esPARALLEL PROCESSING OF WIND SPEED DATA DURING YEARS 2012-2013 IN SHKODRA REGION

Sidita Duli, Faculty of Natural Sciences, University of Shkoder, Albania
Božo Krstajić, Department of Electrical Engineering, University of Montenegro

Abstract: *The Weibull distribution statistical tool is one of different techniques of analyzing the wind speed data. Estimating the Weibull parameters, when the sample of data is given, requires some mathematically complex calculations. Nowadays, most of these estimations are performed in grid systems, where the data can be processed in parallel from different processors. This helps reaching a faster result for the estimation. The aim of this article is to analyze the performance of the parallel implementation in MPI programming, applied on the concrete wind speed data of Shkodra city during the last two weeks.*

Abstrakt : *Shpërndarja statistikore Weibull është një nga disa teknikat që përdoren për të analizuar të dhënat e shpejtësisë së erës. Vlerësimi i parametrave të Weibull, për një mostër të dhënë, kërkon përllogaritje komplekse matematikore. Në ditët e sotme, shumica e këtyre vleresimeve realizohen në sisteme grid, në të cilët të dhënat mund të përpunohen në disa proçesorë njëherësh. Kjo ndihmon në marrjen e një rezultati më të shpejtë. Qëllimi i këtij artikulli është analizimi i performancës të implementimit në programimin MPI, i aplikuar për të dhëna konkrete të shpejtësisë së erës gjatë dy viteve të fundit.*

1 INTRODUCTION

The analysis of the wind speed is useful in many fields of industry, in agriculture and in meteorology. Generating energy from the wind power is a mature technology. This is the main reason that highlights the important role of statistics of the wind speed in a specific location.

The Weibull distribution is a statistical tool used to model the wind speed. The wind speed data collected in a weather station tend to have the form of an Weibull distribution. The shape of the graphics depends on the parameters of the distribution, which are estimated by the set of wind speed data. Once the Weibull distribution can be used to calculate the probability of a particular wind speed at a particular location, it can be used to work out the number of hours per year that certain wind speeds are likely to record and therefore the likely total power output of a wind turbine per year [1].

The Weibull distribution exists in two main forms: the two-parameter and three-parameter Weibull distribution. But the study is focused in the two parameters form, as it is the form applied in meteorology.

The two parameter Weibull Distribution has the following density and distribution functions:

$$f(x) = \left(\frac{b}{a}\right)\left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}\right)^{b}} \quad (1)$$

$$F(x) = 1 - e^{-\left(\frac{x}{a}\right)^{b}} \quad (2)$$

The parameter is the Weibull scale parameter in m/s; a measure for the characteristic wind speed of the distribution. This parameter is proportional to the mean wind speed. The b is the Weibull shape parameter. It specifies the shape of a Weibull distribution and takes on a value of between 1 and 3. A small value for b signifies very variable winds, while constant winds are characterized by a larger b [2].

2 METHODS

The application makes an estimation of the parameters of the Weibull distribution function. Two main methods that can be used in this estimation are the maximum likelihood and the least squares. In the implementation used in this study it is applied the maximum likelihood method, as it is usually considered to be more robust and produces more accurate results.

The computer application that estimates the Weibull parameters takes too much time to produce a result, especially when the sample contains too much data. This is one reason to use the grid architecture and to process the data in multiple cores simultaneously. The principle of parallel programming is to divide the problem in smaller pieces and to distribute the calculations between many processors. Using the technique of parallelizing the code by two different processors that use the same memory space, it makes possible that complex tasks might be computed in a shorter time as they were computed by a single processor architecture.

There are different ways of parallelization the code of this application. Two of them are the message passing interface (MPI) and Posix threads (Pthread). In this research it is analyzed the MPI implementation of the parameter estimation of Weibull parameters.

MPI is a standard developed by the Message Passing Interface Forum (MPIF). It specifies a portable interface for writing message-passing programs, and aims at practicality, efficiency, and flexibility at the same time. MPI is implemented on a great variety of machines, including those "machines" consisting of collections of other machines, parallel or not, connected by a communication network [3]. The programs using MPI libraries may run on distributed-memory multicomputer, shared-memory multiprocessors, networks of workstations, and combinations of all of these.
The algorithm is implemented in the C programming language. It is imported the mpi.h library which contains the parallel procedures and function of the communication and passing the messages between cores of the system. The system where it is executed is a grid cluster, a Linux based system and portable enough to run with consistent result any implementation of message passing model.
The application takes a censored set of data, which might be ordered or unordered, taken from a sample of N data. The data analyzed in this research are the wind speed data collected from the weather station of the team of learning technique of University “Luigj Gurakuqi” for a two year period 2012-13. It is given the location parameter of the Weibull distribution. The data sample and the location parameter are read from a file by each process that is created. The algorithm makes an estimation of the scale parameter and shape parameter of this distribution.

4.RESULTS AND DISCUSSIONS

It is of interest to study the performance of this implementation in parallel of Weibull parameter estimation for a concrete wind speed data in our location. The weather station is installed in the city of Shkodra. It measures the wind speed every 30 minutes. Also it is of interest to analyze the distribution of the wind speed of the last two years. For this reason the parallel implementation is applied to three special data samples. The first data sample is the wind speed data of the 2012. This sample contains 17568 data. The second sample is the wind speed data during the year 2013, and contains 17520 data. The third sample is that containing both wind speed data of years 2012 and 2013. It contains 35088 data.
The samples in these tests are not the best to show the parallel programming advantage of decreasing the time of calculations because the samples are not containing too much data to be complex calculations. One reason of performing the tests on these samples is because of the meaning of the data, as they are concrete wind speed data of a particular location, and not random data. This is a first step to a concrete application of this parallel version.
The MPI parallel implementation is tested for different number of cores, from two up to twelve cores. The results are compared to the serial version represented one cored version in the above tables.

a) Tests of performance when the data sample contains wind speed data of the year 2012 in Shkoder

Cores	1	2	3	4	5	6	7	8	9	10	11	12
Time in seconds	35.7	19.1	13.6	9.8	7.7	6.3	5.4	4.8	4.3	3.9	3.6	3.5
Speed-up	1	1.86	2.62	3.64	4.63	5.66	6.61	7.43	8.3	9.15	9.91	10.02

Table 1: Performance in time and speed up of the MPI implementation proccessing wind speed data during 2012.

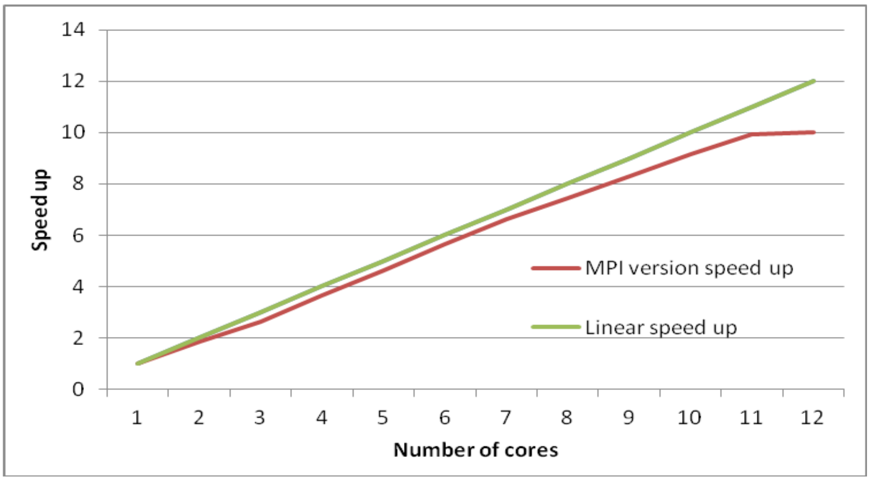

Figure 1 : Speed up of the MPI version for the sample wind speed data of the year 2012

Results show that the time spent to perform the calculations is reduced while increasing the number of cores. For this estimation, the serial version is executed in 35.7 seconds. The table1 shows the time spent to estimate the Weibull parameters of a dataset containing 17568 data. For this dataset, it is shown that the MPI version reduces the time spent to estimate the result. Executing the serial program it is needed 35.7 s to get the parameters of the Weibull distribution for this sample. There results show how parallel programming helps in reducing this time. The MPI technique the result 1.86 times faster as soon as a second processor is added. The time is reduced up to only 3.5 s when is executed the MPI version in 12 cores.

This table shows the speed-up graphic for these performance tests. This speed up is the number of times that the parallel implementation is faster than the serial implementation. The tests show that in this case the speed up increases in a linear form. According to the Gustavson law [4], the linear graphic depends also by an alpha coeficient which is the non-parallelisable fraction of any parallel process. The efficiency shows how well-utilized are the processors in the parallel implementation. It is a value between zero and one. A parallel efficiency of one corresponds to ideal, linear speedup. In this case, the alpha coeficient is small, because the implementation is made with a small fraction of non-pallelizied.

Results of parameter estimation of Weibull distribution for this sample show that the shape parameter is b=3.358281 and the scale parameter is a=2.628137.

b) Tests of performance when the data sample contains wind speed data of the year 2013 in Shkoder

Cores	1	2	3	4	5	6	7	8	9	10	11	12
Time in seconds	34.9	18.6	13.3	9.3	7.3	6.1	5.1	4.6	4.2	3.9	3.6	3.4
Speed-up	1	1.87	2.62	3.75	4.78	5.72	6.84	7.58	8.31	8.94	9.69	10.26

Table 2: Performance in time and speed up of the MPI implementation proccessing wind speed data during 2013.

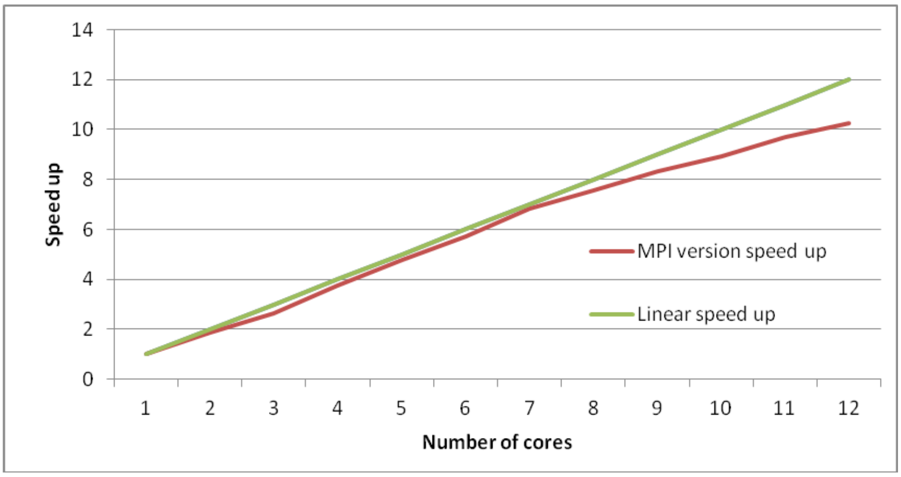

Figure 2 : Speed up of the MPI version for the sample wind speed data of the year 2013

For this estimation, the serial version is executed in 34.9 seconds. The table 2 shows the time spent to estimate the Weibull parameters of a dataset containing 17520 data. The serial implementation needs 34.9 s to get the parameters of the Weibull distribution for this sample. Time is reduces as soon as the second procces is added. When the number of cores is two, it takes 18.6 s to perform the calculations, and the speedup is 1.87. As the number of cores is increased, it reaches the best performance in the seventh core, as it is almost linear speedup.

Results of parameter estimation of Weibull distribution for this sample show that the shape parameter is b=3.358281 and the scale parameter is a=2.628137.

c) Tests of performance when the data sample contains wind speed data of both years 2012 and 2013 in Shkoder

Cores	1	2	3	4	5	6	7	8	9	10	11	12
Time in seconds	66.7	34.3	23.3	17.2	14.7	12.6	10.5	9.4	8.1	7.3	6.6	6.4
Speed-up	1	1.94	2.86	3.87	4.53	5.27	6.35	7.09	8.23	9.11	10.1	10.4

Table 3: Performance in time and speed up of the MPI implementation proccessing wind speed data during both years 2012 and 2013.

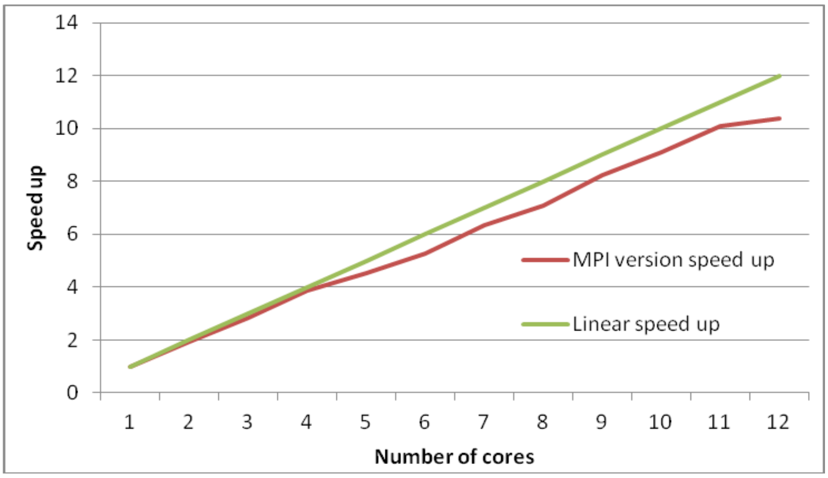

Figure 3 : Speed up of the MPI version for the sample wind speed data of the year 2012 and 2013

For this estimation, the serial version is executed in 66.7 seconds. The table3 shows the time spent to calculate the Weibull parameters of a dataset containing 35088 data. The serial program needs 66.7 s to get the parameters of the Weibull distribution for this sample. The time is reduced up to only 6.4 s when is executed the MPI version in 12 cores. The figure 3 shows that the speed up decreases when the number of cores is larger, but still it tends to be close to the linear speed up.

Results of parameter estimation of Weibull distribution for this sample show that the shape parameter is b=3.570758 and the scale parameter is a=1.936218.

5.CONCLUSIONS

The paper introduces the benefits of implementing the parallel processing in analysing concrete data in meteorology.
In this case the Weibull distribution parameters are estimated to represent the distribution of the wind speed data for two last years. It is tested the performance in time of the MPI parallel version. Also it is measured the speed up of the execution, from two up to twelve cores. Time of the estimation is reduced as soon as adding the second core. When the application is run in twelve parallel processes, time is only few seconds.
The cost of adding new procceses, instead of using only one process in the serial version, is the complexity of the code, and the effort of the programmer to make it work by using the message passing. Adding new procceses slows the execution as a part of this time takes the communication between procceses.
As future work might be testing the parallel implementation with wider range of concrete data, which might be including more years in analyze, or comparing different locations. In this way, having wider range of data, it is highlighed the importance of introducing the parallel programming in the statistical estimations.

6. REFERENCES

1. http://www.reuk.co.uk/Wind-Speed-Distribution-Weibull.htm.
2. The Swiss wind power data website http://wind-data.ch/
3. Y.AOYAMA, J. NAKANO: *Practical MPI Programming*, *International Technical Support Organization,* 1999
4. T. RAUBER, G. RUNGER: *Parallel Programming for Multicore and Cluster Systems*. Springer, 2011.