LINKING EARTH OBSERVATION DATA AND SPATIAL DATABASE FOR MONITORING AND ESTIMATING SUGARCANE YIELD

Michael Andrew Manalili[1, 2*], Trisha Leigh Lunas[2], Melany Palpal-latoc[2], Crispin Tar-odan[2]
[1]Training Center for Applied Geodesy and Photogrammetry, University of the Philippines, Diliman, Quezon City 1101
[2]Department of Geodetic Engineering, University of the Philippines, Diliman, Quezon City 1101

mikelmanalili21@gmail.com, tloleighlunas@gmail.com, lanypalpallatoc18@gmail.com, chris.tards@gmail.com

Keywords: *Spatial database, sugarcane, yield, UML, postgis*

ABSTRACT

The Philippine Sugarcane Regulatory Administration (SRA) is mandated to promote the growth and development of the sugar industry in the country. One of its objectives is to develop a geospatial information database covering the entire areas of operation. A spatial decision support system for the organization is necessary to address environmental and economic considerations for sugarcane yield estimation and sugar price regulation. Diverse environmental, spatial, temporal and social factors results in varying data types, I/O and information extraction and thus limits the possibility being able to query large database of related data where decision can be based. This research proposed a Unified Modelling Language (UML) design to aid the organization in data creation, input and update. This research used PostgreSQL as the Object Relational Database Management System with PostGIS extension to handle spatial objects. The organization responded very well to this simple yet powerful object relational and spatial database structure and has helped them retrieve and query vital information necessary for everyday decision making.

1. INTRODUCTION

The Sugarcane Regulatory Administration (SRA) is an attached agency of the Department of Agriculture of the Philippines with the mandate to promote the growth and development of the sugar through Research and Development (R&D) for continual improvement of the industry. As part of its R&D work, SRA maintains and develops a geospatial information for all its areas of operation and stakeholders. For the longest time since the establishment of the organization, COBOL was the main database being used by SRA to store and manage data. The existing database structure has difficulty and challenges in handling spatial information and thus this limits geographically represented remote sensing and field data to correlate with actual mill data.

Figure 1 shows the general workflow on how earth observation data will be integrated with field, meteorological data and store it in a central database to resolve various decision making processes. Since SRA depend heavily on geospatial information both historical and present data collected to produce a yield data estimates, a relational spatial database is needed to unify all information. This research proposed a conceptual and physical model to aid the organization in data creation, I/O, and update and in performing query operations.

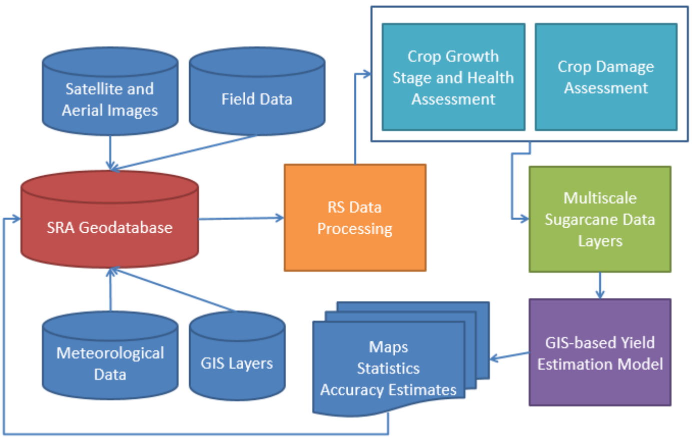


Figure 1. The general workflow for Sugarcane Yield Estimation

1.1 REVIEW AND RELATED LITERATURE

The sugar industry provides substantial socio-economic contribution to the Philippines. According to the SRA website, in the past five consecutive crop years (2010-2015), the Philippines has produced an average of 2.4 million metric tons of sugar which is sufficient to cover the traditional domestic demand of 2.1 million metric tons annually. One of the main functions of the SRA is to institute an orderly system in sugarcane production for the stable, sufficient and balanced production of sugar for local consumption, exportation and strategic reserve.

There are studies that discussed several strategies or methodologies to estimate crop production or yields. Like for instance, Fermont and Benson, 2011 in their research entitled *"Estimating Yield of Food Crops Grown by Smallholder Farmers: A Review in the Uganda Context"* enumerated wide range of methodologies that has been developed to estimate crop production, area, and, ultimately, yields in the fields of smallholder farmers. The information collected has been collated into a crop yield database which is accessible thru excel file upon request from IFPRI-Kampala office. The database gives full information on recorded crop yields and still subject for development.

Other researchers explored the suitability of the Normalized Difference Vegetation Index (NDVI) from the Moderate Resolution Imaging Spectrometer (MODIS) obtained for six sugar management zones, over nine years (2002–2010), to forecast sugarcane yield on an annual and zonal base. Mulianga et al, 2013 concluded in their research that remote sensing technology together with environmental information has potential to be used to estimate crop yield and evaluate the impact of environmental conditions to crop production as opposed to physical methods. *(Forecasting Regional Sugarcane Yield Based on Time Integral and Spatial Aggregation of MODIS NDVI)*

Yield estimation do not depend solely on the methods or strategies on how data and information are collected and validated but it is also affected by how organized the data in a database. SRA as an organization depends heavily on geospatial information both historical and field data collected to produce a yield data estimate for the year, an organized spatial database is required to unify all information not only with respective locality or mill district of operation but for the whole country in which they operate. This can be done through the aid of a Unified Modelling Language (UML).

2. OBJECTIVES

To be able to design a spatial database for data standardization and for estimating sugarcane yield

Specifically, the objectives of this research are to:

a. Standardization of data names and types
b. Create a conceptual model for Sugarcane Yield
c. Implement conceptual model to meet the needs of stakeholders
d. Generate a standard spatial database design for sugarcane yield estimation
e. Perform spatial analysis and table queries

3. METHODOLOGY

This research presents a conceptual and logical model for spatial database design to address varying data types, storage capability and access to large database. Non-spatial database related to yield data has also been integrated to be able to address the needs of different stakeholders. An open source spatial database (PostgreSQL) with spatial functions (PostGIS) was used to handle raster and vector data and the PostgreSQL function to access relational data. A use case scenario was carried out between researchers and the stakeholders to be able to design the functionalities and relationships of objects and entities within the object-relational database management system (ORDBMS). Finally, vegetation indices, meteorological data fromearth observation raster and GIS vector data and attributes were loaded into the open-source software to test spatial queries that is required by the end user.

3.1 USE CASE SCENARIOS

A use case scenario was conducted between the researchers and the stakeholders in order to capture the all the concerns of the end users of the spatial database. Most of the use cases involves asking basic questions and how the spatial database will be able to address basic data and information extraction from related data both from field acquired and remotely-sensed data. The SRA as the regulatory body will be the end user of the spatial database and the mill district officers will be the secondary users to which majority of the inputs and updates will come from. Among the use case, the most important information that can be easily useful to end user are the vegetation indices (NDVI and EVI) as these can already be correlated to sugarcane health, vigor and temporal analysis can be carried out.

3.2 DATA COLLECTION AND CLEANING

Data were collected from various sources but bulk of the data for this research came from the SRA's field personnel. To be able to address problem on data handling and standardization, the rawest form of spatial and non-spatial data were requested to SRA. Bulk of the time were spent on data standardization and data types and formats were standardized for all sites.

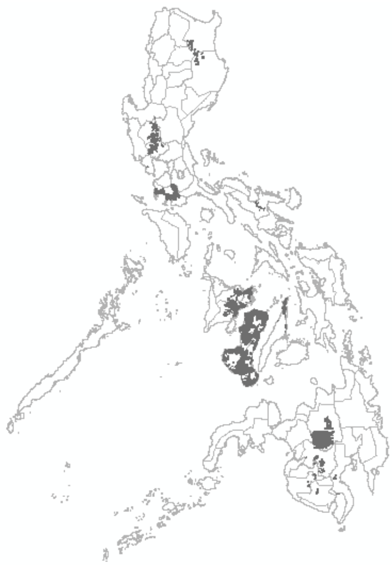


Figure 2. Distribution of sugarcane plantations in the Philippines

Figure 2 above shows six (6) major areas of operations. The sugarcane plots are digitized using a high-resolution satellite imagery (SPOT 5) and was verified using google earth and field visits. All sugarcane plots have spatial and non-spatial attributes with varying data entry and character sets between and among sites of operations. Basic information include area of sugarcane plot, Name of planter, crop variety, farming style and plots planted with cane (figure 3).

GIS_IIDD	PL_NAME	Shape_Leng	Frame_ID_1	GIS_ID	Area	Planter_Na	Variety
386	MERCADO, GERTRUDEZ	520.770365	106	386	1.66	MERCADO, GERTRUDEZ	PHIL 1793-99
387	MERCADO, GERTRUDEZ	572.361941	106	387	1.94	MERCADO, GERTRUDEZ	PHIL 1793-99
388	MERCADO, GERTRUDEZ	367.351471	106	388	0.75	MERCADO, GERTRUDEZ	PHIL 1793-99
389	MERCADO, GERTRUDEZ	295.948063	106	389	0.49	MERCADO, GERTRUDEZ	PHIL 1793-99
390	AMURAO, LOLITA	561.704252	106	390	1.28	MERCADO, GERTRUDEZ	PHIL 1793-99
392	GAMBOA, DOLORES	406.471325	106	392	0.88	MERCADO, GERTRUDEZ	PHIL 1793-99
393	MERCADO, GERTRUDEZ	244.070675	106	393	0.29	MERCADO, GERTRUDEZ	PHIL 1793-99

Figure 3. Sample raw data provided by the mill district officer

4. RESULTS AND DISCUSSION

4.1 RS Derived Metrics

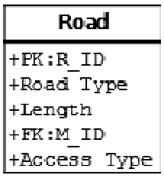
RS Derived Metrics
+PK:RID +Landsat NDVI +Landsat EVI Value +MODIS NDVI (t) +fPAR +Leaf Area Index +LST +SAVI +SAR Area Estimates (t) +SAR Inundation

Vegetation Index is one of the critical considerations being used by the SRA to determine plant growth, infestation and damage due to typhoon or extreme condition. The SRA uses free and available satellite images to generate vegetation indices of a plot per mill district or province of operations. This vegetation index is being collected, processed and analyzed to provide useful information in sugarcane phenology in particular. The Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) are one of the few indices utilized for growth maturity and age extraction of the annual sugarcane crop. This comes in a time series of vegetation analysis and is being explained to farmers by field technicians. This RS based metrics will feed in values to the sugarcane plots but is limited to the detection limit of the spatial resolution of the image being referenced to.

4.2 Sugarcane Plots

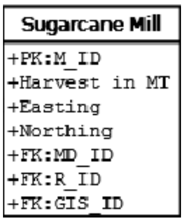
Sugarcane class is the most important class in this conceptual modeling as this contains most of the information that will feed into the mill and eventually will determine the yield or production. The sugarcane has attributes of its location, area of coverage, date of planting, the status of the crop as noted by the farmers and the vegetation index specific to the plant area or plot. This class is the smallest unit of carrier of information from the database represented as a polygon with which the canes are planted.

4.3 Road



The road class contains information on how the sugarcane plot connects to the mills. The spatial data collected for this research did not include the entire farm roads within the farm lots as this is still a work in progress. This class is vital as we want to determine the type of access road to the sugarcane areas for harvesting, planting and monitoring purposes for operation.

4.4 Sugarcane Mill

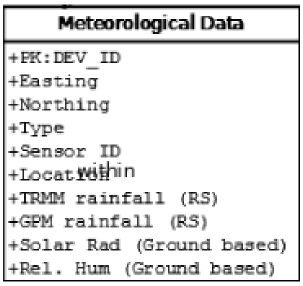


Mill class was identified to be included in the conceptual model as mills produce different harvest at different mill districts. The location of the mill also determines proximity to the sugarcane plots and will have an impact on the transportation cost of harvesting and planting. The class mill also can produce data on harvest in the rawest form and can be compared from one mill to another.

4.5 Sugarcane Plots

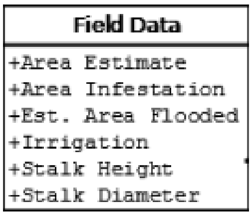
This class is the largest unit of analysis in the conceptual model. Mill district aggregates data from mills where in sugarcane data are feed in. The mill district also carries administrative boundary information for the sugarcane locations for legal and administrative purposes. Climate also is aggregated in this class as climate doesn't change so much from sugarcane plot to another. Climate information is carried by this class as comparison of yield/harvest will be done on a mill basis which may be in one or more mill districts. Soil types are also aggregated in this class as local data on soil have not been acquired by the organization and that regional soil classification is used as basis for delineating soil classification bounded by the mill districts.

4.6 Meteorological Data



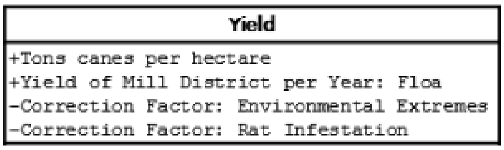
This class will be a floating class from the rest of the database. Floating because the data coming from this class is not controlled by the organization but from a third party software or application. This class will be an important factor that will feed data to the mill district to be able to predict weather scenarios that might help explaining how yield is affected by environmental processes. The type attribute in this weather station will determine the frequency and type of information that can be used from the nearest weather station that would approximate the climatic condition affecting the sugarcanes at a point in time.

4.7 Field Data



The organization collects a number of field data across the country. This data are saved in own personal desktop and hard drives and thus no network serves as a bridge to access the field data within and among mill districts. The geospatial data is manually edited and saved according to assigned field personnel. Data type varies floating, integer and varying characters without standard naming convention or coding system that would identify who, when and where the data points to at a given time. This field data is summarized by mill districts but data organization. This primary field information being vital input to the yield estimation system should be well documented in the database system and the figure above shows some of the information that need to be included.

4.8 Yield



This conceptual design integrates all the data within the database. This is basically a single formula that will estimate the yield of sugarcane per ton canes at plot level using Remote Sensing derived metrics coupled with meteorological and field data. Correction factors like infestation and environmental extremes not directly detectable by remotely-sensed information are being estimated spatially by field personnel.

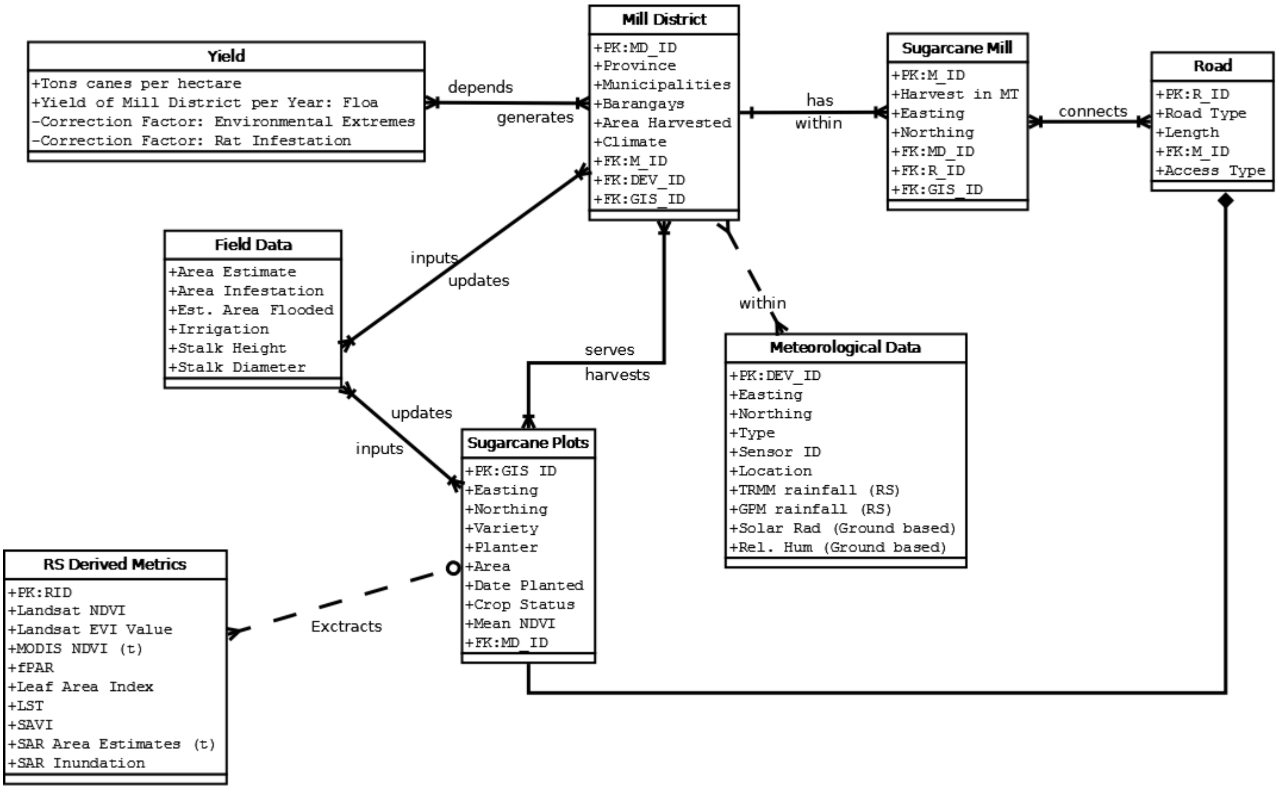


Figure 4. Conceptual / Logical Model design for Sugarcane Harvesting

As illustrated in figure 4, the relationships of each class are explained in this conceptual model. Beginning from the earth observation data where various vegetation metrics and meteorological data will be extracted from the raster to the sugarcane plots, this shows a partly weak association in anticipation of the unavailability of RS imagery and/or when visibility due to cloud cover does not permit full coverage of the area. Since the database is designed to be dynamic for multi-sourced RS images, vegetation metrics can be derived from high spatial and/or high temporal resolution images. The vegetation indices has a capability to handle historical datasets through time series observation using MODIS Terra data. SAR data will be primarily used to determine total area planted since SAR datasets are not affected by clouds and can acquire even at night time conditions. The sugarcane plot class will serve as the smallest carrier of information for this database. RS derived metrics and GIS data populated by field observations and meteorological data will be an input to the sugarcane plots for further analysis in the yield estimation model at the mill district level.

Meteorological data class has a weak relationship with the mill district class because of data availability and time delay before ground measurements are acquired and directly explain phenological response of vegetation to climate. Although the mill district has its own automatic weather stations, the local variability in weather play an important role in sugarcane growth. This limitation in ground measurement can be complimented by RS data like Tropical Rainfall Measurement Mission and Global Precipitation Model (TRMM and GPM).

Mill district has a very strong relationship with yield as all the data here are used as input to estimate the yield with the estimated correction factor being used by SRA. The correction factor accounts for environmental extremes (drought and La Nina) and rat infestation as a negative coefficient being multiplied to the total yield of sugarcanes per hectare before actually milling the sugarcane. This computes the yield in the rawest form.

The SRA collects and manages a collection of GIS data for all its Mill Districts (Province) for its reporting and estimation purposes. As shown in figure 4 below, the coverage of sugarcane database and all its growth factors varies from where the mill district is located.

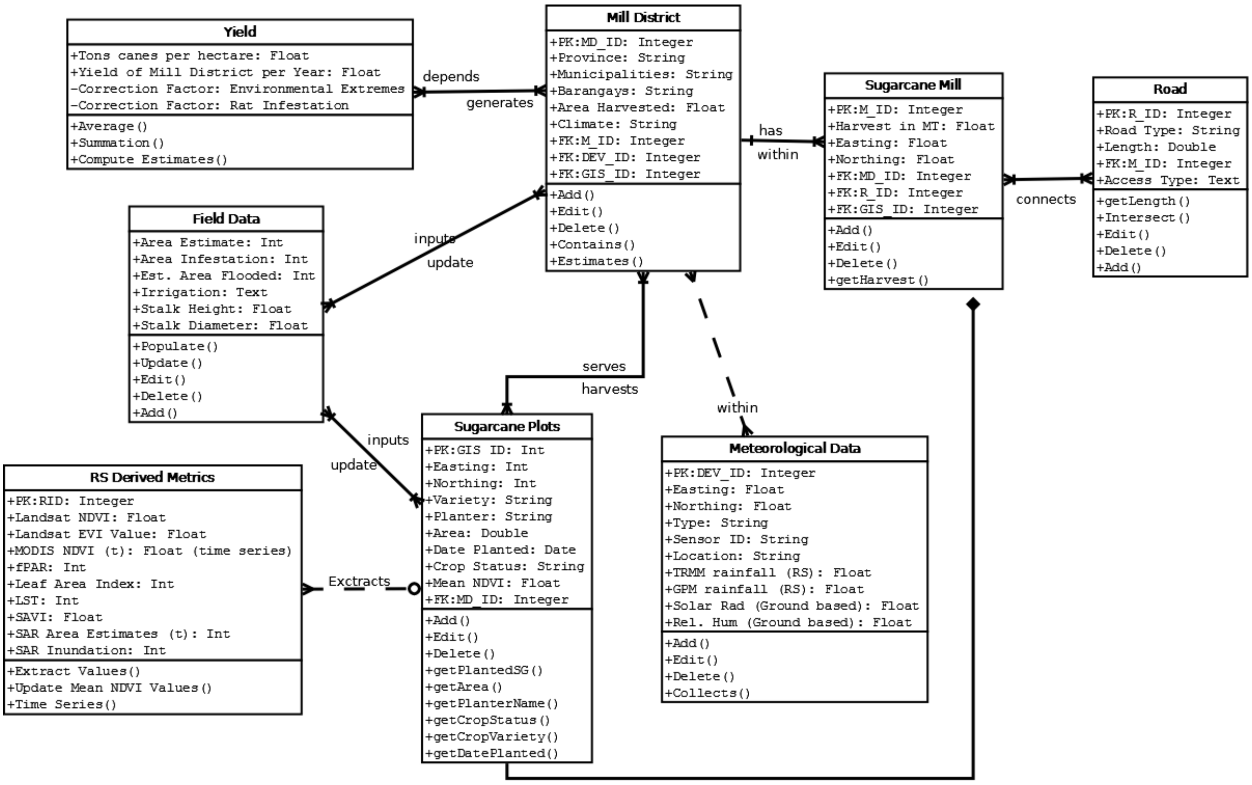
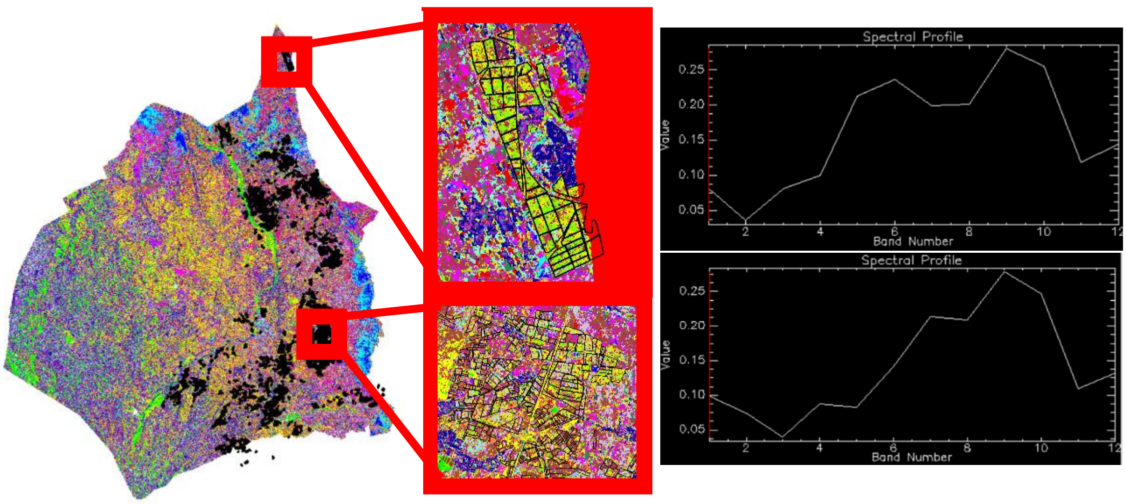


Figure 5. The Physical Model of the Sugarcane Yield Database

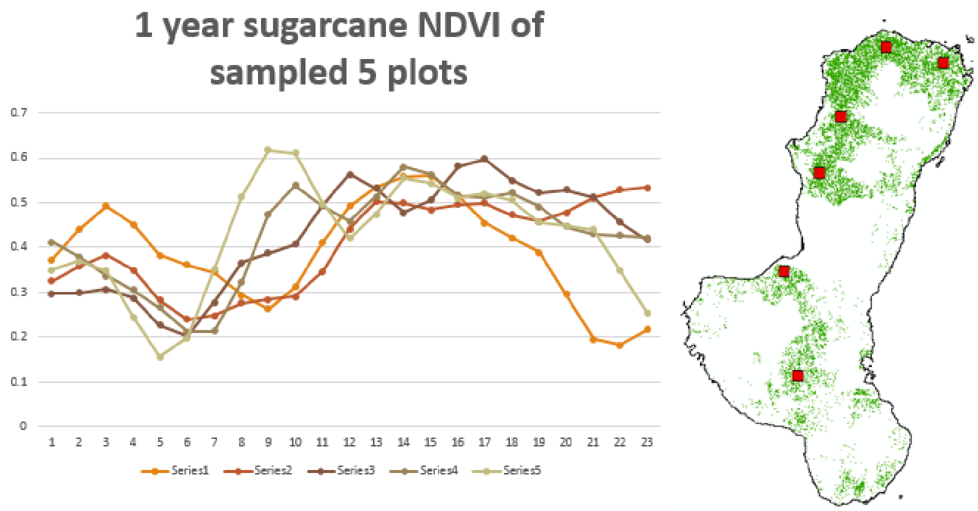
Conceptual model was converted into a physical model by adding the attribute operations and data types for all the class and attributes. Correct data types are vital for database input as this would determine if query operation can be performed in the attribute. The basic functions such as data entry, editing and how this relates to other class are also given in the physical model. The physical model was implemented in PostgreSQL for the spatial database design. All fields and attributes followed the standard data types specified in figure 5. Actual data were used to test the functionalities and identify if problems exists. The relationships in the physical model was modeled using an Entity-Relationship model. This modeled relationships uses the primary and foreign keys of each attributes and class to test if relationship exist and query operation can be performed. Result below show some query operations following the physical model above. The sample query used were from actual use-case scenarios between the researchers and the client (SRA).

4.9 Temporal RS data and database

Results from the physical modeling are promising as query operation are effectively carried out and relationship do exists between modelled attributes and classes. This query operations were demonstrated and showed to be effectively returning the query results as illustrated below.



(A)



(B)

Figure 6. (A) Synthetic Aperture Radar temporal backscatter with classified areas of sugarcane and (B) MODIS Time-series of NDVI for a single crop year

4.10 Spatial and Table Queries

The following query operations are demonstrations of some use case scenarios from the stakeholders. The following queries were executed using PostgreSQL and PostGIS commands:

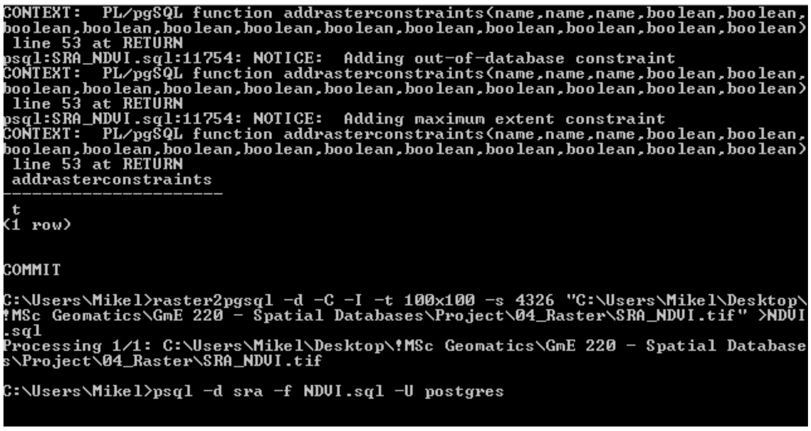


Figure 7. Loading the raster data (NDVI) in Postgresql using the raster2pgsql command

Sample Query Operation: How many ground based meteorological stations are within a mill districts?

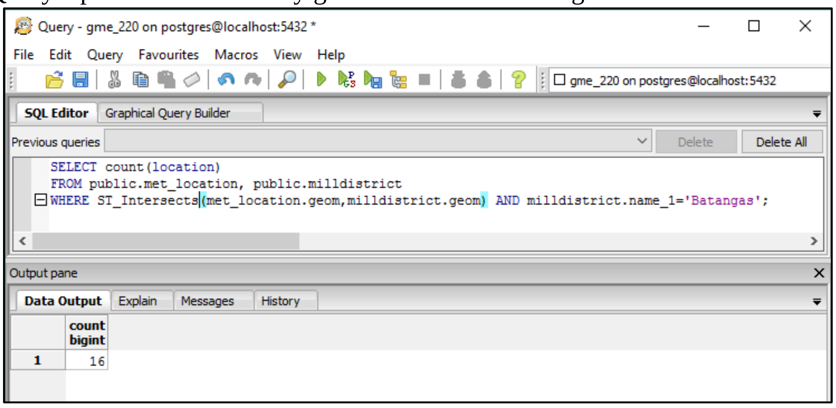


Figure 8. Result of the number of returned meteorological stations

Query Operation: How many sugarcane farms are within 10 km away from the mill in Batangas?

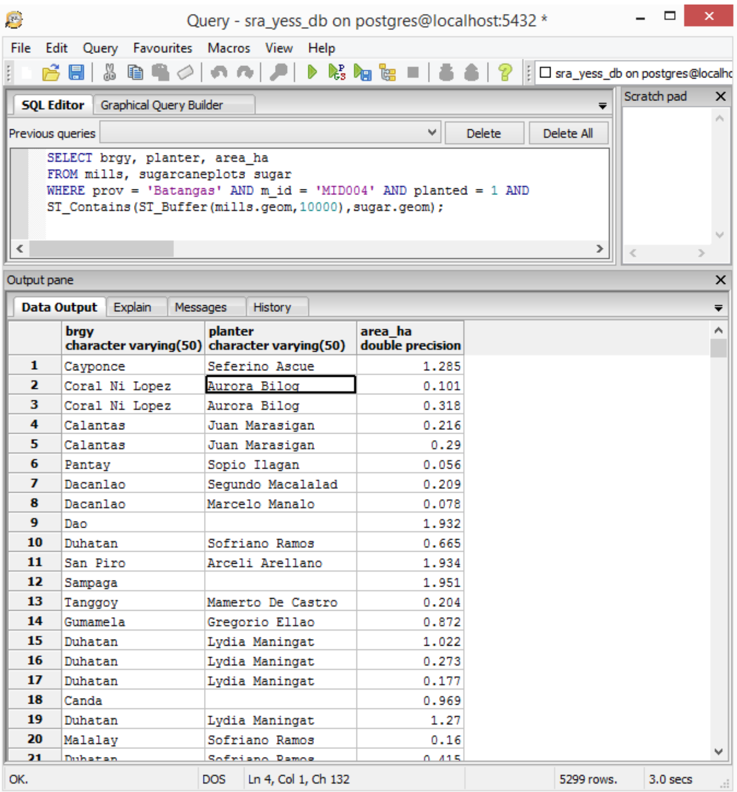


Figure 9. Results of the returned number of farms within 10 kilometer

Sample Query operation: Sum of total sugarcane per year for the last five years

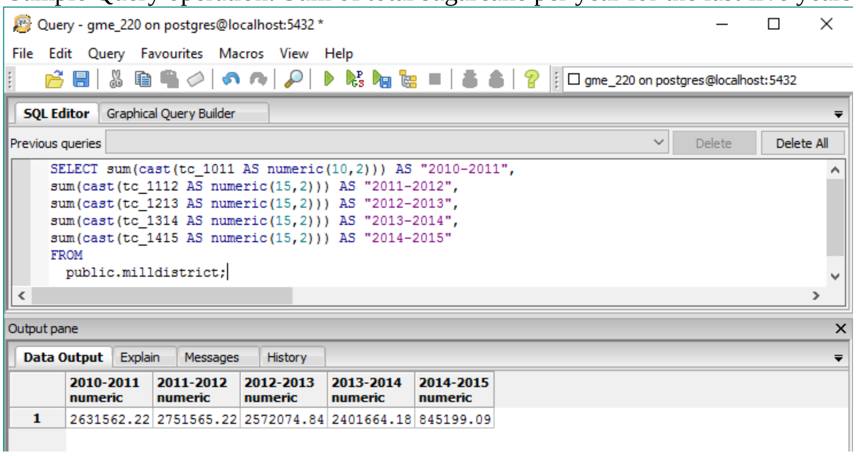


Figure 10. Results of the sum of total sugarcane yield for 5 years

All software used for this research are open source and are licensed to creative commons. Diag ver. 0.97.2 was used for the modeling, Postgresql ver 9.3.9 for the database with PostGIS extension for spatial database and QGIS ver 2.0 for the spatial display and direct connectivity with postgresql database.

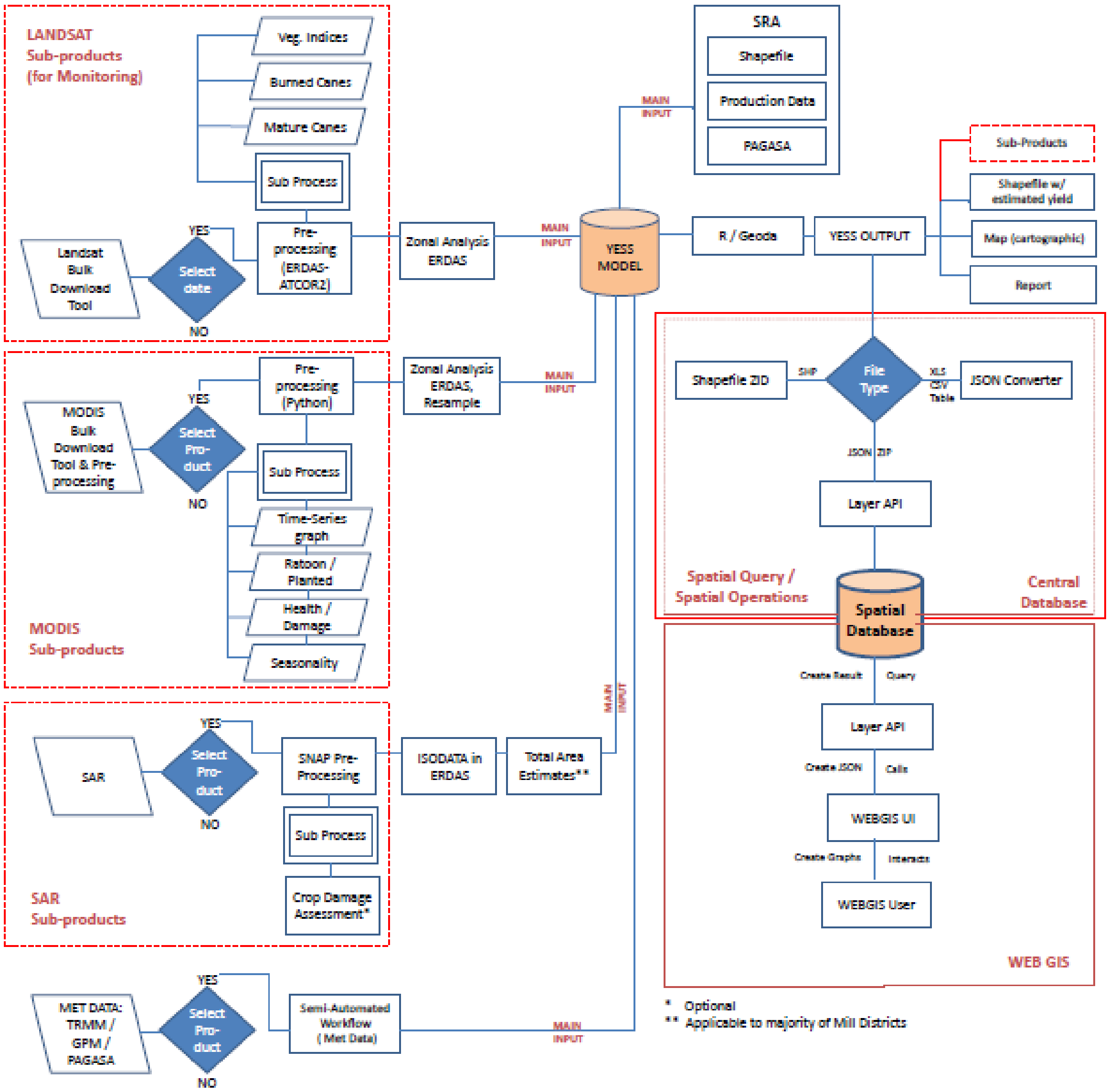


Figure 11. The workflow for the web interface of sugarcane yield estimation highlighting the spatial database

Finally, the spatial database can serve as a link between RS based earth observation data for yield modelling and the IT infrastructure as the physical model can be used to communicate with the web user interface easily. As illustrated in figure 10. The spatial database will serve as an intermediate link between the database and the earth observation data and the user can interact with the system, extract information and use it for decision making.

5 CONCLUSION

Based on the results, conceptual modeling followed by physical modeling showed an effective way of operationalizing the concepts of entities and relationship of real world physical and spatial attributes and that the relationship found therein can provide useful information. The spatial database being one of the most vital part of object modeling proved to have actual use and the importance in agricultural application utilizing open source earth observation data and database tools. Creating a decision support system for an agricultural application is not expensive and that information technology can now be utilized at the farm level at very low to no cost with the advancement of geospatial technology.

REFERENCES

Mulianga B. et. al,. 2014. Journal on Remote Sensing, Forecasting Regional Sugarcane Yield Based on Time Integral and Spatial Aggregation of MODIS NDVI.

Fermont,A and Benson, T. 2011. Estimating yield of food crops grown by smallholder farmers: A review in the Uganda context

Tan, Z., Ding, et al, 2008. The design and implementation of sugar-cane intelligence expert system based on EOS/MODIS data inference model.

www.SRA.gov.ph