Towards the Development of a Global, Satellite-based, Terrestrial Snow Mission Planning Tool

Co-authors: **Sujay Kumar**[1], **Jacqueline Le Moigne**[2], **and Sreeja Nag**[2,3]

1=NASA GSFC - Hydrological Sciences; 2=NASA GSFC - Software Engineering; 3=Bay Area Environmental Research Institute

Bart Forman

Assistant Professor, University of Maryland
The Deborah J. Goodings Professor of Global Sustainability
Department of Civil and Environmental Engineering

June 27^{th}, 2017

Importance of Snow

- Snow is a significant contributor to terrestrial freshwater supply
  - Vital resource for ~billion people worldwide
- Not exactly sure how much snow is out there
  - Significant uncertainty
- Leverage suite of remotely-sensed satellite observations
  - Visible Spectrum ⇒ primarily measures snow extent
  - Microwave Spectrum ⇒ primarily measures snow mass
  - Gravimetry ⇒ large spatiotemporal resolution, not an imager
- Need for computationally efficient observation operator
  - Eventual use in data assimilation framework
  - $\underbrace{\mathbf{y}_i^+}_{\text{posterior}} = \underbrace{\mathbf{y}_i^-}_{\text{prior}} + \underbrace{\mathbf{K}}_{\text{Kalman gain}} \left[\underbrace{Z_{T_b}}_{\substack{\text{observation}\\ \text{via satellite}}} - \underbrace{h(\mathbf{y}_i^-)}_{\substack{\text{prediction via}\\ \text{``machine learning''}}} \right]$
- Goal is to improve SWE estimation at regional and continental scales based on conditional probability, $p(y|z)$

Importance of Snow

- Snow is a significant contributor to terrestrial freshwater supply
  - Vital resource for ~billion people worldwide
- Not exactly sure how much snow is out there
  - Significant uncertainty
- Leverage suite of remotely-sensed satellite observations
  - Visible Spectrum ⇒ primarily measures snow extent
  - Microwave Spectrum ⇒ primarily measures snow mass
  - Gravimetry ⇒ large spatiotemporal resolution, not an imager
- Need for computationally efficient observation operator
  - Eventual use in data assimilation framework
  - $\underbrace{\mathbf{y}_i^+}_{\text{posterior}} = \underbrace{\mathbf{y}_i^-}_{\text{prior}} + \underbrace{\mathbf{K}}_{\text{Kalman gain}} \left[\underbrace{Z_{T_b}}_{\substack{\text{observation}\\ \text{via satellite}}} - \underbrace{h(\mathbf{y}_i^-)}_{\substack{\text{prediction via}\\ \text{"machine learning"}}} \right]$
- Goal is to improve SWE estimation at regional and continental scales based on conditional probability, $p(y|z)$

Importance of Snow

- Snow is a significant contributor to terrestrial freshwater supply
  - Vital resource for ~billion people worldwide
- Not exactly sure how much snow is out there
  - Significant uncertainty
- Leverage suite of remotely-sensed satellite observations
  - Visible Spectrum ⇒ primarily measures snow extent
  - Microwave Spectrum ⇒ primarily measures snow mass
  - Gravimetry ⇒ large spatiotemporal resolution, <u>not</u> an imager
- Need for computationally efficient observation operator
  - Eventual use in data assimilation framework
  - $\underbrace{\mathbf{y}_i^+}_{\text{posterior}} = \underbrace{\mathbf{y}_i^-}_{\text{prior}} + \underbrace{\mathbf{K}}_{\text{Kalman gain}} \left[\underbrace{Z_{T_b}}_{\text{observation via satellite}} - \underbrace{h(\mathbf{y}_i^-)}_{\text{prediction via "machine learning"}} \right]$
- Goal is to improve SWE estimation at regional and continental scales based on conditional probability, $p(y|z)$

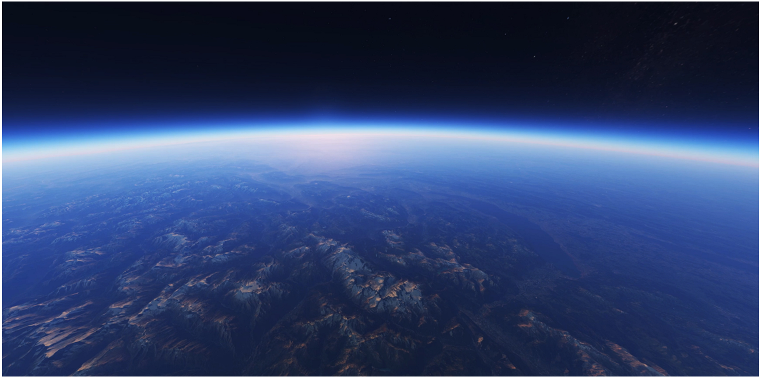
Importance of Snow

- Snow is a significant contributor to terrestrial freshwater supply
  - Vital resource for ~billion people worldwide
- Not exactly sure how much snow is out there
  - Significant uncertainty
- Leverage suite of remotely-sensed satellite observations
  - Visible Spectrum ⇒ primarily measures snow extent
  - Microwave Spectrum ⇒ primarily measures snow mass
  - Gravimetry ⇒ large spatiotemporal resolution, not an imager
- Need for computationally efficient observation operator
  - Eventual use in data assimilation framework
  - $\underbrace{\mathbf{y}_i^+}_{\text{posterior}} = \underbrace{\mathbf{y}_i^-}_{\text{prior}} + \underbrace{\mathbf{K}}_{\text{Kalman gain}} \left[\underbrace{Z_{T_b}}_{\substack{\text{observation} \\ \text{via satellite}}} - \underbrace{h(\mathbf{y}_i^-)}_{\substack{\text{prediction via} \\ \text{“machine learning”}}} \right]$
- Goal is to improve SWE estimation at regional and continental scales based on conditional probability, $p(y|z)$

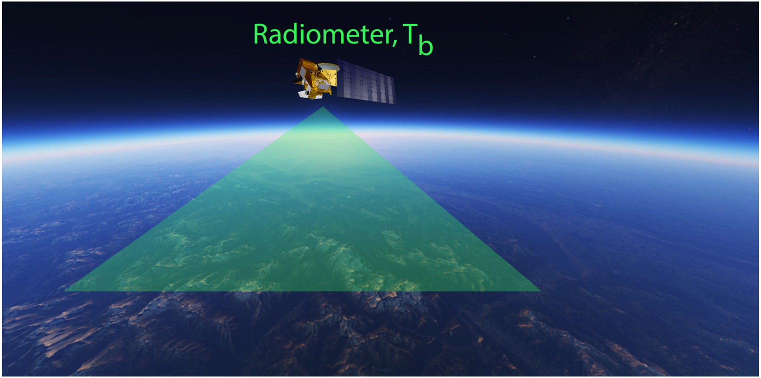
Importance of Snow

- Snow is a significant contributor to terrestrial freshwater supply
  - Vital resource for ∼billion people worldwide
- Not exactly sure how much snow is out there
  - Significant uncertainty
- Leverage suite of remotely-sensed satellite observations
  - Visible Spectrum ⇒ primarily measures snow extent
  - Microwave Spectrum ⇒ primarily measures snow mass
  - Gravimetry ⇒ large spatiotemporal resolution, not an imager
- Need for computationally efficient observation operator
  - Eventual use in data assimilation framework
  - $\underbrace{\mathbf{y}_i^+}_{\text{posterior}} = \underbrace{\mathbf{y}_i^-}_{\text{prior}} + \underbrace{\mathbf{K}}_{\text{Kalman gain}} [\underbrace{Z_{T_b}}_{\text{observation via satellite}} - \underbrace{h(\mathbf{y}_i^-)}_{\text{prediction via “machine learning”}}]$
- Goal is to improve SWE estimation at regional and continental scales based on conditional probability, $p(y|z)$

Satellite-derived Snow "Information"



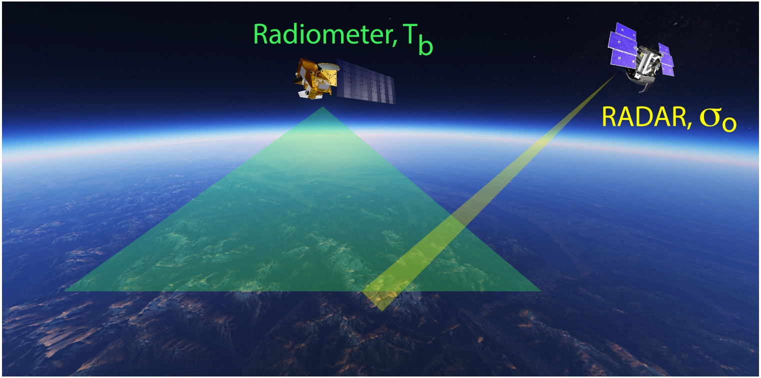
Satellite-derived Snow "Information"



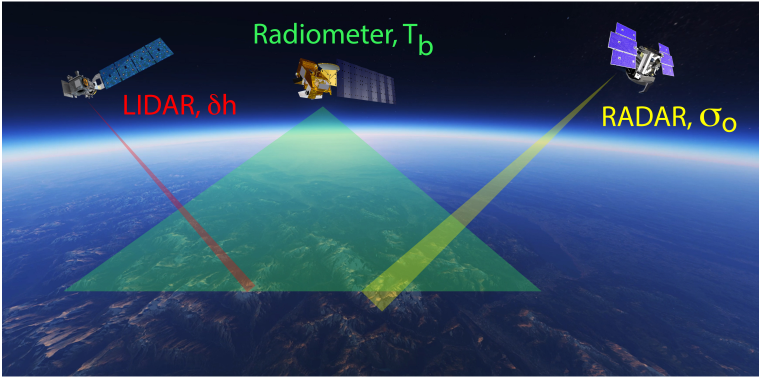
Satellite-derived Snow "Information"

UMD-PSU DA 2017

Bart Forman



Satellite-derived Snow "Information"



Research Objectives

Science and mission planning questions:

1. What observational records are needed (in space and time) to maximize terrestrial snow experimental utility?
2. How might observations be coordinated (in space and time) to maximize this utility?
3. What is the additional utility associated with an additional observation?
4. How can future mission costs be minimized while ensuring Science requirements are fulfilled?

Science and mission planning questions:

1. What observational records are needed (in space and time) to maximize terrestrial snow experimental utility?
2. How might observations be coordinated (in space and time) to maximize this utility?
3. What is the additional utility associated with an additional observation?
4. How can future mission costs be minimized while ensuring Science requirements are fulfilled?

Research Objectives

Science and mission planning questions:

1. What observational records are needed (in space and time) to maximize terrestrial snow experimental utility?
2. How might observations be coordinated (in space and time) to maximize this utility?
3. What is the additional utility associated with an additional observation?
4. How can future mission costs be minimized while ensuring Science requirements are fulfilled?

Research Objectives

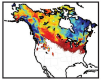
Science and mission planning questions:

1. What observational records are needed (in space and time) to maximize terrestrial snow experimental utility?
2. How might observations be coordinated (in space and time) to maximize this utility?
3. What is the additional utility associated with an additional observation?
4. How can future mission costs be minimized while ensuring Science requirements are fulfilled?

Observing System Simulation Experiment

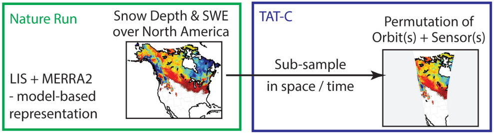
Nature Run

Snow Depth & SWE over North America

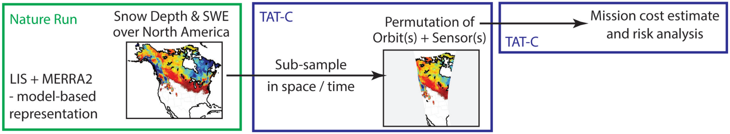


LIS + MERRA2 - model-based representation

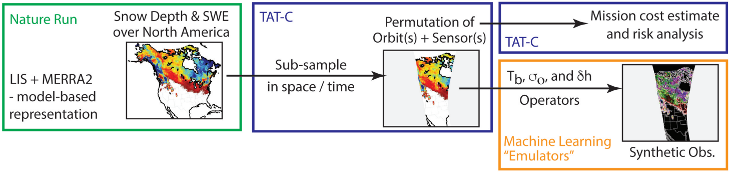
Observing System Simulation Experiment



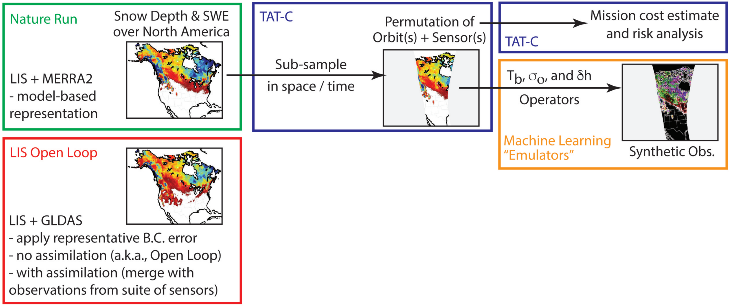
Observing System Simulation Experiment



Observing System Simulation Experiment



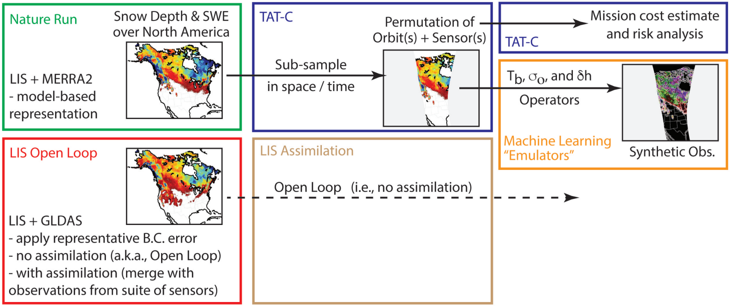
Observing System Simulation Experiment



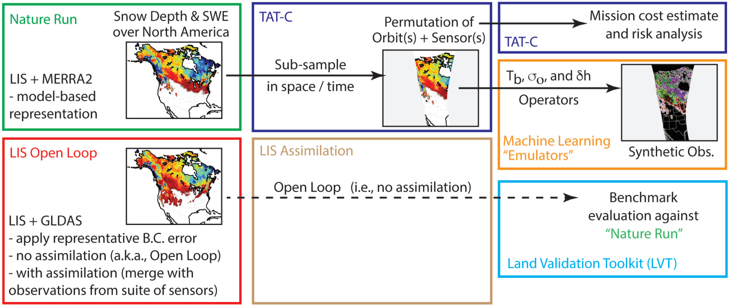
Observing System Simulation Experiment

UMD-PSU DA
2017

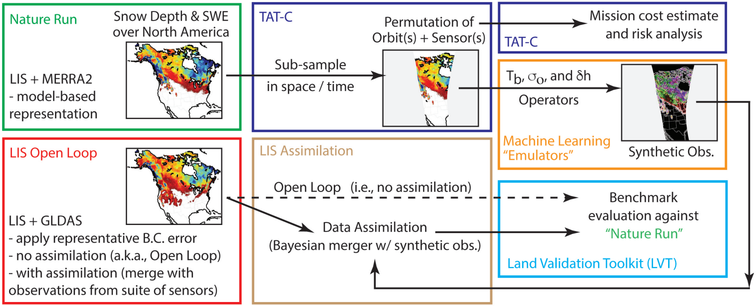
Bart Forman



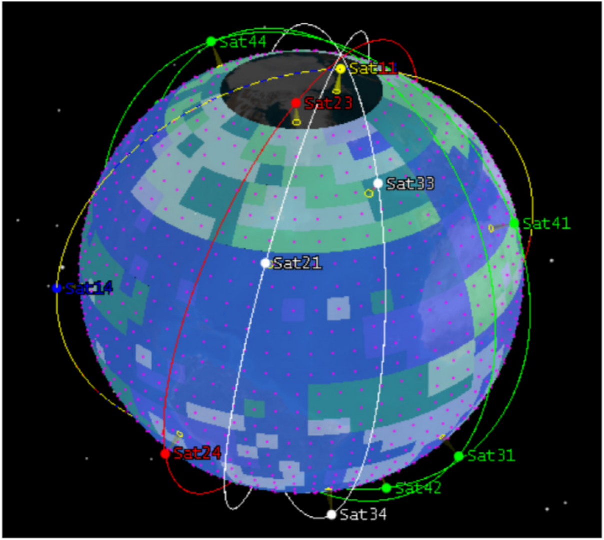
Observing System Simulation Experiment



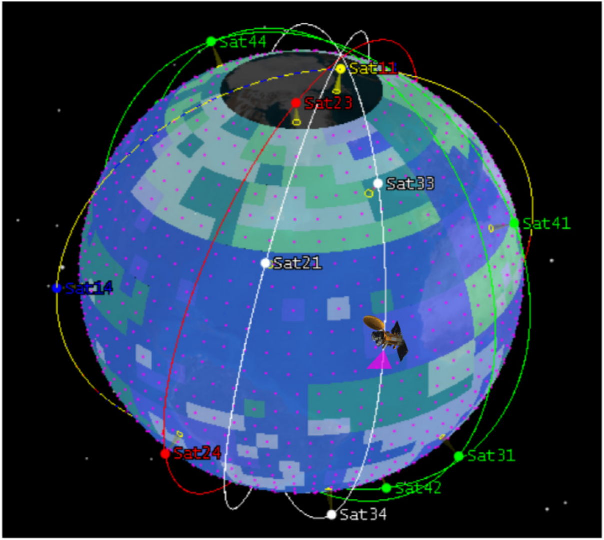
Observing System Simulation Experiment



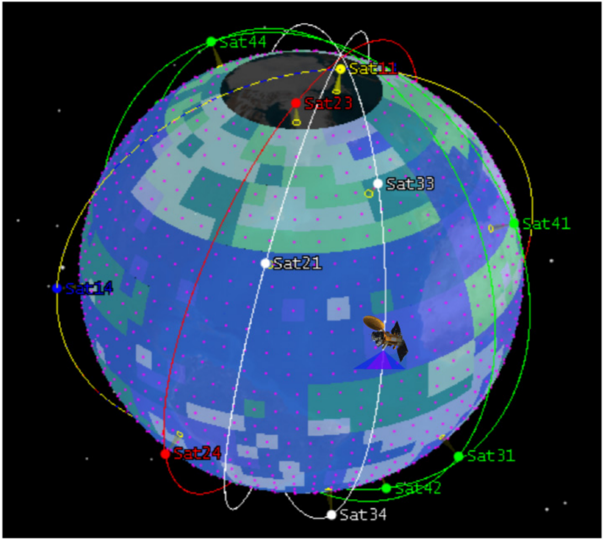
TAT-C Orbital Simulator



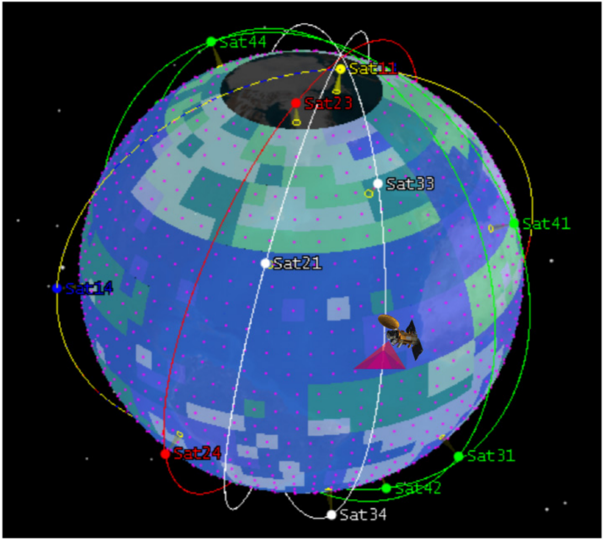
TAT-C Orbital Simulator



TAT-C Orbital Simulator



TAT-C Orbital Simulator

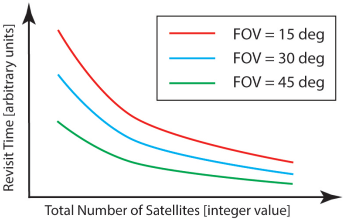


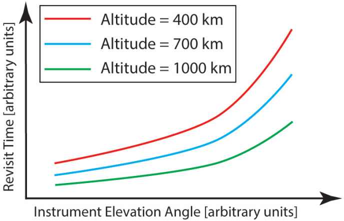
TAT-C Orbital Simulator

“Comb” Viewing $\mapsto$ Single Platform

“Comb” Viewing $\mapsto$ Constellation

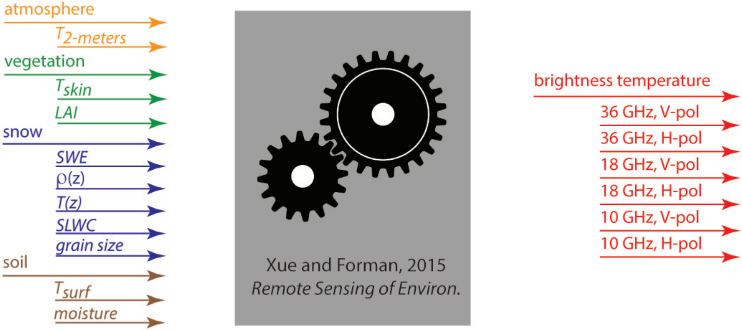
Trade-off Space: Coverage vs. Resolution





- Explore trade-off between engineering and science
  - Field-of-View (FOV)?
  - Platform altitude?
  - Repeat cycle?
  - Single platform vs. constellation?
  - Orbital configuration(s)?
- How do we get the most scientific bang for our buck?

Machine Learning "Emulators"



Physically-based
Land Surface Model(s)

Observation Operator
(Forman et al., 2013;
Forman and Reichle, 2014;
Forman and Xue, 2016)

Multi-frequency,
Multi-polarization
Training Targets

Machine Learning "Emulators"

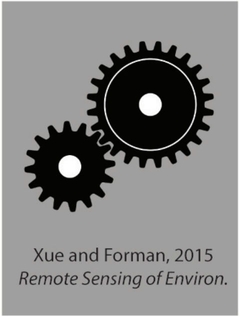
atmosphere
- $T_{2\text{-}meters}$

vegetation
- T_{skin}
- LAI

snow
- SWE
- ρ(z)
- T(z)
- SLWC
- grain size

soil
- T_{surf}
- moisture

Physically-based
Land Surface Model(s)



Observation Operator
(Forman et al., 2013; Forman and Reichle, 2014; Forman and Xue, 2016)

spectral difference
- 18V - 36V
- 18H - 36H
- 10V - 36V
- 10H - 36H

Multi-frequency,
Multi-polarization
Training Targets

Machine Learning "Emulators"

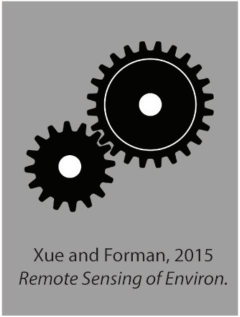
atmosphere
- $T_{2\text{-meters}}$

vegetation
- T_{skin}
- *LAI*

snow
- *SWE*
- $\rho(z)$
- *T(z)*
- *SLWC*
- *grain size*

soil
- T_{surf}
- *moisture*

Physically-based
Land Surface Model(s)

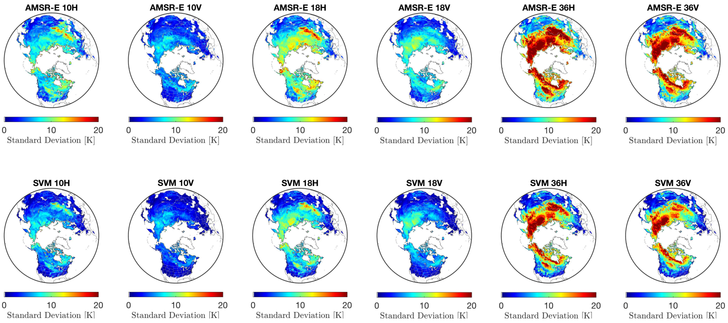


Observation Operator
(Forman et al., 2013;
Forman and Reichle, 2014;
Forman and Xue, 2016)

backscatter
- σ_{VH}
- σ_{VV}

Multi-frequency,
Multi-polarization
Training Targets

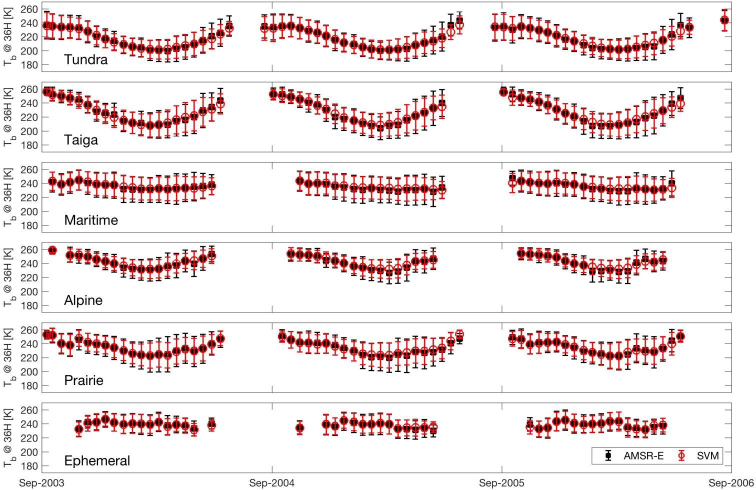
Spatiotemporal Variability



Spatiotemporal Variability

UMD-PSU DA 2017

Bart Forman



Relevancy Scenarios

- **Scenario 1**: Benchmark Analysis
  - Passive MW Assimilation only
- **Scenario 2**: Comparative Analysis
  - Passive MW vs. Active MW vs. LIDAR
- **Scenario 3**: Multi-sensor Analysis
  - single-sensor platform
  - multi-sensor platform
  - constellation of sensors

Research Summary

- Global snow mission will require evidence of achievable science via OSSE . . . or some other means
- NASA LIS provides "nature run" plus assimilation framework
- TAT-C provides spatiotemporal sub-sampling of observations, including cost estimates and risk assessments
- Machine learning maps model state(s) into observation space (i.e., T_b and σ_0)
  - Enables integration of T_b, σ_0, and δh in geophysical realm (i.e., SWE and snow depth)
  - Multiple frequencies/polarizations/observations allow for flexibility and modularity in DA framework
- Snow OSSE is on-going $\longrightarrow$ open to ideas + suggestions!

Research Summary

- Global snow mission will require evidence of achievable science via OSSE . . . or some other means
- NASA LIS provides "nature run" plus assimilation framework
- TAT-C provides spatiotemporal sub-sampling of observations, including cost estimates and risk assessments
- Machine learning maps model state(s) into observation space (i.e., T_b and σ_0)
  - Enables integration of T_b, σ_0, and δh in geophysical realm (i.e., SWE and snow depth)
  - Multiple frequencies/polarizations/observations allow for flexibility and modularity in DA framework
- Snow OSSE is on-going $\longrightarrow$ open to ideas + suggestions!

Research Summary

- Global snow mission will require evidence of achievable science via OSSE . . . or some other means
- NASA LIS provides "nature run" plus assimilation framework
- TAT-C provides spatiotemporal sub-sampling of observations, including cost estimates and risk assessments
- Machine learning maps model state(s) into observation space (i.e., T_b and σ_0)
  - Enables integration of T_b, σ_0, and δh in geophysical realm (i.e., SWE and snow depth)
  - Multiple frequencies/polarizations/observations allow for flexibility and modularity in DA framework
- Snow OSSE is on-going $\longrightarrow$ open to ideas + suggestions!

Research Summary

- Global snow mission will require evidence of achievable science via OSSE . . . or some other means
- NASA LIS provides "nature run" plus assimilation framework
- TAT-C provides spatiotemporal sub-sampling of observations, including cost estimates and risk assessments
- Machine learning maps model state(s) into observation space (i.e., T_b and σ_0)
  - Enables integration of T_b, σ_0, and δh in geophysical realm (i.e., SWE and snow depth)
  - Multiple frequencies/polarizations/observations allow for flexibility and modularity in DA framework
- Snow OSSE is on-going $\longrightarrow$ open to ideas + suggestions!

Research Summary

- Global snow mission will require evidence of achievable science via OSSE ... or some other means
- NASA LIS provides "nature run" plus assimilation framework
- TAT-C provides spatiotemporal sub-sampling of observations, including cost estimates and risk assessments
- Machine learning maps model state(s) into observation space (i.e., T_b and σ_0)
  - Enables integration of T_b, σ_0, and δh in geophysical realm (i.e., SWE and snow depth)
  - Multiple frequencies/polarizations/observations allow for flexibility and modularity in DA framework
- Snow OSSE is on-going $\longrightarrow$ open to ideas + suggestions!

Thank You.

Questions and/or Comments?

Financial support provided by:
NASA New Investigator Program (NNX14AI49G)
NASA GRACE-FO Science Team (NNX16AF17G)
NASA High Mountain Asia Science Team (NNX17AC15G)

High-performance computing support provided by
UMD's Division of Information Technology

SVM Mathematical Framework (1 of 2)

For parameters $C > 0$ and $\varepsilon > 0$, the **standard (primal)** form is:

$$\begin{aligned}
\underset{\mathbf{w},\,\delta,\,\boldsymbol{\xi},\,\boldsymbol{\xi}^*}{\text{minimize}} \quad & \frac{1}{2}\langle \mathbf{w}\cdot\mathbf{w}\rangle + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \\
\text{subject to} \quad & \langle \mathbf{w}\cdot\phi(\mathbf{x}_i)\rangle + \delta - z_i \le \varepsilon + \xi_i \\
& z_i - \langle \mathbf{w}\cdot\phi(\mathbf{x}_i)\rangle - \delta \le \varepsilon + \xi_i^* \\
& \xi_i, \xi_i^* \ge 0, i = 1, 2, \ldots, m.
\end{aligned}$$

where m is the available number of T_b measurements in time (for a given location in space), z_i is a T_b measurement at time i, and $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*$ are slack variables.

SVM Mathematical Framework (2 of 2)

Primal optimization is commonly solved in **dual form** as:

$$\begin{aligned}
\underset{\alpha_i,\,\alpha_i^*}{\text{minimize}} \quad & \frac{1}{2}\sum_{i,j=1}^{m}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\left\langle\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)\right\rangle \\
& + \varepsilon\sum_{i=1}^{m}\left(\alpha_i+\alpha_i^*\right)-\sum_{i=1}^{m} z_i\left(\alpha_i-\alpha_i^*\right) \\
\text{subject to} \quad & \sum_{i=1}^{m}\left(\alpha_i-\alpha_i^*\right)=0, \\
& \alpha_i\,,\,\alpha_i^*\in[0\,,\,C]\,,\; i=1,2,\ldots,m
\end{aligned}$$

where α_i and α_i^* are Lagrangian multipliers, $\langle\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)\rangle$ is the inner dot product of $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$, ε is the specified error tolerance, and C is a positive constant that dictates a penalized loss during training.