

Supplementary Information for “Predicting Avalanche Danger in Northern Norway Using Statistical Models”

Kai-Uwe Eiselt¹ and Rune Grand Gravervsen^{1,2}

¹Department of Physics and Technology, University of Tromsø

²Norwegian Meteorological Institute

S1 Artificial neural network

The artificial neural network (ANN; e.g., LeCun et al., 2015) is essentially a generalisation of the logistic regression (e.g., Dreiseitl and Ohno-Machado, 2002). ANNs consist of an input layer, an output layer, and an arbitrary number of hidden layers in between. The layers are made up of so-called perceptrons whose activation is regulated by a specific activation function (e.g., the sigmoid function as in Eq. 2). Note that in the case of no hidden layer and if the sigmoid function is used as the activation function, the neural network is equivalent to a logistic regression (e.g., Dreiseitl and Ohno-Machado, 2002). The weights of the individual perceptrons and layers are determined during training via the minimisation of the cross-entropy error. The activation function of the hidden layer used here is the rectified linear unit (ReLU):

$$f(z) = \max(z, 0), \quad (1)$$

which has become the most popular choice as it accelerates the learning process compared to other methods (e.g., LeCun et al., 2015). The activation function for the output layer depends on the number of classes in the target variable. In the case of a binary classification, the sigmoid function is used. It predicts the probability $p(x)$ of the binary outcome for a set of predictors x_i (i.e., the probability that the outcome is 1 or true)

$$p(x_i) = \frac{1}{1 + e^{-z}}, \quad (2)$$

where z is a linear combination of the predictor values x_i and the model’s predicted coefficients β_i (for $i > 0$) and intercept β_0 :

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_r x_r. \quad (3)$$

The β_i are obtained by maximum likelihood estimation, maximising the log-likelihood function. Solving Eq. 2 for z , one may interpret z as the logarithm of the *odds* of outcome 1 (i.e., the probability of success over the probability of failure):

$$z = \log\left(\frac{p(x)}{1 - p(x)}\right). \quad (4)$$

Eq. 2 gives a probability between 0 and 1, and values ≤ 0.5 are categorised as 0 and values > 0.5 as 1 for our binary case.

If the target variable consists of multiple classes the number of nodes in the output layer must equal the number of classes. In this case we follow Sharma et al. (2023) in using the SoftMax function as the activation function for the output layer:

$$\text{SoftMax}(c_j) = \frac{e^{c_i}}{\sum_j e^{c_j}}, \quad (5)$$

25 where $\text{SoftMax}(c_j)$ corresponds to the probability of the class c_j of the target variable. Hence, in our 4-level case, we need an output layer with four nodes and j ranges from 1 to 4. Since our focus here is on the full integer danger levels, the ADL with the highest predicted probability is considered the predicted ADL.

A weakness of the ANN is that the weights are difficult to interpret, making it a “black-box” model (e.g., Dreiseitl and Ohno-Machado, 2002), similar to the RF. An advantage of the ANN is that it may find and represent structures in the data that
30 simpler linear models are missing. Several studies have applied ANNs to predict ADL (Schirmer et al., 2009; Dekanová et al., 2018; Fromm and Schönberger, 2022; Sharma et al., 2023; Blagovechshenskiy et al., 2023), however with varying success, as discussed in section 1.

Here we apply the ANN structure suggested by Sharma et al. (2023) (see their Table 4). Note that in contrast to the RF, the data were zero-transformed by subtracting the mean and dividing by the standard deviation for their usage with the ANN. The
35 ANN is generated using the implementation in the Python library Keras version 2.15.0 (<https://keras.io/>) built on Tensorflow version 2.15.0 (<https://www.tensorflow.org/>).

S2 Artificial neural network – Evaluation

The ANN exhibits a performance similar to the random forest (RF) for both the 4-level case and the binary case. Because of the inherent randomness of an ANN we trained 500 realisations on our training dataset for both cases, respectively.

40 Focussing on the 4-level case, the accuracies vary between 0.54 and 0.64, with a mean and median of 0.59 and a $1-\sigma$ standard deviation of ± 0.015 (2.5 %; see Fig. S1a for a histogram of the accuracies). However, the best overall accuracy does not imply the best accuracy per individual avalanche danger level (ADL). In fact, considering the confusion matrices of the “best” and “worst” of the 500 trained ANNs in terms of overall accuracy shown in Fig. S2a and b, it appears that the “worst” model performs much better in predicting ADL 4, while the higher accuracy of the “best” model results from better prediction of
45 ADLs 2 and 3. The better performance of the “worst” model in predicting ADL 4 is accompanied by a much higher tendency to misclassifying ADL 3 as ADL 4 (Fig. S2a, b). Moreover, the misclassification difference (i.e., the difference between true and predicted ADL) by the “worst” model is considerably larger than by the “best” model (Table S1).

In the binary case, the overall accuracies across the 500 ANNs vary slightly less than in the 4-level case (0.74 to 0.80) with a mean and median of 0.78 and a $1-\sigma$ standard deviation of ± 0.01 (1.3 %; see Fig. S1b for a histogram of the accuracies). In
50 contrast to the 4-level case, the “best” binary case model performs better than the “worst” model for both binary-case levels (BCLs), although the higher overall accuracy mostly results from a higher accuracy in predicting BCL 0 (Fig. S2c, d).

In summary, our optimised RF and the ANN of the structure suggested by Sharma et al. (2023) perform similarly when comes to predicting both ADL and BCL.

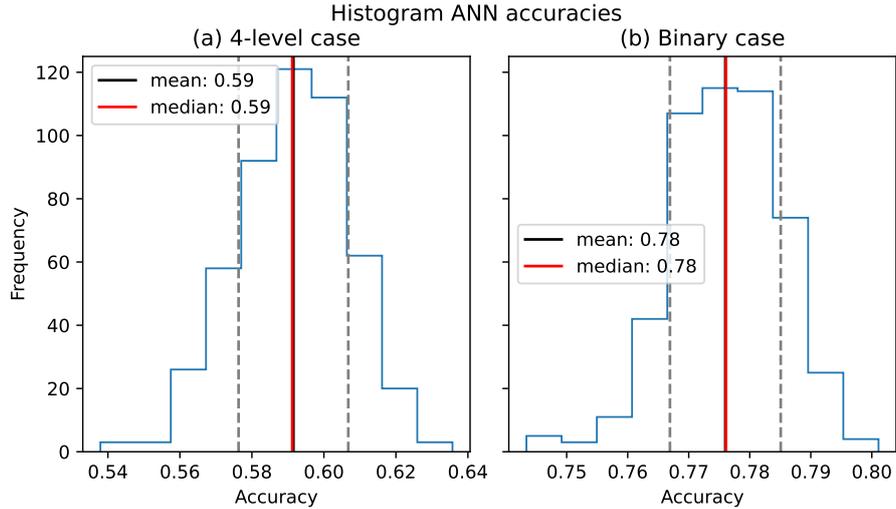


Figure S1. Histogram for the accuracies of 500 artificial neural networks (ANNs) trained for the (a) 4-level case and (b) binary case. The black and red vertical lines indicate the mean and median accuracy, respectively. The vertical dashed gray lines indicate the $1\text{-}\sigma$ standard deviation.

Table S1. Misclassification difference for the “best” and “worst” of 500 trained artificial neural networks (ANNs) in the 4-level case. See text S2 for details.

difference	best	worst
-3	0.0	0.0
-2	0.33	1.04
-1	16.7	20.11
0	63.57	53.79
1	18.57	23.57
2	0.77	1.48
3	0.05	0.0

S3 Artificial neural network – Hindcast

55 Like the RF (section 6), we employ the binary-case ANN to perform a hindcast of the BCL for the NORA3 period (1970-2024). The “best” ANN in terms of overall accuracy of the 500 trained binary-case ANNs is employed to perform the hindcast. We define the binary-case frequency (BCF) as the number of days with BCL 1 per season.

The general results are consistent with the results from the RF: There is no trend in the full-season BCF with partly significant weak negative and positive trends in winter (December through February) and spring (March through May), respectively (not

Confusion matrices – Best and worst ANNs

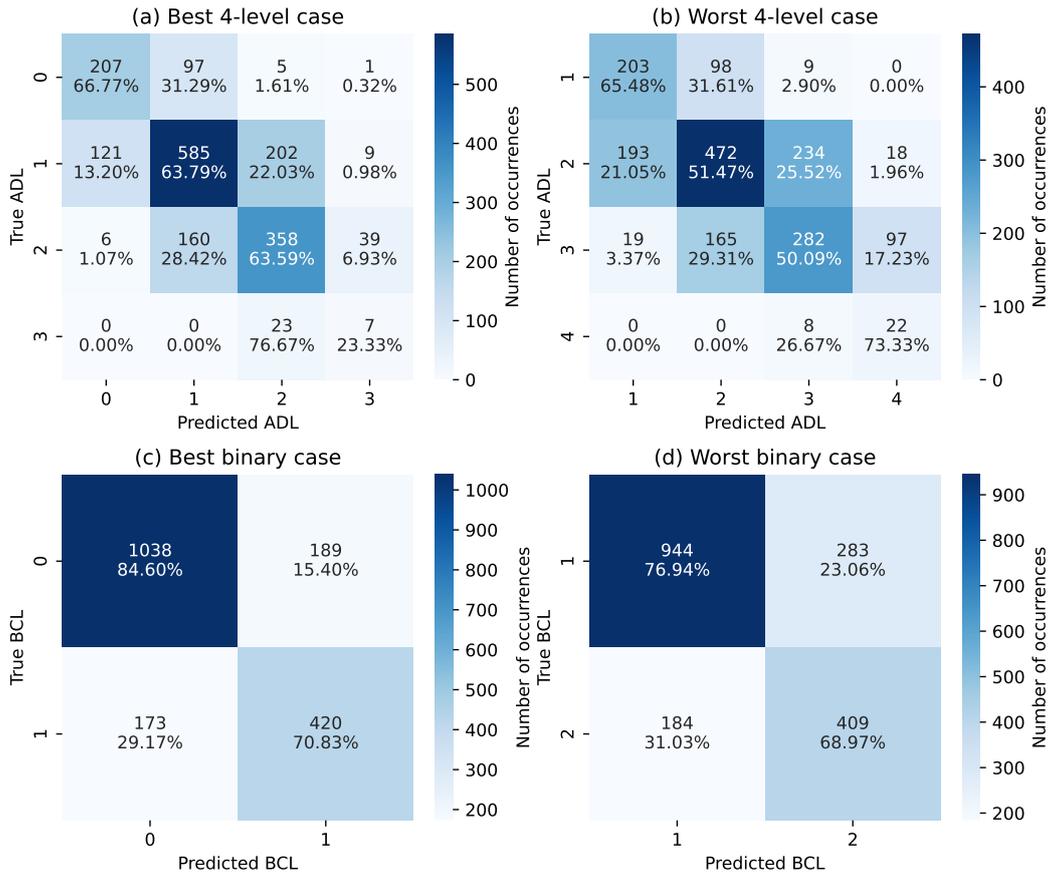


Figure S2. Confusion matrices for the (a, c) “best” and (b, d) “worst” of 500 trained artificial neural networks (ANNs) in the (a, b) 4-level case and (c, d) binary case. See text S2 for details.

60 shown). The BCF peak in the 1990s (especially in winter) is also evident in the results from the ANN (Fig. S3). However, there are some differences. For example, the ANN generally hindcasts a higher BCF (compare Figs. S3 and S11), which is mainly due to higher winter BCF. Notably, the correlation of the Arctic Oscillation (AO) with the ANN-derived BCF is even stronger than with the RF-derived BCF on an annual basis (Pearson R = 0.35-0.55) and similar on for 7-year rolling means (compare Figs. 8b, d and S4b, d). Finally, it appears that the ANN-based BCF hindcast exhibits more of an increasing trend in the decade
65 2010-2020 than the RF-based hindcast, being more consistent with the AO trend (compare Figs. 8c and S4c). This may be connected to the ANN-based BCF following more closely the accumulated snow than the RF-based BCF (compare Figs. 9 and S5).

In summary, the ANN-based BCF hindcast is similar to the RF-based hindcast and appears to confirm the impact of the AO on northern Norwegian avalanche activity. However, given the differences especially in the decade 2010-2020, the robustness

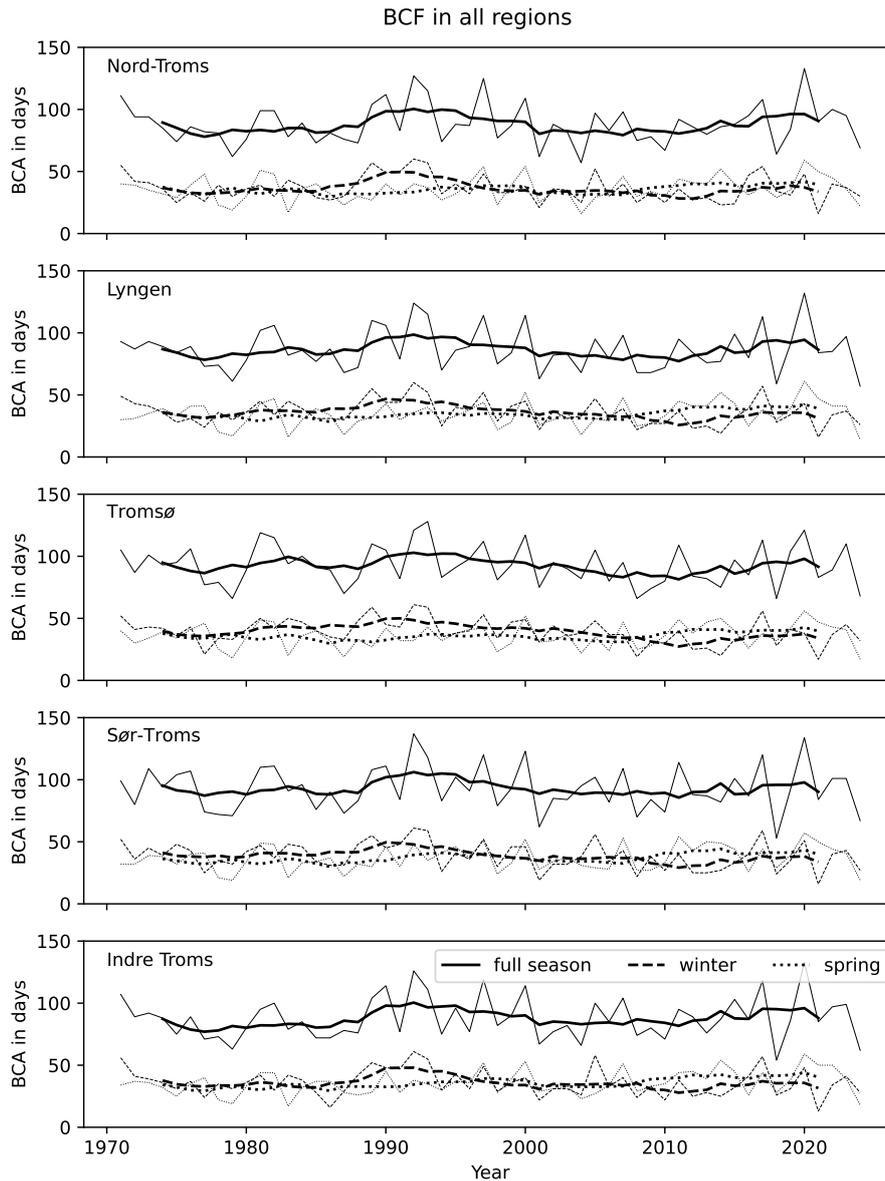


Figure S3. As Fig. 7, but for all regions and based on the artificial neural network (ANN) instead of the random forest (RF).

70 of this relationship remains uncertain. Decomposing the ADLs and BCLs according to the different avalanche problems may provide more clarity regarding the connection between northern Norwegian BCF and the AO, since some types of avalanches are likely more strongly determined by conditions related to, e.g., a high AO index than other types.

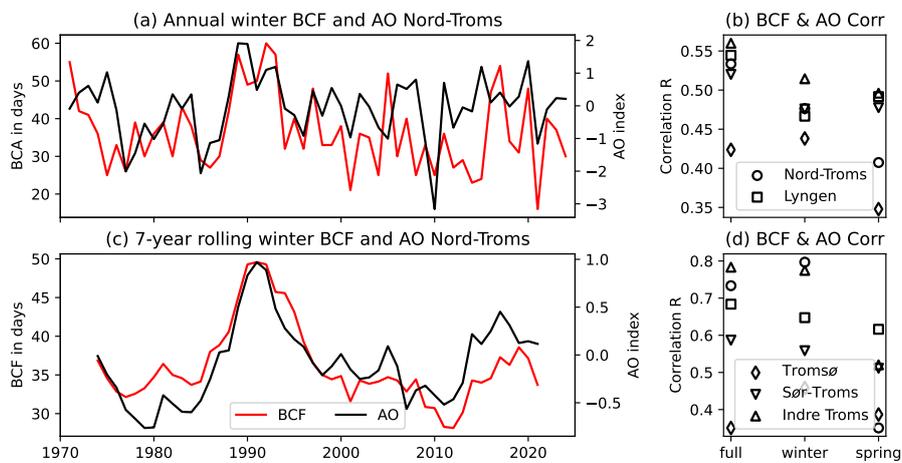


Figure S4. As Fig. 8, but for the artificial neural network (ANN) instead of the random forest (RF).

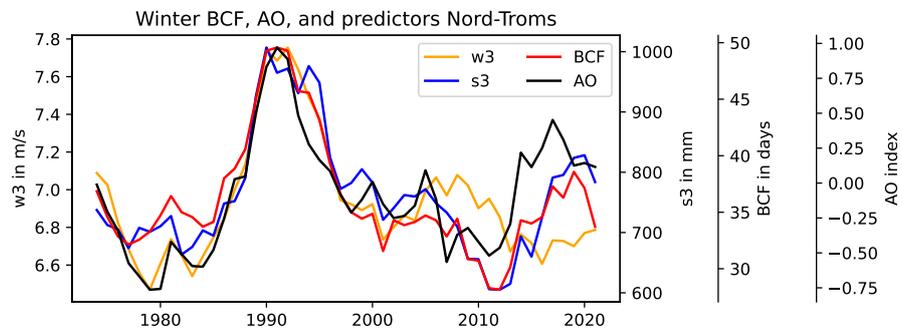


Figure S5. As Fig. 9, but for the artificial neural network (ANN) instead of the random forest (RF).

Table S2: List of the Norwegian avalanche warning regions (excluding Svalbard). For type-A regions daily warnings are issued, while for type-B regions warnings are only issued if the danger level exceeds 3. The area is given in km². See Fig. S6 for a map of the individual regions.

region code	region name	type	area
3005	Øst-Finnmark	B	17954
3006	Finnmarkskysten	A	19964
3007	Vest-Finnmark	A	13728
3008	Finnmarksvidda	B	15062
3009	Nord-Troms	A	9392
3010	Lyngen	A	2842
3011	Tromsø	A	6195
3012	Sør-Troms	A	8203
3013	Indre Troms	A	7372
3014	Lofoten og Vesterålen	A	12105
3015	Ofoten	A	7580
3016	Salten	A	10279
3017	Svartisen	A	13718
3018	Helgeland	A	18111
3019	Nord-Trøndelag	B	27304
3020	Sor-Trøndelag	B	21009
3021	Ytre Nordmøre	B	4566
3022	Trollheimen	A	8099
3023	Romsdal	A	6312
3024	Sunnmøre	A	7970
3025	Nord-Gudbrandsdalen	B	6036
3026	Ytre Fjordane	B	6253
3027	Indre Fjordane	A	5716
3028	Jotunheimen	A	6586
3029	Indre Sogn	A	6679
3030	Ytre Sogn	B	2703
3031	Voss	A	7779
3032	Hallingdal	A	4241
3033	Hordalandskysten	B	8161

Continued on next page

Table S2 continued from previous page

region code	region name	type	area
3034	Hardanger	A	7134
3035	Vest-Telemark	A	8083
3036	Rogalandskysten	B	8526
3037	Heiane	A	8636
3038	Agder sør	B	13343
3039	Telemark sør	B	8600
3040	Vestfold	B	3132
3041	Buskerud sør	B	10535
3042	Oppland sør	B	13078
3043	Hedmark	B	26999
3044	Akershus	B	4989
3045	Oslo	B	474
3046	Østfold	B	4802
	average study region	A	6801
	average type-A regions	A	8988
	average type-B regions	B	10712
	average Norway	A & B	9850

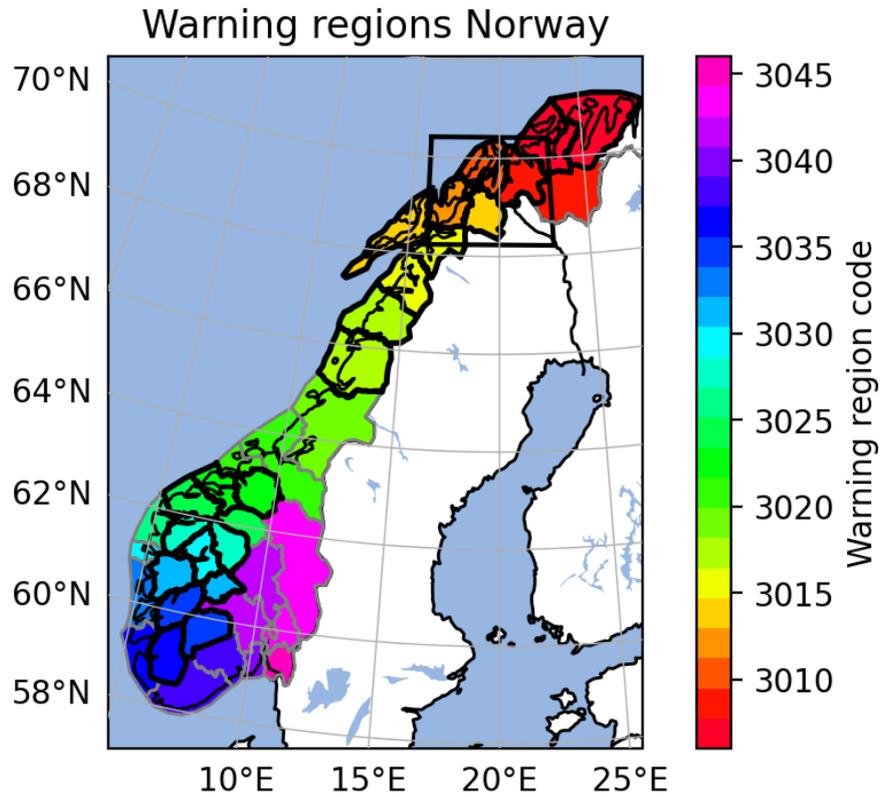


Figure S6. Map of the Norwegian avalanche warning regions. The thick black regions correspond to type A (warnings are issued daily) and the thinner gray regions correspond to type B (warnings only issued when the danger level exceeds 3). The black rectangle indicates our study region in northern Norway. The colour corresponds to the region code (see Table S2).

Table S3. As Table 3 but with the balanced data.

level	precision	recall	f1-score	support
1	0.74	0.49	0.59	917
2	0.44	0.63	0.52	917
3	0.42	0.60	0.50	917
4	0.88	0.42	0.57	917
accuracy			0.54	3668
macro avg	0.62	0.54	0.54	3668
weighted avg	0.62	0.54	0.54	3668

Table S4. As Table 5 but with the balanced data.

level	precision	recall	f1-score	support
0	0.74	0.82	0.78	1227
1	0.80	0.72	0.76	1227
accuracy			0.77	2454
macro avg	0.77	0.77	0.77	2454
weighted avg	0.77	0.77	0.77	2454

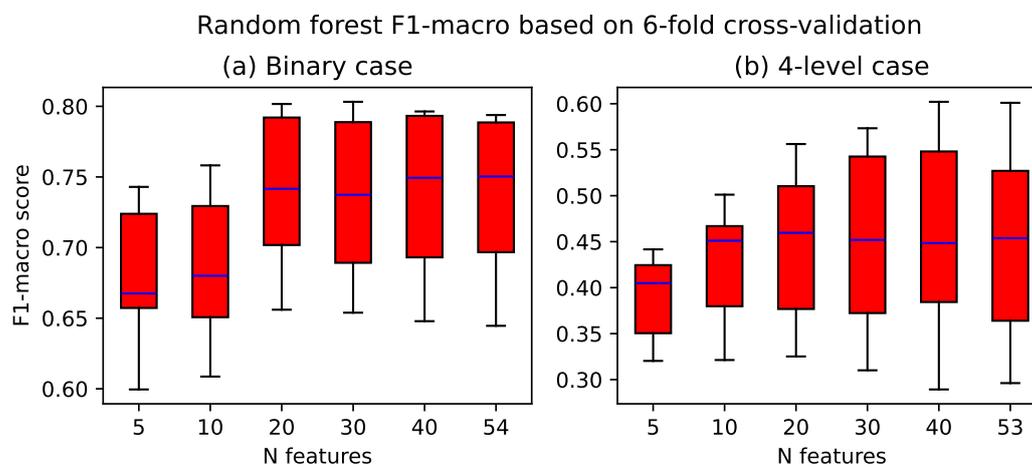


Figure S7. As Fig. 4, but with 6-fold cross-validation.

Table S5. Slope, Pearson correlation R, and p value (Wald test with a t distribution) of linear regressions from of avalanche activity hindcast with the random forest model from 1970 to 2023 for all regions.

	full	winter	spring
Nord-Troms			
slope	-0.04 ± 0.11	-0.14 ± 0.08	0.10 ± 0.08
R	-0.05	-0.24	0.18
p	0.69	0.07	0.20
Lyngen			
slope	-0.03 ± 0.11	-0.17 ± 0.08	0.16 ± 0.07
R	-0.03	-0.29	0.29
p	0.80	0.03	0.03
Tromsø			
slope	-0.09 ± 0.10	-0.21 ± 0.07	0.15 ± 0.07
R	-0.13	-0.39	0.28
p	0.36	0.00	0.04
Sør-Troms			
slope	-0.10 ± 0.11	-0.17 ± 0.08	0.11 ± 0.07
R	0.12	-0.28	0.22
p	0.38	0.04	0.11
Indre Troms			
slope	0.09 ± 0.11	-0.03 ± 0.08	0.16 ± 0.07
R	0.11	-0.06	0.30
p	0.41	0.66	0.03

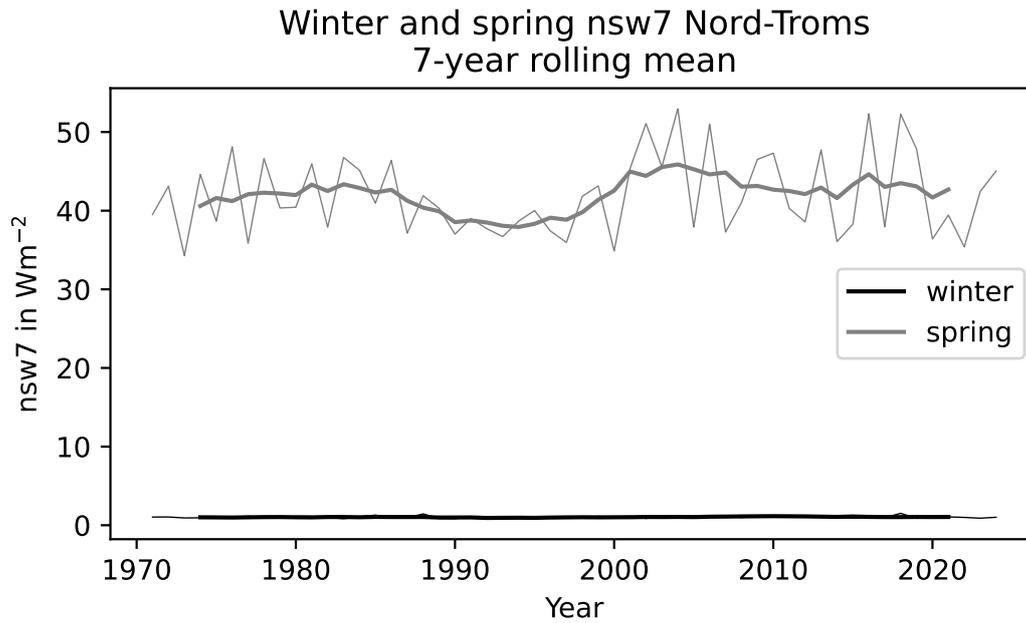


Figure S8. Winter and spring nsw7 in Nord-Troms on (a) an annual basis and (b) for 7-year rolling means.

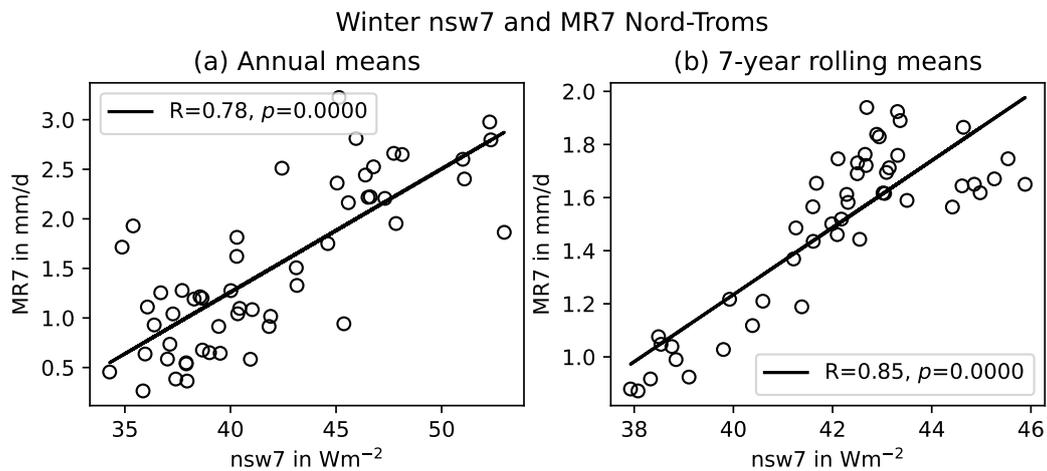


Figure S9. Correlation of spring nsw7 and MR7 in Nord-Troms.

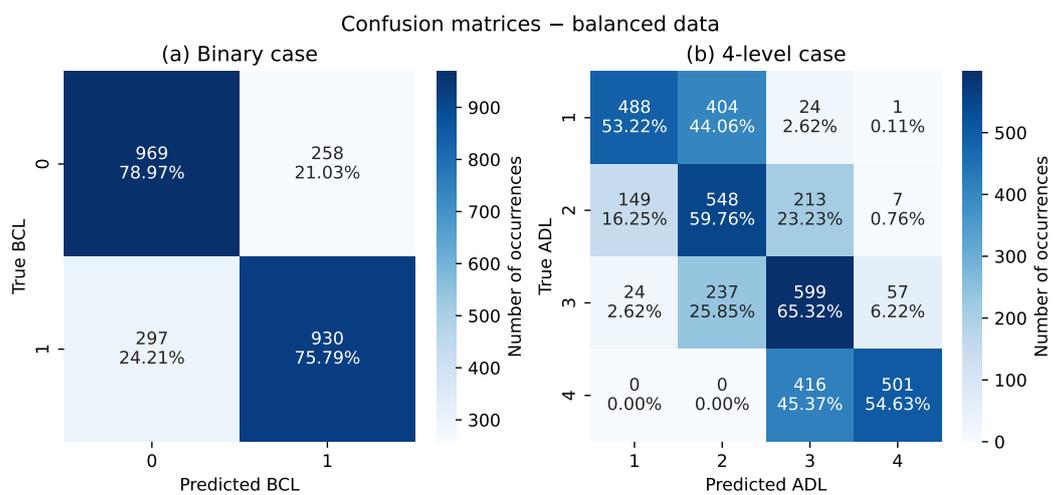


Figure S10. As Fig. 5, but for the balanced data.

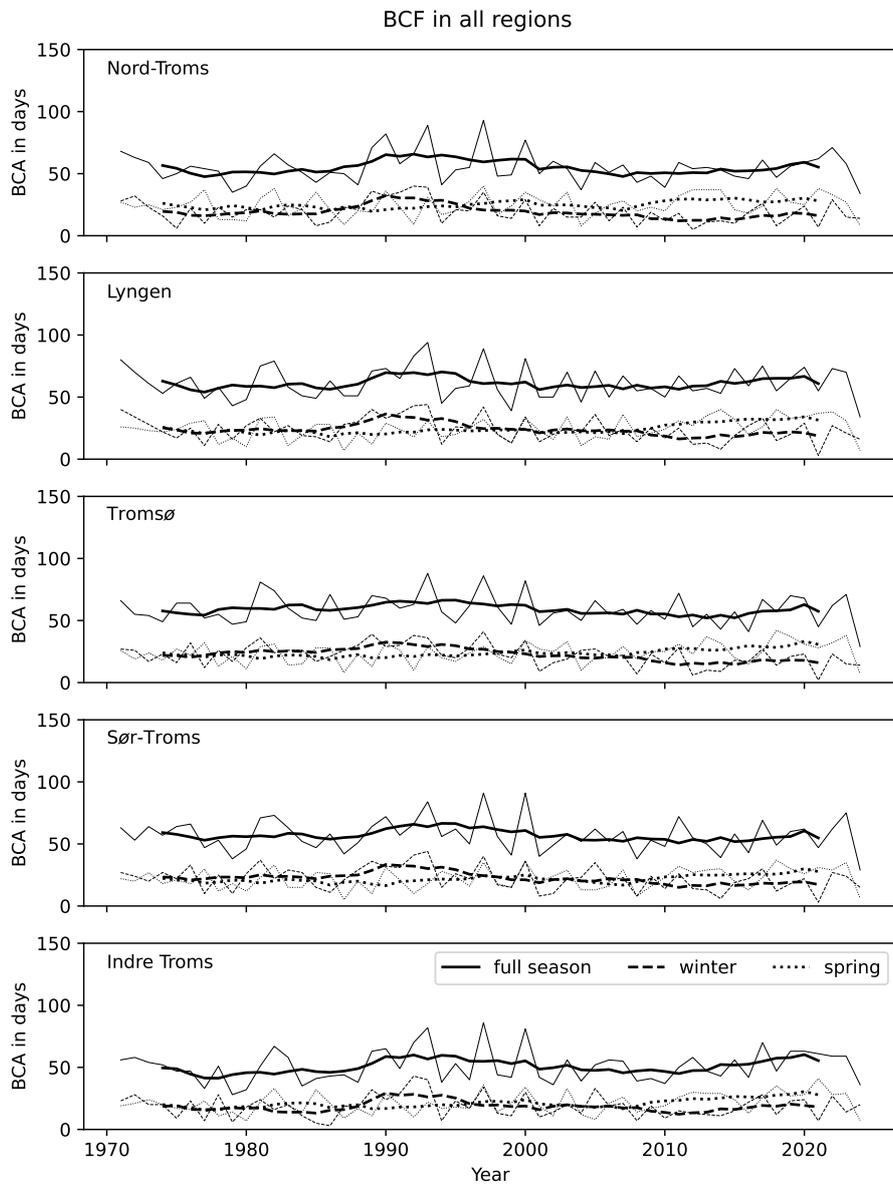


Figure S11. As Fig. S3, but for the random forest (RF).

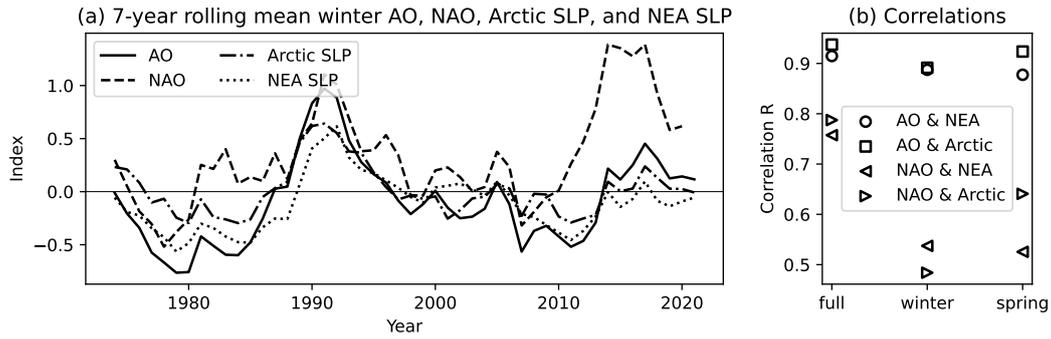


Figure S12. (a) Time series and (b) correlations of Arctic Oscillation (AO) and North Atlantic Oscillation (NAO) with Arctic ($> 75^{\circ}\text{N}$) and north-east Atlantic (NEA; $70\text{--}80^{\circ}\text{N}$, $10\text{--}20^{\circ}\text{E}$) sea-level pressure (SLP). The SLP data is taken from the ERA5 reanalysis. Note that all correlations exhibit p values < 0.01 based on a Wald test with a t distribution.

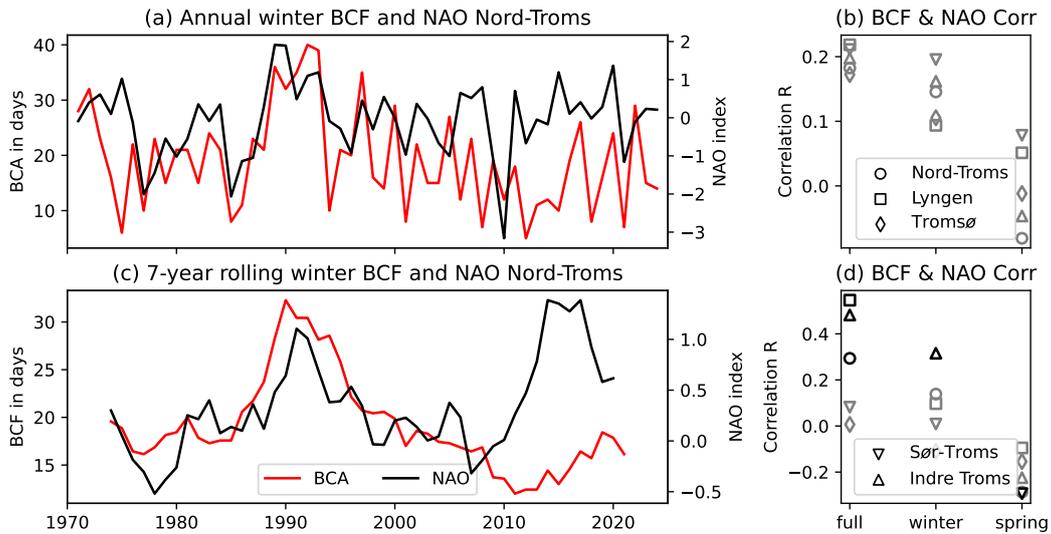


Figure S13. As Fig. 8, but for the North Atlantic Oscillation (NAO) instead of the Arctic Oscillation (AO).

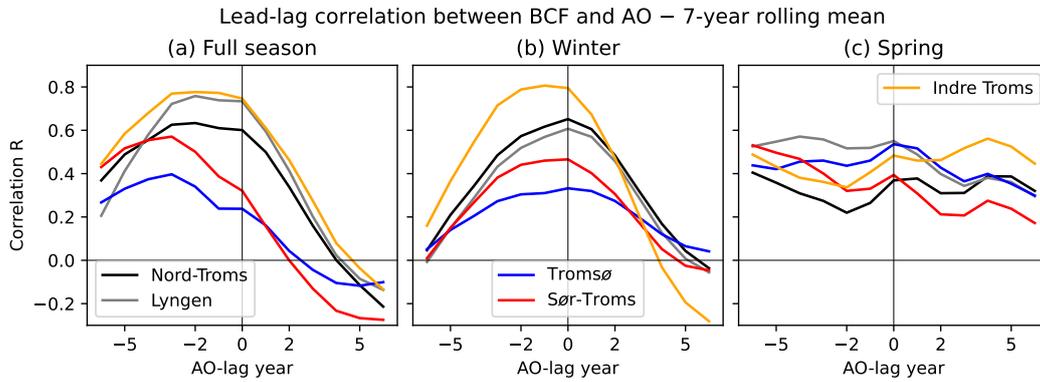


Figure S14. Lead-lag correlations between Arctic Oscillation (AO) and binary-case avalanche activity (BCF) for (a) full season, (b) winter, and (c) spring on an annual basis.

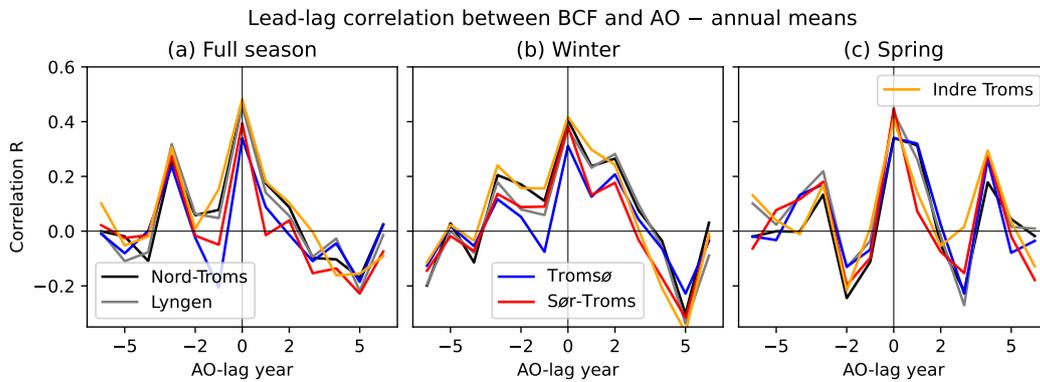


Figure S15. As Fig. S14, but for annual means.

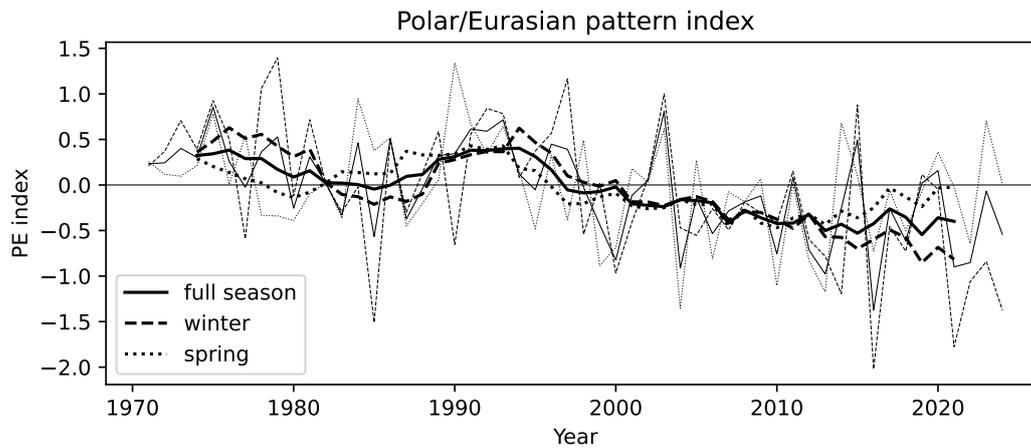


Figure S16. Polar/Eurasian pattern index for full season (continuous), winter (dashed), and spring (dotted). Shown are annual means (thin) and 7-year rolling means (thick). The index was downloaded from <https://www.cpc.ncep.noaa.gov/data/teledoc/poleur.shtml>.

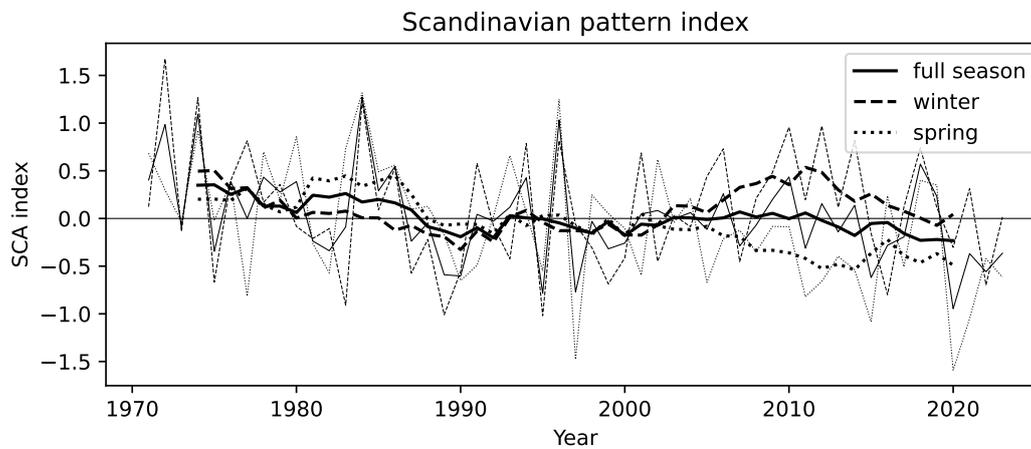


Figure S17. Scandinavian pattern index for full season (continuous), winter (dashed), and spring (dotted). Shown are annual means (thin) and 7-year rolling means (thick). The index was downloaded from <https://psl.noaa.gov/data/timeseries/month/SCAND/>.

References

- Blagovechshenskiy, V., Medeu, A., Gulyayeva, T., Zhdanov, V., Ranova, S., Kamalbekova, A., and Aldabergen, U.: Application of Artificial Intelligence in the Assessment and Forecast of Avalanche Danger in the Ile Alatau Ridge, *Water*, 15, 1438, <https://doi.org/doi.org/10.3390/w15071438>, 2023.
- 75 Dekanová, M., Duchoň, F., Dekan, M., Kyzek, F., and Biskupič, M.: Avalanche forecasting using neural network, In Proceedings of the IEEE ELEKTRO, Mikulov, Czech Republic, <https://doi.org/10.1109/ELEKTRO.2018.8398359>, 2018.
- Dreiseitl, S. and Ohno-Machado, L.: Logistic regression and artificial neural network classification models: a methodology review, *Journal of Biomedical Informatics*, 35, 352–359, [https://doi.org/10.1016/S1532-0464\(03\)00034-0](https://doi.org/10.1016/S1532-0464(03)00034-0), 2002.
- 80 Fromm, R. and Schönberger, C.: Estimating the danger of snow avalanches with a machine learning approach using a comprehensive snow cover model, *Mach. Learn. Appl.*, 10, 100405, <https://doi.org/10.1016/j.mlwa.2022.100405>, 2022.
- LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, *Nature*, 521, 436–444, <https://doi.org/10.1038/nature14539>, 2015.
- Schirmer, M., Lehning, M., and Schweizer, J.: Statistical forecasting of regional avalanche danger using simulated snow-cover data, *J. Glaciol.*, 55, 761–768, <https://doi.org/10.3189/002214309790152429>, 2009.
- 85 Sharma, V., Kumar, S., and Sushil, R.: A neural network model for automated prediction of avalanche danger level., *Nat. Hazards Earth Syst. Sci.*, 23, 2523–2550, <https://doi.org/10.5194/nhess-23-2523-2023>, 2023.