Joint Power and Natural Gas Systems Planning (JPoNG) Documentation and Python Code Guide

Rahman Khorramfar

February 2023

Working Draft. Please Do Not Share.

1 Model Formulation

Our model, referred as JPoNG, determines the minimum cost planning decisions for power and NG systems considering the two systems' interdependency. The proposed model considers a range of generation and storage technologies whose operations are modeled via operational and policy constraints across a set of representative periods. The formulation allows different temporal resolutions for the operation of both systems since the data availability or planning requirements can be different for each system. For example, decisions related to power generation, such as dispatch amounts and unit commitment, require hourly resolution. However, NG system operation does not involve generation decisions and only deals with transmission and storage operations for which daily resolution may be sufficient. Moreover, because of the ability for NG pipelines to also provide some storage via line packing, daily resolution for scheduling NG operation could facilitate management of intra-day variations in NG demand. In the model, the operations of both systems are coupled through two sets of constraints. The first set ensure NG flow to the power system. The second coupling constraint limits the CO_2 emissions incurred by consuming fossil-derived NG in both power and NG systems. This is in contrast with previous studies where the emissions limit is either applied to power system only [21], or are imposed separately for each system [43].

As a common practice in the literature to enhance the scalability of the GTEP [43, 21, 37], the scheduling of power system operations is modeled over a set of representative days. Long-term planning of NG system is usually planned on yearly [29, 2, 34], monthly [15], or daily [7, 43] basis. Some studies use different resolution for power and NG systems. Power and gas systems operate on hourly basis and daily basis over representative time periods in Von Wald et al. [43]. Saedi et al. [33] used half-hour resolution for the power system and daily resolution for the NG. The authors propose a two-phase model in which the the power system operations are planned in the first phase and the second phase determines the optimal gas flow. We take a similar approach here and consider hourly and daily time resolution for the power and NG systems, respectively. However, we only consider representative days for the power system and

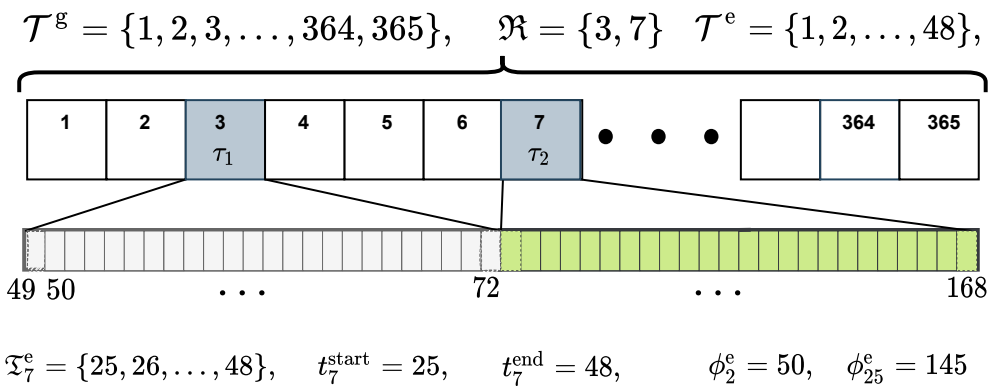

Figure 1: Illustration of the model's temporal resolution for the case of two representative days used for power systems operations. The top row represents planning days in a year, with days 3 and 7 as representative days forming the set $\mathfrak{R}$. The bottom row show the hours corresponding to representative days. The set of planning periods for the NG system ($\mathcal{T}^{\mathrm{g}}$) is the entire year. The set of planning periods for the power system ($\mathcal{T}^{\mathrm{e}}$) is the chronologically ordered set of hours in the representative days whose mapping to their original index is given by $\phi^{\mathrm{e}}_t, t \in \mathcal{T}^{\mathrm{e}}$. The set of hours in their original indexing is denoted by $\mathfrak{T}^{\mathrm{e}}_\tau$ with t^{start}_τ and t^{end}_τ signifying the starting and ending hours for the representative day τ.

model the operations of the NG system across all days of the year. Although the sub-daily representation of NG system can be more advantageous to capture fine-grained operational features, the daily planning in the context of capacity expansion has several merits. First, it enhances the scalability of the model by significantly reducing the number of decision variables and constraints, which allows for introducing complexity in other dimensions of the problem, such as using integer variables for investments decisions, as is done here . Second, it no longer necessitates using nonlinear steady-state gas flow that relates nodal pressure to flow. In other words, a non-compressible formulation can be employed for gas flow in case of daily resolution in which the gas is treated like liquid [23]. The third advantage stems from the data availability. In many cases, data to characterize the NG network operations for real-world systems, such as our case study, is not publicly available at the resolution needed for hourly operations considering pressure variations. Fig. 1 illustrates our modeling approach for the time resolutions of the both systems.

The network representation in the model consists of three sets of nodes as depicted in Figure 2. The first set represents power system nodes and is characterized by different generation technologies (i.e., plant types), demand, storage, and the set of adjacent nodes by which the node can exchange electricity. The second set of nodes are NG nodes each of which are associated with injection amount, demand, and its adjacent nodes. Storage tanks, vaporization and liquefaction facilities, which are commonly used in the non-reservoir storage of NG, collectively form the third set of node referred as storage-vaporization-liquefaction (SVL) nodes. We allow for the possibility of NG storage infrastructure to be located far from demand or injection points in the network, as per existing practice [1] (see 5.1 for detailed discussion). Accordingly, our model makes a distinction between NG and SVL nodes to account for their distinct locations. We also allow the NG system to use LCDF which represents a renewable source of fuel

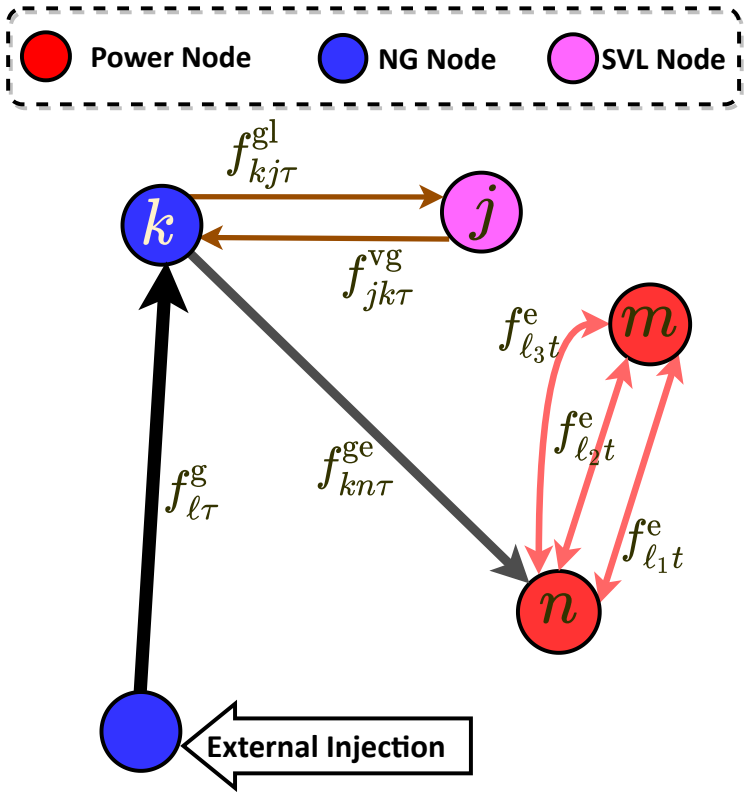

Figure 2: Flow variables between different nodes. In our model, power nodes can be connected by multiple bi-directional transmission lines denoted by $f^{\mathrm{e}}_{\ell t}$. Each power node operates a set of local gas-fired plants by drawing gas from its closest NG node. The variable $f^{\mathrm{ge}}_{kn\tau}$ captures this flow. Each NG node is connected to its closest SVL node through two unidirectional pipelines where one is from NG to SVL's liquefaction facilities denoted by $f^{\mathrm{gl}}_{kj\tau}$; and the other one from SVL's vaporization facility to NG node denoted by $f^{\mathrm{vg}}_{jk\tau}$. The variable $f^{\mathrm{g}}_{\ell\tau}$ denotes the pipe flow between NG nodes. NG nodes can be connected by one or more uni-directional pipelines, but only one connection is depicted here. Candidate transmission lines and pipelines are not shown in this figure.

that is interchangeable with NG and hence can be imported and transported by the NG pipelines [9]. We present power and NG systems' model as well as coupling constraints separately for the ease of exposition. The full description of the mathematical notation used in the formulation is described in 2.

1.1 Power System Model

Objective Function:

$$
\begin{aligned}
\min \sum_{n\in\mathcal{N}^{\mathrm{e}}}\sum_{i\in\mathcal{P}}(C_i^{\mathrm{inv}}x_{ni}^{\mathrm{est}} + C_i^{\mathrm{fix}}x_{ni}^{\mathrm{op}} + \sum_{r\in\mathcal{S}_n^{\mathrm{e}}}(C_r^{\mathrm{pInv}} + C^{\mathrm{pFix}})y_{nr}^{\mathrm{eCD}}) & \\
+ \sum_{n\in\mathcal{N}^{\mathrm{e}}}\sum_{r\in\mathcal{S}_n^{\mathrm{e}}}(C_r^{\mathrm{EnInv}} + C^{\mathrm{EnFix}})y_{nr}^{\mathrm{eLev}} & \quad (1a)\\
+ \sum_{n\mathcal{N}^{\mathrm{e}}}\sum_{i\in\mathcal{P}} C_i^{\mathrm{dec}}x_{ni}^{\mathrm{dec}} & \quad (1b)\\
+ \sum_{n\in\mathcal{N}^{\mathrm{e}}}\sum_{i\in\mathcal{P}}\sum_{t\in\mathcal{T}^{\mathrm{e}}} w_t p_{nti} C_i^{\mathrm{var}} & \quad (1c)\\
+ \sum_{n\in\mathcal{N}^{\mathrm{e}}}\sum_{t\in\mathcal{T}^{\mathrm{e}}}\sum_{i\in\mathcal{P}} w_t C_i^{\mathrm{startUp}} x_{nti}^{\mathrm{up}} & \quad (1d)\\
+ \sum_{l\in\mathcal{L}^{\mathrm{e}}} C_\ell^{\mathrm{trans}} z_\ell^{\mathrm{eInv}} & \\
+ \sum_{\ell\in\mathcal{L}^{\mathrm{e}}:I_\ell^{\mathrm{trans}}=1} C_\ell^{\mathrm{trFix}} + \sum_{\ell\in\mathcal{L}^{\mathrm{e}}:I_\ell^{\mathrm{trans}}=0} C_\ell^{\mathrm{trFix}} z_\ell^{\mathrm{eInv}} & \quad (1e)\\
+ \sum_{n\in\mathcal{N}^{\mathrm{e}}} d_n C_{\mathrm{CO_2}}^{\mathrm{inv}}\kappa_n^{\mathrm{pipe}} + C_{\mathrm{CO_2}}^{\mathrm{str}}\sum_{n\in\mathcal{N}^{\mathrm{e}}}\sum_{t\in\mathcal{T}^{\mathrm{e}}} w_t\kappa_{nt}^{\mathrm{capt}} & \quad (1f)\\
+ \sum_{n\in\mathcal{N}^{\mathrm{e}}}\sum_{i\in\mathcal{P}}\sum_{t\in\mathcal{T}^{\mathrm{e}}} w_t p_{nti}(C_i^{\mathrm{fuel}}h_i) & \quad (1g)\\
+ \sum_{n\in\mathcal{N}^{\mathrm{e}}}\sum_{t\in\mathcal{T}^{\mathrm{e}}} w_t C_n^{\mathrm{eShed}} a_{nt}^{\mathrm{e}} & \quad (1h)
\end{aligned}
$$

The objective function (1) minimizes the total investment and operating costs incurred in the power system. The first term (1a) is the investment and fixed operation and maintenance (FOM) costs for generation and storage. The term (1b) captures the cost of plant retirement or decommissioning. The variable operating and maintenance (VOM) cost is represented by the term (1c). The fourth term (1d) corresponds to startup costs of power plants with unit commitment constraints (i.e. thermal power plants). The network expansion and FOM costs are included in term (1e). The cost of CO_2 transport and storage infrastructure required to accompany CCGT-CCS power generation is incorporated by term (1f) which also captures the cost associated with establishing CO_2 pipelines and storage. Here, we conservatively assume that each CO_2 pipeline connects a power node to the storage site, which ignores the possibility of meshed network design for CO_2 transport. The cost of fuel consumption for non NG-fired power plants (i.e., nuclear) is ensured by term (1g). The term (1h) penalizes the load shedding in the power system.

Investment and Unit Commitment: For every $n \in \mathcal{N}^{\mathrm{e}}, i \in \mathcal{P}$

$$
\begin{aligned}
x_{ni}^{\mathrm{op}} &= I_{ni}^{\mathrm{num}} - x_{ni}^{\mathrm{dec}} + x_{ni}^{\mathrm{est}} && && (2a)\\
x_{nti} - x_{n,t-1,i} &= x_{nti}^{\mathrm{up}} - x_{nti}^{\mathrm{down}} && t\in\mathcal{T}^{\mathrm{e}} && (2b)\\
x_{nti} &\le x_{ni}^{\mathrm{op}} && t\in\mathcal{T}^{\mathrm{e}} && (2c)
\end{aligned}
$$

Constraints (2a) model the number of operating plants. The unit commitment constraints for each node and plant type are presented in (2b) which computes the number of plants committed, started up, or shut down across the hours of the representative days. Constraints (2c) limit the number of committed units to the number of available ones. As per other similar studies [28, 35, 21, 43], we relaxed the integrality of unit

commitment decisions which has been shown to introduce relatively small numerical error while significantly reducing the computational complexity of the problem.
Generation, Ramping, and Load Shedding: For every $n \in \mathcal{N}^{\mathrm{e}}, t \in \mathcal{T}^{\mathrm{e}}$

$$L_i^{\mathrm{prod}} x_{nti} \leq p_{nti} \leq U_i^{\mathrm{prod}} x_{nti} \qquad i \in \mathcal{H} \tag{3a}$$

$$|p_{nti} - p_{n,(t-1),i}| \leq U_i^{\mathrm{ramp}} U_i^{\mathrm{prod}} (x_{nti} - x_{nti}^{\mathrm{up}}) + \max(L_i^{\mathrm{prod}}, U_i^{\mathrm{ramp}}) U_i^{\mathrm{prod}} x_{nti}^{\mathrm{up}} \qquad i \in \mathcal{H} \tag{3b}$$

$$p_{nti} \leq \rho_{nti} U_i^{\mathrm{prod}} x_{ni}^{\mathrm{op}} \qquad i \in \mathcal{R} \tag{3c}$$

$$a_{nt}^{\mathrm{e}} \leq D_{n\phi_t^{\mathrm{e}}}^{\mathrm{e}} \tag{3d}$$

The generation limits are imposed in constraints (3a). Constraints (3b) are the ramping constraints that limit the generation difference of thermal units in any consecutive time periods to a ramping limit in the right-hand-side of the equation. The generation pattern of VREs is determined by their hourly profile in the form of capacity factor; constraints (3c) limit the generation of VRE to hourly capacity factor (i.e. ρ_{nti}) of maximum available capacity (i.e. $U_i^{\mathrm{prod}} x_{ni}^{\mathrm{op}}$). Constraints (3d) state that the load shedding amount can not exceed demand. Note that we use the mapping ϕ_t^{e} to access the demand in the corresponding hour of a representative day.
Power Balance Constraints: For every $n \in \mathcal{N}^{\mathrm{e}}, t \in \mathcal{T}^{\mathrm{e}}$

$$\sum_{i \in \mathcal{P}} p_{nti} + \sum_{m \in \mathcal{N}^{\mathrm{e}}} \sum_{l \in \mathcal{L}_{nm}^{\mathrm{e}}} \mathrm{sign}(n-m) f_{\ell t}^{\mathrm{e}} + \sum_{r \in \mathcal{S}_n^{\mathrm{e}}} (s_{ntr}^{\mathrm{eDis}} - s_{ntr}^{\mathrm{eCh}}) + a_{nt}^{\mathrm{e}} = D_{n\phi_t^{\mathrm{e}}}^{\mathrm{e}} + d_n E^{\mathrm{pipe}} \kappa_n^{\mathrm{pipe}} + E^{\mathrm{cprs}} E^{\mathrm{pump}} \kappa_{nt}^{\mathrm{capt}} \tag{4}$$

Constraints (4) ensures that for each node and for each planning period the generation, the net flow, the net storage, and the load shedding amount should be equal to the net demand. The net demand is defined in the right-hand-side where the first term is the baseline demand, the second term is the electricity consumption by CO_2 pipelines and the last term is the electricity used by compressors. The notation sign$(n-m)$ is the *sign* function that takes value -1 if $n < m$, value 1 if $n > m$, and 0 otherwise. We use this function to ensure that $f_{\ell t}^{e}$ appears with opposite signs (i.e., negative of positive signs) in the balance equations of the nodes connected by transmission line ℓ.
Network Constraints: For every $l \in \mathcal{L}^{\mathrm{e}}, t \in \mathcal{T}^{\mathrm{e}}$, and $n, m \in \mathcal{N}_\ell^{\mathrm{e}}$

$$|f_{lt}^{\mathrm{e}}| \leq I_\ell^{\mathrm{trans}} \qquad \text{if } \mathcal{I}_\ell^{\mathrm{trans}} = 1 \tag{5a}$$

$$|f_{lt}^{\mathrm{e}}| \leq U_\ell^{\mathrm{trans}} z_\ell^{\mathrm{eInv}} \qquad \text{if } \mathcal{I}_\ell^{\mathrm{trans}} = 0 \tag{5b}$$

$$|f_{\ell t}^{\mathrm{e}} - b_\ell(\theta_{mt} - \theta_{nt})| \leq M(1 - z_\ell^{\mathrm{eInv}}) \qquad \text{if } \mathcal{I}_\ell^{\mathrm{trans}} = 0 \tag{5c}$$

$$f_{\ell t}^{\mathrm{e}} = b_\ell(\theta_{mt} - \theta_{nt}) \qquad \text{if } \mathcal{I}_\ell^{\mathrm{trans}} = 1 \tag{5d}$$

$$\theta_{0,t} = 0 \tag{5e}$$

Flow for the existing transmission lines is limited by constraints (5a). Constraints (5b) limits the flow in candidate transmission lines only if it is already established (i.e., z_ℓ^{t}=1). Throughout the paper, we use M to denote a big number. Direct Current (DC) power flow constraints for candidate and existing transmission lines are respectively imposed in constraints (5c) and (5d). The phase angle for the reference node 0 is set in constraint (5e).

Storage Constraints: For every $n \in \mathcal{N}^{\mathrm{e}}, r \in \mathcal{S}_n^{\mathrm{e}}$

$$s^{\mathrm{eLev}}_{nt^{\mathrm{start}}_{\tau}r} = (1-\gamma^{\mathrm{loss}}_r)(s^{\mathrm{eLev}}_{n,t^{\mathrm{end}}_{\tau}r} - s^{\mathrm{rem}}_{n\tau r}) + \gamma^{\mathrm{eCh}}_r s^{\mathrm{eCh}}_{nt^{\mathrm{start}}_{\tau}r} - \frac{s^{\mathrm{eDis}}_{nt^{\mathrm{start}}_{\tau}r}}{\gamma^{\mathrm{eDis}}_r}, \qquad \tau \in \mathfrak{R} \tag{6a}$$

$$s^{\mathrm{eLev}}_{ntr} = (1-\gamma^{\mathrm{loss}}_r)(s^{\mathrm{eLev}}_{ntr}) + \gamma^{\mathrm{eCh}}_r s^{\mathrm{eCh}}_{ntr} - \frac{s^{\mathrm{eDis}}_{ntr}}{\gamma^{\mathrm{eDis}}_r} \qquad t \in \mathcal{T}^{\mathrm{e}} \backslash \{t^{\mathrm{start}}_{\tau} |\ \tau \in \mathfrak{R}\} \tag{6b}$$

$$s^{\mathrm{day}}_{n,\tau+1,r} = (1-24\gamma^{\mathrm{loss}}_r)s^{\mathrm{day}}_{n,\tau,r} + s^{\mathrm{rem}}_{n\Omega_{\tau},r}, \quad \tau \in \mathcal{T}^{\mathrm{g}} \backslash 365 \tag{6c}$$

$$s^{\mathrm{day}}_{n,1,r} = (1-24\gamma^{\mathrm{loss}}_r)s^{\mathrm{day}}_{n,\tau,r} + s^{\mathrm{rem}}_{n\Omega_{\tau},r}, \quad \tau = 365 \tag{6d}$$

$$s^{\mathrm{day}}_{n\tau r} = s^{\mathrm{eLev}}_{nt^{\mathrm{end}}_{\tau}r} - s^{\mathrm{rem}}_{n\tau r}, \quad \tau \in \mathfrak{R} \tag{6e}$$

$$s^{\mathrm{rem}}_{n\tau r} = 0, \quad \tau \in \mathfrak{R}, r \in \mathcal{S}^{\mathrm{sS}} \tag{6f}$$

$$s^{\mathrm{eCh}}_{ntr} \leq y^{\mathrm{eCD}}_{nr} \tag{6g}$$

$$s^{\mathrm{eCh}}_{ntr} \leq y^{\mathrm{eCD}}_{nr} \tag{6h}$$

$$s^{\mathrm{eLev}}_{ntr} \leq y^{\mathrm{eLev}}_{nr} \tag{6i}$$

Recall (see Fig. 1) that representative days are not necessarily consecutive; hence our formulation accounts for the carryover storage level between representative days, which is particularly important when modeling LDES. Li et al. [21] enforce the beginning and ending storage levels of each representative day to 50% of the maximum storage level. Here, we use a similar approach for short-duration batteries in which we time-wrap the beginning and ending hours of a day. That is, we assume the same charging state for the beginning and ending hours of a day, implicitly precluding energy carryover between representative days. For LDES, however, we use the method proposed in [18, 20] in which the unrestricted variable $s^{\mathrm{rem}}_{n\tau r}$ models the carryover from a representative day τ to the next.

Constraints (6a) model battery storage dynamics for the initial hours of each representative day. Constraints (6b) model the storage balance for the remaining hours. The energy transfer across two consecutive representative days is modeled via constraints (6c). The storage in the first and last representative days is related by constraints (6d). Constraint (6e) only applies to representative days and ensures that the storage at the beginning of a day is equal to the storage level at the last hour of the day minus the storage carryover. The storage carryover for short-duration batteries is prevented by constraints (6f). Finally, the charge/discharge limits on storage level are imposed in constraints (6g) to (6i). Analogous to similar studies on power system expansion [35, 21], we do not account for use-dependent storage capacity degradation.

Renewable Portfolio Standards (RPS):

$$\sum_{n \in \mathcal{N}^{\mathrm{e}}} \sum_{t \in \mathcal{T}^{\mathrm{e}}} \sum_{i \in \mathcal{R}} p_{nti} \geq L^{\mathrm{RPS}} \sum_{n \in \mathcal{N}^{\mathrm{e}}} \sum_{t \in \mathcal{T}^{\mathrm{e}}} D^{\mathrm{e}}_{n\phi^{\mathrm{e}}_t} \tag{7}$$

RPS is a policy constraint that requires a minimum share of the total electricity to be met by renewable sources and has been one of the core policies driving decarbonization efforts in many U.S. and global regions [16, 17]. In the model, constraint (7) requires that annual dispatched renewable generation must be at least a pre-specified fraction of total annual demand, specified by the parameter, L^{RPS}.

Resource Availability Constraints:

$$\sum_{n\in\mathcal{N}^{\mathrm{e}}}\sum_{i\in\mathcal{Q}} U_i^{\mathrm{prod}} x_{ni}^{\mathrm{op}} \leq U_{\mathcal{Q}}^{\mathrm{prod}} \qquad \mathcal{Q}\in\mathcal{Q}' \tag{8}$$

We consider resource availability limits for the development of VRE sources. In comparison to thermal plants, the siting of renewable resources is a major challenge due to the relatively large land area footprint per MW, the spatial heterogeneity in their resource availability and land availability limits due to non-energy considerations such as preserving the natural landscape [16]. Therefore, constraint (8) limits the installed capacity of a certain set of power plants to their maximum availability limit. The parameter $\mathcal{Q}$ denotes a generation technology class for which there is a resource availability limit. These classes include solar, onshore wind, offshore wind, and nuclear and are represented by set $\mathcal{Q}'$. Note that each technology class can include multiple plant types. For example, nuclear technology can include existing and new nuclear plant types.

Carbon Capture and Storage (CCS) Constraints: For every $n\in\mathcal{N}^{\mathrm{e}}, t\in\mathcal{T}^{\mathrm{e}}$

$$\kappa_{nt}^{\mathrm{capt}} = \eta^{\mathrm{g}}\eta_i h_i p_{nti} \qquad i\in\mathcal{CCS} \tag{9a}$$

$$\kappa_{nt}^{\mathrm{capt}} \leq \kappa_n^{\mathrm{pipe}} \tag{9b}$$

$$\sum_{n\in\mathcal{N}^{\mathrm{e}}}\sum_{t\in\mathcal{T}^{\mathrm{e}}} \kappa_{nt}^{\mathrm{capt}} \leq U^{\mathrm{CCS}} \tag{9c}$$

The constraint (9a) computes the amount of captured carbon in gas-fired power plants equipped with CCS technology. Constraint (9b) determines the CO_2 pipeline capacity. Finally, constraint (9c) limits the total amount of captured CO_2 to the annual CO_2 storage capacity.

Capacity Reserve Margin:

$$\sum_{n\in\mathcal{N}^{\mathrm{e}}}\left(\sum_{i\in\mathcal{P}}\gamma_{nit}^{\mathrm{CRM}} U_i^{\mathrm{prod}} x_{ni}^{\mathrm{op}} + \sum_{r\in\mathcal{S}_n^{\mathrm{e}}}(s_{ntr}^{\mathrm{eDis}} - s_{ntr}^{\mathrm{eCh}})\right) \geq (1+R^{\mathrm{CRM}})\sum_{n\in\mathcal{N}^{\mathrm{e}}} D_{n\phi_t^{\mathrm{t}}}, \quad t\in\mathcal{T}^{\mathrm{e}} \tag{10a}$$

1.2 NG System Model

We now model the objective function and constraints pertaining to the NG system in the JPoNG.

Objective Function:

$$\min \sum_{l\in\mathcal{L}^{\mathrm{g}}}\left(C_\ell^{\mathrm{pipe}} z_\ell^{\mathrm{g}} + c_\ell^{\mathrm{pipeDec}} z^{\mathrm{gDec}}\ell + c_\ell^{\mathrm{pipeFix}} z_\ell^{\mathrm{gOp}}\right) \tag{11a}$$

$$+ \sum_{k\in\mathcal{N}^{\mathrm{g}}}\sum_{\tau\in\mathcal{T}^{\mathrm{g}}} C^{\mathrm{ng}} g_{k\tau} \tag{11b}$$

$$+ \sum_{j\in\mathcal{N}^{s}}(C_j^{\mathrm{strInv}} x_j^{\mathrm{gStr}} + C_j^{\mathrm{vprInv}} x_j^{\mathrm{vpr}}) \tag{11c}$$

$$+ \sum_{j\in\mathcal{N}^{s}}\left(C^{\mathrm{strFix}}(I_j^{\mathrm{gStr}} + x_j^{\mathrm{gStr}}) + C^{\mathrm{vprFix}}(I_j^{\mathrm{vpr}} + x_j^{\mathrm{vpr}})\right) \tag{11d}$$

$$+ \sum_{k\in\mathcal{N}^{\mathrm{g}}}\sum_{\tau\in\mathcal{T}^{\mathrm{g}}}(C^{\mathrm{LCDF}} a_{k\tau}^{\mathrm{LCDF}} + C^{\mathrm{gShed}} a_{k\tau}^{\mathrm{ng}}) \tag{11e}$$

The objective function (11) minimizes the total investment and operating costs incurred in the NG system. The first term (11a) is the investment, strategic decommissioning and FOM costs for new, existing and operational pipelines, respectively. The second term (11b) is the cost of procuring NG from various sources to the system. For example, New England procures its NG from Canada and its adjacent states such as New York. The term (11c) and (11d) handle the investment and FOM costs associated with NG storage, respectively. The last term (11e) captures the cost of using LCDF and NG load shedding.

NG Balance Constraint: For every $k \in \mathcal{N}^{\mathrm{g}}, \tau \in \mathcal{T}^{\mathrm{g}}$

$$\begin{aligned} & g_{k\tau} - \sum_{l \in \mathcal{L}_k^{\mathrm{gExp}}} f_{\ell\tau}^{\mathrm{g}} + \sum_{l \in \mathcal{L}_k^{\mathrm{gImp}}} f_{\ell\tau}^{\mathrm{g}} - \sum_{n \in \mathcal{A}_k^{\mathrm{e}}} f_{kn\tau}^{\mathrm{ge}} \\ & + \sum_{j \in \mathcal{A}_k^{\mathrm{s}}} (f_{jk\tau}^{\mathrm{vg}} - f_{kj\tau}^{\mathrm{gl}}) + a_{k\tau}^{\mathrm{LCDF}} + a_{k\tau}^{\mathrm{g}} = D_{k\tau}^{\mathrm{g}} \end{aligned} \tag{12}$$

Constraints (12) state that for each node and period, the imported NG (i.e., injection), flow to other NG nodes, flow to power nodes, flow to and from storage nodes, load satisfied by LCDF, and unsatisfied NG load should add up to demand. Unlike power flow, the flow in pipelines is assumed to be unidirectional as it is typical for most long-distance transmission pipelines involving booster compressor stations [43]. We are ignoring the relatively small electricity demand associated with booster compression stations along the NG pipeline network.

Flow on Representative Days:

$$f_{kn\tau_1}^{\mathrm{ge}} = f_{kn\tau_2}^{\mathrm{ge}} \qquad \tau_1, \tau_2 \in \mathcal{T}^{\mathrm{g}} \text{ if } \Omega_{\tau_1} = \Omega_{\tau_2} \tag{13}$$

Given the set of representative days used to model power system operations (see Fig 1), constraint (13) ensures that gas consumption by the power system for all the days represented by the same day is identical.

Gas and LCDF Supply Constraints: For every $k \in \mathcal{N}^{\mathrm{g}}, \tau \in \mathcal{T}^{\mathrm{g}}$

$$L_k^{\mathrm{inj}} \leq g_{k\tau} + a_{k\tau}^{\mathrm{LCDF}} \leq U_k^{\mathrm{inj}} \tag{14a}$$

$$\tag{14b}$$

The fuel import limits are imposed in constraints (14a).

Flow Constraints: For every $\ell \in \mathcal{L}^{\mathrm{g}}, \tau \in \mathcal{T}^{\mathrm{g}}, j \in \mathcal{N}^{s}$

$$f_{\ell\tau}^{\mathrm{g}} \leq U_\ell^{\mathrm{pipe}} z_\ell^{\mathrm{gOp}} \tag{15a}$$

$$\sum_{k \in \mathcal{N}^{\mathrm{g}}: j \in \mathcal{A}_k^{\mathrm{s}}} f_{kj\tau}^{\mathrm{gl}} = s_{j\tau}^{\mathrm{liq}} \tag{15b}$$

$$\sum_{k \in \mathcal{N}^{\mathrm{g}}: j \in \mathcal{A}_k^{\mathrm{s}}} f_{jk\tau}^{\mathrm{vg}} = s_{j\tau}^{\mathrm{vpr}} \tag{15c}$$

The constraints (15a) limit the flow between NG nodes for operational pipelines, respectively. The flow to liquefaction facilities is calculated in constraints (15b). Similarly, the flow out of vaporization facilities is modeled via constraints (15c).

Storage Constraints: For every $j \in \mathcal{N}^s, \tau \in \mathcal{T}^{\mathrm{g}}$

$$s_{j\tau}^{\mathrm{gStr}} = (1-\beta)s_{j,\tau-1}^{\mathrm{gStr}} + \gamma_j^{\mathrm{liqCh}} s_{j\tau}^{\mathrm{liq}} - \frac{s_{j\tau}^{\mathrm{vpr}}}{\gamma_j^{\mathrm{vprDis}}} \tag{16a}$$

$$s_{j\tau}^{\mathrm{vpr}} \leq I_j^{\mathrm{vpr}} + x_j^{\mathrm{vpr}} \tag{16b}$$

$$s_{j\tau}^{\mathrm{gStr}} \leq I_j^{\mathrm{gStr}} + x_j^{\mathrm{gStr}} \tag{16c}$$

Constraints (16a) ensure the storage balance. Constraints (16b) and (16c) limit the capacity of vaporization and storage tanks to their initial capacity plus the increased capacity, respectively.

Operational Pipelines:

$$z_\ell^{\mathrm{gOp}} = \mathcal{I}^{\mathrm{pipe}} + z_\ell^{\mathrm{gInv}} - z_\ell^{\mathrm{gDec}} \tag{17a}$$

Pipeline ℓ is operational if either it is existing and not decommissioned, or newly established.

1.3 Coupling Constraints

The following constraints are coupling constraints that relate decisions of the two systems.

$$\sum_{k \in \mathcal{A}_n^{\mathrm{e}}} f_{kn\tau}^{\mathrm{ge}} = \sum_{t \in \mathfrak{T}_\tau^{\mathrm{e}}} \sum_{i \in \mathcal{G}} h_i p_{nti} \qquad n \in \mathcal{N}^{\mathrm{e}}, \tau \in \mathfrak{R} \tag{18a}$$

$$\mathcal{E}^{\mathrm{e}} = \sum_{n \in \mathcal{N}^{\mathrm{e}}} \sum_{t \in \mathcal{T}^e} \sum_{i \in \mathcal{G}} w_t (1-\eta_i) \eta^{\mathrm{g}} h_i p_{nti}$$

$$\mathcal{E}^{\mathrm{g}} = \sum_{k \in \mathcal{N}^g} \sum_{\tau \in \mathcal{T}^g} \eta^{\mathrm{g}} (D_{k\tau}^{\mathrm{g}} - a_{k\tau}^{\mathrm{LCDF}} - a_{k\tau}^{\mathrm{g}})$$

$$\mathcal{E}^{\mathrm{e}} + \mathcal{E}^{\mathrm{g}} \leq (1-\zeta)(U_{\mathrm{emis}}^{\mathrm{e}} + U_{\mathrm{emis}}^{\mathrm{g}}) \tag{18b}$$

The first coupling constraints (18a) captures the flow of NG to the power network for each node and at each time period. The variable $\mathcal{E}^{\mathrm{e}}$ accounts for CO_2 emission due to the consumption of NG in the power system. The variable $\mathcal{E}^{\mathrm{g}}$ computes the emission from NG system by subtracting the demand from LCDF consumption and gas load shedding. The second coupling constraint (18b) ensures that the net CO_2 emissions associated with the power-gas system are below a pre-specified threshold value, which is defined based on a baseline emissions level. The first term is the emissions due to non-power NG consumption (i.e., NG consumption in the NG system such as space heating, industry use, and transportation). Since the model does not track whether LCDF is used to meet non-power NG demand or for power generation, the first term computes gross emissions from all NG use presuming it is all fossil and then subtracts emissions benefits from using LCDF.

Here we treat LCDF as a carbon-neutral fuel source [9], and thus the combustion emissions associated with its end-use are equal to the emissions captured during its production. The second term captures the emission from NG-fired power plants. Alternatively, the emission constraints can only be applied to the power system as in [35] or separately applied to each system as in [43].

2 Nomenclature

Sets	
$\mathcal{N}^{\mathrm{e}}$	Power system nodes
$\mathcal{P}$	Power plant types
$\mathcal{R} \subset \mathcal{P}$	VRE power plant types
$\mathcal{G} \subset \mathcal{P}$	gas-fired plant types
$\mathcal{CCS} \subset \mathcal{P}$	gas-fired plant types with carbon capture technology
$\mathcal{H} \subset \mathcal{P}$	Thermal plant types
$\mathcal{Q} \subset \mathcal{P}$	Technology with a resource availability limit
$\mathcal{Q}'$	Set of technologies with resource availability limits
$\mathcal{T}^{\mathrm{e}}$	Index set of representative hours for power system
$\mathfrak{R}$	Representative days
$\mathfrak{T}^{\mathrm{e}}_{\tau}$	Hours in $\mathcal{T}^{\mathrm{e}}$ that are represented by day τ
$t^{\mathrm{start}}_{\tau}, t^{\mathrm{end}}_{\tau}$	First and last hour in $\mathfrak{T}^{\mathrm{e}}_{\tau}$
$\mathcal{L}^{\mathrm{e}}$	Existing and candidate transmission lines
$\mathcal{L}^{\mathrm{e}}_{nm}$	Existing and candidate transmission lines
$\mathcal{S}^{\mathrm{eS}}_{n}$	Short-duration energy storage system types
$\mathcal{S}^{\mathrm{eL}}_{n}$	Long-duration energy storage system types
$\mathcal{S}^{\mathrm{e}}_{n}$	All energy storage systems types
$\mathcal{A}^{\mathrm{g}}_{n}$	Adjacent NG nodes for node n between node n and m
$\mathcal{N}^{\mathrm{g}}, \mathcal{N}^{s}$	NG and SVL nodes
$\mathcal{T}^{\mathrm{g}}$	Days of the planning year
$\mathcal{A}^{s}_{k}$	Adjacent SVL facilities of node k
$\mathcal{L}^{\mathrm{g}}$	Existing and candidate pipelines
$\mathcal{L}^{\mathrm{gExp}}_{k}$	Existing and candidate pipelines starting from node k
$\mathcal{L}^{\mathrm{gImp}}_{k}$	Existing and candidate pipelines ending at node k

Indices	
n, m	Power system node
k	NG system node
j	SVL facility node
i	Power generation plant type
r	Storage type for power network
ℓ	Electricity transmission line or gas pipeline
t	Time step for power system's operational periods
τ	Time step for NG system's operational periods

Annualized Cost Parameters	
C^{inv}_{i}	CAPEX of plants, [$/plant]
C^{dec}_{i}	Plant decommissioning cost, [$/plant]
$C^{\mathrm{trans}}_{\ell}$	Transmission line establishment cost, [$/line]
C^{EnInv}_{r}	Storage establishment energy-related cost, [$/MWh]
C^{pInv}_{r}	Storage establishment power-related cost, [$/MW]
C^{pipe}_{ℓ}	Pipelines establishment cost, [$/line]
C^{strInv}_{j}	CAPEX of storage tanks at SVLs, [$/MMBtu]

C_j^{vprInv}	CAPEX of vapor. plants at SVLs, [\$/MMBtu/hour]
C_ℓ^{pipeDec}	Decommissioning cost for pipeline ℓ [\$/line]

Annual Costs

C_i^{fix}	Fixed operating and maintenance cost (FOM) for plants, [\$]
C_r^{EnFix}	Energy-related FOM for storage, [\$/MWh]
C_r^{pFix}	Power-related FOM for storage, [\$/MW]
C_ℓ^{trFix}	Fixed cost of transmission line ℓ [\$/line]
C_j^{strFix}	FOM for storage tanks, [\$/MMBtu]
C_j^{vprFix}	FOM for vaporization plants, [\$/MMBtu/hour]
C_ℓ^{pipeFix}	Fixed cost of pipeline ℓ [\$/line]

Other Cost Parameters

C_i^{var}	Variable operating and maintenance cost (VOM) for plants, [\$/MWh]
C_i^{startUp}	Start-up cost for plants, [\$]
C^{eShed}	Unsatisfied power demand cost, [\$/MWh]
C_i^{fuel}	Fuel price for plants, [\$/MMBtu]
C^{ng}	Fuel price for NG, [\$/MMBtu]
C^{LCDF}	Price of LCDF, [\$/MMBtu]
C^{gShed}	Unsatisfied NG demand cost [\$/MMBtu]

Other Parameters for the Power System

ρ_{nti}	Capacity (availability) factor for renewable plants
D_{nt}^{e}	Power demand, [MWh]
h_i	Heat rate, [MMBtu/MWh]
b_ℓ	Susceptance of line $\ell \in \mathcal{L}^{\text{e}}$
η_i	Carbon capture rate, [%]
U_i^{prod}	Nameplate capacity, [MW]
L_i^{prod}	Minimum stable output, [%]
U_i^{ramp}	Ramping limit, [%]
γ_r^{eCh}	Charge rate for storage
γ_r^{eDis}	Discharge rate for storage
γ_r^{loss}	hourly self-discharge rate for storage
I_ℓ^{trans}	Initial capacity for transmission line ℓ, [MW]
U_ℓ^{trans}	Upper bound for capacity of transmission line ℓ, [MW]
$\mathcal{I}_\ell^{\text{trans}}$	1, if trans. line ℓ exists; 0, otherwise
I_{ni}^{num}	Initial number of plants
$U_{\text{emis}}^{\text{e}}$	Baseline emission of CO_2 in 1990 from generation consumption, [ton]
U^{CCS}	Total annual carbon storage capacity, [ton]
L^{RPS}	Renewable Portfolio Standard (RPS) value
d_n	Distance between node n and CO_2 storage site
E^{pipe}	Electric requirement for CO_2 pipeline operations [MWh/mile/ton/hour]

E^{pump}	Electric requirement for compression of CO_2
E^{cprs}	Number of compressors required in the pipeline from node n to the storage site pipelines [MWh/ton/hour]
$U_{\mathcal{Q}}^{\text{prod}}$	Production capacity for set of plants $\mathcal{Q} \subset \mathcal{P}$, [MW]
ζ	emissions reduction goal
w_t	Weight of the representative period t
ϕ_t^{e}	Mapping of representative period t to its original period in the time series
R^{CRM}	Capacity reserve margin rate
$\gamma_{nit}^{\text{CRM}}$	Capacity derating factor of plant type i at node n at time t

Other Parameters for the NG System

$D_{k\tau}^{\text{g}}$	NG demand, [MMBtu]
η^{g}	Emission factor for NG [ton CO_2/MMBtu]
U_k^{inj}	Upper bound for NG supply, [MMBtu]
γ_j^{liqCh}	Charge efficiency of liquefaction plant
γ_j^{vprDis}	Discharge efficiency of vaporization plant
β	Boil-off gas coefficient
I_ℓ^{pipe}	Initial capacity for pipeline ℓ, [MMBtu/day]
U_ℓ^{pipe}	Upper bound capacity for pipeline ℓ, [MMBtu/day]
$\mathcal{I}_\ell^{\text{pipe}}$	1, if the pipeline ℓ exists; 0, otherwise
I_j^{gStr}	Initial storage capacity, [MMBtu]
I_j^{vpr}	Initial vaporization capacity, [MMBtu/d]
I_j^{liq}	Initial liquefaction capacity, [MMBtu/d]
I_{kj}^{store}	Initial capacity of storage facility
$U_{\text{emis}}^{\text{g}}$	Baseline emission of CO_2 in 1990 from non-generation consumption, [ton]
Ω_n	representative day for day n

Investment Decision Variables

$x_{ni}^{\text{op}} \in \mathbb{Z}^+$	Number of available thermal plants $i \in \mathcal{H}$
$x_{ni}^{\text{op}} \in \mathbb{R}^+$	Number of available VRE plants $i \in \mathcal{R}$
$x_{ni}^{\text{est}} \in \mathbb{Z}^+$	Number of new thermal plants established, $i \in \mathcal{H}$
$x_{ni}^{\text{est}} \in \mathbb{R}^+$	Number of new VRE plants established, $i \in \mathcal{R}$
$x_{ni}^{\text{dec}} \in \mathbb{Z}^+$	Number decommissioned thermal plants, $i \in \mathcal{H}$
$x_{ni}^{\text{dec}} \in \mathbb{R}^+$	Number decommissioned VRE plants, $i \in \mathcal{R}$
$y_{nr}^{\text{eCD}} \in \mathbb{R}^+$	Charge/discharge capacity of storage battery
$y_{nr}^{\text{eLev}} \in \mathbb{R}^+$	Battery storage level
$z_\ell^{\text{eInv}} \in \mathbb{B}$	1, if transmission line ℓ is built; 0, otherwise
$z_\ell^{\text{gInv}} \in \mathbb{B}$	1, if pipeline ℓ is built; 0, otherwise
$z_\ell^{\text{gDec}} \in \mathbb{B}$	1, if pipeline ℓ is decommissioned; 0, otherwise
$z_\ell^{\text{gOp}} \in \mathbb{B}$	1, if pipeline ℓ is operational; 0, otherwise

Other Decision Variables for Power System

$p_{nti} \in \mathbb{R}^+$	Generation rate, [MW]
$x_{nti} \in \mathbb{R}^+$	Number of committed plants

$x_{nti}^{\text{down}} \in \mathbb{R}^+$	Number of plants shut-down
$x_{nti}^{\text{up}} \in \mathbb{R}^+$	Number of plants started up
$f_{\ell t}^{\text{e}} \in \mathbb{R}$	Flow rates, [MW]
$\theta_{nt} \in \mathbb{R}$	Phase angle
$s_{ntr}^{\text{eCh}} \in \mathbb{R}^+$	Storage charged, [MW]
$s_{ntr}^{\text{eDis}} \in \mathbb{R}^+$	Storage discharged, [MW]
$s_{ntr}^{\text{eLev}} \in \mathbb{R}^+$	Storage level, [MWh]
$s_{n\tau r}^{\text{rem}} \in \mathbb{R}$	Storage carry over during day τ for storage type $r \in \mathcal{S}^{\text{sL}}$
$s_{n\tau r}^{\text{day}} \in \mathbb{R}^+$	Storage level at the beginning of day τ for storage type $r \in \mathcal{S}^{\text{sL}}$
$\kappa_{nt}^{\text{capt}} \in \mathbb{R}^+$	Captured CO_2 [ton/h]
$\kappa_{n}^{\text{pipe}} \in \mathbb{R}^+$	CO_2 pipeline capacity [ton/h]
$a_{nt}^{\text{e}} \in \mathbb{R}^+$	Amount of load shedding, [MWh]
$\mathcal{E}^{\text{e}}$	Total emission from power system

Other Decision Variables for NG System (all in MMBtu)	
$x_{j}^{\text{gStr}} \in \mathbb{R}^+$	Installed additional storage capacities
$x_{j}^{\text{vpr}} \in \mathbb{R}^+$	Installed additional vaporization capacities
$f_{\ell\tau}^{\text{g}} \in \mathbb{R}^+$	Flow rates
$f_{kn\tau}^{\text{ge}} \in \mathbb{R}^+$	Flow rates from NG nodes to power nodes
$f_{kj\tau}^{\text{gl}} \in \mathbb{R}^+$	Flow rates from node NG nodes to liquefaction plants
$f_{jk\tau}^{\text{vg}} \in \mathbb{R}^+$	Flow rates from vaporization plants to NG nodes
$g_{k\tau} \in \mathbb{R}^+$	NG supply (injection)
$s_{j\tau}^{\text{gStr}} \in \mathbb{R}^+$	Storage capacities
$s_{j\tau}^{\text{vpr}}, s_{j\tau}^{\text{liq}} \in \mathbb{R}^+$	Vaporization and liquefaction amounts
$a_{k\tau}^{\text{g}} \in \mathbb{R}^+$	Amount of load shedding
$a_{k\tau}^{\text{LCDF}} \in \mathbb{R}^+$	Amount of LCDF consumption
$\mathcal{E}^{\text{g}}$	Total emission from NG system

3 Problem Data

This section describes how we obtained parameters for both networks for the New England Case study. We start with the power system and explain the data preparation process. We then expound on steps we took to construct the natural gas (NG) network and associated data inputs. In this study, our region is New England. However, the modeling framework and the solution approach are applicable to other regions to any spatial resolution.

Fig. 3 shows the name and boundaries of New England counties. The region currently has 67 counties with population ranging from 7 thousand to more than 1.6 million.

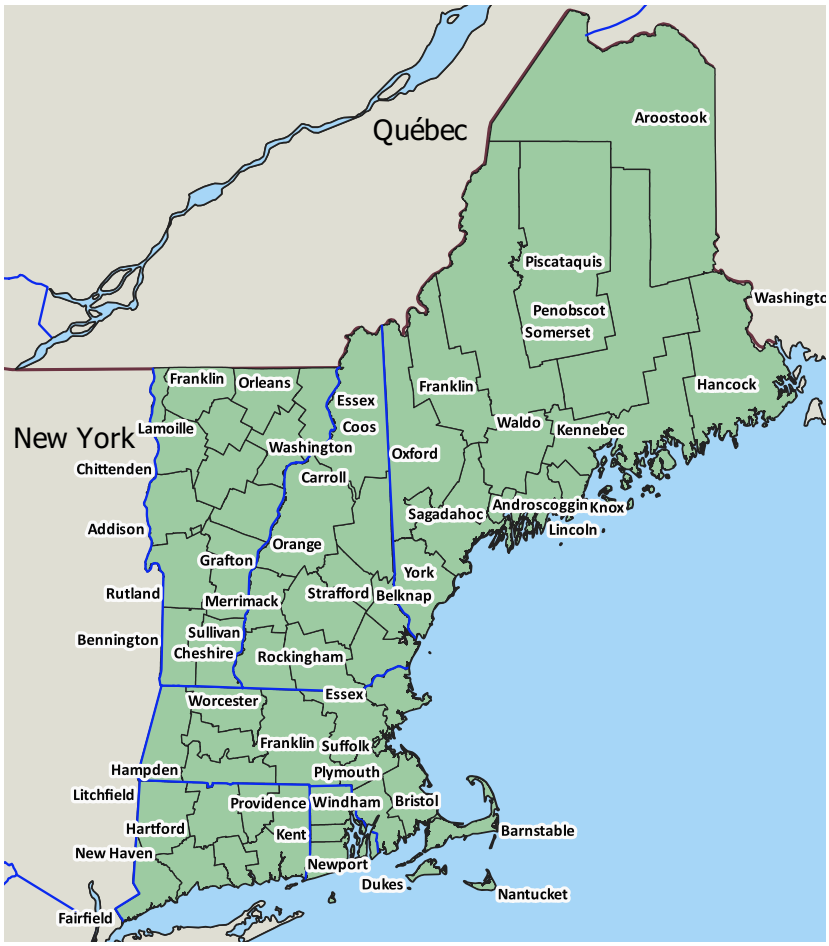

Figure 3: New England counties

4 Power System Data

Our parameters for the power network is based on the *US Test System* developed by *Breakthrough Energy* [45]. The dataset contains high-temporal and spatial resolution load and variable renewable energy (VRE) data for the entire US "base_grid" in the year 2016. It also contains load data for the year 2020 and projections for the year 2030. The dataset provides a test system with detailed information for existing buses, substations, plants, branches, and generation profiles for existing renewable energy capacity, including solar and land-based wind.

The large-scale offshore wind power generation is only projected for coastal regions of Massachusetts and Connecticut, hence we only allow the establishment of offshore wind plants at nodes 1, 2, 3, 4, 5, 14, 15, and 17.

General Note about Tables: The column of tables shows the associated symbol in the formulation. The notations with *tilde* are *crude* cost values whose manipulated forms (e.g., regional update, annualization) are used in the numerical model. For example, $\tilde{C}_i^{\text{inv}}$ in Table 3 denote the value of C_i^{inv} before annualization.

4.1 Power Nodes

The power network has 17 nodes, each corresponding to a group of counties with more than 600,000 aggregate population. The number of nodes is selected to sufficiently capture the network effect without significantly compromising the problem scalability. All counties of a group are in the same state and form a contiguous landmass. Fig 4 shows the grouping of counties and the power node associated with each group in the left figure and illustrates the power node without county grouping in the right figure.

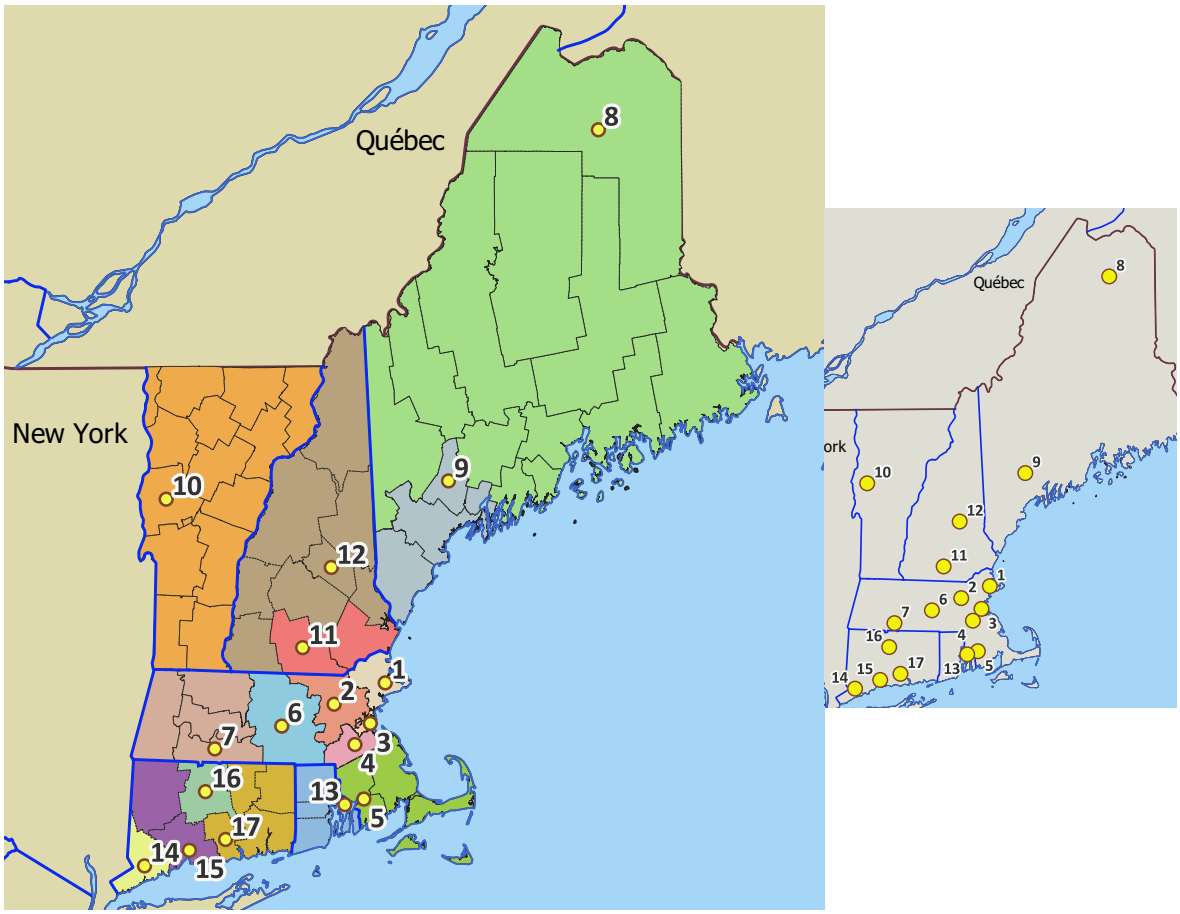

Figure 4: Grouping of New England counties, and the location of power nodes

4.2 Transmission Lines

In the *Breakthrough Energy* [45] data, we consider the base_grid and start by filtering the data for New England region which corresponds to zone 1 to 6 in the dataset. We then filter for high voltage buses [6] i.e., those with voltage greater than or equal to 345kV. This process results in 188 nodes, each of which represents a high-voltage bus. The filtering process results in 192 transmission lines with known susceptance and maximum flow limit. We use this data to identify the existing lines between power nodes. We first assign each of the 188 buses to its nearest power node. An existing line is then assigned between two power nodes if they were

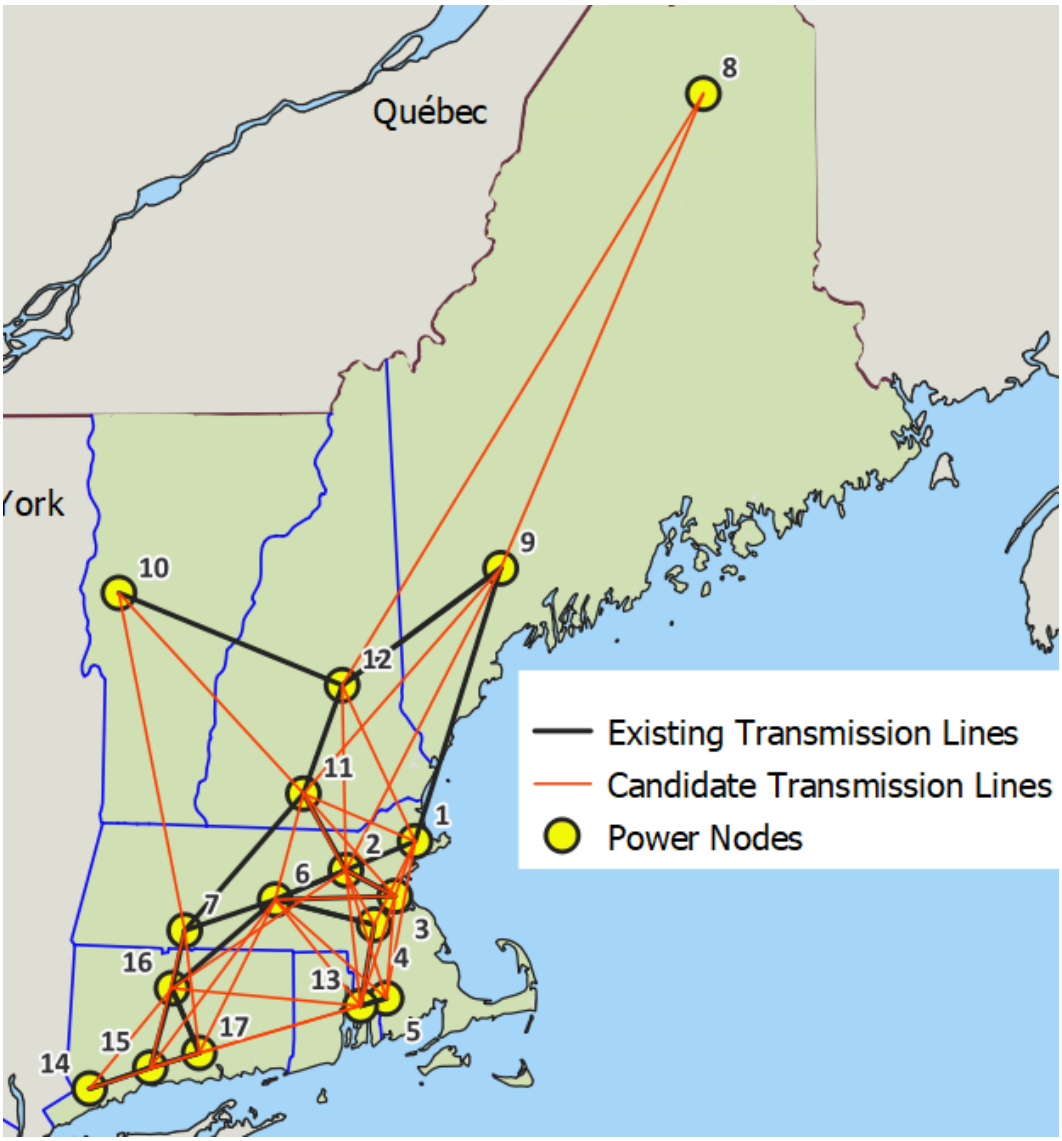

Figure 5: Existing and candidate transmission lines in the power network

assigned two buses that already have an existing transmission line between them. Finally, the length of each existing line is updated according to the distance of the power nodes it connects. This process resulted in 30 existing transmission lines in the power network.

Once the existing transmission lines are identified, we create candidate lines. We assume that each node can be connected to the 2 nearest non-adjacent nodes via candidate transmission liens, thus creating 34 candidate lines. The susceptance and maximum flow limit of these lines is set to their average in the set of existing lines. Fig. 5 shows all the existing and candidate transmission lines in the current data set. It worth noting that multiple transmission lines connect some node pairs, but only one is depicted in the figure.

4.3 Power Plants

The "base_grid" data contains power plant information at each bus. We first remove power plant types of "dfo" (distillate fuel oil) and "coal" as their share is not substantial in the base year, and there are plans to phase out those plant types by 2050 completely. We then calculate the existing generation capacity at each bus from each plant type by considering "in-service" plants with generation capacity greater than 10MW for "ng" (i.e., gas-fired plants)

and "nuclear" type and greater than 2MW for VREs. Note that "ng" plants are treated as a lump since the Breakthrough Energy data set does not provide a breakdown between the capacity of combined-cycle and open-cycle plants. We then assign each plant to the nearest node and eventually aggregate the generation capacity of each plant type for each of the 17 nodes. Table 2 and 3 present the technical assumptions for the existing and new power plants used in this study. Most of the parameters for the existing plants are derived from the Breakthrough Energy data set [5]. The footnote text in Table 2 presents details on the value of each parameter. The technical assumptions for new plants is largely derived from National Renewable Energy Laboratories (NREL) Annual Technology Baseline 2021 edition (ATB 2021) for the year 2045 [25]. Values of parameters not available in ATB 2021 are adopted from the corresponding existing power plants or obtained from Sepulveda et al. [35]. The details of each parameter is provided in the footnote text in Table 3.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
ng	1327	80	3549	1618	284	1780	902	0	1755		1024	0	1454	2052	119	194	796
solar	25	52	9	34	114	137	107	0	0	57	6	0	10	0	0	0	10
wind	2	0	2	0	8	3	56	555	343	115	26	159	51	0	0	5	0
hydro	0	38	0	0	0	4	1780	169	439	92	75	404	0	68	39	17	0
nuclear	1226	0	0	0	617	0	0	0	0	0	0	0	0	0	0	0	1889

Table 1: Aggregate nameplate capacity of existing plants at each node (MW)

Table 2: Parameter Values for Existing Plants

Symbol	Parameter\Type	ng	solar	wind	hydro	nuclear
	FOM [\$/kW/year][1]	21	23	43	78	145
C_i^{var}	VOM [\$/MWh][2]	5	0	0	0	2
η_i	CO_2 Capture [%][3]	0	-	-	-	-
h_i	Heat Rate [MMBtu/MWh][4]	8.7	0	0	0	10.6
$\tilde{C}_i^{\text{dec}}$	Decom. cost per plant [\$][5]	5.0e6	4.5e4	1e6	-	3.0e8
U_i^{prod}	Nameplate capacity [MW][6]	137	4.7	47	15	933
L_i^{prod}	Min stable output [%][7]	31	0	0	0	42
U_i^{ramp}	Hourly Ramp rate [%][8]	96	-	-	-	25
C_i^{fix}	FOM per plant [\$/MW/year][9]	3.6e6	1.45e5	1.8e6	1.8e6	1.4e8
C_i^{startUp}	Startup Cost [\$][11]	4.52e4	-	-	-	4.6e4

[1] and [2] from [25] in year 2019, [3] from [41], [4] approximated from linear and quadratic coefficient of heat rate curve provided in [5], [5] estimated from [30] except for nuclear which is obtained from [39]. Decommissioning of hydro plants is not considered, [6] and [7] approximated from [5], [8] estimated from 30-min ramp rate in [5], [9] $r1 * 1000 * r6$ where $r1$ is the FOM value in $r1$ and $r6$ is nameplate capacity provide in row6, [11] from [35]. Startup cost for "ng" type is assumed to be the same as "CCGT" type.

4.4 Power Load

For all the demand side electrification scenarios of interest, we use the ISO New England Variable Energy Resource data which provides historical time series for load, average solar/wind CFs, and some meteorological data per state level between 2000 and 2021 [12]. All residential power and NG load projections are based on a base year between 2000 and 2021.

To account for non-residential load, we consider the 2050 hourly load profiles provided as part of the NREL's Electrification Future Study (EFS) Load Profile dataset [26]. The repos-

Table 3: Parameter Values for New Plants

Symbol	Parameter\Type	OCGT	CCGT	CCGT-CCS	solar-UPV	windN	wind-offshore	nuclearN
	CAPEX [$/kW][1]	780	935	2167	672	808	2043	6152
	FOM [$/kW/year][2]	21	27	65	15	35	74	145
C_i^{var}	VOM [$/MWh][3]	5	2	6	0	0	0	2
η_i	CO_2 Capture [%][4]	0	0	90	-	-	-	-
h_i	Heat Rate [MMBtu/MWh][5]	9.72	6.36	7.16	0	0	0	10.46
	Lifetime [year]	30	30	30	30	30	30	30
U_i^{prod}	Nameplate capacity [MW][6]	237	573	400	10	10	10	360
$\tilde{C}_i^{\text{inv}}$	CAPEX per Plant [$][7]	1.85e8	5.36e8	8.67e8	6.72e6	8.01e6	2.04e7	2.21e9
L_i^{prod}	Minimum stable output [%][8]	25	33	50	0	0	0	50
U_i^{ramp}	Hourly ramping rate [%][9]	100	100	100	-	-	-	25
C_i^{fix}	FOM per plant [$/yr][10]	5.0e6	1.55e7	2.6e7	1.5e5	3.5e5	5.22e7	5.22e7
C_i^{startUp}	Startup Cost [$][11]	8.0e3	4.52e4	3.79e4	-	-	-	4.6e4

[1−5] from [25] in year 2045. For CCGT-CCS, "Conservative" technology class is considered. For "windN" and wind-offshore "Moderate-Class4" technology class is considered. For all others "Moderate" cost assumption is considered, [6] from [35] for "OCGT", "CCGT", CCGT-CCS, and "nuclearN". For VRE, a modular capacity of 10MW is considered, [7] $r1 * 1000 * r6$ where $r1$ is the CAPEX value in row[1] and $r6$ is nameplate capacity provide in row[6], [8] from [35], [9] from [35], [10] $r2 * 1000 * r6$ where $r2$ is the FOM value in $r1$ and $r6$ is nameplate value provide in $r7$, [11] from [35]

itory contains hourly load profiles for various electrification (Reference, Medium, High) and technology advancement (Slow, Moderate, Rapid) scenarios. The load profiles are provided for several years, including 2050, and are further disaggregated by state, sector (residential, commercial, etc.), and subsectors (space heating and cooling, water heating, etc.). For all scenarios, we consider the aggregated state-level hourly demand profile for the *high* electrification level with *Moderate* technology advancement.

4.5 Regional Cost Multipliers

We incorporate regional capital cost multipliers provided in *ReEDS Model Documentation* [6], summarized in Table 4, to distinguish between the capital cost of new power plants in the different model regions. These multipliers are applied to the baseline capital costs reported in Table 3. Subsequently, the capital costs are annualized to be included in the single-stage investment planning model using the following formula for the annual cost fraction: $\frac{\omega}{1-(\frac{1}{1+\omega})^{lt}}$. Here, lt is the lifetime of the specific technology and ω corresponds to the discount rate of 7.1%. Thus, the annualized CAPEX for new power plants is obtained by multiplying the CAPEX by the annual cost fraction and regional multiplier. For every other investment cost (i.e., transmission lines, pipelines, storage, etc.), no regional cost multiplier is considered and we only multiply the CAPEX by the annual fraction factor to get the annualized CAPEX.

Table 4: Regional CAPEX multipliers for new plant types

State/Technology	OCGT	CCGT	CCGT-CCS	solar-UPV	wind	wind-offshore	nuclear
Connecticut (CT)	1.25	1.3	1.3	1.15	1.4	1.1	1.1
Massachusetts (MA)	1.1	1.1	1.1	1.05	1.35	1.1	1.05
Maine (ME)	1.25	1.3	1.3	1.1	1.35	1.1	1.1
New Hampshire (NH)	1.1	1.1	1.1	1.05	1.35	1.1	1.05
Rhode Island (RI)	1.2	1.25	1.25	1.1	1.35	1.1	1.05
Vermont (VT)	1.1	1.1	1.1	1.05	1.35	1.1	1.05

4.6 Plant Decommissioning Cost

The decommissioning costs for power plants and pipelines are not annualized. Instead, we assume a gradual retirement process starting from the mid-year date of 2040 until the planning year 2050. Since the model only considers a time horizon of a single year, we divide the decommissioning costs for these assets by 10. The distribution of the decommissioning cost over multiple years allows the model to decommission the asset if it is not utilized and its fixed cost is higher than the distributed decommissioning cost. Otherwise, the model keeps the idle asset without decommissioning it.

4.7 Capacity Factors for VREs

We obtain the capacity factors for solar from ISO New England Variable Energy Resource data [12]. We use the data for the base year used for the load projection in 2050 and assume all nodes located in the same state have the same solar and onshore wind capacity factors. We use the data for year 2013 check of [35] for onshore wind. There is no substantial offshore wind generation in the region, hence no historical estimates of offshore wind CF. However, the GenX data set provides a single capacity factor for the entire northeast region [18]. We use this CF for all the nodes that are allowed to establish offshore winds (i.e., coastal nodes of Massachusetts and Connecticut).

4.8 Power Storage

Energy Storage is likely to be an essential part of the future power systems dominated by VRE supply. While many storage technologies are proposed or under development, we model two major technologies: the potential deployment of Li-ion batteries and 'metal-air batteries.' The former technology provides short-term storage, while the latter is a long-duration storage technology (> 12 hours). These technologies are selected due to their commercial availability and available cost projections. However, their cost projection, especially for 'metal-air' is highly uncertain. Therefore we consider two cost projection levels for long-duration battery storage. We sourced Li-ion battery cost and performance assumptions from [25] and for the 'metal-air' from [17] as summarized in Table 5.

Table 5: power storage parameters

Symbol	Parameter	Li-ion	Metal-air (low)[1]	Metal-air (high)[2]
$\tilde{C}^{\text{EnInv}}$	Energy capital cost [$/kW]	129[3]	0.1	3.6
$\tilde{C}^{\text{pInv}}$	Energy power cost [$ /kWh]	156[4]	595	950
γ_r^{eCh}	Charge efficiency	0.92[5]	0.7	0.72
γ_r^{eDis}	Discharge efficiency	0.92[6]	0.59	0.6
C_r^{EnInv}	Energy related FOM ($/kWh/year)	3.22[7]	0	100
C_r^{pInv}	Power related FOM ($/kW/year)	3.9[8]	14.9	23.7
γ_r^{selfD}	hourly self-discharge rate	2.08e-5[9]	2.08e-5	2.08e-5
	Lifetime	15[10]	25	25

columns [1] and [2] from [17],
[3] and [4] from [25] in year 2045 averaged over "Advance", "Moderate" and "Conservative" estimates, [5] and [6] from [25] where the round-trip efficiency is provided at 85%, [7] and [8] 2.5% of energy capital and power cost (row 1 and 2), respectively [8], [9] from [17] the monthly self-discharge provided at 1.5%, [8] from [25]

4.9 Reserve Margin

The capacity derating factor $\gamma_{nit}^{\text{CRM}}$ for VRE plants is set to their capacity factors. For other plants, the value is set to 1. The capacity reserve margin rate R^{CRM} is set at 15%, which is the value recommended by North American Electric Reliability Corporation (NERC) [31].

4.10 Transmission Line FOM

The fixed cost for transmission lines is estimated from [36], in which the operating and maintenance cost for a transmission line of 134 kV is given as about 20% of the investment cost. Therefore, we estimate the FOM cost as 20% of the CAPEX over its lifetime. We currently consider CAPEX of a line at 3500 \$/MW/mile over 30 years so that the annual FOM will be $C_{\ell}^{\text{trFix}} \frac{3500 \times 0.2}{30} \approx 24$\$/MW/mile for all $\ell \in \mathcal{L}^{\text{e}}$. Investment and FOM costs of a line can be calculated by multiplying these values by the length and maximum capacity of a line.

4.11 CCS parameters

We based our estimation of CCS parameters on [38, 3, 4]. We assume that the collected carbon is stored in Appalachian Basin at a Marcellus region located in the middle of Pennsylvania [3]. We then calculate the distance between each node and the storage site. The total capacity of the Appalachian Basin is 1278 Mt (megaton). Assuming the Basin is operable for 100 years, the total annual carbon storage capacity U^{CCS} becomes 12.78 Mt. Other parameters are calculated as follows:

- $C_{\text{CO2}}^{\text{inv}}$: Reference [4] provides CAPEX for 10 and 100 miles pipelines. We consider 100-mile pipelines as all distance values are greater than 200 miles. The CAPEX and FOM for 100 miles are 225 \$M and 1.3 \$M (million dollars), respectively. With 30 years of a lifetime for CO_2 pipelines and WACC=7.1%, the CAPEX becomes 18.31 \$M. The FOM is given at 1.3 \$M, so the per mile investment and FOM is (18.31e6+1.3e6)/100 = 196e3 \$/mile. The reference assumes that the capacity of the pipeline is 10 Mt/y (megaton per year). Therefore, the levelized investment and FOM becomes 196e3/10e6 = 0.0196 \$/mile/ton.

- $C_{\text{CO2}}^{\text{str}}$: The Appalachian basin is an aquifer type storage. The CAPEX is given at 4.3 \$M in [4] for a storage site of type aquifer with 7.3 Mt capacity per year. The CAPEX consists of injection site screening and evaluation, injection equipment, and drilling of 6 wells. Assuming 30 years of lifetime, the annualized CAPEX becomes 350e3 \$. The FOM is given at 600e3 \$, so the levelized investment and FOM is (350e3+600e3)/7.3e6 = 0.13 \$/ton.

- E^{pipe}: For a pipeline of 100 miles long with a capacity 10 Mt/year, reference [4] gives the electric requirement at 32,000 MWh/year or equivalently 32/8760e6 = 0.00365e-6 [MWh/mile/ton/hour].

- E^{pipe}: For a pipeline with a capacity of 10 Mt/year, each pump consumes 4190 MWh electricity annually which accounts for 4190/(8760e6)=0.478e-6 MWh/ton/hour.

- E^{cprs}: Compression pump are located every 3.3 miles along the pipeline [3]. Therefore, the value is $d_n/3.3$.

We assume that CAPEX for compression pumps is negligible. Also note that these cost estimations for CCS storage are conservative as we levelized the investment and FOM costs and do not consider other cost parameters such as labor, FOM of compression pumps and fugitive emission amount which are listed in [3].

4.12 Other Parameters

Other economic and technical assumptions for the power network are presented in Table 6 and Table 7.

Table 6: Other parameters for power network

Symbol	Parameter	Value
	Transmission line lifetime [year]	30
	Weighted average cost of capital (WACC)[1]	7.1%
$\tilde{C}_{\ell}^{\text{trans}}$	Transmission line investment cost [\$/MW/mile][2]	3500
	Li-ion Battery lifetime [year][3]	30
	Transmission line lifetime [year][4]	30
	Uranium price [\$/MMBtu][5]	0.72

[1] from [35], [2] from [6], [3] from [25], [4] from [25] in year 2045, [5] from [35] in year 2045

Table 7: Resource Availability Data

$\mathcal{Q}$	Maximum Available Value
"solar", solar-UPV	22 GW
"wind", "windN"	10 GW
wind-offshore	280 GW
"nuclear", "nuclearN"	3.5 GW

Resource availability amounts are obtained from [16]

5 NG System Data

NG network consists of two types of nodes: 1) NG nodes that have injection capacity and load. We further distinguish between nodes whose load is zero but can be used to inject NG in the system and those that have NG load. The nodes without load are created in the boundary of the region to represent the connection of NG nodes to outside region. We call the former set of nodes "boundary nodes" and the latter "load nodes". Note that this distinction is for the exposition of the data input and has no modeling implications ; 2) SVL nodes: each SVL node consists of Storage, Vaporization, and Liquefaction (SVL) facilities, each with its own capacity. This section provides details of each node type and other parameters associated with the NG system.

5.1 NG Nodes

We construct the NG network based on the data available on Energy Information Administration (EIA) website [10]. The pipeline data provide information on the interstate and major pipelines, and their start and ending counties. Each pipeline is a uni-directional means of transferring natural gas with a daily capacity limit. We first filter all the pipelines whose ending counties are one of the New England counties. We consider each of these counties as a load node. Some load nodes are connected to regions outside New England via pipelines. To capture the import of NG to the region via pipelines, we create boundary nodes in locations where a pipeline connects a load node to an outside region. This process resulted in 6 boundary nodes and 17 load nodes. The import of fuel to the region is realized either

through pipelines from neighboring regions or coastal import facilities from overseas [10, 1]. We consider the boundary nodes as ground injection. The injection to the system can also be realized by LNG facilities located in Everett and Cape Ann, both are in Massachusetts [1]. These are LNG facilities that are equipped with vaporization plants to inject fuel to the network.

by two of the load nodes located in Massachusetts.

NG is usually stored in its liquefied form called LNG. LNG storage usually involves three types of facilities:

- Storage tank: smaller tanks located inland that are filled by truck supply from the import facilities and are used to regulate the pressure of the gas system – these facilities have tanks and vaporization facilities [22]. The capacity of a storage facility is measured in energy/volume level (i.e., MMBtu or MMScf)

- Liquefaction: a facility that receives NG from a pipeline system and liquefies it at a temperature of around -162 Celsius [22, 24]. The capacity of the liquefaction facility is measured in terms of flow rate (i.e., MMBtu/period or MMScf/period)

- Vaporization: The facility evaporates the LNG leaving the storage facility by warming it with seawater or air to produce gas that is injected into the pipeline network. In certain areas, trucks are used to transport LNG from a major storage site to a smaller storage site where they are later evaporated [22]. For ease of transportation, storage and vaporization facilities are built at the same location. The capacity of the vaporization facility is measured in terms of flow rate (i.e., MMBtu/day or MMScf/day).

We construct the SVL network based on the data provided in [1]. We assume that all three facilities are located at the same location. There are currently 5 liquefaction facilities and 43 storage in New England. The exact location of these facilities is not provided, but the source provides a map of the region showing the approximate locations of storage tanks. Given that there are only 5 liquefaction facilities, we first cluster storage tanks into 5 locations and assume a liquefaction and a vaporization facility at each location. This assumption effectively approximates the practice of moving LNG via trucks from the centralized liquefaction facilities to the distributed storage (and vaporization) facilities. Each of these locations is a node in the SVL network. The total liquefaction, vaporization and storage capacities for the New England region is given in [1]. To account for variation in the capacities, we unevenly distribute these capacities over 5 SVL. This process resulted in 6 boundary nodes, 17 load nodes, and 5 SVL nodes as depicted in Fig. 6. Note that node 9 also

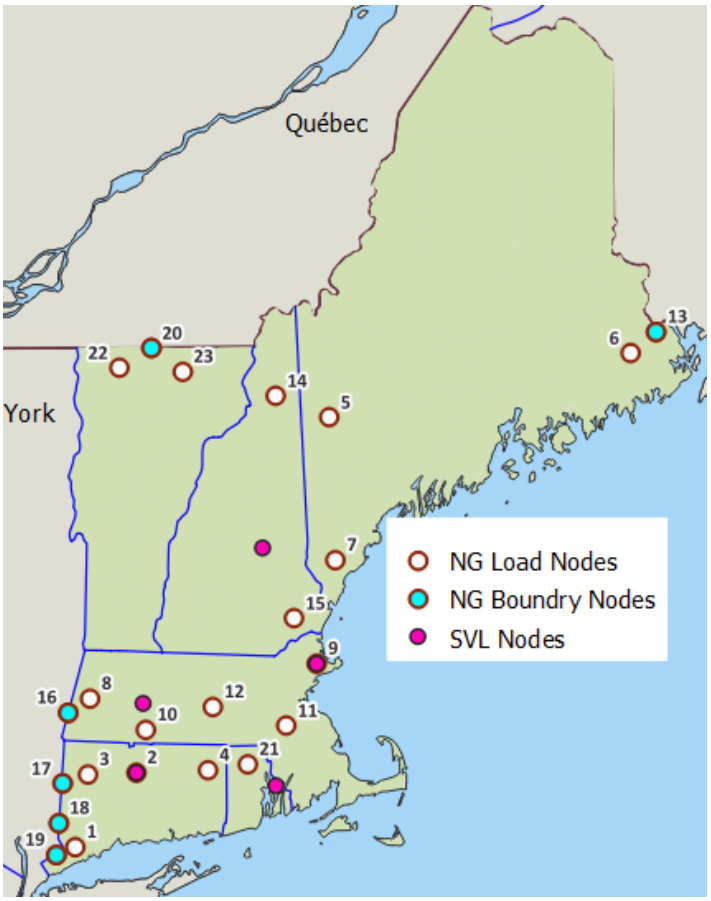

Figure 6: Location of each type of node in the NG system. Note that two SVL and load nodes share the same location. SVL nodes are not numbered here for clarity of the figure.

5.2 Pipelines

We use the pipeline data from [10] to identify the existing pipelines between load nodes and between boundary nodes and load nodes. Each existing pipeline has a known daily capacity limit. We construct candidate lines for boundary nodes by creating a pipeline between them to three of their nearest load nodes.

The transfer of NG between different nodes is assumed to be realized as follows:

- boundary and load nodes to load nodes: The existing interstate pipelines between boundary and load nodes are provided in [10]. We estimate the existing intrastate pipelines based on the map provided in 7 with capacities associated with the interstate pipeline ending in the node from which the intrastate pipeline originates. For candidate pipelines, we assume that each node is connected to 2 nearest non-adjacent, non-boundary load nodes, resulting in 46 candidate pipelines. The capacity of candidate pipelines is set to the average capacity of the existing ones.

- NG load node to power node: The location of NG-fired power plants as well as NG pipelines is provided in [13]. Fig. 7 shows that each power plant is connected to one major pipeline. We use this fact to assume that each power node is already connected to its nearest NG node through distribution pipelines. The pipeline capacity available to these neighboring NG nodes thus limits the amount of flow between NG and power nodes, hence we do not consider a limit to pipelines connection NG and power nodes.

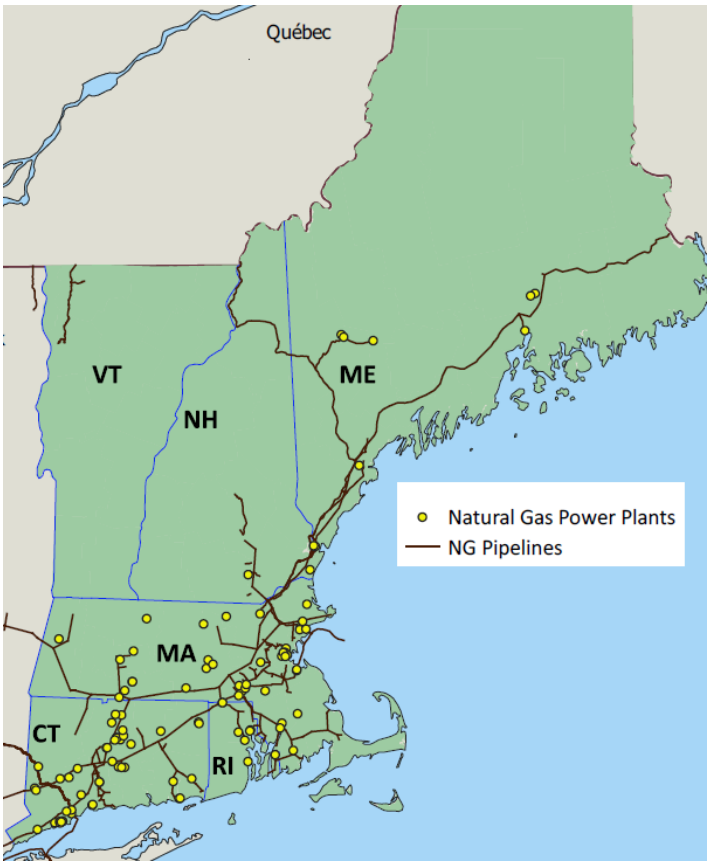

Figure 7: Natural gas pipelines and NG-fired power plants in New England

- NG load nodes and SVL nodes: We assume that each NG load node is already connected to its nearest SVL node through two sets of pipelines. The first set carries NG from the NG node to the liquefaction facility in an SVL, and the second set of pipelines transports NG from the vaporization facility in an SVL to an NG node.

Fig. 8 illustrates the existing pipelines of all node types of the NG system. Currently, in the year 2022, there are 25 pipelines that connect boundary and load nodes. The procedure that we considered to create candidate lines resulted in 69 potential connections that are depicted in Fig. 9. The power nodes draw NG from their nearest load nodes as illustrated in Fig. 10. Finally, Fig. 11 shows all the connections we consider in JPoNG.

5.3 Pipelines FOM Costs

FOM cost of NG pipelines is estimated from [14] in which it is stated that 5-10% of a pipeline's cost across its lifetime is operating cost. Since we assume 30 years of a lifetime for new pipelines, we divide the pipeline CAPEX by 30, hence $C_{\ell}^{\text{pipeFix}} = \frac{20e6 \times 0.1}{30} = 6.6e4\$/\text{mile}$ for all $\ell \in \mathcal{L}^{\text{g}}$.

5.4 Pipelines Decommissioning Cost

The decommissioning cost for the existing pipelines [19]. The reference provides a cost estimate for decommissioning of pipelines in the US Gulf of Mexico between 1995 to 2015. We use the value of 3e5 \$/mile value which is provided in Table 3 of the reference as the average cost of decommissioning. Similar to power plants, we assume a gradual retirement and distribute the value over 10 years, hence $C_{\ell}^{\text{pipeDec}} = 3e4\$/\text{mile}$ for all $\ell \in \mathcal{L}^{\text{g}}$.

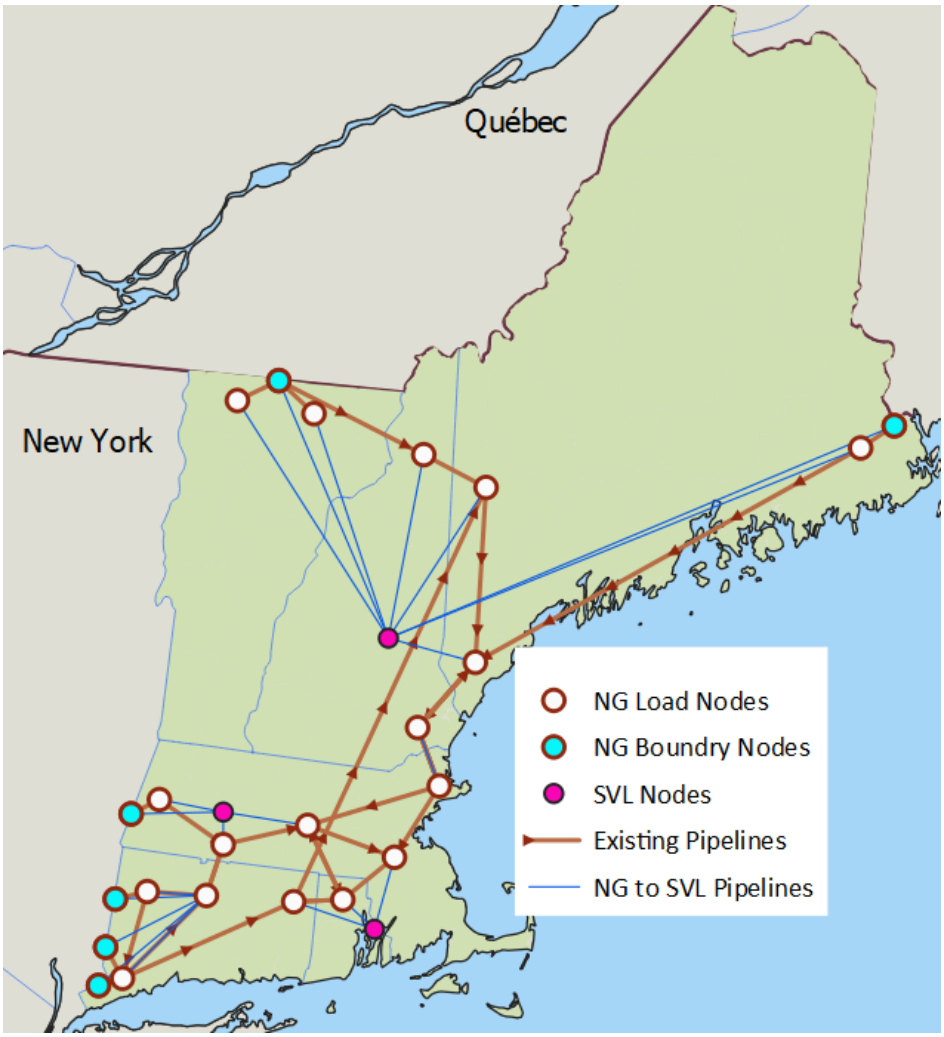

Figure 8: Existing pipelines between three types of NG nodes. The connection between load and SVL nodes are shown by a single line for clarity of the figure.

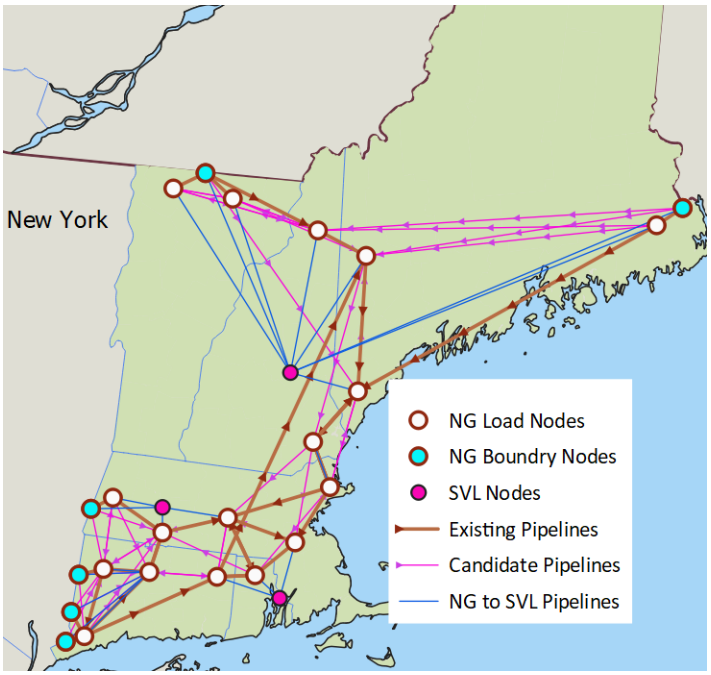

Figure 9: Existing and candidate pipelines between three types of NG nodes

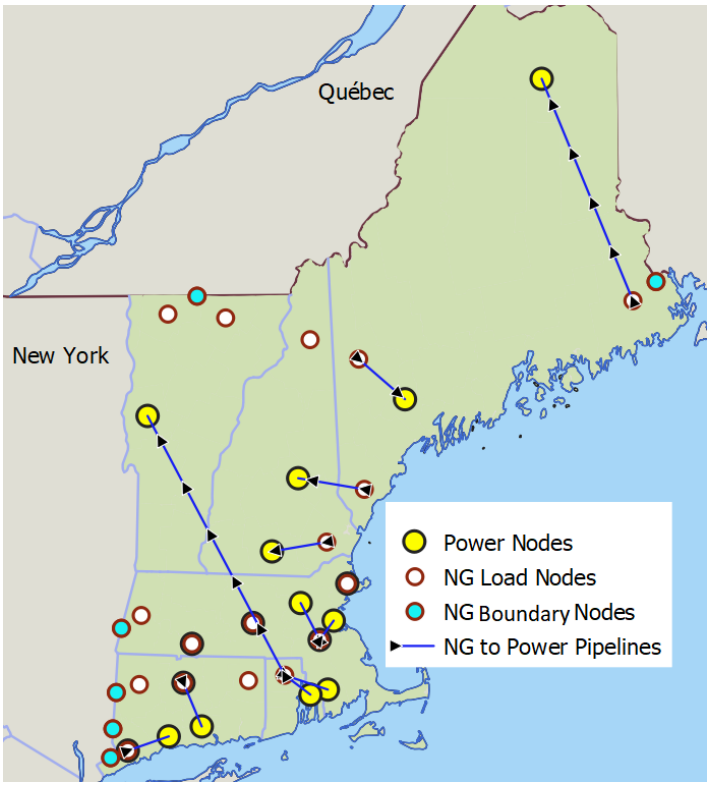

Figure 10: Pipelines between power nodes and NG load nodes

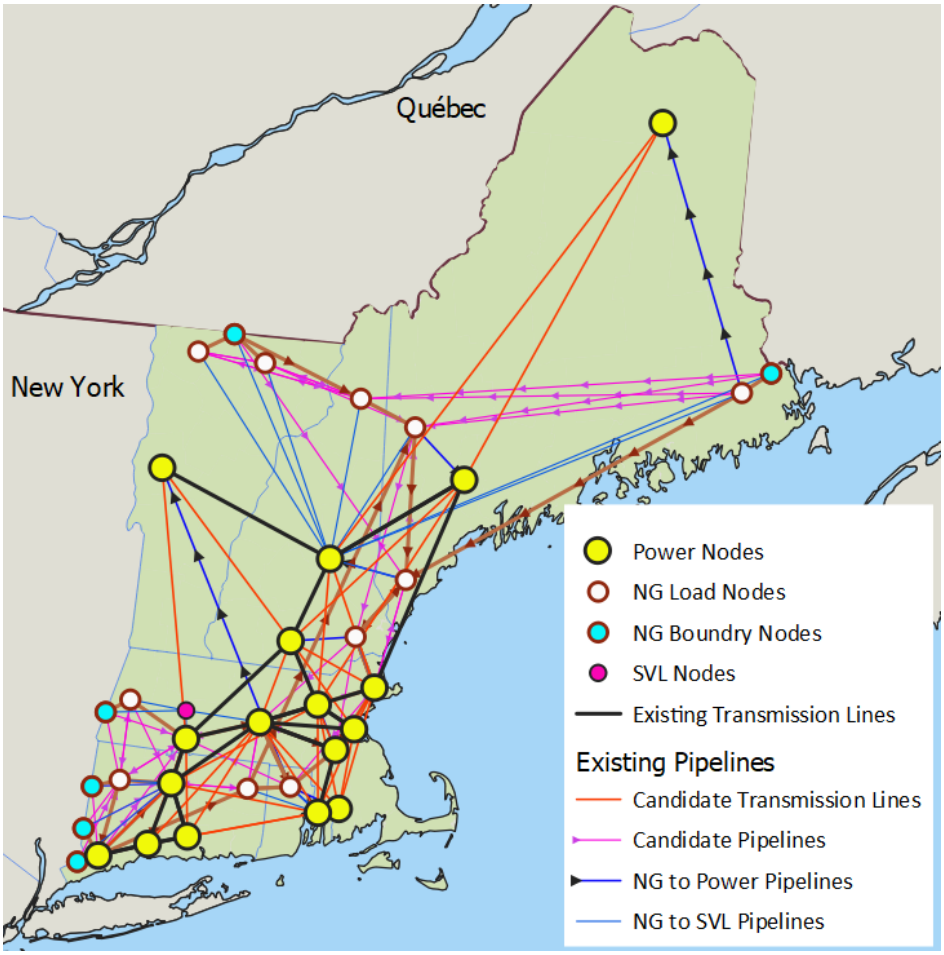

Figure 11: All existing and candidate connections in New England case study

Table 8: Other parameters for NG/SVL network

Parameter	Value	
$\tilde{C}_j^{\text{strInv}}$	Storage tank CAPEX[1] [\$/MMBtu]	729.1
$\tilde{C}_j^{\text{vprInv}}$	Vaporization CAPEX[2] [\$/MMBtu]	1818.31
C_j^{strFix}	Storage tank FOM[3] [\$/MMBtu]	3.6
C_j^{vprFix}	Vaporization FOM[4] [\$/MMBtu/d]	327.3
γ_j^{ligCh}	Liquefaction charge efficiency (%)	100
γ_j^{vprDis}	Vaporization discharge efficiency[5] (%)	98.9
	Pipeline lifetime [year]	30
C_j^{gShed}	NG load shedding cost [\$/MMBtu]	1000
η^{g}	Emission factor for NG [6] [ton/MMBtu]	0.053
C^{alt}	LCDF price [\$/MMBtu][7]	20
	SVL lifetime [year]	30
$\tilde{C}_\ell^{\text{pipe}}$	Pipeline investment cost [\$/mile][8]	20e6
C^{ng}	NG price [\$/MMBtu][9]	5.45

[1] from Table 1 of [11], [2] from Table 1 of [11], [3] 0.5% of CAPEX according to [24], [4] 1.8% of CAPEX according to [24], [5] estimated from [22], [6] from [10], [7] from [9], [8] approximated from [10] for NG pipeline projects. We consider the completed new pipeline projects in New England between 2011 to 2020 and divide the total cost to the length of the pipeline

5.5 Other Parameters

Other economical and technical assumption for the NG system is presented in Table 8.

5.6 NG Load

Similar to the power system, the NG load for the residential sector is projected based on one of the historical years. We estimate the NG load displacement due to space heating electrification by the EFS data [26], NREL's "End-Use Load Profiles for the U.S. Building Stock" project [27], and Energy Information Administration (EIA) [10]. The load for NG is generally disaggregated into five sectors, including residential, commercial, industrial, vehicle fuel consumption, and electric power customers [10]. The consumption for electric power consumers is a decision variable in our model, so our input for NG demand involves the NG demand in the remaining four sectors. The monthly state-level NG load for all five sectors is available on the EIA website [10]. The EIA website does not provide any information on the distribution of the monthly demand over its days (i.e., the load shapes). Therefore, we consider industrial, commercial, and vehicle fuel consumption in 2017 and uniformly distributed the monthly load across their corresponding days. We then scale the values based on the annual industrial and vehicle fuel consumption in 2050 under reference and high electrification scenarios. Finally, we aggregate loads for all sectors and subsectors to obtain the NG consumption of each state in 2050 under the two electrification scenarios. Once the daily NG demand is obtained for each state, we disaggregate the demand over each node in the proportion of the population each node represents in 2019 [40] which implicitly assumes that no change in the proportion of population by 2050.

5.7 Emission Amounts

All New England states have set a goal to reduce the emission of GHG by at least 80% until 2050 below the baseline years which is 1990 for all states except Connecticut [44]. The total

CO_2 emission for the New England states was 171.2 metric tons (mt) in 1990, of which 43.9mt was electricity energy-related emission and 23.6mt was NG energy-related emission [42]. The remaining emission was caused by consuming coal and petroleum that we do not consider in this study. Based on these figures, $U^{\mathrm{g}}_{\mathrm{emis}}$ = 23.6e6 and $U^{\mathrm{e}}_{\mathrm{emis}}$ =43.9e6.

Table 9: Actual emission budget (ton)

	$\zeta = 80\%$	$\zeta = 85\%$	$\zeta = 90\%$	$\zeta = 95\%$
C1	3.51e7	3.73e7	3.95e7	4.17e7
C2 and C3	5.4e7	5.74e7	6.08e7	6.41e7

6 Python Code Guide

JPoNGis coded in Python and with Gurobi solver. The main directory contains 3 folders and several *comma-separated values (CSV)* files and Python scripts. The `CSV` files end with *.csv* and Python scripts have *.py* suffixes. This section gives a brief guide on the code architecture. We start by explaining data files and proceed with Python scripts. For ease of exposition, we assume that load and CFs are based on the 2011 base year. However, this year can be replaced by any other year to reflect another base year.

6.1 Input Data

All required input data are stored in several CSV files in different directories. The main directory (i.e., ...\`JPoNG` folder) contains the file `Other_Parameters.csv` which sets the values for scalar and array-like parameters such as `WACC`, `NG_price`, and `LCDF_price`.

Power_System_Data folder: This folder contains tabular parameters used in the power system:

- `AvailabilityFactor_Solar_2011.csv`: CF for both existing and new solar plants based 2011 data of [12]. The name changes according to the base year.
- `AvailabilityFactor_Wind_Onshore.csv`: CF for both existing and new onshore wind plants.
- `AvailabilityFactor_Wind_Offshore.csv`: CF for new offshore wind plants.
- `CCC_params.csv`: Cost and technological parameters for the carbon pipeline infrastructure. The data also contains the location of the Appalachian basin, the underground storage to bury the captured CO_2
- `Electricity_Load_HE_BaseYear2011.csv`: Nodal load for power system under "HE" electrification scenario and base year of 2011. The files name changes to reflect the scenario and based year.
- `Plant_params.csv`: Technological and cost parameters for all existing and new power plant types
- `Plant_Nodes.csv`: Type, number, and a nameplate capacity of the existing power plants at each power node. The last and three first columns are relevant in this model.
- `Power_Nodes.csv`: ID, location, and other relevant information for each power node. The file also specifies whether an establishment of offshore wind plants is allowed in the node.

- `Regional_Multipliers.csv`: Regional multipliers for new technologies per state. We assume that nodes of the same state share the same multipliers.
- `Storage_Params.csv`: Cost and technological parameters for Li-ion and metal-air batteries.
- `Transmission_Lines.csv`: List of existing and candidate transmission lines along with their susceptance, maximum flow limit, and length.

NG_System_Data folder: This folder contains tabular parameters used in the power system:

- `NG_AdjE_Nodes.csv`: Starting from the second row, each row corresponds to NG nodes (either boundary or load nodes). The values are the ID of the power nodes to which the NG node is connected.
- `NG_Load_HE_BaseYear2011.csv`: Nodal load for NG system under "HE" electrification scenario and base year of 2011. The file's name changes to reflect the scenario and based year. Note that the load for boundary nodes is zeros throughout the year.
- `NG_Nodes.csv`: ID, location, and other relevant information for each NG boundary and load node. The file also specifies which SVL facility each node is connected.
- `NG2NG_Pipelines.csv`: List of existing and candidate pipelines between NG nodes along with their maximum flow limit and length.
- `SVL_data.csv`: ID, location, and capacities of SVL facilities.
- `SVL_params.csv`: Cost and technological parameters for SVL facilities.

joint_CF_with_extreme_days folder: This folder contains the clustering of days based on the number of representative days. For example, the file `joint_CF_with_extreme_days_k=10.csv` assigns each of 365 days to one of the 10 clusters.

6.2 Python Scripts

The Python code is developed with efficiency, agility, readability, and transferability in mind. As such, it both contains elements of object-oriented programming (OOP) and functional programming paradigms. Different aspects of the code are compartmentalized for ease of understanding the coding flow. The following scripts can be found in the main folders directory:

- `Setting.py`: The file has no dependencies on other scripts and defines the "Setting" class to store the problem setting such as emission reduction goal ("emis_reduc_goal) and electrification scenario.
- `SystemClasses.py`: This scrip defines "EV" class for the electricity power sysytem and "GV" class for the NG system. Each class contains fields that are later populated by decision variables and values of those variables after the problem is solved. This script also has no script dependency.
  - `EV()` class: A class to store decision variables and their values for the power system. The class is treated as a static class with no instantiation.
  - `GV()` class: A class to store decision variables and their values for the NG system. The class is treated as a static class with no instantiation.
- `ProblemData.py`: The script recalls necessary packages as well as "Setting" class. It then reads the data from CSV files by defining and populating necessary classes:

- `time_weights` function: Read the representative days and their weights.
- `other_prm()` class: A class to store scalar and array-like parameters of both systems. The class is instantiated by `Other_Input` object.
- `enode()` class: A class to store parameters associated with each power node. The class is instantiated for each power node who are stored in `Enodes` list.
- `plant()` class: A class to store parameters associated with power plant types. The class is instantiated for each power plant type who are stored in `Enodes` list.
- `branch()` class: A class to store parameters associated with the existing and candidate transmission lines. The class is instantiated for each line who are stored in `Branches` list.
- `eStore()` class: A class to store parameters associated with storage technologies. The class is instantiated for each storage type who are stored in `eStore` list.
- `CCS()` class: A class to store parameters associated with carbon capture and store process. The class is instantiated once by `CC_CCS`.
- `gnode()` class: A class to store parameters associated with each NG boundary and load node. The class is instantiated for each power node who are stored in `Gnodes` list.
- `pipe()` class: A class to store parameters associated with the existing and candidate pipelines. The class is instantiated for each line who are stored in `PipeLines` list.
- `exis_SVL()` class: A class to store capacity and ID of the existing SVL facilities. The class is instantiated once by `Exist_SVL`.
- `SVL()` class: A class to store cost and technological parameters associated with the existing SVL facilities. The class is instantiated for each SVL node who are stored in `SVLs` list.

- `Modules.py`: The script recalls all the parameters, variables, and settings classes defined in previous scripts and implements the formulation of both systems along with the coupling constraints. It also contains functions for handling the output:
  - `Power_System_Model(Model)` function: The function takes `Model` object as an input and defines objective function and constraints for the power system. The `Model` object is a Gurobi modeling object.
  - `NG_System_Model(Model)` function: The function takes `Model` object as an input and defines objective function and constraints for the NG system. The `Model` object is a Gurobi modeling object.
  - `Cou;ing_constraints(Model)` function: The function takes `Model` object as an input and defines the coupling constraints. The `Model` object is a Gurobi modeling object. The function also handles different coupling cases (case 1 to 4).
  - `Get_Vars_Vars(Model)` function: The function takes `Model` object as an input and stores the value of major decision variables. The `Model` object is a Gurobi modeling object. This function is only applicable after solving the problem.
  - `Publish_results(s_time,MIP_gap)` function: The function takes `s_time` and `MIP_gap` object as an input publishes the summary of output in `JPoNG_Results.csv` file. It can also publish the detailed variable values in a CSV file appropriately named based on the problem instance. For example, `17-30-HE-1-0.8-0.05.csv` is the detailed solution of an instance whose power system has 17 nodes, number of representative days is 30, scenario is "HE", emissions reduction goal is 80%, and VRE share is 50%. This function is only applicable after solving the problem.

- `Get_marginal_prices()` function: The function is applicable after the problem is solved for a set of representative days. It sets the values of investment variables to their values in the reduced problem, and solves the resulting linear programming problem for an entire year. It then takes calculates the dual for the NG balance equation to compute shadow prices of NG at each NG node. This value is further used to calculate marginal locational prices at each power node.

- `Main.py`: This script is the one that users set the setting and run from the command window. The script starts by recalling necessary packages and classes, sets the values of setting, and solves the problem using functions defined in previous scripts. The setting fields are mnemonically named. If run from the command window, the user should give the values of 14 parameters, each separated by a space.

- `run_parallel.py`: This script is not part of the main code but is used to facilitate running multiple instances in parallel on MIT's Supercloud [32] or any other machine.

- `SC2PC.py`: This script is not part of the main code but is used to facilitate copying the output files from the Supercloud to the local machine.

References

[1] Northeast Gas association. Statistical guide to the northeast u.s. natural gas industry 2021. website, Nov. 2021. Accessed: 2022-3-18.

[2] Fatemeh Barati, Hossein Seifi, Mohammad Sadegh Sepasian, Abolfazl Nateghi, Miadreza Shafie-khah, and João PS Catalão. Multi-period integrated framework of generation, transmission, and natural gas grid expansion planning for large-scale systems. *IEEE Transactions on Power Systems*, 30(5):2527–2537, 2014.

[3] Madalyn S Blondes, Sean T Brennan, Matthew D Merrill, Marc L Buursink, Peter D Warwick, Steven M Cahan, Troy A Cook, Margo D Corum, William H Craddock, Christina A DeVera, et al. *National assessment of geologic carbon dioxide storage resources: methodology implementation*. US Department of the Interior, US Geological Survey, 2013.

[4] Ocean Studies Board, Engineering National Academies of Sciences, Medicine, et al. Negative emissions technologies and reliable sequestration: A research agenda. 2019.

[5] Breakthrough Energy. U.s. test system with high spatial and temporal resolution for renewable integration studies. website, 2020. Accessed: 2022-2-8.

[6] Maxwell Brown, Wesley Cole, Kelly Eurek, Jonathon Becker, David Bielen, Ilya Chernyakhovskiy, Stuart Cohen, Allister Frazier, Pieter Gagnon, Nathaniel Gates, et al. Regional energy deployment system (reeds) model documentation: Version 2019. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2020.

[7] Modassar Chaudry, Nick Jenkins, Meysam Qadrdan, and Jianzhong Wu. Combined gas and electricity network expansion planning. *Applied Energy*, 113:1171–1187, 2014.

[8] Wesley Cole, A Will Frazier, and Chad Augustine. Cost projections for utility-scale battery storage: 2021 update. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2021.

[9] Wesley J Cole, Danny Greer, Paul Denholm, A Will Frazier, Scott Machen, Trieu Mai, Nina Vincent, and Samuel F Baldwin. Quantifying the challenge of reaching a 100% renewable energy power system for the united states. *Joule*, 5(7):1732–1748, 2021.

[10] U.S. Energy Information Administration (EIA). Eia website. Website, Mar. 22 2022. Accessed: 2022-2-18.

[11] Dominion Energy. Green river satellite lng facility basis of cost estimation. website, Mar. 2019. Accessed: 2022-3-18.

[12] ISO New England. 2022 iso new england variable energy resource (ver) data series (2000-2021). website, 2022. Accessed: 2023-1-8.

[13] ISO New England. U.s. energy information administration (eia). website, 2022. Accessed: 2023-1-8.

[14] Manfred Hafner and Giacomo Luciani. *The Palgrave Handbook of International Energy Economics*. Springer Nature, 2022.

[15] Maryam Hamedi, Reza Zanjirani Farahani, Mohammad Moattar Husseini, and Gholam Reza Esmaeilian. A distribution planning model for natural gas supply chain: A case study. *Energy Policy*, 37(3):799–812, 2009.

[16] E3 Inc. and EFI. Net-zero new england: Ensuring electric reliability in a low-carbon future. 2020.

[17] MIT Energy Initiative. The future of energy storage, 2022.

[18] Jesse D Jenkins and Nestor A Sepulveda. Enhanced decision support for a changing electricity landscape: the genx configurable electricity resource capacity expansion model. 2017.

[19] Mark J Kaiser. Ferc pipeline decommissioning cost in the us gulf of mexico, 1995–2015. *Marine Policy*, 82:167–180, 2017.

[20] Leander Kotzur, Peter Markewitz, Martin Robinius, and Detlef Stolten. Time series aggregation for energy system design: Modeling seasonal storage. *Applied Energy*, 213:123–135, 2018.

[21] Can Li, Antonio J Conejo, Peng Liu, Benjamin P Omell, John D Siirola, and Ignacio E Grossmann. Mixed-integer linear programming models and algorithms for generation and transmission expansion planning of power systems. *European Journal of Operational Research*, 297(3):1071–1082, 2022.

[22] Dharik S Mallapragada, Eric Reyes-Bastida, Frank Roberto, Erin M McElroy, Dejan Veskovic, and Ian J Laurenzi. Life cycle greenhouse gas emissions and freshwater consumption of liquefied marcellus shale gas used for international power generation. *Journal of Cleaner Production*, 205:672–680, 2018.

[23] P Duenas Martinez. Analysis of the operation and contract management in downstream natural gas markets. *Pontificial University Comillas (ICAI)(Institute for Research in Technology)*, 2013.

[24] J Mesko and J Ramsey. The use of liquefied natural gas for peaking service. 66. 1996.

[25] NREL. 2021 annual technology baseline (atb). website, 2021. Accessed: 2022-2-8.

[26] NREL. Electrification futures study load profiles. website, Aug. 2022. Accessed: 2022-8-21.

[27] U.S. Department of Energy Office of Scientific and Technical Information. End-use load profiles for the u.s. building stock. Website, June 22 2022. Accessed: 2022-6-18.

[28] Bryan S Palmintier and Mort D Webster. Impact of operational flexibility on electricity generation planning with renewable and carbon targets. *IEEE Transactions on Sustainable Energy*, 7(2):672–684, 2015.

[29] Jing Qiu, Hongming Yang, Zhao Yang Dong, Jun Hua Zhao, Ke Meng, Feng Ji Luo, and Kit Po Wong. A linear programming approach to expansion co-planning in gas and electricity markets. *IEEE Transactions on Power Systems*, 31(5):3594–3606, 2015.

[30] Daniel Raimi. Decommissioning us power plants. *Decisions, costs, and key issues. Resources for the Future (RFF) Report*, 2017.

[31] Andrew Reimers, Wesley Cole, and Bethany Frew. The impact of planning reserve margins in long-term planning models of the electricity sector. *Energy policy*, 125:1–8, 2019.

[32] Albert Reuther, Jeremy Kepner, Chansup Byun, Siddharth Samsi, William Arcand, David Bestor, Bill Bergeron, Vijay Gadepally, Michael Houle, Matthew Hubbell, et al. Interactive supercomputing on 40,000 cores for machine learning and data analysis. In *2018 IEEE High Performance extreme Computing Conference (HPEC)*, pages 1–6. IEEE, 2018.

[33] Isam Saedi, Sleiman Mhanna, and Pierluigi Mancarella. Integrated electricity and gas system modelling with hydrogen injections and gas composition tracking. *Applied Energy*, 303:117598, 2021.

[34] Carlos Saldarriaga-Cortés, Harold Salazar, Rodrigo Moreno, and Guillermo Jiménez-Estévez. Stochastic planning of electricity and gas networks: An asynchronous column generation approach. *Applied energy*, 233:1065–1077, 2019.

[35] Nestor A Sepulveda, Jesse D Jenkins, Aurora Edington, Dharik S Mallapragada, and Richard K Lester. The design space for long-duration energy storage in decarbonized power systems. *Nature Energy*, 6(5):506–516, 2021.

[36] EPRI Research Team. Optimization of transmission line design using life cycle costing. *Technical Update*, (1017712), 2009.

[37] Holger Teichgraeber and Adam R Brandt. Time-series aggregation for the optimization of energy systems: Goals, challenges, approaches, and opportunities. *Renewable and Sustainable Energy Reviews*, 157:111984, 2022.

[38] Gary Teletzke, Jeffrey Palmer, Eric Drueppel, Michael B Sullivan, Kenneth Hood, Ganeswara Dasari, and Gregory Shipman. Evaluation of practicable subsurface co2 storage capacity and potential co2 transportation networks, onshore north america. In *14th Greenhouse Gas Control Technologies Conference Melbourne*, pages 21–26, 2018.

[39] United States Nuclear Regulatory Commission. Backgrounder on decommissioning nuclear power plants. website, 2021. Accessed: 2022-2-8.

[40] US Census 2019. County population totals: 2010-2019. website, 2021. Accessed: 2021-12-16.

[41] U.S. Energy Information Administration (EIA). Emission coefficients. website, 2021. Accessed: 2022-2-8.

[42] U.S. Energy Information Administration (EIA). Energy-related co2 emission data tables. website, 2022. Accessed: 2022-2-8.

[43] Gregory Von Wald, Kaarthik Sundar, Evan Sherwin, Anatoly Zlotnik, and Adam Brandt. Optimal gas-electric energy system decarbonization planning. *Advances in Applied Energy*, page 100086, 2022.

[44] Jurgen Weiss and J. Michael Hagerty. Achieving 80% ghg reduction in new england by 2050. 2019.

[45] Yixing Xu, Nathan Myhrvold, Dhileep Sivam, Kaspar Mueller, Daniel J Olsen, Bainan Xia, Daniel Livengood, Victoria Hunt, Benjamin Rouillé d'Orfeuil, Daniel Muldrew, et al. Us test system with high spatial and temporal resolution for renewable integration studies. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2020.