential basis to derive statements about the possibility of scenarios. This is of interest for scenarios describing unprecedented energy system states (hypothetical scenarios) because it provides an expectation based on what we know to be possible.

The statistical data are used to analyse relations in the target system of relevance for the scenario representation in an energy model. To do so, *classical* statistical methods could be used. These methods have some disadvantages, for example, biases in the choice of explanatory variables by the scenario constructor. Using Bayesian model averaging (BMA) [2, 3] allows for both expert opinion and statistical likelihood and reduces the risk of model misspecification.

The mathematical formulation representing a general relation between the influences (also called explanatory variables) and the dependent variable is chosen, where the input variable of the energy model (the natural gas price) figures as the dependent variable in the BMA model. Evidentially, production data of unconventional gas is considered as an explanatory variable, but for an empirically adequate representation, statistical data of any kind which is suspected to be relevant can be included. Our aim is to find out what phenomena in the target system the input variable actually represents, according to statistical evidence. We employ BMA to evaluate the relevance of different phenomena for the input variable, taking full advantage of the fact that BMA “sorts out” influences that cannot contribute to explaining the dependent variable. Defining the prior distribution also allows us to take into account the expert opinion on the number of explanatory variables expected to influence the dependent variable.

The BMA results are models containing different explanatory variables. The models are ranked according to their explanatory power which is expressed as a posterior model probability (PMP). The relevance of individual explanatory variables is expressed as posterior inclusion probability (PIP). These two values per se already encode highly relevant information for scenario design. In a sense, they are the empirically confirmed counterpart to the expert opinion of scenario design by cross-impact analysis. In contrast to the assumptions of cross-impact analysis, it is observable that the inclusion or exclusion of variables is not reciprocally balancing. Even for the simplest case of one explanatory and one dependent variable (comparable to CIB), we find that changing the role of the variables does not always lead to reciprocal balancing. The balancing is even less observable, given a growing number of explanatory variables. However, the CIB idea that influences are “impacting” each other according to the relations of phenomena in the target system is also captured in the BMA models. Unfortunately, the true data generating process is often unknown, and in systems influenced by human decision, individual relations may be difficult to find in statistical data. What we seek to analyse is the strength of “stability” in the relations of influencing phenomena vis-à-vis the dependent variable.

Briefly described, the BMA method takes statistical data of all influences (i.e. context phenomena of the target system, also called explanatory variables) and uses a Markov Chain Monte Carlo importance sampling method to “build” different models. “Different models” are just equations arranging explanatory variables. The explanatory power of a model is assessed, and the sampler builds another model and assesses the explanatory power of that model for the dependent variable. The number of explanatory variables forms the “model space” defined as the possible combinations of explanatory variables. For example, if an energy model input variable (the dependent variable) is suspected to be influenced by 18 phenomena, the model space ( $2^{18}$ ) contains 262,144 different models that could describe the relations in the target system. A stipulated prior distribution depicts the scenario constructor’s opinion on how many influences she considers relevant. If the constructor has no expertise at all, a flat prior (non-informative prior) reflects ignorance,<sup>7</sup> and the models are determined based on statistical evidence, called likelihood. The posterior model probability (PMP) is proportional to the product of prior model probability and the marginal likelihood of a model. The marginal likelihood of a model in the model space is the probability of the data given the specific model. With increased computational power and advanced importance of sampling techniques, the employment of BMA has risen since the 1990s considerably. Today, different ready-to-use options for mathematical software are available which relinquishes the need for programming skills

and tedious construction of adequate software. For the example in [1], I used the R package BMS by Zeugner [43], which offers abundant possibilities of BMA specification and a complete set of standard features sufficient for a consistent scenario construction.

The BMA approach amounts to saying “It is by at least 93% uncertain that My-model can describe a low-cost natural gas period due to developments in unconventional gas. The scenario is introduced in My-model via the input variable “natural gas price”, but low natural gas prices are also explained by ...”. In the following section, two options for consistent scenario construction are discussed: the consistent scenario construction and the representation quality assessment for (1) a given energy model and (2) an adapted energy model. The latter generally allows for a better representation of scenarios, as the model is adapted to account for the most relevant phenomena influencing the scenario phenomenon.

#### Consistent scenario construction for a given energy model and an adapted energy model

Typically, a scenario is introduced to an energy model by input variable adjustment; in My-model, it was the natural gas price. Having gathered statistical data of all influences, we suspect to be related to the natural gas price and the scenario phenomenon “unconventional gas production” we have investigated the empirical adequacy of the scenario in the fictitious My-model. The BMA analysis delivers an indication of which phenomena are relevant for the input variable and how relevant these are relative to each other (and all phenomena we suspected being relevant).

We can improve the scenario representation quality of an energy model by changing the energy model design. In particular, we can decrease the energy models’ representation uncertainty of a scenario phenomenon by including the relevant input variables and/or parameters in the energy model. To do so empirically adequate, we need to find out which phenomena in the target system have influenced *the scenario phenomenon* in the past. If our intention is to create a consistent scenario, rather than justifying that our input variable assumption represents a scenario, the consistent scenario construction method is suitable.

In consistent scenario construction for adapted energy models, we apply a different reasoning. In the target system, the emergence of a scenario phenomenon is often only possible if different related phenomena develop in a specific way. In other words, the scenario phenomenon, due to the interrelated nature of the target system, *necessitates* other phenomena’s occurrence in a specific way. In the example, the scenario phenomenon of high unconventional gas production necessitates high natural gas prices; otherwise gas production is not profitable.

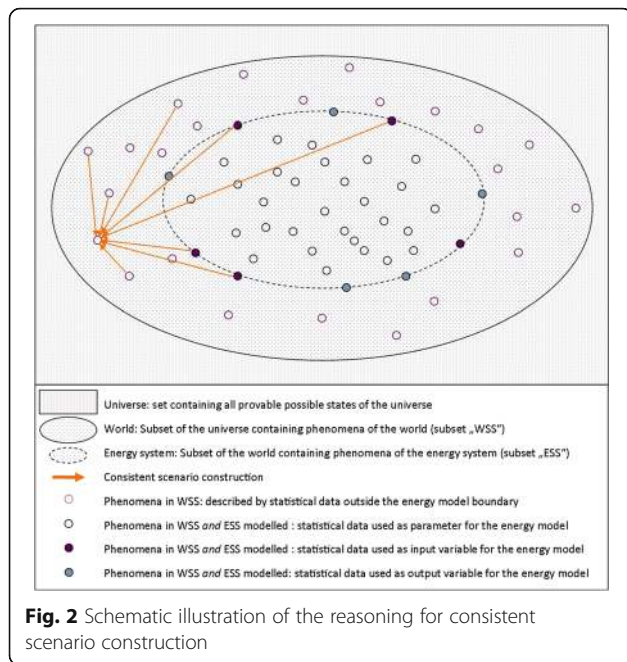
We denote the BMA variable representing the scenario phenomenon as the *dependent variable* and span the model space over all influences suspected to impact the phenomenon, *including* input variables and parameters of the energy model as explanatory variables. In that way, we investigate what potentially explains the scenario phenomenon’s emergence in the target system. If we consistently construct a scenario, our aim is to “present to the model” a world in which the scenario phenomenon is present in the way we are interested in (i.e. *high* unconventional gas production). To do so empirically adequate, we depict the scenario phenomenon “indirectly”, supposing its emergence is dependent on the influences to be in a certain way.

A phenomenon is “presented to the model” in a coherent and consistent way by an adequate choice of numerical values for input variables and the introduction of relevant input variables and parameters in the energy model, called energy model adaption. The more relevant input variables and parameters an energy model contain, the better our representation of the scenario would be. Again, we have BMA select for us which influences are systematically increasing or decreasing the observable value, but now, our dependent variable is the scenario phenomenon. All we have to do is be creative in our suspicions of what target system phenomena influence the scenario phenomenon and perform statistical data handling.

Including the data of energy model input variables as explanatory variables allows us to evaluate how well we can already represent the scenario phenomenon in the energy model. We can evaluate this quite precisely using the PIP of the influences that are energy model input variables relative to phenomena not modelled in the energy model. Also, we find what other influences in the target system are relevant for the emergence of the scenario phenomenon to efficiently *adapt* the energy model with the objective to increase the empirical adequacy of scenario representation. Figure 2 is comparable to Fig. 1 but depicts the different reasoning of consistent scenario construction in contrast to input variable justification (Fig. 2).

Let us suppose we want to construct a scenario representing an increase of unconventional natural gas production with My-model; using statistical data for the quantities of unconventional natural gas produced as the dependent variable, we can find out what phenomena in the target system are most influential. As explanatory variables for BMA, we use both My-model input variable data and data outside My-model boundaries (we suspect to influence the unconventional gas production). We can investigate as many influences on the phenomenon as we want to make a scenario. The best BMA model in terms of posterior model probability is represented in Table 2 (all data and variables are fictitious).





**Fig. 2** Schematic illustration of the reasoning for consistent scenario construction

Using the PMP of the best BMA model, it is possible to evaluate the probability that the phenomenon of our scenario can be described with the influences (let us assume a fictitious PMP of 16%). My-model does not contain all phenomena to describe the scenario with the highest probability. We can base our decision whether to adapt the model to better represent the scenario on that evaluation on the BMA results. Using the influences' PIP, we quickly find which energy model adaptations are most relevant to increase the scenario representation quality. The energy model adaptations may concern both input variables and parameters, depending on the energy model design. If not all influences are used to construct a scenario, the lower bound of uncertainty needs to be adjusted to the PMP of the BMA model consisting of the influences that are used. This is a BMA model with higher uncertainty if it is not the best BMA model in terms of PMP. The scenario construction with *all* relevant influences is a scenario representation according to statistical evidence of the scenario phenomenon with *least* uncertainty for the proven possible states of the world (observed values).

Using the consistent scenario construction approach, we pay attention to the interrelated nature of the target system and design our energy model to represent the scenario phenomenon empirically adequate. This approach is also applicable for phenomena, as for example enactment of a law. This process may at first sight be self-sufficient in that it is a human decision governed by human sovereignty, not grounded in statistically observable interaction with other target system phenomena. Upon second thought, we may find that the decision of actually enacting the law indeed necessitates a specific target system state. For example, the scenario of a law prohibiting the emissions of CO<sub>2</sub> exceeding some defined quantity may necessitate a minimum quantity of CO<sub>2</sub> emissions, to be relevant at all. For human decision-governed scenarios, an investigation of most relevant influences on the decision improves the representation of a scenario.

**Results: consistent numerical value estimation**

With BMA, we can truly *construct* scenarios. Having identified the most influential phenomena, we insert the stipulated values for the time period we want to project in the explanatory variable data record. I would like to remark that this is in the same instance a transparent assumption documentation. We use the BMA model of choice to *compute* the predictive density for the dependent variable based on the stipulated values for the influences. This means, expressed for readers of a scenario study, “the adjusted input values of the energy model correspond statistically to an unconventional gas production of xy”.

The PMP of the BMA model determines the lower bound of uncertainty for the scenarios. In addition, we dispose of statistical criteria, for example, we can assess whether the predicted numerical value of the scenario phenomenon is within the double standard deviation or not. To be clear, the procedure is (1) stipulation of values for influences; we simply fill in the data record for the future period of the scenario. We can orientate on historical “highs and lows”, and include expert knowledge, as mentioned before if the expert knows of signed contracts, the additional available capacity in the year of expected operation is entered as data. Next step (2) is computing the predictive densities for the scenario

**Table 2** Fictitious evaluation of scenario representation for My-model

BMA model-dependent variable	Influences	Relevance of influence (PIP)	Influence modelled in My-model
Unconventional natural gas production	Natural gas price	0.99	Yes
	Oil price	0.85	Yes
	Conventional natural gas production	0.84	No
	Electricity consumption	0.62	Yes
	LNG infrastructure investment	0.59	No
	Gas storage capacity	0.45	Yes



phenomenon. Given the stipulated values, we *compute* the numerical value of the scenario phenomenon. In doing so, we can explicitly explain both the numerical value of the scenario phenomenon (due to observed relations) and the representation quality of the scenario in the energy model, as captured in the input variable(s). Finally, (3) communicate a complete and transparent documentation of the scenario construction.

An exemplary communication for a scenario constructed according to case (b) BMA *scenario construction for an adapted energy model* is available in the [Appendix](#).

### Discussion and limits of the approach

In this section, I would like to discuss the limitations of the approach and address difficulties I experienced when carrying out the method. However, I may be unaware of some problems and limitations, so I do not claim this discussion to be exhaustive.

A practical limitation of the approach is its extensive use of statistical data. This naturally involves both access to statistical data and availability of statistical data in the first place. Today, many statistical data are freely available from official sources, for instance, Eurostat or USA.gov, and often data can be bought. Researchers entertaining energy models are likely to already dispose of a significant amount of relevant data for calibration purposes. Efforts made to improve access to statistical data often prove to be advantageous in more respects than consistent scenario construction. It is my opinion that financial burdens for subscriptions, in particular for research collaborations or larger scientific institutions, are vindicated by the potential impact scenario studies can have, if used for decision support.

Data availability is different from data access. Unprecedented phenomena may have no data record at all. In this case, it is indispensable to transparently communicate that the scenario is incommensurable to any known energy system state. The response of the energy system to a completely unprecedented scenario phenomenon is highly uncertain as the assumption of well-established and long-standing observable relations to hold is uncertain. In less dramatic cases, some methods may be useful. The first method is indirect phenomenon description with different influences data so that as many suspected data generating processes as possible are included in the analysis. Phenomena described with statistical data of various scientific disciplines allow for a multi-dimensional description of the phenomenon if researchers are not primarily focussed on their scientific discipline.

Another method applicable in *some* contexts is data scaling. Scaling is *not suitable* for data known to depend on spatial or temporal relations in the system where they are collected, unless that system is the target system of

the scenario modelling. For example, data of energy efficiency improvements of a technology product are scalable, because the data are regionally independent, i.e. the same technology product would have the same technical efficiency everywhere in the world. In contrast, data of consumer behaviour of a region are not suitable for scaling, because the data depend on regional aspects such as income, social system, political stability, or any cultural aspects of that region, to name just a few.

A third method is data generation. The applicability of data generation is dependent on the phenomena of interest. Consistent scenario construction provides a guideline to survey design, as we can find potentially relevant phenomena for the scenario by statistical evidence. Say, a city wants a scenario of residential CO<sub>2</sub> emission reduction, but the data of the city is lacking. Based on a BMA analysis with data of a comparable city, we find what phenomena are statistically most relevant and hence what data are of primary interest in the city. A survey designed to generate the data for the scenario then contains questions regarding the influence with highest relevance in terms of PIP (e.g. primary heating fuel in the residential housing sector) and questions on statistically relevant data in terms of high PIP (e.g. heating technology efficiency). Of course, other data suspected to be relevant, say as a particularity of the city, should also be generated. If data are partially available, the survey design should account for relevant influences additionally to or for in-depth surveys. Lastly, a method to increase the data availability is open source policy of anonymised data-managing enterprises or research institutions. However, this is more a political question than a method a researcher can apply ad hoc. Nonetheless, it is important to mention it and engage in the discussion.

Some experts hold that some phenomena simply *cannot* be represented statistically, and expert judgement is the only way to evaluate such phenomena. Human psychological phenomena or socially arising phenomena are typical examples showcased. I agree with sincere reservations. Take the example of Weimer-Jehle in [37] “world tensions”. An expert evaluating the impact of world tensions may subjectively have a clear idea of what is meant. However, different experts are likely to have different interpretations of “world tensions” due to the vague definition that may lead to the incommensurability of expert opinions.

I contend that for many psychological or social phenomena, there *are* statistical data with a precise definition of *what* is evaluated by the statistic and *how* the data are generated. Using such data enables recipients of scenario studies and modellers to understand how a phenomenon is interpreted in a scenario. Statistical data, for example, national warfare expenditure [44] or arms trade data<sup>8</sup> [45], or the United Nations Office for Disarmament Affairs UNODA databases, can provide, in my opinion, a suitable

statistical representation of “world tensions”. In addition to the data for armed conflicts [46, 47], corruption indices such as the Corruption Perceptions Index (CPI) by Transparency International [48] or the World Bank’s CPIA database [49] are available, to name just a few. Let us suppose we construct a scenario, and the influence “world tensions” were expressed in statistics. We could combine corruption data and warfare expenditures of relevant nations for the scenario. Let us say for an exemplary scenario in Nicastan, the country imports weapons from Germany. We could consult the United Nations register [50] and find out how many weapons were exported in the past years to Nicastan. Our scenario is constructed such that the *meaning* of “world tensions” corresponds, for example, to an increase of arms imports from Germany by four times the mean over the last years. We can even specify which weapons we include in our assumption and what scope our assumption has (e.g. “transfers between UN member states”). To me, this is a storyline understandable to recipients of scenario studies, and significantly less ambiguous than the CIB evaluation “strong”, “moderate”, or “weak” world tensions p. 339 [37]. For psychological and social phenomena, a large variety of data are available, for example, the World Values Survey “is a global network of social scientists studying changing values and their impact on social and political life” [51]. OECD social databases [52] document and allow access to social data, as well as the United Nations Statistics Division UNDATA databases [53], or research institutions hosting databases [54].

To reduce practical limitations computing “milestone years” and interpolating between those may sometimes be suitable. However, it is an advantage of a BMA consistent scenario construction that the scenario design can significantly differ from linear paths. This is possible because the computation of the corresponding numerical value of the scenario phenomenon (predictive density) takes stipulated numerical values for influences into account. Stipulated values of influences can represent any “shock” or atypical development, in consonance or individually, and the result of the best BMA model will represent the expected value of the scenario phenomenon based on the relations observed in the historical period considered. In this sense, intuitive scenario design has its place in BMA scenarios, as well as expert opinion.

I would like to emphasise that using statistical data for social and psychological phenomena does *not* mean that constructed scenarios are identical to past phenomena. Stipulating the numerical values of relevant influences allows for full creativity and exploration of hypothetical possibilities. Using the BMA model to compute the predictive density means that interrelations and impacts in the past are the basis of expectations for the future. Evaluating the *bearing* of an influences’ hypothetical value on other phenomena is *always* based on system relation assumptions—

be it in the form of expert opinion or observation record. Statistical data are evidentially confirmed system relations (and provable possible) what renders scenario construction (1) consistent with what we observed, (2) intersubjective, (3) systematic, (4) understandable, and (5) comparable. I am aware that statistics can be manipulated and may not mirror the “true” state of the world. Nonetheless, I am convinced that they are a better record of fact than subjective observations. In any case, the data are clearly explicable in contrast to ambiguous classifications (“high”, “low”, etc.) experts can commit to.

Today, scenario studies typically present storylines, but lack of (1) transparent documentation of scenario assumptions *and* implicit assumptions; (2) a (energy) model assessment of scenario adequacy, i.e. input variable aptitude to represent the scenario; and (3) scenario assumption evaluation (stipulation of numerical values). Omitting communication of that information averts scrutinising the representation quality of the scenario for a given (energy) model and the degree of confirmation for a scenario (proven possible or unprecedented hypothetical). In my view, providing that information is the (energy) scenario modeller’s assignment because a scenario study recipient has no means to retrieve that information from (energy) model results or a storyline narrative.

The BMA method allows for the inclusion of any stakeholder choices statistically evident. The behaviour of society members, consumers, the industry, or the government, is statistically recorded in a non-discriminatory way in a multitude of dimensions, for instance, periodicity in time, geographic area, social and cultural categories, or economic benchmarks. The pedigree of these data, their collection, post-processing, financing, hosting, and availability are methodologically rigorous, transparent, and non-discriminative. Scenario construction by the BMA method does not oblige the scenario constructor to claim “expert knowledge” or have skills superior to any person capable of handling statistical data. Some programs suitable to carry out a BMA scenario construction are available without charges, for example, R statistical software package BMS [43]. The scenario constructor need not belong to some exclusive group of “experts”, where the quotes indicate the ambiguous status of such an appointment, as there are no transparent quality criteria for being an expert.

The *opinion* of a scenario constructor can be included by numerical value stipulation. Opinion is systematically and methodologically *levelled* in the light of statistical data. In other words, the advantage of BMA over judgmental scenario construction techniques is the *embedding* of opinions in the relevant context’s empirical record. That levelling helps to alleviate psychological effects reported in the literature. In the BMA method, expert *judgement* is needed to choose the data of the phenomena suspected to be influential, decide computational

procedures (e.g. birth-death-sampler), define priors, and pre- and post-process data (e.g. outlier elimination, reduction of results for interest groups). These choices are contestable facts (not a “feeling”), subject to criticisms, and produce testable results; if the expert changes her judgement and say another sampler routine, the result of BMA changes.

Lastly, the BMA scenario construction method can better account for the real-world complexity of phenomena interactions than the limited cognitive capacity of human reasoners. However, statistical analyses are limited in their capability of detecting relationships. Considering the evaluated uncertainty for scenarios as a *lower bound* expresses that awareness; an uncertainty of *at least* x% corresponds to the best case of actually acknowledging the relevant relations of phenomena. The BMA analysis must not be considered as a tool completely characterising the phenomena present and (variably) interacting in the target system. Rather, the method provides an additional “view” of the target system beneficial for scenario design.

The distinction between provable possible assumptions (based on statistical evidence) and hypothetical assumptions is a novelty and significantly improves the scenario study recipients’ aptitude to evaluate the future scenarios *on their part*. The presentation of scenarios referencing tangible storylines (e.g. the assumption for the gas price in the first year of projection is the mean of the statistical data period from 1995 to 2015), is in my view, more comprehensible than storylines referencing abstract assumptions as for example *moderate* world tensions, *medium* borrowing industrial countries, and *strong* cohesion of OPEC, cf. [37]). So, perhaps the most important novelty of the approach is the possibility to communicate to decision-makers the associated uncertainty in easily understandable terms.

### Conclusion and further research

I would like to conclude the discussion by briefly highlighting the advantages and disadvantages of the proposed approach. The motivation to advance a systematic scenario construction technique for quantitative models is vested in the current practice of opaque scenario design in many disciplines, including energy-economic modelling. Models used in decision support generally aim to *represent* the target system and the relations of phenomena observable in the world, yet they fall short in assessing their representative quality for scenarios.

Scenarios based on intuitive expert opinion or based on mutual agreement of (energy) modellers and sponsors, risk being an inadequate representation of the scenario phenomenon, and most importantly, they defy evaluating their empirical adequacy. This poses a problem for

recipients of energy scenario studies, because the *results* of scenarios depend on a thorough design representing the phenomena of interest, in particular, if the results figure in decision support.

The advantages of the presented BMA approach for consistent model design are:

- Transparent scenario assumption documentation
- Evaluation of the empirical adequacy of assumptions via an uncertainty assessment
- Evaluation of scenario assumptions as statistically confirmed or hypothetical
- Methodological rigour independent of subjective expertise of the scenario constructor, although with the possible inclusion of expert opinion and creativity
- Adaptability for individual energy models (specific input variables, time resolution, geographic scope)
- Formulation of apprehensible scenarios referring to the observed phenomena, e.g. “an economically flourishing period as observed in the years 2003–2005”, rather than general formulations as “high economic growth”
- A clear indication of phenomena influencing the scenario phenomenon in the target system (using PIP’s)
- A clear indication of efficient adaption of an energy model with the aim of increased quality of scenario representation (using PMP’s)
- The possibility to *construct* scenarios in a straightforward sense with concurrent evaluation of the scenario phenomenon’s numerical value and its probability based on empirical evidence (stipulation of numerical values of influences and consequent calculation of the predictive density for a BMA model with given PMP)
- Inclusion of expert opinion (via prior distribution and stipulation of influence values in the projection period) and statistical evidence (via statistical data of the historical period considered)

Opposed to these advantages are practical and conceptual limitations of BMA scenarios. Completeness of this list of limitations is not claimed.

The prerequisites of the presented BMA approach for consistent scenario construction are:

- The need to conceptually embrace that assessing the empirical adequacy of scenarios is important
- Statistical data gathering, handling, or generation, dependent on the phenomenon of interest for a scenario
- Basic understanding of ready-to-use software solutions for BMA, or programming requirements to develop BMA routines
- A clear definition of the primary scenario-phenomena
- Creative and interdisciplinary selection of possible influences on the scenario phenomena

- Adaption and/or adjustment of a quantitative (energy) model if the relevant influences (as evaluated by PIP) are not yet part of the energy model. The necessary workload for adaptations can be weighed against increased representation quality. In any case, the actual representation quality of a given energy model should be communicated via the uncertainty assessment.

The tangible performance edge of Bayesian model averaging for decision-makers (and all recipients of outlooks based on scenario technique) is the possibility of tracing back results to assumptions and communicate the assumptions' statistical confirmation. But we should not take my word for it, so I return to the criteria introduced by Goodwin for a "formal strategy evaluation process within scenario planning" [25] and assess the BMA method for consistent scenario construction in these respects.

The BMA method meets the *transparency* criterion ("the derivation of results can be understood") both on methodological and result levels. The result of the consistent scenario construction is an explicit lower bound of uncertainty and an assessment of the lower bound's legitimacy, in particular for hypothetical assumptions. Data sources, assumption value stipulations, and BMA technical choices leading to the lower bound can be communicated. In my view, BMA is more transparent than expert judgement, though it could be perceived as a "black-box algorithmic procedure" p. 13 [25] by non-statisticians. However, the algorithmic procedures themselves *are* transparent, well-documented, mathematically sound, and understandable with due effort. In contrast, expert judgement remains the virtue of the expert, her experience, knowledge, associations, and her applied heuristics.

The criterion *ease of judgement* ("holistic judgements or decomposed judgements") is met by the uncertainty statement, the identification of relevant influences, and the assessment of the (energy) model's representation quality. The "difficult task of estimating probabilities for states of nature that might prevail in the long term" p.13 is, in contrast to Goodwin and Wright's approach, explicitly *not* avoided. That estimation is highly valuable for recipients basing a decision on scenario results! The estimation of probabilities for states prevailing in the long term is exactly the uncertainty derived from the posterior BMA model probability.

Criterion *versatility* ("evaluate financial and non-financial objectives") is met by the high number of influences the BMA method can accommodate in the form of statistical data of any kind. The *flexibility* criterion ("changes in perspective for different participants of the decision-making process") is met under the condition that the model space is constructed mindful of different perspectives, i.e. including data of various scientific disciplines. In other words, the influences suspected to impact

the dependent variable should account for diverse backgrounds, e.g. social, economic, and technological ones.

Lastly, the *theoretical correctness* criterion defined as "congruency of the strategic options suggested by the method is consistent with the judgments of decision makers" is *not* met, and I am pleased it does not satisfy theoretical correctness defined as such. I entertain the view that it is not the task of decision support to create scenarios anticipating or parroting a decision-maker's judgement nor is it the task of a scenario designer to influence a decision-maker by a strategic choice of phenomena to mirror "consequences" according to experts (and their backgrounds and interests). Including scenario-based information in decisions is sensible with due prudence, *or not at all*, whatever facilitates the decision-makers competency to make a deliberate, independent, and autonomous choice.

## Endnotes

<sup>1</sup>Their five main categories of BMA usage are discussion, model choice, estimation, prediction, revision.

<sup>2</sup>Correspondingly, an assumption propensity to represent empirically adequate personal opinion is subjective confirmation.

<sup>3</sup>For example, the gross domestic product (GDP) of a country is the statistical data representing "an aggregate measure of production, GDP is equal to the sum of the gross value added of all resident institutional units (i.e. industries) engaged in production, plus any taxes, and minus any subsidies, on products not included in the value of their outputs." in the database of Eurostat [40].

<sup>4</sup>Statistical data on stakeholder responses [71] to construction scenarios can in addition account for observable stakeholder interests.

<sup>5</sup>I refrain from discussing the opaque nature of how experts translate their "feeling" into categorical opinion in a given scale, say "+ 2", what is obviously an individual choice and renders a "+ 2" incommensurable with the "+ 2" quantification of another expert. Yet, in the arrangement of scenarios, the "+ 2" judgements are treated equivalently what means that different judgements are expressed in the same number.

<sup>6</sup>For example, [72] defines the industrial production index (IPI) (a kind of ordinal categorical data) with respect to GDP (a kind of quantitative data) as follows: "The IPI is a monthly series that measures output in manufacturing, mining, and electric and gas utilities: Federal Reserve Statistical Release G17. Individual indexes of industrial production are constructed from two types of source data: (1) output measured in physical units and (2) inputs used in the production process (e.g. production worker hours). GDP is a quarterly series that measures the market value of the goods and services produced by labour and property located in the USA.



The aggregate GDP measure that corresponds most closely to the IPI is a GDP for goods measure that consists of durable and nondurable goods within personal consumption expenditures, fixed investment, change in private inventories, and net exports. GDP value production in terms of purchasers' prices, the final prices paid by consumers and by other final demand sectors. The IPI value production in terms of producers' prices paid to manufacturers by wholesalers, by retailers, and, in the case of direct sales, by consumers."

<sup>7</sup>Discussions concerning the effect of flat priors or non-informative priors are provided in particular by [73] and also [74, 75]. However, I do not engage in this discussion here, as I do not interpret the subjective nature of prior specification as disadvantageous in the context of scenario construction.

<sup>8</sup>Providing for example "the total trend-indicator value (TIV) of a country or rebel group's arms imports or exports, broken down by supplier, recipient or type of weapon system." [45]

## Appendix

An exemplary communication for a scenario constructed according to case (b) BMA *scenario construction for an adapted energy model*, could say the following (all numbers and "facts" are fictitious):

### Scenario construction quality—scenario uncertainty

The scenario is constructed based on statistical data of the period 1995–2015 in a yearly time resolution that can be found in the accompanying data bases freely available on [example.ex]. A precise description of the data is available with the cited source. My-Model is a LP with economic optimisation in a yearly time resolution for Nicastan, the scenarios concern the next 20 years. *All scenario statements are conditional on the data used.* That means the reader should add in mind to every sentence the extension "According to the statistical data used."

### Scenarios for unconventional gas production—assessment for My-model

In the historical period considered, the amount of unconventional gas produced was, of all influences suspected to be relevant, most influenced by the natural gas price, the oil price, the conventional natural gas production, the electricity consumption, the LNG infrastructure investment, and gas storage capacity. The best BMA model representing the bearing of these phenomena on the unconventional gas production has a probability of 16% (PMP). Due to the significance of the phenomenon "conventional natural gas production", we adapted the energy model for the representation of the scenario. The impact in the past of "LNG infrastructure" was relevant;

however, we decided not to include that phenomenon in My-model in a trade-off for increased uncertainty by 3% compared to the best performing BMA model. The model can describe the scenarios for unconventional gas with an uncertainty 87%. *Based on the empirical evidence used, it is uncertain by at least 87% that any of the unconventional gas production scenarios presented hereafter is described with My-model.*

### Assessment of the hypothetical scenario assumptions

We consider two scenarios. The "stable increase" scenario represents moderate but stable growth of natural gas production in the next 10 years. The assumption for the gas price is in the first year of projection and is the mean of the statistical data period 1995–2015. We increase the natural gas price linearly until it amounts to half of the maximal natural gas price observed in the historical period by the end of the projection period. We proceed alike for the oil price. The assumption for electricity consumption is moderate growth, expressed numerically by a yearly increase in electricity consumption of 2%, representing a growth rate well within the observations of the historical data. The stipulated value for gas storage capacity is increased once after 5 years, numerically comparable to the additional storage capacity that became available with the launch of the storage facility "South" in Nicastan in 1999. Based on these assumptions, with an uncertainty of at least 87%, the scenario models have an expected annual unconventional gas production per year of 4.47 MMBtu (SD 0.4) in 2015, of 4.53 MMBtu (SD 0.8) in 2016, [...]. The values presented are conditional expected values the standard deviations are indicated in brackets.

All scenario assumptions are within the empirically confirmed possible values so that we would adopt the lower bound of uncertainty for the projection. However, we intentionally do not infer any statement about the actual cause-effect relations in the target system from our analysis. We refrain from an economic or social analysis as to *why* the relations observed in the historical period emerged. The phenomena we include in our analysis are limited (18 influences) and biased towards a supposed relevance for the energy system in Nicastan. To make a meaningful statement about the actual causes of the variables' values observed other system aspects, of which we are ignorant, may prove to be relevant.

The "steep increase and sharp decline" scenario is constructed using *hypothetical* assumptions, Table 3. The values we assume exceed the range of values observed in the historical period in some cases. The value of the unconventional gas production increases in the first 10 years of the projection period steeply, followed by a steep decline towards the 20th year. The peak after 10 years amounts to a production of 12.8 MMBtu in 2025, and the decline settles at the



**Table 3** Scenario presentation example for My-model

My-model input variable	Hypothetical assumption
Natural gas price	Linear increase towards the 10th year. The initial value (2015) is the mean price level of the historical period. The peak price is twice as high as the historical max. value. Linear decrease towards the 20th year to the initial value.
Oil price	Equal to natural gas price.
Conventional natural gas production	Decrease towards the 10th year with a tipping point at a value lower than the historical minimum. Increase towards the 20th year stabilising below the average of the historical period.
Electricity consumption	Flat increase by nearly 0.4% over the projection period.
Storage capacity	Similar to “stable increase scenario”.

lowest value for unconventional gas production in the historical period (0.3 MMBtu). In the My-model, this scenario is represented by the following hypothetical assumptions.

Not all scenario assumptions of the “steep increase and sharp decline” are within the empirically confirmed possible values. We entertain the view that uncertainty is higher than the lower bound of uncertainty for the projection. Both scenarios presuppose a stable bearing of the statistically relevant phenomena on each other (and in compound), as observable in the target system in the past. In the “stable increase” scenario, all values are within the range of empirical possibility of the statistical observation period. That is, we know that these values *can* occur. However, we do not know whether the specific combination of the values assumed in the “stable increase” scenario can occur. If they *do* occur, the probability that the unconventional gas production has the values as represented in the scenario is max. 13%.

Some assumptions of the “steep increase and sharp decline” scenario are hypothetical, i.e. unprecedented in the data record considered. The stability assumption basically amounts to supposing the interaction of phenomena in the target system as observed in the past continues. We assess the interaction solely in terms of increase or decrease (sign) and the strength of the influence (coefficient estimate) on the scenario phenomenon. We are not adopting the lower bound of uncertainty for the steep increase and sharp decline scenario due to a low sign of certainty for the variable “electricity consumption”. This indicates that the phenomenon electricity consumption is relevant and can explain both increase and decrease of the unconventional gas production. The interaction does not seem to be sufficiently stable based on the data employed for the analysis for unprecedented states of the world (hypothetical assumptions). Interactions with other phenomena of the target system we are ignorant of may be involved on whose account the stability assumption may be untenable. We are too sceptical to assume stability for the relation in unprecedented states of the energy system, as modelled in this scenario. We adopt a prudent position and

evaluate the scenario’s uncertainty to be higher than the lower bound of 87%.

Values of all input variables not explicitly discussed in this scenario construction document are set to the last available value statistically confirmed. This represents the situation as observed in 2015 and “locates” the scenarios in the past relative to the publication of this document. Due to data availability restrictions, it is not possible to represent a more recent situation of the Nicastan energy system without loss of representation quality. This implies that all scenarios represent an “alternative” to the actual development in the years up to date if the observable values are not in consonance with the input assumptions.

All results computed with My-model for these two scenarios are uncertain to represent a response of the energy system to the scenario assumptions by at least 87%, where the steep increase and sharp decline scenario is considered not to meet the lower bound.

#### Abbreviations

BMA: Bayesian model averaging; ESS: Energy system subset (subset of WSS; statistical data representing states of the energy system); GDP: Gross domestic product; PIP: Posterior inclusion probability; PMP: Posterior model probability; WSS: World subset (statistical data representing states in the world)

#### Acknowledgements

The author acknowledges the support of Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of Karlsruhe Institute of Technology. I thank the anonymous reviewers and Ms. Fiedler in particular for their suggestions and criticisms that substantially improved the paper.

#### Funding

This work was supported by NICA the New Interdisciplinary Collaboration Association. This work was developed with the support of the Institute of Philosophy and the ITAS of Karlsruhe Institute of Technology (KIT) [Helmholtz School on Energy Scenarios]. Publication charges were kindly covered by KIT open access fund.

#### Availability of data and materials

Examples presented are used to illustrate the methodology and are not based on explicit datasets.

#### Authors’ contributions

The author read and approved the final manuscript.

#### Consent for publication

Not applicable. No personal data used.

**Competing interests**

The author declares that she has no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 December 2017 Accepted: 6 June 2018

Published online: 09 July 2018

**References**

- Culka M (2016) Uncertainty analysis using Bayesian model averaging: a case study of input variables to energy models and inference to associated uncertainties of energy scenarios. *Energy Sustain Soc* 6(1):21. <https://doi.org/10.1186/s13705-016-0073-0>
- Hoeting JA, Madigan D, Raftery AE et al (1999) Bayesian model averaging: a tutorial. *Stat Sci* 14(4):382–417
- Raftery AE, Madigan D, Hoeting JA (1997) Bayesian model averaging for linear regression models. *J Am Stat Assoc* 92(437):179–191
- Wasserman (2000) Bayesian model selection and model averaging. *J Math Psychol* 44(1):92–107. <https://doi.org/10.1006/jmps.1999.1278>
- Volinsky CT, Madigan D, Raftery AE et al (1997) Bayesian model averaging in proportional hazard models: assessing the risk of a stroke. *J R Stat Soc: Ser C: Appl Stat* 46(4):433–448
- Merlise A (1999) Bayesian model averaging and model search strategies. *Bayesian Stat* 6:157–185
- Wintle BA, McCarthy MA, Volinsky CT et al (2003) The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conserv Biol* 17(6):1579–1590
- Madigan D, Andersson SA, Perlman MD et al (1996) Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Commun Stat Theory Methods* 25(11):2493–2519
- Raftery AE, Gneiting T, Balabdaoui F et al (2005) Using Bayesian model averaging to calibrate forecast ensembles. *Mon Wea Rev* 133(5):1155–1174. <https://doi.org/10.1175/MWR2906.1>
- Fragoso TM, Bertoli W, Louzada F Bayesian model averaging: a systematic review and conceptual classification. *Int Stat Rev*. <https://doi.org/10.1111/insr.12243>
- Gass SI, United States. National Bureau of Standards, United States. Energy Information Administration et al. (1980) Validation and assessment issues on energy models: proceedings of a workshop held at the National Bureau of Standards, Gaithersburg, Maryland, January 10–11, 1979. NBS special publication. Dept. of Commerce, National Bureau of Standards : for sale by the Supt. of Docs., U.S. Government. Print. Off
- Munasinghe M, Meier P (1993) Energy policy analysis and modeling. Cambridge studies in energy and the environment. Cambridge University Press, Cambridge [England], New York, NY, USA
- van Notten P (op. 2005) Writing on the wall: scenario development in times of discontinuity. Universal-Publishers, Boca Rotan
- Lindgren M, Bandhold H (2002) Scenario planning: the link between future and strategy. Palgrave Macmillan, Basingstoke
- Rothman DS (2009) Environmental Scenarios. A Survey. in *Environmental Futures The Practice of Environmental Scenario Analysis*, Edited by Joseph Alcamo, Volume 2, Chapter 3, ISBN: 978-0-444-53293-0, Elsevier Science 2009
- Önkal D, Sayim KZ, Gönül MS (2013) Scenarios as channels of forecast advice. *Technol Forecast Soc Chang* 80(4):772–788. <https://doi.org/10.1016/j.techfore.2012.08.015>
- Schrattenholzer L (1981) The energy supply model MESSAGE. RR/ International Institute for Applied Systems Analysis. International Institute for Applied Systems Analysis, Laxenburg, pp 81–31. Austria onlinepure.iiasa.ac.at/1542/1/RR-81-031.pdf
- Loulou R, Remme U, Kanudia A et al. (2005) Documentation for the TIMES model: PART I
- Horridge M, Parmenter BR, Pearson KR (2000) ORANI-G: a general equilibrium model of the Australian economy. Centre of Policy Studies, online [www.academia.edu/download/44224286/ORANI-G\\_A\\_general\\_equilibrium\\_model\\_of\\_t20160330-18629-1jg0ngf.pdf](http://www.academia.edu/download/44224286/ORANI-G_A_general_equilibrium_model_of_t20160330-18629-1jg0ngf.pdf)
- Bhattacharyya SC, Timilsina GR (2010) A review of energy system models. *Int J Energy Sect Manag* 4(4):494–518. <https://doi.org/10.1108/17506221011092742>
- Jebaraj S, Iniyar S (2006) A review of energy models. *Renew Sust Energ Rev* 10(4):281–311. <https://doi.org/10.1016/j.rser.2004.09.004>
- Pilavachi PA, Dalamaga T, Rossetti di Valdalbero D et al (2008) Ex-post evaluation of European energy models. *Energy Policy* 36(5):1726–1735. <https://doi.org/10.1016/j.enpol.2008.01.028>
- International Energy Agency (2011) World energy outlook 2011. International Energy Agency, Paris, p s.I
- Culka M (2014) Applying Bayesian model averaging for uncertainty estimation of input data in energy modelling. *Energy Sustain Soc* 4(1):21. <https://doi.org/10.1186/s13705-014-0021-9>
- Goodwin P, Wright G (2001) Enhancing strategy evaluation in scenario planning: a role for decision analysis. *J Manag Stud* 38(1):1–16. <https://doi.org/10.1111/1467-6486.00225>
- Kosow H, Gaßner R (2008) Methods of future and scenario analysis: overview, assessment, and selection criteria. *Studies/Deutsches Institut für Entwicklungspolitik*, vol 39. Dt. Inst. für Entwicklungspolitik, Bonn
- van Notten PWF, Rotmans J, van Asselt MBA et al (2003) An updated scenario typology. *Futures* 35(5):423–443. [https://doi.org/10.1016/S0016-3287\(02\)00090-3](https://doi.org/10.1016/S0016-3287(02)00090-3)
- Bradfield R, Wright G, Burt G et al (2005) The origins and evolution of scenario techniques in long range business planning. *Futures* 37(8):795–812
- Börjeson L, Höjer M, Dreborg K-H et al (2006) Scenario types and techniques: towards a user's guide. *Futures* 38(7):723–739
- International Energy Agency (2016) Scenarios and projections: world energy outlook. <http://www.iea.org/publications/scenariosandprojections/>. Accessed 14 June 2018
- Bishop P, Hines A, Collins T (2007) The current state of scenario development: an overview of techniques. *Foresight* 9(1):5–25. <https://doi.org/10.1108/14636680710727516>
- Woudenberg F (1991) An evaluation of Delphi. *Technol Forecast Soc Chang* 40(2):131–150
- Bardecki MJ (1984) Participants' response to the Delphi method: an attitudinal perspective. *Technol Forecast Soc Chang* 25(3):281–292. [https://doi.org/10.1016/0040-1625\(84\)90006-4](https://doi.org/10.1016/0040-1625(84)90006-4)
- Bolger F, Stranieri A, Wright G et al (2011) Does the Delphi process lead to increased accuracy in group-based judgmental forecasts or does it simply induce consensus amongst judgmental forecasters? *Technol Forecast Soc Chang* 78(9):1671–1680. <https://doi.org/10.1016/j.techfore.2011.06.002>
- Häder M (2000) Die Delphi-Technik in den Sozialwissenschaften: Methodische Forschungen und innovative Anwendungen. ZUMA-Publikationen. Westdt. Verl., Wiesbaden
- Stewart TR (1987) The Delphi technique and judgmental forecasting. *Clim Chang* 11(1–2):97–113. <https://doi.org/10.1007/BF00138797>
- Weimer-Jehle W (2006) Cross-impact balances: a system-theoretical approach to cross-impact analysis. *Technol Forecast Soc Chang* 73(4):334–361. <https://doi.org/10.1016/j.techfore.2005.06.005>
- Enzer S (1980) INTERAX—an interactive model for studying future business environments: part I. *Technol Forecast Soc Chang* 17(2):141–159. [https://doi.org/10.1016/0040-1625\(80\)90049-9](https://doi.org/10.1016/0040-1625(80)90049-9)
- Amer M, Daim TU, Jetter A (2013) A review of scenario planning. *Futures* 46: 23–40. <https://doi.org/10.1016/j.futures.2012.10.003>
- EUROSTAT (2017) Glossary:Gross domestic product (GDP) - Statistics Explained. <http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:GDP>. Accessed 14 June 2018
- Barrett S, Dannenberg A (2012) Climate negotiations under scientific uncertainty. *Proc Natl Acad Sci U S A* 109(43):17372–17376. <https://doi.org/10.1073/pnas.1208417109>
- Walker WE, Harremoes P, Rotmans J et al (2005) Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integr Assess* 4(1)
- Zeugner S (2014) R package BMS—Bayesian model averaging. <http://bms.zeugner.eu/>. Accessed 14 June 2018
- Fleurant A-EF, Perlo-Freeman S (2016) SIPRI Military Expenditure Database—[www.sipri.org/research/armaments/milex/milex\\_database](http://www.sipri.org/research/armaments/milex/milex_database). Accessed 14 June 2018
- Stockholm International Peace Research Institute (2016) SIPRI Arms Transfers Database — [www.sipri.org/databases/armstransfers](http://www.sipri.org/databases/armstransfers). Accessed 14 June 2018
- Peace Research Institute Oslo PRIO (2016) Data - PRIO. <https://www.prio.org/Data/>. Accessed 14 June 2018
- Maoz Z (2016) About the Correlates of War Project—Corelates of War. <http://www.correlatesofwar.org/>. Accessed 14 June 2018
- Transparency International e.V Research - CPI - Overview. <http://www.transparency.org/research/cpi/overview>. Accessed 14 June 2018

49. World Bank Group (2016) CPIA transparency, accountability, and corruption in the public sector rating. <http://data.worldbank.org/indicator/IQ.CPA.TRAN.XQ>. Accessed 14 June 2018
50. UN-Register (2016) UN-Register: The Global Reported Arms Trade. <http://www.un-register.org/SmallArms/Index.aspx>. Accessed 14 June 2018
51. World Values Survey (2016) WVS Database. <http://www.worldvaluessurvey.org/WVSContents.jsp>. Accessed 14 June 2018
52. OECD Social and welfare issues, OECD: a leader in international measurement and analysis in social policy. <http://www.oecd.org/social/statistics.htm>. Accessed 14 June 2018
53. United Nations Statistics Division UNSD//United Nations Statistics Division - Demographic and Social Statistics. <https://unstats.un.org/unsd/demographic-social/index.cshhtml>. Accessed 14 June 2018
54. Emory University (2016) Social Indicators|Electronic Data Center: Electronic Data Center, Robert W. Woodruff Library. <http://pid.emory.edu/ckhkf>. Accessed 14 June 2018
55. Kahn H (1922-1983) Thinking about the unthinkable. Horizon Press, New York, 1962 (OCOLC)567986632
56. Amanatidou E, Müller AW, Shwarz JO (2016) Assessing the functions and dimensions of visualizations in foresight. *Foresight* 18(1):76–90. <https://doi.org/10.1108/FS-04-2014-0027>
57. Goodwin P, Wright G (2010) The limits of forecasting methods in anticipating rare events. *Technol Forecast Soc Chang* 77(3):355–368
58. Coates JF (2000) Scenario planning. *Technol Forecast Soc Chang* 65(1):115–123
59. European Commission - Joint Research Centre - Institute for Environment and Sustainability (2010) International reference life cycle data system (ILCD) handbook: framework and requirements for life cycle impact assessment models and indicators. EUR (Luxembourg), vol 24586. Publications Office, Luxembourg
60. Hawken P, Ogiwly JA, Schwartz P (1982) Seven tomorrows: toward a voluntary history. Bantam Books, Toronto. ISBN: 055301367X, 9780553013672
61. Gölpinar N, Rustem B, Settergren R (2004) Simulation and optimization approaches to scenario tree generation. *J Econ Dyn Control* 28(7):1291–1315. [https://doi.org/10.1016/S0165-1889\(03\)00113-1](https://doi.org/10.1016/S0165-1889(03)00113-1)
62. Schwartz P (1996) The art of the long view: paths to strategic insight for yourself and your company. Currency Doubleday, New York
63. Harman WW (1979) An incomplete guide to the future. Norton, New York. <https://archive.org/details/incompleteguidet00will>
64. Quist J, Vergragt P (2006) Past and future of backcasting: the shift to stakeholder participation and a proposal for a methodological framework. *Futures* 38(9):1027–1045. <https://doi.org/10.1016/j.futures.2006.02.010>
65. Miola A (2008) Backcasting approach for sustainable mobility. EUR. Scientific and technical research series, vol 23387. Publications Office, Luxembourg
66. Mason D (2003) Tailoring scenario planning to the company culture. *Strateg Leadersh* 31(2):25–28
67. Coyle RG (2003) Morphological forecasting—field anomaly relaxation (FAR). *Futur Res Methodol* 2
68. Gordon TJ, Becker HS, Gerjuoy H (1974) Trend impact analysis: a new forecasting tool. Futures Group, Glastonbury, Conn
69. Saltelli A (2004) Sensitivity analysis practice: a guide to assessing scientific models. Wiley, Hoboken
70. Savage AE, Ward E (1998) Dynamic scenarios: systems thinking meets scenario planning. Learning from the future: competitive foresight scenarios. Wiley, New York
71. Bundesnetzagentur (2015) Stellungnahmen zur Konsultation - Statistik und inhaltliche Auswertung: Netzentwicklungspläne 2024 und Umweltbericht. [https://www.bundesnetzagentur.de/cln\\_112/DE/Home/home\\_node.html](https://www.bundesnetzagentur.de/cln_112/DE/Home/home_node.html). Accessed 14 June 2018
72. U. S. Department of Commerce (2005) Bureau of Economic Analysis: frequently asked questions. [http://www.bea.gov/faq/index.cfm?faq\\_id=73](http://www.bea.gov/faq/index.cfm?faq_id=73). Accessed 14 June 2018
73. Fernández C, Ley E, Steel MFJ (2001) Benchmark priors for Bayesian model averaging. *J Econ* 100(2):381–427. [https://doi.org/10.1016/S0304-4076\(00\)00076-2](https://doi.org/10.1016/S0304-4076(00)00076-2)
74. Gelman A, Jakulin A, Pittau MG et al (2008) A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat* 2(4):1360–1383. <https://doi.org/10.1214/08-AOAS191>
75. Kass RE, Wasserman L (1996) The selection of prior distributions by formal rules. *J Am Stat Assoc* 91(435):1343–1370. <https://doi.org/10.1080/01621459.1996.10477003>

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)