

科学研究費助成事業 研究成果報告書

令和 3 年 6 月 28 日現在

機関番号：82111

研究種目：基盤研究(C)（一般）

研究期間：2017～2020

課題番号：17K08037

研究課題名（和文）深層学習モデリングによる広域穀物単収予測手法の構築

研究課題名（英文）Building an early prediction method of US crop yields based on machine learning algorithm

研究代表者

坂本 利弘（Sakamoto, Toshihiro）

国立研究開発法人農業・食品産業技術総合研究機構・農業環境変動研究センター・上級研究員

研究者番号：20354053

交付決定額（研究期間全体）：（直接経費） 3,700,000円

研究成果の概要（和文）：米国産農作物の作柄早期予測技術の確立を目指し、1．作物フェノロジー把握技術の汎用性向上（8作物36生育ステージに適用可能）、2．気象環境条件を考慮した作柄早期予測手法の高精度化（トウモロコシ・大豆単収予測精度の向上）、3．早期作付分類手法の高度化（ミクセル分解、分類精度向上）を行った。これにより、高頻度観測衛星センサデータ（MODIS）および気象環境データを複合利用する機械学習アルゴリズムを活用した新たな予測手法を構築した。

研究成果の学術的意義や社会的意義

日本は、輸入トウモロコシ・大豆の約7割、輸入小麦の約5割を米国からの輸入に依存している。また、世界的な食料需給情勢の不安定化を背景に、国際的な政策協調として、世界の農業・食料市場に関する正確かつ透明な情報を取得するための衛星リモートセンシング技術を用いた監視ネットワークの構築が推進されている。本研究成果は、作柄早期予測を確立するための基盤的な知見を提供するとともに、国内外の食料安全保障に資する技術としても活用が期待される。

研究成果の概要（英文）：This study aimed to establish early crop yield prediction technology for U. S. crops. Firstly, the versatility of crop phenology detection method was improved to be applicable to 36 growth stages of 8 crops. Secondly, the early crop yield prediction method was improved by considering weather and environmental conditions. Then, the prediction accuracy of corn and soybean yield was improved. Finally, the early crop classification method was modified to enable estimation of crop coverage ratio within a MODIS pixel. Then, the crop classification accuracy was improved. Consequently, a new crop yield prediction method was developed in terms of using machine learning algorithm based on the combined use of high-frequency observation satellite data (MODIS) and meteorological environmental data.

研究分野：リモートセンシング

キーワード：機械学習 食料安全保障 トウモロコシ 大豆 小麦 フェノロジー 作柄予測 作付分類

1. 研究開始当初の背景

国連人口基金によると、世界人口は、2020年～2050年にかけて約20億人増加すると予測されている。世界の食料需給の不安定化にともなう国際食料価格の乱高下は、海外からの食料輸入に依存する国々にとって政情不安を引き起こす外的リスクである。一説によると、2010年の小麦価格の高騰が引き起こした食料不安に対する民衆デモが、大規模な反政府民主化デモに発展し、アラブの春と言われる2010-2011年のアルジェリア騒乱、チュニジアのジャスミン革命、2011年エジプト革命を引き起こしたと言われている。食料需給に関する情勢変化を背景に、2011年のG20農業大臣会合(パリ)において、国際食料価格乱高下に対処するための「食料価格乱高下及び農業に関する行動計画」が宣言され、世界の農業・食料市場に関する正確かつ透明な情報を取得するための手段として、衛星リモートセンシング技術の活用が推進されることとなり、世界農業地理モニタリングイニシアティブ(GEOGLAM)の取組み・運用が開始された。日本の食料供給の海外依存度は、トウモロコシ：100%、大豆：94%、小麦：84%と極めて高く、日本の輸入トウモロコシ・大豆の約7割、輸入小麦の約5割弱が米国で生産されたものである。米国产主要作物の作柄を把握するための衛星リモートセンシング技術の確立は、日本の食料安全保障の強化するための重要な課題である。

高頻度観測衛星センサ(MODIS)データを活用した米国产トウモロコシ・大豆の作柄予測研究は、国内外の多くの研究者が取り組んでおり、植生指数を説明変数とするトウモロコシ・大豆の作柄予測モデルが提案されている。筆者もまた、MODIS植生指数を説明変数とする作柄早期予測モデルを考案するも(Sakamoto et al. 2013)、トウモロコシのみを対象としており大豆に対するモデル検証には至っていなかった。一方、Bolton et al. (2013)およびJhonson (2014)の研究によると、「MODIS植生指数から大豆単収を精度良く推定可能」であることが示されていたが、筆者による予備解析によると、「発芽から59日後の植生指数と大豆単収との間に高い相関関係」があることは追認されたが、その関係式だけでは、2003年産大豆の平均単収を23.6%も過大に予測するという問題を確認した。これは、8月以降に発生した早魃が生殖生長期の大豆を直撃したことに原因があると考え、衛星リモートセンシングデータのみによる収量予測には限界があるという結論にいたった。以上のような背景から、本課題では、気象再解析データ(NLDAS-2)等などの気象環境変動に関する時空間データプロダクトを予測モデルの入力変数として新たに組み込むことで、多様なデータを渾然一体的に処理し、膨大なデータに隠された収量変動と密接な関係にある特徴量を探索することが可能な機械学習型の予測モデルの構築を試みた。

2. 研究の目的

米国产トウモロコシ・大豆・小麦を対象とした早期作柄予測モデルの高度化を目指し、MODISデータおよび気象環境データを入力変数とする機械学習アルゴリズムを活用した新たな予測手法を提案することを目的とする。新たに気象環境データを説明変数として考慮することにより、米国产作物の気象環境ストレス応答を内包した予測モデルを提示し、従来手法との精度比較、多年次での予測結果の検証を行う。本研究では、過去アーカイブデータをフル活用した、1.作物生育ステージ推定アルゴリズム、2.作柄早期予測アルゴリズム、3.早期作付分アルゴリズムを新たに構築し、大陸スケールの作物作柄早期予測を実現するための基盤技術を確立するとともに、その有効性および実用性能について評価する。

3. 研究の方法

(1) データ

米国航空宇宙局の地球観測データ配信サイト(EARTHDATA, <https://search.earthdata.nasa.gov/>)から、MODIS/Terra, Aqua 地表面反射率8日間コンポジットデータ(MOD09A, MOD09Q, MYD09A, MYD09A)を収集した。リモートセンシング解析ソフト(ENVI/IDL)を用いて、フォーマット変換、植生指数(WDRVI)への変換等の前処理を行った。

米国航空宇宙局の気象再解析データ配布サイト(Land Data Assimilation System, <https://ldas.gsfc.nasa.gov/>)から、アメリカ本土を対象とした再解析プロダクト(NLDAS-2 Forcing, Models)を収集した。netCDF形式ファイルのインポート、投影変換、時別データから日別データへの集約等の前処理を行った。NLDAS-2プロダクトから抽出した、降水量、平均気温、短波放射、土壌水分データを用いた。

米国農務省統計データ配布サイト(USDA/NASS Quick Stats, <https://quickstats.nass.usda.gov/>)から、州レベル生育進度データ、郡レベル単収・収穫面積・収穫量データを収集した。shape形式ベクターデータへの変換等の前米国農務省統計局データ配布サイト(USDA/NASS http://www.nass.usda.gov/Research_and_Science/Cropland/Release/)から、米国作付分類マップを収集した。投影変換、クラス再分類等の前処理を行った。

(2) 作物フェノロジー把握手法

従来手法(Shape Model Fitting法 : Sakamoto et al. 2010)は、作物フェノロジーの圃場観

測データをもとにパラメータチューニングを行うため、圃場観測データの無い作物への応用展開が困難であった。本研究は、米国の主要作物 8 種類(トウモロコシ,大豆,冬小麦,春小麦,大麦,ソルガム,水稻,綿花)の生育ステージ(出穂・成熟等)及び農作業(播種・収穫)の時期を把握するために、圃場観測データに依存しないパラメータチューニング方法を新たに考案することで、改良型作物フェノロジー把握アルゴリズム(refined Shape Model Fitting 法)の確立を試みた。

(3) 気象環境データを考慮した作柄早期予測アルゴリズム

従来手法は、次に示す処理過程を経て、トウモロコシの単収を予測する。A. 時系列植生指数(WDRVI)の作成。B. Shape Model Fitting 法による生育ステージの把握。C. トウモロコシ作付地の把握。D. 絹糸抽出日の 7 日前の植生指数(WDRVI)から最終的な子実収量の予測。前述のように、植生指数を説明変数とする単収予測モデルは、大豆の単収推定にも有用であるとされるが、生育期間中の気象ストレス等の外的環境要因による減収効果をは考慮されていないため、開花と葉面積拡大が同時進行する無限伸育型の大豆については、干ばつによる減収被害をトウモロコシよりも受けやすく、予測結果が大きく外れることがある。本研究は、2000~2018 年の主要生産地域 5 州を対象に、機械学習アルゴリズムを活用した、気象環境情報と MODIS 植生指数データおよびフェノロジーデータを融合利用した新たな作柄予測モデルを構築し、トウモロコシに加え、新たに大豆を対象とした早期予測モデルを開発するとともに、衛星リモートセンシングデータのみを説明変数とする従来手法との比較を行った。予測モデルの改良は以下の手順により行った。A) 非灌漑地(灌漑率 10%未満)のみを対象に、郡別統計単収と WDRVI の相関関係が最も高い生育ステージを探索し、3 次近似式を従来型の単収予測モデルとして作成した。B) 従来型の単収予測モデルの予測誤差を目的変数として、植生指数および外部環境要因(再解析気象データ, 灌漑面積率)を説明変数とした誤差推定モデルを機械学習アルゴリズムにより構築した。機械学習アルゴリズムに「Deep Learning(深層学習)」と「Random Forest Regression」の利用を検討した。データセット 19 年分のうち、18 年をモデルキャリブレーション、残り 1 年をバリデーションとして、19 回繰り返し、予測シミュレーションを行った。

(4) 過去の土地利用分類図および機械学習アルゴリズムを活用した早期作物分類手法の高度化

畑作物を対象とした単収早期予測システムには、準リアルタイム作物分類手法の確立が課題となっていた。そこで、2008~2018 年の米国中西部 5 州を対象に、過去の作物分類図(CDL)および rSMF 法による推定発芽日を説明変数とする random forest regression アルゴリズムを利用した早期作物分類手法を検討した。本法は、従来手法では困難であった土地被覆率の推定アルゴリズムを新たに提案し、トウモロコシ・大豆・その他作物・Mixture の作付け分類を準リアルタイムで推定することができるように設計した。準リアルタイム予測を想定した上で 32 日間ごとの作付分類図を作成した上で、従来手法との精度比較を行った。

4. 研究成果

(1) 時空間データベースの構築・可視化データの公開

フォーマットの異なる MODIS/Terra, Aqua 地表面反射率 8 日間コンポジットデータ、気象再解析データ(NLDAS-2)、米国農務省統計データを地理空間情報として一体的に解析できるように、フォーマット変換、投影変換・ジオリファレンス処理等を行い、時空間データベースとして整備拡充を行った。MODIS/Terra, Aqua データについては、短波長赤外、近赤外、赤の反射率の組み合わせで可視化したフォールスカラー画像に変換し、誰でも簡単に米国本土の作柄状態・時系列変化をインターネット地図上で確認することができるように、セキュアな HTTPS プロトコル通信に対応した専用の閲覧サイト(Gaen-View: 世界の農業環境閲覧システム <https://gaenview.rad.naro.go.jp/>)において公開した。2018 年と 2019 年の 7 月下旬頃の状況を比較した例を図 1 に示す。2019 年のオハイオ州北西部(赤枠)において、緑色よりもピンク色の領域が広く見受けられ、前年よりも作物の生育状態が悪いことが推察される。オハイオ州の農業統計データによると、2019 年のトウモロコシ・大豆単収は、前年比で 12.3%, 12.5% 減収であった。高頻度観測衛星データの可視化画像を閲覧するだけでも、大規模不作地域の監視・確認が可能になる。

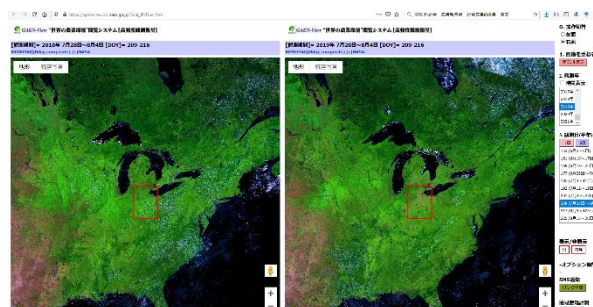


図 1. 2018 年と 2019 年の 7 月下旬頃の可視化衛星画像の比較

(2) 作物フェノロジー把握アルゴリズムの高精度化

本研究では、対象作物が最も盛んに作付けされている地域を 2008 年の CDL から探索し、各作物の平均的な時系列 WDRVI プロファイルを Shape Model とし、2008 年の NASS-CPR に記載された生育ステージを最も精度良く一致するためのパラメータチューニング方法を考案した。rSMF

法により、米国産主要農作物 8 種類の計 36 の生育ステージを推定することが可能になった(図 2)。2009~2016 年の統計データ(NASS-CPR)に対する SMF 法の推定精度を検証した結果、作付が盛んでない州では、推定精度が極端に低くなるという傾向が確認されたものの、春小麦については 8 州(作付面積:全米の 99.8%)において、綿花については 6 州(作付面積:全米の 80.0%)において、各生育ステージを精度良く推定することができた。各生育ステージの二乗平均平方根誤差(RMSE)は次の通り。春小麦-播種日(RMSE=8.1days)、出芽日(5.6days)、出穂日(3.9days)、収穫日(6.7days)、綿花-播種日(RMSE=9.7days)、squaring(6.5 days)、setting bolls(5.8 days)、bolls opening(6.9 days)、収穫日(6.1 days)。MCD12Q2 プロダクトと統計データ(2008~2014年)の比較結果は、春小麦で RMSE=7.6~14.1 days、綿花で RMSE=9.6~17.2 days であり、作物フェノロジーに特化した SMF 法のほうがより精度良く推定可能であるということが示された。作物フェノロジー把握手法を高度化したことで、小麦を含む様々な農作物を対象とした作柄早期予測手法の開発の筋道を確立した。本研究成果は、原著論文として著名な国際誌に掲載され(Sakamoto 2018)、2020 年までの過去 3 年間で急速に被引用件数が伸びるなど(Google Scholar: 23 件、Web of Science:20 件)、国際的な注目を受けている。

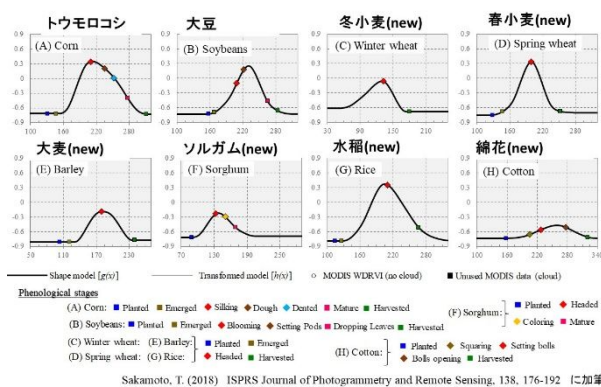


図 2. rSMF 法による 8 作物 36 生育ステージ把握

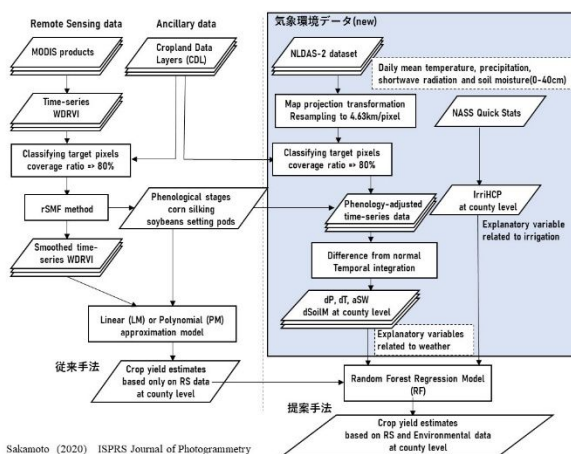
(3) 機械学習アルゴリズムを用いた作柄早期予測モデルの高度化

2000~2018 年の主要生産地域 5 州を本研究の対象地域において、WDRVI と郡別統計データ(単収、灌漑率 10%未満)の回帰モデルを直線近似式から 3 次近似式に変更し、その相関係数を生育ステージ別に調査した結果は以下の通りであった。

[トウモロコシ]: $R^2=0.852$ Yield=21.1WDRVI3+8.7WDRVI2+16.2WDRVI+9.1

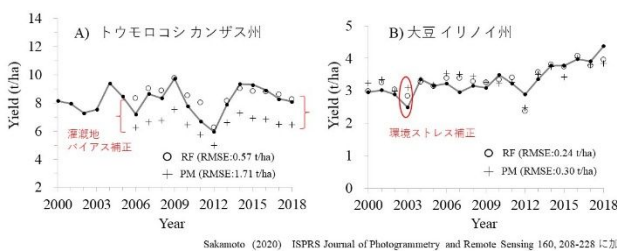
[大豆]: $R^2=0.729$ Yield=24.7WDRVI3+29.1WDRVI2+16.4WDRVI+5.2

説明変数に用いる WDRVI の生育ステージは、トウモロコシ:絹糸抽出期 13 日前、大豆: Setting Pods stage 6 日前であり、いずれも植生指数が最大になる生育ステージよりも早い時期にあたる。予備解析を行ったところ、Deep Learning による単収推定誤差(RMSE)は、トウモロコシ: 0.75t/ha および大豆: 0.35t/ha であるのに対して、random forest regression による推定誤差は、トウモロコシ: 0.70t/ha および 0.26t/ha であった。当初の想定とは異なり、random forest regression を用いた推定モデルのほうが、気象環境データによる精度向上効果も認められた。さらに、random forest regression を用いた推定モデルのほうが、モデル選択の容易さ、計算時間の短縮、推定結果の堅牢性の点においても優位性があり、当初想定していた方針(Deep Learning の利用)を変更し、random forest regression を用いた推定モデルを作成した(図 3)。予測誤差推定に用いる説明変数には、気象環境データすべてを用いる場合が最も良い推定精度を示すわけではなく、州によって最適な説明変数の組み合わせがあることが明らかになった。Kansas 州, Missouri 州, Nebraska 州については灌漑面積率が大きく寄与し、Illinois 州, Iowa 州については平均気温、降水量が推定精度改善に寄与した。予測誤差(RMSE, 州レベル)は、トウモロコシ: 従来型 0.897t/ha, 改良型 0.539t/ha, 大豆: 従来型 0.283t/ha, 改良型 0.206t/ha と、いずれの場合においても



Sakamoto (2020) ISPRS Journal of Photogrammetry and Remote Sensing 160, 208-228 に加筆

図 3. 改良型単収予測モデルのフローチャート



Sakamoto (2020) ISPRS Journal of Photogrammetry and Remote Sensing 160, 208-228 に加筆

図 4. トウモロコシ・大豆単収予測結果と統計値との時系列比較

改良型モデルの予測誤差が小さかった。以上のことから、random forest regression algorithm による外部環境因子を考慮したモデル改良により、トウモロコシ・大豆単収の予測精度の改善が認められた。新たに気象環境変数(灌漑率, 気温, 降水量)を予測モデルに組み込むことで、灌漑地帯におけるバイアス誤差補正が容易になるとともに、一部の地域で大豆の干ばつストレスによる予測誤差を改善できることを確認した(図4)。本研究成果は、原著論文として著名な国際誌に掲載され(Sakamoto 2020)、2020年までの1年間で急速に被引用件数が伸びるなど(Google Scholar: 16件, Web of Science:8件)、国際的な注目を受つつある。

(4) 早期作付分類アルゴリズム

rSMF法から推定された発芽日と過去2年分の土地利用データを入力値とする、作物分類アルゴリズムを作成した(図5)。提案手法は、random forest regression アルゴリズムを使うことで、MODIS 250m 画素内のトウモロコシ、大豆、その他の作物の3種類のミクセル率を推定可能になり、そのミクセル率を基準にトウモロコシ、大豆、その他、Mixtureの4種類に分類することが可能になった。過去2年分のCDLデータとrSMF法による発芽日を説明変数とするType 04 (hybrid)法を用いる場合、画素内のミクセル率推定精度・作付分類マップの分類精度が最も良かった(図6,7)。2018~2019年を対象とした分類精度は、Overall accuracy: DOY215時点で65.3~68.3%、DOY279時点で69.7~72.0%であった。MODISデータ観測期間が長くなるほど分類精度が向上するという傾向を示した。トウモロコシ、大豆のUser's accuracy(UA), Producer's accuracy(PA)は、DOY215

で、トウモロコシ UA: 72.2~72.8%, PA:70.0~76.1%, 大豆 UA: 69.2~77.2%, PA:60.5~62.7%であった。単収早期予測のための早期分類アルゴリズムとして十分な精度を有することを確認した。これら成果は、原著論文としてとりまとめ海外専門雑誌に受理され、現在印刷中である(Sakamoto in press)。

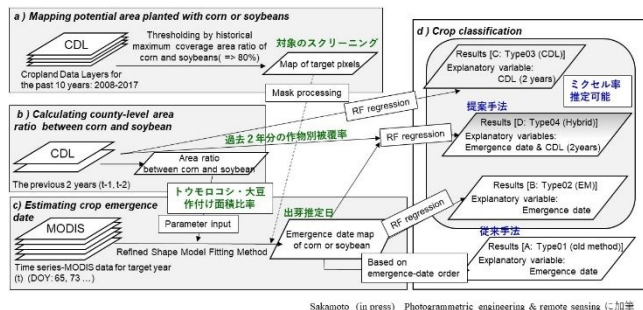


図5. 機械学習アルゴリズムを用いた土地被覆率推定および作付分類フローチャート

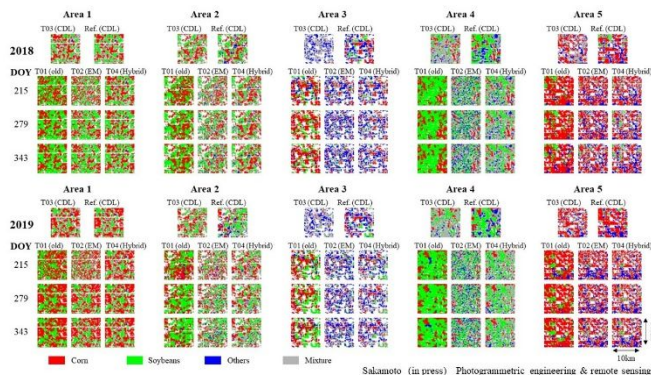


図6. MODIS観測期間およびアルゴリズムによる作物分類結果の拡大比較

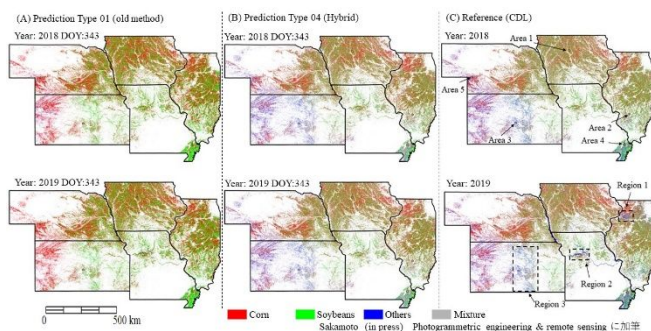


図7. 作物分類図。従来手法(A), 提案手法(B), 参照図(C)

<引用文献>

Sakamoto, T. et al. 2013. MODIS-based corn grain yield estimation model incorporating crop phenology information. Remote Sensing of Environment, 131, 215-231.

Bolton, D. K. et al. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and Forest Meteorology, 173, 74-84.

Johnson, D. M. 2014. An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sensing of Environment, 141, 116-128.

Sakamoto, T. in press. Early classification method for U.S. corn and soybean by incorporating MODIS estimated phenological data and historical classification maps in random forest regression algorithm. Photogrammetric Engineering & Remote Sensing.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Sakamoto Toshihiro	4. 巻 160
2. 論文標題 Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm	5. 発行年 2020年
3. 雑誌名 ISPRS Journal of Photogrammetry and Remote Sensing	6. 最初と最後の頁 208 ~ 228
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.isprsjprs.2019.12.012	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Sakamoto Toshihiro	4. 巻 138
2. 論文標題 Refined shape model fitting methods for detecting various types of phenological information on major U.S. crops	5. 発行年 2018年
3. 雑誌名 ISPRS Journal of Photogrammetry and Remote Sensing	6. 最初と最後の頁 176 ~ 192
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.isprsjprs.2018.02.011	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 坂本利弘
2. 発表標題 機械学習による作物単収予測モデルの高精度化
3. 学会等名 日本写真測量学会 令和元年度秋季学術講演会
4. 発表年 2019年

1. 発表者名 坂本利弘
2. 発表標題 Shape Model Fitting法による米国産作物のフェノロジー推定
3. 学会等名 日本リモートセンシング学会 第63回（平成29年度秋季）学術講演会
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------