CrossMark

ORIGINAL PAPER

A framework to assess quality and uncertainty in disaster loss data

Xavier Romão[1] · Esmeralda Paupério[2]

Abstract There is a growing interest in the systematic and consistent collection of disaster loss data for different applications. Therefore, the collected data must follow a set of technical requirements to guarantee its usefulness. One of those requirements is the availability of a measure of the uncertainty in the collected data to express its quality for a given purpose. Many of the existing disaster loss databases do not provide such uncertainty/quality measures due to the lack of a simple and consistent approach to express uncertainty. After reviewing existing literature on the subject, a framework to express the uncertainty in disaster loss data is proposed. This framework builds on an existing uncertainty classification that was updated and combined with an existing method for data characterization. The proposed approach is able to establish a global score that reflects the overall uncertainty in a certain loss indicator and provides a measure of its quality.

Keywords Loss data · Disaster · Uncertainty · Qualitative method

1 Introduction

In recent years, disaster impact analysis methodologies have grown and their importance has gained worldwide recognition (Okuyama and Santos 2014). Among other aspects of disaster analysis, the importance of estimating economic losses resulting from natural or man-made disasters is well known (De Groeve et al. 2013). Adequate accounts of disaster losses yield valuable information for governments and international organizations to make

✉ Xavier Romão
  xnr@fe.up.pt

[1] CONSTRUCT-LESE, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

[2] Construction Institute, CONSTRUCT-LESE, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

�桜 Springer

decisions about providing disaster relief assistance (e.g., how much, when, to whom and in what form). Reliable disaster loss accounts are also fundamental to establish loss trends and spatial patterns. These are then used to measure the success and failure of global policies related to public health and safety. Disaster loss data are also particularly important for defining priorities on what scientific research fields to fund and for evaluating the contribution and effectiveness of scientific advances for disaster mitigation. Furthermore, insurance companies also require reliable disaster loss accounts in their portfolios. Disaster loss data will enable them to guarantee their solvency or to undertake additional measures to alleviate the risk they may face in case of a disaster (by issuing catastrophe bonds, for example).

Even though the importance of disaster loss analyses is unquestionable, the range of economic costs resulting from natural disasters is still difficult to estimate, both in concept and in practice. For example, difficulties are usually found when trying to define objective spatial and temporal boundaries for a given loss analysis. Moreover, some of the more complex aspects of estimating disaster costs are related to the type and definition of losses in itself. Losses are conventionally classified as either direct or indirect losses. These categories can then be further subdivided into tangibles and intangibles losses, according to whether or not such losses can be valued in monetary terms. Quantifying these loss components is a challenge, especially those of indirect and intangible nature. However, the fact that multiple procedures are available to estimate some of these loss components (e.g., see EMA 2002; ECLAC 2003; Hiete et al. 2012; Koedam 2012) also complicates their unbiased quantification. Other difficulties also arise when trying to obtain damage and loss data from publicly available sources that were not designed for this purpose. In this case, reliable and accurate data are difficult to achieve (Dilley et al. 2005; Gall et al. 2009; Wirtz et al. 2014), and contradictory data for the same event can sometimes be found when different sources of information are used (Serje 2012).

Standardized approaches are therefore required for loss quantification methodologies and loss data collection systems (i.e., databases) (De Groeve et al. 2014; Johansson 2015). The purpose of these standards is to obtain more reliable loss estimates that will provide adequate support for the higher-level strategic objectives of disaster loss analyses. Achieving a high level of reliability in disaster loss estimates is seen to depend on two essential factors. One is the reliability of the procedure that quantifies a given loss component, and the second is the availability of adequate and sufficient data to perform such quantification. Both factors can be associated with a characteristic generally termed as *quality*. There are no specific criteria that data or processes must possess to have *quality*. Instead, *quality* is measured according to the ability of that datum or process to fulfill a certain need or objective (ISO 14040 2006; ASQ 2014). This lack of ability to fulfill needs or objectives is found to be the result of the existing *uncertainty* of the data or processes that are used. *Uncertainty* in these components is therefore a source of inaccuracy, errors, subjectivity and leads to failure in achieving a high level of *quality*. Hence, before grading the *quality* of a certain component of a disaster loss assessment framework, the sources of *uncertainty* that are involved must first be characterized.

In light of this discussion, the growing importance of disaster loss data becomes clear. However, the collection of such data also requires processes to characterize their reliability and quality, as highlighted by Smith and Katz (2013) or De Groeve et al. (2014). Currently there are no standardized processes for the specific characterization of uncertainty in disaster loss data. Although some database providers identify factors that need to be accounted for (Wirtz et al. 2014), no standard approach exists so far addressing all the necessary aspects and formalizing a methodology to quantify the level of uncertainty. The

present article therefore proposes a comprehensive approach to characterize different types of uncertainty in disaster loss data according to their sources and occurrence in the data management process. Furthermore, in order to establish a measure of the overall uncertainty in a certain datum, an uncertainty quantification procedure is also proposed. The proposed approaches are based on existing methods of uncertainty analysis and characterization that are adapted to the specificities of disaster loss data. To illustrate the application of the proposed uncertainty classification framework, two case studies are also presented and discussed in detail.

2 Uncertainty: what is it?

Although the importance of uncertainty has been acknowledged throughout human history, its systematic analysis only started in the twentieth century. Numerous research studies across various fields and disciplines have addressed the issue of uncertainty over the years (Van Asselt 2000). As such, uncertainty can be seen to be a term used to account for many concepts (Morgan and Henrion 1990). However, a simple and unified definition of what is uncertainty was never established. Epistemological differences between research fields added to this difficulty. Therefore, several definitions of uncertainty have been proposed that bear some relation to the field where it has been analyzed. To further complicate this scenario, different lexicons use different names for the same thing, and, in some cases, even the same name for different things. As such, scientific literature contains many definitions, descriptions, and typologies of uncertainty. Existing classifications and their (sometimes) confusing nomenclature reflect the differences between research fields which are inevitably driven by their different objectives and their differences in terms of data availability.

Despite the referred complexities, proposals, such as those in the following, have been made in pursuit of an ideal unified definition of uncertainty:

- A state of incomplete knowledge (Cullen and Frey 1999)
- Any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system (Walker et al. 2003)
- Incomplete information about a particular subject (Ascough et al. 2008)
- Lack of confidence in knowledge related to a specific question (Sigel et al. 2010)

These definitions can be seen to associate uncertainty with a certain state of knowledge or lack thereof. But uncertainty is not simply the absence of knowledge since it can occur in scenarios where there is no shortage of information (Van Asselt 2000). One can picture scenarios where having additional information can either decrease or increase the uncertainty level. Additional knowledge regarding a certain process can reveal the presence of uncertainties that were previously unknown or disregarded. Therefore, having additional knowledge can point out the limitations in our understanding of a given process and increase the uncertainty about it. Nevertheless, knowledge and knowledge-related issues are decisive concepts that must be involved when defining a type of uncertainty which is often termed *epistemic* uncertainty. In different research fields, this type of uncertainty has also been addressed using terms such as incertitude (Carey and Burgman 2008) or epistemological uncertainty (Gillund et al. 2008), or even simply as "uncertainty" (Frey and Burmaster 1999; McCann et al. 2006). More simply, the concept of *epistemic* uncertainty

can be illustrated by recognizing that, for a given person with a certain level of knowledge at a given time, a statement about a fact can either be true, false, or uncertain.

Aside from the referred knowledge-related aspects, uncertainty can also result from another category of factors which are generally found to be associated with randomness. This type of uncertainty, often termed *aleatoric* uncertainty, represents the intrinsic random nature of a certain phenomenon. In different research fields, this type of uncertainty has also been addressed using terms such as irreducible uncertainty (Tucker and Ferson 2003), random variability (Bolker 2008), or ontological uncertainty (Gillund et al. 2008). The time between the consecutive occurrence of disasters such as earthquakes and storms of a given intensity is an example of a random phenomenon. However, when estimating the economic losses of disasters, such randomness is no longer important since losses are assessed after the actual occurrence of this particular random phenomenon. Nevertheless, there are other sources of *aleatoric* uncertainty that can affect disaster loss accounts. The variation in the number of people in a certain area over a certain period of time is often referred as an example of *aleatoric* uncertainty (Aven 2008). Therefore, accounting for such uncertainty is fundamental when estimating the size and characteristics of population affected by a disaster. Another source of *aleatoric* uncertainty is related with observation and/or evaluation errors that may occur when collecting loss data (Gardi et al. 2011). These errors fall within the general category of human errors which have been found to be inherently random (Cuschieri 2006; Der Kiureghian and Ditlevsen 2009).

Most contexts where the scientific research analysis and treatment of uncertainty is addressed assume that uncertainty can be expressed using numerically quantifiable metrics (e.g., see Beck 1987; Smith and Shugart 1994; Paté-Cornell 1996; Charles 1998; Walker et al. 2003). However, there are cases where such numerical quantification of uncertainty is not possible, namely when problems are ill-defined, when information is only partial or not quantifiable. Under such conditions, only qualitative descriptions can be established to express uncertainty. Qualitative descriptions involve language-based terminology which, in many cases, is imprecise such as our use of it. For example, vague and context-dependent terms or expressions can impair our understanding about what is being described. As a result of this lack of accuracy, an additional type of uncertainty then arises, termed *linguistic* uncertainty. *Linguistic* uncertainty differs from *aleatoric* and *epistemic* uncertainties since it is not a property of the data under analysis, and it is not created by processing available data. Instead, *linguistic* uncertainty is created when attempting to express information using non-quantitative metrics.

For completeness, a final category of uncertainty is also briefly addressed herein. This fourth category of uncertainty occurs in a decision-making process based on the interpretation of results that were expressed and communicated following a given analysis. In this process, different individuals can have different interpretations of the same data due to subjective judgment or differences in values, beliefs and preferences. Therefore, the variety of interpretation outcomes is an additional source of uncertainty. This type of uncertainty has been defined by Finkel (1990) who termed it *decision* uncertainty. However, it has also been addressed using terms such as value uncertainty (Morgan and Henrion 1990), volitional uncertainty (Bedford and Cooke 2001), decision-making uncertainty (Ascough et al. 2008), human uncertainty (Maier et al. 2008), or human decision uncertainty (Kujala et al. 2013). This type of uncertainty was also termed ambiguity by Kwakkel et al. (2010), following the terminology introduced by Brugnach et al. (2008) when addressing the fact that different individuals may use different frames of reference to interpret the same data. The term ambiguity in this context must not be mistaken with the same term being used to define one of the components of *linguistic* uncertainty, following the terminology of

Regan et al. (2002). As referred, *decision* uncertainty only appears in the context of decision-making processes after data have been analyzed and results have been expressed. Therefore, additional details regarding its treatment and representation are not addressed herein as it falls outside the scope of the current article.

3 Classifying uncertainty

The quality of results obtained from a disaster loss assessment depends on the uncertainty related with the available data and with data processing operations. Based on the previously described types of uncertainty, *epistemic* and *aleatoric* uncertainties are the general categories found to be the more significant. Still, if these uncertainties need to be expressed in qualitative terms, *linguistic* uncertainty must also be considered relevant. From a practical standpoint, it is also important to categorize these uncertainties according to their potential for being reduced or eliminated. Given its random nature, eliminating *aleatoric* uncertainty is not feasible (although in some cases it can be reduced in the statistical sense of obtaining a lower variance). At best, this type of uncertainty can be described by a statistical model. On the contrary, since *epistemic* uncertainty is presumably caused by having an inadequate level of knowledge, it can be reduced or even eliminated by improving the existing knowledge. With respect to *linguistic* uncertainty, Regan et al. (2002) suggest several ways to reduce this type of uncertainty. Options involve providing precise numerical definitions for vague terms and carefully specifying the context of terms and their meaning when these terms are ambiguous. Establishing precise numerical definitions for terms such as "low," "medium," and "high" is a popular form of reducing vagueness when expressing uncertainty. This approach, however, may impose a level of precision that analysts find difficult to work with. Patt and Dessai (2005) and Budescu et al. (2009), for example, have shown that scientists and policy makers interpret terms such as "likely" or "very unlikely" in very different ways and continue to do so even after they have read a set of numerical definitions for these terms.

3.1 Existing frameworks to classify uncertainty

Even though a general division between *aleatoric* and *epistemic* uncertainties can be conceptually useful, a greater refinement is sometimes necessary (Morgan et al. 2009). To express *aleatoric* and *epistemic* uncertainties using quantitative or qualitative approaches, it is helpful to subdivide these categories of uncertainties into more detailed classes. Several research studies attempted to establish such classes over the years. For example, MacEachren et al. (2005) published a review of models of information uncertainty and imperfect knowledge in the field of geography, while Thomson et al. (2005) proposed a typology of categories of uncertainty for intelligence information analysis. Another taxonomy for the treatment of uncertainty was also proposed by Regan et al. (2002) for ecology and conservation biology. In the field of health care, Han et al. (2011) proposed a three-dimensional taxonomy that characterizes uncertainty according to its fundamental sources, issues, and locus. Another example can be found in the domain of decision support and policy making for which Walker et al. (2003) proposed a framework to express the uncertainty in a model for decision makers. Following some concerns expressed by Norton et al. (2006) about this framework, Kwakkel et al. (2010) proposed a revised and extended version of this uncertainty classification system. With respect to attempts to express

uncertainty in a common way across several domains, reference is made to the taxonomy proposed by Smithson (1989), and later reviewed by Bammer and Smithson (2008). This taxonomy is useful for distinguishing between different kinds of uncertainty and for demonstrating how different disciplines and practice areas focus on different aspects of uncertainty. Reference is also made to the taxonomy of imperfect information developed by Gershon (1998) and also to the empirical classification proposed by Skeels et al. (2010) for the purpose of information visualization. With respect to *linguistic* uncertainty, although it has been addressed and analyzed by many different researchers (e.g., see (Cleaves 1995; Burgman 2005; Auger and Roy 2008; Carey and Burgman 2008)), the taxonomy proposed by Regan et al. (2002) remains a standard reference.

3.2 A classification framework for expressing the uncertainty of disaster loss data

Given the characteristics of the cross-domain approach proposed by Skeels et al. (2010), it is adopted herein to define an uncertainty classification framework suitable for disaster loss data. Still, this classification is adapted and extended in order to include a particular uncertainty component that is not covered by the original proposal. Furthermore, this classification framework is not developed using the general categories of *aleatoric* and *epistemic* uncertainties as a basis. Instead, it establishes a hierarchy and connectivity between five types of uncertainty that can be related to factors that are *aleatoric* and/or *epistemic* in nature. The five uncertainty types that Skeels et al. (2010) propose are:

- Measurement Precision
- Completeness
- Inference
- Disagreement
- Credibility

Based on the description of the "Measurement Precision" uncertainty type provided in (Skeels et al. 2010), it is found that a more adequate designation would be "Measurement" since this category is supposed to cover aspects related to both *precision* and *accuracy*. Therefore, the first category is termed Measurement hereon. The Skeels et al. (2010) classification also establishes that, in a given process (e.g., a disaster loss assessment), uncertainty can exist in different stages of that process. In this context, this framework characterizes a process using three stages, where each one is associated with a more advanced state of data processing. The three stages can be generally defined as:

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

According to the framework of Skeels et al. (2010), each stage is associated with one of the five types of uncertainty. Stage 1 is associated with Measurement, Stage 2 is associated with Completeness, and Stage 3 is associated with Inference. The remaining two types of uncertainty (Disagreement and Credibility) are said to span across all three stages. In addition, it is also found that Disagreement sometimes increases the Credibility uncertainty (Skeels et al. 2010). After a detailed analysis of this classification, it is possible to detect its inability to account for certain mechanisms related to human error. Therefore, the classification framework adopted herein includes a sixth type of uncertainty termed Human Error that is added to the original framework proposed by Skeels et al. (2010). As

Disagreement and Credibility, Human Error also spans across all three previously referred stages. Furthermore, in some occasions, Human Error also leads to an increase in Disagreement and/or Credibility uncertainties. The hierarchy and connectivity between the types of uncertainty covered by the framework adopted herein are illustrated in Fig. 1. To clarify the role of each component of this framework in the global uncertainty of a process, a detailed description of each type of uncertainty is presented in the following. It is noted that such framework assumes that, in a given process, data will need to go through the three stages before being suitable to meet a certain objective (e.g., a subsequent decision-making procedure). However, certain processes may only require Stage 1 (i.e., the collected data are the exact data required for decision-making), or only Stage 1 and Stage 2 (i.e., the collected data needs some manipulation after which it is suitable for decision-making). Further comments on these situations are provided in Sect. 4.

3.2.1 Measurement: Stage 1

This category covers variations, imperfections and limitations in measurements that produce quantitative data. However, the initial category proposed by Skeels et al. (2010) is modified in order to account for two subcategories of uncertainty that were not differentiated originally: accuracy and precision. Accuracy accounts for the closeness between the measurement of a quantity and its true value (JCGM 2008). Hence, accuracy uncertainty addresses the weaknesses of the measurement technique being used and accounts for factors of *epistemic* nature. Precision is related to the closeness of agreement between independent measurements of a quantity under the same conditions. In disaster loss data collection, lack of precision might be due to limitations in the data measurement technique or to random variations found when different persons measure the same data. Based on this description, this type of uncertainty is seen to account for factors of *aleatoric* and/or *epistemic* nature. Both subcategories of uncertainty can, sometimes, be explicitly expressed by a statistical model or a range where the true value is probably in (e.g., using a confidence interval). However, this uncertainty is often not able to be represented since only the measured data that are known to be imprecise are available.

3.2.2 Completeness: Stage 2

This category addresses the uncertainties related to having incomplete data. According to the description from Skeels et al. (2010), this category is represented by three subcategories of uncertainty: sampling, missing values, and aggregation. Sampling is a strategy

Fig. 1 Hierarchy and connectivity between types of uncertainty
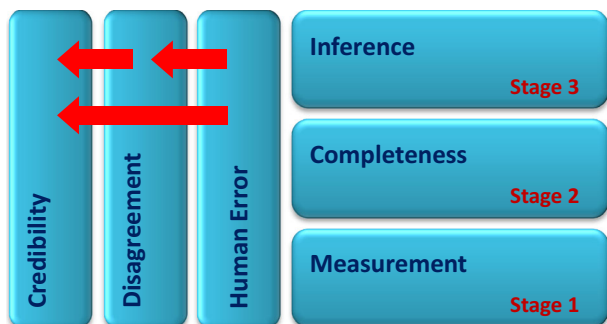
where a subset of individuals from a statistical population is selected in order to estimate characteristics of the whole population. Therefore, completeness uncertainty will inevitably exist when generalizing these estimates to the whole population. Such uncertainty is *aleatoric* if the sample (i.e., the subset of the whole population) is randomly selected. However, if a specific sample is selected instead (e.g., based on a set of pre-defined criteria) the selection procedure may introduce *epistemic* uncertainty due to the potential inadequacy of the criteria. For example, this particular issue can occur when selecting parameters or variables to measure a particular phenomenon that will later be used for analysis (i.e., inference). If the selected parameters are inadequate, an incomplete sample of data will be obtained that will introduce *epistemic* uncertainty into the later analyses. Furthermore, it is noted that, sometimes, the available data represent the entire population. In this case, the data are fully adequate with respect to the completeness uncertainty.

Missing values in the data under analysis also lead to completeness uncertainty, but their effect must be distinguished from those arising from sampling. Missing values are intended to be included but are not present in the data. On the other hand, sampling implies deliberate extrapolation from a few values to cover a larger set of possible values. Datasets with information that is known to be erroneous should be considered incomplete since one obtains a subset of data with missing values after removing the incorrect values. Since this type of uncertainty is related to having inadequate data to perform a given analysis, it is categorized as being of *epistemic* nature.

Aggregating (i.e., summarizing) data is an irreversible procedure also causing uncertainty. Once data have been aggregated, part of the information is lost and data are no longer complete (Skeels et al. 2010). As for the previous case, this type of uncertainty is also related to having inadequate data to perform a given analysis. Therefore, it is categorized as being of *epistemic* nature.

3.2.3 Inference: Stage 3

In a general three-stage process, inference assigns a meaning to the data. Therefore, outcomes of inference are inputs for a decision-making procedure that may follow. Inference is a broad category and may involve fitting the data into a model or transforming the data using a model to estimate new data. According to the description in (Skeels et al. 2010), inference also includes three subcategories of uncertainty: modeling, prediction, and extrapolation into the past.

Modeling uncertainty is introduced when the model being considered is not an adequate representation of the data properties under analysis, i.e., if the model does not reflect the causal relations that produce the phenomenon being examined. This includes models of any kind such as physical models, probabilistic models, hypothesis-testing, diagnostic models, or expert opinions. Prediction involves inferring future events by creating a model for the causal relationship between current or past data and future occurrences. As for the previous case, uncertainty is introduced when the model being considered is not able to represent future outcomes of the phenomenon under analysis. Likewise, uncertainty from extrapolation into the past involves the use of data to reproduce or make inferences about past events. Again, uncertainty is introduced when the model being considered is not able to represent past outcomes of the phenomenon under analysis.

As can be seen, all three categories of uncertainty are directly related to the adequacy of the model being used to establish the required results. The difference between the three types of uncertainty is only at the level of what kind of inference is being performed with the model. Modeling uncertainty occurs when the inference being made is about the

present (i.e., the model is used to reproduce the phenomenon under analysis using the existing data). On the other hand, prediction or extrapolation into the past uncertainties occur when inference is about future or past outcomes of the phenomenon under analysis, respectively, for which there is inadequate data. Since these types of uncertainty reflect the inability to reproduce a phenomenon by lack of capacity or knowledge, there are all found to be of *epistemic* nature.

3.2.4 Human Error: all stages

Human error is a critical element of human activity and professional practice. As previously noted, human errors are considered as a source of *aleatoric* uncertainty and can occur in any activity of the previous three stages that involves people. Even though this uncertainty may be difficult to quantify (Kim and Bishu 2006), its classification and analysis has been addressed using several different approaches (e.g., see reviews by Whittingham (2004) and Rausand (2011)). In order to express more clearly the uncertainty associated with human errors, it is helpful to describe them using a more detailed and categorized approach. Within the scope of the present framework, human errors are considered to be random events that are either unintentional or deliberate. In this context, the taxonomy proposed by Reason (1990) is found to be the more adequate approach to categorize human errors for the proposed uncertainty classification. According to Reason (1990), there are four categories of human errors. The first three categories reflect unintentional events while the last one reflects a deliberate event. These four categories are:

- Slip—An action that is carried out with the correct intention but a faulty execution.
- Lapse—A failure to execute an action because of a distraction or a lapse of memory.
- Mistake—A correct execution of an incorrect intention. A person may believe an action being carried out is correct when, in fact, it is wrong.
- Violation—A person intentionally applies a rule or a procedure that is different than what is known to be required. A violation may be executed with good or bad intention.

3.2.5 Disagreement: all stages

Disagreement can create uncertainty in any of the previously defined three stages. At Stage 1, disagreement happens when a parameter is measured multiple times or is obtained from different sources and the measurements are not the same (as a result of human error or any other cause). At Stage 2, disagreement may occur, for example, when several non-identical but partially overlapping datasets representing the same phenomenon are available. At Stage 3, disagreement can occur when two (or more) different conclusions are drawn from the same data. This can happen when two (or more) experts analyze a certain dataset and come to different conclusions (again, as a result of human error or other causes). In other scenarios, it can happen when different mathematical models are applied to a certain dataset to perform an inference. The *aleatoric* or *epistemic* nature of the disagreement uncertainty depends on the nature of the factors leading to such uncertainty. For example, if the source is related to human error uncertainty, which is *aleatoric*, the resulting disagreement uncertainty will also be of *aleatoric* nature. A similar reasoning can be established for the precision uncertainty of Stage 1 which can be *aleatoric* and/or *epistemic*, thus leading to disagreement uncertainty of the same nature. A similar conclusion can be drawn with respect to the sampling uncertainty of Stage 2. On the other hand, since the remaining uncertainties of Stage 2 (missing values and aggregation), the Stage 1 accuracy uncertainty

and the Stage 3 uncertainties are all *epistemic*, the consequent disagreement uncertainty that may follow is also *epistemic*.

3.2.6 Credibility: all stages

Credibility can also lead to uncertainty in any of the previously defined three stages. This type of uncertainty can be the result of a source of information that produces data in conflict with other data or that produced unreliable data in the past. For example, data with errors can lead to concerns about the correctness of other datasets coming from the same source. Sources of information can be human (e.g., individuals or institutions) or non-human (e.g., machines, measurement tools, models), and credibility issues can be cast on both of them in different forms. For example, credibility could be questioned due to the methods used to get the data or due to concerns involving biases or conflicts of interest with the data creators. A human source may also be considered untrustworthy based on past behavior. Likewise, machines or measurement tools can also be considered untrust-worthy based on past behavior. In this case the credibility appears to be similar to measurement uncertainty. However, the difference is that credibility is a judgment made by the information user about the information source, rather than a known accuracy/precision limitation mathematically expressible. As for the previous case, the *aleatoric* or *epistemic* nature of the credibility uncertainty depends on the nature of the factors that lead to such uncertainty. Furthermore, it is also noted that credibility and disagreement are often associated. When disagreement occurs, whether among people or among measurements, credibility is often called into question. Likewise, when human error occurs, credibility issues are also usually cast.

4 Aspects related to the application of the proposed uncertainty classification framework to disaster loss data

As previously noted, the proposed framework assumes that, in a given process, data go through three stages before being suitable to meet a certain objective. However, for the particular case of characterizing disaster losses, the data being collected in Stage 1 can have two roles. It can represent the actual loss indicator or it can be an auxiliary parameter that will serve as a proxy for the required loss data indicator. For example, when referring to human losses (e.g., the number of deaths), the loss indicator corresponds in many cases to the data being collected (De Groeve et al. 2014). Therefore, the existing uncertainty in the data for this case is only that which comes from Stage 1. Another example of a situation also related to human losses can be defined for the case where the loss indicator now represents the number of affected people (De Groeve et al. 2014). In this case, the final value of the loss indicator can be obtained after Stage 1 (e.g., if the data collection process is rigorous enough) or after Stage 2 if some data manipulation is required, e.g., see (ACAPS 2012; Koedam 2012). In this latter case, the existing uncertainty in the data comes from both Stage 1 and Stage 2.

In another example, if the loss indicator corresponds to the (direct) monetary losses resulting from damaged properties (De Groeve et al. 2014), two possible scenarios can be foreseen. In the first scenario, the total loss data are directly obtained from available sources (e.g., insurance companies) that provide the true monetary losses (e.g., based on insurance claims). In the second scenario, only part of the loss data are obtained as in the

first scenario and the remaining monetary losses must be estimated. In this second scenario, part of the collected data that are available are not the actual loss indicator but a proxy (e.g., damage levels of properties) that needs to be transformed into an estimate in the unit of the required loss indicator (the monetary value). Therefore, in this scenario, part of the loss indicator value will be established in Stage 3 and the uncertainty will come from Stages 1, 2 and 3. In this scenario, it is assumed that some data manipulation in Stage 2 is required before Stage 3; otherwise, only the Stage 1 and Stage 3 uncertainties are involved. Models developed to estimate monetary losses resulting from disasters can be found, for example, for the case of floods (Pistrika 2010; Pistrika et al. 2014; Vozinaki et al. 2015), for the case of earthquakes (Wu et al. 2012; Jaiswal and Wald 2013), or for the case of hurricanes (Hallegatte 2008; Pan 2015; Smith and Matthews 2015).

5 Expressing the uncertainty of disaster loss data

Most standard statistical techniques that have been developed to handle uncertainty assume that it is due to variations in phenomena that can be precisely (i.e., numerically) measured. Such techniques usually consider that a data distribution reflecting this uncertainty is available to allow the use of numerical simulation methods for uncertainty quantification and propagation. For this category of uncertainty analysis, it is, therefore, possible to use methods such as those based on Monte Carlo analysis, Latin Hypercube sampling, importance sampling, variance reduction techniques, perturbation analysis, sensitivity analysis, response surface-based approaches, the Fourier amplitude sensitivity test, the Sobol' variance decomposition, or fast probability integration (e.g., see (Helton and Davis 2003; Saltelli et al. 2004; Sudret 2007; Lemaire 2009; Smith 2014)). In addition, methods using non-probabilistic approaches such as those based on interval analysis or fuzzy analysis (e.g., Ayyub and Klir 2006; Hayes 2011) are also available for this category of problems.

The power and validity of these numerical methodologies is unquestionable. However, they are, usually, only suitable for traditional science fields where sufficient hard data are available for numerical treatment. On the contrary, disaster loss data are often coarse and scattered, thus precluding the use of such refined mathematical manipulations. In other words, available data are frequently insufficient, thus unable to support the meaningful definition of adequate statistical descriptors suitable for mathematical treatment. In such cases, defining qualitative expressions of uncertainty is often the only available option. Qualitative expressions of uncertainty are more difficult to define unequivocally, as well as more difficult to use in a numerical uncertainty propagation analysis. However, they have the potential to be more informative than statistical descriptors since they can include a large number of attributes (Norton et al. 2006).

A suitable methodology to express the uncertainty in disaster loss data must, therefore, be able to accommodate both quantitative and qualitative measures of the uncertainty in a certain datum. To determine how reliable the datum is for a given purpose (i.e., to reflect the quality of the datum), the methodology must be able to express this uncertainty in a clear and meaningful way. Potentially fitting methodologies of this type have been analyzed, for example, by van der Sluijs et al. (2004) and Refsgaard et al. (2007). Based on their descriptions and reviews, the Numeral Unit Spread Assessment Pedigree (NUSAP) method is found to be suitable to characterize the uncertainty in disaster loss data. The NUSAP method has the ability to capture both quantitative and qualitative dimensions of

uncertainty and to represent them in a standardized and self-explanatory way. This method was originally proposed by Funtowicz and Ravetz (1990) to characterize and assess the multidimensional uncertainty in science for policy. Nevertheless, it has also been successfully used and adapted in other research and science fields (van der Sluijs et al. 2005; Costanza 2007; Boone et al. 2009; Colli et al. 2009; Boone et al. 2010; Durugbo et al. 2010; Matschewsky 2013; Henriksson et al. 2014; Lorenz et al. 2015).

The NUSAP method involves five parameters that are used to characterize a certain datum. The five parameters are *Numeral*, *Unit*, *Spread*, *Assessment*, and *Pedigree*. According to Funtowicz and Ravetz (1990), *Numeral*, *Unit*, and *Spread* address the quantitative aspects of the datum, while *Assessment* and *Pedigree* are assigned to describe its more qualitative components. *Numeral* and *Unit* characterize the value of the datum, while *Spread*, *Assessment*, and *Pedigree* characterize its uncertainty and quality. Depending on the datum under analysis, *Numeral* can be defined using an ordinary number representing a mean value or a best estimate. Alternatively, it can also be defined using a more general quantity such as an expression of a number (e.g., a million). Parameter *Unit* usually expresses the scale of *Numeral* by defining its unit of measurement, but it can also contain additional information such as the date of the evaluation. According to Funtowicz and Ravetz (1990), *Spread* is expected to represent the more quantifiable component of the uncertainty of the datum under analysis. Therefore, if sufficient data are available, *Spread* can be defined by the variance of the data, which could be determined by adequate statistical methods. However, data may often be insufficient to establish a meaningful statistic representing the variability of the datum. In some cases, only an interval or a range of variation of the datum can be defined. In this case, the variability can be established using mathematical procedures or expert elicitation.

Assessment is the first parameter of NUSAP expressing qualitative judgments about the datum. *Assessment* can be used to establish a global measure of expert judgement about the overall goodness, reliability or level of confidence associated with the value in *Numeral* or, if desired, in *Spread* instead. For example, this qualitative grade can be defined using qualifiers such as "optimistic/pessimistic", "reliable/unreliable", "official/unofficial", or "exact/accurate/estimate/guess". The final parameter of NUSAP, *Pedigree*, is a concept first introduced in uncertainty analysis by Funtowicz and Ravetz (1990). *Pedigree* involves a set of criteria to assess several aspects related to the information flow and the knowledge used to characterize the datum under analysis. *Pedigree* is a matrix where problem-specific criteria are graded according to a numerical scale with a description assigned to each value of the scale. Therefore, the *Pedigree* matrix quantifies qualitative assessments associated with different components of the uncertainty involved in the process being analyzed. The structure of the *Pedigree* matrix has no formal requirements since the rating scale as well as the number and type of criteria are selected according to the needs of each problem.

6 Pedigree matrices to express the uncertainty in disaster loss data according to the proposed framework

To express the uncertainty in a certain loss data indicator using the framework proposed in Sect. 3.2 with the NUSAP method, a set of *Pedigree* matrices need to be established. Three *Pedigree* matrices in agreement with Stages 1, 2 and 3 are presented in Tables 1, 2 and 3, respectively. For each stage, the corresponding *Pedigree* matrix addresses the uncertainty components that were previously defined. Since each loss indicator being considered might

Table 1 Pedigree matrix for Stage 1—gathering and collecting data

Grade	Criterion				
	Measurement Accuracy	Human error	Credibility 1	Credibility 2	Disagreement
5	Data were 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 are based on measurements and were verified)	There is total agreement among peers regarding the procedure used to collect data	There was agreement between all possible data comparisons
4	Data were 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 are partially based on assumptions or are 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 were 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 cannot be established (data are 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 were 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 are a qualified estimate (data were 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 were collected by a purely subjective method with no discernible rigor	No additional measurements of the data or data-related variables was performed	Data are 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

Table 2 Pedigree matrix for Stage 2—Sorting and manipulating data

Grade	Criterion					
	Completeness			Human error	Credibility	Disagreement
	Sampling	Missing values	Aggregation			
5	Representative data were taken or sampled from datasets collected from a large number of sites/sources	The collected data have no missing values	Aggregated data are fully fit for assessment purposes and capture all the necessary features of the original data	All the procedures for data manipulation were replicated by different people for cross-validation	There is total agreement among peers regarding the procedure used to manipulate the data	There is agreement between all comparable datasets
4	Representative data were taken or sampled from datasets collected from a sufficient number of sites/sources	The collected data have an irrelevant number of missing values	Aggregated data are fit for assessment purposes and capture the most important features of the original data	Some of the procedures for data manipulation were replicated by different people for cross-validation	A large majority of peers (90–100 %) would use this procedure to manipulate the data	There is agreement between the majority of the comparable datasets
3	Representative data were taken or sampled from datasets collected from a small number of sites/sources	The collected data have some missing values but are still fit for use	Aggregated data are fit for basic assessment purposes and capture some features of the original data	Some of the procedures for data manipulation were cross-checked by different people	Many experts (up to 75 %) would use this procedure to manipulate the data	There is agreement between some of the comparable datasets
2	Incomplete data were taken or sampled from datasets collected from a sufficient number of sites/sources	The collected data have a significant amount of missing values and might not be fit for use	Aggregated data are not entirely fit for basic assessment purposes and capture only a few features of the original data	A weak and very indirect validation of some procedures for data manipulation was carried out	Several experts (up to 50 %) would use this procedure to manipulate the data	There is no agreement with any of the comparable datasets
1	Incomplete data were taken or sampled from datasets collected from a small number of sites/sources	The collected data have too many missing values to be fit for use	Aggregated data are unfit for assessment purposes and do not capture features of the original data	No cross-validation of the procedures for data manipulation was performed	Few experts (up to 25 %) would use this procedure to manipulate the data	No datasets can be used to make a comparison

Table 3 Pedigree matrix for Stage 3—transforming data to reach the objectives of the process

Grade	Criterion			
	Inference	Human error	Credibility	Disagreement
5	Inferences were made using an established theory validated by many tests and representing fully understood causal mechanisms	All the inference procedures were replicated by different people for cross-validation	There is total agreement among peers regarding the procedure and model used to make inferences	There is agreement between the results of all comparable analyses
4	Inferences were made using a theoretical model validated by few tests and representing hypothesized causal mechanisms	Some of the inference procedures were replicated by different people for cross-validation	A large majority of peers (90–100 %) would use this procedure and model to make inferences	There is agreement between the results of the majority of the comparable analyses
3	Inferences were made using a computational model involving engineering approximations and approximated causal mechanisms	Some inference procedures were cross-checked by different people	Many experts (up to 75 %) would use this procedure and model to make inferences	There is agreement between the results of some of the comparable analyses
2	Inferences were made using statistical processing or involving simple correlations and representations of causal mechanisms	A weak and very indirect validation of some inference procedures was carried out	Several experts (up to 50 %) would use this procedure and model to make inferences	There is no agreement of results with any of the comparable analyses
1	Inferences were made using definitions involving weak representations of causal mechanisms	No cross-validation was performed	Few experts (up to 25 %) would use this procedure and model to make inferences	No results can be checked by performing comparable analyses

involve different processes of data collection and processing, the set of *Pedigree* matrices that need to be analyzed depend on the indicator. For example, if the value of a loss indicator is defined directly after Stage 1 or Stage 2, only the *Pedigree* matrices of these stages need to be analyzed.

As can be seen, the *Pedigree* matrix for Stage 1 (Table 1) presents two criteria related to credibility. One addresses the credibility of the source of the data while the other addresses the credibility among peers regarding the procedure that is used to collect data. Furthermore, this matrix does not include precision uncertainty in the measurement category. Although both accuracy and precision contribute to the uncertainty in Stage 1, accuracy concerns are currently more relevant, namely due to the lack of standardized approaches for disaster data collection. For example, the issues addressed by Daniell et al. (2013) regarding the death toll of the 2010 Haiti earthquake are found to be related to accuracy problems. Similarly, the concerns in (Molinari et al. 2014) regarding the usability of available flood damage data in Italy for loss assessment purposes are also found to be related to accuracy issues. Furthermore, even though the development of remote sensing technologies for disaster damage collection is growing fast, their reliability and accuracy is still an important factor to consider (e.g., see Joyce et al. 2009; Liou et al. 2012; Zhang et al. 2013; Wang et al. 2015; Foulser-Piggott et al. 2016). As such, uncertainty related to measurement accuracy issues is currently expected to have a larger effect on the reliability of the disaster loss data being collected.

With respect to the aggregation criterion of the Stage 2 *Pedigree* matrix (Table 2), it is noted that its meaning in the context of disaster loss data is twofold. In a first situation, aggregation can represent a procedure where the original data are summarized and transformed into an equivalent dataset. In this case, a model (statistical or other) can also be used as a proxy to represent the equivalent dataset. In a second situation, aggregation can represent the process of transforming data from several sources into a single new dataset. Regarding Stage 2, it is noted that if the collected dataset does not require manipulation before entering Stage 3, only the Stage 1 and Stage 3 *Pedigree* matrices are required.

As can be seen, these *Pedigree* matrices combine different uncertainty identification and criteria specific for loss data and cover the whole process of data collection and recording. Still, potential sources of *linguistic* uncertainty need to be addressed since these matrices express the uncertainty components using qualitative terms. In this case, such sources were carefully analyzed to minimize the influence of *linguistic* uncertainty and to provide a consistent (e.g., between assessments or even between assessors) and objective approach for uncertainty analysis.

Fig. 2 Example of a graphical plot for the *Pedigree* grades of the Stage 1 uncertainty components
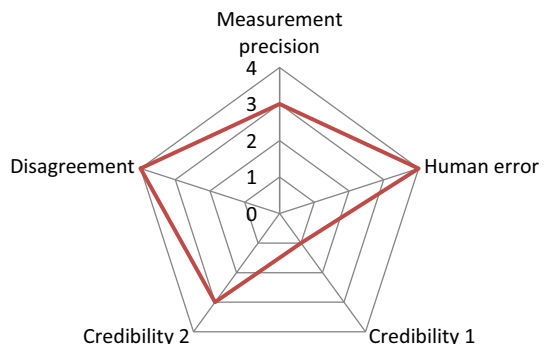
Fig. 3 Global average Pedigree score representing the quality of a given methodology of loss data collection

After grading each criterion, a graphical representation of the grades of each uncertainty component can also be established for each *Pedigree* matrix as presented in Fig. 2. Furthermore, an average score of the *Pedigree* matrix can be established for each stage of the process and a global average *Pedigree* score (i.e., the average of all *Pedigree* matrices scores) can be established to reflect the overall *quality* of the process that lead to the datum under analysis (Costanza 2007). Currently, the average score of the *Pedigree* matrix of each stage is proposed to be obtained without weighting differently the different uncertainty components. Likewise, the global average *Pedigree* score is also obtained without weighing differently the contribution of each stage. This approach was selected because, currently, there is no clear evidence indicating that certain uncertainty components are more influent than others. Furthermore, there is also no indication that a certain stage has a different contribution than another to the overall uncertainty of a certain datum. A possible representation of the global *Pedigree* score is shown in Fig. 3, following the approach suggested by De Groeve et al. (2014).

7 Application examples of the proposed framework

To illustrate the application of the proposed framework for uncertainty classification, two examples are presented in the following. The selected examples are based on recent disasters and both address the uncertainty in the number of deaths resulting from those events. Although the two examples address the same indicator (number of deaths), their uncertainty analysis is considerably different due to the distinct scale of the two disasters. Details about the classifications assigned to the uncertainty criteria of the analyzed *Pedigree* matrices are first presented. The uncertainty of both loss indicators is then also established using the NUSAP method to clarify the application of the proposed framework. In this application, the *Assessment* parameter is rated according to "exact/accurate/estimate/guess".

7.1 The 2010 Xynthia storm: uncertainty classification for the death toll in France

Xynthia was a storm that hit Western Europe between February 27, 2010 and March 1, 2010. The storm reached the west coast of France on February 28 and led to large floods in the coastal areas. In Europe, the storm caused dozens of casualties as well as significant

damage along the Iberian Peninsula, France, Germany, and the Benelux countries. In France, the overall impacts of the storm were 47 deaths, 79 injured, material damage to close to 500,000 people and a total damage cost estimated to be around 2500 million Euros. Detailed analyses and surveys of the impacts of the Xynthia storm can be found, for example, in (Anziani 2010; Bersani et al. 2010; Vinet et al. 2012; Kolen et al. 2013).

The uncertainty analysis of the death toll in France presented herein is mostly based on the data available in (Anziani 2010; Vinet et al. 2012). As referred, 47 people lost their lives in France as a result of the storm: 41 died due to the coastal flooding, while six died from direct or indirect wind effects. Given the scale of the death toll of the Xynthia storm, the number of victims was directly obtained based on the analysis of the corpses. Therefore, as previously referred, the existing uncertainty in the data is only that which comes from Stage 1. Therefore, only the criteria presented in the *Pedigree* matrix of Table 1 are necessary for the uncertainty analysis. The grading of each criterion of the Stage 1 matrix and the corresponding justification is as follows:

- Measurement/Accuracy—Grade 5—Since the Xynthia storm was declared a national disaster by the French authorities, the number of deaths determined by forensic reports is expected to have been established using standard disaster victim identification (DVI) procedures (e.g., see Graham 2006).
- Human error—Grade 5—The use of standard DVI procedures to establish the deaths attributed to the storm complemented by the statements obtained from the fire services (Vinet et al. 2012) are considered sufficient to assign this grade.
- Credibility 1—Grade 5—The official number of deaths is found in a report from the French government (Anziani 2010). Furthermore, the same number of victims was verified during the field survey carried out by Vinet et al. (2012).
- Credibility 2—Grade 5—Collecting the number of deaths based on the report from the French government (Anziani 2010), complemented by the statements obtained from the fire services (Vinet et al. 2012), is considered to be sufficient to assign this grade.
- Disagreement—Grade 5—The statements obtained from the fire services during the field survey carried out by Vinet et al. (2012) are in agreement with the number of deaths determined by forensic reports. As such, it is considered that agreement was found between all possible data comparisons when determining whether the victims could be attributed to the Xynthia storm.

Given the grades assigned to each criterion of the Stage 1 *Pedigree* matrix, the average score that is obtained for Stage 1 is 5. Since only the Stage 1 matrix needs to be analyzed, this score is also the global average *Pedigree* score of the data. Finally, the results of the uncertainty classification according to the NUSAP method are then:

- $N = 47$
- $U =$ number of people
- $S = 0$ (there is no spread since the value of the loss indicator is the exact value)
- $A =$ exact
- $P = 5$

To further illustrate the application of the framework, a second uncertainty analysis is carried out using this example but considering alternative conditions that reflect a common situation in disaster data analysis. In this new analysis, it is assumed that only the number of deaths determined by forensic reports and provided in the report from the French government (Anziani 2010) is available for the uncertainty analysis. In this case, since no

other source of data is available to validate the number of deaths, the classification of the Stage 1 criteria would be as follows:

- Measurement/Accuracy—Grade 5—Since the Xynthia storm was declared a national disaster by the French authorities, the number of deaths determined by forensic reports is expected to have been established using standard disaster victim identification (DVI) procedures (e.g., see Graham 2006).
- Human error—Grade 5—The use of standard DVI procedures to establish the deaths attributed to the storm is still considered to be sufficient to assign this grade.
- Credibility 1—Grade 4—This classification was considered since the number of deaths is only based in the French government report (Anziani 2010) and was not verified.
- Credibility 2—Grade 4—This classification was considered since the number of deaths is only based on the report from the French government (Anziani 2010). According to Kron et al. (2012), every data should be validated for quality assurance purposes.
- Disagreement—Grade 1—The lowest classification is assigned since no data comparison is able to be performed.

Given these new grades assigned to the criteria of the Stage 1 *Pedigree* matrix, the average score that is obtained for Stage 1 is 3.8. As for the previous case, the global average *Pedigree* score of the data is also 3.8. With respect to the results of the uncertainty classification according to the NUSAP method, only the value of P changes to 3.8.

7.2 The 2005 Katrina hurricane: uncertainty classification for the Louisiana death toll

In the morning of August 29, 2005, Hurricane Katrina made landfall along the United States Central Gulf Coast region. The storm was a category 3 hurricane on the Saffir–Simpson scale that caused significant damage. However, the Hurricane Katrina disaster is mostly the result of the cascading effects of the windstorm (Knabb et al. 2005). In particular, numerous failures in the levee infrastructure caused massive flooding of a large urban area (e.g., see (Seed et al. 2006; IPET 2006a; Sills et al. 2008)). As a result of the Hurricane Katrina impact, the Louisiana, Mississippi, Alabama, and Florida states were declared Major Disaster areas (Boyd 2011). The impacts over the greater New Orleans region were particularly severe due to the flooding of approximately 80 % of New Orleans. More than 304,000 houses were damaged in the Greater New Orleans (DHS 2006) and about 1.2 million people were under evacuation orders (Boyd 2011). According to IPET (2006b), flood damage to residential property in New Orleans was estimated at US$16 billion and damage to public structures, infrastructure, and utilities at US$7 billion. With respect to the Louisiana death toll, several estimates varying from 986 (Brunkard et al. 2008) to 3000 (Mutter and Barnard 2009) have been referred across several sources. The death toll established by Boyd (2011) is 1572. Although this number is not the true death toll (Boyd 2011), the process that was used to establish this value is considered herein for the application of the uncertainty classification framework.

Although the loss indicator being analyzed is the number of deaths, the value obtained by Boyd (2011) is the result of a process that is more complex than the one of the previous example. The analysis carried out by Boyd (2011) to establish a more reliable number of deaths caused by Hurricane Katrina started by collecting several data records with potential victims that might be attributed to Hurricane Katrina along with other metadata (e.g., victim attributes, victim recovery location, field survey data). These records were obtained from several sources and are considered herein as sets of raw data with more than just

mortality data. These data were then analyzed, manipulated, and aggregated into a representative dataset that was ultimately used to make inferences about a reliable value for the number of deaths. Given this approach, the number of deaths is considered to have been obtained by a procedure that involves the three stages of the proposed uncertainty classification. Therefore, the *Pedigree* matrices of the three stages are considered for the uncertainty analysis of the Louisiana death toll presented herein. The classification of those *Pedigree* matrices is based on the detailed analysis presented in (Boyd 2011).

For the case of the Stage 1 *Pedigree* matrix, the grading of each criterion and the corresponding justification is as follows:

- Measurement/Accuracy—Grade 4—The raw data that were used to establish the number of deaths were collected from a number of official and unofficial sources (the Louisiana Department of Health and Hospitals (DHH), the State Medical Examiner's Office (SMEO), field surveys, newspaper listings and other media accounts).
- Human error—Grade 5—An independent field survey was carried out to verify cases listed by the SMEO and investigate their characteristics.
- Credibility 1—Grade 4—Most of the data sources are found to be reliable but the field survey was only able to cover part of the data. The field survey was limited to residents, businesses and public places located in the impacted regions of the Orleans and St. Bernard parishes and obtained data from approximately 400 victims.
- Credibility 2—Grade 4—The majority of the data were collected from official sources and the field survey. Still, some data were obtained from unofficial sources.
- Disagreement—Grade 4—According to (Boyd 2011), some of the data obtained from the field survey did not agree with the data from the SMEO (seventeen locations listed did not have markings to indicate a victim had been recovered from that location, and two listed locations had an address that did not exist).

For the case of the Stage 2 *Pedigree* matrix, the grading of each criterion and the corresponding justification is as follows:

- Completeness/Sampling—Grade 4—Data representing the victims and its metadata obtained from the collected records was used to form a global dataset (Boyd 2011). The number of records that was considered is assumed to be sufficient but not large.
- Completeness/Missing values—Grade 3—According to Boyd (2011), the global dataset has missing data for many records. Approximately 200 victims lacked basic information (e.g., cause of death) and detailed victim recovery information is only available for around 800 victims. Nevertheless, the dataset is considered fit for use.
- Completeness/Aggregation—Grade 4—The aggregated dataset is considered to be fit to assess the number of deaths. Even though there are missing data, the dataset represents the most complete listing of Hurricane Katrina victims and corresponding metadata currently available (Boyd 2011).
- Human error—Grade 2—Boyd (2011) refers some of the procedures that were used to compile the global dataset and to identify errors/discrepancies in the data. Since there is no reference to the fact that different people were involved in this compilation, it is considered that only a weak and very indirect validation of the procedures was carried out.
- Credibility—Grade 3—Since the dataset does not include only direct victims of the storm, the criteria to include or not to include certain victims in the dataset were, in some cases, found to be subjective.

- Disagreement—Grade 3—As referred before, this dataset represents the most complete listing of Hurricane Katrina victims currently available. According to Boyd (2011), previous analyses were based on data that only includes victims recovered and processed by the SMEO. Since this dataset includes the SMEO data, it is considered that there is some agreement with other records.

For the case of the Stage 3 *Pedigree* matrix, the grading of each criterion and the corresponding justification is as follows:

- Inference—Grade 2—Boyd (2011) describes the steps used to classify each case according to the circumstances of death established based on the available information. The classification considered three categories: direct flood death, emergency circumstances death, and evacuation/displacement death. Based on the description of the classification process, simple correlations and causal mechanisms were considered for the inference.
- Human error—Grade 3—There is no reference to the fact that different people were involved in the process of defining the value of the final death toll. However, since this study was performed in the context of a PhD research, it is assumed that the supervisor(s) of the research cross-checked some of the results.
- Credibility—Grade 2—Boyd (2011) refers that the death classification process involved some uncertainties and ambiguities.
- Disagreement—Grade 1—As referred before, previous studies were based on data that only includes part of the victims. As such, it is impossible to make a comparison of the results obtained.

Given the grades assigned to each criterion of the Stages 1, 2 and 3 *Pedigree* matrices, the average score that is obtained for each stage is 4.2, 3.2 and 2, respectively. Given these scores, the global average *Pedigree* score of the data is 3.1. Finally, the results of the uncertainty classification according to the NUSAP method are then:

- $N = 1572$
- $U =$ number of people
- $S = +9\ \%$
- $A =$ accurate
- $P = 3.1$

With respect to this classification, it is noted that the *Spread* value (S) is set to 9 % based on the fact that a number of people remain missing and several victims were recovered from unknown locations. According to Boyd (2011), the number of direct flood deaths could be revised to about 9 % more than 1572. Even though the number of deaths established by Boyd (2011) is not the true value, it is still considered accurate.

8 Conclusion

There is a growing interest in the collection of disaster loss data for applications in fields such as disaster loss accounting, disaster forensics or risk modeling (De Groeve et al. 2013). However, the collected data must also follow a set of technical requirements to guarantee its usefulness. One of those requirements is the availability of a measure of the uncertainty in the collected data to express its quality for a given purpose (De Groeve et al. 2014).

Many of the existing disaster loss databases do not provide such uncertainty/quality measures about the data they hold (De Groeve et al. 2013, 2014). The unavailability of a simple and consistent approach to express uncertainty in this field is believed to be one of the reasons for this. After reviewing existing literature on the subjects of uncertainty classification and characterization, a framework to express the uncertainty in disaster loss data was proposed. This framework builds on an existing uncertainty classification (Skeels et al. 2010) that was updated and combined with an existing method for data characterization. This method, the NUSAP method (Funtowicz and Ravetz 1990), includes quantitative and qualitative factors to express the uncertainty of data. These factors were adapted for the case of disaster loss data following the proposed uncertainty classification.

The proposed uncertainty classification aims to be as general and flexible as possible in order to include all possible scenarios of data handling. It includes six types of uncertainty (measurement, completeness, human error, disagreement, and credibility) which can affect the different stages of data collection and processing of a given loss indicator. The uncertainty coming from each stage can then be 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. To illustrate the application of the proposed uncertainty classification framework, two case studies are also presented and discussed in detail.

References

ACAPS (2012) Technical brief—estimation of affected population figures. Assessment Capacities Project (ACAPS), Geneva

Anziani A (2010) Rapport d'information fait au nom de la mission commune d'information sur les conséquences de la tempête Xynthia. Tome I: rapport. No 647, Sénat, session extraordinaire de 2009–2010. http://www.senat.fr/notice-rapport/2009/r09-647-1-notice.html

Ascough J, Maier H, Ravalico J, Strudley M (2008) Future research challenges for incorporation of uncertainty in environmental and ecological decision making. Ecol Model 219(3–4):383–399

ASQ (2014) American society for quality. http://asq.org/glossary/q.html. Accessed Apr 2014

Auger A, Roy J (2008) Expression of uncertainty in linguistic data. 11th International conference on information fusion, Cologne, Germany

Aven T (2008) Evaluation of accident risks—status and trends in risk analysis and evaluation. Swedish Rescue Services Agency, Karlstad

Ayyub B, Klir G (2006) Uncertainty modeling and analysis in engineering and the sciences. Chapman and Hall/CRC, Boca Raton

Bammer G, Smithson M (2008) Understanding uncertainty. Integ Insights 7:1–7

Beck M (1987) Water quality modeling: a review of the analysis of uncertainty. Water Resour Res 23(8):1393–1442

Bedford T, Cooke R (2001) Probabilistic risk analysis: foundations and methods. Cambridge University Press, Cambridge

Bersani C, Gerard F, Gondran O, Helias A, Martin X, Puech P, Rouzeau M, Fleury B, Greff M, Bougere R, Trepos Y (2010) Tempete Xynthia retour d'experience, evaluation et propositions d'action tome I: rapport. http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/104000293/0000.pdf

Bolker B (2008) Ecological models and data in R. Princeton University Press, Princeton

Boone I, Van der Stede Y, Bollaerts K, Vose D, Maes D, Dewulf J, Messens W, Daube G, Aerts M, Mintiens K (2009) NUSAP method for evaluating the data quality in a quantitative microbial risk assessment model for salmonella in the pork production chain. Risk Anal 29(4):502–517

Boone I, Van der Stede Y, Dewulf J, Messens W, Aerts M, Daube G, Mintiens K (2010) NUSAP: a method to evaluate the quality of assumptions in quantitative microbial risk assessment. J Risk Res 13(3):337–352

Boyd E (2011) Fatalities due to Hurricane Katrina's impacts in Louisiana. PhD thesis, University of New Orleans

Brugnach M, Dewulf A, Pahl-Wostl C, Taillieu T (2008) Toward a relational concept of uncertainty: about knowing too little, knowing too differently, and accepting not to know. Ecol Soc 13(2):30

Brunkard J, Namulanda G, Ratard R (2008) Hurricane Katrina deaths, Louisiana, 2005. Disaster Med Public Health Prep 2(4):215–223

Budescu D, Broomell S, Por H (2009) Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. Psychol Sci 20(3):299–308

Burgman M (2005) Risks and decisions for conservation and environmental management. Cambridge University Press, Cambridge

Carey J, Burgman M (2008) Linguistic uncertainty in qualitative risk analysis and how to minimise it. Ann NY Acad Sci 1128:13–17

Charles A (1998) Living with uncertainty in fisheries: analytical methods, management priorities and the Canadian ground fishery experience. Fish Res 37(1–3):37–50

Cleaves D (1995) Assessing and communicating uncertainty in decision support systems: lessons from an ecosystem policy analysis. AI Appl 9(3):87–102

Colli A, Vetere Arellano A, Kirchsteiger C, Ale B (2009) Risk characterisation indicators for risk comparison in the energy sector. Saf Sci 47(1):59–77

Costanza R (2007) Assessing and communicating data quality: toward a system of data quality grading. In: Costanza R, Lisa Graumlich L, Steffen W (eds) Sustainability or collapse? An integrated history and future of people on earth. MIT Press, Cambridge

Cullen A, Frey H (1999) Probabilistic techniques in exposure assessment: a handbook for dealing with uncertainty in models and inputs. Plenum Press, New York

Cuschieri A (2006) Nature of human error: implications for surgical practice. Ann Surg 244(5):642–648

Daniell J, Khazai B, Wenzel F (2013) Uncovering the 2010 Haiti earthquake death toll. Nat Hazards Earth Syst Sci Discuss 1(3):1913–1942

De Groeve T, Poljansek K, Ehrlich D (2013) Recording disaster losses—recommendations for a European approach. scientific and technical research reports, report EUR 26111

De Groeve T, Corbane C, Ehrlich D, Poljansek K (2014) Current status and best practices for disaster loss data recording in EU member states: a comprehensive overview of current practice in the EU member states. Scientific and technical research reports, report EUR 26879

Der Kiureghian A, Ditlevsen O (2009) Aleatory or epistemic? Does it matter? Struct Saf 31(2):105–112

DHS (2006) Current housing unit damage estimates—Hurricanes Katrina, Rita, and Wilma. Department of Homeland, Washington, DC

Dilley M, Chen R, Deichmann U, Lerner-Lam A, Arnold M, Agwe J, Buys P, Kjevstad O, Lyon B, Yetman G (2005) Natural disaster hotspots: A global risk analysis, vol 1. The World Bank, Hazard Management Unit, Washington, D.C.

Durugbo C, Erkoyuncu J, Tiwari A, Alcock J, Roy R, Shehab E (2010) Data uncertainty assessment and information flow analysis for product-service systems in a library case study. Int J Serv Oper Inf 5(4):330–350

ECLAC (2003), Handbook for estimating socio-economic and environmental effects of disasters. United Nations, Economic Commission for Latin America and the Caribbean (ECLAC) and International Bank for Reconstruction and Development (The World Bank)

EMA (2002) Australian emergency manuals series. Part III. Emergency management practice. Volume 3—Guidelines. Guide 11. Disaster loss assessment guidelines. Emergency Management Australia, State of Queensland and Commonwealth of Australia

Finkel A (1990) Confronting uncertainty in risk management: a guide for decision-makers. Center for risk management. Resources for the Future Press, Washington D.C

Foulser-Piggott R, Spence R, Eguchi R, King A (2016) Using remote sensing for building damage assessment: the GEOCAN study and validation for the 2011 Christchurch earthquake. Earthq Spectra 32(1):611–631

Frey H, Burmaster D (1999) Methods for characterising variability and uncertainty: comparison of bootstrap simulation and likelihood-based approaches. Risk Anal 19(1):109–130

Funtowicz S, Ravetz J (1990) Uncertainty and quality in science for policy. Kluwer Academic Publishers, Dordrecht

Gall M, Borden K, Cutter S (2009) When do losses count? Six fallacies of natural hazards loss data. Bull Am Meteorol Soc 90(6):799–809

Gardi A, Valencia N, Guillande R, André C (2011) Inventory of uncertainties associated with the process of tsunami damage assessment on buildings (SCHEMA FP6 EC co-funded project). Nat Hazards Earth Syst Sci 11(3):883–893

Gershon N (1998) Visualization of an imperfect world. IEEE Comput Graph Appl 18(4):43–45

Gillund F, Kjolberg K, von Krauss M, Myhr A (2008) Do uncertainty analyses reveal uncertainties? Using the introduction of DNA vaccines to aquaculture as a case. Sci Total Environ 407(1):185–196

Graham E (2006) Disaster victim identification. Forensic Sci Med Pathol 2(3):203–207

Hallegatte S (2008) An adaptive regional input-output model and its application to the assessment of the economic cost of Katrina. Risk Anal 28(3):779–799

Han P, Klein W, Arora N (2011) Varieties of uncertainty in health care: a conceptual taxonomy. Med Decis Making 31(6):828–838

Hayes K (2011) Uncertainty and uncertainty analysis methods: Issues in quantitative and qualitative risk modeling with application to import risk assessment. Report no EP102467, ACERA project (0705). CSIRO, Hobart, Australia

Helton J, Davis F (2003) Latin hypercube sampling and the propagation of uncertainty in analysis of complex systems. Reliab Eng Syst Saf 81(1):23–69

Henriksson P, Guinée J, Heijungs R, de Koning A, Green D (2014) A protocol for horizontal averaging of unit process data—including estimates for uncertainty. Int J Life Cycle Assess 19(2):429–436

Hiete M, Merz M, Comes T, Schultmann F (2012) Trapezoidal fuzzy DEMATEL method to analyze and correct for relations between variables in a composite indicator for disaster resilience. OR Spectr 34(4):971–995

IPET (2006a) Performance evaluation of the New Orleans and Southeast Louisiana hurricane protection system volume V—the performance—levees and floodwalls. Final report. Interagency performance evaluation task force

IPET (2006b) Performance evaluation of the New Orleans and Southeast Louisiana hurricane protection system volume VII—the consequences. Final report. Interagency performance evaluation task force

ISO 14040 (2006) Environmental management—life cycle assessment—principles and framework. Int Organ Stand, Geneva

Jaiswal K, Wald D (2013) Estimating economic losses from earthquakes using an empirical approach. Earthq Spectra 29(1):309–324

JCGM (2008) JCGM 200 International vocabulary of metrology—basic and general concepts and associated terms (VIM). Joint committee for guides in metrology

Johansson M (2015) Data sources on small-scale disaster losses and response—a Swedish case study of extreme rainfalls 2000–2012. Int J Disaster Risk Reduct 12:93–101

Joyce K, Belliss S, Samsonov S, McNeill S, Glassey P (2009) A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. Prog Phys Geogr 33(2):1–25

Kim B, Bishu R (2006) Uncertainty of human error and fuzzy approach to human reliability analysis. Int J Uncertain Fuzziness Knowl-Based Syst 14(1):111–129

Knabb R, Rhome J, Brown D (2005) Tropical cyclone report: Hurricane Katrina 23–30 August 2005. National Hurricane Center (updated August 2006 and September 2011)

Koedam A (2012) Rapid estimation of affected population figures: desk review. Assessment capacities project (ACAPS), Geneva

Kolen B, Slomp R, Jonkman S (2013) The impacts of storm Xynthia February 27–28, 2010 in France: lessons for flood risk management. J Flood Risk Manag 6(3):261–278

Kron W, Steuer M, Löw P, Wirtz A (2012) How to deal properly with a natural catastrophe database–analysis of flood losses. Nat Hazards Earth Syst Sci 12(3):535–550

Kujala H, Burgman M, Moilanen A (2013) Treatment of uncertainty in conservation under climate change. Conserv Lett 6(2):73–85

Kwakkel J, Walker W, Marchau V (2010) Classifying and communicating uncertainties in model-based policy analysis. Int J Technol Policy Manag 10(4):299–315

Lemaire M (2009) Structural reliability. Wiley, New York

Liou Y, Sha H, Chen T, Wang T, Li Y, Lai Y, Lu L (2012) Assessment of disaster losses in rice paddy field and yield after Tsunami induced by the 2011 Great East Japan earthquake. J Mar Sci Technol 20(6):618–623

Lorenz S, Dessai S, Paavola J, Forster P (2015) The communication of physical science uncertainty in European national adaptation strategies. Clim Change 132(1):143–155

MacEachren A, Robinson A, Hopper S, Gardner S, Murray R, Gahegan M, Hetzler E (2005) Visualizing geospatial information uncertainty: what we know and what we need to know. Cartogr Geograph Inf Sci 32(3):139–160

Maier H, Ascough J, Wattenbach M, Renschler C, Labiosa W, Ravalico J (2008) Chapter five - Uncertainty in environmental decision making: issues, challenges and future directions. Dev Integ Environ Assess 3:69–85

Matschewsky J (2013) Evaluation and optimization of product/service systems within the development process. Technical report. Linköping University, The Institute of Technology, Sweden

McCann R, Marcot B, Ellis R (2006) Bayesian belief networks: applications in ecology and natural resource management. Can J For Res 36(12):3053–3062

Molinari D, Menoni S, Aronica G, Ballio F, Berni N, Pandolfo C, Minucci G (2014) Ex post damage assessment: an Italian experience. Nat Hazards Earth Syst Sci 14(4):901–916

Morgan M, Henrion M (1990) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, Cambridge

Morgan G, Dowlatabadi H, Henrion M, Keith D, Lempert R, McBride S, Small M, Wilbanks T (2009) Best practice approaches for characterizing, communicating, and incorporating scientific uncertainty in decision making. US climate change science program. http://www.climatescience.gov/Library/sap/sap5-2/final-report/default.htm

Mutter J, Barnard K (2009) Climate change, evolution of disasters and inequality. In: Humphreys S (ed) Human rights and climate change. Cambridge University Press, Cambridge

Norton J, Brown J, Mysiak J (2006) To what extent, and how, might uncertainty be defined? Comments engendered by "Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support": Walker et al. Integrated assessment 4: 1, 2003. Integ Assess J Bridging Sci Policy 6(1):83–88

Okuyama Y, Santos J (2014) Disaster impact and input-output analysis. Econ Syst Res 26(1):1–12

Pan Q (2015) Estimating the economic losses of hurricane like in the greater Houston Region. Nat Hazards Rev 16(1):1–12

Paté-Cornell M (1996) Uncertainties in risk analysis: six levels of treatment. Reliab Eng Syst Saf 54(2–3):95–111

Patt A, Dessai S (2005) Communicating uncertainty: lessons learned and suggestions for climate change assessment. Comptes Rendus Geosci 337(4):425–441

Pistrika A (2010) Flood damage estimation based on flood simulation scenarios and a GIS platform. Eur Water 30:3–11

Pistrika A, Tsakiris G, Nalbantis I (2014) Flood depth-damage functions for built environment. Environ Process 1(4):553–572

Rausand M (2011) Risk assessment: theory, methods and application. Wiley, New Jersey

Reason J (1990) Human error. Cambridge University Press, Cambridge

Refsgaard J, Sluijs J, Hojberg A, Vanrolleghem P (2007) Uncertainty in the environmental modeling process—a framework and guidance. Environ Model Softw 22(11):1543–1556

Regan H, Colyvan M, Burgman M (2002) A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecol Appl 12(2):618–628

Saltelli A, Tarantola S, Campolongo F, Ratto M (2004) Sensitivity analysis in practice. A guide to assessing scientific models. Wiley, New York

Seed R, Bea R, Abdelmalak R, Athanasopoulos A, Boutwell G, Bray J, Briaud J-L, Cheung C, Cobos-Roa D, Cohen-Waeber J, Collins B, Ehrensing L, Farber D, Hanemann M, Harder L, Inkabi K, Kammerer A, Karadeniz D, Kayen R, Moss R, Nicks J, Nimmala Pestana J, Porter J, Rhee K, Riemer M, Roberts K, Rogers J, Storesund R, Govindasamy A, Vera-Grunauer X, Wartman J, Watkins C, Wenk E Jr, Yim S (2006) Investigation of the performance of the New Orleans flood protection systems in Hurricane Katrina on August 29, 2005: final report. Independent levee investigation team

Serje J (2012) Data sources on hazards. In: Wisner B, Gaillard J, Kelman I (eds) The Routledge handbook of hazards and disaster risk reduction. Routledge, London

Sigel K, Klauer B, Pahl-Wostl C (2010) Conceptualizing uncertainty in environmental decision-making: the example of the EU water framework directive. Ecol Econ 69(3):502–510

Sills G, Vroman N, Wahl R, Schwanz N (2008) Overview of New Orleans levee failures: lessons learned and their impact on national levee design and assessment. J Geotech Geoenviron Eng 134(5):556–565

Skeels M, Lee B, Smith G, Robertson G (2010) Revealing uncertainty for information visualization. J Inf Vis 9(1):70–81

Smith R (2014) Uncertainty quantification: theory, implementation, and applications. computational science and engineering series. SIAM, Philadelphia

Smith A, Katz R (2013) US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. Nat Hazards 67(2):387–410

Smith A, Matthews J (2015) Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates. Nat Hazards 77(3):1829–1851

Smith E, Shugart H (1994) Uncertainty in ecological risk assessment. Ecological risk assessment issue papers, EPA/630/R-94/009. United States Environmental Protection Agency, Washington DC, USA

Smithson M (1989) Ignorance and uncertainty: emerging paradigms. Springer, New York

Sudret B (2007) Uncertainty propagation and sensitivity analysis in mechanical models—contributions to structural reliability and stochastic spectral methods. Technical report. Université Blaise Pascal, France

Thomson J, Hetzler B, MacEachren A, Gahegan M, Pavel M (2005) A typology for visualizing uncertainty. Proc SPIE 5669:146–157

Tucker W, Ferson S (2003) Probability bounds analysis in environmental risk assessment. Technical report, applied biomathematics, New York, USA

Van Asselt M (2000) Perspectives on uncertainty and risk: the PRIMA approach to decision support. Kluwer Academic Publishers, Dordrecht

van der Sluijs J, Janssen P, Petersen A, Kloprogge P, Risbey J, Tuinstra W, Ravetz J (2004) RIVM/MNP guidance for uncertainty assessment and communication: tool catalogue for uncertainty assessment. Report no NWS-E-2004-37. Utrecht University, The Netherlands

van der Sluijs J, Craye M, Funtowicz S, Kloprogge P, Ravetz J, Risbey J (2005) Combining quantitative and qualitative measures of uncertainty in model based environmental assessment: the NUSAP system. Risk Anal 25(2):481–492

Vinet F, Lumbroso D, Defossez S, Boissier L (2012) A comparative analysis of the loss of life during two recent floods in France: the sea surge caused by the storm Xynthia and the flash flood in Var. Nat Hazards 61(3):1179–1201

Vozinaki A, Karatzas G, Sibetheros I, Varouchakis E (2015) An agricultural flash flood loss estimation methodology: the case study of the Koiliaris basin (Greece), February 2003 flood. Nat Hazards 79(2):899–920

Walker W, Harremoës P, Rotmans J, van der Sluijs J, Van Asselt M, Janssen P, Krayer von Krauss M (2003) Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. Integ Assess 4(1):5–17

Wang J, Qin Q, Zhao J, Ye X, Qin X, Yang X, Sun Y (2015) A knowledge-based method for road damage detection using high-resolution remote sensing image. Geoscience and remote sensing symposium (IGARSS), Milan, Italy

Whittingham R (2004) The blame machine: why human error causes accidents. Elsevier Butterworth-Heinemann, Oxford

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

Wu J, Li N, Hallegatte S, Shi P, Hu A, Liu X (2012) Regional indirect economic impact evaluation of the 2008 Wenchuan earthquake. Environ Earth Sci 65(1):161–172

Zhang Y, Peng C, Li W, Fang X, Zhang T, Zhu Q, Zhao P (2013) Monitoring and estimating drought-induced impacts on forest structure, growth, function, and ecosystem services using remote-sensing data: recent progress and future challenges. Environ Rev 21(2):103–115