# BUILDING THE 4D GIS FOR THE TREND OF HOUSING BUILDING BY MINING THE CONSTRUCTION DATA IN DISTRICT 5TH HOCHIMINH CITY, VIETNAM

**Khuu Minh Canh**[1] **and Le Trung Chon**[2]

[1]HCMC University of Technology, VNU-HCM
268 Ly Thuong Kiet Quan 10, TP.Ho Chi Minh, VietNam
Email: kmcanh@yahoo.com
[2] HCMC University of Technology, VNU-HCM
268 Ly Thuong Kiet Quan 10, TP.Ho Chi Minh, VietNam
Email: ltchon@hcmut.edu.vn

## ABSTRACT

*In this paper, the model of house building in the future will be presented by the 4D GIS, 3D in space and temporal. By mining the building stage 2001-2010 and lidar data in the District 5th of HoChiMinh city, the expecting trend of building is built by combinning some analyses such as point process techniques, Markov, timeseries as well as random network. By this method, building house points will be generated temporally by the statistical random generating method and the trend in changing will controlled by the factors from timeseries analysing. And the geometric model for house is employed just the level of detail 1 (LOD1), block of house*

## 1. INTRODUCTION

Management of urban development related to setting overall architecture of areas problem. At localities, in urban centers, the construction of houses associated with other developments such as the cultural, economic, service, social, business forms, especially in constructions of the people. According to the collected of the civil building data from 2001 to 2010 in District 5, Ho Chi Minh City, there are about 7,000 new housing project started. That means the building density being very large. On average for each day, there are 2 residential buildings which is started. So, there are about 35% of the total built home in the county for 10 years. The data collecting shows that the constructions in the district 5 are almost continuously for 5 years. From 2001 to 2010 there is only 7 over-ten-day-duration times without any new project and all in the Lunar New Year of years 2002 to 2007 and 2010 year.

In this study, the process of building will be studied by using statistical methods. It consider the building as a random process in space. Houses built are a set of random points in available on a map. Accordingly, the statistical model will calculate the parameters of the construction process for urban data in District 5th. And then, the parameters will be applied to calculate the random points of new construction in the year of later. Accordingly, the statistical procedures will be studied:

- To statistic the trend for urban construction in District 5 in the construction in dry season (December to April), in Vietnamese called 'mùa khô', and rain season (May to November), in Vietnamese called 'mùa mưa'.
- Statistic about construction trends over space, as trend to build belong points.
- Build the method to simulate the process of building for future in District 5 respectively in 2001-2010 which results in LOD1.

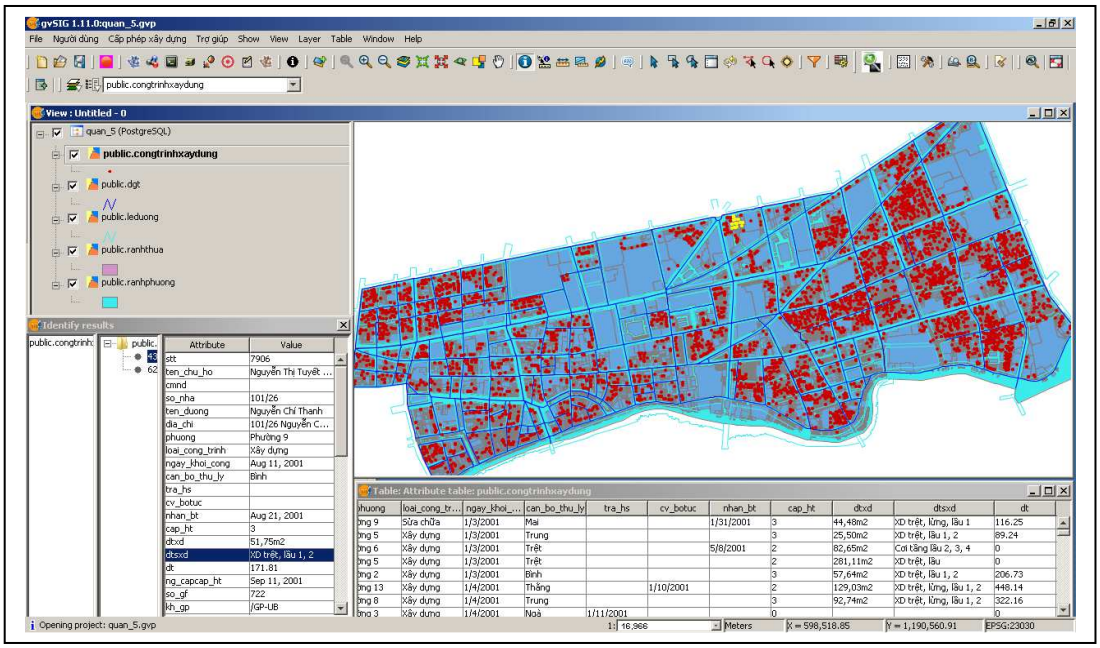


**Figure 1. Map of construction points from 2001 to 2010 in District 5th of Ho Chi Minh city in VietNam.**

## 2. PROCESS AND DATA PREPARATION

### 2.1 Process

Setting up trends in residential building of the people is done by exploiting real data and performing statistical analysis and forecasting of future construction. The development of the construction is the result of many developing interaction each other within and outside local such as economic, cultural and social factors Thus, setting up the urban development, especially in the construction need to examine many factors impact.

The process proposing based on the analysis the properties that house are constructed and belong to the main season of building. The procedures are proposed as follows:

- Step 1: to gather the real data about the construction.
- Step 2: to statistic laws of space-time and trending to build houses.
- Step 3. To Develop areas according to statistics from the data set on the base map layer.
- Step 4: To Perform that random constructions in the future from the above parameters. Random function receives the parameters: enforcement regional functions (from step 3), random time (input parameters), the number of construction projects are generated randomly (from step 2 were analyzed by time include the following information: seasonal rules, number of floors)
- Step 5: To Extracts of the current height of the elevation data from topographic surface model (DSM) and the corresponding number of existing floors.
- Step 6: To a random point setting in step 4 will develop high-level map 1 (LOD1).
- Step 7: To storage and display data sets over time (suggested data structure: NetCDF format).

### 2.2 Data collection

The data of collecting includes data on the floor in District 5, data of constructions. Data of hight surface on elevation DSM (extracted from Lidar data). In particular, performing

of the above procedure,the data of constructions include the following information:

- Location (address, coordinates to locate on map) of the building
- Start date, number of curent floor/forey, the changing in number of the floor (new construction).

THỐNG KÊ VỀ LĨNH VỰC XÂY DỰNG MÙA

Mùa mưa: tháng 5 đến tháng 11; mùa khô: tháng 12 đến tháng 4

| TT | Thời điểm | Mùa | Cả quận (Yt) Số Công trình | MA (trung bình trượt) | baseline CMA (tâm trung bình trượt) | Yt/CMA St x It | St | Yt/St Deseasonalize | Tt | Dự báo |
|---|---|---|---|---|---|---|---|---|---|---|
| TT | | | 7060 | | | | | Khử mùa | | |
| 1 | 2001 - 2002 | Mùa mưa | 329 | | | | 1.21698864 | 270.3394 | 305.1499 | 371.364 |
| 2 | | Mùa khô | 279 | 304 | 315.5 | 0.884311 | 0.79906443 | 349.1583 | 315.0144 | 251.7168 |
| 3 | 2002 - 2003 | Mùa mưa | 375 | 327 | 323 | 1.160991 | 1.21698864 | 308.1376 | 324.8788 | 395.3739 |
| 4 | | Mùa khô | 263 | 319 | 333.25 | 0.789197 | 0.79906443 | 329.1349 | 334.7433 | 267.4815 |
| 5 | 2003 - 2004 | Mùa mưa | 432 | 347.5 | 347.25 | 1.24406 | 1.21698864 | 354.9746 | 344.6078 | 419.3838 |
| 6 | | Mùa khô | 262 | 347 | 356.5 | 0.734923 | 0.79906443 | 327.8834 | 354.4723 | 283.2462 |
| 7 | 2004 - 2005 | Mùa mưa | 470 | 366 | 386.5 | 1.216041 | 1.21698864 | 386.1992 | 364.3367 | 443.3937 |
| 8 | | Mùa khô | 344 | 407 | 403.75 | 0.852012 | 0.79906443 | 430.5035 | 374.2012 | 299.0109 |
| 9 | 2005 - 2006 | Mùa mưa | 457 | 400.5 | 389 | 1.174807 | 1.21698864 | 375.5171 | 384.0657 | 467.4036 |
| 10 | | Mùa khô | 298 | 377.5 | 371.5 | 0.802153 | 0.79906443 | 372.9361 | 393.9302 | 314.7756 |
| 11 | 2006 - 2007 | Mùa mưa | 433 | 365.5 | 362.5 | 1.194483 | 1.21698864 | 355.7963 | 403.7946 | 491.4135 |
| 12 | | Mùa khô | 286 | 359.5 | 394 | 0.725888 | 0.79906443 | 357.9186 | 413.6591 | 330.5403 |
| 13 | 2007 - 2008 | Mùa mưa | 571 | 428.5 | 456.25 | 1.251507 | 1.21698864 | 469.1909 | 423.5236 | 515.4234 |
| 14 | | Mùa khô | 397 | 484 | 483.75 | 0.820672 | 0.79906443 | 496.831 | 433.3881 | 346.305 |
| 15 | 2008 - 2009 | Mùa mưa | 570 | 483.5 | 480.75 | 1.185647 | 1.21698864 | 468.3692 | 443.2525 | 539.4333 |
| 16 | | Mùa khô | 386 | 478 | 492.75 | 0.783359 | 0.79906443 | 483.0649 | 453.117 | 362.0697 |
| 17 | 2009 - 2010 | Mùa mưa | 629 | 507.5 | 480.75 | 1.308372 | 1.21698864 | 516.8495 | 462.9815 | 563.4432 |
| 18 | | Mùa khô | 279 | 454 | | | 0.79906443 | 349.1583 | 472.846 | 377.8344 |

**Figure 2. Timeseries analysis for dry ('khô') and rain (mưa) seasons.**

Besides, in order to analyze 3D, Data of hight surface on elevation (the DSM) are used. From lidar data sources, the hight of the existing building is determined.

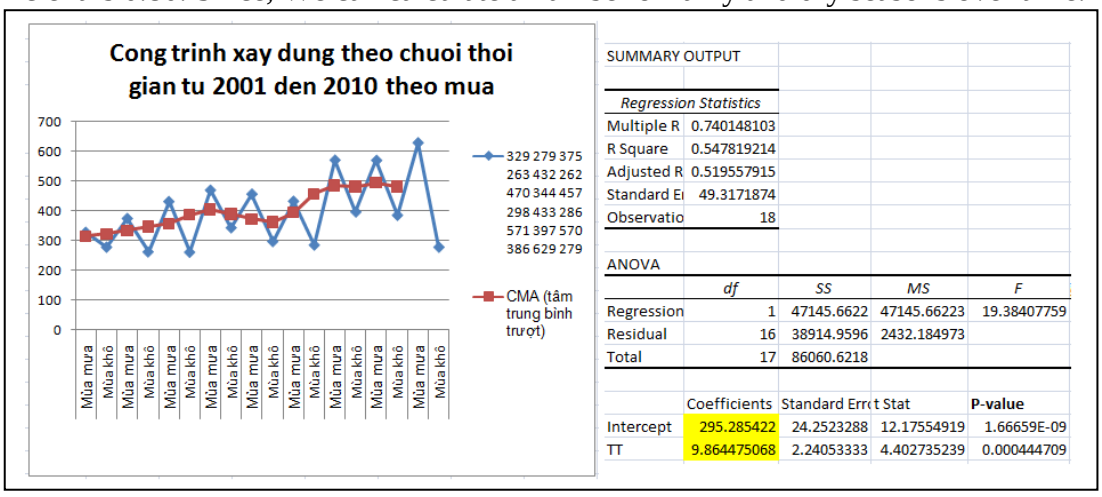
## 3. METHODS

To implement above processes, the statistical techniques and data mining is used to analyze to rule over time, the spatial distribution of builds over time.

### 3.1 Statistical analysis methods

#### 3.1.1 *Timeseries analysis*

Statistical data to analyze time-series method is done by calculating average regression. This data are analyzed into two seasons, the rainy season (May to November each year) and dry season (December to April). The results of the analysis showed that the rainy season constructions regression coefficient is 1.22 and in the dry season the building with a coefficient is 0.80. Since, We can calculate a number of rainy and dry seasons over time.




SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.740148103 |
| R Square | 0.547819214 |
| Adjusted R | 0.519557915 |
| Standard Ei | 49.3171874 |
| Observatio | 18 |

ANOVA

| | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Regression | 1 | 47145.6622 | 47145.66223 | 19.38407759 |
| Residual | 16 | 38914.9596 | 2432.184973 | |
| Total | 17 | 86060.6218 | | |

| | Coefficients | Standard Erro | t Stat | P-value |
|---|---|---|---|---|
| Intercept | 295.285422 | 24.2523288 | 12.17554919 | 1.66659E-09 |
| TT | 9.864475068 | 2.24053333 | 4.402735239 | 0.000444709 |

**Figure 3. Graph and regression statistical output.**

### 3.1.2 *Analysis the space of construction points*

During construction spatial analysis and statistical data to change the floor of the building is an essential problem. To Statistics from the floor changes and the current of high, anticipating of the new work is more realistic because the new buildings are built on the works of the lower floors. According to result of statistics, the more changing 2 to 3 floors are the most of majority of the work. the number of floors in the building change in foollow:

- Non-floor-increasing cases: 1206 buildings.
- 1-floor-increasing cases: 21 buildings.
- 2-floor-increasing cases: 1607 buildings.
- 3-floor-increasing cases: 2608 buildings.
- 4-floor-increasing cases: 850 buildings.
- 5-floor-increasing cases: 409 buildings.
- Over-6-floor-increasing cases: (approx) 200 buildings.

In addition, analysing time-series shows the number of new buildings that constructed will vary by season, typical of construction in the Vietnam. However, to create a random distribution of the new buildings, the spatial distribution must be considered. By method of establishing statistical search criteria number of new construction in the next-09-month-following period (called the near buldings), and external (call far buldings) that they has distance within 500 meters. The number of near buildings and far buldings is statistics computed.

**Table 1. Results of analysis the new construction site in the next 9 months and in 500 mets radius.**

| Type of analysis | Notes |
|---|---|
| Less near buildings | Almost in the $10^{th}$ to $15^{th}$ ward of District 5 (the cluster of western wards of district) |
| Most near buildings | Almost in the $1^{st}$ -$4^{th}$ ward (the cluster of eastern wards of district), which sometimes more than 200 works “near” the building. |
| Less far buildings | Wards $2^{nd}$,$3^{rd}$, $12^{th}$, $13^{th}$, $14^{th}$. |
| Most near buildings | Wards 2,3, 12-15 and the most at the ward 9. Within 9 months there are 628 building that started and far from center of district |

```
import os
import math
import arcpy

def query_new_close_house(sf_name, met_radius, year_radius):
        myVector = [[0 for j in range(6)] for i in range(8000)]
        rows = arcpy.SearchCursor(sf_name,"","","Shape;khoicong_n;khoicong_t;khoicong_1;phuong","")
        i = 0
        for row in rows:
                print i
                myVector[i][0] = row.Shape.centroid.X
                myVector[i][1] = row.Shape.centroid.Y
                myVector[i][2] = row.khoicong_1 * 1.0 + row.khoicong_t * 1.0/12
                myVector[i][3] = 0
                myVector[i][4] = 0
                myVector[i][5] = row.phuong
                i = i +1
        soluong = i-1
        for k in range(soluong-1):
                r = soluong
                for l in range(k+1, r):
                        dx2 = (myVector[k][0] - myVector[l][0])**2
                        dy2 = (myVector[k][1] - myVector[l][1])**2
                        khoangcachkg = math.sqrt(dx2 + dy2)
                        dt = (myVector[l][2] - myVector[k][2])
                        if (khoangcachkg < met_radius) and (dt < year_radius) and (dt >= 0):
                                myVector[k][3] = myVector[k][3] + 1
                        if (khoangcachkg > met_radius) and (dt < year_radius) and (dt >= 0):
                                myVector[k][4] = myVector[k][4] + 1
        file_out = open("d:\\congtrinhgan_xa.txt", "w")
        for k in range(soluong):
                file_out.writelines(myVector[k][5] + ";" +str(myVector[k][0]) + ";" + str(myVector[k][1]) +
                    ";" + str(myVector[k][2]) + ";" + str(myVector[k][3]) + ";" + str(myVector[k][4]) +";\n")
        file_out.close()
```

**Figure 4. Python code to extract the "near-far" in time and space from one construction point in shapefile.**

### 3.1.3 *Generate random point tool*

When the number of points (construction) are determind in a random moment which can arise centralized location. The expression of random points can use the software tools. Example: Random Point tool in ArcGIS will support user that generated random points from the available point. The collection available points is selected which based on criteria such analysis

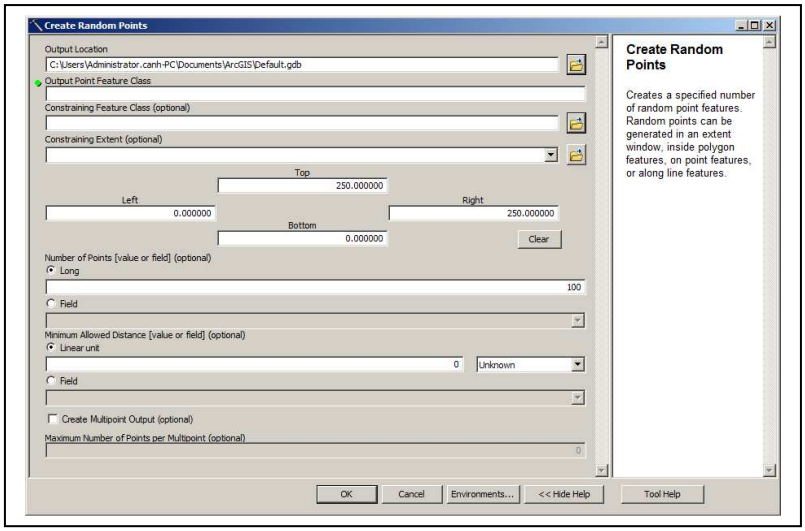



**Figure 5. Random point tool in ArcGIS 10.**

## 3.2 Process for creating the 4D-scene

The creation model of building 4DGIS simulate the development housing construction of people which based on the basis of random points generated over time and generate network cluster as well as trends of hight as above statistical analysis. displaying and storagring about lidar data that taken at the same time will provide information about the digital surface model (DSM) of the District. Then, the elevation of each the house is extracted by interpolation through the number of current storey. The interpolation consists of two processes:

● The first is the process of interpolation for pass time: the current elevation of parcel is archived from lidar data with the number of storey. So the elevation for LOD1 (level of detail 1st) of the parcel is calculated.

● And the second process is interpolate at future time also based on the current lidar data. The elevation of the point in stochastic set is randomly increased for the predicted site of new building in future. Then LOD1 is built.

## 4. RESULT

The process of model building 4D GIS for construction is proposed. The parameters for the time, space in the model include parameters on the amount of the amount incurred, arising cluster location, hight generated by statistical calculations that based on real data on buildings in the previous year. The model will demonstrate 4D spatial data base on 3D surface elevation model (DSM) to extract data from Lidar. 3D elevation data would play a role in the new surface interpolation. On storage of physically, model of the data format much length NetCDF.

In fact, the construction can be affected by many other factors. Therefore, the analysis of data collected about the location and date of starting old buildings is only reflect one of local trends. Accordingly, the following of studies have to supply the factors of economic, social and cultural affecting of the local neighborhood and the whole city in statistical data.

## 5. REFERENCES


Nguyễn Duy, Tiến, 2008. *Các mô hình xác suất và ứng dụng – Phần 1 – Xích Markov và ứng dụng*. Nhà xuất bản đại học quốc gia Hà Nội.

Thạch Thanh Tiền, *luận văn Thạc sỹ Toán "Xích Markov và ứng dụng"*, Trường Đại học Khoa học Tự nhiên, (2013).

Adrian Baddeley, *Spatial Point Processes and their applications*, School of Mathematics & Statistics, University of Western Australia.

HCMGIS, *Dự án bay chụp Lidar khu vực thành phố Hồ Chí Minh*, 2012, HCMGIS, hcmgisportal.vn.

http://www.esri.com

http://vi.wikipedia.org/wiki/Nam_Bộ_Việt_Nam