

数値シミュレーションモデル

- ・ 数値シミュレーションモデルは、物理法則を数式（微分方程式など）で表現し、コンピューターで演算することで、実現象をコンピューター上で再現するモデルである。再現性が十分なモデルであれば、実際には起きていない現象（気象条件など）が起こった場合にどういった現象が起こるのか等、数値実験が可能となる。

諏訪湖貧酸素水塊モデル

- ・ 「平成 29 年度 湖沼水質保全対策調査検討業務」（環境省委託業務）において諏訪湖の物理場及び水質の再現計算を行い、その再現性について客観的に妥当性の評価を行ったモデルを基本とした数値シミュレーションモデルである（第 17 回世界湖沼会議（いばらき霞ヶ浦 2018）分科会にてポスター発表）。

貯留関数法

- ・ 河川などの流量を推定する流出解析モデルの一つで、流量観測がされている大河川での実績が多い。流出過程を降雨が流域の各部分に貯留されながら流れると考える。計算過程に経験値から作られたパラメータなど、ブラックボックスがあるものの、合理式に比べると精度は高い。一般財団法人 国土技術研究センターHP より流出解析シミュレータを入手することもできる。

主成分分析、次元圧縮※

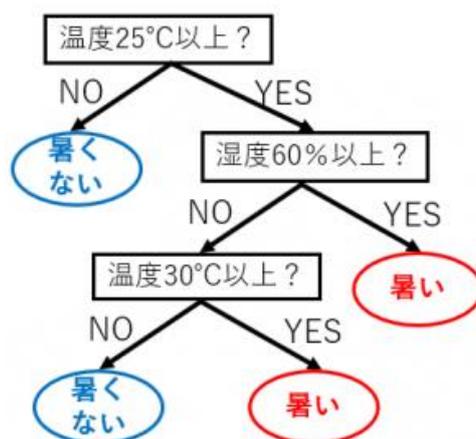
- ・ 次元圧縮（次元削減などとも言う）とは、入力データの次元を圧縮することである。簡単な例では、時系列のデータが「気温」、「海面水温」、「降水量」の 3 つ（3 次元）あった場合、このデータに次元圧縮を行うことで、気温と水温が近い特徴をもっている（季節変動などの共通した特徴を持つ）ため、出力結果として「気温+海面水温」、「降水量」の 2 つ（2 次元）に圧縮される。
- ・ データの次元数が大きいと、次元が増えるにしたがって解析時の計算量が指数関数的に増大し、解析が困難となる現象が起こる（＝次元の呪い）。
- ・ 主成分分析（PCA:Principal Component Analysis、海洋や気象関連の分野では EOF 解析とも呼ばれる）は次元圧縮の手法の一つで、入力データに対してばらつき（分散）を最大化する軸（第 1 主成分）と、これに直交する条件下でのばらつきを最大化する軸（第 2 主成分）、さらにこれらに直交する条件下での…として各主成分を選定し、線形で組み合わせる手法である。

FFT (Fast Fourier Transform : 高速フーリエ変換)

- ・ 波形データ (時系列データ) に対して、周波数成分を抽出可能なフーリエ変換をコンピュータ上で高速に演算する手法。どんな複雑な波形であっても複数の単調な波形 (三角関数の \sin 、 \cos) で表現することができるという理論に基づく (フーリエ変換)。Cooley and Tukey (1965) にて考案され、スペクトル解析・周波数解析手法を始めとして、現在でもよく使われている手法である。
- ・ 参考図書
 - 日野幹雄 (1977) 「スペクトル解析」
 - トランスナショナルカレッジオブレックス編 (1988) 「フーリエの冒険」

RF (ランダムフォレスト) と決定木

- ・ 決定木は木のような構造をもった機械学習モデル。このモデルにより、分類問題 (分類木) や回帰問題 (回帰木) を解く (図 1)。
- ・ RF (Random Forest : ランダムフォレスト) は、アンサンブル機械学習法の一つであり、決定木 (弱学習器) をランダムに複数用意して、同時 (並列) に分類または回帰を行う (これをバギングという)。例えば、図 1 のような決定木を複数作って多数決または平均値をとることにより、1 つの決定木のみで問題を解決するより効率よく分類可能または最適解を見つけやすい。名前はランダムに複数の決定木を使うことから、森 (forest) となることが由来とされている。



引用 : <https://mathwords.net/ketteigi>

図 1 決定木 (分類木) の例

状態空間モデル

- ・ 状態空間モデル (State Space Model) とは、観測された値と、観測の裏にある「状態」に分解してモデル化する手法である。
- ・ 他の用語で使われる「モデル」のイメージとは少し異なり、統計モデルの一つの枠組みを指す。有名な時系列モデルである ARIMA モデル、VAR モデル等もこの枠組みの一つである。
- ・ 基本は以下のような「状態方程式」と「観測方程式」に分けて考えるところから始まり、データの特徴を把握しながら、トレンドや季節成分等を加えてモデル化していく。

状態 = 前時点の状態を用いた予測値 + 過程誤差

観測値 = 状態 + 観測誤差

- ・ 参考図書
 - ✓ 馬場真哉 (2018) 「時系列分析と状態空間モデルの基礎 R と Stan で学ぶ理論と実装」

ニューラルプロセス

- ・ ニューラルプロセス (Neural Process, Garnelo *et al.*, 2018¹², Kim *et al.*, 2019³) とは、ガウス過程 (Gaussian Process) とニューラルネットワーク (NN: Neural Network) の利点を組み合わせたような手法である。
- ・ ガウス過程とは、あるパラメータに対して関数形をランダムに出力する箱のようなものである。一般的に、ガウス過程は観測値に対する確率分布を生成でき (誤差範囲 (不確実性) も示すことができる)、柔軟に観測値を追加できる一方で、計算量が膨大になりやすい。代表的な用途としては、回帰である (ガウス過程回帰)。
- ・ ニューラルネットワークは、脳神経回路を数理モデル化したものであり、このネットワーク (層) を数多く (深く) 連ねた機械学習モデルとして近年、「深層学習」という言葉で注目されているモデルである。ニューラルネットワークは、関数の近似が得意である一方で、新しい関数ごとにゼロから学習しなければならないことが課題の一つである (この課題に対し、近年様々な方法が考案されている)。
- ・ 参考図書等
 - ✓ <https://www.slideshare.net/DeepLearningJP2016/dlconditional-neural-processes>

¹ <https://arxiv.org/pdf/1807.01613.pdf>

² <https://arxiv.org/pdf/1807.01622.pdf>

³ <https://arxiv.org/pdf/1901.05761.pdf>

- ✓ 持橋大地、大羽成征（2019）「ガウス過程と機械学習」

交差検証（クロス・バリデーション）

- ・ 交差検証（Cross Validation）とは、一つのデータセットを分割し、一部のデータは学習に利用し、残りのデータを用いて検証を行う検証手法である。
- ・ 交差検証には k-分割交差検証（k-fold Cross Validation）などいくつかの種類がある。
- ・ 学習器は、学習時は入力データに対する答えをセットで見て学習する。検証時には、入力データに対して推定を行うのみで、答えは一切見ない。検証では答えを見たことのないデータに対してどれだけの精度で推定できるかどうか（汎化性能）を評価する。

ブルーム*

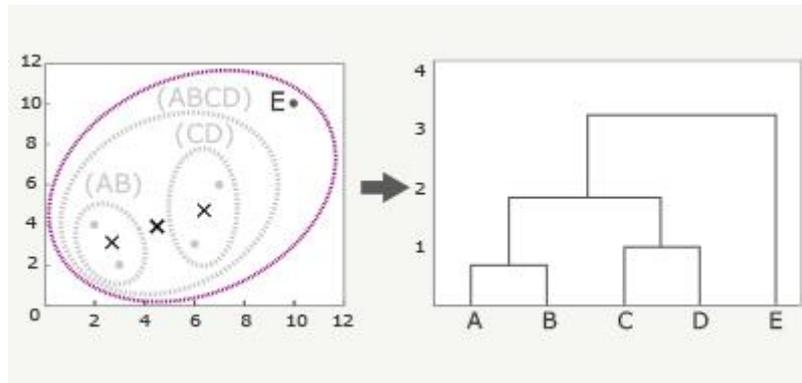
- ・ ブルーム（Bloom）とは、植物プランクトンが大量に発生し、花が咲いたように水面付近が変色した状態を指す。「水の華」とも呼ばれる。

沈降フラックス*

- ・ 沈降フラックス（Deposition Flux）とは、水底へ沈み落ちる物質の量。通常、単位面積・単位時間当たりの沈殿量として定義される（例：全有機炭素フラックス = mg 炭素/m²/日）。

クラスター分析*

- ・ クラスター分析（Cluster Analysis、クラスター解析・クラスタリングとも呼ばれる）は、入力したデータから、互いに似た性質を持つものを集め、このクラスター（集合）を複数作成して分類を行う分析手法である。
- ・ クラスター分析には、「階層的手法」と「非階層的手法」があるが、本検討では「階層的手法」を用いている。階層的手法では樹形図（デンドログラム）が結果として可視化される（図 2）。



引用 : https://www.albert2005.co.jp/knowledge/data_mining/cluster/hierarchical_clustering

図 2 樹形図の例