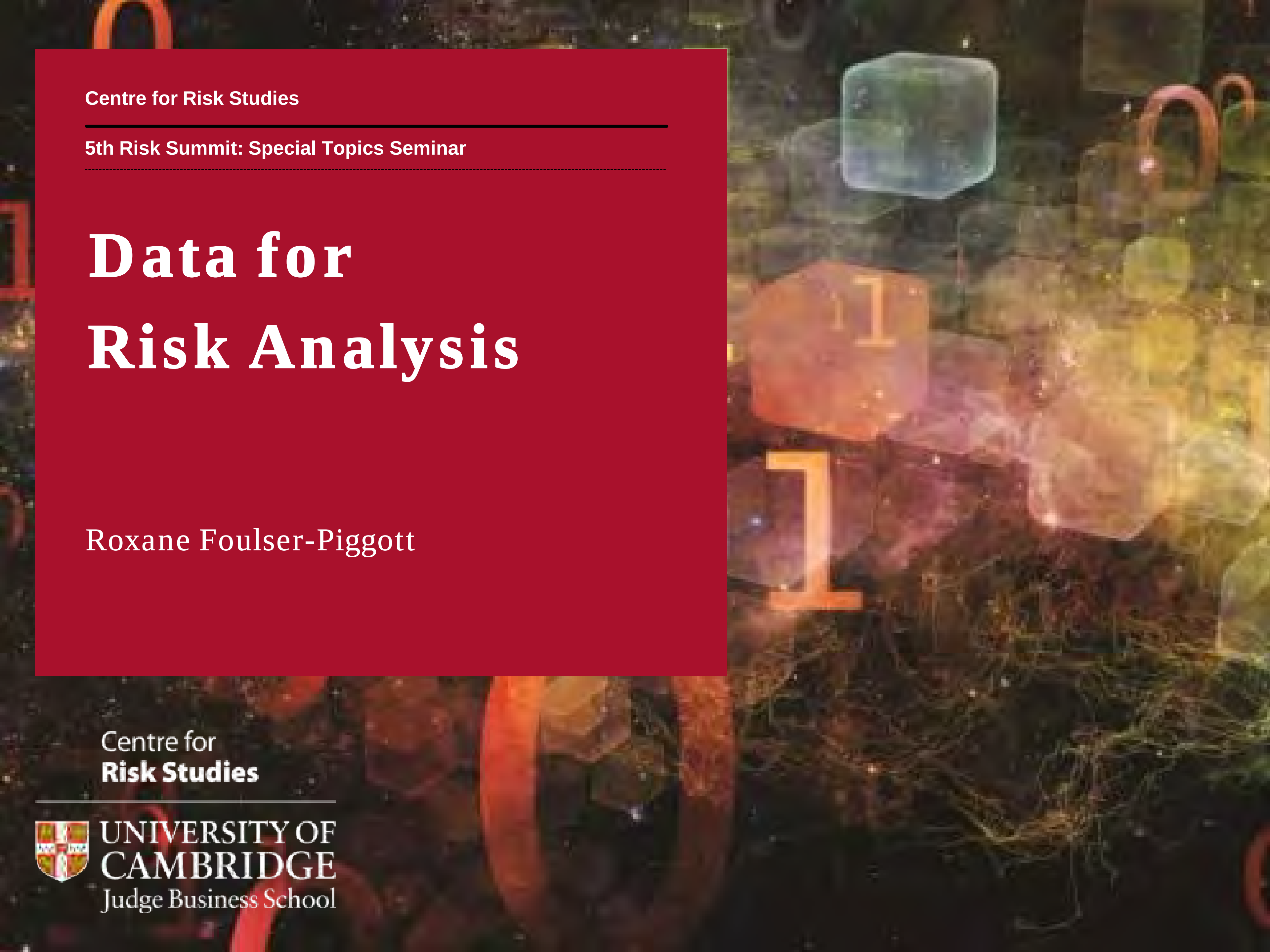


Centre for Risk Studies

5th Risk Summit: Special Topics Seminar

Data for Risk Analysis

Roxane Foulser-Piggott

Centre for
Risk Studies

UNIVERSITY OF
CAMBRIDGE
Judge Business School

Big Data

- Datasets that are too large or complex to manipulate or interrogate with commonly used methods or tools (Snijders et al., 2012) e.g. Boston “street bump” app, Google flu trends.
- “N = all” – no longer need to sample, the sample contains everyone.
- Challenges include capture, curation, storage, search, sharing, transfer, analysis and visualisation.
- Big data is difficult to work with using most relational database management systems and desktop statistics and visualisation packages.

Uncertainties

- Uncertainties in models can be divided into aleatory and epistemic uncertainties.
- Aleatory uncertainties are due to natural randomness and cannot be reduced.
- Epistemic uncertainties are due to lack of data or knowledge.
- Uncertainties in models can be reduced by acquiring more, better quality data.
- Does big data offer the capacity to reduce epistemic uncertainty?

Partitioning Uncertainty

$$Y = b_1 X_1 + b_2 X_2$$

$$Y = f(\mathbf{B}, \mathbf{X}_\mathbf{k})$$

$$Y - \hat{Y} = \sigma$$

$$\sigma_T = \sqrt{\overset{\text{Aleatory}}{\sigma_A^2} + \overset{\text{Epistemic}}{\sigma_E^2}}$$

$$\sigma_{T_{reduced}} = \sqrt{\sigma_T^2 - \sigma_{T_{vu}}^2}$$

$$\sigma_{T_{vu}}^2 = \left|\frac{\partial Y}{\partial X_1}\right|^2 \sigma_{X_1}^2 + \left|\frac{\partial Y}{\partial X_2}\right|^2 \sigma_{X_2}^2$$

- σ is often assumed to be aleatory.
- σ has epistemic components resulting from:
  1. Inexact form of the model and selection of particular model formulation
  2. Selection of a particular database
  3. Input variable measurement error
  4. Statistical errors in the estimation of parameters
- Points 2 and 3 in this list are directly related to the data used:
  - 2: Use decision tree techniques.
  - 3: Use MC simulations or FOSM.

Source: Foulser-Piggott R. (2014) Quantifying the epistemic uncertainty in ground motion models and prediction, Soil Dynamics and Earthquake Engineering

Big Data Issues

- Boston “Street Bump” app, Google Flu trends, Twitter trends.
- “There are a lot of small data problems that occur in big data” Spiegelhalter[1].
- “N = all” Can big data ever give information about the whole population?
- Does measurement error still exist and is it significant?
- Can we be any clearer about causation and correlation in models using big data?

[1] Financial Times article: Big Data: are we making a big mistake, Tim Harford (March 28 2014)

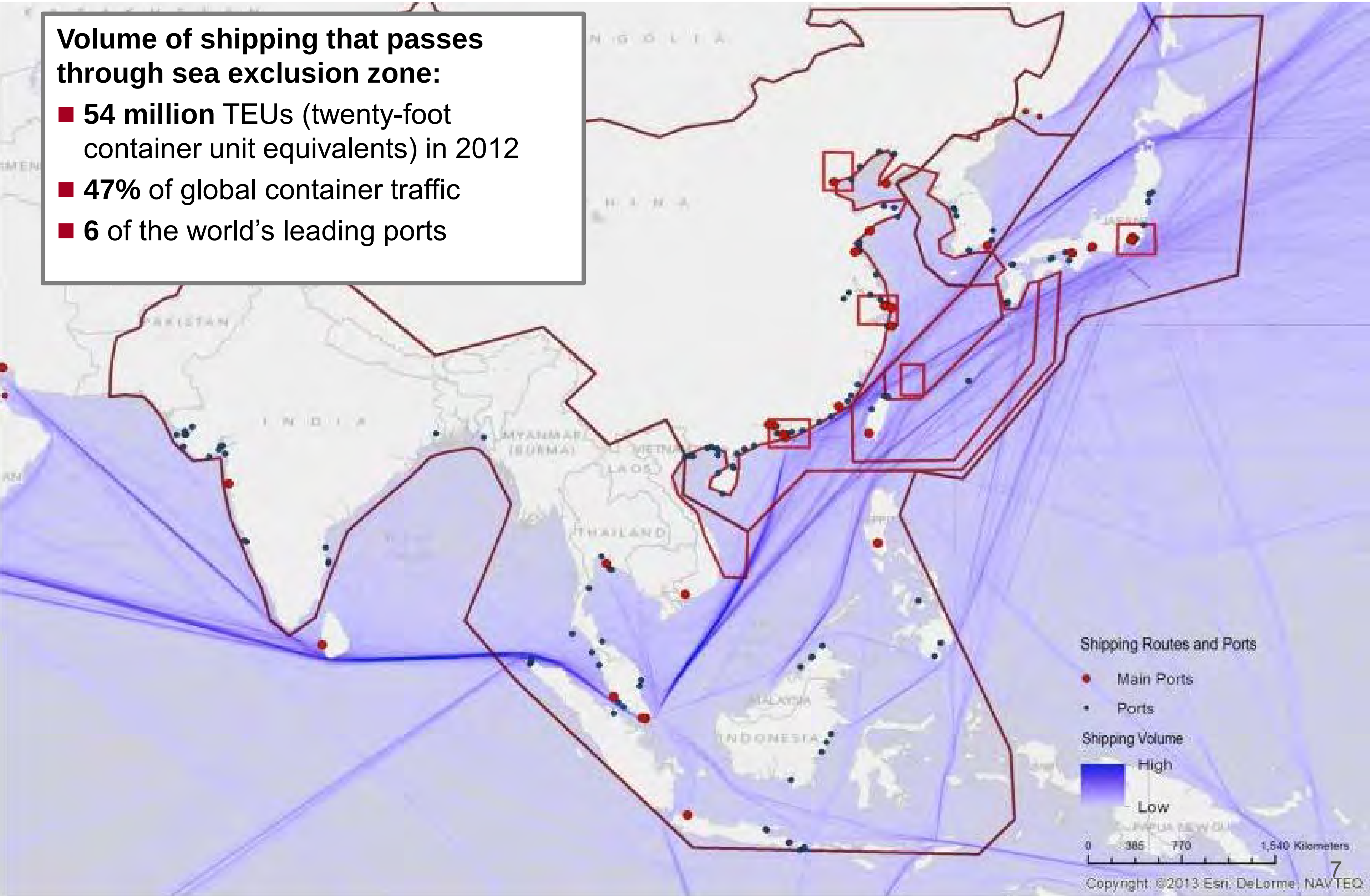
CRS Data

- CRS holds georeferenced datasets which can be used in risk analysis.
- So far, not “big data” but global, low resolution datasets.
- CRS database based on open source data.
- Allows broad understanding of connections between threats (hazards) and different networks (exposure).
- Data analysis example: China-Japan Conflict Scenario

Shipping Lanes Affected by Sea Exclusion Zones

Volume of shipping that passes through sea exclusion zone:

- **54 million** TEUs (twenty-foot container unit equivalents) in 2012
- **47%** of global container traffic
- **6** of the world’s leading ports

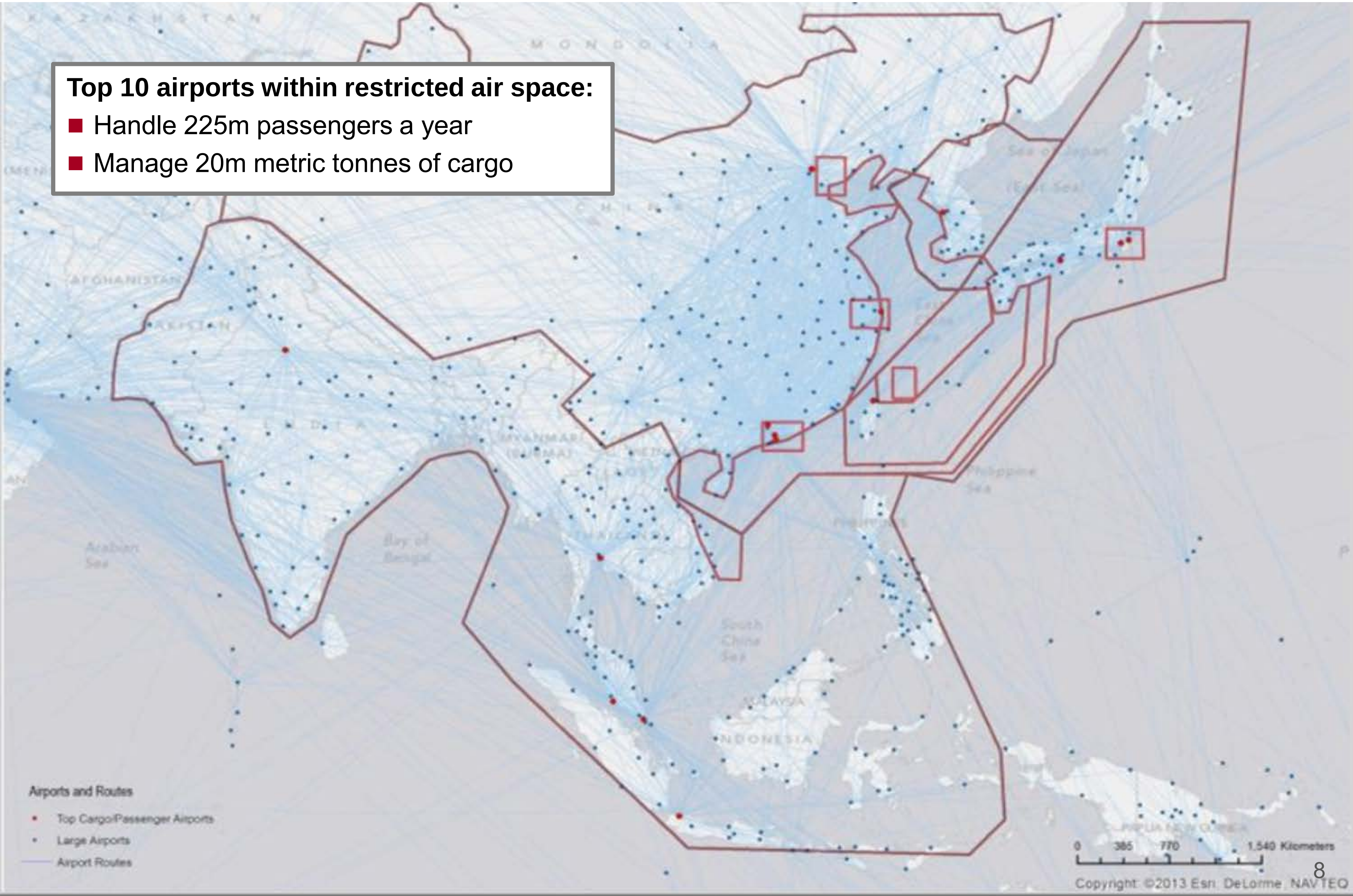


Copyright: ©2013 Esri, DeLorme, NAVTEQ

Commercial Flight Routes Affected by Restricted Air Space

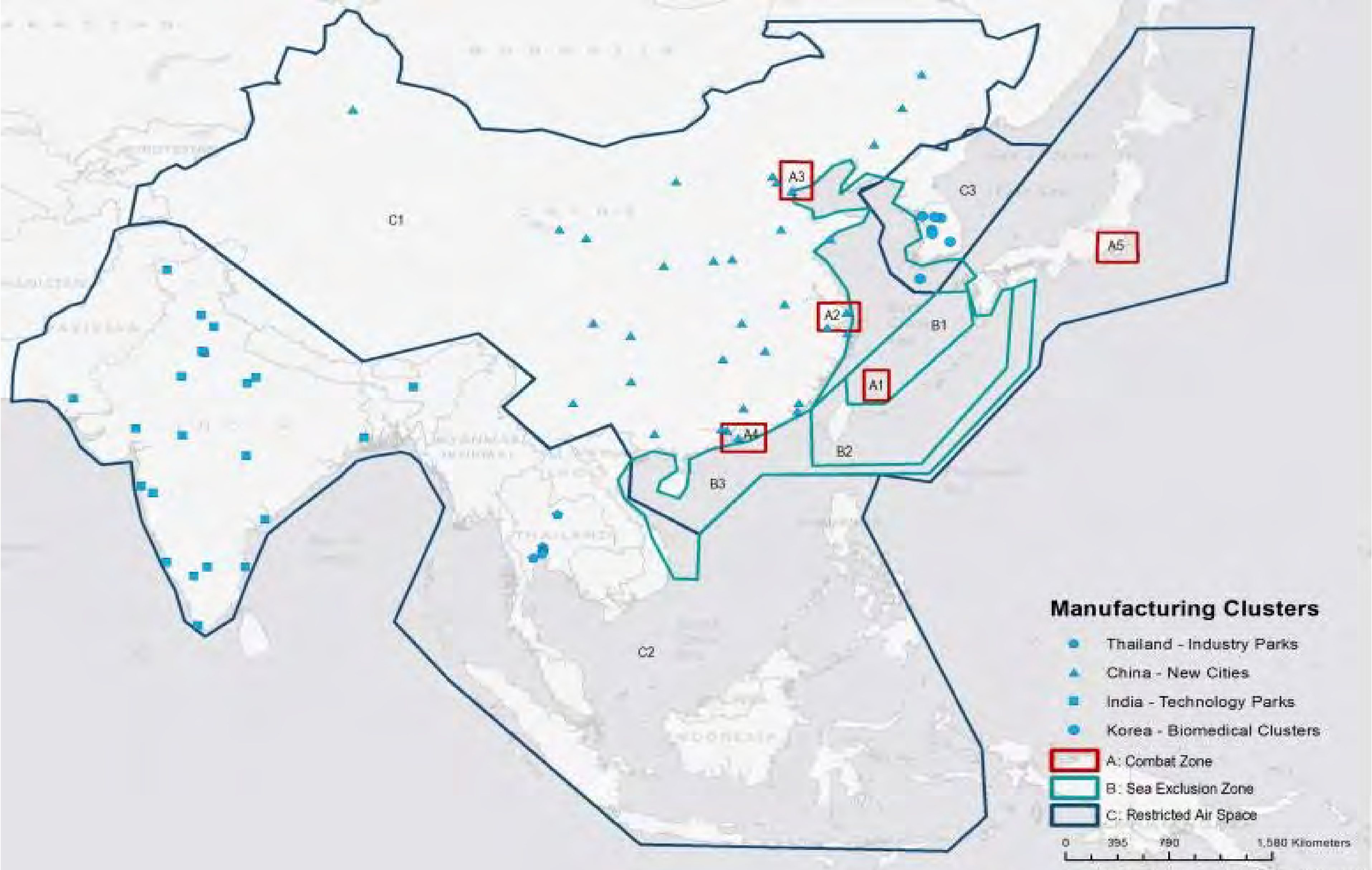
Top 10 airports within restricted air space:

- Handle 225m passengers a year
- Manage 20m metric tonnes of cargo



Copyright: ©2013 Esri, DeLorme, NAVTEQ

Hi-Tec Manufacturing in the War Zone

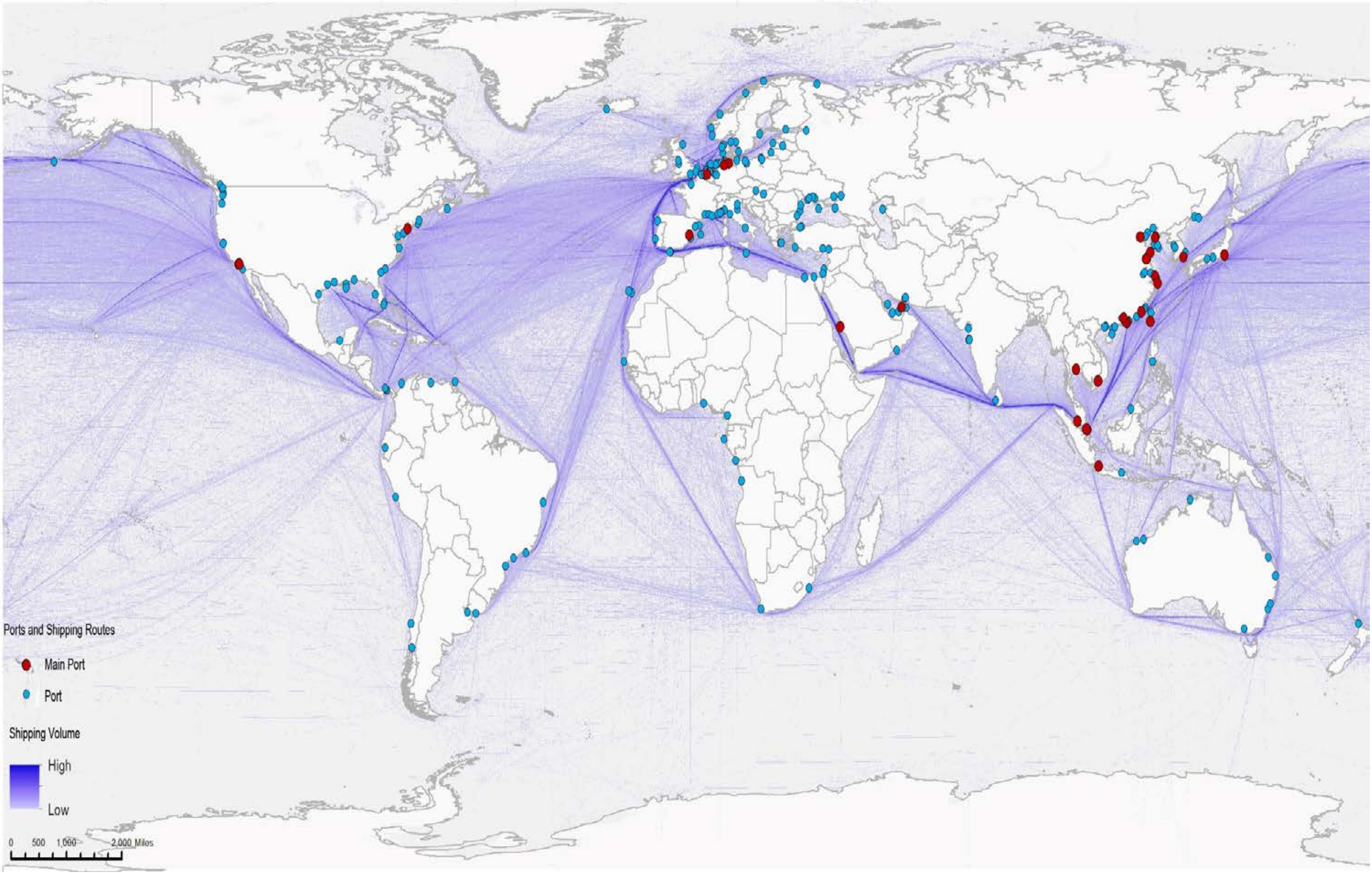

C1
C2
C3
A1
A2
A3
A4
A5
B1
B2
B3
THAILAND
Manufacturing Clusters
● Thailand - Industry Parks
▲ China - New Cities
■ India - Technology Parks
● Korea - Biomedical Clusters
🟥 A: Combat Zone
🟩 B: Sea Exclusion Zone
🟦 C: Restricted Air Space
0 395 790 1,580 Kilometers

Copyright: ©2013 Esri, DeLorme, NAVTEQ

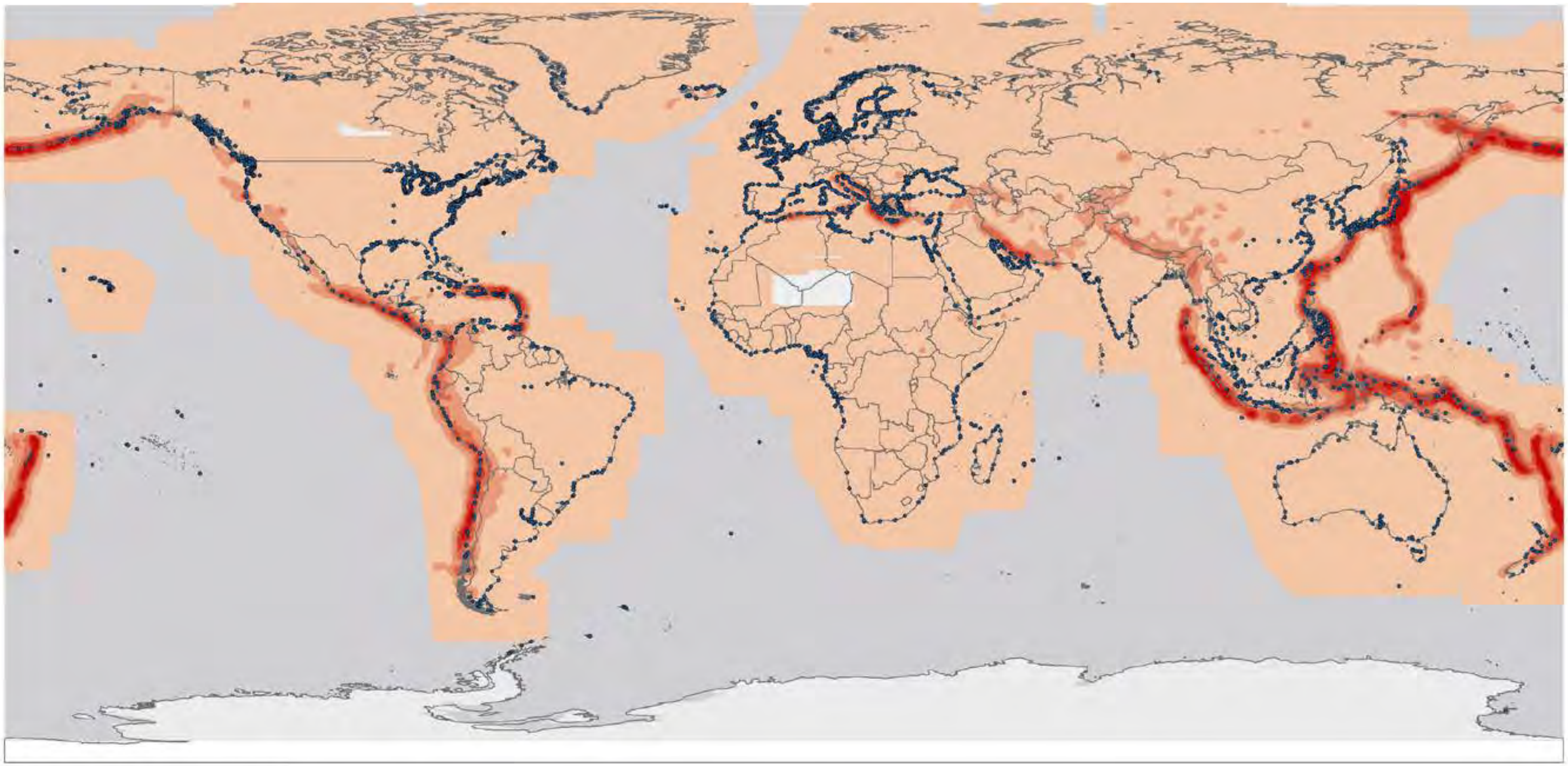
CRS Data – Future Directions

- An example problem: Risk analysis of Singapore port. Research conducted in collaboration with NTU Singapore and Imperial College London.
- CRS Data available:
  - Port location
  - Approximate shipping routes
  - Potential hazards
- Will collecting more data improve our ability to model systemic shocks?
- Will collecting more data reduce uncertainty?
- What data do we need to collect?
- Can we collect or use big data?

Global Shipping Network Routes and Port Locations



Earthquake Hazard



Ports and Peak Ground Acceleration

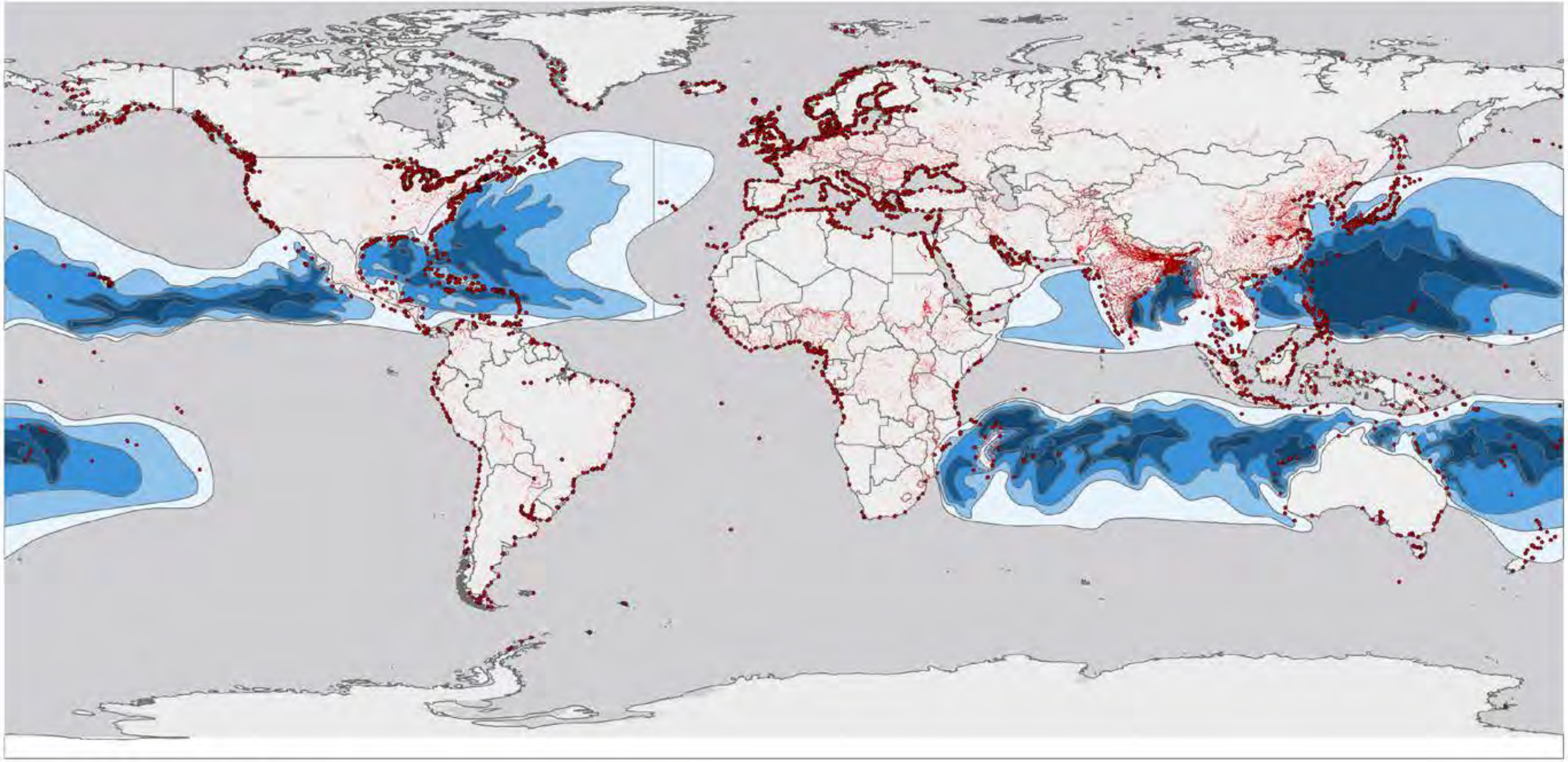
● Port

Peak Ground Acceleration for a 475 year return period (cm/s2)

▨	0.1 - 100.0
▨	100.1 - 250.0
▨	250.1 - 400.0
▨	400.1 - 600.0
▨	600.1 - 1,050.0

Esri, HERE, DeLorme, MapmyIndia, © OpenStreetMap contributors, and the GIS user community

Flood and Storm Hazard



Ports, Flood Risk and Storm Zones

Ports

• Port

Global Flood Risk

Global Estimated Flood Risk Index for Flood Hazard
1 = Low 5=High

🟥 1.0
🟥 2.0
🟥 3.0
🟥 4.0
🟥 5.0

Storm Zones

Zones with 10% probability of wind speed categories 1-5 within the next ten years

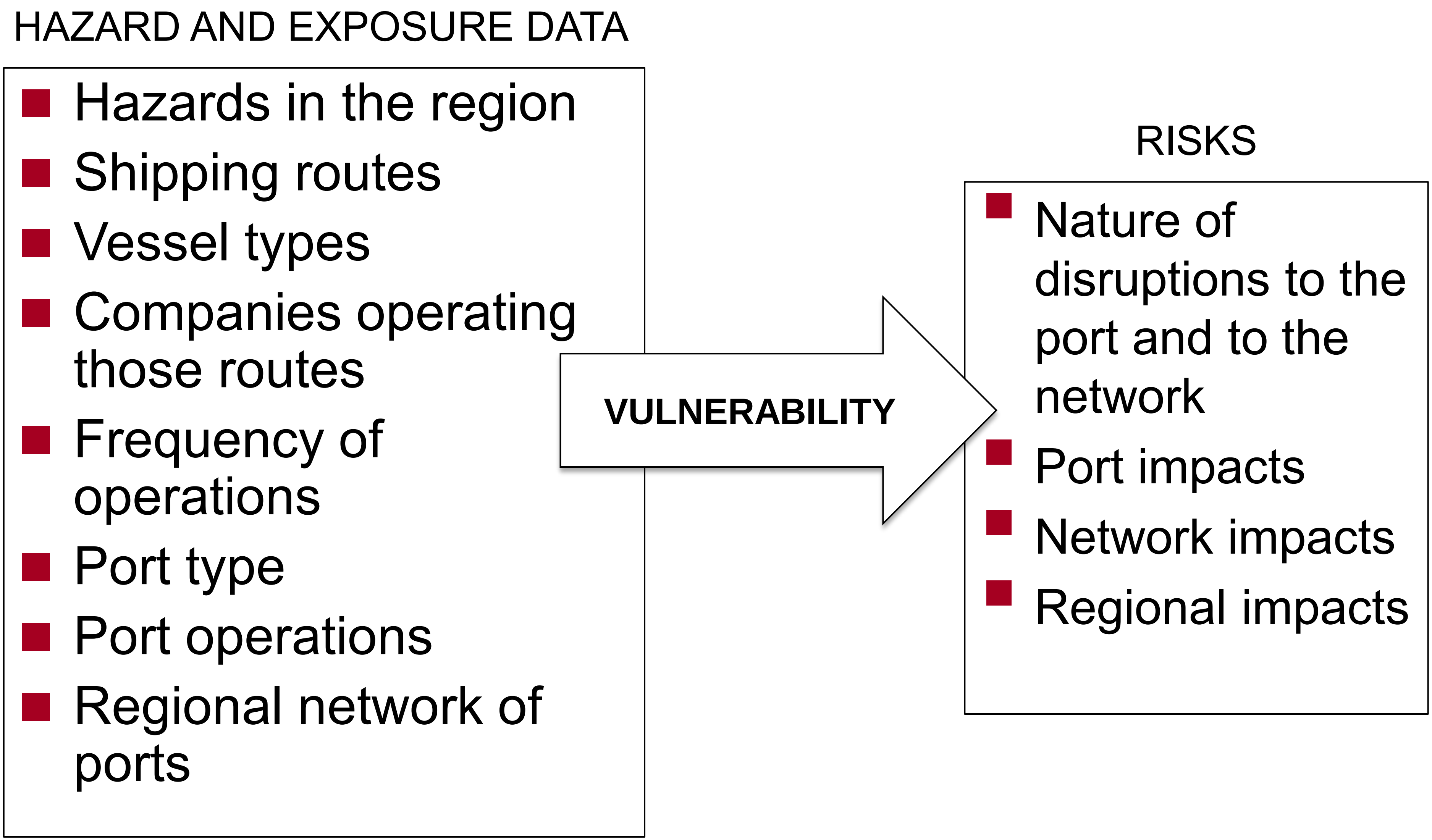
🟦 1) 118 - 153km/h
🟦 2) 154 - 177 km/h
🟦 3) 178 - 209km/h
🟦 4) 210 - 249km/h
🟦 5) 250+ km/hr

Esri, HERE, DeLorme, MapmyIndia, © OpenStreetMap contributors, and the GIS user commur

Defining the Problem

- The choice of methods for collecting data will depend on the variables to be measured, the source and the resources available.
- Aim is to reduce uncertainty in our model by acquiring more data.
- Data for risk analysis falls in three categories:
  - Hazards
  - Vulnerability
  - Exposure
- Detailed understanding of port operations necessary to model exposure and understand impacts of a hazard.

Data Collection



HAZARD AND EXPOSURE DATA

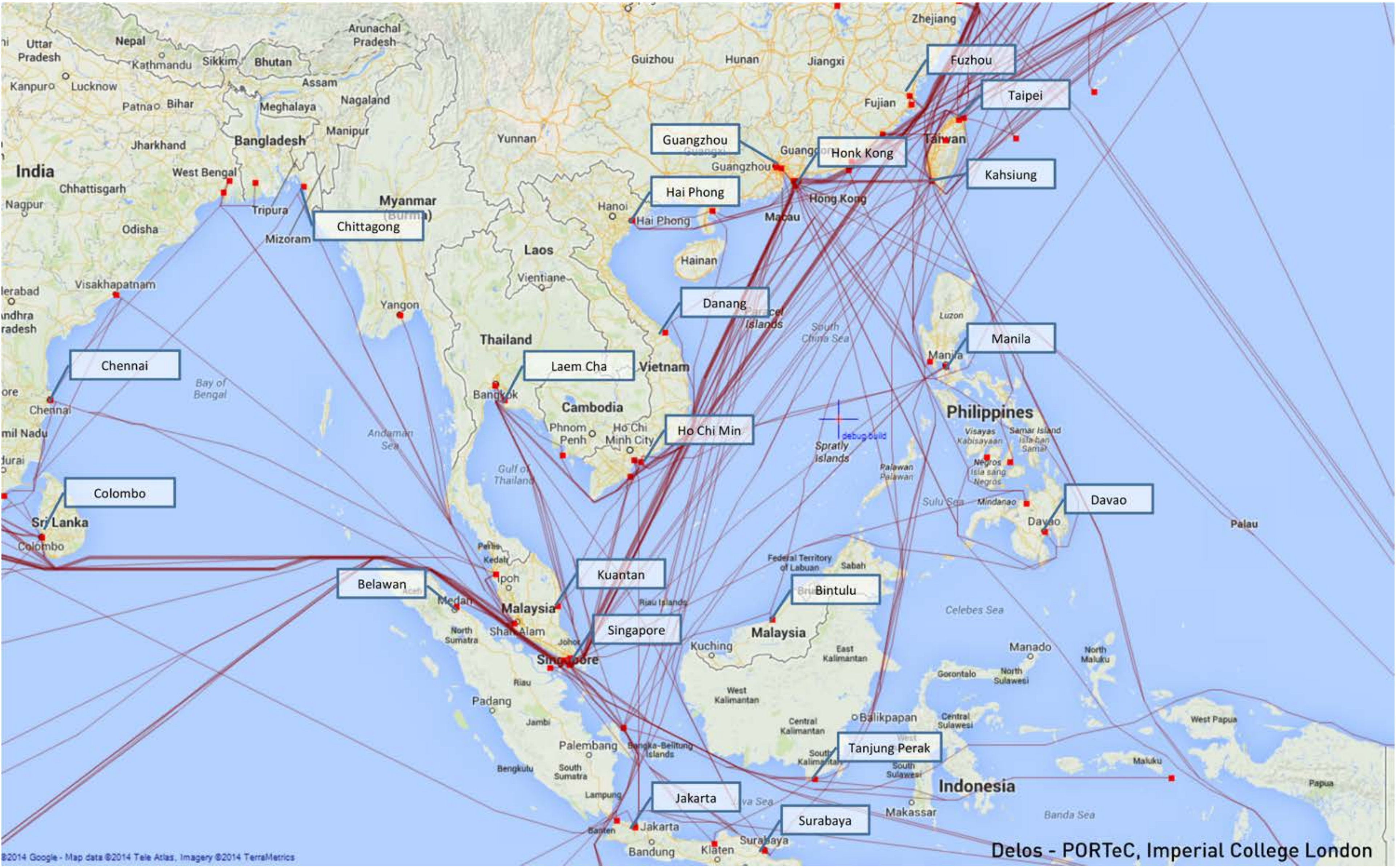
- Hazards in the region
- Shipping routes
- Vessel types
- Companies operating those routes
- Frequency of operations
- Port type
- Port operations
- Regional network of ports

VULNERABILITY

RISKS

- Nature of disruptions to the port and to the network
- Port impacts
- Network impacts
- Regional impacts

Detailed Shipping Routes - Singapore



©2014 Google - Map data ©2014 Tele Atlas, Imagery ©2014 TerraMetrics

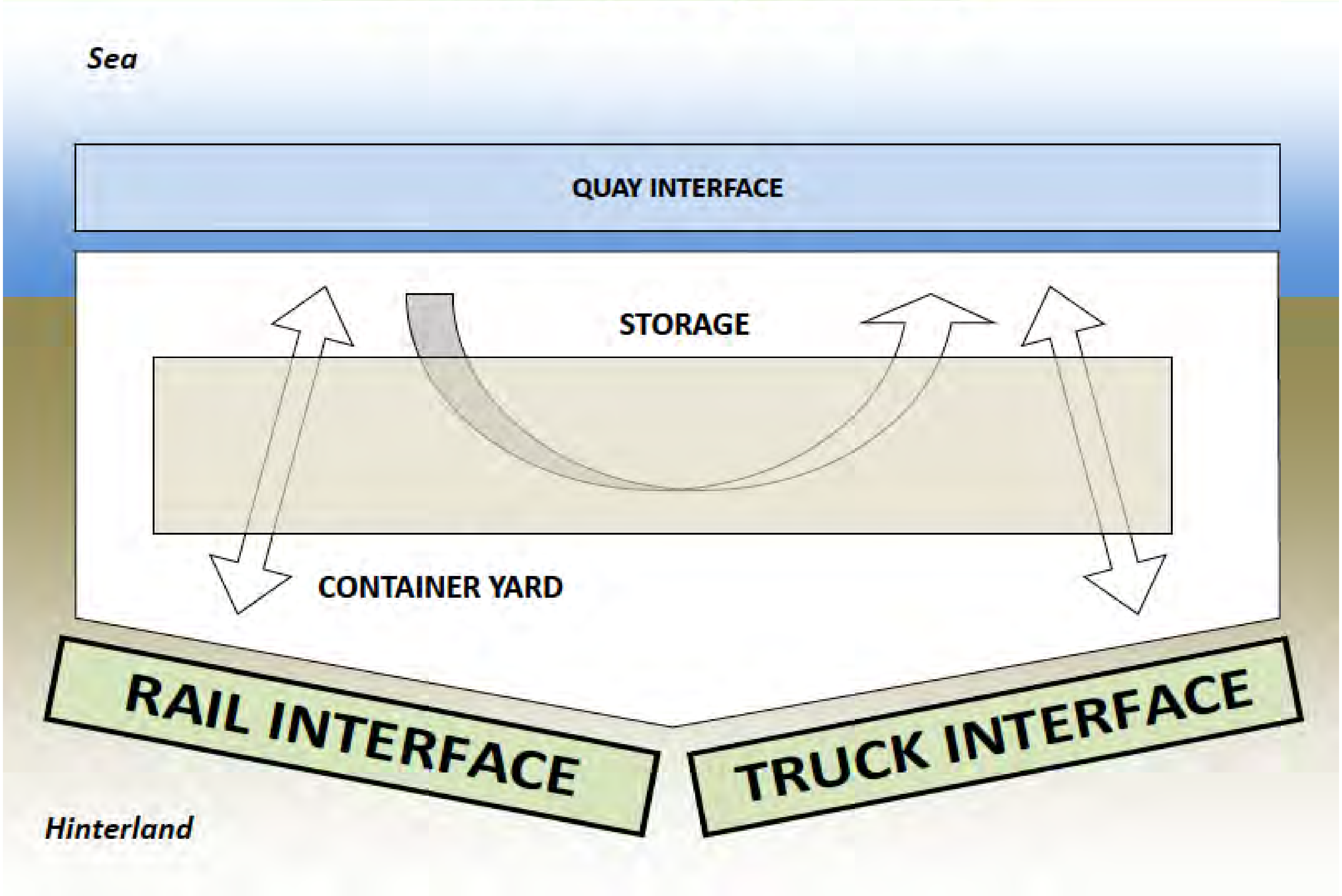
Delos - PORTeC, Imperial College London

With thanks to Matteo Novati and Dr Panagiotis Angeloudis, Imperial College London

Slide courtesy of Dr Panagiotis Angeloudis, Imperial College.

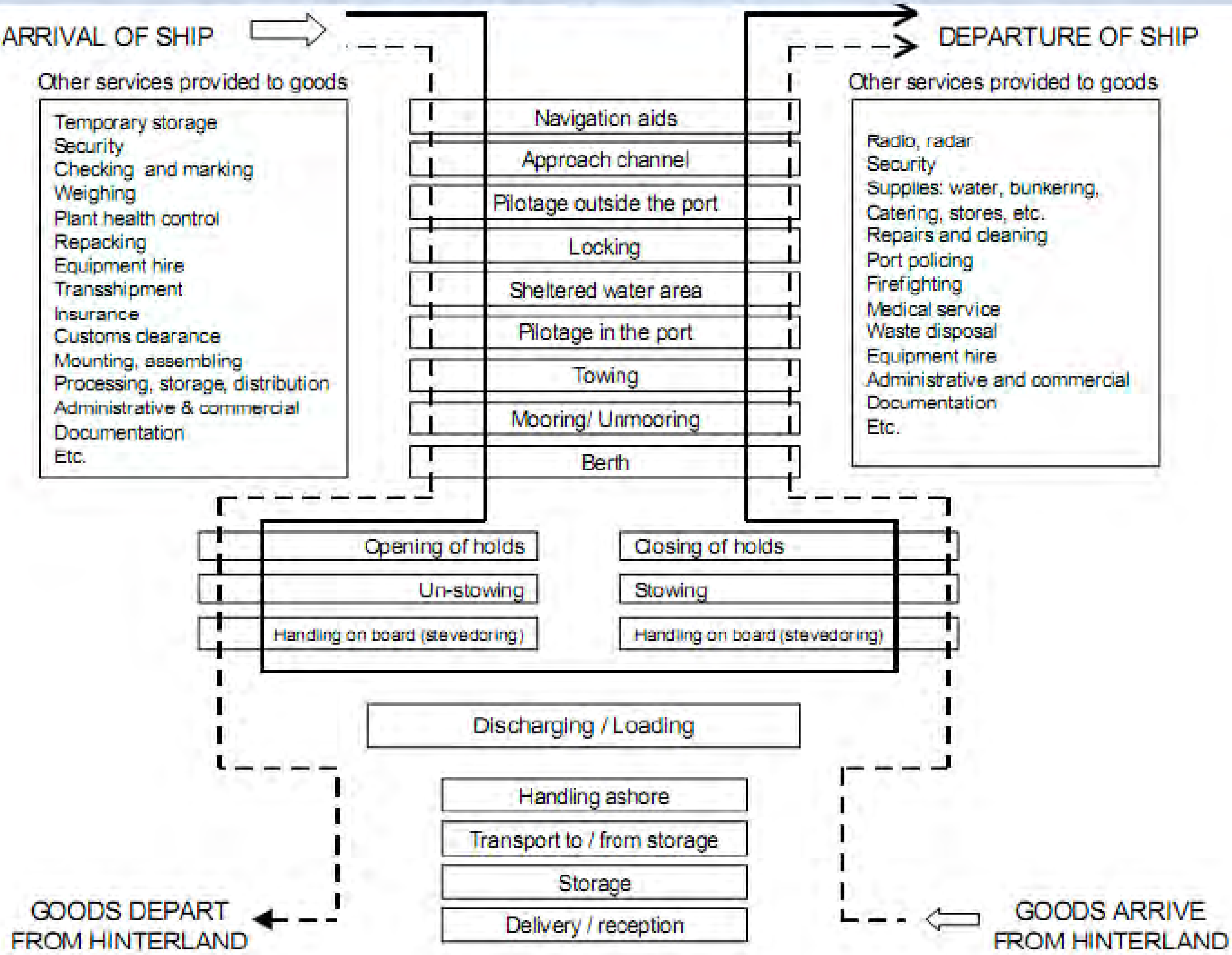
Slide courtesy of Dr Panagiotis Angeloudis, Imperial College.

Summary of operations



Slide courtesy of Dr Panagiotis Angeloudis, Imperial College.

Sum of operations



ARRIVAL OF SHIP

Other services provided to goods

Temporary storage
Security
Checking and marking
Weighing
Plant health control
Repacking
Equipment hire
Transshipment
Insurance
Customs clearance
Mounting, assembling
Processing, storage, distribution
Administrative & commercial
Documentation
Etc.

Navigation aids
Approach channel
Pilotage outside the port
Locking
Sheltered water area
Pilotage in the port
Towing
Mooring/ Unmooring
Berth

DEPARTURE OF SHIP

Other services provided to goods

Radio, radar
Security
Supplies: water, bunkering,
Catering, stores, etc.
Repairs and cleaning
Port policing
Firefighting
Medical service
Waste disposal
Equipment hire
Administrative and commercial
Documentation
Etc.

Opening of holds
Un-stowing
Handling on board (stevedoring)

Closing of holds
Stowing
Handling on board (stevedoring)

Discharging / Loading

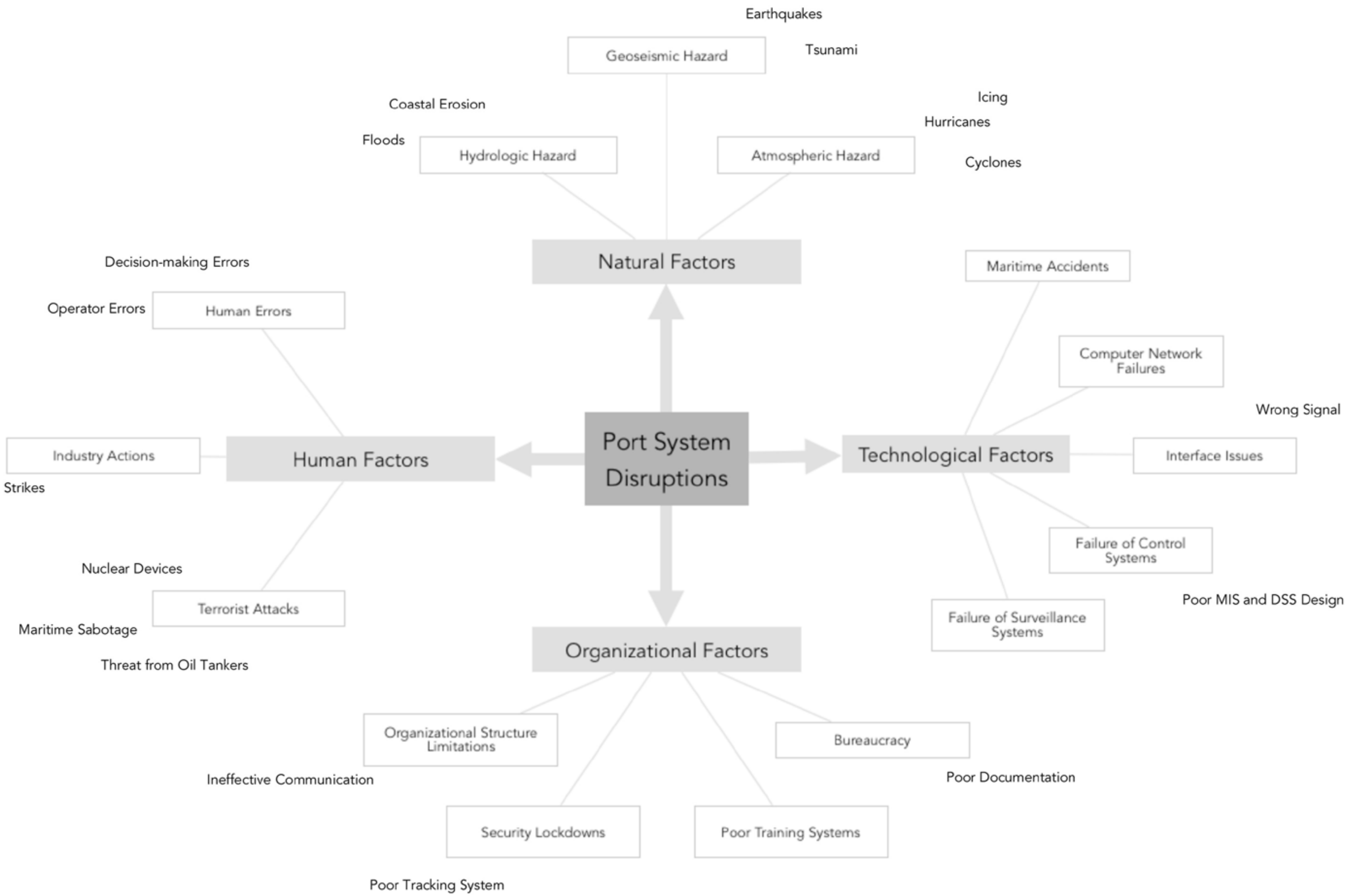
Handling ashore
Transport to / from storage
Storage
Delivery / reception

GOODS DEPART FROM HINTERLAND

GOODS ARRIVE FROM HINTERLAND

Slide courtesy of Dr Panagiotis Angeloudis, Imperial College.

Port system disruptions



UNIVERSITY OF CAMBRIDGE
Judge Business School

Centre for **Risk Studies**

With thanks to Matteo Novati and Dr Panagiotis Angeloudis, Imperial College London

Big Data – Filling in the Gaps

- What “big data” datasets could improve our model for risk analysis of Singapore port?
  - Real-time tracking of ships?
  - Information on each individual container movement?
  - Real-time weather information?
  - Tesco clubcard information on truck drivers?
  - Tweets of dock workers?
- Can “big data” help answer the question?
- Is gathering more data reducing uncertainty or helping us partition uncertainty?

Summary

- Current CRS dataset allows broad understanding of connections between threats (hazards) and different networks (exposure).
- Using traditional datasets vs. big data to better understand a problem.
- Using traditional datasets vs. big data to better understand uncertainties and potentially reduce uncertainties in our models.
- Uncertainties in the data and their propagation through the model requires further research.
- There are challenges with storing, analysing and visualising high resolution data with global coverage.

Centre for
Risk Studies

UNIVERSITY OF CAMBRIDGE
Judge Business School

Dr. Roxane Foulser-Piggott
r.foulser-piggott@jbs.cam.ac.uk