

COVID-19 流行動態の再構成による メタ・ポピュレーションモデルの記述性能評価

斎藤 正也¹・竹内 昌平²・山内 武紀³・内田 満夫⁴

(受付 2021 年 8 月 3 日；改訂 10 月 26 日；採択 11 月 26 日)

要 旨

新型コロナウイルスは大都市を中心に流行が維持されている。地方での流行を分析するには交通等による人の移動を考慮することが必要である。人の移動による感染の導入とその後の流行動態を記述するモデルとして、メタ・ポピュレーションモデルとよばれる、地域毎の SIR モデルに人員の移動を追加することで地域間の相互作用を導入したものがある。このモデルはインフルエンザの流行の記述に使われた実績がある。

本研究では、メタ・ポピュレーションモデルを使つての 2020 年 10 月末までの COVID-19 の国内流行の記述可能性を検討する。そのために各都道府県の流行曲線をモデルに同化することで状態推定を行うとともに、短期間の予測誤差を計測した。地域に分解した記述のためにパラメータ数が増大し、その制約が運用上の課題である。本研究では補助モデルによる地域別の実効再生産数をデータ同化に先立って行うことで対応した。さらに、流行動態に適合させたモデルの利用例の提示として、地方への流入リスクを算定し、規模と頻度の関係として整理した。

キーワード：メタ・ポピュレーションモデル、COVID-19、流入リスク分析、データ同化。

1. はじめに

COVID-19 は 2021 年 7 月現在でも依然として大きな疾病負担を日本および世界にもたらしめている。一方、日本では安定的な流行は大都市圏に限られており、地方においては散発的な小規模流行が発生するに留まっている。このような状況下で数理モデルを使って流行分析をする場合、人の地域間移動と再生産の非決定性の考慮が重要になる。

本研究では、上述の状況をふまえ、人の地域間移動を考慮したメタ・ポピュレーションモデルを使った流行の記述可能性および短期間の予測性能を調べ、類似の新興感染症が導入された場合に役立つ応用上の技術的課題を整理する。この目的のために、流行曲線(具体的には都道府県別の新規感染者数時系列)をデータ同化することで流行動態に適合するようにモデルを校正する。また、その利用例を提示するために特に地方への流入リスク分析を行う。このモデルは SIR と同様にコンパートメント・モデルと呼ばれる属性・病態・地域ごとの人数の推移につ

¹ 長崎県立大学シーボルト校 情報システム学部：〒851-2195 長崎県西彼杵郡長与町まなび野 1-1-1

² 長崎県立大学シーボルト校 看護栄養学部：〒851-2195 長崎県西彼杵郡長与町まなび野 1-1-1

³ 昭和大学 医学部：〒142-8555 東京都品川区旗の台 1-5-8

⁴ 群馬大学大学院 医学系研究科：〒371-8511 群馬県前橋市昭和町 3-39-22

いて記述するもので、病態遷移や地域間の人の移動は二項分布でモデリングされ、全体として二項連鎖モデル(西浦, 2009)になっている. はじめ Sattenspiel and Dietz (1995) がインフルエンザの流行分析に用いたものは決定論モデルで、後に Colizza et al. (2006) が確率論的な拡張を行い、アメリカでのインフルエンザの空間分布の時間的推移を再現するのに必要なネットワーク構造のモデリングに用いた. 著者ら(Saito et al., 2018)はこのモデルを採用して 2012–13 年の国内風疹流行の再構築を行った. しかし、モデルの校正にはスーパーコンピュータを使った大規模な計算が必要であり、迅速にモデルを設計・改良してリスク等の評価を行う場合には課題がある. 本稿では、Saito et al. (2018) のモデルを採用しながら、モデルの流行動態への適合は少ない計算量で実現できる方法を検討する.

2. 手法

都道府県 $i = 1, \dots, 47$ における感受性者数 S_i , 感染者数 I_i , 除外人口 R_i に関する確率過程で流行を記述する. シミュレーションのタイムステップは利用できるデータ間隔に合わせて 1 日とする ($\Delta t = 1$ 日). 時点 $t-1$ から t の間の人数の変化を $\Delta X_{i,t} = X_{i,t} - X_{i,t-1}$, そのうち $X_{i,t-1}$ から $Y_{j,t}$ へ移動した分を $\Delta[XY]_{ij}$ とすると、この確率過程は

$$\begin{aligned}\Delta S_{i,t} &= -\Delta[SI]_{ii} - \sum_{j \neq i} \Delta[SS]_{ij} + \sum_{j \neq i} \Delta[SS]_{ji}, \\ \Delta I_{i,t+1} &= \Delta[SI]_{ii} - \Delta[IR]_{ii} - \sum_{j \neq i} \Delta[II]_{ij} + \sum_{j \neq i} \Delta[II]_{ji}, \\ \Delta R_{i,t+1} &= \Delta[IR]_{ii} - \sum_{j \neq i} \Delta[RR]_{ij} + \sum_{j \neq i} \Delta[RR]_{ji}, \\ \Delta[SI]_{ii} &\sim \text{Binom}\left(S_{i,t}, 1 - \exp\left(-\frac{\mathcal{R}_{i,t} I_{i,t} \Delta t}{N_{i,t-1} T_{\text{inf}}}\right)\right), \\ \Delta[IR]_{ii} &\sim \text{Binom}(I_{i,t}, \Delta t / T_{\text{inf}}), \\ \Delta[XX]_{ij} &\sim \text{Binom}(X_{i,t}, p_{ij}) \quad \text{for } X \in \{S, I, R\},\end{aligned}$$

で与えられる.

上記の確率過程にはパラメータとして i 県から j 県への移動確率 p_{ij} , 感染性期間 T_{inf} , 実効再生産数 $\mathcal{R}_{i,t}$ が含まれている. これらをどのように設定するか, 以下で述べる.

パラメータ p_{ij} は, Δt の間に, i 県の居住者の内 j 県に移動するものの割合である. ここでは「旅客地域流動調査」に基づく「府県相互間輸送人員表」のうち鉄道, 航空などを含む全交通機関を集計したものに依拠した(国土交通省, 2011). この表は i 県から j 県の年間輸送人員数 M_{ij} を与える. ただし, 本表による人の移動よりも少ない(または多い)設定の方が実際の流行動態により適合することが考えられるため, 最尤推定で決定する経験的な係数 Conn とともに,

$$p_{ij} = (1 - s_i) \frac{\rho_{ij}}{\sum_{j \neq i} \rho_{ij}} \quad \text{ただし} \quad \rho_{ij} = \frac{M_{ij}}{365} \times \text{Conn}, \quad s_i = \left(1 - \sum_{j \neq i} \rho_{ij}\right)^{\Delta t}$$

と定義した. Δt が小さい場合には, 単純に $p_{ij} = \rho_{ij} = M_{ij} \Delta t / 365$ とすることと同じである. ここでは Δt を大きく取った場合でも正しく移動量を計算できる競合リスクモデル(Gelfand et al., 2000)に従った. s_i はどこかに移動させる試行を毎日行ったときに, Δt 日目(Δt 回目の試行)を行った時点でなお居住地 i に留まる確率である. その余事象を分配したものが別の県 j に移動する確率 p_{ij} となるが, これは ρ_{ij} に比例するので, 表式のとおりととなる.

感染性期間の設定 発症間隔など世代時間に関係する数値は文献によってさまざまである. これらの値を勘案して, 本論文ではおおよその世代時間を再現するように, 平均感染性期間 T_{inf} は 4 日と設定した. 系統的な T_{inf} の推定を行わなかった理由には, SIR と実際の間の

ギャップもある．感染状態を1段階で表す SIR モデルでは，感染性期間と世代時間を同一視した上で，これが指数分布に従うことが暗に仮定されている．他方，COVID-19 の感染者の発症間隔は（ピークを持つ）ガンマ分布が当てはまる．

世代時間に関する推定結果をいくつか引用する．Imperial College COVID-19 Response Team (2020) ではガンマ分布を仮定し，発症間隔は平均 6.5 日，分散 0.62 日と推定している．接触状況が明らかなペアからなるデータが存在すればさらに潜伏期間や感染性の時間依存性が推定できる．He et al. (2020) の推定では，44% の 2 次感染者が 1 次感染者の発症前に現れ，発症の 12.3 日前から感染性を持ち始めるとしている．別の推定値として，感染者発生間隔が中央値で 4.6 日 (Nishiura et al., 2020)，潜伏期間が平均値で 5 日 (Linton et al., 2020) もある．感染者発生間隔が隔離などの介入により短縮されうることとも報告されており，2020 年 1 月 9 日から 2 月 13 日の中国での感染ペアの解析では (Ali et al., 2020) 7.8 日から 2.6 日に短縮された．

採用した $T_{\text{inf}} = 4$ 日という設定値は引用した論文の数値より少し短い，介入によって世代時間が短縮されることと，5 日以上の設定では後で議論するようにデータへの追従が悪化することを考慮した．

実効再生産数の擬似ダイナミクス 実効再生産数 $\mathcal{R}_{i,t}$ は地域内での小規模流行を仮定した予備解析(次節で述べる)を別途行い，プラグインする．ただし，感染報告がない期間は未定義になるためランダムウォークによる進化

$$\mathcal{R}_{i,t} = \mathcal{R}_{i,t-1} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, 0.1^2)$$

を導入し，確率的な外挿によりアウトブレイクとアウトブレイクの間でも実効再生産数が定義されるようにした．なお，分散 0.1^2 はデータに追従できて，かつ計算の破綻が起きない十分な値を選んだ．

観察モデルおよび参照するデータ 実効再生産数のランダムウォークを追加したメタ・ポピュレーションモデルによる予測と実際の新規感染者数 $J_{i,t}^{\text{obs}}$ との間に

$$J_{i,t}^{\text{obs}} \sim \text{Poisson}(\Delta[SI]_{ii})$$

との関係を仮定して構成される状態空間モデルでデータ同化を行う．新規感染者時系列 $J_{i,t}^{\text{obs}}$ にはジャックジャパン株式会社 (2020) の提供による令和 2 年 1 月 15 日～11 月 11 日の毎日の都道府県別の報告(300 時点)を用いる．

2.1 地域毎の実効再生産数の予備推定

継続した感染報告があるのは東京を中心とする大都市に限られ，流行の波と波の間では散発的な小規模流行のみが見られる地域も多い．そのため全体の流行動態だけでは実効再生産数に意味のある制約を与えるのは困難である．

そこで，単発の小規模流行を記述するモデルを使って，流行があるかぎり実効再生産数を推定し，フル・シミュレーションにプラグインするアプローチを取った．ここで，小規模流行とは T_{inf} 以上続く 0 人報告を区切りとする連続する感染者の報告列である．はしか・風しんのクラスタ分析に Yoshikura and Takeuchi (2016) が用いた方法を参考にした．このように定義した感染報告列 $J_{i,t}$ が負の二項分布

$$(2.1) \quad J_{i,t} \sim \text{NegBin} \left(\mu = \frac{I_{i,t} \mathcal{R}_{i,t} \Delta t}{T_{\text{inf}}}, \sigma^2 = \mu + \frac{I_{i,t} \sigma_0^2 \Delta t}{T_{\text{inf}}} \right).$$

に従うと仮定する．一般に，負の二項分布は母平均がガンマ分布に従うゆらぎを持つポアソン分布として定式化される．この式の導出では，感染者と感受性者の有効接触数を試行回数とす

る二項分布の近似としてのポアソン分布と1日の有効接触数の感染者毎のばらつきがガンマ分布に対応する．分岐過程を用いて少数の感染者による再生産過程を記述することにより精密な導出が可能である (Nishiura et al., 2012)．

3. 結果

地域毎の実効再生産数の予備推定 再生産に負の二項分布を用いた，式(2.1)による都道府県別の再生産数 $R_{i,t}$ の推定結果を図1に示す．いわゆる第1波(3～4月)，第2波(7～8月)を除き，多くの地方では再生産数が定義できるような流行が起こっていないことが確認できる．また，大都市圏に限っても，例えば北海道で9月以降，東京に先駆けて流行が拡大するなど(9月上旬から中旬にかけて北海道では $R_{1,t} > 1$)，流行の進行には地域差が見られる．

交通流のキャリブレーション 流動表をそのままプラグインした移動頻度 (Conn = 1 のときの p_{ij}) は，実際の流行動態に照らして地域間の相互作用が過大である．理由のひとつには，緊急事態宣言等における行動自粛の影響が考えられる．そこで，移動量の係数 Conn を 0.01～1 の範囲で変えて，尤度を評価した．2020年10月末までのデータに対する尤度は図2の通りである．図によると Conn = 0.1 程度で最大となることがわかる．以下の解析では Conn = 0.1 を用いる．

データ同化の結果とモデルの記述力 モデルのデータへの追従性能を確認するためにデータ全体を同化し，フィルタ分布の平均軌道を観察した．大都市圏では概ね各都道府県の動態に追従しているが，いくつかの小規模流行の見逃しがある．例えば，図3では広島県の6月末の立ち上がりや長崎県の6月末および7月中旬をピークとする感染者の列が捕捉されていない．

図3では $T_{\text{inf}} = 4$ 日のほかに $T_{\text{inf}} = 2, 5, 8$ 日でのデータ同化結果も示している． $T_{\text{inf}} = 5, 8$ 日では大都市圏での動態に追従できていない．しかし，これがただちにこれらの長い感染性期間を排除するものとは考えにくい．SIR モデルでは感染者数のスケールリングによりほぼ新規感染者数の流行曲線を変えることなく感染性期間を任意に設定できるからである．追従性能の悪化原因として，式(2.1)による再生産数の推定において，1次感染者数が感染性期間にわたって平滑化されることにより $R_{i,t}$ の変動が正しく捕捉されないことが予想される．

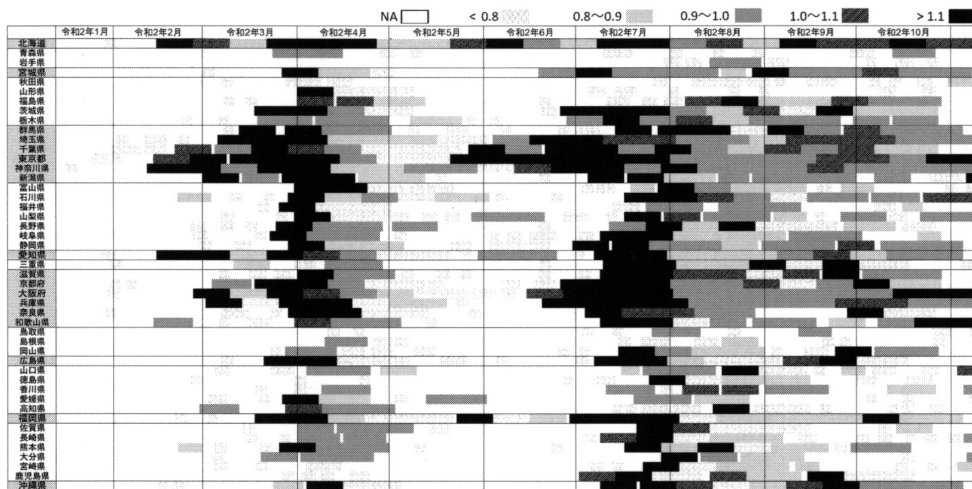


図1. 都道府県別の実効再生産数の推移(令和2年1月～10月)．

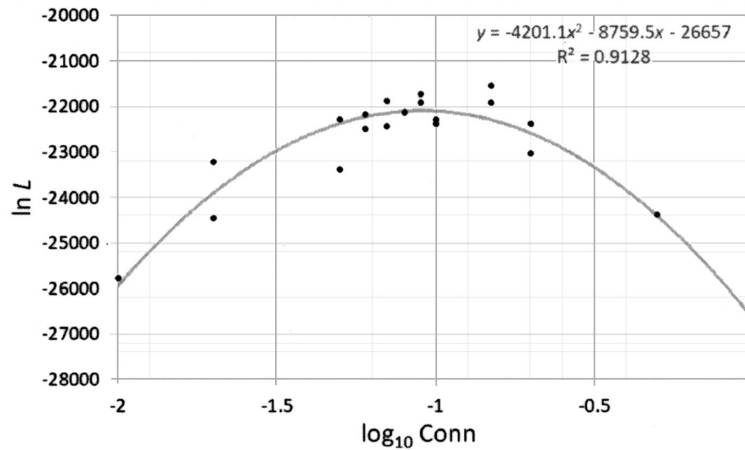


図 2. 都道府県間結合(流動表と等倍のとき $\text{Conn} = 1$)に対する尤度. 灰色の曲線はモンテカルロ法による推定値の 2 次関数による補間.

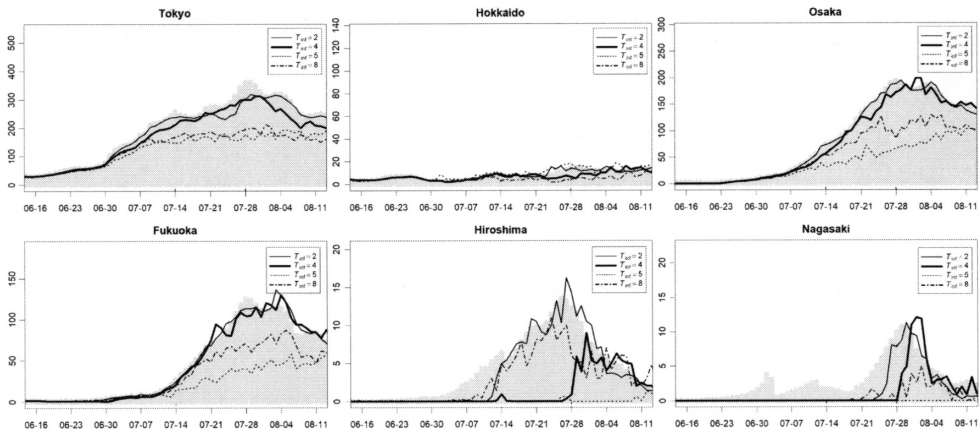


図 3. 6 都道府県のデータ同化結果(270 日間, 10 月 10 日まで). 観測データを網掛けで, 分布の各分位点(2.5%, 50%, 97.5%)を線で表示.

つぎにモデルの予測性能を評価する. SIR モデルが含む感染性の変動に関する作用機序は集団免疫者の増加に限られるので, 予測は最新の実効再生産数に固定した外挿に限られる. そこで, 実効再生産数がほぼ一定である上昇トレンドに限定してモデルとデータを比較し, 誤差を評価することとした. 60 日間のデータ同化により状態を調整した上で, 7 月 14 日を起点とする予測を行い, 2 週間の平均自乗誤差(RMSE)を計算した. 誤差計算のもとになったフィルタリングおよび予測結果を観察すると(図 4), 予測が実際の報告をカバーできるのは東京, 大阪, 福岡などの大都市に限られている. 実際, 流行規模と自乗誤差平方根の散布図(図 5)を描くと, これらは逆相関になっており, 特に小規模流行ではデータと計算の間に数倍の違いがあることがわかる.

予測例: 長崎県の場合 地方では継続した流行がないため予測分布の平均や分位はリスクや負荷の表示として適切な指標となりえない. そこで, 都道府県別の $R_{i,t}$ の推定を行ったときと

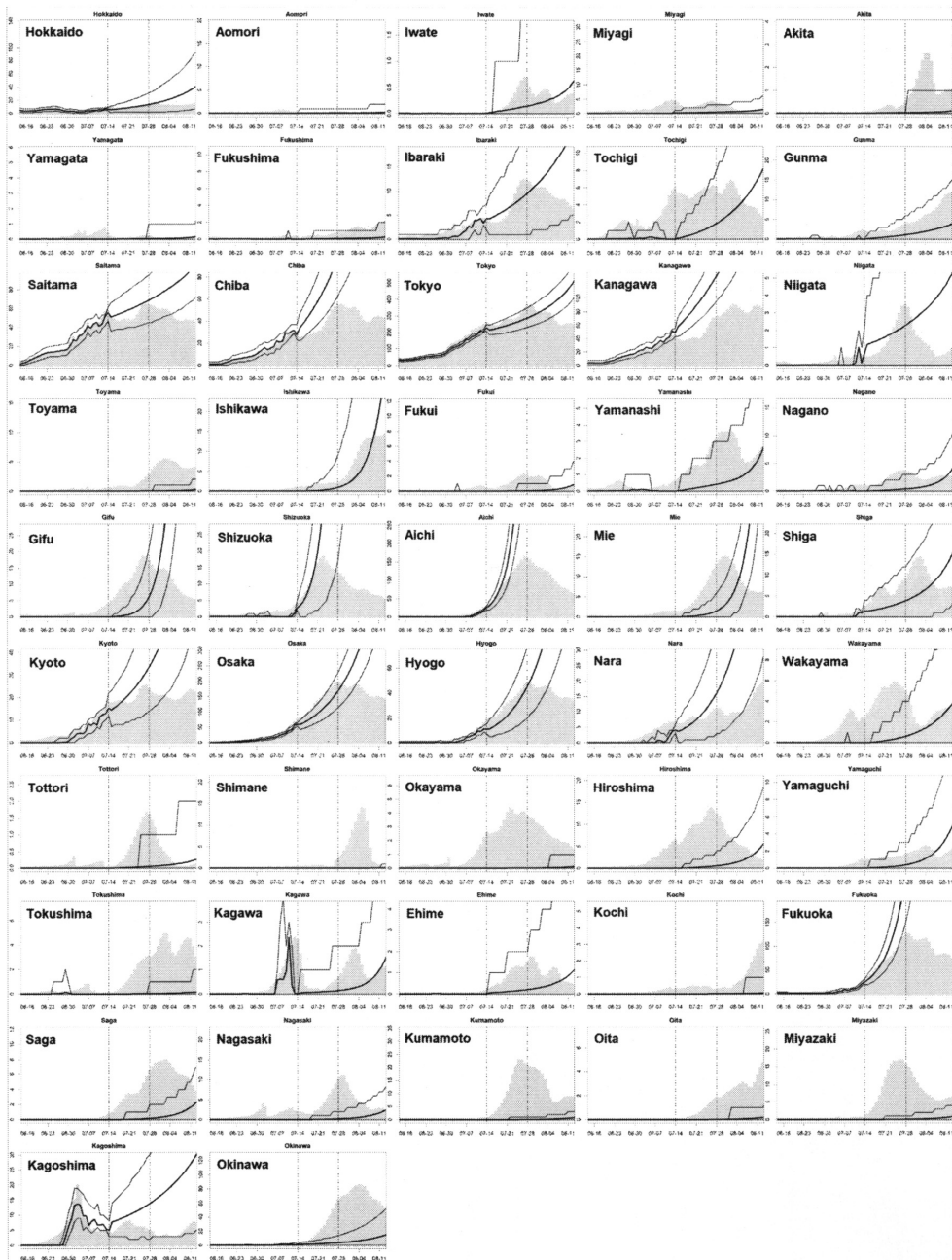


図 4. 47 各都道府県における 7 月 14 日までのフィルタリングとその後の予測。観察データを網掛けで、分布の各分位点 (2.5%, 50%, 97.5%) を線で表示。

同じく、0 でない新規感染者数が連続する期間をひとつの流行と数え、その期間に発生した新規感染者数の合計をその流行の規模として、待ち時間と発生頻度の関係を各時点における予測分布を構成するサンプルをもとに計算した。計算例として長崎県の場合を図 6 に示す。待ち時

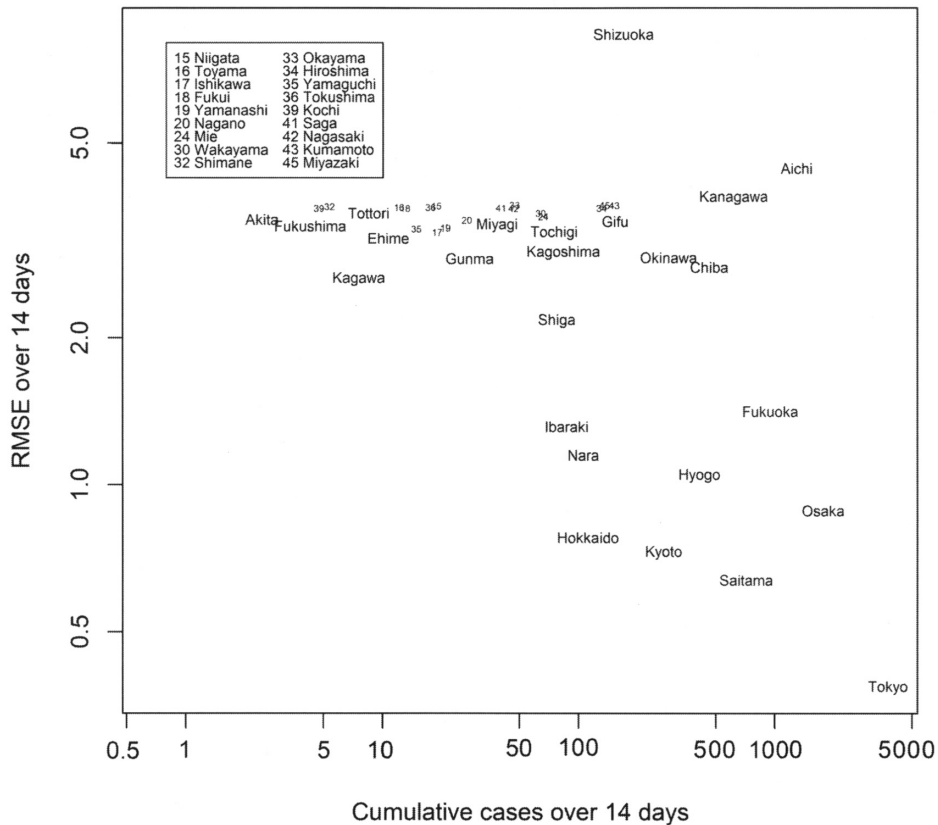


図 5. 7 月 14 日を起点とする 2 週間の予測に対する、自乗誤差平方根(横軸)とデータによる累積感染者数(縦軸)。

間を 90 日とした場合に、5 人以上が 20%，10 人以上が 6%，20 人以上が 2% の確率で発生すると試算された。

4. おわりに

本報告では、確率的メタ・ポピュレーションモデルとよばれる地域毎の確率的 SIR モデルを人の流入に基づいて連結したモデルを用いて、COVID-19 の日本での令和 2 年 10 月までの流行状況を分析した。その分析に基づくリスク評価の例として長崎県にその後発生しうる流行規模を発生確率とともに計算した。

インフルエンザのような感染者数がどの地域でも連続した流行が見られる感染症と異なり、モデルに含まれる力学パラメータを十分に制約するだけの情報が得られないという問題があった。ややアドホックではあるが、可能な限り観察データから情報を抽出するために、今回の解析では地域毎の実効再生産数の事前推定値を使う方法を取った。完全にバイズ的な方法と比較すると不確かさの評価にバイアスが入るなどの問題もあるが、確率的メタ・ポピュレーションモデルのようなフレキシビリティの高い力学モデルを限られた計算資源で活用する場合には現実的な方法と考えられる。実際、同じモデルを用いた風しん国内流行の分析では、流行曲線の変動を類型化することで区間毎の再生産数の推定問題に帰着し、パラメータ数を減らしたが、

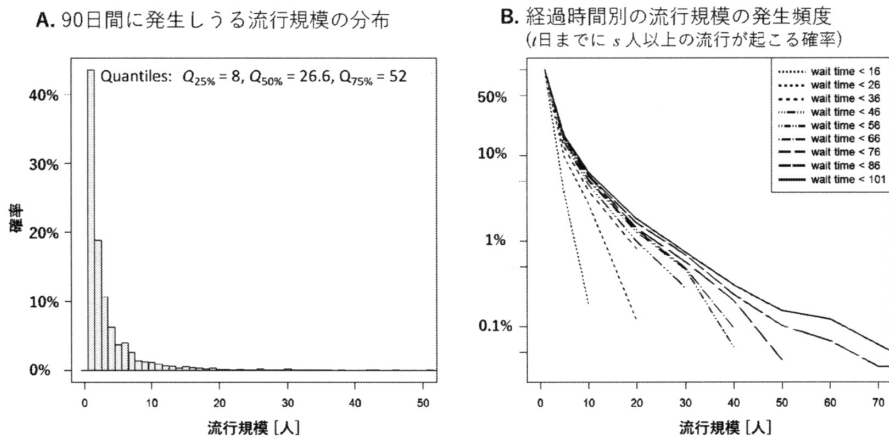


図 6. 長崎県における流行規模と発生頻度. A. 予測時点から 90 日間の間に発生しうる流行規模の頻度分布. B. 予測時点から t 日後 (図中の waiting time) に s 人以上の流行が発生する確率.

それでも一部の区間では MCMC が収束しないなど識別性の困難があった。

本研究の到達地点はリスク分析にメタ・ポピュレーションモデルが応用できることを概念的に示したに留まり、実際の対策立案に用いる場合には必要であるが本モデルには含まれていない要素がある。実際の COVID-19 では流行の担い手が高齢者である時期と若年層であった時期がある。また、人の移動パターンはアンケートに基づく特定時点でのスナップショットである。流行動態は緊急事態宣言等における行動変容を反映しているため、予測には地域間移動の経時変化を捉えたデータが必要である。今後、既報の次世代行列をベースとしてモデルパラメータの爆発を抑えつつ年令構造を取り入れ、移動量の経時変化にはモバイル空間統計を利用することを検討している。

謝 辞

本研究は情報・システム研究機構「COVID-19 対策研究プロジェクト」、JSPS 科研費 JP18K11541, AMED の課題番号 JP20fk0108140 の支援を受けて実施した。また、査読者の助言によりさらなる解析を行うことができた。ここに感謝の意を表明する。

参 考 文 献

- Ali, S.T., Wang, L., Lau, E.H.Y., Xu, X.K., Du, Z., Wu, Y., Leung, G.M. and Cowling, B.J. (2020). Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions, *Science*, **369**, 1106–1109.
- Colizza, V., Barrat, A., Barthe'lemy, M. and Vespignani, A. (2006). The modeling of global epidemics: Stochastic dynamics and predictability, *Bulletin of Mathematical Biology*, **68**, 473–481.
- Gelfand, A.E., Ghosh, S.K., Christiansen, C., Soumerai, S.B. and McLaughlin, T.J. (2000). Proportional hazards models: A latent competing risk approach, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **49**, 385–397.
- He X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19, *Nature Medicine*, **26**, 672–675, doi: 10.1038/s41591-020-0869-5.

- Imperial College COVID-19 Response Team (2020). Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries, doi: 10.25561/77731.
- ジャグジャパン株式会社 (2020). 都道府県別新型コロナウイルス感染者数マップ, <https://gis.jag-japan.com/covid19jp/>.
- 国土交通省 (2011). 旅客地域流動調査—府県相互間旅客輸送人員表(全機関), <https://www.e-stat.go.jp/stat-search/file-download?statInfId=000027669878&fileKind=0>.
- Linton, N.M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A.R., Jung, S., Yuan, B., Kinoshita, R. and Nishiura, H. (2020). Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data, *Journal of Clinical Medicine*, **9**, 538, doi: 10.3390/jcm9020538.
- 西浦博 (2009). 感染症の家庭内伝播の確率モデル：人工的な実験環境, 「特集：確率過程の統計解析」, 統計数理, **57**(1), 139–158.
- Nishiura, H., Yan, P., Sleeman C.K. and Mode, C.J. (2012). Estimating the transmission potential of supercritical processes based on the final size distribution of minor outbreaks, *Journal of Theoretical Biology*, **294**, 48–55, doi: 10.1016/j.jtbi.2011.10.039.
- Nishiura, H., Linton, N.M. and Akhmetzhanov, A.R. (2020). Serial interval of novel coronavirus (COVID-19) infections, *International Journal of Infectious Diseases*, **93**, 284–286.
- Saito, M.M., Nishiura, H. and Higuchi, T. (2018). Reconstructing the transmission dynamics of rubella in Japan, 2012–2013, *PLOS ONE*, **13**(10), e0205889, doi: 10.1371/journal.pone.0205889.
- Sattenspiel, L. and Dietz, A. (1995). Structured epidemic model incorporating geographic mobility among regions, *Mathematical Biosciences*, **128**(1-2), 71–91, PMID: 7606146.
- Yoshikura, H. and Takeuchi, F. (2016). Measles and rubella: Scale free distribution of local infection clusters, *Japanese Journal of Infectious Diseases*, **69**, 293–299.

Reconstruction of COVID-19 Epidemic in Japan Using a Meta-population Model and Evaluation of Its Expressibility

Masaya M. Saito¹, Shouhei Takeuchi², Takenori Yamauchi³ and Mitsuo Uchida⁴

¹Faculty of Information Systems, University of Nagasaki, Siebold

²Faculty of Nursing and Nutrition, University of Nagasaki, Siebold

³Public Health and Preventive Medicine, Showa University

⁴Graduate School of Medicine, Gunma University

COVID-19 epidemic in Japan has been droved mainly in Tokyo and other metropolitan areas. Small outbreaks observed in the rest of region will not be well described by a mathematical model unless the mobility of people is taken into consideration.

This study aims to evaluate the expressibility of the meta-population against the COVID-19 epidemic over Japan's prefectures. For this purpose, we have carried out the state estimation using data assimilation and evaluated the short-term prediction errors. A key issue in employment of this model is configuration of a large number of parameters, that is, the effective reproduction number in each prefecture. Since the transmission chains are not sustained in many local areas, it is hard to integrate the estimation into the data assimilation procedure in the view of identifiability. For this reason, we have incorporated a simplified local outbreak model. We discuss a limitation involved by this secondary model. As a demonstration of the utilization of the constructed predictor, we also have assessed the size and frequency of small outbreaks induced by importation from other areas.