Contribution ID: **13** Type: **Poster Presentation**

Importance of sampling strategy on diffuse solar energy prediction in Cameroon

A key role of the modeling process is to identify the relationship between the inputs and the target variable(s) (outputs) set to study a given phenomenon and also to predict the outcome of that phenomenon. In the context of machine learning, in addition to the concern about feature selection, there is a second concern about the sample size needed for model design.
However, in the literature only few projects have focused on determining optimal sampling approaches. Thus, in this study, the objective is to propose a robust validation procedure for the daily prediction of direct normal irradiation under clear sky conditions (DNI) in Cameroon. So, considering the data of the different types of aerosols (sea salt, sulphate, desert dust, black carbon and organic) and other meteorological variables collected and 181 points of the study area, an automatic learning model based on the Extreme gradient boosting (XGBoost) has been designed using a training set collected from the 9 different clusters.
After what we invested in proposing some optimal approaches for the selection of the validation set based on the derivatives of the K-median. In this respect the methods named maxkmed4 and maxkmed5 have respectively allowed to form a more and less compact data set independently of the sample size. It was also concluded that for a local prediction the sample obtained from maxkmed5 will be the best adapted because being the less compact, it reflects the most the characteristics of the study area according to the test n 7 different areas.

Author: DOUANLA ALOTSE, yaulande (Institute of Mathematics and physics of Porto-Novo, Benin)

Presenter: DOUANLA ALOTSE, yaulande (Institute of Mathematics and physics of Porto-Novo, Benin)

Session Classification: Poster Session

Track Classification: Ambient air pollution