

研究報告書

「時系列生長データに基づく植物生長の統計的予測技術の開発」

研究タイプ: 通常型

研究期間: 2016年10月～2020年3月

研究者: 松井 秀俊

1. 研究のねらい

農作物の生育においては、生育中の気温や日射量といった環境要因や植物の遺伝子発現が、農作物の収量や品質といった形質に関わっていると考えられている。本研究では、これらの環境要因や遺伝子発現量と、農作物の形質との関係を表すための統計モデルの構築を通して、環境変化から生理状態の変化、そして生長への一連の因果関係を明らかにする。そして、得られた結果を元に、将来の収穫量の予測や、生育方法についての新たな指針を提供することを研究のねらいとする。本研究では特に、施設栽培のトマトに着目し、その生長や遺伝子発現の1点ではなく経時的に測定されたデータ、栽培期間における環境変動を計測し、数理的解析を通して収量を予測するための統計モデルを開発する。

時間の経過とともに繰り返し計測値を得た経時測定データは、データが高次元になることや、観測の欠損や観測時点・時点数の不一致がしばしば生じることにより、一般的な多変量解析や機械学習手法の適用が困難になる場合がある。これに対して、離散時点で観測されたデータを関数化処理し、得られた関数集合をデータとして扱う関数データ解析とよばれる方法を適用することで、これらの問題点を解消できる。関数データ解析に基づいて、環境要因と収量との関係を表現する回帰モデルを構築し実際のデータ分析に適用することで、環境要因が「いつ」「どのように」収量に影響を与えているかを明らかにする。

得られたモデルから、収量だけでなく、収穫期の早晩のデータとの関係についても探索する。出荷時期におけるトマトの成熟が早すぎても遅すぎても、商品価値が高いトマトを過不足なく出荷することは難しい。そこで、生産者にとって最適な出荷量や栽培管理法を決定できる指標を与えられるモデルの構築を目指す。

2. 研究成果

(1) 概要

本研究では、さまざまな農作物に対して、その生育環境に関する情報と、農作物の収量といった生育結果に関する情報との関係を、統計モデルを通じて明らかにした。農作物としては、主として施設栽培における多段栽培トマトを対象にし、年間のほとんどの期間を通して毎日収穫されるトマトの収量と環境要因との関係を明らかにするための統計モデルとその推定法を構築した。その結果、収穫までのいつの環境要因が年間のどの時期に、どのように収量に関わっているかを定量化できた。その他の農作物として、コムギやオオムギの品質に関わると考えられている出穂迄日数(播種から出穂するまでの日数)と気温との関係を分析し、生育期間中の気温がどのように出穂迄日数に関わっているかを明らかにした。また、イネのデータについても同様に、気温の収量への寄与の経時変化を調査した。これらの分析を通じて、さまざまな品種の植物に対して本手法が適用可能なことを示した。

(2) 詳細

研究テーマ A 「多段栽培トマトの収量分析モデルの構築」

施設園芸栽培で取得されたトマトの収量と、施設内外で経時的に計測された気温や日射量といった栽培環境のデータとの関係を統計モデルによって表現することで、両者の関係を説明するための方法について検討した。

本研究では、神戸市にある農園で取得された、多段栽培トマトの生育環境と収量に関するデータを分析対象とした。トマトの果実は、苗の植え替え時期を除く 10 月から翌 7 月まで、毎日収穫される。また果実の収量については、対応する部位における開花から結実の期間における環境要因が関わっていると考えられる。ここでは、この期間を収穫日から遡った 80 日間とした。加えて、この関わり方も、年間の時季によって異なる。これらの関係性を考慮に入れるために、収量と環境要因との関係を表す回帰モデルとして、変化係数関数線形モデルを導入した。このモデルは、回帰係数が、収穫日から遡った日数と、年間の時季に依存して変化する、つまり関数で表されるものである。

図 1 は、導入したモデルで推定された、気温に対する回帰係数関数(曲面)である。図の s 方向は収穫日から遡った日数を表しており、右下から左上の方向に、収穫 1 日前～80 日前に対応する。また t 方向は収穫日を表しており、左下から右上の方向に 1 月 1 日～12 月 31 日を表す。この曲面から、春から夏にかけては収穫日からおよそ 40 日前の気温が高いほど収量も高いといった傾向が見て取れる。

しかし、変化係数関数線形モデルは、固定された時点においては環境要因と収量との関係が線形であり、柔軟性に欠けるという問題点がある。例えば、変化係数関数線形モデルでは、特定の日にける特定の日数前の気温が高いほど収量は高い(または低い)という線形の関係性しか表現することができない。そこで、より柔軟なモデルとして、線形の部分を非線形に拡張した変化係数関数加法モデルを提案した。このモデルを適用することで、例えば、ある時点の気温が一定量まで高ければ収量も高いが、その値を超えると逆に収量は低くなる、といった非線形な傾向まで捉えることができる。図 2 は、変化係数関数加法モデルを適用したことで得られた推定結果である。図 2 上の曲線は、収穫日から遡った 80 日間の気温の関数データに対する第 2 主成分の固有関数を、下は対応する主成分得点と収穫日における非線形関数の推定値を表したもので、横軸は主成分得点、縦軸は収穫日(10 月～翌 7 月)に対応する。固有関数は、収穫 80 日前から収穫前日までの気温の上昇具合を表しており、図 2 下の曲面の橙～白色部分から、春から夏にかけては気温の上昇具合がある程度強

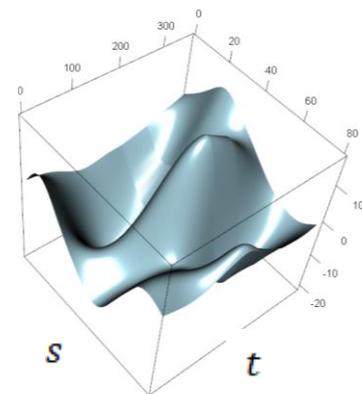


図 1: 変化係数関数線形モデルの回帰係数曲面

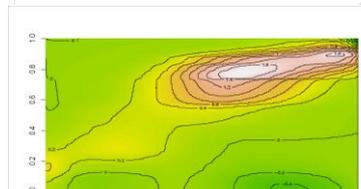
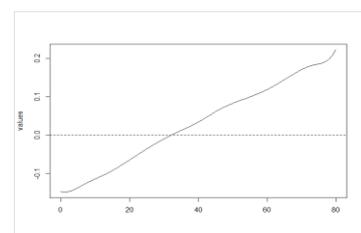


図 2: 変化係数関数加法モデルによる推定結果

いほど収量も多いが、その傾向が強すぎると少なくなるという傾向が捉えられた。このように、関数データに基づく統計モデルを適用することで、環境要因の経時変化による収量への関係性を定量化できた。

また、これらのモデルを用いて、将来の収量の予測も行った。現場の農家では、需要に合わせた次の1週間の収量の予測が重要となる。本研究では、一定期間の収量および環境要因を学習データとみなして、そこから1週先までの7日間の総収量の予測を行った。その結果、既存の関数線形モデル等と比べて予測精度は向上したが、実際の運用に適用するには不十分であった。予測精度の向上のためには、果実の成熟に関わるより直接的な要因をデータとして取得する必要があると考えられる。

研究テーマB「関数線形モデルを用いたムギ・イネの品質分析」

ムギやイネの品質に関わる情報と、環境要因との関係を、統計モデルを用いてモデル化することで明らかにすることを試みた。本研究では、以下に示す2件のデータについて分析を行った。なお、環境要因(気温)のデータについては、気象庁のウェブサイトから取得可能なデータを用いた。

1件目は、岡山県の圃場で生育され取得された、世界中の274系統のオオムギと69品種のコムギに関する20年分の出穂迄日数のデータを分析した。出穂迄日数は1年間で1点のみの観測値である一方で、生育期間中の気温は経時的に観測されるため、ここでは気温を時間の関数データとして扱い、関数線形モデルを適用した。図3は、274品種のオオムギと69品種のコムギそれぞれに対して関数線形モデルを当てはめることで得られた、気温の関数データに対する係数曲線である。横軸は生育期間である11月から翌5月に対応する。この結果から、冬は気温が低いほど、春は気温が高いほど出穂迄日数が長いという傾向が得られた。また、品種によって気温感受性の大小に違いがあることも分かった。

2件目は、日本全国の各地におけるイネの生育期間中の気温と、収量のデータを扱った。これらの関係を、同じく生育期間中の気温の経時変化を関数データとして扱うことで関数線形モデルとして表現し、モデルの係数関数を推定した。図4は気温の係数関数の推定値を示したもので、生育期間中において、気温の収量への影響度合いが変化していることが分かる。特に、幼穂形成期から登熟期にかけて、気温が高いほど収量も高くなるという傾向が得られた。

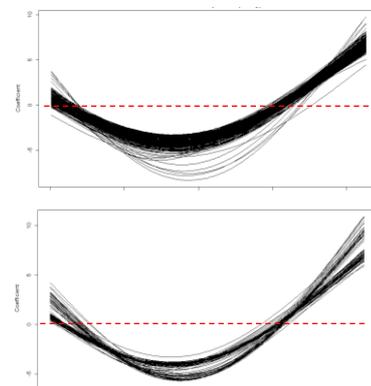


図3:ムギのデータに対する関数線形モデルの係数曲線の推定値. 上は274品種のオオムギ、下は69品種のコムギに対するもの。

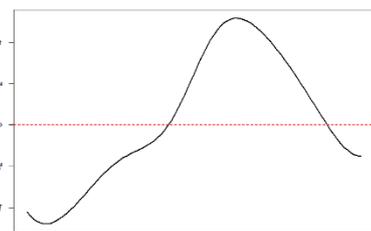


図4:イネのデータに対する関数線形モデルの係数曲線の推定値。

3. 今後の展開

環境要因と農作物の品質との関係を表す関数回帰モデルは、今回分析したデータだけでなく、様々な品種や農場におけるデータにも汎用的に適用可能である。本研究で提案した手法をさまざまなデータの分析へ適用することで、農作物のより良い品質のための栽培の指針となることを目指す。また、本研究は、品質に関わる環境要因を固定して分析を行ったが、実際には分析に用いなかった情報が関わっている可能性もある。そこで、観測されている環境要因を、スパース推定などを利用して自動的に選択するための方法を構築したい。また、経時的に測定されたトマトの遺伝子発現データを活用し、環境要因による生物学的なメカニズムの経時変化を明らかにすることも検討している。

ムギのデータについては、同じ生育環境であっても品種によって出穂迄日数に傾向があることが分かっている。今回得られた気温感受性の違いを足掛かりにして、作物の品質に関わる遺伝子の発見に繋げたい。

4. 自己評価

当初の研究計画にあったトマトのデータ分析については、環境要因と収量との関係に関数データの枠組みでモデル化することができた。特に、収穫日と、収穫日から遡った時点という2種類の時間を考慮に入れた変化係数関数線形モデルを適用した点と、これを拡張した変化係数関数加法モデルを新たに提案した点は評価に値すると考えている。一方で、取得した遺伝子発現データを利用して、環境要因と遺伝子発現、収量という植物生長のメカニズムに沿ったモデリングまでを計画していたが、その包括的なモデルの構築には至らなかった。遺伝子発現データを取得したタイミングが開花時であり、果実の収穫が行われる成熟時と遺伝子発現との関係性が薄かったことが原因である。

環境要因の品質への寄与をモデル化する方法は、トマトやムギに限らず、さまざまな品種や環境下におけるデータに対しても適用可能と考えており、栽培管理における指標として役立つものと考えている。その点で、本研究は幅広い適用可能性を持つものと考えられる。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

1. Matsui, H. Sparse group lasso for multiclass functional logistic regression models. Communications in Statistics- Simulation and Computation 48, 1784-1797.
2. Koda, S., Onda, Y., Matsui, H., Takahagi, K., Yamaguchi-Uehara, Y., Shimizu, M., Inoue, K., Yoshida, T., Sakurai, T., Honda, H., Eguchi, S., Nishii, R., and Mochida, K. Diurnal Transcriptome and Gene Network Represented Through Sparse Modeling in *Brachypodium distachyon*. Frontiers in Plant Science 2017 Nov 28;8 :2055.
3. 松井秀俊. 関数データに基づく統計的モデリング. 統計数理 67, 73-96.
4. Matsui, H. Quadratic regression for functional response models. Econometrics and Statistics, in press.
5. Matsui, H. and Umezui, Y. Variable selection in multivariate linear models for functional data via sparse regularization. Japanese Journal of Statistics and Data Science, in press.

(2)特許出願

研究期間累積件数:0件(公開前の出願件名については件数のみ記載)

(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)

- | |
|---|
| 1. 松井秀俊. 経時測定データに対する統計的モデリング手法とゲノム, 農業データへの応用. 園芸学会小集会「次世代の園芸研究を見据えた先端ゲノム研究」, 近畿大学. |
| 2. Matsui, H. and Mochida, K. Statistical modeling for functional data and its application to crop yield data. 第60回植物生理学会年会, 名古屋大学. |
| 3. 松井秀俊, 持田恵一. 変化係数関数回帰モデルによる作物収穫データの分析. 2019年度統計関連学会連合大会, 滋賀大学. |
| 4. Matsui, H. Varying-coefficient functional additive models. CMStatistics 2019, University of London. |
| 5. 松井秀俊. 応用統計学会奨励論文賞. 「非線形混合効果モデルに基づく関数データクラスタリング」. 2017年. |