Assessment of the quality and uncertainty in disaster loss data

Xavier Romão (FEUP)
Esmeralda Paupério (IC-FEUP)




U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO
institutodaconstrução

IVth International Congress on Risks
23-26 May 2017

Why we need disaster loss data

Disaster loss data include human (casualties, injured, displaced persons, etc), physical (buildings, infrastructures, agriculture, environment, etc) and economic losses resulting from disasters.

We need reliable loss data for

Loss Accounting	Disaster forensics	Risk Modelling
Recording of impact	Identify causes	Model future losses
Measuring trends	Lear from the past	DRR and mitigation

The reliability of disaster loss estimates reflects the **quality** of the data:

- Quality is the ability of data to fulfil a certain need or objective
- Lack of quality is the result of the uncertainty of the data or processes being used to quantify loss estimates

Types of uncertainty and terminology

Typically we define 2 types of uncertainty: aleatoric and epistemic uncertainties

This distinction is useful… but sometimes we need a greater refinement

To express aleatoric and epistemic uncertainties, it helps to subdivide the uncertainty analysis into more detailed classes

A general framework for measuring and expressing uncertainty

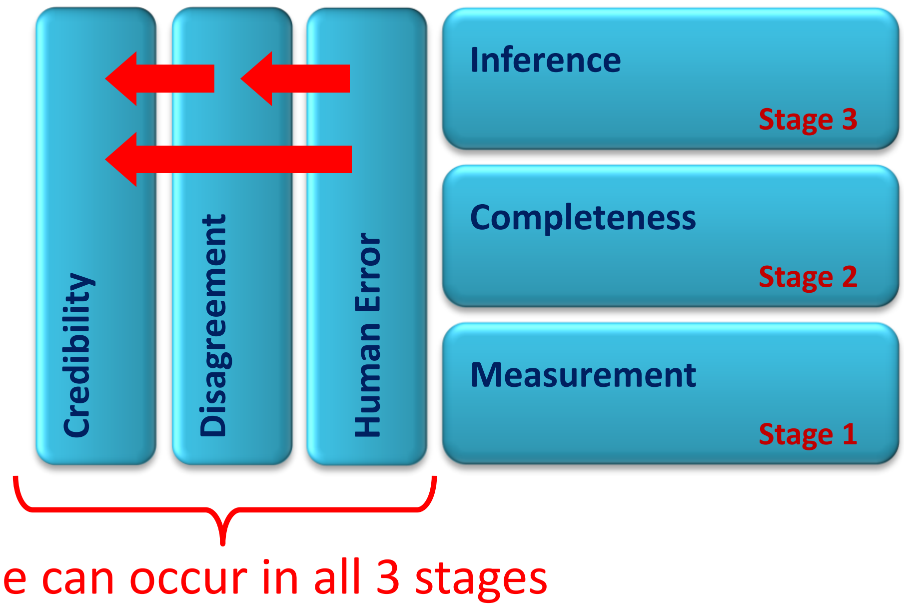
It assumes that uncertainty can exist in 3 different stages (i.e. different states of data processing):

- Stage 1 - Gathering and collecting data
- Stage 2 - Sorting and manipulating data
- Stage 3 - Transforming data to reach the objectives of the process

Types of uncertainty and terminology

The framework defines a relation between 6 types of uncertainty that are connected to aleatoric and/or epistemic factors:

- Measurement (Accuracy, Precision)
- Completeness (Sampling, Missing Values, Aggregation)
- Inference (Modelling, Prediction, Extrapolation into the past)
- Human error
- Disagreement
- Credibility



It assumes data can go through the 3 stages before being suitable to meet a certain objective (e.g. a subsequent decision-making procedure)

This framework and disaster loss indicators

- In disaster loss assessment, some loss indicators may involve data that goes through Stage 1 only (the collected data is the exact data required for decision-making)
  - the number of people killed by an event
    - uncertainty is just that of Stage 1
- In other cases, loss indicators may involve data that goes through Stage 1 and Stage 2 (the collected data needs manipulation before being suitable for decision-making)
  - the number of affected people
    - It is obtained after Stage 1 if data collection is rigorous enough or after Stage 2 if data manipulation is required

      For this last case, uncertainty comes from both Stage 1 and Stage 2

This framework and disaster loss indicators

In a case where the loss indicator is the (direct) monetary losses resulting from damaged properties, there are 2 possible scenarios:

a) the loss data is directly obtained from available sources (e.g. insurance companies based on insurance claims)
b) only part of the loss data is obtained as in a) and the remaining losses must be estimated

In b), part of the data is a proxy for the monetary value of the loss (e.g. damage levels of properties) that has to be transformed into an estimate of the loss

€€€

In this case, the loss indicator will be obtained from Stage 3 and the uncertainty comes from Stages 1, 2 and 3 (assuming that some data manipulation in Stage 2 is also required)

Expressing uncertainty

expressing uncertainty

- quantitative methods
  - The usual approach in traditional science fields where sufficient hard data is available for numerical treatment
- qualitative/quantitative methods
- qualitative methods
  - The only option when data is insufficient and unable to support the meaningful definition of adequate estimates of its variability

Expressing uncertainty

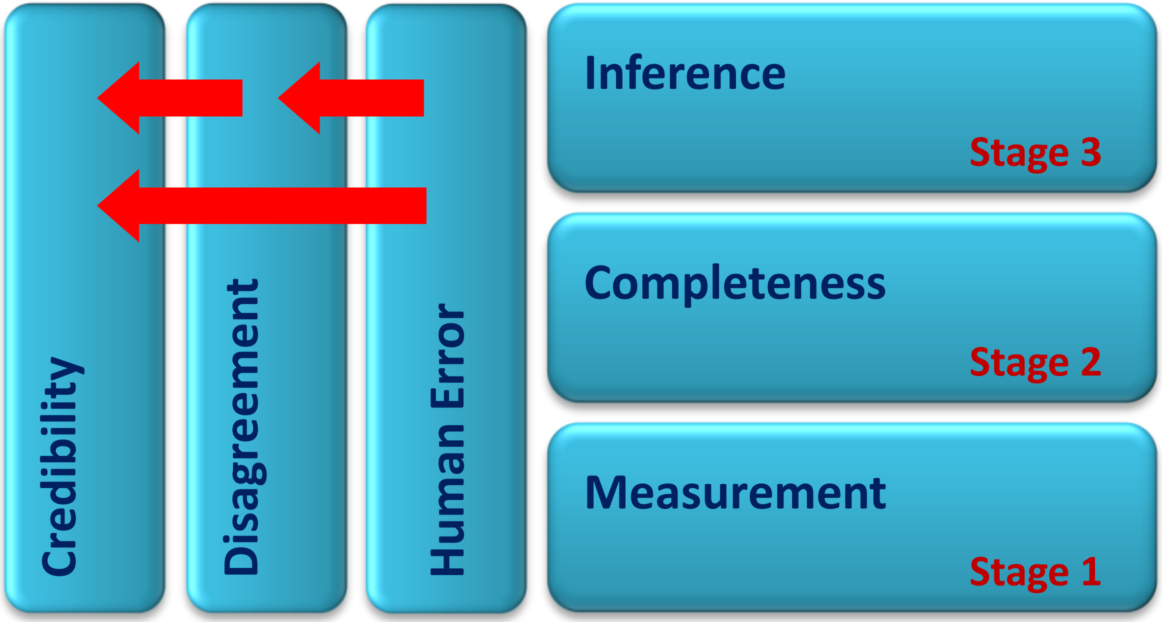
A qualitative/quantitative method: the **NUSAP** method

It uses 5 parameters to characterize data:

- **N**umeral – the value of the data (*quantitative*)
- **U**nit – the unit of measurement of the data (*quantitative*)
- **S**pread – a measure of the uncertainty in the data (e.g. a variance or a range) (*quantitative*)
- **A**ssessment – a global measure of the reliability or confidence in the data (e.g. “reliable/unreliable”, “exact/accurate/estimate/guess”) (*qualitative*)
- **P**edigree – a matrix with a set of criteria used to grade several aspects related to the information flow and the knowledge used to characterize the data (*qualitative*)

Expressing uncertainty – Pedigree Matrix

The Pedigree matrix can be defined to measure qualitatively the several types of uncertainty in the different stages of data processing

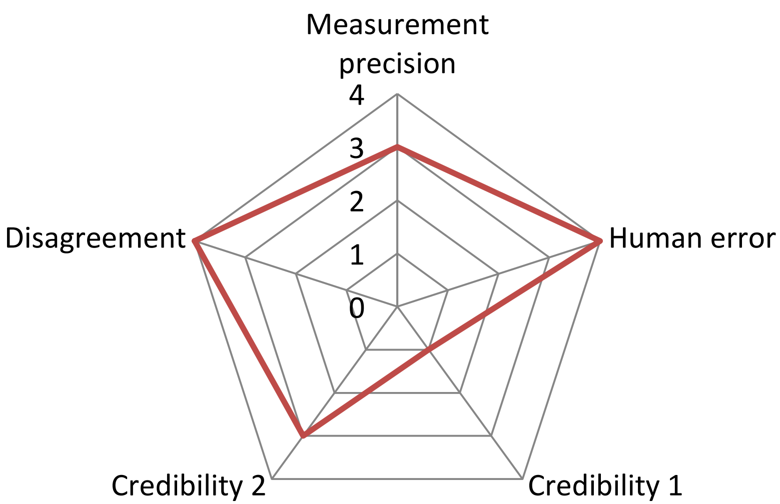


Expressing uncertainty – Pedigree Matrix – Stage 1

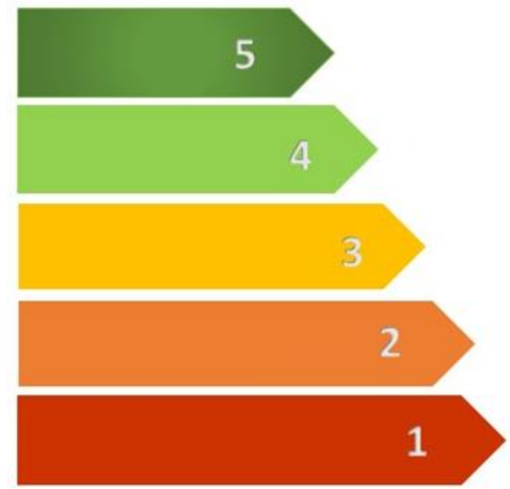
	Criterion				
Grade	Measurement Accuracy	Human error	Credibility 1	Credibility 2	Disagreement
5	Data was collected following an approved standard or by the best available practice	Independent measurements of the same data were performed	The reliability of the source(s) of the data is(are) indisputable (data is based on measurements and was verified)	There is total agreement among peers regarding the procedure used to collect data	There was agreement between all possible data comparisons
4	Data was collected by a reliable method commonly used	Independent measurements of variables closely related to the data were performed	The source(s) of the data is(are) found to be reliable by most people (data is partially based on assumptions or is unverified based on measurements)	A large majority of peers (90–100%) would use this procedure to collect data	There was agreement between the majority of possible data comparisons
3	Data was collected by an acceptable method but there is limited consensus on its reliability	Non independent measurements of the same data were performed	The trustworthiness of the source(s) of the data can't be established (data is unverified and partly based on assumptions)	Many experts (up to 75%) would use this procedure to collect data	There was agreement between some of the possible data comparisons
2	Data was collected by a preliminary or unproven method with unknown reliability	Weak and very indirect measurements of part of the data or data-related variables was performed	Data is a qualified estimate (data was obtained from an expert)	Several experts (up to 50%) would use this procedure to collect data	There was no agreement in any of the possible data comparisons
1	Data was collected by a purely subjective method with no discernible rigour	No additional measurements of the data or data-related variables was performed	Data is a non-qualified estimate or of unknown origin	Few experts (up to 25%) would use this procedure to collect data	No comparison of data was able to be performed

Expressing uncertainty – Pedigree Matrix

By grading each uncertainty component we get:



Average Pedigree score = 3.6



This average score can also be connected to a qualitative expression to define the **A**ssessment part of NUS**A**P

Final remarks

Many of the existing disaster loss databases do not provide uncertainty and quality measures about their data.

The proposed framework builds on an existing uncertainty classification that was combined with the NUSAP method for data characterization. These methods were then adapted for the case of disaster loss data.

The proposed framework is general and flexible in order to include all possible scenarios of data handling. It includes six types of uncertainty that can affect the different stages of data collection and processing

The uncertainty coming from each stage is expressed by one of the components of the NUSAP method: the Pedigree matrix. A global average Pedigree score can then be obtained to reflect the overall uncertainty in a certain loss indicator and to provide a measure of its quality.

Reference: Romão, X., Paupério, E. (2016) A framework to assess quality and uncertainty in disaster loss data. Natural Hazards, 83(2), 1077-1102.