

機械学習を用いた地域農業動向予測モデルの
構築に関する研究

寺谷 諒

機械学習を用いた地域農業動向予測モデルの構築に関する研究

寺谷 諒

要旨

近年、農業従事者の高齢化や、後継者の不足、海外との価格競争による販売額の減少などが問題となっており、農家数は年々減少するとともに、農地の転用や耕作放棄地化が進んでいる。今後もこの傾向は今後も続くとみられ、農地保全に向けた対策を早急に実施していくことが求められている。

適切な農地保全対策を実施していくためには、まず、転用や耕作放棄地の発生という農地の土地利用状況の変化や、離農や規模縮小・拡大といった農家の営農状況の変化について、現状を把握するとともに、それらに関係している要因について詳細な分析を行っていく必要がある。また、それらの分析をふまえたうえで、将来の動向について予測を実施することが非常に重要である。

そこで、本研究では、農地の土地利用状況の変化や農家の営農状況の変化といった地域農業の動向に関する高精度な予測モデルを構築し、関連する要因を明らかにするとともに、将来予測を通じて、今後とり得るべき農地保全対策について検討を行うこととした。

また、農地保全の取組の推進は、全国共通の課題であり、それぞれ状況の異なる各地域で要因分析や予測を行い、有効な対応策を検討できるように、一連の手順に基づく定量的・統計的な手法の開発が必要である。

そこで、本研究では主に機械学習の手法を用いて、データを入手さえすれば、各地域で同様の分析や予測が可能であるような高精度かつ汎用性のあるシミュレーションモデルを提案することを目的とし、その手法に関する検討を行った。なお、分析は複数のスケールを考慮し、様々な粒度のデータを活用しながら多角的な分析を試みた。

まず、第一の分析では、大阪府の北大阪、東大阪地域を対象として、都市郊外部における農地の転用について、社会的要因を考慮に入れたうえで、関連する要因についての詳細な分析を行った。また得られた分析結果をもとに、ナイーブベイズと決定木を用いた農地転用予測モデルを構築し、将来予測シミュレーションを実施した。

結果として、区域区分や近傍の土地利用(農地、建物)が農地の変化に大きな影響を与えていることが分かった。また、構築したモデルの精度の検証のため、近

年の実際の農地の分布を予測する再現シミュレーションを実施したところ、正答率は70%以上の値をとり、先行研究と比較して高い精度となった。さらに、将来予測シミュレーションの結果、現在の状況が続いた場合、今後も多くの農地が転用され、大阪府の掲げる農地の保全目標値とは大きく乖離する可能性が高いことが示唆された。一方で、集団的な農地を優先的に規制するように区域区分の見直しを行った場合、転用される農地は少なくなり、一定の効果がみられることが分かった。

また第二の分析では、京都府綾部市を対象に、農村部における自己保全管理農地の発生に関する要因分析と予測モデルの構築を行った。農地一筆単位のデータとランダムフォレストの手法を用いて、より詳細な分析や予測モデルの構築を試みた。

結果、農地を保有している農家のタイプや、DID(人口集中地区)までの距離、農地貸借の有無が重要な要因であり、自己保全管理農地の発生に大きな影響を与えていることが分かった。また、要因分析によって得られた重要度の高い14変数を用いた予測モデルでは、交差検定による平均正答率が97%となり、自己保全管理農地の発生を高精度に予測できることが分かった。

さらに第三の分析では、農家の営農状況の変化について、離農に焦点をあて、茨城県の5市町村を対象に、市町村単位での高精度な離農予測モデルの構築を試みた。加えてモデルの構築を通じて、農家の離農と関連が高く、離農予測の際に重要である変数についても分析を行った。また、ロジスティック回帰とマイクロシミュレーションを組み合わせた新たな予測手法を提案した。

結果として、経営主の年齢、農業従事者数、総経営耕地面積や農作物の販売金額などが農家の離農に大きく影響していることが分かった。また、交差検定を通じてモデルの予測精度を検証した結果、いずれの市町村でも誤差は小さい値となり、高精度に農家の離農を予測できることが分かった。さらに将来予測シミュレーションの結果、多くの市町村で、これまでよりもさらに離農が進む可能性があることが示唆された。

以上のように、農地の土地利用状況の変化や、農家の営農状況の変化といった地域農業の動向に関して予測モデルを構築し、要因分析や将来予測を実施した。機械学習の手法を用いることで、高精度な予測モデルの構築が可能となり、本研究で提案した手法は、地域農業動向の予測を行ううえで、非常に有効であると考えられる。

また、本研究で提案した複数の手法を組み合わせることによって、さらに詳細な予測を実施でき、より具体的で実効的な農地保全対策の検討が可能になると考えられる。

例えば、100m メッシュによる農地の転用の予測と、農地一筆単位での自己保全管理農地の発生予測の手法を組み合わせることで、マクロとミクロなモデルを統合した農地の土地利用状況の変化の予測を実施することができる。また、農地一筆単位での自己保全管理農地の発生予測と、個々の農家の離農予測の手法を組み合わせることで、ミクロなレベルでのより精緻な地域農業動向予測が可能となる。

また、今後の農地保全対策の方向性として、広域レベルにおいては、区域区分を適切に見直し、特に集団的な農地を優先的に保全するための規制を行っていく必要があると考えられる。また、よりミクロなレベルでは、営農意欲の高い農家や将来の担い手への農地の集積を積極的に進めていくことが重要である。

さらに、各農家の離農の抑制に向けて、若い経営主や農業従事者を増加させていくために、農業経営の継承や、農業経営の維持・拡大に対する政策的支援ならびに新規就農支援をより重点的に実施していく必要がある。しかし一方で、現在の農業従事者の年齢構成では、農家が一定程度離農することはやむを得ない部分もあるため、離農に伴い出てくる多くの農地を、担い手に円滑に移行させるための対策を早急に実施し、地域に存在する農地を保全していくことが重要である。

また、実際には、地域ごとに農地の土地利用状況の変化や農家の離農の傾向に異なる点も見られるため、機械学習によるモデルを用いて各地域固有の特徴をふまえた分析や予測を行い、地域に応じた対応策を実施していくことが重要であると考えられる。

そして、本研究の意義としては、これまで定量的な分析が不十分であった農地の土地利用状況の変化や、農家の営農状況の変化について、上記に示したような機械学習の手法を用いて高精度な予測モデルを構築したことである。また本研究では、より詳細なモデルとして、農地一筆単位での土地利用状況の変化や、農家単位で離農の予測を行うような新たなモデルを提案した。このようなモデルを用いることで、従来よりもきめ細やかな農地保全対策の検討が可能になると考えられる。

さらに、本研究で提案したモデルは、データさえあれば、他地域においても同

様の分析や予測が可能な汎用性を有している。加えて、機械学習の手法を用いているため、地域固有の特徴をうまく抽出し予測を実施できるという利点があり、各地域の実情をふまえた対応策の検討とその実施に役立てることができると考えられる。

以上のように、本研究で提案したモデルを用いることで、適切な農地保全対策の検討が可能となり、今後の地域農業の在り方をより望ましい方向にもっていくことに貢献できると考えられる。

機械学習を用いた地域農業動向予測モデルの構築に関する研究

目次

第1章 はじめに	1
1.1 国内の農業の動向	1
1.2 農地保全に対する政策の動向	3
1.3 適切な農地保全対策の実施に向けて	3
1.4 既存研究	4
1.5 本研究の目的	5
第2章 都市郊外部における農地の転用の分析と予測シミュレーション	8
2.1 背景	8
2.2 目的	9
2.3 材料と手法	9
2.3.1 対象地域	9
2.3.2 使用データ	9
2.3.3 農地の変化に関連する要因の分析	11
2.3.4 予測モデルとシミュレーション	12
2.3.5 決定木分析による変数選択	12
2.3.6 ナイーブベイズモデル	13
2.3.7 モデル構築と再現シミュレーション	14
2.3.8 将来予測シミュレーション	15
2.4 結果	18
2.4.1 近傍の土地利用	18
2.4.2 区域区分	20
2.4.3 農地の変化と有意な相関があった項目	21
2.4.4 予測モデルによる再現シミュレーション	23
2.4.5 予測モデルによる将来予測シミュレーション	25
2.5 考察	27
第3章 農村部を中心とした自己保全管理農地の発生に関する分析と予測シミュレーション	30
3.1 背景	30

3.2	目的	31
3.3	材料と手法	32
3.3.1	対象地域	32
3.3.2	使用データ	33
3.3.3	自己保全管理農地の発生に関する要因分析	34
3.3.4	ランダムフォレストモデル	38
3.3.5	モデルの精度の検証	40
3.4	結果	41
3.4.1	自己保全管理農地の発生状況	41
3.4.2	自己保全管理農地の発生に関する重要な要因の算出	41
3.4.3	ランダムフォレストモデルの精度の検証	46
3.5	考察	48
第4章	農家の離農予測シミュレーション	52
4.1	背景	52
4.2	目的	54
4.3	材料と手法	54
4.3.1	対象地域	54
4.3.2	使用データ	55
4.3.3	離農予測シミュレーションのアルゴリズム	58
4.3.4	離農予測モデルの構築	59
4.3.5	各農家の離農予測	61
4.3.6	対象地域全体の離農農家数の算出	61
4.3.7	シミュレーションの試行回数と予測精度の検証	61
4.3.8	将来の離農農家数の予測	62
4.3.9	都道府県全体の離農農家数の予測手法に関する比較	63
4.4	結果	63
4.4.1	農家の離農の現状	63
4.4.2	離農予測において重要な変数	64
4.4.3	シミュレーションモデルの予測精度	73
4.4.4	将来における離農農家数と離農率	73
4.4.5	都道府県全体の予測精度	74

4.5 考察.....	78
第 5 章 結論	82
謝辞	93
引用文献	94

第1章 はじめに

1.1 国内の農業の動向

近年、農業従事者の高齢化や、後継者の不足、海外との価格競争による販売額の減少などが問題となっており、農家は年々減少するとともに、農地の転用や耕作放棄地化が進んでいる。農地の転用とは、農地を建物等の他の土地利用に変えることである。また、耕作放棄地とは、以前耕作していた土地で、過去1年以上作物を作付けせず、この数年の間に再び作付けする意思のない土地のことである(農林業センサスにおける定義)(農林水産省 2018a)。

農林業センサスで定義されている農業経営体の数は、2005年は200.9万経営体、2010年は167.9万経営体、2015年は137.7万経営体と一貫して減少傾向にある。

また、農林業センサス、耕地及び作付面積統計(農林水産省 2018b)によると、耕地面積は、1975年の557.2万haから2015年には449.6万haまで減少している。一方で、耕作放棄地面積は、1975年の13.1万haから2015年には42.3万haまで増加している(図1.1)。

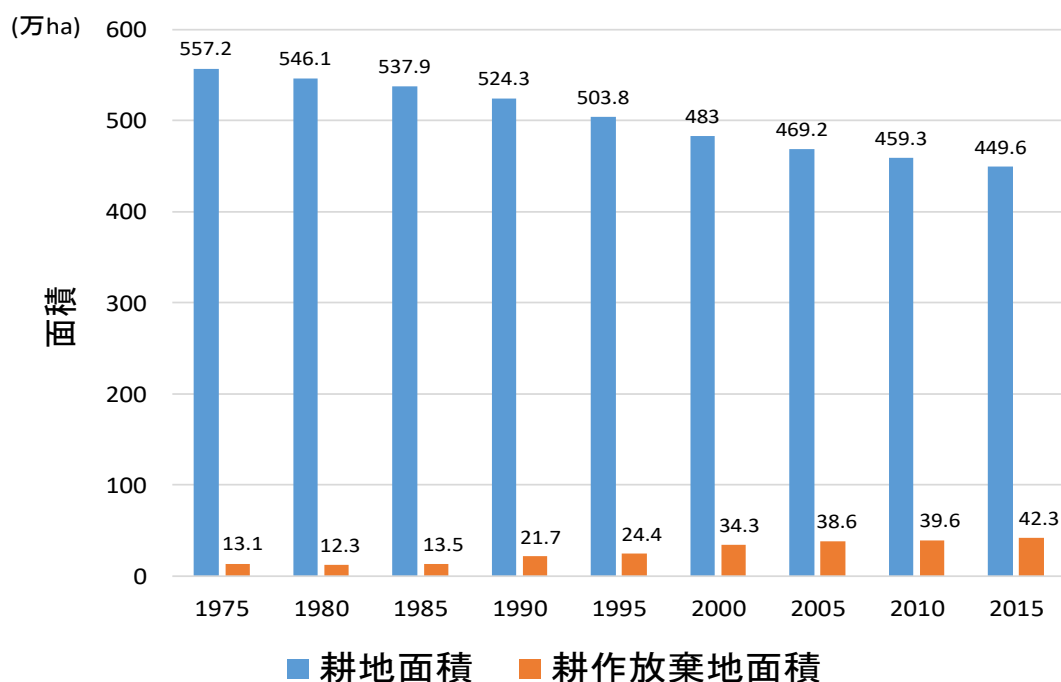


図 1.1 国内の耕地面積と耕作放棄地面積の推移

農業や農村、農地は食料の生産のみならず、都市環境保全、防災、文化的機能などの多面的機能を有しているため、保全を図っていく必要がある。

1999年に施行された「食料・農業・農村基本法」(農林水産省 2008)においても、国土の保全、水源のかん養、自然環境の保全、良好な景観の形成、文化の伝承等農村で農業生産活動が行われることにより生ずる食料その他の農産物の供給の機能以外の多面にわたる機能として、農業の多面的機能がとりあげられ、条文に明記されている。また、具体的には、以下のような機能があげられる(農林水産省 2018c, 農林水産省 2018d)。

①国土の保全機能

- 洪水や地すべり、土砂崩れなどの発生の防止

②水源のかん養機能

- 地下水のかん養や、川の流れの安定化

③自然環境の保全機能

- 微生物による有機物の分解や、生物の生息場所の提供

④良好な景観の形成機能

- 農村における景観の維持・保全

⑤文化の伝承機能

- 農業と関係の深い芸能・祭りや、農業上の技術や地域独自の知恵等の伝承

⑥保健休養機能

- 人々が農業文化や農村の文化・自然にふれる機会の提供

⑦地域の維持活性化

- 農業生産・加工・流通・販売等の活動を通じた地域社会の維持活性化

⑧食料安全保障

- 自国での農作物の生産を通じた、現在及び将来の食料供給リスクの軽減

このように、農業は様々な機能を有している。農業経営体数や耕地面積が減少し、転用や耕作放棄地が増加することは、これらの機能が徐々に失われていることを意味し、その影響は決して小さくない。

1.2 農地保全に対する政策の動向

農業経営体数や耕地面積の減少の傾向は今後も続くともみられ、農地の保全に向けた対策を早急に実施していく必要がある。

今後の農地保全に対する政策の方向性として、法律等に基づく農地転用制度の適切な運用を図っていくことや、農家の離農に伴い出てきた農地や、現存する耕作放棄地を担い手に集積していくということが考えられる(担い手とは、一定の農地を保有し市町村から認定された認定農業者や、集落等の単位で農家が共同で営農する営農組合や、農業法人を指す)。

特に、担い手への集積については、2013年に「農地中間管理事業の推進に関する法律」(総務省 2018a)が制定され、農地の貸借・売買の仲介を行う組織として、農地中間管理機構が設置された。この組織の活用によって、農地の集積を促進させるとともに、耕作放棄地の減少と発生を防止することが狙いとなっている。なお、政府は、今後10年間で担い手への集積面積を全農地面積の80%に引き上げる目標を立てている。しかし、2017年3月末時点では、54%となっており(農林水産省 2018e)、今後、更なる農地集積の加速化が必要である。

また、農業者の所得向上を図るためには、農業者が自由に経営展開できる環境を整備するとともに、農業者の努力では解決できない構造的な問題を解決していくことが必要であるとして、2016年に「農業競争力強化プログラム」が策定された(農林水産省 2017a)。このプログラムでは、生産資材価格の引下げ、農作物の流通・加工の構造改革、収入保険制度の導入、人材力の強化などを実施することとしている。

さらに、耕作放棄地の解消を促進させることを狙いとして、2017年度から、条件に合致した遊休農地(耕作放棄地)の固定資産税を約1.8倍に引き上げるといった政策が実施されている。

1.3 適切な農地保全対策の実施に向けて

政府が実施している上記のような政策を活用しながら、農地の保全を進めていく必要がある。

しかし一方で、各地域において、今後、農地や農家の状況はどのように変化していき、それに対応した真に有効な対策は何であるか、また、対策を実施した場合に、具体的にどの程度の効果が見込めるのかといったような点は十分に明ら

かにされていないという現状もある。

適切な農地保全対策を明らかにしていくためには、転用や耕作放棄地の発生という農地の土地利用状況の変化や、農家の離農や規模縮小・拡大といった営農状況の変化について、現状を把握するとともに、それらに関連している要因について詳細な分析を行っていく必要がある。また、それらの分析結果をふまえながら、将来の動向について予測を行うことが非常に重要である。

つまり、過去の土地利用を分析し、土地利用変化がどのように生じているのかというメカニズムを把握するとともに、それらをふまえたうえで将来の土地利用変化の予測を行っていくことが必要だと考えられる。そして、予測をもとに適切な土地利用計画・都市計画を策定し、望ましいと考えられる土地利用形態に向けて誘導を行っていくことが重要である。

由良ら(2002)が述べるように、土地利用予測は、都市計画や各種政策を策定するのに極めて重要な資料となる。

また、農家の営農状況の変化についても、栗原ら(2014)が述べるように、どのような農家が離農しているかを把握することは、農地集積のために離農を促進する政策にしる、集落機能維持のために離農を防ぐ政策にしる地域が含有する特徴や課題に沿った政策を策定するために不可欠である。また、将来的に離農する可能性が高い農家を予測することで、今後どの程度、耕地面積が減少するリスクがあり、それに対してどのような対応策を実施していけばよいかという点の検討が可能となる。

つまり、農地の土地利用状況の変化や、離農等の農家の営農状況の変化について、現状や要因分析ならびに将来予測を実施することによって、各地域において、今後とり得るべき農地保全対策についての十分な検討が可能になり、有効な対策の実施につながると考えられる。

1.4 既存研究

転用や耕作放棄地化といったいわゆる農地の変化について分析している研究は、国内外に一定程度存在する(Valbuena *et al.* 2010, Ligtenberg *et al.* 2001, 山本ら 1999a)。しかし、既存研究の問題点として、国内の農地を対象とした研究については、農地の変化に関連する社会的要因(社会基盤、法律、人口、農家の属性等)に関する分析が十分に行われていないことがある。

また、農地の変化に関する定量的な予測モデルを構築した研究もいくつか見られるが、対象地域全体における転用や耕作放棄地面積の総量を予測する研究が多く(松井ら 2014, 吉田ら 2005)、どのような場所のどのような農地で転用や耕作放棄地が発生するかという空間的・地理的な面まで考慮した研究は少ない。

さらに、そのような空間的・地理的な面を考慮した既存研究のモデルでは、予測精度が低く、現実の土地利用を再現した場合の正答率は、60%前後(山本ら 1999a, 山本ら 1999b)や、50%以下(池田ら 2007)となっており、モデルの実用性に課題がみられる。

また、農家の営農状況の変化については、農家の離農を対象にした研究がいくつか見られる。しかし、実態調査によって接近した研究は散見されるが(山本 2003, 井上 1989-1990)、定量的に離農要因を分析した研究は少ないことが指摘されている(栗原ら 2014)。

そのような中で、定量的アプローチによる研究として、主に農林業センサス個票を用いて、離農要因の分析(栗原ら 2014, Kurihara *et al.* 2013)や離農予測(将来の農家数の予測)を行った研究(安武 2018, 安武・千田 2013, 渡部ら 2012, 細山 2013)も見られる。しかし、分析のスケールが全国、地方、都道府県や都道府県内の地域エリアといった広範なレベルであり、よりスケールの小さい市町村以下の単位での詳細な分析や予測はほとんど行われていない。また使用する変数も、最初から決め打ちをしているか、分析者が試行錯誤的に選択している研究が多く、統計的手法を用いて厳密な変数選択を行っている研究は非常に少ない。

さらに既存研究では、構築したモデルについて、予測精度の検証を行っている研究も少ない。そのため、モデルの妥当性や、各市町村でどのような変数が離農と高い関連性を有しているかという点も明確に把握できていない。

1.5 本研究の目的

以上をふまえ、本研究の目的は大きく分けて2つとした。

1. 高精度かつ汎用性のある地域農業動向予測モデルの構築
2. 各地域において、今後とり得るべき農地保全対策の解明

まず、1については、農地の土地利用状況の変化や農家の営農状況の変化といった地域農業の動向について、現状把握を行ったうえで、要因分析や将来予測を実施するための予測モデルを構築した。

当然ではあるが、適切な農地保全対策の検討を行うためには、高い精度を有する予測モデルが必要となる。そのため、本研究では、より高精度な予測モデルを構築することを目指し、そのための手法について検討を行った。

また、農地保全の問題については、全国共通の課題であり、それぞれ状況の異なる各地域で要因分析や予測を行い、有効な対応策を検討できるように、一連の手順に基づいた定量的・統計的な手法の開発が必要である。つまり、それぞれの地域でモデルを構築し、個別に予測が実施できるような手法が求められる。

そのため、本研究では、データさえ入手すれば、各地域で同様の分析や予測が可能であるような、より汎用性の高いモデルの構築を目的とした。

また、2については、1で構築したモデルにより、農地の土地利用状況の変化や農家の営農状況の変化に関する要因分析や将来予測を行い、その結果をふまえたうえで、今後とり得るべき農地保全対策について明らかにすることを目的とした。

そして、本研究の目的をふまえ、具体的には、以下のようにいくつかの地域スケールや問題を取りあげて、分析を実施した。

- ①都市郊外部の広域を対象として、農地の土地利用状況の変化(農地の転用)について、社会的要因を考慮に入れたうえで、関連する要因について詳細な分析を行った。また得られた分析結果を用いて、より精度の高い農地変化の予測モデルを構築し、シミュレーションを実施した。
- ②また、よりミクロなモデルとして、農村部が大半を占める市町村を対象に、農地一筆単位のデータを用いて、農地の土地利用状況の変化(自己保管理農地の発生)に関する詳細な分析を行うとともに、予測についても農地一筆単位で実施が可能なモデルの構築を試みた。
- ③さらに、農家の営農状況の変化について、離農に焦点をあて、個々の農家の詳細な属性をふまえた市町村単位での離農予測モデルを構築した。また、モデルの構築を通じて、農家の離農と関連が高く、離農予測の際に重要である変数についても分析を行った。

そして最終的には、本研究において構築した予測モデルの手法や分析結果について整理を行い、全体としての考察を行った。

なお、本論文の構成は以下の通りである。上述のとおり、第1章で研究にあたっての背景や目的を示す。第2章では、都市郊外部における農地の土地利用状況の変化(農地の転用)を対象にした分析の結果を示す。第3章では、農村部における農地の土地利用状況の変化(自己保全管理農地の発生)に関する分析の結果を示す。第4章では、農家の離農に関する分析の結果を示す。そして、第5章でこれまでの章の内容・結果を総括し、高精度な地域農業動向予測モデルの構築のための手法、ならびに今後とり得るべき適切な農地保全対策についての全体的な考察を行う。

第2章 都市郊外部における農地の転用の分析と予測

シミュレーション

2.1 背景

近年、都市の近郊においては、多くの農地が転用され、宅地や道路等の他の土地利用に変化している。都市近郊に存在する農地は、食料の生産に加え、都市のヒートアイランド現象の緩和や緑地空間の提供、農業体験等による生産者と消費者の交流の場の提供などの多面的機能を有している(農林水産省 2018d)。そのため、保全を図っていくことが求められており、農地の減少は、非常に深刻な問題となっている。

農地の減少には、多くの要因が影響していると考えられるが、1つには、市街地が開発によって、無秩序に拡大していくスプロール現象が生じていることがあげられる。平地が少ない日本では、このスプロール現象による土地の混在は、大半の都市の周辺部でみられている(田代 2005)。さらに、他の要因としては、農業従事者の高齢化や後継者の不足、海外との価格競争による販売額の減少などの農業の衰退に関する問題があるといえる。

国内の耕地面積は第 1 章で述べたように、年々減少している一方で、宅地面積は、1961 年の 60 万 ha から 2003 年の 150 万 ha まで増加している(八木 2005)。

そして、農地の転用等による耕地面積の減少を防止する施策を考える際には、耕地面積の減少がどのように生じているのかというメカニズムを把握するとともに、それを踏まえたうえで将来の土地利用変化の予測を行っていくことが重要である。

既存研究でも、農地を対象に含めた土地利用変化の分析や、予測モデルの構築により、土地利用変化の予測を行っている研究はいくつか存在する。

しかし第 1 章で述べたように、既存研究の問題点として、国内の農地の変化に関連する社会的要因(人口、社会基盤、法律等)に関して、分析が十分に行われていないこと、また土地利用変化予測モデルを構築する際の変数選択においても、社会的要因を詳細に考慮していないという点があげられる。また、構築されたモデルの予測精度も決して高いとはいえない。

2.2 目的

そこで、本研究の目的として、都市郊外部を対象に、農地の転用に関連する要因について、社会的要因も考慮に入れたうえで詳細な分析を行うこととした。また得られた分析結果を用いて、より精度の高い農地の転用予測モデルを構築し、シミュレーションを通じて、将来の動向予測と今後の農地保全対策の在り方について検討を行った。

2.3 材料と手法

2.3.1 対象地域

本研究では対象地域として、大阪府をとりあげた。大阪府でも耕地面積の減少が生じており、耕地面積は 1990 年に 18,080ha であったものが、2016 年には 13,100ha にまで減少している(大阪府 2018)。大阪府の野菜生産額は、年間約 81 万人分の消費量に相当しているといわれ、果たしている役割は決して小さくはない(樋口 2008)。

また、大阪府が策定している「大阪府国土利用計画(第 3 次)」では、府内の各地域において土地利用構成が異なる特徴を示していることから、地域の特性をふまえて、府全体を 5 つに区分している(大阪府 2001)。本研究でもその地域区分に従うこととした。本研究では、5 つの区分のうち、大阪市の都心部に比較的近い、北大阪地域、東大阪地域を分析の対象地域とした(図 2.1)。

2.3.2 使用データ

分析には、主に国が公表している統計資料(表 2.1)を用いた。土地利用のデータとして、国土数値情報(国土交通省 2018)の中の「土地利用細分メッシュ」を使用した。この「土地利用細分メッシュ」のデータは、1 辺 100m の正方形のメッシュを 1 単位とし、各メッシュそれぞれにおいて、農地や建物といった土地利用の属性が 1 つずつ与えられている。本研究では、この 1 メッシュを 1 セルと呼び、土地利用の基本単位とした形で分析を行った。データは 3～11 年の間隔で公表されており、本研究では 1997 年、2006 年、2009 年のデータを用いた。「土地利用細分メッシュ」では、土地利用を 11 種類に分類しているが、本研究では、6 種類に再分類を行い、分析を実施した。6 種類の内訳は表 2.2 の通りである。

また、国勢調査(総務省 2018b)や農林業センサスのデータについては、上述の

土地利用細分メッシュと同時期のデータが無いため、2000年、2005年、2010年のデータを用いた。

また、データの地理的・空間的な解析には、Esri社のArcGIS10.2.2 for Desktop、統計解析やシミュレーションには、オープンソースソフトウェアのR3.2.1(The R Foundation 2018)や、決定木による変数の重要度の算出アルゴリズム(Breiman *et al.* 1984)を実装しているSelford System社のCART6、株式会社構造計画研究所のArtisoc2.0を使用した。

本研究での農地の変化の分析では、農地が農地のまま維持されるか、農地が他の土地利用に転用されるかという農地からの遷移のみに焦点をあてた。実際は、他の土地利用が農地に変わるという遷移も起こり得るものの、そのようなケースは非常に少ないため、今回の分析では対象としなかった。

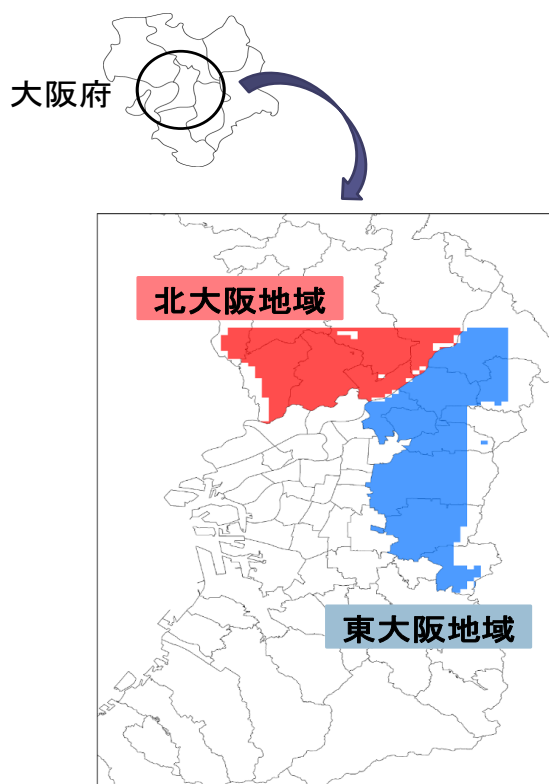


図 2.1 対象地域(大阪府)

表 2.1 使用する統計データ

	データの内容(例)
国土数値情報	土地利用細分メッシュ、標高、地価公示 etc.
国勢調査	人口 500m メッシュ
農林業センサス	農家数、農業従事者の平均年齢 etc.

表 2.2 土地利用の分類

	具体的な土地利用(例)
農地	田・その他の農用地
建物	建物用地・幹線交通用地・その他の用地
森林	森林
荒地	荒地
水域	河川地および湖沼・海浜・海水域
その他	ゴルフ場

2.3.3 農地の変化に関連する要因の分析

本研究では、まず、農地の変化に関連する要因についての分析を実施した。本研究では、地理的要因、社会的要因あわせて計 33 項目について、農地の変化との関連性についての分析を行った。分析した項目の一覧は表 2.3 のとおりである。

農地の変化は、1997 年、2006 年のデータを用いて、1997 年に農地であったセルが、2006 年に農地のまま維持されているか、他の土地利用に転用されているかという 2 つのパターンを分析対象とした。

表 2.3 分析項目一覧

地理的要因：

近傍の土地利用(農地、建物、森林、荒地、水域)、標高、傾斜度

社会的要因：

[法律・施設関係]

区域区分、駅までの距離、主要幹線道路までの距離、地価変化率、工業用地までの距離、人口変化率、医療施設までの距離、役所までの距離、教育施設までの距離、文化施設までの距離、その他の公共施設(警察、消防、福祉施設等)までの距離、都心中心部(地価最高点)までの距離

[農業関係]

農業従事者の平均年齢、基幹的農業従事者の平均年齢、法人化している農家数、1農家あたりの平均耕地面積、耕作放棄地面積割合、主業農家の割合、準主業農家の割合、副業的農家の割合、同居農業後継者がいる農家の割合、同居農業後継者の平均年齢、他出農業後継者がいる農家の割合、他出農業後継者の平均年齢、1農家あたりの農業投下労働の平均、1農家あたりの事業収入の平均

2.3.4 予測モデルとシミュレーション

次に、農地の変化に関連する要因の分析結果をもとに、予測モデルを構築した。既存研究では、ニューラルネットワークや判別分析、セル・オートマトン、決定木などを用いた土地利用変化予測モデルが見られるが(山本ら 1999a, 山本ら 1999b, 池田ら 2007, 三木ら 2004)、本研究では、ナイーブベイズの手法を用いた予測モデルを構築し、また、モデルの変数の選択に決定木分析を使用することを提案した。

2.3.5 決定木分析による変数選択

2.3.3 で述べた農地の変化に関連する要因の分析にて、有意な相関がみられた項目をモデルに使用する変数の候補とし、さらにその中から、特に重要な変数を抽出してモデルに使用する変数を絞り込んだ。本研究では、決定木分析を用いて、各変数の重要度を比較し、変数選択を行った。

決定木は、機械学習のアルゴリズムの 1 つであり、木構造の形をした規則を作成し、データを分類する手法である。決定木の利点としては、学習過程や結果の理解が容易であることがよくあげられるが、それに加えて各変数の相対的な重要度を算出できるということがある。本研究では、この決定木の特徴を生かし、

算出された変数の重要度をもとに変数選択を行った。

そして、決定木に用いるアルゴリズムとしては、CARTを用いた(Breiman *et al.* 1984)。CARTでは、ジニ係数(Gini Index)と呼ばれる指標を用いて分岐変数を求め、木を分割していく。例えば、 K 個の分類クラスがあるとして、あるノード t で、 j 番目のクラス c_j となる確率を $P(c_j|t)$ とすると、ジニ係数は以下の式で算出される。

$$GI(t) = 1 - \sum_{j=1}^K P^2(c_j|t) \quad (2.1)$$

そして、決定木全体のジニ係数が最小となるように、分岐変数を選択し、木を分割していく。

変数の重要度の算出は、決定木で実際に採用された分岐(第一分岐)と、第一分岐とほぼ同じような結果を与える代理分岐を使って算出される。変数ごとに第一分岐や代理分岐として採用された各分岐でのジニ係数の改善度(減少量)の値を合計する。そして、改善度の合計値が最も高かった変数の重要度の値を 100 とし、各変数の相対的な重要度の値を算出することで、変数間の比較を行う(Breiman *et al.* 1984)。

2.3.6 ナイーブベイズモデル

ナイーブベイズとは、機械学習のアルゴリズムの 1 つであり、ベイズの定理を用いるとともに、変数(特徴ベクトル)間に条件付き独立性の仮定をおいた分類アルゴリズムである(Mitchell 1997)。

ナイーブベイズのアルゴリズムの概要を以下に示す。分類クラスを c 、分類に用いる変数(特徴ベクトル)を \mathbf{x} とすると、 $P(c|\mathbf{x})$ を最大化するクラス \hat{c} を選べばよい。なお、以下の式での $\operatorname{argmax}_c f(c)$ は、 $f(c)$ が最大となる時の c を意味する。

$$\hat{c} = \operatorname{argmax}_c P(c | x_1, \dots, x_n) \quad (2.2)$$

$$= \operatorname{argmax}_c \frac{P(c)(x_1, \dots, x_n | c)}{P(x_1, \dots, x_n)} \quad (2.3)$$

$$\propto \operatorname{argmax}_c P(c)P(x_1, \dots, x_n | c) \quad (2.4)$$

ただし、式(2.4)にある尤度を推定するのは困難なため、各変数間に条件付き独立性を仮定し、式(2.5)の右辺で尤度を置き換える。式(2.5)の関係をを用いると式(2.4)は式(2.6)のように変形できる。分類されるクラスの決定は式(2.6)を用いて求める。

$$P(x_1, \dots, x_n | c) = \prod_{i=1}^n P(x_i | c) \quad (2.5)$$

$$\hat{c} = \operatorname{argmax}_c P(c) \prod_{i=1}^n P(x_i | c) \quad (2.6)$$

ナイーブベイズのモデルの利点としては、単純なモデルであるため、アルゴリズムの理解が容易であること、また、モデルが推定すべきパラメータが少ないため、比較的少数の訓練データでも十分に学習が可能であること、さらに、訓練データの偏りに影響されにくく、安定したモデルが構築されやすいことである。また、変数間の依存性を考慮しないモデルであるものの、実際には高い精度の予測が可能なモデルであり、その理由についても理論的な裏付けがなされている(Zhang 2004)。本研究では、上述したナイーブベイズによるモデルを構築し、農地の変化のシミュレーションを実施した。

2.3.7 モデル構築と再現シミュレーション

上述した手法を用いてモデルを構築するが、本研究では、決定木分析によって重要度の高かった上位5項目を変数として選択し、使用することとした。

また、ナイーブベイズモデルは、変数間に独立性を仮定したものであるため、変数として採用する上位5項目の中で、変数間で相関係数が0.9を超えるものについては、重要度の低い一方の変数を除去した。

本研究では、ナイーブベイズモデルの分類クラス c は、農地が転用されたか転用されなかったかの2クラスとし、式(2.6)の $P(c)$ (事前分布)については、転用ありとなしのそれぞれのクラスで0.5とする一様分布とした。これは、訓練データにおいて、分類するクラスごとのデータ数に大きな偏りがあると、一方のクラスに過度に分類されやすいモデルが構築される可能性があるためである。

さらに、変数(特徴ベクトル) \mathbf{x} の分布については、質的変数の場合は、訓練データでの各値(カテゴリ)の相対度数を、母集団での各値の生起確率とみなし

分布を推定した。また、量的変数の場合は、カーネル密度推定を用いて分布を推定した。

カーネル密度推定は、母集団の分布に特定の分布型を仮定しない手法であり、確率密度関数は以下の式で表される。

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (2.7)$$

ここで、 n はサンプルサイズ、 h はバンド幅と呼ばれる平滑化パラメータ、 K はカーネル関数、 X は観測値である。本研究では、バンド幅については、サンプルサイズと標準偏差と四分偏差で計算される Silverman の経験則(Silverman 1986, pp.47-48 (3.30) (3.13)式)をもとに決定し、カーネル関数は、正規分布の確率密度関数を使用した。上記のような一連の工夫を行い、モデルを構築した。

シミュレーションについては、まず過去から現在までの土地利用変化の予測を行う再現シミュレーションを実施し、モデルの精度を検証した。モデルの構築にあたり、1997年から2006年にかけての農地の変化のデータを訓練データとして用いてモデルを作成し、2006年から2009年にかけての実際の土地利用変化を予測した。

シミュレーションにより、各農地セルにおいて、転用されるか転用されないかが予測されるが、それらのうち、現実の土地利用の変化(転用の有無)と一致したセルの割合を正答率という形で算出し、モデルの精度を評価した。予測の対象となった農地セルの数を対象セル数、そのうち、現実の土地利用と一致したセル数を正解セル数として、正答率は以下の式で算出される。

$$\text{正答率 (\%)} = \text{正解セル数} / \text{対象セル数} \times 100$$

2.3.8 将来予測シミュレーション

2.3.7 で示した再現シミュレーションでモデルの精度を検証した後に、将来の農地の変化を予測するシミュレーションを実施した。

大阪府では、2010年に「大阪府国土利用計画(第四次)」を策定し、表2.4のように、将来の農地面積の数値目標を定めている(大阪府2010)。農地法に基づく農地転用許可制度の運用などを通じ、農空間の保全・活用を推進し、2007年から2020年までの大阪府全体における農地の減少率を7%程度に抑えることを目標としている。

本研究では、上述した大阪府の実際の目標をふまえたうえで、表 2.5 のように 4 つのシナリオを作成し、将来予測を実施した。訓練データに 1997 年から 2009 年のデータを用い、2009 年以降の将来の農地の変化(2021 年頃まで)を予測した。

今回は、農地のまとまり具合(近傍の農地セル数)に基づき区域区分の見直しを行うシナリオを作成した。近傍の農地セル数と区域区分を取り上げたのは、次節の分析結果で示すように、2 地域のいずれにおいても、重要度の高い変数であり、また、大阪府の国土利用計画でも区域区分の厳格な適用が謳われているためである。

シナリオ 1 は現在の区域区分のままである場合、シナリオ 2~4 は近傍の農地セル数に応じて各農地の区域区分を見直した場合である。2~4 のいずれも、集団的な農地について優先的に転用規制を実施し、一方で孤立分散した農地は転用規制を弱くするというシナリオである。優先的な規制の対象とする農地のまとまり具合(近傍の農地セル数)と、適用する区域区分を変えてシナリオを作成した。各シナリオでの規制の強弱は、表 2.5 に記載のとおりである。区域区分の変更は、次節で示す分析結果により、農用地区域、市街化調整区域、市街化区域の順に転用規制の効果が強いと考え、このようなシナリオとした。

表 2.4 大阪府全体の農地面積の目標

	面積(ha)		減少率(%) (減少面積(ha))
	2007 年	2020 年	
全農地	14360	13360	7.0 (1000)
市街化区域	4130	3320	19.6 (810)
市街化調整区域	10230	10040	1.9 (190)

表 2.5 予測シミュレーションにおけるシナリオ

シナリオ 1	現在の区域区分を変更しない場合
シナリオ 2 (転用規制・弱)	近傍の農地セル数が 2 以上の農地のうち、市街化区域にあるものを 市街化調整区域 に、市街化調整区域にあるものを 農用地区域 に編入させる。 また、近傍の農地セル数が 1 以下の農地を全て 市街化区域 に編入させる。
シナリオ 3 (同・中)	近傍の農地セル数が 1 以上の農地のうち、市街化区域にあるものを 市街化調整区域 に、市街化調整区域にあるものを 農用地区域 に編入させる。 また、近傍の農地セル数が 0 の農地を全て 市街化区域 に編入させる。
シナリオ 4 (同・強)	近傍の農地セル数が 1 以上の農地を全て 農用地区域 に編入させる。 また、近傍の農地セル数が 0 の農地を全て 市街化区域 に編入させる。

注) 近傍の農地セル数が 2 以上(シナリオ 3、4 は 1 以上)の農地で、すでに農用地区域にあるものは、そのままとする。また、同 1 以下(シナリオ 3、4 は 0)の農地で、すでに市街化区域にあるものはそのままとする。

2.4 結果

2.4.1 近傍の土地利用

農地の変化に関連する要因の分析のうち、主な分析結果として地理的要因である近傍の土地利用(農地、建物)をとりあげる。

近傍のセルの範囲は、セル・オートマトンにおけるムーア近傍を採用し、図 2.2 のように周囲の 8 セルとした。これら周囲の土地利用が、農地の変化に影響を与えているかについて、分析を行った。

図 2.3 は近傍の農地セル数ごとの転用率、図 2.4 は近傍の建物セル数ごとの転用率を示したものである。どちらも、1997 年から 2006 年にかけての転用率を示しているが、それらは以下の式によって算出される。

転用率 (%) = (1997 年時点の農地セルのうち、2006 年までに他の土地利用に転用されているセル数) / (1997 年時点の農地セル数) × 100



図 2.2 近傍の範囲(ムーア近傍)
(周囲 8 セルを近傍の範囲とする)

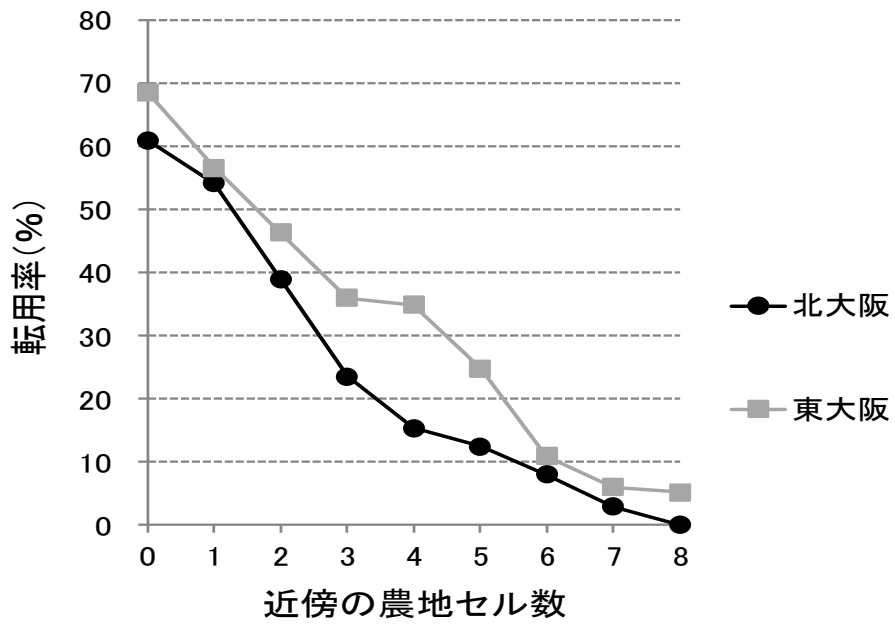


図 2.3 近傍の農地セル数と転用率

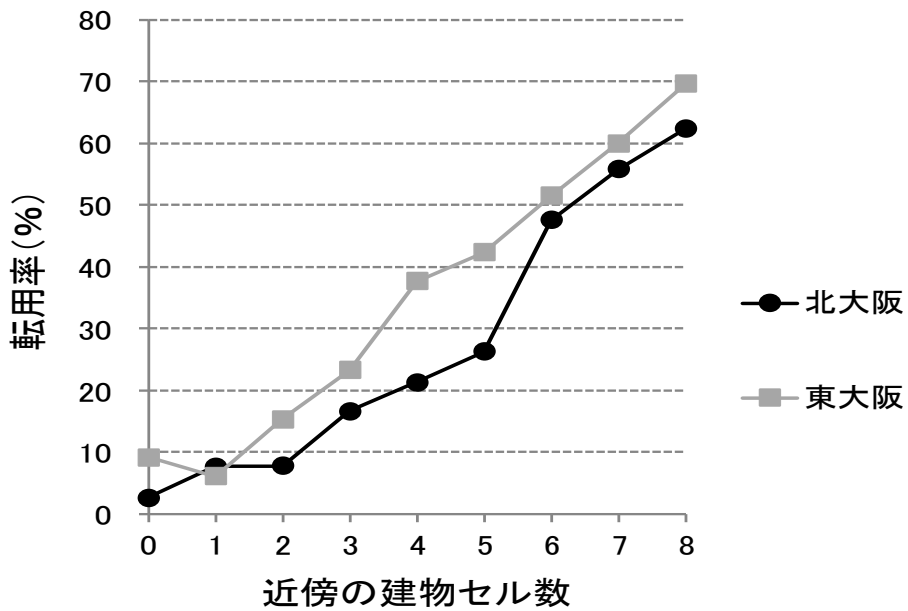


図 2.4 近傍の建物セル数と転用率

グラフより、近傍の農地セル数が多いほど転用率が低くなっている。また一方で、近傍の建物セル数が多いほど、転用率が高くなっている。また、転用されたセルと転用されなかったセルとで、近傍の農地セル数、近傍の建物セル数に差があるかを、マン・ホイットニーのU検定を用い、「帰無仮説 H0：転用されたセルと転用されなかったセルで近傍の農地(建物)セル数に差がない」、「対立仮説 H1：転用されたセルと転用されなかったセルで近傍の農地(建物)セル数に差がある」として検定を実施した。検定の結果、北大阪、東大阪の両地域で近傍の農地セル数、近傍の建物セル数ともに有意水準 5%で有意となった。

2.4.2 区域区分

次に、社会的要因である区域区分と農地の変化についての分析結果を述べる。区域区分とは、土地の開発や用途の規制のためにゾーニング(区分分け)を行ったもので、国内の農地に関しては、「都市計画法」(1919年制定)(総務省 2018c)と「農業振興地域の整備に関する法律」(1969年制定、以下農振法)(総務省 2015)によって規制がなされている。

都市計画法では、都市計画区域に設定されている地域を、市街化区域と市街化調整区域に区分することができるとしている。市街化区域は、おおむね 10 年以内に市街化を図るべき地域とされ、農地の転用は届出のみで原則自由となっている。市街化調整区域は、市街化を抑制すべき区域であり、転用には行政の許可が必要で、定められた条件を満たしていると判断された場合にのみ転用を行うことができる。また、農振法においても、農業の振興が必要であるとされる地域を農業振興地域に指定し、さらに農業振興地域の内部に農用地区域を定めることができるとしている。農用地区域では、原則として農地の転用が不可となっている。

本研究では、市街化区域、市街化調整区域、農用地区域の 3 つの区分と農地の変化の関連性について分析を実施した。なお農用地区域は、すべて市街化調整区域内に存在しているため重複が発生するが、本研究では、市街化調整区域かつ農用地区域に属しているセルを「農用地区域」に分類し、市街化調整区域に属しているものの、農用地区域に属していないセルを「市街化調整区域」に分類した。図 2.5 は区域区分別の 1997 年から 2006 年にかけての転用率を示したものである。転用率の算出式は、2.4.1 に示したものと同様である。

図より、市街化区域で最も転用率が高く、市街化調整区域、農用地区域とな

るにしたがって、転用率が下がっていくことがわかる。また、「帰無仮説 H0 : 区域区分の違いによって転用率に差がない」、「対立仮説 H1 : 区域区分の違いによって転用率に差がある」としてカイ二乗検定を用いて検定を行った。検定の結果、両地域でともに有意水準 5%で有意となった。

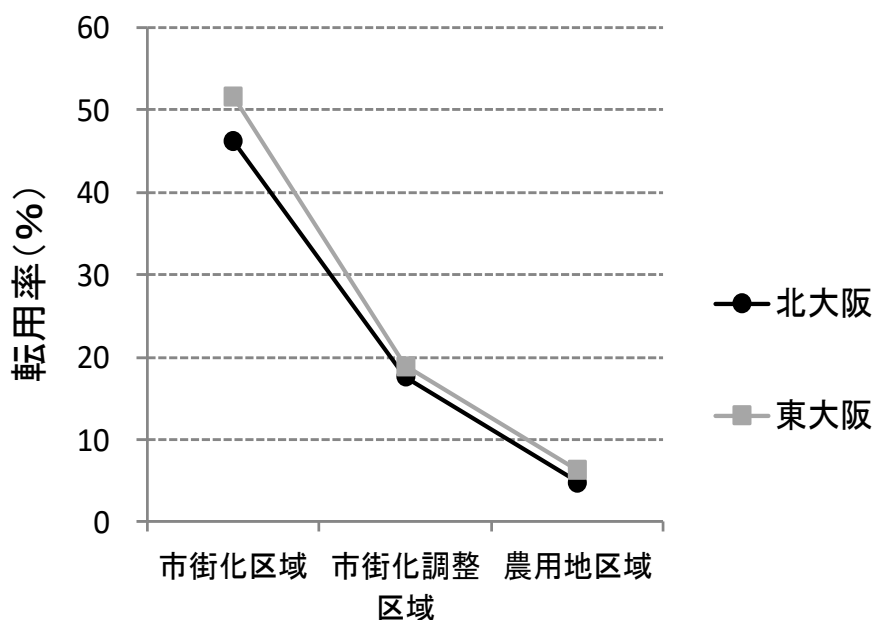


図 2.5 区域区分と転用率

2.4.3 農地の変化と有意な相関があった項目

上記の項目と同様の方法で、残りの項目についても分析を行い、検定を実施した結果、対象とした 33 項目のうち、いくつかの項目で農地の変化と有意な相関がみられた。表 2.6 は、1997 年から 2006 年までの農地の変化と有意な相関がみられた項目を地域別に示したものである。なお、表 2.6 に示した分析項目のうち、区域区分については、カイ二乗検定を実施し、地理的要因と区域区分を除く社会的要因(法律・施設関係)の各項目については、マン・ホイットニーの U 検定を実施した。また、社会的要因(農業関係)の項目については、農業集落単位のデータとなるため、各農業集落の転用率と相関があるかを、相関係数の t 検定(データが正規分布に従わない場合は、スピアマンの順位相関係数の検定)を用いて検定を行った。いずれも、有意水準 5%で検定を実施した。

表 2.6 農地の変化と有意であった項目
 (○：検定の結果、有意であった項目)

	北大阪	東大阪
近傍の建物セル数	○	○
近傍の農地セル数	○	○
近傍の森林セル数		○
近傍の荒地セル数	○	
近傍の水域セル数	○	
標高	○	○
傾斜度	○	○
区域区分	○	○
駅までの距離	○	
人口変化率		○
地価変化率		○
工業用地までの距離	○	
役所までの距離	○	○
医療施設までの距離	○	○
教育施設までの距離	○	○
文化施設までの距離	○	
その他の公共施設までの距離	○	○
都心までの距離	○	○

2.4.4 予測モデルによる再現シミュレーション

モデル構築にあたって、表 2.6 に示した有意な相関があった項目を用いて、決定木分析を実施し、さらなる変数の絞り込みを行った。図 2.6、2.7 は、決定木分析によって算出された各変数の重要度を示したものである。

ナイーブベイズモデルの構築には、重要度の高かった上位 5 項目を変数として使用するが、2 地域とも、近傍の農地セル数と近傍の建物セル数の相関係数が 0.9 を超えたため、これら 2 変数のうち重要度の値が小さい変数を取り除いた。なお、ナイーブベイズモデルにおける説明変数の分布については、区域区分は、訓練データにおける各値の相対度数による分布の推定、それ以外の変数は、カーネル密度推定により分布を推定した。

構築したモデルを用いて 2009 年の土地利用を予測した再現シミュレーションの結果を表 2.7 に示す。正答率は、北大阪地域で 78.2%、東大阪地域で 73.1% と、2 地域とも 70%以上となった。

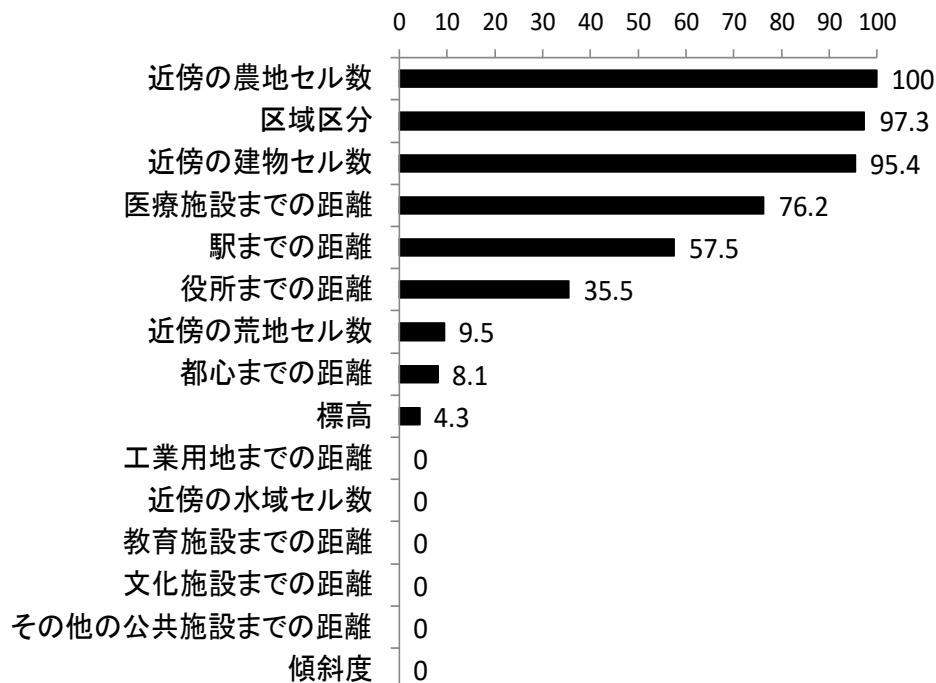


図 2.6 変数の重要度(北大阪地域)

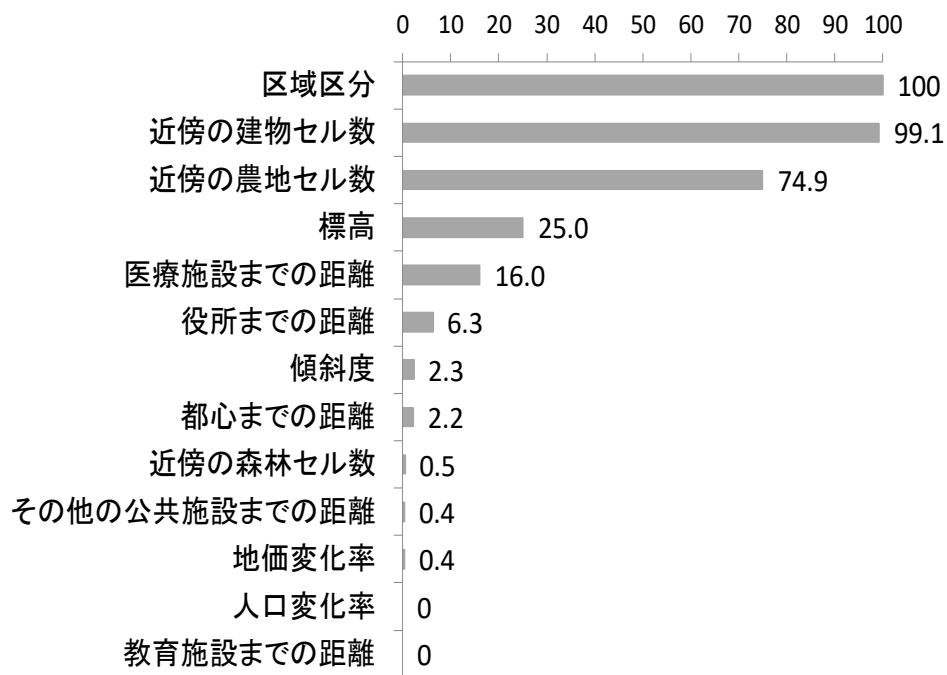


図 2.7 変数の重要度（東大阪地域）

表 2.7 再現シミュレーションの結果

	正答率 (正解セル数/対象セル数)	対象セル数	正解セル数
北大阪地域	78.2 %	509	423
東大阪地域	73.1 %	1452	1062

2.4.5 予測モデルによる将来予測シミュレーション

再現シミュレーションにて、一定の予測精度が得られることが確認できたため、同様の手法を用いて、将来の農地の変化について予測した。

シナリオごとのシミュレーション結果を表 2.8 に示す。また、図 2.8、2.9 に 2009 年時点と、シナリオ 1 及び 4 において予測された農地の分布を示した。両地域とも、区域区分を変更しないとしたシナリオ 1 では転用率が高く、将来的に多くの農地が転用されるという予測結果となった。一方、集団的な農地を優先的に規制したシナリオ 2~4 では、シナリオ 1 に比べて転用率が低くなっている。

大阪府の目標の減少率と比較すると、各地域の転用率は、近傍の農地セル数が 1 以上の農地をすべて農用地区域に編入させるという最も規制の強いシナリオ 4 の場合に、目標値に近い値になるが、それでも目標値に少し及んでいない。

表 2.8 将来予測シミュレーションの結果

(シナリオごとの農地の転用率を示す、()は(転用されたセル数/対象セル数))

	シナリオ 1	シナリオ 2	シナリオ 3	シナリオ 4
北大阪地域	25.7 % (100/388)	17.8 % (69/388)	16.2 % (63/388)	11.1 % (43/388)
東大阪地域	29.7 % (282/949)	18.4 % (175/949)	14.9 % (141/949)	8.7 % (83/949)



図 2.8 北大阪地域の 2009 年の農地(上)とシナリオ 1(中)、シナリオ 4(下)の予測結果
(緑：維持されている農地、肌色：転用された農地)

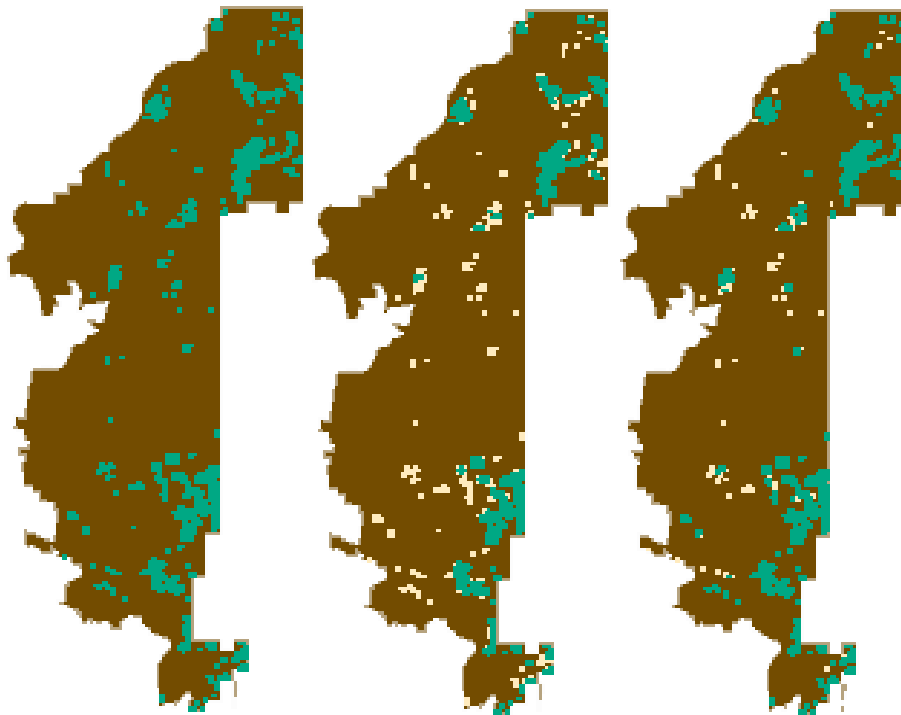


図 2.9 東大阪地域の 2009 年の農地(左)とシナリオ 1(中)、シナリオ 4(右)の予測結果
(緑：維持されている農地、肌色：転用された農地)

2.5 考察

大阪府を対象地域として、本研究の目的である社会的要因を考慮した農地の変化の分析、また高精度な予測モデルの構築および将来の農地の変化に関するシミュレーションを実施した。

農地の変化の要因の分析では、いくつかの社会的要因において、農地の変化と有意な相関がみられた。2 地域でともに有意な相関がみられる要因もあり、例えば、区域区分では同様の傾向がみられた。

地理的要因においても、近傍の農地セル数、近傍の建物セル数では、2 地域で同様の傾向がみられ、これらは地域に関係なく影響を与えていると考えられる。近傍の農地セル数が多いということは、農地がまとまって存在しているということであり、農業の生産性を高く保つことができるため、転用が少ないことが考えられる。一方で、近傍の建物セル数が多い場合は、周囲の市街化が進んでいる

ということであり、当該農地の開発需要も大きくなるため、転用が多くなることが考えられる。この点については、先行研究でも、農家が転用の行動をとる際に、基盤整備水準・宅地需要動向が大きな要因となっていることや(清水 1997)、農地周辺の都市的土地利用が影響を与えている(崔 2001)といったことが指摘されている。

また、社会的要因のなかでも農業関係の項目は、検定の結果、有意にならなかった。この点に関しては、入手できる詳細なデータが限られており、農業集落単位のデータを用いたことが原因のひとつであるかもしれない。

さらに、モデル構築にあたっては、決定木分析によって、複数の変数の中でも特に重要である変数について特定し、モデルの変数選択に活用した。分析の結果として、2地域いずれにおいても上位3項目は、区域区分、近傍の農地セル数、近傍の建物セル数で占められており、これらは、農地の変化に関連する要因のなかでも非常に大きな影響を与えているものであるといえる。また、医療施設までの距離も両地域で上位5項目の中に入っており、比較的重要な変数であると考えられる。また、北大阪地域では駅までの距離も上位5項目に入っている。医療施設や駅までの距離が重要な変数となっているのは、これらが農地法に基づく農地転用許可制度における転用基準の要件の一部になっているからだと考えられる(総務省 2018d)。

予測モデルについては、上述した決定木分析で変数選択を実施したうえで、ナイーブベイズの学習アルゴリズムを用いてモデルを構築するという新しい手法を提案した。実際の農地の変化を再現するシミュレーションを実施した場合の精度は、2地域でともに正答率が70%以上の値をとり、7割以上の農地の変化を正しく予測できていた。既存研究のモデルよりも高い精度であり、より実用的なモデルであると考えられる。精度を上げることができた要因としては、既存研究で見られるような、使用する変数をあらかじめ恣意的に決定してしまう方法ではなく、農地の変化と関連のある要因について分析し、モデル構築のための変数を絞り込むとともに、さらに決定木分析を行い、特に重要な変数のみをモデルに使用したことが大きいと考えられる。

また、モデルを用いて将来を予測した結果、区域区分を変更しないなどといった現在の状況が続く場合、今後も多くの農地が転用される可能性があることが示唆され、大阪府全体として掲げる目標値を達成することは難しいことが分かった。一方で、集団的な農地を優先的に規制するように区域区分の見直しを行っ

た場合、農地の転用率は低くなり、一定の効果があることが分かった。今後の農地の保全に向けて、集団的な農地を優先的に規制していくことは、ひとつの方策であると考えられる。しかし、最も規制を強くしたシナリオにおいても、大阪府の目標値には及ばないという結果となり、区域区分の見直しだけでは、限界があることも同時に示唆された。大阪府の掲げる目標値を達成するには、農業活動を促進していくような農業振興政策等の他の対策ともあわせて実行する必要があると考えられる。

本研究の課題として、上述したように、農家の属性等の農業関係の項目に関しては、より詳細なデータを分析していく必要があると考えられる。また、将来の予測シミュレーションに関しては、区域区分の見直しだけでなく、他の対策を実行した場合のシナリオも作成し、さらなる対策案について検討していく必要がある。さらに、今回提案した手法を大阪府以外の他地域にも適用し、モデルの有効性についても検証していきたいと考えている。

上記のような課題があるものの、本研究で示した一連の手法を用いることで、農地の保全のための今後の対策について検討を行うことができ、将来の農地の保全に役立てることができると考える。

第3章 農村部を中心とした自己保全管理農地の発生 に関する分析と予測シミュレーション

3.1 背景

近年、農地の転用や耕作放棄地化が進んでおり、耕地面積の減少が問題となっている。これからも、この傾向は続いていくとみられ、耕地面積の減少防止に向けた対策を早急に実施していくことが求められている。なお近年は、耕作放棄地面積の増加が転用面積の増加を上回るような形で急速に進行しており(農林水産省 2006)、耕作放棄地の発生が耕地面積減少の大きな要因となっている。

耕作放棄地の発生については、営農環境が厳しい中山間地域が多くを占める農村部において特に深刻な問題となっている。そのため、第2章で示した、都市郊外部における農地の転用についての分析と同時に、農村部における耕作放棄地の発生等の土地利用状況の変化の分析も必要であると考えられる。

また、実質的に耕作放棄地となっているものや、今後耕作放棄地となる可能性のあるものが含まれるものとして、自己保全管理農地がある。自己保全管理農地とは、何も栽培していない不作付地の農地で、耕作が可能な状態で管理されている農地のことである。耕作放棄地の増加を防ぐためには、その予備軍と考えられる自己保全管理農地について、発生を抑制していく必要がある。

高田(2007)が指摘するように、自己保全管理農地は、定義上は農林業センサスの耕作放棄地には含まれていないが、実態として耕作放棄地であるものが存在しており、農林業センサスでは耕作放棄地の実態が過小にとらえられている。

よって、耕作放棄地の実態を正確にとらえるためには、自己保全管理農地についても分析を行うことが不可欠であり、自己保全管理農地の発生の状況とその要因を的確に把握するとともに、発生を防止し、農地を保全していくための有効な対応策について明らかにしていくことが求められる。

また、耕作放棄地化するしないにかかわらず、不作付地である自己保全管理農地が多く存在するということは、農地を有効に活用できていないということであり、自己保全管理農地が増加すること自体も問題である。

既存研究では、高田(2007)が自己保全管理農地を対象に分析を行っているが、既存統計資料や現地調査等のデータをもとにした現状分析にとどまっており、

モデルの構築による要因分析や、自己保全管理農地の発生の予測にまでは至っていない。

また、自己保全管理農地を含んでいない耕作放棄地について、機械学習やシステムダイナミクスなどの定量的手法によって、要因分析や将来予測を行った研究もいくつか見られる(松井ら 2014, 吉田ら 2005, 稲葉 2006)。しかし、第1章でも述べたように、いずれの研究においても、市町村単位での耕作放棄地率や耕作放棄地面積をもとに要因分析を行い、将来予測も各市町村における耕作放棄地率や耕作放棄地面積の総量を予測しているため、対象地域である市町村内のどのような場所のどのような農地が耕作放棄地になる可能性が高いかといった空間的な面は考慮されていない。加えて、モデルの予測精度も決して高いとはいえない。

また、鄭ら(2004)や野村ら (2003)は、地区や集落を対象地域として、地域内における耕作放棄地の発現位置の特徴について分析しているが、実態を把握するレベルにとどまっており、統計的手法やモデルを用いた厳密な要因分析や予測にまでは至っていない。

各地域において、農地保全に向けたきめ細やかな対応策を実施するためには、具体的にどの農地がどういった要因で耕作放棄地や自己保全管理農地になっており、また今後、どの農地が耕作放棄地や自己保全管理農地になるかといった、農地単位のスケールでの分析や将来予測が必要であると考えられる。

3.2 目的

そこで、本研究では、市町村単位の耕作放棄地率や自己保全管理農地率のデータを用いるのではなく、各農地が自己保全管理農地であるか否かという、農地一筆単位のデータを用いて、自己保全管理農地の発生に関する詳細な分析を実施した。また、モデルにおいても、空間的な面を考慮し、農地一筆ごとに自己保全管理農地の発生を高精度に予測するようなモデルの構築を試みた。

また、農地保全の取組の推進は、全国共通の課題であり、それぞれ状況の異なる各地域において、要因分析や予測を行い、有効な対応策を検討するためには、一連の手順に基づいたモデルによる定量的な分析が有効である。今後、そのようなモデルの開発が必要になってくると考えられる。

そこで、本研究では、機械学習の手法を用いてモデルを構築し、今回分析対象

とした地域以外でも適用が可能であるような分析・予測に関する一連の手法を提案することとした。機械学習は、人工知能の分野を中心に研究が進められており、データから自動的に学習を行い、新たな知識を導出したり、また学習済みのアルゴリズムに新たなデータを入力して予測を行う手法である。

以上のことから、本研究の目的をまとめると、農地一筆単位のデータと機械学習の手法を利用して、自己保全管理農地の発生の現状とその要因について詳細に分析を行うとともに、自己保全管理農地の発生を予測する高精度なモデルの構築を試みることである。

3.3 材料と手法

3.3.1 対象地域

本研究の対象地域は、京都府綾部市とした(図 3.1)。綾部市は、古くから農業が盛んであり、多くの農地を有する。市の西部は、平地が広がる地域も存在する一方で、東部には、中山間地域も多く存在している。各地域で農家の高齢化も進んでおり、現状の農業・農地の状況の把握とともに将来の農業活動の維持に向けての対策が必要とされている。なお、図に示すように、綾部市には13の地区が存在する。

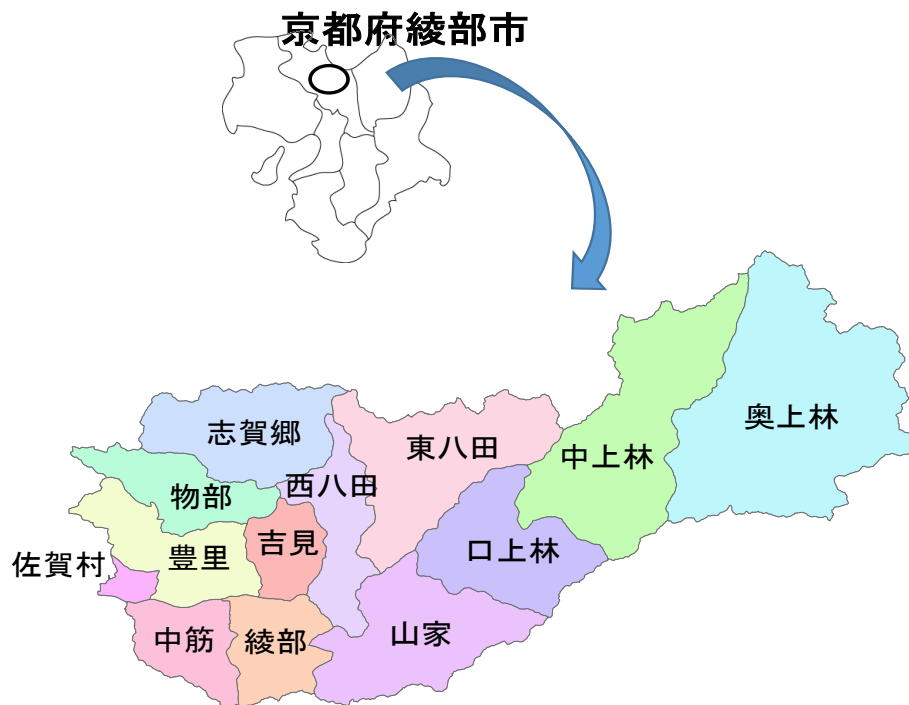


図 3.1 対象地域(京都府綾部市)

3.3.2 使用データ

本研究では、表 3.1 にあるように、国や自治体の統計資料を用いた。

表 3.1 使用するデータ

	データの内容(例)
水稻共済細目書 (GIS データ)	農地一筆ごとに農地の利用状況(栽培作物等)が記載されている
国土数値情報	標高、年降水量、地価、区域区分 etc.
基盤地図情報	道路、建築物
農林業センサス	農業従事者の平均年齢、販売金額 etc.

本研究で対象とする農地は、基本的に水田(転作後も含む)とし、農地のデータとして、2010年から2014年の水稲共済細目書のデータを用いた。水稲共済細目書には、各農家の所有農地とそこで栽培されている作物が記載されている。栽培されている作物のカテゴリは「水稲」、「転作」、「自己保全管理」に大きく分けられている。「水稲」はイネを栽培している場合、「転作」はイネ以外の作物を作っている場合があてはまる。そして、「自己保全管理」は何も栽培されていない不作付地の農地が該当する。

本研究においては、作物を栽培している「水稲」、「転作」を「耕作地」としてひとつにまとめ、この「耕作地」と「自己保全管理」の2種類の形に再分類して、分析を実施した。

実際の分析にあたっては、農地ごとに水稲共済細目書の情報が格納されているGISデータを活用し、綾部市全域における16,394筆の農地を対象とした。また、データの地理的・空間的な解析には、Esri社のArcGIS 10.2.2 for Desktop、統計解析やモデルの構築には、オープンソースソフトウェアのR 3.4.0を使用した。

3.3.3 自己保全管理農地の発生に関する要因分析

本研究では、2010年、2011年、2012年、2013年、2014年の農地の土地利用状況(耕作地であるか自己保全管理農地であるか)を目的変数として、モデルを構築し、そのモデルをもとに自己保全管理農地の発生に関する要因分析を行った。モデル構築に用いるサンプルサイズは、対象農地 $16,394 \times 5$ ヶ年分=81,970である。

目的変数(各年の農地の土地利用状況)に対応する説明変数は、前年の状況のデータを用いた。例えば、2010年の農地の土地利用状況のデータに対しては、説明変数のデータは2009年の値を使用した。今回は、2010～2014年の農地の土地利用状況を目的変数とするため、説明変数のデータは、2009～2013年の値を用いた。なお、説明変数について、国土数値情報、基盤地図情報(国土地理院 2017)、農林業センサスで1年ごとのデータが無いものについては、近い時期のデータを割り当てることで代用した。

要因分析では、表 3.2 にあるとおり、地理的・自然的条件(13項目)、社会的条件(7項目)、農家に関する属性(101項目)の計 121 項目を対象とした。

表 3.2 要因分析の項目

地理的・自然的条件(13 項目) :

面積、形状指数、近隣農地数、近隣農地の自己保全管理農地の数、農家が保有する他農地との地理的關係、標高、傾斜度、森林地域までの距離、年降水量、年平均気温、年最深積雪、年合計日照時間、年平均全天日射量

社会的条件(7 項目) :

農業地域の指定の有無、都市地域の指定の有無、DID(人口集中地区)までの距離、建築物までの距離、道路までの距離、地価、農地貸借の有無

農家に関する属性(101 項目) :

農地を保有している農家のタイプ、農林業センサスの項目 (計 100 項目・2010 年—販売農家のデータ)[注 1]

データの粒度として、基本的に農地一筆単位で一つの値を持つが、地理的・自然的条件の「年降水量」、「年平均気温」、「年最深積雪」、「年合計日照時間」、「年平均全天日射量」は 1km メッシュデータを用いているので、同一メッシュ内にある農地は、同じ値を割り当てている。また、農家に関する属性のうち、「農林業センサスのデータ(計 100 項目)」に関しては、地区単位(全 13 地区)のデータしか入手が出来なかったため、同じ地区にある農地にはすべて同じ値を割り当てている。

なお、地理的・自然的条件の項目である「近隣農地数」の定義としては、当該農地から周囲 5m 以内にある農地の数とした。また、「農家が保有する他農地との地理的關係」については、当該農地を保有する農家の他の保有農地が「近隣に存在」、「同一集落内に存在」、「同一集落内に全く存在しない」の 3 カテゴリーのいずれかを割り当てている。さらに、「形状指数」については、農地の面積を X 平方メートル、外周を l として、以下の式によって算出した値を用いた。

$$s = \frac{l}{4\sqrt{X}} \quad (3.1)$$

農地が正方形であれば、値は 1 となる。正方形でない長方形であれば 1 より大きい値をとることになる。

また、社会的条件の項目である「農地貸借の有無」については、水稻共済細目書上の農地の所有者(耕作者)が変更になった場合に、貸借有とすることとした。さらに、農家に関する属性の項目の「農地を保有している農家のタイプ」は、各農家のデータを用いてクラスター分析を実施し、農家を類型化した結果を用いた。

クラスター分析に用いた変数は、各年における農家の「保有農地数」、「1農地あたりの平均面積」、「1農地あたりの平均形状指数」、「保有農地における水稻農地の割合」、「保有農地における転作農地の割合」、「保有農地における自己保全管理農地の割合」の計6個の変数とした。

また、同じ農家でも年によって、保有農地数など上記の変数の値は変動する。そのため、同じ農家でも各年(2009年～2013年)によって値が異なる場合には、それぞれ別のデータとしてカウントした(用いた農家データのサンプルサイズは、 $n = 6950$)。

手法としては、階層的クラスター分析を用い、各サンプル間の距離はユークリッド距離、クラスター間の距離の測定方法は、Wald法を採用した。最適なクラスター数の決定には、Calinski-Harabasz index(Calinski and Harabasz 1974)を用いた。計算式は、クラスター数を k としたときの Calinski-Harabasz index を C_k とおくと、以下の式で表される。

$$C_k = \frac{(N - k) \cdot SS_B}{(k - 1) \cdot SS_W} \quad (3.2)$$

(N : サンプルサイズ, k : クラスター数, SS_B : クラスター間の分散,
 SS_W : クラスター内の分散)

SS_B 、 SS_W については、以下の式で表される。

$$SS_B = \sum_{i=1}^k n_i \cdot \|m_i - m\|^2 \quad (3.3)$$

(n_i : クラスター i のサンプルサイズ, m_i : クラスター i の重心,
 m : サンプルデータ全体の平均)

$$SS_w = \sum_{i=1}^k \sum_{x \in c_i} \|x - m_i\|^2 \quad (3.4)$$

(x : データ点, c_i : i 番目のクラスター, m_i : クラスター i の重心)

なお、Calinski-Harabasz index が最大となるクラスター数を最適なクラスター数とした。結果として、最適なクラスター数は 5 個となった。

表 3.3 にクラスターごとの該当農家数と各変数の値の平均を示す。各クラスターの特徴として、クラスター 1 は保有農地数が少なく小規模で、大半が転作農地である、2 は小規模で、大半が自己保全管理農地である、3 は小規模だが 1 農地あたりの面積が大きく、大半が水稻農地である。4 は小規模だが、クラスター 1～3 に比べて保有農地数がやや多く、水稻農地の割合が多い。5 は保有農地数が多い大規模な農家で、水稻農地の割合が多いといったことが見てとれる。

「農地を保有している農家のタイプ」の項目は、このクラスター 1～5 から該当するいずれかの値を割り当てている。ただし、農地貸借が行われ、前年まで水稻共済細目書に登録のなかった新たな農家に貸出された場合は、その年は「新規」というカテゴリを割り当てている。

表 3.3 クラスターごとの農家数と変数の値の平均

クラスター NO.	該当農家数	保有農地数 (平均)	1 農地あたりの面積 (a) (平均)	1 農地あたりの形状指数 (平均)	水稻農地の割合 (平均)	転作農地の割合 (平均)	自己保全管理農地の割合 (平均)
1	701	2.5	687	1.2	1 %	96 %	3 %
2	1170	3.3	756	1.2	3 %	7 %	90 %
3	484	3.8	2351	1.1	97 %	3 %	0.1 %
4	4042	5.7	1061	1.2	74 %	16 %	10 %
5	553	39.6	1103	1.1	80 %	15 %	4 %

3.3.4 ランダムフォレストモデル

本研究では、モデルの作成にランダムフォレストの手法を用いた。ランダムフォレストは、機械学習のアルゴリズムの1つであり、多数の決定木を構築し、各決定木の結果の多数決をとることで、分類・予測を行う手法である(Breiman 2001)。

ランダムフォレストのアルゴリズムは以下の通りである。

1. 構築する決定木の数を B 個とする。

For $b = 1$ to B :

- (a) N 個の訓練データからブートストラップ標本集合 Z_d を取り出す。
- (b) Z_d を学習データとして用いて、決定木 T_b を成長させる。そのために、木の終端頂点について以下の処理を再帰的に繰り返す。終端ノードのデータ数の下限は 1 とする。
 - i. p 個の変数から m 個をランダムに選択する
 - ii. m の中から最良の変数と分割点を選ぶ
 - iii. その頂点を二つの子頂点に分割する

2. 木のアンサンブル T_b^B を出力する。

3. 新しいデータ x に対して、以下の形で予測を行う。

b 番目の決定木によるクラス予測結果を $\hat{C}_b(x)$ とし、全ての決定木の予測

結果の集合 $\{\hat{C}_b(x)\}_1^B$ の中で、クラス j と予測した木の数を $|C_j|$ とすると、

ランダムフォレストの最終的な予測結果 C_{rf} は、 $C_{rf} = \underset{j}{\operatorname{argmax}} |C_j|$ となる。

ランダムフォレストの利点として、変数間の相互作用を考慮できることや、ブートストラップサンプリングにより一部のデータのみを使用することで過学習を防ぎ、高い汎化性能を実現できるということがある。さらに、大きな特徴として、ランダムフォレストには変数の重要度を算出できるといった点がある。

本研究では、自己保全管理農地の発生について、どのような要因が大きく影響しているかといった要因分析も目的としている。そのため、変数の重要度について算出することができ、それに基づき重要な要因の特定が可能なランダムフォレストを用いることとした。また、変数の重要度を算出できる手法としては、決

決定木やロジスティック回帰等も候補として考えられるが、ランダムフォレストはそれらの手法に比べて、予測の面でも高い汎化性能を持つため、本研究のもう一つの目的である高精度な予測モデルの構築という点をふまえ、今回はランダムフォレストを採用することとした。

本研究では、要因分析のために、各農地が耕作地となるか、自己保全管理農地となるかを予測するランダムフォレストモデルを作成し、各説明変数の重要度の算出を行った。重要度の算出の指標としては、Mean Decrease Accuracy(MDA)を用いた(Breiman 2001)。MDA は当該変数の値をサンプルデータ間でランダムに入れ替えた時に、本来よりもどれくらい予測精度が下がるかを算出し、重要度の指標としたものである。つまり、値をランダムに入れ替えた時に、予測精度が大きく下がる変数ほど、重要な変数であるとみなされることになる。

変数の重要度は、「(1)全てのデータを用いたモデル」に加えて、より細かい分析のために、前年の農地の土地利用状況が耕作地の場合と、自己保全管理農地の場合にデータを分け、「(2)前年の農地の土地利用状況が耕作地であったデータを用いたモデル」、「(3)前年の農地の土地利用状況が自己保全管理農地であったデータを用いたモデル」についても作成し、計 3 種類のモデルのそれぞれにおいて、重要度の算出を行った。

なお、ランダムフォレストでは乱数を使用するため、毎回、結果が同じになるとは限らない。そのため、複数回の試行を行い、結果を評価することが望ましい。また、決定木の数を十分に大きくすることで、なるべく乱数に依存しない結果を得ることができる(金 2009)。

そこで、各変数の重要度の値については、20 回分のモデル構築を繰り返し、それぞれのモデルで算出された値の平均をとった。

また、いずれのモデルにおいても、生成する決定木の数を 500(今回用いるデータでは、左記の数でモデルの推測誤差が十分に安定することを事前に確認している)、決定木の分岐に使用する変数の候補として、ランダムに選択される変数の数は、 \sqrt{p} (p : モデルに使用する説明変数の総数)とした。なお、モデルの構築には、R の randomForest パッケージを使用した。

3.3.5 モデルの精度の検証

また、上述したランダムフォレストモデルの予測精度を検証するために、交差検定(Cross Validation)を行い、平均正答率を算出した。交差検定の分割数は、 n をサンプルサイズとした場合の以下のスタージェスの公式 (Sturges 1926)をもとに決定した。本研究での分割数は 17 とし、17 分割交差検定を実施した。

$$k = 1 + \log n / \log 2 \quad (3.5)$$

なお、正答率については以下の式で算出した。

$$\text{正答率(\%)} = \frac{\text{予測した土地利用状況が現実の土地利用状況と一致した農地数}}{\text{対象農地数}} \times 100$$

また、本研究では、モデルに用いる説明変数の数を、2~121 個(要因分析で対象とした全変数)まで変更しながら、予測精度の検証を行った。

モデルに用いる変数については、3.3.4 で述べた「(1)全てのデータを用いたモデル」における各変数の重要度の算出結果をもとに、重要度の高い変数から順にモデルに加えていくこととした。例えば、説明変数の数が 2 個のモデルでは、重要度の高かった上位 2 項目をモデルの説明変数として採用した。同様に、説明変数の数が n 個のモデルの場合、重要度の高かった上位 n 項目を説明変数として採用した。

そして、説明変数の数が異なるそれぞれのモデルの中から、最終的なモデルを決定するにあたっては、1-SE Rule(新妻・斎藤 2009)をもとに、最も平均正答率の高かったモデルの「平均正答率－標準誤差(SE)」以上の平均正答率を有するモデルの中で、最も説明変数の数が少ないモデルを最良のモデルとすることとした。

さらに、市内であっても地域によって営農環境が異なると考えられる。そこで、農林水産省が定める農業地域類型区分(都市的地域、平地農業地域、中間農業地域、山間農業地域)(農林水産省 2017b)ごとに、各農業地域に属する農地についての平均正答率を算出し、精度の比較を行った(なお、農業地域類型区分では、地区単位でいずれかの地域類型に区分される形になっており、綾部市では 13 地区のうち、都市的地域に 1 地区、平地農業地域に 1 地区、中間農業地域に 8 地区、山間農業地域に 3 地区が指定されている)。

3.4 結果

3.4.1 自己保全管理農地の発生状況

まず、対象地域における自己保全管理農地の発生の現状を示す。図 3.2 は、2010 年～2014 年における自己保全管理農地数と、全農地に対する自己保全管理農地率の推移をみたものである。グラフを見ると、年々少しずつであるが、自己保全管理農地が増加していることが分かる。

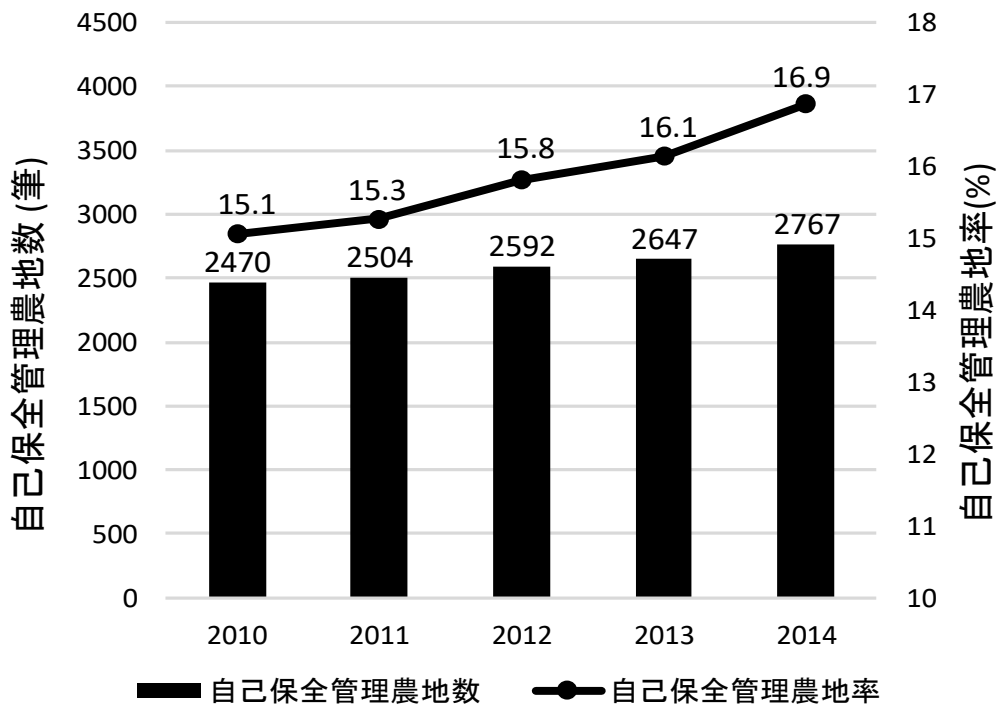


図 3.2 自己保全管理農地数の推移

3.4.2 自己保全管理農地の発生に関する重要な要因の算出

次に、自己保全管理農地の発生に関する要因分析の結果について述べる。前節で述べたように農地の土地利用状況を予測するランダムフォレストモデルを作成し、変数の重要度を算出した。

まず、「全てのデータを用いたモデル」によって算出された重要度の高い変数の上位 15 項目を図 3.3 に示す。

図を見ると、「農地を保有している農家のタイプ」が最も重要度が高く、保有

する農家の違いが自己保全管理農地の発生に大きく影響を与えているということが推察される。

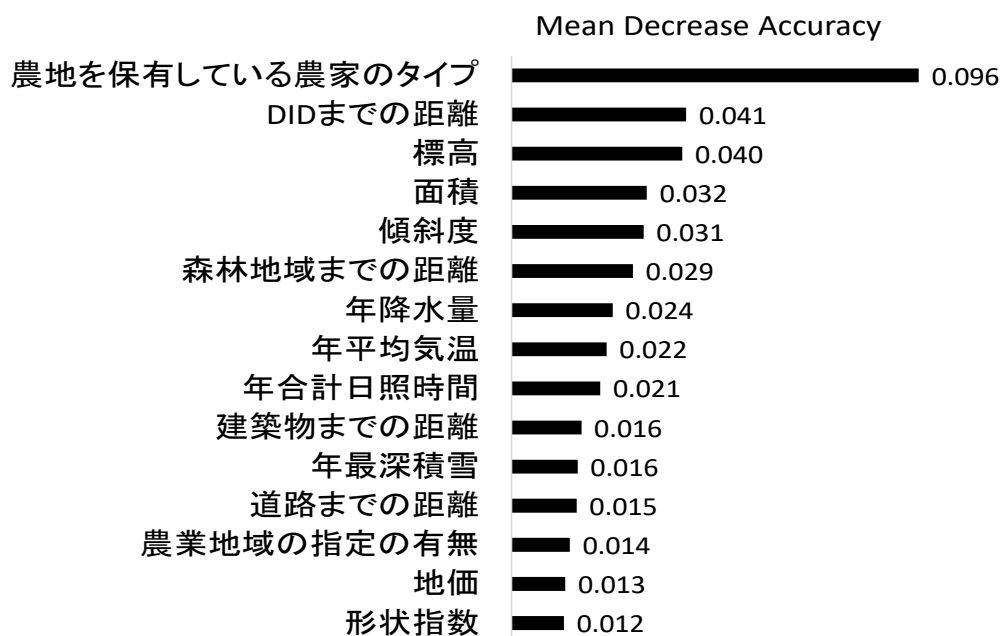


図 3.3 重要度の高い上位 15 項目

また、前年の農地の土地利用状況別にデータを分けてモデルを作成し、重要度の算出を行った場合の結果を示す。図 3.4 は「前年が耕作地であった場合」、図 3.5 は、「前年が自己保全管理農地であった場合」のモデルにおいて、重要度の高い変数の上位 15 項目について示したものである。

前年が耕作地である場合は、「DID までの距離」が最も重要な変数となっている。また、それ以降は、「標高」や「年降水量」等の地理的・自然的条件が比較的上位に来ていることが見てとれる。なお、上位 15 項目は、図 3.3 の全てのデータを用いたモデルにより算出された項目と、順位の違いはあるが、全く同じ項目が並ぶ結果となった。

一方で、前年が自己保全管理農地である場合は、「農地貸借の有無」が最も重要な変数となっている。これは、全てのデータを用いたモデルと、前年が耕作地であったデータを用いたモデルの上位 15 項目では出てこなかった項目である。2 位以下については、これまでの 2 つのモデルでも登場した項目が並んでいる。

自己保全管理農地が、そのまま自己保全管理農地であるか、あるいは耕作地に変化するかについては、農地の貸借の実施の有無が大きく影響していることが推察される。

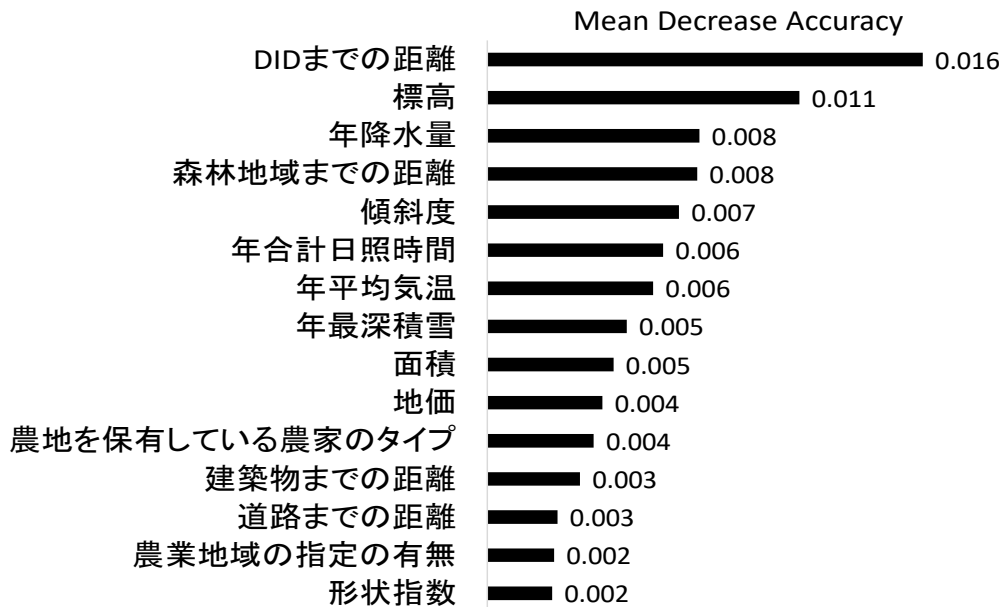


図 3.4 重要度の高い上位 15 項目
(前年の土地利用状況が耕作地の場合)

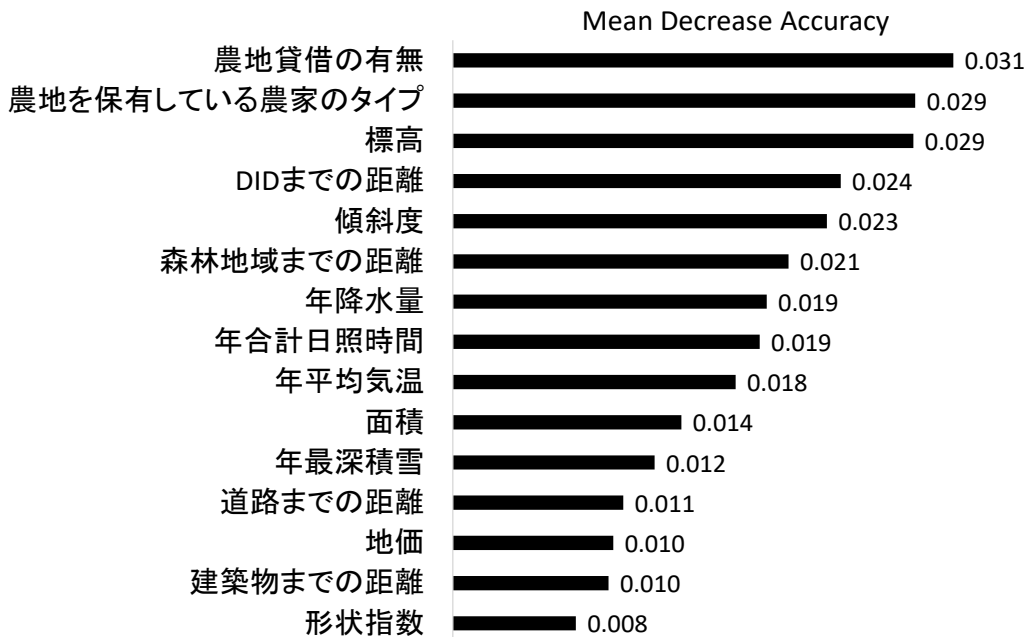


図 3.5 重要度の高い上位 15 項目
(前年の土地利用状況が自己保全管理農地の場合)

さらに、3つのモデルの算出結果(図 3.3~3.5)のそれぞれにおいて最も重要度の高い変数であった、「農地を保有している農家のタイプ」、「DID までの距離」、「農地貸借の有無」について、農地の土地利用状況との関係をより詳細に分析した。なお、以降に示す値は、2010~2014年の全期間の土地利用状況のデータ(説明変数は、2009~2013年)を合計して算出したものである。

図 3.6 は、前年に農地を保有していた農家のタイプ毎に翌年の農地の土地利用状況がどうなったかを示したものである。保有農地が耕作地になった割合と、自己保全管理農地になった割合をそれぞれグラフで示している。ただし、各年で農地の貸借があり、所有者が変わった場合、その農地は貸借後に保有する農家の農家タイプに紐づけてカウントをしている。また、前年まで水稲共済細目書に登録のなかった新たな農家に貸出された農地については、農家のタイプの新規にまとめてカウントしている。

グラフを見ると、クラスター4、クラスター5 が保有している農地は、耕作地になる割合が非常に高いことが分かる。一方で、クラスター2 が保有している農地は、大半が自己保全管理農地になっていることが分かる。

また、帰無仮説 H_0 : 農家のタイプの違いによって、耕作地と自己保全管理農地になる農地数の割合に差がない、対立仮説 H_1 : 農家のタイプの違いによって、耕作地と自己保全管理農地になる農地数の割合に差がある、としてカイ二乗検定を用いて検定を行った。結果として、有意水準 5% で有意となった。

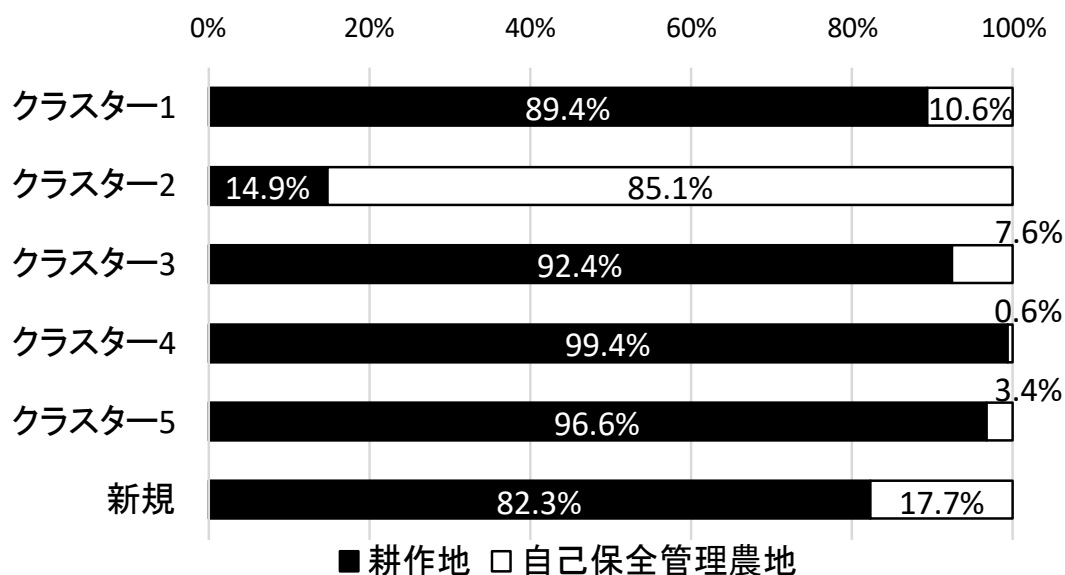


図 3.6 農家のタイプ別の保有農地の土地利用状況

また、前年が耕作地であった農地について、DID までの距離と翌年の農地の土地利用状況の関係を分析した。

耕作地になった農地群と自己保全管理農地になった農地群のそれぞれで、DID までの距離の中央値を算出すると、耕作地になった農地群では 5719m(四分位偏差：2680)、自己保全管理農地になった農地群では 6588m(四分位偏差：2114)であった。自己保全管理農地になる農地は、耕作地になる農地に比べて、DID までの距離が長い農地が多い傾向にあることが分かる。

また、帰無仮説 H_0 ：耕作地になった農地群と自己保全管理農地になった農地群で DID までの距離に差がない、対立仮説 H_1 ：耕作地になった農地群と自己保全管理農地になった農地群で DID までの距離に差がある、として Brunner-Munzel 検定を用いて検定を実施した。結果として、有意水準 5%で有意となった。

さらに、図 3.7 は、前年が自己保全管理農地であった農地について、農地貸借の有無別に、翌年の農地の土地利用状況をグラフで示したものである。

グラフを見ると、貸借が行われた場合、7割程度の農地が、自己保全管理農地から耕作地に変わっていることが分かる。一方で、貸借が行われなかった場合、大半の農地が自己保全管理農地のままであることが分かる。

また、帰無仮説 H_0 ：農地貸借の有無で、耕作地と自己保全管理農地になる農地数の割合に差がない、対立仮説 H_1 ：農地貸借の有無で、耕作地と自己保全管理農地になる農地数の割合に差がある、としてカイ二乗検定を用いて検定を実施した。結果として、有意水準 5%で有意となった。

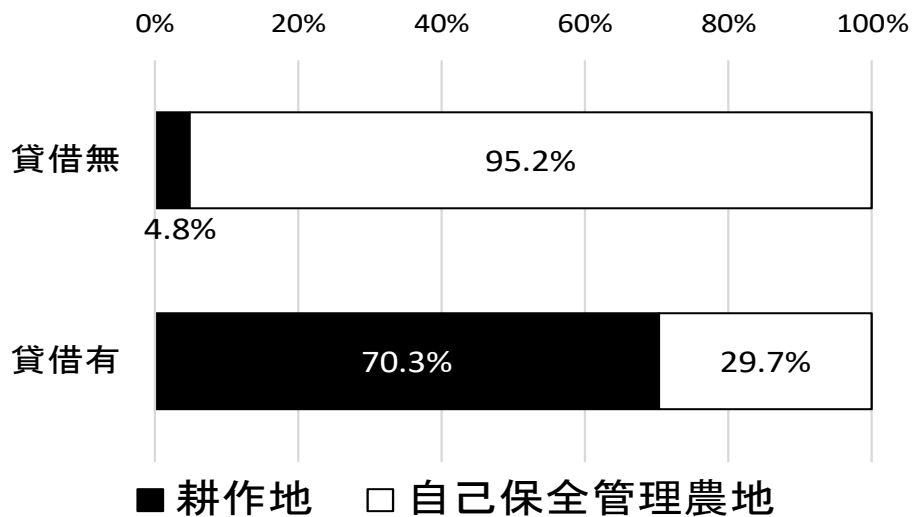


図 3.7 貸借の有無別の農地の土地利用状況
(前年が自己保全管理農地であった農地のみ)

3.4.3 ランダムフォレストモデルの精度の検証

作成したランダムフォレストモデルについて予測精度を検証するため、交差検定により平均正答率を算出した。

本研究では、モデルに用いる説明変数の数を段階的に変更しながら、精度の検証を行った。図 3.8 は、説明変数の数が 2~15 個(用いる変数は図 3.3 の重要度の高い上位の項目から順に採用)と 121 個(全変数)の場合の平均正答率を示したものである。

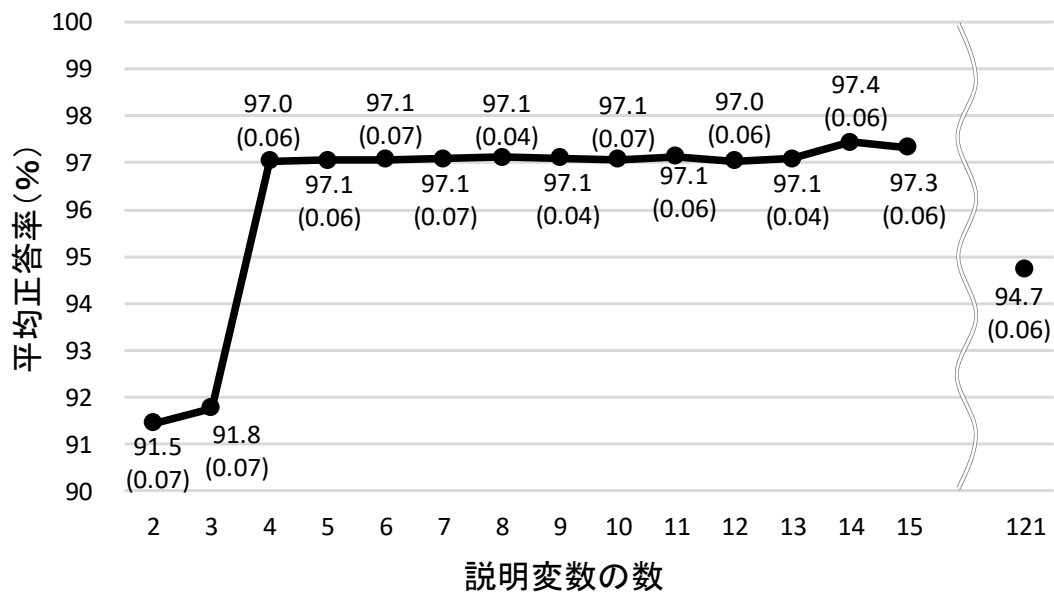


図 3.8 交差検定によるモデルの平均正答率
(各プロットの () の値は平均正答率の標準誤差(SE))

結果として、1-SE Rule に基づく最良のモデルは、説明変数の数が 14 個のモデルとなり、平均正答率は 97.4% と非常に高い精度となった(図 3.8 では、説明変数が 16 個以上の場合の結果は記載していないが、説明変数が 14 個のモデルの「平均正答率－標準誤差(SE)」を上回る平均正答率のモデルは存在しなかった)。また、説明変数の数が 14 個のモデルは、121 個(全変数)のモデルよりも 3% 程度高い正答率となっている。

一方、説明変数の数が 3 個以下のモデルでは平均正答率が低くなっている。これは、ランダムフォレストにおいて、3.3.4 で述べたように、決定木の分岐に使用するためにランダムに選択される変数の候補の数を、 \sqrt{p} (p : モデルに使用する説明変数の総数)としたため、説明変数の数が 3 個以下の場合、選択される変数の数が 1 個となってしまい、決定木の分岐の際に、その一意に選択された変数を使用するしかないためだと考えられる。

また、説明変数の数が 4 個のモデルでは、平均正答率は 97.0% であり、最良とされる 14 個のモデルとそれほど変わらない精度となっている。よって、これら

4個の変数(農地を保有している農家のタイプ、DIDまでの距離、標高、面積)だけでも、ある程度の高い精度で予測が可能なが分かる。

また、表 3.4 は、最良のモデルである説明変数が 14 個のモデルにおける交差検定の結果について、農業地域類型区分別での平均正答率を示したものである。

表を見ると、平均正答率は、地域ごとに多少の違いはあるが、いずれの地域においても高い値となっており、綾部市内の様々な農業地域の農地について、うまく予測できていることが分かる。

表 3.4 農業地域類型区分別の平均正答率
(説明変数の数が 14 個のモデル)

	平均正答率 (±標準誤差) (%)
都市的地域	96.3 (±0.3)
平地農業地域	98.5 (±0.1)
中間農業地域	97.6 (±0.1)
山間農業地域	96.8 (±0.1)

3.5 考察

京都府綾部市を対象地域として、本研究の目的である自己保全管理農地の発生の現状とその要因について分析を行い、さらに自己保全管理農地の発生を予測する高精度なモデルについて構築を試みた。

まず現状として、年々、自己保全管理農地が増加していることが分かった。農林業センサス等から耕作放棄地が増加していることは明らかになっているが、京都府綾部市では、自己保全管理農地も増加していることが分かり、耕作放棄地と同様に発生防止や解消に向けた対策を実施していく必要があることが分かった。

また、自己保全管理農地の発生に関連している要因について分析を行うため、ランダムフォレストモデルを構築し、重要度の高い変数を算出した。結果として、

農地を保有している農家のタイプが重要な要因となっていることが分かった。農家のタイプによって、保有している農地の土地利用状況が大きく異なり、例えば、所有農地が自己保全管理農地になる割合が高いクラスター2の農家は、営農に関して意欲のない農家や、今後リタイアを考えている農家であると推測される。一方で、クラスター4やクラスター5の農家は、営農意欲の高い農家であると考えられ、特にクラスター5の農家は、保有農地数も多い比較的大規模な農家であり、今後、地域の営農を中心的に担っていく可能性を持つ農家である。自己保全管理農地の数を減少させていくためには、クラスター2が保有する農地などを、クラスター5のような農家に集めていくことが重要であると考えられる。

さらに、DIDまでの距離や標高、傾斜度といった変数も重要度が高く、農地の存在する場所の条件も大きく影響していることが分かった。このことから、空間的な面を考慮に入れた農地一筆単位での詳細な分析や予測モデルの構築が重要であることが分かる。

また、前年が耕作地である場合のデータを用いて、重要度の算出を行った場合、DIDまでの距離が最も重要な変数であった。耕作地から自己保全管理農地に変化する農地は、DIDまでの距離が長い傾向にあり、これはDIDから遠いところに位置する中山間地域などで、より自己保全管理農地が発生しやすいことを示唆しているといえる。

一方、前年が自己保全管理農地である場合のデータを用いて、重要度の算出を行った場合、農地貸借の有無が最も重要な変数であった。貸借を行った場合、多くの農地が自己保全管理農地から耕作地に変わっており、自己保全管理農地の解消にあたっては、農地の貸借をより積極的に進めていくことが必要であることが分かった。耕作放棄地を対象とした松井ら(2014)や稲葉(2006)の研究でも農地の流動化や集約化を進展させることが重要であることは指摘されており、自己保全管理農地でも同様の結果が得られた。

また、自己保全管理農地の発生を予測するランダムフォレストモデルについて、その有効性を確認するために、交差検定を通じてモデルの精度を検証した。本研究では、モデルに用いる説明変数の数を段階的に変更しながら精度の検証を行った。また、変数選択については、ランダムフォレストで算出された変数の重要度の値をもとに実施し、重要度の値が高い上位の変数から順にモデルの説明変数として採用した。結果として、説明変数が14個のモデルが最良のモデルとなり、平均正答率は97.4%と非常に高い精度が得られた。

また、要因分析で対象とした全ての変数(121 変数)を用いて構築したモデルと比べて、変数選択を行った場合のモデルは、精度がやや高くなっており、モデルの構築にあたって、変数の重要度の値をもとに変数選択を行う手法が有効であることも分かった。

さらに、説明変数が 4 個のモデルの場合でも、最良のモデルとそれほど変わらない精度となっており、より少ない変数でも、ある程度の高い精度で自己保全管理農地の発生をうまく予測できることが分かった。

また、それぞれ状況が異なると考えられる農業地域類型区分別に精度を見た場合、いずれの地域においても、高い正答率が得られており、綾部市の様々な農業地域において有効なモデルであることが分かった。

本研究では、これまで十分に分析がなされていなかった自己保全管理農地を対象に、発生に関連する要因分析や、予測モデルの構築を行った。

また、市町村単位の耕作放棄地面積や自己保全管理農地面積を用いた研究が多い既存研究とは異なり、対象地域内における空間的な面を考慮し、農地一筆単位のデータによる要因分析と、農地一筆ごとに自己保全管理農地の発生を予測するモデルを構築した。

結果として、予測モデルの精度は非常に高い値となり、本研究で提案した一連の分析やモデル構築手法は、より実用的なものとして、自己保全管理農地の発生予防や解消に役立てることが出来ると考えられる。

また、本研究で提案した手法は、データさえ入手できれば、他の地域においても同様に分析が可能であり、今後、汎用性のある手法として利用できる可能性を有している。

今後の課題としては、構築した予測モデルを活用して、実際に将来の農地の土地利用状況の予測を行うことや、上述した手法の汎用性について検証するために、他地域においても同様の手法で分析を行い、その有効性を確認していく必要がある。

また、要因分析の際に用いた農林業センサスのデータについては、データの入手の関係上、地区別のデータという粒度の粗いデータを用いた。要因分析の結果として、農地を保有している農家のタイプが重要な要因であったということもあり、今後、農林業センサスの個票データ等の農家の属性に関するより詳細なデータを用いて分析を実施することも検討していきたいと考えている。

注 1)

■農業生産関連事業を行っている農家の割合、■水稲作作業を委託した農家の割合、■雇用者を雇い入れた農家の割合、■水稲作受託農家の割合、■農作物販売金額規模別の農家の割合(3項目) …販売なし・50万円未満・50万円以上、■主業農家の割合、■主業農家で65歳未満の農業専従者がいる農家の割合、■準主業農家の割合、■準主業農家で65歳未満の農業専従者がいる農家の割合、■副業農家の割合、■属性別の平均年齢(6項目) …経営者・農業従事者・農業就業人口・基幹的農業従事者・自営農業従事日数が150日以上の人(農業専従者)・同居後継者、■1農家あたりの機械保有台数(3項目) …動力田植機・トラクター・コンバイン、■農業従事者の年齢別の割合(15項目) …15～19歳・20～24歳・25～29歳・30～34歳・35～39歳・40～44歳・45～49歳・50～54歳・55～59歳・60～64歳・65～69歳・70～74歳・75～79歳・80～84歳・85歳以上、■農業就業人口の年齢別の割合(15項目) …※項目は、上述の農業従事者の年齢別の割合と同様、■基幹的農業従事者の年齢別の割合(15項目) …※項目は、上述の農業従事者の年齢別の割合と同様、■同居後継者がいる農家の割合、■同居後継者はいないが、他出農業後継者がいる農家の割合、■同居後継者も他出農業後継者もない農家の割合、■同居後継者の年齢別の割合(8項目) …15～19歳・20～24歳・25～29歳・30～34歳・35～39歳・40～44歳・45～49歳・50歳以上、■農業従事者の自営農業従事日数別の割合(7項目) …29日以下・30～59日・60～99日・100～149日・150～199日・200～249日・250日以上、■同居後継者の自営農業従事日数別の割合(8項目) …従事しなかった・29日以下・30～59日・60～99日・100～149日・150～199日・200～249日・250日以上、■一世代家族経営農家の割合、■一世代家族経営農家のうち、経営者が65歳以上である農家の割合、■二世代家族経営農家の割合、■二世代家族経営農家のうち経営者が65歳以上である農家の割合、■三世代等家族経営農家の割合、■三世代等家族経営農家のうち経営者が65歳以上である農家の割合、■実行組合のある集落の割合、■1集落あたり開催した寄合の回数

第4章 農家の離農予測シミュレーション

4.1 背景

近年、農家数は年々減少するとともに、農地の転用や耕作放棄地化が進んでいる。そのため、農地の保全に向けた対策を早急に実施していく必要がある。

第2章、第3章では、農地の土地利用状況の変化に関する分析を行った。しかし、各章の考察でも述べたように、農家の属性に関しては、粒度の粗いデータを用いており、詳細な分析ができていないという課題があった。

現実的には、耕地面積の減少や、転用ならびに耕作放棄地の増加といった農地の土地利用状況の変化は、当該農地の保有者である農家の営農状況の変化を通じて生じているといえる。そのため、個々の農家の営農状況の変化を分析することは非常に重要であり、前章までの農地の土地利用状況の変化の分析とも密接に関連するものである。

特に、適切な農地保全対策の実施のためには、過去にどのような農家がどの程度離農しているかを把握するとともに、将来における農家の離農の動向について予測を行うことが重要になってくる。栗原ら(2014)が述べるように、どのような農家が離農しているかを把握することは、農地集積のために離農を促進する政策にしる、集落機能維持のために離農を防ぐ政策にしる地域が含有する特徴や課題に沿った政策を策定するために不可欠である。また、将来的に離農する可能性が高い農家を予測することで、今後の耕作放棄地の発生リスクと、それに対する適切な対応策の検討が可能となる。そこで、農家の離農に焦点をあて、定量的な離農予測モデルの構築を試みた。

農家の離農に関する既存研究については、海外において、統計データ等を用いて定量的に分析を実施している研究(Katchova and Ahearn 2017, Landi *et al.* 2016, Mishra *et al.* 2014, Möllers and Fritsch 2010)がいくつか見られる。これらの研究による知見は重要であるが、日本と海外では、農業を取り巻く状況が大きく異なるため、海外で得られた知見を直接的に、国内に適用するのは難しいという問題がある。

一方、我が国を対象とした研究については、実態調査によって接近した研究は散見されるが(山本 2003, 井上 1989-1990)、定量的に離農要因を分析した研究は少ないことが指摘されている(栗原ら 2014)。

そのような中で、定量的アプローチによる研究として、主に農林業センサス個票を用いて、離農要因の分析(栗原ら 2014, Kurihara *et al.* 2013)や離農予測(将来の農家数の予測)を行った研究(安武 2018, 安武・千田 2013, 渡部ら 2012, 細山 2013)も見られる。しかし、第 1 章でも述べたように、分析のスケールが全国、地方、都道府県や都道府県内の地域エリアといった広範なレベルであり、よりスケールの小さい市町村以下の単位での詳細な分析や予測はほとんど行われていない。また使用する変数も、最初から決め打ちをしているか、分析者が試行錯誤的に選択している研究が多く、統計的手法を用いて厳密な変数選択を行っている研究は非常に少ない。

さらに、上記に示した離農予測の既存研究では、予測モデルとしてマルコフモデルを用いているが、マルコフモデルでは、各変数を層別して解析する必要があり、変数の数を増やしていくと、層別後の変数間の組み合わせが指数関数的に増大し、モデルが非常に複雑になるという問題がある。そのため、候補となる変数が多数あったとしても、おのずと使用可能な変数は少数に限られてしまう。実際のところ、農家の離農には様々な要因が複雑に影響していると考えられるため、多数の変数を導入できるようなモデルが必要である。

また、市町村レベルで離農予測を行っている研究(門間 2016)もあるが、経営主の年齢が一定段階(75 歳)に達したら離農するといった非常に単純なアルゴリズムによる予測にとどまる(実際は、75 歳になっても離農しない農家が一定数存在する)。そのため、より現実の状況を反映した詳細な離農予測モデルが必要である。

さらに既存研究では、構築したモデルについて、予測精度の検証を行っている研究も少ない。そのため、モデルの妥当性や、各市町村でどのような変数が離農の予測に有効であるかという点も明確に把握できていない。

農家の離農に関して、大まかな傾向を把握するには、既存研究のような分析や予測でよいと考えられる。しかし、実際に市町村や地区単位でのきめ細やかな対策を実施しようと考えれば、よりミクロなスケールでの分析や、予測精度の高い精緻なモデルの構築が必要になってくる。

また、農家の離農の問題は、全国共通の課題であり、それぞれ状況の異なる各地域で要因分析や予測を行い、有効な対応策を検討できるように、一連の手順に基づく定量的・統計的な手法の開発が必要である。今後は、そのような形で、地域毎にモデルを構築し、個別に予測が実施できるモデルが求められる。

加えて、門間(2016)が述べるように、我が国の農業は多様な地域特性の上に歴史的に形成されており、本来は各地域における個々の農家の特性を反映した構造予測手法が望ましい。

4.2 目的

以上をふまえ、本研究では、市町村単位での離農予測モデルの構築を行うこととした。また、多数の変数を分析対象として用いるとともに、統計的な手続きに基づいた変数選択や予測を実施し、各地域で同様の分析や予測が可能であるような高精度かつ汎用性のあるシミュレーションモデルを提案することを目的とした。また、モデルの構築を通じて、農家の離農と関連が高く、離農予測の際に重要である変数についても分析を行った。

さらに手法として、統計的機械学習(ロジスティック回帰)とマイクロシミュレーションを組み合わせて、農家の多様な特性をふまえながら、個々の農家ごとに離農予測を行うような新たな手法の提案を試みた。

4.3 材料と手法

4.3.1 対象地域

本研究の対象地域は、茨城県とした。また詳細な分析を実施するための市町村として、神栖市、水戸市、筑西市、稲敷市、大子町の5市町村を選定した(図4.1)。上記の5市町村の選定にあたっては、農林水産省が定める農業地域類型区分を参考にした。農業地域類型区分とは、DID(人口集中地区)面積や耕地率などに基づいて農業地域を区分分けしたものであり、第1次分類として、「都市的地域」、「平地農業地域」、「中間農業地域」、「山間農業地域」の4つに分けられる。各農業地域によって営農環境は異なり、それに伴って農家の離農の傾向も異なる可能性があるため、4つの地域それぞれに属する市町村に関して、バランスを考慮し選定を行った。

なお、選定した5市町村のうち、神栖市は大半が都市的地域であり、水戸市は都市的地域と平地農業地域がそれぞれ半数程度であり、筑西市と稲敷市は大半が平地農業地域であり、大子町は中間農業地域と山間農業地域がそれぞれ半数程度を占めている。また、県内における位置関係(地域エリア)も考慮に入れ、県

北地域として大子町、県央地域として水戸市、県西地域として筑西市、県南地域として稲敷市、鹿行地域として神栖市と、各地域エリアから 1 か所ずつ選定するよう配慮した。

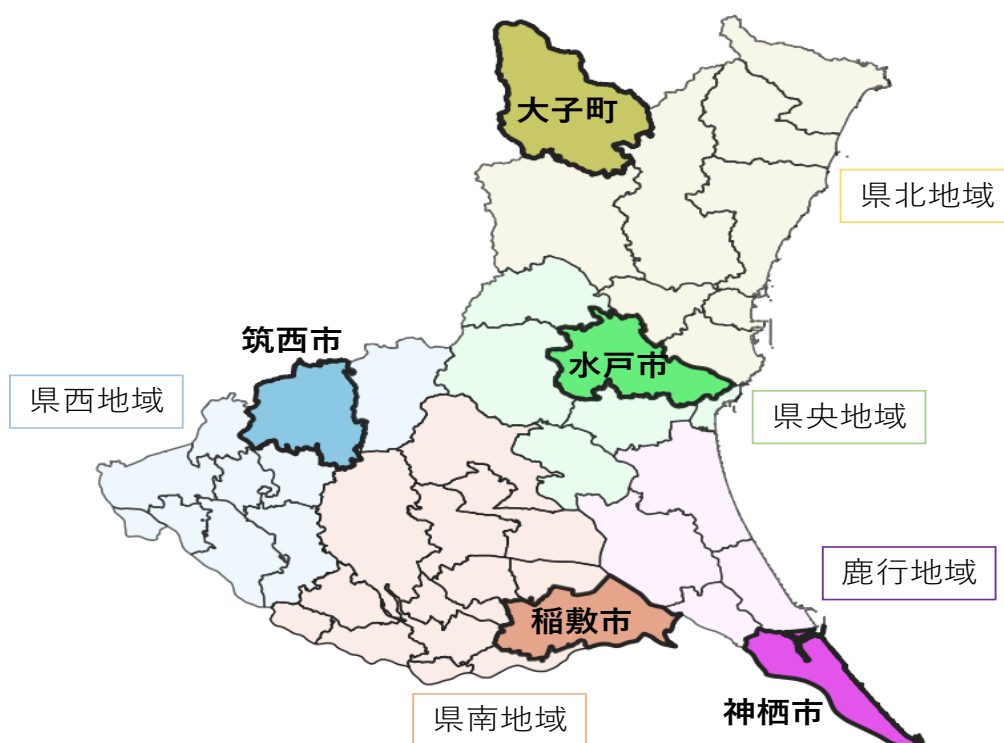


図 4.1 対象地域(茨城県)

4.3.2 使用データ

本研究では、2010年と2015年の農林業センサス個票データを用いた。当データには、個々の農家(経営体)ごとに経営耕地面積や販売金額、農業従事者の年齢等の様々なデータがある。さらに2010年から2015年にかけて、経営体の調査票情報をパネルデータ化したデータがあり、各農家の営農状況の変化が把握可能である(農林業センサスでは、このデータをもとにして、第6巻農林業経営体調査報告書—構造動態編—(農林水産省 2018f)が作成されている)。

本研究では、このデータを用いて、2010年時点で存在する農家が、5年後の2015年において営農を継続しているか、あるいは離農したかという情報(データ)をもとに、モデルを構築した。

さらに、栗原(2014)も述べているように、集落営農に関する情報については、農林業センサス個票でカバーしきれていないため、集落営農へ権利を移管する場合も離農として扱わざるを得ないという点には注意が必要である。また、農林業センサス個票には、各種団体(農協、森林組合等)や地方公共団体・財産区のデータも存在するが、これらは分析の対象からは除外した。

また、分析対象については、農林水産省で定義されている農業経営体(農林水産省 2018a)に該当する農家とした(つまり家族経営体のことを指す)。また、組織経営体(農林水産省 2018a)については対象外とした。理由としては、農林業センサス個票での世帯に関するデータ(経営主の年齢や世帯員数等)については、組織経営体の場合には値が入っておらず、家族経営体とあわせて分析することが難しいため、また、組織経営体は国内の全農業経営体のうちの 1.8%(2010 年)と非常に数が少ないためである。

なお、既存研究にならい、本研究で離農として扱うのは、農林水産省で定義されている農業経営体に該当しなくなった場合とした。これは、農林業センサス個票にて販売金額等の詳細な属性データが存在するのが農業経営体のみであり、農業経営体の定義に属さない農家などは離農の詳細な分析が難しいためである。そのため、農業経営体が、自給的農家(農林水産省 2018a)に移行した場合においても、離農として扱うこととなる。

そして本研究において、離農予測モデルの構築に用いる説明変数の候補は、表 4.1 のとおり計 18 個の変数とした。なお、「農作物の販売金額」、「販売金額が 1 位の部門」については、農林業センサス個票の分類ではカテゴリ数が非常に多いため、カテゴリ数が少なくなるように再分類して、分析を実施した。

また、本研究における分析やモデルの構築には、オープンソースソフトウェアの R 3.4.3 を使用した。

さらに、本論文で示す分析の結果は、いずれも 2010 年と 2015 年農林業センサス調査票情報を元に独自集計を行ったものであることを注記しておく。

表 4.1 離農予測モデルの構築に用いた説明変数の候補

変数名	変数の種類(質的/量的変数)と値の内容	
経営主の年齢	量的	
経営主の性別	質的	0: 男性 1: 女性
後継者の有無	質的	0: 後継者がいない 1: 同居後継者がいる 2: 同居後継者はいないが、他出後継者がいる
農業従事者数	量的	
農外所得の状況	質的	0: 農外所得はない 1: 農外所得があるが、自営農業による所得の方が多い 2: 農外所得があり、自営農業以外の所得の方が多い
世帯員数	量的	
総経営耕地面積	量的	単位は a
貸付農地面積割合	量的	単位は % 算出式は (貸付農地面積/所有している農地面積) ※所有している農地面積(分母)が 0 の場合、割合は 1 とする
耕作放棄地面積割合	量的	単位は % 算出式は (耕作放棄地面積/所有している農地面積) ※所有している農地面積(分母)が 0 の場合、割合は 1 とする
借りている農地面積割合	量的	単位は % 算出式は (借りている農地面積/総経営耕地面積) ※総経営耕地面積(分母)が 0 の場合、割合は 0 とする
何も作らなかった農地面積割合	量的	単位は % 算出式は (何も作らなかった農地面積/総経営耕地面積) ※総経営耕地面積(分母)が 0 の場合、割合は 1 とする
農産物の販売金額	質的	0: 販売無し 1: 50 万円未満 2: 50 万円以上～700 万円未満 3: 700 万円以上
販売金額が 1 位の部門	質的	0: 販売無し 1: 稲作 2: 畑作(麦類、雑穀・いも類・豆類) 3: 園芸作(露地野菜、施設野菜、果樹類、花き・花木) 4: 畜産(酪農、肉用牛、養豚、養鶏、養蚕、その他の家畜) 5: その他(工芸農作物、その他の作物)
農業機械の保有の有無	質的	0: 農業機械を保有していない 1: 農業機械を保有している ※農業機械とは、動力田植機、トラクター、コンバインのいずれかを指す
常雇いの有無	質的	0: 常雇いがいない 1: 常雇いがいる
農作業委託の有無	質的	0: 農作用委託を行っていない 1: 農作業委託を行っている ※委託は、一部の作業も含む
環境保全型農業の取組の有無	質的	0: 環境保全型農業に取り組んでいない 1: 環境保全型農業に取り組んでいる ※環境保全型農業の取組とは、化学肥料の低減、農薬の低減、堆肥による土作りを指す
農業生産関連事業の取組の有無	質的	0: 農業生産関連事業に取り組んでいない 1: 農業生産関連事業に取り組んでいる ※農業生産関連事業の取組とは、農作物の加工、貸し農園・体験農園、観光農園、農家民宿、農家レストラン、海外への輸出等を指す

4.3.3 離農予測シミュレーションのアルゴリズム

本研究では、統計的機械学習(ロジスティック回帰)とマイクロシミュレーションの手法を用いて、市町村単位でのモデルを構築した。

統計的機械学習は、データ駆動型のアプローチをとるため、本研究における地域ごとの特徴の抽出や予測に適している手法といえる。

なお本研究では、統計的機械学習の一手法であるロジスティック回帰を用いた。ロジスティック回帰は、目的変数が0か1かといった2クラスの場合に用いる。本研究では、農家が離農するか、営農継続するかといった2クラスのデータを目的変数として与えるため、当手法を用いた。

またマイクロシミュレーションは、個々の活動主体の行動を確率的にモデル化し、その個々の主体の行動の結果を積み上げて、最終的な予測を行う手法である。農業分野において、マイクロシミュレーションを用いた研究は少ないが、例えば、海外では、農家の収入や農作物市場価格、課税政策などに焦点をあてたO'Donoghue(2017)などの研究が、国内では棚田における農地利用予測を対象とした遠藤(2004)などの研究がある。

本研究では、個々の農家ごとに各々の属性をふまえながら、離農するか営農継続するかを予測し、それら各農家の離農予測結果を積み上げて、地域全体での離農農家数(営農継続農家数)を算出するというモデルを構築した。

マイクロシミュレーションの手法を用いることで、個々の農家の離農予測結果をベースとしたボトムアップ的な解析や予測が可能になるとともに、確率的なモデリングを通じて、各農家の離農行動の不確実性についても考慮することができる。

なお、本研究における離農予測シミュレーションのアルゴリズムの概要は図4.2のとおりである。アルゴリズムは大きく分けて、①離農予測モデルの構築、②各農家の離農予測、③対象地域全体の離農農家数(営農継続農家数)の算出というステップに分かれる。以下、それぞれのステップごとに詳細を説明する。

図 4.2 離農予測シミュレーションモデルのアルゴリズム

①離農予測モデルの構築

対象地域における過去の農家のデータ集合 $D = \{(\mathbf{X}, y)\}$
 $y \in \{0: \text{継続}, 1: \text{離農}\}$ を用いて、ロジスティック回帰による
離農予測モデル $f(\mathbf{x})$ を構築する。

②各農家の離農予測

予測を行う対象地域の農家の数を N とし、ある農家 i の
属性データを \mathbf{x}_i とする。

また、離農農家数をカウントする変数を $n(\text{離農})$ とおく。

農家の分だけ、以下の処理を行う。

For $i = 1$ to N :

- (a)- 農家 i の属性データ \mathbf{x}_i をロジスティック回帰モデル $f(\mathbf{x})$ に
入力し、離農確率 $p(\text{離農})$ を算出する。

$$p(\text{離農}) = p(C_{\text{離農}} | \mathbf{x}_i) = f(\mathbf{x}_i) \quad (C_{\text{離農}}: \text{離農クラス})$$

- (b)- 一様乱数 r ($0 \leq r \leq 1$)を発生させる。

$$\text{if } p(\text{離農}) \geq r: \\ n(\text{離農}) = n(\text{離農}) + 1$$

③対象地域全体の離農農家数(営農継続農家数)の算出

$$[\text{離農農家数}] = n(\text{離農})$$

$$[\text{営農継続農家数}] = N - n(\text{離農})$$

4.3.4 離農予測モデルの構築

本研究では、2010年と2015年の農林業センサス個票のデータを学習データとし、ロジスティック回帰の手法を用いて、各農家が5年後(2010→2015)に離農するか営農継続するかを予測(判別)するモデルを構築した。具体的には、2010年に現存している農家を対象に、2010年時点での属性データを説明変数として用い、2015年における営農状況(離農 or 営農継続)を目的変数として用いることで、モデルを構築した。

また、最終的にモデルで使用する説明変数については、以下のステップ(手順)により変数選択を実施し、決定した。

- I- 説明変数ごとに単変量のロジスティック回帰モデル(1つの説明変数+切片)を構築する。そして、それぞれについて、切片のみのモデルとの尤度比検定を実施し、検定の結果、有意水準 5%で有意であった変数を候補として残し、それ以外は候補から除外する。
- II- Iで候補として残った変数間で総当たりに相関係数を算出し、相関係数が 0.5 以上である変数の組については、どちらか一方を除外する(除外する変数の決定については、Iの単変量のロジスティック回帰モデルで算出された最大対数尤度をもとに、2つの変数のうち、最大対数尤度が小さい変数を除外する)。
- III- IIで残った変数を用いて AIC(赤池情報量規準)による変数選択(変数増減法)を実施し、最終的にモデルで使用する説明変数を決定する。

まず、Iの単変量での分析によって、離農との関連性が低いと考えられる変数について除外する。これを行う理由としては、ロジスティック回帰によるモデルを構築する際に、関連性が低い変数が入っていると、偏回帰係数などのモデルの推定結果が不安定になりやすいためである。また、IIについては、多重共線性の発生を防ぐ目的で実施する。さらに、IIIの AIC は、汎化性能(予測の良さ)を重視したモデル評価指標であり、モデルに投入する変数を変えて算出される AIC が最小となるモデルを、最良のモデルとして採用する。本研究では、離農予測を行うことが目的であるため、予測に重点を置いた AIC を変数選択の基準として用い、最終的にモデルに用いる変数を決定した。

なお、ある説明変数や複数の説明変数の組によって、目的変数を完全に分離可能である(どちらのクラスに属するかを完全に予測(判別)できる)場合、あるいは、ほぼ全てのデータを分離可能な場合は、完全分離・準完全分離と呼ばれ、パラメータの最尤推定量が一般的には存在しない(Albert and Anderson 1984)。このような状態の場合、パラメータの推定が適切に行われず、非常に大きな偏回帰係数や係数の標準誤差が算出される。そのため本研究では、構築されたモデルを確認し、ある変数について完全分離・準完全分離が生じている可能性が高い場合は、その変数を除いて再度モデルの構築を行った。

そして、構築した離農予測モデルでは、ある説明変数の組 \mathbf{x} を入力することによって、以下のように離農クラスに所属する確率が事後確率として算出される(目的変数について離農クラスを 1、営農継続クラスを 0 として学習した場合)。

$$p(C_{\text{離農}} | \mathbf{x}) = \frac{1}{1 + \exp[-(\mathbf{w}^T \mathbf{x} + b)]} \quad (4.1)$$

($C_{\text{離農}}$: 離農クラス, \mathbf{w} : 偏回帰係数, b : 切片, \mathbf{x} : 説明変数)

また本研究では、市町村単位でモデルを構築するため、市町村ごとにモデルに採用される変数や偏回帰係数の値は異なってくる。

4.3.5 各農家の離農予測

次に、上述の手順により構築した離農予測モデルを用いて、各農家の離農を予測する。具体的には、式(4.1)に各農家の属性情報(説明変数の値)を入力することで算出される離農クラスの事後確率を、各農家が離農行動をとる離農確率 $p(\text{離農})$ と設定する。そして、一様乱数 $r(0 \leq r \leq 1)$ を発生させて、 $p(\text{離農}) \geq r$ の場合、離農と予測する。逆に、 $p(\text{離農}) < r$ の場合は、営農継続と予測する。

なお、農家ごとに属性情報が異なるため、各農家で異なる離農確率が設定されて予測が行われる。

4.3.6 対象地域全体の離農農家数の算出

各農家の離農予測結果を合算し、対象地域全体の離農農家数(営農継続農家数)を算出する。離農農家数は、対象地域において離農と判定された農家の数を足し上げることで算出を行う。

4.3.7 シミュレーションの試行回数と予測精度の検証

本研究における離農予測シミュレーションのアルゴリズムでは、②各農家の離農予測において乱数を用いているため、予測値が毎回異なる。また、このような乱数を用いたモンテカルロシミュレーションでは、予測値の標準誤差は試行回数の平方根に反比例して減少するため、高い推定精度を得るためには、十分な試行回数が必要だとされる。

そこで本研究では、シミュレーションの試行回数を 10000 回とし、最終的な予測値は、10000 回の試行の平均をとった(なお、真値である離農農家数は整数であるため、予測値も最終的に小数点以下を四捨五入し、整数の値とした)。つまり、予測の際は、②各農家の離農予測から③対象地域全体の離農農家数(営農継続農家数)の算出までの一連の流れを 10000 回繰り返し、予測値の平均をとる。

なお、①離農予測モデルの構築では、毎回構築されるモデルは同じであるため、最初の1回しか実行しない。

また、シミュレーションモデルの妥当性の確認のために、未知のデータを当てはめた場合の予測精度について検証を行う必要がある。そこで本研究では、交差検定(Cross Validation)を用いて精度を検証した。交差検定ではデータの分割数 k を決定し、分割で生じた k グループのうち $k-1$ 個のグループを学習データとして、残り1つのグループを予測データとして使用する。そして、予測データに使用する1つのグループを変えて k 回繰り返したときの予測結果の平均を評価指標として用いる。なお、分割数 k については、 n をサンプルサイズとした場合の以下のスタージェスの公式をもとに決定した。

$$k = 1 + \frac{\log n}{\log 2} \quad (4.2)$$

なお、データの分割は、分割後の2クラス(目的変数)の比率を考慮せずに、完全にランダムに分割した。

精度の評価には、予測した離農農家数(営農継続農家数)と実際の離農農家数(営農継続農家数)の平均絶対誤差(MAE)を用いた。平均絶対誤差の算出式は以下のとおりである。

$$MAE = \frac{1}{k} \sum_{i=1}^k |\hat{y}_i - y_i| \quad (4.3)$$

(k : 試行回数(分割数), \hat{y}_i : i 番目の試行の予測値, y_i : i 番目の試行の真値)

4.3.8 将来の離農農家数の予測

さらに本研究では、将来予測として、2020年(2015年の5年後)の離農農家数の予測を行った。上述した離農予測モデルに、2015年時点で現存している農家の属性データを入力して離農確率を算出し、予測を行った。なお、将来予測に関しては、過去から現在までのデータを用いて構築したモデルにより予測を行うため、離農の状況についてこれまでと同じ傾向が続いた場合であること仮定している。

また、各農家の属性の値(例えば、経営主の年齢)を更新しながら、離農予測シミュレーションを繰り返し行うことで、5年後より先の離農予測についても当然ながら可能である(値をどのように更新するかは別途検討が必要である)。

4.3.9 都道府県全体の離農農家数の予測手法に関する比較

最後に参考として、都道府県全体の離農農家数(営農継続農家数)を予測する手法に関して2つの手法(モデル)を提案し、精度の比較を行った。

1つ目は、これまでに述べたような市町村単位での離農予測モデルを用いて、各市町村別に離農農家数を予測し、最終的にそれらを合算して都道府県全体の離農農家数を予測する手法である(以下、略称として「市町村累計モデル」と呼ぶこととする)。

2つ目は、都道府県の全ての農家のデータを用いて、都道府県として1本の離農予測モデルを構築し、それを用いて予測を行う手法である(以下、略称として「都道府県モデル」と呼ぶこととする)。つまり、この手法では、1本のモデルを都道府県内の全ての農家にあてはめて離農予測を行い、その結果を積み上げて、都道府県全体の離農農家数を予測する。

本研究では、2つの手法について、対象地域である茨城県全体の離農農家数の予測を行い、精度の比較を行った。精度の算出には交差検定を使用し、厳密な比較を行うため、いずれの手法でも交差検定は10分割とし、各分割データの中身も全く同じものを用いた。

4.4 結果

4.4.1 農家の離農の現状

まずは、対象地域における農家の離農の現状を示す。表4.2は、茨城県の5市町村と茨城県全体における2010年から2015年の間における営農継続農家数、離農農家数と離農率を示したものである。

離農率は、各市町村でいずれも20%前後となっており、かなりの割合で農家が離農していることが見てとれる。また稲敷市は、他の市町村に比べると離農率がやや低いことが分かる。

表 4.2 農家の離農の状況 (2010~2015 年)

	全農家数 (2010 年時点)	営農継続 農家数 (2015 年時点)	離農 農家数 (2015 年時点)	離農率
神栖市	1124	874	250	22.2%
水戸市	3248	2552	696	21.4%
筑西市	3640	2825	815	22.4%
稲敷市	2359	1935	424	18.0%
大子町	1195	928	267	22.3%
茨城県全体	71001	56102	14899	21.0%

4.4.2 離農予測において重要な変数

本研究では、ロジスティック回帰の手法を用いて離農予測モデルを構築したが、このモデルにおいて、最終的に採用された説明変数とその変数の偏回帰係数の値(またはオッズ比)や有意性検定の結果等を見ることによって、離農予測の際に重要な変数について把握することが可能である。

表 4.3 は、前述した変数選択のステップを経て、最終的にモデルで採用された変数や途中のステップまでは候補として残った変数について、対象地域である茨城県 5 市町村別に示したものである。

表 4.3 変数選択により候補として残った変数

	神栖市	水戸市	筑西市	稲敷市	大子町
経営主の年齢	○	○	○	○	○
経営主の性別		△	△	○	△
後継者の有無		○	○	○	○
農業従事者数	○	○	○	○	○
農外所得の状況	○	○	△	△	△
世帯員数		■	■	■	■
総経営耕地面積	○	■	■	○	○
貸付農地面積割合	○	○	○	○	
耕作放棄地面積割合		○	○	○	△
借りている 農地面積割合		△	○	○	△
何も作らなかつた 農地面積割合	○	○	△		
農産物の販売金額	○	○	○	■	■
販売金額が1位の 部門	■	■	■	■	■
農業機械の保有の 有無	△	○	○	○	○
常雇いの有無	△	△	△		△
農作業委託の有無		○	○	○	
環境保全型農業の 取組の有無	△	○	○	○	○
農業生産関連事業の 取組の有無	※	△	△		△

注)表中の記号は以下の点を示す。

○：最終的にモデルで採用された変数

△：ステップⅡまでの処理では候補として残ったが、ステップⅢのAICによる変数選択で除外された変数

■：ステップⅠによる尤度比検定では候補として残ったが、ステップⅡで他の変数との相関係数が高かったために除外された変数

※：当初モデルの変数として採用されたが、準完全分離の状態であったため除外された変数

(空白の場合は、ステップⅠで除外されたことを示す)

また表 4.4~4.8 は、各市町村にて構築された離農予測モデルの結果を示したものである。なお、表に示すモデルは市町村内の全ての農家(家族経営体)のデータを用いて構築したものである。

表には、モデルで採用された説明変数のオッズ比、オッズ比の 95%信頼区間、偏回帰係数に対する Wald 検定による p 値を示している。オッズ比とは説明変数が 1 単位増加した場合(質的変数では、そのカテゴリに該当した場合)に、オッズが何倍になるかを表したものである。なお、オッズは、離農クラスの事後確率 $p(C_{\text{離農}} | \mathbf{x})$ と営農継続クラスの事後確率 $p(C_{\text{継続}} | \mathbf{x})$ に対して、以下のような式で表される。

$$\frac{p(C_{\text{離農}} | \mathbf{x})}{1 - p(C_{\text{離農}} | \mathbf{x})} = \frac{p(C_{\text{離農}} | \mathbf{x})}{p(C_{\text{継続}} | \mathbf{x})} \quad (4.4)$$

つまり、オッズ比が 1 より大きい場合、その説明変数が増加(質的変数ではそのカテゴリに該当)すると離農確率が上昇することを意味する。逆にオッズ比が 1 より小さい場合は、離農確率が低下することを意味する。

モデルの結果を見ると、経営主の年齢、農業従事者数は全ての市町村のモデルで採用され、オッズ比の値も類似し、かつ検定結果も有意であり、各地域で農家の離農に大きく影響していることが推察される。経営主の年齢が高く、農業従事者数が少ないほど、離農する確率は高くなり、各農家における労働力の状況が離農に関係していることが分かる。

また、総経営耕地面積と農作物の販売金額については、全ての地域で、少なくともどちらか一方はモデルの変数として採用され、かつ有意になっている。なお神栖市を除く 4 市町村(水戸市、筑西市、稲敷市、大子町)では、上記 2 つの変数間の相関が強く、変数選択のステップⅡにて、一方が除外され、候補として残った側の変数が採用される結果となっている。また神栖市では両方とも変数として採用され、総経営耕地面積のみ有意になっている。これら 2 つの変数は、農家の規模を表す変数であるといえ、規模が大きい(総経営耕地面積が大きい、または農作物の販売金額が大きい)ほど離農確率は低く、逆に規模が小さいほど離農確率は高くなることが分かる。このように、農家の経営規模も離農に大きく影響していることが見てとれる。

また、農業機械の保有の有無や環境保全型農業の取組の有無は、神栖市を除く

4市町村で採用され、オッズ比の値も似ており、かつ有意である。これらの変数も、離農との関連性が高いことが見てとれる。オッズ比から、農業機械を保有している農家や環境保全型農業の取組を行っている農家は離農確率が低くなる傾向にある。一方、後継者の有無も神栖市を除く4市町村で採用されているが、一部の市町村やカテゴリ(同居・他出後継者有)でのみ有意になっているため、オッズ比にややばらつきが見られる。なお以上のように、平地農業地域・中間農業地域・山間農業地域が中心である4市町村と都市的地域が大半を占める神栖市で採用される変数にやや異なる傾向が見られる。

一方で、神栖市や、都市的地域を半数程度含む水戸市では、他市町村とは異なり、農外所得の状況や何も作らなかつた農地面積割合が採用されているといった特徴が見られる。農外所得の状況については、有意にはなっていないものの、都市的地域は、農外所得が他地域に比べて比較的得られやすい環境にあり、その影響の存在がモデルの変数の選択にも反映されていると考えられる。

さらに、農作業委託の有無や貸付農地面積割合は、平地農業地域を持つ3市町村(水戸市、筑西市、稲敷市)でいずれも採用されており(貸付農地面積割合は、稲敷市では有意でないが)、平地農業地域では、このような農地の流動化に関連する変数も農家の離農に影響していることが見てとれる。平地農業地域では、大規模化が比較的行いやすく、農地の受け皿となる農家を中心とした農地の流動化の進展が、離農にも影響を与えていることが考えられる。

また以上のように、市町村や農業地域によって採用される変数にやや違いが見られ、離農の傾向が一部異なる可能性があることが推察される。

表 4.4 ロジスティック回帰による離農予測モデル(神栖市)

	オッズ比 (95%信頼区間)	p 値
経営主の年齢	1.024 (1.009-1.040)	0.0020*
経営主の性別 (1: 女性)		
後継者の有無 (1: 同居後継者有)		
後継者の有無 (2: 他出後継者有)		
農業従事者数	0.855 (0.734-0.996)	0.0437*
農外所得の状況 (1: 農外所得有、自営農業所得の方が多い)	0.880 (0.549-1.411)	0.5955
農外所得の状況 (2: 農外所得有、農外所得の方が多い)	1.489 (0.995-2.229)	0.0530
総経営耕地面積	0.996 (0.993-0.998)	0.0004*
貸付農地面積割合	1.770 (0.863-3.632)	0.1193
耕作放棄地面積割合		
借りている農地面積割合		
何も作らなかった農地面積割合	2.766 (1.243-6.158)	0.0127*
農作物の販売金額 (1: 50 万円未満)	0.913 (0.488-1.707)	0.7745
農作物の販売金額 (2: 50 万円以上～700 万円未満)	0.628 (0.357-1.106)	0.1070
農作物の販売金額 (3: 700 万円以上)	0.259 (0.128-0.524)	0.0002*
農業機械の保有の有無 (1: 保有している)		
農作業委託の有無 (1: 委託を行っている)		
環境保全型農業の取組の有無 (1: 取り組んでいる)		

注) *は有意水準 5%で有意であることを示す。

表 4.5 ロジスティック回帰による離農予測モデル(水戸市)

	オッズ比 (95%信頼区間)	<i>p</i> 値
経営主の年齢	1.025 (1.016-1.034)	<0.0001*
経営主の性別 (1: 女性)		
後継者の有無 (1: 同居後継者有)	0.876 (0.685-1.120)	0.2922
後継者の有無 (2: 他出後継者有)	0.707 (0.553-0.903)	0.0054*
農業従事者数	0.754 (0.680-0.837)	<0.0001*
農外所得の状況 (1: 農外所得有、自営農業所得の方が多い)	0.628 (0.411-0.960)	0.0316*
農外所得の状況 (2: 農外所得有、農外所得の方が多い)	0.941 (0.726-1.219)	0.6436
総経営耕地面積		
貸付農地面積割合	1.845 (1.133-3.005)	0.0138*
耕作放棄地面積割合	1.737 (1.021-2.954)	0.0416*
借りている農地面積割合		
何も作らなかった農地面積割合	2.053 (1.343-3.137)	0.0009*
農作物の販売金額 (1: 50 万円未満)	0.454 (0.353-0.583)	<0.0001*
農作物の販売金額 (2: 50 万円以上～700 万円未満)	0.217 (0.164-0.288)	<0.0001*
農作物の販売金額 (3: 700 万円以上)	0.037 (0.009-0.156)	<0.0001*
農業機械の保有の有無 (1: 保有している)	0.664 (0.520-0.848)	0.0010*
農作業委託の有無 (1: 委託を行っている)	1.239 (1.025-1.498)	0.0270*
環境保全型農業の取組の有無 (1: 取り組んでいる)	0.699 (0.562-0.869)	0.0012*

注) *は有意水準 5%で有意であることを示す。

表 4.6 ロジスティック回帰による離農予測モデル(筑西市)

	オッズ比 (95%信頼区間)	p 値
経営主の年齢	1.036 (1.028-1.045)	<0.0001*
経営主の性別 (1: 女性)		
後継者の有無 (1: 同居後継者有)	0.602 (0.492-0.736)	<0.0001*
後継者の有無 (2: 他出後継者有)	0.687 (0.516-0.913)	0.0097*
農業従事者数	0.865 (0.793-0.944)	0.0011*
農外所得の状況 (1: 農外所得有、自営農業所得の方が多い)		
農外所得の状況 (2: 農外所得有、農外所得の方が多い)		
総経営耕地面積		
貸付農地面積割合	2.340 (1.634-3.351)	<0.0001*
耕作放棄地面積割合	3.834 (1.797-8.178)	0.0005*
借りている農地面積割合	0.444 (0.277-0.713)	0.0008*
何も作らなかった農地面積割合		
農作物の販売金額 (1: 50 万円未満)	0.544 (0.370-0.800)	0.0020*
農作物の販売金額 (2: 50 万円以上～700 万円未満)	0.268 (0.183-0.392)	<0.0001*
農作物の販売金額 (3: 700 万円以上)	0.053 (0.027-0.105)	<0.0001*
農業機械の保有の有無 (1: 保有している)	0.732 (0.570-0.940)	0.0144*
農作業委託の有無 (1: 委託を行っている)	1.297 (1.089-1.545)	0.0036*
環境保全型農業の取組の有無 (1: 取り組んでいる)	0.679 (0.557-0.828)	0.0001*

注) *は有意水準 5%で有意であることを示す。

表 4.7 ロジスティック回帰による離農予測モデル(稲敷市)

	オッズ比 (95%信頼区間)	p 値
経営主の年齢	1.027 (1.015-1.039)	<0.0001*
経営主の性別 (1: 女性)	1.747 (1.101-2.773)	0.0180*
後継者の有無 (1: 同居後継者有)	0.630 (0.481-0.824)	0.0008*
後継者の有無 (2: 他出後継者有)	1.131 (0.798-1.603)	0.4900
農業従事者数	0.841 (0.749-0.945)	0.0036*
農外所得の状況 (1: 農外所得有、自営農業所得の方が多い)		
農外所得の状況 (2: 農外所得有、農外所得の方が多い)		
総経営耕地面積	0.998 (0.997-0.999)	0.0003*
貸付農地面積割合	1.801 (0.946-3.430)	0.0734
耕作放棄地面積割合	2.080 (0.903-4.788)	0.0853
借りている農地面積割合	0.541 (0.283-1.034)	0.0631
何も作らなかった農地面積割合		
農作物の販売金額 (1: 50 万円未満)		
農作物の販売金額 (2: 50 万円以上～700 万円未満)		
農作物の販売金額 (3: 700 万円以上)		
農業機械の保有の有無 (1: 保有している)	0.665 (0.501-0.881)	0.0045*
農作業委託の有無 (1: 委託を行っている)	1.344 (1.062-1.700)	0.0140*
環境保全型農業の取組の有無 (1: 取り組んでいる)	0.622 (0.481-0.804)	0.0003*

注) *は有意水準 5%で有意であることを示す。

表 4.8 ロジスティック回帰による離農予測モデル(大子町)

	オッズ比 (95%信頼区間)	p 値
経営主の年齢	1.031 (1.016-1.046)	0.0001*
経営主の性別 (1: 女性)		
後継者の有無 (1: 同居後継者有)	0.685 (0.463-1.014)	0.0589
後継者の有無 (2: 他出後継者有)	0.539 (0.357-0.814)	0.0033*
農業従事者数	0.827 (0.694-0.986)	0.0340*
農外所得の状況 (1: 農外所得有、自営農業所得の方が多い)		
農外所得の状況 (2: 農外所得有、農外所得の方が多い)		
総経営耕地面積	0.982 (0.976-0.987)	<0.0001*
貸付農地面積割合		
耕作放棄地面積割合		
借りている農地面積割合		
何も作らなかった農地面積割合		
農作物の販売金額 (1: 50 万円未満)		
農作物の販売金額 (2: 50 万円以上～700 万円未満)		
農作物の販売金額 (3: 700 万円以上)		
農業機械の保有の有無 (1: 保有している)	0.515 (0.374-0.710)	<0.0001*
農作業委託の有無 (1: 委託を行っている)		
環境保全型農業の取組の有無 (1: 取り組んでいる)	0.675 (0.468-0.972)	0.0346*

注) *は有意水準 5%で有意であることを示す。

4.4.3 シミュレーションモデルの予測精度

続いて、交差検定によりシミュレーションモデルの予測精度を検証した結果を示す。表 4.9 は、対象地域である各市町村について、交差検定によって算出された離農農家数(営農継続農家数)の平均絶対誤差を示したものである。また参考として、交差検定の各テストデータ(分割された 1 グループのデータ)における営農継続農家数と離農農家数の平均値を示している。

対象地域によって、農家の総数が異なるため、それに応じて誤差の値も変わってくるが、いずれの市町村でも農家数の平均絶対誤差は、2~4 程度となっており、一定程度の高い精度で予測が出来ていることが分かる。

表 4.9 対象地域における交差検定の結果

	交差検定 分割数	平均絶対誤差 (±標準誤差)	各テストデータの 真値の平均	
			営農継続 農家数	離農 農家数
神栖市	11	2.73 (±0.65)	79.5	22.7
水戸市	13	4.31 (±0.80)	196.3	53.5
筑西市	13	4.08 (±0.93)	217.3	62.7
稲敷市	12	3.08 (±0.93)	161.3	35.3
大子町	11	2.09 (±0.59)	84.4	24.3

4.4.4 将来における離農農家数と離農率

また、シミュレーションモデルを用いて、2020 年の離農農家数を予測した結果を表 4.10 に示す。表には、2015 年から 2020 年の間における営農継続農家数、離農農家数と離農率を示している。

結果を見ると、いずれの市町村においても、離農率は 20%前半から中盤の高い値となっている。また、2010 年から 2015 年の間における離農率(表 4.2)と比べると、いずれの市町村でも、ほぼ同程度かそれを上回る値となっている。今後も

これまでと同様に、一定数の農家が離農するだけでなく、市町村によっては、さらに離農が進む可能性があることが推察される。特に筑西市と大子町は、離農率が25%台と高い値となっており、2010年から2015年の期間と比較しても、離農率が大きく上昇している。これらの市町村では、これまでよりも離農がより一層進む可能性があることが見てとれる。

表 4.10 将来予測シミュレーションの結果(2015~2020年)

	全農家数 (2015年時点)	営農継続 農家数 (2020年時点)	離農 農家数 (2020年時点)	離農率
神栖市	899	699	200	22.2%
水戸市	2644	2038	606	22.9%
筑西市	2887	2145	742	25.7%
稲敷市	1989	1580	409	20.6%
大子町	952	706	246	25.8%

4.4.5 都道府県全体の予測精度

最後に、市町村累計モデルと都道府県モデルの2つの手法によって、茨城県全体の離農農家数(営農継続農家数)の予測を行った場合の精度の検証結果について示す。

表 4.11 は、交差検定によって算出された市町村累計モデルと都道府県モデルの平均絶対誤差を示したものである。結果として、僅かな差ではあるが、市町村累計モデルの方が都道府県モデルに比べて誤差が小さく、精度の良い予測が出来ている。

表 4.11 2つの手法(モデル)における交差検定の結果

	平均絶対誤差 (±標準誤差)	各テストデータの真値の平均	
		営農継続 農家数	離農 農家数
市町村累計モデル	32.5 (±7.53)	5610.2	1489.9
都道府県モデル	34.9 (±6.13)		

注)交差検定は、両モデルで10分割とし、全く同じデータを使用している。

さらに、交差検定による予測結果を市町村別に詳しく分析した結果を以下に示す。具体的には、茨城県全体の予測結果をさらに掘り下げて、市町村ごとに離農農家数の予測値と真値とを比較した。なお各市町村の予測値について、市町村累計モデルでは、市町村ごとのモデルの予測結果をそのまま用い、都道府県モデルについては、1本のモデルで予測した各農家の結果を、その農家が属する市町村ごとに分けて、離農農家数を市町村別に再集計した。

表 4.12 は、茨城県内の全 44 市町村のそれぞれにおいて、市町村累計モデルと都道府県モデルの誤差を比較し、市町村累計モデルの方が誤差の小さかった市町村と、都道府県モデルの方が誤差の小さかった市町村の数をそれぞれ集計したものである。また、誤差の小さいモデルは、他方のモデルと比較して、誤差がどの程度減少しているか(式で表すと、[他方のモデルの誤差－当該モデルの誤差])についても示している(なお表の値は、該当市町村の誤差の減少分の平均をとった値であり、市町村累計モデルは、市町村累計モデルの方が誤差の小さかった 30 市町村の平均、都道府県モデルは、都道府県モデルの方が誤差の小さかった 12 市町村の平均をとっている)。さらに参考として表 4.13 に、各市町村における、2つのモデルの実際の平均絶対誤差を示している。

結果を見ると、市町村累計モデルの方が誤差の小さかった市町村が、都道府県モデルのそれに比べて多いことが見てとれる。また、誤差の減少分の値についても市町村累計モデルの方が 2.60 と大きく、都道府県モデルは 0.47 と誤差の減少はほんの僅かであった。表 4.13 を見ても、下妻市、常総市、常陸太田市、筑西市などでは、市町村累計モデルの方が都道府県モデルよりも非常に誤差が小さく、予測精度が良い。これは、市町村累計モデルが各市町村固有の離農の傾向を

学習できているのに対して、都道府県モデルではそれらを学習できていないと
いうことを示していると考えられる。一方で、都道府県モデルの方が誤差の小さ
い市町村もあるが、市町村累計モデルとの誤差の違いは非常に小さく、予測結果
はそれほど変わらないと考えられる。

結果として、市町村累計モデルの方が、都道府県全体の予測精度も市町村ごと
の予測精度も良好な傾向にあることが見てとれる。

表 4.12 2つの手法(モデル)における市町村別の誤差の比較

予測誤差の市町村別の比較 (市町村累計モデル vs 都道府県モデル)	該当 市町村数	他方のモデルに対する 誤差の減少分 (平均)
市町村累計モデルの誤差の方が小さい	30	2.60
都道府県モデルの誤差の方が小さい	12	0.47

注) 他方のモデルに対する誤差の減少分の値は、(他方のモデルの誤差－当該
モデルの誤差)について該当市町村数の平均をとったものである。

また、市町村累計モデルと都道府県モデルの誤差が等しい(上記のどちらに
も該当しない)市町村は、2市町村存在した。

表 4.13 各市町村における2つの手法(モデル)の平均絶対誤差

	市町村累計	都道府県
水戸市	6.8	7.3
日立市	3.2	6.2
土浦市	3.3	4.5
古河市	4.8	5.6
石岡市	5.1	9.6
結城市	2.3	1.6
龍ヶ崎市	3.3	6.0
下妻市	4.8	10.1
常総市	5.1	11.2

常陸太田市	4.3	17.2
高萩市	2.4	2.3
北茨城市	4.8	3.9
笠間市	4.5	15.6
取手市	2.6	2.8
牛久市	2.2	2.2
つくば市	5.2	4.7
ひたちなか市	2.3	2.4
鹿嶋市	2.6	3.1
潮来市	3.3	3.7
守谷市	1.3	1.6
常陸大宮市	4.5	6.2
那珂市	5.4	9.1
筑西市	8.6	15.1
坂東市	4.0	4.1
稲敷市	5.9	8.2
かすみがうら市	3.3	5.8
桜川市	4.9	5.7
神栖市	2.1	4.2
行方市	5.3	4.6
銚田市	2.5	3.8
つくばみらい市	4.0	3.8
小美玉市	4.7	7.4
茨城町	3.7	3.5
大洗町	2.0	1.8
城里町	5.5	4.5
東海村	1.6	3.1
大子町	3.5	3.3
美浦村	3.1	3.9
阿見町	1.7	1.8
河内町	3.2	4.2
八千代町	4.3	3.9
五霞町	2.4	2.4
境町	3.1	4.4
利根町	2.3	1.8

4.5 考察

本研究では、全国、地方、都道府県や都道府県内の地域エリア単位でのモデル構築が多い既存研究とは異なり、市町村単位での離農予測シミュレーションモデルの構築を試みた。また、統計的機械学習(ロジスティック回帰)とマイクロシミュレーションを組み合わせた新たな予測手法を提案した。

まず、農家の離農の現状として、対象地域とした茨城県の5市町村の離農率はいずれも20%前後となっており、高い割合で農家が離農していることが分かった。

また、ロジスティック回帰による離農予測モデルでは、多数の変数を候補とし、統計的な手続きに基づいて変数選択と予測を行うモデルを構築した。

構築されたモデルの結果から、いずれの市町村においても、経営主の年齢や農業従事者数が離農に対して強く影響を与えていることが分かった。また、農家の規模に関連した変数(総経営耕地面積、農作物の販売金額)も同様に、各市町村共通で離農に大きく関係していることが示唆された。

経営主の年齢、農業従事者数、総経営耕地面積については、地方や都道府県を対象に分析した研究(栗原ら 2014, 安武・千田 2013, 渡部ら 2012)の結果とも一致する。一方で、農作物の販売金額と離農との関係について、本研究のように指摘したものはほとんど見られないが、総経営耕地面積と相関が強いため、代わりに総経営耕地面積で分析を実施している研究が多いからだということも考えられる。

一方、農業機械の保有の有無、環境保全型農業の取組の有無、後継者の有無、農作業委託の有無、貸付農地面積割合、何も作らなかった農地面積割合、農外所得の状況などは、農業地域類型区分が異なる市町村間によって、モデルに採用されるか(かつ有意であるか)どうかに違いがあり、市町村や農業地域によって離農の傾向が一部異なることが分かった。営農環境が異なる各地域によって、離農に影響を与える要素は違ってくると考えられ、それらをモデルに取り込み、予測を行うことが重要であると考えられる。

本研究は、既存研究とは異なり、市町村単位でモデルを構築しているため、市町村レベルでのより細かなスケールで、各地域の離農の特徴について把握が可能である。実際、上述のように各市町村で離農と関連の強い変数について異なる点も見られた。今後も各市町村や各農業地域での離農の傾向について、さらに掘り下げて分析していく必要があると考える。

なお、対象地域のモデルでは、多くの説明変数が採用され、かつ検定の結果、有意であった。このことから、実際に多数の要素が離農に影響を与えていると推察され、本研究で提案したような多数の変数を扱うことのできるモデルを用いることが有効であると考えられる。

また本研究では、離農予測シミュレーションモデルの妥当性を確認するために、交差検定を通じてモデルの精度を検証した。農業地域類型区分が異なる複数の市町村を対象地域として選定し予測を実施したが、いずれの市町村でも平均絶対誤差は2~4程度となっており、様々な地域において、一定程度の高い精度で予測が可能なが分かった。

さらに、構築したモデルを用いて、将来(2020年)の離農農家数の予測シミュレーションを行った。結果として、多くの市町村でこれまでよりも離農率が上昇し、今後、農家の離農がさらに進むことが示唆された。農業従事者の年齢構成等から考えても、離農が進むことはある程度やむを得ないと考えられるが、一方で地域農業を維持していくための対策を早急を実施する必要があると考えられる。

また参考として、都道府県全体の離農農家数(営農継続農家数)を予測する手法に関して、市町村累計モデルと都道府県モデルという2つの手法を提案し、それらの精度の比較を行った。結果として、茨城県全体の離農農家数を予測した場合、市町村累計モデルの方が、僅かではあるが、誤差が小さかった。また、予測結果を市町村ごとに掘り下げて分析すると、市町村累計モデルの方が誤差の小さかった市町村が多く、都道府県モデルに対する誤差の減少分の値も大きかった。これは、市町村累計モデルが、各市町村の固有の特徴を学習できているのに対し、都道府県モデルでは、都道府県全体の傾向しか学習ができていないためであると考えられる。都道府県の離農農家数を予測する場合、既存研究では、都道府県全体の1本のモデルで予測を行っている場合が多いが、本研究で提案したような、市町村累計モデルにより各市町村の離農農家数を予測し、それらを積み上げる手法が有効であることが分かった。

なお、本研究で提案した手法の利点をまとめると以下の通りになる。

まず、上述のように、市町村単位でのモデルを構築しているため、各市町村の特徴をふまえた予測を実施できるとともに、予測結果をもとにした、よりきめ細やかな対策が可能になるという利点がある。また、個々の農家の離農予測を行い、その結果を積み上げるといったマイクロシミュレーションの手法を用

いているため、例えば市町村以外のスケールでの予測も可能である。より広いスケールである都道府県の予測は本論文で示したとおりであり、さらに地区や集落といったより小さいスケールでも予測を実施することができる。ただし、小さなスケールの場合は、本研究のような市町村単位のモデルを用いてどの程度の予測精度が得られるかは別途、検証が必要であると考えられる(また、地区や集落単位でモデルを構築することも考えられるが、サンプルサイズが少ない場合は、構築されるモデルが不安定になる可能性がある)。

さらに、ロジスティック回帰の手法を用いて、離農予測モデルを構築することで、離農に強く影響している変数についても把握することができる。この点は、安武(2018)が述べるように、因果関係を反映することができないマルコフモデルとは大きく異なる点である。また、多数の変数をモデルに組み込むことが可能であるという利点も有している。さらに、個々の農家ごとに離農確率(事後確率)を算出できるため、算出された離農確率を地域全体の農家数の予測に使用するだけでなく、個々の農家の離農リスクとしてとらえることで、各地域において、農家ごとのきめ細やかな対策も可能になると考えられる。

加えて、本研究で提案した手法は、統計的な手続きによる変数選択と予測を行っているため、他の地域でも同様に適用が可能な汎用性を有している。

今後の課題として、本研究では対象外とした組織経営体のモデルの構築について検討を行う必要がある。ただし、1市町村における組織経営体の数は非常に少ないため、市町村単位でのモデルの構築は現実的でない。そのため、都道府県等のより大きなスケールを単位としてモデルを構築せざるを得ないと考えられる。

また、今回は離農を対象とし、現存する農家のうち、将来的にどの程度の農家が離農するかについて予測を行ったが、将来的な農家の数を正確に算出しようと思うと、新規農家の数についても考慮する必要がある。新規農家の数は、それほど多くはないが、今後は、新規農家の数についても予測が可能な手法を考える必要があると考えられる。

さらに、本研究では、農林業センサス個票のデータのみを用いているため、離農予測モデルの構築に際して、外的要因(政策の影響等)を考慮できていない。さらなるモデルの精緻化のためには、そのような外的要因をモデルに取り込むことが必要であると考えられる。

あわせて、本研究で対象とした地域以外においても、同様の手法で予測を実

施し、精度の検証を行うことで、モデルの汎用性や有効性についてさらに検証していく必要がある。

以上のような課題があるものの、本研究で提案した手法は、各地域における農地保全の取組の推進に役立てることができるものであると考える。

第5章 結論

以上のように、本研究では、農地の土地利用状況の変化や、農家の営農状況の変化といった地域農業の動向について、予測モデルを構築し、要因分析や将来予測を実施した。ここで、第1章で示した本研究の目的の「1. 高精度かつ汎用性のある地域農業動向予測モデルの構築」をふまえて、本研究における分析の手順や、使用したデータ、手法等について整理し、それらの特徴や用いることの意義について述べる。

本研究の分析の手順をまとめたものを、図5.1左部に示している。まず、対象地域の決定を行い、分析に必要なデータを収集する。次に、収集したデータをもとに分析に使用する変数を決定する。その後、現状の把握を行うとともに、モデルを構築し、要因分析や将来予測を実施する。また最後に、得られた分析・予測結果から農地保全のための適切な対応策を検討するといった流れになる。

そして、地域農業動向の分析に用いるデータや手法は、様々なものが考えられる。特に、対象地域や分析のスケールによっても、それらは大きく変わってくる。本研究においても、分析について複数のスケールを考慮した。今回は、それらについて、地域レベルと農地・農家レベルの2つに分類して、整理を行った(図5.1)。具体的には、第2章の分析が地域レベル、第3章と第4章の分析が農地・農家レベルに該当する。

図では、地域レベルと農地・農家レベルにおける、対象地域のスケール、使用したデータ、モデル構築に用いた変数と手法、ならびに対応策の検討の内容について示している。

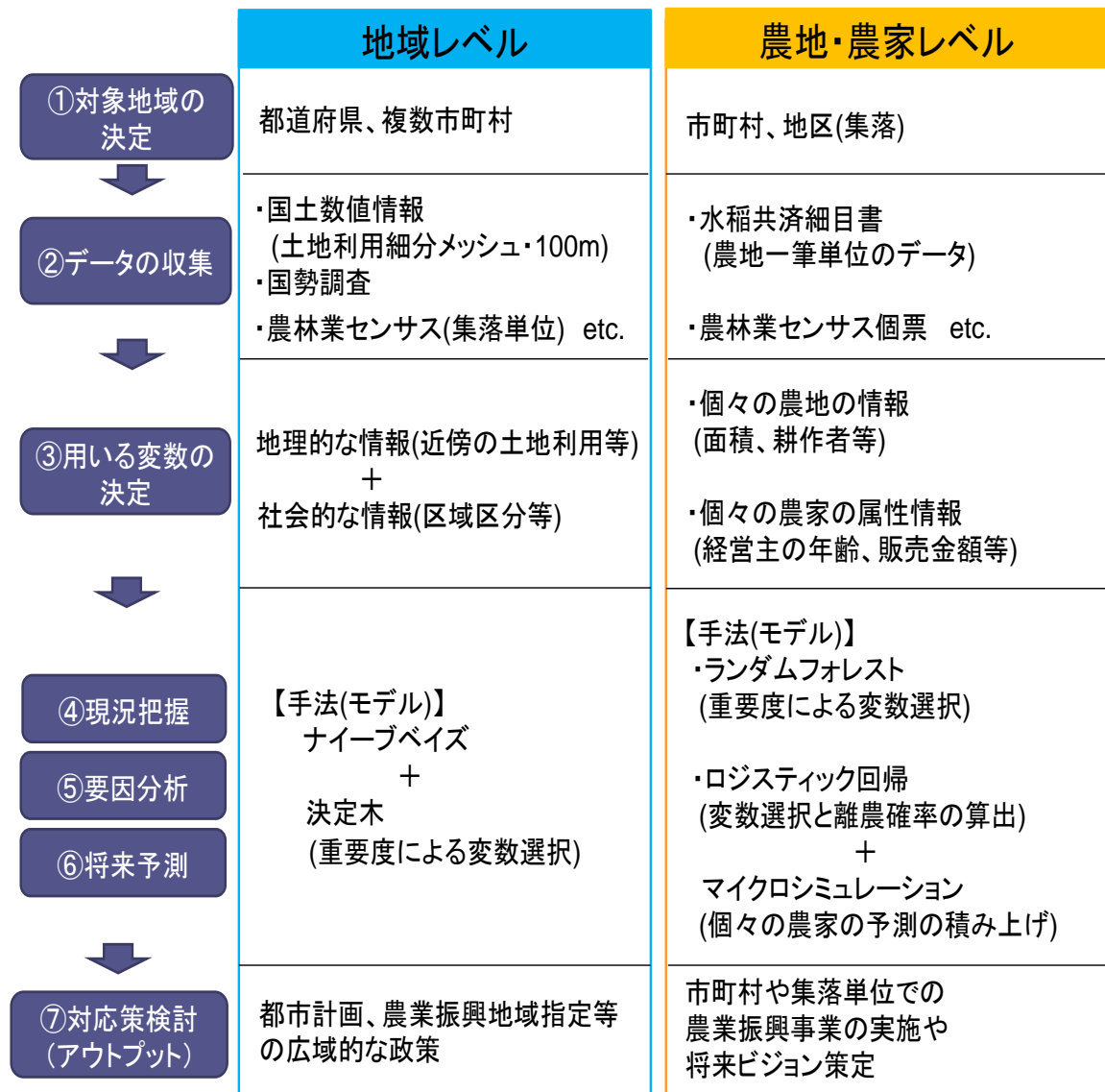


図 5.1 本研究における分析の手法・手順

まず、①対象地域については、地域レベルでは、都道府県あるいは複数市町村、農地・農家レベルにおいては、市町村や地区(集落)となる。地域レベルは、全国や地方といったより広範囲なレベルも考えられるが、各地域で状況が異なるため、詳細な農地保全対策の検討の際に、広い範囲をまとめて分析することはそれほど有効ではないと考えられる。そのため、本研究も複数市町村を対象とするにとどめた。

また、②データについては、地域レベルでは、国の統計データを収集し、土地利用のデータとしては、土地利用細分メッシュ(100mメッシュ)を使用した。

一方で、農地・農家レベルでは、農地一筆単位のデータがある水稻共済細目書や、個々の農家ごとのデータがある農林業センサス個票のデータを収集し、分析に使用した。本研究では、このように非常に粒度の細かいデータを用いたことが大きな特徴であるといえる。

また、③分析やモデル構築に用いた変数としては、地域レベルでは、地理的な情報(近傍の土地利用等)と社会的な情報(区域区分等)の両方を活用した。この点は本研究の大きな特徴である。既存研究においては、セル・オートマトンによるモデルとして、近傍の土地利用の情報のみを用いた研究(山本ら 1999a, 山本ら 1999b)や、逆に、近傍の土地利用は全く考慮せず、主に社会的な情報を用いた研究(池田ら 2007, 三木ら 2004)は見られるが、両者をあわせて用いた研究はほとんど見られない。

しかし、本研究の分析では、近傍の土地利用と区域区分の双方が重要な変数であるという結果が出ており、両者を用いること(地理的情報+社会的情報)が高精度な予測モデル構築に必要であるといえる。本研究でも、両者を用いたことが予測モデルの精度向上の要因のひとつであると考えられる。

また、農地・農家レベルでは、個々の農地の情報ならびに個々の農家の属性情報を変数として用いたことが大きな特徴である。既存研究では、市町村全体における耕地面積や耕作放棄地面積、主業農家率、農業就業人口平均年齢などといった地域全体の情報を使用する研究が多い(稲葉 2006, 吉田ら 2005)。

しかし、より現実的で精緻なモデルを構築するためには、より粒度の細かい農地一筆単位や農家単位の情報を用いることが重要である。

以上をふまえると、本研究で分析に使用した変数に関しては、地域レベルでは、変数として考慮に入れる情報の対象範囲を広げたこと、農地・農家レベルでは、より粒度の細かいデータを用いて、分析対象をさらに綿密に分析したことが大きな特徴であり、その点が高精度なモデルの構築につながっていると考えられる。

さらに④現況把握、⑤要因分析、⑥将来予測では、主に機械学習の手法を用い、地域レベルでは、ナイーブベイズと決定木を組み合わせたモデルを提案した。一方で、農地・農家レベルでは、農地の土地利用状況の変化についてはランダムフォレストを、農家の離農については、ロジスティック回帰とマイクロシミュレー

ションを組み合わせた手法を提案した。

ナイーブベイズについては、データが少数でも、モデルの訓練が可能な点が利点であり、少ないデータで広域的に分析する際には、有効な手法であると考えられる。しかし、ナイーブベイズはアルゴリズムとしては非常に単純なため、大量で詳細な(サンプルサイズや説明変数が多い)データがあり、より高い精度を求め場合には、農地・農家レベルで用いたような複雑なモデルであるランダムフォレストのほうが適していると考えられる。ランダムフォレストは、ブートストラップサンプリングの手法を用いながらアンサンブル学習を行うことにより過学習を防ぎ、高い汎化性能を実現させる。実際に、本研究における自己保全管理農地の発生を予測するランダムフォレストモデルでも非常に高い精度が得られた。

また、農地・農家レベルでは、マイクロシミュレーションも有効な手法であると考えられる。個々の活動主体の行動を確率的にモデル化したうえで、各主体の行動の予測を行い、その結果を積み上げるといった手法は、より現実に近いモデリングであるといえる。本研究においても、農家の離農予測モデルにマイクロシミュレーションの手法を活用し、高い精度を得ることができた。

例えば、既存研究の離農予測モデルでよく用いられているマルコフモデル(安武 2018, 安武・千田 2013, 渡部ら 2012, 細山 2013)では、各農家のデータがあるカテゴリごと(例えば、経営耕地面積の規模別)に再集計するため、最終的には、個々の農家の情報は、要約された形で使用されることになる。

しかし、マイクロシミュレーションでは、個々の農家ごとに予測を行うため、個々の農家の情報をそのまま用いることができるという利点がある。このことから、農林業センサス個票などの粒度の細かいデータが得られる場合には、当手法を活用することが非常に有効であると考えられる。また、最小単位である個々の活動主体の予測をベースとしてそれを積み上げていく方法をとるため、対象地域のスケールを柔軟に変更することができ、例えば、よりマクロなスケールにも適用が可能であるというメリットもある。

しかし一方で、マイクロシミュレーションを用いる際には、個々の活動主体の行動をどのように確率的にモデリングするかという問題がある。

そこで、本研究では確率的なモデリングに機械学習を用いるという手法を提案した。マイクロシミュレーションにおける既存研究では、例えば農業従事者の離脱行動をモデリングする際に、単純に過去の 2 時点間の離脱率をそのまま確率値として用いている(遠藤 2004)。離農予測でも、この方法を用いるとすると、

過去の2時点間の離農率を使用するという単純なモデリングになる。

しかし、機械学習の手法を用いることで、多数の変数の中から離農に関係している変数(例えば、経営主の年齢)を抽出したうえで、それらを用いて離農確率を算出するという、離農の要因を考慮に入れたモデリングが可能となり、より複雑かつ現実的なモデリングを実施することができる。

そして、今回は様々な機械学習の手法がある中で、ロジスティック回帰を用いたが、この理由としては、ロジスティック回帰が、0から1の間の値をとる事後確率を算出可能であり、かつ説明変数の効果を偏回帰係数やオッズ比という形で算出することや、AICなどによる変数選択を実施することが可能であった利点を有するためである。その点で、ロジスティック回帰はマイクロシミュレーションと相性の良い手法であるといえる。

そして、本研究の地域レベルと農地・農家レベルの手法に共通する点として、変数選択を実施しているということがある。地域レベルでは、決定木で算出される変数の重要度を用い、農地・農家レベルでは、ランダムフォレストで算出される変数の重要度や、ロジスティック回帰のモデルにAICを適用して変数選択を行った。

農地の土地利用状況の変化や、農家の営農状況の変化を分析する際には、様々な変数が候補として考え得るが、その中で関連性の高い変数のみを抽出し、それらを用いてモデルを構築することが重要である。変数選択のプロセスは、モデルの予測精度の向上には不可欠であると考えられる。実際、ランダムフォレストによる自己保全管理農地の発生を予測するモデルにおいても、候補とした説明変数について、全ての変数を用いたモデルよりも、変数選択を実施し、少数の変数を用いたモデルのほうが良い精度が得られた。特に本研究のように、非常に多数の説明変数を候補として用いる場合などには、変数選択を実施することが有効であると考えられる。

そして、⑦対応策の検討については、地域レベルでは、都市計画といった広域的な政策が対象となり、農地・農家レベルでは、市町村単位や集落単位での農業振興事業の実施や、将来ビジョンの策定が中心となる。

以上、本研究で用いた分析手法や手順について示した。

また、本研究における各分析の位置づけや関係性を示したものが、図 5.2 である。図は、横軸に分析のスケール、縦軸に分析の対象をとったマトリックスであり、本研究の各章の分析を、該当するセルに記載している。

そして、それぞれの分析やモデルを組み合わせることで、より高度で詳細な地域農業動向の予測が可能になると考えられる。

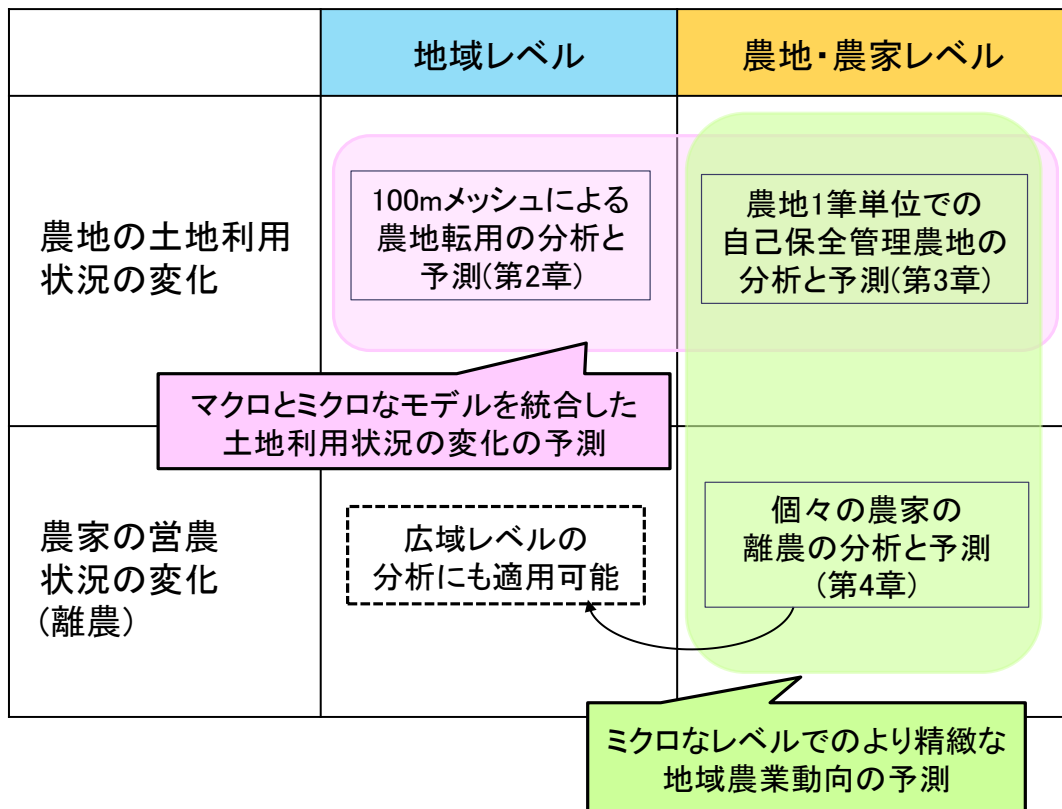


図 5.2 本研究の各分析の位置づけと関係性

まず 1 つ目に、100m メッシュによる農地転用の分析と予測(第 2 章)ならびに農地一筆単位での自己保全管理農地の分析と予測(第 3 章)で提案した手法をあわせることで、マクロとミクロなモデルを統合した農地の土地利用状況の変化の予測が可能となる。

例えばまずは、100m メッシュによるマクロなモデルを用いて、今後の農地転用等の土地利用状況の変化の予測を行い、区域区分の見直しなどを通じて、今後保全すべき農地を検討する。そのうえで今度は、保全すべき農地が存在する地域

を対象に、個々の農地の情報を用いたマイクロなモデルで、農地一筆単位での自己保全管理農地や耕作放棄地の発生予測シミュレーションを実施し、当該地域で実施すべき農地保全対策について検討を行うといったことが可能となる。

また、逆に、マイクロなモデルを用いて、各地域の農地の土地利用状況の変化の分析や予測を行い、今後何らかの農地保全対策を実施した場合に、それらがマクロレベルにおいてどのような影響を与える可能性があるかという点について、分析や予測を行うこともできると考えられる。

そして2つ目としては、農地一筆単位での自己保全管理農地の分析と予測(第3章)と個々の農家の離農の分析と予測(第4章)で提案した手法を組み合わせることによって、マイクロなレベルでのより精緻な地域農業動向の予測が可能となる。

自己保全管理農地の発生の予測に関するモデルにおいては、農家の属性について、保有農地数などの少数の変数をもとに簡略的に類型化した農家タイプと農林業センサスの地区ごとの集計データを用いており、より詳細な農家の属性情報は考慮できていなかった。一方で、農家の離農予測に関するモデルにおいては、各農家の保有農地の地理的な条件(農地一筆ごとの存在する場所、面積、形状等)までは考慮できていなかった。

自己保全管理農地の発生には、農地を保有する農家タイプが大きく影響しているという分析結果が出たことから、両方のモデルを組み合わせることによって、各農地を保有(耕作)している農家の詳細な属性をふまえたうえでの、土地利用状況の変化の予測が可能となり、さらに高精度な予測が実現できると考えられる。また、逆に、農家の離農などの営農状況の変化に関して、各農家の保有農地の地理的な条件をふまえたうえでの予測も同様に可能となる。例えば、経営耕地面積が同じ農家であっても、保有している農地が平地に存在するか山間部に存在するかといった農地の立地条件によって、離農などの傾向は異なってくると考えられる。

さらに、農家の離農予測をふまえて、今後、どの場所のどのような農地が離農に伴い出てくる可能性があるかという点を把握することによって、地域農業の維持に向けて、将来的に担い手がどの場所のどの程度の面積の農地を引き受けて耕作する必要があるかといった点の予測が可能となる。さらに、そのような担い手への農地集積が現実的に可能かどうか、また可能でなければ、どのような対策を打っていくことが求められるかといったことについてもシミュレートする

ことができる。

さらに、個々の農家の離農の分析と予測(第4章)で提案した手法は、マイクロシミュレーションを用いて、個々の農家の予測結果を積み上げる手法であるため、例えば都道府県単位といった市町村以上のより広域なスケールの分析にも適用可能である。

本研究の分析結果においても、市町村単位のモデルで予測した結果を積み上げて、都道府県の離農農家数を予測する手法が、都道府県単位の1本のモデルで予測する手法よりも精度が良かった。このことは、ミクロなモデルによって、ボトムアップ的にマクロなスケールの予測も可能であることを示している。特に、農家の離農のように市町村や農業地域によって、その傾向が異なるような場合に非常に有効であり、本研究で提案した農家の離農予測の手法は、対象地域が広域な場合の分析においても活用できるという利点を有する。

以上のように、本研究で提案した複数の手法を組み合わせることによって、さらに詳細な地域農業動向予測を実施でき、より具体的で実効的な農地保全対策の検討が可能になると考えられる。

また、本研究のもう一つの目的である「2. 各地域において、今後とり得るべき農地保全対策の解明」をふまえ、対象地域における分析や将来予測の結果から、今後の農地保全対策の方向性について述べる。

まず、農地、農家ともに、現在の状況が続いた場合は、さらに減少が続き、地域によってはその減少速度も増していく可能性が高い。

そこで、広域レベルにおいては、区域区分を適切に見直し、保全すべき農地の転用を防ぐことが重要であり、特に、集団的な農地を優先的に保全するように規制を実施していくことが必要である。また農地法で定められている転用許可の基準に関係している駅や医療施設の設置についても、それらが農地の転用に与える影響について事前に十分に把握したうえで、慎重な計画のもとで設置を行うことが求められる。

本研究の分析の結果をふまえると、まとまった農地を優先的に転用規制することで、農地の保全に関して一定の効果が得られると考えられるが、しかしそれだけでは対策として十分でない。特に、よりミクロな各市町村や集落単位での対策とあわせて実行する必要がある。

そして、ミクロなレベルの対策においては、まず、営農意欲の高い農家や、将来の担い手への農地の集積を積極的に進めていくことが重要である。また、自己

保全管理農地の発生には、各農地の面積や存在する場所の条件が大きく影響しているため、そのような各農地の特徴をふまえたうえで、農地集積の計画を立てていく必要がある。特に、面積が小さい農地や中山間地域に存在する農地などで自己保全管理農地が発生しやすく、その後の耕作放棄地化のリスクも高いため、このような農地を将来的にどのように維持保全していくかということを含めたうえで計画を立てておくことが必要である。

また、各農家の離農の抑制に向けて、これからの農業を担う若い経営主や農業従事者を増加させるために、農業経営の継承や、農業経営の維持・拡大に対する政策的支援ならびに新規就農支援をより重点的に実施していく必要がある。また一方で、現在の農業従事者の年齢構成を考えると、農家が一定程度離農することはやむを得ない部分もあるため、前述のように、離農に伴い出てくる多くの農地を担い手に円滑に移行させるための対策を早急の実施し、地域に存在する農地を保全していくことが重要である。

これらのことから、政策としては、農地中間管理機構のさらなる活用や農業競争力強化プログラムにおける人材力の強化の実施などが特に重要になると考えられる。

以上、本研究の分析結果から、上記のような対応策が必要であると考えられるが、実際には、地域ごとに農地の土地利用状況の変化や農家の離農の傾向に異なる点も見られるため、各地域固有の特徴をふまえた分析や予測を行い、地域に応じた対応策を実施していくことが重要である。その際に、本研究で提案したような機械学習を用いた予測手法を、有効に活用できるのではないかと考えられる。

本研究の意義としては、これまで定量的な分析が不十分であった農地の土地利用状況の変化や、農家の営農状況の変化について、上記に示したような機械学習の手法を用いて高精度な予測モデルを構築したことである。この点については、本研究で提案したいずれのモデルにおいても、予測精度を検証した結果、非常に高い精度が得られた。本研究で構築した予測モデルを活用することにより、将来的な地域農業動向を高精度に予測が可能であり、また定量的に数値として、今後どの程度農地や農家が減少するかという点の詳細な把握が可能となる。

また本研究では、より詳細なモデルとして、農地一筆単位での土地利用状況の変化や、農家単位での離農の予測を行うような新たなモデルを提案した。このようなモデルを用いることで、どの場所のどのような農地で自己保全管理農地が発生するか、あるいはどのような属性を持つ農家が離農するのかといったこと

が把握できるようになるため、従来よりもきめ細やかな農地保全対策の検討が可能になると考えられる。

さらに、本研究で提案したモデルは、データさえあれば、他地域においても同様の分析や予測が可能な汎用性を有している。また、機械学習の手法を用いているため、地域固有の特徴をうまく抽出し予測を実施できるという利点もある。実際に、農家の離農予測においても、市町村や農業地域類型区分ごとに離農の傾向に異なる点が見られたが、それぞれの地域の離農の特徴をふまえたモデルを構築し、一定程度の高い精度で予測が可能であった。何度も言うように、農地保全については、全国共通の問題であるため、本手法は各地域の実情をふまえた対応策の検討に役立てることができると考えられる。

以上のように、本研究で提案したモデルを用いることで、適切な農地保全対策の検討が可能となり、今後の地域農業の在り方をより望ましい方向にもっていくことができると考えられる。

今後の課題や展望としては、まず、今回提案した複数の手法を組み合わせ活用する際に、データの紐づけをどのようにするかという問題がある。例えば、農林業センサス個票の農家のデータと水稻共済細目書の農地や耕作者に関するデータは、直接的には紐づけることができない。そのため、その方法について検討していく必要がある。ただし、農林業センサス個票と同様の農家の属性データや、水稻共済細目書と同様の各農地とその所有者のデータなどが地域で一元的に整備・管理されていれば、それらで代用して分析を行うことも可能である。

また、農家の営農状況の変化については、今回は離農のみを対象とした。しかし、規模縮小や規模拡大、現状維持といったような離農以外の営農状況の変化についても予測していくことが必要である。そのことによって、農家の営農状況の変化と農地の土地利用状況の変化に関するそれぞれの予測モデルを統合した際に、より詳細な地域農業動向予測が可能となる。

また、今後の農地保全対策として、上述のように担い手への農地の集積を促進させていくことが重要であるが、集積の促進あるいは停滞に関連している要因については分析が出来ていない。今後は、それらに関する分析も実施し、集積を促進させるために必要な条件や対策について明らかにしていく必要があると考えられる。

さらに、現場においては、圃場管理システムや作業管理システムといった農業経営を支援するソフトウェアは数多く存在するが、本研究で示したような地域

農業の動向について詳細な分析や予測が実施できるツールはほぼ皆無に等しい。そのため、今後は、提案した手法を現場で活用できるようなソフトウェアを開発し、提供していくことが求められる。

以上のような課題はあるが、本研究で提案した一連の手法は、各地域における農地保全の取組の推進に役立てることができるものであると考える。

謝辞

本研究を行うにあたって、指導教官として、多大なるご指導をいただきました守屋和幸教授に深く感謝いたします。また、研究に関していつも気にかけてくださった生物圏情報学講座の先生方に深く感謝をいたします。さらに、アドバイザーとして助言をしてくださった新山陽子教授、畑山満則教授と松原繁夫准教授に深く感謝をいたします。あわせて、社会人として博士課程に通学するにあたり、仕事の面でも色々ご配慮いただきました農業・食品産業技術総合研究機構の中央農業研究センター農業経営研究領域の皆様方に深く感謝をいたします。最後に、研究生活をいつも見守ってくれていた家族・友人にも深く感謝をいたします。

引用文献

- Albert, A. and J. A. Anderson, 1984, On the Existence of Maximum Likelihood Estimates in Logistic Regression Models, *Biometrika*, Vol. 71, No. 1, pp. 1-10.
- Breiman, L., 2001, Random Forests, *Machine Learning*, Vol. 45, No. 1, pp. 5-32.
- Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., 1984, Classification and Regression Trees, Chapman & Hall.
- Calinski, T., and Harabasz, J., 1974, A dendrite method for cluster analysis, *Communications in Statistics*, Vol. 3, pp. 1-27.
- 鄭会勲・淀川智之・矢沢正士, 2004, 耕作放棄地を有する農業集落の空間構造的
特性 —北海道の2町村を事例として—, 農村計画学会誌, Vol. 23, No. 23-suppl,
pp. 259-264.
- 崔龍文, 2001, 東京大都市圏の市街化調整区域における農地転用のプロセスとそ
のメカニズム —伊勢原市下小稲葉地区の事例—, 地学雑誌, Vol. 110, No. 5, pp.
665-688.
- 遠藤和子, 2004, 棚田を対象とする農地利用予測シミュレーション —土地利用
計画策定のために—, 農村計画学会誌, Vol. 23, No. 1, pp. 29-40.
- 樋口修, 2008, 都市農業の現状と課題 —土地利用制度・土地税制との関連を中心
に—, 調査と情報, No. 621, pp. 1-11.
- 細山隆夫, 2013, 北海道中核農業地域における構造変化の地域性と将来展望 —
2010年農業センサス個票組み替え集計分析—, 農業経営研究, Vol. 51, No. 1, pp.
143-148.
- 池田寛之・渡辺公次郎・近藤光男, 2007, ニューラルネットワークとセル・オー
トマトンを用いた土地利用予測, 土木計画学研究・講演集, No. 36, pp. 7-10.
- 稲葉弘道, 2006, 耕地面積と耕作放棄地の変化の要因分析, 千葉大学経済研究,
Vol. 20, No. 4, pp. 725-752.
- 井上裕之, 1989-1990, 離農形態の変化と農地市場 —北海道畑作地帯を対象にし
て—, 農業経営研究, Vol. 27, No. 1, pp. 46-54.
- Katchova, A. L. and M. C. Ahearn, 2017, Farm entry and exit from U.S. agriculture,
Agricultural Finance Review, Vol. 77, No. 1, pp. 50-63.
- 金明哲, 2009, 統計的テキスト解析(15) ～テキストの分類(2)～, ESTRELA, No.
182, pp. 44-49.

- 国土地理院, 2017, 基盤地図情報ダウンロードサービス,
< <https://fgd.gsi.go.jp/download/menu.php>>, 2018年7月31日参照.
- 国土交通省, 2018, 国土数値情報ダウンロードサービス,
<<http://nlftp.mlit.go.jp/ksj/>>, 2018年7月31日参照.
- 栗原伸一・霜浦森平・丸山敦史, 2014, 農業センサス個票を用いた離農要因の地域別分析, *農業情報研究*, Vol. 23, No. 2, pp. 72-81.
- Kurihara, S., S. Shimoura, A. Maruyama and A. E. Luloff, 2013, Factors Associated with Dynamic Change in Japanese Agriculture: Farm Discontinuation and Entry Using the Individual 2005 and 2010 Census of Agriculture, *Asian Journal of Agriculture and Rural Development*, Vol. 3, No. 5, pp. 263-282.
- Landi, C., G. Stefani, B. Rocchi, G. V. Lombardi and S. Giampaolo, 2016, Regional Differentiation and Farm Exit: A Hierarchical Model for Tuscany, *Journal of Agricultural Economics*, Vol. 67, No. 1, pp. 208-230.
- Ligtenberg, A., Arnold, K.B., and Lammeren, R.V., 2001, Multi-actor-based land use modelling: spatial planning using agents, *Landscape and Urban Planning*, Vol. 56, No. 1-2, pp. 21-33.
- 松井孝典・宇賀田徹・町村尚, 2014, 機械学習アルゴリズムによる耕作放棄の要因分析および予測モデルの開発, *土木学会論文集 G(環境)*, Vol. 70, No. 6, pp. II_131-II_139.
- 三木 隆輝・有馬 隆文・出口 敦, 2004, 決定木分析を用いたシナリオ型開発予測に関する研究 —福岡県糸島地域を対象として—, *日本建築学会計画系論文集*, No. 584, pp. 99-105.
- Mishra, A. K., J. M. Fannin and H. Joo, 2014, Off-Farm Work, Intensity of Government Payments, and Farm Exits: Evidence from a National Survey in the United States, *Canadian Journal of Agricultural Economics*, Vol. 62, No.2, pp. 283-306.
- Mitchell, T.M, 1997, *Machine Learning*, McGraw-Hill.
- Möllers, J. and J. Fritsch, 2010, Individual farm exit decisions in Croatian family farms, *Post-Communist Economies*, Vol. 22, No. 1, pp. 119-128.
- 門間敏幸, 2016, わが国の水田農業の構造変動とその対応方向, *東京農業大学農学集報*, Vol. 61, No. 1, pp. 6-16.
- 新妻雅弘・斎藤博昭, 2009, Random Forest を用いた音楽ジャンル分類, *情報処理学会論文誌*, Vol. 50, No. 12, pp. 2910-2922.

野村亘・村上暁信・渡辺達三, 2003, 都市近郊集団農地における耕作放棄地に関する研究, 農村計画学会誌, Vol. 22, No. 22-suppl, pp. 91-96.

農林水産省, 2006, 平成 18 年度食料・農業・農村白書,
<http://www.maff.go.jp/j/wpaper/w_maff/h18_h/trend/1/t1_2_2_03.html>, 2018 年 7 月 31 日参照.

農林水産省, 2008, 食料・農業・農村基本法,
<<http://www.maff.go.jp/j/kanbo/kihyo02/newblaw/newkihon.html>>, 2018 年 7 月 31 日参照.

農林水産省, 2017a, 農業競争力強化プログラム,
<http://www.maff.go.jp/j/kanbo/nougyo_kyousou_ryoku/>, 2018 年 7 月 31 日参照.

農林水産省, 2017b, 農業地域類型区分について,
<http://www.maff.go.jp/j/tokei/chiiki_ruikei/setsumei.html>, 2018 年 7 月 31 日参照.

農林水産省, 2018a, 農林業センサス,
<<http://www.maff.go.jp/j/tokei/census/afc/about/index.html>>, 2018 年 7 月 31 日参照.

農林水産省, 2018b, 面積調査,
<<http://www.maff.go.jp/j/tokei/kouhyou/sakumotu/menseki/>>, 2018 年 7 月 31 日参照.

農林水産省, 2018c, 農業の多面的機能とは?,
<http://www.maff.go.jp/j/nousin/tyusan/siharai_seido/s_about/cyusan/tamen/>, 2018 年 7 月 31 日参照.

農林水産省, 2018d, 農業・農村の有する多面的機能,
<http://www.maff.go.jp/j/nousin/noukan/nougyo_kinou/>, 2018 年 7 月 31 日参照.

農林水産省, 2018e, 平成 29 年度 食料・農業・農村白書,
<http://www.maff.go.jp/j/wpaper/w_maff/h29/index.html>, 2018 年 7 月 31 日参照.

農林水産省, 2018f, 2015 年農林業センサス報告書,
<<http://www.maff.go.jp/j/tokei/census/afc2015/280624.html>>, 2018 年 7 月 31 日参照.

O'Donoghue, C., 2017, Farm-Level Microsimulation Modelling, Palgrave Macmillan.

大阪府, 2001, 大阪府国土利用計画(第三次),
<<http://www.pref.osaka.lg.jp/sokei/kokudokeikakusannji/index.html>>, 2018 年 7 月 31 日参照.

大阪府, 2010, 大阪府国土利用計画(第四次)策定について,
<<http://www.pref.osaka.lg.jp/sokei/kokudokeikakuyoji/index.html>>, 2018 年 7 月 31 日参照.

- 大阪府, 2018, 大阪府の農業データ,
<<http://www.pref.osaka.lg.jp/kannosomu/nourinsuisan/index.html>>, 2018年7月31日参照.
- Silverman, B.W., 1986, Density Estimation for Statistics and Data Analysis, Chapman and Hall.
- 清水千弘, 1997, 農地所有者の土地利用選好に関する統計的検討 —生産緑地法改正による農地転用問題を課題として—, 総合都市研究特別号, Vol. 62, pp. 31-45.
- 総務省, 2015, 農業振興地域の整備に関する法律,
<<http://law.e-gov.go.jp/htmldata/S44/S44HO058.html>>, 2018年7月31日参照.
- 総務省, 2018a, 農地中間管理事業の推進に関する法律,
<http://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=425AC0000000101&openerCode=1>, 2018年7月31日参照.
- 総務省, 2018b, 統計データ,
<<http://www.stat.go.jp/data/index.html>>, 2018年7月31日参照.
- 総務省, 2018c, 都市計画法, <<http://law.e-gov.go.jp/htmldata/S43/S43HO100.html>>, 2018年7月31日参照.
- 総務省, 2018d, 農地法, <http://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=327AC0000000229&openerCode=1#74>, 2018年7月31日参照.
- Sturges, H. A., 1926, The Choice of a Class Interval, *Journal of the American Statistical Association*, Vol. 21, No. 153, pp. 65-66.
- 高田明典, 2007, 群馬県吉井町上奥平における耕作放棄地の拡大とその背景, 地理学評論, Vol. 80, No. 4, pp. 155-157.
- 田代洋一, 2005, 新版農業入門, 大月書店.
- The R Foundation, 2018, The R Project for Statistical Computing,
<<https://www.r-project.org/>>, browsed on Jul.31, 2018.
- Valbuena, D., Verburg, P.H., Bregt, A.K., and Ligtenberg, A., 2010, An agent-based approach to model land-use change at a regional scale, *Landscape Ecology*, Vol. 25, No. 2, pp. 185-199.
- 渡部博明・安武正史・千田雅之, 2012, 中国地方における離農農家の属性と耕地面積の将来予測, 近畿中国四国農研農業経営研究, No. 23, pp. 73-84.

- 八木洋憲, 2005, 土地利用計画論－農業経営学からのアプローチ－, 養賢堂.
- 山本昌弘, 2003, 都市近郊水田地帯における離農構造－利根川下流域・茨城県龍ヶ崎市を事例として－, 村落社会研究, Vol. 9, No. 2, pp. 36-48.
- 山本秀一・奥俊信・谷川寛樹, 1999a, 滋賀県琵琶湖沿岸域における土地利用の推移分析とシミュレーション－セル・オートマトンを用いた土地利用分析とシミュレーション－, 日本建築学会大会学術講演梗概集, Vol. 25, pp. 185-199.
- 山本秀一・奥俊信・谷川寛樹, 1999b, 南大阪地域における土地利用の推移分析とシミュレーション－セル・オートマトンを用いた土地利用分析とシミュレーション－, 日本建築学会大会学術講演梗概集, Vol. 25, pp. 113-114.
- 安武正史, 2018, マルコフモデル利用による予測結果の検証と 2015 年農林業センサスデータの有効性, 関東東海北陸農業経営研究, No. 108, pp. 53-57.
- 安武正史・千田雅之, 2013, 稲作農家の動向解析と水田作経営の将来展望－農林業センサス個票組替集計による－, 農業経営研究, Vol. 51, No. 3, pp. 103-108.
- 吉田晋一・佐藤豊信・駄田井久, 2005, 耕作放棄の要因分析と将来予測－システムダイナミックスを用いて－, 農林業問題研究, Vol. 41, No. 1, pp. 56-59.
- 由良智彦・奥俊信・足立啓・山本秀一・谷川寛樹, 2002, セル・オートマトンとニューラルネットワークを用いた「空地」から「住宅地」への土地利用転換シミュレーション, 日本建築学会大会学術講演梗概集 F-1 都市計画 建築経済・住宅問題, pp. 577-578.
- Zhang, H., 2004, The Optimality of Naive Bayes, American Association for Artificial Intelligence.