Supplemental materials for “Tracking physical delivery of electricity from generators to loads with power flow tracing”

Gailin Pease, Greg Miller, Wenbo Shi

July 18, 2023

Contents

- Supplemental methods (1.x)
- Supplemental discussion (2.x)
- Supplemental figures S1 - S22
- Supplemental tables S1 - S3

1 Supplemental methods

1.1 Variable reference

Each of these variables is defined at every timestamp t. For clarity in the text, we do not use the subscript t except when summing over timesteps.

- N is the number of nodes or buses in the network. We assume every node has a load attached, so this is also the number of loads.
- N_G is the number of generators in the network, here equal to the number of quantities we trace through the network (N_T or the number of tracers in the derivation given in the supplement).
- $P \in \mathbb{R}^{N \times N_T}$ (MW) is bilateral delivery
- $P_D \in \mathbb{R}^{N \times N_T}$ (MW/MW) is downstream delivery density
- $\vec{g} \in \mathbb{R}^{N_T}$ (MW) is the generation vector
- $\vec{l} \in \mathbb{R}^{N}$ (MW) is the load vector

1.2 Terminology reference

- **Bilateral delivery** (MW): Power delivered from a single generator, or from all generators connected to a bus, to a single load in a single timestep.
- **Downstream delivery density** (MW/MW): Power delivered from a single generator G to a single load L in a single timestep, normalized by the load at L.
- **Upstream**: Conceptual term referring to viewing delivery from the perspective of a load looking ”upstream” to see which generators serve the load
- **Downstream**: Conceptual term referring to viewing delivery from the perspective of a generator looking ”downstream” to see which loads it serves
- **Bilateral**: Conceptual term referring to any metric on the level of a single generator/load relationship

- **Delivery, deliverability**: Terms used here and in previous work with a range of definitions, generally attempting to capture something about generator/load relationships. To distinguish when we are talking about flow-based delivery/deliverability, we use the terms "physical delivery" and "physical deliverability"
- **Physical delivery**: Conceptual term referring to all bilateral delivery relationships across a grid as calculated using power flow tracing. Can be summarized using delivery distance or regional connectivity metrics.
- **Physical deliverability**: Conceptual term referring to *whether* power can be delivered between a generator and load, as determined using power flow tracing.
- **Expected deliverability** (% or fraction): A cutoff-based metric to measure physical deliverability, defined in text.
- **Delivery distance** (distance units, here km): Metric to measure physical delivery, defined in text.
- **Regional connectivity**: Metric calculated using graph modularity to measure physical delivery in a boundary set, defined in text.
- **Boundary set**: Boundaries defining regions which together cover the grid without overlap.
- **Region**: One of the regions defined by a boundary set.

1.3 Tracing a vector of quantities

The equations here are extended from those in Kang et al. 2015.

We start with the tracer intensity $E_G \in \mathbb{R}^{N_G \times N_T}$ where N_T is the number of tracers in the system.

We then calculate

$$R_G = P_G \times E_G \tag{1}$$

where $P_G \in \mathbb{R}^{N \times N_G}$ is the injection matrix for N nodes and N_G generators. Each element $P_{Gjk} = p$ if generator k is connected to node i and power injection from generator k to node j is non-zero, else $P_{Gjk} = 0$.

We calculate

$$P_N = diag\Big(\zeta_{N+K} \times \begin{pmatrix} P'_B \\ P_G^T \end{pmatrix}\Big) \tag{2}$$

where $P'_B \in \mathbb{R}^{N \times N}$ describes the power flow between each pair of nodes in the network. For each element P'_{Bij}:

- $P'_{Bij} = p, P'_{Bji} = 0$ if there is active power flow from i to j
- $P'_{Bij} = P'_{Bji} = 0$ otherwise.

We can then calculate the downstream delivery density $P_D \in \mathbb{R}^{N \times N_T}$:

$$P_D = (P_N - P_B'^T)^{-1} \times R_G \tag{3}$$

Instead of a single vector P_D with length N as in Kang et al. 2015, we now have a matrix where each row j describes the intensity (in units of tracer / MW consumed at node) of tracer j.

1.3.1 Bilateral delivery from downstream density

Delivery from generator i to load j can be calculated from delivery density by scaling by the load at j:

$$P_{i,j} = P_{Di,j} \times l_j \tag{4}$$

1.3.2 Application to generator tracing

The algorithm described above can be applied to any quantity associated with electricity generation, including fuel type (eg solar, wind, NG, etc), generator type (eg combined cycle, CHP, etc), generator location, or generator emissions (including CO2, NOX, PM2.5, etc).

To trace delivery from every generator in the grid, we use the algorithm above and set the variables as follows: $N_T = N_G$; and E_G is an identity matrix, since each generator injects 100% of its net generation.

1.3.3 A note on power flow tracing theory

The central assumption of power flow tracing, used here and throughout the power flow tracing literature, is that power is uniformly mixed on each line and at each node Bialek and Kattuman 2004. While this assumption is intuitive, it is not actually provable in reality, since the power from a solar plant (for example) is identical to the power from a natural gas plant. There is therefore no way to directly measure how much of the power serving a load "came from" the solar plant; instead, the source of the power serving a load is a virtual attribute that must be derived from the power flow (which can be measured) and the perfect mixing assumption.

1.4 Models

1.4.1 PyPSA-Eur

In Europe, we use the PyPSA-Eur model Hörsch et al. 2018. The PyPSA-Eur model is widely used, with 165 citations and an active user community. It uses real transmission topology data sourced from ENTSO-E and generator location and capacity data from the aggregated across sources Gotzens et al. 2019.

However, because it is mostly used for capacity expansion modeling and research, there has been relatively little work to benchmark the model to a specific year, although the model does use 2020 generator capacities. In addition, using real transmission line location data scraped from maps has resulted in some regions with unrealistic transmission constraints, especially near large population centers like Paris. The PyPSA-Eur team recommends using a clustered version of the model, where nodes are aggregated and lines between them combined, to avoid these transmission constraints. Here, we use the model clustered to 1024 nodes, the most nodes recommended by the PyPSA-Eur team.

1.4.2 Breakthrough model of Eastern Interconnect

In the US, we use the Eastern Interconnect of the Breakthrough energy model Wu et al. 2021. The Breakthrough authors benchmarked their model to actual generation data, ensuring that the model topology and parameterization, including line capacities, marginal costs, and capacity by fuel type, were in line with actual data. The model is benchmarked to two existing years, 2016 and 2020, and has two hypothetical future cases for 2030.

In this work we use the 2020 grid configuration. Although already slightly out-of-date given the ongoing expansion of wind and solar capacity in the US, the 2020 model is current enough to capture delivery patterns that still exist on today's grid.

The Breakthrough model has two drawbacks for our application. First, unlike PyPSA-Eur, it is a synthetic model, and does not represent the actual topology of the transmission grid. Although it was designed such that that the overall behavior of the grid aligns with reality, including when broken down by region, the underlying data is not based on real data. For example, there is no relationship between the model generators and lines and real generators and lines.

A second and more minor drawback is that the Breakthrough network is developed and available in the Matpower framework. For consistency with PyPSA-Eur, we use the PyPSA modeling framework in this work. In translating the Breakthrough network from its original Matpower format to PyPSA, we make several assumptions when forced to translate between incompatible Matpower and PyPSA models:

1. Line types: Line types are required for PyPSA-Eur network simplification functions which we apply to the Breakthrough network, but are not included in the Matpower Breakthrough model.

We assume one line type per parameter population in the Breakthrough Matpower network, and use the median parameters of each population to parameterize each line type.

2. Marginal cost functions: The Matpower Breakthrough model uses quadratic generation cost functions, while PyPSA accepts only linear marginal costs (and optionally a separate start-up cost). We assume the marginal cost of each generator is the derivative of the quadratic generation cost function when the generator is operating at capacity.

1.4.3 Open source model limitations

Both models share some limitations common among most open source grid models. First, renewable generation and load profiles are not specific to our target year, 2020 (Breakthrough uses 2016 profiles, while PyPSA-Eur uses 2013 by default). Second, load is distributed uniformly across nodes by population, which ignores the effect of loads unrelated to population, for example data centers. Third, both models we use are transmission models. PyPSA-Eur is limited to lines at or above 220 V, while the Breakthrough includes lines as low as 60 V. However, neither network includes distribution networks. Practically, this makes these models easier to work with, since including distribution networks would increase model size and complexity by orders of magnitude. However, it means that do not consider distribution-level generation assets. Next, these models are limited to the generators and loads within their bounds, and do not consider electricity traded with other networks.

1.4.4 Model validation

The models we use include renewable profiles but no generation data for dispatchable generators. This is because the models are designed to support capacity expansion research, for which 2020 generation profiles would not be relevant. To get full generation data, we configure each model to estimate the 2020 grid and run a linear optimal power flow (L-OPF) in PyPSA to dispatch the model's generators in all 8760 hours of the year. OPF dispatches generators to meet model demand while respecting the system's constraints (including but not limited to transmission constraints, the operational constraints of each generator, security constraints) and while minimizing the total cost (where cost can, depending on the grid, include a carbon cost). This is standard practice when using these and similar models, and we perform validation to ensure our dispatch is realistic. However, we do not expect to capture all of the behavior of real-world dispatch systems, and there is no 1:1 correspondence between modeled hours and the actual state of the grid in specific hours of 2020.

We ran validation (Figure S2) to compare modeled fuel mix over the course of the modeled year to actual 2020 generation. 2020 generation data comes from OGE Miller et al. 2023 in the EI and Eurostat in Europe. Fuel types are within reasonable errors to actual generation. Future research and validation may improve model performance when compared to real data. However, we do not expect the models to perfectly match the 2020 generation because of the limitations discussed above. Instead, our goal is to approximate the overall patterns of dispatch closely enough to capture power flow patterns and trends across the network.

The OPF approach also comes with limitations. First, we may not capture operational constraints or deviations from optimal dispatch, for example, ISOs individually running OPF to dispatch their generators or US coal plants self-scheduling by bidding in at lower than their operational costs are not captured. Additionally, some limitations stem from the OPF implementation, which prioritizes computational efficiency. First, we run linear OPF (LOPF) instead of a full nonlinear OPF, which means that line losses are not considered in our dispatch or power flow results. Second, we run each day of the model year in parallel, and disregard generator startup and shutdown costs. This may result in unrealistic dispatch patterns for generators with high startup costs. Future work may explore the Breakthrough OPF approach, which is to run OPF on each day sequentially and constrain each day's OPF run to start with the end state of the prior day.

2 Supplemental discussion

2.1 Robustness of results to clustering

As discussed in the main text methods, we cluster both models to 1024 nodes. Because the European model covers a larger area than the Eastern U.S. model, its resulting node density (5019 km^2) is lower than that of the Eastern U.S. model (4581 km^2). However, because the original European network is at a lower resolution (limited to higher power lines) than the original U.S. model, fewer original model nodes are aggregated to each clustered model node in the European model (5.3 original nodes per clustered node) than in the Eastern U.S. (68 original nodes per clustered node). See Table S2.

To test the impact of these varying clustering levels on our results, we explore an additional grid, the Texas interconnect. The model we use is part of the same original model as our Eastern U.S. model, and we use the same configuration. We select this grid because it is small enough to run unclustered. We run physical delivery analysis on the unclustered Texas grid and on versions clustered to 512, 128, and 37 nodes. These clustering levels range across the node densities in the clustered models used in our main results (Table S3). We create four arbitrary regions, dividing the grid along its median latitude and longitude, to test our regional and boundary-set level metrics (Figure S16).

Nodal delivery distance. We find that spatial patterns of upstream (Figure S15) and downstream (Figure S14) delivery distance show consistent spatial patterns and magnitudes across clustering levels. Note that in an earlier version of our delivery distance metric, we explored using average delivery distance (in the final metric we use median). We found that average delivery distance magnitudes were sensitive to the level of clustering, perhaps because in less clustered networks, a generator will deliver to a higher number of nearby nodes.

Regional metrics. We find that between-region patterns of median delivery distance and expected deliverability are generally consistent across clustering levels, with some exceptions (Figure S17). Systematic changes in the magnitude of expected deliverability are discussed below in 'Boundary set level metrics'.

In delivery distance, we find that the regions with the highest downstream delivery distance (region B) and the highest upstream delivery distance (region D) are the same across all clustering levels, except the most clustered 37-node model. This indicates that there is a minimum resolution needed to get accurate delivery distance results.

We believe inconsistencies in regional results between clustering levels are largely driven by nodes which are in different regions in higher resolution models being clustered together in lower resolution models (Figure S16). Future work could improve this by using models where nodes are never clustered across a boundary. In the current models, European nodes are never clustered across country boundaries. U.S. nodes have no boundary-related guarantees during clustering.

Boundary set level metrics. The magnitude of boundary set metrics (expected deliverability and regional connectivity) varies systematically with clustering (Figure S18). Deliverability increases with increased clustering, since the network has fewer overall nodes and each pair is more likely to pass the same delivery cutoff. Modularity decreases with increased clustering, since fewer nodes mean that there are fewer strong bilateral delivery relationships between close-by nodes that have been clustered into a single node. Because these changes are systematic and should happen uniformly across the network with clustering, we do not expect them to affect the relative performance of boundary-set level metrics.

2.2 Robustness of results to model dispatch

We know from comparison of our modeled dispatch to real data (Figure S2) that coal and oil dispatch have some systematic error in both networks, with natural gas overdispatched relative to coal when compared to ground truth data. To evaluate the potential impact of this or other differences in dispatch between modeled and real data, we compare the European physical delivery metrics of the model used in our main results with a model whose fuel prices have been altered to shift coal and natural gas dispatch (Figure S19). The modified network has lignite dispatch 20 % higher than in the original network, and CCGT (natural gas) dispatch 10 % lower.

Note that we do not use this altered dispatch in our main results, even though at the category level it better matches actual 2020 dispatch, because the change to fuel prices was arbitrary and may not accurately reflect dispatch in time and space, even if the annual total dispatch is closer to ground truth. For the results presented in the main text, we rely on validation by the teams that developed each model, including the fuel prices.

We find that regional metrics are sensitive to changes in dispatch, with delivery distance (Figure S21) being the most sensitive and expected deliverability (Figure S20 being sensitive in some regions (in particular, the Balkans). We evaluate both metrics on the European country boundary set. These sensitivities to dispatch are likely due to spatial biases in coal and natural gas resources. With increased coal dispatch, regions with more coal will have more generation and therefore longer downstream delivery distances and potentially higher deliverability, while regions with decreased natural gas generation will see the opposite effect. This sensitivity means that caution should be taken when evaluating regional metrics on models with known discrepancies from real-world data. Even with this sensitivity, broad trends in delivery distance and expected deliverability, for example, longer delivery distances in Scandanavia compared to mainland Europe, are preserved.

Boundary set level metrics (regional connectivity and load-weighted average expected deliverability) are very robust to changes in dispatch (Figure S22). This may be because the regional effects described above average out at the boundary set level, resulting in the same relative conclusions. This gives us confidence when evaluating boundary sets, even on models which may differ somewhat from the real-world systems they represent.

2.3 Correlations between delivery distance and grid variables

We explore correlations between delivery distance and nodal and area-average generation, load, exported generation, and line capacity.

The analysis here is limited to an exploration of two-way linear relationship strength using correlation. Future work could include using a PCA analysis and linear regression to test whether combinations of these or other explanatory variables could explain more of the variance in physical delivery than the individual correlations considered here. Another direction for future work is to use an experimental approach where a specific feature of the grid model is changed to explore the resulting changes to delivery distance.

The correlation between delivery distance and each variable is generally small, with the largest magnitude around 0.4, meaning that at least 60% of the variance in nodal delivery distance is not explained by any individual explanatory variable. The nodal-level grid features are generally more predictive of delivery distance than area-average features, with the only exceptions happening at the smallest aggregation radius (50 km). This may mean that the heterogenaity in delivery distance is best explained by differences in the characteristics of individual nodes, with regional patterns playing less of a role.

Downstream delivery distance has relatively strong relationships with load and exported generation that are consistent across spatial averaging scales and across our two grid models.

Across both grids, downstream delivery distance is positively correlated with generation at or close to (in the Eastern Interconnect) the node and negatively correlated with generation in the area 150-300 km from the node. This may be because regions that have lots of generation over a large area are more likely to also have large load centers, which is negatively correlated with downstream delivery distance.

Upstream delivery distance has a negative correlation with both generation and exported generation, indicating that when a region has a lot of generation, or more generation than load, its loads are more likely to be served by nearby generation. The relationship between upstream delivery distance and load is smaller and varies in sign between the two grids, indicating that it may be a poorer predictor of delivery distance.

Line capacity is the most variable of the grid features we explored, with correlations that vary in sign and magnitude between the two grid models and across spacial scales. This may indicate that line capacity may not be a good predictor of delivery distance on its own, and a more nuanced or operational metric, like line congestion or through-node power flow, may be a better tool for explaining nodal physical delivery.

2.4 Evaluating boundary proposals in Germany and Denmark

Our regional physical delivery metrics could also be used to evaluate proposed changes to regions. Germany and Denmark provide examples of how physical delivery metrics could be used to evaluate proposed changes to boundaries. Germany is currently one bidding zone (which also includes Luxembourg). There are multiple proposals to divide the country into bidding zones which more accurately reflect transmission constraints across the country. We find support for these arguments when looking at delivery distance across Germany (See Figure S9b), which is only 100 km. While splitting Germany into two bidding zones (Panel B, "German Proposal") would not lengthen the delivery distance in either new bidding zone, reducing the size of the bidding zones would bring each more in line with the actual delivery of electricity. When we consider expected delivery (Panel A), we see that splitting Germany into two bidding zones would slightly increase expected deliverability, with 67.6 % and 64.5% of load-generator pairs in each of the two new bidding zones now deliverable (deliverability cutoff 1kw and 168 hours).

Denmark (red bars of Figure S9) currently has two bidding zones, one for each of Eastern and Western Denmark. We see that these bidding zones currently score very differently from one another, with Eastern Denmark much more deliverable and Western Denmark having longer delivery distances. This trend appears to be caused by stronger directional patterns in power flow in Western Denmark decreasing deliverability but increasing delivery distance. Combining the bidding zones would cause a neutralizing effect (i.e. the scores on both metrics are between the scores of the original bidding zones).

Comparing Denmark and Germany demonstrates the variability in delivery and deliverability between regions, even when using the same set of boundaries. Denmark (across proposed and actual bidding zones) has longer delivery distances than German, even though it is a smaller country. It also has higher expected deliverability. These differences can be even more extreme across some balancing authorities (BAs) in the US, which are extremely variable in size, ranging from individual townships in some Florida BAs to multi-state BAs in the middle of the country.

Note that the analysis here does not consider operational changes that would result from changes to bidding zones, which can be significant Brouhard et al. 2023.

References

Bialek, J.W. and P.A. Kattuman (2004). "Proportional sharing assumption in tracing methodology". en. In: *IEE Proceedings - Generation, Transmission and Distribution* 151.4, p. 526. ISSN: 13502360. DOI: 10.1049/ip-gtd:20040351. URL: https://digital-library.theiet.org/content/journals/10.1049/ip-gtd_20040351 (visited on 12/01/2022).

Brouhard, Thomas et al. (2023). "A clustering approach to the definition of robust, operational and market efficient delineations for European bidding zones". In: *IET Generation, Transmission & Distribution*.

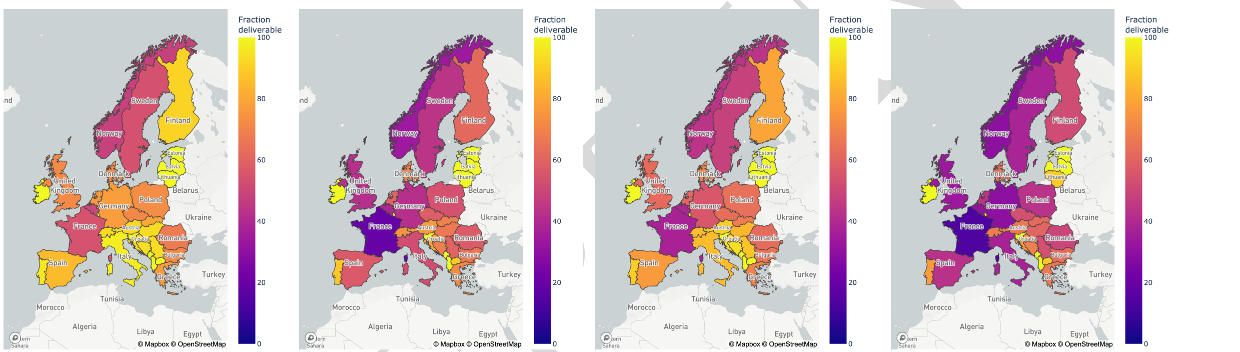
Gotzens, Fabian et al. (2019). "Performing energy modelling exercises in a transparent way-The issue of data quality in power plant databases". In: *Energy Strategy Reviews* 23, pp. 1–12.

Hörsch, Jonas et al. (Nov. 2018). "PyPSA-Eur: An Open Optimisation Model of the European Transmission System". en. In: *Energy Strategy Reviews* 22. arXiv:1806.01613 [physics], pp. 207–215. ISSN: 2211467X. DOI: 10.1016/j.esr.2018.08.012. URL: http://arxiv.org/abs/1806.01613 (visited on 10/05/2022).

Kang, Chongqing et al. (Sept. 2015). "Carbon Emission Flow From Generation to Demand: A Network-Based Model". en. In: *IEEE Transactions on Smart Grid* 6.5, pp. 2386–2394. ISSN: 1949-3053, 1949-3061. DOI: 10.1109/TSG.2015.2388695. URL: http://ieeexplore.ieee.org/document/7021901/ (visited on 09/26/2022).

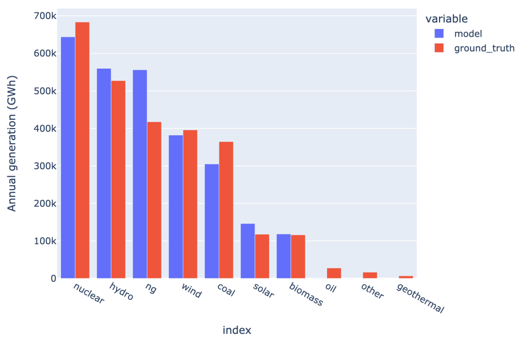
Miller, Gregory J et al. (2023). "Evaluating the hourly emissions intensity of the US electricity system". In: *Environmental Research Letters* 18.4, p. 044020.

Wu, Dongqi et al. (2021). "An open-source extendable model and corrective measure assessment of the 2021 texas power outage". In: *Advances in Applied Energy* 4, p. 100056.

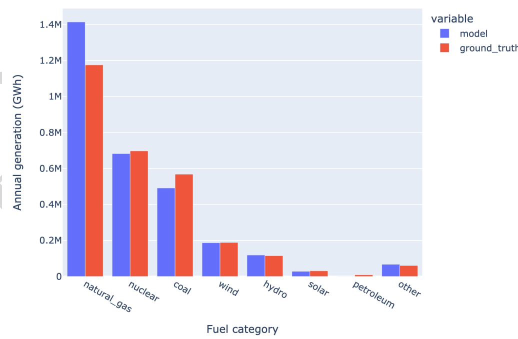


(a) 1kw, 24 hours cutoff (b) 1MW, 24 hours cutoff (c) 1kw, 168 hours cutoff (d) 1MW, 168 hours cutoff

Figure S1: Regional cutoff-based deliverability across four cutoff values (C_p = 1kw and 1 MW; C_t = 24 and 168 hours). Relative relationships between regions remain stable, though stricter cutoffs result in lower overall cutoff-based deliverability scores.

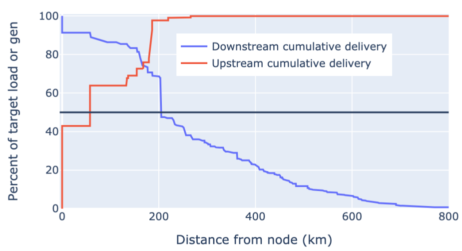


(a) Europe

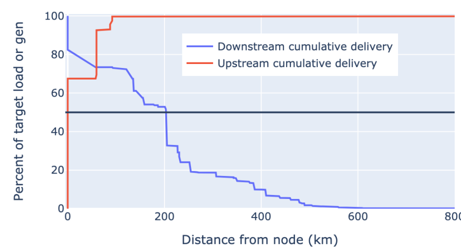


(b) Eastern interconnect

Figure S2: Comparison of 2020 ground truth and model fuel mixes over the model year. Fuel categories are different because each comparison uses the more granular possible category matching between ground truth and model categories.

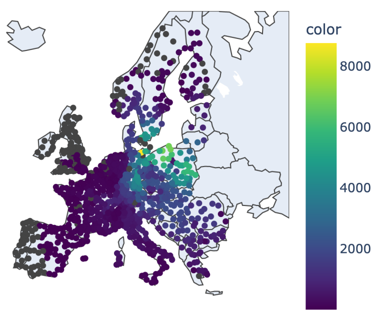


(a) Time 1, Delivery distances to DE1 17

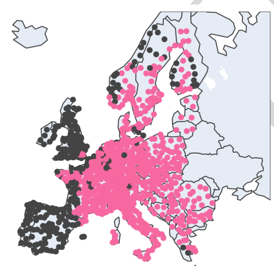


(b) Time 2, Delivery distances to DE1 17

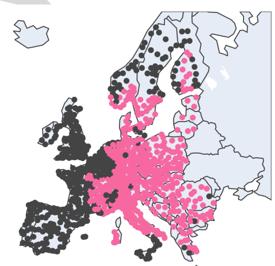
Figure S3: Cumulative consumption and delivery from node DE1 17 at two timestamps. The blue curves correspond to downstream delivery distance, or the distance traveled by power from generators at DE1 17. At a given distance x from DE1 17, y_{blue}% of power generated at DE1 17 remains to be consumed at more distant loads. The red curves correspond to upstream delivery distance, or the distance traveled by power serving the load at DE1 17. At a given distance x from DE1 17, y_{red}% of power generated at DE1 17 has been generated at that distance or further. The median delivery distance is indicated by the horizontal line at 50%. At distances close to the node (0-100 km for the upstream delivery distance curves), there is significant delivery from or to nearby nodes, resulting in the steps seen here, each of which corresponds to a large amount of consumption of the node's power (downstream) or generation for the node's load (upstream). a corresponds to the first row of main text Figure 1, while b corresponds to the second row of main text Figure 1.



(a) Number of hours where nodes deliver at least 1kwh to DE1 17



(b) Deliverable nodes to under a 24 hour, 1kwh cutoff



(c) Deliverable nodes to under a 168 hour (1 week), 1kwh cutoff

Figure S4: Bilateral deliverability from node DE1 17. a shows the count of hours over the model year where each node delivers at least 1 kwh to the load at DE1 17. b and c shows the nodes (green) passing a 24 and 168 hour cutoff, respectively.

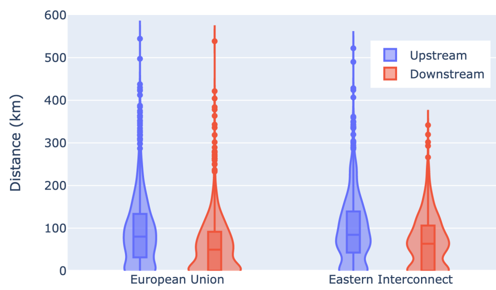
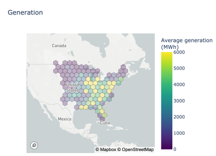
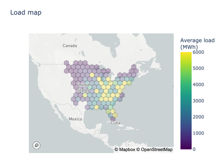


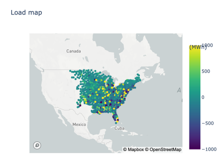
Figure S5: Violin and box and whisker plots of median nodal delivery distances across the US Eastern Interconnect and European grids.



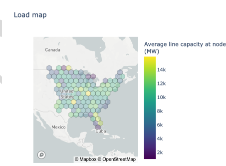
(a) EI generation (Spatial average)



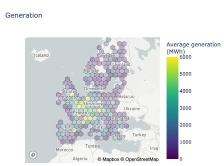
(b) EI load (Spatial average)



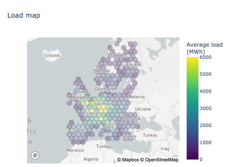
(c) EI exported gen. (Spatial average)



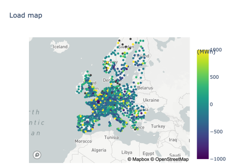
(d) EI line capacity (Nodal)



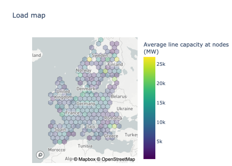
(e) European generation (Spatial average)



(f) EUR load (Spatial average)

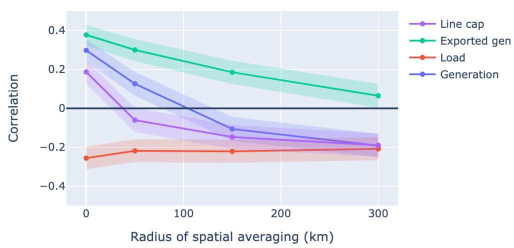


(g) EUR exported gen. (Spatial average)

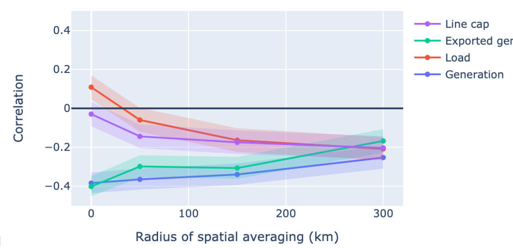


(h) EUR line capacity (Nodal)

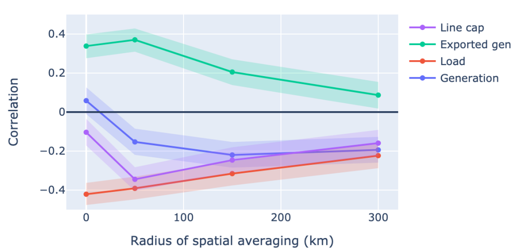
Figure S6: Generation, load, excess generation (generation - load), and connected transmission capacity across Eastern Interconnect (EI) and European (EUR) grid models.



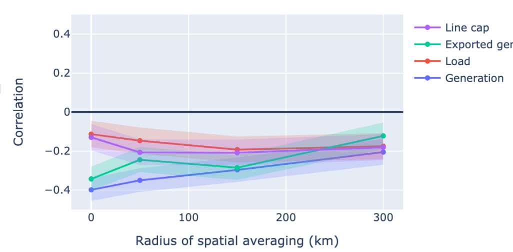
(a) EUR, downstream delivery distance



(b) EUR, upstream delivery distance

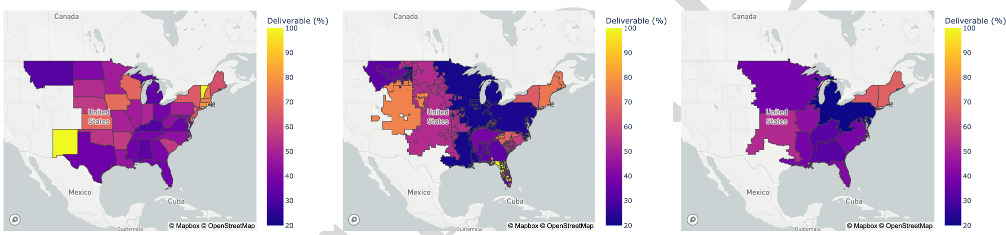


(c) EI, downstream delivery distance



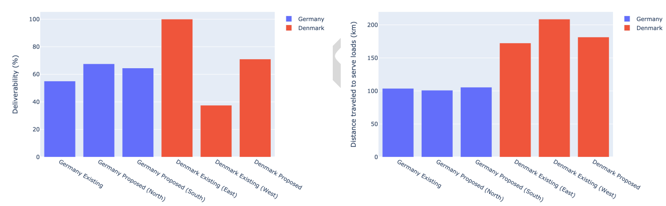
(d) EI, upstream delivery distance

Figure S7: Delivery distance across the US Eastern Interconnect and European grids. Bands show 95% CI, calculated using a Fisher transformation and z test.



(a) States (b) Balancing Authorities (c) NERC regions

Figure S8: Regional variability across boundaries in the US grid.



(a) Expected deliverability (b) Upstream delivery distance

Figure S9: Expected expected deliverability (1kw, 168 hour cutoff) and upstream delivery distance across existing and proposed bidding zones in Germany and Denmark.

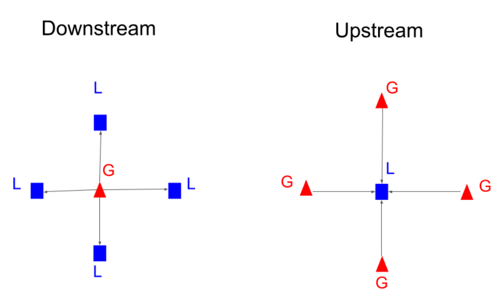
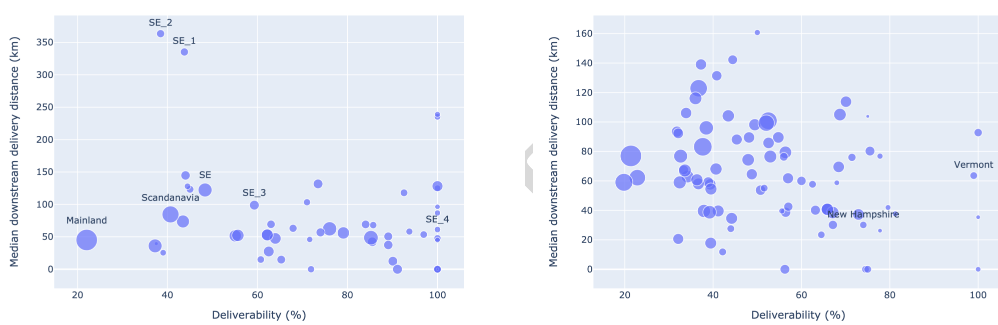


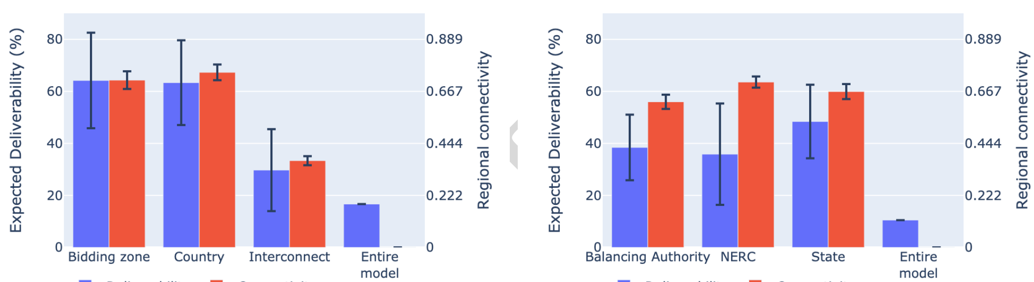
Figure S10: Downstream (left) and upstream (right) perspectives of delivery.



(a) European grid

(b) Eastern Interconnect

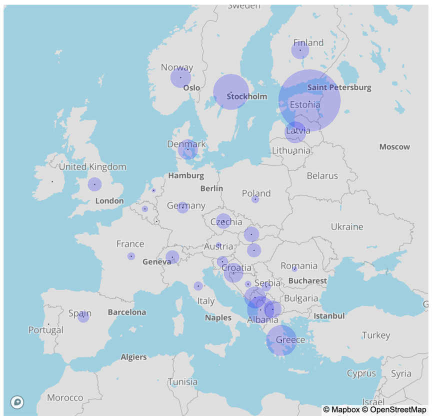
Figure S11: Regional expected deliverability scores (x) compared to median regional downstream delivery distances (y) and mean between-node distances (marker size) across our two grid models. Small regions towards the upper left of the graph have low deliverability relative to their delivery distance, indicating that directional biases in power flow may influence deliverability. Labels are centered immediately above the corresponding point.



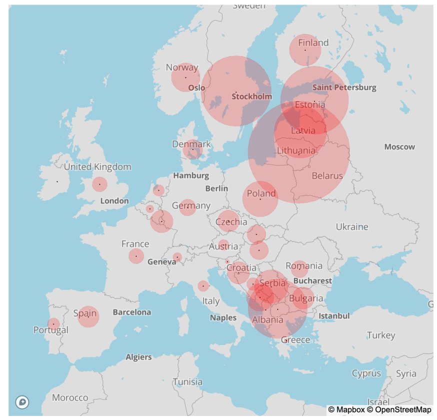
(a) European grid

(b) Eastern Interconnect

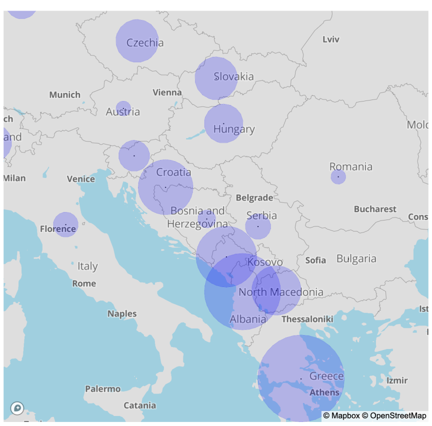
Figure S12: Standard deviation in regional connectivity scores (red) and expected deliverability scores (blue). For regional connectivity scores, which are calculated at every time stamp and averaged to show an overall score, we show variability over model time. For expected deliverability, which is calculated in each region and averaged to the boundary set, we show variability over regions in each boundary set (weighted by annual regional load). Note that these are not error bars on the average socre itself, instead, they indicate the variability of data averaged over.



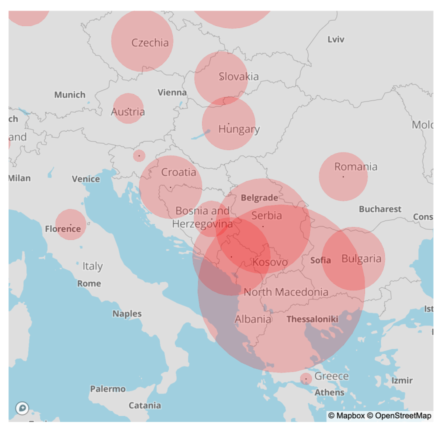
(a) Downstream delivery distance



(b) Upstream delivery distance

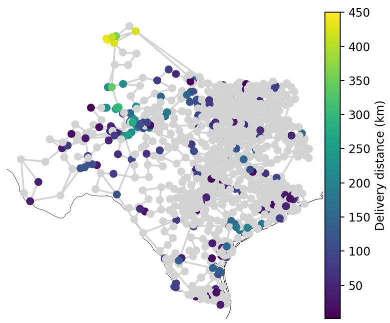


(c) Downstream delivery distance (zoomed)

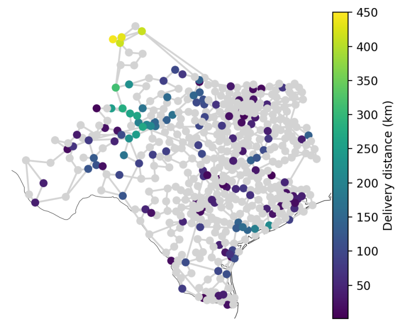


(d) Upstream delivery distance (zoomed)

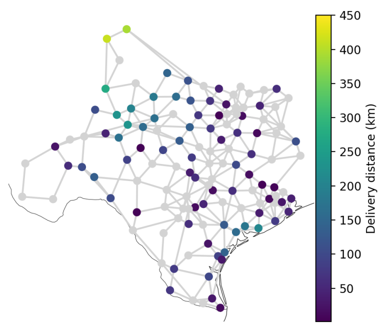
Figure S13: Median upstream (red, left) and downstream (blue, right) delivery distances for each country at two map zoom levels.



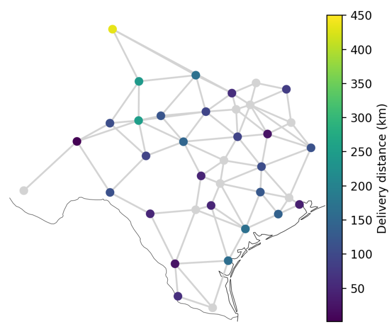
(a) Original network



(b) Clustered to 512 nodes

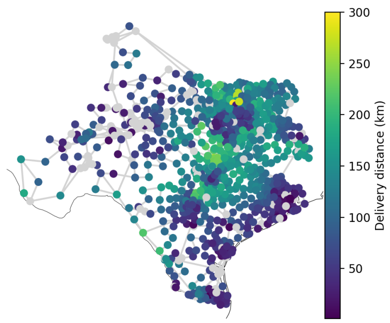


(c) Clustered to 128 nodes

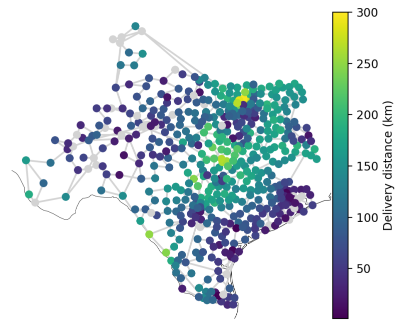


(d) Clustered to 37 nodes

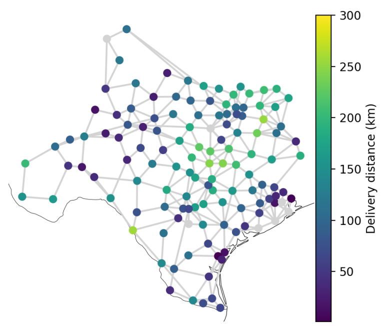
Figure S14: Downstream delivery distances across three levels of clustering in the Texas grid. Node color indicates the median distance traveled by power serving generators at that node.



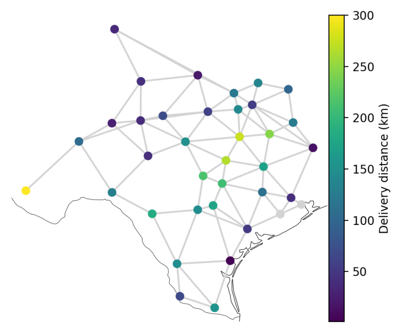
(a) Original network



(b) Clustered to 512 nodes

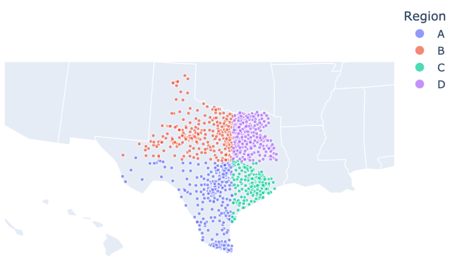


(c) Clustered to 128 nodes

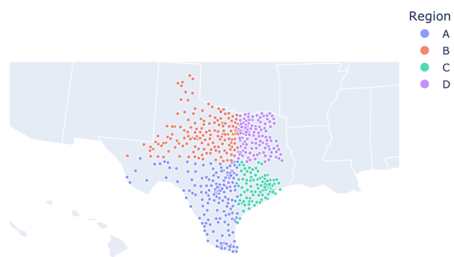


(d) Clustered to 37 nodes

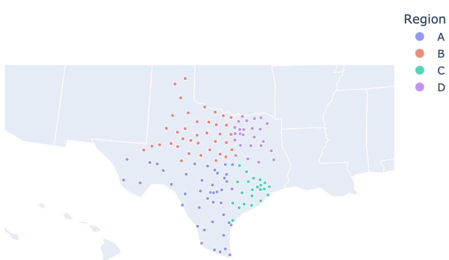
Figure S15: Upstream delivery distances across three levels of clustering in the Texas grid. Node color indicates the median distance traveled by power serving that load.



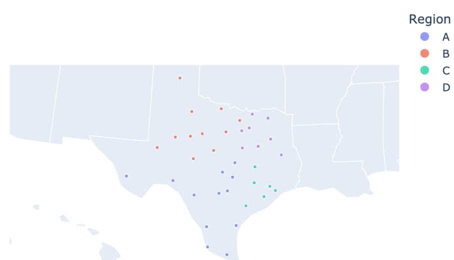
(a) Original network



(b) Clustered to 512 nodes

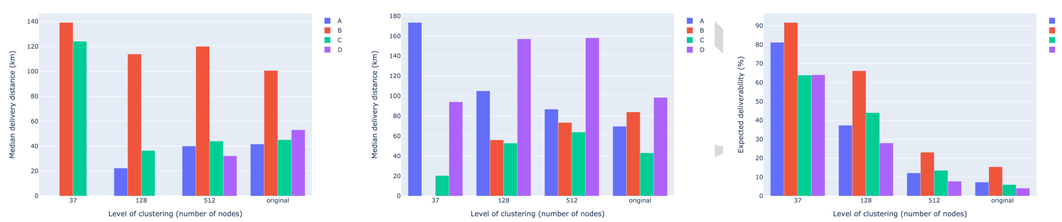


(c) Clustered to 128 nodes



(d) Clustered to 37 nodes

Figure S16: Example Texas boundaries under four levels of clustering



(a) Downstream delivery distance (b) Upstream delivery distance (c) Expected deliverability

Figure S17: Regional delivery distance and deliverability across clustering levels using arbitrary grid boundaries in Texas. While the strongest trends are consistent over clustering levels, the highest level of clustering (to 37 nodes) produces significant changes in regional metrics.

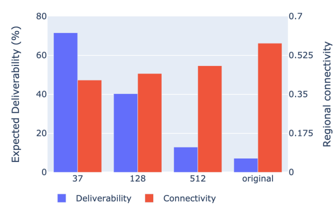


Figure S18: Boundary set level metrics in Texas across clustering levels.

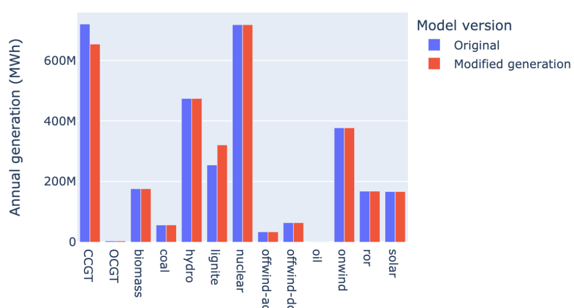


Figure S19: Compare generation between the Europe model used in the main text (blue) and the model with altered dispatch used for evaluating metric sensitivity to dispatch (red).

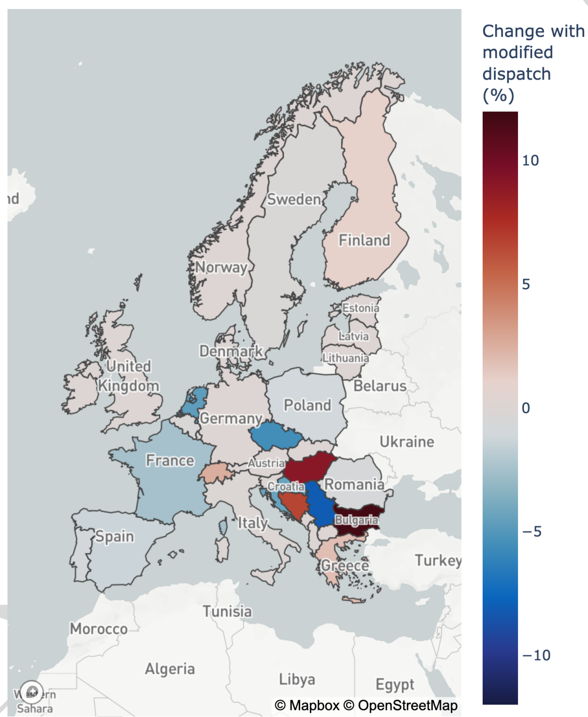
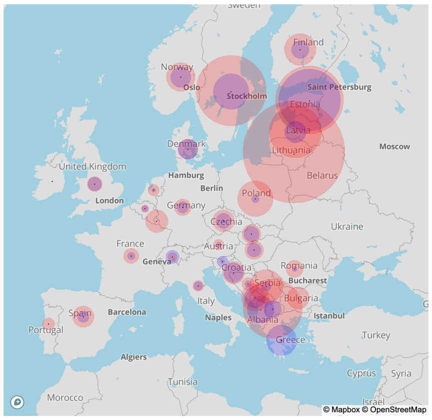
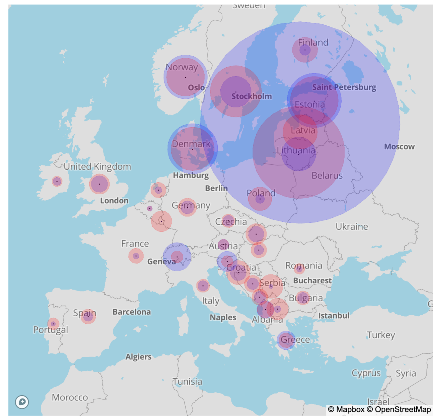


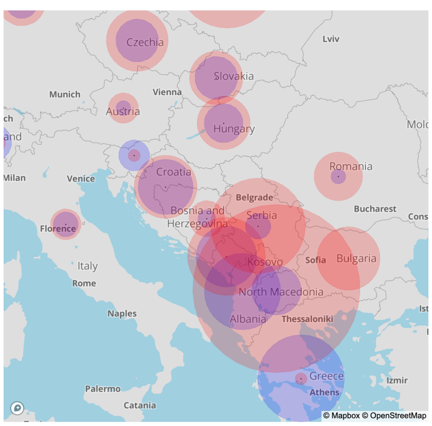
Figure S20: Differences between expected deliverability (using country boundary set) after altering generation, as a percentage of the expected deliverability scores of the model used in the main text.



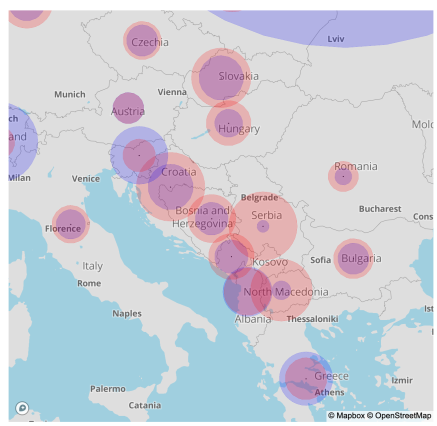
(a) Main text dispatch



(b) Altered dispatch

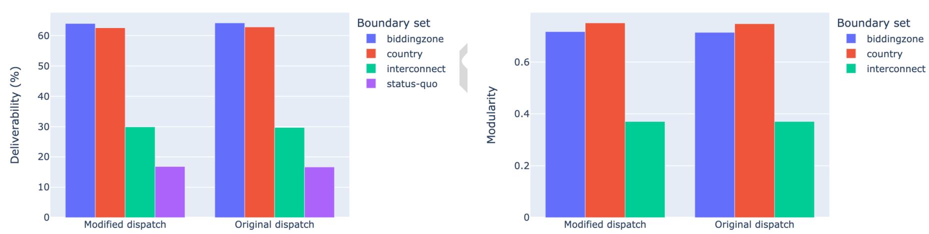


(c) Main text dispatch (zoomed)



(d) Altered dispatch (zoomed)

Figure S21: Median upstream (red) and downstream (blue) delivery distances for each country in the modeled dispatch used in the main text (left) and altered dispatch (right). See Figure S19 for differences in dispatch between models. Delivery distance is the most sensitive of our three metrics to dispatch.



(a) Expected deliverability (b) Regional connectivity

Figure S22: Boundary set level metrics across

	Quantile	Upstream	Downstream
EUR	0.25	31.71	0.00
	0.50	80.26	49.55
	0.75	133.29	91.35
	0.90	196.12	133.52
	0.99	371.42	289.20
US	0.25	42.62	0.00
	0.50	84.41	63.40
	0.75	139.19	106.13
	0.90	200.14	148.03
	0.99	342.05	240.94

Table S1: Upstream and downstream delivery distance quantiles in the EI and European grid

Network	Original node density (km^2/node)	Clustered node density (km^2/node)	Ratio of clustered to original nodes
Eastern Interconnect	66	4581	68
Europe	952	5019	5.3
Texas	198		

Table S2: Resolution of Eastern Interconnect, European, and Texas networks. The first two are used in the results, the Texas network is used only in clustering tests. Texas clustered node density and ratio depend on the level of clustering, and are shown in Table S3

Cluster level	Clustered node density (km^2/node)	Ratio of clustered to original nodes
512	772.6	3.9
128	3090.4	15.6
37	10691.2	54.1

Table S3: Resolution of three clustered versions of the Texas network (512, 128, and 37 nodes, respectively). These span the resolution of the networks used in our main networks (Table S2) and are used along with the unclustered Texas network to test the impact of clustering on physical delivery measurements.