

COVID-19新規感染者数の時系列分析

Time Series Analysis on the Number of COVID-19 Newly Infected People

小松 隆行*

Takayuki Komatsu

概要

本稿では、COVID-19（新型コロナウイルス）に関するオープンデータを、データサイエンスにおける時系列分析的な視点から分析することを試みる。対象とするデータは、日本の都道府県毎に報告されている新規感染者数の日次データ、及び世界保健機関（WHO）が報告している世界各国の新規感染者数の日次データである。これらを時系列データと捉え、これらの移動平均とファイナンス分野でのテクニカル分析を応用した手法を用いた分析結果について報告する。

1. はじめに

COVID-19（新型コロナウイルス）のパンデミックは現在も継続しており、変異ウイルスの出現が繰り返され、新規感染者数の急激な増加と減少により形成される波は世界各国で発生している。この波は、新規感染者が少ない安定期から増加することもある。波のピークから減少したものの0近くに収束する前の途中でリバウンドして次の波となることもある。これらの様子は、世界各国の様々な機関がインターネット上に公開している感染状況のオープンデータ⁽¹⁾を、グラフなどで正確に可視化することで容易に確認できる。日本でも厚生労働省や自治体などが、Excel形式やCSV形式で日次データやそれらを集計加工したデータを公開してダウンロード可能としており、利用者が独自に可視化と分析が可能である⁽²⁾⁽³⁾⁽⁴⁾。これらのデータを可視化したグラフは、様々なサイトや報道などで見られるが、単純なグラフが多く、時系列分析を行ったデータのグラフは殆ど見受けられない。人工知能などを使った予測サイトも開設されているが、従来の統計学的手法でも感染状況の様相の把握や予測も可能と考えられるが、そのような報告は見受けられない。

このようなことから本報告では、これら日本と世界のオープンデータを、時系列分析における基本的な手法である移動平均と、それを基礎とする様々な分析方法を用いて考察を行う。用いる分析手法は、ファイナンス分野でテクニカル分析⁽⁵⁾として利用

されている s 日移動平均、ボリンジャーバンド⁽⁶⁾、そのバンド幅、バンド内の相対的位置をプロットした%bチャート、MACDをCOVID-19の時系列データに応用したもの⁽⁷⁾である。以下では、北海道のデータ、大阪府のデータ、世界各国のデータに関してこれらを用いて順に分析を行う。

2. 北海道の新規感染者数に関する分析

ここでは、まず北海道の新規感染者数のオープンデータ⁽³⁾をダウンロードして分析してみる。今回は、2020年1月28日から2021年12月29日までの日次データを使用した。

2.1 移動平均を基礎とした指標による傾向分析

新規感染者数の日次データは、曜日依存した変化や検査の状況などの要因で、連続する時系列データの増減が大きく激しい変化になる傾向がある。この影響を緩和するために、時系列データ分析の手法である移動平均を考える。一般的には、ある一定期間（ n 日とする、例えば7日： n は自然数）の各日のデータの総和を、この期間 n で割った値である単純移動平均（SMA, Simple Moving Average）が用いられる。報道などでは、7日平均（7MA）で感染状況が報じられることも多い。COVID-19の新規感染者数の7日移動平均の値を7日の最終日にプロットすることで可視化する。 $n=7$ であれば短期移動平均であり、 $n=14, 28, 58$ などと、 n の値を大きくしてゆくと、 n 日移動平均は長期移動平均になってゆく。直

*北海道科学大学未来デザイン学部メディアデザイン学科

前のデータほど重みを大きくして、直近のデータに敏感に反応する指数平滑移動平均(EMA)も用いられることがある。ファイナンス分野では、これらは予測や傾向を分析するためのテクニカル分析で多用されている。他のいくつかのテクニカル分析をCOVID-19の新規感染者数のデータに適用した結果グラフを図1に示す。上から順に、①日次データとそのn日移動平均(nMA, n=7, 14, 28, 56), ②指数平滑移動平均(nEMA, nは①と同様), ③MACDcとMACDcシグナル⁽⁷⁾, ④7日平均基準のボリンジャーバンド⁽⁶⁾, ⑤ボリンジャーバンドを基にした%bチャートとそのn日平均(n=7, 14), ⑥ボリンジャーバンド幅とそのnMAとなっている。⑤の指標以外は、文献⁽⁷⁾において2021年3月2日までの分析において使用されている。一方、⑤による分析は初めてである。

今回使用したデータは、初感染確認から第5波の収束までのデータであり、第5波までの波をすべて含んでいるが、各波のピークでの最大感染者が異なるものの、MACDc⁽³⁾のダウクロスは、感染拡大期のピークアウトを捉えていると言える。これは、短期移動平均線が長期の移動平均線をダウクロスするよりも早く明確に見える。また第4波と第5波においても、ボリンジャーバンド幅⁽⁶⁾は、感染の急拡大縮小に伴い同様に増加減少していると言える。その移動平均のダウクロスは明確さに欠ける。一方で、感染拡大の兆候(起点)や下げ止まりのサインとしては、⑤の%bチャートの移動平均線の動きや状況が、そのようなサインとなる可能性があると考えられる。次節では、このようなサインの可能性について詳しく考察をしてみる。

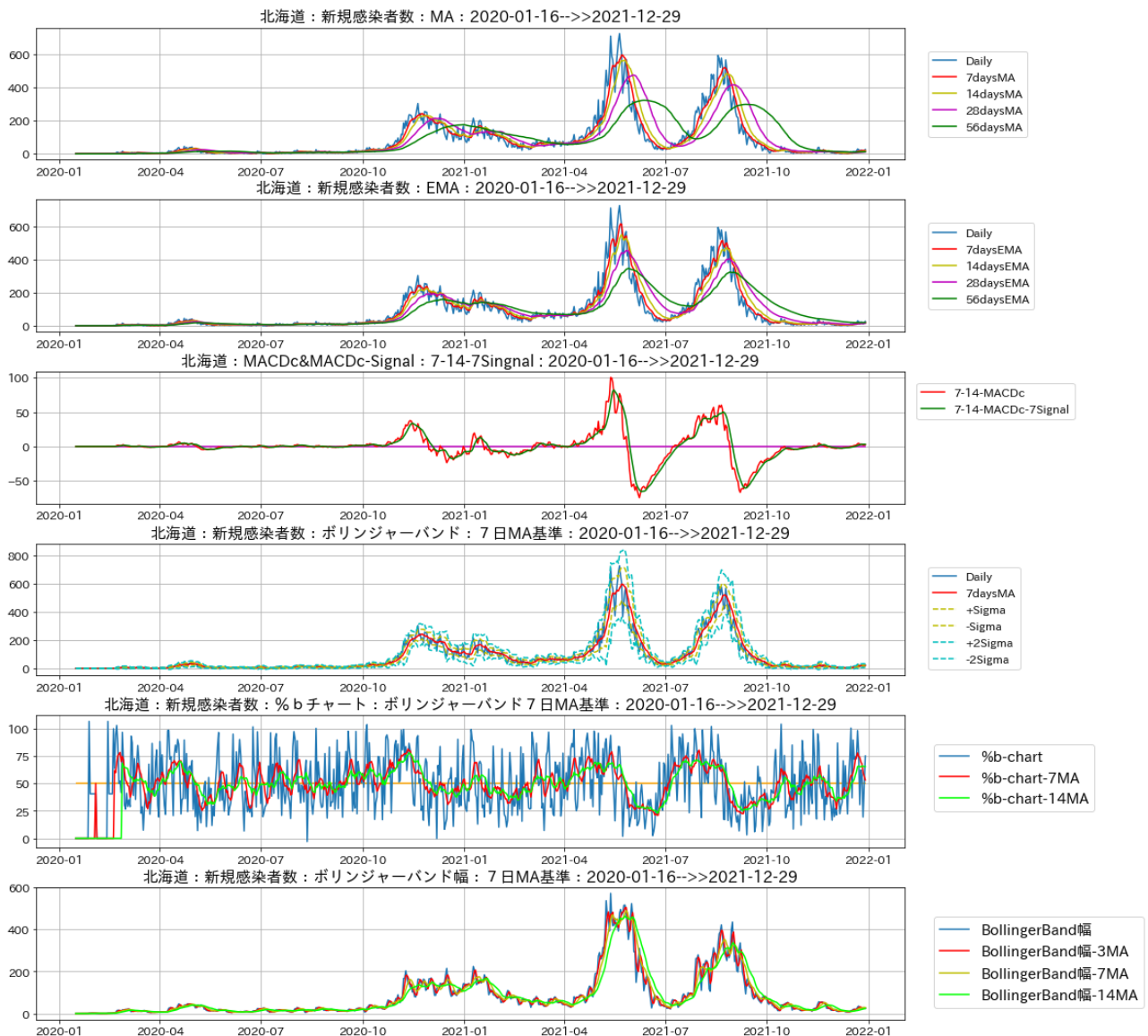


図1 北海道における新規感染者数の移動平均線を基礎とした指標のグラフ (2020/1/28~2021/12/29)

2.2 感染拡大期毎の主要グラフを複数用いた考察

ここでは、第4波と第5波を含む2021年3月5日から2021年12月29日までの北海道のデータで、2.1と同様のグラフ(図2)を用いた詳しい考察について述べる。3番目のグラフでは、MACDcのダウクロス(MACDcがそのシグナルを上から下へ突き抜ける)は感染拡大期のピークアウト(2021年5月中旬と8月下旬)を捉えていると言えるが、「揺れ」が大きくピンポイントでのサインとは言い難い。一方、6月や9月のMACDcのアップクロス(ダウクロスの逆)は、明確な意味のあるポイントを捉えてはいない。6番目のグラフでは、急拡大縮小に追従したボリンジャーバンド幅の増減が、またそれ以外の期間である安定期では小さいことが確認できる。ボリンジャーバンド幅の7MA-14MAのダウクロスがピークアウトを捉えているが、サインとしてはや

や遅れており明確ではない。このボリンジャーバンドにおいて、ある時点の日次データのバンド内での位置をプロットしたものが5番目のグラフの%bチャートである。その7日移動平均と14日移動平均もプロットした。この2つが第4波と第5波の各拡大開始点付近で7MAと14MAが50%線をアップクロスし、各波のピーク付近まで50%~80%間に停留し、その後ピークアウト辺りから減少し始め、縮小期後半で50%線をダウクロスし、安定期では20%~50%の間で停留していることが分かる。第4波後の下げ止まりから第5波へリバンドする拡大開始点付近では、%bチャート7MAが50%線を揺れることなく明確にアップクロスしている。このことから、%bチャートの7MA(および14MA)の50%線アップクロスが急拡大のサインととらえることができる。また、MACDcはこれより少し遅れて0をアップクロスしている。

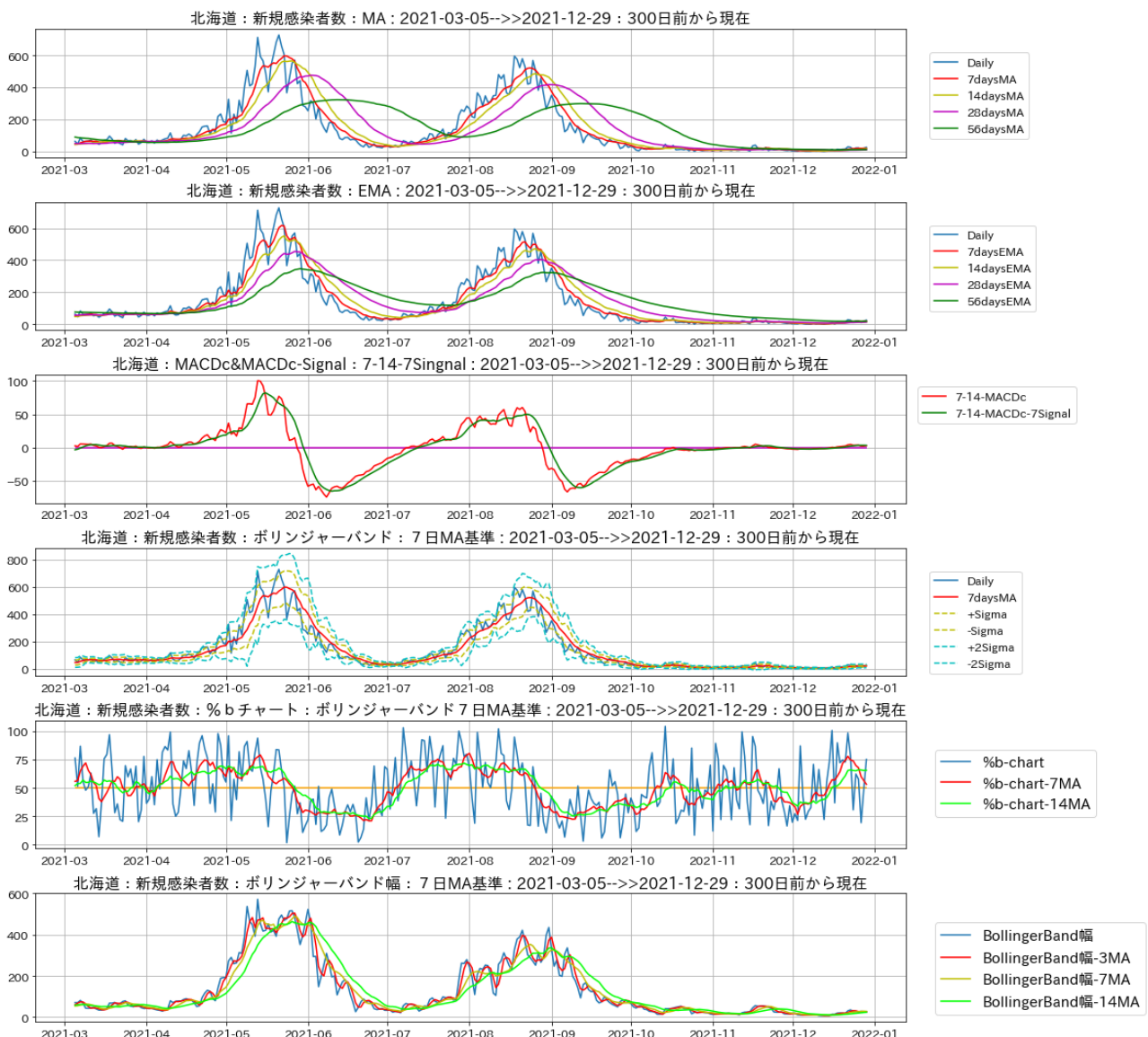


図2 北海道における新規感染者数の移動平均線を基礎とした指標のグラフ (2021/3/5~2021/12/29)

次に、7日平均の値を加工した指標を考え考察してみる。具体的には7日平均の7日平均、及び7日平均の変化率を考え、これまでと同様の考察を試みる。図3において上から順に、①日次データとその移動平均およびMACDcとそのシグナル、②%bチャートとその移動平均、③7日平均の7日平均とそのMACDc、④7日平均の変化率(前日との差)とその7日平均、⑤7日平均の変化率のMACDc、⑥7日平均の変化率の7日平均とそのMACDcである。④と⑤の7日平均の変化率は「日次データにおける当日と7日前の差(の7分の1)」であり「曜日毎の1週間での増減(の7分の1)」という意味でもある。また、⑥の7日平均の変化率の7日平均は「7日平均における

当日と7日前の差(の7分の1)」であり「曜日の起点に無関係な1週間での7日平均の増減(の7分の1)」という意味である。③では、7日平均の7MAのMACDc(赤線)が日次データピーク付近でシグナル(黄線)を揺れずに明確にダウクロスしている。④では、7日平均の変化率(青線)が急拡大期のピーク付近において、0付近で7日平均をダウクロスしている。⑤では、そのMACDc(赤線)のシグナル(黄色)とのダウクロスが、感染拡大ピーク付近で見られる。⑥では、7日平均の変化率の7日平均(緑線)のMACDcのシグナルとのダウクロスが、ピークアウトを捉えており、また緑線の0との明確なアップクロスが7日平均の意味の増減の転換点を捉えている。

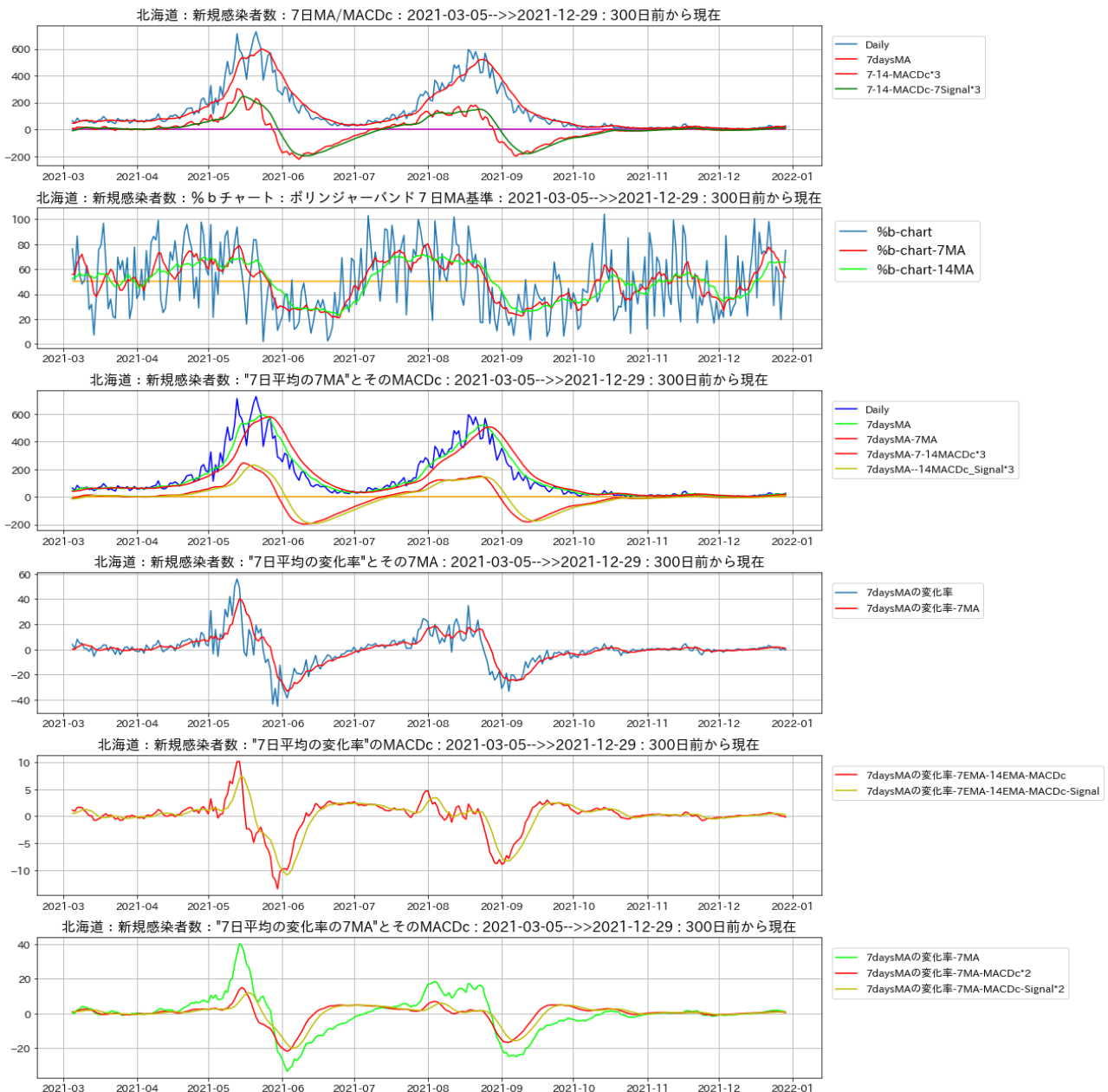


図3 北海道における新規感染者数の7日移動平均に関わる指標のグラフ (2021/3/5~2021/12/29)

3. 大都市圏の新規感染者数の分析と考察

次に、他の都府県として大阪府のデータについて分析する。図4は、図3と同様のグラフにボリンジャーバンドのグラフを2番目に加えたものである。波の開始時期では、%bチャートの2つの移動平均が50%をアップクロスし、7日平均の変化率とその7日平均が正値で増加傾向になり、波のピーク付近では、7日平均の7日平均のMACDcがピーク付近で

シグナルを明確にダウクロスし、7日平均の変化率が0付近で揺れはあるがダウクロスしている。また、7日平均の変化率の7日平均が波の開始時期に同時に増加しはじめ、また増減の転換点では0とアップクロスしている。7日平均の変化率の7日平均のMACDcのシグナルとのダウクロスが、ピークアウトを捉えている。このように、他のデータにも前章での知見が同様に適用できると考えられる。

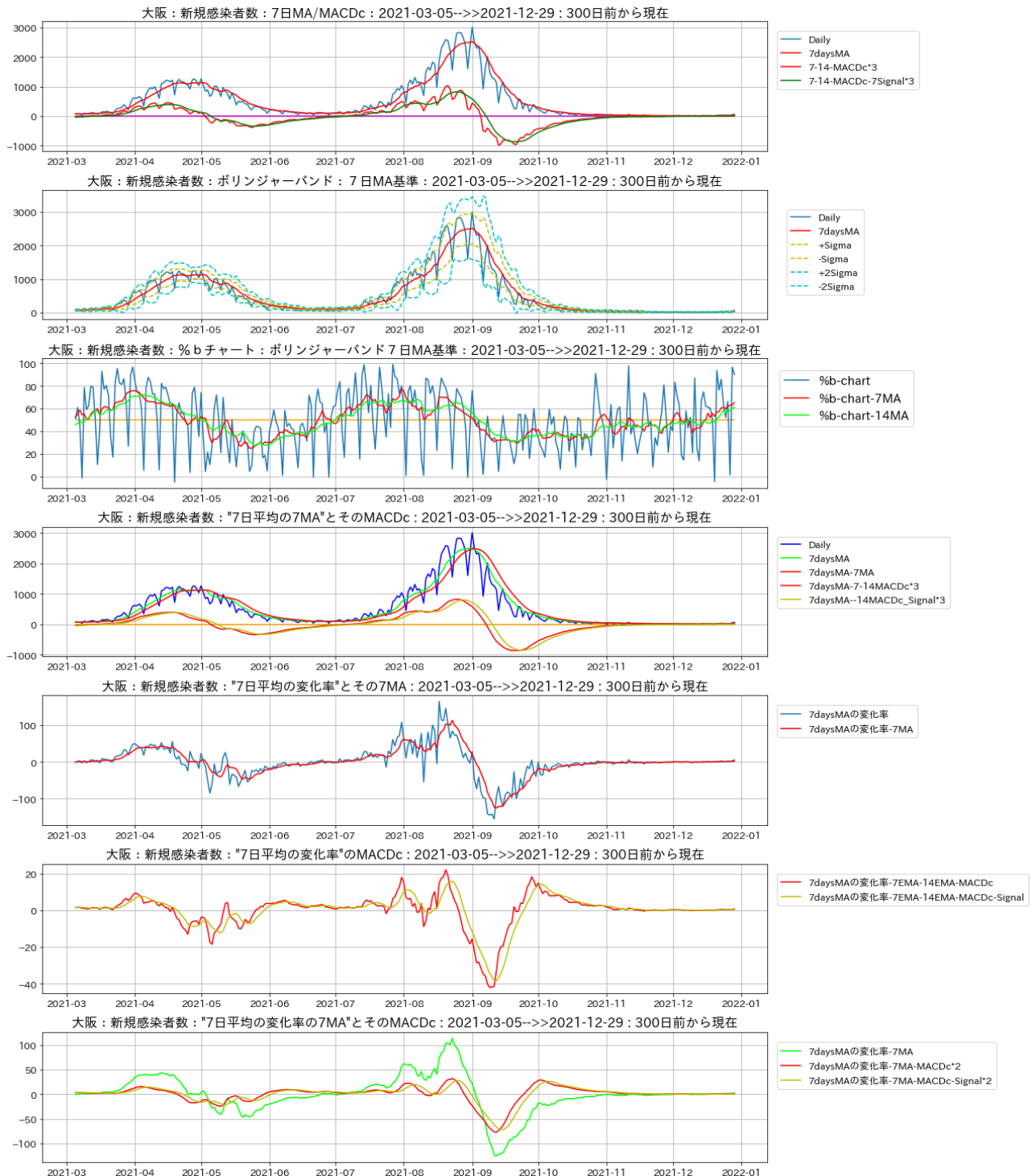


図4 大阪府における新規感染者数に関する主要グラフ (2021/3/5~2021/12/29)

4. 世界各国の分析と考察

ここでは、同様の分析を日本全体と他国(イスラエル、韓国、ロシア)のデータにおいても試みた。そ

れらの結果を以下の図5と図6に示す。2021年3月5日から同年12月29日までのデータを使用した⁽¹⁾。

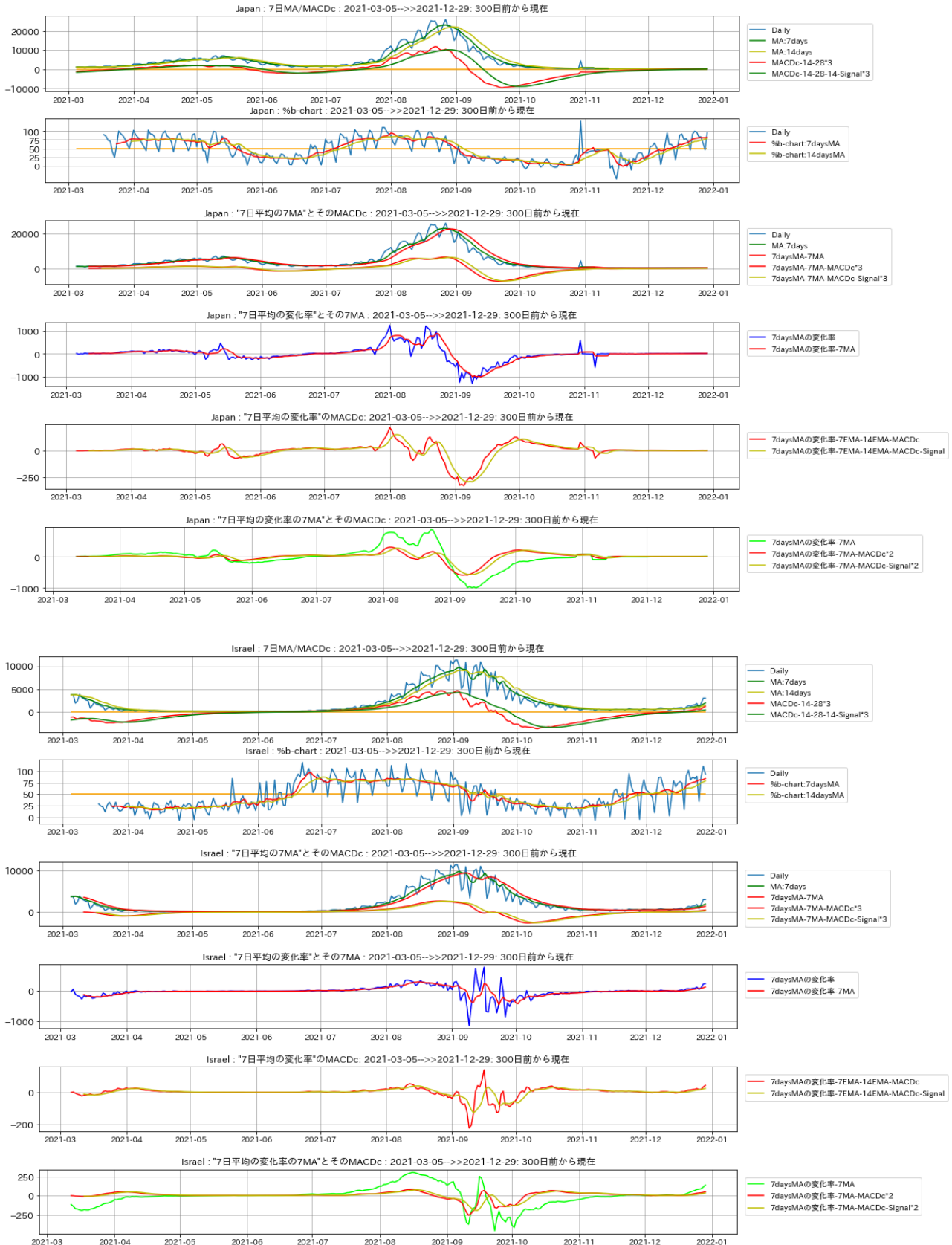


図5 世界の新規感染者数に関する主要グラフ (上：日本，下：イスラエル，2021/3/5～2021/12/29)

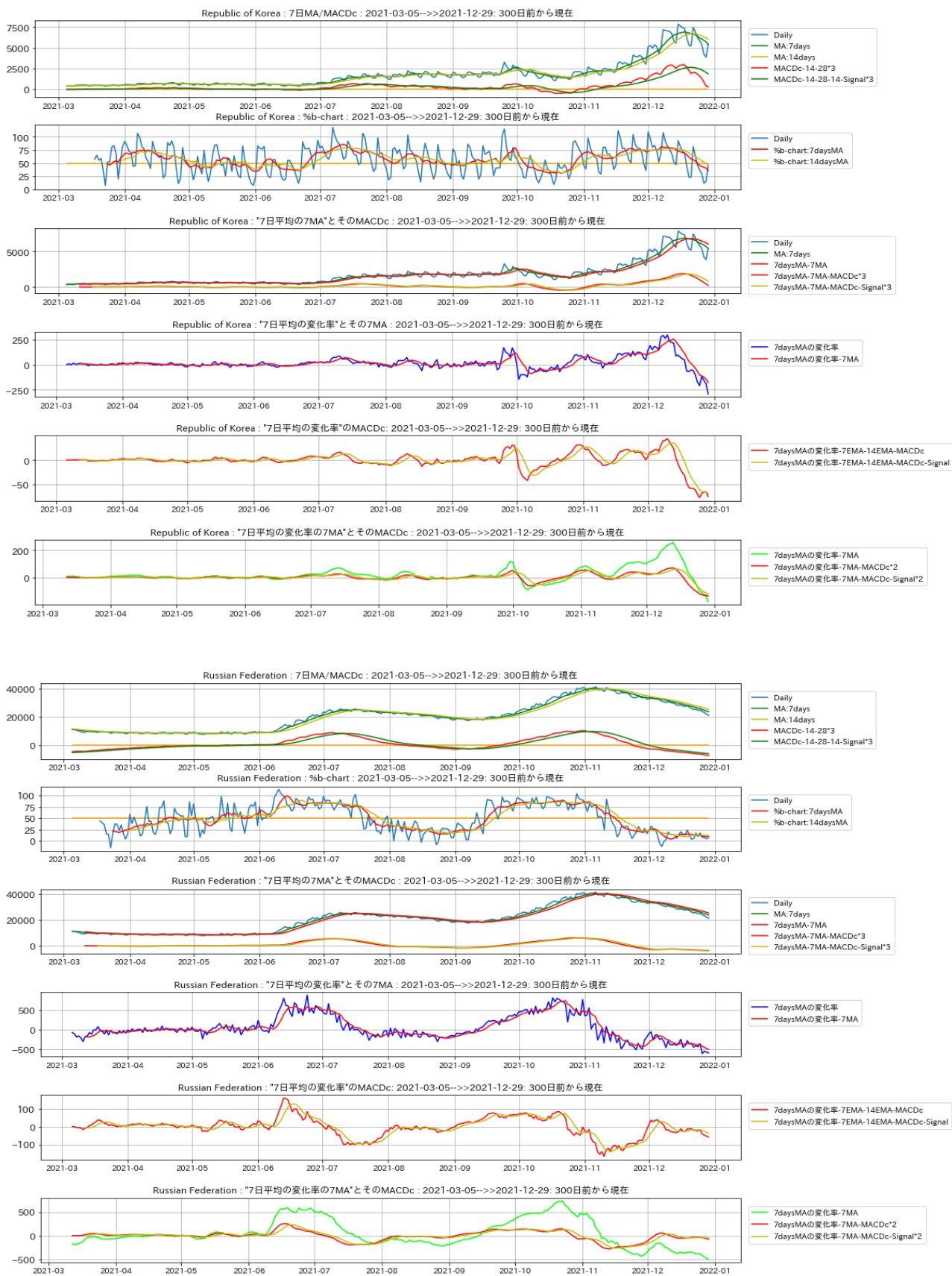


図6 世界の新規感染者数に関する主要グラフ（上：韓国，下：ロシア，2021/3/5～2021/12/29）

図5と図6から、国毎に感染の急拡大縮小期の波の大きさや期間は異なっているが、前章までで得られた知見と同様の知見が適用できると考えられる。日本以外の他国のデータで確認してみると、感染者数の急拡大期としては、図5のイスラエルの2021年6月中旬から8月にかけての時期、図6の韓国の2021年10月下旬から12月中旬、図6のロシアの2021年6月上旬から7月中旬、および9月中旬から10月上旬である。また、ピークアウト付近では、図5のイスラエルの2021年9月上旬、図6の韓国の2021年12月中旬、図6のロシアの2021年11月上旬で同様の知見が確認できる。

5. まとめ

COVID-19の新規感染者数のオープンデータを時系列データと捉えて、それを時系列分析の手法やその応用手法で分析し、いくつかの知見が得られた。メディアで多用される7日平均を基にした指標とそれらの可視化により、時系列分析の視点からの理解がしやすくなり、感染者数の増減の動向を理解したり予測したりする手がかりやサインが得られたと考えられる。

本報告で得られた知見の主なものとしては、感染増大期では、%bチャートの7日平均の50%線アップクロスと50%超での停留、7日平均の変化率の7日平均の増加および0でのアップクロスがある。波のピークでは、7日平均の7MAのMACDcが日次データ

のピーク付近でシグナルを揺れずに明確にダウンクロス、7日平均の変化率の0付近でのダウンクロスや、7日平均の変化率のMACDcのシグナルとのダウンクロスが挙げられる。これらは、感染者数の増減の起点や転換点を推測するサインとなり得ると考えられる。特に%bチャートの7日平均の動きは重要と考えられ、図3と図4における2021年12月中旬からの上昇の様子が、その直後から日本全国で起きることになる爆発的な感染の第6波の予兆と捉えることもできる。この知見は図5と図6における、日本全体のグラフと他国のグラフにおける大きな波の開始時期においても当てはまると考えられる。したがって、%bチャートの7日平均の拡大開始のサインに続いて起きる、他の指標の拡大開始のサインを監視することにより、感染拡大の開始時期を明確に判断できる可能性があると考えられる。

しかし、他の知見も多数あるが、分析結果の定量的な理解や解釈など十分ではないことは多々あるため、今後も増えてゆくデータを加えながら、今回の分析手法の改良や他の様々な手法の応用を行い、新たな知見を増やしていく必要がある。また知見が増えるに従い、それらを組み合わせてアンサンブル的に利用するための手法も必要となる。さらに、その知見をどのように有効利用してゆくかを十分に検討し、より多くのデータ（世界各国や経済圏、日本国内の都道府県や地方など）について検証することが重要であると考えられる。

参考文献

- (1) World Health Organization: WHO Coronavirus (COVID-19) Dashboard, 2022年2月27日, <https://covid19.who.int/info/>.
- (2) 厚生労働省 : 2022年3月3日, <https://www.mhlw.go.jp/stf/covid-19/open-data.html>.
- (3) NHK特設サイト新型コロナウイルス: 新型コロナデータ一覧 : 都道府県ごとの感染状況, 2021年6月30日, <https://www3.nhk.or.jp/news/special/coronavirus/data-widget/>.
- (4) 北海道庁: 新型コロナウイルス感染症 (COVID-19) に関する情報 : 2022年3月3日, http://www.pref.hokkaido.lg.jp/ss/ssa/sin_gatakoronahaien.htm.
- (5) Gerald Appel: Technical Analysis: Power Tools for Active Investors, Financial Times Prentice Hall, p.166, 2005.
- (6) John Bollinger's Official Bollinger Band Website: 2022年3月3日, <https://www.bollingerbands.com/>.
- (7) 機械学習を用いたCOVID-19新規感染者数の分析, 小松隆行, 北海道科学大学研究紀要, (49), 49-61 (2021-09-30).