

と畜検査データの時系列分析に関する研究

酪農学園大学大学院  
獣医学研究科研究生

足立 泰基

獣医疫学  
指導教員 教授 蒔田浩平

2018 年度

## 目 次

目次 .....	I
凡例 .....	III
学位論文要旨 .....	V
ABSTRACT .....	IX
緒言 .....	1
第 I 章 季節自己回帰和分移動平均モデルによると畜検査データの時 系列分析法 .....	5
1. 小緒 .....	5
2. 材料と方法 .....	9
3. 結果 .....	15
4. 考察 .....	29
5. 小括 .....	33
第 II 章 と畜場における小腸廃棄数の時系列分析を利用した豚の抗酸 菌症農場アウトブレイク迅速検出法 .....	35
1. 小緒 .....	35
2. 材料と方法 .....	36
3. 結果 .....	41
4. 考察 .....	60
5. 小括 .....	63
第 III 章 豚エキノコックス症の農場アウトブレイク検出のため のゼロ過剰カウントデータを用いた two-part モデルに基づくと畜検査 データの時系列分析法 .....	65
1. 小緒 .....	65

2. 材料と方法.....	67
3. 結果.....	74
4. 考察.....	104
5. 小括.....	112
第 IV 章 総合考察.....	113
謝辞.....	116
参考文献.....	117

## 凡 例

ACF	: Autocorrelation Function (自己相関関数)
ADF 検定	: Augmented Dickey-Fuller 検定 (拡張ディッキーフラー検定)
AIC	: Akaike's Information Criterion (赤池情報量規準)
ARIMA モデル	: Autoregressive Integrated Moving Average モデル (自己回帰和分移動平均モデル)
FFT	: Fast Fourier Transform (高速フーリエ変換)
GLM	: Generalized Linear Model (一般化線形モデル)
HNB 分布	: Hurdle Negative Binomial 分布 (ハードル負の二項分布)
HP 分布	: Hurdle Poisson 分布 (ハードルポアソン分布)
MAIC	: <i>Mycobacterium avium-intracellulare</i> complex (マイコバクテリウム・アビウム・イントラセルラ ーエ複合体)
MAPE	: Mean Absolute Percentage Error (絶対平均比率誤差)
PACF	: Partial Autocorrelation Function (偏自己相関関数)

- SARIMA モデル : Seasonal Autoregressive Integrated Moving Average  
モデル(季節自己回帰和分移動平均モデル)
- SEP : Swine enzootic pneumonia  
(豚流行性肺炎、豚マイコプラズマ肺炎)
- ZINB 分布 : Zero Inflated Negative Binomial 分布  
(ゼロ過剰負の二項分布)
- ZIP 分布 : Zero Inflated Poisson 分布  
(ゼロ過剰ポアソン分布)

## 【学位論文要旨】

と畜検査データとは、食肉衛生検査機関の獣医師であると畜検査員がと畜場に搬入された家畜を検査し、その結果をまとめたものである。と畜検査データは、衛生的な食肉生産促進のための生産者へのフィードバック事業や、調査研究に利用されてはいるものの、数値の羅列やグラフ表示を行う以上の処理が行われている例は非常に少ない。しかし、羅列された数値や折れ線グラフの傾きのみから飼養条件の悪化や改善の有無を客観的に判断することは困難である。すなわち、生産者、獣医師や行政官が合理的に意思決定するための指標を与える統計学的手法が必要とされていた。

と畜検査データは時間とともに出現する時系列データであり、時系列分析に適していると考えられる。しかし、時系列分析には種々の方法があり、データの分布に応じた適切なモデルを用いる必要がある。そこで本研究は、データの統計学的分布に応じてモデルを使い分け、網羅的に分析できる手法を見出すことを目的とした。第一章では、常時発生がみられる疾病の分析方法として、代表的なガウス型時系列モデルである季節自己回帰和分移動平均モデル(SARIMAモデル)によると畜検査データの分析方法と、これを利用して農場における衛生状態や対策の効果を評価する方法について述べた。第二章では、発生が稀な疾病の時系列分析を行う方法として、周期的要素について擬似ポアソン回帰モデルを推定し、その残差についてARIMAモデルを推定したのち、両モデルを加える方法について述べた。第三章では、発生が稀な疾病の時系列分析を行うもう一つの方法として、two-partモデルを用いた分析方法について検討した。

第一章では、SARIMAモデルを用いてと畜検査における廃棄率データの異常な増加や減少を検出できる好適条件を検討した。寄生虫性肝炎による肝臓廃棄率が30%以上に増加した後、駆虫剤の投与により1.8%まで低下した1農場のデータを用い、廃棄率の増加および減少を検出できる最大の信頼水準を調べたところ、85%が至適条件であった。さらに、豚で常時発生がみられる4疾病について10農場からの搬入豚の検査データを用い、SARIMAモデルの当てはまりと対照モデルである指数平滑法の当てはまりを比較した。40モデル中39モデルで対照モデルより逸脱度が小さく、27モデルで有意であった。本法は生産者に対して農場の衛生状態の合理的な判断指標を提供することによって、衛生的な食肉生産を促進するツールとして利用可能であることが明らかとなった。

第二章では、人獣共通感染症である豚の非定型抗酸菌症による毎日の肝臓廃棄数データを時系列分析し、農場アウトブレイクの疑いがある生産者からの搬入の有無を検出する方法について検討した。豚の非定型抗酸菌症は、比較的稀な疾病であり、廃棄数がゼロであることも少なくないため、第一章の分析方法は適さない。そこで、擬似ポアソン回帰モデルとARIMAモデルを組み合わせる方法を試みた。10年間の日次廃棄数値のうち最初の8年分のデータを用いてモデルを作成し、残りの2年分の日次廃棄数がモデルによる期待値の95%信頼限界を超えた日の搬入生産者を調べたところ、当該搬入日には年次廃棄数集計値の廃棄率から見て抗酸菌症の農場アウトブレイクが疑われる生産者による搬入が必ず行われていた。すなわち、本モデルにより発生が稀な疾病の時系列分析を行い、農場アウトブレイクを検出することが可能となった。

第三章では、人獣共通感染症である豚のエキノコックス症による月次肝臓廃棄数データを時系列分析し、農場アウトブレイクを検出する方法について検討した。北海道東藻琴食肉衛生検査所では、月次集計データにおける廃棄率が1%を超えるか、3か月連続でエキノコックス症による廃棄が認められた場合に、対策を促す情報提供を行っている。しかし、アウトブレイクの定義を考慮し、判定は統計学的に算出された期待値との比較によって行われるべきである。ここで問題となるのは、エキノコックス症は比較的稀な疾病であるために、ゼロデータが多く、毎月のルーチン業務で用いるには二章の方法のような煩雑な方法は適さないことである。そこで、two-partモデルによって計算されたパーセンタイル値をアウトブレイク検出の基準とすることを検討したところ、従来の基準よりも迅速なアウトブレイク検出が可能であることが明らかとなった。すなわち、本モデルにより発生が稀な疾病の時系列分析をワンステップで行い、アウトブレイクを検出することが可能であることが明らかとなった。

以上をまとめる。廃棄率が極端に0%や100%付近に偏っていなければ、ロジット変換した廃棄率は、正規分布することを仮定できるため、第一章に述べたSARIMAモデルを用いた分析を行うのが適切と考えられる。ゼロの多いデータについては、第二章の方法と第三章の方法を用いることができるが、ルーチンの検査業務で用いることを考慮すると、ワンステップでモデリングが可能なtwo-partモデルの方が更新を容易に行うことができ、実用的であろう。すなわち、廃棄数がゼロであることが稀な疾病については、第一章のSARIMAモデルで分析し、ゼロが多い場合はtwo-partモデルで分析することで、いかなる疾病に対しても客観的指標に基づく情報提供を可能にする道筋を付ける

ことが出来た。

## 【ABSTRACT】

Meat inspection data summarizes and integrates the results of meat inspections at slaughterhouses conducted by official veterinarians with local government meat inspection centers. Although meat inspection data has been utilized for surveys and research and fed back to producers to improve hygiene in meat production, few cases have been reported in which anything more than listing or drawing graphs of meat inspection data is performed. However, it is difficult to determine objectively whether livestock breeding conditions tend to become better or worse based on the enumeration of figures or quantity of change in graphs. That is, statistical methods that provide criteria on which the decisions made by producers, veterinarians, and administrators are based are needed.

Time series analysis is a suitable method to analyze meat inspection data as it comprises sequential time series data. Since time series analysis involves a variety of methods, to analyze meat inspection data, it is necessary to select a proper method based on the statistical distribution of data. Therefore, the purposes of the present study were to identify the optimal method for using multiple models properly according to the distribution of data and to analyze data with different types of distributions comprehensively. Chapter I describes the time series analysis method for meat inspection data using the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, which is a typical Gaussian time series model for the analysis of diseases regularly seen that describes methods for estimating the sanitation or effectiveness of hygienic measures at farms. Chapter II describes the method for analyzing diseases rarely seen by estimating, and subsequently merging, a quasi-Poisson regression model for cyclic components and the ARIMA (Autoregressive Integrated Moving Average)

model for residuals. Chapter III describes another method for analyzing diseases rarely seen using the two-part model.

In Chapter I, investigations are carried out to obtain the favorable conditions for the SARIMA model to detect anomalies in the condemnation rate. The data of one farm, in which the condemnation rate of liver due to parasitic hepatitis was elevated above 30% and reduced to 1.8% by anthelmintic administration, was used to determine the maximum confidence level value able to detect the elevation and reduction of the condemnation rate. As a result, the optimum confidence level value was 85%. In addition, the degree of fit between the condemnation rates and the models estimated using the SARIMA model and those estimated using the exponential smoothing method as a control were compared for four diseases regularly seen in swine from 10 producers. The deviance for 39 of 40 models was smaller than the control, and statistical significance was shown for 27 of 40 models. It became clear that this method can provide an index for determining the sanitation of farms for producers and be utilized as a tool to promote hygienic meat production.

In Chapter II, a time series analysis of the daily number of liver condemnations due to atypical mycobacterial disease in swine was conducted and the optimal method to determine whether swine were brought to the slaughterhouse on a specific day from the farm at which the mycobacteriosis outbreak was suspected was investigated. Because atypical mycobacteriosis in swine is relatively rare and the number of condemnations due to atypical mycobacteriosis is often zero, the method described in Chapter I was not suitable for a time series analysis of the disease. Accordingly, the method to merge the quasi-Poisson regression and ARIMA models was tested. The model was developed using daily inspection data from the first 8 of 10 years. Then it was investigated which producers brought pigs to slaughterhouses when the daily number of condemned livers exceeded the 95% confidence

interval of the expected values calculated by the model during the remaining two years. Subsequently, on the day on which the confidence interval was exceeded, the producers bringing pigs to the slaughterhouses always included the producers at farms at which farm-level outbreaks were highly suspected because of the high annual rates of condemnation due to mycobacteriosis. The above results revealed that the time series model enabled the detection of farm-level outbreaks using a time series analysis of rarely seen diseases.

In Chapter III, the method to detect farm-level outbreaks of swine echinococcosis, which is one of the most lethal zoonoses, using a time series analysis of the number of monthly condemned livers due to echinococcosis was investigated. Hokkaido Higashi-Mokoto Meat Inspection Center provides farms with a brochure detailing preventive measures to take when the condemnation of livers due to echinococcosis exceeds 1% of carcasses or when more than one pig carcass is condemned for three consecutive months. However, considering the definition of an outbreak, the determination must be made based on a comparison between the observed and expected values calculated using statistical methods. The problems are that the monthly echinococcosis data for each farm contains many zeros because swine echinococcosis is a rare disease, and that the method described in Chapter II is too complex to perform routine modeling every month. To solve these problems, the method for utilizing the percentiles calculated by the time series model based on a two-part model as criteria to detect outbreaks was investigated. As a result, the time series model was able to detect outbreaks earlier than the conventional criteria. Therefore, the model described could perform a time series analysis of a rare disease in a single step and detect farm-level outbreaks.

In conclusion, the SARIMA model described in Chapter I is suitable for analyzing condemnation rate data that do not deviate excessively around 0% or 100%, as the logit-transformed condemnation rate is assumed to be normally distributed. These results suggest

that data containing many zeros can be analyzed using the methods described Chapters II and III. To analyze data in routine meat inspections, the two-part model described in Chapter III is more suitable because it is easier to renew models that are estimated in a single step.

Therefore, the present study paves the way to enable providing administrative information regarding various diseases based on objective criteria using the SARIMA model for diseases in which the data rarely contain zeros and the two-part model for diseases in which the data contain many zeros.

## 緒 言

と畜場に搬入された家畜は、食肉衛生検査機関の獣医師であると畜検査員によって一頭ずつ検査される。疾病にかかり、若しくは異常があり、食用に供することができないと検査で認めた時等は、と殺禁止・全部廃棄・一部廃棄のいずれかの措置が取られる。このと畜検査の結果をまとめたものが本研究で言うのと畜検査データであり、本研究は、と畜検査データを有効利用するための分析方法に関するものである。

衛生的な食肉生産を促進するため、と畜検査データを希望する生産者に還元する事業が複数の自治体で行われており、これらの事業は、データ還元事業やフィードバック事業などの名前と呼ばれている。国内のと畜検査データを利用した先行研究もみられる[1, 35, 45, 57]。しかし、これらの研究においてと畜検査データは、病理学的検査結果や血液生化学的検査結果を補完するために用いられており、羅列するもしくはグラフ表示する以上の処理がなされていない。

一方、海外の先行研究においては、近年統計学的処理を行う試みが進んでいる。Sanchez-Vazquezら[55, 56]は、局所回帰による季節トレンド分解(Seasonal-trend decomposition based on loess: STL)によると畜検査データの分析方法について報告した。Neumannら[50]は、自己回帰和分移動平均モデル(ARIMAモデル)を用いたと畜検査データの解析方法を報告した。さらに、Vialら[64]は、一般化加法モデル(Generalized Additive Model)を用いたと畜検査データの解析方法について報告した。しかし、これらの報告[50, 55, 56, 64]はいずれも行政区画内の疾病や廃棄の発生数の経時的な傾向と周期性を分析すること

が目的であり、各農場における飼養状況の悪化や改善を判断し、農場アウトブレイクを検出することを目的としていない。

上述のとおりと畜検査データを扱う国内外の先行研究の中には、と畜検査データの悪化や改善を判断するための分析方法について扱う研究は見られない。たとえば薬効試験の場合には、適切な対照群から取得したデータと処置群から取得したデータを統計学的検定により比較して客観的な判断を行うことができるが[48]、と畜検査データの場合は、介入研究を実施しないかぎり、比較可能な対照群といえるものは存在しない。一般に、生産者、獣医師あるいは行政官は、と畜検査データに示された臓器の廃棄率から飼養条件の悪化または改善の有無の判断を試みるが、羅列された廃棄率(数)データやグラフを俯瞰するだけで以上のような結論を下すのは、通常困難である。したがって、と畜検査データをより多くの者が有効に利用するためには、廃棄率の有意な増減の有無を判断するための客観的な指標を与える統計学的手法が必要である。

時系列分析は、時間とともに出現するデータの列を分析するための統計学的処理方法であり、過去のデータの列をもとに数学モデルを推定することにより将来の期待値と分布を予測することができる。さらに計算された分布を最新の観測値と比較することによって、観測値と期待値の乖離に有意性があるか否かを判断することもできる。すなわち、飼養条件の悪化や改善によって家畜の健康状態に変化が起き、と畜検査結果に変化が起きる[1, 35, 45, 57]ことから、と畜検査データの観測値と期待値の乖離の有意性の有無を調べることにより、飼養条件の悪化や改善を推測し、疾病アウトブレイクの有無を推定することができると考えられる。

時系列分析法には種々の方法があり、用いるデータの分布に応じて使い分けられるべきである。本研究の最終的な目的は、データの分布に応じて複数の方法を使い分けることにより、と畜検査データを網羅的に時系列分析できることを示すことであり、本論文では当該目的を達成するために検討した複数の方法について方法ごとに章分けし、述べることとした。と畜検査では日々の検査で常時見られる疾病と稀にしか見られない疾病があり、常時見られる疾病の廃棄率は、適切な変数変換を行うことにより正規分布とみなすことができる。このような、正規分布とみなすことができる廃棄率データに適合するガウス型時系列モデルを用いた分析法について第 I 章で扱った。一方、稀にしか見られない疾病の廃棄率は頻繁に 0% となり、正規分布への変数変換は困難である。稀な事象の統計モデル化では、通常ポアソン分布のような離散分布のあてはまりがよい[14]ことから、離散変数である廃棄率を確率変数としてモデリングを試みたが、過剰なゼロデータのためにポアソン分布を用いたモデルでは過分散となり、適切なモデルが得られなかった。そこで、離散分布モデルの誤差部分に第 I 章で扱ったガウス型時系列モデルをあてはめる方法と、異なる二種類の離散分布を混合させた two-part モデルを用いる方法を検討した。本論文では、離散分布モデルの残差部分にガウス型時系列モデルをあてはめる方法について第 II 章で扱い、異なる二種類の離散分布の混合モデルをあてはめる方法について第 III 章で扱った。各章の内容について以下に簡単に述べる。

第一章では、常時発生がみられるいくつかの疾病による臓器廃棄率を代表的なガウス型時系列モデルである季節自己回帰和分移動平均モデル(SARIMA モデル)によってモデル化する方法と、これを利用して

農場における衛生状態や対策の効果を評価する方法について述べた。

第二章では、発生が稀な疾病の一つである豚非定型抗酸菌症による廃棄数を題材とした。非定型抗酸菌症による肝臓廃棄数の周期的変動要素について最初に擬似ポアソン回帰モデルを推定し、次にそれを観測データから差し引いた残差について ARIMA モデルを推定し、これら両モデルの和を取ることによって最適なモデルを得る方法について述べた。次に、得られたモデルを用いてと畜場に搬入する生産者の中に非定型抗酸菌症のアウトブレイク発生生産者が含まれているかどうかを判定する方法について述べた。

第三章では、発生が稀な疾病である豚エキノコックス症による廃棄数を題材とした。本章においては、2種類の離散分布が混合された two-part モデルを一回の推定でモデリングする方法を検討した。two-part モデルはゼロが多い過分散データのモデリングに適しているが [36, 47]、two-part モデルを時系列分析で用いることはほとんど行われていない。本章では two-part モデルを用い、豚エキノコックス症による月次廃棄数の時系列モデリングを用いて豚エキノコックス症の農場アウトブレイクの判定に用いる方法について述べた。

# 第 I 章 季節自己回帰和分移動平均モデルによると畜検査データの の時系列分析法

## 1. 小 緒

時系列分析は、時間とともに出現するデータの列を分析する方法であり、指数平滑法[24,69]や自己回帰移動平均モデル(ARMA モデル)[11]などのモデルで分析が行われる。

指数平滑法は(1)式であらわされる。

$$\hat{y}_{t+1} = \beta \times y_t + (1 - \beta) \times \hat{y}_t \dots (1)$$

ここで  $y_t$ : 時点  $t$  における観測値       $\hat{y}_t$ : 時点  $t$  における予測値       $\beta$ : 平滑化定数

この方法は一つのパラメータ ( $\beta$ ) を推定するだけでよいため、簡易性の面からよく用いられている。一方、ARMA モデルは自己回帰モデル(AR モデル)と移動平均モデル(MA モデル)の和であり、多様なモデリングが可能である。しかし、ARMA モデルは傾向変動や季節変動を含むデータをモデル化することに難がある[14]。ARIMA モデル[11]は傾向変動を含む時系列データにも対応でき、SARIMA モデル[11]は、通常の ARIMA モデルと季節階差に関する ARIMA モデルを合わせたものであり、季節変動があるものに用いられる。

$p$  次の AR モデルを AR( $p$ )モデルと呼び、以下の(2)式であらわされる[14]。

$$y_t = m + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t, \quad a_t: N(0, \sigma^2) \dots (2)$$

ここで、 $m$  は定数、 $\phi_1, \phi_2, \dots, \phi_p$  は自己回帰項の係数であり、 $a_t$  は、正規分布する誤差項であり、 $\sigma^2$  は、誤差項の分散である。また、 $q$  次の MA モデルを MA( $q$ )モデルと呼び、以下の(3)式であらわされる[14]。

$$y_t = m + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}, \quad a_t: N(0, \sigma^2) \dots (3)$$

ここで、 $\theta_1, \theta_2, \dots, \theta_q$  は、移動平均項の係数である。

AR( $p$ )モデルと MA( $q$ )モデルの和である ARMA( $p, q$ )モデルは以下の(4)式であらわされることとなる[14]。

$$y_t = m + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}, \quad a_t: N(0, \sigma^2) \dots (4)$$

時系列分析では、時系列データの異なる時点間の関係、たとえば  $k$  時点離れた時系列データの 2 つの系列  $y_t$  と  $y_{t-k}$  の関係を記述し、分析する。この際、2 つの系列がどれだけ離れているかを表す  $k$  を「ラグ」と呼び、このラグを以下に定義するラグ演算子( $B$ )を用いて記述することにより、複雑でわかりにくいモデルを簡潔に記載することができる。以下の(5)式のように、演算子  $B$  は、時系列データに作用して 1 時点ずれた時系列データを返す[14]。

$$By_t = y_{t-1} \dots (5)$$

$B^k$  とは、ラグ演算子を複数回作用させることにより 2 時点以上ずらすことであり、その場合には、(6)式のように累乗記号を使って簡潔に記される。

$$B^2 y_t = B(By_t) = By_{t-1} = y_{t-2}$$

$$B^k = y_{t-k}, \quad k = 1, 2, \dots \dots (6)$$

ラグ演算子は、時系列データをシフトさせる作用をもつ演算子であるが、数や変数のように演算子間での交換、結合および分配法則が成り立つ。たとえば、以下の関係が成り立つ[14]。

$$\begin{aligned} (aB + b)(cB + d)y_t &= \{acB^2 + (ad + bc)B + bd\}y_t \\ &= acy_{t-2} + (ad + bc)y_{t-1} + bdy_t \\ &= (cB + d)(aB + b)y_t \quad \dots (7) \end{aligned}$$

この性質から、ラグ演算子  $B$  を変数のように利用し、自己回帰項、移動平均項の係数とラグ演算子の多項式を以下の(8)および(9)式のよ

うに定義することができる[14]。

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad \dots (8)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad \dots (9)$$

以上のように定義することにより、ARMA( $p, q$ )モデルは以下の(10)式のように簡潔に記述できる。

$$\phi(B)y_t = m + \theta(B)a_t, \quad a_t: N(0, \sigma^2) \quad \dots (10)$$

ARMAモデルは、定常条件(すなわち、平均、分散、自己共分散および自己相関係数が一定)を満たす時系列データ(定常過程)とみなせる場合にのみ用いることができるが、定常過程とみなせないものを差分処理することにより、定常過程に変換できる場合がある。この差分処理とは、隣り合う時系列データ間の差からなる時系列データに変換することであり、以下の(11)式であらわされる[14]。

$$\Delta y_t = y_t - y_{t-1}$$

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$$

$$\Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1} \quad \dots (11)$$

$d$ 回差分処理した時系列データ $\Delta^d y_t$ をさらにARMA( $p, q$ )モデルで書き表すモデルをARIMA( $p, d, q$ )モデルという。ARIMA( $p, d, q$ )モデルは以下の(12)式のように書き表すことができる[14]。

$$\Delta^d y_t = m + \phi_1 \Delta^d y_{t-1} + \dots + \phi_p \Delta^d y_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad \dots (12)$$

ARIMA( $p, d, q$ )モデルは、ラグ演算子を併用することにより、以下の(13)式のようにさらに簡潔に記述される[14](表記を簡潔にするため、 $m=0$ の場合について、 $m$ を省略して記述されることがある)。

$$\phi(B)\Delta^d y_t = m + \theta(B)a_t \quad \dots (13)$$

ARIMAモデルにおいて、特に1年分のラグ、月次データであるならば、ラグが12の倍数となる構造を想定したものがSARIMAモデル

である。ARIMA モデルの自己回帰項、移動平均項とラグ演算子の多項式に加えて、下の(14)および(15)式に示す季節自己回帰項、季節移動平均項と 12 の倍数のラグ演算子の多項式を定義する。

$$\Phi_S(B) = 1 - \Phi_1 B^{12} - \Phi_2 B^{24} - \dots - \Phi_P B^{12P} \quad \dots (14)$$

$$\Theta_S(B) = 1 + \Theta_1 B^{12} + \Theta_2 B^{24} + \dots + \Theta_Q B^{12Q} \quad \dots (15)$$

同様に、12 個離れたデータ間の差分処理を(16)式のように定義する。

$$\Delta_S^D y_t = \Delta_S^{D-1} y_t - \Delta_S^{D-1} y_{t-12} \quad \dots (16)$$

以上のラグ演算子と階差演算子を組み合わせることにより、これらの演算子なしでは一般式として記載することが困難な SARIMA モデルを簡潔に記述することができる。たとえば、ARIMA( $p, d, q$ )モデルに季節自己回帰項数  $P$ 、季節移動平均項数  $Q$ 、季節差分処理回数  $D$  を追加した SARIMA( $p, d, q$ )( $P, D, Q$ )モデルは以下の(17)式のとおり簡潔に記述することができる(定数を省略している)。

$$\phi(B)\Phi_S(B)\Delta^d \Delta_S^D y_t = \theta(B)\Theta_S(B)a_t \quad a_t \sim N(0, \sigma^2) \quad \dots (17)$$

ARIMA モデルまたは SARIMA モデルは医学・生物学や公衆衛生分野を含む自然科学の広い分野における時系列分析で使用されている [73]。北海道東藻琴食肉衛生検査所(以下検査所とする)では、10 種類の臓器または疾病による廃棄率について希望する生産者に還元しており、その際、実際の廃棄率とともに送付月の 12 か月前から送付月の 3 か月後までの毎月の廃棄率期待値の信頼限界を SARIMA モデルによって計算してグラフ表示したものを送付している。本研究では、農場における罹患動物の有意な増加やデータ還元後の薬物投与による農場有病率の有意な低減の有無を判定するための SARIMA モデルの条件を検討するとともに、モデルの当てはまりのよさを指数平滑法(対照モデル)と比較検討した。

## 2. 材 料 と 方 法

### 研究対象農場とデータ収集

2003年4月から2014年3月までの11年間に検査所管轄のと畜場に搬入された道東地区に存在する10戸の養豚場(A~J農場)産の6ヶ月齢肉豚1,193,837頭のと畜検査結果のうち、肺、大腸、小腸および肝臓の毎月の臓器廃棄率と Swine enzootic pneumonia(SEP)様肺炎、肺胸膜炎、心膜炎、肝包膜炎、寄生虫性肝炎および肝変性による毎月の臓器廃棄率の全10項目を以下に説明する時系列分析システムのデータ処理項目とし、4疾病(SEP様肺炎、肺胸膜炎、心膜炎及び寄生虫性肝炎)の毎月の廃棄率を以下に説明する廃棄率モデルの妥当性および対照モデルとの比較のために用い、以下に説明する特定農場の寄生虫性肝炎による肝臓廃棄率を信頼水準の選択のために用いた。

### 駆虫剤投与農場の廃棄率データによる信頼水準の選択

最初に、と畜検査結果から農場における衛生状態の変動を示唆する臓器廃棄率の著明な増減を検出するための時系列モデルの信頼水準を選択する。データ収集を行った期間に、研究対象農場において疾病のアウトブレイクによる著明な廃棄率の増加と、衛生対策による終息が見られた事例のデータを用い、当該事例の廃棄率増減の検出できるように信頼水準を選択することとし、そのような事例を有するI農場を対象として選択した。

検査所管轄と畜場に2003年4月~2014年3月に搬入されたI農場産の6ヶ月齢肉豚409,067頭の寄生虫性肝炎による廃棄率データを用いた。当該農場産肉豚の寄生虫性肝炎による肝臓の廃棄率は2011年8月

に 30%を超えたが、2011 年 11 月からのフェンベンダゾール製剤(メイポール 10: Meiji Seika ファルマ、東京、以下メイポール 10 とする)の飼料添加後、2013 年 6 月に 1.8%まで減少している。この廃棄率下落局面を薬物投与の有効性として検出できるように、以下に述べる手順で信頼水準を選択した。

2003 年 4 月を時点  $t=1$  として毎月の廃棄率  $p_t$ =廃棄数/と畜頭数を  $t=1\sim 132$  について計算し、各々の  $p_t$  を以下の(18)式

$$\text{logit}(p_t) = \ln\left(\frac{p_t}{1-p_t}\right) \quad \dots(18)$$

によって対数ロジット変換した。変換後のデータを用いて、(17)式における  $P+Q+p+q$  の最大次数が 5 となる全ての組み合わせについて SARIMA モデルの最尤推定を行った。SARIMA モデルの階差の次数  $D$  および  $d$  を決定するのに必要な単位根の有無の判定を Augmented Dickey-Fuller(ADF)検定[17, 18, 53]によって行った。モデルの選択には、最小の赤池情報量規準(AIC)値を基準とした。なお、最尤推定を行う際に、まず 2003 年 4 月から 2004 年 3 月までのデータを用いて 2004 年 4 月の期待値計算のためのモデリングを行い、2004 年 4 月のデータを追加してから 2004 年 5 月の期待値計算のためのモデリングを行う一歩先予測(one step prediction)を 120 回連続することによって 2004 年 4 月から 2014 年 3 月のモデリングを行った。推定されたモデルから計算された標準偏差と期待値から信頼水準を 80%、85%および 90%とした場合について、それぞれ上方および下方信頼限界を計算した。計算された信頼限界を、以下の(19)式

$$f(x) = \frac{e^x}{1+e^x} \quad \dots(19)$$

によってロジスティック逆変換し、グラフに表示した(図 1a,b)。用いた統計処理ソフトウェアは、R(version 3.0.2)[29]であり、<http://cran.r-project.org/>から Windows32/64bit 版をダウンロードし、コンピュータにインストールして用いた。また、SARIMA モデルの最尤推定には auto.arima 関数[27]を用いた。指標として用いる信頼水準を小さな値にすれば、感度は高くなるが、特異度が低下してばらつき等に起因するノイズによって観測値が信頼限界を超え、誤検出することとなり、信頼水準を大きくしすぎると、アウトブレイクを全く検出できなくなってしまう。そこで、I 農場で実際に発生した寄生虫性肝炎のアウトブレイクを検出できる感度を有する信頼水準のうち、最も特異度が高い信頼水準を選択することにした。具体的には、I 農場の廃棄率が下落した 2011 年 8 月～2013 年 6 月において、観測された廃棄率が最大頻度で下方信頼限界を下回る信頼水準のうち、最大のものを選択した。

I 農場における廃棄率下落局面と同時期に寄生虫性肝炎による肝臓廃棄率の顕著な増減が見られなかった D 農場についても I 農場のデータを用いて行った処理を行い、グラフに表示した(図 2)。但し信頼水準については I 農場の検討で選択されたもの(85%)を用いた。

## 各調査農場における廃棄率モデルの妥当性および対照モデルとの比較

2003 年 4 月から 2014 年 3 月までの 11 年間に検査所管轄のと畜場に搬入された 10 農場(A～J 農場)産の 6 ヶ月齢肉豚 1,193,837 頭の 4 疾病(SEP 様肺炎、肺胸膜炎、心膜炎及び寄生虫性肝炎)の月次集計データを用いた。

毎月の廃棄率( $p_t$ =廃棄数/ $l$ と畜頭数)を計算し、各々の  $p_t$  を(18)式によって対数ロジット変換した後、11 年分のデータ全てを用いて信頼水準

(85%)の選択時と同じ条件で SARIMA モデルの最尤推定と(1)式に示した指数平滑法の最尤推定を行った。推定した SARIMA モデルのパラメータを、 $SARIMA(p,d,q)(P,D,Q)_s+C$  の形式で記載することとした(表1)。ここで、 $p$ :自己回帰項の次数  $d$ :階差の次数  $q$ :移動平均項の次数  $P$ :季節自己回帰項の次数  $D$ :季節階差の次数  $Q$ :季節移動平均項の次数  $s$ :季節変動の期間  $C$ :定数項を意味している。

推定された  $p,q,P,Q,C$  の各項をそれぞれの標準偏差で除した値を  $t$  値とし、パラメータの有意性検定として有意水準 0.05 の  $t$  検定(両側検定)を行った。

さらに、残差の無作為性検定(Ljung-Box 検定[40])として(20)式で計算される  $Q$  値が自由度  $h$  の  $\chi^2$  分布の 5%分位点を上回る場合に、帰無仮説を棄却した。具体的には統計ソフト R の `tsdiag` 関数のデフォルト条件( $h \leq 10$ )で表示される  $p$  値のチャートが 0.05 を下回っている場合に残差の無作為性なしと判定し、その時の  $h-1$  を残差の無作為性が認められた最大の時間差(月)とした。残差とは、観測データとモデルの差であり、残差が無作為とは残差に時間依存性がないことを意味し、残差が無作為ではないとは、観測データの時間依存性を十分にモデル化できていないことを意味する。

$$Q = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}_j^2}{n-j} \quad \dots (20)$$

ここで  $n$ : 標本数

$h$ : 時間差

$\hat{\rho}_j$ : 時間差  $j$  における自己相関係数

以下に示す(21)式によって、各モデルの尤度  $L$  を計算し、(22)式によって逸脱度  $Dev$  を計算した。

$$L = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\varepsilon_t^2}{2\sigma^2}} \quad \dots (21)$$

(21)式において、 $\varepsilon_t = \text{logit } p_t - \text{logit } \hat{p}_t$

ここで  $p_t$ :  $t$  時点における実際の廃棄率

$\hat{p}_t$ :  $t$  時点におけるモデルによる廃棄率期待値

$\sigma$ : モデルの標準偏差

$n$ : 時点数

である。

$$Dev = -2\log L \quad \dots (22)$$

尤度比検定として、SARIMA モデルと指数平滑法によるモデルの逸脱度  $Dev$  について、モデルと対照モデルのパラメータ数の差を自由度とし、 $p < 0.05$  を有意水準とするカイ二乗検定を行った。

(23)式により、絶対平均比率誤差(MAPE)を計算し、農場及び疾病ごとに指数平滑法と SARIMA モデルで数値の大小比較を行い、さらに  $t$  検定(両側検定)を行った。

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{\text{logit } p_t - \text{logit } \hat{p}_t}{\text{logit } p_t} \right| \quad \dots (23)$$

#### 分析システムの構築と還元:

日々のと畜検査の結果は、と畜場でタブレット端末に入力後、オンラインで北海道庁本庁のサーバーに送信され、蓄積されている。著者は、と畜検査結果をサーバーからコンピュータ端末に取り込んで時系列分析する処理を、多数の生産者および疾病について自動的に行うシステムを構築した。なお、生産者への還元は、便宜上グラフ印刷した紙媒体の配布によって実施されている。当該システムの処理手順を以下に説明する。

毎月分の廃棄率( $p_i$ =廃棄数/と畜頭数)を計算し、各々の  $p_i$  を(18)式によって対数ロジット変換した後、信頼水準の選択時と同じ条件で SARIMA モデルの最尤推定を行った。但し、一步先予測により直近 12 か月分の期待値計算用のモデリングを行い、さらに最終月のデータを加えて推定したモデルを今後 3 か月間の期待値計算用モデルとした。

生産者にとって最も有用と考えられる情報の一つは、最終月のデータを使わずに推定された時系列モデルによって計算された信頼区間の間に最終月の観測値があるか否かがわかることである。たとえば、現在が 2018 年 2 月末日であるならば、2018 年 1 月末日までのデータを用いてモデリングし、2 月末日の信頼限界を計算し、2018 年 2 月の 1 か月間の集計データと比較すれば、最終月の観測値が信頼区間の間にあるかどうかができる。しかし、この 1 点の比較のみでは、これまでの推移が生産者にはわからない。そこで、過去 1 年間の推移についても生産者が把握できるように、過去の信頼区間と観測値を表示することにした。具体的には 2017 年 2 月末日までのデータを用いて計算された 2017 年 3 月末日の信頼区間から、2018 年 1 月までのデータを用いて計算された 2018 年 2 月末日の信頼区間までの 12 個の信頼区間を計算し、グラフに表示することにした。また、予測される今後の廃棄率の推移を示すことにより、生産者の早期の対応を促すことができる可能性があるため、最終月の観測値を含めてモデリングを行い、翌月から 3 か月後までの予測値、上の例で具体的に年月を示すと、現在が 2018 年 2 月末日であるならば、2018 年 2 月の観測値も用いて作成したモデルにより 3 月～5 月の予測区間を表示することにした。推定されたモデルから計算された標準偏差と期待値から上方および下方信頼限界を計算し、(19)式によって%値に変換した後、折れ線グラフに表

示した。

### 3. 結 果

#### 駆虫剤投与農場の廃棄率データによる信頼水準の選択

I 農場における寄生虫性肝炎による肝臓廃棄率の経時変化(実線)と期待値の上方および下方信頼限界(80%,85%,90%)ならびに医薬品使用報告書に基づくメイポール 10 の投与豚搬入期間を図 1a,b に示した。

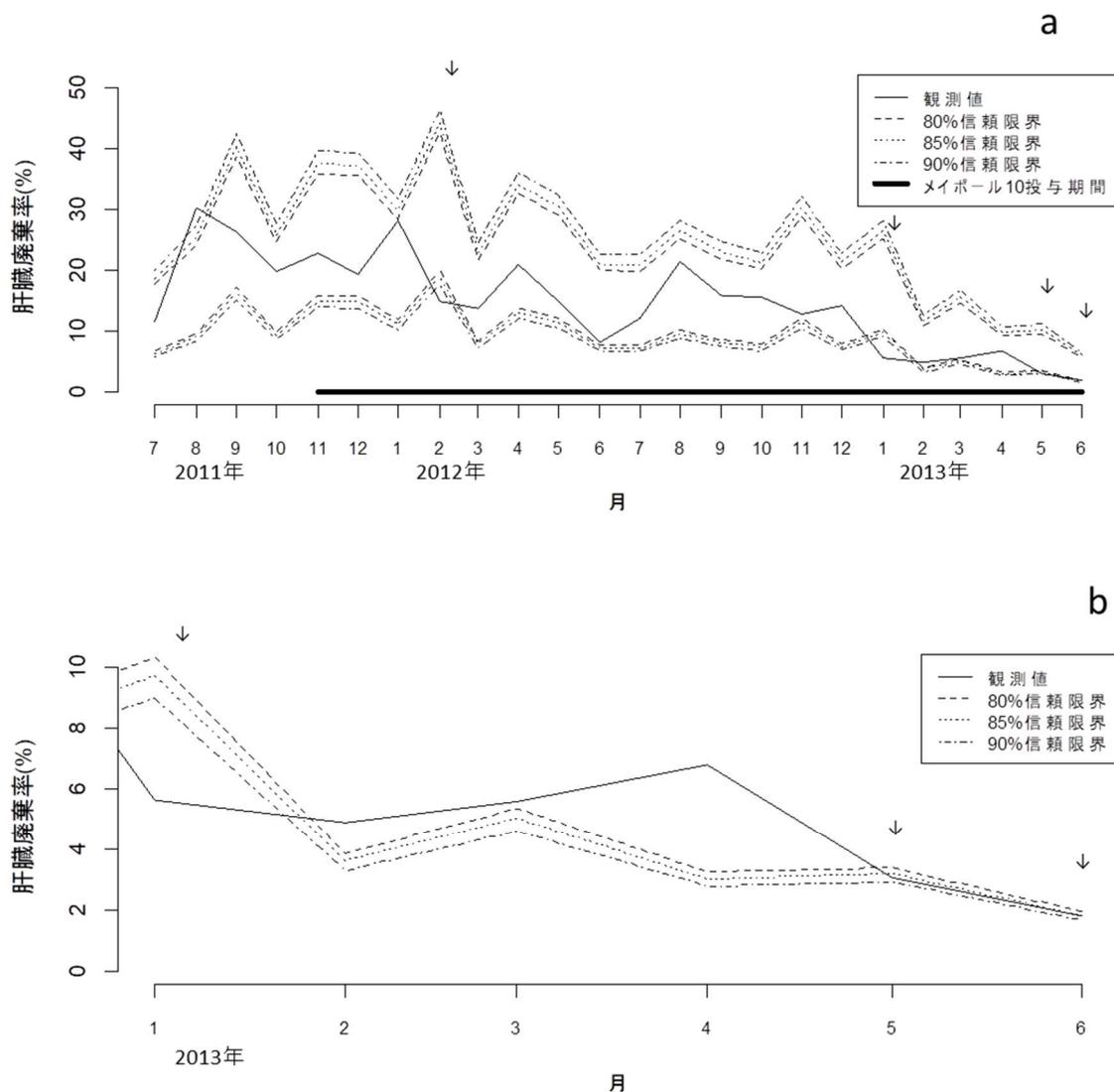


図 1. I 農場におけるメイポール 10 投与期間、寄生虫性肝炎による廃棄率と信頼限界の推移

図中の実線は、寄生虫性肝炎による実際の廃棄率を示す。間隔の異なる破線は、SARIMA モデルによって計算された 80、85 および 90% 信頼限界を示す。廃棄率がいずれかの信頼限界を下回った時を↓で示す。

a: 2011 年 7 月～2013 年 6 月の上方および下方信頼限界

b: 2013 年 1 月～6 月の下方信頼限界

肝臓廃棄率は、2011年8月に上方信頼限界(80~90%)を超えて増加した。2011年11月から2013年6月までの間に廃棄率は、30.2%から1.8%に下落し80%および85%信頼限界を2012年2月、2013年1、5、6月の4回下回ったが、90%信頼限界の場合は、2012年2月および2013年1月の2回下回ったのみであった。以上の結果より、最大頻度で下方信頼限界を下回る信頼水準のうち、最大の信頼水準として85%を選択した。選択した信頼水準85%を用いてI農場とは対照的に駆虫対策を実施していないD農場における寄生虫性肝炎による肝臓廃棄率データをI農場と同様に処理した結果を図2に示す。廃棄率(実線)は上方および下方信頼限界(点線)の間にあり、両信頼限界は概ね横ばいであった。

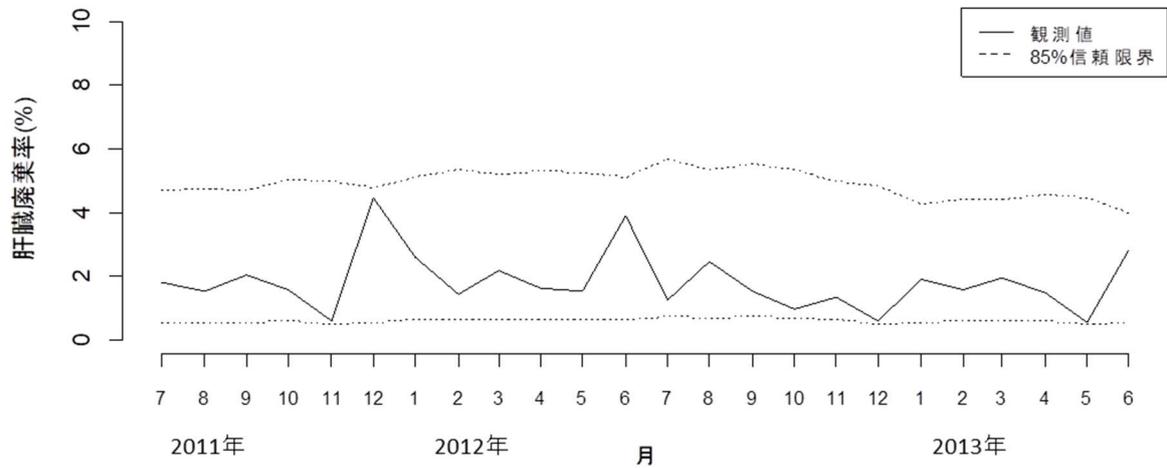


図 2. 図 1a と同時期の D 農場における寄生虫性肝炎による廃棄率と信頼限界の推移

図中の実線は、寄生虫性肝炎による実際の廃棄率を示す。破線は SARIMA モデルによって計算された 85%信頼限界の値を示す。

各調査農場における廃棄率モデルの妥当性および対照モデルとの比較:

推定された各モデルにおけるパラメータ数と有意差が認められたパラメータ数を表 1 に示した。全パラメータ数 162 個のうち、有意なパラメータは 104 個(64.2%)であり、疾病別ではそれぞれ、SEP 様肺炎では 26/40(65.0%)、肺胸膜炎では 23/39(59.0%)、心膜炎では 30/43(69.8%)、寄生虫性肝炎では 25/40(62.5%)であった。

表 1. 推定された各モデルにおけるパラメータ数と有意差が認められたパラメータ数

農場	SEP様肺炎			肺胸膜炎			心膜炎			寄生虫性肝炎		
	モデル <sup>a)</sup>	有意パラ メータ数 <sup>b)</sup>	全パラ メータ数	モデル	有意パラ メータ数	全パラ メータ数	モデル	有意パラ メータ数	全パラ メータ数	モデル	有意パラ メータ数	全パラ メータ数
A	SARIMA(2,1,1) (2,0,0) <sub>12</sub>	1	5	SARIMA(1,0,2) (2,0,0) <sub>12</sub> +C	3	6	SARIMA(0,1,3) (1,0,1) <sub>12</sub>	2	5	SARIMA(1,0,1) (1,0,2) <sub>12</sub> +C	3	6
B	SARIMA(1,1,1)	1	2	SARIMA(2,1,1) (2,0,0) <sub>12</sub> +C	2	6	SARIMA(0,1,1)	1	1	SARIMA(0,0,1) (2,0,0) <sub>12</sub>	1	3
C	SARIMA(4,1,0)	2	4	SARIMA(0,1,1) (2,0,1) <sub>12</sub>	3	4	SARIMA(1,0,1) (1,0,1) <sub>12</sub> +C	3	5	SARIMA(1,0,1)+C	3	3
D	SARIMA(3,1,0) (2,0,0) <sub>12</sub>	3	5	SARIMA(0,1,1) (2,0,2) <sub>12</sub>	1	5	SARIMA(1,1,1) (2,0,1) <sub>12</sub>	3	5	SARIMA(0,1,2) (1,0,0) <sub>12</sub>	1	3
E	SARIMA(0,1,1)	1	1	SARIMA(1,1,1)	2	2	SARIMA(3,0,0) (2,0,0) <sub>12</sub> +C	3	6	SARIMA(1,0,0)+C	2	2
F	SARIMA(1,0,1) (2,0,0) <sub>12</sub> +C	4	5	SARIMA(0,1,1) (1,0,1) <sub>12</sub>	1	3	SARIMA(4,1,1)	5	5	SARIMA(2,0,1)+C	3	4
G	SARIMA(1,0,1) (2,0,0) <sub>12</sub> +C	4	5	SARIMA(0,0,1) (0,1,1) <sub>12</sub>	2	2	SARIMA(1,1,1)	2	2	SARIMA(4,0,0) (1,1,0) <sub>12</sub>	3	5
H	SARIMA(1,1,1)	2	2	SARIMA(0,0,2) (0,0,1) <sub>12</sub>	3	3	SARIMA(0,0,2) (0,0,2) <sub>12</sub>	3	4	SARIMA(0,1,4)	3	4
I	SARIMA(1,0,0) (2,0,2) <sub>12</sub> +C	4	6	SARIMA(2,0,0) (2,0,1) <sub>12</sub> +C	4	6	SARIMA(2,1,3)	6	6	SARIMA(3,1,0) (1,1,1) <sub>12</sub>	3	5
J	SARIMA(3,1,0) (2,0,0) <sub>12</sub>	4	5	SARIMA(2,1,0)	2	2	SARIMA(4,1,0)	2	4	SARIMA(1,0,1) (2,0,0) <sub>12</sub> +C	3	5
	合計	26	40		23	39		30	43		25	40

a) SARIMA( $p, d, q$ ) ( $P, D, Q$ )<sub>s</sub>+Cの形式で記載。 $p$ :自己回帰項の次数  $d$ :階差の次数  $q$ :移動平均項の次数  $P$ :季節自己回帰項の次数  $D$ :季節階差の次数  $Q$ :季節移動平均項の次数  $s$ :季節変動の期間  $C$ :定数項

b)  $t$ 検定、 $p < 0.05$ 、両側検定

残差の無作為性が認められた最大時間差を表 2 に示す。Ljung-Box 検定における時間差  $h$  の設定には明確な基準はないが、10 以上を選択する事が多く、本研究で使用した `tsdiag` 関数の通常条件では時間差 10 までの  $p$  値が表示される。本研究では、残差の無作為性が認められる最大時間差数が 10 未満であったのは、G 農場 SEP 様肺炎のモデルのみであった。

表 2. 推定された各モデルにおける残差の無作為性が認められた最大時間差

農場	SEP 様肺炎	肺胸膜炎	心膜炎	寄生虫性肝炎
A	10 ≦	10 ≦	10 ≦	10 ≦
B	10 ≦	10 ≦	10 ≦	10 ≦
C	10 ≦	10 ≦	10 ≦	10 ≦
D	10 ≦	10 ≦	10 ≦	10 ≦
E	10 ≦	10 ≦	10 ≦	10 ≦
F	10 ≦	10 ≦	10 ≦	10 ≦
G	4	10 ≦	10 ≦	10 ≦
H	10 ≦	10 ≦	10 ≦	10 ≦
I	10 ≦	10 ≦	10 ≦	10 ≦
J	10 ≦	10 ≦	10 ≦	10 ≦

逸脱度および尤度比検定の結果を表 3 に示す。逸脱度の差は、同じ農場の同じ疾病について指数平滑法の逸脱度から SARIMA モデルの逸脱度を差し引いたものである。SARIMA モデルの逸脱度が大きいのは D 農場における SEP 様肺炎の場合のみで、他の 39 農場・疾患の比較では（10 農場 × 4 疾病 = 40 農場・疾患、 $40 - 1 = 39$  農場・疾患）、SARIMA モデルの逸脱度が小さく、そのうち 27 の農場・疾患において有意差が認められた。

表 3. 推定された各モデルにおける逸脱度<sup>a)</sup>および尤度比検定<sup>b)</sup>の結果

農場	SEP 様肺炎		肺胸膜炎		心膜炎		寄生虫性肝炎	
	SARIMA モデル	指数平滑法	SARIMA モデル	指数平滑法	SARIMA モデル	指数平滑法	SARIMA モデル	指数平滑法
A	<u>261.6*</u>	272.4	<u>186.7*</u>	196.3	<u>199.1</u>	204.9	<u>267.8</u>	275.9
B	<u>224.7</u>	225.7	<u>74.8</u>	83.9	<u>154.8</u>	155.9	<u>232.8</u>	235.5
C	<u>112.8*</u>	120.7	<u>43.3***</u>	81.4	<u>121.9*</u>	132.5	<u>271.5</u>	274.3
D	226.1	<u>225.1</u>	<u>36.4*</u>	44.8	<u>160.1***</u>	203.6	<u>272.5</u>	275.7
E	<u>149.3</u>	152.6	<u>83.8*</u>	98.2	<u>131.0</u>	138.3	<u>288.7***</u>	310.0
F	<u>140.6***</u>	163.7	<u>26.2</u>	29.8	<u>81.2***</u>	100.1	<u>269.2***</u>	285.0
G	<u>120.0***</u>	145.9	<u>-46.7***</u>	-35.4	<u>-4.0**</u>	5.2	<u>171.4</u>	178.8
H	<u>57.3*</u>	63.7	<u>-51.2***</u>	-39.3	<u>-72.9**</u>	-60.6	<u>184.3***</u>	202.1
I	<u>30.5***</u>	54.8	<u>-173.3***</u>	-131.7	<u>-64.0**</u>	-50.8	<u>155.8***</u>	200.5
J	<u>253.1**</u>	263.6	<u>160.0**</u>	170.7	<u>194.0*</u>	200.2	<u>340.3**</u>	349.8

a) (21)式と(22)式によって求められた逸脱度を表に示した。

b) 両モデルの逸脱度差のカイ二乗検定での有意差。 \*:  $p < 0.05$  \*\*:  $p < 0.01$  \*\*\*:  $p < 0.001$

*MAPE*を表4に示す。38農場・疾患でSARIMAモデルの*MAPE*が小さく、40農場・疾患中11農場・疾患で有意差( $p < 0.05$ )が認められた。尤度比及び*MAPE*のいずれの検定においても、指数平滑法モデルの方が有意に小さいと判断されたものはなかった。

表 4. 推定された各モデルにおける絶対平均比率誤差(MAPE)<sup>a)</sup>

農場	SEP 様肺炎		肺胸膜炎		心膜炎		寄生虫性肝炎	
	SARIMA モデル	指数平滑法	SARIMA モデル	指数平滑法	SARIMA モデル	指数平滑法	SARIMA モデル	指数平滑法
A	<u>24.52</u>	25.64	<u>20.37</u>	20.75	<u>22.26</u>	22.31	<u>13.67</u>	14.05
B	<u>17.39</u>	17.44	<u>12.10<sup>*b)</sup></u>	12.75	<u>10.04</u>	10.11	<u>12.37</u>	12.49
C	<u>11.78</u>	12.00	<u>9.89<sup>***</sup></u>	12.47	<u>9.45</u>	9.59	<u>12.07</u>	12.27
D	<u>14.36</u>	14.49	<u>13.01<sup>*</sup></u>	13.59	<u>9.41<sup>***</sup></u>	12.73	<u>12.30</u>	12.56
E	<u>17.42</u>	17.52	<u>15.49</u>	16.07	10.96	<u>10.75</u>	<u>13.05<sup>*</sup></u>	14.83
F	<u>12.99<sup>*</sup></u>	14.16	<u>11.94</u>	12.17	<u>8.13</u>	8.61	<u>11.77<sup>**</sup></u>	12.72
G	<u>13.35</u>	14.50	<u>11.57</u>	11.86	<u>7.04</u>	6.96	<u>8.46</u>	8.69
H	<u>12.78<sup>**</sup></u>	13.67	<u>10.71<sup>*</sup></u>	11.43	<u>5.77</u>	5.83	<u>10.92</u>	11.49
I	<u>9.58</u>	10.44	<u>5.93<sup>***</sup></u>	7.76	<u>5.71</u>	5.95	<u>14.02<sup>**</sup></u>	19.02
J	<u>21.85</u>	22.18	<u>19.39</u>	19.96	<u>13.2</u>	13.55	<u>19.07</u>	19.35

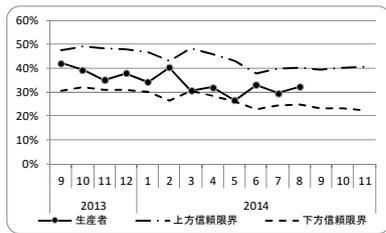
a) MAPE の値を指数平滑法と SARIMA モデルで比較し、小さい数値に下線を付した。

b) 絶対比率誤差 ((25) 式の絶対値記号部分の値) の  $t$  検定での有意差 \*: $p<0.05$  \*\*: $p<0.01$  \*\*\*: $p<0.001$

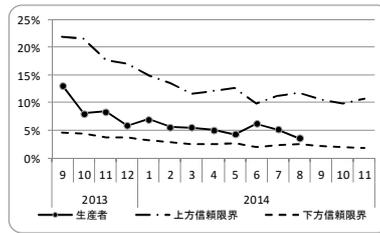
### 分析システムの構築と還元:

システムからの出力例として 2014 年 8 月末日における D 農場の解析結果を示した(図 3)。図 3 に例示した処理結果には、10 の臓器または疾病による廃棄率についてデータ送付月の 12 カ月前から 3 か月先までの毎月の廃棄率期待値の信頼限界を表示しており、検査所では 2014 年 4 月より希望する生産者にはこの様式の処理結果を郵送している。

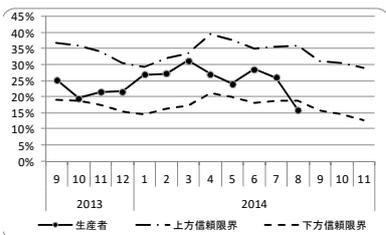
肺の廃棄率



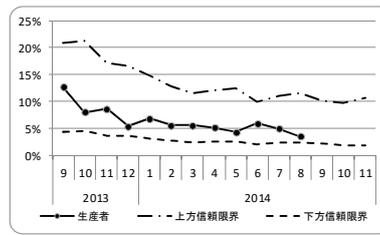
小腸の廃棄率



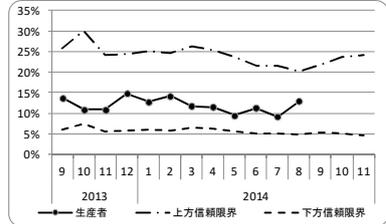
肝臓の廃棄率



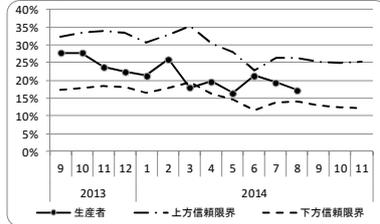
大腸の廃棄率



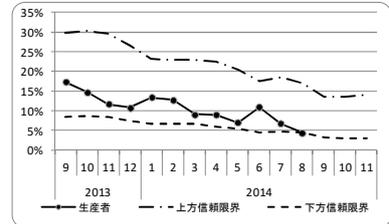
SEP肺炎



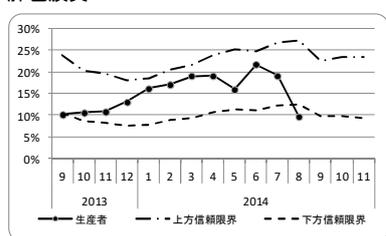
肺胸膜炎



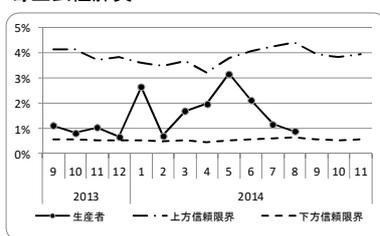
心膜炎



肝包膜炎



寄生性肝炎



肝変性

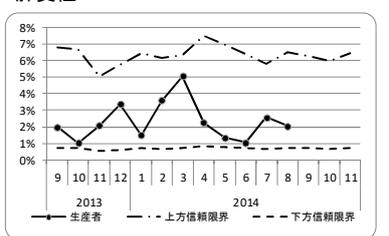


図 3. 分析システムの出カ例(D 農場)

図中の実線は、実際の廃棄率を示す。一点鎖線は SARIMA モデルによって計算された上方信頼限界(85%)の値を示し、破線は、SARIMA モデルによって計算された下方信頼限界(85%)の値を示す。

#### 4. 考 察

本研究では、と畜検査データを SARIMA モデルによって分析するための条件検討を行い、推定したモデルの統計学的評価を行った。公衆衛生分野の研究では、感染症患者数の解析に ARIMA モデルまたは SARIMA モデルが用いられている。そのような研究の例として、カンピロバクター食中毒[5]、重症急性呼吸器症候群[20]、腎症候性出血熱[39]、コレラ[4]、オンコセルカ症[37]、インフルエンザ[58]、マラリア[68]やデング熱[10]の研究があげられる。また救急外来患者数の解析に用いられることもある[51, 60]。これらの中には、バイオテロリズムの発生[51]や、未知の食中毒感染源[5]の検知を目的とするものもある。Reis ら[51]は、救急外来者数に基づく症候群サーベイランスシステムで、バイオテロリストによる攻撃発生の早期シグナルの可能性を有する異常な救急外来者数の増加を発見でき、例えば、炭疽病のアウトブレイクは、最初に救急外来部門への呼吸器症状を伴う患者の報告数の異常な増加として検知されるであろうとしている。このような背景で、Reis ら[51]は、症候群サーベイランスシステム構築のために ARIMA モデルを用いて時系列分析を行っている。時系列分析が、時間とともに出現するデータの分析方法であることからと畜検査データの解析に応用できることは明らかである。しかし、先行研究では、経時変化をトレンドや周期性などの要素に分解する [50, 55, 56, 64]に留まっている。これらの先行研究の中で Neumann ら[50]は、ARIMA モデルを利用し、期待値と廃棄率を比較することによって合理的評価が可能であると言及している点では注目に値するが、評価そのものについては未検討であった。

と畜検査における廃棄率の経時的羅列(プロット)から把握される数

値変動の中から衛生上の問題として捉えるべき変動とそうでない変動を識別する根拠は、分析者の経験に依存するのが普通であった。しかし、このような主観的な判断は分析者ごとに異なる危険性があり、客観的な手法が望まれていた。本分析法では、観測された廃棄率と SARIMA モデルを利用して求められた期待値の信頼限界との大小比較に基づいて上記の識別を行うため、この客観的な手法であることが期待される。

本分析法では、信頼水準を適宜選択することにより、感度や特異度を調整することが可能となった。また、一般的に信頼水準として 95% が選択される例は多いが、気象庁の台風の予報円[28]のように 70% が選択される場合もある。感度と特異度はトレードオフの関係にある。すなわち、信頼水準を大きくすると両信頼限界間が非常に大きくなり、感度が低下する。一方、信頼水準を小さくすると、特異度が低くなり、些細な変動をも有意な現象として誤検出される危険性が高くなる。本研究では、後述するように駆虫剤投与による肝臓廃棄率の下落過程における検出頻度をもとに 85% を選択した。しかし、場合によっては、この値に限定せず、その他の事例も参照して決定することが望ましいであろう。たとえば、農場における真の有病率が変動している蓋然性が高い時の廃棄率変動を検出する最大の信頼水準の選択などである。すなわち、実際の応用では農場における真の有病率を把握後、感度と特異度のバランスを考慮して信頼水準を決定することが望ましい。

今回、I 農場の廃棄率データ分析の信頼水準を 85% とした理由は、駆虫剤投与時の廃棄率が最大頻度で下方信頼限界を下回る信頼水準のうち、最大のものであり、感度と特異度のバランスがこの数値で最適

だったためである。そして、2011年8月に見られるようにこの信頼区間の上方信頼限界を超え廃棄率増加が認められた場合を廃棄率の著明な増加と判定し、2012年2月、2013年1、5、6月に見られるように廃棄率が同区間の下方信頼限界を下回る場合、廃棄率の著明な減少と判定した。以上のように、信頼限界を超えるまたは下回ることを判断指標にすれば、著明な廃棄率の増加や減少があったことを客観的に識別することができ、疾病蔓延や生産者による衛生対策の有効・無効の指標とすることが期待された。

通常、SARIMAモデルの推定を行う際には、モデルを構成する多項式の各係数がゼロと有意差があるかを確認し、さらに残差の無作為性を確認することによってモデルが妥当であると判断される。しかし、検査所が毎月送付している約30生産者の還元データを処理する過程で生成される約3,900個のモデルには、約20,000のパラメータが含まれており、それら全てのパラメータの有意性と全てのモデルの残差の無作為性を確認することは処理手順の煩雑化による作業時間の著しい増大を招くため、事実上不可能である。よって、検査所の毎月のデータ還元作業で用いる分析システムでは、パラメータの有意性と残差の無作為性の確認を行わず、AICが最も低いモデルを自動処理で選択することによって実用的な処理時間で稼働するようにした。このようなルーチン作業による処理結果の信頼性を担保するために、通常SARIMAモデルの推定で行われることが多い処理を省略し、AICのみを指標としてモデルを選択しても、妥当な時系列モデリングがなされていることを確認することが必要となる。そこで、本研究ではデータ還元対象生産者から一部抽出されたデータを用い推定された40のモデルを構成する多項式におけるパラメータの有意性と残差の無作為性

を検討した。モデルには最低 1 つ以上の有意パラメータが存在すれば SARIMA モデルとして機能はする。AIC を規準としてパラメータの個数を選択していることから有意差のないパラメータがモデルの当てはまりを悪化させることはないが、有意パラメータの数は可能な限り多いことが望ましい。今回検討した 40 のモデルは全て有意パラメータを含み、さらに各モデルで全体の 6 割以上のパラメータが有意であり、疾病ごとの偏りは認められなかった。また、残差の無作為性が認められなかったのは 40 モデル中 1 モデル(2.5%)のみであった。以上の結果から、AIC のみを基準としてモデルを選択しても、有意パラメータを含み、残差が無作為となるモデルが選ばれていることが確認できた。

本研究では、尤度比検定を実施し、さらに *MAPE* を計算することによって、SARIMA モデルと指数平滑法によるモデルの当てはまりの良さと精度の比較を行った。尤度比検定は、誤差項を含めたモデルの当てはまりの良さを比較するのに用いられるのに対し、*MAPE* は予測の精度を示す。比較の結果、尤度比検定(39/40 で良好、27 農場・疾患で有意)と *MAPE*(38/40 で良好、11 農場・疾患で有意)の両方で SARIMA モデルが良好であり、当てはまりの良さ及び精度の両面で SARIMA モデルが指数平滑法によるモデルより優れていることがわかった。

以上から、本研究で用いた分析方法は、自動処理を行うために処理手順が簡略化されているにもかかわらず、有意パラメータを含み、残差は概ね無作為であり妥当な時系列モデルが選択されていた。また、選択されたモデルは指数平滑法などの簡便に推定できるモデルよりも精度が高く当てはまりが良好であった。本方法で選択された SARIMA モデルは、妥当性のある時系列モデルであり、高い精度を有している

ことから、と畜検査データの分析に利用することは妥当であると考え  
る。

これまで検査所では、廃棄率を経時的に羅列したグラフを生産者宛  
に郵送していたが、廃棄率の増減のみから対策を講じるべき廃棄率の  
変動を明確に識別するのは困難であった。しかし、信頼限界を超える  
廃棄率の増減を生産者に示すことにより生産者が容易に理解できるこ  
とになったと考えられる。また、最終データ月の翌月から3か月間の  
廃棄率の予測区間の提示は、生産者に早めの対応を促すことも期待さ  
れた。

本研究はと畜検査データの統計処理方法の例として SARIMA モデル  
を示し、主に直感で判断されていた廃棄率の傾向変化を科学的に判断  
する指標として利用できることを示した。本指標をより所にして生産  
現場の衛生向上を行う動機とし、投薬や消毒などの成果を確認できる  
ことにより、高品質で衛生的な食肉生産を支援するツールとして本法  
による分析結果が生産者に利用されることを期待している。特に投薬  
や消毒などの行動はコストアップ要因であるために生産者によっては  
躊躇する可能性があるが、効果をより合理的に判断できることにより  
積極的な対応を支援できる可能性がある。

## 5. 小 括

これまで、と畜検査データに対して表やグラフを用いて経時的に羅  
列される以上の解析がなされていないことが多く、有病率の大小や増  
減の判断を客観的に行うことが難しかった。本章で述べたと畜検査デ  
ータの時系列分析方法によって得られる信頼限界を、有病率の増減の  
指標とすることにより、客観的な判断を行うことができる。

しかし、ロジット変換された有病率にガウス分布を仮定できるのは常時みられる疾病に用いる場合に限られる。稀に見られる疾病に対して時系列分析を行うには、第二章以降に述べる他の方法を用いる必要がある。

## 第 II 章 と畜場における小腸廃棄数の時系列分析を利用した豚の抗酸菌症農場アウトブレイク迅速検出法

### 1. 小 緒

*Mycobacterium avium-intracellulare* Complex (MAIC)は自然界に広く存在する病原菌であり、*M. tuberculosis* と同様に人に結核様病変を引き起こし、後天性免疫不全症候群(AIDS)患者に日和見感染を引き起こす[16, 32, 65]。MAICによる豚の抗酸菌症は慢性感染症であり、顎下リンパ節と腸間膜リンパ節に乾酪壊死を呈する。豚で最もよく見られる病変は、腸間膜リンパ節にみられる乾酪壊死である[70]。抗酸菌症は、人や動物に乾酪壊死を形成する唯一の疾病である。日本国内の食肉検査機関では公式のマニュアル[77]に従い、リンパ節に乾酪壊死が認められた豚には抗酸菌症の診断が下される。腸のみに乾酪壊死病変が認められる場合には腸のみが廃棄されるが、全身性の感染と認められる場合には全部廃棄処分となって、養豚生産者に大きな経済的負担を生じさせる。

前章に述べた通り検査所は、希望する生産者に対してと畜検査データを還元している。加えて、抗酸菌症が認められた全ての生産者に対してその旨を情報提供している。豚の抗酸菌症は年間を通じて散発的に発生しており、毎日のと畜検査データや月次集計データの俯瞰のみからアウトブレイクの有無を判断することは容易ではない。したがって、散発的発生とアウトブレイクを区別するための合理的方法が必要であった。しかし、そのような方法は未だ報告されていない。

本章では、ARIMAモデルによる時系列分析法を用いて抗酸菌症発生数予測モデルの検討を行うが、毎日の抗酸菌症の発生数データはゼ

ロが多く、第1章で用いたような廃棄率のロジット変換値を時系列分析するという方法は適さない。そこで、本章では、ゼロを多く含むカウントデータに対応する誤差構造での時系列分析を試みる。ARIMAモデルの推定を行うには、時系列データはトレンドや周期変動がなく、定常的でなければならない。BrockwellとDavis[14]は、トレンドと周期変動を除去するための方法として次の3つを提唱している。すなわち、(i)トレンドと周期成分を推定し、オリジナルデータから差し引く、(ii)階差  $d$  の差分を行う ( $d$  は、周期成分における1周期分の長さ)、(iii)三角関数と多項式の和からなるトレンドと周期成分の推定を行う。

本章では、曜日を説明変数とする擬似ポアソン回帰モデルを推定してその各曜日の期待値を経時的な廃棄数データの周期成分とし、この周期成分を差し引いた残差に対してARIMAモデルの推定を行った。次に、周期成分とARIMAモデルを加えることによって時間依存的な期待値を推定した。最後に、農場におけるアウトブレイクを検出するために、モデル推定のために用いられた期間より後に実際に観測されたデータと、モデルによって予測された期待値の95%信頼限界を比較した。

## 2. 材 料 と 方 法

### 研究の対象地域とデータ

北海道網走郡に位置する検査所は、年間150,000~200,000頭の肉豚(主として道東産)を受け入れる2つのと畜場を管轄している。検査所では、豚の解体時に肉眼検査を行い、腸に結節性の乾酪壊死を認めた場合は抗酸菌症と診断している。

本研究では、検査所が管轄する2つのと畜場で2003年4月1日から2013年5月27日までの間に得られたデータを用いた。豚の抗酸菌症によって廃棄された小腸の数(以後、観測データと呼ぶ)、当該豚の生産者名と所在する市町村名を、豚抗酸菌症の高発生農場を特定するために用いた。なお、これらの生産者には検査所より、アウトブレイクの恐れがある旨を情報提供される。

データを用いた期間のうち最初の約8年間にあたる1,922操業日のデータを学習期間としてモデル生成に用い、残りの約2年間にあたる788操業日を検証期間としてモデルの検証に用いた。

## 周期性の検討

廃棄数の周期性を理解するために、高速フーリエ変換(FFT)[63]を行い、学習期間に得られた日次廃棄数の観測データからピリオドグラムを計算した。観測データは、非操業日を含んでおらず、完全に毎日連続したデータではないため、生データに対してそのままFFTで処理することはできなかった。そこで、FFTでの処理に先立ち、以下の手順で前処理を行った。すなわち、(i)週末のデータを含まない平日のみの観測データを結合し、(ii)平日の非操業日における欠損データをダミー値である平日の各曜日のメジアンで置換し、1週間が5日からなる名目上の週となるようにした。各曜日における小腸の廃棄数の変化を調べる際には、擬似ポアソン分布を仮定した一般化線形モデル(GLM)を用いた。本章におけるすべての統計学的処理を統計ソフトウェアであるR(version 3.0.1)を用いて行った。

## ARIMA モデルの推定

ARIMA モデルの推定を行うために、擬似ポアソン回帰モデルで計算した各曜日の期待値を曜日別に観測データから差し引き、欠損値を含む残差を用いた。ARIMA モデルの推定を行うときは、3つの主要なパラメータを選択する必要がある。すなわち、自己回帰(AR)の次数、差分また和分の次数、および移動平均(MA)の次数である。自己回帰項は、時系列データの時点  $t$  における観測値と、ラグ  $k$  遡った時点  $t-k$  における観測値との相関を示す。移動平均項は、時点  $t$  における誤差(観測値と推定値の差)と、ラグ  $k$  遡った時点  $t-k$  における誤差との相関を示す。時系列データが確率的トレンドを含んでいた場合(すなわち、データが単位根を含んでいた場合)、差分処理を行うことによってトレンドが除去され、定常過程の時系列データが得られる。単位根について、以下に簡単に説明する。

時系列データは確率過程からの標本の実現値であるが、各時点からは1個のデータしか得られない。ところが、確率過程においては期待値や分散は時点ごとに異なりうるから、追加的な情報なり制約がなければ、各時点で1個のデータから各時点の平均や分散を推定することができない。そこで意味のある統計的推測を行うためには、確率過程を制約した確率モデルを想定する必要がある、平均、分散、自己共分散および自己相関係数が一定であるという定常条件を満たす定常過程であることを制約としている。MAモデルは、モデルの性質上常に定常条件を満たすが、 $p$  次のAR( $p$ )モデルは、以下に示す特性方程式の解が一定の範囲内にあることが必要である。すなわち、

$$1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0 \quad \dots (24)$$

の解がすべて1より大きい、もしくは

$$x^p - \phi_1 x^{p-1} - \phi_2 x^{p-2} - \dots - \phi_p = 0 \quad \dots (25)$$

の解がすべて 1 より小さいことである。ここで、 $\phi_1 \sim \phi_p$  は、AR( $p$ ) モデルにおける自己回帰項の各係数である。本研究では、特性方程式を(24)式と定義する。定常条件を満たさない確率過程を ARMA モデルのような線形過程でモデリングすることはできないが、特性方程式の解に 1 を含むとき(多項式  $f(x) = 0$  の解を「根」と一般的に呼ぶが、特性方程式の解に 1 が含まれるときには「単位根」を持つという)、差分処理によって、定常条件を満たす変換データを得ることができる。

初歩的な ARIMA モデルの原理について、Trottier ら[61]が簡潔に説明しており、詳しい方法に関する説明が Brockwell と Davis[14]の教科書に記載されている。次に、単位根の有無を調べるために ADF 検定[17, 18, 53]を行った。ARIMA モデルの各係数の推定には R の Arima 関数[27]を用い、AIC を同時に求めた。各係数の推定には、2003 年 4 月から 2011 年 3 月までのデータを用い、 $0 \leq p \leq 5$  の範囲で自己相関項の係数 AR( $p$ )を推定し、 $0 \leq q \leq 5$  の範囲で移動平均項の係数 MA( $q$ )を推定した。擬似ポアソン回帰モデルによって周期成分を計算し、Brockwell と Davis の方法[14]に従って当該周期成分を差し引いており、これには全体の平均が含まれているため、ARIMA モデルの推定を定数項なしで行った。さらに、ドリフト項(トレンド)ありとなしの条件で推定を行い、AIC を基準として比較および選択を行った。

AIC を基準として選択された ARIMA モデルを Ljung-Box 検定[40]で調べることにより、モデルの残差に自己相関がなく、各時点において均等に分散していることを確認した。

さらに、パラメータの有意性と AR( $p$ )過程の定常性を調べた。ARIMA モデルの各係数の値が近似的に多変量正規分布に従う[14]こと

に基づき、各係数を標準偏差で除した値を  $t$  値とした  $t$  検定を実施し、各パラメータの有意性を判定した。また、定常性の確認のために、(24)式に示す自己回帰項の各係数を用いた特性方程式の自明でない解の絶対値が 1 より大きいことを確認した。なお、特性方程式の解の計算には R の `polyroot` 関数[31]を用いた。最後に、週周期性(結果に示す)以外の時間依存的効果の有無について、ARIMA モデルの残差を自己相関関数(ACF)および偏自己相関関数(PACF)[14]を計算することによって確認した。なお、ACF および PACF の計算には、欠損値を取り除いた後の残差を用いた。

### 豚小腸廃棄数期待値の予測

上述の方法で作成したモデルを用いて 2011 年 4 月から 2013 年 5 月までの期待値と信頼限界を計算した。次に、計算された信頼限界と同時期における観測データを比較することによる検証を行った。特定の日における廃棄数の期待値の計算は、擬似ポアソン回帰モデルによって計算した各曜日の期待値と ARIMA モデルによる予測値の和をとることによって行った。抗酸菌症による実際に観測された全部または一部廃棄の数が期待値を上回ったときに、当該豚をと畜場に搬入した生産者を特定し、さらに各年度における特定された各生産者の廃棄率を計算した。

### モデルの頑健性試験

ARIMA モデルは、AR または MA の次数のうち大きい方の値によって決定された期間である過去数日の間に記録されたデータを用いて次の日の値を予測するものである。そのために、重要な疑問が提起され

うる。すなわち、直近の廃棄数の変化に対してモデルによる予測がどのくらい影響を受けるのかという問題である。このモデルの頑健性に関する問題を明らかにするために、学習期間の最後の (AR または MA によって決定される) 数日間に実際に廃棄された数と同じ日数分の乱数との置換を次のような手順で行った。学習期間の 8 年間に観測された値に基づく形状パラメータ  $\lambda$  を持つポアソン乱数を発生させ、擬似ポアソン回帰モデルによって計算した各曜日の期待値を差し引いた後に、最少の AIC 値を持つ AR および MA 項数を持つ ARIMA モデルの推定を行った。次に、以下に示す 2 つのモデル間 (model1: 実際に観測された値から各曜日の期待値を差し引いたデータから生成されたモデル、model2: 学習期間の最後の数日を上述の方法で乱数と置換し、生成したモデル) について検証期間における逸脱度差を計算し、有意水準  $p < 0.05$  でカイ二乗検定を行った。以上の統計処理を 100 回試行した。

### 3. 結 果

#### 周期性の検討

FFT を用いて生成したピリオドグラムは、1 週間周期の極大値を示した (図 4)。ダミーデータとして平日のメジアンと置換されたデータの数は 145 であった。

表 5 に曜日ごとの小腸廃棄数の変化とともに、各曜日の平均と殺頭数、2.5 及び 97.5 パーセンタイル値を示した。抗酸菌症による小腸廃棄数として、1,000 頭あたりの廃棄数と分散パラメータ (dispersion parameter は 7.1) を仮定した GLM によって推定された 1 日あたりの平均廃棄数を示し、さらに有意差のある曜日の組み合わせも示した。

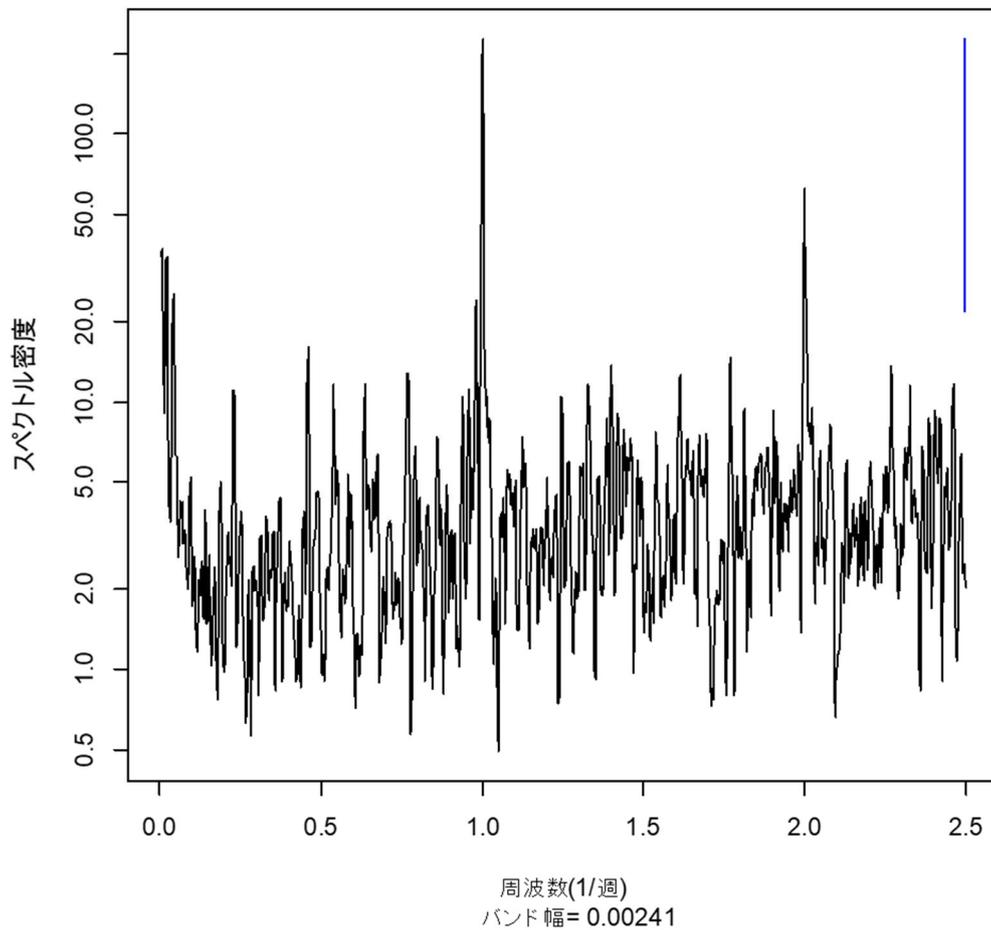


図 4. 平日のみとなるよう前処理された日次廃棄数のピリオドグラム  
 X 軸は週あたりの頻度を示しており、最大値を示すピーク(周波数  
 =1.0/週)は 1 週間周期に対応する。Y 軸は、日次廃棄数変動のスペク  
 トル密度を示す。右上の青線は、95%信頼限界を示す。バンド幅は、  
 スペクトル平滑化を行うウィンドウのバンド幅を示す。

表 5. 2003 年 4 月 1 日～2011 年 3 月 31 日における曜日ごとの小腸廃棄数の変化

曜日	と畜頭数	操業日における一日あたり のと畜頭数 <sup>a)</sup>	1,000 頭あたりの 廃棄数	1 日あたりの 廃棄数 <sup>b)</sup>	有意差のある曜日の 組み合わせ <sup>c)</sup>
月	267,000	640(578-912)	9.5	7.1	火、水、木、金
火	257,000	650(515-872)	5.9	3.8	月、水、木、金
水	293,000	740(572-915)	2.2	1.6	月、火、木、金
木	276,000	696(543-878)	3.3	2.3	月、火、水
金	267,000	674(502-851)	4	2.7	月、火、水
土	23,000	525(400-620)	2.3	1.2	月、火、金
日	1,000	277(230-322)	0	<0.01	-
合計	1,385,000		4.8		

a) 括弧内の数値は、2.5 および 97.5 パーセンタイル値を示す。

b) GLM による推定値

c) GLM による  $p$  値に基づく判定。有意水準は  $p < 0.05$ 。

図 5 および 6 には、1 日あたりの小腸廃棄数と廃棄率の曜日ごとの変化がわかるようにするために、箱ひげ図を示した。1 日あたりのと畜豚頭数はほぼ一定であるが(表 5)、小腸の廃棄数には、曜日によって差がある(表 5、図 5)。小腸の廃棄数は月曜に最も多く、次に火曜が多い。平均廃棄率(0.485%)を超えた日からなる、一部分のデータを用いて解析すると、平均廃棄率を超えた日の数には、類似した曜日の偏りが認められ、さらに曜日ごとの小腸廃棄数についても曜日の偏りが認められた(表 6)。

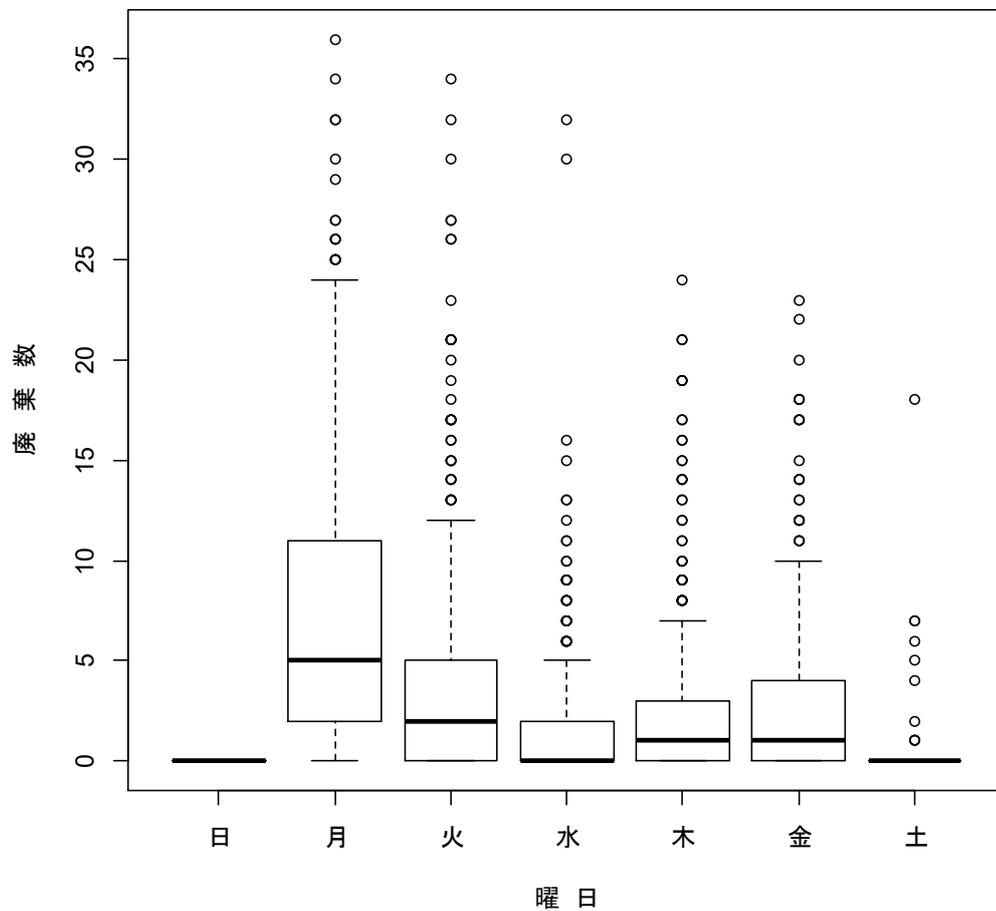


図 5. 各曜日における 1 日あたり小腸廃棄数の箱ひげ図

X 軸は曜日を示し、Y 軸は、2 か所の管内と畜場における 1 日あたりの抗酸菌症による小腸廃棄数を示す。

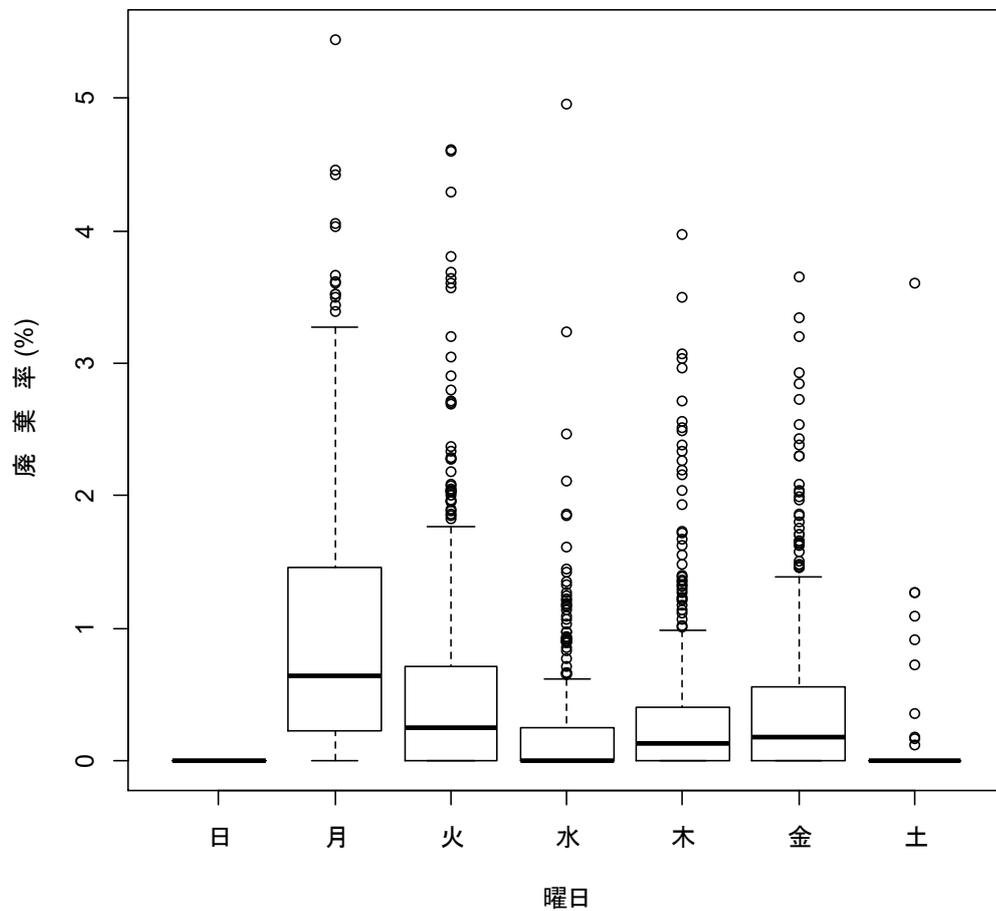


図 6. 各曜日における 1 日あたりの小腸廃棄率の箱ひげ図

X 軸は曜日を示し、Y 軸は、2 か所の管内と畜場における 1 日あたりの抗酸菌症による小腸の廃棄率(%)を示す。

表 6. 平均廃棄率を超えた日における廃棄率の変化

曜日	日数	1日平均廃棄率 <sup>a)</sup>
月	207	11.3(4.0-28.7)
火	134	9.7(3.0-27.0)
水	56	7.7(3.4-24.8)
木	81	8.5(3.0-21.0)
金	111	7.4(3.0-18.5)
土	6	7.8(4.1-16.6)
日	0	0

a) 括弧内の数値は、2.5 および 97.5 パーセンタイル値を示す。

表 7 には、2003 年 4 月 1 日から 2011 年 3 月 31 日までの、生産者の総数、豚搬入日数および 1 日あたりの平均豚搬入数を示した。月曜日から金曜日までについては 95%信頼限界が重なり合っており、特定の曜日への明らかな集中は認められなかった。図 7 には、年間豚搬入頭数別生産者数のヒストグラムを示した。年間搬入頭数のメジアンは、365.6 頭で平均は、1,717.0 頭である。大半の生産者の年間搬入頭数は、500 頭未満である。

表 7. 2003 年 4 月 1 日から 2011 年 3 月 31 日までの、曜日ごとの豚搬入生産者数の変化

曜日	搬入生産者数	搬入日数	1日あたりの搬入生産者数*
月	6,711	357	18.8 (13-25)
火	5,237	396	13.2 (6-24)
水	6,144	396	15.5 (9-23)
木	6,349	397	16 (9-24)
金	6,589	397	16.6 (10-23)
土	283	44	6.4 (3-12)
日	4	2	2.0 (0-3)

\*括弧内の数値は、2.5 および 97.5 パーセンタイル値である。

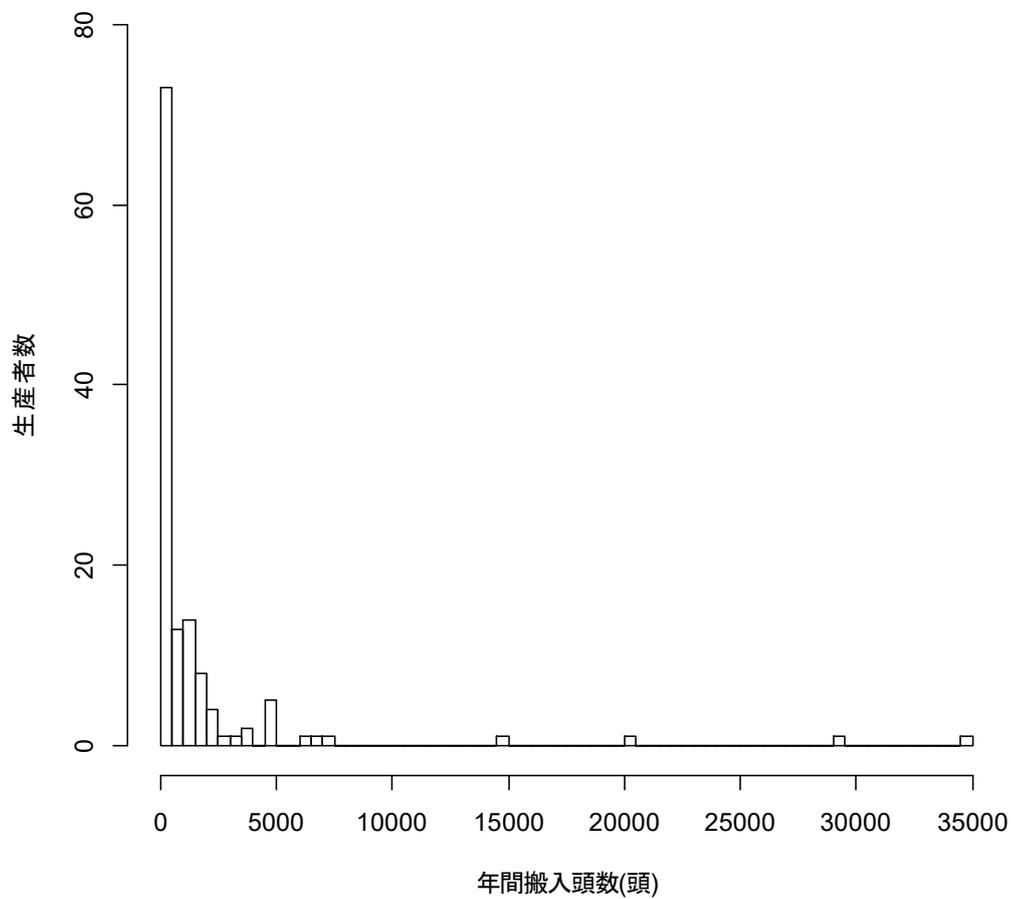


図 7. 2003 年 4 月 1 日から 2011 年 3 月 31 日までの、年間豚搬入頭数別生産者数ヒストグラム

### ARIMA モデルの推定:

ADF 検定の結果、単位根が存在するという帰無仮説は、棄却された。したがって、差分もしくは和分(I)の次数は 0 であることが判明した。また AIC を基準として推定した結果、定数項がゼロである ARIMA(3,0,4)モデルが選択された。ARIMA(3,0,4)モデルの 3,0,4 の各数値は、AR の次数が 3、差分の次数が 0、MA の次数が 4 であることを意味している。Ljung-Box 検定( $p < 0.05$ )の結果、残差は独立であり、時間非依存的に分布していることが示された。モデルの各係数の有意性のための  $t$  検定は、3 つの AR 項および 4 つの MA 項の全ての係数は有意にゼロと異なることを示した(表 8)。

表 8. ARIMA(3,0,4)モデルにおける各係数の推定値

パラメータ	係数 <sup>c)</sup>	標準誤差 <sup>d)</sup>	<i>t</i> 値 <sup>e)</sup>	<i>p</i> 値
AR <sup>a)</sup> (1)	1.44	0.026	54.8	< 0.001
AR (2)	-1.37	0.042	-32.4	< 0.001
AR (3)	0.91	0.025	36.8	< 0.001
MA <sup>b)</sup> (1)	-1.52	0.033	-45.5	< 0.001
MA (2)	1.51	0.05	30.2	< 0.001
MA (3)	-1.07	0.038	-28.2	< 0.001
MA (4)	0.18	0.023	7.6	< 0.001

a) AR: 自己回帰(auto-regressive)

b) MA: 移動平均(moving average)

c) Arima 関数を用いた最尤推定で計算した係数

d) 最尤推定の標準誤差

e) 各係数を標準偏差で除した値

さらに、特性方程式の非自明解は 1 以上であったことから、この ARIMA モデルの定常性が確認された。擬似ポアソン回帰モデルによって計算された各曜日の期待値を差し引いたあとの ARIMA モデルの残差の ACF および PACF のグラフは、明確なピークを示さず、残差は自己相関がない無作為なデータ列であるとみなされた(図 8a および 8b)。

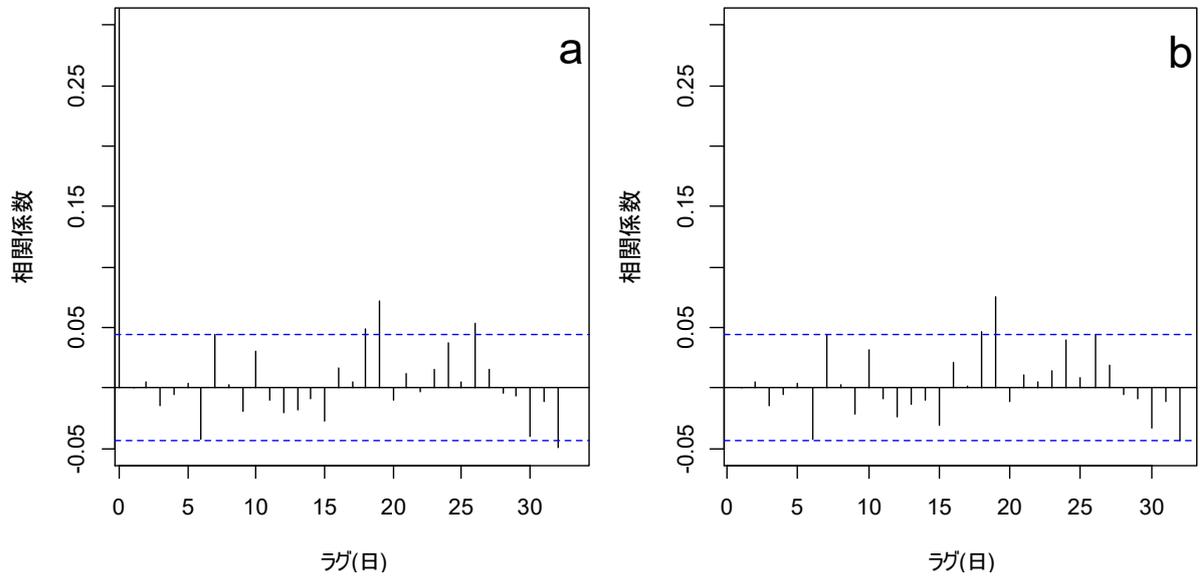


図 8.各曜日の期待値を差し引いたあとの ARIMA モデルの残差の ACF(a)と PACF(b)

X 軸は Lag の日数を示し、Y 軸は相関係数を示す。残差は、Lag 値に関わらずランダムに分布している。

## 豚小腸廃棄数期待値の予測

実際に廃棄された小腸の数が95%信頼限界を上回った日数は、2011年度では3回で、2012年度では17回であった(図9、10)。95%信頼限界の下限值は0を下回っているため、図中には示されていない。これら20日の各操業日における小腸廃棄数の最小値、メジアンおよび最大値は、13、18.5および49であり、これら全ての値は、2011年度および2012年度の全期間における1日あたりの平均廃棄数(2.05)を超えていた。抗酸菌症による実際に観測された小腸の廃棄の数が95%信頼限界を上回ったときに、当該豚をと畜場に搬入した計3生産者の総搬入日数、廃棄数の最小、メジアンおよび最大値を表9に示す。

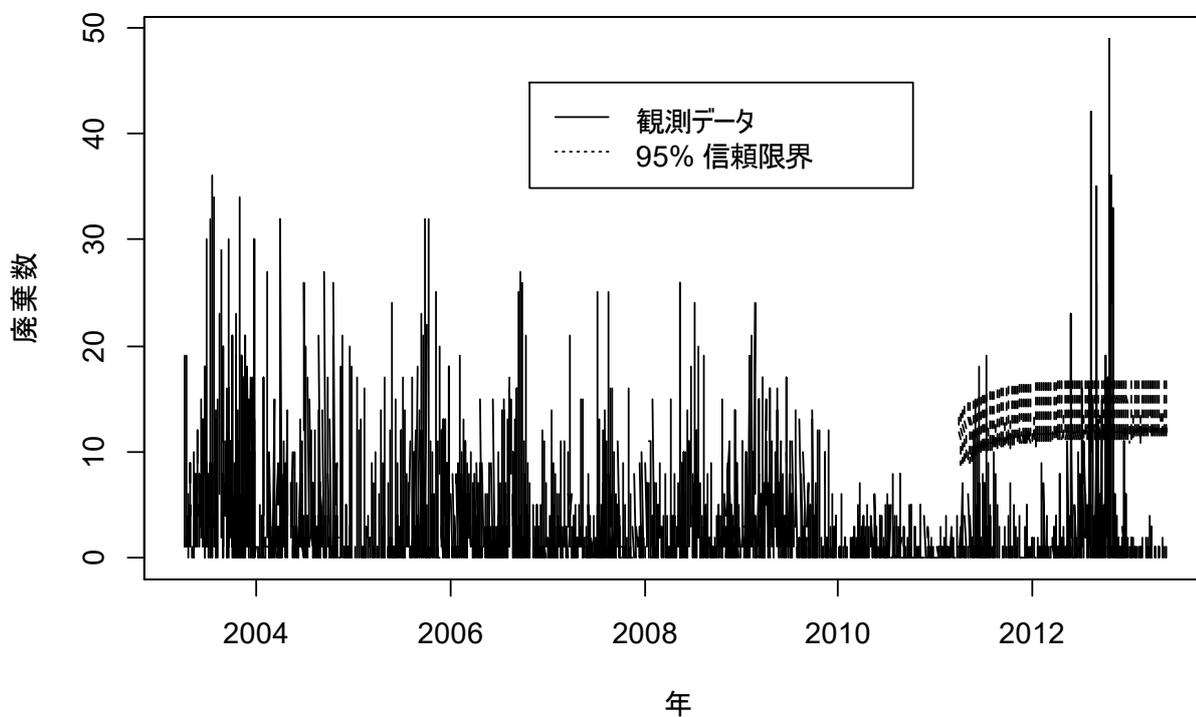


図 9.実際に観測された抗酸菌症による小腸廃棄数と ARIMA モデルによって予測された廃棄数の 95%信頼限界  
 実際の廃棄数が信頼限界を超えたことの判断は、図 9 からではなく、計算値に基づき行う。

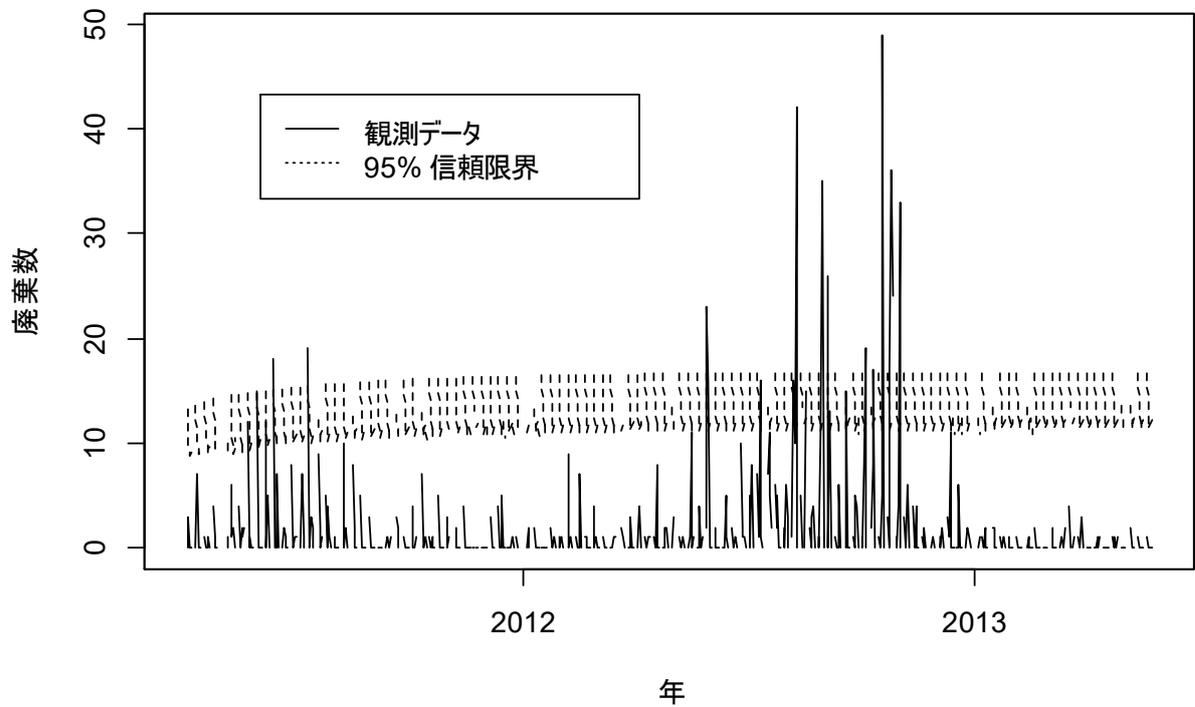


図 10. 2011 年 4 月から 2013 年 5 月までの期間について図 9 を拡大表示したもの

表 9.モデルによって検出された 3 生産者の廃棄率

年度	順位	生産者名	年間廃棄率 <sup>b)</sup>	最小 <sup>b)</sup>	メジアン <sup>b)</sup>	最大 <sup>b)</sup>	搬入日数 <sup>a)</sup>
2011	1	P	13.6%	0.0%	0.0%	77.3%	45
	2	Q <sup>c)</sup>	10.0%	0.0%	0.0%	50.0%	49
	3	R	8.0%	0.0%	0.0%	53.8%	24
2012	1	S	32.9%	32.9%	32.9%	32.9%	1
	2	Q <sup>c)</sup>	30.5%	0.0%	23.5%	89.5%	49
	3	T	23.1%	0.0%	24.0%	49.0%	13

a) 当該生産者が豚を搬入した日数

b) 抗酸菌症による小腸廃棄率

c) 2011 年と 2012 年の生産者 Q は同じ生産者である。

廃棄数が信頼限界を超えた全ての日において、各年度の抗酸菌症による廃棄率が上位3位以内に属する生産者のいずれかが、と畜場に豚を搬入していた。これら3生産者の年間廃棄率は2011年度ではそれぞれ8.0%、10.0%および13.6%であり、2012年度では23.1%、30.5%および32.9%であった。したがって、検査所における年間廃棄率が1%未満であることを考慮すると、これらの生産者の農場で抗酸菌症のアウトブレイクが起きていた可能性は極めて高い。これらの高い廃棄率を示す生産者の中に、両年度で上位3位に含まれている1生産者が含まれていた。

### モデルの頑健性試験

AICを基準に選択されたARIMAモデルは、ARIMA(3,0,4)モデルであり、モデルによる期待値の計算には、AR(3)またはMA(4)の大きい方の次数と等しい日数のデータが用いられる。したがって、検証期間の最初の期待値を計算するために、学習期間の最後の4日間の観測データが用いられる。そこで、それら4つの数値をポアソン分布( $\lambda=3.37$ )する乱数と置換した。置換したデータを用いたARIMAモデルの推定を100回行った結果、最小のAICをとるものとしてARIMA(3,0,4)が99回選択され、ARIMA(4,0,3)が1回選択された。従って、ARIMA(3,0,4)モデルを次の検討に用いることが適当であることが明らかとなった。次に乱数に置換されたデータから生成したモデルと観測値からのデータから生成したモデルの逸脱度の差のカイ二乗検定(自由度4)を100回繰り返した結果、99回で統計学的有意差( $p<0.05$ )は認められず、一回のみ有意差が認められた。以上の結果より、ARIMA(3,0,4)モデルが統計学的に頑健であると結論した。

#### 4. 考 察

本研究は、豚生産農場における抗酸菌症のアウトブレイクと散発的発生をと畜検査データの時系列分析から識別することを企図したものである。結果に示したとおり個体レベルの有病率は極めて低いが、農場レベルにおける散発はよく見られるものである。Moritaら[46]は、群馬県内において1988年から1990年までの間にMAICの散発的な感染が1,200の養豚場中870(72.5%)で検出されたことを報告している。従って、毎日のと畜検査の生データのみから散発例とアウトブレイクを区別することは容易ではない。しかし、本研究で用いられた時系列分析は、検査当日に区別が可能な実用的な方法であることが示された。本章小緒で述べた通り、本疾病の消費者への公衆衛生上の影響の大きさや、生産者への経済的影響の大きさを考慮すると、アウトブレイクの検出は重要である。Komijnら[34]は、豚が、ヒトにおける*Micobacterium avium*感染の重要な媒介動物となっている危険性があることを報告している。解体当日の内臓検査からの検出症例数をデータ処理することによって農場アウトブレイクを迅速に検出することは、公衆衛生上のリスクを最小限にし、当該生産農場の速やかな衛生的改善を促す可能性がある。また、と畜検査と家畜防疫が協力できれば人獣共通感染症への効果的な対応も可能であろう[1]。このように本研究は、豚の抗酸菌症という一つの疾病を対象として扱っているが、本研究手法は、食用家畜における疾病のアウトブレイクを生産農場レベルで検出するのに広く用いられうる。

本研究では、小腸の毎日の廃棄数が95%上方信頼限界を上回ったかどうか特に注目し、と畜場に搬入される豚の疾病による廃棄率の上

昇が起こっている生産者を特定することができた。これらの外れ値事象は、検査当日の豚搬入元である各々の生産農場におけるアウトブレイクに対応していると推定できる。2003年から2011年までの全てのデータと廃棄数が平均を上回った時のみのデータにおける毎日の小腸廃棄数の時間依存性パターンが類似していることは、本研究の時系列モデリングの有効性を支持している。これまで用いられてきたと畜検査データの統計学的な分析手法は、月間および年間の廃棄数総計を計算することのみであった。したがって、従来の方法では、廃棄総数が算出されるまでアウトブレイクの有無は把握されず、よく見られる散発発生影に隠されたアウトブレイクが実際に明らかになるまで長時間が経過する可能性があると思われる。しかし、本研究で試みた時系列分析によってアウトブレイクを従来よりも早期に検出できる可能性が示された。

本研究は、上述の利点がある一方で三つの制約がある。一つは、抗酸菌症診断の感度と特異度について考慮されていないことである。と畜検査における豚の腸抗酸菌症の診断は、肉眼および組織学的検査などの画像検査のみで行われるため、誤診が発生する場合もある。しかし、費用と時間の制限により、通常肉食肉検査で実施している組織学的検査の感度と特異度を確認することはできなかった。実際に、分子生物学的技法を利用せずに、古典的な技術のみに依存することによる感染の見落としは、存在する。たとえば、乾酪壊死を有する239個の豚と体由来のリンパ節を用いた検討では、チールネルゼン染色による検査の感度は79.1%、一方、培養試験の感度は69.5%であった[34]。さらに、出版されているデータ[34]からカッパー係数(一致度)[19]を計算すると0.016となり、「わずかに一致」と解釈される。二つめの制約

は、抗酸菌症が原因でと畜場で廃棄された小腸の数を、本研究では目的変数として用いていることである。疾病の罹患は二値変数である[22]ことから、特定の日にと殺された豚の有病率の方がアウトブレイクを疑う農場を特定するためのモデリングの変数として適しているであろう。しかし、抗酸菌症は年間の廃棄率が低い(2012年度 0.36%:北海道東藻琴食肉衛生検査所、未発表)稀な疾患であるために、毎日の廃棄数データはゼロが多く過分散しており、本研究のモデリングでは有病率を目的変数としては用いていない。三つ目の制約は、抗酸菌症の農場アウトブレイクを検出するうえで、対象農場の規模を考慮していなかったことである。そのため、本研究で用いたモデルは、小規模農場のアウトブレイクを見落とす危険がある。その一方で、大規模農場での疾病発生に対しては感度が高く、散發的発症例を検出してしまう可能性もある。しかし、図9に示した通り、検出レベルは十分に低く小規模農場におけるアウトブレイクであっても概ね検出できると考えられた。さらに、大規模農場は、より小さな独立した複数の生産豚舎群から構成されており、実際のアウトブレイクはその中の一つで起こるはずなので、農場の規模によらずアウトブレイクは検出可能と考えられ、三つ目の制約も重大なものではないであろう。本研究では、廃棄数が一週間周期の変動を示していることをピリオドグラムによって示した。特に、非操業日のデータを除去して実施したFFTの結果から、この周期性が示されており、月曜に特に高い廃棄数が観測されている(表5、図5)。と畜場に搬入される豚の数に曜日の偏りが無い(表4)ことから、この廃棄数の周期性は、と畜場への搬入数とは相関しない。そこで、抗酸菌症の有病率が高い生産者が特定の曜日(月曜日)に搬入する傾向があると仮定したが、そのような事実を確認することが

できなかった(著者、未発表)。現時点で、月曜日に高率となる周期性の原因については不明であり、今後の課題とされた。

ARIMA モデルは、時系列データの最後の数個のデータを用いてそれに続く最初の期待値を予測し、この期待値を次に続く期待値の計算に用い、この処理を同様に続けていく。従って、終わりの数個のデータを統計学的に分布する乱数に置き換えることによる影響を頑健性試験の実施により評価した。その結果、最後の数個の観測値を乱数に置換することは、モデルに対して有意に影響せず、このことから(最後の数個の観測値に先立つ)長期間にわたる情報を十分に織り込んだものであることが明らかとなった。

## 5. 小 括

上記のとおり制約はあるが、時系列分析は、抗酸菌症の潜在的なアウトブレイクを迅速に検出するのに有用である。本方法は、ゼロが多く過分散したデータに対して ARIMA モデルを利用し、時系列分析を行うことができることから、発生が稀な他の食品由来の人獣共通感染症に対しても利用可能であり、利用によって食品の安全性を一層向上させるものである。

本章で用いた方法は、上記の利点を時系列分析法としてよく用いられている ARIMA モデルで実現している点で意義があると考えられるが、周期性の解析と ARIMA モデルの推定の二段階でモデリングを行っており、一旦作成したモデルを持続的に使うにはよいが、食肉検査業務のルーチンとして新たなデータの追加の度にモデルを更新する作業を行うことは容易ではない。そこで、ゼロの多いデータの分布を精査し、あてはまりのよい分布を用いてワンステップでモデリングを行

う方法について、第三章で述べる。

### 第 III 章 豚エキノコックスの農場アウトブレイク検出のためのゼロ過剰カウントデータを用いた two-part モデルに基づくと畜検査データの時系列分析法

#### 1. 小 緒

*Echinococcus multilocularis* は、エキノコックス症を引き起こす条虫であり、このエキノコックス症は、北半球の各地で認められ、病原性が高く致死性の寄生虫性人獣共通感染症である [21, 66]。豚は、自然界における中間宿主ではなく [41, 42]、ヒトも感染した豚を摂食することによって感染することはない。しかし、*E. multilocularis* に感染した豚を検出することによって、大半の豚が 6 ヶ月齢程度の若齢でと殺されるために、感染性の虫卵汚染が豚舎内に生起した時期と場所を示すことができる [71]。そのような点で、公衆および家畜衛生上、重要な知見を披瀝する。一方、ヒトでは感染した時期と場所を特定することは、感染から発症までの期間が長いために極めて困難である。豚では、虫卵を含むアカギツネの糞便を経口摂取することにより *E. multilocularis* 感染が起こり、虫卵を排泄するアカギツネは、飼料、胎盤や死産胎仔の組織を食べるために養豚施設内外を徘徊する [62, 71]。

検査所では、主として道東産の家畜のと畜検査を行っており、豚のと体は、肉眼病理学的検査を受けている。肝臓に境界明瞭な白色結節を呈すると体は豚エキノコックス症と診断され、肝臓は廃棄される。豚エキノコックス感染を鑑別診断するために、過ヨウ素酸 Schiff 陽性のクチクラ層の確認が行われる [54]。 *Ascaris suum* および非定型抗酸菌症は類似した結節を形成するが、組織学的鑑別が可能である。回虫症の場合は、グリソン鞘の線維化と好酸球の集簇が認められ、抗酸菌

症の場合には、チールネルゼン染色で結節が染色される。1カ月間に豚エキノコックス症と診断された豚の数が所定の基準を上回っている場合には、検査所は生産者に対して豚エキノコックス感染症が豚に認められたことを情報提供するとともに、対策方法を記載したパンフレットも送付している。ただし、所定の基準は、有病率と感染が豚から検出される期間に関する定数値であるが、アウトブレイクの定義が考慮されておらず、この定数値を用いることによる速報性がどの程度なのかを確認せずに使用している。したがって、合理的かつ客観的な方法論に基づく新基準を作成することが必要であった。

本研究では、豚エキノコックス感染症によって1カ月間に廃棄された肝臓の数の期待値が時系列分析法によって算出された。時系列分析は、等間隔で測定された連続データ評価するのによく用いられている[2, 3, 49, 64]。しかし、時系列分析の方法は非常に多様であるため、と畜検査データに応用する際は、適切な方法がデータの分布に基づいて入念に選択されなければならない。

一般的に、豚の頭数のようなカウントデータをモデリングして統計学的分析を行う際にはポアソン分布のような離散確率分布を仮定する[14]。ポアソン分布は平均と分散が等しいが、豚エキノコックスに感染する豚の頭数データは、ゼロを多く含むカウントデータであり、この多数のゼロにより平均よりも分散がはるかに大きい過分散となっている。したがって、豚エキノコックスに感染する豚の頭数データにポアソン分布をあてはめることは適切ではない。この過剰なゼロの問題を取り扱うために、ゼロ過剰モデルやハードルモデルなどの「two-part モデル」と呼ばれるものが使われることが多い[36, 47]。これらのtwo-part モデルの背景には、アウトカム変数が混合分布するという考

え方がある[36, 47]。ゼロ過剰モデルは、観察されるゼロ値が2種類の発生源のうちの一つであると仮定している。すなわち、通常のポアソン分布(または負の二項分布)由来のサンプリングゼロと、二項分布由来の構造的ゼロである。ハードルモデルは、ゼロの発生源は1つと仮定している。すなわち、二項分布由来の構造的ゼロである[25]。two-partモデルは多様なデータを解析するのに用いられている[15, 23, 33, 44, 59]が、two-partモデルを利用した時系列分析に焦点を当てた研究は少ない[26, 30, 72]。さらに、two-partモデルを用いたと畜検査データの時系列分析に関する研究論文は未だ見られない。

本研究の目的は、最も適切な統計学的分布をもつ時系列分析法を利用してと畜検査データから生産者ごとに毎月の豚エキノコックス感染数を予測する方法を開発することである。研究の最終目的は、感染豚の数の異常な増加を検出することを可能にする統計学的基準の確立である。

## 2. 材 料 と 方 法

### 研究のデザイン

本研究は、豚エキノコックス症の検出と、生産者への農場衛生改善目的の情報提供における迅速性と効率に関し、時系列モデルと従来の基準を比較したものである。最初に、毎月の豚肝臓廃棄数のモデルとして適切な分布を選択した。次に、時系列データ(原系列)と原系列から特定期間(ラグ)ずらしたデータとの間の自己相関を調べることによって時系列モデルを作成した。なお本研究では、原系列を特定期間ずらすことをラグ処理と呼ぶ。例えば、原系列の時系列データが $\{y_i, y_{i+1}, y_{i+2}, \dots, y_{i+p}\}$ であったときに、ラグ=1でラグ処理したデータは

$\{y_{i-1}, y_i, y_{i+1}, \dots, y_{i+p-1}\}$ であり、ラグ= $m$ でラグ処理したデータは $\{y_{i-m}, y_{i+1-m}, y_{i+2-m}, \dots, y_{i+p-m}\}$ である。最後に、84生産者のデータを用いて農場アウトブレイク検出の迅速性について二つの基準を定量的に比較するために統計学的分析を行った。これらの84生産者のうち、50生産者は小規模生産者(毎月の平均豚搬入数が100頭未満)であり、34生産者は大規模生産者(毎月の平均豚搬入数が100頭以上)である。データ源となる集団は、北海道東部の養豚生産者であり、解析結果の適用対象は日本国内全ての養豚生産者である。しかし、本研究の結果は、過剰なゼロを有するデータの種々の時間的解析に外挿できる。本研究の全ての統計学的分析は、統計ソフトウェア R[29]を用いて実施された。

#### 研究対象の地域とデータ収集

検査所は、北海道網走郡に位置しており、管轄している2つのと畜場では年間150,000~200,000頭の主として道東産の豚がと殺されている。豚のと体は検査所のと畜検査員による肉眼病理学的検査を受け、境界明瞭な白色結節を有する肝臓(図11)は豚エキノコックス症と診断されている。



図 11 豚のエキノコックス症による肝臓の白色結節[76]  
直径 5mm 程度のものであることが多い。

本研究では、2つの検査所管轄と畜場からのと畜検査データを用いており、データ収集の期間は、2003年4月から2015年11月までである(同期間に、143生産者によって総頭数2,317,430頭の豚が搬入されている)。これらのデータのうち、84生産者(全期間で1,397,608頭)の廃棄肝臓数月次集計データを本研究で用いた。これらの84生産者は、観察期間のうち合計で12か月以上の期間に豚を搬入し、かつ搬入した豚の肝臓が豚エキノコックス症で1個以上廃棄されている生産者である。

### モデルのための統計学的分布の選択

時系列モデルに用いるのに適切な分布を選択するために、6種類の分布(ポアソン分布、負の二項分布、ゼロ過剰ポアソン分布[ZIP]、ゼロ過剰負の二項分布[ZINB]、ハードルポアソンモデル[HP]およびハードル負の二項分布[HNB])を最尤推定法によってデータに当てはめた。データとモデルの逸脱度をこれらの6種類のモデルについて計算し、比較した。ZIP、ZINB、HPおよびHNB分布について、psclライブラリ[75]のzeroinfl関数とhurdle関数を用いて計算した。ZINBおよびHNB分布を選択し(結果に示す)、以下の検討で用いた。

### 時系列モデルの作成

次に、説明変数として自己相関のあるラグ処理データを選択(どのラグが自己相関しているか)し、説明変数の係数の推定を行い、時系列モデルを作成した。

本研究では、以下の(26)式の確率質量関数であらわされるZINB分布を有する時系列モデルを仮定した。

$$P(Y_t = y_t) = \begin{cases} \omega_t + (1 - \omega_t)f(0) & (y_t = 0) \\ (1 - \omega_t)f(y_t) & (y_t > 0) \end{cases} \quad \dots (26)$$

ここで  $P(Y_t = y_t)$  は、時点  $t$  における肝臓廃棄数  $Y_t$  の実現値が  $y_t$  である確率を表している。 $\omega_t$  は、時点  $t$  におけるゼロ過剰パラメータであり、 $f(y_t)$  は、(27)式で与えられる負の二項分布を表している。

$$f(y_t) = \frac{\Gamma(\theta + y_t)}{\Gamma(\theta)\Gamma(y_t + 1)} \left(\frac{\theta}{\theta + \lambda_t}\right)^\theta \left(\frac{\lambda_t}{\theta + \lambda_t}\right)^{\lambda_t} \quad \dots (27)$$

ここで  $\theta$  は、負の二項分布のディスパージョンパラメータであり、 $\lambda_t$  は、時点  $t$  における負の二項分布の期待値を表している。さらに、HNB 分布を持つ時系列モデルとして、(28)式で与えられる確率質量関数を仮定した。

$$P(Y_t = y_t) = \begin{cases} \omega_t & (y_t = 0) \\ (1 - \omega_t) \frac{f(y_t)}{1 - f(0)} & (y_t > 0) \end{cases} \quad \dots (28)$$

ここで  $f(y_t)$  は、(27)式であらわされる負の二項分布である。

ZINB および HNB 分布のパラメータ  $\lambda_t$  と  $\omega_t$  について、以下の(29)式および(30)式を用いてモデリングした。

$$\log \lambda_t = \beta_0 + \sum_{i=1}^q y_{t-i} \beta_i \quad (q \leq 12) \quad \dots (29)$$

$$\text{logit}(\omega_t) = \gamma_0 + \sum_{i=1}^q y_{t-i} \gamma_i \quad (q \leq 12) \quad \dots (30)$$

ここで、 $\beta_0$  および  $\gamma_0$  は、回帰式の切片を表している。 $\beta_1, \beta_2, \beta_3, \dots, \beta_{12}$  および  $\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_{12}$  は、回帰係数を表している。(29)式および(30)式を(31)式および(32)式のように書き直すことができる。

$$\lambda_t = e^{\beta_0 + \sum_{i=1}^q y_{t-i} \beta_i} \quad (q \leq 12) \quad \dots (31)$$

$$\omega_t = \frac{1}{1 + e^{\gamma_0 + \sum_{i=1}^q \gamma_t - i \gamma_i}} \quad (q \leq 12) \quad \dots (32)$$

すなわち、(29)式はオートエクスポネンシャル回帰の式であり、(30)式はオートロジスティック回帰の式である。

回帰式の項に用いるラグ処理データを次の(i)~(iv)の方法で選択した。(i) 84生産者の肝臓廃棄数の時系列データから計算されたACFを調べ、有意なラグが存在するか確認した。ACFを計算するために、acf関数[63]を用いた。季節性の変化を表すために、以下に示す回帰モデルの説明変数として、実際の数値から1月単位で遅延した過去の値のうち1年以内のラグ処理データ(すなわち、lag=1,2,3...,12)であって、有意な自己相関を有するものを用いた。一般的に、肝臓廃棄数とラグ処理データとの間に有意な自己相関があるとき、説明変数に用いるラグ処理データの数を増やすことによって残差は減少する。(ii)モデルに用いる説明変数の数を $\alpha$ とし、説明変数として用いるラグ処理データの数をこれ以上増加させても残差が減少しないときの最小の $\alpha$ を調べた。 $\alpha$ の最小値を求めるために、ラグの最も小さなラグ処理データから最もラグの大きいラグ処理データへステップワイズでモデルに加えていった。例を挙げると、廃棄肝臓数の時系列データから計算したACFの有意なラグが2、4および8だった場合、 $\alpha=1$ のときは、lag=2の過去の値が説明変数であり、 $\alpha=2$ のときはlag=2および4の過去の値が説明変数であり、 $\alpha=3$ のときは、lag=2,4および8の過去の値が説明変数となる。(iii) 84生産者の各々の時系列データへのZINBおよびHNB回帰モデルのあてはめを $1 \leq \alpha \leq 8$ で行った。(iv) モデルの残差のACFを計算した。84生産者の時系列データを用いて計算した

ACF の有意なラグの最大数が 8 だったため、 $\alpha$  の最大値を 8 とした。

ただし、ACF チャートが有意な自己相関係数ではないラグ値  $t$  における回帰係数 ( $\beta_t, \gamma_t; t \geq 1$ ) をゼロとした。 $\beta$ 、 $\gamma$  および  $\theta$  を計算するために、`zeroinfl` 関数および `hurdle` 関数を用いた。各係数を計算後、ZINB モデルと HNB モデルのあてはまりの良さを `nested Vuong test`[67] によって比較した。

さらに、時系列モデルの有効性を確認するために、AIC によって修正した `non-nested Vuong test` を用いて、時間依存的な係数のあるモデルと切片だけのモデルを比較した。`Non-nested Vuong test` の計算には、`pscl` ライブラリの `vuong` 関数を用いたが、`pscl` ライブラリに `nested Vuong test` が含まれていないため代替として `vuongtest` 関数[43] を用いた。

## 時系列モデルと従来の基準との比較

作成された時系列モデルの性能の検証において、検出の迅速性は重要である。このため時系列モデルと従来基準のどちらがより迅速に豚エキノコックスの農場アウトブレイクを検出できるか比較を行った。モデルの累積分布関数のパーセンタイル値(以後パーセンタイル値と呼ぶ)を実際の廃棄肝臓数が超えた回数を調べた。R の `dbinom` 関数を用いて(26)式または(28)式からパーセンタイル値を計算した。最初に、ZINB と HNB 分布の 50 パーセンタイル値から 90 パーセンタイル値まで 5 パーセンタイル刻みで計算し、実際の肝臓廃棄数がパーセンタイル値を超えた回数を調べた。従来基準(廃棄肝臓数がと畜数の 1%を超えるか、3 か月連続で 1 個以上の廃棄肝臓がある)と本研究の時系列モデルによる方法を比較するために、84 生産者のデータを用い

て廃棄肝臓数の異常な増加をどちらの基準が早く検出したか、符号検定を用いて判定した。符号検定では  $p$  値を(33)式によって計算した。

$$p = 2 \sum_{i=0}^m {}_N C_i \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{N-i} \quad \dots (33)$$

ここで、 $N$  は双方の基準を用いて検出されたアウトブレイクの総数であり、 $m$  はより少なくアウトブレイクを検出した基準によって得られた検出数の総数である。

### 3. 結 果

#### モデルのための統計学的分布の選択

84 生産者の月次集計データと各モデルの推定値との逸脱度を比較したところ、HNB 分布がもっとも良好な数値を示した。二番目が ZINB 分布、三番目が HP 分布、四番目が ZIP 分布であった(表 10)。全ての生産者において、HNB 分布の逸脱度が最も低かったのに対し、ZINB、HP および ZIP 分布は、HNB 分布とそれぞれ 42、35 および 22 生産者で等しい値であった。計算された逸脱度が最低値だった頻度は、HNB 分布が 84 生産者、ZINB 分布が 42 生産者、HP 分布が 35 生産、ZIP 分布が 22 生産者、負の二項分布が 13 生産者およびポアソン分布が 11 生産者であった。

これらの結果は、ゼロ過剰部分またはハードル部分が必要であることを示している。以上の結果から、ゼロ過剰モデルとして ZINB 分布のみを、ハードルモデルとして HNB 分布を以降の解析で用いた。

表 10. 6つの異なる分布の逸脱(ポアソン分布、負の二項分布、ゼロ過剰ポアソン分布 (ZIP)、ゼロ過剰負の二項分布(ZINB)、ハードルポアソン分布(HP)、ハードル負の二項分布 (HNB))

肝臓廃棄数の順位	ポアソン分布	負の二項分布	ゼロ過剰ポアソン分布 (ZIP)	ゼロ過剰負の二項分布 (ZINB)	ハードルポアソン分布 (HP)	ハードル負の二項分布 (HNB)
1	2,287.6	602.3	1,468.5	599.2	1,468.5	599.2
2	1,551.0	725.9	1,209.0	725.9	1,209.0	725.3
3	1,597.7	564.2	1,123.5	564.2	1,123.5	563.7
4	993.2	695.2	857.8	695.2	857.8	694.3
5	627.1	449.1	518.4	449.1	518.4	447.9
6	804.9	336.0	500.9	336.0	500.9	334.1
7	454.2	319.5	365.4	319.4	365.4	319.4
8	445.5	256.9	343.4	256.9	343.4	255.9
9	506.2	406.2	447.1	406.2	447.1	403.4
10	418.0	370.7	398.8	370.7	398.8	365.7
11	413.4	363.0	385.8	363.0	385.8	360.3
12	412.5	306.1	356.5	306.1	356.5	302.3
13	281.1	264.2	270.6	264.2	270.6	264.0
14	474.4	222.4	289.2	222.4	289.2	221.7
15	277.6	232.3	250.9	232.3	250.9	231.6
16	342.3	267.8	282.0	267.8	282.0	267.8
17	386.4	239.6	271.4	239.6	271.4	239.6
18	388.2	269.9	296.8	269.9	296.8	269.6
19	321.5	254.6	264.5	254.6	264.5	254.6
20	270.0	77.9	142.8	77.9	142.8	77.6
21	236.8	226.0	230.6	226.0	230.6	224.5
22	237.8	208.2	208.5	207.7	208.5	207.7

表 10.6 つの異なる分布の逸脱(続き)

肝臓廃棄 数の順位	ポアソン分布	負の二項 分布	ゼロ過剰 ポアソン 分布 (ZIP)	ゼロ過剰 負の二項 分布 (ZINB)	ハードル ポアソン 分布 (HP)	ハードル 負の二項 分布 (HNB)
23	178.3	123.3	113.6	113.6	113.6	113.6
24	244.4	194.0	200.8	194.0	200.8	194.0
25	221.6	187.0	187.4	186.5	187.4	186.5
26	197.7	163.6	171.7	163.6	171.7	163.1
27	200.6	155.5	166.8	155.5	166.8	154.8
28	182.8	163.4	163.8	163.3	163.8	163.3
29	179.9	101.0	118.8	101.0	118.8	100.8
30	147.5	136.4	137.4	136.4	137.4	136.4
31	178.0	152.2	156.3	152.2	156.3	152.1
32	195.7	81.5	91.4	81.4	91.4	81.4
33	111.1	67.3	73.9	67.0	73.9	67.0
34	170.8	124.7	138.4	124.7	138.4	124.3
35	154.6	126.9	138.8	126.9	138.8	126.0
36	152.9	94.4	108.8	94.4	108.8	94.1
37	126.0	111.9	115.6	111.9	115.6	111.5
38	96.1	68.1	75.1	68.1	75.1	67.9
39	104.1	104.1	40.4	40.4	40.4	40.4
40	118.4	101.4	100.8	100.8	100.8	100.8
41	122.2	84.2	87.8	84.2	87.8	84.2
42	73.3	59.3	67.2	59.3	67.2	58.1
43	85.8	81.7	83.7	81.7	83.7	81.0
44	106.5	71.8	74.6	71.8	74.6	71.8
45	127.9	62.4	68.8	62.4	68.8	62.4
46	110.0	67.1	71.2	67.1	71.2	67.1

表 10.6 つの異なる分布の逸脱(続き)

肝臓廃棄 数の順位	ポアソン分布	負の二項 分布	ゼロ過剰 ポアソン 分布 (ZIP)	ゼロ過剰 負の二項 分布 (ZINB)	ハードル ポアソン 分布 (HP)	ハードル 負の二項 分布 (HNB)
47	71.0	50.7	52.7	50.7	52.7	50.7
48	85.7	77.0	76.5	76.5	76.5	76.5
49	67.8	65.3	66.0	65.3	66.0	65.2
50	61.3	39.0	43.0	39.0	43.0	38.9
51	49.4	27.6	31.8	27.6	31.8	27.6
52	56.4	51.5	52.0	51.5	52.0	51.4
53	57.1	57.1	57.1	57.1	56.8	56.8
54	57.1	57.1	57.1	57.1	56.8	56.8
55	63.4	46.2	47.5	46.2	47.5	46.2
56	56.0	49.9	49.0	49.0	49.0	49.0
57	28.7	26.5	25.9	25.9	25.9	25.9
58	52.3	50.9	50.8	50.8	50.8	50.8
59	44.1	44.1	44.1	44.1	44.0	44.0
60	43.9	43.9	43.9	43.9	43.8	43.8
61	41.5	38.9	38.6	38.6	38.6	38.6
62	43.4	43.4	43.4	43.4	43.2	43.2
63	37.1	37.1	37.1	37.1	37.0	37.0
64	29.7	29.7	29.7	29.7	29.4	29.4
65	18.0	18.0	18.0	18.0	16.8	16.8
66	36.2	32.4	32.0	32.0	32.0	32.0
67	36.1	25.7	23.9	23.9	23.9	23.9
68	27.0	27.0	27.0	27.0	26.9	26.9
69	27.9	27.9	27.9	27.9	27.9	27.9
70	24.4	20.9	20.4	20.4	20.4	20.4

表 10.6 つの異なる分布の逸脱(続き)

肝臓廃棄 数の順位	ポアソン分布	負の二項 分布	ゼロ過剰 ポアソン 分布 (ZIP)	ゼロ過剰 負の二項 分布 (ZINB)	ハードル ポアソン 分布 (HP)	ハードル 負の二項 分布 (HNB)
71	18.7	18.7	18.7	18.7	18.3	18.3
72	19.8	19.8	19.8	19.8	19.5	19.5
73	20.3	20.3	20.3	20.3	20.2	20.2
74	17.3	17.3	17.3	17.3	17.3	17.3
75	15.2	15.2	15.2	15.2	15.1	15.1
76	12.0	12.0	12.0	12.0	12.0	12.0
77	11.5	11.5	11.5	11.5	11.5	11.5
78	11.3	11.3	11.3	11.3	11.3	11.3
79	11.5	11.5	11.5	11.5	11.5	11.5
80	7.1	7.1	7.1	7.1	7.1	7.1
81	8.5	8.5	8.5	8.5	8.5	8.5
82	9.8	9.8	9.8	9.8	9.8	9.8
83	10.1	10.1	10.1	10.1	10.1	10.1
84	9.8	9.8	9.8	9.8	9.8	9.8

## 時系列モデルの作成

図 12 に、ある生産者の  $\alpha = 0 \sim 5$  および 8 における月次集計肝臓廃棄数から推定された ZINB 回帰モデルからの残差の ACF の相関係数を示した。 $\alpha = 0$  のとき回帰式は切片のみとなり、観測値の ACF と  $\alpha = 0$  におけるモデルの残差の ACF は同じものとなるため、観測値の ACF の回帰係数を示していない。 $\alpha$  を増加させるにしたがって有意なラグの数は減少していくが、 $\alpha > 5$  では減少傾向がみられなくなった。

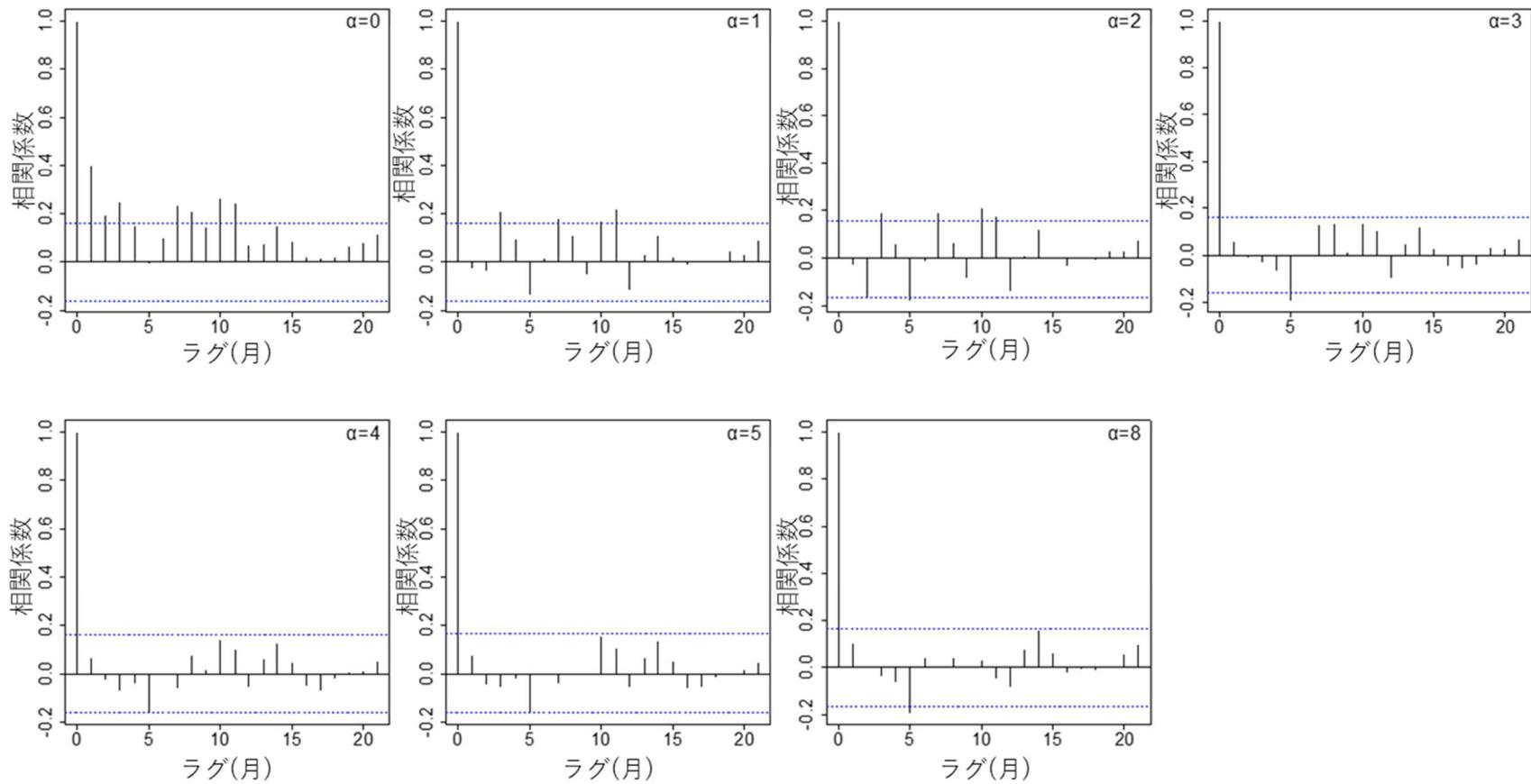


図 12. ある生産者の  $\alpha = 0 \sim 5$  および 8 における豚エキノコックスによる月次集計肝臓廃棄数から推定された ZINB 回帰モデルからの残差の自己相関関数(ACF)  
 X 軸はラグ(月)、Y 軸は相関係数を示す。

図 13～15 に、 $\alpha$  の値とモデルからの残差の ACF における有意なラグの数の関係を示す。モデルからの残差の ACF における有意なラグの数の平均値が最小となるのは図 13 と図 14 に示したように、ZINB および HNB 分布のいずれの場合も  $\alpha = 5$  の時であり、そのときの生産者あたりの有意なラグの数の平均は 0.46 または 0.38 であった。R の acf 関数のデフォルトであるラグの最大値 22 の条件で ACF を計算した。有意なラグの数が 22 の 5% を下回った生産者の数は、 $\alpha = 5$  のとき 76 (ZINB:  $76/84 = 90.5\%$ ) および 79 (HNB:  $79/84 = 94.0\%$ ) であり、 $\alpha = 6$  のとき 75 (ZINB:  $75/84 = 89.2\%$ ) および 77 (HNB:  $77/84 = 91.7\%$ ) であった。84 生産者の総ラグ数の最大値は、1,848 (すなわち、 $84 \times 22$ ) であり、84 生産者のモデルからの残差の ACF における有意なラグの総数は、 $\alpha = 5$  のとき ZINB 分布では 2.1% (すなわち、1,848 個中 39 個) であり、HNB 分布では 1.8% (すなわち、1,848 個中 32 個) であった。有意水準 5% の条件下では、ランダムな変動によって 20 個のラグのうち 1 個は有意となることが期待される [49]。したがって、残差の ACF における有意なラグの数が有意水準 (5%) を下回っていることから、 $\alpha = 5$  で選択された時系列モデルが最適であるという結果が得られた。以後、84 生産者のモデルで  $\alpha = 5$  を用いた。

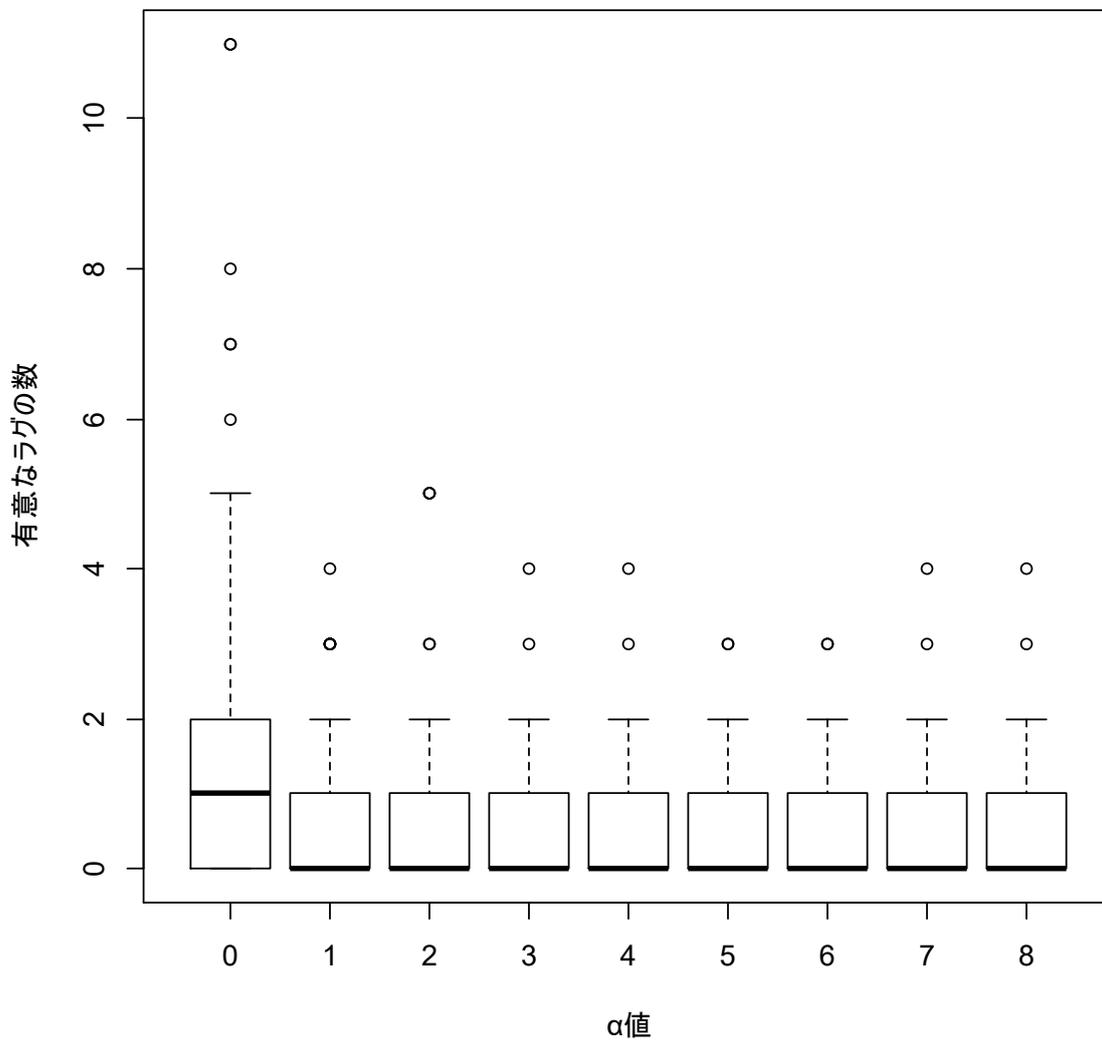


図 13. HNB モデルにおける  $\alpha$  値とモデルの残差の ACF の有意なラグの数の関係

X 軸は  $\alpha$  値を示し、Y 軸は、各  $\alpha$  値の条件の下で推定されたモデルの残差の ACF における有意なラグの数の分布を示す。

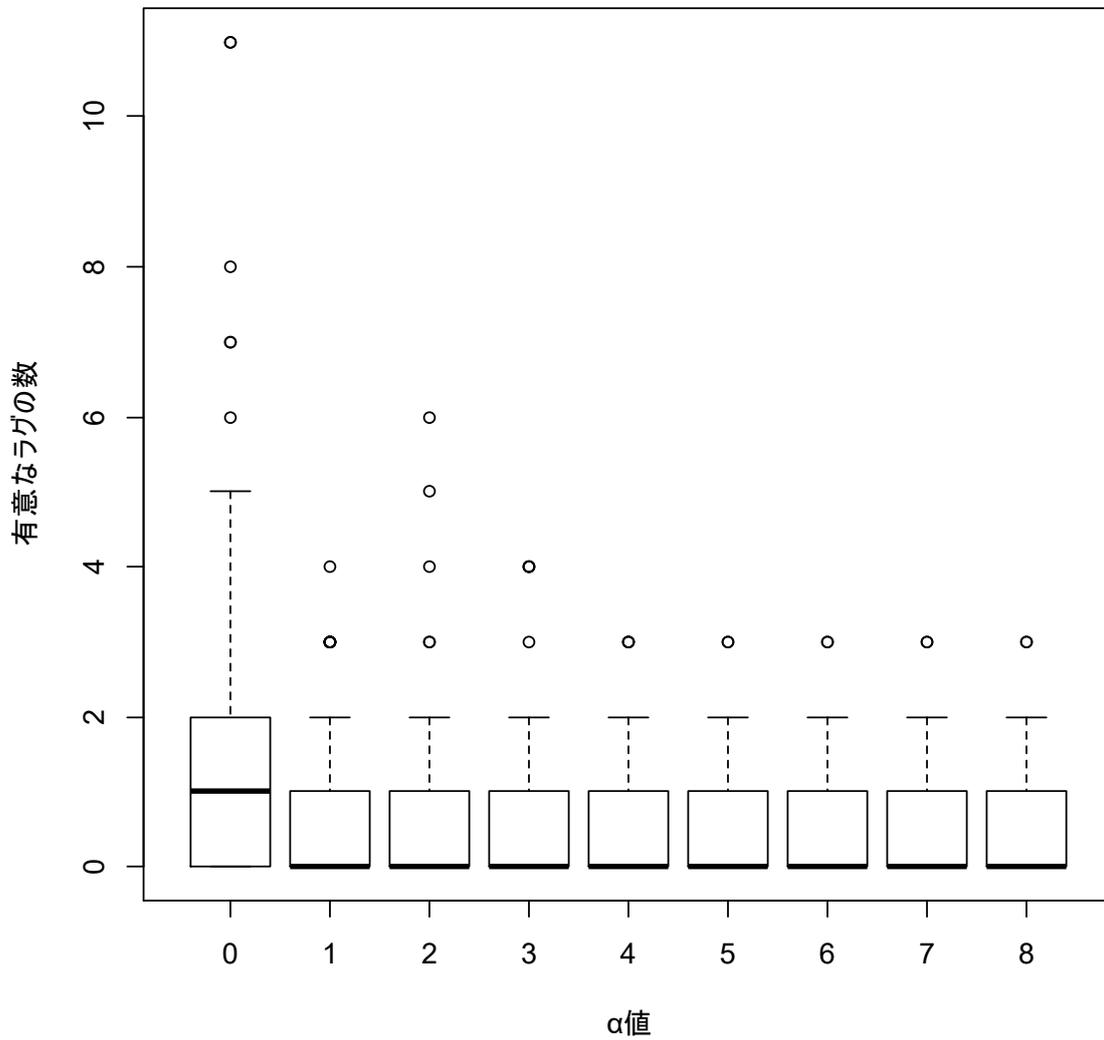


図 14 ZINB モデルにおける  $\alpha$  値とモデルの残差の ACF の有意なラグの数の関係

X 軸は  $\alpha$  値を示し、Y 軸は、各  $\alpha$  値の条件の下で推定されたモデルの残差の ACF における有意なラグの数の分布を示す。

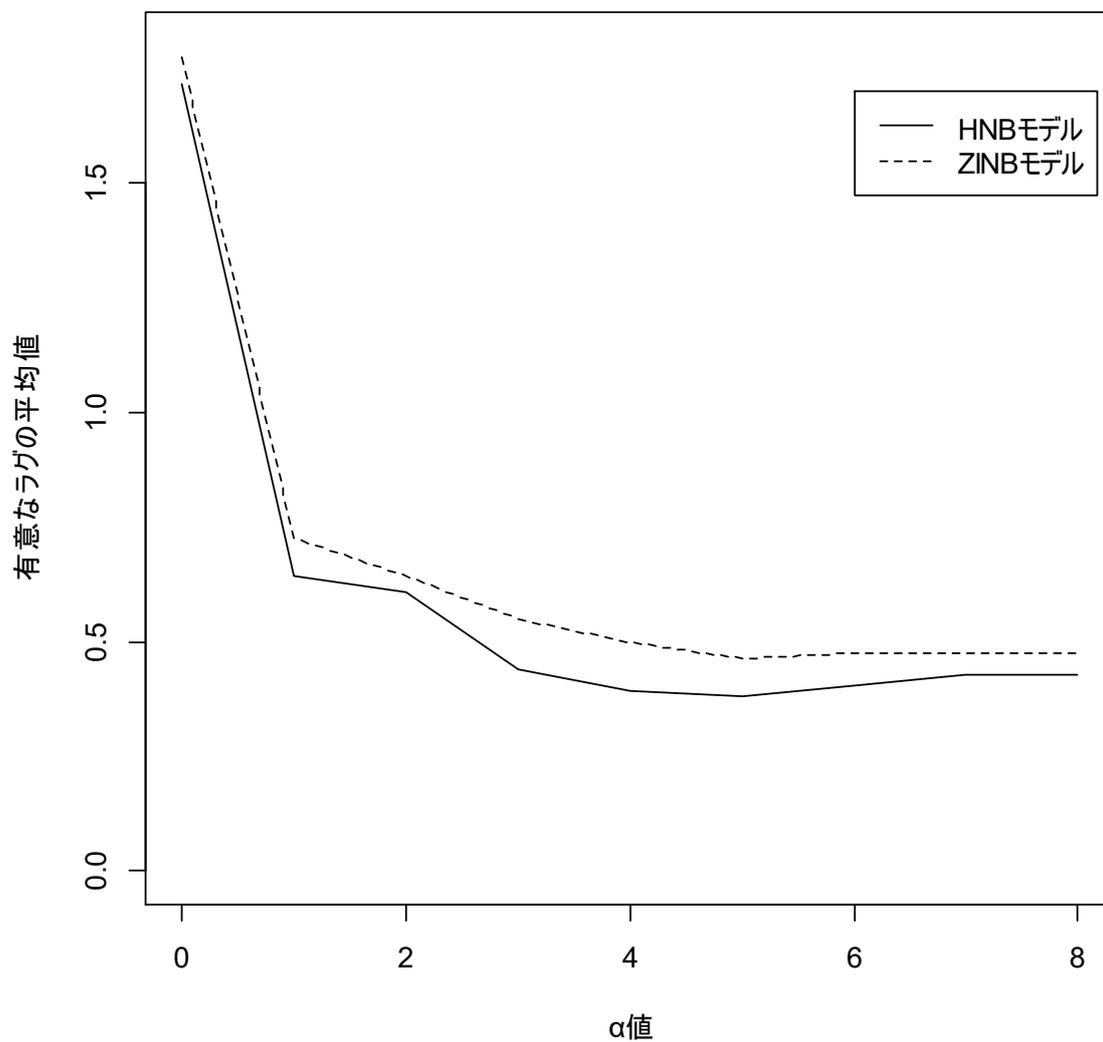


図 15.  $\alpha$  値とモデルの残差の ACF における有意なラグの平均値の関係  
 X 軸は  $\alpha$  値を示し、Y 軸はモデルの残差の ACF における有意なラグの平均値を示す。

ZINB モデルと HNB モデルを比較するための nested Vuong test の結果を表 11 に示す。55 生産者で両方のモデルの当てはまりに有意な差は認められなかった。13 生産者のデータに対して ZINB モデルへの当てはまりがよいのに対し、5 生産者のデータに対して HNB モデルへの当てはまりがよかった。しかし、11 生産者のデータについて、HNB モデルの係数の標準誤差の推定が完了できなかった。係数の値の推定の計算は完了できたため、以下で実施するパーセンタイル値の計算を問題なく行うことができた。

表 11. ゼロ過剰負の二項(ZINB)モデルとハードル負の二項(HNB)モデルの当てはまりの良さを nested Vuong test で比較した結果

肝臓廃棄数の農場順位	ZINB モデルにおける対数尤度	HNB モデルにおける対数尤度	尤度比	p 値
1	-266.00	-265.90	0.30	0.400
2	-305.30	-303.00	4.50	0.152
3	-256.13	-254.11	4.05*	0.042
4	-303.10	-308.53	10.86*	0.023
5	-195.40	-198.40	5.94*	0.010
6	-135.00	-132.40	5.20	0.994
7	-143.10	-142.90	0.34	0.637
8	-114.70	-116.40	3.49*	0.023
9	-172.40	-174.10	3.46*	0.041
10	-164.20	-163.80	0.88	0.962
11	-177.00	-175.20	3.47	0.996
12	-135.10	-132.80	4.50	0.993
13	-127.60	-123.70	7.74	0.313
14	-93.39	-94.53	2.28*	0.010
15	-111.50	-112.10	1.27	0.097
16	-123.40	-122.40	1.87	0.776
17	-96.24	-94.45	3.59	0.079
18	-112.00	-113.70	3.48*	0.029
19	-118.20	-119.80	3.13*	0.012
20	-33.22	-31.53	3.39	0.990
21	-103.70	-103.40	0.57	0.998
22	-85.64	-85.50	0.26	0.700
23	-51.95	-51.86	0.18	0.178

表 11. (続き)

肝臓廃棄数の農 場順位	ZINB モデルにお ける対数尤度	HNB モデルにお ける対数尤度	尤度比	p 値
24	-79.93	-85.09	10.31***	<0.001
25	-81.46	-81.40	0.11	0.456
26	-71.12	-70.56	1.12	0.997
27	-62.15	-65.28	6.25**	0.005
28	-79.26	-79.36	0.21	0.461
29	-41.19	-40.66	1.05	0.996
30	-51.93	-61.95	5.65***	<0.001
31	-63.57	-65.00	2.87*	0.026
32	-23.12	-25.91	5.58*	0.046
33	-27.39	-29.49	4.19**	0.002
34	-52.67	-50.01	5.27	0.995
35	-63.44	-63.02	0.84	0.999
36	-43.18	-43.19	0.02	0.387
37	-50.84	-50.31	Error	Error
38	-30.94	-30.56	0.75	0.996
39	-15.81	-14.97	1.70	0.271
40	-50.39	-50.39	0.00	0.235
41	-36.78	-36.30	0.96	0.941
42	-29.64	-29.06	1.15	0.998
43	-40.83	-40.52	0.62	0.998
44	-26.38	-24.94	2.88	0.976
45	-31.21	-31.21	0.00	0.956
46	-31.52	-30.82	1.39	0.992
47	-25.36	-25.36	0.00	0.347

表 11. (続き)

肝臓廃棄数の農 場順位	ZINBモデルにお ける対数尤度	HNBモデルにお ける対数尤度	尤度比	p 値
48	-33.90	-33.87	0.06	0.410
49	-32.67	-32.58	0.18	0.999
50	-19.49	-19.45	0.08	1.000
51	-13.82	-13.78	0.07	1.000
52	-19.72	-19.17	1.10	0.995
53	-28.55	-28.38	0.33**	0.008
54	-28.55	-28.38	Error	Error
55	-11.36	-18.22	9.78	0.052
56	-24.52	-24.52	0.00	0.126
57	-12.95	-12.95	0.00*	0.030
58	-22.10	-21.90	0.40	0.279
59	-21.00	-20.84	Error	Error
60	-19.13	-18.82	0.62	0.306
61	-14.75	-13.36	2.78***	<0.0001
62	-16.63	-16.04	1.18	0.143
63	-16.98	-16.81	0.34	0.262
64	-8.22	-7.59	1.25	0.117
65	-9.01	-8.38	Error	Error
66	-16.02	-16.02	0.00	0.085
67	-11.96	-11.96	0.00	0.500
68	-13.52	-13.47	0.09	0.082
69	-13.96	-13.93	Error	Error
70	-7.30	-6.38	1.83***	<0.0001
71	-9.36	-9.17	Error	Error

表 11. (続き)

肝臓廃棄数の農 場順位	ZINB モデルにお ける対数尤度	HNB モデルにお ける対数尤度	尤度比	<i>p</i> 値
72	-9.91	-9.75	Error	Error
73	-10.14	-10.12	0.03	0.156
74	-8.66	-8.63	0.07	0.155
75	-7.61	-7.55	0.12	0.153
76	-6.02	-6.02	0.01	0.316
77	-5.75	-5.74	0.01	0.316
78	-5.64	-5.64	0.01	0.315
79	-5.76	-5.76	Error	Error
80	-3.57	-3.53	0.08	0.311
81	-4.26	-4.24	Error	Error
82	-4.91	-4.90	0.02	0.316
83	-5.04	-5.03	Error	Error
84	-4.89	-4.88	Error	Error

\*:  $p < 0.05$  \*\*:  $p < 0.01$  \*\*\*:  $p < 0.001$

Error: HNB モデルの推定が不完全であるため nested Vuong test の計算が完了せず。

時間依存的な変数を持つモデルと切片のみを持つ時間非依存的な帰無モデルを non-nested Vuong test で比較した結果を表 12 に示す。

ZINB モデルと HNB モデルは 32 生産者のモデルで時間依存的な変数がなく、non-nested Vuong test をすることができなかった。残りの 52 生産者のデータで作成したモデルを用いた検討では、20 生産者のモデルで帰無モデルよりも有意に当てはまりが良好であった ( $p < 0.1$ )。上記生産者以外の 32 生産者のモデルでは、帰無モデルよりも尤度の値が良好であった。

表 12. 時間依存的な変数を持つモデル(時系列モデル)と切片のみを持つ帰無モデルの当てはまりの良さを比較するための non-nested Vuong test の結果

肝臓 廃棄 数の 農場 順位	帰無モデルの 対数尤度 (ZINB)	時系列モデル の対数尤度 (ZINB)	AIC 補正の Vuong test の p 値 (null:ZINB)	帰無モデルの 対数尤度 (HNB)	時系列モデル の対数尤度 (HNB)	AIC 補正の Vuong test の p 値 (null:HNB)
1	-297.3	-266.0***	<0.001	-297.3	-265.9***	<0.001
2	-357.5	-305.3***	<0.001	-357.2	-303.0***	<0.001
3	-280.4	-256.1**	0.0028	-280.2	-254.1***	<0.001
4	-334.7	-303.1***	<0.001	-334.1	-308.5**	0.002
5	-222.8	-195.4**	0.0019	-222.2	-198.4**	0.004
6	-162.1	-135**	0.004	-161.4	-132.4***	<0.001
7	-155.6	-143.1*	0.015	-155.6	-142.9*	0.014
8	-127.0	-114.7*	0.030	-126.4	-116.4+	0.067
9	-192.9	-172.4+	0.060	-191.6	-174.1+	0.087
10	-172.8	-164.2	0.21	-170.8	-163.8	0.24
11	-178.9	-177	0.49	-177.5	-175.2	0.45
12	-148.9	-135.1	0.23	-146.8	-132.8	0.25
13	-131.4	-127.6	0.26	-131.3	-123.7*	0.041
14	-110.4	-93.39+	0.06	-110.0	-94.53+	0.066
15	-114.6	-115.5	0.36	-114.4	-112.1	0.46
16	-133.5	-123.4*	0.025	-133.5	-122.4*	0.012
17	-116.4	-96.24*	0.028	-116.4	-94.45*	0.016
18	-131.8	-112.0	0.12	-131.6	-113.7	0.16
19	-126.7	-118.2	0.14	-126.7	-119.8	0.22

表 12. (続き)

肝臓 廃棄 数の 農場 順位	帰無モデルの 対数尤度 (ZINB)	時系列モデル の対数尤度 (ZINB)	AIC 補正の Vuong test の p 値 (null:ZINB)	帰無モデルの 対数尤度 (HNB)	時系列モデル の対数尤度 (HNB)	AIC 補正の Vuong test の p 値 (null:HNB)
20	-38.68	-33.22+	0.090	-38.50	-31.53*	0.030
21	-108.8	-103.7	0.19	-108.1	-103.4	0.21
22	-102.5	-85.64	0.11	-102.5	-85.50	0.11
23	-56.52	-51.95	0.19	-56.52	-51.86	0.18
24	-96.65	-79.93*	0.016	-96.64	-85.09+	0.065
25	-92.09	-81.46+	0.099	-92.09	-81.40+	0.099
26	-81.47	-71.12+	0.075	-81.23	-70.56+	0.067
27	-73.75	-62.15	0.30	-73.46	-65.28	0.49
28	-80.72	-79.26	0.39	-80.72	-79.36	0.37
29	-43.17	-41.19	0.50	-43.10	-40.66	0.42
30	-67.72	-59.13	0.15	-67.70	-61.95	0.31
31	-75.82	-63.57+	0.073	-75.75	-65.00	0.102
32	-40.54	-23.12*	0.020	-40.54	-25.91*	0.038
33	-33.14	-27.39	0.11	-33.14	-29.49	0.23
34	-62.18	-52.65+	0.099	-61.97	-50.01+	0.066
35	-63.44	-63.44	N.T.	-63.02	-63.02	N.T.
36	-46.99	-43.18	0.33	-46.86	-43.19	0.33
37	-53.31	-50.84	0.40	-53.14	-50.31	0.32
38	-33.77	-30.94	0.37	-33.65	-30.56	0.31
39	-19.96	-15.81	0.49	-19.96	-14.97	0.39
40	-50.39	-50.39	N.T.	-50.39	-50.39	N.T.

表 12. (続き)

肝臓 廃棄 数の 農場 順位	帰無モデルの 対数尤度 (ZINB)	時系列モデル の対数尤度 (ZINB)	AIC 補正の Vuong test の p 値 (null:ZINB)	帰無モデルの 対数尤度 (HNB)	時系列モデル の対数尤度 (HNB)	AIC 補正の Vuong test の p 値 (null:HNB)
41	-42.07	-36.78	0.19	-42.06	-36.30	0.18
42	-29.64	-29.64	N.T.	-29.06	-29.06	N.T.
43	-40.83	-40.83	N.T.	-40.52	-40.52	N.T.
44	-35.77	-26.38	0.13	-35.76	-24.94	0.101
45	-31.21	-31.21	N.T.	-31.21	-31.21	N.T.
46	-33.47	-31.52	0.49	-33.47	-30.82	0.41
47	-25.36	-25.36	N.T.	-25.36	-25.36	N.T.
48	-34.83	-33.90	0.27	-34.83	-33.87	0.28
49	-32.67	-32.67	N.T.	-32.58	-32.58	N.T.
50	-19.49	-19.49	N.T.	-19.45	-19.45	N.T.
51	-13.82	-13.82	N.T.	-13.78	-13.78	N.T.
52	-22.76	-19.72	0.34	-22.75	-19.17	0.30
53	-28.55	-28.55	N.T.	-28.38	-28.38	N.T.
54	-28.55	-28.55	N.T.	-28.38	-28.38	N.T.
55	-18.66	-3.633+	0.099	-22.99	-18.22	0.40
56	-24.52	-24.52	N.T.	-24.52	-24.52	N.T.
57	-12.95	-12.95	N.T.	-12.95	-12.95	N.T.
58	-25.33	-22.10	0.31	-25.33	-21.90	0.32
59	-22.04	-21.00	0.29	-21.96	-20.84	0.32
60	-21.66	-19.13	0.29	-21.57	-18.82	0.34
61	-19.01	-14.75	0.24	-19.01	-13.36	0.20

表 12. (続き)

肝臓 廃棄 数の 農場 順位	帰無モデルの 対数尤度 (ZINB)	時系列モデル の対数尤度 (ZINB)	AIC 補正の Vuong test の p 値 (null:ZINB)	帰無モデルの 対数尤度 (HNB)	時系列モデル の対数尤度 (HNB)	AIC 補正の Vuong test の p 値 (null:HNB)
62	-21.29	-16.63	0.42	-21.19	-16.04	0.39
63	-18.42	-16.98	0.39	-18.36	-16.81	0.42
64	-11.44	-8.215	0.32	-11.35	-7.590	0.29
65	-9.011	-9.011	N.T.	-8.380	-8.380	N.T.
66	-16.02	-16.02	N.T.	-16.02	-16.02	N.T.
67	-11.96	-11.96	N.T.	-11.96	-11.96	N.T.
68	-13.52	-13.52	N.T.	-13.47	-13.47	N.T.
69	-13.96	-13.96	N.T.	-13.93	-13.93	N.T.
70	-10.04	-7.296	0.39	-10.04	-6.382	0.32
71	-9.361	-9.361	N.T.	-9.173	-9.173	N.T.
72	-9.908	-9.908	N.T.	-9.752	-9.752	N.T.
73	-10.14	-10.14	N.T.	-10.12	-10.12	N.T.
74	-8.664	-8.664	N.T.	-8.628	-8.628	N.T.
75	-7.607	-7.607	N.T.	-7.545	-7.545	N.T.
76	-6.024	-6.024	N.T.	-6.021	-6.021	N.T.
77	-5.745	-5.745	N.T.	-5.741	-5.741	N.T.
78	-5.644	-5.644	N.T.	-5.640	-5.640	N.T.
79	-5.762	-5.762	N.T.	-5.758	-5.758	N.T.
80	-3.565	-3.565	N.T.	-3.525	-3.525	N.T.
81	-4.258	-4.258	N.T.	-4.239	-4.239	N.T.
82	-4.912	-4.912	N.T.	-4.902	-4.902	N.T.

表 12. (続き)

肝臓 廃棄 数の 農場 順位	帰無モデルの 対数尤度 (ZINB)	時系列モデル の対数尤度 (ZINB)	AIC 補正の Vuong test の p 値 (null:ZINB)	帰無モデルの 対数尤度 (HNB)	時系列モデル の対数尤度 (HNB)	AIC 補正の Vuong test の p 値 (null:HNB)
83	-5.043	-5.043	N.T.	-5.034	-5.034	N.T.
84	-4.892	-4.892	N.T.	-4.882	-4.882	N.T.

+ :  $p < 0.1$  \* :  $p < 0.05$  \*\* :  $p < 0.01$  \*\*\* :  $p < 0.001$

NT: モデルが時間依存的な変数を持たないために検定できなかった。

## 時系列モデルと従来の基準との比較

時系列モデルと従来基準との早期検出性の比較は、ZINBとHNB回帰モデルの異なる条件で行うとより詳細に特徴が把握できる。このため、両モデルについて( $\alpha = 5$ )における分布の50パーセンタイル値から90パーセンタイル値まで5パーセンタイル刻みで期待値を計算した。実際の肝臓廃棄数がこれらパーセンタイル値を超えた回数と、時系列モデルと従来基準との比較結果を表13に示す。検出の迅速性という観点では、時系列モデル基準(ZINBおよびHNBの50-75パーセンタイル値)の方が従来基準よりも有意に早く検出することができた( $p < 0.01$ , 表13)。しかし、80および85パーセンタイル値では、両基準の差は有意ではなくなり、90パーセンタイル値では、従来基準のほうが有意に早かった。毎月の豚の平均搬入数が100未満の小規模生産者では、従来基準は65パーセンタイル値およびそれ以上において有意に早く検出した(表14)。毎月の豚の平均搬入数が100頭以上の大規模生産者では、90パーセンタイル値以下の全ての場合において時系列モデル基準が有意に早く検出した(表15)。

表 13. 肝臓廃棄数増加の検出の迅速性に関する異なるパーセンタイル値選択時の従来の基準と時系列モデルの比較(全 84 生産者)

モデル	パーセン タイル	従来の基 準検出数	従来の基準 が早く検出 した回数	時系列モ デル検出 数	時系列モデ ルが早く検 出した回数	<i>p</i> 値
HNB	50	1,108	20	1,229	303	<0.001
	55	1,108	26	1,162	298	<0.001
	60	1,108	34	1,094	291	<0.001
	65	1,108	46	976	238	<0.001
	70	1,108	52	898	214	<0.001
	75	1,108	75	818	192	<0.001
	80	1,108	124	688	154	0.08
	85	1,108	173	566	136	0.04
	90	1,108	252	381	83	<0.001
ZINB	50	1,108	21	1,174	293	<0.001
	55	1,108	25	1,111	282	<0.001
	60	1,108	32	1,048	275	<0.001
	65	1,108	42	932	226	<0.001
	70	1,108	62	858	203	<0.001
	75	1,108	86	781	183	<0.001
	80	1,108	132	648	146	0.43
	85	1,108	171	536	120	0.003
	90	1,108	243	384	78	<0.001

HNB: ハードル負の二項モデル; ZINB: ゼロ過剰負の二項モデル

表 14.肝臓廃棄数増加の検出の迅速性に関する異なるパーセンタイル値選択時の従来の基準と時系列モデルの比較(月平均搬入数 100 頭未満)

モデル	パーセンタイル	従来の基準検出数	従来の基準が早く検出した回数	時系列モデル検出数	時系列モデルが早く検出した回数	p 値
HNB	50	640	19	543	17	0.87
	55	640	24	522	17	0.35
	60	640	30	498	17	0.08
	65	640	38	468	15	0.002
	70	640	41	448	15	<0.001
	75	640	58	408	14	<0.001
	80	640	88	356	9	<0.001
	85	640	126	287	7	<0.001
	90	640	188	200	6	<0.001
ZINB	50	640	20	527	16	0.62
	55	640	23	506	16	0.34
	60	640	28	481	16	0.1
	65	640	34	457	16	0.02
	70	640	49	426	14	<0.001
	75	640	61	391	12	<0.001
	80	640	100	326	6	<0.001
	85	640	114	287	4	<0.001
	90	640	172	211	4	<0.001

HNB:ハードル負の二項モデル; ZINB: ゼロ過剰負の二項モデル

表 15.肝臓廃棄数増加の検出の迅速性に関する異なるパーセンタイル値選択時の従来の基準と時系列モデルの比較(月平均搬入数 100 頭以上)

モデル	パーセンタイル	従来の基準検出数	従来の基準が早く検出した回数	時系列モデル検出数	時系列モデルが早く検出した回数	p 値
HNB	50	468	1	686	286	<0.001
	55	468	2	640	281	<0.001
	60	468	4	596	274	<0.001
	65	468	8	508	223	<0.001
	70	468	11	450	199	<0.001
	75	468	17	410	178	<0.001
	80	468	36	332	145	<0.001
	85	468	47	279	129	<0.001
	90	468	64	181	77	0.31
ZINB	50	468	1	647	277	<0.001
	55	468	2	605	266	<0.001
	60	468	4	567	259	<0.001
	65	468	8	475	210	<0.001
	70	468	13	432	189	<0.001
	75	468	25	390	171	<0.001
	80	468	32	322	140	<0.001
	85	468	57	249	116	<0.001
	90	468	71	173	74	0.87

HNB:ハードル負の二項モデル; ZINB: ゼロ過剰負の二項モデル

前述の方法で数値を計算し、肝臓廃棄数、パーセンタイル値を廃棄数が超えた月および行政による情報提供が行われた月を、3生産者(以下、生産者 A,B および C と記載する。これら生産者 A,B および C は、第 1 章で述べた生産者 A,B および C とは異なる生産者である)を例として図 16~18 に表示する。図 16~18 に表示されているパーセンタイル値は、ZINB モデルを用いて計算されたものである。従来の基準よりも有意に早く検出することができる 60 パーセンタイル値に加え、10 パーセント間隔でさらに 2 つのパーセンタイル値を表示した。HNB モデルを用いて計算されたパーセンタイル値の推移は ZINB モデルのものと同様である。生産者 A(平均月搬入豚数 = 440) は、累積肝臓廃棄数が最も多い生産者である( $n = 1,185$ )。生産者 B は、月あたり 100 頭未満の豚を搬入しており(平均月搬入豚数 = 54)、累積肝臓廃棄数が 4 番目に多い生産者である( $n = 482$ )。生産者 C は、平均月搬入豚数が最も多い生産者であり(平均月搬入豚数 = 3,185)、累積肝臓廃棄数が 12 番目に多い生産者である。

図 17b に示されたように、生産者 B の場合、1 頭以上が豚エキノコックス症と診断された全ての月に行政からの情報提供を受けている。このことは、1 個でも肝臓が廃棄されると廃棄率が 1%を超えるために起こる。それに対し、生産者 C は、図 18a および図 18b に示されたように、肝臓廃棄数が大きく増加しない限り肝臓廃棄率が 1%を超えなかった。したがって、生産者 C の場合には、3 カ月以上連続して豚エキノコックス症による廃棄があった場合に限り行政からの情報提供が行われていた。

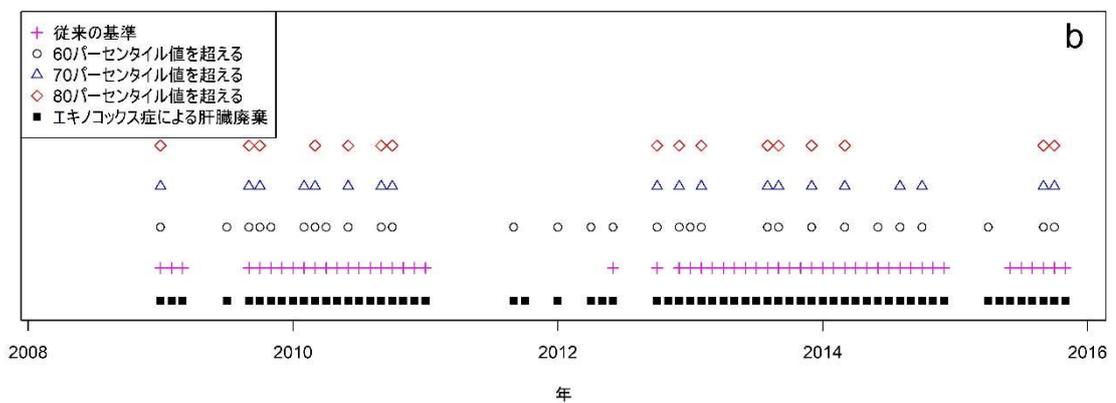
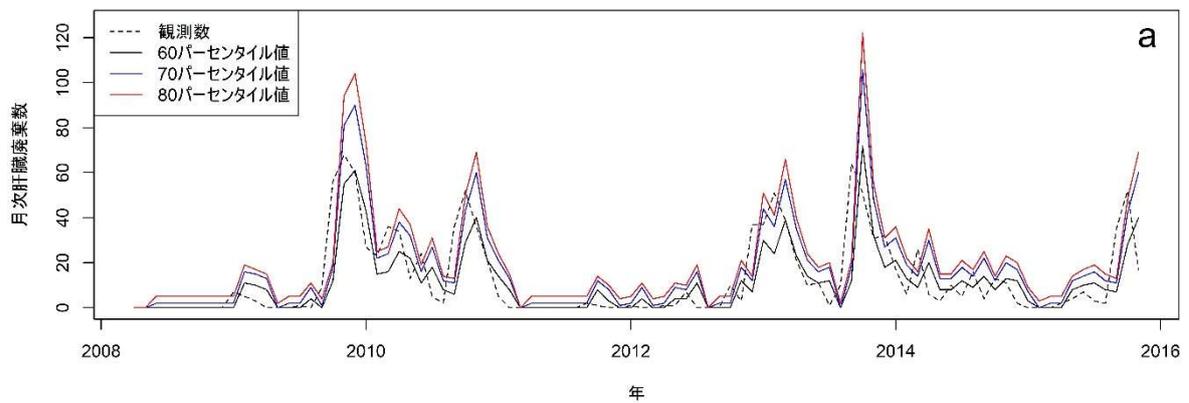


図 16. 生産者 A の豚エキノкокクス症による肝臓廃棄の月次観測数の推移

(a) ZINB 回帰モデルの 60、70 および 80 パーセンタイル値

(b) ZINB 回帰モデルの各パーセンタイル値および従来の基準を超えることによってアウトブレイクが検出された月ならびに豚エキノкокクス症による肝臓廃棄が 1 個以上観測された月

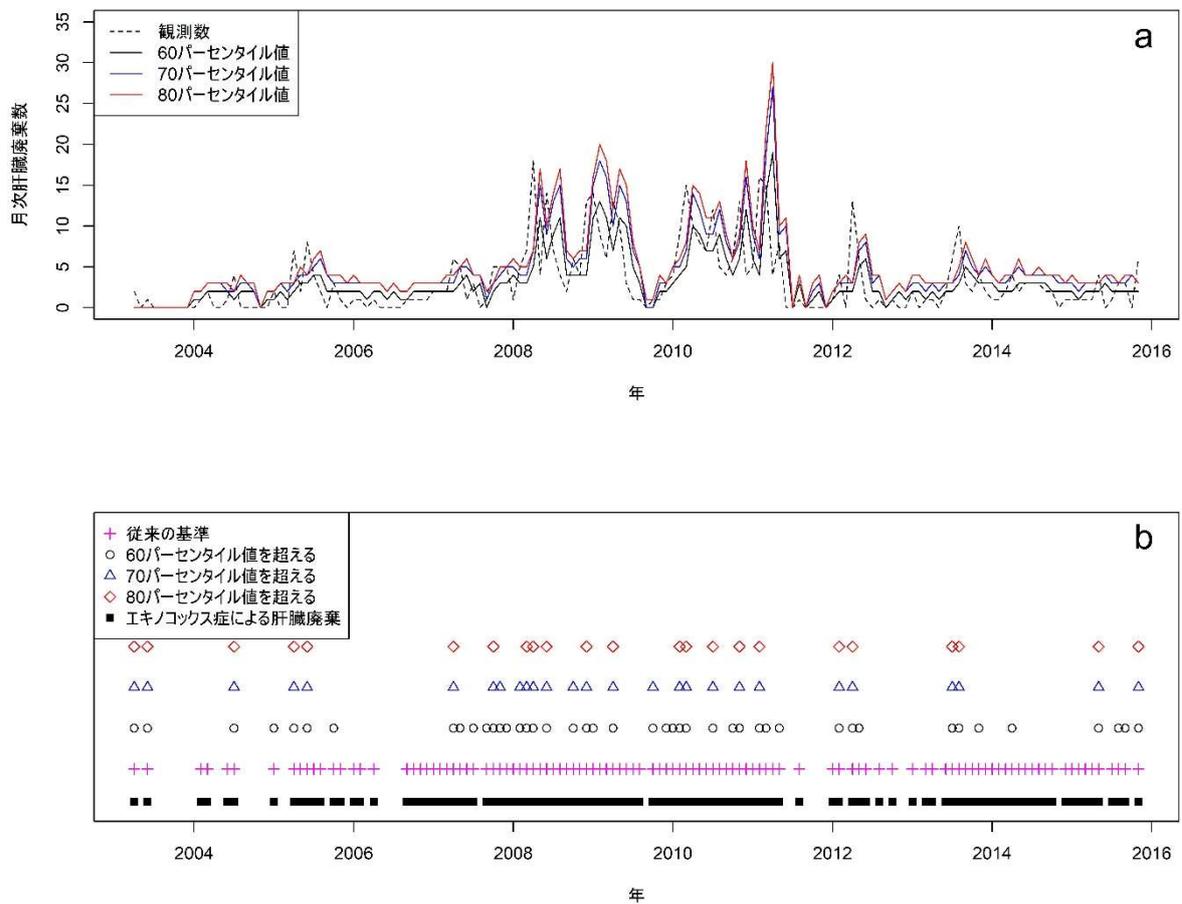


図 17. 生産者 B の豚エキノコックス症による肝臓廃棄の月次観測数の推移

(a)ZINB 回帰モデルの 60、70 および 80 パーセンタイル値

(b)INB 回帰モデルの各パーセンタイル値および従来の基準を超えることによってアウトブレイクが検出された月ならびに豚エキノコックス症による肝臓廃棄が 1 個以上観測された月

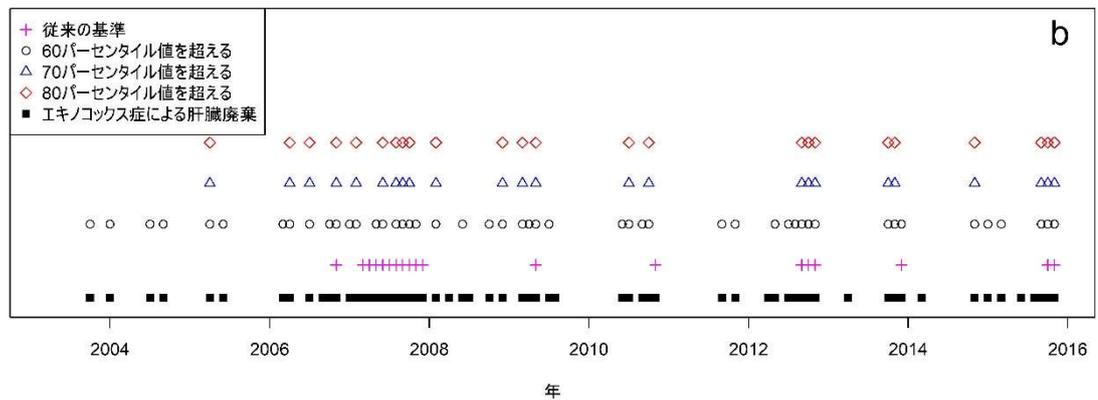
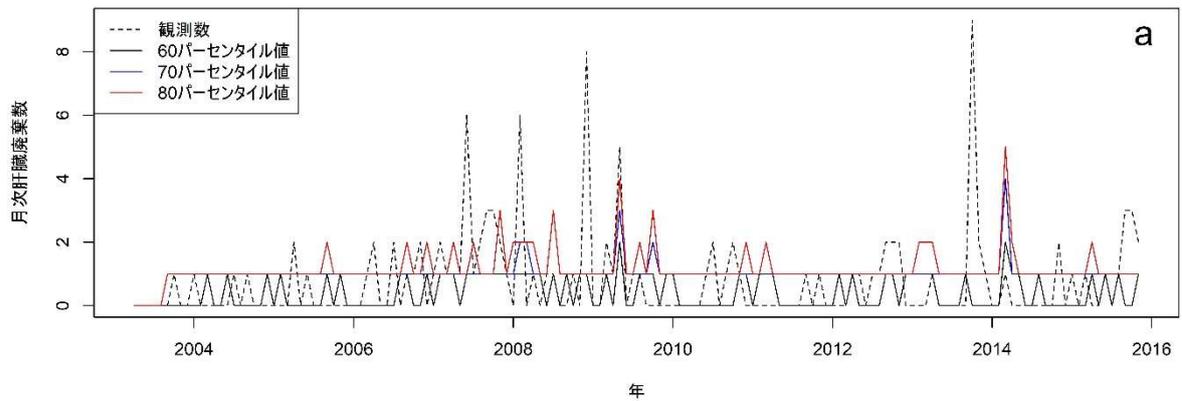


図 18. 生産者 C の豚エキノコックス症による肝臓廃棄の月次観測数の推移

(a) ZINB 回帰モデルの 60、70 および 80 パーセンタイル値

(b) ZINB 回帰モデルの各パーセンタイル値および従来の基準を超えることによってアウトブレイクが検出された月ならびに豚エキノコックス症による肝臓廃棄が 1 個以上観測された月

#### 4. 考 察

本研究の目的は、豚エキノコックス症の農場アウトブレイクを最も早く検出できる統計学的な基準を確立し、適切な時期に農場が疾病の制御手段を講じることができるよう注意を促すことができるようにすることである。本章の研究に先立ち、I章ならびにII章に述べたように二つの方法でと畜検査データの時系列分析に関する研究を進めてきた。一方は、と畜場の臓器廃棄理由として四種類のよくみられる疾病のデータを用い、臓器の廃棄率をロジット変換した後に SARIMA モデルを用いて解析したものであり[2]、もう一方は、豚の抗酸菌症のデータを用い、周期的変動を擬似ポアソンモデルで回帰させた残差を ARIMA モデルで解析したものである[3]。

豚エキノコックス症のように多くのゼロデータと少数の非ゼロデータからなるカウントデータを時系列分析する場合に、二つの問題が生じる。一つ目は、変換したデータを用いてガウス分布に基づくモデリングを行うと、正確ではない予測分布となる可能性があるという問題である。二つ目は、多数のゼロを含むデータを対数変換することは、非ガウス分布を用いたとしても多くの問題を生じる可能性があることである。スリランカ保健省のマラリア対策活動理事会は、SARIMAモデルを用いたマラリア予測システムの試験を行っており、このモデルでは、マラリア月次症例数のカウントデータを対数変換したデータは近似的にガウス分布すると仮定している[13]。Briëtら[12]は、BenjaminとRigby[8]によって考案された一般化自己回帰移動平均モデル(GARMAモデル)を用いることによって近似的にガウス分布を用いる方法より良好な結果を得ている。GARMAモデルは、ポアソン分布や負の二項分布するデータを取り扱うことができるが、説明変数として

用いられるラグ処理された観測値がリンク関数を介して対数変換されることから、観測値がゼロを含むと処理不能となってしまう。ZegerとQaqish[74]は、この問題を解決する一つの方法として、全てのデータもしくはゼロのみに小さな定数を加えることを検討している。Briëtら[12]は、ZegerとQaqish[74]の方法に従って全てのゼロに定数を加えて検討をしたところ、用いたデータにゼロが3つしか含まれていなかったために、GARMAモデルは良好に機能することを見出した。Huangら[26]は、対数変換した過去の Deng 熱症例数を自己回帰項として用い、時系列 HNB モデルによって分析することによって Deng 熱の輸入症例数の時系列分析を行っている。しかし、同報告にはゼロ値をどのように扱ったか記載されておらず、おそらくゼロ値に定数を加えたものと推測される。ラグ処理されたカウントデータの観測値は、上述のとおり、非ガウス分布に基づく時系列分析では一般的にはリンク関数として対数が用いられるためゼロ値の扱いが問題となる。実際に、ゼロを多く含む豚の非定型抗酸菌症のデータを対数変換し、GARMAもしくはtwo-partモデルを利用して解析することを試みた。しかし、残差が自己相関しないモデルを得ることはできなかった。そこで、データを変換することなしに、多数のゼロを扱うことができ、残差が自己相関しないモデルが必要である。

この目的を達成するために、本章では、期待値を計算するためのオートエクスポネンシャル回帰式とゼロ過剰パラメータを計算するためのオートロジスティック回帰式を用いることで、過去のデータの変換を行わないtwo-partモデルを作成した。得られたモデルは、変換なしに多くのゼロを取り扱うことができるので予測分布が不正確にならない。Besag[9]は、マルコフ確率場を解析するために、オートエクスポ

ネンシャル(オートポアソン)モデルとオートロジスティックモデルからなるオートモデルを導入している。このオートモデルは、時間的データの解析にも良好に利用できる[78]。隣接する領域でゼロとなる場所が持続するカウントデータを処理するには、オートポアソンモデルのみならず、元来ゼロ部分の概念が含まれている負の二項分布を応用したオート負の二項モデルも用いることができる[7]。本章の研究では、オートエクスポネンシャルとオートロジスティックモデルがゼロ過剰であると畜検査における臓器廃棄データの時系列分析に利用可能であることを明らかにした。

本章の研究では、時系列モデルを作成する最初の段階として6種類の離散分布(すなわち、ポアソン、負の二項、ZIP、ZINB、HPおよびHNB分布)を比較している。あてはまりの良さという点で、HNBモデルが最も優れており、次にZINBモデルであった。これらのtwo-partモデルの中から一つを選択するとき、ゼロデータが一つの構造的な発生源に由来し、その部分が二項分布している場合にはHNBモデルを選択するべきで、ゼロデータが構造的な発生源とサンプリングによる発生源の2つに由来する場合にはZINBモデルを選択するべきである[25, 52]。本章の研究におけるゼロデータの発生原因は、農場が*E. multilocularis*の虫卵に汚染されていないのか、あるいは虫卵汚染があるにも関わらずと畜検査で病変が見られないという2つがあると解釈でき、この解釈が妥当であるならば、ZINBモデルが最適と考えられる。しかし本研究において、農場の汚染状況と肝臓廃棄の頻度の関連性が明らかではなく、前述の解釈を支持する根拠は十分とは言えない(これらの関連性は、本研究の目的ではなく、今後の課題とされた)。Arab[6]は、ゼロデータ発生機序が不明であり、発生源が1つ(構造的

発生源)だけなのか、あるいは発生源が2つ(構造的発生源とサンプリング発生源)あるのか明らかではない場合、モデル選択のための手順が導入されるべきであると報告している。そこで、本研究では、HNBモデルとZINBモデルの両方の分布を用いてモデリングを行い、Vuong testを用いて両者を比較した。その結果、あてはまりの良さの点でZINBモデルの方がHNBよりも若干優れていた。したがって、アウトブレイクの検出を今後実際に行っていく場合には、ZINBモデルを用いるのがよいと考えられる。

Huら[25]は、HIVの感染リスクを増大させる行動を抑制するための介入に関する臨床試験データを解析するために、ZINBモデルとHNBモデルを比較し、ZINBモデルが良好であったことを報告している。この報告では、データに含まれる多数のゼロは、セクシャルパートナーが存在しないために発生する構造的なゼロと、セクシャルパートナーが存在するが無防備な性行為(高リスクセックス)を行わないことによって発生するサンプリング上のゼロの二つの発生源があると述べている。Cairnsら[15]も、マラリアのコホート研究においてマラリア発症歴のない小児の集団由来のゼロの多いデータへのモデルのあてはめを行うために、4種類の離散分布を比較している。Cairnsら[15]は、ZINB分布が最も良好と結論しているが、ハードルモデルについては検討していない。と畜検査データにおける稀な臓器廃棄理由について時系列分析を行うときは、上述の研究と同様に、ZINBモデルが一般的に有効であろう。

本章の研究で検討した時系列モデルは、自己相関のあるラグ処理データを説明変数として用いている。さらに、モデル選択は、残差の自己相関に基づいている。時系列モデル(たとえばARIMAモデル)のパ

ラメータ推定では、あてはまりの良さを基準として自己回帰項と移動平均項の種々の考えられる次数の組み合わせについて包括的に係数が推定され、モデルの残差の ACF を調べることによりモデルの有効性が確認される。仮にこの手順を本章の研究に用いた場合、12 カ月もしくはそれ以下のラグで処理された過去の肝臓廃棄数の説明変数として可能な組み合わせは 4,095 通りあり、343,980 回の最尤推定が必要となる(84 生産者の各々について 4,095 回の推定)。モデルに用いるラグの数を増加させるに従い推定しなければならないモデルの数が増加するため、モデルに用いるラグの数には上限が必要である。しかし、季節変動(1 年周期の変動)を表すために 12 カ月のラグで処理されたデータが必要である。したがって、本章の研究では、ラグの上限を 12 カ月とし、ルーチンの食肉検査に用いるための実用性を考慮し、前述の手順を用いることとした。時系列モデルが適切であれば、モデルの残差には自己相関がないはずである。したがって、偶然に自己相関を有する 5%未満の項を除けば、残る 95%以上の項は自己相関がないことになる[49]。以上の観点より、本章の研究における  $\alpha$  の選択は有効である。本章で述べた時系列モデルの有効性は、**nested Vuong test** の結果からも支持されており、**nested Vuong test** の結果は、時間依存的な共変数がモデルのあてはまりの良さを改善していることを示している。

アウトブレイクの検出の早さという点では、本研究で設定した時系列分析基準(HNB および ZINB 分布の 50~75 パーセンタイル値)は、従来の基準よりも有意に早く検出することを可能にしている( $p < 0.01$ , 表 13)。しかし、月平均搬入数が 100 頭未満の小規模生産者の場合、60 パーセンタイル値より大きい値を基準とすると時系列分析基準よりも従来基準の方が有意に早く検出する。60 パーセンタイル値もしくはは

それ以下であれば、従来の基準と比較しても検出の早さで劣ることはない。

表 13 および 14 に示したように、時系列分析基準はパーセンタイル値を適切に選択すれば従来の基準よりも有意に早く検出できた。しかし、月間搬入数が 100 頭を下回る小規模生産者の場合、従来の基準は、1 頭でも豚エキノコックス症病変による肝臓廃棄があると必ず検出することとなる(すなわち、エキノコックス症の有無を判定しており、アウトブレイクの危険性を判定していない)。そのような場合、時系列分析基準は従来の基準より早く検出することは決してできない。表 14 に示したように、従来の基準よりも時系列分析基準の方がアウトブレイクを早く検出した例が複数あった。これらの例は、当該小規模生産者がその月に 100 頭以上を搬入していたために豚エキノコックスを検出しなかったことによって生じたものである。すなわち、このような場合には廃棄率が 1%を超えないことから、固定された廃棄率で検出する従来の基準を用いることは適切ではなく、アウトブレイクを見落とすことを意味している。

従来の基準と時系列分析基準の比較の結果、次の 3 点において時系列分析基準が優れていることが明らかとなった。すなわち、1)時系列モデルは、特に大規模農場において農場アウトブレイクを従来法よりも迅速に検出できる、2)少ない見落としで生産者に情報提供できる客観的かつ合理的なシステムである、3)アウトブレイクが持続している生産者に行政からの情報提供を過度に行わないように基準(パーセンタイル値)を選択することができる、という 3 点である。現在用いている基準の下では、3 カ月連続で豚エキノコックスが検出されるか肝臓の廃棄率が 1%を超えなければ生産者は行政からの情報提供を受け取る

ことがない。この状態は、アカギツネの農場への侵入防止措置が遅延するリスクをもたらしている。さらに、豚エキノコックス症の農場アウトブレイクは、同一の生産者によって搬入された豚と体に1カ月もしくはそれ以上の間、豚エキノコックス症病変による肝臓廃棄数の上昇がみられることと定義されるべきである。ここでいう廃棄数の上昇とは、統計学的方法を用いて計算された期待値もしくはパーセンタイル値を実際の廃棄数が超えることである。本章に記載した時系列分析法は、上記の条件の下にアウトブレイクを検出することが可能である。しかし、従来の基準は、廃棄数が固定された閾値を超えた時を判定するのみであり、本章で定義したアウトブレイクを検出することはできない。さらに、従来の基準は各々の農場の状況を考慮していない固定された閾値であるためにアウトブレイクを見落とすリスクが存在する。したがって、本章に記載した時系列分析法を豚エキノコックス症のアウトブレイクを情報提供するためのシステムとして取り入れることは、養豚農場の衛生状況を大きく改善する可能性がある。

上述の時系列モデルの第三の優位性は、生産者とのコミュニケーションの状況を改善する可能性がある。天候についての行政の意思決定に関するある研究は、行政からの偽陽性の情報提供が頻繁になると狼少年効果(cry-wolf effect)または誤報作用(false-alarm effect)と呼ばれる反応を引き起こすことを指摘している[38]。本章の研究では、豚エキノコックス症に関する頻繁な情報提供の全ては誤報ではない。しかし、このような生産者(特に豚エキノコックス症と診断される豚が一頭以上存在する全ての月に情報提供を受ける生産者[図 17])へ頻繁な情報提供を行うことによって、提供された情報や豚エキノコックス症のリスクの無視という行動、すなわち誤報効果に似た行動を引き出してい

ることも懸念される。このような生産者に対しては、情報提供の頻度を下げた基準(パーセンタイル値)を選択することもできる。食肉検査が実施されている状況下では感染した臓器が食用にされる可能性は低く、さらに、ヒトは感染した豚を摂食することによって豚エキノコックス症に罹患することはない。検査所から生産者に対して送付される情報提供用パンフレットでは、次のような対策を講じるよう勧めている。すなわち、i)豚の飼料や後産などの、キツネの餌となりえる廃棄物を片付けること、ii)豚舎に入る際には舎内に虫卵を持ち込まないように長靴を充分洗浄すること、iii)犬の放し飼いを避け、エゾヤチネズミ(*Myodes rufocanus bedfordiae*)を食べさせないようにすること、の3点である。さらに、北海道が定めた「キツネの駆虫に関するガイドライン」に従い、生産者に対して駆虫剤を含有する餌の散布について情報提供することができる。

時系列モデルによる判定結果をと畜検査結果に基づく情報提供文書にわかりやすい文言として取り入れることにより、エキノコックス虫卵汚染に対する適切な対策を後押しすることも考えられる。たとえば、60パーセンタイル値(図15~18の黒線)を超えたときに「増加傾向」、70パーセンタイル値(図15~18の青線)を超えたときに「強い増加傾向」、80パーセンタイル値(図15~18の赤線)を超えたときに「非常に強い増加傾向」などの説明を付加することなどである。さらに、農場における対策を家畜保健衛生所と連携して進めることが出来れば、上記のような情報はより有用なものとなろう。

## 5. 小 括

本章に述べた豚エキノコックス症の農場アウトブレイクを検出する方法は、稀に発生する他の疾病のアウトブレイクを検出する方法としても有用であると考えられる。さらに、本方法によって得られる情報は、家畜衛生部門との連携を介して農場衛生を著しく改善し得ると考えられる。

## 第 IV 章 総 合 考 察

本研究の目的は、データの分布に応じて複数の方法を使い分けることにより、と畜検査データの時系列分析を網羅的に行うことができることを示すことである。

まず第一章では、SARIMA モデルによる分析方法について述べた。常時みられる疾病の廃棄率をロジット変換したデータからモデリングし AIC を基準として選択された SARIMA モデルは、時系列モデルとして妥当性を有し、対照モデルの指数平滑法よりもモデルのあてはまりがよいことがわかった。

また、第二章では、カウントデータの周期的変動要素を擬似ポアソン分布モデルで推定し、その残差について ARIMA モデルで推定した後、両者の和からなるモデルを用いた分析方法について述べた。本モデルが、抗酸菌症の潜在的なアウトブレイクを迅速に検出するのに有用であることを示したが、周期性の解析と ARIMA モデルの推定の二段階でモデリングを行ったものであり、作成したモデルを継続的に用いるのであれば問題はないが、新たなデータを追加しモデルを更新する場合に煩雑となる可能性があった。

さらに、第三章では、two-part モデルによる分析方法について述べた。本モデルは、豚エキノコックス症の農場アウトブレイクを検出するのに有用であり、ゼロを多く含むために過分散となっている時系列データをワンステップでモデル化できることがわかった。

実際にと畜検査データの時系列分析を網羅的に行うために有用な手法の組み合わせについて以下に述べる。廃棄率が極端に 0% や 100% 付近に偏っていなければ、ロジット変換された廃棄率に対して正規分布

を仮定できるため、このような廃棄率を示す疾病に対して第一章に述べた SARIMA モデルを用いた分析を行うのが適切と考えられた。ゼロの多いデータについては、第二章の方法と第三章の方法を用いることができるが、ルーチンの検査業務での応用を考慮すると、ワンステップでモデリングが可能な two-part モデルの方が、モデルの更新がより容易なことから、実用的と目された。すなわち、廃棄数がゼロであることが稀な疾病については、第一章の SARIMA モデルで分析し、ゼロが多い場合は two-part モデルで分析することにより、網羅的に畜検査データの時系列分析を行うことができると考えられた。

本研究の分析方法は、と畜検査員がと畜検査データを分析するためのものであるが、分析結果を有効利用するために、実際に分析結果に基づいて衛生対策を生産者自ら積極的に行うだけでなく、家畜衛生や産業動物臨床に従事する獣医師の監督指導のもとに当該衛生対策が実施されることが望まれる。と畜検査データが有効利用されるためには、種々の分野の獣医師が「one world-one health」の理念に基づいて連携協力することが必要である。たとえば食肉衛生検査所が、生産者のみならず家畜保健衛生所にと畜検査データを提供し、家畜保健衛生所は生産者への衛生指導に当該データを活用すること[1]は、望ましい連携協力の一つである。

本研究は、と畜検査データの時系列分析に関する研究であるが、時間とともに出現するデータ群の分析方法であれば、疫学分野のみならず他の広範な分野のデータ解析に用いることができる。Box と Jenkins[11]が、ARIMA モデルを発表してからすでに 40 年以上経過しているが、獣医学分野において時系列分析法を利用する研究は未だ少なく、ARIMA モデル以外にも数多くの時系列モデルが考案されてい

ることを考慮すると、時系列分析を活用可能な新たな課題も多いと考えられる。

ARIMA モデルが考案された当初、この種の高度な数値計算は、多数のユーザーで共用されている大型コンピュータ上で専門的な言語によって書かれたプログラムを実行する必要であったため、大規模な設備と高額な費用を要する作業であった。しかし今日では、市販のコンピュータと無償で入手できる R[29]があれば事足り、実際にそれだけで本研究で用いた全ての分析法を実施することができる。このような変化は、近年のコンピュータ技術の著しい進歩に加え、それ自体無償で入手でき、新たな処理プログラムが無償のパッケージとして随時追加されるという R の特質によるところが大きい。R のようなフリーウェアの充実とともに、今後、広範な分野の獣医師が安価な費用でデータ分析できる新たな研究課題が増加していくものと思われる。さらに、獣医学分野で有用な処理方法を獣医師自ら R の関数としてプログラミングし、パッケージ化して公開することができれば、獣医学分野におけるデータ処理技術の進歩を積極的に加速することが可能となるであろう。

## 謝 辞

本研究では、酪農学園大学獣医学群獣医学類衛生・環境教育分野獣医疫学ユニット 蒔田浩平教授に多大なご指導とご協力を戴いた。ここに感謝の意を表す。

## 参考文献

1. 阿部隆司 2013. と畜検査データの有効活用—食検と家保の連携で生産現場に活かす—. *獣医疫学雑誌*. **17**: 10-14.
2. 足立泰基, 蒔田浩平. 2015a. 季節自己回帰和分移動平均モデルによると畜検査データの時系列分析法. *日獣会誌*. **68**: 189-197.
3. Adachi, Y. and Makita, K. 2015b. Real time detection of farm-level mycobacteriosis outbreak using time series modeling of the number of condemned intestines in abattoirs. *J. Vet. Med. Sci.* **77**: 1129–1136.
4. Ali, M., Kim, D. R., Yunus, M. and Emch, M. 2013. Time series analysis of cholera in Matlab, Bangladesh, during 1988-2001. *J. Health Popul. Nutr.* **31**: 11-19.
5. Allard, R. 1988. Use of time-series analysis in infectious disease surveillance. *Bull. World Health Organ.* **76**: 327-333.
6. Arab, A. 2015. Spatial and spatio-temporal models for modeling epidemiological data with excess zeros. *Int. J. Environ. Res. Public Health.* **12**: 10536–10548.
7. Augustin, N.H., McNicol, J. and Marriott, C.A. 2006. Using the truncated auto-Poisson model for spatially correlated counts of vegetation. *J. Agric. Environ. Stat.* **11**: 1–23.
8. Benjamin, M.A., Rigby, R.A. and Stasinopoulos, D.M. 2003. Generalized autoregressive moving average models. *J. Amer. Stat. Assoc.* **98**: 214–223.
9. Besag, J. 1974. Spatial Interaction and the statistical analysis of lattice system. *J. R. Stat. Soc. Series B. Stat. Methodol.* **36**: 192–236.

10. Bhatnagar, S., Lal, V., Gupta, S. D. and Gupta, O. P. 2012. Forecasting incidence of dengue in Rajasthan, using time series analyses. *Indian J. Publ. Health.* **56**: 281-285.
11. Box, G. E. P. and Jenkins, G. M. 1976. Time series analysis: forecasting and control, rev. ed., Holden-Day, San Francisco.
12. Briët, O.J., Amerasinghe, P.H. and Vounatsou, P. 2013. Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers. *PloS One.* **8**: e65781.
13. Briët, O.J., Vounatsou, P., Gunawardena, D.M., Galappaththy, G.N. and Amerasinghe, P.H. 2008. Models for short term malaria prediction in Sri Lanka. *Malaria J.* **7**: 76.
14. Brockwell, P. and Davis, R. 2002. Introduction of Time Series and Forecasting, 2nd ed., Springer, New York.
15. Cairns, M.E., Asante, K.P., Owusu-Agyei, S., Chandramohan, D., Greenwood, B.M. and Milligan, P.J. 2013. Analysis of partial and complete protection in malaria cohort studies. *Malaria J.* **12**: 355.
16. Corti, M. and Palmero, D. 2008. *Mycobacterium avium* complex infection in HIV/AIDS patients. *Expert Rev. Anti. Infect. Ther.* **6**: 351–363.
17. Dickey, D. A. and Fuller, W. A. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica.* **49**: 1057–1072.
18. Dickey, D. A. and Fuller, W. A. 1979. Distributions of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **75**:

427–431.

19. Dohoo, I., Martin, W. and Stryhn, H. 2009. *Veterinary Epidemiologic Research*, 2nd ed., VER Inc., Charlottetown.
20. Earnest, A., Chen, M. I., Ng, D. and Sin, L. Y. 2005. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. *BMC Health Serv. Res.* **5**: 36.
21. Eckert, J., Conraths, F.J. and Tackman, K. 2000. Echinococcosis: an emerging or re-emerging zoonosis? *Int. J. Parasitol.* **30**: 1283–1294.
22. Feinstein, A. R. 2001. *Principles of Medical Statistics*, 17th ed., Chapman and Hall/CRC, Boca Raton.
23. Ghosh, S.K., Mukhopadhyay, P. and Lu, J.C. 2006. Bayesian analysis of zero-inflated regression models. *J. Stat. Plan. Infer.* **136**: 1360–1375.
24. Holt, C. 2004. Forecasting seasonals and trends by exponentially weighted averages. *Int. J. Forecast.* **20**: 5-13.
25. Hu, M.C., Pavlicova, M. and Nunes, E.V. 2011. Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *Am. J. Drug Alcohol Abuse.* **37**: 367–375.
26. Huang, X., Williams, G., Clements, A.C. and Hu, W. 2013. Imported dengue cases, weather variation and autochthonous dengue incidence in Cairns, Australia. *PloS One.* **8**: e81887.
27. Hyndman, R. J. and Khandakar, Y. 2008. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **27**: 3.

28. 市沢成介 1975 台風予報の現状について. *天気*. **22**: 281-286.
29. Ihaka, R. and Gentleman, R. 1996. R: A language for data analysis and graphics. *J. Comp. Graph Stat.* **5**: 299-314.
30. Imai, C., Brooks, W.A., Chung, Y., Goswami, D., Anjali, B.A., Dewan, A., Kim, H. and Hashizume, M. 2014. Tropical influenza and weather variability among children in an urban lowincome population in Bangladesh. *Glob. Health Action.* **7**: 24413.
31. Jenkins, M. A. and Traub, J. F. 1972. Algorithm 419 – Zeros of a complex polynomial. *Commun. ACM* **15**: 97–99.
32. Karakousis, P. C., Moore, R. D. and Chaisson, R. E. 2004. *Mycobacterium avium* complex in patients with HIV infection in the era of highly active antiretroviral therapy. *Lancet Infect. Dis.* **4**: 557–565.
33. Karazsia, B.T. and van Dulmen, M.H. 2008. Regression models for count data: illustrations using longitudinal predictors of childhood injury. *J. Pediatr. Psychol.* **33**: 1076–1084.
34. Komijn, R. E., De Haas, P. E. W., Schneider, M. M., Eger, T., Nieuwenhuijs, J. H. M., Van den Hoek, R. J., Bakker, D., Van Zijderveld, F. G. and Van Soolingen, D. 1999. Prevalence of *Mycobacterium avium* in slaughter pigs in The Netherlands and comparison of IS1245 restriction fragment length polymorphism patterns of porcine and human isolates. *J. Clin. Microbiol.* **37**: 1254–1259.
35. 今千晴, 渡辺正基, 作井睦子, 古崎典子 2002. 北海道南部の3農場で生産された肥育豚における寄生虫性肝炎の発生状況および豚回虫 *Ascaris suum* の感染状況. *日獣会誌*. **55**: 240-243.

36. Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**: 1–14.
37. Lara-Ramírez, E. E., Rodríguez-Pérez, M. A., Pérez-Rodríguez, M. A., Adeleke, M. A., Orozco-Algarra, M. E., Arrendondo-Jiménez, J. I. and Guo, X. 2013. Time series analysis of onchocerciasis data from Mexico: a trend towards elimination. *Plos Neglect. Trop. D.* **7**: e2033.
38. LeClerc, J. and Joslyn, S. 2015. The cry wolf effect and weather-related decision making. *Risk Anal.* **35**: 385–395.
39. Liu, Q., Liu, X., Jiang, B. and Yang, W. 2011. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. *BMC Infect. Dis.* **11**: 218.
40. Ljung, G. M. and Box, G. E. P. 1978. On a measure of lack of fit in time series models. *Biometrika* **65**: 297-303.
41. Lukashenko, N.P. 1968. Comparative biologic and pathologic studies of *Alveococcus multilocularis*. *Arch. Environ. Health* **17**: 676–678.
42. Lukashenko, N.P. 1971. Problems of epidemiology and prophylaxis of alveococcosis (multilocular echinococcosis): a general review-with particular reference to the USSR. *Int. J. Parasitol.* **1**: 125–134.
43. Merkle, E.C., You, D. and Preacher, K. 2016. Testing non-nested structural equation models. *Psychol. Methods.* **21**: 151–163.
44. Moineddin, R., Meaney, C., Agha, M., Zagorski, B. and Glazier, R.H. 2011. Modeling factors influencing the demand for emergency department services in Ontario: a comparison of methods. *BMC Emerg. Med.* **11**: 13.
45. 森千恵子, 前田博之, 山田英清, 浦木増太郎, 湯浅亮 1987. と畜

- 場で廃棄処分される肉用豚のいわゆる肝臓変性の調査とその発生防止対策. *日獣会誌*. **40**: 183-189.
46. Morita, Y., Maruyama, S. and Katsube, K. 1994. Prevalence of atypical mycobacteriosis in slaughtered swine in Gunma Prefecture and the serovars of the isolates. *J. Vet. Med. Sci.* **56**: 475-479.
47. Mullahy, J. 1986. Specification and testing in some modified count data models. *J. Econom.* **33**: 341-365.
48. 中野重行 2013. 医薬品の臨床試験におけるプラセボの誕生とプラセボ対照群の必要性. *Jpn. Pharmacol. Ther.* **41**: 617-624.
49. National Institute of Standards and Technology U.S. Department of Commerce. 2013. Autocorrelation plot: random data. e-Handbook of Statistical Methods. Chapter 1.3.3.1.1. at: <http://www.itl.nist.gov/div898/handbook/eda/section3/autocop1.htm> (Accessed 29 November 2016).
50. Neumann, E., Hall, W., Stevenson, M., Morris, R. and Ling Min Than, J. 2014. Descriptive and temporal analysis of post-mortem lesions recorded in slaughtered pigs in New Zealand from 2000 to 2010. *N. Z. Vet. J.* **62**: 110-116.
51. Reis, B. Y. and Mandl, K. D. 2003. Time series modeling for syndromic surveillance. *BMC Med. Inform. Decis.* **3**: 2.
52. Rose, C.E., Martin, S.W., Wannemuehler, K.A. and Plikaytis, B.D. 2006. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *J. Biopharm. Stat.* **16**: 463-481.
53. Said, S. E. and Dickey, D. A. 1984. Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika*

- 71: 599-607.
54. Sakui, M., Ishige, M., Fukumoto, S., Ueda, A. and Ohbayashi, M. 1984. Spontaneous Echinococcus multilocularis Infection in swine in north-eastern Hokkaido, Japan. *Jpn. J. Parasitol.* **33**: 291–296.
55. Sanchez-Vazquez, M. J., Nielen, M., Gunn, G. G. and Lewis, F. I. 2012. Using seasonal-trend decomposition based on loess (STL) to explore temporal patterns of pneumonic lesions in finishing pigs slaughtered in England, 2005-2011. *Prev. Vet. Med.* **104**: 65-73.
56. Sanchez-Vazquez, M. J., Nielen, M., Gunn, G. J. and Lewis, F. I. 2012. National monitoring of Ascaris suum related liver pathologies in English abattoirs: a time-series analysis, 2005-2010. *Vet. Parasitol.* **184**: 83-87.
57. 白石光伸, 久保尚, 田中博, 高橋喜男, 松田均 1992. と畜検査データに基づく養豚農家における豚マイコプラズマ肺炎の防除対策. *日獣会誌*. **45**: 881-884.
58. Soebiyanto, R. P., Adimi, F. and Kiang, R. K. 2010. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *PLoS One* **5**: e9450.
59. Spriensma, A.S., Hajos, T.R., de Boer, M.R., Heymans, M.W. and Twisk, J.W. 2013. A new approach to analyse longitudinal epidemiological data with an excess of zeros. *BMC Med. Res. Methodol.* **13**: 27.
60. Sun, Y., Heng, B. H., Seow, Y. T. and Seow, E. 2009. Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emergency Medicine* **9**:1.

61. Trottier, H., Philippe, P. and Roy, R. 2006. Stochastic modeling of empirical time series of childhood infectious diseases data before and after mass vaccination. *Emerg. Themes Epidemiol.* **3**: 9.
62. Uruguchi, K. and Takahashi, K. 1997. The influence of livestock industrial waste on the foraging behavior of red foxes. *Jpn. J. Livest. Manage.* **32**: 75–82.
63. Venables W. N. and Ripley, B. D. 2002. Modern Applied Statistics with S-PLUS, 4th ed., Springer, New York.
64. Vial, F. and Reist, R. 2014. Evaluation of Swiss slaughterhouse data for integration in a syndromic surveillance system. *BMC Vet. Res.* **10**: 33.
65. Viehman, J. A., Khalil, D., Barhoma, C. and Hanna, R. M. 2013. *Mycobacterium avium-intracellulare* otomastoiditis in a young AIDS patient: case report and review of the literature. *HIV AIDS (Auckl)* **5**: 61–66.
66. Vuitton, D.A., Zhou, H., Bresson-Hadni, S., Wang, Q., Piarroux, M., Raoul, F. and Giraudoux, P. 2003. Epidemiology of alveolar echinococcosis with particular reference to China and Europe. *Parasitology* **127**: S87–107 sul.
67. Vuong, Q.H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**: 307–333.
68. Wangdi, K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N.J. and Kaewkungwal, J. 2010. Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: a case study in endemic districts of Bhutan. *Malaria J.* **9**: 251.

69. Winters, P.R. 1960. Forecasting sales by exponentially weighted moving averages. *Management Sci.* **6**: 324–342.
70. Yachida, S. and Shimizu, K. 1973. Studies on atypical mycobacteria isolated from tuberculous lesions of the mesenteric lymph nodes of slaughtered pigs. *Jpn. J. Vet. Sci.* **35**: 459–471.
71. Yagi, K., Uraguchi, K. and Sakui, M. 2014. Swine alveolar echinococcosis: the significance of its detection. *Jpn. J. Vet. Parasitol.* **13**: 46–53.
72. Yang, M., Zamba, G.K.D. and Cavanaugh, J.E. 2013. Markov regression models for count time series with excess zeros: a partial likelihood approach. *Stat. Methodol.* **14**: 26–38.
73. Zeger, S.L., Irizarry, R. and Peng, R.D. 2006. On time series analysis of public health and biomedical data, *Annu. Rev. Publ. Health* **27**: 57–79.
74. Zeger, S.L. and Qaqish, B. 1988. Markov regression models for time series: a quasi-likelihood approach. *Biometrics* **44**: 1019–1031.
75. Zeileis, A., Christian, K. and Simon, J. 2008. Regression models for count data in R. *J. Stat. Softw.* **27**: 8.
76. 全国食肉衛生検査所協議会編. 1985. 食肉衛生検査病理学カラーアトラス. 学窓社. 東京.
77. 全国食肉衛生検査所協議会編. 2011. 新・食肉衛生検査マニュアル. 中央法規. 東京.
78. Zhu, J., Rasmussen, J.G., Møller, J., Aukema, G.H. and Raffa, K.F. 2008. Spatial-temporal modeling of forest gaps generated by colonization from below-and above-ground bark beetle species. *J. Am.*

*Stat. Assoc.* **103**: 162–177.