

施設トマト栽培における積算日射量と温度分布を用いた収量予測モデルの開発と気象変化適用

逢坂未羽† 南野謙一†

岩手県立大学ソフトウェア情報学部†

1. はじめに

近年、センサーで計測して集めたビッグデータを解析し、収量を予測することで、効率的に栽培管理するスマート農業が進められている。ハウス野菜栽培では見込まれる収量に基づいた販売計画作成などの観点より、収量予測への需要が高まっており、センサーを設置し収集したデータから機械学習により収量予測モデルの開発が行われている。しかしながら、気温や日射量など多くの外的要因により定まる農作物の収量を予測するには、センサーデータの選別、データの前処理などを行い、対象圃場に適合させなければ精度の高いモデル開発はできない。そこで、本研究では、(株)いわて若江農園との共同研究による、実用的な施設トマト栽培の収量予測モデルの開発と気象変化適用について述べる。対象となる圃場は、同農園の中棟であり、敷地面積は 1935 m²、栽培方法はロックウール養液栽培、作型は長期多段どり栽培である。

2. 収量予測モデルの開発手法

2.1 重回帰モデル

機械学習では、施設の環境データを用い、重回帰モデルにより収量予測を行う。重回帰は教師あり学習の手法の 1 つであり、1 つの目的変数を複数の説明変数で予測する。本研究では、説明変数に環境データの日射量と気温、目的変数に収量データをセットし、重回帰モデルを開発する。

収量予測モデルには、訓練データにより学習を行い、訓練データとは異なるテストデータにより評価を行う。訓練データには、2020 年 6 月から 2021 年 5 月までのデータを、テストデータには、2021 年 6 月からの環境データを用いて 2021 年 8 月からの収量を予測する。精度は決定係数、予測値と実測値の平均二乗誤差から評価する。これらの結果を基にして、モデル開発に用いるデータの見直しを行い、気象変化に適用させるモデルをいくつか開発する。

2.2 説明変数

説明変数は施設の日射量、日平均気温を加工したものをを用いるが、日射量は一定期間積算した値とする。積算期間は、1 週刻みの積算日射量と収量の相関を調査した結果、最も相関が高くなった収



図 1 2020 年の訓練データによる重回帰モデル

穫日から前 4 週間積算した値とする。日平均気温は一定期間の温度分布を表した Histogram 関数（以下、H 関数）を用いる[1]。H 関数は、定植から収穫開始日までの日平均気温を対象とし、1 度刻みの度数分布表を作成して気温ごとの相対度数を求め、その値を用いる。H 関数を導入した先行研究を参考にしたところ定植日から収穫日までの期間は 7 週から 9 週となることが多かった。本研究では始めに間をとって 8 週を採用した。しかし、収穫までの 8 週間の気温データの影響を受けながら予測するため訓練データにはない気温変化がある場合には誤差が生じやすい。そのため、予測値をみながら H 関数の期間や度数分布の対象となる気温の範囲を変化させて 2021 年 8 月からの予測を行う。

2.3 目的変数

目的変数には収量データを用いる。しかしながら、作業が休日の場合、収量は 0g となり、休日に採れるはずの収量は別の日の収量にカウントされるため、トマトが生育した環境本来の収量とは異なってしまう。そこで 1 週間積算した値を取ることにした。過去の収量よりも未来の収量を予測できる方が収穫時に必要とされる人員の調整や見込まれる収量に基づいた販売計画作成などに役立つため、収穫日の翌日から 7 日間積算した値を採用する。

3. 評価実験

3.1 モデル評価

2020 年 6 月から 2021 年 5 月の訓練データで学習させた重回帰モデルと収量データ（訓練データ）の適合を示す（図 1）。このとき、説明変数には H 関数の期間 8 週で、1 度刻みの温度分布の相対度数を

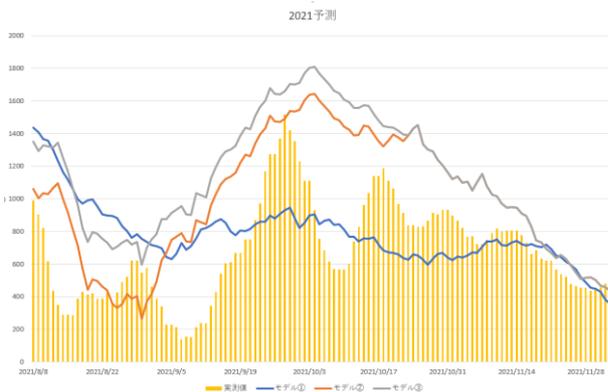


図 2 温度分布を改良した3つの重回帰モデル

用いた。その結果、決定係数は 0.85 となり、1 年を通じて大きなずれがない。

しかし、この重回帰モデルを 2021 年 6 月からのテストデータを用いて評価を行ったところ、予測値が 8~9 月の期間マイナスの値を取った。2021 年 7~9 月の気温が 2020 年の同月よりも高く推移し、訓練データには見られない高温傾向があったため予測精度が低下した。なおかつ 8~10 月中旬まで同農園でのホルモン処理の不手際による不着果が原因で収量が減少した。重回帰モデルの改良として、大きくマイナス値を取っている係数が 24℃, 28℃, 29℃の温度分布（相対度数）であり、それらの収量低下の影響を抑える必要がある。

3.2 気象変化適用

マイナス値を取っている 8~9 月の予測値をプラスにするために訓練データを用いて新たに 3 つのモデルを開発した。モデル①は、24℃, 28℃, 29℃の影響を小さくするために H 関数を、20℃~29℃を 5℃刻み、その他の温度を 1℃刻みにした温度分布の相対度数を説明変数として開発した。モデル②, ③は 2021 年のテストデータで予測してマイナス値をとったモデルから、モデル②は 28, 29℃を、モデル③は 24, 28, 29℃を除いた温度に対して 1 度刻みの温度分布の相対度数を説明変数として開発した。モデル②, ③は係数の値が大きくマイナス値を取る 24℃, 28℃, 29℃を除くことで、収量の値をプラスにできる。3 つの重回帰モデルで 2021 年 6 月からのテストデータを用いて予測を行った(図 2)。ホルモン処理の不手際により 10 月中旬まで収量の減少が続いていることから 3 つのモデルとも予測値と実測値のずれが大きいが予測値が全期間プラス値を取っている。また、9 月下旬から 11 月中旬まで 3 つの予測値の範囲内に実測値を捉えることができた。

しかしながら、予測値の範囲が大きく、実用的に収量の目安がつけにくい。そこで予測値の範囲の幅を狭めるために H 関数の期間を 5 週にして、1 度刻みの温度分布で作成した相対度数を説明変数にした重回帰モデルを開発し、テストデータを用いて予測を行った(図 3)。訓練データでは、決定係数が 0.88 となった。グラフからも 10 月中旬から 11 月上

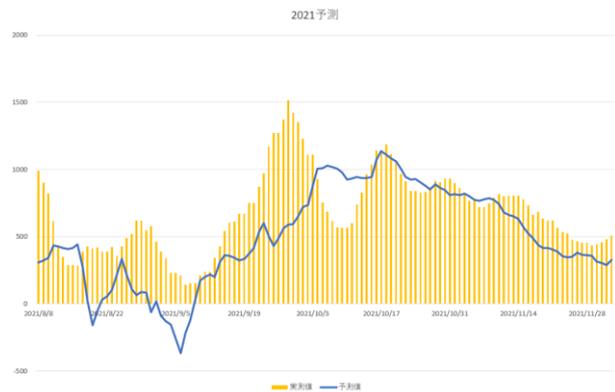


図 3 H 関数の計測期間を改良した重回帰モデル

旬の期間は、ある程度実測値を追えていることが分かる。これは H 関数の期間を 5 週にしたことで少ないデータで重回帰モデルを開発したことから 7~9 月の高温の影響を大きく受けずに予測ができたと推測される。

3.3 平均二乗誤差

実測値と予測値の平均二乗誤差による評価を行った。10 月の平均二乗誤差はモデル①は 1537.3, モデル②は 3206.1, モデル③は 3794.0, H 関数の期間が 5 週のモデルは、1239.0, H 関数の期間が 8 週で 1 度刻みの温度分布で作成した相対度数を説明変数としたモデルは、3175.4 となった。10 月の期間は、7~9 月の高温の影響を大きく受けずに予測できているため、H 関数の期間が 5 週のモデルが平均二乗誤差から見ても実測値とのずれが小さい。11, 12 月の平均二乗誤差は、モデル①は、893.0 モデル②は、1156.1 モデル③は、1156.1, H 関数の期間が 5 週のモデルは、1062.4, H 関数の期間が 8 週で 1 度刻みの温度分布で作成した相対度数を説明変数としたモデルは、1156.1 となった。

4 まとめ

本研究ではハウストマト栽培の収量を対象に、2020 年 6 月から 2021 年 5 月の訓練データを用いて重回帰モデルを開発し、2021 年 8 月からの収量を予測した。2021 年 7~9 月は 2020 年の同月と比べて高温傾向があったため、特定の温度を除いた度数分布（H 関数）と H 関数の計算期間を縮小させた予測モデルを開発することで、7~9 月の高温の影響を小さくして予測ができた。

今後の課題として、1 年分のデータで学習させてモデルを開発したが、年ごとに気象が異なるため、データを蓄積していき数年分のデータで学習させたモデルの開発が必要となる。さらに精度向上のため、収穫前の果実の着果数、大きさ等を説明変数に加えてモデル開発する。

参考文献

- [1] 遠藤隆也・割澤伸一・山田一郎, “トマト低密度植栽培における収穫日・収量の予測手法（計算モデル）の検討”, 日本生物環境工学会大会講演要旨, 2017.