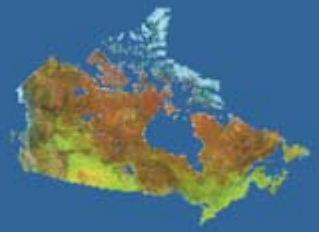
GEMS



A REMOTE PREDICTIVE MAPPING (RPM) APPROACH for the MAPPING of SURFICIAL MATERIALS NORTH of 60º

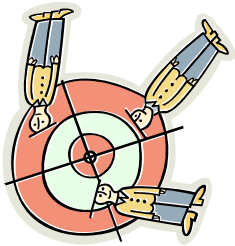
J.R. Harris
Geological Survey of Canada

Geomapping for Energy and Minerals

- Remote Predictive Mapping (RPM)
- Mapping large areas of Canada's North
- Bedrock and surficial geology to aid mineral and energy exploration

A newer approach to image classification

- An new algorithm for classifiying features – the **“ROBUST CLASSIFICATION METHOD” (RCM)**
  - A pixel-based classification method
  - This is an **OPERATIONAL** process!
  - Published in the Canadian Journal of Remote Sensing (CJRS) – Vol.38, No. 2 pp. 1-22



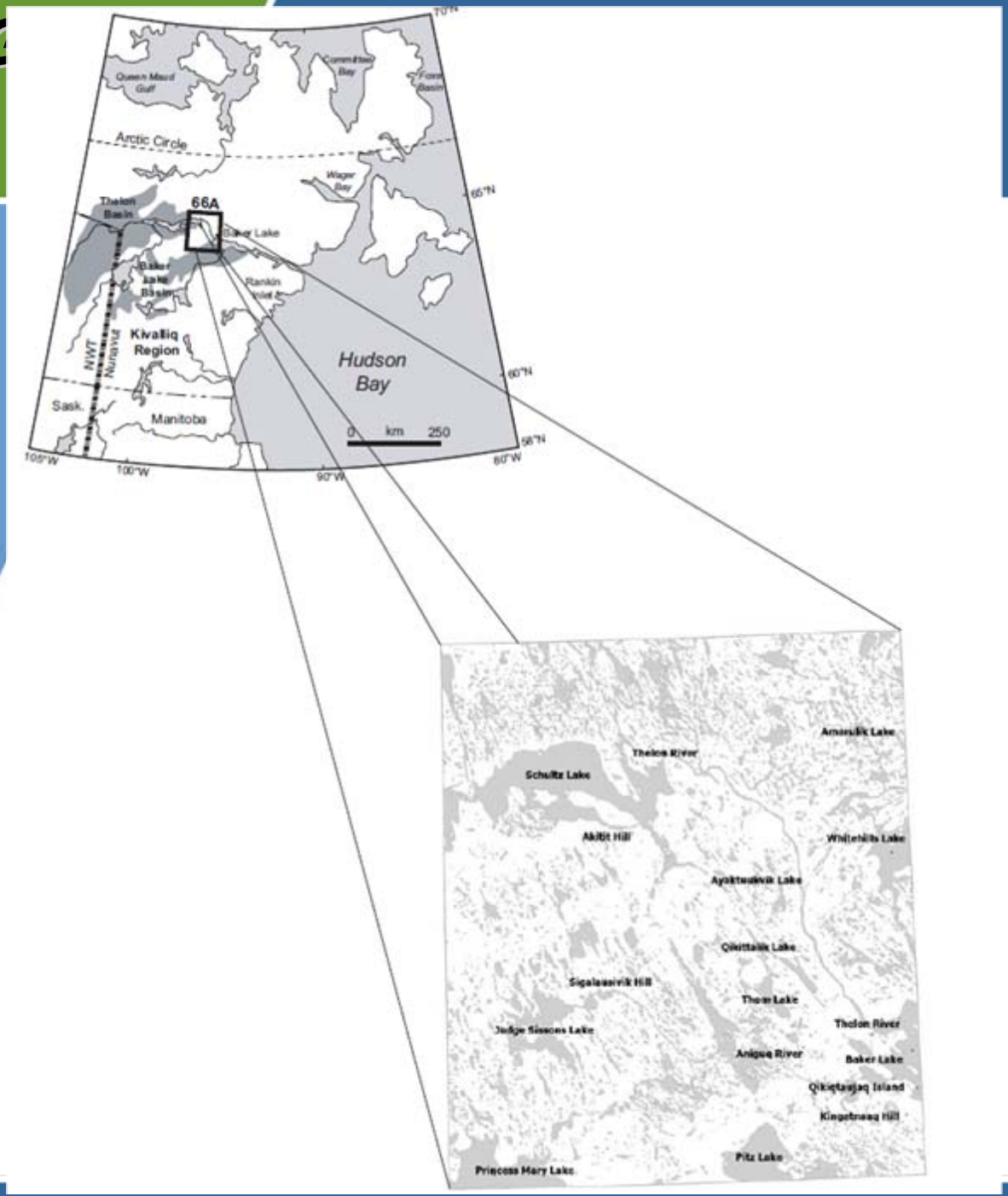
SUMMARY

- Algorithm is based on a randomized and repeated sampling of a training dataset in concert with traditional cross-validation of the classification results
- A series of predictions (classified maps) and associated uncertainty maps and statistics are produced.

SUMMARY (con't)

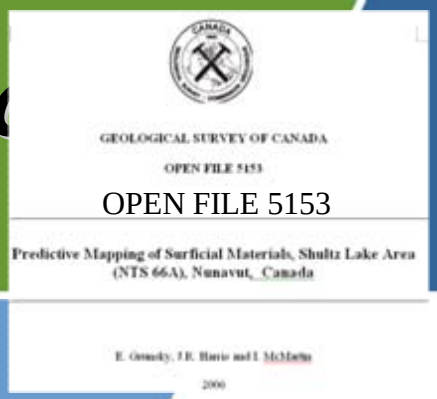
- Similar to Monte Carlo simulation – iteration vs. repetition
- Incorporates aspects of bagging, boosting and ensemble classification methods
  - Boosting/bagging – overall classification accuracy can be improved by repetitive sampling (with replacement) of a training dataset
  - Aggregated prediction (majority classification) is produced (i.e. voting)
  - In addition – an uncertainty map is also produced

A CASE STUDY



Predictive mapping of surficial materials using LANDSAT 7 data over NTS66A (1:250,000) SHULTZ LAKE , Nunavut

Previous work – Grunsky, McMartin and Harris


GEOLOGICAL SURVEY OF CANADA

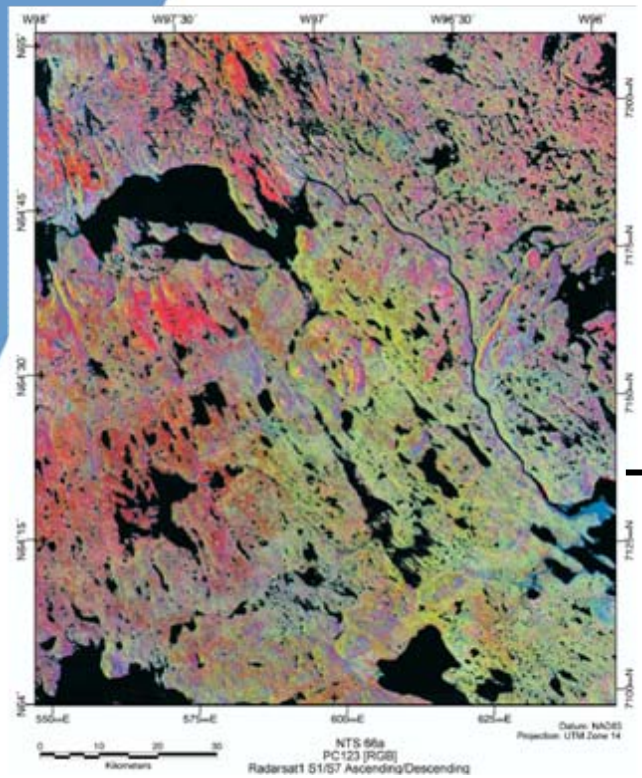
OPEN FILE 5153

OPEN FILE 5153

Predictive Mapping of Surficial Materials, Shultz Lake Area (NTS 66A), Nunavut, Canada

E. Grunsky, J.E. Harris and I. McMartin

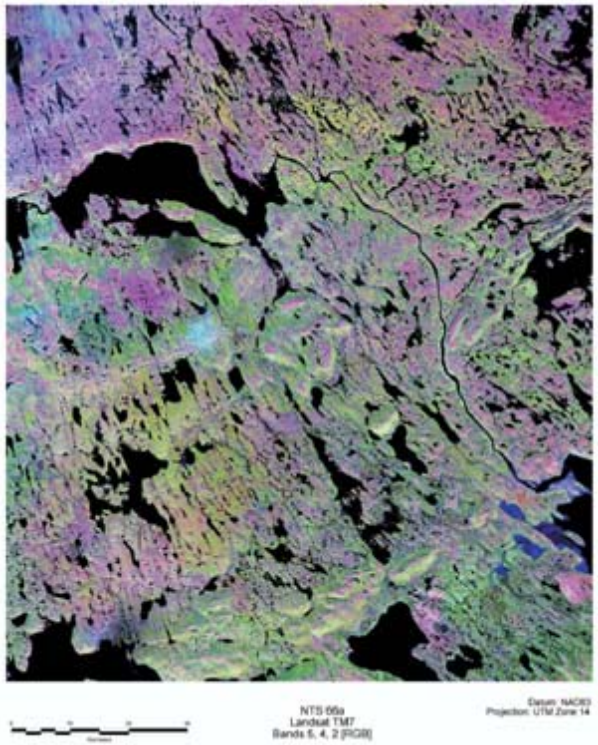
2000



Multi-beam RADARSAT I

83.1% accuracy

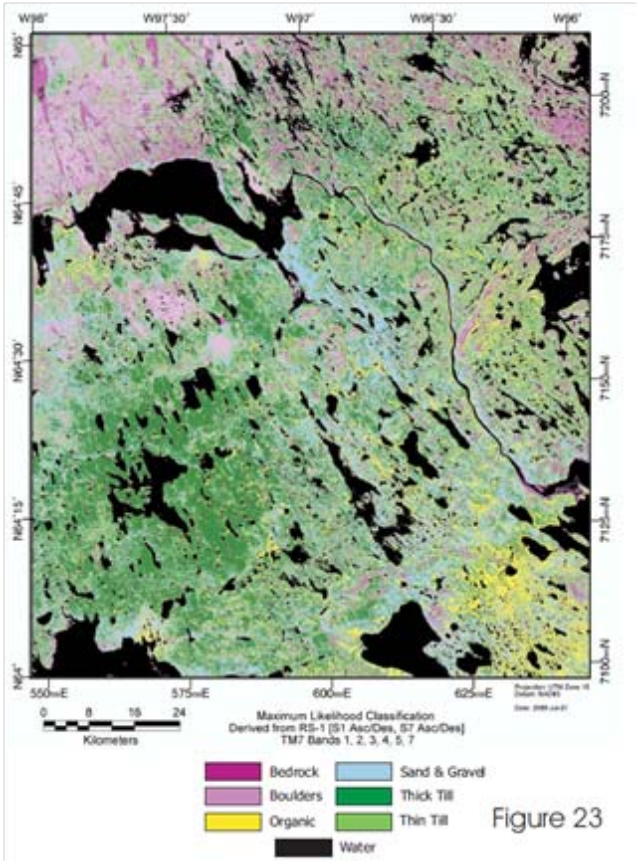
+



LANDSAT TM

84.8% accuracy

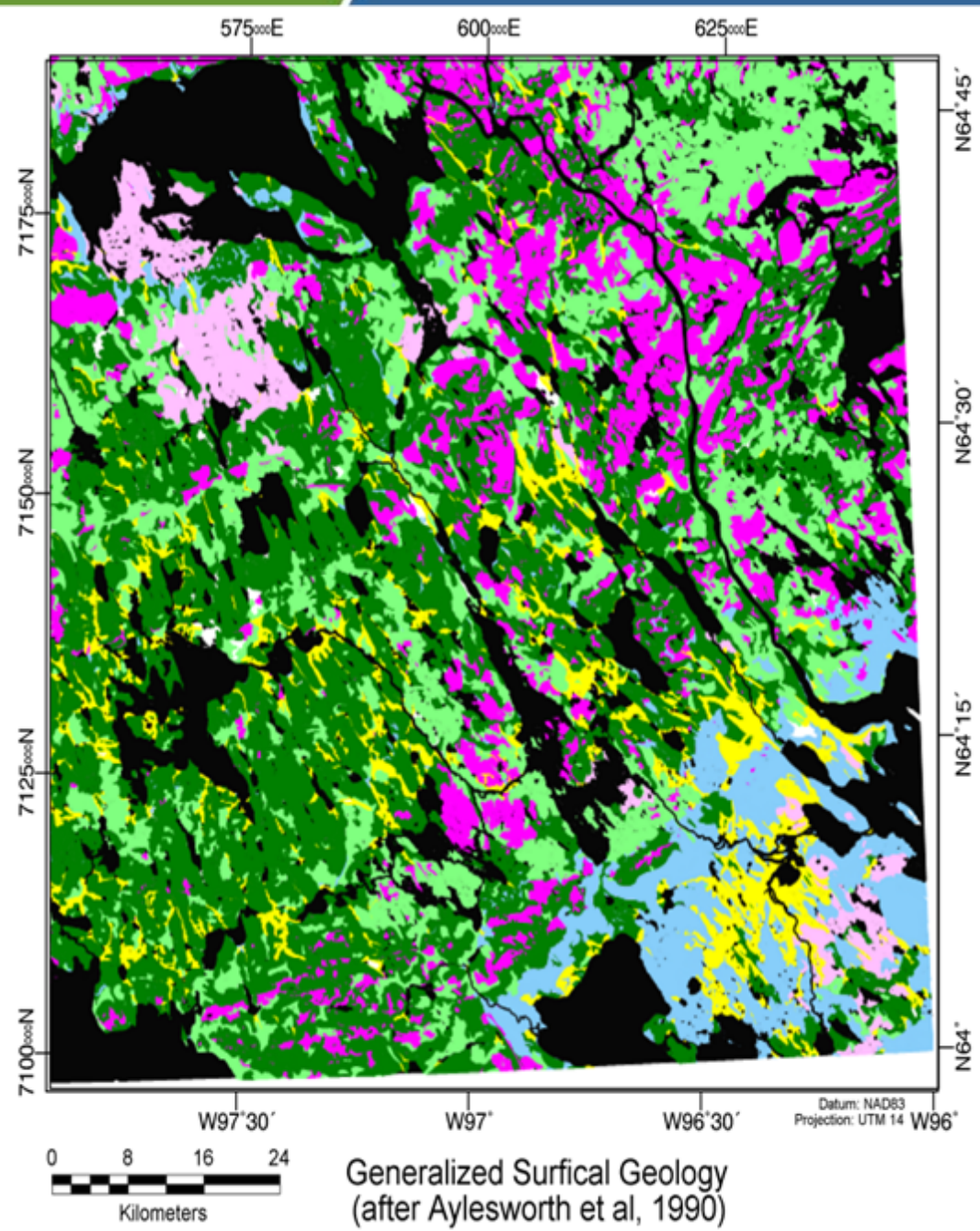
➡



Classification (Maximum Likelihood)

87.8% accuracy

What are we trying to predict (classify)?



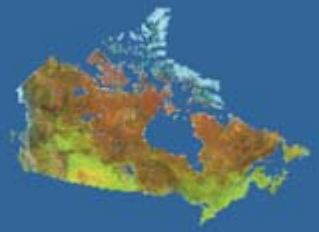
Generalized Surfical Geology
(after Aylesworth et al, 1990)

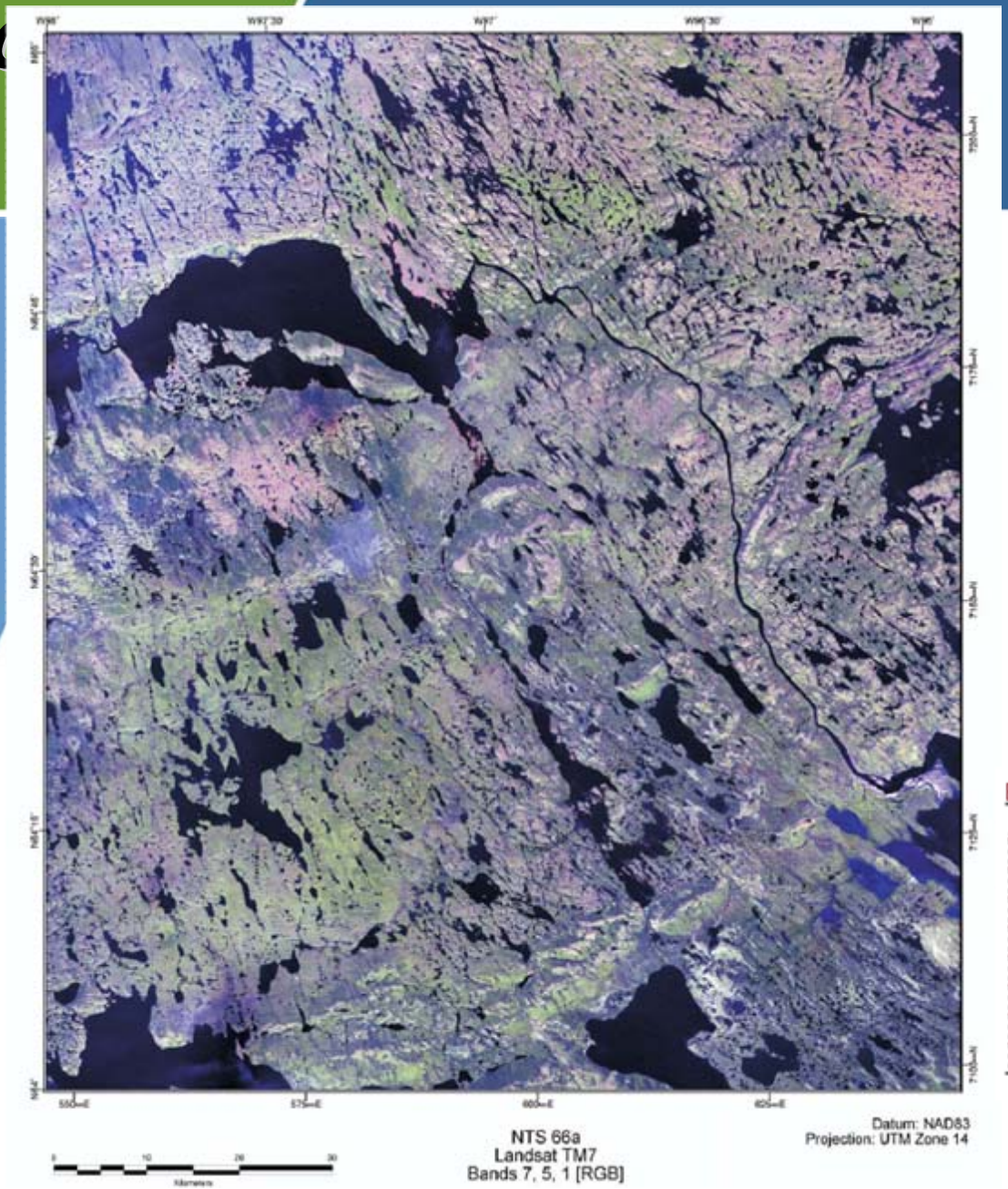
Generalized Surficial Materials (not process)!

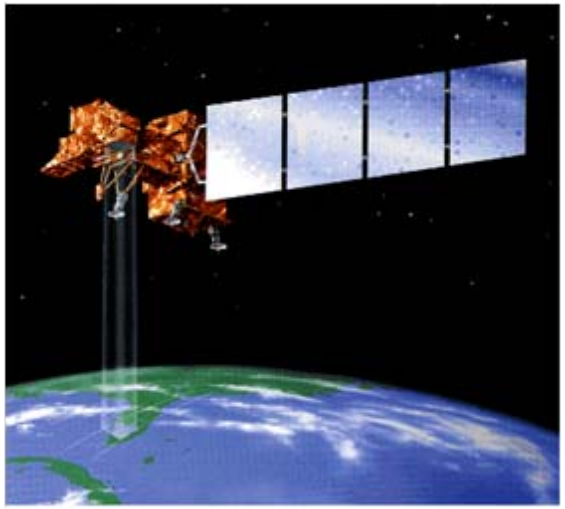
Training Areas

Surface material	Training Area (# of pixels)	Training Area (# of polygons)
Bedrock	26,492	8
Boulders	115,319	12
Organic	53,002	8
Sand and gravel	219,895	10
Thick till	317,863	11
Thin till	149,330	12
Water	1,537,784	12
Clouds	46,092	1

THE DATA: LANDSAT 7 TM



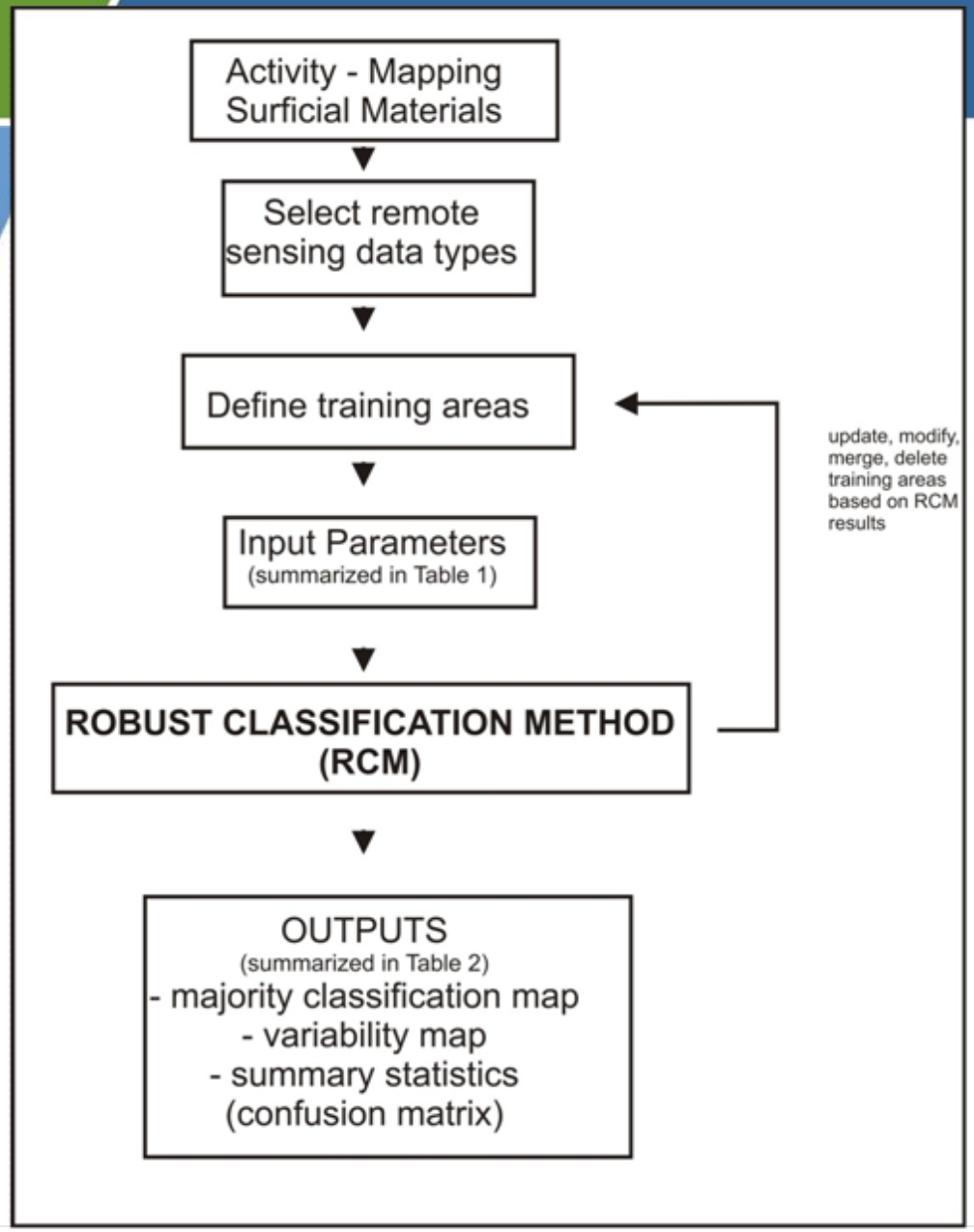




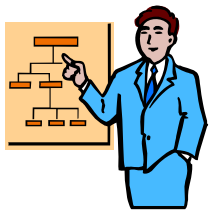
ETM+ Bands

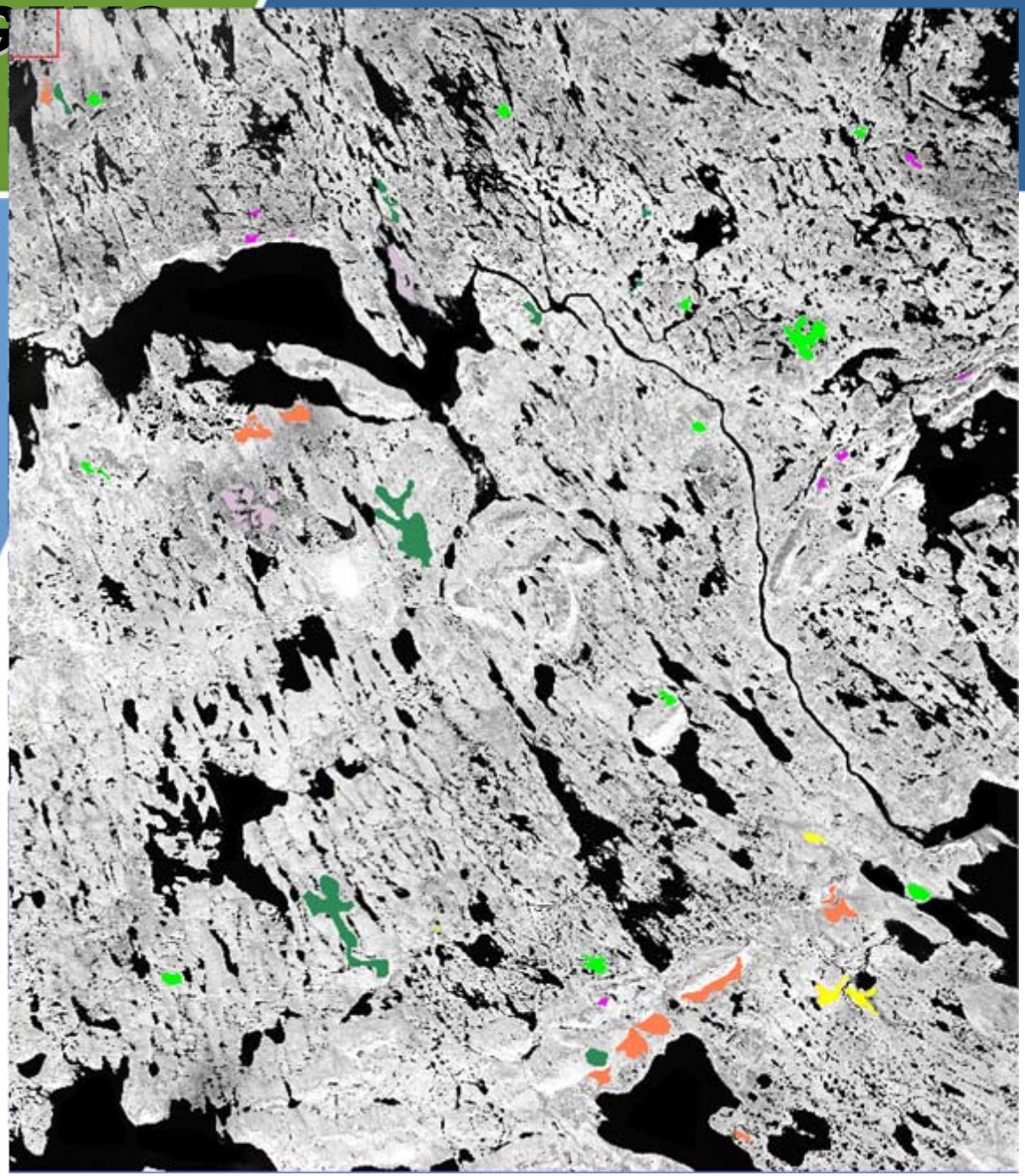
Band Number	µm	Resolution
1	0.45-0.515	30 m
2	0.525-0.605	30 m
3	0.63-0.69	30 m
4	0.75-0.90	30 m
5	1.55-1.75	30 m
6	10.4-12.5	60 m
7	2.09-2.35	30 m
8	0.52-0.9	15 m

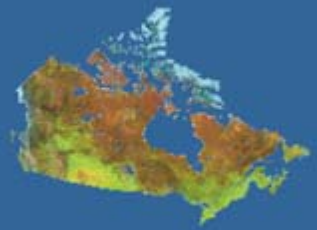
THE PROCESS



The adjacent flow chart summarizes the approach we have adopted. It is based on supervised classification of remotely sensed data using training areas of each surficial material type



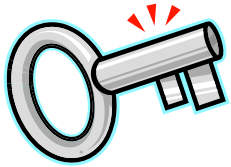




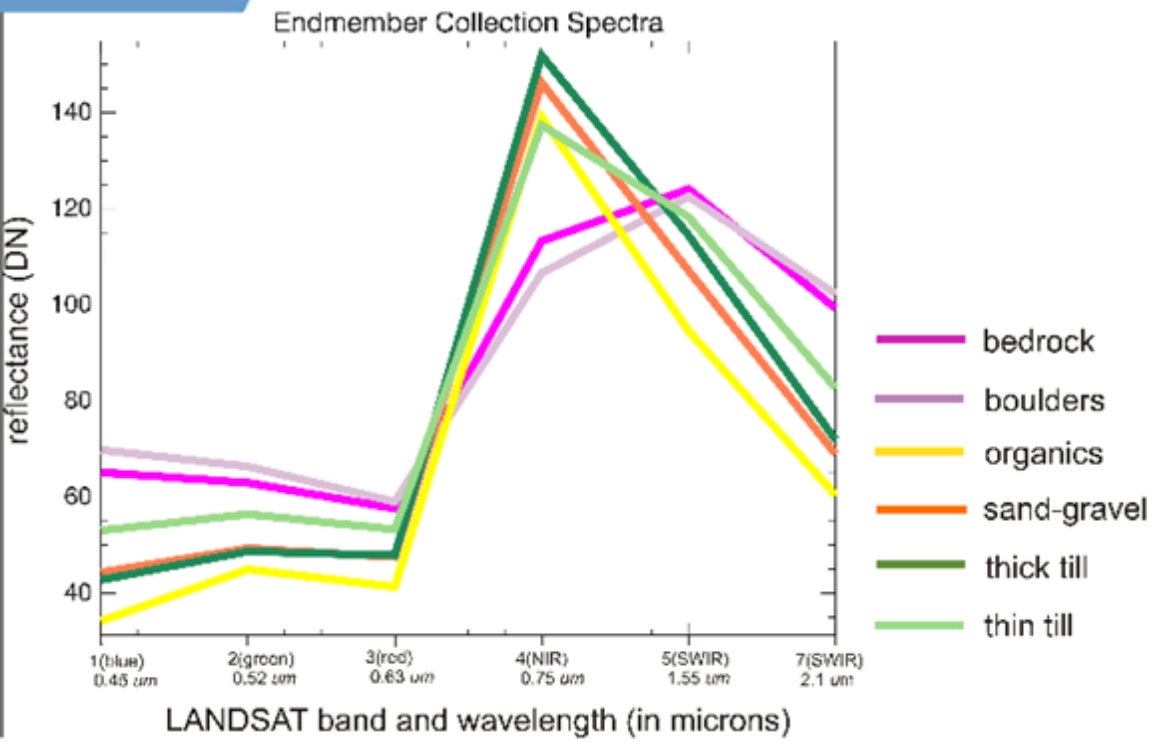
The KEY INPUT !

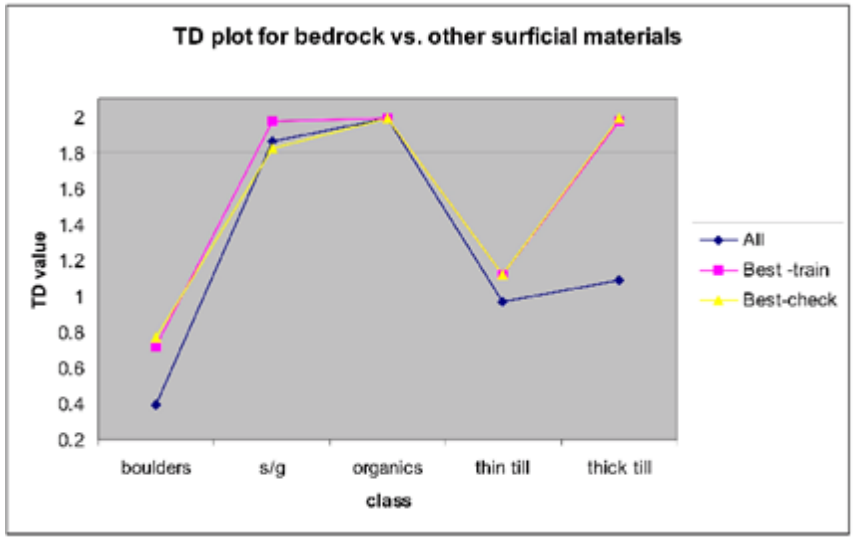
Representative **training areas** (ROIS) for each surficial material

bedrock	Magenta	26,492	8/26,492
boulders	Thistle	115,319	12/115,319
cloud	White	46,092	1/46,092
organics	Yellow	53,002	8/53,002
sand_gravel	Coral	219,895	10/219,895
thick_till	Sea Green	317,863	11/317,863
thin_till	Green	149,330	12/149,330
water	Black	1,537,784	12/1,537,784

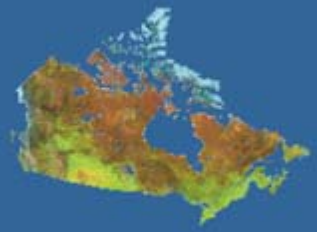


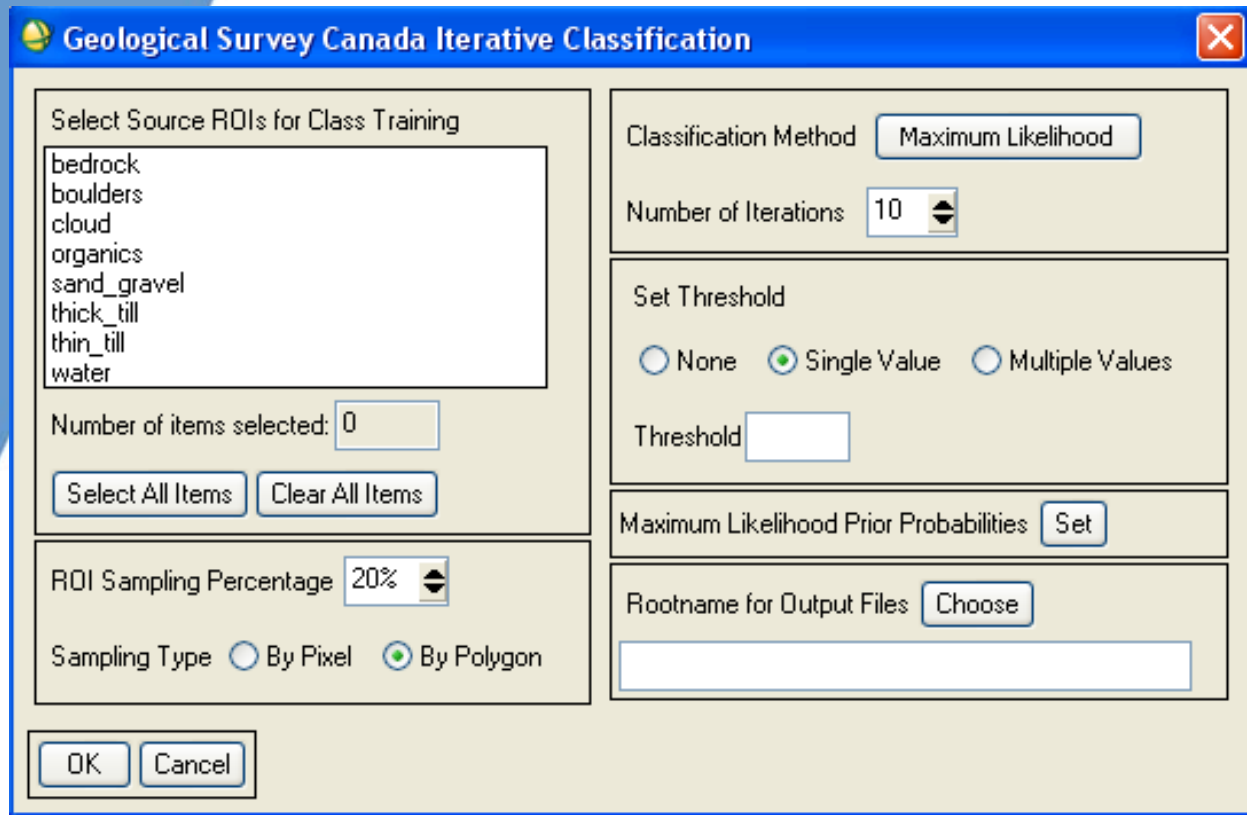
Separability of training areas

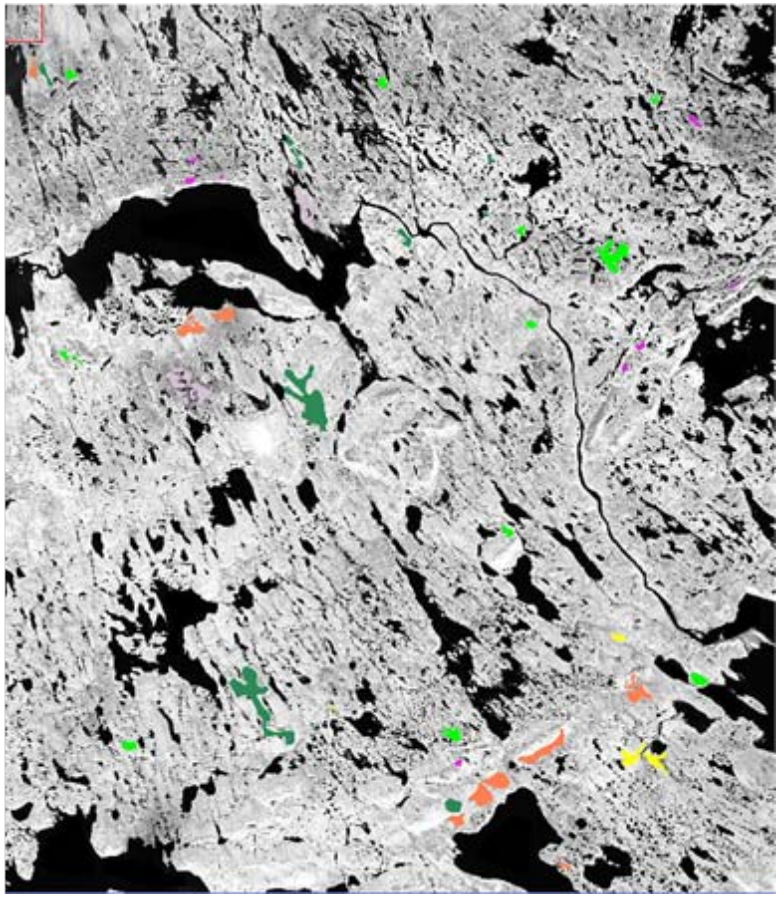




THE ALGORITHM

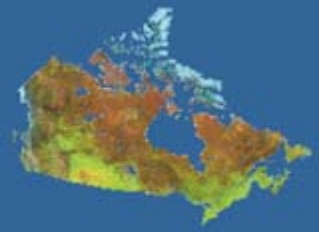






Programmed in IDL within the ENVI image analysis software environment

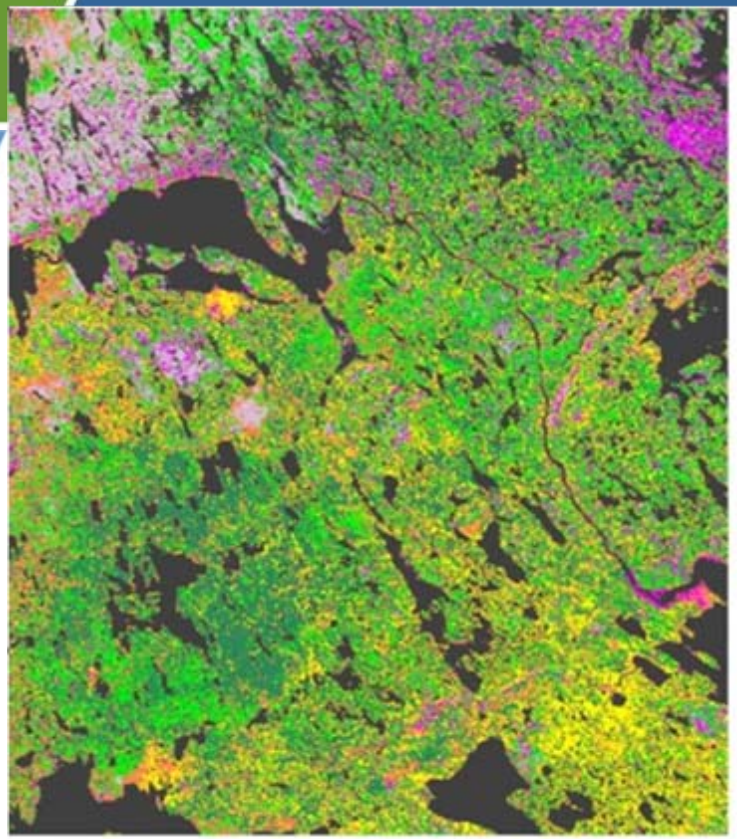
RCM – OUTPUT PRODUCTS



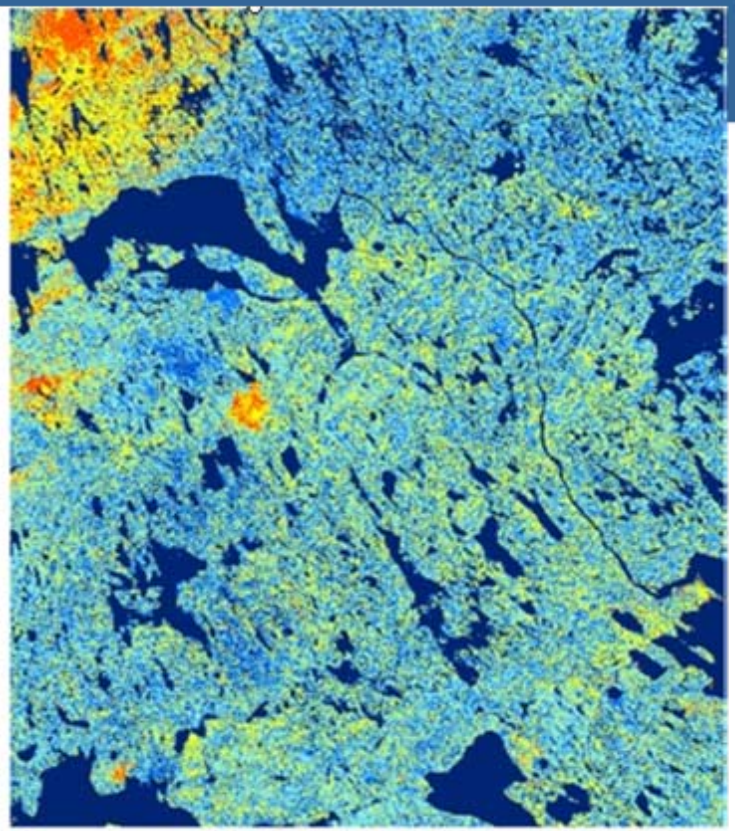
- **Majority Classification Map** (class and percent)
- **Variability Map** – (uncertainty map)
- **Rule Maps** (probability images) for each class – (average, best and worst) – (uncertainty maps)
- **Statistics** – confusion analysis (between classified map and independent training (check) areas

GEMS

Majority Classification and Uncertainty Map



a

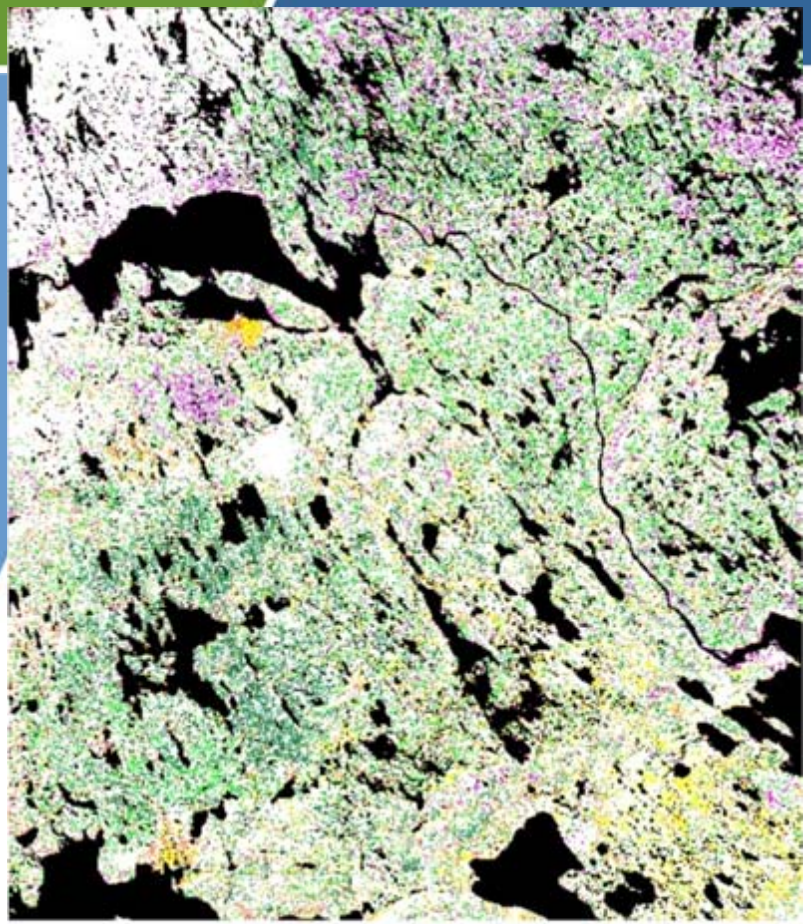


b

Surface Material	Colour	Surface Material	Colour
Bedrock		Thick till	
Boulders		Thin till	
Sand - gravel		Water	
Organics			

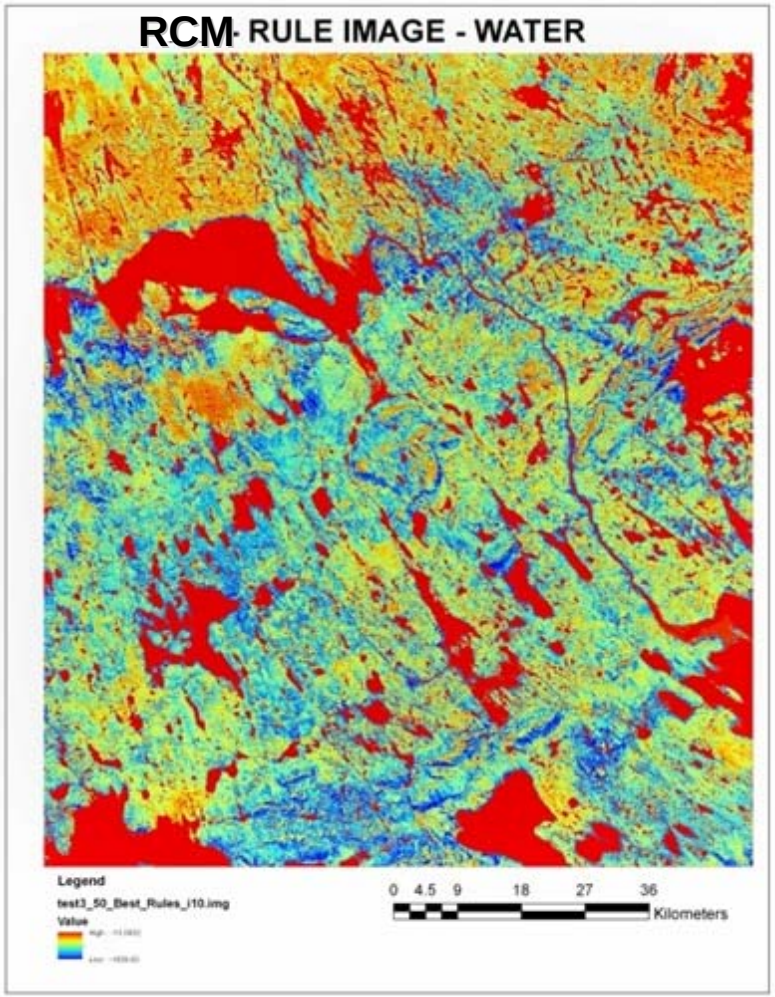
	Class Variability of 1	54.4%
	Class Variability of 2	18.9%
	Class Variability of 3	10.5%
	Class Variability of 4	7.0%
	Class Variability of 5	5.1%
	Class Variability of 6	3.9%

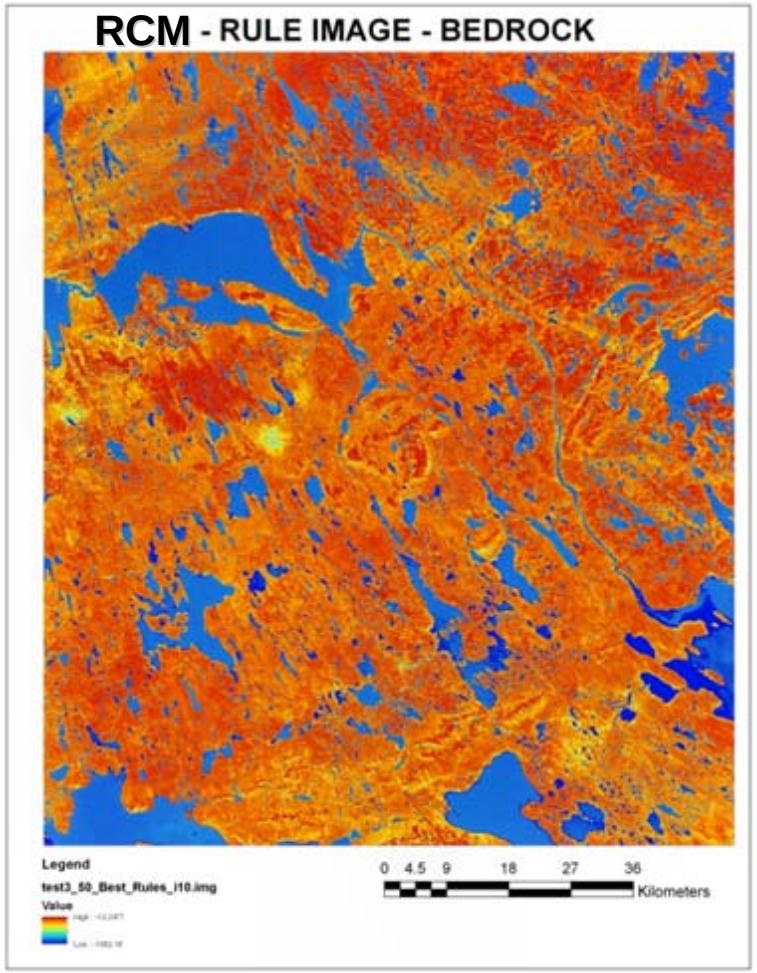
Uncertainty removed



The variability image was used to mask out areas of high uncertainty in the classification process (i.e. areas that were characterized by greater than 4 (> than 2 standard deviations) on the variability image (these pixels were classified as greater and equal to 4 different surficial material classes through the 90 repetitions of RCM and thus are considered to be highly variable and thus uncertain – these areas are shown in white).

Surface Material	Colour	Surface Material	Colour
Bedrock		Thick till	
Boulders		Thin till	
Sand - gravel		Water	
Organics			



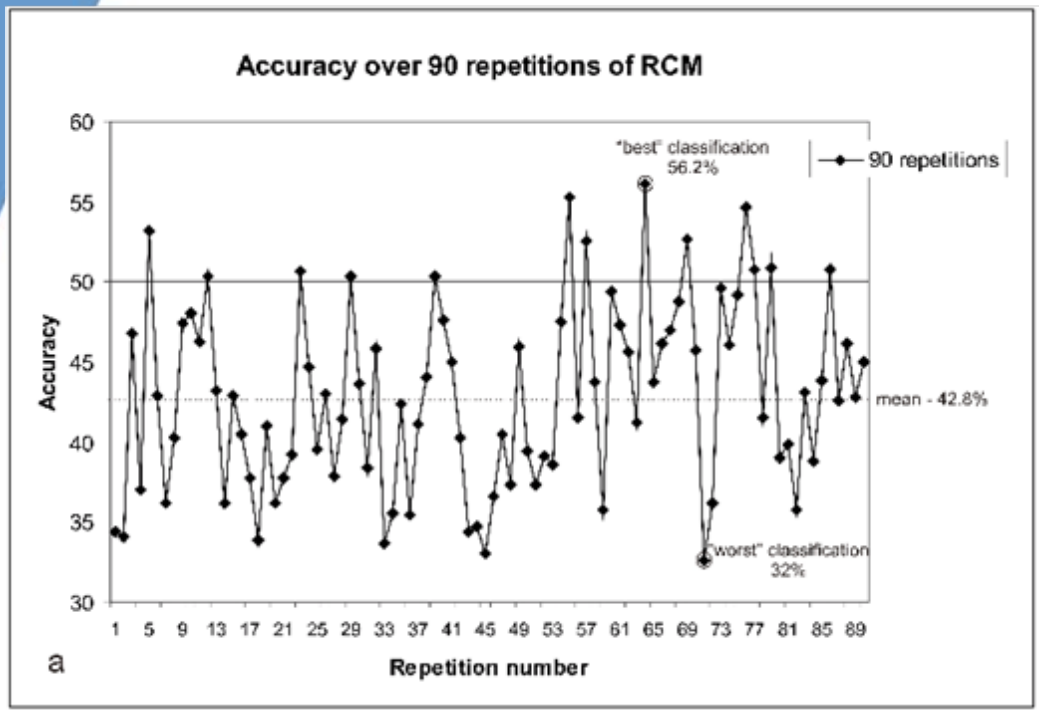


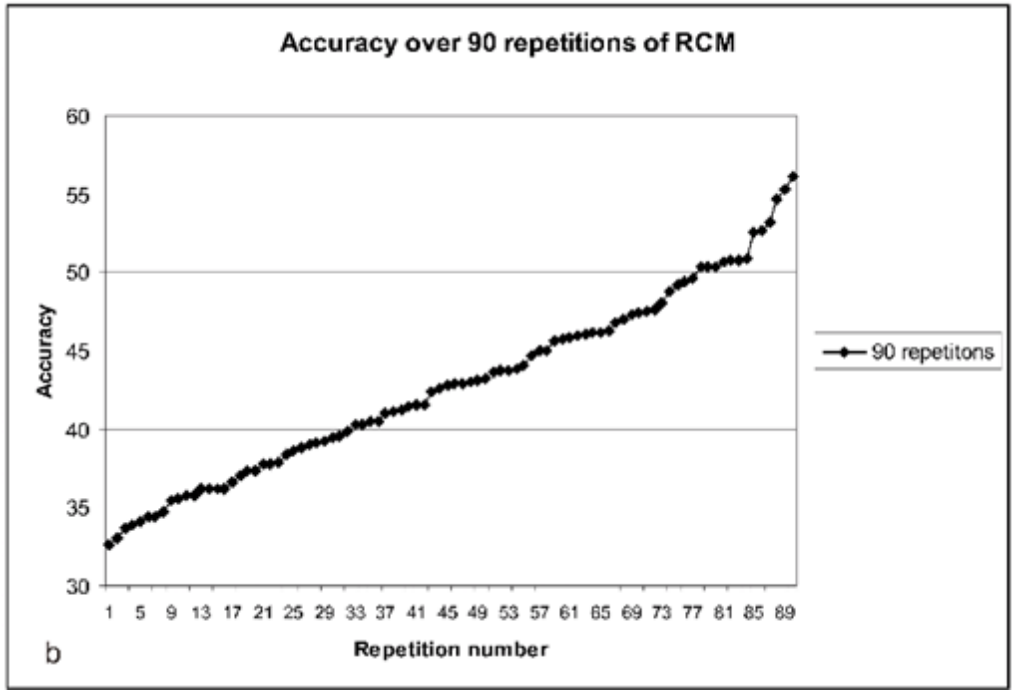
Rule (probability) images)

– dealing with
the certainty of
the classification
of each surficial material
- Fuzzy RGB maps!

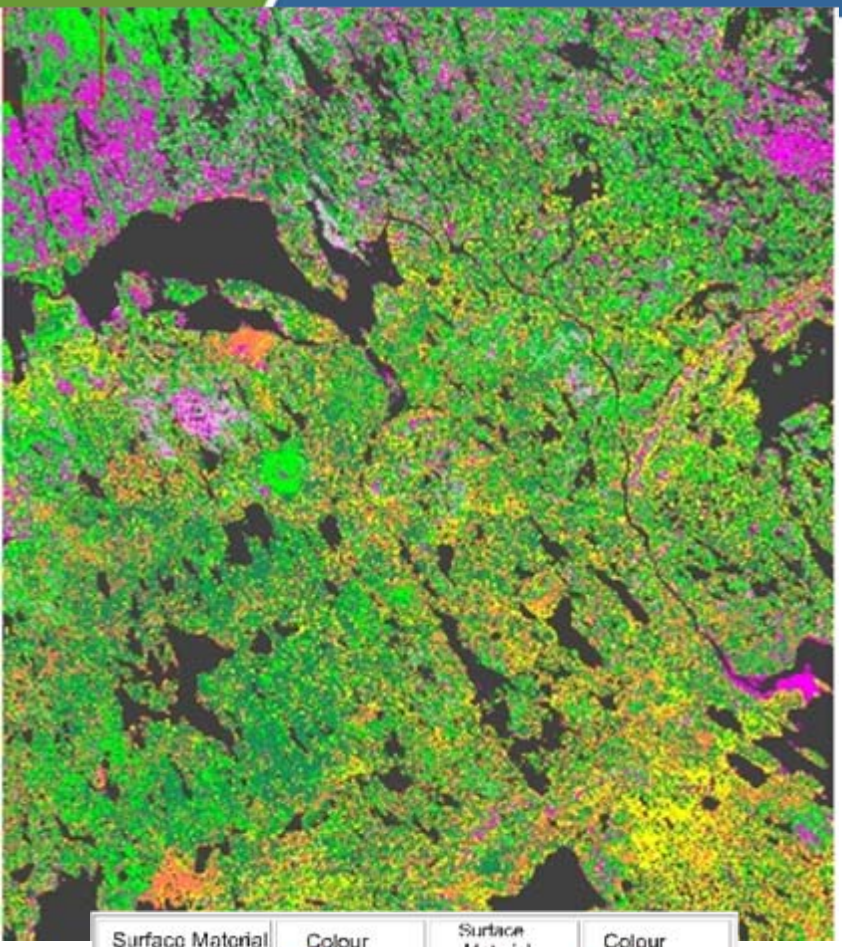
Variability in classification accuracy!

Results of 90 repetitions using randomly chosen training areas for classification and validation



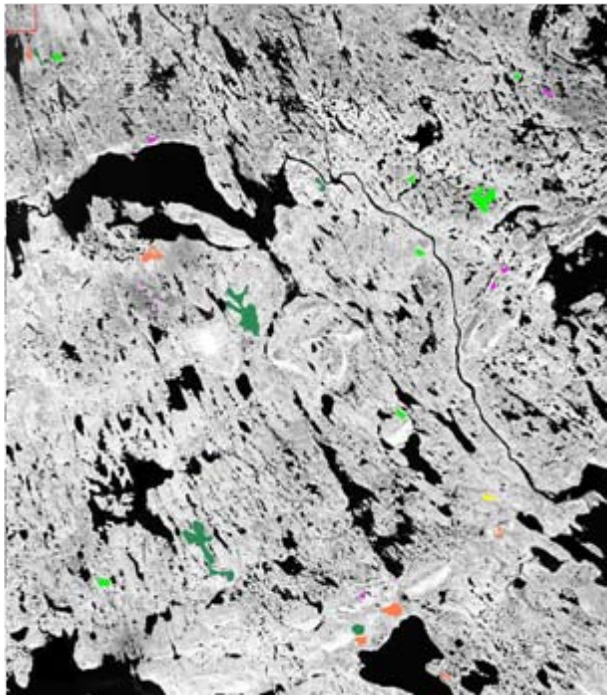


BEST – 56.2%

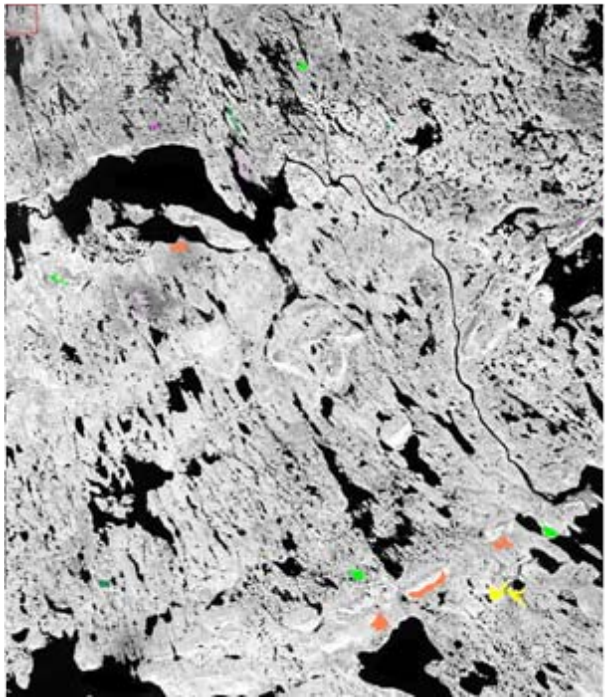


Surface Material	Colour	Surface Material	Colour
Bedrock		Thick till	
Boulders		Thin till	
Sand - gravel		Water	
Organics			

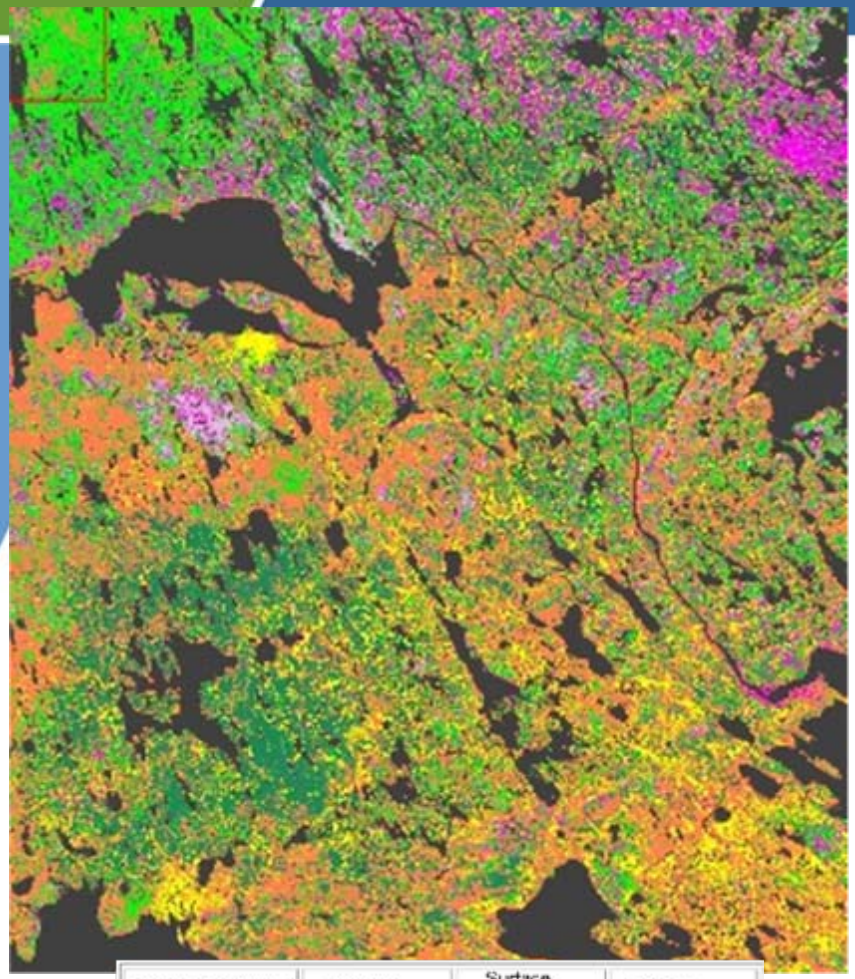
Training areas used for classification



Training areas used for validation

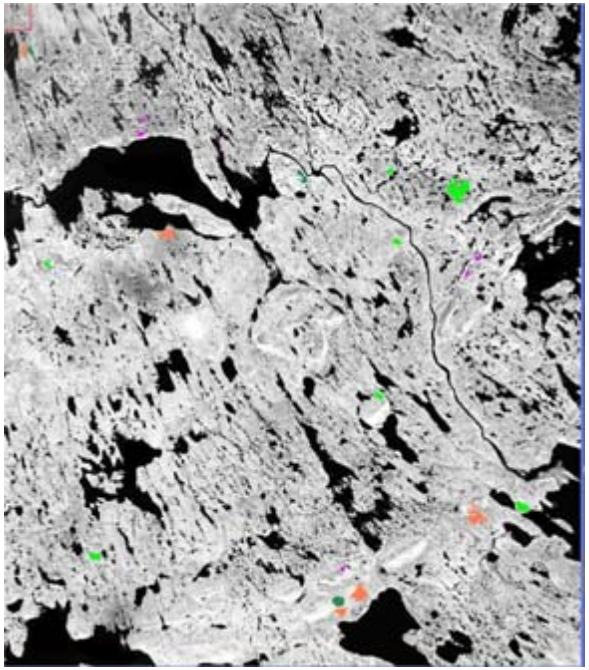


WORST – 32%

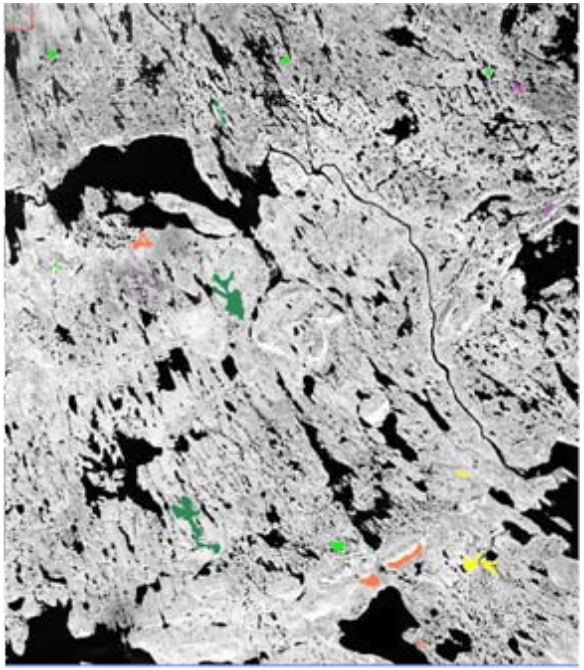


Surface Material	Colour	Surface Material	Colour
Bedrock		Thick till	
Boulders		Thin till	
Sand - gravel		Water	
Organics			

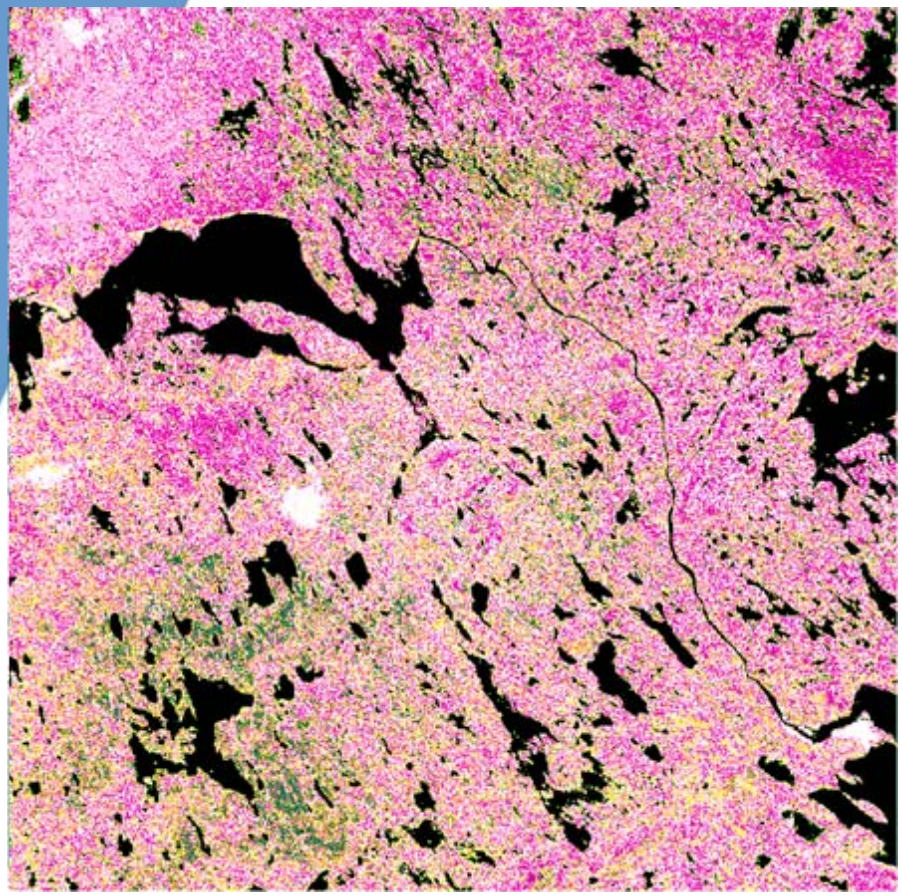
Training areas used for classification



Training areas used for validation



An AGREEMENT MAP



Class	Bedrock	Boulders	Organics	Sand/gravel	Thick Till	Thin Till	Total
Bedrock	**72.3**	13.92	1.72	2.13	0.28	13.16	9.87
Boulders	7.37	**78.86**	0	0.17	0	5.6	4.49
Organics	1.42	0.02	**31.39**	21.38	2.55	0.88	11.97
Sand/gravel	0.14	0	53.25	**26.69**	6.6	6.35	17.96
Thick Till	0	0	0.41	38.44	**78.34**	2.87	31.84
Thin Till	18.77	7.2	13.23	11.19	12.24	**71.14**	23.87
Total	100	100	100	100	100	100	100

Table 4 – Comparison between best and worst classification as determined from RCM. *Best* and *worst* are defined by the highest and lowest overall classification accuracy over 90 repetitions of RCM (see Fig. 5). X and Y axes represent the worst and best classifications (Fig 6a vs. 7a), respectively. Numbers are in percent. The overall agreement between the best and worst classifications is 59.7%. (sum of diagonal numbers (in bold) / 6) (72.3 + 78.86 + 31.39 + 26.69 + 78.34 + 71.4 / 6) = 59.7%.

Much variability between best and worst!
~ 17%!

Assessment of classification accuracies

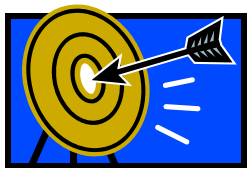
iteration	bedrock	boulders	organic	s/g	Thick till	thin till	Average Accuracy - material
1	31.6	76.9	74.8	28.4	34	23.2	44.8
2	51.7	67.4	54.9	22.6	54.5	33	47.3
3	36.9	50.3	51.3	24.9	89.4	8.3	43.5
4	56	37.2	54.1	33.4	56	41	46.2
5	56.9	40.2	59.3	30.8	68.7	38.7	49.1
6	33.9	71.9	60.5	36.5	56.4	22.7	46.9
7	64.6	49.2	61.2	34.2	39	33	46.8
8	37.5	70.5	21	26.7	53.2	29.8	39.7
9	57.5	61	37.4	44.1	59.9	34.3	49.0
10	27.2	55.7	64.5	32.8	71.1	30.2	46.9
Average Accuracy - iteration	45.3	58.0	53.9	31.4	58.2	29.4	**46.0**
Average Accuracy using variable training areas (see red numbers above)							64.7

Water and clouds are distinct and easily separated from a spectral perspective!

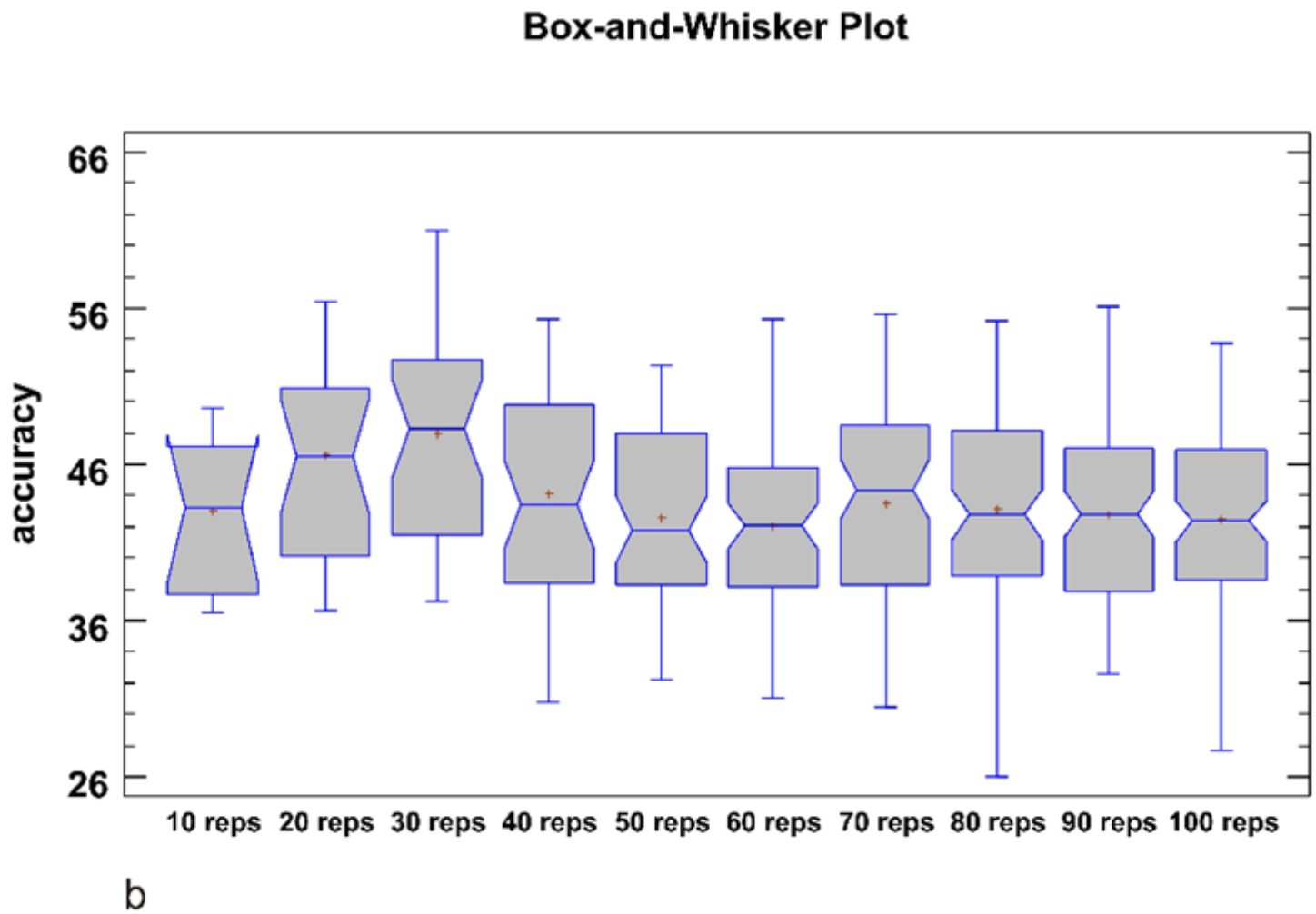
-With clouds and water = **79%**
-With just surficial materials = **46%**
-Variable training areas = **64.7%**

best – boulders, thick till, organics, bedrock
worst – s/g, thin till

Variable training areas! (64.7%)



How many repetitions are required?

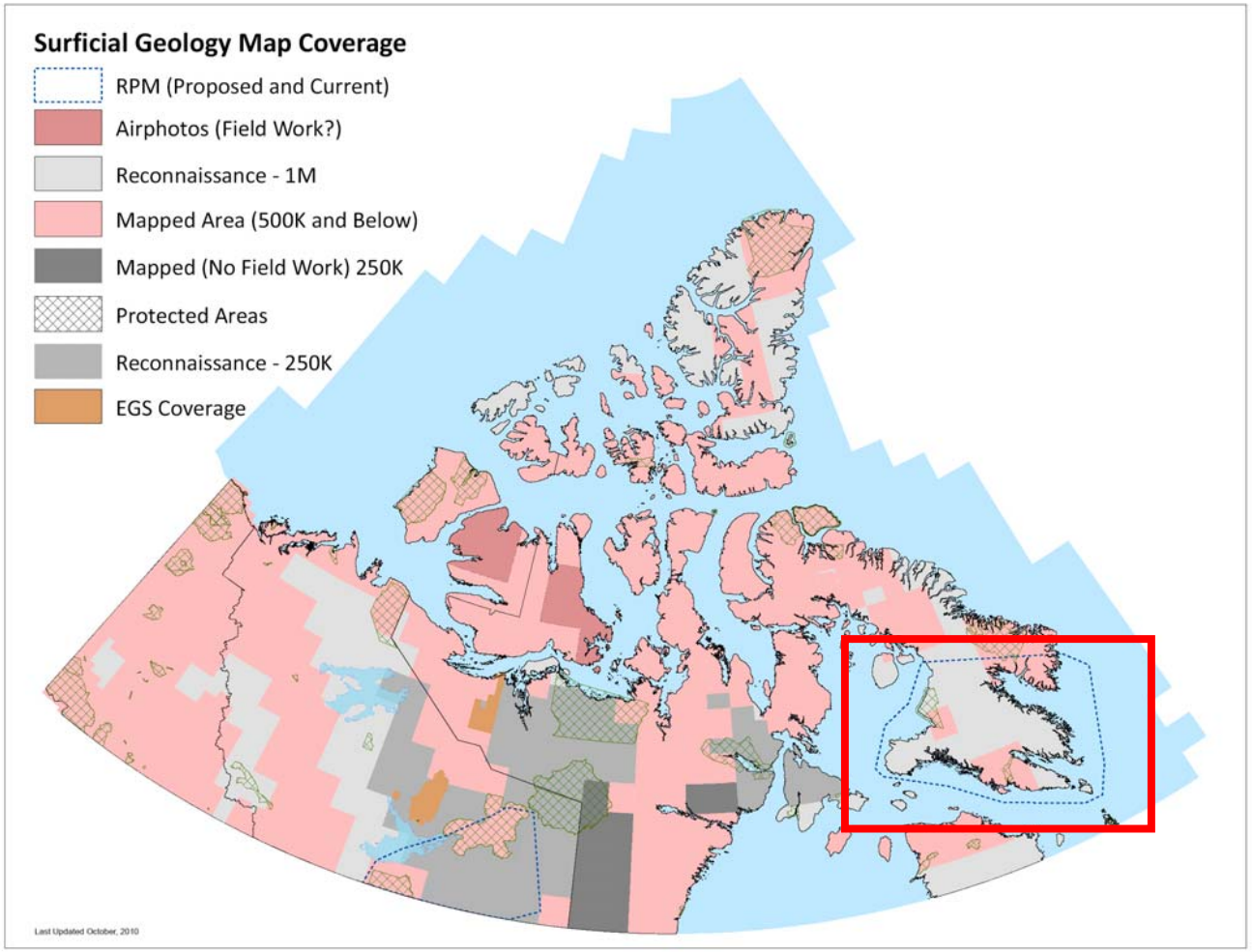


Mean and median values for accuracy stabilizes nominally after 40 repetitions at an average accuracy
of 43.0%

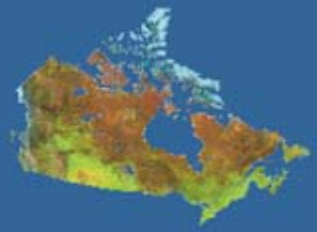
For 90 repetitions of RCM the average classification accuracy was 42.8% with a standard deviation of 5.8%. The 95% confidence around the mean represented an accuracy from 41.5 to 44%.

GEMS

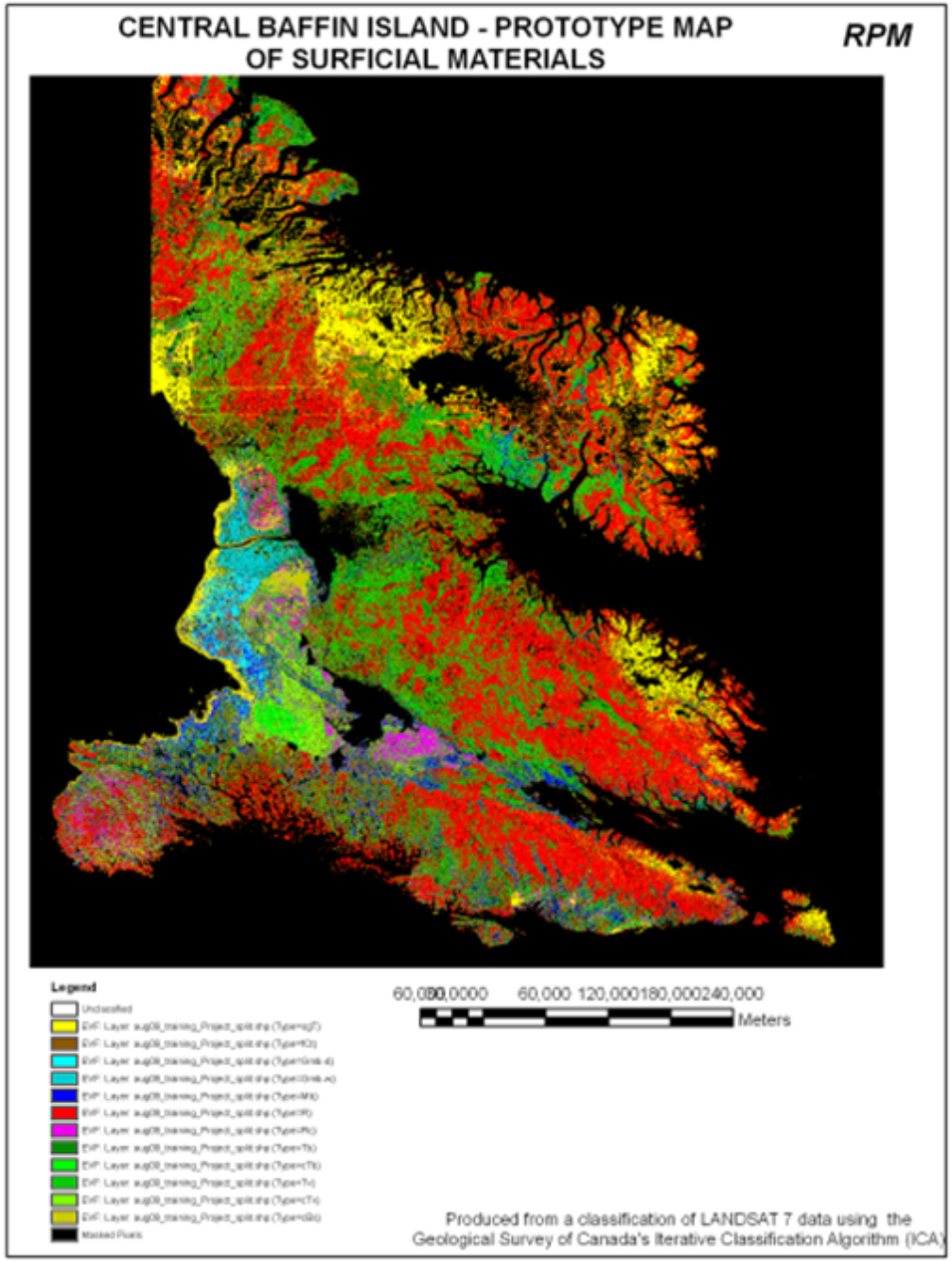
A “GREY SPACE” surficial materials proto-type map

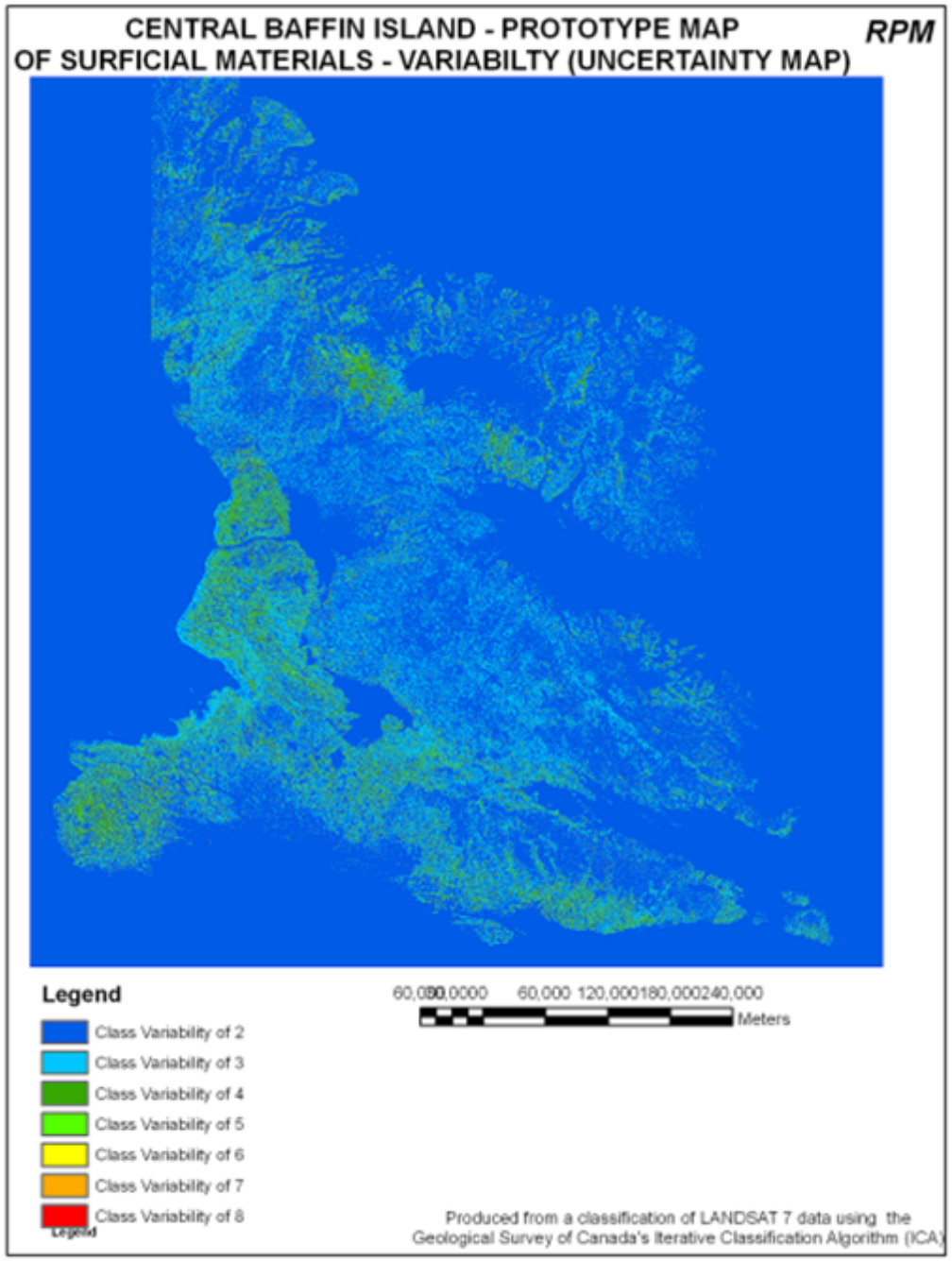


Surficial materials classified



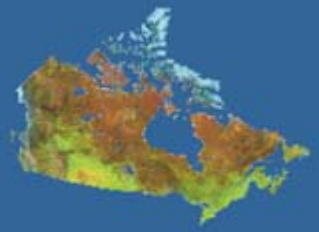
ROI Name	Color	Pixels	Polygons
Till Veneer - GSC- Tb	Green2	1,607	27/1,607
Rock and Rubble - GSC -R	Red	4,504	35/4,504
Deep water proglacial silt (Mb, Mv) - GSC - M	Blue	442	7/442
Sand and Gravel Terrace + GP) - GSC - GF	Yellow	598	19/598
Till Blanket (Th, Tm, Tb)) - GSC -Tb	Green	4,335	52/4,335
Carbonate (till and rock) - GSC - T	Magenta	597	8/597
Carbonate Beach Sediment (Gravel and Sand)) - GSC - Mr	Yellow3	5,024	17/5,024
Glaciomarine Blanket, Dry) - GSC -GMb	Cyan	2,753	3/2,753
Fen Over Carbonate Till (Organics)) - GSC - T	Orange4	254	6/254
Glaciomarine Blanket, Wet)	Cyan3	10,401	4/10,401

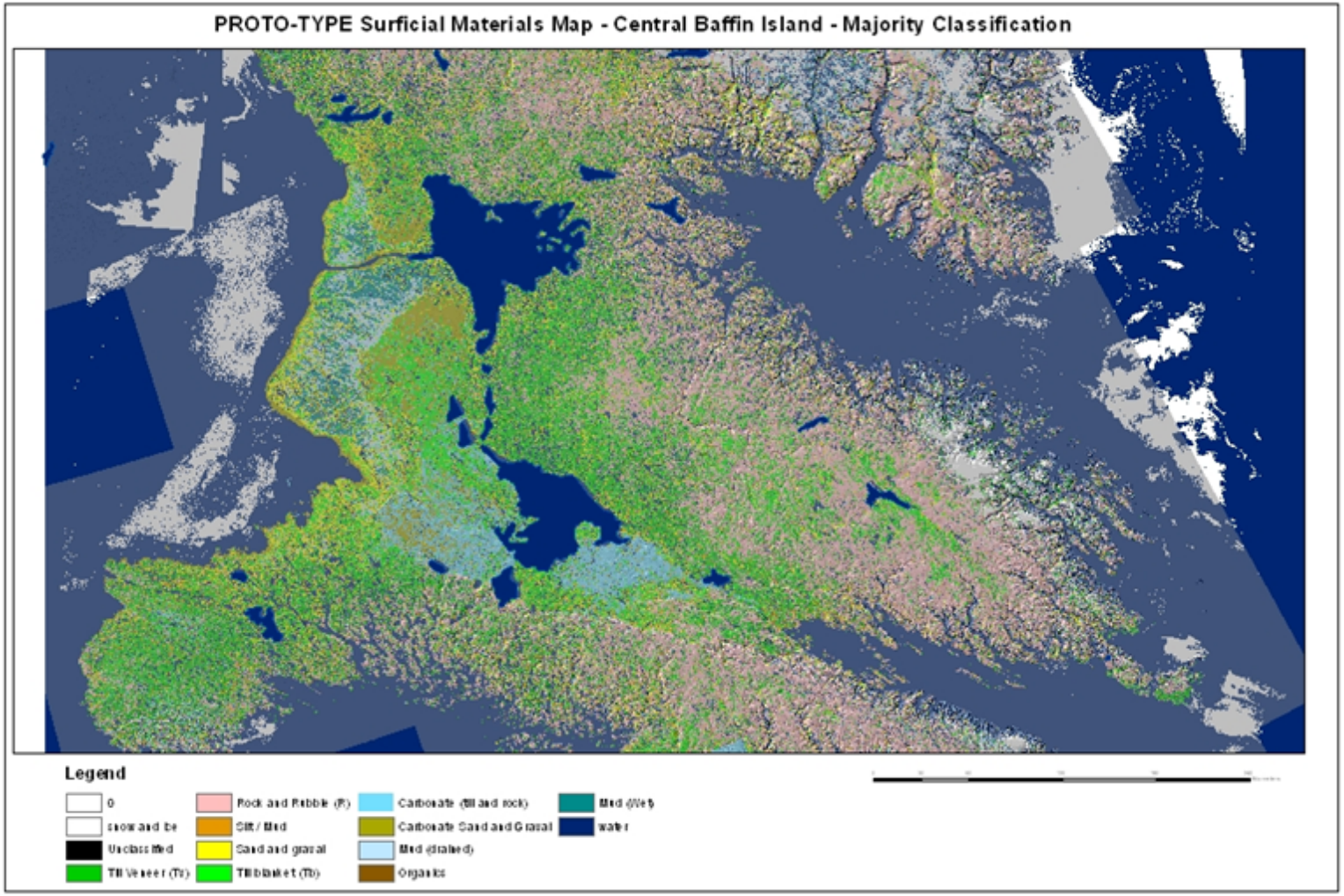

CENTRAL BAFFIN ISLAND - PROTOTYPE MAP
OF SURFICIAL MATERIALS
RPM
Legend
60,000 0 60,000 120,000 180,000 240,000
Meters
Produced from a classification of LANDSAT 7 data using the
Geological Survey of Canada's Iterative Classification Algorithm (ICA)


CENTRAL BAFFIN ISLAND - PROTOTYPE MAP
OF SURFICIAL MATERIALS - VARIABILTY (UNCERTAINTY MAP)
RPM
Legend
Class Variability of 2
Class Variability of 3
Class Variability of 4
Class Variability of 5
Class Variability of 6
Class Variability of 7
Class Variability of 8
60,000 0 60,000 120,000 180,000 240,000
Meters
Produced from a classification of LANDSAT 7 data using the
Geological Survey of Canada's Iterative Classification Algorithm (ICA)

GEMS

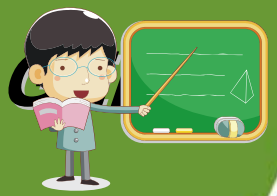
Predictive Map o Surficial Materials – Hall Peninsula, Baffin Island GSC OPEN FILE 7038



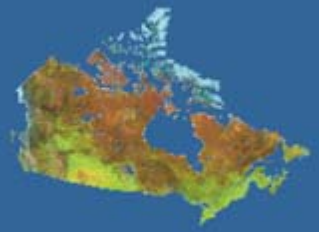


Average Accuracy of 76%

Landsat
Dem derivatives
Landsat texture

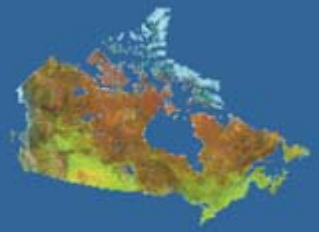


SUMMARY

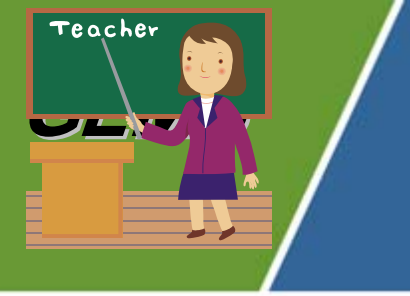


- Both statistical and spatial uncertainty is part of the RPM process!
- **RCM** allows us to evaluate the uncertainty and express in various ways (maps, statistics)
- **RCM** the study and subsequent improvement of training areas with an aim to improving classification accuracies
- Predictive (classifications) maps, as with all geologic maps, are a model of reality…**RCM** provides measures of uncertainty to accompany predictive maps
- Don't fully trust the statistical approach – look at the map!

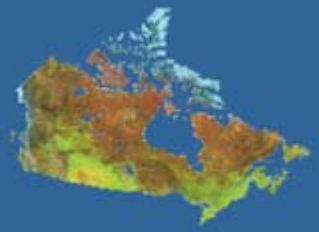
Surficial mapping approach…



- Supervised classification for surficial materials
  - Training area strategy – database by ecozones
- Landforms / Terrain types
  - DEM's and derivatives, lake shape and density analysis
- Combine layers within a GIS
- Incorporation of process information …..



We are still learning!!



- Experimenting with different data types:
  - MERIS, RADARSAT I and II, SPOT, DEM derivatives
- Experimenting with different classification algorithms:
  - Neural Nets, SVM, **Decision Trees (see5)**, Fuzzy Classifiers etc
- Incorporation of **Object Oriented Classification**!
- Training area purification – capturing diversity (spectral, spatial, topographic) !

THANKS !