Using a support vector machine and satellite-based passive microwave brightness temperature observations within a land data assimilation system to improve snow characterization in North America

Yuan Xue and Barton A. Forman

Department of Civil and Environmental Engineering
University of Maryland College Park

June 27, 2017

Snow matters

- 1/6 of the world's population
- 3/4 of Western U.S. water supply
- 1/3 of California's water supply $\Longrightarrow$

  300 crops + 1/2 of the nation's fruits and nuts
- Snow is life = Water + Food + Power + Recreation
  (quoted from NASA snow workshop, 2016)

Snow risks

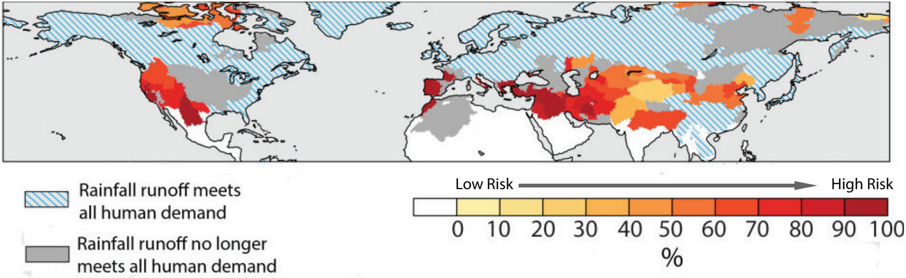


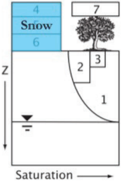
Figure: Projections of decreased potential for snowmelt water to supply human water demand by 2080. (Mankin et al., 2015 Environmental Research Letter)

Objective and Methodology

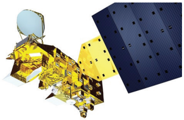
- Goal: Better characterize snow water equivalent (SWE) and snow depth across regional and continental scales.
- Study domain: North America
- Methodology: Data assimilation (DA) approach — Ensemble Kalman filter (EnKF) $\Longrightarrow$ 1D-EnKF
- Merge observations and model estimates

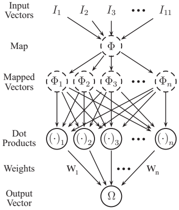
DA Formulation

Land surface model



Advanced Microwave Scanning Radiometer

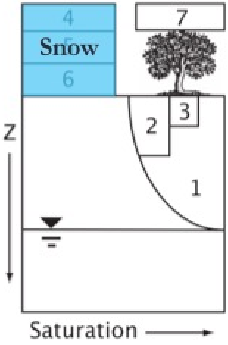




$$\underbrace{x^{i+}}_{\text{Posterior SWE}} = \underbrace{x^{i-}}_{\text{Prior SWE}} + \underbrace{K}_{\text{Kalman Gain}} \left[\underbrace{(y + v^{i})}_{\text{obs. + obs. error}} - \underbrace{\Phi(x^{i-})}_{\text{Machine learning algorithm based predictions}} \right]$$

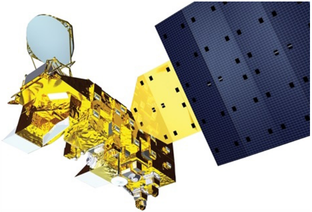
Figure: A conceptual representation of the update step in the EnKF, where the superscript i is the i^{th} replicate in the ensemble.

Forward model



- NASA Catchment model, forced by Modern-Era Retrospective-Analysis for Research and Applications (MERRA)
- three-layer snow model
- 25-km Equal-Area Scalable Earth Grid (EASE-Grid)
- initialize when seasonal snow cover at minimum

Passive microwave observations



- Advanced Microwave Scanning Radiometer EOS (AMSR-E)
- measures passive microwave emissions (i.e., brightness temperatures (Tb))
- from 01 September 2002 to 01 July 2011
- multi-frequency (10 GHz, 18 GHz, and 36 GHz), multi-polarization (H pol., and V pol.) Tbs
- 25-km Equal-Area Scalable Earth Grid (EASE-Grid)

Passive microwave observations over snow covered land

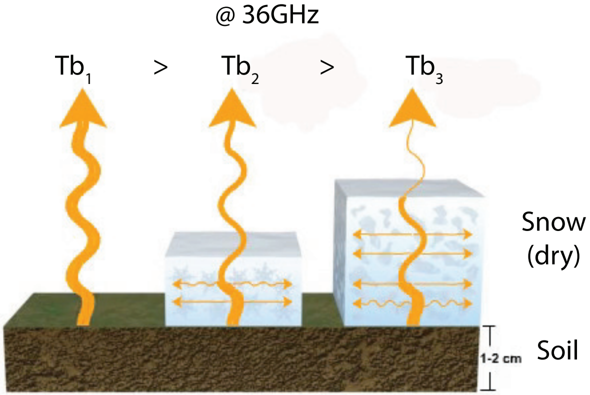
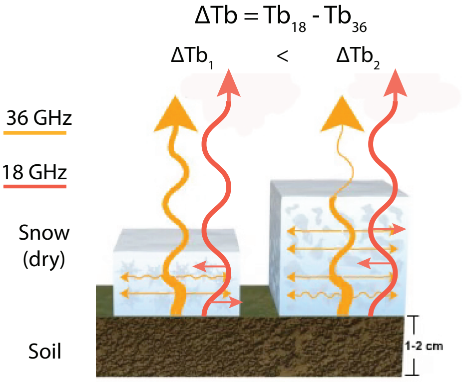


Figure: Radiance emissions from the surface at 36 GHz at either H. pol or V. pol.

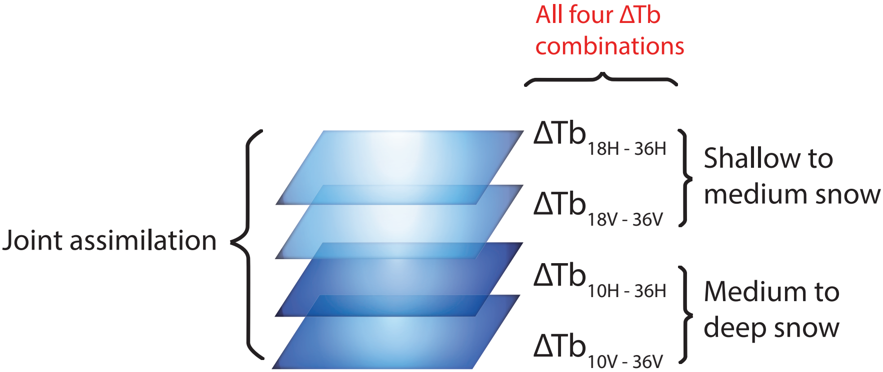
Relationship between SWE and ΔTb



- first-order approximation: ΔTb $\Uparrow$, SWE (or snow depth) $\Uparrow$
- similar approximation for $\Delta Tb = Tb_{10} - Tb_{36}$
- longer wavelength, deeper snow information

Assimilated Observations

- All ΔTbs (at both H and V pol.) contain SWE information, leading to synergistic effects when all observations are assimilated



Observation operator

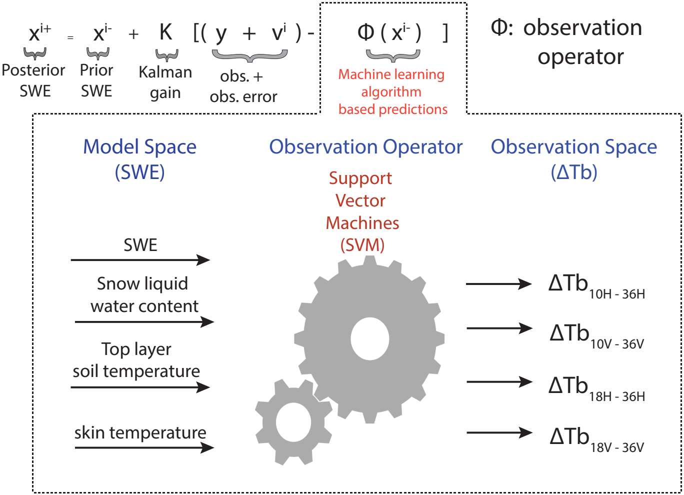


Figure: A conceptual representation of the machine learning algorithm based ΔTb prediction framework.

Others Terms in the EnKF Formulation

- Ensemble size: 40
- Perturb precipitation, shortwave radiation, and longwave radiation
- ΔTb observation error: $\sim \mathcal{N}(0, 3\mathrm{K}) \Longrightarrow$ same for all four channels
- SWE is the only state variable in the EnKF update
- Other snow-related modeled states were adjusted accordingly

SVM-DA evaluation

- Model-derived results vs. ground-based stations
  - National Water and Climate Center Snow Telemetry (SNOTEL) SWE
  - SNOTEL snow depth
  - NOAA Global Summary of the Day (GSOD) snow depth
  - USGS daily and cumulative discharge
- Model-derived results vs. available snow products
  - Canadian Meteorological Centre (CMC) snow depth
  - AMSR-E SWE
  - European Space Agency (ESA) GlobSnow SWE

**Note: Model-derived results:

(1) open-loop derived (OL; without assimilation) ensemble mean

(2) data assimilation derived (DA; with assimilation) ensemble mean

DA experiment results (Example)

DA reduced the bias by ~56% (from 0.13m to 0.05m) in SWE relative to OL;
DA reduced the RMSE by ~37% (from 0.18m to 0.11m) in SWE relative to OL;

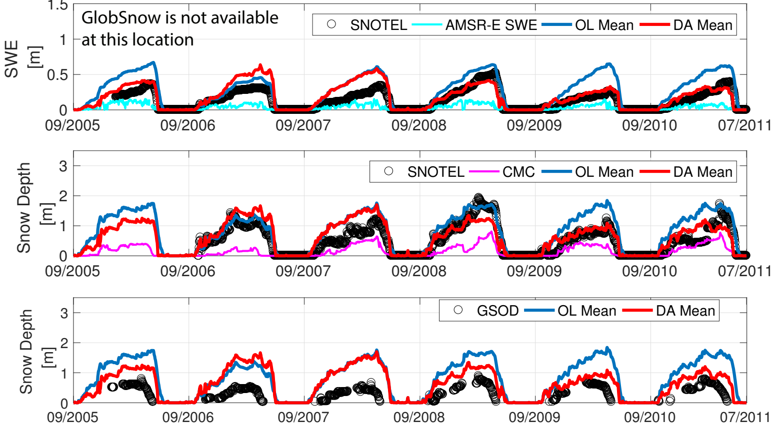
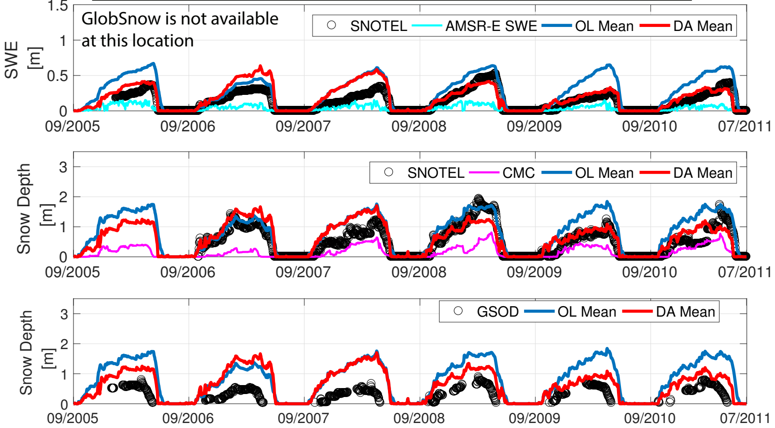


Figure: Comparison of OL-derived, DA-derived SWE and snow depth estimates, and estimates obtained from various snow products against colocated ground-based SNOTEL, and GSOD observations.

DA experiment results (Example)

DA reduced the bias by ~56% (from 0.13m to 0.05m) in SWE relative to OL;
DA reduced the RMSE by ~37% (from 0.18m to 0.11m) in SWE relative to OL;



- On average, slight improvements (relative to OL) were witnessed in DA experiments via comparison against ground-based stations (not shown).

Comparison against USGS discharge

- Normalized information contribution (NIC) (Kumar et al., 2009)
- NIC > 0: DA outperforms OL
- NIC $= 0$: DA is comparable to OL
- NIC < 0: DA degrades OL
- 13 major USGS gauged basins in Alaska (> 625 km^2, with at least two consecutive years of measurements record)

% of basins	**NICs > 0**	**NICs < 0**
vs. daily discharge	68.3%	31.7%
vs. cumulative discharge	84.7%	15.3%

- Relatively good snow estimates obtained from DA (relative to OL) also has the potential to translate into improved runoff estimates.

Comparison against snow products

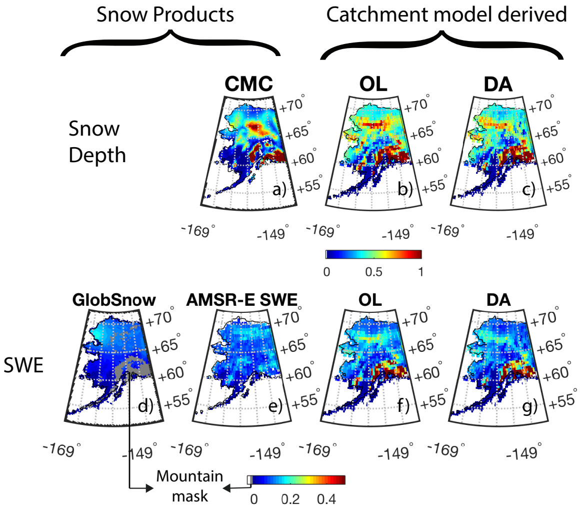
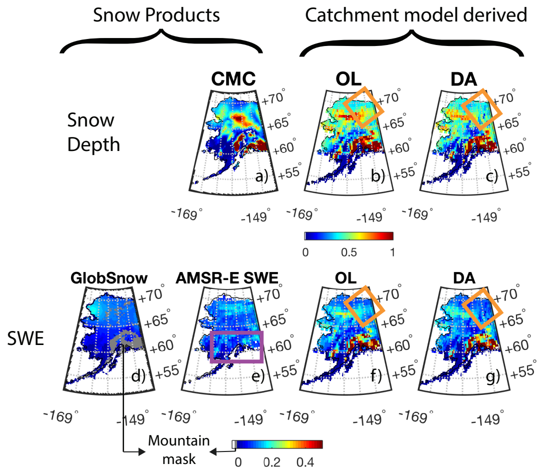


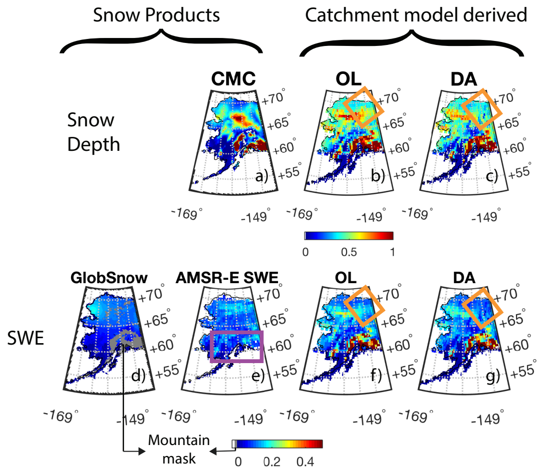
Figure: Comparison between OL, DA, and various snow products on 16 March 2003 in Alaska (Example).

Comparison against snow products



- AMSR-E SWE product underestimates SWE
- Compared with OL, DA moves closer to CMC and GlobSnow

Comparison against snow products



- Regrid all snow products onto the same 25-km EASE-Grid
- Compute time-averaged (from 2002-2011) difference between products and model derived estimates

Comparison against snow products

- Temporally-averaged estimates differences in Alaska, from 2002 to 2011

Domain-averaged mean difference $\pm$ standard deviation [m]	OL	DA
vs. CMC snow depth	0.023$\pm$0.468	0.003$\pm$0.460
vs. GlobSnow SWE	0.013$\pm$0.055	0.008$\pm$0.051
vs. AMSR-E SWE	0.034$\pm$0.110	0.030$\pm$0.108

- Comparing against AMSR-E SWE product, there is no significant difference between OL and DA
- Comparing against GlobSnow SWE product, DA SWE reduced the difference by 38% (relative to OL)
- Comparing against CMC snow depth product, DA snow depth reduced the difference by 87% (relative to OL)

Comparison against snow products

- Temporally-averaged estimates differences in Alaska, from 2002 to 2011

Domain-averaged mean difference $\pm$ standard deviation [m]	OL	DA
vs. CMC snow depth	0.023$\pm$0.468	0.003$\pm$0.460
vs. GlobSnow SWE	0.013$\pm$0.055	0.008$\pm$0.051
vs. AMSR-E SWE	0.034$\pm$0.110	0.030$\pm$0.108

- Comparing against AMSR-E SWE product, there is no significant difference between OL and DA
- Comparing against GlobSnow SWE product, DA SWE reduced the difference by 38% (relative to OL)
- Comparing against CMC snow depth product, DA snow depth reduced the difference by 87% (relative to OL)

Comparison against snow products

- Temporally-averaged estimates differences in Alaska, from 2002 to 2011

Domain-averaged mean difference $\pm$ standard deviation [m]	OL	DA
vs. CMC snow depth	0.023±0.468	0.003±0.460
vs. GlobSnow SWE	0.013±0.055	0.008±0.051
vs. AMSR-E SWE	0.034±0.110	0.030±0.108

- Comparing against AMSR-E SWE product, there is no significant difference between OL and DA
- Comparing against GlobSnow SWE product, DA SWE reduced the difference by 38% (relative to OL)
- Comparing against CMC snow depth product, DA snow depth reduced the difference by 87% (relative to OL)

Comparison against snow products

- Temporally-averaged estimates differences in Alaska, from 2002 to 2011

Domain-averaged mean difference ± standard deviation [m]	**OL**	**DA**
vs. CMC snow depth	0.023±0.468	0.003±0.460
vs. GlobSnow SWE	0.013±0.055	0.008±0.051
vs. AMSR-E SWE	0.034±0.110	0.030±0.108

- Comparing against AMSR-E SWE product, there is no significant difference between OL and DA
- Comparing against GlobSnow SWE product, DA SWE reduced the difference by 38% (relative to OL)
- Comparing against CMC snow depth product, DA snow depth reduced the difference by 87% (relative to OL)

Filter sub-optimality assessment

- Normalized innovation (NI) sequence
- NI sequence approximates white noise $\Longrightarrow$ "optimal" assimilation
- Violation of assumptions: nonlinear observation model operator $+$ nonlinear model dynamics $+$ non-Gaussian model errors

Domain-averaged	10H-36H	10V-36V	18H-36H	18V-36V
$\overline{NI}$ [-]	0.00	0.03	-0.04	-0.02
σ_{NI} [-]	1.56	1.46	1.16	1.14

- higher σ_{NI}: might underestimate observation errors and/or forecast errors

Conclusions and Future directions

- SVM can serve as an efficient and effective observation model operator within a radiance assimilation system.
- Investigate the effects of removing non-SWE related signals from the observations prior to integrating into the DA (IGARSS, July 2017)
- Compare model derived results with satellite-based terrestrial water storage information
- Conduct feasibility test of the current DA system on assimilating other satellite-based Tb observations

Conclusions and Future directions

- SVM can serve as an efficient and effective observation model operator within a radiance assimilation system.
- Investigate the effects of removing non-SWE related signals from the observations prior to integrating into the DA (IGARSS, July 2017)
- Compare model derived results with satellite-based terrestrial water storage information
- Conduct feasibility test of the current DA system on assimilating other satellite-based Tb observations

Conclusions and Future directions

- SVM can serve as an efficient and effective observation model operator within a radiance assimilation system.
- Investigate the effects of removing non-SWE related signals from the observations prior to integrating into the DA (IGARSS, July 2017)
- Compare model derived results with satellite-based terrestrial water storage information
- Conduct feasibility test of the current DA system on assimilating other satellite-based Tb observations

Conclusions and Future directions

- SVM can serve as an efficient and effective observation model operator within a radiance assimilation system.
- Investigate the effects of removing non-SWE related signals from the observations prior to integrating into the DA (IGARSS, July 2017)
- Compare model derived results with satellite-based terrestrial water storage information
- Conduct feasibility test of the current DA system on assimilating other satellite-based Tb observations

Thank you and Acknowledgment

Thank you!

- Financial support was provided by the NASA Earth and Space Science Fellowship Program (NNX14AN23H)
- LIBSVM library provided by National Taiwan University
- University of Maryland supercomputing resources

Additional slides

Cases where no analysis increments occurred

- No SVM (i.e., summertime, ocean)
- Error covariance between Catchment-SWE and SVM-Tb are zeros for all frequency combinations
- No Tb observations (i.e., 09/13/2002-09/19/2002, 10/30/2003-11/05/2003, 11/19/2004, 11/17/2005, 11/18/2006, 11/28/2007, 02/03/2010-02/04/2010, 01/26/2011)
- (Tb obs. + obs. error - Tb forecast) == 0 for all frequency combinations
- Strong radio frequency interference (RFI) hotspots (Tb @ 6 GHz - Tb @ 10 GHz $<$ -10K, if happens $>$ 20% of the entire time series)
- Invalid value of snow temperature, snow specific heat content, snow depth after re-distributing SWE into three layers
- Non-land grids (i.e., contain land ice or significant water bodies)
- Innovation is too large [innovation**2 $<$ (fac ** 2) * (2 * R)], R=4 K^2 or 9 K^2, fac=5, innovation max = 14 K or 21 K

Why to assimilate radiance?

- Avoid inconsistencies in the use of ancillary data between the assimilation system and pre-processed geophysical retrievals
- e.g., soil or vegetation conditions

Why to assimilate radiance?

- What SWE observations (which SWE observation product) to use for assimilation?
  - point-sale SWE observations from ground-based stations (maybe yes, maybe no)
  - AMSR-E SWE product (x)
  - ESA SWE product (x)
- How to evaluate the SVM-SWE-DA work if assimilating point-scale SWE observations?
  - Compare with ground-based SWE observations itself (x)

Why not bring LAI into SVM training?

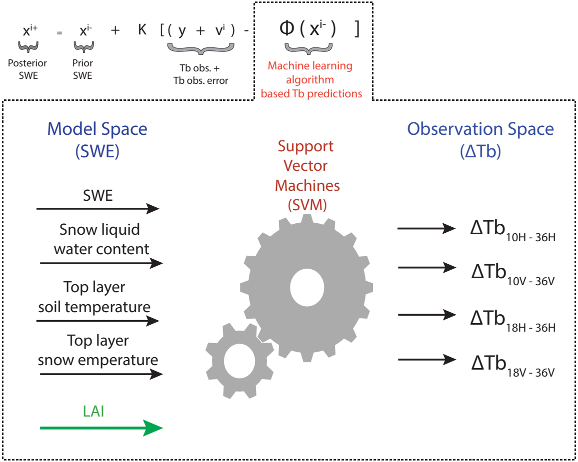


Figure: A modified conceptual representation of the machine learning algorithm based SWE prediction framework with LAI as the input.

Why not bring LAI into SVM training?

- Go back to the "Godilocks" graph... trade-off between SWE sensitivity and prediction accuracy
  - Model complexity $\Uparrow$, SWE sensitivity $\Downarrow$, close-to-zero Kalman gains, no update will take place
- Different from SWE, snow temperature etc., LAI is an 8-day product, which barely has day to day variations (i.e., step-function-like)

Why to select SVM with four input states?

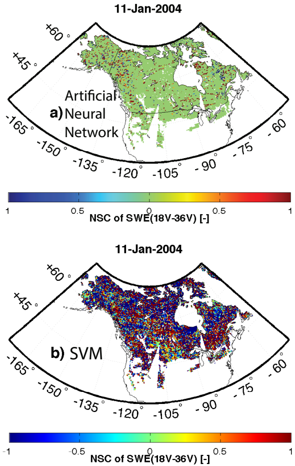


Figure: Comparisons of machine learning algorithms in SWE sensitivity.

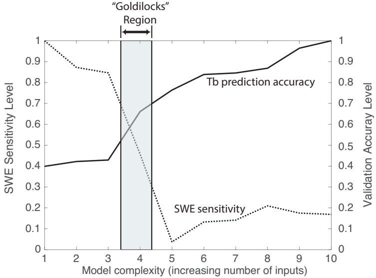


Figure: Trade-off between SWE sensitivity level and prediction accuracy level.

Are the state-of-the-art snow products really state-of-the-art?

- Q: If we have already got these state-of-the-art snow products, why do you still need DA to improve your model estimates?
- We call these snow products as "state-of-the-art" because many previous studies use these products as their references
- Comparison against snow products become more useful at regions without in-situ observations
- "state-of-the-art" products do not equal to "perfect" products
  - ESA GlobSnow SWE product does not have estimates at high mountains
  - CMC SWE product has a fixed snow density everywhere
  - CMC snow depth product assimilate in-situ dataset from WMO directly without considering whether the ground-based stations are representative of the 24km resolution grids
  - AMSR-E SWE product has significant negative bias
- Our efforts to integrate model estimates with observations are useful — new global snow product

How to compare model results with point-scale observations?

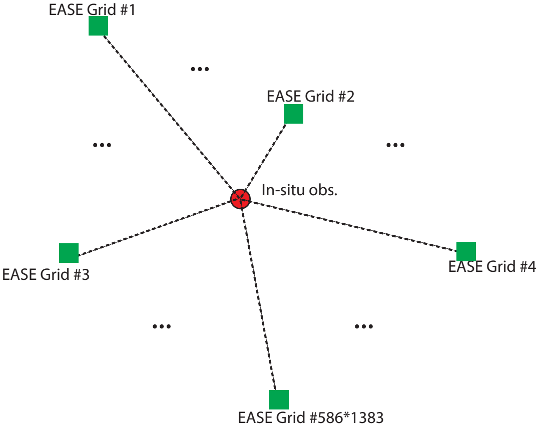


Figure: A conceptual representation of the comparison scheme used in SVM-DA evaluation.

Computational time

- Training procedure is independent of Catchment model run
- Training time varies as a function of the sample size (large domain vs. small domain; winter vs. summer)
- Using fortnight 01 (01 Jan to 14 Jan) in Quebec and Newfoundland, Canada as a example, 2914 Catchments, 9 year (2002 - 2011)
- Training time: $\sim$6-7 hours (6 cores)

Computational time

- Run time: ∼12-20 hours (1 core) — one pixel (from 2002 to 2011)
- Run time: ∼1.5-2.0 days (20 cores) — North America (per year)

Soil effects in SWE retrieval?

- Not too sensitive to soil water content ($>$ 10 GHz)
- The use of a spectral difference algorithm is assumed to minimize many of the errors in the retrieval, such as the dielectric constant of the soil, and the soil surface roughness. (Clifford et al., 2002)
- Error bars for each (LAI-transmissivity) pair account for soil effects

NICs computation

- $NIC_{RMSE} = \frac{RMSE_{OL} - RMSE_{DA}}{RMSE_{OL}}$
- $NIC_{NSE} = \frac{NSE_{DA} - NSE_{OL}}{1 - NSE_{OL}}$
- Kumar et al., 2009, 2014

Thank you!

Questions and/or Comments?