

人流ビッグデータを内包する感染症モデルの データ同化による新型コロナウイルス感染動態の解析

Infectious disease model encompassing human flow big data Analysis of the dynamics of novel coronavirus infection by data assimilation

ビジネスデータサイエンス専攻 樋口研究室
22N7100001I 荒井 智史

1 序論

1.1 研究背景

2020年1月に国内で初めて感染者が確認された新型コロナウイルスの流行が始まってから約4年の月日が経過し、その影響は我々の生活のあらゆる面に影響をもたらした。2023年5月には、新型コロナの分類が5類感染症へ移行され、この移行に伴いコロナ以前の生活様式が戻ってくるようになった。本研究を開始した2022年の6月付近では、以前の株よりも感染力が強いオミクロン株が全世界的に拡大し、国内では第6波が到来していた。このような状況を背景に、新型コロナウイルスという社会問題に数理科学の立場から貢献したいと考え、本研究を開始した。

1.2 研究目的

本研究の目的は、微分方程式による新型コロナウイルス感染症に対応した数理モデルを作成し、ビッグデータを取り入れたシミュレーションによる各感染波の再現を実施し、人流が感染動態に与える効果を数理モデルの観点から考察することである。

1.3 使用データについて

1.3.1 沖縄県における新型コロナウイルス感染者数データ

本研究では、厚生労働省が公開している新型コロナウイルス感染者数オープンデータを使用した [1]。このデータには、2020年1月16日から2023年5月8日の期間における日毎の感染者数が含まれている。”このデータを用いて、新型コロナウイルスの感染状況の詳細な分析を進めた。

1.3.2 V-RESAS 人流比率データ

本研究では、沖縄県の滞在人口に関する人流比率データを扱った。新型コロナの感染拡大に人の移動や人流が重要な効果の一つと考えられていることに焦点を当て「人流の増加により、感染者数が増える」事を仮説にした。これを検証するために本研究では、内閣府が地方創生を目的に公開している人流ビッグデータを扱うこととした [2]。

また、この人流データはユーザーごとに推定される居住地によって市区町村内、都道府県内、都道府県外の3種類のカテゴリに分類される。

- 市区町村内 (当該市区町村に滞在している人口のうち、推計居住地が同じ市区町村である人口)
- 都道府県内 (当該市区町村に滞在している人口のうち、推計居住地が同じ都道府県内の他の市区町村である人口)
- 都道府県外 (当該市区町村に滞在している人口のうち、推計居住地が他の都道府県である人口)

2 感染症モデル (SEIR モデル, SEIR+ モデル) について

2.1 SEIR モデル

感染症数理モデルの一つである SEIR モデルについて解説する [3]。SEIR モデルは、限られた空間内での人から人への直接伝搬する感染症の感染動態をシンプルに表現するための感染症モデルである。SEIR モデルでは、総人口を一定と仮定し、人口を「感受性のある者 (S: Susceptible)」「隔離者 (E: Exposed)」「感染者 (I: Infectious)」そして「回復者 (R: Recovered)」の4つのグループに分類する。SEIR モデルにおける状態遷移図を図1に示す。SEIR モデルの微分方程式を式 (1) に示す。このモデルにおける「限られた空間」とは、一定の人口集団が存在する閉じた空間を指し、総人口 N は $S+E+I+R$ の合計に等しい。

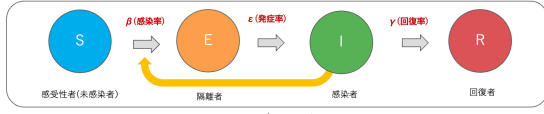


図 1:SEIR モデルにおける状態遷移図

$$\begin{aligned}
 \frac{dS}{dt} &= -\beta S(t)I(t) \\
 \frac{dE}{dt} &= \beta S(t)I(t) - \sigma E(t) \\
 \frac{dI}{dt} &= \sigma E(t) - \gamma I(t) \\
 \frac{dR}{dt} &= \gamma I(t)
 \end{aligned}
 \tag{1}$$

ここで β は感染率、 σ は発症率、 γ は回復率を表している。SEIR モデルは、SEIR の各カテゴリ間の変化を微分方程式で表現しており、時間とともに各カテゴリの個体数がどのように変化するかを表現できる。このモデルを利用することで、感染症の初期の拡大速度や最大の感染者数、感染が収束するまでの時間を予測し、感染症の特性を推察することができる。

2.2 SEIR+ モデル

先の SEIR モデルに、本研究で扱う人流データを取り込める様に独自の改良を施したモデルを SEIR+ モデルを提案する。SEIR+ モデルは以下の式で表すモデルである。

$$\begin{aligned}
 \frac{dS}{dt} &= -\hat{\beta}(t; \theta)S(t)I(t) \\
 \frac{dE}{dt} &= \hat{\beta}(t; \theta)S(t)I(t) - \sigma E(t) \\
 \frac{dI}{dt} &= \sigma E(t) - \gamma I(t) \\
 \frac{dR}{dt} &= \gamma I(t)
 \end{aligned}
 \tag{2}$$

$$\hat{\beta}(t; \theta) = \beta + p_1 \cdot PF_1(t) + p_2 \cdot PF_2(t) + p_3 \cdot PF_3(t)
 \tag{3}$$

$$\theta = (\beta, p_1, p_2, p_3)
 \tag{4}$$

感染率の項に対して追加項として人流データを線形に加えるよう改良を施し、感染率 β を時間変化するパラメータとして扱うようにした。 $PF_n(t)$ ($n = 1, 2, 3$) はそれぞれ市区町村内、都道府県内、都道府県外の人流データを表し、それぞれの係数 p_n ($n = 1, 2, 3$) は $(0,1)$ の範囲を取るパラメータである。

2.3 SEIR,SEIR+ モデルのシミュレーションについて

シミュレーションの進め方として、各モデルを用いて第 2 波から第 8 波までの各感染波を再現出来るようシミュレーション及びパラメータ最適化を行った。各モデルを、時間差分を $\Delta t = 1$ 日として 4 次のルンゲクッタ法で解き、パラメータ最適化は負の対数尤度の最小化を目的関数として、Python の最適化ライブラリ *Optuna* を用いて実施した。その後、各モデルの結果を *AIC*(*Akaike Information Criterion*: 赤池情報量基準) を用いて比較することで感染波に対する人流データの効果を測り、再現度向上に向けてラグの考慮といった追加の施策を実施した。

3 シミュレーション結果

3.1 SEIR,SEIR+ モデルによるシミュレーション結果

各モデルを用いて、第 2 波から第 8 波まで各感染波ごとにシミュレーションを実施した。SEIR モデル,SEIR+ モデルそれぞれの *AIC* の値をまとめたものを表 1 に示す。表 1 から、第 3 波を除いて既存の SEIR モデルよりも SEIR+ モデルの方が *AIC* のスコアが低く、感染波の再現度が高いことが分かる。

既存モデルにおいて感染率は一定の値として扱われるが、SEIR+ モデルは、新型コロナウイルスが持つ一定の感染率 β に人流データを 3 つのカテゴリ (市区町村内、都道府県内、都道府県外) ごとにそれぞれ係数 p_1, p_2, p_3 を導入し、人流の要素を補正項として加える事で、感染率を時間変化するパラメータとしている。実際、感染率は人の接触回数やマスク着用の有無によって動的な変数であり、この部分を考慮したことが SEIR+ モデルが既存のモデルより

も AIC のスコアが改善された結果に繋がったと考えられる。また、図 3 から、第 6 波における人流データの係数 p_n ($n = 1, 2, 3$) を見てみると、 p_2, p_3 が高く、この期間においては特に県外からの人流の効果を感染率に強く反映している事が分かる。この結果から、県外からの感受性者または感染者の増加に伴い、県内での感受性者と感染者の接触率(感染率)及び感染者数が増加すると言う事象が考察される。

次に感染第 6 波に対して SEIR モデル,SEIR+ モデルのシミュレーションを実施した図 2 を見てみると、感染者数のピーク時期において、実際の感染者数とシミュレーションの出力結果にずれがあることが分かる。これは、感染力が強く、免疫逃避の性質を持つオミクロン株が流行した第 6 波では、感染の開始日から感染波のピーク日までが他の波と比べて短いことが原因として考えられる。

表 1 感染波ごとの AIC スコア

AIC 評価基準	SEIR モデル (k=5)	SEIR+ モデル (k=8)
第 2 波	-588.6	-696.4
第 3 波	-650.5	-631.4
第 4 波	-831.0	-839.9
第 5 波	-1012.1	-1038.3
第 6 波	-286.5	-369.3
第 7 波	-1117.2	-1176.6
第 8 波	-876.7	-913.7

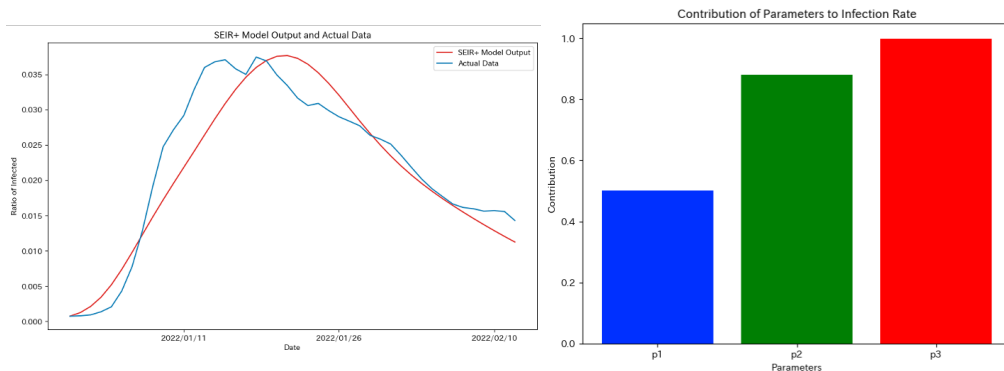


図 2 SEIR+ モデル - 第 6 波

図 3 SEIR+ モデル-第 6 波の p_1, p_2, p_3 の結果

3.2 再現度向上に向けた施策: ラグの追加

3.1 節で先述したように、SEIR+ モデルのシミュレーションでは、感染第 6 波以降の変異株が流行している感染波の再現が難しい事がわかった。この問題を解決するために、シミュレーションの単位時間に合わせて人流データをずらす事で、感染波の開始日より前の時点に位置する人流データを取り入れるようにした。ラグの探索範囲は、0 日から 14 日の間として、AIC が最も低いラグを最適なラグとしてシミュレーションを実施した。各感染波の最適なラグの日数を表 2 に示す。ラグなしのシミュレーションにおいて再現が難しかった第 6 波に対して、最適なラグ (4 日) を加えたシミュレーション結果を図 4 に示す。図 5 には、ラグありでのシミュレーションの際に使用した人流データを示す。図 4 と表 2 の第 6 波における結果から、ラグを入れたことで再現度が改善された事が分かる。改善された要因として、図 5 から分かるように、ラグの追加により感染動態に強い影響を持つ県外からの人流のピークを取り込んだことがあると考えられる。

第 6 波以外の感染波に対しても、最適なラグでのシミュレーションを実行した結果、ラグなし、ラグありの SEIR+ モデルの AIC のスコアを比べると、表 2 の結果になった。表 2 から、どの感染波においても最適なラグの日数が 7 日以下となり、第 3-5 波を除く感染波ではラグを入れたことで AIC のスコアが改善されたことが分かる。この結果から、第 6 波以降の後期の感染波の期間において、変異株によって変化した感染症の伝播パターンや感染から発症までの時間に対応するためには、最大 7 日間のラグを考慮することが重要であると考えられる。

表2 ラグなしとラグありの AIC スコアによる比較表

AIC 評価基準	ラグなし SEIR+ モデル	ラグあり SEIR+ モデル	最適なラグの日数
第 2 波	-696.4	-711.9	4
第 3 波	-650.5	-631.7	0
第 4 波	-839.9	-797.2	3
第 5 波	-1038.3	-1030.5	4
第 6 波	-369.3	-464.3	4
第 7 波	-1176.6	-1207.4	6
第 8 波	-913.7	-974.3	7

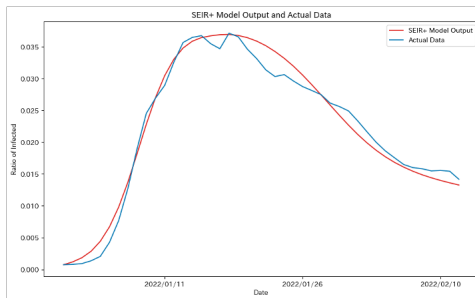


図 4 SEIR+ モデル-第 6 波 (ラグ 4)

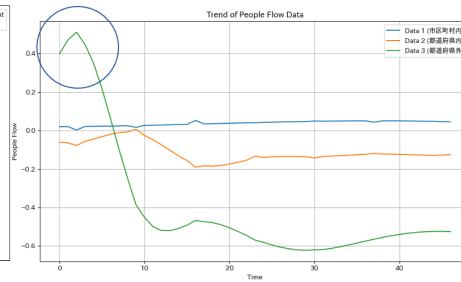


図 5 第 6 波の際の人流データ (ラグあり)

4 結論

本研究では、沖縄県を対象に既存の感染症モデルに人流データを内包した SEIR+ モデルを提案し、新型コロナウイルスの感染動態の解析を目的としたモデリングを行った。既存のモデルと SEIR+ モデルの結果を比較し、既存モデルよりも SEIR+ モデルの方が感染波の再現度が高く、流行初期から中期における感染波を再現するようなシミュレーション結果を得ることができた。また、時間遅れの要素を考慮するためにラグを入れ、感染波の開始日より前にある人流の効果を取り入れるようにシミュレーションを実施した。結果として、変異株が流行していた時期の第 6 波以降の後期の感染波において再現度の改善が確認された。県内の人流よりも県外からの人流の効果の方が新型コロナウイルスの感染者数の増加に影響を与えること、後期の感染波において変異株による感染症の伝播パターンの変化や感染から発症までの時間遅れに対応するには最大 7 日程度のラグの導入が有効である事の 2 点が本研究で示唆された。

5 今後の課題

課題として、今後の感染波や他の感染症に対しても有効なモデルになるよう、感染者数が多い都市部を対象にする、先行研究に倣って都市間の関係性を考慮したモデルの拡張が挙げられる。この拡張によって、より広域に感染症の流行動態を捉えることができると考えられる。新型コロナウイルスの感染に関するモデリング方法は多岐に渡るため、関連した研究を深く調査し、それぞれのモデリングの利点と問題点を検討し、新たなモデリング手法の開発へと役立てる事が重要と言える [5]。

参考文献

- [1] 厚生労働省, “オープンデータ-新規陽性者数の推移 (日別)” 厚生労働省オープンデータ, 最終更新日: 2023/05/08
- [2] 内閣官房デジタル田園都市国家構想実現会議事務局, 内閣府地方創生推進室, “沖縄県のサマリー - 人流 - 沖縄県の滞在人口の動向”, <https://v-resas.go.jp/prefectures/47>.
- [3] Kermack W. O., McKendrick, A. G. “A Contribution to the Mathematical Theory of Epidemics” Vol.115, No. 772, pp. 700 – 721 1927.
- [4] 斎藤正也, 井元清哉, 山口類, 宮野悟, 樋口知之. “連結 SIR モデルを使った日本のインフルエンザ流行における地域間相互作用の推定” 理論応用力学講演会 講演論文集, 63, 2014.
- [5] 徐海燕, 史一華, “数理モデルを用いた感染症のシミュレーションと考察”, 情報処理学会研究報告, 2022