

自然災害リスク評価のためのロジスティック回帰モデルと変数選択

Logistic Regression Model and Model Selection for Estimating the Risk of Natural Disaster

加茂 憲一・柳原 宏和・嘉戸 昭夫・吉本 敦

Kamo, K., Yanagihara, H., Kato, A. & Yoshimoto, A.

キーワード: ロジスティック回帰, 変数選択, 情報量規準, 冠雪害

要約: ロジスティック回帰モデルを用いて, 2004 年小矢部市のスギ林における冠雪害に対する発生リスクの解析を行った. 回帰モデルの構築には冠雪害に影響を与える要因を選択する必要があるが, ここでは 情報量規準による変数選択法を用いて要因を決定した. また, 規準量についても, AIC, BIC, CAIC の 3 つを用いた. 更に, ここで得られた結果と, 先行研究である回帰木による予測結果との比較を行い, ロジスティック回帰モデルの有効性を実証した. その結果, 冠雪害リスク判別能力は, 50% を基準とした場合は大差ないが, ロジスティック回帰モデルが被害リスクを連続的に表現することができる点を考慮すれば, ここでのモデルの方が有効であることが分かった.

Abstract: We apply a logistic regression model to the data of snow damage occurred in Oyabe city in 2004 in order to analyse the damage risk on sugi (*Cryptomeria japonica*) stand on the continuous basis. For specifying factors effecting the snow damage, we rely on the following three information criteria, 1) Akaike's information criterion, 2) Bayesian information criterion and 3) Bias-corrected Akaike's information criterion for selecting appropriate ones. Our results are compared with those from the previous work with the use of the discrete regression tree model. Our experiments show that the ability to judge the degree of the snow damage risk is almost the same for

both approaches according to a critical value by 50%. This implies that considering the amount of information on the risk as well as automatic specification of appropriate variables for the model, we can conclude that the proposed method is superior to the previous one.

Keywords: Information criterion, logistic model, model selection, regression, snow damage

1. はじめに

ロジスティック回帰モデルは、ある事象の生起の有無を表す 2 値変数を応答変数とするものであり、その生起確率を説明変数により予測する場合に適用される。例えば、疾病の発症や、被害の発生といった事象を応答変数とすれば、これら事象の発生確率が様々な要因により変化する場合、考慮される要因を説明変数としてロジスティック回帰モデルが構築され、発生確率の予測が可能になる。本研究では、このロジスティック回帰モデルを用いることにより、森林における自然災害の発生について分析を行うことを主な目的としている。今回は、分析対象とする災害を、富山県のスギ林における冠雪害とした。ここで冠雪害とは、葉や枝などの樹冠に付着した雪の荷重によって幹折や根返りする災害のことを示す。本稿では、この冠雪害の発生を応答変数とし、その被害率が森林の特性や地形因子などを説明変数とする回帰モデルを考える。また、回帰モデルを用いる場合、観測値の平均変動は必要最小限の説明変数により説明されることが理想的であるため、今回、情報量規準を用いた変数選択問題も取り扱う。なお、今回のケースにおいては、変数選択問題は災害リスク要因の特定ということも同時に意味する。次章では、離散的なリスク評価を行った嘉戸・図子 (2007) による回帰モデル結果を紹介し、続いてここで用いる回帰モデルの構築及び分析結果を示す。分析には、嘉戸・図子 (2007) で用いた同じデータに対し候補となる説明変数の全ての組み合わせに対し、ロジスティック回帰モデルを適用しリスク評価を連続的に行い、更に情報量規準を用いて変数選択を行うことにより最終的な回帰モデルを特定する。そして被害リスク要因の特定を試みる。最後に既往研究結果との比較を行い、考察を加える。

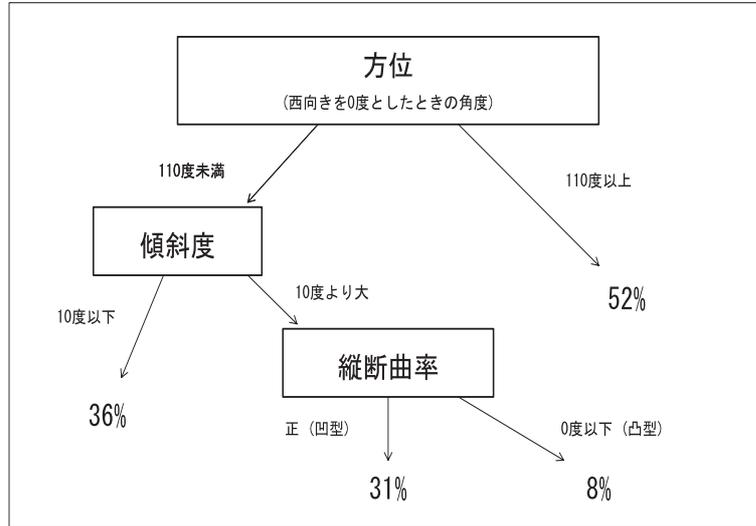


図 1. 回帰木モデル (嘉戸・図子, 2007)

注) 嘉戸・図子 (2007) による回帰木モデルの流れ. 被害リスクを, 斜面方位, 傾斜度, 土地の凹凸の 3 要因により, 4 段階に区分する.

2. 回帰木モデルによる既往研究

森林の冠雪害に対する先行研究としては, 嘉戸・図子 (2007) による回帰木モデルを用いたアプローチがある. このモデルによる冠雪害リスクの予測は, 以下の通りである. まず, 西から 110 度以上の斜面方向にある地点の被害リスクは 52% である. 次に, それ以外の斜面方向に関しては, 傾斜度が 10% 以下の地点では被害リスクは 36%, それ以上の地点に関しては, 更に縦断曲率が 0 以上 (凹型) の地点と 0 未満 (凸型) の地点に分けられ, 被害リスクはそれぞれ 31% と 8% である (図 1). すなわち, 斜面方位, 傾斜度, 土地の凹凸の 3 つで被害リスクを 4 段階に区分している. その結果, 連続に変化している変数にも関わらず, 被害リスクはこれら 4 段階に集約される. 災害リスク判別に用いた 3 つの要因は, 冠雪害に大きな影響を与える風に関するものである.

3. ロジスティック回帰モデルと変数選択

地点 i ($1, 2, \dots, m$) における立木の総数を n_i (ただし $\sum_{i=1}^m n_i = n$), その地点での被害発生確率を p_i とする. この中から y_i 本に被害が発生し, 地点 i での被害発生確率を p_i とすると, y_i は次の二項分布に従う.

$$[1] \quad y_i \sim \text{Bin}(n_i, p_i)$$

ここで $\mathbf{y} = (y_1, \dots, y_m)'$ とすると y_i の密度関数 f は

$$[2] \quad f(\mathbf{y}_i) = n_i C_{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i}$$

となる. 被害発生確率 p_i は, 林分の特性や地形因子によって変動する事が予想されるので, これらの要因を説明変数とする. すなわち地点 i における r 個の説明変数を $\mathbf{x}_i = (x_{i1}, \dots, x_{ir})'$ とし, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m)'$ とする. ロジスティック回帰モデルでは, p_i のロジット変換が, 説明変数の線形結合で表される. つまり

$$[3] \quad \text{logit } p_i = \log \left(\frac{p_i}{1 - p_i} \right) = \sum_{j=1}^r \beta_j x_{ij} = \mathbf{x}_i' \boldsymbol{\beta}$$

であり

$$[4] \quad p_i = \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})}$$

となる. ここで, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)'$ は未知パラメータベクトルである.

次に, 変数選択問題, 説明変数の候補から最適な変数の組み合わせを探索する事を考える. 変数選択は, カルバック・ライブラー情報量 (Kullback and Leibler, 1951) に基づくリスク関数

$$[5] \quad R = -2E_{\mathbf{y}} E_{\mathbf{u}} [\ell(\hat{\boldsymbol{\beta}}; \mathbf{u})]$$

が最小になる変数の組み合わせを探索する事により行う. ここで \mathbf{y} と \mathbf{u} は独立に同一の分布に従い, 期待値 $E_{\mathbf{y}}$ と $E_{\mathbf{u}}$ はそれぞれ, 設定したモデルと, 真のモデルでの期待値を表すものである. また $\hat{\boldsymbol{\beta}}$ は, 設定した変数の組み合わせの下での, 未知パラメータベクトル $\boldsymbol{\beta}$ の最尤推定量であり,

対数尤度 (同時確率密度)

$$\begin{aligned}
 \ell(\boldsymbol{\beta}; \mathbf{y}) &= \sum_{i=1}^m \log f(y_i) \\
 [6] \quad &= \sum_{i=1}^m \{\log {}_{n_i}C_{y_i} + y_i \log p_i + (n_i - y_i) \log(1 - p_i)\}
 \end{aligned}$$

の最大化により求められる。つまり $\hat{\boldsymbol{\beta}}$ は

$$\begin{aligned}
 \hat{\boldsymbol{\beta}} &= \arg \max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}; \mathbf{y}) \\
 [7] \quad &= \arg \max_{\boldsymbol{\beta}} \sum_{i=1}^m \{y_i \mathbf{x}'_i \boldsymbol{\beta} - n_i \log(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}))\}
 \end{aligned}$$

で得られ、この $\hat{\boldsymbol{\beta}}$ を用いて、地点 i の被害発生確率の推定量は、 $\text{logit } \hat{p}_i = \mathbf{x}'_i \hat{\boldsymbol{\beta}}$ と表される。

さて、リスク関数 [5] は、真のモデル (未知) の下での期待値を含んでいるために陽に計算することが出来ず、推定することになる。Akaike (1973) は、その推定量として赤池情報量規準 AIC (Akaike's information criterion) を

$$[8] \quad \text{AIC} = -2\ell(\hat{\boldsymbol{\beta}}; \mathbf{y}) + 2r$$

で提案した。この規準量は、カルバック・ライブラーのリスク関数 [5] を対数尤度で推定した際のバイアスを、説明変数の個数で補正したものである。すなわち、モデルの当てはまりの良さを対数尤度の部分で、また説明変数の増加により過度に当てはまりが良くなる事に対する罰則をその個数で評価している。一方、Schwarz (1978) は、未知パラメータに事前分布を導入して、その周辺尤度の最大化によるモデル選択規準量 BIC (Bayesian information criterion) を

$$[9] \quad \text{BIC} = -2\ell(\hat{\boldsymbol{\beta}}; \mathbf{y}) + r \log n$$

で提案した。また、近年 Yanagihara *et al.* (2003) は、確率展開を用いて

AIC のバイアスを補正した規準量 CAIC (bias Corrected AIC) を

$$[10] \quad \text{CAIC} = -2\ell(\hat{\beta}; \mathbf{y}) + 2r + \frac{1}{n}(a_1 + a_2 + a_3)$$

で提案した. ここで

$$[11] \quad a_1 = (\hat{\mathbf{p}} - \hat{\mathbf{q}})' \mathbf{V}(\mathbf{W}_{(3)} + \mathbf{W}_{(0)} \mathbf{W} \mathbf{W}_{(0)}) \mathbf{V}(\hat{\mathbf{p}} - \hat{\mathbf{q}})$$

$$[12] \quad a_2 = (\hat{\mathbf{p}} - \hat{\mathbf{q}})' \mathbf{W}_{(0)} \mathbf{V} \mathbf{W}_{(0)} (\hat{\mathbf{p}} - \hat{\mathbf{q}})$$

$$[13] \quad a_3 = 2\hat{\mathbf{p}}' \mathbf{W}_{(0)}^2 \hat{\mathbf{q}}$$

であり,

$$[14] \quad \hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_m)',$$

$$[15] \quad \hat{\mathbf{q}} = (\hat{q}_1, \hat{q}_2, \dots, \hat{q}_m)' = (1 - \hat{p}_1, 1 - \hat{p}_2, \dots, 1 - \hat{p}_m)',$$

$$[16] \quad \boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_m)' = \frac{1}{n}(n_1, n_2, \dots, n_m)'$$

$$[17] \quad \mathbf{V} = \text{diag}(\rho_1 \hat{p}_1 \hat{q}_1, \dots, \rho_m \hat{p}_m \hat{q}_m),$$

$$[18] \quad \mathbf{W} = [w_{ij}] = \mathbf{X}(\mathbf{X}' \mathbf{V} \mathbf{X})^{-1} \mathbf{X}',$$

$$[19] \quad \mathbf{W}_{(0)} = \text{diag}(w_{11}, \dots, w_{mm}),$$

$$[20] \quad \mathbf{W}_{(3)} = [w_{ij}^3]$$

である. AIC は, バイアスを $O(1)$ のオーダーで補正しているが, CAIC は $O(n^{-1})$ のオーダーで補正しているのが特筆すべき点である. この他にも, 変数選択に用いられる規準量には様々なものが提唱されているが, 今回は代表的なこれら 3 つの規準量でのモデル選択を行い, 最終的に被害リスク推定を行うモデルを特定した.

さらに, 規準量により異なるモデルが選ばれた場合, それらの比較には尤度比統計量であるデビアンズ

$$[21] \quad D = -2 \log \left(\frac{\text{モデルの尤度}}{\text{完全にフィットしたモデルの尤度}} \right) \\ = 2 \sum_{i=1}^m \left\{ y_i \log \left(\frac{y_i}{n_i \hat{p}_i} \right) + (n_i - y_i) \log \left(\frac{n_i - y_i}{n_i - n_i \hat{p}_i} \right) \right\}$$

を用いた (McCullagh and Nelder, 1989). ここで, 完全にフィットしたモ

デルとは、推定量とデータが完全に一致したモデルのことである。デビアンスは、推定したモデル（説明変数の個数を k とする）が正しいという仮定の下で、漸近的に自由度 $(m - k - 1)$ の χ^2 分布に従うことを用いて適合度の良し悪しを判断するものである。

4. 解析結果と考察

ここでの分析には、2004年1月の富山県小矢部市子撫川ダム近傍のスギ林における冠雪害データを用いた。実際の被害発生時の気象条件や、被害発生地域における森林の特性については嘉戸・凶子(2007)によるもので、そこでは20m×20mの観測地47箇所における冠雪害(幹折れ, 幹曲がり, 根返り)と気象要因が観測されている。これら47地点におけるスギの総本数は1761本であり、そのうち599本に冠雪害が発生したので、全体の被害発生率は約34%である。また、これらの地域における森林の特性や地形因子については、森林の特性としては、林齢, 平均胸高直径, 平均樹高, 形状比, スギ品種の5つ, 地形因子としては、標高, 土壤湿潤度, 横断面係数, 縦断面係数, haあたりの材積, 斜面傾斜度, 地上開度, 地下開度, 流出寄与域, 斜面方位の10個と、次に示す計15個の説明変数をここでの候補として用いた。

1. 林齢: 樹木の年齢。対象地域が人工林であるため、同じ観測地ではほぼ同一である。
2. 平均胸高直径: 胸の高さにおける樹木の直径の平均。今後 DBH (diameter at breast height) と表す。
3. 平均樹高: 樹木の高さの平均。
4. 形状比: 平均樹高 ÷ 平均胸高直径。細長い木程大きな値となる。
5. スギ品種: ボカスギ・カワイダニスギか、それ以外かを表すダミー変数。これらの品種は挿し木品種であり、成長が早い反面、冠雪害に弱い。今後、品種と表す。
6. 標高: 海面からの高さ。
7. 土壤湿潤度: 土壤の水分状況を表す指標。後述の流出寄与域を a , 斜面

表 1. 方位を表すダミー変数

	東	西	南	北
北	0	0	0	1
北東	1	0	0	1
東	1	0	0	0
南東	1	0	1	0
南	0	0	1	0
南西	0	1	1	0
西	0	1	0	0
北西	0	1	0	1

注) 8つの方位を, 東西南北 4つのダミー変数の組み合わせで表す. 例えば, 「北」は (東, 西, 南, 北)=(0, 0, 0, 1) で, 「北東」は (東, 西, 南, 北)=(1, 0, 0, 1) である.

傾斜度を b とすると, $\log(a/\tan b)$ で定義される. 湿潤なほど大きな値を示す (Beven, 1997). 今後, 湿潤度と表す.

8. 横断面係数: 等高線方向の凹凸を表す指標. その数値は曲率. 凹型の地形では正の値を, 凸型の地形では負の値, 平衡地形では 0 に近い値となる (Moore *et al.*, 1993).
9. 縦断面係数: 縦方向の凹凸を表す指標で, 考え方は横断面係数と同じ.
10. ha あたりの幹材積: 森林の生育状況を表す指標. 林齢が高く, 立木密度が高く, 土地が肥沃なほど大きな値となる. 今後, 材積と表す.
11. 斜面傾斜度: 水平方向からの角度を表す.
12. 地上開度: 空の見通しの度合いを表す指標であり, 山頂や尾根部で大きな値をとる.
13. 地下開度: 空が地表に遮られる度合いであり, 窪地や谷底で大きな値をとる.
14. 流出寄与域: その土地に流れ込む水の集水域を数値実験により算出したもの. 尾根で 0 となる.
15. 斜面方位: 東西南北を表すダミー変数により 8 つの方位を表す (表 1).

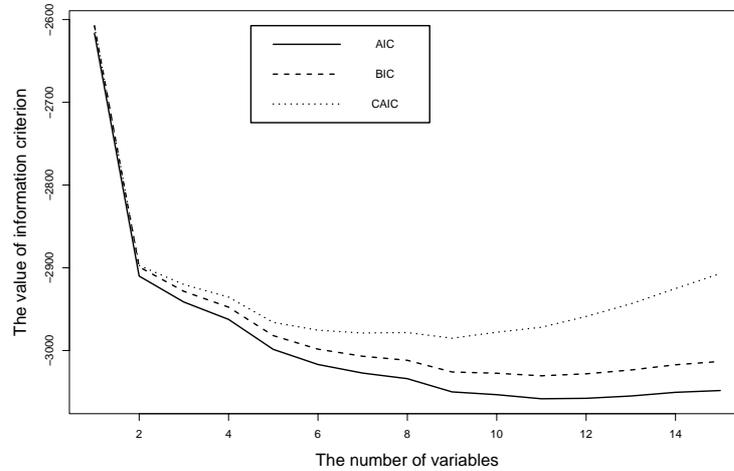


図 2. 情報量規準の値

注) 横軸は変数の個数 (定数は除く) を表す。縦軸は、与えられた変数の個数の中で、規準量を最小にするモデルにおける規準量の値を表す。

これら 15 種類の説明変数全ての組み合わせに対しロジスティック回帰モデルをあてはめ、3 つの規準量 (AIC, BIC, CAIC) により変数選択を行った結果を表 2 に示す。ただし、方位に関しては、ダミー 4 変数を 1 セットとみなしてそれぞれの方位が関わるか否かにより全ての方位を表現した。つまり、ある特定の方位に限っての必要性は考えないこととする。表 2 においては、変数の個数を 1 から 15 に固定し、その中で推定された最適な変数の組み合わせを列挙した。例えば、変数の個数 4 においては、全ての変数の組み合わせの数は ${}_{15}C_4 = 1365$ 通りとなり、その中で、品種、標高、地下開度、方位という組み合わせが、全ての規準量 (AIC, BIC, CAIC) において最適と推定されたことを示している。次に変数の個数については、AIC, BIC では 11 個のモデルが、CAIC では 9 個のモデルが最適と推定された (図 2)。変数の組み合わせは、AIC, BIC によると林齢, DBH, 品種, 標高, 湿潤度, 縦断面係数, 斜面傾斜度, 地上開度, 地下開度, 流出寄与域, 方位, CAIC では、林齢, DBH, 品種, 標高, 湿潤度, 地上開度, 地下開度, 流出寄与域, 方位となった (表 2, 3)。すなわち、CAIC を規準に選ばれた変数は全

表 2. 変数選択の結果

変数の個数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
林齢															
DBH						*									
平均樹高															
形状比															
品種															
標高															
湿潤度															**
横断面係数															
縦断面係数															
材積															
斜面傾斜度															
地上開度															
地下開度															
流出寄与率															
方位															

注) 「変数の個数」は、定数以外の変数の個数を意味する。記号「^{*}」は全ての規準量 (AIC, BIC, CAIC) で選択された変数、記号「^{**}」は AIC と BIC のみ、「^{**}」は CAIC のみで選択された変数を表す。すなわち、変数の個数 6 個においては、林齢は全ての規準量で選択されたが、湿潤度は CAIC でのみ選択されたことを意味する。

表 3. 最適モデルにおける回帰係数の推定結果

	AIC, BIC		CAIC	
	係数	<i>p</i> 値	係数	<i>p</i> 値
定数	-6.199	0.23	-10.08	0.00
林齢	0.049	0.00	0.048	0.00
品種	-2.105	0.00	-1.991	0.00
DBH	-0.037	0.04	-0.046	0.01
標高	0.008	0.00	0.007	0.00
湿潤度	0.281	0.01	0.218	0.02
縦断面係数	-22.75	0.06	-	-
斜面傾斜度	-0.014	0.50	-	-
地上開度	-0.080	0.05	-0.033	0.09
地下開度	0.110	0.00	0.115	0.00
流出寄与域				
北	-0.207	0.37	0.011	0.96
東	1.343	0.00	1.539	0.00
南	-0.104	0.50	-0.075	0.61
西	0.085	0.73	0.259	0.23

注) CAIC では、縦断面係数と斜面傾斜度は選択されなかったため、該当欄には - を記載してある。

て、AIC, BIC でも選ばれていることが分かる。これらの推定結果に基づき、被害リスクの予測が可能となる。そのロジット曲線への適合具合を図 3 に、実測被害リスクと予測被害リスクの関係を図 4 に示す。

同じロジスティック回帰モデルにおいて AIC, BIC と CAIC では異なるモデルが選ばれたため、デビアンズ [21] による比較を行った。AIC, BIC により選択されたモデルのデビアンズは 7.81 であり、CAIC により選択されたモデルのデビアンズは 11.50 であった。従ってその差は 3.69 であり、自由度は変数の個数の差の 2 である。対応する *p* 値は 0.16 であるので、変数を 2 つ増やすことによる有意な改良はないと判断できる。従って、CAIC

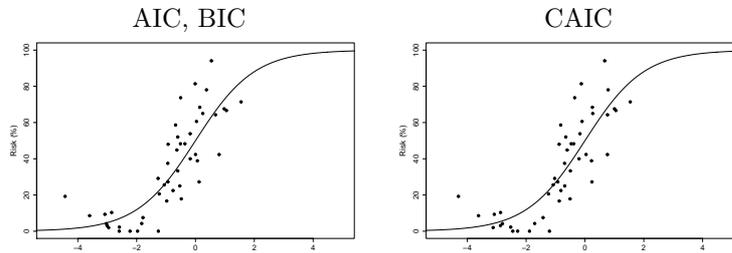


図 3. ロジット曲線への当てはまり

注) 横軸は $x_i'\beta$ の値、縦軸は実際のリスクを表す。また、曲線は推定されたロジット曲線を表す。

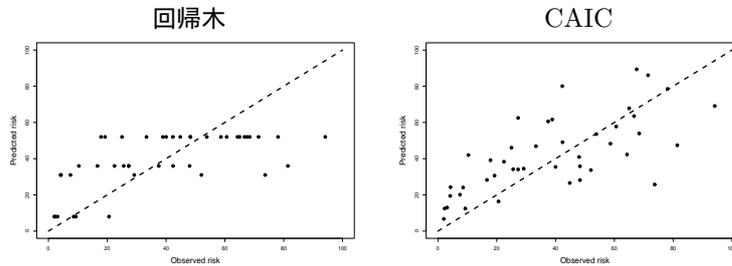


図 4. 被害発生確率の実測値と予測値の比較

注) 横軸は観測リスク、縦軸は各モデルにより予測されたリスクを表す。また、破線は観測されたリスクと予測リスクが等しい状態を表す。

により選ばれたモデルの方が良いと結論できる。

ここで、被害リスクの判別がうまく機能しているかどうかを調べるために、カットオフ値(高リスクと低リスクの分岐点)を50%に設定した判別能力分析を行った(表4)。ここで、感受性とは、実測被害リスク50%以上の群において予測被害リスクも50%以上であった割合、特異性とは、実測被害リスクが50%未満の群において予測被害リスクも50%未満であった割合で定義される(森實, 2007)。例えば、回帰木モデルにおける感受性は $400/(112 + 400)$ である。感受性・特異性共に100%が理想的な判別法である。また、総誤判別率は、カットオフ50%で誤った判別を行った割合であり、低いほど理想的である。回帰木モデルに比べ、ロジスティック回帰モ

表 4. 被害発生確率の判別能力の比較

	回帰木モデル (嘉戸・図子, 2007)		ロジスティック回帰 (CAIC による変数選択)	
	実測		実測	
	50% 未満	50% 以上	50% 未満	50% 以上
予測 50% 未満	945	112	1095	239
50% 以上	304	400	154	273
感受性	78%		53%	
特異性	76%		88%	
誤判別率	24%		22%	

注) 感受性とは、実測被害リスク 50% 以上の群において、予測被害リスクも 50% 以上である割合を、特異性とは、実測被害リスク 50% 未満の群において、予測被害リスクも 50% 未満である割合を表す。

デル (CAIC による変数) の方が誤判別率が低いことが分かる。

説明変数の被害リスクへの影響についても見てみる。推定された未知パラメータの符号は、その変数がリスクを高めるか否かを表すものであるが、CAIC により選ばれた変数では、正の係数を取るものは、林齢、品種、標高、湿潤度、地下開度、方位の北、東、西であり、負の係数を取る係数は DBH、品種、地上開度、流出寄与域、方位の南であった。まず、林齢については、若い木に関しては外圧に対する被害発生が低いいため、林齢は被害リスクを上げる要因となっている。標高については、同じ気候条件下では、一般的に標高が高いほど積雪量が多いため冠雪害が発生しやすい。湿潤度については、高ければそれは樹高成長が良い土地であることを意味し、結果として形状比の高い(樹幹が細長い)冠雪害に弱い木が育ちやすい。地下開度が高い地域は谷底付近であり、凹型の地形である。従って風の影響を受けにくく、被害リスクを上げる効果がある。方位については、北、東、西と異なる向きの変数が被害リスクを上げると推定されたが、係数の値並びに p 値から東向き被害リスクが高い事が分かる。東向き斜面は、被害発生時に風下側であることから、今回のケースでは風が被害リスクを軽減する効果があっ

たことが分かる。

その一方で、被害リスクを下げるものとして、DBH は、太い樹木ほど大きな値となり、これらの樹木における冠雪害リスクが低い事は容易に想像がつく。品種については、ボカスギ・カワイダニスギか否かのダミー変数であり、ボカスギ・カワイダニスギを 0 としているので、これらの品種の被害リスクが高い事を示している。地上開度については、値が大きくなる事は山頂や尾根である事を意味するので、山頂付近では被害リスクが低くなる。これは標高に対する係数が正値であることに矛盾しているように思えるが、標高が高なくても尾根型の地形は存在し、そのような地形においては風の影響が強いと考えられる。今回は、風が被害リスクを低くする効果が期待されたので地上開度の係数が負となったと考えられる。流出寄与域については、この値が大きくなるということは、十分な水分や養分が流れ込み、木々の生育が良く、結果として被害リスクの軽減に繋がっている。方位の南については、 p 値が高く、係数の値自体も 0 に近いので、本質的に効いている変数ではないと考えられる。

本研究と同じデータを用いた先行研究である回帰木モデル(嘉戸・凶子, 2007)の結果を見てみると次のことが指摘できる。この回帰木モデルで使用された説明変数は全て風の影響に関するものである。冠雪害に対する風の影響としては、被害リスクを弱める要因として立木を揺らし雪を落下させる点が考えられ、逆にリスクを強める要因として、風圧により樹木へ雪を付着させる効果がある点と、冠雪により傾いた樹木に更なる風圧が加わる点が考えられる。すなわち、風に関しては冠雪害発生に対して被害リスクを上げる効果と下げる効果の両方が考えられるわけである。今回の被害発生時の風向は西南西～南西であったため、西向き、南向きの斜面は被害発生時に風を受け続けたと考えられた。回帰木モデルにおける変数は斜面方位(特に西向きかどうか)、傾斜度と土地の凹凸であり、斜面の方位は風を受けるか否か、傾斜度と土地の凹凸は風が樹木に対して影響を与えるかを判別するためのものであった。具体的には被害発生当時、西南西～南西からの風が吹いていたため、西向きか否かは、風を直接受ける斜面か否かを判定する要因となる。東向きが被害リスク 52% で最も高いので、今回のケースで

は風が強ければ被害リスクは軽減される傾向にある。次に風の強い西向き斜面においては、傾斜度が高ければ傾斜が急であることを意味し、風を直接的に受けるかどうかを判定する要因である。斜度が緩い地域の被害リスクが高いということは、被害リスクの低い西向き斜面の中でも風の影響は被害リスクを軽減させる傾向にあることを意味する。そして、西向きかつ急斜面という、上記2つの観点において最も風の影響を受けやすい地帯においては、縦断曲率が負値、すなわち凸型の地帯の被害リスクが低くなる。凸型の地帯は、凹型の地帯に対して風の影響が更に強いことが予想されるので、やはり風は被害リスクを軽減させる傾向があることが分かる。

5. 結論

今回は、先行研究である嘉戸・図子(2007)と同じデータに対し、連続的なリスク評価を試みるためにロジスティック回帰モデルを用いて、変数選択のアプローチにより被害リスクの予測を行った。その目的の1つは、地点の持つ説明変数を連続的に捉え、回帰モデルにおける変数選択法により、被害リスクに寄与している要因を特定することにある。変数選択規準量として、今回はAIC, BIC, CAICの3つを用いた。その結果推定された最適モデルはAICとBICによる結果同じになったが、CAICでは異なる組み合わせが推定された。しかしながら、CAICで選ばれた変数は全ての規準量で共通に選ばれた変数、すなわち、林齢, DBH, 品種, 標高, 湿潤度, 地上開度, 地下開度, 流出寄与域, 方位であり、これに縦断面係数, 斜面傾斜度を加えたモデルがAIC, BICで選ばれたものである。今回のケースでは、先行研究である回帰木モデルの結果と同様に、基本的に風を受けると被害リスクが下がるという結果が得られたが、逆に風の影響が被害リスクを高める場合もあることに注意が必要である。

ここで提示したロジスティック回帰モデルと変数選択のアプローチにより、被害リスク要因の特定と、被害リスクの予測値の連続的な評価が可能になることが分かった。また、選ばれた変数の種類並びに係数に関しても、既存の研究結果と比較して、妥当な結果が得られた。このことは、冠雪害リスク解析において、本モデルが十分に有効であることを示しており、今後更な

る応用分析を試みることにより、その精度を上げて行く必要がある。今後の課題としては、他の標本地のデータによる被害リスクの予測を行い、更なるデータの収集により、連続的なリスク評価による詳細な被害リスクマップの作成がある。そうすることにより、冠雪害の危険地域に対する管理制御、冠雪害に強い品種の植林などリスクマネジメントを考慮した森林管理の対策が可能にあるものとする。

引用文献

- Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (B. N. Petrov & F. Csáki eds.), pp. 267–281.
- Beven, K. (1997) Topmodel: A critique, *Hydro. Proc.* 11: 1069–1085.
- 嘉戸 昭夫・関子 光太郎 (2007) 富山県のスギ林における冠雪害と地形要因の関係, *FORMATH* Vol. 6: 77–88.
- Kullback, S. and Leibler, R. (1951) On information and sufficiency. *Ann. Math. Statist.* 22: 79–86.
- McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models* (2nd ed.), Chapman & Hall, London, 511p.
- Moore, I. D., Gessler, P. E., Nielsen, G. A. and Peterson, G. A. (1993) Soil attribute predicting using terrain analysis, *Soil Sci. Amer. J.* 57: 443–452.
- 森實 敏夫 (2007) 入門医療統計学, 東京図書, 344p.
- Schwarz, G. (1978) Estimating the dimension of a model, *Ann. Statist.* 6: 461–464.
- Yanagihara, H., Sekiguchi, R. and Fujikoshi, Y. (2003) Bias correction of AIC in logistic regression models, *J. Statist. Plann. Inference* 115: 349–360.