

# 穀物の生育管理におけるデータ活用の検討

舛田 栄次\*・森 信彰\*<sup>2</sup>

## Study on Application of Big Data Analysis for Grain Growth Management Eiji Masuda and Nobuaki Mori

### 1. 緒 言

近年のIoT技術の進展により、データの収集が容易に行える環境が整いつつある<sup>1)</sup>。また、気象データをはじめとした様々なデータのオープン化も進んでおり<sup>2)</sup>、ビッグデータとして自由に活用できる事例も増えている。一方で、データ解析に関するライブラリもオープンソースとして充実してきており<sup>3)</sup>、これらを活用して新たな価値を創出することが求められる。

農業分野においても、データを活用した新たな農業への変革、いわゆる農業DXが推進されている<sup>4)</sup>。作物の中でも、麦は自給率が低く国内生産の安定化が望まれており<sup>5)</sup>、農業DXによる生産性向上への期待は大きい。関連研究として、収量の安定化と増大を目指した気象変動との関係の解析や<sup>6)</sup>、管理作業の実施性を判断するための発育モデルの構築<sup>7)</sup>も行われており、今後ますますデータ活用が進むと考えられる。

麦の栽培においては、高収量を目指すため、適切に茎数を管理することが重要となる。そのためには、例えば、生育が早い場合に踏圧の時期を早め、回数も増やすなど、適切なタイミングで適切な作業を実施する必要がある。このとき、将来の茎数が事前に予測できれば、はやい段階で作業の実施性を判断でき、より高度な麦生産が可能になると考えられる。本研究では、苗立数および気象データを活用して4種類の茎数回帰モデルを構築し、それぞれの予測精度について比較を行った。

### 2. 解析方法

本研究では、茎数を目的変数、苗立数および気象データを説明変数としたモデルの構築を目的とする。

#### 2.1 利用データ

本研究では、以下の2種類のデータを利用する。

- 山口県内ほ場の定点調査データ
- 1kmメッシュ気象データ

山口県内ほ場の定点調査データは、山口県農林総合技術センターが取りまとめたデータで、山口県内数十地点における播種後1ヶ月程度(12月頃)の苗立数、それ以降の1月から3月にかけての茎数のデータである。1kmメッシュ気象データは、農業・食品産業技術総合研究機構が提供して

いる<sup>8)</sup>。本研究では、2019年度産の6地点、2020年度産の5地点、2021年度産の23地点、2022年度産の18地点の計52地点分の調査データを利用した。茎数は初回の調査(播種後2か月程度の1月頃)のものを利用した。

#### 2.2 説明変数の選択方法

茎数を目的変数とする回帰モデルの説明変数を決定するため、茎数と苗立数および気象(気温、降水量、および日長)に関する基本統計量(積算、平均、最大、および最小)に対する相関分析を行った。ここで、これらの基本統計量は播種日から茎数計測日までの日別時系列データに対して算出したものであり、積算は播種日から茎数計測日までの日別平均値の総和を意味する。相関分析において、最少降水量は無視した。また、前処理として正規化を行った。

#### 2.3 回帰モデル

本研究では、苗立数および気象データを説明変数、茎数を目的変数とする以下の4種類の回帰モデルの構築を行った。

- 重回帰モデル<sup>9)</sup>
- ニューラルネットワーク(NN: Neural Network)<sup>10)</sup>
- LSTM(Long Short-Term Memory)<sup>11)</sup>
- マルチモーダルNN(MNN: Multimodal Neural Network)<sup>12)</sup>

重回帰モデルおよびニューラルネットワークは、説明変数としてスカラ多変量を扱うことができ、LSTMは可変長の時系列を扱うことができる。MNNはスカラ多変量と可変長の時系列の双方を扱うことができる。ここで、NN、LSTM、およびMNNでは、予測精度に影響するノード数、学習率、バッチサイズなどのハイパーパラメータや活性化関数および最適化手法について、ベイズ最適化<sup>13)</sup>を用いて最適な組合せを探索した。このとき、目的関数は訓練誤差と評価誤差の和とし、探索範囲は表1に示すとおり設定した。アルゴリズムの反復回数は50回とした。

また、本研究では、2020年度産から2022年度産の46地点分のデータを訓練データとし、2019年度産の6地点分のデータをハイパーパラメータ調整のための検証データとした。ただし、重回帰モデルに対しては、ハイパーパラメータ調整が不要であるため、すべてを訓練データとした。

#### 2.4 開発環境

データ分析およびモデルの構築にはLinux(Ubuntu 20.04.4 LTS)を利用した。回帰モデルの構築はPython(Ver. 3.9.12)上で行い、機械学習ライブラリとして重回帰モデルにはscikit-learn(Ver.1.2.2)、NN、RNN、およびMNNには

\* 技術支援部

\*2 プロジェクト推進部

TensorFlow(Ver. 2.12.0)を用いた。ベイズ最適化には、bayesian-optimization(Ver. 1.4.3)を用いた。

ノード数	[1, 20]上の整数
活性化関数	ReLU Sigmoid tanh
最適化手法	SGD RMSprop Adagrad Adadelta Adam
学習率	[0.001, 0.1]上の実数
バッチサイズ	[1, 32]上の整数

表1 ベイズ最適化によるハイパーパラメータ、活性化関数、および最適化手法の探索範囲

### 3. 解析結果と考察

#### 3・1 相関分析の結果と説明変数の選択

図1に、茎数に対する苗立数および気象に関する基本統計量の散布図を、表2に、これらの相関係数を示す。苗立数、積算気温、最高気温、積算日長、および最長日長に対してはやや大きい正の相関があることがわかる。一方、平均降水量および最多降水量に対してはやや大きい負の相関があることがわかる。

表3に、本研究で検討した回帰モデルと用いた説明変数を示す。重回帰モデルおよびNNはスカラ多変量を扱うため、相関係数の絶対値が0.3以上となる項目、すなわち、苗立数、積算気温、最高気温、平均降水量、最多降水量、積算日長、および最長日長を説明変数として採用した。これらの説明変数を用いて推定した重回帰モデルの回帰係数を表4に示す。LSTMは可変長時系列データを扱えるため、播種日から茎数計測日までの気温、降水量、および日長の日別時系列データを説明変数とした。MNNはスカラ多変量と時系列といった複数のモダリティを説明変数として扱うため、苗立数および播種日から茎数計測日までの気温、降水量、および日長の日別時系列データを説明変数とした。ただし、苗立数を入力とする部分にはNN、時系列データを入力とする部分にはLSTMを用いた。

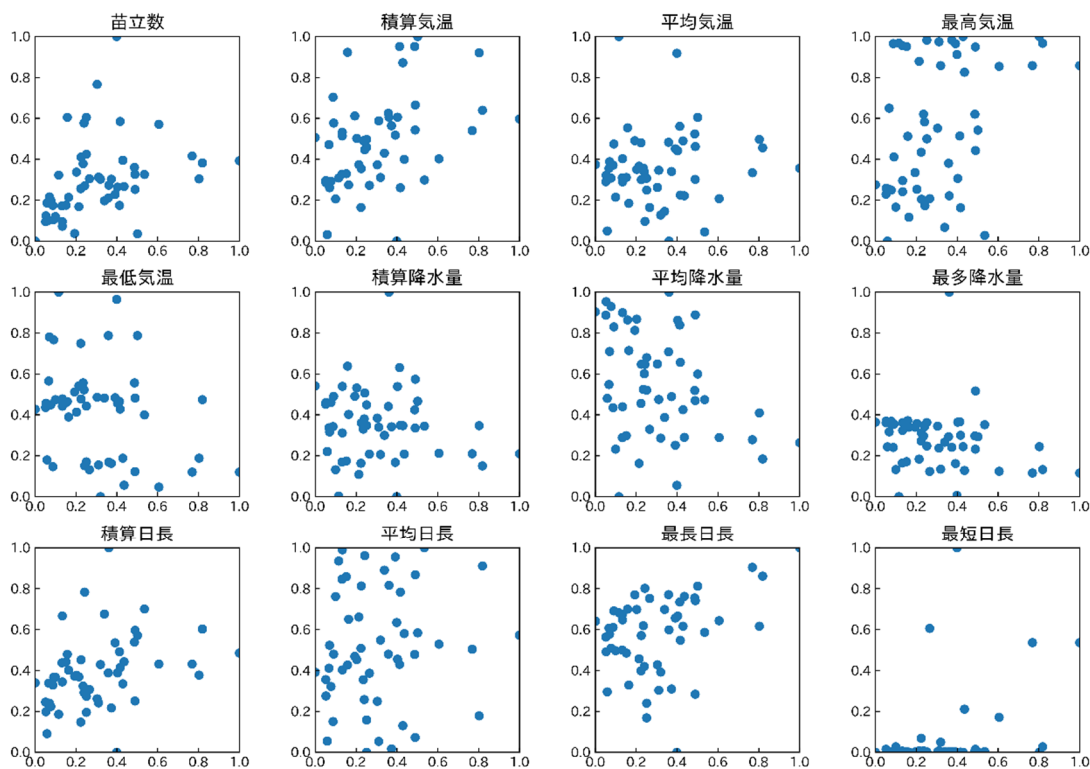


図1 茎数(横軸)に対する苗立数および気象に関する基本統計量(縦軸)の散布図

表2 茎数に対する苗立数および気象統計量との相関係数

苗立数	気温				降水量				日長			
	積算	平均	最高	最低	積算	平均	最多	最少	積算	平均	最長	最短
0.53	0.42	0.08	0.33	-0.25	-0.09	-0.33	-0.31	-	0.45	0.16	0.31	0.27

表3 本研究で用いた回帰モデルと説明変数

モデル	初期生育量に関する説明変数	気象に関する説明変数
重回帰	苗立数	積算気温 最高気温 平均降水量 最多降水量 積算日長 最長日長
NN	苗立数	積算気温 最高気温 平均降水量 最多降水量 積算日長 最長日長
LSTM	-	日別気温 日別降水量 日別日長
MNN	苗立数	日別気温 日別降水量 日別日長

表4 重回帰モデルの回帰係数

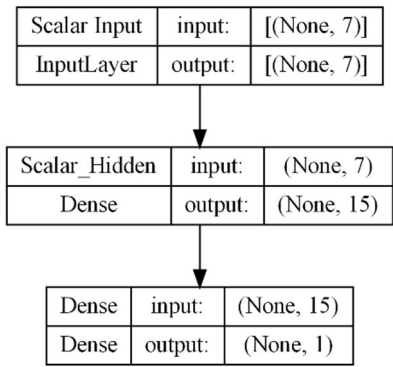
説明変数	回帰係数
苗立数	0.49
積算気温	0.24
最高気温	0.09
平均降水量	-0.2
最多降水量	-0.06
積算日長	0.27
最長日長	0.33

3・2 ハイパーパラメータ最適化および各種モデル

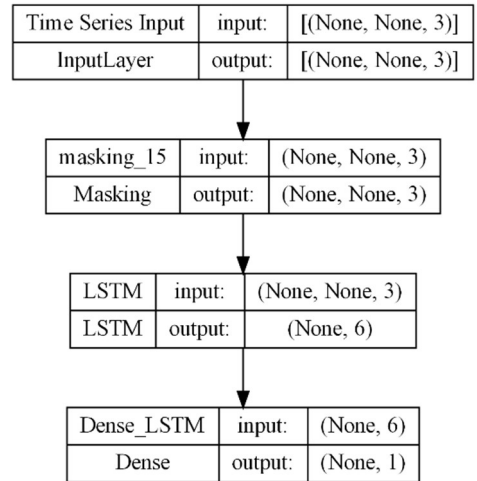
表5に、ベイズ最適化に基づいて求めたNN, LSTM, およびMNNそれぞれのモデルのハイパーパラメータを示す。ただし、各モデルの中間層数は1層とした。図2に、これらのハイパーパラメータを用いて構成したNN, LSTM, およびMNNのネットワーク構造を示す。

表5 ハイパーパラメータ最適化の結果

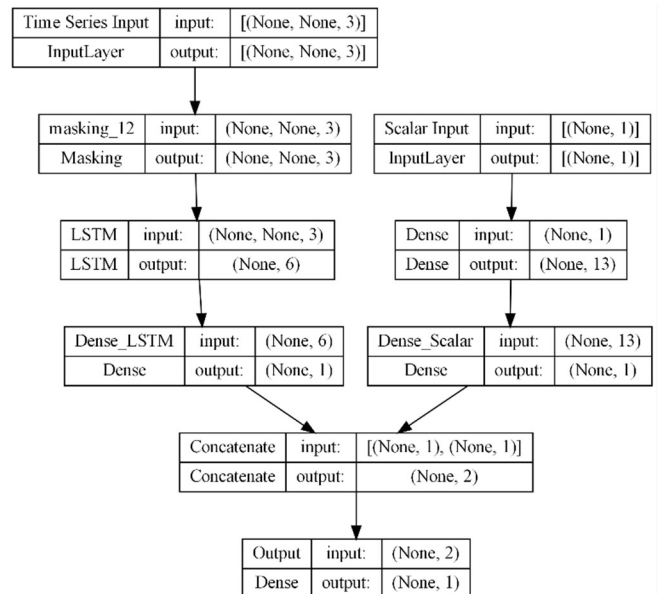
	NN	LSTM	MNN
ノード数	15	6	13 (NN部) 6 (LSTM部)
活性化関数	ReLU	tanh	Sigmoid (NN部) tanh (LSTM部)
最適化手法	SGD	RMSprop	SGD
学習率	0.014	0.018	0.014
バッチサイズ	1	8	19



(a) NN



(b) LSTM



(c) MNN

図2 ネットワーク構造

3・3 予測結果

図3に、表3で示した各回帰モデルによる予測値と実測値の比較を示す。ここで、青の点は訓練データに対する予測値、橙の点は検証データに対する予測値である。図3か

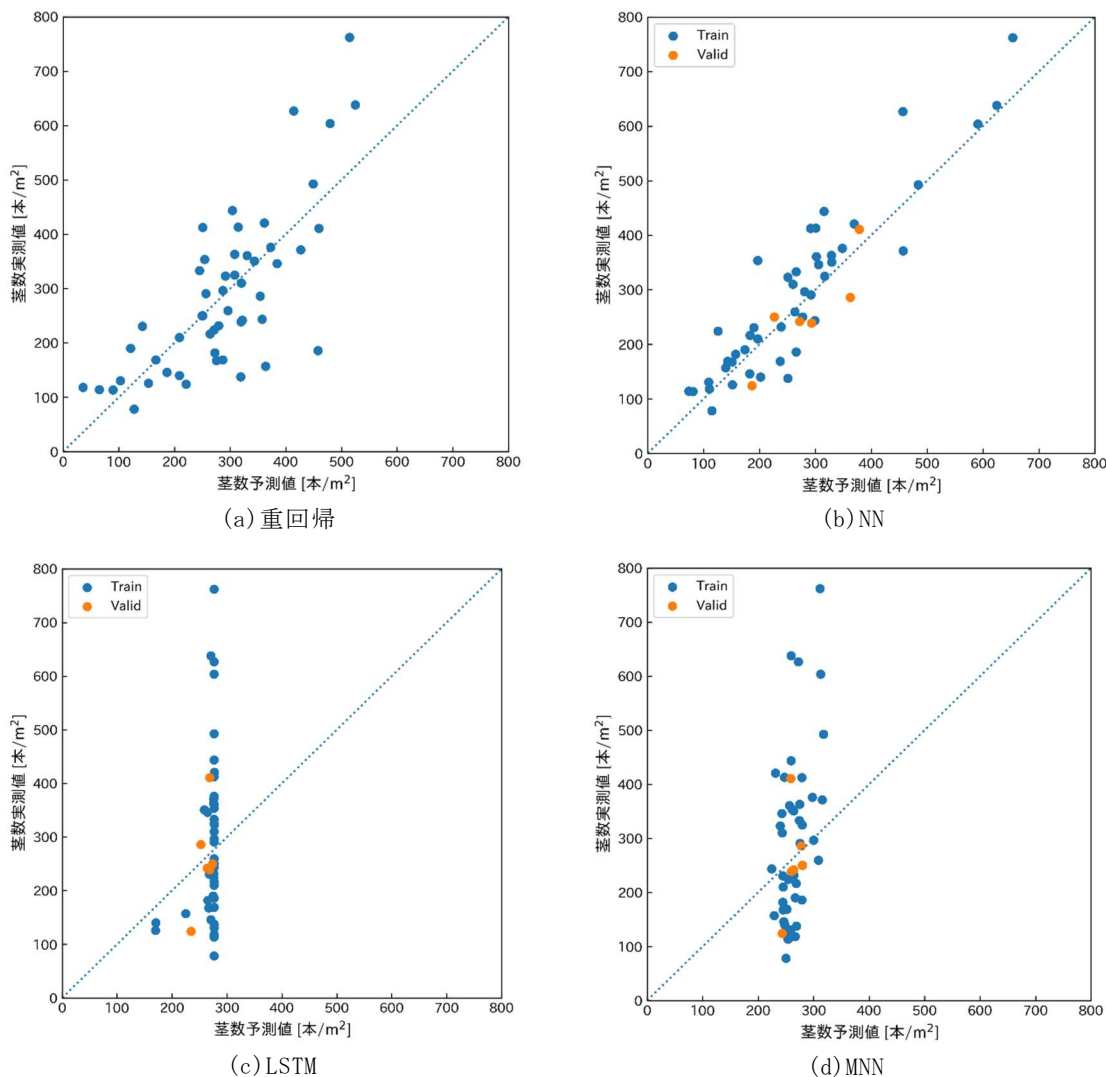


図3 茎数予測値と茎数実測値の比較

ら、重回帰モデルおよびNNは、50-600 本/m<sup>2</sup>内で広く予測されているのに対し、LSTMは100-300 本/m<sup>2</sup>、MNNは200-300 本/m<sup>2</sup>と予測範囲が狭く、入力に関わらず同じような予測を行っていることがわかる。また、重回帰モデルおよびNNによる予測は、RMSEがそれぞれ97.1と63.2、R<sup>2</sup>がそれぞれ0.57と0.82であった。一方、LSTMおよびMNNによる予測は、RMSEがそれぞれ144.5と140.7、R<sup>2</sup>がそれぞれ0.04と0.09であり、時系列データを用いたモデルはスカラ多変量を用いたモデルよりも予測精度が低かった。

これらの結果から、LSTMおよびMNNは気象時系列データの特徴を十分に学習できていないと推察される。この要因として、気象時系列データが播種日から茎数計測日までの約50日分のデータであるのに対し、データセットが46地点分と少なかったため、学習に必要なバリエーションデータを十分に含んでいなかったことが考えられる。これを解決するためには、サンプリング地域や年次の拡充、前処理による気象時系列データ長の削減などの工夫が考えられる。

#### 4. 結 言

本研究では、苗立数および気象データを活用した茎数回

帰モデルの構築について検討した。重回帰モデルおよびNNに基づくモデルは、苗立数および気象データに関する基本統計量(積算, 平均, 最大, および最小)を説明変数とした。一方、LSTMは日別気象時系列データを、MNNは、苗立数および日別気象時系列データの双方を用いた。その結果、重回帰モデルおよびNNによる基本統計量を用いた予測に対し、LSTMおよびMNNによる時系列データを用いた予測の方が精度が低かった。ただし、この差はモデル性能の違いではなく、学習データ数が少なかったことが原因だったと考えられる。今後は、学習データの増加や前処理等による必要学習データ数の低減などの工夫を行い、より高精度なモデルを検討する必要がある。さらに、具体的な管理作業へ反映させられる情報提供の仕組みを検討していくことも必要である。

本研究で使用したデータは、各農林事務所で作業により計測され山口県農林総合技術センターが取りまとめたものである。計測者および山口県農林総合技術センター(現山口県農林水産部農業振興課)の河村剛英氏に深く感謝いたします。

## 参考文献

- 1) 鈴木一哉, 他:IoT 技術の最新動向, 電子情報通信学会通信ソサイエティマガジン, **12**(1), p. 12-20(2018), DOI:<https://doi.org/10.1587/bplus.12.12>.
- 2) 高度情報通信ネットワーク社会推進戦略本部: オープンデータ基本指針, <https://cio.go.jp/sites/default/files/uploads/documents/kihonsisin.pdf>(参照日 2025 年 6 月 29 日).
- 3) 新村正明:JupyterLab による Python 学習のはじめかた, 電子情報通信学会 通信ソサイエティマガジン, **15**(4), p. 333-338(2022), DOI: <https://doi.org/10.1587/bplus.15.333>.
- 4) 農林水産省:「農業 DX 構想」と今後のデジタル戦略の推進について, [https://www.maff.go.jp/j/kanbo/dx/attach/pdf/nougyou\\_dxxkousou-64.pdf](https://www.maff.go.jp/j/kanbo/dx/attach/pdf/nougyou_dxxkousou-64.pdf)(参照日 2025 年 6 月 29 日).
- 5) 農林水産省:麦・大豆生産技術向上事業の概要, [https://www.maff.go.jp/j/syouan/keikaku/soukatu/attach/pdf/mugimame\\_kokusanka-33.pdf](https://www.maff.go.jp/j/syouan/keikaku/soukatu/attach/pdf/mugimame_kokusanka-33.pdf)(参照日 2025 年 6 月 29 日).
- 6) 箕田豊尚, 他:埼玉県におけるコムギの生育, 収量および収量構成要素と気象変動の関係—畑作試験圃場におけるコムギ「農林 61 号」の 45 年間の栽培試験に基づく解析—, 日作紀, **84**(3), p. 285-294(2015).
- 7) 中園江, 他:コムギの発育段階の推定モデル, 日作紀, **83**(3), p. 249-259(2014).
- 8) 農研機構:農研機構メッシュ農業気象データ, [https://amu.rd.naro.go.jp/wiki\\_open/doku.php?id=start2](https://amu.rd.naro.go.jp/wiki_open/doku.php?id=start2)(参照日 2025 年 6 月 29 日).
- 9) 阿部真人:データ分析に必須の知識・考え方 統計学入門 仮説検定から統計モデリングまで重要トピックを完全網羅, ソシム, p. 187-191(2021).
- 10) 岡谷貴之:深層学習, 講談社, p. 7-21(2015).
- 11) Hochreiter S. and Schmidhuber J.: Long Short-Term Memory, *Neural Computation*, **9**(8), p.1735-1780(1997), DOI:<https://doi.org/10.1162/neco.1997.9.8.1735>.
- 12) 柳井啓司, 中鹿亘, 稲葉通将:深層学習, オーム社, p. 234-252(2022).
- 13) Shahriari B., et al.: Taking the Human Out of the Loop: A Review of Bayesian Optimization, *Proceedings of the IEEE*, **104**(1), p.148-175(2016), DOI:<https://doi.org/10.1109/JPROC.2015.2494218>.