Overview

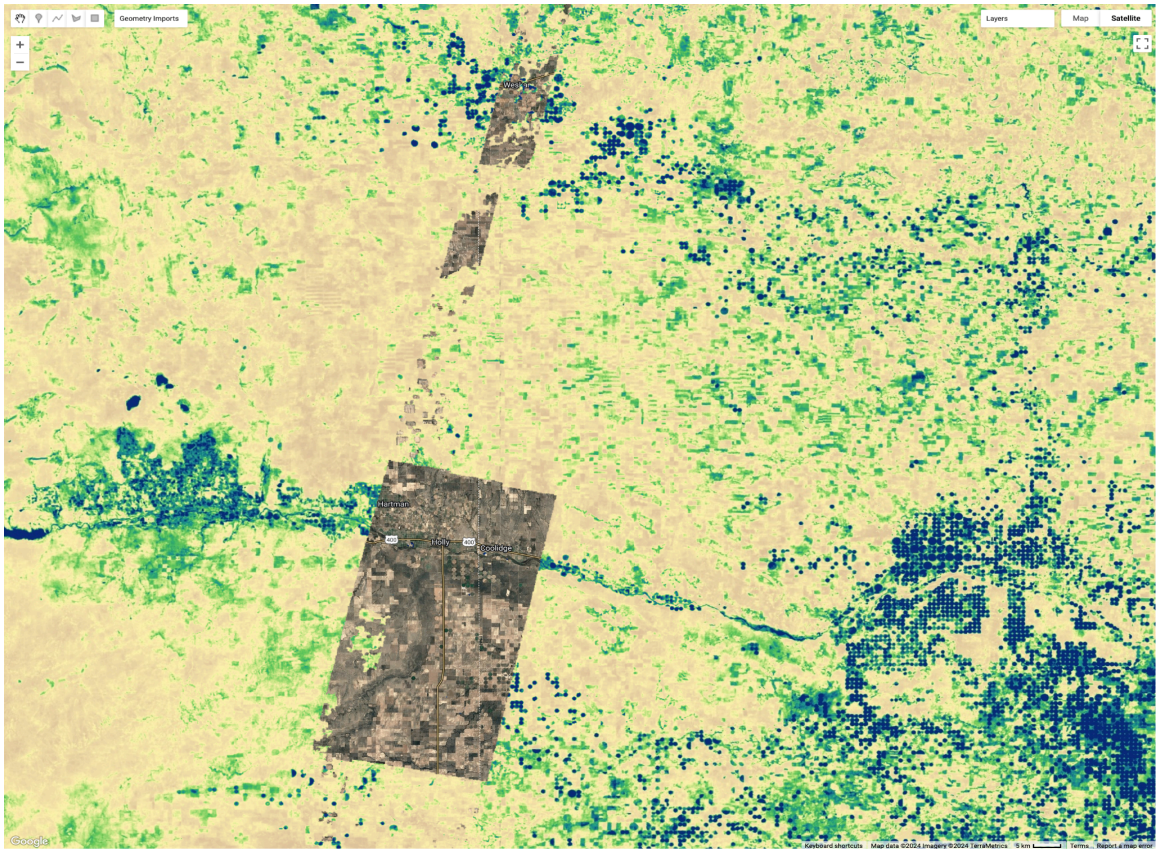
The goal of this page is to document known issues that impact OpenET data. This is an ongoing effort and the information here will be updated regularly as the OpenET team and community identify new issues or make progress in resolving existing ones. The following lists document issues resulting from Landsat retrievals or ET models. For each issue, we provide a brief description, describe the impact on OpenET data, and detail actions the OpenET consortium is taking to address it. If you encounter additional artifacts or issues, please reach out to ***support@openetdata.org***.

Landsat Collection v2.0 Known Issues Impacting OpenET

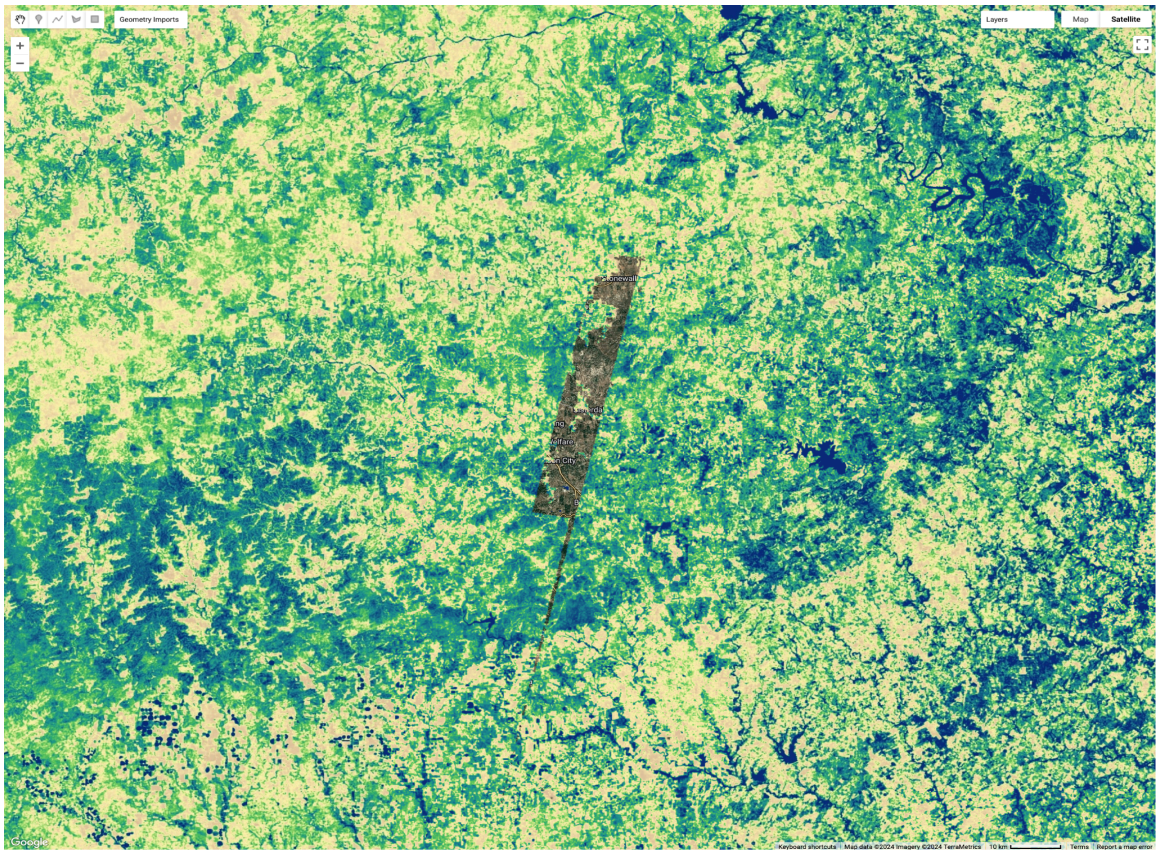
1) Data gaps/holes over Colorado/Kansas and Texas

In certain areas of Colorado, Kansas, and Texas, some ET images include small regions of missing data. These gaps are linked to missing information in the Landsat Collection 2 Level 2 land surface temperature (LST) product, which is caused by data gaps in the ASTER Global Emissivity Dataset (GED). For more details, please refer to the Landsat Collection 2 Known Issues.

Example



Large region of missing data over the Colorado/Kansas border (largest hole is ~30 x 60 km). GEE code editor link that shows the issue and was used to generate the figure: https://code.earthengine.google.com/cb6d419bf1c859350223faea45cbb8fe



Large region of missing data over central Texas (~5 x 45 km)
https://code.earthengine.google.com/9bc4c8e6538822888b2a79e7512add40

Impact

The DisALEXI, GEESEBAL, PTJPL, and SSEBop models are affected by these data gaps and will therefore show no data in the impacted regions. EEMETRIC is not affected because it internally generates a sharpened LST product using filled emissivity data and raw thermal radiance rather than relying on the Level 2 LST product. SIMS is also unaffected since it does not use the LST product in its calculations. As a result, in these areas, the OpenET ensemble is based on data from only one or two models (EEMETRIC and SIMS over agricultural land cover types, and EEMETRIC alone over non-agricultural land cover types).

In addition to the two large affected areas mentioned, smaller clusters of missing data are scattered throughout the central United States. The full extent of these regions can be explored using the ASTER GED emissivity layer here.

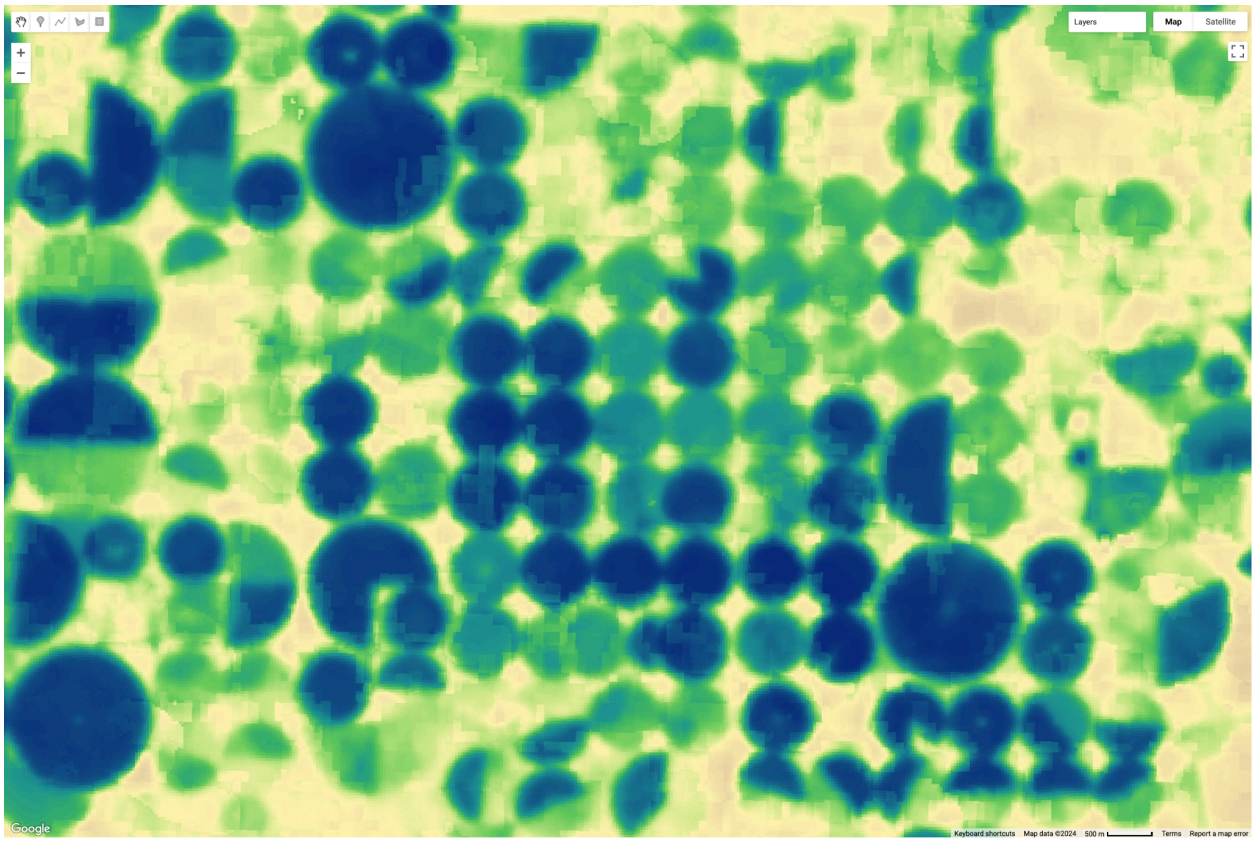
Action

OpenET has developed a solution to address these data gaps, building upon the methodology used in the EEMETRIC model. This fix will be incorporated into the next OpenET data collection release.

2) Blockiness and invalid ET values resulting from anomalies in LST calculation

Some ET images for certain models may exhibit pronounced “blockiness” and invalid ET values. This issue stems from problems with the emissivity data used to generate the Landsat Collection 2 Level 2 LST product. Detailed information on these emissivity issues can be found on the Landsat Collection 2 Known Issues page.

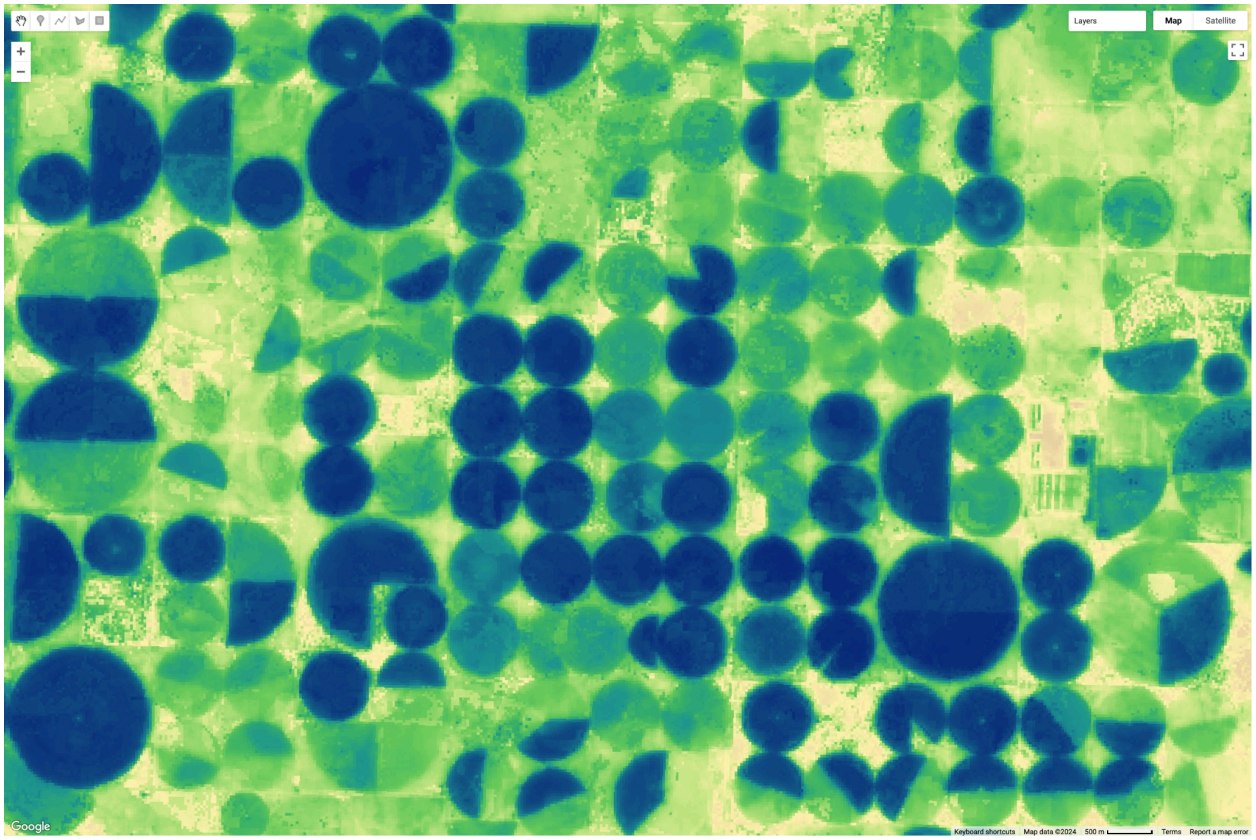
Example



Example of blockiness and artifacts in SSEBop ET for image from May 2023 near Amarillo, TX. https://code.earthengine.google.com/a5acaec172b1858cebe954c1c827ec7b

Impact

These artifacts and blockiness are most evident in the SSEBop and GEESEBAL model outputs because both rely directly on the uncorrected LST data as a primary input. SIMS is unaffected since it does not use LST in its calculations, and PTJPL is only indirectly influenced by LST, resulting in less noticeable impacts. EEMETRIC and DisALEXI are either unaffected or minimally affected because they utilize sharpened LST products that mitigate the impacts from degraded emissivity data. The OpenET ensemble is only minimally impacted, as it integrates data from all six models and thus dilutes the effect of any single model’s input anomalies.



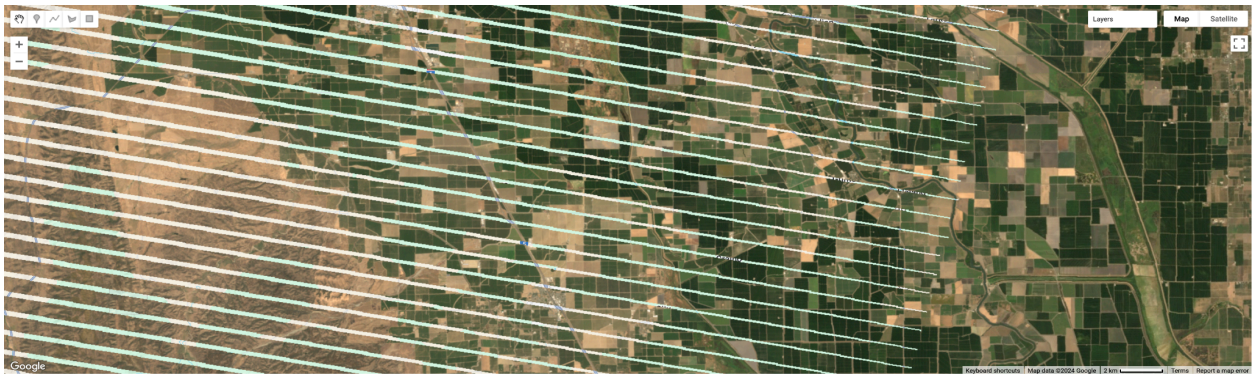
Example of the ensemble ET for the same location and date as the image above. The blockiness and artifacts are still present but not as prominent.
https://code.earthengine.google.com/6d9a046fe582a4a10f0e095f44b7f3fd

Action

OpenET has developed a fix for this issue that will be applied in the next OpenET data collection release.

3) Landsat 7 Scan Line Corrector (SLC) Off Data Gaps

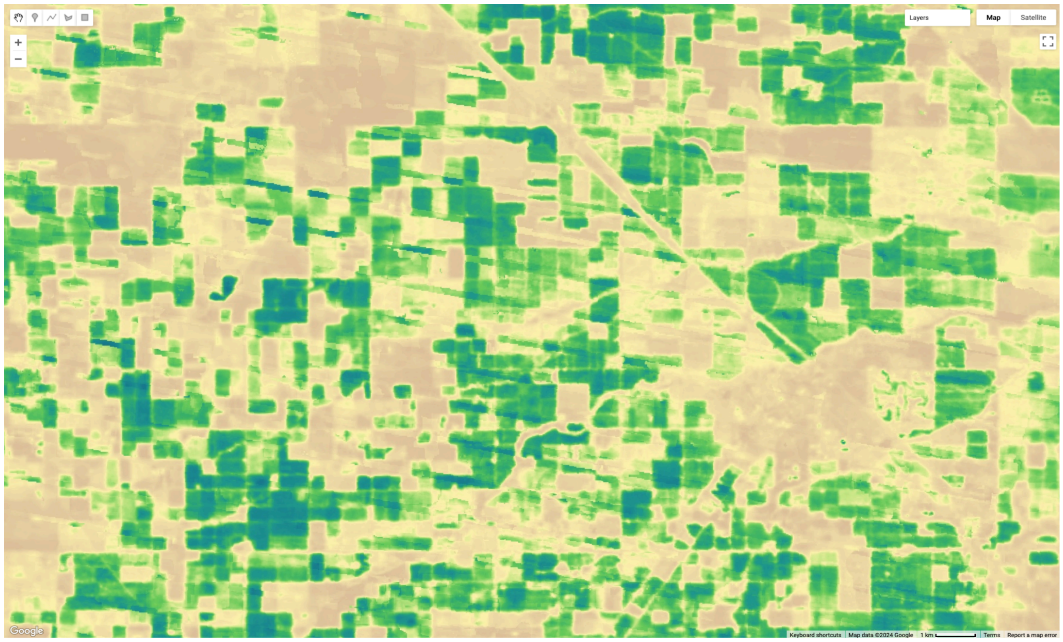
The scan line corrector (SLC) on Landsat 7 failed on May 31st, 2003, resulting in wedge shaped gaps (or “stripes”) in all images collected after that date. The SLC-off images contain data for ~75% of the nominal scene area, with the center of the scene having full coverage while the gaps increase in size toward the edge of the scene.
https://www.usgs.gov/landsat-missions/landsat-7



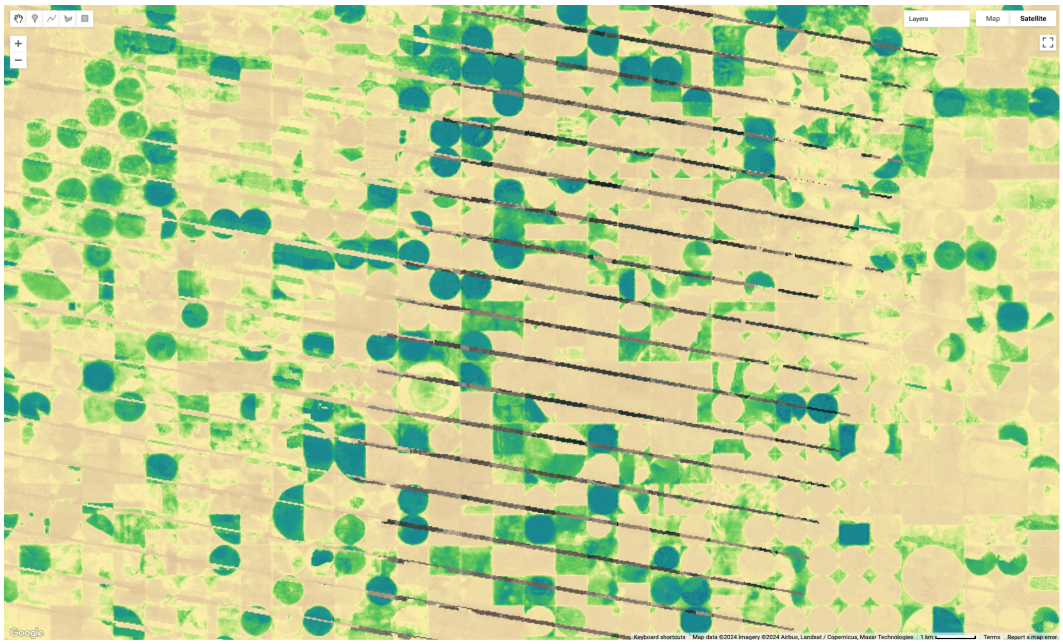
Example of Landsat 7 SLC-off striping in a single overpass image.
https://code.earthengine.google.com/68f03a39a7795ca95f2b07ede0be651b

Impact
All models and the OpenET ensemble are affected by the presence of the Landsat 7 SLC-off data gaps. Artifacts resulting from SLC-induced striping may be present in all monthly ET images from May 2003 to Jan 2022, and may result in a decrease in the number of clear observations and an increase in uncertainty for pixels for which the SLC issue resulted in fewer Landsat observations within each month during this period. The impact of the striping in the Landsat overpass images is moderated in the monthly ET images, since the interpolation process uses all available Landsat observations, including images in neighboring WRS2 “paths”. Also, there was a second Landsat satellite operating during most of this time period providing additional observations to interpolate from.

The data gaps resulting from the SLC failure are especially problematic in scenes for the period from November 2011 through May 2013, after the Landsat 5 mission ended and before Landsat 8 was launched, and there may be insufficient cloud free observations to compute a monthly ET image for some regions. Users are encouraged to consult the metadata layers that provide information on the number of cloud-free scenes that were used to generate monthly data for each pixel during this time period. Pixels with fewer cloud-free observations per month are generally expected to have higher uncertainty from temporal interpolation and users are advised to account for this higher uncertainty in applications of OpenET data that require data during this time period.



Example of SLC-off striping artifacts in the OpenET ensemble monthly ET.
https://code.earthengine.google.com/88d85a0b9445cc4f53175fc1e318c927



Example of SLC-off striping resulting in “no data” gaps for monthly ET images in 2012.
https://code.earthengine.google.com/35a92a7cb7d8ff8c215193429fb956bf

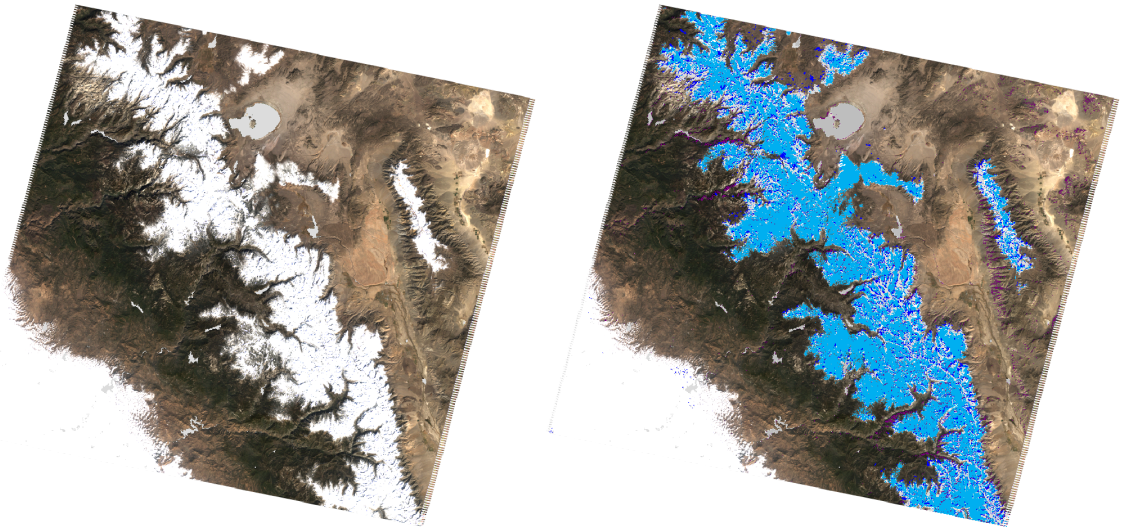
Action

There is currently no plan to perform any additional or separate filling of the Landsat 7 SLC-off data gaps in the next collection release. Users are encouraged to conduct additional evaluation of the OpenET monthly ET data for 2012 for applications where absolute accuracy of the monthly data is an important consideration.

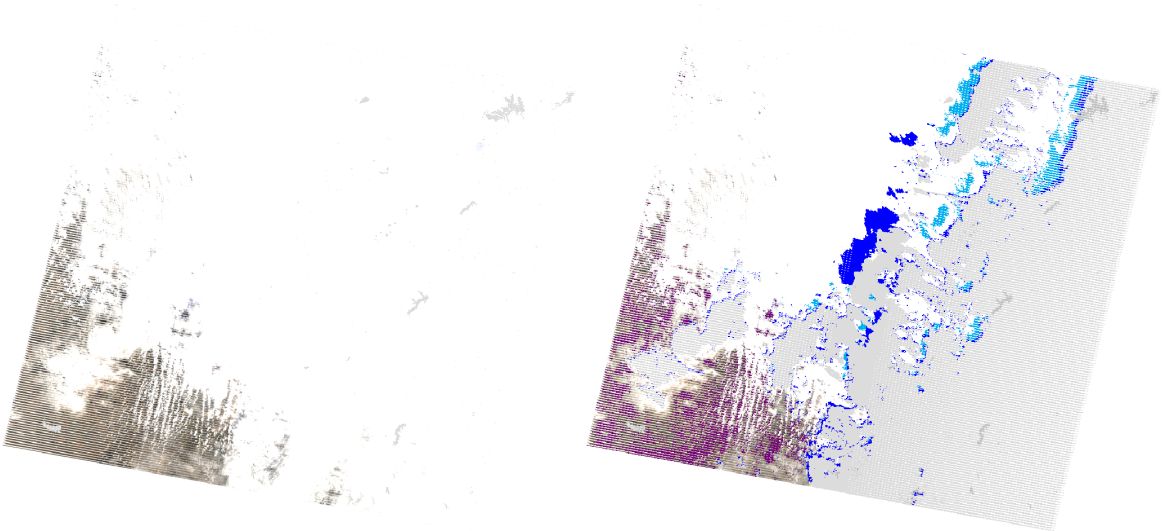
4) Clouds not always flagged as cloud in QA_PIXEL band

In OpenET Collection v2.0, the cloud mask bits from the QA_PIXEL band were used directly to remove pixels identified as cloud, shadow, dilate, cirrus (Landsat 8/9), and snow (for all models except EEMETRIC). If a cloud is not flagged in the QA_PIXEL band, it will not be masked and will be included in model calculations unless the individual model applies its own additional masking.

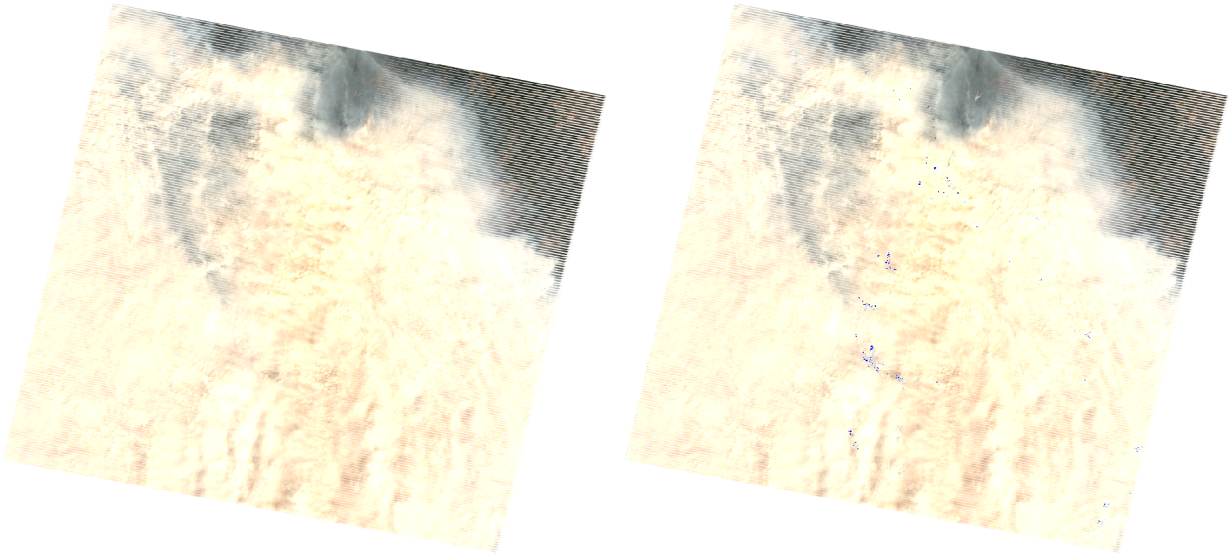
Example
Examples of unmasked clouds shown with the true color Landsat image on the left, and the cloud masked image is shown on the right. The cloud mask images are light blue for snow, dark blue for shadow, gray for cloud, and purple for water.



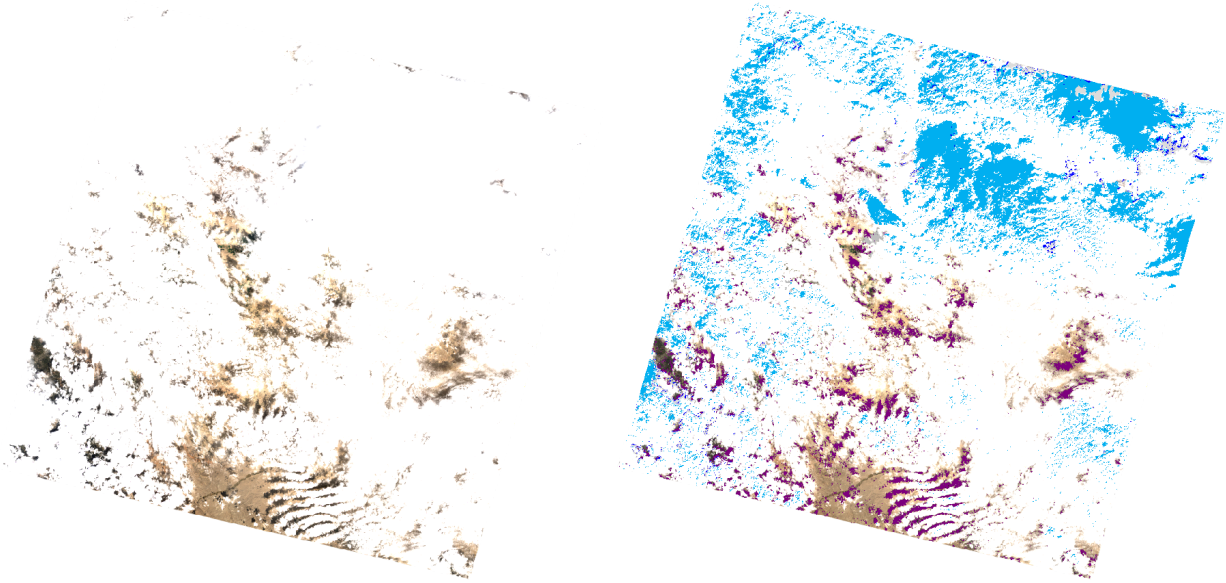
Unmasked clouds/fog in California Central Valley (in bottom left corder of image)
https://code.earthengine.google.com/065620054df817078375149713038556



Unmasked clouds over Texas
https://code.earthengine.google.com/69f84e94678109e7d4041b6918dc6993



Smoke is often not masked correctly
https://code.earthengine.google.com/40e3774b6b282959bd2623908fd4459b



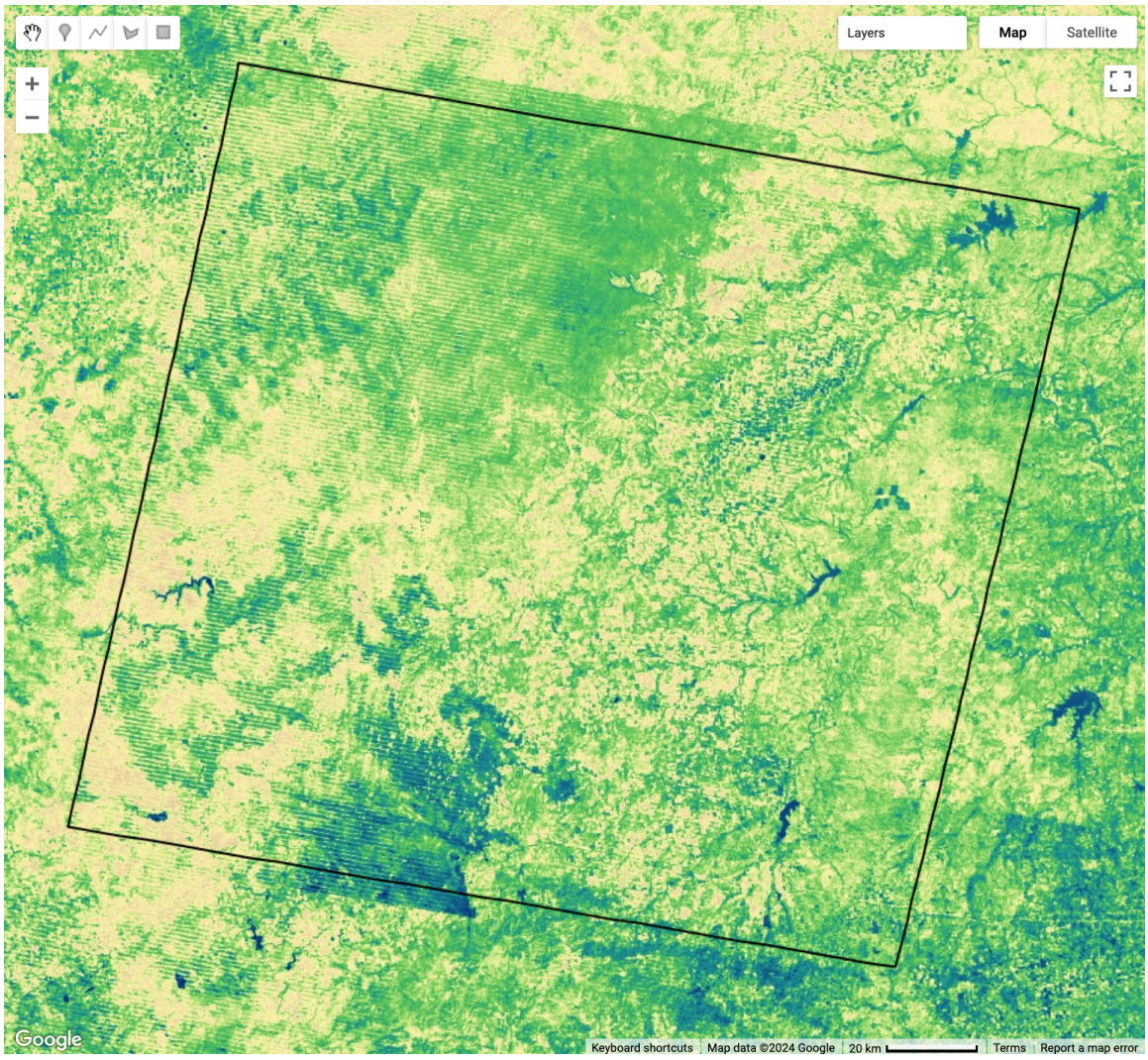
Another example of unmasked clouds
https://code.earthengine.google.com/b784c0aca15343f75d695dca50598e1a

Impact

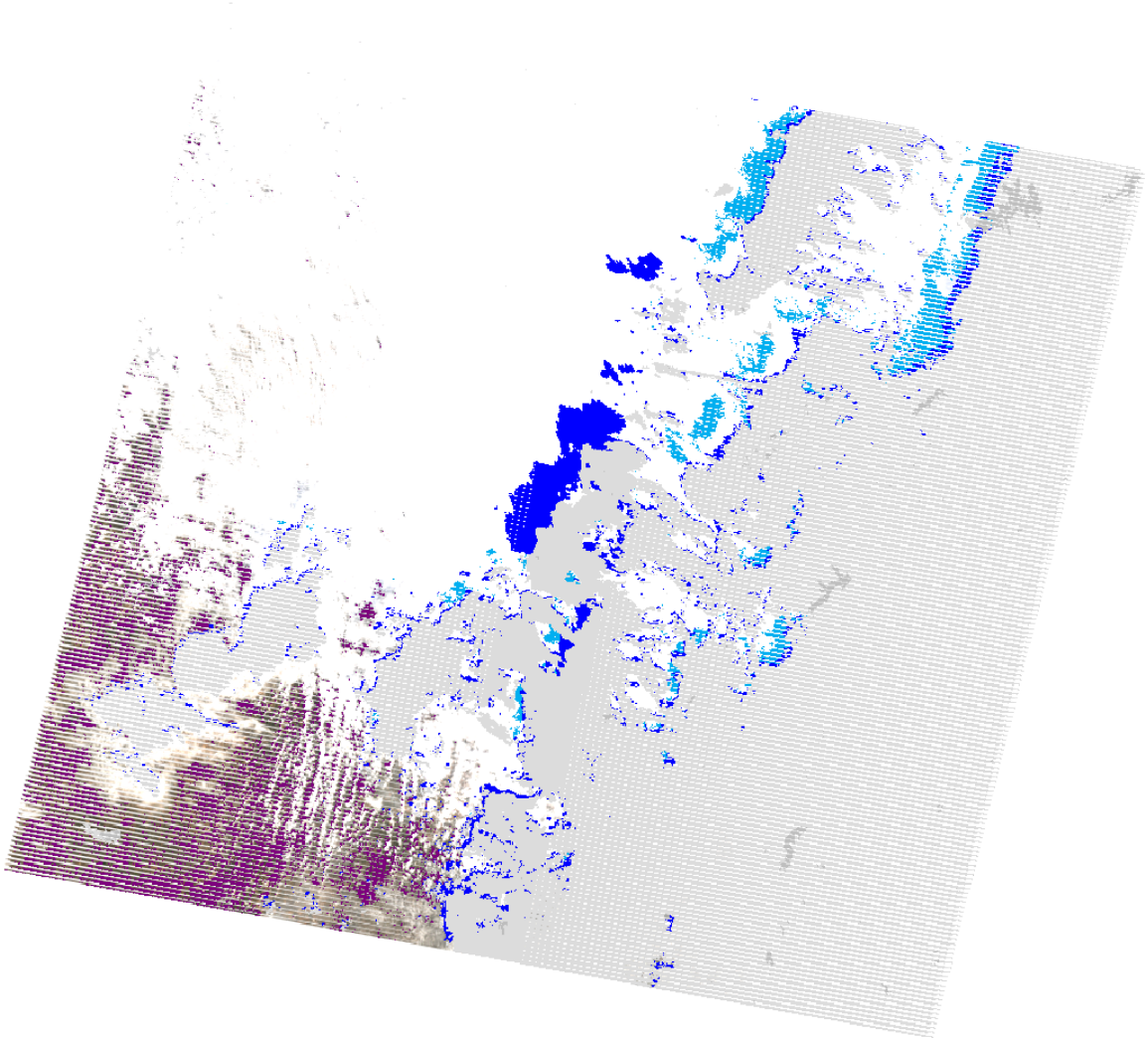
All models and the OpenET ensemble are affected by the presence of unmasked clouds. Although the full magnitude and significance of this impact are not yet well understood, models do not handle these cloud-affected pixels as intended. Unmasked clouds often result in extreme ET data

values—either unusually high or low—because the land surface temperature (LST) values are cold, while the associated reflectance values are frequently saturated and unreliable.

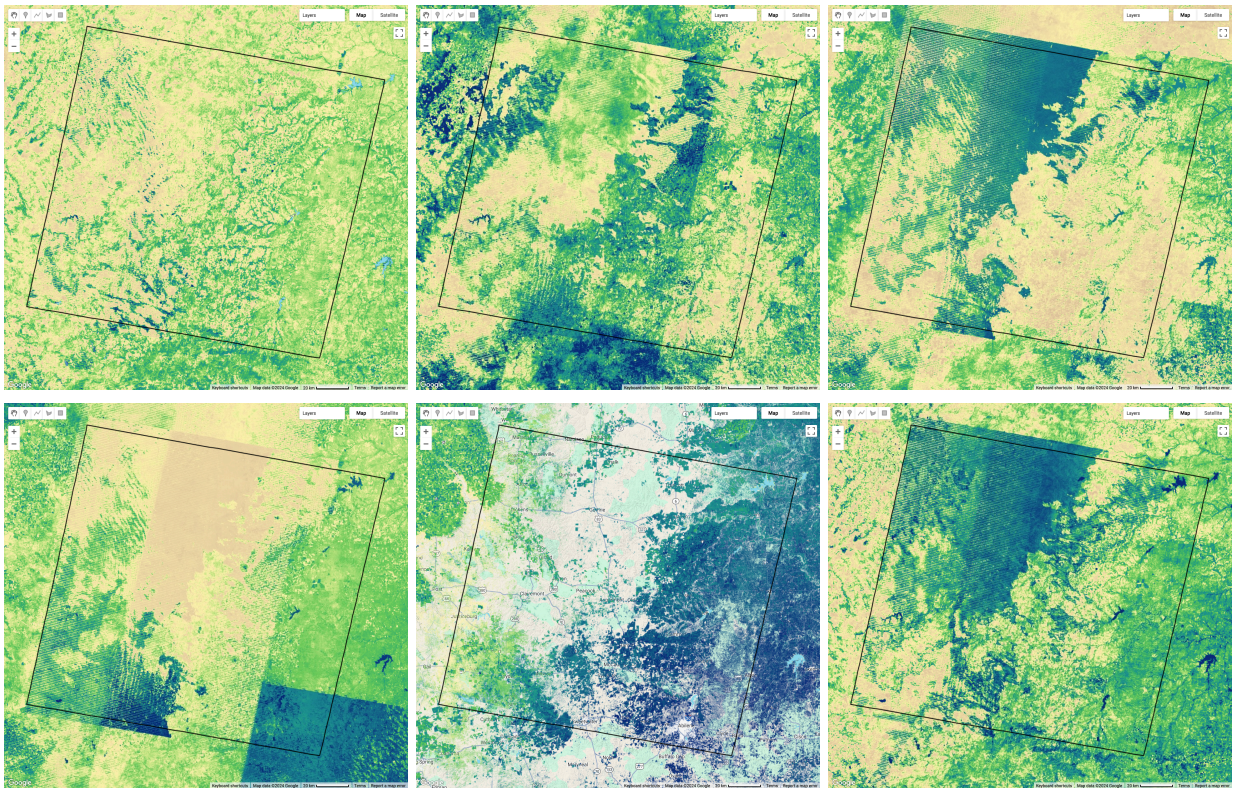
While monthly interpolation, and outlier flagging and removal used in calculation of the ensemble ET value, helps to moderate these extreme values, some of the resulting artifacts may still propagate into the ensemble. For example, in the figure below there are noticeable regions of high ET (darker green) in the OpenET ensemble that appear clipped to the Landsat WRS2 tile boundary. The model ET images for the same date show large differences in the model values that lead to much more variation in the ensemble value.



OpenET Ensemble Monthly ET for Oct. 2009 over central Texas with the WRS2 path 27 row 39 tile boundary shown. https://code.earthengine.google.com/e6a324360dc4d99f4d9576b1b3d95679



Cloud masked true color Landsat image for this same tile collected on Oct 13th, with more than half of the image having unmasked clouds. The CLOUD_COVER_LAND property that is used to screen for cloudy images is computed using the QA_PIXEL band, so this image has a CLOUD_COVER_LAND value of 41% and would not be automatically excluded.
https://code.earthengine.google.com/69f84e94678109e7d4041b6918dc6993



Model ET images for the same month and region for (a) DisALEXI, (b) EEMETRIC, (c) GEESEBAL, (d) PTJPL, (e) SIMS, (f) SSEBOP. GEESEBAL and SSEBop both show very high ET for the unmasked cloud portions, while PTJPL returns ET values near 0.
https://code.earthengine.google.com/e6a324360dc4d99f4d9576b1b3d95679

Action

In OpenET Collection v2.1, enhanced quality control measures are being implemented to address the issue of unmasked clouds. Specifically, Landsat scenes containing extensive unmasked cloud cover will be excluded from use in calculation of ET values. Additionally, for a small subset of scenes that are otherwise mostly clear but contain unmasked clouds, a similar ACCA-based cloud score calculation will be applied to provide supplemental masking and improve data quality.

5) Misclassified cloud and snow pixels in QA_PIXEL band (EEMETRIC Only)

Cloud and snow pixels are frequently misclassified in the QA_PIXEL band. For most models, this is not problematic since both cloud and snow are masked out based on the QA_PIXEL classification. However, EEMETRIC does not mask out snow pixels, making it more susceptible to misinterpretations—either treating cloudy pixels as snow or failing to properly include snow-covered pixels.

Examples

September image over Wyoming with clouds flagged as snow shown in light blue.
https://code.earthengine.google.com/b784c0aca15343f75d695dca50598e1a

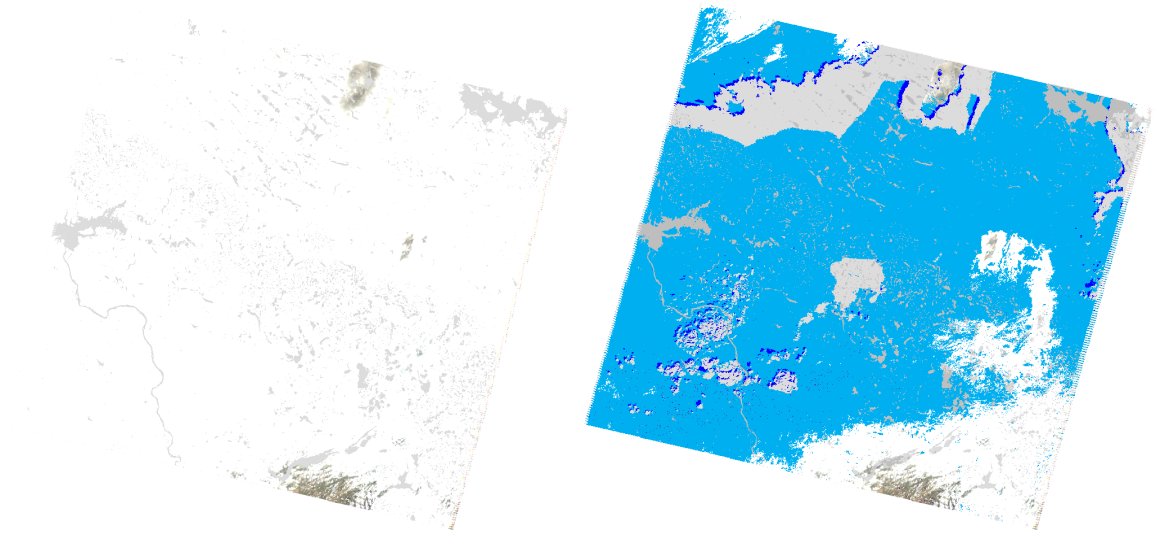
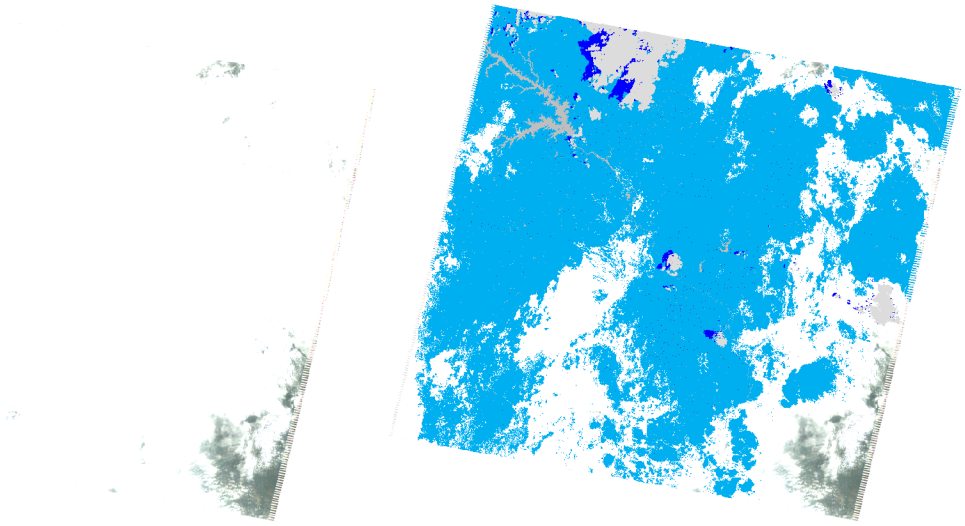


Image from May over North Dakota with clouds flagged as snow shown in light blue.
https://code.earthengine.google.com/c36c555a6af08e785e1fefa31b62f246



Another image from May over Georgia (that was not used in Collection v2.0) but clearly shows clouds that are misclassified as snow (the light blue areas)
https://code.earthengine.google.com/8b285e17198f0d648463a1c82c5687e2

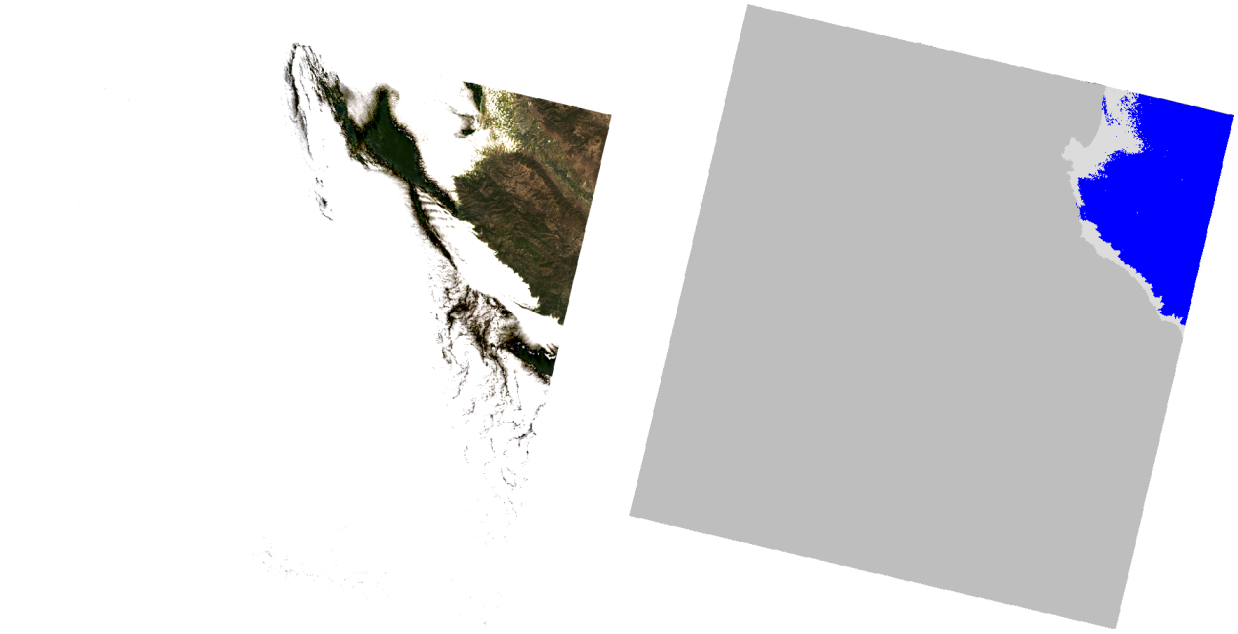
Impact
Only the EEMETRIC model is directly affected by the misclassification of snow and cloud pixels because all other models mask both of these conditions. The exact impact on ET values is currently unknown, but misclassified pixels can produce ET fractions anywhere from 0 to 1, potentially increasing uncertainty. Moreover, the OpenET ensemble may also be influenced, as it can produce a value even when only a single model (such as EEMETRIC) is contributing data (see Issue 8). This may occur when other models report NoData due to snow cover, and eeMETRIC is the only model that is run.

Action
No action is being taken at this time, but this issue may be addressed if EEMETRIC is modified to mask snow.

6) Clear scenes along the coast sometimes flagged as all shadow in QA_PIXEL

Coastal scenes are sometimes flagged as all shadow in the QA_PIXEL band if the ocean is very cloudy and/or bright. All of the models are masking using the shadow bit in the QA_PIXEL band and will exclude these areas or scenes.

Example



Example of clear coastal pixels being flagged as shadow
https://code.earthengine.google.com/480b466c177f87d35e6c6c7b71f98fbc

Impact
For all models and the ensemble, coastal regions may have increased uncertainty as a result of having fewer Landsat scenes to interpolate between, and there may be fewer monthly values overall.

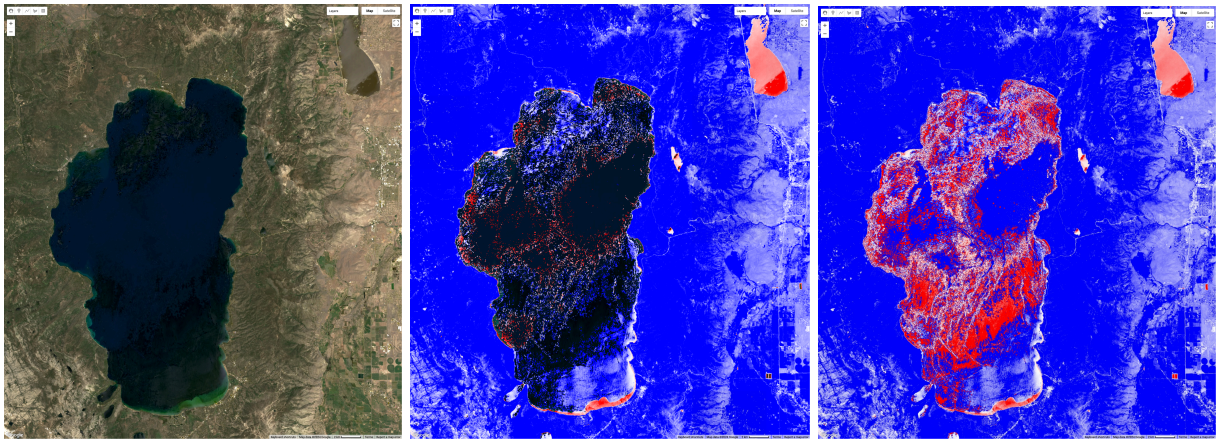
Action
OpenET is identifying Landsat scenes with this issue and may consider manually building the cloud mask for some or all of these images so they can be included in the OpenET data Collection v2.1.

7) Water pixels may be masked out due to negative at-surface reflectance values

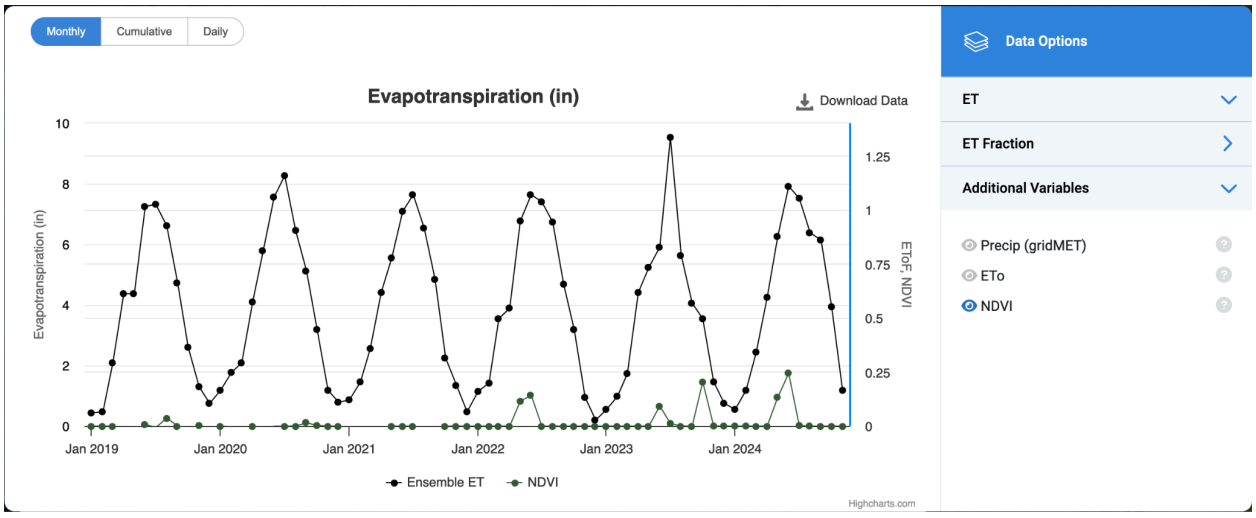
Clear water pixels can sometimes yield negative at-surface reflectance values in Landsat Collection 2 Level 2 images. When using the ee.Image.normalizedDifference() function in Google Earth Engine (GEE) on these values, the function will return masked pixels if the inputs are negative. In contrast, manually computing the normalized difference index (e.g., NDVI) will produce unmasked pixels. However, these values may vary widely, swinging between extremely positive and negative, due to the small (or negative) denominators in the calculation.

Example

The example below shows conditions over Lake Tahoe, where large portions of the lake have very low or negative at-surface reflectance values. When NDVI is computed using the GEE .normalizedDifference() function (middle image), negative values become masked. By manually computing NDVI (image on the right) the resulting pixels are not masked, but the NDVI values can be extremely high or low because of the small values in the denominator.



https://code.earthengine.google.com/1f846432f0180bb3eb52df59fe18d3de



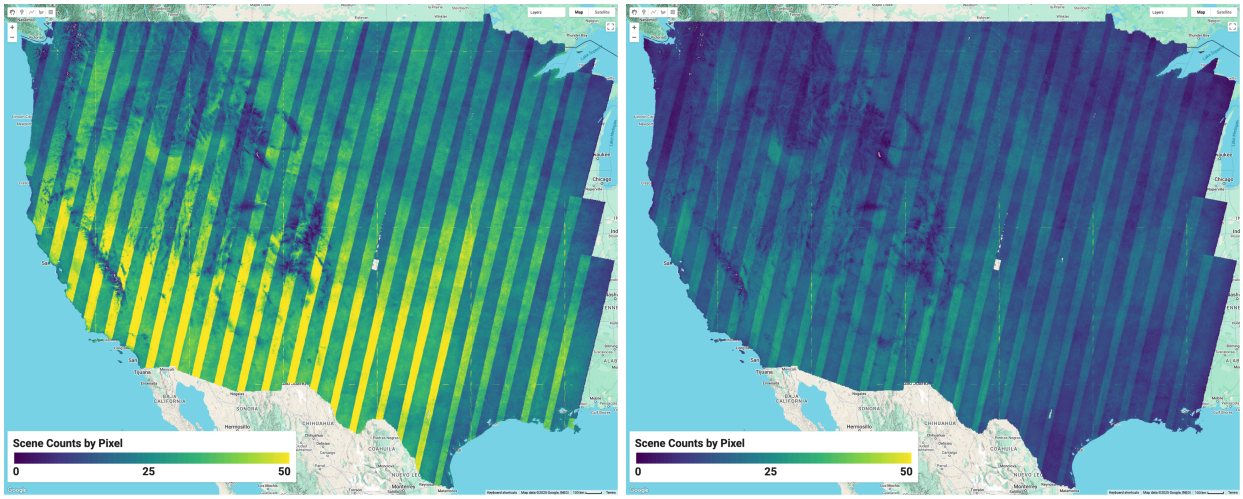
The missing data can be seen in the NDVI time series for Lake Tahoe using the OpenET Data Explorer.

Impact
This issue has been confirmed in PTJPL and SSEBop and may also affect other models, including DisALEXI and EEMETRIC. The extent of the impact is not fully understood, but it may contribute to increased uncertainty in areas with clear water pixels.

Action
In future model versions, water pixels may be identified more reliably by leveraging the QA_PIXEL band water mask, either in place of or as a supplement to normalized difference index methods. If normalized difference indices continue to be used, additional filtering or processing will be implemented to reduce extreme value swings. Improved detection of water pixels will help determine where and when open-water-specific routines or masking should be applied. See Issue 10 for more information on open water estimates.

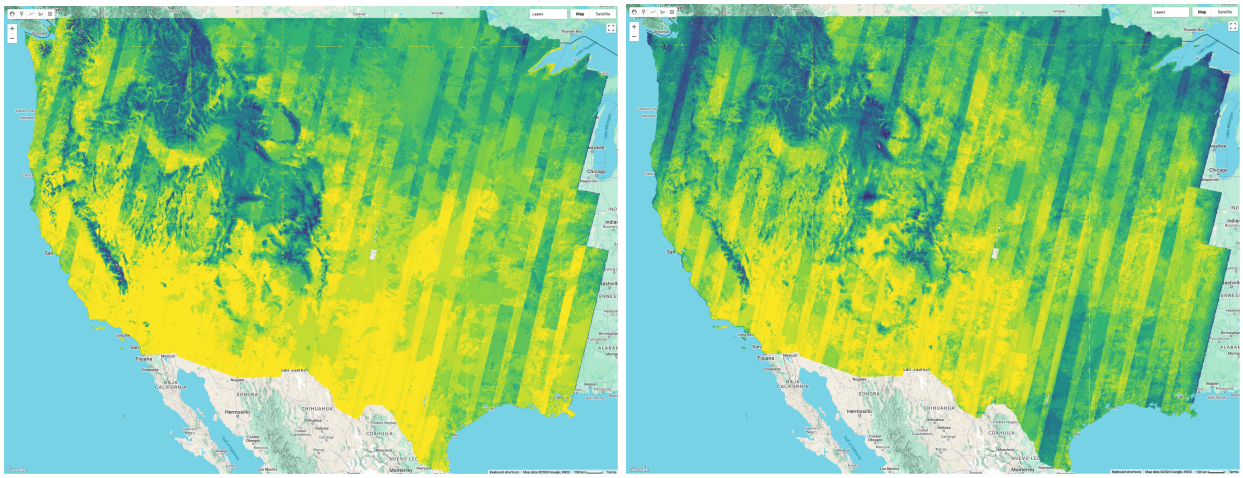
8) Single Landsat coverage for 2012 and early 2013

There have been two Landsat satellites operating concurrently for the full time period of OpenET Collection v2.0, except for the approximately one and half year gap in between the end of Landsat 5 observations in November 2011 and the launch and commissioning of Landsat 8 in May 2013. A single Landsat satellite will revisit the same location on the Earth every 16 days. When there are two Landsats operating, the orbits are offsets such that they will revisit the same location every 8 days. The actual number of useful observations is greatly impacted by clouds, fog, smoke, sensor gaps, etc. and tends to be much lower than the nominal values, except for more arid regions such as the southwestern United States.



Number of cloud free observations in a) 2011 from Landsat 5 and Landsat 7 and b) 2012 from only Landsat 7. The areas of image overlap between the Landsat "paths" can be clearly seen, and may have up to twice as many cloud free observations.
https://code.earthengine.google.com/88b8f17c509c83708c2acbac1405b52a
https://code.earthengine.google.com/3aeae462dabb38929279e98ac529e4ac

Impact
All models and the OpenET ensemble are affected and will have fewer cloud free observations, more data gaps, and potentially higher uncertainty. This issue is magnified by the interpolation issue discussed in Issue 9 below.



Number of SSEBop monthly ET images in 2011 and 2012 after masking all ET values with a count value of 0 to avoid the bad data discussed in Issue 9 below.

Action

Users are encouraged to consult the metadata layers that provide information on the number of cloud-free scenes that were used to generate monthly data for each pixel during this time period. Pixels with less than one cloud-free observation per month are generally expected to have higher uncertainty, and users are advised to account for this higher uncertainty in applications of OpenET data that require daily or monthly data during this time period.

OpenET Collection v2.0 Known Issues

9) Monthly ET values with a “count” band value of 0 may be invalid

Due to a known issue in the interpolation process, some monthly ET values may have been computed using only a single day of data from outside the target month. This scenario can occur when no cloud-free Landsat observations exist within the month itself, resulting in monthly ET values with a “count” band value of 0, and thus the Landsat scenes used to compute ET may not accurately reflect conditions within that month.

Background

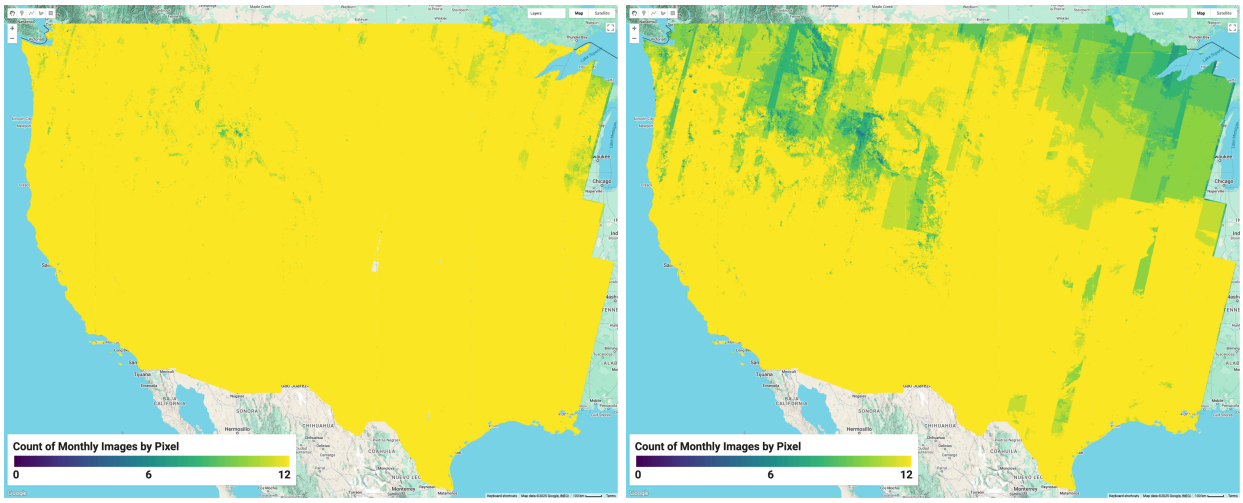
The OpenET interpolation process generates daily ET values for each day in a given month by referencing all available model ET values derived from cloud-free Landsat images within a ±32-day window of the target date. When cloud-free data points exist both before and after the target date, daily ET values are linearly interpolated. This allows for a continuous time series, even if there are no valid Landsat observations during the month in question. For example, if data from January 30 and March 3 are available, ET values for all days in February can still be interpolated.

However, if no valid ET values exist both before and after a given date, the interpolation defaults to using the single closest ET observation within that 32-day window. While this ensures that daily ET values are generated even for months without any direct cloud-free observations, it can lead to a situation where every day in a month is assigned the same ET value derived from a single observation outside that month. Consequently, the monthly ET is summed from daily values that were never truly “interpolated” from observations inside the month, and the “count” band—intended to track the number of cloud-free Landsat observations used in the interpolation—will register zero.

In such cases, the monthly ET values are considered unreliable since they are not truly representative of the target month’s conditions. This issue is being addressed, and future releases will ensure that monthly ET estimates are only calculated when suitable data are available within or adjacent to the month.

Impact

Monthly ET values with a “count” band value of one or more are unaffected by this issue. Monthly ET data values for any model with a count of zero may be affected, but the overall impact on total growing season ET for recent years is likely minimal, since zero-count months typically occur during winter or outside the primary growing season. However, regions and time periods characterized by extensive cloud or snow coverage, or where Landsat coverage is limited, such as single-satellite coverage (Issue 8) or the SLC-off period (Issue 3), may experience a more pronounced effect. In these cases, users may notice more monthly ET values than would be expected and lower-than-expected ET values, as they are generated from fewer than 30 or 31 days of data.



Comparison of the number of monthly SSEBop images in 2021 before and after fixing the interpolation issue. The coverage on the left corresponds to what is currently in Collection v2.0. https://code.earthengine.google.com/d409b06ac5575962f60ace2248860598

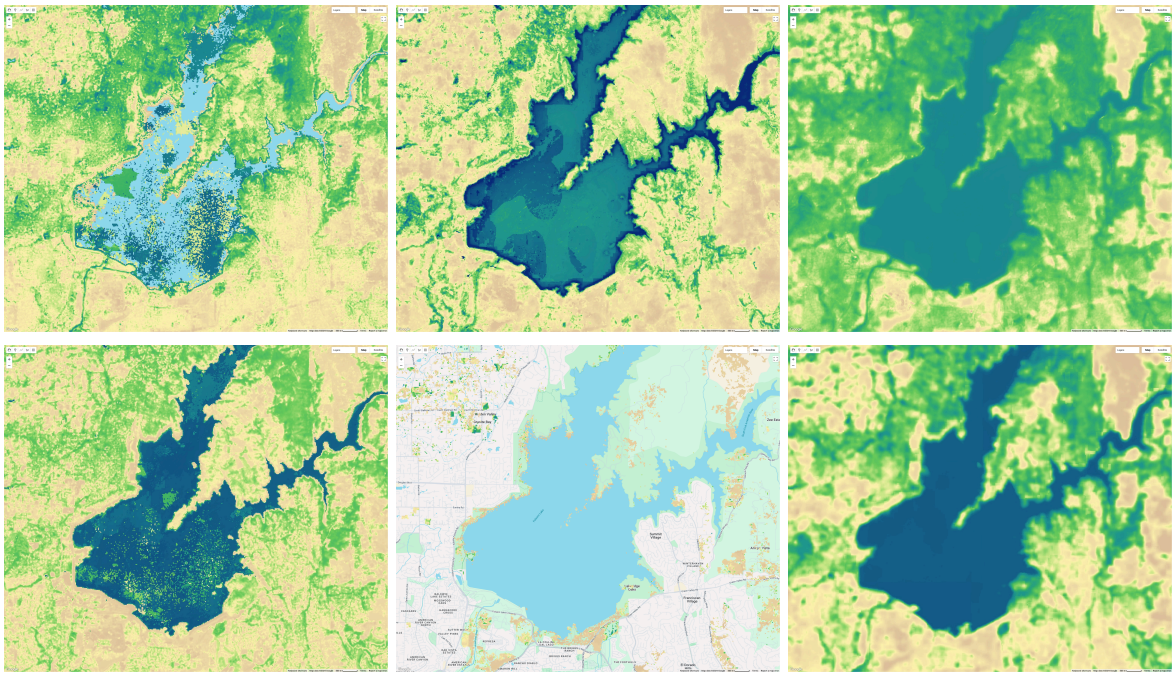
Action

OpenET has developed a fix that will be implemented in the next collection release. Until the fix is available, users are advised to rely only on monthly ET values with a “count” band of at least 1. Additionally, OpenET is evaluating further refinements to the interpolation process for potential inclusion in future releases.

10) High uncertainty for evaporation estimates for open water

OpenET models show high variability, large spread, and inconsistent estimation over large open water bodies (lakes, reservoirs, etc.).

Examples



Model ET images for July 2022 of Folsom Reservoir for (a) DisALEXI, (b) EEMETRIC, (c) GEESEBAL, (d) PTJPL, (e) SIMS (is not run for water pixels), (f) SSEBOP.
https://code.earthengine.google.com/30e085dc6c99568562771f5cbdda8e38

Lake Mead: https://explore.etdata.org/#11/36.1732/-114.4199

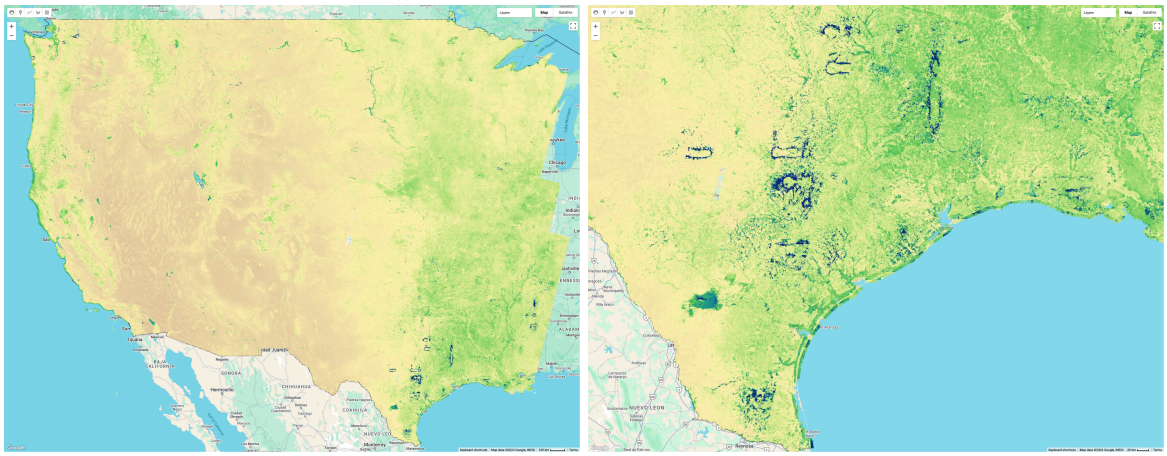
Action

Energy partitioning in water bodies complicates evaporation calculation due to the large heat storage capacity of lakes creating a lag between incoming solar radiation and LE. Additionally, diurnal variation in LE over lakes is typically driven by wind speed and vapor pressure gradients complicating scaling from instantaneous to 24-hour ET values. While some models in OpenET use open water subroutines (i.e. PT-JPL/AquaSEBS, eeMETRIC), many rely on core ET algorithms that were not developed for calculation of ET over open water. Last, most gridded weather datasets do not consider water bodies and their influence on local atmospheric conditions. Similar to evaporative cooling effects in agricultural fields, conditions over water bodies are typically cooler and more humid. Furthermore, wind speeds differ from land surfaces over open water due to changes in surface roughness.

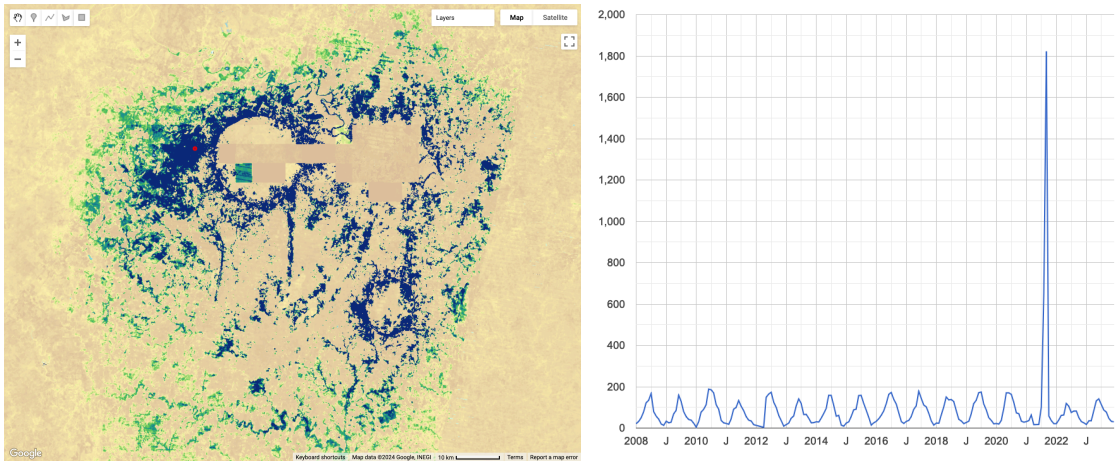
At this time, OpenET data values over large water bodies should be considered experimental and not used for operational decision making or planning. Additional research and review to improve calculation of ET values over open water is underway and will be incorporated in future releases.

11) Extreme values in DisALEXI monthly ET

There are regions of invalid extreme high and negative ET values in the DisALEXI monthly ET.



Maximum of all DisALEXI monthly ET values for the full dataset (a) and zoomed in on Texas (b), with scaling adjusted to highlight the high ET values.
https://code.earthengine.google.com/2d5ed29ae76b5e523855f07fcb5e59a0



Example of extreme values in the DisALEXI monthly ET image for Sept. 2021.
https://code.earthengine.google.com/3902ea72bab14f35b6277e086edc990b

Impact

The DisALEXI monthly ET values are impacted, but for the most part, these high ET values are being flagged as outliers and filtered/dropped in the ensemble calculation process. The majority of the extreme values are clustered in Texas, but smaller patches can be found throughout the dataset. The extreme values tend to result from a single bad overpass image and only impact one or two months.

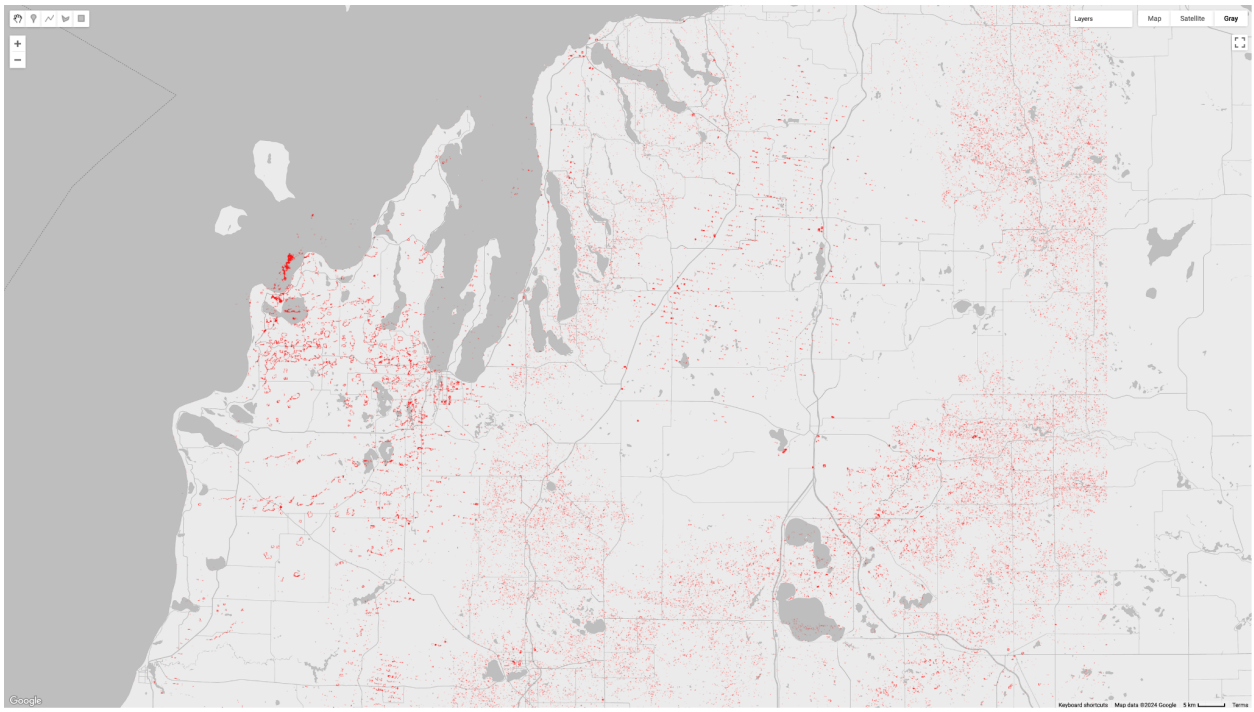
Action

The DisALEXI modeling team is working on updates to the model to remove these values, with the goal of incorporating these improvements into the next collection release. Users are encouraged to check for extreme values if they are using the DisALEXI monthly ET data.

12) Spurious data in ensemble collection over portion of Michigan Lower Peninsula

There are a relatively small number of spurious or unmasked pixels in the ensemble image collection over Michigan that are not present in any of the models.

Example



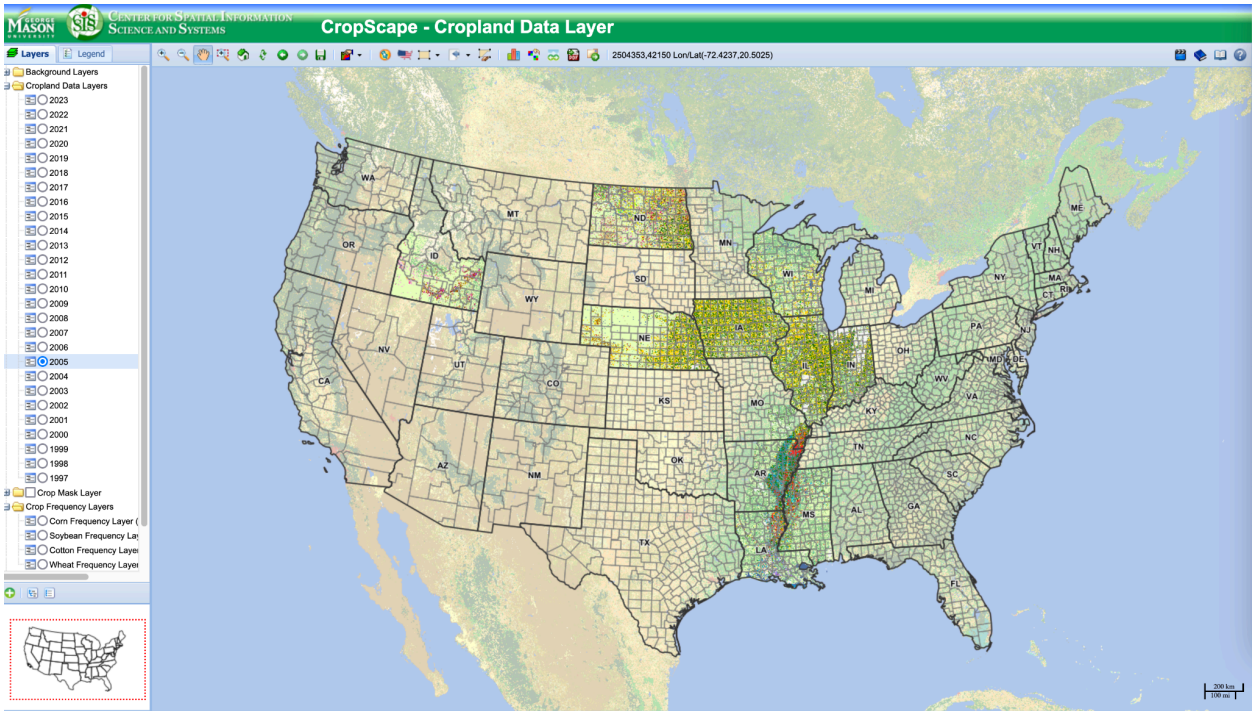
https://code.earthengine.google.com/ba3aea06e52d7fde71413ae458f635b4

Impact
Only the OpenET ensemble monthly ET values for a subset of months in 2020 over the Michigan Lower Peninsula are impacted.

Action
These values will either be masked in the next collection release, or the study area will be expanded to include some or all of the eastern U.S. Until that time, users should only use OpenET Collection v2.0 data within the CONUS portions of UTM zones 10-15.

13) Limited availability of crop type and field boundary information prior to 2007

Prior to 2007, the USDA Cropland Data Layer was produced using different methods and with limited geographic coverage. This limits the ability of OpenET to identify perennial crops and hay pasture crops reliably prior to 2008. OpenET currently uses a combination of the USGS National Land Cover Database for 2001 and 2006 in combination with the USDA Cropland Data Layer for 2008 to identify agricultural pixels and identify perennial crops and grass/hay, creating additional uncertainty in the annual crop type data layers used by OpenET prior to 2008.



Example of Cropland Data Layer coverage for 2005 showing coverage for a subset of the states.

Impact

This impact mostly impacts the SIMS and EEMETRIC models, which have the ability to run specific functions within the models to increase expected accuracy for perennial crops. Applications of data from these models for perennial crops prior to 2008 should assume higher uncertainty relative to the period from 2008 to present.

Action

OpenET will undertake analyses to assess expected changes in accuracy resulting from the limited availability of crop type information prior to 2008, and to assess any trends or patterns in ET data prior to 2008 relative to the period from 2008 to present. One key challenge, however, is that few accurate ground-based ET datasets exist to rigorously evaluate model performance prior to 2008 at the scale of individual fields.