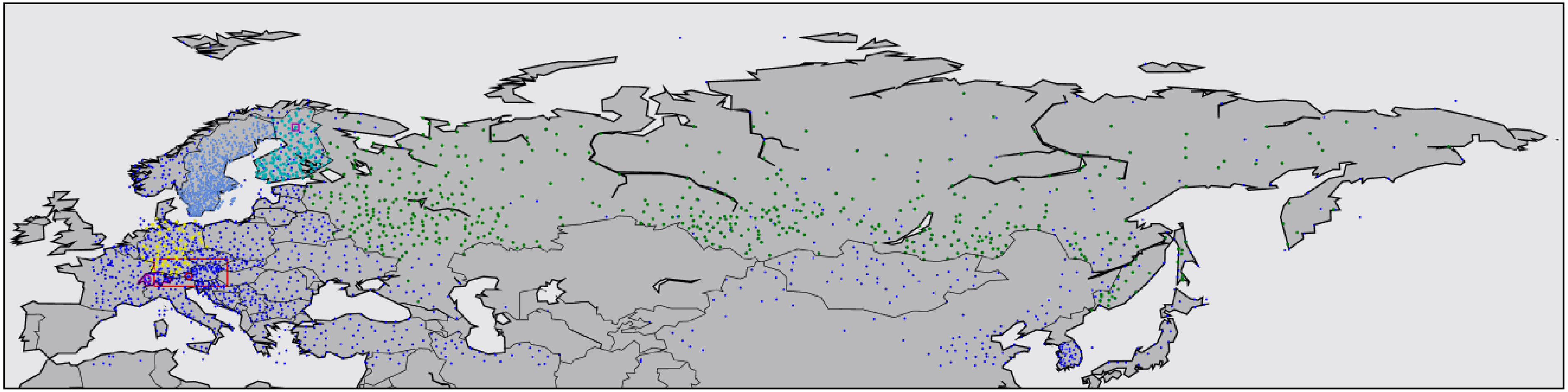
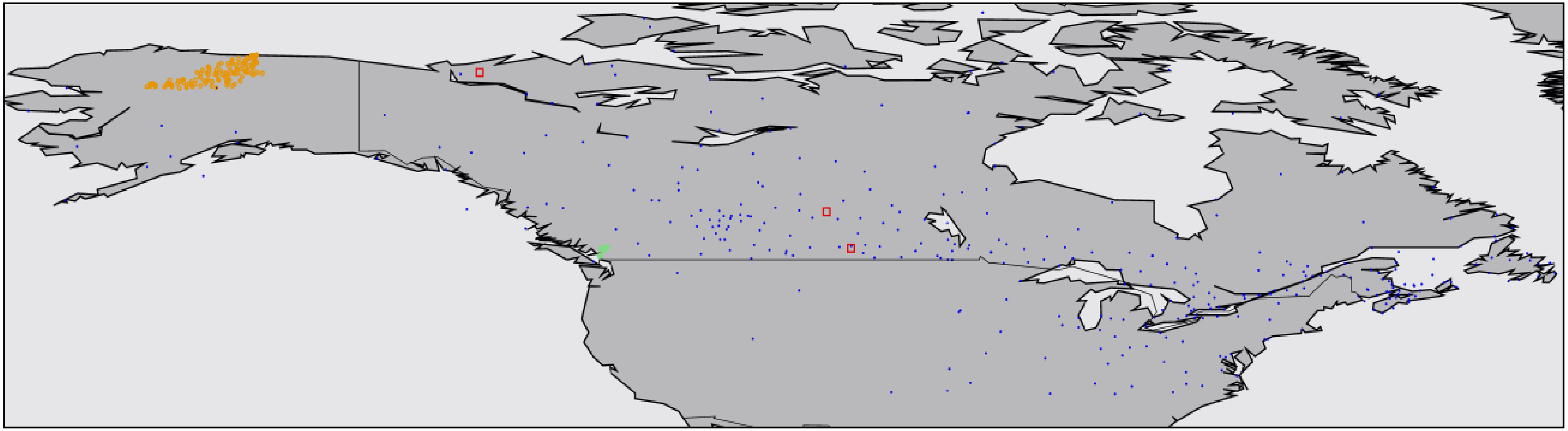
SnowPEx in-situ datasets and the first validation results

Sari Metsämäki, Olli-Pekka Mattila, Gabriele Bippus, Kari Luojus,Chris Derksen, Elisabeth Ripper et al.

enveo SYKE FMI NR EC Natural Resources Canada

In-situ data coverage for validation





In-situ datasets for validation

snowpex

SYKE

Dataset	Region	Snow Class	Method	Available Time Period	Temporal Resolution	Contact	Param.	Data Policy	Samples @ FTP
Pointwise data									
ECMWF Weather stations	Europe/ North America	All	Sonic snow dept, Manual surveys	1978-2014	Daily	ECMWF in SnowPEx K. Luojus, FMI	SD	Restricted	
RIHMI Weather stations	Russia and former USSR	All	Manual surveys	1966-2011	Daily	O. Bulygina, RIHMI	SD, FSC	Open (registration at RIHMI web page)	All seasons from RIHMI database
FMI Weather stations (Finland)	Finland	All	Sonic snow depth, Manual surveys	1978-2014	Daily	K. Luojus, FMI	SD	Restricted (sample data available on FTP)	2003-2004 2011-2012
ECA&D Weather stations (Germany)	Germany (+ Europe)	All	Sonic snow depth, Manual surveys	2000-2012	Daily	ECA&D in SnowPEx, S. Metsämäki, SYKE	SD	Open	All seasons*
SMHI Weather station data (Sweden)	Sweden	Mountains, taiga	Sonic snow depth, Manual surveys	1980-2015	Daily	SMHI in SnowPEx, S. Metsämäki, SYKE	SD	Open	All seasons*
NVE snow stations (Norway)	Norway	All	Automated stations	1967-2015	Hourly/ Daily	Rune Solberg, NR	SD, SWE	Open	Not yet but will be, season has to checked

* All seasons: 2000-2001, 2003-2004, 2005-2006, 2007-2008, 2011-2012

In-situ datasets for validation

Dataset	Region	Snow Class	Method	Available Time Period	Temporal Resolution	Contact	Parameter	Data Policy	Sample @ FTP
Pointwise data									
Environment Canada, Olympics 2010	Southern coast mountains	Alpine	Sonic snow depth	2008-2010	Daily	C. Derksen, Environment Canada	SD	On request	
Environment Canada, Bratt's Lake	Saskatchewan	Prairie	Manual surveys	2002-2005	Bi-weekly	C. Smith, Environment Canada	SWE, SD, Density	On request	
Environment Canada, Trail Valley Creek	Northwest Territories	Tundra	Manual surveys	1991-2014	End of season	P. Marsh, Wilfrid Laurier Univ.	SWE, SD, Density	On request	
University of Saskatchewan, Boreal Ecosystem Research and Monitoring Sites	Saskatchewan	Taiga	Sonic snow depth	1997-2011	Daily	H Wheater, Univ. Saskatchewan	SD	On request	
University of Saskatchewan, Boreal Ecosystem Research and Monitoring Sites	Saskatchewan	Taiga	Manual surveys	1995-2011	Monthly	H Wheater, Univ. Saskatchewan	SWE, SD, Density	On request	
University of Alaska, Kuparuk Basin snow surveys	Alaska	Tundra	Snow surveys	2006-2013	Snapshot	S. Stueffer, Univ. of Alaska – Fairbanks	SWEmax	On request	

In-situ datasets for validation

Dataset	Region	Snow Class	Method	Available Time Period	Temporal Resolution	Contact	Parameter	Data Policy	Data @ FTP
Snow course data									
SYKE Snow Surveys	Finland	Taiga	Manual snow course	2002-2014	Monthly	S. Metsämäki SYKE	SD, FSC (course mean)	Restricted (sample data available on FTP)	10/2003-05/2004 10/2007-05/2008
RIHMI Snow Surveys	Russia	Taiga and tundra	Manual snow course	1966-2014	Bi-weekly	O. Bulygina, RIHMI	SD, SWE, FSC, Density	Open* (registration at RIHMI web page)	All seasons from RIHMI database
Interpolated data									
Hydro-Quebec Krigged SWE	Southern Quebec	Agricultural, forest	Interpolated snow course	1999-2010	Bi-weekly	R. Brown, Environment Canada	SWE	Restricted	
WSL Institute for Snow and Avalanche Research SLF	Switzerland	Mountains	Interpolated snow observations using distributed hydrological model	1998-2014	Daily	T. Jonas, SLF	SWE	Restricted	
SNOWGRID	Alps	Mountains	Gridded snow cover model	2011-2012	Daily	M. Olefs, ZAMG	SWE, SD	Restricted	10/2011-05/2012

* RIHMI web page: http://meteo.ru/english/climate/cl_data.php

Global validation data on Snow Depth (ECMWF + supplementary data)

snowpex

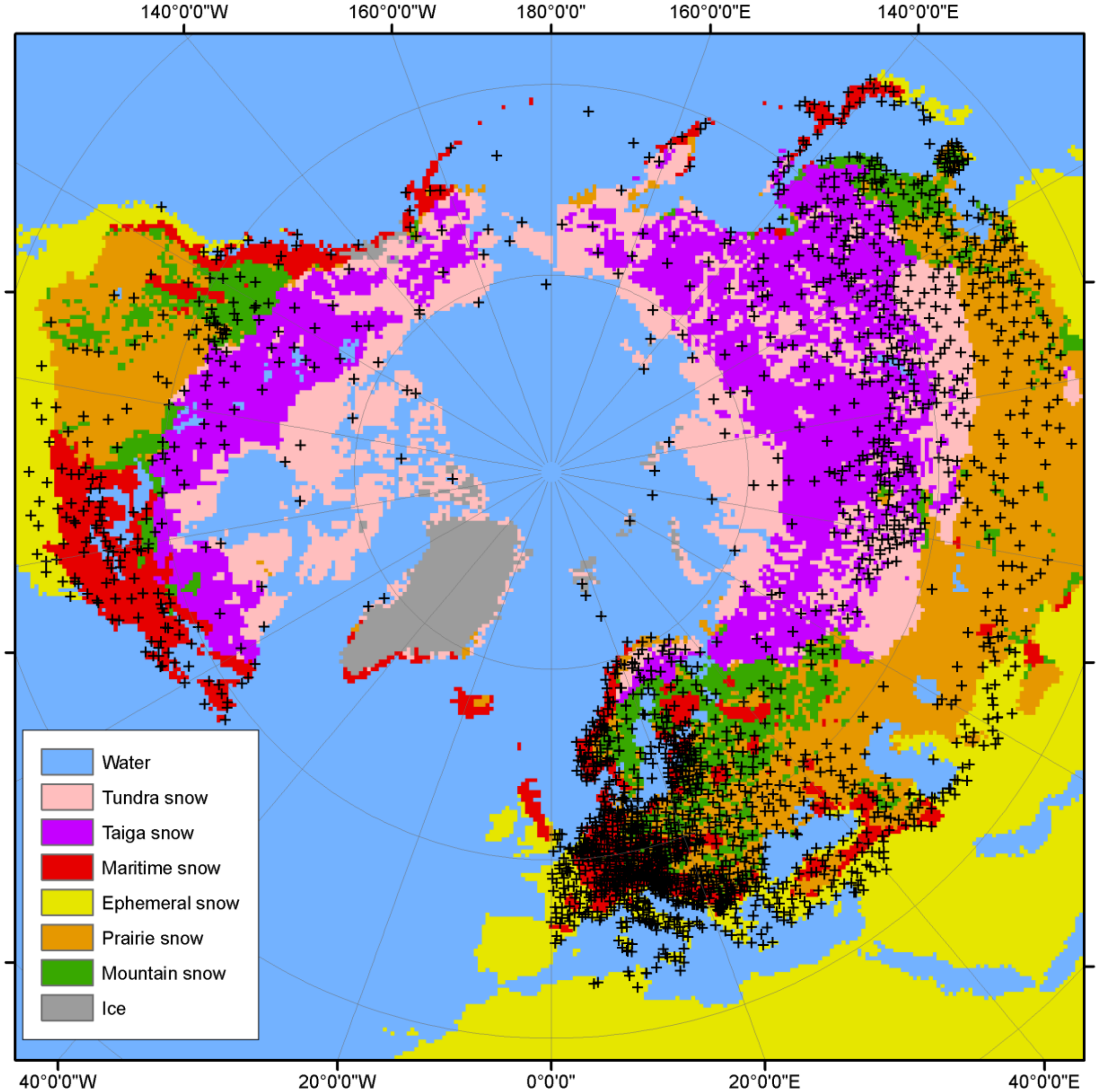
SYKE

- The global dataset used in validation contains ECMWF-dataset as core data and is supplemented, especially in North America, from various contributors (e.g. Dave Robinson)

- Sample from dataset (all dataset in the same format .csv):

```
ID;latitude;longitude;ease2north;ease2east;elevation;timeStamp;snowDepth
NA;54.18;7.9;-3894000;540300;-9999;20031001;0
NA;54.53;9.55;-3840300;646100;-9999;20031001;0
NA;54.53;11.07;-3821800;747700;-9999;20031001;0
NA;53.63;9.98;-3929200;691400;-9999;20031001;0
NA;54.1;13.4;-3832600;913000;-9999;20031001;0
```

Global validation data coverage (Eurasia and North-America)



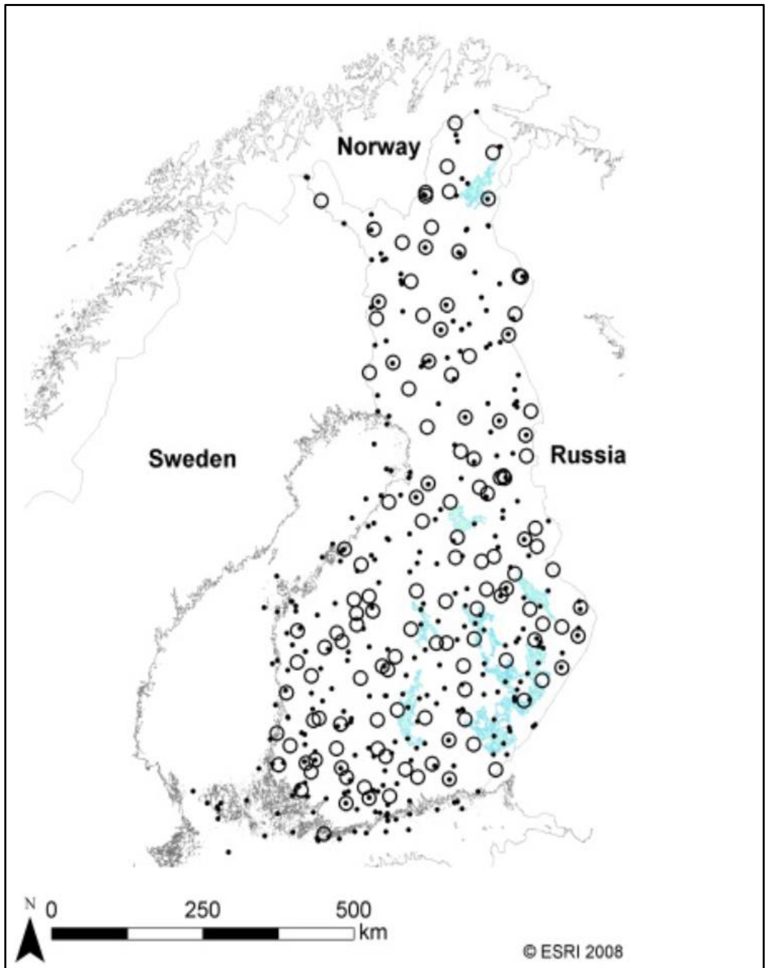
Datasets on Snow Cover Fraction

- Direct SCF observations can be used from two datasets:
  1) SYKE snow courses
  2) RIHMI (Russian Research Institute for Hydro-meteorological Information) weather stations
- Sample from RIHMI data:

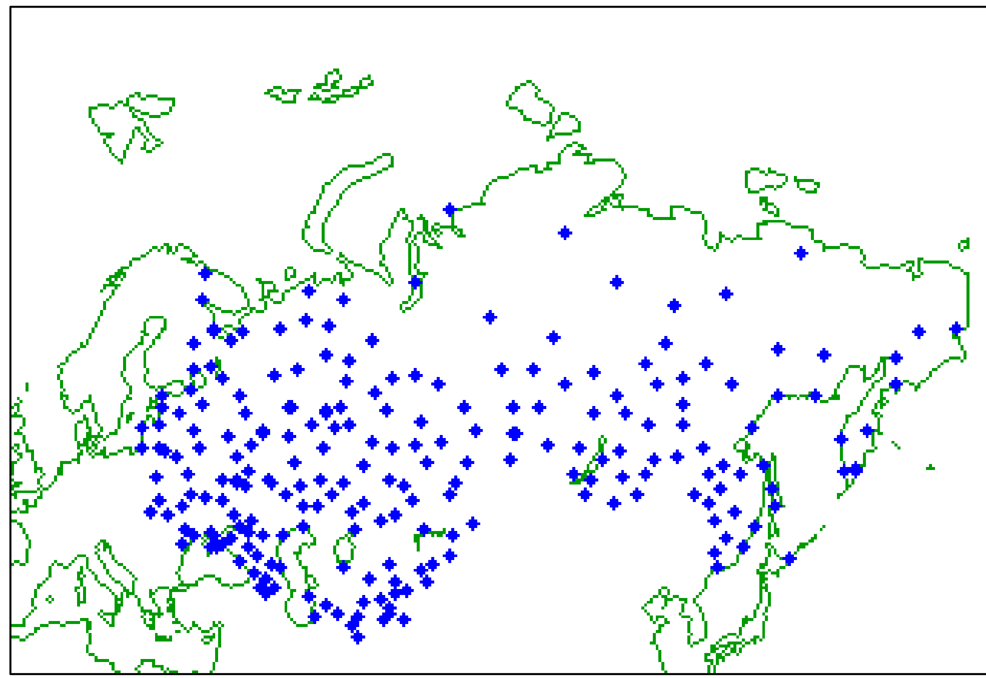
```
ID;latitude;longitude;ease2north;ease2east;elevation;timeStamp;snowDepth;fractionalSnowCover;qualityFlagGeneral;qualityFlagSD;qualityFlagTemp
20046;80.62;58.05;-553700;887900;21;20001001;2;100;0;0;0
20046;80.62;58.05;-553700;887900;21;20001002;2;100;0;0;0
20046;80.62;58.05;-553700;887900;21;20001003;2;100;0;0;0
20046;80.62;58.05;-553700;887900;21;20001004;2;100;0;0;0
```

SCF Datasets
(SYKE Snow course and RIHMI snow coverage)

SYKE snow courses



RIHMI weather stations



Preparing the data for the in-situ validation

1) Reprojecting and aggregating the original product into EASE-grid 2.0 (5km and 25km)
2) Producing the proportion of valid areas (associated to the original product) for EASE-grid cells → VAA
3) Producing the proportion of Mapped (and valid) areas for EASE-grid cells → MAA
4) Producing EASE-grid coordinates for all insitu observations (originally in WGS-84 lat,long system)
5) Generation of auxiliary data for stratifications (LandCover map, Climate map etc.) in EASE-grid 2.0

Validation protocol (CCRS)

We are not directly comparing the pointwise observation and the pixel value of the snow product. Instead, for each site:

1) We generate a Probability Distribution Function (pdf) and the corresponding Cumulative Distribution function (cdf) of the Snow Cover Fraction (SCF) for products and in-situ observations

- ***pdf_product*** is generated based on data provider-given precision and the value (SCF or Binary) of the snow product
- ***pdf_in-situ*** is generated either

    1) from the direct observation on SCF (only Finnish Snow courses and Russian RIHMI-data

    2) from Snow Depth (SD) observation using a statistical relationship (empirically derived) between SD and SCF

2) we generate N realizations of SCF for both in-situ SCF and Product SCF

3) These N realizations are combined → N*N datapairs (in-situ, product) are produced

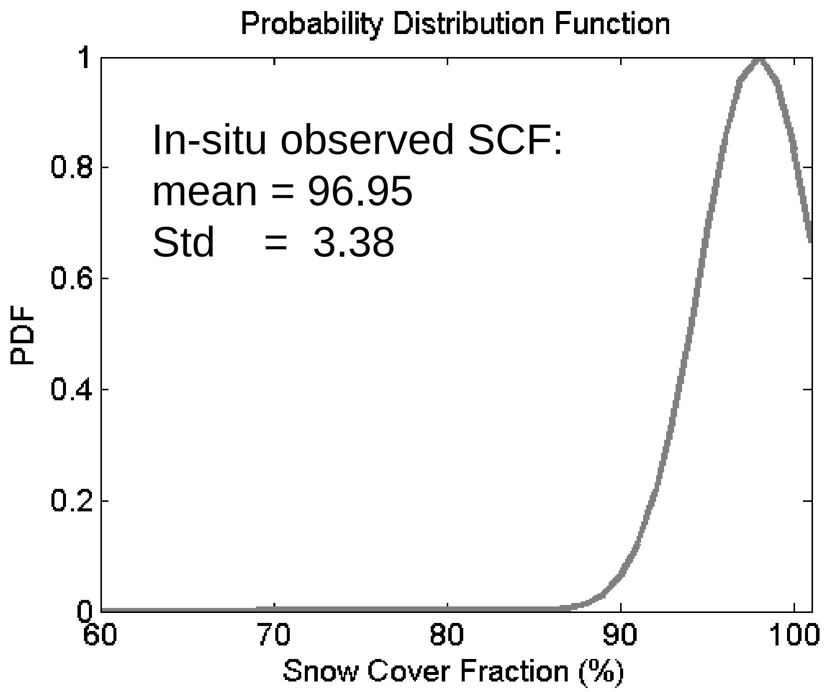
Generation of samples from direct SCF in-situ observations. A true example:

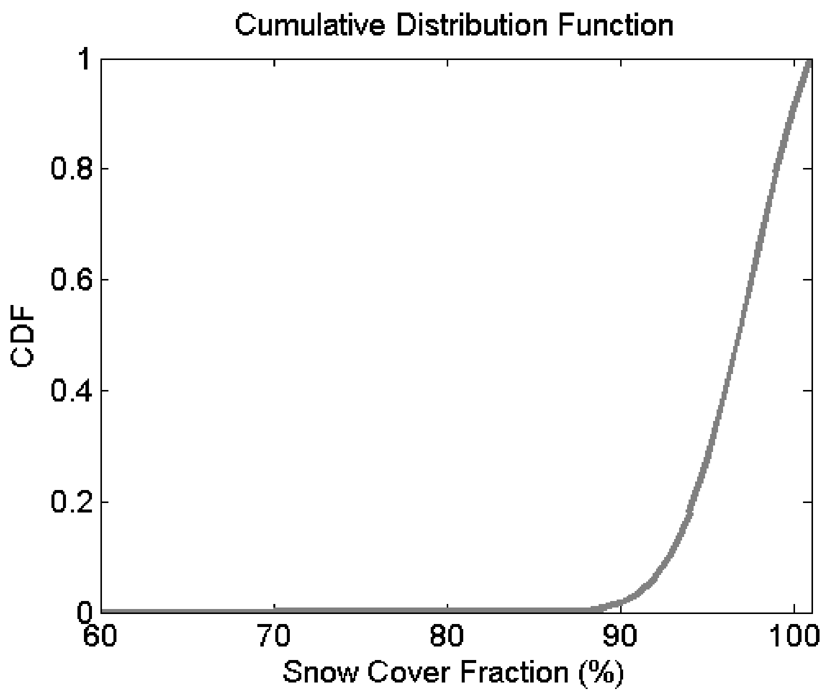
from the snow course SCF (ID 1510, April 15, 2004):

The provided average SCF and standard deviation are used to generate

1) probability distribution function (pdf)

2) Cumulative distribution function (cdf)





Generation of samples from direct SCF in-situ observations. A true example:

from the snow course SCF (ID 1510, April 15, 2004):

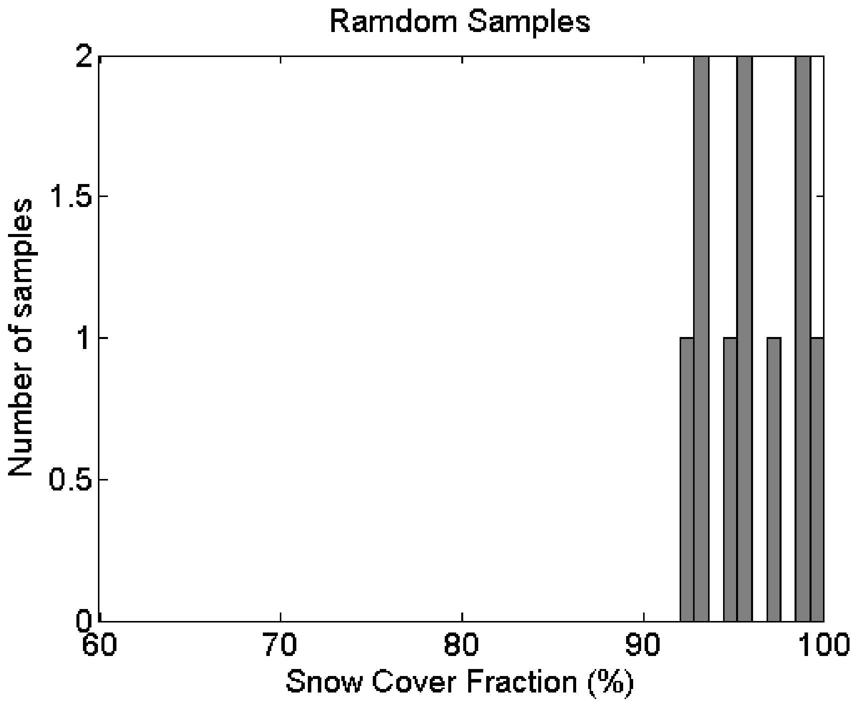
The provided average SCF and standard deviation are used to generate

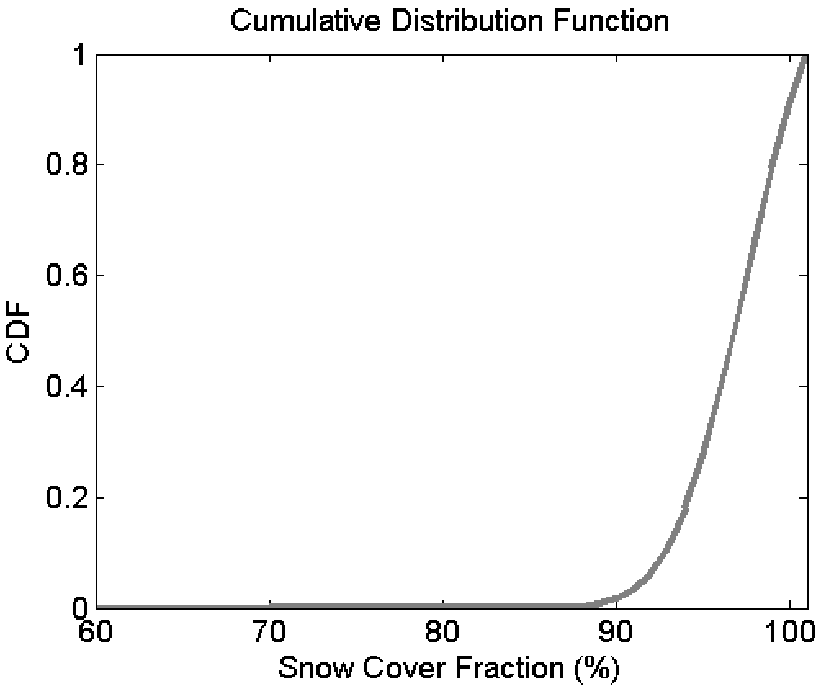
1) probability distribution function (pdf)

2) Cumulative distribution function (cdf)

from CDF:

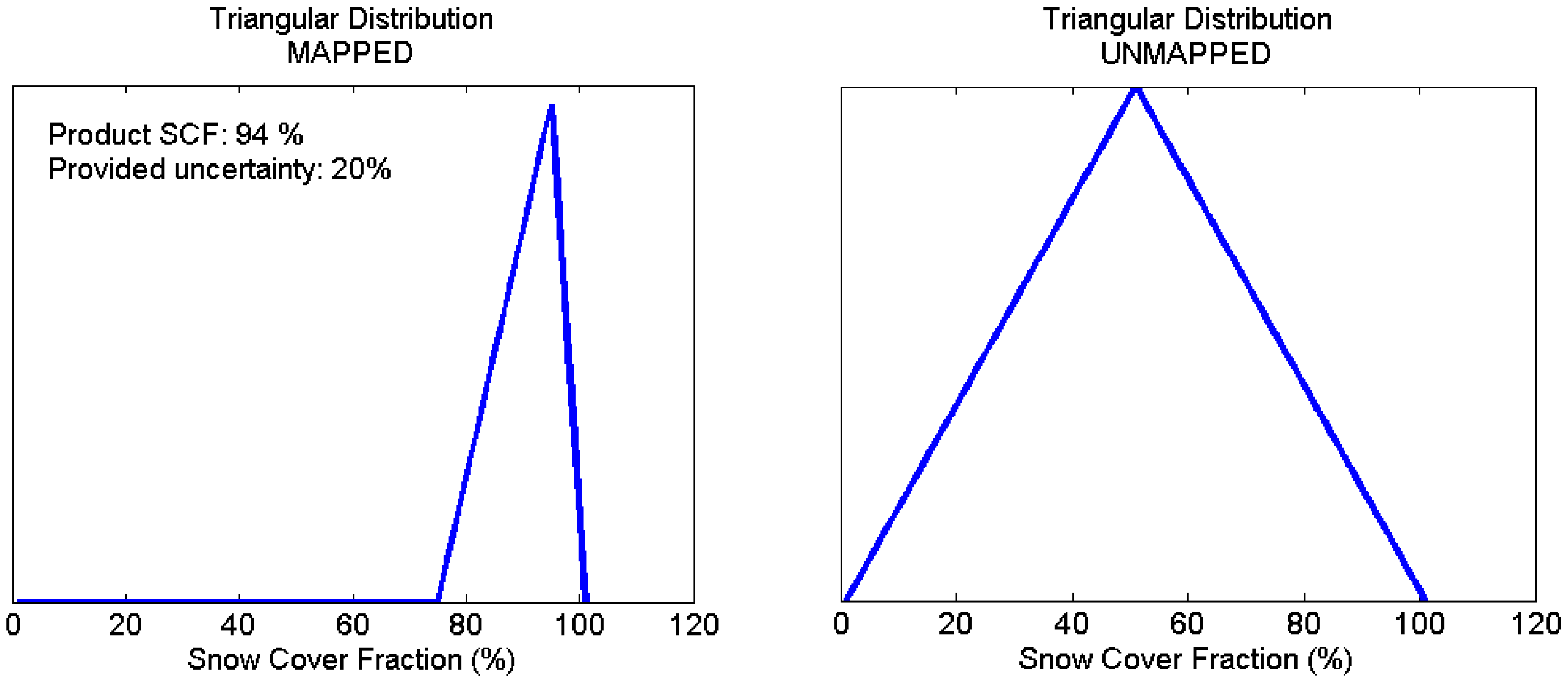
N samples are randomly generated (10 samples according to the protocol)





Generation of samples from direct SCF snow product for a validation cell

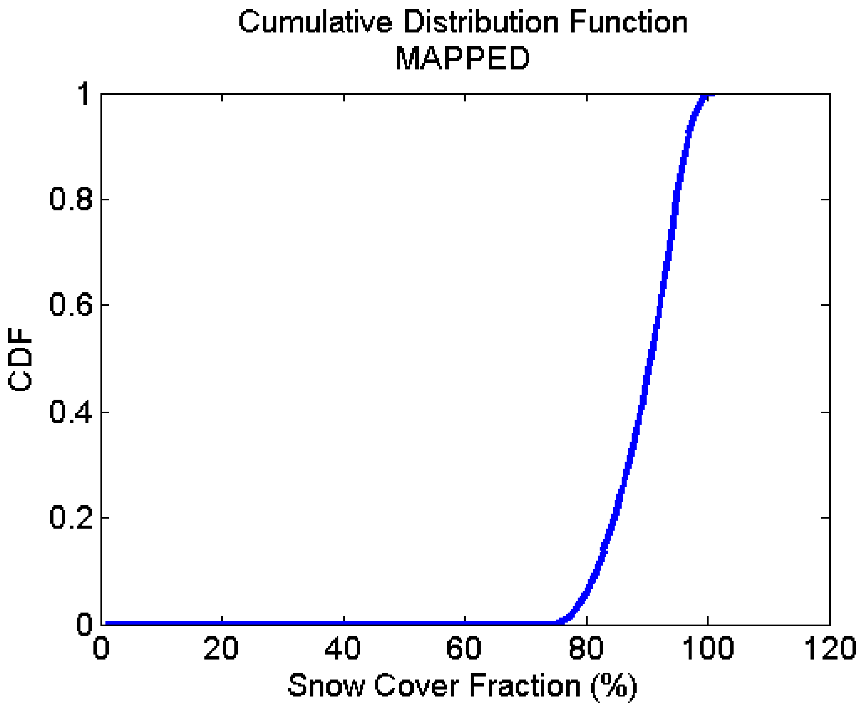
- **PDF for MAPPED** portion is generated using triangular function with mean at product SCF, and vertices according provider-gven precision
- **PDF for UNMAPPED** portion is generated using triangular function with mean at 50%, and vertices at 0 and 100%
- In the example case, mapped portion = 0.8889 → unmapped = 0.1111

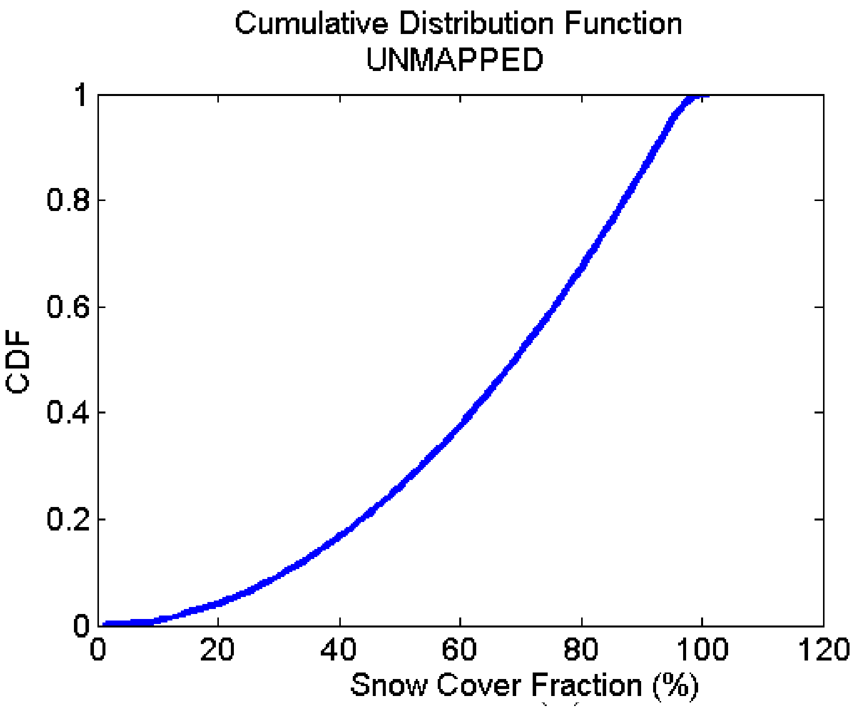


Generation of samples from direct SCF snow product for a validation cell

- **PDF for MAPPED** portion is generated using triangular function with mean at product SCF, and vertices according provider-gven precision
- **PDF for UNMAPPED** portion is generated using triangular function with mean at 50%, and vertices at 0 and 100%
- In the example case, mapped portion = 0.8889 → unmapped = 0.1111

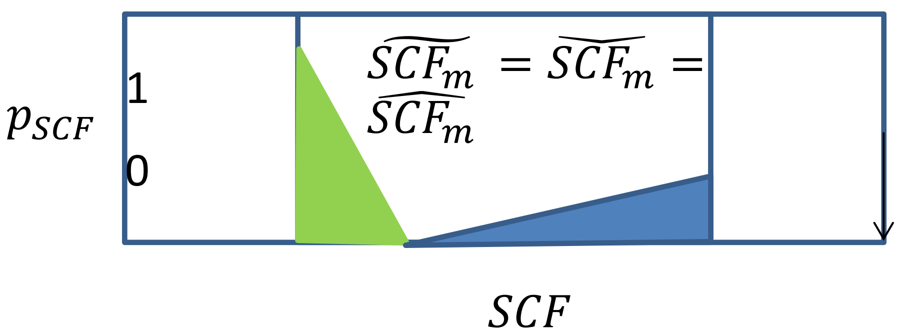
Cumulative Distribution Functions are generated from PDF's, then 10 random samples are generated from these, weighting proportionally to %mapped and %unmapped





Generation of samples from binary snow product for a validation cell

- **PDF for MAPPED** portion is generated using triangular function, vertices according provider-given limits: e.g. ’no-snow’ in SCF range [0-50%], ’snow’ in SCF range [50-100%] and mode in between of these. There may as well be two threholds non-overlapping instead of just one
  - It is important to have the threshold for ’snow’ and ’no-snow’ from the product provider
- **PDF for UNMAPPED** portion is generated using triangular function with mean at 50%, and vertices at 0 and 100%

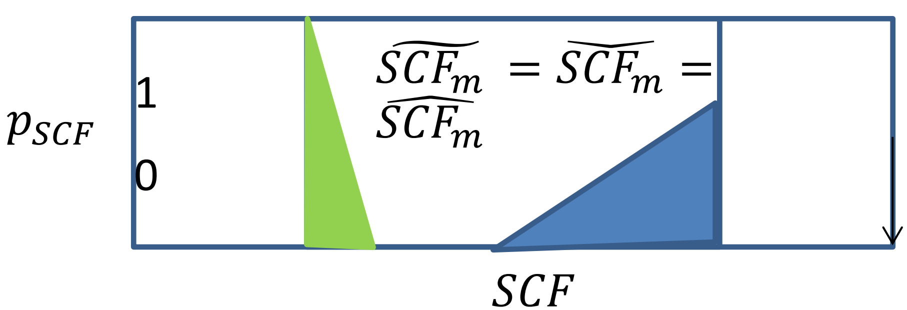


$$\widehat{SCF}_M = \frac{N_{snow}}{N_{snow} + N_{no-snow}}$$

$$p\left(SCF \mid \widehat{SCF}_M, \widehat{SCF}_M, \widehat{SCF}_M, f_{mM}\right) =$$

$$f_M\left[\widehat{SCF}_M p\left(SCF_s \mid \widehat{SCF}_s, \widehat{SCF}_s, \widehat{SCF}_s\right) + \left(1 - \widehat{SCF}_M\right) p\left(SCF_{ns} \mid \widehat{SCF}_{ns}, \widehat{SCF}_{ns}, \widehat{SCF}_{ns}\right)\right] *$$

$$\left(1 - f_M\right) p\left(SCF_U \mid \widehat{SCF}_{mM}\right)$$

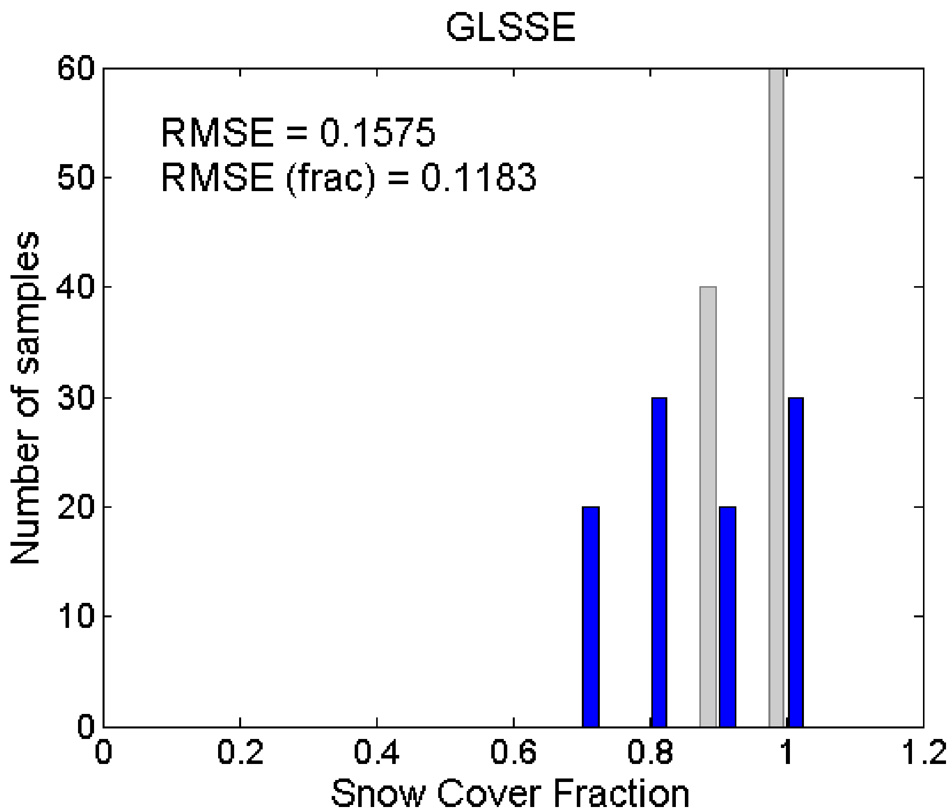


Generation of N*N cases from N in-situ realizations and N product realizations

10*10 combinations are taken from in-situ samples and product samples → 100 datapairs are generated.

Product samples	In-situ samples
0.8900	0.9200
0.7522	0.9700
0.6978	0.9500
0.7778	0.9900
0.8333	0.9300
0.9311	0.9900
0.8478	0.9600
0.8489	1.0000
0.9044	0.9300
0.7944	0.9600

100 Samples
(binned in the figure)



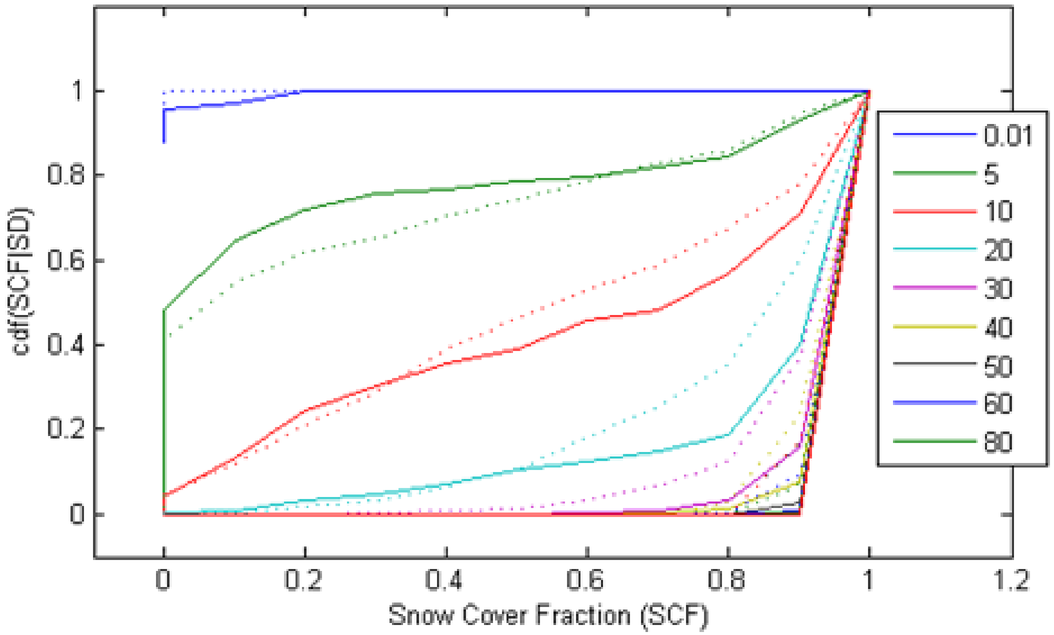
Generation of in-situ samples from Snow Depth Observations

snowpex

SYKE

The ultimate majority of in-situ snow observations are on Snow Depth (SD) not on Snow Cover Fraction.

- SD is converted to FSC using the statistical relationship of these two, based on Finnish Snow course observation covering both these variables



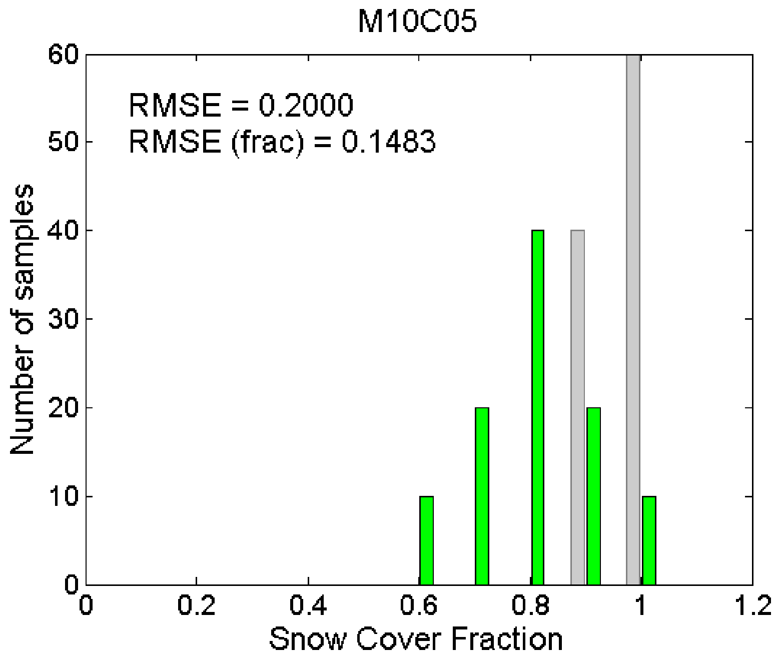
Cumulative Distribution Function is determined according to the observed snow depth. After that, validation continue as previously described.

This is repeated for different products, site by site, day by day, with required stratifications

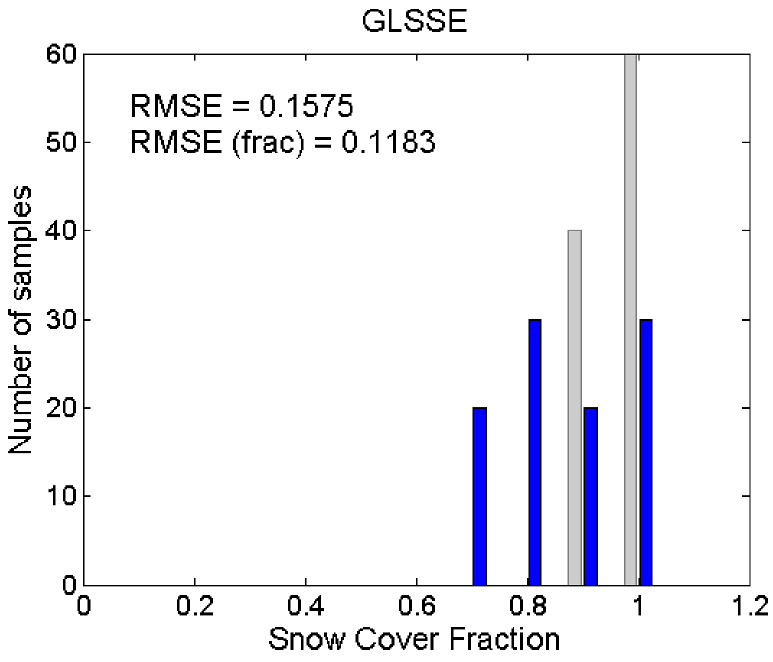
A site&day-specific RMSE can be calculated from the 100 datapairs (insitu & product).

An example on results for Finnish Snow course ID 1510, April 15, 2004.

Product CSF = 82%



Product CSF = 94%

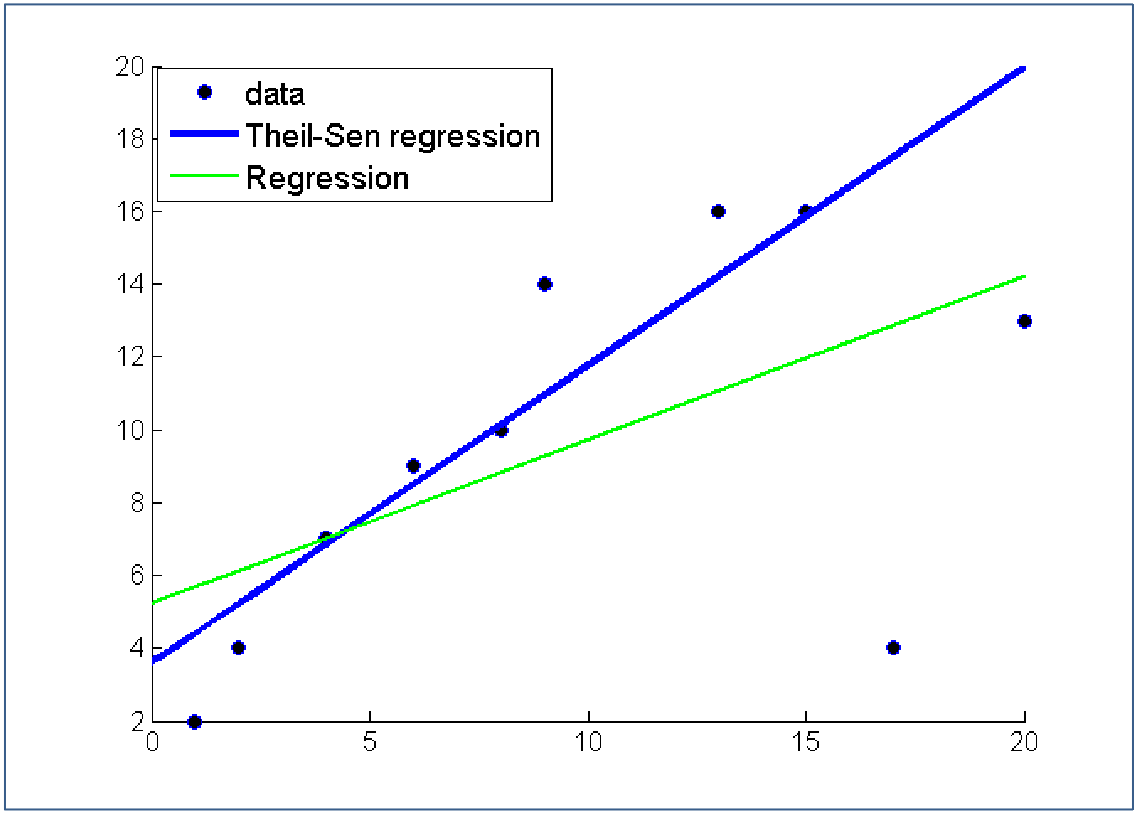


Validation employing different in-situ dataset and different products

- A site&day-specific RMSE can be calculated from the 100 datapairs (insitu & product).
- However, to obtain a general accuracy measures for a product type, all samples are combined into a large dataset (at this phase, separately for each yearly season.
- This dataset can be stratified a varying ways:
  - By month
  - By SCF bins
  - By Landcover (e.g. forest, non-forest)
  - By climate zone
- A number of different measures are proposed within SnowPEx:
  - RMSE
  - Bias (general Bias or as Slope and Intercept from Theil-Sen regression line Non-biased RMSE
  - Correlation
  - …see (Tables 2.2 and 2.3 in the Validation Protocol)

Theil-Sen regression employed

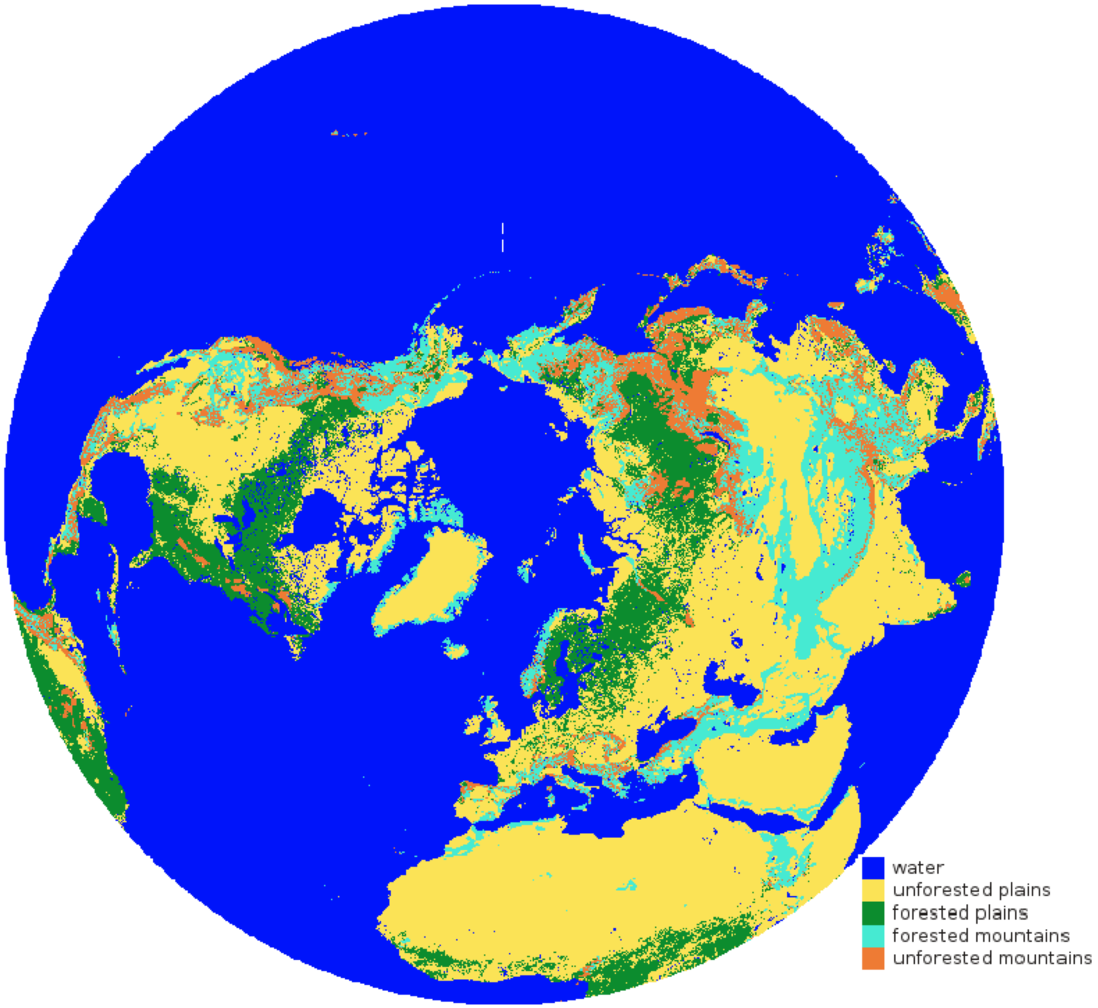
- Theil-Sen utilized for non-biased RMSE, Correlation and SCF-dependent bias



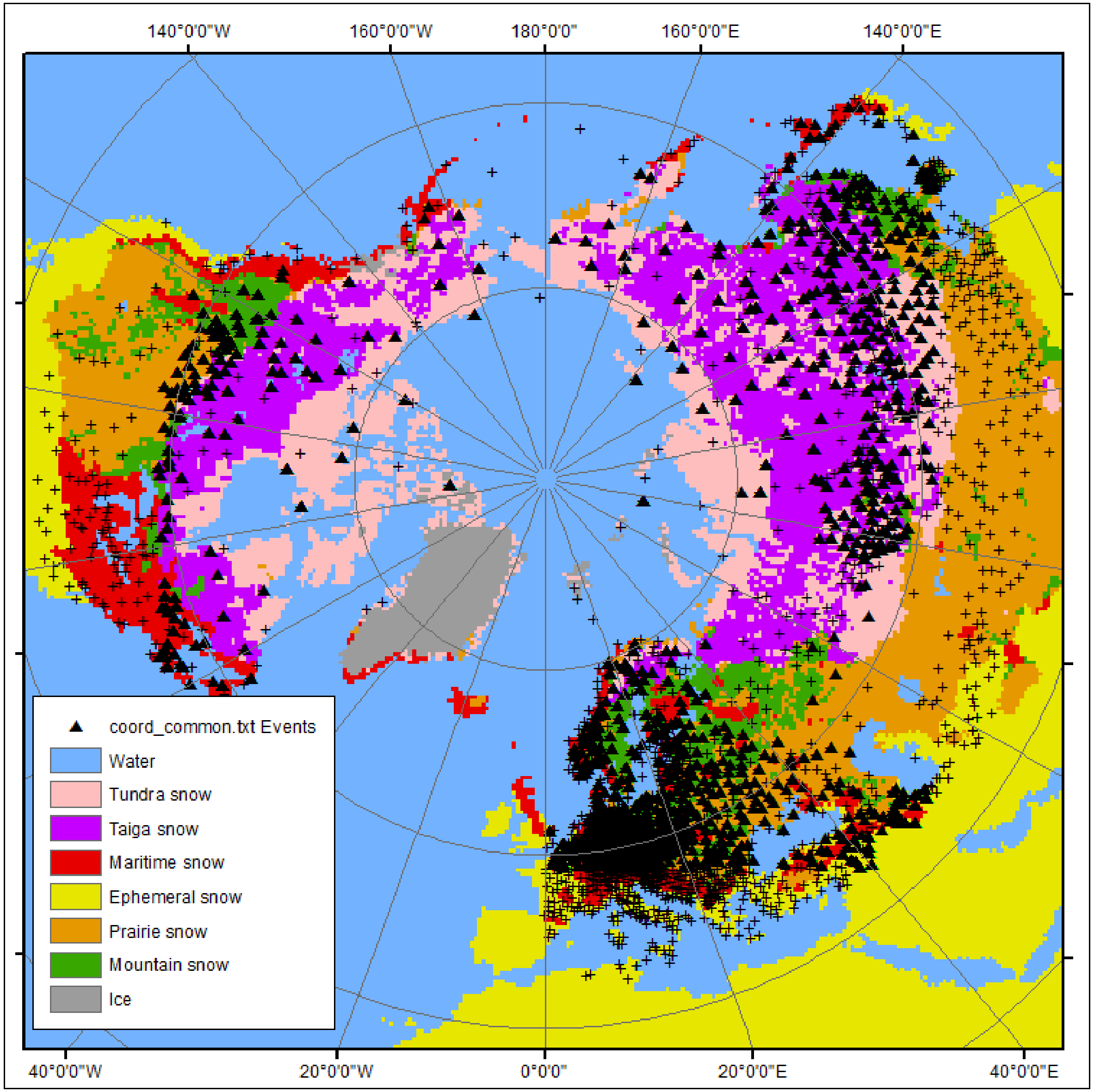
Status of the validation:

- Analyses have been made for the SCF products, seasons 2003-2004 and 2007-2008
  - M10C05
  - PATHF
  - GLSSE
  - CryoLand
  - ASNOW (originally binary, but converts for SCF when aggragated to EASE-2 grid)
- SCF-products were evaluated against:
  - NH ECMWF SD-observations
  - German ECA&D SD-observations
  - Finnish CSF-observations (snow course data)
- RMSE and bias are produced for each of the products, but for comparing the results , a subset of spatially/temporally intersecting cases should be selected
  - GLSSE provides lowest spatial coverage per day, so when comparing GLSSE against the others, number of cases is strongly reduced

The employed LandCover data for stratification



Results from the intersecting cases for three CSF-product: GLSSE, M10C02 and PATHF



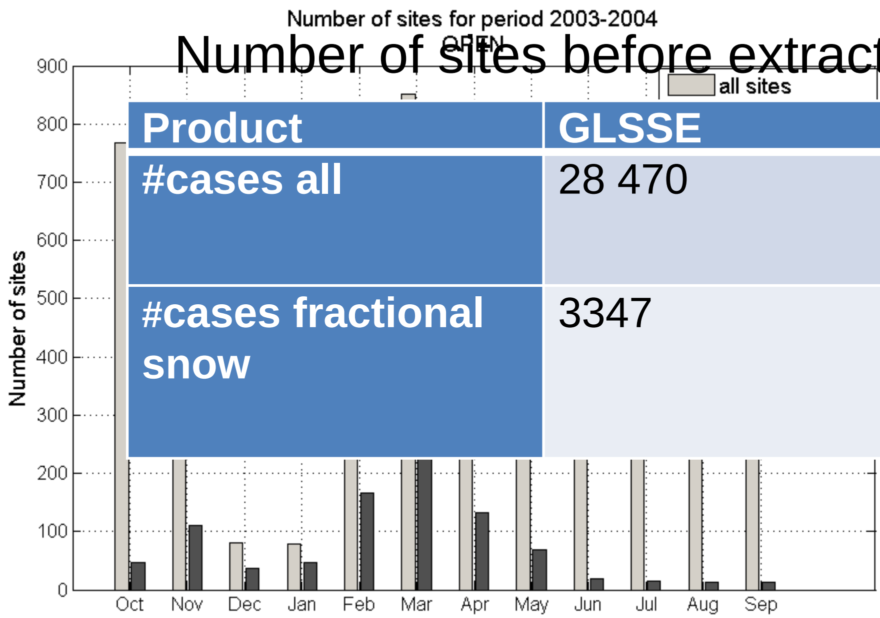
Not all the in-situ sittes are not utilized because of the need of overlap:

Black triangles describe the applied in-situ sites (crosses are the whole ECMWF set)

Number of intersecting sites for the analysis with ECMWF in-situ data (period 2003-2004)

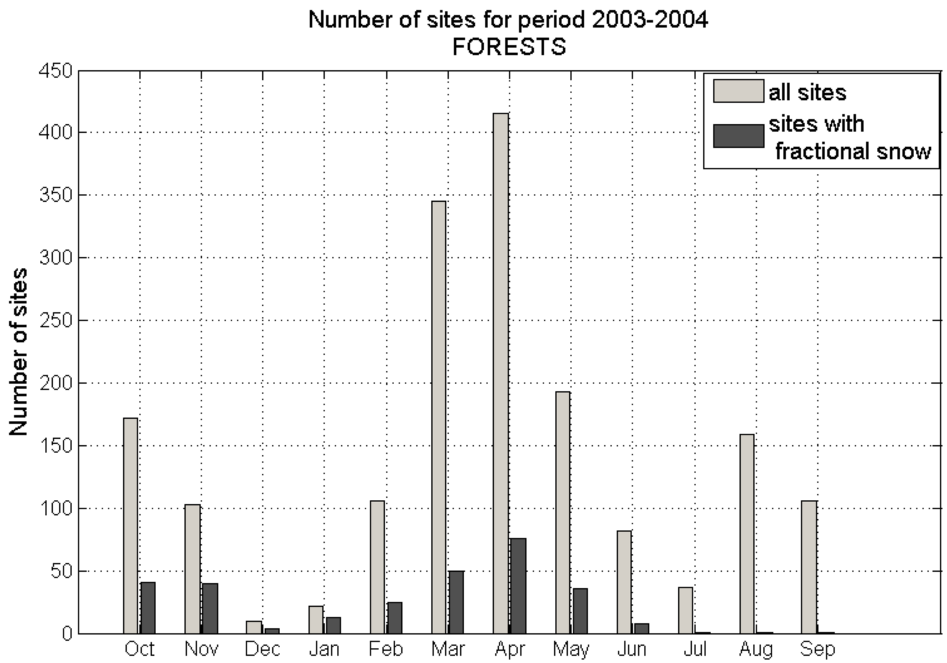
Number of sites before extracting only the overlapping data

Product	GLSSE	M10C05	PATHF
#cases all	28 470	142 109	225 402
#cases fractional snow	3347	17954	55857



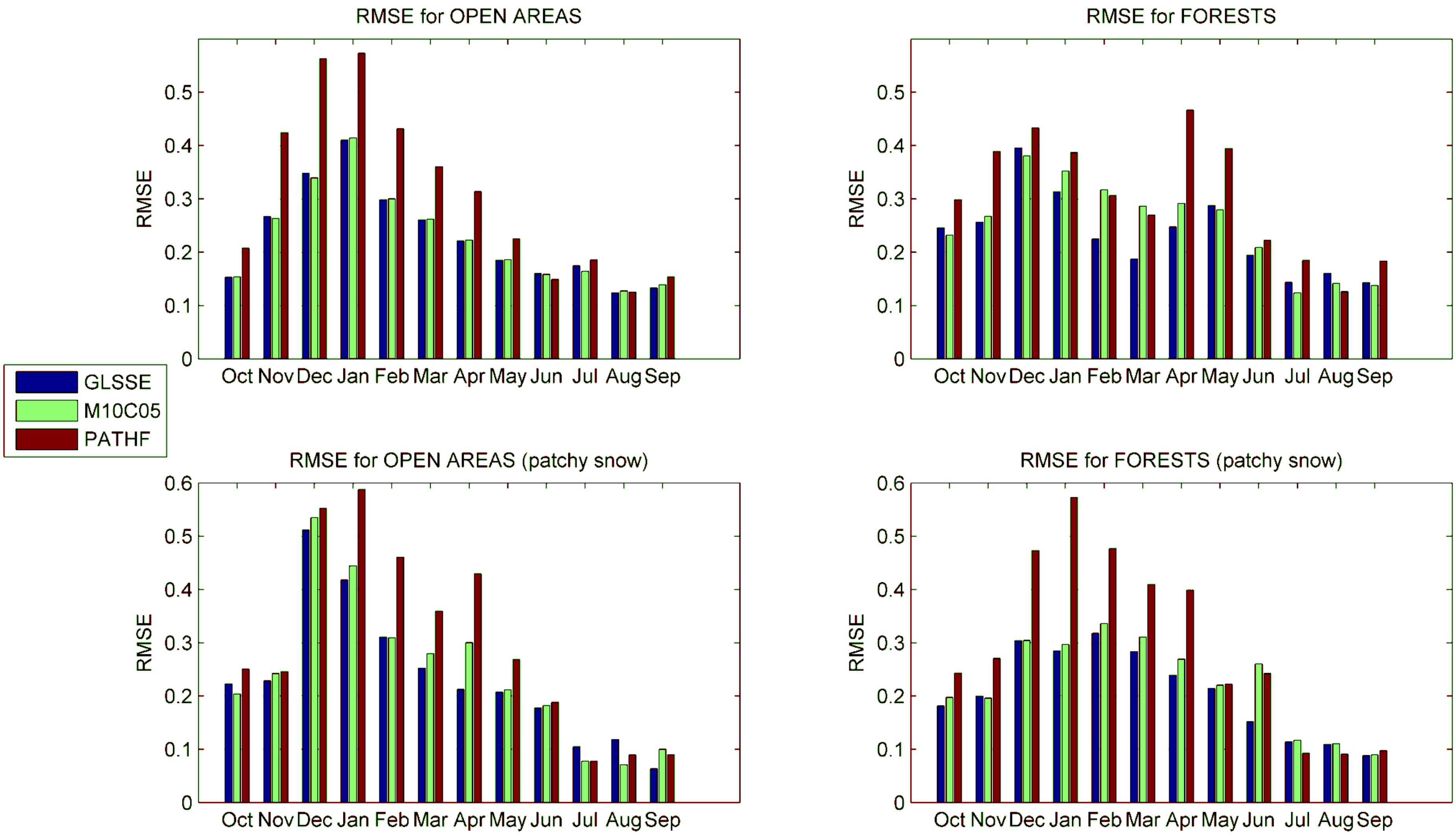
N Cases = 12 189

N cases (frac) = 1803



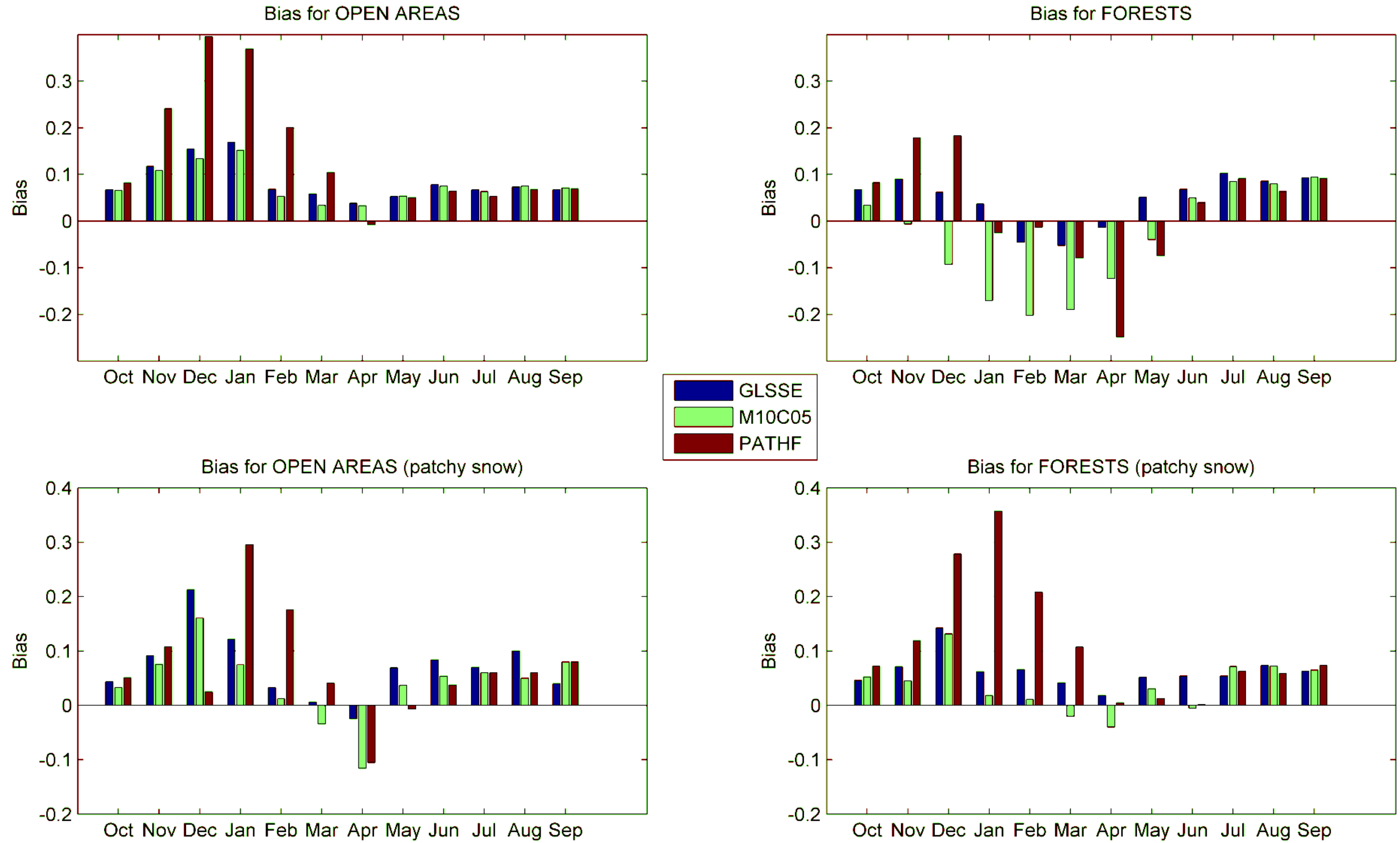
Preliminary validation results using the Extended ECMWF insitu data on SD

The three products providing direct SCF are evaluated: GLSSE, MOD10C05 and PATHF. Analysis made for temporally and sptatially intersecting cases (2003-2004)



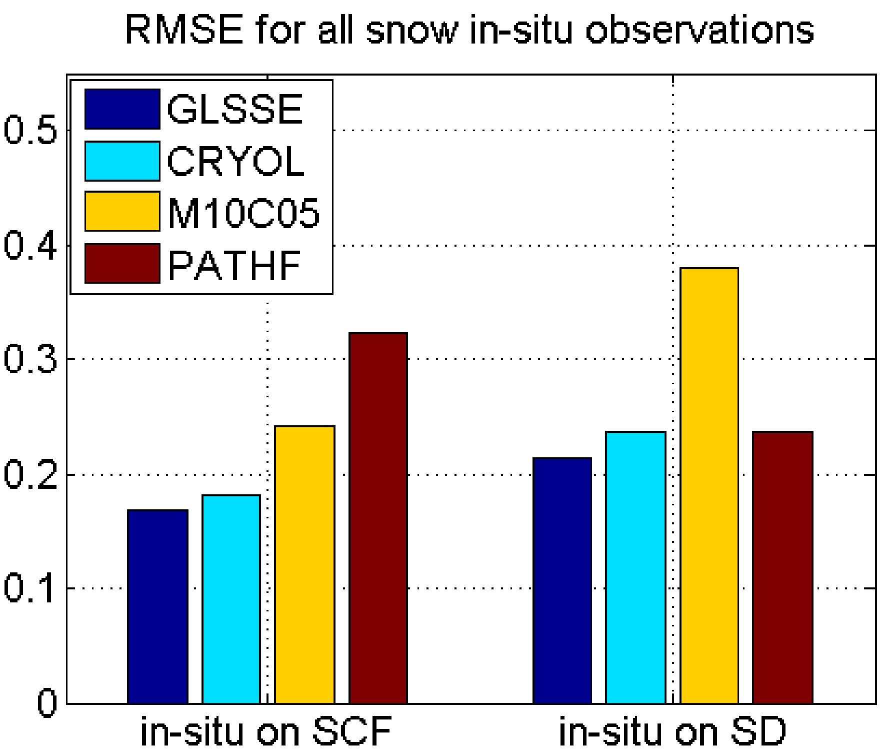
Preliminary validation results using the Extended ECMWF insitu data on SD

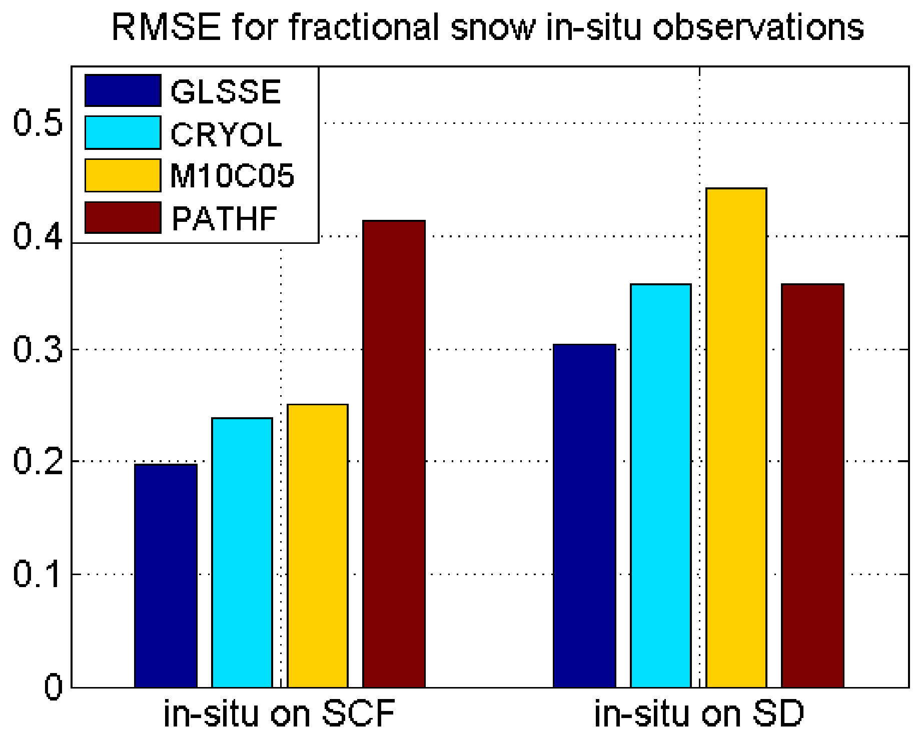
The three products providing direct SCF are evaluated: GLSSE, MOD10C05 and PATHF. Analysis made for temporally and sptatially intersecting cases (2003-2004)



The effect on SD → SCF conversion on the validation results

- An analysis made for Finnish Snow course data (period 2003-2004)
- The switch between the applied data type changes the result particularly for fractional snow





→ We should discuss the validity of the conversions on NH scale