

# ロバスト回帰法による土地利用の長期時系列変化の検出： ベトナム北部の棚田開発の持続可能性の研究

磯田 弦・小林 祥子・サンガーンゴイ カザディ・  
神田 竜也・グエン グー フー・金 料哲

## **Identifying long-term land use changes using robust regression method: a study of sustainability of terraced paddy development in Northern Vietnam**

**Yuzuru ISODA, Shoko KOBAYASHI, Kazadi SANGA-NGOIE,  
Tatsuya KANDA, Ngu Huu NGUYEN, Doo-Chul KIM**

**Abstract:** This paper applies robust logistic regression to use the present land use map as training data in identifying the past land used based on satellite images. Ground data of the past is difficult to obtain, but using this method, past land use can be classified with the current land use map provided that less than half of the land use have changed. The paper presents the case of identifying terraced paddy fields in Northern Vietnam since the 1970s.

**Keywords:** 土地利用判別 (land use classification)、衛星画像 (satellite images)、長期時系列変化 (long-term changes)、ロジスティック回帰 (logistic regression)、ロバスト推定 (robust regression estimation)

### 1. はじめに

中国雲南省から北部ベトナムへと続く山岳地帯には、標高 1,000m 以上にわたる部分にまで棚田が開かれていが、棚田は必ずしも伝統的なものとは限らない。我々によるベトナム北部ラオカイ省サバ州ラオチャイにおける聞き取り調査（2009 年夏）によれば、棚田の開発が盛んになったのは 1980 年代以降ということである。この地域のモン族は山岳民

---

磯田 弦 〒874-8577 大分県別府市十文字原 1-1

立命館アジア太平洋大学

Phone: 0977-78-1111

E-mail: isoda-y@apu.ac.jp

族であるにもかかわらず年率約 3% (1989-99 年) の人口増加を経験しており、伝統的な焼畑農業から水田耕作に移行し、農業集約の高度化を図ることにより人口増加に対応したと考えられる。

この研究は、現地で得られたこの証言を確認し、棚田開発の面積と場所を特定するために、リモートセンシングによって過去 40 年間の水田分布の変化を調べる。過去のグラウンド・データが存在しないため、現況土地利用図をトレーニング・データとして過去の水田分布を推定する方法として、ロバスト・ロジスティック回帰法を検討する。

## 2. 過去の土地利用変化の分析方法

衛星画像データの蓄積により、リモートセンシングは土地利用の変化を調べる最も有力な方法となった。しかし、一般に過去のグラウンド・データを得ることが困難であり、この困難を克服すべく様々なアプローチがとられている (Coppin et al., 2004)。二時点（以上）の衛星画像から、土地利用の変化を検出するには、異なる時点の衛星画像の放射輝度が比較可能でなければならない。この点に着目すると、アプローチは次の3つに分類できる。

(1) 絶対的補正 absolute radiometric correction  
日射条件、大気条件、地形条件、センサ特性について補正し、物理的に意味のある単位に完全に直すことができれば、現在のグラウンド・データから過去の土地利用を判別することができる。しかし、絶対的補正には気象や地形のデータが必要であり (Kobayashi and Sanga-Ngoie, 2008)、任意の時期と場所で行うのは困難である。

(2) 相対的補正 relative radiometric correction  
統計学的な方法で、過去の衛星画像を現在のものに相対的に補正すれば、現在のトレーニング・データを用いて過去の土地利用を判別することができる。この方法として、ヒストグラムを一致させる方法や、ピクセル単位で回帰する方法がある。ただし、いずれの方法でも、土地利用が変化したと考えられるピクセルを除去して補正のパラメーターを得なければならない。そもそも土地利用の変化が未知であるため、土地利用の変化がない観測を特定する方法が必要があり、いくつかの方法が考案されている。

(3) 分類後比較 post-classification comparison  
二時点の土地利用判別を独立に行ったのちに、変化を検出すれば、二時点間の放射輝度の不整合の問題を回避できる。しかし、そのためには、過去の空中写真や土地利用図など、過去のトレーニング・データが必要である。また、分類後比較は各時点の土地

利用判別の精度が低い場合には、誤った土地利用変化を検出する可能性が高くなるといった問題もある (Coppin et al., 2004)。しかし、二時点の衛星画像を完全に比較可能にするのは難しく、また、土地利用の変化がなくても年々の変化があるため、分類後比較のほうが土地利用変化を適切に検出できるとの報告もある (Mas, 1999)。

この研究では、現在の土地利用図をトレーニング・データとして、過去の土地利用を判別し、分類後比較により土地利用変化を特定する。当然、過去の土地利用は現在の土地利用図とは異なるため、後述するロバスト・ロジスティック回帰を用いて、水田を特定する。

## 3. 研究方法

統計学におけるロバスト推定 robust estimation とは、サンプルに含まれる外れ値を、除去あるいはダウンウェイトすることで、サンプルに依存しない統計量を推定する方法のことである。リモートセンシングの分野でも、例えば maximum likelihood classifier で用いる平均と分散にロバスト推定値が用いられる (Mather, 2004)。また、ロバスト回帰 robust regression は、回帰残差の大きいサンプルを除去またはダウンウェイトすることで、サンプルに依存しない回帰式を推定するもので、外れ値の影響を除去しつつ効率を高く（推定誤差のを小さく）保つ様々な方法が研究されている (Andersen, 2008)。例えば、二時点の衛星画像の相対的補正方法として、ロバスト回帰法を用いて、土地利用の変化によって生じると考えられる外れ値をダウンウェイトして補正を行うといった利用が考えられる (Furby and Campbell, 2001)。

本研究では、ロバスト・ロジスティック回帰法を用いて、現在の土地利用図をトレーニング・データ（従属変数）として過去の衛星画像（独立変数）に

回帰することで、過去の水田分布を特定する。ロジスティック回帰とは、独立変数にもとづき、観測があるクラスに属するか否かの確率を求め、その確率が通常 0.5 以上の場合にそのクラスに属すると判別する方法である。ロバスト法は一般に外れ値の影響を除外するために使われるが、この研究の文脈では、主として土地利用変化の結果発生するトレーニング・データ中の誤分類の影響を除外するために用いる。ロバスト・ロジスティック回帰についても、高度な方法が開発されているが(Gervini, 2004), ここでは次の簡便なアルゴリズムを使用する。

- (1) 通常のロジスティック回帰を行う
- (2) すべてのサンプルについて残差を計算する
- (3) 残差のヒストグラムより、正または負の高い残差の部分で密度が高くなっているれば、外れ値（誤分類された従属変数）を判断する閾値（上限・下限）を決める。
- (4) 閾値内のサンプルのみで再び回帰を行う。
- (5) 閾値（上限・下限）が収束するまで、(2)–(4) を繰り返し、最終的な回帰結果を採用する。

ここで、繰り返し計算を行うのは、初期段階の回帰式には外れ値の影響が含まれており、外れ値の判定も真的（ロバストな）回帰式からの残差にもとづいて行わなくてはならないからである。ステップ(4)で暫定的に外れ値と考えられるサンプルを除外するが、永久に分析から除外するわけではなく、次のステップの回帰式の推定に用いないだけである。

以下の分析では、SPSS を用いてロジスティック回帰を行った。アルゴリズムのステップ(3)はヒストグラムを目視で、密度がもっとも低くなっている値を閾値の上限と下限としたが、本来的には残差の高い部分で密度が高くなっているか否かを判定し、閾値を求めるアルゴリズムを挿入すべきであろう。外れ値が存在しないと判定されるとときは、ロバスト推定ではなく、通常の回帰を行ったほうが、効率の

よい推定が行える。

次節では、ロバスト・ロジスティック回帰の有用性を検討するために、まず現況土地利用図から現在の水田分布を推定し、高解像度衛星画像をテスト・データとして精度検証を行う。

### 3. 現況土地利用図にもとづく現在の土地利用判別

土地利用図は一般に総括されているため、観測数は多いが、誤分類を多く含むソースである。そこで、土地利用図を従属変数、衛星画像を独立変数としたロバスト・ロジスティック回帰を行うことで、誤分類の影響を除去して水田を特定する。対象地域はベトナム北部ラオカイ省サパ州ラオチャイの面積約 29 km<sup>2</sup> で標高 1000–1900m にわたる地域である。水田は標高 1000–1300m の傾斜地に棚田として分布している。トレーニング・データとして使用する土地利用図は、2004 年 10 月の現況を示したもので、縮尺 1:10000 相当の GIS データとして提供された。衛星画像は 2007 年 9 月 20 日撮影の Landsat5 の TM センサ 6 バンドで、解像度は 30m である。これは、USGS の Global Land Survey (GLS) コレクションの一部であり、オルソ化され、投影された状態で提供された。精度検証ためのテスト・データには、2009 年 2 月 13 日撮影の ALOS の AVNIR-2 センサ 4 バンド（解像度 10m）を PRISM パンクロマティック画像（解像度 2.5m）でパンシャーピングしたカラー画像を用いる。このように参照する年や季節が異なるが、これらがアーカイブされたデータ中でもっとも時期の近いセットであった。

ロジスティック回帰には、従属変数として土地利用図の水田を 1、それ以外を 0 とコーディングした変数を、独立変数として、地形の影響を抑制するために Band 1, 2, 4, 5, 7 を Band3（赤色相当）で除した 5 つの変数を用いた。観測は対象地域内の 32010 ピクセル（30m × 30m）である。

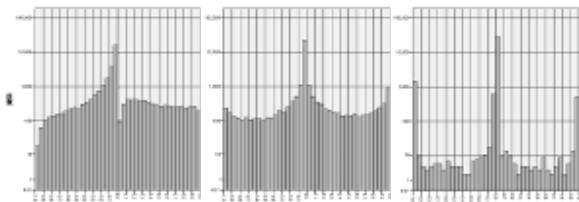


図-1 回帰残差の分布：初回（左）、2回目（中）、最終回（右）。度数は対数目盛で示されている。

回帰結果は十分に有意で、Band7(を Band3 で除したもの) 以外はすべて 1%で有意であったが、紙面の都合上、パラメター推定値の掲載は割愛し、ロジスティック回帰の繰り返しごとの回帰残差のヒストグラムと分類表のみを示す。

ロジスティック回帰の残差(図-1)は特殊であり、-1～+1 の値をとるが、-1～-0.5 が false negative, -0.5～0 が true negative, 0～0.5 が true positive, 0.5～1 が false positive の領域であり、各領域内の度数を見るだけで、回帰のパフォーマンスがわかる。例えば、正しい予測がなされるには大部分の観測が-0.5～0.5 に入っていなくてはならないが、図-1 の初回（全サンプル使用の回帰結果）の残差では、この外側の度数も高く、予測の精度が低いことがわかる。ここでは、独立変数が従属変数を十分に予測できないと考えるのではなく、従属変数に誤分類があると考える。初回の残差には、0.85 以上の部分に小さな密度の高まりがあり、これを除いて回帰し残差を求めるとき、2回目の残差のヒストグラムでは、両端にピークが出現する。これを十数回繰り返すと、最終回（外れ値を除いた回帰結果）の残差が得られ、ここでは大部分の観測の残差が 0 付近にあり、両端に明瞭に表れる高い残差は、従属変数の誤分類によるものであると判断できる。

分類表（表-1）をみると、初回の通常のロジスティック回帰の正分類（全体）は 87.7% にとどまるが、最終回のロバスト・ロジスティック回帰の結果の正分類は、ほぼ 100% になる。もっとも、このア

表-1 分類表：初回と最終回、2007 年

|     | 観測値    | 予測値           |      | 正分類 % |
|-----|--------|---------------|------|-------|
|     |        | 非水田           | 水田   |       |
| 初回  | 非水田    | 24744         | 1410 | 94.6  |
|     | 水田     | 2517          | 3339 | 57.0  |
|     | 全体の正分類 |               |      | 87.7  |
|     | 観測数    | 32010 (100%)  |      |       |
| 最終回 | 非水田    | 24661         | 29   | 99.9  |
|     | 水田     | 30            | 3543 | 99.2  |
|     | 全体の正分類 |               |      | 99.8  |
|     | 観測数    | 28263 (88.3%) |      |       |

ルゴリズムは規則性に従わないサンプルを除くため、予測精度があがるのは当然であるが、使用された観測数をみると、最終回でも 28263 ピクセル (88.3%) が使用されていることがわかる。

予測値の精度検証のためのテスト・データは ALOS のパンシャーピング化したカラー画像を目視で土地利用を判別して作成した。ここでは特に、土地利用図と回帰予測値が一致しない部分で、どちらが正しいかに关心があるため、土地利用図と回帰予測値のクラス（水田・非水田）の組み合わせごとに、約 100 サンプルが得られるように、空間的にランダムに層別抽出を行った。目視による土地利用判別では、Landsat のピクセルと一致する 30m×30m の正方形を用意し、その中の土地利用（水田・非水田）の過半数のものとした。

精度検証結果をまとめた表-2 中の、土地利用図と回帰予測値が一致しない二つの組み合わせでは、すべての場合で正答率が低くなっているが、これは主として土地利用の境界ゾーンに土地利用図と回帰予測値が一致しない地点が多いからである。しかし、いずれの場合でも回帰予測値の正答率が土地利用図の正答率を 10% 程度高くなっている。全体の正答

表-2 精度検証結果

| 土地<br>利用図 | 回帰<br>予測値 | テストデータ       |       |     |     | 利用図<br>正答率 | 回帰<br>予測値<br>正答率 |       |
|-----------|-----------|--------------|-------|-----|-----|------------|------------------|-------|
|           |           | 総数           | (シェア) | 非水田 | 水田  | 合計         |                  |       |
| 非水田       | 非水田       | 24691 (77%)  |       | 111 | 8   | 119        | 93%              | 93%   |
| 非水田       | 水田        | 1463 (5%)    |       | 45  | 52  | 97         | 46%              | 54%   |
| 水田        | 非水田       | 2283 (7%)    |       | 54  | 42  | 96         | 44%              | 56%   |
| 水田        | 水田        | 3573 (11%)   |       | 28  | 91  | 119        | 76%              | 76%   |
| 総計        |           | 32010 (100%) |       | 238 | 193 | 431        | 85.7%            | 86.9% |

率については、回帰予測値が土地利用図よりも 1% 程度高くなっているにとどまるが、ここでの要点は、ロバスト・ロジスティック回帰によって、トレーニング・データよりも高い精度の土地利用判別ができるということである。なお、表-2 の結果を抽出率で加重して集計すると、回帰予測値の正答率は、水田について 70%、非水田については 90%（全体では 87%）となった。

#### 4. 1970 年代以降の土地利用変化

ロバスト・ロジスティック回帰で、トレーニング・データに多少の誤りがあっても、土地利用が判別できるのであれば、現在の土地利用図をトレーニング・データとして用いて過去の衛星画像から、過去の土地利用を判別することができる。そこで、前節で紹介した方法で、過去の水田を判別し、過去 40 年間の水田分布の変化を調べる。使用する衛星画像は、前述 GLS コレクションからの 1993 年 2 月 1 日撮影 Landsat 5 の TM センサ 6 バンド（解像度 30m）と、1973 年 11 月 26 日撮影 Landsat 1 MSS センサ 4 バンド（解像度 60m）、である。トレーニング・データには前節同様の土地利用図を使用する。

ロジスティック回帰の独立変数には前節同様に赤色に相当するバンドで他のバンドを除した変数用いたため、1973 年については 3 変数、1993 年については 5 変数となった。ロジスティック回帰の結果は十分に有意であったため、ここでも分類表のみを提示する（表-3）。

表-3 および表-1 より、通常のロジスティック回帰（「初回」の結果）では、トレーニング・データの参考年（2004 年）から離れるにしたがって正分類率が下がるが、ロバスト・ロジスティック回帰（「最終回」の結果）では、土地利用変化によって生じるトレーニング・データ中の誤分類を除去することにより、1973 年においてもほぼ 100% の正分類

表-3 分類表：初回と最終回、1973 年、1993 年

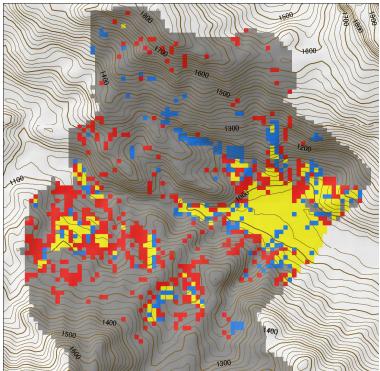
|     | 観測値      | 予測値 1973     |     |       | 予測値 1993      |      |       |
|-----|----------|--------------|-----|-------|---------------|------|-------|
|     |          | 非水田          | 水田  | 正分類 % | 非水田           | 水田   | 正分類 % |
| 初回  | 非水田      | 6887         | 363 | 95    | 24874         | 1280 | 95.1  |
|     | 水田       | 938          | 698 | 42.7  | 2839          | 3017 | 51.5  |
|     | 全体の正分類 % |              |     | 85.4  |               |      | 87.1  |
| 最終回 | 観測数      | 8886 (100%)  |     |       | 32010 (100%)  |      |       |
|     | 非水田      | 7100         | 3   | 100   | 25286         | 20   | 99.9  |
|     | 水田       | 1            | 513 | 99.8  | 16            | 2671 | 99.4  |
|     | 全体の正分類 % |              |     | 99.9  |               |      | 99.9  |
|     | 観測数      | 7617 (85.7%) |     |       | 27993 (87.5%) |      |       |

が達成できる。外れ値として除外される観測数も、時期が古くなるにつれ大きくなるが、それでも 8 割以上の観測が使われており、前述の高い正分類率は規則性に従う極少数のサンプルから得られた結果ではないことがわかる。

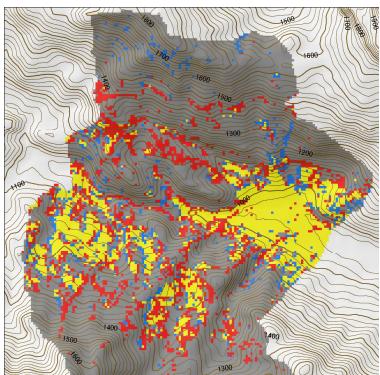
本節および前節のロバスト・ロジスティック回帰の予測値から、各時点の水田分布をそれぞれ独立に求めると、1973 年には 239ha、1993 年には 319ha、2007 年には 453ha となる。現在の水田の約半分が、1970 年代以降に開発されたものである。また、変化率を計算すると 1973–93 年は 33% 増（年率 1.5%）1993–2007 年は 42%（年率 2.6%）となっており、1990 年代以降に開発が加速し、モン族の人口増加と比較できる値になっていることがわかる。三時点の推定結果を分類後比較（図-2）すると、新たな水田は、より標高の高い、傾斜の強い斜面へと外延的に広がっていることがわかる。これらは、現地で得られた証言とおおむね一致する。

#### 5. おわりに

本研究では、サンプル数は多いが精度の低いトレーニング・データを使って土地利用判別する方法として、ロバスト・ロジスティック回帰を検討した。その事例の 1 つは、総描されている現在の土地利用図からの現在の土地利用の判別であり、精度検証では 87% の正答率が得られ、これは若干であるが、トレーニング・データの精度よりも高かった。また、現在の土地利用図をトレーニング・データとして過



(a) 1973-1993 年



(b) 1993-2007 年

図-2 水田分布の変化：同期間に開発された水田（赤）と廃止された水田（青）

去の土地利用を判別した。過去のグラウンド・データが存在しないために定量的な精度検証は不可能であったが、現地で得られた証言とおおむね一致する結果が得られた。なお、我々は本研究の推定結果にもとづき、現地で定性的な情報をを集め、この方法の有効性を確認する予定である。

このように、ロバスト推定は、過去の土地利用変化を知る有力な方法である可能性がある。本研究では水田を特定さえすればよかつたため、ロジスティック回帰を用いたが、多クラスの判別をする場合には、多項ロジスティック回帰への拡張が可能であるに違いないし、maximum likelihood classifier でも高度なロバスト推定を行うものであれば、現在のトレーニング・データで過去の土地利用を推定することが可能であろう。

## 謝辞

この研究の初期の段階でご教示いただいた立命館大学准教授中谷友樹氏に感謝申し上げます。なお、この研究には立命館アジア太平洋大学 2010 年度学術研究助成「東南アジアの棚田開発の持続可能性：ベトナム北部サバの事例」研究代表者：磯田弦、を使用した。

## 参考文献

- Andersen, R., 2008. *Modern Methods for Robust Regression*. Los Angeles: Sage Publications.
- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B. and Lambin, E., 2004, Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, **25**, 1565–1596.
- Furby, S. L., Campbell, N. A., 2001. Calibrating images from different dates to ‘like-value’ digital counts. *Remote Sensing of Environment*, **77**, 186–196.
- Gervini, D., 2004. Robust adaptive estimators for binary regression models. *Journal of Statistical Planning and Inference*, **131**, 297–311.
- Kobayashi, S. and Sanga-Ngoie, K., 2008. The integrated radiometric correction of optical remote sensing imageries. *International Journal of Remote Sensing*, **29**, 5957–5985.
- Mas, J. F., 1999. Monitoring land-cover changes: a comparison of change detection techniques. *International Journal of Remote Sensing*, **20**, 139–152.
- Mather, P. M., 2004, *Computer Processing of Remotely-Sensed Images*, third edition, Chichester: John Wiley & Sons.