Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 119 (2017) 315–324

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International Young Scientists Conference in HPC and Simulation, YSC 2017, 1-3 November 2017, Kotka, Finland

# Quality control and data restoration of metocean Arctic data

Jose Luis Araya Lopez*, Amir Uteuov, Anna V. Kalyuzhnaya

*ITMO University, 49 Kronverksky Pr. St. Petersburg, 197101, Russian Federation*

**Abstract**

We analyze the problem of data availability and how this situation (which is very common in practice) is a limiting factor for doing quality control and data restoration of MetOcean fields. In order to support our claim, we analyzed the case of data availability in the Arctic Region, since it is true that climate historical archives have always been scarce in this area, due to the harsh climate conditions and logistic difficulties that involve the deployment of in-situ weather systems in the region. In addition to the problem of data scarcity, we also have the problem of inconsistent and missing observations. In this paper, we highlight this problem by the analysis of 125 hourly temperature and hourly wind speed time series, which have been all in common their spatial distribution and present an evaluation of those data in terms of quality and availability. We also discuss the way the presence of gaps in time series can condition the ability of estimating synthetic data and discuss possible solutions to overcome this problem in areas where no neighboring stations are available.



Peer-review under responsibility of the scientific committee of the 6th International Young Scientist conference in HPC and Simulation

*Keywords: quality control, meteorology; Arctic region; surface air temperature; wind speed; data restoration.*

## 1. Introduction

The increasing amount of MetOcean data generated by observing weather systems involves a demand for having complete data and with an adequate quality control [1]. This is particularly true when dealing with climate change studies, or using these data for decision-making [2]. The goal of this work is to suggest methodologies for quality control and data restoration in the context of areas with low data coverage, such as we find in the northern regions. In particular, we are interested in suggesting a data integration approach that makes use of data generated by different sources, so that we can overcome the data scarcity problem to some extent. In order to achieve this, we propose the use of other data sources from reanalysis and model outputs as an aid to quality control and data restoration of real measurements. We also make a complete characterization of real datasets and characterize the problems that real-time series have.

* Corresponding author.

*E-mail address: jlaraya1@gmail.com*

1877-0509 

Peer-review under responsibility of the scientific committee of the 6th International Young Scientist conference in HPC and Simulation

10.1016/j.procs.2017.11.190

This paper is divided as follows: first, we define the problem statement and in what context the results of this work should be used. Second, we explain the data sources and the methods used in our study. Third, we present the analysis of the results that we have obtained so far. Finally, we make conclusions on the results and make some recommendations for further improvement.

**Nomenclature**

| | |
|---|---|
| ARCTIC | Arctic System Reanalysis (ASR) |
| ERA | ERA Interim reanalysis data (European Reanalysis) |
| JRA | JRA-55 (Japanese 55-year Reanalysis) |
| NCEP | NCEP reanalysis data (National Centers for Environmental Prediction) |
| NCEP FNL | NCEP FNL (Final) Operational Global Analysis |
| U | eastward component of wind speed |
| UTC | Coordinated Universal Time |
| V | northward component of wind speed |
| WRF | weather research and forecasting model |

## 2. Related works

The topic of quality control observations has a long tradition in Meteorology and Oceanography. The idea objective quality control has evolved over the decades hand in hand with rapid advances in microelectronics, computers, and communications technologies, which has produced an explosion of meteorological networks across all over the world [3]. The idea of objective quality control proposes the use of semi-automatic procedures that can isolate outliers from the datasets. However, the integral idea of quality control does not only involve the detection aspect of the problem, but the prevention aspect as well. This means that a number of measures have to be taken to prevent the outliers from being generated. The list of preventive measures is: proper station sitting, adequate routine maintenance, sensor calibration, archival of original data, use of standard time and observation units, use of similar instruments and configuration of sites and the redundant sensors [4].

There are different quality control algorithms and each one offers advantages and disadvantages. None of them resolves the problem definitely, they are good at spotting outliers based on different premises, and their usefulness lies on the possibility of using them together along with adequate professional judgment of the results. Sensor range based tests detect observations that are out of range according to sensor hardware specifications or theoretical/physical limits [5] [6]. Climate based range tests make use of historic climatic values to impose limits to the values. This approach has been applied by [7]. Temporal checks evaluate the rate of change in time series. The thresholds are more subjective than range tests and usually, they depend on the climate regime, which means that they have to be properly defined according to the problem in question. Step tests typically compare the rate of change among sequential observations. The threshold values used for step tests are also dependent on the station location, time interval between observations, meteorological variable, and tendency. In this category, one can define step tests and persistency tests [8]. Several of the quality control algorithms for the analysis of wind data was proposed by , who developed a sub-routine for the control of data quality of wind in hourly resolution. [6] have also performed quality controls for wind data at hourly resolution. Other quality control algorithms for high-resolution data have also been proposed by [9]. Finally, the spatial allow to detect observations that are inconsistent with data from neighboring sites. Data are compared to expected values (which are calculated using a spatial objective analysis algorithm). Observations that differ by more than a predetermined threshold value are considered to be non-typical.