Evaluation of Data Quality and Imputation Methods for the U.S. Energy Information Administration's *Monthly Liquefied Natural Gas Storage Report*

Makayla S. Cowles[1], Preston A. McDowney[1], Hongbin Weng[1]
Pushpal K. Mukhopadhyay [1]

[1]U.S. Energy Information Administration, Washington, DC 20585

Abstract
In January 2022, the U.S. Energy Information Administration (EIA) launched a new census survey to collect natural gas inventory storage data from all operating non-marine liquefied natural gas (LNG) storage facilities in the U.S. The EIA-191L, *Monthly Liquefied Natural Gas Storage Report*, collects data on injections, withdrawals, total LNG in storage, total LNG storage capacity, and maximum deliverability for operators of LNG facilities across 31 states. EIA uses these data to publish state-level monthly estimates on LNG storage in EIA's *Natural Gas Monthly*. The data are also used in several other EIA publications such as the *Natural Gas Annual*, *Monthly Energy Review*, and *Short-Term Energy Outlook*.

To account for unit nonresponse in the 2022 survey, we developed a donor-based imputation method. It creates imputation cells using the monthly activities for the LNG facilities and selects donors based on the donors' expected total LNG in storage and the recipient's reported total LNG in storage. In this paper, we discuss data quality metrics and statistical methodologies used in the EIA-191L, emphasizing statistical editing and imputation methods.

Key Words: Energy statistics, clustering, hot-deck imputation

1. Background

In 1979, the U.S. Energy Information Administration (EIA) assumed responsibility for the collection, processing, and publication of natural gas storage data in Form EIA-191, *Monthly Underground Natural Gas Storage Report* (U.S. Energy Information Administration 1993, 102). EIA-191 is mandatory under the Federal Energy Administration Act of 1974 (U.S. Congress, House 1974, 107); it collects natural gas storage data from all underground natural gas storage facilities in the United States. Form EIA-191L, *Monthly Liquefied Natural Gas Storage Report*, collects aboveground natural gas inventory storage data from approximately 100 non-marine liquefied natural gas (LNG) storage facilities. Volumes of injections, withdrawals, total LNG in storage, total LNG storage capacity, and maximum deliverability are reported by operators of LNG facilities across 31 states (U.S. Energy Information Administration, n.d.). EIA-191L is a new mandatory survey under Federal Energy Administration Act of 1974 (U.S. Congress, House 1974, 107), related to EIA-191, that started collecting data January 2022. Storage data information is collected at the facility level to allow comparisons of the utilization of the individual natural gas storage facilities. The monthly LNG storage data collected in EIA-191L, combined with data collected from EIA-191, "provides reliable baseline data on storage operations necessary for analyses, modeling, and comparison with normal industry operations in case of severe weather,

[1] The analysis and conclusions expressed in this paper are those of the authors and do not represent the official position of the U.S. Energy Information Administration (EIA) or the U. S. Department of Energy (DOE).

natural disaster, or other extreme circumstances" (U.S. Energy Information Administration 2024, 10).

The U.S. Department of Transportation (DOT) conducts an annual survey of aboveground LNG facilities collecting information on location, function, liquefaction rates, and capacity for each storage facility. The DOT dataset is released annually, but it offers no insight into seasonal movements of the LNG facilities. The purpose of EIA-191L is to collect data that are used by EIA to track seasonal movements of LNG into and out of aboveground storage (U.S. Energy Information Administration 2024, 17). Our work with the EIA-191L data quality and imputation methods helps provide reliable data to understand seasonality of LNG storage movements. This allows energy producers, utilities, and policymakers to anticipate fluctuations in LNG demand. The total LNG in aboveground storage facilities, total storage capacity, and storage activities, including injections and withdrawals, will be published at the state-level and will appear in the EIA publications *Monthly Energy Review*, *Natural Gas Annual*, and *Natural Gas Monthly*.

Collecting monthly information on EIA-191L not only allows EIA to track seasonal movements of aboveground LNG facilities but will also allow EIA to augment the information collected on EIA-191 for a more complete accounting of the LNG market. The information collected on EIA-191L is similar to information collected from underground storage facilities on EIA-191. Unlike the underground storage facilities where the total capacity is calculated based on the physical characteristics of the underground cavern, the total capacity of an aboveground storage facility is based on the fixed capacity storage of a manufactured storage tank. The volume of injections, withdrawals, total LNG in storage, and total storage capacity are reported in thousand cubic feet per month (Mcf/month). The net injections and withdrawals should correspond to the difference between total LNG volume for the current and prior report period. The total storage capacity is based on a facility's design specifications, installed compression equipment, and operating procedures particular to a site. The maximum deliverability of a facility is the maximum Mcf per day (Mcf/day) LNG can be withdrawn or delivered from a storage facility when filled to maximum capacity (U.S. Energy Information Administration, n.d.). Although none of these measures for a given storage facility are fixed, it is not often that the total storage capacity and maximum deliverability change.

EIA-191 was used as a template to design the EIA-191L survey form. Similarly, the initial edits and imputation methods were based on methods used by EIA for the EIA-191, as well. We evaluated the current EIA-191 edits and imputation procedures for adaptation for the new EIA-191L survey form. We used processed, aggregated, and validated monthly data from January 2022 to December 2023 collected by the EIA-191L survey, current edits programmed in the Standard Energy Processing System (STEPS), and third-party sources of aboveground storage for the evaluation. The analysis consisted of two main objectives. First, we evaluated the current edits being implemented in STEPS and recommended modifications or additional edits for EIA-191L. The second objective was to recommend imputation procedures for edit failures and missing data.

2. Statistical Methodologies

The statistical methodologies used in the EIA-191L, *Monthly Liquefied Natural Gas Storage Report* closely follow those used in the EIA-191, *Monthly Underground Natural Gas Storage Report*. We evaluated the statistical edits and implemented the imputation methods for each edit by analyzing processed, aggregated, and validated EIA-191L data. In addition to the validated EIA-191L data, we used the following relationships to evaluate the statistical edits and imputation methods for month T, where the range of T is 1 (Jan 2022) to 24 (Dec 2023):

- Expected Total LNG in Storage

$$expected\ Total\ LNG\ in\ Storage_T = reported\ Total\ LNG\ in\ Storage_{T-1} + Injections_T - Withdrawals_T$$

- Percentage Difference between reported Total LNG in Storage and expected Total LNG in Storage

$$percentage\ difference\ between\ reported\ and\ expected\ Total\ LNG\ in\ Storage_T = \frac{|expected\ Total\ LNG\ in\ Storage_T - reported\ Total\ LNG\ in\ Storage_T|}{reported\ Total\ LNG\ in\ Storage_T} \times 100$$

It is important to note that, for the first reported month (January 2022), no facility has a value for expected Total LNG in Storage because there are no previous reported gas levels that can be used to calculate expected Total LNG in Storage.

2.1 Statistical Edits

The data collection stage is a potential source that may introduce errors in a survey because a respondent may provide incorrect information either intentionally or unknowingly. Some incorrect information can be identified and corrected by using logical edits that use mathematical relationships to detect inconsistencies. In addition to logical edits, almost all surveys require establishing statistical edit rules that identify a set of suspicious observations which must be verified by the respondents or subject-matter specialists. This verification process requires agency resources and adds burdens on respondents. Thus, it is important to establish a set of statistical edit rules that balance between data quality and timeliness. For more information see "Statistical Data Editing" by Waal (2009).

LNG storage activities are similar to underground natural gas storage activities. As a result, we implement eight editing rules for EIA-191L based on editing rules for EIA-191. Specifically, these edit rules flag data when:

1. Total LNG Storage Capacity is zero
2. Total LNG Storage Capacity changes
3. Total LNG in Storage is greater than Total LNG Storage Capacity by 5% or more
4. Expected Total LNG in Storage has a percentage difference of 5% or more from reported Total LNG in Storage
5. Total LNG in Storage is zero
6. Maximum Deliverability changes
7. Injections are greater than monthly Maximum Deliverability, which is defined by the number of days in the month × Maximum Deliverability
8. Withdrawals are greater than monthly Maximum Deliverability, which is defined by the number of days in the month × Maximum Deliverability

To assess data quality, we evaluated the edit rules by using 24 months of reported EIA-191L data and following up with the respondents. We developed programs to reedit the data based on the pre-established eight edit rules. We found the initial reported data contained numerous records that violated the edit rules. Many of those records were corrected or verified by the respondents throughout the analysis. We also recommended modifications to the edit rules and evaluated their effects on the number of false positives (edit failures that were determined to be valid data) and the number of records requiring imputation for edit failure. From our analysis, we determined, with minor modifications, that the initial eight edit rules used for EIA-191 caught all suspected outliers for EIA-191L. Therefore, no additional edits were recommended for EIA-191L.

We determined that the following edit rules work well for EIA-191L without requiring modifications: Total LNG Storage Capacity is zero (Edit Rule #1), changes in Total LNG Storage Capacity (Edit Rule #2), changes in Maximum Deliverability (Edit Rule #6), Injections are greater than monthly Maximum Deliverability (Edit Rule #7), and Withdrawals are greater than monthly Maximum Deliverability (Edit Rule #8).

The following three edit rules needed minor modifications to prevent false edit flags. The modifications are as follows:

- Total LNG in Storage is greater than Total LNG Storage Capacity:

We recommend increasing the threshold, based on the percentage that Total LNG in Storage can be greater than Total LNG Storage Capacity, from 5% to 10%. Increasing this threshold removes all false positive flags based on this edit.

- Difference between the expected Total LNG in Storage and reported Total LNG in Storage:
  It is possible to have imbalances between the expected Total LNG in Storage and reported Total LNG in Storage resulting from losses from liquefaction, transportation, and transfer of LNG. As a result, we recommend increasing the threshold, based on the percentage difference in expected Total LNG in Storage and reported Total LNG in Storage, from 5% to 10%.

- Total LNG in Storage is zero:
  Active underground storage facilities must preserve a volume of base gas that is permanently kept in a storage reservoir to maintain pressure and deliverability rates during withdrawal. Therefore, the Total LNG in Storage in active underground facilities can never fall to zero. However, some aboveground LNG storage facilities operate seasonally. These seasonal facilities can either shut down operations all together or reduce their activities during off-peak periods. During these periods, storage facilities may empty their storage tanks. In addition to seasonal operations, storage facilities may undergo maintenance and need to empty their storage tanks. We recommend replacing the original Edit Rule #5 with the following edit: expected Total LNG in Storage is greater than 10% or more of the Total LNG Storage Capacity when Total LNG in Storage is zero and expected Total LNG in Storage is nonzero. Modifying this edit, helps limit the number of false positives that would result from the expected imbalance between expected Total LNG in Storage and reported Total LNG in Storage.

Our assumption was that the relationships found for the underground reservoirs will also hold for the aboveground LNG facilities, but we learned that allowing for a larger imbalance reduces the number of responses that failed edits resulting in less burden to reach back out to respondents. The decrease in the number of responses that fail the edit checks using our recommended changes from the pre-established thresholds from EIA-191 are given in Table 1.

Table 1. Summary of recommended data edit rule changes

Edit Rule	Pre-established threshold	Data failures	Recommended changes	Data failures
Total LNG in Storage is greater than Total LNG Storage Capacity (Edit Rule #3)	5%	23[1]	10%	0
Difference in expected Total LNG in Storage vs. reported Total LNG in Storage (Edit Rule #4)	5%	114	10%	87
Total LNG in Storage is zero (Edit Rule #5)	0%	68	expected Total LNG in Storage is greater than the Total LNG Storage Capacity by 10%, when Total LNG in Storage is zero and expected Total is nonzero	6

[1] Confirmed to be false positives

2.2 Item Imputation

Item imputation methods are used when information about some items is available for a respondent but information about other items is missing or the information failed edits. Although there are no item nonresponses in EIA-191L, item imputation methods are used for imputing missing values due to edit failures for Total LNG Storage Capacity, Maximum Deliverability, Injections, and Withdrawals. Total LNG in Storage is not imputed for because imputing could cause a cascading effect to the following months, possibly causing more responses to fail edits than beforehand. The relationship between reported and expected Total LNG in Storage is primarily used for item imputation. Because a response can fail multiple edit rules for the same reason e.g., a zero Total LNG Storage Capacity can lead to a response failing a total of three edit rules, it is important to make sure the correct imputation method is being used. Therefore, we also recommended a hierarchy to impute for responses that failed edit checks. The following are the recommended imputation procedures for the failed edit check responses and the order in which they should be imputed:

1. Total LNG Storage Capacity is zero (Edit Rule #1)
   For imputing Total LNG Storage Capacity when respondents report a value of zero, we recommended using the "last observation carried forward" method, since the total storage capacity rarely changes over time. For month T, replace
   $$\text{Total LNG Storage Capacity}_T$$
   with
   $$\text{Total LNG Storage Capacity}_{T-k}$$
   where $T-k$ is the most recent month such that
   $$\text{Total LNG Storage Capacity}_{T-k} > 0$$

2. Total LNG Storage Capacity changes (Edit Rule #2)
   It is possible for a facility to add or remove storage tanks resulting in a change in total storage capacity. If a facility's total storage changes, it is important to reach out to the respondent first to confirm the change. However, if no response is received, then we recommended two imputation methods for Total LNG Storage Capacity when total storage capacity changes. The first method is to replace
   $$\text{reported Total LNG Storage Capacity}_T$$
   with the
   $$\text{mode(reported and imputed Total LNG Storage Capacity)}$$
   if
   $$\frac{|\text{mode(reported and imputed Total LNG Storage Capacity)} - \max(\text{reported Total LNG in Storage})|}{\text{mode(reported and imputed Total LNG Storage Capacity)}} \leq .10$$

   The second method is using the "last observation carried forward" method (explained above). The first method ensures that the mode will pass Modified Edit Rule #3 for the reported month and should be used for fixing data reported by facilities who have already been established in the survey (reported data for 3 or more months). The second method is recommended for newly reporting facilities or when there is a change in ownership.

3. Total LNG in Storage is greater than Total LNG Storage Capacity by 10% or more (Modified Edit Rule #3)
   We recommended following up with the respondents if the reported Total Gas exceeds the Total Capacity by 10% or more. At the time of completing this analysis, no responses failed this edit that required imputation.

4. Expected Total LNG in Storage has a percentage difference of 10% or more from reported Total LNG in Storage (Modified Edit Rule #4)
   We recommended imputing for Injections or Withdrawals when the expected Total LNG in Storage is different from the reported Total LNG in Storage by 10% or more. If reported Total LNG in Storage is imputed for, then it would change the

expected Total LNG in Storage for the following month and could cause additional edit failures. With the assistance of the energy experts, it was determined that Injections should be the primary variable to impute for since the volume of withdrawals is physically measured, and the volume of injections is estimated. If the imputation results in a negative injection value, then we recommended imputing for Withdrawals. We recommended the following methods to ensure that the imputations will result in only positive values (for month T):
If

$$\text{reported Total LNG in Storage}_T + \text{Withdrawals}_T > \text{reported Total LNG in Storage}_{T-1},$$

then impute Injections by

$$\text{Injections}_T = \text{reported Total LNG in Storage}_T - \text{reported total LNG in Storage}_{T-1} + \text{Withdrawals}_T,$$

otherwise impute Withdrawals by

$$\text{Withdrawals}_T = \text{reported Total LNG in Storage}_{T-1} + \text{Injections}_T - \text{reported Total LNG in Storage}_T$$

5. Expected Total LNG in Storage is greater than 10% or more of the Total LNG Storage Capacity and reported Total LNG in Storage is zero (Modified Edit Rule #5)
   Similar to Modified Edit Rule #3, this edit rule checks the relationship between Total LNG in Storage and Total LNG Storage Capacity. The notable difference is Modified Edit Rule #3 checks if reported Total LNG in Storage is greater than Total LNG Storage Capacity, whereas Modified Edit Rule #5 checks if expected Total LNG in Storage is greater than a percentage of the Total LNG Storage Capacity. Hence, we recommended imputing either Injections or Withdrawals using the same imputation method described above when the expected Total LNG in Storage is different than the reported Total LNG in Storage by 10% or more.

6. Total Maximum Deliverability changes (Edit Rule #6)
   Much like Total LNG Storage Capacity, it is possible for a facility's maximum deliverability to change. If a facility's maximum deliverability changes, it is important to reach out to the respondent first to confirm the change. However, if no response is received, then we recommended imputing for Maximum Deliverability when maximum deliverability changes by the "last observation carried forward" method.

7. Injections/Withdrawals are greater than monthly Maximum Deliverability, defined as the number of days in the month × Maximum Deliverability (Edit Rule #7 & Edit Rule #8)
   Injections and withdrawals are reported at a monthly level, but the maximum deliverability is reported at a daily level. Thus, the total injections or withdrawals that is reported for a month should not exceed the facilities' maximum deliverability times the number of days in the month. If this edit fails, it is best to follow up with the respondent. If no response is received, then we recommended replacing

$$\text{Maximum Deliverability}_T$$

   with

$$\max\left(\frac{\text{reported monthly Injections/Withdrawals}}{\text{number of days in the month}}\right)$$

   over the last 12 months, where we increase the left most significant digit (the left most nonzero digit) from the previous maximum by one.

2.3 Unit Imputation

A total of 17 facilities across 11 states were added to the EIA-191L frame in 2023. These newly added facilities were operational in 2022, but only have reported data for 2023. We developed an imputation method for imputing the missing 2022 data for these 17 facilities. Because LNG

storage capacity and maximum deliverability do not change often, we imputed missing Total LNG Storage Capacity and Maximum Deliverability for these facilities by using their most recently reported values from 2023. To impute missing values in Total LNG in Storage, Injections, and Withdraws for 2022, we evaluated three different donor-based imputation methods. Donor-based imputation methods such as the hot-deck imputation method use observed values from the same sample as imputed values to avoid generating impossible responses and are widely used for imputing missing values in survey data. Hot-deck imputations are often used within imputation cells such that the observations within the imputation cell closely follow a cell mean model. A random hot-deck procedure gives an imputed data set the correct expectation under the model. For more information, see Fellegi and Holt (1976), Lohr (2022, Section 8.7.3), Fuller (2009, Section 5.2.1), and Särndal and Lundström (2005).

We focused on hot-deck imputation methods because the hot-deck method preserves the relationship between the expected and reported Total LNG in Storage, thus, ensuring that the Total LNG in Storage is consistent with Injections and Withdrawals. We created imputation cells based on all reported units in 2023 so that the facilities within the imputation cells are homogenous. A donor facility is then selected for a recipient facility by comparing their monthly LNG activities. To determine the best imputation method, we compared two nearest neighbor hot-deck methods and one random hot-deck imputation method. For comparing the three methods, a set of observed facilities are randomly removed from the 2022 data and their reported values are treated as the "truth". In particular, the imputation and comparison steps are described below:

Step 1. Create imputation cells by computing k-means clusters using the difference between expected and reported Total LNG in Storage (excluding the month of January), Total LNG in Storage, and Maximum Deliverability for the 2023 data. The optimum number of clusters is determined to be four using the ratio of the between- and within-cluster sum of squares. All facilities from the year 2023 are used for creating the imputation cells. There were no missing values for the three variables used in creating the imputation cells and all facilities were assigned to an imputation cell.

Step 2. Remove the 17 facilities that are missing for 2022 from the comparison dataset.

Step 3. Randomly select 15 facilities that have reported data for both 2022 and 2023 and remove their 2022 reported values from the comparison dataset. These facilities are treated as recipients and their reported values are treated as the "truth" for this comparison.

Step 4. Compute the Euclidean distance for Total LNG Storage Capacity and Maximum Deliverability within each of the four clusters.

Step 5. If there is one facility with a zero Euclidean distance, then use that facility as the donor. If there are no (or multiple) facilities with a zero Euclidean distance, then the closest five facilities are used as potential donors.

Step 6. The three imputation methods use three donor selection methods to select a donor from the five nearest neighbors. Method 1 is a random hot-deck method that selects a donor randomly. Method 2 selects the donor whose expected Total LNG in Storage is the closest to the recipient's reported Total LNG in Storage for January 2023, and Method 3 selects the donor for which the average expected Total LNG in Storage is the closest to the recipient's average reported Total LNG in Storage for January (winter), April (spring), and July (summer) 2023. The use of expected and reported Total LNG in Storage preserves the relationship between them and captures the existing difference between expected and reported Total LNG in Storage in the imputed data.

Step 7. Compute the root mean squared error (RMSE) from the imputed values and the original reported values ("truth"). Then repeat steps 2 to 6 100 times and compute the average RMSE at the state and U.S. levels.

Table 2: Three donor selection methods for unit imputation

Donor selection	*Method 1*	*Method 2*	*Method 3*
Within imputation cells	X	X	X
Links Total LNG in Storage for January		X	X
Links average seasonal Total LNG in Storage			X

Table 2 summarizes the three donor selection methods. Table 3 shows the ratio of the RMSE for Total LNG in Storage, Injections, and Withdrawals at the U.S. and state levels. The RMSE for Method 1 is used in the denominator in computing the RMSE ratios. Therefore, the RMSE ratios for Method 1 are 1.00 for each state and are excluded from Table 3. Method 3 has the lowest RMSE for the U.S. totals. RMSE ratios for some states shown in Table 3 might be high for methods 2 and 3, such as those for State 25. Since facilities are randomly selected to be recipients at the imputation and comparison Step 3, it is possible that the optimal donor for a chosen facility was also randomly selected to be a recipient, as well, resulting in a random selection of donors (Method 1) performing better.

Based on the RMSE ratios for the United States, Method 2 and Method 3 both perform better than Method 1 for imputing Total LNG in Storage with Method 3 generally performing the best overall. In addition, Method 2 and Method 3 do not add any variability due the selection of donors. However, Method 2 and Method 3 require information from 2023 to impute 2022 data. It is worth noting that Method 1 also performs reasonably well at least for the U.S. total. Because Method 1 does not rely on future observations, it can be used for unit imputation when data for the future months are not available.

Table 3: Ratio of the RMSEs for Methods 2 and 3 compared to those for Method 1

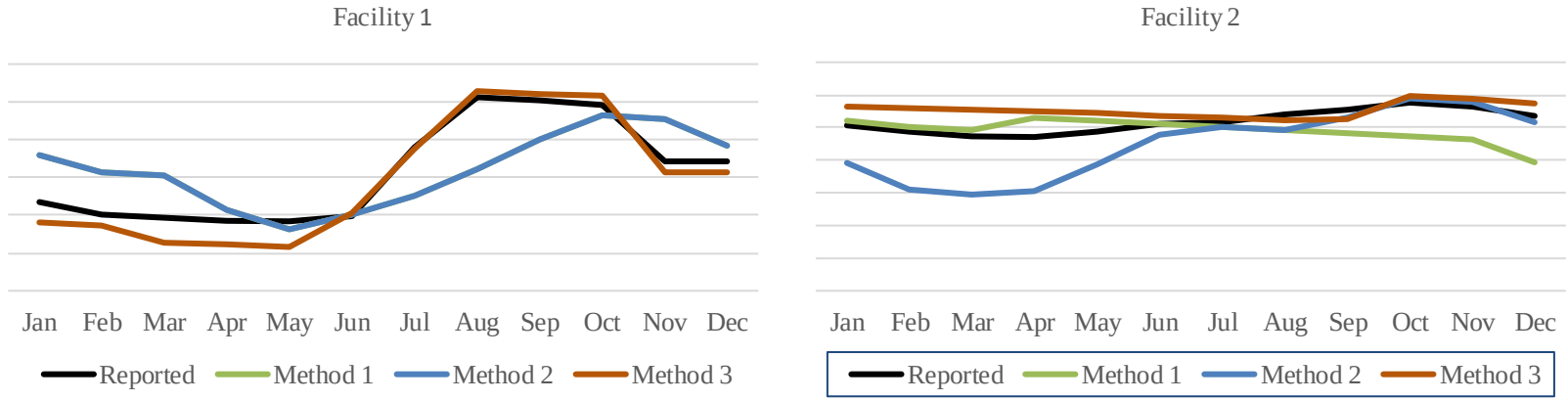
States[1]	*Method 2*			*Method 3*		
	Total LNG in Storage	*Injections*	*Withdrawals*	*Total LNG in Storage*	*Injections*	*Withdrawals*
State 1	0.84	1.06	1.21	0.84	1.09	1.24
State 2	0.88	1.03	0.85	0.90	0.98	0.93
State 3	1.33	1.03	0.95	0.92	1.03	0.95
State 4	0.44	0.55	0.51	0.49	0.54	0.50
State 5	0.97	1.14	0.94	0.84	0.92	0.79
State 6	1.30	1.25	1.02	0.59	1.75	0.93
State 7	0.91	0.97	0.84	0.94	0.55	0.56
State 8	0.74	0.85	0.87	0.75	0.86	0.93
State 9	0.77	1.08	1.06	0.72	0.98	0.92
State 10	0.72	1.35	1.12	0.68	1.26	1.04
State 11	1.03	0.88	0.74	0.71	0.92	0.74
State 12	0.84	0.72	0.88	0.79	0.72	0.89
State 13	0.91	1.03	0.89	0.85	1.15	0.94
State 14	0.84	1.67	0.84	0.68	0.88	0.49
State 15	1.05	1.30	1.09	0.84	1.29	1.09

States[1]	*Method 2*			*Method 3*		
	Total LNG in Storage	*Injections*	*Withdrawals*	*Total LNG in Storage*	*Injections*	*Withdrawals*
State 16	0.92	1.61	1.55	0.98	1.63	1.79
State 17	1.46	1.38	1.30	1.38	1.46	1.19
State 18	1.17	0.97	0.97	1.17	1.33	1.22
State 19	0.88	1.09	1.44	0.84	1.13	1.63
State 20	0.79	1.15	0.98	0.67	0.80	0.85
State 21	0.58	0.89	0.67	0.90	0.80	0.99
State 22	0.88	0.92	1.30	0.76	0.97	0.91
State 23	1.04	1.02	1.01	1.06	1.01	1.07
State 24	0.92	0.92	0.91	0.91	1.09	1.10
State 25	1.30	0.91	1.87	1.19	1.03	2.47
State 26	0.82	0.54	1.04	0.77	0.52	1.02
United States	0.93	1.05	1.01	0.89	0.99	1.00

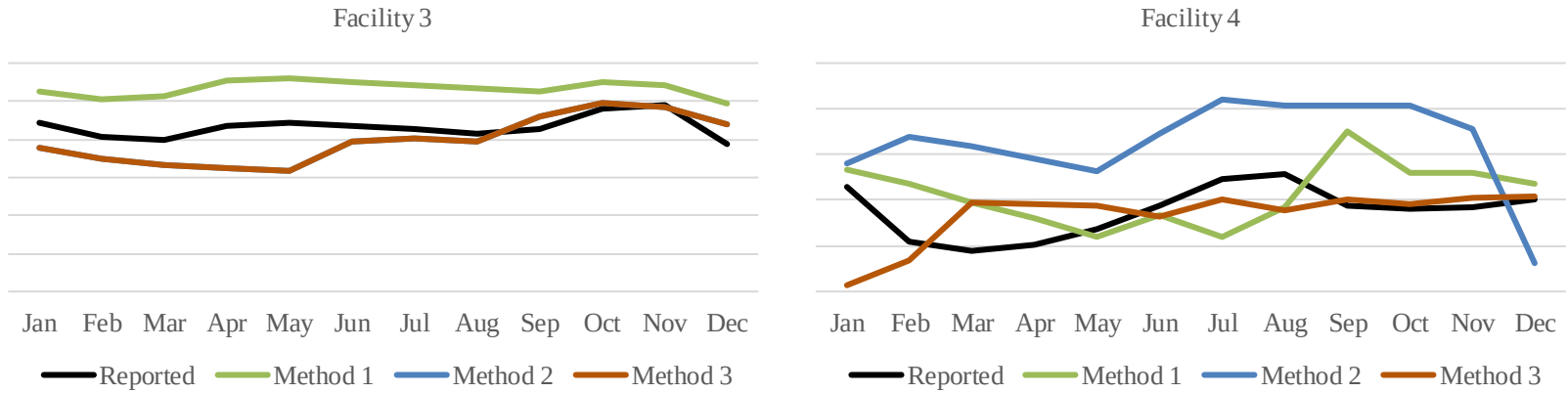
[1] EIA has not yet published the data from EIA-191L. Hence, we cannot disclose specific values or states found in the reported data.

Figure 1, below, shows the reported vs. imputed Total LNG in Storage values for four chosen facilities. Overall, the imputed values from Method 3 for Total LNG in Storage follows closely with the reported values.

Figure 1. Reported vs. imputed Total LNG in Storage for four different facilities



Method 1 and Method 2 chose the same donor for this facility.



Method 2 and Method 3 choose the same donor for this facility

3. Summary

For any new survey it is important to evaluate the quality of the data being reported and develop edit rules, as well as imputation methods for item and unit nonresponse. Since the EIA-191L, *Monthly Liquefied Natural Gas Storage Report*, is closely related to the EIA-191, *Monthly Underground Natural Gas Storage Report*, we assumed similar relationships when evaluating the EIA-191L reported data. The same data edit rules from the EIA-191 were applied to the EIA-191L, and our analyses suggested that modifications were needed for three of the edit rules. These recommended modifications reduced the false positives and the number of records requiring imputation for edit failures. In addition to evaluating the data quality, we developed item imputation methods for responses failing the data edits, and unit imputation methods for newly added facilities with reported 2023 data but missing 2022 data. We evaluated three donor-based imputation methods and found that the method that links expected and reported average seasonal Total LNG in Storage performs the best.

For future research, we would like to explore imputation methods that also consider Injections and Withdrawals. Total LNG in Storage, Injections, and Withdrawals might each have different trends for a facility, and incorporating all three trends could possibly improve the overall results. Additionally, when the Euclidean distance of Total LNG Storage Capacity and Maximum Deliverability is being used to define nearest neighbors, we would like to use a few other definitions for nearest neighbors including geographic regions. Furthermore, if there is change in LNG storage operations or the EIA-191L survey form, then we would need to reevaluate our edit and imputation methods.

Natural gas storage data are a critical link in understanding the overall natural gas system operations. The data collected in EIA-191L will be used to publish state-level monthly estimates on LNG storage in EIA's *Natural Gas Monthly*, as well as EIA's *Monthly Energy Review*, and several EIA dashboard products: *New England Dashboard*, the *Southern California Daily Energy Report*, and the *Natural Gas Storage Dashboard*. The data will also be useful for the forecasts contained in the EIA *Short-Term Energy Outlook* (U.S. Energy Information Administration 2024, 10).

Acknowledgements

The authors gratefully acknowledge the contributions of EIA colleagues Kenneth Pick and David Kinyon to this research.

References

Fellegi, I.P., and D. Holt. 1976. "A Systematic Approach to Automatic Edit and Imputation." *Journal of the American Statistical Association* 71 (353): 17-35. JSTOR.

Fuller, W. A. 2009. *Sampling Statistics*. 2nd ed. Wiley Series in Survey Methodology. John Wiley & Sons, Inc.

Lohr, S.L. 2022. *Sampling: Design and Analysis*. 3rd ed. Texts in Statistical Science. Chapman & Hall/CRC.

Särndal, C.-E., and S. Lundström. 2005. *Estimation in Surveys with Nonresponse*. Wiley Series in Survey Methodology. John Wiley & Sons, Inc.

U.S. Energy Information Administration. n.d. *Monthly Liquefied Natural Gas Storage Report, Form EIA-191L*. Accessed March 29, 2023. https://www.eia.gov/survey/form/eia_191l/form.pdf.

U.S. Energy Information Administration, Office of Oil and Gas. 1993. "Natural Gas Monthly". Office of Scientific & Technical Information Technical Reports. UNT Digital Library. June 22. https://digital.library.unt.edu/ark:/67531/metadc1342384/m1/4/.

U.S. Energy Information Administration. 2024. "Supporting Statement for EIA's Natural Gas Data Collection Package, Part A: Justifications". Office of Management and Budget. March. https://www.reginfo.gov/public/do/PRAViewDocument?ref_nbr=202404-1905-001.

U.S. Congress, House. 1974. *Federal Energy Administration Act of 1974*, H.R. 11793, 93rd Congress. Introduced in House December 5, 1973. https://www.congress.gov/93/statute/STATUTE-88/STATUTE-88-Pg96.pdf.

Waal, T.D. 2009. "Statistical Data Editing." In *Sample Surveys: Design, Methods, and Applications*, edited by D. Pfeffermann and C.R. Rao, 187-214. Vol 29A of Handbook of Statistics. North-Holland