

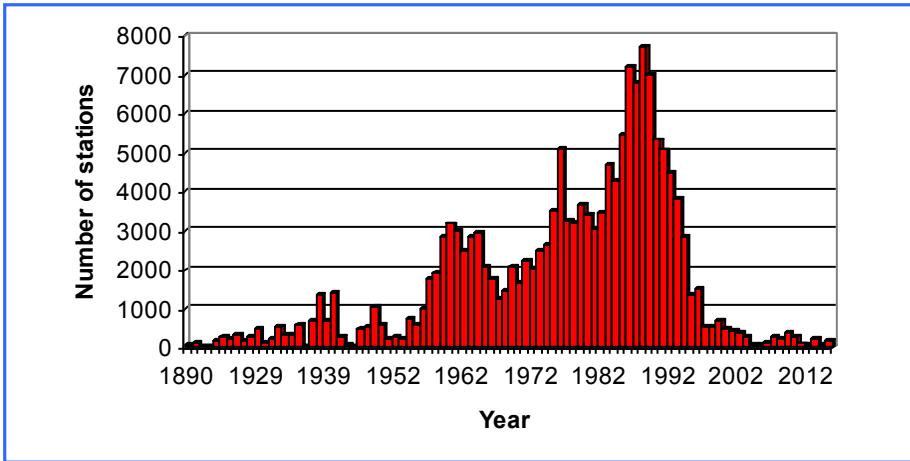


Marine Hydrophysical Institute and All-Russian Research Institute of Hydrometeorological Information – World Data Center: Comparison and Development of Black Sea Databases

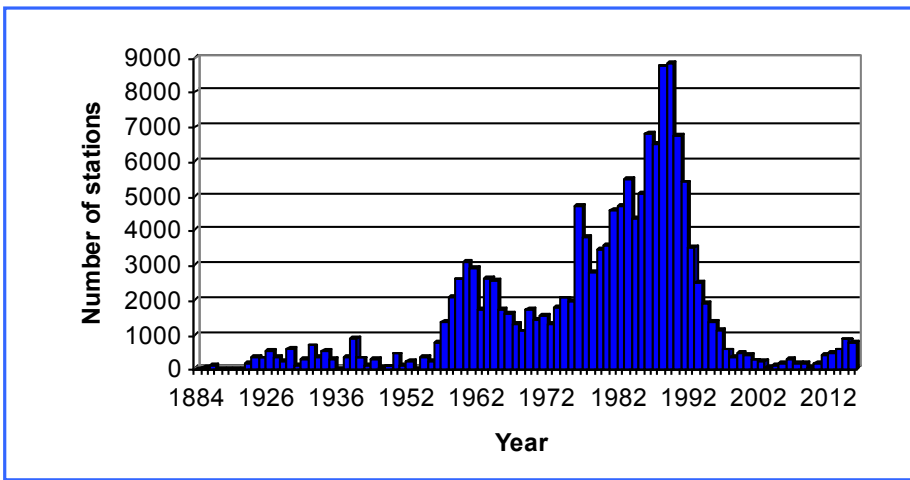
Eugeny Godin, Andrey Ingerov, Elena Isaeva, Tatiyana Kasyanenko
Marine Hydrophysical Institute, Russian Academy of Sciences
Alexander Vorontsov

All-Russian Research Institute of Hydrometeorological Information – World Data Center

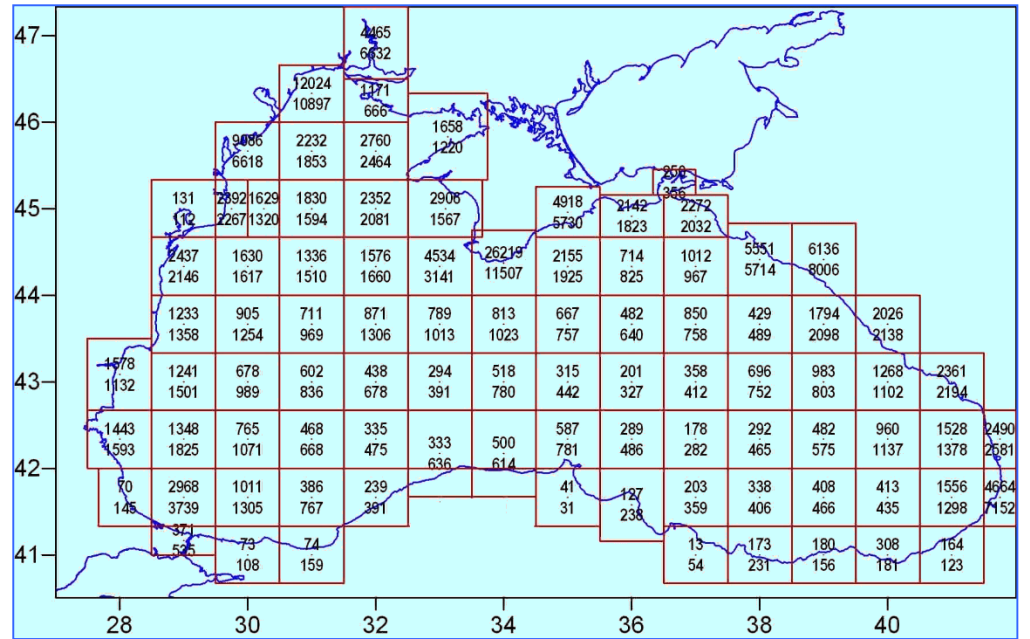
The Black Sea is one of the most studied areas of the World Ocean by the number of oceanographic stations accomplished there. At present, there is a series of Black Sea oceanographic databases, which were developed within both national and international projects. Databases created in All-Russian Research Institute of Hydrometeorological Information – World Data Center and Marine Hydrophysical Institute, Russian Academy of Sciences, should be attributed to the most comprehensive ones. Joint work to compare the Black Sea oceanographic data arrays held in the institutions has been carried out since 2017. A preliminary analysis showed that the two arrays differ not only in total number of stations (148,329 stations for 1884–2015 and 156,980 stations for 1890–2015, respectively) but in their distribution both in time and throughout the Black Sea.



Marine Hydrophysical Institute RAS, Distribution of oceanographic stations on years



All-Russian Research Institute of Hydrometeorological Information – World Data Center
Distribution of oceanographic stations on years

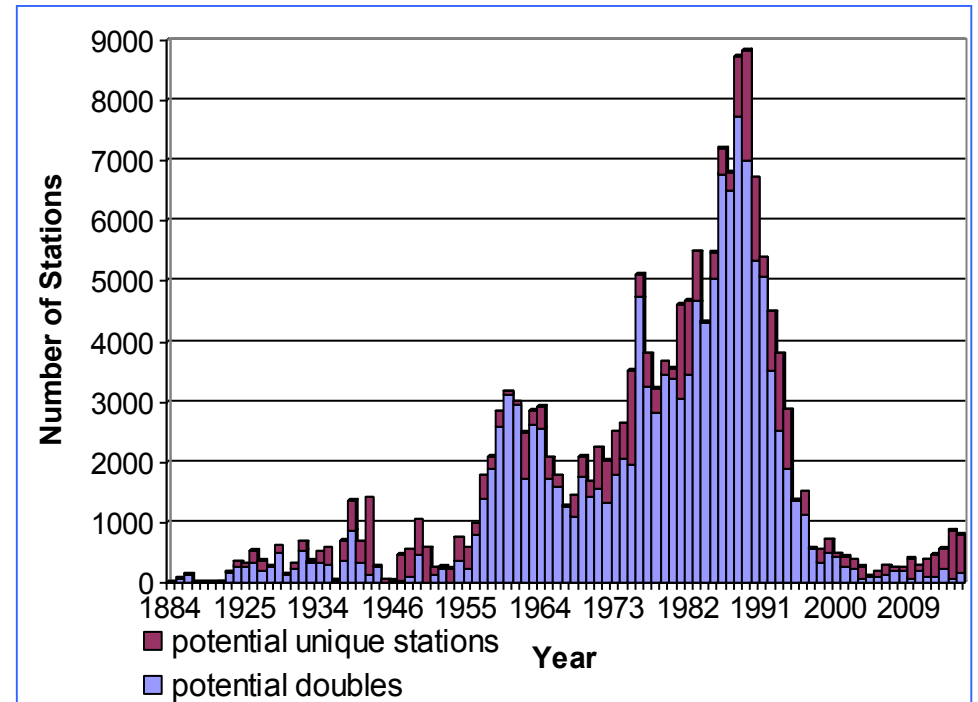


Distribution of oceanographic stations on squares in the Black Sea:
lower –All-Russian Research Institute of Hydrometeorological Information – World Data Center
upper –Marine Hydrophysical Institute RAS

Duplicate check is a standard stage of data collection process and, for this course, searching and fixing duplicates procedure is a part of most oceanographic data projects such as SeaDataNet, ARGO, PERSEUS, GTSP and so on. We have analyzed an available information and took into consideration their approaches to duplicate check. We believe that both automatic and manual procedures for detecting “exact” and “near” metadata and data duplicates is more effective and try to realize our procedure on these principles.

Special software was developed using Lazarus IDE in the Oceanographic Data Bank of Marine Hydrophysical Institute to optimize the procedure of comparing the two data arrays. The software allowed to look through metadata, visualize them on a map, and, based on comparing coordinates and time of stations and number of measurements, set “flags” to mark stations as “doubles”, “doubles in question”. Data marked as “doubles” or “doubles in question” (after further consideration) compose the overlapping part of the databases. It is kept as an array of coupled numbers of coinciding stations. After that, data, which do not coincide (“unique” data), are detected by exclusion. However, its use showed that the comparison of only metadata did not give sufficient information to classify stations as “unique” or “doubles”. A new version of the software was designed to eliminate the arising ambiguities. It offers more functions, including those of selecting, displaying, and comparing both metadata and data. The current version of the software provides:

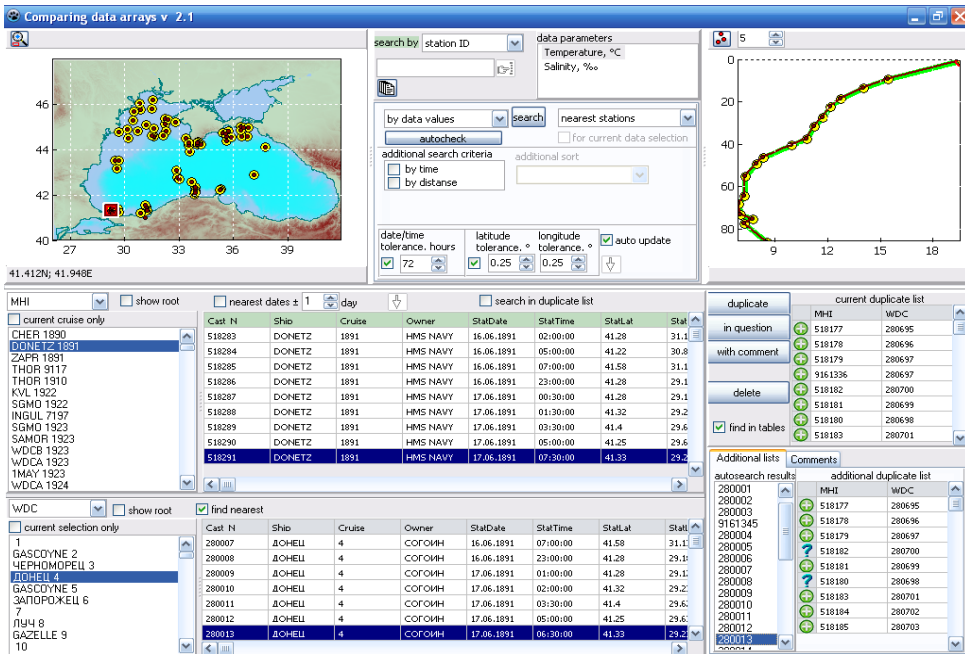
- Metadata selecting both by cruises (data sets) and by spatial and temporal criteria, as well as viewing them as graphs and tables
- Displaying lists of measured parameters for a couple of stations under consideration from the compared arrays; presenting corresponding vertical profiles for a chosen parameter
- Possibility of shifting one of the arrays in time by a value determined by user (it makes the comparison easier if one array contains UTC/GMT and the other has local time), etc.
- Current cruises or near date stations viewing modes, map zoom, profile zoom and scroll possibilities, setting of spatial and temporal tolerances for automatic search etc.



At the same time, it was found out that in the compared arrays data completeness of stations with identical time and coordinates often varies. The software allows creating a file that contains information on measured parameters, depth range, and number of levels in each profile. The information can be used to form the most complete data set for every station of that kind.

In future, forming a joint oceanographic database on the Black Sea seems to be expedient. Now we are testing an automatic double search procedure comparing stations data values, location and date/time. According to preliminary estimates, the database will include above 185,000 oceanographic stations that will serve as an informational basis for analysis of climatic changes in the Black Sea and creation of various informational products using great experience in this field accumulated by the two institutions.

The work is carried out in the framework of the contract on creative collaboration between Marine Hydrophysical Institute, Russian Academy of Sciences, and All-Russian Research Institute of Hydrometeorological Information – World Data Center, and within the state task on theme No. 0827-2018-0002 “Operative Oceanology”.



A screenshot of the software operation

