

Development of Near-Infrared Spectroscopy Module for Process Analytical Technology

Daisuke Oida, Noriaki Yamasaki, Masaru Koshihara, Takanori Saitoh

[Summary]

Due to chronic pharmaceutical supply shortages, continuous manufacturing is being explored as a means of achieving stable, high-quality drug production. We developed a near-infrared (NIR) spectroscopy module for process analytical technology (PAT) applications supporting continuous manufacturing. This article first describes the module's fundamental architecture, key features, and data analysis methods. We then evaluate module's performance by using pharmaceutical excipients to demonstrate that the module achieves high-accuracy estimates of concentration and is suitable for deployment in continuous manufacturing processes.

1 Introduction

Pharmaceutical supply shortages have persisted since around 2020, triggered by a contamination scandal in which numerous patients suffered adverse health effects. Investigations revealed systemic failures in quality control, including non-compliance with approved manufacturing procedures and falsification of quality test records. In response, Good Manufacturing Practice (GMP) regulations governing pharmaceutical production and quality management were revised. As a result, quality control deficiencies were identified at other pharmaceutical manufacturers, and subsequent business suspension orders led to widespread shipment halts across many drug products. Compounding this, the COVID-19 pandemic that emerged the same year drove a sharp surge in demand for certain medications, including cold remedies, further exacerbating the supply shortages. Supply shortages have continued during resurgences of COVID-19 and influenza. Against this backdrop, establishing a manufacturing framework capable of stably supplying high-quality pharmaceuticals has become an urgent priority.

Pharmaceuticals are manufactured by the following sequential processes. First, the active ingredient is precisely weighed (weighing step). It is then mixed with a liquid such as water, granulated, and dried (granulation step). Next, excipients and other additives are blended in (mixing step), and finally the mixture is compressed into tablet form (tableting step). At each step, all material is removed from the manufacturing equipment and inspected for quality before proceeding to the next stage. If the material passes inspection, it is transferred to the equipment for the

subsequent step. This currently predominant approach is known as batch manufacturing. This approach has notable drawbacks because material transfer and inspection between steps are performed manually, making the process labor-intensive and introducing the risk of contamination.

As a solution, continuous manufacturing is being explored whereby each process step is linked directly to enable uninterrupted production. Performing all steps from weighing through to tableting in one continuous flow, is expected to cut costs, shorten production time, and prevent contamination. However, continuous manufacturing requires monitoring product quality inline rather than removing material for offline inspection. For example, moisture content is a critical parameter at the granulation step, while blend uniformity is key at the mixing step. The technologies used to measure these parameters inline are collectively called Process Analytical Technology (PAT)¹⁾. The most widely used PAT is near-infrared (NIR) spectroscopy²⁾⁻⁴⁾.

Near-infrared spectroscopy is an analytical technique that determines sample composition by analyzing the transmission or reflectance spectra of near-infrared light (wavelength range: 800 to 2500 nm) irradiated onto a sample. When NIR radiation is incident on a sample, molecular bonds predominantly found in water and organic compounds, such as O–H, N–H, and C–H groups, undergo vibrational excitation, resulting in absorption of the incident radiation. Since the degree of light absorption is dependent on the chemical composition of the sample, qualitative and quantitative compositional information can be extracted through spectral analysis.

In NIR spectroscopy, broadband white light is directed onto

the sample, and the transmitted or reflected light is dispersed before being detected by a photodetector to capture the spectrum. The white light sources used in NIR spectroscopy typically operate at relatively low irradiance, enabling non-destructive measurement of samples without causing photodegradation or structural alteration. Furthermore, the use of high-speed photodetectors allows rapid spectral acquisition, thereby facilitating real-time compositional analysis. Based on these characteristics, NIR spectroscopy can be regarded as an optimal inline inspection technology for PAT applications.

NIR spectroscopic instruments used in PAT applications are integrated directly into existing pharmaceutical manufacturing lines. Consequently, a compact form factor is required to accommodate line space constraints. In addition, lightweight construction is necessary to facilitate easy handling during maintenance. Furthermore, at blending processes where the instrument itself rotates along with the blending vessel, high vibration resistance is an essential requirement.

We have developed an NIR spectroscopy module that satisfies the requirements of PAT applications and can be readily integrated into pharmaceutical manufacturing lines. This article describes the configuration and key features of the developed NIR spectroscopy module. Furthermore, the spectral acquisition methodology and data analysis procedures are discussed in detail. Finally, by presenting the results of compositional analysis on samples prepared by blending pharmaceutical excipients used in actual drug formulations, we demonstrate the applicability of the developed module to quantitative analysis of pharmaceutical ingredients.

2 Near-Infrared Spectroscopy Module

2.1 External and Internal Structure

Figure 1 shows the external appearance of the developed NIR spectroscopy module. The module is cylindrical with a diameter of 44.6 mm and a length of 111 mm, a mass of about 550 g, and is constructed from stainless steel. The upper surface has an M25 threaded fitting for direct mounting onto pharmaceutical manufacturing equipment.

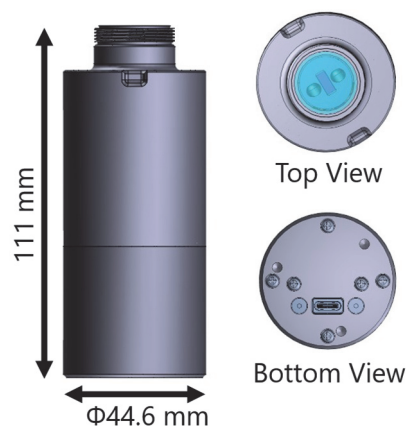


Figure 1 External Appearance of the NIR Spectroscopy Module

Figure 2 and Table 1 present the internal configuration diagram and specifications of the NIR spectroscopy module, respectively. The module primarily consists of two tungsten lamps for NIR illumination of the sample, a Linear Variable Filter (LVF) for dispersing the reflected light from the sample, and a line sensor that converts the dispersed light into electrical signals to capture the spectrum. Tungsten lamps are used in NIR spectroscopy due to their broad output wavelength coverage across the NIR region, low cost, and stable output power.

The LVF is an optical filter used for light dispersion, and Figure 3 shows its characteristics. The wavelength of light transmitted through the LVF is linearly related to the position along the filter. Consequently, the spectral intensity distribution, or the light intensity as a function of wavelength, can be measured by positioning a line sensor directly beneath the LVF. The module design can be simplified and made compact by exploiting the spectral characteristics of the LVF. Furthermore, unlike FT-NIR (Fourier Transform Near-Infrared) spectroscopy, which requires moving components to acquire spectra via an interferometer, the LVF-based design contains no moving parts, thereby providing superior vibration resistance.

The wavelength resolution of the LVF is defined as the FWHM (Full Width at Half Maximum) of the transmitted intensity profile when monochromatic light is incident on the filter. The wavelength resolution of the LVF in the present module is 2% of the central wavelength. For example, when monochromatic light at 1500 nm is incident on the filter, the FWHM of the transmitted spectrum is 30 nm. An InGaAs sensor, which exhibits high sensitivity in the NIR region, is used as the line sensor, with a pixel count of 256.

The module can be powered and controlled by connecting it to a PC via the USB port on the underside. The user can adjust the spectral acquisition timing and configure parameters such as the integration time (6 μ s or greater) and the on/off state of the tungsten lamp by using the bundled spectral capture software.

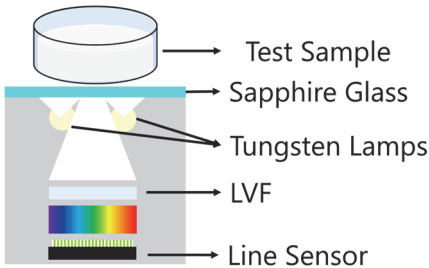


Figure 2 Near-Infrared Spectroscopy Module Internal Structure

Table 1 Near-Infrared Spectroscopy Module Specifications

Model	Standard Wavelength	Long Wavelength
Wavelength Range	1000 to 1700 nm	1250 to 2100 nm
Wavelength Resolution	2% of wavelength (30 nm @ 1500 nm)	
Pixel Count	256	
Integration Time	>6 μ s	
Max. Connections	8 per PC (limited by PC power supply capacity)	
Measurement Speed	>1 kHz (with 1 connection, 100 μ s integration time)	
Mass	<550 g	
Dimensions	$\phi 44.6 \times 111$ mm	
Material	SUS316	
Protection Class	IP67-equivalent	
I/F and Power Supply	USB3.0	
Power Consumption	2.5 W (5 V, 500 mA)	
Ambient Temperature Range	0 to 40°C	
Supported OS	Windows 10, 11	

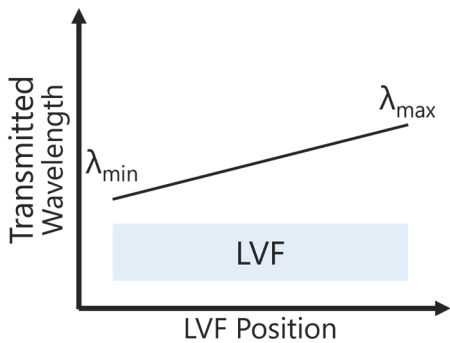


Figure 3 LVF Characteristics

2.2 Features

There are two NIR spectroscopy modules with different wavelength ranges. The first standard wavelength model covers a wavelength range of 1000 to 1700 nm, which is commonly used in NIR spectroscopy. The second long-wavelength model operates over an extended wavelength range of 1250 to 2100 nm, shifted toward longer wavelengths relative to the standard model. The long-wavelength model is better suited for moisture content measurement compared to the standard wavelength model. The primary absorption peaks of water molecules in the NIR region are at about 1450 and 1940 nm. However, the peak at about 1940 nm exhibits significantly greater absorbance. Consequently, the long-wavelength model, which encompasses this higher-absorbance peak, determines moisture content with higher accuracy.

In addition to the different wavelength ranges, both modules are designed to be compact and lightweight for easy integration into pharmaceutical manufacturing lines, while also offering superior vibration resistance. Furthermore, the spectral acquisition rate of 1 kHz enables real-time compositional estimation. Moreover, since up to eight modules can be controlled from one PC, multiple manufacturing lines can be monitored simultaneously from one workstation.

3 Applications in Pharmaceutical Production Lines

This section explains how the NIR spectroscopy module with the described features is used in pharmaceutical manufacturing lines. Figure 4 shows typical application examples in granulation and blending processes.

At granulation, a liquid such as water is mixed with the active pharmaceutical ingredient to form uniform granules. To this end, the moisture content is measured using the NIR spectroscopy module, and the mixing duration and liquid discharge volume are controlled based on the measured values, maintaining a constant moisture content in the granules.

Similarly, in the subsequent blending process, the NIR spectroscopy module is used to measure the concentration of the active pharmaceutical ingredient and excipients, and the mixing duration and excipient quantities are controlled based on the measured values, thereby enabling production

of pharmaceuticals with uniform composition. In this manner, use of the NIR spectroscopy module makes it possible to maintain quality control at each manufacturing stage.

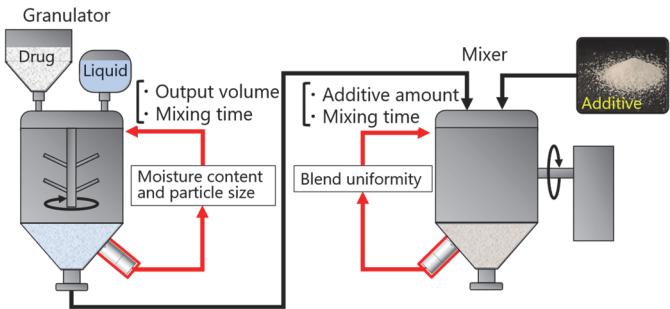


Figure 4 Module Use in Granulation and Mixing

4 Spectral Measurement and Analysis

Figure 5 shows the workflow for spectral measurement and analysis. First, three types of spectra are captured using the near-infrared spectroscopy module: the sample spectra (Figure 5(a)), the reference spectrum (Figure 5(b)), and the offset spectrum (Figure 5(c)). Subsequently, the spectral analysis stage performs two processing steps on the captured spectra: preprocessing (Figure 5(d)) and component estimation (Figure 5(e)). The preprocessing step reduces spectral noise and enhances peaks, thereby improving the accuracy of the downstream component estimation. Component estimation quantifies sample constituents by

constructing a calibration model using Partial Least Squares (PLS)⁵⁾ regression.

4.1 Spectral Measurement

The sample spectra are captured by measuring multiple samples with varying compositions. The reference spectrum is captured by measuring a standard reflector characterized by high reflectance and low wavelength dependence in the near-infrared region. The offset spectrum is captured without a sample.

The integration time must be carefully considered during spectral measurement. An excessively short integration time causes spectral peaks to be buried in noise, while an excessively long integration time results in sensor saturation. Since spectral intensity also depends on the sample type and distance between the near-infrared spectroscopy module and the sample, the integration time must be configured appropriately for the given measurement conditions.

4.2 Spectrum Analysis

4.2.1 Preprocessing

Preprocessing is performed using the spectra captured as described in the preceding section. The preprocessing pipeline applies four sequential operations to the spectra: temporal averaging, reflectance conversion, smoothing and differentiation, and normalization. Figure 6 shows the resulting spectra after each processing step.

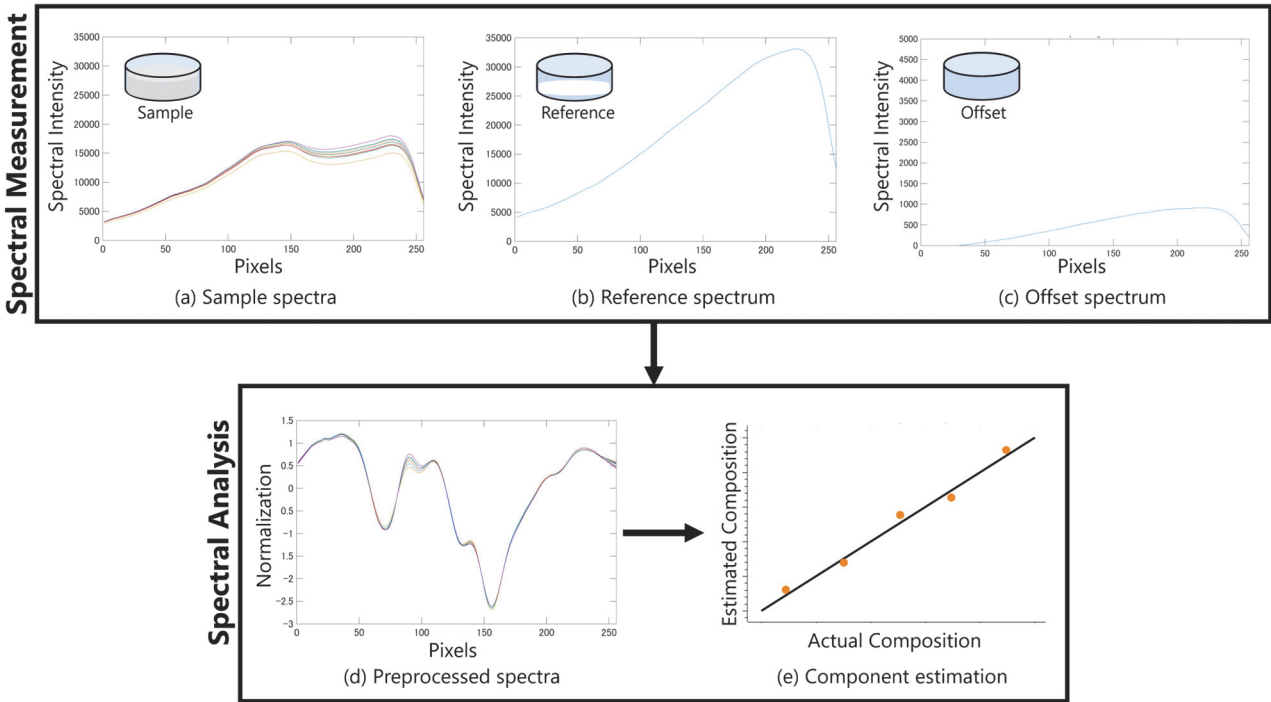


Figure 5 Spectral Measurement and Analysis Flow Sequence

First, multiple captured spectra are averaged (Figure 6(a)), which reduces random noise attributable to factors such as sensor sensitivity.

Next, conversion to reflectance R (Figure 6(b)) is performed using Eq. (1),

$$R = \frac{I_S - I_O}{I_R - I_O} \quad (1)$$

where, I_S , I_R , and I_O are the sample spectrum, reference spectrum, and offset spectrum, respectively. Converting the spectra to reflectance compensates for spectral variability arising from unit-to-unit variation in the light source and sensor.

Next, smoothing and differentiation are applied (Figure 6(c)). Smoothing reduces noise, while differentiation extracts spectral peaks. Both operations can be performed simultaneously using the Savitzky–Golay method⁽⁶⁾, in which a polynomial is fitted within a defined window and the central value of the fitted polynomial replaces the data within that window. Smoothing is achieved over the full spectrum by applying this operation repeatedly while sliding the window across the entire spectral range. Furthermore, differentiation can also be performed by analytically differentiating the fitted polynomial.

Finally, normalization is applied (Figure 6(d)), which compensates for spectral drift caused by factors such as temperature variation. The normalized spectrum X is computed using Eq. (2),

$$X = \frac{I - \mu}{\sigma} \quad (2)$$

where, I is the differentiated spectral data, μ is the mean, and σ is the standard deviation. All spectral data are standardized to zero mean and unit standard deviation using normalization. As shown in the enlarged insert in Figure 6(d), preprocessing enhances peak variations attributable to differences in sample composition.

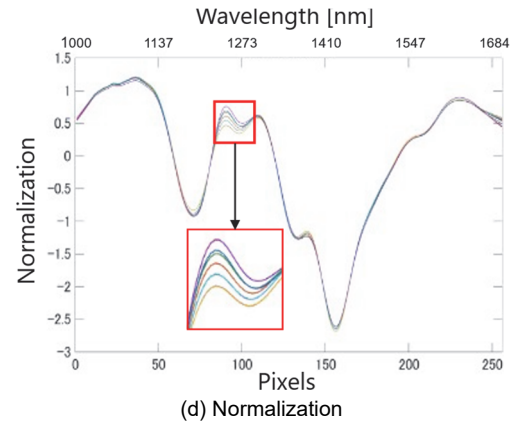
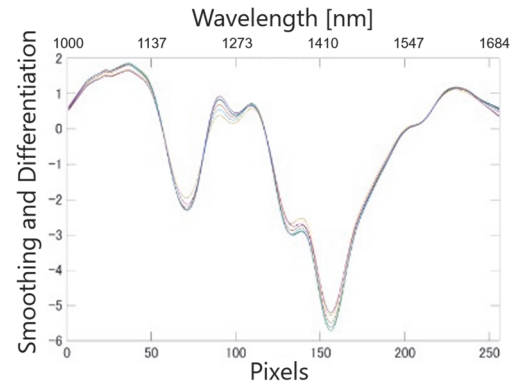
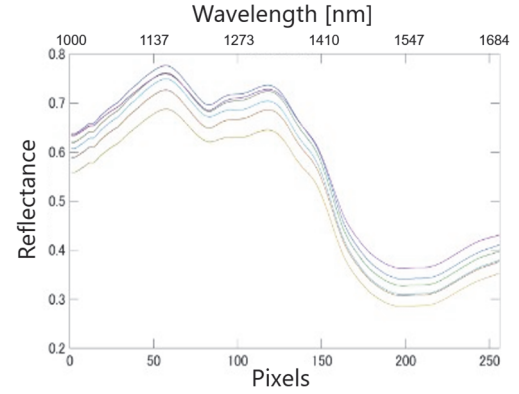
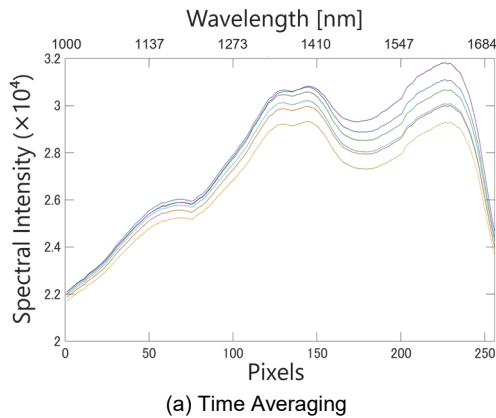


Figure 6 Preprocessing Spectral Conversion

4.2.2 Composition Estimation

At component estimation, calibration data is created from the preprocessed spectral data. The calibration data B satisfies the following multiple regression equation,

$$y = XB + D \quad (3)$$

where, y is the component matrix, X is the standardized spectral data obtained from Eq. (2), and D is the residual. The calibration data B can be analytically determined under the condition that the residual D is minimized. However, since the spectra captured by near-infrared spectroscopy exhibit multicollinearity with high correlation between wavelengths, component estimation becomes unstable with calibration data based on Eq. (3).

Therefore, we used PLS, which creates new variables (latent variables) from the preprocessed data that contribute to component estimation. This method stabilizes component estimation.

Using PLS, $\mathbf{X}$ and $\mathbf{y}$ in Eq. (3) are expressed by the following equations,

$$\mathbf{X} = \sum_{n=1}^N \mathbf{t}_n \mathbf{P}_n^T + \mathbf{E} = \mathbf{T} \mathbf{P}^T + \mathbf{E} \quad (4)$$

$$\mathbf{y} = \sum_{n=1}^N \mathbf{t}_n \mathbf{q}_n + \mathbf{f} = \mathbf{T} \mathbf{q} + \mathbf{f} \quad (5)$$

where, N is the number of principal components, $\mathbf{t}_n$ is the n -th principal component, $\mathbf{T}$ is the latent variable, $\mathbf{P}$ and $\mathbf{q}$ are the loading matrices of $\mathbf{X}$ and $\mathbf{y}$, respectively, and $\mathbf{E}$ and $\mathbf{f}$ are the residuals of $\mathbf{X}$ and $\mathbf{y}$, respectively. Also, the superscript T denotes transpose of a matrix. The latent variable $\mathbf{T}$ is determined so that the covariance with the component matrix $\mathbf{y}$ is maximized as,

$$\mathbf{T} = \frac{\mathbf{X} \mathbf{X}^T \mathbf{y}}{|\mathbf{X}^T \mathbf{y}|} = \mathbf{X} \mathbf{W} \quad (6)$$

where, $\mathbf{W}$ is the weight vector. $\mathbf{P}$ and $\mathbf{q}$ are analytically determined under the condition that $\mathbf{E}$ and $\mathbf{f}$ in Eq. (4) and (5) are minimized. Using these, the calibration data $\mathbf{B}$ when using latent variables is expressed as,

$$\mathbf{B} = \mathbf{W} (\mathbf{P}^T \mathbf{W})^{-1} \mathbf{q} \quad (7)$$

Figure 7 shows the calibration data created using the preprocessed spectra from Figure 6. The component can be estimated by calculating the inner product of this calibration data and the preprocessed spectra.

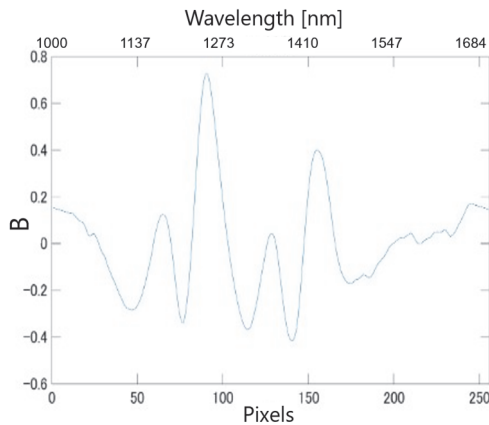


Figure 7 Calibration Data

4.2.3 Evaluation of Calibration Data

Before applying the created calibration data to an actual pharmaceutical line, it is necessary to evaluate the component estimation accuracy. Therefore, the preprocessed

spectra are split into training data and test data. Next, calibration data is created using the training data, and the component estimation accuracy is evaluated. Finally, the test data component estimation accuracy is evaluated using the same calibration data.

There are two indicators for evaluating the component estimation accuracy of the calibration data. The first is the coefficient of determination R^2 , which is calculated by the following equation,

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

where, n is the number of samples, y_i is the actual component, y'_i is the estimated component, and $\bar{y}$ is the mean value of the actual components. The closer the coefficient of determination is to 1, the higher the estimation accuracy. Figure 8 shows the change in the coefficient of determination with respect to the number of principal components. It is evident that increasing the number of principal components brings the coefficient of determination closer to 1, but caution is needed because too many principal components cause overfitting, where accuracy improves for training data while accuracy decreases for test data. In practice, parameters (such as the number of principal components and the order of differentiation) are adjusted so that the coefficient of determination approaches 1 for both the training data and the test data.

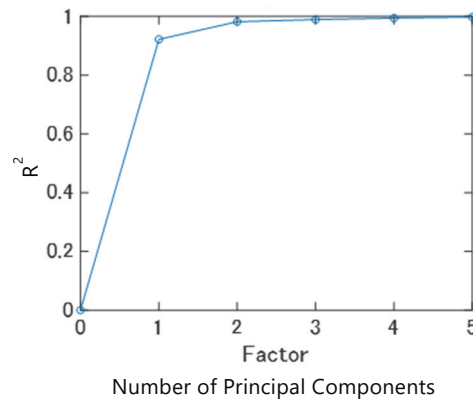


Figure 8 Change in Coefficient of Determination R^2 with Number of Principal Components

Another indicator is the Root Mean Square Error (RMSE), which is calculated by the following equation,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (9)$$

where, n , y_i , and y'_i are the same as defined above. RMSE

is an indicator showing how much error the estimated concentration has relative to the actual concentration.

5 Evaluation of Component Estimation Accuracy

To evaluate the component estimation accuracy of the developed module, samples were prepared by mixing three types of additives, and concentrations were estimated. In addition, an automated verification device was fabricated to efficiently measure all samples.

5.1 Samples

The three additives were lactose, Ceolus (Asahi Kasei Corporation), and calcium stearate, which are widely used in pharmaceuticals. In this evaluation, the calcium stearate concentration was estimated. For this purpose, ten samples with different concentrations (0%, 1%, 2%, 3%, 3.5%, 4%, 5%, 6%, 8%, and 10%) of calcium stearate were prepared. The concentrations of the other two additives were kept constant across all samples so as not to affect the measurements. The samples were transferred to Petri dishes at a thickness of approximately 1 cm.

5.2 Automated Verification Device

Figure 9 shows the appearance of the automated verification device used to measure the samples. The rotating table at the top of the device has 12 equally spaced holes. A total of 12 Petri dishes is placed above these holes: 10 containing samples, one containing a standard reflector for reference spectrum measurement, and one empty dish for offset spectrum measurement. The developed near-infrared spectroscopy module is mounted under the rotating table to measure spectra from the bottom surface of the Petri dishes. Since the automated verification device has a built-in rotary stage, the 12 samples were measured continuously.

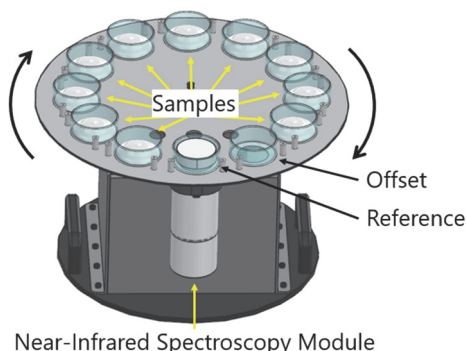


Figure 9 External View of Automated Verification Device

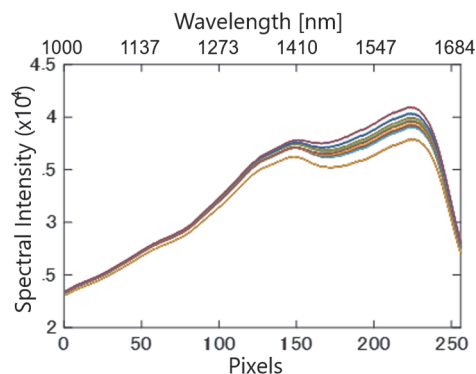
5.3 Measurement Conditions

Both the standard wavelength and long-wavelength

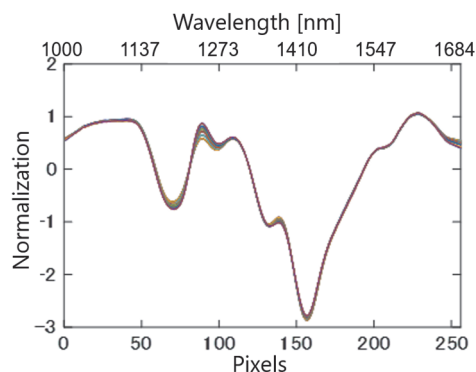
models were used for the measurements. The number of spectra measurements was set to 800 per sample. The integration time was set to 1280 μs for the standard wavelength model and 1270 μs for the long-wavelength model. These values were set so that the maximum value of the reference spectrum would be 80% of the maximum intensity of the sensor.

In addition, to determine any spectral variation due to sample position, spectra were measured over a wide sample area while rotating the stage. The number of time averages per sample was set to 80. As a result, the number of data points after time averaging was 10 per sample. Furthermore, the smoothing window size was set to 13 and the order of differentiation was set to 1.

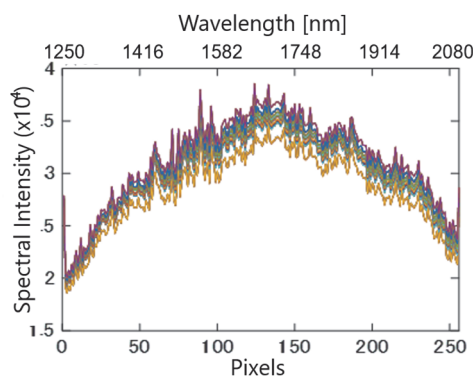
Six samples of 0%, 2%, 4%, 6%, 8%, and 10% were used as training data, and four samples of 1%, 3%, 3.5%, and 5% were used as test data. Figure 10 shows raw and preprocessed spectra measured by both models. The raw spectra of the long-wavelength model in Figure 10(c) show large intensity variations between pixels due to the sensitivity characteristics of the sensor, but the influence of these sensitivity characteristics is removed by preprocessing, as shown in Figure 10(d).



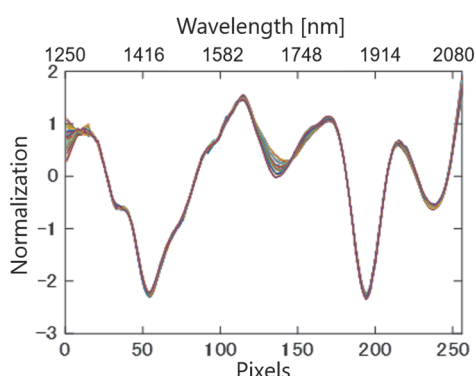
(a) Raw Spectra (Standard Wavelength Model)



(b) Preprocessed Spectra (Standard Wavelength Model)



(c) Raw Spectra (Long-Wavelength Model)



(d) Preprocessed Spectra (Long-Wavelength Model)

Figure 10 Raw and Preprocessed Spectra Measured by Standard and Long-Wavelength Models

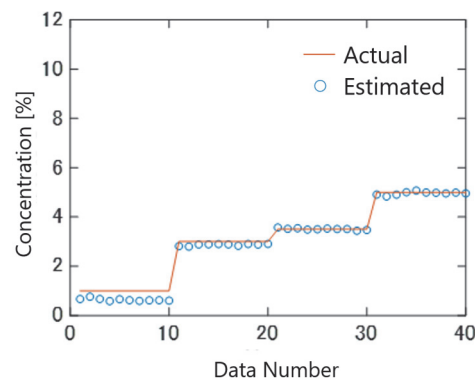
5.4 Concentration Estimation Results

Figure 11 shows the estimated calcium stearate concentration results. Figures 11(a) and 11(b) show the results for the standard and long-wavelength models, respectively. The y-axis is concentration, the x-axis is the data number after time averaging; the orange line indicates actual concentration, and the blue circles indicate estimated concentration. The concentration errors (RMSE) for the standard wavelength model and the long-wavelength model were 0.20% and 0.24%, respectively, demonstrating high estimated concentration accuracy. In addition, the estimation speed was several times per second, facilitating real-time concentration estimation.

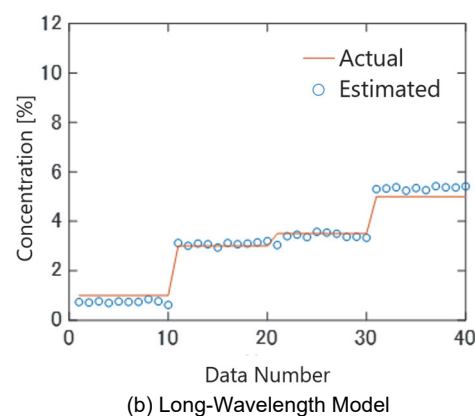
6 Conclusions

This study focused on near-infrared spectroscopy among PAT technologies to develop a near-infrared spectroscopy module that can be easily integrated into pharmaceutical manufacturing lines. A compact and lightweight module with high resistance to vibration was achieved by using an LVF for spectroscopy. Evaluation using common pharmaceutical additives demonstrated that the developed module has high

estimated concentration accuracy and can estimate concentration in real-time. These results show the developed module has practical applications in pharmaceutical component analysis on continuous pharmaceutical lines.



(a) Standard Wavelength Model



(b) Long-Wavelength Model

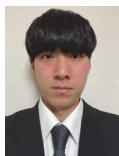
Figure 11 Estimated Calcium Stearate Concentration Results

References

- 1) Wu S, Liu Y, Fan X, Shen Y, Qu H. Trends and new process analytical technologies in pharmaceutical manufacturing. *Int J Pharm.* 2025 Sep 15;682:125957.
- 2) Yukihiro Ozaki, "Near-Infrared Spectroscopy," Spectroscopy Series, vol. 2, Kodansha, 2015 (in Japanese).
- 3) Sierra-Vega NO, Karry KM, Románach RJ, Méndez R. Monitoring of high-load dose formulations based on co-processed and non co-processed excipients. *Int J Pharm.* 2021 Sep 5;606:120910.
- 4) Kiricenko K, Klinken S, Kleinebudde P. Proof of a LOD prediction model with orthogonal PAT methods in continuous wet granulation and drying. *J Pharm Sci.* 2025 Jan;114(1):176-184.
- 5) Rosipal, R., Krämer, N. (2006). Overview and Recent Advances in Partial Least Squares. In: Saunders, C., Grobelnik, M., Gunn, S., Shawe-Taylor, J. (eds) Subspace, Latent Structure and Feature Selection. SLSFS 2005. Lecture Notes in Computer Science, vol 3940. Springer, Berlin, Heidelberg.

- 6) Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639.

Authors



Daisuke Oida
1st Development Dept.
Development Div.
Sensing & Devices Company



Noriaki Yamasaki
1st Development Dept.
Development Div.
Sensing & Devices Company



Masaru Koshihara
1st Development Dept.
Development Div.
Sensing & Devices Company



Takanori Saitoh
1st Development Dept.
Development Div.
Sensing & Devices Company