

# 研究終了報告書

## 「解釈可能な AI による土壌・作物系モデルの開発」

研究期間：2020年12月～2023年3月

研究者：田中 貴

### 1. 研究のねらい

農業生産の中でも、水稻や小麦などの農作物は、その年の気象や地域ごとに異なる土壌など不確定な要素から影響を受けやすい。このように変動しやすい環境要因に対して、安定的に効果のある最適な栽培管理を実施するためには、農業者が長年にわたって培ってきた経験や勘に頼らざるを得なかった。一方、農業の現場レベルで、近年、人工衛星・ドローンを用いたりリモートセンシングや土壌センサ、収量センサ付きコンバインなどが普及しつつあり、多様なセンサ類から取得される膨大な時空間データを、どのように営農現場における意思決定に活用するのが課題となっている。また、実際に作物モデルを農業者の意思決定支援に用いるためには、地域に応じたテーラーメイドなモデルが求められ、気象や土壌に対する作物の環境応答を定量的に解析する必要がある。既存の作物モデルで広く知られている WOFOST や APSIM などは、気候変動や品種選択、栽培技術の作物収量への影響を大まかに比較する際にしばしば用いられる研究者向けのツールとなっている。これらの作物モデルは、気象データを主なドライビングフォースとしており、実際の生産現場で不確定な土壌要因は、ほぼブラックボックスな状態で収量推定を実行する。つまり、実際の栽培環境が反映されていないため、農業者が意思決定を支援できるほどの精度は有していない。そこで、第一にブラックボックスだった土壌特性が経時的・空間的にどのように分布するのかを迅速かつ安価に定量化できる手法を確立する必要がある。次に、土壌の時空間変動の推定ができるようになると、膨大な数の環境因子の作物への影響評価が必要となる。ここで、深層学習を活用することで、環境因子と作物収量との間にある複雑な因果関係をモデル化できると期待されるが、そのモデルの複雑さから因果関係を解釈することは非常に困難である。また、環境因子を高精度に測定できるセンサ類がいくら発達したとしても、作物の収量変動を支配する全ての要因を網羅することは不可能だろう。つまり、実験において未観測な要因が、どの程度、モデルの不確実性に影響するのかを考慮する必要がある。そこで、本研究では、農家圃場における土壌特性の時空間分布をリアルタイムに推定する技術を開発し、その推定値を予測因子とする深層学習を用いて作物の収量予測モデルを構築し、最終的に解釈可能な土壌・作物系モデルに変換する手法を確立する。

## 2. 研究成果

## (1) 概要

本研究は、農家圃場において土壌特性の時空間変動を推定する技術を開発し、その推定値を予測因子とする深層学習を用いて作物の収量予測モデルを構築する。最終的には、本モデルの感度分析によって、栽培管理に新たな知見を付与することである。岐阜県海津市の農家圃場において、本研究の対象作物であるコムギ 2 作期分、ダイズ 1 作期分のデータを収集した。播種前に土壌試料の採取と深さ別土壌硬度の測定を実施し、可変施肥ブロードキャスターによる施肥量をランダムに変化させる施肥試験を行った。土壌の空間分布を推定するには、多点の土壌分析が必要となるが、迅速かつ低コストな推定手法を開発しなければならない。本研究では中赤外スペクトルから土壌の理化学性を推定する手法を開発した。複数の学習器をスタッキングするアンサンブル学習を用いることで、土壌特性を推定するモデルの精度を向上できた。また、コムギ収量データセットおよびマルチスペクトルカメラを搭載したドローンによるセンシング画像から、コムギ収量を推定する深層学習モデルを構築した。撮影時期における推定精度を比較した結果、出穂期が最も適した撮影時期であることが明らかとなった。これらの予測収量データと土壌や施肥量との関係性を複数の機械学習モデルを用いて学習と感度分析を行ったところ、いずれのモデルも大幅な減肥が生産者の収益増大に寄与することを明らかにした。しかし、学習器によって経済的最適施肥量の空間分布が大きく異なることが示され、機械学習モデルによる因果推論の不確実性が懸念された。そこで、気象や土壌養分、水分に対する作物の生理反応を考慮して、作物の生育や収量をシミュレーションできる作物モデルを用いて、合成データを生成する手法を開発した。これにより、農家圃場で取得した実データでは検証できなかった機械学習モデルの因果推論の妥当性を評価できる。また、本データセットは、作物の収量応答モデルの学習データとしても利用できると思われる。一方、深層学習による収量応答モデルの構築を行うことで、収量推定の精度を向上できた。しかし、研究当初の期待に反して、深層学習は、栽培管理に対する収量応答性の因果関係を全く評価できてなかった。今後は、因果推論を重視する Causal machine learning などの新たな手法の適応を検討する必要がある

## (2) 詳細

## 研究テーマ A「土壌特性値の時空間変動の推定」

土壌の分析コストを削減するために、迅速かつハイスループットに土壌特性値を推定する技術が求められる。本テーマでは、中赤外のスペクトルデータ(図1)から土壌特性値を推定する機械学習モデルの構築を目指した。研究期間中に、約 3000 点の土壌データを収集することができた。スペクトルデータを解析する手法として、研究開始の当初は multi-task CNN が有力な手法として期待されていたが、より単純なアルゴリズムである勾配ブースティングやランダムフォレストなどをアンサンブル学習させたモデルの方が推定精度は高かった。pH や全炭素などの項目では rRMSE が 10%以下と非常に高い推定精度を達成できた。一方、有効態リン酸や交換性 K などは、研究期間内において、高精度な推定という目標を達成することはできなかった。スペクトルデータの前処理、解析アルゴリズム等の最適化に関して、論文投稿の準備を進めている。また、土壌水分の時空間変動を推定するために、現地圃場において、土壌

水分センサを設置し、経時的な土壌体積含水率の推移データをモニタリングした。SAR 衛星の後方散乱係数や植生指数を説明変数とする推定モデルの構築を試みたが、推定に十分な精度を達成することはできなかった。今後は、気象データや時系列変化を考慮できるLSTMなどのモデルを用いて推定できるかどうかを試行する必要がある。

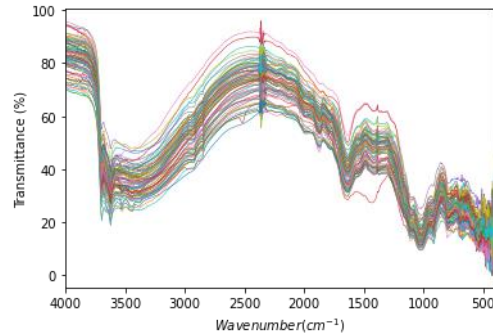


図1. 中赤外スペクトルの例

#### 研究テーマB「収量推定モデルの高精度化」

本テーマでは、ドローンに搭載したマルチスペクトル画像を用いて、CNNによりコムギの収量を予測するモデルを構築した。マルチスペクトル画像の時系列的な変化を学習することで、収量予測精度の向上を図れると、研究当初において想定していたが、出穂期の一時期の画像データから推定した精度と、複数次期のデータを用いたモデルとの精度には差が認められなかった。すなわち、コムギの収量推定には、複数時期の撮影が不要なことを明らかにした。なお、従来の植生指数を説明変数とする線形回帰モデルと比較して、RMSEにして1.0 t/haから0.9 t/haに推定誤差を下げることに成功したが、達成目標としていた0.5 t/haの推定誤差には及ばなかった。取得画像の高解像化を通じて収量推定における飽和問題を解決できる可能性がある。

#### 研究テーマC「土壌・作物系モデルの構築」

土壌データ(研究テーマAで取得)や農家圃場の施肥試験(図2)における施肥量を説明変数とし、収量(研究テーマBで予測)を応答変数とする収量応答モデルを構築した。



図2. 大規模区画圃場(各2ha)における施肥試験の空撮画像

学習モデルに対して、施肥量を可変させる感度分析を実施し、生産者への栽培指導の指針となる経済的最適施肥量を推定した。予備実験的にランダムフォレストやサポートベクター回帰

など複数のアルゴリズムを用いて収量応答モデルを構築し、同様に感度分析を実施したが、アルゴリズムや説明変数の組み合わせによって、推定される経済的最適施肥量が大きく変動する(図 3)ことから、機械学習モデルによる因果推論には、大きな不確実性が存在すること、農家への処方箋提示が困難であることを明らかにした。研究計画当初では、空間的な土壌特性や収量変動を考慮するために、CNN による収量応答モデルの構築を想定していたが、本研究におけるデータ収集を待たずして、University of Illinois Urbana-Champaign との共同研究によって、大規模な合成データセットを整備することができたため、これを対象に CNN による収量応答モデルを構築した。その結果、収量予測精度は優れていたが、実際の因果推論には全く用いることができないことを明らかにした。機械学習の分野において、モデルによる因果推論の能力を重視するアルゴリズムとして Causal machine learning と呼ばれるアプローチが近年、開発・整備されている。今後は Causal machine learning の手法を、作物の収量応答性の評価に用いることができるかを検討する必要がある。また、実際の農家圃場データにおいては、物理的に同じ場所に異なる処理を設定できないため、因果推論の精度を検証できない。そこで、作物モデル WOFOST を用いて、解釈可能なパラメータを可変させることで、より現実的な合成データを生成する手法を開発した。本データは、機械学習の交差検証や学習データとしても利用することができるため、その有用性について今後は検討する必要がある。

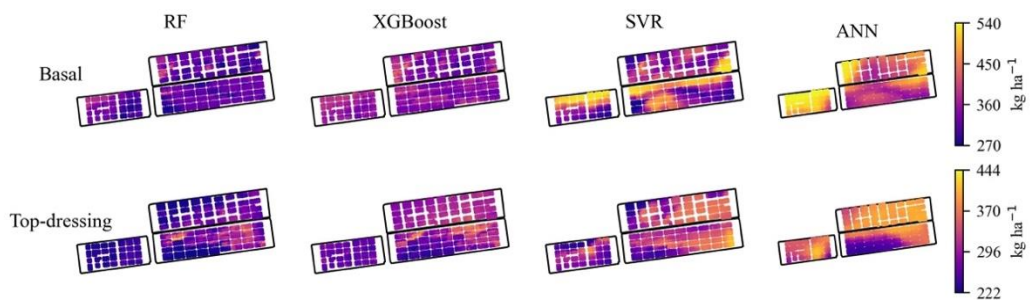


図 3. 異なる学習器が算定した経済的最適施肥量マップ。各学習器が、基肥と追肥に関して経済的に最適であると判断した施肥量を空間的に示したが、いずれも異なるシナリオを生み出している。

### 3. 今後の展開

2022年6月の一ヶ月間、オランダの Wageningen University & Research に渡って、空間統計モデルの専門家である Heuvelink 氏や作物モデルの開発者である de Wit 氏らから本研究を遂行する上で、貴重な指針・助言を受けることができた。今後もコンタクトを継続しつつ、機械学習モデルの不確実性評価や作物モデルとの統合的な解析手法の開発など、従来の研究計画にはなかった新たな着想を検証する予定である。また、生産者等を対象とするアウトリーチ活動を積極的に行ってきたが、それに興味を示す生産者も少なくない。本研究は1農家のみを対象としたパイロット的な試験研究であった。今後は、データを効率的に回収・解析するプラットフォームを構築することで生産者のネットワークを広く形成し、農学研究と開発技術の普及という相反する両輪を同時に駆動できる体制を構築する予定である。将来的には自動的に生産者への栽培指導が可能となるデータ解析プラットフォームを構築していきたい

と考えているが、特定の地域や作物種に限定したテーラーメイドな意思決定支援ツールの開発を10年の試行を経て開発していきたいと考えている。

#### 4. 自己評価

研究途上において、海外の研究機関との共同研究への発展もあり、研究手法に変更点も発生したが、概ね当初の研究目的の80%を達成できたと考える。実施体制として、オランダや米国との研究者との共同実施体制は今後も継続する予定である。また、農学分野においてもAI・機械学習の浸透が進みつつあるものの、医療や工学などの他分野と比較すると、そのスキルをフルに活用できるフィールドや人材は枯渇している。自分自身の研究成果の発信を通じて、AIや機械学習を専門としない農学研究者との交流が増えた。今後は、そのような研究者との共同研究に発展させ、そこで育成した農学とデータサイエンスに精通した人材を社会に輩出することで、広く社会・経済への波及効果が期待される。また、その対象は、研究者だけでなく、生産者や営農指導に携わる方々にも関連する。これまで幾度か本研究内容を紹介する一般向けセミナーの招待講演等も国内外で行ってきたが、実際に営農データを活用する現場の人々への普及や教育に若干なりにも貢献できたが、農業生産の現場と研究フィールドとの乖離は非常に大きく、アウトリーチが不十分であると考えている。

#### 5. 主な研究成果リスト

##### (1) 代表的な論文(原著論文)発表

研究期間累積件数: 4件

1. Tanaka, T.S.T., Tanabe, R., Matsui, T. Can on-farm experiments benefit intensive farming systems with small- to moderate-scale fields? Precision Agriculture '21, 2021, 861-867.
収量センサや可変管理を用いた農家圃場における栽培試験は、大規模農家の施肥や播種などの栽培管理を最適化することに貢献する。一方で、中小規模の農家を対象とした研究は広く行われてこなかった。本研究では、日本において UAV リモートセンシングによる圃場内の収量予測技術と簡易的な施肥試験による取り組みを実践した。その結果、多くの農家圃場において、農家慣行よりも施肥量を削減することで経営の最適化に貢献することを明らかにした。
2. Tanaka, T.S.T., Matsui, T. Explainable machine learning for site-specific yield response modeling using Moran's eigenvector maps. Proceedings of #OFE2021, 2022, 357-362.
近年の研究において、農家圃場における局所作物収量応答モデリングには機械学習がよく用いられる。局所的な収量応答性を機械学習モデルに学習させるために、土壌要因をサロゲートする座標データを用いた場合の説明可能性に及ぼすリスクについて言及した。座標データを直接的に説明変数に用いたモデルは過学習すること、座標データの変換値である Moran's eigenvector maps を説明変数とするモデルは比較的、研究者の仮説の範囲内での収量応答性を示した。
3. He, J., Zhou, X., Matsui, T., Li, F., Tanaka, T.S.T. Critical reevaluation of an efficient sampling

公開

design for assessing soil properties using bootstrap sampling and geostatistical analysis in Japanese large-scale paddy fields. *Soil Science and Plant Nutrition*, 2022. 68(5-6), 536-546

土壌分析は、作物の栽培管理を指導する上で重要な情報となる。圃場の大区画化に伴い、従来の小規模なサンプリングでは、圃場を代表する分析値を得られるとは言い難い。本研究では、サンプル数やサンプリング距離などに応じて、分析結果にどれほどのエラーを生じさせるのかを検討した。その結果、多くの土壌特性値において、従来のサンプリング手法では十分な分析精度を得ることが困難であることを示し、分析値の不確実性を念頭に栽培の計画や指導を行う必要がある。

(2)特許出願

該当なし

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

田邊良也, 松井勤, 田中貴. UAV リモートセンシングと深層学習によるコムギ収量予測モデルの精度検証. 日本作物学会第 251 回講演会. 2021.

田邊良也, 何景雲, 松井勤, 田中貴. 機械学習を用いたコムギにおける可変量施肥の有効性の検証. 日本作物学会第 253 回講演会. 2022.

何景雲, 松井勤, 田中貴, 李伏生. FTIR スペクトルと地球統計学的分析による水田における土壌特性の空間変動の評価. 日本作物学会第 253 回講演会. 2022.

田中貴. 大規模畑作における局所管理の最適化に向けた精密農業の現状. 園芸学会東海支部シンポジウム. 2022.

田中貴, 横山結衣. 局所作物収量応答モデリングの因果推論の交差検証に向けた WOFOST モデルによる合成データの生成. システム農学会 2022 年度一般研究発表会