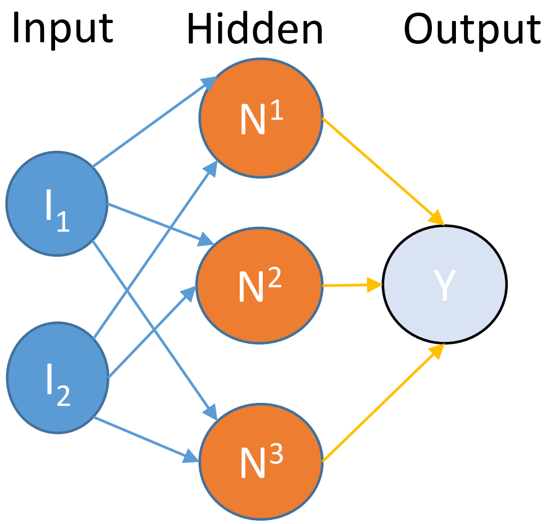
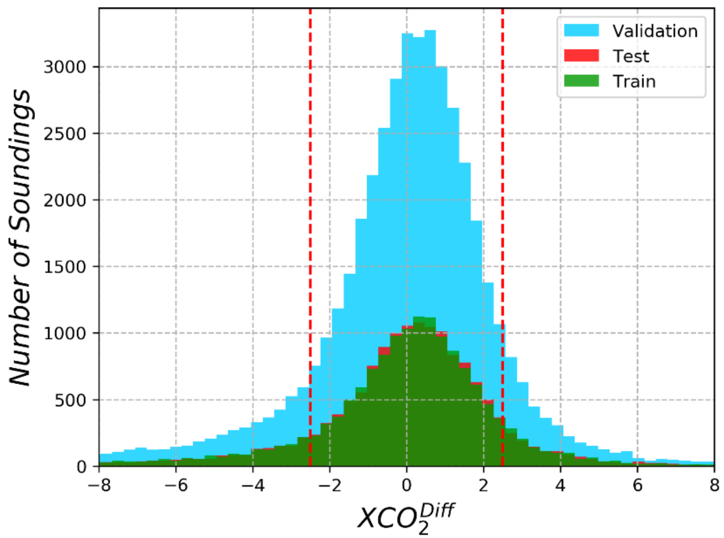
Assessing the Feasibility of Using a Neural Network to Filter OCO-2 Retrievals at Northern High Latitudes

Joseph Mendonca, Ray Nassar, Chris O'Dell, and Debra Wunch

- Created a 3 layer neural network (NN) to filter OCO-2 B10 data at norther high latitudes.
- Trained the NN using a portion of OCO-2 data coincident with TCCON sites at northern high latitudes.
- Validated the NN filter using OCO-2 data that was not used to train the NN.
- Compared the bias, precision, and throughput to the B10 qc flag filter.
- Found that the NN filter improved the bias, precision, and throughput at most TCCON sites.



Train NN



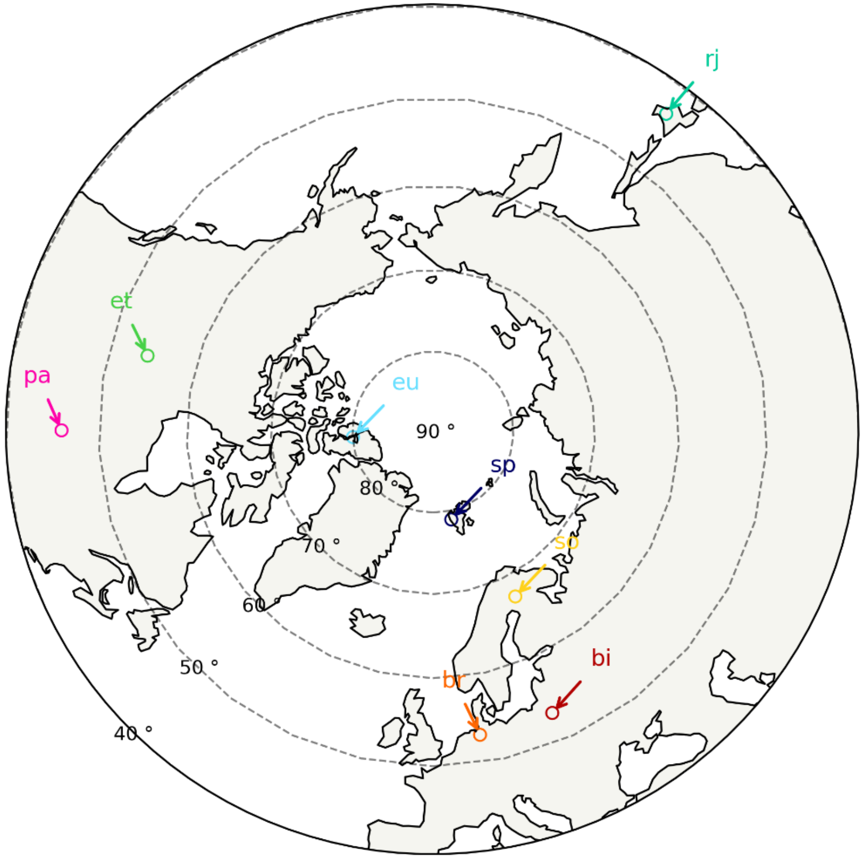
Validate NN

	Neural Network		B10 qc_flag	
Site	Bias ± precision	Number of retrievals	Bias ± precision	Number of retrievals
All	0.25 ± 1.27	23429	0.52 ± 1.45	20198
Eureka (eu)	-0.53 ± 2.35	59	0.34 ± 2.94	634
Ny Ålesund (sp)	0.87 ± 2.30	91	2.09 ± 2.65	92
Sodanklyä (so)	0.34 ± 1.23	5118	0.64 ± 1.30	4736
East Trout Lake (et)	0.01 ± 1.34	5261	0.44 ± 1.48	3186
Białystok (bi)	0.27 ± 1.16	6237	0.40 ± 1.18	5609
Bremen (br)	0.42 ± 1.19	4066	0.85 ± 1.28	3672
Rikubetsu (rj)	0.23 ± 1.39	2597	0.13 ± 1.70	2269
Park Falls (pa)	-0.12 ± 1.27	14859	-0.12 ± 1.18	12406

Environment and Climate Change Canada
Environnement et Changement climatique Canada
Canada

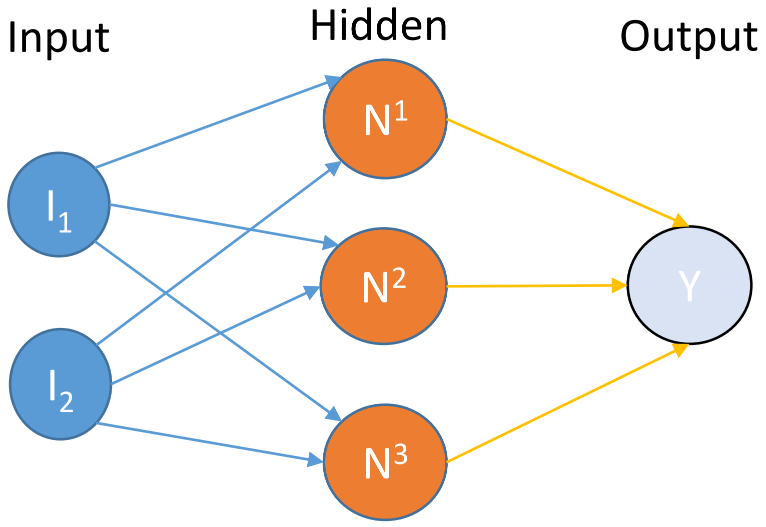
TCCON Data Coincident Criteria

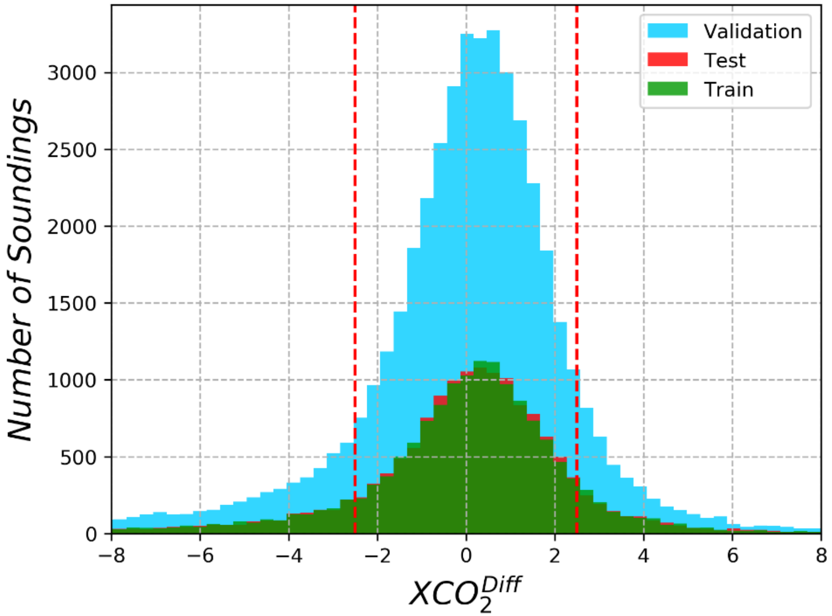
- Total Carbon Column Observing Network (TCCON) retrievals are used as the truth proxy.
  - Eureka (eu), East Trout Lake (et), Park Falls (pa), Sodanklyä (so), Ny Ålesund (sp), Białystok (bi), Rikubetsu (rj), and Bremen (br)
- Filtering TCCON Data:
  - Flag = 0
  - XHF <= 150 ppt (Avoid polar vortex)
  - XCO <= 125 ppb (Avoid biomass burning)
- Filtering OCO-2 data:
  - Using all data (nadir and glint)
  - Measurement type = land
- Coincident Criteria:
  - Distance <= 150 km
  - T700 hpa diff <= 2K
  - Time diff <= 2 hrs (Avoid diurnal cycle)
- Bias calculation is OCO-2 minus TCCON.
  - Take different ak into account
  - Take different apriori account
- The data is split into training, testing and validation randomly.



Training the Neural Network

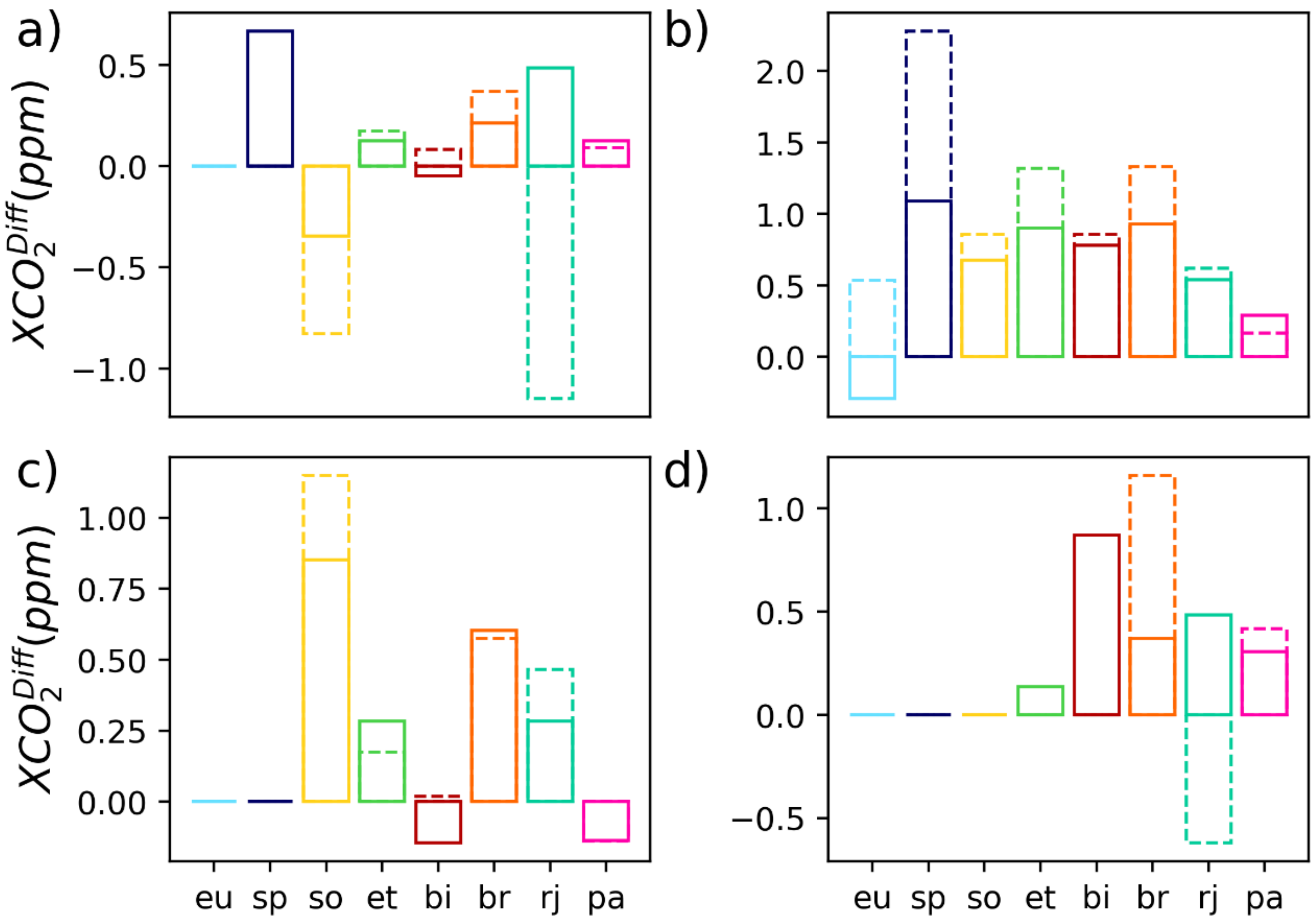
- This is 3-layer NN with multiple neurons
- Input layer (I) are the values of the features.
  - Feature values have been standardize $z_i = \frac{I_i - \mu_i}{\sigma_i}$ to take into account units.
  - Each input is connected to each neuron.
- Hidden layer (N) is the calculation done by each neuron.
  - Each neuron is connected to the output.
- Output (Y) is the output value.
  - Output values range between 0.0 to 1.0
  - If Y <= 0.1 the retrieval is "good" and if Y > 0.1 the retrieval is "bad"
- Train the NN by assigning OCO-2 data as "good" or "bad" based on bias with TCCON. All data in between the red lines is assigned "good" while the rest is "bad" .
- Used the Train and Test data set to train the NN.
- Used the Validation data set to assess the NN filter.
  - Park Falls data is only in the Validation data set.





Validating the Neural Network – Bias

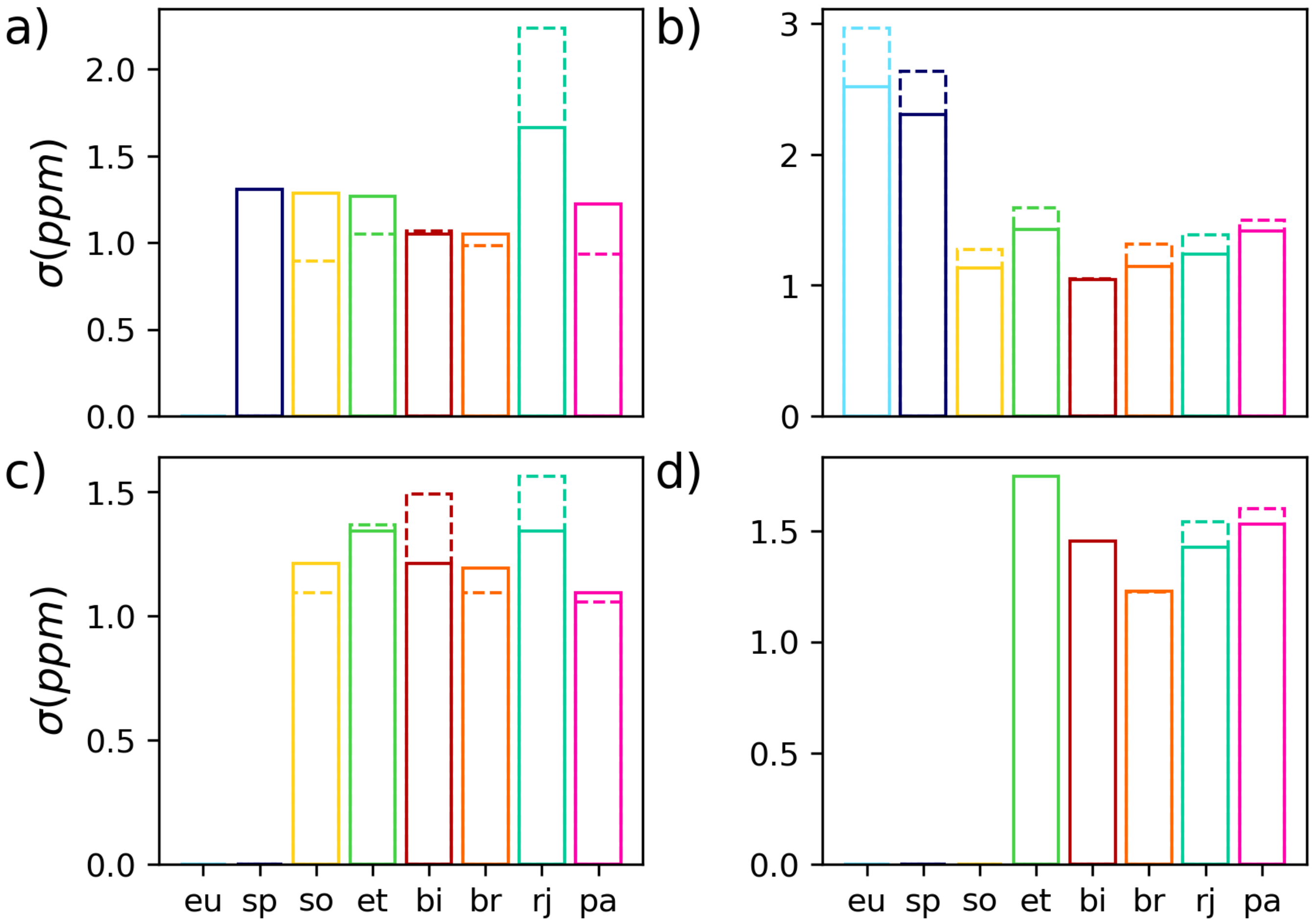
- Figure shows the seasonal bias at each TCCON site.
  - Bias is only shown if there are more than 10 coincident OCO-2 retrievals.
- The solid bars are the bias using the NN filter and the dashed bars are the bias using the B10 qa filter.
- In most cases where there is data from both filters the NN filter shows a lower bias compared to the B10 qa filter.



a) Spring (Mar, April, and May)
b) Summer (June, July, and Aug)
c) Fall (Sept, Oct, and Nov)
d) Winter (Dec, Jan, and Feb)

Validating the Neural Network – Precision

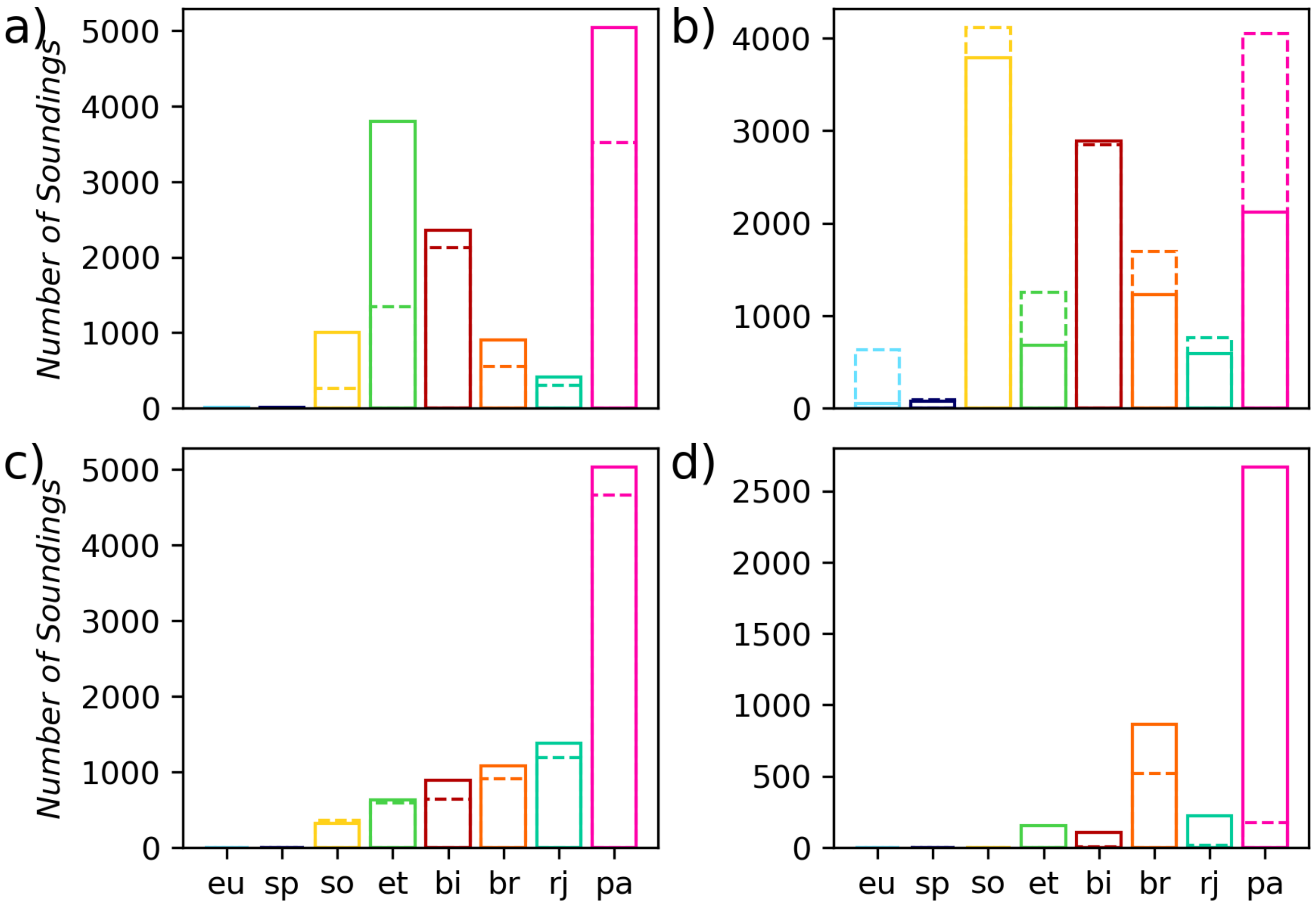
- Figure is the same as the previous slide except it shows the precision.
- In most cases the precision between the two filters are within ~0.3 ppm.
- The NN filter improves the precision at all sites in summer, except at Białystok (where the precision is the same).



a) Spring (Mar, April, and May)
b) Summer (June, July, and Aug)
c) Fall (Sept, Oct, and Nov)
d) Winter (Dec, Jan, and Feb)

Validating the Neural Network – Throughput

- Figure is the same as the previous slides except it shows the throughput.
- In most cases of spring, fall and winter the throughput is greater with the NN filter compared to the B10 qa filter.
- The throughput has decreased with the NN filter in summer but this results in improved bias and precision at the TCCON sites.



a) Spring (Mar, April, and May)
b) Summer (June, July, and Aug)
c) Fall (Sept, Oct, and Nov)
d) Winter (Dec, Jan, and Feb)