

# 気象観測データの異常値判別手法の比較検討

1220180 吉岡昌輝

高知工科大学 システム工学群 建築都市デザイン専攻

E-mail: 220180j@ugs.kochi-tech.ac.jp

気象庁は気象データを観測し、品質管理を行っている。社会気象工学研究室が佐岡地区にて観測している観測データについて品質管理を行うために、観測データの異常値判別をしていく必要がある。そのため本研究では、既往研究をもとに2つの手法で異常値判別を行った。そして、本研究で得られた結果と既往研究の結果を比較し、より良い異常値判別方法を検討した。

**Key Words** : 気温, 相対湿度, 平滑化, 相関関係

## 1. 序論

### 1-1. 背景

気象庁は気温や相対湿度、降水量や風速といった気象要素を観測しているが、度々誤った値の観測データが記録されてしまう事例が発生している。そのため、観測データの品質管理は重要とされており、気象庁では品質管理が行われている<sup>1)</sup>。現在、社会気象工学研究室では香美市土佐山田町佐岡地区にて気象データを観測している。佐岡の観測データの信頼性を担保するためにも品質管理を行う必要がある。既往研究として昨年、有本が2つの方法を用いて佐岡の気象データの異常値判別を行い、品質管理情報の作成をした<sup>2)</sup>。

まず1つ目は、平均と標準偏差での異常値判別である。2つ目は、佐岡の観測データのみを用いた平滑化による異常値判別である。しかし、この2つの異常値判別が正しいとは限らない。佐岡と気象庁の差をホテリングの $T^2$ 法を用い、佐岡の観測データと佐岡の平滑化した観測データの差の平均標準偏差を使い、異常値判別を行うことで同じような傾向がみられるかを検討する必要がある。また、差の平均と標準偏差を求める際に、 $\pm 3\sigma$ の範囲で行ったが、果た

してその範囲で正確な異常値検出を行えるかどうかにも検討する必要がある。

そこで本研究では、既往研究をもとに佐岡における観測データの信頼性を向上させるためにも、既往研究とは異なる異常値判別の方法も検討することで佐岡地区における異常値判別の方法の確立を目指す。また、気温と相対湿度の関係や、既往研究で用いた異常値判別で計算した異常値の頻度との関係も検討する。

### 1-2. 目的

既往研究<sup>2)</sup>とは異なる統計的手法を用いて異常値判別を行う。既往研究で行われた異常値判別の結果と本研究で行う異常値判別の結果を比較し、実際に佐岡での異常値判別に用いることができる手法を検討する。

## 2. 手法

### 2-1. 使用するデータ・期間

対象地域は佐岡地区で、使用するデータは、佐岡の気温と相対湿度、気象庁が観測している気温と相対湿度を使用する<sup>3)</sup>。気象庁の気温データに関しては

後免地域気象観測所にて観測されたデータを用いる。また、相対湿度に関しては高知地方気象台にて観測されたデータを用いる。佐岡と気象庁ともにそれぞれ10分毎に観測されたデータを使用する。

データの期間は2018年の2月1日から2019年の10月4日である。図-1に各観測点の位置関係を示しており、図-2に佐岡の2点の位置関係を示す。本研究では佐岡WS1を仮社殿とし、佐岡WS2を東屋と呼ぶ。



図-1 観測点の位置関係<sup>2), 4)</sup>



図-2 佐岡地区の「佐岡WS1：仮社殿」と「佐岡WS2：東屋」の位置関係<sup>2), 5)</sup>

## 2-2. 気象庁の観測データを用いた異常値判別

佐岡の観測データを気象庁の観測データと比較して異常値判別を行う。

有本は、佐岡と気象庁の観測データの差を計算し、それぞれの差の平均と標準偏差 $\sigma$ を求めて平均から $\pm 3\sigma$ より大きい値と小さい値を異常値として判別を行った。

本研究では、佐岡と後免の気温差と佐岡と高知の相対湿度差をそれぞれ計算し、ホテリングの $T^2$ 法の計算により異常値判別<sup>6)</sup>を行う。計算した差のF統計量は自由度1のカイ二乗分布に従うと仮定し、デ

ータ異常度を計算する。本研究では、そのカイ二乗分布の99%点を異常値の閾値とする。

## 2-3. 佐岡の観測データの平滑化による異常値判別

佐岡の観測データのみを用いて異常値判別を行う。

有本は、佐岡の観測データをSavitzky-Golay(SG)法により1日毎のデータを平滑化し、ホテリングの $T^2$ 法の計算により異常値判別を行った。

本研究では、まず同じように佐岡の観測データに対して、Savitzky-Golay(SG)法により1日毎のデータを平滑化する。観測データと平滑化されたデータの差を気温、相対湿度それぞれ計算する。計算された佐岡の差の平均と、差の標準偏差 $\sigma$ を用いて、平均から $\pm 3\sigma$ より大きい値と小さい値を異常値として判別する。対象期間のデータを月毎時間毎にわけて差の平均、標準偏差を計算し、異常値判別を行う。したがって、1地点あたり12か月×24時間分のデータができることになる。

## 3. 結果・考察

### 3-1. 気象庁の観測データを用いた異常値判別結果

佐岡と気象庁の観測データの差をとり、異常値判別を行った。図-3、図-4は佐岡と気象庁の観測データより異常値判別したものを気温、相対湿度にそれぞれ分け、異常値判別の時間帯別頻度を示したものである。気温は、20~9時頃にかけて東屋、仮社殿ともに異常値が多くみられ、9~20時頃の太陽が昇っている時間帯に異常値が少ない傾向にあることがみてとれる。相対湿度に関しては、東屋と仮社殿で異常値の頻度が大きくなるどころが、東屋では0~8時の間、仮社殿では9~20時の間と異なっていたものの、24時間通しての異常値の頻度のばらつきは、あまり目立たなかった。

有本の手法で行った異常値判別による結果は、図-5、図-6のようになった。このグラフは、佐岡と後免の気温差と佐岡と高知の相対湿度差をそれぞれ計算し、差の平均と差の標準偏差 $\alpha$ を求めて平均から $\pm 3$

σより大きい値と小さい値を異常値として判別したものである。図-3、図-4と比較すると異常値の頻度は異なるものの、異常値の時間帯別頻度の傾向が類似していることが分かった。

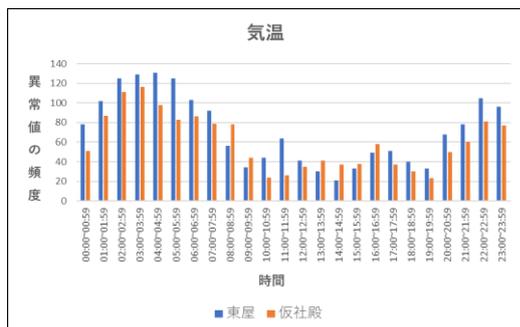


図-3 2-2の手法による気温の異常値の時間帯別頻度

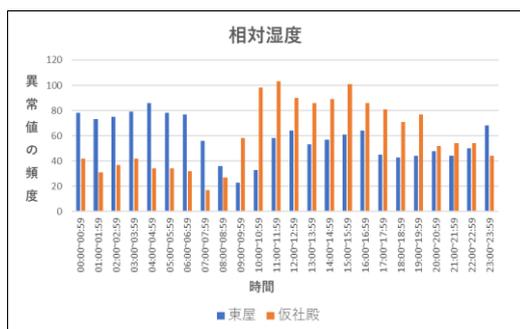


図-4 2-2の手法による相対湿度の異常値の時間帯別頻度

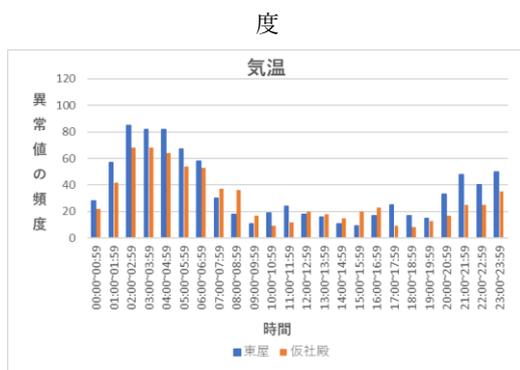


図-5 有本の手法による気温の異常値の時間帯別頻度

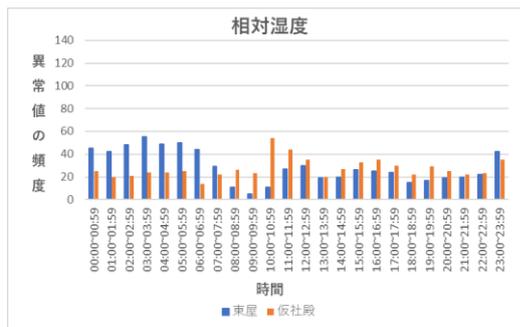


図-6 有本の手法による相対湿度の異常値の時間帯別頻度

度

### 3-2. 佐岡の観測データの異常値判別の結果

東屋と仮社殿それぞれの1日毎のデータを用いて平滑化による異常値判別を行った。図-7、図-8には平滑化による気温、相対湿度の異常値判別の時間帯別頻度を示した。気温と相対湿度ともに、24時間通して異常値の頻度に大きな差がなかった。また、東屋と仮社殿において、気温も相対湿度もそれぞれ異常値判別の時間帯別頻度がほとんど同じ値を示し、傾向が類似していた。

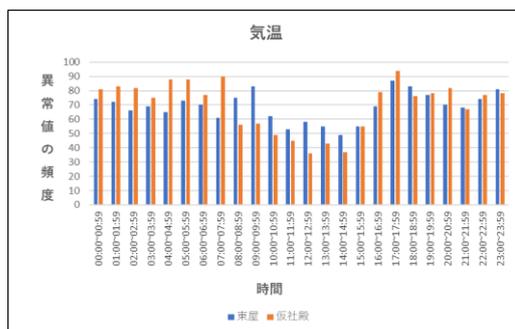


図-7 平滑化による気温の異常値の時間帯別頻度

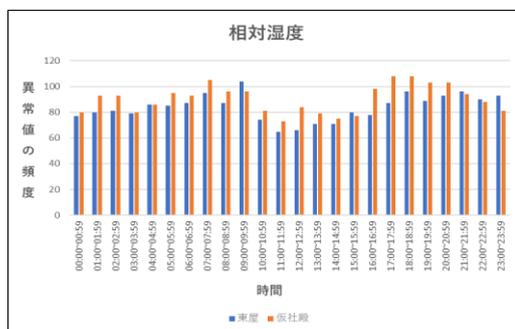


図-8 平滑化による相対湿度の異常値の時間帯別頻度

有本の手法で行った異常値判別による結果は、図-9のようになった。佐岡の東屋と仮社殿の気温の観測データそれぞれに対して平滑化し、ホテリングの $T^2$ 法の計算により異常値判別したものである。

このグラフは、図-7と比較すると図-7では、24時間通して異常値の頻度のばらつきがほとんど見られなかったのに対して、図-9では、8~17時頃の昼間の時間を中心に異常値が多く検出された。

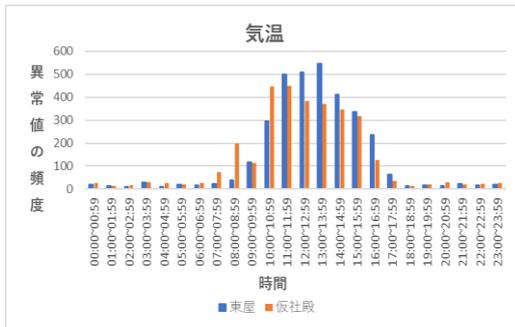


図-9 有本の平滑化による異常値の時間帯頻度

### 3-3. 気象データの相関関係の比較

図-10、図-11は、気象庁の観測データを用いた異常値判別の結果をもとに作成した各時間帯の気温と相対湿度の異常値検出数の散布図である。図-7のように東屋では、気温での異常値の数が多い時間帯では相対湿度での異常値の数も多くなる傾向にある正の相関に見えるグラフとなった。逆に仮社殿では負の相関がみられた。

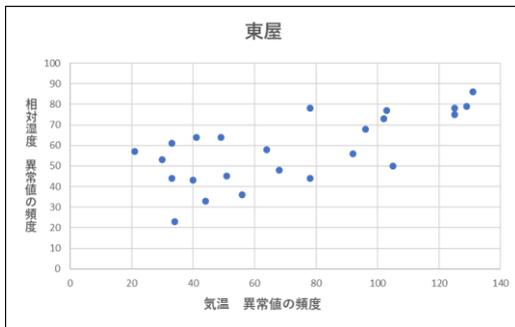


図-10 東屋における気温と相対湿度の相関関係

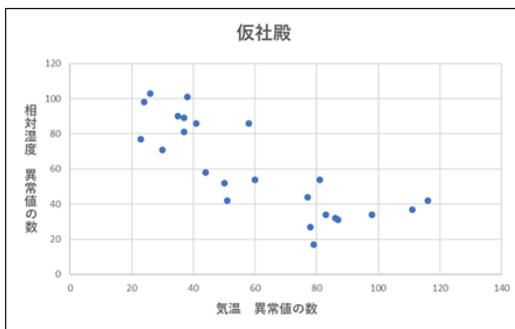


図-11 仮社殿における気温と相対湿度の相関関係

また、平滑化による異常値判別の結果をもとに作成した各時間帯の気温と相対湿度の異常値検出数の散布図を作った結果は、気温での異常値の数が多い時間帯では相対湿度での異常値の数も多くなる傾向

にある正の相関に見えるグラフとなった。

## 4. まとめ

既往研究をもとに2つの手法を用いて東屋と仮社殿の観測データの異常値判別を行った。気象庁の観測データを用いた異常値判別では、既往研究で得られた結果を比較してみても同じような傾向がみられた。また、佐岡の観測データのみを用いた平滑化による異常値判別では、本研究では24時間通して検出数に変化があまりなく、既往研究では昼間に異常値の検出数が多いという大きな違いがみられた。

気温が上昇すれば相対湿度が減少するという関係があるため、2つの手法において気温の異常値検出数が多い時間帯では相対湿度の異常値検出数も多くなる傾向が確認できたが、逆の傾向を示す場合もあった。このような場合、適切に異常値判別が行えていない可能性があるため、他の異常値判別手法の検討が必要になる。

## 5. 参考文献

- 1) 観測部計画課情報管理室：測候時報.82，観測データの品質管理強化と高度化の現状 - 気温に関する自動品質管理機能、対話的品質管理ツール、観測所運用記録情報の業務への利用 - ，気象庁，2015
- 2) 有本陸矢，「気象観測データの品質情報作成手法の検討」2020年度高知工科大学システム工学群卒業研究，2021.
- 3) 気象庁 <https://www.jma.go.jp/jma/index.html>
- 4) 国土地理院：地理院タイル <https://maps.gsi.go.jp/development/ichiran.html>
- 5) Google Map <https://www.google.co.jp/maps/?hl=ja>
- 6) 井手剛：入門機械学習による異常検知-Rによる実践ガイド，コロナ社，2015