Real-Time Monitoring and Control of Pharmaceutical Production Processes Using Spectroscopic Data

分光データを利用した医薬品生産プロセスのリアルタイムモニタリングと制御

Dr. Sanghong Kim
金 尚弘

To realize the continuous production in the pharmaceutical industry, it is essential to monitor the production process in real time, but it is often difficult. In addition, conventional methods for predicting pharmaceutical quality from near-infrared spectra have drawbacks, such as a decrease in prediction accuracy over time. In this study, a stable and accurate method for predicting the quality of pharmaceuticals from near-infrared spectra was developed using data science techniques. As a result, a new analysis method was established that can contribute to realize real-time quality management and control, and to improve the efficiency of the pharmaceutical production processes.

医薬品連続生産の実現には生産プロセス内の医薬品の情報をリアルタイムにモニタリングする技術が必要不可欠であるが，それは困難であることが多い。また，従来の近赤外スペクトルから医薬品品質を予測する手法には，予測精度が経時的に低下するなどの欠点があった。本研究では，近赤外スペクトルから医薬品品質を安定的かつ高精度に予測する手法の開発を目的とし，データサイエンスの技術を活用して品質予測のためのデータ解析手法を開発した。その結果，従来法とは異なる新しい解析手法を確立することができた。本手法を用いたリアルタイムな品質管理・制御を行うことで，医薬品生産プロセスの効率化を実現することが期待される。

Introduction

In the pharmaceutical production process, strict quality control based on Good Manufacturing Practice (GMP) is required. In recent years, there has been a shift from batch processes to continuous processes and the introduction of Quality by Design (QbD) and Process Analytical Technology (PAT) [1]-[4] to improve the efficiency of pharmaceutical production processes. In order to achieve the above goals, it is necessary to monitor and control the quality of pharmaceuticals in real time. However, it is often difficult to measure the quality of a drug directly in real time. In order to solve this problem, this research developed and put into practical use a method for constructing a statistical model to estimate drug quality from variables that can be measured relatively easily in real time, such as near-infrared spectra.

Main

Process Nonlinearity, Model Maintenance, Input Variable Selection

There are many studies on methods for predicting drug quality from near-infrared spectra, but most of them use the classical method of partial least squares (PLS) [5]-[8]. However, PLS is known to have some problems such as inability to cope with nonlinearity and high maintenance load because the estimation accuracy decreases with time. In this research, a new method called locally-weighted PLS to improve PLS was developed. Locally-weighted PLS can cope with nonlinearity by selectively using historical data that are similar to those around the conditions for which quality prediction is required. In addition, since a new model is automatically constructed for each quality prediction, the estimation accuracy can be maintained for a long time. Figure 1 shows the results of PLS and locally-weighted PLS prediction of the concentration of the active ingredients in the powder after mixing for a mixing process. Locally-weighted PLS reduced the root mean square error of prediction (RMSEP) from 1.84 to 1.13. This allows for more strict quality control and more efficient drug production processes.

In addition to the development of locally-weighted PLS, a method to select the variables that are useful for estimation was developed. It is said that various physicochemical

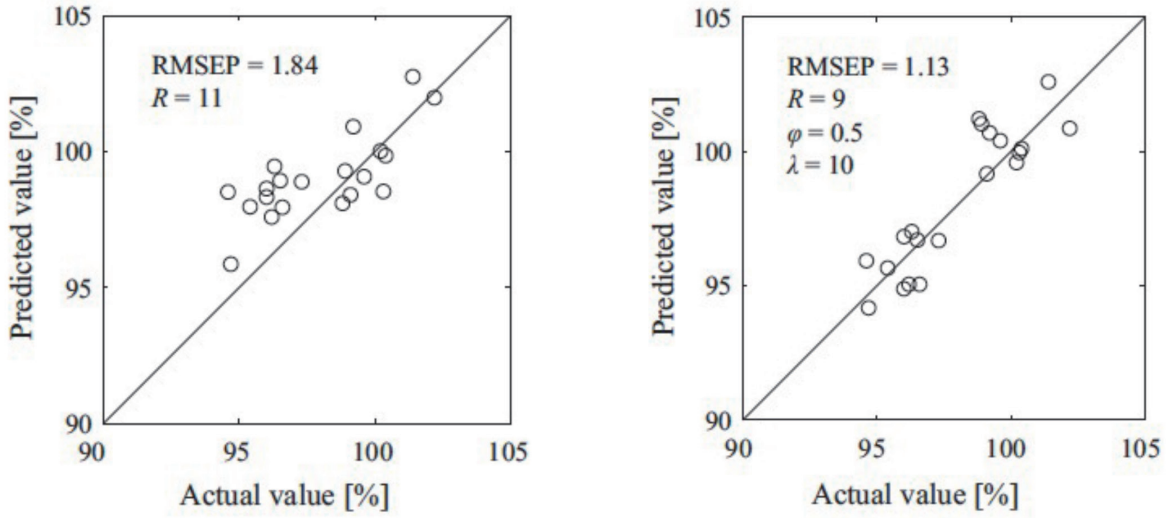

Figure 1 Prediction results by PLS (left) and locally-weighted PLS (right)

factors coexist in near-infrared spectra, and it is not always possible to make good predictions by selecting functional groups and their corresponding wavelength regions, or by selecting wavelength regions by referring to the spectra of active ingredients and other ingredients as pure substances. The wavelength regions selected by the developed method are shown in Figure 2 as the blue region. The target granule is composed of API (Active Pharmaceutical Ingredient) and five other components. The peaks of API and other components do not exist in the selected region, which may seem useless for estimation. However, the RMSEP was reduced to 1.13 from 1.84 by using absorbance at those wavelengths compared to use the absorbance at wavelengths selected by conventional method: variable importance on the projection (VIP) [12].

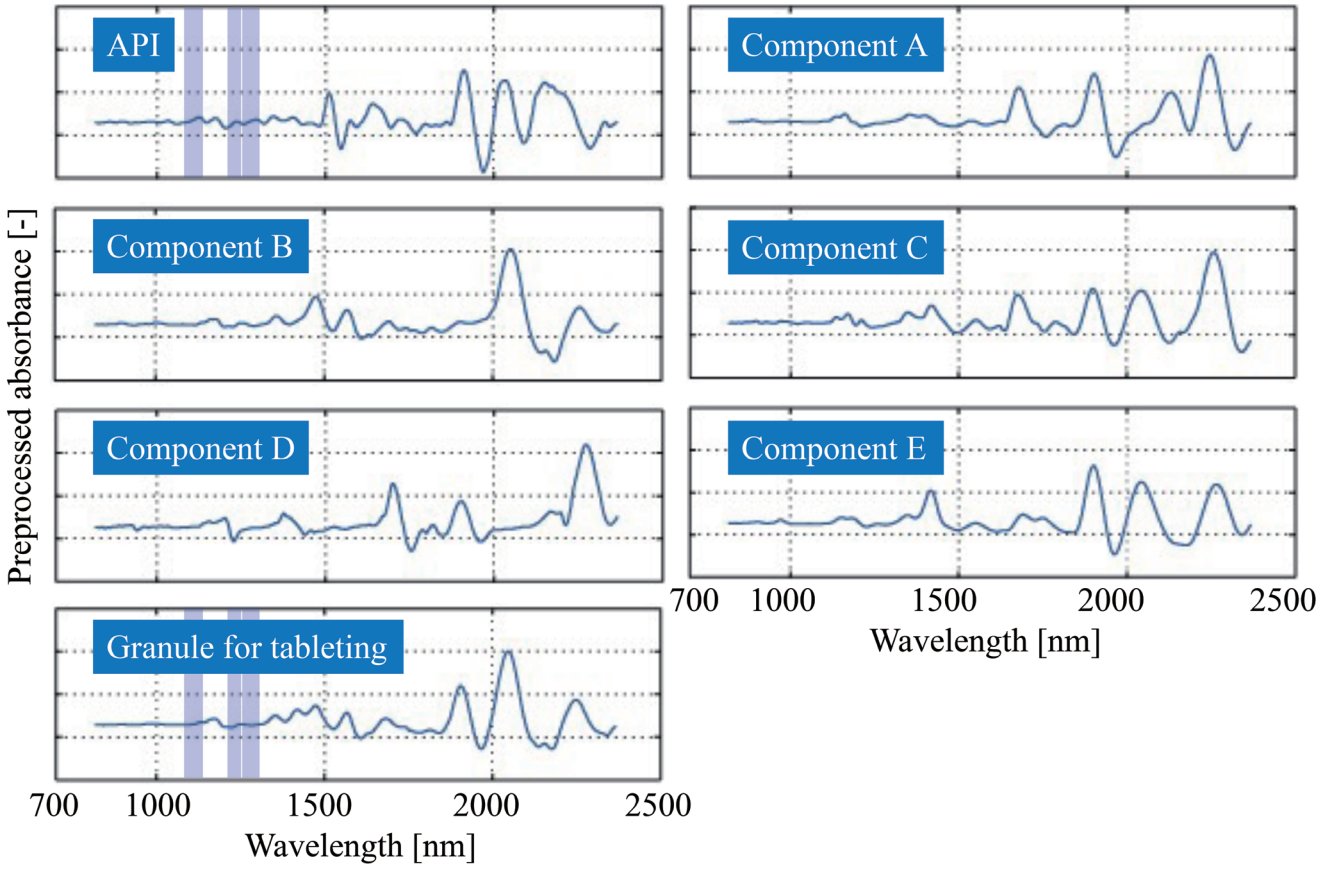

Figure 2 Wavelength selection result of the proposed method and preprocessed absorbance of each component of a drug

The detailed information can be found in the paper [9].

Input Variable Scaling

When using statistical methods to build a quality prediction model, it is necessary to pre-process the input variables. Although input variable scaling, a data preprocessing method in which the values of each input variable are multiplied by the scaling factor of the input variable, can have significant effect on the estimation performance of soft sensors, research on input variable scaling has not been actively conducted. Hence, this research focuses on input variable scaling. In past research, autoscaling was commonly used [11-13]. In addition, Pareto scaling, level scaling, Poisson scaling, range scaling, and VAST scaling [14] have been considered.

In this research, a new method of preprocessing near-infrared spectra was developed, and it was found that it can improve the prediction accuracy of the impurity concentration inside the device. In the case of spectral data, it is difficult to evaluate the importance of each variable because hundreds to thousands of variables need to be handled simultaneously.

An example of the application of the proposed method to the prediction of magnesium stearate concentration is shown in Figure 3. Figure 3 shows the near-infrared spectra of magnesium stearate and methanol solutions of magnesium stearate with different concentrations. Unlike Section 2-1, it was known in advance that the blue region in the figure is important for concentration prediction because it is a solution system. The evaluation results are shown in Figure 4. From Figure 4, it can be seen that the proposed method is able to correctly determine the important regions for concentration prediction. Accordingly, the prediction error was reduced by 45%.

The detailed information can be found in the paper [15].

Model and Parameter Selection

Research on statistical modeling methods has been actively conducted, however, the usefulness of the methods has usually been evaluated by using a single dataset in most of the research, and the robustness of the methods have not

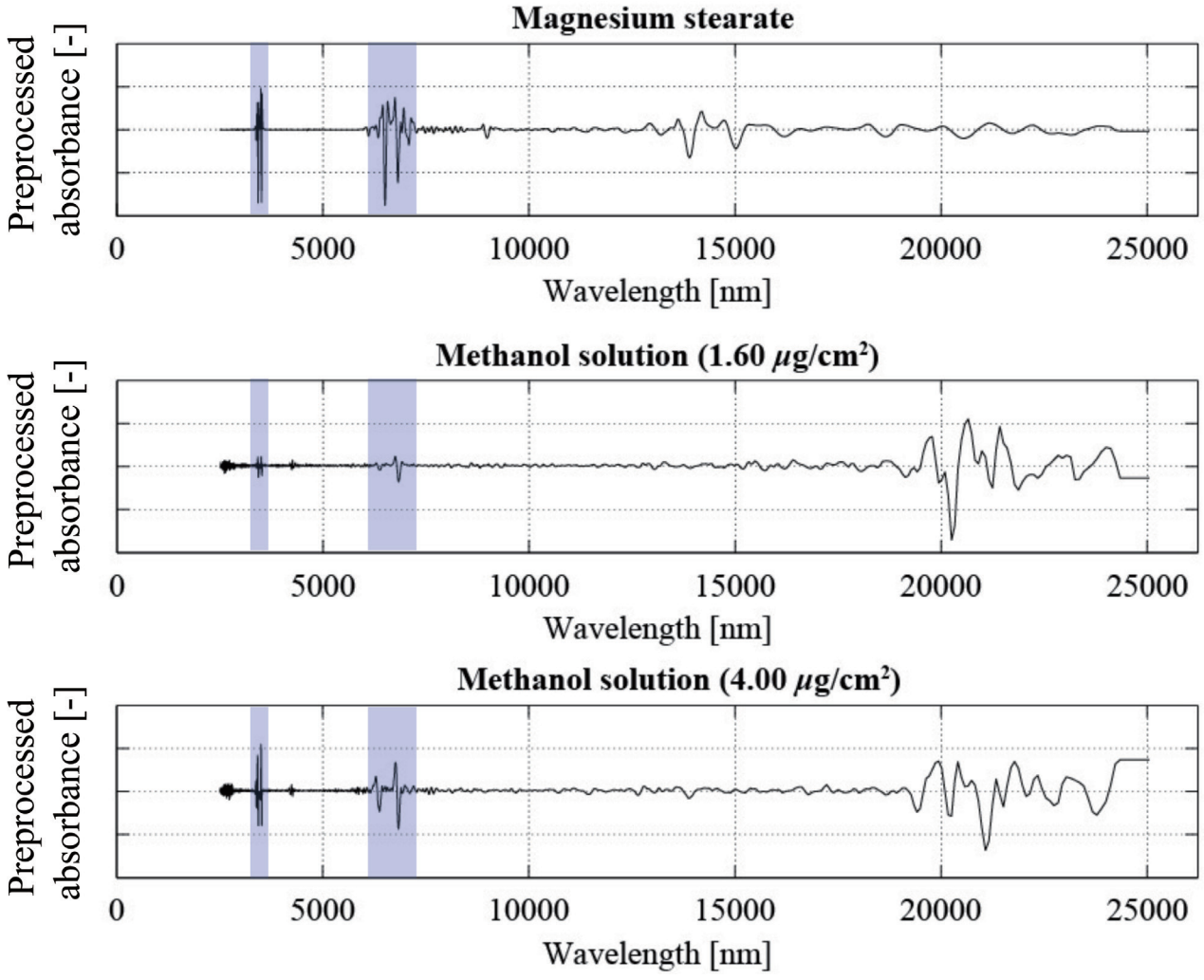

Figure 3 Preprocessed absorbance of magnesium stearate and its methanol solution

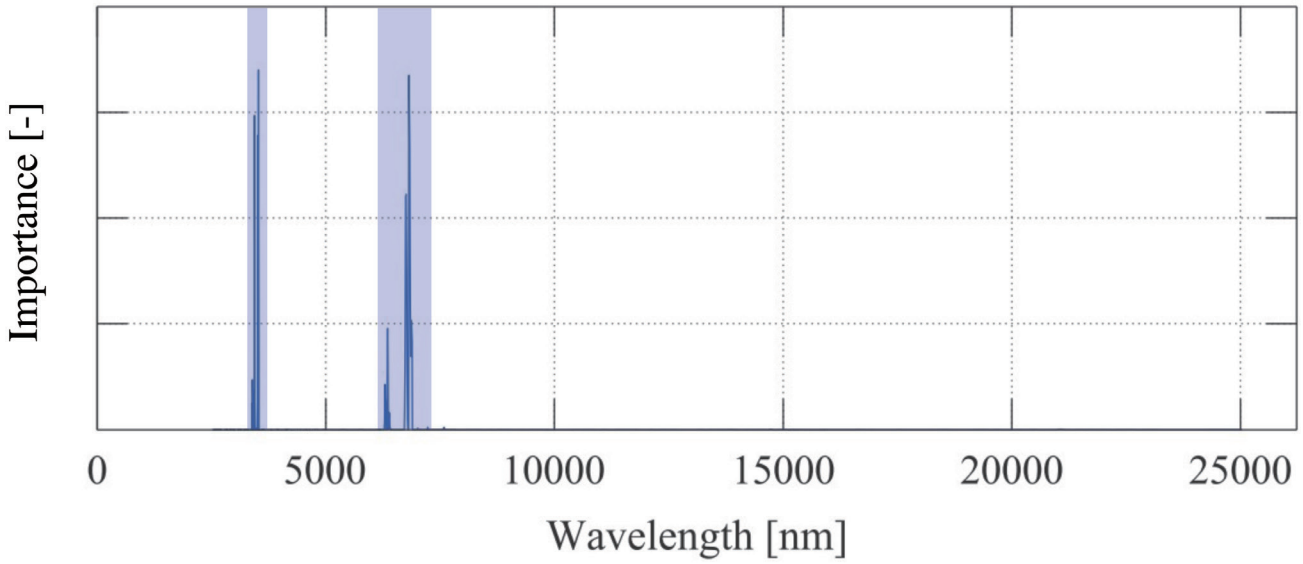

Figure 4 Importance evaluation result of the proposed method

been evaluated. To propose a highly reliable method, it is necessary to evaluate whether the developed soft-sensor can stably estimate the difficult-to-measure variables for various processes and a wide range of operating conditions. In this research, partial least squares (PLS), locally weighted PLS, support vector regression (SVR), and artificial neural network (ANN), which are widely used for soft-sensor design, are compared using twelve real-world datasets. The result of comprehensive comparative study in 6 datasets is show in Figure 5. In Figure 5, p denotes the ratio of the number of samples used to model construction to the number of all the samples in each dataset. The result showed that locally weighted PLS outperformed the other methods. At the same time, the drawback of locally weighted PLS has also been made clear. locally weighted PLS is a modified version of PLS which is a linear regression method, and locally weighted PLS can deal with nonlinearity. However, while the conventional parameter tuning method can adapt to the nonlinearity of the target process, it has also a risk of excessively adapting to variations in measured values due to the influence of noise. Thus, a new parameter tuning method is proposed to improve the accuracy while guaranteeing the minimum prediction accuracy, and showed that the prediction accuracy of active ingredient and impurity concentrations can be improved.

The detailed information can be found in the paper [16].

Conclusion

In this research, real-time monitoring methods of continuous pharmaceutical processes were investigated. Some of the developed technologies have already been put to practical use and have achieved social contribution. In addition to the content of the former section, the research on measurement and automation using data science and process control technology for many processes in the pharmaceutical production process, including the granulation process, drying process, and tableting process have been conducted. This research is expected to make a significant contribution to society, as it is planned to be put to practical use by our joint research partners.

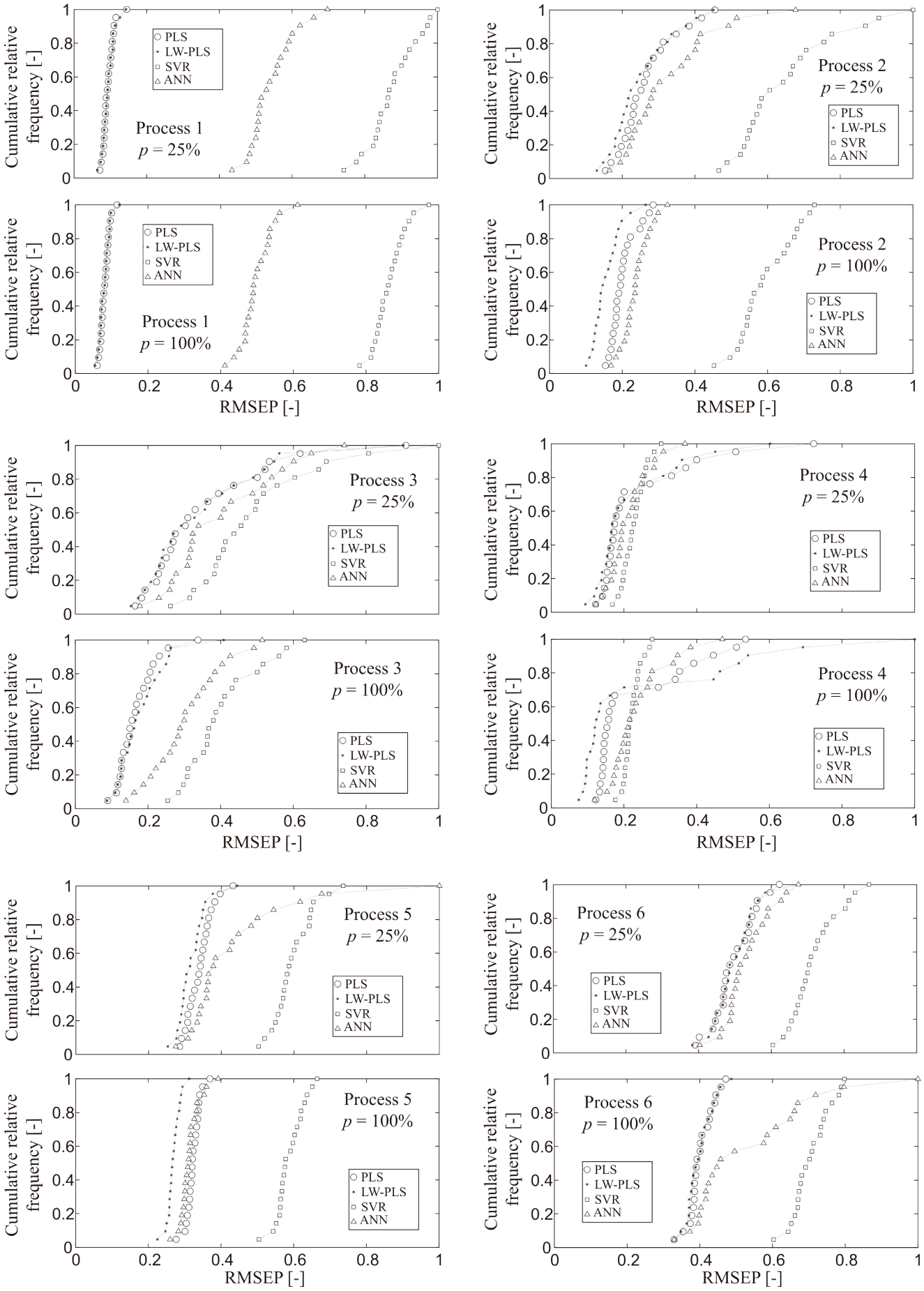

Figure 5 Result of comprehensive comparative study

* Editorial note: This content is based on HORIBA's investigation at the year of issue unless otherwise stated.

References

[1] FDA 2004. Pharmaceutical cGMPs for the 21st century—a risk-based approach final report.

[2] ICH 2005. ICH harmonised tripartite guideline—pharmaceutical development Q8 (R2).

[3] ICH 2005. ICH harmonised tripartite guideline—quality risk management Q9.

[4] ICH 2008. ICH harmonised tripartite guideline—pharmaceutical quality system Q10.

[5] Moes, J.J., Ruijken, M.M., Gout, E., Frijlink, H.W., Ugwoke, M.I., 2008. Application of process analytical technology in tablet process development using NIR spectroscopy: blend uniformity, content uniformity and coating thickness measurements. Int. J. Pharm. 357, 108–118.

[6] Berthiaux, H., Mosorov, V., Tomczak, L., Gatumel, C., Demeyre, J., 2006. Principal component analysis for characterising homogeneity in powder mixing using image processing techniques. Chem. Eng. Process. 45, 397–403.

[7] Virtanen, S., Antikainen, O., Yliruusi, J., 2007. Uniformity of poorly miscible powders determined by near infrared spectroscopy. Int. J. Pharm. 345, 108–115.

[8] Wu, H., Tawakkul, M., White, M., Khan, M., 2009. Quality-by-design (QbD): an integrated multivariate approach for the component quantification in powder blends. Int. J. Pharm. 372, 39–48.

[9] Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 2001. Multi- and Megavariate Data Analysis. Principles and Applications. Umetrics Academy.

[10] Kim, S., Kano, M., Nakagawa, H., Hasebe S., 2011. Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection, Int. J. Pharm., 421, 269–274

[11] Engel J, Gerretzen J, Szyma´nka E, Jansen JJ, Downey G, Blanchet L, et al. 2013. Breaking with trends in pre-processing? Trends Anal. Chem. 50, 96–106.

[12] van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ., 2006. Centering, scaling, and transformations improving the biological information content of metabolomics data. BMC Genomics, 7, 142.

[13] Todeschini R, Consonni V, Maiocchi A., 1999/ The k correlation index theory development and its application in chemometrics. Chemom Intell Lab Syst 46, 13–29.

[14] Keun HC, Ebbels TMD, Antti H, Bollard ME, Beckonert O, Holmes E, et al., 2003. Improved analysis of multivariate data by variable stability scaling application to NMR-based metabolic profiling. Anal Chim Acta, 490 265–76.

[15] Kim, S., Kano, M., Nakagawa, H., Hasebe S., 2015. Input variable scaling for statistical modeling, Comput. & Chem. Eng., 74, 4, 59–65

[16] Matsuyama, Y., Kim, S., Hasebe S., 2021 Robust parameter tuning method of LW-PLS and verification of its effectiveness by twelve industrial processes", Comput. & Chem. Eng. 146 107224

Dr. Sanghong Kim

金 尚弘

Associate Professor
Department of Applied Physics and Chemical Engineering,
Tokyo University of Agriculture and Technology.