薬 剤 学, **78** (2), 67–71 (2018)

≪若手研究者紹介≫

近赤外分光法による製剤品質管理に向けた情報処理技術の活用

宮 野 拓 也* Takuya Miyano
第一三共株式会社 バイオマーカー推進部

1. は じ め に

進路選択を控えた高校 3 年生のある日，薬の研究をしようと思い立った．

その理由の一つは，世の人を病から守るという大きな取り組みに参画したいと考えたからである．一人の人が病によって健康な生活が送れなくなると，その人が健康であったなら享受できたであろう沢山の幸せを得る基盤を損なってしまう．さらに，その人の周囲にいる家族や友人もまた，大きな悲しみを背負うことになる．当時，そうした光景を目の当たりにして，ただ口惜しいと思った．薬の研究は，目の前の病に対して無力であった私に，これから何ができるのだろうと考えた一つの結論である．

もう一つの理由は，ヒトの生体中の情報に作用する“薬”という存在への興味である．幼少期の私は，初めて触れたパソコン（1993 年，OS は Windows 3.1 であった）で，ごく簡単なプログラミングをしながら遊んでいた．その中で「設定値を一つ変えるだけで，パソコン上の世界を変えられる」ことの楽しさを知った．それと同じ感覚でヒトの生体を眺めたとき「薬を一つ投与するだけで，ヒトの生体内の世界を変えられる」ことが，単純に面白そうに思えた．どちらも，その世界は膨大な情報に基づいていて，適切な情報処理技術があれば世界を変えられるのだろうと．

薬の研究に対する意志は大学卒業後も変わらず，第一三共株式会社に入社した．配属された製剤技術研究所で私に与えられた業務は，固形製剤の製造プロセス管理に用いられる Process analytical technology（PAT）の開発，特に近赤外分光法（near-infrared spectroscopy：NIRS）の解析であった．NIRS の解析は，私が興味を持つ「情報処理」に合致した．

NIRS に関する研究として，NIRS の波数選択法[1,2]及び検量モデルメンテナンスの運用化[3]を検討したが，本稿では波数選択法の詳細を述べる．

2. 近赤外分光法（NIRS）

NIRS は，試験サンプルを溶解する等の前処理が不要及び迅速かつ非破壊の測定が可能であるため，PAT の代表的な手段として利用されている．近赤外（near-infrared：NIR）スペクトルは様々な化学的及び物理的特性を反映するため，各種の物質特性（e.g., 顆粒中の薬物含量や水分含量）を推定できる．しかし，それは同時に NIR スペクトルが様々な物質特性の情報が入り混じったものであることを意味する．

3. NIRS の検量モデル

NIR スペクトルから目的とする物質特性を推定するためには，NIR スペクトルと物質特性との関係を表す検量モデルが必要となる．この検量モデルがあって初めて NIR スペクトルから物質特性を推定することができる．すると，検量モデルの推定精度が品

*2009 年 3 月京都大学薬学研究科修士課程修了．同年 4 月第一三共株式会社入社，製剤技術研究所にて固形製剤研究に従事．2017 年 4 月同社バイオマーカー推進部に異動．2016 年 3 月東京薬科大学大学院より博士（薬学）学位を授与．2014 年パーティクルデザイン賞受賞．趣味：ピアノ．連絡先：〒140–8710 東京都品川区広町 1–2–58 E-mail: miyano.takuya.iu@daiichisankyo.co.jp

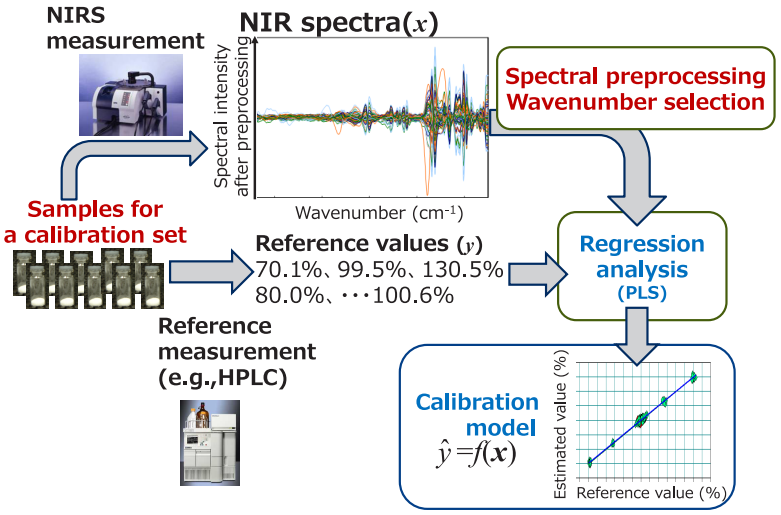

Fig. 1.　Workflow to develop a calibration model.

質管理の精度に直接影響を与えることになる．

検量モデルを構築する流れを Fig. 1 に図示する．まずキャリブレーションセット用のサンプルを選択する．これらのサンプルを NIRS とその参照法（e.g., 高速液体クロマトグラフィー）で測定し，NIR スペクトルと参照値を得る．次に，NIR スペクトルに対して，スペクトル前処理をかけてノイズを除去する．そして，NIR スペクトルのうちどの波数を使用するかを選択する．最後に，こうして得られたスペクトル情報と参照値を用いて回帰分析を実施し，検量モデルを作製する．回帰分析法としては，多重共線性に対応可能な partial least squares（PLS）が一般的に用いられる．

高い推定精度を有する検量モデルを構築するために，キャリブレーションセット，スペクトル前処理，及び波数をいかに適切に選択するか．これが検量モデルを構築する担当者に与えられる課題である．ここで特に困難と私が感じたのは波数選択であった．

4. 波数選択法

波数選択を実施するための方策を考えよう．

まず波数の全組合せは膨大であり，それを網羅的に評価することは非現実的である．後述の事例研究で使用した NIR スペクトルは 2,202 点の波数点で構成されるため，2^{2202} 通りの波数組合せが存在する．これを全て評価すると 10^{658} 年の計算時間がかかる．

分光学的知識を頼りにすれば，より効率的に波数を評価できるものの，体系的な方法とは言い難い．というのも，NIRS では複数の官能基に基づく吸収

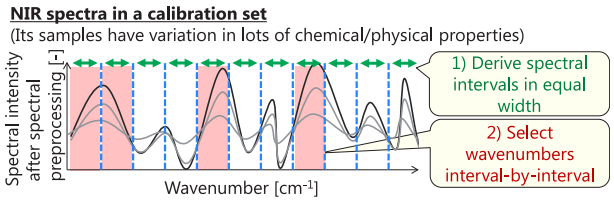

Fig. 2.　An illustrative example of iPLS.

帯が複雑に重なり合っているため，その物質特性の変動を適切に反映する波数を分光学的知識のみに基づいて特定することは困難である．それ故，高い推定精度が得られる波数組合せを見つけるために，試行錯誤で波数の組合せ評価を幾度も繰り返すことになる．これは作業工数が大きく，しかも適切な波数が選択できるか否かが個人の能力や運に依存する．

情報処理的な手法を用いれば，より体系的に波数を評価できる．これの目的は，高い推定精度を得る波数組合せを短い計算時間で選択することである．ただし，従来法は計算負荷が非常に大きいため，NIR スペクトルのデータ量や計算機の性能によっては膨大な（時に数日や数年といった単位で）計算時間を要する．また，推定精度の観点でも従来法には改善の余地があると私は感じた．

4.1　従来法：interval PLS（iPLS）

代表的な従来法として iPLS を概説する．Fig. 2 に示すように，iPLS は全波数領域を等幅のスペクトル領域に分割し，そのスペクトル領域単位で波数を取捨選択する．こうしたスペクトル領域に着目した波数選択は有望な方策である．なぜなら，隣接する波数におけるスペクトル強度は強く相関することが

多く，それが特定のスペクトル領域における吸収ピークとして表れるためである．ただし，iPLS が用いる等幅のスペクトル領域では各吸収ピークを正確に捉えられるとは限らない．また，その幅を最適化するために試行錯誤の評価が必要となる．では，各吸収ピークをより正確かつ効率的に捉えるにはどうしたらよいだろうか．

4.2 提案法：Spectral fluctuation dividing-PLS（SFD-PLS）

私はまず，スペクトル全体を各吸収ピークに対応するスペクトル領域に分割することを意図した SFD-PLS を開発した．Fig. 3 に示すように，吸収ピークの端点を各波数のスペクトル強度の標準偏差が極小になる点として捉えることができると考えた．SFD-PLS は，その極小点においてスペクトル全体を複数のスペクトル領域に分割し，スペクトル領域単位で波数を選択する．

この SFD-PLS を実際の NIR スペクトルに適用したところ，2,202 個の波数点を約 200 個のスペクトル領域にまとめることができた．しかし，それでも約 200 個というスペクトル領域の数は膨大であり，本手法には改善の余地があると感じた．

4.3 改良した提案法：SFD-nearest correlation spectral clustering-PLS（SFD-NCSC-PLS）

そこで SFD-PLS を改良し，スペクトル領域をその関連性に基づいてグループ化することを意図した SFD-NCSC-PLS を開発した．NIRS では化学物質が有する一つの官能基が複数の吸収帯を有するため，その含量が変動すると複数の吸収ピークが同時に変動する．こうした複数のスペクトル領域をまとめて評価できれば，より効率的に波数を選択できるだろう．Fig. 4 に示すように SFD-NCSC-PLS は，SFD-PLS で作製したスペクトル領域につき，そのスペクトル強度の相関に基づいてスペクトル領域をクラスタリングし，数個のスペクトル領域グループを作成する．そしてスペクトル領域グループ単位で波数を選択する．この方策により，約 200 個のスペクトル領域を数個のスペクトル領域グループにまとめることができる．提案法の妥当性を確認するために，実データを用いた事例研究を実施した．

4.4 事例研究：顆粒中の薬物含量推定

波数選択の事例研究として，製剤顆粒中の薬物（化合物 X）含量を推定する検量モデルを構築した．

キャリブレーションセット及び薬物単味の NIR スペクトル並びに各波数選択法により選択された波数を Fig. 5 に示す．Fig. 5（a, b）に示すように，薬物（化合物 X）は 6,500 及び 5,150 cm^{-1} 付近において主要な吸収ピークを示す．また水分は 7,150，5,300，及び 5,000 cm^{-1} 付近に主要な吸収ピークを

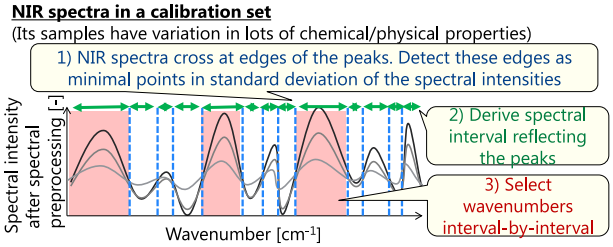

Fig. 3. An illustrative example of SFD-PLS.

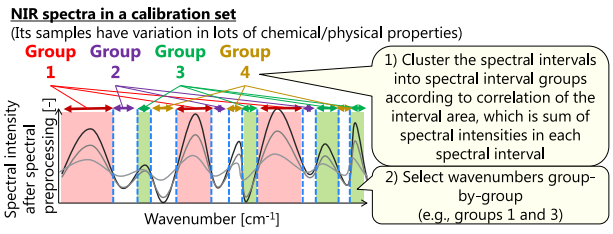

Fig. 4. An illustrative example of SFD-NCSC-PLS.

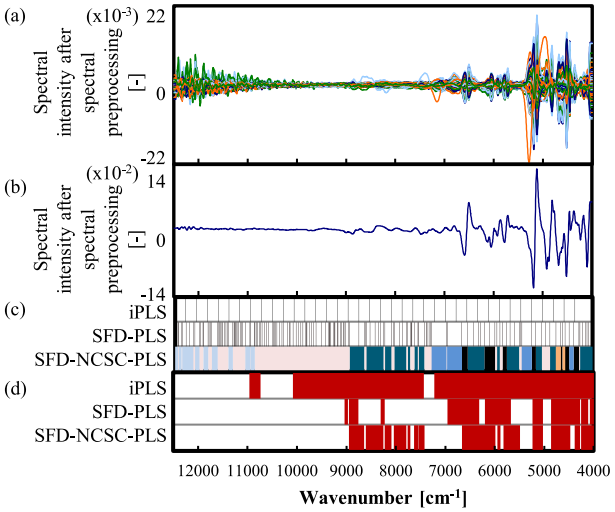

Fig. 5. Wavenumber clustering and selection results. (a) preprocessed NIR spectra in calibration set. (b) preprocessed NIR spectrum of drug substance. (c) generated spectral intervals and wavenumber groups; the same color means the same group in each method. In iPLS and SFD-PLS, vertical lines represent boundaries of spectral intervals. In SFD-NCSC-PLS, shaded regions represent wavenumber groups. (d) selected wavenumbers (shaded regions).

示す（これらの波数領域において顕著な変動を示すNIRスペクトルは高い水分含量を有するキャリブレーション用サンプルに相当する）．以上を踏まえて選択された波数を眺めると，提案法であるSFD-PLS及びSFD-NCSC-PLSは薬物（化合物X）の主要な吸収ピークを選択する一方で，水分に関連する波数領域を除外したと考えられる．

波数選択に要した計算時間及びその検量モデルで得られた推定精度をFig. 6に示す．SFD-PLSは従来法であるiPLSよりも高い推定精度（より低いstandard error of prediction：SEP）を短い計算時間で達成した．さらにSFD-NCSC-PLSはSFD-PLSよりも高い推定精度（低いSEP）を短い計算時間で達成した．従って，SFD-NCSC-PLSは従来法及びSFD-PLSよりも推定精度及び計算負荷の観点で優れている．本事例研究において提案法は1時間未満で波数選択を完了させたが，この程度の短い計算時間であれば実務担当者として許容できると感じた．

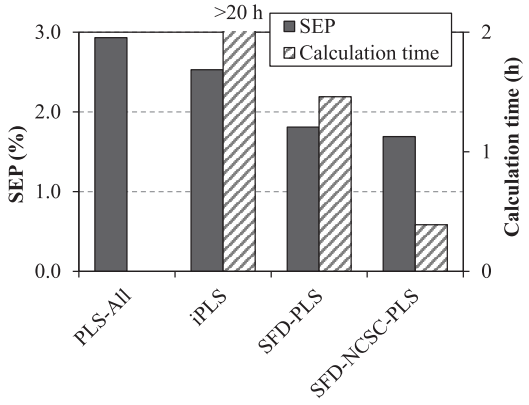

Fig. 6. Comparison of wavenumber selection methods in terms of SEP and calculation time. PLS-All uses all the wavenumbers without wavenumber selection.

5. 波数選択ソフトウェアの配布

上記の事例研究では，提案法をプログラミングソフト上で動作させた．しかし，プログラミング言語を取り扱える担当者は限られる．仮にそのプログラミング言語を取り扱えたとしても，他人が書いたプログラムコードを読み解くのは骨の折れる作業である．このような状況では，自分以外に提案法を使ってくれる担当者はいないだろう．

そこで，プログラミングの知識がない担当者でも提案法を実行できるように，実行形式のソフトウェアを作成した．Fig. 7にそのソフトウェア画面を示す．本ソフトウェアを用いることで，単純なパソコン操作のみで提案法を実行し，その結果を得ること

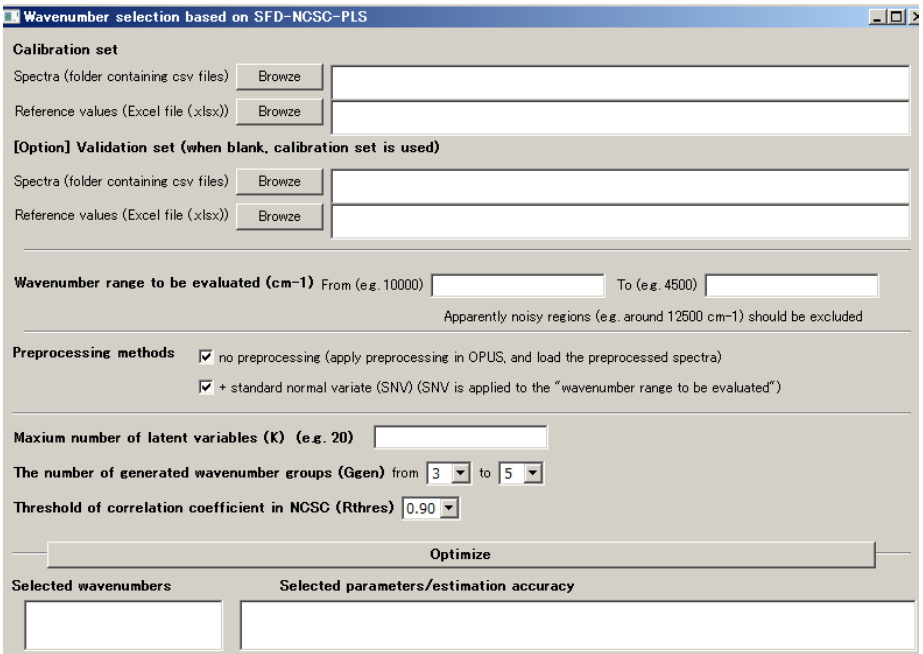

Fig. 7. Appearance of software to conduct SFD-NCSC-PLS. (procedure) 1. Load spectra and reference data. 2. Enter wavenumber range to be evaluated. 3. Select preprocessing method and tuning parameters. 4. Click 'Optimize', then results are shown.

ができる．実際，本ソフトウェアを社内で配布したところ，何人かの担当者に実務で提案法を活用してもらうことができ，とても便利だというフィードバックを頂いている．自分が作成した手法が誰かの役に立つのは嬉しい．

なお，本ソフトウェアの作成にあたって，無料のプログラミング言語である Python を使用した．Python で作成したプログラムであれば，PyQt（Graphical User Interface を作成するツール）や py2exe（Python のコードを実行形式に変換するツール）等のツールを使用して簡便にソフトウェアを作成可能である．こうしたツールを活用することで，新規技術を容易に共有できる．

6. おわりに

情報処理技術を用いた新規の波数選択法を提案した．提案法は従来法よりも高い推定精度を短い計算時間で達成した．さらに，実行形式のソフトウェアを作成したことで，プログラミングの知識がない担当者でも提案法を使用することができるようになった．本研究は，PAT を適切かつ効率的に導入することに貢献する．こうした PAT の推進が，製品品質及び生産効率を向上する一助になることを望む．

情報処理技術が活用される他の研究領域として，バイオインフォマティクス（生命情報科学）がある．冒頭で述べた通り私はヒトの生体内の情報に興味を持っており，本研究で培った情報処理技術をバイオインフォマティクスにも活かしたいと希望していた．その希望が受け入れられ，2017 年 4 月バイオマーカー推進部に異動し，現在はバイオインフォマティクス研究に着手したところである．今後は，本領域で新しい薬を創出するための研究を推進し，世の人を病から守るという大きな取り組みに貢献していきたい．

最後に，終始懇切なる御指導を賜りました京都大学・加納学教授，東京大学・杉山和弘准教授，東京薬科大学・瀬田康生教授，及び第一三共株式会社・渡部知行博士に深謝致します．多大な御協力を賜りました第一三共株式会社・中川弘司博士及び田邊秀章氏，並びに京都大学・藤原幸一助教，ルイトポルド・ファーマシューティカルズ・南秀実氏及び第一三共プロファーマ株式会社・脇山尚樹博士に心より感謝申し上げます．

引用文献

1) T. Miyano, M. Kano, H. Tanabe, H. Nakagawa, T. Watanabe, H. Minami, Spectral fluctuation dividing for efficient wavenumber selection: Application to estimation of water and drug content in granules using near infrared spectroscopy, *Int. J. Pharm*., **475**, 504–513 (2014).
2) T. Miyano, K. Fujiwara, M. Kano, H. Tanabe, H. Nakagawa, T. Watanabe, H. Minami, Efficient wavenumber selection based on spectral fluctuation dividing and correlation-based clustering for calibration modeling, *Chemom. Intell. Lab. Syst*., **148**, 85–94 (2015).
3) T. Miyano, H. Nakagawa, T. Watanabe, H. Minami, H. Sugiyama, Operationalizing maintenance of calibration models based on near-infrared spectroscopy by knowledge integration, *J. Pharm. Innov*., **10**, 287–301 (2015).