Measuring Safety Through the **Distance Between System States** with the RiskSOAP Indicator

Dr Maria **Mikela** CHATZIMICHAILIDOU[a],

Dr Nektarios KARANIKAS[b], Dr Ioannis DOKAS[c]

[a]Imperial College London, United Kingdom

[b]Amsterdam University of Applied Sciences, The Netherlands

[c]Democritus University of Thrace, Greece

Imperial College London

Lloyd's Register Foundation

ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

HVA

Outline

1. Context
2. Problem statement
3. Solution
4. Methodology
5. Applications
6. Results
7. Conclusions

Context

- Modern engineering systems = complex socio-technical structures
  - mission to offer services of high quality, ensure owners' profitability
  - safety & security
  - consist of many parts, controlled by human or automated agents spread throughout different hierarchical levels & networks
  - **accidents are inevitable**; acknowledged need for the use of **systems-theoretic tools to support safety-driven design & operation**
  - **agents be enabled** to perceive & comprehend threats/vulnerabilities & project what they may entail → ***risk-focused Situation Awareness (risk SA)***

Problem statement

- Risk SA presupposes that an agent is offered an **indication of the variability of the system states** to update his/her mental model & adjust the system processes accordingly

- Most critical high-level risk: large **gap between work-as-imagined & work-as-done**

- To maintain the safety levels that the system was originally planned for, controllers must be **aware of the distance between system design & operation**, i.e. maintain a high **risk SA**

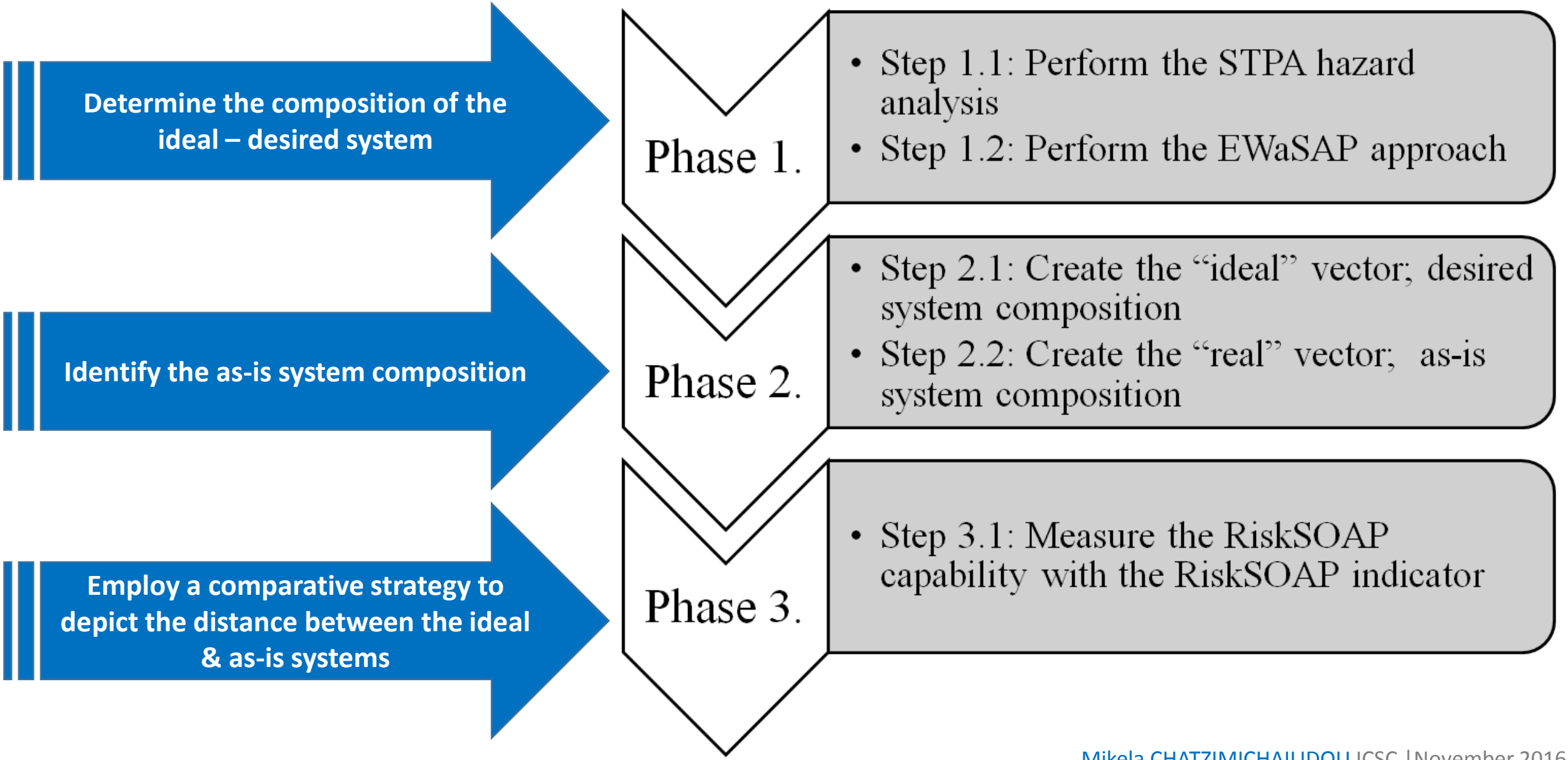
Solution

- Agent's risk SA supported by the **system components & composition**
  - ***system's risk SA provision (RiskSOAP) capability***: operationalised through a **quantification of the differences of various system versions** in regard to safety

- RiskSOAP may be increased or decreased by including or excluding, upgrading, downgrading or maintaining **system parts & elements**, or their properties, throughout the system's lifecycle

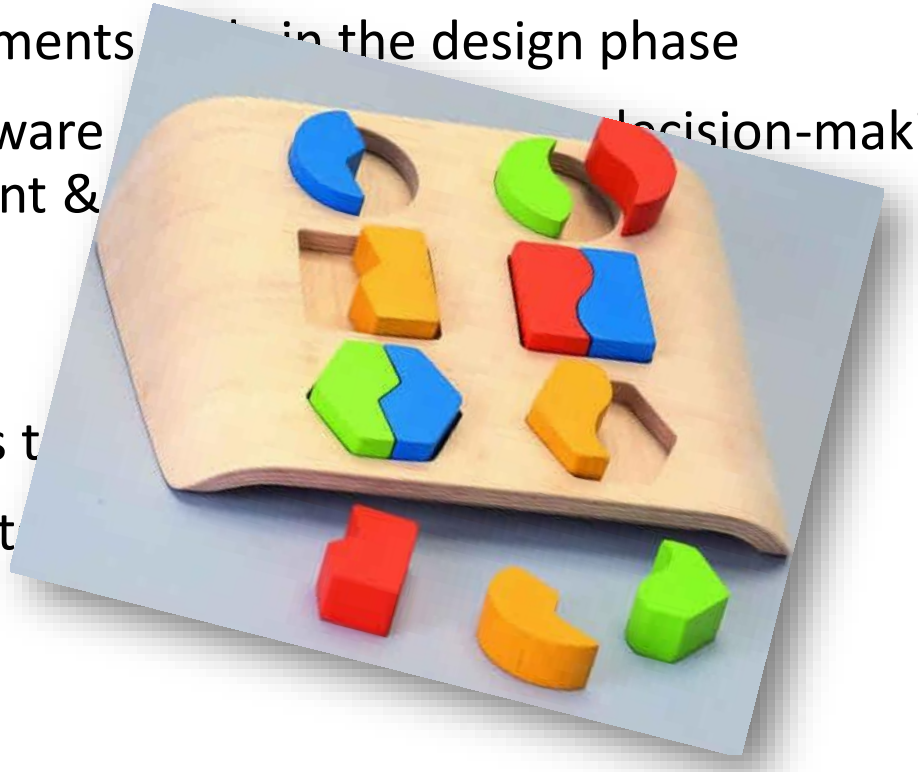
- **RiskSOAP methodology**
  - summary of previous publications as a means to provide an overall view & a comprehensive demonstration of its applicability

RiskSOAP methodology



Methods & tools

- **STPA** (Leveson, 2011)
  - Safety = a dynamic control problem & an emergent property
  - Accidents ≠ chains of event & component failures
  - Non-probabilistic top-down approach; safety requirements in the design phase
  - Inadequate control actions & causal factors (e.g. software decision-making flaws, conflicts between controllers, poor management &



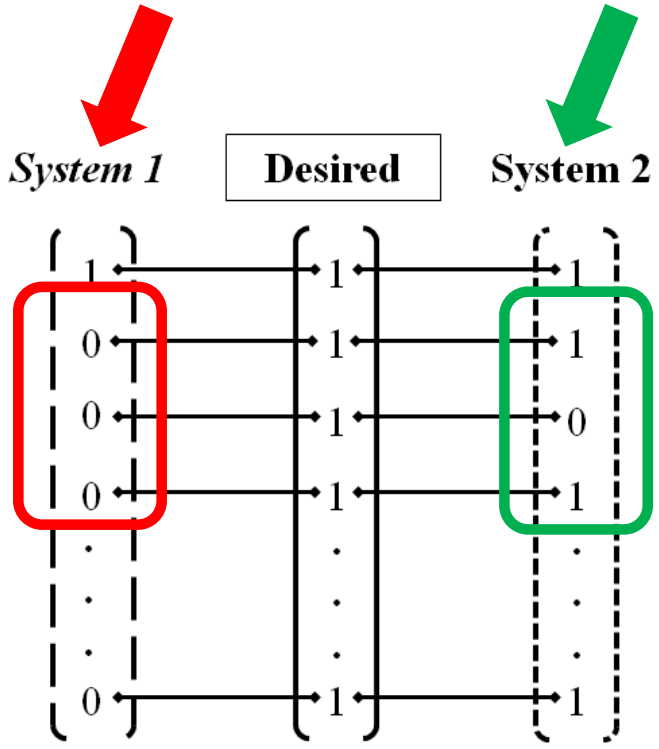
- **EWaSAP** (Dokas et al., 2013)
  - Structured method for identifying early warning signs t
  - Awareness actions: provide warning messages & alert

- **Dissimilarity measure** (pattern matching)
  - Rogers-Tanimoto:

$$RTd(i,r) = \frac{2S10 + 2S01}{S11 + S00 + 2S10 + 2S01}$$

Hypothesis checked

"Provided that there are **more than one versions** of the same system that differ in their composition, the RiskSOAP methodology is adopted and the **RiskSOAP** indicator is calculated **as many times as the different alternative versions** of the system. After obtaining these values, it is expected that the **lowest value** for the RiskSOAP indicator will be returned **for the system version that is proclaimed as less vulnerable**, and vice versa."



RiskSOAP indicator for *System 1*

RiskSOAP indicator for **System 2**

IF

RiskSOAP indicator for **System 2**

<

RiskSOAP indicator for *System 1*

THEN

System 2 closer to the desired system version than *System 1*

3 Case studies

- **ACROBOTER**
  - robotic installation for manipulating small objects autonomously/in cooperation with humans
  - original system failed to meet its purposes & deliver the tasks as described in the project scope

- **Überlingen mid-air collision accident**
  - Johnson, 2004; BFU, 2002: technical (e.g. optical STCA, phone connection, TCAS downlink) & organisational (e.g. detailed/unified directives, aid to the ATC) deficiencies

- **Road Tunnel in Greece**
  - planned renovation
  - tunnel not yet monitored by exclusive tunnel control centre

Results I - ACROBOTER

System Elements	STPA & EWaSAP	Operated system	Designed system
Present	213	96	92
Absent	0	117	121
Vectors' length	213	213	213
RiskSOAP indicator		=2*117 + 2*0 / 96 + 0 + 2*117 +2*0 =0.7091	=2*121 + 2*0 / 92 + 0 + 2*121 +2*0 =0.7246
Difference between the 'as-is' system versions		=0.7246 - 0.7091 =0.0155	

Results II - Überlingen mid-air collision

System Elements	STPA & EWaSAP	Operated system	BFU (2002) & Johnson (2004)
Present	279	74	134
Absent	0	205	145
Vectors' length	279	279	279
RiskSOAP indicator		=0.8471	=0.6840
Difference		=0.1631	

Results III – Road tunnel

System Elements	STPA & EWaSAP	Operated system	Directive 2004/54/EC & PIARC
Present:	191	109	153
Absent:	0	82	38
Vector's length:	191	191	191
RiskSOAP indicator		=0.6007	=0.3319
Difference		=0.2688	

Observations

- RiskSOAP **numerical** expression; differences between system versions communicated in a **simple & understandable** manner → **lower RiskSOAP indicator**:
  - **safer system**; system components & characteristics effectively operationalised
  - safety **risk SA** of the system agent(s) **enhanced**

- Calculation before & after changes; **management can act upon system changes** based on: the different values, their fluctuation & the priorities/constraints of the industry

- 3 apps; **improved system compositions** correspond to **lower indicator** values

- Potential for **safety improvements** even for **systems** contemplated as **safe enough**

Final remarks I

- RiskSOAP measurement: **distance** of systems **from** their **ideal** composition & potential for further **improvements** in terms of safety constraints generated through STPA & EWaSAP

- RiskSOAP indicator
  - **selection criterion** between **alternative design versions** of the same system; by comparing the vectors corresponding to different design versions
  - **decision-making tool** when evaluating **system changes** with reference to the ideal system; system composition **modifications** to shorten the distance between vectors & 'lessen' the operated-ideal system dissimilarity
    - **criterion for evaluating modifications** that might affect the minimum acceptable safety level of the system under consideration; **threshold value** can be set

Final remarks II

- STPA & EWaSAP require **experienced & well qualified** analysts, supported by interdisciplinary team with common & complementary understanding of the system under study

- **Subjective** interpretation of the value of the indicator; may differ across systems & designers, affecting the degree of interventions in the systems

- **Binary**-based indicator; no intermediate case between absence-presence, but variables may have true value ranging between ‘0’ & ‘1’

Future work

- Consider dissimilarity measures which account for **intermediate** states of system components

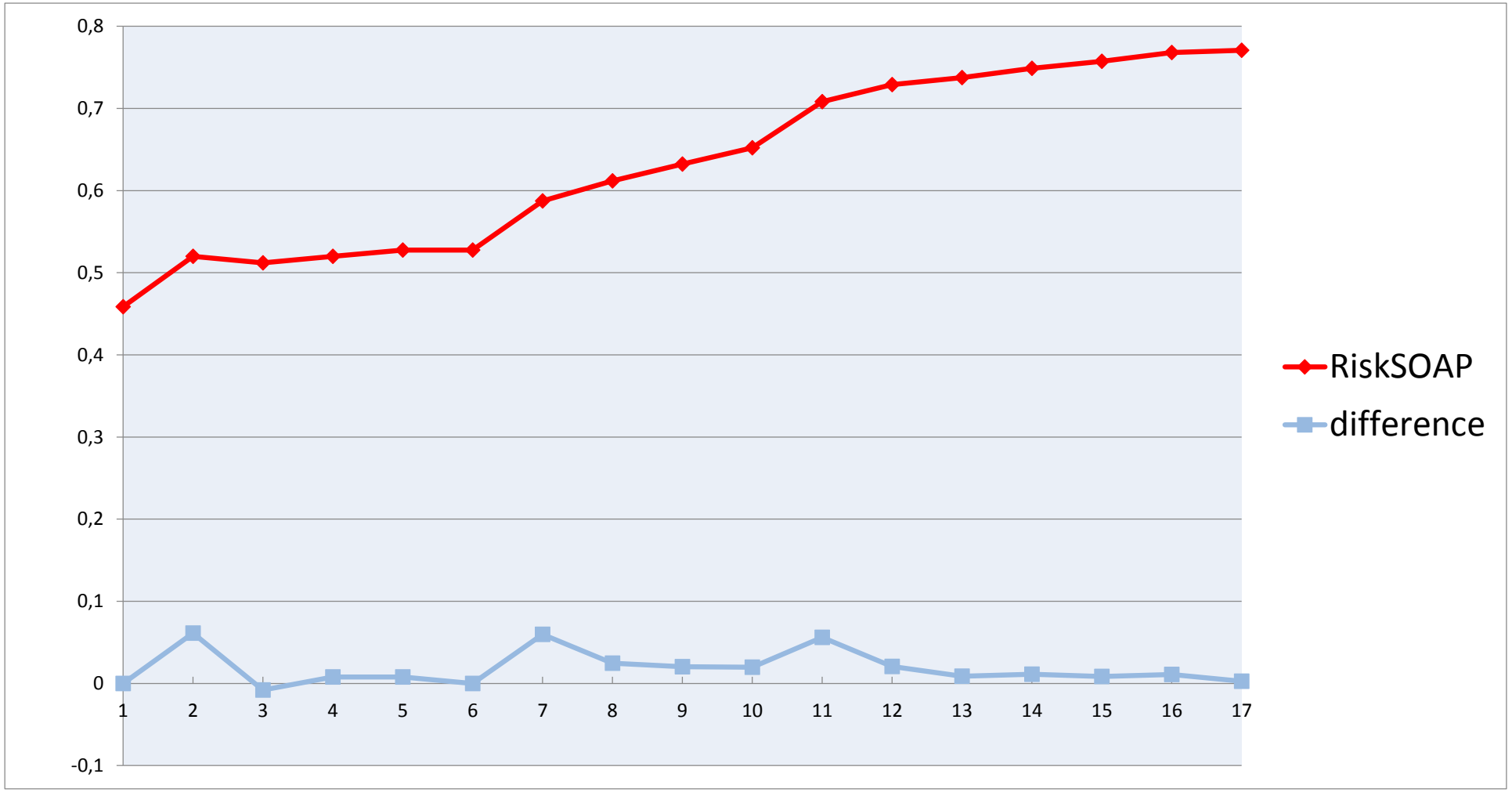
- STPA analysis applicable to:
  - emergent properties (other than safety) such as **security & quality**
  - various business functions such as production & **finance**

    → if purpose of system related to economics, then system elements decided by economic/**econometric** models

- RiskSOAP as a **leading safety indicator**; awareness of system deviations from its desired state → (re)design & operation of safer systems

Leading Safety Indicator.. *maybe? how?*

Überlingen RiskSOAP values along the accident's timeline [preliminary results]



Thank you!

contact:

m.chatzimichailidou@imperial.ac.uk

mikelachatzimichailidou@gmail.com

http://www.imperial.ac.uk/people/m.chatzimichailidou

ResearchGate, LinkedIn, Twitter, fb

Imperial College London

Lloyd's Register Foundation

ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

HVA