
予測・プロジェクトの評価の仕方

2021年12月20日

仲田泰祐(東京大学)

背景

- コロナ禍において感染の予測・プロジェクションは多くの人々の注目を集めた
- 予測は「こうなるのでは」を提示
 - 例: 平田研究室(名古屋工業大学)
- プロジェクションは「こういう仮定を置くとこうなるのでは」を提示
 - What if...の例
 - 「何も対策を取らないと」42万人死亡
 - 「ワクチン接種ペースが1日100万本だと」再度宣言回避
 - 「人流5割削減しないと」感染減少は起きない
 - 例: 藤井仲田研究室(東京大学)、倉橋研究室(筑波大学)、西浦研究室(京都大学)

予測・プロジェクトの評価の仕方

- ここでは、予測・プロジェクトの評価の仕方を説明
 - **政策分析としての有用性**に関する評価基準を説明
 - 筆者の中央銀行での政策分析経験に基づく
 - 中央銀行は数理モデルを政策分析に活用してきた経験が非常に長いため、他の世界での数理モデルの政策活用にも何かしらの示唆は与えうると、個人的には感じている
 - もちろん、中央銀行・経済分析の世界での評価基準を別の分野に当てはめることは必ずしも理にかなっているとは言えない事には留意
- 理解を深めるために「評価の仕方」を4つのチームの予測・プロジェクトに応用
 - 筑波大学倉橋研究室(AI-K)
 - 東京大学藤井仲田研究室(AI-FN)
 - 京都大学西浦研究室(AB-33)
 - 名古屋工業大学平田研究室(AI-H)
 - この4チームは2021年を通して定期的にモデルによる予測・プロジェクトを提示してきた

プロジェクト・予測の評価の仕方

- AI-K
 - 2020年11月から定期的に中長期プロジェクトを公表。ほぼ毎週、もしくは隔週
- AI-FN
 - 2021年1月から定期的に中長期プロジェクトを公表。9月まではほぼ毎週、9月以降隔週もしくは隔月
- AB-33
 - 2021年3月から定期的に短期実効再生産数プロジェクトを公表。ほぼ毎週
 - 6月30日から9月16日までは中期病床プロジェクトを毎週公表
 - 1月13日、6月9日には中長期プロジェクト
 - 12月8日から中長期の次世代行列の最大固有値プロジェクト
- AI-H
 - 2021年8月から定期的に中長期予測を公表。ほぼ毎週

考えられる評価ポイント

■ 再現性

- 使用しているモデルは過去のデータをどのくらい説明出来るか。プロジェクション開始時点での現実との乖離はどのくらいか
- パラメーターを毎回推定・調整しているか

■ 透明性

- モデル・パラメーターの値・コード・過去の分析は公開されているか

■ モデルの点検・過去の分析の検証・(条件付き)予測精度の記録

- 過去の予測・プロジェクションと現実の乖離を定期的に検証しているか
- 予測と現実の乖離は「予測誤差」、プロジェクションと現実の乖離は「条件付き予測誤差」で測定できる
- 条件付き予測誤差:「変数Aがこうなったら変数Bはこうなる」の「変数Aがこうなったら」の部分に変数Aの実現値を入れた時のモデルのプロジェクションと変数Bの実現値の乖離
 - 例:変数Aが人流、変数Bが新規感染者数
 - 「人流の実現値を正しく仮定していたら、感染プロジェクションはこうであった」と試算と現実の感染者数の推移の違いを計算
 - その違いが何故生まれるかを考えることで、モデルを改善するヒントを得ることが出来る

考えられる評価ポイント

- 感染症分野における学術的価値
 - 出しているプロジェクションは感染症分野の学術雑誌に掲載されうる質であるか
- 置いている仮定が妥当か否か
 - 尤もらしい仮定を置いているか。悲観的・楽観的なケースで置かれていた仮定は悲観的・楽観的と言えたか。推定していないパラメター設定の根拠の説得力
- 予測精度・条件付き予測精度

まとめ

	AI-K	AI-FN	AB-33	AI-H
再現性	高い	高い	高い	高い
透明性	高い	高い	中	高い
*コード	要請ベースで公開	HPで公開	学術雑誌掲載までは 原則非公開	要請ベースで公開
*過去の分析の記録	有	有	有	有
点検・検証	中	高い	低い	高い
*予測精度の記録	無	有	無	有
*過去の分析の検証	有	有	無	有
学術的価値	中	低い	高い	中
仮定の妥当性	評価はある程度主観的			
(条件付き)予測精度	N/A	MAE約1,600人 (1週間先・全国・21 年2月-11月)	N/A	MAE約2割 (1か月先・東京・21 年2月-6月)

参考資料: モデルの違い

	AI-K	AI-FN	AB-33	AI-H*
ワクチンの影響	Yes	Yes	Yes/No*	Yes
変異株の影響	Yes	Yes	Yes/No*	No
人流データ	Yes	Yes	Yes/No*	Yes
経済データ	No	Yes	No	No
年齢別異質性*	Yes	Yes/No	Yes/No	No
べき乗則*	Yes/No	No	No	No
医療逼迫による リスク回避行動*	No	Yes/No	No	No

*年齢別異質性: AI-FNはパラメータ調整で考慮。最近の分析では必要に応じて明示的に考慮したモデルを使用。AB-33は病床プロジェクションでは考慮、実行再生産数プロジェクションでは考慮せず

*AB-33は9月10月に一時的にワクチンの影響を考慮、8月-10月に一時的にデルタ株の影響を考慮、7月-9月に一時的に人流の影響を考慮。12月8日からの次世代行列最大固有値予測ではワクチンを考慮

*べき乗則: 内閣官房AI-Simの大澤モデル、栗原モデルではべき乗則が存在(複雑系ネットワークが考慮されているから)。倉橋の2本柱のもう一つであるマルチエージェントモデルではべき乗則を考慮

*医療逼迫によるリスク回避行動: AI-FNはYabu and Watanabe分析等を根拠にシナリオとして頻繁に考慮

*AI-Hモデルは機械学習モデルなので、過去の感染の推移が変異株・年齢別異質性・べき乗則・リスク回避行動を反映している限りにおいて、それは予測に反映される

再現性

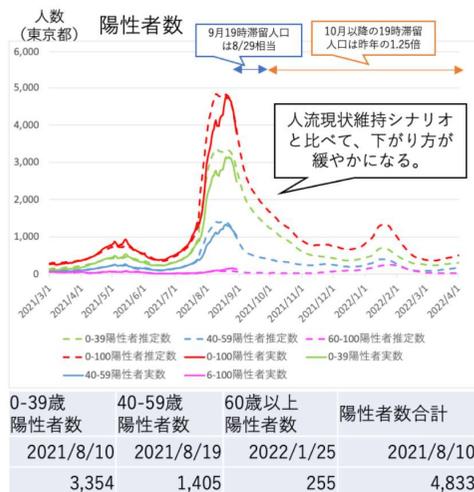
- AI-K: 高い
 - 毎回パラメターを推定。時には使用する人流データを変更
- AI-FN: 高い
 - 毎回パラメターを推定
 - パラメターがTime-Varyingなので定義的に過去のデータは完全に説明
- AB-33: 高い
 - 毎回Scalingパラメターを推定(病床プロジェクション)
 - 毎回過去の実効再生産数パスを推定(実効再生産数プロジェクション)
- AI-H: 高い
 - 毎回パラメターを推定

人流が25%増加した場合の陽性者数・重症者数の推定

➤ 19時の繁華街人流が昨年の1.25倍に増加した場合、12月以降に重症者数が急増する。

- 9/1以降の19時滞留人口は8/29相当を維持、10/1以降の19時滞留人口は昨年の1.25倍
- 180日後ワクチン感染予防効果70%、260日以降50%
- 重症者数は、東京都基準で実数から統計モデルで推定

赤：新規陽性患者数 (0歳以上)
 緑：新規陽性患者数 (0歳～39歳)
 青：新規陽性患者数 (40歳～59歳)
 紫：新規陽性患者数 (60歳～)
 実線：実測数 波線：推定数
 *数値は7日間移動平均



図：19時滞留人口と感染変化率



$$D_t = b_0 + \sum_{t=1}^2 f(b_t, x_{t+5})$$

D_t : 感染変化率 = $\sum_{t=1}^7 I_t / \sum_{t=1}^7 I_{t-1}$ (I:感染者数)

x1 : 19時の都内繁華街滞留人口
 x2 : 都外からの流入人口

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
 b0 1.011725 0.001768 572.1 <2e-16 ***
 edf Ref.df F p-value
 f(x1) 6.154 7.308 21.16 <2e-16 ***
 f(x2) 7.869 8.596 10.96 <2e-16 ***
 R-squared: 0.594
 GCV = 0.00049665
 b1 b2
 VIF 2.19 2.19

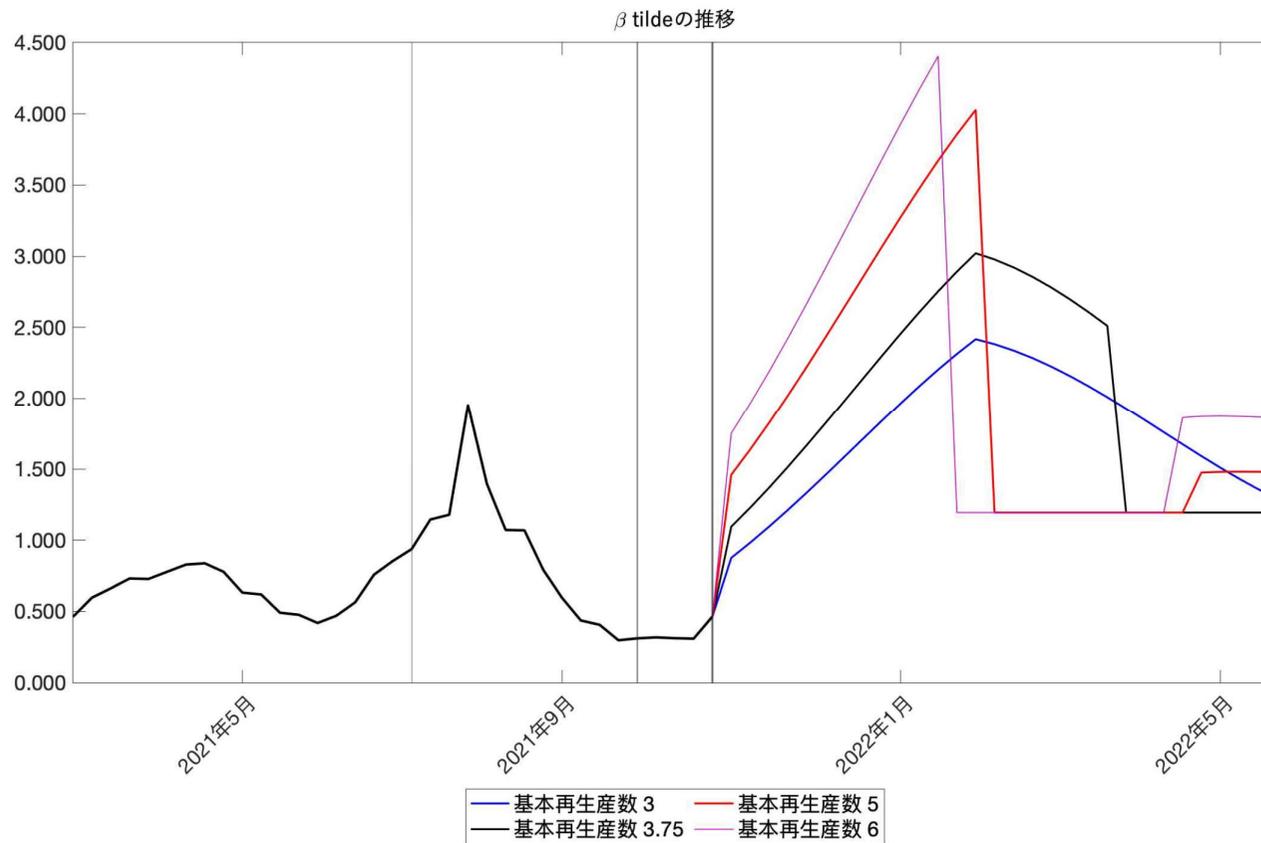
19時滞留人口が1.25倍になった場合の感染変化率を非線形モデルで推定し、シミュレーションに活用した。

1. SEIR数理モデルとAI最適化手法による感染モデル

人口流動を考慮したSEIRモデルとAI技術（進化的最適化+準ニュートン法）を用いて感染モデル推定の最適化を行うことで、0歳～39歳、40歳～59歳、60歳以上の3つの年代内および年代間での感染推定を行った。県外からの陽性患者流入者数をモバイル空間統計データ(NTTドコモ)およびLocationMind xPop*1から推定してモデルに組み込み、2021年3月1日～9月12日のデータからモデルを学習させた。

AI-FN

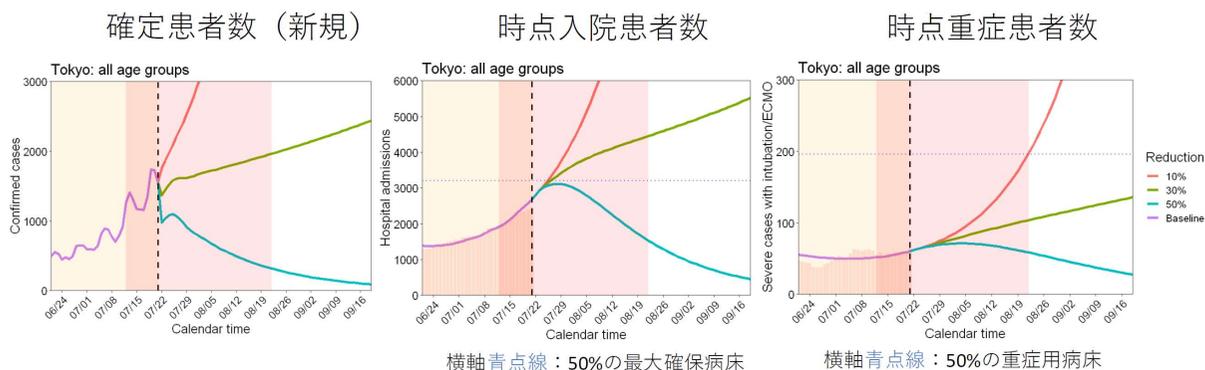
- 過去のTime-Varyingパラメターを毎回推定



https://covid19outputjapan.github.io/JP/files/FujiiNakata_Outlook_20211101.pdf

AB-33

東京都におけるトータル患者数推移に関するリアルタイムプロジェクション



東京都：
確保病床 5,967床
新計画の最大確保病床 6,404床
(7月26日時点)

東京都：
重症用病床 392床
重症者のための病床 1,207床
(7月26日時点)

Learning periodのデータ出典：
HER-SYSデータ

Learning periodは7月21日時点までの発症者データ
- 115- (9月19日までプロジェクション)

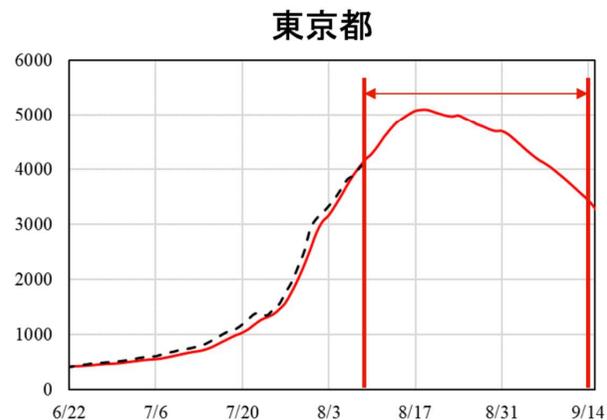
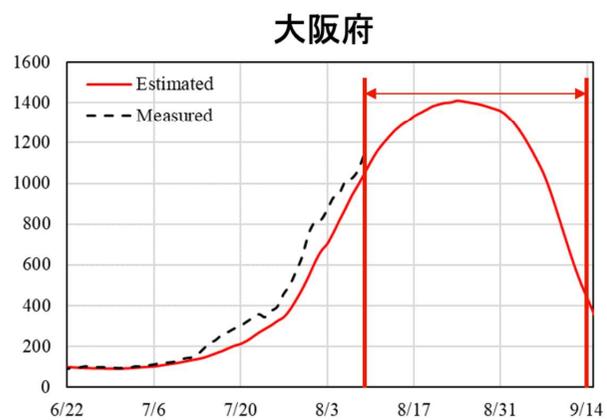
115

- Bayesian MCMCによって年齢群別患者数に適合。加えて、既知の入院率および重症化率を年齢に独立にスケールしつつ東京都データに適合。

<https://www.mhlw.go.jp/content/10900000/000812898.pdf>

AI-H

学習期間① 2020/02/15—2021/08/09
(推定期間: 8/10-9/14)



1週間ごとに学習を繰り返し予測モデルを更新。

https://www.covid19-ai.jp/ja-jp/presentation/2021_rq3_countermeasures_simulation/articles/article150/

透明性

■ AI-K: 高い

- コードは要請ベースで公開。過去の分析は本人HPと内閣官房HPに
- <https://www.covid19-ai.jp/ja-jp/researcher/setsuya-kurahashi/>

■ AI-FN: 高い

- HPでコードを公開。過去の分析は本人HPと内閣官房HPに
- <https://covid19outputjapan.github.io/JPL/>, <https://github.com/Covid19OutputJapan/Covid19OutputJapan.github.io/tree/main/archives/>

■ AB-33: 中

- コードは原則非公開(将来分析が論文として学術雑誌に掲載された場合は公開)。過去の分析はアドバイザリーボードHPに
 - コードの公開方針をご教示して頂いた西浦教授に感謝しております
- https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_00294.html

■ AI-H: 高い

- コードは要請ベースで公開。過去の分析は内閣官房HPに
- <https://www.covid19-ai.jp/ja-jp/researcher/akimasa-hirata/>

点検・検証

■ AI-K: 中

- 過去のプロジェクションと実現値の誤差を適宜検証

- https://www.covid19-ai.jp/ja-jp/presentation/2021_rq3_countermeasures_simulation/articles/article186/

■ AI-FN: 高い

- 条件付き予測精度を毎回記録・時おり過去のプロジェクションと実現値の誤差の検証

- 過去まで遡ったモデルの予測精度検証、<https://covid19outputjapan.github.io/JP/nationwide.html>

- (8月17日)過去の見通し検証、https://covid19outputjapan.github.io/JP/files/FujiiNakata_PastProjections_20210817.pdf

- (8月20日)五輪開催の感染への影響: 振り返り、https://covid19outputjapan.github.io/JP/files/FujiiNakata_OlympicsReview_Slides_20210820.pdf

- (12月14日)基本シナリオと実現値の乖離、https://covid19outputjapan.github.io/JP/files/FujiiNakata_ProjectionReview_20211214.pdf

■ AB-33: 低い

■ AI-H: 高い

- 予測精度を検証(8月15日)・過去の予測と実現値の誤差を頻繁に分析

- https://www.covid19-ai.jp/ja-jp/presentation/2021_rq3_countermeasures_simulation/articles/article175/

- https://www.covid19-ai.jp/ja-jp/presentation/2021_rq3_countermeasures_simulation/articles/article200/

AI-K

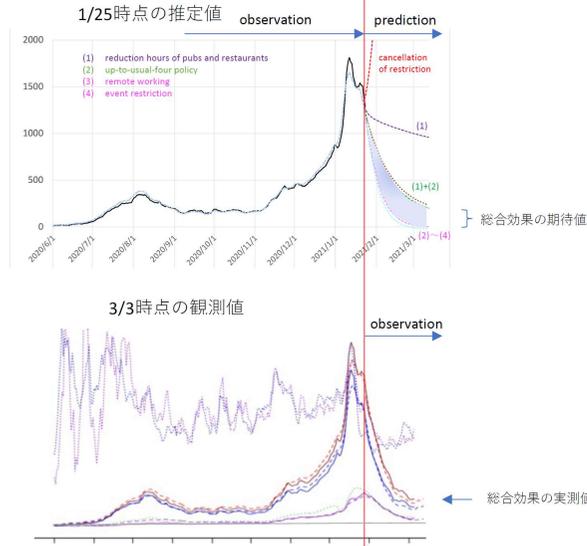
「東京都第2次緊急事態宣言効果の推定値と観測値の比較」

東京都第2次緊急事態宣言効果の推定値と観測値(2021/3/3)

- 進化的探索アルゴリズムと準ニュートン法を組み合わせた最適化法で、感染率を逆シミュレーション推定
(黒:観測値、青点線:推定値、過去データでの誤差1.2人/day)
- 東京郊外市街地モデルで推定した感染予防策の実効再生産数減少率を利用
- 基本予防策だけの場合は感染者急増 (赤点線)

感染予防策

- 飲食店利用客数を50%に制限した場合 (紫点線)
- テレワークの実施率を強化し出社率を50%とした場合 (茶点線)
- 上記+斜め席とアクリル板飛沫防止徹底で (緑点線)
- 飲食店利用客数を25%に制限した場合は (ピンク点線)
- 2+3+4の予防策を実施できた場合は (水色点線)
→政府対策の期待値

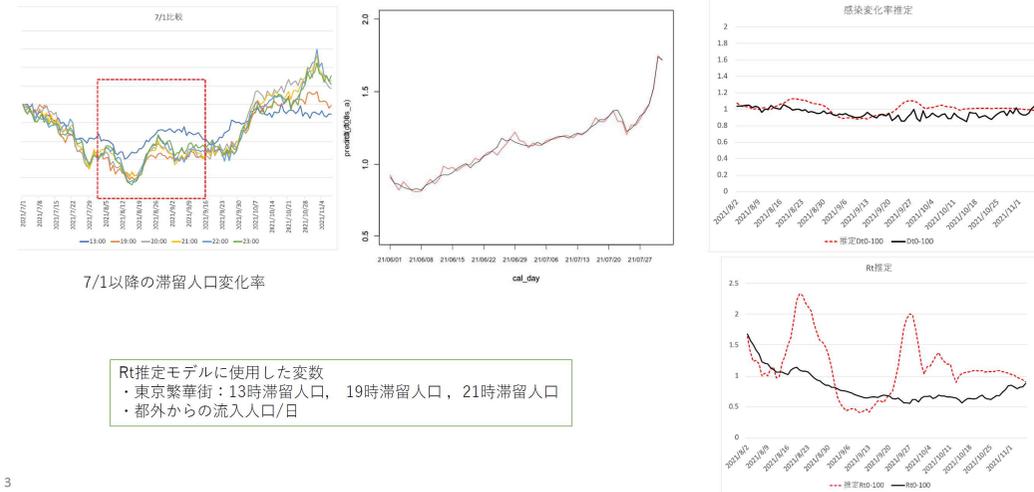


3

「人流と感染変化率との関係(東京都)」

https://www.covid19-ai.jp/ja-jp/presentation/2021_rq3_countermeasures_simulation/articles/article186/

8/1時点での予測を検証し仮説生成

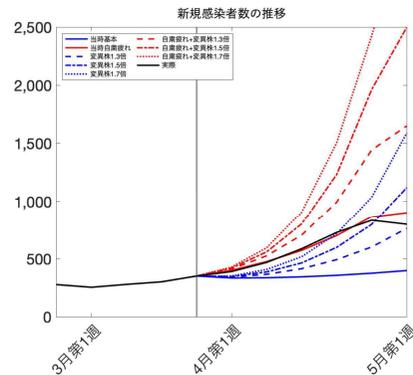


AI-FN

「(8月17日)過去の見直し検証」

https://covid19outputjapan.github.io/JP/files/FujiiNakata_PastProjections_20210817.pdf

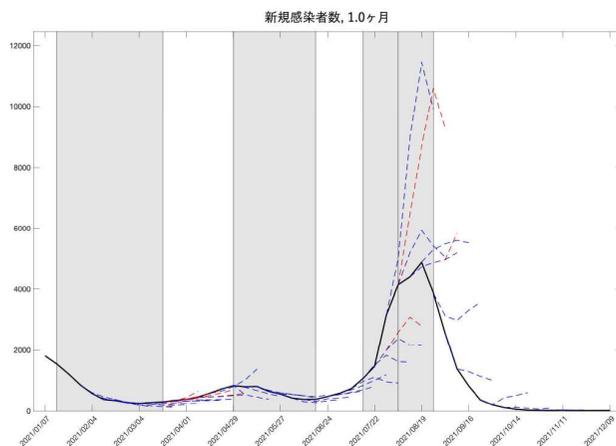
3月30日



- 3月29日の見直しではアルファ株を考慮したシナリオも提示。ここでは考慮していなかったシナリオに、現実と整合的な割合推移・1.5/1.7倍の感染力を追加した場合を分析

「(12月14日)基本シナリオと実現値の乖離」

https://covid19outputjapan.github.io/JP/files/FujiiNakata_ProjectionReview_20211214.pdf



「(8月20日)五輪開催の感染への影響:振り返り」

https://covid19outputjapan.github.io/JP/files/FujiiNakata_OlympicsReview_Slides_20210820.pdf

海外入国者の影響

- 5月21日の分析では「**悲観的に見積もって1日新規感染者数15人・重症患者数(国基準)3人**」をベースラインとして提示
 - 5月21日のベースライン分析で置かれていた仮定と現実との違い
 - 現実には五輪入国者約5万人(5月には約10万人と仮定)
 - 現実にはデルタ株の蔓延(5月の分析では考慮せず)
 - 日本で初めてのデルタ株の今後の推移を考慮した見直し:「(5月21日)インド型変異株リスク」
 - https://covid19outputjapan.github.io/JP/files/FujiiNakata_Variant_Slides_20210521.pdf
 - 初めて今後のデルタ株割合の推移・デルタ株の感染力に関する試算がADBから示されたのは6月9日
 - 現実には開催時の1日感染者数約1500人(5月の分析では約500人と仮定)
- 5月の分析を以上の条件でやり直すと、おそらく「**悲観的に見積もって1日平均新規感染者数40人・重症患者数(国基準)8人**」

海外入国者の影響

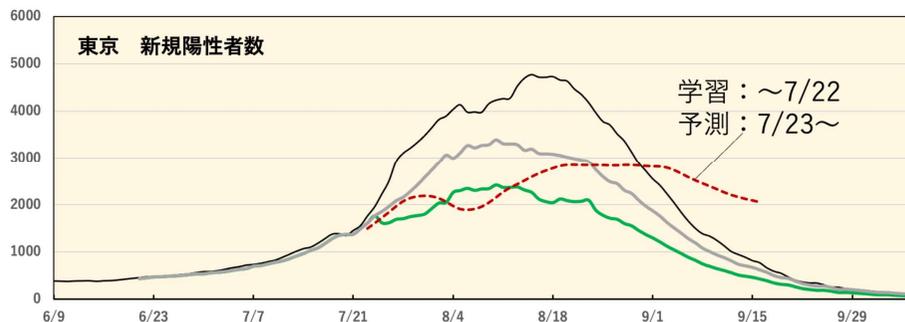
- 実際に約5万人の海外入国者によってどのくらい感染が拡大したのか?
 - きちんと検証するのは非常に難しい
 - 空港検疫ですり抜けた陽性海外入国者がどのくらい感染を広げたか
 - 日本に入国した後に、どのくらいの海外入国者が感染した日本居住者と接触して感染したか
 - そこから広がる2次感染、3次感染...
- が、公表されている五輪関係者の感染者数である程度の規模感は掴める
 - 7月1日から8月18日までの累計感染者数602人:1日平均約12人
 - この中の半数以上は日本人の委託業者であり、その中には海外入国者との接触で感染していない人もいてであろうことに留意
 - 重症患者数(国基準)0人

AI-H

「7-9月における新規陽性者数の増加と減少について」

https://www.covid19-ai.jp/ja-jp/presentation/2021_rq3_countermeasures_simulation/articles/article175/

東京における新規予測者数（報告値）、補正値の予測値との比較



○機械学習による推定値と近似直線に基づく補正値の新規陽性者数はほぼ合致。

7月の連休、お盆の活動に伴う実効再生算数の上昇が、予測以上であった。

○9月中旬以降の機械学習による予測の相違

7月22日時点で、ワクチンの効果を十分に見積もれていなかった。

考えられる理由：イスラエルと日本では、ワクチンの種類に加え、実効再生算数がゼロとなる人流増減率も異なる。デルタ株に対するワクチンの感染予防効果。

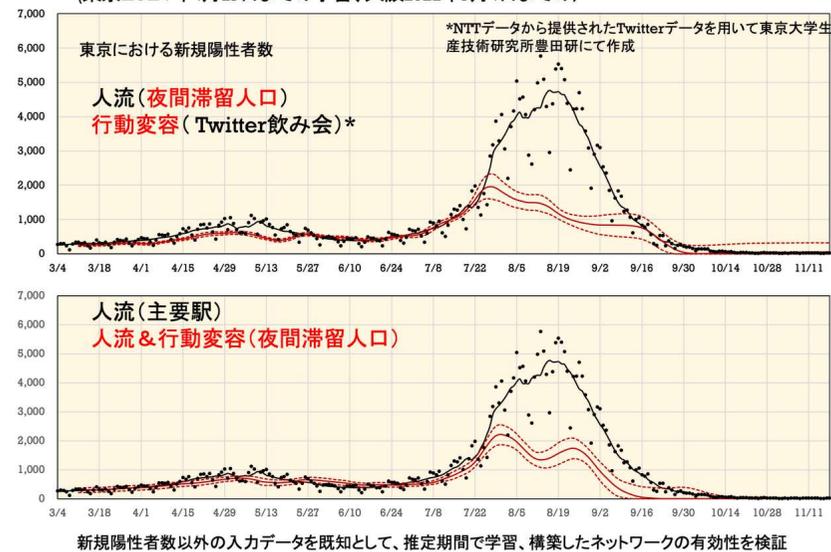
機械学習：東京(7/22まで) およびイスラエルのデータを利用。モデル(1)、(2)の併用。ワクチン効果は実際の東京のデータを使用。

「新規陽性者数の高精度化に向けて」

https://www.covid19-ai.jp/ja-jp/presentation/2021_rq3_countermeasures_simulation/articles/article200/

東京、大阪における第4, 5波の推定例

(東京2021年4月15日までの学習、大阪2021年3月4日までの)



*それ以外の入力パラメータは検証中 右軸: 駅における人流、ウイルスの感染力、Twitter(飲み会)データの規格化値

コメント

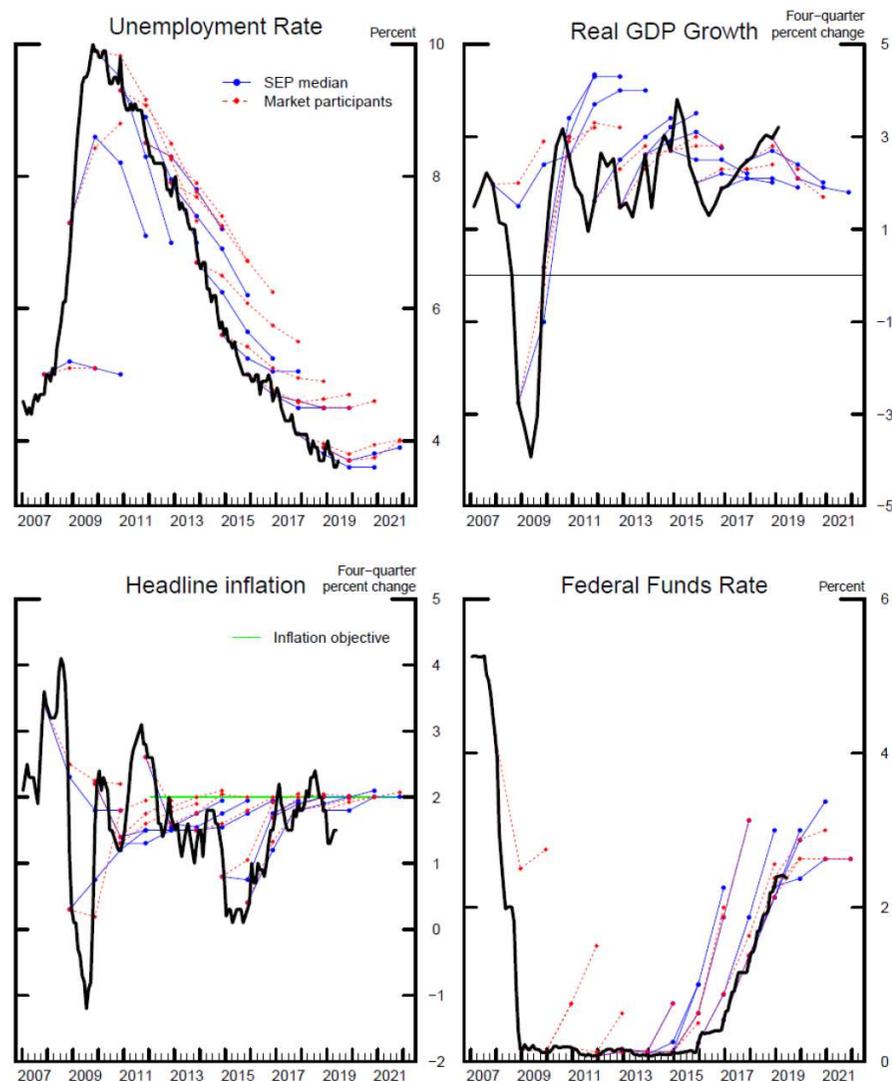
(1) 定期的なモデルの点検、(2) 過去の(条件付き)予測誤差の検証/可視化、(3) それらによって得られた学びによるモデルの改善、はモデルを中長期に渡って運用する世界では当たり前のように行われている

例: 中央銀行

中央銀行

- 予測誤差の可視化
- 図は「スパゲティーチャート」と呼ばれる

Figure 4: Forecast Revisions



<https://www.federalreserve.gov/econres/feds/files/2020065pap.pdf>

中央銀行

■ 予測の記録

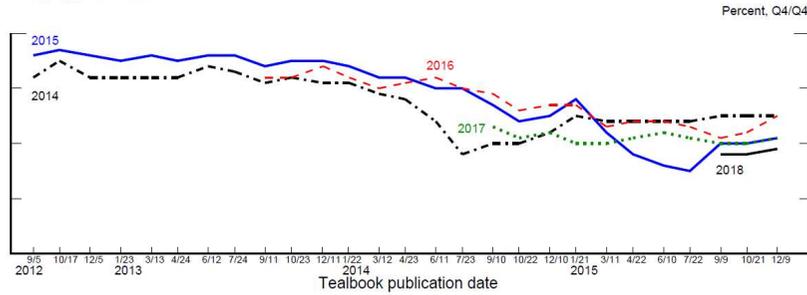
Class II FOMC - Restricted (FR)

Authorized for Public Release

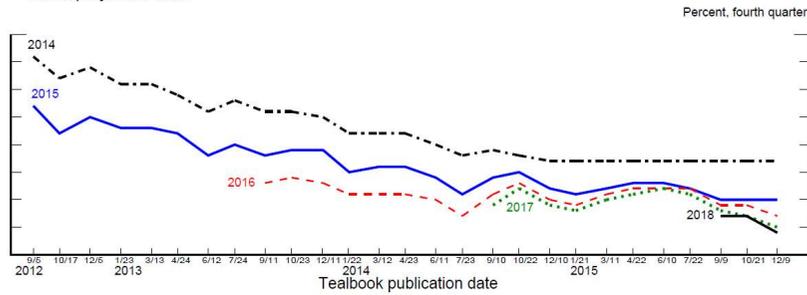
December 9, 2015

Evolution of the Staff Forecast

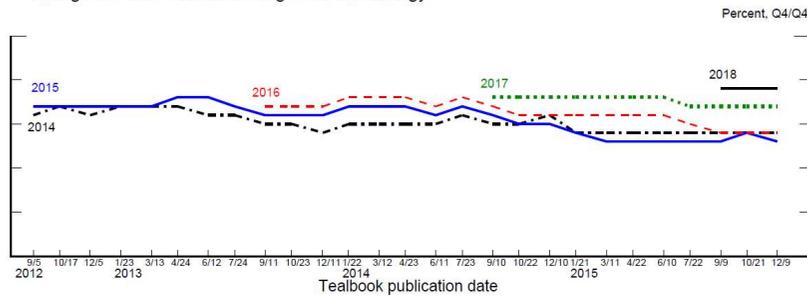
Change in Real GDP



Unemployment Rate



Change in PCE Prices excluding Food and Energy



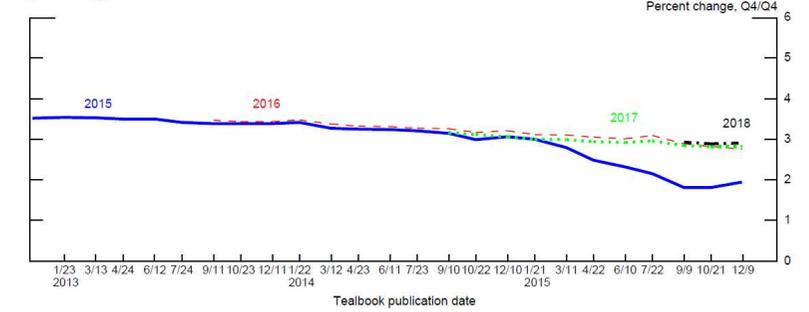
Class II FOMC - Restricted (FR)

Authorized for Public Release

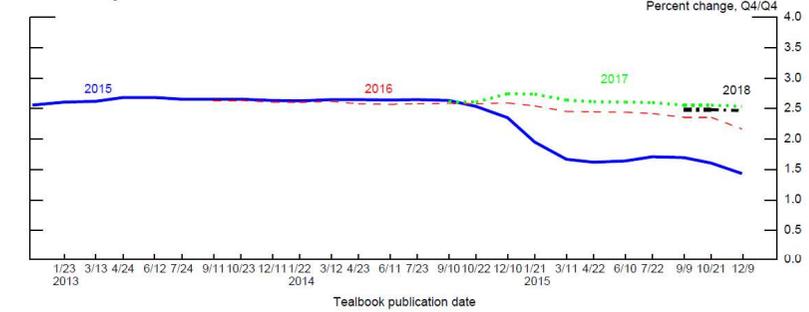
December 9, 2015

Evolution of Staff's International Forecast

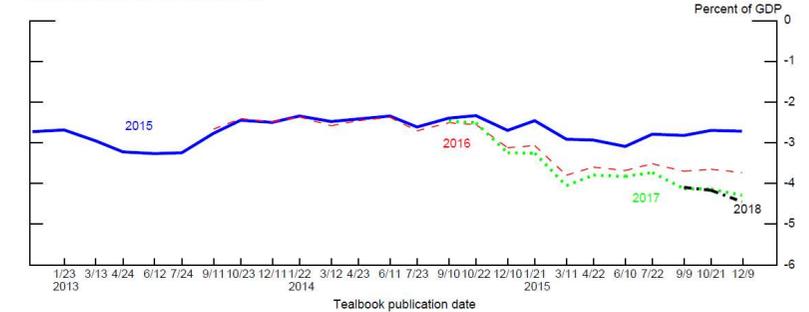
Total Foreign GDP



Total Foreign CPI



U.S. Current Account Balance

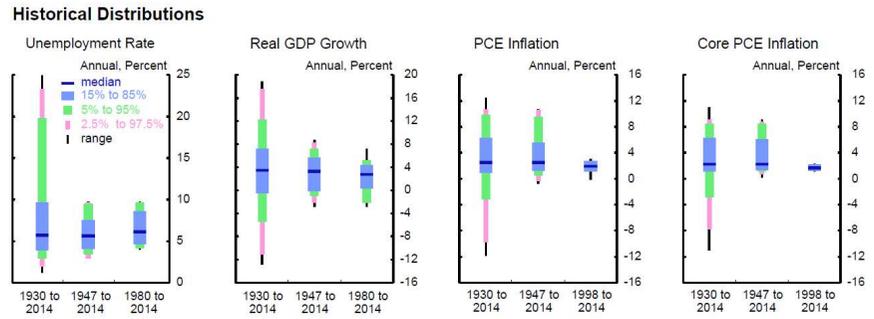
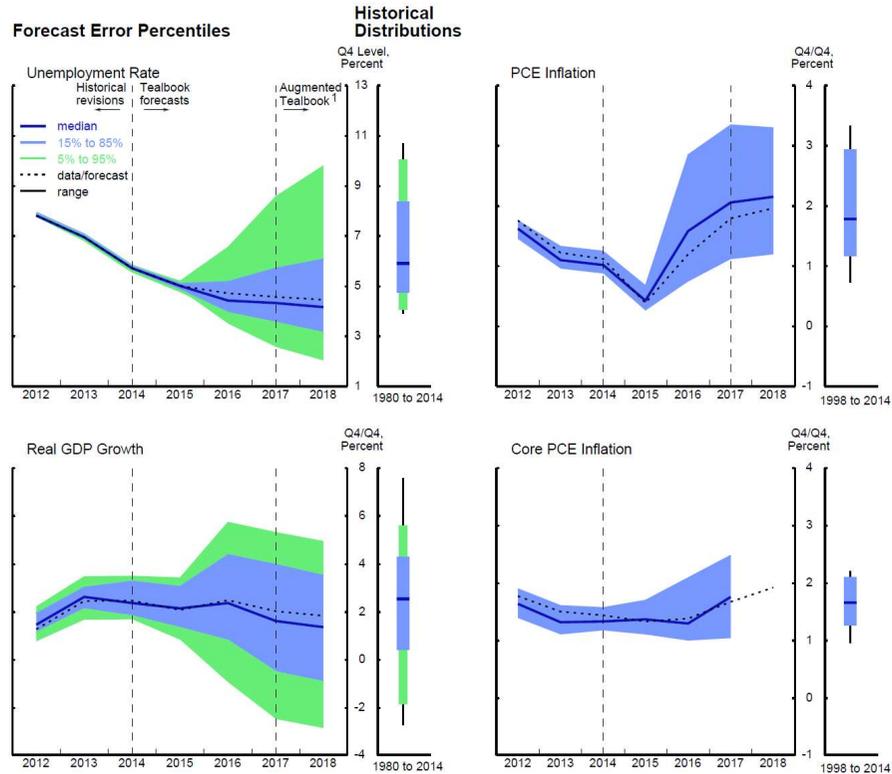


<https://www.federalreserve.gov/monetarypolicy/files/FOMC20151216tealbooka20151209.pdf>

中央銀行

- 過去の予測誤差に基づく信頼区間の計算

Prediction Intervals Derived from Historical Tealbook Forecast Errors



Note: See the technical note in the appendix for more information on this exhibit.
 1. Augmented Tealbook prediction intervals use 2- and 3-year-ahead forecast errors from Blue Chip, CBO, and CEA to extend the Tealbook prediction intervals through 2018.

<https://www.federalreserve.gov/monetarypolicy/files/FOMC20151216tealbooka20151209.pdf>

Risks & Uncertainty

中央銀行

- 定期的な戦略検証

Review of Monetary Policy Strategy, Tools, and Communications

2019-2020 Review: Overview
2020 Statement on Longer-Run Goals and Monetary Policy Strategy
Guide to changes in the 2020 Statement on Longer-Run Goals and Monetary Policy Strategy
Fed Listens
System Analytical Work
Q&As
Related Speeches

2019-2020 Review: Overview

In 2019, the Federal Reserve launched its first-ever comprehensive and public review of the monetary policy framework—the strategy, tools, and communication practices—it employs to achieve its congressionally mandated goals of maximum employment and price stability. At the time the Fed announced its review, employment and inflation were near the Fed's objectives, making it a good time to step back and consider whether the U.S. monetary policy framework could be improved to better meet future challenges.

Moreover, good institutional practice suggests that routine self-evaluation is healthy for any organization. For the Fed, however, a fresh look had become particularly important because changes in the economic environment suggested that revisions to the existing framework could be helpful to policymakers in addressing future challenges. In particular, the neutral level of the

Related Information

Fed Listens: Perspectives from the Public (PDF)

June 12, 2020

Conference on Monetary Policy Strategy, Tools, and Communication Practices (A Fed Listens Event)

June 4-5, 2019

Federal Reserve Bank of Chicago

<https://www.federalreserve.gov/monetarypolicy/review-of-monetary-policy-strategy-tools-and-communications.htm>

感染症分野における学術的価値

- AI-K: 中
 - ある程度Richなモデルを使用
- AI-FN: 低い
 - (非常に)シンプルなモデルを使用
- AB-33: 高い
 - 感染症数理モデル専門
- AI-H: 中
 - ある程度Richなモデルを使用

仮定の妥当性

- ある程度主観的に判断せざるを得ない
 - その後結局どうなったかではなく、**当時の情報**と照らし合わせて判断する必要
 - **分析の目的**と照らし合わせて評価することが必要
 - 悲観的なシナリオと楽観的なシナリオでは違う仮定が使われていて当然
 - 「悲観的なシナリオ」として提示されていたシナリオで置かれていた仮定は悲観的と言えたか？
 - 「楽観的なシナリオ」として提示されていたシナリオで置かれていた仮定は楽観的と言えたか？
 - 変異株・ワクチン接種その他様々な**現実的な要素**を考慮しているか
 - アルファ株・デルタ株の今後の推移・感染力の根拠は妥当だったか
 - ワクチン接種ペースの仮定は妥当か
 - ワクチンの効果に関するパラメターの根拠は妥当か
 - 感染予防効果減退スピードの仮定は妥当か

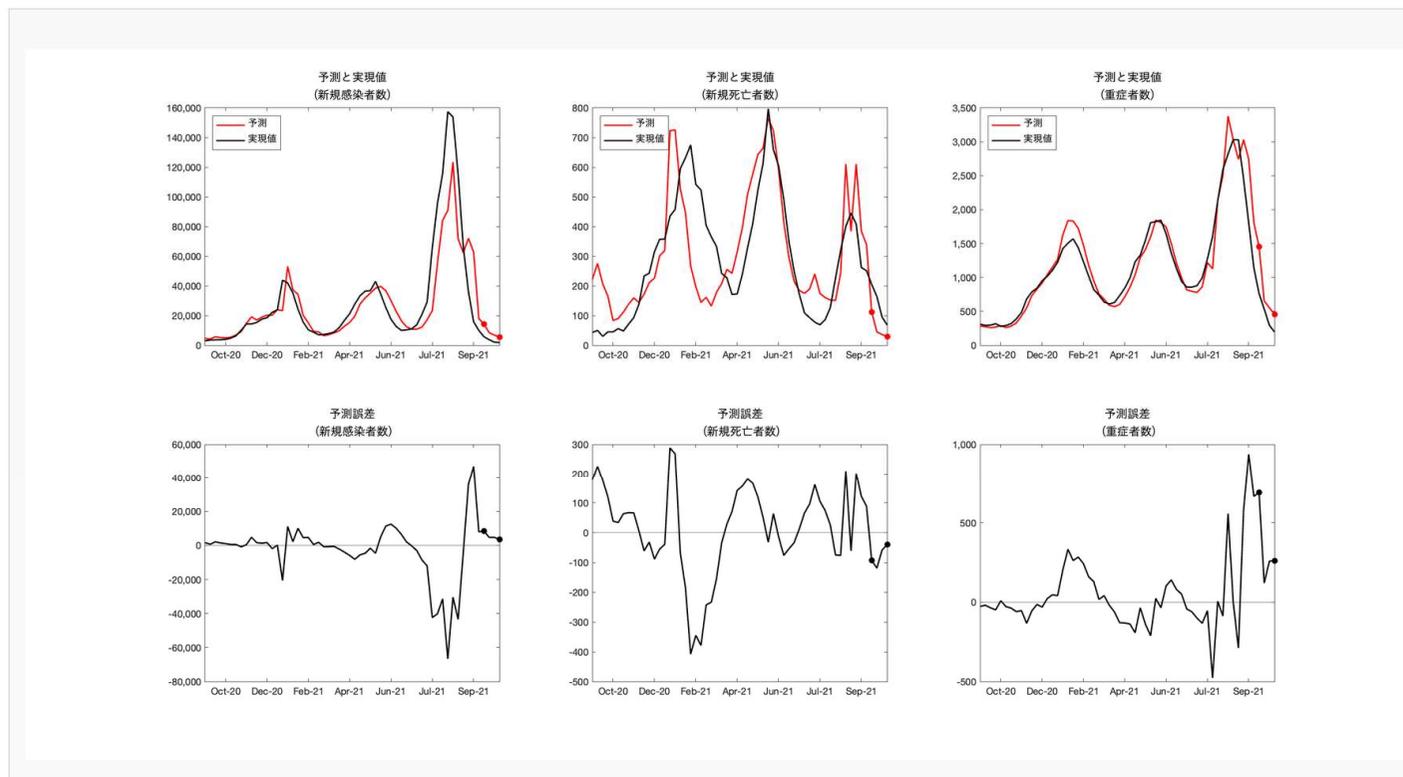
予測精度・条件付き予測精度

- AI-K・AB-33
 - 条件付き予測精度が記録されていないので、検証できず
 - 上述した通り、AI-Kはスナップショットで条件付き予測精度を何度か検証
- AI-FN
 - 更新の度に(条件付き)予測精度を記録
 - 1週間先の全国での新規感染者数: MAE(Mean Absolute Error)は約1,500人(2021年1月末から11月上旬まで)
 - 実現値の一日平均は約5,000人
- AI-H
 - 10月19日のレポートで過去の予測精度の記録を提示
 - 1か月先の東京都での新規感染者数: MAEは約2割(2021年1月末から6月末まで)

AI-FN

2. 過去まで遡ったモデルの予測精度検証

i. 1週間前に計算された現在までの予測

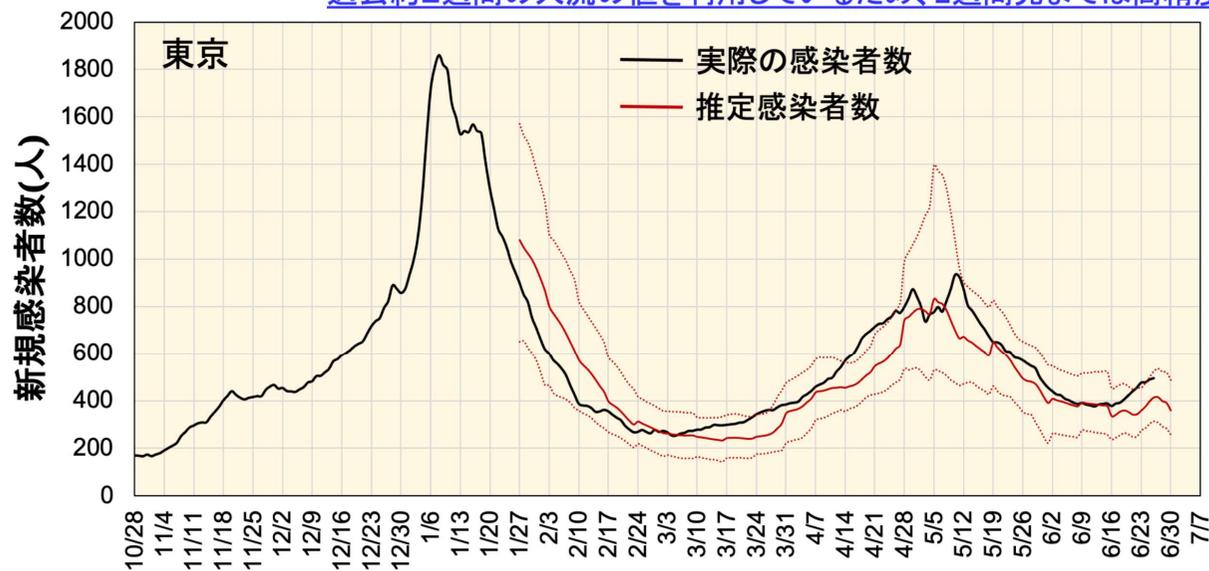


<https://covid19outputjapan.github.io/JP/nationwide.html>

AI-H

1週間ごとに学習を繰り返し予測モデルを更新。人流がわかれば、**1か月先まで**
の一日当たりの新規感染者数(1週間平均値※)を予測可能(平均20%):**第4波まで**

過去約2週間の人流の値を利用しているため、**2週間先までは高精度**



6都府県のデータを学習したため、初期値がずれる場合あり

1. E. A. Rashed and A. Hirata, "Infectivity upsurge by COVID-19 viral variants in Japan: evidence from a deep learning modeling." Int. J. Environ. Res. Public Health, 2021.

- 火曜日に分析更新・Zoom説明会 : <https://Covid19OutputJapan.github.io/JP/>
- 参考資料 : <https://covid19outputjapan.github.io/JP/resources.html>
- Zoom説明会動画 : <https://covid19outputjapan.github.io/JP/recording.html>
- 経済セミナー連載
 - <https://note.com/keisemi/n/n9d8f9c9b72af>、<https://note.com/keisemi/n/n7f38099d0fa2>
 - <https://note.com/keisemi/n/nd1a6da98f00e>、<https://note.com/keisemi/n/n430f8178c663>
- 論文 : <https://link.springer.com/article/10.1007%2Fs42973-021-00098-4>
- Twitter: <https://twitter.com/NakataTaisuke>
- 質問・分析のリクエスト等
 - taisuke.nakata@e.u-tokyo.ac.jp