



Novel indicators for identifying critical
INFRAstructure at RISK from Natural Hazards

Deliverable D5.1

Integrated Spatio-Temporal Database



Primary Author	Tao Cheng and Khaled Taalab/ University College London (UCL)
WP	5
Submission Date	12/11/2014
Primary Reviewer	Bryan Adey and Jürgen Hackl/ Eidgenoessische Technische Hochschule Zurich (ETH Zurich)
Dissemination Level	PU

This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 603960.

Project Information

<u>Project Duration:</u>	1/10/2013 - 30/09/2016
<u>Project Coordinator:</u>	Professor Eugene O' Brien Roughan & O' Donovan Limited eugene.obrien@rod.ie
<u>Work Programme:</u>	2013 Cooperation Theme 6: Environment (Including Climate Change).
<u>Call Topic:</u>	Env.2013.6.4-4 Towards Stress Testing of Critical Infrastructure Against Natural Hazards-FP7-ENV-2013-two stage.
<u>Project Website:</u>	www.infrarisk-fp7.eu

Partners:



Roughan & O' Donovan Limited, Ireland



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Eidgenössische Technische Hochschule Zürich, Switzerland.



Dragados SA, Spain.



Gavin and Doherty Geosolutions Ltd., Ireland.



Probabilistic Solutions Consult and Training BV, Netherlands.



Agencia Estatal Consejo Superior de Investigaciones Científicas,
Spain.



University College London, United Kingdom.



PSJ, Netherlands.



Stiftelsen SINTEF, Norway.



Ritchey Consulting AB, Sweden.



University of Southampton (IT Innovation Centre), United
Kingdom.

Document Information

Version	Date	Description	Primary Author
Rev01	31/03/2014	Draft for Comment	Cheng & Taalab
Rev02	21/08/2014	Revision	Cheng & Taalab
Rev03	03/10/2014	Second Revision	Cheng & Taalab
Rev04	11/11/2014	Final revision for submission	Cheng & Taalab

This document and the information contained herein may not be copied, used or disclosed in whole or part except with the prior written permission of the partners of the INFRARISK Consortium. The copyright and foregoing restriction on copying, use and disclosure extend to all media in which this information may be embodied, including magnetic storage, computer print-out, visual display, etc.

The information included in this document is correct to the best of the authors' knowledge. However, the document is supplied without liability for errors and omissions.

All rights reserved.

Executive Summary

This report describes the deliverable 5.1 – the description of example information to be included in an integrated spatio-temporal database that can be used to evaluate infrastructure related risk due to natural hazards. It first gives an overview of the desired spatio-temporal database, then outlines its structure and the data content using an example with publically available data sources from Switzerland. Finally, it discusses how such a database would accommodate other data sources and support future modelling work of INFRARISK.

Table of Contents

1	OVERVIEW OF AN INTEGRATED SPATIO- TEMPORAL DATABASE.....	1
2	DATA OBJECTS and DATA MODEL.....	3
2.1	Data objects	3
2.2	Data Model.....	4
3	DATA CONTENT and FORMATS.....	8
3.1	Description of Data Types	10
3.1.1	Hazard data	10
3.1.1	Environment Data	11
3.1.2	Infrastructure Data.....	17
3.1.3	Infrastructure Behaviour.....	19
4	USEFULNESS OF STDB.....	20
5	SUMMARY.....	23
6	REFERENCES	24

1 OVERVIEW OF AN INTEGRATED SPATIO- TEMPORAL DATABASE

In recent years, both the number of natural hazards occurring worldwide and the number of people affected by these events has risen significantly (Tschoegl et al., 2006). The effects include fatalities, injuries, damage to buildings and network infrastructure, economic losses and a multitude of other socioeconomic impacts. Since 1980, natural hazards have claimed over two million lives and caused losses worth US\$ 3,000 billion worldwide (Wirtz et al., 2014). Clearly, there is a need to reduce the vulnerability of both people and infrastructure, as well as to mitigate against future hazards, however, any vulnerability reduction strategies must overcome challenges such as climate change, urbanisation and environmental degradation. In order to reduce the risks related to natural hazards it is useful to collect and collate historical data. This data can be used for analysis such as risk assessment, loss modelling and planning mitigation strategies.

The INFRARISK project is concerned with the behaviour of critical infrastructure when subjected to natural hazards. Specifically, how the road and rail networks behave when they are subjected to landslides, floods, earthquakes or a combination of all three. The occurrence of these natural hazards, the behaviour of the infrastructure when the natural hazards occur and the consequences they have on the road and rail networks will vary both spatially and temporally. In order to predict future events (be they hazard occurrence or infrastructure response to hazards), it is useful to have data from previous events. This data should include the time and date that the hazard occurred as well as the location of the hazard (e.g. epicentre of an earthquake) or area affected (extent of flood and landslide). The response of the infrastructure should also be recorded. This may be the extent of damage caused or measure of performance, such as the traffic flow on a road. Other data of interest including land use, topography and climatic conditions can be used to model hazard source events, meaning those which can trigger a hazard, through to consequence events, which are events resultant from changes in the infrastructure's behaviour. Once data has been gathered, it should be collated. This means that the hazards, infrastructure and ancillary environmental data can be matched to a specific location and in some instance to a specific time.

An integrated spatio-temporal database (STDB) is used to store information that has both spatial and temporal attributes. As the name suggests, the STDB share characteristics of both spatial and temporal databases. The spatial element stores data which has an attribute value and a spatial location. For example, the location of a bridge on a road network is time independent. The fundamental spatial data types are points, lines and polygons, which can be used for analysis such as 'intersects', 'lies within' and 'is contained by' as well as operations such as length and area calculations. Temporal databases are used to store data which change over time. This can be in the form of functions with a continuous time range (e.g. times or dates) or with distinct values at various points in time (e.g. daily measurements). Distinct temporal points are usually used to record the spatio-temporal elements of infrastructure reaction to hazards. For example, the extent of damage caused by a flood would change as the spatial extent of the flood changed over time. Figure 1 shows an example of spatio-temporal data as it relates to flooding. Generally, the flood would be measured at regular time intervals (represented by T1, T2, T3). In this example, T1 is the earliest measurement and T3 is the latest measurement. Over time the extent of the flood changes, meaning the data has both spatial and temporal elements. In comparison, in the context of natural hazards, environmental

parameters such as land use or soil properties might be considered static despite the fact that they are constantly changing, as this is occurring at a much slower rate.

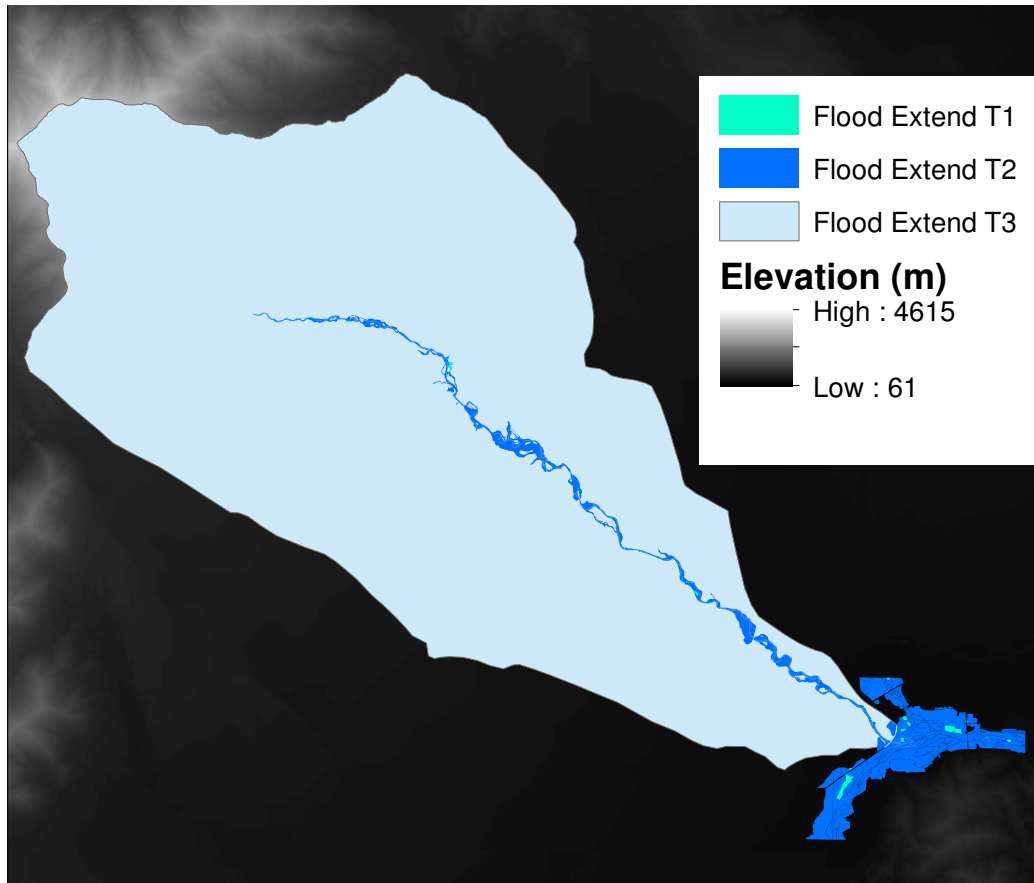


Figure 1: Example of spatio-temporal data. Here flood extent is measured at three time intervals (T1, T2, T3)

The value of some variables can be seen as static in time, while some can be seen as dynamic. For instance, hazard occurrence might be recorded as a single date and extent or as a time series with a measurement/extent recorded at regular intervals. Climatic data, such as rainfall, is generally recorded at fixed locations (e.g. monitoring stations) on a daily or hourly basis. As both spatial and temporal element needs to be supported, the database is referred to as spatio-temporal.

2 DATA OBJECTS and DATA MODEL

2.1 Data objects

Environmental data comes in many different formats and it is not always desirable to convert them into the same format. As such, using a geodatabase is a suitable method of storing a large number of different types of data that have spatial and temporal characteristics. The types of data which are commonly stored in a geodatabase include geographic features, attribute data, satellite images, GPS coordinates, 3D surface data and network data. Figure 2 shows some of the most commonly used data objects, including:

Feature Class: A group of geographic features with the same geometry type (either point, line or polygon). The feature class also has the same attributes and spatial reference. For examples, the road network, made up of motorways, primary and tertiary roads can be stored as a single 'roads' feature class.

Feature Dataset: A collection of feature class objects which have the same spatial reference and coordinate system and cover a similar geographical extent. The feature dataset is a method of representing the spatial relationship between feature classes.

Geometric Network: A set of nodes and edges used to model directed flow systems (e.g. utility networks)

Network dataset: A set of nodes and edges used to model undirected flow (e.g. road network). Each element has associated attributes.

Raster Catalog: A collection of raster datasets which can display overlapping and touching rasters without the need to perform an operation that joins the data.

Relationship Class: Manage the relationships between objects (tables or feature classes) of different classes.

Table: Columns and rows of data, with each row representing a single record and each column a field. Tables generally store information associated with other spatial data in the database, for example the type, trigger and date of a landslide event.

Data in the database can be queried directly in the GIS (the geodatabase was assembled using ArcMap 10) using SQL and a query layer in ArcGIS (ESRI, 2011). This allows relevant data to be extracted. Typical GIS operations such as overlay, proximity queries and data extraction can be performed using the inbuilt GIS functionality.

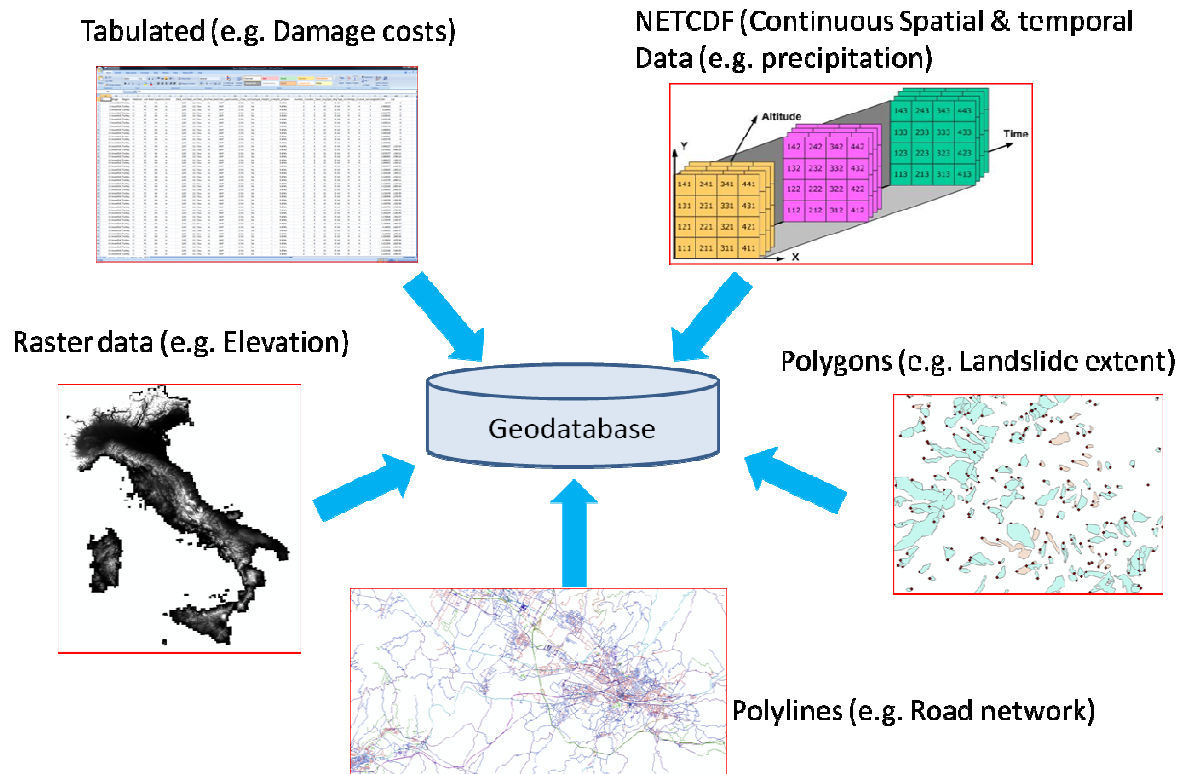


Figure 2: Conceptual format of the Geodatabase

2.2 Data Model

Within a geodatabase, it is possible to apply logical rules and relationships to the data, define geospatial relationships and integrate many data types. To better understand the relationships between objects to be included in a database that would be used to model the relationship between the environment, natural hazard occurrence and infrastructure behaviour, an Object Role Modelling (Halpin, 2006) data model has been developed. Figure 3 shows a possible database structure at the highest level, which is the relationship between the three basic data groups: infrastructure, hazard and environment. In the database, these groups are represented by feature classes. As this is an example, it is important to note that the specifics of the data model can be amended to reflect the data which is being stored.

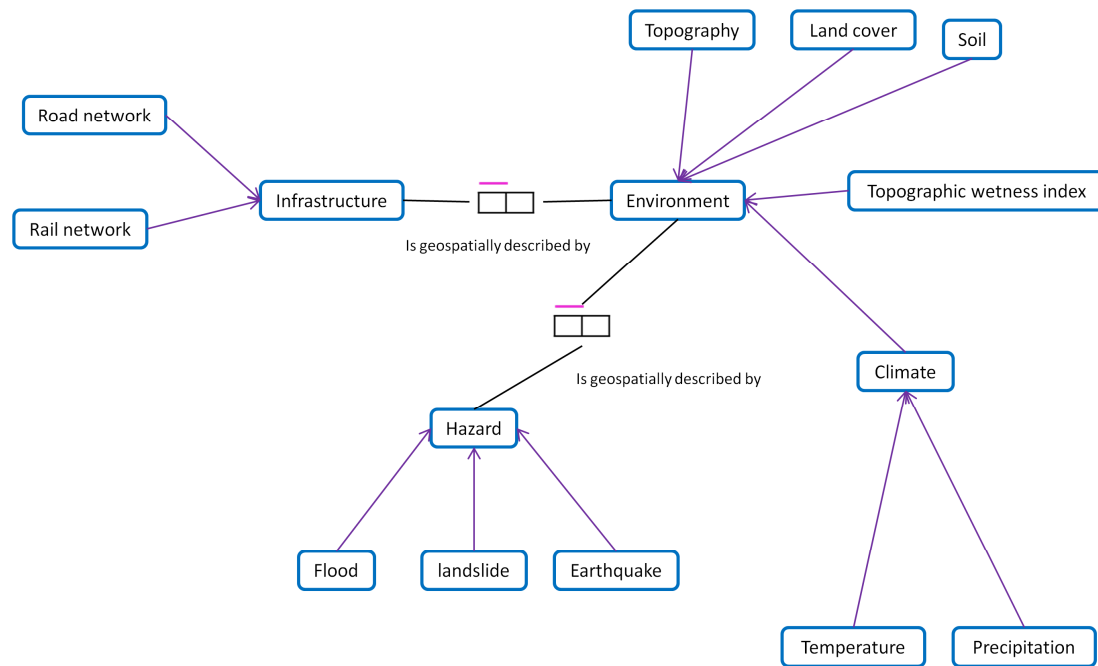


Figure 3: ORM database overview

The following diagrams show the relationship between the data which makes up each of these three groups. Figure 4 shows how road and rail network data is related to form infrastructure. The structure of the environment data is shown in Figure 5. There are many more environmental metrics which could be added to the data. Any additional variables would follow the same environment ORM structure. The Hazard ORM (Figure 6) shows how data on floods, landslides and earthquakes would be stored. If more data about floods, landslides or earthquakes became available (e.g. landslide triggers), it would be added to the database using the same structure.

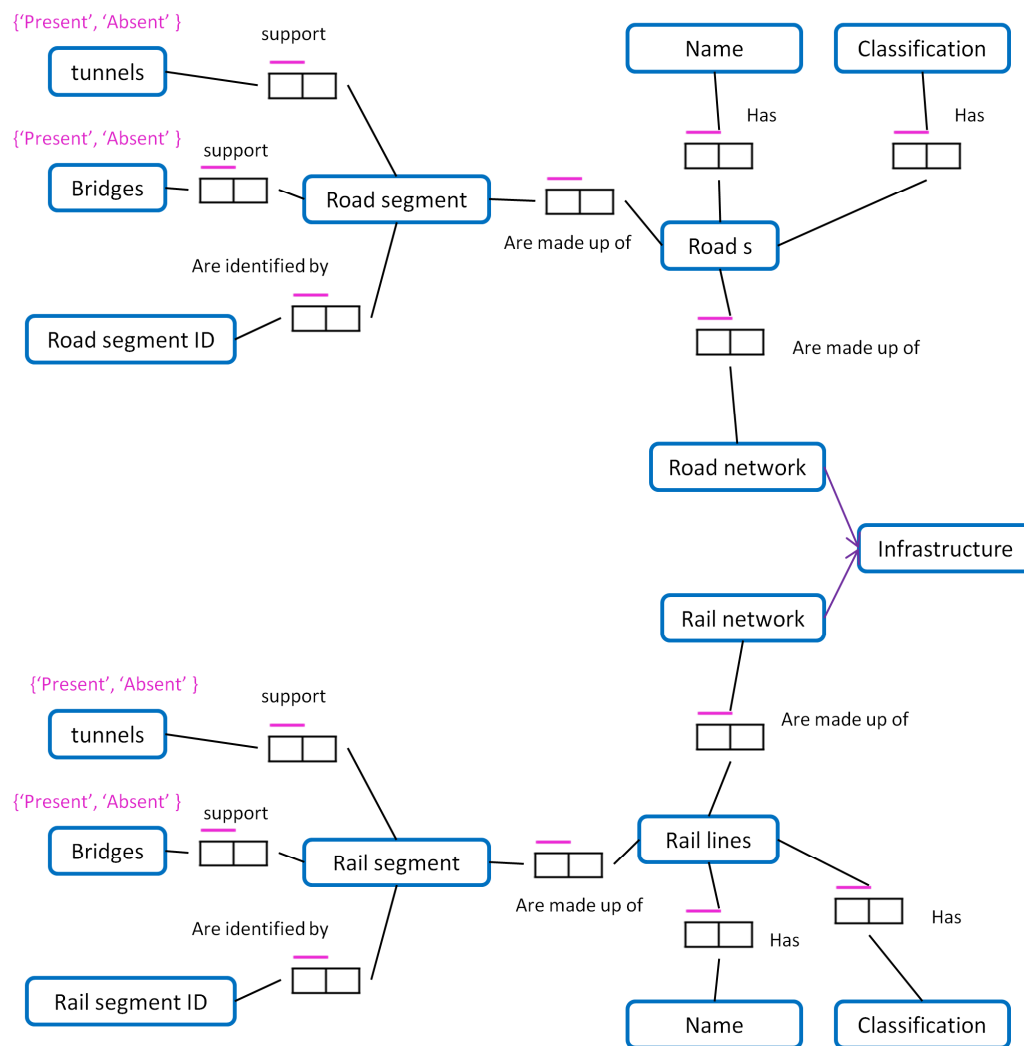


Figure 4: Infrastructure ORM

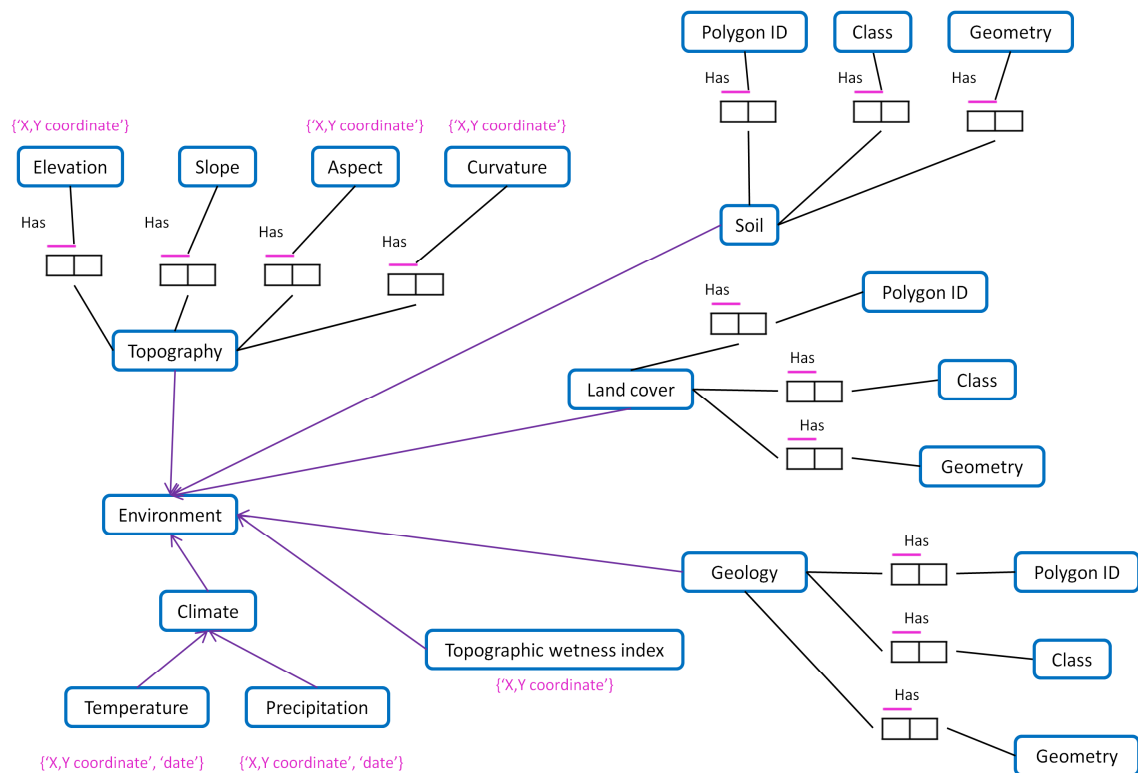


Figure 5: Environment ORM

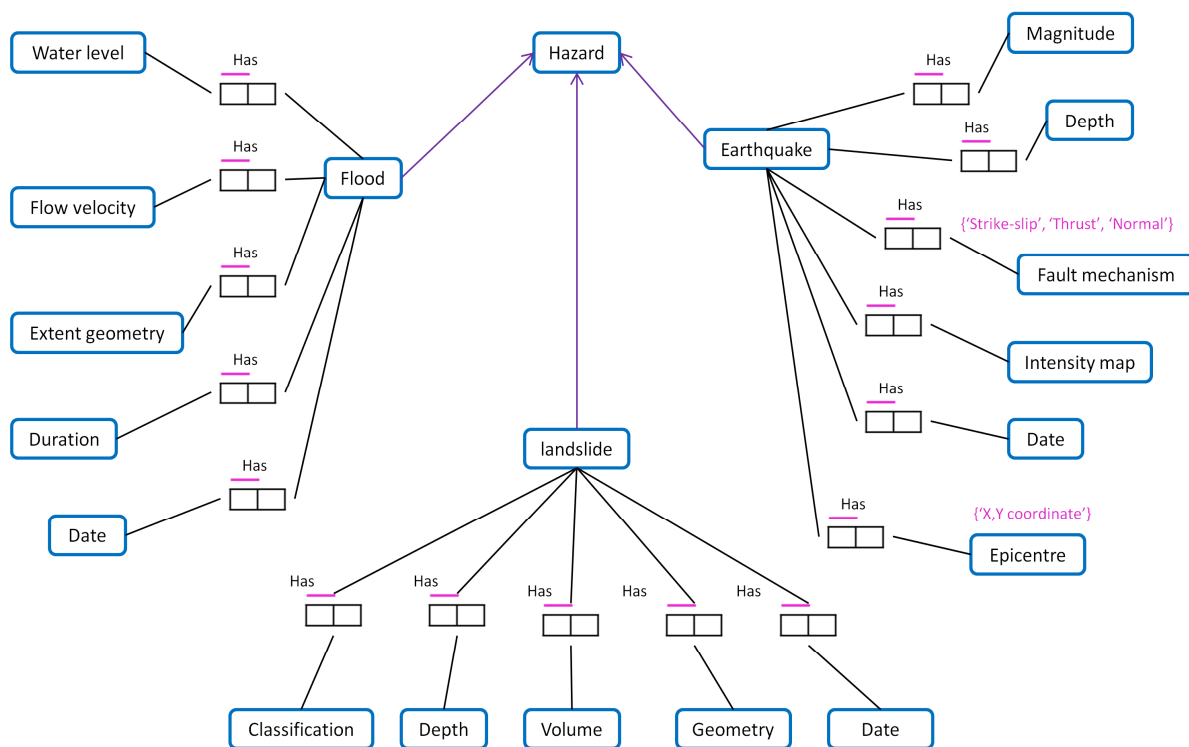


Figure 6: Hazard ORM

Infrastructure behaviour is described by either the condition of the road segment (which should be known both before and after the hazard has occurred) or the level of service it is able to provide, i.e. the amount of traffic which should be able to drive over the road. If, for example, daily traffic flow was known for a given road segment, it would be possible to calculate ‘normal’ flow for a given time and day. This would allow the disruption, for instance, the measured change between expected and observed traffic volume, caused by a hazard to be quantified. This measure would be indicative of the severity of the natural hazard. See Figure 7.

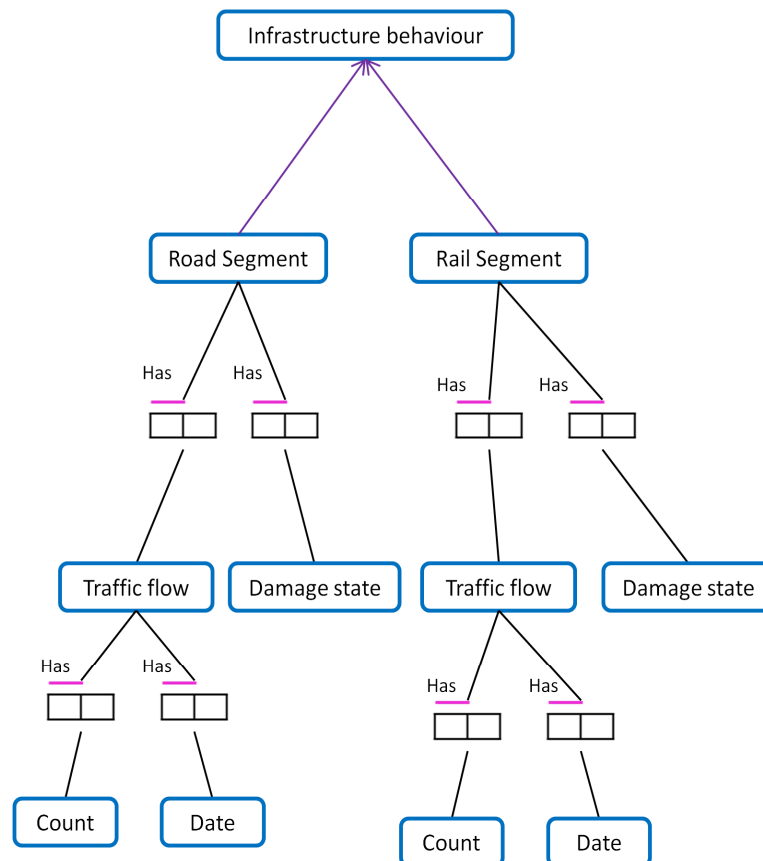


Figure 7: Infrastructure behaviour

3 DATA CONTENT and FORMATS

If there is no data restriction, such a database would contain empirical data on infrastructure behaviour given natural hazards. This behaviour would then consist of damage states (or any other quantifiable measure of damage) relating to the functionality of, for example, a bridge. Another measure of infrastructure behaviour would be flow along the network. This would be in the form of traffic counts (the number of cars passing a certain point over a defined time period) on numerous roads in the network. As a minimum, this would need to be a daily traffic count (average weekly or monthly traffic flows would not be sufficiently detailed). An alternative to direct measures of critical infrastructure behaviour is to measure the economic cost of repairing or rebuilding the damaged road and rail network. This relates specific pieces of infrastructure (roads sections, bridges etc.) with

hazard events and the cost (€ millions). Table 1 shows the example content of a STDB for the evaluation of the impact of floods and landslides on the road and rail network.

Variable	Purpose
Transport networks	The spatial location and geographic extent of the road and rail networks are required to relate critical infrastructure behaviour to natural hazards. Details on the structural parameters of the network (e.g size of road, material) will determine extent of/cost of damage or traffic flow
Natural Hazard	The location, magnitude and extent of the landslide/flood/earthquake is required to predict the effects of the hazard.
River network	Proximity to the river network will increase the risk of flood and landslide
Digital elevation model	This allows the derivation of a number of topographic feature such as elevation, slope, aspect
Precipitation	Daily rainfall across the study area is required to determine 'trigger' events for floods and landslides. As the trigger events can be sudden heavy rainfall, or prolonged rainfall and can be localised or catchment-wide, spatio-temporal rainfall data is required.
Soil classification	The occurrence of floods and landslide will be affected by the soil type. Physical attributes such as porosity and the presence of impermeable layers will dictate the movement of water across the landscape and contribute to landslide susceptibility.
Geological classification	Bedrock geology will affect soil formation and as well as the propensity of land to flood given different patterns of precipitation.
Land cover/use	Differences in land cover will affect the susceptibility of flooding and landslides in an area. The presence of different land covers within a catchment will affect the probability that certain events (e.g. precipitation) lead to the occurrence of hazards.
Traffic flow data	To model the impact of natural hazards on the functionality of the road network, traffic flow data can be used. This gives a representation of how hazards impact on those who use the network. This requires daily data as it is important to establish what 'normal' flow is to assess whether this is affected by.
Damage data	A damage state or percentage can be assigned to a structure after an event. The road may be functional for drivers to use however, damage may still have occurred.
Economic losses due to infrastructure repair	To model the impact of the hazard in terms of economic losses. Here, we do not get an understanding of the physical effects on infrastructure, but rather the cost of repairs required to restore infrastructure to its previous 'pre-hazard' state. For the purposes of this study, the economic losses should be infrastructure specific.

Table 1: The contents of an ideal database

All the data in the database would be organised in the Geodatabase, part of a GIS software package (ArcGIS), which combines spatial and temporal information in a range of formats (ESRI, 2011) (Figure 2).

In the following sections we use publically available from Switzerland, as well as other examples, to illustrate the types of data that should be included in the STDB as well as possible formats of this data

3.1 Description of Data Types

3.1.1 Hazard data

Natural hazards have the capacity to affect infrastructure of all types. The location, magnitude and impact of these hazards will be influenced in part by the environment in which they occur. The natural hazards considered in this database are floods, landslides and earthquakes.

3.1.1.1 Hazard data description

The STDB should contain information on the natural hazards and the probability of their occurrence, as in the Swiss Flood and Landslide Damage Database developed by the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL) and the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss). The hazards should be classified based on the type of hazard, e.g. a flood, landslide or rock fall. Their political authority which has jurisdiction over this region should be listed. The starting event, or trigger, should be listed. For example, the starting event for a flood may be a specific rain fall pattern or snow melt, both can be further classified on the basis of magnitude and duration. Before the construction of such a database it is necessary to have a through classification of all events and consequences to be considered as well as the descriptive attributes to be used in further analysis.

3.1.1.2 Hazard Data format

The Swiss Flood and Landslide database contains data about damage caused by flood and landslides but it contains very little data about the hazards themselves. Ideally, to model risk to infrastructure, more hazard data is required. Figure 6 outlines some basic data requirements for each of the hazards. Data on Landslides should include a depth (m), volume (m³), extent (m²), flow type, date and trigger mechanism. Data on the rock and/or soil type would also be useful, although this can be sampled from other environmental data layers. Figure 8 shows an example of landslide data in a shapefile format. The red polygons represent the spatial extent of the landslides, which can be linked with a host of landslide attribute data used to describe the event. Floods should have a water level (m), flow velocity (m/s), extent (m²), date and duration. A polygon shapefile would also be an appropriate format to represent data on floods. Different geometry shapefiles could be linked to different time steps to represent changes in the flood extent (and other properties) over time (e.g. Figure 1). Earthquakes require data on magnitude, depth (m), fault mechanism, epicentre and shake map (extent). A point shapefile would be a suitable format to represent Earthquakes. This could be linked to an associated ground shaking map to represent changes in intensity of ground motion away from the epicentre. In the case of floods and landslides, this data would allow predictive modelling of hazard occurrence. These models could be used to identify the triggering mechanisms which cause the hazards and hence identify infrastructure at risk before the hazard has occurred.

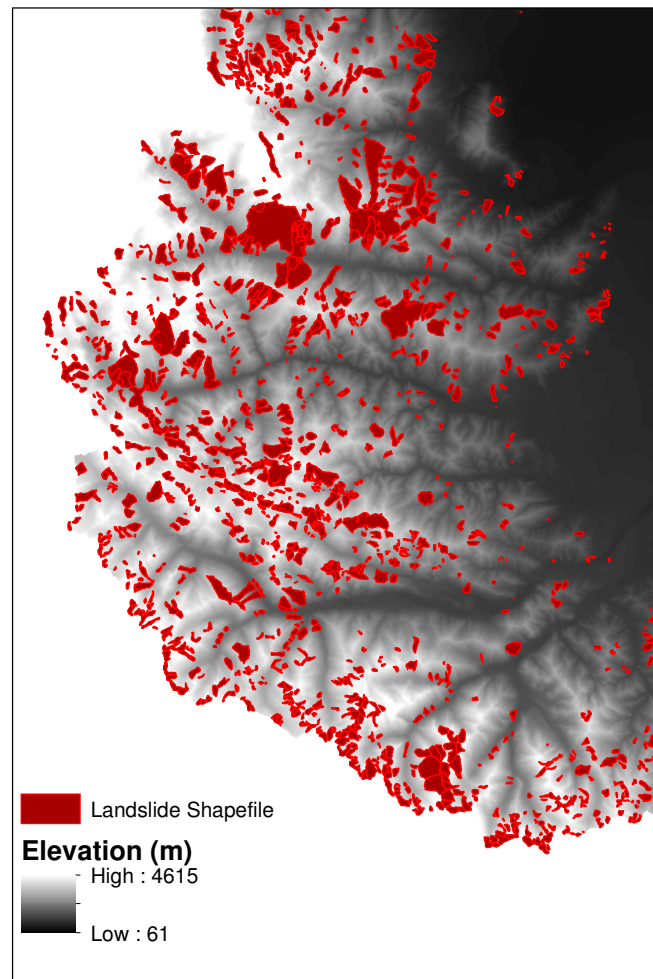


Figure 8: Example of landslide data in a shapefile format

3.1.1 Environment Data

The STDB should contain data that describes the physical environment that contains the infrastructure and the hazard. This is important as it will affect both how the hazard is triggered, the spatial and temporal occurrence of the hazard and the impact on the infrastructure. The following section describes a number of relevant environmental datasets which should be included in the STDB.

3.1.1.1 Topographic data description

Topography data should be included. For example, it could be represented using a digital elevation model (DEM) which is a gridded dataset that represents the earth's surface. As well as elevation, this model can be used for the derivation of other topographic metrics such as slope, aspect and curvature. Moreover, the DEM can be used to derive other metrics relating hydrological processes in the landscape, such as topographic wetness index, which can be linked to the occurrence of floods and landslides (Pourghasemi et al., 2013).

3.1.1.2 Topographic data format

The most common format of a DEM is a raster grid. This format is widely used and easily manipulated in all GIS software. Here the landscape is divided into a set of grid cells and each cell has an associated elevation value, creating a continuous representation of the earth's surface. In Switzerland, for example it is represented by a 200 x 200 m resolution DEM in a raster format ranging from 193-4557 m elevation (Figure 9). The basis of this model is a 1:25000 map which has been digitised and interpolated by the Swiss Federal Office of Topography. The average error associated with the DEM varies between ± 1.5 m in the Central Plateau region and between ± 3 -8 m in the Alps. As well as being used to calculate elevation, there are GIS operations which can use the DEM to calculate slope, aspect, flow paths and topographic wetness index characteristics, which may be key determinants in flood and landslide susceptibility.

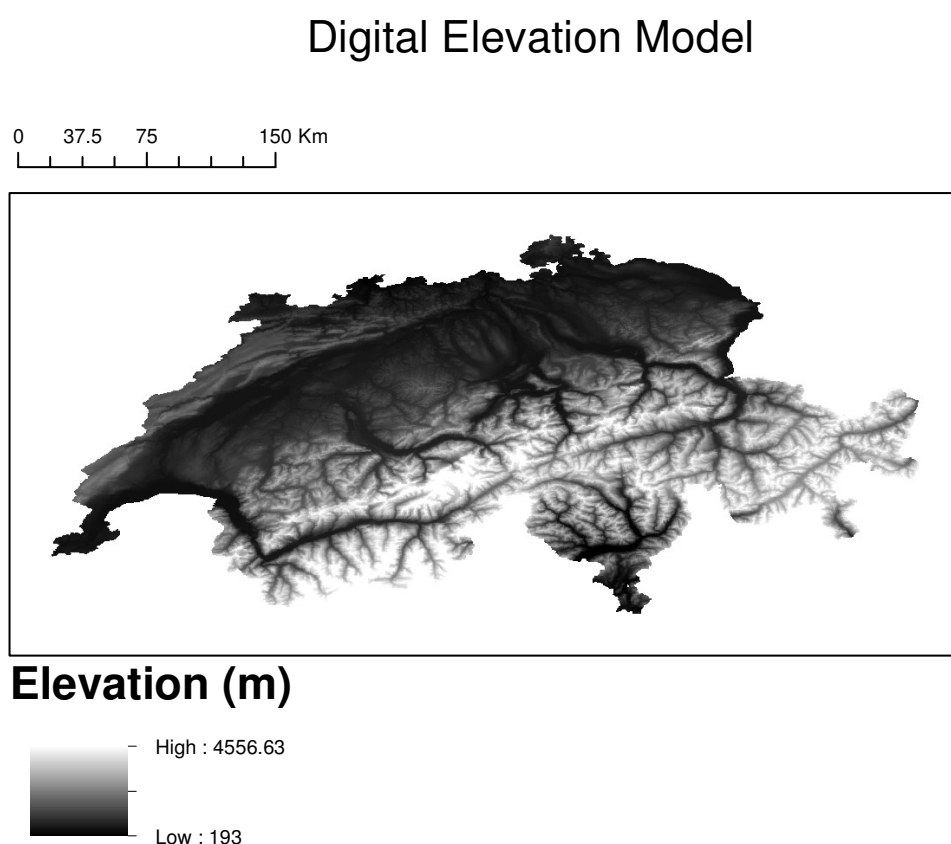


Figure 9: Digital Elevation Model

3.1.1.3 Land Cover data description

Land cover data should be included in the STDB, as it will affect the occurrence of natural hazards. For example, land cover will influence the movement of water across the landscape which will subsequently affect both the location of a flood and the rate of flooding. For example, as part of the Europe-wide CORINE land cover mapping project the landscape is divided into 44 classes derived from satellite imagery. Bossard et al. (2000) provide a comprehensive guide to the derivation of the classes and the taxonomy of the classification. In some cases, individual countries or regions will have their own land cover classification. This may include classes that are more specific to the

region. In Switzerland, there is an incomplete land use data layer available from Swiss Federal Office of Topography (swisstopo) (Figure 10c). The land use which has been classified, can fall into one of over 70 categories. This data can be downloaded from

<http://www.swisstopo.admin.ch/internet/swisstopo/en/home/products/landscape/vector200/vector200.html>

3.1.1.4 Land Cover data format

The format of land cover data is a polygon shapefile. Each polygon represents a distinct land cover group. Figure 10b shows part of the CORINE land cover classification, which is a Europe-wide land cover mapping project which provides data at a 1:100000 scale.

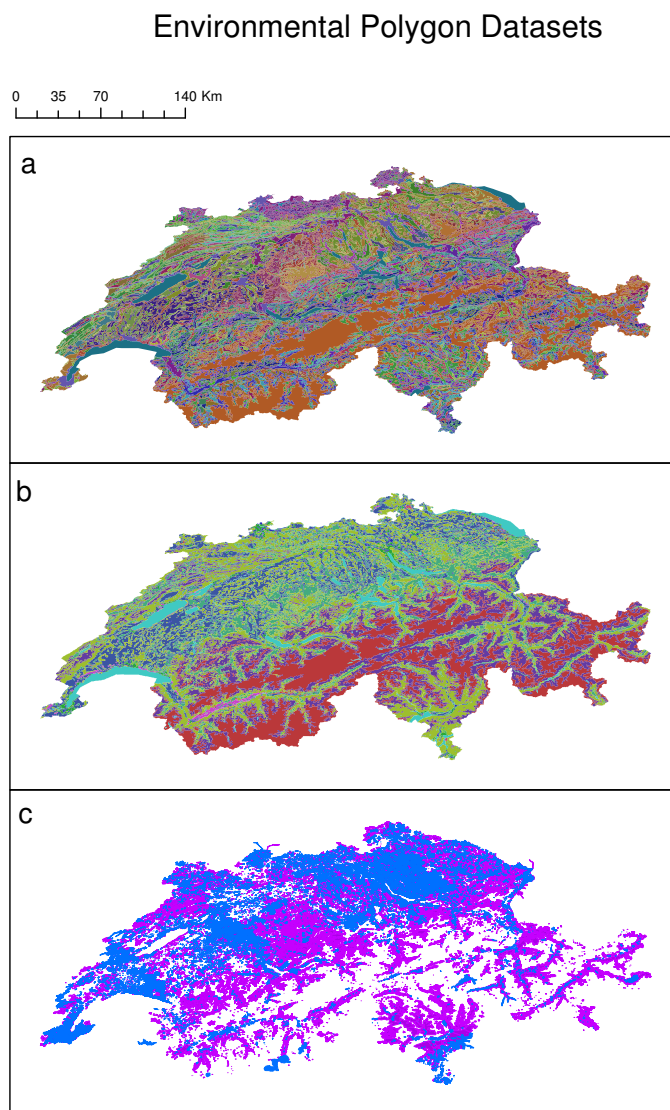


Figure 10: Environmental polygon datasets. 10a Soil classification; 10b CORINE land cover; 10c natural land use (pink) and man-made land use (blue)

3.1.1.5 Soil data description

Data on the type of soil present in the study area is of interest as soil classification and soil property data will influence the propensity of an area to produce landslides. Structural properties of soil, such as porosity and chemical properties such as clay content will influence how water moves through the soil. This will impact on the way in which flooding occurs, for example the presence of an impermeable, clay-rich layer below the topsoil may lead to increased overland flow during periods of intense rainfall. Areas with this soil type may be more susceptible to flash flooding than those where the soil is more permeable.

3.1.1.6 Soil data format

The format of land cover data is a polygon shapefile. Each polygon represents a distinct land cover group. The soil classification is a 1:200000 scale map published by the Swiss Federal Office of Spatial Planning, Agriculture and Forestry (Figure 10a). The soils are classified from 144 mapping units into 18 groups. The source reference for this data is "Federal Statistical Office (FSO) GEOSTAT". Soil data can also be in the raster format. The European Commission provides a European Soil Portal from which it is possible to download 1 km grid resolution raster datasets of soil classification and other soil metrics (such as hydrological, chemical and structural properties). This data can be downloaded from <http://eusoils.jrc.ec.europa.eu/>.

3.1.1.7 Geology data description

Bedrock geology and soil parent material are very relevant to flood, landslide and earthquake hazards. It will affect the intensity of ground movement during an earthquake as well as the type of mass movement during a landslide. Moreover, the distance from geological faults is an important metric when modelling the occurrence of landslides (Pourghasemi et al., 2013).

3.1.1.8 Geology data format

Geological data can be represented as a series of polygons, showing bedrock geology and parent material. These would be two distinct layers. Each polygon represents a bedrock/ parent material class. Geological faults are best represented as a polyline. For the UK, geological data is available to download from <http://digimap.edina.ac.uk/>.

3.1.1.9 Waterways data description

The location of streams, rivers and canals is particularly relevant to flooding events. The occurrence of floods at one location can be due to heavy rainfall elsewhere, increasing the volume of water in a number of tributaries, which can cause flooding further downstream. The proximity to rivers has also been shown to influence the occurrence of landslides (Pourghasemi et al., 2013).

3.1.1.10 Waterways data format

Rivers, streams and canals would best be represented as a polyline shapefile with attributes such as depth, width and flow measurements associated with each river. Any lakes or water bodies should

be stored as polygons. Figure 13c shows an example of a 1:200000 polyline vector layer for Switzerland produced by the Federal office of Topography.

3.1.1.11 Built environment and administrative boundaries data description

In this instance, the built environment refers to the location of settlements (cities, town and villages and their associated populations) as well as specific buildings of interest such as schools, hospitals, fire stations and police stations. The administrative boundaries demarcate the division between regions and, in some cases, local authorities. The location of settlements is of interest because this can be used to estimate the number of people affected by damage and disruption to infrastructure. Recording the location of buildings such as hospitals is necessary as, in the event of a natural hazard which affected the road and rail infrastructure, emergency services may need to be re-routed. Recording the administrative boundaries of the hazard and infrastructure is useful as the authorities responsible for hazard mitigation and response may need to be advised on mitigation strategies or warned about probably future hazard events.

3.1.1.12 Built environment and administrative boundaries data format

Administrative boundaries should be a polygon dataset, such as in the Swiss example (Figure 11a). The buildings dataset (Figure 11b) is a set of polygons that show the building's footprint. In this data, buildings of interest, such as those that house the emergency services, are identified in the polygons attributes. Towns, villages and cities can be represent by a polygon showing the extent or a point showing the centre of the settlement. Attributes such as population, should be associated with either the point of polygon files.

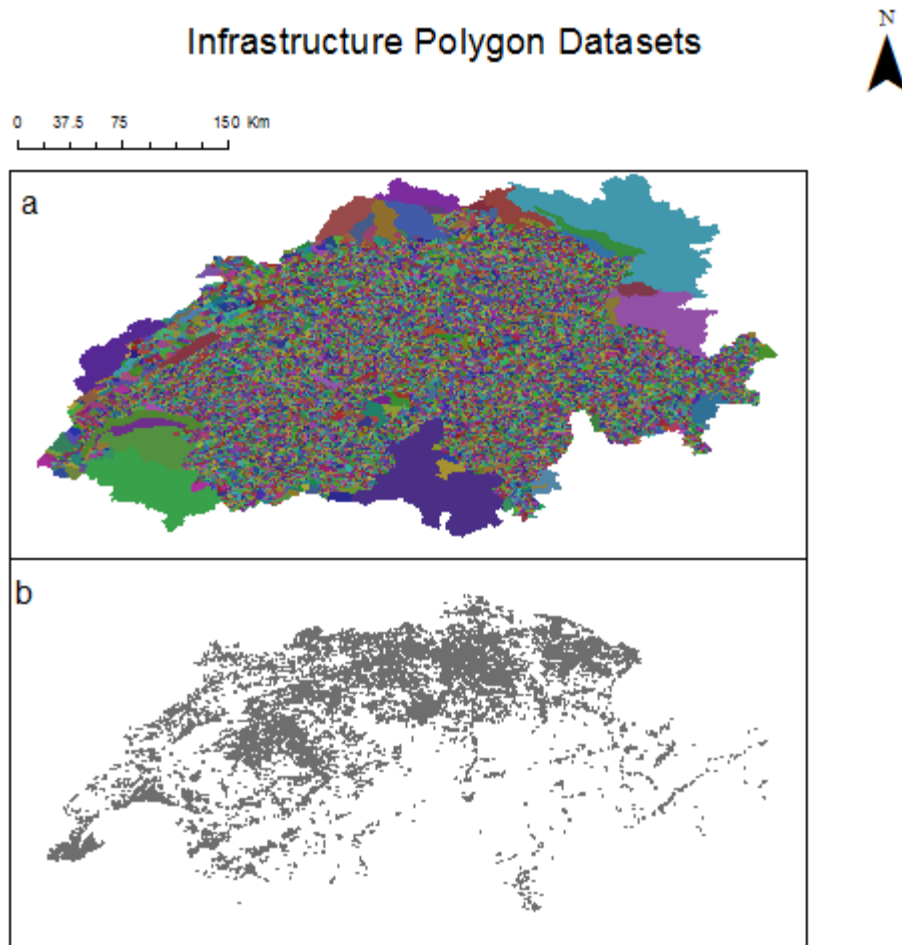


Figure 11: Infrastructure polygon datasets. 11a Communities; 11b Buildings

3.1.1.13 Climatic data description

Many flood and landslide hazard events are triggered by climatic variables, in particular precipitation. One example would be heavy rainfall causing a flood. The relationship between climate and hazard is highly complex, as the patterns of rainfall over a prolonged period of weeks or months can cause a hazard event. This makes predicting the temporal occurrence of hazards very difficult. Where the trigger of a hazard event is known (e.g. heavy rainfall), it is possible to use climatic data to quantify the amount of rainfall that caused the hazard. As climatic variables do not act in isolation, climatic data of interest include temperature, wind and solar radiation. In some regions, seasonal events, such as snow melt will also be of interest.

3.1.1.14 Climatic data format

Climatic data has both a spatial and temporal dimension. Generally, climatic variables are measured on a daily (or sometimes hourly) basis at a number of monitoring stations (points). In order to provide a continuous surface of the climatic attribute being measured, these point measurements are frequently interpolated. To store the interpolated dataset, the NetCDF file format is used. This is specifically developed to store spatio-temporal datasets, which are large files (in comparison with a single polygon or raster layer). It is possible to sample from this layer directly, or transform this data

into a series of rasters (a single raster for each time step). Figure 12 shows a single day's rainfall taken from daily rainfall readings for the entire alpine region over the period 1971-2008. This data was derived from 5500 daily rainfall measurements across the alpine region. Using a distance-angular weighting scheme to describe the relationship between precipitation and topography, rainfall data was extrapolated across a 5x5 km grid. Isotta et al (2014) provides an explanation of the methods used to derive the dataset.

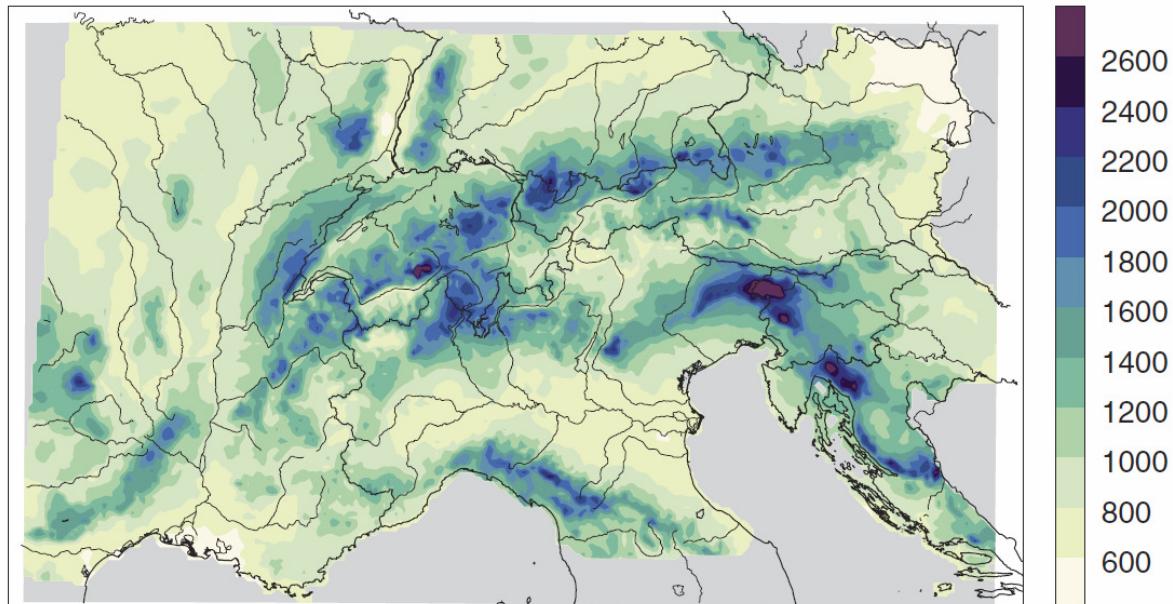


Figure 12: Mean annual average precipitation (mm per year) for the period 1971-2008 for the entire alpine region. (reproduced from Isotta et al., 2014)

3.1.2 Infrastructure Data

Infrastructure can be defined as the network man-made systems which function interdependently to produce and distribute a continuous flow of essential goods and services. Infrastructure is deemed critical when it is necessary for a country's economic and national security (President's Commission on Critical Infrastructure Protection, 1997). Typically, critical infrastructure will include transport, energy, water supply, telecommunications, finance and both government and emergency services (Rinaldi et al., 2001). Transport Infrastructure is critical to the economic and social prosperity of a country. In the context of natural hazards, its importance is magnified, as transport networks must be in place in order to facilitate repairs to all other lifeline systems.

3.1.2.1 Road and Rail Network data description

The road and rail network should be integrated as GIS data layers for the entire region being investigated, e.g. a country. For example, the road and rail network for most countries in Europe are available from the Openstreetmap project (Haklay & Weber, 2008). Using the Openstreetmap example, the roads are classified by type, the most significant in terms of this study are motorways and primary roads (although all roads down to paths are included in the database). There are also additional details such as the presence of bridges and tunnels on the roads, junctions, signals, crossings, speed cameras, level crossings and roundabouts. Attributes such as speed limit and

number of lanes should also be recorded as this data can be used to establish flow on the network (see section 3.2.3). The rail network is classified between rail, narrow gauge, funicular (cable railway) and platforms.

3.1.2.2 Road and rail data format

The roads and rail network should be represented by shapefiles. The roads and rail tracks are best represented by polylines. A typical feature of most GIS road networks is that the roads are divided into road segments. A segment being the portion of road between junctions. The junctions are defined as any place on the network where it is possible to make a route choice. Features on the road and rail network, such as signals, roundabouts and level crossings are best represented by a point shapefile. Using the OpenStreetmap example, the presence of tunnels and bridges on the network are indicated by the number of tunnels and/or bridges in a given road segment. This data can be downloaded from http://wiki.openstreetmap.org/wiki/OSM_Map_On_Garmin/Download. Where more information on bridges and tunnels is required, it would be better to represent them their own segment. This would allow the addition of structural attribute data, such as building material, to be associated with bridges and tunnels on the network. Figure 13a and 13b shows the Swiss road and rail network respectively. This data was created by the Swiss Federal Office of Topography.

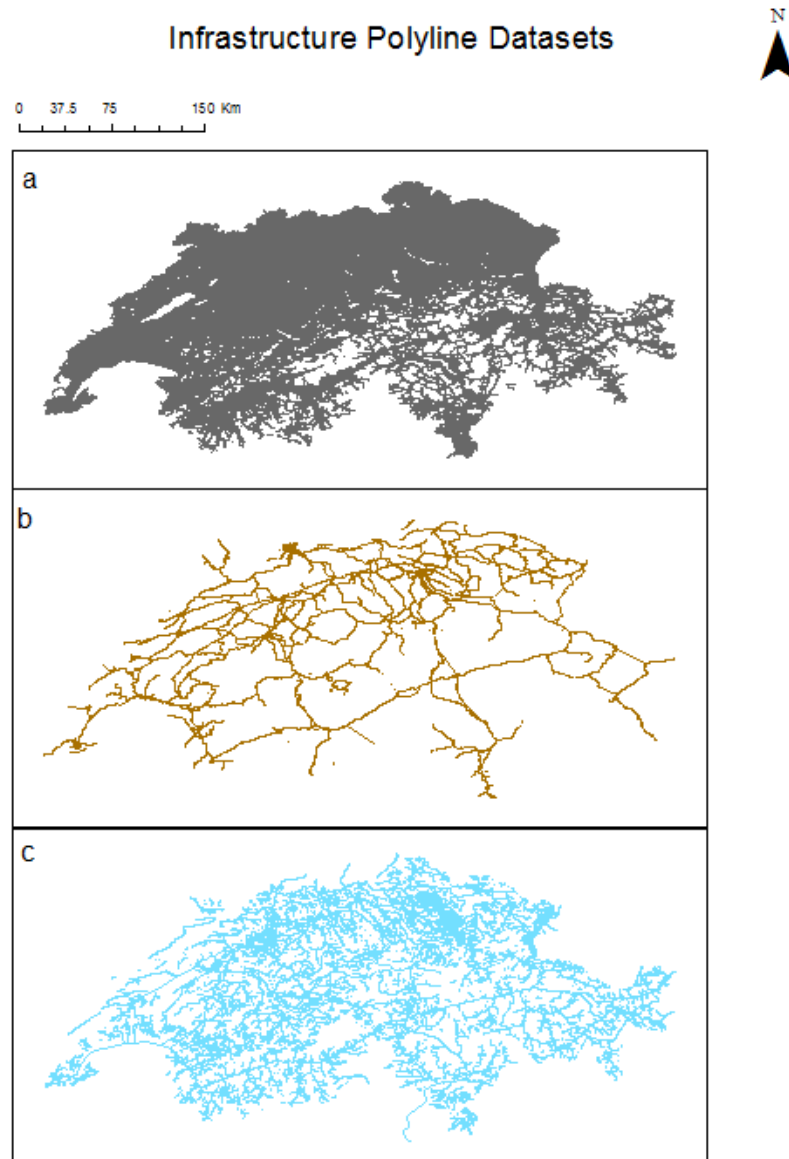


Figure 13: Infrastructure Polyline Datasets. 13a The road network; 13b The rail network; 13c Waterways

3.1.3 Infrastructure Behaviour

We are interested in the affect natural hazards have on infrastructure. This may be a single piece of infrastructure, such as a bridge or the road network as a whole. Changes in infrastructure behaviour due to a natural hazard is a consequence of the hazard which can be used to help assess risk.

3.1.3.1 Infrastructure Behaviour data description

To quantify the effect of hazards on infrastructure, there are a number of metrics which we would like to know. One measure would be the classification of the damage caused. For example, the seismic risk posed to bridges is often represented by fragility curves, which estimate the probability of a number of damage classes (usually no damage, mild, moderate, extensive and collapse) that relate to increasing measures of ground motion. These damage classes correspond to various service

levels that the bridge can provide after the seismic event. Data recording post-earthquake damage to infrastructure would be useful to model future damage events. An alternative to damage classification is the monetary cost to repair or rebuild the infrastructure. This damage metric is used in The Swiss flood and landslide damage database 1972–2007 (Hilker et al., 2009). In this instance, the cost of damage was aggregated to the community level. In terms of impact, cost may be a better metric to measure impact than damage state as many parties interested in the cost of damage. For example, stakeholders such as insurers and local and national authorities will be interested in damage cost as it compares to the cost of various mitigation strategies. Another metric is to consider the flows along the road and rail network. When a hazard affects the network it will affect the traffic flow. For instance, if an earthquake caused a bridge to collapse, traffic that would normally travel over that bridge must use alternative routes. This can cause congestion in parts of the network that have not been directly damaged by the hazard. These are known as cascading effects. Traffic flow data can be used to assess the effect of natural hazards on the network as a whole. As traffic flow differs between roads due to capacity and location and on the same road depending on time and day of the week, historic traffic data can be used to represent expected flows. This expected flow can be used to quantify the change in flow on other parts of the network after the hazard has occurred.

3.1.3.2 Infrastructure Behaviour data format

Data on damage, repair cost and traffic flow will typically be stored in an excel-style spreadsheet format. As long as there is an identifier that can link the damage or traffic count to the relevant road segment, as well as a date, it is straightforward to link it to the shapefiles (points, polylines and polygons) that represent the road and rail network. For example historic traffic flow data can be attached to the road network shapefile. This is a feature of many route guidance systems, which use a record of previous traffic flow at a specific time and day to estimate the fastest route between two points on the network.

4 USEFULNESS OF STDB

An appropriately designed STDB, with sufficient data, can be used to help determine risk related to infrastructure due to natural hazards. We are interested in the spatial and temporal relationship between hazards, infrastructure and the environment. Figure 14 shows the results of three example queries that the proposed database would support. Figure 14a shows where road and rail segments intersect. This could be used as part of a pre-screening process to help identify vulnerable parts of the road and rail network as a whole because a single, localised hazard event (such as a landslide) could disrupt both networks. Figure 14b shows the number of damaging events aggregated to the community level, ranging from low (green) to high (red). This might be of interest to strategic planners who would like to allocate finances in the future on hazard mitigation strategies and require an initial screening to identify where they should investigate things in more detail. Figure 14c shows the parts of the road and rail network that are within 200 m of a waterway, which might be useful for flood risk analysis. This could be refined to only include major waterways or waterways which have flooded within the last five years, for example.

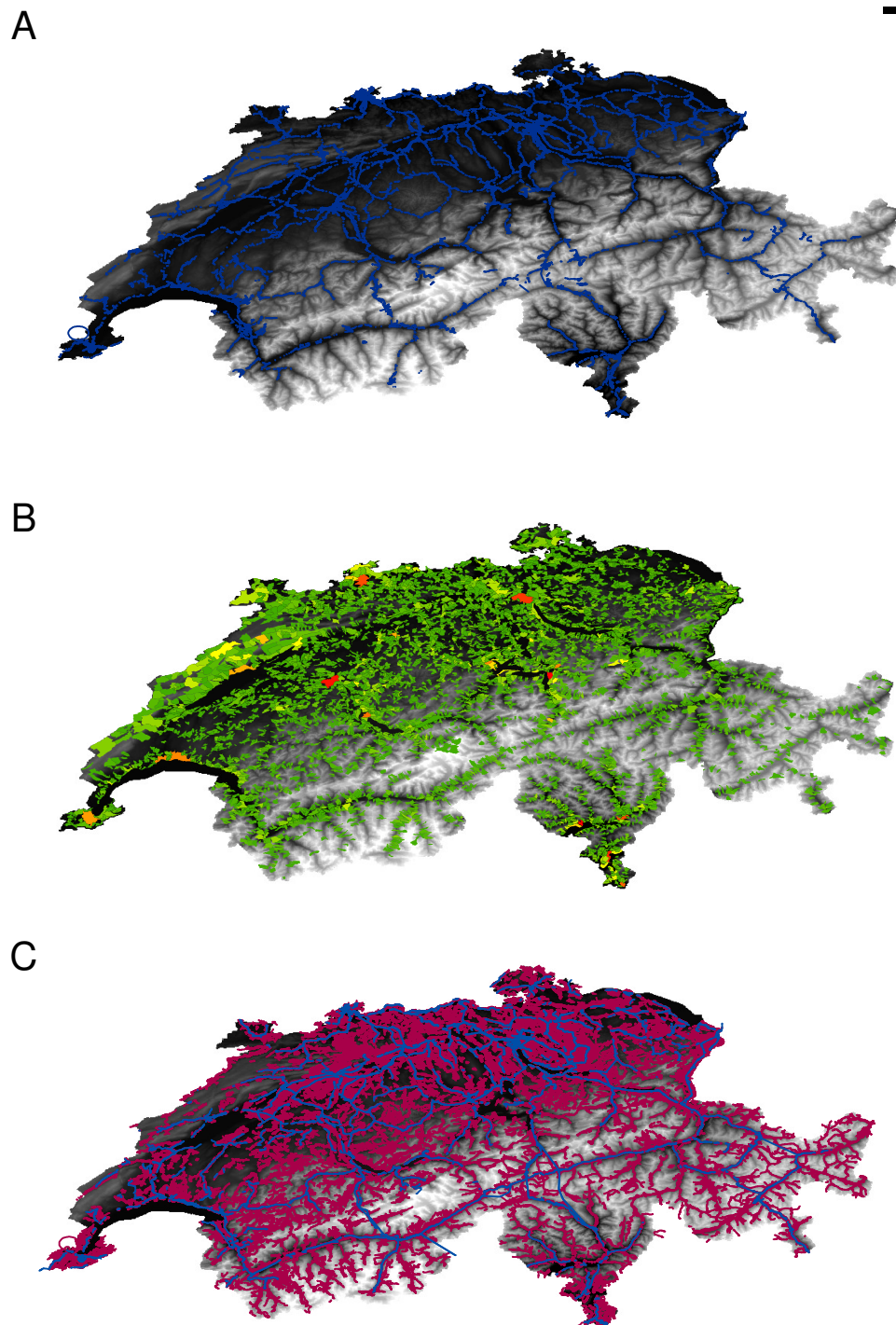


Figure 14: Examples of potential uses of the database. 14a) Intersecting road and rail segments. 14b) Number of flood and landslide incidences by community. 14c) Road (purple) and Rail (blue) networks which are within 200 m of waterways.

Ultimately, such a STDB can be used to support the quantitative modelling of infrastructure risk. One quantitative method of modelling the relationship between infrastructure, natural hazards and the

environment is to use data mining. Data mining uses historic records of events to make predictions. A critical feature of data mining modelling is the need for there to be hundreds (as a minimum) of records (individual instance of, for example, damage to bridges) to produce models which can make accurate predictions. These predictions could be in the form of hazard susceptibility (e.g. Kavzoglu et al., 2014; Tehrany et al., 2014), Spatio-temporal hazard prediction (e.g. Dai & Lee, 2003) damage caused to infrastructure due to natural hazards (e.g. Petrucci & Pasqua, 2012) or interruptions to the flow on a network (e.g. Poljanšek et al., 2012). The basic premise of this modelling approach is that there is a variable which you are interested in predicting and a number of explanatory variables which can be used for prediction (Figure 15). For example, you may wish to predict the location of hazard occurrence. The logic behind the data mining approach is that environmental factors (the predictor variables) control the location of hazard occurrence and the environmental attributes in places where hazards have occurred in the past, are the attributes that lead to landslide occurrence. For instance, if most recorded landslides have occurred on slopes of 25° or above then this is where landslides will likely occur elsewhere in the region. Using GIS to sample all the environmental predictors in a range of locations (both locations where landslides have occurred previously, and locations where have not) it is possible to use a data mining algorithm to develop a numerical relationship, between the environmental variables and the location of landslides. This relationship can be applied to map the areas which are more susceptible to landslide occurrence for a given region. Susceptibility is the relative spatial likelihood of the occurrence of landslides.

In many cases, you would like to predict the location of the hazard before it occurs so to implement mitigation strategies. A data-mining approach can be used for the spatio-temporal prediction of hazards. Here data-mining can be used to find the pattern of variable which trigger hazards. To extend the landslide example, the previous model will have identified some areas that are more susceptible to landslide occurrence, however, this does not give a risk estimate. That requires a temporal element (e.g. the probability that there will be a landslide in the next month). This requires a model of the triggering factors. Again, GIS can be used to sample environmental hazard triggers (e.g. spatio-temporal climatic data) and develop models which predict spatio-temporal hazard occurrence. Where these data are predicted (e.g. rainfall or temperature), the model can predict future hazard occurrence.

Where infrastructure behaviour is known, it is possible to develop data mining models which can predict hazard impact. For instance, when there are records of damage to road segments due to earthquakes, it is possible to estimate the location and severity of damage given and epicentre, magnitude and suite of environment data.

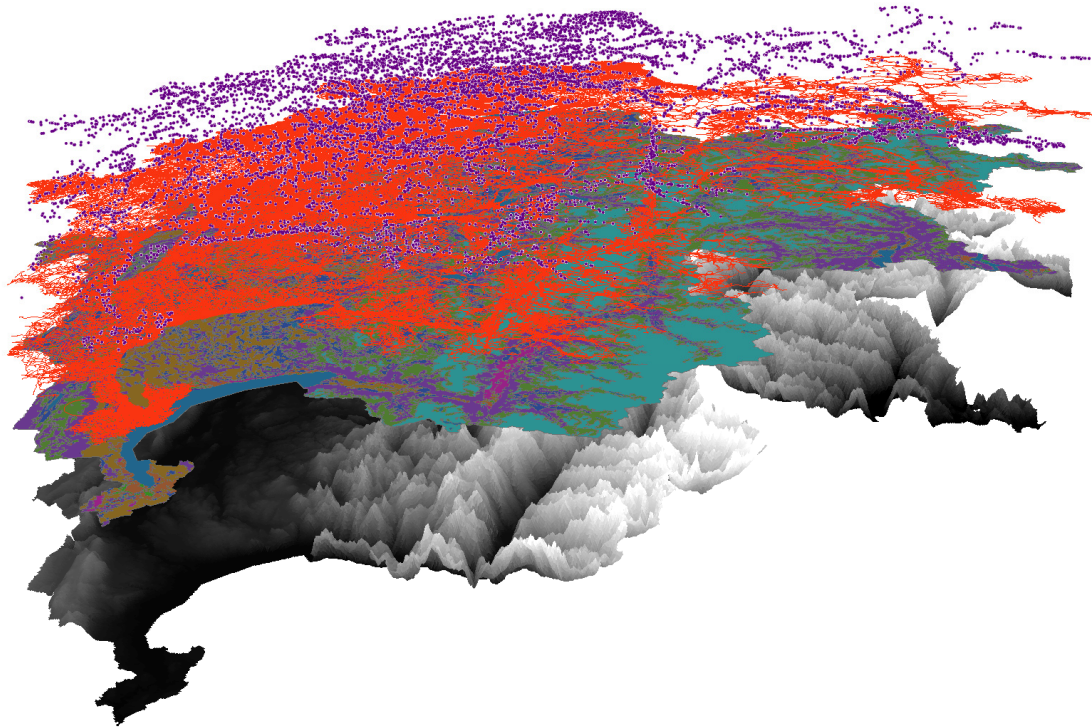


Figure 15: Example of combining data to use for SVM modelling

5 SUMMARY

The relationship between natural hazards, infrastructure and the environment is of interest to a range of disciplines, at a range of scales. This database should contain spatial data, in many forms including polygons (e.g. land use), networks (e.g. road and rail), spatio-temporal data (e.g. daily rainfall) and tabulated data (e.g. hazard attributes). This data can be interrogated using both GIS operations and SQL queries which allows data to be used for modelling purposes. This database would support quantitative modelling of the events which can trigger hazards, these ‘source events’ can include rainfall and temperature. Source events can lead to different ‘hazard events’. The magnitude, location and type of hazard which occurs are determined by both the source event and the environment. The impact on road and rail network can be either seen as either an ‘infrastructure event’, where we are interested in the damage caused to a specific element of the network, such as a road segment, bridge or tunnel or a ‘network event’, where we look at how the hazard has affected the network. This might be in terms of the effect on traffic flow. A reduction in flow can indicate damage or blockage, whereas an increase in flow and subsequent congestion can indicate that the network is not resilient to hazards. This analysis can inform other models interested in ‘consequence events’, which may be the economic losses due to the network being inaccessible or the loss of life due to the damage to routes used by the emergency services.

6 REFERENCES

- Bossard, M., Feranec, J., & Otahel, J. (2000). CORINE land cover technical guide: Addendum 2000.
- Dai, F. C., & 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(5), 527-545
- ESRI 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.
- Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *Pervasive Computing, IEEE*, 7(4), 12-18.
- Halpin, T. (2006). Object-role modeling (ORM/NIAM). In *Handbook on architectures of information systems* (pp. 81-103). Springer Berlin Heidelberg.
- Hilker, N., Badoux, A., & Hegg, C. (2009). The Swiss flood and landslide damage database 1972–2007. *Nat. Hazards Earth Syst. Sci*, 9(3), 913-925.
- Isotta, F.A. et al. 2014: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *Int. J. Climatol.*, 34: 1657-1675. doi: 10.1002/joc.3794. <http://onlinelibrary.wiley.com/doi/10.1002/joc.3794/abstract>
- Kavzoglu, T., Sahin, E. K., & Colkesen, I. (2014). Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, 11(3), 425-439.
- Petrucci, O., & Pasqua, A. A. (2012). Damaging events along roads during bad weather periods: a case study in Calabria (Italy). *Natural Hazards and Earth System Science*, 12(2), 365-378.
- Poljanšek, K., Bono, F., & Gutiérrez, E. (2012). Seismic risk assessment of interdependent critical infrastructure systems: The case of European gas and electricity networks. *Earthquake Engineering & Structural Dynamics*, 41(1), 61-79.
- Pourghasemi, H. R., Jirandeh, A. G., Pradhan, B., Xu, C., & Gokceoglu, C. (2013). Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *Journal of Earth System Science*, 122(2), 349-369.
- President's Commission on Critical Infrastructure Protection. (1997) *Critical Foundations: Protecting America's Infrastructures* <http://www.ciao.gov>
- Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001) Identifying, understanding, and analyzing critical infrastructure interdependencies. *Control Systems, IEEE*, 21(6), 11-25.
- Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology*, 512, 332-343.

Tschoegl, L., Below, R., & Guha-Sapir, D. (2006). An analytical review of selected data sets on natural disasters and impacts. *Centre for Research on the Epidemiology of Disasters*.

Wirtz, A., Kron, W., Löw, P., & Steuer, M. (2014). The need for data: natural disasters and the challenges of database management. *Natural Hazards*, 70(1), 135-157.