

## 放電加工機の時系列解析（その1）

### 1.1 はじめに

放電加工機の時系列解析に取り組みます。

時系列解析は、以下の4ステップから構成されます。

#### ① ディスクリプション（description）

時系列の特徴を把握します。定常性の確認、成分分解等を行います。

#### ② モデリング（modeling）

時系列モデルを構成し、パラメータを推定します。

#### ③ 予測（prediction）

現在までの情報から今後の変動を予測します。

#### ④ 信号抽出

必要な信号や情報を取り出す。（異常検知とか）

今回の報告書は、上記の **ディスクリプション（description）** についてのまとめです。

### 1.2 どのようなデータを扱ったか

放電加工機は、加工中に障害が発生すると自己診断して3段階のレベルに分け、レベルに応じたメッセージを表示します。内容は以下です。

- ・エラーメッセージ：続行不可能な障害が発生したときに表示し、動作を中断する。
- ・ハルトメッセージ：再開可能な障害が発生したときに表示し、一時停止する。
- ・コメントメッセージ：続行可能な障害で発生し、注意を促す。

また、メッセージの記録は、USBによりCSVデータとして取り出すことが可能です。

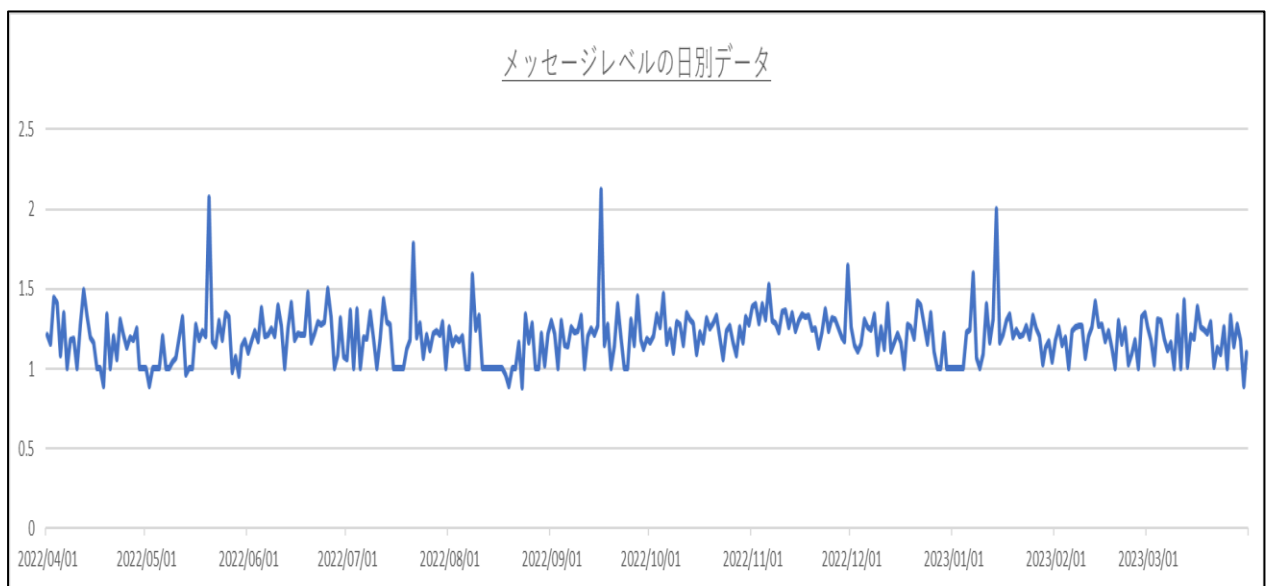
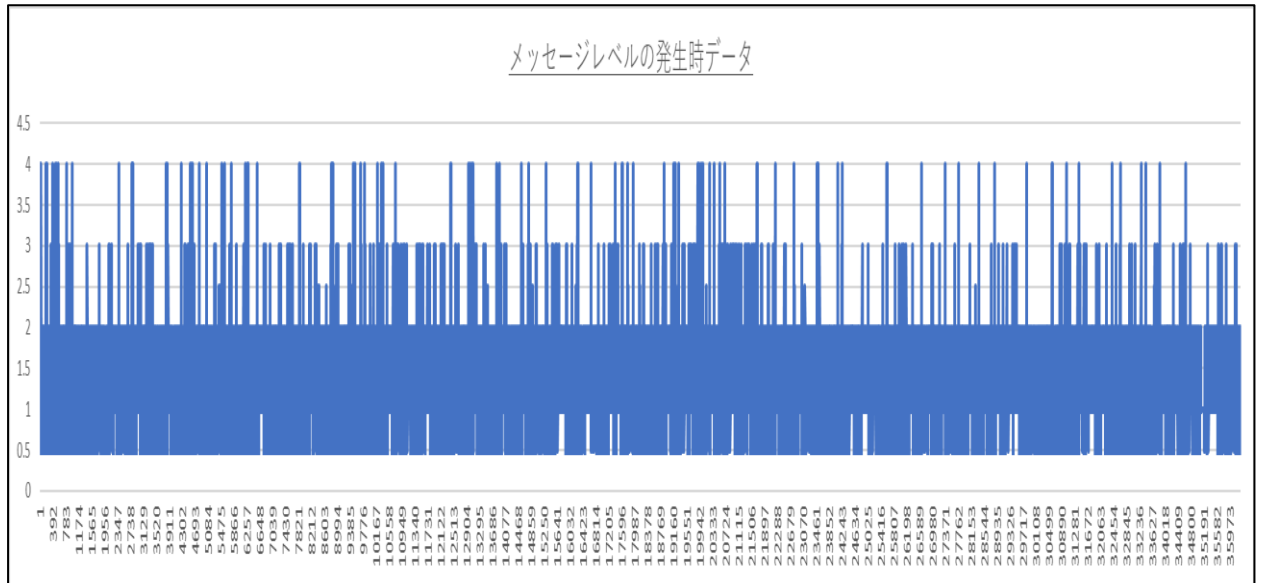
レベルに対応してデータ処理できるように、3段階のレベルについて障害のレベルが大きいと、点数は大きくなるように点数を設けました。

これらは、メッセージが発生した時の時間毎のデータで、点過程データと呼ばれるもので、時系列データではありません。故に、一日分の平均値を取って時系列データとして取り出すようしました。本データを以下に示します。2022年4月から2023年3月までのデータです。

上段が発生時毎のデータ（点過程データ）で、1年で37000回発生しています。

1点の発生は、一定時間の間隔ではありません。

下段は点過程データを、一日毎に平均して、平均したデータです。点の間隔は1日毎です、ので、時系列データと言えます。



### 1.3 プログラムについて

Python の statsmodels というライブラリー内の以下の関数で計算しています。

自己相関・偏自己相関コレグラム → acf, pacf

ADF 検定 → adfuller

成分分解 → seasonal\_decompose

詳細は以下で確認ください。

<https://www.statsmodels.org/stable/api.html#statsmodels-api>

本報告書にはプログラムは記載していません。

～参考文献、WEB～

「時系列解析入門」 北川源四郎 著 岩波書店

時系列データ分析 <https://qiita.com/tk-tatsuro/items/16ce74fc954b5a58df00>

Python で時系列解析・超入門（その1）

<https://www.salesanalytics.co.jp/datascience/datascience085/>

## 2 時系列データのディスクリプション (description) について

### 2.1 時系列データの分類

時系列データの種類は、以下の点から類別されます。

- ① 連続時間時系列 か 離散時間時系列(等間隔で観測されたデータ)

時系列データは、ほぼ 離散時間時系列です。

- ② 一変量時系列 か 多変量時系列

時点での採取データが、1 個か多数個かということです。

- ③ 弱定常時系列 か 非定常時系列

時系列データを、確率モデルで表す時に、データの水準やばらつき、自己相関の関係が  
時点によらず一定のもの、すなわち、時間で変化しない確率モデルのものを弱定常時系  
列と呼びます。「分析しやすいデータ」です。

非定常時系列データは、前処理で、弱定常時系列データに変換したりします。

(非定常のままの解析もあります)

弱定常時系列を数式で示すと

$$E(y_t) = \mu \quad \rightarrow \quad t \text{ 時点の期待値}$$

$$\text{Cov}(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k \quad \rightarrow \quad t \text{ 時と } t-k \text{ 時の相関 (自己共分散)}$$

となります。

- ④ ガウス型時系列と非ガウス型時系列

時系列の分布が正規分布 (ガウス分布) に従うものがガウス型時系列、そうでない場合  
が非ガウス型時系列となります。今回は、ガウス型時系列と仮定しています。

- ⑤ 線形時系列と非線形時系列

時系列モデルとして、線形モデルと非線形モデルに分類されます。

時系列データを入手したら、確認すべきは、③の定常性関連であり、

その後、時系列データを成分別に分解して、確認します。

順を追って、説明します。

## 2.2 定常性の確認

定常性の確認には、自己相関・偏自己相関コレグラムでの確認、ADF 検定による判定があります。まずは、データの「平均」と「分散」は以下でした。

平均	1.2007606312849317
分散	0.02551541556461745

### 2.2.1 自己相関・偏自己相関コレグラムでの確認

#### 自己相関(Autocorrelation)

始めに自己相関について説明します。2つの変数間の関係性を表す相関係数がありますが、自己相関(係数)は、1つの変数において、現在の状態と過去の状態の相関係数を計算します。lag(ラグ) = 0,1,2 とすると、0個ずらす、1個ずらす、2個ずらすという意味で相関をみます。

横軸が lag、縦軸が自己相関係数のこのグラフは、コレログラムと呼ばれています。

lag を 1 から 80 まで変えた時の元データとの相関関係をグラフ化したものです。網掛けの部分は 95%信頼区間を表しています。簡単に言えば網掛け内は相関係数が 0 である可能性が高いという事です。

グラフの見方ですが、lag=0 の自己相関係数が 1 なのは、元データ同士の相関係数のためです。(仮に、lag=12,24,36...と 12 の倍数で山が出来ていると、このデータは 12 の周期性がありそうと解釈します。)

#### 偏自己相関(Partial autocorrelation)

先ほどの自己相関は、例えば、現在と 2 時期前の自己相関を見たい時に、実は 1 時期前との関係性が含まれています。偏自己相関は、1 時期前との関係性を取り除き、現在と 2 時期前の直接的な関係を確認する事が出来ます。

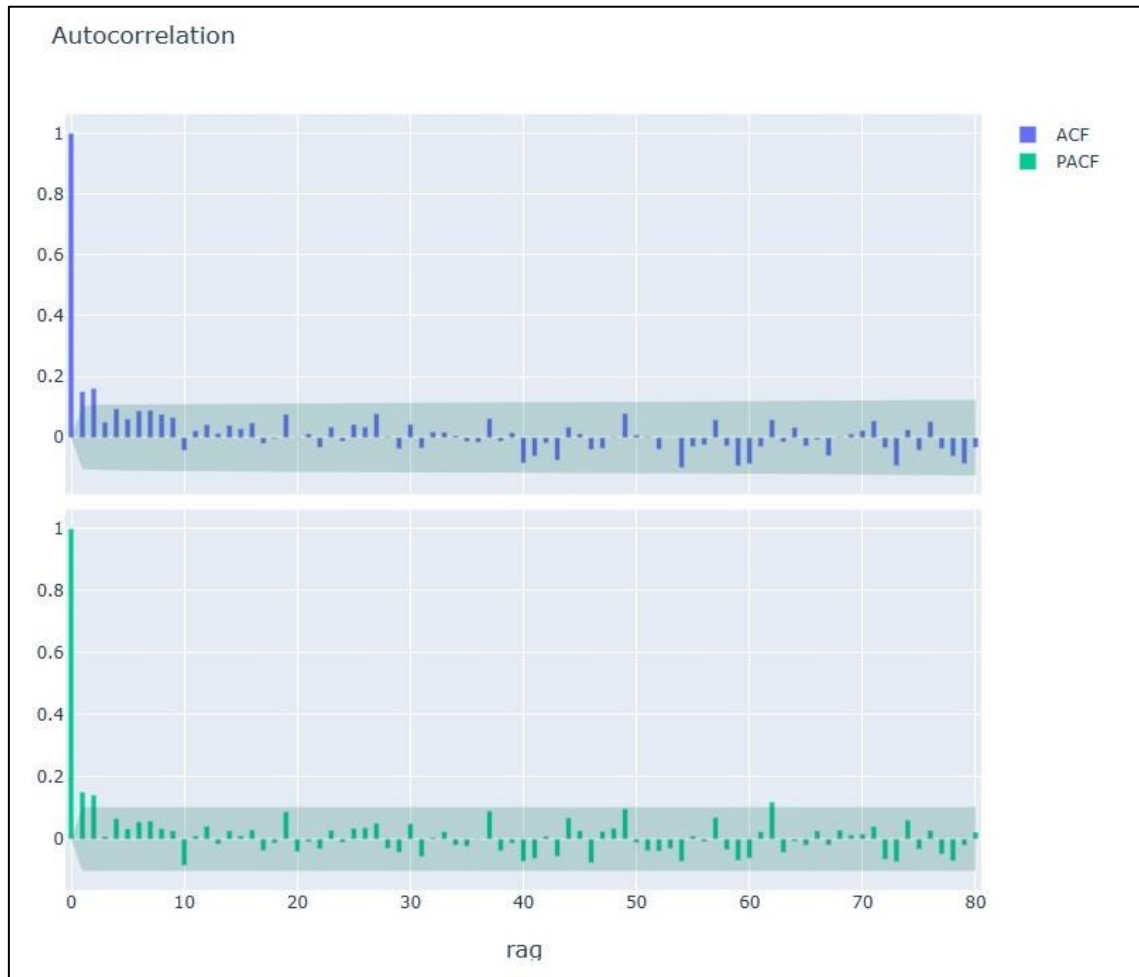
グラフは偏自己相関のコレログラムとなります。

lag=0,1 は自己相関と同じになります。lag=0 は自己相関と同じですが、lag=1 の場合は間に値がないので関係性を取り除く必要がないため自己相関と同じになります。

lag=2,3 を見ると自己相関のコレログラムと比べ小さい事が分かります。

これは、lag=1 や 2 の関係性が取り除かれている事が分かります。(仮に、lag=12 に相関があると、今回のデータではやはり 12 の周期のようだと解釈します。)

データでのグラフを提示します。



網掛け内にあるので、相関係数が 0 である可能性が高いです。

### 2.2.2 ADF 検定

ADF 検定 (augmented Dickey–Fuller test) は統計的仮説検定手法です。

p 値 (p-value) : 帰無仮説のもとで検定統計量はその値となる確率のことです。P 値が小さいほど、検定統計量はその値となることはあまり起こりえないことを意味します。

一般的に P 値が 5% または 1% 以下の場合に帰無仮説 (時系列は非定常である) を偽として棄却し、対立仮説 (時系列は定常である) を採択します。

ADF 統計量 (ADF Statistic) と棄却限界値 (critical value) の比較 : ADF 統計量 (ADF Statistic) が棄却限界値 (critical value) より小さいとき対立仮説 (時系列は定常である) を採択します。どの基準 (1% 有意、5% 有意、10% 有意) を使うかという問題があります。1% 有意が最も厳しく、10% 有意が最もゆるいです。伝統的に、5% 有意を使うケースが多いです。

データの結果は以下です。

ADF 統計量	-10.629988477502625
p 値	5.244540594121452e-19
棄却限界値	'1%': -3.448493650810824, '5%': -2.8695352280356556, '10%': -2.5710293341377715

ADF 統計量、p 値 から 定常性はあります。

### 3.時系列データの成分分解

ここからは時系列データの構成要素について見ていきます。

時系列データは以下の3つの成分から構成されています。

- ・トレンド成分

時系列の長期的傾向のことです。時間の経過とともに増加・減少する傾向とも言えます。

- ・季節成分（周期変動）

周期的に繰り返される変動のことです。

- ・残差成分（不規則変動、時間の経緯と関係ない）

上の二つ以外の成分です。

成分分解するための方法は、以下の3つの成分分解方法があります。

- ・移動平均法を利用した分解（加法モデルを仮定）

- ・移動平均法を利用した分解（乗法モデルを仮定）

- ・STL 分解（LOESS 平滑化を利用した分解）

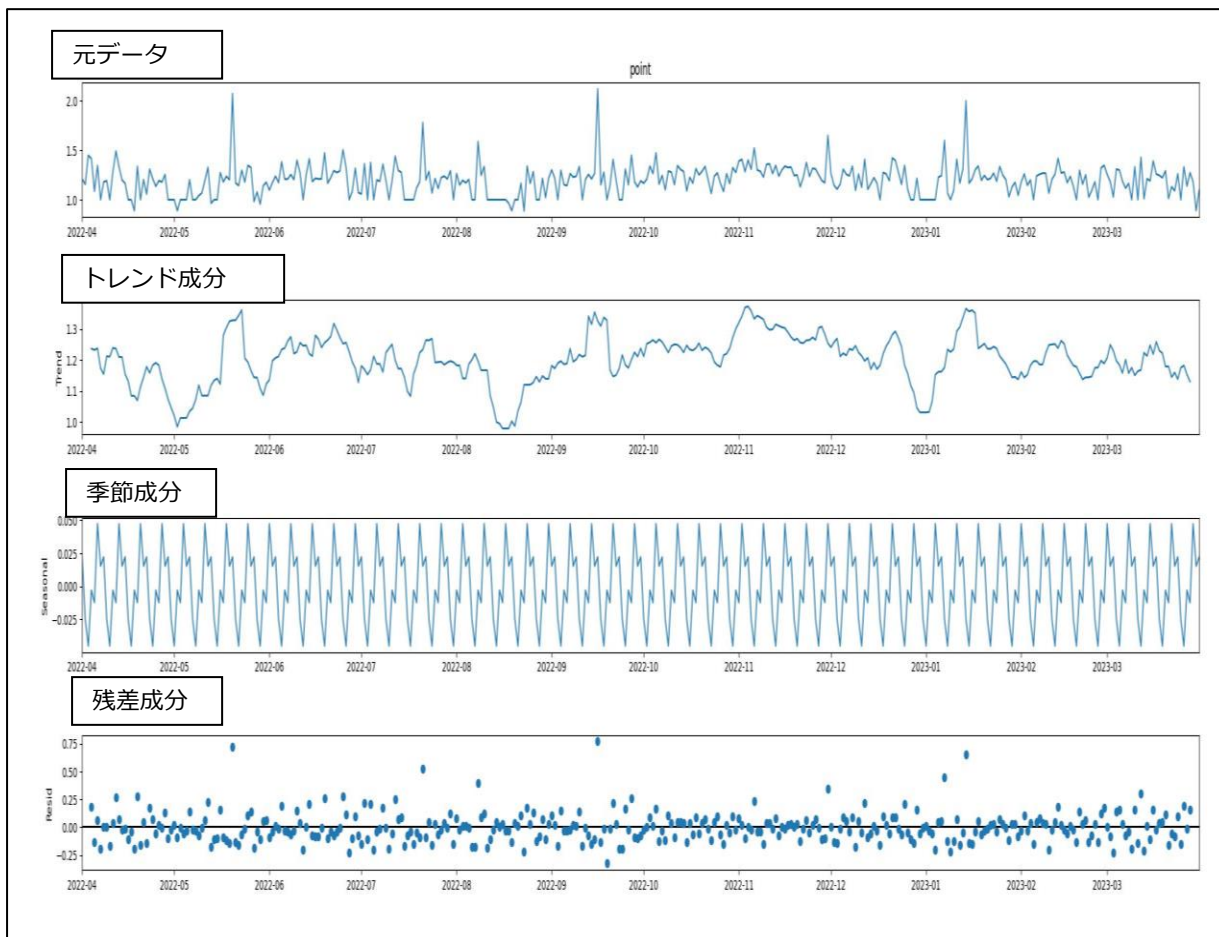
今回は、移動平均法を利用した分解（加法モデル）で行いました。

以下のステップですが、ライブラリーで行います。

- 1.設定した季節性の周期の移動平均を計算しトレンド成分とする
- 2.原系列からトレンド成分を除去し季節成分（平均値を0に調整）を求める
- 3.原系列からトレンド成分と季節成分を除去し残差を求める



データの結果は以下です。



#### 4. 次回のお取り組み

今回で、データの性格が見えてきたので、

モデリングに取り組む予定です。

## おまけ 時系列データの前処理

時系列データが定常性が無いと判断された場合、データの関係性を正しく解析ができない為、定常性を持たせるための変換を行う必要があります。各種の変換方法を述べます。

### 1.差分変換

差分変換とは、ある時点  $t$  のときの値とその直前の時点  $t-1$  のときの値の差分  $\delta t = y_t - y_{t-1}$  を計算し、その差を新しいデータとする手法です。1 時点離れたデータとの差を取る手法です。その結果を差分系列、または階差系列と言います。

### 2.対数変換

対数変換とは、変動が著しく大きな時系列データに対し、対数値変換を施すことで原系列の傾向を保持したまま、値を小さく変換する手法です。

変動の分散を一様にし、複雑な時系列でも変数変換で分析が簡単になることがあります。

対数変換された時系列データは対数系列と呼ばます。数式では  $\log(y_t)$  と表します。

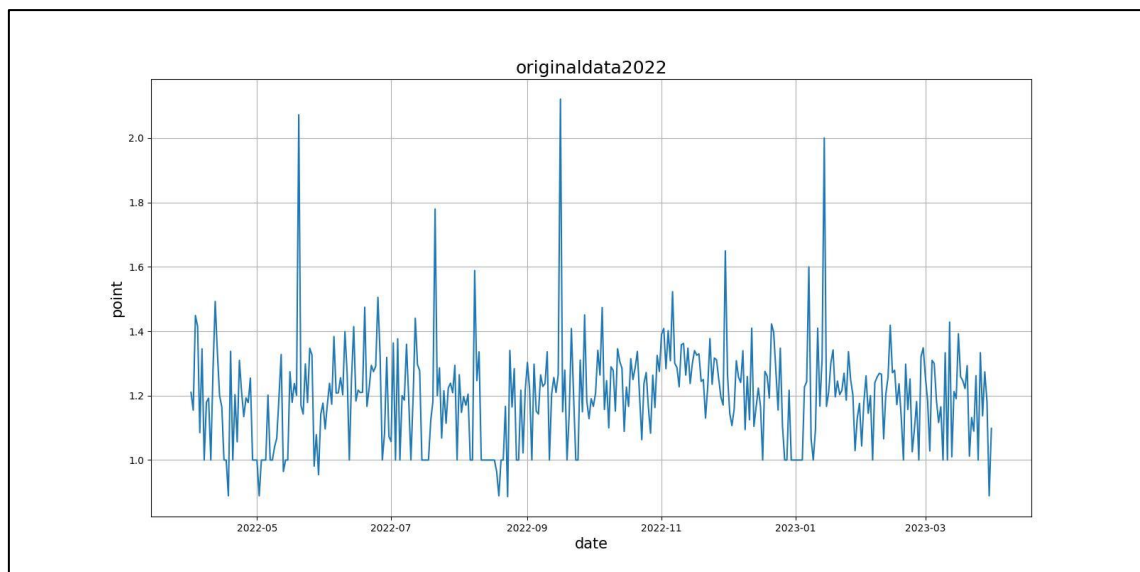
### 3.対数差分変換

対数差分変換とは、対数変換と差分変換の両方の変換を施す手法です。

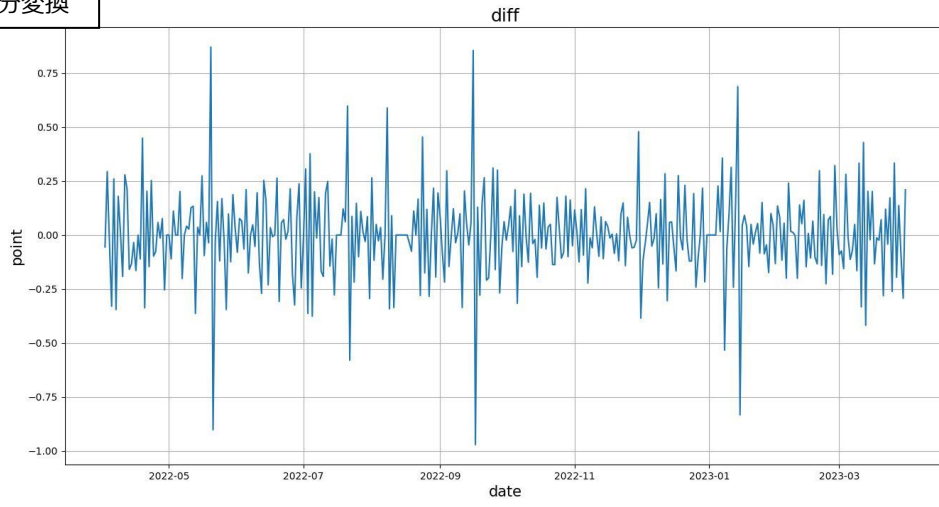
時系列データが変化率や成長率などである場合、対数差分  $\Delta \log(y_t) = \log(y_t / y_{t-1})$

を計算して解析することがあります。

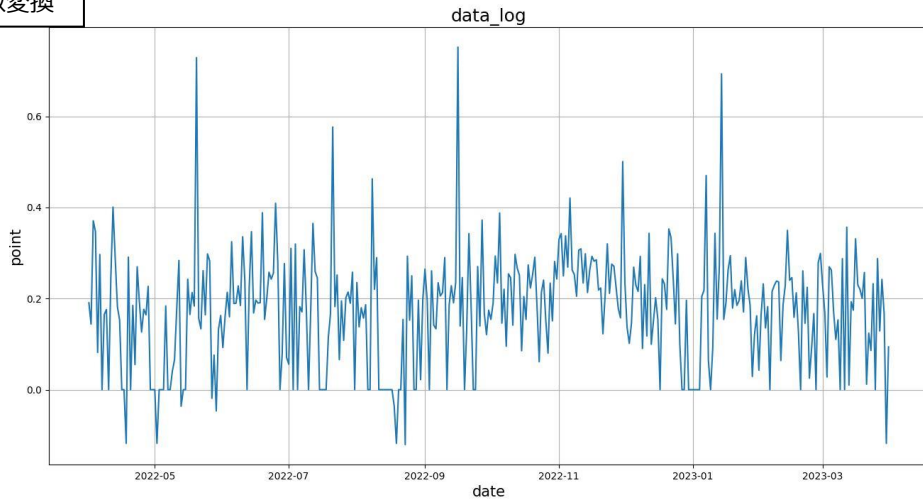
今回は、元データの状態で、定常性は確保されたので、変換の必要ではないのですが、グラフにしてみました。



差分变换



对数变换



对数差分变换

