

全国版の土砂流出発生確率マップの構築に向けた研究

(国研)防災科学技術研究所 秋田寛己[○], 遊佐暁, 横山仁, 佐野浩彬, 川崎正貴, 上田啓瑚, 臼田裕一郎

1. 背景と目的

近年, 2022 年 8 月新潟・山形県北部豪雨のように行政界を越えた広域スケールの土砂災害が我々の人命や財産に甚大な被害を及ぼしている。これら土砂災害につながる斜面崩壊・土石流・地すべりに伴う土砂流出の発生確率は素因である地形・地質とのリスク分析がなされ, 全国版マップが試行されている¹⁾。このような土砂流出に関連した発生確率マップ(以下, 確率マップと略述)は平時の地域防災対策の一つとして有効である。一方では気候変動による極端降雨の発生が問題視される中, 素因と降雨などの誘因双方の将来変化を確率マップ上に表現できる手法の開発が求められる。

これら将来変化を確率マップに考慮するにはまず, 素因・誘因を網羅した因子全ての重要度を相対的に明らかにする必要がある。そこで本研究は過去 70 年程度の全国の土砂災害発生実績オープンデータ²⁾を活用し, 将来変化する気候変動に関わる降雨や気温などの因子が土砂流出発生確率にどの程度影響を及ぼすのかを機械学習モデルの構築・予測により検討した。

2. 方法

解析領域は一部島嶼を除く日本全国とし, 確率マップの空間分解能は4分の1地域メッシュ(250m)とした。機械学習に用いる教師データは図-1に示すように都道府県別の1953~2020年までの土砂災害発生実績オープンデータ²⁾に加え, 著者らが2020~2024年までの土砂災害を対象に光学衛星データから広域抽出した発生箇所を合算処理し作成した。メッシュ配分した教師データは7,450個であり, 解析対象を降雨による発生実績のみとし大規模崩壊は対象外とした。

確率計算には土砂災害が稀な現象と仮定し, (1)式のポアソン分布を適用し土砂流出発生確率(P_s)を計算して目的変数に設定した。土砂災害発生に関わる整理期間と件数は都道府県単位で集計されているため, P_s には地域差が生じる。

$$P_s(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

ここに, λ : 稀な事象の単位時間あたりの生起確率, k : 単位時間あたりにその事象が何回起こるかをあらわす変数, e : 自然対数の底

メッシュ内での n 回以上発生する P_s を都道府県別に計算し, 再起期間 10 年の発生確率(以下, P_{s10})を採用した。なお, 過去に1メッシュ内で複数の土砂流出があった場所は極めて稀なケースであるが, そのメッシュの P_{s10} が高くなるように設定した。

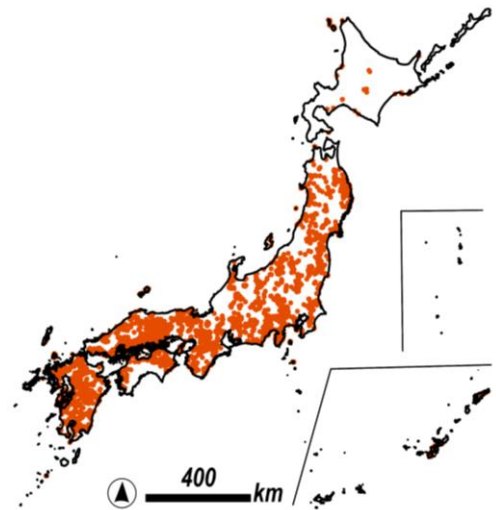


図-1 全国の土砂災害発生実績のあるポイント

表-1 説明変数の一覧

No.	説明変数	空間分解能(m)	データ区分 ¹⁾	内容	出典
地形ファクター					
1	標高	10	V	セルの標高値(m)	本研究
2	傾斜	10	V	セルの平均傾斜角(°)	本研究
3	集水面積	10	V	セルに流れ込む面積の合計値(m ²)	本研究
4	SPI	10	V	ストリームパワー指数。土壌侵食の強さをあらわす指標。 $\ln(A \times \tan \beta)$ で計算。A:集水面積, β はラジアン単位の傾斜	本研究
5	TWI	10	V	地形的湿潤指数。地下水の溜まりやすさをあらわす指標。 $\ln(\tan \beta)$ で計算	本研究
6	断面曲率	10	V	斜面上で、最大傾斜角の方向をあらわす。侵食や沈殿につながる流れの加速と減速に影響を与える	本研究
7	平面曲率	10	V	最大傾斜角の方向と垂直な方向をあらわす。流れの収束と分岐に影響を与える	本研究
8	地すべり地形の面積	1:25,000	V	空中写真や地形図から過去の地すべりの痕跡を判読した結果。メッシュ内の面積の合計値を計算	防災科学技術研究所
9	土砂災害警戒区域の面積	1:25,000	V	降雨の土砂災害発生による影響区域を地形から判読した結果。メッシュ内の面積の合計値を計算	国土数値情報(国土交通省)
10	深層崩壊発生危険確率	1:25,000	V	地すべりや大規模崩壊について地質と地形から危険度を評価した結果。メッシュ内に占める面積比率を計算	国総研資料(2022)
地質ファクター					
11	岩相	1:200,000	C(26)	20万分の1縮尺で作成された全国の岩石・岩相。本研究では類似したグループに分類処理	産総研地質調査総合センター
12	地質年代	1:200,000	C(20)	20万分の1縮尺で作成された全国の地質年代。本研究では類似したグループに分類処理	産総研地質調査総合センター
土地利用・植生ファクター					
13	土地利用分類(2020年)	100	C(10)	Random Forestを適用した2020年のLCM土地利用予測結果	東京大学
14	植生分類	1:25,000	C(25)	第6-7回植生調査結果(1/2.5万縮尺, 2024.6版)と土地利用凡例が連結されたデータを使用	生物多様性センター(環境省), 国立環境研究所
15	最大樹木被覆率	500	V	衛星画像(2008年のTerra/MODIS)から計算されたversion 2の結果を使用	国土地理院, 千葉大学
気候ファクター					
16	年降水量(正規化)	1,000	V	降水量の年合計値(0.1mm単位)。全国平均値を求め, メッシュごとに平均値で除して正規化	国土数値情報(国土交通省)
17	年平均気温(正規化)	1,000	V	年平均気温(0.1℃単位)。全国平均値を求め, メッシュごとに平均値で除して正規化	国土数値情報(国土交通省)
18	Langの雨量因子(正規化)	1,000	V	年降水量/年平均気温で求める。気候の乾燥の程度を表す指数。全国平均値を求め, メッシュごとに平均値で除して正規化	本研究
19	年最深積雪(正規化)	1,000	V	積雪の年最大値(cm単位)。全国平均値を求め, メッシュごとに平均値で除して正規化	国土数値情報(国土交通省)
20	年合計日照時間(正規化)	1,000	V	日照時間の年合計値(0.1時間単位)。全国平均値を求め, メッシュごとに平均値で除して正規化	国土数値情報(国土交通省)
21	年平均全日射量(正規化)	1,000	V	全日射量の年合計値(0.1MJ m ⁻²)。全国平均値を求め, メッシュごとに平均値で除して正規化	国土数値情報(国土交通省)
22	再起期間10年 1時間積算雨量の確率値	1,000	V	3次メッシュにおける再起期間10年の1時間積算雨量のメッシュ別確率値	国総研資料(2022)
23	再起期間10年 24時間積算雨量の確率値	1,000	V	3次メッシュにおける再起期間10年の24時間積算雨量のメッシュ別確率値	国総研資料(2022)
24	再起期間10年 48時間積算雨量の確率値	1,000	V	3次メッシュにおける再起期間10年の48時間積算雨量のメッシュ別確率値	国総研資料(2022)

1) 機械学習モデルのカテゴリ区分。Vは連続変数, Cはカテゴリであり括弧内はカテゴリ数。

説明変数は表-1のように地形、地質、土地利用・植生、気候に大別した因子を網羅的に設定した上で、多重共線性を考慮し機械学習モデルに導入する。各因子は空間分解能が10～1,000 mと幅があるが、全て250 mメッシュサイズに統一しサンプリングした。

以上の目的変数と説明変数を使用し、機械学習モデルを構築する。機械学習にはR言語(ver.4.4.2)の各種パッケージを使用し、ランダムフォレスト(RF)・サポートベクタマシン(SVM)・ k 近傍法(knn)・ニューラルネットワーク(nnet)・勾配ブースティング(xgboost)の中からRMSEや R^2 値の計算結果に基づきモデルを選定する。モデルに変数を導入するにあたり、通常の連続変数は正規化を行い、カテゴリ変数(土地利用など)は0もしくは1を割り当てるダミー変数化の処理を行った。

モデル構築のための教師データは学習用70%と検証用30%の割合とし、10分割クロスバリデーションとOne SEルールを適用しハイパーパラメータをチューニングした。最後に構築したモデルを使用し Ps_{10} の予測値を計算し、日本全国の発生確率マップを作成した。

3. 結果と考察

3. 1 機械学習モデルの性能比較

表-2に各モデルのRMSEと R^2 値、MAEを示し、図-2にRFモデルの Ps_{10} の実績値と予測値の関係を示す。比較したモデルの中ではRF、xgboost、nnet、knn、SVMの順序でRMSEが低いため、RFモデルを採用する。一方でRFを含め、比較したどのモデルでも Ps_{10} の0.5を超過する範囲の予測値が実績値よりも全体的に低いことがわかる。これは1メッシュ内で過去に複数の土砂流出があった場所をモデルが予測しきれないことが原因と考えられる。いずれの機械学習モデルでも0.5以下を示す Ps_{10} の予測に適する一方で、前述の極めて稀な発生確率の予測が困難であることが明らかになった。

3. 2 説明変数の重要度の分析

次に各説明変数の相対的な重要度(Importance)をStrobl and Zeileis (2008)の手法³⁾で計算した。図-3にRFモデルで採用された27の説明変数を示す。地形、地質、土地利用・植生、気候に大別して重要度の積算値(Σ Importance)を計算すると、気候(33.7)が目的変数に強く影響しており、次に地形(21.3)、地質(3.2)、土地利用・植生(1.9)の順序であった。個別には年降水量・24時間積算雨量確率値・年平均気温などが上位であり、降雨の多い地域で発生確率が高い。本研究のように全国広域に計算・比較した場合、地形・地質といった素因よりも降雨などの誘因の方が土砂流出の発生確率に影響を与えている可能性が示唆された。

表-2 各モデルの計算結果

Method	RMSE	R^2	MAE	hyperparameter
RF	0.079	0.524	0.055	mtry = 27, mse = 500, rsq = 500
SVM	0.102	0.201	0.067	sigma = 0.01035188, and C = 4
knn	0.093	0.312	0.066	k = 10
nnet	0.091	0.334	0.067	size = 8 and decay = 0.1
xgboost	0.084	0.430	0.058	nrounds = 50, lambda = 0, alpha = 0.1 and eta = 0.3

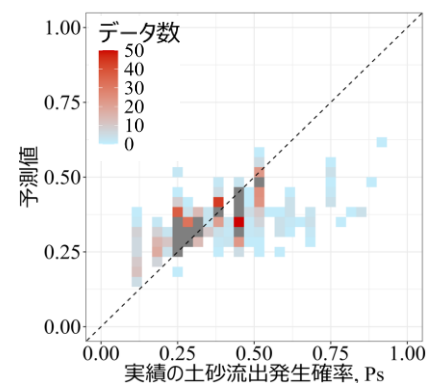


図-2 RFモデルの Ps_{10} の実績値と予測値

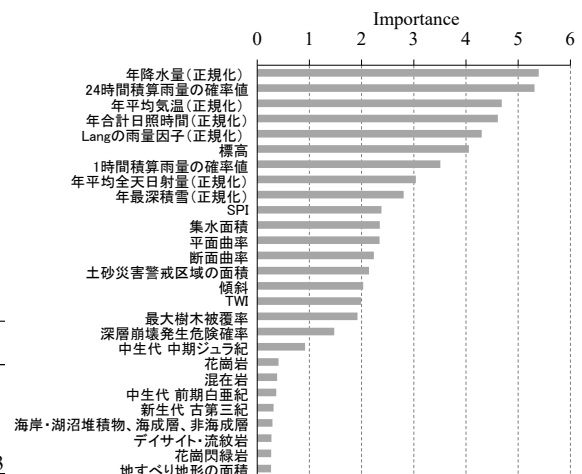


図-3 RFの説明変数における重要度

謝辞: 本研究は、環境省・(独)環境再生保全機構の環境研究総合推進費(S21,JPMEERF23S12117)のプロジェクトにより実施した。このプロジェクトの一環で土地利用分類データは東京大学の橋本禪教授・Huang Wanhui 特任研究員からご提供頂き、植生分類データは国立環境研究所の石濱史子主幹研究員・丹野夕輝専門員からご提供頂いた。各積算雨量確率値は国土技術政策総合研究所が公開する「土砂災害分野におけるL積率法を用いた解析雨量プロダクトの確率化手法」⁴⁾のGISデータを使用させて頂いた。末尾ではあるが、ここに付記し謝意を表する。

参考文献: 1) 国土交通省国土技術政策総合研究所(2020): 地形・地質に関する主題図を用いた全国における土砂災害発生リスク推定法に関する考察, 国土技術政策総合研究所資料, No. 1120, 97pp. 2) 国土交通省国土技術政策総合研究所(2022): 降雨による土砂災害に関する全国集計データ, 国土技術政策総合研究所資料, No. 1231, 199pp. 3) Strobl, C., Zeileis, A. (2008): Why and how to use random forest variable importance measures (and how you shouldn't), Use! R 2008, <https://www.scribd.com/document/334476264/rf> 4) 国土交通省国土技術政策総合研究所(2022): 土砂災害分野におけるL積率法を用いた解析雨量プロダクトの確率化手法, 国土技術政策総合研究所資料, No. 1222, 35pp.