Establishment of a site-wide meteorological database for Level 3 PSA and EPZ analysis

Seunghwan Kim [a], Sung-yeop Kim [a,b*], Sunghyun Park [c]
[a]*Korea Atomic Energy Research Institute, 111 Daedeok-daero 989beongil, Yuseong-gu, Daejeon 34057, Republic of Korea*
[b]*University of Science and Technology, 217 Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea*
[c]*Central Research Institute, Korea Hydro & Nuclear Power Co., Ltd., 70 Yusung-daero 1312beongil, Yuseong-gu, Daejeon 34101, Republic of Korea*
**Corresponding author: sungyeop@kaeri.re.kr*

***Keywords :** meteorological data, weather sequence, offsite consequence analysis, Level 3 PSA

1. Introduction

Level 3 Probabilistic Safety Assessment (PSA) focuses not on predicting the outcomes of accidents, but rather on evaluating the potential impacts of accidents that have not actually occurred. Since weather conditions directly influence offsite consequence analyses, it is essential to acquire accurate and undamaged meteorological data. For example, if one year of site-specific meteorological data is available at hourly intervals, there are 8,760 possible accident initiation times.

The meteorological data set spanning from the initiation point to the end point is referred to as a weather sequence (or weather trial). This weather sequence plays a crucial role in assessing the progression of the accident and its impacts. Therefore, acquiring and maintaining site-specific weather sequences is essential for safety assessment of nuclear power plant, as they provide indispensable data for predicting the dispersion pathways and impacts of radioactive materials during accidents. However, perfectly securing and maintaining weather sequences presents several challenges. An analysis of meteorological data collected from five domestic nuclear power plant sites (Hanul, Kori, Saeul, Wolsong, and Hanbit) over a five-year period (2019− 2023) revealed that 0.40% of the total data was missing.

This study describes the development of a site-wide meteorological database for domestic nuclear power plants, incorporating gap-filling techniques. As a case study, it presents the meteorological input data generated for the Hanul site and discusses the selection of representative meteorological years.

2. Missing Data Imputation Methods

When missing weather data occurs, it is crucial to fill these gaps using appropriate methods to create a complete weather sequence. For instance, missing data can be estimated using statistical methods, or supplemented by referencing historical data under similar conditions.

Imputation refers to the process of estimating and filling in missing values within a dataset, serving as a critical preprocessing step in statistics and data science. This process enhances the accuracy and completeness of analyses by compensating for data loss and reducing bias that would arise from simply deleting missing values. Major imputation methods include Simple Imputation, Constant Value Imputation, Regression Imputation, Multiple Imputation, K-Nearest Neighbors (KNN) Imputation, and Time Series Imputation, all of which contribute to the reliability of data analysis. For datasets collected over time, such as meteorological records, time series imputation methods are commonly employed, including Last Observation Carried Forward (LOCF), Next Observation Carried Backward (NOCB), and linear interpolation [1, 2].

In particular, when compensating for missing meteorological data at nuclear power plant (NPP) sites to complete weather sequences, the following criteria and methods are typically applied.

- **Maintain Data Quality**: Compensate for missing data without compromising the quality of the weather data. However, weather data with more than 10% loss should not be used.
- **Meteorologist Review**: If meteorological data is lost for an extended period, it should be reviewed by a meteorologist.
- **Compensation Methods**:
  - Interpolation: Use interpolation when the time interval is not large.
  - Refer to Nearby Meteorological Tower: Fill in the missing data by referring to the time data of the nearby meteorological tower.
  - Use Surrounding Data: Fill in the missing data by referring to the data from the day before and the day after.
  - Create a Reasonable Algorithm: Supplement the data by creating a reasonable algorithm.

3. Establishment of Meteorological Database for Domestic NPP Sites

3.1 Collection Status of Meteorological Data at Domestic Nuclear Power Sites (5 Years)

Table 1 presents the collection status of meteorological data at five domestic nuclear power sites over a five-year period (2019-2023). The values in each cell represent the number of missing data records for each site on a yearly basis. For example, in 2019 at the Hanul site, 8760 meteorological records should have been collected, but 18 hours (18 weather records) of data were missing and thus not collected.

Table 1: Meteorological Data at Domestic Sites (5 Years)

Year \ Site	Hanul	Kori	Saeul	Wolsong	Hanbit
2019	18	32	11	25	84
2020	43	6	6	22	45
2021	13	10	133	10	199
2022	10	4	13	21	32
2023	21	9	17	37	65
Missing Hours	105	61	180	115	425
Missing Rate	**0.24%**	**0.14%**	**0.41%**	**0.26%**	**0.97%**
Average	**0.40%**				

As shown in Table 1, the average data loss rates over five years for the Hanul, Kori, Saeul, Wolsong, and Hanbit nuclear power plants are 0.24%, 0.14%, 0.41%, 0.26%, and 0.97%, respectively. The overall average data loss across all sites is 0.40%.

Table 2 illustrates the hourly meteorological data gaps at the representative Hanul site over five years. A total of 105 hours of meteorological data were missing, which posed challenges in developing continuous weather sequences.

Table 2: Status of Missing Weather Data at Hanul Site (Hourly)

Month \ Year	2019	2020	2021	2022	2023	Total (hours)
Jan						
Feb						
Mar						
Apr					▽4	**4**
May	▽13	▽6	▽5	▽4	▽7	**35**
Jun						
Jul		▽2	▽1			**3**
Aug			▽1			**1**
Sep			▽2			**2**
Oct					▽6	**6**
Nov	▽5	▽3	▽4	▽6	▽2	**20**
Dec		▽32			▽2	**34**
Sum (hours)	**18**	**43**	**13**	**10**	**21**	**105**

As observed in the meteorological data status at the Hanul site, meteorological data is an essential input for offsite consequence analysis. Therefore, it is crucial to supplement missing meteorological data to create a complete weather sequence.

Based on the time series imputation methods described in the previous section, the following approaches were employed to address the gaps in meteorological data.

- Interpolation Method: For relatively minor data deficiencies, interpolation was used to fill in the gaps.
- Near-Time Data Supplementation: For larger data deficiencies, data from similar patterns on the previous or next day was utilized to supplement the missing information.
- Referencing Nearby Weather Towers: When weather data was missing, data from nearby weather towers at the same time was referred to for supplementation.

The meteorological data for the Hanul site was prioritized as the pilot case study and completed using this approach, which was subsequently extended to construct a comprehensive weather sequence database for all sites, including Saeul, Wolsong, and Hanbit.

The weather sequences stored in the meteorological database were converted into the input format required by the MACCS code using KMAC (KHNP Meteorological Data Analyzer & Converter for MACCS Code) [3]. KMAC not only handles data format conversion but also includes functionality for selecting representative meteorological years from multiple data sets.

Fig. 1 demonstrates the weather data conversion interface for the Hanul nuclear power plant using KMAC, while Figure 2 shows an example of the converted data in MACCS input format.

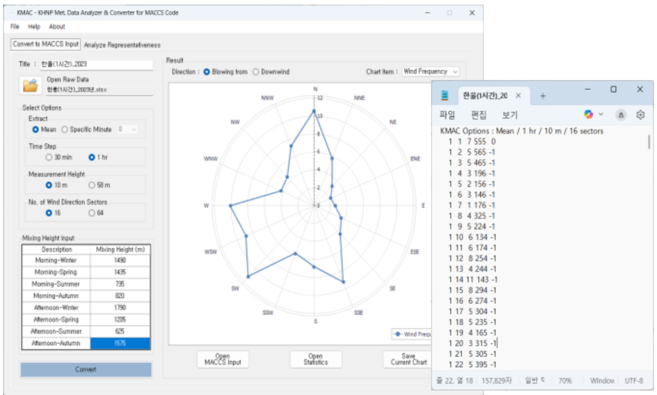

Fig. 1. KMAC weather data analyzer and MACCS input produced by KMAC

Selecting representative weather data from multi-year datasets stored in the meteorological database is a critical process. Conservative approaches may involve selecting data that shows the highest impact results, while other methods could focus on data with average impact results or the most qualitative data with minimal data loss. This selection process requires careful consideration from multiple perspectives.

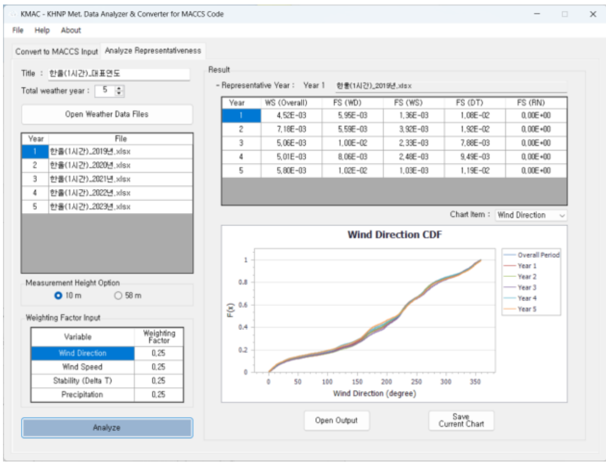

Fig. 2 Example of Representative Meteorological Year Selection for Hanul Site

Fig. 2 demonstrates the selected representative meteorological data for the Hanul site, chosen using the KMAC representative year evaluation module.

4. Conclusions

It is ideal to have meteorological data with complete data integrity, achieved through maintaining equipment reliability and establishing redundant data transmission paths, along with enhanced security measures to ensure stable data collection and transmission. However, when such measures are insufficient and result in missing meteorological data, it is essential to consider effective methods for supplementing the missing data to ensure the continuity of the meteorological sequence.

This study addresses methodologies for meteorological data gap compensation to complete weather sequences and implements them in a database for five Korean NPP sites. The methodology was initially validated at the Hanul site, where 105 hours of missing meteorological data were compensated using time-series analysis techniques to enable full-period analysis.

This approach was subsequently extended to develop a comprehensive meteorological database for all domestic NPP sites. The developed meteorological database for all plant sites is projected to function as essential input data for Level 3 PSA and EPZ analysis applications.

ACKNOWLEDGEMENT

This work was supported by an Innovative Small Modular Reactor Development Agency grant funded by the Korean Government (MCEE: Ministry of Climate, Energy and Environment) (No. RS-2023-00258118).

REFERENCES

[1] D. Atkinson, R.F. Lee, Procedures for Substituting Values for Missing NWS (National Weather Service) Meteorological Data for Use in Regulatory Air Quality Models, 1992.

[2] S. J. Lee et al., Probabilistic Safety Assessment of Nuclear Power Plants, Korea eBook Publishing Company, Seoul, pp. 89-143, 2021.

[3] FNC. Tech, KMAC (KHNP Meteorological Data Analyzer & Converter for MACCS Code) User Guide, FNT. Tech, 2018.

[4] SNL, MACCS User Guide – Version 4.2, SAND2023-01315, Sandia National Laboratories, 2023.