

2021年度 修士論文

論文題目

COVID-19 流行予測のための
免疫獲得者数を考慮したSEIRモデルの拡張

(Extension of SEIR model considering the number of
immunized people for COVID-19 epidemic prediction)

指導教員

舟橋 健司 准教授

伊藤 宏隆 助教

名古屋工業大学大学院 工学研究科 工学専攻 情報工学系プログラム

2020年度入学 32414086番

福田 太一

目次

第 1 章	はじめに	1
第 2 章	SEIR モデルの概要	3
第 3 章	免疫獲得者数を考慮した拡張 SEIR モデルの提案	5
3.1	重症者を考慮した SEIR モデル	5
3.2	自然免疫を考慮した SEIR モデル	6
3.3	自然免疫とワクチンによる免疫を考慮した SEIR モデル	7
3.4	重症者, 自然免疫, ワクチンによる免疫を考慮した SEIR モデル	8
第 4 章	SEIR モデルのパラメータ推定法	9
4.1	過去の最適パラメータの導出	10
4.1.1	乱数で疑似的に導出	10
4.1.2	勾配法を用いた最適化アルゴリズムによる導出	10
4.2	未来の最適パラメータの推定	13
4.2.1	ホルトウィンタース法	14
4.2.2	カルマンフィルタ	14
第 5 章	SEIR モデルおよび拡張 SEIR モデルを用いた検証実験	21
5.1	実験概要	22
5.1.1	過去の最適パラメータの導出	23
5.1.2	未来の最適パラメータの推定	24
5.1.3	実験環境	25
5.2	実験結果	25
5.2.1	ワクチン接種が進んでいない時期 (2020 年 12 月 ~ 2021 年 4 月)	25
5.2.2	ワクチン接種の影響を考慮する時期 (2021 年 8 月 ~ 2021 年 12 月)	31
第 6 章	拡張 SEIR モデルを用いた感染者数の傾向予測	37
第 7 章	むすび	44
	謝辞	45
	参考文献	46
	発表論文リスト	49

第1章 はじめに

2019年に12月に中国・武漢市にて、新型コロナウイルス感染症 (COVID-19) が新たに検出された [1]. それ以降世界中に感染が拡大するパンデミックが発生し, 2021年現在でも日本をはじめとした世界各国において流行が続いている. これまで日本では緊急事態宣言・蔓延防止重点措置などの様々な政策と規制や複数にわたるワクチン接種によって, 感染の規模を幾度か抑えることには成功しているが未だに収束の兆しは見えない. これらの試みによって国内の COVID-19 の感染を完全に終わらせることは難しい. しかし, その影響力を最小限に留めることは可能であることが言える. 無論, COVID-19 対策は日別の感染者が増加傾向になる前に意思決定し, 実行に移すことによって高い効果を発揮するもののワクチン接種や特効薬のような科学的なアプローチは医学の研究の進捗度に依存するため実施は不定期であり, 緊急事態宣言を含む政策と規則によるアプローチは感染者が増加傾向にあることをデータで確認してから敷かれることが多い. 特に後者は情報技術などを用いて感染規模の拡大を事前に予測することで従来よりも早いタイミングでの意思決定を可能にできる. 従って, 感染規模を事前に予測することは COVID-19 の被害の最小化に寄与すると言える.

そこで, 本研究では従来感染症の流行予測に有効とされる SEIR モデル [2] と呼ばれる数理モデルの1つを用いた日本における COVID-19 の感染拡大の規模を予測する手法を模索した. SEIR モデルで高い予測精度を出すためにはこのモデルのパラメータをどのように推定すればよいか, SEIR モデルを時期に応じてどのようなシチュエーションを仮定して構築すればよいかに焦点を当てて議論を進める. そこで得られた結論をもとに, 日本国内の感染拡大の予測に有効な SEIR モデルについて考察する.

SEIR モデルを用いた関連研究について, 2020年1月の COVID-19 の感染拡大シ

ミュレーションを行ったものが存在する [3]. シミュレーションの結果, 中国の地域の中でも武漢での新型コロナウイルス (SARS-Cov2) の感染規模が最も大きくなるであろうと推測された. また 2020 年の春のインドにおける予測シミュレーションを行った研究 [4] も行われており, そこでは線形回帰モデルと SEIR モデルとの精度比較の結果後者の方が優れていることが示されていた. スペインとイタリア, パキスタンとマレーシアにおけるシミュレーション [5, 6] においてもパラメータの調整次第で高い精度が確認されている. 上述の研究の共通点として既知である過去のデータとモデルの出力との比較において, 誤差を最小化するパラメータが導出されており, 未知のデータ, 即ち未来の感染拡大を予測する試みは行われていない.

2章で後述する拡張 SEIR モデルをベースに未来の COVID-19 の流行を予測する研究も行われている [7, 8, 9]. しかし, これらにおける重篤患者・集中治療室の患者 [7], 出生による総人口の変化 [8, 9] といった拡張要素が元々の SEIR モデルと比較してどれだけ精度向上への寄与するかに関する記述はない. そこで, 未来の COVID-19 の流行予測に必要な SEIR モデルを拡張するにあたって拡張する因子はどのようなものを選択すべきかを検証する必要があると考えた. 本研究では基本的な SEIR モデルと複数の拡張 SEIR モデルを用意し, それぞれの精度比較をもって予測精度向上に必要な要素について考察する. 各 SEIR モデルに投入するパラメータを過去のデータを利用して推定し, その結果自然免疫とワクチンによって獲得される免疫が精度向上に有効であった. また, それらを考慮した拡張 SEIR モデルの予測結果によって一部の時期に限り感染者数の傾向を予測した.

本論文では第2章で SEIR モデルの簡単な概要, 第3章では本研究で精度の比較の為に用意した複数の拡張 SEIR モデルの説明, 第4章では高い精度で COVID-19 の流行を予測するための SEIR モデルのパラメータ推定法について, 第5章では複数の拡張 SEIR モデルの精度検証実験についての説明及び実験結果と SEIR モデルに取り入れるべき要素の考察, 第6章では拡張 SEIR モデルの予測結果を用いた未来の新規感染者の傾向の予測, 第7章で本研究の総括と更なる精度向上に向けた今後の課題について述べる.

第2章 SEIRモデルの概要

ある地域にて感染症が流行した場合、人口は基本的に次の通りに分類される。

- Susceptible(S): ウィルスに感染する可能性のある人口
- Exposed(E): ウィルスに感染したが症状を発していない人口
- Infected(I): ウィルスに感染し症状を発した人口
- Recovered(R): 症状から回復または死亡した人口

また、各人口の1日当たりの変化は次の常微分方程式にて表現される。 [2].

$$\frac{dS}{dt} = -\frac{R_0}{e} * \frac{I(t)}{N} S(t) \quad (2.1)$$

$$\frac{dE}{dt} = \frac{R_0}{e} * \frac{I(t)}{N} S(t) - \frac{1}{e} * E(t) \quad (2.2)$$

$$\frac{dI}{dt} = \frac{1}{e} * E(t) - \frac{1}{l} * I(t) \quad (2.3)$$

$$\frac{dR}{dt} = \frac{1}{l} * I(t) \quad (2.4)$$

R_0 : 基本再生産数, e : 平均潜伏期間, l : 平均発症期間, N : その地域の総人口

SEIRモデルはこれらの常微分方程式を積分演算することで、図2.1のように各人口をSusceptibleから順にボトムアップで予測する工程を繰り返し、時系列的变化を逐次的に予測する。 R_0 は「基本再生産数」と呼ばれるパラメータで、感染者がウィルスに対して免疫を持たない人口に対して直接感染させる人数の期待値を表した数値である。 e は感染者の体内にウィルスが潜伏する期間の期待値、 l は症状を発している期間の期待値を表すパラメータである。 $S(t)$ の係数にあたる $-\frac{R_0}{e} * \frac{I(t)}{N}$ は1日当たりにSusceptibleからExposedに遷移する人口の割合を表しており、同様に e^{-1} , l^{-1} は

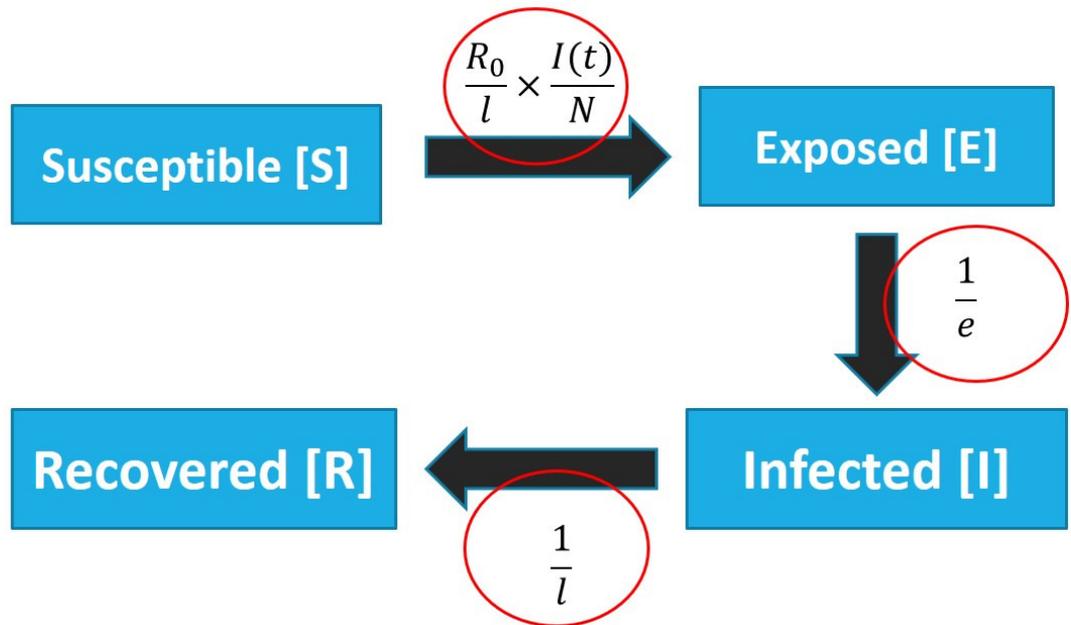


図 2.1: SEIR モデルの構造

それぞれ Exposed から Infected へ, Infected から Recovered への遷移割合を示す. 加えて N は $N = S(t) + E(t) + I(t) + R(t)$ を満たすと同時にモデルの各常微分方程式の総和が 0 に等しいため, SEIR モデルはその地域の総人口が時間の経過とともに変化しないという仮定の下で計算を行うという性質を持っている. また, 上述の SEIR モデルの基本的な概念に加え, ワクチン, 自然免疫, 重篤患者といった感染症に関わるその他の要因を考慮しパラメータや予測する人口を追加した拡張 SEIR モデルも存在する. 本研究で用意した拡張 SEIR モデルについては次章で述べる.

第3章 免疫獲得者数を考慮した拡張SEIRモデルの提案

本研究では通常のSEIRモデルに加え次の4種類の拡張SEIRモデルを用意した.

3.1 重症者を考慮したSEIRモデル

感染症流行下では一般的に Infected に属する人口の持つ症状が重症化する場合も考えられる. このモデルでは通常のSEIRモデルで扱う人口に重症者患者を表す人口である Severely を新たに追加した (図 3.1). 構造式は以下の通りになる.

$$\frac{dS}{dt} = -\frac{R_0}{e} * \frac{I(t)}{N} S(t) \quad (3.1)$$

$$\frac{dE}{dt} = \frac{R_0}{e} * \frac{I(t)}{N} S(t) - \frac{1}{e} * E(t) \quad (3.2)$$

$$\frac{dI}{dt} = \frac{1}{e} * E(t) - \frac{1}{l} * I(t) - v * I(t) \quad (3.3)$$

$$\frac{dV}{dt} = v * I(t) - v_r * V(t) \quad (3.4)$$

$$\frac{dR}{dt} = \frac{1}{l} * I(t) \quad (3.5)$$

Severely の追加に付随して新たなパラメータ v, v_r を導入した. これらはそれぞれ Infected から Severely へ, Severely から Recovered へ遷移する確率を表している.

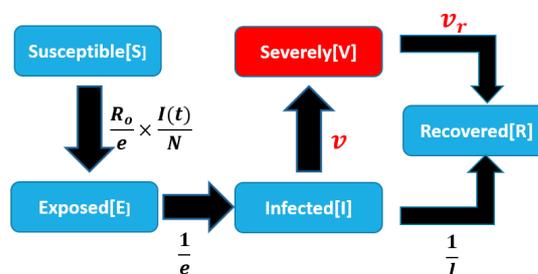


図 3.1: 重症者を考慮した SEIR モデル

3.2 自然免疫を考慮した SEIR モデル

人間には一度体内に侵入した病原菌の情報を記憶する自然免疫という機能が働くとされている。COVID-19 流行下の現在においては一度症状を克服したのち SARS-CoV-2 に対して自然免疫を獲得する場合も考えられる。従って、このモデルでは免疫を獲得した人口である Quarantined を新たに追加した (図 3.2)。構造式は以下の通りになる。

$$\frac{dS}{dt} = -\frac{R_0}{e} * \frac{I(t)}{N} S(t) \quad (3.6)$$

$$\frac{dE}{dt} = \frac{R_0}{e} * \frac{I(t)}{N} S(t) - \frac{1}{e} * E(t) \quad (3.7)$$

$$\frac{dI}{dt} = \frac{1}{e} * E(t) - \frac{1}{l} * I(t) \quad (3.8)$$

$$\frac{dR}{dt} = \frac{1}{l} * I(t) - \alpha_r * R(t) \quad (3.9)$$

$$\frac{dQ}{dt} = \alpha_r * R(t) \quad (3.10)$$

Quarantined の追加に付随する新たなパラメータに α_r を導入した。これは Recovered から Quarantined に遷移する確率を表している。なお、本研究では Quarantined に属する人口がウィルスに感染するブレイクスルー感染は起こらないものとする。

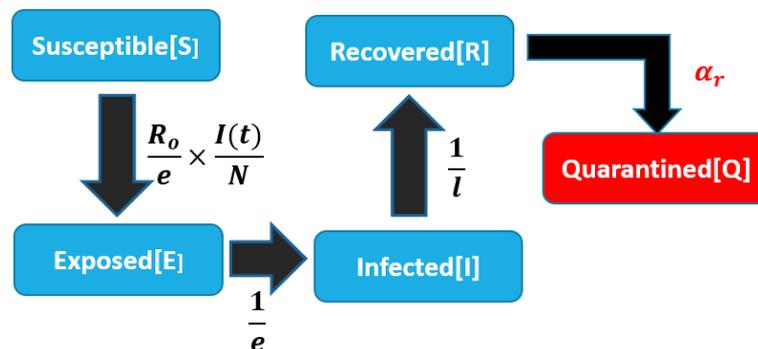


図 3.2: 自然免疫を考慮した SEIR モデル

3.3 自然免疫とワクチンによる免疫を考慮した SEIR モデル

このモデルでは先に述べた自然免疫に加えワクチンによって獲得した免疫の影響を考慮している (図 3.3). ここではワクチンを接種するのは Susceptible に属する人口に限定する. 構造式は以下の通りになる.

$$\frac{dS}{dt} = -\frac{R_0}{e} * \frac{I(t)}{N} S(t) - \alpha_s * S(t) \quad (3.11)$$

$$\frac{dE}{dt} = \frac{R_0}{e} * \frac{I(t)}{N} S(t) - \frac{1}{e} * E(t) \quad (3.12)$$

$$\frac{dI}{dt} = \frac{1}{e} * E(t) - \frac{1}{l} * I(t) \quad (3.13)$$

$$\frac{dR}{dt} = \frac{1}{l} * I(t) - \alpha_r * R(t) \quad (3.14)$$

$$\frac{dQ}{dt} = \alpha_r * R(t) + \alpha_s * S(t) \quad (3.15)$$

新たに導入した α_s は Susceptible から Quarantined への遷移確率を表している.

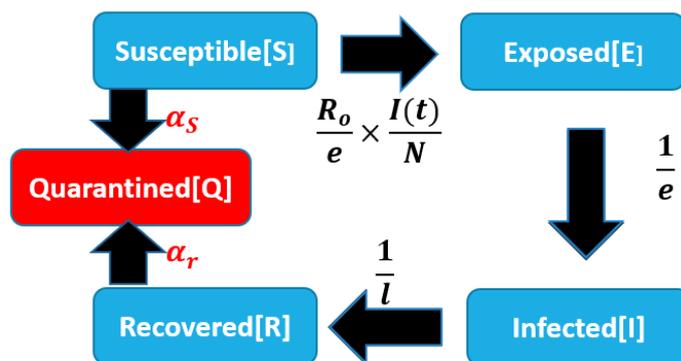


図 3.3: 自然免疫とワクチンによる免疫を考慮した SEIR モデル

3.4 重症者, 自然免疫, ワクチンによる免疫を考慮した SEIR モデル

このモデルは重症者, 自然免疫, ワクチンによる免疫全てを考慮した SEIR モデルとなっている (図 3.4). その構造式は以下の通りになる.

$$\frac{dS}{dt} = -\frac{R_0}{e} * \frac{I(t)}{N} S(t) - \alpha_s * S(t) \quad (3.16)$$

$$\frac{dE}{dt} = \frac{R_0}{e} * \frac{I(t)}{N} S(t) - \frac{1}{e} * E(t) \quad (3.17)$$

$$\frac{dI}{dt} = \frac{1}{e} * E(t) - \frac{1}{l} * I(t) - v * I(t) \quad (3.18)$$

$$\frac{dV}{dt} = v * I(t) - v_r * V(t) \quad (3.19)$$

$$\frac{dR}{dt} = \frac{1}{l} * I(t) + v_r * V(t) - \alpha_r * R(t) \quad (3.20)$$

$$\frac{dQ}{dt} = \alpha_r * R(t) + \alpha_s * S(t) \quad (3.21)$$

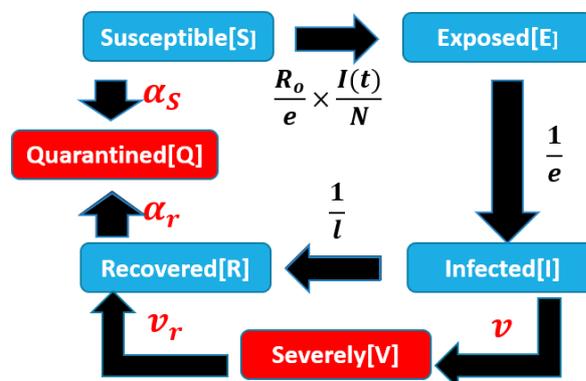


図 3.4: 重症者, 自然免疫, ワクチンによる免疫を考慮した SEIR モデル

第4章 SEIRモデルのパラメータ推定法

本研究ではSEIRモデルを用いて未来のCOVID-19の流行を予測するにあたって、次の三つの手順を踏む。

- Step1 過去の最適パラメータを導出する。
- Step2 過去の最適パラメータから未来の最適パラメータを逐次的に推定。
- Step3 未来の最適パラメータをSEIRモデルに投入。

まず本研究では、予測開始日のインデックスを i としたときのSEIRモデルの最適パラメータ P_i を以下の通りに定義する。

$$P_i = \arg \min_{\theta} [\text{Recoveredの} RMSE(i)]$$

$$RMSE(i) = \sqrt{\frac{1}{n-i+1} \sum_{t=i}^n (y_t - \hat{y}_t)^2}$$

θ : パラメータの集合, y_t : 観測値, \hat{y}_t : モデルの予測値

先行研究 [7] ではSEIRモデルのパラメータ推定を地域の平均収入や健康管理の徹底度などの国勢調査のデータを共変量とした回帰モデルによって行っていた。日本国内ではそのようなデータは容易に入手できないことから、本研究では P_i を導出するにあたってSEIRモデルの予測値と [10] から入手できるRecoveredの観測値とのRMSE(最小二乗平方根誤差)の最小化を目標とする。Step1では流行を予測したい時期よりも過去の時期から予測シミュレーションを開始し、その予測値と既知である過去のデータとのRMSEを最小にするパラメータ、「過去の最適パラメータ」を導出する。Step2では、未来の流行をSEIRモデルで予測した場合にその時は未知だが

後に得られる観測値, すなわち未来のデータと予測値の RMSE が最も小さくなるであろうパラメータ, 「未来の最適パラメータ」を時系列分析を用いて推定する. 最後に Step3 では Step2 で得られたパラメータを SEIR モデルに投入し予測値を算出する. この章では Step1 と Step2 についてそれらの手順の詳細について述べる.

4.1 過去の最適パラメータの導出

SEIR モデルを用いて予測する期間の長さを S (日) とした時, 予測値のサンプル数は S と等しい. この場合, RMSE を算出する際に参照される過去のデータのインデックスは $i \sim i + S$ になる. 過去のデータの長さを T とすると, 前述の範囲の最大値が T を越えない予測開始日 i の範囲は $0 \leq i \leq T - S$ となる. 従って, ここで求める過去の最適パラメータ P_{past} は次の通りになる.

$$P_{past} = (P_1, P_2, \dots, P_{T-S})$$

次に, P_{past} の各要素を一つずつ導出するにあたり次の2通りの方法を試している.

4.1.1 乱数で疑似的に導出

一様分布によって生成されたパラメータの組み合わせを一つずつ SEIR モデルに投入し, 最も RMSE の小さかったものを採用する. 本研究におけるサンプル数は 10000 通り用意した.

4.1.2 勾配法を用いた最適化アルゴリズムによる導出

パラメータ θ に従う目的関数 $J(\theta)$ が存在するときに, $J(\theta)$ が最小化に向かう降下方向に θ を更新する. この時, $J(\theta)$ の降下方向は $-\nabla J(\theta)$ で与えられ, パラメータの更新は次の通りに行われる.

$$\theta \leftarrow \theta - \alpha * \nabla J(\theta_t) \tag{4.1}$$

α : 学習率

α は「学習率」と呼ばれるパラメータで、一度の更新におけるパラメータの変化の大きさを表している。 P_{past} の導出においては RMSE の最小化を目標としているため、本研究で設定した目的関数 $J(\theta)$ を

$$J(\theta) = \frac{1}{2n} \sum_{t=i}^n (y_t - f_t(\theta))^2 \quad (4.2)$$

y_t : 観測値, $f_t(\theta)$: パラメータ θ に従う SEIR モデルの予測値

とした場合を考える。 $J(\theta)$ の降下方向 $-\nabla J(\theta)$ は $-J(\theta)$ の一次微分によって導出することができるため、

$$\begin{aligned} -\nabla J(\theta) &= -\frac{dJ(\theta)}{d\theta} \\ &= \frac{1}{n} \sum_{t=i}^n (y_t - f_t(\theta)) * \frac{df_t(\theta)}{d\theta} \end{aligned} \quad (4.3)$$

となる。また $\frac{df_t(\theta)}{d\theta}$ について SEIR モデルの出力を示す数式 (2.1) ~ (2.4) の微分が難しいことを考慮し、本研究は数値微分によって近似的に導出する。

$$\begin{aligned} \frac{df_t(\theta)}{d\theta} &= \lim_{h \rightarrow 0} \frac{f_t(\theta + h) - f_t(\theta)}{h} \\ &\doteq \frac{f_t(\theta + \epsilon) - f_t(\theta)}{\epsilon} \end{aligned} \quad (4.4)$$

$\epsilon : 1.0 * 10^{-10}$ などの極小値

上述の方法を繰り返すことによって、任意の θ とその制約 S 対して

$$J(\theta^*) \leq J(\theta) \quad (4.5)$$

$$\theta, \theta^* \in S \quad (4.6)$$

を満たす最適解 θ^* を求める方法を「最急降下法」 [11] と呼ぶ。

最急降下法には局所最適解に陥りやすいという欠点が存在する。局所最適解とは、点 θ の ϵ 近傍を

$$B(\theta, \epsilon) = \{y \in \mathbb{R} \mid \|\theta - y\| < \epsilon\} \quad (4.7)$$

と定義したときに、 $\epsilon > 0$ について

$$\theta \in B(\theta^*, \epsilon) \cap S \text{ ならば } J(\theta^*) \leq J(\theta) \quad (4.8)$$

を満たす θ^* の集合($\theta^* \in S$)を表しており、 $J(\theta^*)$ を極小値と呼ぶ。最適解ならば局所最適解になる。しかし、その逆は一般的に成立しない。局所最適解となる θ^* において、コスト関数の勾配 $\nabla J(\theta^*)$ が0になることが知られている[11, 12]。最急降下法においても、 θ の初期値によっては局所最適解付近で $\nabla J(\theta)$ が徐々に0に近づくにつれ θ の更新が停滞してしまうことがある。加えて、学習率 α が決め打ちになる関係上この値が必要以上に小さい場合は局所最適解にたどりつけず、必要以上に大きい値を設定した場合は θ が局所最適解付近を振動してしまうこともあり、適切な α の見極めが難しい欠点も挙げられる。これらの欠点を克服するために、最急降下法を応用したRMSProp[13]、Adagrad[14]などの最適化アルゴリズムが開発されている。

本研究では、そのようなアルゴリズムの中でもAdam[15]を採用した。パラメータの更新は次の通りに行われる。

$$\beta_1, \beta_2 \in [0, 1) \quad (4.9)$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla J(\theta) \quad (4.10)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla J(\theta))^2 \quad (4.11)$$

$$\hat{m}_t = \frac{m_t}{(1 - \beta_1)} \quad (4.12)$$

$$\hat{v}_t = \frac{v_t}{(1 - \beta_2)} \quad (4.13)$$

$$\theta \leftarrow \theta - \alpha * \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (4.14)$$

t : 更新を繰り返した回数 (*iteration*)

ϵ : $1.0 * 10^{-10}$ などの極小値

このアルゴリズムの特徴はパラメータの勾配 $\nabla J(\theta)$ の移動平均 m_t, v_t を用いることで勾配の振動を抑制するAdagradと、これらを利用してパラメータ更新の度に学習率を動的に変化させるRMSPropの特性を併せ持つ点にある。これにより、最急降下法よりも α の初期値に対する依存度の小さい安定したパラメータ更新を行うことができると考え、Adamの採用に至った。

また、局所最適解に陥りやすい欠点に対しては、 θ の初期値の組み合わせを一様分布による乱数で複数用意した上でパラメータ更新を行い、最終的に $J(\theta)$ が最も小さくなったものを一時的な解とするといった改善策を試行している。

4.2 未来の最適パラメータの推定

未来の流行を予測するにあたって、予測開始日を過去のデータの示す最終日の次の日 ($i = T + 1$) に設定した場合を仮定する. 過去の最適パラメータを導出する過程で得られるのは予測開始日 $0 \leq i \leq T - S$ までの最適パラメータであるため、ここでは $T - S + 1 \leq i \leq T + 1$ の最適パラメータ

$$P_{past}' = (P_{T-S+1}, P_{T-S+2}, \dots, P_{T+1}) \quad (4.15)$$

を推定する必要がある. この時、過去のデータはこれ以上参照できないため、 P_{past} と同じ導出法を用いることはできない. 従って、 P_{past} の時系列的变化に着目して、これを利用した時系列モデルによる P_{past}' を図 4.1 のように推定した.

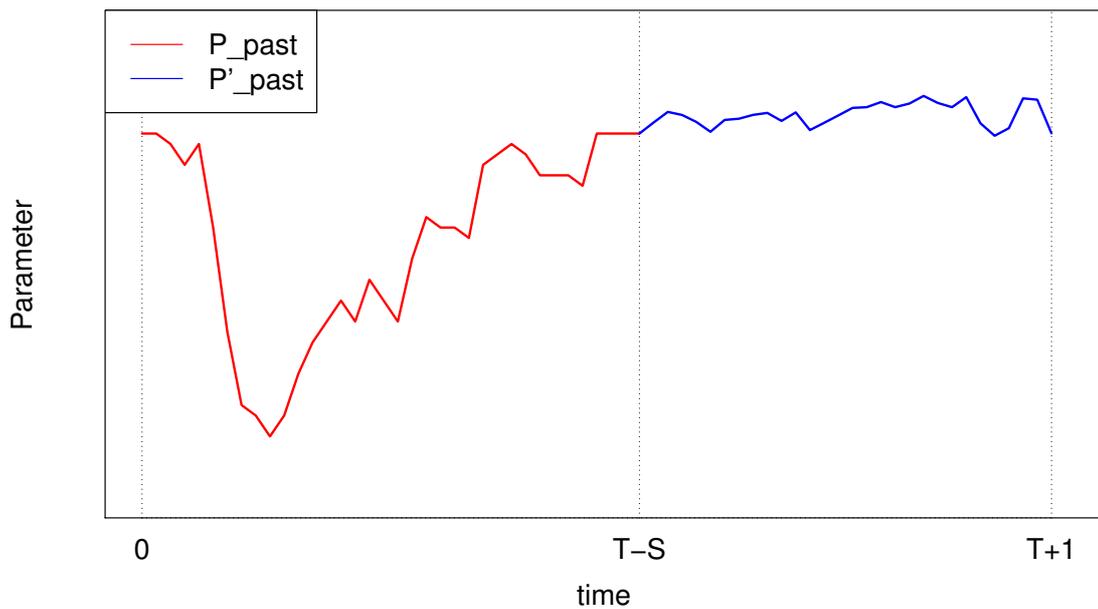


図 4.1: P_{past} の時系列変化を利用した P_{past}' の推定の例

4.2.1 ホルトウィンターズ法

ホルトウィンターズ法 [16] は時系列データを構成する3つの要素を利用して未来の時系列の予測値を逐次的に推定するモデルを構築する. 1つ目の要素は *trend* と呼ばれるものでこれは時系列データの上昇傾向または下降傾向, 2つ目は *season* で, 時系列データが反復する季節変動, 3つ目の要素 *level* は *trend* に依存しない時系列データの期待値をそれぞれ意味している. ある時刻 t における観測値 y_t とすると, そのデータが周期 L で反復している場合の m ステップ後の未来の期待値 \hat{y}_{t+m} は

$$\hat{y}_{t+m} = level_t + m * trend_t + season_{t-L+1+(m-1)modL} \quad (4.16)$$

$$level_t = \alpha(y_t - season_{t-L}) + (1 - \alpha)(level_{t-1} + trend_{t-1}) \quad (4.17)$$

$$trend_t = \beta(level_t - level_{t-1}) + (1 - \beta)trend_{t-1} \quad (4.18)$$

$$season_t = \gamma(y_t - level_t) + (1 - \gamma) \quad (4.19)$$

で表される. *level*, *trend*, *season* の初期値については導出が複雑なためここでは割愛する. α, β, γ は SSE の最小化問題をネルダーミード法 [17] で解いて導出される.

$$SSE = \sum_{t=i}^n (y_t - \hat{y}_t)^2 \quad (4.20)$$

y_t : 観測値, \hat{y}_t : 予測値

4.2.2 カルマンフィルタ

カルマンフィルタは, 状態空間モデルと呼ばれる時系列モデルの一種である. ここで言う「状態」とは, 直接的に観測される観測値とは逆の潜在的な確率変数を意味している [18]. 状態空間モデルにおいて, 状態に関してはマルコフ性に従い, 観測値に関してはその時点における状態によってのみ決まるとそれぞれ仮定する. ある時系列 t における状態 x_t , 観測値 y_t は次の方程式によって表現される.

$$x_t = G_t x_{t-1} + w_t \quad w_t \sim \mathcal{N}(0, W_t) \quad (4.21)$$

$$y_t = F_t x_t + v_t \quad v_t \sim \mathcal{N}(0, V_t) \quad (4.22)$$

$G_t : p \times p$ の状態遷移行列, $F_t : 1 \times p$ 観測行列, $w_t : 状態雑音, v_t, 観測雑音$
 $W_t : w_t$ の $p \times p$ の共分散行列, $V_t : v_t$ の共分散行列

状態の初期値 x_0 は p 次元のベクトル $m_0, p \times p$ の共分散行列 C_0 の正規分布に従う. すなわち, $x_0 \sim \mathcal{N}(m_0, C_0)$ である. カルマンフィルタを用いて未来の期待値 ($y_T (T > t)$) を予測するにあたって, $1 < T \leq t$ における観測値に対する状態の条件付き分布であるフィルタリング分布 $p(x_t | y_{1:t})$ を推定する必要がある, この過程をカルマンフィルタリングと呼ぶ. また, カルマンフィルタリングに必要な一期先予測分布 $p(x_t | y_{1:t-1})$, 一期先予測尤度 $p(y_t | y_{1:t-1})$ は次の通りになる.

フィルタリング分布 $p(x_t | y_{1:t}) \sim \mathcal{N}(m_t, C_t)$: 平均ベクトル m_t , 共分散行列 C_t
 一期先予測分布 $p(x_t | y_{1:t-1}) \sim \mathcal{N}(\alpha_t, R_t)$: 平均ベクトル α_t , 共分散行列 R_t
 一期先予測尤度 $p(y_t | y_{1:t-1}) \sim \mathcal{N}(f_t, Q_t)$: 平均値 f_t , 共分散行列 Q_t

カルマンフィルタリングを行う手続きについて考える. 一期先予測分布について, その平均ベクトル α_t は

$$\alpha_t = E[x_t | y_{1:t-1}] \quad (4.23)$$

となり, 反復期待値の法則とマルコフ性より, 以下が成り立つ.

$$\alpha_t = E[E[x_t | x_{t-1}, y_{1:t-1}] | y_{1:t-1}] \quad (4.24)$$

$$= E[E[x_t | x_{t-1}] | y_{1:t-1}] \quad (4.25)$$

状態方程式において, $p(x_t | x_{t-1}) = \mathcal{N}(G_t x_{t-1}, W_t)$ が成り立つため

$$= E[G_t x_{t-1} | y_{1:t-1}] \quad (4.26)$$

$$= G_t * E[x_{t-1} | y_{1:t-1}] \quad (4.27)$$

$$= G_t m_{t-1} \quad (4.28)$$

共分散行列 R_t について,

$$R_t = \text{Var}[x_t | y_{1:t-1}] \quad (4.29)$$

となり, 総分散の法則を用いて

$$= E[\text{Var}[x_t|x_{t-1}, y_{1:t-1}]|y_{1:t-1}] + \text{Var}[E[x_t|x_{t-1}, y_{1:t-1}]|y_{1:t-1}] \quad (4.30)$$

$$= E[\text{Var}[x_t|x_{t-1}]|y_{1:t-1}] + \text{Var}[E[x_t|x_{t-1}]|y_{1:t-1}] \quad (4.31)$$

α_t と同様に

$$= E[W_t|y_{1:t-1}] + \text{Var}[G_t x_{t-1}|y_t] \quad (4.32)$$

$$= W_t + G_t C_{t-1} G_t^\top \quad (4.33)$$

次に一期先予測尤度について, 平均ベクトル f_t は

$$f_t = E[y_t|y_{t-1}] \quad (4.34)$$

反復期待値の法則とマルコフ性より,

$$= E[E[y_t|x_t, y_{1:t-1}]|y_{1:t-1}] \quad (4.35)$$

$$= E[E[y_t|x_t]|y_{1:t-1}] \quad (4.36)$$

状態方程式において, $p(y_t|x_t) = \mathcal{N}(F_t x_t, V_t)$ が成り立つため

$$= E[F_t x_t|y_{1:t-1}] \quad (4.37)$$

$$= F_t * E[x_t|y_{1:t-1}] \quad (4.38)$$

$$= F_t \alpha_t \quad (4.39)$$

共分散行列 Q_t について,

$$Q_t = \text{Var}[y_t|y_{t-1}] \quad (4.40)$$

総分散の法則を用いて

$$= E[\text{Var}[y_t|x_{t-1}, y_{1:t-1}]|y_{1:t-1}] + \text{Var}[E[y_t|x_{t-1}, y_{1:t-1}]|y_{1:t-1}] \quad (4.41)$$

$$= E[\text{Var}[y_t|x_t]|y_{1:t-1}] + \text{Var}[E[y_t|x_t]|y_{1:t-1}] \quad (4.42)$$

f_t 同様に

$$= E[V_t|y_{1:t-1}] + \text{Var}[F_t x_t|y_{1:t-1}] \quad (4.43)$$

$$= V_t + F_t R_{t-1} F_t^\top \quad (4.44)$$

最後にフィルタリング分布について考える. ベイズの定理を用いて

$$p(x_t|y_{1:t}) = p(x_t|y_{1:t-1}) \frac{p(y_t|x_t)}{p(y_t|y_{1:t-1})} \quad (4.45)$$

$$\propto p(y_t|x_t)p(x_t|y_{1:t-1}) \quad (4.46)$$

$$= \mathcal{N}(F_t x_t, V_t) \mathcal{N}(\alpha_t, R_t) \quad (4.47)$$

この時のフィルタリング分布の平均ベクトル m_t , 共分散行列 C_t について, ノイズ ϵ が正規分布に従う次のような一般の線形回帰モデル

$$y = X\beta + \epsilon, \quad \epsilon \sim \mathcal{N}(0, V) \quad (4.48)$$

に対して, $p(\beta|y) \propto \mathcal{N}(X\beta, V)\mathcal{N}(m_0, C_0)$ を満たす $p(\beta|y)$ の平均ベクトル m , 共分散行列 C は次の通りに導かれることが知られている.

$$m = m_0 + C_0 X^\top (X C_0 X^\top + V)^{-1} (y - X m_0) \quad (4.49)$$

$$C = C_0 - C_0 X^\top (X C_0 X^\top + V)^{-1} X C_0 \quad (4.50)$$

これをフィルタリング分布 $p(x_t|y_{1:t})$ に当てはめた場合, 式 (3.39), (3.43) を用いて

$$m_t = \alpha_t + R_t F_t^\top (F_t R_t F_t^\top + V_t)^{-1} (y_t - F_t \alpha_t) \quad (4.51)$$

$$= \alpha_t + R_t F_t^\top Q_t^{-1} [y_t - f_t] \quad (4.52)$$

$$C_t = R_t - R_t F_t^\top (F_t R_t F_t^\top + V_t)^{-1} F_t R_t \quad (4.53)$$

$$= [I - R_t F_t^\top Q_t^{-1} F_t] R_t \quad (4.54)$$

となる.

上記の導出法に基づくカルマンフィルタリングの疑似コードを図 4.2 に示す. ここではカルマン利得と呼ばれる $p \times p$ の行列が出現する. これは, 観測値と一期先予測尤度の平均ベクトルとの差 $y_t - f_t$ をフィルタリング分布にどれだけ反映させるかの重みを表している. また, この状態空間モデルのパラメータを $\theta = \{G_t, F_t, W_t, V_t, m_0, C_0\}$ と置く時, その対数尤度 $l(\theta)$ は

$$l(\theta) = \sum_{i=1}^t \log p(y_i|y_{1:i-1}, \theta) \quad (4.55)$$

$$:= -\frac{1}{2} \sum_{i=1}^t t \log |Q_t| - \frac{1}{2} (y_t - f_t)^2 / Q_t \quad (4.56)$$

Require: 時点 $t-1$ におけるフィルタリング分布: m_{t-1}, C_{t-1}

```

1: for  $T = 1$  to  $t$  do
2:   { • 一期先予測分布 }
3:    $\alpha_t \leftarrow G_t m_{t-1}$ 
4:    $R_t \leftarrow G_t C_{t-1} G_t^\top + W_t$ 
5:   { • 一期先予測尤度 }
6:    $f_t \leftarrow F_t \alpha$ 
7:    $Q_t \leftarrow F_t R_{t-1} F_t^\top + V_t$ 
8:   { • カルマン利得 }
9:    $K_t \leftarrow R_t F_t^\top Q^{-1}$ 
10:  { • 状態の更新 }
11:   $m_t \leftarrow K_t [y_t - f_t]$ 
12:   $C_t \leftarrow [I - K_t F_t] R_t$ 
13: end for
14: return  $m = \{m_0, m_1, \dots, m_t\}, C = \{C_0, C_1, \dots, C_t\}$ 

```

図 4.2: カルマンフィルタの疑似コード

となり [19], パラメータ θ は $l(\theta)$ を最大にする最尤推定によって補正されることになる. 次に観測値の k 期先の期待値 y_{t+k} を推定するカルマン予測について述べる. k 期先予測分布もフィルタリング分布同様に正規分布に従う.

k 期先予測分布 $\mathcal{N}(a_t(k), R_t(k))$: 平均ベクトル $\alpha_t(k-1)$, 共分散行列 $R_t(k-1)$

ここで, $k=1$ から $k=s$ ($s \geq 1$) における時点 $t+k$ 期先予測分布を求めることを考える. $\alpha_t(k)$ について, 反復期待値の法則の法則とマルコフ性より以下が成り立つ.

$$\alpha_t(k) = E[x_{t+k} | y_{1:t}] \quad (4.57)$$

$$= E[E[x_{t+k} | x_{t+k-1}, y_{1:t}] | y_{1:t}] \quad (4.58)$$

$$= E[E[x_{t+k} | x_{t+k-1}] | y_{1:t}] \quad (4.59)$$

また, k 期先予測分布 $p(x_{t+k} | x_{t+k-1})$ は正規分布 $\mathcal{N}(G_{t+k} x_{t+k-1}, W_{t+k})$ に従うため, $\alpha_t(k)$ は

$$= E[G_{t+k} x_{t+k-1} | y_{1:t}] \quad (4.60)$$

$$= G_{t+k} * \alpha_t(k-1) \quad (4.61)$$

Require: 時点 $t+k-1$ における $k-1$ 期先予測分布: $\alpha_t(k-1), R_t(k-1)$

- 1: **for** $k=1$ to $s(s \leq 1)$ **do**
- 2: { \cdot k 期先予測分布 }
- 3: $\alpha_t(k) \leftarrow G_{t+k}\alpha_t(k-1)$
- 4: $R_t(k) \leftarrow G_{t+k}R_t(k-1)G_{t+k}^\top + W_{t+k}$
- 5: **end for**
- 6: **return** $m = \{m_0, m_1, \dots, m_t\}, C = \{C_0, C_1, \dots, C_t\}$

図 4.3: カルマン予測の疑似コード

となる. $R_t(k)$ について, 総分散の法則とマルコフ性より以下が成り立つ.

$$R_t(k) = \text{Var}[x_{t+k}|y_{1:t}] \quad (4.62)$$

$$\begin{aligned} &= E[\text{Var}[x_{t+k}|x_{t+k-1}, y_{1:t}]|y_{1:t}] \\ &\quad + \text{Var}[E[x_{t+k}|x_{t+k-1}, y_{1:t}]|y_{1:t}] \end{aligned} \quad (4.63)$$

$$= E[\text{Var}[x_{t+k}|x_{t+k-1}]|y_{1:t}] + \text{Var}[E[x_{t+k}|x_{t+k-1}]|y_{1:t}] \quad (4.64)$$

$\alpha_t(k)$ 同様に $R_t(k)$ は以下の通りになる.

$$= E[W_{t+k}|y_{1:t}] + \text{Var}[G_{t+k}x_{t+k-1}|y_{1:t}] \quad (4.65)$$

$$= W_{t+k} + G_{t+k}R_t(k-1)G_{t+k}^\top \quad (4.66)$$

上記の推定法に基づくカルマン予測の疑似コードを図 4.3 に示す. 総じて, このアルゴリズムにおける演算は k 期先予測分布を $(k-1)$ 期先予測分布を状態遷移することで推定を行っている.

最後に、カルマンフィルタのパラメータ $\theta = \{G_t, F_t, W_t, V_t, m_0, C_0\}$ の初期設定について述べる [18]. 本研究では観測値の持つトレンドと周期 s の存在を仮定し、トレンドを考慮したローカルトレンドモデルと周期性を考慮した周期モデルの出力を足し合わせたモデルを使用する. この場合、ローカルトレンドモデルの初期パラメータは以下の通りになる.

$$x_t = [x_t], G_t = [1], W_t = W \quad (4.67)$$

$$F_t = [1], V_t = V \quad (4.68)$$

周期モデルのパラメータは一部に $(s-1)$ 次元のベクトルと行列を用いて以下の通りになる.

$$x_t = \begin{bmatrix} x_{t-s} \\ x_{t-s+1} \\ \vdots \\ x_{t-1} \end{bmatrix}, G_t = \begin{bmatrix} -1 & \cdots & -1 & -1 \\ 1 & & & \\ & \ddots & & \\ & & & 1 \end{bmatrix}, W_t = \begin{bmatrix} W & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \quad (4.69)$$

$$F_t = [1, 0, \dots, 0], V_t = V \quad (4.70)$$

また、本研究では $W = 0, V = 0$ として実験を行った.

第5章 SEIRモデルおよび拡張SEIRモデルを用いた検証実験

本章では, SEIRモデルのパラメータ推定法を用いた実験について述べる. 最初に, COVID-19の流行予測に関しては次の二通りの時期に分けて行う.

- ワクチン接種が進んでいない時期 (2020年12月 ~ 2021年4月)
- ワクチン接種の影響を考慮する時期 (2021年8月 ~ 2021年12月)

これらの時期で予測をするにあたって, 2章で述べた通常のSEIRモデルに加え, 自然免疫・重症者・ワクチンによる免疫といった因子を考慮した複数の拡張SEIRモデルの精度比較を行う. ここで用いる精度の評価指標について, 予測する時期によって Recovered の観測値の大小が異なり, 観測値と予測値の差分が大きく影響する絶対誤差である RMSE では観測値と予測値のずれの度合いを判断するのは難しい. 例えば得られたサンプル数が1であると仮定した場合, 観測値が50で予測値が49であった時の RMSE と観測値が20で予測値が19であった時の RMSE は両方とも1である. しかし, 前者と後者とではずれの重みが異なる. そこで, 本研究では精度の評価指標として観測値に対して差分を相対的に評価する相対誤差である MAPE(平均絶対誤差率)を採用した.

$$MAPE = \frac{1}{n-i+1} \sum_{t=i}^n \frac{|y_t - \hat{y}_t|}{|y_t|} \quad (5.1)$$

y_t : 観測値, \hat{y}_t : モデルの予測値

先ほどの例を MAPE で評価した場合, 前者の MAPE は0.02, 後者は0.05となる. これはそれぞれの予測値と実測値との絶対誤差は実測値に対して平均で2.0%, 5.0%ず

れていることを表している。すなわち、後者のずれの度合いの方が大きいと判断する。ワクチン接種が進んでいない時期、ワクチン接種の影響を考慮する時期それぞれにおいて用意した SEIR モデルの MAPE を比較し、その結果から高い精度で COVID-19 の流行を予測するのに考慮すべき因子について考察する。

5.1 実験概要

この節では実験の設定を述べる。実験における COVID-19 流行予測の期間と予測に用いるモデルを表 5.1 に示す。ワクチン接種が進んでいない時期の予測には Susceptible から Quarantined への遷移を含まない通常の SEIR モデル、重症者を考慮した SEIR モデル、自然免疫を考慮した SEIR モデルを用いる。日本の全人口におけるワクチン接種率が 20% を上回る [20]、2021 年 8 月から 2021 年 12 月の時期の予測にはこれらのモデルに自然免疫、ワクチンを考慮した SEIR モデルと重症者、自然免疫、ワクチンを考慮した SEIR モデルを加えた計 5 種類のモデルの MAPE を評価指標とした精度を検証する。未来の流行予測の期間は 1 ヶ月とし、各時期における月の始めを予測開始日に設定した。例えば、ワクチン接種が進んでいない時期の場合の予測開始日は 2020 年 12 月 1 日、2021 年 1 月 1 日といったように設定される。また、Step1 での 2 種類の過去の最適パラメータ導出法と、Step2 で行われる 2 種類の未来の最適パラメータ推定法の合計 4 通りの手法を用いた場合の予測値についても比較する。

表 5.1: COVID-19 流行予測の期間と予測に用いるモデル

予測期間	予測に用いるモデル
ワクチン接種が進んでいない時期 (2020 年 12 月 ~ 2021 年 4 月)	通常の SEIR モデル, 重症者を考慮した SEIR モデル, 自然免疫を考慮した SEIR モデル,
ワクチン接種を考慮する時期 (2021 年 8 月 ~ 2021 年 12 月)	上記の三つのモデルに加え 自然免疫、ワクチンを考慮した SEIR モデル, 重症者、自然免疫、ワクチンを考慮した SEIR モデル

表 5.2: 乱数で疑似的に導出する際の設定

	$R_0 \sim U(0.1, 3.0)$
	$e \sim U(4, 10.0)$
	$l \sim U(6, 14.0)$
パラメータの分布	$v \sim U(1e-7, 1e-2)$
	$v_r \sim U(1e-4, 1e-2)$
	$\alpha_r \sim U(1e-4, 1e-2)$
	$\alpha_s \sim U(1e-6, 1e-3)$
パラメータの組み合わせ	$n = 10000$
予測期間の長さ (日)	$S = 28$
過去のデータ	ワクチン接種が進んでいない時期は 2020年7月1日以降, ワクチン接種の影響を考慮する時期は 2021年2月17日以降のデータを使用.

5.1.1 過去の最適パラメータの導出

次に, 4章で述べた推定法を実験でどのように設定したのかについて述べる. 過去の最適パラメータの導出法の1つとしている「乱数で疑似的に導出」する方法について, 実験時の設定を表5.2に示す. パラメータについては各パラメータは最大値 a , 最小値 b の一様分布 $U(a, b)$ に従う乱数を生成し計 10000 通りの組み合わせを用意する. また, 予測期間については事前検証の結果, 最も有効であった 28 日に設定した. 過去の最適パラメータの導出に必要な過去のデータは, ワクチン接種が進んでいない時期は 2020 年 7 月 1 日, ワクチン接種を考慮する時期は日本国内でワクチン接種が始まる 2021 年 2 月 17 日以降 [20] のものを使用し, その末尾は未来の流行開始日の前日までとした.

最適化アルゴリズム (Adam) で導出する際の設定を表 5.3 に示す. パラメータの初期値の組み合わせは計算時間を考慮して 10 通りとし, この際 R_0, e, l は一様分布に従う乱数によって生成されるが $v, v_r, \alpha_r, \alpha_s$ は一定の数値で固定した. 過去のデータの範囲は前述の方法と同じである. Adam の更新式に用いるパラメータは $\alpha = 0.5, \beta_1 = 0.9, \beta_2 = 0.99$ とし, 更新を繰り返す回数は 50 回とした.

表 5.3: 最適化アルゴリズム (Adam) で導出する際の設定

パラメータの分布	$R_0 \sim U(0.2, 1.0)$
	$e \sim U(6, 10)$
	$l \sim U(6, 10)$
	$v = 1e + 4$
	$v_r = 1e + 4$
	$\alpha_r = 1e + 4$
	$\alpha_s = 1e + 6$
パラメータの組み合わせ	$n = 10$
予測期間の長さ (日)	$S = 28$
過去のデータ	ワクチン接種が進んでいない時期は 2020年7月1日以降, ワクチン接種の影響を考慮する時期は 2021年2月17日以降のデータを使用.
Adam のパラメータ	$\alpha = 0.5, \beta_1 = 0.9, \beta_2 = 0.99$
更新を繰り返す回数	$t = 50$

5.1.2 未来の最適パラメータの推定

ホルトウィントース法 (以下, HW) に関しては観測値以外に必要な設定は存在しない. カルマンフィルタ (以下, KF) のパラメータ $\theta = \{G_t, F_t, W_t, V_t, m_0, C_0\}$ の初期設定については4章で述べたが再掲する. ローカルトレンドモデルの初期パラメータは以下の通りになる.

$$x_t = [x_t], G_t = [1], W_t = W \quad (5.2)$$

$$F_t = [1], V_t = V \quad (5.3)$$

周期モデルのパラメータは一部に $(s-1)$ 次元のベクトルと行列を用いて以下の通りになる.

$$x_t = \begin{bmatrix} x_{t-s} \\ x_{t-s+1} \\ \vdots \\ x_{t-1} \end{bmatrix}, G_t = \begin{bmatrix} -1 & \cdots & -1 & -1 \\ 1 & & & \\ & \ddots & & \\ & & & 1 \end{bmatrix}, W_t = \begin{bmatrix} W & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \quad (5.4)$$

$$F_t = [1, 0, \dots, 0], V_t = V \quad (5.5)$$

本研究ではこれら2つのモデルの出力を足し合わせて予測値を算出した。これらのモデルにおける定数 W, V に関しては $W = 0, V = 0$ として実験を行った。

5.1.3 実験環境

データの管理は Excel, プログラミング言語は R, OS は Windows10, CPU は Intel(R) Core(TM) i5-10300H CPU@2.50GHz 2.50GHz を使用した。プログラム実行時には doParallel パッケージ, snow パッケージによるマルチコアの並列処理によって計算時間を短縮した。

5.2 実験結果

この節では、複数の SEIR モデルの精度比較実験を行った結果について述べる。

5.2.1 ワクチン接種が進んでいない時期 (2020年12月～2021年4月)

通常の SEIR モデル (モデル1), 重症者込みの SEIR モデル (モデル2), 自然免疫込みの SEIR モデル (モデル3) について Step1 で過去の最適パラメータを乱数を用いて導出した際の Recovered の予測値を図 5.1～図 5.5, Step1 で過去の最適パラメータを最適化アルゴリズムを用いて導出した際の Recovered の予測値を図 5.6～図 5.10, これらの実測値との MAPE を表 5.4, 表 5.5 に示す。2020年の12月においては MAPE が 0.1 を越えるモデルの予測値が複数確認でき, 図 5.1, 図 5.6 ではそれらが実測値に対して大幅に乖離していた。表 5.4, 表 5.5 からは 2021年1月, 2月では一つの Step1 と Step2 の組み合わせに対して固定したモデルの構造に関わらず各モデルの MAPE がそれぞれが近い値となり, 図 5.3～図 5.10 では各モデルの予測値が重なり合うようにプロットされるパターンも確認された。また, 全体的に予測期間が 2020年12月から 2021年4月に近づくほど MAPE が小さくなる傾向を示していた。未来の予測開始日が先になるほど Step1 で参照できる過去のデータの長さが増え, その分より多くの過去の情報を取り込むことで正確な未来の最適パラメータの推定に繋がり, 予測精度の向上に寄与することが示唆される。

表 5.4: 乱数によって過去の最適パラメータを導出した際の MAPE

予測期間 (1ヶ月)	通常		重症者込み		自然免疫込み	
	HW	KF	HW	KF	HW	KF
2020年12月	43.021706	0.085055	3.079404	0.069454	0.097939	0.055811
2021年1月	0.051279	0.020858	0.024198	0.019391	0.057257	0.020994
2021年2月	0.029544	0.068476	0.048759	0.070224	0.207246	0.078385
2021年3月	0.014810	0.008262	0.005092	0.008234	0.005256	0.041694
2021年4月	0.007263	0.001694	0.006879	0.012841	0.007013	0.001496
平均	8.627393	0.038686	0.632867	0.036029	0.077380	0.041401

Step1 の手法について、最適化アルゴリズムの処理時間は約7時間、乱数によって過去の最適パラメータを導出した際の処理時間は約30分であった。実験を始める前は処理時間の差から最適化アルゴリズムを用いた方が MAPE は全体的に小さくなると予想していた。しかし、実際は予想に反して乱数を用いた方が MAPE が小さくなっているケースを確認した。このことより、Step1 で Adam による最適化が十分ではなく、乱数を用いた場合よりも最適なパラメータを導出できていない場合もあったと言える。十分な最適化のためには、パラメータの初期値の範囲を広くする、学習率 α や更新を繰り返す回数 t を更に大きくするなどの改善策が挙げられる。

MAPE の平均値が最も小さかったモデルと手法の組み合わせは、予測モデルにモデル3、Step1 に最適化アルゴリズム、Step2 に KF を選択し予測を行った場合で MAPE の平均値は 0.035387、次点で予測モデルにモデル2、Step1 に乱数、Step2 に KF を選択して予測を行った場合で MAPE の平均値 0.036029 であった。これら2つの結果を比較した場合、ワクチン接種が始まっていないまたはワクチン接種率が低い時期に SEIR モデルを用いて COVID-19 の流行を予測する際には自然免疫の影響を考慮した SEIR モデルの精度が高くなる。

表 5.5: 最適化アルゴリズムによって過去の最適パラメータを導出した際の MAPE

予測期間 (1ヶ月)	通常		重症者込み		自然免疫込み	
	HW	KF	HW	KF	HW	KF
2020年12月	0.322164	0.078721	0.297037	0.074060	0.399111	0.080209
2021年1月	0.034329	0.020867	0.032356	0.020196	0.034380	0.020812
2021年2月	0.052354	0.063254	0.062307	0.074968	0.052610	0.063581
2021年3月	0.013998	0.009542	0.013415	0.009168	0.015103	0.002200
2021年4月	0.010361	0.012603	0.010363	0.012615	0.010929	0.011871
平均	0.086392	0.036619	0.082846	0.037822	0.102134	0.035387

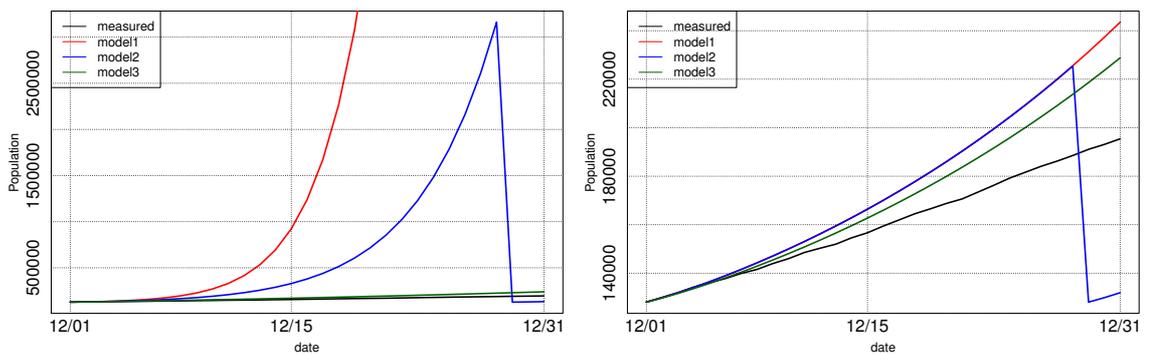


図 5.1: Step1 で乱数, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (12月)

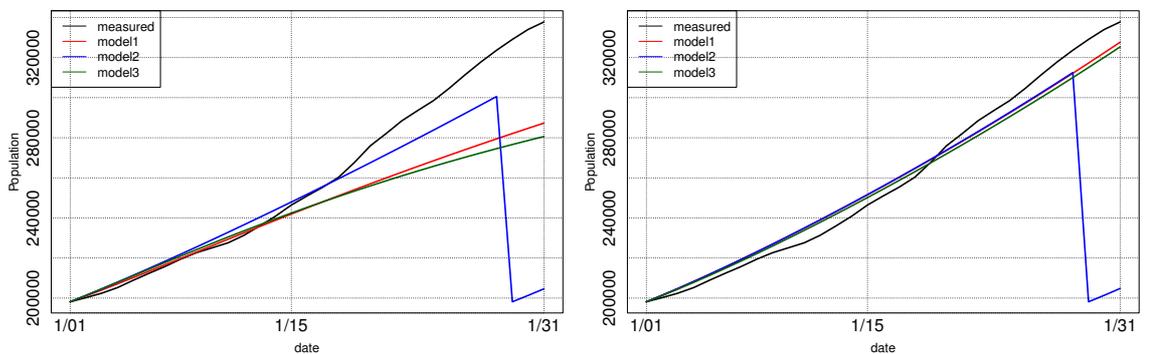


図 5.2: Step1 で乱数, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (1月)

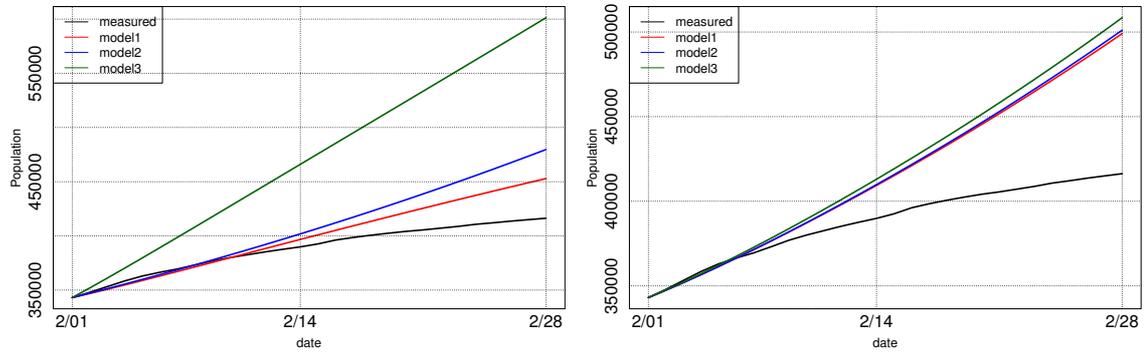


図 5.3: Step1 で乱数, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2月)

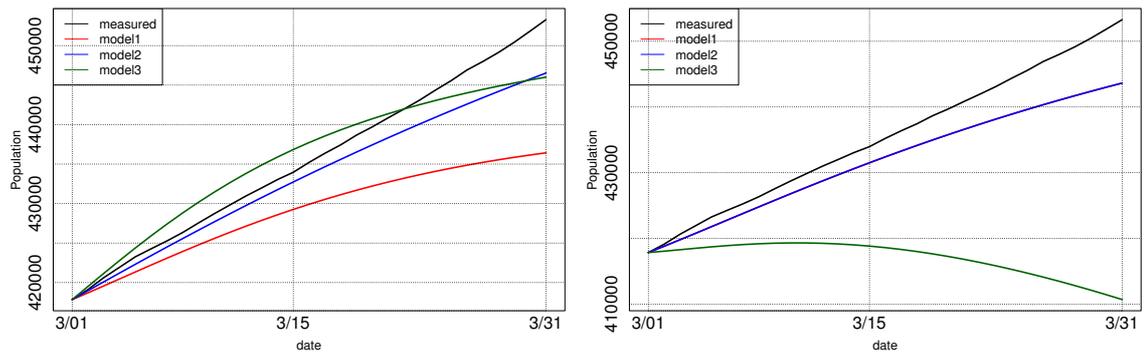


図 5.4: Step1 で乱数, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (3月)

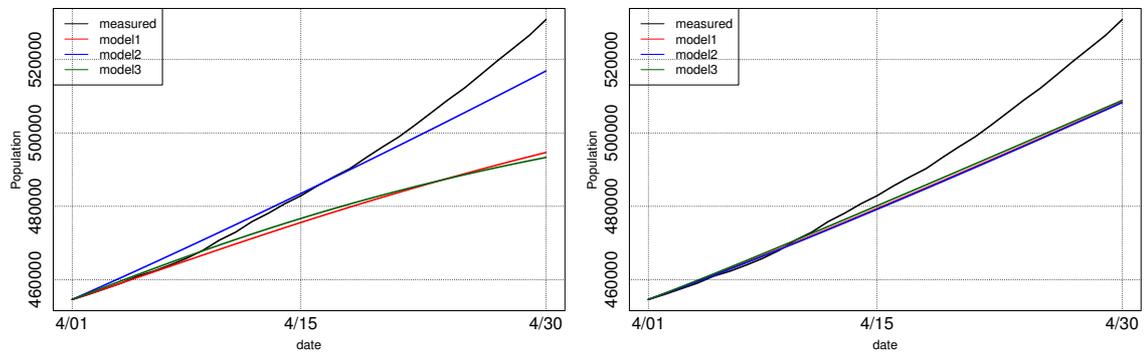


図 5.5: Step1 で乱数, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (4月)

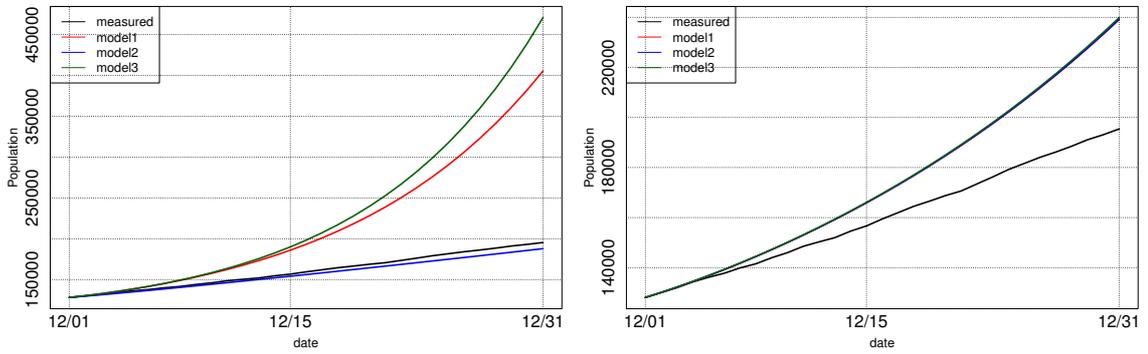


図 5.6: Step1 で最適化アルゴリズム, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2020 年 12 月)

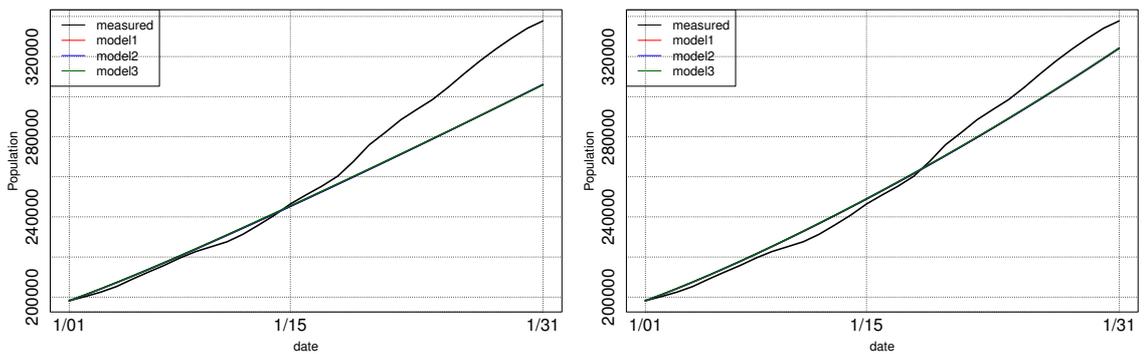


図 5.7: Step1 で最適化アルゴリズム, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2021 年 1 月)

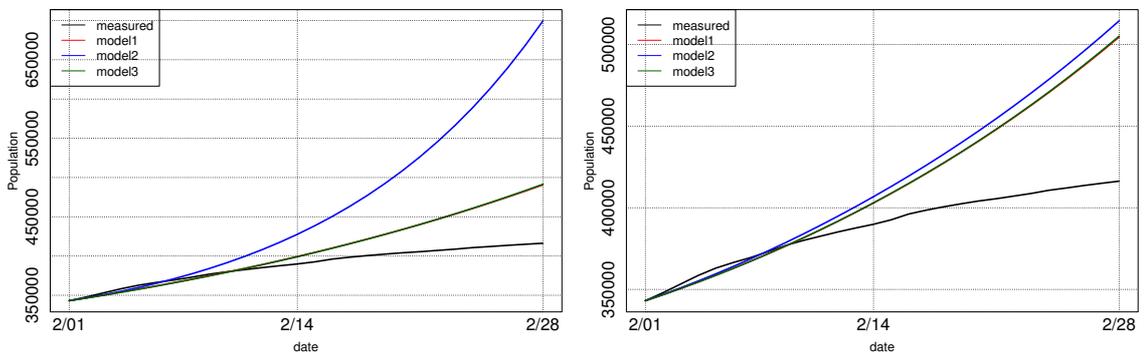


図 5.8: Step1 で最適化アルゴリズム, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2021 年 2 月)

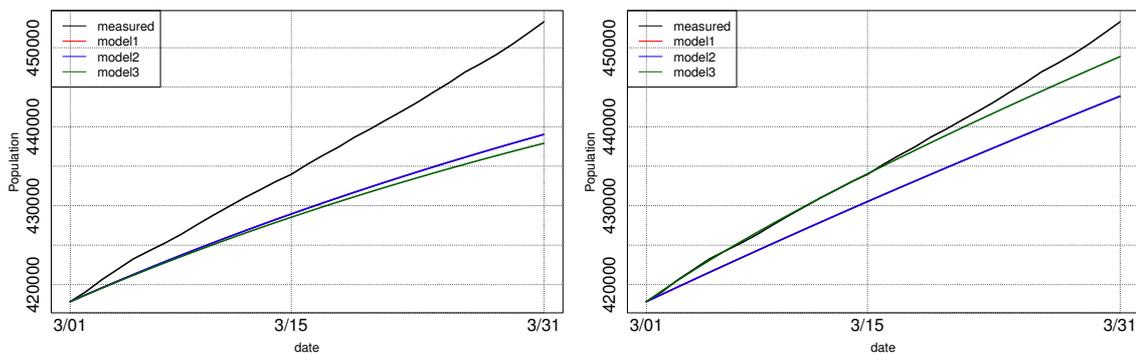


図 5.9: Step1 で最適化アルゴリズム, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2021年3月)

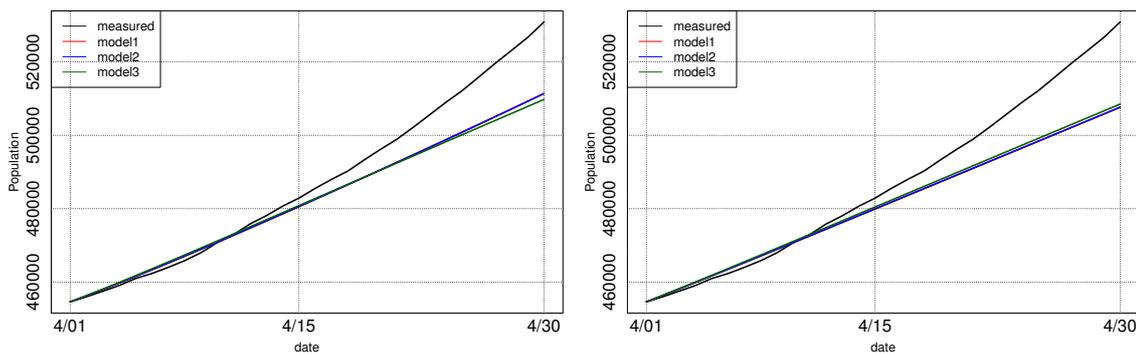


図 5.10: Step1 で最適化アルゴリズム, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2021年4月)

5.2.2 ワクチン接種の影響を考慮する時期 (2021年8月～2021年12月)

通常の SEIR モデル (モデル 1), 重症者込みの SEIR モデル (モデル 2), 自然免疫込みの SEIR モデル (モデル 3), 自然免疫とワクチンによる免疫込みの SEIR モデル (モデル 4), 重症者, 自然免疫, ワクチンによる免疫込みの SEIR モデル (モデル 5) における Step1 で過去の最適パラメータを乱数を用いて導出した際の Recovered の予測値を図 5.11～5.15, Step1 で過去の最適パラメータを最適化アルゴリズムを用いて導出した際の Recovered の予測値を図 5.16～5.20, これらの実測値との MAPE を表 5.6, 表 5.7 に示す. 2021年8月 COVID-19 の新規感染者の急激な増加傾向にあった時期であった. MAPE も後の予測期間と比べて大きく, 図 5.11, 図 5.16 では月末においては予測値と実測値との絶対誤差が実測値の 2 倍以上ずれているモデルを複数確認した. 現状この期間における Recovered 予測精度は低く, 新規感染者の急増にも対応し十分な精度を確保できる手法の模索が今後の課題となる. 2021年10月, 11月, 12月に関しては先の期間とは対称的に1日当たりの新規感染者が少なく COVID-19 の流行が一時的に落ち着きつつあった時期であり, MAPE も12月に近づくにつれ減少する傾向が確認できたため, このような時期の予測精度は十分高い. 特に12月においては Step1 に最適化アルゴリズム, Step2 に HW を用いたモデル 5 を除いて予測値と実測値とのずれが最大で1万未満であった様子が図 5.15, 図 5.20 から確認できた. また, 表 5.6, 表 5.7 ではワクチン接種が進んでいない時期同様に最適化アルゴリズムによる最適化が十分ではないケースもあった.

MAPE の平均値が最小となったモデルと手法の組み合わせは, 予測モデルにモデル 3, Step1 に乱数, Step2 に KF を選択して予測を行った場合で MAPE の平均値は 0.022514 であった. これより, ワクチン接種率が高まりその影響を考慮する時期に Recovered を高い精度で予測する上では, SEIR モデルに重症者を考慮せずに自然免疫とワクチンによる免疫獲得の影響を加味した方がよい.

表 5.6: 乱数によって過去の最適パラメータを導出した際の MAPE

予測期間 (1ヶ月)	通常			重症者込み			自然免疫込み			自然免疫+ ワクチン			重症者+自然免疫 +ワクチン		
	HW	KF	HW	HW	KF	HW	HW	KF	HW	KF	HW	KF	HW	KF	
2021年8月	0.479484	0.022574	0.183129	0.022068	0.048158	0.017796	0.035537	0.019963	0.071699	0.021297					
2021年9月	0.089222	0.092761	0.932803	0.096201	0.235366	0.126534	0.222015	0.082462	0.245273	0.119006					
2021年10月	0.003134	0.001692	0.003259	0.001575	0.006096	0.006647	0.000991	0.003155	0.030952	0.027888					
2021年11月	0.000843	0.000955	0.000887	0.000984	0.003409	0.005837	0.003861	0.006099	0.004142	0.003966					
2021年12月	0.000634	0.000187	0.000829	0.000186	0.001435	0.001099	0.000186	0.000893	0.000474	0.000786					
平均	0.114663	0.023634	0.224181	0.024203	0.058893	0.031583	0.052518	0.022514	0.070508	0.034589					

表 5.7: 最適化アルゴリズムによって過去の最適パラメータを導出した際の MAPE

予測期間 (1ヶ月)	通常			重症者込み			自然免疫込み			自然免疫+ ワクチン			重症者+自然免疫 +ワクチン		
	HW	KF	HW	HW	KF	HW	HW	KF	HW	KF	HW	KF	HW	KF	
2021年8月	0.073596	0.029233	0.073847	0.020174	0.057179	0.017854	0.057447	0.012741	1.886301e+25	0.015160					
2021年9月	0.387674	0.104865	0.388388	0.104879	0.281025	0.104600	0.281708	0.163851	0.054263	0.118535					
2021年10月	0.001892	0.002590	0.001892	0.002590	0.002477	0.001802	0.002446	0.005536	0.003026	0.001837					
2021年11月	0.000965	0.000961	0.000965	0.000960	0.002148	0.002178	0.002147	0.002178	0.001599	0.002174					
2021年12月	0.000354	0.000186	0.000362	0.000189	0.001926	0.000877	0.001958	0.000871	0.015412	0.000829					
平均	0.092896	0.027567	0.093091	0.025759	0.068951	0.025462	0.069141	0.037035	3.143835e+24	0.027707					

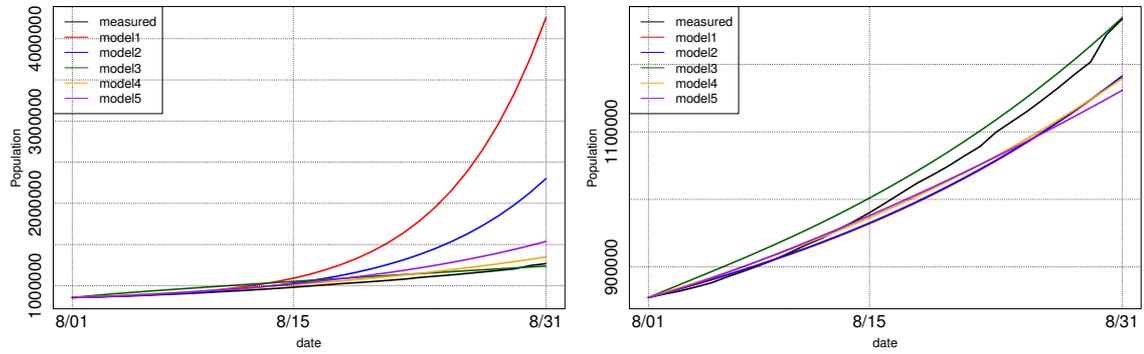


図 5.11: Step1 で乱数, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2021 年 8 月)

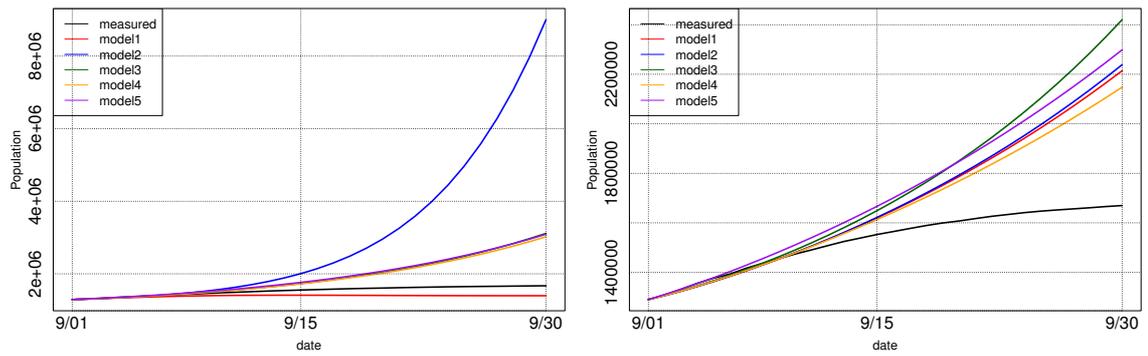


図 5.12: Step1 で乱数, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2021 年 9 月)

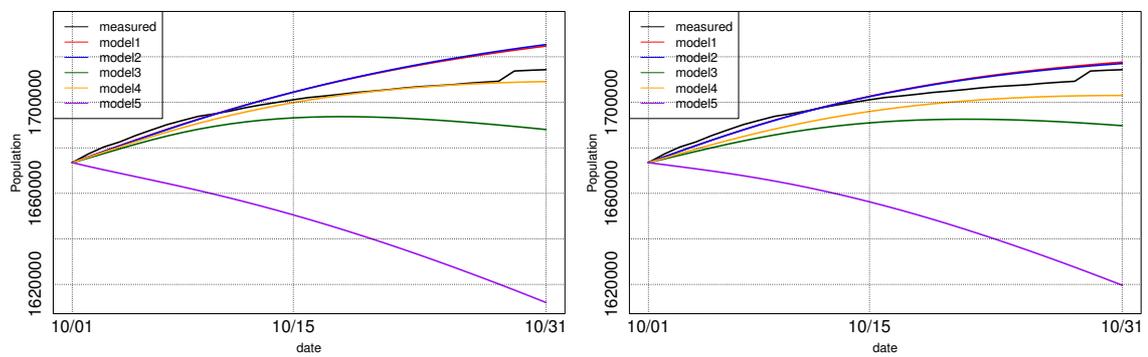


図 5.13: Step1 で乱数, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2021 年 10 月)

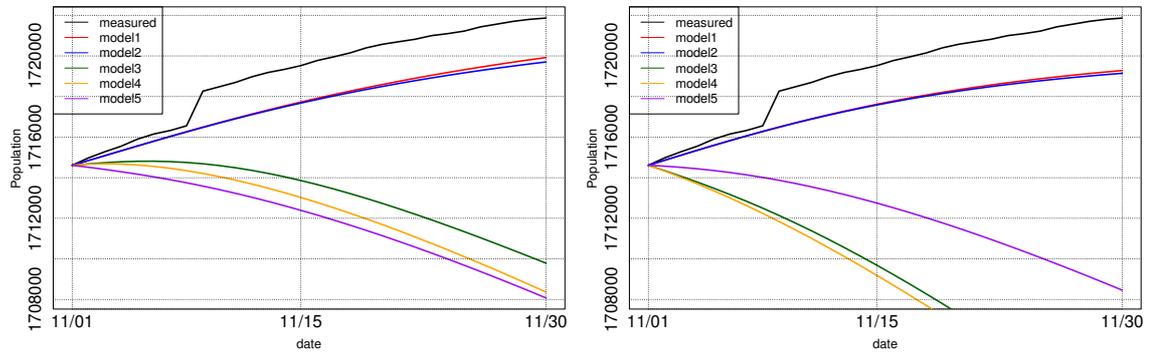


図 5.14: Step1 で乱数, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2021 年 11 月)

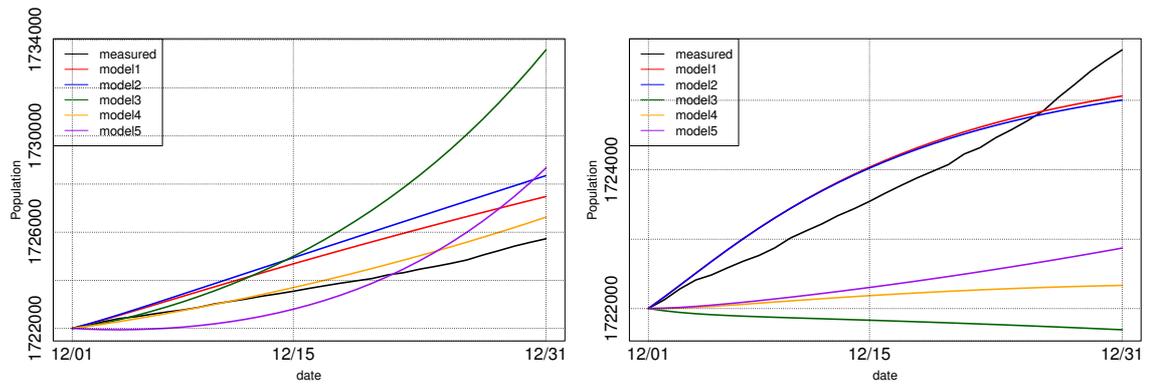


図 5.15: Step1 で乱数, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2021 年 12 月)

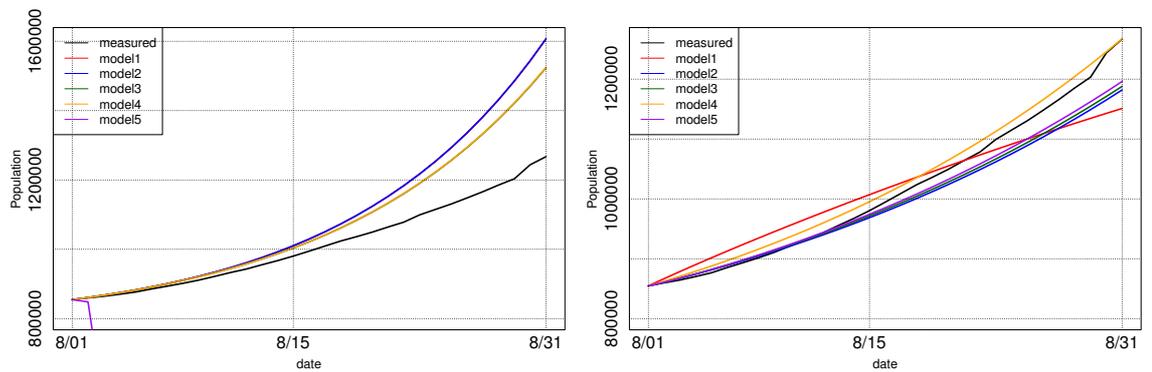


図 5.16: Step1 で最適化アルゴリズム, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (8 月)

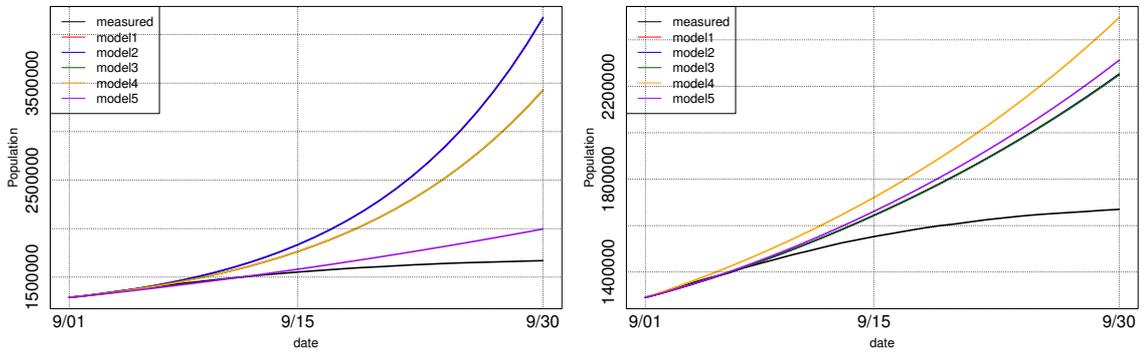


図 5.17: Step1 で最適化アルゴリズム, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (9月)

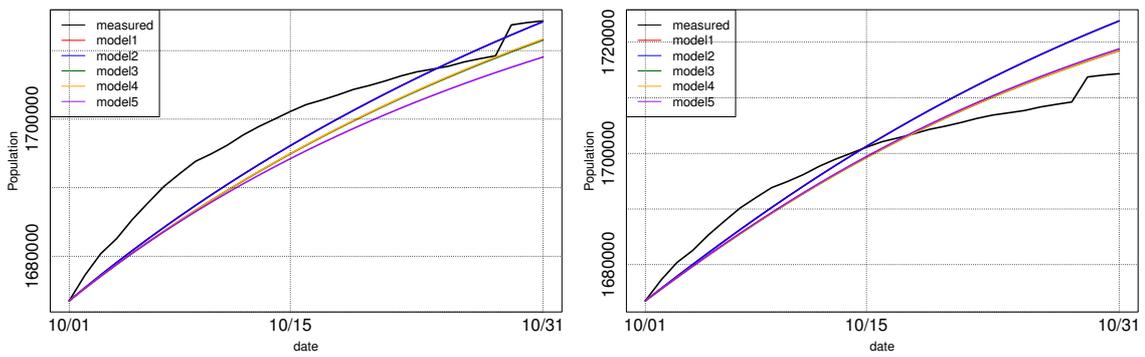


図 5.18: Step1 で最適化アルゴリズム, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (10月)

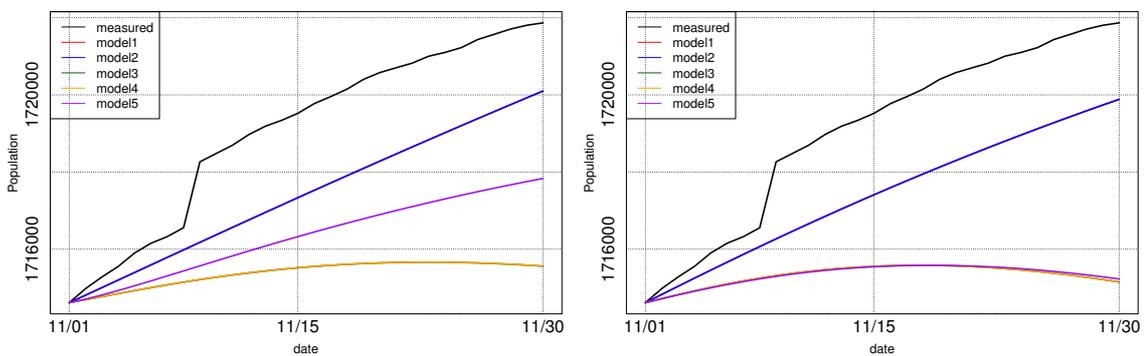


図 5.19: Step1 で最適化アルゴリズム, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (11月)

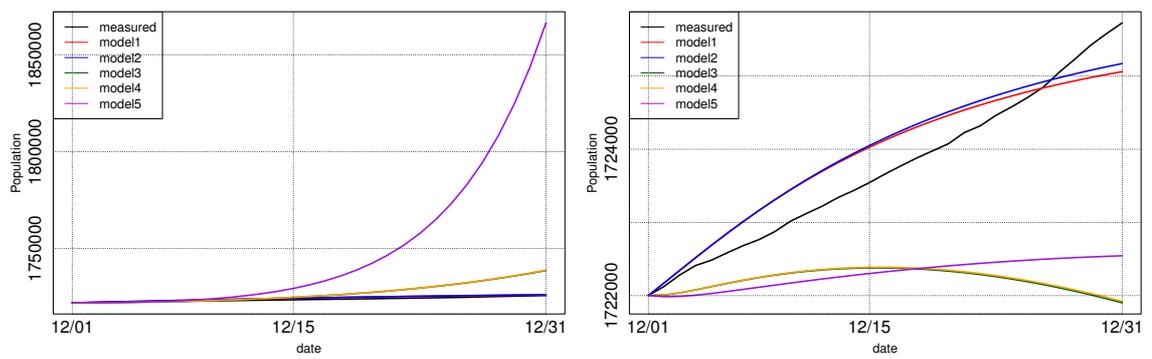


図 5.20: Step1 で最適化アルゴリズム, Step2 で HW(左), KF(右) を用いた場合の Recovered の予測値 (2021 年 12 月)

第6章 拡張SEIRモデルを用いた感染者数の傾向予測

日本における Infected の実測値 [10] を図 6.1, 日本における Recovered の実測値の差分, 即ち前後の要素間の差 ($y_{t+1} - y_t$) を図 6.2 に示す. これらの 2 つの値の間には強い相関があるため ($r = 0.956457$), SEIR モデルによって Recovered を高い精度で予測できている場合, その予測値の差分は Infected の実測値と正の相関関係にあり Infected の増加傾向ないしは減少傾向の予測が可能になるという仮説を立てた. そこで, 本章では 5 章でワクチン接種が進んでいない時期とワクチン接種を考慮する時期それぞれにおいて MAPE の平均値が最小となったモデルと手法の組み合わせ (表 6.1) を用いて Recovered の予測値の差分を求め, Infected の実測値との相関関係を調べる. また, そのモデルによる Infected の予測値も同様に Infected の実測値との相関関係を調べ, 2 つの調査結果を比較する.

表 6.1: 5 章の実験で MAPE の平均値が最小であったモデルと手法の組み合わせ

予測を行う時期	使用したモデル	Step1	Step2
ワクチン接種が進んでいない時期 (2020 年 12 月 ~ 2021 年 4 月)	自然免疫を考慮した SEIR モデル	最適化アルゴリズム	KF
ワクチン接種を考慮する時期 (2021 年 8 月 ~ 2021 年 12 月)	自然免疫とワクチンを 考慮した SEIR モデル	乱数	KF

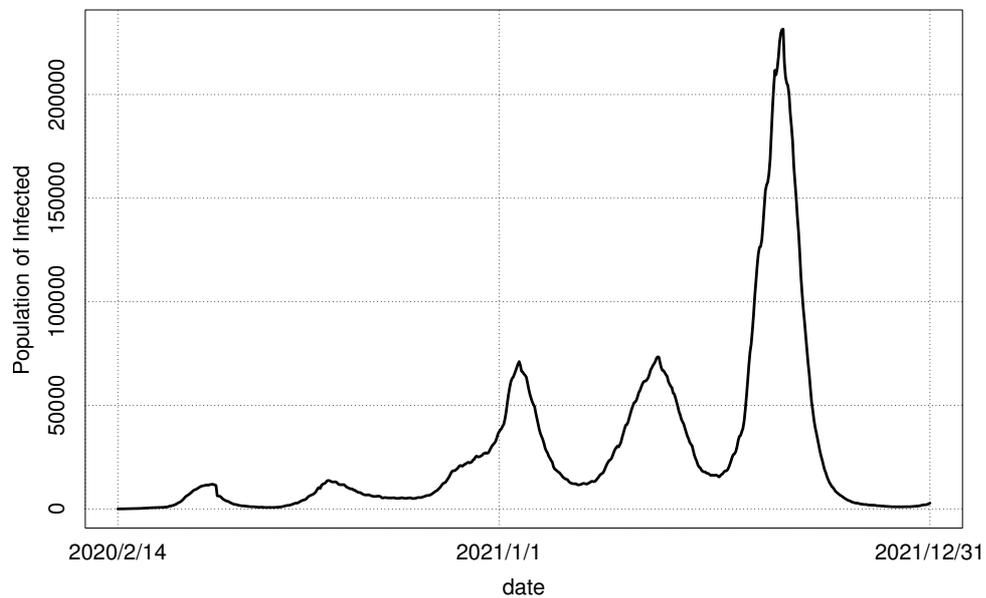


図 6.1: 日本における Infected の実測値

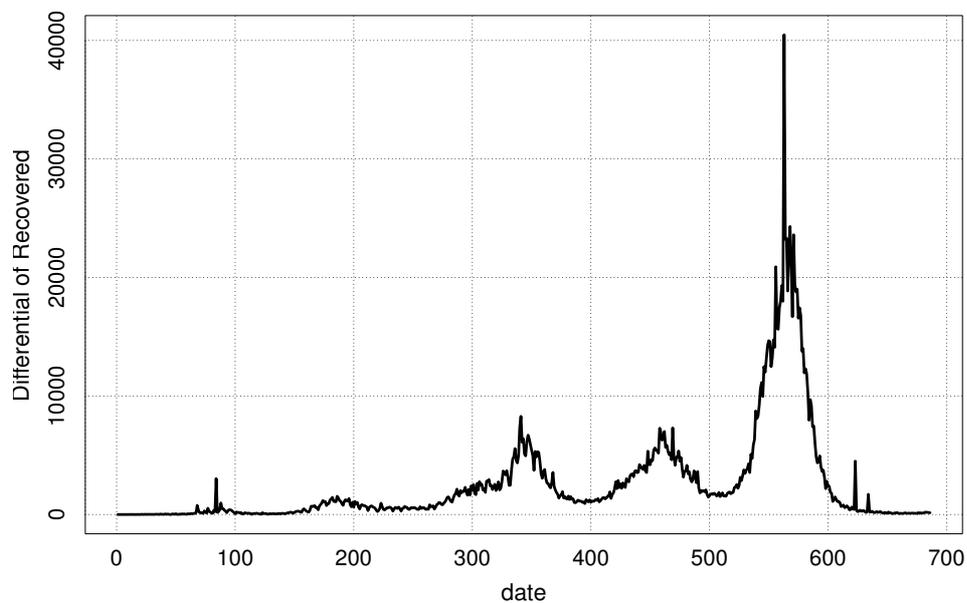


図 6.2: Recovered の実測値の差分 (2020年2月14日～2021年12月31日)

Recovered の予測値の差分及び Infected の予測値と Infected の実測値との相関係数を表 6.2, Infected の実測値と予測値及び Recovered の予測値の差分を図 6.3 ~ 図 6.12 に示す. 表 6.2 より, Infected の実測値と強い正の相関 ($r > 0.90$) を示した予測期間だけではなく, 逆に強い負の相関 ($r < -0.90$) を示した予測期間も確認できた. 2021 年 3 月, 2021 年 4 月を除いて 2 種類の相関係数の絶対値に大きな差はないため, Recovered の予測値の差分または Infected の予測値によって Infected の実測値の傾向を予測することは一部の予測期間で可能である. 表 6.2 と図 6.3 ~ 図 6.12 から, 相関係数と Infected の実測値の傾向との関係は次の通りになる.

- 相関係数が 0.90 に近い

モデルが Infected の増加傾向, 減少傾向を正確に予測できていた. (図 6.3, 図 6.8, 図 6.10, 図 6.11)

- 相関係数の絶対値が 0.50 に近い

モデルは最初の数日の Infected の傾向を予測できていたが, その後は実際の Infected が Recovered の予測値の差分とは違う変化をしていた (図 6.4, 図 6.6). あるいは Recovered 予測値の差分は停滞していたが実際の Infected は増加していた. (図 6.7, 図 6.12)

- 相関係数が -0.90 に近い

モデルが実際の Infected の傾向とは真逆の傾向を予測していた. (図 6.5, 図 6.9)

これらを考慮すると, SEIR モデルを用いて Infected の傾向を正確に予測するのであれば最終的に Recovered の予測値の差分と Infected の実測値が強い正の相関になる可能性の高い時期に行うことが望ましい. 従って, そのような時期はどのような特徴を持っているのかについて考察する.

今回の検証で用いた SEIR モデル (表 6.1) は 2020 年 12 月, 2021 年 8 月, 10 月, 11 月における Infected の傾向を予測できていた. 2020 年 12 月, 2021 年 8 月について, それぞれの前の月にあたる 2020 年 11 月, 2021 年 7 月には Infected の実測値の急増が始まっていた. また 2021 年 10 月, 11 月においてはそれぞれの前の月は逆に Infected が下がり始めていた. 即ち, 2020 年 12 月, 2021 年 8 月, 10 月, 11 月は予測開始日の 1ヶ

表 6.2: Infected の実測値との相関係数

予測期間	Recovered の 予測値の差分	Infected の 予測値
2020年12月	0.973004	0.978865
2021年1月	0.470719	0.527949
2021年2月	-0.941027	-0.937491
2021年3月	0.426782	-0.576844
2021年4月	0.590570	-0.911632
2021年8月	0.952419	0.953557
2021年9月	-0.981297	-0.981933
2021年10月	0.942157	0.933452
2021年11月	0.990732	0.992435
2021年12月	-0.636404	-0.538579

月前から Infected の傾向が変わり始め続いていた. 従って, このような時期に SEIR モデルを用いて予測を行うことで未来の Infected の傾向を予測できるであろう.

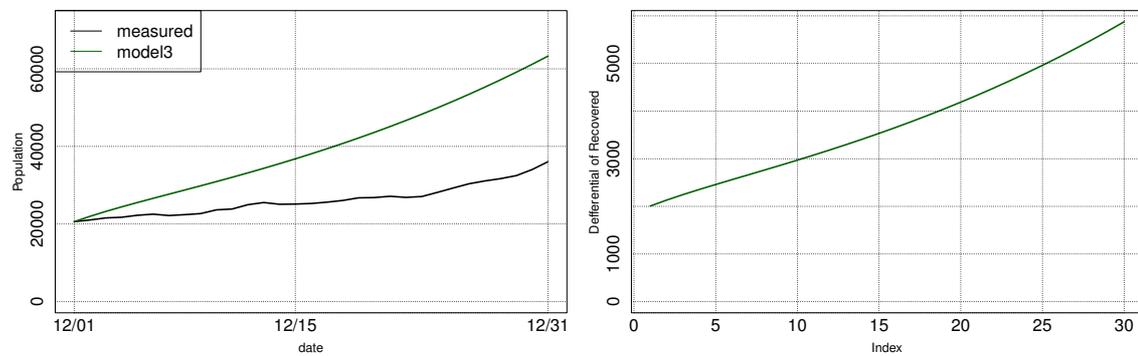


図 6.3: Infected の実測値と予測値 (左), Recovered の予測値の差分 (右) のプロット (2020年12月)

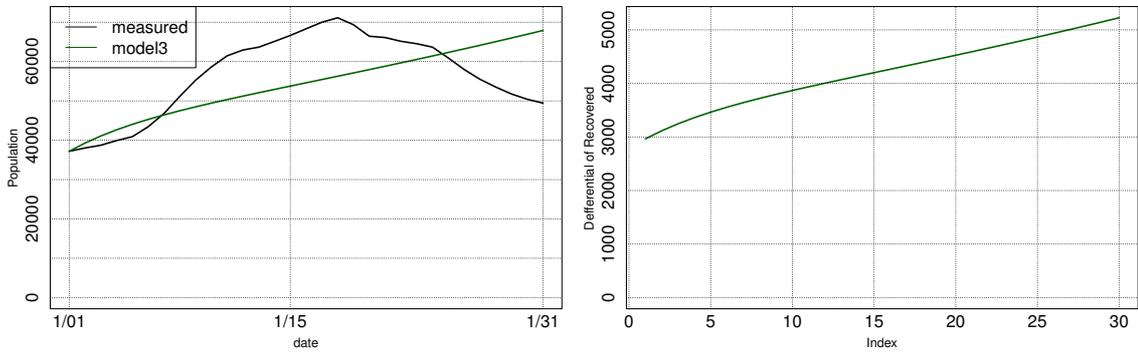


図 6.4: Infected の実測値と予測値 (左), Recovered の予測値の差分 (右) のプロット (2021年1月)

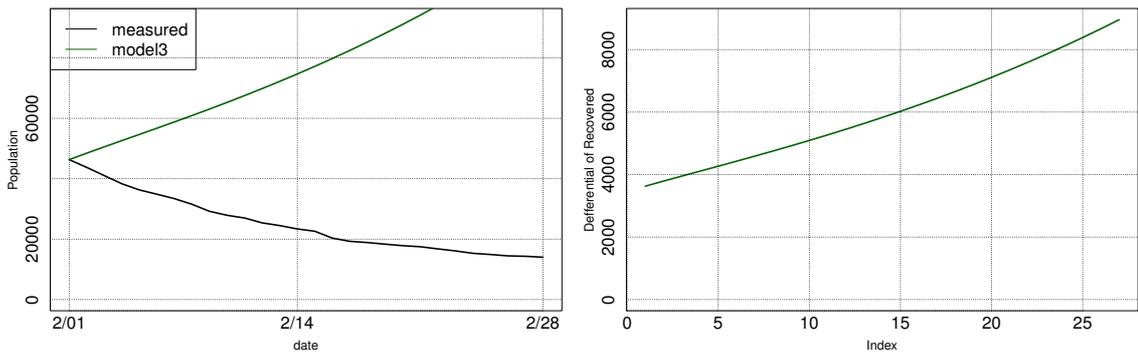


図 6.5: Infected の実測値と予測値 (左), Recovered の予測値の差分 (右) のプロット (2021年2月)

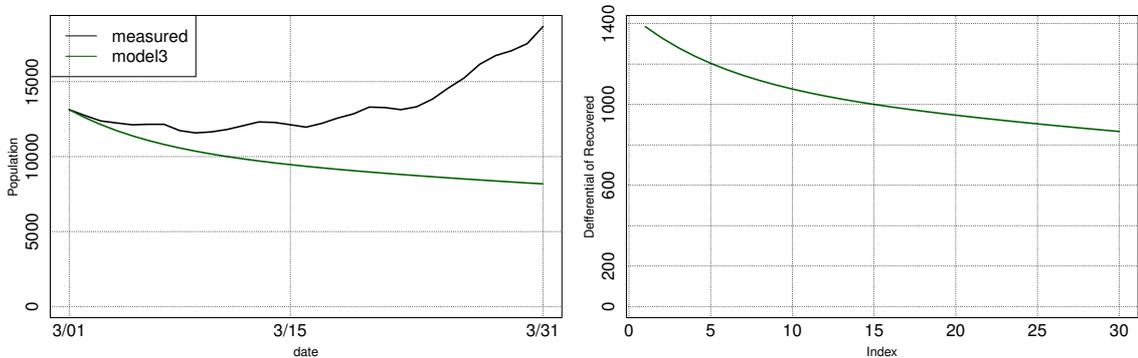


図 6.6: Infected の実測値と予測値 (左), Recovered の予測値の差分 (右) のプロット (2021年3月)

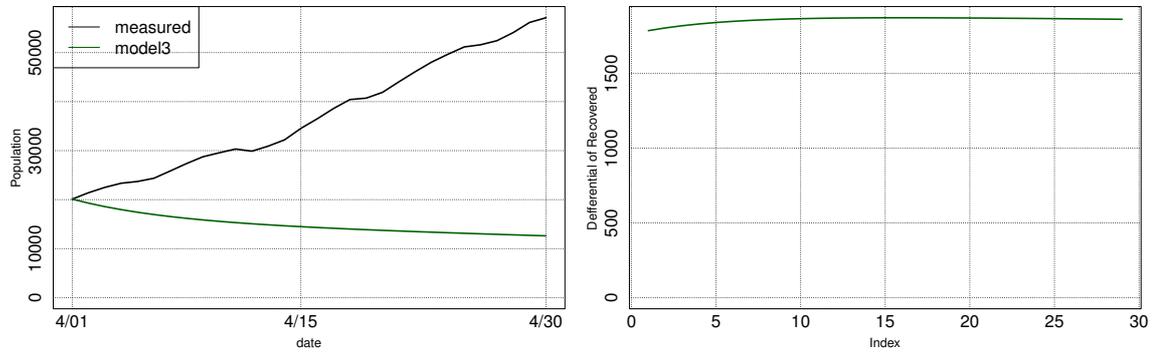


図 6.7: Infected の実測値と予測値 (左), Recovered の予測値の差分 (右) のプロット (2021年4月)

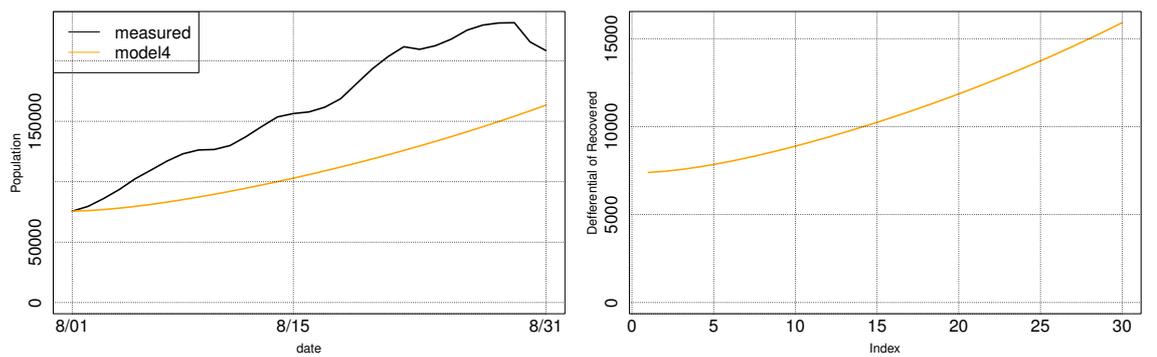


図 6.8: Infected の実測値と予測値 (左), Recovered の予測値の差分 (右) のプロット (2021年8月)

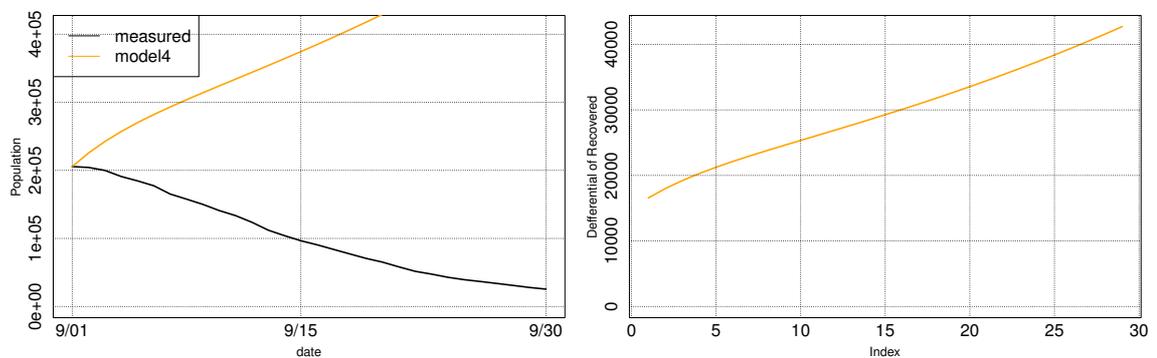


図 6.9: Infected の実測値と予測値 (左), Recovered の予測値の差分 (右) のプロット (2021年9月)

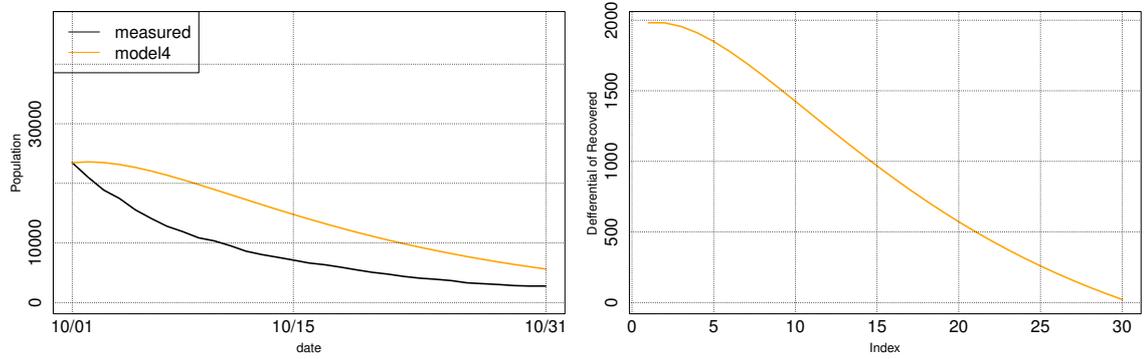


図 6.10: Infected の実測値と予測値 (左), Recovered の予測値の差分 (右) のプロット (2021 年 10 月)

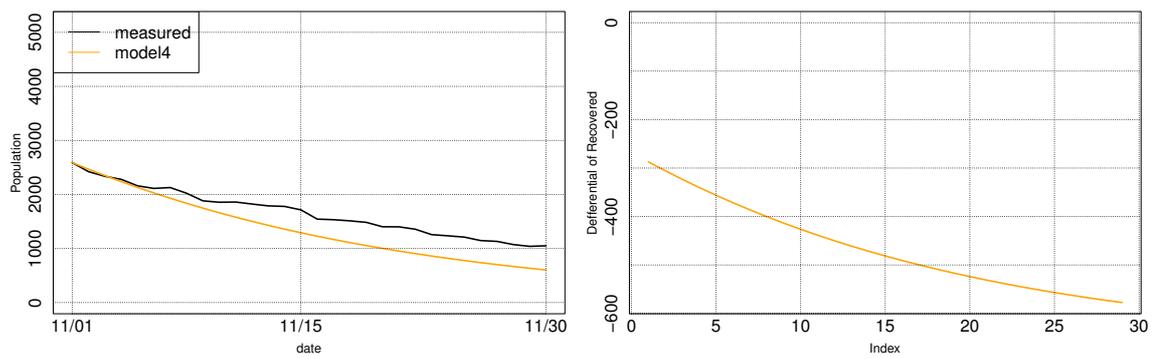


図 6.11: Infected の実測値と予測値 (左), Recovered の予測値の差分 (右) のプロット (2021 年 11 月)

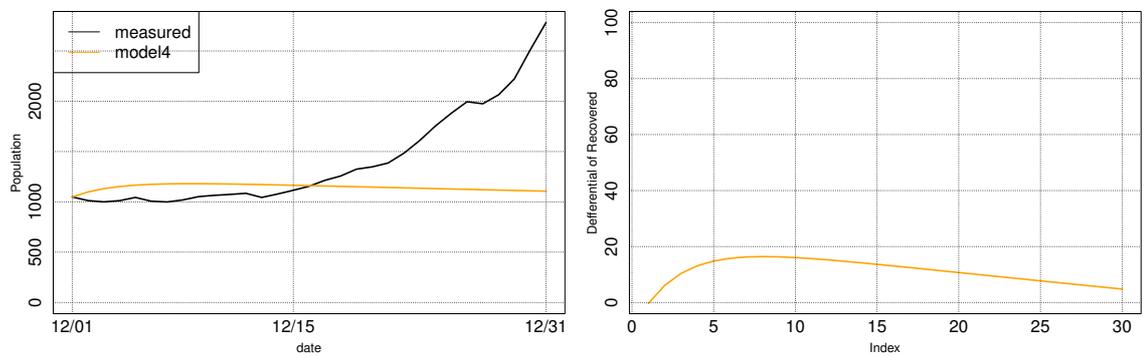


図 6.12: Infected の実測値と予測値 (左), Recovered の予測値の差分 (右) のプロット (2021 年 12 月)

第7章 むすび

本研究では, 感染症の流行予測に有効な SEIR モデルを用いて日本における COVID-19 の流行予測を行った. その結果, ワクチン接種が進んでいない時期である 2020 年 12 月 ~2021 年 4 月に Recovered の予測を行う場合は通常の SEIR モデルに自然免疫の影響を加味して拡張された SEIR モデルとそのパラメータ推定法に最適化アルゴリズム及び KF の組み合わせ, ワクチン接種率が 20%を越え, その影響を考慮する時期である 2021 年 8 月 ~12 月の場合は通常の SEIR モデルに自然免疫とワクチンによる免疫の影響を加味して拡張された SEIR 本モデルとパラメータ推定法に乱数及び KF の組み合わせが有効であった. また, 予測した Recovered の差分を用いることで, 予測開始日の 1ヶ月前に Infected の増加傾向または減少傾向が始まった場合に未来の Infected の傾向を予測した.

今後の課題としては, まず第一に予測精度の向上が挙げられる. ワクチン接種を考慮する時期に Recovered の予測を行った際に, 2021 年 8 月, 9 月のような COVID-19 の流行が急激に変化する時期における予測精度が他の予測時期と比較して低かった. 解決策としては過去の最適パラメータ導出の際に最適化アルゴリズムにおけるパラメータの初期点の範囲・学習率・更新を繰り返す回数を増大する, 得られた過去の最適パラメータの特徴を調査しその結果を未来の最適パラメータ推定に利用する, 新たな拡張 SEIR モデルを用意するなどの方法を検討している. 第二に, 現状 SEIR モデルによって全ての時期において Infected の傾向を予測することはできていないため, その問題点の改善が挙げられる. 具体的な方法としては, 最初から SEIR モデルのパラメータを Recoveredのみだけではなく, Infected を含む予測に向けて最適化などの方法を検討している.

謝辞

本研究を進めるにあたり, 日頃から多大な御指導, 御尽力, 御協力を賜りました名古屋工業大学, 舟橋 健司准教授, 伊藤 宏隆助教および舟橋研究室諸氏に心から感謝致します. 本研究の一部は JSPS 科研費 JP19K03028 の助成を受けたものです.

参考文献

- [1] 新型コロナウイルス感染症診療の手引き,
https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_001111.html
- [2] 西浦博、稲葉寿：“感染症流行の予測：感染症数理モデルにおける定量的課題”，
統計数理, 第54巻, 第2号, 2006.
- [3] L. Peng, W. Yang, T. Zhang, C. Zhuge, and L. Hong, “Epidemic analysis of COVID-19 in China by dynamical modeling”, arXiv preprint arXiv, February 2020.
- [4] G. Pandey, P. Chaudhary, R. Gupta, and S. Pal, “SEIR and Regression Model based COVID-19 outbreak predictions in India”, Pandey et al, 2020.
- [5] L. Lopez, and X. Rodo, “A modified SEIR model to predict the COVID-10 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemic”, Results in Physics 21, No.103764, 2021.
- [6] H. Alrabaiah, M. Afran, K. Shah, Ibrahim Mahariq, and A. Ukkah, “A comparative study of spreading of novel corona virus disease by using fractional order modified SEIR model”, Alexandria Engineering Journal (2021), vol. 60, pp. 573–585, 2021.
- [7] S. O. Arik et al., “Interpretable Sequence Learning for COVID-19 Forecasting”, arXiv:2008.00646 [cs, stat], Aug. 2020.
- [8] N. b. Khedher, L. Kolsi, H. Alsaif, “A multi-stage SEIR model to predict the potential of a new COVID-19 wave in KSA after lifting all travel restrictions”, Alexandria Engineering Journal (2021), 60, pp. 3965–3974, 2021.

- [9] R. Ghostine, M. Gharamti, S. Hassrouny, I. Hoteit, “An Extended SEIR Model with Vaccination for Forecasting the COVID-19 Pandemic in Saudi Arabia Using an Ensemble Kalman Filter”, *Mathematics* 9, 636, 2021.
- [10] 新型コロナウイルス国内感染の状況,
<https://toyokeizai.net/sp/visual/tko/covid19/en.html>
- [11] 金森敬文, 鈴木大慈, 竹内一郎, 佐藤一誠 : “機械学習のための連続最適化”, 講談社, 東京都, 2014.
- [12] 福島雅夫 : “新版 数理計画入門”, 朝倉書店, 東京都, 2011.
- [13] D. John, H. Elad, and S. Yoram, “Adaptive subgradient methods for online learning and stochastic optimization”, *The Journal of Machine Learning Research*, 12:21212159, 2011.
- [14] T. Tieleman and G. Hinton. Lecture 6.5 - RMSProp, COURSERA: Neural Networks for Machine Learning. Technical report, 2012.
- [15] D. P. Kingma, J. L. Ba, “ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION”, *ICLR 2015*, 2015.
- [16] P. Goodwin. “The Holt-Winters approach to exponential smoothing: 50 years old and going strong”, *The International Journal of Applied Forecasting*, Vol. 19, No. Fall, pp. 30-33, 2010.
- [17] J. A. Nelder, and R. Mead, “A simplex method for function minimization”, *The computer journal*, 7(4), pp. 308-313.
- [18] 萩原純一郎, 瓜生真也, 牧山幸史, 石田基広 : “基礎からわかる時系列分析 R で実践するカルマンフィルタ・MCMC・粒子フィルタ”, 技術評論社, 東京都, 2017.
- [19] G. ペトリス, S. ペトローネ, P. カンパリオーニ : “Rによるベイジアン動的線形モデル”, 和合肇 (監訳), 萩原淳一郎 (訳), 朝倉書店, 東京都, 2013.

- [20] Statistics and Research Coronavirus (COVID-19) Vaccinations,
<https://ourworldindata.org/covid-vaccinations?county=JPN>

発表論文リスト

1. 福田太一, 伊藤宏隆, 舟橋健司: “学生你的生活及び修学データを用いたロジスティック回帰分析による要注意学生の推定”, 令和2年度電気・電子・情報関係学会東海支部連合大会, I1-3, 2020.
2. Taichi Fukuta, Hiroataka Ito, Kenji Funahshi: “Estimating the Optimal Parameters of SEIR model for Predicting the Trajectory of COVID-19 Pandemic in Japan”, IEEE-GCCE 2021, pp.567-570, 2021.