International Journal of Data Analytics
Volume 2 • Issue 1 • January-June 2021

Data Analysis and Visualization in Python for Polar Meteorological Data

V. Sakthivel Samy, National Centre for Polar and Ocean Research, India
https://orcid.org/0000-0002-5352-3118
Koyel Pramanick, Pondicherry University, India
Veena Thenkanidiyoor, National Institute of Technology, Goa, India
Jeni Victor, Indian Institute of Tropical Meteorology, India

ABSTRACT

The aim of this study is to analyze meteorological data obtained from the various expeditions made to the Indian stations in Antarctica over recent years and determine how significantly the weather has shown a marked change over the years. For any time series data analysis, there are two main goals: (a) the authors need to identify the nature of the phenomenon from the sequence of observations and (b) predict the future data. On account of these goals, the pattern in the time series data and its variability are to be accurately identified. This paper can then interpret and integrate the pattern established with its associated meteorological datasets collected in Antarctica. Using the data analytics knowledge the validity of interpretation for the given datasets a pattern has been identified, which could extrapolate the pattern towards prediction. To ease the time series data analysis, the authors developed online meteorological data analytic portal at NCPOR, Goa http://data.ncaor.gov.in/.

KEYWORDS

Meteorological Datasets, Time Series Data, Weather Prediction

INTRODUCTION

It is well known that global warming (Jonathon, 2011) has led to significant climate changes. Empirical measurements of the *Earth's heat (Johnson et al., 2019)* content show the planet *is* still accumulating *heat* and *global warming is* still happening. The ice at both South Pole and North Pole encounters the first effect of any such climatic disturbances (Flohn, 1978). Thus, the Polar climate system is paid more attention to atmospheric scientific community as it is regarded to be sensitive to anthropogenic induced climate changes, ozone depletion and melting of ice shelf (Bintanja et al., 2013). The temperature difference between equator-tropical region and polar region is the major driver of the general circulation of whole atmosphere.

Observations of Automatic Weather Station (AWS) parameters / features like air Temperature (Temp), Air Pressure (AP), wind speed(WS), wind direction(WD) and relative humidity(RH) and also several other observations carried out at Indian research base stations such as Maitri (70°45'52"S, 11°44'03"E) and Bharati (69°24'41"S, 76°11'72"E) in Antarctica. Using these datasets collected over the years, data analytics and visualization have been performed through analysis methods like trend, behavior, seasonal and wavelet spectrum analysis. For this study, we used AWS datasets from Maitri station covering the period 2012-2015. These datasets are available in our National Polar Data Center (NPDC) - http://data.ncaor.gov.in/newhtml.

DOI: 10.4018/IJDA.2021010102

IDENTIFYING AND ELIMINATING ERRORS IN THE DATA

The statistical data of hourly temperature, wind speed, wind direction and air pressure used in this study are summarized in histograms (Figure 1). These statistical parameters are calculated separately using all yearly available datasets, polar day (from 10 November–24 January) and for polar night (from 19 May–27 July). Statistical properties of data summarized in the histogram of all year temperature (Figure 1a) is skewed to the left. The abrupt end of the right tail (Figure 1a) is associated with the melting barrier of the snow and ice in the region. In response to a favorable temperature increase, ice melting may take place. The associated modifications in the snow-ice surface energy budget lead to a limitation of temperature increase. The histogram of temperatures shows two relatively small local maxima which are related to the temperature values corresponding to cold and warm seasons in the region (Figure 1b & c). The temperature distribution for polar day (Figure 1b) is unimodal and asymmetric.

The distribution of temperature for polar night (Figure 1c) presents less asymmetry than the corresponding distribution for polar day (Figure 1b). Since cold season, the decrease in the frequency in the right tail of the histogram, towards positive temperature, is less abrupt than in the case of the polar day histogram. The histogram of polar night temperatures (Figure 1c) suggests also a possible multimodality in the distribution. The air pressure histograms all show bell-shaped distributions (Figure 1d–f) suggesting a Gaussian probability distribution function of this variable. During polar night, due to the absence of the solar radiation, the surface temperature is negative and hence the pressure shows various range from 970-980 hPa. The histograms of wind speed (Figure 1g–i) show a typical Weibull distribution. The double mode that is seen in wind speed (Figure 1g) is more likely to appear on polar night wind speed than polar day. The histograms of the wind direction (Figure 1j–l) show one prominent peak at about 250° that is related to the typical synoptic disturbances in the region, which are associated with the katabatic and supergeostrophic winds, respectively.

ANALYZING THE TIME SERIES DATA

In the time series data analysis it is often assumed that dataset has a systematic pattern, which are usually a set of identifiable components. It is also assumed that the dataset may consist of random noise. Together, these make the pattern in the datasets to identify. Thus the time series analysis techniques adopted must involve some form of filtering out noise or elimination of outliers in order to make the pattern more noticeable. As most time series patterns of meteorological data can be described in terms of four basic classes of components: ***trend***, ***behavior***, ***seasonality*** and ***wavelet spectrum analysis***, we have adopted the mentioned data analysis for this study. The trend of the data represents a general systematic linear or (most often) nonlinear component that changes over time and may not repeat or at least within the time range captured by our data. The behavioral analysis of the data represents the dependence of the various statistical parameters over one another. Those three general classes of time series components may coexist in real-life data. The wavelet analysis is adopted as it helps in the exploration of cyclic patterns of data. The purpose of this analysis is to decompose a complex time series, which may contain cyclic components, into a few underlying sinusoidal functions of particular wavelengths.

Trend Analysis

The trends in the data for several years can be used for prediction of future data. Moreover, the trends show how global atmospheric phenomenon has had an effect on the weather in the recent years. Trend analysis is based on the idea that what was the (non) linearity of the measured/observed parameter in the past that gives the scientists an idea of what will be the future projection of the same parameter.

Using the data from Maitri station at Antarctica, preprocessing or cleaning of data is done first using python. In this all the data sheets are read, the blank data or all Nan/NA are recorded as "-999"