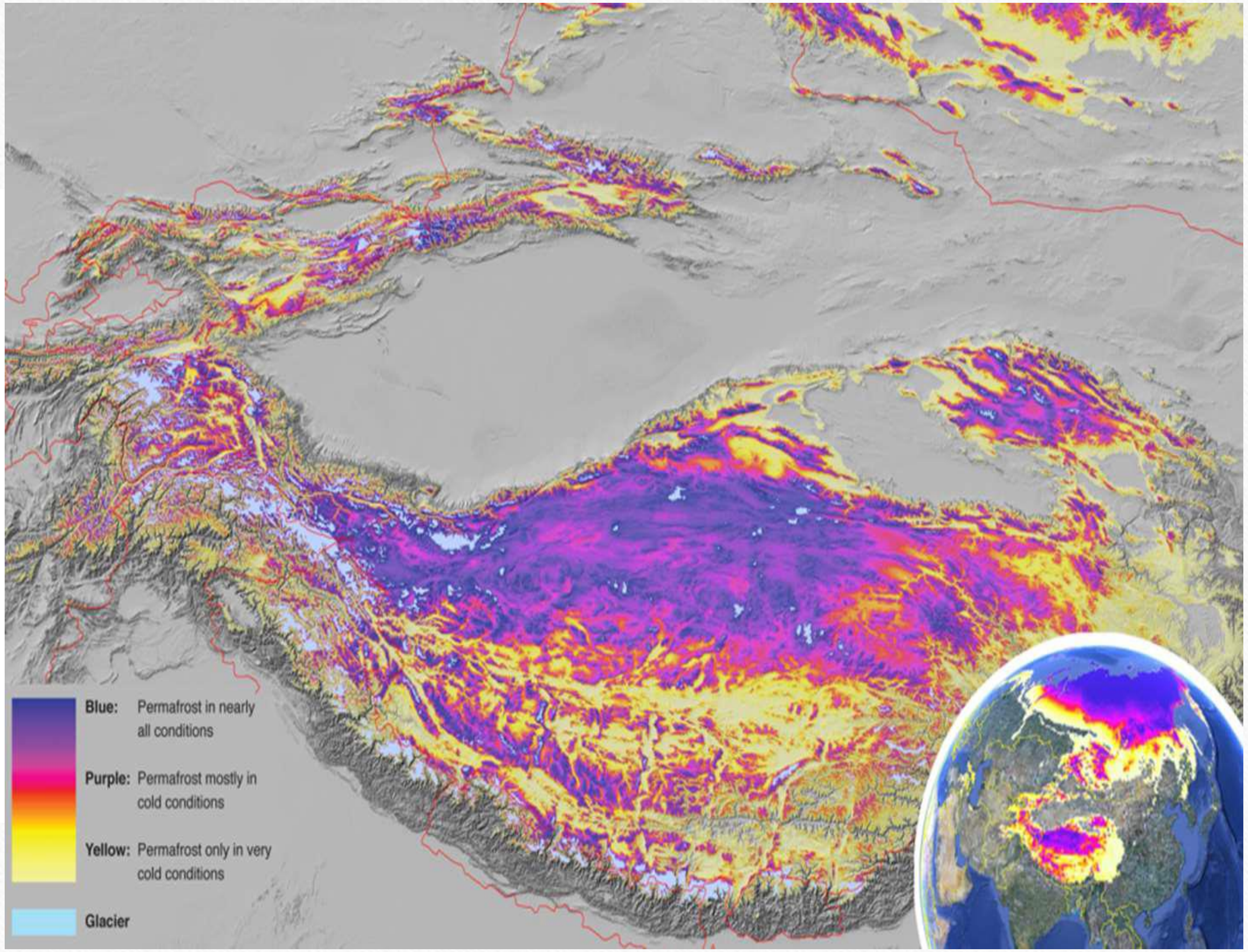
Modeling Permafrost Distribution using Machine Learning models in Solukhumbu, Nepal

Arnab Singh[1], Darwin Rana[1], Sangya Mishra[1], Dibas Shrestha[1]
singharnab78@gmail.com
Central Department of Hydrology and Meteorology, TU
10th EARSeL Workshop on Land Ice and Snow

Estimated extent of alpine permafrost by region[17]

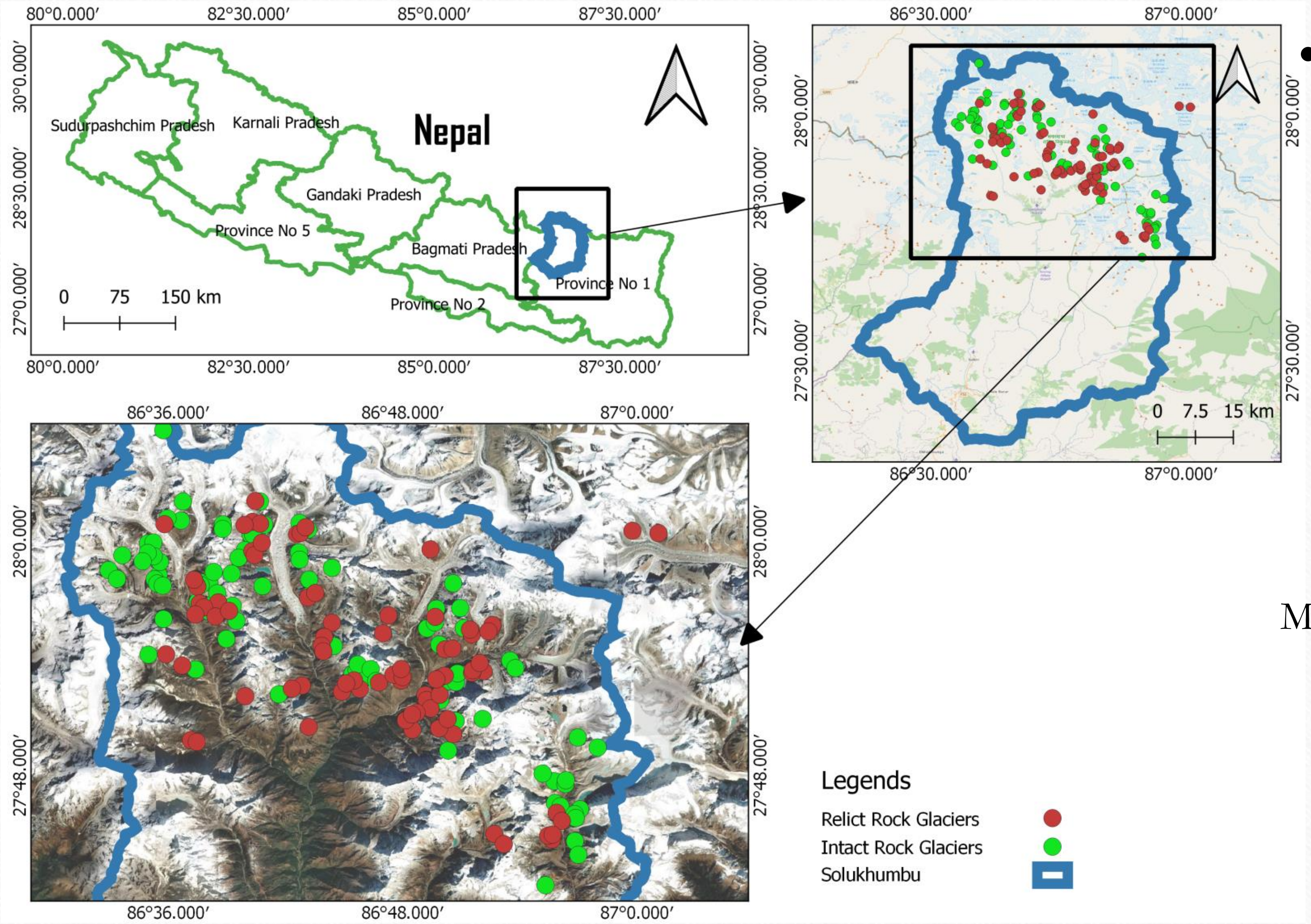
Locality	Area
Qinghai-Tibet Plateau	1,300,000 km^2 (500,000 sq mi)
Khangai-Altai Mountains	1,000,000 km^2 (390,000 sq mi)
Brooks Range	263,000 km^2 (102,000 sq mi)
Siberian Mountains	255,000 km^2 (98,000 sq mi)
Greenland	251,000 km^2 (97,000 sq mi)
Ural Mountains	125,000 km^2 (48,000 sq mi)
Andes	100,000 km^2 (39,000 sq mi)
Rocky Mountains (US and Canada)	100,000 km^2 (39,000 sq mi)
Fennoscandian mountains	75,000 km^2 (29,000 sq mi)
Remaining	<100,000 km^2 (39,000 sq mi)



Credit: Gruber et al, 2012

Assumptions

- Intact Rock Glaciers are an indicator of presence of permafrost (Active and Inactive Rock Glaciers are grouped into Intact Rock Glaciers (Baral and Haq, 2020)
- Relict Rock Glaciers are an indicator of absence of permafrost
- Topo-climatic variables Mean Annual Air Temperature (MAAT) (1960-2018), Potential Incoming Solar Radiation (PISR), Elevation, Aspect and Slope favorable to rock glaciers is also favorable to permafrost

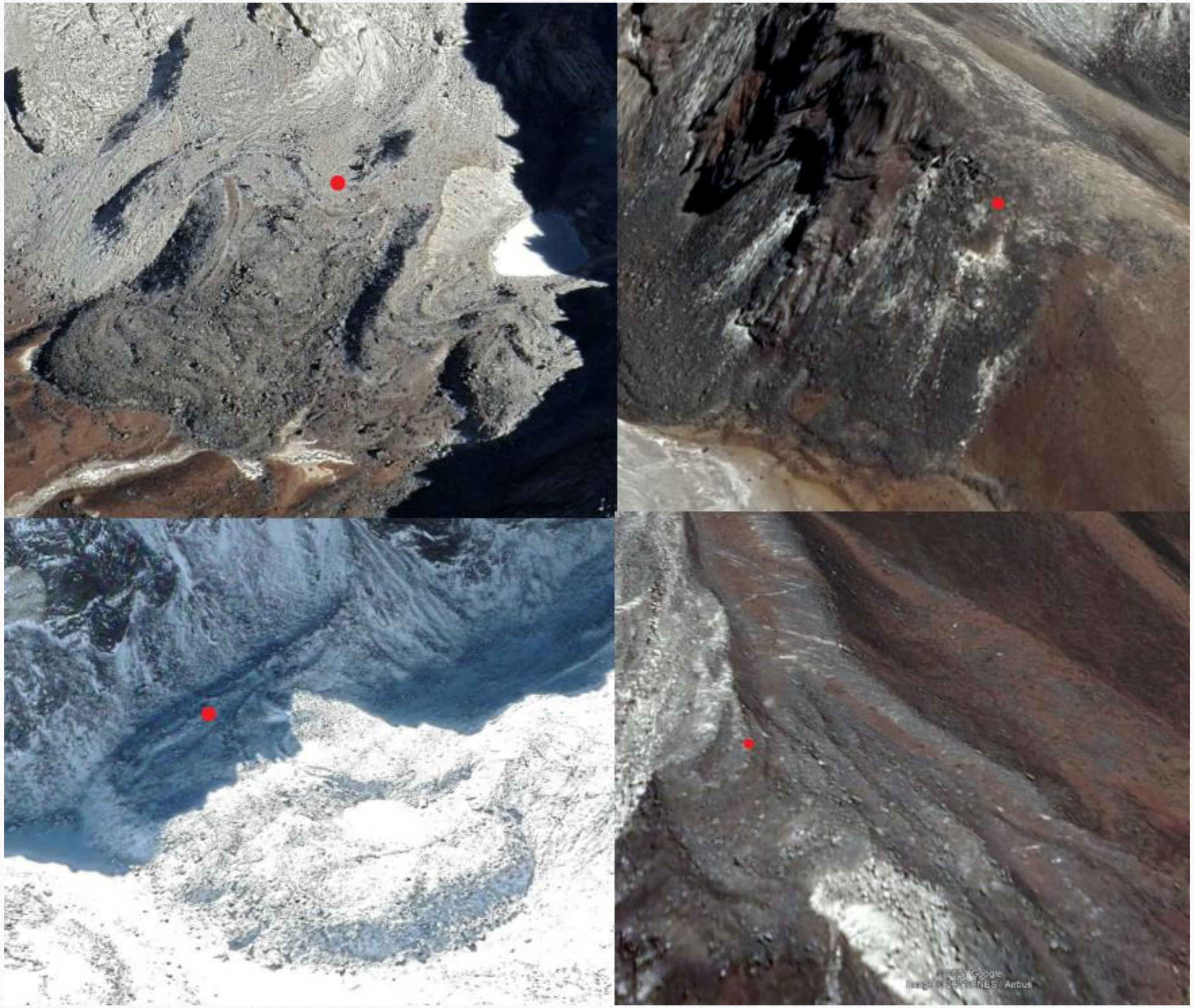


- Regional estimates of permafrost distribution are frequently based on rock glaciers (Haeberli et al, 2006)

Mapping based on Schmid et al, 2015

Rock Glacier	Number
Intact	89
Relict	78
Total	167

Intact Rock Glaciers

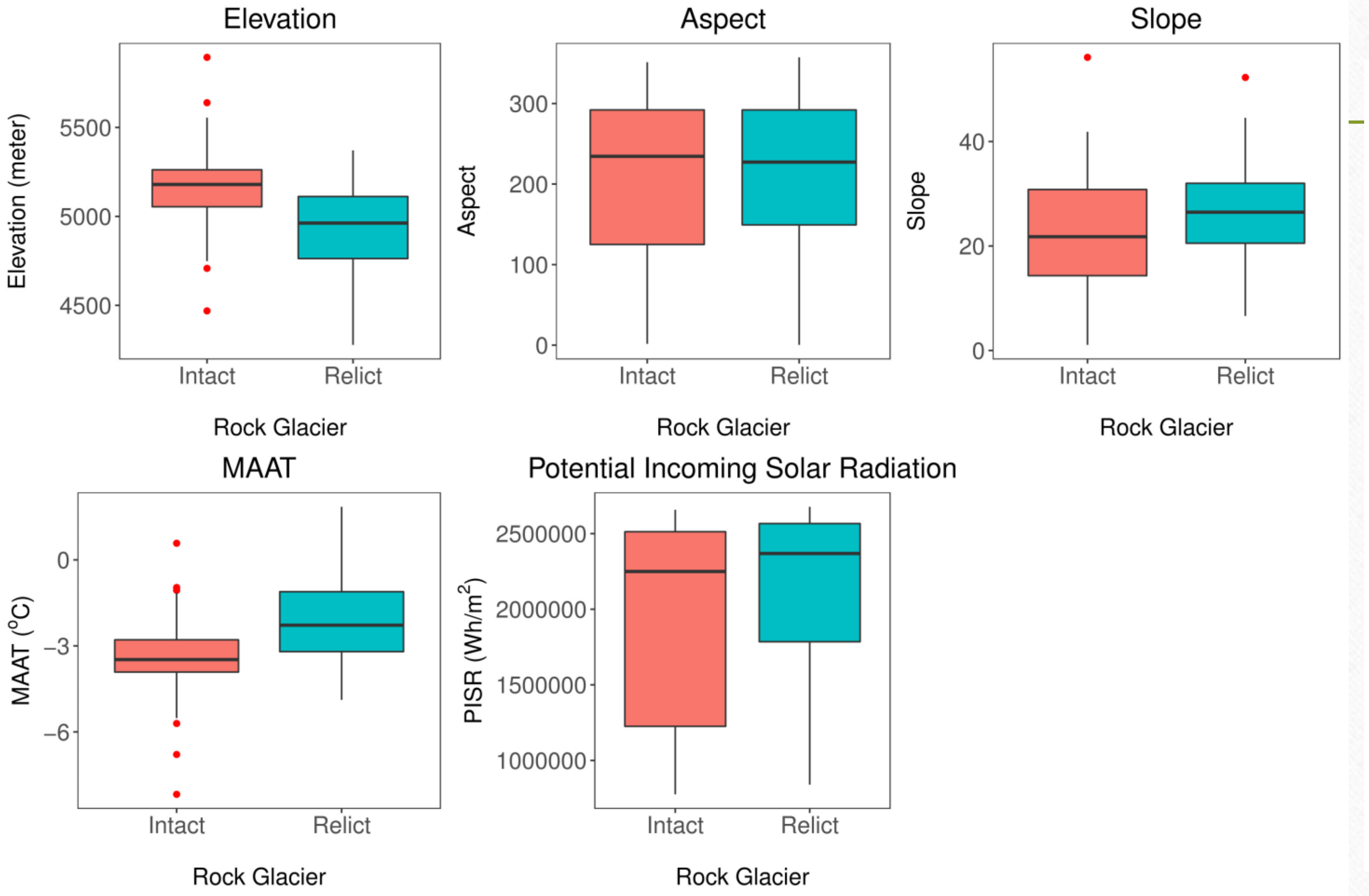


Relict Rock Glaciers

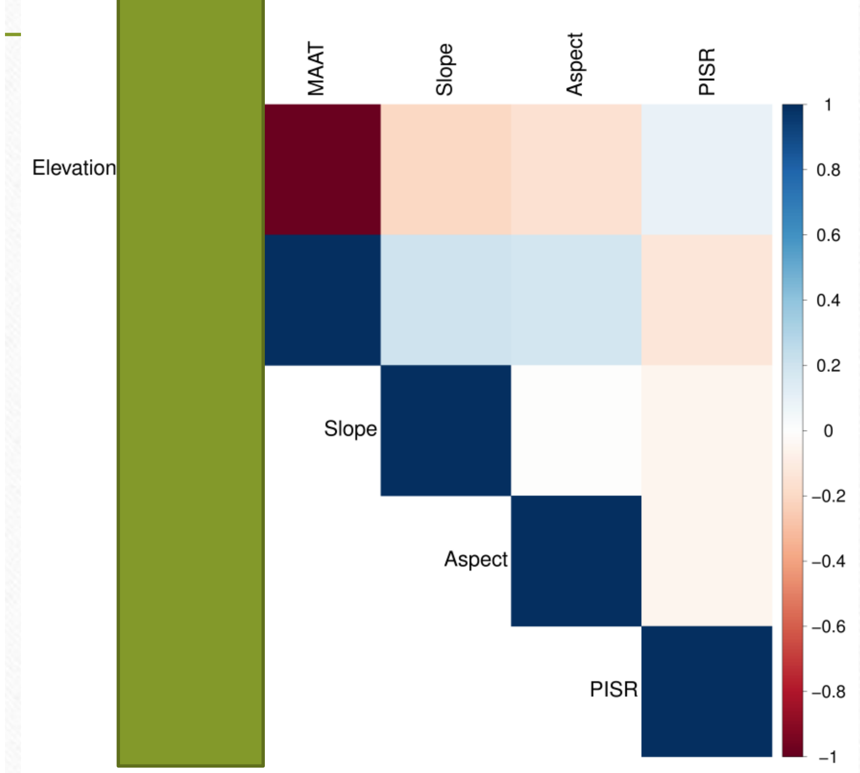
Objective

- Find the region favorable for permafrost for future monitoring
  - For that, data must be free and widely used and available
  - Models must not be too complex

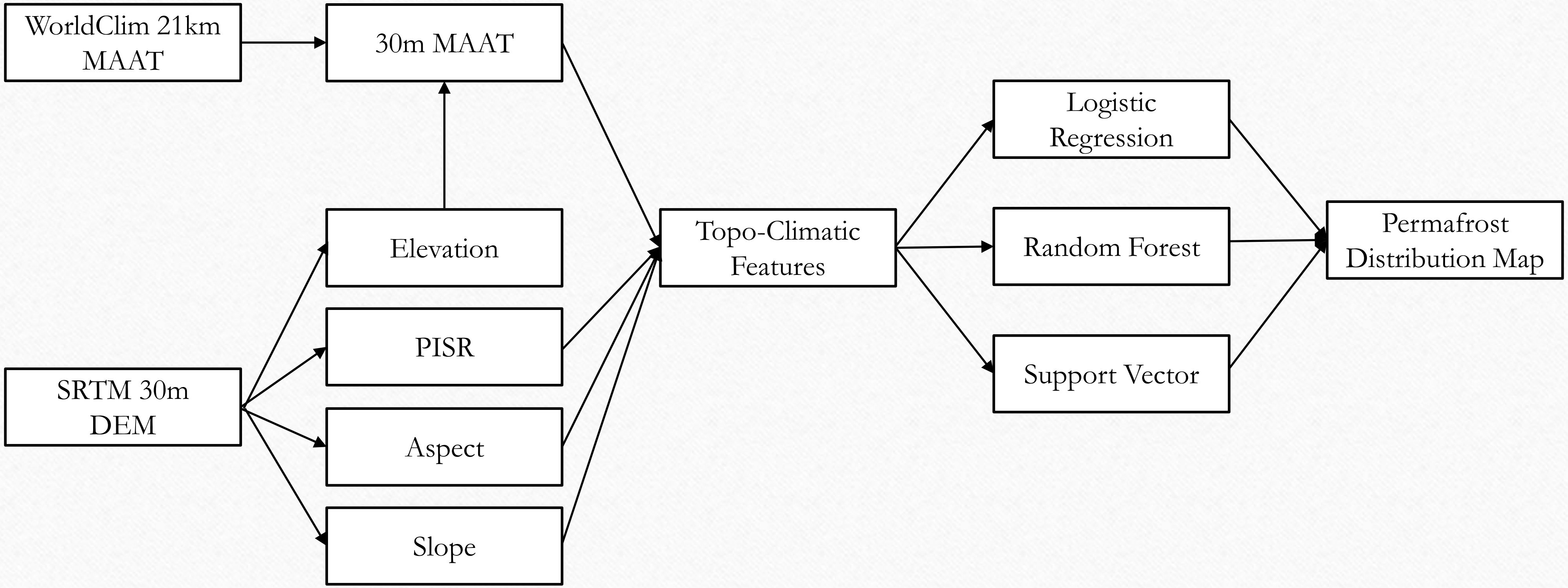
Dataset



Correlation Matrix



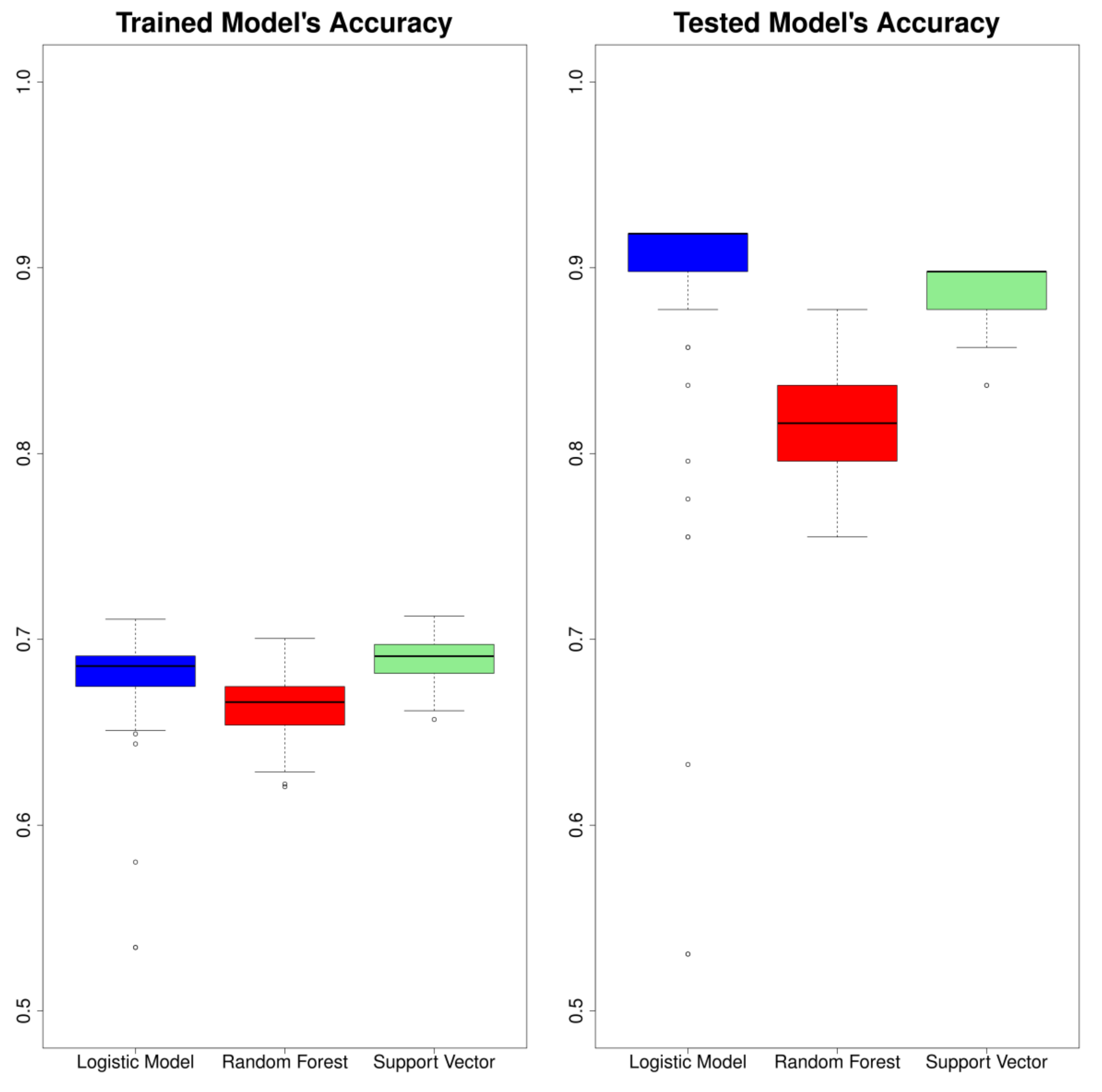
Methodology



Results

- Model's Accuracy

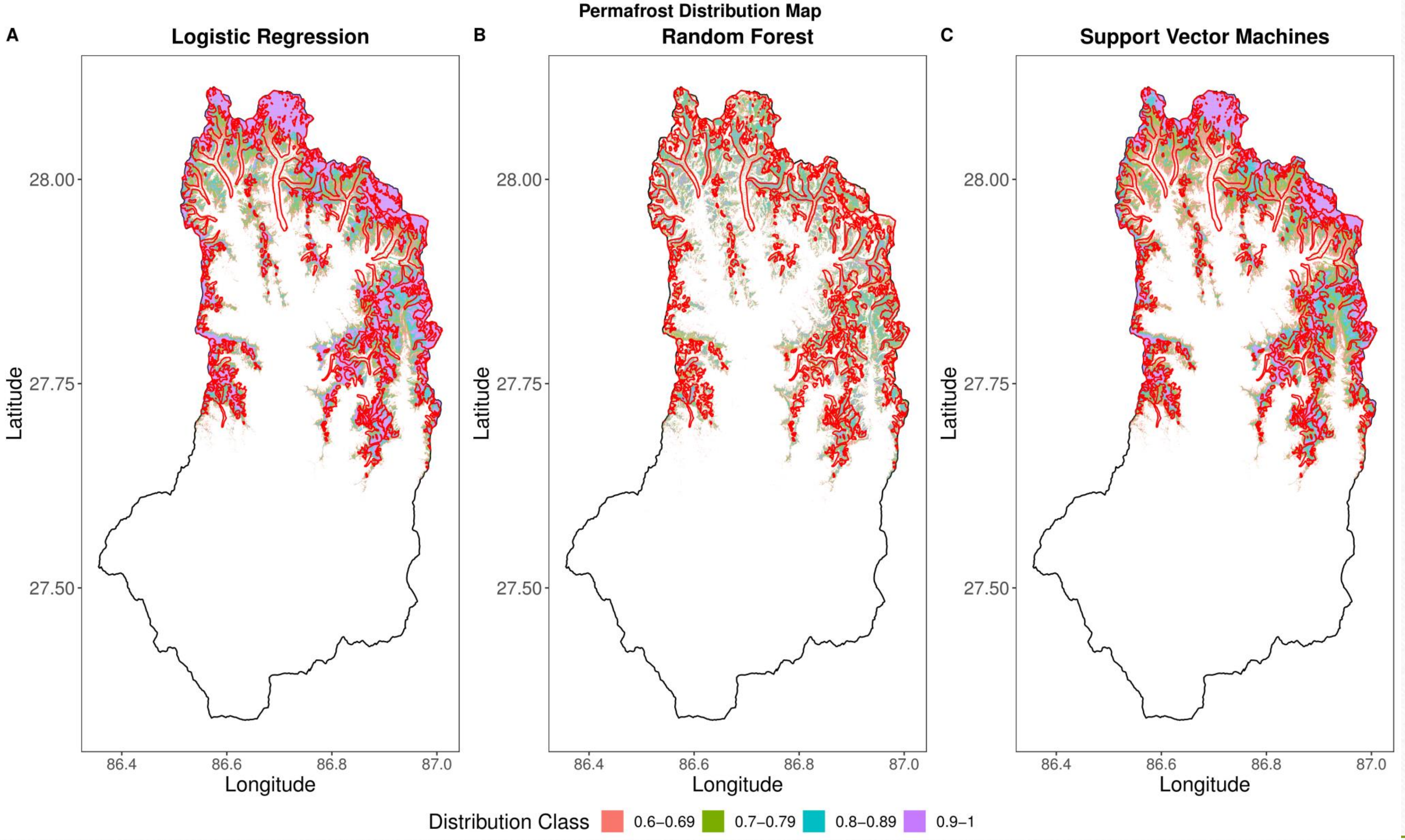
Models	LR	RF	SVM
Accuracy (Training)	0.68	0.66	0.69
Accuracy (Testing)	0.91	0.81	0.89



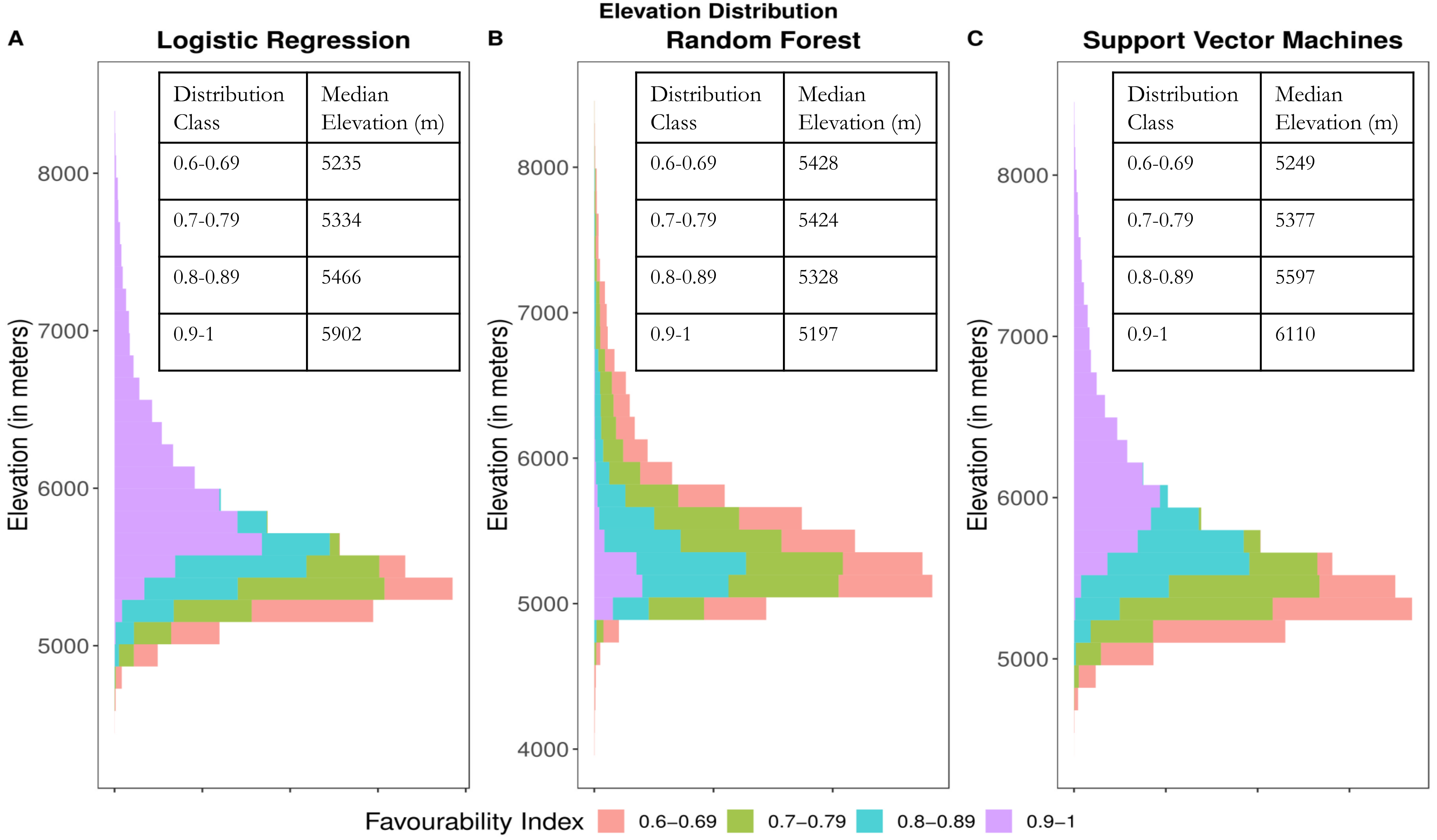
Cutoff Uncertainty

- 0.5 in Haq and Baral, 2019 and Lewkowicz and Edine, 2009
- 0.6 in Sattler et al., 2016

	LR		**RF**		**SVM**	
Uncertainty Interval	Miss Classification Error	Uncertainty Percent	Miss Classification Error	Uncertainty Percent	Miss Classification Error	Uncertainty Percent
0.49 - 0.51	0.27	0.04	0.29	0.014	0.282	0.0144
0.48 - 0.52	0.25	0.08	0.286	0.026	0.2758	0.0322
0.47 - 0.53	0.22	0.11	0.276	0.0488	0.268	0.059
0.46 - 0.54	0.19	0.172	0.271	0.062	0.2612	0.0922
0.45 - 0.55	0.18	0.2	0.267	0.076	0.251	0.122
0.44 - 0.56	0.18	0.24	0.258	0.1	0.23	0.149
0.43 - 0.57	0.17	0.25	0.256	0.108	0.22	0.181
0.42 - 0.58	0.15	0.29	0.252	0.128	0.208	0.2094
0.41 - 0.59	0.14	0.32	0.245	0.145	0.187	0.262
0.4 - 0.6	0.12	0.35	0.243	0.171	0.167	0.306
0.39 - 0.61	0.12	0.358	0.239	0.172	0.149	0.34
0.38 - 0.62	0.12	0.37	0.238	0.192	0.135	0.363
0.37 - 0.63	0.11	0.39	0.231	0.226	0.12	0.381
0.36 - 0.64	0.1	0.41	0.224	0.249	0.106	0.4012
0.35 - 0.65	0.09	0.45	0.225	0.257	0.095	0.42
0.34 - 0.66	0.093	0.48	0.217	0.283	0.085	0.447



Permafrost Distribution Map
A
Logistic Regression
B
Random Forest
C
Support Vector Machines
Latitude
28.00
27.75
27.50
Longitude
86.4
86.6
86.8
87.0
Distribution Class
0.6–0.69
0.7–0.79
0.8–0.89
0.9–1



A — Logistic Regression

Distribution Class	Median Elevation (m)
0.6-0.69	5235
0.7-0.79	5334
0.8-0.89	5466
0.9-1	5902

B — Random Forest

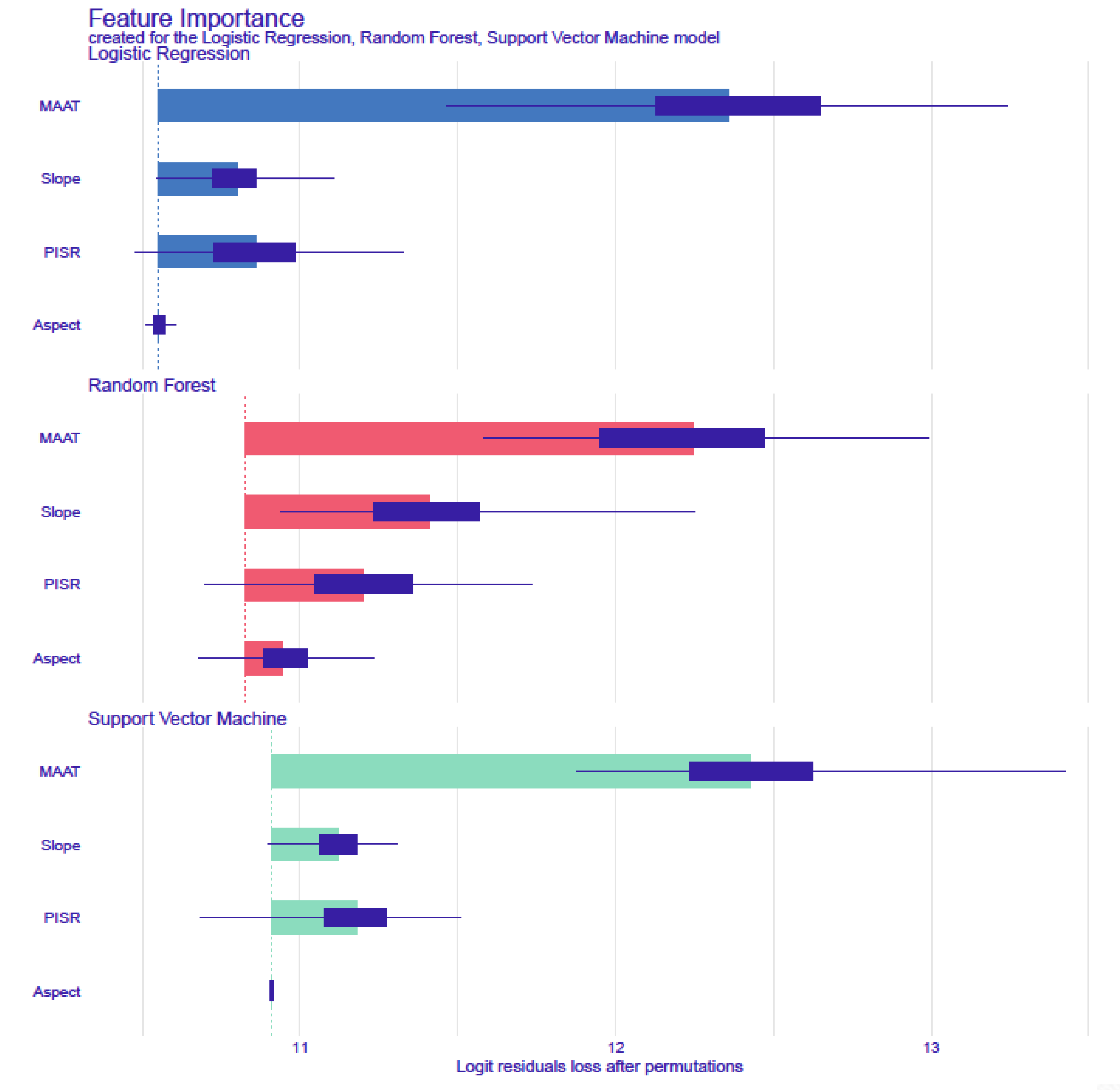
Distribution Class	Median Elevation (m)
0.6-0.69	5428
0.7-0.79	5424
0.8-0.89	5328
0.9-1	5197

C — Support Vector Machines

Distribution Class	Median Elevation (m)
0.6-0.69	5249
0.7-0.79	5377
0.8-0.89	5597
0.9-1	6110

Sensitivity Analysis

- To understand how model's are predicting
- Which features are important and which are not



Conclusion

- Logistic Regression performed the best, followed by Support Vector Machines and Random Forest
- Random Forest' result vary wildly with other two
- No North-South distinction of permafrost zones; necessity to include precipitation in future permafrost distribution modeling