Integrated strategic energy mix and energy generation planning with multiple sustainability criteria and hierarchical stakeholders

Chandra Ade Irawan[a,b,*], Dylan Jones[c], Peter S. Hofman[a,d], Lina Zhang[a,e]

[a] *Nottingham University Business School China, University of Nottingham Ningbo China*
[b] *Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute*
[c] *School of Mathematics and Physics, Centre for Operational Research and Logistics, University of Portsmouth, UK*
[d] *James Cook University, Singapore*
[e] *Centre for Simulation, Analytics & Modelling, University of Exeter Business School, University of Exeter, UK*

Abstract

This paper proposes a combination of two optimization models for simultaneously determining strategic energy planning at both national and regional levels. The first model deals with a single-period energy mix where the electricity production configuration at a future date (e.g., 2050), based on the available generation sources, is optimally obtained. An optimization model, based on a non-linear goal programming method, is designed to ensure a mixed balance between national and regional goals. The desired energy mix configuration, which is the solution obtained by solving the first model, is then fed into the second model as the main data input. In the second model, a multiple-period generation expansion plan is designed which optimizes the energy transition over the time horizon from the present until the future planning date (2050). The model considers uncertain parameters, including the regional energy demand, fuel cost, and national peak load. A two-stage stochastic programming model is developed where the sample average approximation approach is used as a method of solution. The practical use of the proposed models has been assessed through application to the electricity generation system in China.

Keywords: Multiple objective programming; Energy planning; Goal programming; Two-stage stochastic programming

1. Introduction

Global total electricity consumption has continuously grown and increased from 10,897.94 TWh in 1990 to 24,738.92 TWh in 2018, and it is expected that world energy consumption will continue to rise by nearly 50% in 2050 (Energy Information Administration, 2020). Electricity can be generated from several generation sources, including fossil fuel-based, nuclear, and renewable sources. There are advantages and disadvantages when using each source of power, which can be measured over a multiple set of economic, environmental, social, and technical criteria. The fossil fuel-based sources (i.e., coal,

* Corresponding author
Email address: chandra.irawan@nottingham.edu.cn (Chandra Ade Irawan)

petroleum, and natural gas) generate a large quantity of relatively cheap energy at the expense of a large amount of greenhouse gas (GHG) emissions. Nuclear generation is a good option for lowering the GHG emissions, however it has low public acceptance due to significant security concerns (Lugovoy et al., 2021). Renewable energy sources (i.e., hydro, solar, wind, and geothermal) produce clean energy without GHG emissions, but their cost and feasibility are highly dependent on the potential of different regions and their prevailing weather conditions (Thangavelu et al., 2015).

The concept of Generation Expansion Planning (GEP) represents a type of strategic energy approach that determines optimal generation sources, their capacity and location, together with time construction/decommissioning of new/existing power plants in order to minimize total cost over a long-term planning horizon, while satisfying high demand growth for electricity at national and regional levels. A series of constraints can also be considered, including the penetration of renewable energy, the GHG (CO_2) emissions produced and energy mix targets. This means that a wide range of aspects, including economic, environmental, regulatory, technical, operational and social, can be included in the GEP model. The GEP has been widely investigated in the literature, and for more details we refer to survey papers such as Koltsaklis and Dagoumas (2018), Dagoumas and Koltsaklis (2019), and Sadeghi et al. (2017).

In this paper, we propose an integration of two optimization models for determining generation expansion planning at both national and regional levels. In the first proposed model, a single-period energy mix at a future date (e.g., 2050) is obtained using a non-linear goal programming method where a balance between national and regional decisions is also considered. The use of goal programming for solving the GEP problem is still limited, although this method has the ability to address multiple goals across conflicting criteria. There are several related works in energy planning where goal programming is used. San Cristóbal (2012) investigates the GEP problem for the renewable energy plants. A goal programming model is proposed to obtain the optimal mix of different power plant types over the planning horizon. Moreover, the power plant locations, together with their capacity expansion decisions, are optimally determined. The application of the proposed model is assessed by locating five renewable energy plants in five places located in Cantabria, Spain. Jones et al. (2016) propose a relevant extension to the classic goal programming model called extended network goal programming (ENGP). This method aims to ensure that a parametric mix of optimization and balance (efficiency and equity) is achieved across a hierarchical decision network. A model is designed which gives consideration to the balance and efficiency of multiple objectives and stakeholders at each network node level. The methodology is evaluated on a small example concerning the regional development of renewable energy sources, where the model aims to find an optimal number of renewable energy projects. Özcan et al. (2017) study a combination of goal programming with an Analytical Hierarchy Process (AHP) and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) for maintenance strategy selection in hydroelectric power plants.

In the second model, we propose a stochastic multi-period GEP model that determines the energy transition over the time horizon from the present until the end of the horizon period (2050). The energy mix configuration at the future period (2050) is obtained from the first model, which can be considered an energy mix target at the end of the planning horizon. The proposed model takes into account uncertain parameters, including the regional energy demand, fuel cost, and national peak load. The GEP under uncertainty has been reported in the literature where Malcolm and Zenios (1994) and Mulvey et al. (1995) are among the first to deal with this problem. They put forward robust optimization models to attain capacity expansion plans in the presence of uncertain energy demand. Robust optimization frameworks are also proposed for the GEP under uncertainty by recent researchers including Costa et al. (2017), Moret et al. (2020), and Moreira et al. (2021).

Two-stage and multi-stage stochastic programming models are also used to address the GEP under uncertainty. It is common that uncertain parameters are tackled by the scenario tree configuration or Monte Carlo simulation. The two-stage stochastic programming model is studied by, among others, Krukanont and Tezuka (2007), Feng and Ryan (2013), Park and Baldick (2015), Irawan et al. (2022), and Kim et al. (2021). The multi-stage stochastic programming can be considered the most popular tool to solve the GEP under uncertainty. Recent researchers who use this approach include Li et al. (2010), Li and Huang (2012), Min and Chung (2013), Thangavelu et al. (2015), Betancourt-Torcat and Almansoori (2015) and Ioannou et al. (2019). A fuzzy stochastic programming has also been implemented for the GEP under uncertainty, including the works of Hu et al. (2014) and Li et al. (2014).

In this paper, we design a new methodology that integrates two optimization models for the generation expansion planning at both national and regional levels. The first model is an ENGP model where multiple objective functions are considered based on economic, environmental, and social dimensions. We also take into account the target set by the stakeholders (national and regional level) for each objective function. A Non-Linear Programming (NLP) model is built and a commercial non-linear solver (Baron) are used to solve the model. The main output of the first model is the energy mix at both national and regional levels for a single future date (2050). This energy mix is then used in the second model as a target that needs to be achieved at that future point. The second model can be considered as a stochastic GEP where a two-stage stochastic programme is designed. Moreover, we also develop an algorithm based on the sample average approximation method (Kleywegt et al., 2002) to solve the model. To the best of our knowledge, our methodology is new and has not been used in the current literature. The proposed methodology is assessed on a realistic instance based on power system data in China.

This paper makes novel contributions to both the application and methodology of operational research. Firstly, in terms of application it tackles the important problem of determining targets for China's electricity generation and developing an energy transition plan to meet those targets in the period until 2050. This is of high strategic importance, not just to China but on a worldwide scale due the economic and environmental implications, as China is set to become the World's largest economy by around 2030, with a consequent large electricity generation requirement (Energy Information Administration, 2022). Environmentally, China is currently the world's largest consumer of energy, the largest producer and consumer of coal, and the largest emitter of carbon dioxide. Hence, the need to provide more accurate forecasting and strategies to ensure economically and environmentally efficient future Chinese energy generation is the prime motivation for this paper. Compared to other papers in literature, our paper has the novel methodological aspects of (i) developing a model that can accurately preferentially achieve both national and regional energy generation goals as closely as possible, (ii) developing a stochastic energy transition model that is robust under multiple, uncertain future policy scenarios and (iii) developing an overall methodology that combines the setting of national and regional targets with an energy transition plan to meet those targets. To the best of our knowledge, these aspects have not been investigated comprehensively in the extant literature and hence we believe that our paper is methodologically novel and original with respect to the above three points. To summarize, the principal novel contributions of this work are as follows:

- a novel methodology that combines two optimization models for simultaneous determination of the strategic energy planning at both national and regional levels
- a non-linear extended network goal programming model to obtain an energy mix at a future date
- a stochastic energy transition model and its solution method based on sample average approximation
- the construction and analysis of the results for the electricity generation system in China.

The paper is organized into five main sections. In Section 2, we present the novel optimization models that can be used for the generation expansion planning at both regional and national levels. Section 3 describes the energy data in China together with the instances that we use. In Section 4, results of computational experiments are presented together with their discussion. The final section provides conclusions, limitations and suggestions for future work.

2. Optimization Models

The proposed methodology comprises two optimization models where the first model aims to obtain the energy mix for the national level together with its regions in a future single period (e.g. 2050). The model is developed based on the extended goal programming network, and formulated based on the Mixed Integer Nonlinear Programming (MINLP). The second model is then designed and solved to generate an energy expansion planning (GEP) starting from the present period to the future period (multi-period) in order to achieve the energy mix for the future period obtained by the first model. The second model can represent the energy transition planning to achieve the energy mix that has been set. The stochastic model is proposed based on the Mixed Integer Linear Programming (MILP), where uncertain parameters are taken into account, including the fuel cost, annual energy demand for each region, and the national peak load. To solve the proposed stochastic model, the Sample Average Approximation (SAA) method is put forward. Figure 1 shows the flowchart of the proposed methodology used in this paper, while the detailed description of each model is provided in the following subsections.

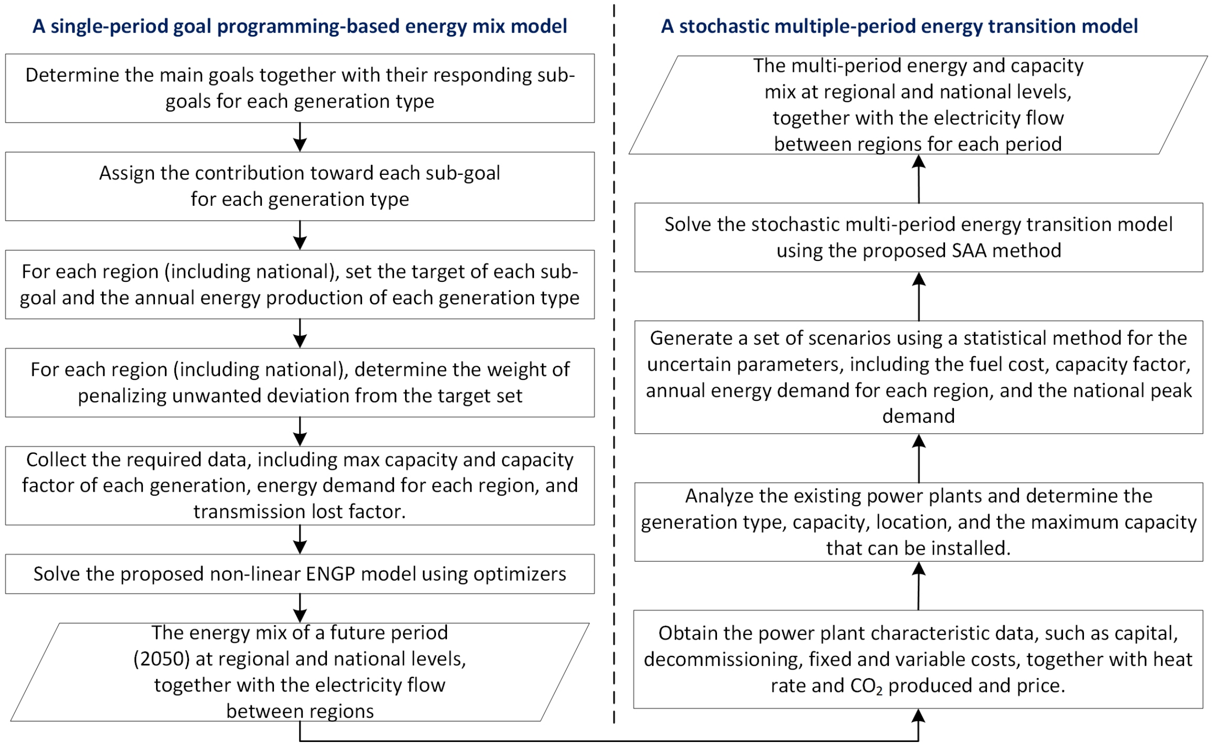

Figure 1: The flowchart of the proposed methodology

2.1. A single-period goal programming-based energy mix model

The single-period energy mix is designed based on a non-linear ENGP methodology which is developed to ensure that balance and optimization are achieved across a hierarchical decision network. The model has two levels, namely national and regional, where a country (national level) may consist of several regions (m regions). The country and each region are treated as stakeholders that have their own preferential data with respect to a set of sustainability objectives. This implies a decision-making network comprising multiple layers (levels) with multiple objectives and multiple stakeholders. The proposed model aims to optimize deployment of the diversity of energy sources in such way that the energy demand at national and regional levels is satisfied when considering economic, environmental, and social sustainability aspects.

The economic, environmental, and social aspects are considered goals, where each goal comprises several sub-goals. In this paper, the sub-goals are taken from Streimikiene et al. (2012) who has developed evaluation criteria based upon an in-depth review of the literature around energy technology evaluation criteria and major international research projects that focused on evaluation of energy technology options and had participation from leading energy research groups that developed the criteria by considering electricity generation technology options and their impacts in a range of European countries and China. It is therefore very well suited for the purpose of this research.

Table 1 present the sub-goals (set S) which is divided into two subsets, namely the subset of sub-goals with underlying maximization functions ($S^+ \subset S$), and those with underlying minimization functions ($S^- \subset S$). A set of generation sources (K) is considered, including coal, gas, hydro, nuclear, biomass, wind, and solar. Each generation source $k \in K$ has a different assessment (a_{ks}) on sub-goal $s \in S$. A target of sub-goal $s \in S$ (τ^S_{rs}) needs to be set for each stakeholder r. Here, $r = 0$ indicates the national level, whereas $r = 1, \ldots, m$ represents regions, from region $r = 1$ to $r = m$. Penalties (u^S_{rs} and v^S_{rs}) are given for each unit of negative deviation from the target of the sub-goal $s \in S^+$ and each unit of positive deviation from the target of sub-goal $s \in S^-$ respectively.

Table 1: Main goals and their sub-goals

Main Goal	Sub-goals	Abbreviation	Function
Economic dimension	Private costs (investments and operation costs)	PR-COST	Minimize
	Average availability (load) factor	AVAILAB	Maximize
	Security of supply	SECURE	Maximize
	Cost of grid connection	GRID-COST	Minimize
	Peak load response	PEAKLOAD	Maximize
Environmental dimension	GHG emissions	CO2eq.	Minimize
	Environmental external costs	ENV	Minimize
	Radionuclide external costs	RADIO	Minimize
	Human health impact	HEALTH	Minimize
Social dimension	Technology-specific job opportunities	EMPL	Maximize
	Food safety risk	FOOD	Minimize
	Fatal accidents from past experience	ACC-PAST	Minimize
	Severe accidents perceived in future	ACC-FUT	Minimize

Each region r has a set of generation sources (K) to satisfy its energy demand (d_r). In the event that the total energy production of region r ($\sum_k E_{rk}$) is not sufficient, the demand can be met by importing energy ($G_{rr'}$) from other regions that has surplus production (L_r), considering a transmission lost factor ($l_{rr'}$). Note that the surplus region does not receive any transmission from other regions. At both regional and national level an annual energy production target (MWh) is set from its generation source $k \in K$ (τ^E_{rk}). Negative deviation from the target is penalized (u^E_{rk}). The energy production of the generation source $k \in K$ is affected by its capacity factor (ψ_k), and the maximum potential capacity (κ_{rk}) available in each region r. A diagrammatic illustration of the proposed energy mix model is presented in Figure 2.

Each stakeholder (national and regional) may have a different perspective regarding each objective. Here, the proposed model takes into account balance and efficiency among objectives and stakeholders. Three parameters are used where the first is a relative importance national-regional weight (w), where $0 \leq w \leq 1$. A larger w indicates a greater degree of decision-making at the national level. Secondly, Parameter α_r controls the mix of optimization and balance of stakeholder $r, r = 0, 1, \ldots, m$, where $0 \leq \alpha_r \leq 1$. When $\alpha_r = 0$, stakeholder r focuses on the efficiency of the objectives, otherwise, if $\alpha_r = 1$ the balance of the objectives is considered. The third parameter is β ($0 \leq \beta \leq 1$) which controls the mix of optimization and balance when considering a set of regions. Increasing the β value increases the importance given to the maximal stakeholder dissatisfaction, whereas decreasing it increases the

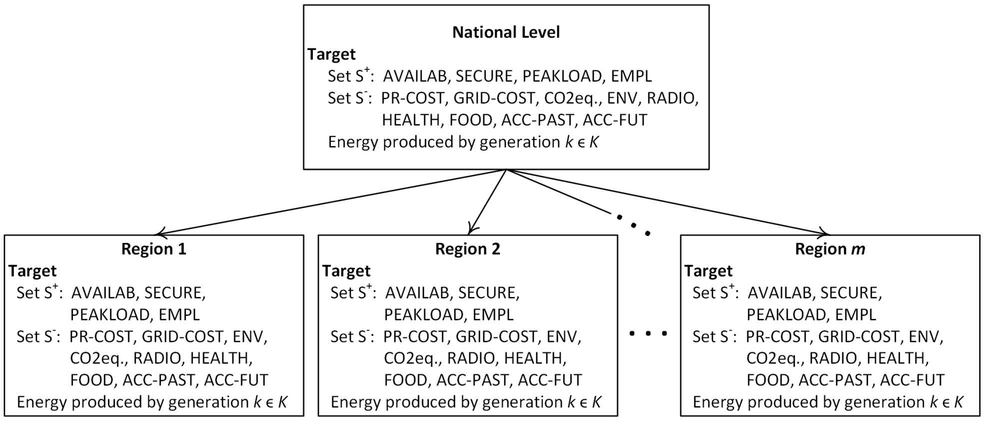

Figure 2: Diagrammatic illustration of the energy mix model

importance given to the average stakeholder dissatisfaction. In summary, the following notations are used to describe the sets and parameters of the non-linear ENGP energy mix model.

Sets and indices

r: index of stakeholders where $r = 0$ is for national, and $r = 1, \ldots, m$ for regions 1 to m
K: set of electricity generation sources indexed by k (coal, gas, wind, etc.)
S: set of sub-goals (i.e., installation cost, operational cost, dispatch ability, energy generated, etc.)
S^+: set of sub-goals with underlying maximization functions ($S^+ \subset S$)
S^-: set of sub-goals with underlying minimization functions ($S^- \subset S$)

Parameters

a_{ks}: the contribution towards sub-goal $s \in S$ of generation source $k \in K$
τ^S_{rs}: the target of sub-goal $s \in S$ set by stakeholder $r = 0, 1, \ldots, m$
τ^E_{rk}: the annual energy production target (MWh) from generation source $k \in K$ set by stakeholder $r = 0, 1, \ldots, m$
d_r: the annual energy demand of region $r = 1, \ldots, m$
ψ_k: capacity factor (%) of generation source $k \in K$ representing the ratio of its actual output over a period of time to its potential output which is calculated based on the installed capacity
κ_{rk}: maximum potential capacity (MW) of generation source $k \in K$ for stakeholder $r = 0, 1, \ldots, m$
ϕ: total operating hours (it is assumed the same for all power plants, i.e., 8,760 hours per year)
$l_{rr'}$: the transmission loss factor from region r to r', where $(r, r') = 1, \ldots, m$ and $r \neq r'$
u^S_{rs}: weight associated with penalizing negative deviation from the target of sub-goal $s \in S^+$ for stakeholder $r = 0, 1, \ldots, m$
v^S_{rs}: weight associated with penalizing positive deviation from the target of sub-goal $s \in S^-$ for stakeholder $r = 0, 1, \ldots, m$
u^E_{rk}: weight associated with penalizing negative deviation from the annual energy production target from generation source $k \in K$ for stakeholder $r = 0, 1, \ldots, m$
w: controls national-regional weighting
α_r: controls mix of optimization and balance in stakeholder $r = 0, 1, \ldots, m$
β: controls mix of optimization and balance when considering the set of regions

Decision Variables

E_{rk}: amount of energy produced (MWh) by generation source $k \in K$ in stakeholder $r = 0, 1, \ldots, m$
N_{rs}^S: negative deviation from the target of sub-goal $s \in S$ for stakeholder r
P_{rs}^S: positive deviation from the target of sub-goal $s \in S$ for stakeholder r
N_{rk}^E: negative deviation from the annual energy target of generation source $k \in K$ for stakeholder $r = 0, 1, \ldots, m$
P_{rk}^E: positive deviation from the annual energy target of generation source $k \in K$ for stakeholder $r = 0, 1, \ldots, m$
$G_{rr'}$: electricity transmitted (MWh) from region r to r', where $(r, r') = 1, \ldots, m$ and $r \neq r'$
$L_r = \begin{cases} 1 & \text{if region } r \text{ has surplus production, where } r = 1, \ldots, m, \\ 0 & \text{otherwise} \end{cases}$
λ_r: maximal deviation from set of normalized weighted goals for stakeholder $r = 0, 1, \ldots, m$
D: maximal measure from amongst the set of regions (the worst performing region)

Mathematical Model

The energy mix problem can be formulated as the mixed integer non-linear programming extended network goal programme (MINLENGP), as follows:

$$\min Z = w\left[\alpha_0 \cdot \lambda_0 + (1-\alpha_0)Z^0\right] + (1-w)\left[\beta \cdot D + (1-\beta)Z^R\right] \tag{1}$$

Subject to

$$Z^0 = \sum_{s \in S^+} \frac{u_{0s}^S \cdot N_{0s}^S}{\tau_{0s}^S \sum_{k \in K} E_{0k}} + \sum_{s \in S^-} \frac{v_{0s}^S \cdot P_{0s}^S}{\tau_{0s}^S \sum_{k \in K} E_{0k}} + \sum_{k \in K} \frac{u_{0k}^E \cdot N_{0k}^E}{\tau_{0k}^E} \tag{2}$$

$$Z^R = \sum_{r=1}^{m}\left[\alpha_r \cdot \lambda_r + (1-\alpha_r)\left(\sum_{s \in S^+} \frac{u_{rs}^S \cdot N_{rs}^S}{\tau_{rs}^S \sum_{k \in K} E_{rk}} + \sum_{s \in S^-} \frac{v_{rs}^S \cdot P_{rs}^S}{\tau_{rs}^S \sum_{k \in K} E_{rk}} + \sum_{k \in K} \frac{u_{rk}^E \cdot N_{rk}^E}{\tau_{rk}^E}\right)\right] \tag{3}$$

$$\sum_{k \in K}(a_{ks} \cdot E_{rk}) + N_{rs}^S - P_{rs}^S = \tau_{rs}^S \sum_{k \in K} E_{rk}, \qquad \forall s \in S, r = 0, 1, \ldots, m \tag{4}$$

$$E_{rk} + N_{rk}^E - P_{rk}^E = \tau_{rk}^E, \qquad \forall k \in K, r = 0, 1, \ldots, m \tag{5}$$

$$\sum_{k \in K}\sum_{r=1}^{m} E_{rk} \leq \sum_{k \in K} \tau_{0k}^E \tag{6}$$

$$E_{rk} \leq \kappa_{rk} \cdot \psi_k \cdot \phi, \qquad \forall k \in K, r = 1, \ldots, m \tag{7}$$

$$\sum_{k \in K} E_{rk} - \sum_{\substack{r'=1 \\ r' \neq r}}^{m} G_{rr'} + \sum_{\substack{r'=1 \\ r' \neq r}}^{m} G_{r'r}(1 - l_{r'r}) \geq d_r, \qquad r = 1, \ldots, m \tag{8}$$

$$\sum_{k \in K} E_{rk} \geq d_r \cdot L_r, \qquad r = 1, \ldots, m \tag{9}$$

$$\sum_{\substack{r'=1 \\ r' \neq r}}^{m} G_{rr'} \leq L_r \cdot \sum_{k \in K} E_{rk}, \qquad r = 1, \ldots, m \tag{10}$$

$$\sum_{\substack{r'=1\\ r'\neq r}}^{m} G_{r'r} \le d_r \cdot (1 - L_r), \qquad r = 1, \ldots, m \tag{11}$$

$$\frac{u_{rs}^S \cdot N_{rs}^S}{\tau_{rs}^S \sum_{k\in K} E_{rk}} \le \lambda_r, \qquad \forall s \in S^+, r = 0, 1, \ldots, m \tag{12}$$

$$\frac{v_{rs}^S \cdot P_{rs}^S}{\tau_{rs}^S \sum_{k\in K} E_{rk}} \le \lambda_r, \qquad \forall s \in S^-, r = 0, 1, \ldots, m \tag{13}$$

$$\sum_{k\in K} \frac{u_{rk}^E \cdot N_{rk}^E}{\tau_{rk}^E} \le \lambda_r, \qquad \forall r = 0, 1, \ldots, m \tag{14}$$

$$\alpha_r \cdot \lambda_r + (1-\alpha_r)\left(\sum_{s\in S^+} \frac{u_{rs}^S \cdot N_{rs}^S}{\tau_{rs}^S \sum_{k\in K} E_{rk}} + \sum_{s\in S^-} \frac{v_{rs}^S \cdot P_{rs}^S}{\tau_{rs}^S \sum_{k\in K} E_{rk}} + \sum_{k\in K} \frac{u_{rk}^E \cdot N_{rk}^E}{\tau_{rk}^E}\right) \le D, \qquad \forall r = 1, \ldots, m \tag{15}$$

$$E_{rk} \ge 0; N_{rk}^E \ge 0; P_{rk}^E \ge 0, \qquad \forall k \in K, r = 0, 1, \ldots, m \tag{16}$$

$$N_{rs}^s \ge 0; P_{rs}^s \ge 0, \qquad \forall s \in S, r = 0, 1, \ldots, m \tag{17}$$

$$G_{rr'} \ge 0, \qquad \forall (r, r') = 1, \ldots, m; r \neq r' \tag{18}$$

$$\lambda_r \ge 0, \qquad \forall r = 0, 1, \ldots, m \tag{19}$$

$$D \ge 0 \tag{20}$$

$$L_r \in \{0, 1\}, \qquad \forall r = 1, \ldots, m \tag{21}$$

The objective function (1) represents the overall stakeholder dissatisfaction that needs to be minimized. The dissatisfaction is incurred at the national level (2) and regional level (3) where percentage normalization is used in order to overcome incommensurability. Constraints (4) determine the unwanted deviational variables for each sub-goal and for each stakeholder (N_{rs}^S and P_{rs}^S). Constraints (5) give the national and regional level goals for energy generation. Constraint (6) determines the amount of energy produced at the national level for each generation source, which is based on energy produced by each region. This constraint also ensures that the total energy produced at the national level does not exceed the target that has been set. Constraints (7) impose that the energy produced by each generation source and each region must consider the maximum capacity installed and the capacity factor of the generation. Inequality set (8) ensure that the energy demand for each region must be satisfied by a combination of their production and importation from other regions. Here, the transmission loss factor needs to be considered.

Constraints (9) indicate the regions that have surplus production. Constraints (10) ensure that only surplus regions may transmit electricity to other regions, whereas Constraints (11) enforce that surplus regions do not receive electricity from others. Constraints (12–13) guarantee that for each region and for each sub-goal, the weighted, normalized, unwanted deviation from each stakeholder goal target is less than, or equal to, the maximal value for that stakeholder (λ_r). The interpretation of the inequality set (14) is similar to (12–13); however, it is for the energy produced by each stakeholder. Constraints (15) ensures that the parametric combination of the worst case and average deviations for each region is less than or equal to the worst case regional score (D). Constraints (16–20) indicate the non-negativity of the decision variables, whereas binary variables L are defined in Equation (21). The ENGP energy mix model expressed by Equations(1)–(21) is considered as a Mixed Integer Nonlinear Programming (MINLP). The sources of non-linearity are Equations (2), (3), (10), (12), (13), and (15), where Variables E_{rk} are mainly used to normalize the goals. Here, goal target parameters or other forms of normalization constant could also be used, however they do not give a normalizing factor as accurate as the level of energy production.

2.2. A multiple-period generation expansion planning (energy transition) model

The multi-period stochastic generation expansion planning model is developed where the main input is the energy mix to be achieved in the future period for each region (χ_{rk}); it is obtained by the goal programming-based energy mix model described in Subsection 2.1. By implementing the proposed generation expansion planning model, we can analyze the energy mix transition over the planning horizon. The objective of this model is to determine the optimal generation sources, level of expansion, and timing of new generator construction in order to minimize the total cost, while satisfying the electricity demand. The model considers the existing installed capacity (MW) at the beginning of the period for each generation source and for each region (b_{rk}). The expansion and reduction of the installed capacity for generation source k and region r are limited to φ^{+}_{rk} and φ^{-}_{rk} respectively. Similar to the previous model, the installed capacity must not exceed κ_{rk}. We also consider a set of parameters for each generation source including the capital cost, decommissioning cost, fixed operating cost, variable operating cost, fuel cost, heat rate, capacity factor and amount of CO_2 produced. The CO_2 price (ρ_t) is taken into account in order to penalize the amount of CO_2 generated. The electricity can be transmitted from one region to others in order to satisfy demand, considering the transmission loss factor and the transmission cost.

A two-stage stochastic program model (Birge and Louveaux, 2011) is developed to deal with the uncertain parameters which are captured by their probability distributions. Scenarios are used to capture a realization of these parameters and we denote Ω to be the set of scenarios indexed by $\omega \in \Omega$. It is also assumed that we can estimate the probability associated with a scenario (p_ω). In this study, the uncertain parameters include the fuel cost ($o_{kt\omega}$), the annual energy demand for each region ($\hat{\tau}^{E}_{rt\omega}$), and the national peak load demand ($\delta_{t\omega}$). It is assumed that the additional ($\hat{X}_{krt}$) and reduced ($\hat{Y}_{krt}$) installed capacities decision variables for all periods must be determined before the realization of uncertain parameters, hence, these are determined in the first stage. The second stage decision variables include annual electricity generated ($\hat{E}_{krt\omega}$), electricity transmitted from one region to another ($\hat{G}_{rr't\omega}$), and the ones indicating whether or not a region has surplus production ($\hat{L}_{rt\omega}$). The second stage are recourse decisions, because variables $\hat{E}_{krt\omega}$, $\hat{G}_{rr't\omega}$, and $\hat{L}_{rt\omega}$ are obtained so that the best response on the occurring scenario is given to the setting defined by the first-stage decisions. We also introduce auxiliary variable $\hat{A}_{krt}$ to represent the maximum electricity produced by each generation among the scenarios. This variable may determine the total power plant's capacity for each period. Excess capacity for each scenario is penalized based on parameter $\bar{c}$. Two parameters denoted by $\hat{c}$ and $\tilde{c}$ are used to penalize the unmet annual energy demand and unmet peak load, respectively. These parameters aim to achieve the feasibility of the obtained solutions over the set of scenario realizations (Mulvey et al., 1995). Subsequently, two decision variables $U^a_{rt\omega}$ and $U^p_{t\omega}$ are introduced to represent unmet annual energy demand and unmet peak load, respectively. Here, we define $\mathbf{X} = \{\hat{X}_{krt}, \hat{Y}_{krt}, \hat{A}_{krt}\}$ as the first-stage decisions and $\mathbf{Y} = \{\hat{E}_{krt\omega}, \hat{G}_{rr't\omega}, \hat{L}_{rt\omega}, U^a_{rt\omega}, U^p_{t\omega}\}$ as the second-stage decisions for each scenario $\omega \in \Omega$. The notations used in this model are similar to the ones in the previous model, with additional notations as follows:

Sets and indices

T: set of time periods indexed by $t \in T$
Ω^F: set of scenarios for the fuel cost
Ω^D: set of scenarios for the annual energy demand for each region
Ω^P: set of scenarios for the national peak load
Ω: combination of scenarios ($\Omega = \{\Omega^F, \Omega^D, \Omega^P\}$) indexed by $\omega \in \Omega$

Parameters

b_{rk}: existing installed capacity (MW) of generation source $k \in K$ in region $r = 1, \ldots, m$ at the initial period
c_k: unit capital cost (\$/MW) to build new generation source $k \in K$
q_k: unit decommissioning cost (\$/MW) to retire generation source $k \in K$
f_k: unit fixed operating cost (\$/MW) of generation source $k \in K$
v_k: unit variable operating cost (\$/MWh) of generation source $k \in K$
$o_{kt\omega}$: unit fuel cost (\$/MMBtu) of generation source $k \in K$ in period $t \in T$ under scenario $\omega \in \Omega$
h_k: unit heat rate (MMBtu/MWh) of generation source $k \in K$ representing the efficiency of the power plant that converts a fuel into electricity. The heat rate indicates the amount of energy used by a power plant to generate one MWh of power.
ψ_k: capacity factor (%) of generation source $k \in K$
ϕ: total operating hours
$l_{rr'}$: the transmission loss factor from region r to r', where $(r, r') = 1, \ldots, m$ and $r \neq r'$
$\bar{l}$: the average transmission (distribution) loss factor
$l^c_{rr'}$: the transmission cost per MWh from region r to r', where $(r, r') = 1, \ldots, m$ and $r \neq r'$
ε_k: unit CO_2 emissions (tons of CO_2/MWh) caused by generation source $k \in K$
ρ_t: unit CO_2 emissions price (\$/ton CO_2) in period $t \in T$
$\hat{\tau}^E_{rt\omega}$: annual energy demand (MWh) in period $t \in T$ under scenario $\omega \in \Omega$ for region $r = 1, \ldots, m$
$\delta_{t\omega}$: peak load (MW) in period $t \in T$ at the national level under scenario $\omega \in \Omega$
$\hat{p}_\omega$: probability that scenario ω occurs
κ_{rk}: maximum total capacity of generation $k \in K$ that can be installed in region $r = 1, \ldots, m$ over the planning horizon
φ^+_{rk}: maximum capacity of generation $k \in K$ that can be installed per year in region $r = 1, \ldots, m$
φ^-_{rk}: maximum capacity of generation $k \in K$ that can be retired per year in region $r = 1, \ldots, m$
χ_{rk}: minimum percentage of electricity produced by generation source $k \in K$ at the end of planning horizon (i.e., 2050) in region $r = 1, \ldots, m$ (obtained from the first model)
$\bar{c}$: unit penalty for excess capacity (\$/MW)

$\hat{c}$: unit penalty for unmet annual energy demand (\$/MWh)
$\tilde{c}$: unit penalty for unmet peak load (\$/MW)

The First-Stage Decision Variables

$\hat{X}_{rkt}$: installed capacity (MW) of new generation $k \in K$ during period $t \in T$ in region $r = 1, \ldots, m$
$\hat{Y}_{rkt}$: total capacity (MW) of generation $k \in K$ that are retired during period $t \in T$ in region $r = 1, \ldots, m$
$\hat{A}_{rkt}$: maximum annual electricity production (MWh) of generation $k \in K$ generated during period $t \in T$ in region $r = 1, \ldots, m$

The Second-Stage Decision Variables

$\hat{E}_{rkt\omega}$ = annual electricity generated (MWh) by generation source $k \in K$ in region $r = 1, \ldots, m$ during period $t \in T$ under scenario $\omega \in \Omega$
$\hat{G}_{rr't\omega}$: electricity (MWh) transmitted from region r to r' in period $t \in T$, where $(r, r') = 1, \ldots, m$ and $r \neq r'$, under scenario $\omega \in \Omega$

$$\hat{L}_{rt\omega} = \begin{cases} 1 & \text{if region } r \text{ has surplus production in period } t, \text{ where } r = 1, \ldots, m, \text{ under scenario } \omega \in \Omega, \\ 0 & \text{otherwise} \end{cases}$$

$U^a_{rt\omega}$: unmet annual energy demand (MWh) during period $t \in T$ in region $r = 1, \ldots, m$, under scenario $\omega \in \Omega$
$U^p_{t\omega}$: unmet peak load (MW) during period $t \in T$ under scenario $\omega \in \Omega$

In the proposed two-stage stochastic programming model, the first-stage model optimizes the strategic capacities decision, whereas the second-stage model minimize the expected operational costs. The first-stage objective function and constraints can be expressed as follows:

$$\min_{\mathbf{X}} \left\{ \sum_{t \in T} \sum_{r=1}^{m} \sum_{k \in K} \Big(c_k \cdot \hat{X}_{rkt} + q_k \cdot \hat{Y}_{rkt} + f_k \big(b_{kr} + \sum_{t'=1}^{t} (\hat{X}_{rkt'} - \hat{Y}_{rkt'}) \big) \Big) + \sum_{\omega \in \Omega} p_\omega \cdot \mathbb{Q}(\mathbf{X}, \omega) \right\} \tag{22}$$

subject to:

$$b_{rk} + \sum_{t \in T} \Big(\hat{X}_{rkt} - \hat{Y}_{rkt} \Big) \leq \kappa_{rk}, \qquad \forall k \in K; r = 1, \ldots, m \tag{23}$$

$$\sum_{t \in T} \hat{Y}_{rkt} \leq b_{rk}, \qquad \forall k \in K; r = 1, \ldots, m \tag{24}$$

$$\hat{X}_{rkt} \leq \varphi^+_{rk}, \qquad \forall t \in T; k \in K; r = 1, \ldots, m \tag{25}$$

$$\hat{Y}_{rkt} \leq \varphi^-_{rk}, \qquad \forall t \in T; k \in K; r = 1, \ldots, m \tag{26}$$

$$\hat{A}_{rkt} = \psi_k \cdot \phi \cdot \big(b_{rk} + \sum_{t'=1}^{t} (\hat{X}_{rkt'} - \hat{Y}_{rkt'}) \big), \qquad \forall t \in T; k \in K; r = 1, \ldots, m; \tag{27}$$

$$\hat{X}_{rkt} \geq 0, \qquad \forall k \in K; t \in T; r = 1, \ldots, m; \tag{28}$$

$$\hat{Y}_{rkt} \geq 0, \qquad \forall k \in K; t \in T; r = 1, \ldots, m; \tag{29}$$

$$\hat{A}_{rkt} \geq 0, \qquad \forall k \in K; t \in T; r = 1, \ldots, m; \tag{30}$$

The objective function (22) aims to minimize the total costs, where the first term of this equation determine the total construction, decommissioning, and fixed operating costs, whereas the second term calculate the expected operational costs. Constraints (23) ensure that the total capacity installed for each generation source and each region cannot exceed the maximum capacity available. Constraints (24) guarantee that only existing capacity can be terminated. Constraints (25) and (26) aim to limit the capacity expansion and the capacity reduction respectively. Constraints (27) determine the maximum annual electricity production (MWh) that can be generated by power plant $k \in K$ during period $t \in T$ in region $r = 1, \ldots, m$ by considering its capacity factor and the total operating hours per year ($\phi =$ 8,760 hours). Constraints (28)–(30) indicate that the first-stage model uses non-negative continuous variables.

The second-stage model is implemented after the strategic capacities decisions from the first-stage model have been applied and the uncertain parameters are revealed. Here, the second-stage model consists of a set of deterministic models over all possible scenarios. The objective function of this model for a single scenario $\omega \in \Omega$ is associated with all incurred operational costs. The second-stage model is formulated as follows:

$$\min_{\mathbf{Y}} \mathbb{Q}(\mathbf{X}, \omega) = \sum_{t \in T} \Bigg\{ \Big(\sum_{r=1}^{m} \sum_{k \in K} \Big[\hat{E}_{rkt\omega} \big(v_k + o_{kt\omega} \cdot h_k + \varepsilon_k \cdot \rho_t \big) + \bar{c} \big(\frac{\hat{A}_{rkt} - \hat{E}_{rkt\omega}}{\gamma \cdot \psi_k} \big) + \sum_{r'=1}^{m} \hat{G}_{rr't\omega} \cdot l^{c}_{rr'} \Big] \Big) + \Big(\sum_{r=1}^{m} \Big[U^{a}_{rt\omega} \cdot \hat{c} \Big] + U^{p}_{t\omega} \cdot \tilde{c} \Big) \Bigg\} \tag{31}$$

subject to:

$$\hat{E}_{rkt\omega} \leq \hat{A}_{rkt}, \qquad \forall t \in T; k \in K; r = 1, \ldots, m; \omega \in \Omega \tag{32}$$

$$\sum_{k \in K} \hat{E}_{rkt\omega} + \sum_{\substack{r'=1 \\ r' \neq r}}^{m} \hat{G}_{r'rt\omega} \cdot (1 - l_{r'r}) - \sum_{\substack{r'=1 \\ r' \neq r}}^{m} \hat{G}_{rr't\omega} + U^{a}_{rt\omega} \geq \hat{\tau}^{E}_{rt\omega}, \quad \forall t \in T; r = 1, \ldots, m; \omega \in \Omega \tag{33}$$

$$\sum_{k \in K} E_{rkt\omega} \geq \hat{\tau}^{E}_{rt\omega} \cdot \hat{L}_{rt\omega}, \qquad t \in T; r = 1, \ldots, m; \omega \in \Omega \tag{34}$$

$$\sum_{\substack{r'=1 \\ r' \neq r}}^{m} \hat{G}_{rr't\omega} \leq \hat{\tau}^{E}_{rt\omega} \cdot \hat{L}_{rt\omega}, \qquad t \in T; r = 1, \ldots, m; \omega \in \Omega \tag{35}$$

$$\sum_{\substack{r'=1 \\ r' \neq r}}^{m} \hat{G}_{r'rt\omega} \leq \hat{\tau}^{E}_{rt\omega} \cdot (1 - \hat{L}_{rt\omega}), \qquad t \in T; r = 1, \ldots, m; \omega \in \Omega \tag{36}$$

$$(1 - \bar{l}) \sum_{k \in K} \psi_k \sum_{r=1}^{m} \Big[b_{rk} + \sum_{t'=1}^{t} \Big(\hat{X}_{rkt'} - \hat{Y}_{rkt'} \Big) \Big] + U^{p}_{t\omega} \geq \delta_{t\omega}, \qquad \forall t \in T, \omega \in \Omega \tag{37}$$

$$\hat{E}_{rk|T|\omega} \geq \chi_{rk} \sum_{k' \in K} \hat{E}_{rk'|T|\omega}, \qquad \forall k \in K; r = 1, \ldots, m; \omega \in \Omega \tag{38}$$

$$\hat{E}_{rkt\omega} \geq 0, \qquad \forall k \in K; t \in T; r = 1, \ldots, m; \omega \in \Omega \tag{39}$$

$$U^a_{rt\omega} \geq 0, \quad \forall t \in T; r = 1, \ldots, m; \omega \in \Omega \tag{40}$$

$$U^p_{t\omega} \geq 0, \quad \forall t \in T; \omega \in \Omega \tag{41}$$

$$\hat{G}_{rr't\omega} \geq 0, \quad \forall t \in T, (r, r') = 1, \ldots, m; r \neq r', \omega \in \Omega \tag{42}$$

$$\hat{L}_{rt\omega} \in \{0, 1\}, \quad \forall t \in T; r = 1, \ldots, m; \omega \in \Omega \tag{43}$$

The objective function (31) aim to minimize the expected operational costs that consist of the variable, fuel, emission, transmission and penalty costs. Three types of penalty costs are computed, which are associated with the under production capacity, unmet annual energy demand, and unmet annual peak load. Constraints (32) ensure that the amount of energy generated by each generation source for each scenario is less than the maximum capacity. Constraints (33) impose that the annual energy demand for each region must be satisfied otherwise a penalty is incurred. Each region's demand can be met by generation within that region and/or by importing energy from other regions. Constraints (34) determine the regions that have surplus production for each period under scenario $\omega \in \Omega$. Only surplus regions may export electricity to other regions which is expressed by Constraints (35). Constraints (36) ensure that surplus regions do not receive electricity from others. Constraints (37) state that the total installed capacity should be able to meet the national peak load energy demand for each period, taking into account the average transmission loss. A penalty cost is also incurred if the peak load is not satisfied. Constraints (38) enforce the energy mix in the final period which is determined by the output of the first model (the energy mix model). Constraints (39–42) indicate that the model uses non-negative continuous variables, whereas Constraints (43) define the binary variables used in the model.

2.3. Sample average approximation for the stochastic generation expansion planning model

The proposed stochastic generation expansion planning model considers several uncertain parameters. When a large number of scenarios considered in the model is used, the model can be challenging to solve using an exact method, mainly due to memory issues. We implement the sample average approximation (SAA) approach in this study to address the proposed problem. The SAA method has been widely used to solve problems under uncertainty (Bertsimas et al., 2018). This method is suitable for tackling the problem with an enormous number of possible scenarios, where the uncertain parameters follow a continuous distribution (e.g., normal or exponential distributions). A comprehensive guide to this method is provided by Kim et al. (2015).

Algorithm 1 presents the steps of the proposed SAA for solving the proposed model. M independent samples are first randomly generated, and we refer to each sample as a SAA problem. Each SAA problem consists of N scenarios generated from set Ω based on the given distribution function of the uncertain parameters using the Monte-Carlo sampling method (Shapiro, 2003), which is a mathematical technique used to estimate the possible outcomes of an uncertain event. These scenarios represent the realization of uncertain parameters, namely the fuel cost, annual energy demand, and peak load. Algorithm 2 describes the steps of generating a scenario ω. In the SAA, the expected value of the objective function is estimated by the average over objective values $\mathbb{Q}(\mathbf{X}, \omega)$. Therefore, the original two-stage stochastic problem (22)–(43) is replaced by the equivalent deterministic programming model expressed as follows:

$$\min Z = \sum_{t\in T}\sum_{r=1}^{m}\sum_{k\in K}\Big(c_k\cdot \hat{X}_{rkt} + q_k\cdot \hat{Y}_{rkt} + f_k\big(b_{kr} + \sum_{t'=1}^{t}(\hat{X}_{rkt'} - \hat{Y}_{rkt'})\big)\Big)+ \tag{44}$$

$$\frac{1}{N}\sum_{\omega=1}^{N}\sum_{t\in T}\Bigg\{\sum_{r=1}^{m}\sum_{k\in K}\Big[\hat{E}_{rkt\omega}\big(v_k + o_{kt\omega}\cdot h_p + \varepsilon_k\cdot\rho_t\big) + \bar{c}\big(\frac{\hat{A}_{rkt} - \hat{E}_{rkt\omega}}{\gamma\cdot\psi_k}\big) + \sum_{\substack{r'=1\\ r\neq r'}}^{m}\hat{G}_{rr't\omega}\cdot l^{c}_{rr'}\Big]+$$

$$\sum_{r=1}^{m}\Big[U^{a}_{rt\omega}\cdot\hat{c}\Big] + U^{p}_{t\omega}\cdot\tilde{c}\Bigg\}$$

$$\text{s.t. } (23)\text{–}(30) \text{ and } (32)\text{–}(43)$$

The SAA model (44), (23)–(30) and (32)–(43) is considered as a Mixed Integer Linear Programming (MILP) and can be optimally solved by an exact method using a commercial solver (e.g., CPLEX). In summary, the model consists of $m|T||\Omega|$ binary variables and $m(3|K||T| + |K||T||\Omega| + (m+1)|T||\Omega|) + |T||\Omega|$ continuous variables, together with $m(2|K| + 3|T||K| + |T||K||\Omega| + 4|T||\Omega| + |K||\Omega|) + |T||\Omega|$ constraints. We denote Z_i^N and $\mathbf{X}_i^N$ as the objective value and the first-stage solution of the i^{th} SAA problem, respectively.

Algorithm 1 Procedure SAA for the stochastic generation expansion planning model

1: Set parameters M, N, N' and α
2: Set $Z^N = \infty$
3: **for each** sample $i = 1$ to M **do**
4: Generate N scenarios to be included in the i^{th} SAA problem
5: Solve the i^{th} SAA problem expressed by Model (44), (23)–(30) and (32)–(43)
6: Let Z_i^N be the objective value and $\mathbf{X}_i^N$ be the first-stage solution
7: **if** $Z_i^N < Z^N$ **then** $Z^N = Z_i^N$ and $\hat{\mathbf{X}} \leftarrow \mathbf{X}_i^N$
8: **end for**
9: Obtain the average and variance of the M SAA objective functions using (45) and (46), respectively
10: Calculate an approximate $100(1-\alpha)\%$ confidence lower bound (LB) using (47)
11: **for each** scenario $\omega = 1$ to N' **do**
12: Generate a scenario (ω) for the ω^{th} second-stage problem expressed by Model (31)–(43)
13: Solve the ω^{th} second-stage problem by fixing the first-stage solution $\hat{\mathbf{X}}$
14: Let $\hat{Z}_\omega$ be its objective value
15: **end for**
16: Compute the average and variance of N' second-stage problems using (48) and (49), respectively
17: Obtain the approximate $100(1-\alpha)\%$ confidence upper bound (UB) using (50)
18: Determine the optimality gap using (51).

Determination of lower and upper bounds

The optimality gap is used to evaluate the robustness of the proposed SAA where statistical confidence intervals are used to estimate lower (LB) and upper (UB) bounds (Shapiro et al., 2009). To estimate a valid LB, we first calculate the average and standard deviation of the objective values of the M the SAA problems which are denoted by $\bar{Z}^{MN}$ and $\hat{\sigma}^{MN}$, respectively, formulated by the following expressions:

Algorithm 2 The main steps of generating a scenario ω

for each period $t \in T$ **do**
  Generate the national peak load during period t ($\delta_{t\omega}$) based on the given distribution
  for each region $r = 1$ to m **do**
    Generate the electricity demand of region r during period t ($\hat{\tau}^{E}_{rt\omega}$) based on the given distribution
  end for
  for each generation $k \in K^F$, where $K^F = \{$coal, gas, nuclear, and biomass$\}$ **do**
    Generate the fuel cost of generation k during period t ($o_{kt\omega}$) based on the given distribution
  end for
end for

$$\bar{Z}^{MN} = \frac{1}{M}\sum_{i=1}^{M} Z_i^N \tag{45}$$

$$\hat{\sigma}^{MN} = \sqrt{\frac{1}{M(M-1)}\sum_{i=1}^{M}(Z_i^N - \bar{Z}^{MN})^2} \tag{46}$$

An approximate $100(1-\alpha)\%$ confidence LB can be determined using the average and standard deviation of M SAA models as follows:

$$LB_{1-\alpha} = \bar{Z}^{MN} - \mathbf{t}_{\alpha,M-1} \cdot \hat{\sigma}^{MN} \tag{47}$$

where $\mathbf{t}_{\alpha,M-1}$ is the α-critical value of the $\mathbf{t}$-distribution with $(M-1)$ degrees of freedom.

From the M solutions obtained, we take the first-stage best solution ($\hat{\mathbf{X}}$) that consists of variables $\hat{X}_{krt}$ and $\hat{Y}_{krt}$. Solution $\hat{\mathbf{X}}$ provides the best objective function value (Z^N). Then, multiple (N') second-stage problems are formed, where $N' >> N$. The second-stage model aims to minimize the total operational cost expressed by Equations (31)–(43), with fixed first-stage solution ($\hat{\mathbf{X}}$). Each second-stage problem consists of only one scenario generated using the Monte-Carlo sampling method. This problem can easily be solved by an exact method (e.g., CPLEX), with $\hat{Z}_\omega$ be the optimal objective function value of the ω^{th} second-stage problem. To estimate the upper bound (UB), the average ($\bar{Z}^{N'}$) and the standard deviation $\hat{\sigma}^{N'}$ of objective values are then calculated using the following equations:

$$\bar{Z}^{N'} = \frac{1}{N'}\sum_{\omega=1}^{N'} \hat{Z}_\omega \tag{48}$$

$$\hat{\sigma}^{N'} = \sqrt{\frac{1}{N'(N'-1)}\sum_{\omega=1}^{N'}(\hat{Z}_\omega - \bar{Z}^{N'})^2} \tag{49}$$

The approximate $100(1-\alpha)\%$ confidence upper bound (UB) is determined using the following expression:

$$UB_{1-\alpha} = \hat{Z}^{N'} + \mathbf{z}_\alpha \cdot \hat{\sigma}^{N'} \tag{50}$$

where $\mathbf{z}_\alpha$ is the standard normal critical value with a $100(1-\alpha)\%$ confidence level. The statistical optimality gap (%Gap) can be formulated based on UB and LB values as follows:

$$\%\text{Gap} = \frac{UB_{1-\alpha} - LB_{1-\alpha}}{UB_{1-\alpha}} \times 100 \tag{51}$$

It is worth noting that the measurement unit used for $\bar{Z}^{MN}, \bar{Z}^{N'}$, $LB_{1-\alpha}$, and $UB_{1-\alpha}$ is cost which is the same as the objective function given in Equation (22). Furthermore, it should be noted that the SAA works by minimizing a sample average function (44). Alternative methods are available, for instance those based on the Minimax or Maximin functions, that minimize the affect of the worse-case scenario. However, given the nature of the problem studied in our paper, the SAA is preferred in order to not give over-emphasis to extreme or unusual generated scenario settings.

3. The Electricity Generation System in China

In this study, the proposed methodology is assessed using China's energy system. China has the highest total energy consumption worldwide (Global Energy Statistical Yearbook, 2021) and aspires to be carbon neutral before 2060 (Mallapaty, 2020). According to the 2018 China Electric Power Yearbook (CEPY, 2018), China's power grid can be divided into six regions (see Table 2). For convenience of presentation, we first outline three energy demand/production settings introduced by the Economics and Technology Research Institute (ETRI) of the China National Petroleum Corporation (CNPC) to envision electricity demand and production levels by 2050 (CNPC ETRI, 2019). The baseline (BL, hereafter) setting is used as a reference setting, whilst the beautiful China (BC) setting emphasizes a higher degree of environmental and social sustainability with an efficient, clean, and low-carbon energy system. The intelligence connected (IC) setting, however, expects an interconnected system powered by advanced digital and intelligent technology with higher productivity and faster economic growth and, therefore, potentially increased energy demand. Input data were estimated based on, or directly sourced from, the existing literature and industry documentation and records. Whenever regional data is not available, we use accessible national data to gauge for each region based on the regional energy production ratio for each generation source indicated in CEPY (2018).

Table 2: Regional division in the model

Regions	Provinces, municipalities and autonomous regions included	Regional power grids
Northeast China	Liaoning, Jilin, Heilongjiang, East Inner Mongolia	Northeast power grid
North China	Beijing, Tianjin, Hebei, Shandong, Shanxi, West Inner Mongolia	North power grid
East China	Zhejiang, Shanghai, Jiangsu, Fujian, Anhui	East power grid
Central China	Hubei, Hunan, Jiangxi, Chongqing, Henan, Sichuan	Central power grid
Northwest China	Gansu, Qinghai, Ningxia, Shaanxi, Xinjiang, Tibet	Northwest power grid
South China	Guangxi, Yunan, Guizhou, Guangdong, Hainan	Southern power grid

3.1. The Energy Mix Model

3.1.1. Maximum capacity of each generation source for each region

Table 3 shows the detailed parameters for the maximum capacity available for each generation source across all regions, based on Zhang et al. (2018) (for coal, nuclear, hydro, wind, and solar) and Yi et al. (2016) (for biomass). It is assumed that the maximum capacity for gas for each region is set to 100,000 MW.

Table 3: Maximum capacity of each generation source for each region (unit: MW)

Region	Coal	Gas	Hydro	Nuclear	Wind	Solar	Biomass
Northeast China	105,794.2	100,000.0	25,780.2	24,444.4	211,875.0	236,767.8	29,000.0
North China	279,206.4	100,000.0	15,164.8	19,555.6	508,500.0	153,628.0	30,000.0
East China	258,179.8	100,000.0	40,945.1	68,444.4	52,262.5	79,525.1	17,000.0
Central China	140,868.9	100,000.0	286,615.4	48,888.9	16,950.0	222,308.7	44,000.0
Northwest China	132,143.6	100,000.0	89,472.5	0.0	310,750.0	477,150.4	20,000.0
South China	122,807.0	100,000.0	232,022.0	58,666.7	29,662.5	200,620.1	19,000.0

3.1.2. The annual energy production target

The annual energy production target for each generation source across all regions is summarized in Table 4, referring to the 2050 energy demand projections for the BL setting in CNPC ETRI (2019).

Table 4: The annual energy production target of each generation source for each region (unit: 10^9 MWh)

Region	Coal	Gas	Hydro	Nuclear	Wind	Solar	Biomass
Northeast China	0.397110	0.142552	0.025325	0.133119	0.347780	0.149481	0.044023
North China	1.048032	0.376217	0.014622	0.000000	0.492384	0.463701	0.093199
East China	0.969106	0.347884	0.113725	0.701692	0.150209	0.356561	0.141999
Central China	0.528767	0.189814	0.902471	0.000000	0.112657	0.212908	0.068988
Northwest China	0.496015	0.178057	0.178710	0.000000	0.404164	0.713123	0.006135
South China	0.460970	0.165476	0.665147	0.565189	0.192806	0.104226	0.045657

3.1.3. The transmission loss factor and capacity factor

The transmission loss factor between regions was derived from the product of the transmission loss parameter (i.e., 2.75%/1000 km from Yi et al. (2016)) and the distance between regions. We estimate the distances among regions by the point-to-point distances among the central cities or states of each of the six regions. Specifically, the central city or state of each region is enumerated as follows: Qiqihar (Heilongjiang) for the Northeast China, Shijiazhuang (Hebei) for the North China, Hangzhou (Zhejiang) for East China, Zhangjiajie (Hunan) for Central China, Bayingol Mongolian Autonomous Prefecture for Northwest China, and Hechi (Guangxi) for South China. The point-to-point distance between the central cities or states are subsequently evaluated using Google Maps.

The specific parameters for the capacity factor of each generation source were obtained from the U.S. Energy Information Administration summarized by Irawan et al. (2022) and Thangavelu et al. (2015).

3.1.4. The assessment and target of sub-goals

Each generation source has been assessed with regard to the three main goals and their corresponding sub-goals as illustrated in Table 1. This is based on the assessment of Streimikiene et al. (2012) where normalization using a scale in range [0.1–1] is applied. The value of a_{ks} is formulated as follow: $a_{ks} = 0.1 + 0.9(\hat{a}_{ks} - \min_{k' \in K}\{\hat{a}_{k's}\})/(\max_{k' \in K}\{\hat{a}_{k's}\} - \min_{k' \in K}\{\hat{a}_{k's}\})$, where $\hat{a}_{ks}$ represents the original assessment value provided by Streimikiene et al. (2012) for sub-goal $s \in S$ of generation $k \in K$. Table 5 shows the assessment of each generation source on the sub-goals.

Table 5: The assessment of each generation source on the sub-goals

Generation Source	Minimizing Function								
	PR-COST	GRID-COST	CO2eq.	ENV	RADIO	HEALTH	FOOD	ACC-PAST	ACC-FUT
Coal	0.1272	0.1000	1.0000	0.5564	0.1062	0.9169	0.1000	1.0000	1.0000
Gas	0.2936	0.1000	0.8409	0.3983	0.1000	0.5195	0.1000	0.5846	0.4000
Hydro	0.3326	0.1000	0.1061	0.1331	0.1000	0.1598	0.1000	0.1000	0.4000
Nuclear	0.1000	0.1000	0.1036	0.1204	1.0000	0.1279	0.1000	0.1000	1.0000
Wind	0.2667	1.0000	0.1000	0.1000	0.1012	0.1000	0.1000	0.1000	0.1000
Solar	1.0000	0.1000	0.2190	0.2453	0.1000	0.6462	0.1000	0.1000	0.1000
Others	0.2039	1.0000	0.1717	1.0000	0.1168	1.0000	1.0000	0.5846	0.4000
Average	0.3320	0.3571	0.3630	0.3648	0.2320	0.4958	0.2286	0.3670	0.4857
25% Pctile	0.1656	0.1000	0.1049	0.1268	0.1000	0.1439	0.1000	0.1000	0.2500

Generation Source	Maximizing Function			
	AVAILAB	SECURE	PEAKLOAD	EMPL
Coal	0.8875	0.6400	0.5500	0.1978
Gas	0.8875	0.1000	1.0000	0.1685
Hydro	0.9550	1.0000	0.3700	0.2453
Nuclear	0.9438	0.8200	0.1900	0.1000
Wind	0.2575	1.0000	0.1000	0.1280
Solar	0.1000	1.0000	0.1000	1.0000
Others	1.0000	1.0000	1.0000	0.6925
Average	0.7188	0.7943	0.4729	0.3617
75% Pctile	0.9494	1.0000	0.7750	0.4689

It is assumed that four strategic plans are put forward for all stakeholders at national and regional levels. These strategic plans are represented by the sub-goal target settings, which is presented in Table 6. The first setting is called Avg, which is determined based on the average assessment value of the sub-goal (Table 5). Avg-Env, Avg-Cost and Avg-Soc are the target settings which emphasize environmental, economic, and social dimensions, respectively. Here, the targets are also set based on the average assessment value, except for the prioritized dimension. For this dimension, the sub-goal values for minimizing and maximizing functions are set based on the 25% and 75% percentiles, respectively, of the assessment values in Table 5. For example, in the target setting Avg-Env, the value of sub-goal CO_2eq is set to 0.1049 which is the 25% percentile of the assessment values of this sub-goal, whereas in other target settings, it is set to 0.3648 (the average value).

Table 6: Sub-goal settings for national and regional levels

Main Goal	Sub-goals	Sub-goal Target Settings			
		Avg	Avg-Env	Avg-Cost	Avg-Soc
Economic	PR-COST	0.3320	0.3320	0.1656	0.3320
	AVAILAB	0.7188	0.7188	0.9494	0.7188
	SECURE	0.7943	0.7943	1.0000	0.7943
	GRID-COST	0.3571	0.3571	0.1000	0.3571
	PEAKLOAD	0.4729	0.4729	0.7750	0.4729
Environmental	CO2eq.	0.3630	0.1049	0.3630	0.3630
	ENV	0.3648	0.1268	0.3648	0.3648
	RADIO	0.2320	0.1000	0.2320	0.2320
	HEALTH	0.4958	0.1439	0.4958	0.4958
Social	EMPL	0.3617	0.3617	0.3617	0.4689
	FOOD	0.2286	0.2286	0.2286	0.1000
	ACC-PAST	0.3670	0.3670	0.3670	0.1000
	ACC-FUT	0.4857	0.4857	0.4857	0.2500

Parameters for penalizing unwanted deviations of sub-goals are taken from Yu et al. (2020) and determined by aggregating provincial data into regional ones. The penalties for unwanted deviations of energy production are assumed to be the same for each generation source and each region.

3.1.5. The Setting of Instances

Twelve instances have been constructed from four sub-goal target settings (ie., Avg, Avg-Env, Avg-Cost, Avg-Soc) and the three aforementioned energy demand/production settings (i.e., BL, BC, IC). Instance 1 combines the Avg sub-goal target and the baseline demand (BL) settings, whereas Instance 12 is the combination of Avg-Soc and IC settings. Section 4.1 discusses the experimental results from the first model for each of the 12 instances. The energy mix result for each instance will be used as an input for the second model (Generation Expansion Planning model).

3.2. The Generation Expansion Planning Model

3.2.1. Power plants of different generation sources

Detailed parameters for the costs of operating and decommissioning power plants of different generation sources, together with their heat rates, CO_2 emissions, and CO_2 prices, are shown in Table 7. The data for the capital cost, fixed and variable operation costs, heat rate, and CO_2 emission rate are set by referring to the official data from Energy Information Administration (2020). According to EIA, variable operations and maintenance (O&M) costs, such as ammonia, water, and miscellaneous chemicals and consumables, are directly proportional to the plant generating output. For the hydro, solar, and wind cases, all O&M costs are treated as fixed costs. Fixed O&M costs also include costs directly related to the equipment design, such as labor, materials, contract services for routine O&M, and administrative and general costs. The decommissioning cost for the nuclear power plant was determined based on OECD Nuclear Energy Agency (2016), whereas for other power plants, it is calculated based on Kutani et al. (2016). There are also other interesting studies in formulating the power plant decommissioning cost, including for nuclear (Sorgulu and Dincer, 2018) and wind (Topham and McMillan, 2017). The CO_2 price data are taken from Rentizelas et al. (2012). The transmission cost was estimated at 23.2 $/MWh (Lin et al., 2019).

Table 7: Cost parameters for each generation source

	Coal	Gas	Hydro	Nuclear	Wind	Solar	Biomass
Capital cost ($/kW)	3,676	1,084	5,316	6,041	1,265	1,313	4,097
Decommissioning cost ($/kW)	183.8	54.2	265.8	730	63.25	65.65	204.85
Fixed operating cost ($/kW-y)	40.58	14.1	29.86	121.64	26.34	15.25	125.72
Variable operating cost ($/MWh)	4.5	2.55	-	2.37	-	-	4.83
Heat rate (Btu/kWh)	8,638	6,431	-	10,608	-	-	13,300
CO2 emission (lb/MMBtu)	206	117	0	0	0	0	206
Year		2025	2030	2035	2040	2045	2050
CO_2 price ($/tons)		37.416	47.76	60.96	77.796	99.288	126.72

3.2.2. Existing power plant capacity in 2020

We have collected data on the existing capacity of hydro, nuclear, and coal power plants in China (endcoal.org), whilst the pursuit of necessary data for other generation sources proved difficult. To this end, we estimated the capacity of power plants based on historical data regarding the electricity production level and the capacity factor for each generation source. Specifically, we first computed the electricity production growth rate from 2017 to 2020. Given the regional electricity production level for each generation source (CEPY, 2018) and its capacity factor (as indicated for the first model), we then calculated the capacity of existing power plants in 2017, and adjusted it for the year 2020 by incorporating the electricity production growth rate. The final values considered in the model (see Table 8) were derived by retrieving the available collected data, or our estimations when no data are available. Triangulating data from different sources also helps to check and ensure data reasonableness and integrity.

Table 8: Existing power plant capacity in 2020 (unit: MW)

Region	Coal	Gas	Hydro	Nuclear	Wind	Solar	Biomass
Northeast China	65,565.34	3,204.17	2,304.42	3,458.41	28,219.55	7,667.11	-
North China	173,036.53	8,456.26	1,330.48	-	39,953.07	23,783.86	-
East China	160,005.43	7,819.44	10,348.16	18,229.90	12,188.30	18,288.51	-
Central China	87,302.67	4,266.47	82,118.60	-	9,141.23	10,920.35	-
Northwest China	81,895.19	4,002.20	16,261.40	-	32,794.71	36,577.03	13.88
South China	76,108.94	3,719.43	60,523.70	14,683.57	15,644.68	5,345.87	-

3.2.3. The data of Energy Demand and Fuel costs

In the second model, the uncertain parameters include the annual energy demand for each region, the national peak load, together with the fuel cost for each generation source and capacity factor for renewable energy (solar and wind). Based on Zhang et al. (2018), we obtained the estimated regional electricity demand from 2025 to 2050 with a step size of five years under the BL setting. Note that Zhang et al. (2018) allocated three provinces/municipalities (Sichuan, Chongqing, and Tibet) to a separate region named Southwest. Therefore, we adjusted the data by obtaining the demand levels for the three provinces/municipalities, respectively, using the energy production ratio from CEPY (2018), and subsequently adding the demand for Sichuan and Chongqing to Central and Tibet to Northwest to be consistent with the regional division in our study. The regional demand for the BL setting is then leveraged to estimate the demand level for the BC and IC settings, using the demand/production ratios between the three settings suggested in CNPC ETRI (2019). To estimate the peak load during the expansion planning period, we deployed historical data on the yearly electricity demand from the National Bureau of Statistics of China (2020) and the corresponding peak load from 2005 to 2020 to develop a regression model. This model was used to project the peak load given the aforementioned energy demand estimation for the three settings.

The estimated fuel cost data for coal and gas from year 2025 to 2050 are taken from the Energy Information Administration (EIA, 2021). For nuclear, historical fuel cost data from 1970 to 2019 is also obtained from the EIA, it is then forecasted using an ARIMA model to estimate the price from 2025 to 2050. A zero fuel cost is applied for renewable generations including hydro, solar and wind. In this study, we assume that all the uncertain parameters follow the normal distribution with the forecasted/estimated value treated as the average value (μ). The standard deviation (σ) value of an uncertain parameter is estimated based on the value of μ, with σ assumed to be $0.05 \cdot \mu$. Random numbers for uncertain parameters are generated based on this normal distribution, $N(\mu, \sigma)$. Figure 3 illustrates the estimated energy demand for the Northeast region in the BL setting where the demand of each period is generated based on the normal distribution.

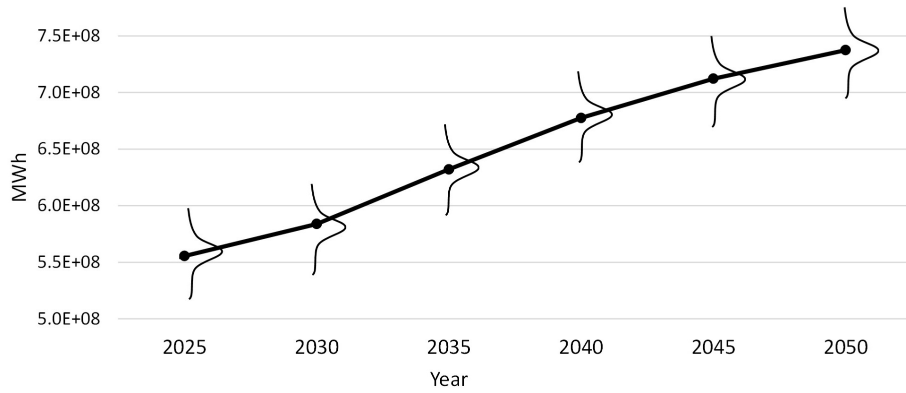

Figure 3: Energy demand for the Northeast region in the BL setting

4. Results and discussion

4.1. Results of the first model

This subsection presents the experimental results of the analysis carried out to assess the performance of the first model (the single-period goal programming-based energy mix model). The non-linear ENGP energy mix model is implemented in a non-linear optimizer that can solve the MINLP, namely Baron solver (https://minlp.com/baron-solver). The tests were run on a workstation with an Intel Xeon W-2133 CPU @3.60 GHz processor, 128.00 GB of RAM. We limited the computational time (CPU) to 24 hours (86,400 seconds) only, and set a relative gap to a very small value (i.e., 1e-10). To evaluate the performance of the solver, we propose two indicators, namely %Gap and %Gap^n. The gap (%Gap) is determined based on the upper and lower bounds (UB and LB) obtained by Baron solver. To calculate the %Gap^n, we use an NLP (non-linear programming) solver, namely IPOPT, to solve the ENGP energy mix model without including the binary variables L_r (a pure NLP model). In other words, Constraints (9)–(11) and (21) are removed from the model, which is relatively easier to solve than the original ENGP energy mix model. The solutions obtained by this solver can also be used for the lower bound (LB^n) to assess the quality of solutions produced by Baron solver. Then, %Gap^n is computed based on the following expression:

$$\%\text{Gap}^n = \frac{\text{UB} - \text{LB}^n}{\text{UB}} \times 100 \tag{52}$$

Table 9 presents the summary of experimental results of the goal programming-based energy mix model consisting of the 12 instances generated and which are based on the demand and sub-goal settings. For each instance, the %Gap, %Gap^n, and CPU time are provided. According to the computational experiments, IPOPT solver can optimally solve the pure NLP model for all the instances. Therefore, the solutions produced by IPOPT solver can be used as good lower bounds. When solving the MINLP, Baron produces optimal solutions for all instances, except Instance IC-Avg. However, the %Gap produced by Baron for this instance is very low (0.0003%). The average %Gap of 0.000% is also obtained by Baron. Compared to solutions attained by IPOPT in solving the pure NLP model, Baron yields a small average %Gap^n of 0.147%. This indicates that the overall Baron's performance in solving the MINLP model is very good, and its solutions can be used as an input to the second model.

Table 9: The experimental results of the first Model

Instance	The performance of Baron				Energy Mix at National Level (%)						
	UB	%Gap	%Gap^n	CPU(s)	Coal	Gas	Hydro	Nuclear	Wind	Solar	Biomass
The Energy Mix Target for the BL setting					30.70	11.20	14.80	11.00	13.00	15.60	3.70
BL-Avg	0.326	0.000	0.305	406	28.87	11.20	17.84	11.00	13.00	14.17	3.91
BL-Avg-Env	0.940	0.000	0.083	14743	7.99	11.20	34.47	10.62	18.48	13.54	3.70
BL-Avg-Cost	0.589	0.000	0.165	2995	30.70	11.66	20.93	11.00	8.70	13.31	3.70
BL-Avg-Soc	0.756	0.000	0.112	7928	7.34	11.20	39.58	11.00	13.00	14.17	3.70
The Energy Mix Target for the BC setting					8.70	7.10	15.40	13.80	19.20	27.70	8.10
BC-Avg	0.414	0.000	0.232	184	10.57	15.33	18.43	13.55	19.20	14.06	8.87
BC-Avg-Env	0.936	0.000	0.087	27667	8.70	7.10	33.11	10.54	20.46	13.08	7.00
BC-Avg-Cost	0.732	0.000	0.151	5501	12.98	16.20	24.93	13.55	10.55	13.70	8.10
BC-Avg-Soc	0.764	0.000	0.099	1226	8.05	7.10	30.35	13.29	20.82	14.06	6.33
The Energy Mix Target for the IC setting					20.00	8.00	14.00	11.00	19.00	23.00	5.00
IC-Avg	0.418	0.000	0.217	86400	20.00	10.99	19.47	11.00	19.00	12.18	7.36
IC-Avg-Env	1.018	0.000	0.092	6078	8.31	8.00	36.89	11.00	19.00	11.79	5.00
IC-Avg-Cost	0.713	0.000	0.115	27739	20.00	15.40	23.71	11.00	12.71	12.18	5.00
IC-Avg-Soc	0.848	0.000	0.101	5347	12.99	8.00	32.65	11.00	19.00	12.18	4.18
Average		0.000	0.147	15518							

Avg: the target setting using the average assessment value of the sub-goal
Avg-Env, Avg-Cost and Avg-Soc: the target settings which emphasize environmental, economic, and social dimensions, respectively.

For each instance, Table 9 also presents the single-period energy mix at national level in 2050 for each demand setting (i.e., the baseline (BL), beautiful China (BC), and intelligence connected (IC)) and for each sub-goal setting (i.e., Avg, Avg-Env, Avg-Cost, and Avg-Soc). From the table, it is noticed that when the Avg setting is used, the electricity production for each generation source tends to meet the target that has been set. As expected, when the cost dimension is prioritized (Avg-Cost), the fossil generations (coal and gas) are mainly used to cover the energy demand. The wind and solar generations are not significantly built. In contrast, the use of non-fossil generations, especially hydro, is dominant when prioritizing either environmental (Avg-Env) or social (Avg-Soc) dimensions. In this case, the electricity produced by wind and solar generations also increases.

The energy transmitted for one region to others is presented in Figure 4, whereas the energy mix at regional level for each instance is provided in Figure B.8 in Appendix. We can notice that South, Northwest, and Northeast regions have surplus production, whereas East, North, and Central regions need to import energy supply from other regions. In general, the flow of electricity transmission occurs as follows: from the South to Central and East; from Northwest to North and Central; from Northeast to North and East. We also observe interesting results regarding the hydro and solar power plants. The energy production generated by hydro exceeds the target that has been set for all instances. This is mainly because hydro has a smaller operating cost compared to other renewable energy types. The hydro generations built are mainly located in South and Central China Regions. In contrast, the energy production of solar power plants does not meet the target for all instances. This is mainly because of its maximum capacity and the high investment/operating costs.

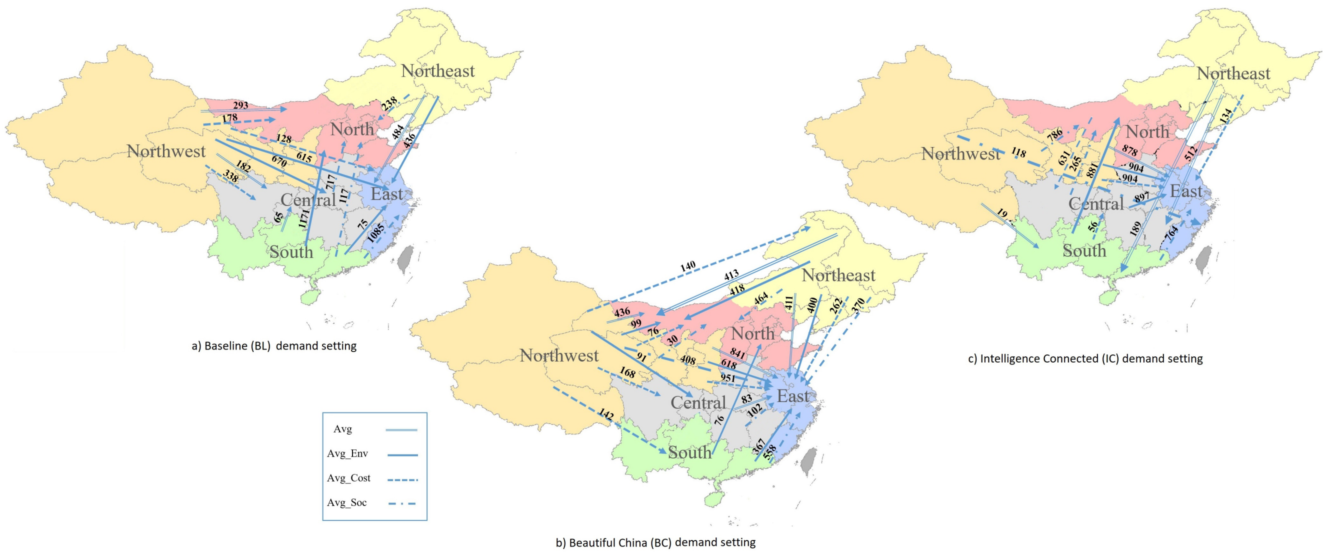

Figure 4: The flow of electricity transmission for Different Demand Settings

4.2. Results of the second model

In this subsection, we present the results of computational experiments of the second model (the multiple-period Generation Expansion Planning model). The proposed SAA algorithm was coded in C++ .Net 2019, where the IBM ILOG CPLEX version 20.10 Concert Library was also used to solve the SAA models with an exact method. We use the same workstation used for the experiments conducted in Subsection 4.1. In the SAA method, we vary the sample size (N) and set it to 50, 100, 200. The number of samples (M) is set to 10, whereas the large sample size (N') is assigned to 10,000. A 95% confidence interval is used to estimate the statistical gap (%). As previously discussed, the main input for the second model is the output generated by the first model, which is the energy mix for national and regional levels in 2050.

Table 10 presents the summary of experimental results on the second model, where %Gap for each instance is presented based on the sample size used. The table also reveals the computational time (CPU) needed to solve the problem. In the table, we also provide %Dev^c and %Dev^e, where the former represents the deviation of the total cost obtained using different sub-goal settings compared to the setting that yields the best total cost for each demand setting. We also recorded the total amount of CO_2 produced by the solution, generated by solving each instance. Here, the environmental performance of implementing sub-goal settings is measured by %Dev^e, indicating the deviation of the total amount of CO_2 produced using different sub-goal settings compared to the setting that yields the smallest total CO_2 for each demand setting. $\%\text{Dev}^c$ and $\%\text{Dev}^e$ are calculated based on the following expression:

$$\%\text{Dev}^c = \frac{\text{UB}_i - \text{UB}^b}{\text{UB}_i} \times 100; \qquad \%\text{Dev}^e = \frac{\epsilon_i - \epsilon^b}{\epsilon_i} \times 100 \tag{53}$$

where UB_i and ϵ_i are the upper bound of the total cost and the amount of CO_2 produced by solving instance i, whereas UB^b and ϵ^b are the best upper bound and the amount of CO_2 among the instances for each demand setting (i.e., BL, BC, and IC).

Table 10: The experimental results of the second model

Instance	$N = 50$				$N = 100$				$N = 200$			
	%Gap	CPU	$\%\text{Dev}^c$	$\%\text{Dev}^e$	%Gap	CPU	$\%\text{Dev}^c$	$\%\text{Dev}^e$	%Gap	CPU	$\%\text{Dev}^c$	$\%\text{Dev}^e$
BL-Avg	1.126	217	1.72	55.84	0.622	347	1.76	55.17	0.443	1029	1.75	55.43
BL-Avg-Env	1.541	216	21.85	0.66	1.255	365	21.77	0.71	1.006	1074	21.75	1.00
BL-Avg-Cost	1.149	264	0.00	59.76	0.644	418	0.00	59.17	0.474	1193	0.00	59.43
BL-Avg-Soc	0.783	280	21.72	0.00	0.437	387	21.74	0.00	0.265	975	21.77	0.00
BC-Avg	0.992	222	1.69	24.33	0.472	363	1.66	24.85	0.317	1162	1.65	25.01
BC-Avg-Env	0.848	245	2.29	0.00	0.433	379	2.28	0.00	0.277	1044	2.26	0.00
BC-Avg-Cost	0.893	272	0.00	25.57	0.468	410	0.00	25.95	0.312	1129	0.00	26.08
BC-Avg-Soc	0.878	304	4.44	0.09	0.450	424	4.42	0.34	0.301	1108	4.41	0.28
IC-Avg	1.133	343	4.38	33.41	0.615	474	4.37	33.49	0.365	1217	4.38	33.59
IC-Avg-Env	1.021	381	14.39	0.00	0.558	474	14.43	0.00	0.333	1037	14.42	0.00
IC-Avg-Cost	1.220	439	0.00	36.12	0.648	551	0.00	35.21	0.384	1261	0.00	34.83
IC-Avg-Soc	1.043	489	9.04	8.16	0.601	596	9.12	7.90	0.359	1252	9.10	7.95
Average	1.052	306			0.600	432			0.403	1124		

Avg: the target setting using the average assessment value of the sub-goal

Avg-Env, Avg-Cost and Avg-Soc: the target settings which emphasize environmental, economic, and social dimensions, respectively.

According to Table 10, %Gap is decreasing with the sample size, meaning that the larger sample size produces better %Gap at the expense of a longer computing time. The average %Gap obtained by $N = 200$ (0.403%) is significantly lower than that obtained by $N = 50$ (1.502%). In general, the proposed SAA method yields a relatively small gap. Moreover, the results among instances are also consistent with a small standard deviation. The CPU time required to solve the problem is acceptable as the average CPU time are 306, 432, and 1124 seconds when $N = 50, 100, 200$ respectively. Figure 5 presents the upper and lower bounds obtained by different sample sizes for different sub-goal settings in the BL demand setting. Here, the UB and LB values are determined based on the total costs of six periods (2025, 2030, 2035, 2040, 2045, and 2050) to represent all the periods in the planning horizon. The figure indicates that increasing the sample size consistently reduces the %Gap and improves the values of upper bound (UB) and lower bound (LB), leading to better solutions.

According to Table 10, as expected, instances with cost priority (Avg-Cost) yield the smallest

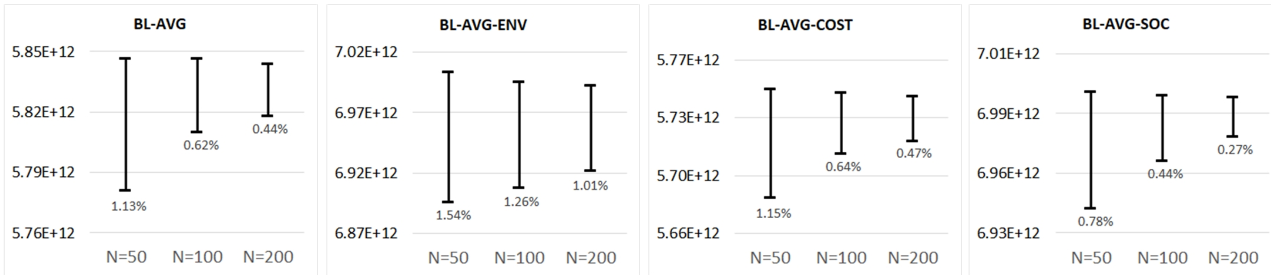

Figure 5: The upper and lower bounds for the BL demand setting, obtained using different sample sizes

total cost for all demand settings (BL, BC and IC). When $N = 200$, compared to the Avg setting, implementing cost priority reduces the total cost by an average of 1.75%, 1.65%, and 4.36% for the BL, BC, and IC demand settings, respectively. On the other hand, the total cost of applying the Avg-Env setting significantly increases the total cost by an average of 21.77%, 2.28%, and 14.43% for BL, BC, and IC demand settings respectively. The use of non-fossil generations reduces the amount of CO_2 produced which can be achieved by implementing the Avg-Env and Avg-Soc. When $N = 200$, the implementation of the Avg-Env yields the smallest amount of CO_2 produced for the BC and IC demand settings. Compared to the ideal conditions (smallest CO_2 produced), implementing the Avg-Cost (minimizing cost) increases the amount of CO_2 by 55.43%, 25.01%, and 33.59% for the BL, BC and IC demand settings, respectively. In the BL demand setting, implementing the Avg-Env or Avg-Soc (priority on Environment or Social) yields a significantly smaller amount of CO_2 than implementing the Avg-Cost. The BC demand setting has the smallest deviation between those implementations. This is most likely because the production target for fossil power plants (coal and gas) has already been set to a low level in this setting. The deviation in the IC setting is slightly higher than that in the BC demand setting. In the IC setting, more coal and gas plants are needed to satisfy high energy demand, as the capacity of non-fossil power plants is limited. It is clear that there is a trade-off between economic and environmental issues. Here, the decision makers are required to determine the solutions they prefer.

Figure 6 presents the percentage of energy generated by non-fossil fueled power plants (i.e., Hydro, Nuclear, Wind, Solar, and Biomass) for each instance, which is obtained by using the SAA method with $N = 200$. As seen in the figure, for all instances the percentage of energy produced by non-fossil power plants increases significantly at the beginning of the period. However, the increment rate will gradually decrease because of the capacity limitation. Still, they will be become the main source at the end of period, especially when implementing the Avg-Env and Avg-Soc setting. To minimize the total cost, the fossil-fueled generations are mainly used at the beginning of the period. The renewable energy power plants are then gradually built and their electricity production steadily increases.

Figure B.9 in the Appendix shows the national energy mix for 2025 to 2050, generated by solving each instance using the SAA approach with $N = 200$. The figure reveals that the energy production generated by the coal power plants decreases significantly over the planning horizon when implementing Avg-Env and Avg-Soc (priority on environment and social aspects). At the beginning of the period,

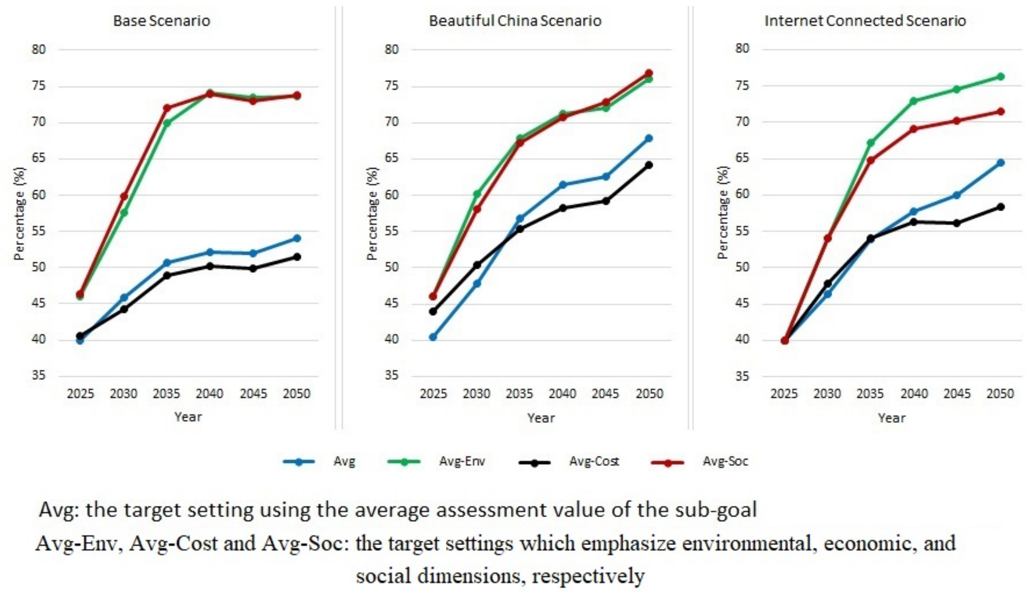

Figure 6: The percentage of energy produced by renewable energy for each instance using $N = 300$

the coal power plants are the main energy producers, however they are gradually reduced and replaced by hydro power plants. However, when Avg and Avg-cost settings are implemented, the amount of energy produced by coal power plants is not significantly decreased till the end of the period, as it is considered a cheap energy producer. In all instances, the use of natural gas is increasing, as it produces lower CO_2 emissions, although it is still considered a fossil-fueled power plant. The figure also shows that the use of renewable energy (hydro, solar, wind, and biomass) and nuclear gradually increases for all instances. At the end of period, the renewable energy will be the main source of energy production.

Figure B.10 in the Appendix presents the national capacity mix from 2025 to 2050 for all instances. Similar to the previous figure, it is also generated by the SAA method using $N = 200$. The figure shows that the decreasing capacity over the planning horizon is for coal power plants only. A substantial decrease occurs when the Avg-Env or Avg-Soc sub-goal setting is implemented. The capacity of other generation sources rises, especially for non-fossil fueled power plants. It is interesting to note that there is a significant increase in the solar generation capacity and it will be the largest one at the end of the period. However, as the capacity factor of this generation source is relatively low, its energy production is also relatively low. A significant capacity growth also occurs for the wind power plants, especially near the end of the period.

We also carry out an analysis to assess the quality of two benchmark solutions in a stochastic environment, when the uncertain parameters are considered. The first benchmark solution is obtained by solving the deterministic version of the second model using an exact method (CPLEX), where the average values are used for the uncertain parameters (annual demand, peak load, and fuel cost). The second benchmark solution is generated by solving the original two-stage stochastic programming problem, expressed by Model (22)–(43), using an exact method (CPLEX), with a high number of scenarios ($|\Omega|$) (e.g., $|\Omega| = 2000$, 3000, and 4000). Note that the scenarios are generated using the Monte-Carlo method used for the SAA method and the probability of each scenario ($\hat{p}_\omega$) is equally set. From these two benchmark solutions, we take the first-stage solution ($\hat{\mathbf{X}}$), consisting of variables $\hat{X}_{krt}$ and $\hat{Y}_{krt}$, and execute the second stage of the SAA method (Line 11–15 of Algorithm 1). Parameter N' is set to 10000, which is similar to previous experiments. Instance BL-Avg-Env is used in this analysis.

Figure 7 presents the histograms of the total costs (6 periods) from the different solutions in the stochastic environment, where the UB value is calculated based on Equation (50). The figure reveals that the deterministic solution yields a higher average total cost than those generated by solving the stochastic model using either the SAA method or the exact method on the original model. In addition, the stochastic solutions also produce a much smaller standard deviation. The figure also shows that the solution obtained by the SAA with $N = 200$ generates the smallest UB value, average cost and standard deviation, with an acceptable computing time (1074 seconds). Solving the original two-stage stochastic model with $|\Omega| = 4000$ requires 232,309 seconds to complete, with the obtained total cost slightly worse than the one obtained by the SAA method. When $|\Omega|$ is set to 2000 and 3000, CPLEX needs respectively 29,157 and 64,307 seconds to finish, with slightly higher total costs. Here, the CPU time increases significantly with the number of scenarios ($|\Omega|$). In general, the proposed SAA method produces more robust solutions than the other solutions.

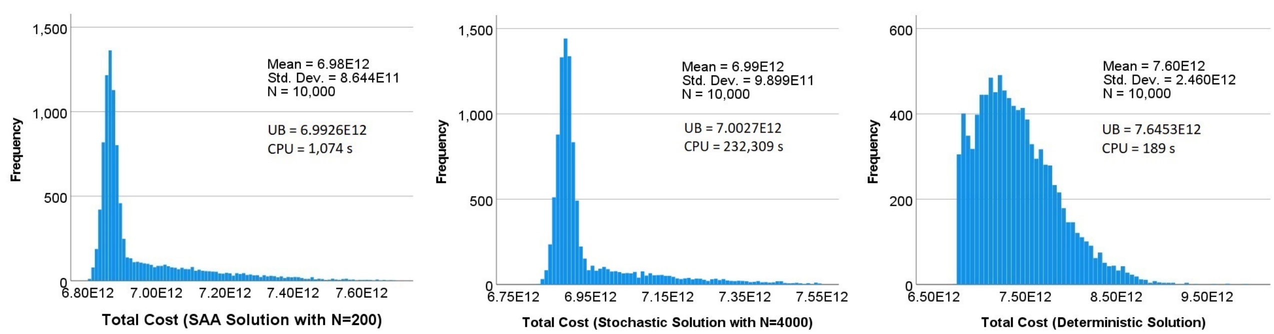

Figure 7: The total cost of different solutions in the stochastic environment for Instance BL-Avg-Env

4.3. Managerial and policy implications

The results and findings of our integrated optimization models generate insightful implications for relevant government agencies and policymakers. Firstly, our models and analysis help policymakers to make better informed decisions for selecting the optimal energy mix generated from typical fossil fuel-based and renewable energy power generation to meet the rising energy demand. We also provide guidance on the corresponding capacity planning given the optimal energy mix, i.e., whether and when to demolish existing power plants, scale up or down the capacity of existing plants, or invest in new plants, and the magnitude of such capacity changes. Secondly, our models account for multiple criteria (i.e., economic, environmental, and social dimensions), the targets set by different levels of stakeholders, and the uncertainty of critical factors (e.g., regional annual energy demand, fuel cost). Based on the models, appropriate sensitivity analysis can be conducted to enable policymakers to evaluate what-if scenarios and identify the optimal energy mix configuration and the energy generation planning under various conditions. For example, the impact of the capacity factor for renewable energy on the energy mix and the minimal cost to satisfy the projected energy demand can be investigated. In this way, some quantitative analysis can be derived to understand the cost reduction effects by increasing the capacity factor for renewable energy, which facilitates the cost-benefits analysis for investing in technologies to enhance the capacity factor for renewable energy generation facilities. Likewise, policymakers can also explore the impact of the target of sub-goals set by different regions since regions may shift their prioritized goals during development. By committing to environmental related sub-goals at the national level, policymakers may use positive or negative incentives to shape the sub-goal target set by different regions to best meet the requirements at the national level. The choices cities and regions make for pilot reforms concerning energy generation planning (e.g., prioritizing the environmental dimension) can, therefore, be better planned. Thirdly, our results suggest the pattern of inter-regional energy transmission under different circumstances. This serves policymakers with meaningful information on the direction and scale of regional transmission and, therefore, helps them to plan investment in building transmission lines between regions. In countries like China where different power corporations supply electricity for different regions, analysis on the inter-regional energy transmission also facilitates power trade planning. Regions with insufficient capacity may sign mid- or long-term contracts with regions with surplus capacity strategically to meet the regional energy demand at a reasonable price, which contributes to avoidance of the electricity price instability.

5. Conclusion, limitations and future research directions

This paper investigates an integrated strategic energy mix and energy generation planning with multiple sustainability criteria and consideration of stakeholder demands at national and regional levels. A novel combination of two optimization models is proposed, the first is built based on a non-linear extended network goal programming model to determine energy mix at a future period, with consideration of economic, environmental, and social sustainability factors, and targets set by stakeholders at both national and regional level. The second model is developed based on a stochastic multi-period generation expansion planning model to determine the energy transition over a time horizon of 30 years. The results obtained from the first model are used as the main input for the second.

We assess the practicality of the proposed models using data on the power system in China and model a transition towards a more sustainable electricity system. The first model is solved using a commercial non-linear solver (Baron), whereas the second model is addressed by an algorithm based on the sample approximation approach (SAA), where a linear solver (CPLEX) is used. Our model results cast light on pressing issues regarding the energy mix configuration from typical fossil fuel-based and renewable energy generation sources to meet the mounting demand for energy sustainably and the corresponding capacity planning in response to such an energy transition, as well as providing insights on future demands for inter-regional power transmission pattern. We also summarize the significant implications from our model results for policymakers to better plan energy generation and expansion under different conditions. Furthermore, the methodology is designed to provide a plan for a strategic energy transition over an extended planning period, if political, social, environmental, economic or technical conditions significantly change over that period then it is possible to re-run the planning model with new goals and/or data in order to provide adjusted plans. For example, recent debates and adjustments in regarding carbon reduction goals in the medium and long term can be taken into account in the models.

Our research and approach has a number of limitations that also provide potential avenues for further research. First of all our approach of combining two optimization models for long term energy planning can be further advanced. This is important both academically and practically as countries increasing need to consider the long term implications on climate change of their decisions on power generation. One area where the models can be improved is in terms of the assumptions related to the efficiency of the various options for power generation. Technological development for various energy options show that the efficiency of renewable energy technology such as solar and wind energy have advanced more rapidly in the past decade than more traditional options based on fossil fuel combustion. Models could use extend past learning curves for the various energy options to improve the prediction of future efficiency levels. Furthermore similar methods could be used to look at the dynamics of fixed and variable costs. Other parameters that can be further refined include various sustainability criteria and stakeholder demands. Moreover given the current debate on carbon reduction goals in the coming decades a more stringent set of carbon reduction goals could be used within the models and potentially over an even longer time period. Also further scenarios could be explored based on centralized versus distributed energy system strategies and various pathways towards low-carbon transitions such as discussed in other studies (Hofman and Elzen, 2010; Geels et al., 2020).

As we apply the models for China, the proposed methodology can also be applied for other nations with a national-regional energy generation structure, particularly those with large populations and imbalances in power generation and consumption. In the proposed models, stakeholders comprise of two layers, namely national and regional levels. The model can be extended by considering sub-regional levels so that three layers of stakeholders are taken into account. Moreover, it is worth performing further investigations on the effect of centralized versus decentralized energy strategies and taking into account efficiency improvement in energy technologies in the proposed methodology.

Acknowledgments

The authors would like to thank all the referees for their constructive comments which improve both the presentation as well as the content of the paper.

Appendix A. Glossary

ACC-FUT	Severe accidents perceived in future
ACC-PAST	Fatal accidents from past experience
AHP	Analytical Hierarchy Process
AVAILAB	Average availability (load) factor
Avg	Average Setting
Avg-Cost	the target settings which emphasize economic dimension
Avg-Env	the target settings which emphasize environmental dimension
Avg-Soc	the target settings which emphasize social dimension
BC	Beautiful China Energy Demand Scenario
BL	Baseline Energy Demand Scenario
CEPY	China Electric Power Yearbook
CNPC	China National Petroleum Corporation
CO2eq.	GHG emissions
EMPL	Technology-specific job opportunities
ENGP	Extended Network Goal Programming
ENV	Environmental external costs
ETRI	Economics and Technology Research Institute
FOOD	Food safety risk
GEP	Generation Expansion Planning
GHG	Greenhouse Gas
GRID-COST	Cost of grid connection
HEALTH	Human health impact
IC	Intelligence Connected Energy Demand Scenario
LB	Lower bound
MINLENGP	mixed integer non-linear programming extended network goal programme
MMBtu	Million British thermal unit
MWh	Megawatt hours
MWh	Megawatt
NLP	Non-Linear Programming
Pctile	Percentile
PEAKLOAD	Peak load response
PR-COST	Private costs (investments and operation costs)
RADIO	Radionuclide external costs
SAA	Sample Average Approximation
SECURE	Security of supply
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
UB	Upper bound

Appendix B. Charts of the Computational Results

Figure B.8: The regional energy mix generated by solving the non-linear ENGP energy model

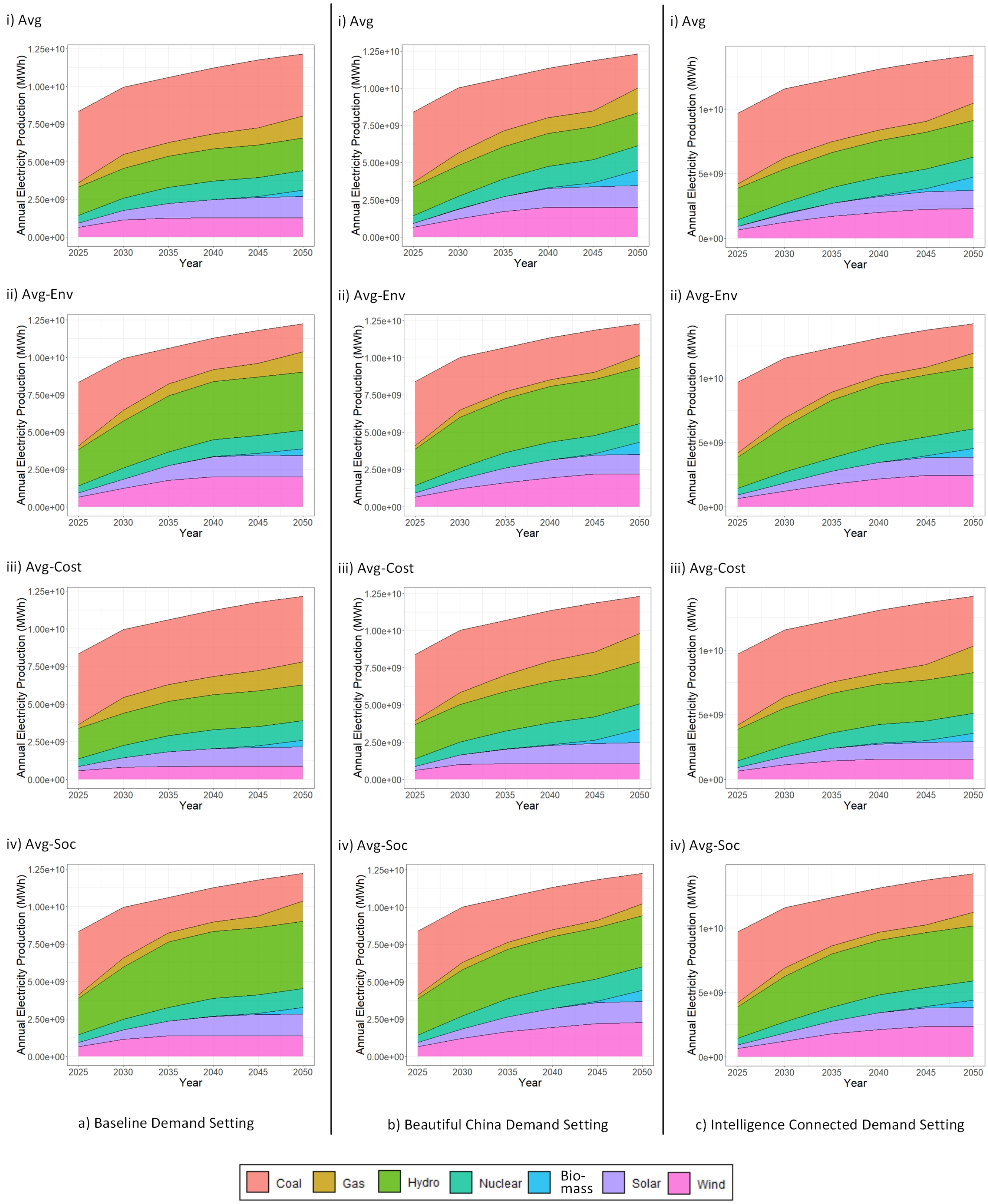

Figure B.9: The national energy mix for each instance using the SAA method with $N = 300$

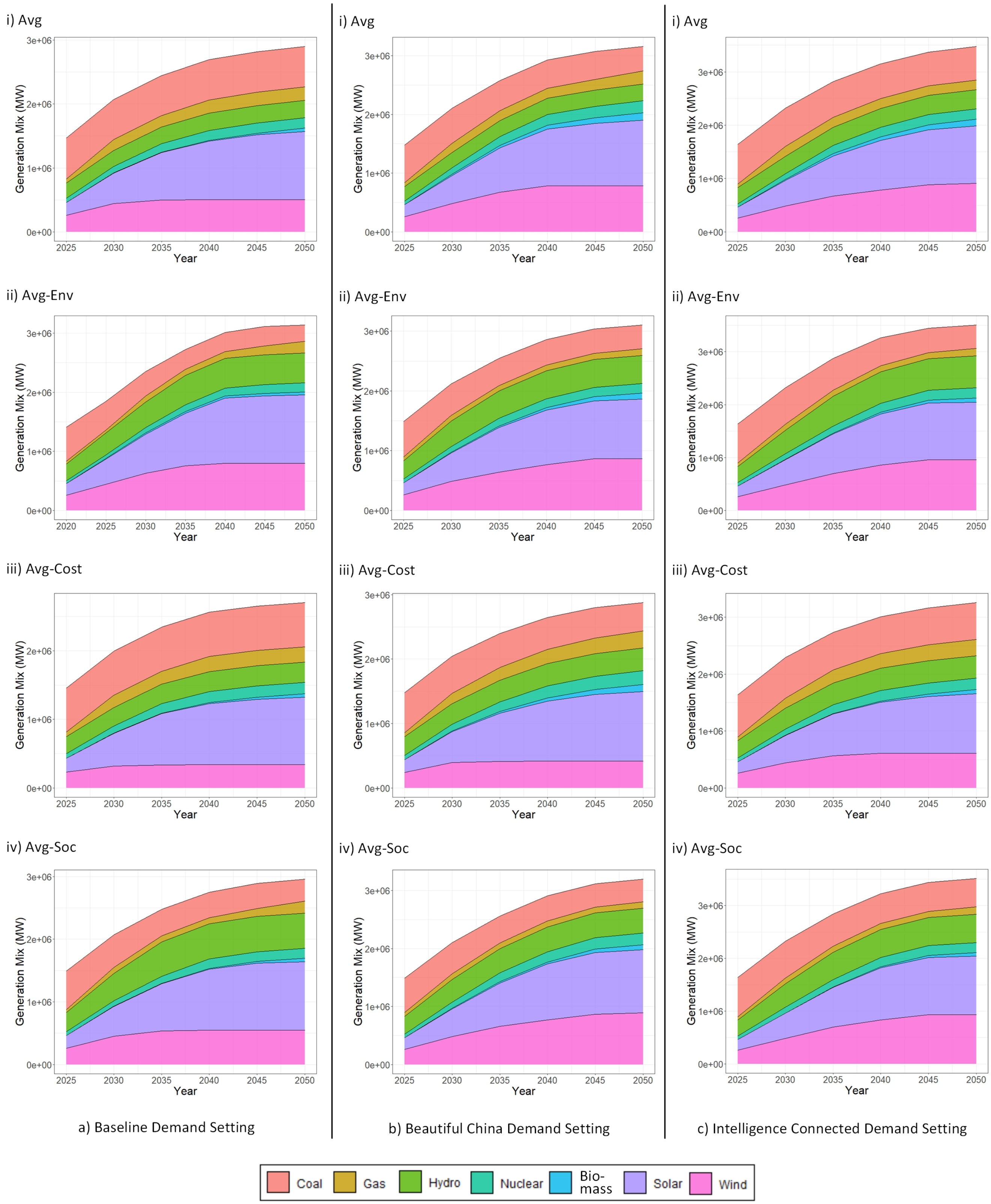

Figure B.10: The national capacity mix for each instance using the SAA method $N = 300$

References

Bertsimas, D., Gupta, V., Kallus, N., 2018. Robust sample average approximation. Mathematical Programming 171, 217–282.

Betancourt-Torcat, A., Almansoori, A., 2015. Design multiperiod optimization model for the electricity sector under uncertainty – a case study of the emirate of abu dhabi. Energy Conversion and Management 100, 177 – 190.

Birge, J.R., Louveaux, F., 2011. Introduction to stochastic programming. Springer Science & Business Media.

CEPY, 2018. 2018 China Electric Power Yearbook. China Electric Power Press.

CNPC ETRI, 2019. World and China Energy Outlook 2019. URL: https://doi.org/10.13140/RG.2.2.23399.88482.

Costa, O.L., de Oliveira Ribeiro, C., Rego, E.E., Stern, J.M., Parente, V., Kileber, S., 2017. Robust portfolio optimization for electricity planning: An application based on the brazilian electricity mix. Energy Economics 64, 158 – 169.

Dagoumas, A.S., Koltsaklis, N.E., 2019. Review of models for integrating renewable energy in the generation expansion planning. Applied Energy 242, 1573–1587.

Energy Information Administration, 2020. Capital Cost and Performance Characteristic Estimates for Utility Scale Electric Power Generating Technologies. Technical Report.

Energy Information Administration, 2022. World Energy Statistics and Balances. Technical Report.

Feng, Y., Ryan, S.M., 2013. Scenario construction and reduction applied to stochastic power generation expansion planning. Computers & Operations Research 40, 9 – 23.

Geels, F., McMeekin, A., Pfluger, B., 2020. Socio-technical scenarios as a methodological tool to explore social and political feasibility in low-carbon transitions: Bridging computer models and the multi-level perspective in uk electricity generation (2010–2050). Technological Forecasting and Social Change 151, 119258.

Global Energy Statistical Yearbook, 2021. World Energy Consumption Statistics. Enerdata. URL: https://yearbook.enerdata.net/total-energy/world-consumption-statistics.html.

Hofman, P.S., Elzen, B., 2010. Exploring system innovation in the electricity system through sociotechnical scenarios. Technology Analysis & Strategic Management 22, 653–670.

Hu, Q., Huang, G.H., Cai, Y.P., Sun, W., 2014. Planning of electric power generation systems under multiple uncertainties and constraint-violation levels. Journal of Environmental informatics 23, 55–64.

Ioannou, A., Fuzuli, G., Brennan, F., Yudha, S.W., Angus, A., 2019. Multi-stage stochastic optimization framework for power generation system planning integrating hybrid uncertainty modelling. Energy Economics 80, 760–776.

Irawan, C.A., Hofman, P.S., Chan, H.K., Paulraj, A., 2022. A stochastic programming model for an energy planning problem: formulation, solution method and application. Annals of Operations Research 311, 695–730.

Jones, D., Florentino, H., Cantane, D., Oliveira, R., 2016. An extended goal programming methodology for analysis of a network encompassing multiple objectives and stakeholders. European Journal of Operational Research 255, 845 – 855.

Kim, D., Ryu, H., Lee, J., Kim, K.K., 2021. Balancing risk: Generation expansion planning under climate mitigation scenarios. European Journal of Operational Research doi:https://doi.org/10.1016/j.ejor.2021.06.001.

Kim, S., Pasupathy, R., Henderson, S.G., 2015. A Guide to Sample Average Approximation. Springer New York, New York, NY. pp. 207–243.

Kleywegt, A.J., Shapiro, A., Homem-de Mello, T., 2002. The sample average approximation method for stochastic discrete optimization. SIAM Journal on Optimization 12, 479–502.

Koltsaklis, N.E., Dagoumas, A.S., 2018. State-of-the-art generation expansion planning: A review. Applied Energy 230, 563–589.

Krukanont, P., Tezuka, T., 2007. Implications of capacity expansion under uncertainty and value of information: The near-term energy planning of japan. Energy 32, 1809 – 1824.

Kutani, I., Motokura, M., Okubo, N., 2016. Cost Assessment of Energy Security Improvement in East Asia Summit Region. Technical Report. ERIA. URL: https://www.eria.org/publications/cost-assessment-of-energy-security-improvement-in-east-asia-summit-region/.

Li, G., Huang, G., Sun, W., Ding, X., 2014. An inexact optimization model for energy-environment systems management in the mixed fuzzy, dual-interval and stochastic environment. Renewable Energy 64, 153 – 163.

Li, Y., Huang, G., 2012. Electric-power systems planning and greenhouse-gas emission management under uncertainty. Energy Conversion and Management 57, 173 – 182.

Li, Y., Huang, G., Li, Y., Xu, Y., Chen, W., 2010. Regional-scale electric power system planning under uncertainty - a multistage interval-stochastic integer linear programming approach. Energy Policy 38, 475 – 490.

Lin, J., Kahrl, F., Yuan, J., Chen, Q., Liu, X., 2019. Economic and carbon emission impacts of electricity market transition in china: A case study of guangdong province. Applied Energy 238, 1093–1107.

Lugovoy, O., Gao, S., Gao, J., Jiang, K., 2021. Feasibility study of china's electric power sector transition to zero emissions by 2050. Energy Economics 96, 105176.

Malcolm, S.A., Zenios, S.A., 1994. Robust optimization for power systems capacity expansion under uncertainty. The Journal of the Operational Research Society 45, 1040–1049.

Mallapaty, S., 2020. How china could be carbon neutral by mid-century. Nature 586, 482–483. doi:10.1038/d41586-020-02927-9.

Min, D., Chung, J., 2013. Evaluation of the long-term power generation mix: The case study of south korea's energy policy. Energy Policy 62, 1544 – 1552. URL: http://www.sciencedirect.com/science/article/pii/S0301421513007477, doi:https://doi.org/10.1016/j.enpol.2013.07.104.

Moreira, A., Pozo, D., Street, A., Sauma, E., Strbac, G., 2021. Climate-aware generation and transmission expansion planning: A three-stage robust optimization approach. European Journal of Operational Research 295, 1099–1118.

Moret, S., Babonneau, F., Bierlaire, M., Maréchal, F., 2020. Decision support for strategic energy planning: A robust optimization framework. European Journal of Operational Research 280, 539–554.

Mulvey, J.M., Vanderbei, R.J., Zenios, S.A., 1995. Robust optimization of large-scale systems. Operations Research 43, 264–281.

National Bureau of Statistics of China, 2020. China Statistical Yearbook. China Statistics Press. URL: http://www.stats.gov.cn/tjsj/ndsj/.

OECD Nuclear Energy Agency, 2016. Costs of Decommissioning Nuclear Power Plants. Technical Report NEA No. 7201.

Özcan, E.C., Ünlüsoy, S., Eren, T., 2017. A combined goal programming – ahp approach supported with topsis for maintenance strategy selection in hydroelectric power plants. Renewable and Sustainable Energy Reviews 78, 1410–1423. URL: https://www.sciencedirect.com/science/article/pii/S1364032117305440, doi:https://doi.org/10.1016/j.rser.2017.04.039.

Park, H., Baldick, R., 2015. Stochastic generation capacity expansion planning reducing greenhouse gas emissions. IEEE Transactions on Power Systems 30, 1026–1034.

Rentizelas, A.A., Tolis, A.I., Tatsiopoulos, I.P., 2012. Investment planning in electricity production under co2 price uncertainty. International Journal of Production Economics 140, 622–629.

Sadeghi, H., Rashidinejad, M., Abdollahi, A., 2017. A comprehensive sequential review study through the generation expansion planning. Renewable and Sustainable Energy Reviews 67, 1369–1394.

San Cristóbal, J.R., 2012. A goal programming model for the optimal mix and location of renewable energy plants in the north of spain. Renewable and Sustainable Energy Reviews 16, 4461–4464.

Shapiro, A., 2003. Monte Carlo Sampling Methods. Monte Carlo Sampling Methods. Elsevier, pp. 353-425.

Shapiro, A., Dentcheva, D., Ruszczynski, A., 2009. Lectures on stochastic programming: modeling and theory, MOS-SIAM series on optimization. Technical Report. Society for Industrial and Applied Mathematics.

Sorgulu, F., Dincer, I., 2018. Cost evaluation of two potential nuclear power plants for hydrogen production. International Journal of Hydrogen Energy 43, 10522–10529.

Streimikiene, D., Balezentis, T., Krisciukaitiene, I., Balezentis, A., 2012. Prioritizing sustainable electricity production technologies: Mcdm approach. Renewable and Sustainable Energy Reviews 16, 3302 – 3311.

Thangavelu, S.R., Khambadkone, A.M., Karimi, I.A., 2015. Long-term optimal energy mix planning towards high energy security and low ghg emission. Applied Energy 154, 959–969.

Topham, E., McMillan, D., 2017. Sustainable decommissioning of an offshore wind farm. Renewable Energy 102, 470–480.

Yi, B.W., Xu, J.H., Fan, Y., 2016. Inter-regional power grid planning up to 2030 in china considering renewable energy development and regional pollutant control: A multi-region bottom-up optimization model. Applied Energy 184, 641–658.

Yu, S., Zheng, Y., Li, L., Wang, K., 2020. Ranking provincial power generation sources of china: a decision-maker preferences based integrated multi-criteria framework. Environmental Science and Pollution Research 27, 36391–36410.

Zhang, Y., Ma, T., Guo, F., 2018. A multi-regional energy transport and structure model for china's electricity system. Energy 161, 907–919.