MEDEA

Multidimensional seismic risk assessment through explainable artificial intelligence

Deliverable 4.2.: Data Fusion

MULTIDIMENSIONAL SEISMIC RISK ASSESSMENT COMBINING STRUCTURAL DAMAGES AND PSYCHOLOGICAL CONSEQUENCES USING EXPLAINABLE ARTIFICIAL INTELLIGENCE

Co-funded by the European Union

eCAMPUS UNIVERSITÀ
UNIVERSITÀ DI PISA
GASILSKA ZVEZA SLOVENIJE
MEĐIMURSKA ŽUPANIJA
Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Deliverable Title	**Data Fusion**
Deliverable number	**4.2**
Deliverable Lead:	UNIPI
Related Work Package:	WP4: Benchmark structures, EDPs, and representative families
Author(s):	Francesco Pistolesi (UNIPI), Michele Baldassini (UNIPI)
Dissemination Level:	Public
Due Submission Date:	31/05/24
Actual Submission:	03/06/24
Project Number	101101236
Status	Version 1.0 (29/02/2024)
Reviewed (Authors)	Francesco Focacci (eCampus University)
Start Date of Project:	01/12/22
Duration:	24 months
Abstract	This deliverable describes the integration of risks associated with earthquakes, combining structural risks with the psychological consequences for affected populations. We explored advanced data fusion techniques to merge these distinct risk dimensions. Structural risk is quantified by evaluating the number of buildings reaching the D3 state based on peak ground acceleration (PGA). Psychological risk is assessed by considering factors such as the number of people with disabilities, the number of elderly individuals, and other vulnerable groups within the population living in the affected area. The deliverable first introduces the most appropriate data fusion techniques. We investigated several data fusion techniques, including Bayesian Inference, Kalman Filtering, Dempster-Shafer Theory, and Weighted sum with the Analytic Hierarchy Process to derive weights. The deliverable also shows the application of these techniques.
Status changes history	Version 1.0 (29/02/2024) – First release

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Contents

eCAMPUS UNIVERSITÀ | UNIVERSITÀ DI PISA | GASILSKA ZVEZA SLOVENIJE | MEĐIMURSKA ŽUPANIJA | Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

1 Data fusion

Data fusion is a multidisciplinary field combining information from multiple sources to produce more consistent, accurate, and valuable data than any individual data source. This process is integral to various applications across numerous fields, including sensor networks, robotics, defense, environmental monitoring, medical diagnostics, and disaster management. Data fusion aims to enhance the overall understanding and decision-making capabilities by exploiting the strengths and compensating for the weaknesses of individual data sources.

Data fusion is deeply rooted in the principle that integrating diverse datasets can reveal patterns and insights that are not evident when considering each dataset in isolation. This is particularly relevant in complex systems where different types of data collected from various sensors or sources need to be synthesized to form a coherent picture of the situation. For instance, in earthquake risk assessment, data fusion can integrate structural damage, geological, and demographic data to provide a comprehensive risk profile.

Mathematically, data fusion can be approached through various methodologies suited to different data types and application scenarios. Bayesian inference is one such method, grounded in Bayes' theorem, which provides a probabilistic framework for updating the likelihood of a hypothesis as new evidence is obtained. This approach efficiently handles uncertainty and incorporates prior knowledge into the analysis. By updating the prior probabilities with new data, Bayesian inference produces posterior probabilities that better reflect the current state of knowledge.

Another prominent data fusion technique is the Kalman filter, which is widely used in dynamic systems to estimate the state of a process over time. The Kalman filter operates recursively, using a series of measurements observed over time that contain noise and other inaccuracies. It predicts the current state based on the previous state and corrects this prediction using the new measurement. This method is especially useful in applications such as navigation and tracking, where real-time data from multiple sensors must be integrated to accurately estimate an object's position and velocity.

Dempster-Shafer theory, also known as the Theory of Evidence, offers an alternative approach to managing uncertainty and combining evidence from different sources. Unlike Bayesian inference, which relies on prior probabilities, Dempster-Shafer theory assigns degrees of belief to various hypotheses based on the available evidence. These beliefs are combined using Dempster's rule of combination, which accounts for the conflict and agreement between different pieces of evidence. This method is advantageous in scenarios where information is incomplete or ambiguous.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

The weighted sum method is another data fusion technique that deals with combining multiple criteria or factors into a single comprehensive score by assigning weights to each criterion based on its relative importance. This method is straightforward and intuitive, making it widely used in various decision-making processes. Each criterion is normalized to ensure comparability, and the weighted sum is calculated by multiplying each normalized criterion by its assigned weight and summing the results. The simplicity of this approach allows for easy interpretation and implementation, making it suitable for initial assessments.

Integrating multiple data sources through these techniques can significantly enhance the reliability and robustness of the resulting information. In disaster management, for example, data fusion can combine seismic data, structural health monitoring data, and demographic information to comprehensively assess the physical and human impacts of an earthquake. This integrated approach enables more effective resource allocation, emergency response planning, and long-term mitigation strategies.

Despite its advantages, data fusion also presents several challenges. These include issues related to data heterogeneity, where different sources may provide data in varying formats and resolutions; data quality, where the reliability and accuracy of each source must be assessed; and computational complexity, as integrating large volumes of data from multiple sources can be resource-intensive. Addressing these challenges requires sophisticated algorithms and robust computational frameworks capable of handling the intricacies of real-world data.

In conclusion, data fusion is a powerful means that will continue to help information integration. By using various mathematical and computational techniques, data fusion enables the synthesis of comprehensive and accurate information from disparate datasets. This capability is particularly crucial in fields such as disaster management, where timely and accurate information can significantly impact the effectiveness of response and mitigation efforts. As data continues to grow in volume and complexity, the importance of advanced data fusion techniques will only increase, driving further innovation and application in this critical field.

2 Bayesian inference

2.1 Introduction

Bayesian inference is a powerful statistical method used to update the probability estimate for a hypothesis as more evidence or information becomes available.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

This approach is based on Bayes' theorem, which describes the probability of a hypothesis given prior knowledge and new data. The theorem is a fundamental tool in the field of statistics and is widely used in various applications, including data fusion, machine learning, and decision making under uncertainty.

Bayes' theorem provides a mathematical formula for updating probabilities. It is expressed as:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

where: $P(H|E)$ is the posterior probability of the hypothesis H given the evidence E; $P(E|H)$ is the likelihood, which is the probability of the evidence E given that the hypothesis H is true; $P(H)$ is the prior probability of the hypothesis H, representing the initial degree of belief in H before observing the evidence; $P(E)$ is the marginal likelihood, also known as the evidence, which is the total probability of the evidence under all possible hypotheses.

The marginal likelihood $P(E)$ can be computed using the law of total probability:

$$P(E) = \sum_i P(E|H_i) \cdot P(H_i)$$

where H_i represents all possible hypotheses. The prior probability $P(H)$ represents the initial belief about the hypothesis before any new evidence is taken into account. It encapsulates existing knowledge or assumptions about the hypothesis.

The likelihood $P(E|H)$ is the probability of observing the evidence given that the hypothesis is true. It measures how well the hypothesis explains the observed data.

The posterior probability $P(H|E)$ is the updated probability of the hypothesis after considering the new evidence. It combines the prior probability and the likelihood to provide a revised estimate of the hypothesis's probability.

The marginal likelihood $P(E)$ is a normalizing constant that ensures the posterior probabilities sum to one. It is the probability of the evidence under all possible hypotheses and is computed by integrating or summing over the likelihoods of all hypotheses.

Suppose we have a hypothesis H about the occurrence of an event, and we observe new evidence E. We want to update our belief in H using Bayesian inference.

- Let $P(H) = 0.6$ be the prior probability of H.
- Let $P(E|H) = 0.7$ be the likelihood of observing E given that H is true.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

- Let $P(E) = 0.5$ be the marginal likelihood of E.

Using Bayes' theorem, the posterior probability $P(H|E)$ is calculated as follows:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} = \frac{0.7 \cdot 0.6}{0.5} = 0.84$$

Thus, the updated probability of H after observing E is 0.84.

In the continuous case, where the evidence and hypotheses are described by continuous variables, Bayes' theorem is expressed using probability density functions (pdfs):

$$f_H(h|E) = \frac{f_E(E|h) \cdot f_H(h)}{f_E(E)}$$

where $f_H(h|E)$ is the posterior density of the hypothesis given the evidence, $f_E(E|h)$ is the likelihood function, $f_H(h)$ is the prior density, and $f_E(E)$ is the marginal density of the evidence.

In the multivariate case, where multiple hypotheses and evidence variables are considered, Bayes' theorem extends to joint probability distributions. Suppose we have a set of hypotheses $\{H_i\}$ and evidence $\{E_j\}$. The posterior probability for a specific hypothesis H_i given the evidence $\{E_j\}$ is:

$$P(H_i|\{E_j\}) = \frac{P(\{E_j\}|H_i) \cdot P(H_i)}{P(\{E_j\})}$$

where $P(\{E_j\})$ is the joint marginal likelihood of the evidence, computed as:

$$P(\{E_j\}) = \sum_i P(\{E_j\}|H_i) \cdot P(H_i)$$

Bayesian inference provides a robust framework for updating the probability of a hypothesis based on new evidence. By incorporating prior knowledge and the likelihood of the observed data, Bayesian methods offer a principled approach to probabilistic reasoning and decision making under uncertainty. This technique is widely applicable in various domains, including data fusion, where it plays a crucial role in integrating information from diverse sources to improve accuracy and reliability.

2.2 Case Study: Multidimensional Risk Assessment Using Bayesian Inference

In this case study, we analyze the risk to both structures and people in an area that has already been affected by an earthquake. The earthquake has resulted in

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

a specific number of buildings reaching the D3 damage status. We aim to calculate the multidimensional risk using Bayesian Inference by integrating observed data on structural damage and various human risk factors, such as the number of people with disabilities, the number of elderly people, types of disabilities, the number of families with low income, and the number of large families.

2.2.1 Variables and Definitions

Consider the definitions as follows:

- Structural Risk (S): The number of buildings (B) that have achieved the D3 damage status.
- Human Risk Factors:

  D: Number of people with disabilities.

  E: Number of elderly people.

  T: Types of disabilities among the disabled population.

  L: Number of families with low income.

  F: Number of large families (with more than three children).

Given that the earthquake has already occurred, we will use the observed data directly without needing to predict future values. Bayesian Inference will allow us to update our understanding of the risk based on this observed data.

2.2.2 Bayesian Inference

After the earthquake, we can use Bayesian Inference to update our beliefs about the overall risk based on the observed data. We start with our prior knowledge about each risk factor and combine it with the observed data to calculate the posterior probabilities, which reflect the updated risk assessment. In particular:

- Prior Probabilities:

  $P(S)$: Prior probability distribution of structural risk.

  $P(D)$: Prior probability distribution of the number of people with disabilities.

  $P(E)$: Prior probability distribution of the number of elderly people.

  $P(T)$: Prior probability distribution of types of disabilities.

  $P(L)$: Prior probability distribution of the number of low-income families.

  $P(F)$: Prior probability distribution of the number of large families.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

- Observed Data:

  O_S: Observed number of buildings with D3 damage.

  O_D: Observed number of people with disabilities.

  O_E: Observed number of elderly people.

  O_T: Observed types of disabilities.

  O_L: Observed number of low-income families.

  O_F: Observed number of large families.

- Posterior Probabilities:

  $P(S|O_S)$: Posterior probability of structural risk given the observed data.

  $P(D|O_D)$: Posterior probability of the number of people with disabilities given the observed data.

  $P(E|O_E)$: Posterior probability of the number of elderly people given the observed data.

  $P(T|O_T)$: Posterior probability of types of disabilities given the observed data.

  $P(L|O_L)$: Posterior probability of the number of low-income families given the observed data.

  $P(F|O_F)$: Posterior probability of the number of large families given the observed data.

2.2.3 Calculation of Multidimensional Risk

We start by defining the prior distributions for each variable based on historical data and expert knowledge. Let's assume we have the following priors:

- $P(S) \sim$ Normal(μ_S, σ_S^2)
- $P(D) \sim$ Poisson(λ_D)
- $P(E) \sim$ Poisson(λ_E)
- $P(T) \sim$ Multinomial(p_T)
- $P(L) \sim$ Poisson(λ_L)
- $P(F) \sim$ Poisson(λ_F)

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union

Next, we use the observed data from the earthquake event, for example: $O_S = 12$, $O_D = 25$, $O_E = 18$, $O_T = (10, 8, 5, 2)$, $O_L = 12$, and $O_F = 9$.

Using Bayes' theorem, we update the prior probabilities with the observed data to obtain the posterior probabilities. For example, the posterior probability of structural risk is calculated as:

$$P(S|O_S) = \frac{P(O_S|S) \cdot P(S)}{P(O_S)}.$$

Given the normal distribution assumption for structural risk, the likelihood $P(O_S|S)$ is also a normal distribution centered around the observed value $O_S = 12$. We then update the mean and variance based on the observed data and the prior distribution.

Similarly, for the Poisson-distributed human risk factors, we update the rate parameters (λ) using the observed counts, yielding new posterior distributions.

2.2.4 Example Calculation

Assume the following prior distributions and observed data: $P(S) \sim$ Normal$(10, 2^2)$, $P(D) \sim$ Poisson(20), $P(E) \sim$ Poisson(15), $P(T) \sim$ Multinomial$(0.4, 0.3, 0.2, 0.1)$ for four types of disabilities, $P(L) \sim$ Poisson(10), $P(F) \sim$ Poisson(8).

Observed data are as follows: $O_S = 12$, $O_D = 25$, $O_E = 18$, $O_T = (10, 8, 5, 2)$, $O_L = 12$, and $O_F = 9$.

Using Bayesian updating, we calculate the posterior probabilities for each risk factor. For instance, the posterior probability for structural risk might be as follows:

$$P(S|O_S) = \frac{P(O_S|S) \cdot P(S)}{P(O_S)}.$$

Given the normal distribution assumption for structural risk, the likelihood $P(O_S|S)$ is also a normal distribution centered around the observed value $O_S = 12$. We then update the mean and variance based on the observed data and the prior distribution.

Similarly, for the Poisson-distributed human risk factors, we update the rate parameters (λ) using the observed counts, yielding new posterior distributions.

2.2.5 Combined Multidimensional Risk

To assess the overall multidimensional risk, we combine the posterior probabilities of all risk factors. This integration can be represented as a multidimensional

eCAMPUS UNIVERSITÀ
Università di Pisa
GASILSKA ZVEZA SLOVENIJE
MEĐIMURSKA ŽUPANIJA
Co-funded by the European Union
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

distribution that considers both structural and human factors, allowing for a comprehensive risk assessment.

In this case study, Bayesian Inference was used to integrate observed data on structural damage and various human risk factors to assess the multidimensional risk after an earthquake. This approach allows us to update our understanding of the overall risk based on observed data, providing a robust framework for post-event risk assessment and decision-making. By combining information from multiple sources, we can obtain a more accurate and nuanced understanding of the impacts of the earthquake, leading to better resource allocation and mitigation strategies.

3 Kalman filtering

3.1 Introduction

Kalman filtering, introduced by Rudolf E. Kálmán in 1960, is a powerful mathematical technique used for estimating the state of a dynamic system from a series of noisy measurements. The filter is designed to operate in real-time and is extensively utilized in various applications, including navigation, signal processing, control systems, and data fusion. This introduction aims to provide a comprehensive understanding of Kalman filtering, its principles, and its relevance to data fusion.

The Kalman filter operates on a recursive estimation framework. At its core, it combines predictions from a mathematical model of the system with actual measurements to produce estimates that are more accurate than those obtained from either the model or the measurements alone. This process involves two key stages: prediction and update.

- **Prediction Stage** In the prediction stage, the filter uses the current state estimate and the system's dynamic model to predict the state at the next time step. The dynamic model is typically represented by linear equations, though extensions of the Kalman filter, such as the Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF), handle nonlinear systems. The predicted state estimate and its associated uncertainty (error covariance) are calculated in this stage.

- **Update Stage** - In the update stage, the filter incorporates new measurements to correct the predicted state. This involves calculating the Kalman gain, which determines the weighting between the predicted state and

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

the new measurement. The state estimate and its uncertainty are then updated based on the new measurement and the Kalman gain.

The Kalman filter is mathematically described by a set of equations that govern the prediction and update stages. These equations are derived from the principles of Bayesian inference and linear algebra.

Prediction equations are as follows. Let $\mathbf{x}_k$ represent the state vector at time step k, and $\mathbf{P}_k$ denote the error covariance matrix associated with the state estimate. The dynamic model of the system is given by:

$$\mathbf{x}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{x}_{k-1|k-1} + \mathbf{B}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}_{k-1}^{\top} + \mathbf{Q}_{k-1}$$

Here, $\mathbf{F}_{k-1}$ is the state transition matrix, $\mathbf{B}_{k-1}$ is the control input matrix, $\mathbf{u}_{k-1}$ is the control input vector, $\mathbf{w}_{k-1}$ is the process noise with covariance $\mathbf{Q}_{k-1}$, and $\mathbf{x}_{k|k-1}$ represents the predicted state estimate.

When a new measurement $\mathbf{z}_k$ is available, the update equations correct the predicted state estimate. The measurement model is given by:

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k$$

$$\mathbf{y}_k = \mathbf{z}_k - \mathbf{H}_k\mathbf{x}_{k|k-1}$$

$$\mathbf{S}_k = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^{\top} + \mathbf{R}_k$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\top}\mathbf{S}_k^{-1}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k\mathbf{y}_k$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1}$$

Here, $\mathbf{H}_k$ is the measurement matrix, $\mathbf{v}_k$ is the measurement noise with covariance $\mathbf{R}_k$, $\mathbf{y}_k$ is the measurement residual (innovation), $\mathbf{S}_k$ is the innovation covariance, $\mathbf{K}_k$ is the Kalman gain, and $\mathbf{x}_{k|k}$ represents the updated state estimate.

Data fusion involves combining information from multiple sources to achieve a more accurate and reliable understanding of a system. Kalman filtering is particularly well-suited for data fusion due to its ability to recursively process noisy measurements and produce optimal estimates.

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

While the standard Kalman filter assumes linear models and Gaussian noise, many real-world systems exhibit nonlinearity and non-Gaussian noise. To address these challenges, several extensions of the Kalman filter have been developed:

- **Extended Kalman Filter (EKF):** - The EKF linearizes the nonlinear system dynamics and measurement models around the current state estimate. It then applies the standard Kalman filter equations to the linearized models. - This approach works well for mildly nonlinear systems but can introduce errors if the system exhibits significant nonlinearity.
- **Unscented Kalman Filter (UKF):** - The UKF uses a deterministic sampling technique known as the unscented transform to approximate the state distribution. It avoids linearization and provides more accurate estimates for highly nonlinear systems. - By propagating a set of sigma points through the nonlinear functions, the UKF captures the true mean and covariance of the state distribution more effectively than the EKF.
- **Particle Filter:** - Particle filters, also known as sequential Monte Carlo methods, represent the state distribution using a set of particles. Each particle represents a possible state of the system, and the filter updates the particles based on the measurements. - This approach can handle highly nonlinear and non-Gaussian systems but is computationally more intensive than Kalman filters.

Kalman filtering is a a robust and efficient method for estimating the state of dynamic systems. Its recursive nature, mathematical rigor, and ability to handle various types of data make it indispensable in fields. The extensions of the Kalman filter, such as the EKF and UKF, further enhance its applicability to complex, nonlinear systems. By using Kalman filtering, we can achieve accurate and reliable multidimensional risk estimates.

3.2 Case Study: Multidimensional Risk Assessment Using Kalman Filtering

In this case study, we analyze the risk to both structures and people in an area that has already been affected by an earthquake using Kalman filtering. The earthquake has resulted in a specific number of buildings reaching the D3 damage status. We aim to calculate the multidimensional risk by integrating observed data on structural damage and various human risk factors, such as the number of people with disabilities, the number of elderly people, types of

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

disabilities, the number of families with low income, and the number of large families. This approach will utilize the Kalman filter to provide real-time updates and more accurate risk assessments.

3.2.1 Variables and Definitions

Variables are as follows:

- Structural Risk (S): The number of buildings (B) that have achieved the D3 damage status.
- Human Risk Factors:

  D: Number of people with disabilities.

  E: Number of elderly people.

  T: Types of disabilities among the disabled population.

  L: Number of families with low income.

  F: Number of large families (with more than three children).

Given that the earthquake has already occurred, we will use the observed data directly and update our risk assessment in real-time as new data becomes available.

3.2.2 Kalman Filtering

The Kalman filter operates through a cycle of prediction and update stages, which will be adapted to our specific variables.

- **Prediction Stage:**

  State Vector ($\mathbf{x}_k$): Includes all risk factors.

$$\mathbf{x}_k = \begin{bmatrix} S_k \\ D_k \\ E_k \\ T_k \\ L_k \\ F_k \end{bmatrix}$$

  State Transition Model ($\mathbf{F}_k$): Describes how the state evolves over time.

$$\mathbf{F}_k = \mathbf{I} \quad \text{(Identity Matrix)}$$

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

-

Control Input Model ($\mathbf{B}_k$): Describes external influences (assumed zero here).

$$\mathbf{u}_k = 0$$

Process Noise ($\mathbf{w}_k$): Represents uncertainties in the model, with covariance matrix $\mathbf{Q}_k$.

The predicted state and error covariance are:

$$\mathbf{x}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{x}_{k-1|k-1}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}_{k-1}^{\top} + \mathbf{Q}_{k-1}$$

- **Update Stage:**

Measurement Vector ($\mathbf{z}_k$): Represents observed data.

$$\mathbf{z}_k = \begin{bmatrix} O_S \\ O_D \\ O_E \\ O_T \\ O_L \\ O_F \end{bmatrix}$$

Measurement Model ($\mathbf{H}_k$): Maps the true state to the observed data (identity matrix for direct observation).

$$\mathbf{H}_k = \mathbf{I}$$

Measurement Noise ($\mathbf{v}_k$): Represents uncertainties in the measurements, with covariance matrix $\mathbf{R}_k$.

The innovation (residual) and innovation covariance are:

$$\mathbf{y}_k = \mathbf{z}_k - \mathbf{H}_k\mathbf{x}_{k|k-1}$$

$$\mathbf{S}_k = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^{\top} + \mathbf{R}_k$$

The Kalman gain, updated state estimate, and updated error covariance are:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\top}\mathbf{S}_k^{-1}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k\mathbf{y}_k$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1}$$

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

3.2.3 Observed Data and Initial Conditions

Assume the following initial conditions and observed data:
Prior Estimates:

$$\mathbf{x}_0 = \begin{bmatrix} 10 \\ 20 \\ 15 \\ 8 \\ 10 \\ 7 \end{bmatrix}$$

Here, the state vector components could represent:

- $x_{0,1}$: Initial estimate of structural damage.
- $x_{0,2}$: Initial estimate of the number of people with disabilities.
- $x_{0,3}$: Initial estimate of the number of elderly people.
- $x_{0,4}$: Initial estimate of types of disabilities (aggregated score).
- $x_{0,5}$: Initial estimate of the number of families with low income.
- $x_{0,6}$: Initial estimate of the number of large families.

The error covariance matrix represents the initial uncertainty associated with each element of the state vector. This matrix is typically initialized as a diagonal matrix, where each diagonal element represents the variance (uncertainty) of the corresponding state variable.

$$\mathbf{P}_0 = \begin{bmatrix} 2^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3^2 \end{bmatrix}$$

This means that the variance (uncertainty) in the initial estimate of structural damage is $2^2 = 4$; the variance in the initial estimate of the number of people with disabilities is $5^2 = 25$. And so on for the other state variables.

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Observed Data (Post-Earthquake):

$$\mathbf{z}_1 = \begin{bmatrix} 12 \\ 25 \\ 18 \\ (10, 8, 5, 2) \\ 12 \\ 9 \end{bmatrix}$$

3.2.4 Kalman Filter Implementation

Prediction Stage:

$$\mathbf{x}_{1|0} = \mathbf{F}_0\mathbf{x}_0 = \mathbf{x}_0$$
$$\mathbf{P}_{1|0} = \mathbf{F}_0\mathbf{P}_0\mathbf{F}_0^\top + \mathbf{Q}_0$$

Assume $\mathbf{Q}_0$ is a small diagonal matrix representing low process noise.
Update Stage:

$$\mathbf{y}_1 = \mathbf{z}_1 - \mathbf{H}_1\mathbf{x}_{1|0}$$
$$\mathbf{S}_1 = \mathbf{H}_1\mathbf{P}_{1|0}\mathbf{H}_1^\top + \mathbf{R}_1$$

Assume $\mathbf{R}_1$ is a diagonal matrix representing measurement noise covariance.
Calculate the Kalman gain:

$$\mathbf{K}_1 = \mathbf{P}_{1|0}\mathbf{H}_1^\top\mathbf{S}_1^{-1}$$

Update the state estimate:

$$\mathbf{x}_{1|1} = \mathbf{x}_{1|0} + \mathbf{K}_1\mathbf{y}_1$$

Update the error covariance:

$$\mathbf{P}_{1|1} = \mathbf{P}_{1|0} - \mathbf{K}_1\mathbf{H}_1\mathbf{P}_{1|0}$$

3.2.5 Results and Interpretation

Using the observed data and the Kalman filter equations, we obtain the updated state estimates and error covariances. These provide us with the multidimensional risk assessment, combining the structural risk and various human risk factors. For example, after the update step, the state vector may be as follows:

$$\mathbf{x}_{1|1} = \begin{bmatrix} 11.5 \\ 23 \\ 16.5 \\ (9.5, 7.5, 5, 2) \\ 11 \\ 8 \end{bmatrix}$$

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union

This updated state vector reflects the integrated risk assessment, incorporating the new measurements and providing a more accurate and comprehensive view of the situation post-earthquake.

In this case study, Kalman filtering was applied to integrate observed data on structural damage and various human risk factors to assess the multidimensional risk after an earthquake. The filter effectively combined predictions from the dynamic model with actual measurements, providing real-time updates and more accurate risk assessments. This approach demonstrates the utility of Kalman filtering in multidimensional risk assessment, enhancing decision-making and resource allocation in post-disaster scenarios.

4 Dempster-Shafer Theory for Data Fusion

4.1 Introduction

The Dempster-Shafer Theory (DST), also known as the Theory of Evidence, is a mathematical framework for modeling epistemic uncertainty—situations where information is incomplete or imprecise. Developed by Arthur P. Dempster and Glenn Shafer in the 1960s and 1970s, DST extends the classical probability theory to provide a more flexible and nuanced way of handling uncertainty and combining evidence from multiple sources. This introduction describes the fundamental concepts, mathematical formalism, and applications of Dempster-Shafer Theory in the context of data fusion.

The core of Dempster-Shafer is the concept of belief functions, mass functions, and the combination of evidence. Unlike traditional probability theory, which assigns probabilities directly to events, DST operates by assigning belief masses to sets of events, allowing for a representation of both uncertainty and ignorance.

- **Frame of Discernment** Θ: The frame of discernment is the set of all possible outcomes or hypotheses. For example, in a diagnostic system, Θ might represent the set of possible diseases. Each element in Θ is mutually exclusive and collectively exhaustive.

- **Mass Function (Basic Probability Assignment, BPA)**: The mass function $m : 2^{\Theta} \to [0, 1]$ assigns a belief mass to each subset of Θ, including the empty set. The sum of the masses of all subsets of Θ is 1:

$$\sum_{A \subseteq \Theta} m(A) = 1$$

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

A mass $m(A)$ represents the exact amount of belief that is committed to the subset A and does not imply any belief in any of its subsets.

- **Belief Function (Bel)**: The belief function $Bel : 2^{\Theta} \rightarrow [0, 1]$ measures the total belief committed to a subset A:

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

  where $Bel(A)$ represents the degree of belief that the true state of the world is contained within A.

- **Plausibility Function (Pl)**: The plausibility function $Pl : 2^{\Theta} \rightarrow [0, 1]$ is the complement of the belief in the complement of A:

$$Pl(A) = 1 - Bel(\neg A) = \sum_{B \cap A \neq \emptyset} m(B)$$

  where $Pl(A)$ represents the degree to which A is plausible, given the evidence.

One of the most significant aspects of Dempster-Shafer Theory is Dempster's rule of combination. This rule combines two independent mass functions, m_1 and m_2, into a new mass function m. The combined mass function m is defined as:

$$m(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) \cdot m_2(C)$$

where K is the conflict coefficient, representing the degree of conflict between the two sources of evidence:

$$K = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C)$$

The factor $\frac{1}{1-K}$ normalizes the combined mass function to ensure that the total mass is 1.

Dempster-Shafer Theory offers several advantages over traditional probability theory, making it particularly useful for data fusion:

- Handling Uncertainty and Ignorance: DST can represent both uncertainty (through belief and plausibility) and ignorance (through non-specific mass assignments), providing a more comprehensive representation of uncertainty.

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

- Flexibility: DST allows for partial belief in multiple hypotheses without requiring prior probabilities, making it flexible in situations where information is sparse or ambiguous.
- Conflict Resolution: Dempster's rule of combination effectively resolves conflicts between different pieces of evidence, offering a systematic way to integrate multiple sources of information.

Dempster-Shafer Theory is widely applied in various fields requiring data fusion, where it enhances decision-making by combining evidence from diverse sources

The mathematical model is as follows. For a subset $A \subseteq \Theta$, the belief and plausibility functions are defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B).$$

Given two mass functions m_1 and m_2 defined on the same frame of discernment Θ, the combined mass function m using Dempster's rule is:

$$m(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) \cdot m_2(C)$$

where the conflict coefficient K is:

$$K = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C)$$

The normalization factor $\frac{1}{1-K}$ ensures that the sum of the combined mass function is 1.

Dempster-Shafer Theory provides a robust and flexible framework for handling and combining uncertain and imprecise information. Its ability to represent both belief and plausibility, along with the systematic approach to combining evidence, makes it a powerful tool for data fusion. The applications span various fields including risk assessment, highlighting its versatility and effectiveness in dealing with complex, real-world problems.

4.2 Multidimensional Risk Assessment Using Dempster-Shafer Theory

In this case study, we aim to evaluate the multidimensional risk associated with an earthquake that has caused significant damage to buildings and affected the

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union

population. We will use the Dempster-Shafer Theory (DST) to combine different risk factors and assess the overall risk. The risk factors include the structural damage to buildings and various human risk factors such as the number of people with disabilities, the number of elderly people, the types of disabilities, the number of families with low income, and the number of large families.

4.2.1 Scenario and Data

Assume the following observed data after an earthquake in a given area:

- Number of buildings (B) with D3 damage status: 12
- Number of people with disabilities (D): 25
- Number of elderly people (E): 18
- Types of disabilities (T): 10 mobility impairments, 8 visual impairments, 5 hearing impairments, 2 cognitive impairments
- Number of families with low income (L): 12
- Number of large families (F): 9

We will apply DST to combine these different data to obtain a multidimensional risk assessment.

4.2.2 Steps of Dempster-Shafer Theory

DST uses the steps as follows:

1. Define the Frame of Discernment (Θ): The frame of discernment represents all possible states of the world we are interested in. In this case, it includes:

   S: Structural risk (damage to buildings);

   H: Human risk factors (disabilities, elderly, low income, large families).

2. Assign Basic Probability Assignments (BPAs): We need to assign belief masses to subsets of Θ based on the observed data. These assignments represent our confidence in different states of the world.

3. Combine Evidence Using Dempster's Rule: We will combine the BPAs from different sources of evidence to obtain a comprehensive risk assessment.

Based on the observed data, we assign belief masses to different subsets of Θ:

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

- Structural Risk (S): The number of buildings with D3 damage status indicates a significant structural risk. We assign a high belief mass to S:

  $$m_S(S) = 0.8 \quad \text{(High belief in structural risk)}$$

  The remaining belief mass accounts for uncertainty:

  $$m_S(\Theta) = 0.2 \quad \text{(Uncertainty)}$$

- Human Risk Factors (H): The number of people with disabilities, elderly people, and families with low income and large families indicate significant human risk. We assign a high belief mass to H:

  $$m_H(H) = 0.75 \quad \text{(High belief in human risk)}$$

  The remaining belief mass accounts for uncertainty:

  $$m_H(\Theta) = 0.25 \quad \text{(Uncertainty)}$$

4.2.3 Combining Evidence

Using Dempster's rule, we combine the BPAs from the structural risk and human risk factors:

- Combination of BPAs: The combined mass function m is computed using:

  $$m(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_S(B) \cdot m_H(C)$$

  where the conflict coefficient K is:

  $$K = \sum_{B \cap C = \emptyset} m_S(B) \cdot m_H(C)$$

- Calculating Combined Masses: For the intersection $S \cap H = S \cup H = \Theta$, we have:

  $$m(S \cap H) = 0.8 \times 0.75 = 0.6$$

  For the intersection $S \cap \Theta = S$, we have:

  $$m(S \cap \Theta) = 0.8 \times 0.25 = 0.2$$

  For the intersection $\Theta \cap H = H$, we have:

  $$m(\Theta \cap H) = 0.2 \times 0.75 = 0.15$$

  For the intersection $\Theta \cap \Theta = \Theta$, we have:

  $$m(\Theta \cap \Theta) = 0.2 \times 0.25 = 0.05$$

eCAMPUS UNIVERSITÀ | UNIVERSITÀ DI PISA | GASILSKA ZVEZA SLOVENIJE | MEĐIMURSKA ŽUPANIJA | Co-funded by the European Union
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

- Conflict Coefficient K: - There is no conflict since S and H are not disjoint:

$$K = 0$$

- Normalizing Combined Masses: The combined masses are already normalized since $K = 0$. Therefore:

$$m(S) = 0.2, \quad m(H) = 0.15, \quad m(\Theta) = 0.05, \quad m(S \cap H) = 0.6$$

The combined mass function provides the following data: $m(S) = 0.2$, there is a 20% belief in structural risk alone; $m(H) = 0.15$, there is a 15% belief in human risk factors alone; $m(S \cap H) = 0.6$, there is a 60% belief in the combined structural and human risk, indicating that both types of risks are significant; $m(\Theta) = 0.05$, there is a 5% uncertainty.

Using Dempster-Shafer Theory, we have combined structural damage and various human risk factors to obtain a multidimensional risk assessment. This approach allows us to handle uncertainty and provides a comprehensive view of the risk situation after the earthquake. The results indicate a significant combined risk, highlighting the importance of considering both structural and human factors in post-disaster assessments.

5 Weighted Sum for Data Fusion

5.1 Introduction

Data fusion is the process of integrating multiple sources of information to produce more consistent, accurate, and useful data than that provided by any individual data source. Among the various techniques used for data fusion, the weighted sum method is one of the simplest and most intuitive. This method combines different pieces of information by assigning weights to each source and summing the weighted values to produce a final result. This introduction will explore the fundamental concepts, mathematical formulation, advantages, and applications of the weighted sum method in data fusion.

The weighted sum method for data fusion relies on the principle of assigning importance or reliability to each data source through weights. These weights reflect the confidence in each data source and are used to balance the contributions of each piece of information in the final fusion result. In

eCAMPUS UNIVERSITÀ | UNIVERSITÀ DI PISA | GASILSKA ZVEZA SLOVENIJE | MEĐIMURSKA ŽUPANIJA | Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

more detail, data Sources are the various pieces of information or measurements that need to be fused. In practice, these typically come from different sensors, databases, or experts. Weights are numerical values assigned to each data source that reflect the relative importance or reliability of the information provided by that source. The sum of the weights is typically normalized to 1.

The final fused value (so-called weighted sum) is calculated as the sum of the products of each data source and its corresponding weight. This can be expressed as follows:

$$F = \sum_{i=1}^{n} w_i x_i$$

where F is the fused result, x_i is the value from the i-th data source, w_i is the weight assigned to the i-th data source, and n is the number of data sources.

The weighted sum method can be formally described using the following steps:

1. Data Representation: suppose we have n data sources, each providing a measurement x_i (for $i = 1, 2, \dots, n$).
2. Weight Assignment: assign a weight w_i to each data source such that:
$$\sum_{i=1}^{n} w_i = 1$$
3. Weighted Sum Calculation: The aggregate result F is given by:
$$F = \sum_{i=1}^{n} w_i x_i.$$

The weighted sum method offers several advantages, making it a popular choice for data fusion in many applications. First, the weighted sum method is straightforward to understand and implement. Its simplicity makes it accessible for a wide range of applications.

The method also allows for easy adjustment of weights[1] to reflect changes in the reliability or importance of different data sources. This flexibility is valuable in dynamic environments where data quality may vary over time.

[1]Weights can also be set by using multi-criteria decision-making techniques. For example, the next section describes the Analytic Hierarchy Process, one of the most widely used methods to derive weights that reflects the importance of various criteria in real-world problems.

eCAMPUS UNIVERSITÀ
UNIVERSITÀ DI PISA
GASILSKA ZVEZA SLOVENIJE
MEĐIMURSKA ŽUPANIJA
Co-funded by the European Union
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Another advantage is that calculating the weighted sum is computationally efficient, making it suitable for real-time applications where quick data fusion is required.

Finally, the method provides an intuitive way to combine data, as the final result is a weighted average that can be easily understood and interpreted.

The weighted sum method is widely used in various fields requiring data fusion, including decision support systems where it can be used to combine expert opinions, historical data, and real-time information to support decision-making processes. This method is also effective when implementing environmental monitoring systems often rely on data from multiple sensors to assess conditions such as air quality, water quality, and weather. The weighted sum method helps integrate these diverse data sources.

5.2 Multidimensional Risk assessment using the Weighted Sum Method

In this case study, we will evaluate the multidimensional risk associated with an earthquake by applying the weighted sum method. We will combine evidence from structural damage to buildings and various human risk factors such as the number of people with disabilities, the number of elderly people, the types of disabilities, the number of families with low income, and the number of large families.

Assume that after the earthquake, we have the following observed data:

- Number of buildings (B) with D3 damage status: 12
- Number of people with disabilities (D): 25
- Number of elderly people (E): 18
- Types of disabilities (T): 10 mobility impairments, 8 visual impairments, 5 hearing impairments, 2 cognitive impairments
- Number of families with low income (L): 12
- Number of large families (F): 9

We will apply the weighted sum method to combine these risk factors into a single multidimensional risk assessment.

First, we need to assign weights to each risk factor based on their relative importance or impact on the overall risk. For this example, we will assume

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union

the following weights directly expressed by experts[2]:

- Structural damage (W_B): 0.4
- Number of people with disabilities (W_D): 0.2
- Number of elderly people (W_E): 0.15
- Types of disabilities (W_T): 0.1
- Number of families with low income (W_L): 0.1
- Number of large families (W_F): 0.05

The weights are chosen such that they sum up to 1:

$$W_B + W_D + W_E + W_T + W_L + W_F = 1$$

Before calculating the weighted sum, we need to normalize the data to ensure that each risk factor is on a comparable scale. We will use a simple normalization method by dividing each value by the maximum value observed for that factor. Suppose the maximum values for normalization are as follows:

- Maximum number of buildings with D3 damage status: 20
- Maximum number of people with disabilities: 50
- Maximum number of elderly people: 30
- Maximum number of types of disabilities: 10 (maximum of total types considered, here the sum is 25)
- Maximum number of families with low income: 20
- Maximum number of large families: 15

Normalized values (x_i) are calculated as follows: $x_B = \frac{12}{20} = 0.6$; $x_D = \frac{25}{50} = 0.5$; $x_E = \frac{18}{30} = 0.6$; $x_T = \frac{25}{25} = 1.0$ (Total types considered are normalized to 1); $x_L = \frac{12}{20} = 0.6$; $x_F = \frac{9}{15} = 0.6$.

Using the weights and normalized values, we calculate the weighted sum to obtain the overall risk assessment:

$$F = W_B x_B + W_D x_D + W_E x_E + W_T x_T + W_L x_L + W_F x_F$$

[2]Associating weights with criteria is a delicate task as it is difficult to generate a number to express the level of danger that a specific criterion has compared to the others. The next chapter explains how the Analytic Hierarchy Process can help with this task.

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Substituting the values:

$$F = (0.4\times0.6)+(0.2\times0.5)+(0.15\times0.6)+(0.1\times1.0)+(0.1\times0.6)+(0.05\times0.6)$$

$$F = 0.24 + 0.1 + 0.09 + 0.1 + 0.06 + 0.03$$

$$F = 0.62$$

The weighted sum method gives us a final risk score of 0.62. This score represents the combined risk from structural damage and various human factors. The closer the score is to 1, the higher the overall risk. In this context, a score of 0.62 indicates a significant level of combined risk due to the earthquake.

The weighted sum method provides a straightforward and efficient way to combine different risk factors into a single multidimensional risk assessment. By assigning appropriate weights to each factor and normalizing the data, we can derive an overall risk score that reflects the relative importance of each factor. In this case study, the weighted sum method effectively combines the structural damage and human risk factors, yielding a comprehensive assessment of the earthquake's impact. This method's simplicity and computational efficiency make it a valuable tool for real-time risk assessment and decision-making in disaster management.

6 The Analytic Hierarchy Process (AHP) to Infer Weights

The Analytic Hierarchy Process (AHP) is a structured technique for organizing and analyzing complex decisions, based on mathematics and psychology. It was developed by Thomas L. Saaty in the 1970s and is used worldwide in decision-making processes where multiple criteria are involved. AHP helps decision-makers derive ratio scales from paired comparisons. These scales are derived from the principal eigenvectors and eigenvalues of comparison matrices, making it a rigorous and reliable method to determine the weights of various criteria.

The AHP involves the following main steps:

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

1. **Define the Problem and Structure the Hierarchy:** Identify the goal of the decision-making process. Break down the problem into a hierarchy of more easily comprehended sub-problems, each of which can be analyzed independently. The hierarchy typically has three levels: the goal, the criteria, and the alternatives.
2. **Perform Pairwise Comparisons and Build the Comparison Matrix:** Compare the criteria (and sub-criteria, if any) in pairs, judging which element is more important and to what extent. Use a scale[3] of 1 to 9, where 1 represents equal importance and 9 represents extreme importance of one criterion over another.
3. **Calculate the Weights (Eigenvalues):** Construct a comparison matrix where each element a_{ij} represents the relative importance of criterion i over criterion j. Normalize the comparison matrix and calculate the priority vector (eigenvector) by finding the principal eigenvalue. This vector provides the relative weights of the criteria.
4. **Check Consistency:** Calculate the Consistency Index (CI) and Consistency Ratio (CR) to ensure that the judgments are consistent. A CR of 0.1 or less is generally considered acceptable.

6.1 Deriving Weights for the Weighted Sum Method

Let us apply AHP to derive the weights for our case study involving the risk factors after an earthquake.

1. Define the Problem and Structure the Hierarchy:

   Goal: Determine the overall risk after an earthquake.

   Criteria: Structural damage (B), number of people with disabilities (D), number of elderly people (E), types of disabilities (T), number of families with low income (L), and number of large families (F).
2. Pairwise Comparisons and Comparison Matrix: After performing the pairwise comparisons for the six criteria using a scale of 1 to 9, sup-

[3] Saaty's scale is the most widely used.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

pose the Pairwise Comparison Matrix **A** is as follows:

	B	D	E	T	L	F
B	1	3	5	7	5	9
D	1/3	1	3	5	3	7
E	1/5	1/3	1	3	3	5
T	1/7	1/5	1/3	1	3	5
L	1/5	1/3	1/3	1/3	1	3
F	1/9	1/7	1/5	1/5	1/3	1

3. Calculate the Weights:

   Normalize the comparison matrix by dividing each element by the sum of its column. Then, average the normalized values in each row to get the priority vector.

   The Normalized matrix is as follows:

	B	D	E	T	L	F
B	0.445	0.536	0.526	0.438	0.357	0.438
D	0.148	0.179	0.316	0.313	0.214	0.341
E	0.089	0.060	0.105	0.188	0.214	0.244
T	0.064	0.036	0.035	0.063	0.214	0.244
L	0.089	0.060	0.035	0.021	0.071	0.146
F	0.049	0.026	0.021	0.012	0.071	0.049

   Priority vector (average of rows):

$$W_B = \frac{0.445 + 0.536 + 0.526 + 0.438 + 0.357 + 0.438}{6} = 0.457$$

$$W_D = \frac{0.148 + 0.179 + 0.316 + 0.313 + 0.214 + 0.341}{6} = 0.252$$

$$W_E = \frac{0.089 + 0.060 + 0.105 + 0.188 + 0.214 + 0.244}{6} = 0.150$$

$$W_T = \frac{0.064 + 0.036 + 0.035 + 0.063 + 0.214 + 0.244}{6} = 0.109$$

$$W_L = \frac{0.089 + 0.060 + 0.035 + 0.021 + 0.071 + 0.146}{6} = 0.070$$

$$W_F = \frac{0.049 + 0.026 + 0.021 + 0.012 + 0.071 + 0.049}{6} = 0.043$$

4. Check Consistency:

   Calculate the consistency index (CI):

$$CI = \frac{\lambda_{\text{max}} - n}{n - 1}$$

eCAMPUS UNIVERSITÀ
UNIVERSITÀ DI PISA
GASILSKA ZVEZA SLOVENIJE
MEĐIMURSKA ŽUPANIJA
Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

where $\lambda_{\max}$ is the principal eigenvalue, and n is the number of criteria. To calculate the principal eigenvalue in the context of the Analytic Hierarchy Process (AHP) example, we first compute the sum of each column:

Column	Sum
1	$1+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{5}+\frac{1}{9}\approx 2.515$
2	$3+1+\frac{1}{3}+\frac{1}{5}+\frac{1}{3}+\frac{1}{7}\approx 4.810$
3	$5+3+1+\frac{1}{3}+\frac{1}{3}+\frac{1}{5}\approx 9.967$
4	$7+5+3+1+\frac{1}{3}+\frac{1}{5}\approx 16.633$
5	$5+3+3+3+1+\frac{1}{3}\approx 15.333$
6	$9+7+5+5+3+1\approx 30.000$

Then, we normalize the Pairwise Comparison Matrix:

$$\mathbf{A}_{\text{norm}}=\begin{bmatrix}\frac{1}{2.515} & \frac{3}{4.810} & \frac{5}{9.967} & \frac{7}{16.633} & \frac{5}{15.333} & \frac{9}{30.000}\\ \frac{1/3}{2.515} & \frac{1}{4.810} & \frac{3}{9.967} & \frac{5}{16.633} & \frac{3}{15.333} & \frac{7}{30.000}\\ \frac{1/5}{2.515} & \frac{1/3}{4.810} & \frac{1}{9.967} & \frac{3}{16.633} & \frac{3}{15.333} & \frac{5}{30.000}\\ \frac{1/7}{2.515} & \frac{1/5}{4.810} & \frac{1/3}{9.967} & \frac{1}{16.633} & \frac{3}{15.333} & \frac{5}{30.000}\\ \frac{1/5}{2.515} & \frac{1/3}{4.810} & \frac{1/3}{9.967} & \frac{1/3}{16.633} & \frac{1}{15.333} & \frac{3}{30.000}\\ \frac{1/9}{2.515} & \frac{1/7}{4.810} & \frac{1/5}{9.967} & \frac{1/5}{16.633} & \frac{1/3}{15.333} & \frac{1}{30.000}\end{bmatrix}$$

We calculate the Priority Vector (Principal Eigenvector)

$$\mathbf{w}=\begin{bmatrix}0.457\\0.252\\0.150\\0.109\\0.070\\0.043\end{bmatrix}$$

and the Principal Eigenvalue ($\lambda_{\max}$)

$$\mathbf{A}\mathbf{w}=\begin{bmatrix}1 & 3 & 5 & 7 & 5 & 9\\ \frac{1}{3} & 1 & 3 & 5 & 3 & 7\\ \frac{1}{5} & \frac{1}{3} & 1 & 3 & 3 & 5\\ \frac{1}{7} & \frac{1}{5} & \frac{1}{3} & 1 & 3 & 5\\ \frac{1}{5} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 1 & 3\\ \frac{1}{9} & \frac{1}{7} & \frac{1}{5} & \frac{1}{5} & \frac{1}{3} & 1\end{bmatrix}\begin{bmatrix}0.457\\0.252\\0.150\\0.109\\0.070\\0.043\end{bmatrix}\approx\begin{bmatrix}2.742\\1.505\\0.891\\0.651\\0.412\\0.251\end{bmatrix}$$

Next, we divide each element of **Aw** by the corresponding element of **w**:

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

$$\begin{bmatrix} \frac{2.742}{0.457} \\ \frac{1.505}{0.252} \\ \frac{0.891}{0.150} \\ \frac{0.651}{0.109} \\ \frac{0.412}{0.070} \\ \frac{0.251}{0.043} \end{bmatrix} \approx \begin{bmatrix} 6.00 \\ 5.97 \\ 5.94 \\ 5.97 \\ 5.89 \\ 5.84 \end{bmatrix}$$

Finally, we calculate the average of these values to obtain λ_{max}:

$$\lambda_{\text{max}} = \frac{6.00 + 5.97 + 5.94 + 5.97 + 5.89 + 5.84}{6} \approx 5.935$$

Thus, the principal eigenvalue (λ_{max}) for this pairwise comparison matrix is approximately 5.935.

We then have to calculate the consistency ratio (CR):

$$CR = \frac{CI}{RI}$$

where RI is the random index, which depends on n.
Given $n = 6$, RI = 1.24. If $CI \leq 0.1$, the consistency is acceptable.

6.2 Application to the Case Study

With the derived weights:

- Structural damage (W_B): 0.457
- Number of people with disabilities (W_D): 0.252
- Number of elderly people (W_E): 0.150
- Types of disabilities (W_T): 0.109
- Number of families with low income (W_L): 0.070
- Number of large families (W_F): 0.043

we can apply the weighted sum calculation for the case study as follows:
Normalized values:

$$x_B = 0.6,$$
$$x_D = 0.5,$$
$$x_E = 0.6,$$
$$x_T = 1.0,$$
$$x_L = 0.6,$$
$$x_F = 0.6$$

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Calculate the weighted sum:

$$F = (0.457\times0.6)+(0.252\times0.5)+(0.150\times0.6)+(0.109\times1.0)+(0.070\times0.6)+(0.043\times0.6)$$

$$F = 0.2742 + 0.126 + 0.09 + 0.109 + 0.042 + 0.0258$$

$$F = 0.667$$

The overall risk score of 0.667 indicates a significant combined risk from both structural damage and human factors. This score, derived using the AHP to determine appropriate weights, provides a more structured and justified assessment of the risk, ensuring that each factor's contribution is accurately reflected in the final result.

The Analytic Hierarchy Process (AHP) offers a robust method for determining the weights of criteria in the weighted sum method for data fusion. By structuring the problem, performing pairwise comparisons, calculating weights, and checking consistency, AHP ensures that the weights are derived systematically and reflect the relative importance of each criterion. This method enhances the reliability and accuracy of the final risk assessment, making it a valuable tool in multidimensional risk analysis, especially in scenarios involving complex decision-making processes like post-earthquake evaluations.

7 Comparison and concluding remarks

The comparison of various data fusion methods reveals a range of strengths and weaknesses, especially in the context of our case study on assessing multidimensional risk post-earthquake. Bayesian inference, Kalman filtering, Dempster-Shafer theory, and the weighted sum method each offer distinct approaches to integrating diverse risk factors, such as structural damage and human vulnerability, into a comprehensive risk assessment.

Bayesian inference excels in incorporating prior knowledge and updating risk estimates with new data, providing a probabilistic framework that effectively manages uncertainty. This method is particularly valuable when historical data and expert knowledge are available, enabling a dynamic risk assessment that evolves with additional information. However, Bayesian

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

inference can be computationally intensive and requires accurate prior distributions, which may be challenging to determine in practice.

Kalman filtering offers a robust solution for real-time data fusion, particularly suitable for environments where continuous monitoring and updates are crucial. It is highly efficient in processing noisy data and providing real-time estimates of risk. Nevertheless, Kalman filtering assumes linearity and Gaussian noise, which may not always hold true in complex, multidimensional risk scenarios. The method's effectiveness diminishes when these assumptions are violated, potentially leading to less accurate risk assessments.

Dempster-Shafer theory provides a powerful framework for handling uncertainty and combining evidence from multiple sources, offering a flexible approach that can accommodate incomplete and imprecise information. This theory's ability to manage conflicting evidence makes it particularly useful in situations where data sources may be unreliable or contradictory. However, the computational complexity of Dempster-Shafer theory can be a drawback, especially in large-scale applications where numerous variables and pieces of evidence need to be combined.

The weighted sum method, complemented by the Analytic Hierarchy Process (AHP) for determining weights, is straightforward and intuitive. It allows for the explicit inclusion of expert judgment in assigning relative importance to different risk factors. This method is computationally efficient and easy to implement, making it suitable for initial assessments. However, the simplicity of the weighted sum approach may overlook the interactions between variables and fail to capture the nuances of complex risk scenarios. The determination of weights through AHP relies heavily on subjective judgments, which can introduce bias if not carefully managed.

In conclusion, each data fusion method has its advantages and limitations in the context of multidimensional risk assessment post-earthquake. Bayesian inference is ideal for scenarios with rich historical data and the need for probabilistic updates, while Kalman filtering shines in real-time applications with linearity and Gaussian noise assumptions. Dempster-Shafer theory is unparalleled in managing uncertainty and conflicting evidence, albeit at a computational cost. The weighted sum method, with AHP-derived weights, offers simplicity and ease of implementation but may oversimplify complex interactions.

The most appropriate method for our multidimensional risk assessment framework will be selected based on the findings of a pilot study using actual data from existing areas. This empirical evaluation will help optimize

eCAMPUS UNIVERSITÀ · UNIVERSITÀ DI PISA · GASILSKA ZVEZA SLOVENIJE · MEĐIMURSKA ŽUPANIJA · Co-funded by the European Union

the framework, ensuring that the chosen method aligns with the practical realities and specific needs of the case study. By rigorously testing these methods in a real-world context, we aim to develop a robust and reliable risk assessment tool that effectively integrates the diverse dimensions of risk associated with earthquakes.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.