Overcapacity-enabled seasonal flexibility benefits energy-intensive industry and decarbonizing energy systems

Ruike Lyu[a,b], Anna Li[a], Jianxiao Wang[c], Hongxi Luo[a], Yan Shen[b], Hongye Guo[b], Ershun Du[b], Chongqing Kang[b], Jesse Jenkins*[a]

[a] *Princeton University, Princeton, 08540, US*
[b] *Tsinghua University, Beijing, 100084, China*
[c] *Peking University, Beijing, 100084, China*

Preprint submitted to Nature Energy *November 18, 2025*

Contents

1. Supplementary Figures and Tables

Table 1: Aluminum Demand Projection (unit: 10,000 t/year).

Scenario	Year	Total Demand	Domestic Scrap	Imported Scrap	Recycled Aluminum Production	Primary Aluminum Demand	Overcapacity Rate (%)
Low Demand	2030	3706	2177	218	1796	1910	58
Low Demand	2035	3847	3179	318	2623	1224	73
Low Demand	2040	3947	3772	377	3112	835	81
Low Demand	2045	3719	3827	0	2870	849	81
Low Demand	2050	3440	3749	0	2812	628	86
Low Demand	2055	3390	3880	0	2910	480	89
Low Demand	2060	3222	3884	0	2913	309	93
Mid Demand	2030	4873	2559	256	1971	2902	36
Mid Demand	2035	5028	3876	388	2984	2044	55
Mid Demand	2040	5210	4808	481	3702	1508	66
Mid Demand	2045	4881	4915	0	3440	1440	68
Mid Demand	2050	4642	4965	0	3476	1167	74
Mid Demand	2055	4593	5153	0	3607	986	78
Mid Demand	2060	4224	4997	0	3498	726	84
High Demand	2030	4953	2812	0	1828	3125	31
High Demand	2035	5770	3989	0	2593	3176	29
High Demand	2040	5871	4527	0	2943	2928	35
High Demand	2045	5941	5063	0	3291	2650	41
High Demand	2050	6136	5726	0	3722	2414	46
High Demand	2055	5837	5880	0	3822	2014	55
High Demand	2060	5513	5971	0	3881	1632	64

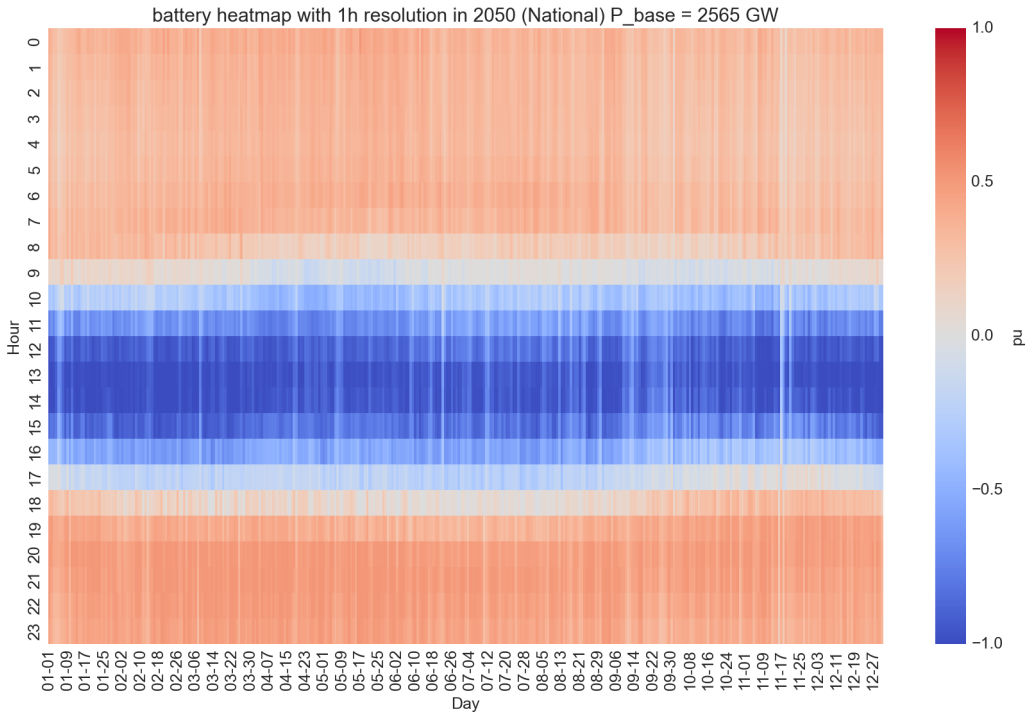

Figure 1: Heatmap of battery operation (2050 net zero; core scenario). Positive numbers for discharging (red); negative numbers for charging (blue).

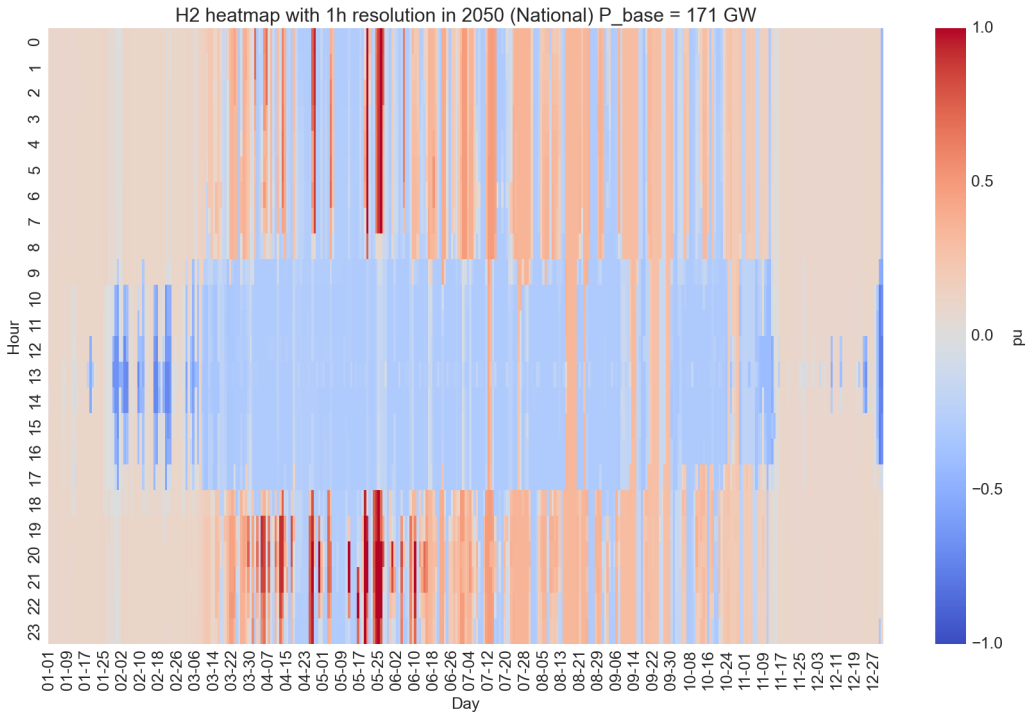

Figure 2: Heatmap of H2 storage operation (2050 net zero; core scenario). Positive numbers for discharging (red); negative numbers for charging (blue).

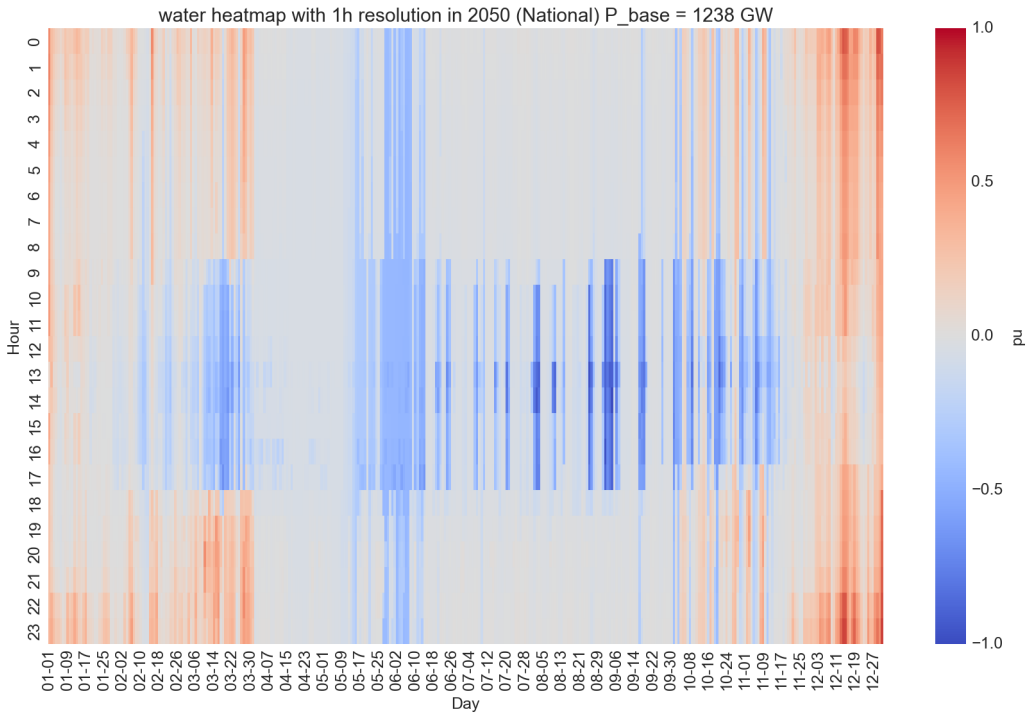

Figure 3: Heatmap of water storage operation (2050 net zero; core scenario). Positive numbers for discharging (red); negative numbers for charging (blue).

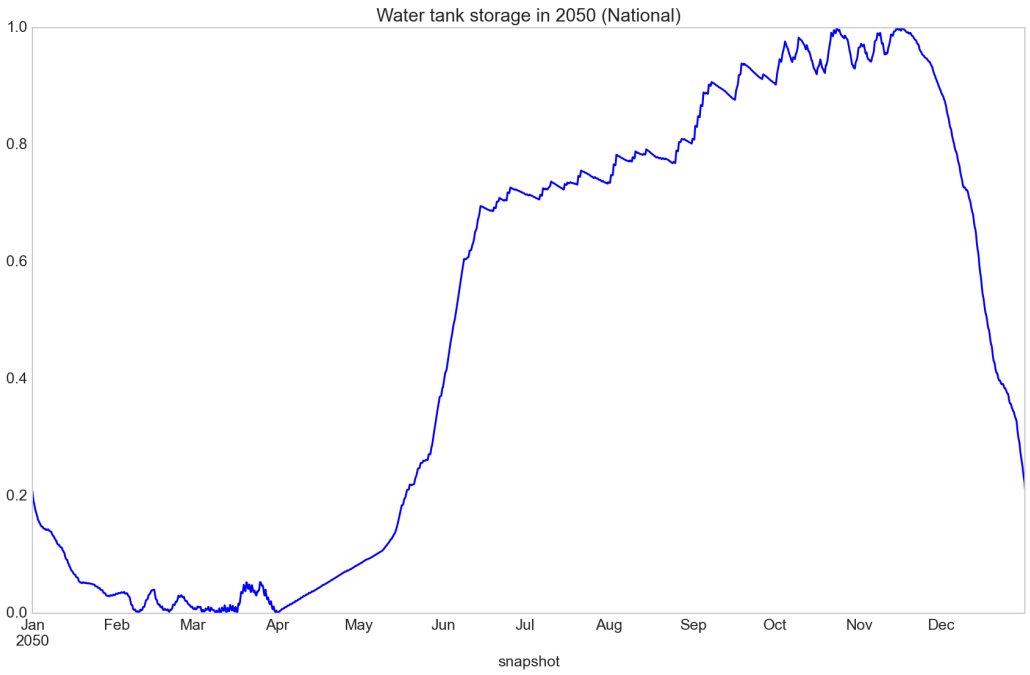

Figure 4: Stored energy (normalized with respect to rated value) in water tanks (2050 net zero; core scenario).

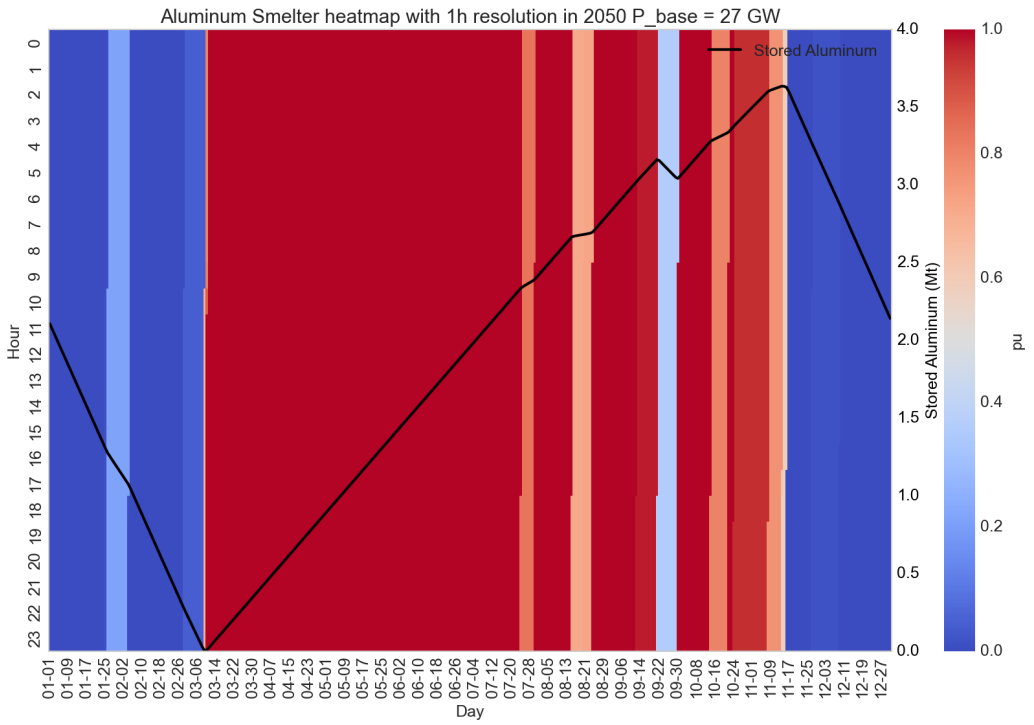

Figure 5: Aluminum smelter operation heatmap (2050 net zero; High-flexibility; 36% overcapacity). 1 for consuming electricity at rated power (red); 0 for shutting down (blue).

2. Primary Aluminum Demand Projection

Our objective is to forecast China's aluminum demand over the coming decades while simultaneously predicting the availability of recycled aluminum. Our methodology is based on the assumption that per capita aluminum stock levels tend to saturate with economic development, which is a commonly used approach for forecasting demand for metallic raw materials such as steel and aluminum [1, 2]. Having obtained the annual aluminum demand projections, we assume that the hourly aluminum demand remains constant throughout each year. This demand must be satisfied either through production from aluminum smelters or by drawing from existing aluminum inventories. This assumption simplifies the temporal distribution of aluminum demand and allows us to focus on the operational flexibility of aluminum smelters in meeting the steady hourly demand requirements. A similar approach has been adopted by Li et al. [3]. The constant hourly demand assumption eliminates the complexity of modeling temporal demand variations while imposing stricter supply-demand balance constraints, as any production shortfall must be compensated by inventory drawdown, and any production surplus must be stored, both of which incur additional costs.

2.1. Historical Aluminum Stock Calculation

To obtain historical aluminum stock data, which is not explicitly reported in official statistics, we extracted historical aluminum production data from the *China Statistical Yearbook* and calculated aluminum stock by accumulating historical production while accounting for product aging. Specifically, utilizing primary and secondary aluminum production data from 1990 to 2023 in China, and assuming that aluminum product lifetimes follow a Weibull distribution, we calculated the annual in-use aluminum stock through 2024. The historical aluminum stock calculation is based on a dynamic material flow analysis approach that considers both primary aluminum production and recycling effects. The methodology employs a Weibull distribution to model the lifetime and decay patterns of aluminum products in use. For each target year t, the aluminum stock $S(t)$ is calculated as:

$$S(t) = \sum_{y=1950}^{t} P_{adj}(y) \cdot \phi(t-y) \tag{1}$$

where $P_{adj}(y)$ represents the adjusted aluminum production in year y (including recycled content), and $\phi(t-y)$ is the survival probability function

for aluminum products of age $(t - y)$.

The survival probability is modeled using a Weibull distribution with the following parameters: Shape parameter $k = 2$, Scale parameter $\lambda = \frac{\mu}{\Gamma(1+1/k)}$, where μ is the mean product lifetime, Mean lifetime μ varies by scenario: 14-18 years depending on the configuration. Here, $k = 2$ implies a linearly increasing hazard with age and captures wear-out behavior typical of capital equipment and infrastructure; this value is widely adopted in reliability and stock-turnover studies, lies within the empirical range reported for comparable technologies (approximately k≈1.5-3), and avoids imposing extreme skewness. We treat k=2 as an explicit modeling assumption and our results confirm consistent with existing studies [4, 5], which do not change the study's qualitative conclusions.

The survival probability function is defined as:

$$\phi(age) = 1 - F_{Weibull}(age; k, \lambda) \tag{2}$$

where $F_{Weibull}$ is the cumulative distribution function of the Weibull distribution.

The methodology incorporates recycling effects through an iterative two-pass calculation: *First Pass:* Calculate initial stock without considering recycling effects using historical production data. *Second Pass:* For each year $t > 1950$, calculate the decay amount from previous years' production:

$$D(t) = \sum_{y=1950}^{t-1} P_{adj}(y) \cdot f_{Weibull}(t - y; k, \lambda) \tag{3}$$

$$= \sum_{y=1950}^{t-1} P_{adj}(y) \cdot \frac{k}{\lambda} \cdot \left(\frac{t-y}{\lambda}\right)^{k-1} \cdot e^{-\left(\frac{t-y}{\lambda}\right)^k} \tag{4}$$

where $f_{Weibull}$ is the probability density function of the Weibull distribution. The recycled aluminum amount is then calculated as:

$$R(t) = D(t) \cdot \rho \tag{5}$$

where ρ is the recycling rate (ranging from 0.65 to 0.75 depending on the scenario). The adjusted production for year t becomes:

$$P_{adj}(t) = P_{primary}(t) + R(t) \tag{6}$$

2.2. EnergyPathways Model Configuration

To integrate aluminum demand forecasting into the EnergyPathways framework [6], we configured the model using the stock-based method with national population data as the primary driver. The national population data was processed from provincial population projections by Chen et al. [7] and aggregated to create a national population driver. The historical aluminum stock values from 2003-2025 were input into EnergyPathways as the baseline data for fitting an S-shaped growth curve. The model applies logistic regression to fit a sigmoid function to the historical stock data, capturing the characteristic saturation behavior where per capita aluminum stock levels approach an asymptotic maximum as economic development progresses. This S-curve fitting approach allows the model to project future aluminum stock levels beyond the historical data period by extrapolating the fitted growth trajectory. The EnergyPathways model then employs the following formula to calculate the projected national aluminum demand [8]:

$$D_{yrc} = \sum_{t \in T} \sum_{v=1}^{V} U_{tvryc} \cdot f_{tvc} \cdot d_{yr} \cdot (1 - R_{yrc}) \tag{7}$$

where D_{yrc} represents the aluminum demand in year y of carrier c in region r, U_{tvryc} is the normalized share of service demand for technology t of vintage v, f_{tvc} is the efficiency of technology t using carrier c, d_{yr} is the total service demand input for year y in region r, and R_{yrc} represents unitized service demand reductions.

After running the EnergyPathways model, aluminum scrap availability and primary aluminum demand were calculated using the model-generated outputs (Figure 6). The model produces two key outputs: total aluminum demand by year, and aluminum production and sales by year. Aluminum scrap is calculated as the difference between total demand and stock increase using the formula:

$$S_y = D_y - \Delta\text{Stock}_y \tag{8}$$

where $\Delta\text{Stock}_y = \text{SD}_y - \text{SD}_{y-1}$, S_y represents scrap in year y, D_y represents sales, and SD_y represents service demand. Primary aluminum demand accounts for recycling and import scenarios through the following calculation:

$$P_y = D_y - (S_{d,y} + S_{i,y}) \cdot R_y \tag{9}$$

where P_y is primary demand, D_y is domestic demand, $S_{d,y}$ and $S_{i,y}$ are domestic and imported scrap, and R_y is recycling rate. The calculation incorporates

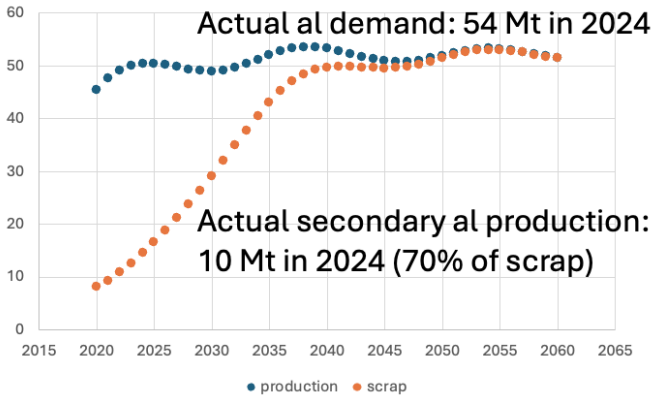

Figure 6: EnergyPathways model outputs in core scenarios (unit: Mt/year). The model validation against 2024 actual data shows strong agreement: projected total aluminum demand of 50 Mt closely matches the actual domestic demand of 52 Mt (54 Mt total demand minus 2 Mt primary aluminum imports). The projected scrap aluminum availability of 15 Mt aligns with the model's historical stock calculation, and the 70% recycling rate used in the model is consistent with the actual recycled aluminum production of 10 Mt (15 Mt $\times$ 70% = 10.5 Mt).

scenario-specific parameters where domestic scrap is calculated from model outputs.

2.3. Scenario Configuration for Aluminum Demand

In our scenario configuration, the parameter that most directly influences national aluminum demand is the empirical saturation value used when fitting per capita aluminum stock demand with a logistic curve. As the baseline scenario (Mid), we set the per capita aluminum stock demand saturation at the level of U.S. demand (approximately 600 kg per capita), while setting the saturation values for primary aluminum demand in the Low and High scenarios at 500 and 700 kg per capita, respectively. The former represents a scenario where, despite China's overall economic scale potentially exceeding that of the United States, per capita consumption demand may not reach U.S. levels due to relatively constrained per capita resources, with this value being close to the settings in the literature [5]. The latter represents a scenario where the development of electric vehicles and renewable energy-related industries drives aluminum demand, but these industries have not yet dominated in developed countries such as the United States, while potentially becoming a major component of China's economy, thus China's future per

capita aluminum stock demand may exceed the current U.S. per capita aluminum stock level.

China is also a major importer and exporter of aluminum-related products. Based on current export levels, we have also established scenarios for reduced and increased exports. Historically, due to insufficient domestic scrap aluminum availability, China has long relied on importing scrap aluminum from abroad for recycled aluminum refining to supplement domestic aluminum demand. In 2005, imported scrap aluminum accounted for up to 60% of the scrap aluminum used in national recycled aluminum production [4]. Since then, both the total volume and proportion of imported scrap aluminum have declined year by year. Over the past decade since 2015, China's scrap aluminum import volume as a proportion of domestic available scrap aluminum has decreased from approximately 30% to around 10%. We believe that in the future, as domestic scrap aluminum availability increases, the proportion of imported scrap aluminum will further decline. Therefore, in the baseline scenario (Mid), the proportion of imported scrap aluminum to domestic available scrap aluminum is set at 5%, while in the Low scenario, the scrap aluminum import proportion remains unchanged at 10%, representing a setting with more available scrap aluminum compared to the more likely baseline scenario, thus representing a lower primary aluminum production demand. In the High scenario, conversely, the imported scrap aluminum proportion is set at 0%, as abundant domestic scrap aluminum supply may make imported scrap aluminum uneconomical, representing a setting that may reduce the impact of scrap aluminum availability on primary aluminum demand reduction.

Two parameters that indirectly affect primary aluminum demand through multiplier effects are the scrap aluminum recycling rate and the average lifetime of aluminum products. In developed countries such as the United States, the scrap aluminum recycling rate can reach as high as 60% [9]. In recent years, China's scrap aluminum recycling rate (recycled aluminum production/available scrap aluminum) has been around 70% (estimated at 10.5 million tonnes/15 million tonnes in 2024), although China's recycled aluminum accounts for only about 25% of total aluminum production, this is primarily due to insufficient available scrap aluminum rather than low recycling rates. Meanwhile, the average lifetime of aluminum products also affects total aluminum demand, as new products need to be produced to replace products that age and become scrap aluminum in a given year. In the Low primary aluminum demand scenario, we assume that with the improvement of sus-

tainable development concepts and technologies in Chinese society, the scrap aluminum recycling rate further increases to 75%, while the average lifetime of aluminum products increases from 16 years in the baseline scenario (the baseline value used in the literature) to 18 years, thus producing more recycled aluminum output with the same level of available scrap aluminum while reducing aluminum demand, thereby further reducing primary aluminum demand. In contrast, in the High Demand scenario, considering that the increased scrap aluminum may partially consist of aluminum materials previously invested in the construction sector, which are relatively difficult to recycle or may not meet the quality requirements for aluminum products, the recycling rate is set at 65%, and a shorter average product lifetime is set.

In our study, we did not explicitly set the impact of aluminum ingot import and export volume changes on China's primary aluminum demand in the scenario configuration. The most important reason is that in recent years, China's aluminum ingot exports have accounted for less than 5% of national primary aluminum production, having a relatively small impact on results. Our considerations also include that the proportion of aluminum ingot exports is more viewed as a result rather than a cause of domestic aluminum demand and supply relationships, and factors that could lead to both increases and decreases in this proportion may become dominant. On one hand, the export volume of industrial raw materials such as aluminum ingots is likely to further decrease in the future with China's economic development and transformation, and aluminum ingot exports from countries rich in mineral and energy resources such as Australia may be more competitive, so China's aluminum ingot export volume is unlikely to increase significantly in the future. On the other hand, as the proportion of recycled aluminum increases, China's domestic average aluminum ingot costs and energy consumption will decrease, and China is unlikely to significantly import aluminum ingots. In summary, we believe that overall, the impact of aluminum ingot imports and exports on China's national electrolytic aluminum demand may be neutral, or there is no reason to assume a positive or negative impact on primary aluminum demand in specific scenario settings, with domestic demand and scrap aluminum availability and recycling being the dominant factors.

3. Smelter Operation Modeling and Capacity Distribution

3.1. Electrolytic Cell Flexibility Modeling

The model captures the two most important characteristics of electrolytic cell operation through limited parameters: temperature constraints and stoichiometry. This model represents the constraints of electrolytic cell operational flexibility, reflecting the academic and industrial understanding for the complex dynamics of the thermal process in electrolytic cell operation processes and the need for expensive restart procedures once the internal temperature falls below critical points.

This study employs a mixed-integer linear model, an initial version of which was developed by Shen et al. [10] as a linear model for production and energy consumption decisions in smelter operations. Refer to the original paper for more details including the detailed mathematical expressions and linearization methods. The model is capable of accurately reproducing the relationships between internal temperature changes and electrolysis rates with applied voltage and current as observed in numerical simulations of electrolytic cell operation (since the electrical energy consumed per unit of aluminum production increases non-linearly as the electrolytic cell deviates from its rated operating point), while being transformed into a linear programming model through piecewise linearization and numerical approximation. It optimizes the input voltage and current of electrolytic cells at hourly intervals while ensuring that the internal temperature of electrolytic cells remains within allowable ranges.

The improved modeling method for electrolytic aluminum loads (EALs) addresses the limitations of prior models by incorporating more accurate thermal balance constraints and efficient mathematical approximations. The model aims to achieve a computationally efficient optimization framework that accounts for the complex energy-material-temperature conversion mechanisms within EALs.

The model starts with the fundamental *power-current coupling constraint*, which models a potline as an equivalent circuit with internal electromotive force (EMF) and resistance. This leads to a quadratic relationship between power and current, given by $P_{j,t}^{\mathrm{al}} = {I_{j,t}^{\mathrm{al}}}^2 R_{j,m} + I_{j,t}^{\mathrm{al}} E_{j,m}$. To improve computational efficiency, a first-order Taylor expansion approximation is applied, resulting in the following linear relationship:

$$P_{j,t}^{\mathrm{al}} = (2I_{j,\mathrm{N}}^{\mathrm{al}} R_{j,m} + E_{j,m}) I_{j,t}^{\mathrm{al}} - {I_{j,\mathrm{N}}^{\mathrm{al}}}^2 R_{j,m} \tag{10}$$

where $P_{j,t}^{\text{al}}$ and $I_{j,t}^{\text{al}}$ are power and current of potline j at time t.

A key improvement is the more accurate *power-temperature coupling constraint*, which recognizes that input power perturbations disrupt the thermal balance of the electrolytic cell. Unlike the original model, which oversimplified the dynamic thermal response by neglecting changes in reaction energy and heat dissipation, this improved method incorporates the thermal inertia of both the electrolyte and the cell shell. The energy balance is described by a set of discretized differential equations:

$$P_e(t) - k\eta(W_h + W_r)I(t) - S_1 h_1(T_e(t) - T_s(t)) = c_e m_e(T_e(t+1) - T_e(t)) \quad (11)$$

$$S_1 h_1(T_e(t) - T_s(t)) - S_2 h_2(T_s(t) - T_0(t)) = c_s m_s(T_s(t+1) - T_s(t)) \quad (12)$$

Here, T_e and T_s represent the electrolyte and cell shell temperatures, respectively, and the equations account for input power, reaction energy, heat dissipation, and the change in internal energy over time. Furthermore, the model addresses inherently nonlinear *yield-current coupling* and *temperature-current relationships*. Instead of imposing rigid yield constraints as in the original model, the improved model removes them, allowing the optimization to naturally determine the most economical yield level through cost function minimization.

The improved model is a Mixed Integer Nonlinear Programming (MINLP) model that accounts for the piecewise, discontinuous nature of EAL regulation costs. The cost function includes electricity costs, production losses, and the risk of cell shutdowns, with different cost structures depending on the potline's current level. To handle the *discontinuity at piecewise points*, the model introduces additional weight variables (w_k) to perform a weighted sum of the function values on both sides of each point. For *time-coupled variables*, the model simplifies the calculation of potline shutdown risk costs, which are typically tied to the duration of operation in a "Cooling State". The original formulation required tracking a binary state variable across multiple time periods, leading to complexity and the multiplication of binary variables. The improved model addresses this by establishing a lower bound for the duration of the Cooling State and calculates the shutdown risk cost as a simple sum

over all time steps where the potline is in this state:

$$S_j^{\text{wtr-cool}} = (1 - e^{-\gamma}) S_{j,on}^{\text{al}} \sum_t^T y_{1,j,t} \tag{13}$$

This simplification is justified by the observation that in optimal operation, potlines rarely enter the Cooling State due to economic considerations, which validates the practical viability of this lower-bound estimation.

In this study, to model the shutdown decisions and associated costs of smelters, we introduced startup and shutdown variables and costs into the aforementioned linear model, transforming it into a mixed-integer model. This allows our model to optimize smelter operational decisions over a larger time scale, but the introduction of integer variables also makes it computationally infeasible to directly embed a plant-level operational model into a power system planning model. Therefore, this paper designs a decomposition-based iterative algorithm for the solution, as detailed in Note 4.

Although our baseline model (Mid-flexibility) has adopted the most advanced models currently available in academia, real parameters that can be obtained, and precise numerical simulation results, the parameters still have considerable uncertainty due to the limited implementation of deep grid interaction by aluminum smelters, particularly involving actual cases of adjusting internal temperature to near critical levels or even restarting. To encompass the results that these uncertain parameters may lead to, we established Low and High flexibility scenarios beyond the baseline scenario:

Using performance data from numerical simulations of aluminum smelter operations, as well as existing literature on the shutdown, restart processes, and costs of real-world aluminum smelters, we developed a reduced-order model that represents the key characteristics of aluminum electrolytic cell temperature constraints and restart costs. However, the cost uncertainty is large because there are few historical cases of restarts of aluminum smelters, and we also evaluated three comparative scenarios: Low: No progress has been made in smelter operational flexibility, and restart costs in historical high-cost reports are used. High: New technologies and methods have significantly improved smelter operational flexibility, and restart costs have dropped substantially. This represents the goal of ongoing research, development, and demonstration efforts to enhance smelter-grid interaction [11]. Un-constrained: This is an idealized scenario that disregards the temperature constraints, start-up/shutdown constraints, and restart costs of potlines. It

only considers rated power constraints, and the value of grid interaction in this scenario represents the theoretical maximum.

3.2. Province-level Capacity Distribution

Table 2: Capacity distribution by province (unit: 10,000 t/year).

Node #	Province	Aluminum Smelter Capacity	Node #	Province	Aluminum Smelter Capacity
1	Heilongjiang	15	17	Anhui	161
2	Inner Mongolia	506	18	Xizang	0
3	Jilin	8	19	Shanghai	0
4	Liaoning	56	20	Hubei	73
5	Xinjiang	536	21	Sichuan	106
6	Beijing	0	22	Chongqing	38
7	Hebei	5	23	Zhejiang	43
8	Tianjin	0	24	Jiangxi	161
9	Gansu	300	25	Hunan	123
10	Shanxi	93	26	Guizhou	117
11	Ningxia	91	27	Fujian	62
12	Shandong	527	28	Yunnan	461
13	Qinghai	186	29	Guangxi	308
14	Shaanxi	124	30	Guangdong	41
15	Henan	298	31	Hainan	0
16	Jiangsu	59			

Based on the Chinese government's policy that primary aluminum capacity will not exceed 45 million tons per year, we assume that the production capacity of aluminum smelters remains unchanged at the 2024 year-end level (45 million tons per year) in the baseline scenario, with annual electricity consumption accounting for approximately 6.8% of national power consumption. Without early retirement, we assume this capacity will continue until 2060. This assumption is based on the typical lifespan of an aluminum smelter exceeding 50 years, and the fact that more than 90% of China's aluminum smelting capacity was built after 2000.

Since not all provinces have complete statistical data on aluminum smelting capacity, we distributed the national aluminum smelting capacity to each province based on provincial aluminum production in 2023 (Table 2), and assumed that the capacity share of each province remains unchanged across all scenarios. Theoretically, the optimal retirement strategy should prioritize retiring capacity in provinces with high operational costs (primarily electricity

costs) or low renewable energy potential. However, such an optimized retirement strategy would require treating aluminum smelter capacity as decision variables to be co-optimized with power system capacity expansion, which is computationally intractable within the PyPSA-China framework. Given that this study focuses on the value of overcapacity rather than optimal capacity distribution, and considering that distributional optimization represents a second-order improvement, we leave this optimization for future work, which may require more computationally efficient energy supply-demand co-optimization models for effective solution.

To systematically explore the trade-offs between overcapacity maintenance costs and flexibility benefits, we designed multiple overcapacity retention scenarios across different combinations of aluminum demand, smelter flexibility, and market opportunity settings. In addition to the two extreme cases of complete overcapacity elimination (where capacity equals demand) and complete capacity retention (maintaining 2024 levels), we also established intermediate scenarios with 5%, 10%, 20%, ..., 90% overcapacity retention rates. The retention rate refers to the proportion of overcapacity retained relative to the initial overcapacity when no early retirement occurs. For example, if demand is 5 Mt and initial capacity is 45 Mt, then a 10% overcapacity retention corresponds to an actual capacity of 5 + (45-5) $\times$ 10% = 9 Mt. We assume that the capacity share of each province remains unchanged across all scenarios, meaning that capacity reductions are proportionally distributed among provinces. Given the declining trend in aluminum demand, we do not consider capacity expansion scenarios.

3.3. Smelter Costing

Table 3: Typical costing of aluminum smelting in 2020 (In 2020 CNY).

Item	Unit Consumption	Unit Price (CNY)	Cost (CNY/tonne aluminum)
Alumina	1.925	2885	5553.625
Electricity	13500	0.463	6250.5
Anode Carbon Block	0.45	5800	2610
Aluminum Fluoride	0.025	10650	266.25
Cryolite	0.003	7100	21.3
Labor			150
Maintenance Costs			400
Other Expenses			1500
Depreciation Costs			300
Total			17051.675

As a traditional industry that has developed for many years, the investment, operation, and maintenance costs of aluminum smelting can be assessed from numerous reports based on actual cases. In this study, we primarily base our calculation of aluminum electrolysis cost changes across various scenarios on the aluminum smelting production cost composition disclosed in the 2020 annual report of China Aluminum Corporation (Table 3). In the optimization model, we treat raw material costs as constant across all scenarios since they do not affect production operation optimization results. Investment costs (amortized as depreciation costs or retirement losses, with different proportions depending on retirement decisions) and maintenance costs are only related to the maintained capacity, therefore modeled as fixed costs. Other costs such as management expenses, sales expenses, and financial expenses are treated as variable costs, proportional to production volume. We model labor costs as standby costs, meaning they are only paid in full when smelters are operating and not paid when smelters are shutdown. The electricity cost for aluminum electrolysis operation is a result of the optimization model, where the energy component cost is settled at nodal marginal electricity prices, and we assume that the proportion of transmission and distribution fees to energy costs remains constant (averaging 25% in 2020, lower than residential electricity due to aluminum smelting plants typically operating at higher voltage levels with corresponding lower transmission and distribution rates).

However, it is challenging to obtain sufficient reports on idling costs and startup/shutdown costs, as there are not many historical cases where smelters were idled for long periods or shut down due to insufficient demand. This is partly because aluminum smelting has high investment costs, typically requiring high capacity utilization to recover investment costs. On the other hand, shutting down smelters, which requires substantial labor to perform complex operations and maintenance to avoid irreversible damage to electrolytic cells from temperature changes and internal liquid solidification during this process (Figure 7), is extremely expensive. According to several publicly available reports on electrolytic cell restarts following demand declines after the 2008 Great Recession and the 2020 pandemic, the average cost of a single electrolytic cell restart can range from less than 110,000 CNY/MW to 760,000 CNY/MW [13, 14, 15], which is 1000 times higher than restarting coal-fired units [16]. This study does not attempt to predict the future trajectory of smelter restart costs, but rather sets up extreme cases in addition to the baseline case to represent the two ends of the spectrum for the

Figure 7: If the restart process is not done properly, it may cause permanent damage to the cell [12]. The restart process of an aluminum electrolytic cell requires a lot of manual work, including complex steps such as cleaning the cell, preheating, adding electrolyte, and powering on.

development of flexible aluminum smelting technologies, resulting in a total of three smelter flexibility cases:

In the *Low flexibility* and *Mid flexibility* cases, the smelter flexibility parameters are set based on current literature and reports. The former's restart cost comes from a case of a U.S.-owned smelter restart, estimated at 760,000 CNY/MW. The latter's restart cost is from a documented restart event at a smelter in Yunnan, China, estimated at 110,000 CNY/MW.

In the *High flexibility* case, we assume that the aluminum smelting industry has improved electrolytic cell insulation and restart technologies to enhance grid interaction flexibility. Restart costs have significantly decreased due to technological improvements and cost learning curve benefits from frequent restarts, representing the best-case scenario for aluminum smelting and grid interaction technology development.

3.4. Warehouse Costs

For smelters to conduct seasonal production to help adapt to seasonal supply and demand fluctuations in the power grid, they may need to increase the storage period of aluminum products to act as a buffer for balancing production and demand. Therefore, we also need to model the warehouse costs of aluminum products. The warehouse cost is determined by the density

of aluminum, which defines how much space one metric ton occupies. The calculation method follows: Storage demand equals cumulative production minus cumulative demand, and storage area equals storage demand divided by density, warehouse height, and storage efficiency. For example, the warehouse costs in Yunnan Province are based on short-term rental rates [17], which may overestimate actual costs for two reasons: aluminum smelters typically build their own warehouses, and long-term storage is expected to be cheaper than short-term rentals of general-purpose warehouses.

Using Kunming rental rates of 0.8 CNY/m^2·day (including tax) as an example, assuming flat storage warehouses with good load-bearing capacity, 7-meter height, and 70% storage efficiency, the volume of 1 ton of primary aluminum is 0.37 m^3 (1/2.7), requiring a floor area of 0.14 m^2 (0.37 m^3 / 7 m / 0.7). The storage cost per hour is 4.7×10^{-3} CNY (0.14 m^2 × 0.8 CNY/m^2·day / 24 hours). For 1 ton of aluminum, the cost for storing 2 weeks is less than 1.5 CNY, compared to electricity costs of approximately 6000 CNY, indicating that storage costs are negligible relative to electricity costs in aluminum production. Currently, the actual inventory levels of steel, cement, and aluminum fluctuate around weekly production volumes. Therefore, the current proportion of warehouse costs to total costs for electrolytic aluminum is almost negligible. Even after optimization and shifting to seasonal storage as suggested in our results, warehouse costs have not become a significant component. Additionally, aluminum products have relatively low density and are easy to transport, therefore we don't specifically consider changes in transportation costs.

3.5. Workforce Estimation

The distribution of industrial workers required for aluminum smelters and coal/gas power plants across different months is estimated based on the monthly capacity utilization rate, maximum output, and the average number of workers required per unit of operating smelting capacity or generation capacity. For aluminum smelters, we use data from the 2020 annual report of China Aluminum Corporation (CHALCO) [18] as a reference: with 6.79 million tons of aluminum production capacity, the company employs 104,801 workers, approximately 100,000 people. In 2024, China's aluminum industry employed 600,000-700,000 workers. We assume that the number goes down with the decrease in aluminum smelting capacity and use the average number of workers required per unit of operating smelting capacity to estimate the

maximum labor force required for aluminum smelters and coal/gas power plants.

Gas-fired power plants have a high degree of automation. A typical 1 GW combined cycle gas turbine unit typically requires only 100-300 personnel for operation and maintenance. We use 200 workers per GW for gas-fired power plants based on expert experience. For coal-fired power plants, the staffing requirements are higher, typically 600-1,000 workers per GW, and we use 800 workers per GW based on expert experience. These numbers are higher than the global average due to China's relatively low labor costs. While the actual number of workers at each plant varies depending on when it was built and its level of automation, we use these broad technology-based estimates without distinguishing between different coal/gas power technologies. Generally, 60-85% of the workforce at a smelter or power plant is fixed, independent of operational status. However, for seasonal operations where facilities are completely shut down for certain months, maintenance typically requires only 10-15% of the total workforce. We use 10% in our estimates, as this type of seasonal operation may become common practice in the future, leading industries to optimize their operational structure to further reduce staffing needs during shutdown periods.

When considering capacity factors, for the winter months of November to February, we use the maximum monthly output as the reference for employment estimation. This is because in low-carbon power systems, to cope with winter load peaks and heating demand, coal and gas power plants operate at high intensity and are likely to be fully staffed, even when capacity factors are less than 50%. In this case, estimating based on output hours would significantly underestimate the required workforce.

4. Power System Capacity Expansion Model

This study aims to evaluate the load flexibility value of aluminum smelters for China's decarbonized power system. A provincial resolution is needed because China's aluminum industry is unevenly distributed across multiple provinces, and the highly interconnected nature of China's power grid means that the impact of this flexibility is likely to span across provincial boundaries. We set the net-zero year of 2050 though China's overall carbon neutrality target is 2060 [19], as the power sector is generally considered the primary and critical area for achieving this goal and requires deep decarbonization at an earlier stage. According to several studies and scenario analyses, China's power sector is expected to achieve net-zero emissions between 2045 and 2055 [20]. To compare the effects of different carbon reduction constraints, we set up scenarios with 20%, 60%, and 100% reductions in carbon emissions for 2030, 2040, and 2050, respectively, based on 2020 emission levels. We consider three main power system scenarios (see Table 1) that reflect different levels of power generation technology development. All scenarios assume that a variety of competitive, clean, and stable technologies are available, covering a range from high fixed costs and low variable costs to low fixed costs and high variable costs. We also assume the availability of cost-competitive long-duration storage technologies (in the form of hydrogen and hot water) and a certain degree of flexibility in residential heating.

This study uses PyPSA, an open-source model that has been described in detail elsewhere [21], to explore China's pathway to carbon emission reduction. Our model operates at hourly resolution and adopts a framework of simultaneous operation and investment optimization for capacity expansion planning, with investment periods corresponding to 2030, 2040, and 2050. Within each investment period, both operational and investment costs are minimized. The model covers thirty-one provincial administrative regions of China (excluding Hong Kong, Macau, and Taiwan), with each province aggregated into a single independent node. Although we refer to the model as an electricity system in this paper, each node actually resolves electricity, heat, hydrogen, biomass, methane, and coal carriers, making it a multi-energy coupled system where electricity serves as the core energy exchange and conversion medium (see Figure 8).

The electricity buses are connected via high-voltage direct current (HVDC) or high-voltage alternating current (HVAC) lines, representing transmission. Hydrogen buses are connected via hydrogen pipelines. Different sector buses

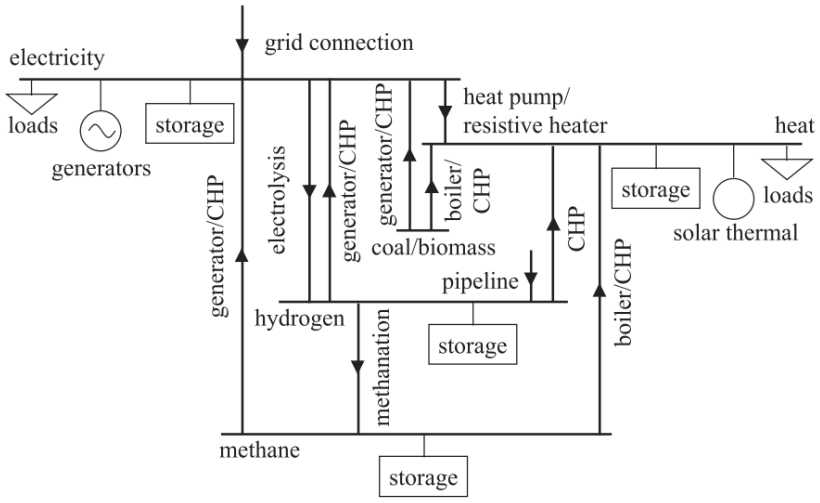

Figure 8: Energy flow at a single node. In the model, a node represents a provincial-level administrative region in China. Within each node, there is a bus (thick horizontal line) for each energy carrier (electricity, heat, hydrogen, biomass, methane, and coal) to which different loads (triangles), energy sources (circles), storage units (rectangles), and converters (lines connecting buses) are attached. The notation "boiler/CHP to heat" indicates that coal and biomass can either be converted directly to heat through boilers, or simultaneously produce both electricity and heat through combined heat and power (CHP) plants.

are connected within nodes through energy converters (as shown in Figure 8). Our model utilizes the PyPSA modeling framework, which can model the operation and optimal investment of energy systems across multiple periods. The model employs linear optimization methods to minimize annual operational and investment costs under technical and physical constraints, assuming perfect competition and perfect foresight of weather conditions. The capacity of generators, storage, converters, and transmission, as well as the hourly operational dispatch of each unit, are subject to optimization.

In the model, demand curves (for different sectors), the ratio of district heating to decentralized heating, and the hydro capacity of reservoirs and run-of-river generators and pumped hydro storage are exogenous variables that are not optimized. Transmission network assets can be reinforced but not removed. The detailed mathematical modeling formulations are outlined in the original paper. The detailed input data for the model is primarily from the PyPSA China-specific project, originally established and open sourced by Zhou et al. [22], and has been used for peer-reviewed publications. This section details the input data and key assumptions used in the model, covering network infrastructure, electricity and heat demand, and the techno-economic characteristics of various energy technologies. Most of the technical data regarding costs, efficiencies, and lifetimes are based on the database used in PyPSA-Eur [23], which includes technology cost projections from 2020 to 2050. We assume these parameters remain constant after 2050.

4.1. Mathematical Formulation

The model's objective is to determine the optimal capacities of generators, storage units, and transmission lines, as well as their hourly dispatch, to minimize the total annual system cost. This cost is a combination of fixed investment costs and variable operational costs. The model assumes a multi-node system, where buses are denoted by n, generation and storage technologies at a given bus by s, the hour of the year by t, and branches (including transmission lines and energy converters) by ℓ.

The objective function in each investment period is to minimize the sum of all costs:

$$\min_{G_{n,s},E_{n,s},F_\ell,g_{n,s,t},f_{\ell,t}} \sum_{n,s} c_{n,s}G_{n,s} + \sum_{n,s} \hat{c}_{n,s}E_{n,s} + \sum_{\ell} c_\ell F_\ell + \sum_{n,s,t} o_{n,s,t}g_{n,s,t} + \sum_{\ell,t} o_{\ell,t}f_{\ell,t} \tag{14}$$

The terms in the objective function represent the following costs:

- $c_{n,s}G_{n,s}$: Fixed annualised cost of generation and storage power capacity, where $c_{n,s}$ is the specific cost and $G_{n,s}$ is the power capacity.
- $\hat{c}_{n,s}E_{n,s}$: Fixed annualised cost of storage energy capacity, where $\hat{c}_{n,s}$ is the specific cost and $E_{n,s}$ is the energy capacity.
- $c_\ell F_\ell$: Fixed annualised cost of branches, where c_ℓ is the specific cost and F_ℓ is the branch capacity.
- $o_{n,s,t}g_{n,s,t}$: Variable operational cost for generator and storage dispatch, with $o_{n,s,t}$ as the specific cost and $g_{n,s,t}$ as the hourly dispatch.
- $o_{\ell,t}f_{\ell,t}$: Variable operational cost for power flow through branches, with $o_{\ell,t}$ as the specific cost and $f_{\ell,t}$ as the hourly power flow.

We assume a social discount rate of 2% in this study [24], which is consistent with assumptions in Zhou et al. [22]. The model is subject to several technical and physical constraints. The first is the energy balance constraint, ensuring that the inelastic energy demand $d_{n,t}$ at each bus n is met at every time step t. This is achieved by local generation, storage dispatch, or energy flow from connected branches.

$$\sum_s g_{n,s,t} + \sum_\ell \alpha_{\ell,n,t} \cdot f_{\ell,t} = d_{n,t} \quad \forall n, t \tag{15}$$

where $\alpha_{\ell,n,t} = -1$ if branch ℓ starts at bus n, representing an outflow, and $\alpha_{\ell,n,t} = \eta_{\ell,t}$ if branch ℓ ends at bus n, representing an inflow. The efficiency factor $\eta_{\ell,t}$ can be time-dependent to account for technologies whose efficiency, such as that of a heat pump, varies with external conditions.

The dispatch $g_{n,s,t}$ of each generator and storage unit is bounded by its installed capacity $G_{n,s}$ and time-dependent availability factors, $\underline{g}_{n,s,t}$ and $\bar{g}_{n,s,t}$:

$$\underline{g}_{n,s,t} \cdot G_{n,s} \leq g_{n,s,t} \leq \bar{g}_{n,s,t} \cdot G_{n,s} \quad \forall n, s, t \tag{16}$$

For conventional flexible generators, the availability factors are typically constant with $\underline{g}_{n,s,t} = 0$ and $\bar{g}_{n,s,t} = 1$. For variable renewable energy technologies, such as wind and solar, the time-varying $\bar{g}_{n,s,t}$ represents the weather-dependent available power per unit, and since curtailment is allowed, $\underline{g}_{n,s,t} = 0$. For storage units, the availability factors are $\underline{g}_{n,s,t} = -1$ and $\bar{g}_{n,s,t} = 1$, where the negative value of the lower bound denotes the available charging power.

The power capacity $G_{n,s}$ itself is subject to minimum $\underline{G}_{n,s}$ and maximum $\bar{G}_{n,s}$ installable potentials:

$$\underline{G}_{n,s} \leq G_{n,s} \leq \bar{G}_{n,s} \quad \forall n, s \tag{17}$$

Similarly, the power flow $f_{\ell,t}$ on each branch is constrained by its capacity F_ℓ and time-dependent per-unit availabilities, $\underline{f}_{\ell,t}$ and $\bar{f}_{\ell,t}$:

$$\underline{f}_{\ell,t} \cdot F_\ell \leq f_{\ell,t} \leq \bar{f}_{\ell,t} \cdot F_\ell \quad \forall \ell, t \tag{18}$$

For bidirectional HVDC links, $\underline{f}_{\ell,t} = -1$ and $\bar{f}_{\ell,t} = 1$. For unidirectional branches, such as a resistive heater converting electricity to heat, the flow is constrained to be non-negative with $\underline{f}_{\ell,t} = 0$ and $\bar{f}_{\ell,t} = 1$.

Finally, the total CO_2 emissions from the power and heat sectors are limited by a predefined cap, CAP_{CO_2}. This is enforced using the specific emission factors ϵ_s (in tCO_2/MWh_{th}) for each fuel type s, the efficiency $\eta_{n,s}$ of the generator, and its dispatch $g_{n,s,t}$:

$$\sum_{n,s,t} \frac{\epsilon_s g_{n,s,t}}{\eta_{n,s}} \leq CAP_{CO_2} \tag{19}$$

Further details on the specific constraints and modelling framework can be found in the documentation for PyPSA [21].

4.2. Provinces and Network

The model adopts a one-province, one-node network representation, including all thirty-one provincial administrative regions of mainland China. The existing transmission infrastructure (Figure 9(a)), as of September 2022, consists of both High-Voltage Alternating Current (HVAC) and High-Voltage Direct Current (HVDC) interconnections between provinces. For future transmission expansion, we use a fully connected national grid (Figure 9(b)) as the candidate network, which is a strategy strongly supported by the State Grid Corporation of China and connects regional grids via Ultra-High-Voltage Alternating Current (UHVAC) lines.

4.3. Electricity and Heat Demand

Historical and projected demand for electricity (Figure 10) and heat are used as exogenous inputs. The heat demand considered includes low-temperature space heating (SPH) and domestic hot water (DHW) for the residential and service sectors. Industrial heat demand is not included. Within

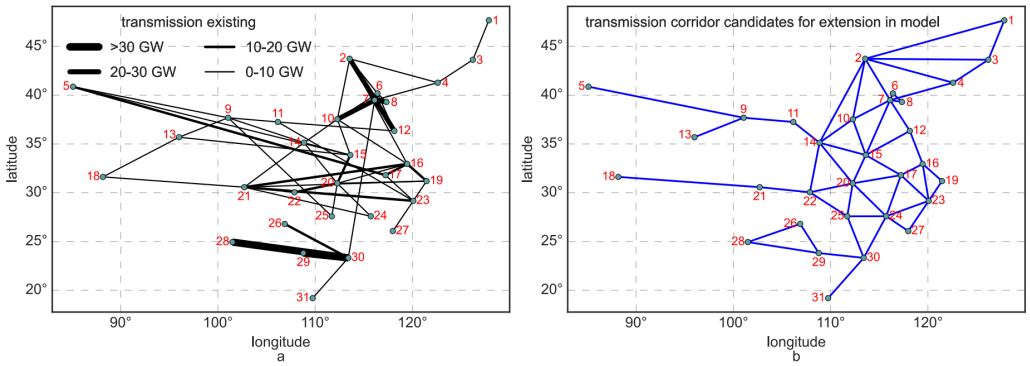

Figure 9: Chinese transmission networks. (a) Shows the included existing transmission lines. (b) Shows the transmission corridor candidates for extension in the model referring to the fully connected national grid (FCG). The province name of each node is listed in Table 2.

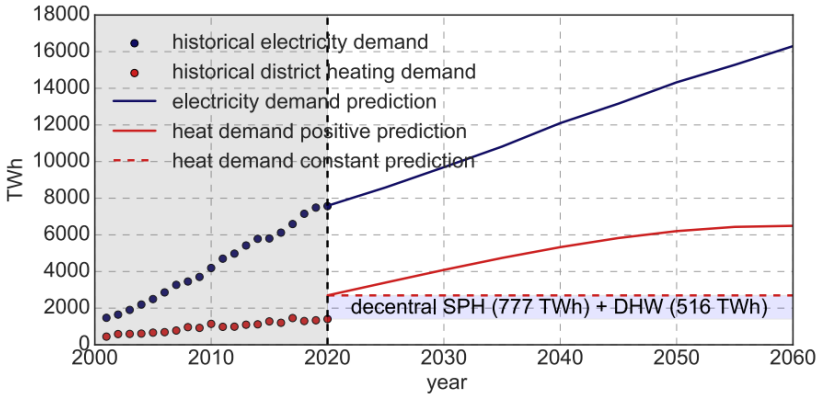

Figure 10: Historical and projected demand for electricity and heat.

each province, heat demand is split between decentralized heating for rural areas and district heating for urban areas, based on 2020 urban-rural population ratios.

Historical heat demand is estimated to be 2,695 TWh in 2020, composed of district heating (1,402 TWh), decentralized SPH (777 TWh), and DHW (516 TWh). Electricity demand projections from 2020 to 2060 are based on China's economic development forecasts. We assume annual growth rates that decrease over time, from 2.7% (2020–2030) to 1.3% (2050–2060), leading to a projected electricity demand of approximately 16,000 TWh in 2060, consistent with IEA reports. Electrification of the heating and transport sectors is not explicitly considered in these demand forecasts.

Hourly demand curves for SPH are derived from ERA5 reanalysis temperature data using a degree-day approximation, while DHW curves are constructed using daily heat needs and intra-day patterns. Hourly electricity demand curves are built from 2018 standard load profiles and scaled to match the annual total demand for each projection year.

4.4. Power and Heat Generation Technologies and Costs

The model includes a diverse range of technologies to meet electricity and heat demands (Table 4). Retrofitted coal power plant represents the coal power plant retrofitted with carbon capture. The carbon capture rate is assumed as 90%. The retrofitting of coal power plants reduces their efficiency to 90% of the original value but does not reduce operational lifespan. Heat pumps include both air-sourced and ground-sourced heat pumps. The CHP, OCGT, and CC stand for combined heat and power, open cycle gas turbine, and carbon capture, respectively. The costs of power and heat generation technologies are based on the database used in PyPSA-Eur [23] and other studies (Table 5). All costs are given in 2020 CNY. Here, fixed Operation and Maintenance (FOM) costs are given as a percentage of the investment cost per year. The investment cost of solar photovoltaic (PV) systems is determined as the average of costs associated with the utility-scale solar installation and those of the rooftop solar installation. The efficiency of Combined Heat and Power (CHP) systems is the electricity generation efficiency. Refer to the supplementary material of Zhou et al. [22] for more details.

4.4.1. Renewable Energy Generation

Hourly wind and solar generation time series are generated by combining ERA5 reanalysis weather data (wind speed, solar radiation, temperature)

Table 4: Power and heat generation technologies.

Electricity	Central Heating	Decentral Heating
Coal power plant	Coal boiler	Coal boiler
Retrofitted coal power plant	CHP coal	Gas boiler
OCGT	CHP coal	Biomass boiler
Onshore wind	CHP gas	Resistive heater
Offshore wind	CHP biomass	Heat pump
Hydroelectricity	CHP biomass CC	Solar thermal
CHP coal	CHP hydrogen	Short-term TES
CHP gas	Resistive heater	
CHP biomass	Heat pump	
CHP biomass CC	Solar thermal	
CHP hydrogen	Long-term TES	

with technology specifications (Figure 12). The potential for onshore and utility-scale solar PV is constrained by eligible land use types and protected areas, as identified by the *World Database on Protected Areas (WDPA)* (Figure 11). We assume specific installation densities for different technologies: 10 MW/km^2 for onshore wind, 10 MW/km^2 for offshore wind (in water depths up to 50 m), and 170 MW/km^2 for solar PV.

4.4.2. Fossil Fuels with Carbon Capture

Existing coal-fired power plants can be retrofitted with carbon capture facilities, assuming a constant 90% CO_2 capture rate from 2020 to 2060. The retrofit does not extend the operational life, and reduces the plant efficiency to 90% of its original value.

4.4.3. Combined Heat and Power (CHP)

The CHP model is based on extraction condensing units, which define a feasible operational area for simultaneous power and heat production. When operating in condensing mode without heat extraction, the electrical efficiencies vary by plant type and vintage. Existing CHP coal power plants achieve 37% electrical efficiency in this mode, while newly constructed CHP coal facilities demonstrate improved performance at 48%. CHP gas power plants operate at 40% electrical efficiency when functioning solely for electricity generation. The flexibility to extract steam for heat production allows these units to adapt their operational configuration based on demand requirements, creating a versatile dispatch capability within the energy system.

Table 5: Costs of power and heat generation technologies. (In 2020 CNY)

Technology	Investment Cost [CNY]	Unit	FOM [%/a]	Lifetime [a]	Efficiency
Wind onshore	8728.2	kW_{el}	1.25	27	1
Wind offshore	14079	kW_{el}	2.51	27	1
Solar PV	5717.4	kW_{el}	1.58	35	1
Solar thermal collector decentral	2106	m^2	1.3	20	Variable
Solar thermal collector central	1092	m^2	1.4	20	Variable
Open cycle gas turbine (OCGT)	3541.2	kW_{el}	1.78	25	0.4
Gas boiler decentral	2433.6	kW_{th}	6.56	20	0.97
Gas boiler central	468	kW_{th}	3.25	25	1.03
Combined heat and power (CHP) gas central	4602	kW_{el}	3.31	25	0.4
Coal power plant	29998.8	kW_{el}	1.6	40	0.33
Coal power plant retrofit	1287	kW_{el}	0	40	0.9
CHP coal central	14820	kW_{el}	1.63	25	0.48
Coal boiler central	2254.2	kW_{th}	2.5	25	0.9
Coal boiler decentral	1302.6	kW_{th}	3.49	20	0.9
Hydro electricity	17222.4	kW_{el}	1	80	See text
Nuclear power plant	43001.4	kW_{el}	1.4	40	0.33
Methanation + DAC	13221	kWH_2	3	20	0.8
Biomass boiler decentral	5327.4	kW_{th}	7.39	20	0.82
CHP biomass central	26371.8	kW_{el}	3.61	25	0.3
CHP biomass carbon capture	25740000	tCO_2/h	3	25	0.9
Air-sourced heat pump decentral	7332	kW_{th}	2.96	18	Variable
Air-sourced heat pump central	7417.8	kW_{th}	0.21	25	Variable
Ground-sourced heat pump decentral	11700	kW_{th}	1.85	20	Variable
Hydrogen electrolysis	5070	kW_{el}	2	25	0.66
CHP hydrogen central	10140	kW_{el}	5	10	0.5
Hydrogen storage with tanks	93.6	kWh_{H_2}	0	20	1
Hydrogen storage underground	23.4	kWh_{H_2}	0	100	1
Hydrogen pipeline	1762.8	MW/km	4	50	1
Resistive heater decentral	780	kW_{th}	2	20	0.9
Resistive heater central	546	kW_{th}	1.52	20	0.99
Battery inverter	2106	kW_{el}	0.2	10	0.97*0.97
Battery storage	1809.6	kWh	0	20	1
Hot water tank decentral	140.4	kWh_{th}	1	20	τ=3 days
Hot water tank central	4.524	kWh_{th}	0.52	20	τ=180 days
Hot water tank (dis)charging	0		0		0.84

Notes: [a]Our technology classification may ignore the specific technology difference among each category. For example, for coal power plants, supercritical (SC) and ultra-supercritical (USC) coal power plants can achieve higher efficiencies than the values shown here. [b]We assume a uniform 10% efficiency penalty for all power generation technologies equipped with carbon capture. However, for coal and biomass power plants, the actual efficiency reduction could reach 15% [25]. Therefore, we adopt a best-case scenario for carbon capture/DAC that likely underestimates the costs of clean firm energy, which may lead to an underestimation of the value of aluminum load flexibility in our results.

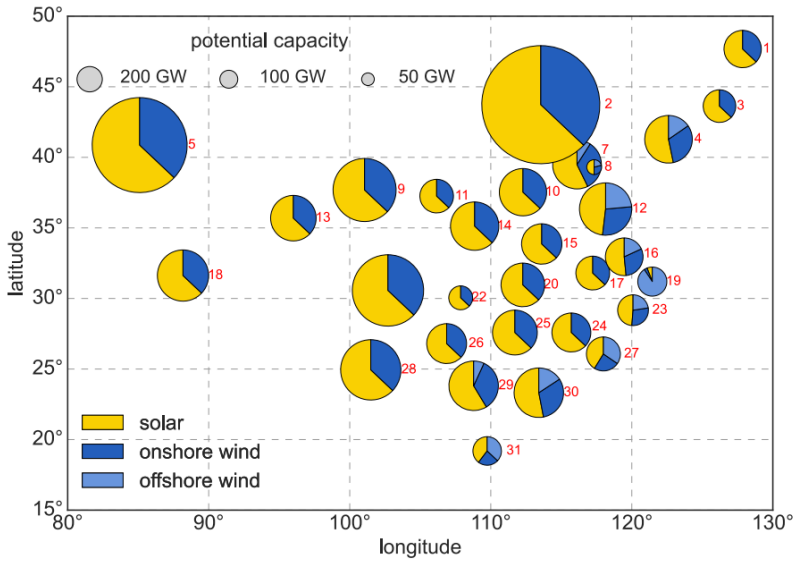

Figure 11: Potential for onshore and utility-scale solar PV constrained by eligible land use types and protected areas. Yearly capacity factor represents the ratio of average energy output of a resource over a year to its rated/nameplate value, indicating the utilization rate of that resource. In this figure, renewable energy resources capacity factors are displayed in circular insets to save space.

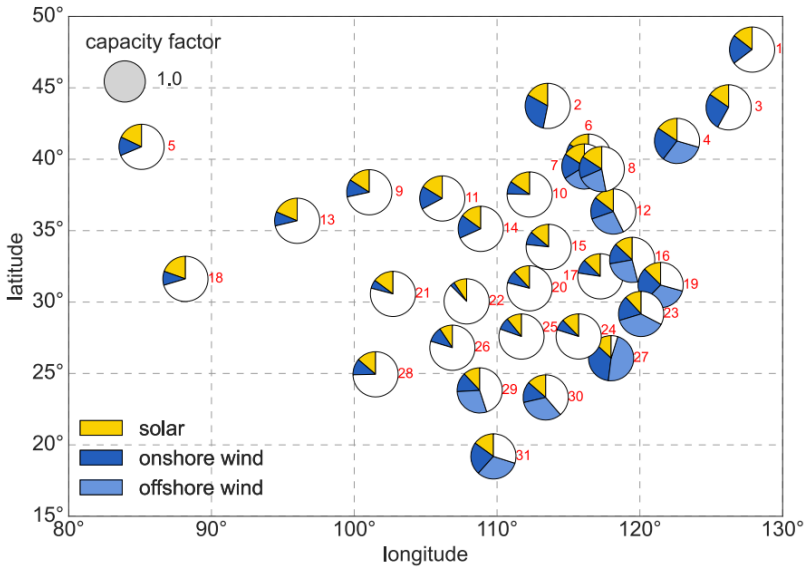

Figure 12: Capacity factors of different technologies.

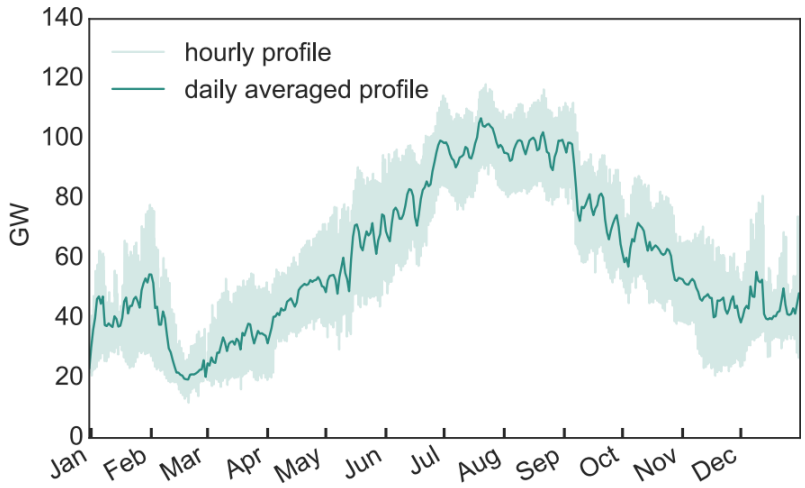

Figure 13: Averaged hydroelectricity time series of 41 large-scale hydro stations.

4.4.4. Hydropower

The model includes 41 large-scale hydropower plants, with their hourly inflow and potential generation calculated from gridded surface runoff data (Figure 13).

4.4.5. Heat Pumps

The performance of heat pumps is modeled with a temperature-dependent Coefficient of Performance (COP). The COP decreases significantly when the temperature difference between the heat source and sink increases, which is particularly important since heating demand tends to peak during cold periods when the COP is lowest. For air-source heat pumps, the COP is calculated as a quadratic function of the temperature difference ΔT (in °C) between the sink and source: COP = 6.81 - $0.121\Delta T + 0.00063\Delta T^2$. Ground-source heat pumps follow a similar relationship but with different coefficients: COP = 8.77 - $0.150\Delta T + 0.00073\Delta T^2$. The model assumes a constant sink water temperature of 55°C.

4.4.6. Biomass

The model considers agricultural residues that are typically burned as waste (0.19 Gt/year) as a potential biomass source. Biomass technologies, including boilers and CHP, can be equipped with carbon capture, which

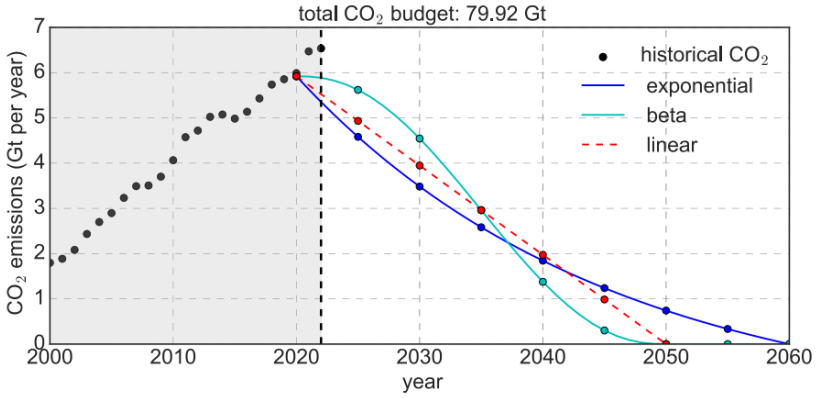

Figure 14: Carbon budget and pathways.

results in negative net CO_2 emissions but reduce the efficiency of the plants to 90% of the original value.

4.4.7. Thermal Storage

Both long-term (LTES) and short-term (STES) hot water storage are available for district and decentralized heating, respectively. These storage systems are modeled with specific hourly energy losses.

4.4.8. Electricity Storage

Two scalable stationary storage options are included: batteries and hydrogen storage. For hydrogen, the capacity for electrolysis, storage (in steel tanks or underground salt caverns), and fuel cells can be optimized independently. The model assumes only provinces with salt caverns can invest in underground hydrogen storage.

4.4.9. Synthetic Methane and DAC

In certain scenarios, synthetic methane can be produced from hydrogen using the Sabatier process. The CO_2 required for this process is sourced from the atmosphere via *Direct Air Capture (DAC)*. This captured CO_2 is distinct from the CO_2 sequestered from CHP or retrofitted coal plants.

4.5. Carbon Budget and Pathways

The study explores China's power and heat sector decarbonization pathways within a total carbon budget of 79.9 $GtCO_2$, which corresponds to

27.3% of the global remaining budget for a "well below 2°C" target. This budget is based on the sector's 2020 emissions and is designed to achieve net-zero emissions by 2050.

The CO_2 emission reduction pathway follows a linear trajectory from 2020 to 2050 (Figure 14). No emission constraints are imposed in 2025, as emissions are expected to peak around this time. Subsequently, emission reductions of 20%, 60%, and 100% relative to 2020 levels are implemented in 2030, 2040, and 2050, respectively. Net CO_2 emissions are calculated based on the consumption of natural gas (0.2 tCO_2/MWh_{th}) and coal (0.34 tCO_2/MWh_{th}), with biomass assumed to be carbon-neutral. Our analysis focuses on fuel combustion emissions and excludes upstream emissions (e.g., coal mining, biomass growing, harvesting, and transportation).

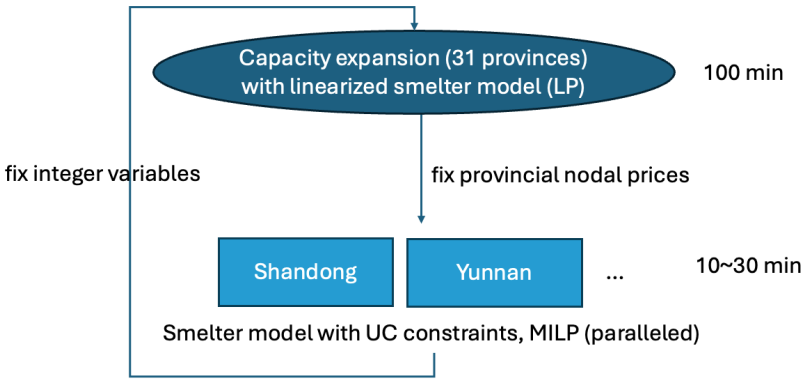

Figure 15: Framework of the iterative solution algorithm.

5. Dimensionality Reduction and Tailored Iterative Solution

Our objective is to address the issue of computational intractability that arises when integrating aluminum smelter operation model into the power system planning model. We adopt the D3R framework proposed by Lyu et al. [26], which is designed for the dimensionality reduction and linearization of industrial load constraints. Based on simulations of results, we determine the parameters for a linear approximation model of the original precise model, which is then embedded into the linear capacity expansion model, and its optimal solution serves as the initial value. After determining the local marginal price at each province from the optimal solution of the linear capacity expansion model, all smelters run their detailed 8760-hour operational models (which include integer variables) using the hourly electricity price as an exogenous input. This yields the optimal operating results for each smelter. At this point, the operation of each plant is decoupled from the power system and other smelters, and the problem can be solved in parallel. Then, we fix the smelter's operational decisions and optimize the remaining planning and operational variables of the power system. This process is iterated until the objective function of the planning model converges, which proves that we have simultaneously obtained the optimal planning and smelter operation results.

5.1. Data-Driven Dimension Reduction (D3R) Framework

The integration of detailed aluminum smelter operational models into power system planning models presents significant computational challenges

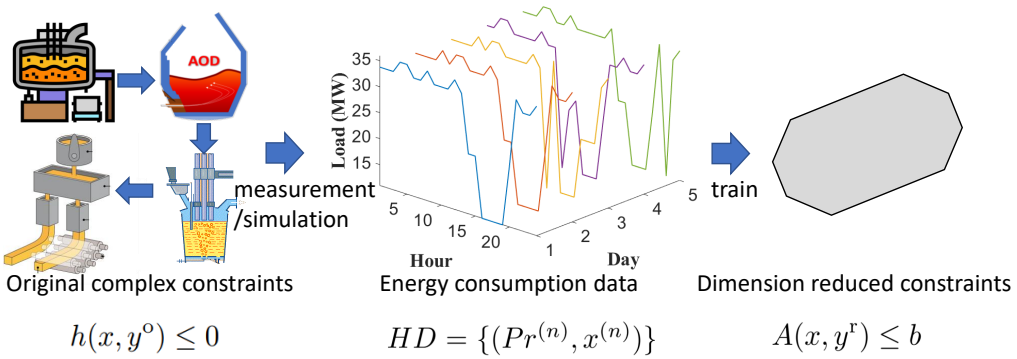

Figure 16: Data-driven dimension reduction framework for industrial load modeling, illustrated with a steel plant example. h: high-dimensional constraints; HD: historical dataset; A: low-dimensional constraints.

due to the complex mixed-integer constraints inherent in industrial load modeling. Traditional analytical dimension reduction methods, which rely on pre-specified geometric templates (such as boxes, ellipses, or zonotopes) to approximate original constraints, face two key limitations: they lack adaptability across various industrial loads with different constraint structures, and they require convexity in the original constraints, making them incompatible with industrial load models containing integer variables.

To address these limitations, we adopt the Data-Driven Dimension Reduction (D3R) framework (Figure 16) proposed by Lyu et al. [26]. D3R leverages optimal energy consumption data from industrial loads to train a dimension-reduced model that best fits the original constraints through inverse optimization. The framework exhibits enhanced adaptability and is compatible with constraints involving integer variables, making it particularly suitable for aluminum smelter modeling.

5.1.1. Problem Formulation

The D3R approach works by replacing the original high-dimensional constraints $h(x, y^{\mathrm{o}}) \leq 0$ with simplified lower-dimensional linear constraints $A(x, y^{\mathrm{r}}) \leq b$, where x represents hourly energy consumption, y^{o} represents variables in the original constraints (including integer variables for operating modes), and y^{r} represents variables in the reduced constraints (continuous only). Rather than requiring strict equivalence between the reduced and original constraints, D3R introduces a loss function $J(h(\cdot), A, b)$ to quantify

the approximation error and determines appropriate parameters A and b to minimize this error.

The training process uses historical energy usage data $HD = \{(Pr^{(n)}, x^{(n)})\}$, where $Pr^{(n)}$ and $x^{(n)}$ represent hourly electricity prices and energy consumption on day n, respectively. The optimization problem is formulated as:

$$\min_{A,b}. J = \frac{1}{N}\sum_{n=1}^{N} ||x^{(n)} - x_n||^2 \tag{20a}$$

$$\text{s.t.} \quad x_n = \text{argmin.}\{Pr^{(n)\intercal} x_n \ : \ A(x_n, y^{\text{r}}) \leq b\}, \ \forall n \tag{20b}$$

where x_n is the optimal energy consumption for day n based on the reduced constraints. This inverse optimization approach has been validated across multiple industrial load datasets, demonstrating superior performance compared to analytical methods with normalized root mean square errors typically ranging from 4% to 10% in terms of hourly load profile and less than 0.2% in terms of objective value (e.g., optimal electricity cost).

5.1.2. ALF-based Reduced Constraints

The D3R framework employs an adjustable load fleet (ALF) composed of multiple adjustable loads (ALs) as the form of the reduced constraints to be trained. The motivation for choosing ALF is that it can directly model the actual composition and behavior of industrial loads without requiring baseline load definitions or assuming bidirectional energy exchange capabilities as in virtual battery models. The ALF model captures three key physical characteristics of industrial loads through linear constraints:

Equipment Composition: Multiple pieces of equipment in industrial facilities are represented by parallel ALs, where the total energy consumption is the sum of individual device consumption.

Equipment Power Rating: Each device's power consumption is bounded by its rated capacity.

Production Goals: Daily energy consumption limits reflect production targets, as they are typically proportional to energy usage through product energy intensity factors.

The reduced constraints in the form of an ALF are then:

$$x_t = \sum_{i=1}^{I} p_{t,i}\Delta t, \quad \underline{P}_i \leq p_{t,i} \leq \overline{P}_i \ : \ \underline{\mu}^{\text{P}}_{t,i}, \overline{\mu}^{\text{P}}_{t,i}, \ \forall t \tag{21a}$$

$$\underline{E}_i \leq \sum_{t=1}^{T} p_{t,i}\Delta t \leq \overline{E}_i \ : \ \underline{\mu}_i^{\mathrm{E}}, \overline{\mu}_i^{\mathrm{E}} \tag{21b}$$

where the subscript n is omitted, I is the number of ALs as a hyperparameter that balances model complexity and approximation accuracy. A larger I allows reduced constraints to better approximate original constraints but increases computational costs, while limited training data may constrain the choice of I to avoid overfitting. x_t is the net energy consumption at time t, which is the summation of $p_{t,i}\Delta t$, representing the average consumption power of AL i at time t multiplied by the length of time interval Δt. $\theta = \{\underline{P}_i, \overline{P}_i, \underline{E}_i, \overline{E}_i\}$ include the power and energy limits of AL i, respectively, which are the parameters to be fitted. The Lagrange multipliers $\underline{\mu}_{t,i}^{\mathrm{P}}, \overline{\mu}_{t,i}^{\mathrm{P}}, \underline{\mu}_i^{\mathrm{E}}, \overline{\mu}_i^{\mathrm{E}}$ are the dual variables associated with the power and energy limits, respectively. The above linear constraints can be easily transformed into the form of $A(x, y^{\mathrm{r}}) \leq b$, where $x = [x_t]$ and $y^{\mathrm{r}} = [p_{t,i}]$.

While this simplified model may overlook some temporal coupling between production processes and relax integer variables, it effectively balances computational complexity and approximation accuracy. Moreover, the D3R framework is not limited to the ALF model - other reduced constraint forms can be chosen based on specific needs while maintaining the framework's core methodology.

5.1.3. Training Algorithm for the ALF Model

Substituting the form of the ALF, the optimality conditions (Karush-Kuhn-Tucker conditions) for day n include primal feasibility, stationarity condition derived from the Lagrangian function, dual feasibility, and complementary slackness:

$$Pr_t^{\mathrm{E}} - \underline{\mu}_{t,i}^{\mathrm{P}} + \overline{\mu}_{t,i}^{\mathrm{P}} - \underline{\mu}_i^{\mathrm{E}} + \overline{\mu}_i^{\mathrm{E}} = 0, \ \forall t, \forall i \tag{22a}$$

$$(\underline{\mu}_{t,i}^{\mathrm{P}}, \overline{\mu}_{t,i}^{\mathrm{P}}, \underline{\mu}_i^{\mathrm{E}}, \overline{\mu}_i^{\mathrm{E}}) \geq 0, \ \forall t, \forall i \tag{22b}$$

$$\underline{\mu}_{t,i}^{\mathrm{P}}(\underline{P}_i - p_{t,i}) = 0, \quad \overline{\mu}_{t,i}^{\mathrm{P}}(p_{t,i} - \overline{P}_i) = 0, \quad \forall t, i \tag{22c}$$

$$\underline{\mu}_i^{\mathrm{E}}(\underline{E}_i - \sum_{t=1}^{T} p_{t,i}) = 0, \quad \overline{\mu}_i^{\mathrm{E}}(\sum_{t=1}^{T} p_{t,i} - \overline{E}_i) = 0, \quad \forall i \tag{22d}$$

The complementary slackness conditions involve bilinear constraints, which can be transformed via the Fortuny-Amat transformation into a more easily

solvable form. For example, the first term can be replaced by the following:

$$\underline{\mu}^{\mathrm{P}}_{t,i} \leq M(1 - z_{t,i}), \quad p_{t,i} - \underline{P}_i \leq M z_{t,i}, \quad \forall t, i \tag{23}$$

where $z_{t,i}$ is a binary variable and M is a large positive number. The transformed conditions are used in inverse optimization, forming a mixed-integer linear programming (MILP) problem that can be solved by commercial solvers.

Since the number of integer variables scales with dataset size N, direct solving becomes computationally intractable. We employ zeroth-order stochastic gradient descent (ZOSGD) for iterative solving, which is suitable because it: 1) avoids explicit gradient computation, 2) reduces computational burden through batch processing, and 3) maintains tractability via iterative optimization. In each iteration, certain days of data are randomly selected and parameters are updated using estimated gradients:

Initialize $j = 0$ and $\theta^{(0)} = \{\underline{P}_i, \overline{P}_i, \underline{E}_i, \overline{E}_i\} = [0]$.

Randomly select B days from the dataset and solve the batch problem: $\min_{\theta} J = \frac{1}{B}\sum_{n=1}^{B} ||x^{(n)} - x_n||^2$ s.t. ALF constraints and optimality conditions to obtain θ^*.

Update $\theta^{(j+1)} = (1 - \alpha)\theta^{(j)} + \alpha\theta^*$; if $\theta^{(j+1)}$ converges, break; otherwise, set $j = j + 1$ and go to step 2.

where α is the learning rate, which can be adaptively adjusted. The ZOSGD algorithm can be easily implemented by using commercial solvers, and the computational complexity is independent of the number of days in the dataset. For more detail about the D3R method, refer to the original paper with the open-sourced codes.

5.1.4. Implementation for Aluminum Smelting

The integration of aluminum smelter operational models into the PyPSA-China framework is achieved through a sophisticated multi-bus architecture that separates electrical and aluminum production systems while maintaining their operational coupling. The system creates dedicated aluminum buses for each province with significant aluminum production capacity, establishing a parallel network structure that mirrors the electrical grid topology. This design allows for independent optimization of aluminum production while maintaining the necessary electrical connections for power supply.

Aluminum smelters are modeled as specialized Link components that convert electrical energy into aluminum production, with each smelter connected

between the electrical bus of its respective province and a dedicated aluminum bus. We employ the first-order adjustable load model (linearized) from the D3R framework to approximate the original smelter's complex flexibility operational model. The adjustable load model can be effectively modeled using PyPSA's Link tools, with all parameters calculated and provided by the D3R framework. The aluminum modeling framework incorporates sophisticated operational constraints that reflect the real-world characteristics of aluminum smelting operations. These constraints are implemented through configurable parameters that define the operational flexibility of aluminum smelters across different scenarios, supporting multiple flexibility levels ranging from low flexibility (90% minimum power output) to high flexibility (50% minimum power output), with corresponding startup costs that reflect the technical and economic challenges of varying smelter operations.

Algorithm 1 Iterative Optimization Algorithm for Aluminum Smelter Integration

Require: Initial aluminum production targets Q_n^0 for each province n
Require: Convergence tolerance $\epsilon = 1 \times 10^{-4}$, Maximum iterations $k_{max} = 10$, $k = 0$, $f^0 = \infty$
while $k < k_{max}$ **do**
  Step 1: Power System Optimization
  Solve linear power system capacity expansion model with D3R-approximated aluminum constraints
  Extract nodal marginal electricity prices λ_n^k for each province n
  Step 2: Aluminum Smelter Optimization (Parallel)
  for each province n in parallel **do**
    Solve detailed aluminum operational MILP model:
    Extract optimal energy consumption pattern x_n^k
  end for
  Step 3: Update and Convergence Check
  Fix aluminum energy consumption patterns x_n^k in power system model
  Solve updated power system optimization to get objective f^k
  Calculate relative change: $\delta = |f^k - f^{k-1}|/|f^{k-1}|$
  if $\delta < \epsilon$ **then**
    Converged: Return optimal solution
    BREAK
  else
    Update iteration counter: $k = k + 1$
    Update objective value: $f^{k-1} = f^k$
  end if
end while
Return: Optimal capacity expansion decisions and aluminum operational schedules

The iterative optimization algorithm (Algorithm 1) addresses the computational intractability that arises when integrating aluminum smelter operation models into power system planning models. The algorithm employs a decomposition-based approach that alternates between solving the power system optimization problem and the aluminum smelter operational optimization problems until convergence is achieved (Figure 15). This iterative

process allows for the coordination of system-wide optimization with detailed industrial process optimization while maintaining computational tractability.

The algorithm begins by solving the power system capacity expansion model with simplified linear constraints representing aluminum smelters, using the D3R framework to approximate the complex mixed-integer constraints. The solution provides nodal marginal electricity prices for each province, which serve as input parameters for the detailed aluminum smelter operational models. These detailed models, which include integer variables for startup/shutdown decisions and temperature constraints, are then solved in parallel for each province using the nodal electricity prices as exogenous inputs.

The detailed aluminum optimization subproblems for each province are formulated as mixed-integer linear programming problems that minimize the total cost of aluminum production, including electricity costs and startup costs, subject to operational constraints such as minimum and maximum power limits, production targets, and temperature constraints. The optimization considers the binary variables for smelter operation status and startup decisions, ensuring that the thermal constraints and restart costs are properly accounted for in the operational decisions. At this stage, the smelter operations in each province are decoupled from the power system and other provinces' smelter operations, as we assume that the annual aluminum production for each province remains constant during each iteration in both the centralized and distributed problems. This decoupling enables rapid solution of the optimization problems, and although the 8760-hour MILP problems still contain numerous integer variables, they can typically be solved within 10-30 minutes. Also, the MILP problems can be solved in parallel across provinces, significantly reduces computational time, making the approach feasible for large-scale systems with multiple provinces.

After solving the detailed aluminum operational models, the resulting optimal energy consumption patterns are fixed in the power system model, and the remaining planning and operational variables are optimized. This process creates a feedback loop where the power system optimization provides electricity prices that influence aluminum production decisions, while the aluminum operational decisions affect the overall system load and generation requirements. The algorithm continues this iterative process until the objective function of the planning model converges, typically within 3-5 iterations, ensuring that both the power system and aluminum smelter operations are simultaneously optimized. The typical runtime for one iteration

is approximately 2 hours, with the centralized capacity expansion model requiring about 100 minutes and the distributed smelter operational models taking 10-30 minutes to solve in parallel across all provinces.

The convergence criterion is based on the relative change in the objective function between consecutive iterations, with a typical threshold of 1e-4 indicating satisfactory convergence. The algorithm includes mechanisms to prevent infinite loops and ensure numerical stability, such as maximum iteration limits and convergence monitoring.

The optimization problems are solved using the Gurobi solver (version 12.0.0) with carefully tuned parameters to ensure computational efficiency and numerical stability. For the linear programming problems in the power system optimization, we use the barrier method with 192 threads, crossover disabled, and a barrier convergence tolerance of 1e-5. For the mixed-integer linear programming problems in the aluminum smelter optimization, we employ automatic method selection (typically dual simplex for branch-and-bound), with a MIP gap of 0.01, moderate heuristic search intensity of 0.5, and minimal cut generation to balance solution quality and computational speed. The solver configuration includes 192 threads for parallel processing, feasibility tolerance of 1e-6, and disabled dual reductions to maintain numerical stability across the iterative optimization process. The computational experiments are conducted on a high-performance computing cluster using SLURM job scheduling system. Each simulation is allocated 1 node with 40 CPU cores and 1 TB of total memory (25 GB per core), with a maximum runtime of 72 hours. The simulations are executed using Snakemake workflow management system with 40 parallel cores.

It is important to note that our model aggregates aluminum smelting capacity at the provincial level, which in reality represents multiple distinct facilities within each province. If facility-level operational data were available, modeling individual facilities separately would likely reveal greater flexibility benefits than our current aggregated approach. This is because when production demand is lower than the total capacity of multiple facilities in a province, some facilities could be completely shut down while others operate at near full capacity - a more efficient operational pattern than forcing all facilities to operate at reduced capacity simultaneously. Our current provincial-level modeling approach, which requires all capacity within a province to ramp up or down together, likely underestimates the true operational flexibility potential of aluminum smelters. Therefore, the economic benefits of overcapacity flexibility identified in our analysis should be consid-

ered conservative estimates.

References

[1] L. Song, P. Wang, M. Hao, M. Dai, K. Xiang, N. Li, W.-Q. Chen, Mapping provincial steel stocks and flows in China: 1978-2050, J. Clean. Prod. 262 (2020) 121393.

[2] S. Li, Z. Wang, Q. Yue, T. Zhang, Analysis of the quantity and spatial characterization of aluminum in-use stocks in China, Resour. Policy 79 (2022) 102979.

[3] J. Li, R. Lyu, Y. Zhang, X. Cha, K. Zheng, H. Guo, Value-based industrial load shedding with supply chain coordination reduces electricity cost, 2026 IEEE Power and Energy Society International Meeting (PESIM) (in press).

[4] Y. Li, Q. Yue, J. He, F. Zhao, H. Wang, When will the arrival of China's secondary aluminum era?, Resour. Policy 65 (2020) 101573.

[5] Y. Yang, H. Zhang, L. Wu, M. Wang, Supply potential, carbon emission reduction, energy conservation, and sustainable pathways for aluminum recycling in China, Sustain. Prod. Consum. 50 (2024) 239–252.

[6] C. Farid, H. Luo, A. Gunawan, A. Pascale, E. Larson, EnergyPATHWAYS Model User Guide, Revision 1, Technical Report, Andlinger Center for Energy and the Environment, Princeton University, 2024. Accessed: 2025-09-16.

[7] Y. Chen, F. Guo, J. Wang, W. Cai, C. Wang, K. Wang, Provincial and gridded population projection for China under shared socioeconomic pathways from 2010 to 2100, Sci. Data 7 (2020) 83.

[8] B. Haley, R. Jones, B. Preneta, A. T. Murray, Annual Decarbonization Perspective 2024 Technical Documentation, Technical Report, Evolved Energy Research, 2024. URL: https://carbonfreeeurope.imgix.net/2024-ADP-TechnicalDocumentation.pdf, accessed: 2025-09-16.

[9] The Aluminum Association, Key Sustainability Performance Indicators for the Aluminum Can, Technical Report, The Aluminum Association, 2024. URL: https://www.aluminum.org/sites/default/files/

2024-12/FINAL-2024_Aluminum-Can-KPI-Report.pdf, accessed: 2025-09-10.

[10] Y. Shen, R. Lyu, H. Guo, C. Kang, An Improved Modeling Method for Electrolyte Aluminum Loads Considering Thermal Balance and Flexible Regulation Cost, in: 2025 IEEE International Conference on Power Systems and Smart Grid Technologies (PSSGT), 2025, pp. 331–337.

[11] M. Lukin, R. Jeltsch, Reduction Cell Restart Method Influence on Cell Life Evolution, in: B. A. Sadler (Ed.), Light Metals 2013, Cham, 2016, pp. 719–724.

[12] H. A. Øye, POWER FAILURE, TEMPORARY POT SHUT-DOWN, RESTART AND REPAIR, 27th Int. Alum. Conf. (2011).

[13] H. Xueen, W. Xinshe, W. Jianpei, The silver declaration on baiyinpo: The record of the secondary startup of the electrolysis cell of luoyang yugang longquan aluminum co., ltd., China's Nonferr. Met. 15 (2009) 54–57.

[14] N. Lazzaro, High costs challenge smelter responses to aluminum oversupply: Alcoa, 2020. URL: https://www.spglobal.com/commodity-insights/en/news-research/latest-news/metals/051420-high-costs-challenge-smelter-responses-to-aluminum-oversupply-alcoa.

[15] J. Coyne, Alcoa to restart 268,000 mt/year of aluminum smelting capacity in Brazil: company, 2021. URL: https://www.spglobal.com/commodity-insights/en/news-research/latest-news/metals/092021-alcoa-to-restart-268000-mtyear-of-aluminum-smelting-capacity-in-brazil-company.

[16] R. K. SMITH, Analysis of hourly generation patterns at large coal-fired units and implications of transitioning from baseload to load-following electricity supplier, J. Mod. Power Syst. Clean Energy 7 (2019) 468–474.

[17] Colliers, Map of Logistics and Warehouse Market Rental Prices in China for the Second Half of 2023, Colliers official website, 2024. Published: March 11, 2024. Retrieved from https://www.colliers.com.cn/zh-cn/research/20240312logisticsrental. Accessed: 2025-09-10.

[18] Aluminum Corporation of China Limited, 2020 Annual Report, Technical Report, Chinalco Finance and Asset Department (Board of Directors Office), 2021. URL: https://www.chalco.com.cn/tzzgx/yjbg/ndbg/202104/P020210420669662364743.pdf, accessed: 2025-09-10.

[19] F. Yang, X. Yang, X. Li, China's diverse energy transition pathways toward carbon neutrality by 2060, Sustain. Prod. Consum. 47 (2024) 236–250.

[20] J. Liu, T. Wang, Y. Wang, X. Lin, R. Zhou, K. Wang, China's 1+N policy system supports an earlier peak in carbon emissions, Renew. Sustain. Energy Rev. 215 (2025) 115626.

[21] T. Brown, J. Hörsch, D. Schlachtberger, PyPSA: Python for Power System Analysis, J. Open Res. Softw. 6 (2018). arXiv:1707.09913.

[22] X. Zhou, K. Strunz, T. Brown, H. Sun, F. Neumann, Multi-energy system horizon planning: Early decarbonisation in China avoids stranded assets, Energy Internet 1 (2024) 81–98.

[23] J. Hoersch, F. Hofmann, D. Schlachtberger, T. Brown, Pypsa-eur: An open optimisation model of the european transmission system, Energy Strategy Reviews 22 (2018) 207–215. doi:10.1016/j.esr.2018.08.012. arXiv:1806.01613.

[24] M. Victoria, K. Zhu, T. Brown, G. B. Andresen, M. Greiner, Early decarbonisation of the european energy system pays off, Nature communications 11 (2020) 6223.

[25] T. Schmitt, S. Leptinsky, M. Turner, A. Zoelle, C. W. White, S. Hughes, S. Homsy, M. Woods, H. Hoffman, T. Shultz, et al., Cost and Performance Baseline for Fossil Energy Plants Volume 1: Bituminous Coal and Natural Gas to Electricity, Technical Report, National Energy Technology Laboratory (NETL), Pittsburgh, PA, Morgantown, WV, and Albany, OR (United States), 2022. URL: https://www.osti.gov/biblio/1893822. doi:10.2172/1893822.

[26] R. Lyu, H. Guo, G. Strbac, C. Kang, Data-Driven Dimension Reduction for Industrial Load Modeling Using Inverse Optimization, IEEE Trans. Smart Grid 16 (2025) 2695–2698.