

ヘイズ災害のリスク解除情報の 作成方法に関する研究

平成 23 年 3 月 11 日

京都大学大学院工学研究科

都市社会工学専攻

原田 大輔

要 旨

本研究は、インドネシアのスマトラ島での森林火災によって発生する煙害ヘイズ災害のリスク解除情報を作成することを目的とする。インドネシアのスマトラ島の森林火災は、周辺各国における日常的サービスの中断、呼吸器系疾患の多発等の深刻なヘイズ（煙害）災害をもたらしている。マレーシア政府は、大気中の汚染物質量をリアルタイムに観測し、ヘイズ警報を発令するシステムを開発している。ヘイズ警報を発令（解除）するためには、衛星情報と地上観測データに基づいて、将来時刻における大気汚染物質の滞留量を迅速に予測するような統計的予測モデルが有用である。本研究では、ヘイズ災害に関する危機管理情報を作成するための方法論として、インドネシアの火災発生・終息メカニズムを Duration モデルを用いてモデル化すると共に、それに基づくレジーム変化 (regime switching) と長期記憶性を考慮した統計的時空間モデルを構築する。さらに、マレーシア半島部を対象としたケーススタディにより、本研究で提案した方法論の有効性を実証的に検証する。

目次

第 1 章	はじめに	1
第 2 章	本研究の基本的な考え方	3
2.1	本研究の枠組み	3
2.2	Duration Model に基づく森林火災モデル	3
2.3	ARFIMA モデルと RS モデル	4
2.4	RS-P-STARFIMA モデルの構成	6
2.5	本研究における分析目的	8
第 3 章	Duration モデルを用いた森林火災モデル	10
3.1	Geo Additive 関数	10
3.2	hotspot 数予測モデル	12
第 4 章	P-STARFIMA モデル	13
4.1	ARFIMA モデル	13
4.2	周期的実数値と分時系列過程	14
4.3	P-STARFIMA モデル	15
第 5 章	RS-P-STARFIMA モデル	18
5.1	レジームの定義	18
5.2	RS モデル	19
5.3	RS-P-STARFIMA モデル	20
第 6 章	モデルの推計方法	22
6.1	RS-P-STARFIMA モデルの推計手順	22
6.2	RS モデルのセミパラメトリック推計	22
6.3	RS-P-STARFIMA モデルの推計	26
第 7 章	適用事例	29
7.1	データベース	29
7.1.1	ホットスポット数	29
7.1.2	PM10 濃度	29
7.2	duration モデルを用いた森林火災モデルによるホットスポット数の推計	30

7.3 RS-P-STARFIMA モデルの特定化	31
7.4 RS モデルの推計	34
7.5 P-STARFIMA モデルの推計	36
7.6 危険予告情報の作成	39
7.7 危険解除情報の作成	42
 第 8 章 おわりに	 47
 第 9 章 付録 補足説明	 49
 参考文献	 53

第1章 はじめに

インドネシアのカリマンタン島、スマトラ島における森林火災により、東南アジア周辺諸国では、毎年のように深刻なヘイズ（haze: 煙害）災害¹⁾が発生している。特に、1997年、2005年に大規模なヘイズ災害が発生し、周辺諸国で呼吸器系疾患の多発、日常生活の中断等の深刻な被害が発生した²⁾。特に、マレーシアにおけるヘイズ災害は深刻であり、1997年、2005年にはマレーシア半島の中部地域の2カ所において非常事態宣言が発令された。この経験を踏まえ東南アジア地域各国において、ヘイズ災害に関する正確な予測と予告システムの構築が望まれている。

1997年、2005年のヘイズ災害では、大気汚染物質が南西モンスーン風により運ばれ、マレーシア半島に12時間程度で到着したことが報告されている。一方、過去の経験より、マレーシア国内の経済活動や地域住民が、ヘイズ災害に対して事前の予防措置を講じるためには、少なくとも6時間程度の時間的余裕が必要となる。このために、マレーシア政府は、ヘイズ災害の発生に先だって、ヘイズに関する危険予告情報を発令する危機管理システムを開発している。さらに、ヘイズ災害が長期化することから、ヘイズ終息時刻の予測も重要な課題となっている。

ヘイズ災害の特徴として、1) 森林火災発生からヘイズ発生まで時間的余裕が極めて乏しいこと、2) 小規模なヘイズは頻繁に発生しており、ヘイズの規模に関する判断が必要なこと、3) 汚染物質が空気中に滞留することから、ヘイズ災害が長期化することがあげられる。さらに、ヘイズ災害は国境を越える環境汚染であり、現在のところマレーシア政府がスマトラ島における森林火災の発生や、大気汚染物質の移動に関する詳細なモニタリング情報を獲得することは不可能である。ヘイズ災害に関する危機管理を行う上で、衛星情報やマレーシア国内の観測情報に基づいて、ヘイズの発生と終息を短期間に予測できるような簡便な予測モデルが必要とされている。

以上の問題意識に基づいて、本研究では、ヘイズ発生の時空間的特性を表現しうるような時空間統計的予測モデルを開発することとする。すでに、Matsuta等はマレーシア半島におけるヘイズの時空間特性を詳細に分析し、ヘイズ現象の1) 長期記憶性、2) 周期変動性、3) 空間相関性について考察している³⁾。さらに、4) ヘイズ発生期には、高濃度の大気汚染物質が急激に増加することも指摘している。また、塚井ら^{?)}は、インドネシアの森林火災の発生メカニズムについて、duration modelに基づく火災発生・終息メカニズムについて定式化し、衛星画像に基づく地表面温度が閾値を越える地点（以下、hotspot）データを用いてその有効性を示している。

本研究では、これら既往の研究を統合させ、インドネシアの火災発生・終息メカニズムと

マレーシアのヘイズ災害発生・終息メカニズムを統計的手法によりモデル化し、ヘイズ災害のリスク解除情報の作成方法に関する方法論を提示する。具体的には、火災発生・終息メカニズムを空間加法的ハザード関数を用いてモデル化すると共に、大気汚染物質の時空間特性である、長期記憶性、周期変動性、空間相関性を表現するために、周期的時空間実数と分自己回帰移動平均モデル (Periodically Spatio-Temporal Auto-Regressive Fractionally Integrated Moving-Average model: 以下、P-STARFIMA モデルと略す) を定式化する。さらに、ヘイズ発生による大気汚染物質の急激な増加現象を、レジーム変化モデル (regime switching model: 以下、RM モデルと略す) を用いて表現する。その上で、ヘイズ発生期と通常期を通じた大気中における汚染物質濃度の時空間変動をレジーム変化を考慮した P-STARFIMA モデル (Regime Switching P-STARFIMA model: 以下、RS-P-STARFIMA と略す) を提案する。以下、2. で、本研究の基本的な考え方を説明する。3. で duration model を用いた森林火災モデルを定式化する。4. で、P-STARFIMA モデルを定式化し、5. では、RS モデルを定式化した後、RS-P-STARFIMA モデルに拡張する。6. で、モデルの推計方法を説明する。7. で、マレーシアを対象とした適用事例について考察する。

第2章 本研究の基本的な考え方

2.1 本研究の枠組み

大気汚染の予測手法として、汚染物質の時空間的な動態の支配方程式を、大気の力学モデルに基づいて同定する力学的アプローチと、大気観測データの時空間相関構造を統計的に同定する統計的アプローチがある。前者は、力学理論に基づいて構造モデル^{4), 5)}を作成する方法であるが、広範囲をカバーするような動学的モデルを構築するためには多くの入力情報が必要となり、計算時間と観測費用が膨大になるという難点がある。一方、統計的な予測手法は、データの持つ時間的・空間的な相関構造に基づいたモデルを作成する方法であり、将来予測の信頼水準を議論することが可能である。しかし、モデルの予測精度は、モデルや観測誤差項の特定化に高度に依存している。確率誤差構造の特定化が不適切であれば、パラメータ推計量に系統的バイアスが生じたり、パラメータの推計効率が低下するという問題が発生する⁶⁾。本研究では、1. で言及したようなヘイズの特性や、ヘイズ生起に関する危険情報の作成という予測モデルの目的を考慮して、時空間モデルを用いた統計的アプローチを採用する。

本研究で提案するモデルは、1) Duration Model に基づく森林火災モデル、2) レジーム変化モデル (Regime Switching model)、3) 周期的時空間実数値自己回帰移動平均モデル (Periodically Spatio-Temporal Auto-Regressive Fractionally Integrated Moving-Average model, P-SARFIMA モデル) の三つの要素モデルから構成される。ヘイズ災害のリスク解除情報を作成するためには、現在のインドネシアの火災状況から、今後の火災の拡散・鎮火過程をモデル化し、その結果を用いて、マレーシアにおいてヘイズ災害が発生または終息するかを予測する必要がある。本研究では、そのような観点から上記の三つのモデルの統合を行い、ヘイズ災害のリスク解除情報、すなわちヘイズの継続時間や終息時期に関する情報を提供するための方法論を提示する。なお、図-2.1 は、本研究のモデルの構造を表している。

2.2 Duration Model に基づく森林火災モデル

森林火災の発生と延焼過程の予測について、先行研究ではセルラートオートマタモデル (CA モデル) が提案されている⁷⁾。CA モデルでは地表面のローカルな被覆特性を、対象領域をメッシュ分割した各セルの特性として与え、任意の出火点から火災が延焼する過程の支配パラメータを、既往火災データに基づいて同定する。CA モデルを用いれば、対象領域内

の各地点・各時点の森林火災の有無を、シミュレーションによって求めることができる。しかし、そのためには地表面の被覆特性に関する情報が必要であり、データ収集負荷が大きくなる。このため、本研究では、よりデータ収集負荷の少ない煙害予測・警報システムを構築するために、火災地点数（以下、ホットスポット数）に基づく森林火災モデル及びヘイズ災害モデルを構築する。ホットポイント数については、その時点で観測したスマトラ島の衛星写真に基づいて、地表面の温度が一定以上の閾値を超える地点をスマトラ島全体で集計することによって求めることが可能である。後に説明するレジームスイッチングモデルでは、入力情報としてホットポイント数の時系列データを用いる。また、本研究では、火災の特性をより反映するため、森林火災が発生するまでの期間、および火災が発生してからの森林火災継続期間のそれぞれを確率変数とみなし、ある時点までに得られる情報に基づいて森林火災規模を予測するモデルとして Duration model の適用を検討する。なお、Duration model を用いた火災延焼・消滅モデルについては幾つか先行研究が存在する。例えば、Reed ら⁹⁾は、4つのパラメータを有する competing hazard model を用いて火災の発生・消滅メカニズムを考慮した上で、森林火災の規模の推定モデルを定式化し、大規模火災を除いて、統計的に優位な結果を得ている。ただし、既往の研究では火災の発生・消滅過程の空間的相互作用を取り扱っていない。本研究では火災の空間的特性をより反映するために、空間加法的ハザード関数を用いて火災の発生・消滅メカニズムのモデル化を行う。このようなモデル化により、火災の空間的な拡散過程を反映したシミュレーションが可能となる。筆者の知る限り、火災の発生・消滅メカニズムを分析した既往研究において、空間加法的ハザード関数を用いて分析した事例は存在しない。

2.3 ARFIMA モデルと RS モデル

1. で言及したようなヘイズの時空間特性を適切に記述するためには、1) 長期記憶性、2) 周期変動性、3) 空間相関性を考慮したモデルの開発が必要である。しかし、これら三点を考慮したモデルは筆者が知る限り見当たらない。ここで既往研究を俯瞰しておこう。時空間モデルに関する研究は、Cliff and Ord⁷⁾等が先鞭を付けた。さらに、伝統的な Box and Jenkins 流の ARMA 時系列モデル¹⁰⁾に、空間構造を導入した時空間自己回帰移動平均 (Space-Time AutoRegressive Moving Average (STARMA) モデルが提案された¹¹⁾。Pfeifer and Deutsch¹²⁾は、STARMA モデルに和分を導入した STARIMA (Space-Time AutoRegressive Integrated Moving Average) モデルを提案している。これらの時空間モデルは、環境モデリングの分野^{13), 14)}を中心に、多くの適用事例が報告されている^{15), 16)}。また、空間構造を特定化することにより、STARIMA モデルを拡張する試みもされている。例えば、Kamarianakis and Prastacos^{17), 18)}は、道路ネットワークの構造特性を階層的隣接構造で表現したような STARIMA モデルを提案している。さらに、Niu 等^{19), 20)}

は、STARIMA モデルに周期的タイムラグ構造を導入した Seasonal STARIMA モデルを定式化し、気象・天候データの周期的時空間変動パターンをモデル化している。また、近年、これらの研究を踏まえて、系列相関関数の収束速度が非常に緩慢な長期記憶性を持つ時系列データに関する研究が発展しつつある。時系列データに含まれる長期記憶性は、Box and Jenkins¹⁰⁾ 流の和分アプローチでモデル化することが不可能である。長期記憶性を持つ非定常確率過程をモデル化するために、原系列の実数と分を考慮する自己回帰実数と分移動平均 (AutoRegressive Fractionally Integrated Moving Average: 以下、ARFIMA と略す) モデルが提案された^{21)–23)}。Woodward 等²⁴⁾ は、実数と分パラメータ d が $|d| < 1/2$ を満たす場合、確率過程が定常性を有し、反転可能であることを示した。さらに、Katayama²⁵⁾ は季節和分自己回帰移動平均移動 (Seasonally Integrated AutoRegressive Moving Average: SIARMA)²⁶⁾ モデルに実数と分を導入した SFIARMA (Seasonally Fractionally Integrated AutoRegressive Moving Average) モデルを提案している。

一方、レジーム変化 (regime switching: 以下、RS と略す) という時系列データの非線形構造変化に関する研究が発展しつつある。RS モデルは Hamilton²⁹⁾ によって提案され、景気分析や金融計量経済学の分野で応用研究が展開されてきた。従来の時系列モデルは、データ生成プロセス (Data Generation Process: DGP と略す) を表現した単一の構造方程式を仮定するのに対して、RS モデルでは、複数の DGP を表現した構造方程式と各 DGP によってデータが生成されるレジームを仮定し、複数のレジーム間の推移現象によって時系列データの構造変化を表現する。そのため、RS モデルは、各レジームに対応する構造方程式とレジーム間推移確率を含むモデルシステムとして定式化される³⁰⁾。Hamilton²⁹⁾ は、レジーム間の推移確率過程を Markov 過程で表現した Markov switching モデルを提案した。Kim and Nelson³¹⁾ は、Markov switching モデルのパラメータの推計法について詳細な検討を行っている。さらに、Markov 推移確率過程が、外生変数の関数として表現される時間共変型レジーム変化 (Time-varying Regime Switching: 以下、TRS と略す) モデルも提案されている³²⁾。TRS モデルでは異なる期待値を持つレジームを想定するため、レジーム間の推移によって、系列相関関数の収束速度が緩慢な DGP を表現することが可能となる。そこで、これまでに RS モデルと ARFIMA モデルの間で、それぞれが表現する DGP 特性の相違点や、将来予測能力について検討が行われてきた。Ryden は、アメリカの株価指標の分析に RS モデルを適用した結果、レジームの推計が適切ではないとき、RS モデルは系列相関関数の収束速度が緩慢な DGP を十分に表現できない場合があると報告している³³⁾。Bos 等³⁴⁾ は、レジームを与件とした ARFIMA モデルを用いて、G7 各国のインフレ率の時間変動を実証分析している。この研究では、レジーム変化を考慮すると実数と分パラメータの値は減少するものの、依然として有意な長期記憶性が見られると報告している。すなわち、ARFIMA モデルによって表現される DGP の特性は、RS モデルによって代替的に表現できないと結論づけている。

一方, Granger and Hyung³⁵⁾は, レジーム変化が存在する時系列に対してレジーム変化を考慮しないARFIMAモデルを当てはめる misspecification (特定化の誤り)があると, 存在しない筈の長期記憶性, すなわち「見せかけの長期記憶性」が表れることを示唆している. 「見せかけの長期記憶性」については, Diebold and Inoue³⁰⁾が, モンテカルロシミュレーションによって発生させた時系列データに基づいた検討を行っている. この研究では, レジームの変化確率が低い, すなわち, 同一レジームの継続期間が長いほど, 実数と分パラメータの推計値が大きくなると報告されている. さらに, Gabriel and Martins³⁶⁾は, ポルトガルでのインフレ率の変化に対してARFIMAモデルとMarkov switchingモデルを適用して, 両者の予測力の比較を行っている. この研究では, レジーム変化によって期待値が大きく変化する場合にARFIMAモデルを適用すると, Markov switchingモデルと比較して, 将来予測力が非常に低いことを明らかにした. なお, 時空間モデルに関する研究は, Cliff and Ord⁷⁾等が先鞭を付けた. さらに, 伝統的なBox and Jenkins流のARMA時系列モデル¹⁰⁾に, 空間構造を導入した時空間自己回帰移動平均 (Space-Time AutoRegressive Moving Average (STARMA) モデルが提案された¹¹⁾. Pfeifer and Deutsch¹²⁾は, STARMAモデルに和分を導入したSTARIMA(Space-Time AutoRegressive Integrated Moving Average)モデルを提案している. これらの時空間モデルは, 環境モデリングの分野^{13), 14)}を中心に, 多くの適用事例が報告されている^{15), 16)}. また, 空間構造を特定化することにより, STARIMAモデルを拡張する試みもされている. 例えば, Kamarianakis and Prastacos^{17), 18)}は, 道路ネットワークの構造特性を階層的隣接構造で表現したようなSTARIMAモデルを提案している. さらに, Niu 等^{19), 20)}は, STARIMAモデルに周期的タイムラグ構造を導入したSeasonal STARIMAモデルを定式化し, 気象・天候データの周期的時空間変動パターンをモデル化している.

このように長期記憶モデルに関する研究が蓄積されているが, 筆者の知る限り長期記憶性を考慮した時空間モデルに関しては, ほとんど研究が蓄積されていない. ヘイズ災害の特性を捉えるためには, 周期性, 長期記憶性, 空間相関性を表現できる時空間統計モデルを開発する必要がある. このため, 本研究では, RS(Regime Switching)モデルとP-STARFIMAモデルを統合したRS-P-STARFIMAモデルを用いて分析を実施する.

2.4 RS-P-STARFIMAモデルの構成

マレーシア半島におけるヘイズ災害は, 1) スマトラ島における森林火災の発生により, 時間遅れを伴いながら, 大気汚染物質の濃度が急激に上昇する, 2) 一度, マレーシア半島に到着した大気汚染物質は, 空気中に長時間にわたって広域的に滞留し, ヘイズ災害が長期間継続するという特徴的な時空間変動特性を有している. ヘイズ災害の時空間分布予測に関しては, 気象に関する力学モデルに基づく方法も開発されているが, 限られた情報と

時間の中で、マレーシア半島におけるヘイズ災害の発生を精度良く予測するためには、ヘイズ災害の時空間特性を表現しうる統計的予測モデルを開発し、力学モデルとの併用により、信頼性の高い警報システムとする必要がある。しかしこれまでのところ、そのような統計的予測モデルの開発は行われていない。森林火災による急激なヘイズ発生期の DGP は、通常期とは異なる DGP に従っている可能性が高い。また、本研究の目的がヘイズ発生期の将来予測にある以上、大気汚染物質に関する時系列データの発生プロセスのレジーム変化を適切にモデル化する必要がある。

既存研究^{33, 34)}により、RS モデルが表現する DGP 特性は ARFIMA モデルが表現する DGP 特性とは異なっていることが指摘されている。ただし、異なるレジーム間の推移確率が低い場合、Markov switching モデルと長期記憶モデルのいずれが適切であるかの識別は難しく、さらに、レジーム変化によって期待値が大きく変化する場合は ARFIMA モデルは、予測力に問題がある。したがって、レジーム変化と長期記憶性の両方を考慮したモデルの開発を行うことにより、急激な時系列期待値のシフトに対しては RS モデルの DGP 特性を、また、期待値シフト後の観測値が示す、緩やかな系列期待値への回帰に対しては長期記憶モデルの DGP 特性を、それぞれ対応させることが可能と考えられる。すなわち、RS モデルは ARFIMA モデルで十分に表現できないレジームの急激な変化を補完する役割を果たすことが期待される。しかし、上述の統合モデルの適用例は、著者らの知る限り、Haldrup and Nielsen³⁷⁾による電力料金の時間変動にレジーム変化を考慮した Seasonal ARFIMA モデル等に留まるなど、十分ではない。そこで、本研究では、ヘイズ発生機構 (DGP) の構造変化を表現するレジーム変化モデルと、ヘイズ滞留特性を記述する P-STARFIMA を連動したレジーム変化周期的時空間自己回帰実数値と分移動平均 (Regime Switching Periodically Spatio-Temporal AutoRegressive Fractionally Integrated Moving Average: 以下、RS-P-STARFIMA と略す) モデルを提案する。

RS-P-STARFIMA モデルを構成する RS モデルは、衛星情報によるスマトラ島における森林火災に関する先行情報を用いて、マレー半島における大気汚染物質のデータ発生レジームの変動を予測する役割を果たす。すなわち、RS モデルは、マレーシアに運ばれる大気汚染物質の濃度変化の期待値を外生的に予測する。マレーシア半島に運ばれた大気汚染物質は、長期間にわたりマレーシア半島内に滞留することになる。P-STARFIMA モデルは、マレーシア半島に滞留する大気汚染物質濃度の空間相関、長期記憶性、周期性を表現する。このように RS-P-STARFIMA モデルは、スマトラ島における森林火災の発生によって生じる原系列の急激な外生的な変化を RS モデルによって表現すると同時に、マレーシア半島に滞留する汚染物質濃度の空間相関・長期記憶を有する周期的変動を P-STARFIMA モデルによって表現する。

2.5 本研究における分析目的

ヘイズ災害に対する危機管理を行うために、マレーシア政府は国内の経済活動や地域住民に、さまざまな情報を提供している。このうち、本研究では、Duration モデルを用いた森林火災モデルと RS-P-STARFIMA モデルの統合モデルを用いて、ヘイズ災害の終息を予告する危険解除情報を作成する方法論を提案する。

1997 年、2005 年のヘイズ災害では、大気汚染物質が南西モンスーン風により運ばれ、マレーシア半島に 12 時間程度で到着したと報告されている。1. で言及したように、マレーシア国内の経済活動や地域住民が、ヘイズ災害に対して事前の予防措置を講じるためには、少なくとも 6 時間程度の時間的余裕が必要となる。そこで、本研究ではリアルタイムで観測できる大気汚染濃度と、スマトラ島における森林火災の状況に基づいて、マレーシア半島におけるある一定時間後の大気汚染物質濃度をシミュレートするための方法論を提案する。なお、マレーシアのヘイズ災害は、森林火災による大気汚染物質の外生的な供給によって発生する。そこで、森林火災等の外生要因の状態により、大気汚染物質濃度データの生成プロセス (Data Generation Process:DGP と略す) が異なると考え、時系列データが同一の DGP に従う期間をレジームと定義する。本研究では、各 CA 局における各時刻のレジーム状態を決定する外生要因として、インドネシアスマトラ島の森林火災 (以下、hotspot とする) 規模を表す hotspot 数を採用する。すなわち、スマトラ島の森林火災規模を火災発生地点の時間的・空間的な特性に基づいて内生的に求めることのできる統計モデルが必要となる。そこで、hotspot の継続期間に着目し、時空間データに基づく生存関数モデルの適用を試みる。

本研究が提案する方法は以下の通りである。まず、1) スマトラ島での森林火災が発生するまでの期間、および火災発生後の森林火災継続期間のそれぞれを確率変数とみなし、森林火災モデルを用いて hotspot 数を推計する。2) 1) で求めた hotspot 数を用いて、ヘイズ発生の推移確率を RS モデルを用いて予測する。次に、3) 予測される確率が最大となるレジームを与件とした P-STARFIMA モデルを用いて大気汚染物質予測濃度を算出する。以上の方法で求めた、ヘイズ発生レジームの生起確率と大気汚染物質濃度の予測値は、ヘイズ危険予告情報の発令のための有用な基礎情報を提供すると考えられる。

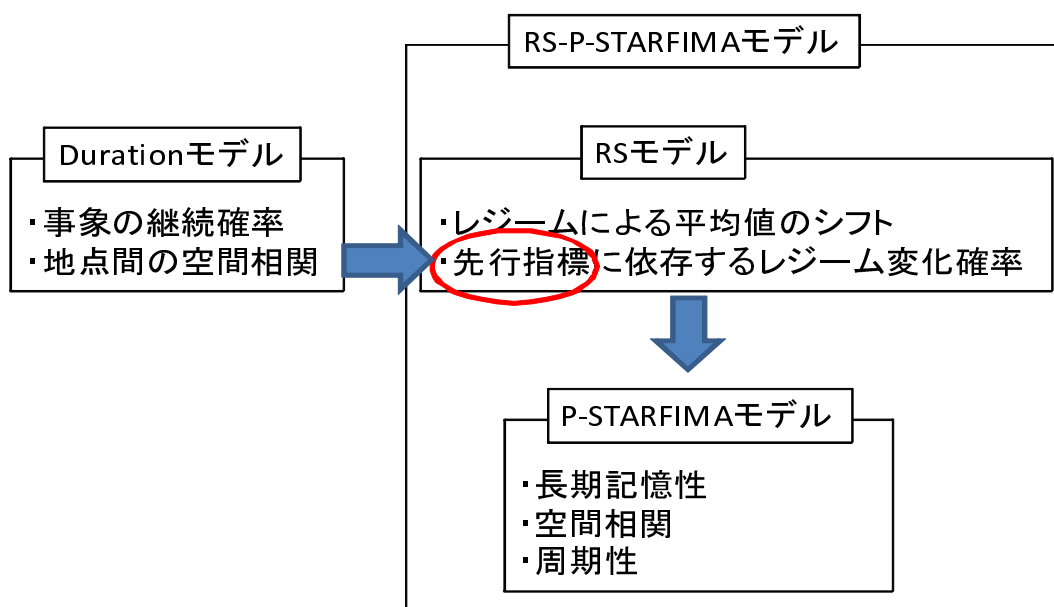


図-2.1 統合モデルの構造

第3章 Durationモデルを用いた森林火災モデル

3.1 Geo Additive 関数

生存分析とはあるイベントが発生するか否か、またイベントが発生するまでにどの時間がかかるかということを同時に処理する手法である。まず生存分析に用いられる基本事項を簡単に説明しておこう。生存分析の基本的な考え方はある事象の終了する時点が確率分布に基づくということにあり、ある事象の発生時点を T とすると、任意の時点 t までにその事象が発生している場合を累積分布関数から

$$F(t) = Pr\{T < t\} \quad (3.1)$$

とする。したがってイベント発生時刻 T が t 以上になる場合には、まだイベントが発生していないことを意味し、あるイベントの生存確率分布 $S(t)$ を用いて表すと、

$$S(t) = Pr\{T \geq t\} = 1 - F(t) \quad (3.2)$$

となる。一方、hazard 関数とは、時刻 t である単一の状態から他の状態に移行する確率を意味する。つまり時刻 t まである一定の状態が継続する条件付き確率と考えられる。このとき時刻 t に従うイベント発生確率密度関数 $f(t)$ 、生存関数 $S(t)$ を用いて以下の関係を得る。

$$\begin{aligned} h(t) &= \lim_{h \rightarrow 0} \frac{Pr(t \leq T < t+h | T \geq t)}{h} \\ &= -\frac{d \ln S(t)}{dt} \\ &= \frac{f(t)}{S(t)} \end{aligned} \quad (3.3)$$

さらに proportional hazard 関数を定義しよう。

$$h(t, x, \beta) = h_0(t) \{\Phi(x, \beta)\} \quad (3.4)$$

ここで、 $h_0(t)$ は時刻 t に依存した基準ハザードであり、ハザード率の基本形状を表す。一方、 $\Phi(x, \beta)$ は時刻 t に依存しないイベント発生要因となる説明変数 x と、そのパラメータ β で構成されており、各時刻でハザード率の大きさに影響を及ぼす。したがって、proportional hazard 関数ではハザード関数が各主体に共通のベースラインハザード (baseline hazard : $h_0(t)$) と各主体に独自の相対ハザード (relative hazards : $\Phi(x, \beta)$) の積で表され、イベント発生確率が基準ハザードに比例する形で決定される。

式 (1) の関係より観測期間内に地点 i において火災が発生しない right censoring の有無を記号 δ_i で表すと、対数尤度関数 L は、

$$L = \sum_i \left\{ \delta_i \log h_i(t) - \int_0^t h_i(u) du \right\} \quad (3.5)$$

となる。ただし $\delta_i=1/0$ のとき right censoring が起こっている/起こっていないことを表す。

kneib and Fahrmeir は、空間データを対象とした duration モデルとしてハザード関数に空間特性をはじめとする柔軟な関数形を設定した geo additive 関数を提案した。

$$\begin{aligned}\Phi^i(x, \beta) = & \exp\{\gamma_0 + \sum_{j=1}^j \gamma_j x_{ij} + \sum_{k=1}^k g(t) x_{ik} \\ & + \sum_{l=1}^l f_l(x_{il}) + f_i(s)\}\end{aligned}\quad (3.6)$$

ここで x_{ij}, x_{ik}, x_{il} は観測開始時刻の地点 i の特性を表わす定数型共変量である。一方、 $\gamma_j, g(t)$ はそれぞれ時刻 t に依存しない定数パラメータ、時刻 t に依存する定数パラメータを表す。また $f_l(x_{il})$ は共変量 x_{il} に関する 3 次スプライン関数などが提案されている。 $f_i(s)$ は地点 i 近傍の固有効果を表す項であり、たとえば場の条件として、マルコフ場や、正規場などのランダム効果が仮定される。また kneib and Fahrmeir はパラメータ推定時に、地点 i 近傍の固有効果を考慮したパラメータ間の制約を表すペナルティ項を追加した。

$$\begin{aligned}f_i(s) &= \frac{1}{N_i} \sum_{s \in D_i} \beta_s + u_s \\ u_s &\sim N\left(0, \frac{\delta_s^2}{N_i}\right)\end{aligned}\quad (3.7)$$

β_s は領域 s の固有効果パラメータ、 D_i は地点 i 近傍であり、 N_i は近傍 D_i に含まれる s の領域数である。また、 u_s はランダム項であり、 δ_s^2 はランダム項の分散を表す。なお各地点 i と領域 s の隣接関係は $N \times S$ の行列 L_i で与えられる。ただし N は地点数、 S は固有効果を持つ領域数である。 $S \times 1$ のパラメータベクトルを β 、領域 S 間の近接度を与える $S \times S$ のペナルティ行列を $K_s = L_i' L_i$ とすれば式 (3.7) のマルコフ場に対応するペナルティ項 P_s は

$$P_s = -\frac{\beta' K_s \beta}{2\delta_s^2}\quad (3.8)$$

で与えられる。このとき地点 i 近傍の固有効果を考慮した対数尤度関数 L_p は

$$L_p = \sum_i^i \{\delta_i \log h_i(t) - \int_0^t h_i(u) du\} - P_S\quad (3.9)$$

となる。また式 (3.7) を未知パラメータについて直接最大化するとペナルティ行列が正則ではない場合は、分散 δ_s^2 の推計値を求めるための周辺分布が得られない。kneib and Fahrmeir は、 δ_s^2 に適当な初期値を与えてハザード関数のパラメータを推計し、推計されたパラメータを所与とする周辺分布関数を解くことで、 δ_s^2 を更新する手順を、パラメータが収束するまで繰り返す計算アルゴリズムを考案した。kneib and Fahrmeir の方法は、周辺尤度関数をラプラス近似すると同時に、式 (3.9) 第 1 項（通常の尤度関数）は δ_s^2 の変化に関する感度が鈍い性質に着目して、以下の集中化周辺尤度関数を最大化する

$$L(\delta_s^2) = -\frac{1}{2} \sum \log |\delta_m^2| - \log |\mathbf{H}| - P_S\quad (3.10)$$

ここで H は, 式 (3.9) のフィッシャー情報量行列である. 式 (3.10) を最大化することにより, δ_s^2 の更新値を得る.

3.2 hotspot 数予測モデル

将来時刻 $t + \tau$ の hotsspot 数 $x_{t+\tau}$ は, 推計された duration モデルを用いて以下の手順によって求める.

- 未出火地点 (時点 t の未出火地点数: N_t^o) については, 現在時点 t で入手できる情報に基づいてハザード関数を更新して, 将来時刻 τ までの出火時点数 $\Delta x_t^o(\tau)$ を,

$$\Delta x_t^o(\tau) = \sum_{i=1}^{N_t^o} \{S_i^o(t) - S_i^o(t + \tau)\} \quad (3.11)$$

によって求められる.

- 出火地点 (時点 t の出火地点数: N_t^d) については, 現在時点 t で入手できる情報に基づいてハザード関数を更新して, 将来時刻 τ までの鎮火時点数 $\Delta x_t^d(\tau)$ を,

$$\Delta x_t^d(\tau) = \sum_{i=1}^{N_t^d} \{S_i^d(t) - S_i^d(t + \tau)\} \quad (3.12)$$

によって求められる.

- 上記, 求められた将来時刻 τ までの出火時点数 $\Delta x_t^o(\tau)$ と鎮火時点数 $\Delta x_t^d(\tau)$ を用いて将来時刻 $t + \tau$ までの hotspot 数は, 以下のように求まる.

$$x_{t+\tau} = x_t + \{\Delta x_t^o(\tau) - \Delta x_t^d(\tau)\} \quad (3.13)$$

ただし, x_t は現在時刻 t の hotspot 数である.

第4章 P-STARFIMA モデル

4.1 ARFIMA モデル

読者の便宜を図るため、P-STARFIMA モデルの定式化に先立って、ARFIMA モデルを簡単に紹介しておく。ARFIMA モデルは、ARIMA モデルで用いられる和分パラメータを実数次元に拡張したモデルである。\$L\$ をタイムラグ作用素 (\$Lx_t = x_{t-1}\$) とすれば、実数和分作用素 \$\nabla^d\$ は

$$\nabla^d = (1 - L)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-L)^j \quad (4.1)$$

と表される。実数和分の意味を説明するために、以下のような簡単な確率過程をとりあげよう。

$$\nabla^d x_t = \varepsilon_t \quad (4.2)$$

ただし、\$\{\varepsilon_t\}_{t=0}^{\infty}\$ は互いに独立で同一の平均 0、分散 \$\sigma^2\$ を持つ正規分布に従うホワイトノイズ過程である。上式において、実数和分パラメータ \$d\$ が \$|d| < 1/2\$ を満足する時、以下の性質が成立する²²⁾。

a) 確率過程 \$\{x_t\}_{t=0}^{\infty}\$ は定常過程であり、\$\varepsilon_t\$ のラグ係数多項式を \$\psi(L)\$ とすれば、無限次元の移動平均表現

$$x_t = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad (4.3)$$

が可能である。ただし、

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \quad (4.4)$$

であり、\$\Gamma(\cdot)\$ はガンマ関数である。\$j \rightarrow \infty\$ の極限において、\$\psi_j \sim j^{d-1}/(d-1)!\$ が成立する。

b) 確率過程 \$\{x_t\}_{t=0}^{\infty}\$ は反転可能であり、\$x_t\$ のラグ係数多項式を \$\pi(L)\$ とすれば、無限次元の自己回帰表現

$$\pi(L)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = \varepsilon_t \quad (4.5)$$

が可能である。ただし、

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} \quad (4.6)$$

であり, $j \rightarrow \infty$ の極限で $\pi_j \sim j^{-d-1}/(-d-1)!$, $\pi_j(d) = \psi_j(-d)$ が成立する.

以上の性質より, 実数分パラメータ d が $-1/2 < d < 1/2$ を満足する時, 確率過程 $\{x_t\}_{t=1}^\infty$ は, 定常かつ反転可能である. さらに, AR 過程ラグ次元 p , MA 過程ラグ次元 q を持つ ARMA 過程の AR パラメータ ϕ_i ($i = 1, \dots, p$) と MA パラメータ θ_j ($j = 1, \dots, q$) の収束速度は指数関数のオーダーを持つが, 実数分パラメータのラグ効果は多項式のオーダーで減衰する. すなわち実数分過程のパラメータの減衰は AR 過程や MA 過程と比較して緩慢であり, 大きなラグの過去から現在への影響が無視できない.

以上の議論を拡張し, 実数分パラメータ $0 < d < 1/2$ と, AR 過程のラグ次元 p , MA 過程のラグ次元 q を持つ長期記憶を持つ定常確率過程

$$\phi_p(L)\nabla^d x_t = \theta_q(L)\varepsilon_t \quad (4.7)$$

を考える. なお, $\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$, $\theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$ によって定義する. このような実数分次元 d を持つ確率過程を ARFIMA(p, d, q) と表そう. 実数分パラメータ d が整数の場合, ARFIMA(p, d, q) は通常の ARIMA(p, d, q) に一致する. $-1/2 < d < 0$ を満足する ARFIMA($0, d, 0$) 過程は, 反長期記憶過程³⁸⁾と呼ばれる. 長期記憶過程に関する詳細は, Beran³⁹⁾を参照して欲しい. 実数分パラメータの推計問題に関しては, Geweke and Porter-Hudak⁴⁰⁾が先鞭をつけ, パラメータ d の推計量が種々提案されている^{41), 42)}. ARFIMA(p, d, q) モデルの推計方法に関しては参考文献^{43)–47)}を参照して欲しい.

4.2 周期的実数分時系列過程

ARFIMA モデルの中に, 周期的変動を組み込もう. 周期的実数分時系列過程の一般形は,

$$\begin{aligned} \phi_p(L)\Phi_P(L^s)\nabla^d \nabla_s^{d_s} x_t \\ = \theta_q(L)\Theta_Q(L^s)\varepsilon_t \end{aligned} \quad (4.8)$$

と表せる. ただし, $\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$, $\Phi_P(L) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_P L^{Ps}$, $\theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$, $\Theta_Q(L) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_Q L^{Qs}$ によって定義する. 時系列過程 (4.8) を P-ARFIMA(p, d, q) $\times$ (P, d_s, Q) と表記しよう. ただし, p, q は非周期的自己回帰 / 移動平均オーダー, P, Q は周期的自己回帰 / 移動平均オーダー, ϕ_i ($i = 1, \dots, p$), θ_j ($j = 1, \dots, q$), Φ_k ($k = 1, \dots, P$), Θ_l ($l = 1, \dots, Q$) は, それぞれ非周期的自己回帰 / 移動平均パラメータ, 周期的自己回帰 / 移動平均パラメータである. また, $\nabla_s = 1 - L^s$ は, 周期 s を有するタイムラグ作用素, d, d_s は, それぞれ原過程, 周期成分に関する実数分パラメータである.

ここで，簡単な周期的実数分時系列過程として，P-ARFIMA(0, d , 0) $\times$ (0, d_s , 0) 過程

$$\nabla^d \nabla_s^{d_s} x_t = \varepsilon_t \quad (4.9)$$

を取り上げる．確率過程 (4.9) は，周期的変動を考慮した実数分過程のプロトタイプであり，確率誤差過程 $\{\varepsilon_t\}_{t=0}^\infty$ は定常過程である．ここで，周期ラグ s が偶数の場合，式 (4.9) のラグ作用素は，

$$\begin{aligned} \nabla^d \nabla_s^{d_s} &= \nabla^{d+d_s} (1+L)^{d_s} \\ &\times \prod_{j=1}^{\frac{s}{2}-1} \left\{ 1 - 2 \cos\left(\frac{2\pi j}{s}\right) L + L^2 \right\}^{d_s} \end{aligned} \quad (4.10)$$

と展開できる（付録 1 参照）． $d + d_s < 1/2$ ， $d_s < 1/2$ が成立する時，確率過程 $\{x_t\}_{t=0}^\infty$ は定常である．また， $d + d_s < -1/2$ ， $d_s > -1/2$ が成立する時，確率過程 $\{x_t\}_{t=0}^\infty$ は反転可能である．この時，周期的実数分過程 (4.9) を，それと等価な無限次元移動平均過程

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad (4.11)$$

に変換できる．ただし， ψ_j ($j \geq 1$) は再帰的に

$$\psi_j = \frac{2}{j} \sum_{i=0}^{j-1} \sum_{l=1}^{1+\frac{s}{2}} D_l \cos[(j-i)\lambda_l] \psi_i \quad (4.12)$$

と表現できる（付録 2 参照）．ただし， $D_1 = (d + d_s)/2$ ， $D_{1+s/2} = d_s/2$ ， $D_i = d_s$ ($i = 2, \dots, s/2$)， $\lambda_1 = 0$ ， $\lambda_{1+s/2} = \pi$ ， $\lambda_l = 2\pi(l-1)/s$ ($l = 2, \dots, s/2$) であり，

$$\psi_j = \sum_{i=0}^j \psi_0(d, i) \psi_s(d_s, j-i) \quad (4.13)$$

が成立する．ここで，

$$\begin{aligned} \psi_0(d, i) &= \frac{\Gamma(i+d)}{\Gamma(d)\Gamma(i+1)} \\ \psi_s(d_s, l) &= \begin{cases} \psi_0(d_s, l/s) & l = 0, s, 2s, \dots \text{の時} \\ 0 & \text{それ以外の時} \end{cases} \end{aligned} \quad (4.14)$$

である（付録 3 参照）． $\psi_0(d, i)$ は，ARFIMA(0, d , 0) 過程の無限移動平均係数である．この周期的実数分過程は $|d + d_s| < 1/2$ ， $|d_s| < 1/2$ が成立する時のみ，定常的であり，かつ反転可能である．

4.3 P-STARFIMA モデル

以下では，P-ARFIMA モデルを空間に拡張した周期変動成分を持つ時空間実数分モデルを定式化する．すなわち，時刻 t ($t = 0, \dots$) と地点 k ($k = 1, \dots, K$) の時空間変数 $x_{k,t}$ が，

過去の時空間変数の実績値の加重線形モデルで表現されるような P-STARFIMA モデルを定式化する．時刻 t における時空間変数ベクトルを $K \times 1$ 次元列ベクトル $\mathbf{x}_t = (x_{1,t}, \dots, x_{K,t})'$ で表現する．記号「 $'$ 」は転置を表す．

$$\begin{aligned} & \phi_{p,\lambda}(L)\phi_{P,\lambda}(L^s)\nabla^d\nabla_s^{d_s}\mathbf{x}_t \\ &= \boldsymbol{\theta}_{q,\mathbf{r}}(L)\boldsymbol{\theta}_{Q,\mathbf{r}}(L^s)\boldsymbol{\varepsilon}_t \end{aligned} \quad (4.15)$$

式 (4.15) の各項は次のように定義される．

$$\begin{aligned} \phi_{p,\lambda}(L) &= \mathbf{I} - \sum_{j=1}^p \sum_{l=0}^{\lambda_j} \phi_{jl} \mathbf{W}_l L^j \\ \phi_{P,\lambda}(L^s) &= \mathbf{I} - \sum_{j=1}^P \sum_{l=0}^{\Lambda_j} \Phi_{jl} \mathbf{W}_l L^{js} \\ \boldsymbol{\theta}_{q,\mathbf{r}}(L) &= \mathbf{I} - \sum_{j=1}^q \sum_{l=0}^{r_j} \theta_{jl} \mathbf{W}_l L^j \\ \boldsymbol{\theta}_{Q,\mathbf{r}}(L^s) &= \mathbf{I} - \sum_{j=1}^Q \sum_{l=0}^{R_j} \Theta_{jl} \mathbf{W}_l L^{js} \end{aligned}$$

ここで， ϕ_{jl}, Φ_{jl} は時間ラグ j ，空間ラグ l の周期 / 非周期の自己回帰パラメータ， θ_{jl}, Θ_{jl} は時間ラグ j ，空間ラグ l における非周期 / 周期の移動平均パラメータ， p, P は非周期 / 周期の自己回帰オーダー， q, Q は非周期 / 周期の移動平均オーダー， $\lambda = \{\lambda_j(j = 1, \dots, p)\}$ ， $\Lambda = \{\Lambda_j(j = 1, \dots, P)\}$ は自己回帰項における非周期 / 周期の空間オーダーベクトル， $\mathbf{r} = \{r_j(j = 1, \dots, q)\}$ ， $\mathbf{R} = \{R_j(j = 1, \dots, Q)\}$ は移動平均項における非周期 / 周期の空間オーダーベクトル， $\mathbf{W}_l$ は l 番目の空間オーダーにおける $K \times K$ 次の空間構造行列（詳細は 6 . (2) 参照のこと）， d, d_s はそれぞれ非周期 / 周期の実数値と分次数， $\nabla^d, \nabla_s^{d_s}$ は非周期 / 周期階差オペレータであり， s は周期を表わす．また， $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{K,t})'$ は正規分布に従うランダム誤差列ベクトルであり，

$$\begin{aligned} E\{\boldsymbol{\varepsilon}_t\} &= \mathbf{0} \\ E\{\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-j}'\} &= \begin{cases} \sigma^2 \mathbf{I} & (j = 0) \\ 0 & (j \neq 0) \end{cases} \\ E\{\mathbf{X}_t \boldsymbol{\varepsilon}_{t-j}'\} &= \mathbf{0} \quad (j \geq 0) \end{aligned}$$

を満足する．本モデルには，合計 $3 + \sum_{j=1}^p \lambda_j + \sum_{j=1}^q r_j + \sum_{j=1}^P \Lambda_j + \sum_{j=1}^Q R_j$ 個のパラメータが含まれている．さらに，本モデルは無限次元移動平均過程

$$\mathbf{x}_t = \frac{\boldsymbol{\theta}_{q,\mathbf{r}}(L)\boldsymbol{\theta}_{Q,\mathbf{r}}(L^s)}{\phi_{p,\lambda}(L)\Phi_{P,\lambda}(L^s)}\boldsymbol{\varepsilon}_t \quad (4.16)$$

に変換可能であると考えられる．ここで，モデル (4.15) を $((p, \lambda), d, (q, \mathbf{r})) \times ((P, \Lambda), d_s, (Q, \mathbf{R}))$ 次元を有する周期的時空間自己回帰実数値と分移動平均 (Periodically Space-Time AutoRe-

gressive Fractionally Integrated Moving-Average : P-STARFIMA) モデルと呼び , P-STRIFIMA $((p, \lambda), d, (q, r)) \times ((P, \lambda), d_s, (Q, r))$ と略記する . 以上のモデル (4.15) は , Pfeifer and Deutsch²⁸⁾ による P-STARIMA モデルの和分パラメータ d, d_s を実数化した拡張になっている .

第5章 RS-P-STARFIMA モデル

5.1 レジームの定義

大気汚染濃度は、同国内に設けられた複数のモニタリング局（以下、地点と呼ぶ）で観測される．過去のヘイズ災害では、ヘイズ発生パターンが地点により多様に異なる³⁾．そこで、レジームを各地点ごとに個別に定義する．

いま、離散的時刻 t ($t = 0, \dots, T$) のレジーム状態を、離散的確率変数ベクトル $S_t = \{S_{k,t} : k = 1, \dots, K\}$ で表現する．レジーム状態 $S_{k,t} = m$ ($m = 1, \dots, M$) は、時刻 t の地点 k におけるレジームを表現している． $S_{k,t} = 1$ のときは通常期であり、 m の値が大きくなるほどヘイズの程度が重くなり、 $S_{k,t} = M$ の時に最も深刻なヘイズが発生していると考えよう．本来、レジーム状態 $S_{k,t}$ は、観測不可能な確率変数である．ここでは記述の便宜上、地点 k ごとに、レジーム状態の点列 $\{S_{k,t}\}_{t=0}^T$ が外生的に与えられていると考えよう．さらに、地点 k ($k = 1, \dots, K$) の時空間変数の観測値 $\bar{x}_{k,t}$ ($t = 0, \dots, T$) を、レジーム状態に着目して、 M 個の排他的クラス $\Omega_{k,m} = \{t | S_{k,t} = m, t = 0, \dots, T\}$ に分割する．ただし、記号「 $\bar{\cdot}$ 」は観測値であることを表す．その上で、地点 k の各レジームに属する観測値の期待値ベクトルを $\mu_k = (\mu_{k,1}, \dots, \mu_{k,M})$ と定義する．ただし、 $\mu_{k,m}$ は、レジーム m に属する時空間データ $\bar{x}_{k,t}$ ($t \in \Omega_{k,m}$) の条件付期待値 $\mu_{k,m} = \mu(S_{k,t} = m) = E[x_{k,t} | t \in \Omega_{k,m}]$ を意味する．また、レジームの標準偏差ベクトルを $\rho_k = (\rho_{k,1}, \dots, \rho_{k,M})$ と定義する．ただし、 $\rho_{k,m}$ は、レジーム m に属する時空間データ $\bar{x}_{k,t}$ ($t \in \Omega_{k,m}$) の条件付標準偏差 $\rho_{k,m} = \rho_k(S_{k,t} = m) = \sqrt{VAR[x_{k,t} - \mu_{k,m} | t \in \Omega_{k,m}]}$ である．

ここで、標準化操作を行った時空間変数（以下、標準化変数と呼ぶ）

$$z_{k,t}(S_{k,t}) = \frac{\bar{x}_{k,t} - \mu_k(S_{k,t})}{\rho_k(S_{k,t})} \quad (5.1)$$

を定義しよう．この時、地点 k のレジーム m に属する標準化変数 $z_{k,t}(S_{k,t})$ ($t \in \Omega_{k,m}$) は、確率密度関数 $g_{k,m}(z_{k,t})$ ($z_{k,t} \in (-\infty, \infty)$) に従うと仮定する．ただし、確率密度関数 $g_{k,m}$ は期待値 0 を中心とする対称的確率密度関数であり、 $g_{k,m}(-u) = g_{k,m}(u)$ が成立する．以上では、レジーム状態の点列 $\{S_{k,t}\}_{t=0}^T$ を所与と考えた．しかし、現実には、時空間データ $\bar{x}_{k,t}$ のみが観測可能であり、その観測値がどのレジーム状態の下で出現したのかは観測できない．また、各レジームにおけるデータの生成構造（期待値ベクトル μ_k 、および確率密度関数 $g_{k,m}$ も未知である）．このように RS モデルでは、観測値の点列に基づいて、レジーム状態とデータの生成構造を同時に推計することが必要となり、モデルの推計方法が複雑になる．RS モデルの推計方法は、改めて 5 (2) で言及する．

5.2 RS モデル

マレーシアの大気汚染物質濃度に関する各地点のレジームが、マルコフ過程に従って変化すると考えよう。ヘイズ災害の場合、スマトラ島における森林火災や南西モンスーン風等の外生的要因によりレジーム状態が変化する。そこで、RS モデルとして、レジーム状態の変化を記述するマルコフ推移確率が、外生変数の影響を受けて変化するような TRS(Time-varying Regime Switching) モデル³²⁾を採用する。レジーム状態 $S_{k,t}$ は、観測不可能な確率変数である。いま、ある地点のレジーム変数 $S_{k,t}$ の時間的推移過程が、斉次マルコフ過程

$$P_k(S_{k,t+1} = n | S_{k,t} = m; \psi_t) = p_{mn}^{k,t}(\bar{v}_{t-\Delta}, \beta_k^m) \quad (5.2)$$

$$(n = 1, \dots, M-1; m = 1, \dots, M) \\ \sum_{n=1}^M p_{mn}^{k,t}(\bar{v}_{t-\Delta}, \beta_k^m) = 1 \quad (5.3)$$

に従うと仮定する^{29), 48)}。 $\beta_k^m = (\beta_k^{m1}, \dots, \beta_k^{mM-1})$ であり、 β_k^{mn} は、地点 k でレジーム m からレジーム n に推移する確率を支配するパラメータである。 $\bar{v}_{t-\Delta}$ はレジーム推移確率に影響を及ぼす外生変数の観測値であり、 Δ は外生変数の先行ラグを表す。たとえば、 $\bar{v}_{t-1}$ は、時刻 t より 1 期前の外生変数（先行指標）の観測値を表す。また、 ψ_t は、時刻 t までに入手可能な観測値集合（情報集合）であり、

$$\psi_t = \{(\bar{v}_{j-\Delta}, \bar{x}_j), j = 0, \dots, t\} \quad (5.4)$$

と定義する。外生変数 $v_{t-\Delta}$ のデータは、初期時刻 $t = 0$ より前の時刻 $t = -\Delta$ より入手可能であると考えよう。情報集合は、時間とともに情報が蓄積されるフィルトレーション過程である。ここで、レジーム推移確率を

$$P_k(S_{k,t+1} = n | S_{k,t} = m; \psi_t) = \frac{\exp(\beta^{mn} \bar{v}_{t-\Delta})}{1 + \sum_{l=1}^{M-1} \exp(\beta^{ml} \bar{v}_{t-\Delta})} \quad (5.5)$$

$$P_k(S_{k,t} = M | S_{k,t-1} = m; \psi_t) = 1 - \sum_{n=1}^{M-1} P_k(S_{k,t} = n | S_{k,t-1} = m; \psi_t) \quad (5.6)$$

と特定化する。マルコフ推移確率 (5.5), (5.6) を用いて、各地点 k におけるレジーム状態の点列 $\{S_{k,t}\}_{t=0}^T$ を記述しよう。ここで、2 つの条件付確率 $P_k(S_{k,t} = m | \psi_{t-1})$, $P_k(S_{k,t} = m | \psi_t)$ を定義する。前者は、時刻 $t-1$ の情報集合 ψ_{t-1} に基づいて、時刻 t のレジーム状態を予測する確率（レジーム予測確率と呼ぶ）を表す。一方、後者は、情報集合 ψ_{t-1} に、時刻 t の観測値 $\bar{v}_{t-\Delta}$, $\bar{x}_t$ を加えた情報集合 ψ_t を用いて、時刻 t のレジーム状態を推定する確率（レジーム生起確率）を表している。

Hamilton-Filtering 法⁴⁸⁾を用いて、情報集合の点列 $\{\psi_t\}_{t=0}^T$ から、レジーム状態 $\{S_{k,t}\}_{t=0}^T$ を生成する過程を記述する。時刻 $t-1$ におけるレジーム生起確率 $P_k(S_{k,t-1} = m|\psi_{t-1})$ ($m = 1, \dots, M$) を与件とする。レジーム予測確率 $P_k(S_{k,t} = n|\psi_{t-1})$ は、レジーム生起確率 $P_k(S_{k,t-1} = m|\psi_{t-1})$ とマルコフ推移確率 $P_k(S_{k,t} = n|S_{k,t-1} = m; \psi_{t-1})$ を用いて、

$$\begin{aligned} P_k(S_{k,t} = n|\psi_{t-1}) &= \sum_{m=1}^M P_k(S_{k,t} = n, S_{k,t-1} = m|\psi_{t-1}) \\ &= \sum_{m=1}^M P_k(S_{k,t-1} = m|\psi_{t-1}) \\ &\quad P_k(S_{k,t} = n|S_{k,t-1} = m, \psi_{t-1}) \end{aligned} \quad (5.7)$$

と表せる。つぎに、情報集合 ψ_{t-1} の下における時空間変数 $x_{k,t}$ の条件付確率密度関数を $f_k(x_{k,t}|\psi_{t-1})$ とする。この時、時刻 t のレジーム生起確率 $P_k(S_{k,t} = n|\psi_t)$ は

$$\begin{aligned} P_k(S_{k,t} = n|\psi_t) &= P_k(S_{k,t} = n|x_{k,t}, \psi_{t-1}) \\ &= \frac{f_k(x_{k,t}, S_{k,t} = n|\psi_{t-1})}{f_k(x_t|\psi_{t-1})} \\ &= \frac{f_k(x_{k,t}|S_{k,t} = n, \psi_{t-1})P_k(S_{k,t} = n|\psi_{t-1})}{\sum_{m=1}^M f_k(x_{k,t}|S_{k,t} = m, \psi_{t-1})P_k(S_{k,t} = m|\psi_{t-1})} \\ &= \frac{g_{k,n}(z_{k,t}(n))P_k(S_{k,t} = n|\psi_{t-1})}{\sum_{m=1}^M g_{k,m}(z_{k,t}(m))P_k(S_{k,t} = m|\psi_{t-1})} \end{aligned} \quad (5.8)$$

と表せる。ここで、初期時刻 $t = 0$ のレジーム生起確率 $P(S_{k,0} = m|\psi_0)$ を与えよう。初期値の与え方に関しては、のちに 5 (2) で議論する。この時、初期レジーム生起確率と情報集合が順次得られれば、式 (5.7), (5.8) を用いて、レジーム推移確率を逐次求めることができる。さらに、各時刻のレジーム状態として、レジーム生起確率をもっとも大きくなるような状態 $S_{k,t}^*$

$$S_{k,t}^* = \arg \max_{n \in \{1, \dots, M\}} \{P(S_{k,t} = n|\psi_t)\} \quad (5.9)$$

を採用しよう。以上の操作をすべての地点 k ($k = 1, \dots, K$) に対して実施することにより、情報集合の点列 $\{\psi_t\}_{t=0}^T$ に対して、レジーム状態 $\{S_t^*\}_{t=0}^T$ を生成できる。ただし、 $S_t^* = (S_{1,t}^*, \dots, S_{K,t}^*)$ である。

5.3 RS-P-STARFIMA モデル

RS モデルと P-STARFIMA モデルを統合した RS-P-STARFIMA モデルを定式化しよう。ここで、時刻 t ($t = 0, 1, 2, \dots$) における K 個の地点の大気汚染物質濃度を時空間変数ベクトル $x_t = (x_{1,t}, \dots, x_{K,t})'$ で表現する。また、時空間変数の期待値ベクトルを $\mu(S_t^*) = (\mu_1(S_{1,t}^*), \dots, \mu_K(S_{K,t}^*))$ 、標準偏差ベクトルを $\rho(S_t^*) = (\rho_1(S_{1,t}^*), \dots, \rho_K(S_{K,t}^*))$ と表す。さ

らに，式 (5.1) で定義される標準化変数ベクトルを $z_t(S_t^*) = \{z_{1,t}(S_{1,t}^*), \dots, z_{K,t}(S_{K,t}^*)\}$ と記述しよう．この時，レジーム状態に関する時空間系列 $\{S_t^*\}_{t=0}^\infty$ に対して，レジーム変化を考慮した時空間変数の変化過程を RS-P-STARFIMA モデル

$$\begin{aligned} & \phi_{p,\lambda}(L)\phi_{P,\lambda}(L^s)\nabla^d\nabla_s^{d_s}z_t(S_t^*) \\ & = \theta_{q,m}(L)\theta_{Q,m}(L^s)\varepsilon_t \end{aligned} \tag{5.10}$$

を用いて表現することが可能となる．

第6章 モデルの推計方法

6.1 RS-P-STARFIMA モデルの推計手順

RS-P-STARFIMA モデルの推計方法として、長期記憶性とレジーム変化を同時に扱い、かつレジームを内生的に推計するような同時推計法を開発することが理想的である。しかし、筆者らの知る限り、現在のところ、このような同時推計法は提案されていない。同時推計法を開発することが困難であることより、本研究では、1) RS モデルのパラメータ推計と各地点・各時刻のレジーム状態の推定、2) レジーム状態の推定結果を与件とした RS-P-STARFIMA モデルのパラメータ推計という段階的推計方法を提案する。

6.2 RS モデルのセミパラメトリック推計

RS モデルにおけるレジーム状態変数 S_t 、レジーム推移確率の支配パラメータ β_k^m 、レジームの期待値 μ_k と標準偏差 ρ_k は、いずれも観測不可能な潜在変数であるため、通常の方法でパラメータを推計できない。本研究では、観測できない隠れたパラメータの最尤推計値を求めるための汎用的手法である EM アルゴリズム⁴⁹⁾を用いる。標準化変数 z_t は、長期記憶性を持つ確率過程から発生しているため、その確率密度関数をパラメトリックな関数として表現することが困難である。そこで、RS モデルの推計には、Bordes 等⁵⁰⁾によって提案されたセミパラメトリックな EM アルゴリズムを用いる。後述するように、この方法は混合分布の確率密度関数をカーネル密度関数を用いて近似するところに特徴がある。なお、セミパラメトリック EM アルゴリズムの有効性はモンテカルロシミュレーションと実証分析によって確認されているが、筆者らの知る限り、パラメータの推計量の極限分布が理論的に導出されておらず、検定統計量を求めることができない。

情報集合 ψ_{t-1} の下における時空間変数 $x_{k,t}$ の条件付確率密度は $f_k(x_{k,t}|\psi_{t-1})$ と表せる。この時、情報集合を与件とした条件付対数尤度関数を、

$$\begin{aligned}\ln \mathcal{L}(\beta) &= \ln \left\{ \prod_{t=1}^T \prod_{k=1}^K f_k(x_{k,t}|\psi_{t-1}) \right\} \\ &= \sum_{t=1}^T \sum_{k=1}^K \ln \left\{ f_k(x_{k,t}|\psi_{t-1}) \right\} \\ &= \sum_{t=1}^T \sum_{k=1}^K \ln \left\{ \sum_{m=1}^M f_k(x_{k,t}, S_{k,t} = m|\psi_{t-1}) \right\} \\ &= \sum_{t=1}^T \sum_{k=1}^K \ln \left\{ \sum_{m=1}^M f_k(x_{k,t}|S_{k,t} = m, \psi_{t-1}) \right\}\end{aligned}$$

$$\begin{aligned}
& P_k(S_{k,t} = m | \psi_{t-1}, \beta_k) \} \\
& = \sum_{t=1}^T \sum_{k=1}^K \ln \left\{ \sum_{m=1}^M g_{k,m}(z_{k,t}(m)) \right. \\
& \quad \left. P_k(S_{k,t} = m | \psi_{t-1}, \beta_k) \right\}
\end{aligned} \tag{6.1}$$

と定義することができる．上式においては，条件付対数尤度関数がマルコフ推移確率 (5.7) に含まれる支配パラメータ $\beta = (\beta_1, \dots, \beta_K)$ の関数になることを明示的に表現している．しかし，対数尤度関数の中に，未知変数であるレジーム状態変数 S_t ，レジームの期待値 μ_k ，標準偏差 ρ_k ，未知関数である確率密度関数 $g_{k,m}(z_{k,t})$ ，およびレジーム予測確率 $P(S_{k,t} = m | \psi_{t-1}, \beta_k)$ が同時に含まれるため，対数尤度関数値を容易に求めることができないという問題がある．

セミパラメトリック法を用いた EM アルゴリズムでは，直接探索法により条件付対数尤度 (6.1) を最大にするようなレジーム推移確率の支配パラメータ β_k ($k = 1, \dots, K$) を求める．図 - 6.1 に RS モデルの推計アルゴリズムの概要を示している．同図に示すように，直接探索法の各ステップ ι において，外生的に与えたレジーム推移確率の支配パラメータ $\beta_k^{(\iota)}$ のそれぞれに対して対数尤度関数値 $\ln \mathcal{L}(\beta^{(\iota)})$ を求める．その中で最大の対数尤度を与えるような支配パラメータ値を最尤推計値 $\hat{\beta}_k$ として採用する．いま，直接探索法のステップが ι 回繰り返され，支配パラメータの候補値 $\beta_k^{(\iota)}$ に対して対数尤度関数値を求める段階を考えよう．これは，図 - 6.1 に示すアルゴリズムの中で，矩形で囲まれた範囲に該当する．対数尤度関数値 $\ln \mathcal{L}(\beta^{(\iota)})$ は以下の手順で求めることができる．

1) ステップ 1 (初期値の設定)

初期分布 $g_{k,m}^{(0,\iota)}$ ，初期レジーム平均ベクトル $\mu_k^{(0,\iota)}$ と標準偏差ベクトル $\rho_k^{(0,\iota)}$ ，初期時刻のレジーム生起確率 $P_k^{(0,\iota)}(S_{k,0} = m | \psi_0)$ ，各時刻のレジーム生起確率 $P_k^{(0,\iota)}(S_{k,t} = m | \psi_t)$ ($t = 0, \dots, T$) を設定する．以下，記述の簡単化のため，直接探索ステップを表す添字 ι を省略する．任意の初期レジーム平均ベクトル $\mu_k^{(0)}$ を与える．初期レジーム平均ベクトル $\mu_k^{(0)}$ によってレジーム状態の点列 $\{S_{k,t}^{(0)}\}_{t=0}^T$ を次式から得る．

$$S_{k,t}^{(0)} = \arg \min_{m \in 1, \dots, M} \{ |\bar{x}_{k,t} - \mu_{k,m}^{(0)}| \}$$

ステップ 3，4 において後述する方法を用いてレジーム状態の点列から初期標準偏差ベクトル $\rho_k^{(0)}$ ，初期分布 $g_{k,m}^{(0)}$ を得る．情報集合の点列 $\{\psi_t\}_{t=0}^T$ を与件とする． $P_k^{(0)}(S_{k,0} = m | \psi_0)$ ($m = 1, \dots, M$) として，連立方程式

$$\begin{aligned}
& \sum_{m=1}^M \pi_{k,m}^{(0)} = 1 \\
& \pi_k^{(0)} = \pi_k^{(0)} P_k^{(0)}(n | m, \psi_0)
\end{aligned}$$

を満足する定常確率 $\pi_k^{(0)} = (\pi_{k,1}^{(0)}, \dots, \pi_{k,M}^{(0)})$ を用いる．ここに， $P_k^{(0)}(n | m, \psi_0)$ は，マルコフ推移確率 $P_k^{(0)}(S_{k,1} = n | S_{k,0} = m; \psi_{t-\Delta})$ を (m, n) 要素とする $(M \times M)$ 次行列である．

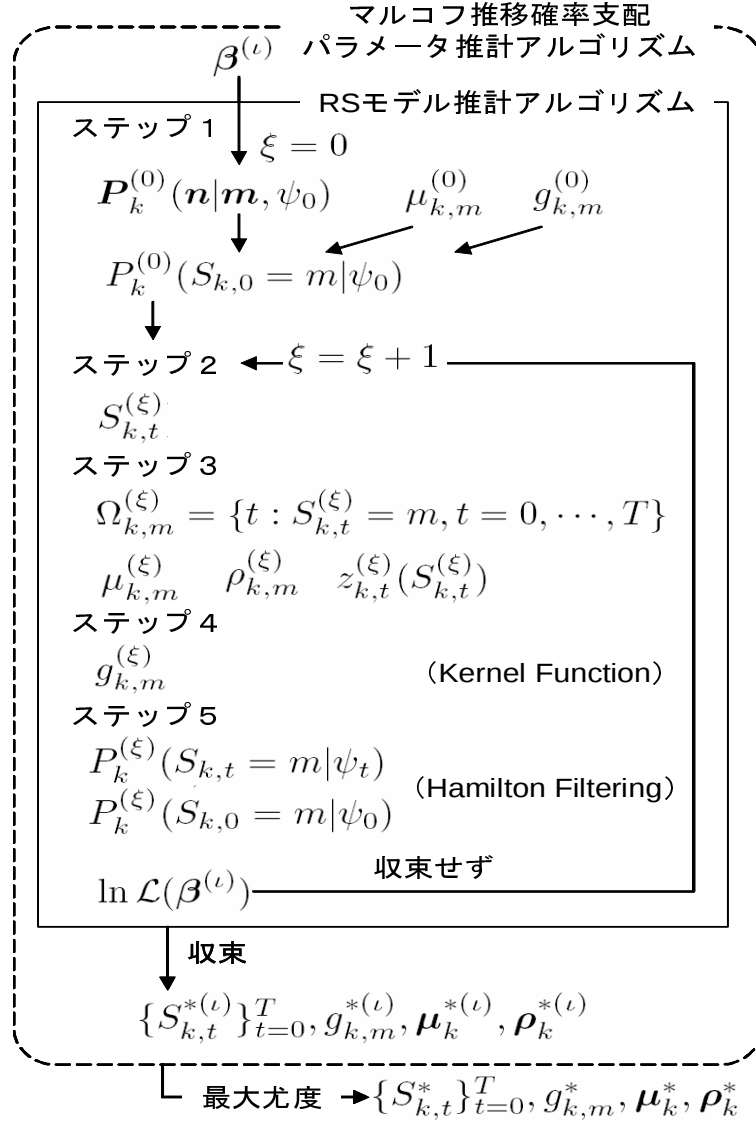


図-6.1 モデルの推計手順

確率密度関数 $g_{k,m}^{(0)}$, および初期時刻レジーム生起確率 $P_k^{(0)}(S_{k,0} = m|\psi_0)$ が得られれば , Hamilton-Filtering 法によって , 各時刻のレジーム生起確率 $P_k^{(0)}(S_{k,t} = m|\psi_t)$ ($t = 0, \dots, T$) を求めることができる . また , セミパラメトリック EM アルゴリズムの反復回数を $\xi = 1$ に設定する .

2) ステップ 2 (レジーム状態の設定)

レジーム状態の点列 $\{S_{k,t}^{(\xi)}\}_{t=0}^T$ を次式で定義する .

$$S_{k,t}^{(\xi)} = \arg \max_{m \in 1, \dots, M} \{P_k^{(\xi-1)}(S_{k,t} = m|\psi_t)\} \quad (6.2)$$

3) ステップ 3 (標準化変数の設定)

レジーム状態の点列 $\{S_{k,t}^{(\xi)}\}_{t=0}^T$ に対して , $T_{k,m}^{(\xi)} = ||\{t : t \in \Omega_{k,m}^{(\xi)}, t = 0, \dots, T\}||$ を定義

する．ただし， $\Omega_{k,m}^{(\xi)} = \{t : S_{k,t}^{(\xi)} = m, t = 0, \dots, T\}$ であり， $T_{k,m}^{(\xi)}$ はレジーム集合 $\Omega_{k,m}^{(\xi)}$ に属するサンプル $\bar{x}_{k,t}$ ($t = 0, \dots, T$) の総数を意味する． $\|A\|$ は集合 A の要素数を意味する．この時，レジームの期待値 $\mu_k^{(\xi)}$ ，標準偏差 $\rho_k^{(\xi)}$ の各要素を，

$$\mu_{k,m}^{(\xi)} = \frac{1}{T_{k,m}^{(\xi)}} \sum_{t=0}^T \delta_{k,t}^{m(\xi)} \bar{x}_{k,t} \quad (6.3)$$

$$\rho_{k,m}^{(\xi)} = \sqrt{\frac{1}{T_{k,m}^{(\xi)}} \sum_{t=0}^T \delta_{k,t}^{m(\xi)} (\bar{x}_{k,t} - \mu_{k,m}^{(\xi)})^2} \quad (6.4)$$

と定義する．ただし， $\delta_{k,t}^{m(\xi)}$ は $\bar{x}_{k,t}$ が，レジーム集合 $\Omega_{k,m}^{(\xi)}$ に属するサンプルである時に 1 を，そうでない時に 0 をとるダミー変数である．この時，レジーム状態の点列 $\{S_t^{(\xi)}\}_{t=0}^T$ に対して，標準化変数を次式で定義する．

$$z_{k,t}^{(\xi)}(S_{k,t}^{(\xi)}) = \frac{\bar{x}_{k,t} - \mu_k^{(\xi)}(S_{k,t}^{(\xi)})}{\rho_k^{(\xi)}(S_{k,t}^{(\xi)})}$$

4) ステップ 4 (確率密度関数の推計)

確率密度関数 $g_{k,m}^{(\xi)}(z_{k,t})$ を，バンド幅を $h_{k,m}^{(\xi)}$ とするカーネル関数 $K(\zeta)$ を用いて次式のように近似する⁵¹⁾．

$$\hat{g}_{k,m}^{(\xi)}(u) = \frac{1}{T_{k,m}^{(\xi)} h_{k,m}^{(\xi)}} \sum_{t=0}^T \delta_{k,t}^{m(\xi)} K\left(\frac{u - z_{k,t}^{(\xi)}}{h_{k,m}^{(\xi)}}\right) \quad (6.5)$$

ここで，カーネル密度関数 $K(\zeta)$ として，

$$K(\zeta) = \frac{1}{2} U(|\zeta| \leq 1) \quad (6.6)$$

を用いる．ただし， $U(|\zeta| \leq 1)$ は， $|\zeta| \leq 1$ の時に 1 を，そうでない時は 0 の値をとる一様分布である．一様分布に用いるバンド幅 $h_{k,m}^{(\xi)}$ として，Silverman⁵²⁾ の公式

$$h_{k,m}^{(\xi)} = \left(\frac{264}{729\pi(T_{k,m}^{(\xi)})^2} \right)^{1/10} \rho_{k,m}^{(\xi)} \quad (6.7)$$

を用いる．確率密度関数の対称性より次式を得る．

$$g_{k,m}^{(\xi)}(u) = \frac{\hat{g}_{k,m}^{(\xi)}(u) + \hat{g}_{k,m}^{(\xi)}(-u)}{2} \quad (6.8)$$

5) ステップ 5 (レジーム予測確率の推計)

対数尤度関数値 (6.1) を求めるには，レジーム予測確率 $P_k(S_{k,t} = m | \psi_{t-1}, \beta_k)$ が必要となる．4 (2) で言及した Hamilton-Filtering 法によって，レジーム予測確率を求める．レジーム予測確率が十分に収束した場合，これまでのステップで求めた $\{S_{k,t}^{(\xi)}\}_{t=0}^T, g_{k,m}^{(\xi)}, \mu_k^{(\xi)}, \rho_k^{(\xi)}$ を $\{S_{k,t}^{*(\iota)}\}_{t=0}^T, g_{k,m}^{*(\iota)}, \mu_k^{*(\iota)}, \rho_k^{*(\iota)}$ と表す．以上の推定結果を用いて，式 (6.1) により対数尤度

関数値 $\ln \mathcal{L}(\beta^{(\iota)})$ を計算する．収束が不十分な場合，反復回数を $\xi = \xi + 1$ とし，ステップ 2 へ戻る．

なお，以上の手順により，設定した $\beta^{(\iota)}$ に対して，条件付対数尤度関数 (6.1) が求まる．グリッドサーチにより，対数尤度関数値を最大とするような最尤推計量 $\hat{\beta}$ を求めることができる．最尤推定量 $\hat{\beta}$ に対して，最終的に収束した $\{S_{k,t}^{*(\iota)}\}_{t=0}^T, g_{k,m}^{*(\iota)}, \mu_k^{*(\iota)}, \rho_k^{*(\iota)}$ を $\{S_{k,t}^*\}_{t=0}^T, g_{k,m}^*, \mu_k^*, \rho_k^*$ と表す．

6.3 RS-P-STARFIMA モデルの推計

第 1 段階の RS モデルの推計により，レジーム状態の点列 $\{S_t^*\}_{t=0}^T$ と，標準化変数の点列 $\{z_t\}_{t=0}^T$ が求まる．第 2 段階では，P-STARFIMA モデルの実数と分パラメータ d, d_s ，自己帰帰・移動平均パラメータ $\eta = \{\phi_{10}, \dots, \phi_{p\lambda_p}, \theta_{10}, \dots, \theta_{qr_q}, \Phi_{10}, \dots, \Phi_{P,\Lambda_P}, \Theta_{10}, \dots, \Theta_{Q,R_Q}\}$ ，分散 σ^2 を推計する．以下では，Chung⁵³⁾の方法に基づいて，レジーム状態の点列 $\{S_t^*\}_{t=0}^T$ を与件とした条件付対数尤度関数を定式化する．

いま，モデル (5.10) で表現される確率過程 $\{z_t\}_{t=0}^\infty$ に関して以下の条件が成立すると仮定しよう．

条件 (A) $t \leq 0$ に対して， $z_t = 0$ が成立する．さらに，確率誤差項の系列 $\{\varepsilon_t\}_{t=1}^\infty$ が，同一で独立な正規分布 $N(0, \sigma^2)$ に従う．ただし， $E[\varepsilon_t^4] < \infty$ が成立する．

条件 (B) 周期性パラメータ s と期待値ベクトル μ_k ，分散ベクトル ρ_k を既知とする．

条件 (C) 周期 / 非周期実数と分パラメータは，条件 $(d + d_s, d_s) \in D_i^s \times D_j^s = D_{i,j}^s, (i, j = 1, \dots, 4)$ を満たす．ただし， $D_1^s = [\alpha, 1/2 - \alpha], D_2^s = [-\alpha, \alpha], D_3^s = [-1/2 + \alpha, -\alpha], D_4^s = [-1/2 - \alpha, -1/2 + \alpha]$ ，および $\alpha \in (0, 1/4)$ である．すなわち，実数と分過程が定常，かつ反転可能である．

条件 (D) パラメータ η の推計値が，AR, MA 作用素 $\phi_{p,\lambda}(L), \phi_{P,\lambda}(L^s), \theta_{q,r}(L)$ および $\theta_{Q,r}(L^s)$ の定常性，反転可能性条件を保証するようなコンパクト集合 D_η に含まれる．さらに，分散 σ^2 は空間 $\mathbb{R}^+$ 内のコンパクト集合に含まれる．

ここで，上記の条件 (C), (D) が成立するとき，モデル (5.10) は反転可能であり，ラグ多項式を

$$\nabla^d \nabla_s^{d_s} \frac{\phi_{p,\lambda}(L) \phi_{P,\lambda}(L^s)}{\theta_{q,r}(L) \theta_{Q,r}(L^s)} = \sum_{j=0}^{\infty} \pi_j(\delta) L^j \quad (6.9)$$

の形に展開できることに着目しよう．ただし， $\pi_j(\delta)$ は，パラメータ $\delta = (d, d_s, \eta)$ に関する多項式である．この時，モデル (5.10) の誤差項 $\varepsilon_{k,t}$ を， δ の関数として

$$\varepsilon_{k,t}(\delta) = \sum_{j=0}^{t-1} \pi_j(\delta) z_{k,t-j} (S_{k,t-j}^*) \quad (6.10)$$

と表すことができる．ただし，条件 (A) より， $l \geq t - 1$ 次以上の標準化変数に関して $z_{k,t-l} = 0$ が成立するため，式 (6.10) の多項式は有限次数 $l = t - 1$ までの和として表されていることを断っておく．この時，条件付対数尤度関数を

$$\ln \tilde{\mathcal{L}}(\delta, \sigma^2) = -\frac{TK}{2} \ln 2\pi - \frac{TK}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T \sum_{k=1}^K \{\varepsilon_{k,t}(\delta)\}^2 \quad (6.11)$$

と定式化することができる．上式を最大にするような σ^2 の最尤推計量は，任意の δ に対して，

$$\hat{\sigma}^2(\delta) = \sum_{t=1}^T \sum_{k=1}^K \frac{\{\varepsilon_{k,t}(\delta)\}^2}{TK} \quad (6.12)$$

で与えられる¹⁰⁾．したがって，条件付対数尤度関数 (6.11) の最大化問題は，

$$Q(\delta) = \sum_{t=1}^T \sum_{k=1}^K \{\varepsilon_{k,t}(\delta)\}^2 \quad (6.13)$$

の最小化問題に帰着される．この時，レジーム状態の点列 $\{S_t^*\}_{t=0}^T$ を与件とするパラメータ δ の CSS 推計量は，式 (6.13) を最小にするようなパラメータ $\check{\delta}$ として与えられる．また，分散 σ の CSS 推計量は，式 (6.12) より $\check{\sigma}^2 = \hat{\sigma}^2(\check{\delta})$ と表される．なお，条件 (C) を満たす場合，以上で求めた CSS 推計量は，一致性を満足する²⁵⁾．

なお，確率過程が長期記憶性を示す場合，自己相関関数や偏自己相関関数の減衰は非常に緩やかであるため，これらに基づいて視覚的に AR / MA 項の次数を決定することはできない．そこで，候補となるモデルを複数推計して，その中から AIC が最大のモデルを選択する．モデルのパラメータ数を K とすると，AIC は条件付対数尤度関数 $\ln \tilde{\mathcal{L}}(\check{\delta}, \check{\sigma}^2)$ を用いて，

$$AIC = -2\{\ln \tilde{\mathcal{L}}(\check{\delta}, \check{\sigma}^2) - K\} \quad (6.14)$$

と定義される⁵⁴⁾．ただし， K はパラメータ数であり， $K = 3 + p + q + P + Q + \sum_{j=1}^p \lambda_j + \sum_{j=1}^q r_j + \sum_{j=1}^P \Lambda_j + \sum_{j=1}^Q R_j$ である．

Site_ID	Site_Location	Site_Area	Comms_Date	Latitude	Longitude
CA0001	Sek. Men. Pasir Gudang 2, Pasir Gudang	Comprehensive	19950110	N01°470	E103°894
CA0003	Inst. Latihan Perindustrian, Perai	Comprehensive	19960101	N05°371	E100°389
CA0005	Jab. Bekalan Air Daerah Gombak	Comprehensive	19960401	N03°262	E101°652
CA0006	Bukit Rambai, Melaka	Comprehensive	19960501	N02°265	E102°176
CA0008	Sek. Men. Jalan Tasek, Ipoh	Comprehensive	19960701	N04°630	E101°116
CA0009	Sek. Keb. Seberang Jaya II, Perai	Comprehensive	19960901	N05°398	E100°403
CA0010	Tmn. Semarak (Phase II), Nilai	Comprehensive	19961001	N02°821	E101°815
CA0011	Sek. Men. (P) Raja Zarina, Kelang	Residential	19960112	N03°010	E101°408
CA0016	Sek. Keb. Seri Petaling, Petaling Jaya	Industrial	19960112	N03°110	E101°705
CA0017	Sek. Keb. Bakar Arang, Sungai Petani	Residential	19970101	N05°631	E100°470
CA0019	Sek. Men. Vok. Perdagangan, Johor Bahru	Residential	19961201	N01°497	E103°727
CA0020	Sek. Keb. Air Puteh, Taiping	Residential	19970101	N04°899	E100°680
CA0038	Universiti Sains Malaysia, Pulau Pinang	Residential	19980111	N05°359	E100°298
CA0046	Sek. Men. Pegoh, Ipoh 2, Perak	Residential	19990112	N04°553	E101°081
CA0054	Sek. Men. Keb. Seri Permaisuri, Cheras	Residential	20040218	N03°106	E101°718

図-6.2 CA 局の概要

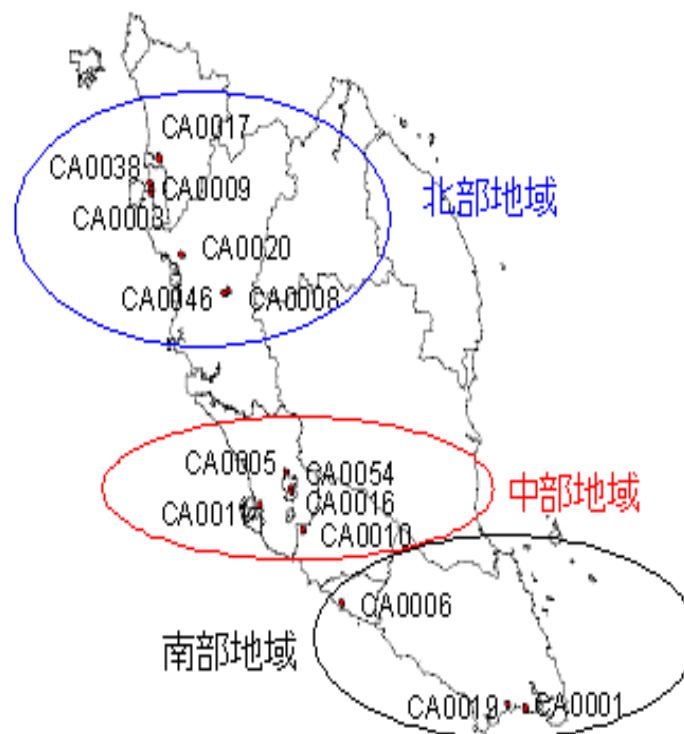


図-6.3 CA 局の設置位置

第7章 適用事例

7.1 データベース

7.1.1 ホットスポット数

本研究で用いる森林火災データは、森林火災予防プロジェクトの web ページに掲載されている、スマトラ島、マレー半島、およびボルネオ島を含む範囲を撮影した衛星写真データである。なお対象地域では1日1回の衛星写真が撮影されているホットスポットデータは衛星写真の画像解析によって作成され、発生地点の緯度経度が記録される。分析対象地域はマレーシア半島の煙害の主要な原因となっている、スマトラ島東岸の Riau 州, Jambi 州, Lampong 州で発生した森林火災とした。分析対象期間は、2006 年 6 月 26 日から 10 月 19 日までの 87 日間とした。なお 6 月 26 日から 6 月 25 日, 10 月 20 から 11 月 2 日までホットスポットは発生していない。集計の結果 3 州における対象期間中のホットスポット発生回数は 22002 回であった

7.1.2 PM10 濃度

1995 年、マレーシアの大気・水質プログラムが民営化された。20 年間のコンセッション契約により、Alam Sekitar Malaysia (ASMA) が、マレーシア全土に 50ヶ所の連続大気汚染モニタリング (Continuous Air quality monitoring: 以下、CA 局と略す) ネットワークの建設、運営・維持を請負った⁵⁵⁾。CA 局には大気分析装置と各種センサーが設置され、一酸化炭素、酸化窒素、粒子状物質、炭化水素、二酸化硫黄、光化学オキシダント等の大気中濃度が連続測定されている。これらのセンサーで測定したデータは、ASMA のデータ収集システムで一括管理されており、これらのデータに基づいてマレーシア環境省が定める総合大気汚染指数 (Air Pollution Index: API と略す) 等が算出される⁵⁶⁾。マレーシアの人口、経済活動は、マレー半島の西海岸地域に集中している。この地域では、南西モンスーン風が卓越する乾期に、スマトラ島の森林火災による煤煙が短期間のうちに到達するヘイズ被害が多発している。本研究では、マレー半島の西海岸地域に設置された 15 カ所の CA 局における観測結果を用いて、マレー半島におけるヘイズ災害の発生状況を分析する。各 CA 局の概要を図 - 6.2 に示す。また、各 CA 局の設置場所を図 - 6.3 に示す。

ヘイズ災害では一酸化炭素、窒素酸化物、粒子状物質等が問題となるが、本研究では粒子状物質 (PM10) に着目する。PM10 とは、大気中に浮遊する粒径 10 マイクロメートル以下の粒子状物質を意味する。PM10 は 1997 年、2005 年のヘイズ災害を引き起こした主

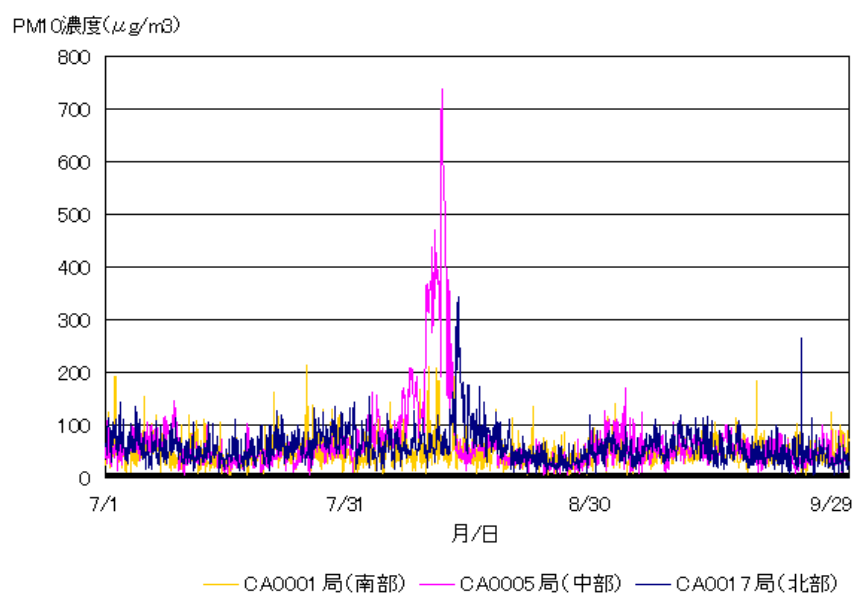


図-7.1 代表的 CM 局における観測結果

要物質であり，API の中でも主要な指数に指定されている．また，PM10 は大気中に滞留し，ヘイズ災害を長期化している原因となっている．分析にあたっては，2006 年 1 月から 9 月に至る 9 ヶ月にわたり 1 時間ごとに観測された PM10 の観測データを用いる．分析対象としたデータベースには，各観測局ごとに 6,552 個（15 観測局全体として合計 98,280 個）の観測データが含まれる．図 - 7.1 は，ヘイズ発生期間を含む 2005 年 7 月から 9 月までの，代表的な CA 局設置地点における PM10 の時間観測値の変動パターンを示している．同図に示すように，マレー半島の中中部・北部地域を中心としてヘイズ災害が発生している．一方，南部地域では，中部地域に近い CA0006 を除いて，他の CA 局のヘイズ発生は小規模である．

7.2 duration モデルを用いた森林火災モデルによるホットスポット数の推計

本研究で開発する森林火災発生モデルの目的は，火災発生地点やその時間的・空間的な拡大の様子を正確に再現することではなく，ホットスポット発生地点周辺の空間特性を考慮した予測モデルを作成することである．そこで以下の分析では対象地域を 30 分割（1 辺約 56km）のメッシュ単位に編成することにより，局所的な特性を安定的に抽出した結果，対象地域のメッシュ数は 100 となった．また各メッシュで発生する火災発生および，鎮火イベントの閾値は，各メッシュ内のホットスポットの有無としている．また，対象地域の

局所的な特性を表す空間ハザード領域 s は、火災の発生しやすいメッシュから発生しにくいメッシュまでを満遍なく含むようにするため、対象期間内のホットスポット発生回数が最大（1 番目）のメッシュから 1 1 番目，2 1 番目，… 9 1 番目までの 10 メッシュとした。また各メッシュの近傍 D_1 は、4 辺を共有する 4 メッシュに加えて 4 隅を共有する対角線方向の 4 メッシュを含む、8 メッシュと定義した。さらに、ハザード関数の分布関数としてはワイブル分布を仮定し、説明変数として、前回出火基幹、既往鎮火回数、さらに 10 分割した地域の固有効果を説明変数としてパラメータの推計を行った。

表－1，表－2 にそれぞれ森林火災発生モデルと森林火災継続モデルの推計結果を示す。形状パラメータ、およびペナルティー、分散を除くその他のパラメータは、イベントのハザード関数に含まれるパラメータであり、正の推計値が得られた場合はイベントが起こりやすく、負の推計値が得られた場合は、イベントが起こりにくいことを表している。たとえば森林火災発生モデルの既往鎮火期間の推計値は正であるので、既往鎮火期間が長いほど火災が起こりやすいことを表している。森林火災モデルの推計結果に着目すると推計値は優位ではないものの、地点別に推計値がばらつく傾向が確認できる。また森林火災継続モデルにおいても、同様の傾向がみられる。両モデルの平均尤度を比較すると、モデルの当てはまりは森林火災継続モデルのほうが良好であることが分かる。

次に得られたモデルに基づいて 3 章で説明した手順にしたかってホットスポット数の予測を行った。表－7.2 は 24 時間後のホットスポット数の予測値と実測値を表している。横軸に表す時刻 t のそれぞれを現在時刻 $\tilde{t}$ とみなして、その時刻 $\tilde{t}$ から 24 時間後にホットスポットが発生した数を予測した結果を示している。同図より、24 時間後のホットスポット数の予測結果は、実測値のピークを的確にとらえている。一方、予測値と実測値の先後関係に着目すればホットスポットの出火数が実測値に対して予測値がやや遅行する分布であることが分かるが、実際の森林火災規模を的確にとらえており早期警報システムとして予測を行う際には重要な指標となる。

7.3 RS-P-STARFIMA モデルの特定化

前述したように、RS モデルと P-STARFIMA モデルのパラメータを同時推計し、かつレジームの変化時刻を内生的に決定できるような推計法を開発することは困難である。そこで、本研究では、1) RS モデルを推計するとともに、レジーム状態の時系列を推計する。2) レジームの時系列を与件として、P-STARFIMA モデルのパラメータを推計するという手順を採用する。

RS モデルを推計する場合、レジーム変化を説明する先行変数 $v_{t-\Delta}$ とその先行ラグ Δ を先験的に決定することが重要である。マレーシア半島で発生するヘイズ災害は、スマトラ島における森林火災で発生した煤煙が、南西モンスーン風により半島部に運ばれることに

説明変数	推定値	t 値
前回出火期間	0.034	1.77
既往鎮火回数	0.022	2.72
定数項	-0.0764	-9.64
固有効果（地点 1）	0.084	0.98
固有効果（地点 2）	-0.192	-0.67
固有効果（地点 3）	0.105	0.65
固有効果（地点 4）	0.054	0.63
固有効果（地点 5）	0.035	0.29
固有効果（地点 6）	0.086	0.27
固有効果（地点 7）	0.057	0.41
固有効果（地点 8）	-0.144	-1.01
固有効果（地点 9）	0.001	0.01
固有効果（地点 10）	-0.048	-0.34
形状パラメータ	0.362	22.92
ペナルティー項分散	1.866	2.77
最終尤度	-2066.3	
平均尤度	-2.364	
サンプル数	874	

表 - 1 森林火災発生モデルの推計結果

より発生する．中でも，スマトラ島中部北東沿岸の Riau 地域は森林火災の多発地域であり，2005 年 7 月～9 月の観測期間中も同地域において大規模な森林火災が発生している．図 - 7.3 に南西モンスーン期の典型的な風向と Riau 地域の位置を示している．森林火災の発生地点は，ホットスポットと呼ばれる．スマトラ島の Riau 地域のホットスポット数は，NOAA 衛星より 1 日 1 回または 2 回の頻度で観測されている．解像度は $1.1 \times 1.1\text{km}$ である．同期間における衛星情報より測定し，内挿によって 1 時間ごとのホットスポット数を求めた．現在のところ，ASMA では NOAA 衛星情報のみ利用可能であるため，RS モデルの説明変数としてホットスポット数のみを用いる．マラッカ海峡上空のモンスーン風の状態（風向と風速）や火災地点の雨量データが入手可能であれば，レジーム変化の予測精度を向上できる．また，過去のヘイズ災害から，スマトラ島の大規模火災から約 12 時間でヘイズ災害が発生していることが判明している．そこで，先行ラグとして 12 時間を採用した．なお，先行ラグを変化させ，RS モデルを推計した結果，先行ラグが 12 時間の時，RS モデルの尤度比が最大になることを確認している．

P-STARFIMA モデルを推計する場合，周期 s と空間重み行列 W_l を特定化する必要がある．周期は，PM10 原系列の標本自己相関関数を用いて決定できる．図 - 7.4 に，PM10 原系列（観測地点 CA0005）のラグ w に対する標本自己相関関数

$$\rho_k(w) = \frac{T}{T-w} \frac{\sum_{t=1}^{T-w} [z_{k,t}]' z_{k,t+s}}{\sum_{t=1}^T z_{k,t} z_{k,t}}$$

説明変数	推定値	t 値
前回出火期間	0.012	2.35
既往鎮火回数	0.023	2.59
定数項	-0.665	-8.35
固有效果（地点 1）	-0.030	-0.23
固有效果（地点 2）	-0.340	1.02
固有效果（地点 3）	-0.093	-0.52
固有效果（地点 4）	-0.151	-1.70
固有效果（地点 5）	0.003	0.05
固有效果（地点 6）	-0.038	-0.15
固有效果（地点 7）	-0.105	-0.76
固有效果（地点 8）	0.137	0.75
固有效果（地点 9）	-0.028	-0.23
固有效果（地点 10）	0.062	0.39
形状パラメータ	0.718	23.08
ペナルティー項分散	1.212	0.76
最終尤度	-1413.3	
平均尤度	-1.717	
サンプル数	823	

表 - 2 森林火災継続モデルの推計結果

を示している．この図より，明らかに原系列に 24 時間単位の周期性を確認できるため，周期を $s = 24$ と設定する．また，同図に示すように，標本自己相関関数の減少速度は緩慢であり，長期記憶性の存在を示唆している．

さらに，空間近接構造を表す空間重み行列 W_l を特定化するために，CA 局間の距離に基づいて各観測地点の近傍（clique）を定義した．すなわち，地点間距離が 1 ～ 100km の場合を 1 次の近傍，101 ～ 200km の場合を 2 次の近傍と呼ぶ．その上で，空間重み行列 W_l の要素 w_{ij}^l を，地点 i に対して地点 j が l 次の近傍であることを表わすダミー変数（すなわち，地点 i と地点 j が l 時の近傍であれば $w_{ij}^l = 1$ ，そうでない時 $w_{ij}^l = 0$ となる）を用いて定義した．この時，

$$W_0 z_t = z_t$$

$$W_l z_t = \left(\sum_{j=1}^K w_{1j}^l z_{j,t}, \dots, \sum_{j=1}^K w_{Kj}^l z_{j,t} \right)'$$

が成立する．なお，大気汚染データの場合，空間構造が地点別の風力や風向の影響を受ける可能性がある．しかし，各 CA 局と他の CA 局で観測されたデータの空間自己相関関数と， W_l を用いて集計化したデータの空間自己相関関数の間に有意な差異が存在しないことが判明した．したがって，以下では，空間構造行列を地点間距離のみで定義することとする．

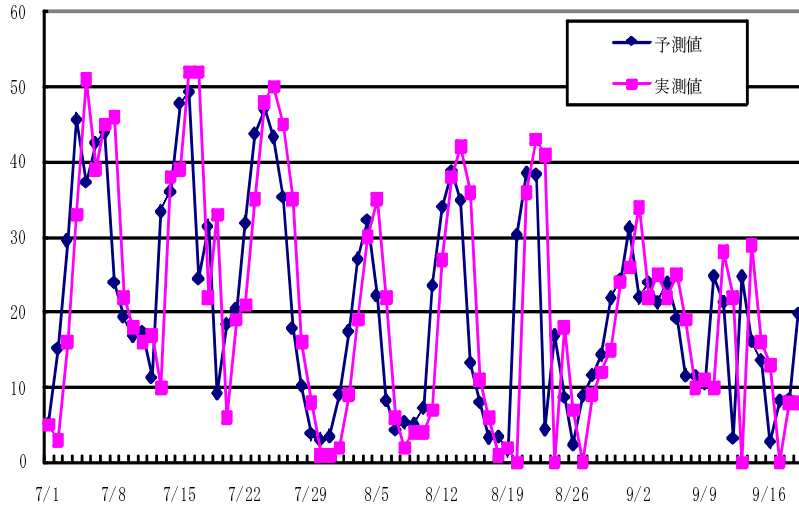


図-7.2 24 時間後のホットスポット数推計結果

前述したように，確率過程が長期記憶性を示す場合，自己相関関数や偏自己相関関数に基づいて視覚的に AR / MA 項の次数を決定できない．そこで，モデル構造を網羅的に設定し，その中から AIC 基準 (式 (6.14) 参照) が最大となるようなモデル構造を選択した．その結果，RS-P-STARFIMA((2, (1, 1)), d , (0, 0))((1, 0), d_{24} , (0, 0)) モデルが採択された．最終的に選択されたモデル構造は

$$\begin{aligned}
 & (I - \sum_{j=1}^2 \phi_{j0} \mathbf{W}_0 L^j - \sum_{j=1}^2 \phi_{j1} \mathbf{W}_1 L^j) \\
 & (I - \Phi_{10} \mathbf{W}_0 L^{24}) \nabla^d \nabla_{24}^{d_{24}} z_t(\mathbf{S}_t^*) \\
 & = \epsilon_t
 \end{aligned} \tag{7.1}$$

と定式化できる．すなわち，選択されたモデルには，1 次の空間ラグを持つ 2 次の AR ラグ，24 時間周期の 1 次の AR ラグ，および非周期，周期性を持つ長期記憶ラグが含まれるが，MA ラグは含まれない．

7.4 RS モデルの推計

まず，4 (2) で考察したセミパラメトリック EM アルゴリズムを用いて RS モデルを推計した．モデル推計にあたっては，15 箇所の CA 局で観測された PM10 の時系列データ $\{\mathbf{x}_t\}_{t=0}^{6552}$ を用いた．標準化変数の確率密度関数 $g_{k,m}$ の推計にセミパラメトリックな方法を用いているため，相当量の計算時間が必要となる．特に，レジーム数を多くすれば，計算

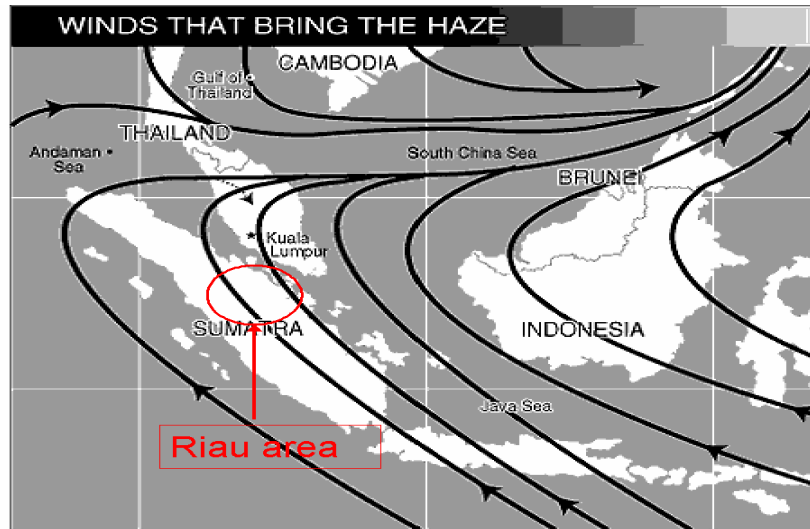


図-7.3 南西モンスーン期の風向と Riau 地域の位置

負荷が飛躍的に増加する．このため，基本的なレジーム数を $M = 2$ と設定し，1) 平常時のレジーム状態を $S_{k,t} = 1$ ，2) ヘイズ発生時のレジーム状態を $S_{k,t} = 2$ に設定した．EM アルゴリズムを用いて推計した CA0005 局 ($k = 2$) におけるレジーム状態 $m = 2$ の確率密度分布 $g_{3,2}(z_{3,t})$ の推計結果を図 - 7.5 に示している．CA0005 局は首都 Kuala Lumpur に位置する観測局であり，ヘイズ災害で深刻な被害が発生している．

同図には，支配パラメータの最尤推計量 $\hat{\beta}$ に対して，初期確率密度関数 $g_{3,2}^{(0)}$ を推定した結果と，アルゴリズムが終了した時点の確率密度関数 $g_{3,2}^*$ の推計結果を同時に示している．確率密度関数 $g_{3,2}^{(0)}$ は双峰的である．式 (6.8) を用いて対称化操作を行うことにより，確率密度関数を求めたことを考慮すれば，時空間変数 $x_{3,t}$ ($t \in \Omega_{3,2}$) の初期分布 $g_{3,2}^{(0)}$ が最尤推計された分布 $g_{3,2}^*$ より相当程度乖離していたことが読み取れる．しかし，EM アルゴリズムを反復することにより，最終的に期待値 $\mu_{3,2}^*$ を中心として対称的に分布する確率密度関数が得られている．表 - 2 には，各 CA 局におけるレジーム状態の PM10 濃度の期待値と，平常時の継続確率 $P_k(s_{k,t} = 1 | s_{k,t-1} = 1; v_{t-1-\Delta})$ の支配パラメータ β_k^{11} とヘイズ発生時におけるヘイズ終了確率 $P_k(s_{k,t} = 1 | s_{k,t-1} = 2; v_{t-1-\Delta})$ の支配パラメータ β_k^{21} の推計結果を示している．ホットスポット数が増加するほど，ヘイズ災害の終了確率は小さくなるため， β_k^{21} は負の値をとっている．

特に，CA0005，CA0010，CA0054，CA0016 局では，ヘイズ発生時の PM10 濃度の期待値が，マレーシア政府がヘイズ災害の基準値と定める $200 \mu\text{g}/\text{m}^3$ を超えている．特に，Kuala Lumpur 近郊に位置する CA0054 局の PM10 濃度は，基準値を 2 倍を超える値になっている．

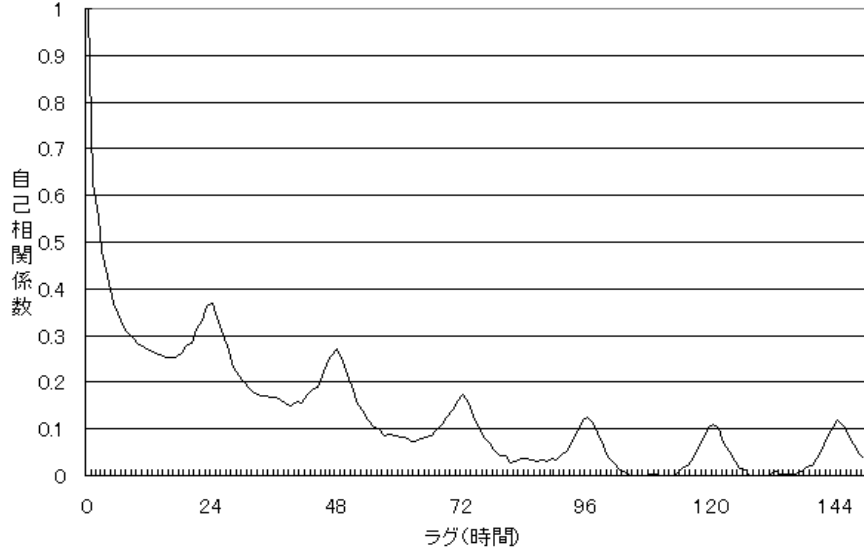


図-7.4 標本自己相関関数

7.5 P-STARFIMA モデルの推計

RS モデルが推計できれば、情報集合の点列 $\{\psi_t\}_{t=1}^T$ に対して、レジーム生起確率 $P_k(S_{k,t} = m|\psi_t, \beta_k)$ を求めることが可能となる。さらに、レジーム生起確率を用いて、情報集合の点列 $\{\psi_t\}_{t=1}^T$ に対して、レジーム状態の点列 $S_t\}_{t=0}^T$ を求める。その上で、標準化変数の推定値点列 $\{z_t(S_t)\}_{t=0}^T$ を求め、P-STARFIMA モデルのパラメータを推計した。P-STARFIMA モデル (7.1) のパラメータに関する CSS 推計量 $\ddot{d}, \ddot{d}_s, \ddot{\eta} = (\ddot{\phi}_{10}, \ddot{\phi}_{20}, \ddot{\phi}_{11}, \ddot{\phi}_{21}, \ddot{\Phi}_{10})$ と各推計量の t -値を表 - 3 に示している。いずれのパラメータも t -値が十分大きい値を示している。さらに、1 次の空間ラグと時間ラグを持つ自己回帰パラメータ ϕ_{11} 、および 1 次の空間ラグと 2 次時間ラグを持つ自己回帰パラメータ ϕ_{21} が、それぞれ 0.0980, 0.1433 であり大きい値を示している。このことより、RM10 が時空間的な拡散を伴う現象であることが理解できる。一方、24 時間の周期性ラグを表すパラメータ Φ_{10} は -0.0924 であり、PM10 の時間減衰特性を表している。長期記憶性パラメータ d, d_s は、それぞれ 0.2534, 0.1711 である。 $0 < d, d_s < 0.5$ が成立することより、長期記憶過程の定常性条件を満足している。ここで、長期記憶パラメータの有意性を単純な t -検定で評価できないため、RS-P-STARFIMA モデル (7.1) の長期記憶成分の存在に関する LM 検定²⁵⁾を試みよう。長期記憶性に関する仮説検定モデルを

$$\begin{cases} H_0 : d = 0 & \text{and} \quad \ddot{d}_s, \ddot{\eta}, \ddot{\sigma}^2 \\ H_1 : d \neq 0 & \text{and} \quad \ddot{d}_s, \ddot{\eta}, \ddot{\sigma}^2 \end{cases} \quad (7.2)$$

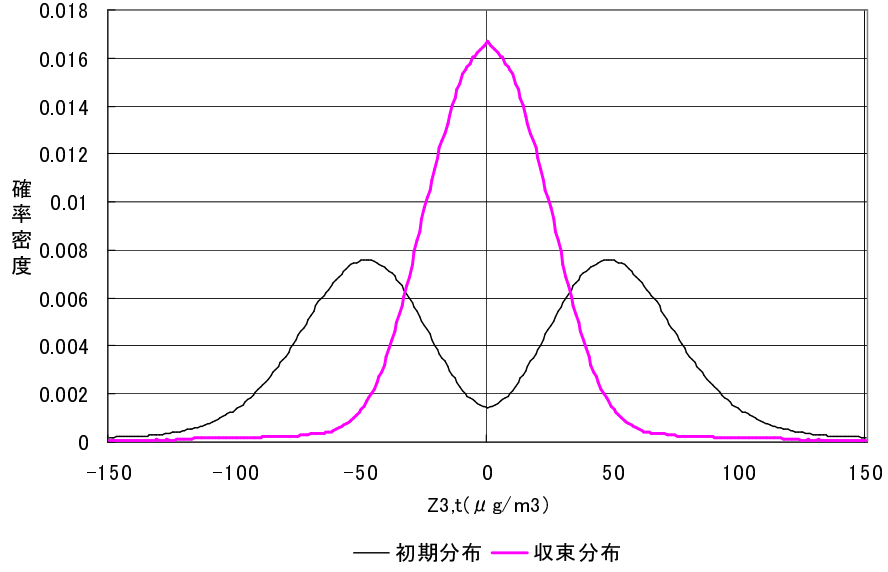


図-7.5 確率密度分布 (CA0005 局)

と定式化する．CSS 関数 (6.11) を $\tilde{\mathcal{L}}(d, d_s, \boldsymbol{\eta}, \sigma^2)$ とすれば，LM 検定統計量は

$$\begin{aligned}\tilde{\mathcal{L}}^\circ(H_0|H_1) &= \left. \frac{\partial \tilde{\mathcal{L}}(d, d_s, \boldsymbol{\eta}, \sigma^2)}{\partial d} \right|_{d=0, \ddot{d}_s, \ddot{\boldsymbol{\eta}}, \ddot{\sigma}^2} \\ &= T \sum_{i=1}^{T-1} \sum_{k=1}^K \frac{\ddot{\gamma}_k(i)}{i} \\ \ddot{\gamma}_k(i) &= \frac{\sum_{t=1}^{T-i} \ddot{\varepsilon}_{k,t} \ddot{\varepsilon}_{k,t+i}}{\sum_{t=1}^T \ddot{\varepsilon}_{k,t}^2}\end{aligned}\tag{7.3}$$

と表される²⁵⁾． $T \rightarrow \infty$ において

$$\frac{1}{\sqrt{TK}} \tilde{\mathcal{L}}^\circ(H_0|H_1) \rightarrow \sqrt{\frac{\pi^2}{6}} N(0, 1)\tag{7.4}$$

となり，標準正規分布を用いて検定できる．有意水準を $\alpha 100\%$ とする仮説 H_0 の帰無仮説 H_1 に対する棄却域は

$$\left| \frac{\sqrt{6}}{\sqrt{\pi^2 TK}} \tilde{\mathcal{L}}^\circ(H_0|H_1) \right| \geq N_{\frac{\alpha}{2}}\tag{7.5}$$

と表される．ただし，標準正規確率密度関数 ϕ に関して $\phi(N_{\frac{\alpha}{2}}) = \alpha/2$ が成立する．同様に，周期実数分パラメータ d_s に関する仮説検定モデルを

$$\begin{cases} \hat{H}_0 : d_s = 0 & \text{and} \quad \ddot{d}, \ddot{\boldsymbol{\eta}}, \ddot{\sigma}^2 \\ \hat{H}_1 : d_s \neq 0 & \text{and} \quad \ddot{d}, \ddot{\boldsymbol{\eta}}, \ddot{\sigma}^2 \end{cases}\tag{7.6}$$

	CSS 推計量	t-値
d	0.2534	
d_s	0.1711	
ϕ_{10}	0.0899	32.292
ϕ_{20}	0.2387	85.741
ϕ_{11}	0.0980	35.220
ϕ_{21}	0.1433	51.481
Φ_{10}	-0.0924	-33.211
初期尤度	-345085	
最終尤度	-299144	
σ^2	546.94	
AIC	598306	

表 - 3 モデルの推計結果

帰無仮説と対立仮説	統計量
$H_0 : H_1$	-35.28
$\hat{H}_0 : \hat{H}_1$	-56.01

表 - 4 LM テスト統計量

と表そう．CSS 関数を $\tilde{\mathcal{L}}^\circ(\beta, \sigma^2)$ とすれば，検定統計量は

$$\begin{aligned}
\tilde{\mathcal{L}}^\circ(\hat{H}_0|\hat{H}_1) &= \frac{\partial \tilde{\mathcal{L}}(d, d_s, \boldsymbol{\eta}, \sigma^2)}{\partial d} \Big|_{d_s=0, \hat{d}, \hat{\boldsymbol{\eta}}, \hat{\sigma}^2} \\
&= T \sum_{i=1}^{[(T-1)/s]} \sum_{k=1}^K \frac{\ddot{\gamma}_k(is)}{i} \\
\ddot{\gamma}_k(i) &= \frac{\sum_{t=1}^{T-i} \ddot{\epsilon}_{k,t} \ddot{\epsilon}_{k,t+i}}{\sum_{t=1}^T \ddot{\epsilon}_{k,t}^2}
\end{aligned} \tag{7.7}$$

と表される．ただし $[(T-1)/s]$ は $(T-1)/s$ を超えない最大の整数とする． $T \rightarrow \infty$ において

$$\frac{1}{\sqrt{TK}} \tilde{\mathcal{L}}^\circ(\hat{H}_0|\hat{H}_1) \rightarrow \sqrt{\frac{\pi^2}{6}} N(0, 1) \tag{7.8}$$

となり， d の場合と同様に，標準正規分布を用いて検定できる．表 - 4 に LM 検定の結果を示す．これより，実数分パラメータ d, d_s に関して有意水準 1% で帰無仮説を棄却することができる．すなわち，RS-P-STARFIMA((2, (1, 1)), $d, (0, 0)$) $\times$ ((1, 0), $d_s, (0, 0)$) の周期成分，非周期成分の双方における長期記憶性の存在を棄却できない．

7.6 危険予告情報の作成

ヘイズ災害に関する危機管理を行う上で、1) ヘイズ災害の発生に関する危険情報と、2) ヘイズ災害の危険解除情報を事前に住民に通知することが重要である。ASMA では、サテライト情報と各 CA 局における 1 時間毎の大気観測結果に基づいて、24 時間後のヘイズ災害の発生を予告する情報とヘイズ災害の終息に関する情報を提供するシステムの開発をめざしている。RS-P-STARFIMA モデルの特徴は、RS モデルと P-STARFIMA モデルを連動することにより、スマトラ島における森林火災の発生によるレジーム状態と PM10 濃度の急激な変化をモデル化できる点にある。本節では、RS-P-STARFIMA モデルを用いて、1) 24 時間後におけるヘイズ災害発生確率、2) ヘイズ災害の深刻度に関する危険予告情報を作成する。

RS モデルの説明変数として、スマトラ半島における 24 時間前のホットスポット数という先行指標を用いている。現在時刻 $\tilde{t}$ において、24 時間先のレジーム状態を予測するためには、先行指標 $\bar{v}_{\tilde{t}-24}, \dots, \bar{v}_{\tilde{t}-1}$ が必要となる。当該時刻 $\tilde{t}$ において、これらの情報は利用可能である。いま、当該時刻 $\tilde{t}$ のレジーム生起確率を $P_k(S_{k,\tilde{t}} = m_0 | \psi_{\tilde{t}})$ とする。この時、当該時刻 $\tilde{t}$ から 24 時間先のレジーム生起確率 $\hat{P}_k(\hat{S}_{k,\tilde{t}+24} = m_{24} | \psi_{\tilde{t}+24})$ は、式 (5.7) を反復的に利用することにより

$$\begin{aligned}
 & \hat{P}_k(S_{k,\tilde{t}+24} = m_{24} | \psi_{\tilde{t}+24}) \\
 &= \sum_{m_{23}=1}^M \hat{P}_k(S_{k,\tilde{t}+23} = m_{23} | \psi_{\tilde{t}+23}) \\
 & P_k(S_{k,\tilde{t}+24} = m_{24} | S_{k,\tilde{t}+23} = m_{23}; \psi_{\tilde{t}+23}) \\
 &= \dots = \sum_{m_0=1}^M \dots \sum_{m_{23}=1}^M P_k(S_{k,\tilde{t}} = m_0 | \psi_{\tilde{t}}) \\
 & \prod_{l=1}^{23} P_k(S_{k,\tilde{t}+l} = m_l | S_{k,\tilde{t}+l-1} = m_{l-1}; \psi_{\tilde{t}+l-1}) \\
 & P_k(S_{k,\tilde{t}+24} = m_{24} | S_{k,\tilde{t}+23} = m_{23}; \psi_{\tilde{t}+23})
 \end{aligned} \tag{7.9}$$

と表される。図 - 7.6 は、CA0005 局 ($k = 3$) において、横軸に表す時刻 t のそれぞれを現在時刻 $\tilde{t}$ とみなして、その時刻 $\tilde{t}$ から 24 時間後にヘイズ災害が発生する (レジーム状態 $S_{3,\tilde{t}+24} = 2$ が生起する) 確率を予測した結果を示している。なお、24 時間以上の将来時刻のレジーム状態を予測するために、は現時刻以降のホットスポット数の推計結果を用いる必要がある。

また、現在時刻 $\tilde{t}$ から、24 時間後の PM10 濃度 $X_{\tilde{t}+24}$ を予測する問題を考えよう。将来時刻の PM10 濃度を予測する場合、その時刻までの PM10 濃度を RS-P-STARFIMA モデルを用いて逐次予測することが必要となる。この場合でも、24 時間将来までのレジーム状態の予測結果 $\{\hat{S}_{k,t}\}_{t=\tilde{t}}^{\tilde{t}+24}$ は利用可能である。したがって、 $\hat{X}_{\tilde{t}+24}$ を予測することは、 z_t の

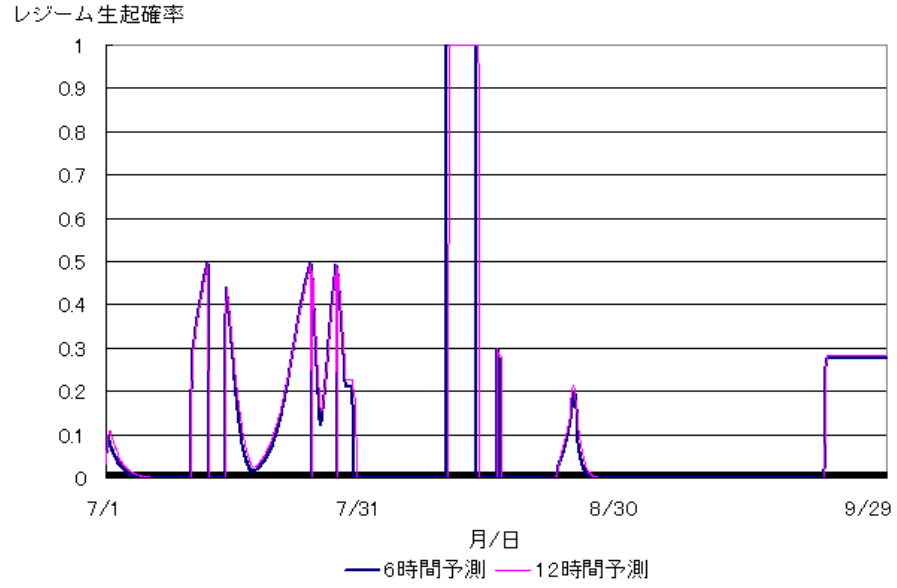


図-7.6 外生的に与えた時の 6, 12 時間後レジーム予測確率 (CA0005 局)

予測量 $\hat{z}_{t+24}$ を求める問題に帰着する．ここで， $S_t(S_t^*)$ を次式のように定義する．

$$\begin{aligned} y_t(S_t^*) = & (I - \sum_{j=1}^2 \phi_{j0} \mathbf{W}_0 L^j - \sum_{j=1}^2 \phi_{j1} \mathbf{W}_1 L^j) \\ & (I - \Phi_{10} \mathbf{W}_0 L^{24}) z_t(S_t^*) \end{aligned} \quad (7.10)$$

この時，確率過程 $\{y_t(S_t^*)\}_{t=0}^{\infty}$ は定常かつ反転可能な周期的時空間実数値和分過程

$$\nabla^d \nabla_{24}^{d_{24}} y_t(S_t^*) = \varepsilon_t \quad (7.11)$$

に従う．さらに，時系列過程 (7.11) の無限次元自己回帰過程，無限次元移動平均過程を，

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j \mathbf{y}_{t-j}(\mathbf{S}_{t-j}) \quad (7.12)$$

$$\mathbf{y}_t(\mathbf{S}_t) = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad (7.13)$$

と表現しよう．現在時刻 $\tilde{t}$ で予測した h ($h \geq 1$) 期先の最良線形予測量 $\hat{\mathbf{y}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h})$ は，過去の残差点列 $\{\bar{\varepsilon}_{\tilde{t}+h-j}(\mathbf{S}_{\tilde{t}+h-j}^*)\}_{j=h}^{\infty}$ を用いて，

$$\begin{aligned} \hat{\mathbf{y}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h}) &= \sum_{j=1}^{\infty} -\pi_j \bar{\mathbf{y}}_{\tilde{t}+h-j}(\mathbf{S}_{\tilde{t}+h-j}^*) \\ &= \sum_{j=h}^{\infty} \psi_j \bar{\varepsilon}_{\tilde{t}+h-j}(\mathbf{S}_{\tilde{t}+h-j}^*) \end{aligned} \quad (7.14)$$

と表せる⁵⁷⁾ (付録 4) 参照)．現在時刻 $\tilde{t}$ での残差は，時刻 $\tilde{t}$ 以前の観測値を用いて

$$\begin{aligned} \bar{\varepsilon}_t(\mathbf{S}_t^*) &= (I - \sum_{j=1}^2 \phi_{j0} \mathbf{W}_0 L^j - \sum_{j=1}^2 \phi_{j1} \mathbf{W}_1 L^j) \\ &\quad (I - \Phi_{10} \mathbf{W}_0 L^{24}) \nabla^d \nabla_{24}^{d_{24}} \bar{\mathbf{z}}_t(\mathbf{S}_t^*) \end{aligned} \quad (7.15)$$

と表される．また，

$$\begin{aligned} \hat{\mathbf{z}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h}) &= \hat{\mathbf{y}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h}) \\ &+ \{(\sum_{j=1}^2 \phi_{j0} \mathbf{W}_0 L^j + \sum_{j=1}^2 \phi_{j1} \mathbf{W}_1 L^j)(I - \Phi_{10} \mathbf{W}_0 L^{24}) \\ &\quad + \Phi_{10} \mathbf{W}_0 L^{24}\} \bar{\mathbf{z}}_{\tilde{t}+h}(\mathbf{S}_{\tilde{t}+h}^*) \end{aligned} \quad (7.16)$$

を用いて $\hat{\mathbf{y}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h})$ を $\hat{\mathbf{z}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h})$ に変換することができる． $\hat{\mathbf{z}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h})$ が求まれば，標準偏差ベクトル $\rho(\hat{\mathbf{S}}_{\tilde{t}+h}) = \{\rho_k(\hat{\mathbf{S}}_{\tilde{t}+h}) : k = 1, K\}$ ，レジーム平均ベクトル $mu(\hat{\mathbf{S}}_{\tilde{t}+h}) = \{\mu_k(\hat{\mathbf{S}}_{\tilde{t}+h}) : k = 1, K\}$ から

$$\hat{\mathbf{X}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h}) = \rho(\hat{\mathbf{S}}_{\tilde{t}+h}) \hat{\mathbf{z}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h}) + \mu(\hat{\mathbf{S}}_{\tilde{t}+h}) \quad (7.17)$$

を得る．すなわち，当該時刻 $\tilde{t}$ までの観測値を用いて最良線形予測量 $\{\hat{\mathbf{X}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h})\}_{h=1}^{\infty}$ を構成することができる．中部地域の CA0005 局 ($k = 3$) における 24 時間後予測値を図 - 7.10 に示している．同図は，横軸に示す各時刻を現在時刻 $\tilde{t}$ と考え，縦軸方向にその時刻 $\tilde{t}$ から 24 時間後の時刻 $\tilde{t} + 24$ における PM10 濃度の予測値 $\hat{x}_{3,\tilde{t}+h}(\hat{\mathbf{S}}_{3,\tilde{t}+h})$ をプロットしたものである．併せて，同図には，現実に横軸の各時刻 $\tilde{t}$ に対して，現実に観測された時刻 $\tilde{t} + 24$ における観測値 $\bar{x}_{3,\tilde{t}+24}$ を示している．この結果より，RS-P-STARFIMA モデルを用いて，現実の PM10 濃度の動向をある程度再現できていると言える．しかしながら幾つかの問題が見受けられる．8/11 前後のピークに着目すると，24 時間後の予測値は実測値に対して遅れて立ち上がっている．つまりレジーム変化の予測精度は良好であるが，ピー

ク時濃度の予測結果に時間遅れが発生していることがわかる．この原因としては、レジーム状態が変化する際にレジーム推移確率の説明変数として採用した hotspot 数に起因していると考えられる．図 - 7.2 は横軸に表す時刻 t のそれぞれを現在時刻 $\tilde{t}$ とみなして、その時刻 $\tilde{t}$ から 24 時間後にホットスポットが発生した数を予測した結果を示している．同図より、24 時間後のホットスポット数の予測結果は、実測値のピークを的確にとらえている．一方、予測値と実測値の先後関係に着目すればホットスポットの出火数が実測値に対して予測値が遅延する分布であることが分かる．同様の傾向が PM10 の大気濃度予測にも見られる．図 - 7.10 の大気濃度予測結果を見ると、レジーム変化の遅れを反映するように予測値として pm10 大気濃度がピークを迎える時刻に遅れがみられる．つまり、観測値と比べた場合、ヘイズ災害の規模が最大になる発生時刻が遅れて到着するというシミュレーション結果が得られている．この原因としてはレジーム変化確率の予測精度によるところが大きい．P-STARFIMA モデルでは、当該時刻 $\tilde{t}$ までの観測値を用いて最良線形予測量を構成するため、ピーク時点においてピーク時濃度に関する観測値を利用できないという限界がある．ピーク時濃度の時間遅れを是正するためには、レジームの細分化を行ない、ピーク時濃度の予測精度を向上させる必要がある．

7.7 危険解除情報の作成

ヘイズ災害のいま 1 つの特徴は、汚染物質が大気中に滞留し、長期間にわたって被害が継続することである．このため、ASMA では、住民にヘイズ災害の終息に関する危険解除情報を提供するシステムの開発もめざしている．危険解除情報としては、1) ヘイズ災害の継続期間の長さに関する情報、2) ヘイズ災害の終息に関する情報が重要な役割を果たす．前者は、ヘイズ災害が発生した（あるいは、継続している）段階で、ヘイズ災害が継続する期間の長さを住民に通知することを目的とする．ヘイズ災害は、スマトラ島の森林火災による有害物質がマレーシアに到達することにより発生する．ヘイズ災害の期間長を予測するためには、スマトラ島における森林火災の継続時間を予測することが必要である．森林火災の継続時間は、たとえば多目的地ハザードモデル⁵⁸⁾を用いて予測することが可能である．一方、スマトラ島の森林火災が終息したことが判明した段階で、1) ヘイズ災害に関する危険解除を宣言する時刻を予告する、あるいは、2) CA 局における観測情報に基づいて、危険解除情報を提供できるか否かを判断する問題である．以下では、これら 2 つの問題について考察しよう．

今、ある時刻 $\tilde{t}$ より各 CM 局において、ヘイズ災害が終息するまでに要する時間を予測する問題を考える．24 時間前から現時刻までの先行指標に関する点列を $\{\bar{v}_{\tilde{t}+l-24}\}_{l=1}^{24}$ と表す．また現時刻以降の先行指標に関する予測値の点列 $\{v_{\tilde{t}+l}^{\infty}\}_{l=1}^{\infty}$ は duration モデルより随時更新可能である．この時、先行指標の点列 $\{v_{\tilde{t}+l-24}\}_{l=1}^{\infty}$ を与件として、式 (7.9), (6.2) を反復

的に繰り返すことにより，レジーム状態の点列 $\{\hat{S}_t^*\}_{t=\bar{t}}^\infty$ を求めることができる．さらに，将来時刻の誤差項ベクトルの点列を $\{\varepsilon_t = \mathbf{0}\}_{t=\bar{t}}^\infty$ と仮定する．この時，式 (7.14), (7.16), (7.17) を用いて，将来時刻における PM10 濃度の予測値ベクトルの点列 $\{\hat{x}_t(S_t)\}_{t=\bar{t}}^\infty$ を求めることができる．図 - 7.11 はヘイズ被害が深刻であった中部地域の CA0016 局を対象として，ヘイズ終息時刻の予測を試みた結果を示している．この事例では，8月10日12時を現在時刻と考え，CA0016 局の PM10 濃度が当該時刻以降にどのように減衰するかをシミュレートしている．8月12日13時を起点とする事例2のシミュレーション結果では，24時間経過した時刻においても PM10 濃度は危機管理基準である $200\mu\text{g}/\text{m}^3$ 程度の値になっている．同図には，同局の平常時における PM10 濃度の期待値を併記している．森林火災が終了してから5日が経過した時刻においても，PM10 濃度は平常時の2倍程度の値になっている．一方，8月10日12時を起点とする事例1のシミュレーション結果では，起点時刻における PM10 濃度は事例2の場合よりも大きい値を示している．しかし，PM10 濃度の減衰速度は事例2の場合よりも早い結果となっている．このように PM10 濃度の変化過程は，過去における PM10 の滞留履歴や他地点からの時空間的影響を受けて多様に変化することが理解できる．これらの結果を踏まえると，ヘイズ災害の危険解除情報を提供する場合には，その時刻までの観測結果に基づく慎重な判断が求められることがわかる．なお，どの時刻においてヘイズ災害に関する警報解除を発令するかに関しては，意思決定者（政府）が，大気汚染物質の残留濃度や警報発令期間長を総合的に勘案して決定する問題である．ASMA に課せられた課題は，あくまでも大気汚染物質の残留濃度に関する長期予測結果（図 - 7.11 参照）を作成することにとどまる．

レジーム生起確率

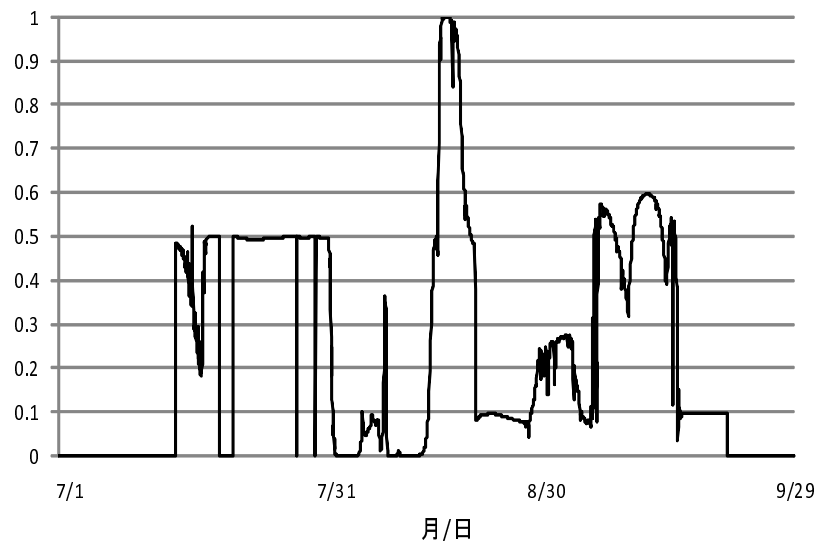


図-7.7 シュミレーションにより与えた時の 24 時間後レジーム予測確率 (CA0005 局)

PM10濃度($\mu\text{g}/\text{m}^3$)

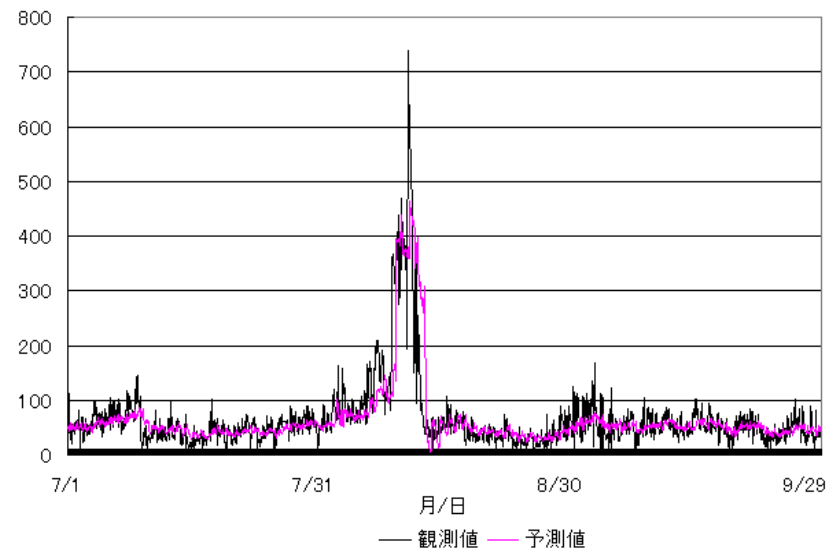


図-7.8 外生的に与えた時の 6 時間後予測値 (CA0005 局)

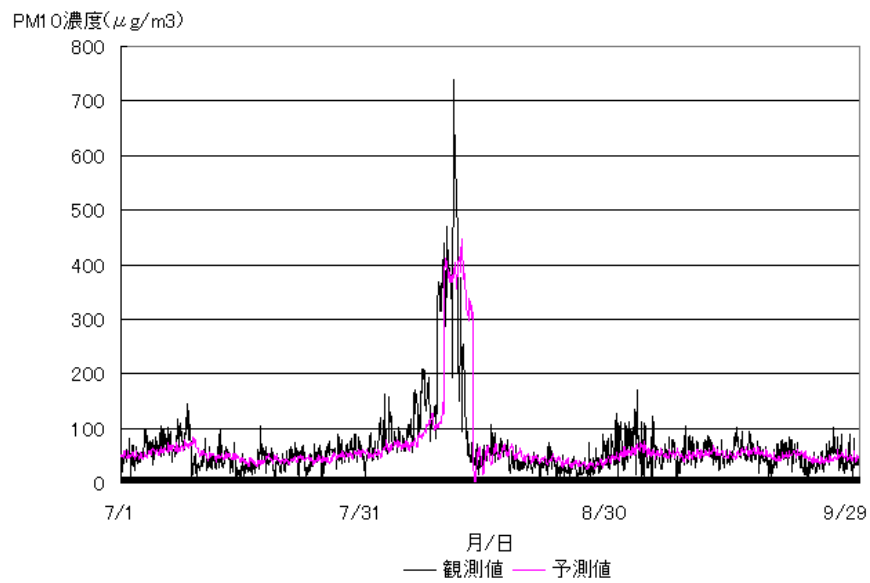


図-7.9 外生的に与えた時の 12 時間後予測値 (CA0005 局)

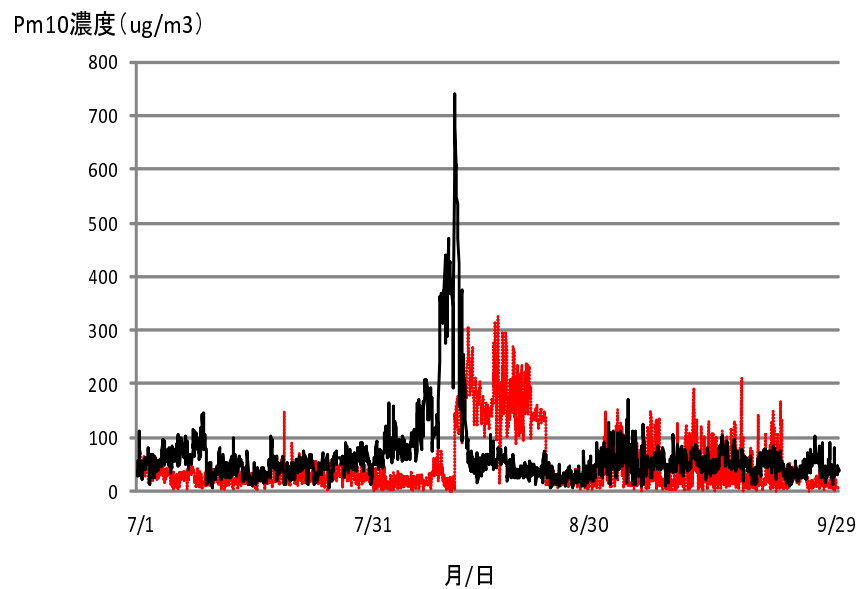


図-7.10 シュミレーションにより与えた時の 24 時間後予測値 (CA0005 局)

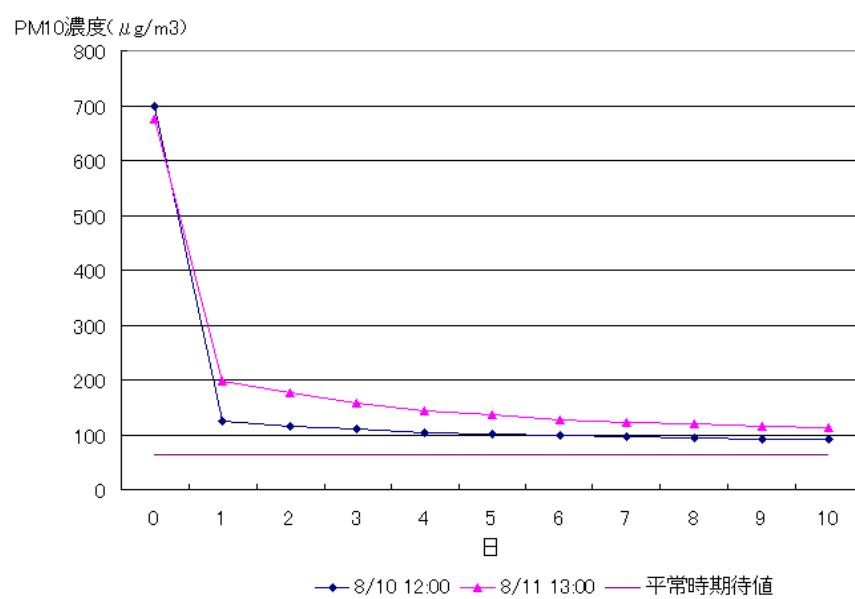


図-7.11 24時間毎長期予測値 (CA0016局)

第8章 おわりに

本研究では、ヘイズ災害のリスク解除情報を作成するため、森林火災発生・終息モデルに基づく、ヘイズ災害の統計的予測モデルの開発を試みた。その際、ヘイズ災害の特徴として、森林火災の発生による大気汚染物質濃度の急激な上昇と、大気中に滞留する汚染物質が緩やかに減衰する時系列相関構造を有することを指摘した。本研究で提案した RS-P-STARFIMA モデルは、ヘイズ発生による急激な PM10 の変動と、大気中に残存する PM10 の長期記憶性、周期性の双方を同時に表現できるという特徴を持っている。さらに、提案したモデルを用いて、スマトラ島における森林火災の発生に関する人工衛星情報に基づいて、マレーシア半島部におけるヘイズ災害の危険予告情報と危険解除情報を作成する方法論を提案した。また、2005 年に発生したヘイズ災害を対象として、RS-P-STARFIMA モデルを実際に作成するとともに、危険管理情報の作成方法の有効性を実証的に検証した。当然のことながら、実証分析で作成した RS-P-STARFIMA モデルは、分析対象とした 2005 年のヘイズ災害にのみ有効な統計的予測モデルである。モデルの汎用性を増加させるためには、火災データ及びヘイズ災害に関するデータの蓄積を図ると同時に、モデルの継続的改善を図ることが必要であることは言うまでもない。

本研究で提案したヘイズ災害予測モデルは、実用性の高いモデルであると考えるが、今後に残された研究課題も少なくない。第 1 に、geo-additive 関数を用いて導出した hotspot 数だが、シミュレーション結果を見る限り、予測する時間間隔が大きくなるほど実測値との乖離が見られる。結果として、推定されるレジーム推移確率に誤差が生じ、シミュレーションで導出した大気汚染濃度と実測値の間に乖離が見られた。現在、空間的相互作用を及ぼす領域を周辺 8 領域と設定しているが、今後、空間的相互作用を及ぼす領域を統計的に見直す必要がある。第 2 に、レジームの設定方法に関する分析が不可欠である。本研究では、レジーム状態のレーティングとその設定を統計的な判断で設定した。PM10 濃度がもたらす被害の深刻度を考慮してレジーム状態を定義することが必要である。特に、レジーム生起確率は、ヘイズ発生確率を住民に告知する上で重要な情報となる。レジーム分割の判断に関しては、現地における政策担当者との緊密な議論が必要であることは言うまでもない。第 3 に、RS モデルの推計精度の向上に関する研究が必要である。そのためには、レジーム変数の設定方法と細分化を試みる必要がある。さらに、RS モデルの説明変数として、ホットスポット数以外の説明変数を組み込むことが必要であり、これらの課題を達成するには、観測データの充実と蓄積が不可欠である。それと同時に、計算負荷を軽減する方法を開発することが必要である。第 4 に、本研究ではレジーム状態の点列を確定的に設定した。今後、レジーム状態の確率分布を許容したような RS-P-STARFIMA モデルの開

発が望まれる．

なお，本研究の遂行にあたって，Alam Sekitar Malaysia（ASMA）の協力を得ている．また，データベースの作成・解析に関して，松田明広氏（経済産業省）によるところが大きい．ここに感謝の意を表します．

第9章 付録 補足説明

読者の便宜を図るため，式 (4.10),(4.12),(4.14),(7.14) を導出する．詳細は，参考文献 10), 25), 26), 57) を参照して欲しい．

1) 式 (4.10) の導出 偶数 s に対して，有限積の公式⁵⁹⁾より，次式が成立する．

$$\frac{1-L^s}{1-L^2} = \prod_{j=1}^{\frac{s}{2}-1} \left\{ 1 - 2 \cos \left(\frac{2\pi j}{s} \right) L + L^2 \right\}$$

$\Lambda = (1-L)^{d+d_s}(1+L)^{d_s}$ と略記すれば，次式を得る．

$$(1-L)^d (1-L^s)^{d_s} \quad (9.1)$$

$$= \Lambda \frac{(1-L^s)^{d_s}}{(1-L)^{d_s}(1+L)^{d_s}} = \Lambda \left\{ \frac{1-L^s}{1-L^2} \right\}^{d_s} \quad (9.2)$$

$$= \Lambda \prod_{j=1}^{\frac{s}{2}-1} \left\{ 1 - 2 \cos \left(\frac{2\pi j}{s} \right) L + L^2 \right\}^{d_s} \quad (9.3)$$

2) 式 (4.12) の導出 式 (4.10) より，以下の無限次元移動平均表現を得る．

$$x_t = (1-L)^{-(d+d_s)}(1+L)^{-d_s} \quad (9.4)$$

$$\prod_{j=1}^{\frac{s}{2}-1} \left\{ 1 - 2 \cos \left(\frac{2\pi j}{s} \right) L + L^2 \right\}^{-d_s} \varepsilon_t \quad (9.5)$$

$$= \sum_{j=0}^{\infty} \psi_j L^j \varepsilon_t \quad (9.6)$$

上式の実数フィルター $F(L)$ は，

$$F(L) = (1-L)^{-(d+d_s)}(1+L)^{-d_s} \quad (9.7)$$

$$\prod_{j=1}^{\frac{s}{2}-1} \left\{ 1 - 2 \cos \left(\frac{2\pi j}{s} \right) L + L^2 \right\}^{-d_s} \quad (9.8)$$

$$= \left\{ 1 - 2 \cos \left(\frac{2\pi 0}{s} \right) L + L^2 \right\}^{-\frac{d+d_s}{2}} \quad (9.9)$$

$$\left\{ 1 - 2 \cos \left(\frac{2\pi \frac{s}{2}}{s} \right) L + L^2 \right\}^{-\frac{d_s}{2}} \quad (9.10)$$

$$\prod_{j=1}^{\frac{s}{2}-1} \left\{ 1 - 2 \cos \left(\frac{2\pi j}{s} \right) L + L^2 \right\}^{-d_s} \quad (9.11)$$

$$= \prod_{l=1}^{1+\frac{s}{2}} \left\{ 1 - 2 \cos(\lambda_l) L + L^2 \right\}^{-D_l} \quad (9.12)$$

$$(9.13)$$

と表せる．ただし， $D_1 = (d + d_s)/2, D_{1+s/2} = d_s/2, D_l = d_s$ ($l = 2, \dots, s/2$)， $\lambda_1 = 0, \lambda_{1+s/2} = \pi, \lambda_l = 2\pi(l-1)/s$ ($l = 2, \dots, s/2$) である． F の j 次導関数より $F^{(j)}(0) = j!\psi_j$ を得る． F の 1 次導関数は，

$$F^{(1)}(L) \quad (9.14)$$

$$= F(L) \left(\sum_{l=1}^{1+\frac{s}{2}} \frac{-2D_l(L - \cos(\lambda_l))}{(1 - 2\cos(\lambda_l)L + L^2)} \right) \quad (9.15)$$

$$= F(L) \left(\sum_{l=1}^{1+\frac{s}{2}} \frac{-2D_l(L - \cos(\lambda_l))}{(L - e^{i\lambda_l})(L - e^{-i\lambda_l})} \right) \quad (9.16)$$

$$\equiv F(L)G(L) \quad (9.17)$$

である．上式で定義した関数 G の j 次導関数を用いることにより，

$$G^{(j)}(0) = \sum_{l=1}^{1+\frac{s}{2}} 2D_l j! \cos[(1+j)\lambda_l] \quad (9.18)$$

が得られる．ここで， $F^{(j)}(0) = \{F(0)G(0)\}^{(j-1)} = j!\psi_j$ である．Leibniz の公式を用いると，

$$\{F(0)G(0)\}^{(j-1)} = \sum_{i=0}^{j-1} \frac{(j-1)!}{i!(j-i-1)!} F(0)^{(i)} G(0)^{(j-i-1)} \quad (9.19)$$

である．したがって，次式を得る．

$$j!\psi_j = \sum_{i=0}^{j-1} \frac{(j-1)!}{i!(j-i-1)!} i!\psi_i (j-i-1)! \quad (9.20)$$

$$\left(\sum_{l=1}^{1+\frac{s}{2}} 2D_l \cos[(j-i)\lambda_l] \right) \quad (9.21)$$

上式を整理すると式 (4.12) が得られる．

3) 式 (4.14) の導出 式 (4.12) の中間項は以下のように整理することができる．

$(j-i) = 0, s, 2s, \dots$ の場合

$$2 \sum_{l=1}^{1+\frac{s}{2}} D_l \cos[(j-i)\lambda_l] = d + sd_s \quad (9.22)$$

$(j-i) \neq 0, s, 2s, \dots$ の場合

$$2 \sum_{l=1}^{1+\frac{s}{2}} D_l \cos[(j-i)\lambda_l] = d \quad (9.23)$$

よって式 (4.12) より，

$$\psi_j = \frac{2}{j} \sum_{i=0}^{j-1} \sum_{l=1}^{1+\frac{s}{2}} D_l \cos[(j-i)\lambda_l] \psi_i \quad (9.24)$$

$$= \frac{1}{j} \sum_{i=0}^{j-1} \left\{ d + I_{\frac{j-i}{s}} s d_s \right\} \psi_i \quad (9.25)$$

$$= \sum_{i=0}^{j-1} \left\{ \frac{1}{j} d + I_{\frac{j-i}{s}} \frac{1}{s} d_s \right\} \psi_i \quad (9.26)$$

$$= \sum_{i=0}^j \frac{(d+i-1)!}{i!(d-1)!} I_{\frac{j-i}{s}} \frac{(d_s + [\frac{j-i}{s}] - 1)!}{[\frac{j-i}{s}]!(d_s-1)!} \quad (9.27)$$

ここで、 $I_{\frac{j-i}{s}}$ は $\frac{j-i}{s}$ が整数のとき 1、それ以外のとき 0 をとるダミー変数、 $[\frac{j-i}{s}]$ は、 $\frac{j-i}{s}$ を超えない最大の非負整数を表わす。したがって、式 (4.14) が成立する。

4) 式 (7.14) の導出 式 (7.12)、式 (7.13) の無限次元自己回帰過程、無限次元移動平均過程を、ラグオペレーター L を用いて

$$\varepsilon_t = \pi(L) \mathbf{y}_t(\mathbf{S}_t) \quad (9.28)$$

$$\mathbf{y}_t(\mathbf{S}_t) = \psi(L) \varepsilon_t \quad (9.29)$$

と表現しよう。上記の 2 つの式より、

$$\pi(L) \mathbf{y}_t(\mathbf{S}_t) = \pi(L) \psi(L) \varepsilon_t \quad (9.30)$$

が得られる。よって、

$$\pi(L) \psi(L) = 1 \quad (9.31)$$

が成立する。これを L についての多項式として解くと、

$$\begin{cases} \pi_1 = -\psi_1 \\ \pi_2 = -\psi_1 \pi_1 - \psi_2 \\ \vdots \\ \pi_j = -\psi_1 \pi_{j-1} - \cdots - \psi_j \\ \vdots \end{cases} \quad (9.32)$$

が得られる。

現在時刻 $\tilde{t}$ より将来時点の誤差項に関して $\{\varepsilon_{\tilde{t}+j} = 0, j \geq 0\}$ が成立すると仮定しよう。現在時刻 $\tilde{t}$ で予測した h ($h \geq 1$) 期先の最良線形予測量 $\hat{\mathbf{y}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h})$ は、将来時刻 $\tilde{t} + h$ を起点とする過去の観測値点列 $\{\bar{\mathbf{y}}_{\tilde{t}+h-j}(\mathbf{S}_{\tilde{t}+h-j})\}_{j=1}^{\infty}$ を用いて、

$$\hat{\mathbf{y}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h}) = \sum_{j=1}^{\infty} -\pi_j \bar{\mathbf{y}}_{\tilde{t}+h-j}(\mathbf{S}_{\tilde{t}+h-j}) \quad (9.33)$$

と表せる。式 (9.33) より $\hat{\mathbf{y}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h})$ は、現在時刻 $\tilde{t}$ までの観測値点系列 $\{\bar{\mathbf{y}}_{\tilde{t}+1-j}(\mathbf{S}_{\tilde{t}})\}_{j=1}^{\infty}$ を用いて

$$\hat{\mathbf{y}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h}) = \sum_{j=1}^{\infty} -\pi_j^{(h)} \bar{\mathbf{y}}_{\tilde{t}+1-j}(\mathbf{S}_{\tilde{t}+1-j}) \quad (9.34)$$

と表すことができる．ここで式 (9.32) の関係を用いると

$$\pi_j^{(h)} = \sum_{i=0}^{h-1} \psi_i \pi_{j+h-1-i} \quad (9.35)$$

である．また式 (9.32) から $\sum_{i=0}^j \psi_{j-i} \pi_i = 0$ であるから

$$\begin{aligned} & \sum_{i=0}^{j-1} \psi_{j-i+h-1} \pi_i \\ &= \sum_{i=0}^{j+h-1} \psi_{j-i+h-1} \pi_i - \sum_{i=j}^{j+h-1} \psi_{j-i+h-1} \pi_i \\ &= - \sum_{i=j}^{j+h-1} \psi_{j-i+h-1} \pi_i = - \sum_{i=j}^{h-1} \psi_{h-i-1} \pi_{i+j} \\ &= - \sum_{i=0}^{h-1} \psi_i \pi_{j+h-i-1} \end{aligned}$$

と変形できる．すなわち

$$\pi_j^{(h)} = - \sum_{i=0}^{j-1} \psi_{j-i+h-1} \pi_i \quad (9.36)$$

となる．よって次式を得る．

$$\begin{aligned} \hat{\mathbf{y}}_{\tilde{t}+h}(\hat{\mathbf{S}}_{\tilde{t}+h}) &= \sum_{j=1}^{\infty} \left\{ \sum_{i=0}^{j-1} \psi_{j+h-1-i} \pi_i \right\} \bar{\mathbf{y}}_{\tilde{t}+1-j} \\ &= \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} \psi_{j+h-1-i} \pi_i \bar{\mathbf{y}}_{\tilde{t}+1-j-i} \\ &= \sum_{j=1}^{\infty} \psi_{j+h-1} \boldsymbol{\varepsilon}_{\tilde{t}+1-j} \\ &= \sum_{j=h}^{\infty} \psi_j \boldsymbol{\varepsilon}_{\tilde{t}+h-j} \end{aligned} \quad (9.37)$$

参考文献

- 1) The Global Fire Monitoring Centre : Forest Fire Situation in Malaysia, *International Forest Fire News Country Notes*, Vol.26, pp.66-74, 2001.
- 2) Nichol, J. : Smoke haze in SE Asia: a predictable recurrence, *Atmospheric Environment*, Vol.32, pp.2715-2716, 1998.
- 3) Matsuta A.: Seasonal Space-Time Fractional ARIMA Modeling for Atmospheric Data in Peninsular Malaysia, Master Thesis submitted to Kyoto University, 2005.
- 4) Venevsky, S., Thonicke, T., Sitch, S. and Cramer, W. : Simulating regimes in human-dominated ecosystems: Iberian Peninsula case study, *Grobal Change Biology*, Vol.8, pp.984-998, 2002.
- 5) Thonicke, T., Venevsky, S., Sitch, S. and Cramer, W. : The role of fire disturbance for global dynamics: coupling fire into a dynamic grobal vegetation model, *Grobal Ecology and Biogeography*, Vol.10, pp.661-677, 2001.
- 6) 刈屋武昭, 矢島美寛, 田中勝人, 竹内啓 : 経済時系列の統計, 岩波書店, 2003.
- 7) Cliff, A.D. and Ord, J.K.: Space-time modeling with an application to regional forecasting, *Transactions of the Institute of British Geographers*, Vol.64, pp.119-128, 1975.
- 8) Hangrove, R., Guardner, R., Tuener, M., Romme, W. and Despain, D.: Simulating firepatterns in heterogeneous landscapes, *Ecological Modelling*, Vol.135, pp.243-263, 2000.
- 9) Reed, W.J. and Mckelvey, K.S.: Power-law behaviour and parametric models for the size-distribution of forest fires, *Ecological Modelling*, Vol.150, pp.239-254, 2002.
- 10) Box, G.E.P., Jenkins, G.M. and Reinsel, G.C.: *Time Series Analysis/Forecasting and Control(third edition)*, Prentice Hall, 1994.
- 11) Whittle, P.: On stationary process in the plane, *Biometrika*, Vol.41, pp.434-449, 1954.

- 12) Pfeifer, P.E. and Deutsche, S.J.: A STARIMA model building procedure with application to description and regional forecasting, *Transactions of The Institute of British Geographers*, Vol.5, pp.330-349, 1980.
- 13) Pfeifer, P.E. and Deutsch, S.J.: Variance of the sample-time autocorrelation function of contemporaneously correlated variables, *SIAM Journal of Applied Mathematics, Series A*, Vol.40, pp.133-136, 1981.
- 14) Stoffer, D.S.: Estimation and interpretation of space-time ARMAX models in the presence of missing data, *Journal of The American Statistical Association*, Vol.81, pp.762-772, 1986.
- 15) Pfeifer, P.E., and Deutsch, S.J.: A three-stage iterative procedure of first-order for space-time modeling, *Technometrics*, Vol.22, pp.35-47, 1980.
- 16) Pfeifer, P.E. and Bodily, S.E.: A test of space-time ARMA modeling and forecasting with an application to real estate prices, *International Journal of Forecasting*, Vol.16, pp.255-272, 1990.
- 17) Kamarianakis, I. and Prastacos, P.: Space-time modeling of traffic flow, *Computers & Geosciences*, Vol.31, pp.119-133, 2005.
- 18) Kamarianakis, I. and Prastacos, P.: Forecasting Traffic flow conditions in an urban network: a comparison of multivariate and univariate procedures, *In Transportation Research Record, Journal of the Transportation Research Board*, No.1857, pp.74-84, 2003.
- 19) Niu, X.-F. and Tiao, G.C.: Modeling satellite ozone data, *Journal of American Statistical Association*, Vol.90, pp.969-983, 1995.
- 20) Niu, X-F, McKeague, I.W. and Elsener, J.B.: Seasonal space-time models for climate systems, *Statistical Inference for Stochastic Processes*, Vol.6, pp.111-133, 2003.
- 21) Granger, C.W.J. and Joyeux, R.: An introduction to long-range time series models and fractional differencing, *Journal of Time Series Analysis*, Vol.1, pp.15-30, 1980.
- 22) Hosking, J.R.M.: Fractional differencing, *Biometrika*, Vol.68, pp.165-176, 1981.
- 23) Gray, H.L., Zhang, N.F. and Woodward, W.A.: On a new definition of the fractional difference, *Mathematical Computation*, Vol.50, pp.513-529, 1988.

- 24) Woodward, W.A., Chang, Q.C. and Gray, H.L.: A k -factor long memory model, *Journal of Time Series Analysis*, Vol.19, pp.485-504, 1998.
- 25) Katayama, N.: Seasonally and Fractionally Differenced Time Series, *Hi-Stat Discussion Paper Series*, No.11, 2004.
- 26) Porter-Hudak, S. : An application of seasonal fractionally differenced model to the monetary aggregates, *Journal of American statistical Association*, Vol.85, pp.338-344, 1990.
- 27) Pfeifer, P. and Deutch, S.: A three-stage iterative procedure for space-time modeling, *Technometrics*, Vol.22, pp.35-47, 1980.
- 28) Pfeifer, P. and Deutch, S.: Seasonal space-time ARIMA modeling, *Geographical Analysis*, Vol.13, pp.117-133, 1981.
- 29) Hamilton, J.: A new approach to the economic analysis of nonstationary series and the business cycle, *Econometrica*, Vol. 57, pp.357-384, 1989.
- 30) Diebold, F.X. and Inoue, A.: Long memory and regime switching, *Journal of Econometrics*, Vol.105, pp.131-159, 2001.
- 31) Kim, C.-J. and Nelson, C.R.: *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, MIT Press, 1999.
- 32) Diebole, F.X., Lee, J.-H. and Weinbach, G.C.: Regime Switching with Time-Varying Transition Probabilities, in: *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press,, pp.283-302, 1994.
- 33) Rydén, T., Teräsvirta, T. and Åsbrink, S.: Stylized facts of daily return series and the hidden Markov model, *Journal of Applied Econometrics*, Vol.13, pp.217-244, 1998.
- 34) Bos, C.S., Franses, P.H. and Ooms, M.: Long Memory and Level Shifts: Re-Analyzing Inflation Rates, *Empirical Economics*, Vol.24, pp.427-449, 1998.
- 35) Granger, C.W.J. and Hyung, N.: Occasional Structural Breaks and Long Memory, *USCD Discussion Paper*, 99-14, 1999.
- 36) Gabriel, V.J. and Martins, L.F.: On the forecasting ability of ARFIMA models when infrequent breaks occur, *Econometrics Journal*, Vol.7, pp.455-475, 2004.
- 37) Haldrup, N. and Nielsen, M.Ø.: A regime switching long memory model for electricity prices, *Journal of Econometrics*, Vol.135, pp.349-376 , 2006.

- 38) Mandelbrot, B.B.: *Fractals; Form, Chance and Dimension*, San Francisco: Freeman, 1977.
- 39) Beran, J.: *Statistic for Long-Memory Process*, Chapman and Hall, 1994.
- 40) Geweke, J. and Porter-Hudak, S.: The estimation and application of long memory time series model, *Journal of Time Series Analysis*, Vol.4, pp.221-238, 1983.
- 41) Reisen, V.A.B.: Estimation of the fractional difference parameter in the smoothed periodogram, *Journal of Time Series Analysis*, Vol.15, pp.335-350, 1994.
- 42) Arteche, J. and Robinson, P.M.: Semiparametric inference in seasonal and cyclical long memory processes, *Journal of Time Series Analysis*, Vol.21, pp.1-25, 2000.
- 43) Reisen, V.A.B. and Lopes, S.: Estimation of parameters in ARFIMA processes: A simulation study, *Communications in Statistics*, Vol.30, pp.787-803, 2001.
- 44) McCoy, E.J. and Walden, A.T.: Wavelet analysis and synthesis of stationary long-memory processes, *Journal of Computational and Graphical Statistics*, Vol.5, pp.26-56, 1996.
- 45) Jensen, M.J.: Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter, *Journal of Forecasting*, Vol.18, pp.17-32, 1999.
- 46) Whitcher, B.: Wavelet-based estimation for seasonal long-memory processes, *Technometrics*, vol.46, pp. 225-238, 2004.
- 47) Lopes, S.R.C., Olbermann, B.P.A. and Reisen, V.A.: Comparison of estimation methods in non-stationary ARFIMA process, *Journal of Statistical Computation and Simulation*, Vol.74, pp.339-347, 2004.
- 48) Hamilton, J.: Analysis of time series subject to changes in regime, *Journal of Econometrics*, Vol.45, pp.39-70, 1990.
- 49) Redner, R.A. and Walker, H.F.: Mixtures densities, maximum likelihood and the EM algorithm. *SIAM Review*, Vol.26, pp.195-249, 1984.
- 50) Bordes, L., Chauveau, D. and Vandekerckhove, P.: An EM algorithm for a semi-parametric mixture model, *Computational Statistics & Data Analysis*, to appear, 2006.
- 51) Bonnie, K.R. and Ruey, S.T. : Bandwidth selection for kernel regression with long-range dependence errors, *Biometrika*, Vol.84, pp.791-802, 1997.

- 52) Silverman, B.W. *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.
- 53) Chung, C.-F.: A generalized fractionally integrated ARMA process, *Journal of Time Series Analysis*, Vol.17, pp.111-140, 1996.
- 54) Maddala, G.S. : *Introduction to Econometrics (Third edition)*, Willey, 2000.
- 55) Abdul, R.K.: Observations of PM10 Readings in Relation to Forest Fire Events from ASMA 's Continuous Air Quality Monitoring Stations, ASMA, 2000.
- 56) Department of Environment Malaysia: *A Guide to Air Pollutant Index (API) in Malaysia*, 2000.
- 57) Katayama, N. : Asymptotic Prediction Mean Squared Error for Strongly Dependent Processes with Estimated Parameters, *Hi-stat discussion paper series*, No.10, 2004.
- 58) Lancaster, T.: *The Econometric Analysis of Transition Data*, Cambridge University Press, 1990.
- 59) 大槻義彦, 室谷義昭: 新数学公式集初等関数, 丸善, 1991.

謝 辞

本論文を結ぶにあたり、本論文作成に際してご協力を頂いた方々へ感謝の意を申し上げます。京都大学工学研究科の小林潔司教授には、本論文を作成するにあたり終始適切なご指導、御助言を頂きました。また、研究活動を進めていく中で、マレーシアへの渡航を通じ今までに経験したことのない貴重な経験を得ることができました。深く感謝の意を申し上げます。京都大学工学研究科の松島格也准教授には、本論文の修正に際して、貴重な御助言、ご示唆を頂きました。厚く御礼申し上げます。京都大学工学研究科の大西正光助教授には、御多忙の中、研究に対する批評から、日常生活に至るまで大変お世話になりました。深謝の意を申し上げます。そして、計画マネジメント論研究室の諸兄には、本研究を取りまとめる上での多大な御協力を頂きました。また、日頃の研究に対する姿勢や生活態度から学ぶ点も多くありました。ここに深く感謝する次第です。