# Downscaling WaPOR data with Sentinel-2 NDVI Using PyDMS

26 Jan 2026

Purpose: Downscale WaPOR data to higher resolution for irrigated agriculture.

Inputs: AOI, WaPOR AETI/NPP (dekadal), Sentinel-2 NDVI

Outputs: Downscaled ETa GeoTIFF + residual GeoTIFF

For questions: Suzan Dehati (email: [s.dehati@un-ihe.org](mailto:s.dehati@un-ihe.org))

## WaPOR Data Download (WaPORDL)

* To download dekadal AETI for your study area (use the bbox)
* Notebook Link: <https://github.com/wateraccounting/WaPORMOOC/blob/main/1_WaPOR_download_colab/Download_WaPORv3_Data.ipynb>

## NDVI as high-resolution data:

* NDVI proxies vegetation density/health → strongly linked to transpiration and canopy processes.
* ETa spatial variability is often driven by vegetation patterns at field scale, which NDVI captures well.
* Notebook for downloading Sentinel-2 data: <https://ocw.un-ihe.org/pluginfile.php/181229/mod_folder/content/0/Downloading_sentinel_Data.ipynb?forcedownload=1>

## PyDMS installation:

* GitHub repository: <https://github.com/radosuav/pyDMS/tree/master>
* Easiest way (if GDAL is already installed in your environment):
  + pip install python\_dms
* New notebook: "PyDMS\_Sakha.ipynb"

Notes:

* WaPOR rasters has -9999 as the NoData value, new script will convert it to 0, so they don’t break the downscaling.
* Negative values in NDVI must be masked (set to 0 / NoData), because they can bias the learned NDVI–ET relationship and reduce the downscaled ET values.

## PyDMS Parameters:

1. **movingWindowSize = 150**

**Meaning:** size of the local window (in pixels of the working grid) used to find training samples / compute local relationships.  
**Effect:**

* Bigger window → smoother, more stable model, less noise, but may miss local variation.
* Smaller window → more local detail, but can become noisy or unstable if there aren’t enough good pixels.

1. **cvHomogeneityThreshold = 0.25**

**Meaning:** “how uniform must the high-res signal be inside an aggregated low-res pixel to be considered a reliable training sample.” Usually CV = std/mean within the area.  
**Effect:**

* Lower threshold (e.g., 0.10) → only very homogeneous areas used → fewer samples but higher confidence; may fail in heterogeneous landscapes.
* Higher threshold (e.g., 0.40+) → more samples, but risk mixing land covers → model can get biased.

1. **perLeafLinearRegression = True (Decision Tree mode)**

**Meaning:** decision tree splits feature space (NDVI ranges etc.), then fits a linear regression in each leaf.  
**Effect:**

* Captures non-linear relationships while staying interpretable.
* Usually improves performance vs a single global regression.

1. **movingWindowSize = 150linearRegressionExtrapolationRatio = 0.0**

**Meaning:** controls how much the model is allowed to extrapolate beyond training ranges.  
**Effect:**

* Higher → can produce extreme predictions (including negatives).
* Lower → safer, less extreme outputs.

1. **disaggregatingTemperature = False**

In pyDMS it’s a switch for cases when the high-res variable is LST/temperature and the relationship needs special handling.  
**What to say:**

* “We’re using NDVI, not temperature, so we leave it False.”

lowResGoodQualityFlags

* If you have quality mask raster