LITERATURE: Waterbody masks and ancillary data sets

Wednesday, January 3, 2024 8:52 AM

1. **LAKE POLYGONS**
   a. **HydroLAKES**
      i. 1.4 million lakes larger than 10 ha (Messager et al., 2016)
      ii. 89 of 113 SuRGE sites are in HydroLAKES
      iii. Poor coverage of Inventory reservoirs
      iv. Used in the two Johnson et al papers
      v. **LakeATLAS** is an enhanced version of HydroLAKES.
      vi. **Lake-TopoCat**
         1) Sikder, M. S., J. Wang, G. H. Allen, Y. Sheng, D. Yamazaki, C. Song, M. Ding, J. F. Crétaux and T. M. Pavelsky (2023). "Lake-TopoCat: a global lake drainage topology and catchment database." Earth Syst. Sci. Data 15(8): 3483-3511.
         2) Paper: https://essd.copernicus.org/articles/15/3483/2023/essd-15-3483-2023-discussion.html
         3) Data: https://zenodo.org/records/7916729
         4) a complimentary dataset including lake typology (flow-through, headwater, endorheic, or coastal) and catchment surface areas for HydroLakes.
      vii. **Global Lake area, Climate, and Population dataset (GLCP)**
         1) Meyer, M. F., M. R. Brousil, A. N. Cramer, B. P. Lanouette, J. C. Padowski and S. E. Hampton (2020). "The Global Lake Area, Climate, and Population Dataset: A New Tool for Addressing Critical Limnological Questions." Limnology and Oceanography Bulletin 29(4): 110-116.
         2) Paper: https://www.nature.com/articles/s41597-020-0517-4
         3) Data: https://portal.edirepository.org/nis/mapbrowse?packageid=edi.394.4
         4) 1.42 + million lakes and reservoirs of at least 10 ha in size from 1995 to 2015 with co-located basin-level temperature, precipitation, and population data.
         5) surface area records from 1995 to 2015 for all HydroLAKES polygons
         6) The GLCP connects data from the Joint Research Centre (JRC) Global Surface Water dataset (Pekel et al. 2016), the HydroLAKES (Messager et al. 2016) and HydroBASINS (Lehner and Grill 2013) datasets, as well as NASA's MERRA-2 (Gelaro et al. 2017) and Gridded Population of the World (GPW) (Doxsey-Whitfield et al. 2015) data. Within the final GLCP, the HydroLAKES and HydroBASINS identifier columns ("Hylak_id" and "HYBAS_ID") have been retained to improve interoperability with additional datasets.
      viii. **Lake trophic state**
         1) Meyer et al 2024. National-scale, remotely sensed lake trophic state, 1984-2020. Scientific Data.
         2) Paper: https://eartharxiv.org/repository/view/5366/
         3) Data: https://portal.edirepository.org/nis/mapbrowse?packageid=edi.1395.1
         4) Annual trophic status for 55,662 lakes >10Ha.
         5) 89 SuRGE lakes
      ix. **GLOBathy**
         1) Khazaei, B., L. K. Read, M. Casali, K. M. Sampson and D. N. Yates (2022). "GLOBathy, the global lakes bathymetry dataset." Sci Data 9(1): 36.
         2) Paper: https://www.nature.com/articles/s41597-022-01132-9
         3) Data: https://gee-community-catalog.org/projects/globathy/
         4) Bathymetric maps based on the waterbody maximum depth estimates and HydroLAKES geometric/geophysical attributes for 1.4+ million HydroLAKES waterbodies
      x. **Ice cover duration**
         1) Doris Domart1, Daniel F. Nadeau1, Antoine Thiboult1, François Anctil1, Tadros Ghobrial1, Yves T. Prairie2, Alexis Bédard-Therrien1, Alain Tremblay. A global analysis of ice phenology for 3702 lakes and 1028 reservoirs across the Northern Hemisphere using Sentinel-2A imagery. Found pre-print online
         2) Also includes GRAND reservoirs
   b. **Global Lake and Wetland Dataset (GLWD)**
      i. Lehner, B. & Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. J. Hydrol. 296, 1–22 (2004).
      ii. Paper: https://www.sciencedirect.com/science/article/abs/pii/S0022169404001404
      iii. Data: https://www.worldwildlife.org/pages/global-lakes-and-wetlands-database
      iv. GLWD-1: lakes greater than 10 ha in size
         1) Extensive attribute data:
         2) 3,067 largest lakes (area ≥ 50 km2) and 654 largest reservoirs (storage capacity ≥ 0.5 km3)
      v. GLWD-2: lakes > 0.1km2 that are not in GLWD-1. ~250,000 polygons
      vi. Missing lots of NWI derived Inventory polygons
   c. **Global Water Bodies database (GLOWABO)**
      i. Verpoorter, C., T. Kutser, D. A. Seekell and L. J. Tranvik (2014). "A global inventory of lakes based on high-resolution satellite imagery." Geophysical Research Letters 41(18): 6396-6402.
      ii. Paper: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2014gl060641
      iii. Data:
      iv. 117 million lakes as small as 0.001km2
      v. Landsat 7
      vi. Attributes restricted to shape
   d. **GLAKES**
      i. Pi, X., Q. Luo, L. Feng, Y. Xu, J. Tang, X. Liang, E. Ma, R. Cheng, R. Fensholt, M. Brandt, X. Cai, L. Gibson, J. Liu, C. Zheng, W. Li and B. A. Bryan (2022). "Mapping global lake dynamics reveals the emerging roles of small lakes." Nature Communications 13(1): 5777.
      ii. Paper: https://www.nature.com/articles/s41467-022-33239-3
      iii. Data: https://doi.org/10.5281/zenodo.7016548.
      iv. 3.4 million lakes with maximum surface area larger than 3 ha (Pi et al., 2022)
      v. Estimates lake size changes from the period (1984–1999) to the end (2010–2019)
      vi. based on the Global Surface Water Occurrence (GSWO) dataset

            1) 30 m resolution Landsat satellite observations between 1984 and 2019
        vii. Georeferenced global Dam And Reservoir (GeoDAR) was used to distinguish reservoirs from natural lakes
    e. **Global Lake Inventories**
        i. Sheng, Y., C. Song, J. Wang, E. A. Lyons, B. R. Knox, J. S. Cox and F. Gao (2016). "Representative lake water extent mapping at continental scales using multi-temporal Landsat-8 imagery." Remote Sensing of Environment 185: 129-141.
        ii. Paper: https://www.sciencedirect.com/science/article/pii/S0034425715302649
        iii. Data:
        iv. circa-2000 and circa-2015 Global Lake Inventories with more than 9 million lake polygons larger than 0.4 ha (Sheng et al., 2016)
        v. Attributes on seasonal variability in Landsat 8 continental lake mapping
    f. **LAGOS-US**
        i. Background and data: https://lagoslakes.org/lagos-us-overview/
        ii. The LAGOS-US Research Platform includes the 479,950 lakes and reservoirs > 1 ha in the 48 conterminous US states and tribes
        iii. LOCUS is based on a snapshot of the high-resolution National Hydrography Dataset
        iv. excludes basins that were entirely artificial or built for high-intensity human use based on our interpretation of labels assigned by the NHD HR--these include, but are not limited to, sewage treatment ponds, aquaculture ponds, or retention pond.
        v. Data modules include reservoirs vs lakes, depth, network status, water chemistry
    g. **NHDPlusV2**
        i. Schmadel et al. (2018) and Gardner et al. (2019) integrated lakes into river networks
    h. **NHDPlus High Resolution**
        i. **Daily water temperature data**
            1) Willard, J. D., J. S. Read, S. Topp, G. J. A. Hansen and V. Kumar (2022). "Daily surface temperatures for 185,549 lakes in the conterminous United States estimated using deep learning (1980–2020)." Limnology and Oceanography Letters 7(4): 287-301.
            2) The NHD (Moore et al. 2019) high-resolution polygons (based on 1 : 24,000 scale data) were downloaded as geodatabase files for each of 48 states in the conterminous United States, as well as the District of Columbia. Lakes and reservoirs were extracted using the "NHDWaterbody" layer from the geodatabase and filtered to values in the "FType" attribute that corresponded to 390, 436, and 361 (lake/pond, reservoir, and playa, respectively). The Great Lakes, several improperly labeled coastal lagoons, and lakes less than 4 ha (based on the value in the "AreaSqKm" NHD attribute) were removed from the dataset, and the remaining 185,549 lakes defined the complete lake coverage used in this data release.
    i. GLOBCOVER
        i. Bontemps, S., Defourny, P., Bogaert, E. V., Arino, O., Kalogirou, V., and Perez, J. R.: Globcover Products Description and Validation
        ii. Used in WETCHARTS
            1) A global wetland methane emissions and uncertainty dataset for atmospheric chemical transport models (WetCHARTs version 1.0)
        iii. NEED TO FOLLOW UP ON THIS
    j. **Hydrolinks**
        i. R package that correlated lat/long with polygons from NHDPlusV2, NHDHighRes or Hydrolakes
2. **POINT DATA**
    a. Global Reservoirs and Dams (GRanD)
        i. Largely replaced by GeoDAR
        ii. Lehner, B., C. R. Liermann, C. Revenga, C. Vörösmarty, B. Fekete, P. Crouzet, P. Döll, M. Endejan, K. Frenken, J. Magome, C. Nilsson, J. C. Robertson, R. Rödel, N. Sindorf and D. Wisser (2011). "High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management." Frontiers in Ecology and the Environment 9(9): 494-50
        iii. Paper: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/100125
        iv. Data: https://www.globaldamwatch.org/directory
        v. GRanD was constructed by harmonizing AQUASTAT and a wide range of regional gazetteers and inventories
        vi. v1.3, contains 7320 dams as well as their reservoir boundaries and over 50 attributes,
        vii. Complimentary ice cover data
            1) Doris Domart1, Daniel F. Nadeau1, Antoine Thiboult1, François Anctil1, Tadros Ghobrial1, Yves T. Prairie2, Alexis Bédard-Therrien1, Alain Tremblay. A global analysis of ice phenology for 3702 lakes and 1028 reservoirs across the Northern Hemisphere using Sentinel-2A imagery. Found pre-print online.
    b. **GlObal geOreferenced Database of Dams (GOOD2)**
        i. Mulligan, M., van Soesbergen, A. and Sáenz, L., 2020. GOODD, a global dataset of more than 38,000 georeferenced dams. Scientific Data, 7(1), pp.1-8.
        ii. Paper: https://www.nature.com/articles/s41597-020-0362-5
        iii. Data: https://www.globaldamwatch.org/directory
        iv. Tabular, location only, many lacking area
        v. This global dataset of more than 38,000 dams was developed by digitising visible dams using Google Earth's satellite imagery
    c. Georeferenced global Dam And Reservoir (GeoDAR)
        i. Wang, J. et al. GeoDAR: Georeferenced global dam and reservoir dataset for bridging attributes and geolocations. Earth Syst. Science Data Discuss 14, 1–52 (2021).
        ii. Paper: https://essd.copernicus.org/articles/14/1869/2022/
        iii. Data: https://zenodo.org/records/6163413
        iv. The GeoDAR dataset, taking advantage of multi-source dam/reservoir inventories, provides global documentation of reservoirs with detailed attribute tables and well-georeferenced spatial locations.
        v. GeoDAR v1.0 is essentially a georeferenced subset of the International Commission On Large Dams World Register of Dams (ICOLD WRD)
        vi. GeoDAR v1.1 consists of (a) nearly 25 000 dam points which harmonized v1.0 and GRanD for an expanded inclusion of the largest dams and (b) the reservoir boundaries for most (87 %) of the dam points based on a one-to-one relationship between dams and reservoirs.
    d. **World Register of Dams (WRD)**
        i. https://www.icold-cigb.org/
        ii. regularly updated by the International Commission on Large Dams
        iii. 60 000 "large" dams, defined as those with a wall higher than 15 m or between 5 and 15 m but with a reservoir storage greater than 3×106 m3 (mcm).
        iv. more than 40 attributes (e.g., reservoir storage capacity, dam height, and reservoir purpose)
        v. Must purchase access!
        vi. Password: biogeo01
    e. Hydropower Infrastructure - LAkes, Reservoirs, and RIvers (HILARRI), Version 2
        i. Web site: https://hydrosource.ornl.gov/dataset/hilarri-v2
        ii. HILARRI is a database of links between major datasets of operational hydropower dams and powerplants, and inland water bodies

- **iii.** Dams from the National Inventory of Dams (2021) and the Global Reservoir and Dam Database (GRanD v1.3
- **iv.** Hydropower plants from the Existing Hydropower Assets dataset (EHA 2022)

3. RASTER IMAGES

- a. **Global Surface Water Explorer (JRC GSW)**
  - i. Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422 (2016).
  - ii. Paper: https://www.nature.com/articles/nature20584
  - iii. Data: https://global-surface-water.appspot.com/
    - 1) Filezilla works well
    - 2) In ArcGIS Pro
      - a) Add WMTS server
      - b) Search GSW in ArcGIS online (catalog pane)
      - c) 30m resolution
  - iv. the entire multi-temporal orthorectified Landsat 5, 7 and 8 archive spanning the past 32 years to map the spatial and temporal variability of global surface water and its long-term changes.
  - v. The data I downloaded are raster images. I don't think these are available as polygons with attributes.
  - vi. Does not distinguish between rivers, lakes, etc
- **b. Wetland Area and Dynamics for Methane Modeling- WAD2M**
  - i. Zhang, Z., Fluet-Chouinard, E., Jensen, K., McDonald, K., Hugelius, G., Gumbricht, T., Carroll, M., Prigent, C., Bartsch, A., and Poulter, B.: Development of the global dataset of Wetland Area and Dynamics for Methane Modeling (WAD2M) , Earth Syst. Sci. Data, 13, 2001–2023, https://doi.org/10.5194/essd-13-2001-2021, 2021.
  - ii. Paper: essd-13-2001-2021.pdf (copernicus.org)
  - iii. Data: Development of a global dataset of Wetland Area and Dynamics for Methane Modeling (WAD2M) (zenodo.org)
  - iv. Gridded 0.25 degree, temporal dynamics at a monthly time step from 2000-2018
  - v. Fuses multiple datasets including ground-based wetland inventories, remote sensing products of open waters, and surface inundation datasets based on optical and active and passive microwave satellite observations
  - vi. Masks for coastal wetlands, cultivated wetlands (like rice paddies), open water (using Pekel et al. 2016 waterbodies covered by open water for >50% of months as the mask)… there is a version 2.0 but it isn't published yet which also adds HydroLakes to the mask
  - vii. They are using the version 2.0 to help reduce double counting in the global methane budget, but there is still a double counting term that is estimated by expert elicitation and includes small lakes and ponds <0.1km2
- c. **Surface WAter Microwave Product Series (SWAMPS)** multi-satellite surface water product
  - i. Schroeder, R., McDonald, K. C., Chapman, B. D., Jensen, K., Podest, E., Tessler, Z. D., Bohn, T. J., and Zimmermann, R.: Development and Evaluation of a Multi-Year Fractional Surface Water Data Set Derived from Active/Passive Mi- crowave Remote Sensing Data, Remote Sensing, 7, 16688– 16732, https://doi.org/10.3390/rs71215843, 2015.
  - ii. Used in WETCHARTS to introduce temporal variability into GLOBCOVER and/or GLWD polygon water masks (see above).
  - iii. The WETCHARTS paper argues that this remote sensing dataset is great for temporal resolution, but that satellite-retrieved inundation data products suffer from spatial bias because vegetation cover remains a major confounding variable (Schroeder et al., 2015). Moreover, satellites cannot directly observe subsurface soil saturation, even though these soils amount to significant CH4 fluxes to the atmosphere (Turetsky et al., 2014).
- **d. Epilimnion Water Temperature**
  - i. Willard, J. D., J. S. Read, S. Topp, G. J. A. Hansen and V. Kumar (2022). "Daily surface temperatures for 185,549 lakes in the conterminous United States estimated using deep learning (1980–2020)." Limnology and Oceanography Letters 7(4): 287-301.
  - ii. Paper: https://aslopubs.onlinelibrary.wiley.com/doi/full/10.1002/lol2.10249
  - iii. Data: http://10.0.19.202/P9CEMS0M
  - iv. daily surface water temperatures for 185,549 lakes and reservoirs (hereafter referred to simply as lakes) in the conterminous United States (the lower 48 states and the District of Columbia) from 1980 to 2020
  - v. Based on NHDHighResolution

4. GRIDDED METHANE EMISSION INVENTORIES

- **a. EDGAR**
- **b. WETCHARTS**
  - i. Bloom et al. (2017). A global wetland methane emissions and uncertainty dataset for atmospheric chemical transport models (WetCHARTs version 1.0). Geosci. Model Dev., 10, 2141–2156, 2017 https://doi.org/10.5194/gmd-10-2141-2017
  - ii. Gridded data using 2 water masks (GLWD and GLOBCOVER).
    - 1) Temporal variability in water extent in these masks was indexed to SWAMPS, a remote sensing based inundation extent data product. Combining these data sources retains the strength of these two water masks and the temporal resolution of remote sensing data. See SWAMPS above for more information.
    - 2) ERA-5 precipitation was independently used to simulate temporal variability in inundation extent.
    - 3) Both of these water masks include lakes and reservoirs, in addition to wetlands. Thus the gridded product lumps these emission sources.
  - iii. Monthly time step
  - iv. They use a scaling factor to make sure that the predicted global emissions match the range of top-down values presented in Sounis et al for 2009 -2010 (166 Tg CH4 yr−1; 125–204 Tg CH4 yr−1).
    - 1) This is kinda groovy. WETCHARTS is used to estimate the spatial distribution of emissions, but not total emissions! Total emissions is taken from Sounois!
  - v. The various scenarios in this study were used as priors for a run of GEOS-Chem, a global atmospheric transport and chemistry model the yield spatially explicit predictions of atmospheric CH4 concentration. This enables a comparison of predicted CH4 concentration to observations at long-term monitoring sites. This comparison then serves as a check for the playability of the simulations run in this paper.
  - **vi. Only made it to Discussion! Should finish this paper.**
- c. **GEOS-Chem wetland CH4 emissions inventory**
  - i. Pickett-Heaps et al., 2011; 2009–2010 derivation described in Turner et al., 2015;
  - ii. Referenced in WetCharts
  - **iii. UPDATE**
  - iv. C