

気象観測データの品質情報作成手法の検討

1210008 有本陸矢

高知工科大学 システム工学群 建築都市デザイン専攻

Email: 210008@ugs.kochi-tech.ac.jp

気象庁は気象データを観測し、独自の手法で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いたいが、過去のデータに対して品質管理を行っても本当に異常値かどうかを確かめる手段がない。そのため本研究では、統計的手法を用いて異常値の可能性のあるデータを判別し、観測データとともに品質情報として提供することで佐岡の観測データの信用性担保を目指す。基準データと観測データとの差の平均と標準偏差を用いた方法と観測データの平滑化による方法により異常値判別を行い、佐岡の観測データに対して品質情報を作成することができた。

Key Words : 気温, 相対湿度, 異常値判別, 平滑化

1. 序論

1-1. 背景

気象庁は気温や相対湿度、降水量といった気象データを観測している。そして私たちは報道機関やインターネットを通じてその観測データを取得している。しかし、気象庁の観測では度々誤った観測値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は必ずしも正確とは限らず、観測データの品質管理は重要となる。

気象庁では¹⁾温度変化の仕方により方法を分けて品質管理を行っている。急速な温度変化については、10分程度の温度変化に対応させた指数 T_a を確立し、気温の上昇時の変化と下降時の変化が共に大きい場合を捉えて異常値かどうかを評価している。また、指数 T_a のみで排除できない判別結果の空振りに対応するために指数 T_b を確立し、前50分間の気温変化に対する観測時刻の気温変化の割合を指数として評価している。この二つの指数を用いて統計的手法に

より閾値を設定したうえで異常値判別を行っている。緩慢な温度変化については、対象観測地とその周辺の10点の観測地を用いて重回帰分析を行い、対象観測地の気温を推定し、実際の観測値との差を求めることで正常・異常を判別している。

社会気象工学研究室では高知県香美市土佐山田町佐岡地区にて気温と相対湿度などの気象情報を観測している。この観測データの信用性を担保するため、気象庁と同様の方法を適用し品質管理を行いたいが、過去のデータに対して異常値判別を行ってもそれが本当に異常値であるかを確かめる手段がない。

1-2. 目的

プログラミング言語の Python にて統計的手法を用いて異常値判別を行う。そして、異常値の可能性のあるデータを判別し、観測データとともに品質情報として提供することで、佐岡における気象観測データの信用性担保を目指す。

2. 手法

2-1. 使用するデータ・期間

佐岡地区にて観測している気温と相対湿度の品質情報を作成する際、佐岡地区の観測データの他に、気象庁にて観測している気温と相対湿度を使用する²⁾。観測点の位置関係を図-1に示す。佐岡地区には二か所の観測地点があり、図-2に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。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。佐岡の観測データと気象庁の観測データは共に10分毎に観測されたデータを使用する。佐岡と気象庁共にデータの期間は2018年の2月1日から2019年の10月4日である。



図-1 観測点の位置関係（地理院地図³⁾から作成）



図-2 佐岡地区の「佐岡 WS1：仮社殿」と「佐岡 WS2：東屋」の位置関係（Google Map⁴⁾から作成）

2-2. 平均と標準偏差での異常値判別

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。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する。そして、それぞれの平均と標準偏差 σ 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する。

(1) 月毎での異常値判別

対象期間のデータを月毎にわけて平均と標準偏差を計算する。そのため2月などは2018年と2019年を混ぜて計算するが、12月は2018年のみとなる。

(2) 月毎と時刻毎での異常値判別

(1)で月毎に分けたものをさらに1時間毎に分けてそれぞれの平均と標準偏差を計算し、異常値判別を行う。

2-3. 平滑化による異常値判別

東屋と仮社殿の観測データそれぞれに対して、Savitzky-Golay(SG)法により1日毎のデータを平滑にし、ホテリングの T^2 理論の計算により異常値判別を行う。SG法により平滑化したデータと観測値の差のF統計量は自由度1の χ^2 分布に従うと仮定し、データ異常度を計算する⁵⁾。本研究では、その χ^2 分布の99%点を異常値の閾値とする。

2-4. 品質情報の作成

平均と標準偏差での異常値判別と平滑化による異常値判別の結果を用いて品質情報を作成する。

3. 結果・考察

3-1. 平均と標準偏差での異常値判別の結果

(1) 月毎での異常値判別

月毎の異常値判別では時間帯によって異常値の頻度にかかなりの差が出た。気温に関しては特に、東屋では7時から10時にかけてと12時から14時にかけて異常値が多く検出された。これは佐岡の観測機器の設置場所が要因であると推測される。東屋の観測機器は山の斜面近くに設置されているため、日陰の影響が大きくなる。東屋付近は日が当たらない時

間が長く、9時ごろは後免では気温が上昇するのに対して、東屋の気温はあまり上昇しないため気温差が大きくなり異常値として検出されると考えられる。同様に昼間の12時から3時頃に関しては後免では気温の上昇が続くのに対して、東屋付近では日当たりの関係で気温が上昇しにくいことが、後免との気温差を生じせる一つの要因として考えられる。

相対湿度に関しては前提条件として、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。そのため気温の方で異常値と判断されたならば相対湿度の値にも少なからず影響があると考えられる。東屋の0時から10時における異常値の頻度の変動に着目すると、気温と相対湿度のグラフはおおよそ似たような変動をしていることが分かった。一方仮社殿の気温と相対湿度との関係はこのグラフでは確認できなかった。

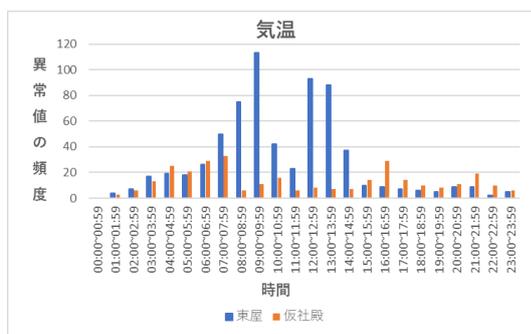


図-3 月毎よる気温の異常値の時間帯別頻度

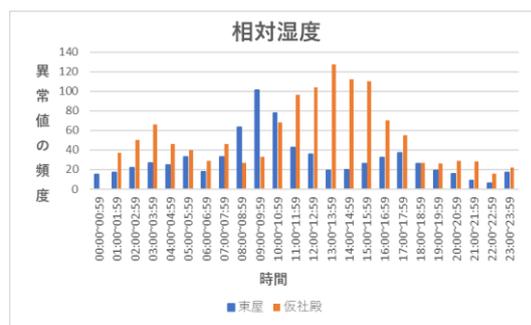


図-4 月毎よる相対湿度の異常値の時間帯別頻度

(2) 月毎と時刻毎での異常値判別

月毎での異常値判別では特定の時間帯で多く異常値と判別されてしまったので、1時間ごとに分けて平均および標準偏差を計算し、異常値判別させることで日当たりの条件によって生じると考えられる異常値を削減することを試みた。気温に関して、東屋では図-5を見ると図-3で多く異常値が判別された8

時、9時、12時、13時台辺りの頻度は減少した。

よって、日当たりの条件によって生じる異常値を緩和することはできた。しかし、東屋と仮社殿共に図-3と図-4では頻度の低かった20時あたりから4時ごろまでの日光が関係しないと考えられる時間帯の夜間の異常値が約5倍から10倍増加した。

また、図-5と図-6では東屋の気温と相対湿度の異常値の頻度はおおよそ似たような変動をしていることがグラフから確認できた。

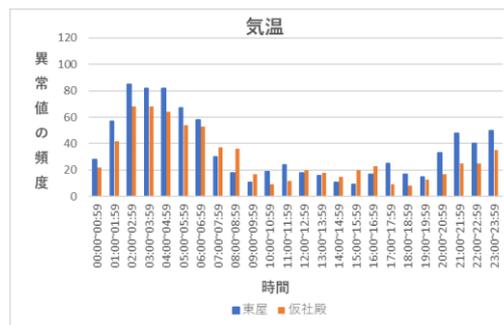


図-5 月毎と時刻毎による気温の異常値の時間帯頻度

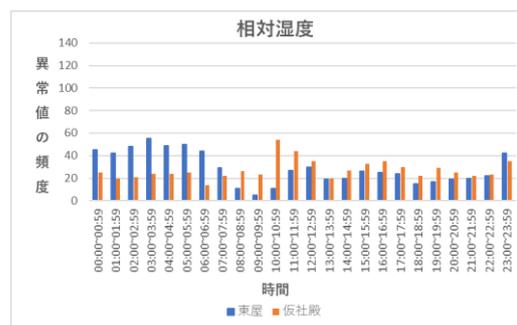


図-6 月毎と時刻毎による相対湿度の異常値の時間帯頻度

3-2. 平滑化による異常値判別の結果

東屋と仮社殿それぞれについて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に示したのは2018年の6月13日の1日における東屋の気温の観測値を平滑化し、平滑化したデータと観測値の差のF統計量(異常度)をプロットしたものである。赤線が閾値を示しており、赤線より上に突出した箇所を異常値と判別する。図-8は2018年の6月13日の東屋の観測データを縦軸気温、横軸時間でプロットしたグラフである。太線が気温の観測値であり、細線が観測値を平滑化したもので、異常値は星マークで表した。グラフより、視覚的に異常値が検出された時間帯を知ることができる。図-7、図-8よ

り異常値と判別されたのは4箇所存在することが確認できた。このようにして1日毎の異常値判別を行った。図-9には平滑化による気温の異常値判別の時間帯頻度を示した。東屋と仮社殿共に8時から17時辺りの時間帯に多く異常値として判別され、夜間の時間帯と比較すると約10倍もの差があった。このことから、平均と標準偏差での月毎の異常値判別の結果と同様に東屋の観測機器の設置場所における日当たりの条件が関係していると考えられる。

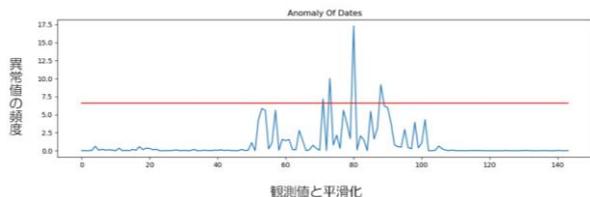


図-7 観測値と平滑化のデータ

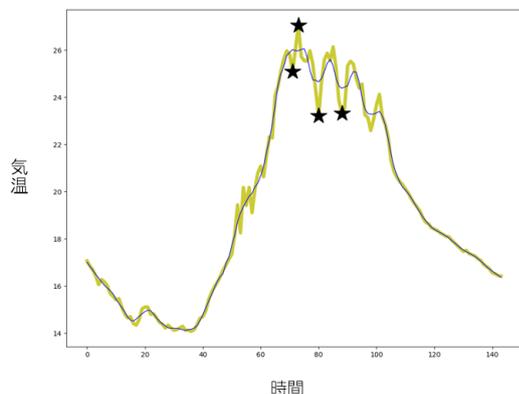


図-8 観測値と平滑化による1日の気温変動

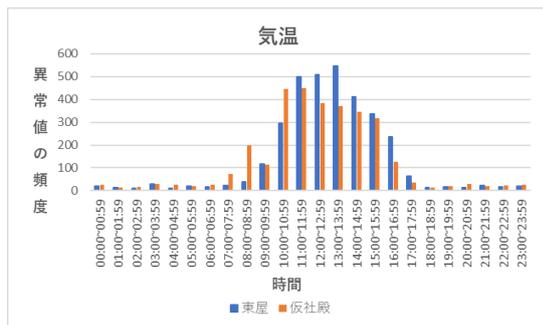


図-9 平滑化による異常値の時間帯頻度

3-3. 品質情報の作成

平均と標準偏差での異常値判別と平滑化による異常値判別で判別した異常値を一つの表にまとめることで佐岡の気象観測データの品質情報の管理を行う。表-1は2018年6月13日の12時から13時50分までの東屋における気温の品質情報を示したものであ

る。異常値を1とし、正常値を0と表記した。このようにして東屋と仮社殿の気温と相対湿度に関して、品質情報を作成した。

表-1 品質情報

東屋			品質情報		
観測値			異常値判別		
日付	時間	気温(°C)	平均・標準偏差(月毎)	平均・標準偏差(月毎と時間毎)	平滑化
2018/6/13	12:00	26.036	0	0	0
2018/6/13	12:10	27.063	1	0	1
2018/6/13	12:20	25.72	0	0	0
2018/6/13	12:30	25.55	0	0	0
2018/6/13	12:40	25.525	0	0	0
2018/6/13	12:50	25.963	0	0	0
2018/6/13	13:00	25.404	0	0	0
2018/6/13	13:10	24.267	0	0	0
2018/6/13	13:20	23.208	0	0	1
2018/6/13	13:30	24.677	0	0	0
2018/6/13	13:40	25.598	0	0	0
2018/6/13	13:50	25.866	0	0	0

4. まとめ

3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。平均と標準偏差の月毎の異常値判別と平滑化による異常値判別では、東屋の観測機器の設置場所の日当たりの条件によって9時から14時辺りまでの気温が変化する時間帯が夜間と比較して多く異常値として判別された。平均と標準偏差の月毎と時刻毎の異常値判別では、月毎の異常値判別の東屋における日当たりの条件によって生じると考えられる異常値を削減することができた。3つの手法より異常値の可能性のあるデータを判別し、品質情報を作成することができた。この品質情報を観測データと共に提供することで、ユーザーは観測データの信用性を評価することができ、観測データを有効に活用できると考える。

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) 気象庁 <https://www.jma.go.jp/jma/index.html>
- 3) 国土地理院：地理院地図- <https://maps.gsi.go.jp/>
- 4) Google Map <https://www.google.co.jp/maps/?hl=ja>
- 5) 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