

Title: GREMLIN CONUS2 Dataset

Abstract: The objective of this research is to develop techniques for assimilating GOES-R series observations in precipitating scenes for the purpose of improving short-term convective-scale forecasts of high-impact weather hazards. Whereas one approach is radiance assimilation, the information content of GOES-R radiances from its Advanced Baseline Imager saturates in precipitating scenes, and radiance assimilation does not make use of lightning observations from the GOES Lightning Mapper. Here, a convolutional neural network (CNN) is developed to transform GOES-R radiances and lightning into synthetic radar reflectivity fields to make use of existing radar assimilation techniques. We find that the ability of CNNs to utilize spatial context is essential for this application and offers breakthrough improvement in skill compared to traditional pixel-by-pixel based approaches. To understand the improved performance, we use a novel analysis method that combines several techniques, each providing different insights into the network's reasoning. Channel-withholding experiments and spatial information-withholding experiments are used to show that the CNN achieves skill at high reflectivity values from the information content in radiance gradients and the presence of lightning. The attribution method, layerwise relevance propagation, demonstrates that the CNN uses radiance and lightning information synergistically, where lightning helps the CNN focus on which neighboring locations are most important. Synthetic inputs are used to quantify the sensitivity to radiance gradients, showing that sharper gradients produce a stronger response in predicted reflectivity. Lightning observations are found to be uniquely valuable for their ability to pinpoint locations of strong radar echoes.

Contact: Kyle Hilburn, kyle.hilburn@colostate.edu

License information or restrictions placed on the data: The material is open access and distributed under the terms and conditions of the Creative Commons CC BY license that allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. (<https://creativecommons.org/licenses/by/4.0/>)

Recommended data citation: Hilburn, K. (2022). GREMLIN CONUS2 Dataset. Colorado State University. Libraries. <http://dx.doi.org/10.25675/10217/235392>

Associated article citation: Hilburn, Kyle A., Imme Ebert-Uphoff, and Steven D. Miller. (2021). Development and Interpretation of a Neural-Network-Based Synthetic Radar Reflectivity Estimator Using GOES-R Satellite Observations, *Journal of Applied Meteorology and Climatology*, **60**, 3-21, <https://doi.org/10.1175/JAMC-D-20-0084.1>.

Format of data files: NetCDF

Location where data were collected: Region bounding box: Latitude 29.577 to 48.002 degrees_north and Longitude -106.766 to -75.184 degrees_east

Time period during which data were collected: 2019-04-18-2019-07-18

File Information: gremlin_conus2_dataset.nc – This is the dataset used to train and test the GREMLIN Version-1 model in Hilburn et al. (2021). It consists of GOES-16 ABI, GOES-16 GLM, and MRMS data resampled to the 3 km HRRR grid and matched in time. The samples consist of 256 x 256-pixel images covering severe storms for 6-hour periods with 15-minute refresh over a 92-day period. The time-period runs from 2019-04-18 to 2019-07-18. The spatial coverage is eastern CONUS ranging from 29.577 to 48.002 degrees_north latitude and -106.766 to -75.184 degrees_east longitude.

Definitions of acronyms, site abbreviations, or other project-specific designations used in the data file names or documentation files:

- ABI – Advanced Baseline Imager

- CIRA – Cooperative Institute for Research in the Atmosphere
- CLASS – NOAA Comprehensive Large Array-data Stewardship System
- CONUS – Contiguous United States
- C07 – Shortwave infrared window band (3.9 μm)
- C09 – Mid-level water vapor band (6.9 μm)
- C13 – Clean longwave infrared window band (10.3 μm)
- GED – Lightning Group Extent Density
- GLM – Geostationary Lightning Mapper
- GOES – Geostationary Operational Environmental Satellite
- GREMLIN – GOES Radar Estimation via Machine Learning to Inform NWP
- HRRR – High-Resolution Rapid Refresh
- MRMS – Multi-Radar Multi-Sensor
- NCEP – National Centers for Environmental Prediction
- NOAA – National Oceanic and Atmospheric Administration
- NWP – Numerical Weather Prediction
- REFC – Composite Radar Reflectivity
- SPC – Storm Prediction Center

Variable information:

- nsamples – number of sample images (2246)
- nx – number of longitude points (256)
- ny – number of latitude points (256)
- latitude – Earth latitude (degree north)
- longitude – Earth longitude (degree east)
- time – time (seconds since 2019-04-17_00:00:00Z)
- GOES_ABI_C07 – C07 brightness temperature (K), fill value -1.E30
- GOES_ABI_C09 – C09 brightness temperature (K), fill value -1.E30
- GOES_ABI_C13 – C13 brightness temperature (K), fill value -1.E30
- GOES_GLM_GROUP – lightning group extent density (groups 5-minutes⁻¹ km⁻²), fill value -1.E30
- MRMS_REFC – composite radar reflectivity (dBZ), fill value -999

Uncertainty, precision, and accuracy of measurements:

- Brightness temperature – The relative accuracy (noise effective differential temperature) of ABI is 0.1 K and the absolute accuracy is 1 K for a 300 K scene.
- Lightning – The nominal flash detection efficiency for GLM is 80% and the false event rate is 5%.
- Radar reflectivity – The nominal calibration accuracy of WSR-88D NEXRAD horizontal polarization radar reflectivity is 1 dB.

Environmental or experimental conditions: severe storms

Method(s): The methodology is described in detail by Hilburn et al. (2021). The ABI, GLM, and MRMS data sets were resampled to a common 3 km grid. A cloud height of 10 km was used for removing parallax displacements. Satellite and radar samples were matched in time with a maximum time difference of 2.5 minutes. GLM lightning groups were accumulated over 15-minute time periods. Regions- and times-of-interest were selected to maximize the number of severe storm reports.

Standards or calibrations that were used: The ABI radiances were calibrated using the coefficients provided in the L1b ABI data files.

Software: GNU Fortran 8.5.0, Python 3.7.3

Quality assurance and quality control that have been applied: Samples have been screened for missing data, and out-of-bounds checks were applied to the variables. The MRMS MergedReflectivityQCComposite field was used.

Limitations to reuse: None

Date dataset was last modified: 15-June-2022

- Are there multiple versions of the dataset? No

Data source: This dataset was created from publicly available datasets.

- The L1b ABI data files are available from NOAA CLASS (https://www.avl.class.noaa.gov/saa/products/search?datatype_family=GRABIPRD).
- The L2 GLM data files are available from NOAA CLASS (https://www.avl.class.noaa.gov/saa/products/search?datatype_family=GRGLMPROD).
- The MRMS composite reflectivity data files are available from NCEP (<https://mrms.ncep.noaa.gov/data/>).

Any other important information:

- Samples 0 to 1797 were used for training and samples 1798 to 2245 were used for testing.
- Funding: GOES-R Program Award NA19OAR4320073