

遺伝的アルゴリズム (GA) を活用した スペクトルの波長選択および時系列 データにおけるプロセス変数かつその時間 遅れ (ダイナミクス) の選択

明治大学 理工学部 応用化学科
データ化学工学研究室 金子 弘昌

波長選択・プロセス変数および時間遅れの選択とは？¹

- ✓ Genetic Algorithm-based Partial Least Squares (GAPLS) や Genetic Algorithm-based Support Vector Regression (GASVR) と似た、スペクトル解析や時系列データ解析 (ソフトセンサー) 用の手法
 - GAPLS, GASVR についてはこちら
<https://datachemeng.com/gapls gasvr/>
- ✓ スペクトルデータ解析では、波長の領域の組み合わせを、回帰モデルの推定性能が高くなるように選択
 - Genetic Algorithm-based WaveLength Selection using Partial Least Squares (GAWLSPLS)
 - Genetic Algorithm-based WaveLength Selection using Support Vector Regression (GAWLSSVR)

波長選択・プロセス変数および時間遅れの選択とは？²

- ✓ 時系列データ解析 (ソフトセンサー) では、プロセス変数の組み合わせとそれらの時間遅れ幅を、回帰モデルの推定性能が高くなるように選択
 - Genetic Algorithm-based process Variable and Dynamics Selection using Partial Least Squares (GAVDSPLS)
 - Genetic Algorithm-based process Variable and Dynamics Selection using Support Vector Regression (GAVDSSVR)
- ✓ スペクトルデータと時系列データは似ているため、同じような手法を用いることができる
 - 参考: <https://datachemeng.com/preprocessspectratimeseriesdata/>

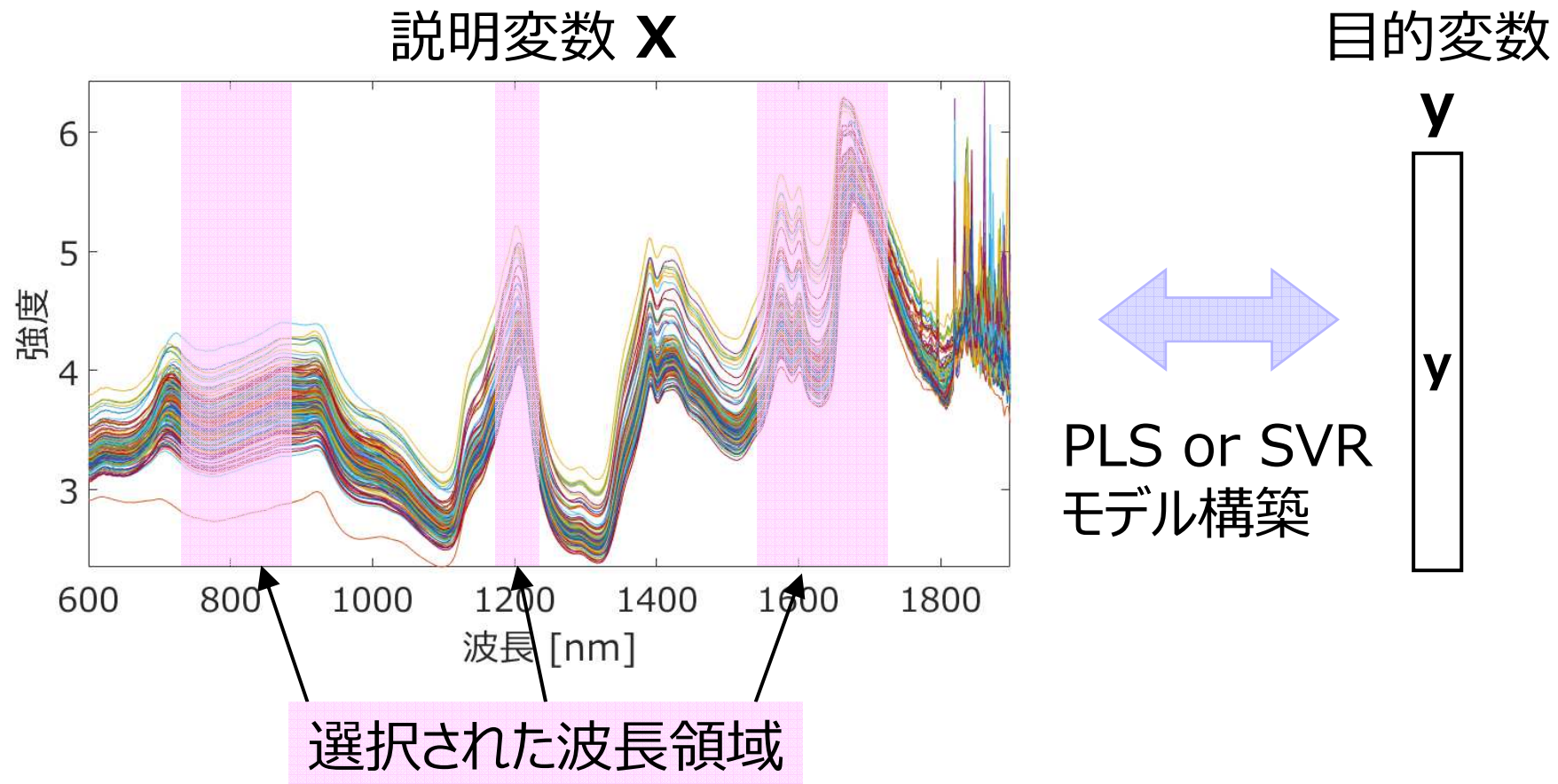
PLS, SVR

✓ PLS, SVRについてはこちら

- PLS : <https://datachemeng.com/partialleastquares/>
- SVR : <https://datachemeng.com/supportvectorregression/>

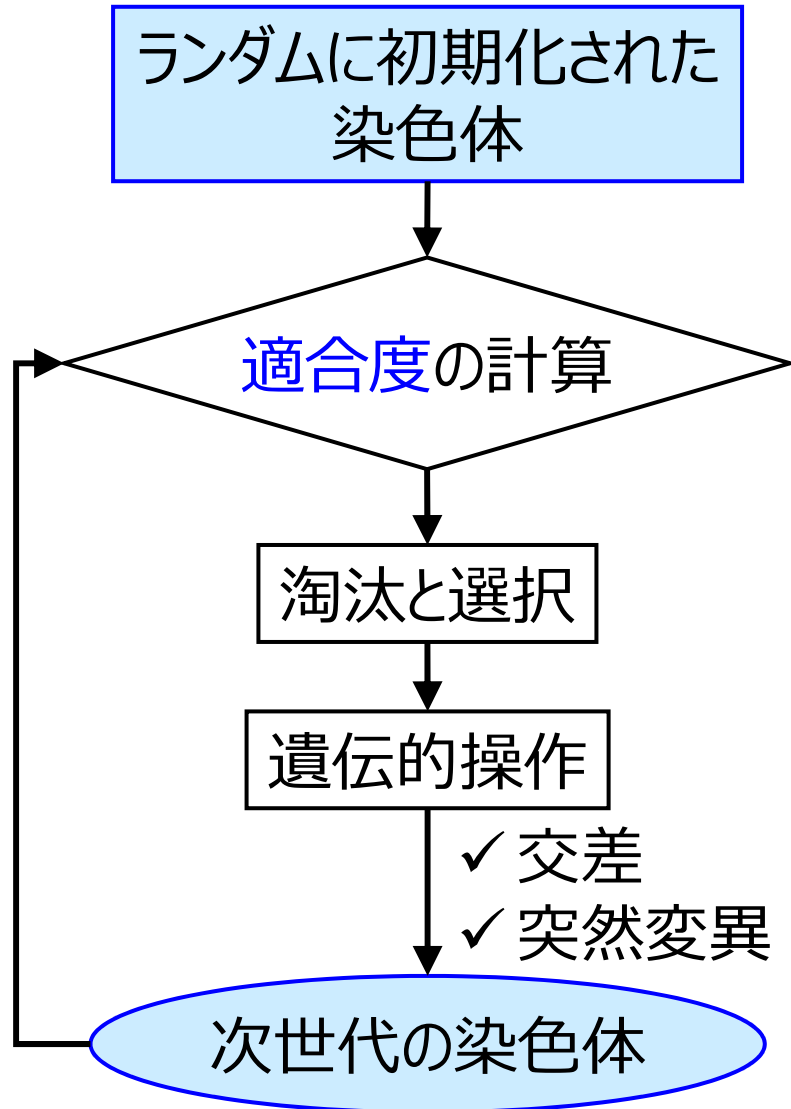
波長を領域で選択するイメージ

選択する波長領域が3つの場合



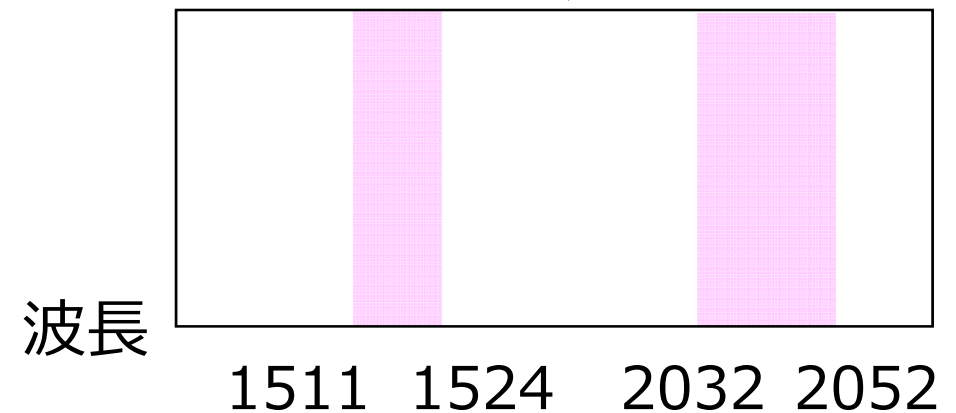
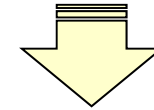
遺伝的アルゴリズム (GA) で最適な領域の組み合わせを選択

波長領域選択



染色体

1511	14	2032	21
------	----	------	----



適合度: 選択された波長領域のみで PLS や SVR を行った際のクロスバリデーション推定値を用いた r^2

波長領域を最適化可能

GAWLSPLSやGAWLSSVRの前に設定すること

6

✓ 選択する領域の数

- いくつかの領域数で波長領域を選択し、モデルの推定性能を確認して適した領域の数を決めてもよい

✓ 選択する領域の幅の最大値

- ある程度大きくしておけばよい

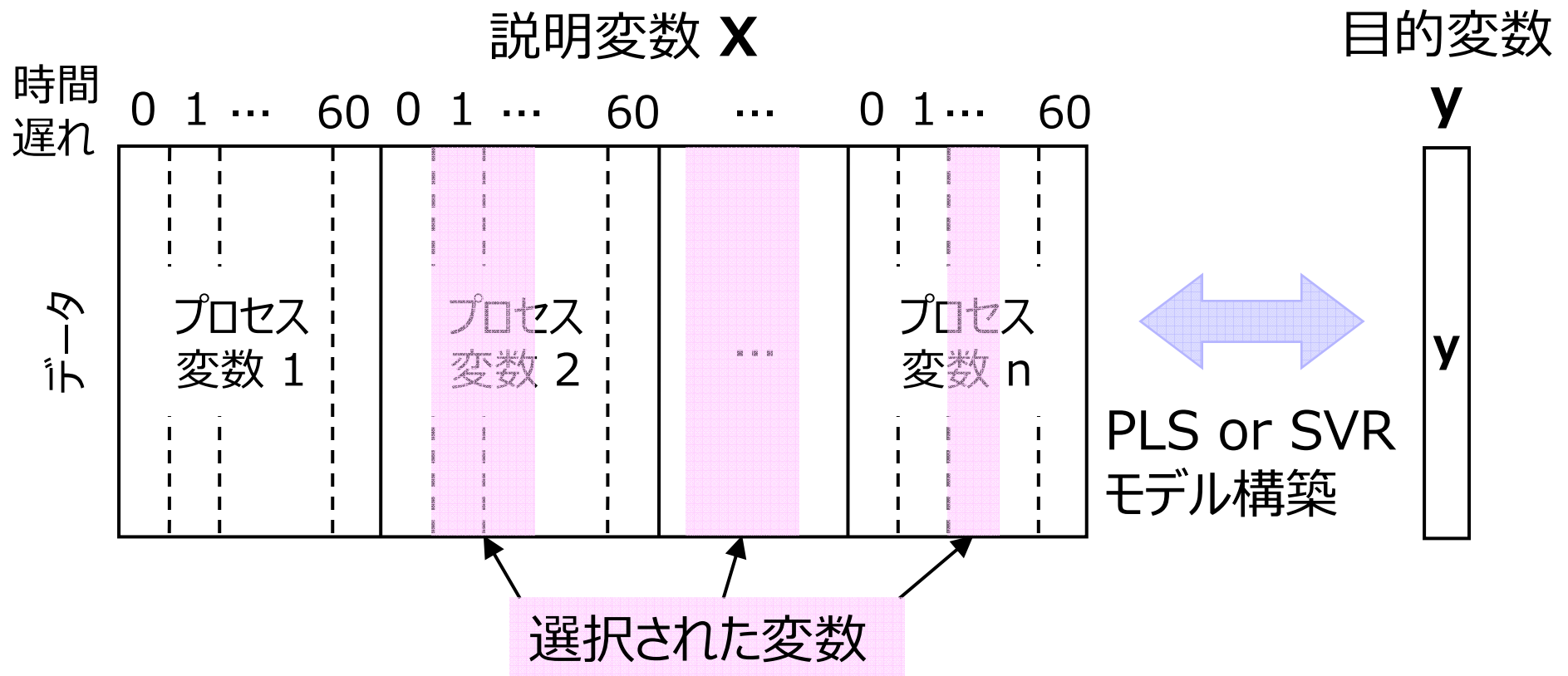
✓ PLS のときの最適成分数、SVR のときの C, ε, γ

✓ 遺伝的アルゴリズム (GA) 関係の設定

- 染色体数、世代数 など

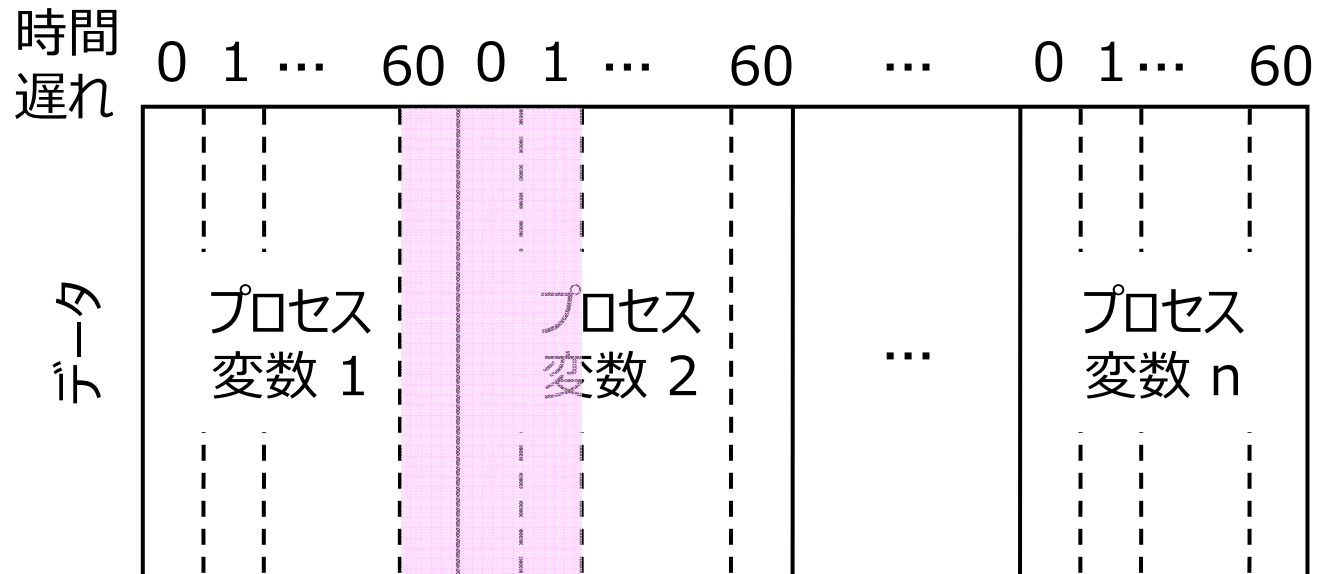
プロセス変数および時間遅れを選択するイメージ⁷

選択するプロセス変数の数が3つの場合



遺伝的アルゴリズム (GA) で最適な組み合わせを選ぶ

波長領域選択と異なること



2つのプロセス変数をまたいで時間遅れ領域が選択されないようにする

GAVDSPLSやGAVDSSVRの前に設定すること

✓ 選択するプロセス変数の数

- いくつかのプロセス変数の数で GAVDSPLS や GAVDSSVR を行い、モデルの推定性能を確認して適した数を決めてもよい

✓ 選択する時間遅れの最大値

- ある程度大きくしておけばよい

✓ PLS のときの最適成分数、SVR のときの C, ε, γ

✓ 遺伝的アルゴリズム (GA) 関係の設定

- 染色体数、世代数 など

どうやって実際にGAWLS, GAVDSを実行するか？

10

✓ scikit-learn や DEAP を用いて、

- GAWLSPLS
- GAWLSSVR
- GAVDSPLS
- GAVDSSVR

のデモを行うプログラムを作成しました

https://github.com/hkaneko1985/gawls_gavds

- ✓ランダム性があるため、GAWLSPLS, GAWLSSVR, GAVDSPLS, GAVDSSVR を行った結果、いつも同じ結果が得られるとは限らない
- ✓クロスバリデーションをして推定した結果がよくなるように変数を選択しているが、モデルがオーバーフィッティングする可能性もある

- ✓ M. Arakawa, Y. Yamashita, K. Funatsu, *J. Chemometr.*, 25, 10-19, 2011
- ✓ H. Kaneko, K. Funatsu, *AIChE J.*, 58, 1829-1840, 2012
- ✓ H. Kaneko, K. Funatsu, *Chemometr. Intell. Lab. Syst.*, 121, 26-32, 2013