新型コロナウイルスの時系列解析

（4）第５波の統計モデル

nino

2023 年 12 月 15 日

１．はじめに

前々報では、第５波は正規分布に似た大きな波にほぼ７日周期の短周期成分が加わる形状を示すことが明らかとなった。そこで、この特性を統計モデルに当てはめて検証するとともに、モデルによる感染状況等の把握を試みた。

２．時系列データ

使用する時系列データは前々報[1)]と同じ厚生労働省のオープンデータ（以降、報告値という）の次の５項目（略称）である[2)]。

①PCR 検査実施人数（検査）、②新規陽性者数（陽性）、③入院治療等を要する者（入院）、④重症者数（重症）、⑤死亡者数（死亡）、　以降は、略称を用いる。

対象期間は、第５波に相当する 2021/7/11(日)～10/23(土)の 105 日間(15 週間)とした。

３．第５波の特徴

まず、準備として、第５波の時系列図を再掲し、その特徴の概要をまとめた[1)]。

図１に第５波の報告値の時系列を示した。縦軸は人数、横軸の日付は１週間間隔である。

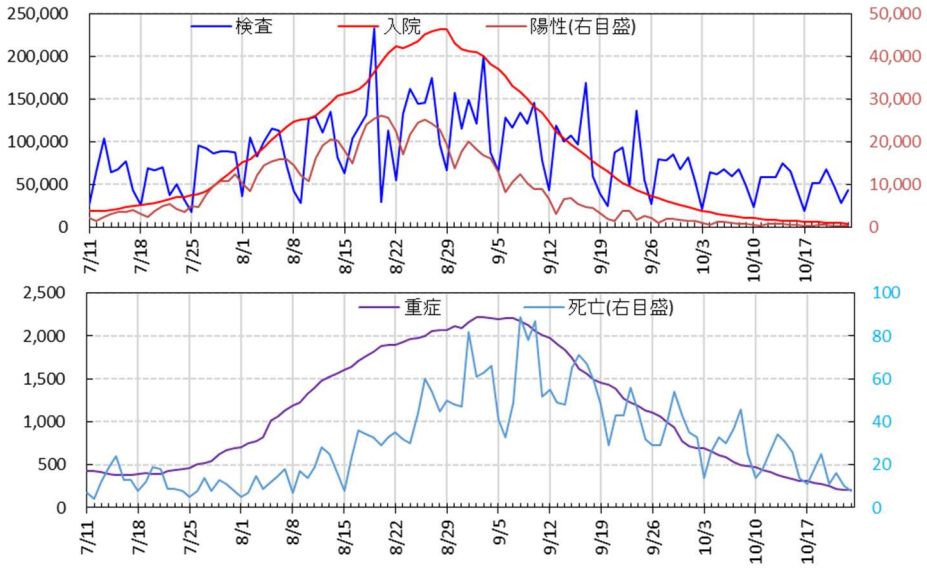

図１　第５波の報告値の時系列

検査、陽性および死亡は大きな波にほぼ７日周期の短周期成分などが加わる形状をしているが、入院と重症では７日周期成分はほとんど見られなかった。

大きな波と７日周期成分に分けて調べるため、図２に７日間移動平均および報告値と７日間移動平均との差（残差）の時系列を示した。

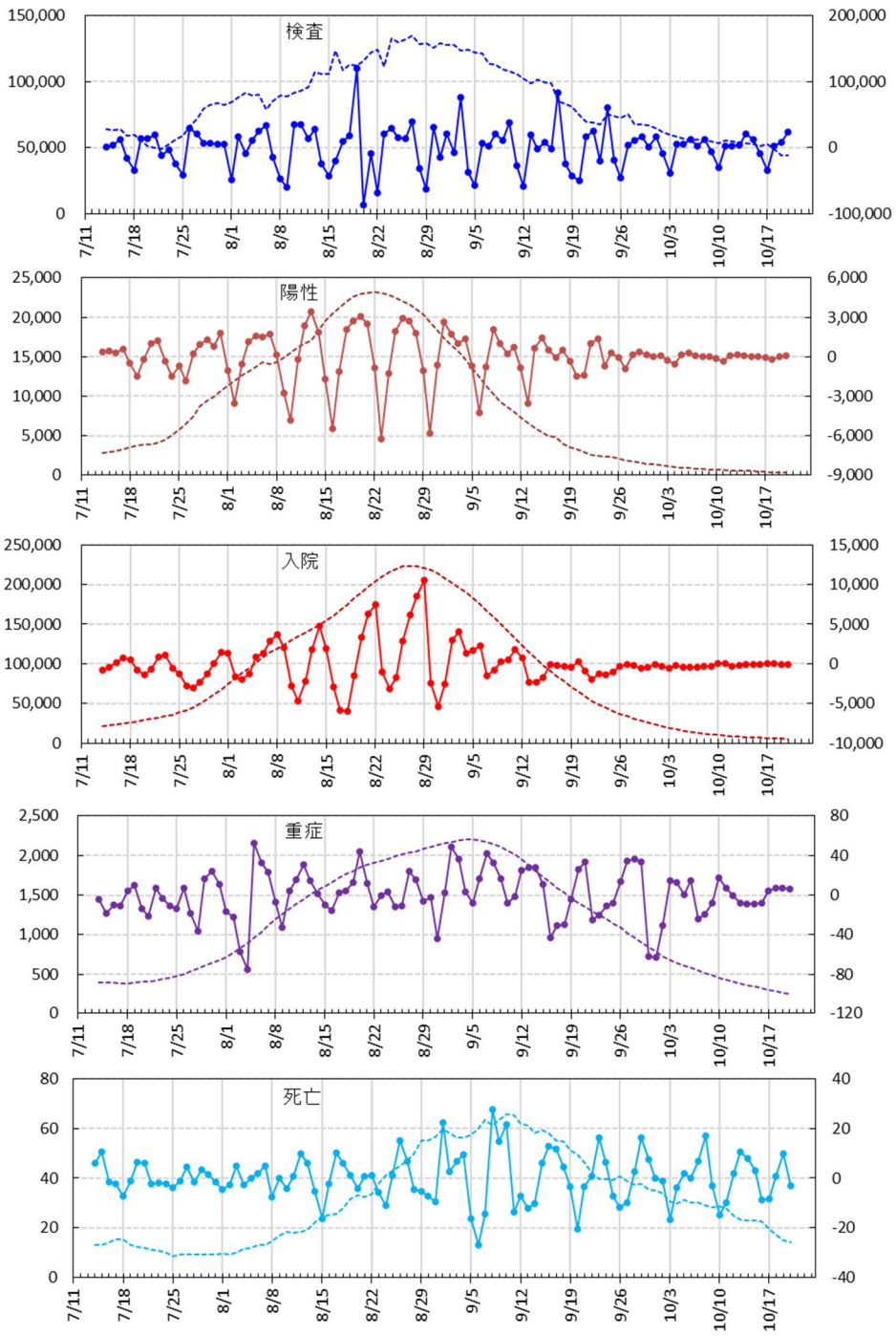

図２　７日間移動平均(点線)と残差(実線：右目盛)の時系列

７日間移動平均については、陽性、入院および重症は正規分布に似た形状を示し、８月上旬には小さな肩ピークが認められた。検査と死亡はそれにノイズが少し加わる波形であった。また、正規分布は一定の人数（バックグラウンド）の上に乗っていた。
残差はほぼ７日周期でかつ縦軸のゼロレベルを中心として変動しており、その振幅は７日間移動平均に比例して増減する傾向が見られた。陽性と入院については、残差の周期特性が前期と後期とでやや異なる形状を示した。

４．統計モデル

4.1　統計モデルの式

統計モデルとして、正規分布に曜日効果が加わる簡易なモデルを当てはめた。曜日効果は、曜日変動の振幅が正規分布に比例して変化するものとした。通常、曜日効果は時間によらず一定と仮定する「固定曜日効果」を用いる場合が多いが、このモデルの曜日効果は時間の経過とともに変化する「時変曜日効果」に相当する[3)]。

モデルでは、報告値 $O(t)$を次式に示す４つの成分に分けた。

$$O(t) = G(t) + b + D(t) + I(\mathrm{t}) \tag{1}$$

ここで、$G(t)$、b、$D(t)$、および $I(\mathrm{t})$は、それぞれ時点 t（t=0,1,2,⋯,104 日）おける正規分布、バックグラウンド（定数）、曜日効果、および不規則変動を表す。全項目とも同じ式である。なお、$G(t) + b$ が図２の７日間移動平均に該当する。

図３に期間を 105 日間（15 週間）としたモデルと４成分の模式図を示した。

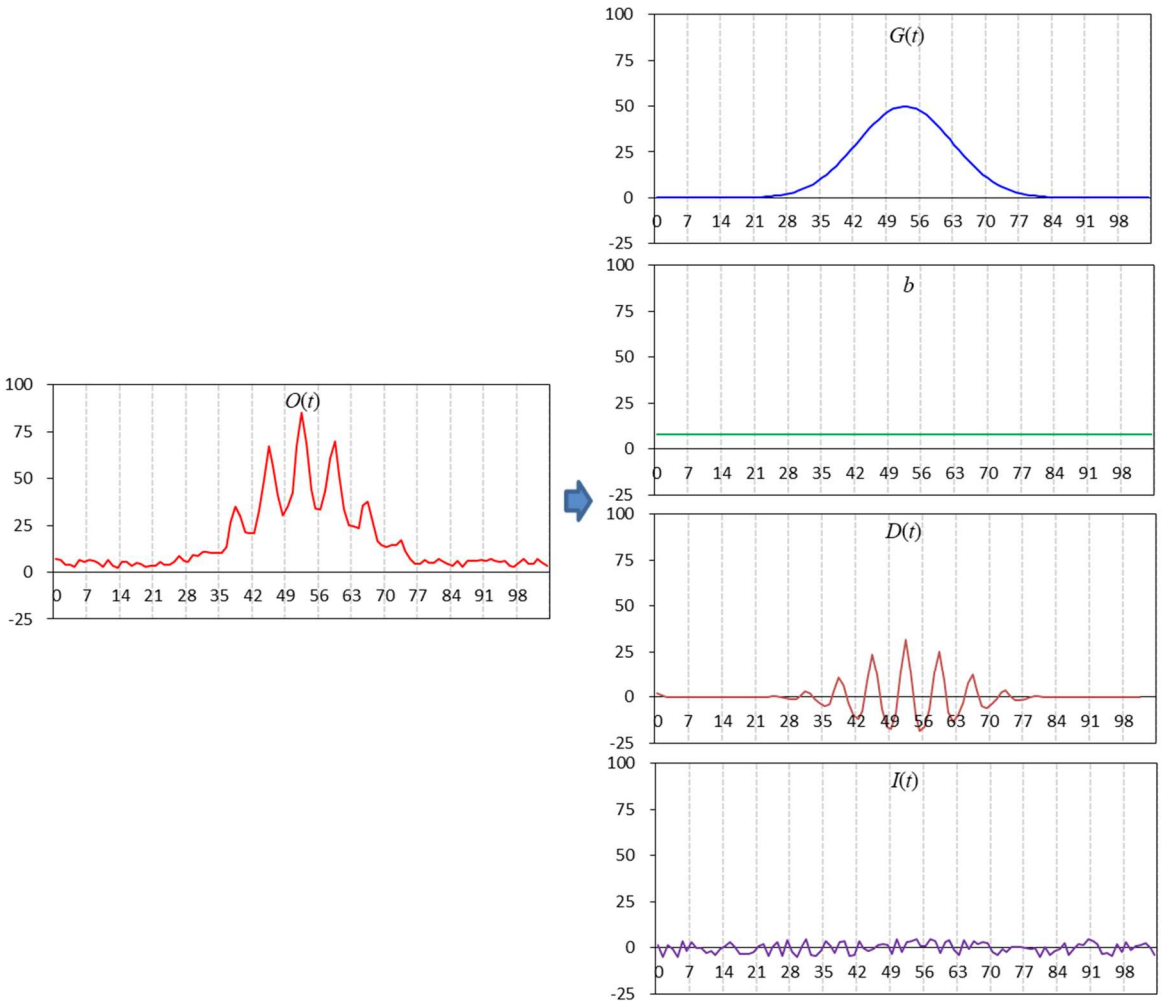

図３　モデルと４成分の模式図

式(1)中の $G(t)$と $D(t)$について説明する。

まず、$G(t)$は正規分布に準じた次式で表す。

$$G(t) = H\exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \tag{2}$$

ここで、H、μ、σはそれぞれピークの高さ、平均（ピーク位置）、標準偏差（ピークの拡がり程度）である。

次に、$D(t)$については、最初に、安定的であり１週間に亘って和をとるとほぼ０になるという曜日係数 d_nを定義する。具体的には、日付 t の曜日は $t \equiv n \pmod 7$となる n に対応する曜日であり、n=0, 1, …, 6 はそれぞれ日曜、月曜、…土曜に対応する[4)]。そして、d_n は、時間によらず一定と仮定すると、$\Sigma d_n = 0$ となる。先述したように、曜日変動の振幅は正規分布 $G(t)$に比例して変化するので、結局、$D(t)$は、

$$D(t) = d_n\, G(t) \tag{3}$$

で表される。

したがって、報告値 $O(t)$は次のように書き換えられる。

$$O(t) = G(t) + b + d_n\, G(t) + I(\mathrm{t}) \tag{4}$$

なお、式(4)の右辺の最初の３項をまとめて、推定値 $E(t)$とした。

4.2　最小二乗法によるパラメーターの推定

最小二乗法を用いて各パラメーターを推定する。そのためには、式(4)より、次の不規則変動（誤差）の二乗和、

$$\Sigma I(\mathrm{t})^2 = \Sigma\{\, O(t) - [\, G(t) + d_n\, G(t) + b\,]\,\}^2 \tag{5}$$

が最小となるように各パラメーターを求めれば良い。

以降、Excel の solver を用いた最小二乗法[5)]を使うことを前提にして話を進める。

solver のパラメーターを以下のように設定した後、solver を実行する。

① 目的セルの設定：式（５）

② 目標値：最小値

③ 変数セルの変更：$G(t)$の H、μ、σ、各曜日係数 d_n、およびバックグラウンド b

④ 制約条件の対象：$\Sigma d_n = 0$

最初に、最小二乗法により得られた各項目の $G(t)$の各係数とバックグラウンド b を表１に示した。表１には μ（ピーク位置）に対応した月/日をあわせて示した。

表１　各項目の $G(t)$の各係数とバックグラウンド b

	検査	陽性	入院	重症	死亡
H	88,234	21,561	206,609	1,977	50
μ	48	41	46	52	61
（月/日）	(8/28)	(8/21)	(8/26)	(9/1)	(9/10)
σ	25	15	16	19	17
b	31,969	758	9,622	224	11

表１より、ピーク H は入院が最も高く、次いで、検査、陽性、重症、死亡の順であった。ピーク位置 μ は陽性が最も早く、次いで、入院、検査、重症、死亡の順であった。陽性に比べて、入院は５日遅れ、検査は７日遅れ、重症は 11 日遅れ、死亡は 20 日遅れを示した。入院と死亡は報告値の場合とほぼ一致しているが、検査と重症はやや乖離していた。後２者は不規則変動が大きいことが影響しているためと考えられた。
ピークの拡がり程度 σ は検査が最も大きく、次いで、重症、死亡、入院、陽性であった。この順も不規則変動の大きさにほぼ対応していると考えられた。
バックグラウンド b は新型コロナウイルスの感染状況が比較的落ち着いた時期の人数を表していると考えられる。ただし、図１からも推察されるように、この期間の前半は後半に比べてやや高かったことから、第４波の影響が少し残っていたと考えられる。その影響がなければ、もう少し低いバックグラウンド値を示していた可能性が高い。各項目のバックグラウンドをみてみると、検査で最も高く、約 32,000 人であった。次いで、入院の 9,622 人、陽性の 758 人、重症の 224 人、死亡の 11 人の順であった。

次に、最小二乗法により得られた各項目の曜日係数 d_n を表２に示した。また、図４には曜日係数 d_n の変化を示した。

表２　各項目の曜日係数 d_n

	検査	陽性	入院	重症	死亡
d_1(日曜)	−0.785	−0.059	0.024	−0.003	−0.254
d_2(月曜)	0.012	−0.295	−0.010	0.002	−0.351
d_3(火曜)	0.157	−0.092	−0.027	0.001	−0.164
d_4(水曜)	0.272	0.096	−0.023	−0.003	0.307
d_5(木曜)	0.269	0.129	−0.001	0.001	0.282
d_6(金曜)	0.388	0.121	0.014	0.004	0.239
d_0(土曜)	−0.313	0.100	0.022	−0.002	−0.058

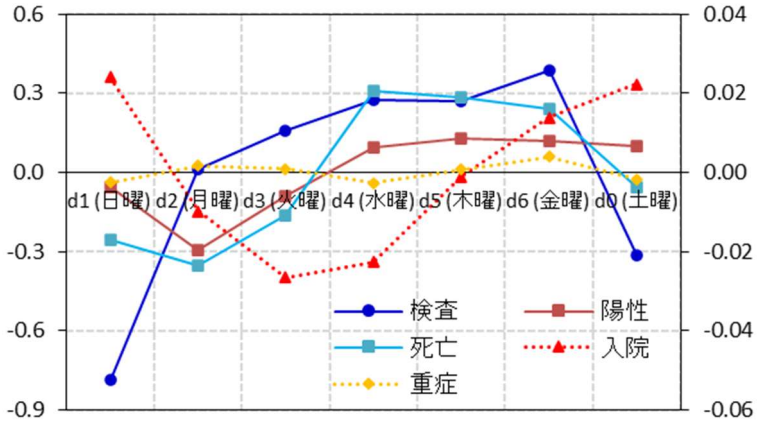

図４　曜日係数 d_n の変化（点線：右目盛）

表２については、最低値を示した曜日に着目すると、検査は日曜、陽性は月曜、入院は火曜から水曜となり、１日ずつずれていた。前々報[1)]で述べたように、検査件数の少ない日曜日の影響が陽性や入院に影響を与えた可能性がある。重症はモデル自体の精度が良

くないため、明確な傾向は見られなかった。死亡は、理由はわからないが、日曜から月曜にかけて低かった。

図４からは、各項目相互の曜日変化に違いが認められた。（詳細は後述する。）

4.3　推定値 E(t)の時系列の特徴

図５に各項目の $E(t)$の時系列を示した。

各項目とも報告値（図１）と比較的良く対応しているが、多少違いも認められた。そこで、正規分布とバックグラウンドの和 $G(t)+b$、および曜日効果 $D(t)$に分けて検討する。

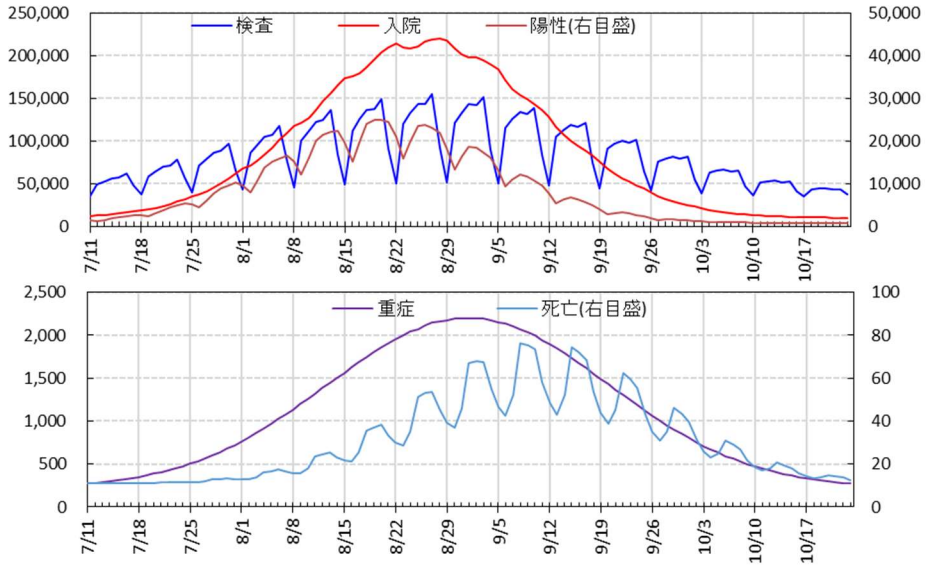

図５　推定値 $E(t)$の時系列

4.4 モデルの $G(t)+b$ と $D(t)$の時系列の特徴

先述したように、報告値の７日間移動平均と曜日効果は、モデルではそれぞれ $G(t)+b$ と $D(t)$に該当する。図６にモデルの $G(t)+b$ と $D(t)$を示した。

まず、$G(t)+b$ についてみると、全項目とも前期の小さな肩ピークを除いて７日間移動平均（図２）とほぼ同じ形状を示し、バックグラウンドレベルもほぼ同じであった。

次に、各項目の曜日効果 $D(t)$について検討した。

検査については、日曜に最も低く、次いで土曜であり、月曜から金曜にかけては徐々に上昇している。検査の残差（図２）では不規則変動が大きいため、土曜と日曜が低いという特徴以外ははっきりした傾向が見られなかったが、$D(t)$によると、検査数は平均的には週明けから徐々に高まることが示唆された。

陽性は、月曜が最も低く、次いでその前後の火曜と日曜であった。水曜から土曜にかけて高い値で推移しており、その中でも木曜が最も高かった。陽性と検査の最低値と比較すると、陽性は検査より１日遅れで現れる可能性が高い。また、報告値の残差（図２）と比較すると、両者はよく似た変化パターンを示している。

入院は、火曜と水曜が低いことから、陽性よりさらに１～２日遅れになっていると考えられた。また、土曜と日曜が高い傾向を示した。この原因はわからないが、入院は新規入院患者数ではなく、入院患者数を積み上げたものであることが一因となっている可能性がある。これまで述べたように、陽性や入院の最低値が検査より数日遅れの曜日に現れ

ていることから、検査数が陽性や入院に影響していることが考えられる。
重症は、前３者のような曜日効果は認められず、図２に比べて振幅やバラツキが極端に小さくなった。この理由として、前々報[1)]でも述べたように、重症は特定の周期成分を持たないためと考えられた。
死亡は、月曜や日曜が少なく、水曜と木曜および金曜が多くなる傾向を示した。死亡は重症と密接に関連していると考えられるが、重症は曜日効果が認められないことから、死亡の曜日効果は他の要因によるものと考えられる。

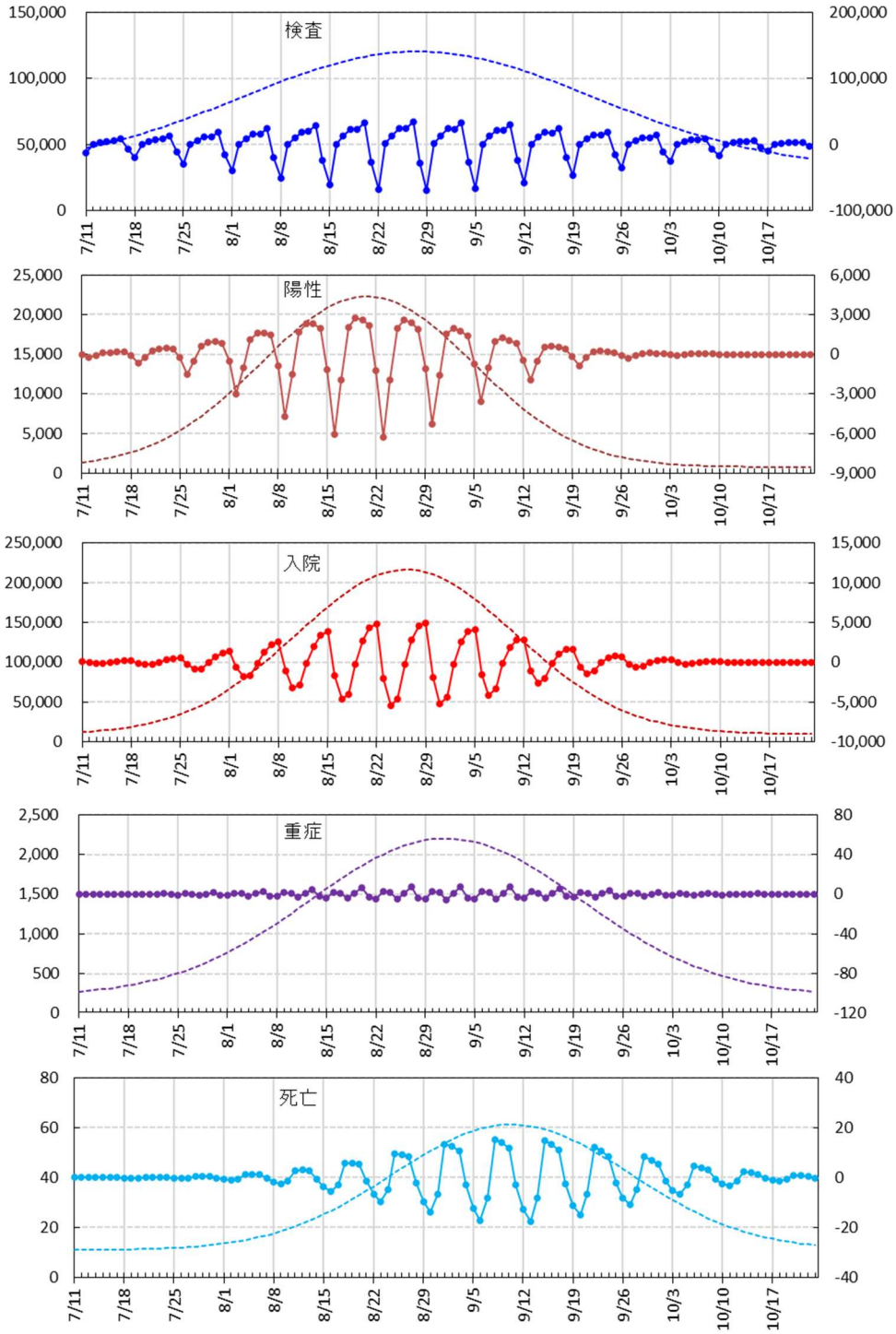

図６　モデルの $G(t)+b$ と $D(t)$ の時系列

4.5 陽性率

モデルの $G(t)$と $E(t)$を用いて陽性率％（＝100×陽性／検査）を調べた[1]（図７）。

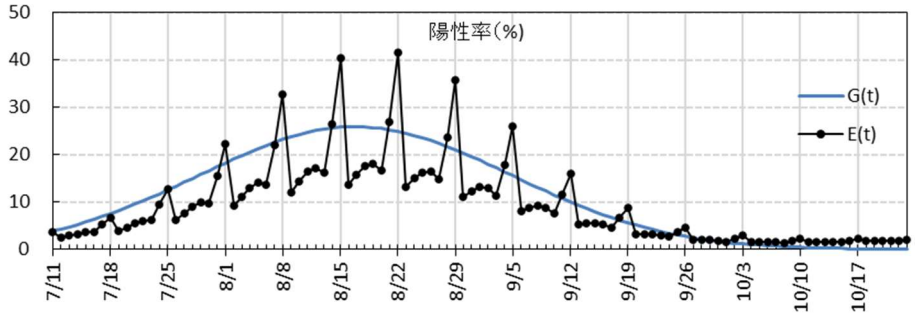

図７　モデルの陽性率の変化

まず、大きな波の変化傾向を示す $G(t)$の陽性率についてみると、8/15 の週に最大を示した。図６によると、この週に陽性もほぼ最大となっているが、検査は翌週の 8/22 に最大となっている。検査が感染状況に追いついてないことが示唆される。

次に、$E(t)$の陽性率は正規分布に曜日効果が加わる形状を示した。日曜の陽性率が最も高く、次いで土曜であった。陽性率の最も高い 8/16～22 の７日間についてみると、日曜は約 42%、土曜は約 27%であったが、週日は約 14%～18%の範囲にあった。日曜と土曜の陽性率は週日のそれに比べて平均してそれぞれ約 2.6 倍と約 1.7 倍であり、新規陽性者の一部が特に日曜と土曜で見逃されている可能性がある。感染拡大時における特に日曜と土曜の検査体制の充実が必要と考えられた。

５．まとめ

報告値の時系列に正規分布と時変曜日効果の和からなる簡易モデルを当てはめて検討したところ、陽性や入院についてはモデルが報告値とほぼ一致した。ただし、８月上旬の小さな肩ピークや前期と後期における曜日効果のわずかな違いが認められるので、次報ではそれらも考慮した詳細モデルについて考察する。

参考文献

１）物理のかぎしっぽ，新型コロナウイルスの時系列解析(2) 第５波の特徴
http://hooktail.sub.jp/contributions/covid_02.pdf
２）厚生労働省
https://www.mhlw.go.jp/stf/covid-19/open-data.html
３）時変曜日効果モデルを用いた季節調整法の適用　大島敬士
統計研究彙報　第 80 号　2023 年３月
https://www.stat.go.jp/training/2kenkyu/ihou/80/pdf/2-2-803.pdf
４）合同式　志甫淳　東京大学大学院数理科学研究科
https://www.ms.u-tokyo.ac.jp/video/open/2021koukai-kouza/op2021-001.pdf
５）物理のかぎしっぽ，周期時系列の統計解析　(2)最小二乗法による近似
http://hooktail.sub.jp/contributions/shuki02.pdf