

ノンパラメトリック回帰を利用した施設園芸作物の生産性の推定

大河内 理貴[†] 青野 雅樹^{††} 川嶋 和子[‡]

[†]豊橋技術科学大学 大学院工学研究科 情報工学専攻

^{††}豊橋技術科学大学 大学院工学研究科 情報・知能工学系 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

[‡]愛知県農業総合試験場 〒480-1193 愛知県愛知郡長久手町大字岩作字三ヶ峯 1-1

E-mail: [†]ohkochi@kde.cs.tut.ac.jp, ^{††}aono@tut.jp, [‡]kazuko_kawashima@pref.aichi.lg.jp

あらまし 農作物の生産性や品質の向上のために生産物に対する環境データが作物にどのような影響をもたらすかについては、品種改良を主眼とした多くの研究が行われている。しかし、園芸作物の環境データが収穫に与える影響に関する研究はあまり行われておらず、これらをどのように取り扱い、また組合せて制御した場合に、生産性を向上させるかについては十分な知見が得られていない。そこで、温室内で栽培されたトマトについて、センサから得られた環境データをもとに、加法モデルや MARS* といったノンパラメトリック回帰によるデータマイニング手法を利用し、収穫量や収穫日の予測を行ったのでそれを報告する。

キーワード データマイニング, 加法モデル, MARS

Estimating production of greenhouse crops by using Nonparametric Regressions

Masaki OUKOUCHI[†] Masaki AONO^{††} and Kazuko KAWASHIMA[‡]

[†]Dept. of Information and Computer Sciences, Toyohashi University of Technology

^{††}Dept. of Computer Science and Engineering, Toyohashi University of Technology

1-1 Hibarigaoka, Tenpaku-cho, Toyohashi-shi, Aichi, 441-8580 Japan

[‡]Aichi Agricultural Research Center

1-1 Sagamine, Yazako, Nagakute-cho, Aichi-gun, Aichi, 480-1193 Japan

E-mail: [†]ohkochi@kde.cs.tut.ac.jp, ^{††}aono@tut.jp, [‡]kazuko_kawashima@pref.aichi.lg.jp

1. はじめに

近年、農業の大規模化や企業の農業への参入が増加している。このような大規模な生産施設では、市場価格の変動による経営的なリスクを最小限に抑える目的や、生産者が直接大量の農作物を販売することが難しいなどの理由により、量販店などと契約栽培を行っている場合が多い。契約栽培をおこなう場合、量販店は生産者の提出する生産量予測をもとに販売計画を立てることになる為、精度の高い生産量の予測が求められている。

また、新しく農業に参入する場合、どのように育てれば作物の生産量が増加するのか、品質が高くなるのかなどの経験則を持っておらず、手探りで行わなくてはならない。そのため、生育条件により収穫量を増やす方法、冷夏や暖冬などの環境災害に遭遇した場合でも、例年通りに品質および収穫量を一定に保つ方法を知りたいがっている。

これまで、農作物の収穫量の予測および収穫量と環境要因の関係については、主要作物である水稲などの

穀物類の研究が盛んに行われてきた。一方で園芸作物は、野菜を始め、果樹や花など多岐にわたっており、野菜だけでも性質が多岐にわたり、野菜一般の解析は難しく、解析研究が進みにくいという背景がある。

しかし、愛知県では園芸作物の生産高は非常に高く、特にトマトの生産高は群を抜いて高い。そのため、園芸作物、特にトマトの栽培に関する知見が必要とされている。

そこで、著者らは愛知県農業総合試験場と連携し、温室内で栽培されたトマトについて得られた環境データと収穫量データを元に生産性についての推定を行った。

2. 関連研究

トマトに環境データが与える影響については、環境や施肥の異なる状態でトマトを育てた場合、どのような影響が現れるか[1]や、夏季の遮光がトマトの収量と品質にどのような影響を及ぼすか[2]などが報告されている。しかし、どのようなデータ処理を行った場合

* MARS は Salford Systems の登録商標です

に予測精度が最も高くなるかについては、調査されていない。

また、久枝ら[3]は大規模な温室でトマトの測定栽培を行い、環境データとの単回帰分析で出荷量の予測を行った。その結果、収穫が行われる8週間前から1週間までの積算日射量を用いた場合に精度が高くなることが報告されており、この期間が着果から収穫までの期間と同じであると結論付けている。この結果の一般的な適用性の確証はないが、トマトの品種や栽培方法に応じて収穫量とそれに先んじる一定期間の積算日射量の間には何らかの関係があると考えられる。また、日射量以外の環境要因が生育に全く関係がないとは考えにくいいため、それらについて調査する必要がある。

データを式やモデルで近似することを回帰と呼び、予測や分類、制御などに利用することが目的である。回帰はパラメトリック回帰とノンパラメトリック回帰に分類できる。パラメトリック回帰とは回帰式に含まれる回帰係数の数を少なくすることを重視する回帰手法である。一方、ノンパラメトリック回帰とは回帰係数の数が多くなっても気にせず目的に沿った回帰を行おうとするものである。[4]

ノンパラメトリック回帰は農作物の生産性の解析に多く使われている。しかし、水稻の出穂期[5]や樹木の開花時期[6]など、露地栽培作物や収穫が一度である作物の推定には多く使われているが、多段栽培作物や施設園芸作物についてはあまり使われておらず知見が得られていない。そこで、ノンパラメトリック回帰を利用し施設で栽培されたトマトについて生産性の推定を行った。

3. 解析手法

気象データと生産性の関係を単純な線形解析だけで捉えることは難しい。そこで、本稿ではトマトの生産性の推定に加法モデルとMARS法の2つのノンパラメトリック回帰を利用した。

3.1. 加法モデル [7][4]

目的変数 y が説明変数 $\{x_1, \dots, x_r\}$ とともに観測されているとする。通常重回帰は以下の式に回帰するものである。

$$y = \sum_{i=1}^r a_i x_i + b \quad (1)$$

この式は、それぞれの目的変数が説明変数に対し線形な関係であることを仮定している。未知データの予測が簡単にできるという利点はあるが、観測される環境パラメータと目的変数との複雑な非線形構造を表現することはできない。

そこで、重回帰を幾分か柔軟にしたものを用いる。

$$y = \sum_{i=1}^r f_i(x_i) + \varepsilon \quad (2)$$

$\{f_i(x_i)\}$ のそれぞれは滑らかなノンパラメトリック関数で、 ε は平均が0、分散が一定のランダムな誤差である。それぞれの説明変数をノンパラメトリックな回帰関数を使い変換したものを足し合わせ目的変数を求める、という、より一般性の高い回帰式を使う手法を加法モデルと呼ぶ。

ノンパラメトリック関数の導出には様々な方法があるが、本稿では平滑化スプラインを用いた。導出の為にペナルティ付き残差平方和PASS

$$PASS = \left(y - \sum_{i=1}^r f_i(x_i) \right)^2 + \lambda \int \{f''(x)\}^2 dx \quad (3)$$

が最小となる $f(x)$ を求める。ここで λ は平滑化パラメータで、 λ が大きくなると滑らかな曲線を求めることができる。

3.2. MARS[8][9]

MARS (Multivariate Adaptive Regression Splines) 法は交互作用項を含む回帰構造を打ち切りベキ乗スプライン関数によって適応的に推定する[8][9]。基底関数のみならず、変数や節点の位置などもデータに基づいて選定され、回帰構造の解釈に有益な情報を与える。

目的変数 y_i が説明変数 $X_i(x_{i1}, \dots, x_{ir})^T$ とともに観測されているとし、ノンパラメトリック回帰モデル

$$y_i = f(X_i) + \varepsilon_i \quad (i = 1, \dots, n) \quad (7)$$

を満たす関数 f を推定する。関数 f は基底関数 $\varphi_k(x)$ ($k = 1, \dots, K$)の線形集合

$$f(x) = a_0 + \sum_{k=1}^K a_k \varphi_k(x) \quad (8)$$

によって表現されると仮定し、係数 a_0, a_1, \dots, a_k を推定する。各基底関数は打ち切りベキ乗スプライン関数

$$\varphi_k(x) = \prod_{l=1}^{L_k} [s_{kl}(x_{v(k,l)} - c_{kl})]_+^m \quad (9)$$

である。ここに、 $[x]_+$ は x の正の部分、 m は所与のスプライン次数である。基底関数の個数 K 、交互作用の次数 L_k 、符号 $s_{kl}(+1$ または $-1)$ 、基底関数に用いられる変数の番号 $v(k,l)$ 、および節点の位置 c_{kl} はすべてデータの基づく不適合度基準によって選定される。

MARS法のアルゴリズムは以下の二つから構成される。

● 前進ステップワイズ過程

1. 定数項に対する基底を $\varphi_0 \equiv 1$ と置く。
2. $K+1$ 個の基底関数 $\varphi_0, \varphi_1(x), \dots, \varphi_K$ があるとし、新たに2個の基底関数

$$\varphi_{K+1}(x) = \varphi_{K^*}(x)[+(x_{v^*} - c_*)]_+^m,$$

$$\varphi_{K+2}(x) = \varphi_{K^*}(x)[-(x_{v^*} - c_*)]_+^m$$

を追加する。親基底関数 $\varphi_{K^*}(x)$ に用いられていな

い変数 x_{p^*} , 節点の位置 c_s はすべて不適合度基準を最小にするように選定される.

3. K が既定の K_{max} を超えるまで基底関数の追加が続けられる.

● 後退ステップワイズ過程

不適合度基準が最小になるよう, 1 個の基底関数を選びこれを削除する. どの基底関数を削除しても不適合度基準が減少しなくなるまで削除が続けられる.

今回は不適合度基準に一般化クロスバリデーションに少し手を加えた GCV' を用いた.

$$GCV' = \frac{\sum_{i=1}^n (y_i - f(X_i))^2}{n * \left(1 - \frac{\sum_{i=1}^n [H]_{ii} + 0.5d(q-1)}{n}\right)^2}$$

q が基底の個数, d がこの項の効果の大きさを決める負でない定数である. 既定の数が多いときに GCV' のは通常の GCV より大きくなる傾向が強まるので, GCV' を用い場合より既定の個数が少なめになる. よって GCV の場合よりもわかりやすい式を得ることができる.

4. 収穫量の推定

4.1. 使用データ

使用するデータは, 愛知県愛知郡長久手町にある愛知県農業総合試験場アイハウスで 2007 年 8 月から 2010 年 7 月までに栽培された桃太郎ヨーク[10]についてのデータである. アイハウスでは, 7 月末から 8 月上旬頃に播種, 8 月末から 9 月上旬頃に定植し, 10 月半ばから 2 月にかけて収穫を行う促成栽培, 12 月頃に播種, 2 月末頃に定植し, 4 月半ばから 7 月頭にかけて収穫を行う半促成栽培が行われている. データとしては 2007 年度の促成栽培, 2008 年度の半促成栽培は灌水量を変化させ, 栽培を行った. そのため, それぞれ 4 作分のデータがある. そのうち 3 作分を学習データ, 1 作分をテストデータとし, 重回帰, 加法モデル, MARS 法による解析を行った.

4.2. 前処理

1 週間ごとの積算収穫量を目的変数する. 収穫量は 1 週間という短い期間の積算値の為, 何処で区切るかにより解析結果に大きな影響がある. そのため前処理として, 積算収穫量に対し, ラプラス平滑化を行った. 具体的には t 週目の積算収穫量 $x[t]$ に対して,

$$x[t]' = \frac{x[t-1] + 4x[t] + x[t+1]}{6} \quad (10)$$

のように, 前週と次週, および t 週のデータから, (10) 示し, のような重みで平滑化を行った.

また, 説明変数として 2 パターンを用いた.

パターン 1 として, 一定期間の平均気温, 日中

(6:30~18:00)の平均気温, 夜中(18:30~6:00)の平均気温, 最高気温, 最低気温, 日射量, 湿度, 灌水量の積算値を説明変数とし予測を行う.

パターン 2 として, 一定期間の平均気温, 日射量, 湿度, 灌水量の積算値と同じ期間の各積算値の差を説明変数とし予測を行う.

4.3. 評価方法

評価基準として回帰モデルの評価には寄与率 R^2

$$R^2 = 1 - \frac{\sum(\hat{y} - y)^2}{\sum(y - \bar{y})^2} \quad (11)$$

を, 使用する. この値が 1 に近いほどモデルの当てはまりがよいと考えられる.

また, テストデータの予測評価には残差平均

$$\frac{1}{n} \sum_{i=1}^n \frac{|\hat{y} - y|}{y} \quad (12)$$

を利用する. これは残差の割合の平均であり, 0 に近いほうがよい値である.

説明変数の選択, 重要度の基準には一般化クロスバリデーション値を用いた.

4.4. 実験結果

4.4.1. 促成栽培の場合

最も良かった結果のみを表 1 に示す. 積算期間とは説明変数の積算期間を示している. 各収穫週の前日から示した期間までの各環境データの積算量を説明変数としている. 推定の結果, 重回帰よりもノンパラメトリック重回帰を用いた方が, 寄与率, 残差平均共, 良い結果となった.

加法モデルでは, パターン 1 の 8 つの説明変数より湿度, 灌水量, 最高気温, 日射量の 4 つを説明変数として選択した. 各説明変数の重要度は並び順となっており, 湿度が最も重要である. MARS 法ではパターン 1 の 8 つの説明変数より日射量, 最低気温, 湿度, 平均気温を組み合わせ, 8 つの変換関数を使用した. 説明変数の重要度は並び順で, 日射量が重要だという結果になった.

生産者の経験則では促成栽培では日射量が重要とされており, 日射量が高いほど収穫量が増えると考えられている. 加法モデルでは湿度の重回帰モデルに対する貢献度が高く, 経験則と必ずしも一致していない. 法では日射量が有効な説明変数とされており, こちらの方が経験則に近い. また, MARS 法より得られた日射量と収穫量の関係を図 1 に示す. 図 1 より 11 月から翌年 2 月にかけてのトマトの収穫量は収穫の 1~6 週間前の積算日射量が 411.47MJ/m²以上の時, 日平均では 9.8 MJ/m²以上の時, 収穫量が増加することが分かった. この値を超えるのは 11 月下旬前と 2 月以降の収穫に対

表 1. 促成栽培の収穫量の推定
(A : 2007 灌水少, B : 2007 灌水多, C : 2008, D : 2009)

	重回帰	加法モデル	MARS
学習データ	A, C, D	B, C, D	A, C, D
テストデータ	B	A	B
積算期間	差分 4 週間	8 週間	6 週間
寄与率	0.534	0.592	0.805
残差平均	0.266	0.198	0.327

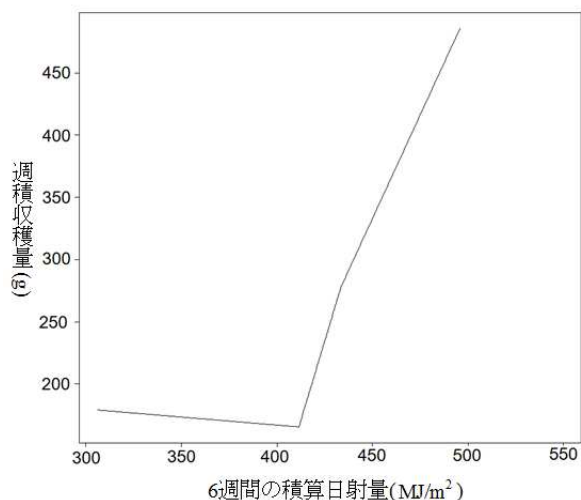


図 1. MARS 法による促成栽培での積算日射量と積算収穫量の関係

する積算日射量に近い。促成栽培の時期に、日射量が収穫量に最も影響を与えるという実験結果は経験則と一致しており、有用なモデルだと考えられる。

4.4.2. 半促成栽培の場合

最も良かった結果のみを表 2 に示す。半促成栽培での予測は促成栽培に比べると全体的に良い結果を得ることができた。これは促成栽培と比べ発育期間が短く、成長が速いため同じような成長になりやすい為だと思われる。また、半促成栽培においても、重回帰よりもノンパラメトリック回帰を用いた方が良い結果を得ることができた。

加法モデルでは説明変数としてパターン 1 の 8 つの説明変数より、夜間の気温、日平均気温、最大気温、最低気温、灌水量の 6 つを選択した。各説明変数の重要度は並び順となっており、夜間の気温が最も重要である。また MARS 法ではパターン 2 の 8 つの説明変数より差分気温、積算気温、積算灌水量、差分湿度を組み合わせて 5 つの変換関数を使用している。各説明変数の重要度は並び順となっており、差分気温が最も重要である。

生産者の経験則では半促成栽培では気温と灌水量

が重要とされており、気温は高すぎると収穫量が減る、灌水量は多いと良く成長し、収穫量が増えると考えられている。加法モデルでは気温の回帰モデルに対する貢献度が高かったが、収穫量との関係を見てみると気温が高く、特に夜間の気温が高いほど収穫量が増加する傾向にあり、経験則と一致しなかった。一方、MARS 法では気温、灌水量が有効な説明変数とされており、経験則と一致することがわかった。また、MARS 法より得られた積算気温と収穫量の関係を図 2 に、積算灌水量と収穫量の関係を図 3 に示す。図 2 より 4 月中旬から 7 月上旬にかけてのトマトの収穫量は収穫の 1~3 週間前の積算気温が 476.26℃ の時、日平均では 22.6℃ の時、収穫量が最も増加することが分かった。この値は 5 月中旬から 6 月上旬の収穫に対する積算気温に近い。また、図 3 より、トマトの収穫量は収穫の 1~3 週間前の積算灌水量が 24430ml 以上の時、日平均では 1163ml 以上の時、収穫量が増加することが分かった。半促成栽培の時期に、気温と灌水量が収穫量に最も影響を与えるという実験結果は経験則と一致しており、有用なモデルだと考えられる。

表 2. 半促成栽培の収穫量の推定
(A : 2008 灌水少, B : 2008 灌水多, C : 2009, D : 2010)

	重回帰	加法モデル	MARS
学習データ	B, C, D	B, C, D	A, C, D
テストデータ	A	A	B
積算期間	8 週間	7 週間	差分 3 週間
寄与率	0.743	0.785	0.822
残差平均	0.329	0.24	0.176

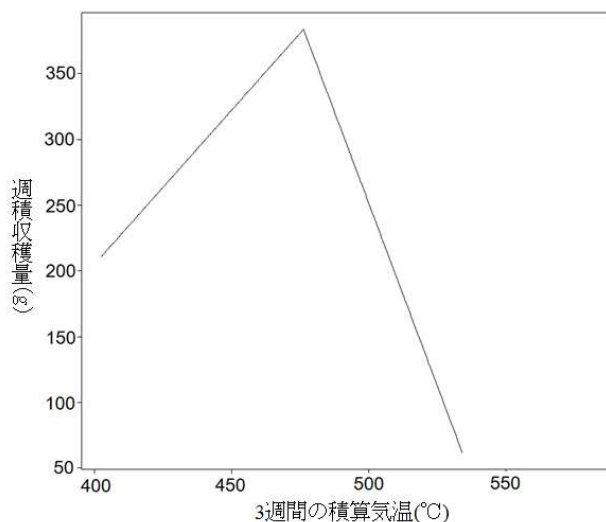


図 2. MARS 法による半促成栽培での積算気温と積算収穫量の関係

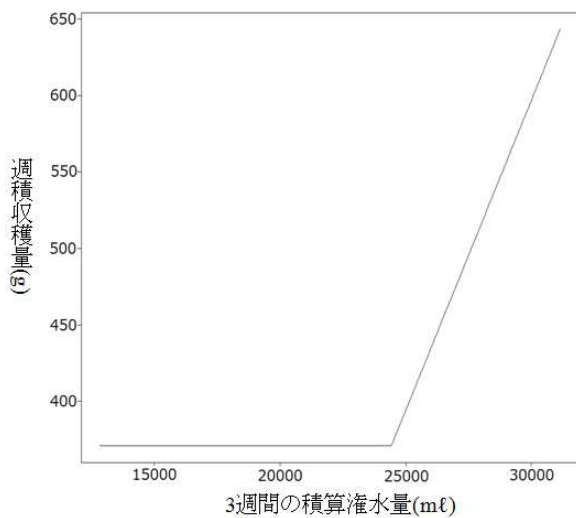


図 3. MARS 法による半促成栽培での積算灌水量と積算収穫量の関係

5. まとめと今後の課題

本稿はトマトの生産性と環境データの間を、ノンパラメトリック回帰を用いて解析し、トマトの収穫量の予測を行ない、重回帰との比較を行い、結果を示した。

結果として、加法モデルと MARS 法は園芸作物の生産性の解析に適しており、ある程度の予測結果を得ることができた。特に、MARS 法は選択された説明変数と収穫量の関係が生産者の経験則に近く、生産量の予測に適していると思われる。

また、本研究の解析に用いた説明変数の積算期間は 6~8 週が最も有効であるという知見が得られた。この期間はトマトの開花から収穫までの必要日数と近く、この期間の環境要因がトマトの生産性に大きな影響を与えると考えられる。さらに、この結果は久枝ら[3]の結論と合致する。

今後の課題としては、今回導出した回帰式を元にどうすれば収穫量が増加、また安定するのかを導き出し、シミュレーション手法を実現させる予定である。

また、この回帰式が正しいのか、有用性があるのか生産の現場で検証していく予定である。

参考文献

- [1] 中野明正, 上原洋一, “トマト生産における施肥, 栽培法が収量, 品質, 環境に与える影響—昔のトマトはおいしかったか?—”農業および園芸 81 巻 2 号, pp291-301, 2006.
- [2] 和田光生, 池田英男, 松下健司, 神原晃, 平井宏昭, 阿部一博, “夏季の遮光が一段栽培したトマト果実の収量と品質に及ぼす影響”, 園芸学会雑誌 75(1), 51-58, 2006.
- [3] 久枝和昇, 仁科弘重, “大規模トマト生産温室における生産性向上に関する研究—積算日射量に

基づいたトマトの出荷量予測—”, 植物環境工学 vol.19, No.1, pp.11-18, 2007.

- [4] 竹澤邦夫, “みんなのためのノンパラメトリック回帰 第3版”, 吉岡書店, 2007.
- [5] 長谷川正俊, 高橋真理, 三沢民男, “ノンパラメトリック DVR 法による山形県作付け主要水稻品種の出穂期予測”, 日本作物学会東北支部会報 (46), 27-28, 2003.
- [6] 大谷義夫, “気象生態反応に基づくニホンナシの開花予測”, 栃木県農業試験場研究報告 (58), 7-16, 2007-02.
- [7] Trevor Hastie and Robert Tibshirani, “Generalized Additive Models”, Statistical Science, Vol.1, No.3, 297-318, 1896
- [8] FRIEDMAN J. H., “Multivariate adaptive regression splines”, The Annals of Statistics 19, 1-141, 1991.
- [9] 坂本亘, “経験 Bayes 法による多変量適応的回帰スプラインの基底および節点の選定”, 日本計算機統計学会大会論文集 (20), 165-168, 2006.
- [10] 住田敦, 加屋隆士, 畠中誠, “完熟トマト‘桃太郎’系品種の育種と普及”, 園学研. 7 (1) : 1-4. 2008.