Skewed Distribution Analysis in Simulation-Based Operation Planning

Dr. Johan Schubert, Dr. Ronnie Johansson, Dr. Pontus Hörling

Swedish Defence Research Agency

Decision Support

- Commanders in operation planning:
  - **Which factors are important for success?**
  - **Within which parameter ranges is success achieved?**

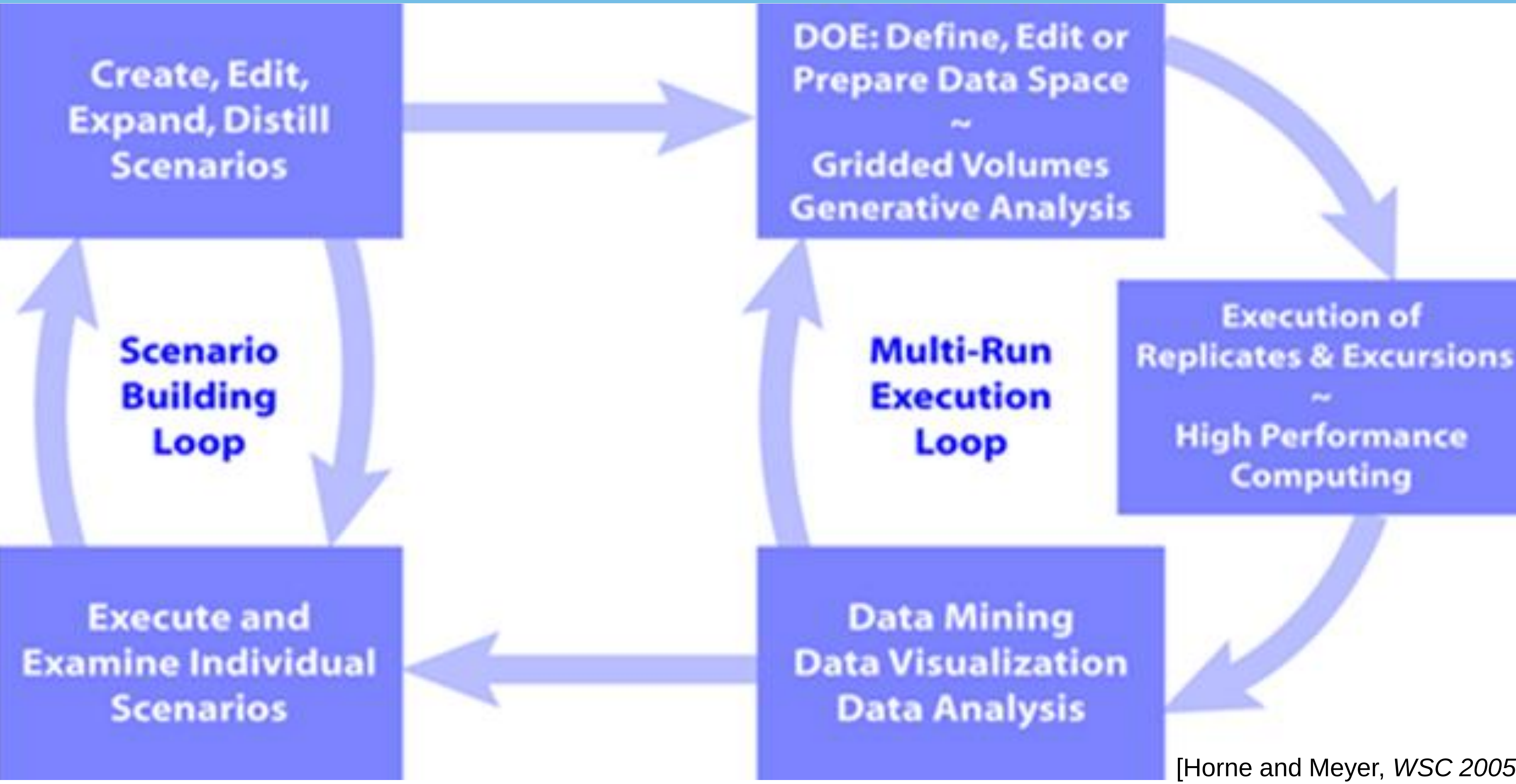
1. Simulate **10 000** instances **with uniform distributions** over all input parameters
2. Select the **1000** **best simulations** where blue forces achieve success

Decision Support

- Observe which factors have **skewed distributions** within the set of **1000**:

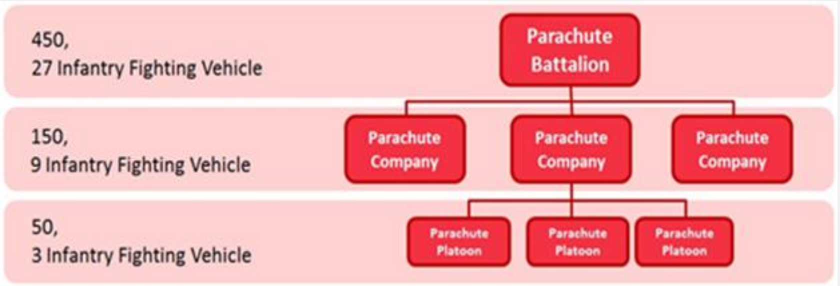
1. These are the **important factors**
2. The **high frequency ranges** corresponding to blue force success

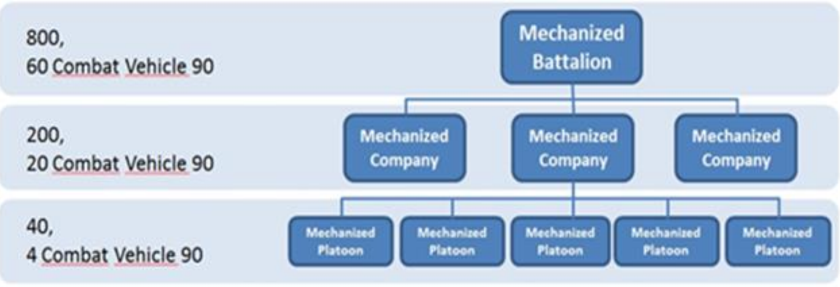
Data Farming: Data Analysis & Viz



[Horne and Meyer, *WSC 2005*]

Scenario





FOI

Decision Support Process

- Evaluate **all plan instance** by all **MOEs**
- Many **MOEs** = Multi Criteria Decision Making
- Difficult for humans to **assign weights** to **MOEs**
- **Use Preference Analysis**

FOI

Decision Support Sub-Processes

- **Analyst View** [A]: statistical analysis, prepare specific questions
- **Commander's Overview** [B]: simulations leading to blue force success are analyzed
- **Commander's Specific Questions** [C]: where subsets of simulations are analyzed

[**A**: Moradi and Schubert, *NMSG 2014*], [**B**: This paper], [**C**: Schubert and Hörling, *Submitted*]

FOI

Commander's Overview

Commander's Overview

- **How many parameters** do we need to achieve blue force success?
- **What are these parameters**?
- **What are the value ranges** for blue force success for these parameters?

Background of Skewed Distribution Analysis

- **Traditional Data Farming**
  - **establish** a target function
  - **analyze** this function using regression trees
- **MSG-088 ”*Data Farming in Support of NATO*”**
  - Traditional analysis **failed**
- **New approach**
  - **Study best subset** of simulations
  - **Observe skewedness** measured by Shannon Entropy

FOI

Conceptual idea

1. **Consider all simulations**: all input parametes uniformly sampled
2. Sort data set after “best” outcome (according to MOEs) and **retain the best 10%**
3. **Study distribution** of remaining values for each input parameter
   - If close to uniform distribution: No or weak effect
   - Skewedness: Strong effect of this parameter on outcome for blue side

Entropy approach

- We use the ***entropy*** as a measure of skewedness of input parameters
  - Study the entropy for input parameters with discrete distributions
  - Continuous distributions are discretized as histograms before analysis

Entropy approach

- **Entropy** H of discrete distribution P of parameter X_i

$$H_i = -\sum_{a=1}^{N_i} P(x_{ia}) \log_2 P(x_{ia})$$

- Uniform distribution gives the **highest entropy**

$$H_{i,max} = \log_2 N_i$$

- **Lowest entropy**

$$H_{i,min} = 0$$

FOI

Normalization

- We adjust the each parameter entropy to be comparable

$$\bar{H}_i = \frac{H_i}{H_{i,max}}$$

Multiple parameter distribution

- Normalized ***joint* entropies** for joint distribution

$$H_{ij} = -\sum_{a=1}^{N_i}\sum_{b=1}^{N_j} P(x_{ia}, x_{jb}) \log_2 P(x_{ia}, x_{jb})$$

- Max entropy

$$H_{ij,max} = \log_2(N_i \cdot N_j)$$

- Normalizing the entropy

$$\bar{H}_{ij} = \frac{H_{ij}}{H_{ij,max}}$$

FOI

Decision support

- **Decide** which **parameters** and associated **values** are most decisive for a successful result
- Simulations and entropy are used in **a two-step procedure**

Decision support procedure

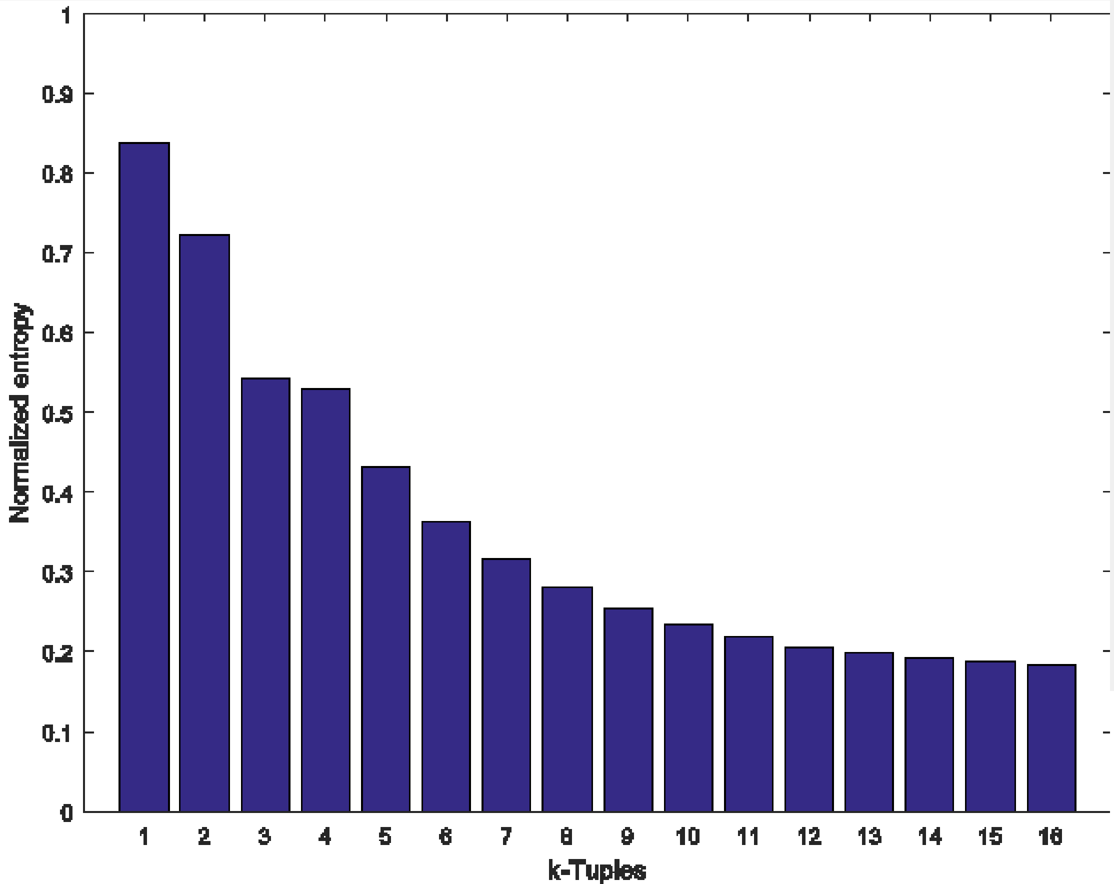
1. Select a suitable **number** of factors
2. Determine the preferred **values** on the selected factors

Decision support procedure: **Step 1**

- **Which factors** has the greatest influence on the result?
- **Step 1:** How many parameters are required to explain blue success?

Decision support procedure: **Step 1**

- Joint normalized entropy decreases
- At some point the decrease is insignificant

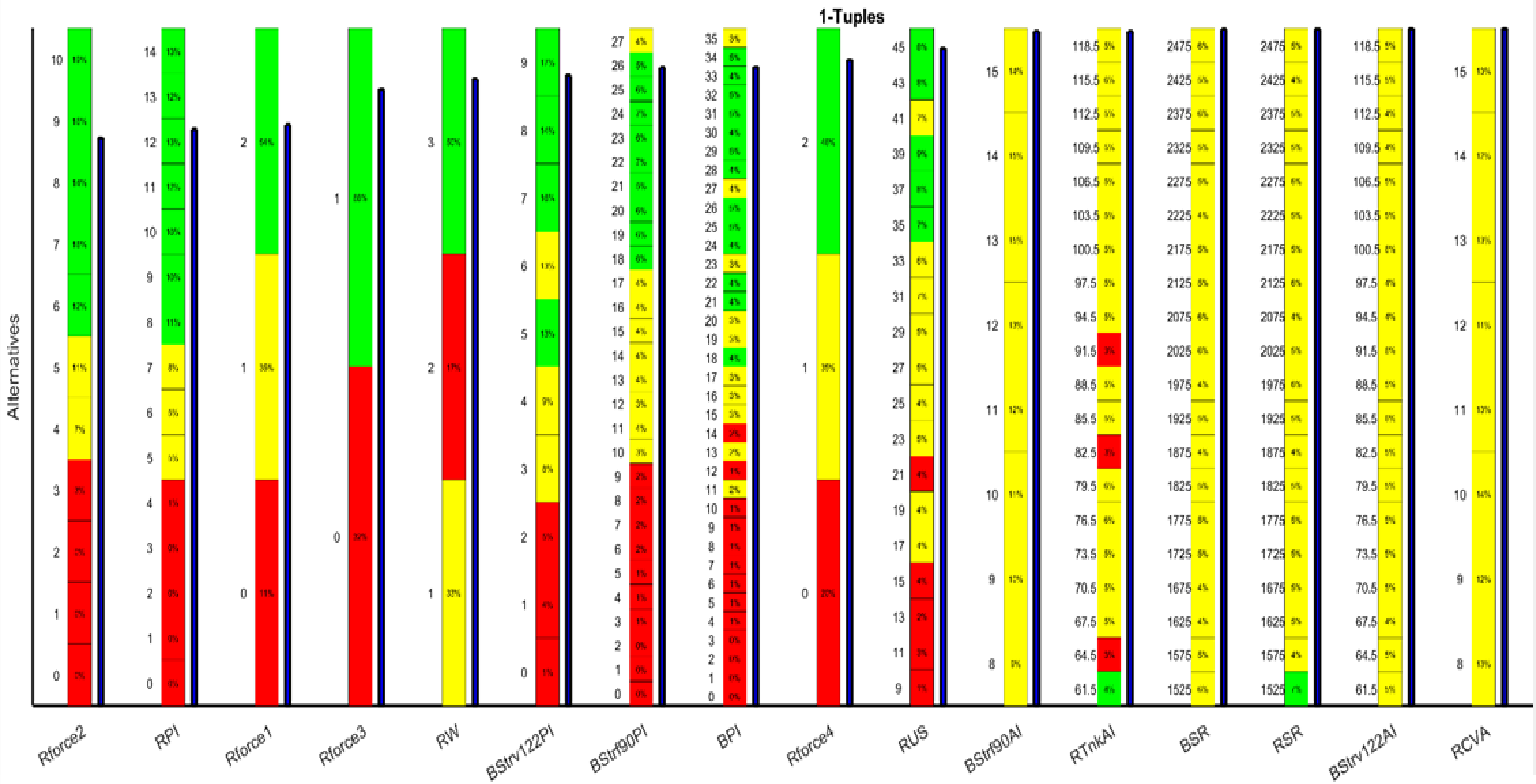


Decision support procedure: **Step 2**

Step 2:

- Which are the **most important parameters**?
- **Which values** give the best result for blue?

Decision support procedure: **Step 2**

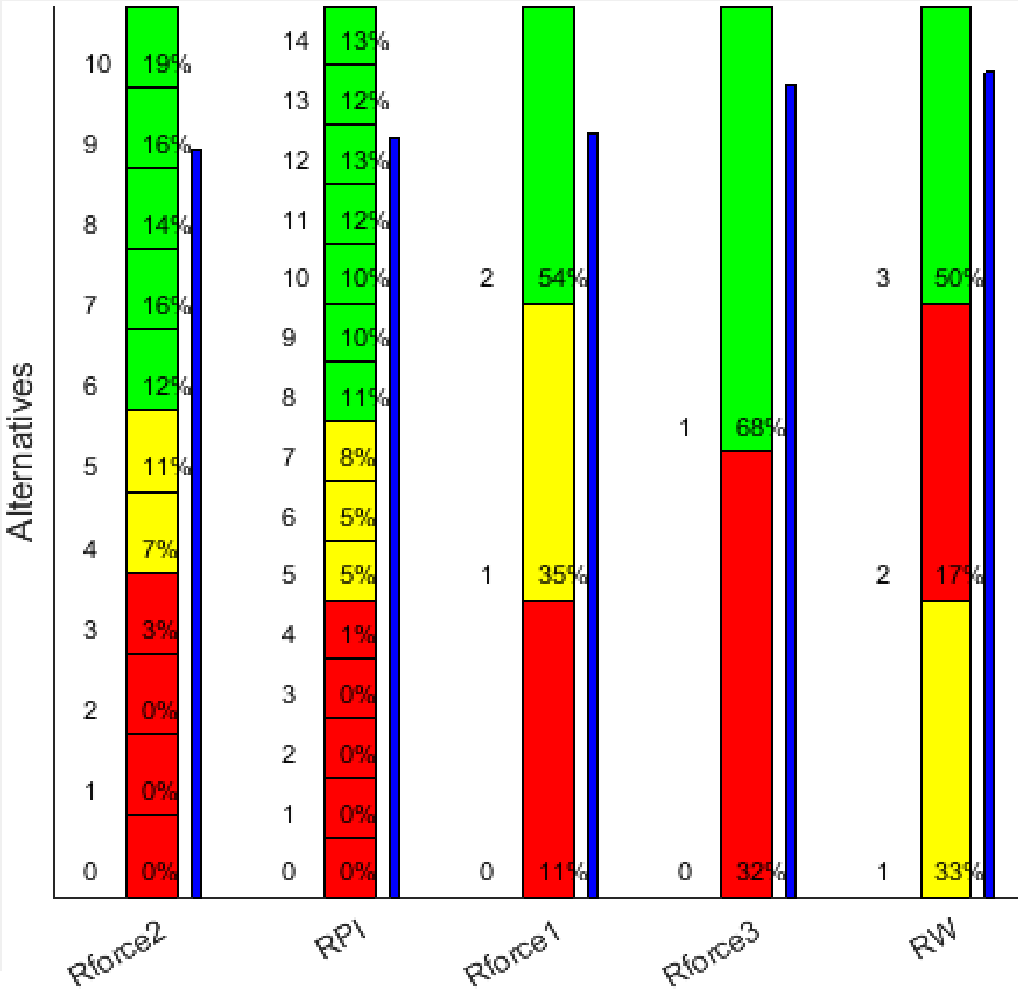


Decision support example

Find the best **factor** and **value**

Look closer at the best **factor**: `Rforce2`

- For `Rforce2` 6-10 are the **best values**
- In contrast, values 0-3 did not yield good result



FOI

Conclusions

- We have presented a methodology for a **Commander's Overview** that presents the big-picture
  - which parameters are important
  - within what ranges must these parameters lie to achieve success in operation planning
- **First step:** taking data farming from its **traditional analytical view** and applying it in a **decision making mode**