

日本におけるインフルエンザ定点観測データの解析

齋藤 正也 データ同化研究開発センター 特任助教

2009年は新型インフルエンザ・パンデミックが発生した年である。日本における2009/2010年シーズンの流行は、パンデミック・ウィルスが支配的であったという分析がなされている。定点観測データによると、各県での動向は、大局的には1回の感染の波を示しているが、より詳細に調べると、小規模な感染の波が重畳されていることがわかり、複数の県で同期した、急激な感染者数の増減も見られる。単一ウィルス株が卓越していたという事実を踏まえ、このような急激な変化が外部との相互作用のみで説明できると考え、確率的なノイズ項を含むSIRモデルの解と照合することで、外部との相互作用が感染伝播に与える影響を調べる。

方法

定点観測データ

定点観測データとは、全国約5000指定医療機関による感染者数データで、都道府県別の値が国立感染症研究所から公開されている。

Fig.1は最近の3年間の北海道、宮城、三重での定点観測データである。ピークのずれや離れた地域での同期した急激な増加が見られる。また、2010/2011シーズンに限っては、感染に複数の波が見られる。このような現象はおそらく他の地域からの感染者の流入によってもたらされると考えられるが、今回の実験ではまずはこれらの変化がSIRモデルに少数回の外力を導入することで再現できるかどうかを調べる。

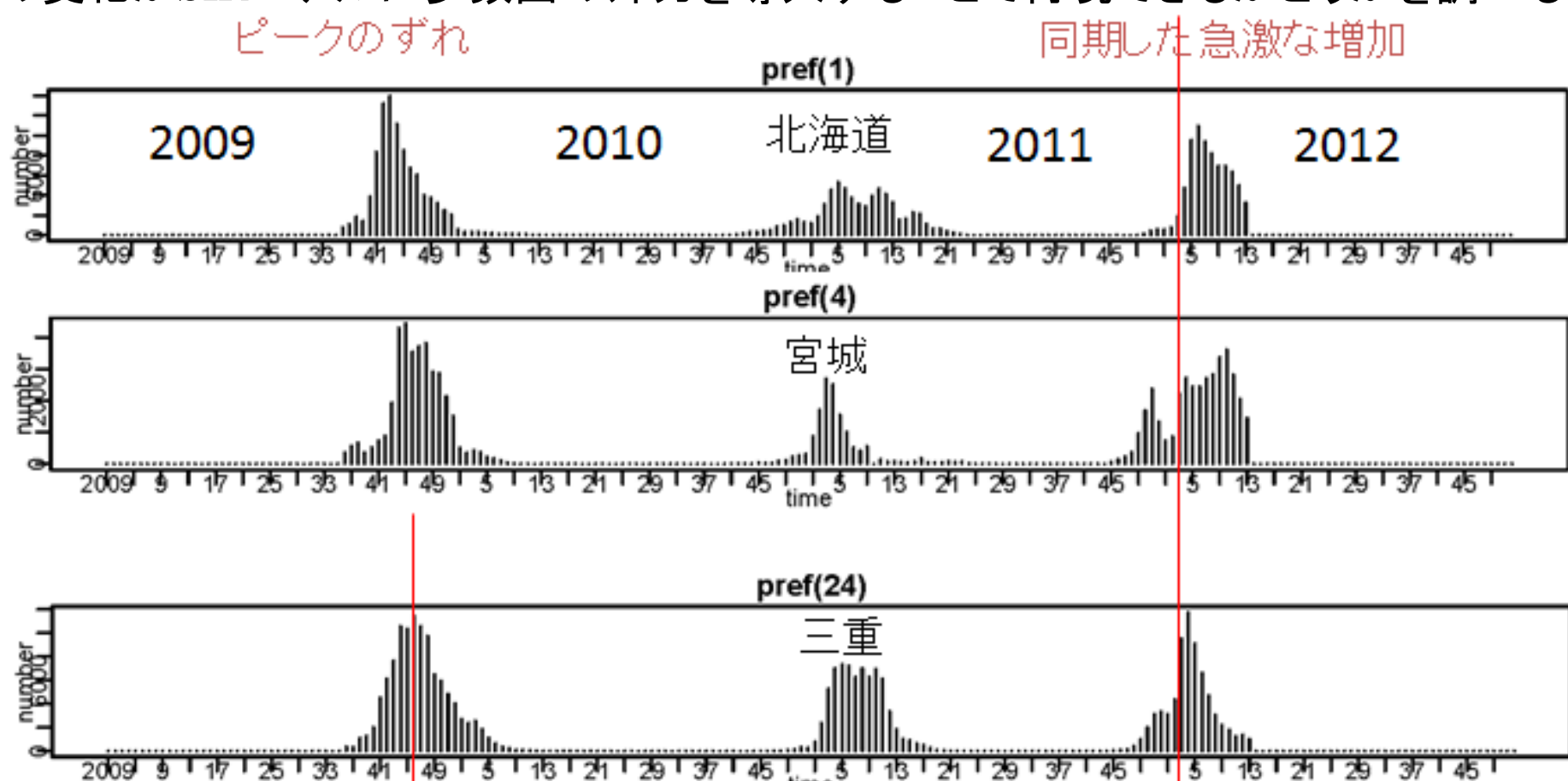


図 1 定点観測データ (2009～2012年, 抜粋)

SEIRモデル

SIRは感受性人口 S 、感染者人口 I 、除外人口 R についての常微分方程式で、

$$\dot{S} = -\lambda SI \quad \dot{I} = \lambda SI - \gamma I \quad \dot{R} = \gamma I \quad (1)$$

で定義される。 R についての方程式は (S, I) の変化に影響を与えないので、以下では無視する。定点観測データは、各県で週ごとに新たに発生した患者の標本抽出したものの件数になっていると考えられる。この件数に J_n^{obs} という変数を対応させる。SIRモデルでの(単位時間当たりの)新規感染者数は $J \equiv -\dot{S}$ で与えられる。そこで、負の対数尤度を

$$L = \sum_{n=1}^{T_{\text{obs}}} (J_n^{\text{obs}} - J_n)^2$$

で与える(添え字は観測時点の番号)。本来、母集団に対する標本の割合と観測データの取得期間(ここでは1週間)に由来する係数を J にかけろべきであるが、SIRモデルの適当なスケールンによってこの係数を消去できるので、直接比較している。

感染者数の確率的変動の導入

$$S_n = S_{n-1} - \lambda SI \Delta t \quad I_n = I_{n-1} + \lambda SI \Delta t - \gamma I \Delta t + \Delta I_n$$

$$p(\Delta I_n) = \left(\frac{T_{\text{shock}}}{T_{\text{obs}}} \right) \mathcal{N}(\Delta I_n | 0, [0.1 I_{\text{max}}]^2) + \left(1 - \frac{T_{\text{shock}}}{T_{\text{obs}}} \right) \delta(\Delta I_n - 0) \quad (2)$$

というモデルを用いる。ここで、簡単のためにSIRを1次の差分で近似した後、感染者数の確率的変動を表す項 ΔI_n を追加する表示を取っている。確率的変動を導入するねらいは、(感染者でありうる)他県の住民が当該県の住民との接触の効果を取り入れるためである。 ΔI_n が従う分布は、0でない値を取る確率が(適当に小さく取った) $T_{\text{shock}}/T_{\text{obs}}$ であることを表している。技術的には確率的変動ですべてが説明されることを避けるためでは、接触した他県民が感染者数である確率が小さいことを表現しているとも解釈できる。

パラメータと確率的変動の推定

式(1)や(2)は小さいモデルなので、パラメータと初期値を状態ベクトルに組み込み粒子フィルタで推定することが考えられる。しかし、パラメータと初期値に強い相関があるため適当な事前分布を設計することが容易でない。そこで、先に観測時系列 $(J_n^{\text{obs}})_n$ を分析してパラメータの初期推量を得ておくという方法を取る。

J_n^{obs} のピーク値を D_0 、その時刻を t_0 、 $\log J_n^{\text{obs}}$ の両裾での傾きを D_{\pm} とすると、観測データとSIRモデルの解との対応関係から以下の代数方程式が得られる。

$$D_0 \equiv -\dot{S}_0 = \lambda S_0 I_0 = \gamma \frac{S_0 I_0}{S_0 - I_0} \quad (3) \quad D_{\pm} S_0 = 2 D_0 + \gamma S_0 \log \left(\frac{S_0 (D_{\pm} + \gamma)}{D_0 + \gamma S_0} \right) \quad (4)$$

理想的には式(4)を例えばニュートン法で解くと S_0, γ が求まり、次いで式(3)から I_0, λ が求まる。しかし、実際には D_0 、 D_{\pm} を精度よく得ることが困難なため式(4)の求解に失敗する場合がある。その場合は式(3)の極めて荒い近似 $D_0 \approx \gamma$ を使う。また、解が得られても対応するSIRモデルの解は観測データをうまく再現できないことがある。そこで、解の周辺でのランダムサンプリングによって改良する。

ノイズ時系列の解釈

推定されたノイズ項大きさから、外部との相互作用がどの程度必要かを見積もろう。対象領域と接触を持つ仮想的な相手領域を考える。相手地域内の感染者 I' のうち $\varepsilon I'$ が対象地域での新規感染者発生に寄与すると考える。これとノイズ項による感染者増加を対応させると、

$$\lambda S(\varepsilon I') = \sum_{i \in T_{\text{obs}}} \delta I_i \times \delta(t - t_i)$$

両辺を積分する。このとき、 I で代用すると、 $\lambda S I = -\dot{S}$ より外部の地域との結合度を表す ε を

$$\varepsilon = \frac{\sum_{i \in T_{\text{obs}}} \delta I_i / \gamma}{\int_{-\infty}^{+\infty} I(t) dt} = \frac{\sum_{i \in T_{\text{obs}}} \delta I_i}{S(-\infty) - S(+\infty)}$$

と見積もることができる。

結果

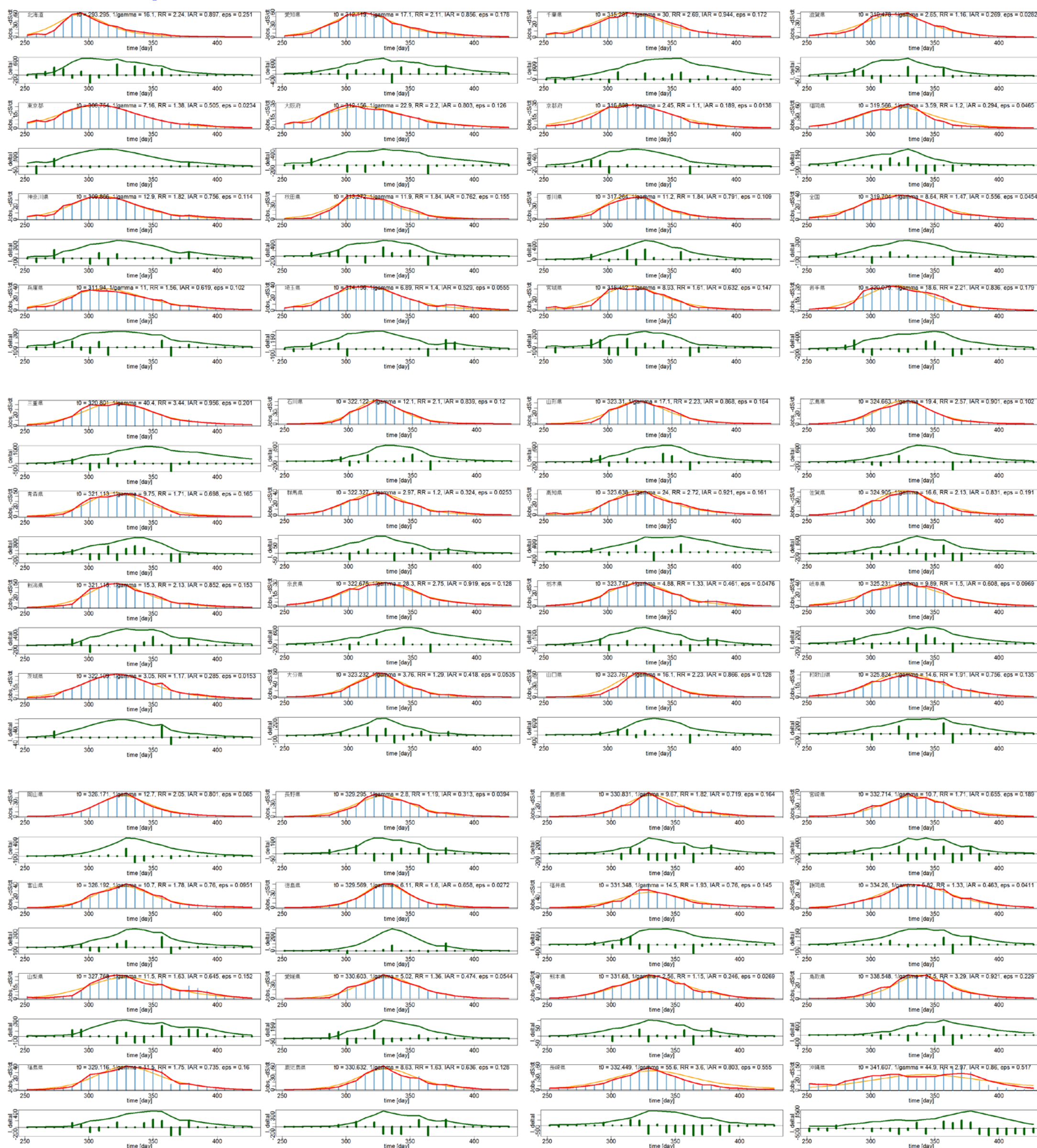


図 2 推定結果

2009/2010年シーズンでの感染動向にフィッティングを行った。全国および各都道府県において、観測データとSIRモデルの解とを比較したものを図2に示す。青が観測データ、赤がシミュレーション、緑が外力(4倍強調表示)である。パラメータとノイズ時系列 (ΔI_n) の推定値から以下のことがわかる:

- 感染のピークに約30日の開きがある。北海道以外のピークが早い都道府県は、パンデミック・ウィルスが上陸した5月上旬ごろに感染が報告されたところに対応する。北海道はさらに早く、他県とは別の感染ダイナミクスが働いている可能性がある。
- 総患者数へのノイズ項からの寄与 ε は20%程度以下である。50%を越える長崎と沖縄県では、時系列の観察からフィッティングに失敗していると考えられる。

参考文献

- 1) 廣瀬 英雄, A(H1N1)インフルエンザパンデミックシミュレーション, 日本計算機統計学会シンポジウム論文集 (23), 127-130, 2009-11-07
- 2) M. M. Saito, S. Imoto, R. Yamaguchi, S. Miyano, T. Higuchi, Estimation of Macroscopic Parameter in Agent-based Pandemic Simulation, Proceedings of FUSION 2011, Chicago.
- 3) M. M. Saito, S. Imoto, R. Yamaguchi, S. Miyano, T. Higuchi, Parallel Agent-based Simulator for Influenza Pandemic, In: Frank Dignum, Francien Dechesne, Hiromitsu Hattori, Adriaan ter Mors, Jose Miguel Such, Danny Weyn (eds.) Advanced Agent Technology; proceeding of workshop s at AAMAS 2011, LNAI 7068, pp.361-370