PROBABILISTIC ASSESSMENT OF EXTREME EVENTS SUBJECT TO EPISTEMIC UNCERTAINTIES

Presentation on the paper of Nishijima et al

Andreas Kurz

ETH Zurich

kurz@ibk.baug.ethz.ch

22.10.2008

Contents

- Introduction
- Eleatory and epistemic uncertainties
- Probabilistic modeling of events subject to uncertainties
- Examples
  - n-year maximum
  - Return period
  - Hazard curve
- Discussion

Introduction

Probabilistic models of events

- Crucial corner stone in risk based decision making
- Base for assessment of probabilities of events of interest
  - e.g. extreme wave heights, current and wind velocities
- Established through joint consideration of
  - knowledge, experience and observations
  - combining statistical assessments with subjective judgments
- Associated with uncertainties
  - aleatory uncertainties, and epistemic uncertainties

Introduction

Literature review

- How uncertainties from different sources may be categorized?
- How these categories be considered in probabilistic risk assessment and risk based decision making?

→relevance of epistemic uncertainties in risk assessment well recognized

→general principles for modeling and assessing probabilistic characteristics well understood

→situations where general principles are violated in practice

Eleatory and epistemic uncertainties

Aleatory uncertainty

- variability of events subject to inherent natural variability

Epistemic uncertainty

- imprecise models
- lack of data
- insufficient knowledge

→Why categorization of the two components of uncertainty?

Eleatory and epistemic uncertainties

Categorization of the two components

- optimization of resource allocations aiming to reduce uncertainty and thereby to enhance ranking of options for the purpose of risk management;
  - epistemic uncertainty reduced by accumulating data and knowledge
- Epistemic uncertainty often may have a profound effect on the probabilistic characteristics of systems

→epistemic uncertainty utilized for reduction of the uncertainty of a whole system performance by inspecting the states of some of the components in the system

Probabilistic modeling of events subject to uncertainties

Probabilistic modeling approach in practice

- statistical modeling relying only on available relevant data
  - preferred by classical statisticians (results are coherent with frequentistic interpretation of probabilities)
  - one to one correspondence between observations and model predictions
  - clear distinction between epistemic uncertainty and aleatory uncertainty
  - Typically formulated as annual extreme value distribution

→This approach has drawbacks!

Probabilistic modeling of events subject to uncertainties

Drawbacks statistical modeling

- direct observations of extreme events are rare (by definition)
    - parameter estimation of the distributions with large statistical uncertainties (epistemic uncertainty)
- available scientific knowledge and/or engineering experiences cannot be included in the modeling

→Do overcome these drawbacks

- by means of engineering probabilistic models
    - for different types of hazards, which enables to integrate into the hazard analysis the available knowledge and engineering experience

Probabilistic modeling of events subject to uncertainties

General principles

- the probabilistic modeling problem can be represented as a problem involving the expectation operation over a function $g(\mathbf{X})$ of aleatory random variable $\mathbf{X}=(X_1,X_2,\ldots,X_{n})$ as: $E[g(\mathbf{X})]=E_{\mathbf{\Theta}}\left[E_{\mathbf{X}}[g(\mathbf{X})\,|\,\mathbf{\Theta}]\right]$

- Where $\mathbf{X}$ are random variables characterized by
- Joint probability function $F_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta})$ conditional on epistemic random variables $\mathbf{\Theta}=(\Theta_1,\Theta_2,\ldots,\Theta_n)$ with probability distribution function $F_{\mathbf{\Theta}}(\boldsymbol{\theta})$
- $F_{\mathbf{X}}(\mathbf{x}\,|\,\boldsymbol{\theta})$ together with $F_{\mathbf{\Theta}}(\boldsymbol{\theta})$ constitutes probabilistic assessment model

Probabilistic modeling of events subject to uncertainties

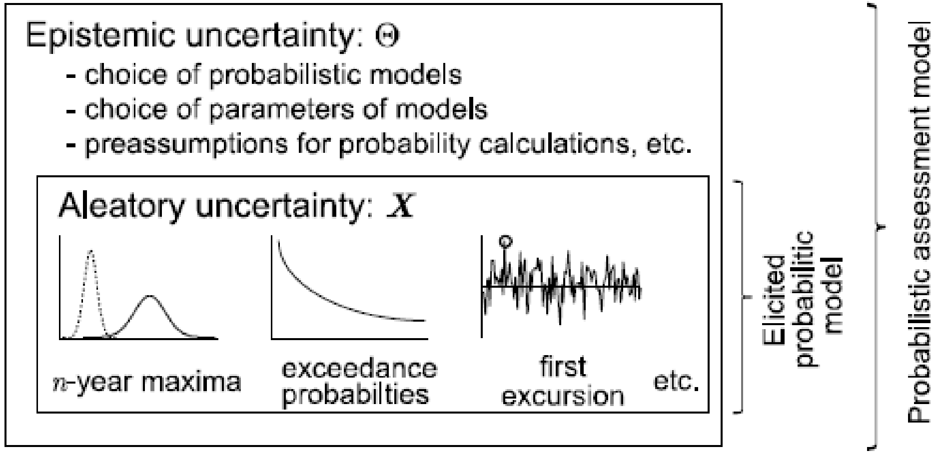

Figure 1. Probabilistic assessment subject to aleatory and epistemic uncertainties.

Epistemic random variables

- uncertainties of the parameters of distributions
- Likelihood (degree of belief) associated with different distribution families and pre-assumptions
- pre-assumptions reflect the modeler’s perception of the phenomena of interest

Examples: n-year maximum

Derivation of the cumulative distribution of the n -year maxima from the annual maximum distribution

$$F_{X,n}(x) = \int \left\{ F(x \mid \theta) \right\}^n p(\theta) d\theta$$

- Where $F(x \mid \theta)$ is conditional cdf of the annual maxima
- $p(\theta)$ is pdf of the epistemic random variable Θ

Sources of the epistemic uncertainty

- statistical uncertainties when cdf established by a pure statistical approach
- model and statistical uncertainties when cdf established based on engineering probabilistic models

→In practice deviations from general principle are observed!

Examples: n-year maximum

Hazard maps

- distribution function of annual maxima $\tilde{F}(x)$ based on characteristic values (e.g. quantile values)
- Calculation n *-year* maximum distribution $F^*_{X,n} = \left\{\tilde{F}(x)\right\}^n$
- Since $\tilde{F}(x)$already contains the effect of epistemic uncertainty

$$\tilde{F}(x) = \int F(x \mid \theta)\, p(\theta)\, d\theta$$

→$F_{X,n}(x)$ and $F^*_{X,n}(x)$ not identical

Jensen‘s inequality for $n > 1$

$$F_{X,n}(x) = E_\Theta\left[\left\{F(x \mid \Theta)\right\}^n\right] \geq \left[E_\Theta\left\{F(x \mid \Theta)\right\}\right]^n = \left\{\tilde{F}(x)\right\}^n = F^*_{X,n}(x)$$

→ n-year maximum events are overestimated!

Examples: n-year maximum

Case of wind hazard analysis

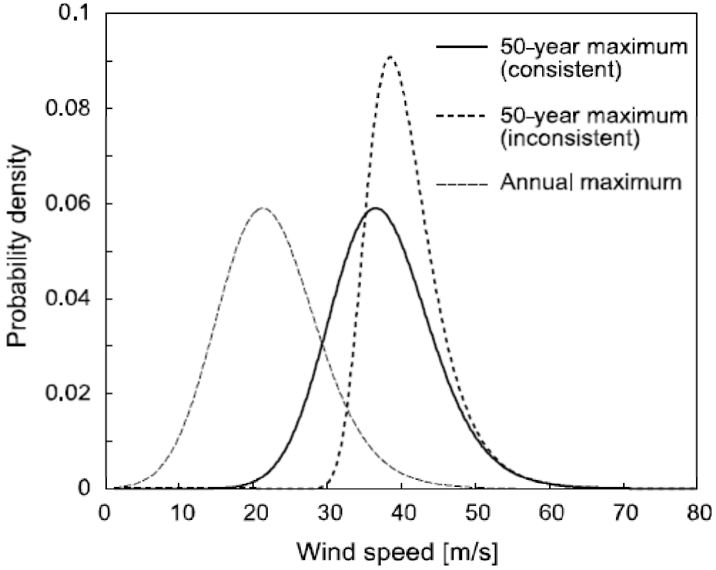

Figure 2. Probability density functions of maximum wind speed.

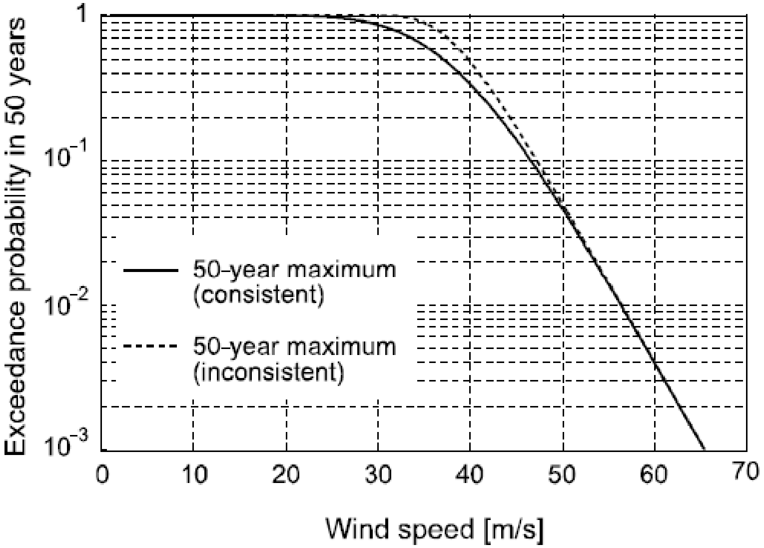

Figure 3. Exceedance probabilities of 50-year maximum wind speed.

→ pdf looks significantly different and mean value of the 50-year maximum wind speed is overestimated

→ Exceedance probability overestimated at the range between 10^{-1} and 1

Examples: Return Period

- Return period may be defined as the expected value of the arrival time of the event of interest (Benjamin, Cornell)
- Event characterized by intensity X (e.g. wind speed):

$$E\left[T(x)\right]^{*} = \frac{1}{1-F(x)}$$

- epistemic uncertainty represented through Θ involved

$$E\left[T(x)\right] = E_{\Theta}\left[\frac{1}{1-F(x\,|\,\Theta)}\right] = \int \frac{p(\theta)}{1-F(x)} d\theta$$

  - Where $F(x\,|\,\theta)$ is conditional cdf on epistemic uncertanty and
  - $p(\theta)$ is pdf of epistemic random variable Θ
- Formulation is coherent with the general principle

Examples: Return Period

Case probabilistic engineering models

- cdf of maximum intensity within a given reference period established by combination of probabilistic models
- Represent natural random nature (aleatory uncertainty) subject to model/statistical uncertainties (epistemic uncertainty)
- cdf obtained already considers the epistemic uncertainty

$$\tilde{F}(x) = \int F(x \mid \theta)\, p(\theta)\, d\theta$$

- Return period $E\left[T(x)\right]^{**} = \frac{1}{1-\tilde{F}(x)}$ not same as $E\left[T(x)\right] = E_{\Theta}\left[\frac{1}{1-F(x \mid \Theta)}\right]$

$$= \int \frac{p(\theta)}{1-F(x)}\, d\theta$$

Examples: Return Period

Jensen‘s inequality $E[T(x)] = E_\Theta\left[\frac{1}{1-F(x|\Theta)}\right] \geq \frac{1}{1-E_\Theta[F(x|\Theta)]} = \frac{1}{1-\tilde{F}(x)} = E[T(x)]^{**}$

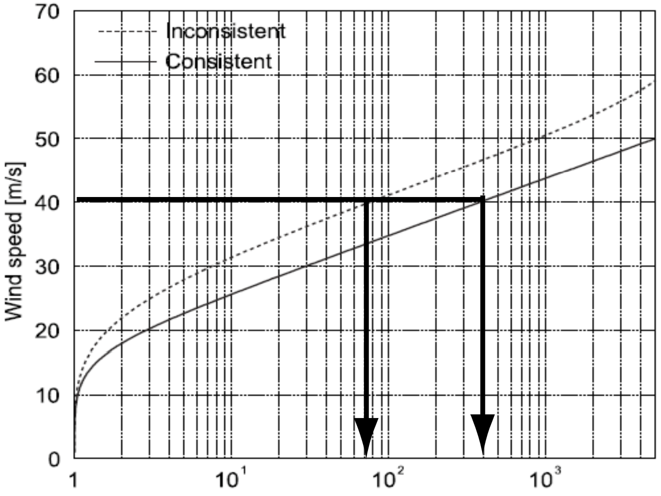

Figure 4. Comparison of return periods.

Discussion

- does not correspond to the expected value of the arrival time

→ the return period assessed under-estimates the expected arrival time!

- return period assessed by Equation (11) should not be used for these purposes.

Example: Hazard curve

- Represents the relationships between the exceedance probabilities for a given uncertain phenomenon represented by the random variable X
- In assessment of probabilities assumptions are made involving epistemic uncertainty
  - imperfection of the postulated models and scarce data available for estimating parameters in the models

→ epistemic uncertainties are often inconsistently considered!

Discussion

Hazard maps

- information provided in typical hazard maps not sufficient to use directly
- do not differentiate the sources of uncertainties
- distributions of maximum values for a given reference period cannot be correctly established

Return period

- Does not correspond to the expected value of the arrival time
- return period assessed by Equation (11) should not be used for these purposes.

Seismic hazard analyses

- justifies the seismic hazard analyses presently made in practice

Discussion

Bayesian probabilistic network

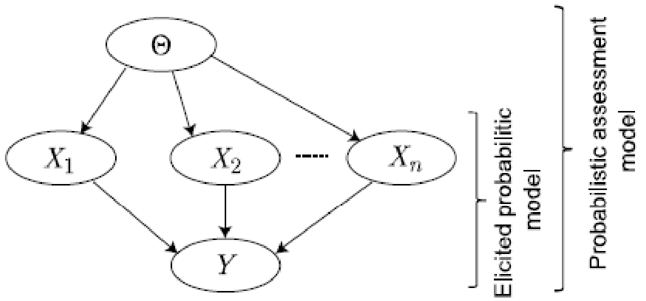

Figure 5. Graphical representation for interrelation between random variables.

- For understanding interrelations between all random variables in the probabilistic assessment
- Θ represents epistemic uncertainty
- X_i represents the annual maximum wind speed at the i^{th} year
- Y represents 50-year maximum wind speed