1st Workshop on Leveraging AI in the Exploitation of Satellite Earth Observations & Numerical Weather Prediction

23-25 April 2019

Developing fine-scale snow cover fraction estimates using Deep Learning

Soni Yatheendradas[1,2], Sujay V. Kumar[2], Daniel Duffy[3]

[1]UMD/ESSIC

[2]NASA / GSFC: Hydrological Sciences Laboratory

[3]NASA / GSFC: NASA Center for Climate Simulation

24 April 2-19, Session 2, 10:50-11:10 am

Introduction and overall goals

- Need fine-resolution, gap-free, satellite-based data (e.g., snow cover fraction, or SCF) for evaluating local hydrology
  - MODIS MOD10A1 v5 SCF @ 500 m: gaps in spatial coverage
  - MOD10C1 v5 SCF daily @ 5-km: near-complete spatial coverage
- Downscale / super-resolution regression from MOD10C1 to obtain MOD10A1 SCF- like product without gaps
- Demonstration of this Deep Learning prototype for a target 1-km resolution

Background

- Computer vision: simple super-Resolution Convolutional Neural Network (SRCNN): for fast, state-of-the-art image restoration
- Emerging geoscientific super-resolutioning applications
  - Use serially stacked SRCNNs etc.;
    - Augmented with auxiliary data channels (e.g., terrain elevation)
  - Cannot adequately handle input data gaps (e.g., at coastal pixels)
- Computer vision innovations in image inpainting: partial convolution
  - Handles <u>gaps consistent across input channels (e.g., RGB in image)</u>

Spatial domain and data

- 3° X 3° over Central California & Nevada
- 3-year training period (2009-2011)
- Similar development (dev) / validation and test set distributions
  - Alternate days of an year (2012) form these 2 sets

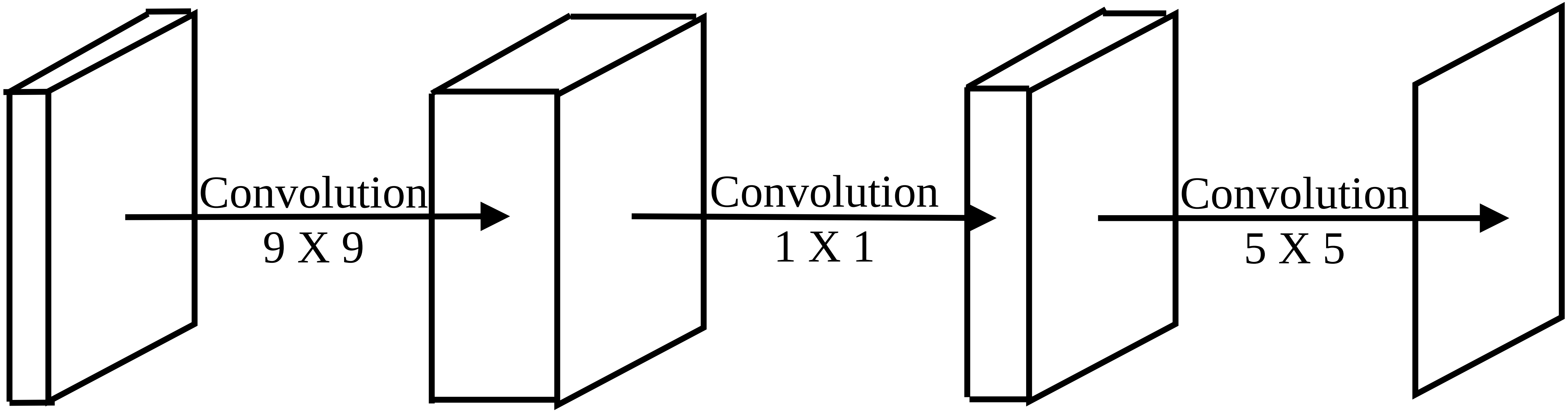
- MOD10C1 SCF input has 3 channels to be considered together: SCF, cloud cover fraction (CCF) and Confidence Index
- Different auxiliary channel types:
  1. Static terrain-related: elevation, slope, aspect
  2. Dynamic satellite-based: MOD10A1 snow albedo, MOD11A1 land surface temperature (LST)
  3. Dynamic LSM-based (land surface model): precipitation, snow water equivalent (SWE), surface radiative temperature, leaf area index (LAI)

Our infrastructure

- Developing infrastructure called MENSA (Machine learning Environment for NASA Scientific data Applications) coded in Python
  - Modified to handle and fill gaps that vary across the input / auxiliary channels
- Successfully implemented on NCCS ADAPT GPUs
- Following accomplishments till date:
  - Keras API layer for modified partial convolution
  - Loss functions in Tensorflow / Keras that consider only valid spatiotemporal data values in target image, e.g., RMSE
- Ongoing work:
  - Additional modified Keras API layers, e.g., batch normalization, activation, addition, pooling.
  - Other loss functions, e.g., that consider visual semantics

Data flow schematic through SRCNN

- Augmentation of traditional Dong et al* [2014] SRCNN
- ‘Same’ padding throughout during convolution



Input stack: MOD10C1 + Auxiliary

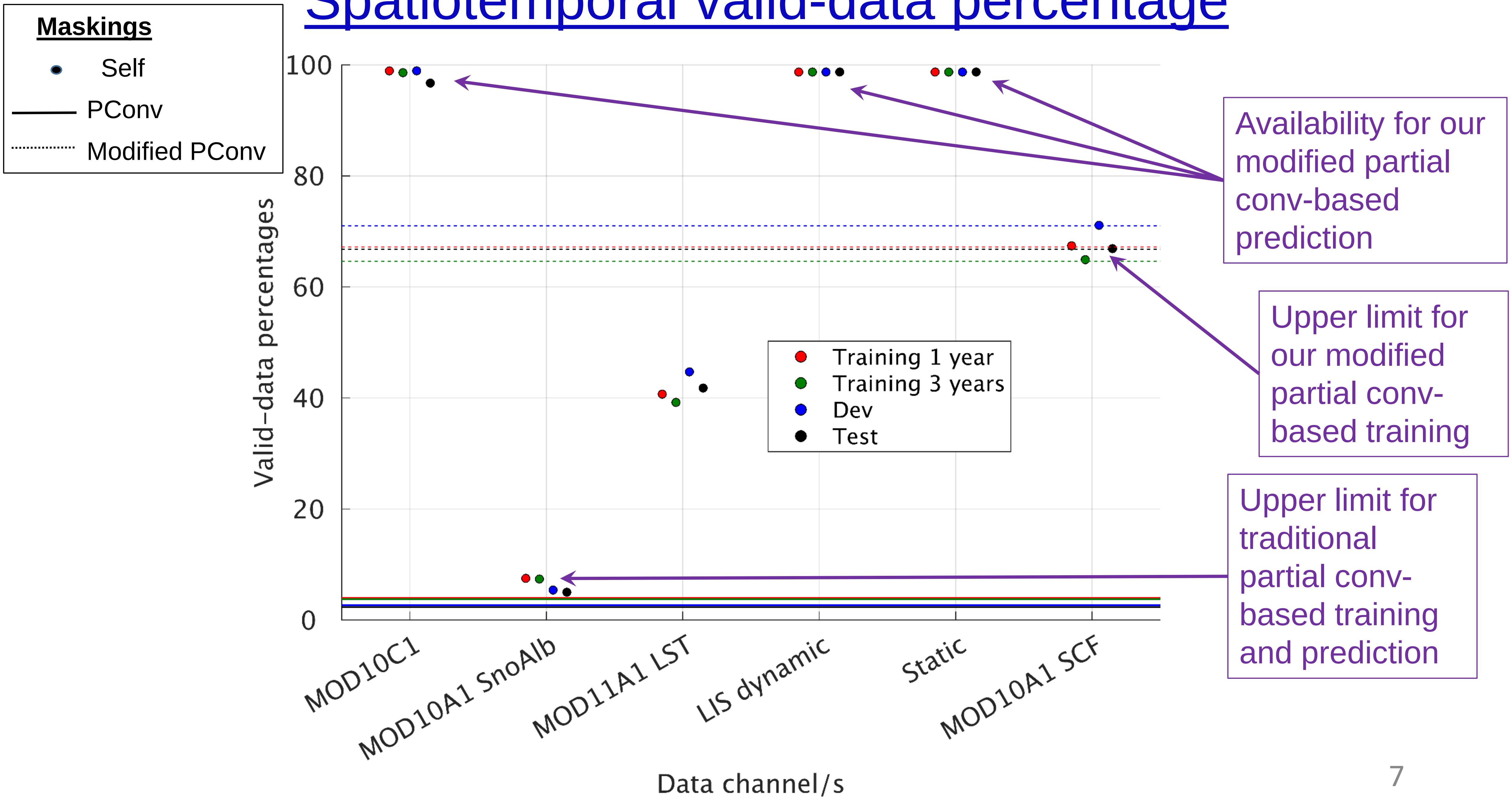
Activation stack 1: 64 channels

Activation stack 2: 32 channels

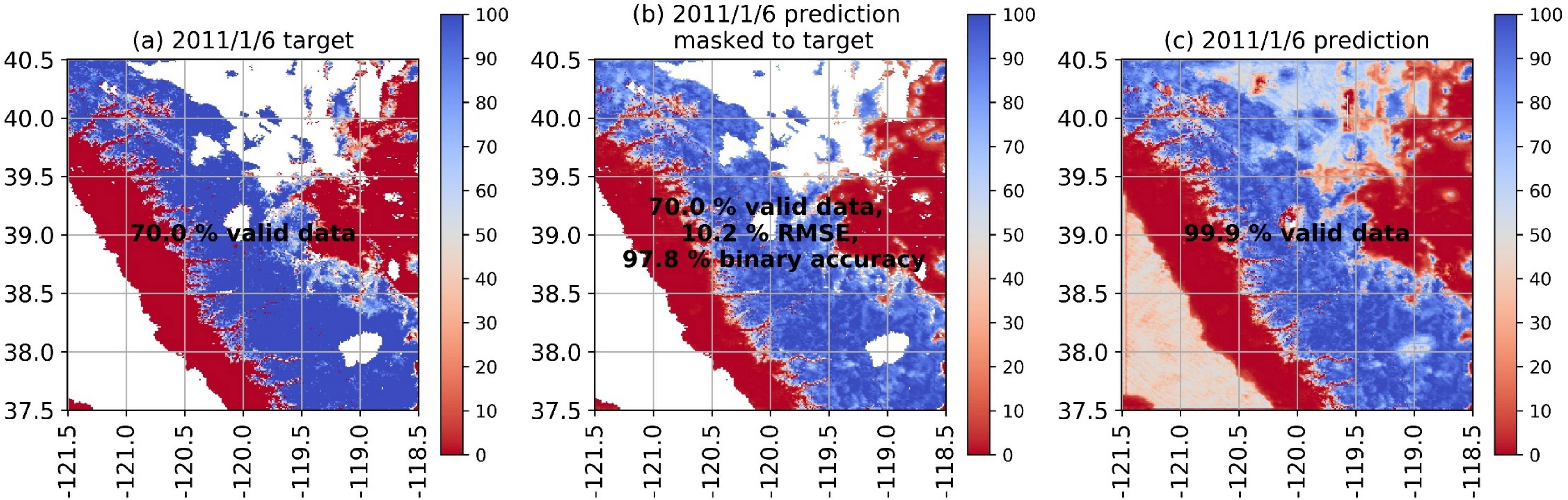
Output MOD10A1 SCF-like grid

* Dong et al [2014]: Learning a deep convolutional network for image super-resolution

Spatiotemporal valid-data percentage



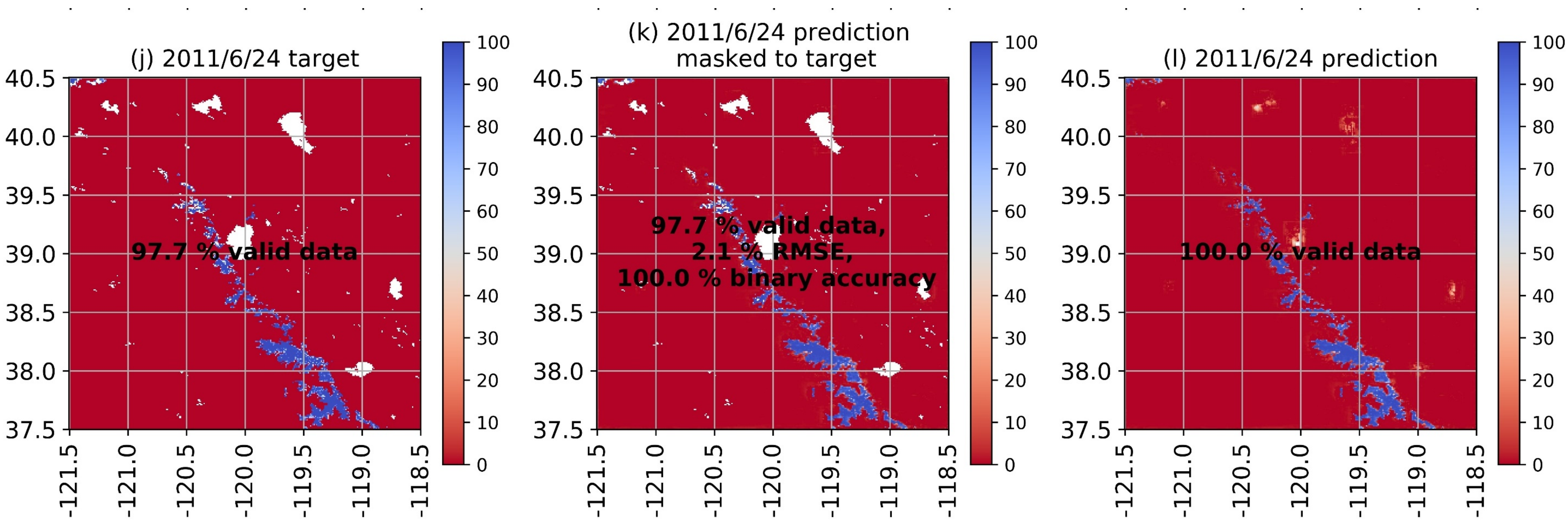
Sample winter day's prediction vs. target



Batch RMSE [%]

- Training: 5.7
- Dev : 5.1
- Test : 5.3

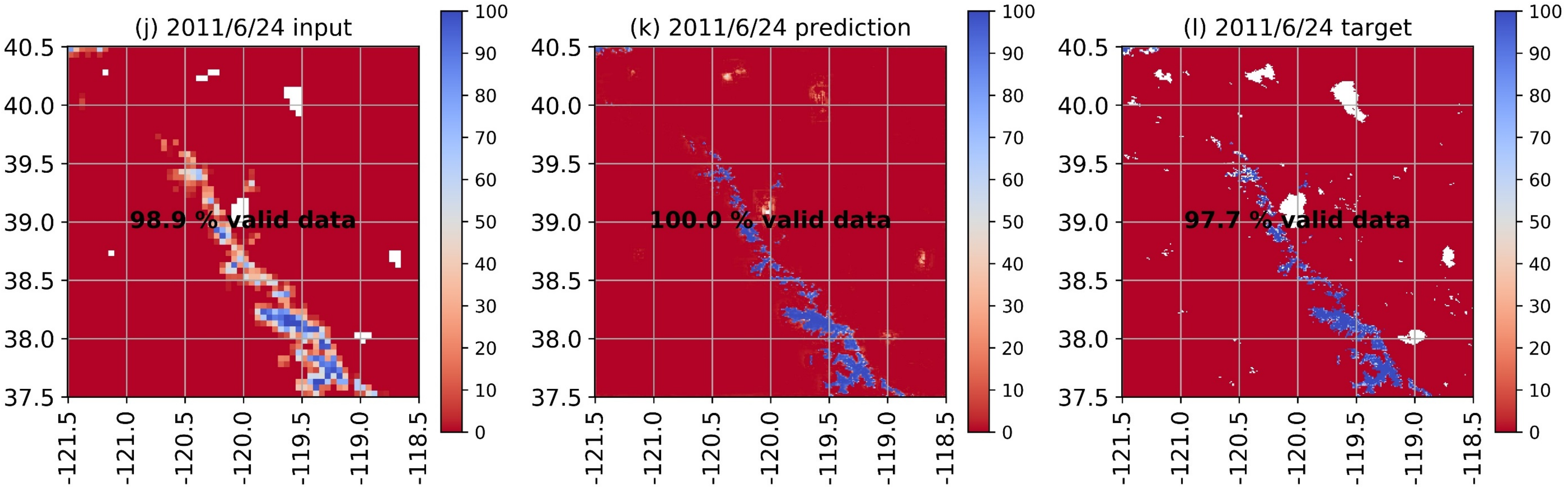
Sample summer day's prediction vs. target



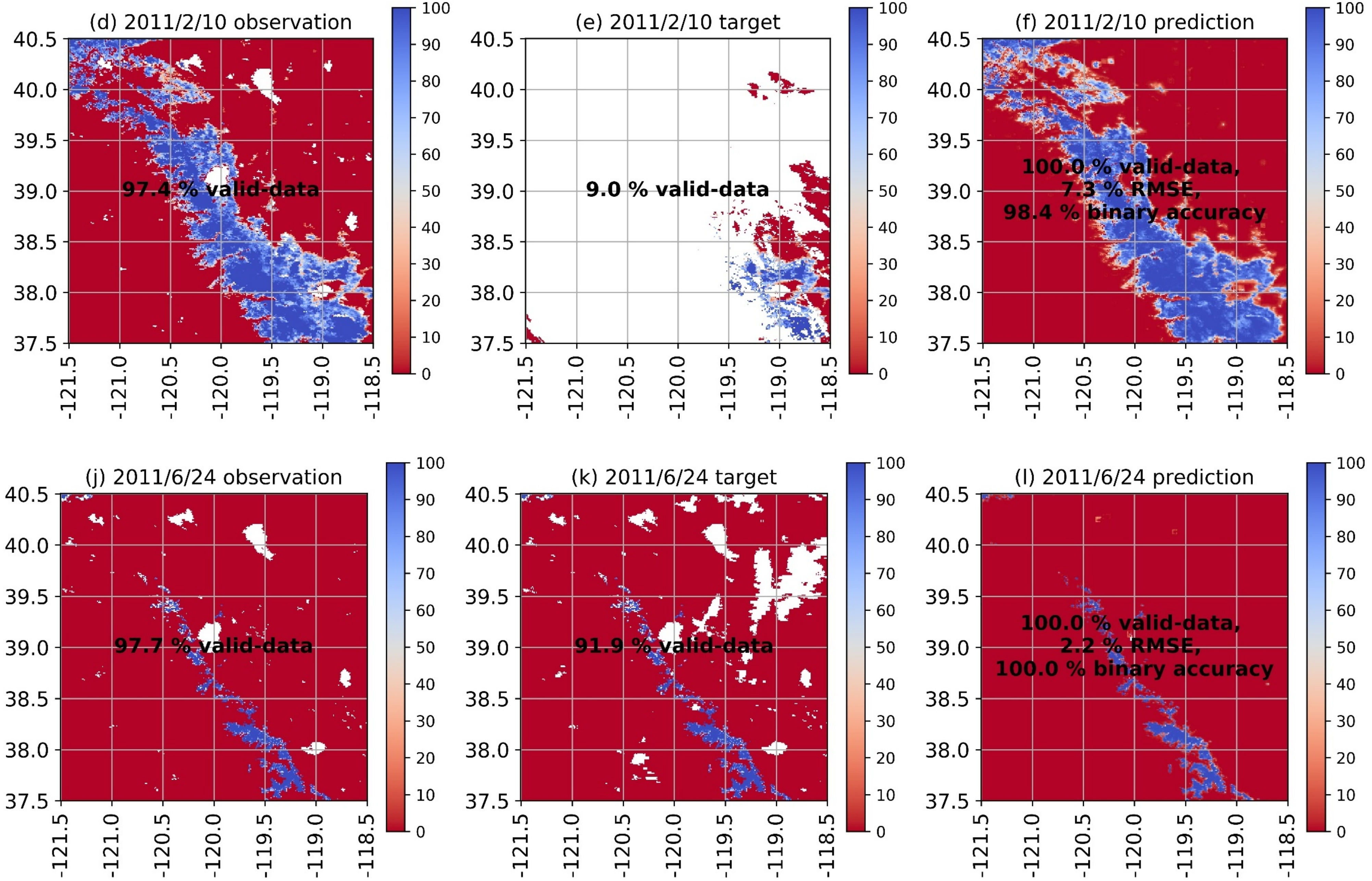
Batch RMSE [%]

- Training: 5.7
- Dev : 5.1
- Test : 5.3

Sample day's gap-filling w.r.t input and target



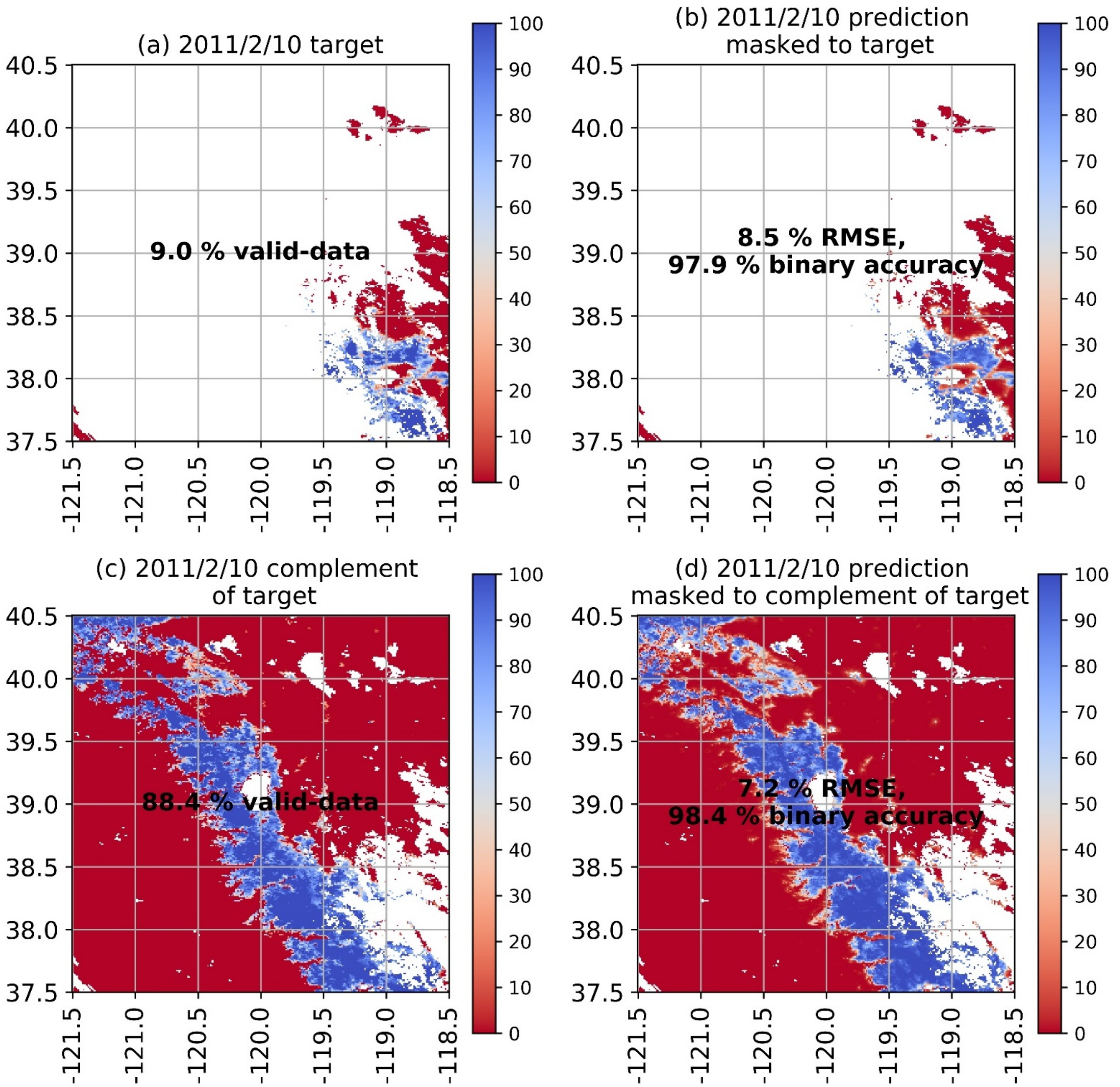
Sample winter / summer days' prediction after further masked training



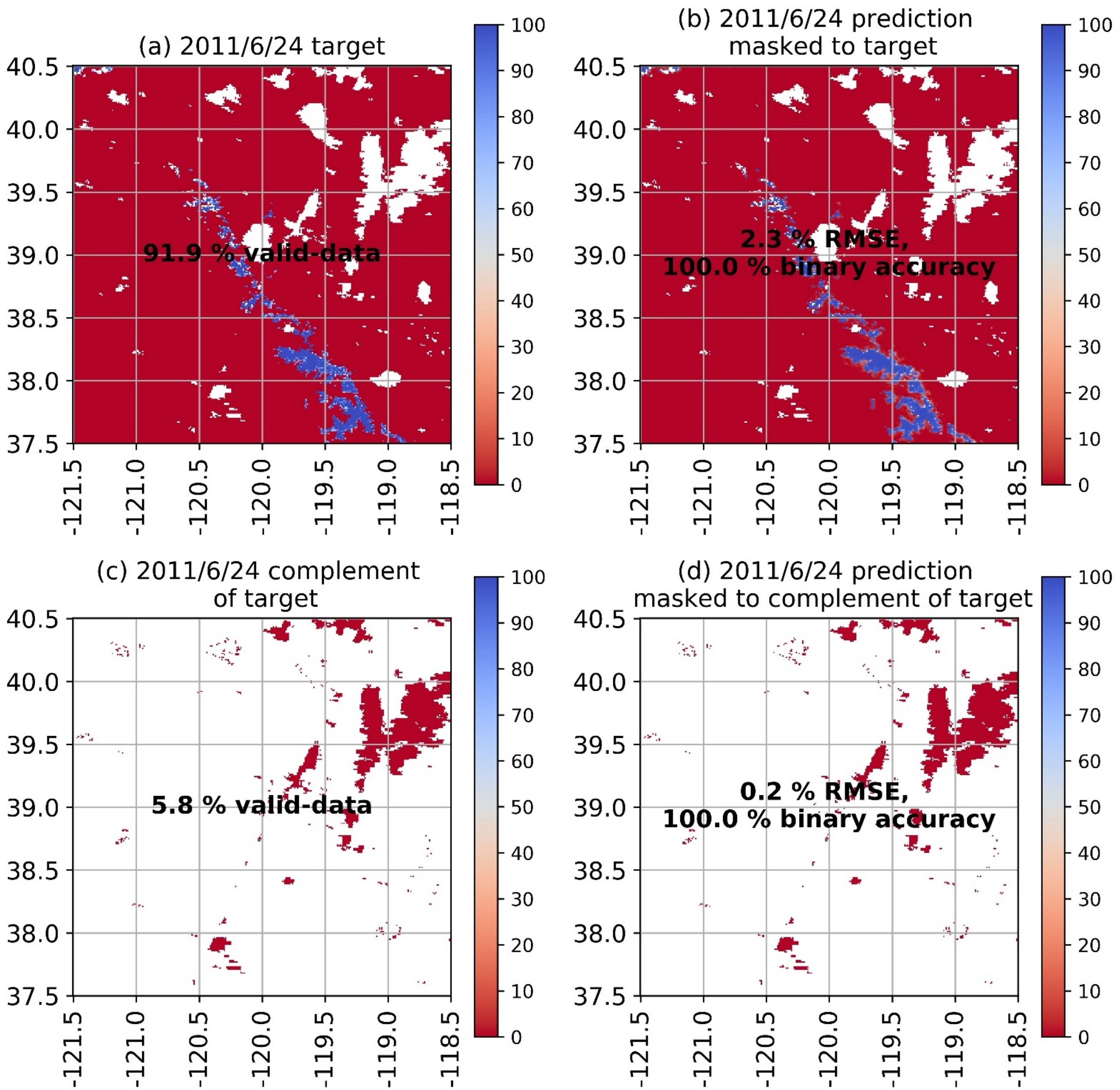
Batch RMSE [%]

- Training: 5.7
- Dev : 5.5
- Test : 5.6

Sample winter day after further masked training: spatial components



Sample summer day after further masked training: spatial components



Discussion and ongoing work

- The simple SRCNN shows good skill in estimating SCF with RMSE estimates around or below 10%
- Preparation of Keras layers that enable using other types of convolutional network architectures
    - Parallel SRCNN stacking (ResNet-style)
    - Encoder-decoder type: e.g., U-Net
- MENSA data processing environment will include for hydrology:
  - Applications (e.g., super-resolution, regression, classification, segmentation)
  - Data analytics