

特集号招待論文

ドローンによる作物の表現型計測と機械学習による作物バイオマス・収量の予測

辰己賢一¹

¹東京農工大学

近年、付加価値の高い農作物の生育状況の把握や収量を予測するためにドローンや衛星リモートセンシングによる空撮画像が使われている。一方、収量に大きな影響を及ぼす空撮画像から得られる有効な説明変数はいまだ明確になっていない。本稿では、空撮画像からトマトの各株における草高や植生指数を計測し、各株ごとにそれらの平均値や分散などの1次元計測値、画像のきめや画素間の空間パターンを考慮できるテクスチャ情報を2次元計測値としてそれぞれ計測した。次に、得られた計測値から計5種類の変数選択法によってトマトの果実部を除く地上部バイオマス重（以下、バイオマス）、果実重、果実数にそれぞれ影響の大きい計測値を選択し、選択された計測値を説明変数とし、3つの機械学習モデルを用いて、バイオマス、果実重、果実数の予測を試みた。その結果、1次元計測値だけでなく2次元計測値を説明変数候補として考慮することで予測精度が大幅に向上すること、収穫の約1カ月前の2次元計測値が果実重や果実数を予測する上で影響の大きい説明変数であることが明らかとなった。この結果は、収量予測におけるデータ取得の効率化に大きく寄与するものである。

1. ドローンの農業活用

トマト (*Solanum lycopersicum* L.) は、世界で広く栽培されている野菜の1つであり、人間の健康維持に重要な役割を果たしている。生鮮トマトの世界生産量は約1億8,000万トンであり、そのうち約4分の1は加工用として栽培され、ケチャップ、サルサ、ジュースなどの形で消費されている[1]。主な生産国は中国、インド、パキスタン、トルコ、アメリカで、これらの国で世界の生産量の約60%を占めており、生産量と収穫面積は年々増加している。

近年、植物の草高や葉の形態などの形質を広範囲で測定する手段として無人航空機（以下、ドローン）が注目されている。ドローンは、操作が比較的容易で、携帯性に優れ、飛行経路や飛行速度、撮影高度や撮影間隔の制御を事前設定でき、圃場^{☆1}における生育状況の把握等に用いられる高分解能の画像を効率よく取得できる利点がある[2],[3]。また、ドローンには、RGBカメラ[4]やマルチスペクトルカメラ[5]など、農業用途に役立つさまざまなセンサを搭載することができ、農業と情報の技術融合に新しい視点をもたらすことが期待されている。ドローンを使ったト

マトの生育状況の把握を試みた研究には、RGBおよびマルチスペクトルの時系列空撮画像を用い、機械学習モデルによって群落における葉面積や成長速度、収量、果実数の予測可能性を調べたものがある[6]。しかしながら、既往の研究は、群落単位での評価を試みたものであり、より詳細な株単位でのトマトの収量予測を試みた研究はない。

ドローン空撮画像により得られた作物の生育期間中における表現形質を用いた収量予測研究では、機械学習アプローチが用いられ、有用な知見が得られている[7]。一方、多時期の表現型データセットを収集するには、一般に時間と人的労力を要する。したがって、トマトのバイオマスや収量の予測にとって重要となる生育時期の主要な表現形質を効率よく得ることができれば、データ収集とその解析処理にかかる労力は大幅に削減可能となる。ドローン空撮撮影により得られる表現形質と機械学習を用いた作物の収量予測のために最適な説明変数の検討は比較的数量多く実施されているが、トマトの株単位におけるバイオマス、果実重、果実数の予測のための重要な変数の探索・選択を試みた研究はほとんどない。したがって、これらの予測にとって重要となる変数を選択し、機械学習アルゴリズムを用いることで、トマト収量の予測を高精度で実施することを可能にする技術の確立が期待できる。また、育種サイクルの短縮に資する情報の抽出が期待できるとともに、圃場におけるデータ収集やその前処理にかかる労力が軽減される。本稿では、ドローン空撮画像から得られる草高^{☆2}や植生指数^{☆3}マップから、バイオマス等の予測に影響の大きい変数群を変数選択法によって選択し、選択した変数群を用いて複数の機械学習アルゴリズムによってトマトのバイオマス、果実重、果実数の予測を試みた結果を報告する。

2. 圃場栽培と空撮画像取得

2.1 栽培実験圃場

栽培試験は、露地栽培に適した加工用トマト（品種：なつのしゅん）を用い、2020年5月13日から7月30日まで、東京農工大学フィールドミュージアム府中の実験圃場において図1に示す3反復（各区5m × 5m）で実施した。なお、移植の1カ月前に温室内で育てた苗床を圃場に移植し、トマトの苗は支柱等で固定せずに栽培を実施した。栽植密度は、畝間^{☆4}0.85m、株間0.40mとした。施肥は、移植前に元肥（N : P : K = 10 : 10 : 10kg 10a⁻¹）を与え、移植後1週間は、各区画に灌漑チューブを用いて500mlの点滴灌漑を朝夕の計2回、30分ずつ実施し、以降は雨水のみによる栽培を行った。また、各プロットには農業用のプラスチックマルチフィルムを用い、定期的に手作業で除草を実施した。

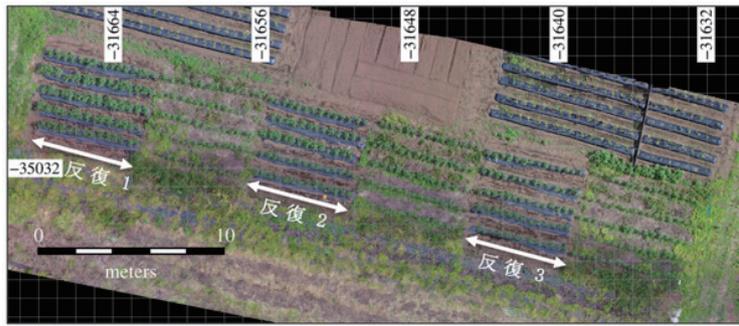


図1 栽培試験圃場（本図の撮影日は2020年6月18日）

2.2 ドローンによる圃場空撮

DJI Matrice 210V2（DJI Co.,Ltd., Shenzhen, China）に搭載したRGBカメラ Zenmuse X5S（DJI Co.,Ltd., Shenzhen, China）とマルチスペクトルイメージセンサカメラAltum（MicaSense Co.,Ltd., SEA, USA）を用いて、RGB画像とマルチスペクトル画像を取得した（図2）。各カメラの静止画解像度は、RGB画像が5,280 × 3,956 pixel，スペクトル画像（長波長赤外線を除く）は2,064 × 1,554 pixelである。



図2 ドローン本体およびセンサカメラ

トマトの株がドローン機のプロペラ回転による風の影響を受けず、かつ高解像度の画像を用いて各株の変数を選択する目的から、撮影高度は地上12mとし、また、オーバーラップ率およびサイドラップ率をそれぞれ90%、70%に設定した。空撮日は2020年の5月24日、5月30日、6月5日、6月11日、6月18日、6月26日、7月2日、7月12日、7月16日、7月24日である。なお、飛

行経路を含む飛行パラメータの設定は、画像から高精度の3次元マップ点群情報の生成を可能にする写真測量ソフトウェアであるPix4Dcapture (Pix4D S.A., Lausanne, Switzerland) を用いて行った。本研究では、RGB空撮画像および地上基準点データから数値地形モデル (DTM) および数値表層モデル (DSM) マップを作成し、スペクトル画像については、5バンド (青, 緑, 赤, レッドエッジ, 近赤外) の反射率マップを作成した。

2.3 バイオマス・収量の計測

トマトバイオマス・収量に影響を及ぼす説明変数の選択, モデルの学習, 予測精度評価のため, 各株ごとにバイオマス, 果実重, 果実数の計測を7月29日と7月30日の両日に実施した。

2.4 草丈と植生指数の計測

各株の草高は, 2.2節で得られたDSMからDTMを減算することにより算出した。また, 反射率マップから, 葉のクロロフィル量や成長の指標としてよく用いられる[8],[9], 3つの植生指数: 緑の正規化植生指数 (Green Normalized Difference Index ; GNDVI) [10], 正規化植生指数 (Normalized Difference Vegetation Index ; NDVI) [11], 加重差植生指数 (Weighted Difference Vegetation Index ; WdVI) [12]を以下の式によりそれぞれ導出した。

$$\text{GNDVI} = (\text{NIR} - \text{Green}) / (\text{NIR} + \text{Green})$$

$$\text{NDVI} = (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$$

$$\text{WdVI} = \text{NIR} - a \times \text{Red}$$

ここで, NIR (Near Infrared Ray) は近赤外, Greenは緑色, Redは赤色領域の反射率を示す。aは, ソイルラインの傾きであり, NIRバンドとRedバンド間の線形関係性から得られる。なお, この3つの植生指数を選択した理由は, 1) 必要以上に多数の植生指数を用いると多重共線性の問題があるため, 必ずしも効率的でないこと, 2) NDVIとGNDVIは葉のクロロフィル量と相関があることが知られており, 収量予測に広く利用されていること, 3) WdVIは土壌バックグラウンド値の影響を考慮することができるため, である。

2.5 表現型計測

2.4節により得られた各株における草高および植生指数からバイオマスおよび収量に関連する可能性が高いと考えられる表現型を計測するために, 以下の前処理を実施した。

(1) トマト株部分の抽出: 5月24日撮影のオルソモザイク画像を用いて, NDVIの値が0.5以上の画素から作物体部分を抽出する。さらに, 雑草と認識できる一部の画素については手動で除去する

(2) 各トマト株の重心の決定: (1) で得られた株の閉じた輪郭ベクトルから, 各株の重心位置を決定する

(3) 各株の関心領域の抽出: (1) および (2) のプロセスにより導出した各株の重心位置を中心に半径20cm内に含まれる株の領域を対象領域として抽出する

以上のプロセスにより得られた撮影日ごとの各株の対象領域から, ピクセル計測値と動的成長率を抽出した。ピクセル計測値については, 草高マップと植生指数マップから平均 (AVE), 標準偏差 (SD), 歪度 (SKEW), 範囲レンジ (RANGE), 最大 (MAX) の計5つを1次元計測

値として算出し、次に、グレーレベル同時生起行列GLCM（Gray-Level Co-occurrence Matrix）によりテクスチャ解析を行い、空間パターンを考慮した13種類の2次元計測値（Sum Average（SA）、Entropy（Ent）、Difference Entropy（DE）、Sum Entropy（SE）、Variance（Var）、Difference Variance（DV）、Sum Variance（SV）、Angular Second Moment（ASM）、Inverse Difference Moment（IDM）、Contrast（Con）、Correlation（Cor）、Information Measures of Correlation-1（MOC-1）、Information Measures of Correlation-2（MOC-2））をそれぞれ算出した（表1）。

表1 本研究で算出したGLCM計測値

GLCM 特徴量	省略形	算出式
Sum Average	SA	$\sum_{k=0}^{2(N-1)} k P_{x+y}(k)$
Entropy	Ent	$-\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_d(i,j) \log(P_d(i,j))$
Difference Entropy	DE	$-\sum_{k=0}^{N-1} P_{x-y}(k) \log(P_{x-y}(k))$
Sum Entropy	SE	$-\sum_{k=0}^{2(N-1)} P_{x+y}(k) \log(P_{x+y}(k))$
Variance	Var	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-\mu)^2 P_d(i,j)$
Difference Variance	DV	$\sum_{k=0}^{N-1} \left(k - \sum_{k=0}^{N-1} k P_{x-y}(k) \right)^2 P_{x-y}(k)$
Sum Variance	SV	$-\sum_{k=0}^{2(N-1)} \left(k - \sum_{k=0}^{2(N-1)} k P_{x+y}(k) \right)^2 P_{x+y}(k)$
Angular second moment (Uniformity)	ASM	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_d(i,j)^2$
Inverse Difference Moment	IDM	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{1+(i-j)^2} P_d(i,j)$
Contrast	Con	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 P_d(i,j)$
Correlation	Cor	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_d(i,j) \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x \sigma_y}$
Information Measure of Correlation-1	MOC-1	$\frac{HXY - HXY1}{\max(HX, HY)}$
Information Measure of Correlation-2	MOC-2	$[1 - \exp\{-2(HXY2 - HXY)\}]^{1/2}$

ただし、 $p_d(i, j)$ は、行列 (i, j) の相対頻度。

$$p_x(i) = \sum_{j=0}^{N-1} p_d(i, j)$$

$$p_y(j) = \sum_{i=0}^{N-1} p_d(i, j)$$

$$p_{x+y}(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_d(i, j), \quad k = i + j = 0, 1, \dots, 2(N-1)$$

$$p_{x-y}(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_d(i, j), \quad k = i - j = 0, 1, \dots, N-1$$

$$HX = - \sum_{i=0}^{N-1} p_x(i) \log(P_x(i))$$

$$HY = - \sum_{j=0}^{N-1} p_y(j) \log(P_y(j))$$

$$HXY = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_d(i, j) \log(P_d(i, j))$$

$$HXY1 = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_d(i, j) \log(P_x(i) P_y(j))$$

$$HXY2 = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_x(i) p_y(j) \log(P_x(i) P_y(j))$$

$$\mu_x = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i p_d(i, j)$$

$$\mu_y = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} j p_d(i, j)$$

$$\sigma_x = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - \mu)^2 p_d(i, j)}$$

$$\sigma_y = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (j - \mu)^2 p_d(i, j)}$$

次に、連続する2撮影時期から得た草高と植生指数の変化量を撮影間隔の日数で除することにより、動的成長率を算出した。以上より、合計756個の計測値（計18種の1次元・2次元計測値 × 10撮影日 × 草高・3植生指数、草高および3植生指数の生育期間中における9種の動的成長率）を変数選択のために各株ごとに抽出した。

2.6 変数選択

計算の複雑さを軽減させ、効率的なデータ解析を実施するため、バイオマス、果実重、果実数に影響を与える重要な計測値やその時期を決定するために変数選択法により変数選択を実施した。この変数選択は、機械学習アルゴリズムや回帰モデリングにおける基本的なステップになる。本研究では、2.5節で計測した計756種の計測値を候補とした。なお、1次元計測値と2次元計測値を変数候補とした理由は、1次元計測値と動的成長率を用いた場合（以下、1次元計測値）と全計測値（1次元・2次元計測値、動的成長率）を用いた場合で、予測精度に差がどの程度生じるかについて調べるためである。

トマトの生育状態とその生育時期を考慮してバイオマス、果実重、果実数に影響の大きい変数を選択するため、Boruta[13]、DALEX[14]、Genetic Algorithm (GA) [15]、LASSO[16]、Recursive Feature Elimination (RFE) [17]の5つの変数選択法を用いた。Borutaは、ランダムフォレストをベースにしたノンパラメトリックな変数選択アルゴリズムで、変数の重要度を評価することができ、統計的に有意な変数の選択に有用である。DALEXは機械学習モデルで使用する変数について、損失関数などの属性を説明するノンパラメトリックな変数分析手法である。GAは、遺伝学や生物進化の仕組みに基づいてモデル最適化を行うためのノンパラメトリックな手法である。LASSOはL1ノルムを用いたペナルティにより、特定の不要な係数を削除することで、予測誤差を最小化するための変数を選択するパラメトリックな変数選択法である。RFEは、指定した変数の数量に達するまで重要度の低い変数を削除していくノンパラメトリックな手法である。本研究では、これらの5つの変数選択法を用い、重要度スコアが最も上位の5変数をトマトのバイオマス、果実重、果実数の予測に有用な変数としてそれぞれ採用した。

変数選択により選択された各変数群は、ランダムフォレスト (RandomForest ; RF) [18]、リッジ回帰 (Ridge Regression ; RI) [19]、サポートベクタマシン (Support Vector Machine ; SVM) [20]の3種の機械学習モデルの入力に用い、バイオマス、果実重、果実数の予測を実施した。なお、全実測データの80%をモデルの学習データとして使用し、残りの20%をモデルの評価に使用した。各モデルの予測性能評価は、決定係数 (R^2) と相対平均二乗誤差 (rRMSE) を用いて実施した。

3. トマト収量の予測結果

3.1 草高と植生指数の時系列マップ

図3および図4に、生育期間における草高とGNDVIの時系列マップをそれぞれ示す。草高は開花期まで直線的に増加し、その後は葉が水平方向に広がるため、草高の増加率は非常に小さく、結実後もほとんど変化しなかった。一方、GNDVIなどの植生指数もシグモイド型の値を取り、生育初期～開花期にかけては指数関数的に植生指数の値は大きくなるが、開花期以降は葉の老化や枯死などが要因となり緩やかに植生指数は低下する。

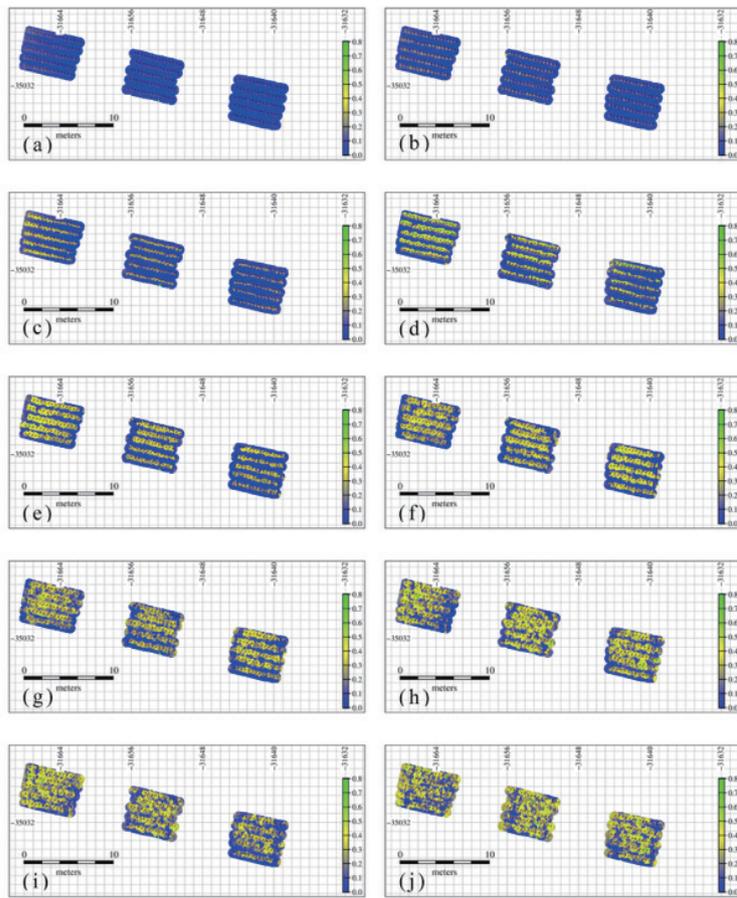


図3 草高 (m) の時系列変化 ((a) 5月24日, (b) 5月30日, (c) 6月5日, (d) 6月11日, (e) 6月18日, (f) 6月26日, (g) 7月2日, (h) 7月12日, (i) 7月16日, (j) 7月24日)

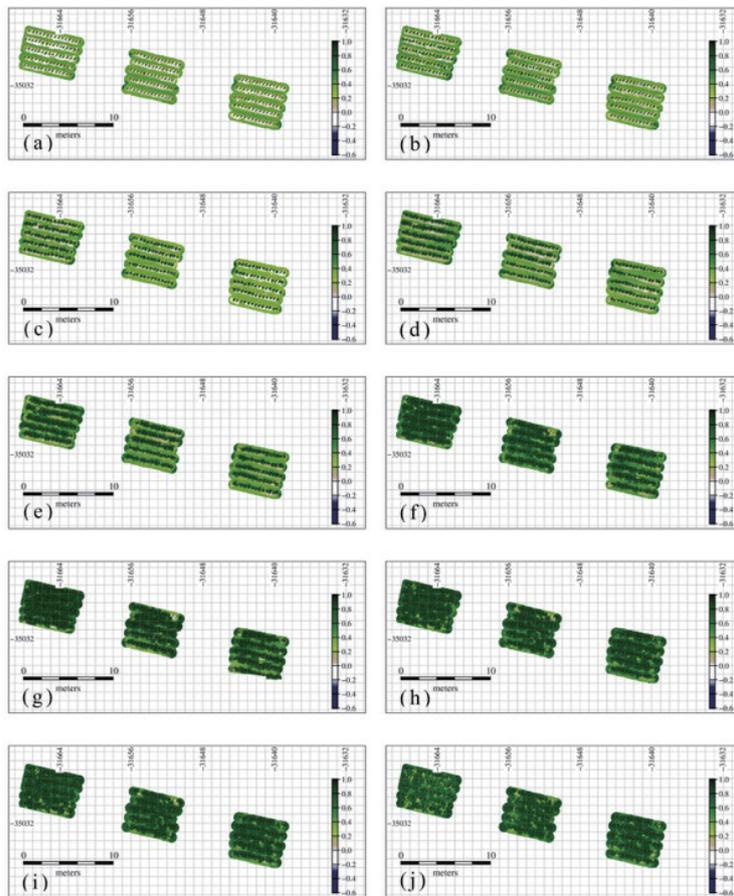


図4 GNDVI (-) の時系列変化 ((a) 5月24日, (b) 5月30日, (c) 6月5日, (d) 6月11日, (e) 6月18日, (f) 6月26日, (g) 7月2日, (h) 7月12日, (i) 7月16日, (j) 7月24日)

3.2 変数変択の結果

5つの変数選択法を用いて重要度スコアに応じて選択されたバイオマス、果実重、果実数の予測に大きな役割を果たすと認められた上位5つの変数をそれぞれ示す(表2, 表3, 表4)。バイオマスについては、果実発育中期(6月下旬から7月中旬)の草高と植生指数に関する1次元および2次元の計測値が選択された。用いた変数選択法により結果は異なるが、1次元計測値として草高のAVEとMAX、植生指数のRANGEが複数の選択法において選択され、1次元計測値として草高のMOC-1, MOC-2, NDVIのSVとDVが多くの変数選択法において選択された(表2)。これらのことから、バイオマスの予測には、草高と植生指数の1次元計測値およびGLCMによるテクスチャ計測値の両方が重要であることが明らかになった。特に後述する果実重と果実数における変数選択の結果と比較し、草高に関する変数が数多く選択されていることが分かる。これらの結果は、今回候補として取り上げた変数や変数選択法によって異なる可能性があるものの、本研究では、バイオマスの予測において、草高の1次元計測値の重要性が明らかになった。バイオマスの予測は、葉の光合成による同化能力を推定する上でも重要であるため、草高とバイオマスの関係性育種家や研究者にとって興味深い情報となる。以上、バイオマスの推定には、果実発育中期以

降の草高および植生指数の1次元および2次元の計測値が相対的に重要度が高い変数であることが分かった。次に、果実重の予測において、すべての計測値から変数選択モデルで選択された変数の多くは植生指数に関するものであった(表3)。Boruta, DALEX, GA, RFEによって1次元計測値のみおよび全計測値から変数選択した結果、6月18日におけるWDVIのRANGEが、重要な変数としてランクインしていることが分かる(表3)。さらに、NDVIのAVEは、GAを除くすべての選択モデルでランクインしていることが明らかになった。以上より、収穫の約1カ月前の植生指数に関する変数が果実重の予測にとって重要であることが明らかとなった。結実開始期の植生指数が最終的な果実重を決定づける重要な因子であることを示す本結果は、圃場管理において果実重を推定する上で注目すべき変数とその栽培時期を示すものであり、栽培管理を行う上で意義深い知見である。果実重と同様に、果実数も収量を決定する上で重要な要素である。上空からの撮影画像だけでは十分に確認することができない各株の果実数を推定する技術は、栽培管理に貢献できる。本研究では、特にNDVIまたはWDVIのAVEが1次計測値および全計測値を対象とした変数選択において、果実数と関係性が高い変数として選択された(表4)。また、果実発育期間の初期から中期にかけての植生指数に関する1次および2次の計測値は果実数の推定に有用な変数となっていることが分かる。一方で、バイオマスとは異なり、草高に関連する計測値は果実数の推定にとって相対的に重要ではないことが明らかとなった。

表2 1次元計測値および全計測値から変数選択法によって選択されたバイオマス予測のための重要度の高い変数

	1次元計測値から選択			全計測値から選択		
Rank	Feature value	Statistics	Date	Feature value	Statistics	Date
Boruta						
1	Plant height	AVE	0626	Plant height	MOC-1	0712
2	GNDVI	RANGE	0716	NDVI	SV	0712
3	Plant height	MAX	0720	NDVI	DV	0712
4	GNDVI	AVE	0712	Plant height	AVE	0702
5	NDVI	SD	0712	GNDVI	DV	0724
DALEX						
1	Plant height	AVE	0626	NDVI	SV	0712
2	Plant height	MAX	0702	Plant height	SE	0712
3	Plant height	AVE	0702	NDVI	SE	0712
4	GNDVI	MAX	0530	NDVI	DV	0712
5	GNDVI	RANGE	0530	Plant height	Ent	0626
GA						
1	Plant height	RANGE	0618	Plant height	RANGE	0712
2	Plant height	AVE	0626	NDVI	SV	0712
3	GNDVI	RANGE	0716	NDVI	IDM	0702
4	WDVI	MAX	0724	NDVI	DV	0712
5	NDVI	SD	0712	GNDVI	MOC-2	0724
LASSO						
1	Plant height	AVE	0626	Plant height	AVE	0626
2	GNDVI	RANGE	0716	GNDVI	DV	0724
3	NDVI	MAX	0716	NDVI	SV	0716
4	Plant height	MAX	0702	GNDVI	Con	0618
5	Plant height	SKEW	0605	Plant height	MAX	0702
RFE						
1	Plant height	AVE	0626	NDVI	SV	0712
2	NDVI	MAX	0716	Plant height	MOC-1	0712
3	Plant height	MAX	0702	NDVI	DV	0712
4	Plant height	AVE	0702	NDVI	MAX	0716
5	NDVI	SD	0712	GNDVI	DV	0724

表3 1次元計測値および全計測値から変数選択法によって選択された果実重予測のための重要度の高い変数

	1次元計測値から選択			全計測値から選択		
Rank	Feature value	Statistics	Date	Feature value	Statistics	Date
Boruta						
1	WDVI	RANGE	0618	WDVI	RANGE	0618
2	NDVI	AVE	0618	NDVI	AVE	0618
3	WDVI	AVE	0618	WDVI	AVE	0618
4	NDVI	AVE	0626	WDVI	SA	0618
5	GNDVI	AVE	0626	NDVI	AVE	0626
DALEX						
1	WDVI	AVE	0618	WDVI	SA	0618
2	NDVI	AVE	0724	NDVI	AVE	0626
3	WDVI	RANGE	0618	NDVI	AVE	0618
4	Plant height	RANGE	0618	WDVI	RANGE	0618
5	NDVI	AVE	0618	GNDVI	IDM	0712
GA						
1	WDVI	RANGE	0618	NDVI	IDM	0716
2	NDVI	MAX	0606	WDVI	RANGE	0618
3	NDVI	AVE	0618	GNDVI	SE	0724
4	NDVI	SD	0716	Plant height	Growth Rate	0530-0605
5	NDVI	SD	0524	WDVI	MAX	0606
LASSO						
1	NDVI	AVE	0618	GNDVI	Con	0618
2	Plant height	MAX	0724	Plant height	MAX	0724
3	NDVI	RANGE	0724	WDVI	SA	0626
4	NDVI	RANGE	0524	NDVI	AVE	0626
5	Plant height	SKEW	0712	NDVI	Cor	0712
RFE						
1	NDVI	AVE	0618	NDVI	AVE	0618
2	WDVI	RANGE	0618	WDVI	RANGE	0618
3	WDVI	AVE	0618	WDVI	AVE	0618
4	-	-	-	NDVI	AVE	0626
5	-	-	-	WDVI	SA	0618

表4 1次元計測値および全計測値から変数選択法によって選択された果実数予測のための重要度の高い変数

	1次元計測値から選択			全計測値から選択		
Rank	Feature value	Statistics	Date	Feature value	Statistics	Date
Boruta						
1	NDVI	AVE	0626	WDVI	RANGE	0618
2	GNDVI	AVE	0626	NDVI	AVE	0618
3	NDVI	MAX	0618	WDVI	AVE	0618
4	GNDVI	MAX	0611	WDVI	SA	0618
5	GNDVI	SD	0626	NDVI	AVE	0626
DALEX						
1	NDVI	RANGE	0606	WDVI	IDM	0618
2	NDVI	AVE	0626	NDVI	RANGE	0626
3	NDVI	AVE	0524	NDVI	AVE	0618
4	NDVI	MAX	0618	WDVI	AVE	0618
5	NDVI	MAX	0618	GNDVI	SA	0712
GA						
1	WDVI	SD	0712	NDVI	IDM	0716
2	NDVI	SD	0524	WDVI	RANGE	0618
3	WDVI	MAX	0606	GNDVI	AVE	0724
4	GNDVI	AVE	0626	Plant height	Growth Rate	0530-0605
5	GNDVI	RANGE	0524	WDVI	MAX	0606
LASSO						
1	GNDVI	MAX	0611	GNDVI	Con	0618
2	GNDVI	AVE	0626	Plant height	MAX	0724
3	WDVI	SD	0606	WDVI	SA	0626
4	GNDVI	AVE	0712	NDVI	AVE	0626
5	NDVI	MAX	0606	NDVI	Cor	0712
RFE						
1	GNDVI	AVE	0626	NDVI	AVE	0618
2	NDVI	AVE	0626	WDVI	RANGE	0618
3	NDVI	MAX	0618	WDVI	AVE	0618
4	NDVI	MAX	0606	NDVI	AVE	0626
5	NDVI	SD	0626	WDVI	SA	0618

3.3 機械学習モデルによる予測結果

変数選択法により選択された変数群を用いて、RF、RI、SVMモデルにより予測したバイオマス、果実重、果実数の実測値とシミュレーション値の関係を図5、図6、図7にそれぞれ示す。また、表5に、変数選択により選択された変数群を用いたRF、RI、SVMモデルによる実測値と予測値間のrRMSEの値を示しており、テストデータに対する検証モデルの予測精度を反映するものである。

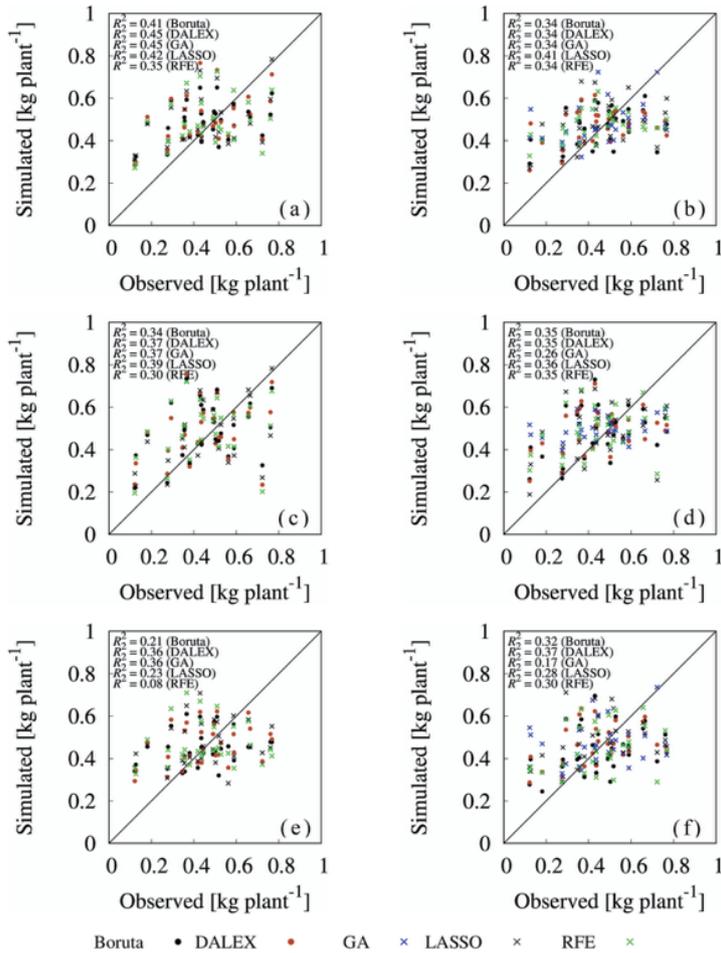


図5 バイオマスの実測値と予測値の相関図 ((a) 1次次元計測値から選択された変数群を用いたRFによる結果, (b) 全計測値から選択された変数群を用いたRFによる結果, (c) 1次次元計測値から選択された変数群を用いたRIによる結果, (d) 全計測値から選択された変数群を用いたRIによる結果, (e) 1次次元計測値から選択された変数群を用いたSVMによる結果, (f) 全計測値から選択された変数群を用いたSVMによる結果))

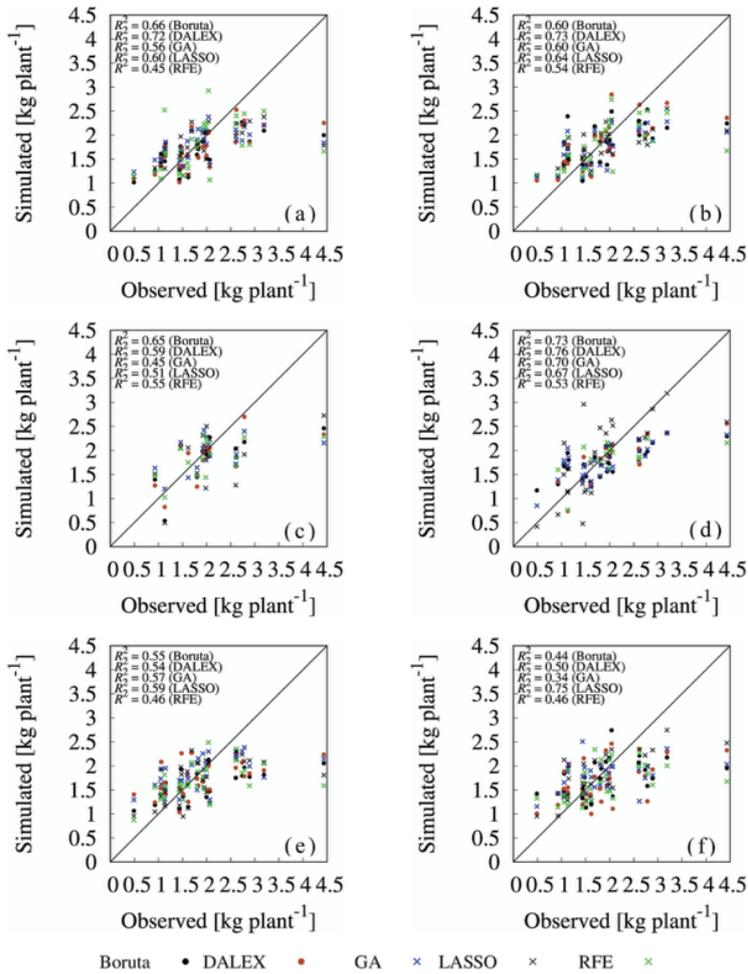


図6 果実重の実測値と予測値の相関図 ((a) 1次元計測値から選択された変数群を用いたRFによる結果, (b) 全計測値から選択された変数群を用いたRFによる結果, (c) 1次元計測値から選択された変数群を用いたRIによる結果, (d) 全計測値から選択された変数群を用いたRIによる結果, (e) 1次元計測値から選択された変数群を用いたSVMによる結果, (f) 全計測値から選択された変数群を用いたSVMによる結果))

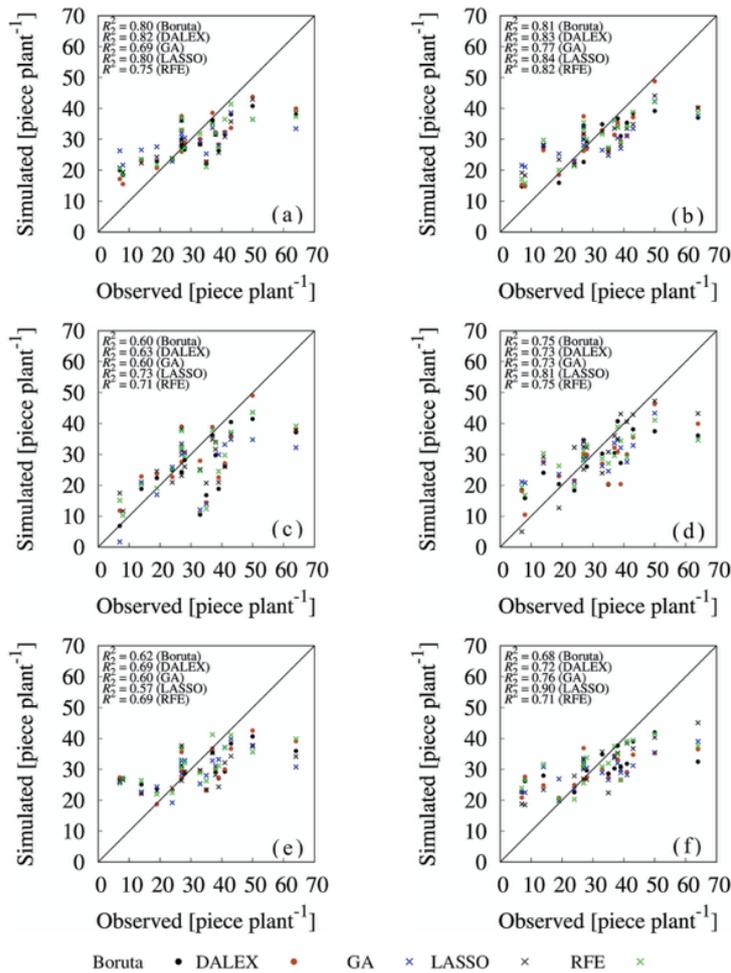


図7 果実数の実測値と予測値の相関図（(a) 1次元計測値から選択された変数群を用いたRFによる結果、(b) 全計測値から選択された変数群を用いたRFによる結果、(c) 1次元計測値から選択された変数群を用いたRIによる結果、(d) 全計測値から選択された変数群を用いたRIによる結果、(e) 1次元計測値から選択された変数群を用いたSVMによる結果、(f) 全計測値から選択された変数群を用いたSVMによる結果）

表5 Random forest (RF), Ridge regression (RI), Support vector machine (SVM) によるバイオマス、果実重、果実数の予測値と実測値間のrRMSE (%)

1次元計測値から選択					
Model	Boruta	DALEX	GA	LASSO	RFE
バイオマス [kg plant ⁻¹]					
RF	17.8	22.2	16.9	22.5	22.9

RI	26.4	26.7	24.9	30.6	26.7
SVM	17.6	18.9	16.7	21.4	21.8
	果実重 [kg plant ⁻¹]				
RF	14.0	13.9	13.2	15.7	24.1
RI	49.6	48.5	48.5	50.1	48.0
SVM	14.3	14.5	14.6	18.7	15.9
	果実数 [piece plant ⁻¹]				
RF	12.6	14.2	10.0	12.4	14.2
RI	30.4	25.5	30.3	18.1	21.2
SVM	13.1	14.2	13.5	13.0	13.6
全計測値から選択					
Model	Boruta	DALEX	GA	LASSO	RFE
	バイオマス [kg plant ⁻¹]				
RF	18.7	16.5	15.0	13.6	13.4
RI	22.7	20.5	11.0	25.8	17.6
SVM	21.4	18.7	11.2	21.6	20.5
	果実重 [kg plant ⁻¹]				
RF	18.0	15.6	12.5	13.8	15.6
RI	11.5	20.6	12.8	22.1	16.7
SVM	14.7	15.4	14.6	15.9	14.5
	果実数 [piece plant ⁻¹]				
RF	14.9	13.5	11.2	11.5	13.5
RI	15.5	17.7	12.7	28.1	13.4

SVM	12.8	10.9	8.8	10.6	14.1
-----	------	------	-----	------	------

バイオマス予測のための5つの変数群を比較すると、1次元計測値を使ってBorutaとDALEXにより選択された変数群を用い、RFモデルによって得られた予測結果は、ほかの変数選択法および予測モデルの組合せと比較して高い R^2 が得られた (Boruta : $R^2=0.41$, DALEX : $R^2=0.45$) (図5a) . rRMSE指標で見た際、全計測値候補からGAによって選択された変数群を用いたRIとSVMを使って得られたバイオマスの予測結果は、ほかの組合せと比較して誤差が相対的に小さかった (表5) . 一方で1次元計測値から変数を選択したケースでは、総じてRFがほかの機械学習モデルと比較して R^2 の値が高く、rRMSEの指標においては、全計測値から抽出した変数群と予測モデルの組合せにおいて、RFを使った結果は総じてrRMSEの値が小さい結果となった。以上より、 R^2 で見た際、バイオマスの予測には、GLCMによるテクスチャ情報の重要性は低く、1次元計測値だけで高い予測精度が得られる結果となった。果実重については、全計測値からBoruta, DALEX, GAを使って選択された変数群を用いたRIモデルによる結果 ($R^2=0.73$ for Boruta; 0.76 for DALEX; 0.70 for GA) , すべての変数からLASSOを使って得られた変数群を用いたSVMモデルによる結果 ($R^2=0.75$) , 1次元計測値からRFEを使って得られた変数群をRIで予測した結果 ($R^2=0.55$) の組合せが優れた予測性能を示した。特にRIに着目すると、RFEを除き、全計測値から変数選択し、予測モデルを適用した結果は、1次元計測値のみから選択した変数群を用いたモデル予測結果と比較し、大幅に予測精度が向上した。たとえば、RIモデルの1次元計測値からGAで選択した変数群を使った結果は R^2 が0.45に対し、全計測値から選択した結果では $R^2=0.70$ となった (図6) . また、1次元計測値、2次元計測値ともに、植生指数から得られる変数の重要度が高いことが分かる。以上の結果より、果実重を予測するには、2次元計測値から得られる特徴量が重要であることが明らかになった (図6, 表5) . 次に、果実数の予測では、全計測値候補から選択された変数群を使って実施した予測モデル結果は、1次元計測値から選択した変数群を用いた結果と比較して、有意に高い適合度を示した (図7) . 特に、Boruta, DALEX, GA, RFEで選択された変数群をRFに適用した結果、およびLASSOで選択した変数群をSVMに適用した結果は、ほかの変数群と予測モデルの組合せと比較して高い予測精度を示した ($R^2=0.81$ for Boruta; $R^2=0.83$ for DALEX; $R^2=0.82$ for RFE; $R^2=0.77$ for GA; $R^2=0.82$ for RFE; $R^2=0.90$ for LASSO) .

本研究では、パラメトリック (ノンパラメトリック) な変数選択法とパラメトリック (ノンパラメトリック) な機械学習モデルの間に、トマトのバイオマス・果実重・果実数の予測精度に対する明確な関連性は見られなかった。一方で、果実重および果実数の予測に関して、全計測値から予測に重要となる変数を選択し、これらを用いたモデルによる予測精度は、1次元計測値のみから変数を抽出し、その変数群をモデルに適用した結果と比較して、予測精度が向上した。さらに、収穫の約1カ月前の植生指数の特徴量がトマトの果実重、果実数の予測にとって重要であることが分かった。

以上、トマトのバイオマス、果実重、果実数の予測に対して影響の大きい表現形質や生育ステージに焦点を絞ることで、効果的に表現型データの収集に貢献することができる可能性があることが明らかとなった。今後は、本研究で用いた変数選択法を多地点かつ複数年において収集した

多くのデータでテストすることにより、ロバスト性の高い予測モデルの構築に必要な特徴量を抽出することを目指していく。

参考文献

- 1) Ramasamy, S. and Ravishankar, M. : Integrated Pest Management Strategies for Tomato Under Protected Structures, In : Sustainable Management of Arthropod Pests of Tomato, pp.313-322 (2018).
- 2) Shi, Y., Thomasson, JA., Murray. SC., Pugh, NA., Rooney, WL., Shafian, S., et al. : Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research, Zhang J, editor, PLoS ONE, Vol.11 (7), e0159781 (2016).
- 3) Barbedo, JGA. : A Review on the Use of Unmanned Aerial Vehicles and Imaging Sensors for Monitoring and Assessing Plant Stresses, Drones, Vol.3 (2) : p.40 (2019).
- 4) Du, M. and Noguchi, N. : Monitoring of Wheat Growth Status and Mapping of Wheat Yield's within-Field Spatial Variations Using Color Images Acquired from UAV-camera System, Remote Sensing, Vol.9 (3) : p.289 (2017).
- 5) Duan, T., Chapman, S. C., Guo, Y. and Zheng, B. : Dynamic Monitoring of NDVI in Wheat Agronomy and Breeding Trials Using an Unmanned Aerial Vehicle, Field Crops Research, Vol.210, pp.71-80 (2017).
- 6) Johansen, K., Morton, M.J.L., Malbêteau, Y., Aragon, B., Al-Mashharawi, S., Ziliani, M., et al. : PREDICTING BIOMASS AND YIELD AT HARVEST OF SALT-STRESSED TOMATO PLANTS USING UAV IMAGERY, Int Arch Photogramm Remote Sens Spatial Inf Sci, XLII-2/W13, pp.407-411 (2019).
- 7) Yang, Q., Shi, L., Han, J., Zha, Y. and Zhu, P. : Deep Convolutional Neural Networks for Rice Grain Yield Estimation at the Ripening Stage Using UAV-based Remotely Sensed Images, Field Crops Research, Vol.235, pp.142-53 (2019).
- 8) Kyratzis, A. C., Skarlatos, D. P., Menexes, G. C., Vamvakousis, V. F. and Katsiotis, A. : Assessment of Vegetation Indices Derived by UAV Imagery for Durum Wheat Phenotyping under a Water Limited and Heat Stressed Mediterranean Environment, Front Plant Sci, Vol.8, p.1114 (2017).
- 9) Guan, S., Fukami, K., Matsunaka, H., Okami, M., Tanaka, R., Nakano, H., et al. : Assessing Correlation of High-Resolution NDVI with Fertilizer Application Level and Yield of Rice and Wheat Crops using Small UAVs, Remote Sensing, Vol.11 (2), p.112 (2019).
- 10) Rouse, J. W., Haas, R. H., Schell, J. A. and Deering, D. W. : Monitoring Vegetation Systems in the Great Plains with ERTS, Third ERTS Symposium, NASA SP-351 I, pp.309-317 (1973).
- 11) Qi, J., Chehbouni, A., Huete, A. R. , Kerr, Y. H. and Sorooshian, S. : A Modified Soil Adjusted Vegetation Index, Remote Sensing of Environment, Vol.48, pp.119-126 (1994).
- 12) Haralick, R. M., Shanmugam, K. and Dinstein, I. : Textural Features for Image Classification, IEEE Trans Syst, Man, Cybern, SMC-3 (6), pp.610-621 (1973).
- 13) Kursu, M. B., Jankowski, A. and Rudnicki, W. R. : Boruta - A System for Feature Selection, Fundamenta Informaticae, 101 (4), pp.271-285 (2010).
- 14) Biecek, P. : DALEX : Explainers for Complex Predictive Models, Journal of Machine Learning Research, Vol.19, pp.1-14 (2018).
- 15) Scrucca, L. : GA : A Package for Genetic Algorithms in R, Journal of Statistical Software, Vol.53 (4), pp.1-37 (2013).

- 16) Tibshirani, R. : Regression Shrinkage and Selection Via the Lasso, Journal of the Royal Statistical Society, Series B (Methodological), Vol.58 (1), pp.267-288 (1996).
- 17) Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. : Gene Selection for Cancer Classification Using Support Vector Machines. Machine Learning, Vol.46 (1/3) pp.389-422 (2002).
- 18) Breiman, L. : Random Forests, Machine Learning, Vol.45 (1), pp.5-32 (2001).
- 19) Hoerl, A. E. and Kennard, R. W. : Ridge Regression, Applications to Nonorthogonal Problems, Technometrics, Vol.12 (1) ,pp.69-82 (1970).
- 20) Cortes, C. : Vapnik V. Support-vector Networks, Machine Learning, Vol.20 (3), pp.273-297 (1995).

脚注

- ☆1 圃場（ほじょう）：農作物を育てる場所
- ☆2 草高（そうこう）：地面からあるがままの状態の作物最上部までの高さ
- ☆3 植生指数（しょくせいしすう）：植物による光の反射特性を使って得られる植物の量や活力を表す指標
- ☆4 畝間（うねま）：畝と畝との間の中心から中心までの距離



辰己賢一（非会員）tatsumi@go.tuat.ac.jp

東京農工大学農学研究院農業環境工学部門准教授，2002年京都大学大学院工学研究科環境専攻修了．専門は農業情報気象学，作物生育シミュレーション．気候変動が農業生産に与える影響の定量的分析に関する研究に従事．

受付日：2021年9月20日

採録日：2021年10月31日
編集担当：飯村 結香子 (NTT)