

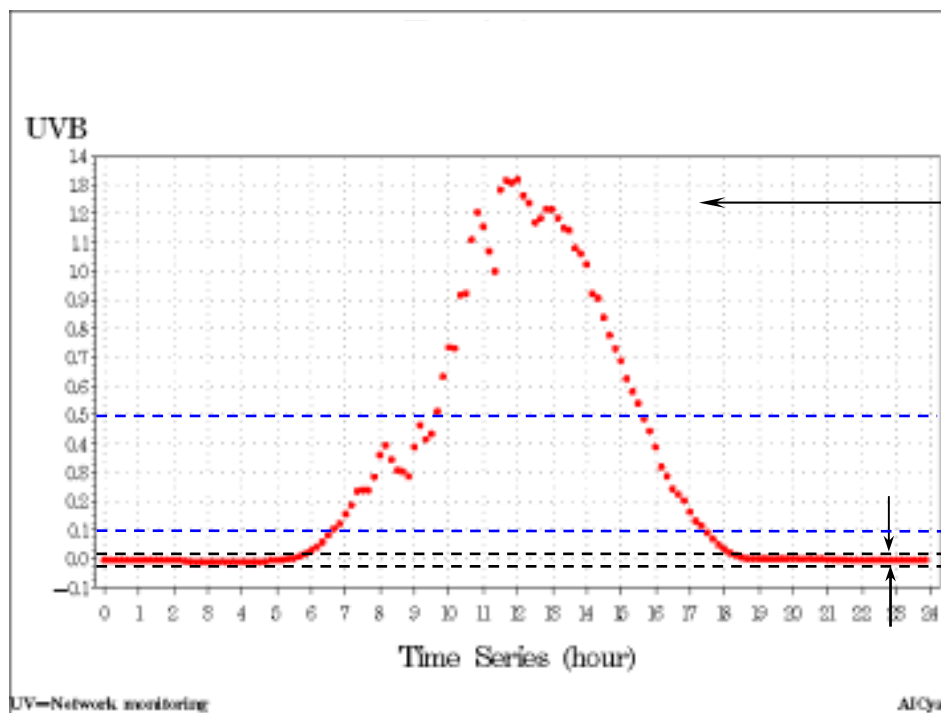
データ検証作業の進捗状況

UVモニタリングネットワーク：データ検証ワーキンググループでは欠測データおよび特異データを分類し、その特徴に基づいて異常データの検討を行った。
その進捗状況の報告と今後の問題点を提起する。

- 目次
1. 放射強度の目安と典型的時系列パターン
 2. 欠測データ及び特異データの分類と特徴
 3. データ解析からみた特徴
 4. データ処理の流れ

1-1. UV B 放射強度の目安と典型的時系列パターン

UV - B 放射強度分布のピーク値の目安 : $0.1 \sim 2.5$ [W / m^2]



晴天時の測定では、1 ~ 2 程度

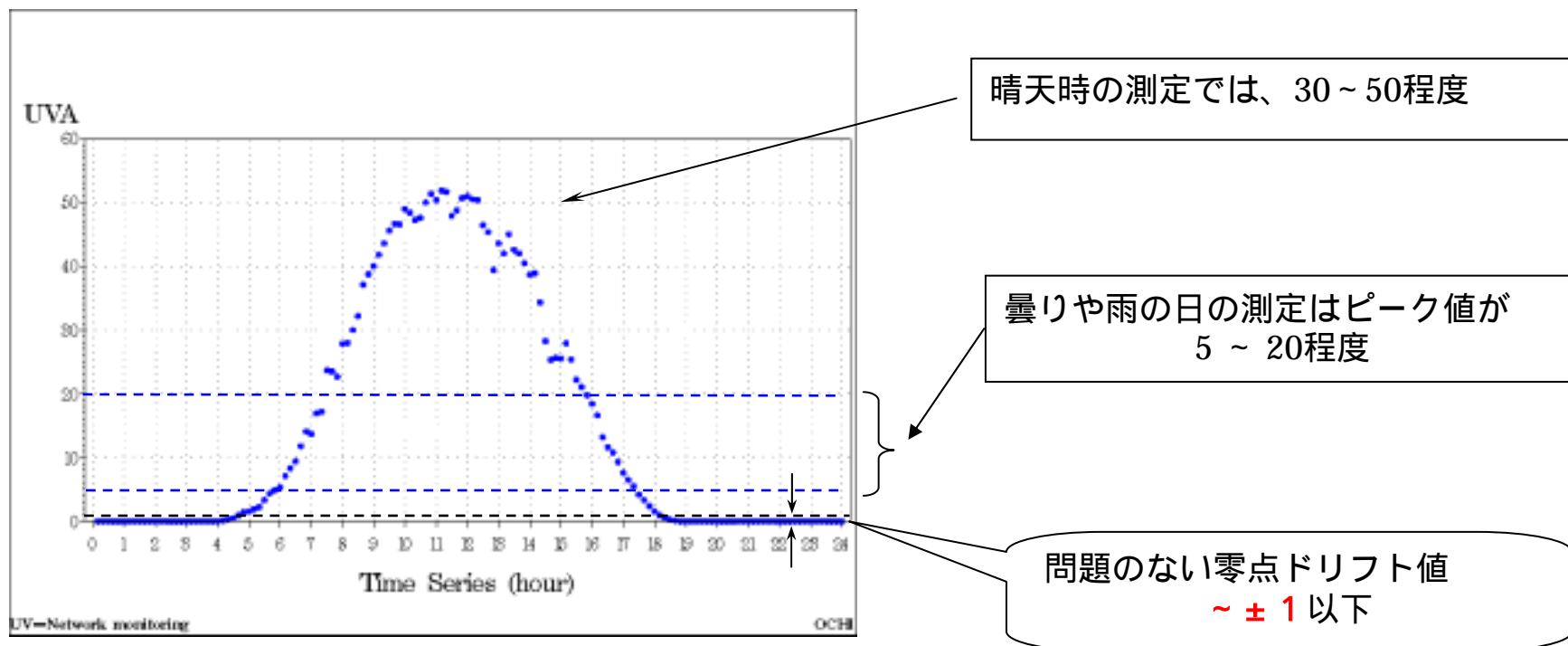
曇りや雨の日の測定はピーク値が
0.1 ~ 0.5程度

問題のない零点ドリフト値
~ ± 0.01 以下

日中でもピーク値が 0.01 程度かそれ以下の場合は降雪などの
問題や機器の不良などが発生している可能性があります。

1-2. UV A 放射強度の目安と典型的時系列パターン

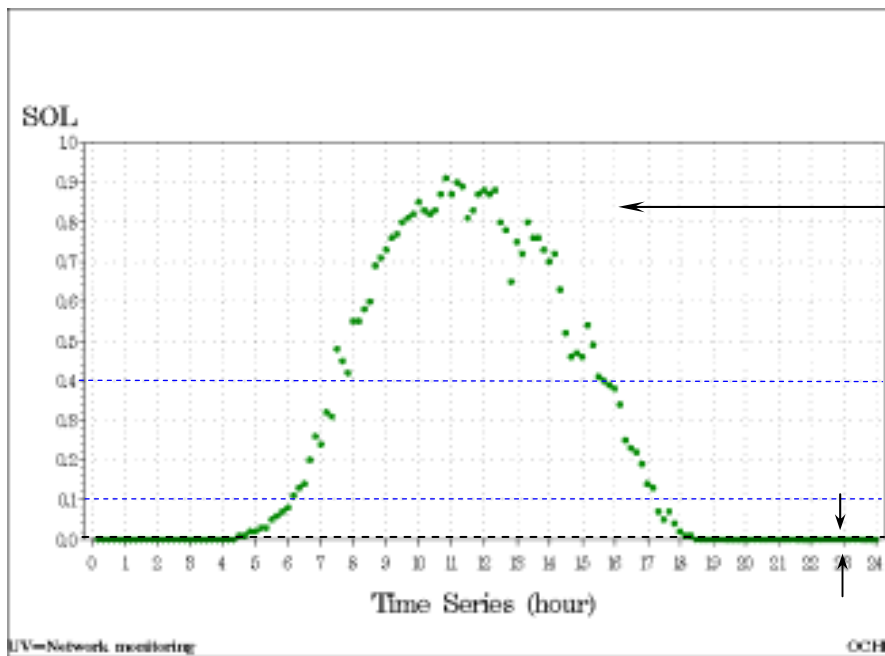
UV - A 放射強度分布のピーク値の目安 : 5 ~ 60 [W / m²]



日中でもピーク値が 1 程度かそれ以下の場合は降雪などの問題や機器の不良などが発生している可能性があります。

1-3. SOL 放射強度の目安と典型的時系列パターン

SOL 放射強度分布のピーク値の目安 : $0.1 \sim 1.2$ [kW/m²]



晴天時の測定では、0.5 ~ 1.0程度

曇りや雨の日の測定はピーク値が
0.1 ~ 0.4程度

問題のない零点ドリフト値
~ ± 0.01 以下

日中でもピーク値が 0.01 程度かそれ以下の場合は降雪などの
問題や機器の不良などが発生している可能性があります。

2. 欠測データおよび特異データの分類とその特徴

欠測データ

欠測報告を受けたデータ、「-999」などの識別数値、また、抜け落ちてしまっているデータや「0値データ（終日0値）」を欠測データと定義し処理しています。

特異データ

特異データは、UVネットワーク・ワーキンググループでの検討結果を踏まえ、データの特性に応じて処理します。特異データには標準的時系列パターンと比較することで容易に確認できるものと、データ間の相互関係や統計的性質から確認できるものとがあります。

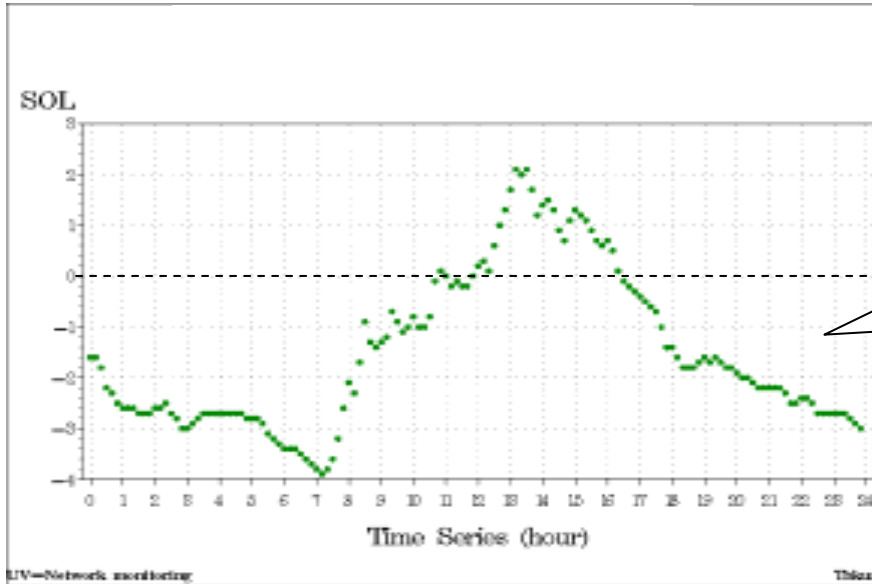
以下、これまでの検証作業で明らかになった特異データの実例を、項目リストと共に挙げてゆきます。その後でデータの相互関係や統計的性質の解析例を示します。

- ベースレベルの異常（零点の異常ドリフト：BL）
- 零値の測定（昼間などに突然計測値が零になる現象：MZ）
- 突出異常（OS）
- 非物理的形状異常（データ処理の過程などから入る異常値全般：NP）
- ノイズ混入（異常ノイズの混入：NS）
- データ属性の異常（日付や時間の付け間違い等：AT）

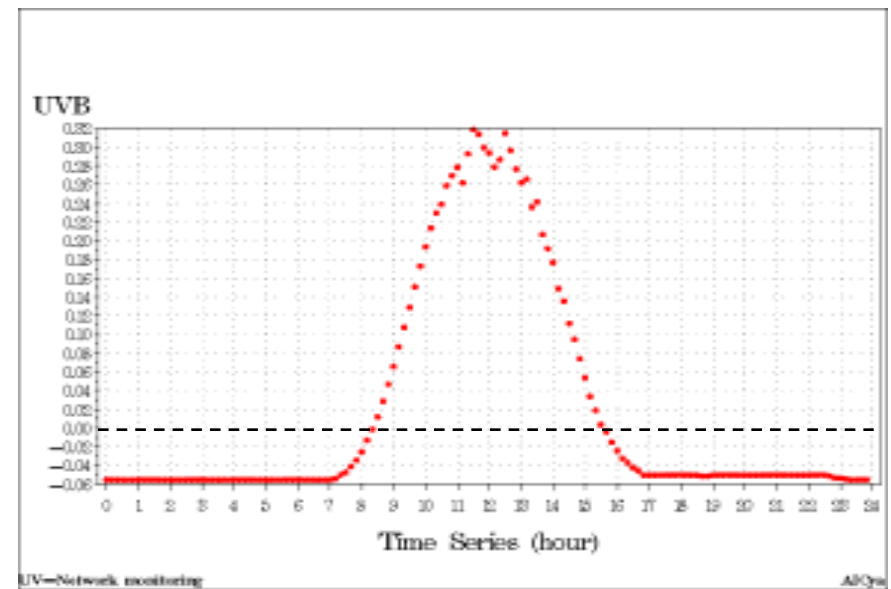
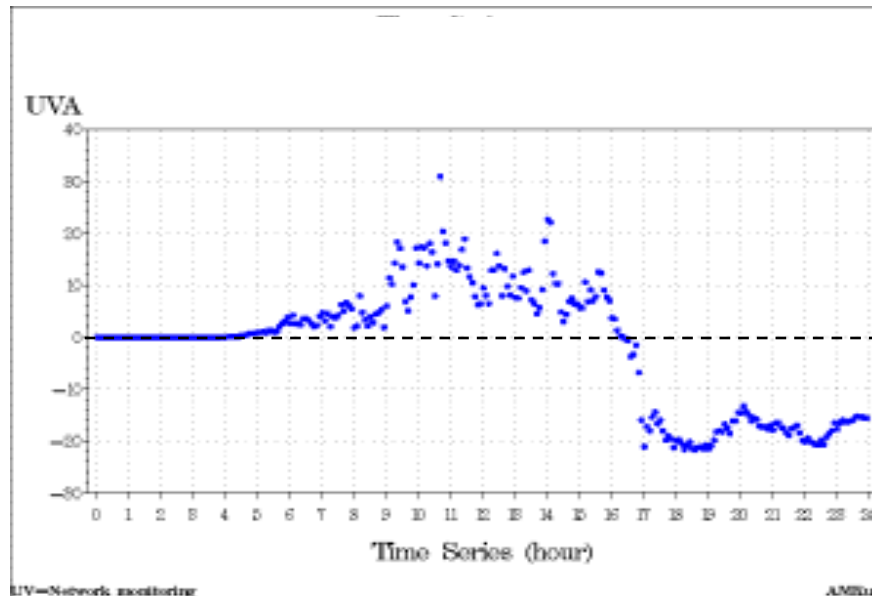
2-1. ベースレベルの異常 (BL)

零点レベルの不安定動作に起因する異常
ドリフト異常、データの揺らぎ、跳び、符号
など…

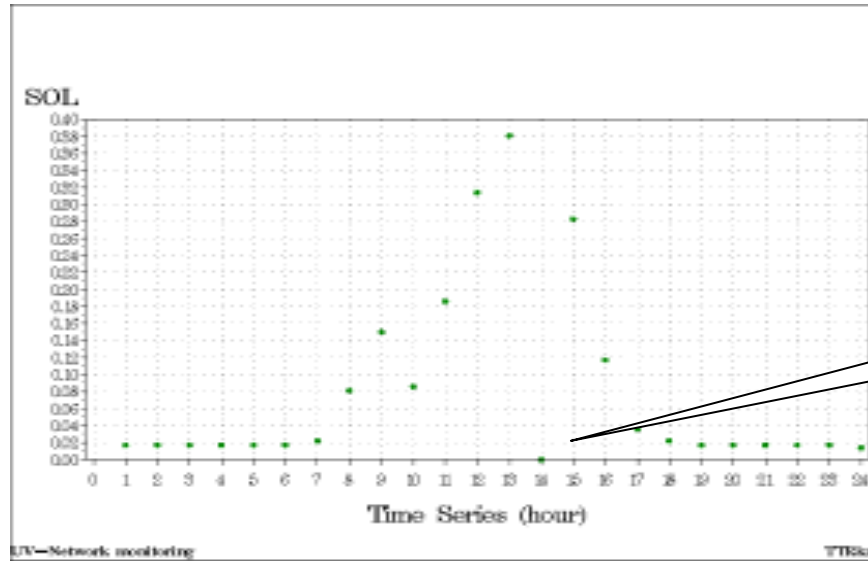
アースの不良等の可能性



典型例



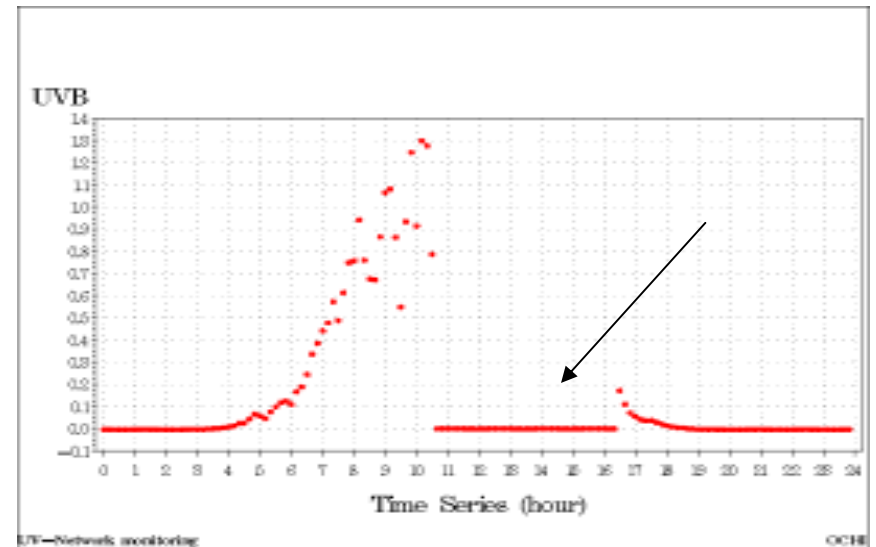
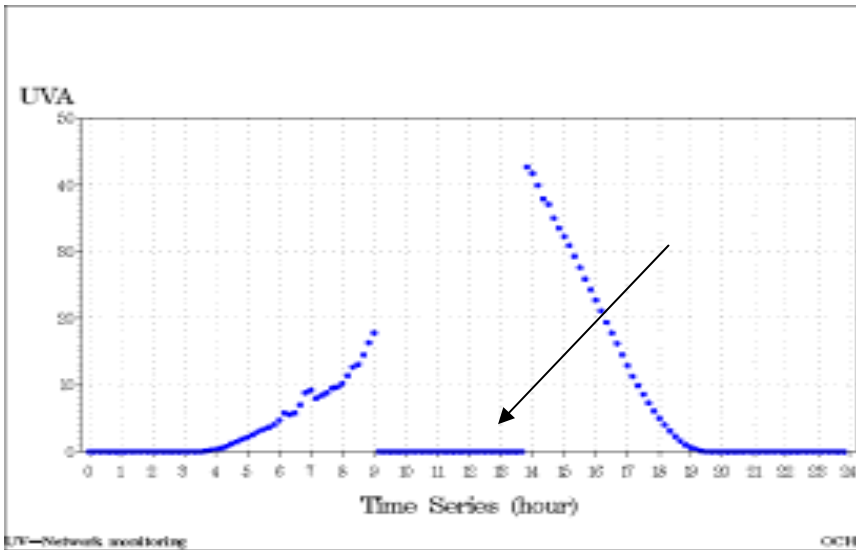
2-2. 零値の測定(MZ)



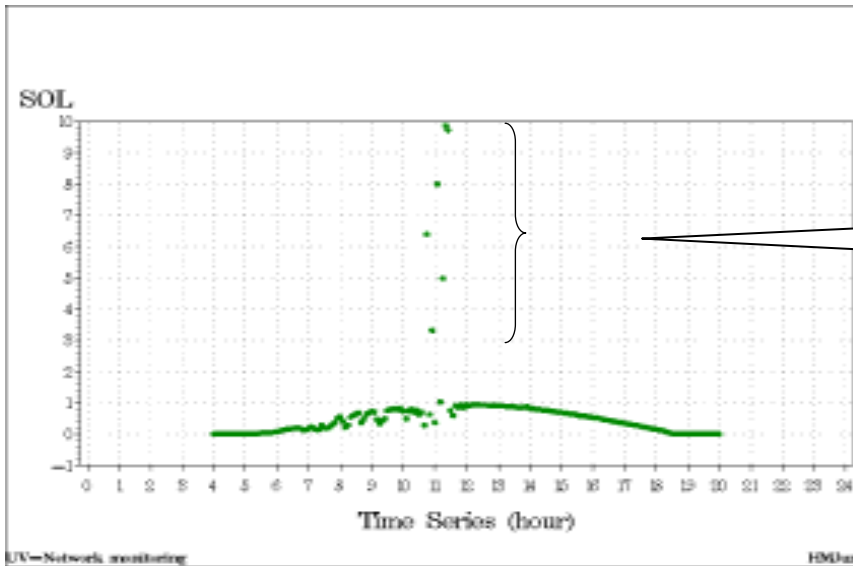
日中、突然計測値が零になる

- ・オーバーフロー、結線不良の可能性

典型例



2-3. 突出異常(OS)

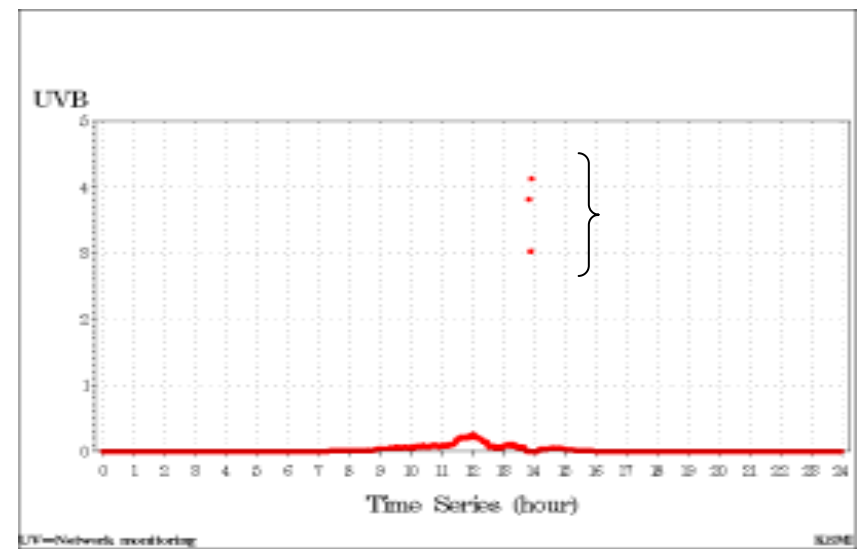
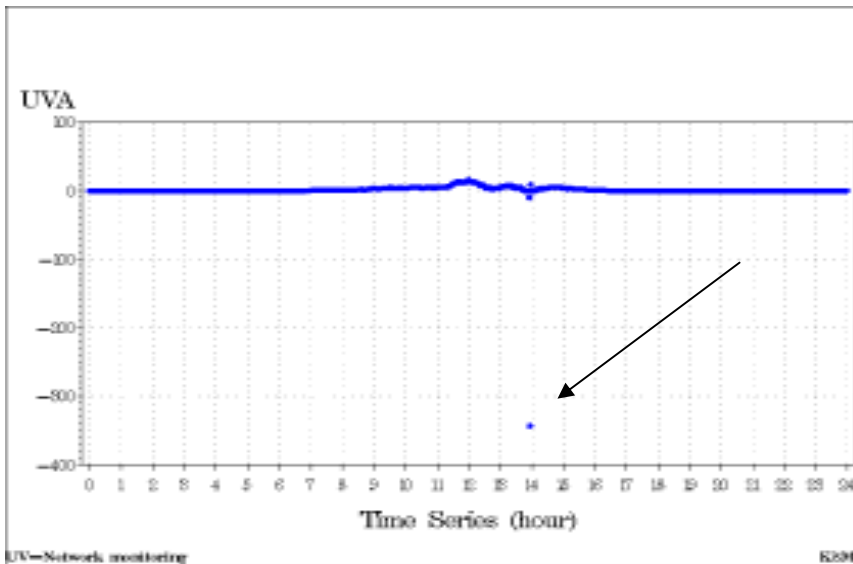


通常の10倍以上の値（正・負）を
突発的に持つデータが現れる

（判断が困難なデータも多い）

一時的な結線不良や
ノイズ混入の可能性

典型例



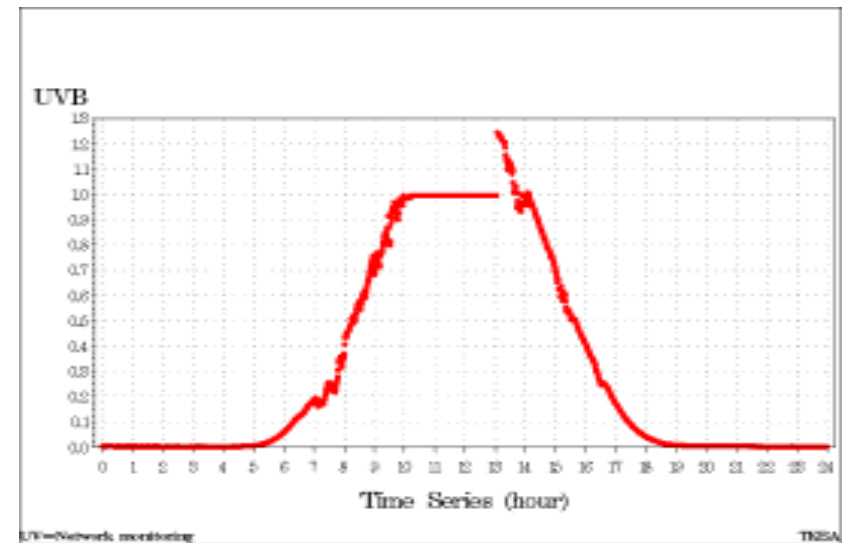
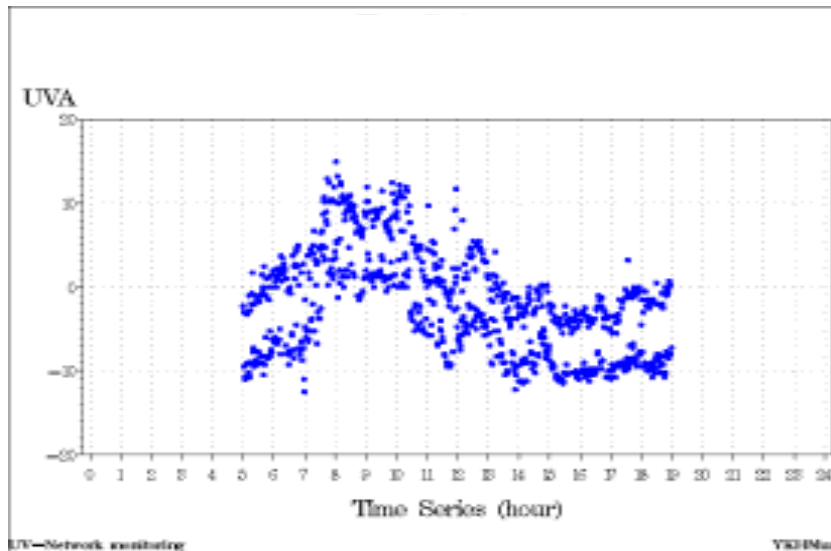
2-4. 非物理的形状異常 (NP)

定期的に起こる日陰の影響は異常値として
処理せず個別に対応する

物理的には困難な形状などの
特徴を有するグラフやパターン

データ処理の問題（データ処理上の問題であれば原因が解り次第修正し処理しなす）

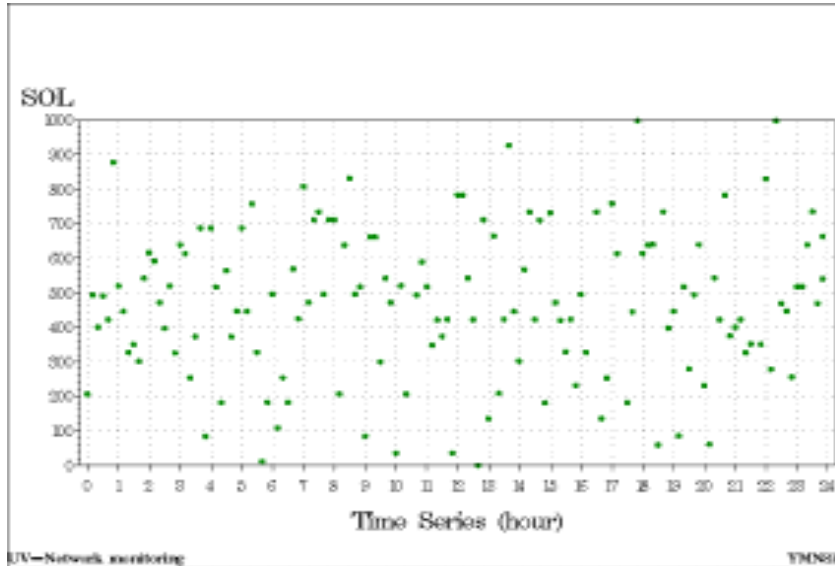
典型例



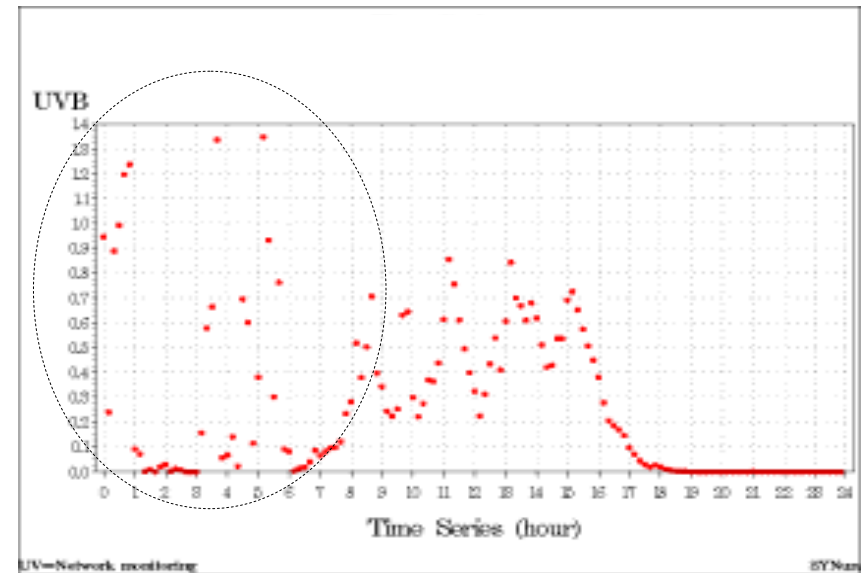
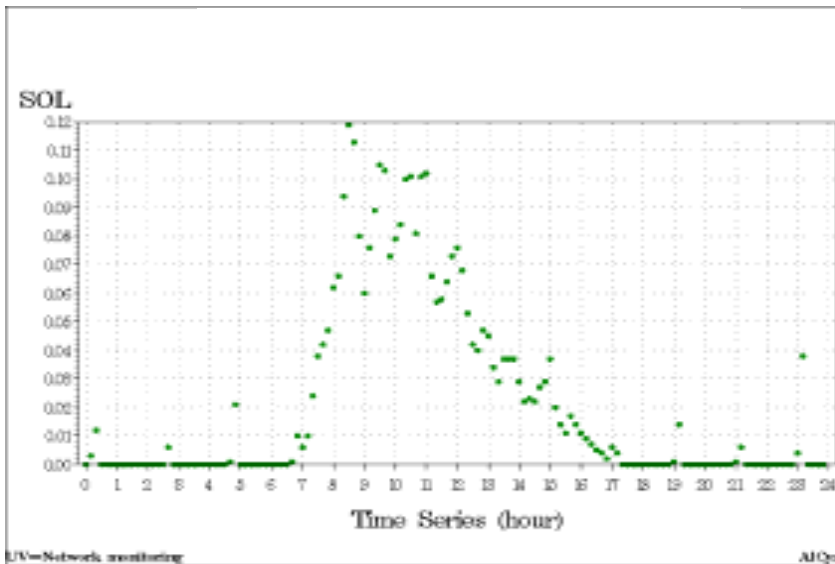
2-5. ノイズ混入(NS)

多様なノイズのパターンがある

結線の不良（開放）や外部機器の（周期的）ノイズ、（人工）光が継続的に混入した可能性



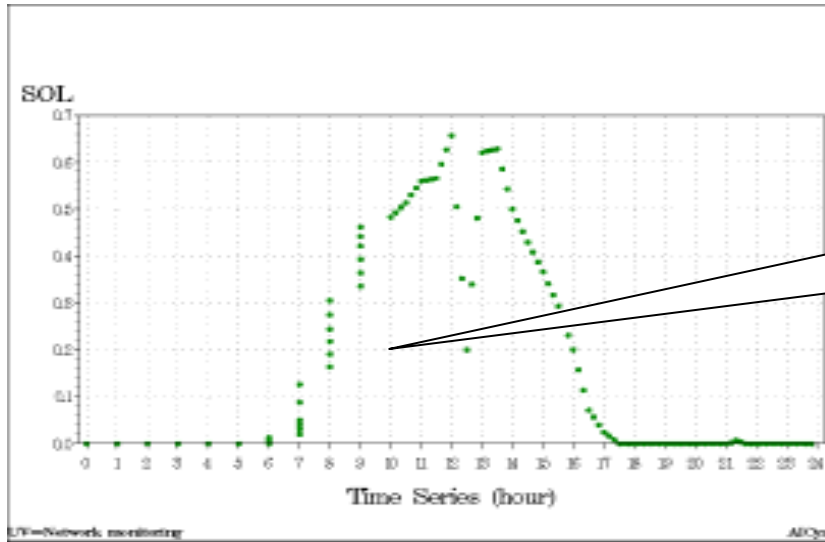
典型例



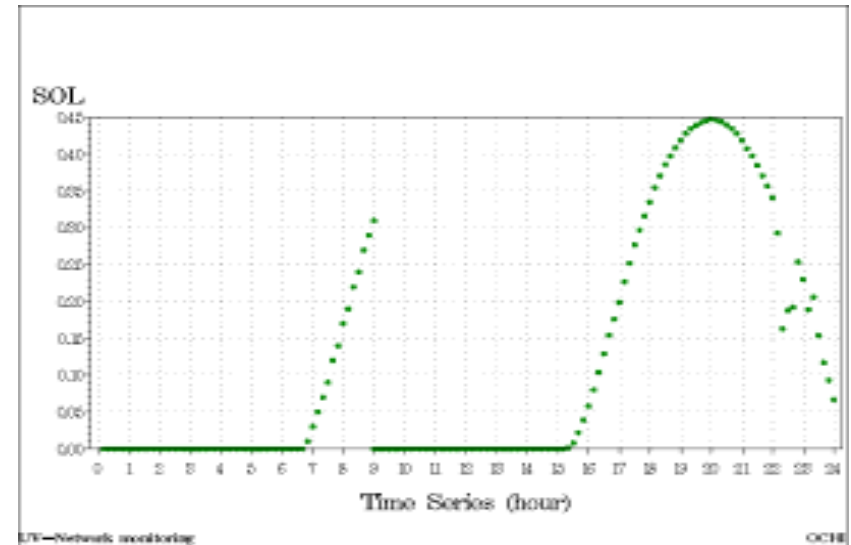
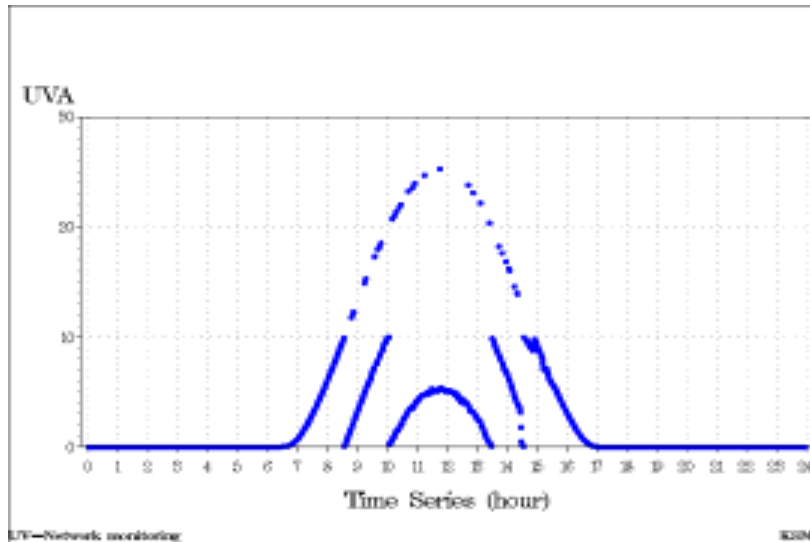
2-6. データ属性の異常(AT)

測定時間がずれたり、時間・日付の情報が欠落するなどの異常
(左の例では「分」の情報が欠落している)

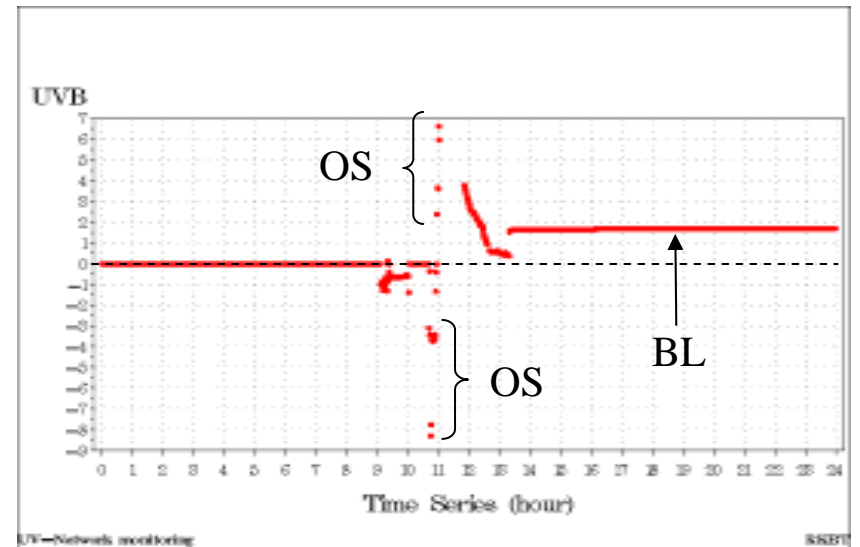
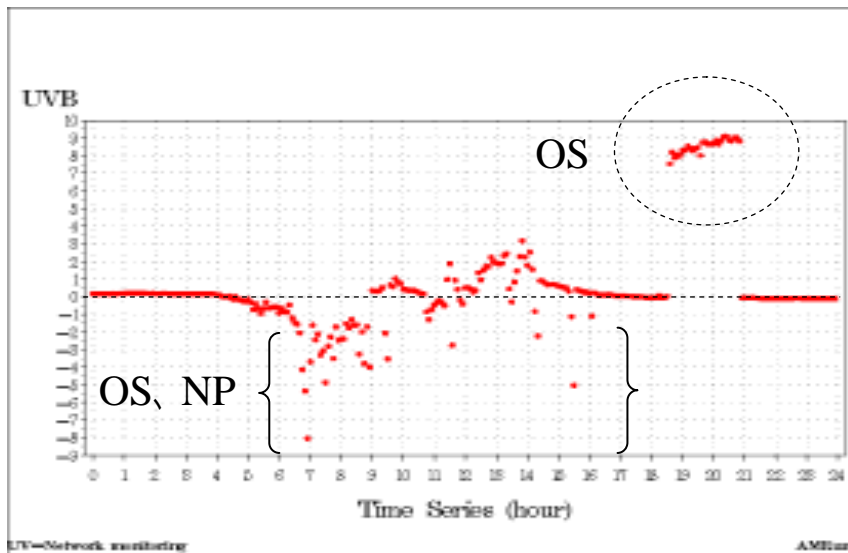
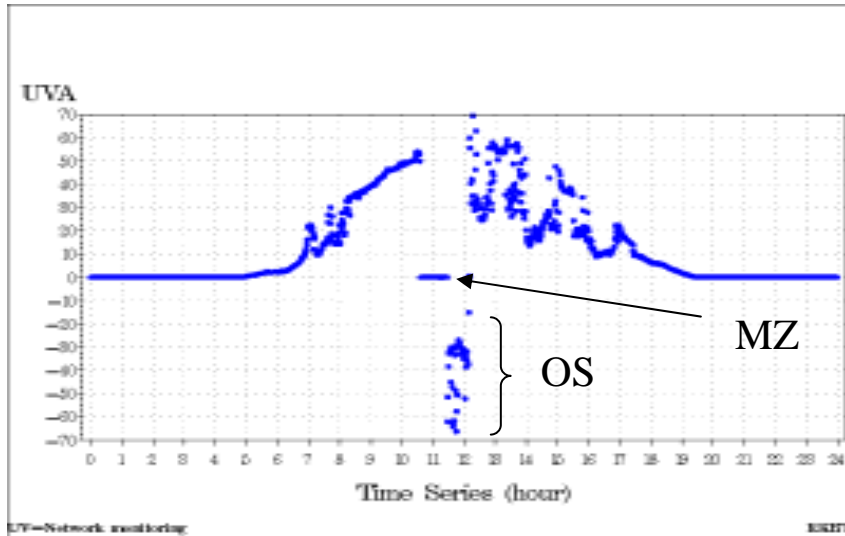
データ処理における問題
の可能性



典型例

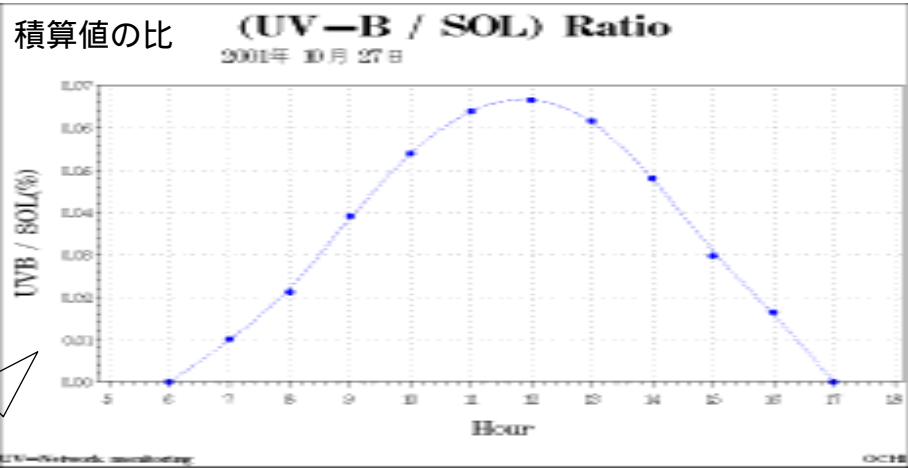
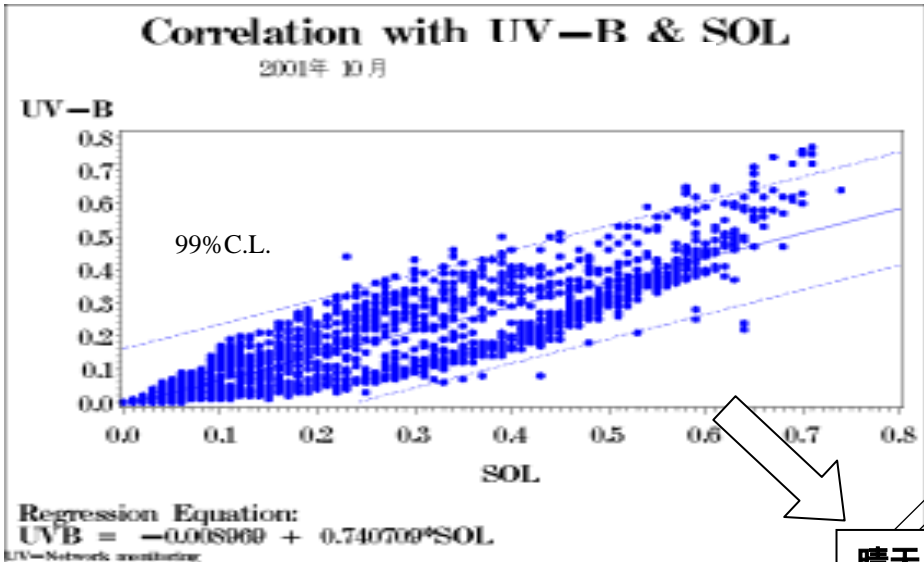


2-7. 複合異常データの例

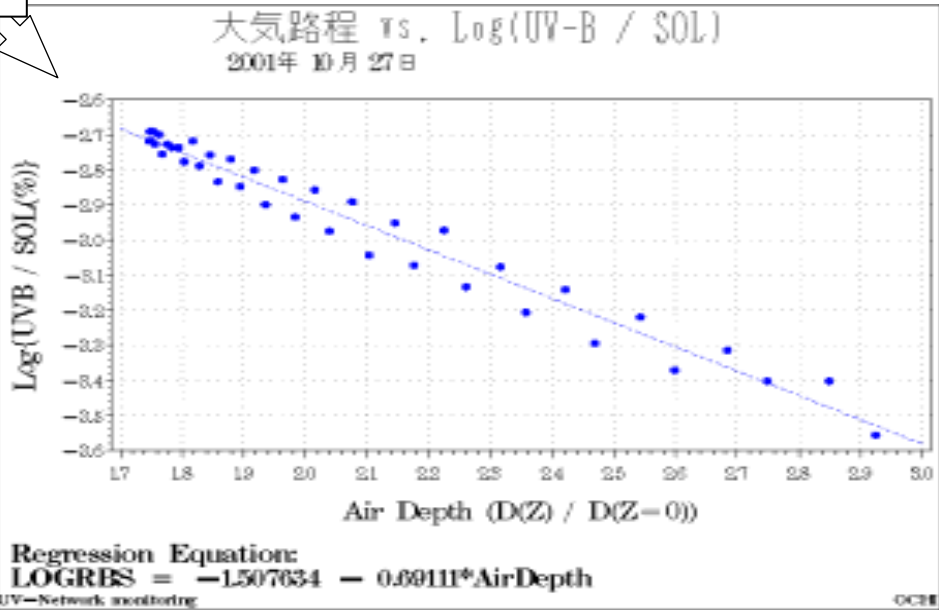
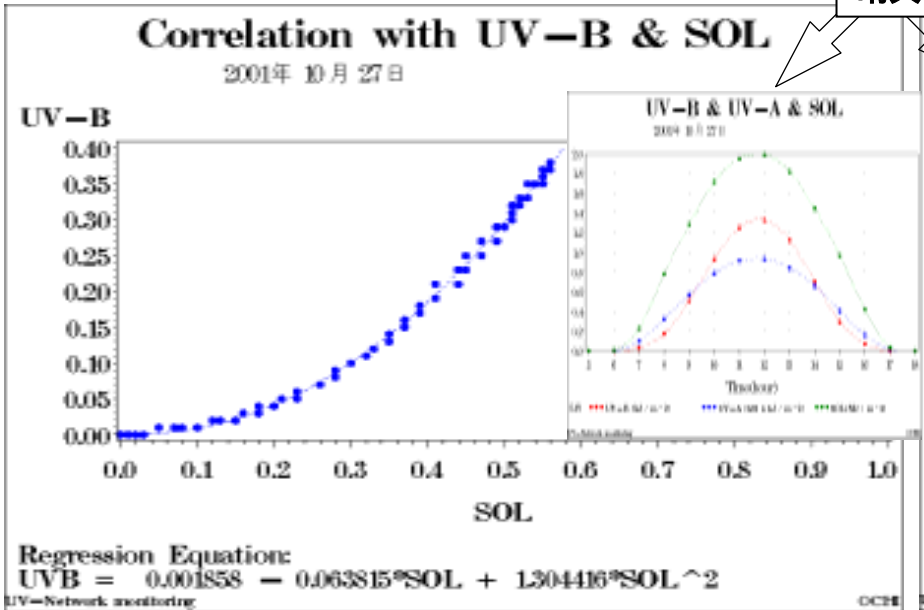


3-1. データ解析から見た特徴

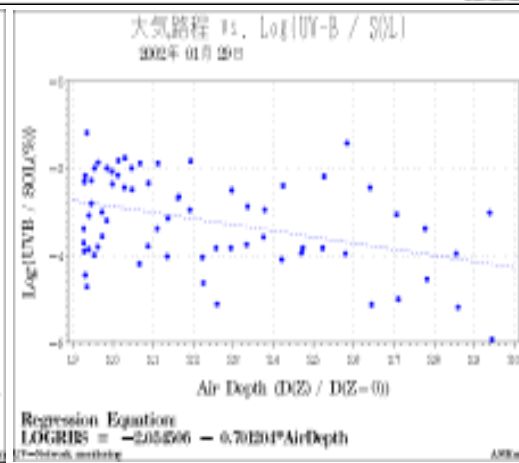
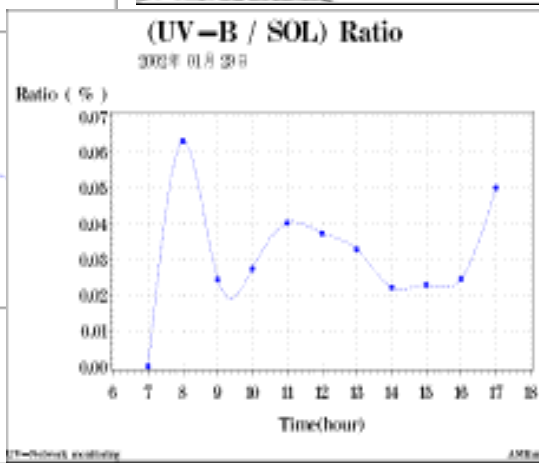
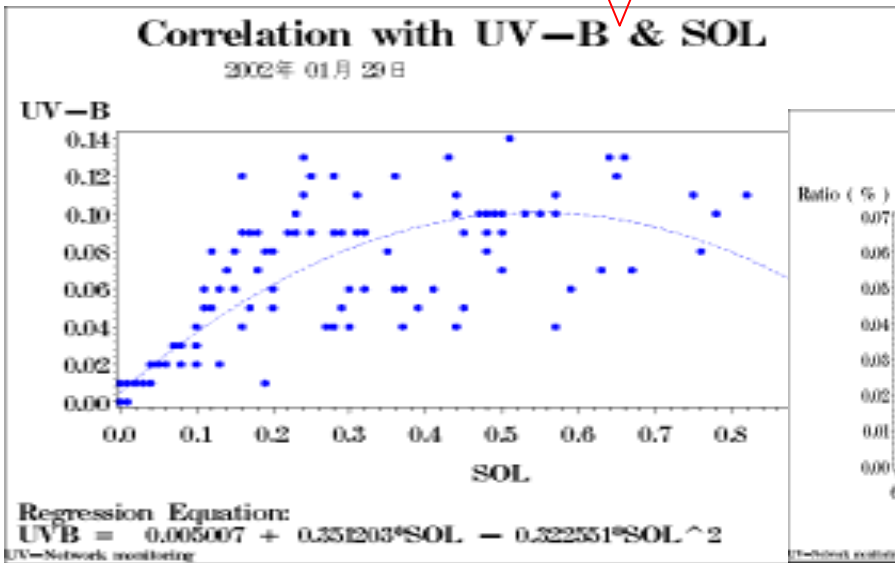
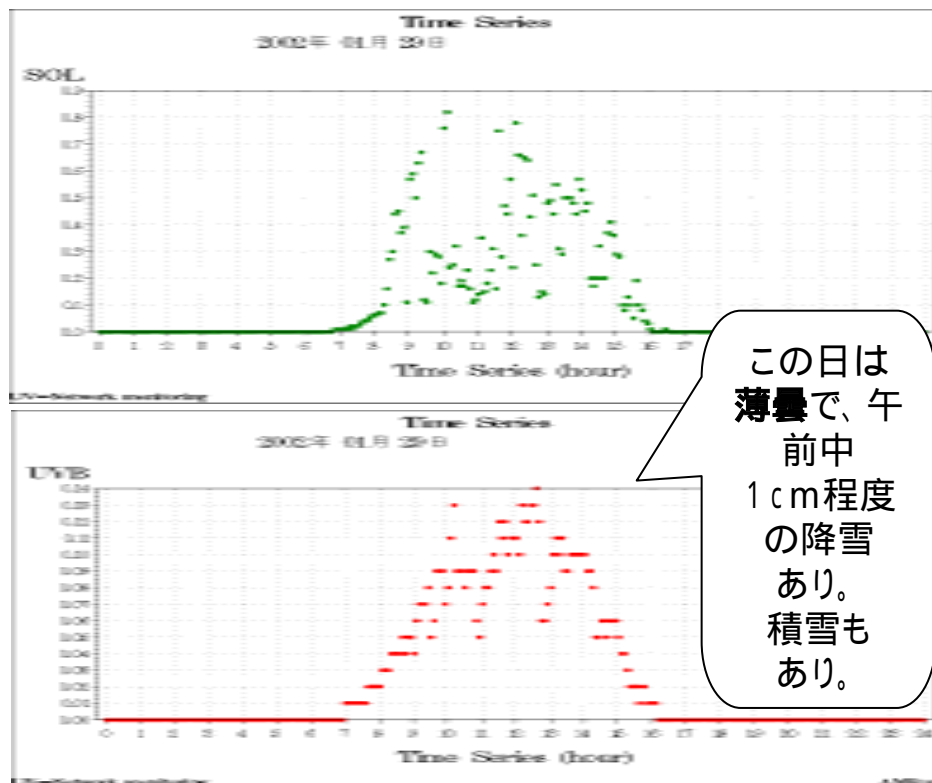
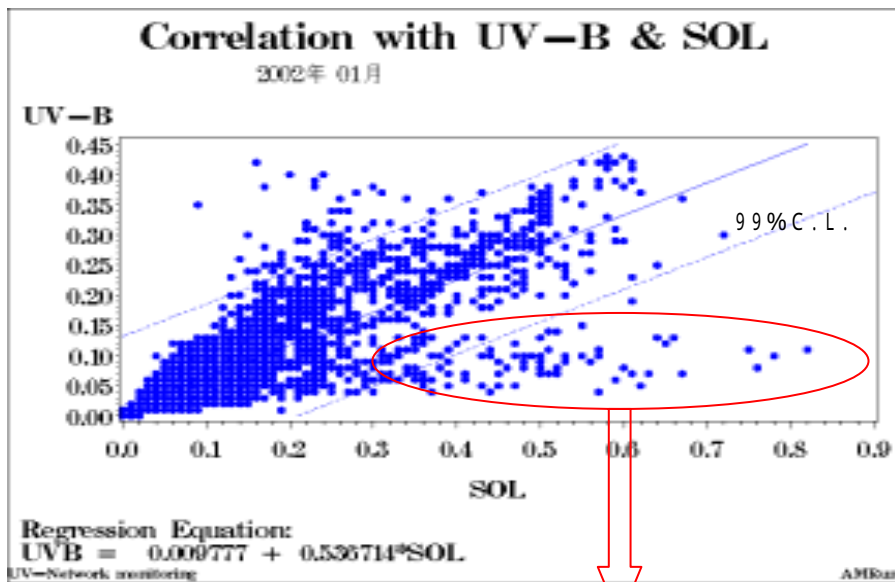
UV B と SOL の相関
(2001年10月 と 10月27日の例)



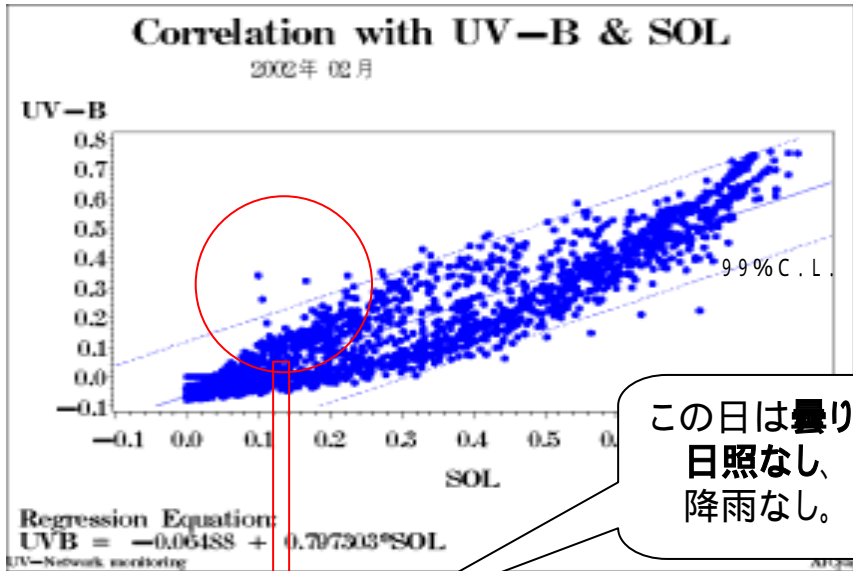
晴天



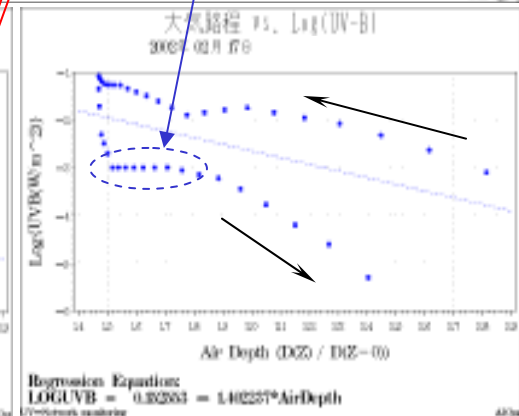
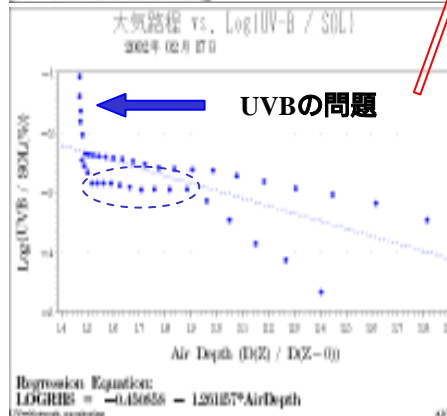
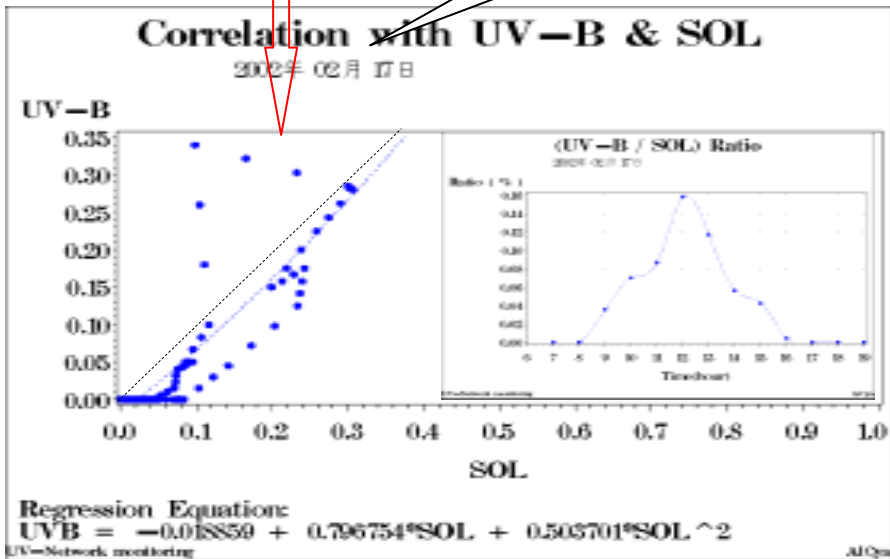
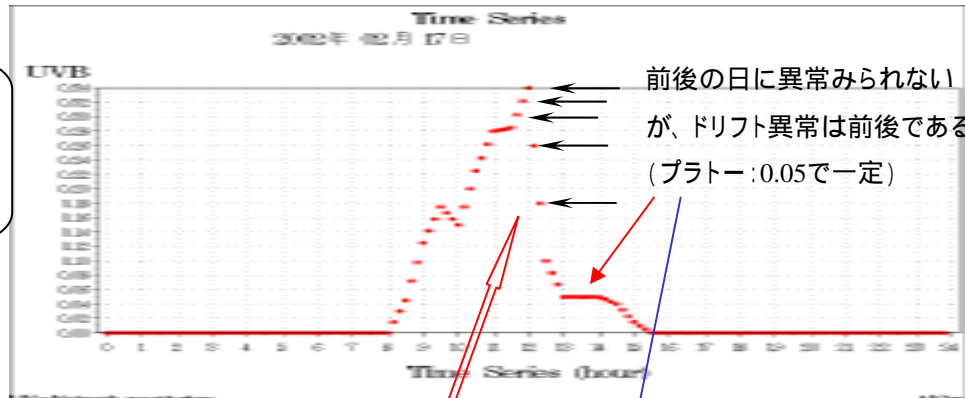
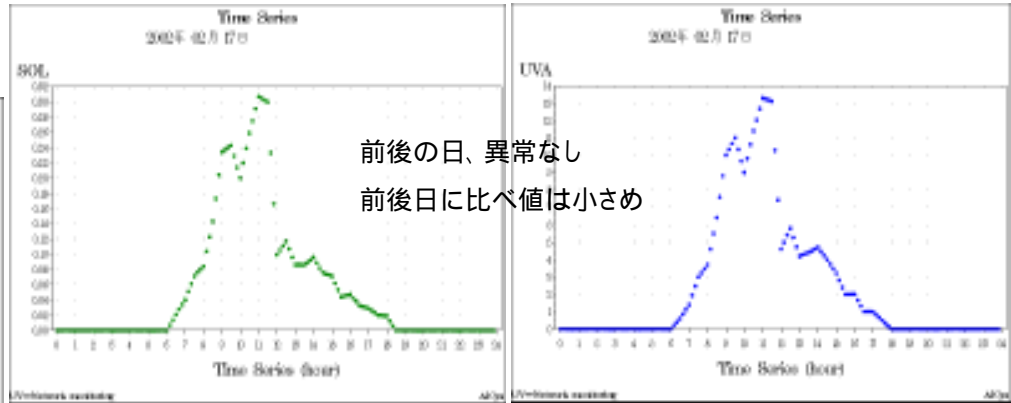
3-2. データ解析から見た特徴



3-3. データ解析から見た特徴



この日は曇り、
日照なし、
降雨なし。



4-1.データ処理の流れ

共通フォーマット化データ

各測定局へ

欠測処理 : 60分中(合計で)10分間より多い欠測データがある場合
その1時間は欠測扱い

特異データ削除 : 特異データの分類に従い削除、その1日は欠測
原則として処理は「日」単位

夜間データの確認 : 異常動作、停電等のチェックはするが、零点ドリフトの補正はしない

データ値の確認 : 平均値やUV-A・UV-B・SOLの時刻相関、またUV-Bと
SOLARの比などから総合的なチェックを行う

この段階で各測定局へ連絡(報告と確認)

固定データ

前1時間平均値に整形

例として、5時の前1時間平均値
とは4:01~5:00までの平均値
を指す

「1~24時」の時間系を採用

機器の校正 : 機器の校正結果を基に、比例分配(「日」単位)で補正

確定データ

数値データの表示方法と精度(有効数字4桁) : \pm . $E \pm$ を採用

共有化、公開に向けて処理(次項データ形式詳細)

4-2.データ処理の流れ(データ形式の詳細)

データはCSV形式で統一し、月単位での処理を基調とする

時別値 [前1時間平均値、1～24時間]

形式: { 年、月、日、時、UV B [W/m^2]、UV A [W/m^2]、SOL [kW/m^2] }

日別値 [日積算値 : 日の出時刻・日の入り時刻を考慮した日積算値()]

形式: { 年、月、日、UV B [kJ/m^2]、UV A [kJ/m^2]、SOL [MJ/m^2] }

但し、時別値に欠測データがあれば、その日の日別値は欠測扱い(*、)

月別値 [日別値の月別平均値、1～12月]

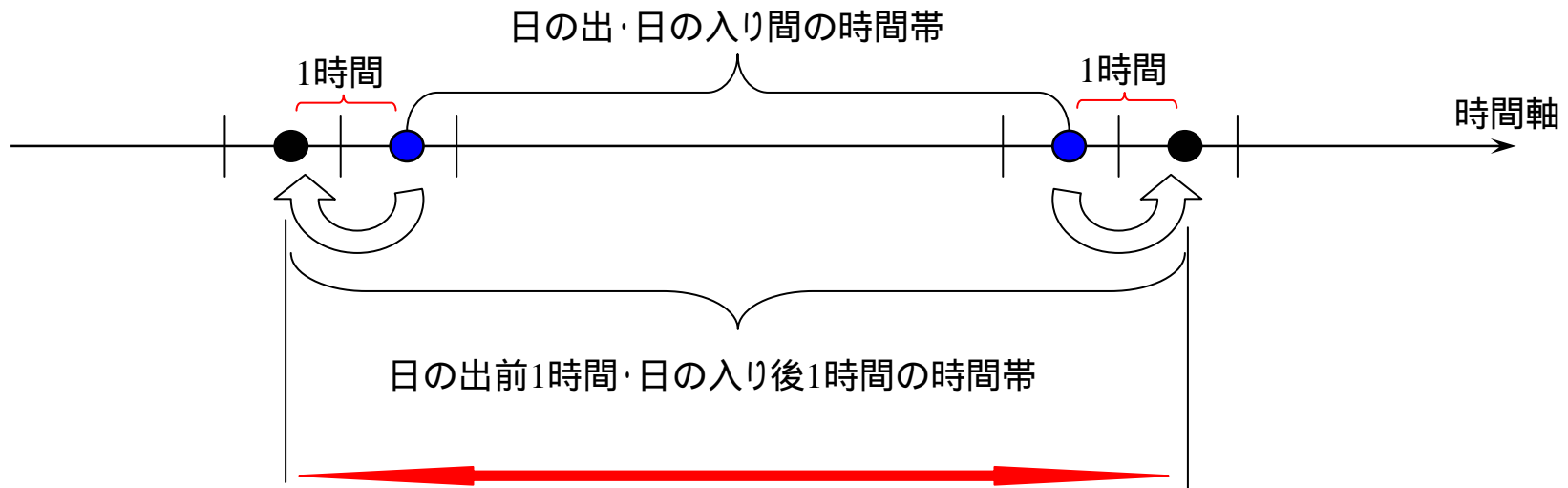
形式: { 年、月、UV B [kJ/m^2]、UV A [kJ/m^2]、SOL [MJ/m^2]、各データ数、各標準誤差 }

月毎に日別値の平均値を計算、1日でも日別値があれば平均処理する(処理データ数を報告(*))

- * … 暫定的な処方であり、今後の更なる検討を要す
- … 次の頁参照

4.3. データ処理の流れ

「日の出前1時間・日の入り後1時間」のデータと「時別値」および「日積算値」



「日積算値」は、この時間帯の時間平均値の積算で算出する

- 注意1** : 欠測ポイントがある場合、欠測ポイントを平均値で補間する必要がある。
注意2 : 日の出・日の入り時刻には数分の誤差が含まれる。

日の出・日の入り時刻の算出について : 太陽の視赤緯、均時差は理科年表を参照し、太陽の「浮き上がり」補正をしてある。高度は全ての局で海拔0m。

データ検証作業の進捗状況

特異データの分類とその処理()

新たに特異データの分類に『**相関の問題 (CO)**』を取り入れた。



UV-A、UV-B、SOLARの瞬時値を時間で直接比較

瞬時値、日別値、UV-B / SOLARの通年表示(異常値の棄却検定)

必要ならJIS K 8402及びISO 5725に規定されている

グラブス・スミルノフ棄却検定法でチェック

まとめ

異常・特異データを判別する為の項目リスト (略記)

特異データ分類 : ベースレベル異常 (BL)

特異データ分類 : 零値測定異常 (MZ)

特異データ分類 : 突出異常 (OS)

特異データ分類 : 非物理的形状 (NP)

特異データ分類 : ノイズ混入 (NS)

特異データ分類 : データ属性の問題 (AT)

特異データ分類 : 原因不明の異常 (AB)

特異データ分類 : 測定感度の問題 (MS)

特異データ分類 : 相関の問題 (CO)

原因と対策

アース等の問題、原則として使用しないが、その旨明記して使用することも可

オーバーフローの可能性あり。
零値は原則として欠測値扱い

結線開放の可能性。突出異常を含む
データセットは原則使用しない

データ処理の問題、
データセットは原則使用せず

結線開放の可能性、外部ノイズ源あり

データ処理での問題

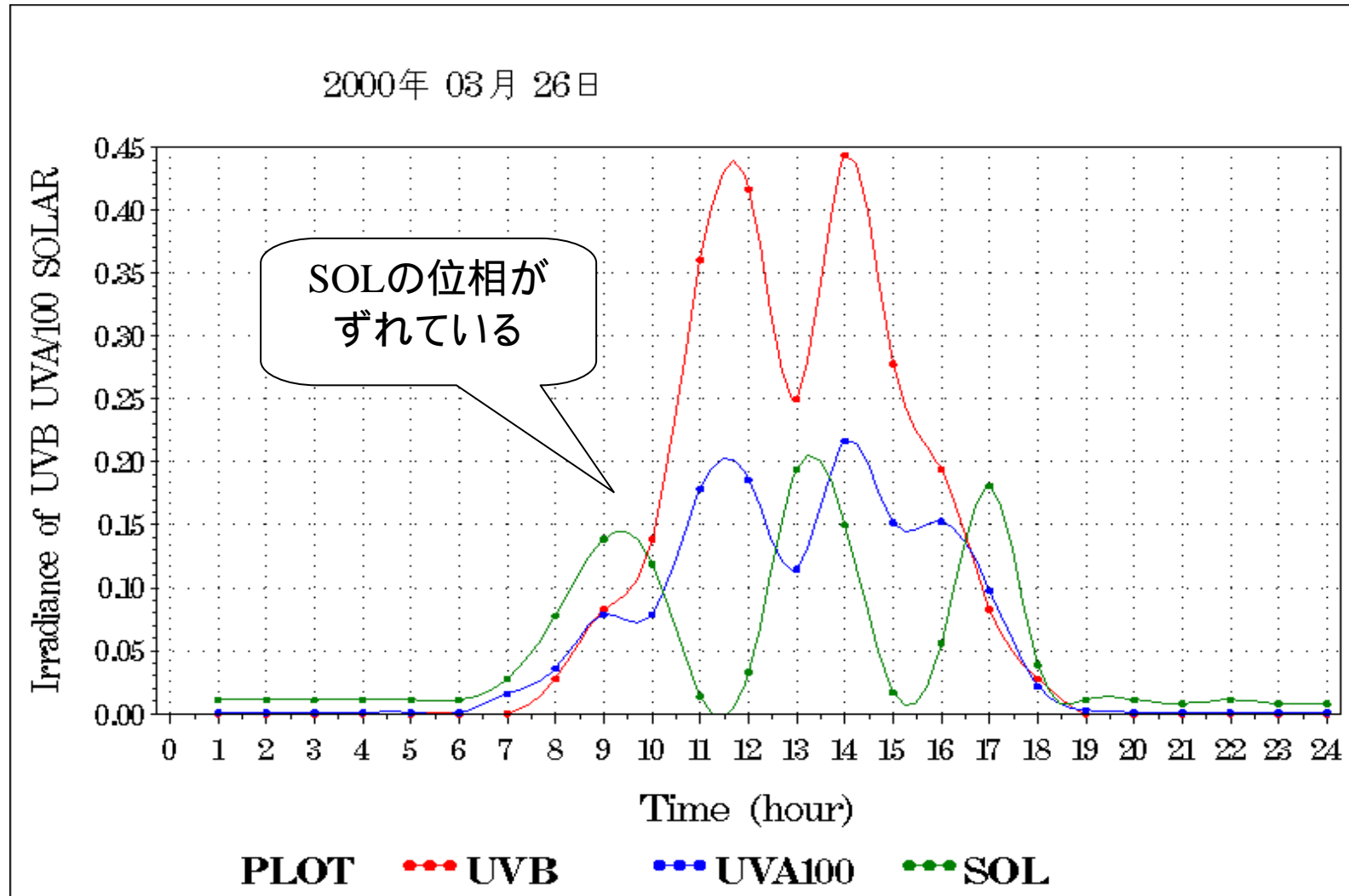
判断が微妙なので、欠測処理できない

SOLAR、UV-A、UV-Bの間の相関の異常
測定環境の確認が必要

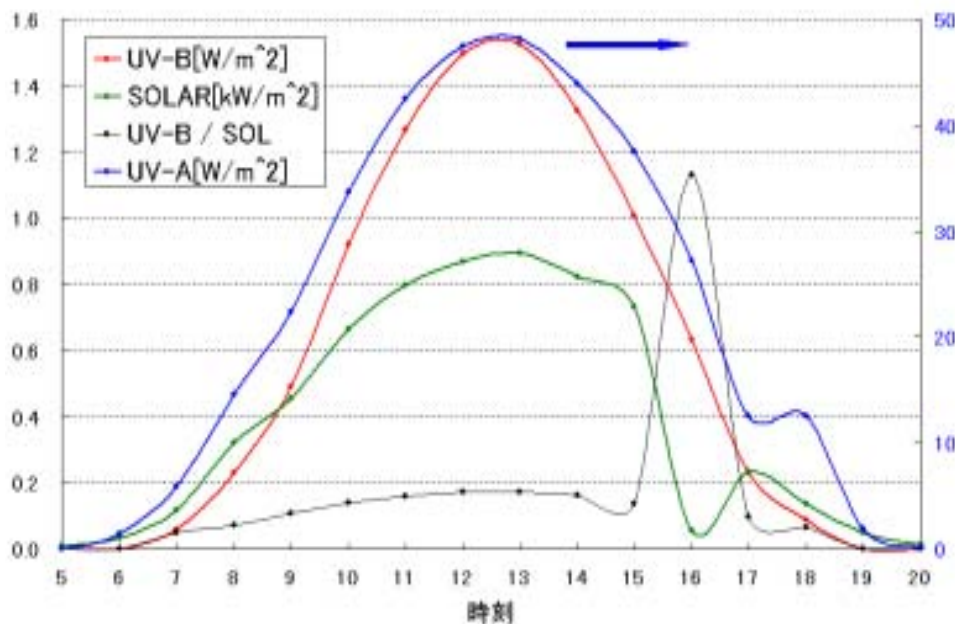
原則的に欠測値として処理

特異データ分類 : 相関の問題 (CO)

・SOLAR、UV-A、UV-Bの間の相関に明らかな問題がある場合。**必ず、観測局の報告を考慮する**



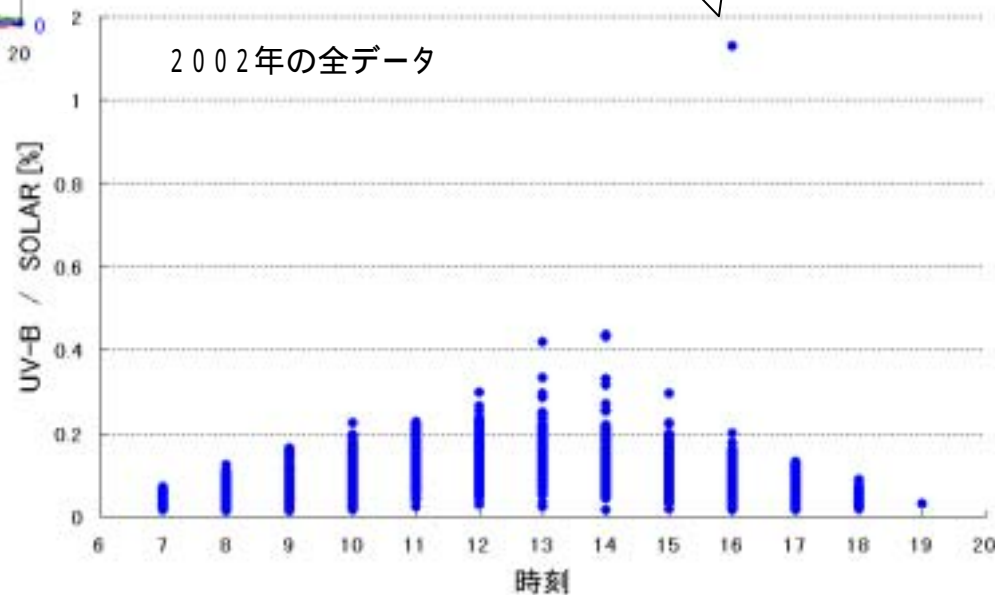
特異データ分類 : 相関の問題 (CO)



統計的には、危険率1%で棄却され得るデータ

しかし、観測局の報告がない場合は削除しない。

UV-A・Bとの関係ではSOLARが一時間ずれている可能性もある

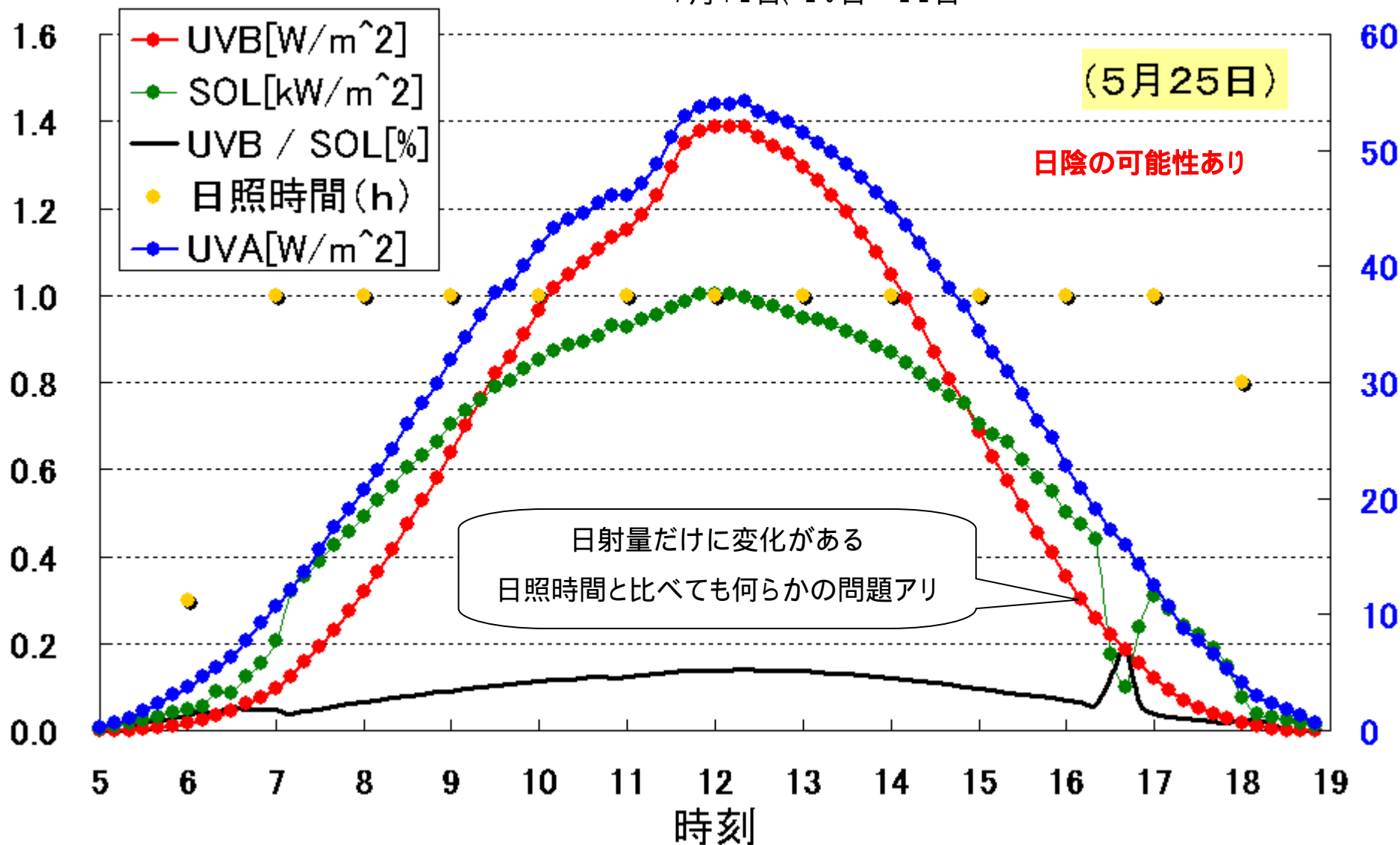


特異データ分類 : 関連の問題 (CO)

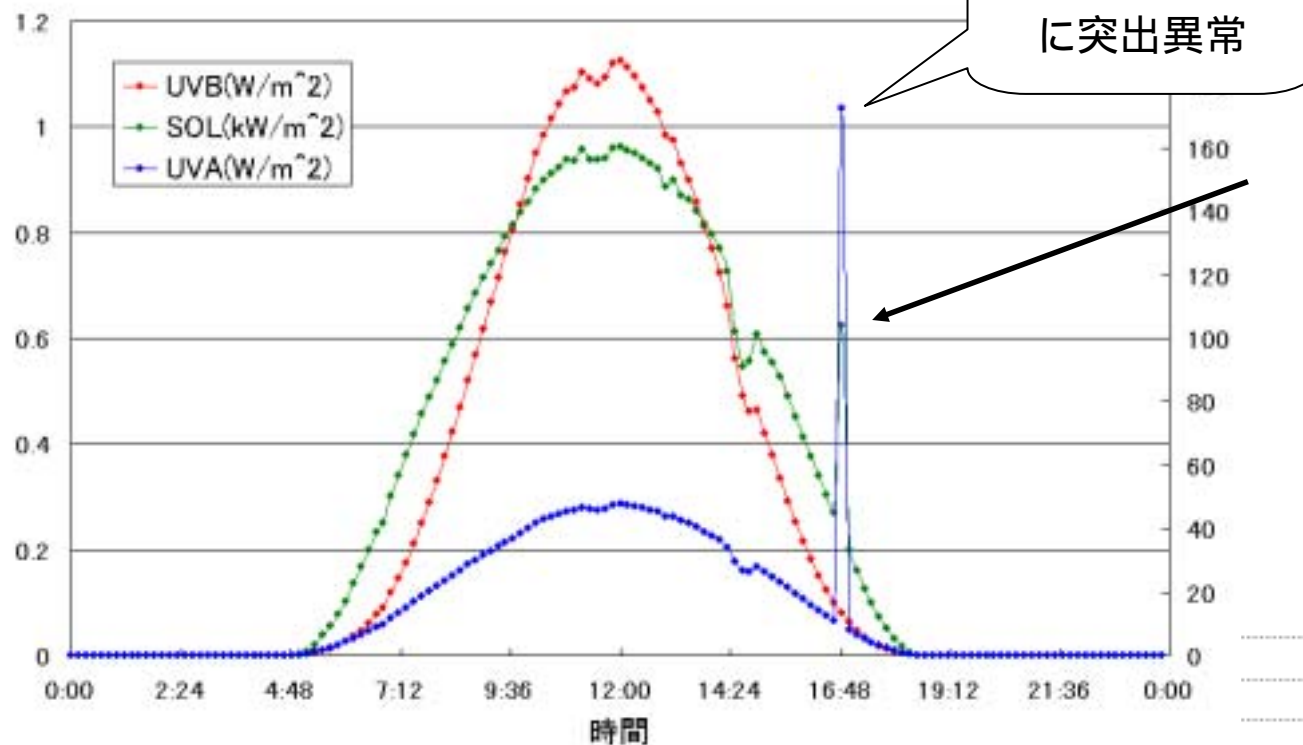
同様の傾向のある日: 5月21日、23日、26日、28日

6月2日 - 7日、16日、19日、22日

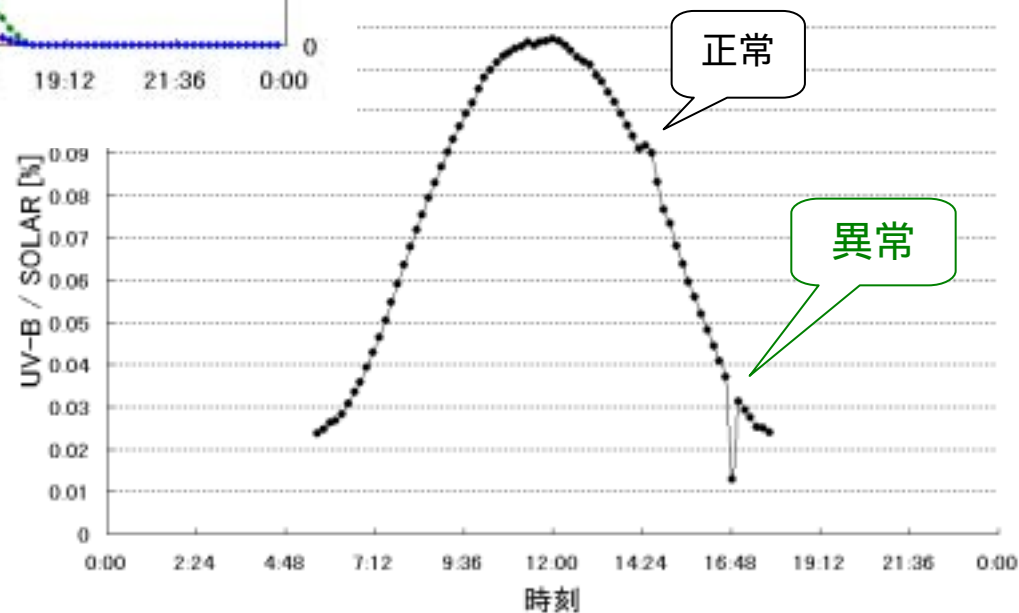
7月12日、20日 - 22日



特異データ分類 : 相関の問題 (CO)



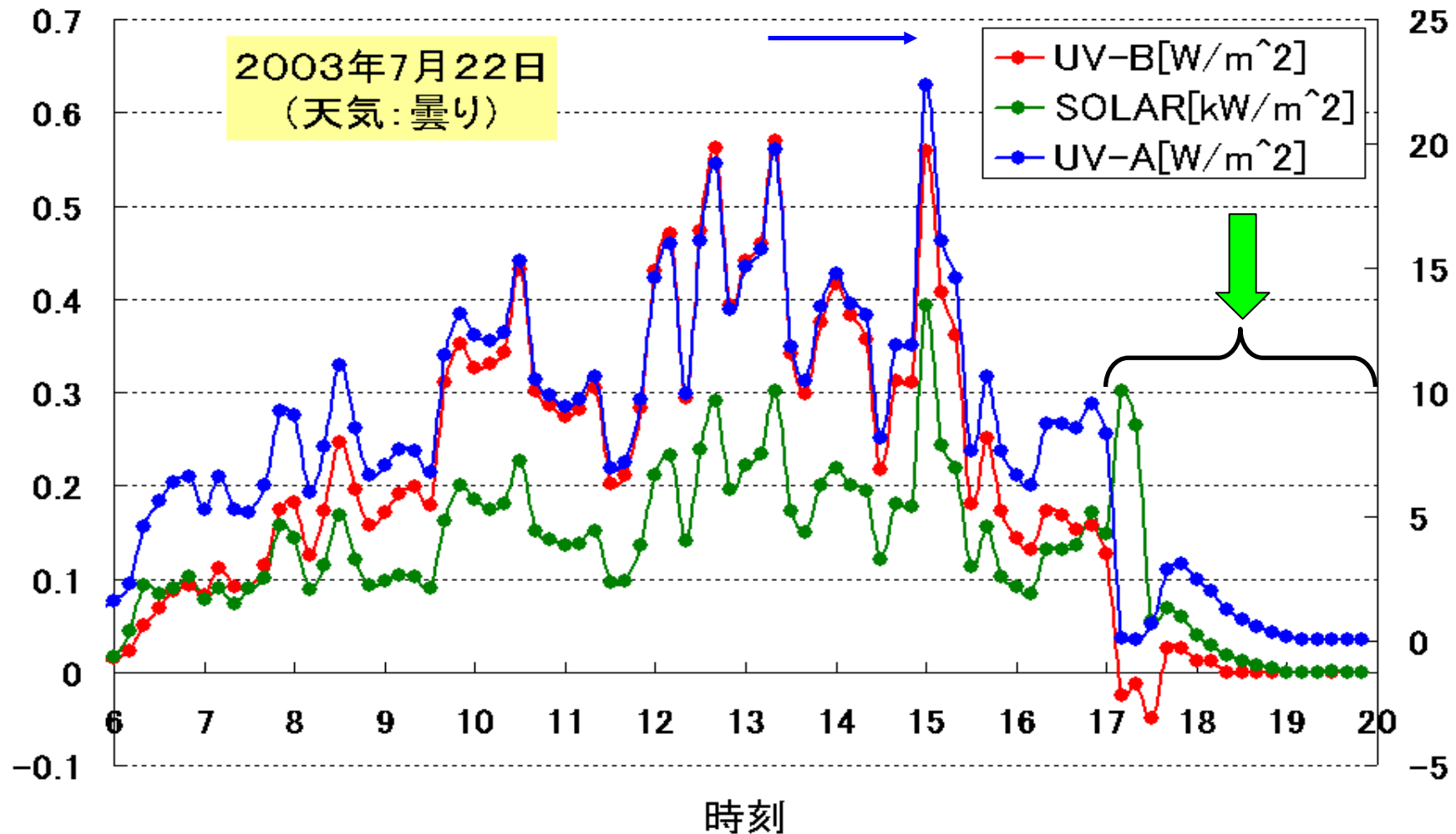
SOLARも突出しているが異常か？
 相対的な大きさの問題



特異データ分類 : 相関の問題 (CO)

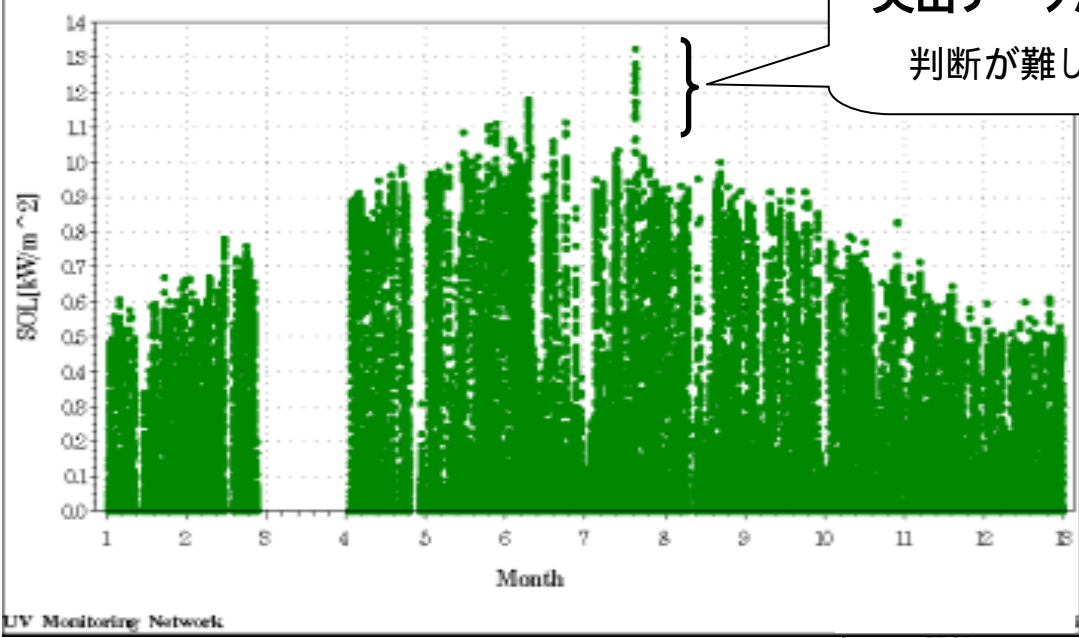
SOLAR、UV-A、UV-Bの位相に問題があるが、
原因の分っているデータ例

観測局からの報告 : 17:00に機器設置の調整
を行ったので以降欠測
データ扱い!



特異データ分類 : 相関の問題 (CO)

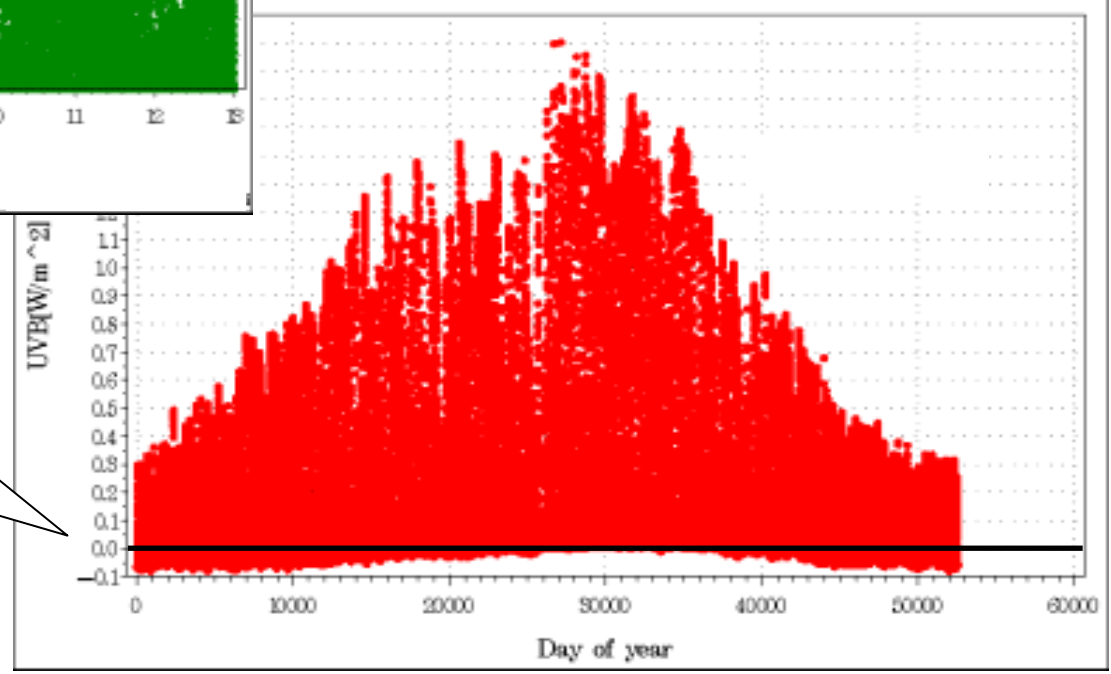
瞬時値の年間表示
2002年



突出データか?
判断が難しい

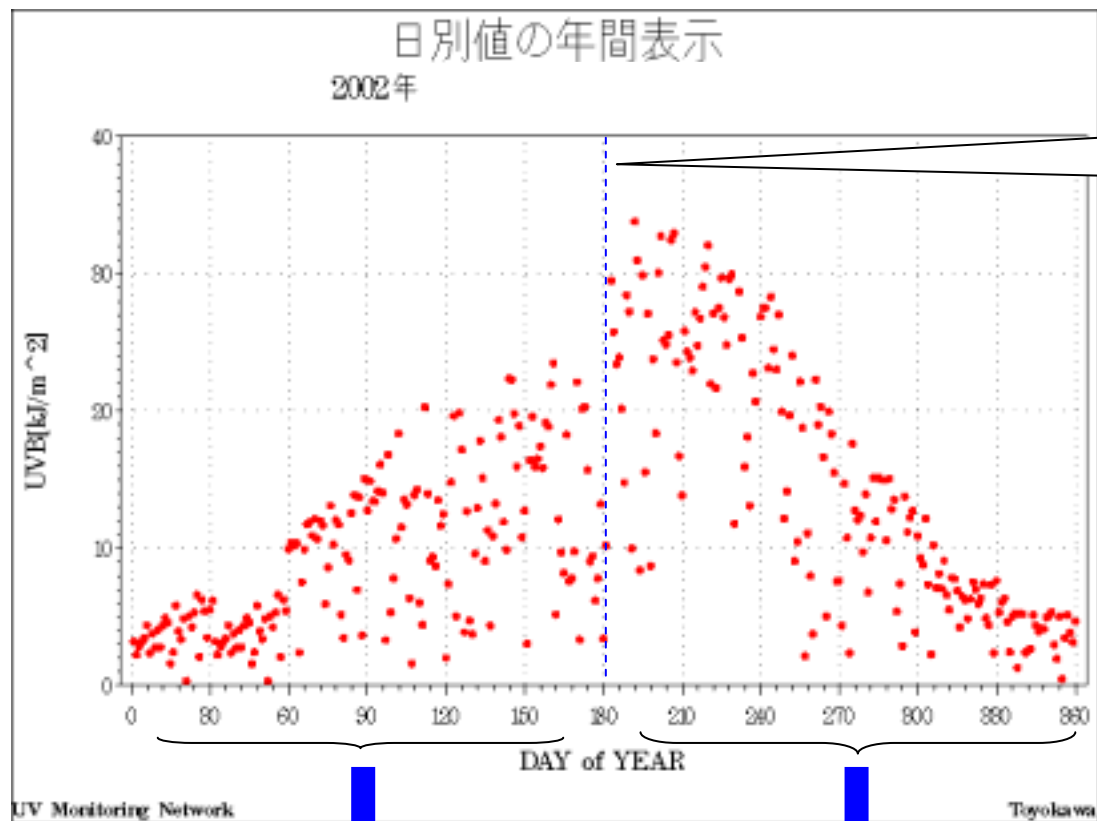
原因のはっきりしない特異データの
処理基準の作成を進める

年間表示
2002年



ベースレベルの揺らぎ
冬季、日中ピーク値の最大30%
もマイナスに振れてしまう観測例

特異データ分類 : 相関の問題 (CO)



7月から測定器が
交換されている！

~つくばの年平均値程度

~鹿児島県の年平均値程度

機器の校正結果待ち(再確認が必要)

放射強度の目安と典型的時系列パターン

UV - B 放射強度分布のピーク値の目安 : 0.1 ~ 2.5 [W / m²]

UV - A 放射強度分布のピーク値の目安 : 5 ~ 70 [W / m²]

SOL 放射強度分布のピーク値の目安 : 0.1 ~ 1.2 [kW / m²]



2000年からの累積データの平均値をまとめ、各地域毎の標準値算出を行う