PICES on the cloud

by Chelle Gentemann, Marisol García-Reyes, Trond Kristiansen and William Sydeman

At the 2019 PICES Annual Meeting, Marisol García-Reyes presented a talk on "*Cloud computing of key NASA oceanographic data: Implications for automating aspects of ecosystem status reports*" at the Topic Session (TS 8) on "*Creating more effective Integrated Ecosystem Assessments (IEAs) in PICES countries*", which introduced a simple tool that provides time series of satellite-based environmental data for all 14 PICES regions, as defined for PICES' third North Pacific Ecosystem Status Report (NPESR). The PICES Regional Ecosystem Tool was created by Earth and Space Research and Farallon Institute with support and funding from NASA's Inter-agency Implementation and Advanced Concepts Team (IMPACT) to facilitate the reporting and comparison of oceanographic conditions across PICES regions while increasing accessibility to satellite data on the cloud.

Connecting environmental and ecological data allows us to better understand the effects of climate change and extreme events on marine ecosystems. About every 5 years PICES publishes a NPESR to provide a review and summary of the marine ecosystems in each of the 14 PICES regions (Fig. 1). The individual chapter leads are responsible for compiling relevant ecosystem and environmental information to create their assessment.

Currently, the relevant and comparable environmental satellite data exist but are at different archive centers, temporal samplings, and spatial resolutions. Creating time series, trends, and maps requires finding, downloading, and storing large amounts of data. Processing these differently-sampled data requires a high-level understanding of software development to extract and analyze them, limiting their usage and hindering comparative ecosystem analyses. Advances in open source software and cloud computing environments can provide access to environmental data without needing to download and store the data. Taking the analysis to the data, rather than the data to the user, makes analysis of the regional environmental conditions easier and reproducible.

As part of NASA's plan to migrate copies of their data onto the cloud and make processing of data reproducible and shareable, we developed an online tool that allows one to easily generate spatially-explicit (customizable) time series of satellite-based data for all of the PICES regions. This tool is interactive and runs fully on the cloud. Moreover, the notebooks can be modified to customize data and regions and to do further analysis or different plots. This approach could be used not only to enhance the timeliness and utility of the PICES NPESR documents, but also to promote biophysical analysis by providing easy access to the data required to do so.
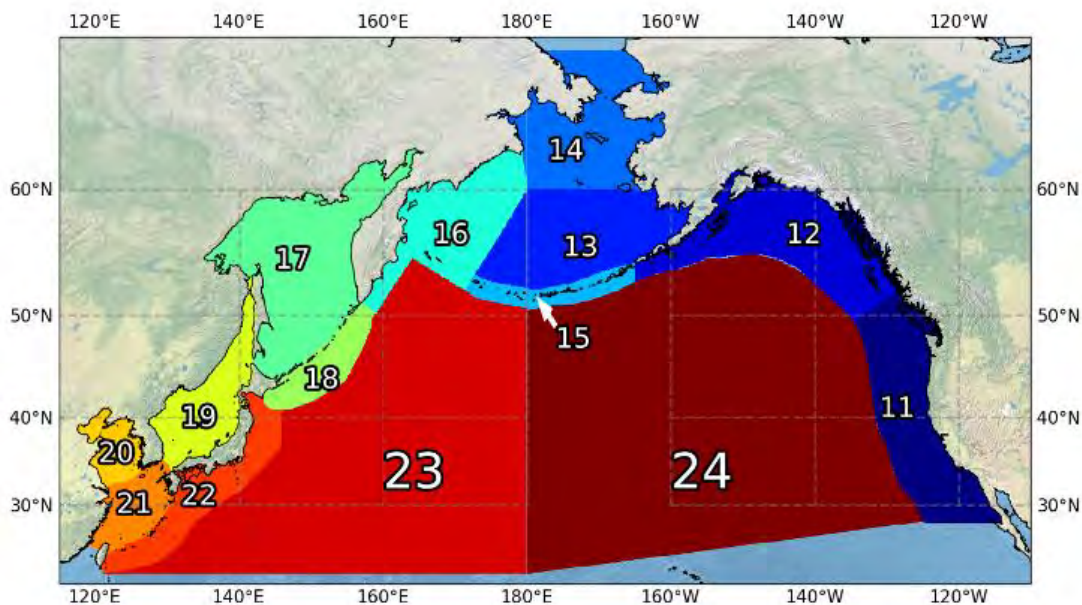

Fig. 1   PICES regions as defined in the 3rd NPESR.

In short, the PICES Regional Ecosystem Tool runs completely on the cloud without need of a user background in cloud computing. To launch the tool, click on this [launch binder] [1] icon. Once the tool is running, simply click on the orange "PICES Regional Ecosystem Tool" icon on the left menu and a script window opens with instructions on how to run the script and customize time series output for particular regions and data types. Plots of the monthly data, the seasonal cycle, and the anomalies are created (Fig. 2). Further analysis can be completed within the tool, or the time series data can be downloaded.

We encourage you to try this tool and contact us with suggestions for features, plots, or other data you would like to see included. As part of this project we have also developed online and easy to follow workshops on Python for oceanographers, which are available here. Our future work on the tool will include newly available cloud-optimized data and climate projections such as the Coupled Model Intercomparison Project Phase 6 (CMIP6) and Community Earth System Model (CESM) Large Ensemble Numerical Simulation (LENS) dataset.
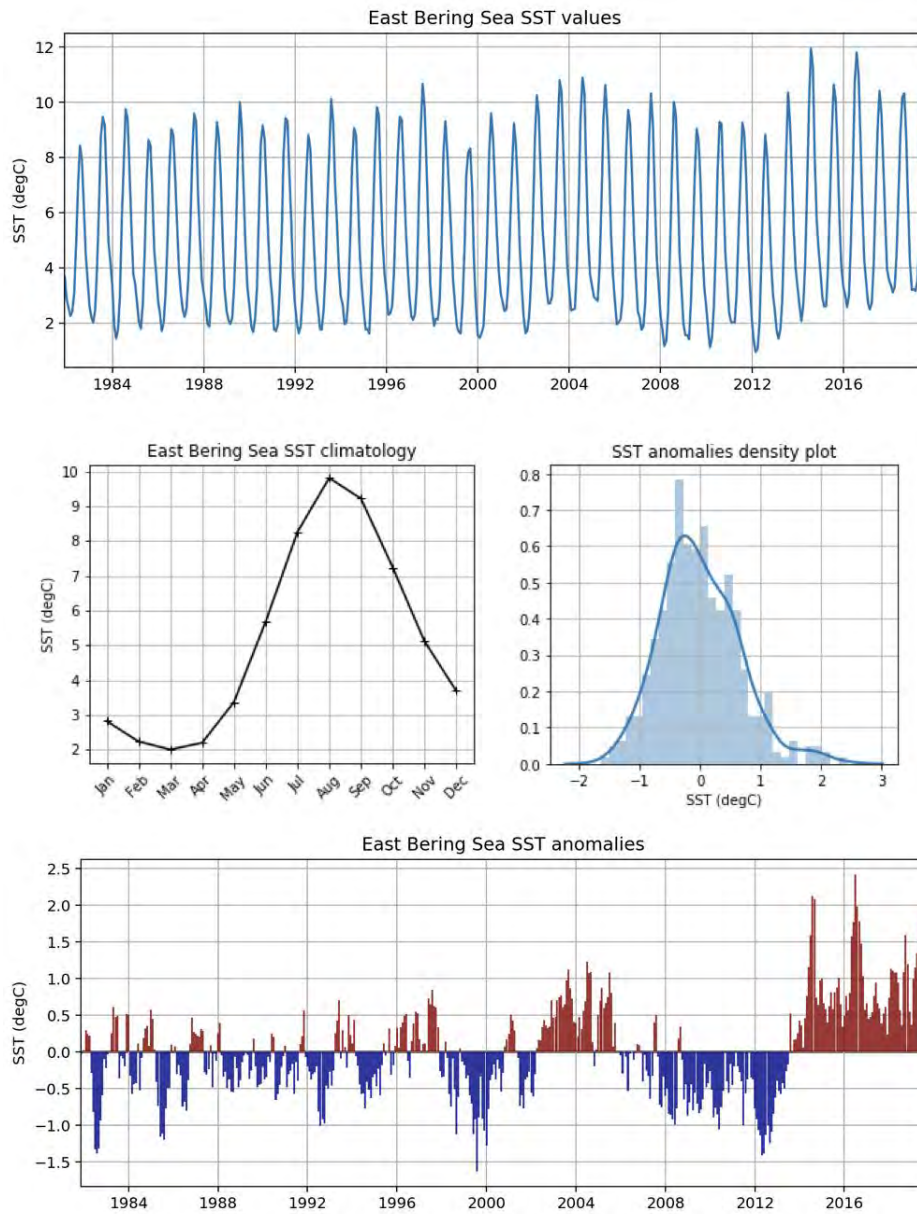

Fig. 2   Figures created by the PICES Regional Ecosystem Tool. PICES Region 13 - East Bering Sea monthly sea surface temperature (SST) data from 1981–present. From top to bottom: SST time series, climatology of the SST annual cycle and density plots of SST anomalies, and time series of monthly SST anomalies.

[1] https://github.com/python4oceanography/PICES-tools

In A BOX
The Technical Details

Technology

The PICES Regional Ecosystem Tool takes full advantage of new developments on open source software (OSS), which not only provides state-of-the-art data processing but more importantly facilitates and promotes the creation of reproducible research environments that can be publicly (or privately) shared. All of the OSS we developed for the PICES tool is hosted on GitHub an online software repository with version control. The tool runs from the GitHub repository, and accesses data included in the same repository. In future versions, as NASA data become available, the data accessed would be also on the cloud. We then utilize the free Binder service which creates a Docker image of the software repository, providing an interactive JupyterHub notebook environment that runs the PICES tool. This notebook is running on the Google cloud. Many of the OSS tools we used here make it almost invisible to a user that they are running their analysis on the cloud. We expect this type of OSS library development to continue, providing a gateway for research scientists who are not experienced software developers to access all the public data and computing power that is available on the cloud.

Data

GIS shapefiles for each region, provided by PICES, were processed using the open source GeoPandas and Xarray Python software libraries to generate each region's masks.

Currently, the tool includes SST (NOAA Optimum Interpolation V2), chlorophyll-*a* concentration (GlobColour SeaWiFs product version 2017.2), wind vectors (NOAA FNMOC 10-m surface winds), and surface ocean current vectors (ESR OSCAR surface currents). In a future version (currently in development) of the PICES Regional Ecosystem tool we will include output data from the recently released CMIP6 climate models that it has been made freely available by Pangeo and Google cloud.

Dr. Chelle L. Gentemann (cgentemann@esr.org) is a Senior Scientist at both Earth & Space Research and the Farallon Institute. Her more recent research focuses on interdisciplinary science using cloud computing, open source software development, air–sea interactions, and upper ocean physical processes. She has worked on the calibration, radiative transfer modeling, algorithm development, validation, and operational near-real-time distribution of multiple satellite sensors.

Dr. Marisol García-Reyes (marisolgr@gmail.com) is a Principal Scientist at Farallon Institute. Her research focus is on how variability of ocean climate impacts the marine ecosystem. One of the most time consuming, and at the same time, basic and important parts of this work is to acquire and process climate, atmospheric, and oceanographic data, and then synthesize them in a way that can be matched with biological and ecological data to look for biophysical relationships.

Dr. Trond Kristiansen (trondkr@faralloninstitute.org) is a Principal Scientist at the Farallon Institute who is currently working on downscaling global climate models for the Arctic Ocean using ROMS, coupled particle-tracking and individual-based models for kelp, larval fish, and plastic, as well as trying to understand how habitats at higher latitudes may be affected by climate change.

Dr. William J. Sydeman (wsydeman@comcast.net) is President and Senior Scientist at Farallon Institute who studies how biological components of the marine ecosystem are affected by the variability of the physical environment. He leads collaborative projects that investigate climate impacts on multiple trophic levels. Bill's expertise is in the pelagic ecosystem, encompassing upwelling processes, spatial ecology, predator-prey interactions, and top predators such as seabirds. In PICES he is a member of the Section on Marine Birds and Mammals.