

重心動揺データの平滑化について - その1

指数平滑、移動平均、重み付き移動平均

慶応義塾大学理工学部
竹内寿一郎

1. はじめに

近年、重心動揺をカオス現象として捉え、いろいろなかたちで解析が行われてきたが、特に埋め込みによるアトラクタのフラクタル次元を求める研究が盛んになってきた。埋め込みを行うにあたって、その遅れ時間を幾つにすべきかという問題は次元の大きさに大きく関係するので無視することはできない。たとえば遅れ時間を極端に短くとると、埋め込み時の次元間に正の相関が強くなり、アトラクタの形は平たくなり次元が小さくなることが予想される。最近の研究では遅れ時間を20点乃至は30点をとる必要があるという文献もあり、従来行われてきた20Hz、60秒間で観測された1200点ではかなりの情報損失となりかねない。そこで我々は1000Hzで60秒間で観測された60000点のデータを用いてサンプリングの影響について調べることにした。その第一ステップとして60000個のデータの平滑化について検討することにした。

右図は1000Hzデータのはじめの100点で、明らかにサンプリングノイズがあることが分かる。このような場合、高周波フィルターを通すのが一般的であるが、ここでは指数平滑、単純移動平均、重み付き移動平均、フーリエ逆変換、自己回帰などの手法を用いて細かい波をろ過することを考える。そしてそれぞれの手法の特徴をふまえた上で次ステップのサンプリングの影響の研究につなげたい。

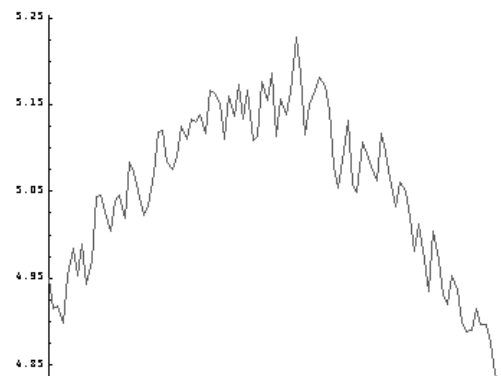


図1. 1000Hzデータのはじめの100点

2. 指数平滑法

最も単純でパラメータが少なく済む方法は指数平滑平均である。 X_t を原系列、 F_t を平滑化された系列とすると、

$$F_{t+1} = \alpha * X_t + (1 - \alpha) * F_t$$

で表される。ここで α は平滑化パラメータで α が大きいほど原系列に近く、0に近いほど滑らかに(上下に動く系列の内側に)なる傾向がある。

下記の動詞 expsm は J ではもっと簡単にかけるはずだが、一応定義通りに計算させてみた。^: を用いると1行で表現できるのではないかと思う。是非挑戦してみよう。

```

expsm=:4 : 0
sr=:Fr:={.y[i=.0[y=.({.@),]y.
label_L1.
if.(#y.)<:i.>:i do. goto_OWARI. end.
Fr=:((1-x.)*Fr)+x.*i{y.
sr=:sr,Fr
goto_L1.
label_OWARI.

```

```

)
data=:+/\(-+/%#)?20#10
data
0.05 _0.9 _5.85 _3.8 _4.75 _3.7 _2.65 _6.6 _2.55 0.5 3.55 3.6 _0.35 _1.3
2.75 _0.2 _1.15 _3.1 _3.05 0
plot >(];0.7&expsm;0.3&expsm)data
    
```

下の図2はパラメータ α の違いによって平滑化された折れ線がどのように変化するか、乱数データを用いて調べたもので、図3は実際に重心動揺データに適用し、はじめの120点について図示したものである。

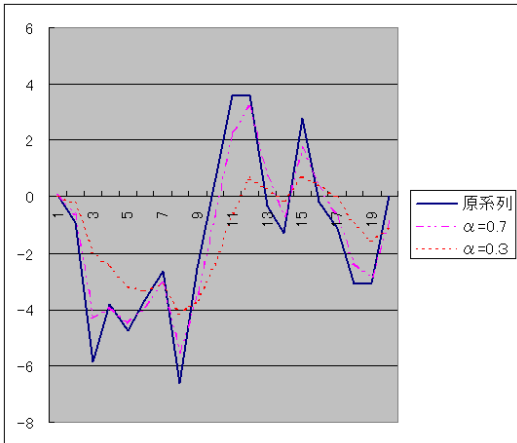


図2 . 乱数による指数平滑法
($\alpha = 0.3, 0.7$)

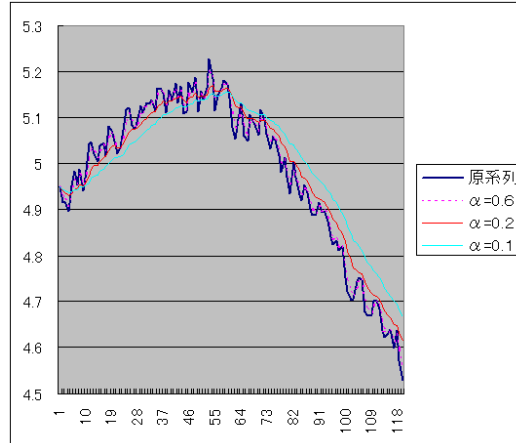


図3 . 1000Hz データの指数平滑法
先頭から120点 ($\alpha = 0.1, 0.2, 0.6$)

図2から α が0.7でもかなり原系列を追っていることが分かる。 α が0.3でも大きな変化にはよく追従している。図3から α が0.5より大きいと、かなり原系列に引っ張られた平滑化曲線になることが予想される。 α が小さくなるにつれて滑らかにはなるが、原系列に対して動きが鈍く、系列が増加のときは下めに、減少のときは上めになることが見てとれる。

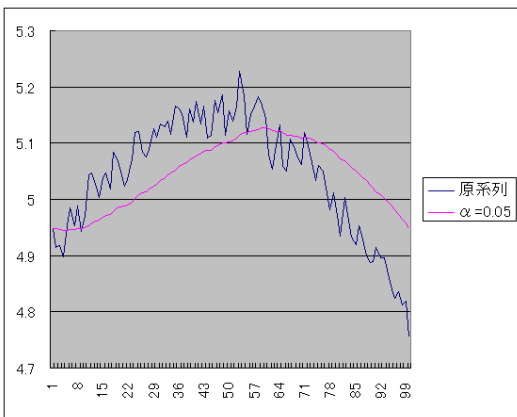


図4 . 1000Hz データの指数平滑法
先頭から100点 ($\alpha = 0.05$)

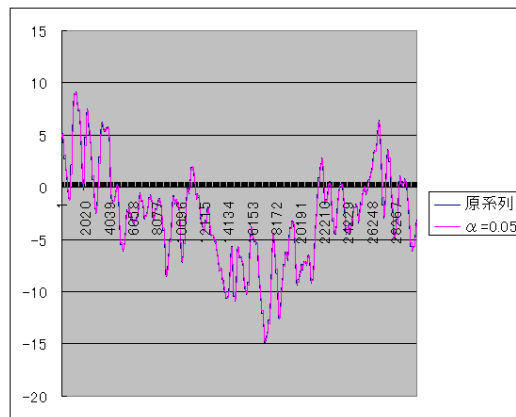


図5 . 1000Hz データの指数平滑法
先頭から30000点 ($\alpha = 0.05$)

図4で α が0.05では原系列から随分ずれているように思えるが、図5のように実際全データをプロットしてみると、見た感じでは、殆ど原系列と差がないくらい追従している。

3 . n項移動平均

n 項移動平均は簡単な手法であり、Jで表現すると下のmaveのように非常に短い関数として書くことができる。

```
mave=:/\%[
mmave=: [mave((<. @-:@[])#{. @[]), ], (<. @-:@[])#{: @]
```

ここで、mmave は両端に両端の同じデータを付加して短くならないようにした関数である。

5 mave data

_3.05 _3.8 _4.15 _4.3 _4.05 _3 _1.55 _0.3 0.95 1.2 1.65 0.9 _0.05 _0.6 _0.95 _1.5

5 mmave data

_1.32 _2.09 _3.05 _3.8 _4.15 _4.3 _4.05 _3 _1.55 _0.3 0.95 1.2 1.65 0.9 _0.05 _0.6
_0.95 _1.5 _1.46 _1.23

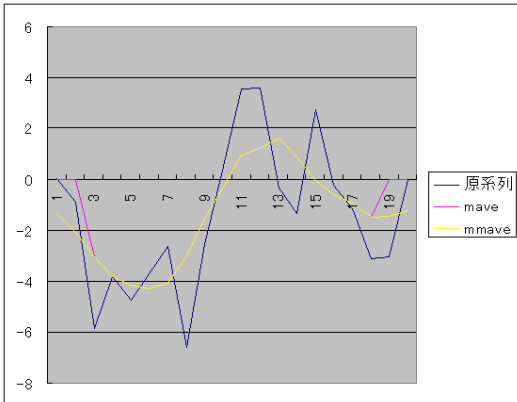


図6．ランダムデータの移動平均
(5 項)

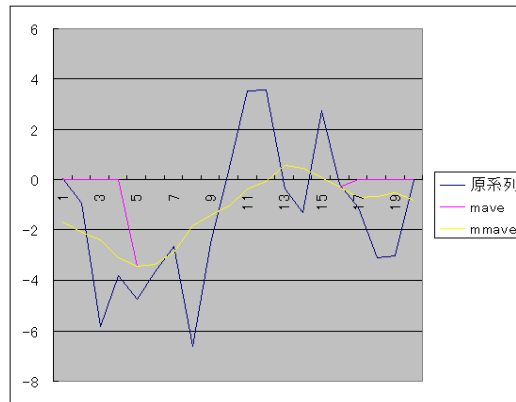


図7．ランダムデータの移動平均
(9 項)

mave と mmave の違いは前後の (移動平均個数-1) 個だけで、真ん中の部分は理論的にも完全に一致する。mmave による補外は、まあまあといったところではないかと思われる。

移動平均を重心動揺データに適応したのが図8～図10である。

これらの図は mave を使用した場合両端が欠けるので、比較のために15項移動平均のときの欠測値の数両端7項ずつを考慮して、8番目のデータから描画してある。

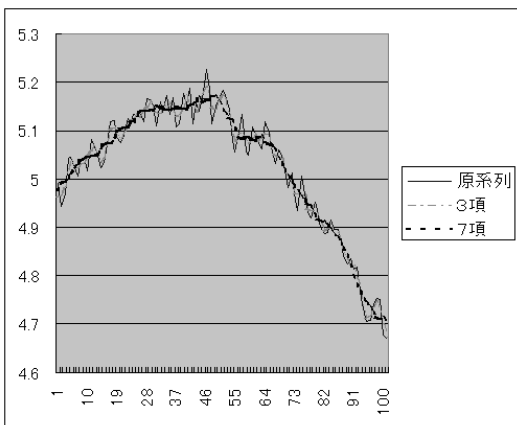


図8．1000Hz データの移動平均、
8番目から100点(3項、7項)

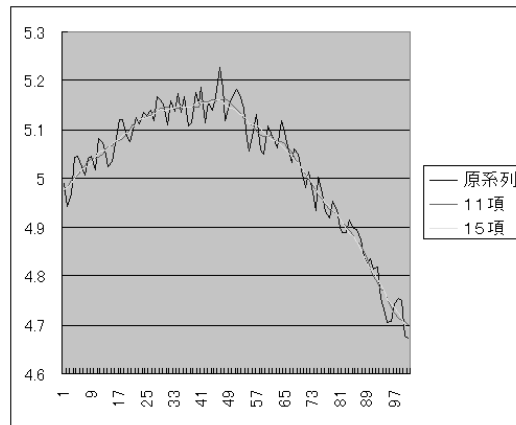


図9．1000Hz データの移動平均、
8番目から100点(11項、15項)

図8～図9は折れ線が重ならないように別図にして描いたものであるが、3項では殆ど原系列そのもの、7項になるとようやく移動平均らしい曲線が得られる。11～15項になるとこれらの曲線は図からは殆ど違いが分からない。

図10は11個のデータを拡大して見た図で、移動平均の項数が少ないと、原系列を追っている様子が見てとれる。項数が増えるに従って原系列が上ならば下、下ならば上、と言う具合になるべく内側になるように、言い換えれば保守的 (Conservative) に原系列を追っている。

比較のために、後で述べるスペンサーの移動平均を並べて描いておいたが、これは15項のわりには原系列の影響を受け、なめらかな曲線?でよく追っていることが分かる。

図11は全データを15項の移動平均をとったものを、先頭から30000点描いたものであるが、巨視的に見るとどんな手法でも結構良く追っているように見えてしまう。が、実際は拡大して見るとかなり違うのである。

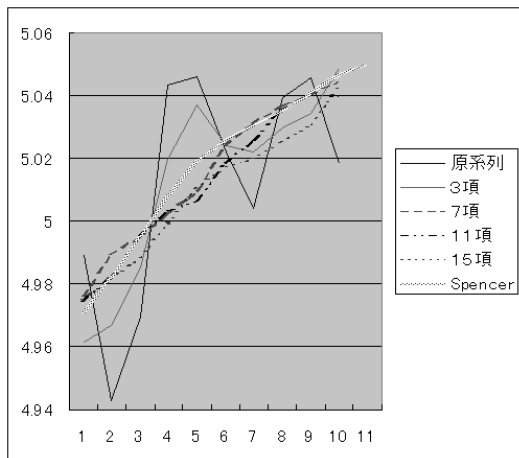


図10 . 1000Hz データの移動平均の比較
8番目~18番目迄を拡大

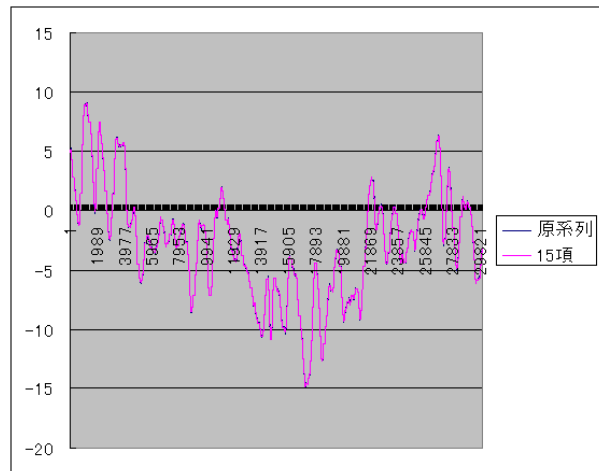


図11 . 1000Hz データの移動平均
はじめから30000点(15項)

4 . 重み付き移動平均

4 . 1 . スペンサーの重み付き移動平均

経済時系列でよく用いられているトレンドを求める方法として、スペンサーの移動平均法がある。

局所的には3次式が当てはめられた、対称な15点からなる重みをもつ移動平均法である。

重み : (-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, -3)/320

志村氏によるスペンサーの移動平均動詞は次の3行からなっている。

```
wts=(|.,}.)74 67 46 21 3 _5 _6 _3%320
locspen=(+/.*)&wts
spencer=:15"_ locspen\{7:#{.},],7:#{:
```

wts は15項のウェイトが対称だから8項の入力で済ませようと工夫したもの。

locspen は各項とウェイトとの内積の動詞を定義しておく。

そして、spencerで一気に計算する。その原理は2<\dataを試してみるとよく分かる。

これはdataのはじめから2項ずつ、1つずつずらしながらボックス(<)してゆく。

また、15"_は定数15を作り出す動詞で、全体をフォークにしている。

なお、後半の(7:#{.},],7:#{:は前後に先頭、および末尾の項を7個付け加える動詞である。

```
plot >(;spencer;7&mmave)(50{x)
plot >(;spencer;15&mmave)(50{x)
```

で図12と図13が描ける。

図12と図13はほとんど変わらないように見えるが、細かくみて見ると7項移動平均では激しく変化している原系列を追っていることがわかる。15項移動平均では全体的に滑らかにゆったりと変化している。7項、15項ともに移動平均の特徴は、激しく変化するところでは完全には滑らかになっていないことである。それに比べスペンサーの移動平均は15項移動平均より滑らかで、かつ、7項移動平均くらい原系列を追っていることが分かる。

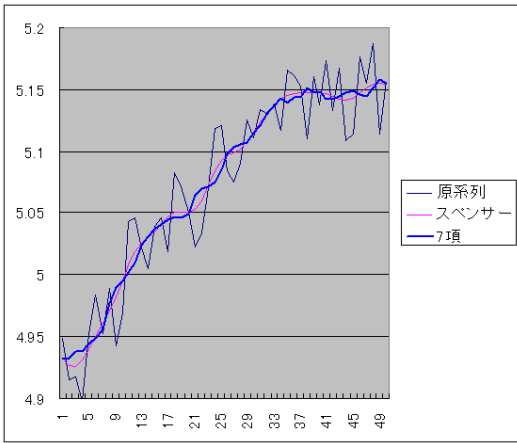


図 12 . スペンサーと移動平均の比較
先頭から 50 点 (7 項)

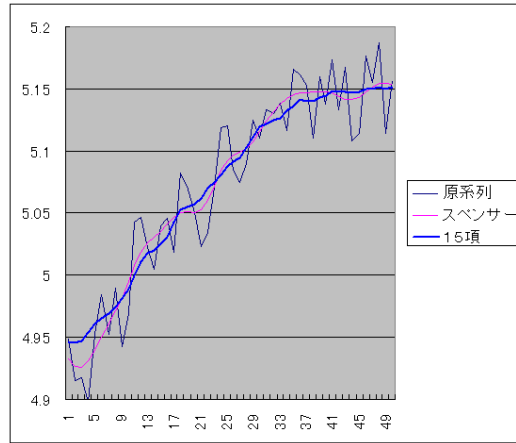


図 13 . スペンサーと移動平均の比較
先頭から 50 点 (15 項)

4 . 2 . ヘンダーソンの重み付き移動平均

季節性が強い経済時系列を解析する手法として、季節変動調整法がいくつか知られている。最も有名なのがセンサス局法 X-11、最近のバージョンは通称 X-12 と呼ばれていて、基本的には全変動を、トレンドサイクル (TC)、季節変動 (S)、不規則変動 (I) の 3 つの成分に分ける手法である。このとき TC を求めるのに使用されるのがヘンダーソンの重みつき移動平均法なのである。

n を奇数として n 項ヘンダーソンの移動平均は次の式による。

$$H^{(n)}(B) = \sum_{i=-(n-1)/2}^{(n-1)/2} h_i^{(n)} B^i, \quad i = 0, \pm 1, \pm 2, \dots, \pm(n-1)/2$$

ここで、 B (Back) は時間遅れのパラメータで負のときは F (Forward) とも書く。

また、 $h_i^{(n)}$ は重み係数で、 $m = (n+3)/2$ として、

$$h_i^{(n)} = \frac{315\{(m-1)^2 - i^2\}(m^2 - i^2)\{(m+1)^2 - i^2\}\{(3m^2 - 16) - 11i^2\}}{8m(m^2 - 1)(4m^2 - 1)(4m^2 - 9)(4m^2 - 25)}$$

である。

従って 5 項ヘンダーソンの移動平均アルゴリズムは

$$H^{(5)}(B) = -\frac{21}{286}B^2 + \frac{84}{286}B + \frac{160}{286} + \frac{84}{286}F - \frac{21}{286}F^2$$

7 項では

$$H^{(7)}(B) = -\frac{42}{715}B^3 + \frac{42}{715}B^2 + \frac{210}{715}B + \frac{295}{715} + \frac{210}{715}F + \frac{42}{715}F^2 - \frac{42}{715}F^3$$

である。

J の関数は志村氏にならって作成したが、残念ながら時間の関係でタシットにより動詞を作成することができなかった。

まず、ヘンダーソンの重み係数を求める動詞 `hpnt` を定義する。

```
hpnt=:4 : 0
m=: -:x.+3
nm=:315*(((m-1)^2)-*:y.)*((m^2)-*:y.)*(((m+1)^2)-*:y.)*(((3**m)-16)-11**y.)
hi=:nm%8*m*((m^2)-1)*((4**m)-1)*((4**m)-9)*((4**m)-25)
)
```

移動平均なので両端が欠ける。その両端を端点を繰り返すことによって補項している。

補項は ((.<.@-:x.)# {.y.), y. , (<.@-:x.)# {.y. で実現させている。

```
henderson=:4 : 0
wt=:x.hpnt i=: -:<:x.
```

```
x.(+/.*)&wt\((<.@-:x.)#\{y.},y.,(<.@-:x.)#\{y.
)
```

項数による違いを見るには、

```
plot >(;7&henderson;15&henderson)120{x
```

データの中程の状態を見るには次を実行する。

```
plot (100+i.20){"1 >(;7&henderson;15&henderson)120{x
```

ヘンダーソンの移動平均の項数による違いを表したのが図 14～図 15 である。

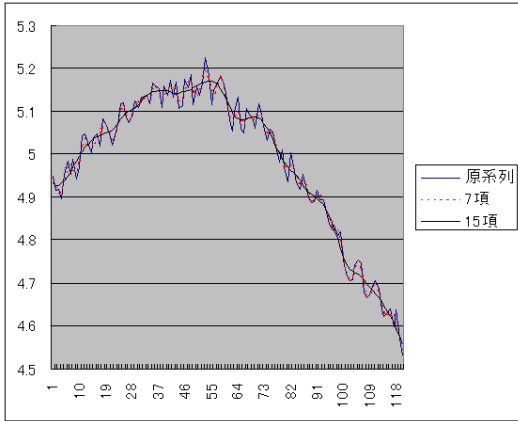


図 14 . ヘンダーソンの移動平均
先頭から 120 点 (7 項、15 項)

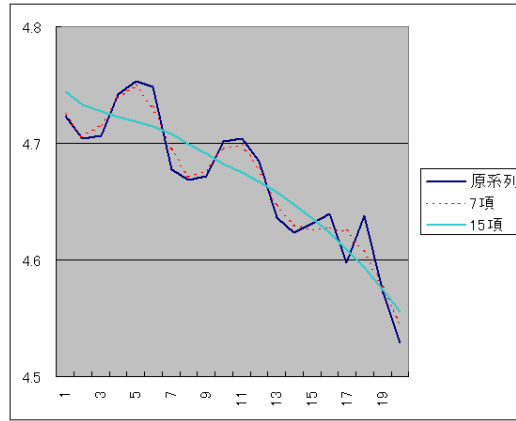


図 15 . ヘンダーソンの移動平均
99 番目から 119 番目 (7 項、15 項)

やはり 7 項を用いると上下のノイズを多少滑らかでは追ってしまっている。重心動揺の 60000 点のデータでは、単純な移動平均もヘンダーソンの移動平均も 15 項ぐらいとるのが妥当ではないかと思われる。ただ、単純な移動平均では急激な変化にはついて行けないでなだらかな曲線となるが、ヘンダーソンの 15 項移動平均では細かい激しい変化に対し、その変化について行きなが、しかも滑らかな曲線となっている。

このことは図 10 のところで述べたように、スペンサーの移動平均と同じような性質をもつものと考えられる。そこで図 16 にスペンサーとヘンダーソンの 15 項の移動平均の比較図を掲げた。

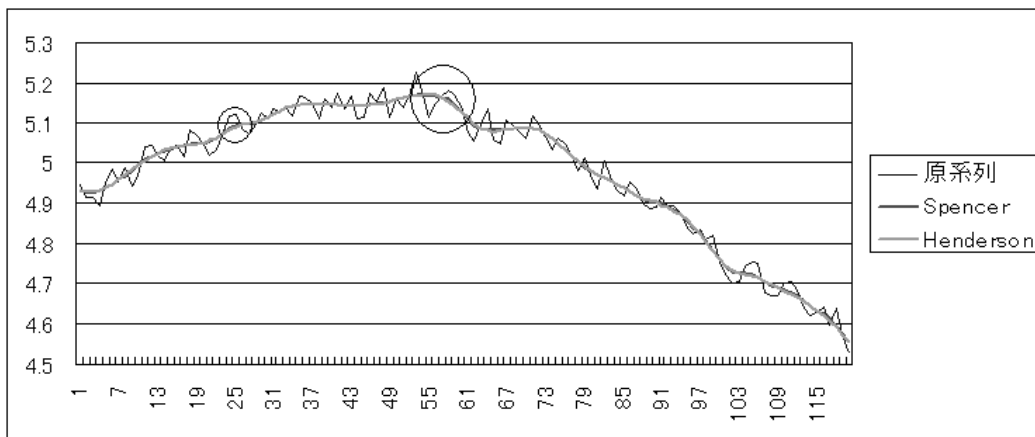


図 16 . スペンサーとヘンダーソンの移動平均の比較 (15 項)

全体として双方とも滑らかな曲線で原系列をよく追っているのが見てとれる。図 16 を注意深くみれば、左の小丸の中ではスペンサーがやや上に、真中の丸の中ではスペンサーが下になったり上になったりして、必ずしも完全には一致していないことが分かる。