Scaling operational peak powerflow cases to create historical powerflow cases

J. SCHWARTZ*, R. BOULTON
AltaLink Management Ltd.
Canada

SUMMARY

Historical powerflow cases (HPCs) are offline powerflow cases where the powerflow values through power system elements (e.g., machines, loads, transmission lines) correspond to a timestamp which occurred in the past. Using HPCs in power system studies over the traditional approach of using seasonal peak powerflow cases has several advantages which are discussed in the literature.

The literature discusses various methods by which to create a HPC. Previous approaches in the literature assign historical powerflow values to machines and loads in an operational peak powerflow case. The case is solved and the resulting powerflows on transmission lines (TLines) and transformers is checked against the historical data. This method works best for a vertically integrated utility, i.e., a single utility having historical data available for all power system elements modeled in the powerflow case. This approach is difficult to use when many utilities operate together in a region (e.g., the province of Alberta in Canada) because not all utilities share their historical data with each other. This is the situation for the authors' utility, i.e., multiple generation owners, transmission owners, distribution owners, and industrial customers.

This paper improves upon the HPC creation methods documented in the literature by relying on the subsystem scaling feature of commercial powerflow software to scale machines and loads. The revised methodology in this paper is robust for equipment that does not have any direct measurements in a single utility's historian.

This paper discusses the scaling methodology for the various scenarios which can occur when multiple owners have facilities operating close together. The basic scenario is when a substation is supplied by TLines where historical data is available for each TLine. The sum of powerflow on these TLines is performed in the solved operational peak powerflow case and compared with the same sum performed on the historical data. The difference of these two sums is the scaling increment required to be applied to loads/machines at this substation only. The case is solved and a validation is included to check the powerflow sum after scaling agrees with the sum obtained from the historical data. This is done for each substation.

Similar scaling is performed for a large area where historical data is available for only a few TLines supplying the area, i.e., an interface. Only the loads/machines in this unknown area are scaled by comparing powerflow on these interface TLines. Another benefit of the revised methodology in this paper is equipment in unknown areas does not get removed or equivalenced as is done by existing methods. This means the resulting HPC has a larger range of applications.

james.schwartz@altalink.ca

Historical data for TLines is the focus of this methodology because the historian at the authors' utility has consistent storage format for TLines: a positive MW/Mvar value indicates power is leaving the substation. Transformers are not consistent because of ownership (e.g., distribution vs industrial customer), winding configuration, and voltage class (transmission vs distribution). This is an important consideration because the HPC creation methodology presented in this paper is automated and focuses on minimizing the amount of manual data entry from the user.

Scaling of every known substation and unknown area in the powerflow case is performed. An overall validation is performed on every TLine comparing the value from the solved powerflow case with the flow in the historian. This paper shows the results of this validation indicate a significant reduction in error from methods in previous publications.

KEYWORDS

1 Introduction

Historical powerflow cases (HPCs) are offline powerflow cases where the powerflow values through power system elements (e.g., machines, loads, transmission lines) and their related voltages correspond to a timestamp which occurred in the past. Using HPCs in power system studies over the traditional approach of using seasonal *peak* powerflow cases has advantages including the following:

- Probabilistic contingency analysis can be performed by considering HPCs at multiple timestamps and including reliability data for the outaged elements.

- Post-disturbance analysis requiring powerflow simulation is based on historical data.

- Post-disturbance analysis requiring transient/dynamic simulation or electromagnetic transient (EMT) simulation can produce more accurate and reliable results because they have a more accurate starting point, i.e., the HPC.

Peak powerflow cases occur one hour out of the given time-period (e.g., season, year) by definition and represent the highest coincident demand from customers. Peak powerflow cases do not necessarily represent worst-case loading through transmission lines (TLines) and transformers; this is only true without further analysis for a two-bus system.

The following are documented methods by which HPCs can be created:

1. A utility can purchase an add-on module to their energy management system (EMS) software suite [1] which can produce an offline powerflow case, i.e., a snapshot, based on the same topology used in the EMS's topology which is likely to be a node-breaker model.

2. Proprietary software [2] may be purchased which connects to a utility's historical measurements database and copies the data into an existing offline powerflow case using a manually specified mapping.

3. A utility may develop in-house software using open-source tools [3] to connect to their historical measurements database and copy the data into an existing operating planning powerflow case using a manually specified mapping.

4. A utility can use an in-house approach with open-source software [4] to automatically create a HPC where manual mapping is only required between PSS®E bus number and substation name.

The authors' utility, their independent system operator (ISO), and the neighboring transmission facility owners (TFOs) use powerflow cases based on a bus-branch topology for planning studies. Therefore even if software cost was not an issue for Method #1 above a significant amount of work would be required to convert the node-breaker topology to the bus-branch topology for studies. Method #1 can create some logistic challenges such as how often snapshots should be created and and how long they should be retained. Method #2 presents a large amount of manual work because the user must specify the mapping between historical data channel names and PSS®E powerflow components in Microsoft Excel. Method #3 follows the approach of Method #2 but uses in-house developed software to copy data from the historical data server (historian) to the powerflow software instead of proprietary software; it manually specifies 3500 mappings between historical data channel names and machine/load components and uses Microsoft Excel to retrieve data from a historian using an Excel add-in provided by the historian software developer.

Manually specifying thousands of historical data channel names to powerflow component mappings inherently creates *more* data and this is a concern because now more data must be maintained following changes to the utility's historian and operations planning powerflow case topology. Validation of the generated HPC is not possible unless TLine and/or transformer mapping is performed; [3] only specifies mapping for machine and load components.

Method #4 reduces the amount of manual mapping required however the HPC validation is negatively affected by equipment (e.g., machines) with missing historical data; this often happens when multiple TFOs and generation facility owners (GFOs) operate in the same province because they do not

always share their historical data with each other. Method #4 does not automatically handle inconsistencies in the format of transformer historical data. At the authors' utility the sign of the stored historical value depends on multiple factors including the following:

- high-voltage (HV) or low-voltage (LV) winding where the current transformer (CT) is located

- transformer location in the network, i.e., transmission-level, distribution-level, generator step-up, industrial customer

This paper improves upon Method #4 as follows:

- only TLine historical data is used to mitigate inconsistencies in transformer historical data

- subsystem scaling is used to mitigate missing historical data (e.g., other TFOs, GFOs)

- no network equivalencing is performed because areas with unknown historical data are scaled at the interfaces

The historian used at the author's utility is the PI System™ [5] by OSISoft therefore this paper uses the following industry terms: *PI server* means historian, *PI tag* means historical data channel name, and *PI data* means historical measured quantities. The only manual mapping required by the approach described in this paper is between PSS®E bus number and the physical substation number used in the PI tag nomenclature; this only happens once, is relatively easy to create, and can be automated to some extent by using existing internal databases. The approach described in this paper uses Python to automate creation of the HPC. All PI tags, PI data, and PSS®E data are represented as Python objects in a customized object-oriented data model designed for the task of creating HPCs. Therefore the benefits of object-oriented programming can be realized which include ease of code maintenance, flexibility for adding new features, and re-usability.

The contribution to knowledge of this paper is the approach by which to create HPCs when historical data for some equipment in the powerflow case is missing (e.g., it is owned by another utility).

2 Methodology

The steps of the approach used in this paper are summarized below and discussed in detail in this section.

1. choose a timestamp
2. select appropriate PSS®E case
3. assign actual substation numbers to PSS®E HV buses
4. download all PI tags
5. assign relevant PI tags to each PSS®E TLine
6. download relevant PI data
7. place TLines out-of-service (OOS) based on their PI data
8. scale generation/load in entire province to adjust powerflow on interconnections
9. scale generation/load in specific areas while adjusting power order on HVDC links
10. scale generation/load at each *known* substation
11. scale generation/load for *unknown* areas using defined interfaces
12. validate solved TLine powerflows in PSS®E with the original PI data

2.1 choose a timestamp

A timestamp (e.g., `2022-08-25 14:40:00`) must be selected for which to create the HPC. This can be of any resolution (e.g., seconds, milliseconds) which is supported by the PI server. This timestamp must have occurred in the past because no data will exist on the PI server for timestamps occurring in the future.

Table 1: seasonal PSS®E case range example

timestamp month	OPBC
2021-Jan → 2021-Apr	2020/2021 winter-peak
2021-May → 2021-Oct	2021 summer-peak
2021-Nov → 2021-Dec	2021/2022 winter-peak

2.2 select appropriate PSS®E case

A solved PSS®E case needs to be selected which represents the topology of the transmission network for the timestamp chosen in Section 2.1. This paper uses peak operations planning base cases (OPBCs) obtained from the ISO of the authors' utility: they provide winter-peak and summer-peak powerflow cases every year. This paper uses winter-peak cases for timestamps between November 1 of a given year and April 30 of the following year. This paper uses summer-peak cases for timestamps between May 1 and October 31 of a given year. Table 1 provides an example of how OPBCs are chosen based on a the given timestamp.

2.3 assign actual substation numbers to PSS®E HV buses

Every bus in the powerflow case is updated with the corresponding substation number as used in the PI tag nomenclature. The substation name is prepended to the PSS®E bus' name string because no other convenient fields exist in the PSS®E v35 network data model for a bus to add in customized data. An example would be bus #1's name would be changed from `NUCPLANT` to `123S-NUCPLANT` if bus #1 represented a bus at a substation referred to as `123S` on the PI server.

A `.csv` file is maintained which is a lookup table between PSS®E bus number and physical substation name. This file is version controlled using `git` to track changes and ensure no unintended editing occurs. Without this file the rest of the procedure is not possible because this effectively links the topology contained in PSS®E with the PI data. Populating this file only needs to happen once and can be automated to some extent using existing data sources.

2.4 download all PI tags

A list of all PI tags on the PI server is downloaded directly into Python memory and stored in a pandas `Series` [6] which supports vector-based pattern matching.

The performance of the methodology discussed in this paper is negatively impacted if PI tags are allowed to be deleted from the PI server (e.g., after physical equipment is salvaged, TLine renumbering/reconfiguration).

2.5 assign relevant PI tags to each PSS®E TLine

Customized Python objects are created which keep track of each PSS®E TLine and corresponding PI tags which can be assigned manually or automatically. Manual PI tag assignments are stored in `.csv` files and are required for PI tags which do not follow standard convention; this is less than 1% of the PSS®E equipment. Automatic PI tag assignments use regular expressions [7] applied to all PI tags downloaded in the previous step.

The PSS®E circuit IDs follow a convention in the OPBCs available to the authors' utility. The PSS®E circuit IDs for TLines are the last two digits in the TLine asset name (e.g., circuit ID **56** for TLine **456L**, circuit ID **42** for TLine **1142L**).

TLine powerflow PI tags for the authors' utility are in the form `123S_456L-10_MW` which would measure active power leaving substation `123S` on segment `10` of TLine `456L`. Active powerflow PI tags of all TLines connecting to substation `123S` are given by matches of applying regex `'^123S_\d+L.+MW$'` to a complete list of PI tags contained on the PI server. This regex matches PI tags starting with `123S_` followed by at least one consecutive digit followed by `L` and ending with `MW`.

The active power PI tags for a PSS®E TLine connecting substations **123S** to **234S** with circuit ID **42** are given by applying regex `'^(?:123|234)S_\d*42L.*MW$'` to a list of all PI tags. This regex matches PI tags starting either **123** or **234** followed by **S_** followed by any number of consecutive digits followed by **42L** and ending in **MW**. Example matches would be **123S_542L-10_MW** and **234S_542L-10_MW** which are the active power measurements at each end of TLine 542L. Active *and* reactive power measurements can be obtained with the following regex:
`'^(?:123|234)S_\d*42L.*(?:MW|MVAR)$'`

For each PSS®E TLine the from- and to- substation numbers are extracted from the PSS®E bus name and put into the following regex template with the circuit ID taken from the PSS®E network data record: `'^(?:`*from_sub*`|`*to_sub*`)S_\d*`*ckt_id*`L.*(?:MW|MVAR)$'`. The matches of the regex are stored in a Python object representing the PSS®E TLine; this object supports writing to disk, i.e., pickling, for later access.

2.6 download relevant PI data

The PI data is downloaded directly into Python memory for the selected timestamp and selected PI tags from the previous step. This is made possible by using OSIsoft's AF SDK .NET library [8] and the **pythonnet** Python package. All PI data is stored in a pandas **DataFrame** [6] which supports powerful indexing and efficient data analysis methods.

2.7 place TLines OOS based on their PI data

Given this methodology starts from *peak* OPBCs the majority of TLines in the case are in-service. The available PI data for each TLine is checked for zero powerflow and the given TLine is appropriately placed OOS in PSS®E.

2.8 scale generation/load in entire province to adjust powerflow on interconnections

The province of Alberta has the following interconnections:

- one $500\,\text{kV}$ AC TLine to British Columbia
- two $138\,\text{kV}$ AC TLines to British Columbia
- one $240\,\text{kV}$ AC TLine to Montana
- one back-to-back HVDC link to Saskatchewan

The PI data corresponding to these interconnections is obtained for the chosen timestamp. The sum of the PI data, i.e., the *setpoint*, is compared with the corresponding value obtained from the peak operation powerflow case, i.e., the *measurement*. The difference of these two values, i.e., the *error*, is used to scale generation and load in a subsystem representing Alberta buses. The case is solved and the error calculation is repeated to ensure the scaling was sufficient. This is made possible by using a particular PI tag representing Alberta internal generation plus imports on the interconnections which the authors' utility receive from their ISO.

The powerflow swing-bus in the OPBCs is located at one of the $500\,\text{kV}$ British Columbia buses. Scaling the Alberta buses as described here allows for accurate powerflows on the interconnections given the swing-bus is not located at an Alberta bus.

2.9 scale generation/load in specific areas while adjusting power order on HVDC links

Alberta has two $1000\,\mathrm{MW}$ bi-directional LCC-HVDC links. The *error* between power order in the OPBC and the PI data may be significant so the loads in the areas adjacent to the converter stations are scaled as follows to ensure the PSS®E case will solve after the HVDC power orders indicated in the PI data are used in the case:

1. obtain *setpoint* as the HVDC power order from PI data
2. obtain *measurement* as the HVDC power order in PSS®E OPBC
3. calculate $error = measurement - setpoint$
4. scale loads in planning area containing rectifier bus by $-\frac{1}{2}error$
5. scale loads in planning area containing inverter bus by $\frac{1}{2}error$
6. revise HVDC link power order in PSS®E to the *setpoint*
7. solve the case

For large magnitudes of *error* the procedure above is applied iteratively in small steps (e.g., $50\,\mathrm{MW}$) so ensure the PSS®E case continues to solve.

2.10 scale generation/load at each *known* substation

In this paper a *known* substation is one where PI data is available for all of the HV TLines connecting to the substation. The PI data is obtained for each HV TLine measured at the given substation. This PI data is summed to obtain the net generation/load at this substation. A positive value for this sum means the substation is providing power to the transmission network. A negative value for this sum means the substation is consuming power from the transmission network. The sum of the same TLine powerflows referenced to the given substation is performed in the PSS®E case. The difference of these two sums is the amount generation/load at this substation needs to be scaled to match the sum indicated by the PI data. PSS®E activity **SCAL** is used to perform the scaling on a subsystem containing only the buses belonging to this substation and then the case is solved. A check is performed to make sure the new net generation/load at this substation in PSS®E powerflow agrees with the PI data.

Each substation in the PSS®E case is checked for whether all the necessary PI data is available. If not all the PI data is available or some is reporting invalid values then scaling of the given substation is skipped.

The PSS®E OPBCs used by this paper use *bus-branch* network modeling instead of the more detailed *node-breaker* network modeling. Specific substations can still be referenced because of the mapping from each PSS®E HV bus to a physical substation as described in Section 2.3.

Figure 1 shows an example substation containing the following equipment:

- four $240\,\mathrm{kV}$ TLines
- three $138\,\mathrm{kV}$ TLines
- generation at the $13\,\mathrm{kV}$ bus
- loads at the $13\,\mathrm{kV}$ bus and $25\,\mathrm{kV}$ bus
- a connection to an independent power producer (IPP) which is otherwise not connected to the bulk transmission network

If the authors' utility owns the $240\,\mathrm{kV}$ and $138\,\mathrm{kV}$ buses then PI data for the connecting TLines will be available. This paper calculates the net generation/load at this substation to be the sum of powerflows on TLines #1-#7. The machine and load components at the IPP bus are considered to belong to the main substation. The authors' utility does not have reliable and consistent access to PI data for IPP facilities
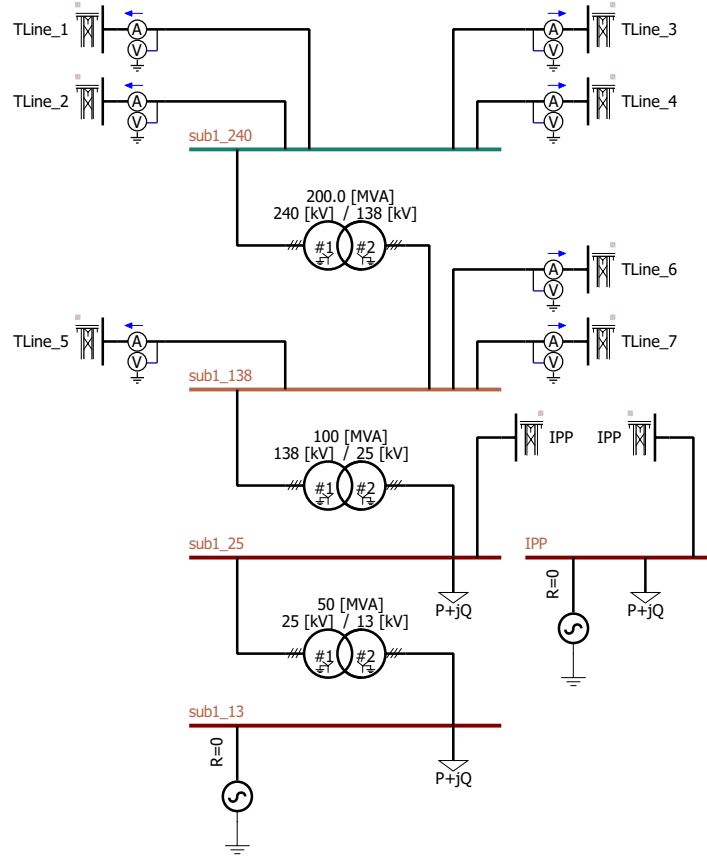
Figure 1: Example substation topology

because they are owned by other companies by definition. The methodology in this paper would proceed as follows:

$$setpoint = \sum_{PI}(HV\ TLines)$$
$$measurement = \sum_{PSSE}(HV\ TLines) \tag{1}$$
$$error = measurement - setpoint$$

If $error > 0$ then generation will be decreased and/or load will be increased. If $error < 0$ then generation will be increased and/or load will be decreased. The max/min active power output limits on subsystem generation are considered. The subsystem to be scaled consists of the two machines and three loads shown in Figure 1.

2.11 scale generation/load for *unknown* areas using defined interfaces

The authors' utility does not have complete access to PI data for GFOs, distribution facility owners (DFOs), IPPs, or other TFOs. This paper uses the PI data for the TLines interfacing between the authors' utility and the neighboring utility, i.e., jointly-owned TLines, to decide how to scale buses owned by other utilities.

Utilities owning buses in only one geographic area (e.g., Calgary and Edmonton distribution) can be scaled automatically by searching for TLines connecting the authors' utility to one of these distribution companies.

Utilities owning buses in separate geographic areas (e.g., other TFOs) can still be automated by specifying substations for which to find jointly-owned TLines.

Table 2: validation results for 8760 HPCs

voltage class [kV]	count	$q = 0.5$ [MW]	$q = 0.9$ [MW]	$q = 0.95$ [MW]
500	94,367	16.7	71.2	91.5
240	1,577,627	13.0	44.7	63.2
138	4,454,077	3.5	15.3	22.4
69	306,694	0.4	4.8	6.7

Table 3: median TLine thermal ratings

voltage class [kV]	summer [MVA]	winter [MVA]
500	2550	2550
240	500	610
138	120	146
69	39	48

2.12 validate the HPC

The HPC is validated by comparing the solved PSS®E TLine powerflows with the corresponding PI data. Results are grouped by voltage class of the TLines and quantity of TLines powerflow, i.e., MW, MVA.

3 Numerical Results

The procedure discussed in Section 2 was executed for 8760 hourly timestamps between 2021-May through 2022-April. These months are chosen based on the seasonal TLine ratings being defined as described in Section 2.2:

- summer ratings: May through October

- winter ratings: November through April

The validation errors, i.e., difference between PI data powerflow and PSS®E powerflow, for every TLine in all 8760 cases are combined into one data structure for further analysis. A statistical overview of the validation is provided in Table 2. Table 2 shows the validation results grouped-by voltage class. For each voltage class the following information is shown:

- number of TLines considered (count)

- median ($q = 0.5$)

- 90th percentile ($q = 0.9$)

- 95th percentile ($q = 0.95$)

where q refers to quantile. The median thermal ratings taken across all TLines represented in an OPBC are shown in Table 3 to provide context for the validation summary in Table 2. The 90th and 95th percentiles of the validation data are roughly 10% of the median thermal rating for each voltage class. The median validation data is insignificant compared to the median thermal ratings. These validation results are an improvement upon previous publications [4]. Given these HPCs are created with incomplete PI data the accuracy of the HPCs should be adequate for the needs discussed in Section 1.

4 Conclusion

This paper discusses a methodology to create HPCs when complete access to historical data is not possible (e.g., when neighboring utilities do not share historical data). The methodology discussed in this paper focuses on minimizing the amount of manual data entry from the user. Validation results of HPCs created with this methodology are shown to be an improvement over previous attempts in the literature.

BIBLIOGRAPHY

[1] Advanced Control Systems. [Online]. Available: https://www.acspower.com/energy-management-system/

[2] Siemens PTI. (2017). [Online]. Available: https://youtu.be/VS6DerBnFDE

[3] A. Alam, V. Frey, S. Larson, S. Mundade, S. Ruelas, and A. Herasimava, "Improving operational planning study models with historical generation and load data," in *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T D)*, May 2016, pp. 1–5.

[4] J. Schwartz, "A novel approach to create historical powerflow cases from PI data using open source software," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*, 2020, pp. 1–5.

[5] OSIsoft PI System. [Online]. Available: https://www.osisoft.com/pi-system/

[6] W. McKinney, "Data structures for statistical computing in python," in *9th Python in Science Conference*, 2010, pp. 51–56.

[7] Wikipedia. Regular expressions. [Online]. Available: https://en.wikipedia.org/wiki/Regular_expression

[8] OSIsoft. PI System AF-SDK. [Online]. Available: https://techsupport.osisoft.com/Documentation/PI-AF-SDK/Index.html