Method to Identify a Uniform Spatial Resolution for Multiple Raster Datasets

Sarah Becker, PhD

Physical Scientist, US. Army Corps of Engineers
Engineer Research and Development Center

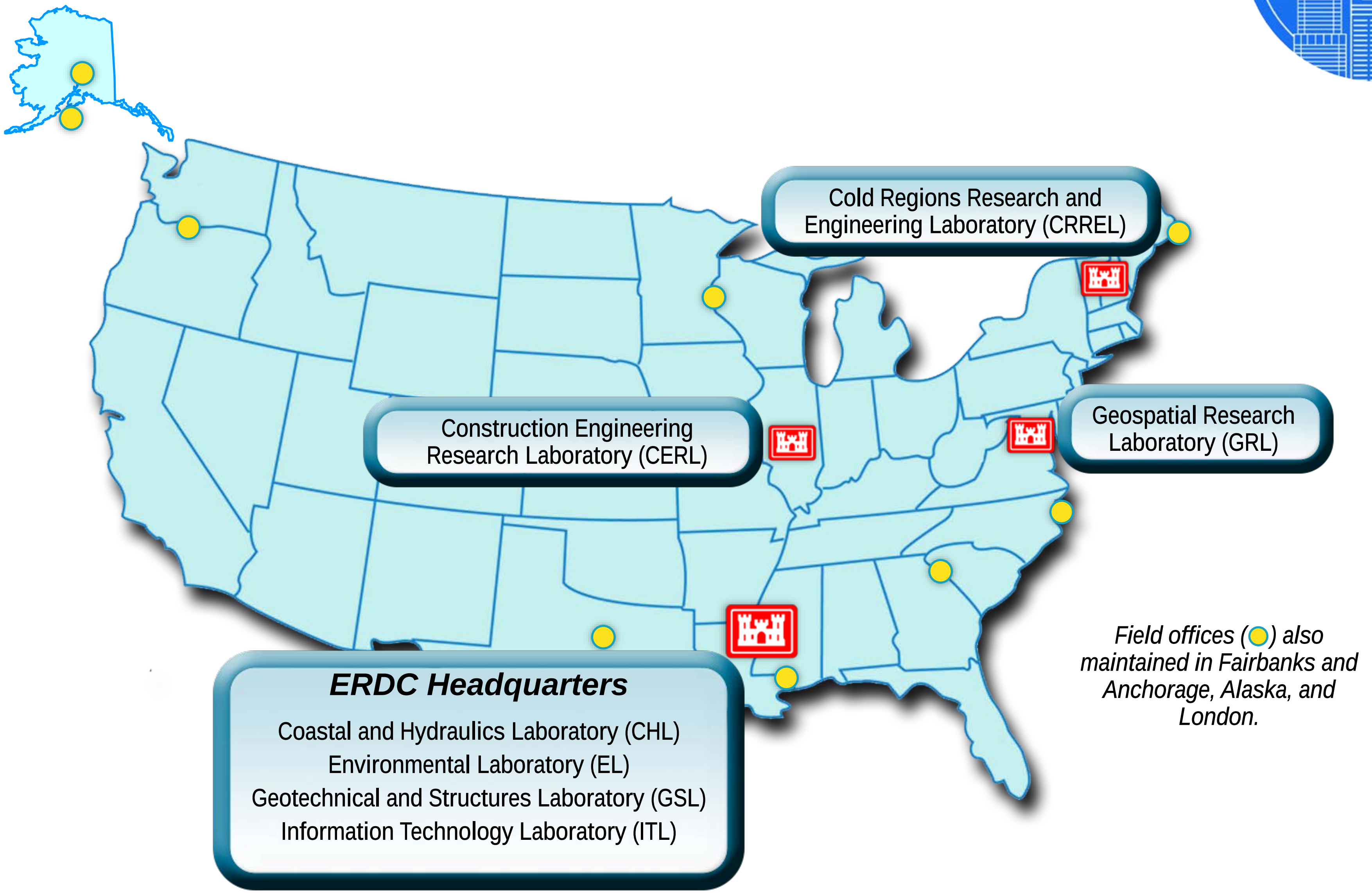
Sarah.J.Becker@usace.army.mil

EAS
ENVI ANALYTICS SYMPOSIUM

Agenda

- U.S. Army Corps of Engineers – Engineer Research and Development Center (ERDC) overview
- Research problem
- Research question
- Definition of key statistic used in methodology
- Methodology
- Results
- Moving forward

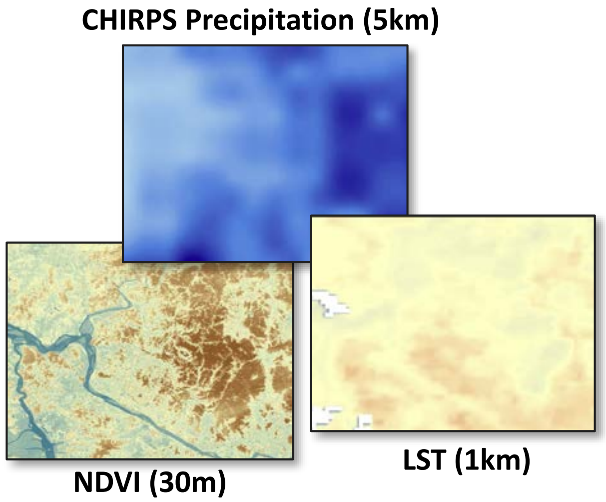
U.S. Army Corps of Engineers
Engineer Research and Development Center (ERDC)



<u>Research problem</u>:

Image scientists often have to contend with comparing images of different spatial resolutions

- Researchers encounter this issue when performing:
  - Species distribution modeling
  - Habitat modeling
  - Route modeling
  - Mosquito/malaria migration mapping
  - Environmental quality studies
  - Land cover mapping



- One example: the *maximum entropy model* is a specific model that requires all raster data to be the same pixel size, or spatial resolution.

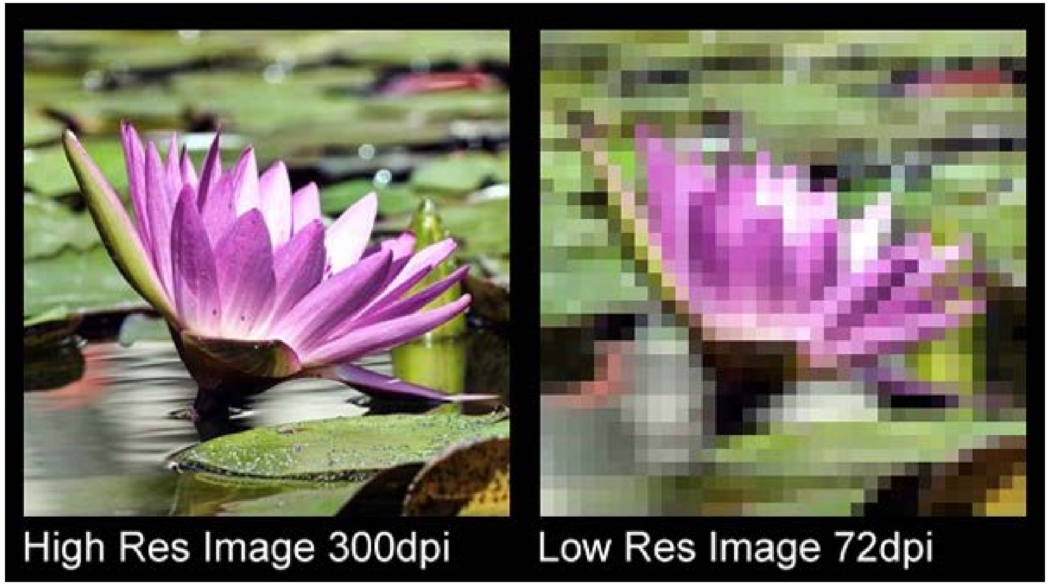
<u>Research problem</u>:

Image scientists often have to contend with comparing images of different spatial resolutions

- Solution: Resample images to a common spatial resolution, or pixel size, when necessary to perform certain types of statistical modeling.

→ How do we determine an optimal pixel size for a <u>single image</u>?

  - Determining a best pixel size is often performed ad hoc. How small of a pixel is small enough? How big is too big?

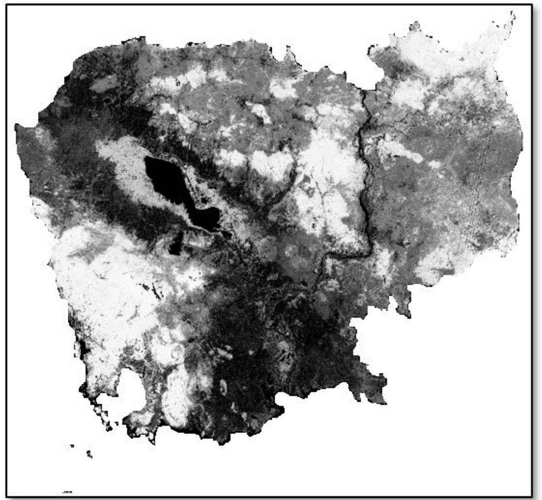


Comparison of high and low resolution renditions of <u>same</u> image.
What are the advantages and disadvantages of each rendition?

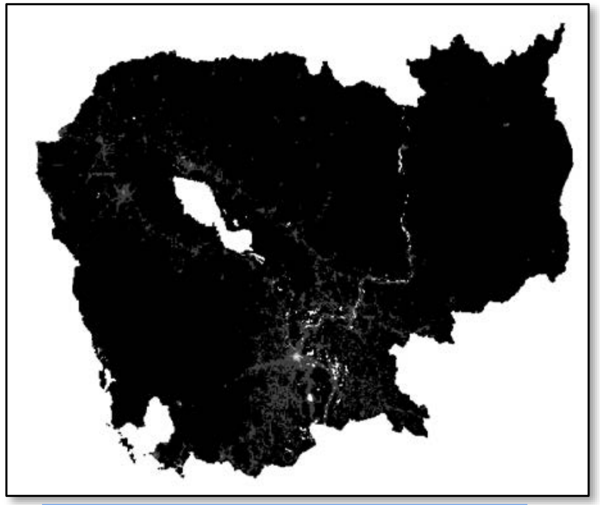
Research problem:
Image scientists often have to contend with comparing images of different spatial resolutions

- Solution: Resample images to a common pixel size (if necessary to perform certain types of statistical modeling).

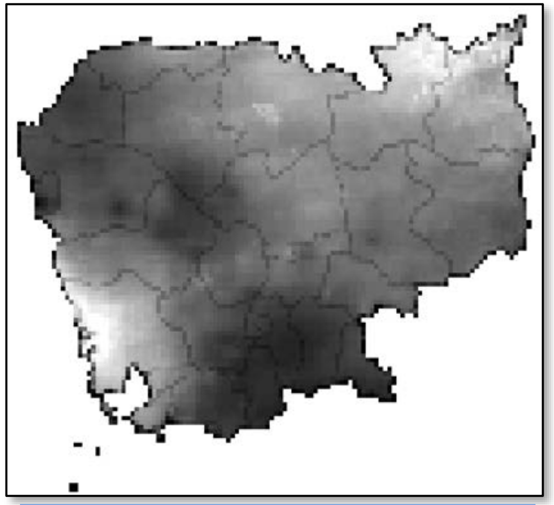
→ When working with different types of imagery of the same location, how do we determine a best uniform pixel size?



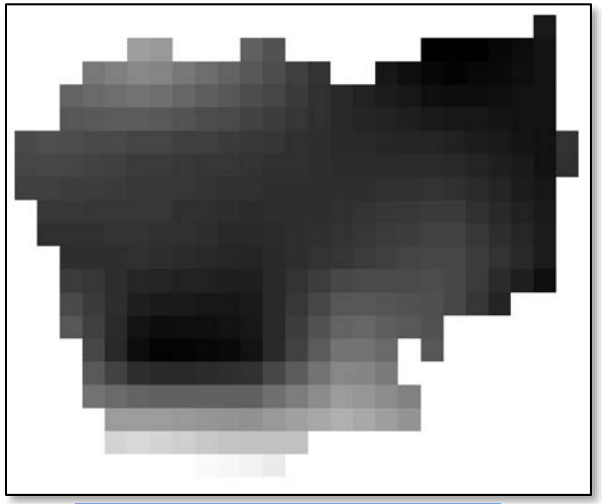
Cambodia percent tree cover data, 30 m



Cambodia population data, 1,000 m



Cambodia precipitation data, 4,500 m



Cambodia wind data, 22,450 m

Research question to move beyond the current state of research

- When resampling data from multiple sources, how can we determine the pixel size with the least amount of error?
  - Establishing the ideal pixel size for raster data from a single source is typically accomplished with analysis of semivariograms.
  - **Researchers have not examined how to determine what pixel size to use when merging datasets of different resolutions.**

Key definition: What is a semivariogram and how is it used to identify an optimal pixel size?

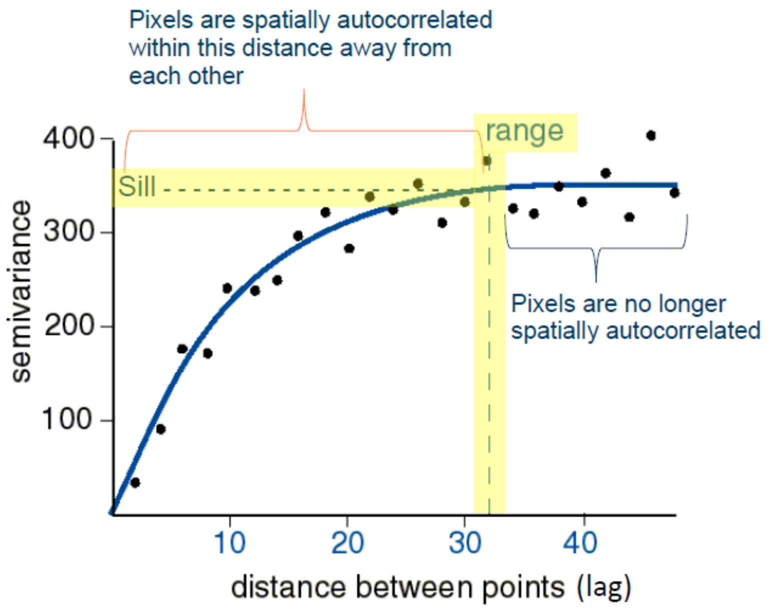
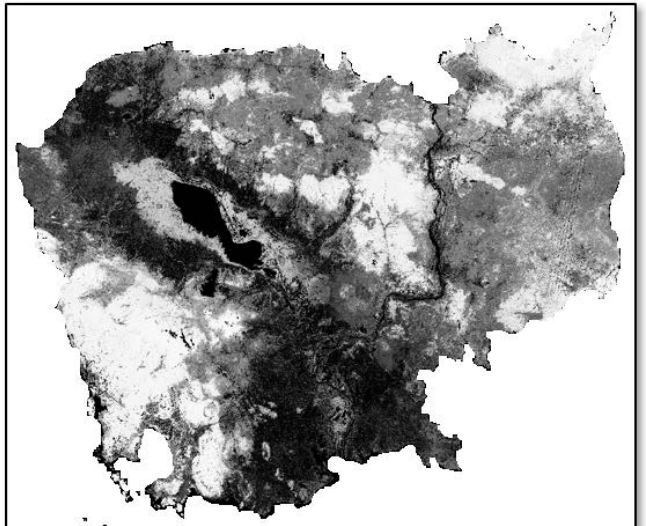


Figure from: http://www.edc.uri.edu/nrs/classes/nrs409509/Lectures/8Models/semivariogram.gif



Cambodia tree data

1. Bottom line: semivariograms identify an ideal pixel size while preserving spatial information and minimizing error.

2. Semivariograms are created for each dataset and optimal pixel sizes are determined for the dataset based on half the range of the semivariograms.

Methodology

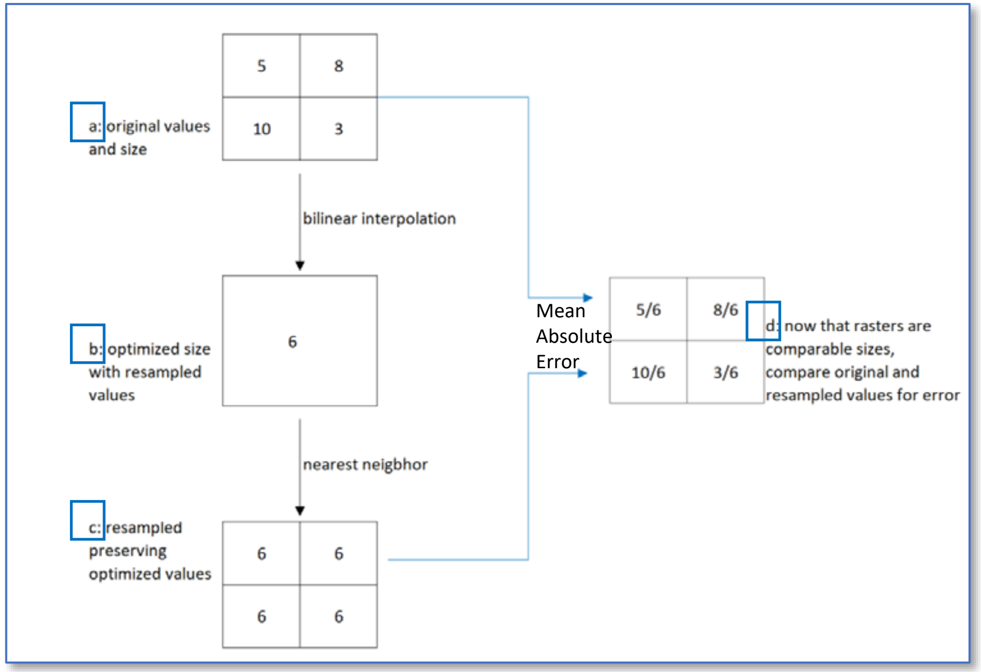
Data: multiple rasters, each one with its own unique ideal pixel size.

Goal: find an ideal pixel size common to all of the rasters.

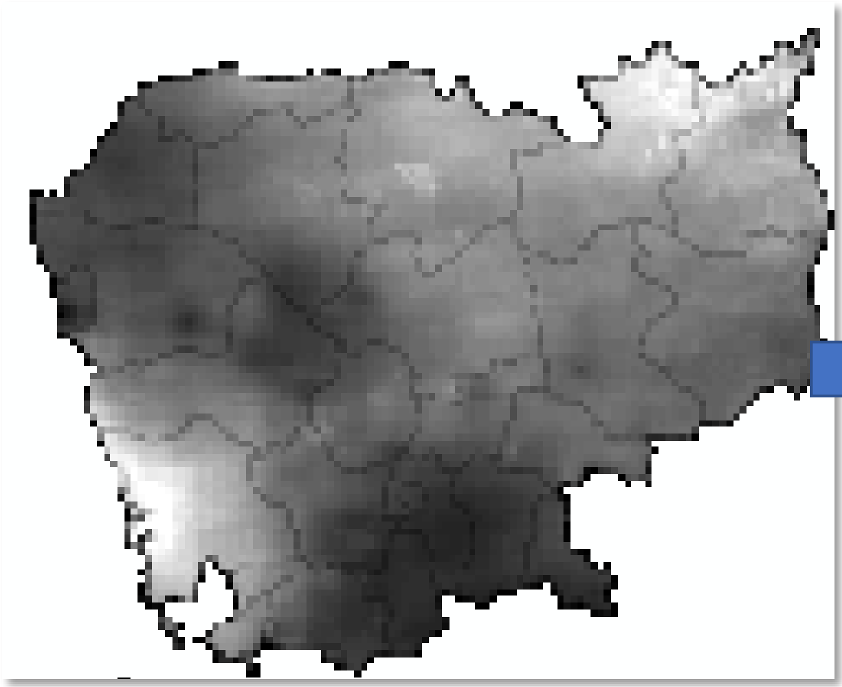
- (**a**) The original datasets are resampled to all of the individual optimal pixel sizes that were calculated from the semivariogram using a (**b**) bilinear interpolation.
- They are then resampled back to their original pixel sizes using a (**c**) nearest neighbor assignment.
- Compare (**a**) to (**c**) using (**d**) Mean Absolute Error (MAE).

$$\mathrm{MAE} = \frac{\sum_{j=1}^{n} |x_j - y_j|}{n}$$

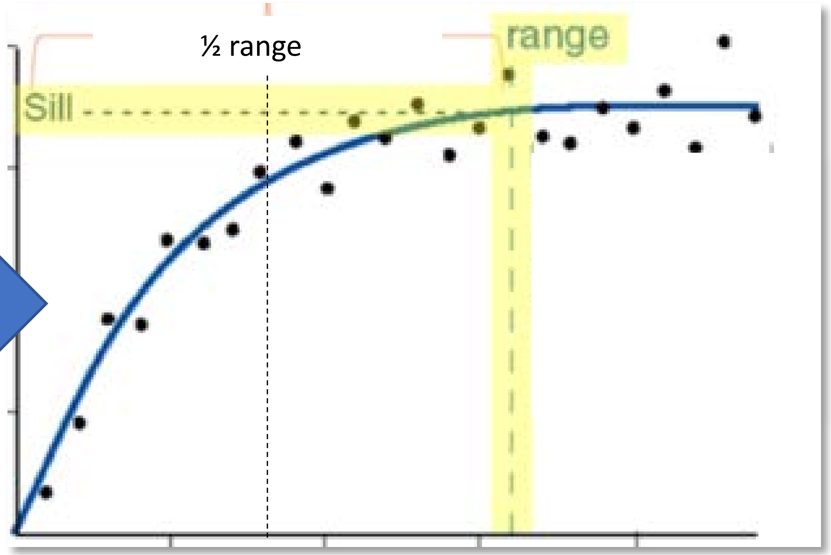
  - The pixel size with the lowest MAE across all datasets is the ideal pixel size if this pixel size is larger than the original pixel sizes.
  - If the ideal pixel size is smaller than the original pixel sizes, resampling each image to that pixel size will introduce spatial autocorrelation. Therefore the pixel size at the next lowest MAE will be considered.



Results

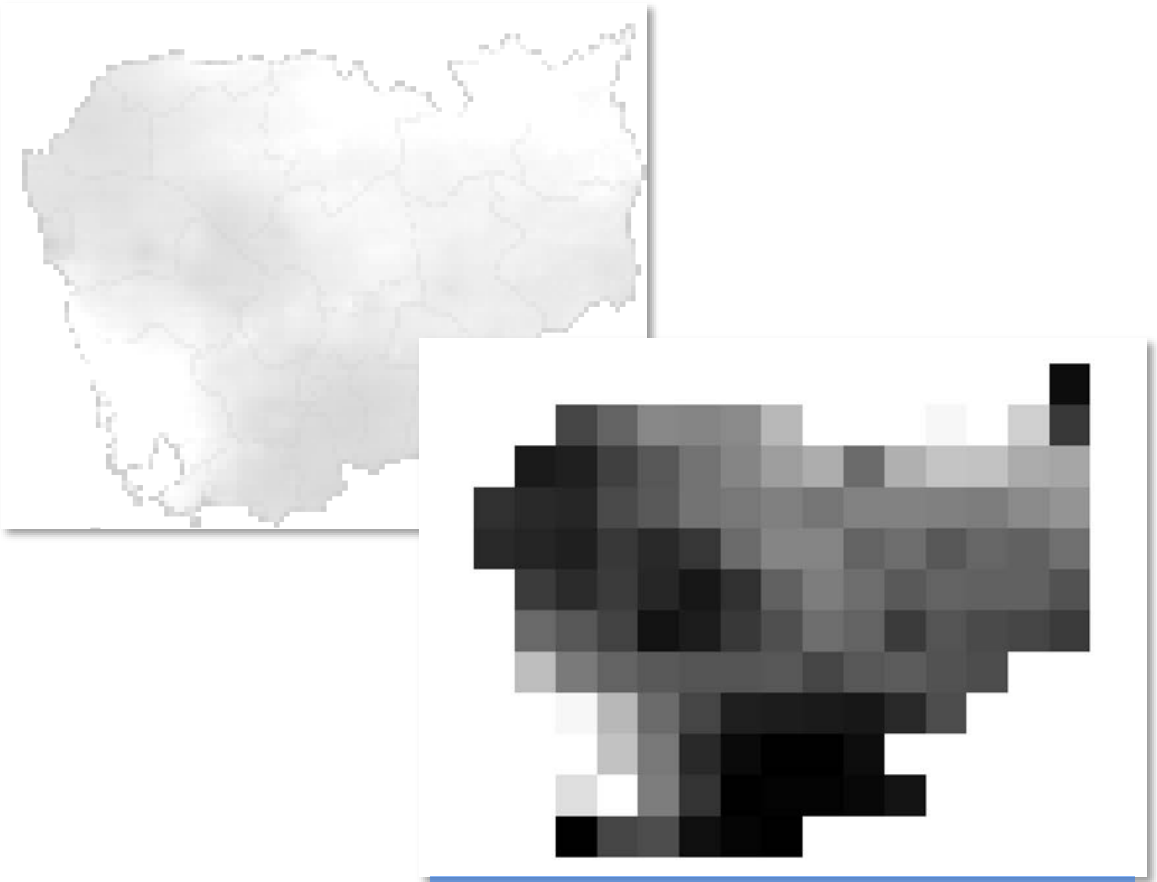


Cambodia precipitation data, 5,000 m



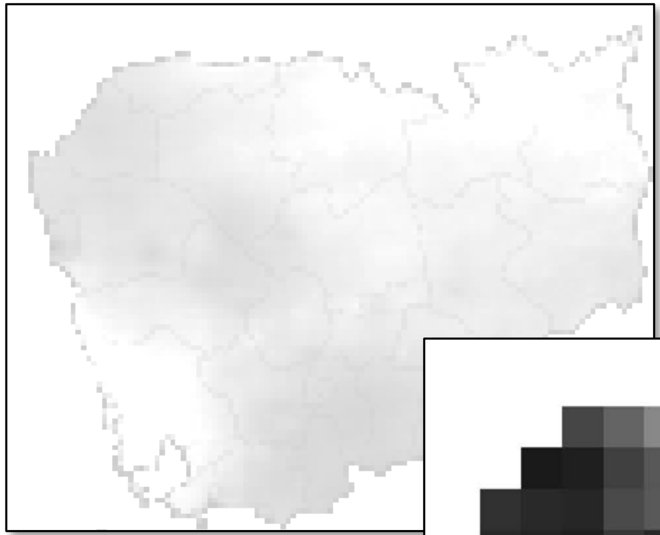
= 96,114 m

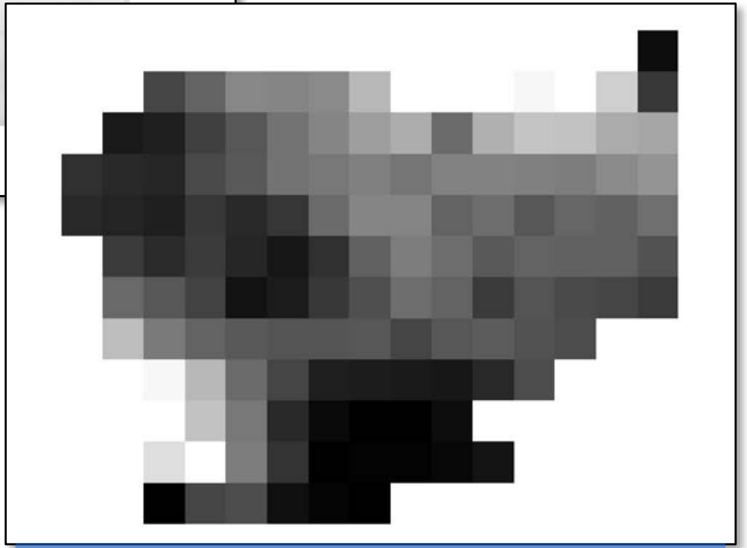
Results



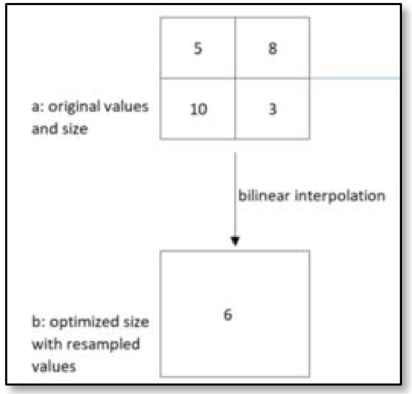
Resampled to 96,114 m using bilinear interpolation

Results

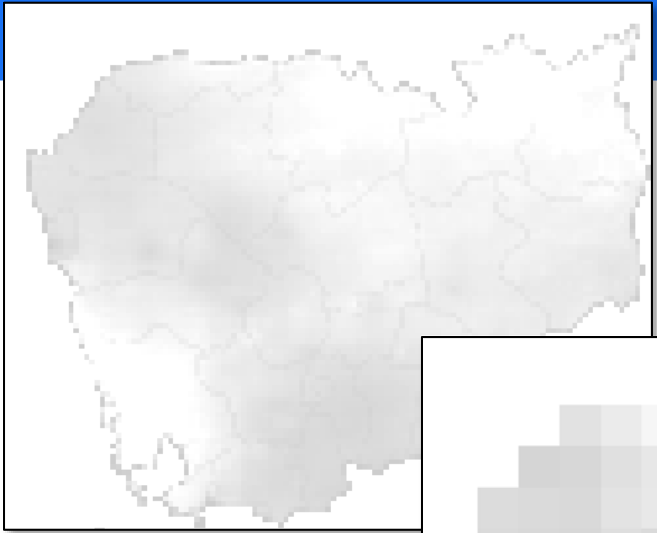


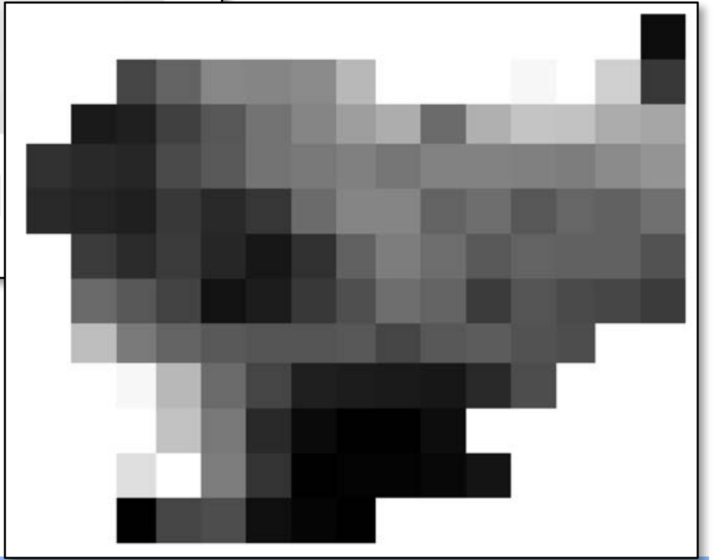


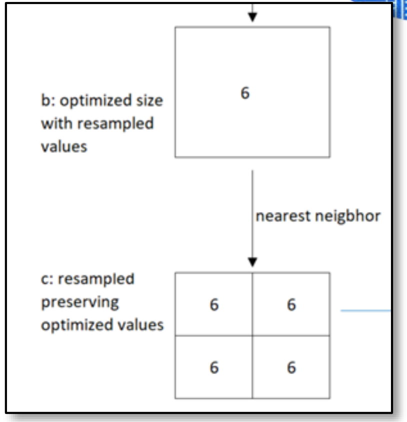
Resampled to 96,114 m using bilinear interpolation



Results

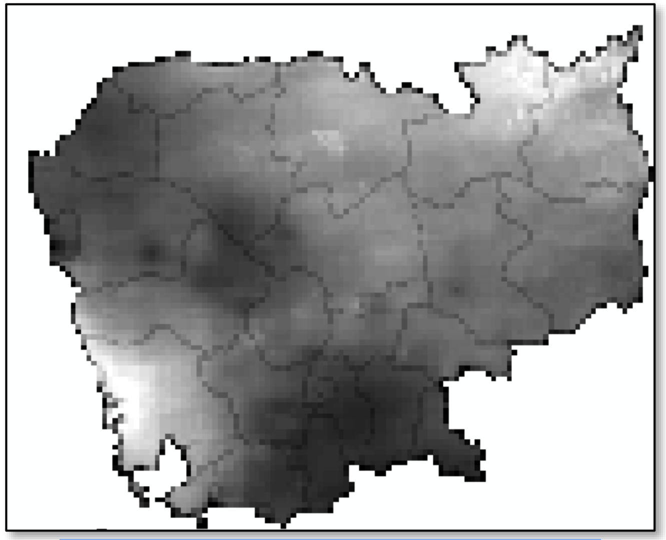




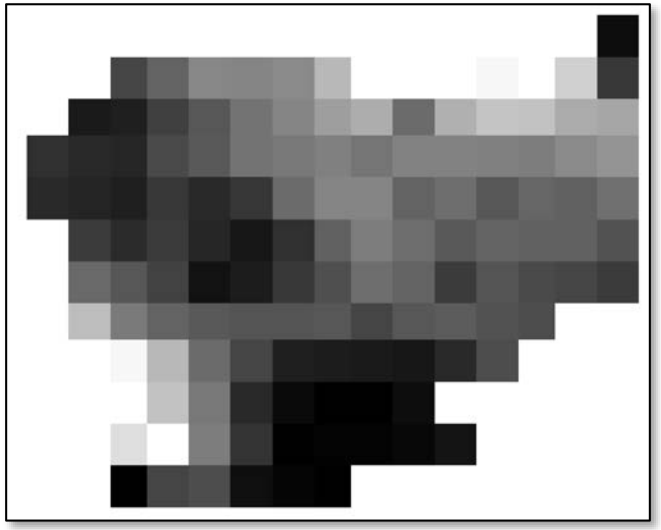


Resampled back to 5,000 m using nearest neighbor

Results



Cambodia precipitation data, 5,000 m



Resampled to 5,000 m using Nearest Neighbor

$$\mathrm{MAE} = \frac{\sum_{j=1}^{n} |x_j - y_j|}{n}$$

$$\mathrm{MAE} = 23.1783$$

Results

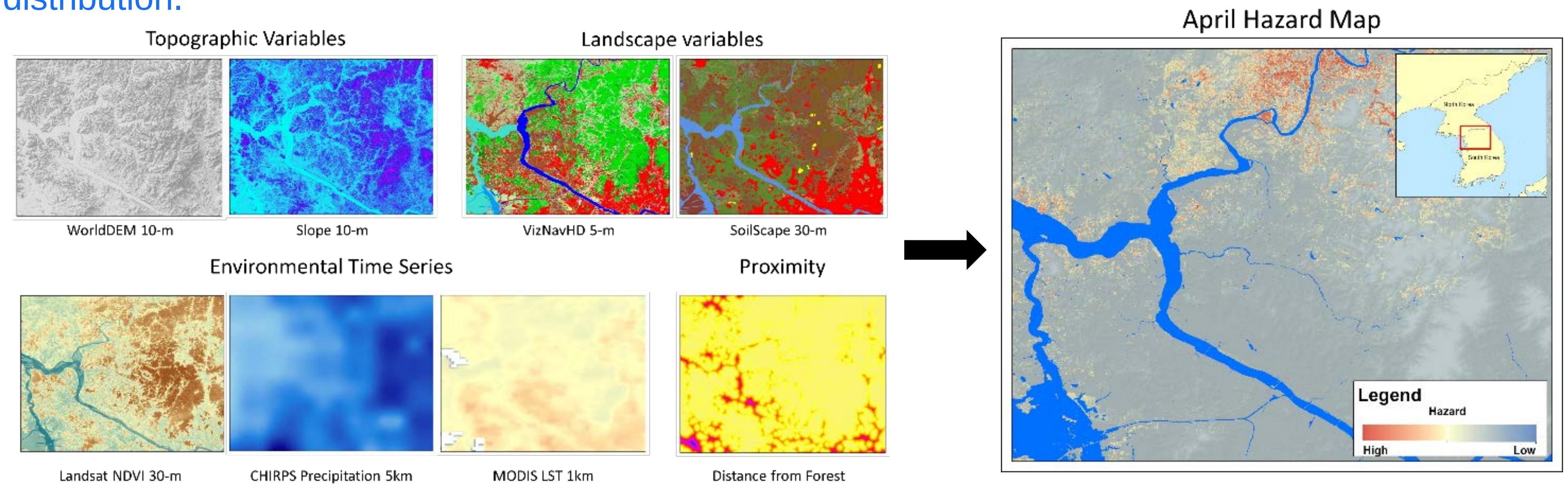
Identify the best uniform pixel size for all of the datasets

		Mean Absolute Error (MAE) at each ideal pixel size			
		470.3 m	20,457.1 m	94,900.6 m	96,114.4 m
Original raster resolution	Trees, 30 m	4.2073	8.6604	12.0908	11.4899
	Population, 1,000 m	0.0342	0.3322	0.3504	0.309
	Precipitation, 5,000 m	0.4525	12.5672	21.8766	23.1783
	Wind, 25,000 m	0.0002	0.0086	2.47E-08	0.0321
	Mean MAE	1.1736	5.3921	8.5795	8.7523

Best pixel size = 20,457.1 m

Moving forward

- Potential use case:
  - Now that the raster datasets are resampled to a matching pixel size, pair the raster data with mosquito presence data at a location.
  - Input the raster and mosquito data into a species distribution model, like the *maximum entropy model*, to determine mosquito distribution.



- Conduct testing using pixel sizes at intervals between smallest and largest pixels, compare errors to errors in our original approach, weigh costs and benefits.
- Explore other statistical methods, like local spatial deviance.

Thank you

- Questions?

- Email: Sarah.J.Becker@usace.army.mil

- Project coauthors: Susan Lyon, Nicole Wayant, and Megan Maloney

- Work package funding: New and Enhanced Tools for Civil and Military Operations (NET-CMO)