Monitoring of riparian vegetation dynamics using multi-source remote sensing and machine learning techniques: a preliminary study

Huiran Jin, Associate Professor
New Jersey Institute of Technology (NJIT)

Project overview

- **Title**: "Integrating multi-scale remote sensing and mechanistic modeling to assess riparian ecosystem dynamics and feedbacks to hydroclimate variability".
- **Sponsor**: NSF's Macrosystems Biology and NEON-Enabled Science program.
- **Overarching goal**: To mechanistically link riparian vegetation dynamics to hydroclimate variations and assess the functional importance of riparian ecosystems to macrosystem fluxes of carbon and water.
- **Specific objectives**:
  - 1) document vegetation dynamics during the past decades,
  - 2) examine how shifts in vegetation emerge from changes in physiological functions, and
  - 3) assess the feedbacks of local-scale changes to macrosystem fluxes.

"a time-series record of riparian vegetation on a seasonal-to-yearly basis"

Multi-type data fusion

– Part I: Review of existing studies

- We conducted a detailed review of existing literature that:
  - were published in peer-reviewed journals from 2000 until Oct 2021,
  - explored pixel- or feature-level fusion of multi-source data collected by different types of sensors (optical, radar, lidar),
  - aimed at mapping land cover (and/or land use) distribution in a wide range of environments from urban, desert and coastal zones, to forest, cropland and rangeland,

  and most importantly,

  - explicitly presented **classification accuracies** from which improvement gained by the fusion could be assessed, a major difference compared to existing reviews of data fusion that are approach oriented.

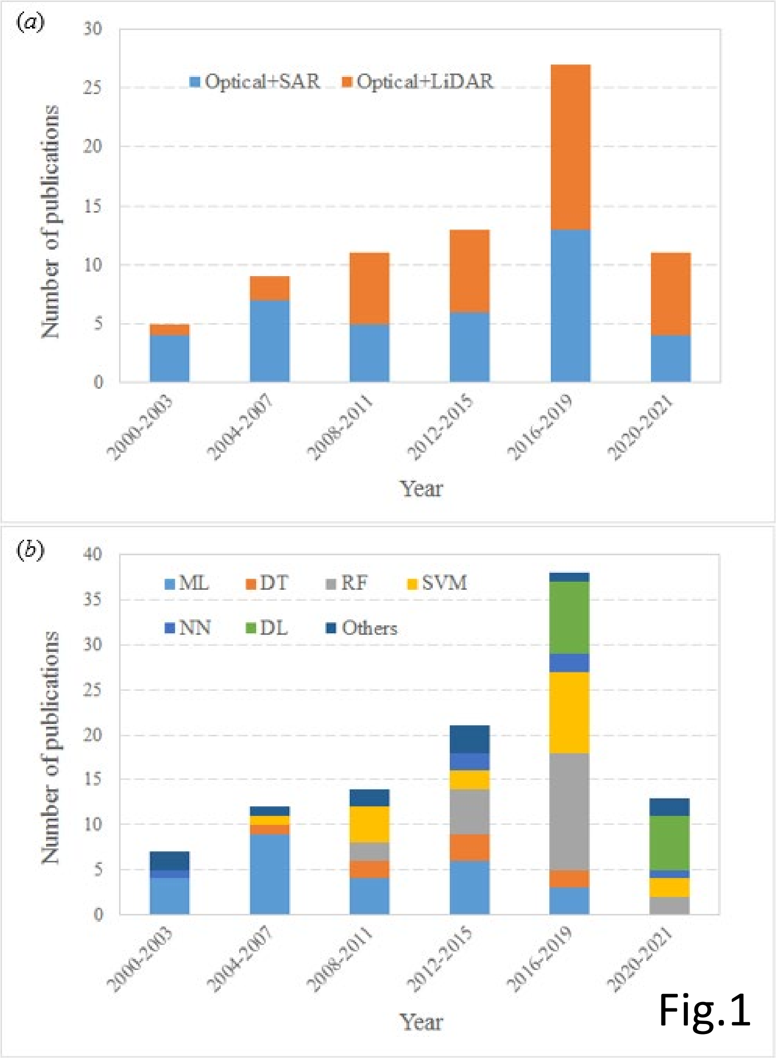


Fig.1

- Fig.1 shows growth of classification research over the past two decades based on (a) optical–SAR and optical–LiDAR fusion, and (b) algorithms applied.

- A more obvious increasing trend over time can be observed for optical–LiDAR fusion studies (a).
- With the advent of big data and in parallel with the rapid computational advances, deep learning, a specialized subset of machine learning and a fastest-growing trend in data sciences, has gained increasing interest in remote sensing image classification (b).

Optical & SAR

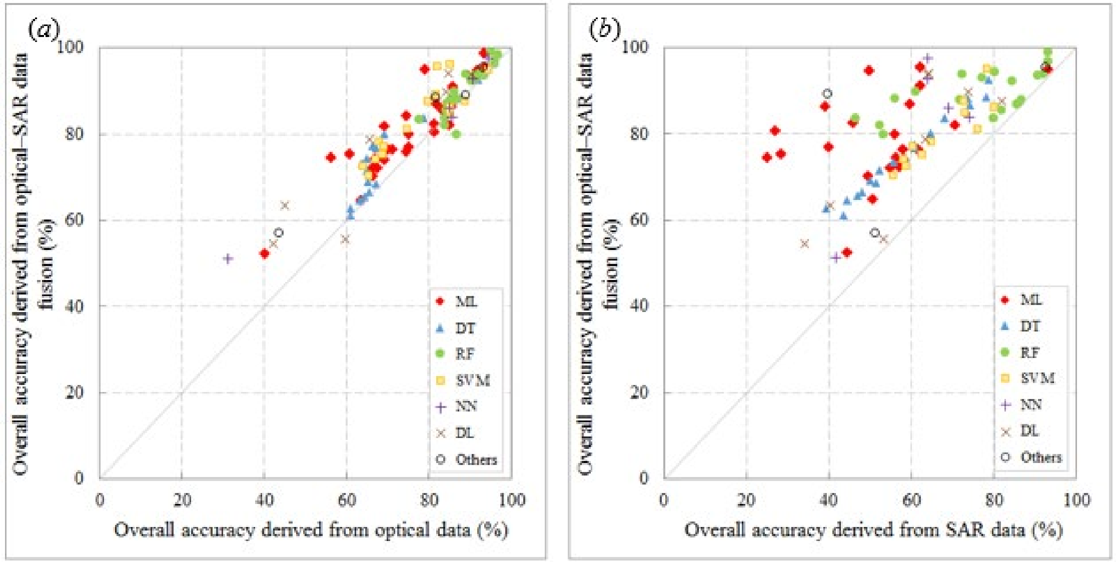


Fig.2. (*a*) optical vs. optical–SAR classification, and (*b*) SAR vs. optical–SAR classification.

Optical & LiDAR

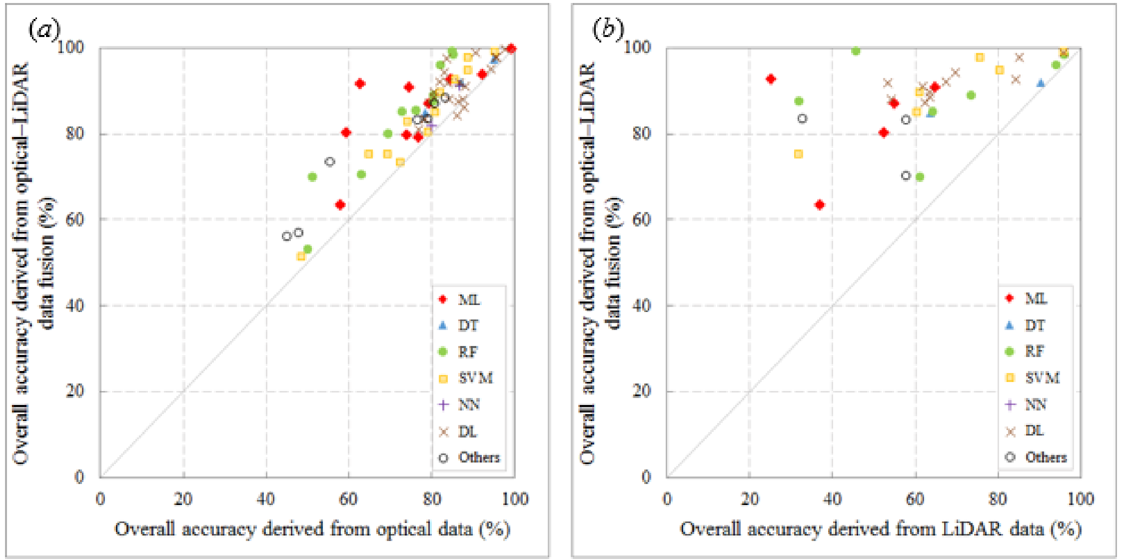


Fig. 3. (*a*) optical vs. optical–LiDAR classification, and (*b*) LiDAR vs. optical–LiDAR classification.

- Only in rare cases did the fusion of optical & SAR data and optical & LiDAR data end up with slightly lower accuracies than those derived from optical data alone (shown as points located below the 1:1 line in Fig.2a, Fig.3a).

- We present here a case study to investigate further the benefits of fusion, and to address two previously unanswered questions:

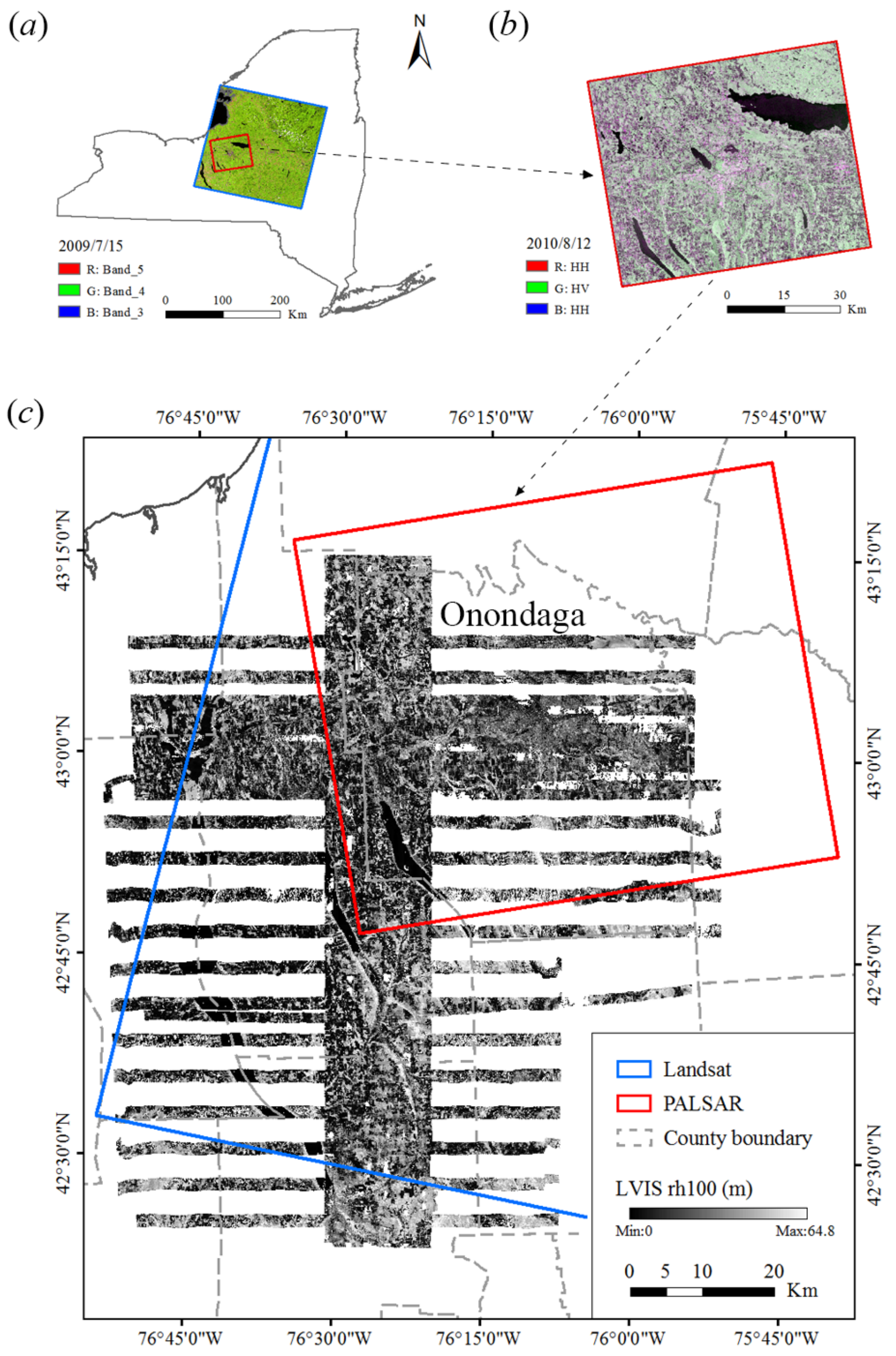
  - Can classification accuracy be improved by combining the three different types of data with a commonly used classifier (random forest), and

  - To what degree the inclusion of seasonal spectral and scattering variations of certain vegetation classes may impact the accuracy when optical, SAR, and LiDAR data are used in combination.

Multi-type data fusion
– Part II: Case study

- Study area: Central NY State (US).
- Data: Multitemporal Landsat (optical) and PALSAR (radar), and LVIS (airborne lidar).
- Features extracted: spectral (52), scattering (132), and structural (4).
- Digital orthoimages at 1- and 2-ft resolution from aerial photos were used to prepare reference data for both training and validation purposes.
- Classification: random forest.
- Accuracy assessment: error matrix.

Fig.4



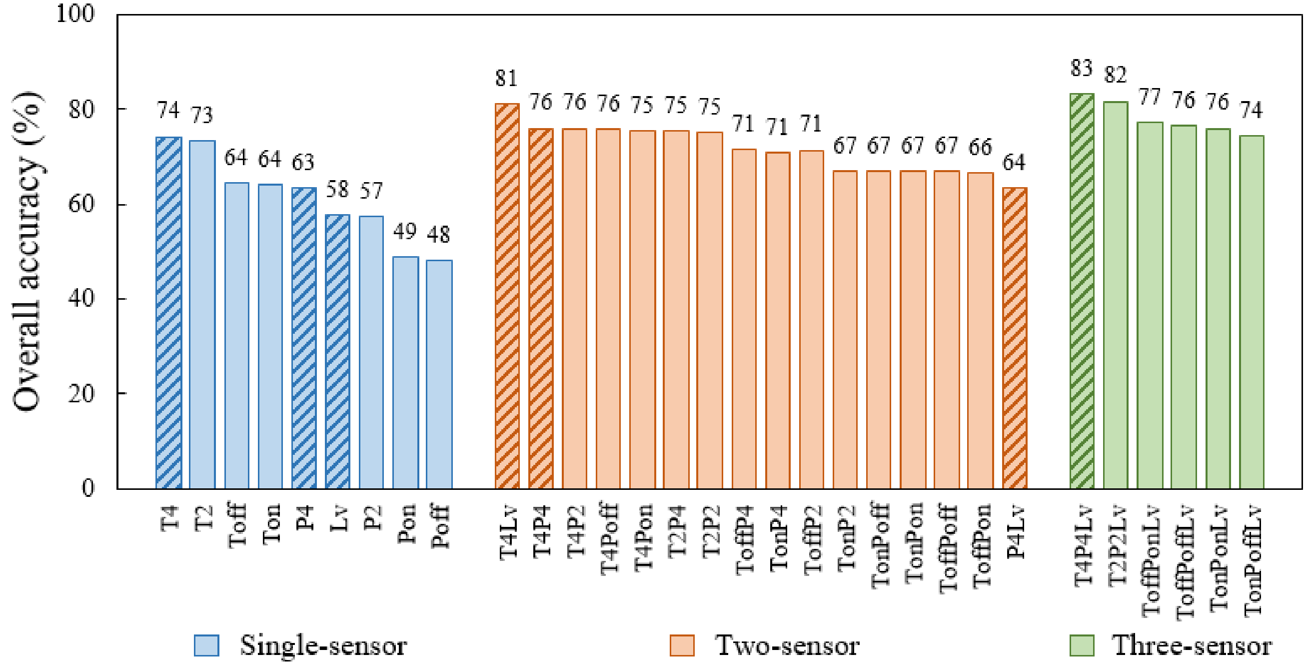


Fig.5

- Fig.5 reports overall accuracy of all classifications performed in this study.

- Performance of three-sensor classifications was relatively close with overall accuracy varying from 74% to 83% ($T_4P_4L_v$).
- More substantial differences in accuracy were associated with certain classes.

Error matrix of the T_4P_4 classification. Overall accuracy is 76% with a standard error of 2%.

Class	Water	Developed	Barren	Deciduous	Evergreen	Shrubland	Herbaceous/ Planted	Total	UAc (SE)
Water	**3.734**	0.000	0.000	0.000	0.000	0.000	0.000	3.734	100 (0)
Developed	0.053	**21.361**	0.372	1.292	0.000	0.036	4.249	27.363	78 (4)
Barren	0.000	1.064	**0.308**	0.000	0.000	0.002	0.778	2.153	14 (5)
Deciduous	0.043	0.311	0.000	**19.362**	0.100	0.201	0.036	20.052	97 (1)
Evergreen	0.387	0.418	0.000	0.998	**3.230**	0.093	0.107	5.234	62 (7)
Shrubland	0.107	2.381	0.000	8.002	0.057	**6.374**	2.431	19.353	33 (4)
Herbaceous/Planted	0.000	0.036	0.000	0.275	0.000	0.179	**21.622**	22.111	98 (1)
Total	4.324	25.572	0.680	29.929	3.387	6.884	29.224	100	
PAc (SE)	86 (3)	84 (4)	45 (22)	65 (4)	95 (2)	93 (2)	74 (4)		

Error matrix of the $T_4P_4L_v$ classification. Overall accuracy is 83% with a standard error of 2%.

Class	Water	Developed	Barren	Deciduous	Evergreen	Shrubland	Herbaceous/ Planted	Total	UAc (SE)
Water	**3.734**	0.000	0.000	0.000	0.000	0.000	0.000	3.734	100 (0)
Developed	0.021	**21.810**	0.336	1.292	0.000	0.036	4.985	28.481	77 (4)
Barren	0.000	1.064	**0.308**	0.332	0.000	0.002	0.556	2.263	14 (5)
Deciduous	0.021	0.729	0.000	**27.768**	0.000	0.468	0.021	29.007	96 (2)
Evergreen	0.468	0.693	0.000	1.579	**3.406**	0.043	0.107	6.297	54 (6)
Shrubland	0.021	0.439	0.000	0.826	0.036	**5.645**	2.345	9.313	61 (6)
Herbaceous/Planted	0.000	0.222	0.000	0.000	0.000	0.036	**20.648**	20.906	99 (1)
Total	4.266	24.958	0.644	31.797	3.442	6.230	28.663	100	
PAc (SE)	88 (3)	87 (4)	48 (25)	87 (2)	99 (1)	91 (5)	72 (4)		

- Vertical structural features extracted from LiDAR waveforms played a crucial role in distinguishing shrubland from deciduous forest.

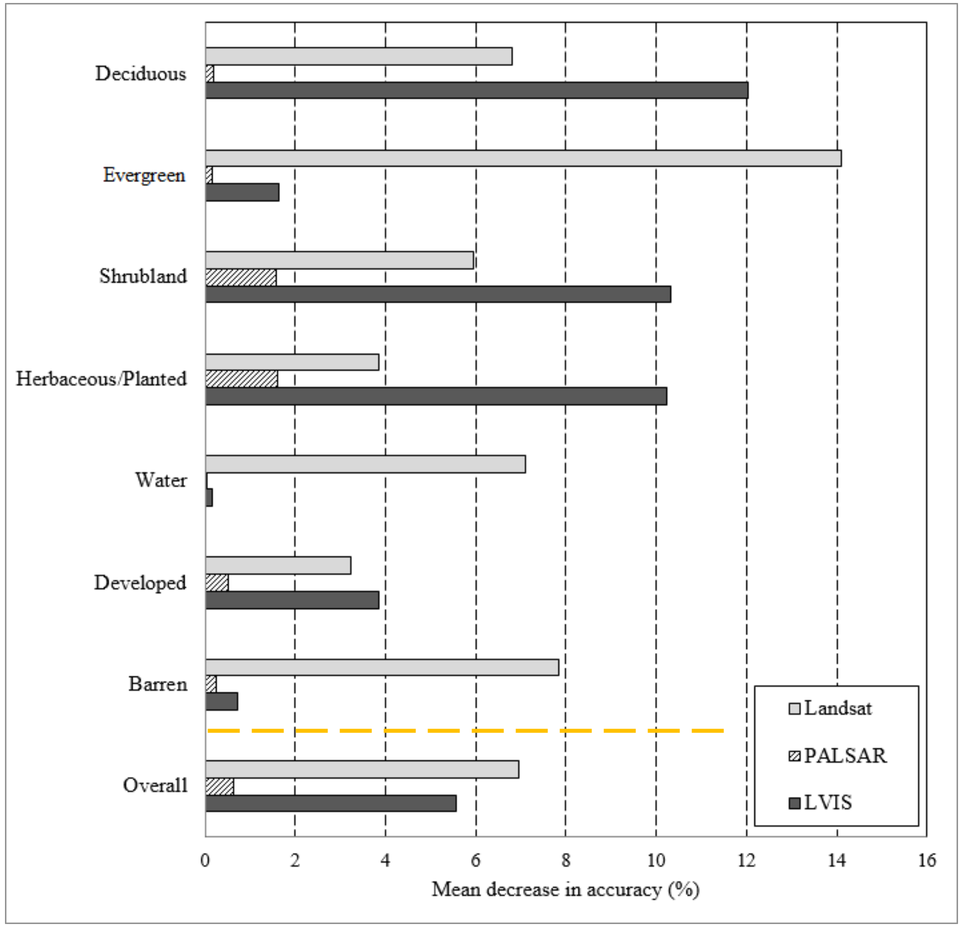


Fig.8. Averaged feature importance for each land cover class and over all classes.

- The Landsat feature set was of greater importance to the overall accuracy.
- Both Landsat and LVIS contributed considerably more than the numerous PALSAR-derived features.
- LVIS was the most indispensable feature set for three out of the four vegetation classes assessed in this study: deciduous, shrubland and herbaceous/planted.
- The integration of the PALSAR feature set was more beneficial to shrubland and herbaceous/planted over the other classes.

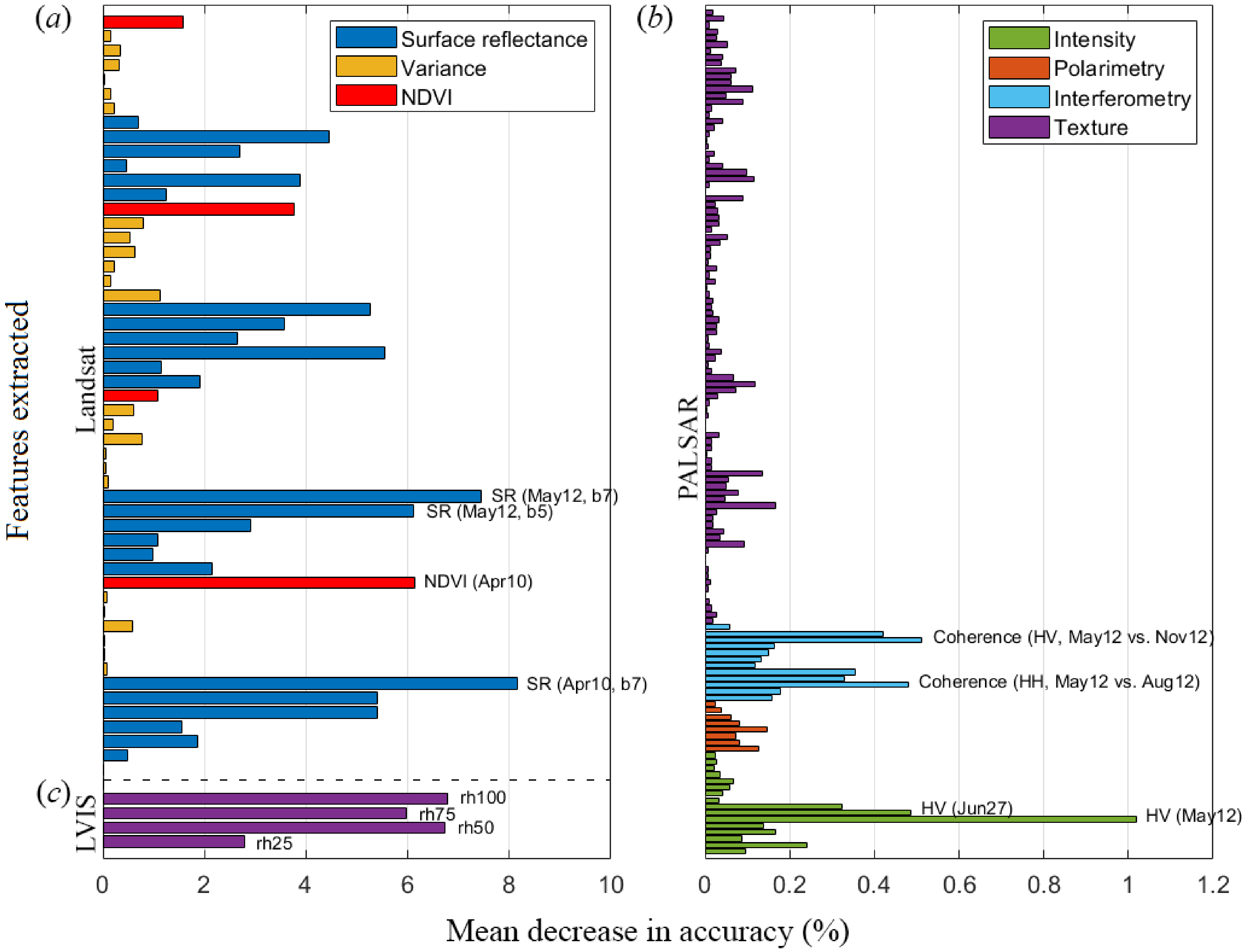
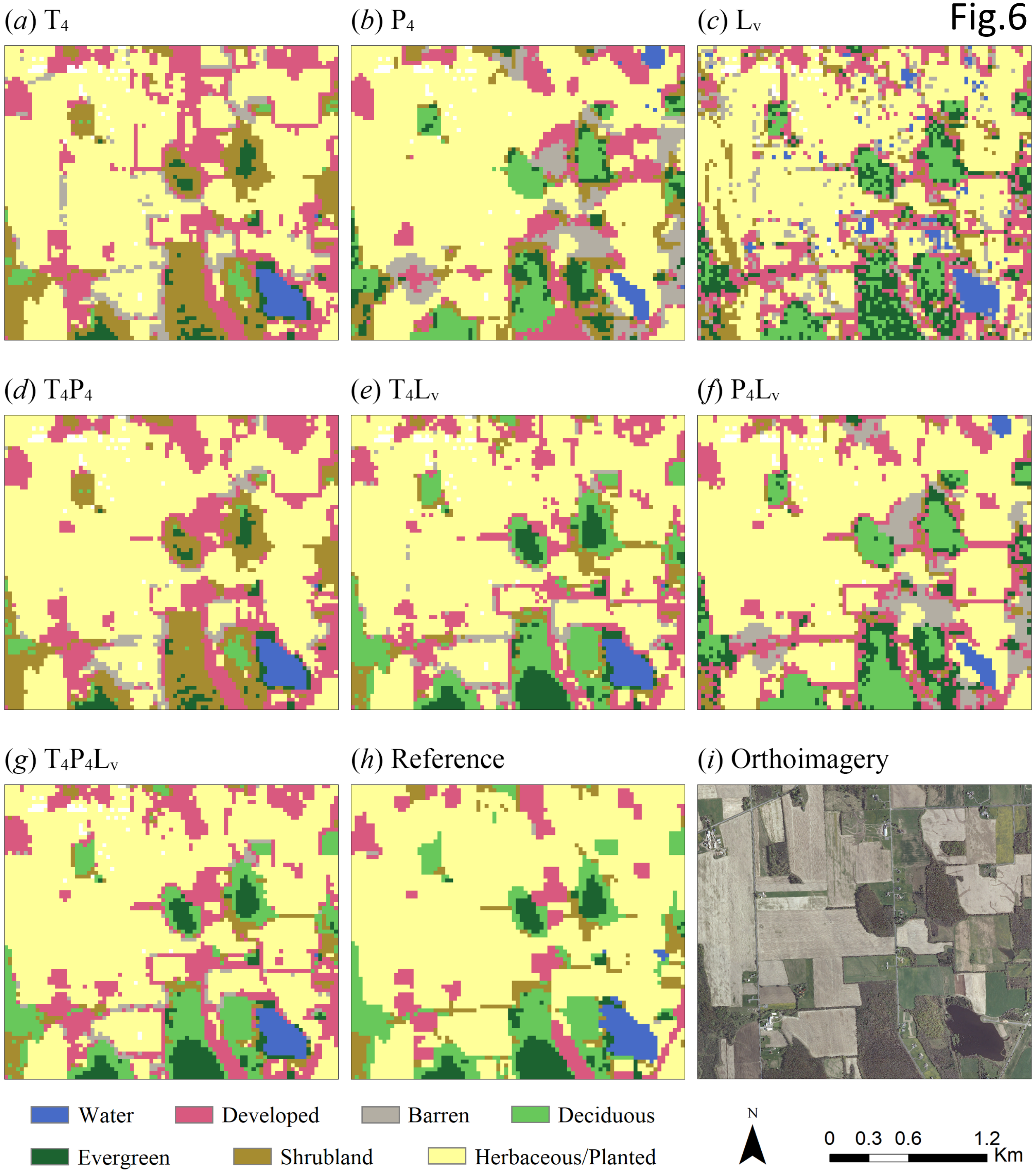

(a)
(b)
(c)
Surface reflectance
Variance
NDVI
Intensity
Polarimetry
Interferometry
Texture
Features extracted
Landsat
PALSAR
LVIS
SR (May12, b7)
SR (May12, b5)
NDVI (Apr10)
SR (Apr10, b7)
rh100
rh75
rh50
rh25
Coherence (HV, May12 vs. Nov12)
Coherence (HH, May12 vs. Aug12)
HV (Jun27)
HV (May12)
0
2
4
6
8
10
0
0.2
0.4
0.6
0.8
1
1.2
Mean decrease in accuracy (%)

Fig.6



- Fig.6 presents classifications over a small aera consisting of 85 by 75 30-m pixels.

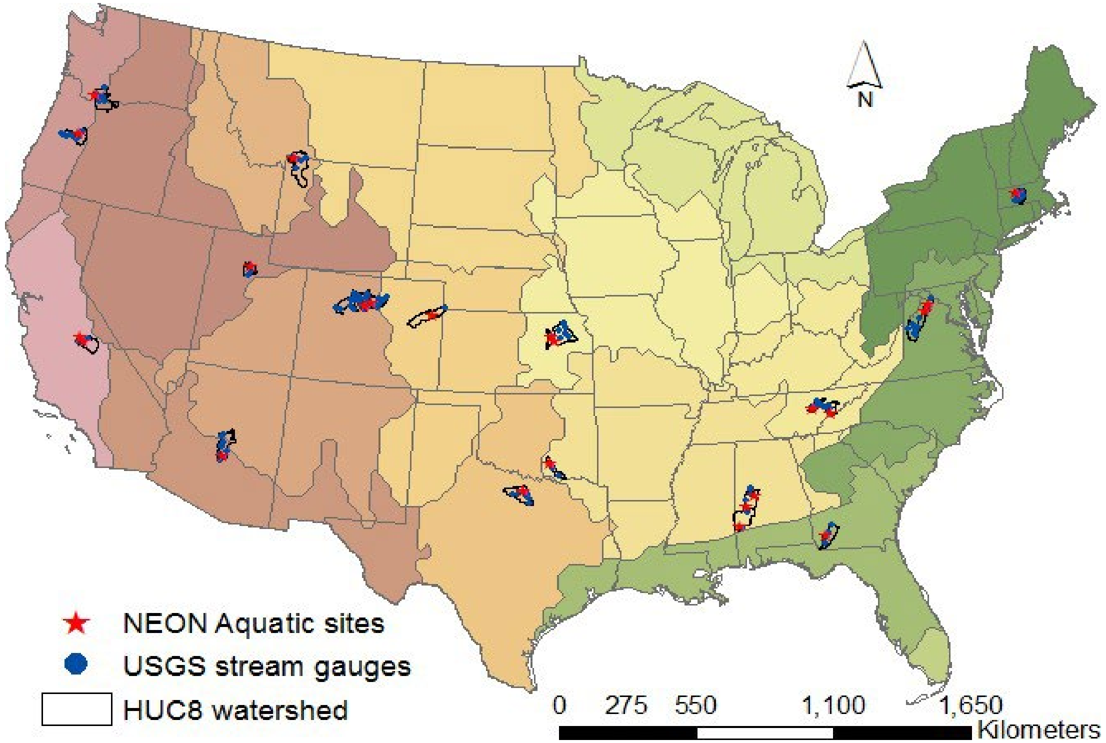
- Substantial differences can be observed when comparing (g), the resultant map of $T_4P_4L_v$ classification, to the other six maps derived from single- and two-sensor classifications (a-f).
- (g) appears more consistent with the real land cover patterns as displayed in (h) and (i), and artifacts and isolated pixels that are most likely to be noise were largely removed.

Takeaway messages

- Integration of spaceborne optical (Landsat-5/TM) and SAR (ALOS-1/PALSAR) and airborne full-waveform LiDAR (LVIS) data provides the most pronounced discriminative power among different land cover types.
- An overall accuracy of 83% was achieved when all spectral, scattering and vertical structural features were used as classification inputs (2–19% and 9–25% higher than the numerous two-sensor and single-sensor scenarios assessed, respectively).
- The inclusion of seasonal spectral and scattering variations of vegetation classes also improved overall accuracy by nearly 10%, when observations from all three sensors were used in combination.
- LVIS-derived canopy height metrics contributed more useful thematic information that is complementary to Landsat data and beneficial to classification performance.

What's next…

- We are extending our current proof-of-concept work to additional sites, sensors and time periods with the use of other machine learning as well as advanced deep learning algorithms that have gained a lot of interest recently to ensure more generalized and statistically representative conclusions.



FIELD SITE ID	AOP (NIS, LIDAR)	LANDSAT (L8)	NLCD CLASSES
ARIK* (CO)	2017-05-22,05-25 2020-06-17,06-20 2021-06-04,06-05	2017/06/10 2020/05/17,07/20 2021/06/05	Emergent Herbaceous Wetlands| Grassland/Herbaceous| Woody Wetlands
BLUE (OK)	2017-05-06	2017/05/06	Grassland/Herbaceous| Pasture/Hay
HOPB (MA)	2016-08-18 2017-08-23 2019-08-26	2016/08/05 2017/08/24 2019/08/30	Evergreen Forest| Mixed Forest
MCDI (KS)	2017-06-05 2020-07-13	2017/05/29 2020/06/22,08/25	Cultivated Crops| Grassland/Herbaceous
MCRA (OR)	2018-07-16 2021-07-22	2018/07/17 2021/07/25	Evergreen Forest
PRIN* (TX)	2016-04-24,04-27 2017-05-02 2021-06-25	2016/05/03 2017/05/06 2021/06/25	Deciduous Forest| Grassland/Herbaceous
REDB* (UT)	2017-05-31 2019-05-10 2021-05-20	2017/06/04 2019/06/10 2021/05/14,05/30	Deciduous Forest
SYCA (AZ)	2019-09-12 2021-04-07,04-09,04-10	2019/08/22 2021/04/05,04/14	Shrub/Scrub
WLOU (CO)	2017-09-13 2019-08-26 2020-07-29,08-07	n/a 2019/08/17 2020/08/03	Evergreen Forest
LIRO* (WI)	2017-09-11 2020-08-16	2017/09/13 2020/09/05	Deciduous Forest| Mixed Forest

* ARIK, PRIN, REDB & LIRO have lidar data acquired in 2022.

Thank You!