A Ph.D. research proposal on

A Knowledge-guided quantitative prediction for landslide hazard and risk --- case study in Darjeeling Himalaya, India

Applicant

Saibal Ghosh

Geologist, GSI, India

Promotor

Prof. Dr. Victor Jetten

Supervisors

Dr. Cees van Westen
Associate Professor

Dr. E.J.M. Carranza
Assistant Professor

Department of Earth System Analysis

ITC, The Netherlands

National Remote Sensing Agency

(Department of Space, Government of India)

ITC

Table of contents

List of tables

List of figures

Abstract

The medium scale (1:25000) landslide hazard prediction fundamentally depends on the understanding of interplay of relevant determining factors (both predisposing and triggering) and their selective combinations for causing different slide types. Such understanding of landsliding processes actually forms the basis of preparation of detailed landslide inventory (both spatial and temporal) and geofactor database required for analysing the spatio-temporal prediction of landslide hazard. Available literature review indicates that various GIS based quantitative landslide hazard prediction methods in vogue are often more focused on the tool than on the quality and selection of input data and frequently involves an extreme simplification of the landslide controlling factors. A reasoned selection of geofactor combination for each slide type and also to the slides of different periods can reduce the basic uncertainty of selection in both spatial as well as temporal landslide and geofactor database. Despite care and efforts, in landslide hazard database (landslide inventory and geofactor), a number of uncertainties and incompleteness of data are commonplace. These represent the inherent fuzziness and inevitable errors associated with any natural system. To model this natural fuzziness and to improve further the model accuracy and performance, a quantitative technique using both knowledge and data is envisaged. The prime objective of this research programme would be to develop a technique where knowledge guides the primary selection and mapping of geodata necessary for spatio-temporal prediction and followed by development of a suitable quantitative modelling technique where during process development stage, both knowledge base or priori and quantitative data treatment are interactively and iteratively involved to create a better prediction method.

The above landslide hazard analysis is always the fundamental step to attempt the quantitative risk analysis of any area and this can be accomplished by supplementing the quantitative hazard information with the information of consequence analysis (vulnerability and worth of risk elements). On medium scale, hazard information is mostly represented by a spatial database, whereas true quantitative risk analysis involves calculation of specific risks of each element at risk for each landslide magnitude/ scenario. Up scaling of this information of specific risk into the medium scale spatial hazard database is difficult because of scale constraints. Another major constraint of consequence analysis is vulnerability assessment, where quantitative estimation is difficult due to non-availability of past landslide damage data. Considering all the above constraints in mind, this research should also aim to attempt to develop a suitable methodology of semi-quantitative to quantitative risk calculation for a semi-urban/ urban mountainous environment in Darjeeling district, India using the above proposed knowledge-guided quantitative hazard prediction model and the available database of elements at risk.

1.0. Introduction

Landslides are major threats to life and property in the mountainous terrains around the world. Due to the growing urbanization and uncontrolled landuse of the limitedly-available mountainous areas, on global scale, there is an increasing trend of landslide hazard and associated risk. A recent global risk assessment study (Nadim et al., 2006) indicates that the regions with the highest risk of such danger can be found in Colombia, Tadjikistan, India, and Nepal, where the estimated number of people killed per year per 100 km^2 was found to be greater than one. Historical record indicates that the greatest number of loss of life due to a single landslide event was the earthquake-triggered loess landslide disaster in Kansu Province, China in 1920, where 100,000 people lost their life (Schuster and Fleming, 1986). One of the best known prominent landslide devastation of the last century was an earthquake-triggered (magnitude 7.5) debris avalanche in 1970 on the slopes of Mt. Huascaran, Peru, with an average speed of 320 km/hour burrying the towns of Yungay and Ranrahirca and killing more than 18000 people**.** Similar disastrous landslide in Europe has been the 1963 Vaiont reservoir slide in north-eastern Italy, resulting in death of 2000 people and a great economic loss of 126 million US dollars (Schuster and Fleming, 1986). According to a report estimate (Schuster, 1996), the annual direct and indirect cost due to landslide damages is in the order of 2-5 billion US dollars. At the global level, all the above disasters raised severe concern within the international community. Even as substantial scientific and material progress are made towards aiming a progress in reducing the loss of lives, but the overall damage costs due to such disasters and specific risk to the vulnerable community have not decreased. In this context, it is also apparent that over 95% of all the landslide disasters occur in the developing countries (Hansen, 1984). Due to the higher relative cost of damage (cost in terms of GDP), those living in developing countries and especially those with limited resources tend to be more adversely affected from such hazards. Thus the recent trend throughout the world and especially for the affected developing countries is to develop effective mitigation measures and safer land utilization regulations rather than cost-intensive projects of slope stabilizations (Guzzetti et al., 1999). For landslide mitigation, spatio-temporal assessment of landslide hazard and risk at national, regional and local scales are being considered as important decision-making tools for making detailed mitigation plans and preparedness for such hazards. Towards achieving this, characterisation of each type of landslide danger is a fundamental step.

As a part of the natural processes, landslides occur as a consequence of a number of predisposing and triggering factors (Varnes, 1978). Thus, characterization of landslide danger broadly means understanding or unravelling the mechanisms or interplay of all the determining factors (both predisposing and triggering), which leads to any slope failure. On medium scale (1:25,000), the landslide hazard assessment in a catchment or part of a hill district is being attempted by several methods ranging from simple geomorphological analysis to complex data treatment (Baeza and Corominas, 2001). The reliability of the geomorphological analysis depends on the appropriate interpretation of the landscape, which is based on subjective expert criteria. In order to reduce subjectivity and quantify the degree of susceptibility, data-driven techniques have been incorporated. All these methods first attempt to spatially disintegrate landslide-susceptible areas by correlating some of the main causative factors that contribute to landslides. But all the above techniques primarily need landslide characterization, which signifies the understanding of the slope processes and the relationship of those processes to the predisposing and triggering factors such as geology, geomorphology, hydrogeology, failure and slide mechanics, climate and vegetation etc. The framework of such landslide characterization forms the basis of landslide inventory data for the successful use in any landslide hazard and risk prediction method. Such methods are founded on the notion that their behaviour is controlled by a) natural, physical or logical laws, and once these laws are understood, the methods can be adopted for representing the phenomena of interest (Chung and Lecrec, 1994) and b) the interplay of causal factors, which was responsible during the present and past landslide processes will be acting in similar manner and will cause similar type of landslide processes in future under the same geofactor setting.

Thus, the identification and mapping of past and present landslide bodies and understanding their processes constitute the fundamental steps for predicting future slope failures, which remain highly subjective. Likewise, many basic instability determinants for landslides cannot be acquired and mapped with adequate accuracy. Most of the current methods for manipulating instability factors and evaluating hazard levels still remain error-prone or questionable (Carrara, 1993). Thus, most scientists recognized the superiority of quantitative techniques due to their rigorous quantitative data treatment framework which promotes objectivity. But a quantitative prediction model is superior only if the conditions of validity and accuracy are testified and the outcome of model results is scientifically explained in the backdrop of landslide causes and processes**.** The possible data uncertainty and incompleteness may also attribute to uncertainty in estimation of accuracy and validation of data-dependent models because a valid and accurate quantitative model respecting a fully data-driven path is possible only if it is supported by quality data sources (Haining, 1990). Thus using valid quantitative methods, which maximize accuracy while they incorporate subjectivity or expert knowledge when data is not sufficient and accurate, may prove to be the best compromise in the selection of a better objective prediction (Chung and Lecrec, 1994). For achieving this, techniques for integration of knowledge in a quantitative prediction model require to be developed through objective process of knowledge transfer using various statistical as well as mathematical interactive data treatment techniques right from the data selection to data integration stage.

Any hazard analysis is incomplete if it does not include the temporal probability. Hazard analysis is thus the process of identification and characterization of the spatial landslide susceptibility together with evaluation of their corresponding frequency of occurrence (Fell et al., 2005). The estimation of the latter on medium scale is always difficult due to the absence of event-based complete historic record of both landslide occurrences as well as triggering factors (van Westen et al., 2006). Reliable temporal prediction or hazard analysis thus needs well correlated event-based, multi-temporal landslide inventory data and correlated rainfall events database. A well validated quantitative landslide hazard prediction is always the basic step to attempt quantitative risk analysis of any area (Chowdhury and Flentje, 2003). Risk can be analysed by introducing another important analysis called consequence analysis to the hazard already assessed. This involves a) identifying and quantifying the elements at risk, b) assessing temporal spatial probability for the particular risk elements vis-à-vis the particular landslide hazard and c) assessing the vulnerability i.e. extent of damage (Fell et al., 2005). Risk can then conceptually be calculated by the following formula:

$$\text{Risk} = \text{Hazard} * \text{Vulnerability} * \text{Cost of element at risk}$$

Since, on medium scale, hazard information is mostly represented by a spatial database, but true quantitative risk analysis involves calculation of specific risks of each element at risk for each landslide magnitude/ scenario and up scaling this information into the medium scale spatial database, which is difficult because of scale constraints. This compels most of the researchers for incorporating qualitative or semi-quantitative risk analysis approach in stead of a quantitative estimation. But the most difficult problem of any consequence analysis is the vulnerability assessment, which are mostly carried out qualitatively because estimation of past damage potential vis-à-vis hazard magnitudes through assessment of vulnerability curve is highly affected by lack of sufficient data on extent of past landslide damages.

1.1. Motivation:

Every year during monsoon (June-September), almost the entire Himalayas and parts of Western Ghats Mountains in India witness several landslide events of variable dimension/ types causing substantial loss to lives and damages to properties. In India the average annual landslide damage costs are estimated to be nearly 1 billion US dollars for the total 89000 Km of slide-prone roads in India (Mathur, 1982). It has also been reported that in 1968 and 1973, two catastrophic damages were caused due to floods and landslides to the roads of West Bengal and Sikkim in India (area where the proposed research will be taken up) and total cost of such damage including restoration costs were estimated to be 53 million and 24 million US dollars respectively (Chopra, 1977). Almost in every monsoon, the densely populated hill towns, tourist spots, religious and mythological places and prominent hydro power sites located in Himalayas, remain cut-off from the rest of the country for days together due to road blockades by some conspicuous landslides. The entire Himalayan orogenic belt is geologically complex, young and active and falls within active seismic zone (Zone – IV to V) of India (BIS, 2002) and also receives a high amount of monsoon rainfall (3000 mm to 5000 mm). Due to its complex geo-environmental setting and strong influence of various triggering factors, all types of slope instability problems are thus abundant in the entire terrain.

Observing the above-mentioned hazard scenarios and increased rate of risk due to excessive rate of urbanization in the Himalayan/ sub-Himalayan terrains (e.g. Darjeeling and Kurseong towns in the proposed study area witnessed a population growth of 47.18% and 49.74% respectively during 1991-2001 period – Census 2001 Report) and parts of Western Ghats over the past couple of years, the Government of India has brought a paradigm shift in the approach to tackle these disasters. The new approach aims for adopting sustainable and multidisciplinary approach for disaster mitigation as the key developmental process (National Disaster Management Report, 2004). To accomplish the above task in the field of landslide hazards, an important scientific step is to develop an effective methodology for determination of landslide hazard and risk, so that the aspects of disaster preparedness for larger areas in those fragile parts of India can well be addressed. Through successful completion of this research, gaps in current level of expertise in the field of landslide spatio-temporal analysis and risk studies would be filled up, which, in turn will facilitate improvement of the current knowledge base of the hazard scientists of the Geological Survey of India.

2.0. Research problems

Landslide hazard zonation is defined as the 'mapping of areas with an equal probability of occurrence of landslides of a given type and magnitude within a specified period of time' (Guzzetti et al., 1999; Varnes, 1984). To do this, the fundamental steps are the spatial prediction of susceptible zones, estimation on the probability of magnitude of future landslide and then temporal prediction of landslide recurrence in different susceptible zones. Landslide hazard estimates in turn, are the most crucial input to risk analysis, the latter being defined as "the expected number of lives lost, persons injured and damage to property and disruption of economic activity due to a particular landslide hazard phenomenon for a given area and reference period" (Varnes, 1984). For calculation of risk, a hazard-consequence matrix approach is generally followed (Chowdhury and Flentje, 2003), where risk per each hazard scenario is computed. Literature reviews pertaining to all the above aspects lead to the some of the following research problems.

2.1. Problems related to scale of spatial prediction

Landslide susceptibility can be determined through deterministic method, which is followed in smaller areas on larger scales (larger than 1:10000). These methods are process-based and give more detailed results, expressing the hazard in terms of factor of safety to each mapping unit. The deterministic method can quantitatively represent the landsliding processes by considering the detailed physical and dynamic in-situ parameters of slope forming material and can easily be used to retrieve temporal probability information by modelling different groundwater scenarios caused by different rainfall event (triggering factor). The deterministic methods highly depend on a large number of detailed site-specific geotechnical and groundwater parameters, otherwise its results are oversimplified (Moon and Blackstock, 2004) and that is why for medium scale (1:25,000 to 1:50,000) analysis in a large area, the use of such deterministic method may not be feasible. Moreover, deterministic models are also difficult to represent as 2D GIS spatial data product because it considers depth wise data variability for calculation of factor of safety.

2.2. Problems in methods for medium-scale spatial prediction

In medium scale landslide susceptibility analysis, knowledge-driven/ heuristic and data-driven quantitative methods are prevalent. The knowledge-driven methods are mostly qualitative (direct) but semi-quantitative methods (indirect) based on heuristics are also followed. The data-driven methods are mostly statistical (bivariate and multivariate) and few are mathematical (artificial neural network).

The knowledge-driven /heuristic direct approaches to spatial prediction of landslide susceptibility involve detailed geomorphological mapping using uniquely coded polygons, which are evaluated one-by-one by an expert to assess the type and degree of hazard (Barredo et al., 2000; Hansen, 1984; Varnes, 1984). Indirect heuristic approach utilizes data integration techniques, including qualitative parameter combination, in which the analyst assigns weighting values to a series of terrain parameters and to each class within each parameter. The relative importance of each terrain parameter as a predisposing determining factor of slope instability is quantitatively determined by pair-wise comparison using the so-called analytical hierarchy process (AHP) (Saaty, 1996) or is incorporated through spatial multi-criteria evaluation (SMCE). In direct heuristic methods, use of detailed geomorphological factor maps in general raised the overall accuracy of the susceptibility maps, though the accuracy of such direct qualitative model largely depends on the experience of the expert using the method. Whereas, in indirect heuristic methods, similar weight values are considered for all locations within the same factor. The addition of such unique weight values tends to "flatten out" the results of indirect methods. Thus, the main limitations of the knowledge-driven methods are the subjectivity involved both in the direct mapping as well as in the assignment of weights in indirect methods (van Westen et al., 2003) and general non-availability of any quantitative technique of model validation.

Since the late eighties, the increasing popularity of geographic information system (GIS) has facilitated development of various quantitative or data-dependent landslide spatial prediction methods (Aleotti and Chowdhury, 1999). GIS is very suitable for such methods, in which all possible landslide contributing terrain geofactors (evidence) are combined with a landslide inventory map (target) using data-integration techniques (Bonham-Carter, 1996; Chung et al., 1995; van Westen, 1993). Thus, in such quantitative methods of hazard estimation, spatial associations of past and present landslides and associated geofactors act as the key parameters to predict future landslides (Carrara et al., 1991; Zezere et al., 2004). These data-dependent methods aim to introduce objectivity in analysis by reducing subjectivity or generalisation of the true knowledge-driven methods.

Amongst the quantitative methods, the application of the *bivariate* statistics (e.g. weight of evidence method) in landslide spatial prediction is common and it needs to be weighed in light of following limitations because of mis-applications by many researchers, which include i) generalisation by assuming that landslides happen under the same combination of factors throughout the study area, ii) ignorance of the fact that each landslide type has its own set of causal factors, and should be analysed individually and iii) lack of suitable expert opinion on different landslide types and processes and of slides of different periods, which may be inevitable if these methods are solely applied by GIS-experts, and not by earth scientists. (Van Westen et al., 2003). Another debate regarding bivariate methods is that, instead of partitioning the study area into unique domains or mapping/ terrain/ slope unit, the conditional probabilities are determined for separate geofactors and then added sequentially under the assumption that such factors are weakly correlated to each other (assumption of conditional independence amongst independent variables). It has been argued by some researchers (Carrara et al., 1995) that this method perhaps holds true where very few environmental factors are only responsible for landslides, and a sound expert knowledge exists about the landslide processes.

The subjectivity in considering and using individual geofactor of landslide occurrence in bivariate statistical methods, results in uncertainties in spatial analysis. It is apparent that landsliding in any area is directly linked with a phenomenon, which is the result of the interplay of several interrelated geofactors, and many of which are sometimes ill-known, extremely general, fuzzy and even unmappable. Therefore, several authors (Baeza and Corominas, 2001; Carrara, 1983; Carrara, 1989; Neuland, 1976) apply multivariate statistics in spatial prediction of landslide hazard to reduce the uncertainties more objectively. The general linear multivariate model assumes the following form:

$$L = B_o + B_1X_1 + B_2X_2 + B_3X_3 + \dots \dots + B_mX_m$$

where L is the predicted degree of presence/ absence (or the areal percentage) of landslides in each terrain unit (which can be a unique condition unit or slope unit), X's are input predictor variables (or instability factors) measured or observed for each terrain unit, and B is the co-efficient estimated from the data through techniques such as multiple regression, discriminant analysis etc (Carrara, 1983; Carrara, 1989). These multivariate methods generally require large amount of data sets and sometimes used as black box methods independent of expert knowledge.

Bivariate or multivariate methods may be found statistically suitable to predict future landslides at medium scales (1:25,000 to 1:50,000), but logical explanations of the results or outputs and exact knowledge about the dependencies of causal parameters with the target are sometimes absent in these type of methods. Since these methods are mostly based on various statistical data treatments focused mainly for objective elimination or reduction of errors and uncertainty in prediction, the aspects of data quality, reasoned selection of input parameters and inherent fuzziness of some geofactor data etc. are frequently overlooked. Multivariate methods, in spite of limitations and pitfalls in applications, are used nowadays as among the most feasible quantitative tools for assessing different levels of landslide susceptibility. For example, when a set of independent variables include both good and bad predictors (the latter having no clear physical relationship with mass movement processes), a step wise regression technique in multivariate statistics is followed with an aim to eliminate statistically non-significant factors, but sometimes the output of these analyses may generate unreliable and meaningless results. In similar way, artificial neural network (ANN) – a mathematical technique is also used for spatial prediction of landslide hazard. The ANN method is not sensitive to any statistical distribution of data, and can integrate both continuous as well as categorical data set. The ANN methods are adaptive and generic in nature. They are construed to handle imperfect or incomplete datasets and can capture non-linear and complex interactions among variables of a system (Lee et al., 2003). Since ANN is almost independent of the quality of input variable; chances of getting unreasonable goodness in results are sometimes highly abstract and misleading. Like multivariate techniques, in ANN method also, the internal processes which train the input dataset and minimise the statistical errors and uncertainties are difficult to follow.

2.3. Problems in spatial prediction of landslides in India

In India, the BIS (Bureau of Indian Standards) guidelines (BIS, 1998) provide a generalized heuristic system for weighting or ranking of environmental factors of landslide susceptibility without directly or indirectly considering landslide inventory data. The weights have been defined by a group of experts using the analytical hierarchy process (Saaty, 1980), and are being used within a rigid framework all over India including Darjeeling Himalayas. As such, the system may not be directly suitable for the application in all landslide-prone parts of India, because the importance or influence of geological factors to landslide occurrence may vary from one area to another due to variability in spatial associations of predictor variables which are often different for different landslide types. The BIS system is as such rigid because, the mutual importance/ influence of the environmental factors in any area can not be tested/ verified or incorporated in the system through objective means.

2.4. Problems in temporal prediction

For hazard evaluation, statistical or quantitative hazard evaluation methods are considered as a suitable method for assessing the spatial probability of large areas on medium scale but there are problems in evaluating either temporal probability or the effects of future environmental changes using statistical models. With the above mentioned quantitative (statistical/ mathematical) susceptibility methods, temporal probability of hazard can not be directly determined. In medium scale, the temporal probability of hazard in a catchment is empirically determined through frequency-magnitude analysis of past landslide inventory along with their mutual relations with the magnitude/ intensity and duration of triggering factor such as rainfall or earthquake (van Westen et al., 2006). And a number of researchers (Terlien, 1998; van Westen et al., 2006) have indicated that incorporation of temporal probability with the spatial information is by far one of the most difficult tasks for landslide hazard and risk analysis because of want of relevant quality data sources on past landslide events and continuous rainfall data.

Landslides are normally initiated by triggering events and usually in tropical terrains, rainfall or storms are one of the most obvious triggering events for landslides (Malamud et al., 2004; Starkel and Basu, 2000). The increased rainfall results in increased moisture content of the slope forming materials along with substantial modification of the ground water scenario, which can raise pore water pressures above the critical value necessary to induce slope failure. This knowledge has motivated many researchers to empirically quantify the relationship between landslide occurrences and rainfall variables and thus the rainfall/landslides relationship forms the basis for temporal prediction of landslide occurrence (Dai and Lee, 2001; Dai and Lee, 2003). These relationships are also established by many other workers (Brand et al., 1984; Caine, 1980; Crozier, 1999; Lumb, 1975) keeping in mind the assumption that there exists some linear relationship between the occurrence of landslides and the quantity of rainfall, in terms of rainfall intensity and duration such as short-term rainfall or antecedent rainfall.

The basic aim of temporal prediction is to statistically evaluate different rainfall event of variable intensity and duration and correlate those rainfall events with the frequency-magnitude distribution of past landslides with the assumption that rainfall intensity has an empirical relation with landslide frequency and magnitude. Literature review indicates that this relationship is extremely site-specific and perhaps non-linear and is also dependent highly on the availability of complete data on rainfall and landslides. Thus, problems arise normally when there are data inadequacies, lack of correlation between the dates of landsliding and rainfall, discontinuous rainfall record, non-availability of rainfall stations and lack of any systematic pattern of events etc. In this situation, adoption of appropriate approaches to transform the susceptibility information into the spatio-temporal prediction becomes higly difficult and problematic (Chowdhury and Flentje, 2002; Gabet et al., 2004; van Westen et al., 2006).

Temporal prediction of future landslide can also be determined through statistical analysis of only landslide frequency data through determination of landslide recurrence probability using some discrete probability distribution models (Coe et al., 2004; Crovelli, 2000). The above analysis solely depends on the availability of a complete historic event-based landslide database and can generally be applicable to large areas. But this method lacks any direct correlation with initiation events (e.g. effect of rainfall etc.) or different landslide types and magnitudes.

2.5. Problems in landslide inventories and geofactor databases

The crucial issue in any of the above data-dependent methods for landslide susceptibility prediction lies with the input data, which sometimes remain inadequate in quantity and quality for the task to be accomplished in many study areas. In this regard, the most fundamental step of spatial prediction and for subsequent temporal probability assessment is to prepare a detailed and reasoned multitemporal landslide inventory and geofactor databases after developing relevant knowledge base about the landsliding processes of the study area, which might be extremely specific to different slide types (e.g. rock slide, debris slide, shallow and deep-seated slides etc.) and even to the slides of different time periods. For landslides, two classes of inventories are generally mapped: (i) landslide events associated with certain triggering factors, and (ii) historical (geomorphological) landslides, which are the sum of one or many landslide events over time in a region (Malamud et al., 2004). A landslide inventory used for any hazard analysis must include information on the type, material involved, degree of activity, location, date of occurrence, aerial extent, volume, relative age, estimated depth, estimated velocity, degree of certainty of landslide mapping etc (Cruden and Varnes, 1996; Varnes, 1978). Although, data certainty and accuracy levels may be high in case of landslide inventory mapped by well-trained and experienced experts, there are certain inevitable limitations in this respect. For example, old and dormant landslides sometimes can not be identified easily (Carrara, 1993). Errors in estimating the dimensions of landslides do exist due to inaccurate base maps (Malamud et al., 2004). Problems in inferring landslide polygons from cluster results errors in the analysis of the frequency–size statistics of landslides, which is an extremely important parameter for frequency-magnitude analysis of hazard. Coalescence of multiple small slope failures into a larger landslide area may locally prevent the correct identification of the smallest failure. Furthermore, for many of the mapped landslides, the exact date of occurrence remains unknown, thus making it difficult to correlate such landslides with triggering events, especially as different landslide types have different magnitude/ intensity of meteorological triggers. The lack of temporal information (exact date of occurrence) in landslide inventories may lead to serious bottlenecks in determination of temporal probability and therefore pose difficulties in quantitative hazard and risk assessment (van Westen et al., 2006).

Data about landslide geofactors collected in field or through RS techniques using GIS may also be affected by inaccuracies and inherent fuzziness, whose magnitude sometimes is difficult to ascertain and thus tremendously affecting the whole hazard evaluation process irrespective of the methods used (Carrara et al., 1995; van Westen et al., 2006). Hazard analysis methods may also be affected by large amount of errors and wrong assumptions, or may generate questionable or unequivocal outcomes. In addition, a mixture of continuous variables (e.g. elevation) and categorical variables (e.g. presence/ absence of any rock type) lead to a solution that is generally not optimal, namely, it should not affect the probability of correct prediction. Moreover, there are several independent variables which might be of extreme importance for causing some specific landslide types, but may be difficult to map due to scale constraints or their direct spatial representation in GIS may not be feasible (e.g. structural parameters of rocks, rock/ soil weathering, soil depth etc.).

2.6. Problems in landslide risk analysis

The ultimate aim of all the above spatial and temporal analyses is to attempt for a landslide risk analysis by incorporating the information of consequence analysis of elements at risk. Risk is expressed as "probability of an adverse event times the consequence if that event occurs" (Fell et al., 2005). Consequence analysis takes into account the identification and quantification of elements at risk (E), their temporal-spatial probability and the vulnerability (V) that is the degree of loss or extent of damage caused by a particular landslide hazard of certain magnitude or intensity. The specific risk can then be calculated by multiplying the above consequence information (V*E) with the spatio-temporal probability of that particular landslide event of certain magnitude/ intensity (H). Finally the total risk of an area is calculated for all the risk elements and for all landslide types which can be represented (Lee and Jones, 2004) as

$$\text{Risk} = \sum (H^* \sum (V^*E))$$

The statistical hazard maps are mostly used in qualitative to semi-quantitative risk assessment only. If it is combined with landslide inventory maps for different triggering events, or for events with different return periods then the quantitative hazard maps can be used for quantitative risk assessment over larger areas (van Westen et al., 2006; Zezere et al., 2004) on a catchment scale. Quantitative risk analysis in any study area is thus largely dependent on the successful completion of the temporal probability estimation, with the help of event based landslide record and rainfall data. But the quantitative assessment of temporal probability may be constrained due to lack of good records on historic landslide data such as frequency, type, volume and damage etc. Sometimes, the historic event-based landslide records may be available but the database may lack the above essential information on its type and magnitude, which are the essential parameters for the frequency-magnitude distribution of events. However, availability of past sets of high resolution aerial photographs and satellite imageries can indirectly facilitate us to collect relevant data on the morphometric part of the old slides but correctly assessing the time of occurrence of those slides with available rainfall events remains a major bottleneck in this respect.

A higher level of uncertainty always occurs while addressing the specific temporal-spatial probability of an element at risk in a larger area with respect to a landslide because it becomes spatially difficult to link the particular element at risk with the exact location of the hazard and also to upscale this information onto a medium scale spatial data (map). Regarding this, the correct spatial estimation of landslide initiation zone, run out zone, spatial intersection and characterization of elements at risk in case of a haphazardly-built hill settlement are extremely important tasks. This problems are often encountered when a particular element at risk is affected by different nature and type of landslides, then estimating the temporal-spatial probability of that element at risk and determination of its vulnerability become really difficult (van Westen et al., 2006). The successful completion of above estimation may be possible in case of availability of an accurate DTM and orthophotos derived from high resolution stereo data (e.g. Cartosat – 1 - 2.5 m resolution) and availability of adequate correlatable data on past landslides and rainfall.

Moreover, in consequence analysis of risk, the problem of vulnerability assessment for landslides of different magnitude always remains an unsolved task because past damage records are rarely available in any area. The landslide damages are only localized and may be represented as a point and may only be available in records when there is sufficient damage to important facility or loss of life. Thus, unlike floods, earthquake, storm hazards, the entire damage attributes caused due to past landslides in general are rarely available in most of the study areas. The above specifications sometimes act as serious bottlenecks towards quantitative estimation of risk for a large area on medium scale.

Because of lack of temporal information, most hazard maps are qualitative in nature and concentrate basically on determining the susceptibility or spatial prediction only. Lack of such temporal probability information in susceptibility or spatial probability maps and lack of quantitative vulnerability information are one of the major bottlenecks in quantitative risk analysis (van Westen et al., 2006). In this respect it is always to be evaluated whether quantitative risk calculation on medium scale for a large area is at all pre-requisite or not because medium scale risk maps are mostly used for regional mitigation planning wherein addressing the risk levels through qualitative/ semi-quantitative means may be equally effective (Australian Geomechanics Society and Sub-committee on landslide risk management, 2000; Lateltin, 1997; van Westen et al., 2006).

3.0. Research hypotheses

1. Spatial prediction of landslide hazard in a large area on medium scale (1:25000 to 1:50000) can be improved by developing and integrating knowledge into the system of data-driven approaches of spatial prediction through development of a knowledge-guided quantitative landslide hazard assessment method. The knowledge required to be used would be based on logical understanding of different landslide processes, knowledge about the interplay or combination of different geofactors and effect of triggering mechanisms etc. to different landslide types (both space and time). This aims at the problem of how the above knowledge base about different landslide processes can be outlined and through which methodology; the same can be integrated or elicitated in the best-suited quantitative method, so that the model becomes effective in hazard prediction. Using such strategy to maximize accuracy while incorporating subjective expert knowledge may prove to be the best compromise in the selection of relevant approaches to any prediction. (Chung and Lecrec, 1994).

2. Thus, better results in all the quantitative methods could be obtained by entering into the model only the variables that the investigator assumes to be most significant (incorporation of knowledge right from the selection stage). Moreover, like other natural system, as discussed before, the landslide hazard database also contains inherent fuzziness both in some of its geofactors and landslide inventory database, which might be extremely important to be modelled or considered for better prediction and that can only be integrated through assimilation of relevant knowledge about the landsliding processes. Though, in this respect subjectivity also can not be ruled out because, such inputs largely depend on the analytical skill and experience of the investigator. Since, the input factors are invariably interrelated, considering all the available variables without the proper knowledge about the landslide processes and causal factors can produce even worse results since some variables may be characterized by meaningless spatial and temporal existence. Thus, the right approach of spatial and temporal prediction would always be the incorporation of relevant expert knowledge and simultaneous statistical or mathematical data treatment for reducing the level of uncertainties and errors to compensate situations of knowledge gaps.

3. Any quantitative prediction method is superior only if validity and accuracy are met (Matthews, 1981). For any method to be valid, it must express the true meaning of what it is attempting to represent while at the time must respect the assumptions of the quantitative techniques applied. Haining (1990) encourages sensitivity analysis in areas where observational data are not always very accurate or precisely measured. This may be used to help construct a better prediction model. An interactive procedure, aimed at modifying some estimators influenced by a small number of extreme values, is an example that may prove valuable. This concept is attractive provided that the degree of subjectivity is proportional to the model validity and to the data properties. For example, a valid and accurate prediction model accompanied by quality data would require little subjectivity and would, thus, respect a fully objective path. Whereas a valid quantitative method, which maximizes accuracy while incorporating subjectivity or expert knowledge is always an effective tool in situation when data is not sufficient. This method proves to be the best compromise in the selection of a knowledge-guided quantitative approach of landslide hazard prediction (Chung and Lecrec, 1994). Such well validated and tested quantitative landslide hazard prediction is always the most fundamental step towards any quantitative risk calculation.

4.0. Research objectives

1) To develop a knowledge-guided quantitative technique for medium scale (1:25000) spatio-temporal prediction of landslides.

a) To develop a conceptual knowledge base regarding different landslide processes and linking the same to build up causal relationships of different combinations of geofactors and intensity of triggering event with different landslide types (both space and time).

b) To develop a GIS-based quantitative technique for selection and preparation of a reasoned geospatial and temporal database of landslides, related geofactors and past daily rainfall record (triggering factor) using interactive treatment of both knowledge base and data.

c) To develop a quantitative methodology for eliciting/ representing and validating the developed knowledge-guided spatio-temporal database into prediction maps for determination of spatial and temporal probabilities through proper process of data integration.

2) To develop a methodology for medium scale (1:25000) risk analysis by using the knowledge-guided spatio-temporal prediction models through consequence analysis of elements at risk.

a) To develop a method for identifying types of elements at risk (both physical and social) and their nature of vulnerability to the available landslide hazard of specific type and magnitude.

b) To develop a methodology for integrating the consequence information of elements at risk with the spatio-temporal probability for attempting assessment of total risk.

5.0. Research questions

Related to Objective – 1

Sub-objective – 1a

1. What are the different landslide processes (both past and present) prevalent in the study area?

2. Which group of geofactors and classes of geofactors are found responsible for different slide types (space and time)?

3. Is there a logical chain or network of interrelationships between groups or classes of geofactor and individual landslide process?

4. Which essential geofactors are difficult to map or difficult to exhibit as geospatial data?

5. Which geofactors have fuzziness inherent and thus require modelling to give a better prediction?

6. What and how much uncertainties and gaps are present in the database of frequency-magnitude analysis of temporal landslides and daily rainfall data?

7. Is there a specific type of empirical relation between different landslide processes (type and magnitude) and rainfall distribution in the study area?

Sub-objective – 1b

1. Which GIS based data treatment can be applied to geofactors that are difficult to map, so that their importance can be indirectly considered in spatial analysis?

2. What and how much uncertainties and errors in geospatial data products are caused by scale-related disparities, data inadequacies and gaps in input data?

3. Can the identified fuzziness in geofactor database be quantitatively incorporated in the geospatial database?

4. Which statistical data treatments are relevant during preparation of reasoned spatial database of geofactors and temporal landslides, so that the required quantitative data treatment can interactively supplement the gap in knowledge base?

5. Which basic mapping unit (for medium scale assessment) is relevant/ appropriate for evaluation of landslide hazard in a Himalayan terrain? Does it influence prediction accuracy?

6. Which statistical functions/ treatments help in determining the rainfall event and reduce the uncertainty in correlating the rainfall event with the frequency-magnitude distribution of temporal landslides?

Sub-objective – 1c

1. Which quantitative modelling technique is suitable for data and knowledge integration process, so that model prediction accuracy can be improved?

2. How sensitive is model prediction accuracy to selective and sequential use of knowledge base?

3. Can accuracy and precision of prediction by a purely data-driven be increased if model parameters are guided by relevant expert knowledge on landslide processes?

4. What are the possible limitations of a knowledge-guided quantitative prediction method?

5. Can a knowledge-guided quantitative prediction model be tested/ validated in other areas having similar physiographic conditions or can it be tested with different temporal test dataset (time and space robustness of models)?

Related to Objective – 2

Sub-objective – 2a

1. Which types of physical and social elements at risk should be involved for calculation of risk at medium scale?

2. What are the possible uncertainties in mapping such elements at risk from the derived data products of stereo CARTOSAT – 1 and other available sources (Census data etc.) in a highly dense and haphazardly-built settlement on steep hill slopes?

3. Can the worth of risk elements be quantitatively identified? Can the factor of "loss of human life" be possible to incorporate in the landslide risk analysis?

4. What are the difficulties for deducting temporal spatial probability of elements at risk with respect to a specific landslide hazard of certain magnitude?

5. Can the vulnerability (both physical and social) of risk elements be quantitatively identified for the test area with the available input data?

Sub-objective – 2b

1. What uncertainties are there in upscaling the information of temporal spatial probability and vulnerability of an element at risk based on a particular landslide danger (specific risk) into a medium scale risk map?

2. Can a standard risk curve be generated when all the individual specific risks per landslide hazard scenario are summed up for all elements at risk to calculate the total risk of an area?

3. Is the quantitative risk analysis possible under the condition of data inadequacy or data gaps? What are the possible limitations/ difficulties in quantitative risk calculation?

6.0. Proposed methodology for Knowledge-guided spatio-temporal prediction of landslides

For quantitative evaluation of medium scale landslide hazard or spatio-temporal prediction, estimation of "where" a landslide will occur, "when" or "how often" it will occur and "how large" the landslide are would the most important aspects to determine (Guzzetti et al., 2005). Mathematically, landslide hazard (H_L) thus can be defined as the conditional probability of landslide size (A_L), conditional probability of landslide occurrence in an established time period t, P_N and of landslide spatial occurrence, S, given the geofactor setting. Accordingly, the hazard (H_L) can be determined by the following formula after assuming the independence amongst the above three different probabilities (Guzzetti et al., 2005).

$$H_L = A_L * P_N * S$$

It is proposed that the above quantitative determination of landslide hazard can be optimized or improved if it is guided by the knowledge base on the landslide processes of any study area. This knowledge guidance in any quantitative method should initiate right from the data collection stage and proceed further to the data treatment and even in the final data integration phase to arrive at a better prediction. The purpose of knowledge-guided data-driven/ quantitative method of hazard prediction is that during all its process execution stages, it must be flexible enough to handle step-wise training and testing of input data sets and will allow incorporation of relevant knowledge base, whenever required. The various statistical or mathematical procedures employed in such methods would aim at physically training the input data sets by applying the knowledge on logical dependencies of various causative and triggering factors and combinations of their specific factor classes with specific type of instability of the study area. The inclusions of the above conceptual knowledge into the model parameters of the quantitative hazard prediction would be the fundamental objective of this research, which will be attempted through the following methodologies starting from data capture to data integration. The general methodology for developing such knowledge-guided quantitative techniques of spatio-temporal prediction of landslide can be schematically illustrated in the following flow diagram (Figure: - 1).

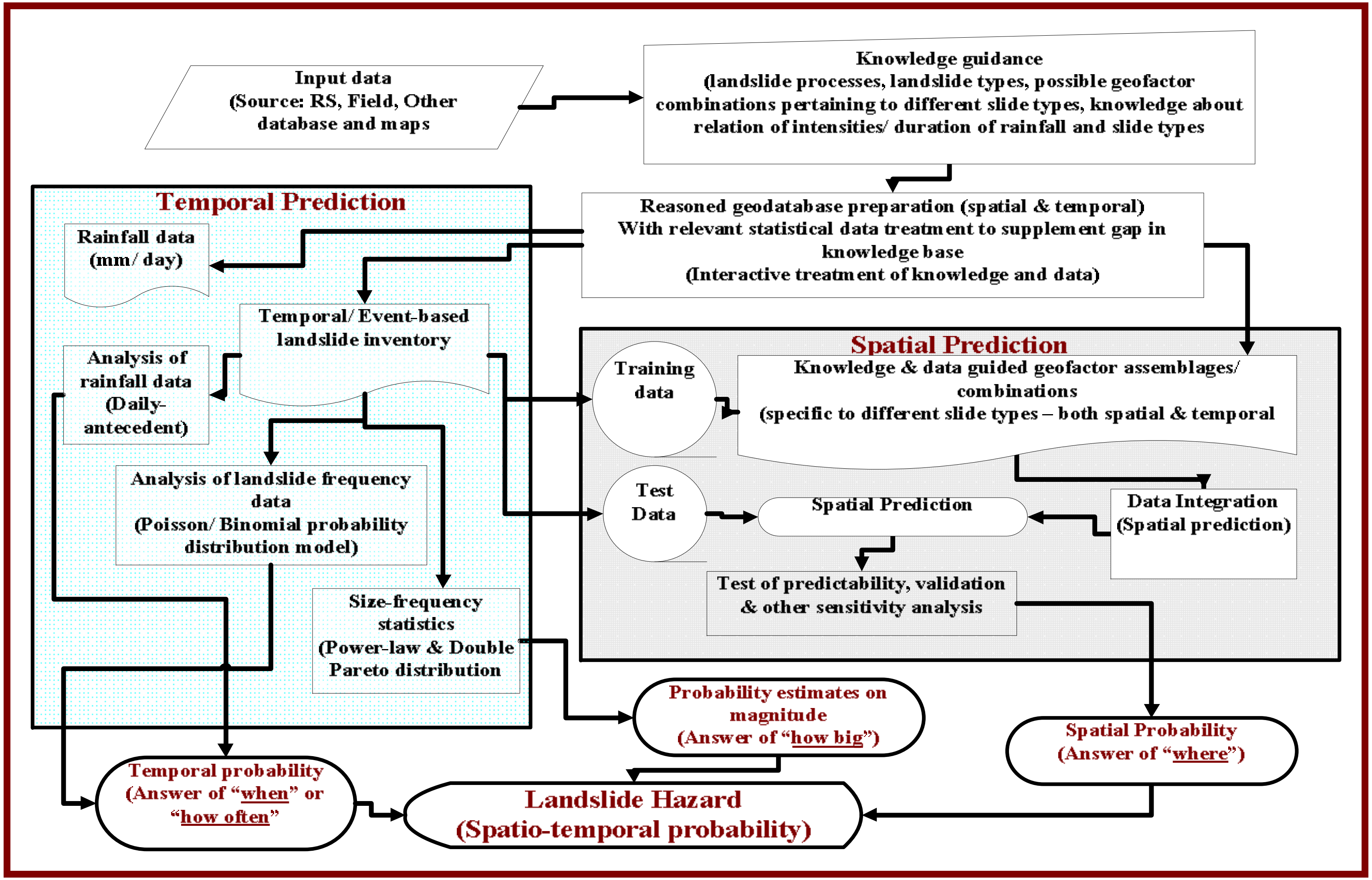

Figure - 1 = A general flow diagram of the knowledge-guided spatio-temporal prediction methods

The tasks described below have been framed up in such a way so that the required spatial prediction studies are linked to the temporal database of various landslide types (including magnitudes), to arrive at the temporal and magnitude prediction of future events. That is why the following '*conceptual knowledge development', 'data capture', 'geodatabase preparations'* and *'data integration'* steps are proposed to be carried out keeping the entire spatio-temporal and risk analysis of the study area in mind.

6.1. Development of conceptual knowledge base: The conceptual knowledge development is the most fundamental step in any knowledge-guided hazard analysis process and this is being accomplished right from taking the decision of which data products are essential for landslide hazard analysis and through interactive and iterative interactions and selections of the relevant input data. During this stage, the possible landslide processes responsible for different slide types (both space and time) and their pre-disposing and triggering factors are outlined, which becomes the fundamental steps to develop the knowledge base for developing the causal dependencies or relation networks amongst the various predictor variables (choice of combinations of independent variables) and target. These combinations may be extremely specific to specific slide types and also to slides of different periods. Ultimately after successful completion of this stage, which is highly dependent on the intricate studies of various input parameters (both independent and dependent) available from detailed field investigations and data from RS and other sources, the following knowledge-bases would be developed:

a) Idea to develop temporal and spatial classification of available landslide inventory and trace the spatial evolution of landslides.

b) Identification of prevalent landslide processes (spatial and temporal) and identification of different slide types, variation in their magnitudes that is variability in size (area or volume).

c) Reasoned identification of geofactors and its factor classes and their specific combinations (specific to individual slide type and even for slides of different periods).

d) Development of causal/ dependency network of geofactors responsible for all the possible landslide types prevalent in the study area.

e) Distribution of rainfall events over discreet or continuous time steps and linking the same to the recorded major landslide events of the study area (linking of intensity-duration of triggering mechanisms to the different slide types).

Conceptually, development of the above-mentioned knowledge base at this stage would be done through a circular and continuous process of repetitive interactions of data and knowledge, which is schematically shown in Figure -2.

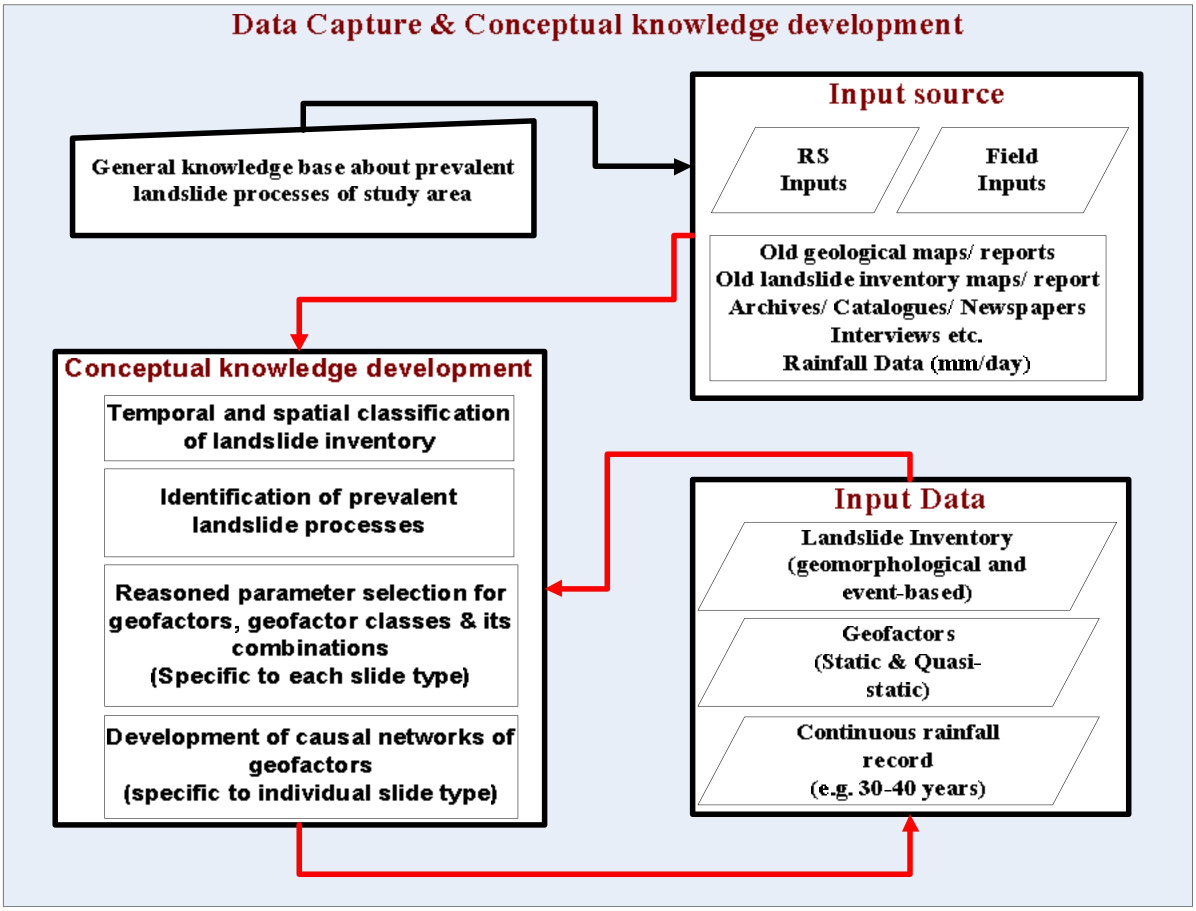

Figure – 2: Flow diagram showing data capture and knowledge development processes

6.2. Selection of sources, types of input data and data capture: The choice of the sources of the relevant input data depends primarily on the basic knowledge about the resolution of specific data product, correctness in identification of specific information from the specified data sources and knowledge about its uncertainties. Despite the subjectivity, selection of sources, quality of input data types, methods of data capture at this stage largely depend on the logical selection of input database through assimilation of suitable knowledge about the terrain conditions, prevalent landsliding processes in the study area. Keeping these aspects in mind, the different input data to be required for the proposed research are listed below:

6.2.1. Sources and types of Input Data:

No.	Primary source	Primary data product	Expected final/ derived data product
1.	Cartosat-1 panchromatic stereo data (2.5 m resolution	Carto DEM and Orthophoto	Slope, Aspect, Relief, Slope shape or slope profile, Drainage, landslide incidence, landuse and land cover, geomorphology, lineament & structure, mapping unit, elements at risk etc.
2.	Large scale (preferably	Orthophoto	Temporal landslide inventory, temporal

	1:20000 or larger) B&W aerial photograph		landuse and land cover pattern
3.	Merged data products of IRS LISS-III PAN (5 m spatial resolution) and IRS LISS-III MSS/ IRS-LISS-IV MSS images of different periods, if required	Orthophoto	To supplement temporal landslide inventory data and temporal landuse/ Landcover data
4.	Old landslide reports/ newspaper/ Old landslide inventory maps		Augmentation in the temporal landslide inventory database
5.	Toposheets (1:25000/ 50000)		Landuse and Landcover, other toposheet elements
6.	1:50000/ 1:25000 geology maps and reports from the Geological Survey of India		Geology map (preferably on 1:25000 scale)
7.	Rainfall data (from rainfall stations maintained by Indian Meteorological Department) and other rainfall data (from tea estates and state government offices), if available		1. Daily record of rainfall of various rainfall stations for the last 30 to 40 years, or whatever available.
8.	Census India		Specific information on details of houses, inhabitants etc. (for risk analysis)
9.	Fieldwork		1. Collection of GCPs through DGPS survey for preparation of DEM and Orthophotos from CARTOSAT -1 data. 2. Augmentation of landslide inventory data through interviews and site visits. 3. Collection of field data on causes of landslides of the area and landslide processes **(development of knowledge base)**. 4. Attempt to prepare spatial prediction map through direct assessment at field, so that clear concept on direct causes of

			slide be assessed through ground survey **(development of knowledge base)**. 5. Identification of training and testing datasets involving both failed and unfailed slopes.

Table -1: Nature and type of input data required for landslide hazard and risk prediction

6.3. Geospatial and statistical database preparation:

6.3.1. Landslide inventory database: After data capture and conceptual knowledge development, the tasks of geospatial and statistical database preparation are to be undertaken in a suitable GIS environment by applying the conceptual knowledge already developed during pre- and syn- data acquisition stages. The most important and fundamental step in this regard is the physical classification of the available temporal landslide inventory database under GIS (both spatial and temporal) into different groups or themes as per the following attributes: **a)** type of slide and material involved, **b)** spatial location, dimension (area, volume etc.) of initiation zone and zone of deposition, **c)** status of activity (active, stable, dormant etc.), **d)** depth and **e)** possible time of occurrence etc.. The proposed landslide inventory would be prepared after systematic interpretation of all the available high resolution past aerial photographs, temporal satellite imageries (merged LISS and PAN of different periods, Cartosat-1 imagery etc.), old inventory maps and old catalogues and reports etc. available in the study area following the standard methologies prevalent in literature (Cardinali et al., 2002; Guzzetti et al., 1994; Guzzetti et al., 2005; Malamud et al., 2004). The main aim of preparation of the landslide inventory would be **a)** preparation of a detailed ***geomorphological inventory***, which would contain all the different types of slide events over the entire period of study (e.g. 30-40 years) and **b)** to attempt to prepare separate ***event-based landslide*** inventories, considering at least 2 or 3 different landslide events in the study as per availability of input data. While preparing the above landslide inventories, knowledge on the morphological appearances of different landslide types (Cruden and Varnes, 1996), local lithological and structural setting and spatial and temporal evolution of individual landslide would be utilised (Guzzetti et al., 2005). The purpose of preparation of both the geomorphological as well as event-based landslide inventories based on the above parameters are to **i)** attempt to establish spatial evolution of landslides caused by different events through comparison of different inventories so that a rationale for the study area can be established that 'where landslide may cause damage in the future based on where landslides have already occurred in the past', which becomes an important parameter for hazard and risk estimation, **ii)** to classify the inventory into different spatial and temporal datasets (training and testing) for using the same for susceptibility prediction and validation, **iii)** to ascertain the size-frequency statistics of past landslides, which is extremely important for assessing the probability on the temporal and magnitude aspects of future landslide.

6.3.2. Rainfall database: For determination of rainfall thresholds, a continuous daily rainfall database for certain period (say 30-40 years or more as per availability of data) should be prepared in such a way so that any further statistical analysis such as daily vs. antecedent rainfall data (e.g. 3-day antecedent, 5-day antecedent, 10-day antecedents etc.) with this basic rainfall data can be carried out later in the data integration phase. During this rainfall data preparation, another aspect is also kept in mind, to determine the effect of the spatial influence zones of respective rainfall stations in the study area. During this stage mostly, the average mean annual rainfall for a certain period (for the period, database is available) would be determined, which in turn can be used to prepare average annual rainfall isohyets maps for using it as a dynamic predictor variable for spatial prediction. Apart from this, the rainfall database should also be made ready in such a way so that the rainfall data is continuous in nature, spatially representative for the study area and the computed extreme annual rainfall peaks in general are correlatable with the known landslide events in the area.

6.3.3. Geofactor database: Another most important task during this stage is to prepare the geospatial database of pre-disposing geofactors. These include all the relevant static/ quasi-static (e.g. lithology, geomorphology, structure, slope derivatives, landuse and land cover, soil characteristics, soil depth etc.) and dynamic (rainfall intensity, found most relevant to trigger a particular type of slide) variables and their possible combinations/ groups specific to different slide types. The specific combinations of geofactors in relation to the specific slide types are finalised using detailed conceptual knowledge base on the prevalent landslide processes in the study area developed during the data collection stage (through sufficient exposure to field work and RS data products). In this stage, classification of factors into continuous and categorical geofactor classes will be made in such a way so that the classifications are meaningful and can be related to the specific target types (e.g. landslide type) after suitably explaining the specific landslide processes. During this stage, various data-dependent techniques may also be adopted, so that the required statistical data treatments can simultaneously guide us to select a reasoned and statistically significant geospatial data especially in situations where knowledge are not adequate. Moreover, the above-mentioned statistical data treatment also can be used for quantitative corroboration of the knowledge base interactively used in this stage. The following statistical data treatments generally used in this stage facilitate us **i)** to reject the highly correlatable variables, since presence of highly correlatable independent variables in population lessen the performance of any data integration modelling techniques, **ii)** allow us to weed out statistically insignificant data and check the preference of variables and **iii)** can enable us to draw some preliminary weights/ ranks of predictor variables. Some examples of such statistical data treatments are **a)** test for normal distribution (K-S test), **b)** Principle component analysis (PCA) **c)** statistical significance test (T-test) and **d)** preference of variable (F-value test), **e)** bivariate analysis for preliminary weight assessment (Baeza and Corominas, 2001) etc. The tasks as described above under the geospatial database preparation stage can be schematically illustrated by the following flow diagram (Figure - 3).

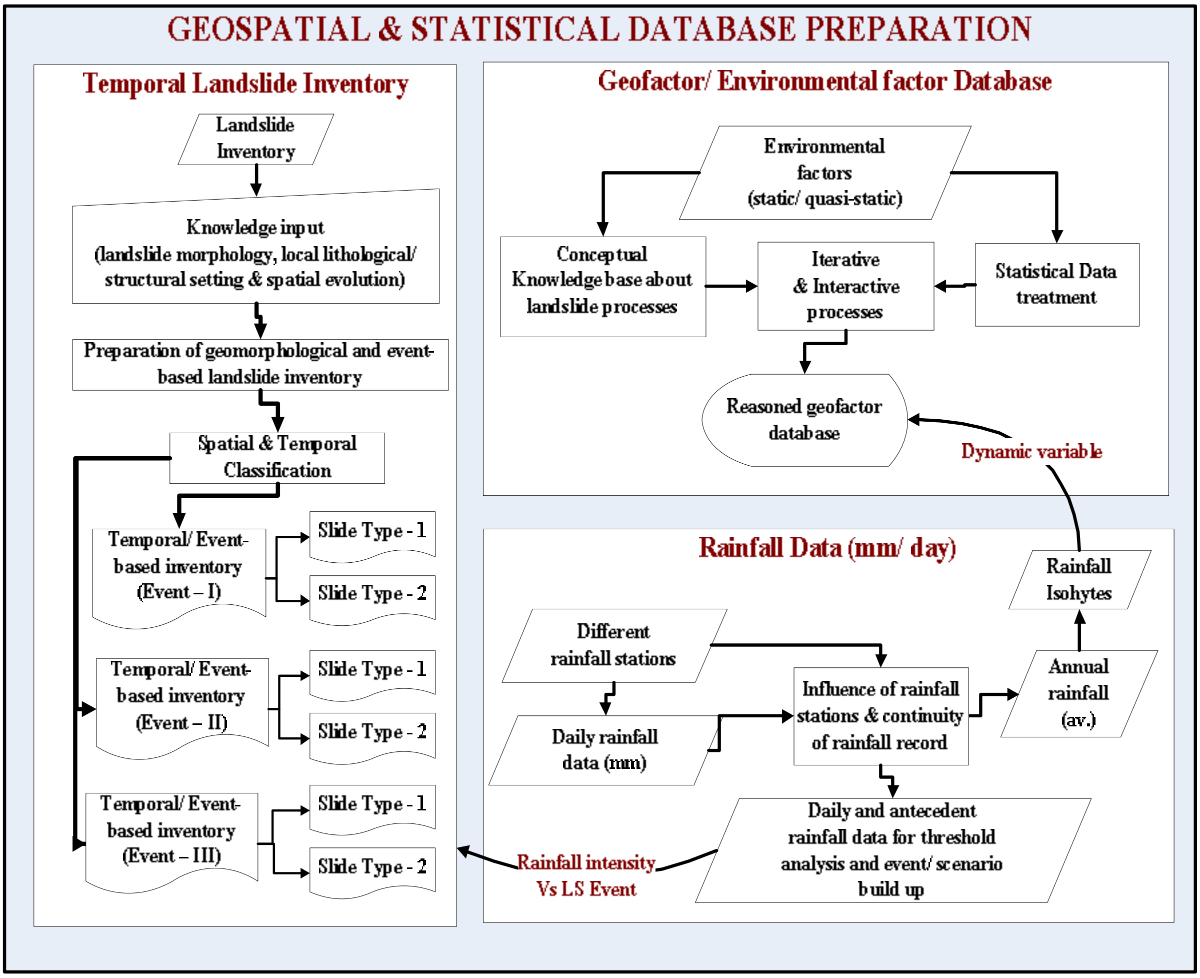

Figure - 3: Flow diagram showing geospatial & statistical data preparation stages

6.4. Data integration: Since the ultimate aim of the present study is to develop methodologies for the spatio-temporal hazard and risk analysis, identification of temporal scenarios or return periods of specific intensity of rainfall event and the frequency-magnitude relations of temporal landslide data are extremely important. As described before, probability of hazard in truest sense means the quantitative information about the expected number and expected size of future landslides of a specific type which may occur within a specific period of time (Zezere et al., 2004) at a specific location. Thus, for this analysis, in the data integration stage, the broad processes to be carried out are **a)** determination of spatial probability, **b)** temporal probability assessment through analysis of rainfall data and determination of rainfall threshold or scenarios or return periods and c) frequency-magnitude analysis of available temporal landslide inventory data for determination on magnitude of future landslides. To attain all the above stages, the following tasks are outlined.

6.4.1. Identification and delineation of suitable mapping unit: One of the most important tasks in the initial phases of data integration stage is to identify and delineate the size and distribution of individual mapping units. By definition, each mapping unit contains a set of ground conditions which differ from the neighbouring units (Hansen, 1984). The entire data integration process in subsequent stages of analysis corresponds to the identified mapping units, which thus represent as spatially homogeneous domains. Here, in the present research, attempts would be made to classify the terrains into three different types of mapping units, so that better applicability of the relevant mapping unit can be tested later after validating the results of data integration using all the mapping unit types. The three mapping unit types as proposed are **i)** grid cell of 10 m X 10 m size derived from Carto DEM for subsequent raster GIS analysis, **ii)** Unique condition units by overlaying the relevant geofactor combination layers (Chung et al., 1995) specific to identified slide types and **iii)** segregation of the terrain into main slope units through semi-automatic extraction using DEMs (Carrara, 1989) or by combining slope and aspect information directly derived from DEMs in the form of slope facets under GIS (Surendranath et al., 2006). In case of all the above three mapping units, the minimum size of individual unit should be determined by observing the minimum dimension of the available landslide incidences mapped in the inventory database. Though, it has been established by Carrara *et al.* (1992) that segregation of the area into slope units can reduce the extent of errors while landslide bodies are aggregated since landslide processes are highly controlled by the slope features of the terrain, but the present research intends to leave a scope to test the utilities of other mapping unit types vis-à-vis the 'slope unit' or 'slope facet' in the study area.

6.4.2. Spatial prediction: The crucial aspect of any landslide hazard analysis is to assess the spatial prediction. The spatial prediction answers the aspect of "where" the future landslide should occur. In the proposed research project, for spatial prediction, both the logistic regression modelling and artificial neural network techniques are preferred as data-dependent models (Figure – 4). The above two quantitative techniques were primarily chosen because of the following premises that **a)** despite several attempts, data uncertainty can not be ruled out in the input database, **b)** multivariate models and ANN give better and robust prediction in comparison to the bivariate models since the chosen two models can handle both continuous as well as categorical data sets and also can handle the mutually dependent predictor variables.

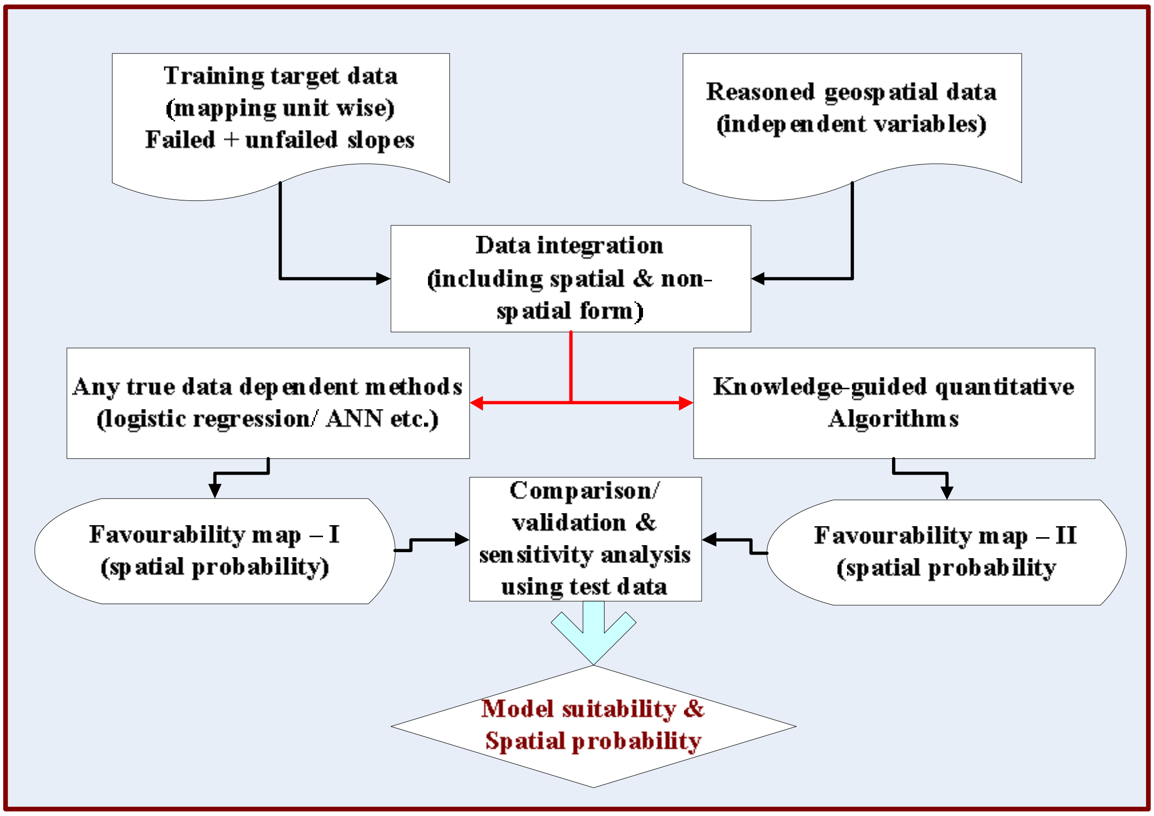

Figure – 4: Schematic flow diagram showing methods to be followed for spatial prediction

In ***logistic regressions***, following the methods adopted by several workers (Chung and Fabbri, 1999; Dai and Lee, 2003; Guzzetti et al., 2005), the probability of landslide occurrence in terms of independent variables $X_1, X_2, \ldots, X_n$ can be expressed as

$$P(Y=1) = \frac{1}{1+e^{-(\alpha+\beta_1 X_1+\ldots+\beta_n X_n)}}$$

Where, β_i = $(i = 1, 2, \ldots, n)$ are the co-efficient estimated from the sample data and α is the intercept. $P(Y=1)$ is the probability of future landslide occurrence in any mapping unit. The parameters of the logistic regression model are estimated through *maximum likelihood* method that is the co-efficient which make the observed results more "likely" are selected. Since the relationship of independent predictor variables and the landslide susceptibility is non-linear, an iterative algorithm is used for parameter estimation. For dependent variable, training data comprising equal number of failed (Y = 1) and unfailed (Y = 0) slopes are considered for this analysis.

In ***artificial neural network (ANN),*** the basic and fundamental element of a neural network is the processing node. The weighted input data is then passed through an activation function to produce the node's output value. A neural network consists of a number of interconnected nodes. Each node is a simple processing element that responds to the weighted inputs it received from other nodes. The corresponding weight, called the bias weight, effectively controls the threshold level of the activation function. The processing nodes are organized into layers, each generally fully interconnected to the following layer. The final processing layer is called the output layer. Any layers in between the input and output layers are termed hidden layers**.** There are two stages involved in using neural networks for a multisource classification; the *training stage* and the *classification* stage. The back propagation algorithm trains the network, typically, until some targeted minimal error is achieved between the desired and actual output values of the network. If still, the desired RMSE goal is not achieved, separate training data set may be used**.** The rating and weight calculated using the above probability and neural network methods are then used for landslide susceptibility analysis (Lee et al., 2003).

In the present research topic, maximum stresses were given to prepare and select a reasoned geospatial database for both the independent as well as dependent variables of landslide prediction. Yet, during actual data integration stage in future, incorporation of some priori or expert knowledge may still be required to improve further the model performance, then separate algorithmic modifications would be tried to supervise the learning process of the relevant quantitative modelling technique by introducing either fuzzy-based concepts or expert weights (e.g. Candidate's method to be developed in course of research). The provision of this is pertinent in the research proposal since as per the proposed research hypothesis, it would always be corroborative if a true data-dependent modelling technique is validated simultaneously against a knowledge-guided quantitative algorithm developed in the proposed study area.

6.4.3. Validation and Sensitivity Analysis of susceptibility models: Any type of quantitative prediction model requires rigorous tests for its reliability, degree of fitting and robustness in skill of its prediction (both space and time). In the proposed research project, for the spatial prediction models, attempts would be made to carry out quantitatively all the possible types of validation and testing. The "goodness of fit" would be tested to find how well the prediction image has classified the mapping units of the study area. The performance of any quantitative prediction model would be validated through "success rate" and "prediction rate" curves as proposed by Chung and Fabbri (2003). For success and prediction rate calculations, time, space and random partitioning of the target data are proposed for the study area for establishing both the time and space robustness of models.

Apart from the above, the model reliability or "goodness of fit" can be tested by creating several susceptibility models by random partitioning of different training data ensembles from the entire study area as proposed by Guzzetti *et al.* (2006). Within the different susceptibility models, the role of different independent variables can also be quantitatively outlined/ tested, which in turn can directly corroborate the strengths/ weaknesses of the proposition of selection of knowledge-guided predictor variables in the parameters. Through the above type of sensitivity analysis, Guzzetti *et al.* (2006) also proposed a method to quantitatively estimate the error associated with the probability of landslide spatial occurrence and also defined a criteria to rank the quality of a particular susceptibility method, which can also be applied in the proposed study area.

6.4.4. Determination of temporal probability or return periods and rainfall thresholds:

Since it has been assumed that all the landslide events in the study area were triggered by rainfall, the statistical analysis of the available daily rainfall data would enable determination of triggering thresholds and calculations of recurrence interval (Aleotti and Chowdhury, 1999). There are various methods through which rainfall thresholds can be determined empirically. Considering the resolution of rainfall data, it has been decided that in the proposed study area, rainfall thresholds would be attempted based on antecedent rainfall account. For reconstruction of such rainfall events, for example, cummulative rains from 1 to 30 days in the forms of various antecedents (1 day, 2 days, 3 days, 5 days, 10 days, 15 days, 30 days etc.) prior to the date or period of identified landslide events would be considered. Here the analysis is being proposed to link different rainfall duration-intensity data to different slide types (such as shallow slides and deep-seated slides), which are related to an event and can be obtained from the inventory database. Some of the rainfall thresholds based on similar analysis of antecedent rainfall data as proposed by some early workers (Chelborad, 2003; Kim et al., 1991) can be illustrated by the following figure (Figure - 5):

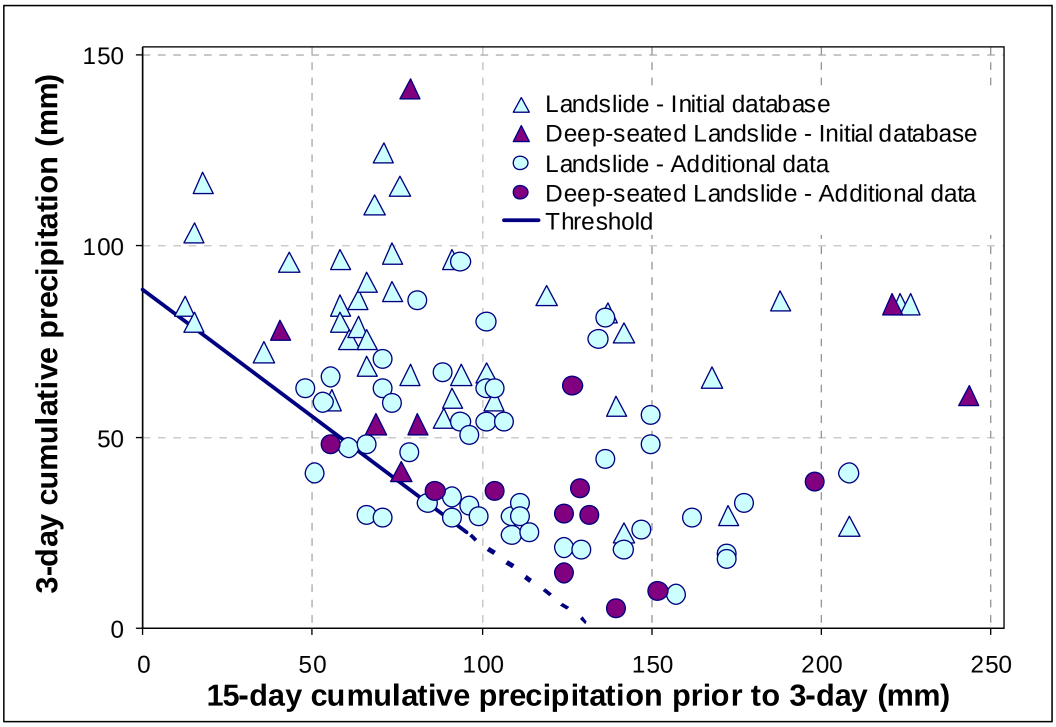

Figure - 5 = 3-day and prior 15-day cumulative precipitation associated with historical landslides that were part of events with 3 or more landslides in a 3-day period, in Seattle. from Chleborad (2003).

The reconstruction of such type of intensity-duration of rainfall and known landslide event would enable us to identify the thresholds. Considering the above theoretical assumptions of rainfall threshold analysis, the following methodology is being proposed for the study area broadly adopting the methodologies used by Zezere *et al.* (2005) in Lisbon area.

1. From the landslide event inventory database, all the landslide events, preferably date-wise should be identified and this reconstruction should also include the particular landslide types and their sizes associated with each landslide events.

2. From the daily record of past rainfall data of the representative rainfall stations of the area, the cummulative absolute antecedent rainfall for say 1 day, 3 days, 5 days, 10 days, 15 days, 30 days etc. prior to dates of confirmed landslide activity during the period of study should be identified. During this process, as explained earlier the critical rainfall combinations

(quantity-duration) responsible for each landslide event should be assessed (identification of thresholds).

3. To understand the overall role of precipitation, the rainfall intensity (mm/ day) and the critical rainfall duration (days) of all the identified landslide events of the study area can be plotted and a regression equation can be developed as proposed by Zezere et al. (2005), which can give the rainfall threshold curve of the study area (Figure - 6).

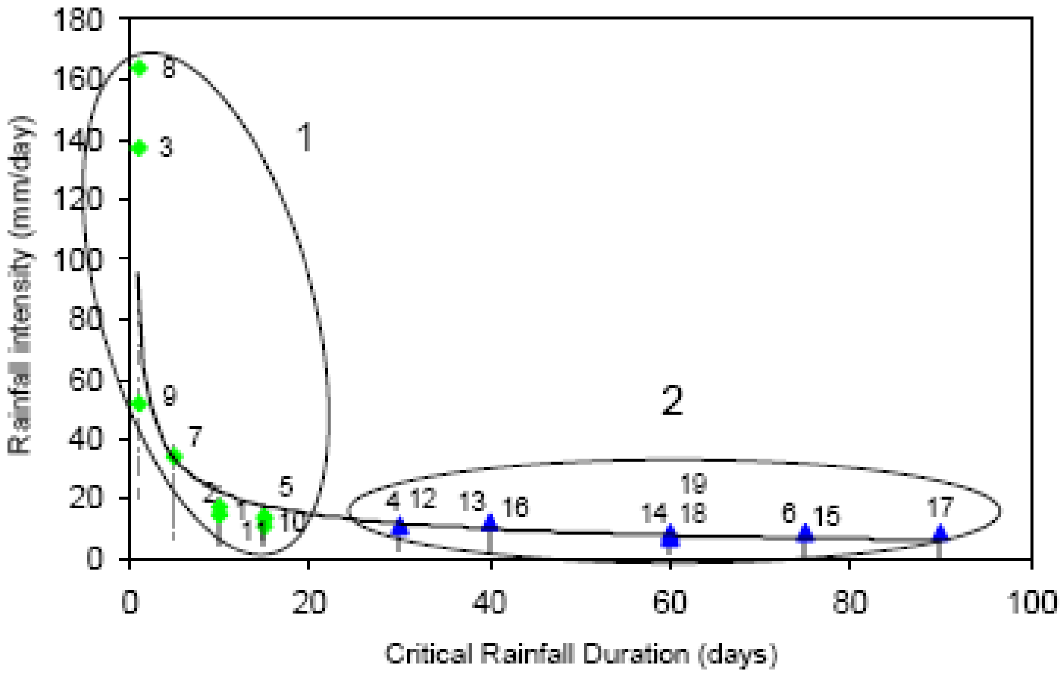

Figure - 6: Rainfall intensity-duration plot for shallow and deep-seated landslide (Zezere *et al., 2005)*

4. The return periods of all the rainfall amount-duration combinations, which had triggered the particular slide event can be computed by using Gumbel distribution, which will give the probability of temporal occurrence of that particular event (Figure – 7).

The overall procedure as described above can be illustrated by the following flow diagram (Figure – 7).

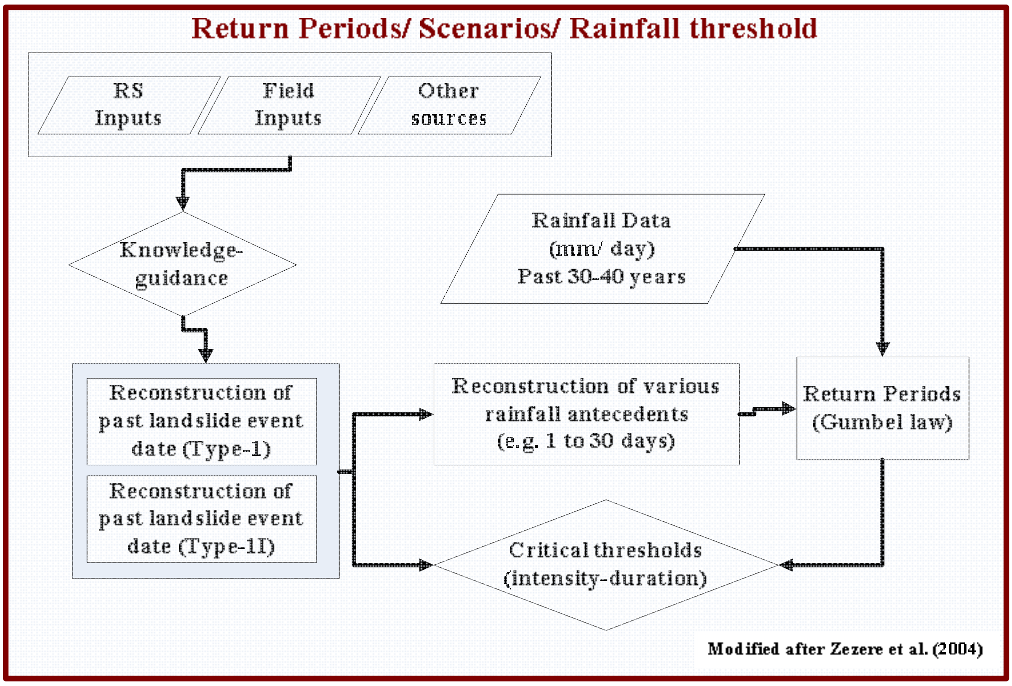

Figure - 7 – Flow diagram showing the proposed rainfall-threshold analysis

6.4.5. Determination of temporal probability from analysis of landslide frequency data:

The temporal probability using landslide frequency estimates through determination of exceedance probability of one or more landslide can also be attempted assuming landslides as random point events. There are two different discrete probability distribution models – the '*Poisson distribution'* and the '*Binomial distribution'*, which are extensively used for such exceedance probability calculations of landslides (Coe et al., 2004; Crovelli, 2000). As per Poisson model, the occurrence of landslides that is experiencing 'n' landslides during time't' is given by

$$P\,[NL\,(t) = n] = e^{(-\lambda t)} * \frac{(\lambda t)^n}{n!} \qquad \text{where } \lambda = \text{average rate of occurrence of landslide}$$

$$n = 0, 1, 2, 3\ldots \ldots, n$$

Thus, the exceedance probability or the probability of experiencing one or more landslides during time't' would then be

$$P\,[\,N_L\,(t) \geq= 1\,] = 1 - P\,[\,N_L\,(t) = 0\,] \;= 1 - e^{-\lambda t} \;= 1 - e^{\frac{t}{\mu}}$$

where, $\mu = \frac{1}{\lambda}$ and μ = mean recurrence interval between successive failure events

Following the binomial probability distribution model, the similar exceedance probability can be calculated through following expression:

$$P\,[NL\,(t) \geq= 1] = 1 - P\,[NL\,(t) = 0] = 1 - (1\text{-}P)^{t} \qquad = 1 - (1 - \frac{1}{\mu})^{t}$$

where, P is the estimated probability of a landslide event in time t, and $\mu = 1/p$ is the estimated mean recurrence interval between successive landslide events. The parameter μ can be obtained from the historical landslide inventory data of the study area.

6.4.6. Determination of probability of landslide size: It has been observed by many earlier workers (Hovius et al., 1997; Malamud et al., 2004) that irrespective of space, time and type of slope instability and their triggering mechanisms, landslide frequency increases with increase in landslide area up to a maximum value and then it rapidly decays down along power law (Figure - 8). Analyzing the above generic distribution pattern of landslide size and frequency, the probability of future landslides to occur can be determined through analyzing the historic landslide inventory database (Guzzetti et al., 2005). This determination of probability of landslide size is essential to infer "how large" the future landslide would be, which is an essential parameter for hazard estimation of an area. Thus, the probability estimation of area of future landslide having an area of a_L or greater can be expressed as

$$P\,A_L = P\,[AL \geq a_L]$$

The above probability expression can be solved through analysis of area-frequency data of all past landslides available in the landslide inventories of the study area. For hazard estimation, in this analysis, area is being considered as a proxy for landslide magnitude. For the above analysis, a sufficiently complete landslide inventory database encompassing all possible size fractions are required. Malamud *et al.* (2004) analysed sufficiently complete set of three landslide inventories (Central Italy, California, USA and Guatemala), which were caused by different triggering mechanisms and found that the probability density function (PDF) – p of landslide area A_L can be expressed as

$$p\,(A_L;\,\rho, a, s) = \frac{1}{a\Gamma(\rho)}\left[\frac{a}{A_L - s}\right]^{\rho+1} \exp\left[-\frac{a}{A_L - s}\right]$$

where, $\Gamma(\rho)$ is the gamma function of ρ and $\rho > 0$, $a > 0$, and $s \leq A_L < \infty$ are the parameter of the above distribution. In the above equation ρ controls the power law decay of distribution for medium and large landslides, a determines the location of maximum PDF distribution and s controls the exponential roll over of small landslide sizes. Using the above expression of PDF, the probability of future landslide size PA_L can be determined as:

$$PA_L = \int_{a_L}^{\infty} p(A_L;\rho,a,s)\,dA_L = \int_{a_L}^{\infty} \frac{1}{a\Gamma(\rho)}\left[\frac{a}{A_L - s}\right]^{\rho+1} \exp\left[-\frac{a}{A_L - s}\right]$$

In similar line, Stark and Hovius (2001), after analyzing landslide inventory dataset of New Zealand and Taiwan, explained that the PDF of landslide area in their case was found to be in good agreement with the double Pareto probability distribution and they gave another set of equation for PA_L, which is as follows:

$$PA_L = \int_{a_L}^{\infty} p(A_L;\rho,a,s)\,dA_L = \int_{a_L}^{\infty} \frac{\beta}{l(1-\delta)}\left[\frac{\left[1+(m/l)^{-\alpha}\right]^{\frac{\beta}{\alpha}}}{\left[1+(A_L/l)^{-\alpha}\right]^{1+(\beta/\alpha)}}\right] \mathrm{X}\ (A_L/l)^{-(\alpha+1)}\,dA_L$$

where, $\alpha > 0$, $\beta > 0$, $0 \leq c \leq l \leq m \leq \infty$ and $\delta = \left[\frac{1+(m/l)^{-\alpha}}{1+(A_L/l)^{-\alpha}}\right]^{\frac{\beta}{\alpha}}$

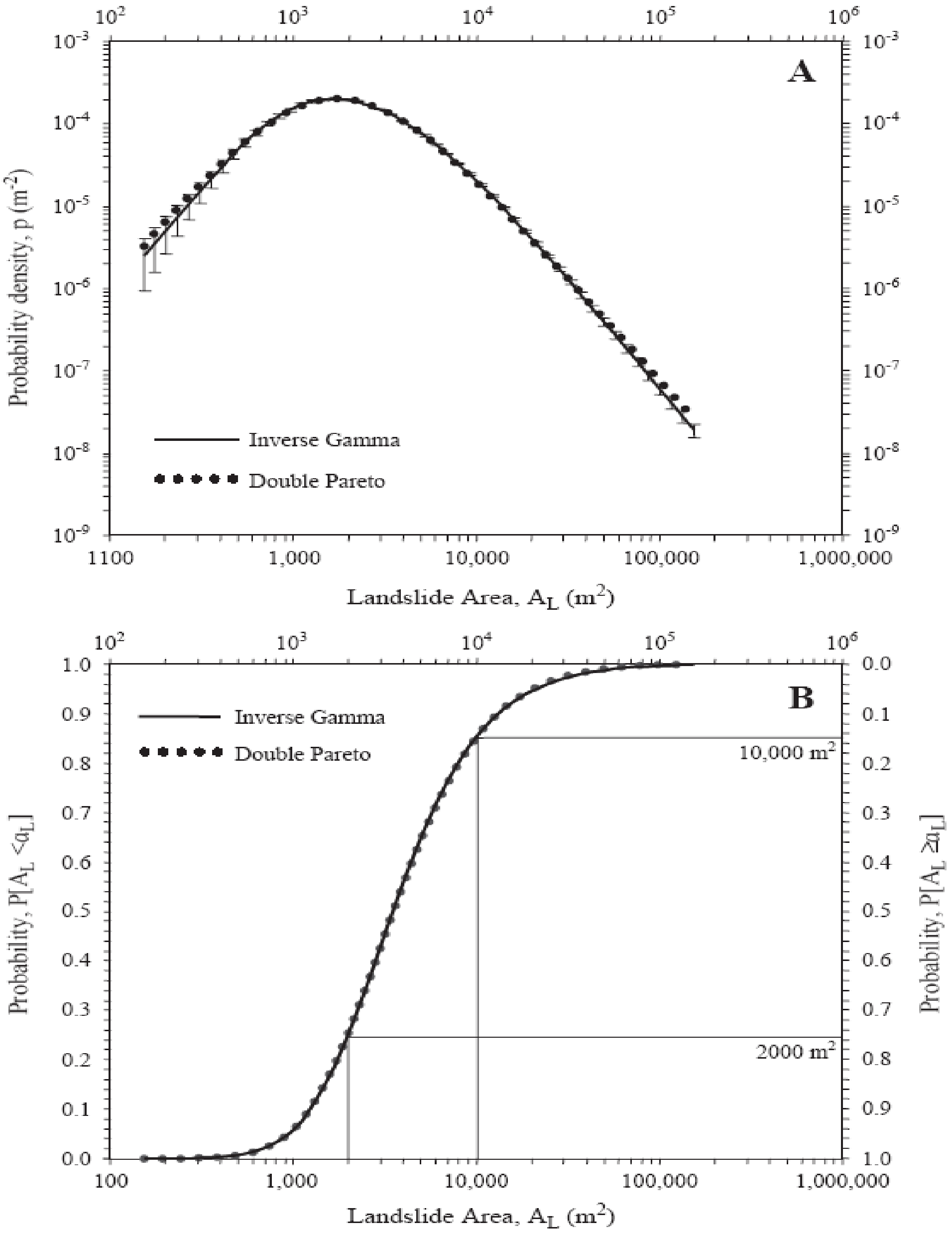

Figure - 8 = Probability density (PDF) and Probability of landslide area plot (The solid line represents the equation of truncated inverse gamma function proposed by Malamud *et al.* (2004) and dotted line is represented by a double Pareto distribution (Stark & Hovius, 2001)

7.0. Proposed methodology of for landslide risk analysis:

To mitigate the ever-increasing physical conflict between the society and natural hazards, availability of risk maps is becoming a strong decision-making tool to the planners and decision-makers. For accomplishing the above task, a framework of risk analysis for a small type area has been proposed in this research project using the medium scale spatio-temporal information derived through the preceding steps. Risk as a whole is a complete phenomenon which is known as *Risk Management*, which encompasses risk analysis, risk evaluation and risk treatment part. The risk evaluation and subsequent treatment part solely depend on the risk analysis, and the latter has a direct link with the hazard evaluation exercise, that is why risk analysis is only being attempted within the scope of current research topic.

For calculation of risk, a hazard-consequence matrix approach can be followed (Chowdhury and Flentje, 2003). Following the above method, a set of different hazard categories (hazard scenarios depending on various magnitudes, return periods, type etc.) can be combined with a set of consequences and several risks (specific risks) can be calculated. Combining all the above specific risks, the total risk of an area is estimated. Again consequence assessment can be separately made on 'damage to properties (economic loss)' and 'loss of life' and there may be two risk assessment matrices. To quantitatively explain the calculation of annual probability in case of "loss of life (P_{life})" and "damage to property (P_{prop})", the following probabilities are to be determined (Morgan, 1992).

$$P_{life} = P(H) * P(S/H) * P(T/S) * P(V/L)$$
$$P_{prop} = P(H) * P(S/H) * P(V/S)$$

where, P (H) = annual probability of a particular hazard scenario

P(S/H) = Conditional probability of spatial impact of a landslide hazard to a particular element at risk (e.g. / building), which depends on run out or travel distance

P(T/S) = Conditional probability of temporal impact (probability of the building occupied by the people at the time of impact)

P(V/L) = the vulnerability (degree of loss of life) of a person given landslide impact

P(V/S) = the vulnerability (degree of damage) of a property given landslide impact

For calculating specific risk to property, the above annual probability (P_{prop}) are multiplied with the cost of elements at risk (E) in case of property and in case of "loss of life" it is expressed as 'annual probability of 'loss of life' (P_{life}) to a person under the impact of a given hazard scenario. For calculation of total risk, all the specific risks pertaining to each hazard scenario are summed up together and a total annualized risk can be estimated (Australian Geomechanics Society and Sub-committee on landslide risk management, 2000).

Considering the above mathematical expression, it is apparent that the parameters such as P(H) will come from hazard information whereas P (T/S) and E can be obtained through elements at risk mapping. In the proposed research project, mapping of elements at risk will be carried out from the orthomap prepared from stereo data of Cartosat-1 (2.5m resolution) and supplemented by Census information and field work at some test locations. For successful completion of risk estimation, still determination of two probabilities which are related to i) travel distance or run out [P(S/H)] and ii) vulnerability [P(V/S) or P(V/L)] are to be assessed during the consequence analysis.

7.1. Estimation of travel distance or Run out (L): The travel distance (L) can be defined as the horizontal projection of the line joining the upper part of the landslide source or initiation zone to the outermost edge of the landslide accumulation zone (Hungr et al., 2005). Finley *et al.* (1999) using multiple regression analyses proposed a number of mathematical expressions (Hungr et al., 2005) for determining travel distances in cut slopes, fills, retaining walls and boulder falls after studying a number of debris slides in Hong Kong.

Hungr *et al.* (2005) also proposed that when a landslide source and potential landslide volume are available, the travel distance can also be empirically obtained by the following expression

$$L = \frac{H}{\tan\alpha}$$; where, H is the vertical drop and α is the reach angle that is tan (*H/L)*.

Similarly, many authors also proposed various other mathematical expressions (Log tan (H/L) = a + b Log V) based on inverse relationship between the tangent of reach angle (H/L) and the landslide volume but in various such regression equations, correlation co-efficient of relationship between the two variables (reach angle and volume) are too weak to be used. In order to improve the empirical relationships, Corominas (1996) attempted to develop regression equations after classifying the volumes of different slide type and could able to reach better mathematical equations using Log tan(H/L) and V (volume) as variables (R^2 ranging from 0.65 to 0.91). All the above empirical methods proposed are simple in approach for calculation of travel distance in GIS environment but it should always be noted that the assumptions implicit in these methods are imprecise and statistical scatter is very large and also they do not portray the kinematic parameters, which could be pivotal during run out processes (Hungr et al., 2005). Similar empirical approaches based on different slide types are proposed for calculation of run out in the proposed research project.

7.2. Estimation of Vulnerability: Assessment of vulnerability is one of the most important and possibly the most difficult task of the consequence analysis in risk estimation. Vulnerability assessment becomes a difficult proposition in cases where no data on past damages are available. Though, assessment of vulnerability in terms of damage functions have been tried by some workers and those can be established for different vulnerable elements. For social vulnerability, Roberds (2005) after working with past landslide data of Hong Kong proposed that potential loss of life (*PLL,* which is the mean of the probability distribution of the number of public fatalities) for various landslide cases can be determined through the following probabilistic expressions:

$$PLL = \{PLL\ for\ ref\ LS\}*\{SF\}*\{PF\}$$

Where, *PLL for ref LS* is the PLL for reference landslide (say 10 m wide and 50 m^3 of volume, based on past landslide inventory data, which has been proposed by Ho *et al.,* (2000) for different landuse types in Hong Kong), SF (for scale factor) is the ratio of width of debris in landslide to the reference landslide i.e. 10 m, PF (proximity factor) is the function of debris mobility (run out) vs. elements at risk location. Roberds (2005) also tried to explain damage functions of static element at risk such as building, tower etc. through examples of spatial intersection. For such physical vulnerability assessment (degree of loss or damage to structure), without the availability of past damage profiles depending on different types of structures, assessment is difficult. Theoretically, a simple empirical physical vulnerability assessment can be made so that if an element at risk is spatially intersected by the travel path of a landslide, vulnerability, depending on the construction type can qualitatively be assessed as ‘completely damaged’ (V=1) to ‘no damage’ (V=0).

8.0. Proposed study area

The proposed study area in Darjeeling district, West Bengal, India is within the eastern part of the Himalayan tract (Figure - 9) which comprises intra-thrusted rock slices of Precambrian to Quaternary ages of the Himalayan Fold Thrust Belt (FTB). The eastern Himalayan tract represents a complicated geological and tectonic milieu, in which rocks of varying ages and metamorphic suite are juxtaposed along certain E-W trending Tertiary regional thrusts. Along the foothills to the south, coarse to very coarse-grained clastics (conglomerate-sandstone-siltstone) of the Siwalik Group of Tertiary age are exposed and are separated from the adjoining Quarternary sediments of the foredeep region further to the south by a frontal thrust (Himalayan Foothill Thrust – HFT or Himalayan Frontal Thrust HFT). The coarser clastics of the Siwalik Group towards the north are thrusted over by sandstone-shale (± coal) sequence of the Gondwanas (Mesozoic) along the Main Boundary Thrust (MBT). Further to the north, low grade meta-psammo-pelitic lithoassemblages of the Precambrian Daling Group are thrusted over the Younger Gondwana / Siwalik sediments (Figure – 10). Further north in the Higher Himalaya, granite gneisses and high-grade meta-sediments belonging to the Central Crystalline Gneissic Complex (CCGC) are thrusted over the low-grade metamorphics of the Daling Group along the Main Central Thrust (MCT). Along the lower part of MCT, a strongly lineated, coarse to medium grained granite gneiss and granite mylonites (Lingtse gneiss) in the form of sheets are conspicuously disposed as thrust wedges (Pawde and Saha, 1982; Raina and Srivastava, 1981).

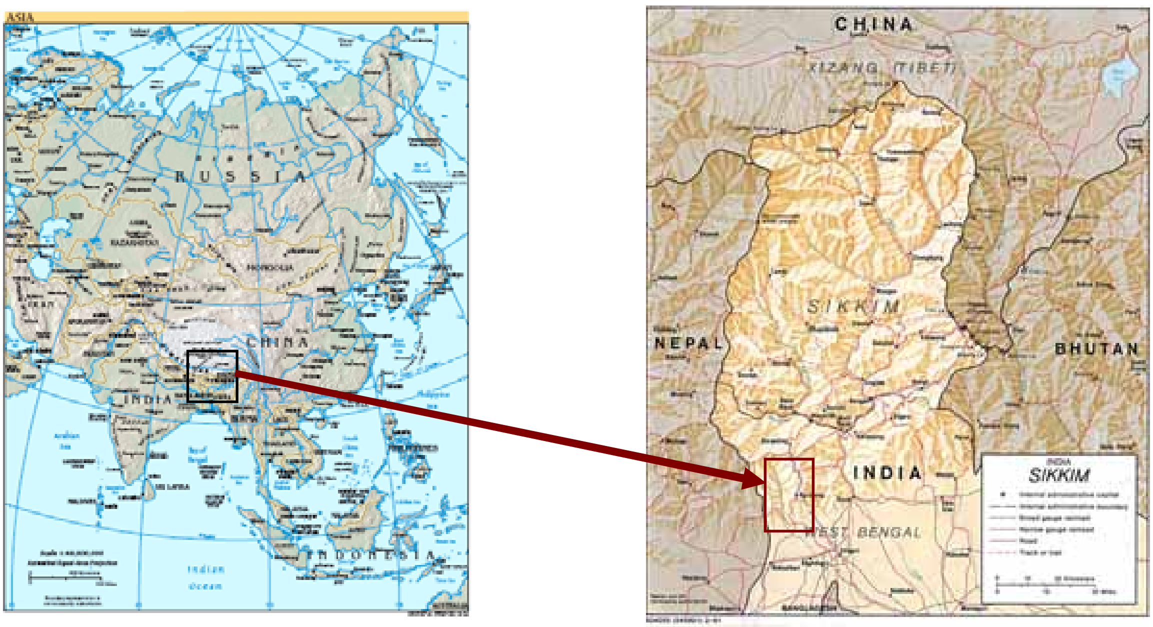

Figure 9 – Location Map of the Study Area

(Source: https://www.cia.gov/cia/publications/factbook/reference_maps/pdf/asia.pdf; http://www.lib.utexas.edu/maps/middle_east_and_asia/sikkim.jpg; http://en.wikipedia.org/wiki/Image:India-locator-map-blank.svg)

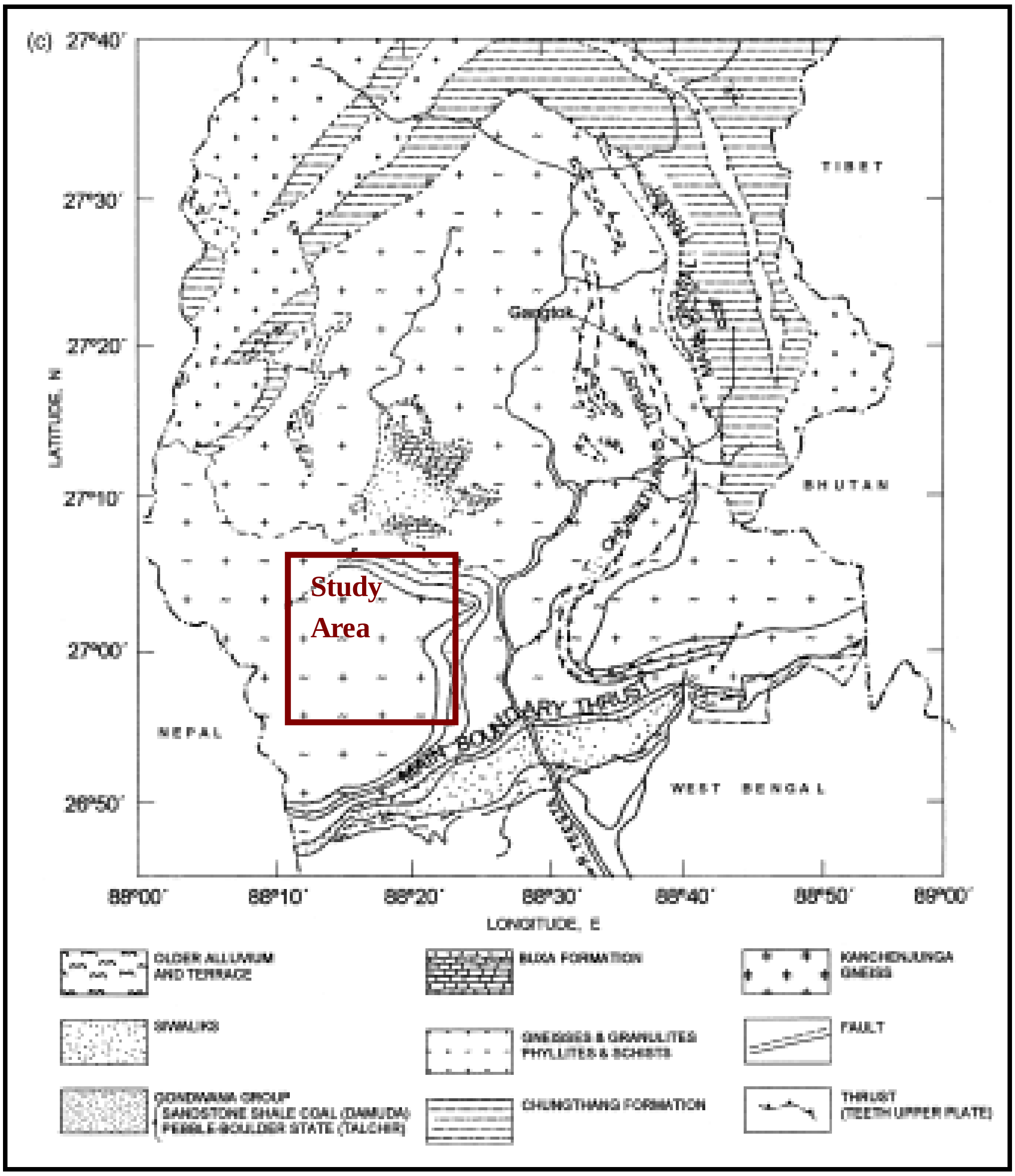

Figure – 10 Geological map of Sikkim-Darjeeling Himalayas
-After Raina & Srivastava (1981)

The overall relief difference in the studied area varies from 250 m to as high as 2650 m. The general trend of the mountain ranges is east -west. A number of NE-SW and NW-SE trending ridges and spurs are carved out of this trend and form high mountain ranges. The average annual rainfall in Darjeeling Himalaya to the west of the Tista River fluctuates between 2000 mm and 4000 mm, although locally and in exceptional years, it passes to 5000 mm. The two-year recurrence interval of rainfall event in Darjeeling is calculated at 2735 mm and the 50-year recurrence interval is at 4178 mm (Soja and Starkel, 2007).

Landsliding triggered due to excessive rate of rainfall are common in this terrain. The entire Darjeeling Himalaya is studded with numerous landslides of variable dimensions, types and morphology. The major communication networks of Darjeeling district such as NH-31A, NH-55 and SH-12A, which are the only lifelines of this mountainous terrain, are severely affected by some of those perennial landslides. Some of them are highly active, causing devastating natural hazards during every monsoon resulting frequent closure of the highways and damage to the townships, loss of life and property. Darjeeling district also acts as the only gateway to enter any parts of India's smallest border state, Sikkim. Thus, any hazard in Darjeeling Himalaya equally affects the access to Sikkim. In this Sikkim-Darjeeling Himalaya, apart from urban growth centres and tourist spots, several important hydro power installations are also located and a number of such projects in Tista valley are presently under construction, which definitely signifies an elevated level of landslide-related risks in those areas, which needs proper attention.

Available historical data indicates that the oldest recorded history of landslide in this area is of a devastating landslip in and around Darjeeling town on 24th September 1899 due to an unprecedented rainfall of 1065.50mm in that area, where 72 people lost their life and enormous amount of land and property was lost (Griesbach, 1900). Apart from that, the Darjeeling district has witnessed several prominent landslide events triggered by incessant rainfall, some of which are mentioned below in Table – 2 (Basu and De, 2003).

Location of event	Date/ Time of Occurrence	Loss/ Damage	Triggering factor
Darjeeling town and its surrounding areas	24 - 09 – 1899	72 people died along with huge loss of property and land	1065.50 mm rainfall in 3 days
Tindharia, Kalimpong, Darjeeling & Kurseong towns	15th January, 1934	Loss of property and land	Reported to be linked with Bihar-Nepal earthquake of 1934
Darjeeling, Kalimpong & Kurseong towns	11-13 June, 1950	127 people died with huge loss of land and property worth 648 Lakhs of rupees; hundreds became homeless; Siliguri-Kalimpong Railway line was closed for ever due to incessant instability along steep hill slope	Caused due to unprecedented continuous rainfall of 834.10mm in 3 days
Darjeeling, Kurseong, Giddapahar, Gayabari, Kalimpong, the entire stretch of NH-55 & NH-31A, adjoining districts of Sikkim	2-5th October, 1968	The most dreadful landslide and flood disaster of Darjeeling-Sikkim Himalaya; 677 official death; unofficial figure more than 1000; total stretch of NH-55 & NH - 31A either washed away or damaged; several bridges such as at Rangpoh, Tista bazaar on Tista river were washed away/ severely damaged; roads were closed for more than a year; as per Indian Tea association 10-15% of Tea cultivation were damaged by this event	Due to incessant rainfall of 1121.4mm within 4 days
Lodhama, Bijanbari, Darjeeling, Sonada, Sukhiapokhri, Kurseong, Paglajhora, Tindharia etc.	3rd to 4th September, 1980	215 people lost their life; loss of property worth 100 million rupees	Due to incessant rainfall of 299.1mm in 2 days.

Darjeeling, Tukvar, Bennockburn, Bloomfield, Paglajhora, Chunavati-Tindharia areas	15th & 16th September, 1991	2 people died and huge land and property got damaged; Darjeeling-Silguri Toy train tract was severely damaged for 5 months	Due to incessant rainfall of 462.5mm in 2 days.
Mungpoh, Takdah, Pesoke, Rongtong, Tindharia, Kurseong, Gayabari, Darjeeling etc.	11-13th July, 1993	15 people died; Properties of several crores were damaged	Due to heavy and concentrated rainfall of 211.3mm in 2 days
Mostly areas along NH-55 (Chunabhati, Tindharia, Paglajhora, Kurseong, Sonada, Darjeeling)	5th & 8th July, 1998	Severe damage and road blockades mainly along NH-55; most affected terrain is Kurseong and its surroundings.	300-600mm cumulative rainfall in 2/3 days caused these slides
Gayabari Tea Estate areas on SH-12A, parts of NH-55 near Kurseong, Tindharia etc.	6-8 July 2003	25 people died with loss of huge property and land, Tea Cultivation etc.	Incessant rainfall of about 500 mm in 2 days

Table – 2: An account of past major landslide events in Darjeeling Himalaya

For preparation of detailed landslide inventory and hazard analysis (1:25000), the parts of the Balason-Mahanadi-Rohini catchments and parts of left abutment of Tista covering hill settlements like Tindharia – Gayabari – Giddapahar – Kurseong – Sonada - Ghum – Darjeeling – Mirik - Sevok – Tista bazar sector (parts of Topographical sheet 78A/8, 78B/1 and 78B/5) and parts of the communication corridor - NH-55 cum UNESCO's World Heritage Himalayan Toy Train Tract, NH-31A and SH-12A are prima facie identified as study area. For detailed risk estimation and preparation of large scale risk maps a type area amongst any one of the following urban agglomeration centers such as Darjeeling, Kurseong, Mirik and Tindharia will be covered depending on the availability of suitable large scale imageries and other high resolution data products.

9.0. Proposed work plan, schedule and budget:

The proposed Ph.D. research programme will be carried out in 'sandwich' mode and the tentative schedule and work plan is enumerated in Table – 3.

Time period/ Place of study	Total duration of stay	Nature & quantum of work (tentative)
Phase – I (August, 2007 to April, 2008) **INDIA**	**8 months**	• Collection of aerial photographs/ CARTOSAT 1 stereo data/ multitemporal IRS PAN and LISS-IV data/ Rainfall data • Collection of ground control points (GCPs) through DGPS Survey (service from Earthquake Geology Division, GSI, E.R. or any other facilities in GSI having DGPS instrument and expertise will be utilised). • Generation of DEMs from Cartosat Stereo data and using GCPs (using LPS suite of ERDAS Imagine) and generation of orthophotos from DEMs (Services from the laboratories of PGRS, ER, CGMT, GSITI, Hyderabad and NRSA, Hyderabad will be required/ utilised). • Study of multi-temporal aerial photograph/ imagery data for multi-temporal landslide inventory (Service from PGRS, ER, GSITI, Hyderabad and NRSA, Hyderabad will be required/ utilised). • Field assessment of failed and unfailed slopes to develop a conceptual knowledge about different landslide types and its relation with combination of geofactors; ancillary data collection: rainfall, landslide incidences and damage (logistic supports from GSI, ER would be sought).
Phase – II (April, 2008 to October 2008) **ITC, The Netherlands**	**6 months**	• Preparation and finalisation of landslide inventory (geomorphological & event-based). • Preparation of relevant geofactor database (spatial and textural). • Analysis of rainfall data in search for rainfall event. • Magnitude - Frequency Distribution of past landslides. • Search for data gaps in the field of geofactors, landslide inventory, rainfall etc. • Writing of scientific papers; conference presentation, • Attending specialized courses

Phase – III (October, 2008 to April, 2009) **INDIA**	**6 months**	• Collection of additional data from field (temporal landslide inventory and temporal geofactor database) as per the gaps identified in the earlier step. • Collection of data pertaining to the risk elements mapped (Preparation of spatial and textural database); collection of census and other information pertaining to the risk elements. • Finalisation of all sorts of landslide inventories and geofactor databases after incorporation of additional data, if any. • Writing of scientific papers
Phase – IV (April, 2009 to October, 2009) **ITC, The Netherlands**	**6 months**	• To outline and develop the quantitative model of spatial prediction using the available knowledge and geospatial database (data integration for spatial prediction). • To attempt for temporal prediction using the spatial prediction map, available landslide inventory, rainfall data (data integration for temporal prediction). • To prepare the risk map after incorporating the information of consequence analysis (vulnerability and risk elements). • Writing of scientific papers. • Conference presentation
Phase – V (October, 2009 to March 2010) **INDIA**	**5 months**	• To test the developed prediction model at field and make final adjustments to model parameters and collection of relevant information. • Model validation and sensitivity analysis • Writing of scientific papers • Thesis chapter writing
Phase – VI (March, 2009 - September, 2010) **ITC, NL**	**6 months**	• Writing of scientific papers. • Preparation, finalisation and submission of draft of Ph.D. thesis for evaluation by examiners.

Table – 3: Tentative work plan and schedule

9.1. Research budget allocation: The proposed ITC research budget of Euro 10,000 for the proposed Ph.D. research programme has tentatively been allocated in the following different heads (Table - 4).

	Year 1	Year 2	Year3	Total
Conference		1000	1000	2000
Travel to India	1000	1000	1000	3000
Fieldwork	1000	1000	-	2000
Data collection	1000	1000	-	2000
Thesis workout	-	-	1000	1000
Total	3000	4000	3000	**10,000**

Table – 4: Tentative allocation of research budget

10.0. Plan of future publications & thesis chapters:

10.1. Tentative topics of future publications

1. Conceptual knowledge development on various landsliding processes, logical selection of reasoned geofactor database using various knowledge and data-guided statistical treatments and establishing their role in causing specific slide types.

2. Role of analysis of rainfall amount-duration and temporal landslide data in predicting different hazard scenarios.

3. Role of multi-temporal landslide inventory data in probabilistic analysis of temporal and magnitude aspects of future landslides.

4. Development and validation of a knowledge-guided quantitative algorithm vs. true data-dependent methods of spatial prediction.

10.2. Tentative chapters of final thesis

Chapter - I Introduction

- Research Problems
- Research Hypotheses
- Objectives of research along with main research questions answered
- Rationale of taking up this research problem
- Methodological framework: How the objectives were achieved
- Structure of thesis chapters

Chapter – II Development of conceptual knowledge base on landslide processes

- Identification and delineation of landslide types (both space and time) and its processes.
- Spatial and temporal disposition of pre-disposing factors and their specific relations to different landslide types.
- Reasons for variation in causal relationships of geofactors & their factor classes to specific slide types.
- Knowledge-guided preference of variables and its rationale; selection of geofactor combinations or groups with special application to the study area.

- Spatial evolution of landslides with time: its type and distribution.
- Development of knowledge regarding intensity and magnitude of triggering mechanisms vis-à-vis various slide processes.
- Knowledge development through creation of direct susceptibility maps at field for specific type locations.
- Selection of a set of training sites comprising equal number of failed and unfailed slopes and its basis of selection.

Chapter - III **Reasoned geospatial database preparation**

- Transformation of knowledge on landslide processes for preparation of reasoned thematic geofactor layers for different slide types (with special emphasis on limitation of spatial data generation)
- Transformation of knowledge on spatial evolution of landslide for the preparation of geomorphological and event-based landslide inventories (with special reference in establishing the spatial relationship between two inventories).
- Interactive statistical data treatments to geofactor database for facilitating selection of statistically significant variables in case of knowledge gaps (with special reference to how far statistical data treatments corroborates knowledge selection).

Chapter – IV **Data integration for spatial prediction**

- Mapping unit: method of selection, rationale and distribution; spatial relation with existing landslide types
- Data integration and spatial predition through true data-dependent modelling techniques (Logistic regression and ANN).
- Validation and sensitivity analysis of such true data-dependent methods to establish whether the above models are insensitive to knowledge-guided geospatial data selection.
- Spatial prediction after developing specific knowledge-guided algorithms in a quantitative technique and validation of its model performance vis-à-vis true data-dependent techniques.
- Validation of all the models with spatial and temporal test datasets and also using different mapping unit types; determination of prediction image and spatial probability.

Chapter – V **Data integration for temporal prediction and prediction of magnitude**

- Analysis of rainfall data and linking the same to temporal landslide events and determination of rainfall thresholds.
- Development of various hazard scenarios depending on different landslide events of differing magnitudes and types
- Temporal prediction using landslide frequency analysis and comparing the same with hazard scenarios developed through rainfall and landslide event data
- Limitations and uncertainties in temporal prediction studies
- Pattern of distribution of landslide area and frequency and estimation of area of future landslide.

Chapter – VI **Data integration for landslide risk analysis**

- Elements at risk mapping: nature of spatial distribution; classification (both physical and social type); selection of mapping unit for risk calculation and preparation of elements at risk database; limitations of elements at risk mapping.
- Empirical determination of travel distance and linking the same to spatial prediction map for using the same for risk calculations.

- Identification of limitations in assessing the spatial interaction of elements at risk and hazard of a particular magnitude.\
- Development of various risks depending on different hazard scenarios; generation of risk curves and estimation of annual risks
- Limitations in calculation of vulnerability and limitations in quantitative calculations of risk;
- Utility of risk maps (qualitative vs. quantitative).

Chapter– VII Discussions and synthesis of information

11.0. References:

Aleotti, P. and Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. Bulletin of Engineering Geology and Environment, 58: 21-44.

Australian Geomechanics Society and Sub-committee on landslide risk management, 2000. Landslide risk management concepts and guidelines. Australian Geomechanics, 35(1): 49-92.

Baeza, C. and Corominas, J., 2001. Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. Earth surface processes and landforms, 26: 1251-1263.

Barredo, J., Benavides, A., Hervas, J. and van Westen, C.J., 2000. Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin, Gran Canaria Island, Spain. International Journal of Applied Earth Observation and Geoinformation, 2(1): 9-23.

Basu, S.R. and De, S.K., 2003. Causes and consequences of landslides in Darjeeling-Sikkim Himalayas. Geographia Polonica 76(2): 37-52.

BIS, 1998. Preparation of landslide hazard zonation maps in mountainous terrains - guidelines. Bureau of Indian Standard:IS 14496 (Part - 2)

BIS, 2002. Seismic Zonation Map of India: IS: 1893 (Part - I); Revised 2002.

Bonham-Carter, G.F., 1996. Geographic Information Systems for Geoscientists: Modeling with GIS. Pergamon, Elsevier Science Ltd., 398 pp.

Brand, E.W., Premchitt, J. and Phillipson, H.B., 1984. Relationship between rainfall and landslides, Proceedings of the fourth International Symposium on Landslides, Vancouver, pp. 377-384.

Caine, N., 1980. The rainfall intensity duration control of shallow landslides and debris flows. Geografiska Annaler, 62A: 23-27.

Cardinali, M. et al., 2002. A geomorphological approach to the estimation of landslide hazards and risks in Umbria, Central Italy. Natural Hazards and Earth System Sciences, 2: 57-72.

Carrara, A., 1983. Multivariate models for landslide hazard evaluation. Mathematical Geology, 15: 403-426.

Carrara, A., 1989. Landslide hazard mapping by statistical methods: a "black box" model approach. In: CNR-USNSF (Editor), Proceed. Int. Workshop Natural Disasters in Europe-Mediterr. Countries, Perugia, pp. 205-224.

Carrara, A., 1993. Potentials and pitfalls of GIS technology in assessing natural hazards. In: Reichenbach&Guzzetti&Carrara (Editor), Geographical Information Systems in Assessing Natural Hazards — Abstracts. CNR, Perugia, pp. 128-137.

Carrara, A. et al., 1991. GIS techniques and statistical models in evaluating landslide hazard. Earth surface processes and landforms, 16(5): 427-445.

Carrara, A., Cardinali, M. and Guzzetti, F., 1992. Uncertainty in assessing landslide hazard and risk. ITC Journal, 2: 172-183.

Carrara, A., Cardinali, M., Guzzetti, F. and Reichenbach, P., 1995. GIS technology in mapping landslide hazard. In: A.C.F. Guzzetti (Editor), Geographical information systems in assessing natural hazards. Kluwer Academic Publishers, pp. 135-175.

Chelborad, A.F., 2003. Preliminary Evaluation of a Precipitation Threshold for Anticipating the Occurrence of Landslides in the Seattle, Washington, Area: USGS Open File Report 00-469
In: USGS (Editor). USGS.
Chopra, B.R., 1977. Landslides and other mass movements along roads in Sikkim and North Bengal. Bulletin of the International Association of Engineering Geology, 16: 162-166.
Chowdhury, R. and Flentje, P., 2002. Uncertainties in rainfall-induced landslide hazard. Quarterly Journal of Engineering Geology and Hydrogeology, 35: 61-70.
Chowdhury, R. and Flentje, P., 2003. Role of slope reliability analysis in landslide risk management. Bulletin of Engineering Geology and Environment, 62: 41-46.
Chung, C.-J.F. and Fabbri, A.G., 2003. Validation of Spatial Prediction Models for Landslide Hazard Mapping. Natural Hazards, 30(3): 451-472.
Chung, C.-J.F., Fabbri, A.G. and Westen, C.v., 1995. Multivariate regression analysis for landslide hazard zonation. In: A. Carrara and F. Guzzetti (Editors), Geographical Information Systems in Assessing Natural Hazards. Kluwer Academic Publishers, Netherlands, pp. 107-133.
Chung, C.F. and Fabbri, A.G., 1999. Probabilistic prediction models for landslide hazard mapping. Photogrammetric Engineering & Remote Sensing, 65(12): 1389–1399.
Chung, C.J.F. and Lecrec, Y., 1994. Use of quantitative techniques for zonning landslide hazards. Proc. Int. Assoc. of Mathematical Geology, Quebec, pp. 87-93.
Coe, J.A. et al., 2004. Probabilistic assessment of precipitation-triggered landslide using historical records of landslide occurrence, Seattle, Whashington. Environmental & Engineering Geoscience, X(2): 103-122.
Corominas, J., 1996. The angle of reach as a mobility index for small and large landslides. Canadian Geotechnical Journal, 33(2): 260-271.
Crovelli, R.A., 2000. Probability models for estimation of number and costs of landslides, USGS, Denver, Colorado.
Crozier, M.J., 1999. Prediction of rainfall-triggered landslide: a test of antecedent water status model. Earth surface processes and landforms, 24: 825-833.
Cruden, D.M. and Varnes, D.J., 1996. Landslide types and processes. In: Turner&Schuster (Editor), Landslides investigation and mitigation (Special Report 247). Transportation Research Board (National Research Council), pp. 36-75.
Dai, F.C. and Lee, C.F., 2001. Frequency-volume relation and prediction of rainfall-induced landslides. Engineering Geology, 59(3-4): 253-266.
Dai, F.C. and Lee, C.F., 2003. A spatiotemporal probabilistic modelling of storm-induced shallow landsliding using aerial photographs and logistic regression. Earth surface processes and landforms, 28: 527-545.
Fell, R., Ho, K.K.S., Lacasse, S. and Leroi, E., 2005. A framework for landslide risk assessment and management. In: E. Eberhardt, O. Hungr, R. Fell and R. Couture (Editors), Proceedings, International Conference on Landslide Risk Management, May 31-Jun 3, 2005, Vancouver, Canada.
Finlay, P.J., Mostyn, G.R. and Fell, R., 1999. Landslide risk assessment: prediction of travel distance. Canadian Geotechnical Journal, 36: 46-62.
Gabet, E.J., Burbank, D.W., Putkonen, J.K., Pratt-Sitaula, B.A. and Ojha, T., 2004. Rainfall thresholds for landsliding in the Himalayas of Nepal. Geomorphology, 63(3-4): 131-143.
Griesbach, R.L., 1900. Annual Report of GSI. In: GSI (Editor). GSI.
Guzzetti, F., Cardinali, M. and Reichenbach, P., 1994. The AVI project: a bibliographical and archive inventory of landslides and floods in Italy. Environmental Management, 18(4): 623-633.
Guzzetti, F., Carrara, A., Cardinali, M. and Reichenbach, P., 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology, 31(1-4): 181-216.
Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M. and Galli, M., 2006. Estimating the quality of landslide susceptibility models. Geomorphology, 81(1-2): 166-184.
Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M. and Ardizzone, F., 2005. Probabilistic landslide hazard assessment at the basin scale. Geomorphology, 72: 272-299.

Haining, R., 1990. Spatial Data Analysis in the Slope and Environmental Sciences. University Press, pp. 409 Pages.
Hansen, A., 1984. Landslide hazard analysis. In: Brunsden&Prior (Editor), Slope Instability. John Wiley & Sons, New York, pp. 523-602.
Ho, K., Leroi, E. and Roberd, W., 2000. Quantitative risk assessmnt - application, myths and future direction, GeoEngg2000, Proceedings of International conference on geotechnical and geological engineering, Melbourne.
Hovius, N., Stark, C.P. and Allen, P.A., 1997. Sediment flux from a mountain belt derived by landslide mapping. Geology, 25: 231-234.
Hungr, O., Corominas, J. and Eberhardt, E., 2005. Estimating landslide motion mechanism, travel distance and velocity. In: H.F.C. Eberhardt (Editor), Landslide risk management. Balkema, Vancouver, pp. 99-128.
Kim, S.K., Hong, W.P. and Kim, Y.M., 1991. Prediction of rainfall-triggered landslides in Korea. In: D.H. Bell (Editor), Landslides, Balkem, Rotterdam, pp. 989-994.
Lateltin, O.E., 1997. Berücksichtigung der Massenbewegungsgefahren bei raumwirksamen Tätigkeiten, Bundesant für Raumplanung (BRP), Bundesant für Wasserwirtschaft (BWW), Bundesant für Umwelt, Wald und Landschaft (BUWAL). Bern.
Lee, E.M. and Jones, D.K.C., 2004. Landslide risk assessment. ThomasTelford, London, 454 pp.
Lee, S., Ryu, J.-H., Lee, M.-J. and Won, J.-S., 2003. Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea. Environmental Geology, 44: 820-833.
Lumb, P., 1975. Slope failures in Hong Kong. Engineering Geology, 8: 31-65.
Malamud, B.D., Turcotte, D.L., Guzzetti, F. and Reichenbach, P., 2004. Landslide Inventories and Their Statistical Properties. Earth Surface Processes and Landforms, 29: 687-711.
Mathur, H.N., 1982. Influence of human activities on landslides, mudflows and slope movements in India and efforts at reducing their negative impact. In: A.I. Sheko (Editor), Landslides and mudflows-Reports of the Alma-Ata International seminar. GKNT, Moskow, pp. 20-44.
Matthews, J.A., 1981. Quantitative and Statistical Approaches to Geography. Pergamon Press, Oxford.
Moon, V. and Blackstock, H., 2004. A Methodology for Assessing Landslide Hazard Using Deterministic Stability Models. Natural Hazards, 32(1): 111-134.
Morgan, G.C., 1992. Quantification of risks from slope hazards. Open File Report No. 1992-15, Geological Survey of Canada, Canada.
Nadim, F., Kjekstad, O., Peduzzi, P., Herold, C. and Jaedicke, C., 2006. Global landslide and avalanche hotspots. Landslides, 3: 159-173.
Neuland, H., 1976. A prediction model of landslips. Catena 3: 215-230.
Pawde, M.B. and Saha, S.S., 1982. Geology of Darjeeling Himalaya. In: GSI (Editor), Miscellaneous Publication No. 41 Part-II (Section-IB: Geology, Stratigraphy and Palaeontology). GSI.
Raina, V.K. and Srivastava, B.S., 1981. Tectonic evolution of Sikkim Himalaya. Contemp. Geosci. Res. Himalaya, 1: 109-115.
Roberds, W., 2005. Estimating temporal and spatial vulnerability. In: H.F.C. Eberhardt (Editor), Landslide risk management. Balkema, pp. 129-157.
Saaty, T.L., 1980. The analytic hierarchy process: planning, priority setting, resource allocation. McGraw-Hill, 287 pp.
Saaty, T.L., 1996. The Analytic Hierarchy Process. McGraw Hill, 1980, New York.
Schuster, R.L., 1996. Socioeconomic significance of landslides. In: Turner&Schuster (Editor), Landslides investigation and mitigation Special Report 247. Transportation Research Board (National Research Council) pp. 12-35.
Schuster, R.L. and Fleming, W.F., 1986. Economic losses and fatalities due to landslides. Bulletin of the Association of Engineering Geologists, XXIII(1): 11-28.
Soja, R. and Starkel, L., 2007. Extreme rainfalls in Eastern Himalaya and southern slope of
Meghalaya Plateau and their geomorphologic impacts. Geomorphology, 84: 170-180.
Stark, C.P. and Hovius, N., 2001. The characterization of landslide size distributions. Geophysical Research Letters, 28(6): 1091-1094.

Starkel, L. and Basu, S.R., 2000. Rains, Landslides and Floods in the Darjeeling Himalaya. In: INSA (Editor), pp. 1-168.
Surendranath, M. et al., 2006. Landslide Hazard Zonation in Darjeeling Himalayas: A case study on integration of IRS and SRTM data, Proc. 2nd Int. Symp. on Geo-Information for Database Management, Goa.
Terlien, M.T.J., 1998. The determination of statistical and deterministic hydrological landslide-triggering thresholds. Environmental Geology, 35(2-3): 124-130.
van Westen, C., 1993. Remote Sensing and Geographic Information Systems for Geological Hazard Mitigation. ITC-Journal, 4: 393-399.
van Westen, C.J., Rengers, N. and Soeters, R., 2003. Use of geomorphological information in indirect landslide susceptibility assessment. Natural Hazards, 30(3): 399-419.
van Westen, C.J., van Asch, T.W.J. and Soeters, 2006. Landslide hazard and risk zonation—why is it still so difficult? Bulletin of Engineering Geology and Environment, 65: 167-184.
Varnes, D.J., 1978. Slope movements types and processes. In: R.L. Schuster and R.L. Krizek (Editors), Landslides: Analysis and Control. Special Report 176. Transportation Research Board, National Academy of Sciences, Washington, D.C., pp. 11-33.
Varnes, D.J., 1984. Landslide Hazard Zonation: a review of principles and practice. UNESCO, Darantiere, Paris, 61 pp.
Zezere, J.L. et al., 2004. Integration of spatial and temporal data for the definition of different landslide hazard scenarios in the area north of Lisbon (Portugal). Natural Hazards and Earth System Sciences, 4: 133-146.
Zezere, J.L., Trigo, R.M. and Trigo, I.F., 2005. Shallow and deep landslides induced by rainfall in the Lisbon region (Portugal): assessment of relationships with the North Atlantic Oscillation. Natural Hazards and Earth System Sciences, 5: 331-344.