Improving Temporal Resolutions of Drought Predictions Using Knowledge Discovery from Database Approach

Getachew Berhan Demisse, Addis Ababa University-Ethiopia, getachewb2012@gmail.com
Samuel Sisay, Jima University, Ethiopia, sis.samuel@gmail.com
Tsegaye Tadesse, National Drought Mitigation Center, University of Nebraska-Lincoln, USA, ttadesse2@unl.edu

Abstract

Attributed to climatic change and uncertainty of weather conditions, drought has become a recurrent phenomenon in many parts of the world. Even though there had been a number of efforts for predicting drought in the past, there were limited approaches for improving the temporal resolutions. Taking this gap into account, the objective of this article was to develop an approach to improve the temporal resolutions of drought predictions using knowledge discovery from database (KDD) approach. The focus of the study was on using higher temporal resolution datasets from satellite, climate, oceanic and biophysical domain sources. A total of 24 years historical data from the years 1983 to 2006, with spatial resolution of 8 km, for selected attributes were used in constructing regression tree models. For constructing the models, 80% of the data were used for training and 20% for testing. After improving the temporal resolution from 30 days to 10 days temporal resolutions, the time lag prediction accuracy has significantly improved from 95% to 99% correlation coefficient. From these experimental analyses, it was concluded that high temporal resolution dataset has a significant impact on the overall accuracy of the prediction models. The result can be used by decision makers at different levels for their early update of drought episodes for their smart decisions in mitigating drought.

Key words: Dekadal Data, Drought Monitoring, Drought Prediction, Model, Regression Tree

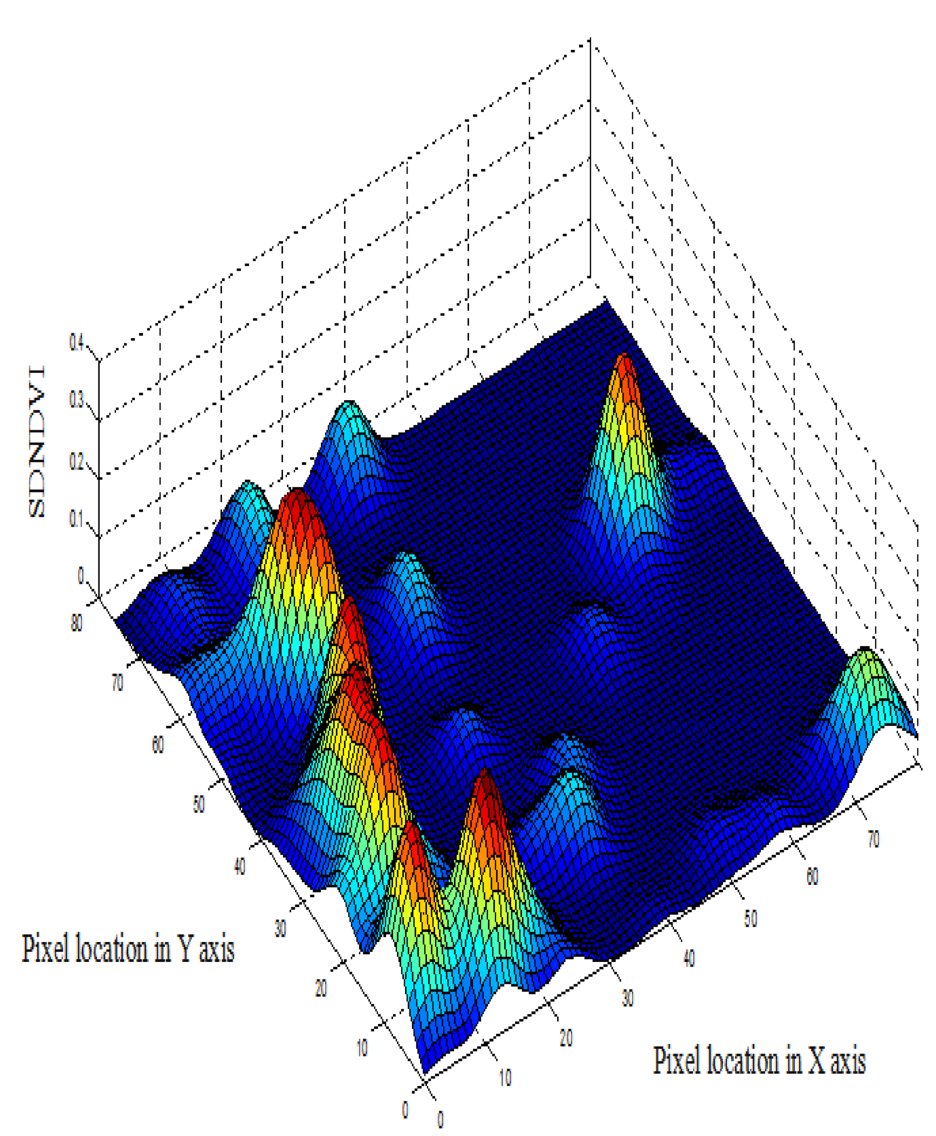

Figure 1: Research conceptual framework.

Methodological Approach

The study area for this research was Ethiopia, which occupies the interior of the Horn of Africa stretching between 3° and 14° N latitude and 33° and 48° E longitude, with a total area of 1.13 million km^2. A total of 11 attributes were iteratively identified for modeling drought. The modeling approach in this research was conducted with the assumption that it is possible to identify drought in space-time dimensions as spatial object (Figure 1). With this assumption, for characterizing drought a total of 11 attributes were selected using selection criteria: AIC (Akaike's information criterion), VIF (variance inflation factor) and Moran's I index. The AIC was used as a model performance by including a given potential attribute; the VIF as parameter for controlling the duplication of the information of the potential attribute with previously selected attributes; and Moran's I index as a parameter for controlling absence of key attributes and for avoiding misspecification of the drought model.

The overall methodological approach for modeling drought object is presented in Figure 2. The objective of the study was to develop an approach to improve the temporal resolutions of drought predictions using KDD approach. The focus of the study was on using higher temporal resolution datasets from satellite, climate, oceanic and biophysical domain sources.

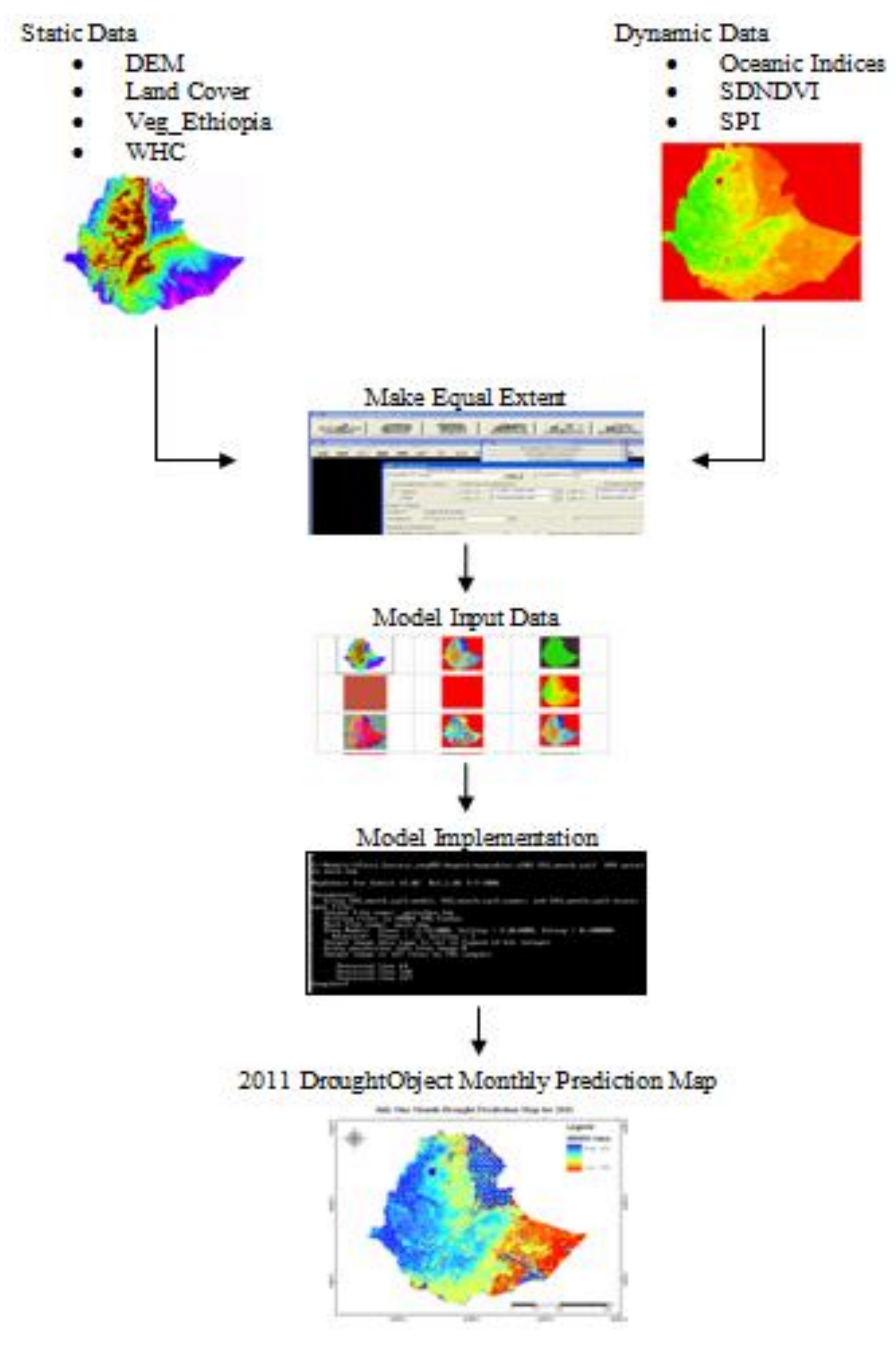

Figure 2: Workflow in the new drought modeling approach.

System Overview

Design Science problem solving approach: Design science in information system has two processes (build and evaluate), and four design artifacts (constructs, models, methods and instantiations). Purposeful artifacts are built to address heretofore unsolved problems, such as modeling drought with object-oriented approach. In the modeling processes (Figure 3), the mathematical relationships of the 11 independent attributes and the dependent attributes were built, and iteratively evaluated (Figure 4). In this process, the constructs are the mathematical combinations of the 11 attributes, the models are the regression equations for retrieving drought patterns from satellite images, the methods are the overall processes of getting the drought objects identified and segmented from satellite images (including all algorithms used). Finally, the instantiation in this research is that all the constructs, models and methods were implemented.

Data from 1983 to 2006 for the 11 key attributes were used for the drought modeling experiment. The years 1983-2006 were used because the complete dataset for the 11 attributes was obtained in these time periods for the whole modeling exercise to develop the knowledge base using machine learning. During the analysis, the machine leaning experiment produced 10–30 models. The criteria for the selection of rules were the accuracy level (quantified by correlation coefficient) and the number of absolute and relative errors. The interpretation for Rule 1 below is that among the 53,990 training cases (80% of the dataset), 124 cases satisfy all three conditions.

Rule 1: [124 cases, mean -59.0, range -251 to 163, est err 62.0]
if
SPI_3month <= -20
SDNDVI <= 0
Landcover in 14 (rain fed croplands), 20 (mosaic cropland)
then
Drought = 57.6 + 0.83 SPI_3month - 0.023 DEM + 0.15 SDNDVI

In this case, drought values range from -251 to 163, with an average value of -59.0. The regression tree model finds that the target value of these or other cases satisfying the conditions can be modeled by the formula: *Drought = 57.6 + 0.83 SPI_3month – 0.023 DEM + 0.15 SDNDVI.*

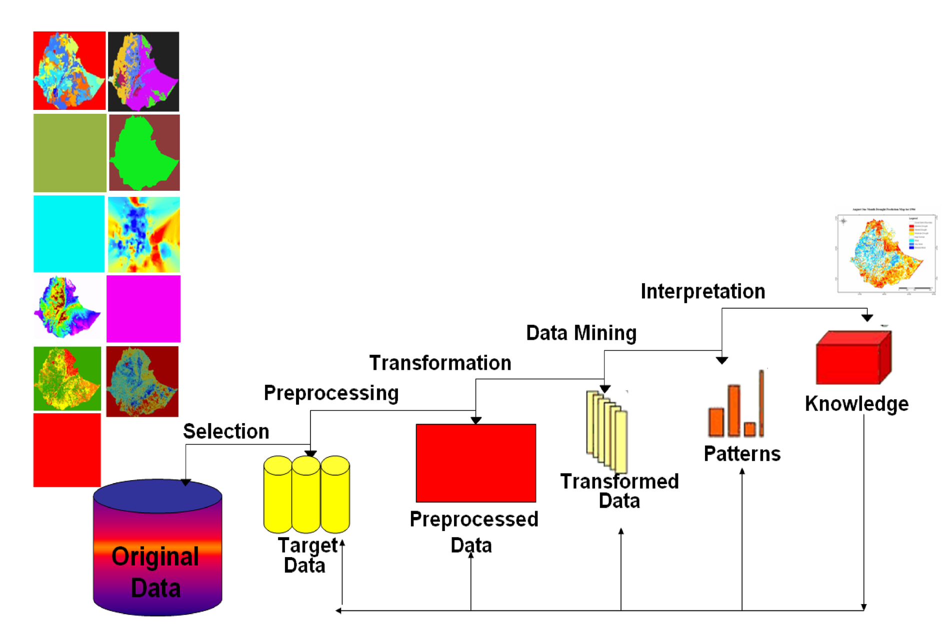

Figure 3: KDD processes in drought modeling.

Conclusions

After improving the temporal resolution from 30 days to 10 days temporal resolutions, the time lag prediction accuracy has significantly improved from 95% to 99% correlation coefficient. From these experimental analyses, it was concluded that high temporal resolution dataset has a significant impact on the overall accuracy of the prediction models. The result can be used by decision makers at different levels for their early update of drought episodes for their smart decisions in mitigating drought.

Drought information extraction framework: Figure 5 presents drought information extraction design, which is a four-tier framework including a data processing component, knowledge base triggering component, drought identification component and knowledge construction and presentation component. Cloud computing approaches in different platforms are being tested for integrating these different components. The method that is developed here is expected to help in exploiting the large datasets available from biophysical, climate and satellite observations for environmental management. It is also expected that the method can be directly applied in scientific data mining of huge amounts of space data gathered by NASA for environmental resources assessment and management.

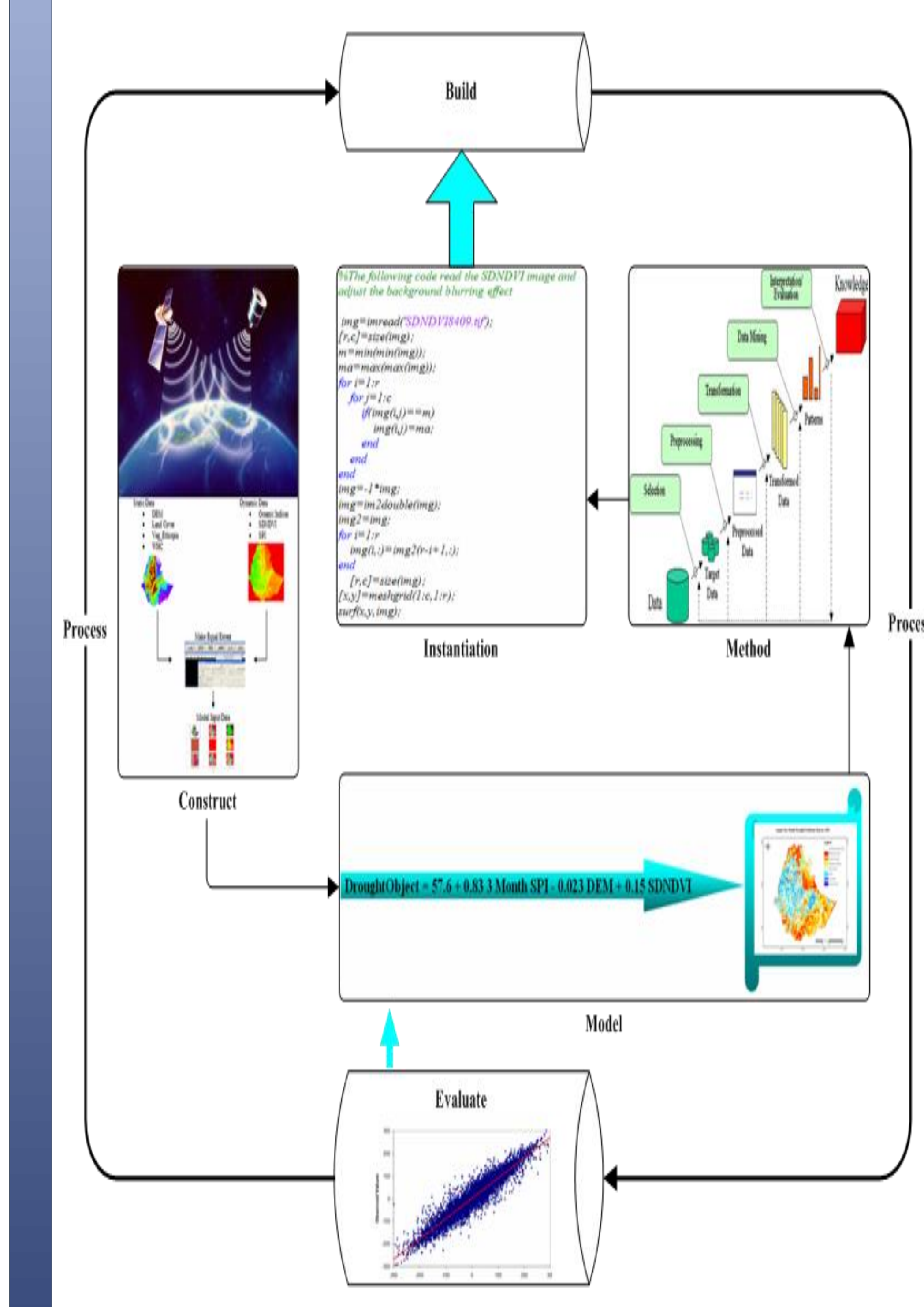

Figure 4: Design science problem solving processes.

Future Research

Future research may focus on investigating and developing a web-enabled big data management approach for improved drought data analysis and information delivery for decision makers at different levels.

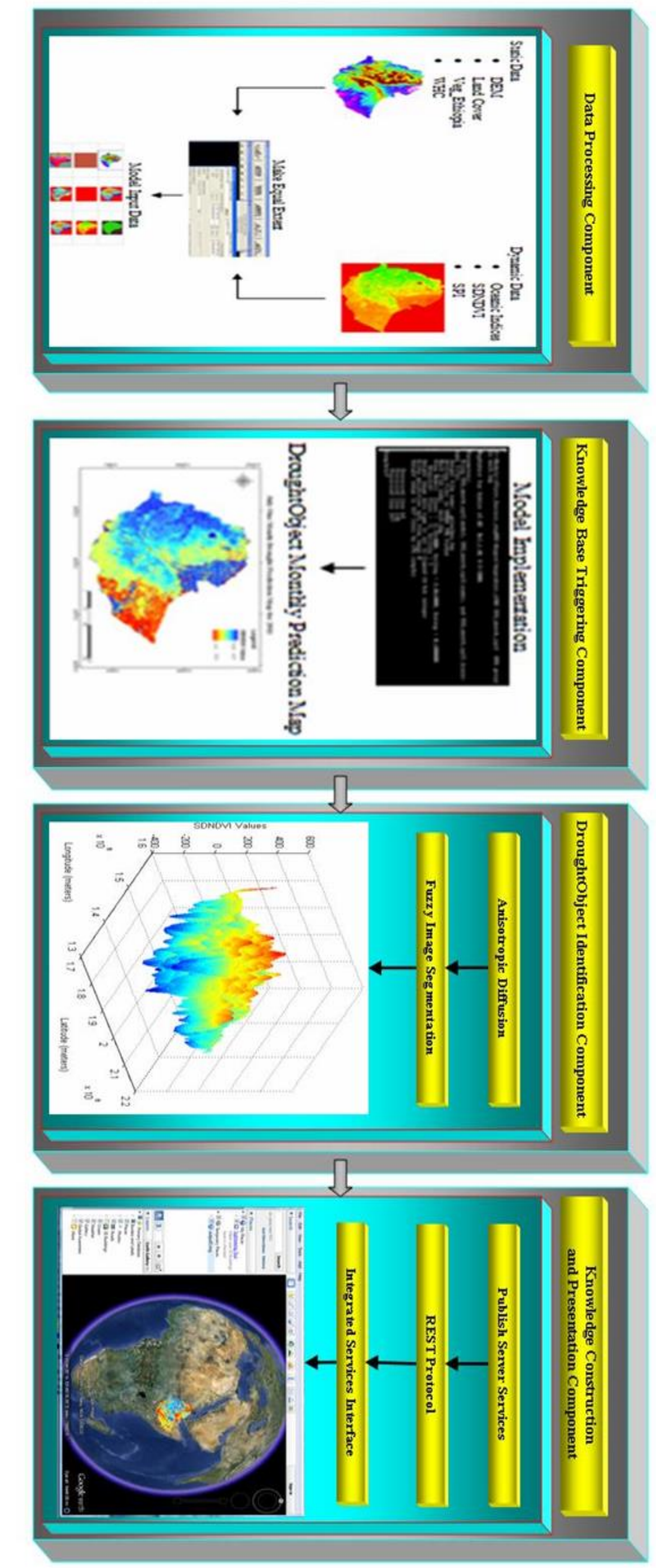

Figure 5: Drought information extraction framework.

For More Information:

Getachew Berhan Demisse
E-mail: getachewb2012@gmail.com
Tel: +251 911 638408
P.O.Box 1176
Climate Science Center,
Addis Ababa University, Ethiopia