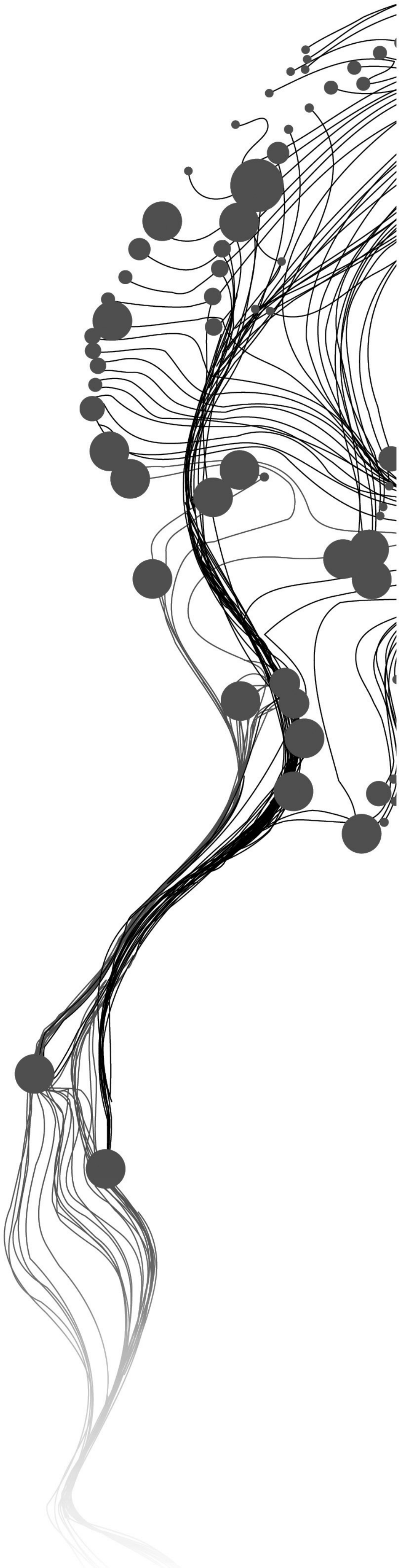
DESIGN AND REALIZATION OF RASTER-BASED SPATIAL UNCERTAINTY CALCULUS AROUND VECTOR OBJECTS

DEEPAK RAJ AWASTHI
March, 2012

SUPERVISORS:

Dr. Ir. R. A. (Rolf) de By
Dr. Ivana Ivanova



DESIGN AND REALIZATION OF RASTER-BASED SPATIAL UNCERTAINTY CALCULUS AROUND VECTOR OBJECTS

DEEPAK RAJ AWASTHI
Enschede, The Netherlands, March, 2012

Thesis submitted to the Faculty of Geo-information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:

Dr. Ir. R. A. (Rolf) de By
Dr. Ivana Ivanova

THESIS ASSESSMENT BOARD:

Dr. C. P. J. M. van Elzakker (chair)
Dr. A. Dilo

ABSTRACT

Design and realization of uncertainty is always an important subject matter in geographical information science. Spatial uncertainty can be associated with spatial objects in various ways. Textual description is one of the causes of uncertainty. It is often used in geographical information science during communication or in spatial queries by using named places and spatial relationships. These descriptions provide uncertain or vague information about a locality. Consequently, there always exit spatial uncertainty like positional or extensional uncertainty. Currently, spatial databases and geographic information systems are unable to deal with such spatial uncertainty due to lack of data types and uncertainty functions. There is a need of a mechanism that can represent spatial objects along with their uncertainty and allows a computational facility. In order to solve such problem, a tandem approach is proposed in the project where uncertainty is represented by tiny rasters. The tandem data model uses the combination of two basic data types vector and raster, where vector is deemed as a crisp representation of the object and preserve its identity as point, line string or polygon feature. Raster provides supplementary information about the vector object and describes the possible spatial position of the object using probability theory. This project mainly focuses on the calculation of uncertainty value in every pixel of the raster by using a technique of calculus and subsequently creation of uncertainty raster for tandem data type. PostgreSQL/PostGIS functions are developed to create uncertainty raster and allow computational facility between tandems to approximate the spatial position of the object.

Keywords

spatial uncertainty representation, tandem approach, spatial database (PostgreSQL/PostGIS) raster support, textual description

ACKNOWLEDGEMENTS

This thesis would not be a reality without the support of various people and organizations.

First and foremost, I would like to express my sincere gratitude to my first supervisor Dr. Ir. R. A. (Rolf) de By for his scholarly advice, continuous suggestion, comments, and encouragement.

A special thanks to my second supervisor Dr. Ivana Ivanova for her comments, advice and recommendation that help me a lot to overcome my academic challenges.

I am indebted and thankful to Dr. Arta Dilo for her support, guidance and providing valuable information regarding spatial uncertainty.

I would like to acknowledge NUFFIC for providing fellowship to pursue M.Sc. degree in Netherland, Government of Nepal for granting permission to commence this academic course as well as providing suitable environment and ITC/UT for providing platform to develop profession in the field of Geo-Information Science.

I appreciate Jorgearevalo for his support in the installation of PostGIS raster on window environment system via online discussion on his blog. Similarly, I would like to thank Mauricio C. M. de Paulo for his online support about QGIS plugin to make it work efficiently in the visualization of PostGIS raster. A special thanks to Hari "Guru" for his continuous support, guidance and discussion on every issue throughout this M.Sc. period.

Thanks to all my classmates for sharing a long period together and helping me directly or indirectly in various aspects of this academic program. I owe sincere and earnest thankfulness to all Nepalese friends for providing a homely environment throughout this period.

Finally, I would like to show my gratitude from the bottom of my heart to my family members for their love, support and blessing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 MOTIVATION AND PROBLEM STATEMENT

Textual description (e.g."25 km NW of Belem") is often used to identify a geographical feature using named place and spatial relationship. Such textual descriptions are associated with a number of uncertainties and these are often inevitable (Liu et al., 2009). Spatial analysis without consideration of the uncertainty factor has limited use (Fisher, 1999). Probable uncertainty in spatial domain and to keep on high trust in the accuracy of data may lead to great loss for GIS users (Shi, 1998). Rowe (2005) found misleading results while studying the impact of locality uncertainty on species distribution patterns. These textual descriptions provide vague or uncertain information about a geographical feature especially its position and extent. Generally, the tools and techniques like gazetteer (Goodchild and Hill, 2008) and Geographical Information Retrieval (GIR) (Jones and Purves, 2008) are used to deal with textual information. However, the representation problem of uncertain objects in spatial databases (Dilo et al., 2007), causes difficulty to GIR while dealing with the spatial queries indicating vague geographic terminology. Spatial uncertainty always comes together with indeterminate boundaries.

There are always two types of indeterminate boundary: sharp boundaries of which the position and shape are unknown or which cannot be measured precisely (called positional uncertainty and measurement uncertainty respectively), and fuzzy boundaries such as between grassland and forest which are not well-defined and are called fuzziness (Schneider, 1999). Dilo et al. (2007) define vagueness as a limitation in border line cases to answer exactly what it is. According to Molenaar (2000), crisp objects can be determined through their boundaries whereas vague objects cannot, because of their unfixed spatial extent. He illustrated extensional uncertainty associated with vague objects to define its potential spatial extent. Several theories have been proposed to deal with uncertainty or vagueness.

Probabilistic model is used to handle positional uncertainty associated with spatial assertions during geo-referencing a locality (Guo et al., 2008). Fuzzy set theory is used to define vague objects and operators to represent and analyze spatial vagueness (Dilo et al., 2007). Fuzzy theory introduced by Zadeh (1965) is the most often used theory to handle vague objects. According to Altman (1994) fuzzy theory is suitable for modelling vagueness in the information which is inherently imprecise such as qualitative information where uncertainty cannot be attributed to randomness. Random set theory is also used frequently to deal with spatial uncertainty in recent times. Zhao et al. (2009) used random sets to model a spatial object with uncertain boundaries and define random data types in the context of two-dimensional Euclidean space. To represent and analyze geographic phenomena they should be modelled and conceptualized.

According to Goodchild (1992), the field and object models are two approaches to conceptualizing and modelling geographic phenomena . Basically, the field-based model is more appropriate in finding uncertainty (Goodchild, 1989). Raster representation is the most common implementation of the field model and can be used to determine spatial uncertainty when uncertainty cannot be modelled in a deterministic way (Guo et al., 2008). Though object and field models are used individually to conceptualize geographic phenomena, there is increasing demand for the integration of these models to represent complex geographic phenomena as such various approaches developed to unify them.

Cova and Goodchild (2002) linked field and object representation of geographical space to show the potential value of combined object-field model in spatial analysis and decision support such as locating facilities like fire stations, transmission lines, and biodiversity reserves. Furthermore, they mentioned that object-field may have application in assessing spatial data uncertainty. Similarly, Kjenstad (2006) presented the common base-model using UML (unified modelling language) for classical field and object model called Parameterized Geographic Object Model (PGOMODEL) to conceptualize the representation of geographic features. Generally, a glacier easily modelled by the object-based model when we only its independent attributes on location such as *size* and *glacier type* are considered. Similarly, glacier with dependent attributes on location such as *velocity* and *thickness* can be modelled by field-based model. Kjenstad used PGOMODEL to model conceptually a glacier including its both dependent and independent attributes. In the same way, Goodchild et al. (2007) introduced a single fundamental construct geo-atom as a foundation for both object and field conceptualizations. Mainly, Goodchild offered a conceptual and theoretical framework to clarify and integrate thinking on geographic representation in GIS. Finally, Voudouris (2010) suggested the object-field model with uncertainty and semantics which particularly used to represent indeterminate geographic phenomena exhibiting combination of field and object like properties. A single foundation termed as 'Elementary-geoParticle' was introduced with associated uncertainty and semantics. Voudouris represented an elementary geoparticle as a raster cells and used it in town center cases where it contains composite values as index of town centredness and object reference i.e., to which town the particle belongs. Thus according to Voudouris (2010), uncertainty can be used to quantify the likelihood of the elementary geo-particle being part of a town center.

In addition to these unified models of object and field, in (Glemeser and Fritsch, 1998), a concept of hybrid model was introduced. The hybrid model can process both raster and vector data in an integrated manner taking uncertainty into account. Based on the concept of aforementioned models to represent indeterminate geographic phenomena, it seems that there is a possibility to represent uncertain objects along with the design and realization of an uncertainty using a spatial database management systems like PostgreSQL/PostGIS.

A tandem approach, as a combination of raster and vector (i.e., double representation of a single uncertain object) seems one of the possibilities. The tandem technique will unite both vector and raster to represent uncertain objects based on the existing structure of geometry and PostGIS raster data type of PostgreSQL/PostGIS. PostGIS Raster is a new data structure and function set developed inside PostGIS as an extension for PostgreSQL. It has the ability to perform spatial analysis and processing of raster data as well as combination of raster and vector data seamlessly by offering a single interface of overlay SQL functions (Obe and Hsu, 2011).

Hence, to achieve the requirement of a mechanism which can represent spatial objects with their uncertainty and provide a facility to compute them, the tandem approach will be developed in the spatial database PostgreSQL/PostGIS environment and is indeed a new endeavor. The realization of raster-based uncertainty with the help of tandem approach will definitely extend the scope of spatial databases, so the research aims to develop such functionality.

1.2 RESEARCH IDENTIFICATION

Although textual descriptions provide vague information as mentioned in the problem statement, such vague information may come across in different forms in GIS. There is a need to handle objects associated with such information in GIS and spatial databases as such the concept of vague data types and related vague operators were introduced by various researchers. Furthermore, the unified approach of object and field models were also introduced to represent objects with indeterminate characteristics.

Based on the concept of unified or hybrid model, the study will try to represent vague or uncertain objects by the combination of vector and raster named 'tandem approach'. Once the objects are represented in a spatial database there is a possibility to use them in various applications. Such as in gazetteer and GIR where user may require the answer of spatial queries like 'Find a geographic feature located 25 km NW of X'. There is a possibility to deliver result of such query with visualization by the design associated uncertainty. While finding uncertainty, the study has targeted to create uncertainty raster as well as to develop the storage and overlay techniques of tandems.

1.2.1 Research objectives

Uncertainty representation in spatial databases is one of the indispensable issues in spatial domain. The primary objective of this study is to develop new functionality to represent uncertainty associated with objects having uncertain spatial extent using tiny rasters in spatial databases and finding their possible spatial position.

This primary objective can be obtained by the following sub objectives.

- To develop functionality to represent objects with uncertain spatial extent in spatial databases by the combination of vector and raster, i.e., tandem.
- To develop functionality for overlays between two tandems.

1.2.2 Research questions

- What is/are appropriate model/s of the spatial uncertainty?
- What precisely does raster cell value mean in such formed raster?
- How can combined data structures of tandem help to represent that uncertainty?
- How can PostgreSQL/PostGIS store raster/vector in combination as tandem?
- How do two uncertain objects determine their uncertain area of overlap?

1.2.3 Innovation aimed at

The following innovations are intended in this research:

- Storage of raster/vector combination in spatial databases to express spatial uncertainty around vector objects.
- Development of functionality for overlay function of such raster/vector combination layers in spatial databases for objects that display spatial uncertainty.

1.3 BACKGROUND

Uncertainty management in spatial database systems is one of the major requirements observed. Not all spatial phenomena which we do need to represent in databases have exactly defined boundaries like cadastral and administrative boundaries. There exists various spatial phenomena whose position and extent can not be defined clearly or vague phenomena which are gradually changing such as population density, pollution, air pressure, temperature, vegetation. Similarly, boundary between grassland and forest, valleys and mountains or hillsides and mountains. In these cases boundary can not be defined in a crisp way. Currently, geographical information systems and spatial databases can only deal with crisp spatial objects which have fixed extent, shape and boundary and they are unable to deal with objects with indeterminate boundaries (Schneider, 2008). Various researchers contributed to conceptualize this issue.

In order to support such kind of data for vagueness or fuzziness, an abstract and conceptual model called fuzzy spatial data type was proposed in (Schneider, 1999, 2008). Major fuzzy spatial data types introduced based on the framework of fuzzy set theory were fuzzy points, fuzzy lines, and fuzzy regions. In the same way, a conceptual and an implementation model of fuzzy spatial objects was defined in (Schneider, 2003) based on a finite collection of elements from a regular grid which form a partition of a bounded subspace of R^2.

Similarly, Dilo et al. (2007) described the requirement of vague objects and operators to represent and analyze vagueness. They presented refined mathematical definitions for vague objects and operators based on fuzzy theory. Vague objects defined are categorized in three parts as a simple type, a general type and a vague partition. Furthermore, the simple type contains objects like vague points, vague lines and vague regions which can't be further divided into components. The general type contains vague multipoint, vague multiline and vague multiregion whereas the vague partitions is a collection of vague multiregions.

In the same way, Zhao et al. (2009) introduced random sets to model spatial objects with uncertain boundaries. Random sets mainly followed the concept of probability theory to model uncertainty along with fuzzy concept. Based on random sets, concept of random data types: random point, random line and random region were defined in two-dimensional Euclidean space (R^2). Similarly, random sets have been used in image space to model uncertainties of spatial objects (Zhao et al., 2010).

These conceptual models for representing spatial objects with uncertainties are the major motive to formulate a data type in spatial databases environment which can represent positional uncertainty of a spatial object.

1.4 THESIS OUTLINE

Chapter 1: Introduction

This chapter includes motivation and problem statement, objectives and research questions,innovation aimed at and background of the research along with this outline.

Chapter 2: Literature review

This chapter describes some basic concepts like spatial uncertainty and uncertainty modeling techniques, object-field model, hybrid model etc. which are necessary for the successful accomplishment of the project.

Chapter 3: Tandem Data Model

This chapter includes a proposed tandem data type with its conceptual data model. It also described required tools and techniques for the creation of uncertainty raster and its visualization. Basic design principles for creating uncertainty rasters are also defined. It also includes methodology adopted to generate raster cell values and the process of creating such raster. Besides this it contains the result and discussion part for the uncertainty raster obtained.

Chapter 4: Overlay between tandems

It illustrates about overlay operation between two tandems, where both tandems represent the same locality along with their implementation in the spatial database. Result of overlay operation is defined and discussed at the end of the chapter.

Chapter 5: Conclusion and recommendations

In this final chapter conclusion about the research, answers to research questions and some further recommendations are described.

Chapter 2
Literature review

The purpose of this chapter is to specify fundamental concepts required during the research. An overview of spatial uncertainty and its different forms and dimensions is provided in Section 2.1.1. A brief introduction about textual description, which is taken as a motivating example in the research as a basis of spatial uncertainty, and its related aspects is given in Section 2.1.2. A general introduction about a data model in spatial database management system is provided in Section 2.1.3. In addition, spatial uncertainty modelling techniques are reviewed in Section 2.2. Uncertainty modelling techniques defined are probability theory, random set and fuzzy theory. As a conceptual model for spatial objects representation, object model, field model and unified object-field model along with the necessity of their unification are described in Section 2.3. Similarly, the concept of hybrid data model is described in Section 2.4.

2.1 CONCEPTS AND DEFINITIONS

2.1.1 Spatial uncertainty

Generally, spatial uncertainty caused by lack of knowledge about spatial objects. Based on English literature, a deficiency in the knowledge of information is termed as *uncertainty*. In geographical information science, uncertainty may appears in different forms termed as *spatial uncertainty*. Generally, terms like error, accuracy, precision, vagueness or fuzziness and ambiguity are often used while talking about spatial uncertainty. As defined in Section 1.1, in the context of spatial databases Schneider (1999) defined spatial uncertainty based on indeterminate boundaries. Similarly, Dilo et al. (2007) defined vagueness as spatial uncertainty. Fisher (1999) defined uncertainty in some different context based on how well objects or classes are defined. The figure below illustrates Fisher's different types of spatial uncertainty existing in the spatial domain, along with the method that can handle them.

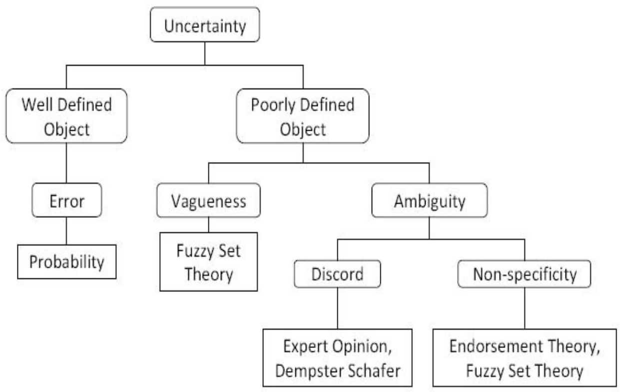


Figure 2.1: A classification of spatial data uncertainty and methods to address them (Fisher et al., 2006)

Based on the above conceptual model, if the object is well-defined (e.g., a cadastral boundary) then uncertainty is termed as error and can be handled by a probability function. If the object is poorly defined (e.g., a soil boundary or vegetation boundary) then uncertainty is defined as vagueness or ambiguity. Generally when the classes, to which objects belong are defined poorly, vagueness arises in the spatial domain. For example, there is vagueness between the classes grassland and forest because there is no distinct boundary. Vagueness can be conveniently handled by fuzzy theory. In the same way, ambiguity is associated with doubt that exists during the classification of an object based on different perceptions about it. Ambiguity is further divided into two parts: discord and non-specificity based on the same conceptual model. If there is disagreement about classification to which class an object belongs to then discord occurs. For example, as mentioned by Fisher (1999) about Kashmir area that both India and China claimed resulted to discord in the political classification of geographic area. Possible solutions defined for discord are expert opinion and artificial intelligence (AI) methods, including Dempster Schafer's theory of the evidence. Non-specificity occurs when a class is formulated in such a way that some of the related objects will not come inside the class. In other words, non-specificity occurs when the appropriate class for assigning a feature is missing. Based on (Fisher, 1999), non-specificity can be described by spatial relationship. Such as an expression, A is north of B provides different meaning in reality. A might be exactly north of B, might be north-west or north-east so non-specificity may occur in such statements. Based on the above taxonomy non-specificity can be treated as discord by AI including endorsement theory as well as fuzzy-set theory.

Similarly, Molenaar (2000) defined extensional uncertainty associated with an object having uncertain spatial extent. Additionally, The extensional uncertainty refers to uncertainty in identifying geometric element that determine the spatial extent of the object.

Spatial objects with uncertain boundaries, where uncertainty is quite related to positional uncertainty and extensional uncertainties, are going to be handled in this project. Mainly point objects they are obtained based on textual descriptions will be experimented with this project to represent it along with its positional uncertainty and to find its possible extent.

2.1.2 Textual description as a motivating example

Textual description is often used in communication to indicate a locality. Similarly, in past when there was no proper way to manage geographical information, these textual descriptions used to handle their record. There is a need to identify a location indicated by such textual description and to identify associated uncertainty because these descriptions provide vague information. According to Liu et al. (2009), the following four uncertainties are mainly considered:

- Uncertainty of textual description itself: Suppose when we say '25km NW of Belem', how much it is certain that the statement is itself true. Though distance is 25km, how probable is that a target object (an object of interest) is exactly 25km far from a reference object i.e., Belem. Similarly what is the probability of target object's position NW of Belem.

- Imperfection of reference object: In case of above example, X is a reference object. A position of the target object is completely affected by imperfection of the reference object. Imperfection of reference object may appear as ambiguity when we indicate it by name. Similarly in above example X ia a place name and it may not have a determinate boundary which also generates uncertainty.

- Imprecision and vagueness of spatial relationship: In the above example, there exist two spatial relationships between target object and reference object. The first is directional (NW) and the second is distance (25km). NW provides vagueness whereas 25km provides measurement error. It may be 24.0125 or 25.9502.

- Distribution range of target object: Generally, each spatial relationship in a textual description provides a particular distribution range of target object. Textual description like target object is in the south of Belem gives a distribution of target object inside Belem.

In this project a textual description is taken as a **motivating example** to express uncertainty of spatial objects. In (Liu et al., 2009), these descriptions were considered during geo-referencing locality, though geo-referencing is beyond the scope of the research.

2.1.3 Data model

A data model in geographical information science and spatial database management system is a conceptualized model that defines how spatial objects are organized, stored and accessed in spatial databases. A Data model is a set of guidelines for the representation of spatial data in databases as a logical units and provide relationship between them (Goodchild, 1992). Furthermore, Goodchild emphasized data model in spatial databases as the fundamental unit that helps to present the real world among database users as ultimate product.

In the project, a tandem data model will be conceptualized to store spatial objects with their uncertainty and allow a mechanism for computation in spatial databases environment along with realization. The proposed tandem data type and its conceptual model is defined in Section 3.1.

2.2 UNCERTAINTY MODELLING TECHNIQUES OF SPATIAL OBJECTS

This section defines some important techniques of modelling spatial uncertainty. A basic concept of probability theory has been defined in the first part of Section 2.2.1. Later, this section describes Gaussian distribution curves and their properties, which are mostly used in probability applications. In addition of this, the concept of joint probability is defined and the formula to find the probability of a point being inside a rectangular region is derived. Generally, probability theory is used to handle positional uncertainty. In addition of positional uncertainty, there are some other uncertainties like vagueness, imprecision and ambiguity which are entirely different from positional uncertainty. They may need to deal with fuzzy techniques or random sets. A random set and its related aspects is defined in the Section 2.2.2. It uses the probability as uncertainty distribution along with acquiring the techniques of fuzzy set and may be used to represent spatial objects with uncertain boundaries. Finally, a general concept of fuzzy theory and some of its uses is defined in Section 2.2.3.

2.2.1 Probability theory

In probability theory, the probability of event A occurring can be estimated by repeating a random experiment n times. If m is the number of occurrences of A out of n observation in the experiment then the ratio m/n is called the relative frequency of event A. Whenever n is sufficiently large, based on limit theory the ratio m/n approaches to constant probability P at event A.

$$\lim_{n \to \infty} \frac{m}{n} = P \tag{2.1}$$

Probability theory is used to study random phenomena and can be used to handle random error in spatial data. When a value of a variable is the outcome of a random experiment that variable is known as a *random variable*. Representing uncertainty using probability theory uses statistically defined probability functions to generate probability value. Probability functions are of two types; discrete and continuous. The probability that the variable X can take a particular value x is denoted by $P(X = x) = P(x)$ and is called discrete probability. The value of $P(x)$ lies between 0 and 1 so it is always positive and the sum of $P(x)$ for all possible values of X is 1. Similarly, the continuous probability function is generally denoted by $f(x)$ and it can be described as follows.

- The probability that x is between two points a and b is $P(a \leq x \leq b) = \int_a^b f(x)\mathrm{d}x$.
- This is non-negative for all real x in (a,b).
- The integral of the probability function from $-\infty$ to ∞ is equal to 1 i.e., $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$.

A common type of continuous distribution function is a normal distribution also called Gaussian distribution. A normal density function for a single variable X can be defined as :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad , \tag{2.2}$$

where π and e are constants, μ and σ are mean and standard deviation of the distribution respectively. This is the typical normal density function where both μ and σ are real numbers and σ is greater then zero. The normal distribution is used in many statistical tests due to its basic mathematical properties. It is illustrated by a symmetric, bell- shaped probability density curve. Based on the above normal density function it is obvious that normal distribution relies on mean and standard deviation of the random variable. Actually, the normal distribution is symmetric about the mean, where peak density occurs. Standard deviation indicates spread of the curve from the mean. The total area under the normal curve is equal to 1. In normal curve, approximately 68% of all observations fall within one standard deviation (σ) of the mean (μ) i.e., in the interval $(\mu - \sigma, \mu + \sigma)$. Similarly, 95% of all data will fall within two standard deviations (2σ) from the mean and 99.7% of the data will fall within three standard deviation (3σ) from the mean. Hence, in normal distribution almost all data lie within three standard deviations from the mean. The figure below shows a normal curve with its class intervals.

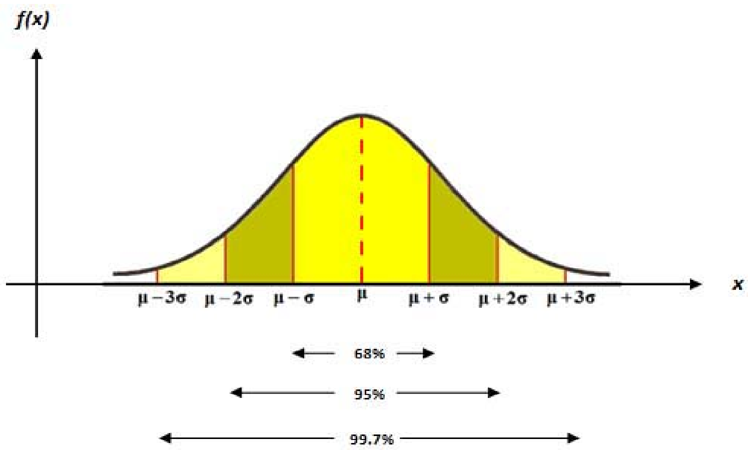


Figure 2.2: Normal (Gaussian) density function

As defined in (Fisher, 1999) probability theory is mainly used in handling error in the case of well-defined objects. Based on Schneider (1999), probability theory can be used to handle positional or measurement uncertainty. Probability theory is used by Guo et al. (2008) and Liu et al. (2009) to find the uncertainty distribution of point object in case of handling textual descriptions or spatial assertions. Similarly, probability theory used in (Glemeser and Fritsch, 1998) to define the geometric uncertainty of an object where geometric uncertainty indicates variation in position of geometric features point, line and polygon.

Joint probability

A joint probability is the likelihood of two events occurring together and at the same point in time. In other word joint probability is the probability of event y occurring at the same time when event x occurs. It is generally denoted by $p(x \cap y)$ or $p(x, y)$.

As defined in (Knisley, 2001), suppose that (x, y) is the outcome of a certain experiment such that it must occurs in some region. The region can be considered as sample space S where sample space indicates all possible outcome of that experiment, then x and y are called *random variables*.

Let Δp be the small probability that (x, y) is in a small region R inside S. In this scenario a function $f(x, y)$ is called a *joint probability density* of the experiment and followed following equation.

$$\Delta p = f(x, y)\Delta A \tag{2.3}$$

where ΔA is the area of the small region R as shown in figure 2.3. Equation (2.3) defines that joint probability density can be defined as the *probability per unit area* of the experiment.

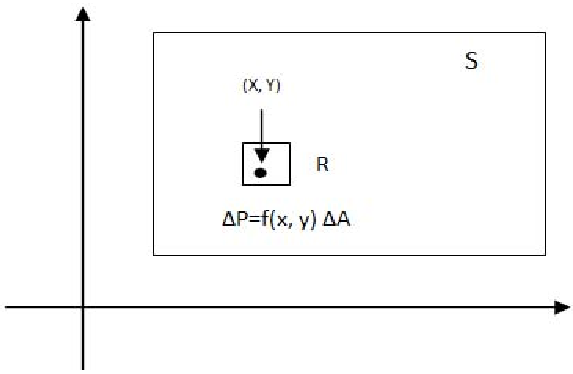


Figure 2.3: Joint probability

Let us suppose that h be a small incremental size considered while doing fine partition of both the x-axis and y-axis, then the probability Δp of (x, y) in R is approximated by the 'Riemann sum' rule as:

$$Probability \approx \sum_j \sum_k \Delta p_{jk} = \sum_j \sum_k f(x, y)\Delta A_{jk} \tag{2.4}$$

The actual probability of (x,y) being in R can be approximated when limit of the h tends to zero.

$$Probability = \lim_{h \to 0} \sum_j \sum_k f(x, y)\Delta A_{jk} = \iint_R f(x, y)\Delta x \Delta y \tag{2.5}$$

Equation 2.5 provides the probability of (x, y) being inside the region R. Since there is 100% certainty that (x, y) lies inside the sample space S, the sum of all possible probability value should be equal to one as follows:

$$\iint_S f(x, y)\mathrm{d}x.\mathrm{d}y = 1 \tag{2.6}$$

In the project, the concept of joint probability will be used to find the probability of a point object being inside a raster cell of the raster part of tandem data model.

2.2.2 Random sets

Random sets can be used to model uncertainty of spatial objects by using the concept of probability theory. In (Zhao et al., 2009), random set is used to model boundary uncertainty. Similarly, it is used to model extensional uncertainty of natural entities extracted from satellite images (Zhao et al., 2010). In addition of these, major random data types are defined in (Zhao et al., 2009). These data types were defined in the context of image space or raster environment where semantics of pixel value is defined for respective data types.

In the project the uncertainty is related to extensional uncertainty and which is being represented by tiny rasters. In such scenario, it seems that the random set can be a suitable uncertainty modelling techniques for the project.

Random sets are random elements taking values as subsets of some space and can be defined by a collection of pair (Zhao et al., 2010). For example a random set Γ from n dimensional space R^n can be defined as: $\Gamma = \{(\gamma_i, m_i) | \gamma_i \subset R^n, m_i \in [0, 1], i \in \{1, \dots, n\}\}$. Where γ_i is called a focal elements of the random set. Similarly, m is an uncertainty assignment value which is associates with each focal element and quantifies how likely these elements belong to Γ, where $m(\emptyset) = 0$ and $\sum_{i=1}^{n} m(\gamma_i) = 1$.

Zhao et al. (2009) defined three major random data types: random Point, random line and a random region.

Let us suppose that the space of a raster be $U_R \subset R^2$ in two dimensional Euclidean space, where the cell or pixel ξ of the raster is a basic element. Then major random data types defined in (Zhao et al., 2009), based on pixel ξ are as follows.

The random point (R_p) is defined as a random set in R^2 that contains a finite collection of ξ with positive probability values, denoted as $R_p(\xi_{p1}, \xi_{p2}, \dots, \xi_{pn})$. The probability value with each ξ is the likelihood that the point is in this location.

The random line (R_l) can be denoted as $R_l\{(l_1, p_1), (l_2, p_2), \dots, (l_n, p_n)\}$. Here l_i is uncertain lines that compose a line object, called focal elements. It is constructed by a set of ξ with a same probability value p_i. The probability value attached with each ξ is that the likelihood of the line goes through the location of this element.

The random region (R_r) can be denoted as $R_r\{(r_1, p_1), (r_2, p_2), \dots, (r_n, p_n)\}$. Here r_i is uncertain region that compose a region object, called focal element. It is constructed by a set of ξ with a same probability value p_i. The probability value attached to each ξ is that the likelihood of the region intersects with the location of this element.

Moreover, Zhao et al. (2010) defined that the random set associates a function $Pr_\Gamma : R^2 \to [0, 1]$ to every $\xi \in R^2$. This function is called the covering function. The covering function of a random set can be estimated from its focal elements $\gamma_i, i = \{1, 2, \dots, n\}$ as:

$$Pr_\Gamma(\xi) = \frac{1}{n}\sum_{i=1}^{n} I_{\gamma_i}(\xi), \xi \in R^2, \gamma_i \in p(R^2) \tag{2.7}$$

where I_{γ_i} is the indicator function of γ_i:

$$I_{\gamma_i} = \begin{cases} 1, & \xi \in \gamma_i \\ 0, & \xi \notin \gamma_i \end{cases} \tag{2.8}$$

A random set can be categorized in the following three major sets based on the value of covering function i.e., $Pr_\Gamma(\xi)$ and are called p-level sets. Thus P-level sets are as :

- The set $\Gamma_c = \{\xi \in R^2 : Pr_\Gamma(\xi) = 1\}$ is called the core set. It describes the certain part of Γ.

- The set $\Gamma_\alpha = \{\xi \in R^2 : Pr_\Gamma(\xi) \geq \alpha\}$ is called the α-level set of Γ. It indicates a combination of core set and the set defined by α value. Such as a median set is the set where $\alpha = 0.5$.

- The set $\Gamma_s = \{\xi \in R^2 : Pr_\Gamma(\xi) > 0\}$ is called the support set of Γ. It indicates the possible part of set Γ.

The formal definition of random sets along with its related aspects can be found in (Nguyen, 2006; Nuñez-Garcia and Wolkenhauer, 2002).

2.2.3 Fuzzy sets

Based on the concept of fuzzy theory Schneider (1999, 2008) defined an abstract, conceptual model of fuzzy spatial object introducing fuzzy point, fuzzy line and fuzzy regions. Similarly, with the concept of fuzzy theory Dilo et al. (2007) provided a mathematical definition for vague object types and operators. These data types and operators have been defined to represent spatial objects with indeterminate boundaries.

Fuzzy set theory is an extension and generalization of Boolean set theory developed by Zadeh (1965). It can be represented by an ordered set of pairs. Fuzzy set A in X (universe of discourse) is as: $A = \{(x, \mu_A(x)) | x \in X\}$. In this, $\mu_A(x)$ is membership value of element x within the closed interval [0, 1] and is generated based on the membership function defined as $\mu_A : X \rightarrow [0, 1]$. The membership value $\mu_A(x)$ represents how much the elements x belongs to the fuzzy set A.

In this project extensional uncertainty is being represented in tiny rasters. Although the uncertainty relates to spatial position of the objects, it prevent us to describe the boundary of spatial objects. In such situation, the fuzzy theory can be used to model uncertainty.

To determine the location of a particular locality identified by linguistic descriptions (similar to textual descriptions) fuzzy theory was applied by Hwang and Thill (2005) to model a locality. Hwang and Thill (2005) categorized geographical space into three components; it means locality divided into three parts: *core*, *boundary* and *exterior* as shown in figure below.

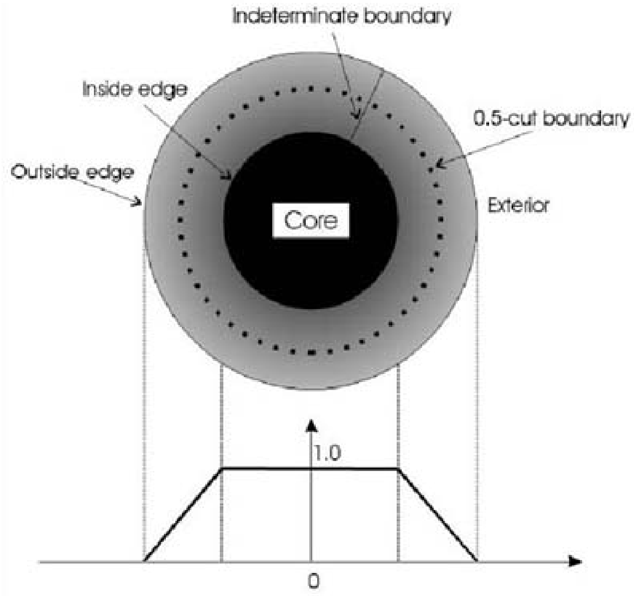


Figure 2.4: Representation of fuzzy region (Zhan and Lin, 2003)

Hwang and Thill (2005) defined a locality in the geographical space as a fuzzy region and defined it formally as follows: Let *Fr* be a fuzzy region. *Fr* is composed of three parts: *Core*, *Boundary* and *Exterior* areas which are considered as crisp regions (regular closed sets). R^2 represents two-dimensional geographic space. μ_{Fr} be the fuzzy set membership function of Fr.

$$\begin{aligned} Core(Fr) &= \{(x,y) \in R^2 | \mu_{Fr}(x,y) = 1\} \\ Boundary(Fr) &= \{(x,y) \in R^2 | 0 < \mu_{Fr}(x,y) < 1\} \\ Exterior(Fr) &= \{(x,y) \in R^2 | \mu_{Fr}(x,y) = 0\} \end{aligned}$$

The membership value 1 represents full membership. The 0 membership value represents no membership and the values between one and zero in decreasing order shows decreasing membership towards the object.

2.3 OBJECT-FIELD CONCEPT

Object and field model have been introduced as two alternative approaches for conceptualizing and modelling geographical phenomena (Goodchild, 1989, 1992). The field model has also been called location-based model and considers each location in space is mapped to a value from attribute domain of spatial objects. Elevation, precipitation and temperatures are common examples of spatial variables that are modelled by field-based model. In the context of object model space is occupied as a number of discrete entities or things with their own identity and attributes. River, city and building can be considered as three different classes of object and are often modelled by an object-based model. In the context of representing complex geographic phenomena with indeterminate boundaries like atmospheric pressure, wild fire, epidemics and mountains which are continuously changing over time, there seems a requirement for the unification of these two models because a single model can not represent them completely. For example, if we represent atmospheric pressure ridges using an object-based model we are unable to show the variation of atmospheric pressure within the ridge whereas if we represent these ridges by a field-based model it shows pressure variation but boundaries and dimensions of the ridges are not clearly modelled. Finally, Voudouris (2010) added uncertainty and semantics to the combined object-field model and introduced conceptual and logical levels of abstraction. He introduced a case of town centers as indeterminate phenomena and conceptualized associated existential and extensional uncertainty.

The current project experiments with a mechanism that can represent spatial objects with uncertain boundaries by the combination of vector and raster quite similar to the concept of unified object-field model.

2.4 HYBRID MODEL

A hybrid model, which can process both raster and vector data in an integrated manner taking uncertainty into account, is introduced in (Glemeser and Fritsch, 1998). It is named hybrid model based on the integration of raster and vector together. In addition of this, a hybrid overlay function is developed which can be applied without consideration of the spatial data type (raster or vector) and spatial objects (point, line or polygon). Besides this the function takes advantage of the fact that the overlay operation is better in raster than vector data representation.

Similar to this hybrid model i.e., combination of vector and raster, the current project uses the tandem approach (vector/raster combination) to represent spatial objects with uncertain spatial extent. The major two models involved in a hybrid model are as follows.

2.4.1 Vector model

In a vector model, each location is represented by a single (x, y) coordinate. A point object is recorded as a single coordinate. A line object recorded as a series of order pairs of coordinates. Similarly, a area or a polygon object is recorded as series of ordered (x, y) coordinates that constitutes line segments and finally enclosed the polygon.

A vector object in the proposed tandem data model considered as the crisp part of spatial objects. In this sense the vector object represents best possible position of the object along with characterizing geometry of spatial object as point, line string or polygon.

2.4.2 Raster model

The raster model divides space into grid cells or pixels. Each grid cell is filled with measured attribute values and generally contains discrete or continuous value. Each cell is identified by their positions on the grid as cell co-ordinates (i, j); where i is the row number and j is the column number in a table. The cell width in ground units such as meter or kilo meter indicates the resolution of raster data.

Raster data in the tandem data model provides supplementary information about the position of the vector object and define the possible spatial position of the object.

2.5 SUMMARY

In this project, an uncertainty associated with spatial objects due to textual descriptions is going to handle. Currently, uncertainty associated only with point objects is considered in the project where a point object defined by a textual description represents a locality. When saying about a locality it is obvious that it should have some spatial extent. However, the locality associated with a point object and described by a textual description has uncertain spatial extent. This is the uncertainty going to handle in the project. Uncertain spatial extent implies a spatial object without precise boundary. In this scenario it is difficult to say about the exact boundary of the locality and can be considered as vagueness. Again, when talking about uncertain spatial extent it is possible to determine its boundary based on the position of the boundary points. In this case it relates to the positional uncertainty.

In Section 2.2 three uncertainty modelling techniques along with uncertainty, which they can handle, are defined. Thus, in the gist it seems that all three uncertainty moddeling techniques can be used to design the uncertainty of the spatial object having uncertain spatial extent. However, the project will follow the probability theory to represent the uncertainty associated with point objects where error will be represented by standard deviation. Detailed description about uncertainty value calculation method using probability theory is defined in Section 3.2.4.

Chapter 3

Tandem data model and methods to create uncertainty rasters

This chapter includes a proposed tandem data type with its conceptual data model described in Section 3.1. Before the creation of uncertainty raster design principle should be formulated as such it has been defined in Section 3.2.1. Tools required to create and display an uncertain raster are defined in Section 3.2.3. The technique used to generate uncertainty value in each pixel of a raster is defined in Section 3.2.4. Textual descriptions handled during the project and details of the procedure followed to determine dimension of uncertainty raster in respective cases are defined in Section 3.2.5. Creation of the tandem data type in the database is defined in Section 3.3. Finally, the results and discussion part of the implementation of raster creation is discussed in Section 3.4.

3.1 CONCEPTUAL DATA MODEL FOR TANDEM DATA TYPE

The conceptual data model for proposed tandem data type is shown in figure 3.1 below.

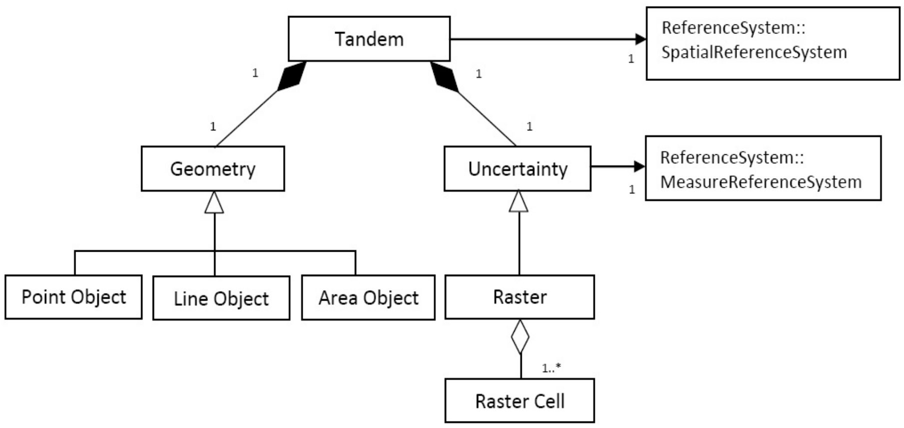


Figure 3.1: Conceptual model of the tandem data type

Tandem is the root class of the model. Tandem is an instantiable class i.e., we can create instances of this class. It is composed of geometry and uncertainty classes. Both geometry and uncertainty are abstract classes and their objects are associated with the same spatial reference system. Spatial reference system describes the coordinate space in which geometric and uncertainty objects are defined. In addition of spatial reference system, uncertainty class object associated with measure reference system which describes the uncertainty value.

Point, line and area are instantiable subclasses of the Geometry class. They are restricted to 0, 1- and 2-dimensional geometric objects and exist in coordinate space R^2 in this data model. The uncertainty class has raster class which is an instantiable class and consists with the class Raster

cell. Raster cell is an abstract and leaf class of the hierarchy.
Methods adopted to create uncertainty rasters are defined in following sections.

3.2 CREATION OF AN UNCERTAINTY RASTER FOR THE TANDEM DATA TYPE

Prior requirements to create an uncertainty raster are as follows:

3.2.1 Raster design principles for the tandem data type

To design an uncertainty raster of a tandem data type following design principles should be considered:

- Spatial reference system(SRS):
  The Universal Transverse Mercator (UTM) projection system is used as spatial reference system.
- Cell shape:
  The shape of raster cells (grids/pixels) taken as square i.e., equal cell width and cell height.
- Origin:
  Origins of the raster and UTM coordinate system are different. For the raster the origin is upper-left corner whereas for the UTM system it is lower-left corner.
- Cell size:
  The equal cell size of $\delta \times \delta$ has been selected in the creation of the raster.
- Dimensions:
  Dimensions of the raster are identified based on the type of textual descriptions going to handle. Techniques used to find the appropriate dimensions of the raster are defined in the Section 3.2.5 below.
- Cell value type:
  Raster cell contains cell value of type double precision as defined in PostGIS raster.

3.2.2 Design principles for raster creation function

During the project raster creation functions will be developed to create raster with desired dimension and cell size. There is need to insert cell values to generate uncertainty raster for tandem data type. Including these prior needs similar type of requirements should be identified first for the raster creation function. Thus, major design principles are as follows:

- Creating an empty raster:
  First requirement must be how to create an empty raster with desired width, height and cell size. *PostGIS 2.0.0SVN* provides a function named "ST_MakeEmptyRaster()" that able to create the desired empty raster. It takes arguments like width, height, cell size both in x and y direction, SRID and coordinate of the origin of the raster.
- Assigning cell values:
  The next step should be assigning proper cell values, however there is need to assign a band to the empty raster first before assigning cell values. A function named "ST_AddBand()" can be used to add band with desired pixel types. Then a function "ST_SetValue()" can be used to assign cell values to cell located at (j, i).

- Conversion between coordinates:
  During the uncertainty raster creation there might require a mechanism that can convert the spatial coordinate system to cell coordinate and vice versa. There are Functions like "ST_Raster2WorldCoordX()" and "ST_Raster2WorldCoordY()" provide X and Y coordinates in a spatial coordinate system from cell coordinate system respectively for a upper left corner of a cell. Similarly, functions "ST_World2RasterCoordX()" provide X and the function "ST_World2RasterCoordY()" provide Y coordinates in a cell coordinate system from spatial coordinate system.

3.2.3 Tools and techniques used

In the research project, "PostgreSQL 9.0" and "PostGIS 2.0.0SVN" are used as spatial databases system. PostgreSQL is an object-relational spatial database management system and an open source software. PostGIS is a free and open source library that spatially enables PostgreSQL. pgAdmin 1.12.3 which is a graphical user interface is used to access postgrSQL . pgAdmin is generally packaged with PostgreSQL and installed automatically along with it but can be obtained individually.

"PostGIS 2.0.0SVN" is the recent version of PostGIS released on "2011-10-06 08:02:01" and raster support is the core part of it. In the prior version of PostGIS, raster support functionality is associated separate package with separate installation. It can be installed in Windows operating system (OS) with its binary file due to unavailability of executable file for windows operating system so far.

For the display purpose of created raster Quantum GIS (QGIS) version 1.7.1 has been used along with the Plugin named "Load PostGIS Raster to QGIS" 0.5.1. version.

During the research there is need to upgrade the PostGIS to more recent version because the above defined version of PostGIS does not contain a function that can take two rasters as argument. The latest version of "PostGIS 2.0.0SVN" released on "2012-01-07 19:35:20" is used to solve the overlay issue. Similarly, QGIS plugin is also upgraded to the latest one which is released on 2012-01-20 with version 0.5.3. Details about the QGIS plugin can be found in (de Paulo and Vinhas, 2011).

The operating system used is Windows 7 home premium with 32-bit system.

In addition of these major software there is need of supportive software which helps to run "PostGIS 2.0.0SVN" in Windows operating system efficiently. "Python 2.7.2", "numpy-1.5.1-win32-superpack-python2.7", "GDAL 1.8" and Python binding "GDAL-1.8.0.win32-py2.7" are used to support PostGIS raster loader.

In the project "MATLAB-7.10.0" is also used to find and analyze the bivariate standard cumulative distribution function.

3.2.4 Uncertainty value calculation method

The concept of joint probability and double integral defined in Section 2.2.1 is used to generate the probability value of the target point object being inside every cell of a raster. Since calculating double integral of probability density function is a difficult procedure. The integral belongs to the branch of mathematics called *CALCULUS*. Thus, an alternative way is used to find probability rather than solving the actual double integral. In fact, the double integral represents a volume under the surface $z = f(x, y)$ over a region D. So, in the project this alternative method of solving double integral is adopted. The overall method is described below.

If two variables X and Y follow the bivariate distribution curve, where mean and standard deviation of X are μ_x and σ_x respectively. The same for Y as μ_y and σ_y. The ρ represents correlation coefficient between X and Y, then $f(x,y)$ represents the standard bivariate density and can be defined as:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{\{-\frac{1}{2(1-\rho^2)}\}\{(\frac{x-\mu_x}{\sigma_x})^2 - 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y}) + (\frac{y-\mu_y}{\sigma_y})^2\}} \tag{3.1}$$

Now, to calculate the probability of a point object being inside a raster cell, the volume of a solid between the probability density function and raster cell ground level is determined. To calculate the volume of the solid, a *five point concept* is applied where these five points are four corner points and the middle point of the raster cell. When the probability density at each five points is calculated using Equation 3.1 then the shape of that solid object can be considered as *pyramid shape* as shown in Figure 3.2.

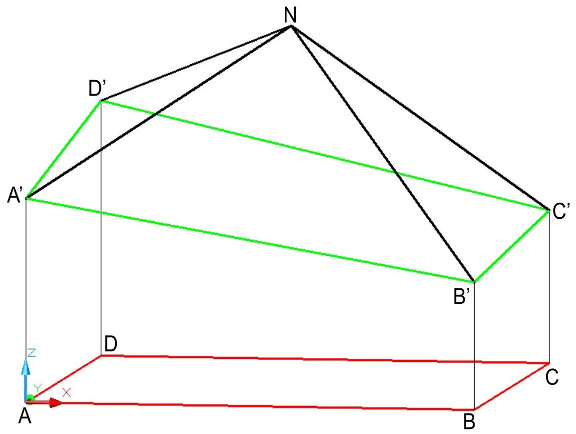


Figure 3.2: Formation of pyramid

As shown in Figure 3.2, $ABCD$ is a raster cell A', B', C' and D' are the points obtained by calculating probability density on the respective corner of the raster cell. The middle point of the raster cell now becomes apex point, i.e., rooftop vertex N and $A'B'C'D'$ becomes the base of the pyramid. Now, the probability of point object being inside cell can be obtained by calculating the volume under this pyramid. Since, the probability densities at each point are generally not equal it is rare chance that the base of the pyramid may form linear plane or is parallel to the xy plane in three-dimensional space. The volume under the pyramid is calculated the volume under each of the four triangles of the pyramid first and their summation as approximate volume under the pyramid which is actually the required probability value. The process followed to calculate uncertainty values is shown in Figure 3.3.

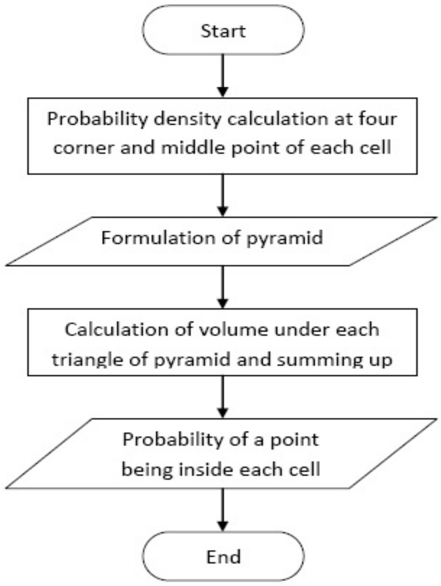


Figure 3.3: A five point technique to calculate uncertainty value

The volume under a triangle can be obtained by projected area of the triangle times the average height of three vertices of the triangle. The derivation of the formula for volume under a triangle is described below. The adopted method for the volume is provided in (`http://mathpages.com/home/kmath393.htm`).

Let us consider points $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$ and $P_3(x_3, y_3, z_3)$ as three vertices of a triangle. Where z-axis represents the vertical heights and our task is to find the volume under this triangle above the plane z=0. Here, these vertices have been numbered as P_1, P_2 and P_3 such that $z_1 \leq z_2 \leq z_3$.

Let us suppose that $P_4(x_4, y_4, z_4)$ be the intersection of the segment P_1P_3 with $z = z_2$ plane then, z_4 will equal to z_2. Now, in this case we have two triangle tetrahedrons as shown in Figure 3.4.

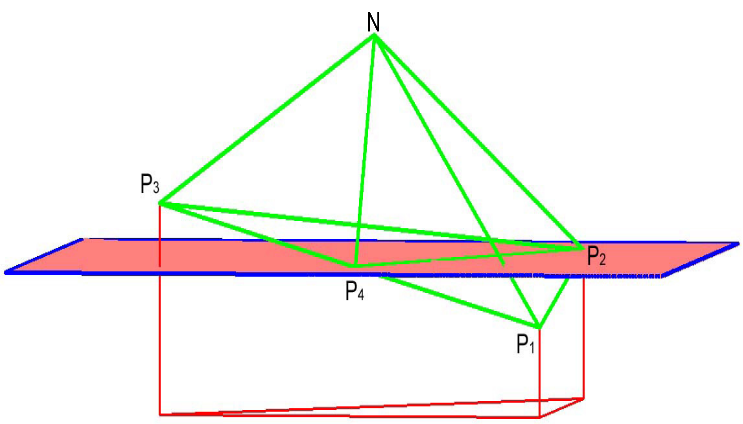


Figure 3.4: Volume under a triangle

Volume for each of two tetrahedrons with respect to $z = z2$ plane can be calculated by one third the projected base area times the altitude. If we consider A_1 and A_2 as areas of triangles $\Delta P_2P_3P_4$ and $\Delta P_1P_2P_4$ projected onto $z = 0$ plane respectively, then the volume under the original triangle can be calculated as:

$$V = \frac{1}{3}\left|A_1\right|(z_3 - z_2) - \frac{1}{3}\left|A_2\right|(z_2 - z_1) + (|A_1| + |A_2|)z_2 \tag{3.2}$$

where,

$$A_1 = \frac{1}{2}\{(x_4y_2 - x_2y_4) + (x_3y_4 - x_4y_3) + (x_2y_3 - x_3y_2)\} \text{ and}$$
$$A_2 = \frac{1}{2}\{(x_1y_2 - x_2y_1) + (x_4y_1 - x_1y_4) + (x_2y_4 - x_4y_2)\}$$

In fact, the volume of original triangle is calculated by adding the volume of each tetrahedron with respect to $z = z_2$ plane and then adding the volume of whole triangle with respect to $z = 0$ plane. Also, the value of x_4 and y_4 can be obtained using the concept of a point which divides a line in a certain ratio such that there values will be as follows:

$$x_4 = x_1 + \left\{(\frac{z_2 - z_1}{z_3 - z_1})(x_3 - x_1)\right\} \text{ and}$$
$$y_4 = y_1 + \left\{(\frac{z_2 - z_1}{z_3 - z_1})(y_3 - y_1)\right\}$$

Now, on substituting above values in Equation 3.2 the required volume is given by:

$$V = \frac{1}{6}(z_1 + z_2 + z_3)(x_1y_2 - x_2y_1 + x_2y_3 - x_3y_2 + x_3y_1 - x_1y_3) \tag{3.3}$$

It means the volume under the triangle is just the projected area times the average of three heights.

3.2.5 Raster creation methods

To create a raster for a provided textual description, separate functions are developed for each textual descriptions. In general, these functions take cell size, coordinates of provided point object and error information of the point object as function parameter and create the raster in a UTM projection system. Function parameters may vary based on the textual description that we are going to handle. The raster dimensions are determined by the technique used to create raster and which also depends on the type of textual description. All sections defined below contain the type of textual description and the techniques used to create raster. Although techniques to create raster may differ, all techniques follow the concept of the five point rule to calculate the uncertainty value of the raster cell as defined in Section 3.2.4.

First template

In this template, the textual description contains the place name, its coordinate values and error information. A raster around the point is created based on these given parameters and the following considerations.

- A raster should have a fixed UTM zone and the zone depends upon the provided point object. The SRID of the zone can be find by a routine based on the coordinate of the given point.

- As defined in Section 3.2.1, the cell width and height should be equal i.e., cell size $\delta \times \delta$ is considered in both x and y direction. In other words, the UTM grid should be fixed in both directions.
- The fixed grid matches with the value of δ and any raster origin is always at coordinate values that are a multiple of δ in both x and y direction.
- The raster should contain an even number of rows and column. Standard deviations σ_x and σ_y are considered equal. Considering even number of rows or column helps to determine the raster dimension. Besides this, equal sigma allows to locate the given point object at more or less in the middle of the raster. Locating the given object at the middle of the raster can able to capture the whole symmetrical distribution of the object.

The overall process to create uncertainty raster for first template is shown in Figure 3.5.

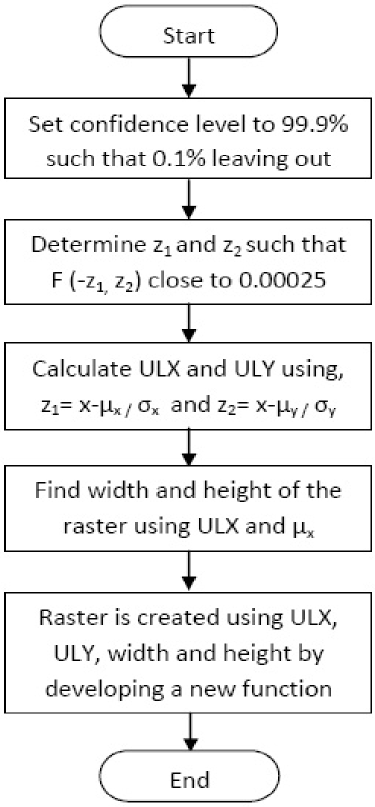


Figure 3.5: Uncertainty raster creation procedure, first template

Figure 3.5 illustrates the following procedure:

1. Set confidence level:
   First, the confidence interval of parameter is defined and it has been considered as 99.9%. It means that the raster we are constructing contains 99.9% of certainty that the true position of the point object is inside the raster and leaving out at most 0.1% certainty. Now the next step is to find the upperleft x and y coordinates, i.e., (ULX, ULY) and dimension of the raster which can capture the above confidence level. Moreover, the (ULX, ULY) and dimension are required for raster creation function in the PostgreSQL/PostGIS.

2. Origin of the raster:
To find the origin or (ULX, ULY), when both σ_x and σ_y are equal it can be considered that the cumulative value for this corner i.e., $F(ULX, ULY)$ is close to 0.00025. In fact the value here 0.00025 is one fourth of the outside certainty 0.001 or 0.1%. The reason for one fourth is the equal σ and spatial symmetry of the uncertainty values by which cumulative values of the all four corner of the raster equivalent to that outside certainty. Cumulative value at a corner represents the probability of a variable being inside a semi infinite region formed by the position that corner. It indicates that these coner points accumulate outside certainty. In this way in the bivariate normal distribution when inner certainty considered 99.9% and 0.1% as outside certainty then there shoud be a coordinate as $(-z_1, z_2)$ such that the value of $F(-z_1, z_2)$ should be close to 0.00025.

3. Finding the coordinate $(-z_1, z_2)$:
To find the co-ordinate pair $(-z_1, z_2)$ where $F(-z_1, z_2)$ is close to 0.00025, a function *mvncdf* () is used. This is the function available in the MATLAB software. The most relevant $(-z_1, z_2)$ pair found using the function is $(-3.4807, 3.4807)$ where $F(-3.4807, 3.4807)$ $= 0.00024999$. The MATLAB software is used only once here throughout the study because we unable to find the required statistical table that provide values for $F(-z_1, z_2)$.

4. Calculation for the origin:
Now, (ULX, ULY) calculated using the equations $z_1 = \frac{x-\mu_x}{\sigma_x}$ and $z_2 = \frac{y-\mu_y}{\sigma_y}$, where μ_x, μ_y, σ_x, σ_y, z_1 and z_2 all are known values. In fact, μ_x and μ_y are the x and y co-ordinates of the given point in UTM projection.

5. Finding dimension of the raster:
Once the (ULX, ULY) calculated there respective value is converted to multiple of scale. Then the distance between ULX and μ_x calculated where the value of μ_x is already converted to multiple of scale. This distance is divided by δ and multiplied by 2 which gives the desired width of the raster. Since width and height are equal so the height will also be same.

6. Creation of the raster and generating uncertainty value:
Finally, a function has been created to built the raster using calculated (ULX, ULY) and dimensions with appropriate SRID. The function also generates the uncertainty values using the five point concept defined in Section 3.2.4 and insert values on respective cells. The density function used in this case is defined in Equation3.1. The function developed is td_create_raster_first_template() and contains following parameters. The source code of this function is provided in Appendix A.

    - scalx: represents width of the raster cell with data type double precision.
    - scaly: represents height of raster cell with data type double precision.
    - point geometry: is the given point with data type geometry.
    - sigmax: indicates standard deviation in X-direction with data type double precision.
    - sigmay: indicates standard deviation in Y-direction with data type double precision.
    - rho: is the correlation coefficient with data type real.

Second template

In this template, the textual description contains the place name, its coordinates, distance and compass direction such as "22km SW of Belem [0127/4829]". Along with this information distance uncertainty and directional uncertainty have been provided. Based on these values a raster around the target point is created. Angles in this template are measured in radians and the north direction is considered as 0 rad. Angles are measured in clockwise direction such that $3 = \frac{\pi}{2}$ rad, $6 = \pi$ rad. The technique used to create raster in this scenario is shown in Figure 3.6.

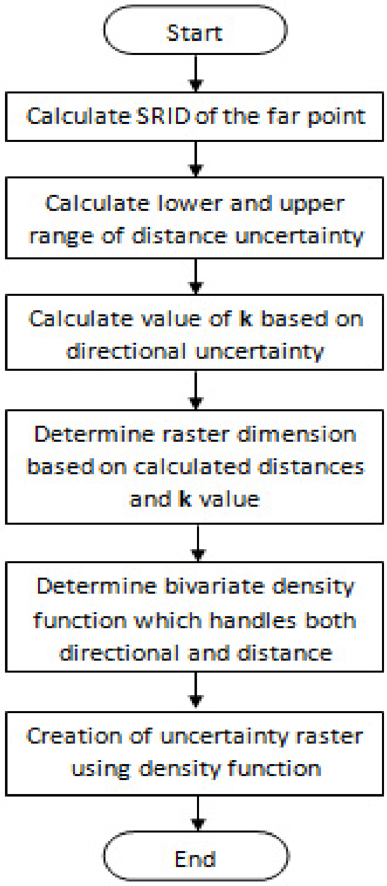


Figure 3.6: Uncertainty raster creation procedure, second template

The figure illustrates the following procedure:

1. Finding SRID of the target point:
   First of all the SRID of the target point is calculated in UTM projection based on the given point, distance and directional information.

2. Distance ranges calculation:
   Lower and upper range of the given distance are calculated based on provided distance, distance uncertainty and z value equal to 3.290527 for 99.9% certainty in class interval $(-z\sigma, z\sigma)$. The equation $z = \frac{x-\mu}{\sigma}$ is used to calculate these ranges.

3. Finding the raster dimension:

   - Two circles C_1 and C_2 are drawn taking lower range and upper range of distance as radius respectively as shown in Figure 3.7. The figure is drawn considering the example of "22km SW of Belem [0127/4829]" where A is given point and B is target point. *ST_buffer()* function is used to draw these circles.

- After that the union of these two circles is obtained using *ST_Union()* function.
- A function is developed to find the co-ordinates of point C' in direction AC such that $AC' = 2 * AC$. Similarly, the co-ordinates of point D' in direction AD is obtained where $AD' = 2 * AD$ using same function. Direction of AC and AD are determined by the textual description as mentioned in Section 3.2.5.
- A function is developed to create a polygon $AC'D'$.
- Intersection area between the polygon and the union of two circle created above is obtained using *ST_intersection()* function.

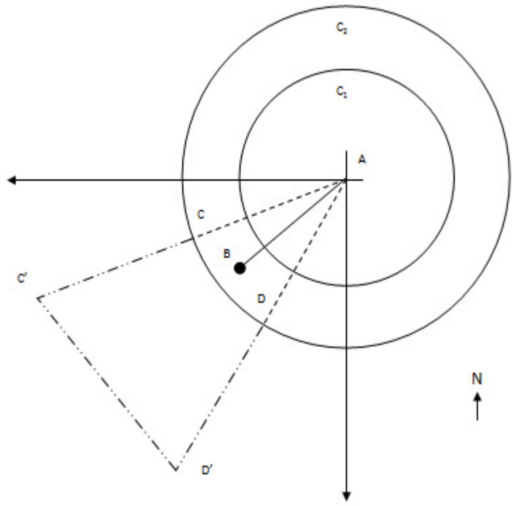


Figure 3.7: Process followed to obtain raster dimension, second template

- A bounding box of the intersected area is calculated using *ST_Envelope()* function which provides the desired raster dimension.

4. Function to create the raster:
   Finally, a function is developed to create the raster with fixed grid and proper SRID which covers the bounding box. To create the raster its origin obtained by *ST_XMin()* and *ST_YMax()* function available for bounding box. Similarly width of the raster obtained using *ST_XMax()* and *ST_XMin()* functions and height by *ST_YMax()* and *ST_YMin()*. The function developed is td_create_raster_second_template() and contains following parameters. The source code of this function is provided in Appendix B.

   - scalx: represents width of the raster cell.
   - scaly: represents height of the raster cell.
   - point: is the given point object.
   - distance: is the given distance.
   - dist_sigma: indicates uncertainty in distance provided.
   - direction_mean: indicates the mean direction of the target object from the reference object and depends on the textual description.

- compass_uncertainty: represents uncertainty in mean direction i.e., how much deviation on either direction from the mean. It also depends on textual descriptions and we define how it can be determined in Section 3.2.5.

5. Required PDF and uncertainty value calculation:
   Finding the required probability density function and calculation of uncertainty value are next requirements and they are defined in the following sections.

After the creation of the raster with desired dimension, the most important thing is to generate uncertainty values. In this template, two types of uncertainty are associated with textual descriptions, namely directional and distance. There is need of a bivariate probability density function (*PDF*) which captures both of these uncertainties at once. Generally, such *PDF* will have two input parameters in the distance and directional angle, along with two more subordinate input parameters associated with measure of spread of the two first input parameters.

Thus, there is requirement to find the solution, how to combine two separate and independent univariate density functions into a single bivariate PDF, which can include all four parameters as input. In the project, to solve this problem a *von Mises distribtion* is taken as a univariate PDF for directional uncertainty, and *univariate normal distribution* is selected to tackle with distance uncertainty. These density functions are finally combined as a single bivariate PDF which can address both uncertainty. These two density functions and the combined bivariate PDF are described below.

von Mises distribution: *von Mises distribution* is considered as the most useful distribution to describe directional or circular data. The probability density function of von Mises distribution is defined as:

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa cos(\theta - \mu)} \quad , \tag{3.4}$$

where θ is the directional variable, μ is the mean direction and κ is the concentration parameter. $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero of κ. The Bessel functions are solutions of a second-order differential equation. According to Evans et al. (2000), the von Mises distribution can be described as a circular analogue of the normal distribution on the line. Moreover, the distribution is unimodal and symmetrical about the mean direction. The larger the value of κ, the greater the clustering around μ and the distribution approaches a normal distribution, with a small the value of κ, the distribution approaches to uniform.

To calculate the von Mises density, it is obvious that there is need to find the appropriate value of κ. In circular distribution, κ i.e., the concentration parameter is related to the directional uncertainty. Based on the directional uncertainty the value of κ varies. More directional uncertainty less the value of κ, and vice versa. As such, to find the value of κ there is need to find the directional uncertainty first. According to Wieczorek et al. (2004), if a locality description contains any direction more specific than the cardinal directions(for example, 'NE'), then 'NE' could mean any direction between ENE and NNE. The directional uncertainty in these cases is $22.5°$ in either direction from the given mean heading. In this way, it is possible to know about a directional uncertainty from the given textual description. Evans et al. (2000) defined the relation between directional uncertainty or circular variance and concentration parameter. They provide the following equation.

$$\sigma^2 = 1 - \frac{I_1(\kappa)}{I_0(\kappa)} \quad , \tag{3.5}$$

where σ is directional uncertainty, I_1 and I_0 are modified Bessel function of first and zero order, respectively. Thus, once directional uncertainty is known it is possible to estimate the respective κ value from the above relation. However, it is also clear that there is need to calculate modified Bessel functions of order first and zero. To calculate these functions, two separate PostgreSQL/PostGIS functions were developed. Finally, a function was developed which uses above two functions together to estimate κ. The function is able to estimate κ value to an accuracy of five decimal places for the given directional uncertainty.

Functions to estimate modified Bessel of order one and zero follow the polynomial approximation defined in (Abramowitz and Stegun, 1964). Since Bessel function shows two different distribution in the ranges $-3.75 \leq x \leq 3.75$ and $3.75 \leq x < \infty$, there are two different error (ϵ) margins for these ranges while using this approximation technique. According to Abramowitz and Stegun (1964) calculating the Bessel value of order zero, within first range the $|\epsilon| < 1.6 \times 10^{-7}$ whereas in second range $|\epsilon| < 1.9 \times 10^{-7}$. Similarly, calculating the Bessel value of order one $|\epsilon| < 8 \times 10^{-9}$ and $|\epsilon| < 2.2 \times 10^{-7}$ in the first and second range respectively.

Univariate normal distribution: The univariate normal distribution is chosen here to handle distance uncertainty. Equation 2.2, as defined in Section 2.2.1 represents univariate density function.

Combined bivariate probability density: In the current scenario, there are two probability density functions for directional and distance uncertainty. There is need of a single bivariate probability density function to address both of these directional and distance uncertainty. Liu et al. (2009) defined a *refinement* operation to find the resultant uncertainty field from the multiple uncertainty fields when individual uncertainty fields are not conflicting with each other. According to them, the resultant uncertainty field is obtained by the product of individual uncertainty fields which is normalized by the integral of their product. Based on the same concept the resultant PDF for the von Mises distribution and univariate normal distribution can be obtained as defined in Equation 3.6.

$$f(\theta, x) = \frac{\frac{1}{2\pi I_0(\kappa)} e^{\kappa cos(\theta - \mu_c)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu_d)^2}{2\sigma^2}}}{\int_0^\infty \int_0^{2\pi} \frac{1}{2\pi I_0(\kappa)} e^{\kappa cos(\theta - \mu_c)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu_d)^2}{2\sigma^2}} d\theta dx} \tag{3.6}$$

In fact, the denominator of the above function is the value used to normalize the probability density obtained by the product of the two probability densities such that the combined PDF generates exact probability values. Since the analytical solution of the denominator is unknown to us, we assume that the value of denominator is one. Then, the final required bivariate density function is as follows.

$$f(\theta, x) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa cos(\theta - \mu_c)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu_d)^2}{2\sigma^2}} \tag{3.7}$$

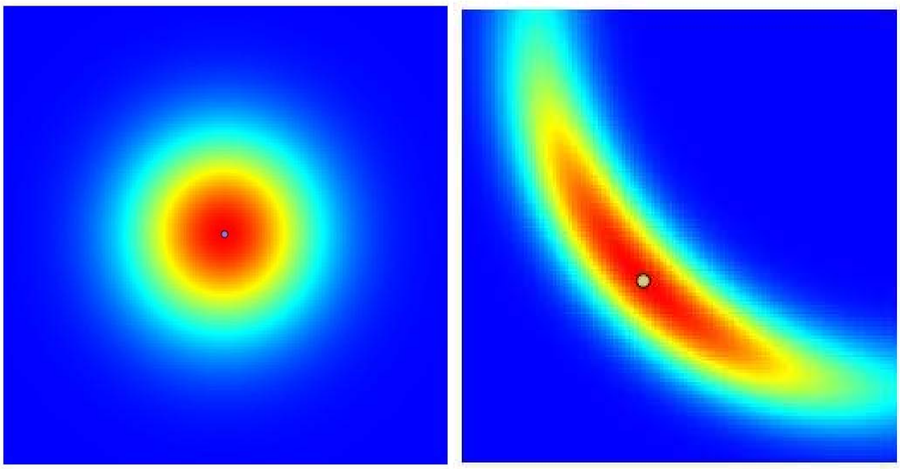
Now, the above PDF can be used to generate the uncertainty values for each cell using the five point concept defined in Section 3.2.4. The uncertainty raster obtained by inserting cell values is finally normalized such that the total sum of all cell values is one. In this PDF, μ_c and μ_d are directional mean and distance mean respectively. μ_c is obtained based on textual description and μ_d is considered zero. Concentration parameter κ is calculated based on directional uncertainty. Uncertainty in distance σ is provided. A directional variable θ is calculated at each five points of a cell by calculating the azimuth of a line connecting them with the reference point. Similarly, distance variate x is calculated at each of five points of a pixel and obtained by subtracting the distance between reference object and these points individually from the given distance.

3.3 CREATION OF THE TANDEM AND ITS STORAGE

As defined in Section 3.1, the tandem is the combination of a vector and raster object. Thus, a data type as tandem is required that can store the values of the vector and raster object together. Fortunately, PostgreSQL/PostGIS provides a composite data type which can combine both vector and raster data types in single structure and allow us to access both data types separately. It means there is facility to compute with vector and raster parts of a single tandem separately which helps in the process of overlay. In this way, a composite data type named “tandem” was created first which contains “geometry” and “raster” data types. Then a table is created with one of its fields as tandem data type. Finally using some SQL syntax, the raster created in Section 3.2 along with the point object are stored as tandem in the table. In fact, the tandem stores a vector object along with its positional uncertainty together.

3.4 RESULT AND DISCUSSION

The project data used for these two templates are from Brazil. Results obtained in both templates are shown below.



(a) Result of first template (b) Result of second template

Figure 3.8: Tandems

Result description, first template: Figure 3.8a belongs to first template where the textual description used is “Juiz de Fora [2145/4320(USBGN)]” and some provided information with it are as follows.

- *Juiz de Fora* is the name of the village, USBGN is the official body which provided coordinates.
- Given coordinate [2145/4320(USBGN)] indicates $43°20'$ West and $21°45'$ South.
- The locality is assumed to be correct to the minute level i.e., $1'$ or $\approx 1850m$ at the Equator.

The result shows the lowest uncertainty at the middle or core part i.e., near the location of point object. The uncertainty increases gradually along with increasing distance away from the point object. The distribution of uncertainty is symmetrical around the object. The result is a tandem representation i.e., combination of a vector and raster object. The procedure followed to obtain the raster dimension, which can cover 99.9% of certainty, and uncertainty value for the result are defined in Section 3.2.5. Table 3.1 shows details about the raster part of the tandem result.

Table 3.1 Raster detail, first template

ULX	ULY	Width	Height	ScaleX	ScaleY	SRID	Number of bands
653000	7613250	156	156	250	250	32723	1

Similarly, Table 3.2 shows details of uncertainty value generated in the raster before normalization of cell values.

Table 3.2 Uncertainty value detail, first template

Count	Sum	Mean	Stddev	Min	Max
24336	0.9991	4.1054e-005	7.0336e-005	1.3560e-009	0.0003228

Discussion, first template: The uncertainty value for every pixel in the result is the probability of the point object being inside that pixel i.e., more probability value denotes less uncertainty and vice versa. Mainly, the concept of joint probability with bivariate normal distribution function defined in Equation 3.1 is used to find the probability of a point object being inside a pixel. The probability is highest near the point object because the given point object is considered as mean of the distribution. This can also be observed from the same probability density function where the value of the density function is maximum when the value of variable x and y are close to μ_x and μ_y. Subsequently, the height of the pyramid from the xy plan is also high and that results in a volume under the pyramid that is also large. Consequently, a large volume indicates more probability.

Standard deviations are considered equal in both x and y directions, resulting the symmetrical distribution of uncertainty and locate the given point object around the middle of the raster. Locating the given point object at the middle allows to capture the whole symmetrical distribution.

Since locality is considered to be correct to the minute level, the value of standard deviation will be at least $1'$. In a similar way, it can be estimated that the required raster should have dimension greater than $1' \times 1'$. For this result, standard deviations are considered $3 * 1850m$ and the raster dimensions calculated based on the predefined confidence level that is 99.9%.

In Table 3.1 ULX and ULY represent coordinates of the origin of the raster. It is obtained as multiple of scalx or scaly where scalx and scaly are cell size in x and y direction respectively. Width and Height indicate column and rows of the raster. The dimension of the desired raster depends on the value of both standard deviations and cell size. If the standard deviation is large then the size of the raster is also large, and vice versa. Similarly, for same standard deviations the size of raster is larger when it has small cell size. In a same way, in Table 3.2 count indicates number of pixels on the raster, Sum indicates the total sum of all uncertainty value in the raster. This should be close to *one* because it is the probability of a point object being inside every pixel of the raster which can be considered as a sample space as defined in Section 2.2.1.

Result description, second template: Figure 3.8b belongs to the second template where the textual description handled is "22km SW of Belem[0127/4829(USBGN)]". The following information is provided along with this textual description.

- Uncertainty range in the distance $22km$ is $\pm 2km$.
- Based on the provided direction "SW", the mean compass direction from the north in clockwise direction is $\frac{5\pi}{4}$ rad.
- The uncertainty range in compass direction is considered $\frac{\pi}{8}$ rad as defined in (Wieczorek et al., 2004).

The result shows combined uncertainties for both directional and distance. The uncertainty value at each pixel is calculated using the density function defined in Equation 3.7. The point object is at given distance i.e., $22km$ apart from the given reference point "Belem" in "SW" direction i.e., mean direction. Uncertainty is lowest near the point object and increases gradually along with increasing distance and direction away from the position of the point object in both directions. The uncertainty distribution is symmetrical in respect of direction as well as distance. This is also a tandem data structure that represents a spatial object with its uncertainty in spatial position. The calculation of raster dimension and uncertainty value calculation methods have been defined in Section 3.2.5. Table 3.3 below shows details of the raster part of the tandem.

Table 3.3 Raster detail, second template

ULX	ULY	Width	Height	ScaleX	ScaleY	SRID	Number of bands
753500	9839750	106	106	250	250	32722	1

Similarly, Table 3.4 shows details of uncertainty value generated in the raster before normalization of the value.

Table 3.4 Uncertainty value detail, second template

Count	Sum	Mean	Stddev	Min	Max
11236	18319.8652	1.6305	2.3647	4.6748e-026	9.02235889434814

Discussion, second template: To handle the directional uncertainty von Mises distribution has been used. This is the distribution which is symmetrical about the mean direction. So in the result the directional uncertainty distribution is symmetrical to the direction mean $\frac{5\pi}{4}$ rad. While using Equation 3.7 to calculate density, a directional variable θ is calculated at each five points of a cell by calculating the azimuth of a line connecting them with the reference point. That shows the directional variation.

The distance uncertainty is handled by univariate normal distribution which is also symmetrical about the mean. Thus distance uncertainty is also symmetrical in the result. The mean distance is considered zero and the distance variable at each five points of every pixel is obtained by subtracting the distance between reference object and these points from the given distance.

In this case also, the probability value depends on the value of the density function defined in Equation 3.7. If the value of function is higher then more probability value, and vice versa. The value of the function depends on x and θ. If the value of x is less the value of univariate part in the function is high. In a same way when the value $\theta - \mu$ is small the value of cosine is high as such the value of directional part is also high. In this way, it is observed that the value of the function is high when there is less deviation in direction as well as distance. That implies near the point object more probability i.e., less uncertainty, and vice versa.

In Table 3.3 details of the raster including its origin, width, height, scale in both x and y direction, SRID and number of band information is provided. Similarly, in Table 3.4 the column Sum has value equal to 18319.8652, however it should be around one since the uncertainty value calculated here is also the probability of a point object being inside each pixel. The reason for such a large value is the consideration of one in the denominator of Equation 3.6. If the value of the denominator can be approximated then, this sum should be around one.

As compared to raster of the first template, it seems that the raster of second template has smaller dimension and it almost depends upon the uncertainty range in direction and distance uncertainty provided. Suppose, for the same case if distance is greater than $22km$ with same distance uncertainty range then raster will be comparatively larger. If the direction uncertainty is smaller than $\frac{\pi}{8}$ rad for the same given distance and distance uncertainty, the raster dimension will be smaller. These things finally affect the locality area.

As defined above, Figure 3.8b shows result obtained with directional uncertainty $\frac{\pi}{8}$ rad. To analyze the variation of distribution in second template let us consider directional uncertainty equal to $\frac{\pi}{32}$ rad smaller than the previous one for the same textual description. As defined in Section 3.2.5, both concentration parameter κ and directional uncertainty are related with each other. Moreover , the section defined that the function is developed to approximate the value of κ for respective directional uncertainty. Values of κ obtained for these two directional uncertainty using the same function are shown in Table 3.5 below:

Table 3.5 Concentration parameter value

Directional uncertainty in radians	κ
$\frac{\pi}{8}$	3.5923957824707
$\frac{\pi}{32}$	52.1240234375

The result obtained for second template when directional uncertainty is $\frac{\pi}{32}$ rad and corresponding value of $\kappa = 52.1240234375$ is as follows.

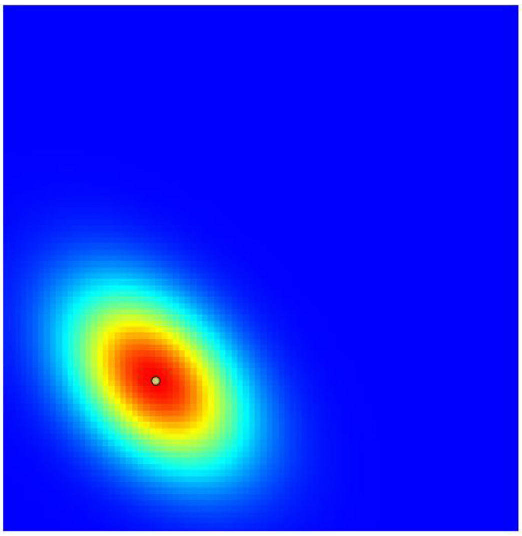


Figure 3.9: Small directional uncertainty, second template

The raster dimension and uncertainty value detail are shown in Table 3.6 and Table 3.7 respectively.

Table 3.6 Raster detail with small directional uncertainty, second template

ULX	ULY	Width	Height	ScaleX	ScaleY	SRID	Number of bands
758000	9839750	88	88	250	250	32722	1

Table 3.7 Uncertainty value detail with small directional uncertainty, second template

Count	Sum	Mean	Stddev	Min	Max
7744	21730.6617	2.8062	6.5446	3.1564e-029	35.7086

While comparing Figure 3.8b and Figure 3.9, it is clear that the distribution is more clustered around directional mean in second result than first one. It verifies that the greater the value of κ, the distribution is more clustered and vice versa. Similarly, the raster dimension is smaller in second result as compared to the first.

Chapter 4

Overlay between tandems

An overview about overlay operation between tandems is described in Section 4.1. The complete procedure followed to implement the overlay operation is described in Section 4.2. The resultant tandem obtained as a result of overlay operation is discussed in Section 4.3.

4.1 OVERLAY OPERATION

In Section 3.3, the creation of the tandem and its storage was defined. A single locality can be addressed by more than one textual description because multiple descriptions can better approximate the target object with reduced uncertainty. The more textual description more tandems we have. To find the target object in such case there is need to do overlay between tandems representing the same locality. The following two textual descriptions are taken as an example to perform the overlay operation between tandems, where both phrases describe the same locality.

- 22km SW of "Belem" [48° 29′W , 1° 27′S].
- 30km NE of "Abaetetuba" [48° 52′W , 1° 43′S].

The uncertainty range in distance for both of the above textual descriptions is $\pm 2km$. About directional uncertainty, in both cases it is $\frac{\pi}{8}$ rad. The implementation part is defined in Section 4.2 below.

4.2 IMPLEMENTATION OF OVERLAY OPERATIONS

Performing overlay between tandems or more specifically saying raster overlay indicates algebraic operation between two sets formed by uncertainty values of two rasters. In this way set values are in the unit interval $[0, 1]$. After implementation of overlay, the resultant raster should also contain cell values in the range $[0, 1]$. If this procedure is analyzed carefully, it is observed that there is binary operation between two sets having values in the range $[0, 1]$ and converted into a single set $[0, 1]$. This process is generally called Triangular norm or in short T-norm (Klement et al., 2004). In the project, *Product T-norm* has been used for overlay operation. The main overlay operation used is *intersection*. Triangular norm is briefly described below along with its basic types and they can be used for raster intersection in the project.

Triangular norm

Menger (1942) introduced first the notion of a space, in which distances are determined by probability distribution functions rather than by real numbers where he employed the function T from I^2, the closed unit square, to I, the closed unit interval, which he called a *triangular norm*. According to Klement et al. (2004), a mapping

$$T : [0, 1] \times [0, 1] \rightarrow [0, 1]$$

is a triangular norm if it is non-decreasing in each argument and satisfies the following conditions for all a, b, c, d in I.

- One identity: $T(a, 1) = a$
- Symmetry: $T(a, b) = T(b, a)$
- Monotonicity: $T(a, b) \leq T(c, d)$ whenever $a \leq c$ and $b \leq d$
- Associativity: $T(T(a, b), c) = T(a, T(b, c))$

Two major types of T-norms are “Product T-norm” and “Hamacher product T-norm”. They can be used for raster overlay in intersection process and defined as follows.

- Product T-norm:
  It is the ordinary product of real numbers and can be defined as $T_p(a, b) = a \cdot b$
- Hamacher product T-norm can be defined as:

$$T_{H_0}(a, b) = \begin{cases} 0, & \text{if } \ a = b = 0 \\ \frac{ab}{a+b-ab}, & \text{otherwise} \end{cases}$$

Both T-norms mentioned above belong to the family of “Hamacher T-Norm.” The Hamacher T-Norm is defined as follows.

$$H_\gamma(a, b) = \frac{ab}{\gamma + (1 - \gamma)(a + b - ab)} \tag{4.1}$$

where, $0 \leq \gamma \leq +\infty$. In Equation 4.1, when $\gamma = 0$ it will become “Hamacher product” and when $\gamma = 1$ it becomes “Product T-norm.” Hamacher T-norms are the only T-norms which are rational functions.

The raster overlay operation is defined as follows:

Implementation procedure

Overlay operation is performed separately on the vector and raster parts of tandems. For the separation of these two data fields from a single tandem, two different functions were developed. To implement the raster overlay operation, a function named “ST_MapAlgebraExpr()” is used. The function can take two rasters as argument and performs the desired overlay operation. In the current overlay operation, the “Intersection” process is done where the resultant raster cell contains the value obtained by performing product T-norm. During the overlay, if one of the rasters has a missing pixel value then the resultant raster takes the value of the other raster for that particular common pixel. Moreover, if both rasters have a missing pixel value then the resultant raster will contain a constant *no data value* for that particular pixel. The resultant raster obtained after raster overlay has the same dimension as the intersected part of rasters and it has resolution as the raster which is used as first argument in the function. In our case both rasters have same resolution so it does not affects which one be first argument in the function. The Detail information about the function can be find in the link (`http://postgis.refractions.net/documentation/manual-svn/RT_ST_MapAlgebraExpr2.html`) for PostGIS Raster documentation manual.

For the point object, the resultant point is obtained by using a function *ST_Collect()* and then *ST_Centroid()* subsequently. Both of these functions are buil-in functions in PostGIS. Finally, individual results of raster overlay and vector overlay are stored as tandem again. The overall process followed in the project starting from the creation of an uncertainty raster, tandem and finally overlay operation is shown in the figure below.

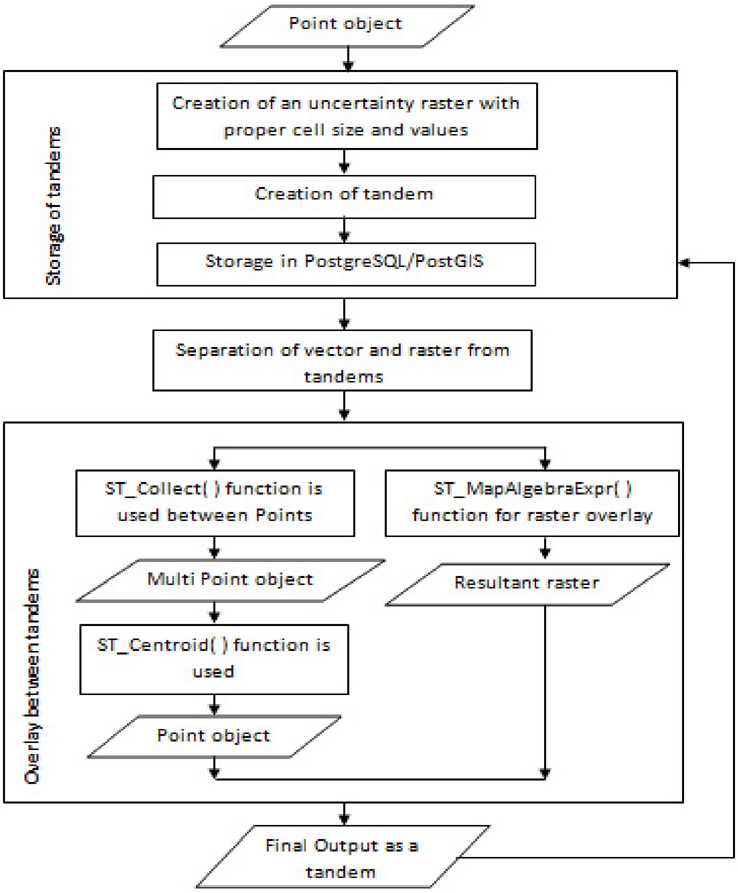
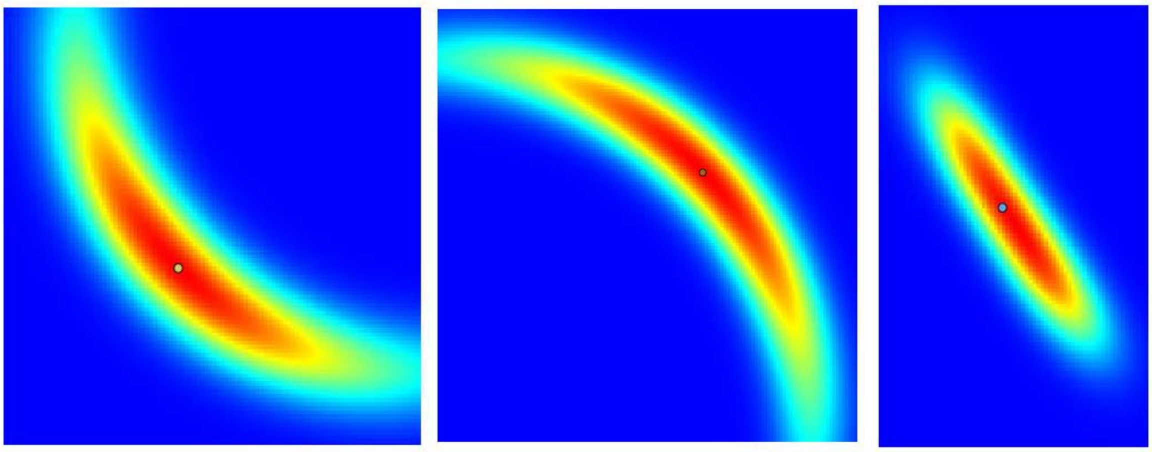


Figure 4.1: Overall procedure followed

4.3 RESULTS AND DISCUSSION

Tandems obtained by following the procedure mentioned in Section 3.2.5 for the textual descriptions currently selected in the overlay operation are shown below along with the result of overlay operation where these two tandems are used as input.



(a) 22km SW of "Belem" (b) 30km NE of "Abaetetuba" (c) Resultant tandem

Figure 4.2: Tandems used for overlay and resultant tandem

Tandems shown in Figure 4.2a and Figure 4.2b are used in overlay operation. Both of these tandems belong to "second template" defined in Section 3.2.5, where both tandems represent directional and distance uncertainty together. The uncertainty values are calculated using the density function defined in Equation 3.7, and the same five point concept. Tandem shown in Figure 4.2c is the resultant tandem obtained after overlay operation. The first row, second row and third row in Table 4.1 below show the raster detail for the first tandem, second tandem and resultant tandem. Similarly, in Table 4.2 first, second and third rows provide uncertainty value detail for the raster in first tandem, second tandem and resultant tandem respectively. In case of first and second tandems, uncertainty value details provided are after normalization, whereas for the resultant raster, uncertainty value detail provided is directly generated after overlay operation.

Table 4.1 Raster detail in tandems

S.N.	ULX	ULY	Width	Height	ScaleX	ScaleY	SRID	Number of bands
1	753500	9839750	106	106	250	250	32722	1
2	737250	9844000	135	135	250	250	32722	1
3	753500	9839750	70	106	250	250	32722	1

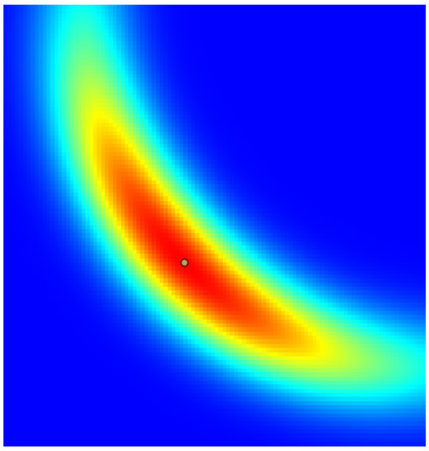
Table 4.2 Uncertainty value detail in tandems

S.N.	Count	Sum	Mean	Stddev	Min	Max
1	11236	0.9999	8.8999e-005	0.00013	2.5518e-030	0.00049
2	18225	1.0	5.4869e-005	9.05287e-005	0	0.00036
3	7420	0.000147	1.9757e-008	3.6844e-008	5.7599e-030	1.5908e-007

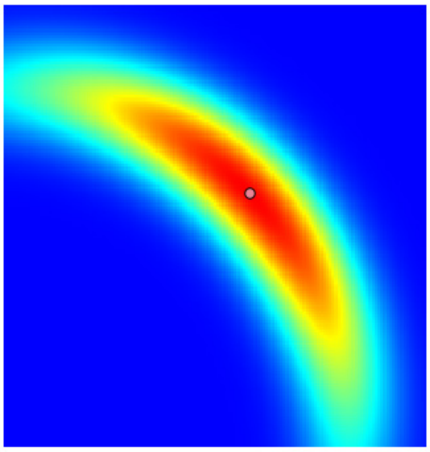
Raster dimensions shown in Table 4.1 show that the resultant tandem has smallest dimension, which is actually the intersected part of both rasters. The intersected part shows the more certain part of the locality because it is common to both tandems. The uncertainty value is obtained as Product T-norm i.e., algebraic product of uncertainty values of both rasters. During the overlay of two rasters a common pixel between two rasters has two values associates with these two rasters. These two values show that to what extent the pixel belongs to two rasters. In fact, uncertainty values associated with these two rasters form two fuzzy sets and overlap indicates a fuzzy intersection process. According to Klement et al. (2004), Product T-norm can be used for the fuzzy intersection. Similarly, algebraic product, currently denoted as product T-norm, is defined as possible candidate for fuzzy intersection process (Zadeh, 1965) and shows that $AB \subset A \cap B$ where A and B are two fuzzy sets.

The selection of union operation instead of intersection in tandem overlay can be another possibility. That requires T-conorms (e.g. probabilistic sum) in place of T-norms. However, in this case the resultant tandem will capture the combined dimension of rasters that leads to broad distribution of uncertainty value and increases uncertainty.

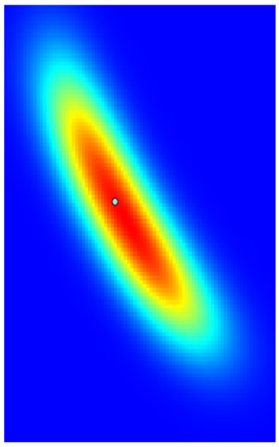
Overlay analysis with different distance uncertainty: Let us take same textual descriptions and perform the overlay operation with 10% distance uncertainty in both cases. This consideration results distance uncertainty $\pm 2.2km$ for the first textual description and $\pm 3km$ for the second. Tandems along the result obtained in this case are shown in Figure 4.3 below.



(a) 22km SW of "Belem"



(b) 30km NE of "Abaetetuba"



(c) Resultant tandem

Figure 4.3: Overlay with another case

Raster dimension details shown in Table 4.3 below.

Table 4.3 Overlay raster detail another case

S.N.	ULX	ULY	Width	Height	ScaleX	ScaleY	SRID	Number of bands
1	753000	9839750	108	108	250	250	32722	1
2	737250	9847000	147	147	250	250	32722	1
3	753000	9839750	84	108	250	250	32722	1

While comparing raster dimension between Table 4.1 and Table 4.3, it observed that dimension increased with increasing distance uncertainty. In this case, the resultant tandem shown in Figure 4.3c is slightly banded in a similar form of Figure 4.3a, whereas in the previous result shown in Figure 4.2c is not banded at all. It is due to equal distance uncertainty in previous case, whereas in this case, unequal and less in the textual description related to Figure 4.3a.

4.4 FURTHER WORK

In addition of first and second template described in Section 3.2.5, a third template was also discussed in the project. However, due to time constraint it could not be completed in the specified time. An overview about a textual description of this template is described below along with some thoughts towards its tandem representation which may help in future work.

4.4.1 Third template

In this template, the textual description is in the form "20km from P along L", where 'P' represents a reference point object and 'L' represents linear feature like highway, river etc. The geometry of linear feature and reference point object are provided. Uncertainty in distance can be approximated similar to second template .

Since the target point object is along the linear feature, a spatial spread of the target object should be around the linear feature and in the similar pattern of it. Such spread indicates the orthogonal spatial distribution of the point object to each line segment. The probable segment of the linear feature, around which the target object lies, can be estimated by finding the upper range and lower range of distance uncertainty. The raster dimension can be approximated by distance uncertainty and the spread on either side of the linear feature.

The segmented part of linear feature estimated by distance uncertainty may contains various line segments. By finding the required probability density function (PDF) for each line segment and summing up them provides the resultant PDF that can be used to find the distribution of the point object along linear feature.

In this way, for each line segment there will be two PDFs, the one associated with distance uncertainty and another with orthogonal distribution in either side of the object. These two PDFs together provide the distribution of the point around the line segment. Now, the PDF defined in Equation 2.2 can be used for distance uncertainty. The orthogonal distribution of the point towards the line segment is same as the distribution of the point in any arbitrary direction. According to Shahin (1997), the distribution of a point object in any arbitrary direction can be obtained as marginal probability density of the bivariate distribution of the point. Moreover, the marginal probability density in the direction perpendicular to the line segment can be obtained by rotating the original XY coordinate to the $X'Y'$ by an angle θ such that the X' axis is parallel to the line segment. The relation between the rotation of axes and Gaussian bivariate distribution is defined in (Schwarzlander, 2011), where the marginal probability density along Y' is given by.

$$f(y') = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_y^2 cos^2\theta + \sigma_x^2 sin^2\theta}} exp\left[-\frac{1}{2}\frac{y'^2}{\sigma_y^2 cos^2\theta + \sigma_x^2 sin^2\theta}\right] \tag{4.2}$$

Now, the combined PDF of the perpendicular PDF defined in Equation 4.2 and the PDF defined in Equation 2.2 can be obtained as:

$$f(y,\theta) = \frac{\frac{1}{\sqrt{2\pi}\sqrt{\sigma_y^2 cos^2\theta+\sigma_x^2 sin^2\theta}} exp\left[-\frac{1}{2}\frac{y^2}{\sigma_y^2 cos^2\theta+\sigma_x^2 sin^2\theta}\right]\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-\mu)^2}{2\sigma^2}}}{\int_0^{\infty}\int_0^{2\pi}\frac{1}{\sqrt{2\pi}\sqrt{\sigma_y^2 cos^2\theta+\sigma_x^2 sin^2\theta}} exp\left[-\frac{1}{2}\frac{y^2}{\sigma_y^2 cos^2\theta+\sigma_x^2 sin^2\theta}\right]\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-\mu)^2}{2\sigma^2}}d\theta dy} \tag{4.3}$$

Since the analytical solution is difficult for the denominator of Equation 4.3, let us consider again 1 for the denominator then the required PDF will be as follows.

$$f(y,\theta) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_y^2 cos^2\theta + \sigma_x^2 sin^2\theta}} exp\left[-\frac{1}{2}\frac{y^2}{\sigma_y^2 cos^2\theta + \sigma_x^2 sin^2\theta}\right] \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-\mu)^2}{2\sigma^2}} \tag{4.4}$$

Thus, using Equation 4.4 the distribution of the point object around line segments can be obtained.

Chapter 5

Conclusion and Recommendations

A brief overview of the project is defined in Section 5.1. Research questions and their answers are given in Section 5.2. Finally, some limitations and recommendation are presented in Section 5.3.

5.1 CONCLUSION

The main objective of the project is "to develop new functionality to represent uncertainty associated with objects having uncertain spatial extent using tiny rasters in spatial databases and finding their possible spatial position". To achieve this objective the tandem approach is used where tandem is the combination of vector and raster objects. Tandem is a novel data type where the vector object considered as a crisp part of a spatial object and it characterize the geometry of the object as point, line or polygon etc. Raster provides supplementary information about the vector object and describes possible spatial position of the object using probability theory. Thus, the combination of both vector and raster as tandem data type provides the comprehensible picture of spatial objects. Different PostgreSQL/PostGIS Functions are developed, which can create an uncertainty raster by generating uncertainty value, based on types of textual descriptions. Finally, the overlay operation is performed between tandems to approximate the spatial position of uncertain objects.

5.2 ANSWERS TO RESEARCH QUESTIONS

To achieve the objective, five research questions were formulated. All these questions along with their answers are described below.

1. What is/are appropriate model/s of the spatial uncertainty?

   Major uncertainty techniques reviewed are probability theory, random sets and fuzzy theory. As defined in Section 2.5, mainly the uncertainty associated with point objects is handled in this project and all modelling techniques are applicable to model the extensional uncertainty. However, probability theory is used with the concept of joint probability by calculating the probability of a point object being inside each cell. As defined in Section 2.2.2, a random point can be described as a random set in R^2 that contains a finite collection of uncertain units with positive probability values. Furthermore, the probability value with each uncertain unit is the likelihood that the point is in that location and obtained by a covering function. This implies that if random sets would be used instead of probability theory, the probability value associated with each random unit can be obtained using the covering function. Similarly, a vague point is a site with known location $P(x, y)$ where it has different membership value that indicates the degree of belonging of the site with phenomena of interest by Dilo et al. (2007). These membership values are calculated based on membership functions. So, using the concept of membership function the uncertainty value can be generated to define the uncertain point objects while using fuzzy theory.

In gist, it is noticed that the uncertainty value calculation methods may differ while using these different modelling techniques.

2. What precisely does a raster cell value mean in such formed raster?

   The raster cell value represents the probability of a point object being inside each cell. These values are calculated based on the concept of joint probability theory. Since calculating such probability is not possible by analytical procedure so the approximation technique is used where this uncertainty value represents the volume under a pyramid formed by the four corner points and one mid-point of a pixel.

3. How can combined data structures of tandem help to represent that uncertainty?

   Spatial objects with uncertain spatial extent are represented in the project with the help of the tandem data type. The uncertain spatial extent shows uncertainties like positional uncertainty or extensional uncertainty. As defined earlier the tandem data model is the combination of vector plus raster. In this way, while representing uncertainty using the tandem data model the vector object deemed as the crisp part of the object where it preserves its geometrical characteristics. The raster defines possible spatial position of the object with gradual changes of uncertainty and provide additional information about the vector object regarding its position. Presence of both vector and raster data type in the tandem provide comprehensible and complete representation of the object with its uncertainty.

4. How can PostgreSQL/PostGIS store raster/vector in combination as tandem?

   PostgreSQL/PostGIS provides a mechanism of composite data type which allow us to formulate a tandem data type as combination of geometry and raster data type in the database. So, a tandem data type as a composite data type has been developed to store a tandem value.

5. How do two uncertain objects determine their uncertain area of overlap?

   The uncertainty in the spatial position of a locality can be estimated by performing intersection operation between two tandems using the product T-norm. Intersection between tandems or more specifically rasters provides the common region and the uncertainty value in the common region is calculated by product T-norm. Product T-norm is like a algebraic product and mostly used in fuzzy intersection.

5.3 RECOMMENDATIONS

Following recommendations are purposed for the project:

- This is the project started as an initiation to represent uncertainty using tiny rasters in spatial databases especially PostgreSQl/PostGIS. Accordingly, simple considerations are taken first for cell size and resolution. The raster cell size has been considered equal in all tandems, fixed grid along with equal cell height and cell width. However, this simple consideration may not be always applicable. Uncertainty rasters for respective tandems should be created with suitable cell size and grid system where these size may not be equal. Finally, that will lead to experimentation of the overlay operation between two tandems with different resolutions.

- In the second template the positional uncertainty of the reference object i.e., "Belem" is not considered and it is assumed as true position. It would be better to consider the uncertainty of the reference object for further approximation on finding the locality of target objects. The probability distribution of reference objects should be calculated first by considering the associated uncertainties to it and then the distribution of the target objects can be found based on the distribution of reference objects.

- In the same way, due to lack of analytical solution the denominator of Equation 3.6 is considered 1. It was a easy way adopted to solve that part because whatever be its value final distribution will be in same ratio for all variates. However, such consideration is not a precise way because this affects the total sum of uncertainty value and needs one more extra step for normalization. Thus, it should be approximated.

- The visualization of tandem data type is one major problem. Although the PostGIS raster can be viewed in QGIS, it is unable to recognize the tandem data type and keep it in no geometry data type. For the temporary visualization of all results as combination of vector plus raster i.e., tandem individual layers have been made in their respective spatial reference system and shown together in QGIS. The visualization of results as tandem data type is essential for uncertainty realization which is one of the important aspects of the project. Therefore it is required to construct a data structure for the visualization of tandem data type such that it is recognized by current GIS softwares.

- Currently, in the project only the uncertainty associated with point objects is represented in tiny rasters. The vector part of the tandem model may be point, line or polygon feature as shown in its conceptual model. Accordingly, representing uncertainty associated with line and polygon features requires new methods and techniques. As an uncertainty modelling technique probability theory can be used again for line and polygon feature. However, random sets can be better alternative for these features.

Bibliography

Abramowitz, M. and Stegun, I. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. Dover publications.

Altman, D. (1994). Fuzzy set-theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems*, 8(3):271–289.

Cova, T. J. and Goodchild, M. F. (2002). Extending geographical representation to include fields of spatial objects. *International Journal of Geographical Information Science*, 16:509–532.

de Paulo, M. and Vinhas, L. (2011). Postgis raster plugin for quantum GIS. *Proceeding of the XII Brazilian Symposium on GeoInformatics, Campos do Jordão,SP,Brazil*.

Dilo, A., de By, R. A., and Stein, A. (2007). A system of types and operators for handling vague spatial objects. *International Journal of Geographical Information Science*, 21:397–426.

Evans, M., Hastings, N., and Peacock, B. (2000). *Statistical distributions*. John Wiley & Sons.

Fisher, P. (1999). Models of uncertainty in spatial data. *Geographical Information systems*, 1(Principles and technical):192–205.

Fisher, P., Comber, A., and Wadsworth, R. (2006). Approaches to uncertainty in spatial data. *Fundamentals of Spatial Data Quality*, pages 43–59.

Glemeser, M. and Fritsch, D. (1998). Data uncertainty in a hybrid GIS. *International Archive of Photogrammetry and Remote sensing, Stuttgart, Germany*.

Goodchild, M. F. (1989). Modeling error in objects and fields. *The accuracy of spatial databases*, pages 107–113.

Goodchild, M. F. (1992). Geographical data modeling. *Computers & Geosciences*, 18(4):401–408.

Goodchild, M. F. and Hill, L. L. (2008). Introduction to digital gazetteer research. *International Journal of Geographical Information Science*, 22(10):1039–1044.

Goodchild, M. F., Yuan, M., and Cova, T. J. (2007). Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science*, 21(3):239–260.

Guo, Q., Liu, Y., and Wieczorek, J. (2008). Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *International Journal of Geographical Information Science*, 22(10):1067–1090.

Hwang, S. and Thill, J.-C. (2005). *Modeling Localities with Fuzzy Sets and GIS Fuzzy Modeling with Spatial Information for Geographic Problems*, pages 71–104. Springer Berlin Heidelberg.

Jones, C. B. and Purves, R. S. (2008). Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3):219–228.

Kjenstad, K. (2006). On the integration of object-based models and field-based models in GIS. *International Journal of Geographical Information Science*, 20(5):491–509.

Klement, E., Mesiar, R., and Pap, E. (2004). Triangular norms. position paper I: basic analytical and algebraic properties. *Fuzzy Sets and Systems*, 143(1):5–26.

Knisley, J. (2001). Applications of the double integral. http://math.etsu.edu/multicalc/prealpha/Chap4/Chap4-3/printversion.pdf, (07/02/2012).

Liu, Y., Guo, Q. H., Wieczorek, J., and Goodchild, M. F. (2009). Positioning localities based on spatial assertions. *International Journal of Geographical Information Science*, 23(11):1471–1501.

Menger, K. (1942). Statistical metrics. *Proceedings of the National Academy of Sciences of the United States of America*, 28(12):535.

Molenaar, M. (2000). Three conceptual uncertainty levels for spatial objects. *In: ISPRS 2000 congress : geoinformation for all : Amsterdam, the Netherlands, 16-23 July, 2000. pp. 670-677.*

Nguyen, H. (2006). *An introduction to random sets.* Chapman & Hall/CRC, Boca Raton.

Nuñez-Garcia, J. and Wolkenhauer, O. (2002). Random set system identification. *Fuzzy Systems, IEEE Transactions on*, 10(3):287–296.

Obe, R. and Hsu, L. (2011). *PostGIS in action*. Manning, Stamford.

Rowe, R. J. (2005). Elevational gradient analyses and the use of historical museum specimens: A cautionary tale. *Journal of Biogeography*, 32:1883–1897.

Schneider, M. (1999). *Uncertainty management for spatial data in databases: Fuzzy spatial data types*, volume 1651 of *Lecture Notes in Computer Science*, pages 330–351. Berlin:Springer-Verlag.

Schneider, M. (2003). Design and implementation of finite resolution crisp and fuzzy spatial objects. *Data & Knowledge Engineering*, 44(1):81–108.

Schneider, M. (2008). *Fuzzy Spatial Data Types for Spatial Uncertainty Management in Databases.*, volume 2, pages 490–515. In Galindo, J. (Ed.).

Schwarzlander, H. (2011). *Probability concepts and theory for engineers*. John Wiley & Sons.

Shahin, F. (1997). Modeling positional uncertainty of linear features in geographic information systems. *An-Najah Univ. J. Res., Nat. Sci*, 11:23–38.

Shi, W. (1998). A generic statistical approach for modelling error of geometric features in GIS. *International Journal of Geographical Information Science*, 12:131–143.

Voudouris, V. (2010). Towards a unifying formalisation of geographic representation: the object-field model with uncertainty and semantics. *International Journal of Geographical Information Science*, 24(12):1811–1828.

Wieczorek, J., Guo, Q. G., and Hijmans, R. J. (2004). The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18(8):745–767.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.

Zhan, F. B. and Lin, H. (2003). Overlay of two simple polygons with indeterminate boundaries. *Transactions in GIS*, 7:67–81.

Zhao, X., Chen, X., and Stein, A. (2009). Random sets model for spatial objects with uncertain boundaries. *In:Proceeding of the 12th AGILE international conference on geographic information science, Hanover, Germany*.

Zhao, X., Stein, A., and Chen, X. (2010). Application of random sets to model uncertainties of natural entities extracted from remote sensing images. *In: Stochastic Environmental Research and Risk Assessment, 24(2010)5, pp. 713-723*.

Appendix A

Source code for uncertainty raster creation function, first template

```
CREATE OR REPLACE FUNCTION deepak.
   td_create_raster_first_template(scalx double precision,
   scaly double precision, point geometry, sigmax double
   precision, sigmay double precision, rho double precision)
  RETURNS raster AS
$BODY$

/*
-This function takes 6 parameters where: scalx, scaly
   represents cell size width and height though in our case
   both are equal.
-point is given point object.
-sigmax and sigmay represent  standard deviation of a variable
    x and y respectively.
-rho represents correlation coefficient between x and y,.
*/
  DECLARE
  easting double precision;             --x-coordinates of the
      point in UTM projection.
  northing double precision;            --y-coordinates of the
      point in UTM projection.

  upperleftx double precision;
  upperlefty double precision;

  result raster;
  srid smallint;
  uncertainty_per_cell double precision;

  ULX double precision;
  ULY double precision;
  ULPD double precision;
  LLPD double precision;
  LRPD double precision;
  URPD double precision;
  MPPD double precision;
```

```
  z1 real:=-3.4807;          --These z1 and z2 are the values
     for which the cell values F(z1,z2)<0.00025 i.e. one
     fourth of error margin 0.1%. Here, F(-3.4807 3.4807)=
     0.00024999
  z2 real:=3.4807;

  i smallint;                   -- used to insert uncertainty
     values in the raster.
  j smallint;

  width_height smallint;

  BEGIN

 IF ST_SRID(point)=4326 THEN

--Finding SRID in UTM of the given point. Also calculating
   easting and northing of the point in the same system.
        srid:=deepak.td_utmzone_from_point(point);
        easting:= st_X(st_transform(point,srid));
        northing:= st_Y(st_transform(point,srid));
 ELSE
 RAISE EXCEPTION 'Given point argument is not in proper SRS
    4326.';
 END IF;

--calculate origin of the raster using values of z1,z2,meanx,
   meany,sigmax and sigmay as follows
upperleftx:= z1*sigmax+easting;                 --The given
   point coordinates values considered as mean.
upperlefty:= z2*sigmay+northing;

--Finding origin as multitude of cell size of the raster.
upperleftx:= deepak.td_upperleftx(upperleftx,scalx) ;
upperlefty:= deepak.td_upperlefty(upperlefty,scalx);
width_height:= ((deepak.td_upperleftx(easting,scalx)-
   upperleftx)/scalx)*2::smallint;  --to find the width and
   height of the desired raster.

--Now create raster with calculated origin and dimension which
    have the error margin of 99.9%. To became more certain +2
   added on both width and height.

result:= ST_MakeEmptyRaster(width_height+2,width_height+2,
   upperleftx,upperlefty,scalx,scaly,0.0,0.0,srid);
result:= ST_AddBand(result,'32BF');
```

```
FOR i IN 1..width_height+2
                            --to fetch each cells of the raster
   LOOP
        FOR j IN 1..width_height+2
        LOOP


        ULX:= ST_Raster2WorldCoordX(result,j,i);
                    --This function returns the x-coordinates
           of upperleft corner
        ULY:= ST_Raster2WorldCoordy(result,j,i);
                    --This function returns the y-coordinates
           of upperleft corner

        --calculate probability density at each corner and
           middle point of each cells

        --For upperleft
        ULPD:= deepak.td_probability_density(ULX,ULY,easting,
           northing,sigmax,sigmay,rho);

        -- For lowerleft
        LLPD:= deepak.td_probability_density(ULX,ULY-scalx,
           easting,northing,sigmax,sigmay,rho);

        --For lowerright
        LRPD:= deepak.td_probability_density(ULX+scalx,ULY-
           scalx,easting,northing,sigmax,sigmay,rho);

        --For upperright
        URPD:= deepak.td_probability_density(ULX+scalx,ULY,
           easting,northing,sigmax,sigmay,rho);

        --For middle point
        MPPD:= deepak.td_probability_density(ULX+(scalx/2.0),
           ULY-(scalx/2.0),easting,northing,sigmax,sigmay,rho)
           ;

        --Now calculating volume under each triangles of
           pyramid and summing up them, the formula used below
            is after simplification.

        uncertainty_per_cell:=1/6.0 *(scalx*scalx)*(ULPD+LLPD
           +2*MPPD+LRPD+URPD);

        result := ST_SetValue(result,j,i,uncertainty_per_cell)
           ;
```

```
        END LOOP;
END LOOP;


Return result;

END
$BODY$
  LANGUAGE plpgsql IMMUTABLE STRICT
```

Appendix B

Source code for uncertainty raster creation function, second template

```
CREATE OR REPLACE FUNCTION deepak.
   td_create_raster_second_template(scalx real, scaly real,
   point geometry, distance double precision, dist_sigma
   double precision, direction_mean double precision,
   compass_uncertainty double precision)
  RETURNS raster AS
$BODY$

/*
-- This function takes 7 parameters
-- scalex and scaley represent cell width and height though
   both are equal here.
-- point represents the given point object.
-- distance represents distnace between the reference object
   and target object.
-- dist_sigma is the distance unceraInty from the given
   distance. For example it is plus/minus 2 in given example.
-- Similarly compass_uncertainty represents uncertainty in
   compass direction.
*/
  DECLARE
  easting double precision;            --easting of the given
      point in UTM projection.
  northing double precision;           --northing of the
     given point in UTM projection.


  easting_farpoint double precision;     --easting of the far
     point.
  northing_farpoint double precision;    --northing of far
     point.

  result raster;

  srid_far smallint;
  srid_given smallint;
```

```
uncertainty_per_cell double precision;

ULX double precision;
ULY double precision;
ULPD double precision;
LLPD double precision;
LRPD double precision;
URPD double precision;
MPPD double precision;

i smallint;                   -- used to insert uncertainty
   values in the raster.
j smallint;

width smallint;               -- width and hieght of the
   raster.
height smallint;

dst double precision;         -- dst represents difference
   of distance of each pixels from the given distance i.e.,
   22 km
upper_range double precision;         --For (d+4sd)
lower_range double precision;         --For (d-4sd)

bearing double precision;
bounding_box geometry;
kappa double precision:=deepak.td_kappa_estimation(
   compass_uncertainty);

BEGIN

 IF ST_SRID(point)=4326 THEN

      --calculating SRID of the far point from the given
         point
      --The function deepak.td_farpoint_in_utm_template2(
         point,distance) return geometry of far point in UTM
         .

      easting_farpoint:= st_x(st_transform(deepak.
         td_farpoint_in_utm_template2(point,distance),4326))
         ;
      northing_farpoint:= st_y(st_transform(deepak.
         td_farpoint_in_utm_template2(point,distance),4326))
         ;

      srid_far:= deepak.td_utmzone_from_point(
         ST_GeomFromText('POINT('||easting_farpoint||' '||
```

```
            northing_farpoint||')',4326));

        --for Given point
        srid_given:=deepak.td_utmzone_from_point(point);
        easting:= st_X(st_transform(point,srid_given));
  northing:= st_Y(st_transform(point,srid_given));
   ELSE
   RAISE EXCEPTION 'Given point argument is not in proper SRS
      4326.';
   END IF;

--First calculate the distance (d-4sd) and (d+4sd) using z
   value equal to 3.290527 for the class interval(-z\sigma,z\
   sigma)

        upper_range:= dist_sigma*3.290527+distance;
                   -- Because z= x-mu/sigma
        lower_range:= distance-dist_sigma*3.290527;

-- Determine the bounding box

        bounding_box:= deepak.td_boundingbox(lower_range,
           upper_range,easting,northing,direction_mean,
           compass_uncertainty,srid_given);

--calculating the coordinates for the origin of the raster as
   (ULX,ULY).
        ULX:= ST_XMin(bounding_box);
        ULY:= ST_YMax(bounding_box);

-- Calculating the upperleftx and upperlefty coordinates as
   the multiple of scale.

   ULX:= deepak.td_upperleftx(ULX, scalx);
   ULY:= deepak.td_upperlefty(ULY, scalx);

-- Calculate the width and height of the desired raster which
   is also the bounding box for the above two circles.
        width:= round((ST_XMax(bounding_box)-ST_XMin(
           bounding_box))/scalx)::smallint;
        height:= round((ST_YMax(bounding_box)-ST_YMin(
           bounding_box))/scalx)::smallint;

--creating raster

        result:= ST_MakeEmptyRaster(width,height,ULX,ULY,scalx
           ,scaly,0.0,0.0,srid_far);
        result:= ST_AddBand(result,'32BF');
```

```
--Generating unceratinty value and inserting on respective
   cells.

      FOR i IN 1..width
      LOOP
                                          FOR j IN 1..height
                                                    LOOP

         --calculate probability density at each corner and
            middle point of each cells
                  ULX := ST_Raster2WorldCoordX(result,j,i);
                  ULY := ST_Raster2WorldCoordY(result,j,i);

                  dst := distance-(sqrt(POWER((ULX-easting),2) +
                       POWER((ULY-northing),2)));
                  bearing := ST_Azimuth(ST_MakePoint(easting,
                     northing), ST_MakePoint(ULX,ULY)); --
                     Returns angle in radian. Angle is computed
                     clockwise such that 12=0,3=pi()/2 etc.

--For upperleft corner of a cell
                  ULPD:= deepak.td_probability_density_circular(
                     dst,dist_sigma,bearing,direction_mean,kappa
                     );

-- For lowerleft corner of a cell
                  dst := distance-(sqrt(POWER((ULX-easting),2) +
                     POWER(((ULY-scalx)-northing),2)));
                  bearing := ST_Azimuth(ST_MakePoint(easting,
                     northing), ST_MakePoint(ULX,ULY-scalx));
                  LLPD:= deepak.td_probability_density_circular(
                     dst,dist_sigma,bearing,direction_mean,kappa
                     );

--For lowerright corner of a cell
                  dst := distance-(sqrt(POWER(((ULX+scalx)-
                     easting),2) +POWER(((ULY-scalx)-northing)
                     ,2)));
                  bearing := ST_Azimuth(ST_MakePoint(easting,
                     northing), ST_MakePoint(ULX+scalx,ULY-scalx
                     ));
                  LRPD:= deepak.td_probability_density_circular(
                     dst,dist_sigma,bearing,direction_mean,kappa
                     );

--For upperright corner of a cell
```

```
                dst := distance-(sqrt(POWER(((ULX+scalx)-
                   easting),2) +POWER((ULY-northing),2)));
                bearing := ST_Azimuth(ST_MakePoint(easting,
                   northing), ST_MakePoint(ULX+scalx,ULY));
                URPD:= deepak.td_probability_density_circular(
                   dst,dist_sigma,bearing,direction_mean,kappa
                   );

--For middle point of a cell
                dst := distance-(sqrt(POWER(((ULX+(scalx/2.0))
                   -easting),2) +POWER(((ULY-(scalx/2.0))-
                   northing),2)));
                bearing := ST_Azimuth(ST_MakePoint(easting,
                   northing), ST_MakePoint(ULX+(scalx/2.0),ULY
                   -(scalx/2.0)));
                MPPD:= deepak.td_probability_density_circular(
                   dst,dist_sigma,bearing,direction_mean,kappa
                   );

--Now calculating volume under each triangles of pyramid and
   summing up them, the formula used below is after
   simplification.
                uncertainty_per_cell:=1/6.0 *(scalx*scalx)*(
                   ULPD+LLPD+2*MPPD+LRPD+URPD);

                result := ST_SetValue(result,j,i,
                   uncertainty_per_cell);

        END LOOP;
    END LOOP;

    Return result;
  END
$BODY$
  LANGUAGE plpgsql VOLATILE
```