International Centre for
Integrated Mountain Development

796

ISSN: 1027-0027

Issues in Mountain Development 99/6

Integration of GIS and Remote Sensing for Crop Acreage Estimation: An Information System Development Approach

This paper presents an approach to the development of a Geographic Information System (GIS) for crop acreage estimation to support the crop forecasting system at regional level. The research adopted the Structured System Development Methodology (SSDM) to develop a GIS that includes complex processes and data models. The overall aim of the system design was to support crop area estimation through Area Frame Sampling (AFS), Remote Sensing (RS), and a combination of both. However, a detailed design was carried out to support AFS only. Based on a system that can support AFS survey design, field work, data processing, and assessment of the quality of data entered, two prototype systems were designed: CAEIS-I and CAEIS-II. CAEIS-I was developed using the Arc Macro Language (AML) of UNIX based Arc/Info, and CAEI-II was developed using MS Access '97 in a PC environment.

Introduction

Crop forecasting involves the prediction of crop yields (tons/ha) and production before the harvest actually takes place; typically a couple of months in advance. Reliable estimates of areas covered by certain crops and their production are essential for proper planning and monitoring as well as improvement of the agricultural development process and the decision-making process.

Information on changes in land cover/use is of particular importance when such changes could result in progressive land degradation. Derivation of land cover/use and estimation of agricultural production through classical methods are costly, time consuming, and subject to a variety of errors in terms of types and sources. Recent developments in GIS and Remote Sensing (RS) technologies and crop modelling have created promising opportunities for improving agricultural statistics' systems. Techniques in crop inventory systems that are based on the application of GIS, RS, and other agro-ecological models provide examples of developments in each discipline. Several processes, such as derivation of improved land cover/use, crop production estimation, and data handling, are involved in each discipline and are the main components of estimation and forecast of agricultural production. When all these components have to be combined into a single system, especially for a large area, the implementation of the system becomes very complex. This is, in fact, due to the different kinds of information needed for each individual component. Such complexity can be solved with an automated Information System (IS) that has the capability of integrating different processes and all the necessary data in the system (Paresi 1995). The IS can be automated by developing a conceptual process model and data model and by linking these two. In this study, an attempt has been made to develop a conceptual process and data model and to develop a GIS for crop area estimation (hereby referred to as CAEIS) to support the Crop Forecasting System (CFS) at regional level.

The main purpose of developing an information system for estimating crop area is to support crop forecasting system.

Crop Acreage Estimation: An Overview

The main purpose of developing an information system for estimating crop area is to support CFS. Therefore, it is important to understand how relevant the information produced from CAEIS will be to CFS as per the requirements of the organization. A crop production forecast is derived by multiplying the area under cultivation for a particular crop (usually stated in hectares or acres) by the yield of the crop per unit area (usually in quintals/hectares or acres) using the formula: **Area x Yield = Production**.

ISSUES IN MOUNTAIN DEVELOPMENT is a series released from time to time to brief planners, development workers, researchers, and donors on recent trends, findings, and issues affecting mountain development. The papers in this series can be quoted with due acknowledgement. They can also be accessed on ICIMOD's World Wide Web pages on the Internet at:

http://www.icimod.org.sg

Comments are Welcome

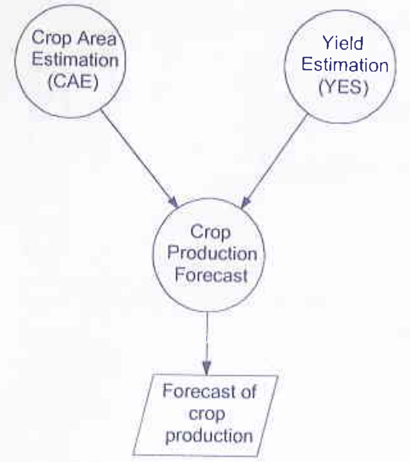

Figure 1: Key Processes of a CFS

The determination of Area and Yield is very complex, because they are based on many factors and variables. So, from the users' perspective, and based upon the definition, CFS, as an information system for crop production, is a combination of three key processes (Figure 1).

Since the main aim of this study is to focus on the design and development of a Crop Area Estimation (CAE) sub-system, the paper focusses only on this sub-system.

Crop Area Estimation (CAE) Sub-system

The objective of the CAE sub-system is to distinguish, identify, measure, and map the area covered by significant crops in the study area (Sharifi 1995). This entails four independent components: a ground survey of randomly sampled areas based on the so-called Area Frame Sampling (AFS) method; a multi-temporal classification of satellite images covering the entire area; a combination of these two; and aggregation/disaggregation to different administrative levels.

Area Estimates through Area Frame Sampling (AFS)

The concept of AFS is based on the development of area estimates through direct measurement of fields within sampled areas. The AFS is generally derived by dividing the total area to be surveyed into small blocks, without any overlap or omission; select a random sample of blocks; obtain the data needed on the population in the sample blocks; and estimate population totals. If the limit of the region is known, the exact elements of the population (the frame) are known. These elements can be mainly of two types: points or pieces of land known as segments. Random selection of segments is carried out for one block only, and a similar pattern of segments is replicated in all other blocks. The main idea of designing an area frame is to locate sample areas for ground survey in a statistically viable way in order to minimise sampling errors. The AFS method includes stratification of the whole region. This process uses datasets such as current land-use maps, topographic maps, Digital Elevation Models (DEMs), and satellite images to stratify the region into homogeneous areas in term of use, type, and pattern of agricultural fields. The area of each crop in the whole area is estimated by statistical methods.

Area Estimates through Remote Sensing (RS)

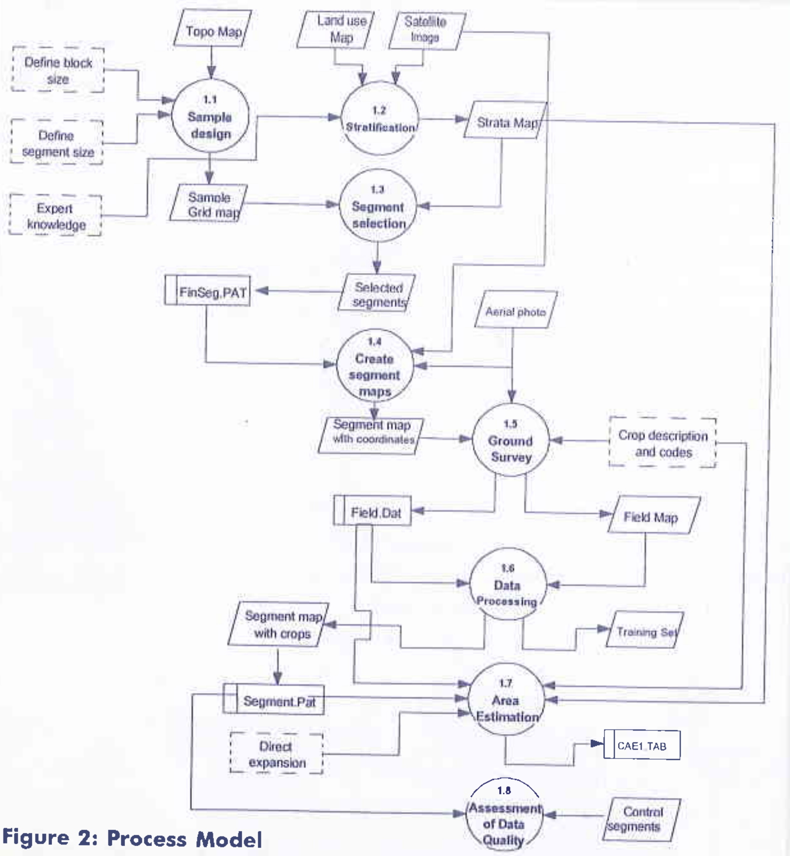

Figure 2: Process Model

A basic step for crop area estimation is crop classification. Satellite remote sensing data have been available since the early 1970s and their characteristics have facilitated their use in improving crop area estimation. Historically, area estimates have been derived through statistical procedures with samples obtained in the field as input data. This is both expensive and time-consuming. Remote-sensing techniques have become useful and attractive for this process.

Aggregation/Disaggregation

The area estimates derived through the above method are at the stratum level and physically do not exist. To transform this into the required administrative units, the result is overlaid on a map showing the lowest administrative unit and later aggregated to the required level.

Design of Proposed System

A system has both static and dynamic components. The static component describes the structure of the system, and the dynamic component describes the process in the system. More specifically, the static properties of a system are concerned with the abstractions of objects and concepts involved, their properties (attributes), and their relationships. The dynamic aspects of the system are concerned with operations on objects and properties required by transactions and queries. Thus, the dynamic model abstracts the behaviour of the system and takes the form of specifications. The proposed system was designed by applying 'process modelling' which involves identifying the processes of the system and defining the relationship between the system and the processes. Data modelling includes a description of how the thematic and geometric data can be linked in a database, how the data about features of the terrain can be abstracted and represented in a database, and, based on this, the database for implementation in which storage structures are specifically designed (Refer to Pradhan1995 for details regarding process and data models.)

Implementation

The implementation phase entails the sample design (process 1.1), stratification (1.2), selection of segments (process 1.3), area estimation (process 1.7), and assessment of the quality of data (process 1.8), (Figure 2).,and refer to Pradhan (1995.)

Sample Design

First of all, a grid of the study area was designed by identifying the minimum and maximum X, Y coordinates to determine the limits of the study area. Then the coordinates of each point, in both X and Y directions, at an interval of 10km (the size of the block) was calculated using the UTM coordinate system. Coverage of the square grid was generated from the coordinates of these points using Arc/Info. The size of the segments was determined to be 1km by 1km, and, therefore, each block was again subdivided into small 1km by 1km grid cells; i.e., each square block contained100 segments. The segments (points) associated with each square block were assigned a unique code from 1 to 100 in each square block as shown in Figure 3.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

10 Km (vertical)

10 Km (horizontal)

Figure 3: Segments (1km x 1km) in a square block (10km x 10km)

Stratification

Geometrically-corrected satellite imagery (SPOT-XS) was used to identify and delineate sub-areas of different agricultural systems. An existing land-use map of the study area was very useful for stratifying the area against other areas under different types of agriculture. Four main strata were defined based on the intensity of agriculture, as given in Table 1.

Table 1: Area of Stratum

Lu_code	Strata	Intensity	Area (km²)
20	Irrigated agriculture area	High intensity	490.40
30	Rainfed agriculture area	Medium intensity	1066.56
60	Mixed agriculture area	Low intensity	525.60
40	Pasture (rangeland)	Very low intensity	465.25

Selection of Segments

After the grid has been defined, the most practical method for locating or selecting sample segments is to locate the maximum number of segments (depending upon the chosen maximum sample rate) that could be in any strata. Later some of them are discarded due to other specific reasons, e.g., lower sampling rate for a strata, the segment which is falling on the border of the region with less than 50% of its area inside the region, etc. These segments are referred to as possible segments. So, the whole segment selection procedure consists of selecting possible segments and then selecting final segments.

The 100 segments produced from the above process in each square block are the total number of the population of the entire area, some of which were selected for a sample survey. To identify the possible segments of each square block, it was necessary to give a unique code number for each segment in each square block. When discarding or deleting the possible segments, the same segments should be deleted in each square block. The next step is to select the possible segments at random with a similar pattern (replicates) in each block. In this selection, two or more segments can be geographically close to each other which, in general, yields redundant information as their behaviour tends to be closer. So, a systematic, aligned sampling method with a distance threshold was applied for random selection of the segments. It generated an aligned sample with a certain distance of replicates, ensuring distribution of the segments avoiding the forbidden areas. In this study, a threshold distance of 3km was defined to ensure that the distance between the centres of two segments was not less than 3km.

The 100 segments produced from the above process in each square block are the total number of the population of the entire area, some of which were selected for a sample survey. To identify the possible segments of each square block, it was necessary to give a unique code number for each segment in each square block. When discarding or deleting the possible segments, the same segments should be deleted in each square block. The next step is to

select the possible segments at random with a similar pattern (replicates) in each block. In this selection, two or more segments can be geographically close to each other which, in general, yields redundant information as their behaviour tends to be closer. So, a systematic, aligned sampling method with a distance threshold was applied for random selection of the segments. It generated an aligned sample with a certain distance of replicates, ensuring distribution of the segments avoiding the forbidden areas. In this study, a threshold distance of 3km was defined to ensure that the distance between the centres of two segments was not less than 3km.

When the samples were selected in a block and their code number identified, replication was carried out by selecting samples with the same code number in each block, and the rest of the segments were deleted. The remaining segments with code numbers 12, 17, 45, 72, and 77 in each block are the possible segments. (Table 2)

Table 2: Selected Possible Segments

Segment-ID	Block -ID	Segment Code	X	Y
172	1	12	298185.81250	396227.50000
177	1	17	303185.81250	396227.50000
385	1	45	301185.81250	395927.50000
592	1	72	298185.81250	395627.50000
597	1	77	303185.81250	395627.50000

The possible segments were overlaid on the stratification map, and it was observed that few segments were existing across the region and strata boundaries. When a segment straddles the stratum boundary and is shared by two strata, there are at least three alternative ways of coping with this kind of situation (Gallego 1995).

- Assigning the segment to the stratum with which it most overlaps
- Splitting border segments into pieces that belong to different strata
- Keeping only the largest piece of the segment in one stratum and discarding all other pieces

According to Gallego, the first alternative is the easiest to manage, but may introduce some bias. The second alternative seems to keep the largest amount of information and might prove to be the best in future, but tests have not given convincing results. The third alternative gives the best practical results, although it is not supported by solid theoretical background , and this still needs to be developed. So, considering the importance of the largest amount of information, the first alternative was selected in this study.

After dealing with the straddle boundaries and overlapping of segments in two strata, final selection of segments was carried out by defining a sampling rate for each stratum. The sample rates can only be equal or less than the maximum sample rate. In this study, the maximum sample rate was 5/100 = 5% (100 segments in a block and 5 of them chosen as possible segments). The sample rates for each stratum were chosen depending upon the intensity of crop production in the area. In this study, the irrigated area was an intensive agricultural area, and, therefore, the highest sample rate ,4%, was chosen; i.e. a maximum of 4 segments per block for the irrigated area. So, out of the possible segments, 4 segments were randomly selected in a block for the irrigated area only and the code numbers (GLu1 - new item) of randomly selected segments were recorded. In this case, the segments with the code number GLu1 = 15, 17, 20, and 23 were randomly selected for irrigated areas. The rule applied was - if the stratum is irrigated area and the code number is GLu1 = 15, 17, 20, and 23, keep the segments, otherwise delete them.

Similarly, the sample rates 3%, 2%, and 1% were chosen for rainfed, mixed, and pasture areas respectively, and the segments were randomly selected for each strata and replicated in each block, applying the same rule. The remaining segments were the final selection for the sample survey. The real sample rate for each stratum was calculated by the system shown in Table 3.

Table 3: Calculated Sample Rates for Each Strata

Stratum	Lu_code	No. of Sample Segments	Area of Strata (km^2.)	Real Sample Rate
Irrigated agriculture area	20	20	490.40	4.08
Rainfed agriculture area	30	32	1066.56	3.56
Mixed agriculture area	60	11	525.60	2.09
Pasture area	40	2	465.25	0.86

The overall sample rate for the whole area = 2.64%

Creation of Sample Map

Aerial photographs on a scale of 1:40000 were used to create sample maps. The scale of the photographs was too large to use directly for drawing tracts. Hence, the photographs had to be enlarged before use. The general method applied for creating representation of sample maps is as follows.

- Finding the centre coordinate of a segment on the 1:50,000 scale map
- Comparing the features on the map and photos to find the matching patterns

- Scanning the whole photo at a resolution of 600 dpi to easily find control points
- Selecting the control points (a minimum of 4) on the map, calculating the coordinates with the help of rulers, and using these coordinates to geo-correct the photograph
- Cutting out a geo-corrected photo of required size depending upon the size of segments (extending the area of the segment by about 50 to 100 metres is preferable)
- Printing the cut-out photo on an A4 size paper, covering as much area as possible. Printing the scanned photo at a very high resolution to be used during field work.

Ground Survey

Field work was carried out in each segment selected to produce an accurate map of crops and other land cover. In some cases, crop yields were estimated. To find the exact location of a segment was the most difficult part. The relevant landmarks shown on the map and the photos were used to locate a segment. In many cases, a GPS was used to find and verify the exact location (co-ordinates) of segments. The field boundaries depicting each crop type were drawn on tracing paper. The crop codes for each polygon were indicated on the tracing paper. Features such as roads, paths, rivers, hedges, and streams were also mapped on it.

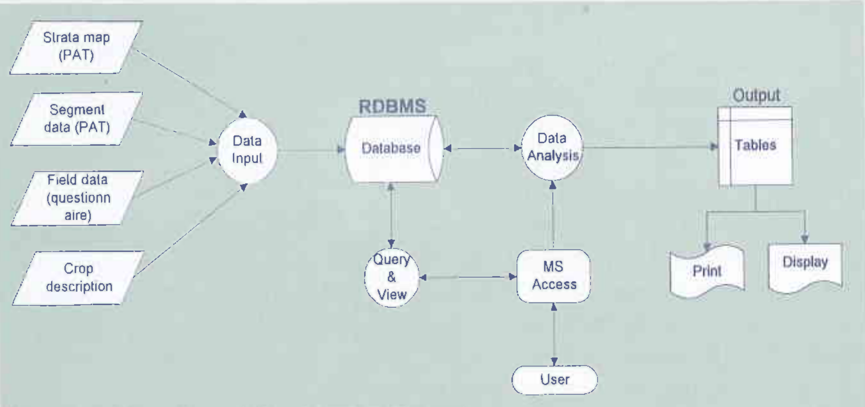

Figure 4: Overview of the CAEIS-II System

Data Processing

Data processing and preparation of the data required two types of data - segment specifics that involved tracts and tract numbers and crop specifics involving area codes of all crops. Data preparation and processing mainly included the digitisation of mapped segments during field work and entry of recorded data into questionnaires. Arc/Info was used to digitise the mapped segments and the crop attribute data were entered in dBase III+. Later these data were managed into Arc/Info DBMS.

CROP_CODE	DESCRIPTION	IRRIGATED	Area {ha}	CV (Zc)%
102	Wheat	y	15739	30.17
103	Wheat	n	41310	16.64
104	Barley	y	7846	42.13
105	Barley	n	3831	43.69
112	Beans	y	929	67.46
114	Chickpea	y	1047	48.16
115	Chickpea	n	364	95.09
116	Lentil	y	95	69.52
117	Lentil	n	158	95.09
144	Sugarbeet	u	9	95.09

Table 4: Estimated Crop Area of the Study Area by Coefficient of Variances, 1997

A substantial part of the study was involved in building a prototype system for area estimation during the implementation phase. A prototype system, CAES-II, was developed for this process.

The CAEIS-II was developed using Microsoft Access '97 database software in a PC environment (process 1.7). The advantage of building a system in a PC environment is that it is not as complicated as it would be if a UNIX environment were to be used; the secondary advantage is that most users who are familiar with a database concept in a PC environment can easily handle the system. An overview of CAEIS-II is given in Figure 4. The detailed statistical formula used in this system can be found in Pradhan(1995).

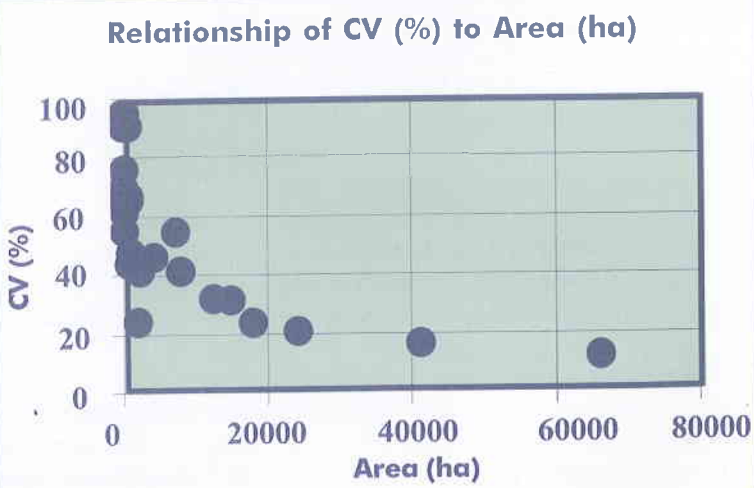

Figure 5: Relationship of CV (%) to Area of Crops (ha)

Results and Discussions

The results given in Table 4 (sample only) are derived from the prototype system CAEIS-II. The coefficient of variance (CV) is a measure of the accuracy of unbiased area estimates made from ground surveys and is calculated by: $CV(\%) = \frac{\sqrt{Var(Z_c)}}{Z_c} x100$

Results obtained from the system were pooled to produce Figure 5 which shows the relationship between the CV and the total estimated area of crops in the region. The CV decreases as the crop area increases because the standard error, $SE = \sqrt{Var(Z_c)}$, decreases. The main crops typically covered areas of between 0 to 15,000ha at this level. Hence, it can be concluded that the system works well for large areas.

Conclusions

CAEIS is an information system that assists in the design and implementation of an area frame, selection of the sample areas within the frame, and estimation of the areas of major crops. The main advantages of the system include the following.

- The concept of GIS is employed to support spatial information management, analysis of spatial data combined with related thematic data to support the decision-making process, and finally to allow the presentation of results of various processes in a manageable and easily understandable form. Spatial presentation is a natural way of approaching any spatial problem. Visualisation of results using GIS technology is one of the most comprehensive and effective forms of presentation and communication.
- Using the geographic information concept has given us the possibility of combining various forms of information from many different sources and relating them through a common spatial location.
- The integration of different data from various sources into a single system has improved the function, operation, and performance of the system.
- The CAEIS is actually a combination of three different systems - AFS, RS, and CAE. When we look at the whole CAEIS, we can see that the system uses different types of data from various sources, involves many processes, and, thus, becomes a very complex system. The CAEIS is based on GIS which also contain RS. All spatial data, including digitised thematic maps, remote sensing images, and tabular data have to be integrated into a single information system to improve the functionality, applicability, and performance of the system as a whole. Hence the necessity for integration of GIS and RS, a very complex task because of the differences in their data acquisition processes and the structure of these data.

References

Gallego, J. 1995. Sampling Frames of Square Segments, Institute for Remote Sensing Application, MARS. Italy: Joint Research Centre.

Paresi, C.M. 1995. Introduction to Information Development and to System Development Methodologies (Ed). The Netherlands: ITC.

Pradhan, S. 1998. Development of a Geo-Information System for Crop Area Estimation Using Area Frame Sampling and Remote Sensing Technique, M.Sc. Thesis, The Netherlands: Wageningen Agricultural University and ITC.

Sharifi, M.A. 1995. Development of Crop Inventory and Forecasting System for the Major Agricultural Commodities in Hamadan Province in Islamic Republic of Iran. The Netherlands: ITC.

Sushil Pradhan
GIS Analyst
Mountain Environment and Natural Resources' Information Systems,
ICIMOD
e-mail: sushil@icimod.org.np

Established in 1983, ICIMOD is dedicated to the cause of poverty alleviation and environmental conservation in the Hindu Kush-Himalayan range of Afghanistan, Bangladesh, Bhutan, China, India, Myanmar, Nepal, and Pakistan. ICIMOD is a focal point for documentation and information exchange, training, applied research, and demonstration on a wide range of issues affecting mountain people.

Editors: **Shahid Akhtar/Archana S. Karki**
ICIMOD, P. O. Box: 3226, Kathmandu, Nepal
Tel. (977 1) 525313 or Fax (977 1) 524509, e-mail: distri@icimod.org.np