Comparison of two spatial interpolation methods for estimating snow distribution in an alpine region

Demontis F., Gabellani S., Rudari R.,

CIMA Research Foundation, ITALY

francesco.demontis@cimafoundation.org
www.cimafoundation.org

RESEARCH FOUNDATION cima
OBSERVE TO PREDICT
PREDICT TO PREVENT

CENTRO FUNZIONALE
REGIONE AUTONOMA VALLE D'AOSTA

ARPA Valle d'Aosta

Outlines

RESEARCH FOUNDATION cima

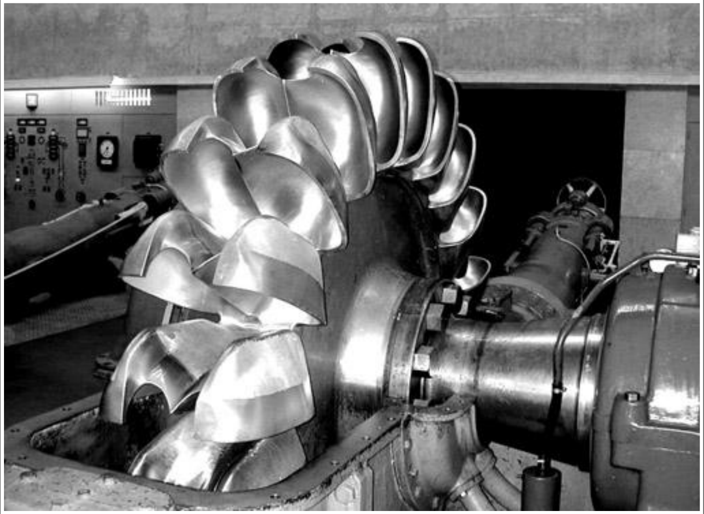
OBSERVE TO PREDICT
PREDICT TO PREVENT

- *Introduction*
- *Study Area & Data Set*
- *Set of independent variables*
- *Models & Comparisons*
- *Results*
- *Conclusions*

Introduction

This work presents a comparison between two different statistical techniques for the interpolation of snow depth measurements.

Multiple Linear Regression (MLR) and Classification and Regression Tree (CART) models are compared to evaluate both the best model and the best combination of independent variables.



Motivation of the work

- Large Data sets, mainly snow depth manual measurements;
- Few studies over Italian Alps about interpolation techniques applied to snow data sets;
- The most part of studies are carried out over restricted flat areas, using measurements dedicated for this purpose, to evaluate Snow Water Equivalent (fine resolution) or for estimating snow cover extent over large areas using coarser spatial resolution.

What have we done?

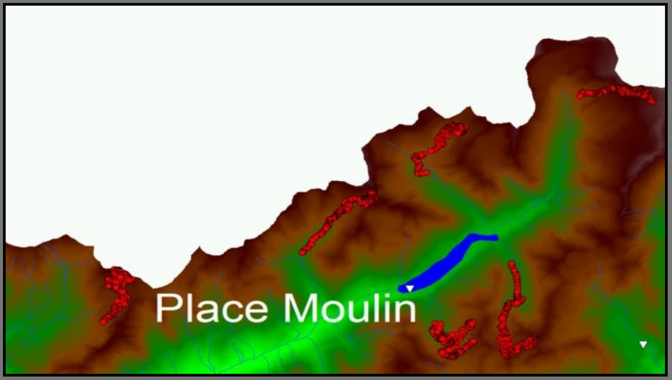
- Although these spatial interpolation methods have already been used in several work, we tried to use data sets , not exactly taken for snow depth estimate purpose, developing regression methods over complex terrain. In fact, an high variability of morphological and climatological parameters can be noted both at the small scale and a regional scale;

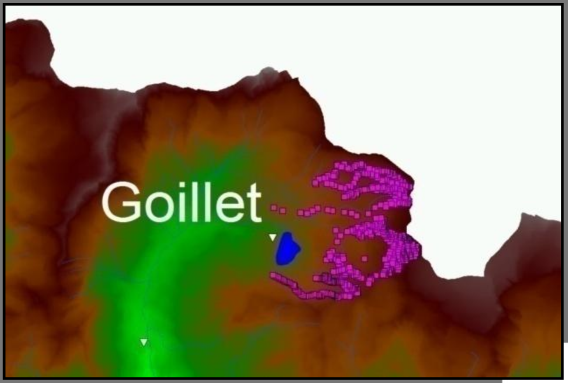
- We try to extend interpolated field range at regional scale, using real time telemetry data.

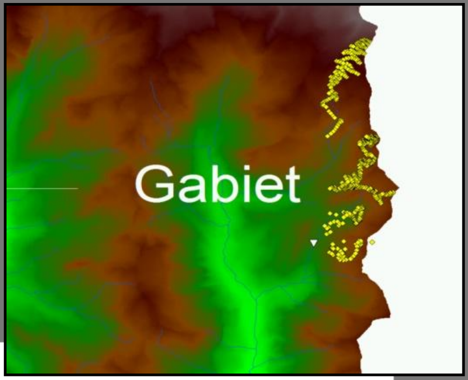
Why ?

To improve hydrological modelling and discharge forecast

Study Area

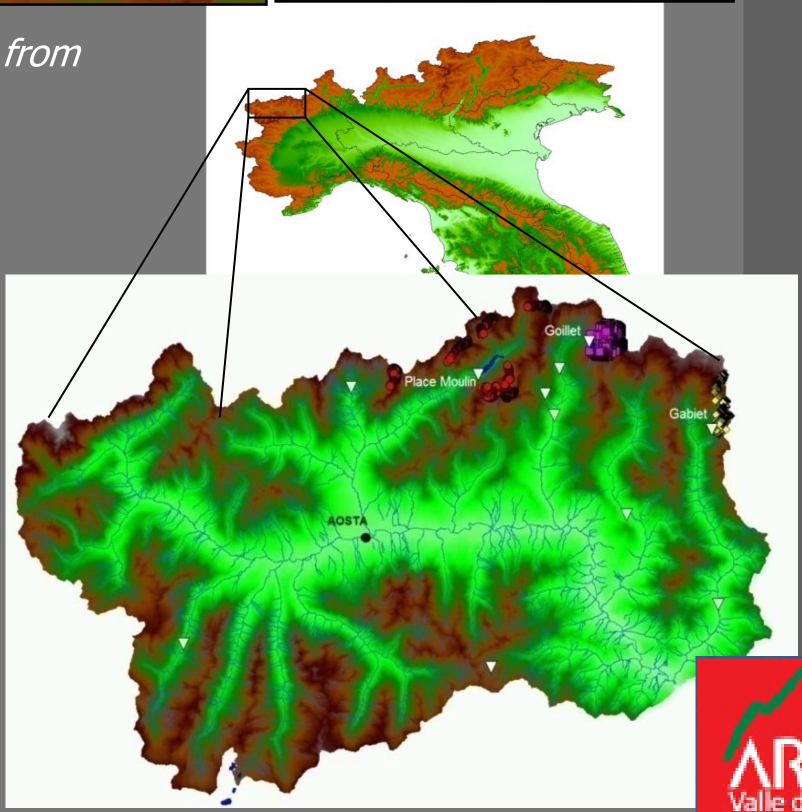






Results of manual measurements have been performed from Regional Agency for Protection of Environment

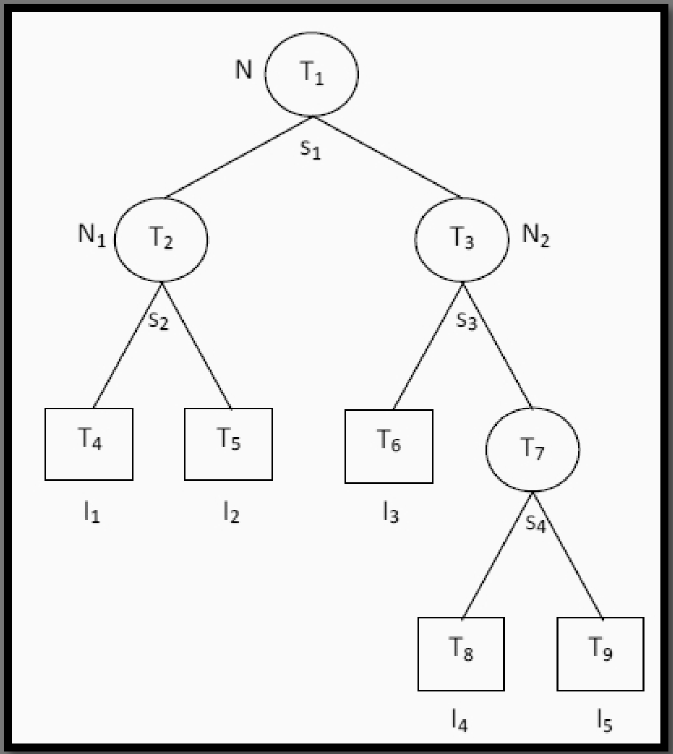
Snow Campaign	n° Courses	n° samples	Elevation range [m]		
Valpelline 2007					
			Min	Mean	Max
23-24 May 2007	6	883	2230	2900	3610
Valpelline 2008					
			Min	Mean	Max
24 June 2008	4	442	2720	3200	3900
Gabiet					
			Min	Mean	Max
28-apr-07	5	560	2370	2870	3750
Goillet					
			Min	Mean	Max
20-apr-07	8	522	2550	3050	3450



Models

We are looking for an empirical relationship between snow depth and some morphological parameters using:

Classification and Regression Tree (CART)



Multiple Linear Regression (MLR)

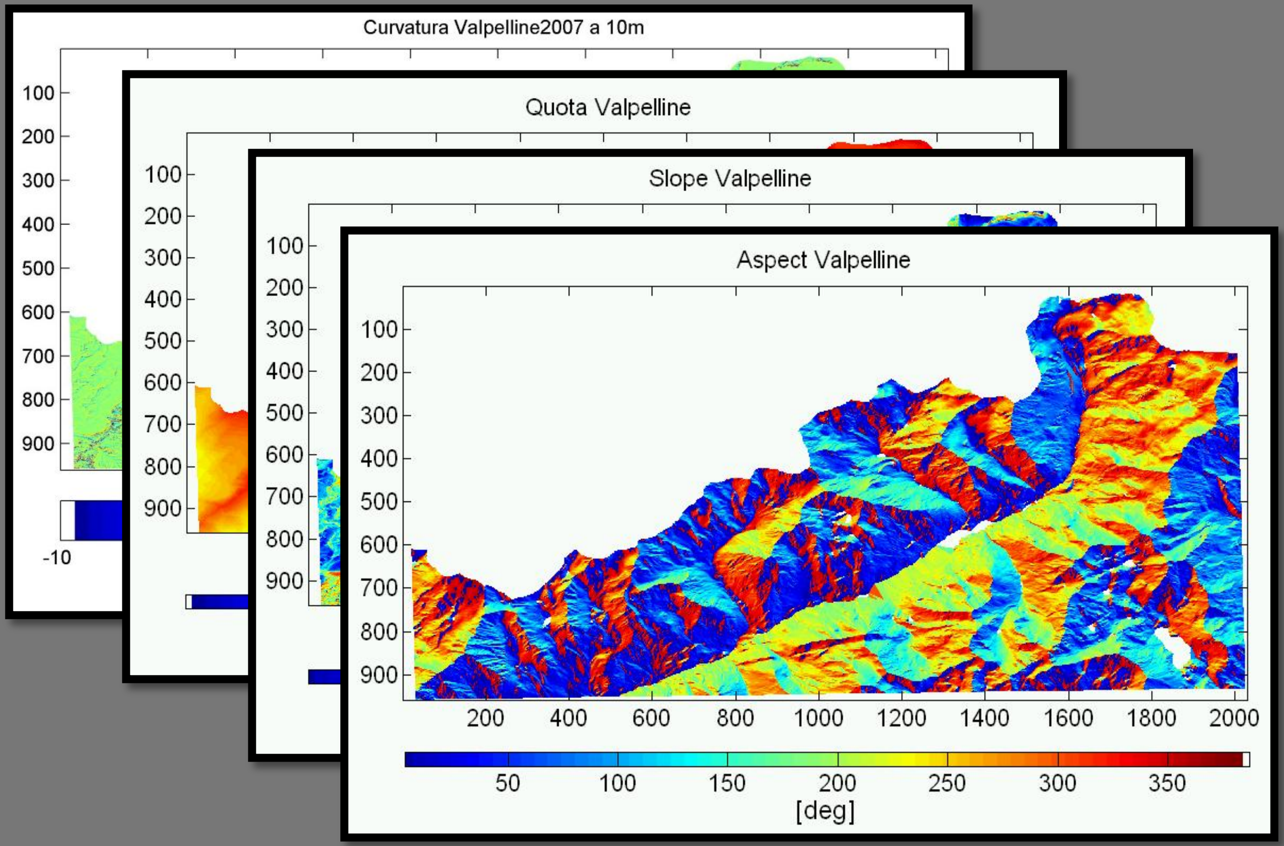
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$$

- Multiple linear regression models predict a value of the Y variable, snow depth in this case, given known values of the X variables using a regression equation.

- CART is a recursive partitioning method, builds classification and regression trees for predicting continuous dependent variables (regression) and categorical predictor variables (classification). Regression attempts to predict the values of a continuous variable from one or more continuous predictor variables.

Independent variables

RESEARCH FOUNDATION cima
OBSERVE TO PREDICT
PREDICT TO PREVENT



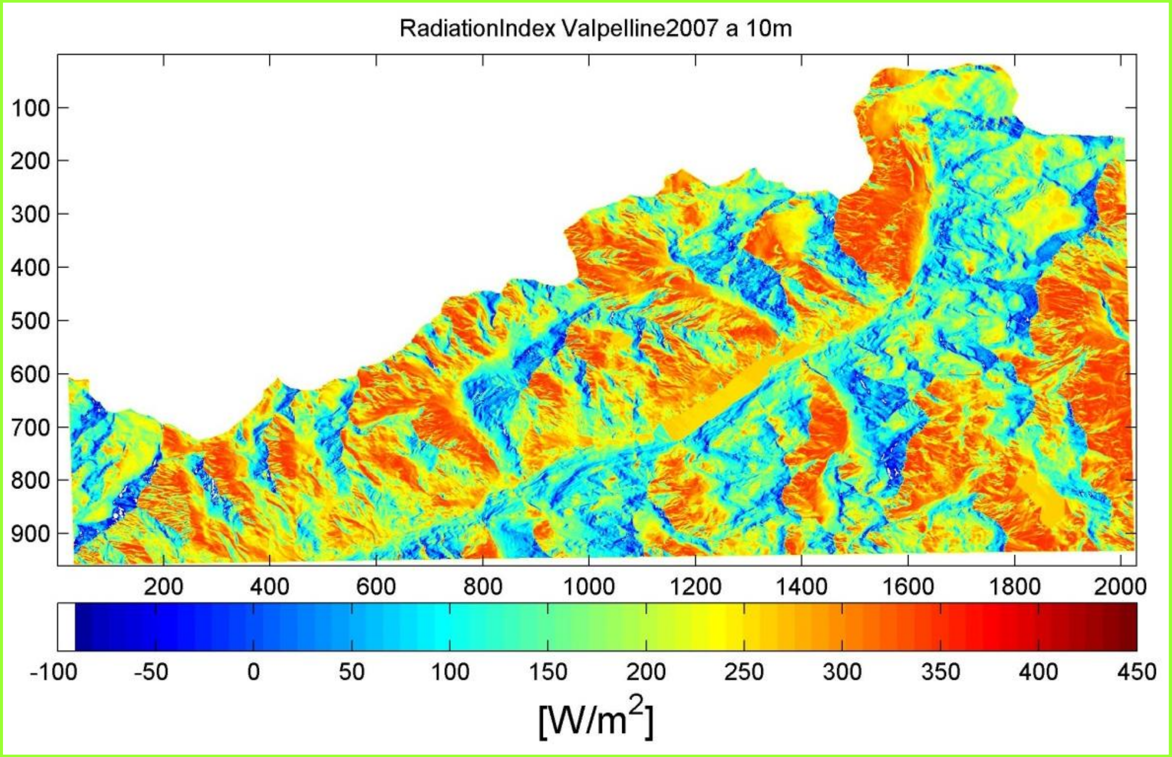
8 Predictors are considered at the start of simulation

- *Elevation*
- *Slope*
- *Aspect*
- *Grad in North and East direction*
- *Concavity/Convexity*
- *Radiation Index*
- *Mean UpwindSlope*

Individuation of best spatial interpolation technique for the evaluation of snow depth distribution on alpine basin

Radiation Index

Radiation Index has been developed following *Yang et al. 2001*, 2006, for radiance components estimation, and *Dozier and Frew, 1989* for direct, diffuse and scattered radiance in complex terrain.

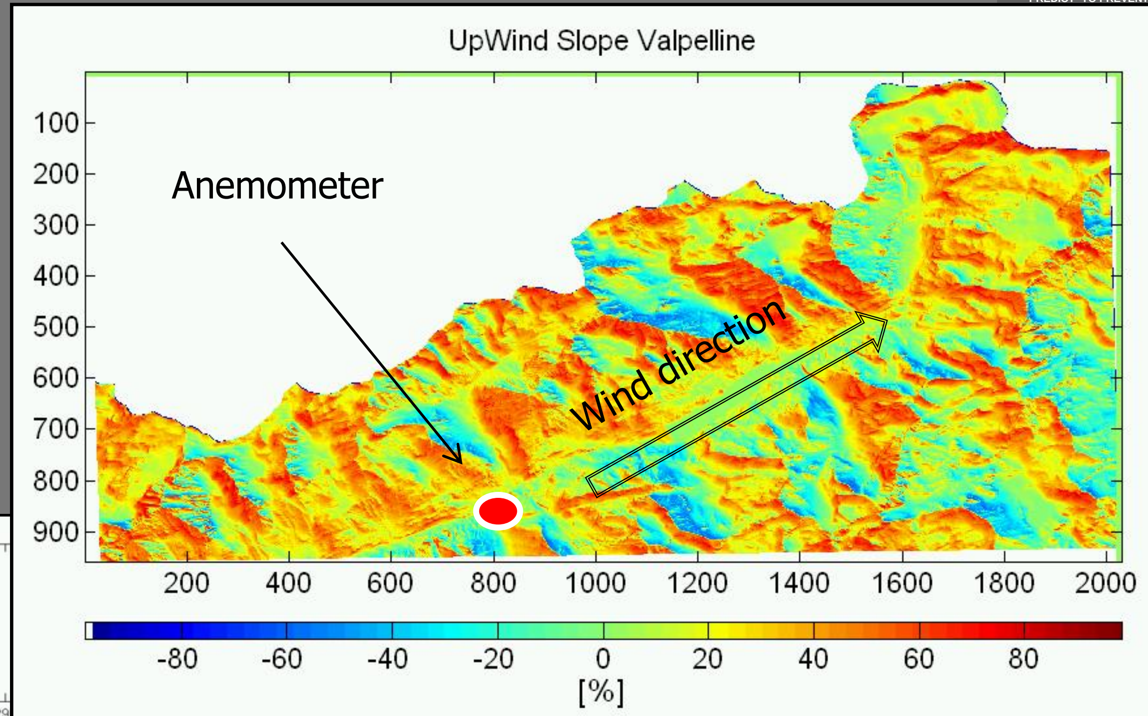


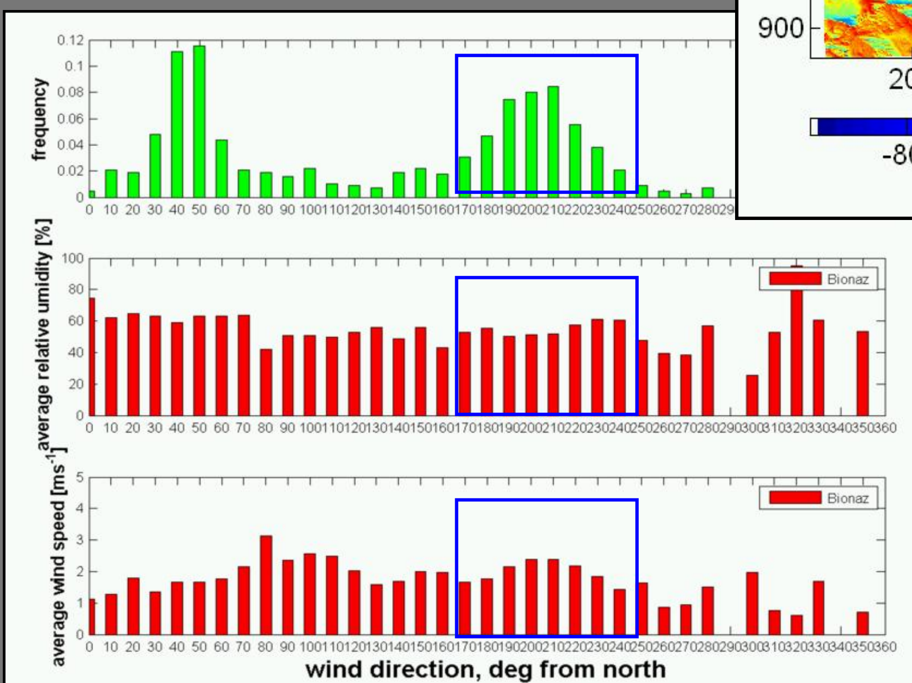
Radiation index shows a mean of shortwave radiance on clear sky hypothesis based on 30 previous day as regards the day of the measurements.

UpWind Slope Index

The wind effects are evaluated using the index proposed by Winstral 2002, based on maximum upwind slopes relative to seasonally dominant winds.

The aim of maximum upwind slope is to quantify the extent of shelter or exposure provided by the terrain upwind of each pixel



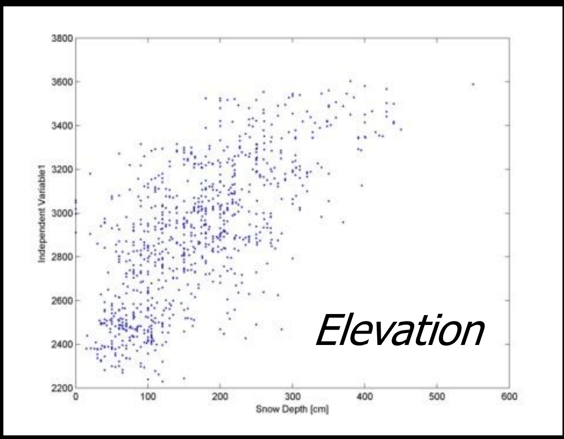


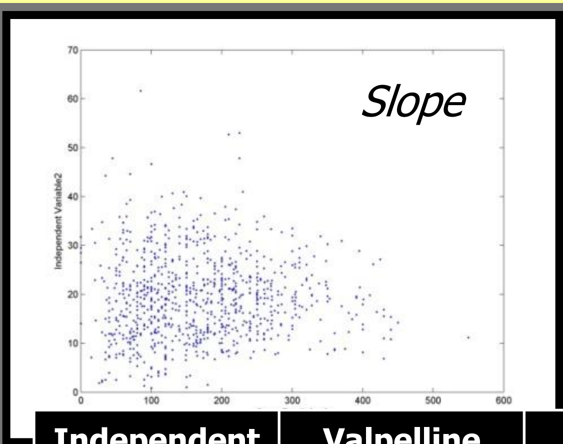
Up wind terrain consist of a pie-shaped area , chosen 60°, centered on the prevailing monthly wind direction.

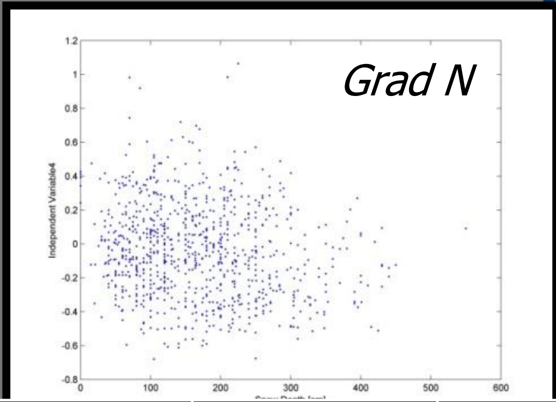
Upwind index is a monthly mean parameters taking into account a mean maximum upwind slope for each pixel.

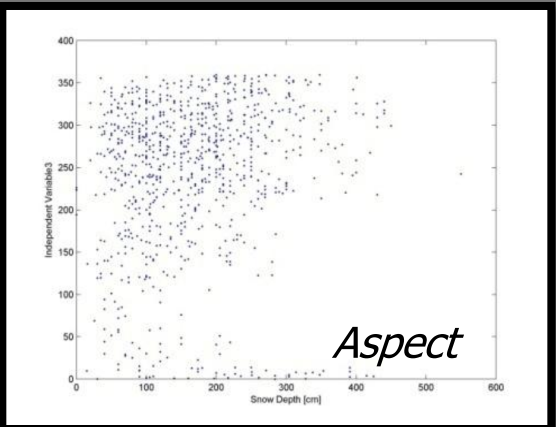
Predictors relationship

Valpelline 2007 snow campaign scatterplots





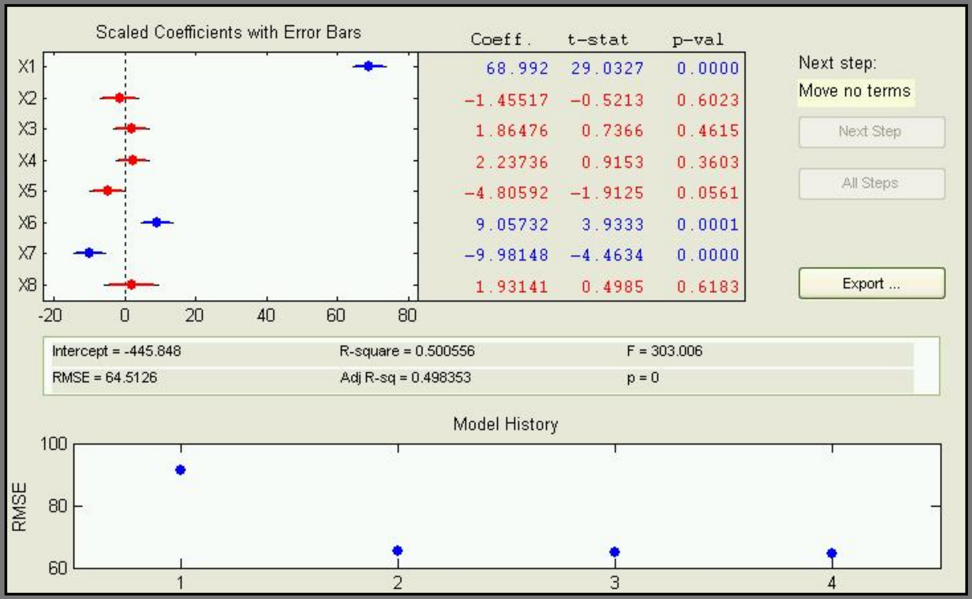




Independent Variable Correlation	Valpelline 2007	Valpelline 2008	Gabiet	Goillet
Elevation	0.6988	0.4643	0.0152	0.2596
Slope	0.0050	0.2738	0.1018	0.1845
Aspect	0.1126	-0.0337	-0.1053	0.0927
Grad North	-0.1711	-0.2862	0.0422	-0.0907
Grad East	0.0190	0.0298	-0.0076	0.2562
Rad Index	-0.0430	-0.1940	-0.0306	-0.2606
Upwind Slope	0.1629	0.2438	0.0500	0.1489
Concavity	-0.0107	0.0351	0.1837	0.1274

Best predictors combination MLR

Stepwise regression *for MLR has been used to test the statistical significance of the selected predictors.*

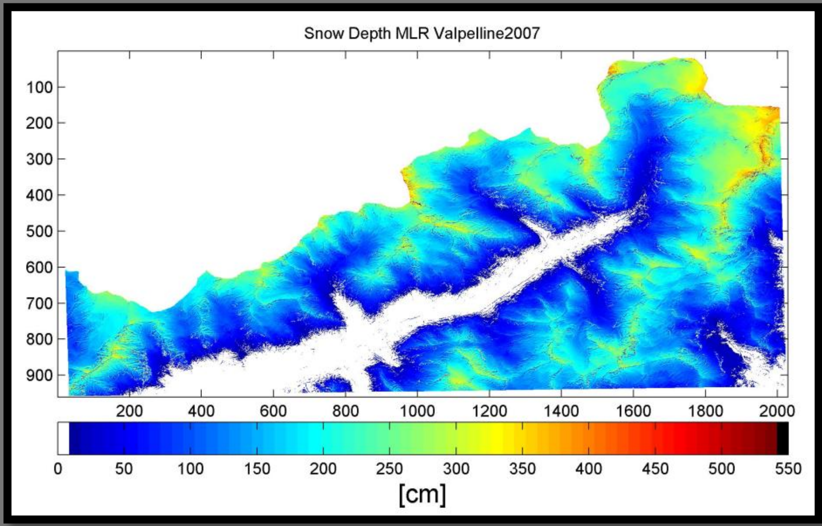


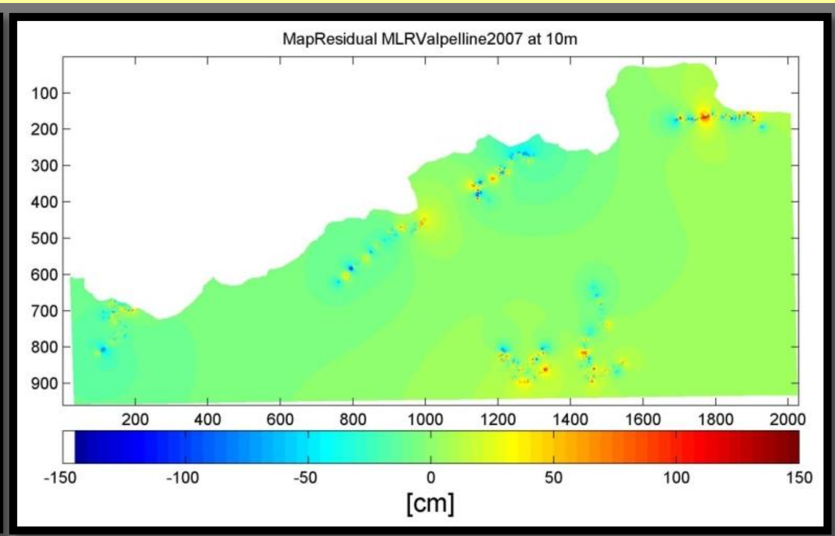
In the stepwise regression the choice of predictive variable is carried out by an automatic procedure based of P-value significance.

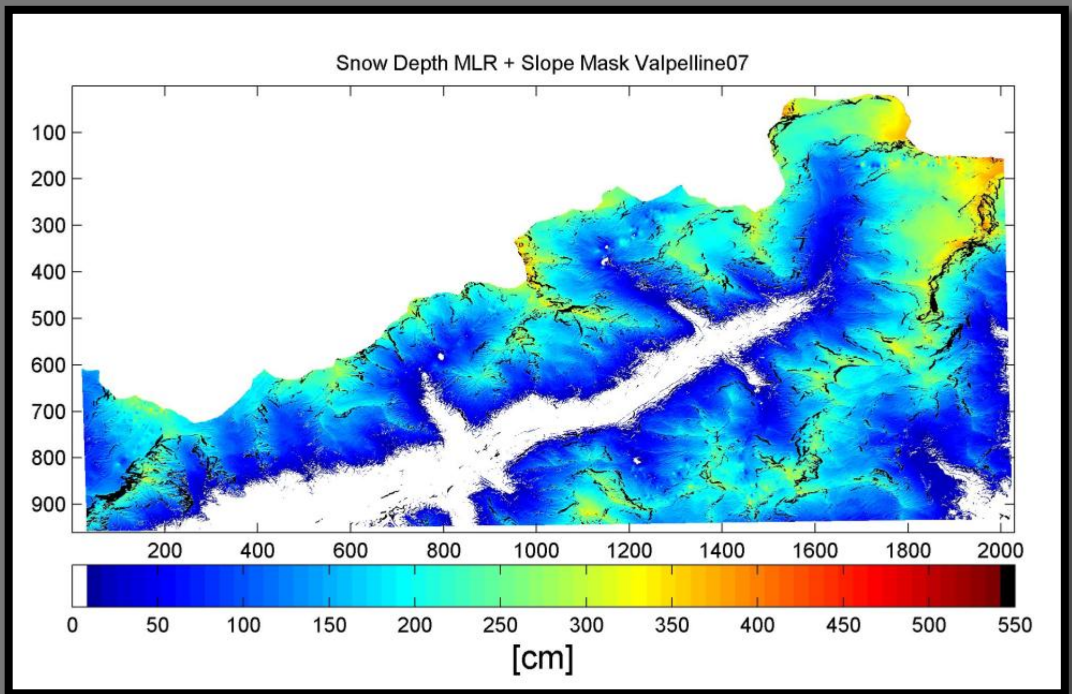
Forward selection and backward selection have been evaluated.

The former starts with no variables in the model, trying the variables to include one by one, the latter starts with all possible predictors in the model, deleting one by one the variables that are not significant

Snow Depth Distribution with MLR







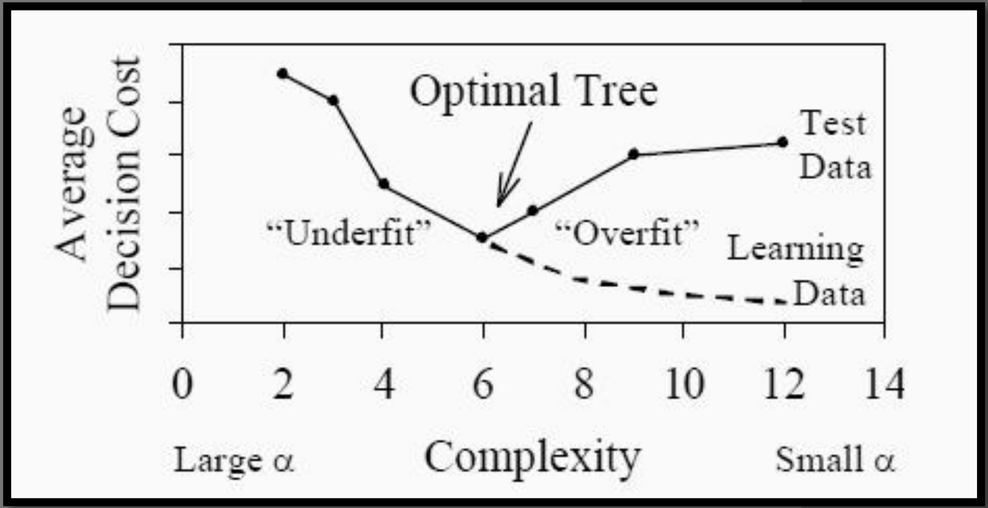
Snow depth distribuited map has been masked with slope over surfaces upwards of 60 degrees and added with IDW interpolated residuals.

RESEARCH FOUNDATION cima
OBSERVE TO PREDICT
PREDICT TO PREVENT

Best predictors combination CART

Cross validation procedure for CART has been used to test the statistical significance of the selected predictors.

Snow depth samples, have been randomly divided into 10 subset for the crossvalidation procedure. Nine of the subsets were used to grow a tree whereas the remaining subset were used to test model cost function through the tree.



Minimum model residual variance (cost) have been used to evaluate the range of optimal number of terminal nodes in the final pruned tree.

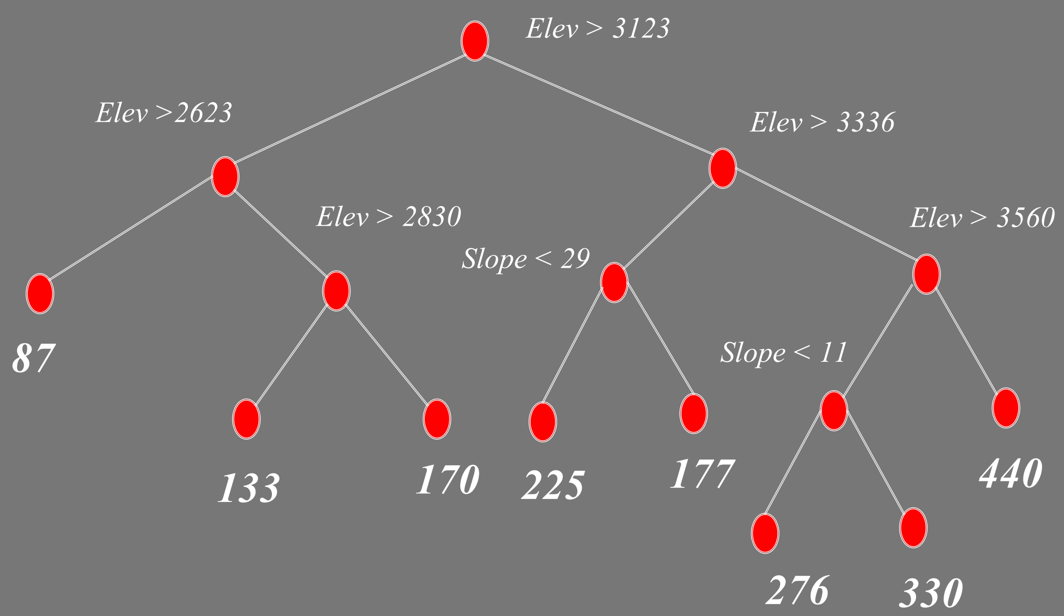
In this work all tree are pruned at the same pruning level considering **8 terminal nodes** as good choice between a mis-classification probability and tree complexity.

To perform a more robust estimation, optimal predictors combination were selected as results of **100 tenfold crossvalidation procedures**.

RESEARCH FOUNDATION
OBSERVE TO PREDICT
PREDICT TO PREVENT

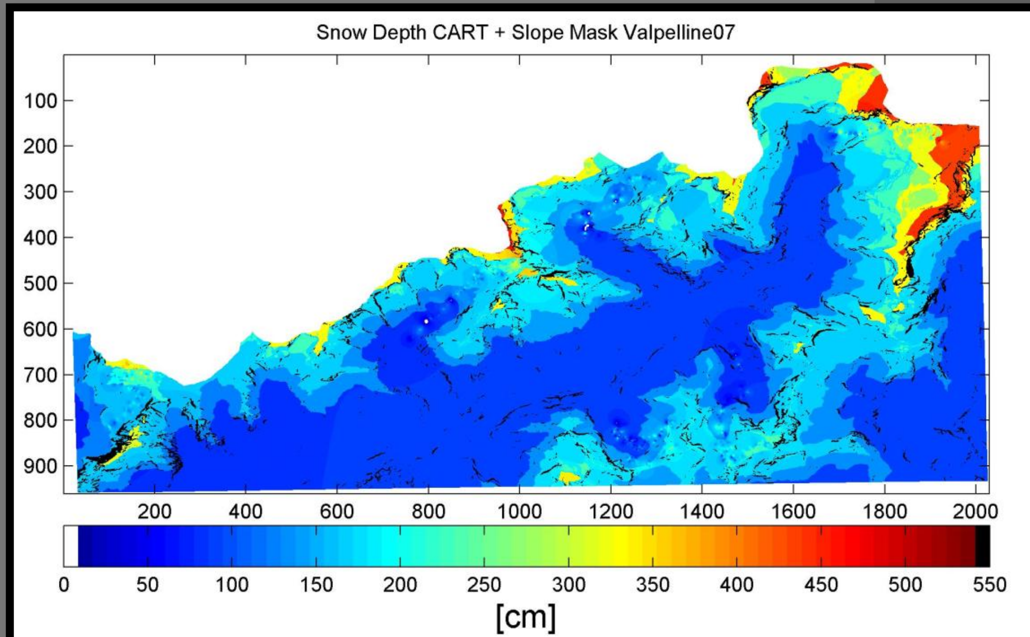
Snow Depth Distribution with CART

RESEARCH FOUNDATION cima
OBSERVE TO PREDICT
PREDICT TO PREVENT

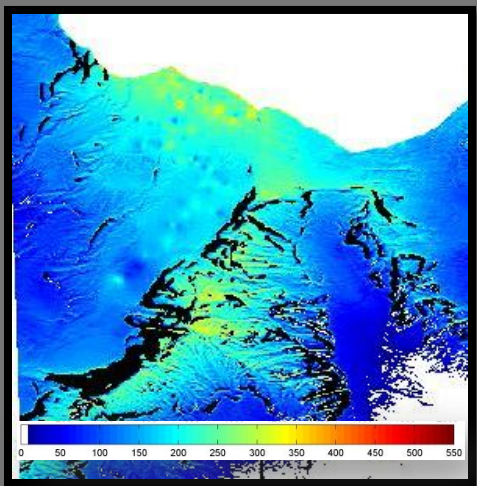


Elevation and Slope are the predictors selected for Valpelline 2007.

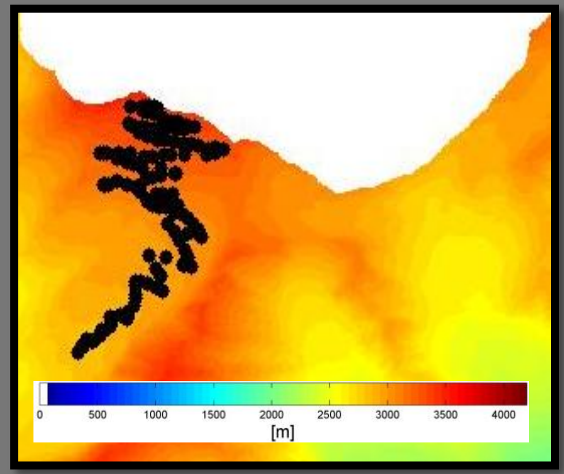
Snow depth distribution obtained with Cart are also affected by relevant correlation with elevation data.



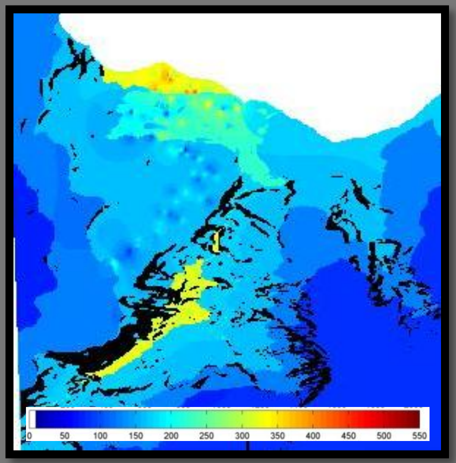
Comparison Snow Depth distributions



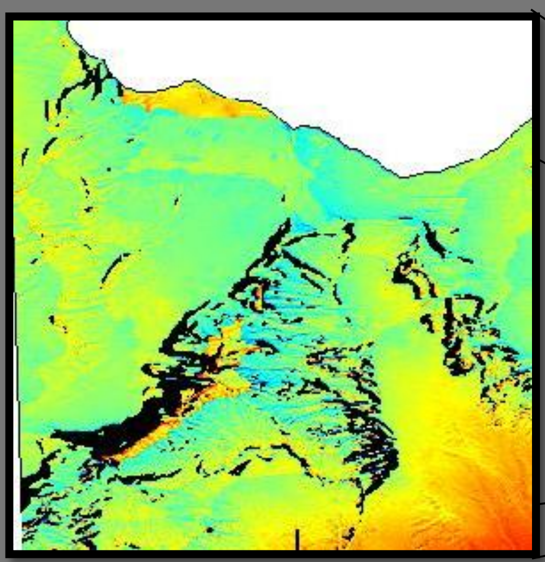
MLR



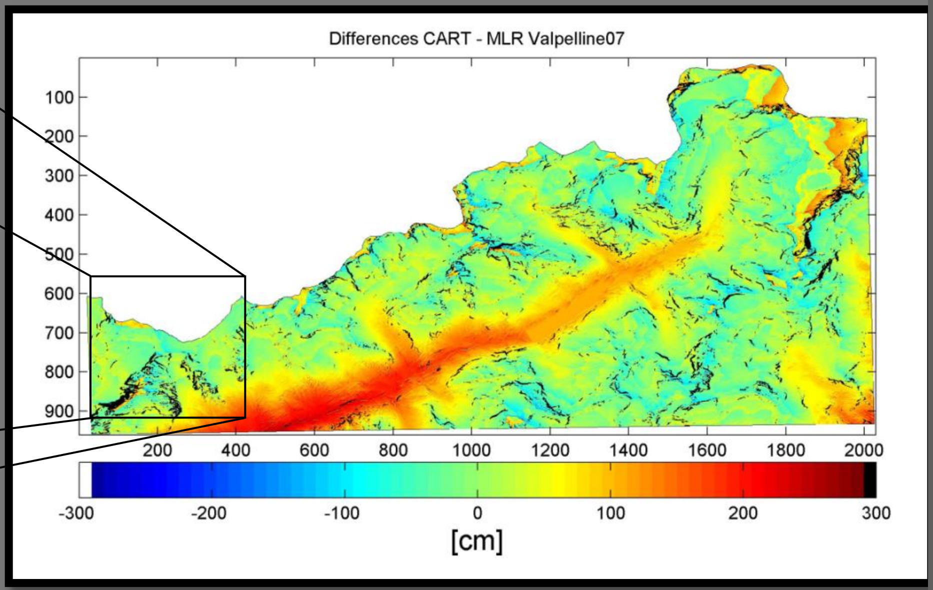
Digital Elevation Model



CART



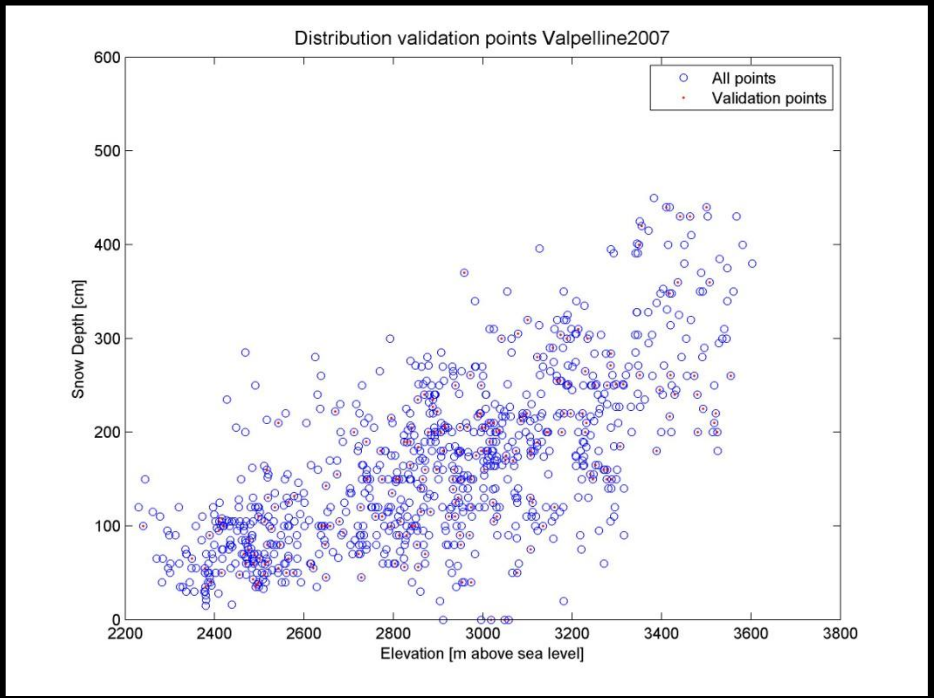
CART - MLR



Validation Procedure

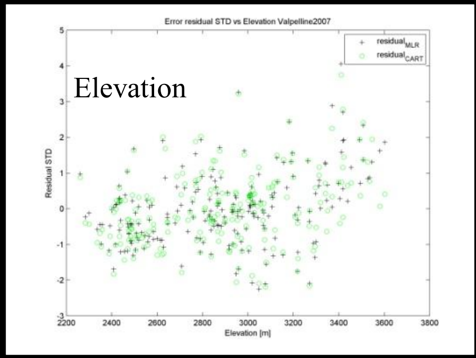
For testing the predictive skills of the models data set were partitioned in two subsamples, about three quarter of whole data-set have been used for fitting the models and the remaining have been tested to evaluate the goodness of fit.

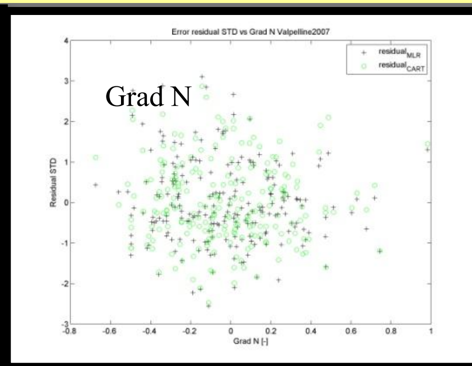
	Valpelline 2007	Valpelline 2008	Goillet	Gabiet
Mean data-set	166.90	208.35	130.70	97.34
Mean val data-sets	167.02	208.67	131.61	97.76
mean val MLR	167.52	208.52	128.73	97.92
mean val CART	167.45	208.47	129.37	97.56
std dev. data-sets	5.42	6.37	5.50	3.89
std dev. MLR	6.37	5.88	3.97	2.81
std dev. CART	5.73	6.06	4.23	2.94

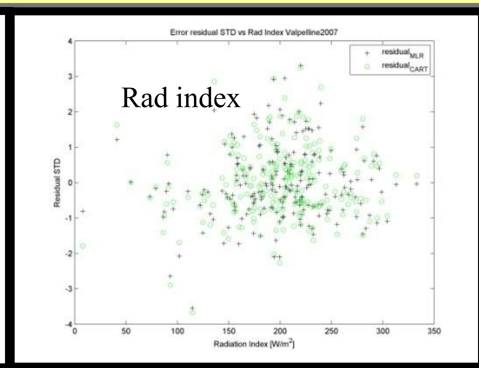


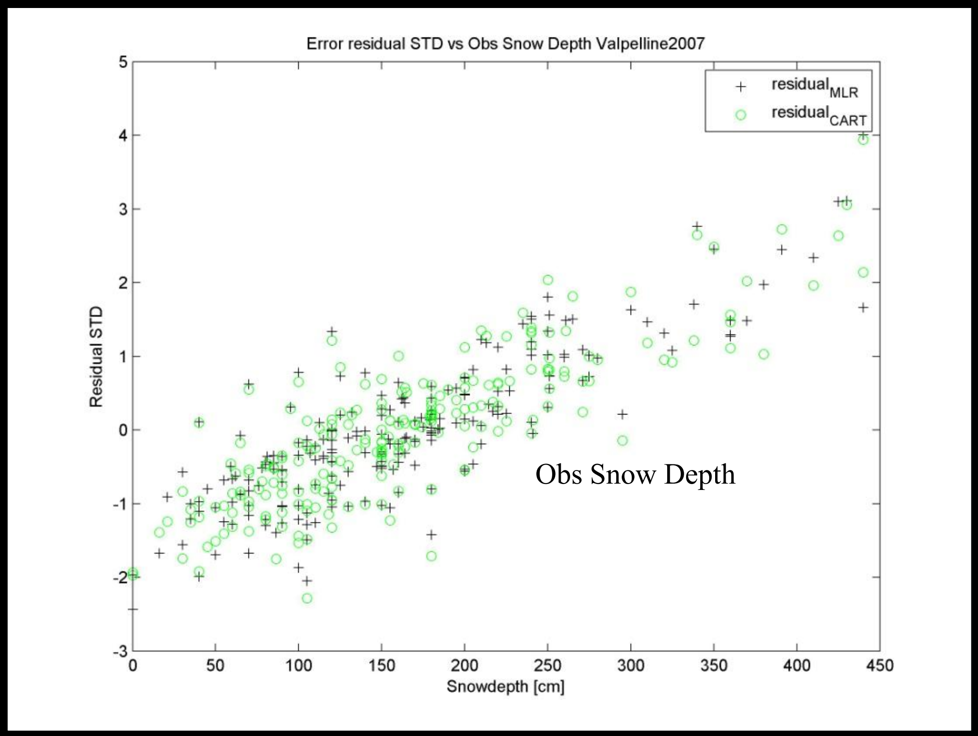
Subsample population has been chosen so that the stability of the mean and std dev does not strongly depends on random selection of fitting data subsample

Residuals Analysis









Scatterplots Standardized Residual

Observing scatter plots between Residuals Std and dependent variables does not appear a relevant residual trend, validating a fitting procedure and the choice of the predictors.

Unfortunately, the scatterplot between Std residuals and observed snow depth shows a remarkable trend of the residuals, underlining a systematic error in the models.

RESEARCH FOUNDATION cima
OBSERVE TO PREDICT
PREDICT TO PREVENT

Results Local Scale

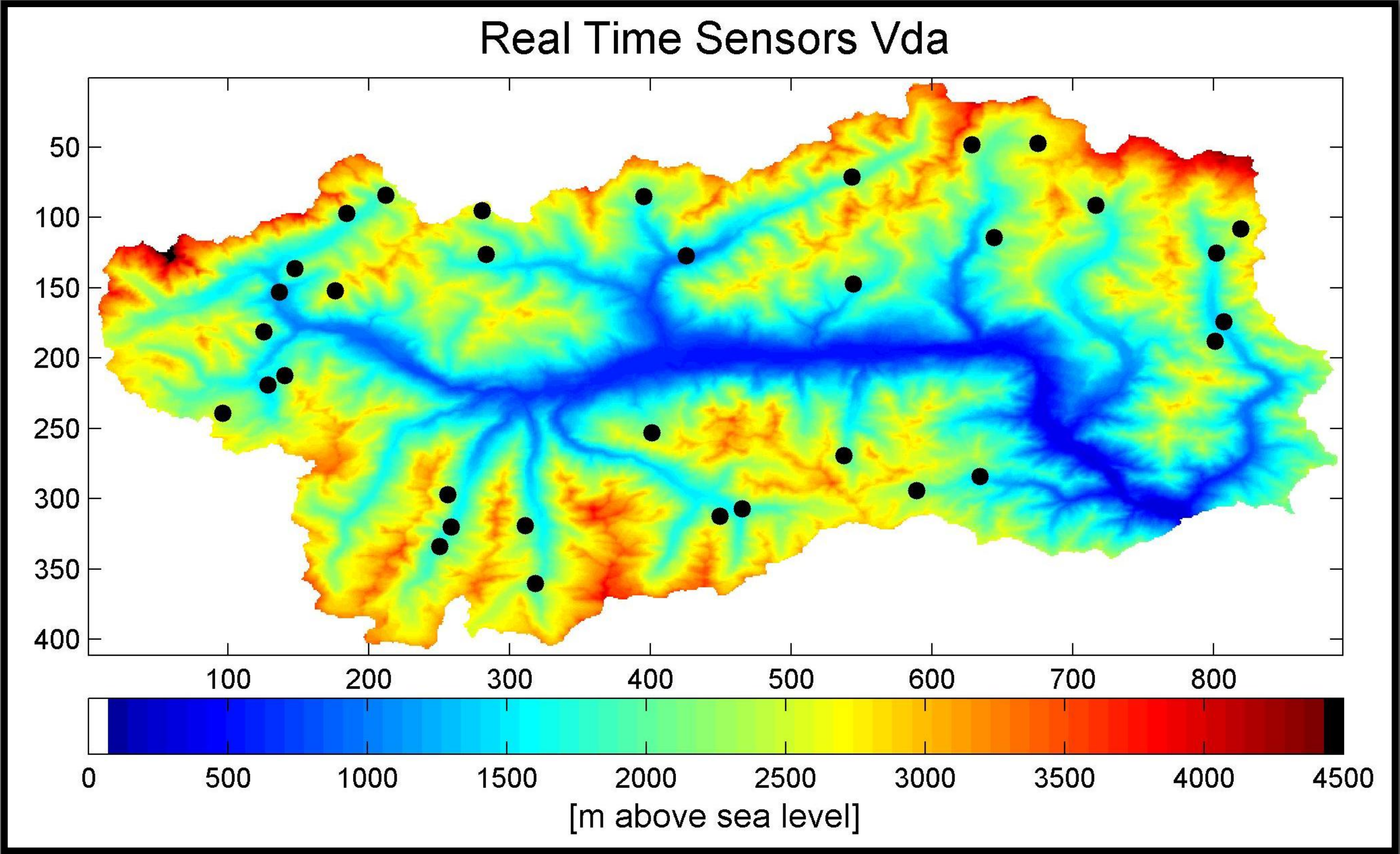
R^2	Adj R^2	Corr	MSE	RMSE	Predictors	Regression
Valpelline 2007						
0,30	0,29	0,54	5843,16	76,37	Elev-Upsl- Concavity *	MLR
					Elev-GradN- Rad- Concavity	
0,33	0,33	0,57	5529,23	74,3	Elev-Slope-Aspect	CART
Valpelline 2008						
0,06	0,05	0,23	6482,94	80,4	Elev-GradN	MLR
					Elev-Grad N	
0,22	0,21	0,46	5343,76	72,99	Elev-GradNE	CART
Gabiet						
0,01	0	0,02	2377,57	48,69	Slope-Aspect-Concavity *	MLR
					Grad NE –Rad-Concavity	
0,01	0	0,01	2374,78	48,66	Elev	CART
Goillet						
0,03	0,02	0,14	4650,79	68,1	Elev-Rad-Concavity *	MLR
					Elev- Grad E - Concavity	
0,04	0,03	0,19	4568,91	67,5	Elev-Aspect-UpSl	CART

Valpelline Model 2007 on 2008 data set

To understand the predictive skills of model built on snow depth data of 2007 snow campaigns, the model has been used to test next year snow depth distribution.

Valpelline 2007 on Valpelline next year						
R^2	Adj R^2	Corr	MSE	RMSE	Predictors	Regression
0.16	0.158	0.40	5717.9	75.61	Elev-Upsl-Concavity	MLR
0.17	**0.17**	**0,415**	**5631.8**	**75.04**	**Elev-Slope-Aspect**	**CART**

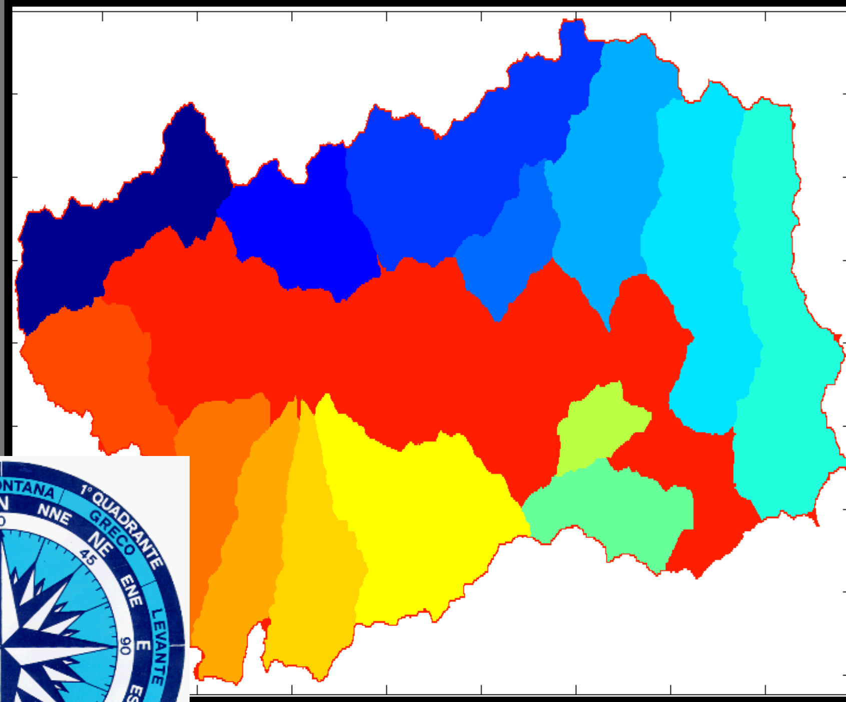
Regional Scale Analysis

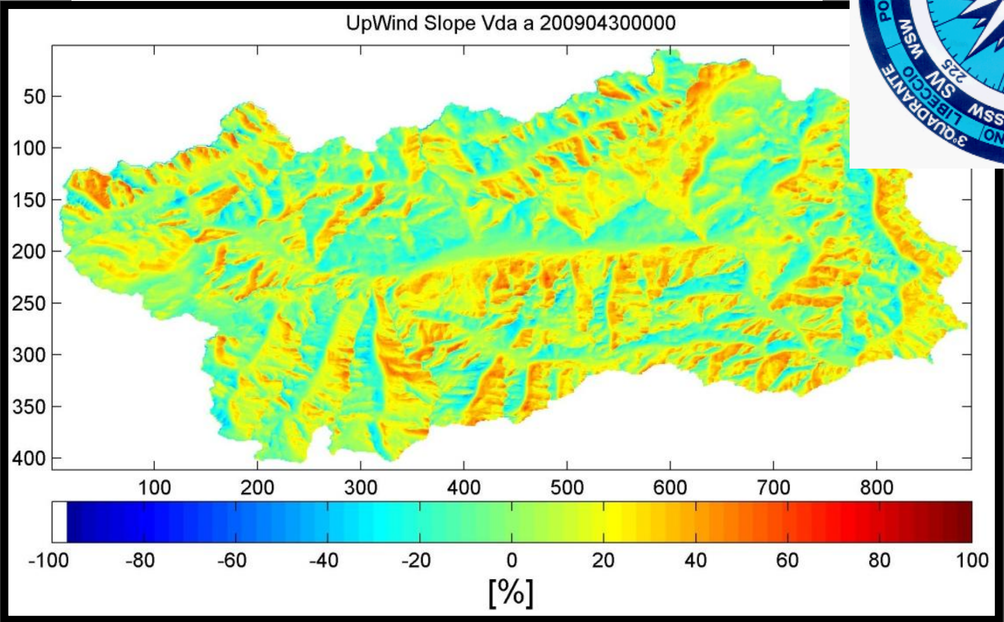


Real time data (mainly ultrasonic depth gauges) have been provided by Functional Centre of Civil Protection of Valle d'Aosta

Regional Upwind Index

Bacino	Direzione prevalente [gradi]
Artanavaz	260-310
Ayasse	280-340
Bianco	200-260
Buthier-BassaValle	15-75
Dora di Rhemes	200-260
Dora La Thuile	140-200
Evancon	280-340
Grand Eyvia	270-330
Lys	220-280
Marmore	120-180
S.Barthelemy	170-230
Savara	190-250
Chalamy	280-340
Dora Valgrisanche	200-260
Dora	270-320

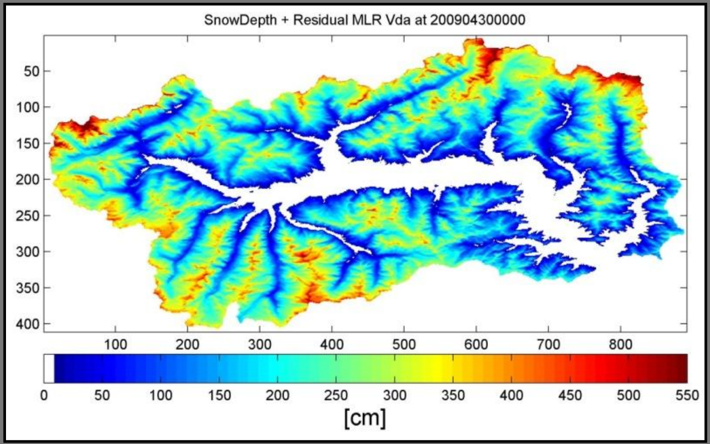




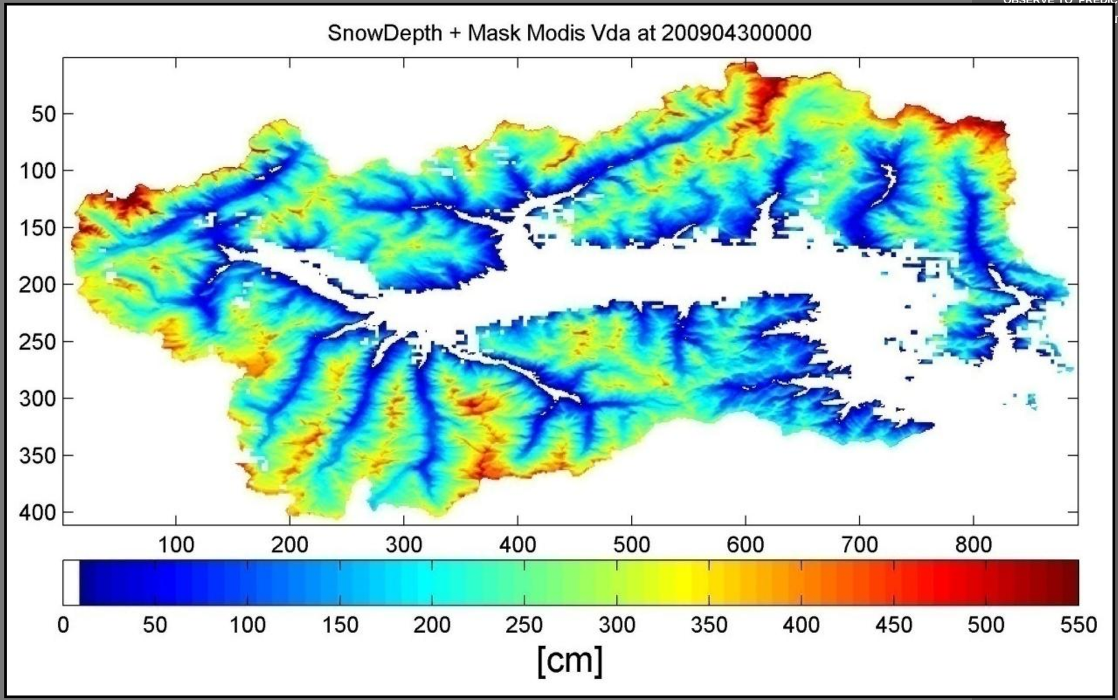
Mean upwind index has been derived partitioning the region surface in different macro-areas which have homogeneus wind features.

Direction on prevailing wind has been selected on anemometric data between 1995-2008

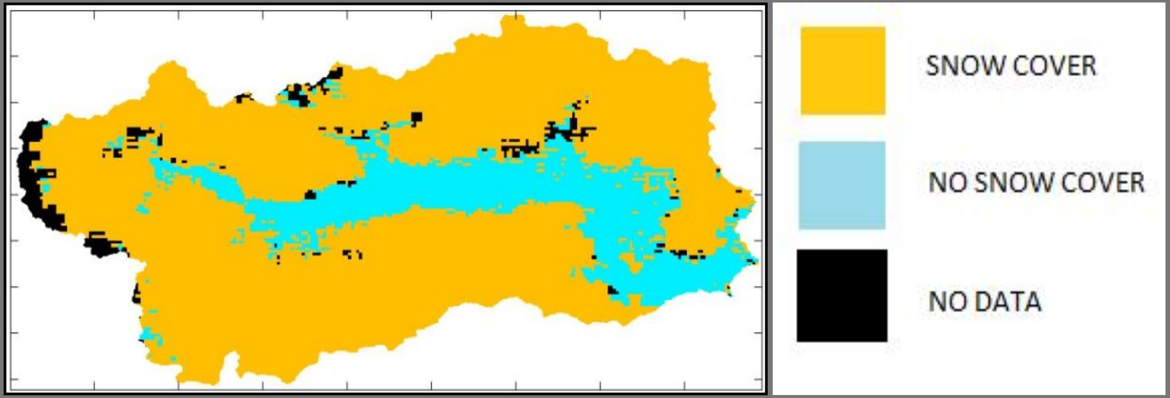
Regional Scale Analysis



MLR Snow Depth 30 Apr 2009

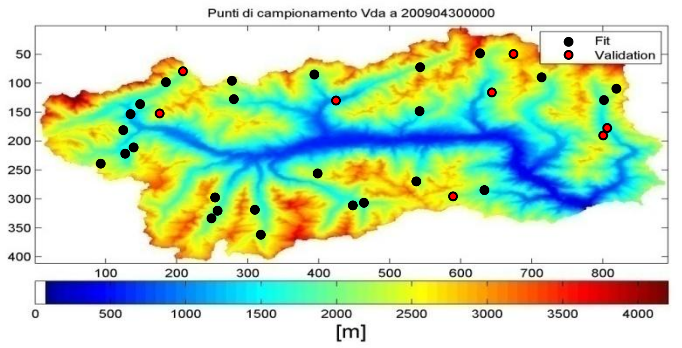


MODIS Mask 30 Apr 2009



Snow depth model was applied for 15 different clear sky days between January and June 2009.

Modis data with negligible cloud cover are added to interpolation maps adjusting snow depth covering at regional scale.



Results Regional Scale

R-square increases during melting period, including other predictors except Elevation, as GradN and Concavity. Rmse values are holded within 65 cm.

R^2	Adj R^2	Corr	MSE	RMSE	Predictors	Date
0.17	0.03	0.26	3099.78	54.53	Elev	**25-feb**
0.24	0.11	0.42	3331.43	55.53	Elev	**27-feb**
0.21	0.08	0.34	3787.13	59.41	Elev	**28-feb**
0.25	0.13	0.35	4394.16	63.53	Elev	**13-mar**
0.25	0.13	0.44	4510.03	64.81	Elev	**16-mar**
0.19	0.06	0.32	4872.91	67.94	Elev	**17-mar**
0.21	0.07	0.35	3892.35	61.08	Elev	**18-mar**
0.27	0.15	0.39	3305.19	55.24	Elev	**19-mar**
0.33	0.21	0.48	4063.97	61.76	Elev	**21-mar**
0.53	0.45	0.71	3755.73	59.88	Elev-GradN	**22-apr**
0.61	0.55	0.77	4511.80	65.33	Elev-GradN	**30-apr**
0.61	0.54	0.77	4228.61	62.77	Elev-GradN	**01-may**
0.50	0.42	0.68	4557.28	65.08	Elev-GradN	**07-may**
0.37	0.26	0.39	1100.90	30.27	Elev-Concavity	**27-may**
0.30	0.15	0.28	3043.87	51.36	Concavity	**28-may**
0.30	0.15	0.38	1529.10	36.30	Elev-Rad-Concavity	**13-jun**

Conclusions

RESEARCH FOUNDATION cima
OBSERVE TO PREDICT
PREDICT TO PREVENT

• Results at small scale have shown a poor predictive skills of the models. In particular for Goillet- Gabiet the models are not able to explain snow depth variability, producing R-square values close to 0. Valpelline 2007 produces better results with R-square values close to 0.3 where there is a strong correlation with Elevation and more snow depth measurements.

• Residual Analysis denote a systematic error for both models showing a trend between standardized residuals and observed snow depth.

• At regional scale, MLR provides better results, with R-square values close to 0.6 during spring. An analysis of the independent variables chosen by the procedure in the regional scale experiment shows that, during the accumulation period, snow depth correlates with elevation only, whereas, during spring, gradient in north direction concavity parameters and radiation index become relevant too.

RESEARCH FOUNDATION

cima

OBSERVE TO PREDICT,
PREDICT TO PREVENT.

Thanks for your attention !