6th Asia-Oceania Group on Earth Observations (AOGEO) Workshop
May 29-31, 2023
Online& offline·Macau, China

Landsat analysis ready data: algorithms and application demonstrations with deep learning based land cover mapping

Hankui Zhang

Geospatial Sciences Center of Excellence, Department of Geography and Geospatial Sciences
South Dakota State University, Brookings, SD 57007, USA

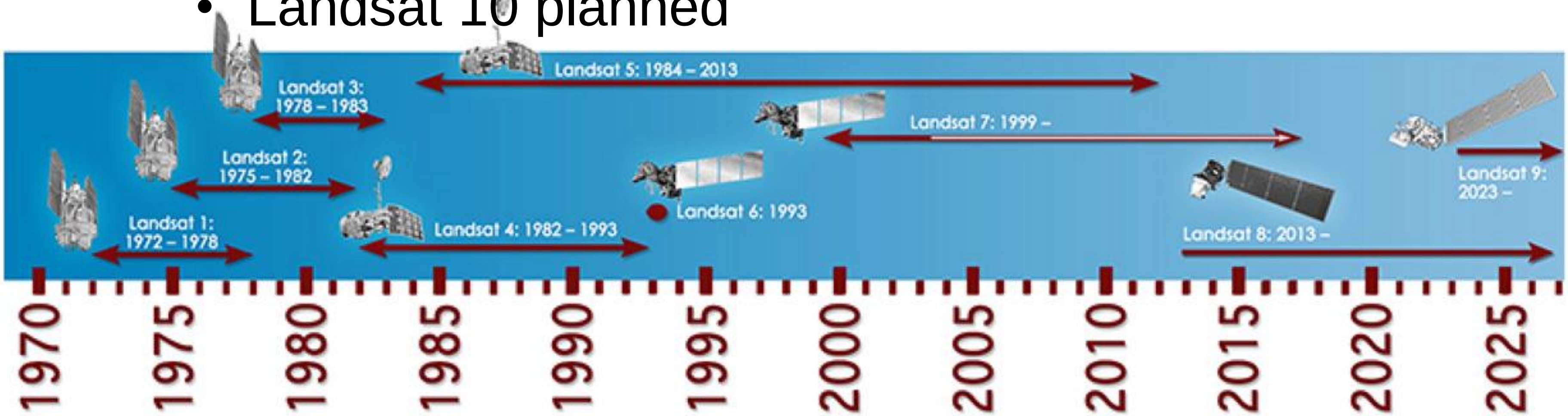
https://sites.google.com/view/hankui

Outline

- **What is analysis ready data (ARD)**
- **ARD algorithms**
- **ARD for deep learning based land cover mapping**
- **Summary and Outlook for Asia-Oceania**

NASA – USGS Landsat

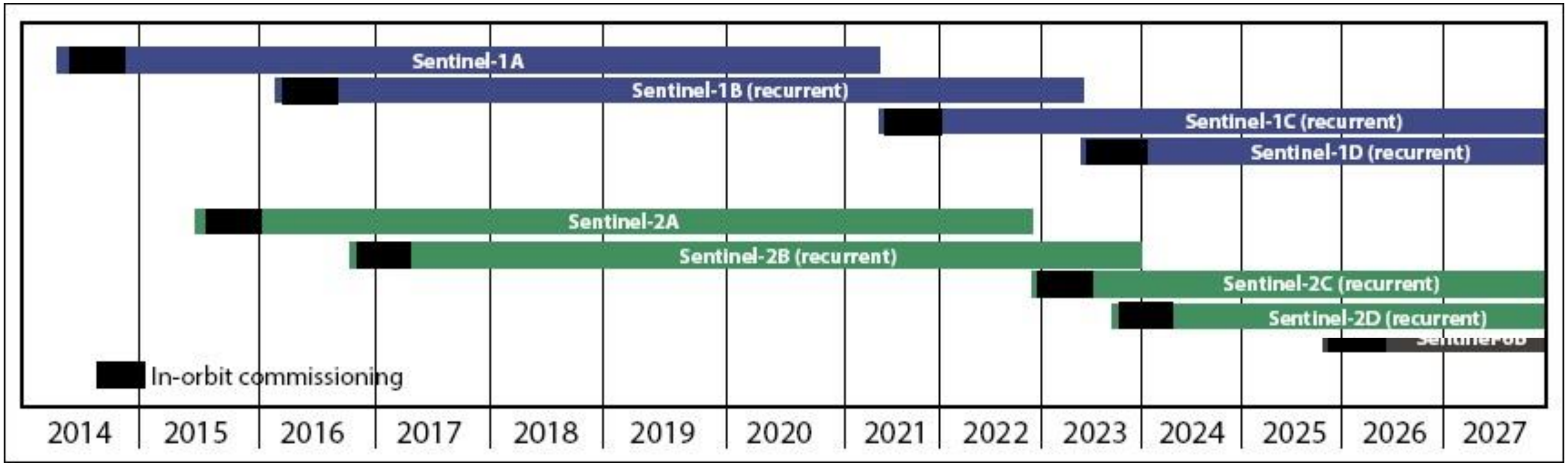
- 30 m spatial resolution global coverage
- 16-day revisit
- Since 1972
  - Landsat 7 just died April 2022
  - Landsat 8 and 9 (launched Sep. 2021) running
  - Landsat 10 planned



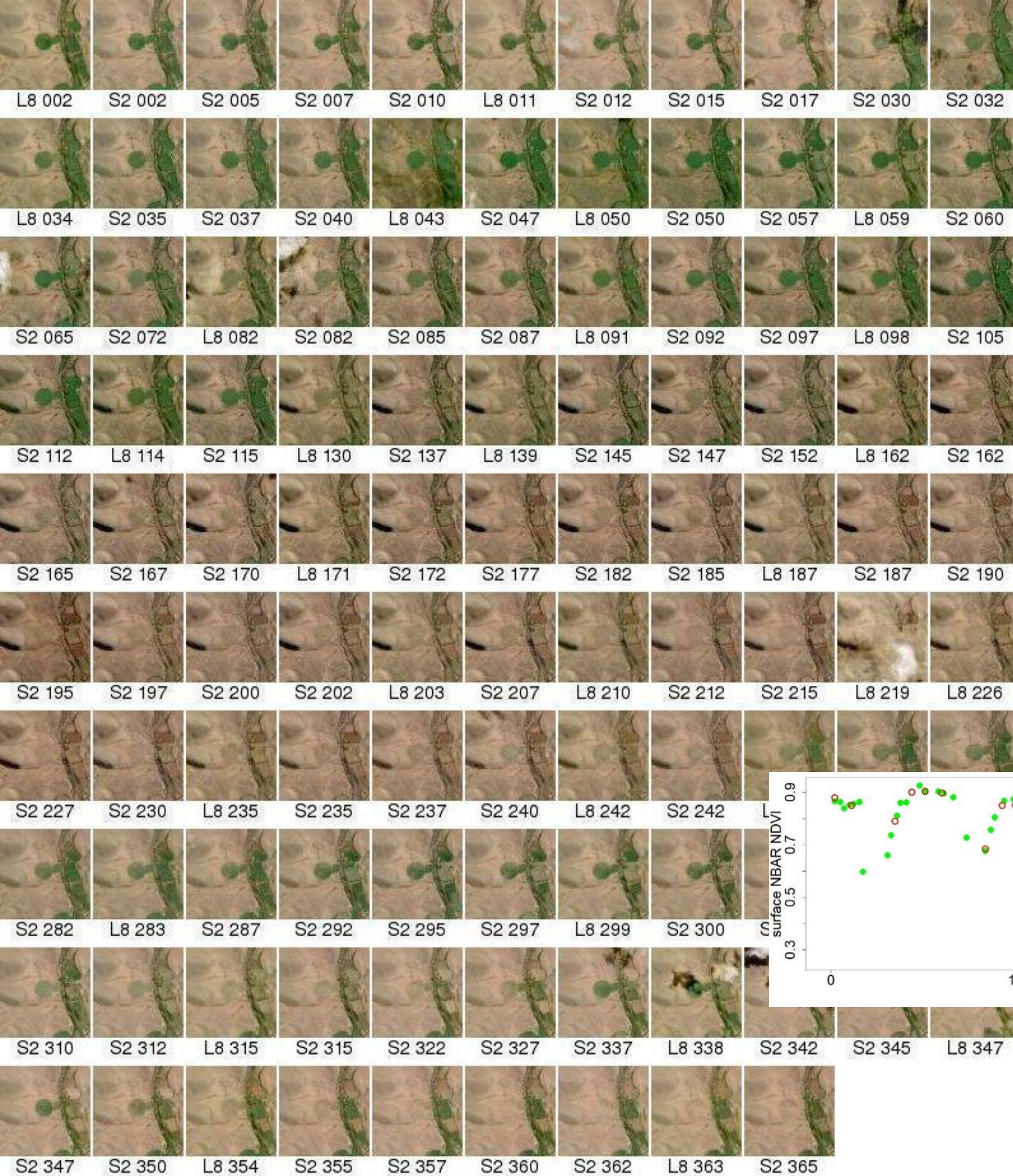
Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., ... & Roy, D. P. (2016). The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, *185*, 271-283.

ESA Sentinel-2

- 10-20 m spatial resolution global coverage
- 10-day revisit
- Since 2015
    - Sentinel-2A and B running
    - <u>Sentinel-2C and D planned</u>



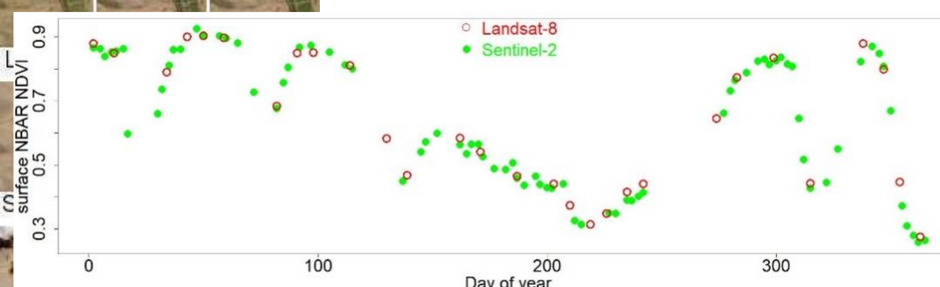
Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment*, *120*, 25-36.



entinel-2A,B evisit

A commercial crop field in South Africa

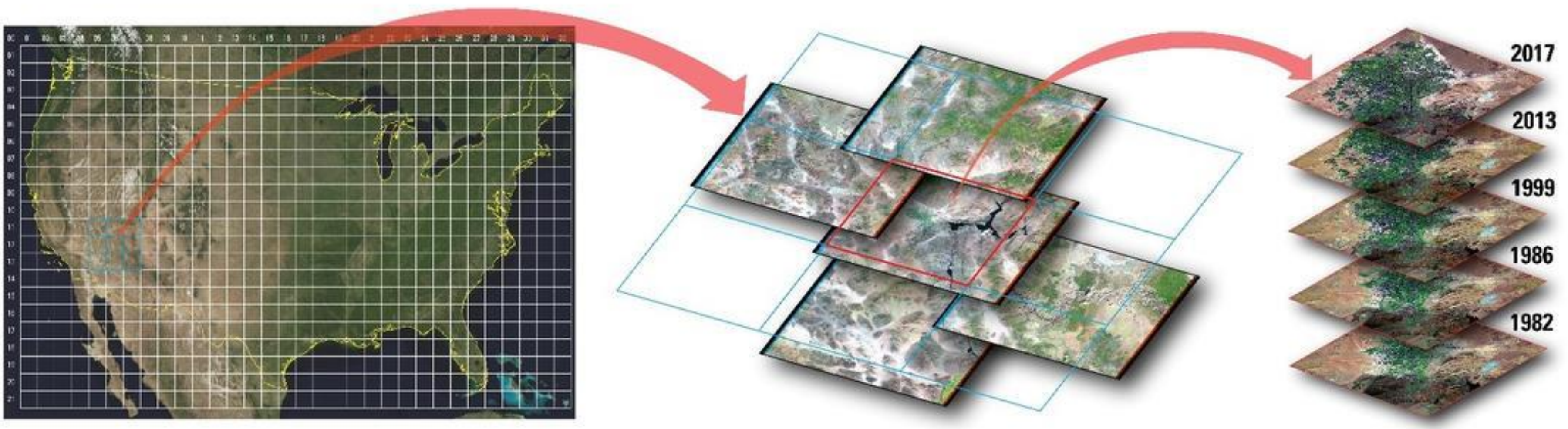
- 108 cloud-free images 2018
- 71 × 71 30 m pixels
- "L8" (Landsat-8)
- "S2" (Sentinel-2)



农田像素的NDVI

Analysis ready data

- Data that have been processed to allow analysis with a minimum of additional user effort are often referred to as Analysis Ready Data (ARD)



ARD ≥ processing level 3 (NASA)

Level	Description
Level 0	Reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artifacts (e.g., synchronization frames, communications headers, duplicate data) removed.
Level 1A	Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, <u>including radiometric and geometric calibration coefficients and georeferencing parameters (e.g., platform ephemeris)</u> computed and appended but not applied to Level 0 data.
Level 1B	Level 1A data that have been processed to sensor units (not all instruments have Level 1B source data).
Level 2	Derived geophysical variables at the same resolution and location as Level 1 source data.
Level 3	Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.
Level 4	Model output or results from analyses of lower-level data (e.g., variables derived from multiple measurements).

ARD algorithms should be

- Robust to spatial and temporal variation

- Automated

- Efficient to handle large volume data
  - 50 TB/year for Landsat 5 & 7
  - 370 TB/year for each of Landsat 8 and 9 (~1TB/day)
  - 511 TB/year for each of Sentinel-2A and 2B sensor

Operational Atmospheric Correction

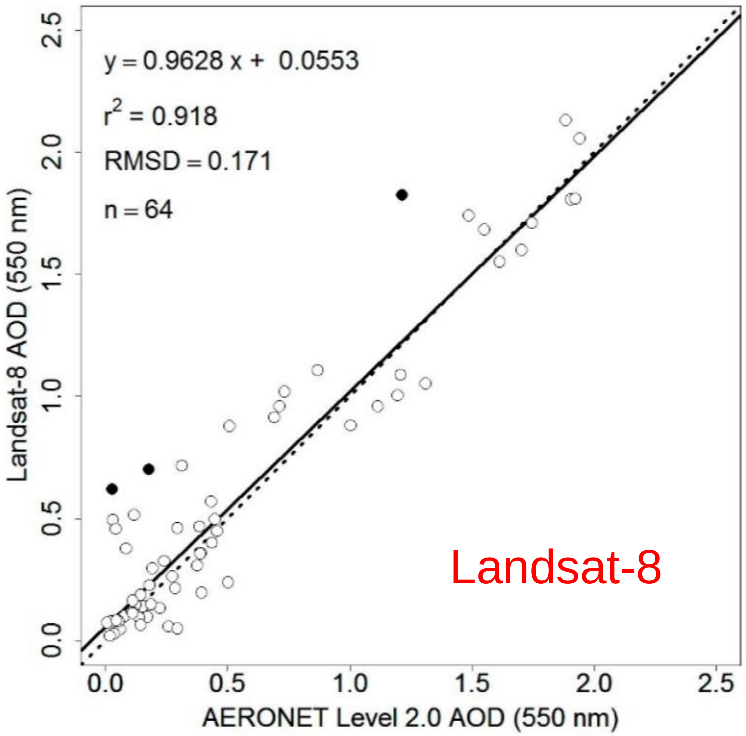
- Radiative transfer codes (Established)
  - 6S
  - MODTRAN
  - LibRadtran
- Characterize atmosphere (Difficult!)
  - Aerosol is most challenging
    - Dark object method
    - Surface reflectance library method
  - Reanalysis water vapor and ozone

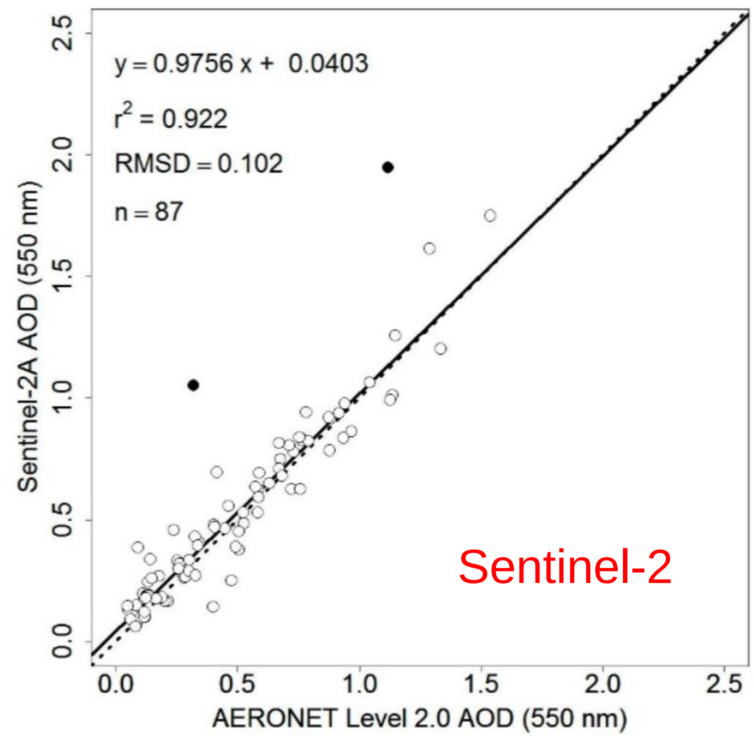
Public available software

- Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS)
  - Applied for Landsat-5/7
  - 6S
  - Dark object
- Land surface reflectance code (LaSRC)
  - Applied for Landsat-8 & Sentinel-2
  - 6S
  - Dark object + MODIS reflectance library (ratio)
- Sen2Cor
  - Applied for Sentinel-2
  - LibRadtran
  - Dark object

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., ... & Lim, T. K. (2006). A Landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters*, *3*(1), 68-72.
Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, *185*, 46-56.

LaSRC derived aerosol optical depth (AOD) accuracy





remote sensing

MDPI

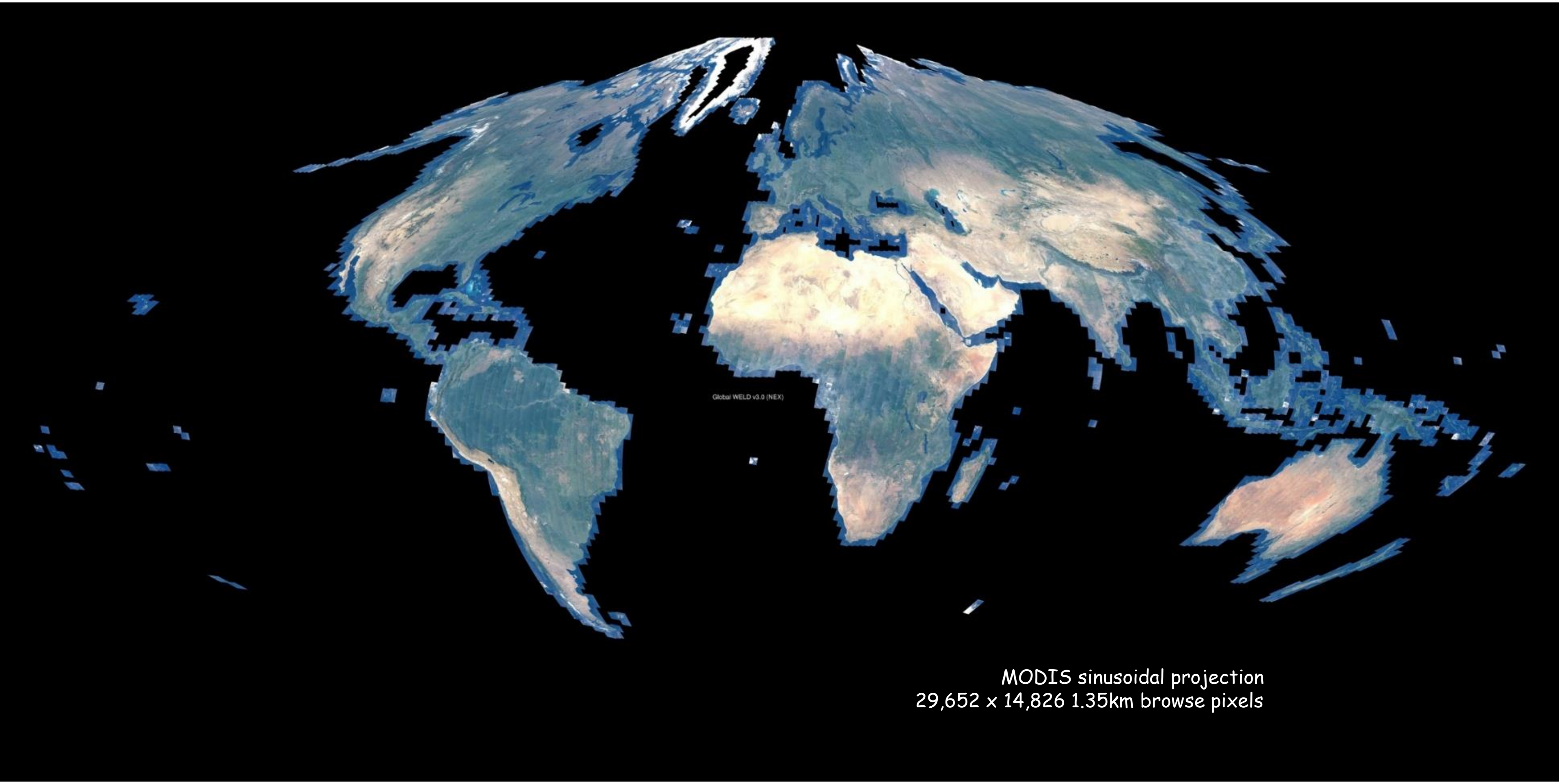
Article

Evaluation of Landsat-8 and Sentinel-2A Aerosol Optical Depth Retrievals across Chinese Cities and Implications for Medium Spatial Resolution Urban Aerosol Monitoring

Zhongbin Li [1], David P. Roy [2], Hankui K. Zhang [1,*], Eric F. Vermote [3] and Haiyan Huang [1]

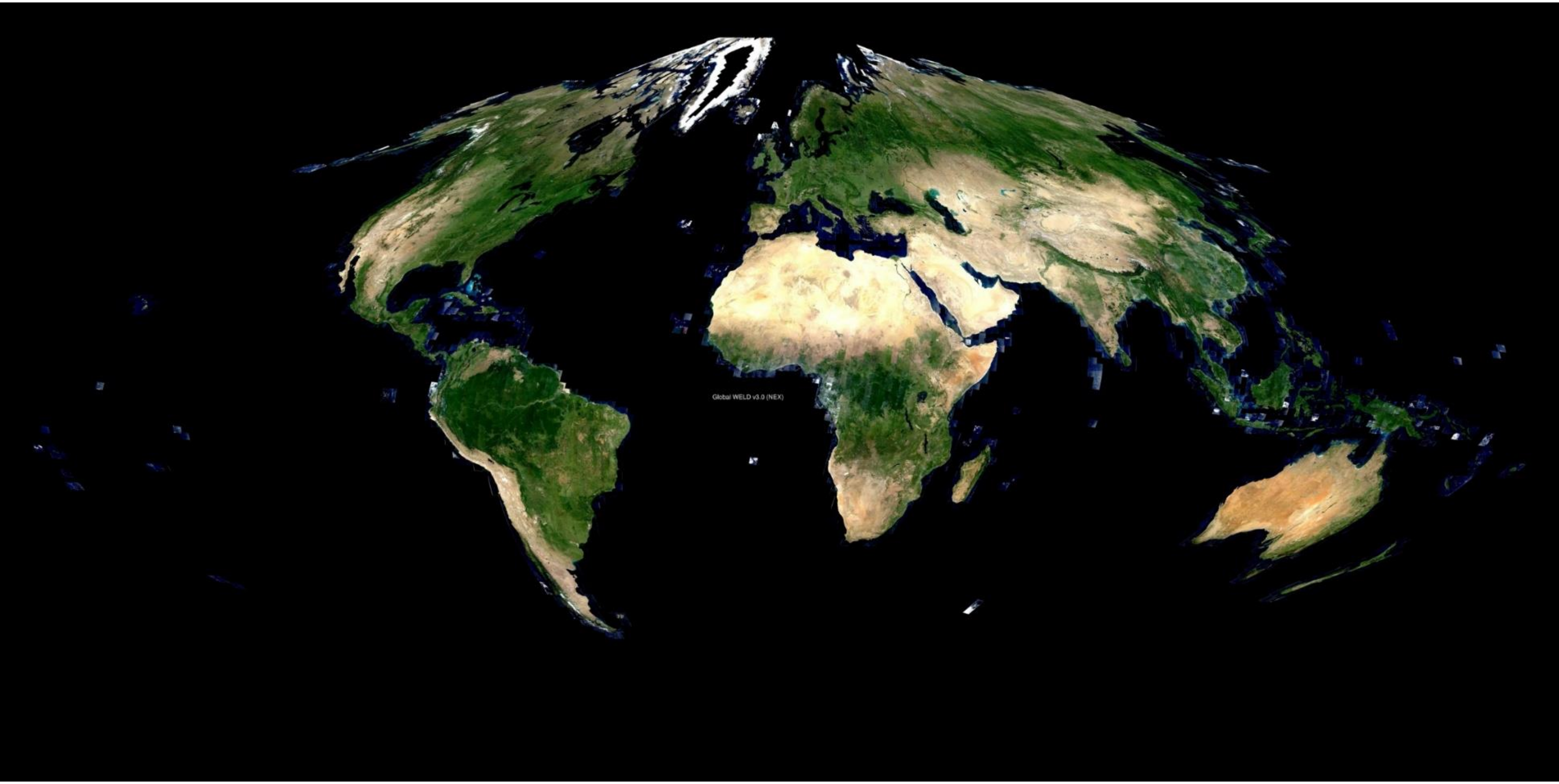
Global Landsat Annual 2010 30 m Top of Atmospheric reflectance

124,433 L1T scenes (45,711 Landsat 5 & 78,722 Landsat 7)



Global Landsat Annual 2010 30 m Surface reflectance

124,433 L1T scenes (45,711 Landsat 5 & 78,722 Landsat 7)



BRDF correction for Landsat & Sentinel-2

1410 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 54, NO. 3, MARCH 2016

Optimal Solar Geometry Definition for Global Long-Term Landsat Time-Series Bidirectional Reflectance Normalization

Hankui K. Zhang, David P. Roy, and Valeriy Kovalskyy

Contents lists available at ScienceDirect
ELSEVIER
Remote Sensing of Environment
journal homepage: www.elsevier.com/locate/rse

A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance

CrossMark

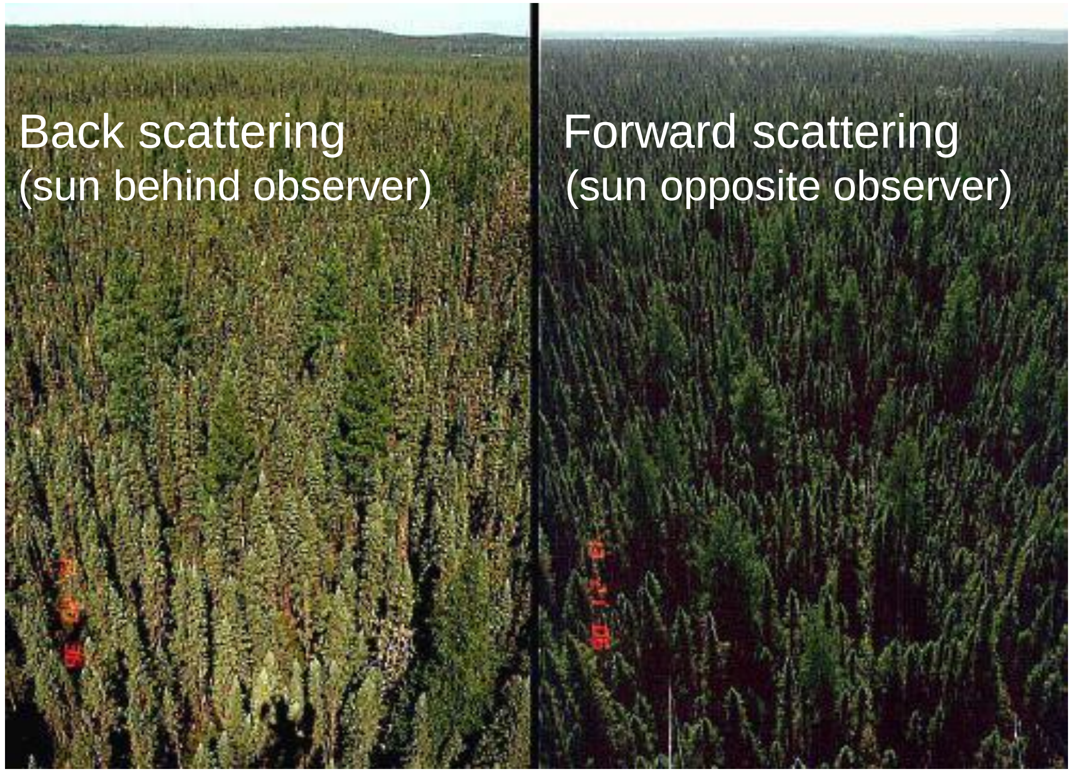
D.P. Roy [a,*], H.K. Zhang [a], J. Ju [b,c], J.L. Gomez-Dans [d,e], P.E. Lewis [d,e], C.B. Schaaf [f], Q. Sun [f], J. Li [a], H. Huang [a], V. Kovalskyy [a]

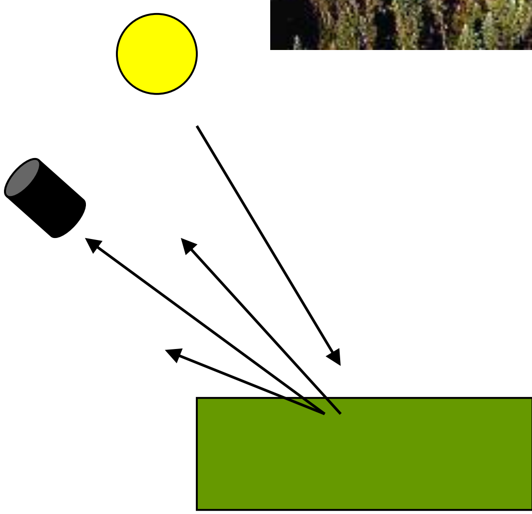
ELSEVIER
Remote Sensing of Environment
journal homepage: www.elsevier.com/locate/rse

Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences

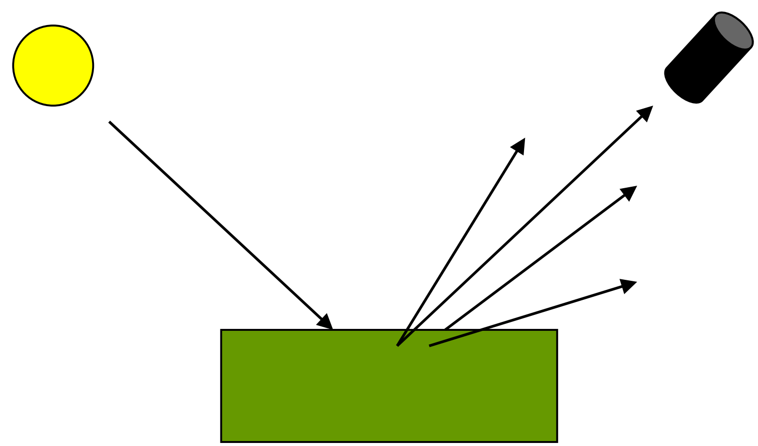
Check for updates

Hankui K. Zhang [a,*], David P. Roy [a], Lin Yan [a], Zhongbin Li [a], Haiyan Huang [a], Eric Vermote [b], Sergii Skakun [b,c], Jean-Claude Roger [b,c]



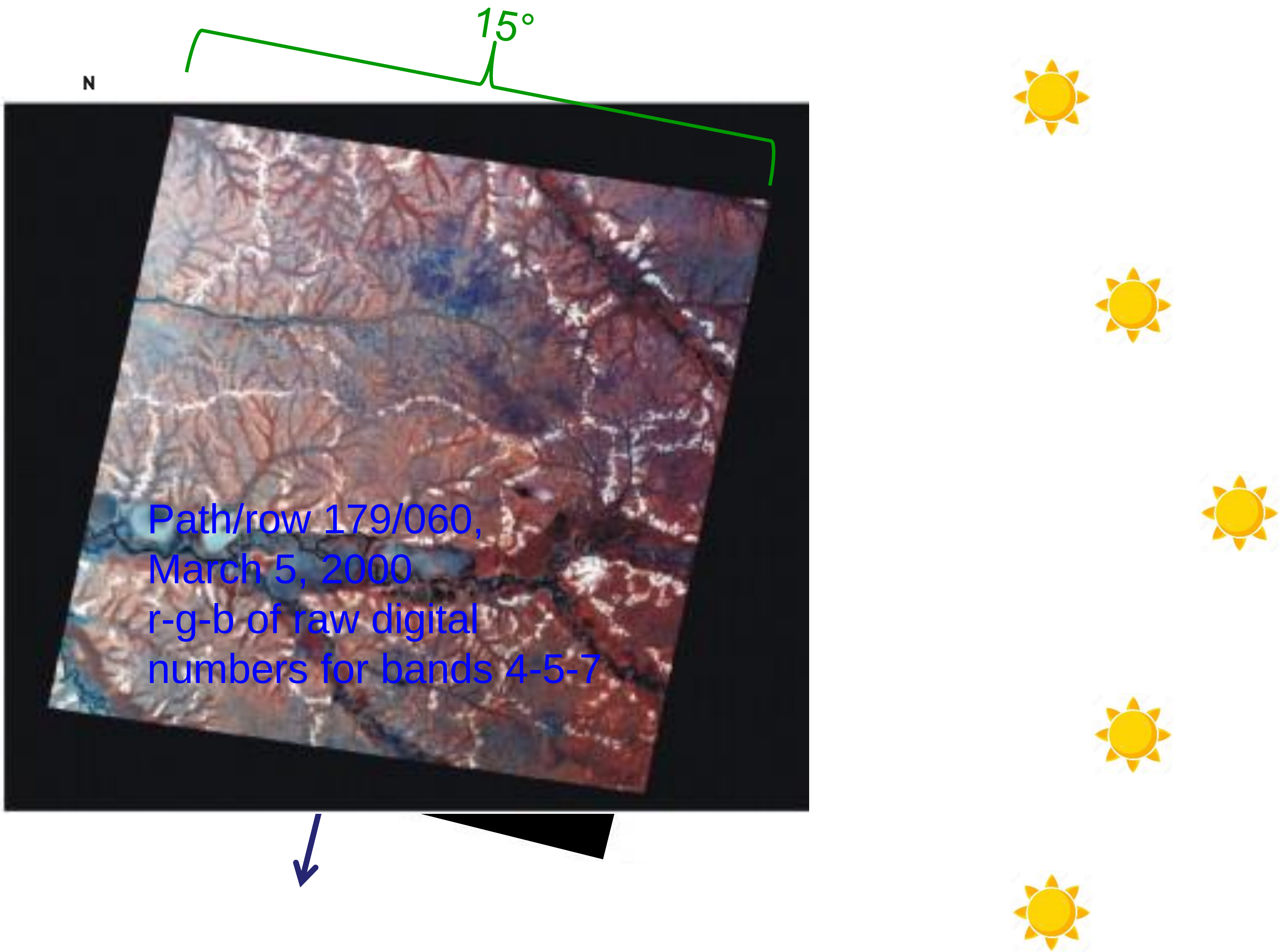


Back scatter direction



Forward scatter direction

Ground projection of Landsat geometry

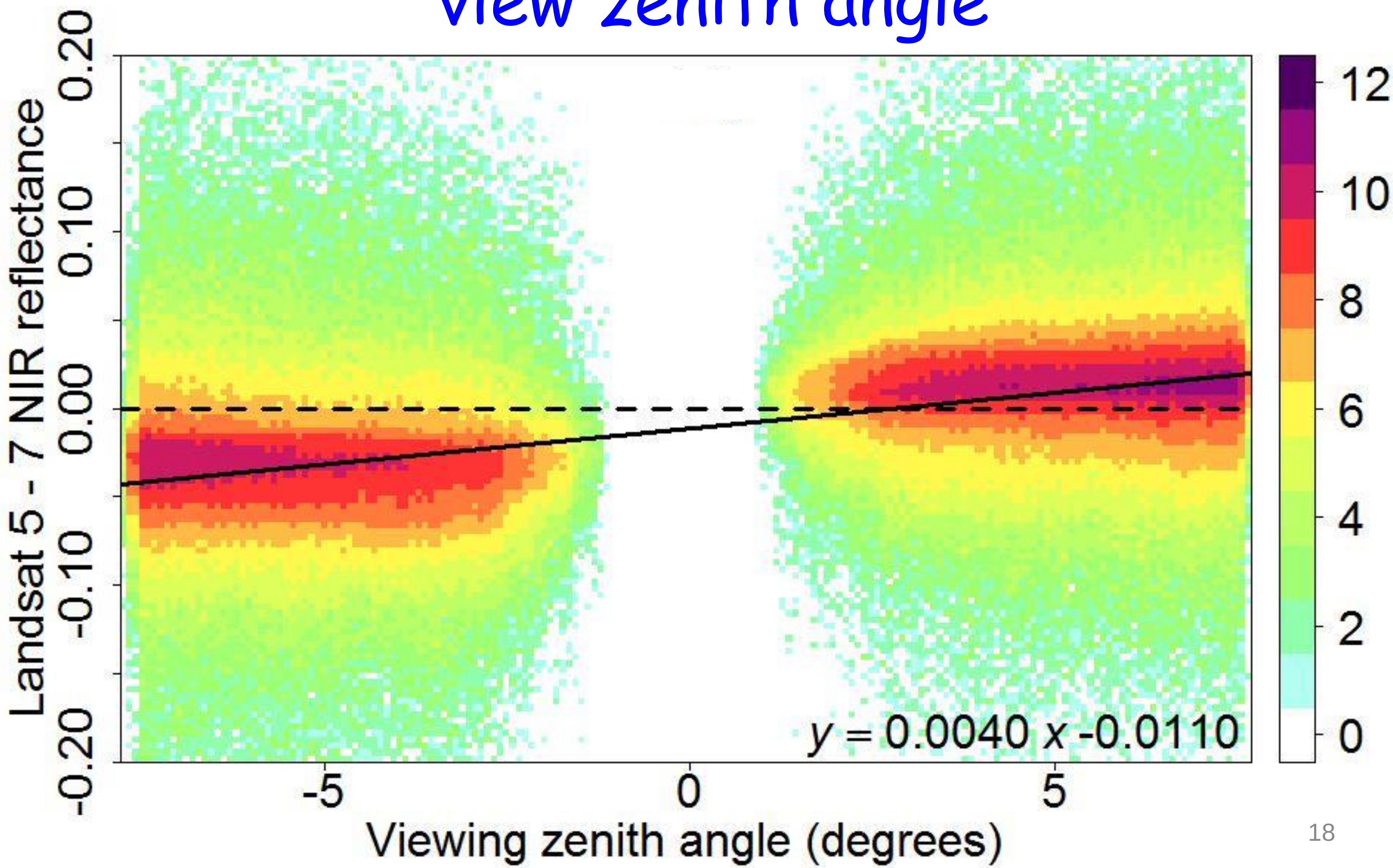


A general method to derive Landsat nadir BRDF adjusted reflectance (NBAR)

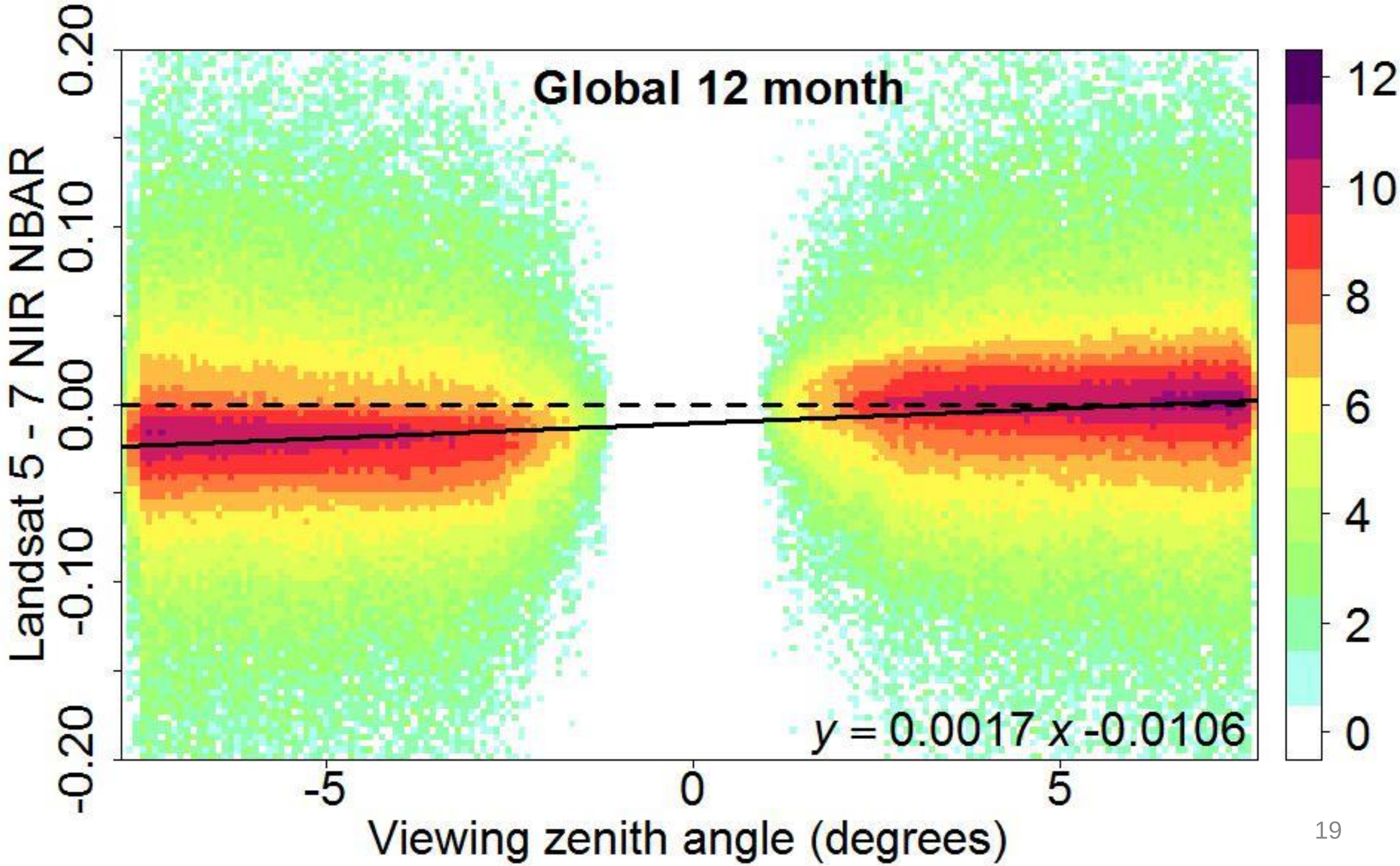
- BRDF shapes of all different cover types are very similar
  - over the narrow 15° Landsat field of view

- Using a single set of global average MODIS BRDF parameters
  - normalize Landsat pixels acquired at any time and location
  - 2010 global MODIS BRDF mean parameters

(Landsat 5 TM - Landsat 7 ETM+) observed surface reflectance versus Landsat 5 TM view zenith angle



(Landsat 5 TM - Landsat 7 ETM+) NBAR derived by the fixed mean global 12 month



Has been tested effective on Sentinel-2 (20.6° field of view)

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 57, NO. 6, JUNE 2019 3591

Investigation of Sentinel-2 Bidirectional Reflectance Hot-Spot Sensing Conditions

Zhongbin Li, Hankui K. Zhang, and David P. Roy

ELSEVIER

Remote Sensing of Environment

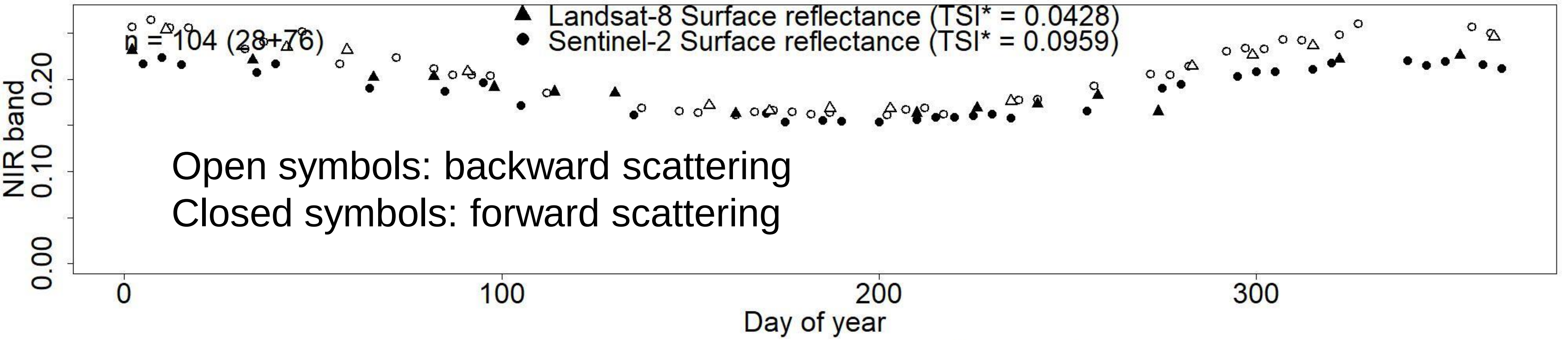
journal homepage: www.elsevier.com/locate/rse

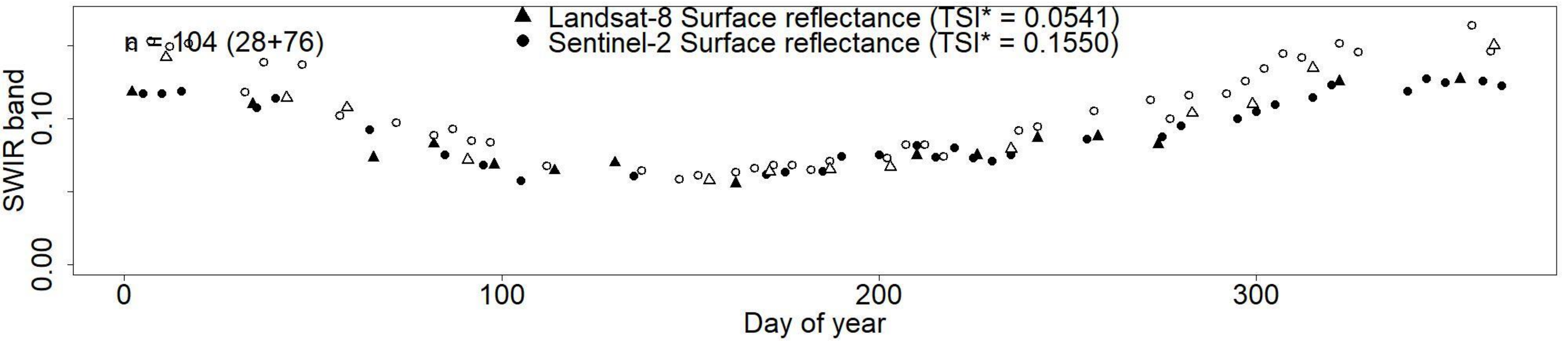
Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance

CrossMark

David P. Roy *, Jian Li, Hankui K. Zhang, Lin Yan, Haiyan Huang, Zhongbin Li

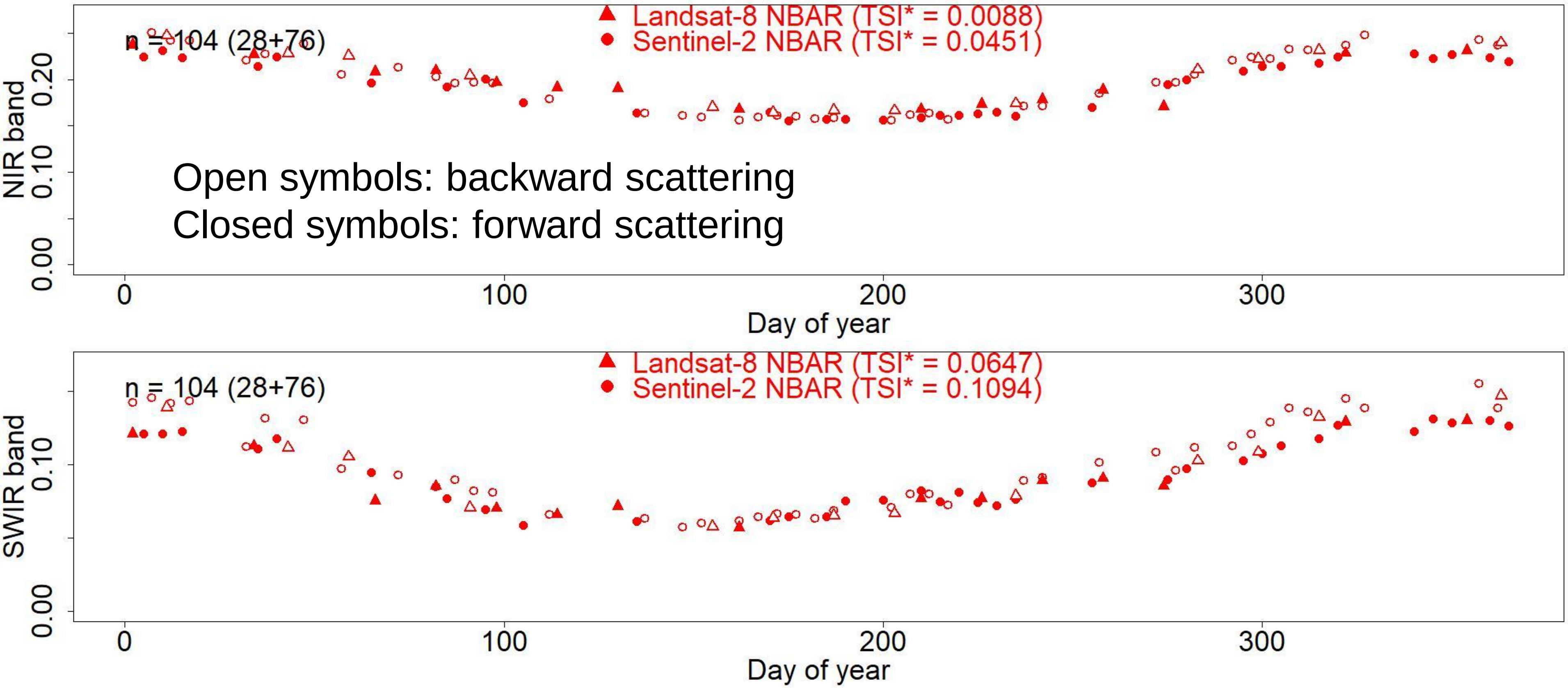
Example Landsat-8 and Sentinel-2 reflectance time series





- Backward reflectance is greater than forward reflectance

Example Landsat-8 and Sentinel-2 NBAR time series



- Nadir BRDF-adjusted reflectance more consistent

BRDF due to orbit drift induced solar angle variation

ELSEVIER

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Landsat 5 Thematic Mapper reflectance and NDVI 27-year time series inconsistencies due to satellite orbit change

CrossMark

H.K. Zhang *, D.P. Roy

Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007, USA

ELSEVIER

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

A conterminous United States analysis of the impact of Landsat 5 orbit drift on the temporal consistency of Landsat 5 Thematic Mapper data

Check for updates

David P. Roy[a,*], Zhongbin Li[a], Hankui K. Zhang[b], Haiyan Huang[a]

USGS Landsat ARD

- Tiled in Albers projection
- Covers only US
- Surface reflectance (LEDAPS & LaSRC)
- Surface temperature (MODTRAN)
- Per-pixel solar and viewing geometry
- Processed and stored in Amazon Cloud
  - User download interface unchanged

Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., **Zhang, H. K.,** & Lymburner, L. (2018). Analysis ready data: enabling analysis of the Landsat archive. *Remote Sensing*, *10*(9), 1363.

USGS Landsat Collection-2 data

- Global Landsat archive processed December 2020
- Surface reflectance (LEDAPS & LaSRC)
- Surface temperature (MODTRAN)
- Per-pixel solar and viewing geometry
- Processed and stored in Amazon Cloud
  - User download interface unchanged
- Not tiled: UTM projection same as Collection-1
  - Not real ARD

USGS Landsat Collection-3 data is planned

- Should correct for BRDF both viewing angle and orbit drift
- Landsat 4, 5, 7 and 8 consistency

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

ELSEVIER

Making Landsat 5, 7 and 8 reflectance consistent using MODIS nadir-BRDF adjusted reflectance as reference

Xianghong Che[a], Hankui K. Zhang[b,*], Jiping Liu[a]

[a] Research Center of Geospatial Big Data Application, Chinese Academy of Surveying and Mapping, 28 Lianhuachi West Road, Beijing 100830, China
[b] Geospatial Sciences Center of Excellence, Department of Geography and Geospatial Sciences, South Dakota State University, Brookings, SD 57007, USA

Landsat ARD application: 30 m land cover mapping using CNN

- Zhang, H.K., Roy, D.P., Luo, D., Large area, single pixel time series, convolutional neural network land cover classification, *Remote Sensing of Environment*, Accepted for publication.

Remote Sensing of Environment 197 (2017) 15–34

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

ELSEVIER

Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification

CrossMark

Hankui K. Zhang *, David P. Roy

Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007, USA

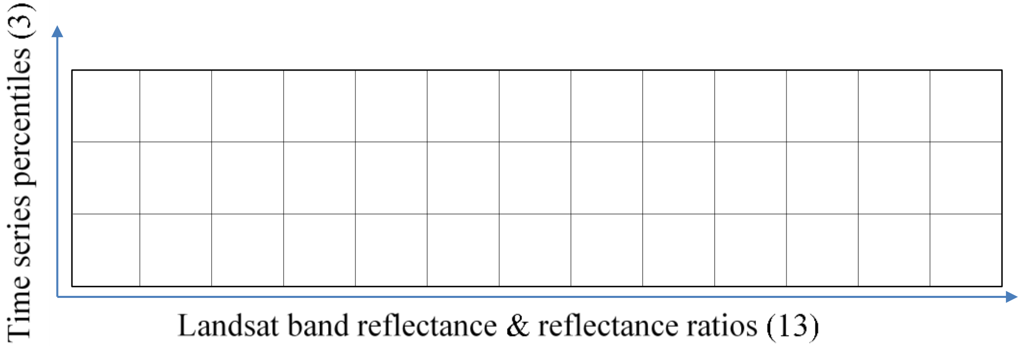
>3.3 million training & evaluation samples from NLCD

- Extracted from CONUS NLCD 2011
  - Purity filtering
  - Systematically sampling
- 15 classes of the 16 over CONUS
  - perennial ice/snow class not included
- A total of 3,314,439 samples

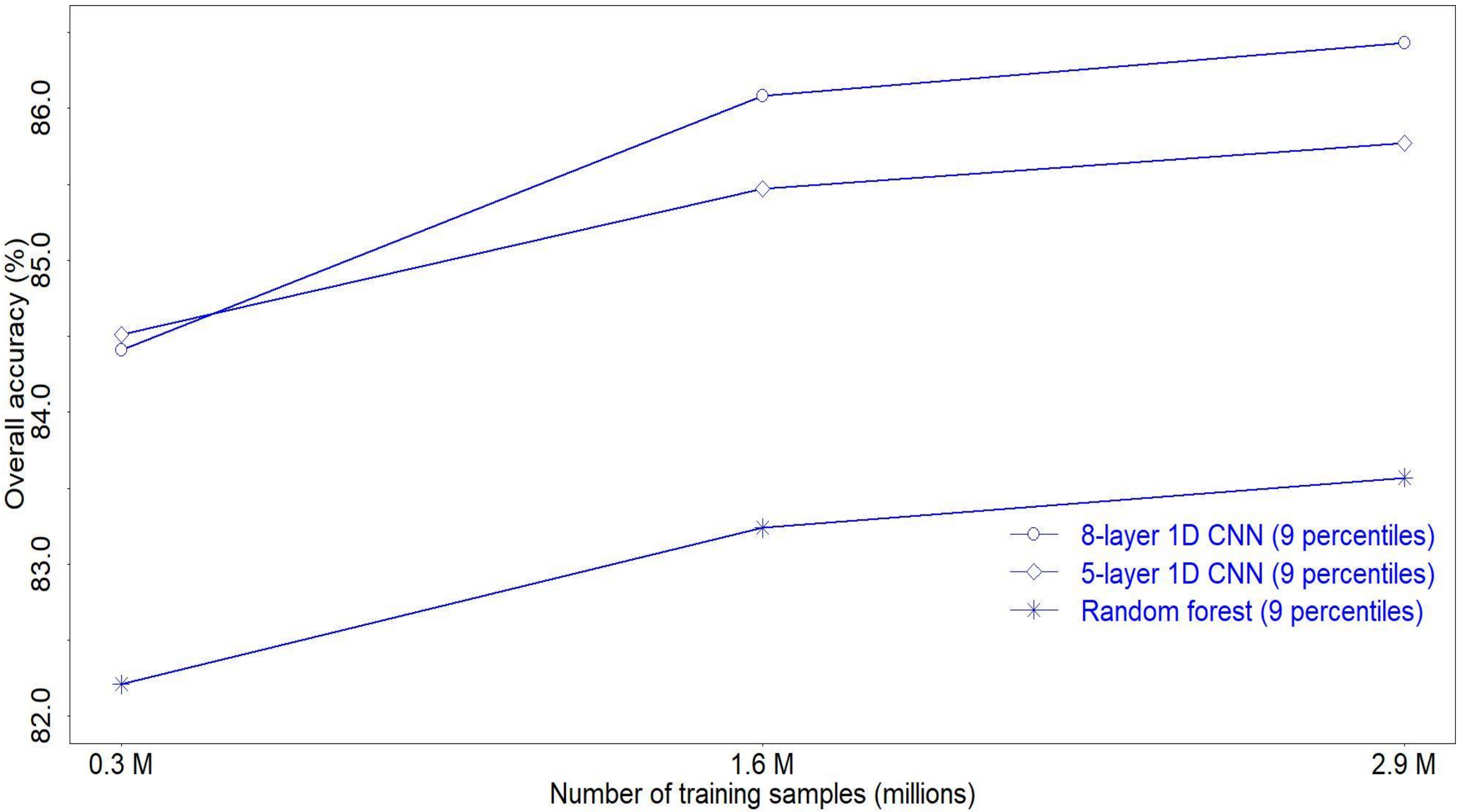
ID	Land cover class	Number of 30m samples	ID	Land cover class	Number of samples
11	Open water	255,725	43	Mixed forest	18,391
21	Developed open space	11,288	52	Shrub/scrub	823,707
22	Developed low intensity	3,427	71	Grassland/herbaceous	493,638
23	Developed medium intensity	2,003	81	Pasture/hay	187,039
24	Developed high intensity	2,135	82	Cultivated crops	644,661
31	Barren land	36,562	90	Woody wetlands	103,608
41	Deciduous forest	321,740	95	Emergent herbaceous wetlands	30,714
42	Evergreen forest	379,801			

CNN applied to single pixel 2D array variables

- A 2D array of a single pixel variables
  - Spectral dimension: 13 = 5 Landsat bands (non-blue) & 8 band ratios
  - Temporal dimension: 3 = 20, 50 and 80 percentiles of time series observations
  - No spatial dimension – a single pixel

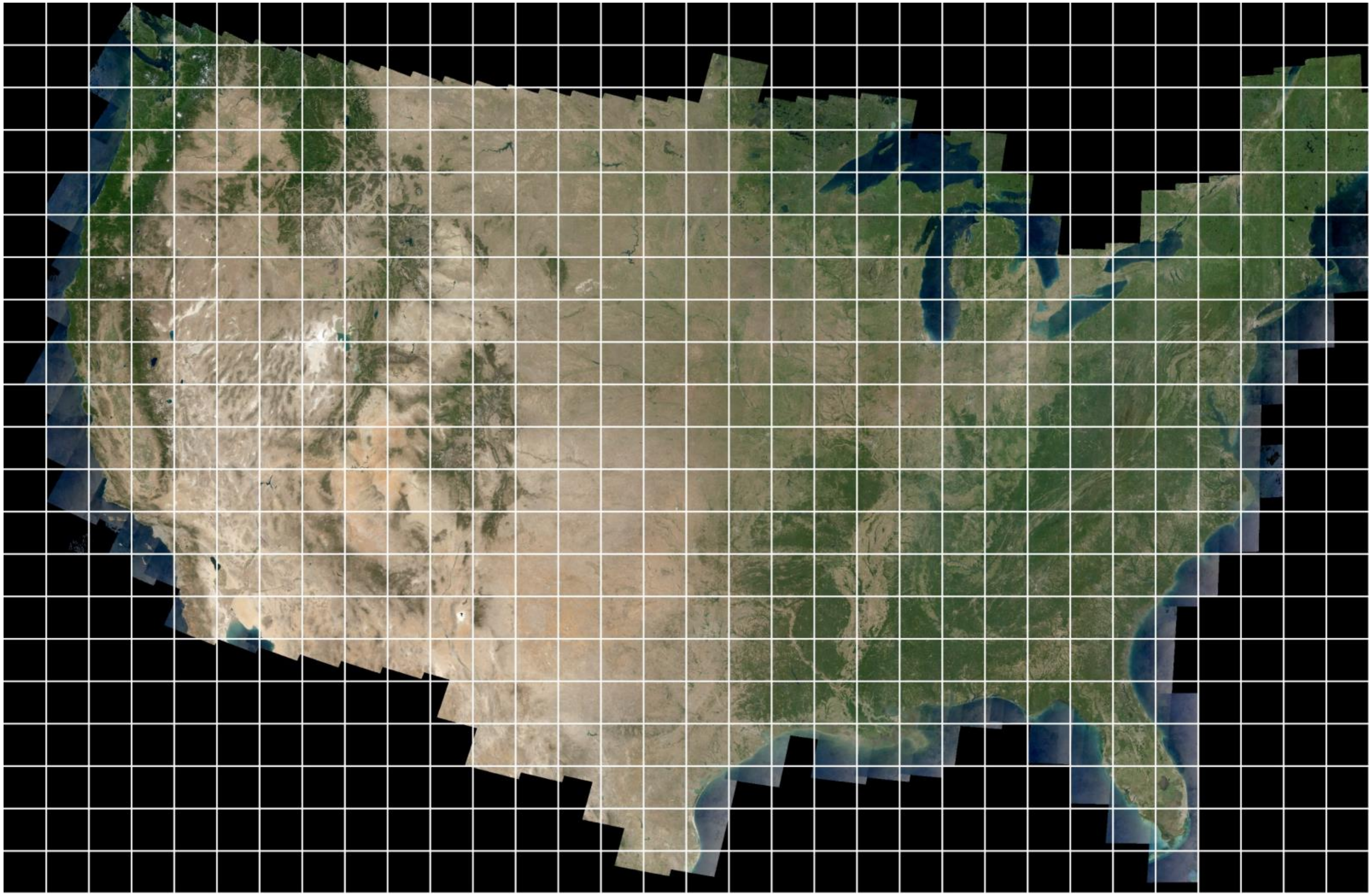


NLCD overall accuracy CNN and random forest

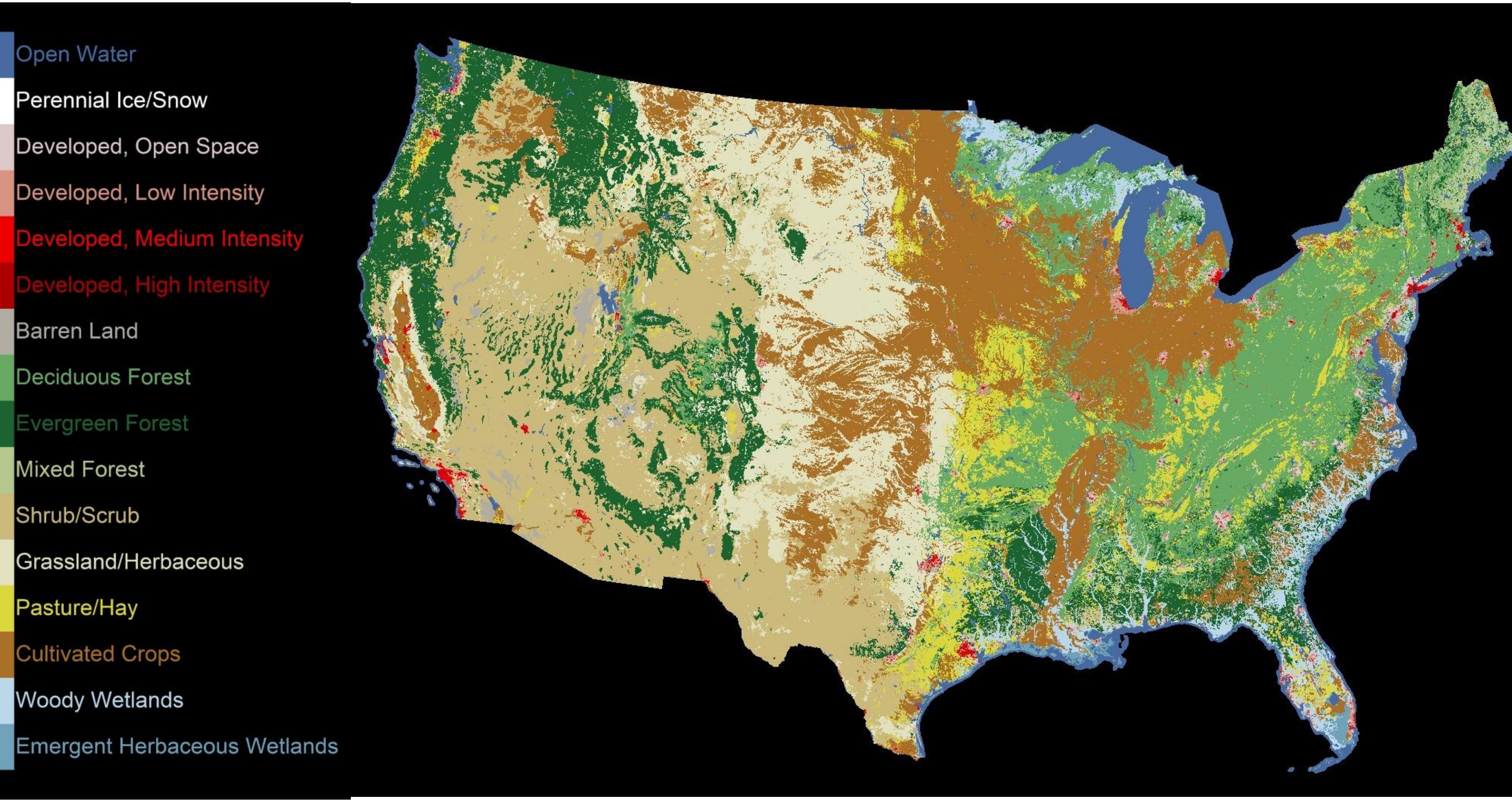


CONUS 50 percentiles of 2011 red, green and blue reflectance

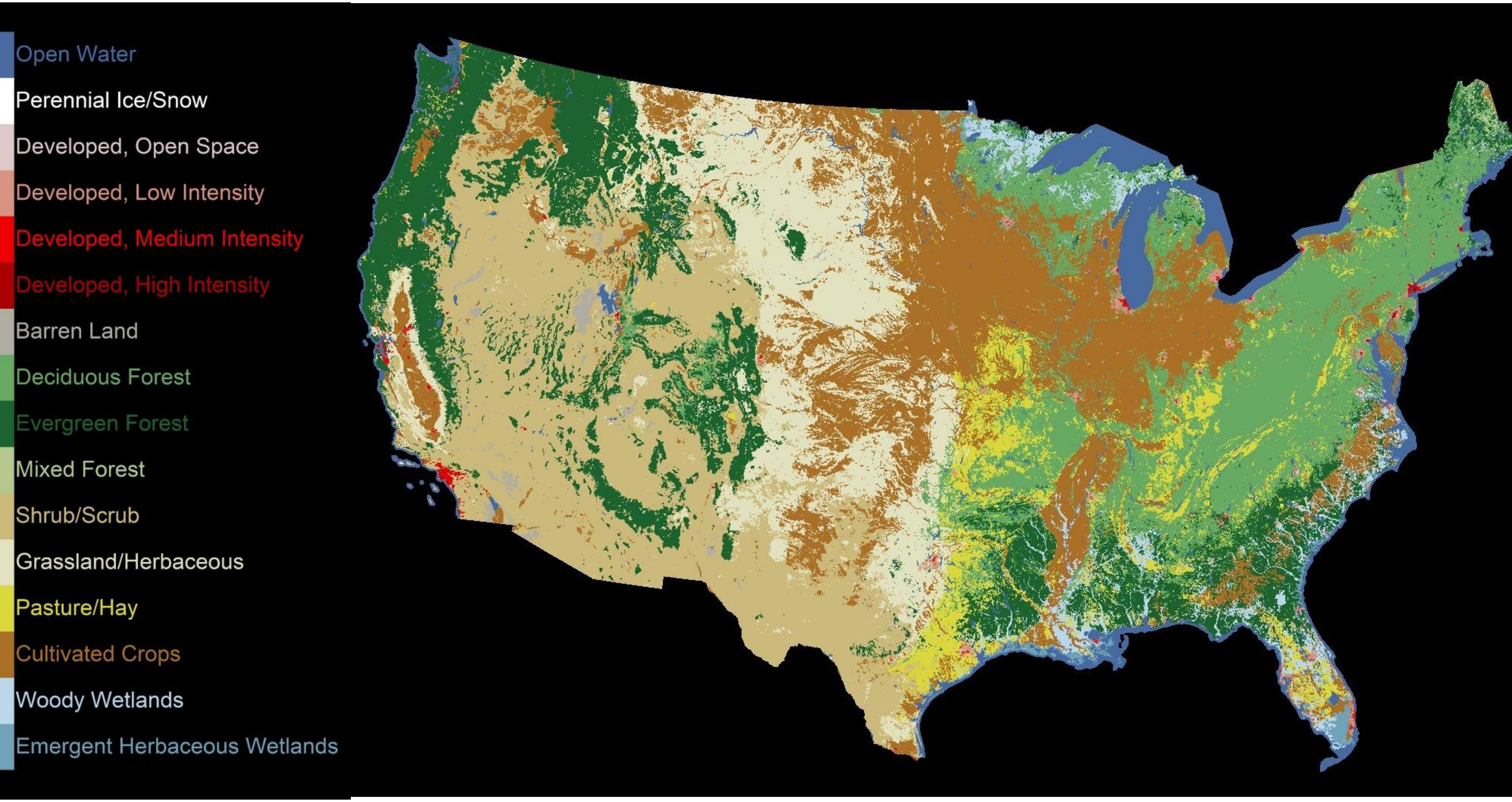
- True color display
- 14785 Landsat-5 & 14680 Landsat-7 ARD tile granules
- From April 1st to October 31st 2011



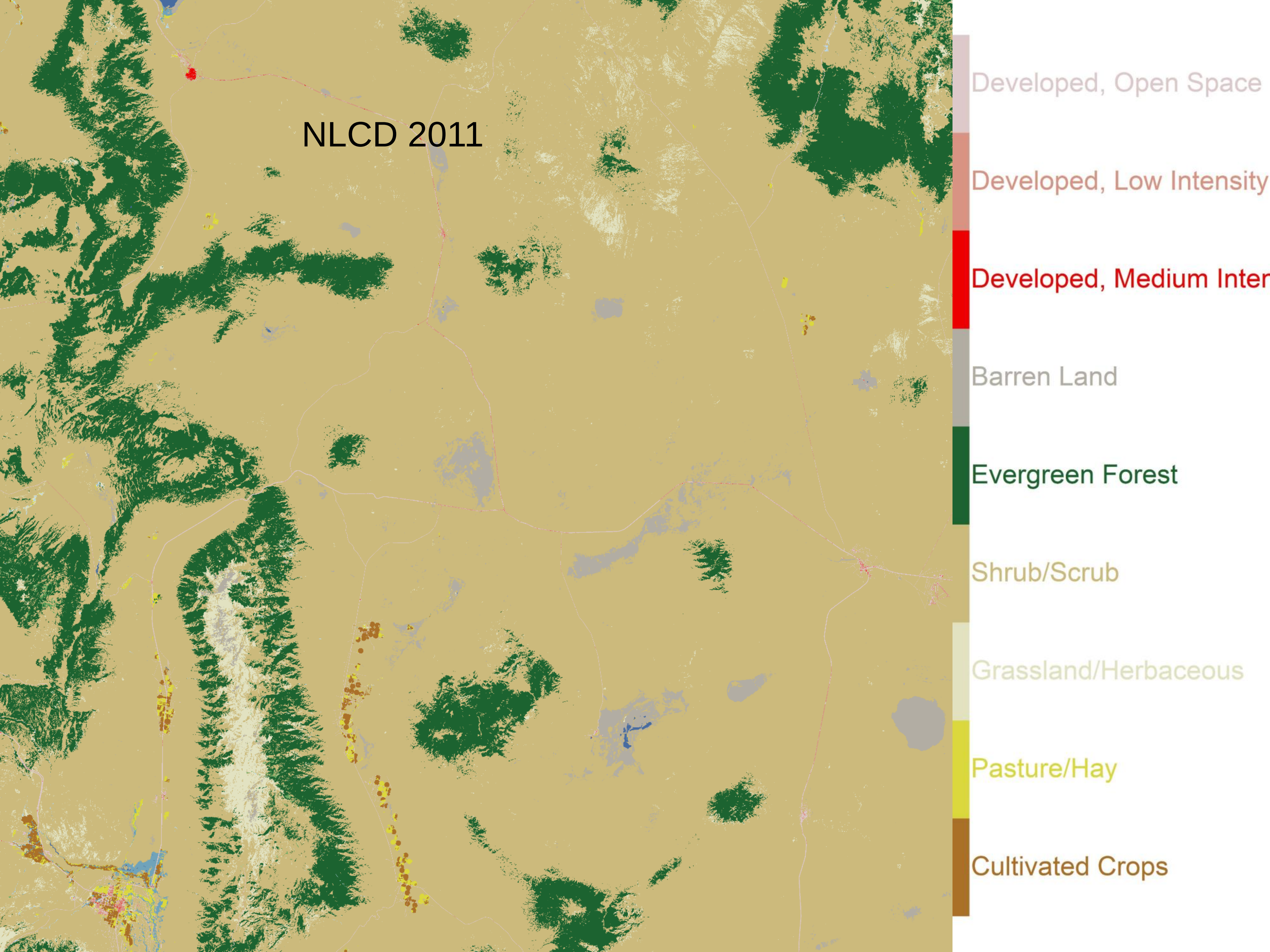
NLCD 2011 – made by USGS

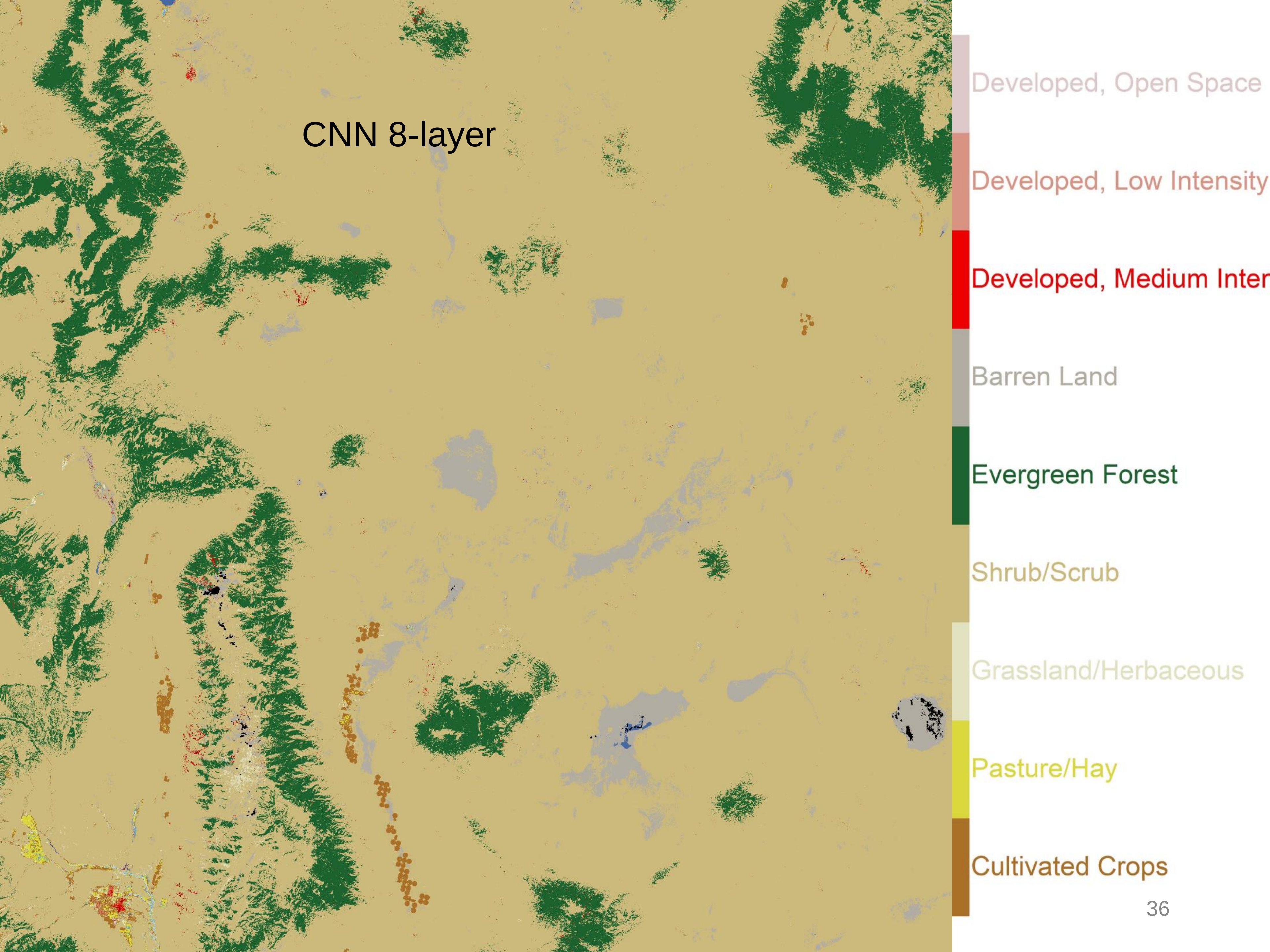


CNN – 8-layer trained using 90% samples



- Landsat 30m true color surface reflectance
- Nevada desert
- 5000×5000 30m ARD tile
- h04v09


NLCD 2011
Developed, Open Space
Developed, Low Intensity
Developed, Medium Inter
Barren Land
Evergreen Forest
Shrub/Scrub
Grassland/Herbaceous
Pasture/Hay
Cultivated Crops


CNN 8-layer
Developed, Open Space
Developed, Low Intensity
Developed, Medium Inter
Barren Land
Evergreen Forest
Shrub/Scrub
Grassland/Herbaceous
Pasture/Hay
Cultivated Crops

Summary and outlook

- ARD is important - Data access and processing is easier

- Deep learning is underexplored for large area and time series applications - ARD is helpful

ISSN: 0034-4257

Remote Sensing of Environment

Submit your Paper | View Articles

Guide for authors Track your paper Order journal

Advancing deep learning for time series analysis

August 2022

Remote sensing time series analysis has been widely used for land cover/use change monitoring and surface parameter inversion. Deep learning models offer operational and practical advantages but should respect remote sensing signal characteristics and application domain pressing challenges. Deep learning for surface parameter estimation that traditionally relied on physical models should be geographically robust, respect physical laws, and/or enable knowledge discovery. Such enhanced deep learning algorithms could substantially advance multiple important disciplines such as land cover studies, time series classification, change detection and continuous monitoring.

This special issue focuses on deep learning to address the following topics:

- Preprocessing or fusion of time series data

Thoughts for Asia-Oceania region

- Landsat ARD is hold in Amazon Web Services in Oregon, US
- Large area data access is difficult
- Consider holding a copy of Landsat & Sentinel-2 data in Asia-Oceania region
  - At least data covers Asia-Oceania
  - Agreement between Amazon or local cloud provider?
  - China should lead?

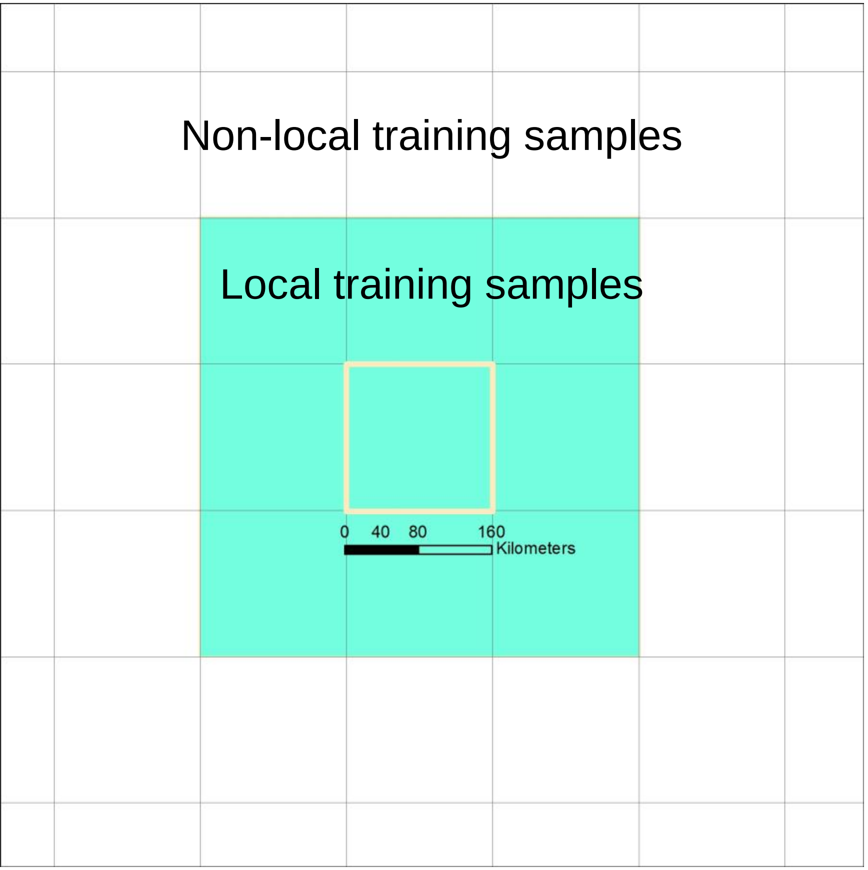
Thoughts for China agencies/satellites

- ‘Open’ policy benefits both users and data providers
  - Good to see Gaofen ARD project

- A systemically maintained satellite series
  - Reflectance consistency (quantitative applications)

Thanks

Locally adaptive random forest classification



For each class

- If more local (3 x 3 WELD tiles) samples than needed, then randomly take needed number from local

- If less local samples than needed, then take all local and randomly select the other needed from the non-local study area

For large volume data computation is getting cheaper

- Compute is cheap & disk storage is getting cheaper

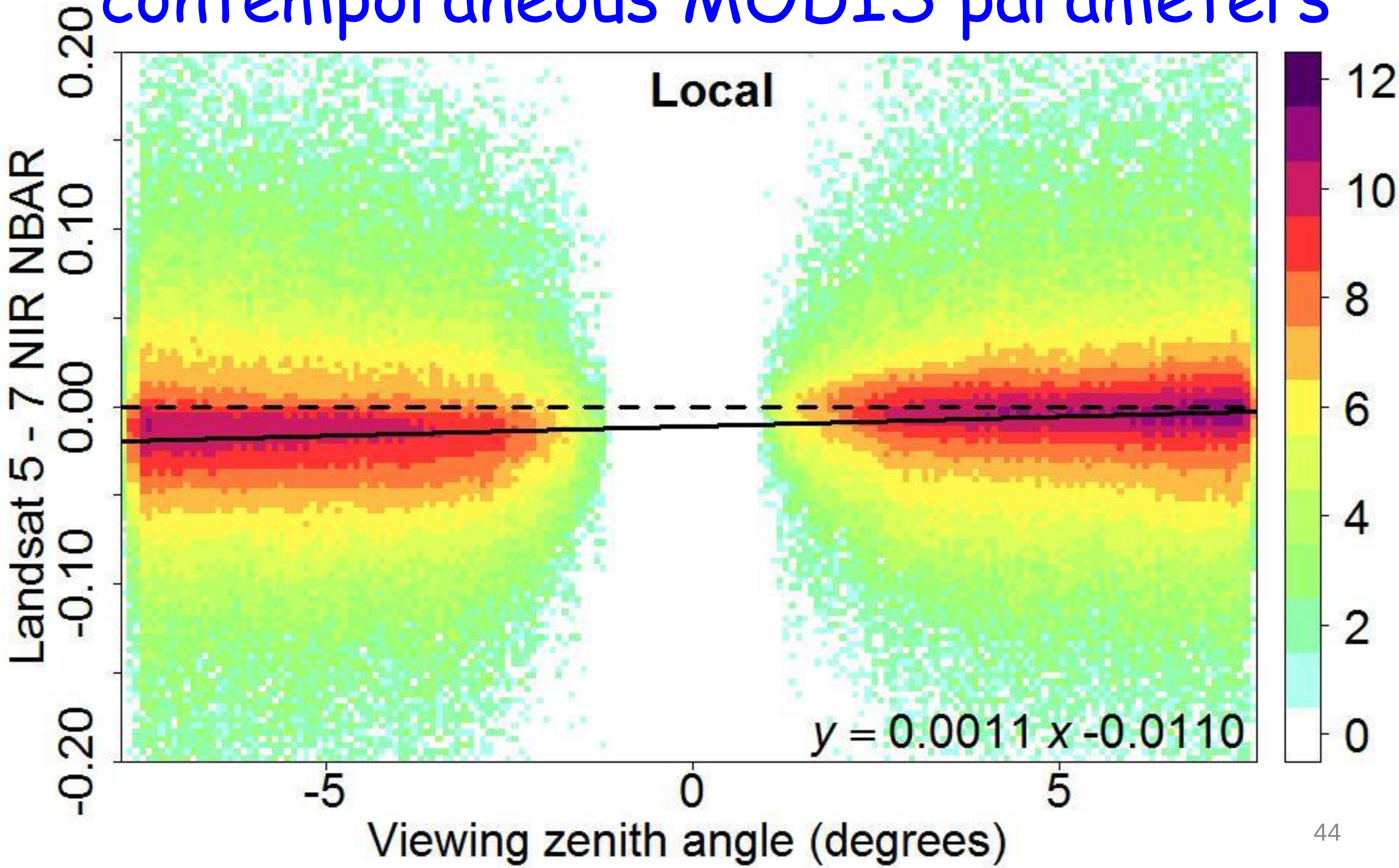
- The North America classification
  - processed on 512 GB of memory, 32 cores 64-bit Linux
  - 6TB input global WELD data
  - 91 hours for single random forest classification
  - 242 hours (10 days) for locally adaptive random forest classification

- Global WELD
  - Processed on NASA NEX super computer
  - A month to process three year GWELD

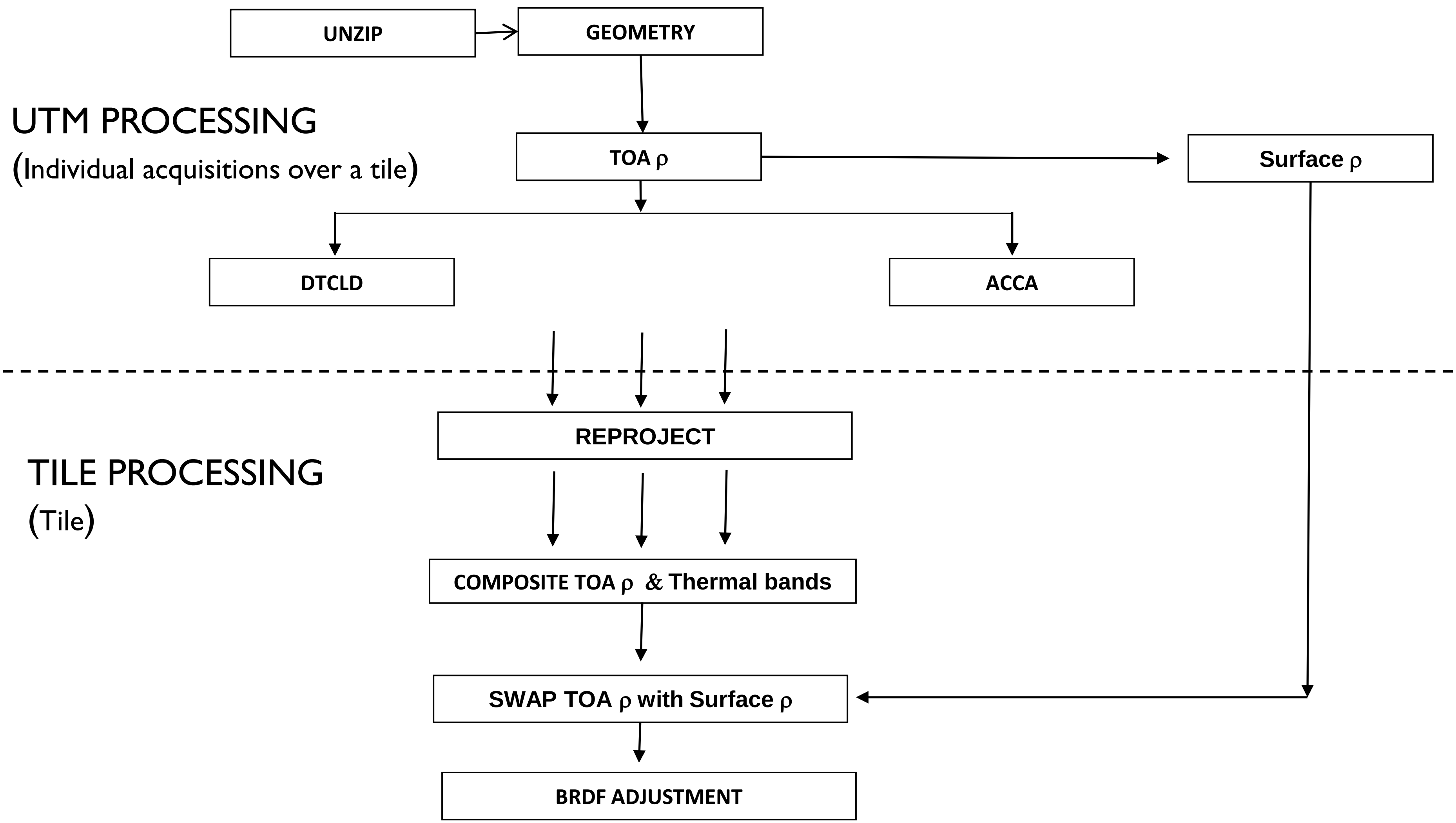
30 m applications using Global WELD

True color 2010 annual 30 m GWELD NBAR product

(Landsat 5 TM - Landsat 7 ETM+) NBAR derived by the local spatially and temporally contemporaneous MODIS parameters



Overview of Global Version 3.x WELD Processing Sequence



Product resources

- monthly version 3.1 at https://lpdaac.usgs.gov/products/gweldmov031/
- annual version 3.1 at https://lpdaac.usgs.gov/products/gweldyrv031/
- monthly version v3.0 at https://lpdaac.usgs.gov/products/gweldmov003/
- annual version v3.0 at https://lpdaac.usgs.gov/products/gweldyrv003/
- ATBD (Algorithm Theoretical Basis Documents) at https://lpdaac.usgs.gov/documents/496/GWELD_ATBD.pdf
- User guide at https://lpdaac.usgs.gov/documents/497/GWELD_User_Guide_V31.pdf

Free data benefits: global Landsat applications

- Global 30 m land cover
  - Gong et al. (2013). Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 34(7), 2607-2654.
  - Supercomputer at Tsinghua University

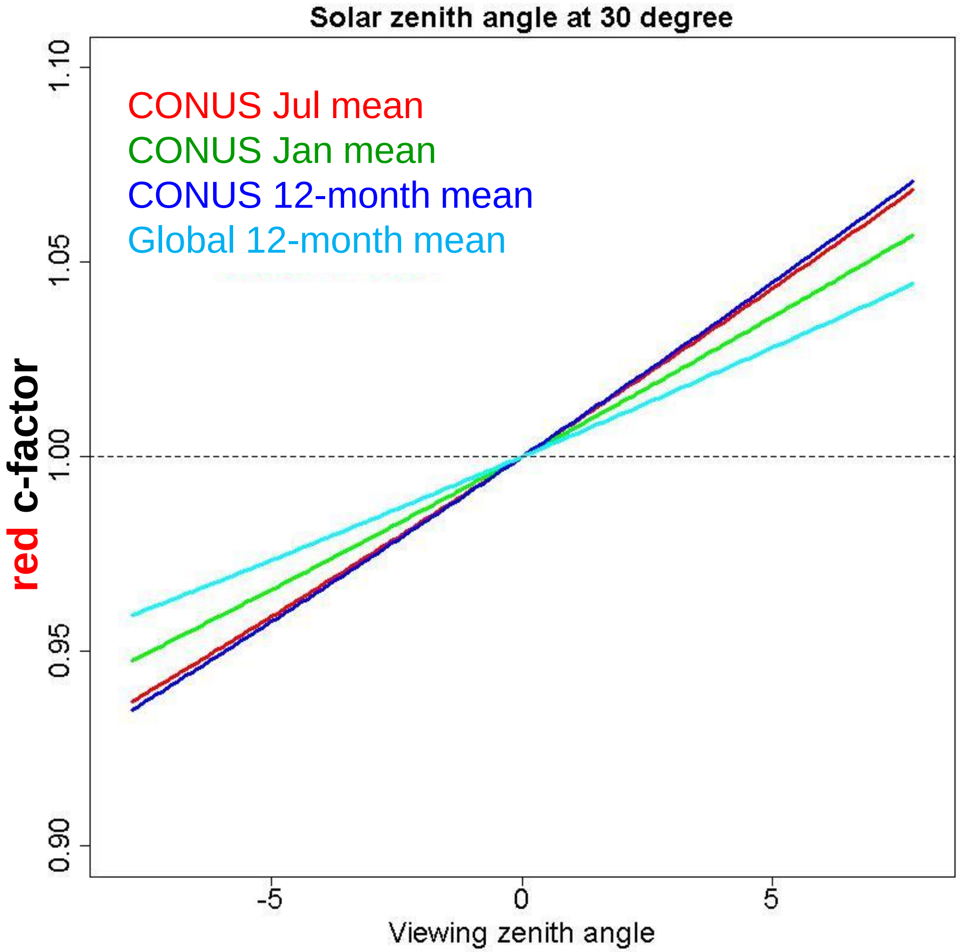
- Global 30 m forest loss
  - Hansen et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850-853.
  - Google earth engine

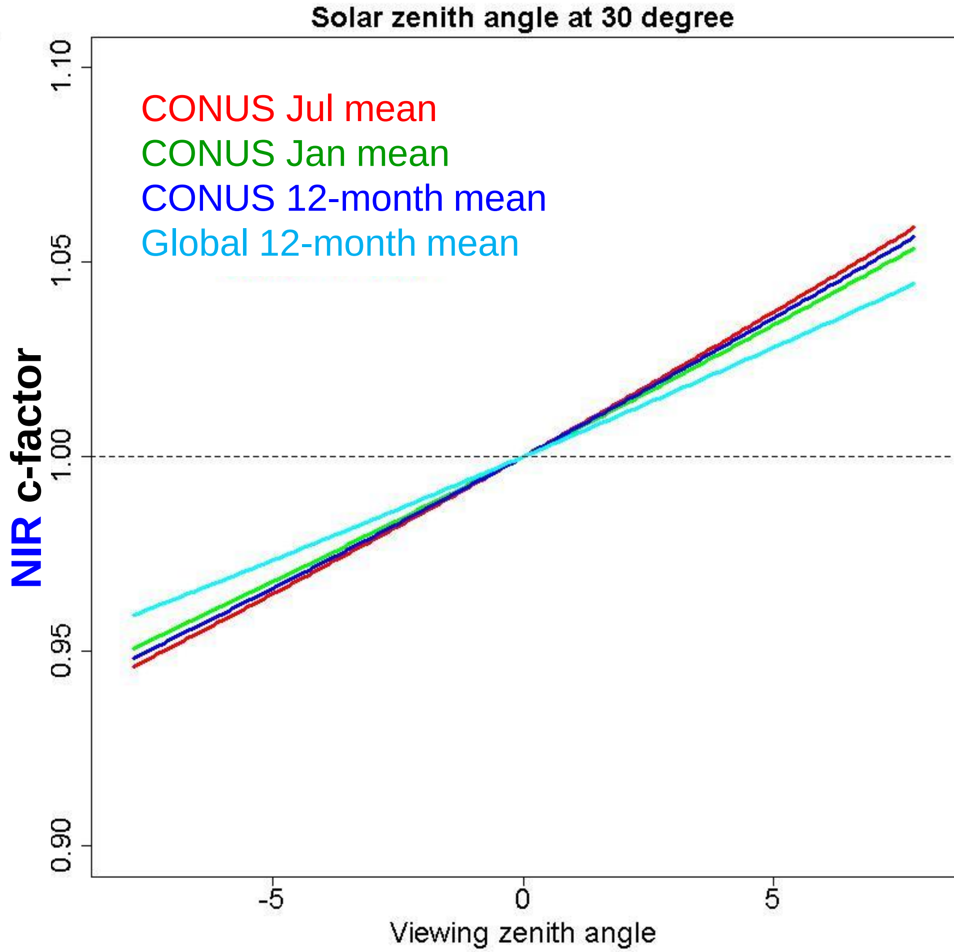
Free data benefits: global 30 m applications

- Global 30 m water mapping
  - Pekel et al. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418.
  - Google earth engine

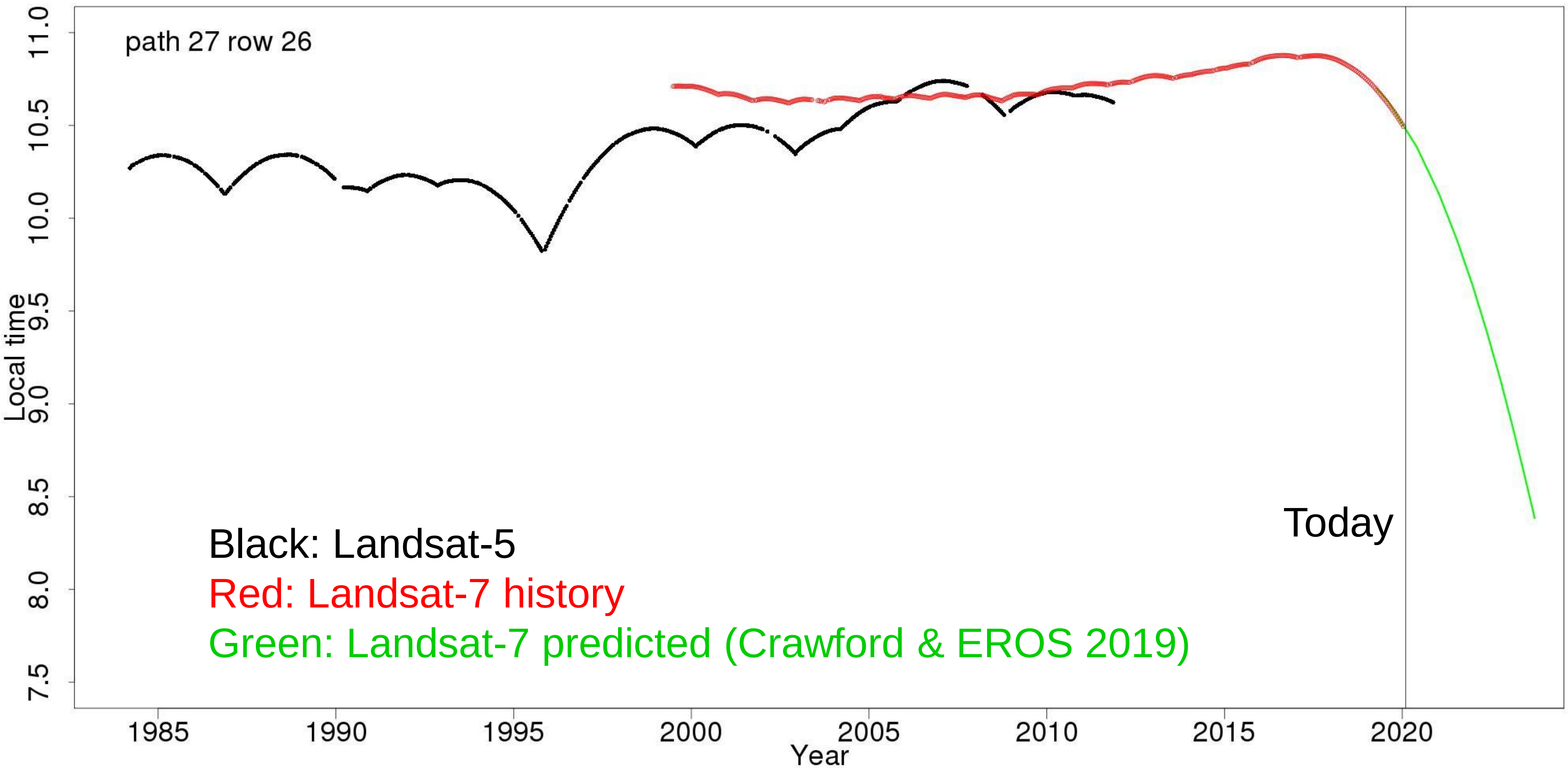
- Global 30 m urban mapping
  - Liu et al. (2020). High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nature Sustainability*, 1-7.
  - Google earth engine

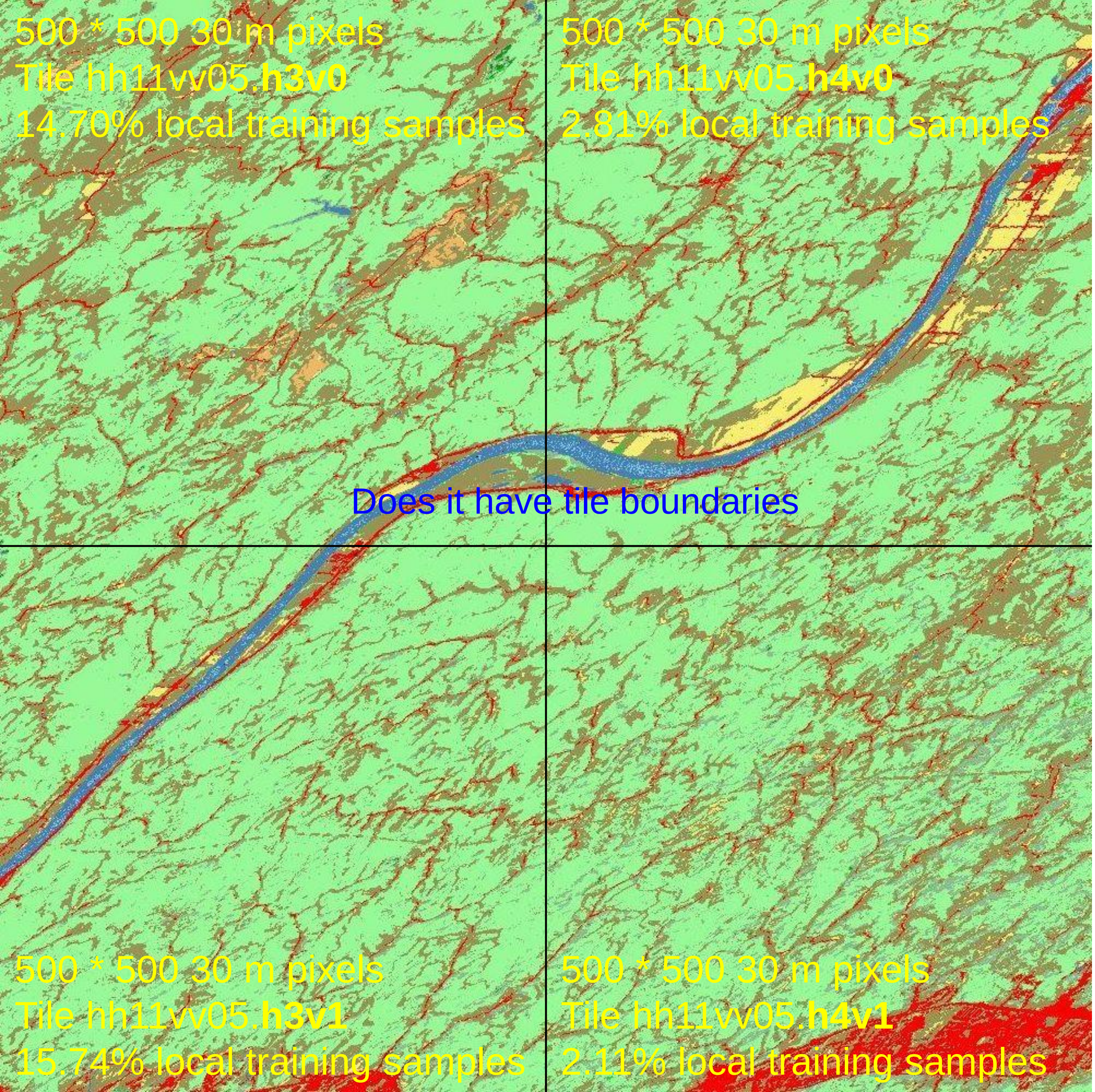
BRDF shape over Landsat 15° FOV
using just 3 mean MDC43 model parameters



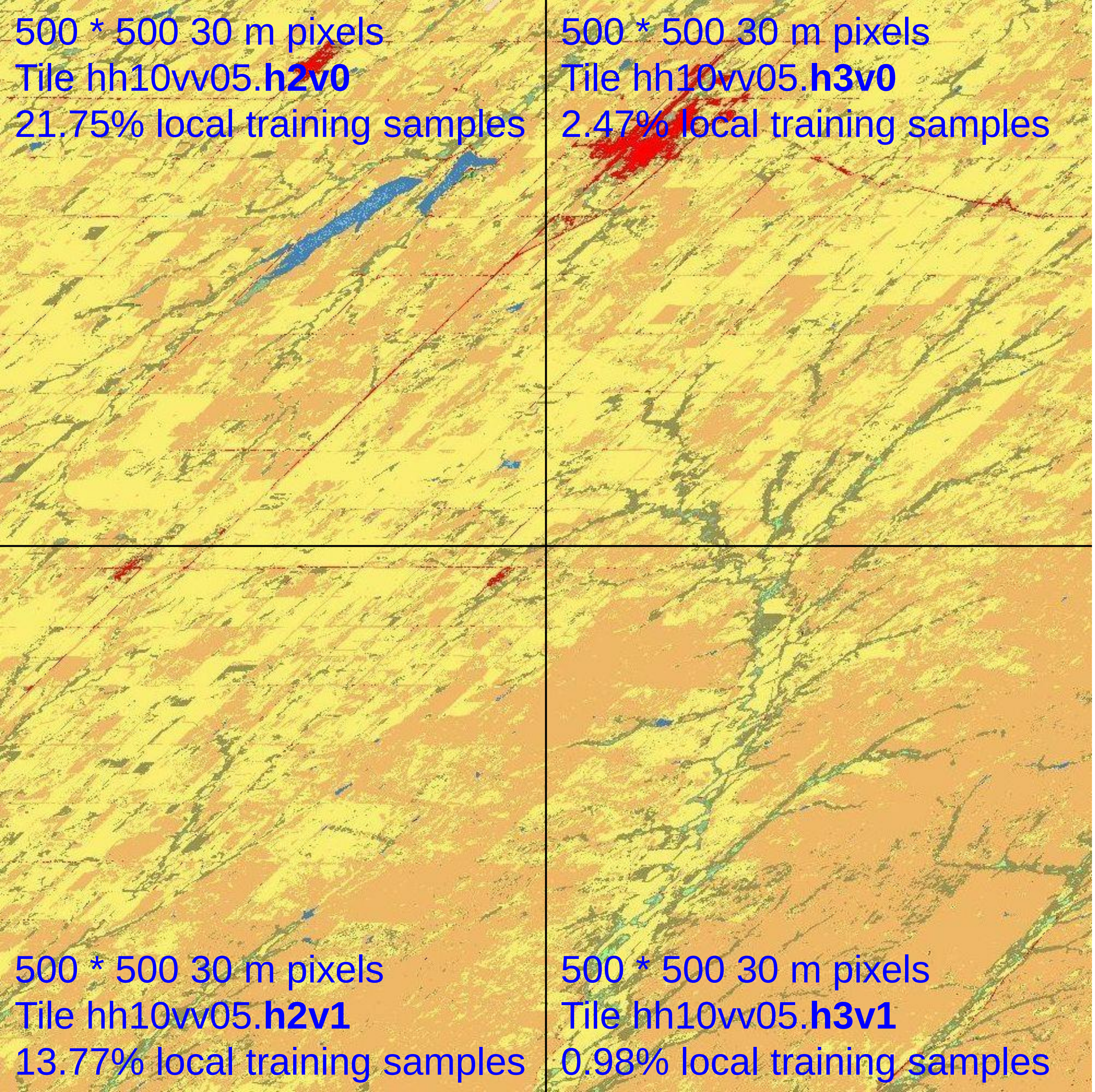


Landsat 7 more orbit drift



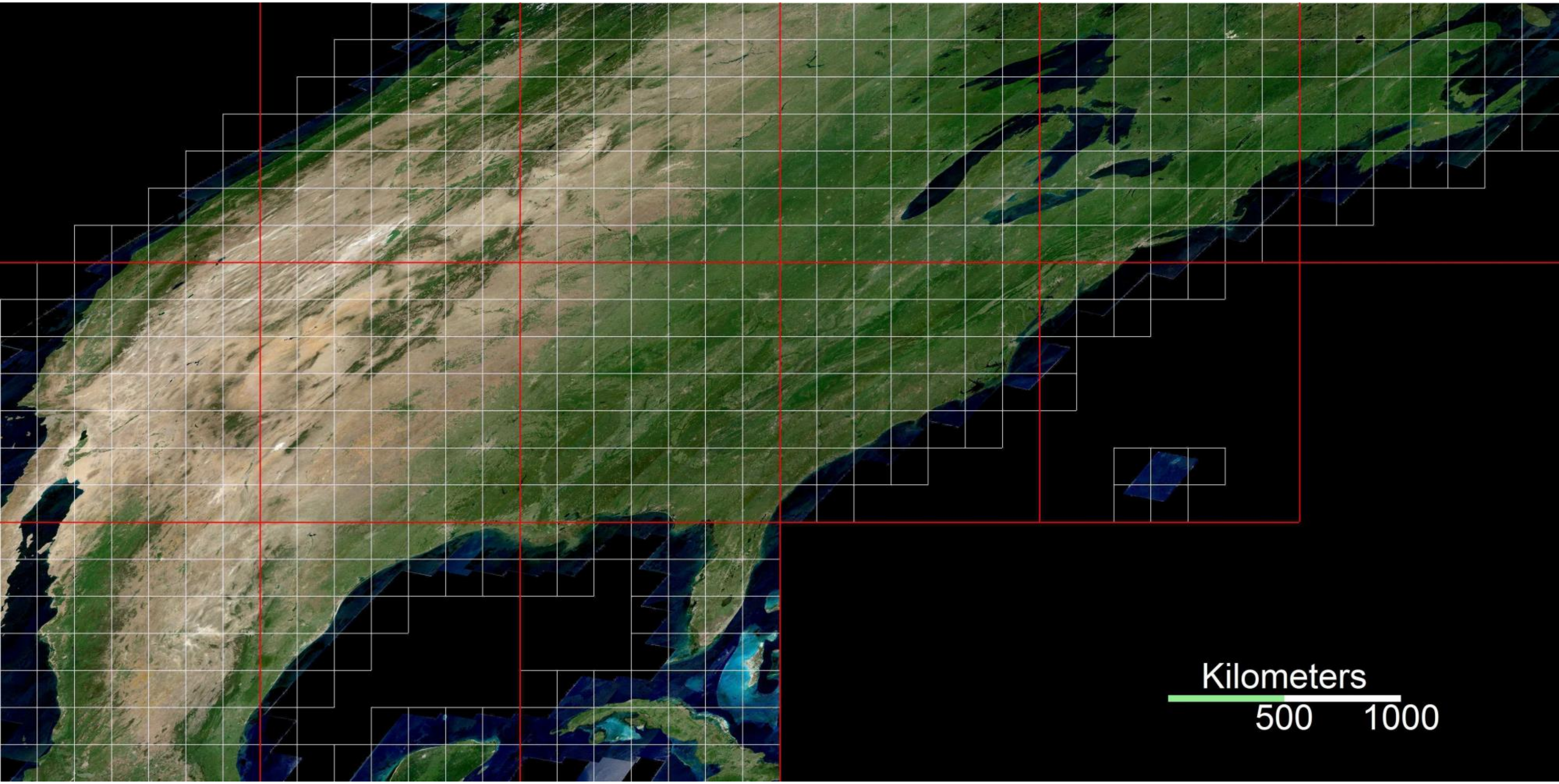


- Four tiles intersection
- Locally adaptive classification results
- 1000 × 1000 30 m
- The Crown City, Ohio



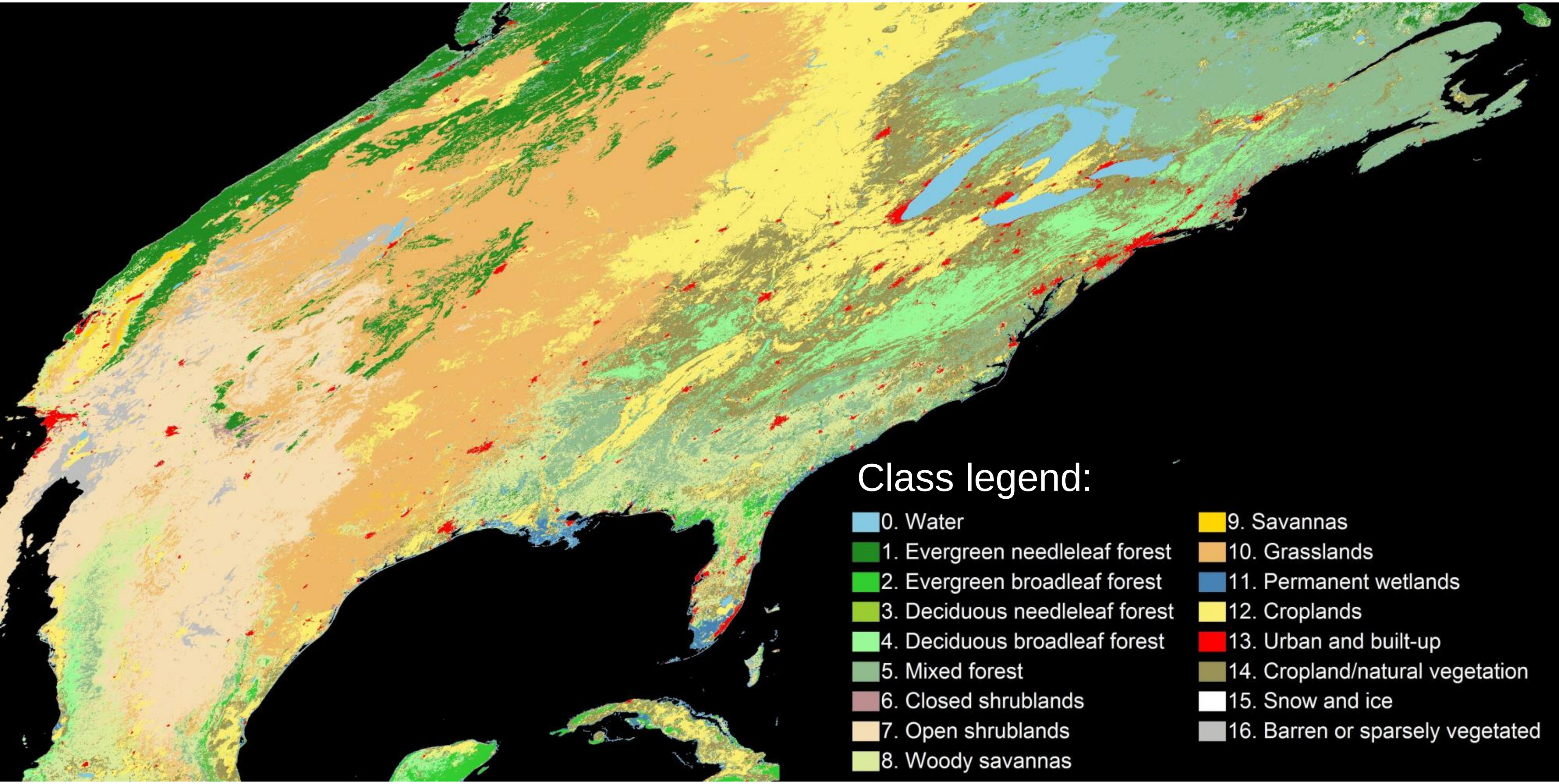
- Four tiles intersection
- Locally adaptive classification results
- 1000 × 1000 30 m
- Herington, Kansas

MODIS land products 30 m aligned with global WELD data



10° x 10° 500m global MODIS Land Cover product (MCD12Q1) tile (17 IGBP land cover classes)

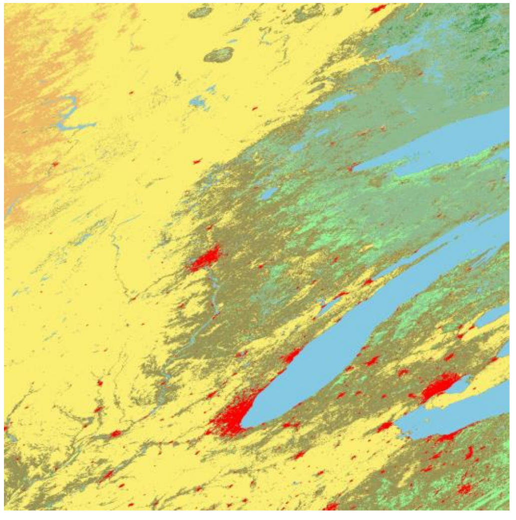
Land cover training class labels: 500 m MODIS IGBP land cover product (MCD12Q1) for 3 years (2009-2011)



MCD12 overall land cover classification accuracy 75% (Friedl et al. 2010)

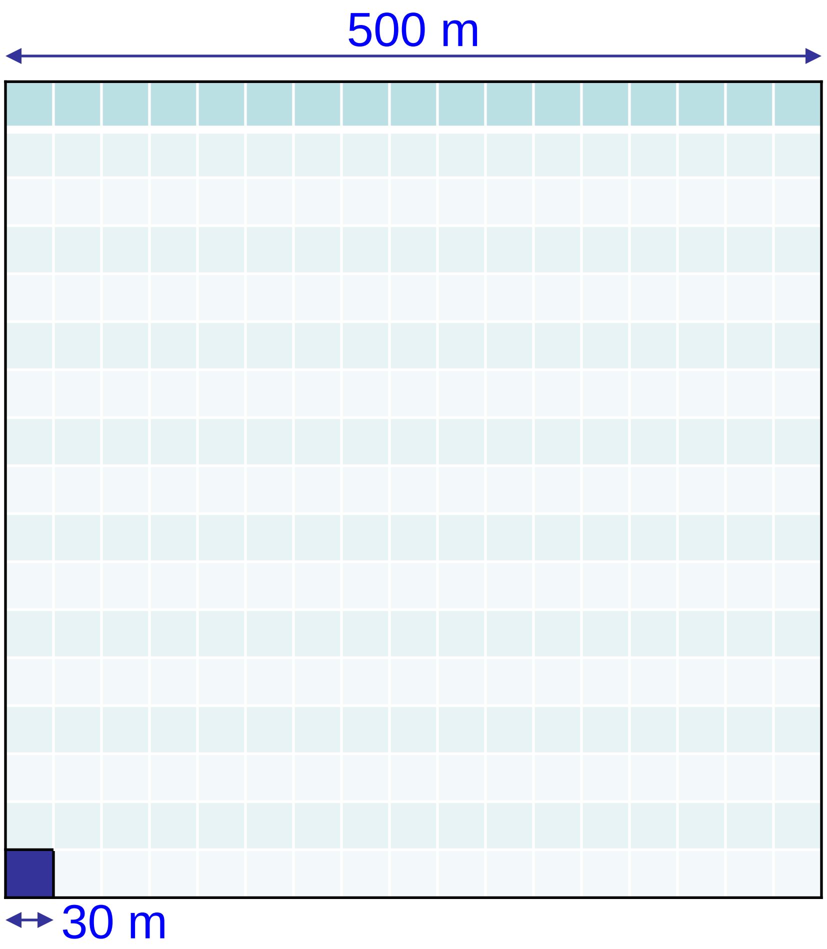
Land cover training class labels: MCD12Q1 filter

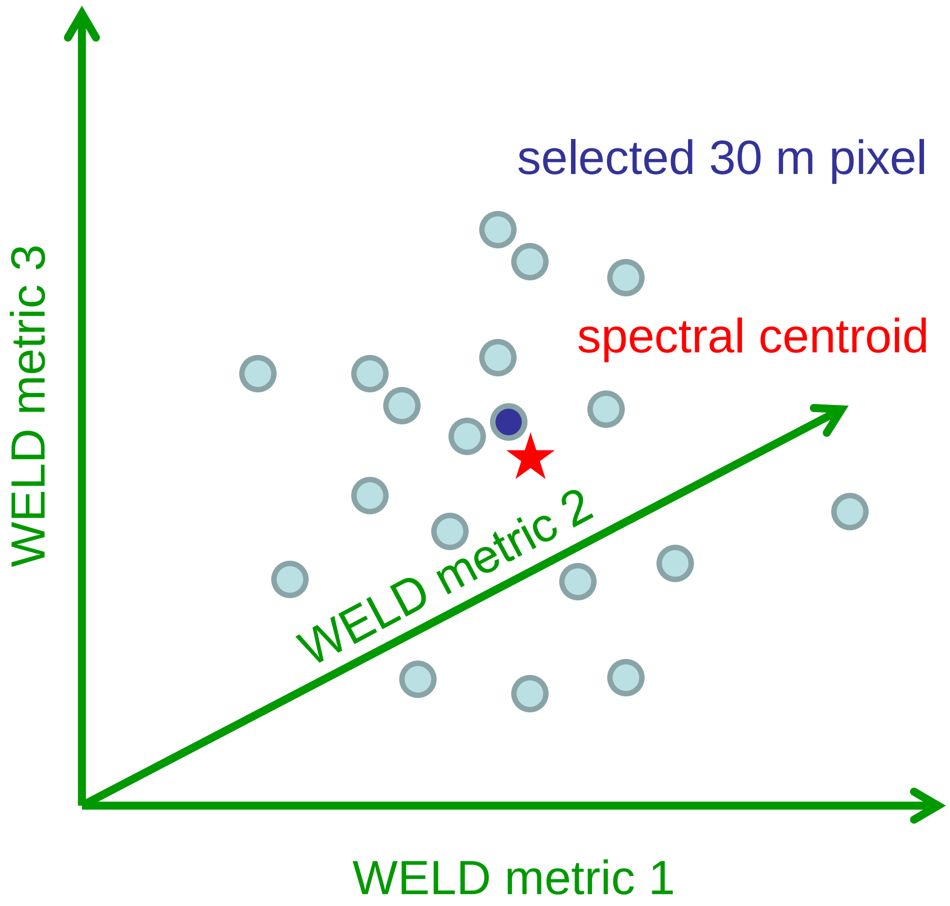
- Only 500 m pixel locations that have
  - same MCD12Q1 land cover class over the three years (2009 to 2011)
  - same land cover type with neighbor pixels
  - classification confidence > 50%
  - quality assessment = “good quality”



8.44% of the North America MCD12Q1 2010 500 m land pixels were retained

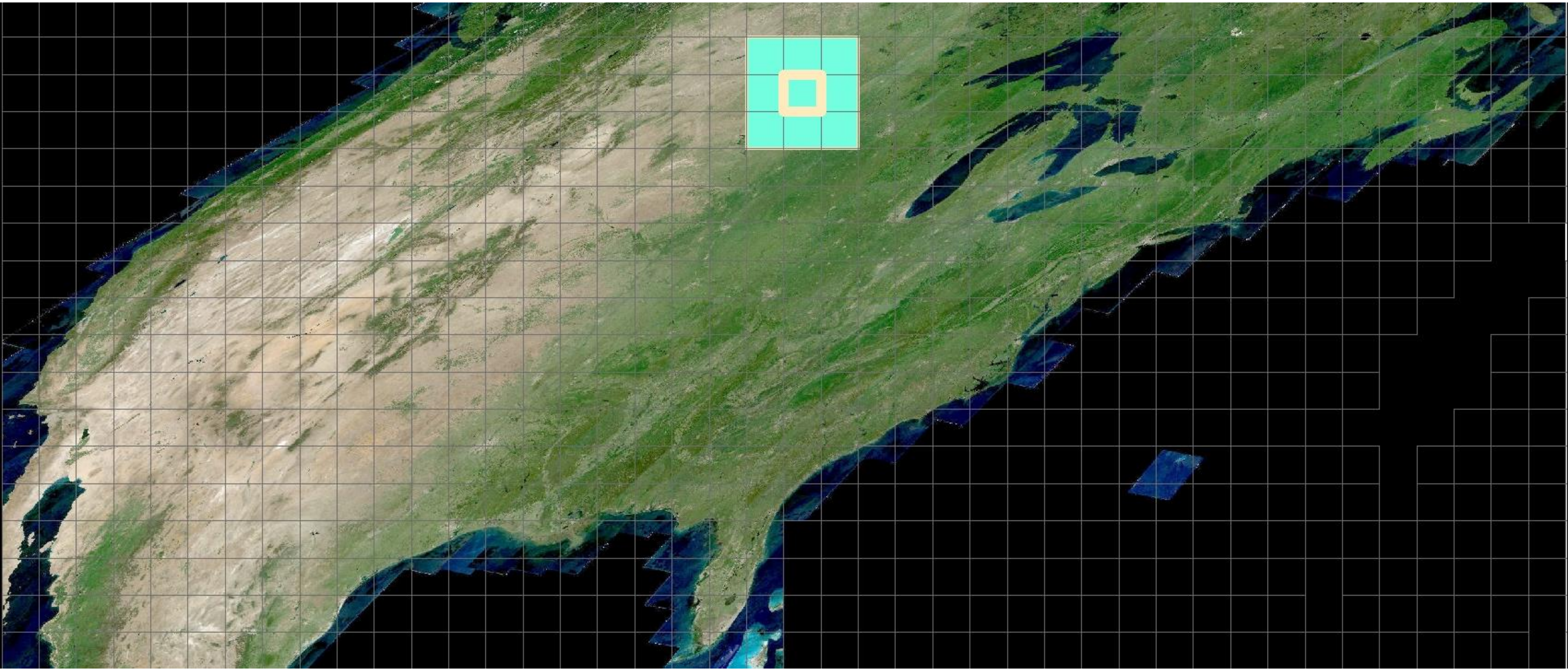
Land cover training class labels: How to associate each selected 500 m pixel class label with the most representative 30 m pixel metric data ?



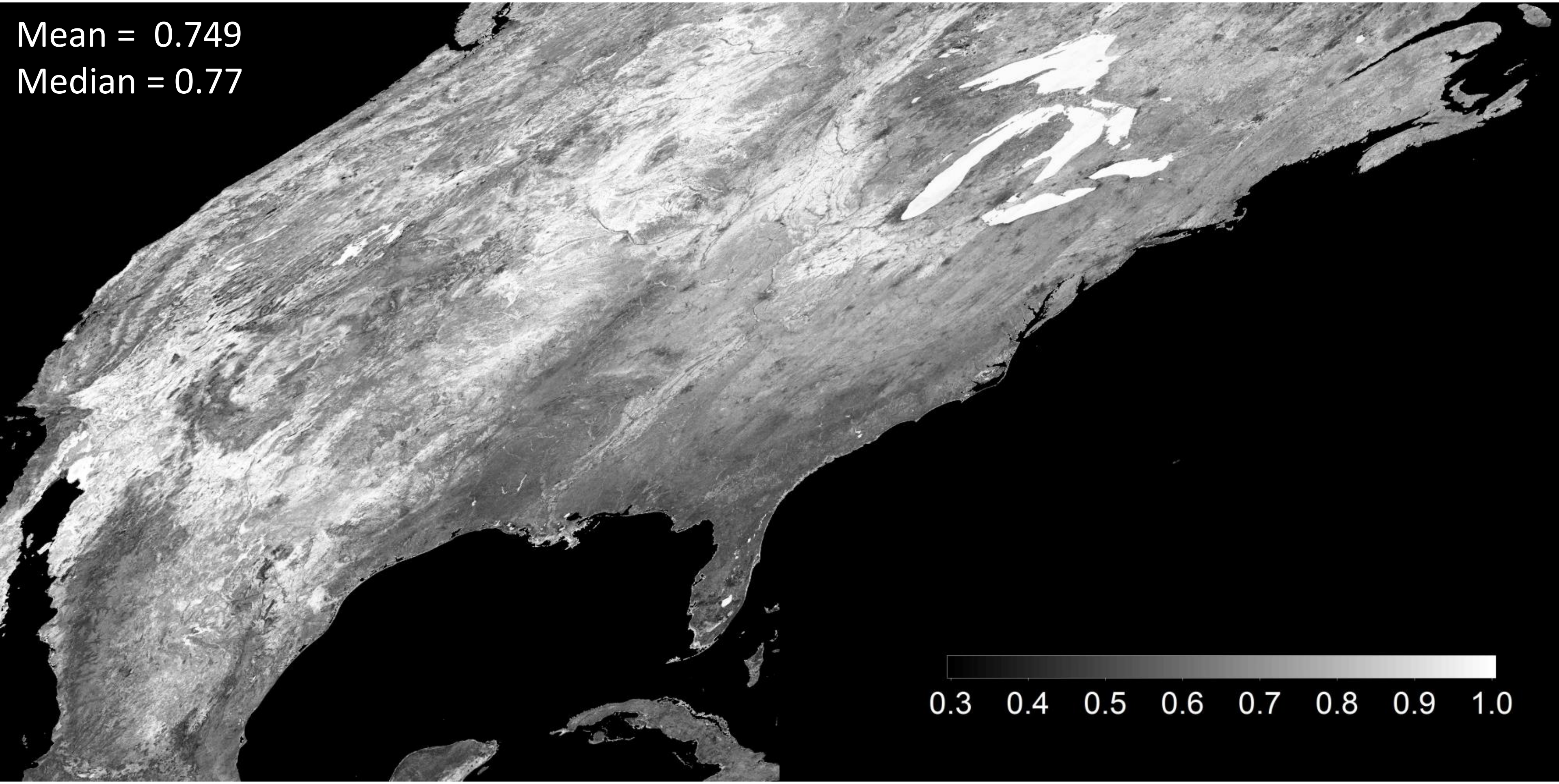


Select the 30 m pixel in 39 dimensional feature space (metrics) that is the most similar to the other 17 x 17 30 m pixels -> spectral centroid

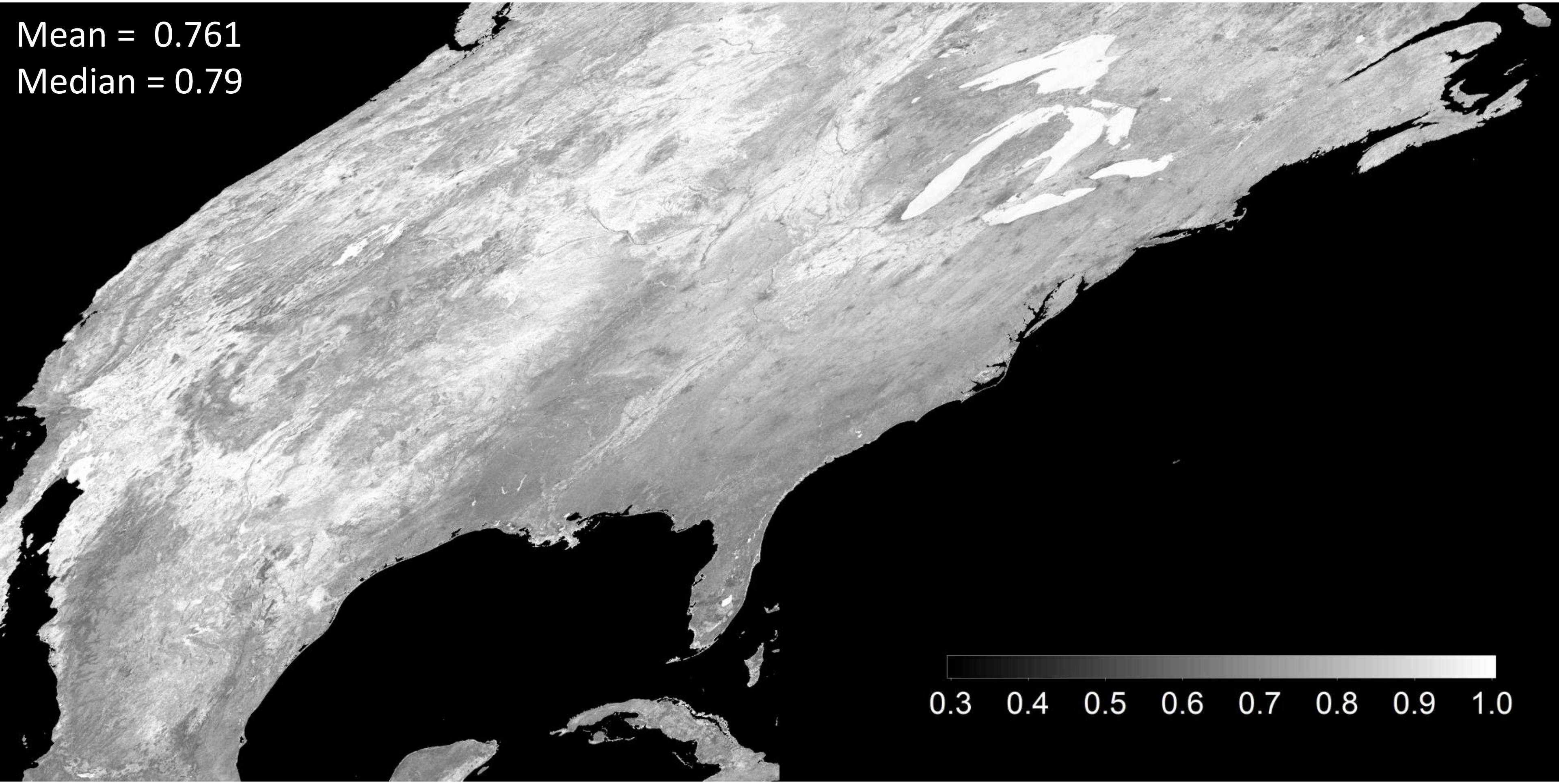
Locally adaptive random forest: 511 random forest



Results: Single random forest confidence map



Results: Locally adaptive random forest confidence map



Results: 30 m Locally adaptive random forest