CIGRE-487 **2022 CIGRE Canada Conference & Expo**
Calgary, Alberta, Oct.31 – Nov. 3, 2022

Interactive Parallel-Coordinates – A Bridge between Probabilistic and Deterministic

Ming Hu
Alberta Electric System Operator
CANADA

SUMMARY

This paper describes the adoption of an interactive data visualization technique – "parallel coordinates" to perform the deterministic and probabilistic analysis for power system historical performance and future performance assessment. It further discusses the concept, application examples and potentials to bridge between the traditional deterministic methods and the evolving probabilistic methods used in power system reliability assessment.
The power system has evolved over past few decades towards more decentralization, more integration of variable renewable sources and distributed energy sources. These trends bring more variability and uncertainty for system operation and planning, that deserves more focus and shift towards using probabilistic methods in power system reliability assessment. Probabilistic methods have been developed and implemented by pioneers in power systems since the 1930s. Accompanied by continual advancements in computing power, "the application of probabilistic methods in power systems has been gradually increasing" as they "are able to respond to the actual stochastic factors that influence the reliability of the system." However, due to being highly abstract and computationally demanding, probabilistic methods are still facing challenges in full implementation and have yet to replace the deterministic methods - especially for the so-called composite system - despite the significant advancements in computing power. Even the pioneering specialists believe that "it is quite likely that deterministic approaches will continue to be used widely in practice. In this case, a bridge between the two approaches (probabilistic and deterministic) may be a valuable way forward."

This paper introduces an interactive data visualization technique – "parallel coordinates", and its implementation examples to perform probabilistic analysis for power system historical performance and future performance assessment and discuss why and how it can help to bridge between abstract probabilistic methods and practical deterministic methods.

The paper at first discusses parallel-coordinates' concept and power when combining with interactive data visualization data technique; then to discuss its connection between probabilistic and deterministic analysis in both concept and application level, and why and

ming.hu@aeso.ca

how the interactive parallel-coordinates technique can be used to bridge the connection; further, to discuss the implementation and actual application and result of the interactive parallel-coordinates approach to perform electric power system historical performance analysis as well as future planning analysis in day-to-day work.

In conclusion, as a single piece of the methods and tools used in reliability assessment, this interactive parallel-coordinates tool can provides: 1) more efficient analysis and better understanding of historical performance and impacting factors; 2) make better study assumptions on operation conditions and model configurations; 3) more efficient results sanity check and analysis, and an efficient screening mechanism to identify needs of further investigation or analysis; 4) a consistent and transparent mechanism to facilitate efficient discussion, communication and coordinated solution developing among different team members and different teams.

KEYWORDS

Parallel Coordinates, Probabilistic, Deterministic, Reliability Assessment, Data Visualization, Big Data

I. INTRODUCTION

The power system has evolved over past few decades towards more decentralization, more integration of variable renewable sources and distributed energy sources. These trends bring more variability and uncertainty for system operation and planning, that deserves more focus and shift towards using probabilistic methods in power system reliability assessment. Probabilistic methods have been developed and implemented by pioneers in power systems since the 1930s. Accompanied by continual advancements in computing power, "the application of probabilistic methods in power systems has been gradually increasing" [3] as they "are able to respond to the actual stochastic factors that influence the reliability of the system." [2] However, due to being highly abstract and computationally demanding, probabilistic methods are still facing challenges in full implementation and have yet to replace the deterministic methods - especially for the so-called composite system [2][3] - despite the significant advancements in computing power. Even the pioneering specialists believe that "it is quite likely that deterministic approaches will continue to be used widely in practice. In this case, a bridge between the two approaches (probabilistic and deterministic) may be a valuable way forward." [2].

This paper introduces an interactive data visualization technique – "parallel coordinates" [1], and its implementation examples to perform probabilistic analysis for power system historical performance and future performance assessment and discuss why and how it can help to bridge between abstract probabilistic methods and practical deterministic methods.

II. PARALLEL COORDINATES

A. Concept of Parallel Coordinate

Parallel coordinates are a common way of visualizing multivariate data's complex relations and impacts among all variables [1], especially for high-dimensional data where interrelations and interactions between variables are too trivial to be communicated and understood through plain descriptions, or too abstract through statistical descriptions.

Like the regular Cartesian coordinates system, parallel coordinates have a separate coordinate (axis) for every dimension of the data and represents each point's value by its position on corresponding dimensions. The key difference is that it arranges axes parallel to each other instead of perpendicular to each other. Figure 1 below Shows the same data represented in a vertical-axis space versus a parallel-axis space for 2-dimension (left) and 3-demension (right).

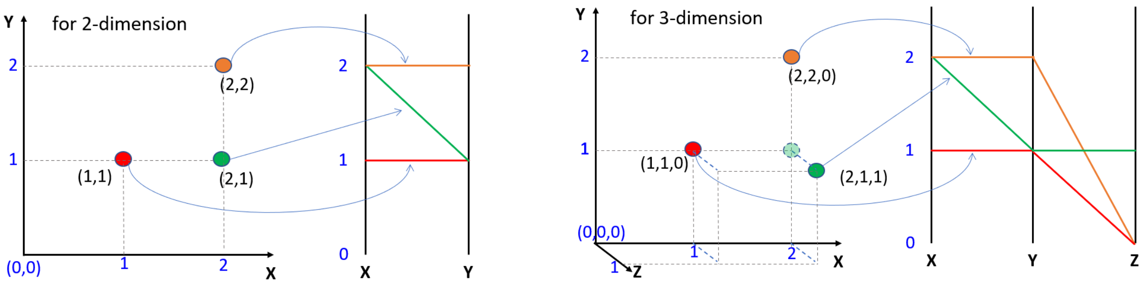

Figure 1. Concept of parallel coordinates

With all coordinates arranged in parallel, all multivariate data can be represented in 2d. It is intuitive and avoids the use of highly abstracted high-dimensional representations that could overwhelm even the brightest minds.

B. Interaction Makes It More Powerful

Further, the recent advancements of interactive computer visualization techniques make it convenient to display, hide or highlight all or any part of the data through filtering the data by one or more dimension(s), as shown in Figure 2. It can also animate dynamic changes from filtering through mouse interactions (move, drag and click …).

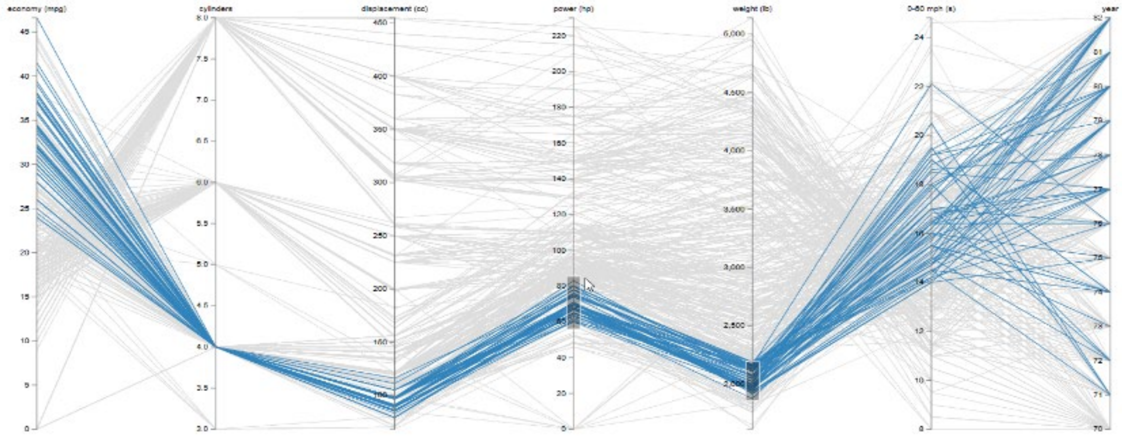

Figure 2. Interactive parallel coordinates

The interactive parallel coordinates enable us to investigate the individual factor (an axis) and its interactions with another factor (lines between any two axes) without losing the overall picture (all axis and all lines) in the background, and its combined interactions (lines between multiple axes) with all other factors on the chart. This bridges the gap between overall top-down methods and detailed bottom-up methods.

On a 20-inch+ monitor screen, an interactive parallel-coordinates chart with 20 separate axes can be clearly laid out. With its interactive features, it provides the equivalent information of 190 (= 20*(20-1)/2) scatter charts in terms of correlation analysis between factors, as shown in Figure 3.

1 N-dimension PC chart ➔ N*(N-1)/2 scatter charts

Dimensions	Scatter $C_{(N,2)}$
5	10
10	45
20	190
30	435

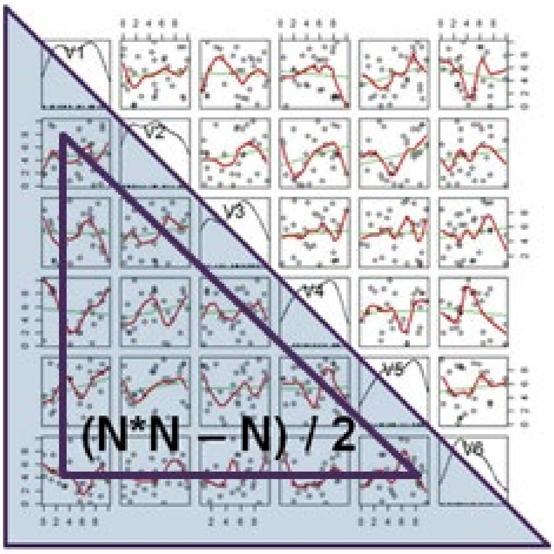

Figure 3. Equivalent number of correlation analysis by using scatter charts

III. CONNECTION BETWEEN PROBABILISTIC AND DETERMINISTIC IN CONCEPT & APPLICATION

A. Deterministic & Probabilistic in Concept

To clearly distinguish between deterministic and probabilistic methods, we adopt the concept of "macroscopic versus microscopic" used in statistical physics [4]. The "deterministic" is to focus on individual "microscopic" states (microstates) and related system operation conditions that includes more practice details. As it involves practice details, the focus of analysis is usually limited to a certain number of credible and significant operation states (or scenarios), due to limitations related to data and computation requirement. This deterministic method is naturally suitable for the operation practice, but for the purpose of strategy and planning, it can be overwhelming with all possible microstates included, or might not be adequate if without as some significant microstates might get missed.

On the other hand, the "probabilistic" is to focus on some aggregated "macroscopic" system state (macro-state) and related aggregated characteristics and patterns driven by all possible underlying microstates. By projecting all individual microstates with overwhelming practice details on an aggregated macro-state with only few aggregated performance indexes, many practice details must be abstracted, simplified (otherwise no point to have the macro-state), in order to refocus on system aggregate characteristics that is more suitable by using probabilistic methods and related abstract indexes.

Therefore, some gaps must exist between the methods used for microstates (bottom-up) and for macro-states (top-down) due to the necessary abstractions and simplifications involved. Conceptually, the parallel-coordinates technique, which can represent both macro-state (overall system) and microstates (with practice details) on the same chart, has good potential to bridge the gap between macro-state-oriented probabilistic methods and microstate-oriented deterministic methods.

B. Connection at Application Level

"Two main approaches exist for evaluating system reliability (by using probabilistic method); 1) analytical and 2) simulation. Simulation techniques, often known as Monte Carlo simulation, estimate the indexes by simulating the actual process and random behaviour of the system." [2] Each simulation represents an individual microstate, so the microstate-oriented analysis is already used in probabilistic methods. However, there are still significant simplifications and abstractions involved: 1) a significant number of factors and their correlations need to be simplified in order to have a practical Monte Carlo process to simulate a large number of scenarios that can represent reasonable conditions of potential reality. This requires context knowledge and expertise to make sure the randomly generated scenarios can correctly capture all expected significant patterns and interactions between related key factors, as well as not create any unrealistic significant pattern; 2) a large number of simulation results also require context knowledge and expertise to check, review, interpret and communicated into meaningful information in order to support related decisions.

Although the Monte Carlo simulation itself can be performed automatically and efficiently by a computer program, the pre-simulation work of simplifications and related data preparation and model configurations, and post-simulation work of sanity checks, analysis, interpretation, and communication cannot be fully automated, and become the bottleneck of the whole probabilistic method process. Also, it is still heavily dependent on abstract indexes to describe, measure and communicate the reliability performance results. The high abstraction involved in representing the system macro-state from simulations of many microstates introduces an abstract layer that blurs the original and underlying connections. This was discussed in reference [3], noting that "difficulties in interpreting probabilistic results … difficulties related to terminology, concepts, and theory modelling hypotheses … causing the slow rate of dissemination of probabilistic methodologies when applied at operation or planning stages of electric systems."

The interactive parallel-coordinates technique has the advantage to examine and to compresence the system macro-state characteristics, patterns, and their interactions between many different driven factors together with all its individual microstates, in an intuitive and visual manner (further discussed and illustrated in next section). For pre-simulation, it can significantly enhance the capability and efficiency to determine limited number of key factors with proper level of simplification. For the simulation results checking, analysis, interpretation, and communication, it can enhance the capability and efficiency to reveal and identify expected or unexpected correlations and patterns, to directly associate abstract probabilistic indexes (of system macro-state) results with correspondent individual operation scenarios (system microstates). Serving as a bridge between macro-state indexes with individual microstate operation conditions, it enables less abstract and more transparent and efficient discussion and communication in conventional wisdom with common sense.

IV. IMPLEMENTATION AND EXAMPLES

A. Adoption and Implementation

The adoption and implementation of the interactive parallel-coordinates technique was inspired by a simple online "Nutrient Explorer" interactive dashboard [5]. The dashboard presents the data of 14 different nutrients for over 7,500 different foods. This high amount of data is overwhelming to most. However, the interactive parallel-coordinates used on the dashboard make it convenient to observe, and easy to comprehend. In order to take the advantage of its convenience, we start to implement and test parallel-coordinates to conduct historical and forecast performance analysis for system with components of generation, transmission and market (composite system). The implementation is mainly based on web-based front-end technology that enables easy sharing with an internal URL link and a web browser. The JavaScript library of "visual toolkit for multidimensional detectives" [6] based on Data-Driven Document JavaScript Library (D3JS) [7] is a key part for this implementation. The data preparation for displaying on the dashboard is as straightforward as preparing a multiple-column data file in simple csv format. The first row of the csv file has the labels for each dimension (coordinate/axis) in each cell, and each row of the csv file represents a data point with the value of each dimension (coordinate/axis) in the corresponding column labelled in the first row. Different data files can be prepared and displayed on the dashboard by linking the file name as the dashboard URL parameter. It is also possible to enable the dashboard to get data in other formats or directly from other data sources, such as a database.

B. Example of Historical Performance Analysis

The parallel-coordinates based historical performance analysis can include significantly large amounts of identified necessary data items, such as loads (total, by region), generators output (total, by region, by type), line flows (by line, cut-plane) and related operation limits, all on a single interactive dashboard for intuitive visualization analysis. Figure 4 shows about one third of the whole dashboard that has about 50 thousand (5 years and 8 month) historical hourly data of 40+ measures that was initially identified as performance measures (supply cushion, line flow margin, etc.) and impacting conditions (levels of loads, flows, generations, interchanges, pool price, calendar data, etc.). It can dynamically calculate basic statistics such as probability, conditional probability, count, min, average and max for every measure and condition, on-the-fly, through mouse interactions. For example, filtering with two conditions of high AIL (load) and low wind generation (labelled as GenWind) by mouse interactions (the two red-yellow blocks) on the dashboard. All data meet the conditions are highlighted as dark-blue lines, with all other data shown as light-blue lines in the background.

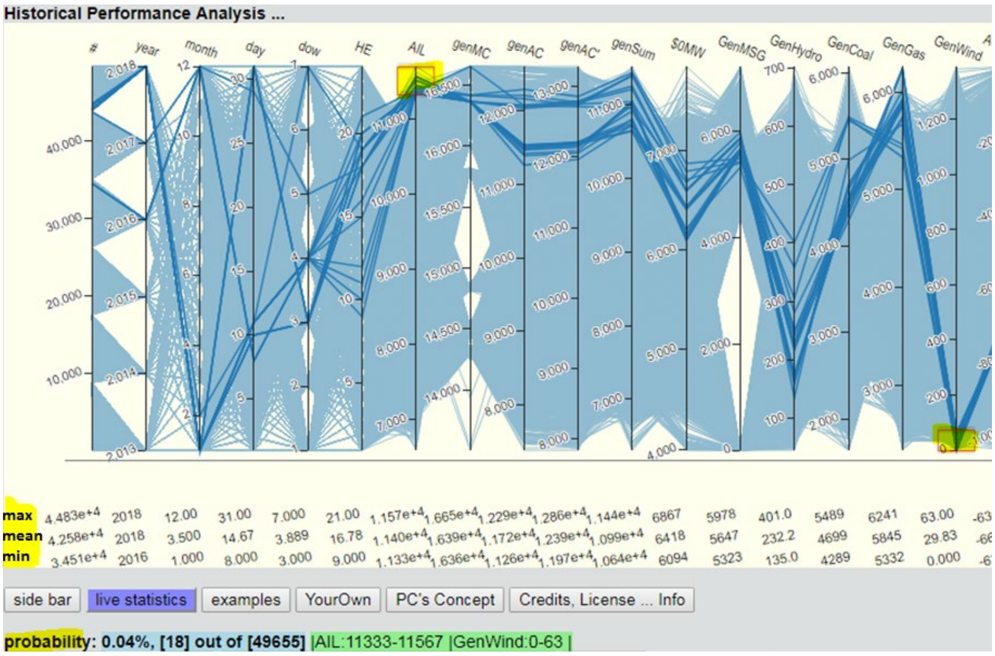

Figure 4. Example screen for historical performance analysis

The dashboard shows the count (18) and probability (0.04%) of the selected points, and the corresponding min, mean and max values of all other measures and conditions, as well as the calendar information are all displayed and highlighted dynamically with the mouse interactions on the dashboard.

Through mouse interactions on one or multiple axis(s), we can quickly: 1) observe the basic statistics and distribution of any one or multiple measure(s) and condition(s); 2) observe the relation, correlation details (for different range) between any two or more measure and condition; 3) understand the system normal or abnormal in terms of related measures and conditions; 4) identify expected or unexpected trends and patterns, or outliers for further investigation.

On top of the above information and knowledges obtained, to have better study assumptions, scope and more proper model simplifications and configurations for future studies, we can: 1) further validate with the study assumptions (of conditions) and result of our past studies in order to have better study assumptions and model for future studies; 2) further validate the simplifications made for the study model and related model algorism/configuration details, and adjust the model simplifications approach accordingly by including strong impacting factors or excluding weak impacting factors.

C. Example of Future Performance Forecast Analysis

The same dashboard can be used for future performance forecast analysis by replacing the linked data source of historical hourly data with the data source (similar csv file) of simulated future hourly data that is generated by study simulations under all different assumptions for the same or similar performance measures and conditions. In addition to the impacting factors used and identified in historical performance analysis, for the future performance analysis, we also include additional factors (dimensions) to describe new scenarios with 1) different future generation development and dispatch scenarios (driven by different assumptions in future market simulation work); 2) different wind generation profiles based on different weather-years; and 3) different transmission project development staging scenarios.

Following the similar process used for historical analysis, the dashboard with simulated future performance measures and conditions can help us: 1) identify and validate expected and unexpected future operation conditions combinations against study assumptions and simplifications; 2) identify the future reliability performance concerns (at low or no reliability margin) and corresponding operation condition details; 3) identify future patterns that enable further simplification regarding study assumptions, scope, and model configurations; 5) identify significant outliers for further investigation or analysis; 6) investigate and develop mitigation alternatives of any identified specific reliability concern in the context of all associate microstates' operation condition details.

V. CONCLUSION AND DISCUSSION

This paper discussed our effort of adopting and implementing the interactive parallel-coordinates technique to bridge the gap between deterministic methods and probabilistic methods for reliability assessment that involve transmission system, renewable development, and generation market. As a single piece of the methods and tools used in reliability assessment, this interactive parallel-coordinates tool provides us: 1) more efficient analysis and better understanding of historical performance and impacting factors; 2) make better study assumptions on operation conditions and model configurations; 3) more efficient results sanity check and analysis, and an efficient screening mechanism to identify needs of further investigation or analysis; 4) a consistent and transparent mechanism to facilitate efficient discussion, communication and coordinated solution developing among different team members and different teams.

As many of the probabilistic reliability indexes and related concepts were established more than half a century ago when today's computing power and related data processing and visualization capabilities were not available or possible. Highly abstract and sophisticated indexes and concepts were the best and only choice. The significant advancement of data visualization capabilities could help to see through the abstract layers intuitively for better understanding, interpretation, and communication.

Further, there might even be opportunity to create and define visualization-based reliability measures to supplement the existing reliability indexes for facilitating better understanding, interpretation, and communication in more intuitive and less sophisticated manner among conventional wisdoms. The organization-wise evolution of the power system toward less-centralization with more marketizations will involve more people and parties that might not be familiar with or equipped to understand abstract and sophisticated concept and indexes on power system reliability. The data visualization techniques such as interactive parallel coordinates discussed in this paper can serve as the common-sense bridge through the abstractions and sophistications.

BIBLIOGRAPHY

[1] Wikipedia, Parallel coordinates [online] https://en.wikipedia.org/wiki/Parallel_coordinates
[2] R. Allan, R. Billinton, Probabilistic assessment of power systems. IEEE PROCEEDINGS OF THE IEEE, VOL. 88, NO. 2, FEBRUARY 2000 https://ieeexplore.ieee.org/document/823995
[3] M. Schilling, J. Souza, M. Filho, Power System Probabilistic Reliability Assessment: Current Procedures in Brazil, IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 23, NO. 3, AUGUST 2008
[4] Wikipedia, Statistical physics [online] https://en.wikipedia.org/wiki/Statistical_physics
[5] K. Zhang, Nutrient Explorer [online] http://bl.ocks.org/syntagmatic/raw/3150059/
[6] K. Zhang, A visual toolkit for multidimensional detectives [online] https://syntagmatic.github.io/parallel-coordinates/
[7] Mike Bostock, Data-Driven Documents [online], https://d3js.org/