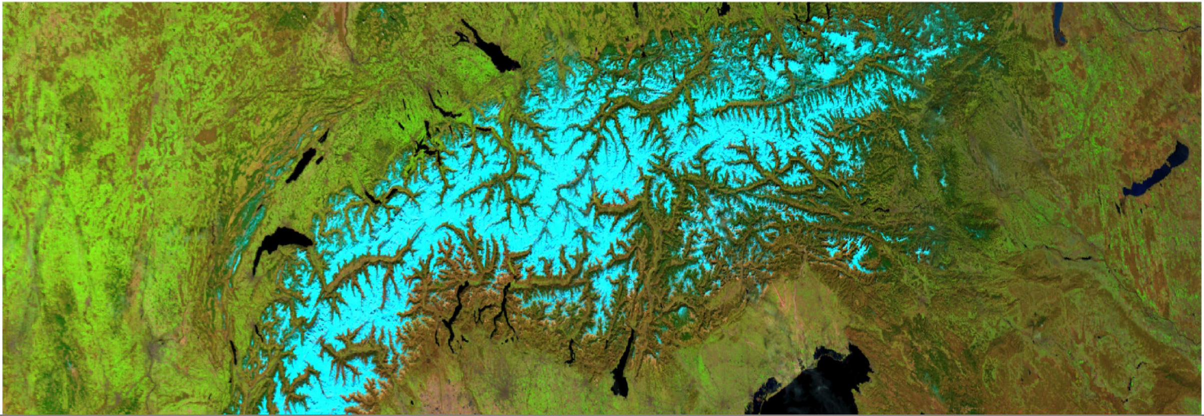
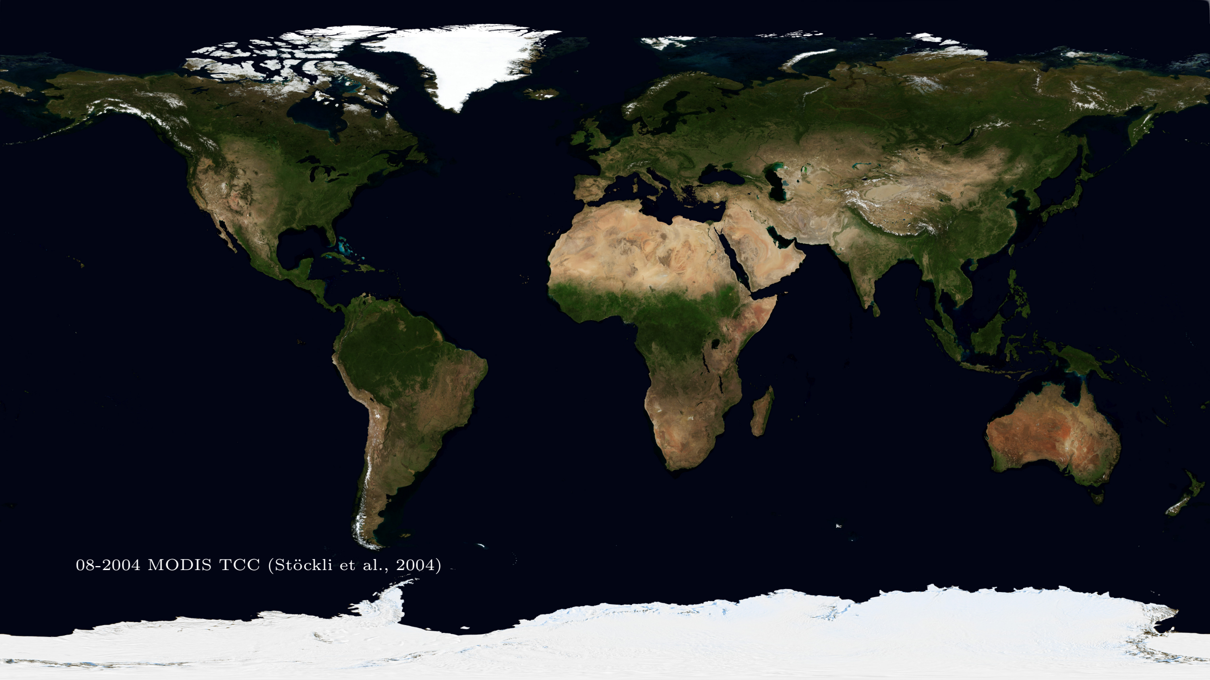


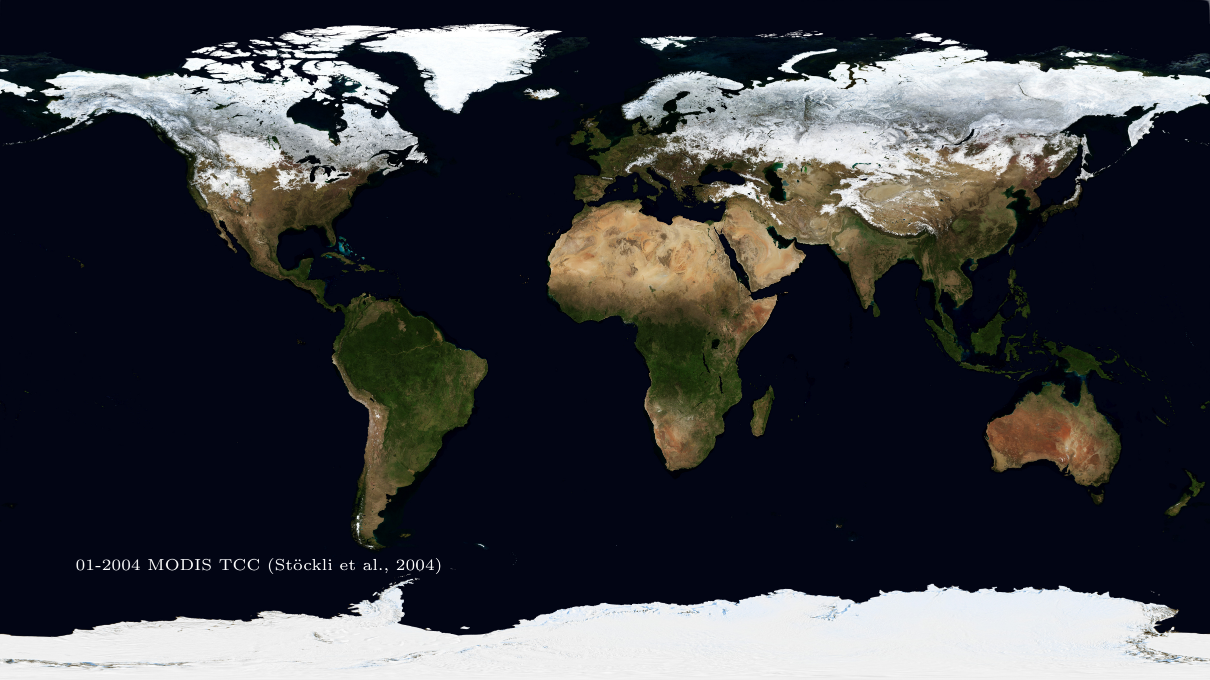
Machine learning techniques for mapping snow and ice over land

Kristoffer Aalstad

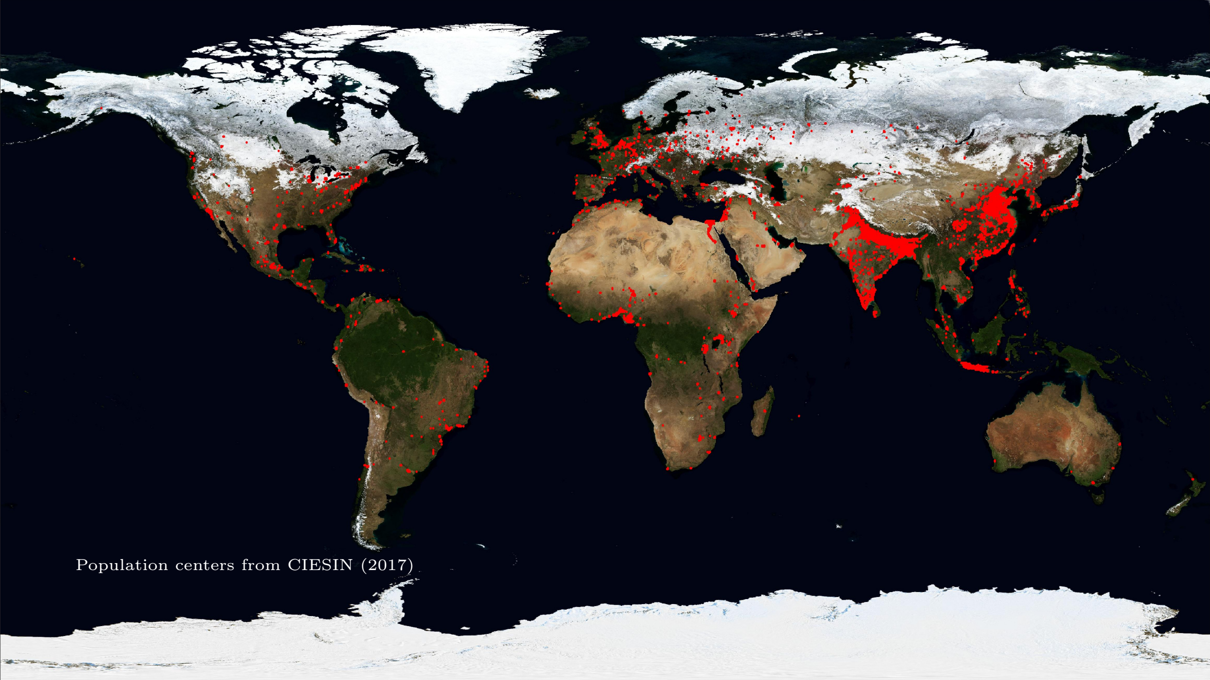
PhD Trial Lecture

UiO : Department of Geosciences
University of Oslo


08-2004 MODIS TCC (Stöckli et al., 2004)



01-2004 MODIS TCC (Stöckli et al., 2004)



Population centers from CIESIN (2017)

Key properties of snow and ice over land:

1. Reflective (high albedo)
2. Stores water (liquid and frozen)
3. Seasonal, ephemeral or perennial
4. High spatial variability
5. Melts (but high latent heat of fusion)

Σ = strongly modulate the surface energy, water, and carbon balance.
$\rightarrow$ Essential variables in the climate system.

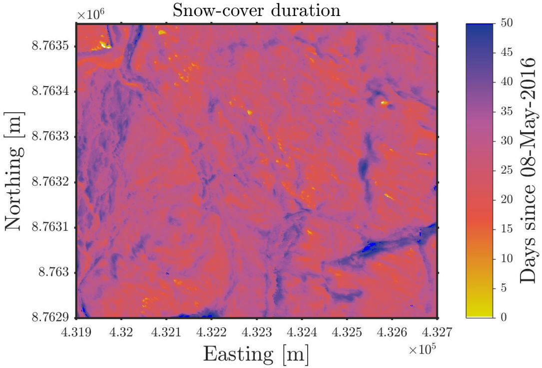


Figure: Snow-cover duration over a small $\simeq$1 km^2 area near Ny-Ålesund, Svalbard. Purple/yellow shows a high/low snow-cover duration.

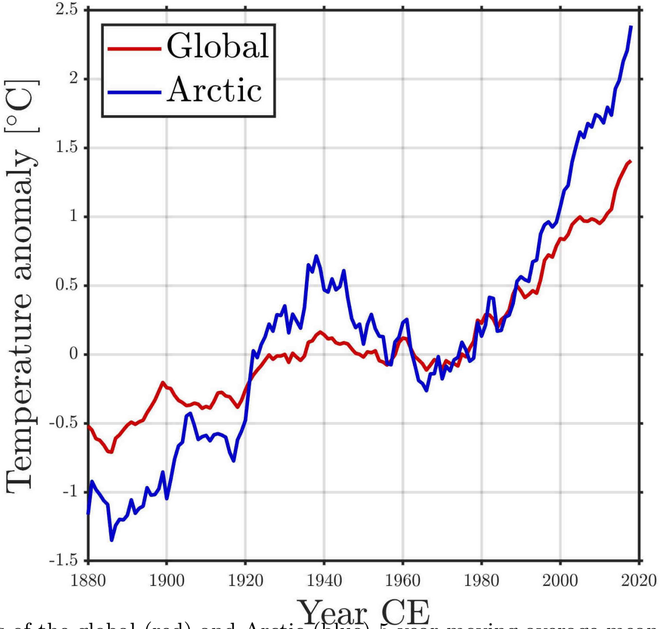
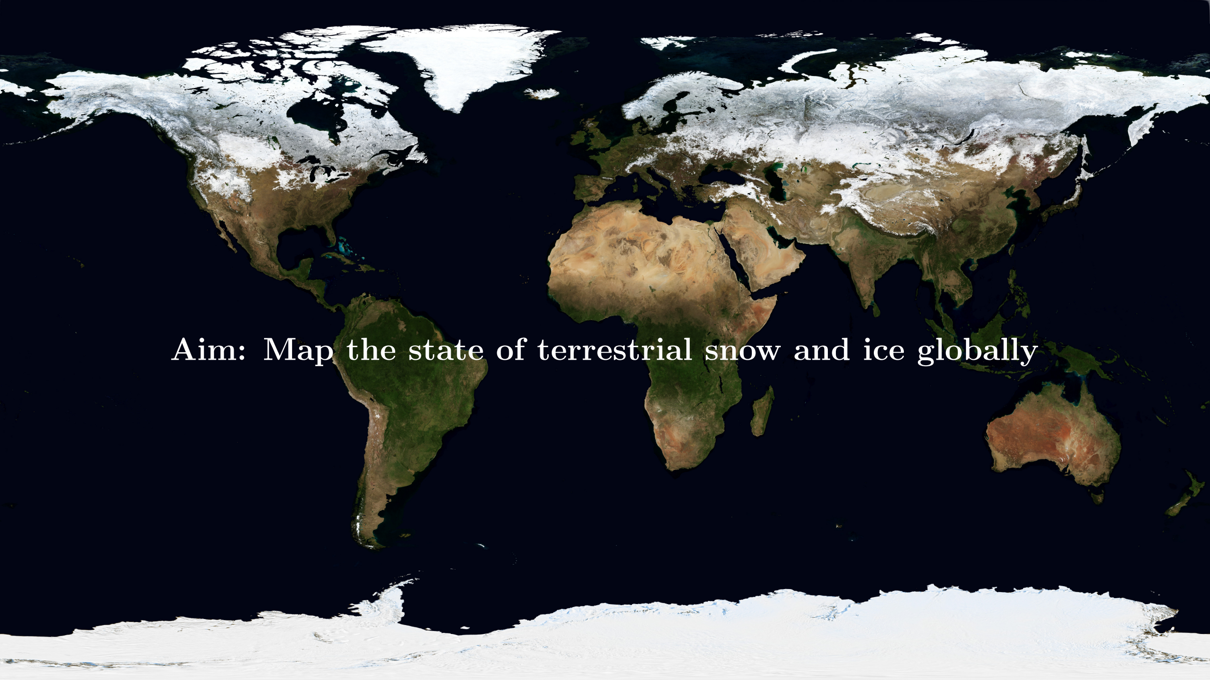


Figure: Time series of the global (red) and Arctic (blue) 5 year moving average mean surface air temperature anomaly in the period 1880-2018. Temperature data from NASA-GISS.


Aim: Map the state of terrestrial snow and ice globally

Mechanistic models

- Directly predict the desired output y given input x.
- Process-based, focusing on the causality in a system.
- Provides gap-free predictions.
- Problems with non-linearity (sensitivity to ICBC), scale, transferability, equifinality, and uncertainty (see Beven, 2001).
- Difficult to constrain.
- Computationally expensive.

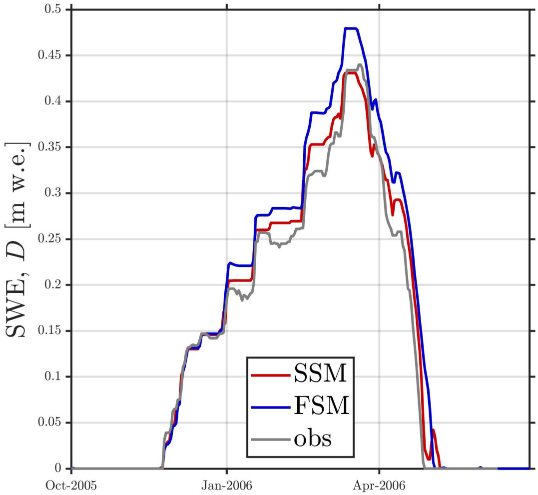


Figure: Snow water equivalent evolution for 2006 at Col de Porte (France) according to observations (gray) and predictions from two snow models.

Terrestrial observations

- Techniques: Snow surveys (courses), snow pillows, sonic rangers, terrestrial laser scanning, ground penetrating radar, time-lapse photography, gamma ray sensors.
- Relatively accurate.
- Often direct.
- Gaps in space and time.
- Limited spatial extent.

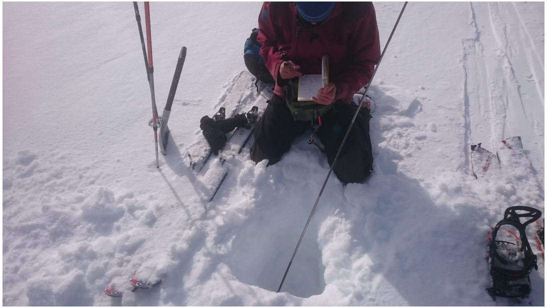


Figure: Measuring snow water equivalent using snow probes and a density sampler along a ski transect at the Steinfåen plateau near Ny-Ålesund, Svalbard 10.05.2016 (Photograph: K. Aalstad).

Airborne remote sensing

- Techniques: Optical spectrometry, light detection and ranging (LIDAR), stereo imaging, structure from motion (SfM), passive microwave radiometry, radar, gamma ray sensing.
- Relatively accurate.
- Usually indirect.
- Gaps in space and time.
- Greater spatial extent.

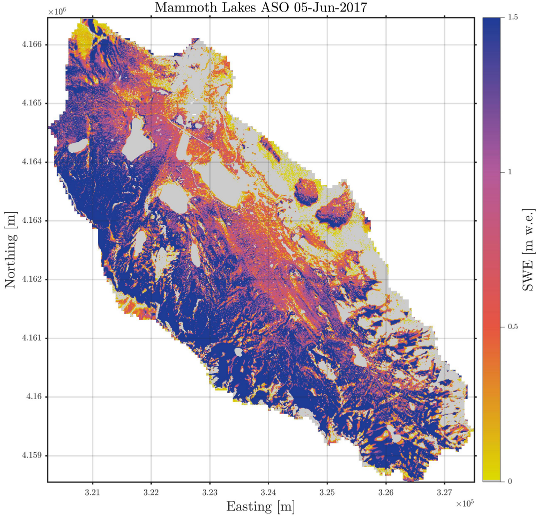


Figure: Snow water equivalent (SWE) distribution over the Mammoth Lakes Basin retrieved from the airborne snow observatory LIDAR.

Satellite remote sensing

- Techniques: Optical spectrometry, light detection and ranging (LIDAR), stereo imaging, passive microwave radiometry, radar, gravimetry.
- Relatively accurate.
- Usually indirect.
- Gaps in space and time.
- Potentially global extent.

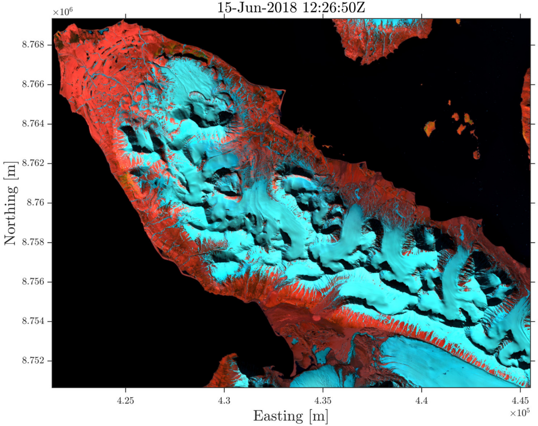


Figure: False color satellite image from the Sentinel-2B satellite over the Brøgger peninsula in the north western part of the Svalbard archipelago.


Method: Make predictions y given input data x

Machine learning

- The field of **machine learning** is a subfield of **artifical intelligence** existing at the intersection of applied statistics and computer science.
- **Machine learning techniques** are computer algorithms that are able to **"learn"** automatically from data (Goodfellow et al., 2016).
- *"A computer algorithm **learns** from experience E with respect to a class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."* (Mitchell, 1997).

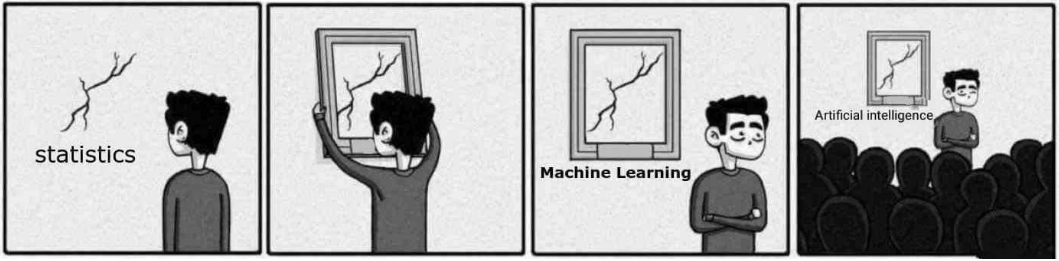


Figure: Original comic by sandserif. Source: towardsdatascience.com.

Example

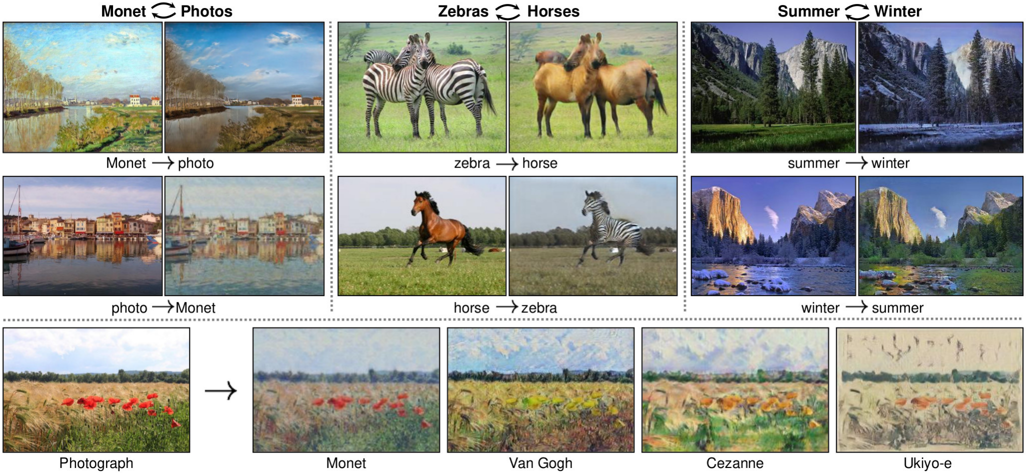


Figure: A trained machine learning model that has learned to translate paintings into photos, zebras into horses, and summer into winter based on unpaired training images (Zhu et al., 2017).

Supervised learning

- Learn a **mapping** from inputs to outputs based on **training data**.
- When the outputs are discrete the problem is known as **classification** whereas when they are continuous the problem is known as **regression**.
- Notable techniques: Neural networks, random forests, support vector machines, Gaussian processes, decision trees, linear reg., logistic reg. ...

Mathematically, a supervised algorithm learns the mapping from inputs $\mathbf{x}$ to outputs $\mathbf{y}$ based on **training data** $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^{N}$ that contains N pairs of **features** $\mathbf{x}_i$ (inputs) and known **targets** (outputs) $\mathbf{t}_i$ (MacKay, 2003).
Algorithms can learn by reducing measures of the **training error**

$$\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{t}$$

where the **prediction**

$$\mathbf{y} = \mathcal{M}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda})$$

is the learned mapping using model $\mathcal{M}(\cdot)$ with (hyper)parameters $\boldsymbol{\theta}$ ($\boldsymbol{\lambda}$).
Loosely, the machine learning model is statistical and not mechanistic.
Thereby, machine learning emphasizes prediction not causality.

Unsupervised learning

- Are **untrained**, i.e. do not experience any targets.
- Learns from the structure of the **features** in the dataset.
- Are **descriptive** as opposed to **predictive**.
- Notable techniques: Neural networks, k-means clustering, hierarchical clustering, principal component analysis, Gaussian mixture models ...

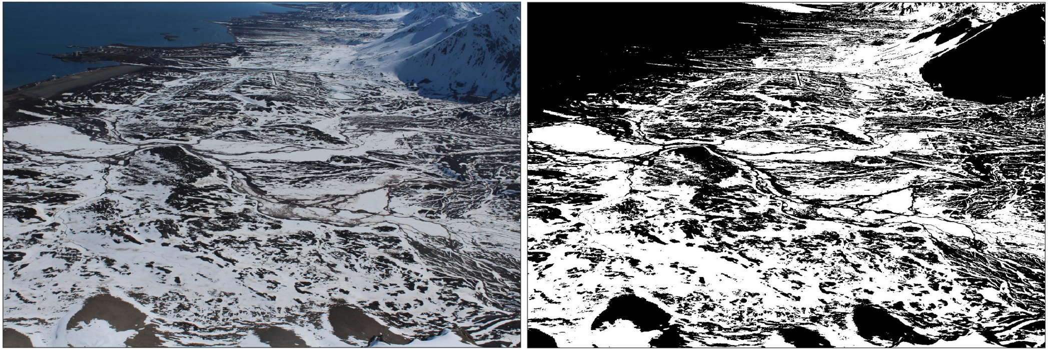


Figure: **Left:** Photo of a patchy snow-cover near Ny-Ålesund, Svalbard. **Right panel:** Unsupervised classification using k-means with 2 clusters. Successfully classifies snow-covered areas outside of shadows.

Generalization

- Models also need to **generalize**: make accurate predictions on new (non-training) data.
- Models with a high **capacity** (complexity) may **overfit** and generalize poorly even if they have minimum training error.
- Seek models that provide both low generalization and training error. Achievable through **regularization**.
- Leave some data out of the training dataset to create a **test** or **validation** dataset.

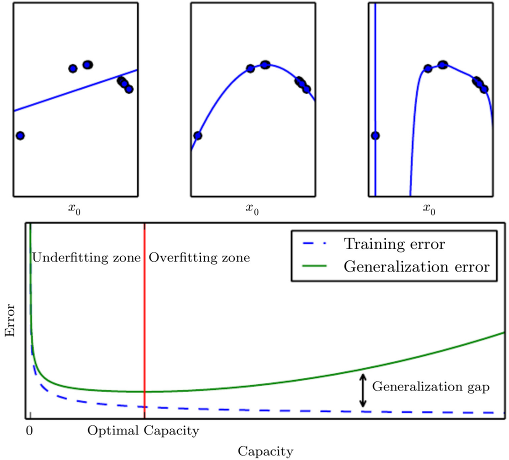


Figure: **Top panels:** Underfitting (left), optimal fit (middle), overfitting (right). **Bottom panel:** More complex models generalize poorly. Adapted from Goodfellow et al. (2016).

(Artificial) Neural Networks

- Modeled loosely on the brain.
- Network of idealized neurons that map from input to output.
- In **feedforward** networks connections are directed from inputs to outputs.
- **Feedback** (recurrent) networks do not have this constraint.
- **Deep** neural networks have several hidden layers.
- Nonlinear curve fitting devices often viewed as **black boxes**.

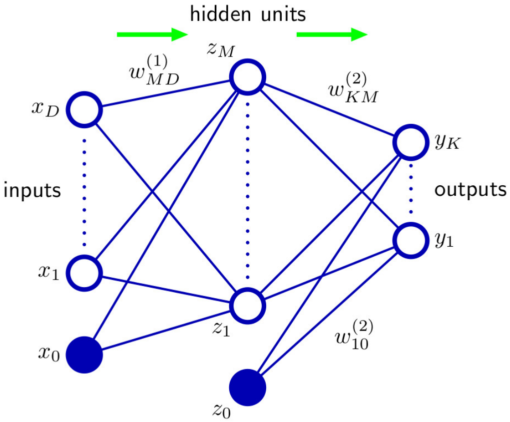


Figure: Architecture of a 2-layer feedforward neural network adapted from Bishop (2006). The arrows indicate the direction of information flow.

Single neuron for binary classification (i.e., 2 classes $t = 0$ or $t = 1$)

Activity rule: The **activation** a, given I-dimensional feature space, is

$$a = w_0 + \sum_{i=1}^{I} w_i x_i$$

where w_i are weights, and the **activity** for the single neuron is

$$y(a) = 1/(1 + e^{-a})$$

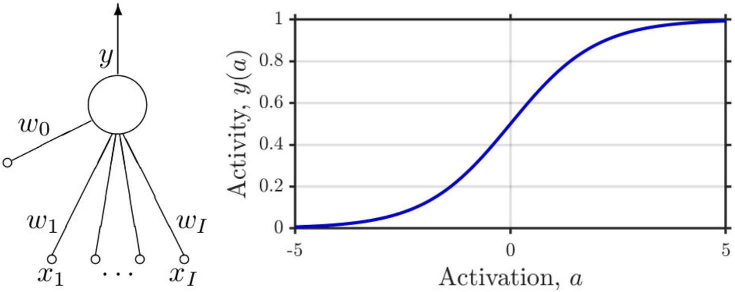


Figure: **Left:** Architecture of a single neuron (MacKay, 2003). **Right:** The logistic activation function.

Learning rule: **Minimize** the cost function

$$\mathcal{J} = \overbrace{-\sum_{n=1}^{N}\left\{t^{(n)}\ln([y(\mathbf{x}^{(n)};\mathbf{w})] + (1-t^{(n)})\ln[1-y(\mathbf{x}^{(n)};\mathbf{w})]\right\}}^{\text{Training error (cross-entropy)}} + \overbrace{\frac{\alpha}{2}\sum_{i=1}^{I} w_i^2}^{\text{Regularization}}$$

regularization (weight decay) punishes overfitting. Take the gradient

$$g_j = \frac{\partial \mathcal{J}}{\partial w_j} = -\sum_{n=1}^{N}\left(t^{(n)} - y^{(n)}\right)x_j^{(n)} + \alpha w_j$$

and perform "**backpropagation**", that is iterate through a **gradient descent** to update the weights until convergence

$$w_j \longleftarrow w_j - \eta g_j$$

η is the learning rate. The prediction y is the probability of being in class 1 (i.e. $1-y$ is the probability of being in class 0).

Toy problem: Single neuron for binary snow classification

- Task: Supervised classification of an orthophoto into snow/no-snow.
- Manually classified orthophoto as a reference "truth" for comparison.
- Raw data: Reflectance in the Red, Green, and Blue band.
- Snow is **white** (flat reflectance) and **bright** (high reflectance) so we use these two features as our x_i (with $I = 2$).
- 10 training data points: 5 snow-covered pixels and 5 bare pixels.
- 10^5 iterations. $<$ 10 lines of simple code, runs in $<$ 1 sec.

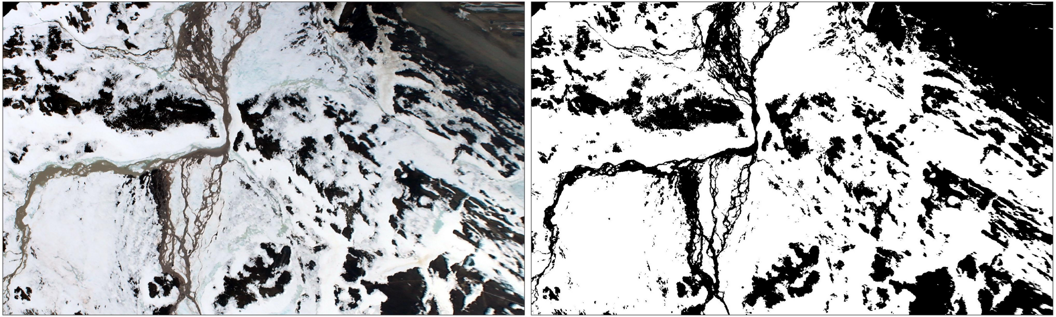
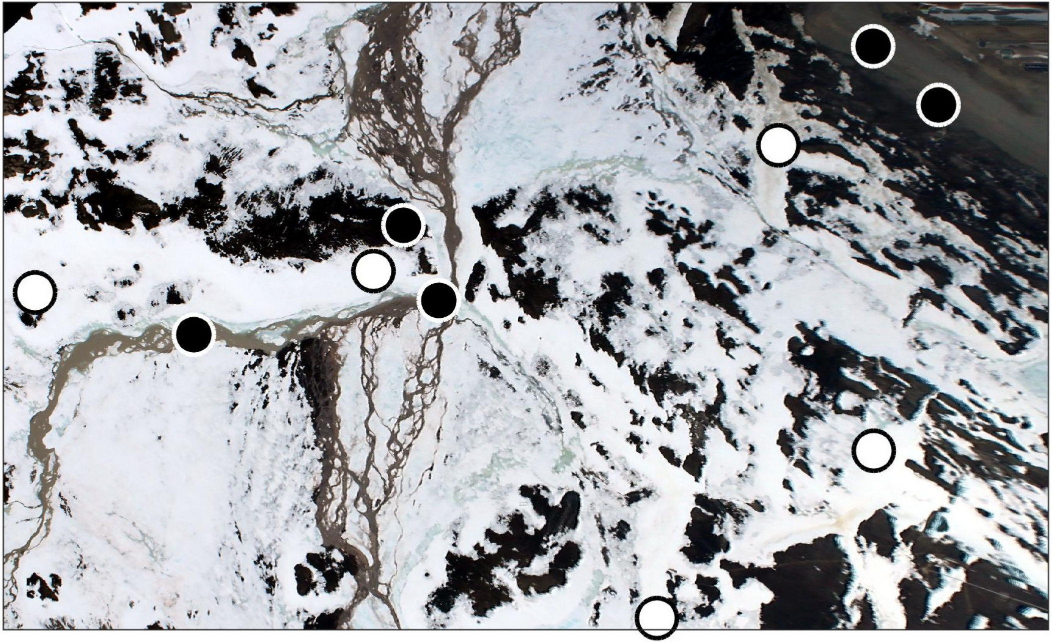
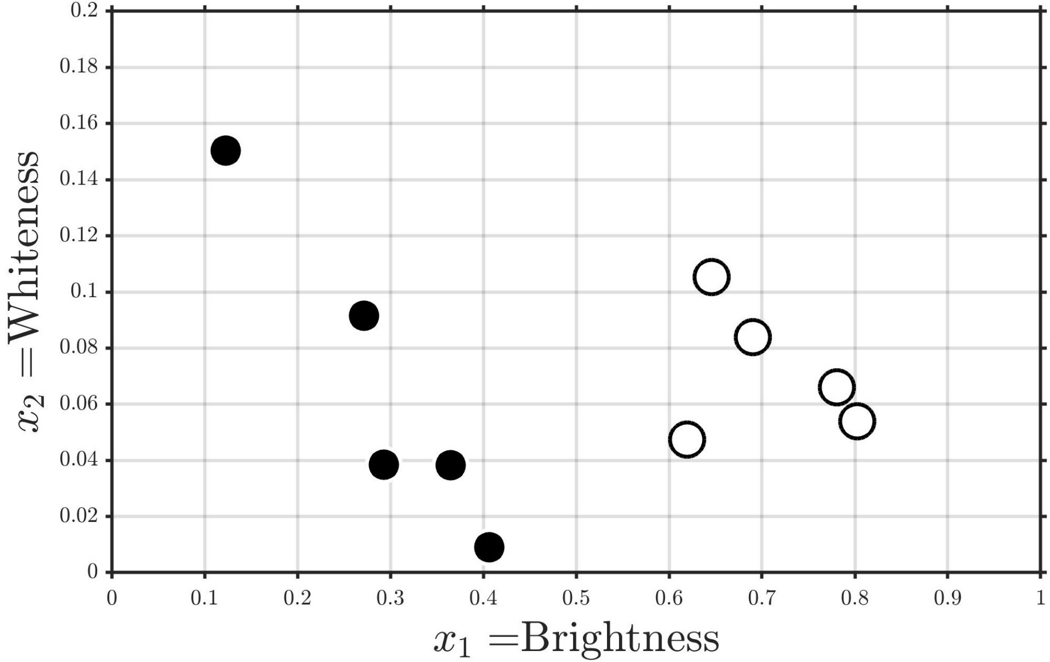


Figure: **Left:** Input orthophoto from Bayelva, Svalbard. **Right:** Manually optimized classified image.

```
%% Train a single neuron
L=10^5;                 % Number of iterations
w=zeros(3,L);           % Initial guess for w (including bias).
X=[ones(N,1) x'];       % Add unit feature for bias.
alpha=0.01;             % Weight decay
eta=0.01;               % Learning rate
for l=1:L-1             % Iterate
    a=X*w(:,l);         % Activation
    y=1./(1+exp(-a));   % Activity
    e=t-y;              % Error signal
    g=-X'*e+alpha*w(:,l); % Gradient vector with weight decay
    w(:,l+1)=w(:,l)-eta*g;% Update weights using gradient descent
end
```

Figure: Example MATLAB code for training a single neuron with backpropagation and weight decay.





x_2 =Whiteness
x_1 =Brightness
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

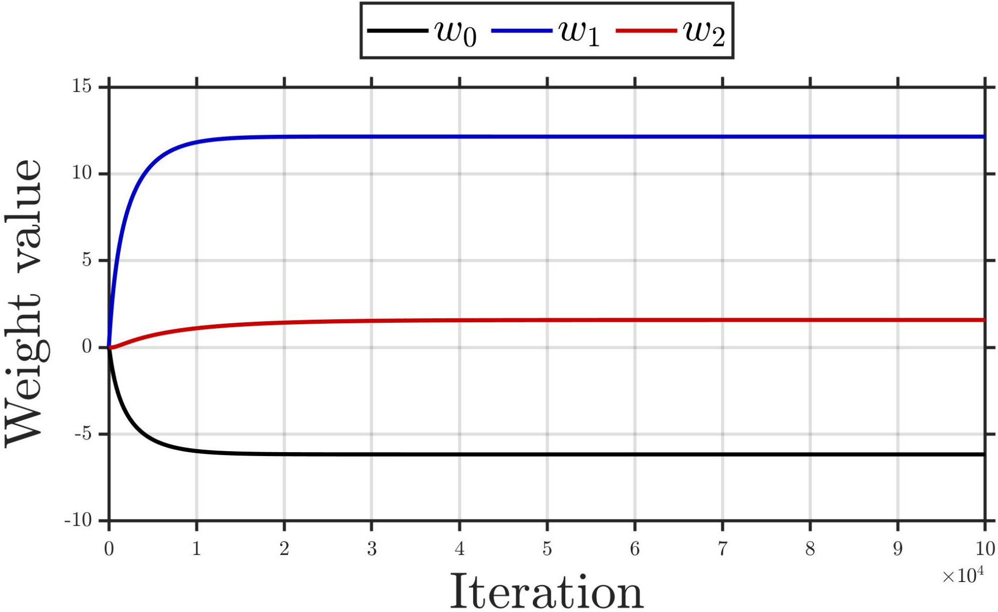
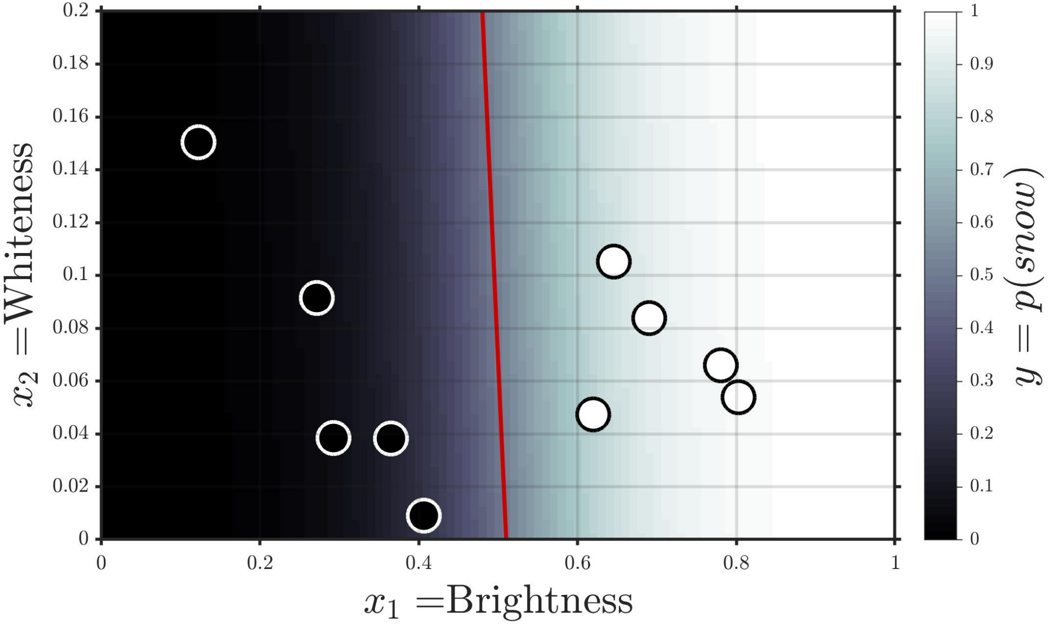


Figure: Training the weights of the single neuron. Convergence after $\sim 3 \times 10^4$ iterations.



0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
0.2
0.4
0.6
0.8
1
x_2 =Whiteness
x_1 =Brightness
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
$y = p(snow)$

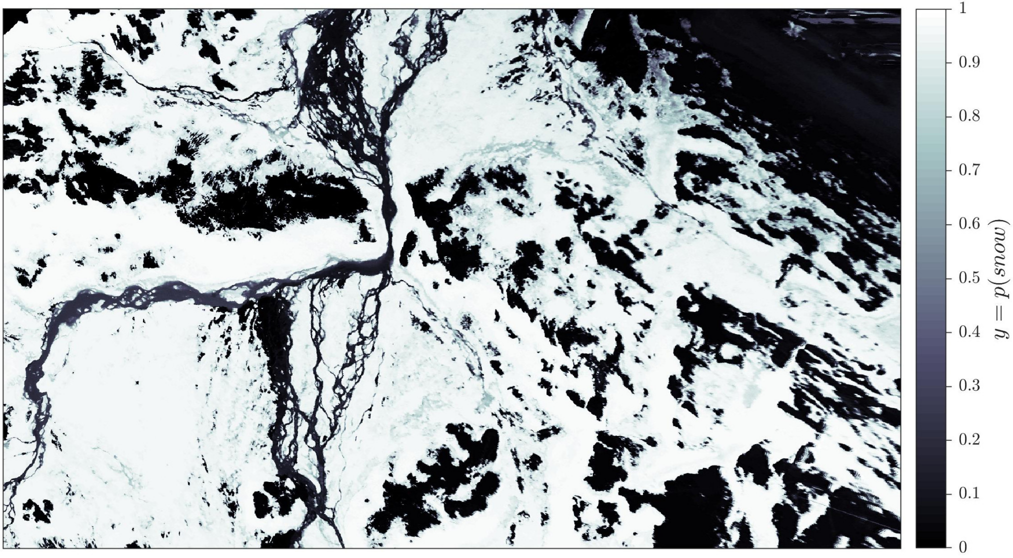


Figure: Probability of snow for all pixels in the image after running them through the single neuron.

Final classification using a threshold $y > 0.5$ for snow-covered pixels.

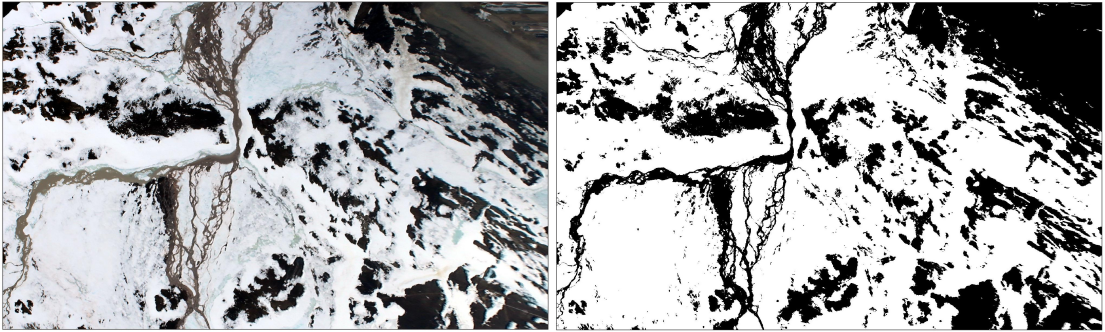


Figure: **Left:** The original image. **Right:** Image classified using the trained neuron.

How does this compare to the manually optimized classified image?

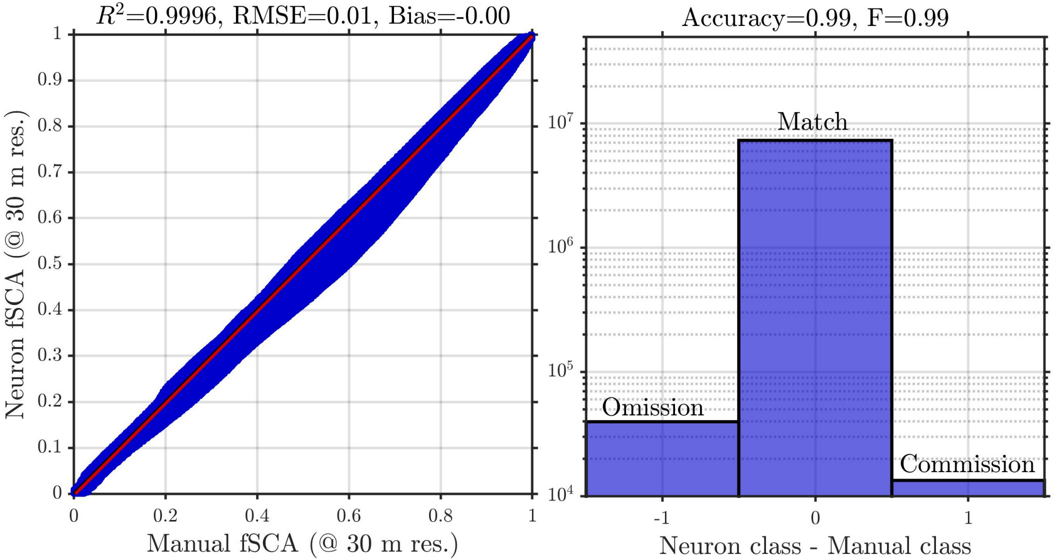


Figure: **Left:** Scatter plot relative to the reference manual fSCA. **Right:** Class difference histogram.

How well does the trained neuron generalize to a different image?
Poorly. The neuron is not robust to new circumstances (extrapolation).

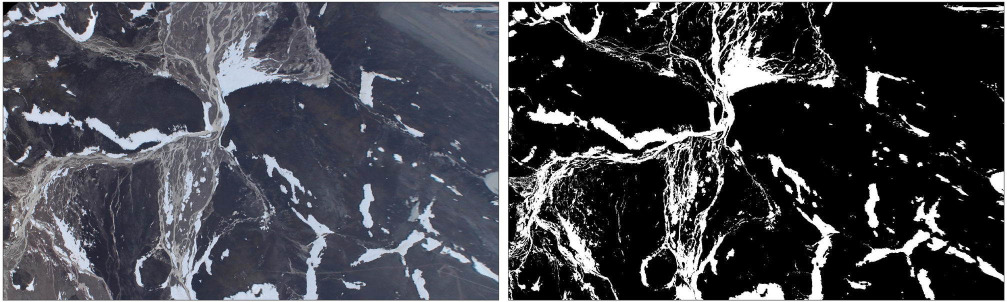


Figure: **Left:** The original image. **Right:** Classified image using a threshold.

Needs training with illumination condition as an additional feature

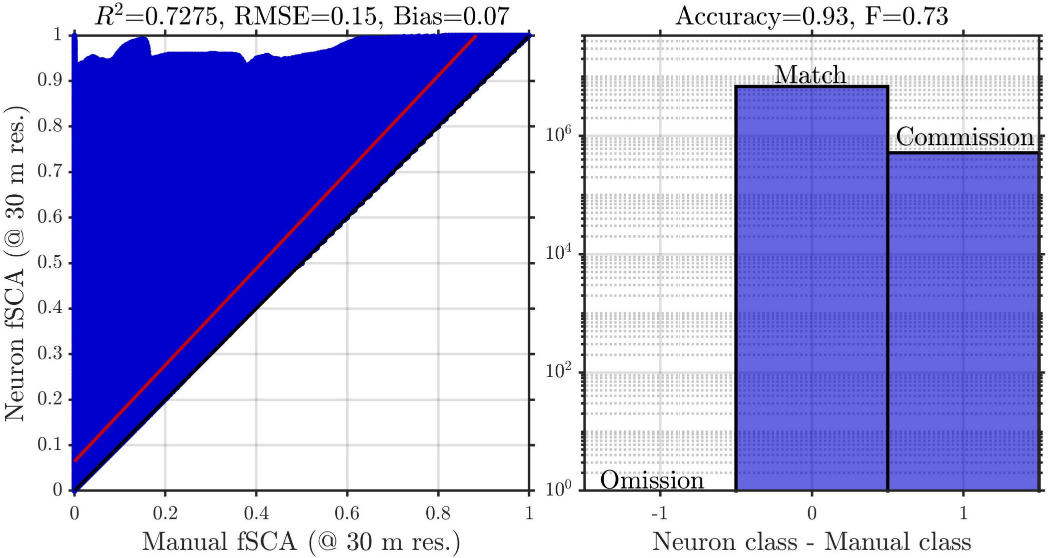


Figure: **Left:** Scatter plot relative to the reference manual fSCA. **Right:** Class difference histogram.

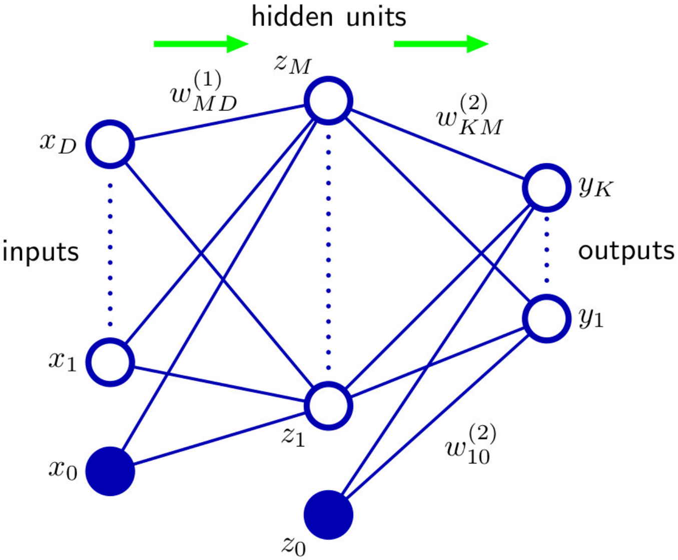


Figure: The general feedforward neural network. Adding more features, neurons, and hidden layers would increase the capacity of the neural network with respect to fitting more complex patterns in the training data, with regularization avoiding overfitting. Adapted from Bishop (2006).

Snow-cover prediction

- Study of Czyzowska-Wisniewski et al. (2015).
- Task: Predict (retrieve) **fractional snow-covered area (fSCA)** using single pixel Landsat (5&7) surface reflectances in VSWIR bands and topographic data.
- Method: 4-layer neural network with a 35-11-7 hidden layer node structure with early stopping.
- Training & validation: Coincident fSCA from IKONOS high resolution imagery.

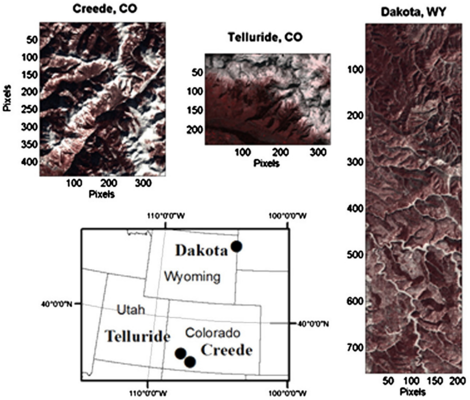
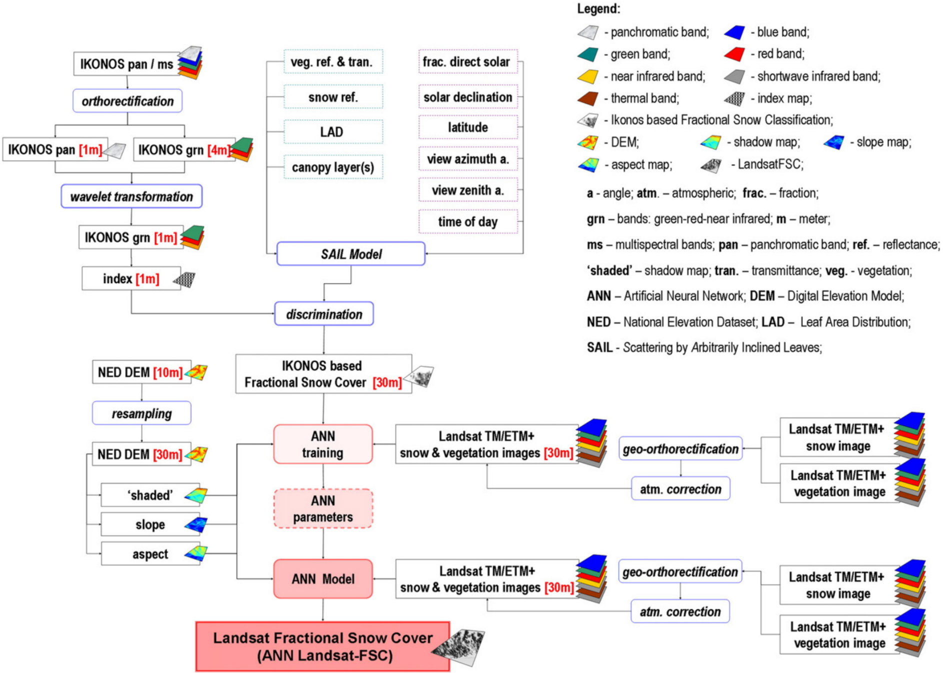


Figure: **Top:** Landsat imagery of the 3 study areas Creede, Telluride, and Dakota. **Bottom:** Location of the study areas in central USA. Adapted from Czyzowska-Wisniewski et al. (2015).



Legend:
- panchromatic band;
- blue band;
- green band;
- red band;
- near infrared band;
- shortwave infrared band;
- thermal band;
- index map;
- Ikonos based Fractional Snow Classification;
- DEM;
- shadow map;
- slope map;
- aspect map;
- LandsatFSC;
a - angle; atm. – atmospheric; frac. – fraction;
grn – bands: green-red-near infrared; m – meter;
ms – multispectral bands; pan – panchromatic band; ref. – reflectance;
'shaded' – shadow map; tran. – transmittance; veg. - vegetation;
ANN – Artificial Neural Network; DEM – Digital Elevation Model;
NED – National Elevation Dataset; LAD – Leaf Area Distribution;
SAIL - Scattering by Arbitrarily Inclined Leaves;
IKONOS pan / ms
orthorectification
IKONOS pan [1m]
IKONOS grn [4m]
wavelet transformation
IKONOS grn [1m]
index [1m]
veg. ref. & tran.
snow ref.
LAD
canopy layer(s)
frac. direct solar
solar declination
latitude
view azimuth a.
view zenith a.
time of day
SAIL Model
discrimination
IKONOS based Fractional Snow Cover [30m]
NED DEM [10m]
resampling
NED DEM [30m]
'shaded'
slope
aspect
ANN training
ANN parameters
ANN Model
Landsat TM/ETM+ snow & vegetation images [30m]
geo-orthorectification
atm. correction
Landsat TM/ETM+ snow image
Landsat TM/ETM+ vegetation image
Landsat TM/ETM+ snow & vegetation images [30m]
geo-orthorectification
atm. correction
Landsat TM/ETM+ snow image
Landsat TM/ETM+ vegetation image
Landsat Fractional Snow Cover (ANN Landsat-FSC)

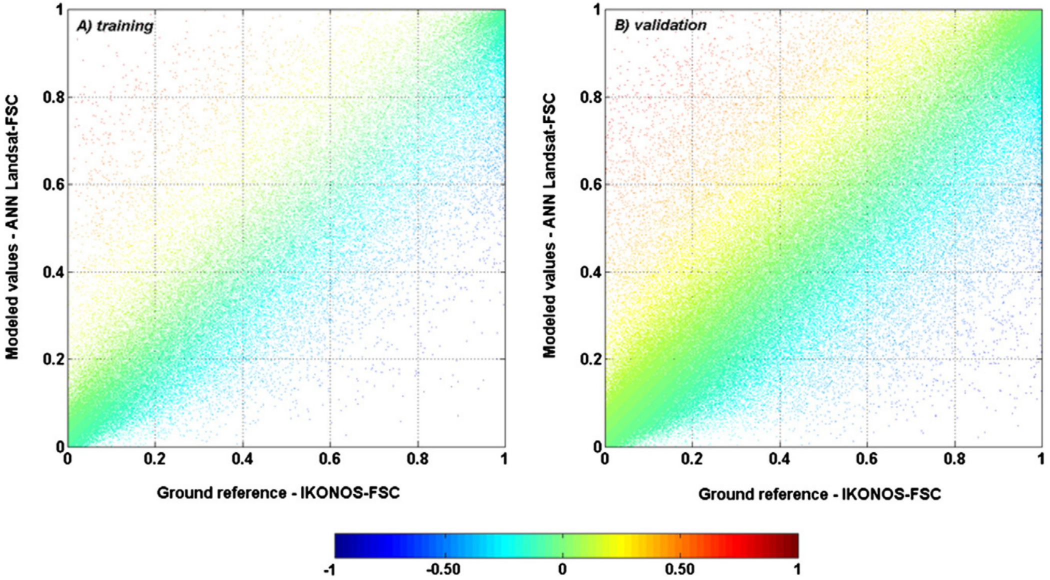


Figure: Landsat vs. IKONOS fSCA during training and validation (Czyzowska-Wisniewski et al., 2015).

			Correlation		Error			Model		Binary Classification				
datasets		N	r	r^2	RMSE	MAE	ME	CRM	EF	Accuracy	Precision	Recall	Specificity	AUC
Whole	tr	98 000	0.95	0.91	0.11	0.07	0.00	0.00	0.91	0.91	0.91	0.95	0.84	0.89
	val	297 728	0.95	0.90	0.12	0.08	0.00	0.00	0.90	0.90	0.92	0.95	0.79	0.87

Figure: Evaluation metrics for the fSCA retrieved from Landsat (Czyzowska-Wisniewski et al., 2015).

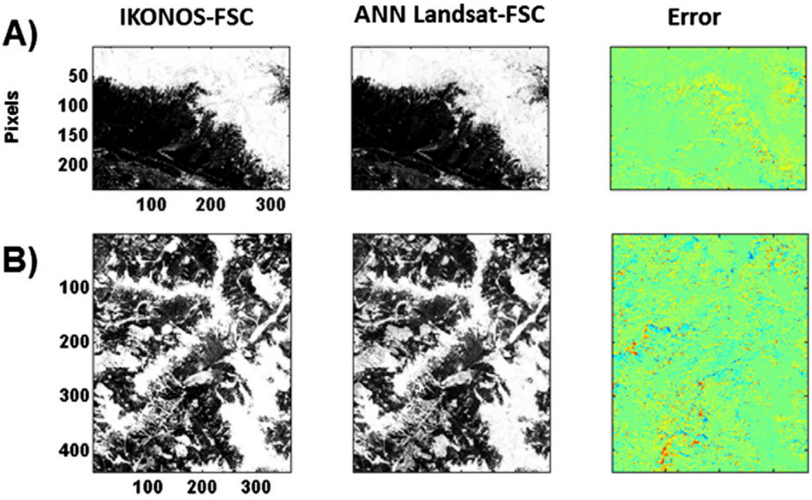


Figure: IKONOS-fSCA (left), Landsat-fSCA (mid), and error (right) (Czyzowska-Wisniewski et al., 2015).

What about clouds?

- Clouds masking is a difficult problem for optical imagery.
- Easy to recognize by eye, harder for an algorithm. Maybe machine learning can help?
- Jeppesen et al. (2019) show how a **deep convolutional neural network** can outperform traditional methods (Fmask) for Landsat imagery with almost double the accuracy and F-score.
- All bands in the VSWIR were found to be important for accurate snow/cloud discrimination.

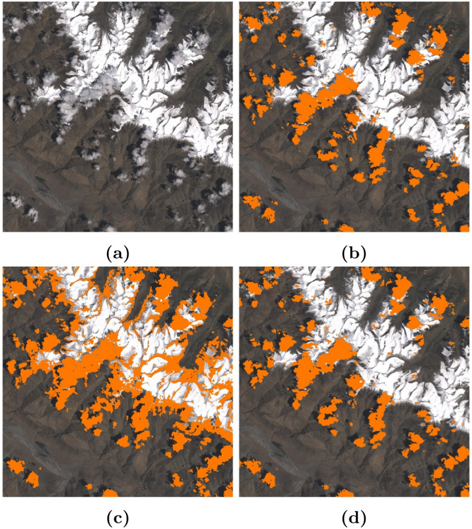


Figure: **(a)** Landsat scene **(b)** "ground truth" (clouds are orange) **(c)** Fmask **(d)** RS-Net. Adapted from Jeppesen et al. (2019).

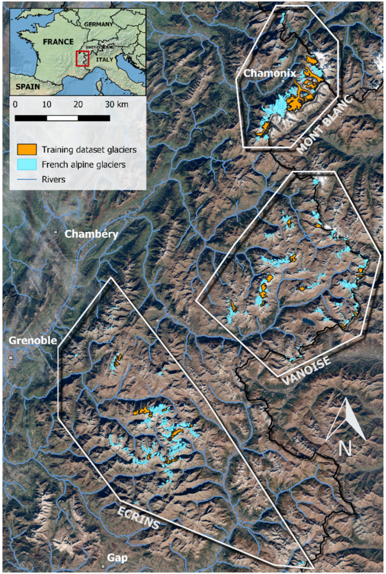


Figure: Glaciers considered by Bolibar et al. (2019).

Predicting the evolution of glaciers with neural networks

- Study of Bolibar et al. (2019).
- Task: Predict annual glacier-wide surface mass balance (accumulation-ablation) using climatic (temperature and snowfall) and topographic inputs.
- Method: A deep neural network with 4 hidden layers with a 40-20-10-5 hidden layer node structure.
- Training & validation: 31 years of reconstructed surface mass balance estimates.

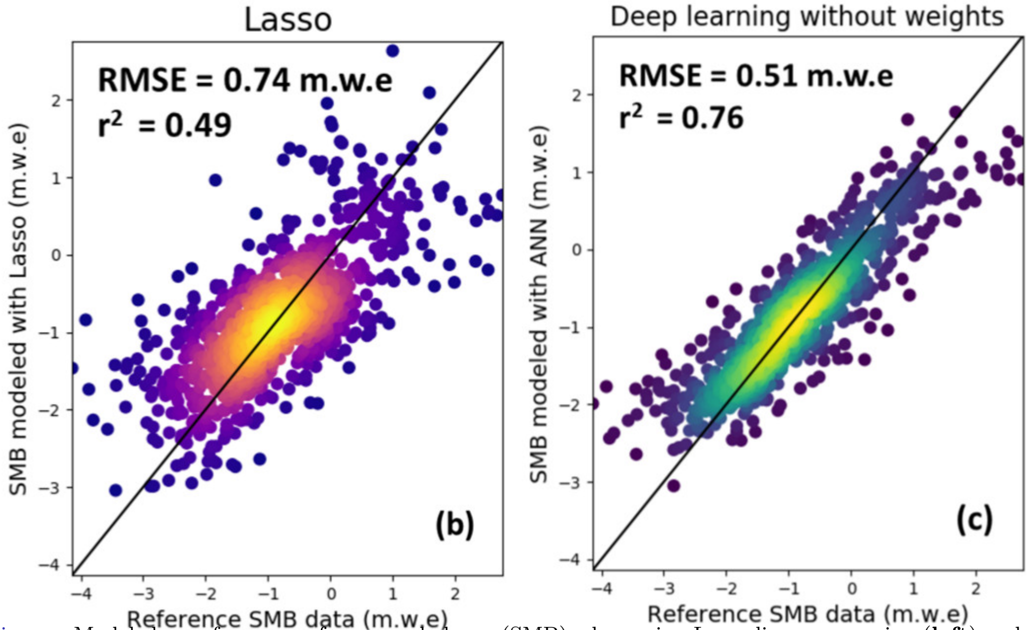


Figure: Modeled vs reference surface mass balance (SMB) when using Lasso linear regression (**left**) and a deep neural network (**right**). Only validation (not training) data is considered.

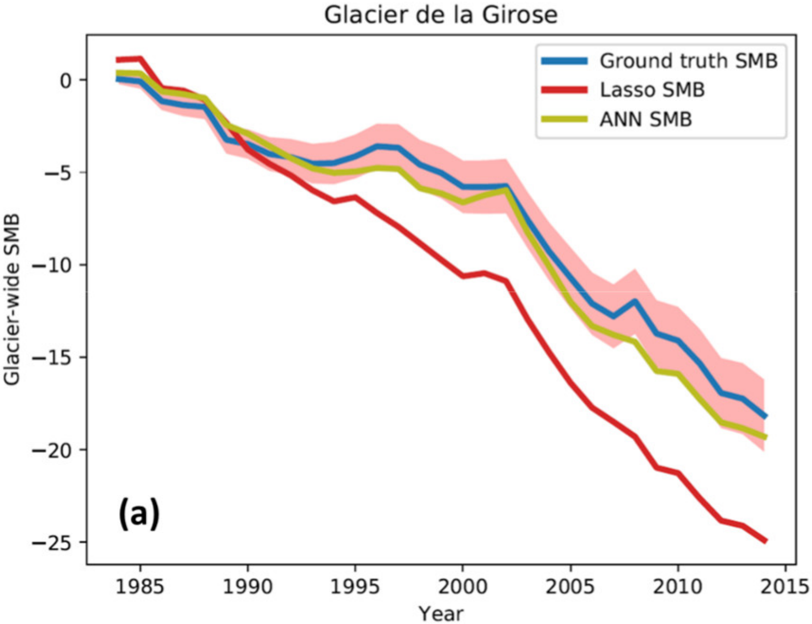


Figure: Cumulative glacier-wide surface mass balance (meters) from the reference data (blue, uncertainty range in pink), the lasso regression (red), and the deep neural network (green) for Glacier de la Girose.

Predicting SWE with machine learning

- The "holy grail" of snow hydrology (Dozier et al., 2016).
- Machine learning-based supervised regressions are a promising approach, given that many remotely sensed features are correlated with SWE.
- Applications date back at least to Chang and Tsang (1992) who used a neural network with passive microwave measurements as inputs to predict SWE.

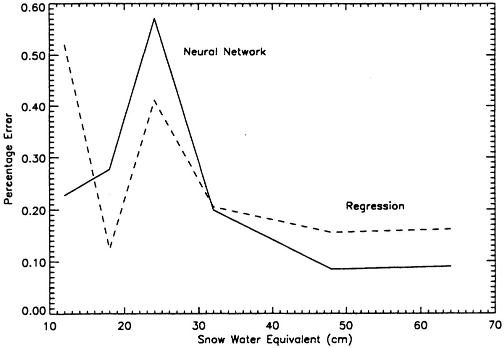


Figure: Precentage error in SWE prediction based on passive microwave satellite measurements using a linear regression (dashed) and neural network (solid line). Adapted from Chang and Tsang (1992).

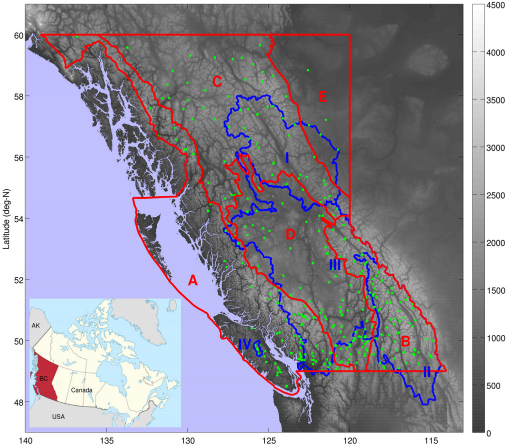


Figure: Location of the snow survey locations (green) in British Columbia, Canada. Surveyed SWE was the target data for the period 1980-2010. Adapted from Snauffer et al. (2018).

Predicting SWE with neural networks

- Study of Snauffer et al. (2018).
- Task: Predict point-scale SWE by fusing SWE estimates from global reanalysis and satellite-based data sets (inputs).
- Method: An ensemble (a collection) of neural networks with one hidden layer and early stopping for regularization.
- Training & validation: 30 years of surveys from 386 locations.

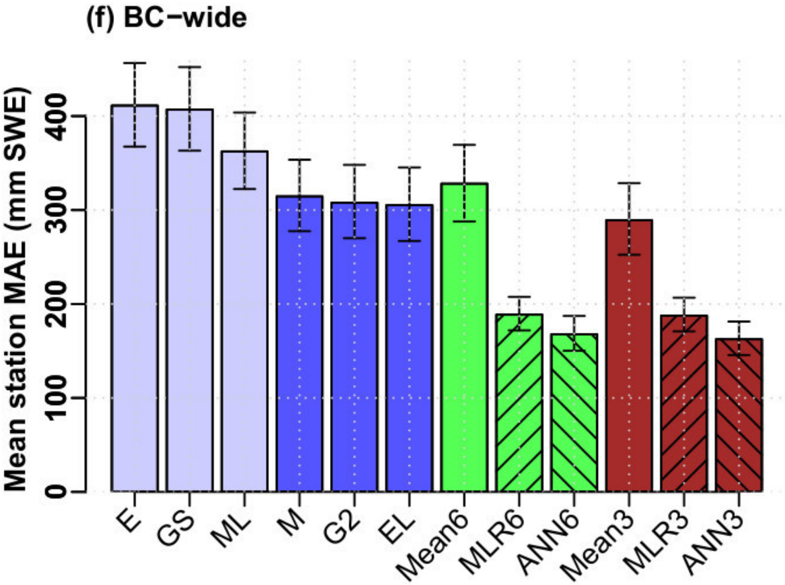


Figure: Mean absolute errors for input SWE, multiple lienar regression (MLR) SWE, and the neural network (ANN) SWE. Adapted from Snauffer et al. (2018). As usual, only validation data is considered.

Random forests

- Prediction based on growing an ensemble (a collection or "forest") of decision trees (Breiman, 2001).
- Takes average (regression) or majority vote (classification) of ensemble members to make a prediction y given features x_i.
- Combines random feature selection and "bagging". So that each tree is exposed to different features and training data.

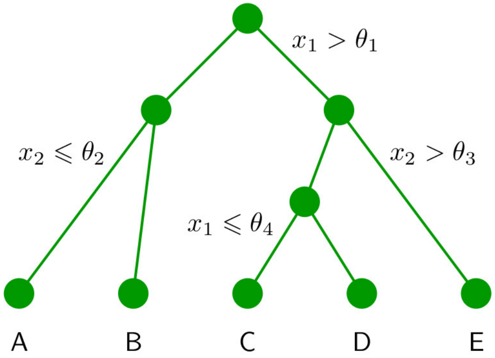


Figure: A standard binary decision tree performing a classification based on the magnitude of input features x_1 and x_2. Adapted from (Bishop, 2006).

Predicting SWE using random forests

- Study of Bair et al. (2018).
- Task: Predict gridded SWE using static (e.g. DOY, elevation, lat, lon) and dynamic (passive microwave ΔT_b, fSCA, mean reconstructed SWE) as inputs (features).
- Method: Random forest (also used neural networks).
- Training & validation: Reconstructed SWE over Afghanistan for 2003-2011.

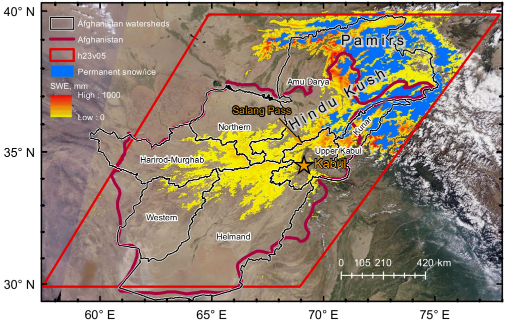


Figure: MODIS true color satellite image of Afghanistan overlaid with 1 April reconstructed SWE. Adapted from Bair et al. (2018).

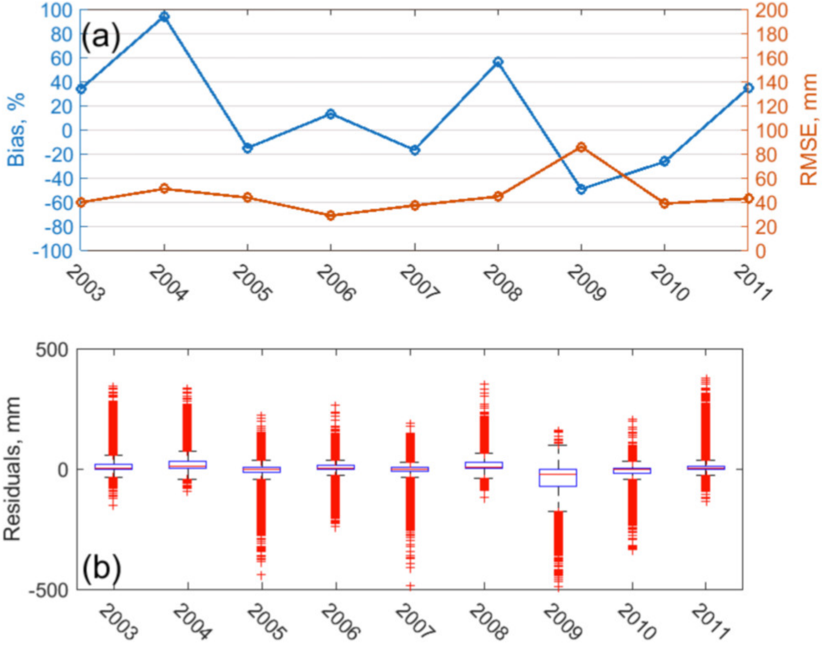


Figure: Validation of the random forest SWE predictions. Adapted from Bair et al. (2018).

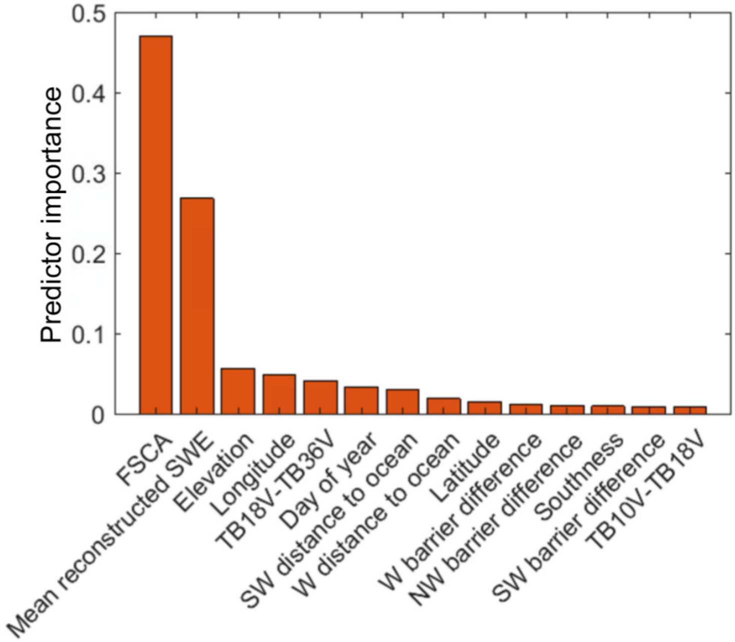


Figure: Predictor (i.e. feature) importance. Adapted from Bair et al. (2018).

Summary: Machine learning techniques for mapping snow and ice over land.

Pros

- Powerful **non-linear curve fitting** devices.
- Provide "evidence-based" (**data-driven**) predictions.
- Can be very **accurate**.
- Makes it possible to **emulate** complex models allowing for more efficient inferences to be made.
- Can help make sense of vast amounts of **unused data**.
- Still a **developing** field.

Cons

- **Black box** systems.
- Difficulty inferring **causes**.
- Not **process-based**.
- Can struggle to **generalize** outside of the training data.
- **Stationarity** assumption: it is not clear how well a model trained on the present climate will perform in a future climate.
- For some problems **generating accurate training data** in large volumes can be challenging.

In the quest for ever more accurate predictions. . .

- Data driven (*inductive*) prediction is great, and ML is a fantastic toolbox for this.
- Should not abandon the hypothesis-based (deductive) scientific framework, giving up on making causal inferences. (Pearl, 2019).
- Some balance must exist between mechanistic and machine learning modeling (Baker et al., 2018). Perhaps **data assimilation** can help to maintain this balance.

Figure: Throwing the baby out with the bathwater. From *Appeal to Fools* by Thomas Murner (1512).

Bair, E. H., Calfa, A. A., Rittger, K., and Dozier, J. (2018). Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. *The Cryosphere*. doi: 10.5194/tc-12-1579-2018.

Baker, R. E., Peña, J. M., Jayamohan, J., and Jérusalem, A. (2018). Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*. doi: 10.1098/rsbl.2017.0660.

Beven, K. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bolibar, J., Rabatel, A., Gouttevin, I., Galiez, C., Condom, T., and Sauquet, E. (2019). Deep learning applied to glacier evolution modeling. *The Cryosphere Discussions*. doi: 10.5194/tc-2019-163.

Breiman, L. (2001). Random forests. *Machine Learning*. doi: 10.1023/A:1010933404324.

Chang, A. T. C. and Tsang, L. (1992). A Neural Network Approach to Inversion of Snow Water Equivalent from Passive Microwave Measurements. *Nordic Hydrology*. doi: 10.2166/nh.1992.0012.

Czyzowska-Wisniewski, E. H., van Leeuwen, W. J. D., Hirschboeck, K. K., Marsh, S. E., and Wisniewski, W. T. (2015). Fractional snow cover estimation in complex alpine-forested environments using an artificial neural network. *Remote Sensing of Environment*.

Dozier, J., Bair, E. H., and Davis, R. E. (2016). Estimating the spatial distribution of snow water equivalent in the world's mountains. *WIREs Water*. doi: 10.1002/wat2.1140.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Jeppesen, J. H., Jacobsen, R. H., Inceoglu, F., and Toftegaard, T. S. (2019). A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sensing of Environment*.

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*. doi: 10.1145/3241036.

Snauffer, A. M., Hsieh, W. W., Cannon, A. J., and Schnorbus, M. A. (2018). Improving gridded snow water equivalent products in British Columbia, Canada: multi-source data fusion by neural network models. *The Cryosphere*. doi: 10.5194/tc-12-891-2018.

Stöckli, R. et al. (2004). https://earthobservatory.nasa.gov/features/bluemarble.

Zhu, J. Y., Park, T., Isola, P., and Efros, A. (2017). Unpaired image-to-image translations using cycle-consistent adversarial networks. Proceedings of The IEEE International Conference on Computer Vision (ICCV).

Thanks for your attention

The Research Council of Norway

Funding: SatPerm project (FRINATEK #239918)