Predicting distribution of *Lilium philippinense* (Liliaceae) over Luzon's Cordillera Central Range, Philippines, using ArcGIS Geostatistical Analyst

Ashlyn Kim D. Balangcod

Department of Mathematics and Computer Science,
College of Science, University of the Philippines Baguio,
2600 Baguio City, Philippines
ashlynbalangcod@yahoo.com

ABSTRACT. The growing importance of geographical information system (GIS)-based output in the analysis of biodiversity data is due to its convenient method of spatial analysis of data and prediction of plant geographical distribution. The Interpolated Distance Weight Method (IDW) of ArcMap ArcGIS 9 was used to determine possible areas of *Lilium philippinense*, endemic to the Cordillera Central Range (CCR) and declining in population due to habitat destruction, swidden activities and over-collecting. The variables considered in this study are soil pH, soil phosphorus content, organic matter, elevation, latitude and longitude. All variables were studied from actual *L. philippinense* sites and, using IDW, prediction maps were generated that identified areas where *L. philippinense* are likely to thrive. The Geostatistical Analyst of ArcGIS is a useful tool for predicting potential sites for introduction of *L. philippinense* as an extended *in-situ* conservation strategy.

Keywords. ArcGIS, Cordillera Central Range, distribution prediction, *Lilium*, Luzon, Philippines, potential geographic distribution

Introduction

The study of plant species distribution is an important aspect of biodiversity science. Currently, modeling distribution is less taxing with the use of computer applications or software. If modeled visually through maps, data taken from the field can be interpreted and analysed more accurately. Geographic Information System (GIS) software, specifically, ArcGIS 9, provides storage of quantitative data for generating visual representation on a geographic reference, and retrieval and analysis of information (Fischer 2009). According to Main et al. (2004), GIS help manage, analyse, and present spatially related information combining multiple layers of environmental and biological information related to a spatial location, to gain a better understanding of a specific location (Main et al. 2004). Additionally, researchers can use GIS to fully investigate data and develop spatially accurate graphical data displays. This is very important, especially in geographic distribution where a more accurate and graphical display of populations can be presented. This graphical display can help in decision-making, such as in conservation. This paper focuses on the use of the

Interpolated Distance Weight Method (IDW) in the interpolation of known data of *Lilium philippinense* Baker (Liliaceae) population sites to predict potential areas for cultivation.

In ArcMap, a data layer can be created for each variable of the observed sites. ArcMap is a map-centric application that supports editing and viewing of maps (Longley et al. 2005) and used in visualising different types of data providing an interactive interface for data manipulation, information retrieval and spatial analysis (Zeiders 2002). A data layer is generated from a database of the values of the variables, with the longitude and latitude of the sites. Each data layer can be used by the IDW method of ArcMap GIS to produce the prediction maps. The prediction map is another layer which predicts the values of the predicted sites and compares it with the other areas of unknown data.

The distribution of *L. philippinense* populations was addressed in this study because, first, it is an endemic species in the southwestern part of the Cordillera Central Range (CCR); second, its populations are already declining due to anthropogenic activities; third, few studies have documented this endemic species; and fourth, a dataset consisting of variables such as latitude, longitude, elevation, and soil parameters such as soil pH, phosphorus and organic matter content has been gathered by Balangcod (2009) and allowed to be used in this study. These variables were used to create data layers using ArcMap GIS 9. Specifically, the dataset was used to predict the potential sites of distribution of *L. philippinense* in the CCR using the IDW method of ArcMap GIS 9. With the population of *L. philippinense* dwindling, the use of a dataset in extrapolating and predicting potential sites for the introduction of the species in an extended *in-situ* conservation programme is helpful.

In using the GIS software, a fair knowledge of computers is a necessity as some parts of the software call for critical analysis, an element taught in computer science.

The features and uses of ArcGIS

GIS is defined by Paul Longley et al (2005) as a computerised tool for solving geographic problems, a mechanised inventory of geographically distributed features and facilities, and a tool for performing operations on geographic data that are too tedious or expensive or inaccurate if performed by hand. Childs (2004) describes GIS as all about spatial data and the tools for managing, compiling and analysing that data. GIS has numerous uses such as census, mapping, modelling and prediction.

ArcGIS is a leading software in the GIS market due to its extensive features and global community of users (Information Management Editorial Staff 2004). Among its features is the interpolation tool in the Spatial Analyst extension. Interpolation is a process used to predict the values of cells at locations that have no information. Using the principle of spatial autocorrelation which measures the degree of dependence between near and distant objects, interpolation determines interrelation of values to also determine the spatial pattern (Childs 2004). Another principle which is the basis of spatial interpolation is Tobler's Law, which states that "all places are related but nearby places are more related than distant places" (Miller 2004). This means that the best guess for a point with no information is the value measured at the nearest observation points (Longley et al. 2005).

Another interesting feature of ArcGIS is the use of interpolation methods such as IDW, Kriging and Spline. The IDW method is a deterministic interpolation technique. Deterministic interpolation uses mathematical formulas or measured points to create surfaces (Childs 2004). Kriging is a popular statistical method based on regionalised variables (Longley et al. 2005) and Spline is a method that uses a mathematical function that minimises overall surface curvature (Childs 2004). IDW is the method most often used by spatial analysts due to its simplicity. It estimates unknown data by getting the average of known measurements of nearby points, with the nearest neighbours getting greater weight in the computation of averages. The formula used by IDW is as follows:

$$z(x) = \sum_i w_i z_i / \sum_i w_i$$

where $\boldsymbol{x}$ is the point of interest, the unknown value is denoted by $\boldsymbol{z(x)}$ and the known measurements as $\boldsymbol{z_i}$. The weights ($\boldsymbol{w_i}$) are defined most often by the inverse square of distances formula:

$$w_i = 1/d_i^2$$

where $\boldsymbol{d_i}$ is the distance from $\boldsymbol{x}$ to $\boldsymbol{x_i}$, with $\boldsymbol{x_i}$ as the points where measurements were taken. The data points run from 1 to $\boldsymbol{i}$. (Longley et al. 2005).

Almost all information that requires mapping or modelling over the Earth's surface can now be effectively stored and retrieved using Information Systems. Specifically, Geographic Information Systems or GIS are used for these tasks. In recent years, modelling potential species distributions using GIS has become popular because it is a powerful tool for researchers involved in vegetation mapping, biodiversity mapping and population distributions (Moreno et al. 2007, Murray 2009, Hasmadi 2010). GIS is also a convenient tool in creating prediction maps (Pallaris 1998, Sergio & Draper 2002, Vargas et al. 2004, Vogiatzakis & Griffiths 2006). The last-mentioned study modeled the potential distribution of 36 endemic and 47 non-endemic species of *Anthurium* (Araceae) in Ecuador based on mean annual temperature and humidity. GIS was also used to identify and analyse the environmental tolerance limits of *Cecropia* (Pallaris 1998).

Methodology

Study area

The Cordillera Central Range (CCR) is located in the northern part of the Philippines. It is a mountainous region, with an estimated total area of 17,500 km (CPA Phil. 2006). It has six provinces, viz., Apayao, Abra, Mt. Province, Ifugao, Kalinga and Benguet. The CCR has a diverse flora and fauna, some of which are endemic to the area. *Lilium philippinense*, a species described in 1880, is endemic there (Elwes 1880).

Lilium philippinense is one of three species of *Lilium* L. found in the CCR (Palima 1988). It is a bulb species with a strikingly white trumpet-like flower. The flower has an aromatic fragrance and is produced singly per stem. Each plant bears 1–2, rarely to 4, stems. This species flowers only once per year and is visible during the rainy season from late May to August (Balangcod 2009).

Between 2007 and 2009, a study indicated 118 population sites of *L. philippinense.* These sites were georeferenced using a GPS receiver, Garmin's GPSmap 60C, where latitude, longitude and elevation were recorded. Out of the 118 sites, soil samples from 45 sites were collected for determining the soil pH, phosphorus and organic matter content. These additional variables together with the longitude, latitude and elevation were used with the IDW of ArcMap Gis 9 to predict potential sites where *L. philippinense* would likely grow.

Two sets of data were used. The first set comprises data on latitude, longitude and elevation taken from 118 sites, and the second data set comprises data on latitude, longitude, elevation, soil pH, phosphorous and organic content taken from 45 sites that were a subset of the 118 sites mentioned. All data gathered from the fieldwork was initially saved in excel files (.xls) but were converted to database files (.dbf) since it is the format needed by ArcMap to plot the data on the map.

Using the four characteristics: elevation, soil pH, phosphorous and organic matter, four colour-filled contour prediction maps were produced. These four maps were overlain on the CCR map to determine the areas where *L. philippinense* Baker is predicted.

Preprocessing

A base map for plotting the data was needed for visualisation in ArcMap. A vector and raster dataset of the Philippine map with provincial boundaries was taken from PhilGIS, a website that provides free Philippine spatial data. Separate maps for each province with municipality boundaries were also obtained from the same website. The Philippine maps in shapefile format (.shp) were set in ArcMap using its default datum World Geodetic System (WGS) of 1984 as its coordinate system which is also the datum used in all layers of data in ArcMap.

Use of GIS in predicting sites for Lilium philippinense

Modelling the distribution of different sites across the CCR employed the two data sets mentioned. Prediction maps were generated for each factor: elevation, soil pH, phosphorous and organic matter content of the soil. The Spatial Analyst IDW was used to generate prediction maps with settings of the default power of two. This power value controls the significance of known points on the interpolated values, based on the distance of the known and the output points. Higher values of power may cause "non-smoothness" of values (Longley et al. 2005). The number of nearest neighbours was set to 15, meaning computation of the unknown data depends on the nearest 15 known data.

Using IDW interpolator, the ArcMap calculates the value of each cell of the map depending on the weight or attributes of its neighbouring data. Given that the observed data were randomly distributed in the CCR, the Variable option was used under the search option (Longley et al. 2005).

Verification

To verify the accuracy of the prediction, specifically the prediction maps for the elevation, a Digital Elevation Map (DEM) was downloaded and was used to cross-check the predicted values against the true values. The presence of the DEM allows comparison of the predicted and the actual values.

For the prediction map for elevation using the 45 study sites, 60 sample areas from the predicted sites were chosen randomly for verification. Ten areas were chosen from each province (Fig. 1). Variables such as latitude, longitude, and elevation for these areas were extracted from the DEM and then compared with the predicted elevation.

To verify further, the GPS readings of the elevation of 33 population sites that were included in the 118 study sites but excluded from the subset of 45 study sites were also noted. These 33 sites or points were found within the coloured contours of the prediction map. Note that the 118 sites and the subset of 45 sites have elevation data gathered from the field survey using a GPS handset.

As for the prediction map for elevation using the 118 study sites, the same 60 sampling areas were used and the actual and predicted elevations noted. The percentage errors for the predictions using only a subset of 45 study sites and using the full complement of 118 study sites were computed and compared.

Results

Spatial analyst – IDW method

There were four data layers used in this study. The first type comprised the shape files for the six provinces of the CCR Region; Abra, Apayao, Benguet, Kalinga, Mt. Province and Ifugao which have data until the municipal level. The second type was a Philippine Digital Elevation Map (DEM) which had the actual elevation for the entire country. The third data layer was made up of the corresponding latitude and longitude of the study sites. Finally, the prediction maps generated by the Spatial Analyst of ArcMap constituted the fourth data layer.

The *first data layer* was taken from PhilGIS, a website that provides free shape files for the Philippines. There are six shape files, all with boundaries until the municipal level. The shape files have information including province, municipal or town and barangay. The *second data layer* was taken from a GTOPO30 website. This layer was a raster component which stored the elevation value for each latitude and longitude coordinate of the Philippines. The *third data layer* comprised the two datasets

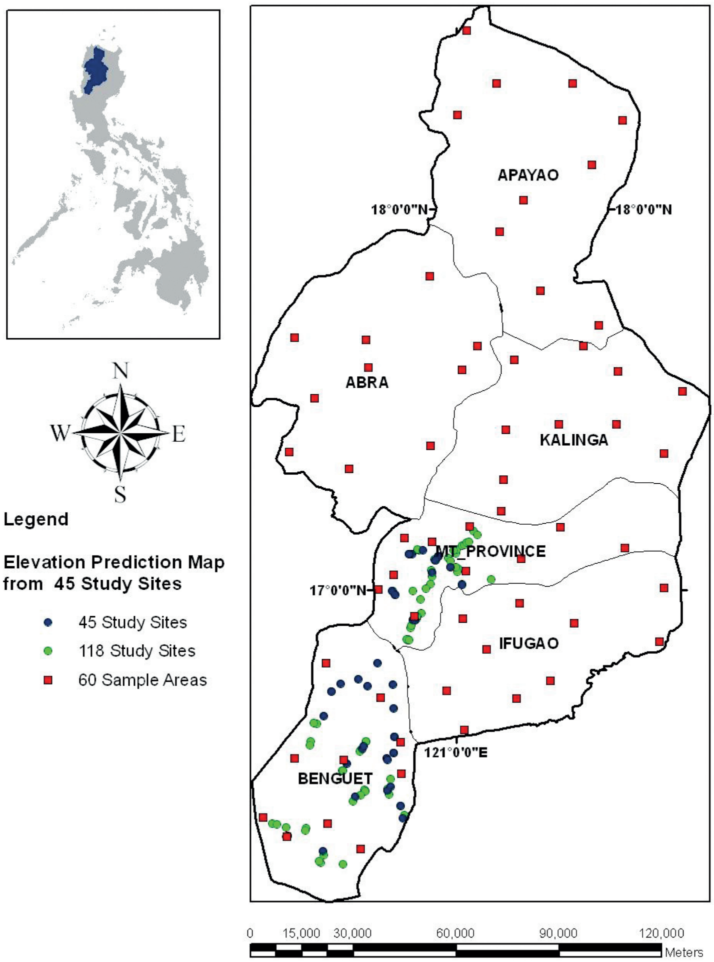

Fig. 1. Sixty sampling areas were randomly chosen to verify the predicted values of elevation generated from the prediction maps that used data from just 45 sites, and all 118 sites. The 45 sites mentioned are a subset of the 118 sites.

taken from the 118 study sites and the subset of 45 study sites. Each dataset had its own separate layer. These layers were superimposed on the others. These study site layers were derived from the table initially saved in Excel format (.xls) that was later changed into a database file (.dbf). The data needed for plotting were latitude, longitude and elevation. The *fourth layer*, consisting of the prediction maps, was generated by the ArcMap IDW Interpolator. There were five prediction maps produced. Two prediction maps using the elevation were created from the subset of 45 sites and from the full complement of 118 sites (Fig. 2 and 3). The other three factors, soil pH, phosphorous content of soil and organic matter were the input used for the other three prediction maps (Fig. 4 to 6).

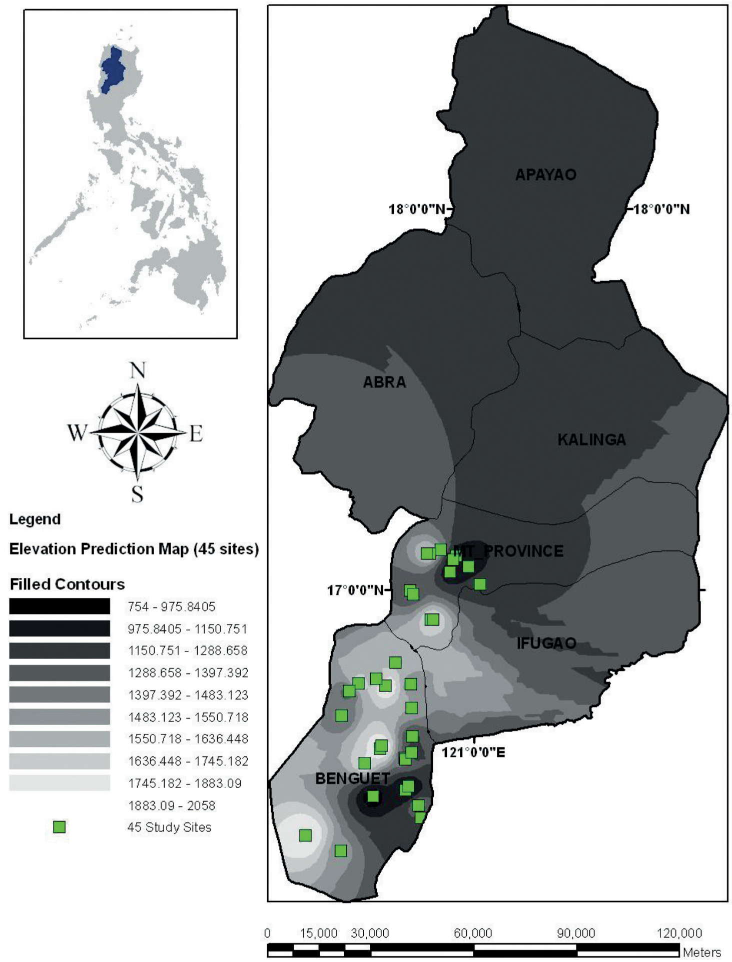

Fig. 2. Prediction Map using elevation of 45 observed sites of *Lilium philippinense*.

However, the ArcMap Spatial Analyst IDW method of prediction uses only values within the range of known data in predicting the values for areas that have unknown data. For example, if the known data are 0, 5, 7 and 10, then unknown data will only have a value within the range of 0 to 10. In this case, the predicted elevation of the unknown areas would only have a predicted value within 754 to 2058 m, when data from only 45 sites were used, and a range of 754 to 2155 m when data from all 118 sites were used. The same applies to prediction of the other factors, namely, soil pH, phosphorous content and organic matter.

To delimit the area of prediction, the range of values where the most number of sites were found was determined. For example, the elevation ranges of 754 to 2058 and 754 to 2155 were divided into ten classes. ArcMap automatically divides the ranges with the specification of the number of classes as input. The top five ranges or 50% of the ranges with the highest number of sites were noted for each data set and

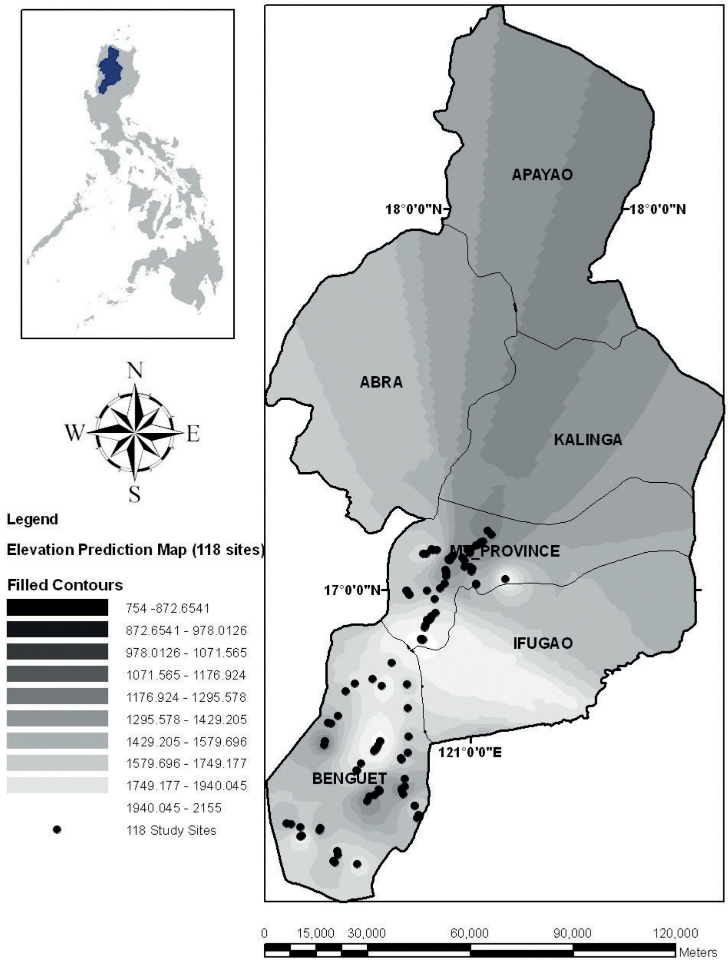

Fig. 3. Prediction Map using elevation of 118 observed sites of *Lilium philippinense.*

were the only data visible in the prediction maps. Table 1 contains the values for the ranges and the corresponding number of sites for each range. Separate prediction maps were created for both sets.

For data on soil pH, phosphorous content and organic matter, their ranges were also divided into ten classes each, and the top five ranges where the most number of sites were found were noted and mapped.

After creating the prediction maps, merging of all maps using all factors (elevation, soil pH, phosphorous content and organic matter) was done and the areas where *Lilium philippinense* was predicted to thrive was determined. This is inferred from the overlain map. The Identify Tool of ArcMap enables a point-and-click query of the location using the shape files of the six provinces.

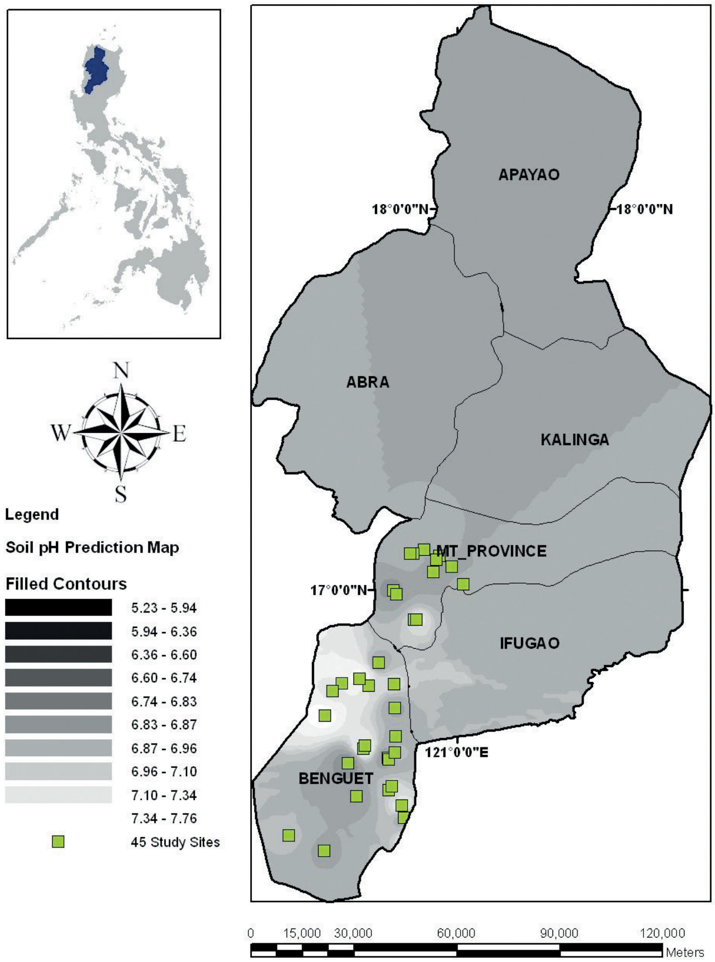

Fig. 4. Prediction Map using soil pH generated from 45 observed sites of *Lilium philippinense*.

There were two output maps from the merging of all prediction maps. The first used the three prediction maps of soil pH, phosphorous content of soil and organic matter with the elevation prediction generated from the subset of 45 study sites. The second map used the same three prediction maps but merging with the elevation map from all 118 sites, instead of from just 45 sites. The areas where the four prediction maps overlap were highlighted in black and isolated.

Percentage Error

Elevation is the factor that can be verified using actual measured GPS values and the DEM. Prediction of elevation generated from 45 study sites, or 45 known points, was computed against the actual values taken from the DEM and was compared with the prediction of elevation generated from 118 study sites. From the 60 sample areas, the elevation from both the predicted elevation generated from 45 study sites, and the

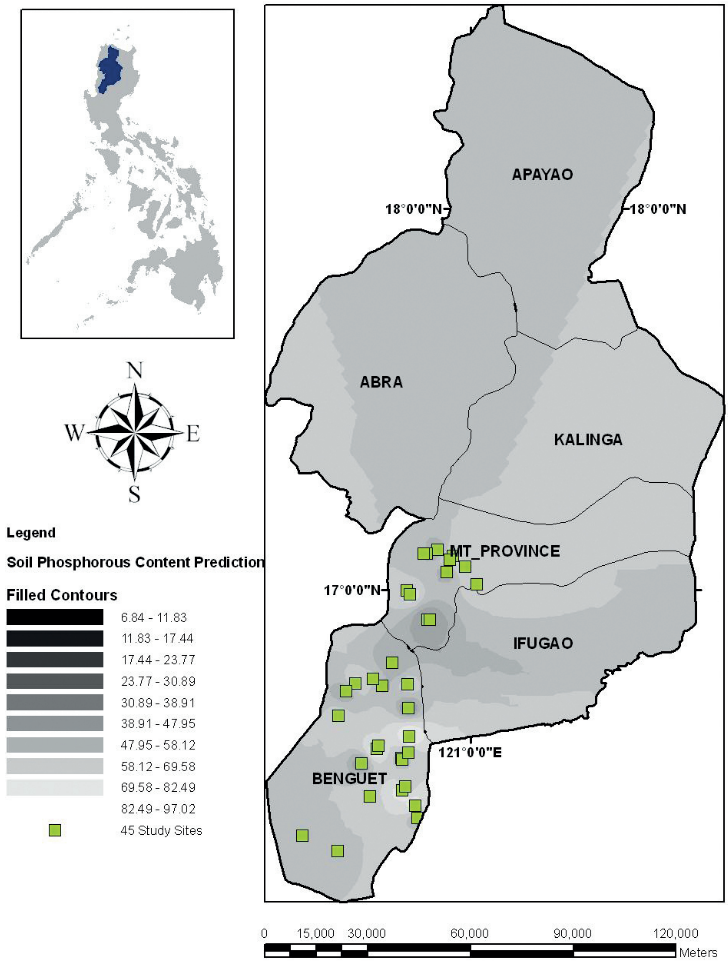

Fig. 5. Prediction Map using phosphorous content of soil generated from 45 observed sites of *Lilium philippinense*.

predicted elevation generated from 118 study sites, were recorded and the percentage error computed. Using the formula below, there is a percentage error of 68.57% for the 45 study sites while the 118 sites had only 47.16% error.

$$\left(\frac{actual - predicted}{actual}\right) * 100$$

The prediction generated from the 45 study sites was further verified by computing the percentage error with GPS readings from the 118 sites. The elevation of 33 random sites not included in the 45 sites were used as the actual value. The percentage error was 7.99%. However, if the DEM of the 33 sites were used as actual values, the percentage error was only 4.31%.

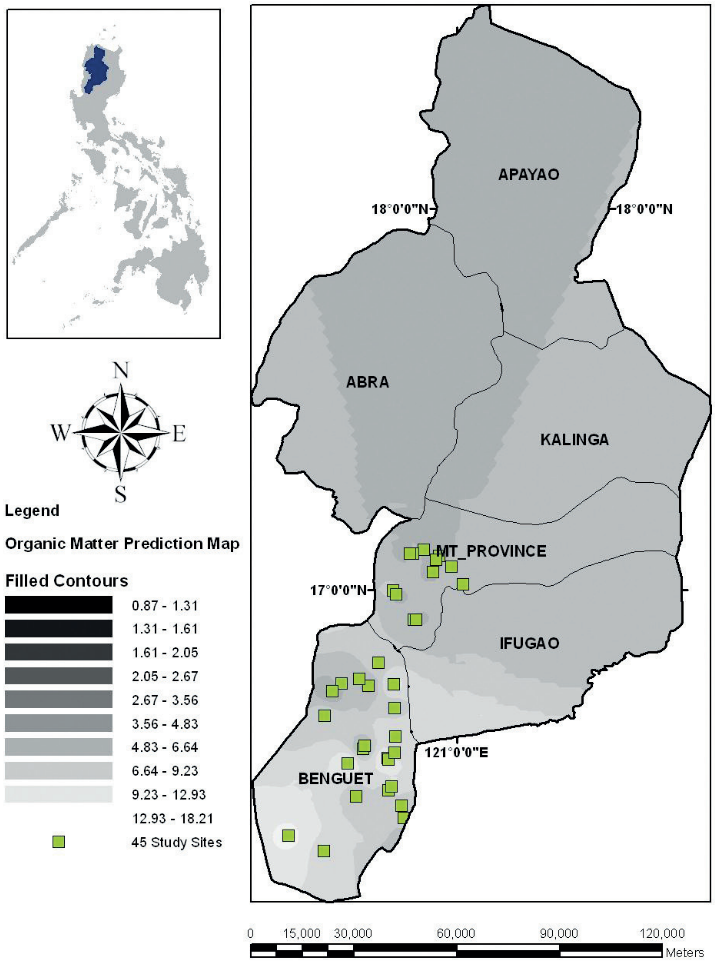

Fig. 6. Prediction Map using organic matter generated from 45 observed sites of *Lilium philippinense*.

From the merged predicted maps of elevation, soil pH, and phosphorous content of soil, a list of predicted municipalities were identified using the Municipal maps taken from PhilGIS website. The areas are given in Table 2.

Discussion

The ArcGIS ArcMap creates a good visualisation of data geographically. It allows modelling of the locations of where *L. philippinense* were observed for users to easily identify the sites and the similarity of sites. This was shown in the different maps produced.

Table 1. Ranges for elevation, soil pH, soil phosphorous content, organic matter and the number of study sites located within each range. * denotes ranges used in the final prediction mapping. 50% of the ranges with the highest number of study sites were taken into account.

Attribute	Range	No. of Study Sites	Range	No. of Study Sites
Elevation (45 sites)	754.0000–975.8405	7*	1483.230–1550.718	4
	975.8405–1150.751	6*	1550.718–1636.448	2
	1150.751–1288.658	5*	1636.448–1745.182	6*
	1288.658–1397.392	5*	1745.182–1883.090	2
	1397.392–1483.123	3	1883.090–2058.000	5
Elevation (118 sites)	754.0000–872.6541	10	1295.578–1429.205	9
	872.6541–978.0126	19*	1429.205–1579.696	14*
	978.0126–1071.565	13*	1579.696–1749.177	10
	1071.565–1176.924	9	1749.177–1940.045	9
	1176.924–1295.578	13*	1940.045–2155.000	12*
Soil pH	5.23–5.94	3	6.83–6.87	3
	5.94–6.36	4*	6.87–6.96	3
	6.36–6.60	9*	6.96–7.10	3
	6.60–6.74	3	7.10–7.34	8*
	6.74–6.83	4*	7.34–7.76	5*
Soil phosphorous content	6.84–11.83	7*	38.91–47.95	5*
	11.83–17.44	4*	47.95–58.12	5*
	17.44–23.77	5*	58.12–69.58	4
	23.77–30.89	4	69.58–82.49	3
	30.89–38.91	4	82.49–97.02	4
Organic matter	0.87–1.31	5*	3.56–4.83	5*
	1.31–1.61	6*	4.83–6.64	6*
	1.61–2.05	4	6.64–9.23	1
	2.05–2.67	4	9.23–12.93	4
	2.67–3.56	8*	12.93–18.21	2

Table 2. Identified municipalities under each province where *Lilium philippinense* was predicted to grow.

Province	Municipalities
Abra	Bangued, Danglas, Langiden, La Paz, Luba, Penarrubia, Pilar, San Isidro, San Quintin, Villaviciosa
Kalinga	Balbalan, Lubuagan, Pinukpuk, Tabuk, Tanudan, Tinglayan
Apayao	Conner
Mt. Province	Bontoc, Sadanga, Sagada, Tadian
Ifugao	Asipulo, Lagawe, Lamut, Tinoc
Benguet	Atok, Bokod, Buguias, Itogon, Kabayan, Kapangan, La Kibungan, Trinidad

The Spatial Analyst IDW was able to predict the values of unknown data given the actual points. However, the limitation of the IDW method, that did not produce values other than the specified range of the known data, has a huge effect on the accuracy of values. The output of the prediction cannot generate values higher or lower than the observed or known values. A possible solution is to acquire more data in sites, not necessarily where *Lilium philippinense* is observed, especially in places on the northern part of the CCR Region. A wider spread of range of values and location would perhaps enable better prediction.

In addition, the IDW method in interpolation to generate prediction maps is based on proximity, and its accuracy in giving a good prediction depends on the number of actual values surrounding that empty space or grid on the map. The nearer and the more the actual values are to an empty space or grid, the better the prediction for that space. This means that if the actual values were found in one specific area, the correctness of the prediction grows less with farther distance (Longley et al. 2005). Due to this limitation of the IDW, prediction is more accurate for areas nearer the observed sites. In this study, the extent of the prediction maps were set to the boundaries of the whole CCR, hence, predicted areas that were farther relative to the actual distribution may not be accurate.

The presence of the majority of known points in the southern part of the CCR Region also affects the prediction. The IDW relies greatly on distance, and therefore the farther the distance of the unknown area from the area with known data, the lower the chances for a good prediction. This is validated with the result of the percentage error of the prediction from the 45 study sites.

Conclusions and recommendations

GIS is a very useful tool in plant distribution. Specifically, the IDW feature of ArcGIS is useful in predicting potential sites for cultivating endangered species like the *L. philippinense.* The accuracy of prediction is dependent on available data and the number of points (populations) that are plotted on the map. The more data and points, the more accurate is the prediction.

The data provided for *L. philippinense* is still limited. Factors such as air temperature and rainfall were not included in the prediction maps since there were no complete measurements for the whole region. There are only three weather stations located in the CCR and these stations are located in Baguio City and Benguet province.

References

Balangcod, T. (2009) *Geographic Distribution, Ecology and Reproductive Biology of Lilium philippinese Baker, an Endemic Species in the Cordillera Central Rang, Luzon, Philippines.* Ph.D. dissertation. Univ. of the Philippines Los Baños.

Childs, C. (2004) Interpolating Surfaces in ArcGIS Spatial Analyst. *ArcUser*, pp. 32–35. www.esri.com

Elwes, J.H. (1880) *A Monograph for the Genus Lilium.* London: Taylor & Francis.

Fischer, F. (2009) Do we still need a desktop GIS? *GeoInformatics: Magazine for surveying, mapping & GIS professionals ESRI* reprint. Vol. 12.

Hasmadi, M., Zaki, M., Ismail, H., Adnan, A.M., Pakhriazad, H.Z. & Fadlli, A.Y.M. (2010) Determining and mapping of vegetation using GIS and phytosociological approach in Mount Tahan. *Malaysian J. Agric. Sci.* 2(2): 80–89.

Information Management (2004) Group 1 Releases Geocoding Solution for ESRI's ArcGIS Software. http://www.information-management.com/news/1002282-1.html. Accessed 13 June 2010.

Longley, P.A., Goodchild, M.F., Maguire, D.J. & Rhin, D.W. (2005) *Geographic Information Systems and Science,* 2nd Ed. London: John Wiley & Sons.

Main, C.L., Robinson, D.K., McElroy, J.S., Mueller, T.C. & Wilkerson, J.B. (2004) A guide to predicting spatial distribution of weed emergence using Geographic Information Systems (GIS) online. *App. Turfgrass Sci.* doi:10.1094/ATS-2004-1025-01-DG.

Miller, H.J. (2004) Tobler's First Law and spatial analysis. *Ann. Assoc. Amer. Geogr.* 94(2): 284–289.

Moreno, E.C., Smith, J.K., Skidmore, A.K., Prins, H.H.T. & Toxopeus, A.G. (2007) Modeling spatial patterns of plant distribution as a consequence of hydrological dynamic processes in a Venezuelan flooding savanna. *Ecotropicos* 20(2): 55–73.

Murray, D.P. (2009) Spatial Distribution of Four Exotic Plants in Relation to Physical Environmental Factors with Analysis using GIS. M.Sc. in Geography Thesis. Virginia Polytechnic Institute & State Univ.

Palima, C.J. (1988) Ornamental vascular flora of Baguio City and vicinity, vol. 1, 2. Unpublished.

Pallaris, K. (1998) *Modelling the Distribution of Cecropia Species Using GIS-based Techniques.* M.Sc. in Geography Thesis. King's College, London.

PhilGIS. (2009) Philippine GIS Data Clearing house. http://www.philgis.org/freegisdata.htm

Sergio, C. & Draper, D. (2002) How to evaluate species when distribution is poorly understood. The use of predictive studies for Iberian bryophytes. *Portugaliae Acta Biol.* 20: 37–48.

Vargas, J.H., Consiglio, T., Jørgensen, P.M. & Croat, T.B. (2004) Modelling distribution patterns in a species-rich plant genus, *Anthurium* (Araceae), in Ecuador. *Diversity Distrib.* 10: 211–216.

Vogiatzakis, I.N. & Griffiths, G.H. (2006) A GIS-based empirical model for vegetation prediction in Lefka Ori, Crete. *Plant Ecol.* 184(2): 311–323

Zeiders, M. (2002) Introduction to ArcMap. GIS Resource Document. http://www.pop.psu.edu/gia-core/pdfs/gis_rd_02-13.pdf