Tracking physical delivery of electricity from generators to loads with power flow tracing

Gailin Pease,[*,†] Greg Miller,[†] and Wenbo Shi[†]

E-mail: gailin.pease@singularity.energy

Abstract

We use power flow tracing to measure the physical delivery of power between all generators and loads in models of the European and Eastern U.S. transmission grids. We believe this is the first work to measure physical delivery on high resolution models of real transmission grids. Physical delivery analysis can inform the integration of delivery requirements into clean energy procurement guidelines, such as those in the Greenhouse Gas Protocol (GHGP) standard, to support credible claims. We propose three new metrics to summarize physical delivery across a transmission grid: delivery distance (median distance traveled by power), expected deliverability (generator/load pairs which meet a physical delivery cutoff), and regional connectivity (which uses modularity, a metric from graph theory, to summarize how well boundaries align with physical delivery). We evaluate these metrics across Europe and the Eastern U.S.. We find that the distance and direction traveled by power is highly variable, and can be partially explained by patterns of power flow from large generation centers towards large load centers. We evaluate how existing boundaries align with physical delivery and find that bidding zones and countries in Europe and states in the Eastern U.S. perform better than other options.

Introduction

Electricity generation accounts for a large fraction of the world's greenhouse gas emissions, contributing to the climate crisis. Governments and electricity consumers are increasingly interested in procuring clean electricity, and understanding where electricity is delivered can be a useful input for credible and effective procurement guidelines. However, definitions of deliverability vary widely through the academic literature and often do not consider the physics of electricity delivery. We use delivery based on power flow tracing as a simple, physics-grounded approach which can provide granular insights into electricity delivery.

Renewable energy markets

Electricity consumers are increasingly interested in tracking their carbon footprint using accounting standards such as the GHG protocol[1] and reducing their emissions through the voluntary procurement of carbon free electricity.[2] While some point to the sheer volume of clean energy transactions as an indicator of the success of this voluntary market model,[3] others[4,5] have criticized the voluntary market as greenwashing which fails to deliver the promised grid decarbonization.[5]

In addition to the fast-growing voluntary market, governments are increasingly introducing standards for clean energy procurement in their jurisdictions, sometimes including stricter location requirements than voluntary procurement guidelines.[6,7]

Grid operators dispatch generation resources to balance load across time and location while maintaining system security and keeping the grid's components within their operational limits.[8] Traditionally, voluntary procurement has not attempted to match the time or location of procured generation with the buyer's load. One emerging voluntary procurement approach, 24/7 procurement, aligns electricity procurement with the times and locations when a buyer is consuming electricity.[9] Modeling suggests that 24/7 matching in the US and European grids, paired with location matching based on operational regions, would reduce system emissions over annual matching, which is the current status-quo.[10,11] Given that the grid is balanced in both time and space, defining the right location matching criteria for 24/7 matching is key for procuring energy that aligns with a specific load.

Defining physical delivery and deliverability

Just as 24/7 is the principle that underpins temporal matching of clean energy supply with demand, *deliverability* underpins locational matching. Conceptually, deliverability is the the ability of generated electricity to serve a load through the grid's transmission and distribution network. The simplest deliverability approaches treat electricity as uniformly mixed (a "copperplate" model) within either a synchronous grid (eg, the Eastern, Western, and Texas interconnects in the US) or an operational region (eg, balancing areas in the US).

Other approaches attempt to identify which loads could be impacted by a generator, including both delivery directly from that generator and second-order changes in power flows due to that generator. Blumsack et al.[12] use electrical distance to define "deliverable" zones. Congestion-based approaches to deliverability similarly attempt to define deliverable zones, often based on locational marginal prices (LMPs), as regions within which there are no congestion constraints limiting the delivery of generated electricity.[13] Neither approach guarantees that generators will actually deliver power to a given load, and may not be stable over time.

Congestion-based delivery rules for procurement have been shown to reduce system emissions. For example, Ricks et al.[14] show, using a small example system, that a congestion-based deliverability requirement is important for building new renewable projects to minimize emissions from hydrogen production loads. One limitation of congestion-based deliverability in real-world procurement is the use of LMP prices to determine deliverable zones,[15] which are an indirect metric that can be affected by factors other than congestion. Physics-based definitions of congestion could be used alongside the physical delivery metrics developed in this work to give a fuller picture of both the physical delivery of dispatched, existing generation (physical delivery) and the potential for the delivery of additional generation (congestion-based metrics).

Finally, some approaches use power flow to define deliverability. Power flow measures the time-varying magnitude and direction of power along each of a grid's lines and buses. Kirschen et al.[16] is an early example of this approach, where the "domain" of a generator is all loads downstream on an acyclic power flow diagram. Achayuthakan et al.[17] proposed extending power flow tracing, originally designed to trace transmission costs,[18] to trace power from generators to loads. Since delivery based on power flow tracing is derived from the grid's underlying physics, we refer to it as *physical delivery* to distinguish it from other definitions of deliverability.

In this work, we measure the physical delivery from generators to loads on models of the European and Eastern US transmission grids, and propose three new metrics for summarizing and analyzing physical delivery. Our physical delivery findings on real grids can serve as a foundation for future work on designing markets and accounting protocols that align participant incentives towards grid decarbonization.

Methods

We extend the algorithm of Kang et al.[19], which used a matrix formulation of power flow tracing to track one feature through a power grid, and the conceptual approach of Achayuthakan et al.[17], which applied power flow tracing to measuring the physical delivery of electricity. (Algorithm provided in supplement section 1.3.) We believe we are the first to apply this approach to high-resolution models of real transmission systems, perhaps because of computational and analytical challenges created by data volume and complexity (representing delivery over a model year in Europe requires approximately 200 GB of data).

Node-specific metrics

Bilateral delivery

Bilateral delivery describes the volume of power delivered from each generator to each load in the network. We call these relationships *bilateral delivery relationships* to distinguish them from metrics which aggregate over many generator/load pairs.

Bilateral delivery at time t is represented by a matrix $P_t \in \mathbb{R}^{N_G \times N}$, where each element $P_{i,j,t}$ describes the delivery in MWh between generator i and load j at time t. N_G is the number of generators in the system, and N is the number of buses (each of which could have a load). Note we exclude the subscript t for simplicity of notation except when aggregating over time.

We can look at bilateral delivery from either the *upstream* perspective, where we focus on a single load and ask which generators contribute to it; or the *downstream* perspective, where we focus on a single generator and ask which loads it serves (Figure S10).

Bilateral delivery relationships are complex and three-dimensional, with a scalar value for every generator, load, and time combination in the model. While this data richness allows for nuanced evaluations of an individual node's generation and load, it can make it difficult to draw insights across the entire network. To address this, we introduce three aggregate metrics, *expected deliverability*, *delivery distance*, and *regional connectivity* (Table 1).

Although each metric provides valuable insights in the two grid models we consider in this work, each has limitations. We see these metrics as a starting point for future refinement to better explore the full range of insights available from bilateral delivery data.

Delivery distance

Delivery distance is the median distance traveled to each load or from each generator in the network; measured at every timestamp or aggregated over time. The delivery distance from a generator is downstream delivery distance, and is the the median distance traveled by power from the generator, or equivalently the distance traveled before 50 % of the generation from a generator has been consumed. The delivery distance to a load is upstream delivery distance, and is the distance within which 50 % of the power to serve a load was generated. We use the median instead of the mean because we found in sensitivity testing that average distances were sensitive to the resolution of a grid model. Using the median has the additional advantage that the metric could be extended in the future to look at other delivery quantiles, for example, the distance within which 90 % of power is consumed.

Here, we use the straight line distance in kilometers between each node pair as our distance function $f_d()$, but our definition of delivery distance can be used with any distance metric. Two potentially relevant metrics could be an electrical distance metric, a simple version of which measures the impedance between two buses on a network;[20] or electrical line distance in km. We use kilometers here because of its simplicity and relevance for procurement decisions, where electrical distance and topology data may not be available.

The downstream delivery distance d for a generator i with generation g_i (MW) is:

$$\begin{aligned} \min \quad & d \\ \text{s.t.} \quad & \textstyle\sum_{j=0}^{N} \mathbf{1}_d P_{i,j} \geq \frac{1}{2} g_i \end{aligned} \tag{1}$$

where we sum over the delivery to all loads in the network. Here $\mathbf{1}_d$ is an indicator variable determining whether node n_i falls within d of n:

$$\mathbf{1}_d := \begin{cases} 1 & \text{if} \quad f_d(i,j) < d, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Table 1: Proposed metrics

Metric	Definition	Scale	Units	Range	Benefits	Drawbacks
Delivery	Power delivered from each gen to each load	Bi	MWh	0, max g/l	Represents nuanced behavior of all Bi pairs	Difficult to draw general conclusions
Delivery distance	Median distance traveled by power	g/l, R	km	0, 500^B	Intuitive. Definition at each gen/load enables network-wide analysis	Relatively sensitive to model clustering and dispatch. Ignores directional trends.
Expected Deliverability	Percent of Bi pairs above a cutoff	Bi, R, BS	%	0, 100	Easy to compare regions and boundary sets	Requires arbitrary choice of cutoff. Ignores magnitude of delivery.
Regional connectivity	Modularity of a BS on the delivery graph	BS	-	0^A, 1	Summarizes bilateral delivery. Accounts for magnitude and direction of delivery.	Limited to BS-level analysis. Less intuitive.

Bi = bilateral (each generator and load pair), BS = boundary set, R = region, g/l = generator or load

^{A} Theoretically modularity can be as low as $-1/2$, but this is not possible in these graphs and boundaries, since the graphs are strongly connected locally and weakly connected at greater distances

^{B} The longest distances in our models were under 500 km, but larger distances are theoretically possible

The upstream delivery distance for a load k is similar, but we now sum over all generators m in the network to determine the minimum distance within which those generators provide at least 50 % of the load at k:

$$\begin{aligned} \min \quad & d \\ \text{s.t.} \quad & \textstyle\sum_{m=0}^{N_G} \mathbf{1}_d P_{m,k} \geq \frac{1}{2} l_k \end{aligned} \tag{3}$$

Expected Deliverability

For some applications, it may be useful to determine whether electricity can be physically delivered between a generator/load pair. We call this concept *physical deliverability*, to highlight that the user is interested in the potential of physical delivery during normal grid operation, not the magnitude of that delivery. We note that physical deliverability differs from other definitions of deliverability, for example, congestion-based deliverability, which ask whether a generator could ramp up (or be built) to meet a load, regardless of whether power is physically delivered between the pair.

We propose expected deliverability as a metric for measuring physical deliverability. Expected deliverability is defined using two cutoffs, C_p, which describes the amount of power for a generator-load pair to be considered "deliverable" in a single hour, and C_t, which describes the number of hours during the model year where the power cutoff must be met. One limitation of this metric is that because we only model one year, we do not capture all possible grid states, and may miss some deliverable generator-load pairs.

We describe whether a generator g and load l meet the cutoff C_p in a single hour t using the indicator variable $\mathbf{1}_{C_p}$:

$$\mathbf{1}_{C_p} := \begin{cases} 1 & \text{if } \; P_{l,g,t} > C_p, \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

We then determine whether g and t meet the cutoff over the model year using the indicator variable $\mathbf{1}_{C_p,C_t}$:

$$\mathbf{1}_{C_p,C_t} := \begin{cases} 1 & \text{if } \; \sum_t \mathbf{1}_{C_p} > C_t, \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Note that this definition focuses on the downstream perspective. This allows us to answer questions like "where is generator x deliverable to?". Because of the asymmetry between upstream and downstream bilateral relationships (see Results), upstream and downstream expected deliverability may not always be the same. It is possible to formulate upstream expected deliverability, but we leave the exploration of this additional metric to future work.

Unlike delivery distance, which is a physical delivery metric that looks at the distance within which most power is served, expected deliverability looks at the widest extent traveled by even a small portion of power from a generator. Although the two are conceptually different, we do expect some alignment between the metrics, with generators with longer downstream delivery distances having larger expected deliverability ranges.

Regional metrics

As recent legislation demonstrates,[6,7] regions are a common and useful way to define procurement requirements. The regional metrics we propose provide quantitative evidence for the relative ability of existing boundary sets to accurately represent underlying physical deliverability.

While the metrics introduced above are useful for evaluating delivery and deliverability in specific cases, individually analyzing every voluntary clean energy market transaction for deliverability may not be practical. As recent legislation demonstrates,[6,7] regions within which generation is considered deliverable are a common and useful way to define procurement requirements. In this section, we introduce how the above metrics can be aggregated to evaluate physical delivery and deliverability within a single region or a set of regions defined by a boundary set.

Each aggregation prioritizes different features of the underlying bilateral delivery relationships.

Aggregating delivery distance

To aggregate delivery distance to the region, we take the median of the delivery distances for all loads (upstream delivery distance) or generators (downstream delivery distance) over the region. This provides *regional delivery distance*, a descriptive statistic of physical delivery in a region, which can be contextualized by comparison to measures of region size. Here, we compare regional delivery distance to the average between-node distance in a region. More complex metrics, for example mean between-node distance weighted by load size, could be explored as more robust measures of region size.

Aggregating expected deliverability

Regional expected deliverability measures how likely a random generator/load pair is to pass a given expected deliverability cutoff. Regional expected deliverability is measured as M_R/N_R, where M_R is the number of generator/load pairs in region R meeting the cutoff out of N_R total generator/load pairs in R.

For the results shown here, we use a cutoff of 1 kW delivered in at least 168 hours (equivalent to one week of hours) of a typical year ($\approx$ 0.3% of hours). While the choice of cutoff is arbitrary, sensitivity testing comparing results under four cutoffs found that the relative performance of regions is robust over cutoff choices, although expected deliverability decreases as cutoffs become stricter (Figure S1). Very strict cutoffs (eg, tens of MW of power over months of the model year) may benefit smaller regions or be sensitive to the number of nodes in a region. Following these findings, we use a cutoff in the middle of the range we tested, and focus on comparative results which are robust to cutoff choice.

Measuring regional connectivity with modularity

Modularity is a metric from graph theory used to measure the quality of a set of boundaries dividing a graph. Conceptually, modularity compares the connectedness of each region in the graph to what the expected connectedness would be if edges were randomly distributed. In our application, boundaries that group sections of the network where generators and loads have high delivery to one another perform well, while boundaries that divide those sections or have many generators and loads which do not deliver to one another score poorly. Modularity is powerful because it can summarize all bilateral delivery connections on a network without reducing each connection to true/false (as with expected deliverability) or to a median distance per generator or load (as with delivery distance).

To calculate modularity, we first construct a graph of all bilateral delivery relationships. Note that this is no longer the topology of the underlying grid; instead, it represents the bilateral delivery $P_{i,j}$ of each generator, load pair i, j as an edge.

Once we have defined the graph, we calculate modularity:[21]

$$\frac{1}{2m}\sum_{ij}[P_{ij} - \gamma\frac{k_i k_j}{2m}]\delta(c_i, c_j) \tag{6}$$

Where m is the total number of edges, k_i is the degree of node i, and $\delta(ci, cj) = 1$ if and only if i and j are in the same region. We call this bilateral delivery-based modularity *regional connectivity.*

Regional connectivity has several drawbacks. It is likely unfamiliar to many power systems experts and has no natural units, which may limit interpretability. Also, we currently calculate modularity only for the entire boundary set, so we cannot identify which regions are most responsible for good or bad performance of a boundary set's regional connectivity.

Models and data

The proposed methods require a complete network topology, along with generation, load, and power flow data. We use open source models of the European[22] and U.S.[23] (Eastern interconnect only) power grids, clustered to 1024 nodes. We use model parameterizations which approximate the behavior of the 2020 grid, and run linear OPF for each day of the model year to generate generator profiles and power flow results. See supplement section 1.4 for more detail on our model configuration.

We compare our modeled dispatch to actual 2020 dispatch from OGE[24] and Eurostat (supplement section 1.4.4) and find that all fuel categories are within 20% of their actual dispatch. These errors in dispatch may impact some regional metrics but we do not expect them to alter our conclusions, which focus on trends and relative metric performance which are robust to changes in model dispatch (supplement section 2.2).

Model limitations

Our models mirror the overall behavior of the 2020 grid but do not capture the actual historic behavior of the grid at specific hours. This limits our analysis to trends and typical behavior, which can be captured by the models, rather than temporally specific results.

We cluster both models to 1024 nodes before analysis. In the European model, clustering is necessary to avoid unrealistic transmission constraints near large population centers. Although clustering is not required in the Eastern U.S. network, we cluster the Eastern U.S. network for consistency with the European network and for computational feasibility. We performed sensitivity testing of our metrics to clustering (supplement section 2.1), and found that while clustering has systematic impacts on some of our metrics (more clustered grids have uniformly higher deliverability and lower regional connectivity), clustering in most cases does not impact the relative performance of boundary sets or regions. Exceptions occur when a grid is so highly clustered that it no longer accurately captures spatial patterns of delivery, or when clustering combines nodes with very different characteristics across boundary borders, neither of which we observe in the models used in our results here.

An additional limitation specific to the Eastern U.S. network is that it is a synthetic model, designed to replicate the behavior of the U.S. grid in a spatially granular manner but not based on real transmission topology. Benchmarking by the Eastern U.S. team[23] gives us confidence that our overall results, including regional analyses, will accurately represent the U.S. grid, but insights at a single node or line may not be accurate.

Regions and boundary sets

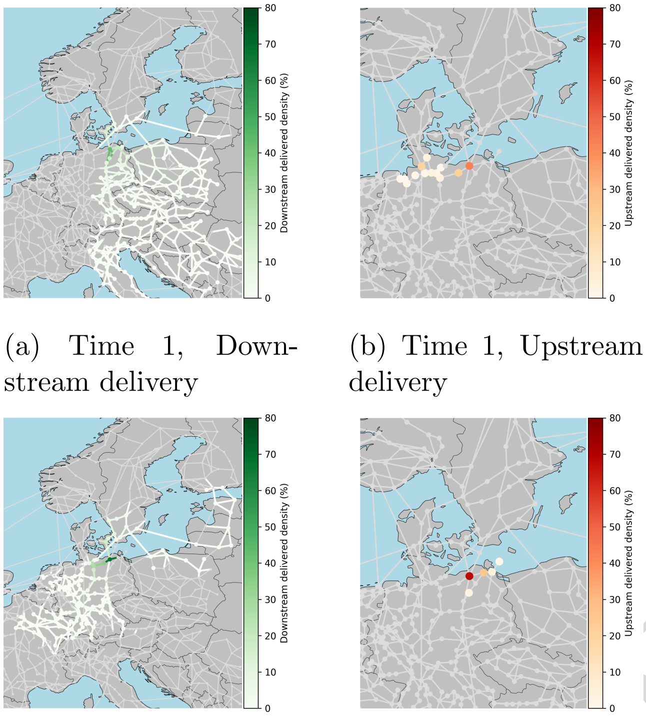
We evaluate our region and boundary set-level metrics on three boundary sets on each grid. Each boundary set defines a set of regions which uniquely and completely cover the grid. In the Eastern U.S., we consider balancing authorities ("BAs"), states, and NERC regions (NERC is a standard setting body whose Regional Entities cover the US). In Europe, we consider interconnects (or synchronous grids), countries, and bidding zones. By comparing metrics across boundary options, we are able to provide insights about which existing boundaries are more aligned with underlying physical delivery.

These boundary sets have varying histories and relationships to the underlying grid. For example, in the US, state boundaries do not define the operational boundaries of the grid, but they can affect the resource mix on the grid through policy. On the other hand, BAs are directly aligned with electricity markets in the U.S., but given the history of how regulated and deregulated regional markets developed, BAs can cover a single city or multi-state regions. This complexity means that physical delivery and deliverability can vary widely between and within regions.

Results

Bilateral delivery

The core result of our physical delivery analysis is bilateral delivery relationships between every load and generator pair in the network. An example of these relationships for a single node, DE1 17, is shown in Figure 1. DE1 17 is a Northern Germany node with a variety of generators and a small load. While limited in scope, analysis of individual bilateral delivery relationships will lay the foundation for un-



(a) Time 1, Downstream delivery

(b) Time 1, Upstream delivery

(c) Time 2, Downstream delivery

(d) Time 2, Upstream delivery

Figure 1: Example delivery to and from northern Germany node DE1 17 over two time stamps. Grey nodes receive (left) or serve (right) less than 1kw to DE1 17. Color indicates volume of power served or received as a percentage of DE1 17 generation (left) or load (right). Left panels show delivery from generators at DE1 17 (darkest green) at an example timestamp. Right panels shows delivery to loads at DE1 17 (darkest red). Note that we do not highlight lines here, because power flow tracing of generators does not provide data on the path traveled by power serving a specific load. The lower panels show delivery at a timestamp 10 hours after the upper panels.

derstanding the dynamics shaping broader patterns of delivery across our models.

Bilateral delivery relationships are highly variable in time and space. The two rows of Figure 1 are separated by only 10 hours, yet the primary direction of power flow has shifted between the two timestamps, from west-to-east in the first time stamp to east-to-west in the second timestamp. These large shifts are not due to the activity of DE1 17 alone; instead, they're the result of the behavior of the entire system.

Delivery distance

Single-node delivery distance

Delivery distance from generators at a node (downstream delivery distance) and to the load at a node (upstream delivery distance) summarize the spatial extent of bilateral relationships. Figure S3 demonstrates how delivery distance is derived from the bilateral relationships of DE1 17, the node shown in Figure 1, by looking at the cumulative delivery to and from the node over distance. The delivery distance curves in Figure S3 are typical, with most power consumed close to the node where it is generated but a long tail of smaller delivery to or from more distant nodes.

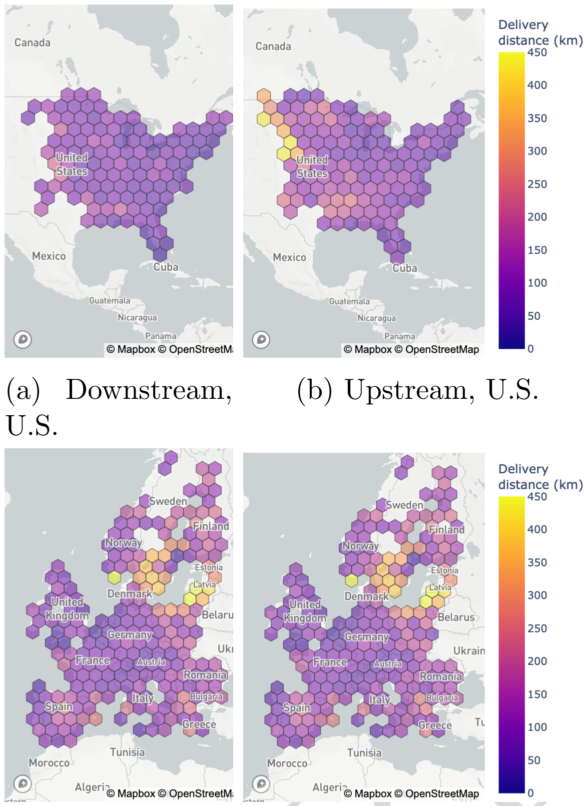
The median distance traveled by power from generators at DE1 17, 205.2 km, is the same between the two timestamps shown here. This is a typical downstream delivery distance for this node (its median downstream delivery distance over the model year is 203.1 km). The median upstream delivery distance changes between the timestamps, from 50.7 km in the first timestamp to 0 km in the later timestamp. A delivery distance of 0 km indicates that at least half of the power serving the load at this node was generated at this node, but it does not mean that no power serves the node from further away, as visible in Figure 1d.

Using delivery distance to understand spatial trends

While the median delivery distances of both models are similar (upstream 80.2 km in Europe and 84.4 km in the Eastern Interconnect; downstream 52.9 km in Europe and 62.4 km in the Eastern Interconnect), there is large variability in delivery distance within each grid model. [1]. Downstream delivery distance has an IQR range (75th percentile - 25th percentile) of 91 km in Europe and 106 km in the Eastern interconnect (Figure Figure S5, Table S1). This variability has implications for clean energy procurement informed by physical delivery, since the radius of generators delivering to a load is dependent on the location of the load. The delivery distance relevant for a specific load could be located on a map of delivery distances (such as Figure 2) or found using a load-specific delivery analysis. Alternatively, regional summaries of delivery distances could guide a procurement policy based on region boundaries similar to current protocols.[6,7]

Because delivery distance provides a scalar summary of physical delivery at each network node, we can easily relate delivery distance to potential drivers of physical delivery defined at the network nodes. While not directly applicable to procurement, this makes delivery distance a useful tool for understanding physical delivery. Here, we explore whether four simple drivers of power flow, specifically nodal and area-averaged annual generation, annual load, annual excess generation ("exports"), and transmission capacity, shown in Figure S6, affect physical delivery as measured by delivery distance.

Although the predictive power of each variable is generally small (maximum magnitude 0.4), we find significant and consistent correlations between downstream delivery distance and two explanatory variables, load and ex-



(a) Downstream, U.S. (b) Upstream, U.S.

(c) Downstream, Europe (d) Upstream, Europe

Figure 2: Downstream (left) and upstream (right) delivery distances for the the Eastern interconnect and European grids. The results are spatially binned to provide easier visualization of regional trends.

[1]Note that the statistics here are of the median distance for each generator across the model year, which allows us to look at the range of typical distances across generators and locations. Statistics of the entire set of delivery distances across all hours would show a wider range. Note also that when we take the median across generators, we treat each generator equally, regardless of its capacity or dispatch over the model year, which means our metrics reflect the typical generator, not the typical generated MW. Since delivery distance is positively correlated with generation, a weighted median of downstream delivery distance would be longer.

ported generation. Downstream delivery distance is negatively correlated with load, indicating that where there are large load centers near a generator, that generator's power is more likely to be consumed nearby. The opposite relationship is true of exported generation, indicating that when generation exceeds load, either at a specific node or across the region close to the node, generation from that node is likely to travel further. Both of these relationships become weaker when we consider larger spatial averages of the explanatory variables, indicating that the grid closest to the node has the largest impact on that node's physical delivery (Figure S7). These results support the intuitive idea that the relative location of load and generation centers is an important driver of physical delivery, a relationship which we leverage to explain regional trends seen in other metrics.

Expected deliverability

Single-node expected deliverability

Bilateral deliverability abstracts away the volume of delivered power, but unlike delivery distance, it preserves directional biases in delivery. A 1kw power cutoff C_p is shown in each panel of Figure 1, where nodes which are not deliverable to DE1 17 in the two example hours are shown in grey. Physical deliverability for this node over the entire model year is shown in Figure S4, which demonstrates how the patterns of delivery seen in Figure 1 are preserved by expected deliverability. Since both the east- and west- direction delivery patterns from DE1 17 are seen in multiple hours over the model year, those regions are included in expected deliverability regions, while uncommon patterns, like power flowing north into Scandinavia, are not.

Regional expected deliverability

Regional expected deliverability provides quantitative evidence for the relative physical deliverability of regions. When combined with regional delivery distance, the two metrics can provide nuanced insights of regional physical delivery patterns.

Regional expected deliverability is a function of region size, delivery distance, and directional biases. Where regions of similar size have similar directional power flow patterns, differences in delivery distance explain differences in regional expected deliverability. For example, Vermont and New Hampshire, two neighboring states in the northeast U.S., are close in size (95.2 km average inter-node distance in Vermont, 88.1 km in New Hampshire) but Vermont has a longer downstream delivery distance (63.6 km to New Hampshire's 30.1 km). The resulting regional expected deliverabilities are 99 % in Vermont, where most nodes are within the median downstream delivery distance, but only 74 % in New Hampshire (Figure S11).

Many regions, however, have diverging expected deliverability and delivery distance, indicating that expected deliverability is capturing asymmetric patterns of delivery not reflected by delivery distance. This illustrates the importance of comparing multiple metrics, each of which highlights a different feature of the underlying bilateral delivery data, to understand physical delivery trends. For example, in Europe, Sweden has a strong directional bias in delivery, with generation in the northern part of the country usually traveling south towards larger load centers in southern Scandanavia or mainland Europe. This leads to low expected deliverability in northern Sweden bidding zones and increasing deliverability towards the south as directional biases in power flows lessen. From north to south, deliverability scores for SE-1, SE-2, and SE-3, and SE-4 are 43.8 %, 38.5 %, 59.3, and 100 % respectively. This trend runs opposite to median downstream delivery distances in these regions (Figure S11 (b)), which are longer in the north where generators are further from large load centers (from north to south, median regional delivery distances are 335.2 km, 363.4 km, 98.9 km, and 61.3 km).

The large regional differences within Sweden are masked when considering Sweden as a whole, which has a 48.3 % expected deliverability. This illustrates a risk of regional physical delivery metrics: where physical delivery is heterogeneous within a region, a region-level statistic will not accurately reflect the physical delivery environment of all loads and generators within that region.

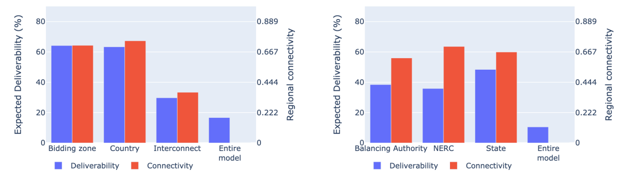
Evaluating boundaries

At the boundary set level, we evaluate physical delivery using expected deliverability and regional connectivity. Expected deliverability considers only whether physical delivery meets a certain threshold, while regional connectivity considers the magnitude of all bilateral delivery relationships.

In Europe, modularity and deliverability scores agree, with bidding zones and countries performing similarly and much better than interconnects (Figure 3a). The differences between bidding zone and country scores are smaller in magnitude than the variability of scores across regions or model time (Figure S12). The similarity between expected deliverability and regional connectivity suggests that both reflect the underlying bilateral delivery of generators and loads in the network.

In the Eastern Interconnect model, the boundary set options score similarly overall, but there is more deviation between the two boundary level metrics (Figure 3b). The largest discrepancy is in NERC regions, which score the best out of the boundary options on regional connectivity and the worst out of the options on the deliverability metric. This may be because the relatively large size of NERC regions results in lower average deliverability, even though the physical delivery relationships within them are still strong relative to the grid as a whole, resulting in a higher regional connectivity score. States, by contrast, score well on both deliverability and modularity. To better understand these differences between modularity and deliverability, future work could break down the modularity score to identify which specific regions have deviating modularity and expected deliverability.

Summarized to the boundary set level, physical delivery metrics can obfuscate significant deviations in performance between regions. BAs in the US are an extreme example of this (Figure S8) In Florida (the Southeastern U.S.), BAs can be as small as a single town. These



(a) European grid (b) Eastern Interconnect

Figure 3: Boundary deliverability and modularity scores in the European and US grids.

small BAs are highly deliverable, with deliverability scores of 75-100%. Elsewhere, large BAs like PJM have much lower expected deliverability (18.12 % for PJM). The low load-weighted average BA deliverability score, 32.3 %, reflects the relative prominence of large, low deliverability ISOs, but obfuscates the fact that some smaller BAs actually have very high deliverability.

Discussion and conclusions

We extend existing power flow tracing algorithms[17,19] to trace bilateral delivery between generators and loads in power networks. Building on this, we define three new metrics for measuring physical delivery and deliverability: delivery distance, expected deliverability, and regional connectivity.

Using these metrics, we explore how physical delivery varies across models of the U.S. Eastern Interconnect and the European grid. We find that most electricity is delivered close to where it is generated, with the median delivery distance across both grids 82.3 km. Physical delivery patterns can be partially explained by familiar characteristics of power grids, with power traveling towards large load centers.

We evaluate physical delivery and physical deliverability at the level of regions and boundary sets, since these are relevant categories for policy and procurement applications. Regional and boundary-set level physical delivery metrics can sometimes obscure within- or between-region variation in physical delivery.

Despite this challenge, we find that countries and bidding zones perform better on physical delivery metrics than synchronous grids in Europe. This may be because the smaller sizes of countries and bidding zones relative to synchronous grids is in better agreement with the short delivery distances we see throughout both network models. In the US, the three options (BAs, NERC regions, and states) perform similarly, with states slightly outperforming the other options using the median of our two metrics. This may be because states are the best compromise between NERC regions (which are relatively large) and BAs (which have highly variable physical deliverability between regions).

Future work: Aligning procurement boundaries with physical delivery

We focus on existing boundaries because they can feasibly be incorporated into procurement requirements. However, our metrics could also be used to evaluate changes to boundaries. We applied our metrics to potential bidding zone changes in Germany and Denmark, and showed that splitting Germany into two zones would improve its deliverability, while combining Denmark into one zone would not substantially impact its deliverability (See supplement section 2.4).

While this work considered only European and U.S. electricity grids, open source models of grids around the world are available, with quickly expanding coverage and quality.[25] Expanding the scope of physical delivery-based boundary evaluation to these grids will be essential for ensuring that findings are broadly relevant.

Future work: evaluating the impact of procurement decisions

We frame physical delivery as a tool for procurement decision making and policy design. A key question in policy design, where the intention is to shape the decisions of many actors, is what the cumulative effect of those decisions will be.

Capacity expansion modeling has shown that 24/7 matching paired with location matching using a 'copperplate' deliverability model (location matching in either the same market region[11] or balancing authority[10]) improves system decarbonization compared to weaker matching requirements. Other work has found that a congestion-based deliverability requirement is necessary for achieving avoided emissions when procuring clean energy for hydrogen production Ricks et al.[14]. Capacity expansion modeling could evaluate the decarbonization impact of procurement using physical delivery requirements based on the metrics here compared to other deliverability definitions.

Conclusions

Physical delivery and deliverability can be a data-rich foundation for electricity policy and procurement. As electricity consumers work to ensure that they are consuming clean electricity, and policy designers work to create guidelines that will be widely adopted and help further the clean energy transition, an understanding of the underlying grid physics will be vital for informed and effective decision making.

Acknowledgement This work was funded by Google. The authors thank Hallie Cramer for extensive discussions on physical delivery and procurement. The authors thank Savannah Goodman, Varun Mehra, and Ian Schneider for feedback on a draft of this paper. The authors thank Xin Chen for discussions on flow tracing algorithm implementations, Ciril Wakounig and Julien Lavalley for feedback on initial metric proposals, Yixing Xu for help with the U.S. model, and Fabian Neumann for help with the European model.

Supporting Information Available

The supplement contains supplemental methods (sections 1.1-1.4) and discussion (sections 2.1-2.4), supplemental figures S1-S22, and supplemental tables S1-S3.

Data is available at doi.org/10.5281/zenodo.8147413.

References

(1) Bhatia, P.; Ranganathan, J., et al. The Greenhouse Gas Protocol. **2004**,

(2) Heeter, J. S.; OShaughnessy, E. J. *Status and trends in the US voluntary green power market (2021 data)*; 2021; https://www.nrel.gov/docs/fy23osti/84419.pdf, accessed 6/26/2023.

(3) O'Shaughnessy, E.; Heeter, J.; Shah, C.; Koebrich, S. Corporate acceleration of the renewable energy transition and implications for electric grids. *Renewable and Sustainable Energy Reviews* **2021**, *146*, 111160.

(4) Gouldson, A.; Sullivan, R. Long-term corporate climate change targets: What could they deliver? *Environmental science & policy* **2013**, *27*, 1–10.

(5) Bjørn, A.; Lloyd, S. M.; Brander, M.; Matthews, H. D. Renewable energy certificates threaten the integrity of corporate science-based targets. *Nature Climate Change* **2022**, *12*, 539–546.

(6) City of Boston, BERDO Ordinance Amending City Of Boston Code, Ordinances, Chapter Vii, Sections 7-2.1 And 7-2.2, Building Energy Reporting And Disclosure. https://www.boston.gov/sites/default/files/file/2022/12/Final\%20Amended\%20Docket\%200775\%20BERDO\%202_0.pdf, Accessed: 2023-06-23.

(7) City of New York, Local Law 97 Chapter 3, Article 301 Maintenance Of Buildings. https://www.nyc.gov/assets/buildings/apps/pdf_viewer/viewer.html?file=2014CC_AC_Chapter3_Maintenance_of_Buildings.pdf&section=conscode_2014, Accessed: 2023-06-23.

(8) Von Meier, A. *Electric power systems: a conceptual introduction*; John Wiley & Sons, 2006.

(9) Miller, G. Beyond 100% renewable: Policy and practical pathways to 24/7 renewable energy procurement. *The Electricity Journal* **2020**, *33*, 106695.

(10) Xu, Q.; Manocha, A.; Patankar, N.; Jenkins, J. System-level impacts of 24/7 carbon-free electricity procurement. *Available at SSRN 4248431* **2021**,

(11) Riepin, I.; Brown, T. System-level impacts of 24/7 carbon-free electricity procurement in Europe. **2022**,

(12) Blumsack, S.; Hines, P.; Patel, M.; Barrows, C.; Sanchez, E. C. Defining power network zones from measures of electrical distance. 2009 IEEE Power & Energy Society General Meeting. Calgary, Canada, 2009; pp 1–8.

(13) Stoft, S. Transmission pricing zones: simple or complex? *The Electricity Journal* **1997**, *10*, 24–31.

(14) Ricks, W.; Xu, Q.; Jenkins, J. D. Minimizing emissions from grid-based hydrogen production in the United States. *Environmental Research Letters* **2023**, *18*, 014025.

(15) Lesieutre, B. C.; Eto, J. H. Electricity transmission congestion costs: A review of recent reports. **2003**, Lawrence Berkeley National Laboratory.

(16) Kirschen, D.; Allan, R.; Strbac, G. Contributions of individual generators to loads and flows. *IEEE Transactions on power systems* **1997**, *12*, 52–60.

(17) Achayuthakan, C.; Dent, C. J.; Bialek, J. W.; Ongsakul, W. Electricity Tracing in Systems With and Without Circulating Flows: Physical Insights and Mathematical Proofs. *IEEE Transactions on Power Systems* **2010**, *25*, 1078–1087.

(18) Bialek, J. Tracing the flow of electricity. *IEE Proceedings - Generation, Transmission and Distribution* **1996**, *143*, 313.

(19) Kang, C.; Zhou, T.; Chen, Q.; Wang, J.; Sun, Y.; Xia, Q.; Yan, H. Carbon Emission Flow From Generation to Demand: A Network-Based Model. *IEEE Transactions on Smart Grid* **2015**, *6*, 2386–2394.

(20) Cuffe, P.; Keane, A. Visualizing the electrical structure of power systems. *IEEE Systems Journal* **2015**, *11*, 1810–1821.

(21) Hagberg, A.; Conway, D. Networkx: Network analysis with python. *URL: https://networkx. github. io* **2020**,

(22) Hörsch, J.; Hofmann, F.; Schlachtberger, D.; Brown, T. PyPSA-Eur: An Open Optimisation Model of the European Transmission System. *Energy Strategy Reviews* **2018**, *22*, 207–215, arXiv:1806.01613 [physics].

(23) Wu, D.; Zheng, X.; Xu, Y.; Olsen, D.; Xia, B.; Singh, C.; Xie, L. An open-source extendable model and corrective measure assessment of the 2021 texas power outage. *Advances in Applied Energy* **2021**, *4*, 100056.

(24) Miller, G. J.; Pease, G.; Shi, W.; Jenn, A. Evaluating the hourly emissions intensity of the US electricity system. *Environmental Research Letters* **2023**, *18*, 044020.

(25) Parzen, M.; Abdel-Khalek, H.; Fedotova, E.; Mahmood, M.; Frysztacki, M. M.; Hampp, J.; Franken, L.; Schumm, L.; Neumann, F.; Poli, D., et al. PyPSA-Earth. A new global open energy system optimization model demonstrated in Africa. *Applied Energy* **2023**, *341*, 121096.

(26) Bialek, J.; Kattuman, P. Proportional sharing assumption in tracing methodology. *IEE Proceedings - Generation, Transmission and Distribution* **2004**, *151*, 526.

(27) Brouhard, T.; Hennebel, M.; Petit, M.; Gisbert, C. A clustering approach to the definition of robust, operational and market efficient delineations for European bidding zones. *IET Generation, Transmission & Distribution* **2023**,

(28) Gotzens, F.; Heinrichs, H.; Hörsch, J.; Hofmann, F. Performing energy modelling exercises in a transparent way-The issue of data quality in power plant databases. *Energy Strategy Reviews* **2019**, *23*, 1–12.