Department of Environmental Science

Zoran Steinmann

Estimating missing data
How to fill data gaps in life cycle inventories?

LCA Discussion forum, Zürich

30 March 2017

Radboud University Nijmegen

Problem setting

LCA: environmental impact of a process/product

Requires considerable amount of data

Common problem: **lack of data in inventory and impact assessment**

Typically make assumptions to address the issue

Questions

- What can we do to fill these data gaps?

- How much data do we really need in impact assessment?

Plant specific carbon footprints of fossil power plants

Dominant source (67%) of global electricity generation

Relatively high CO_2 emissions per kWh generated

Limited data availability on plant specific level

Coal, Oil & Natural gas

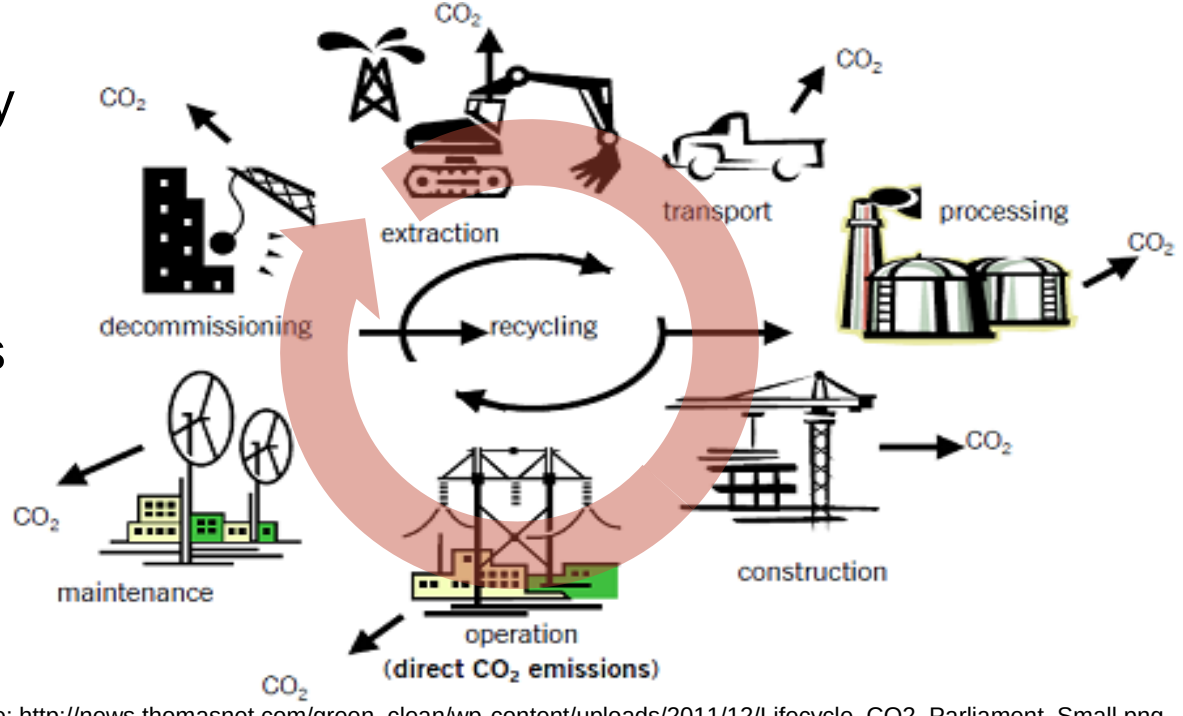
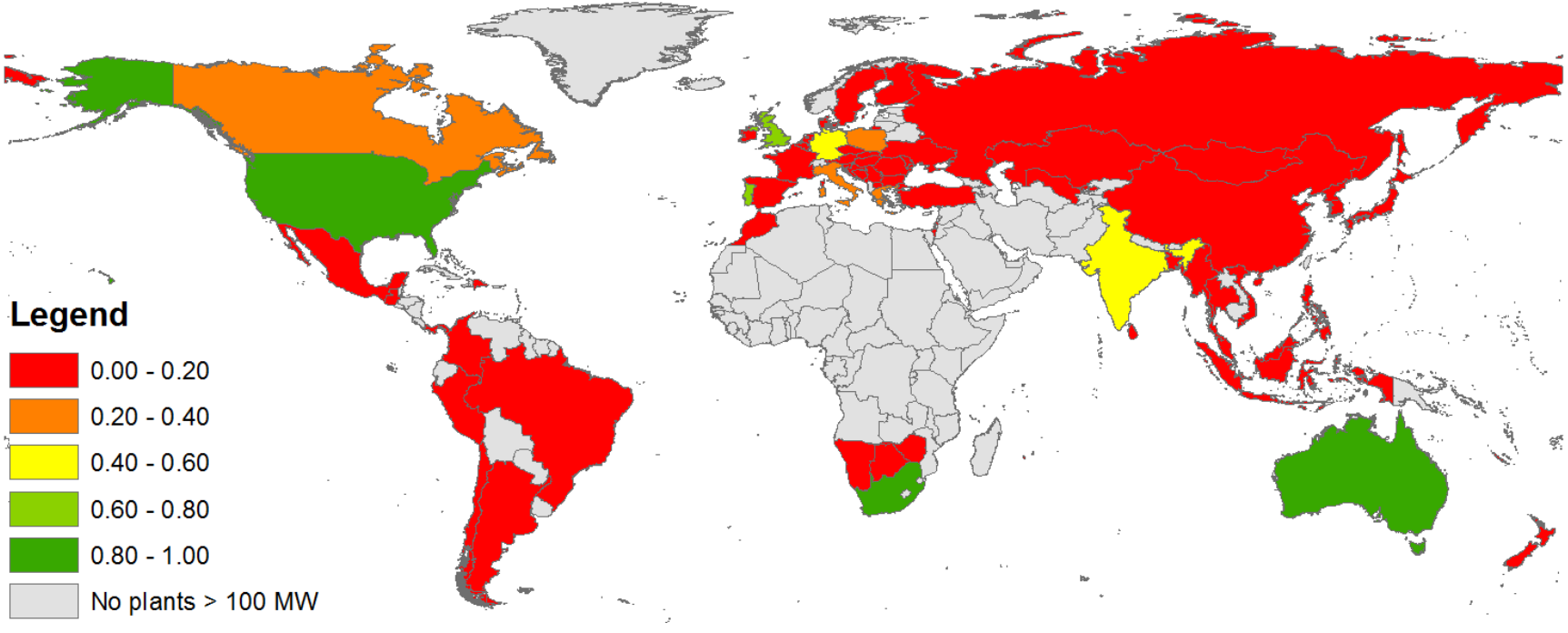


Image Source: http://news.thomasnet.com/green_clean/wp-content/uploads/2011/12/Lifecycle_CO2_Parliament_Small.png

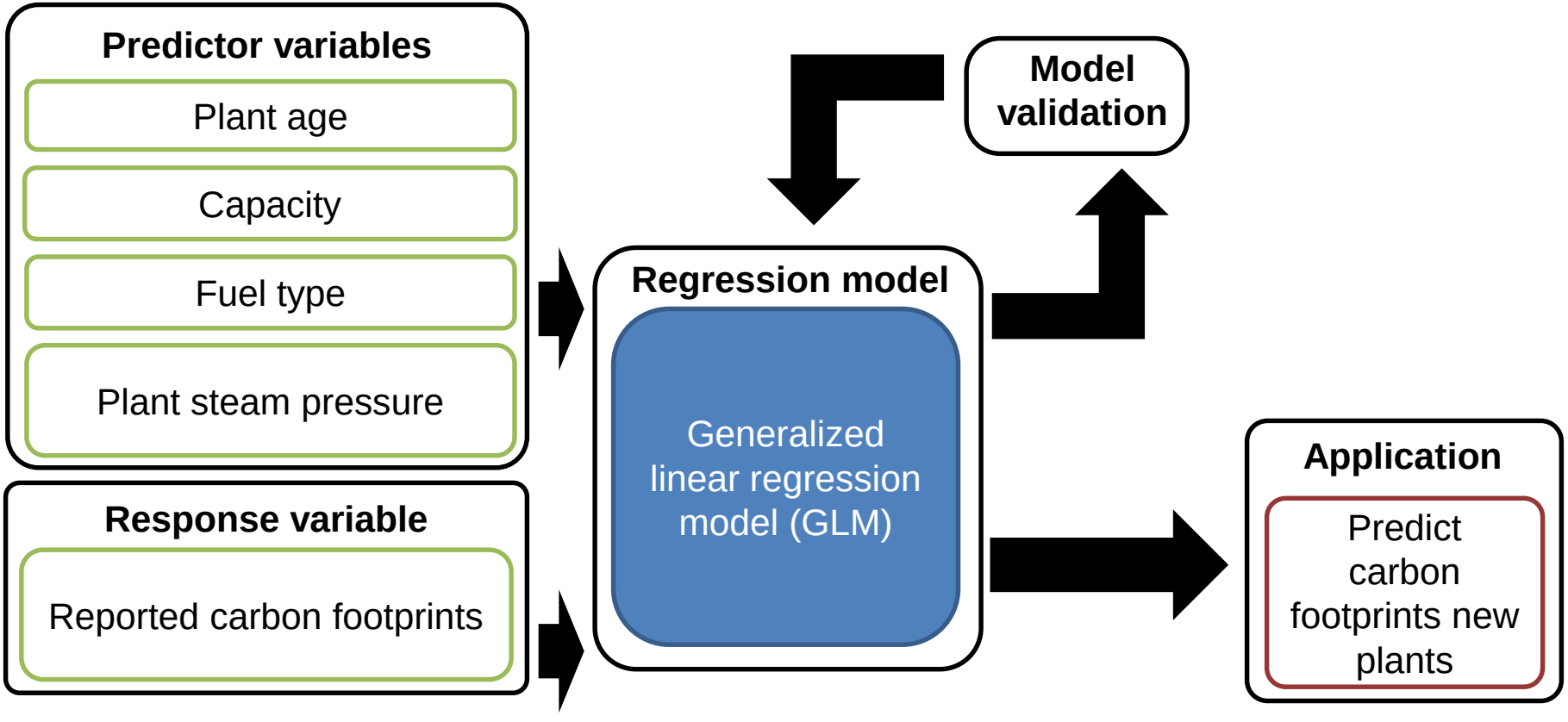
(Lack of) data availability; coal plants

- Emissions data available for 445 plants (495 GW) worldwide
- Sources: government databases, private utilities, operators

Distribution of available emissions data, by plant capacity
[as a fraction of total installed capacity in each country]



Solution → Model based on readily available data



Regression model

Generalized linear regression (GLM) model:

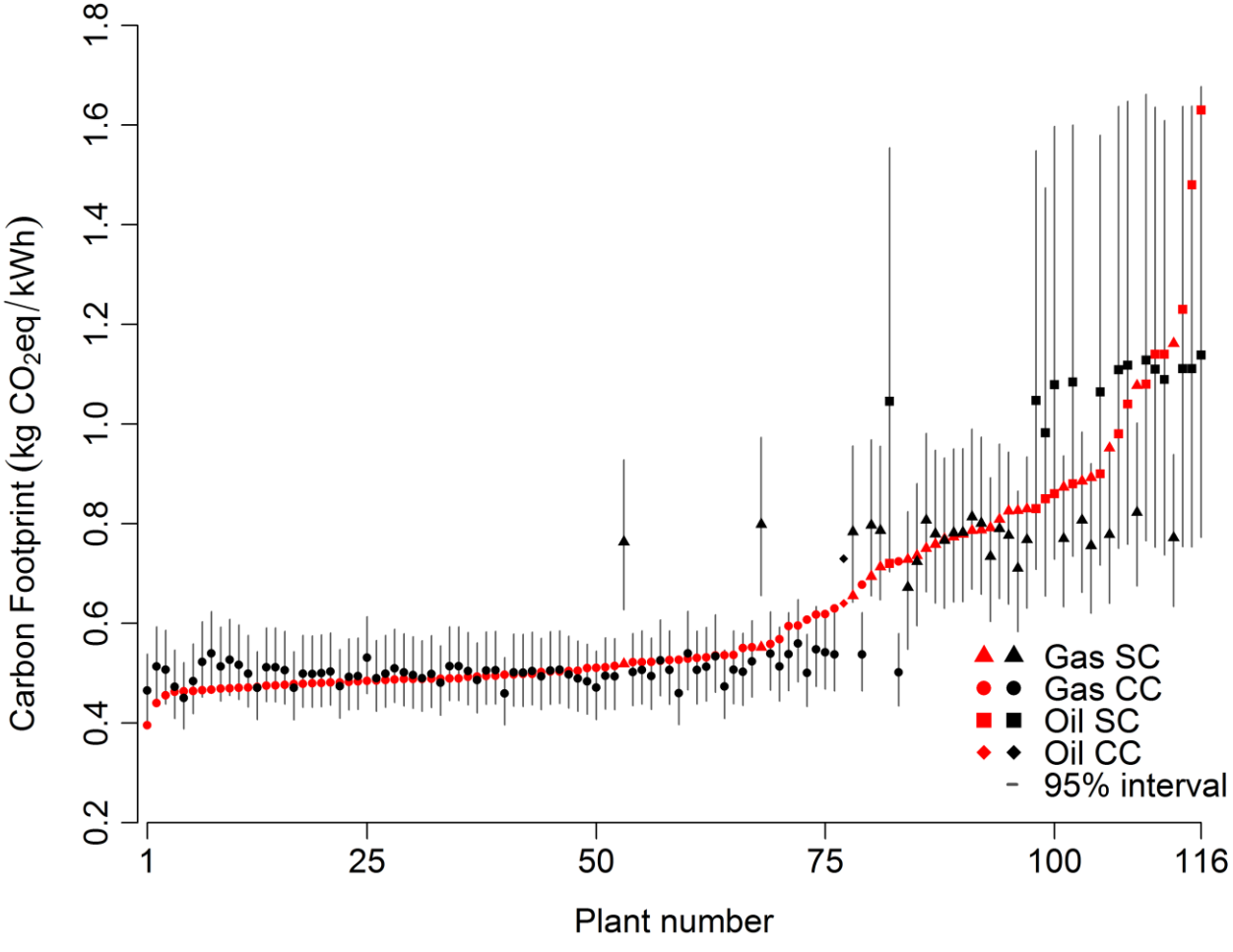
$$\log(CF) = \beta_0 + \beta_1 \cdot \text{Plant age} + \beta_2 \cdot \text{Plant capacity} + \beta_3 \cdot \text{Fuel type} + \ldots$$

Data on specific power plants:

- WEPP (World Electric Power Plants) database: almost all operational power plants >100 MW in the world
- Power plant design details such as: location, capacity, number of generators, year of construction, et cetera…

After fitting and model validation, the model was applied to 2538 gas oil fuelled power plants and 764 coal fuelled power plants

Plant specific predictions for oil and gas



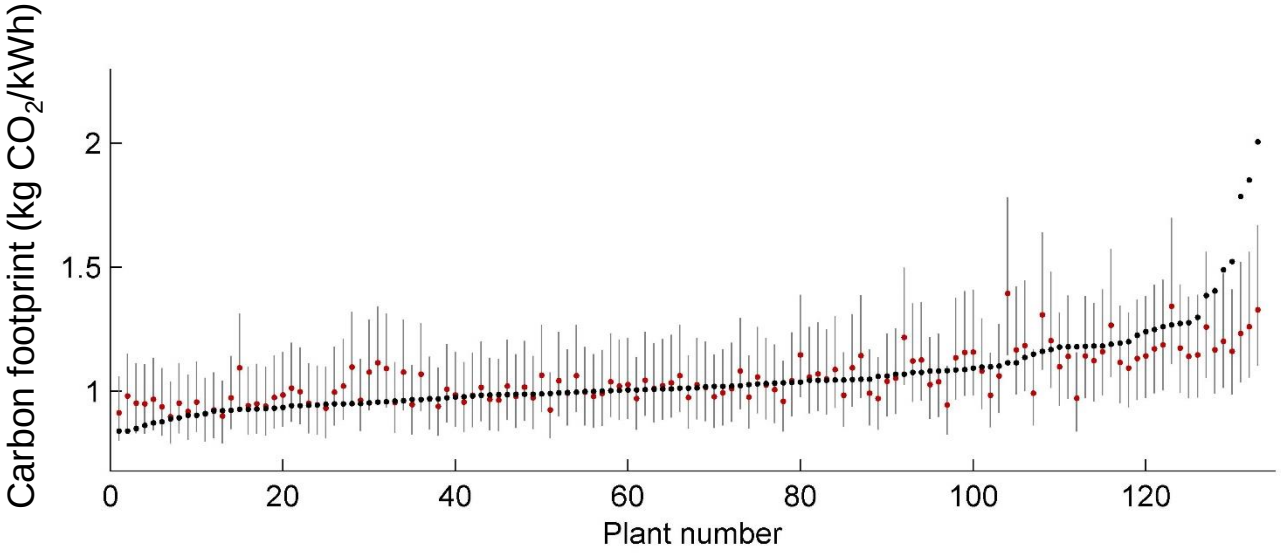
R^2 = 0.81 (validation set)

Type of fuel most important predictor by far

Smallest uncertainty in prediction for Combined Cycle Natural Gas

Largest uncertainties for Single Cycle oil plants

Plant specific predictions coal fuelled power plants

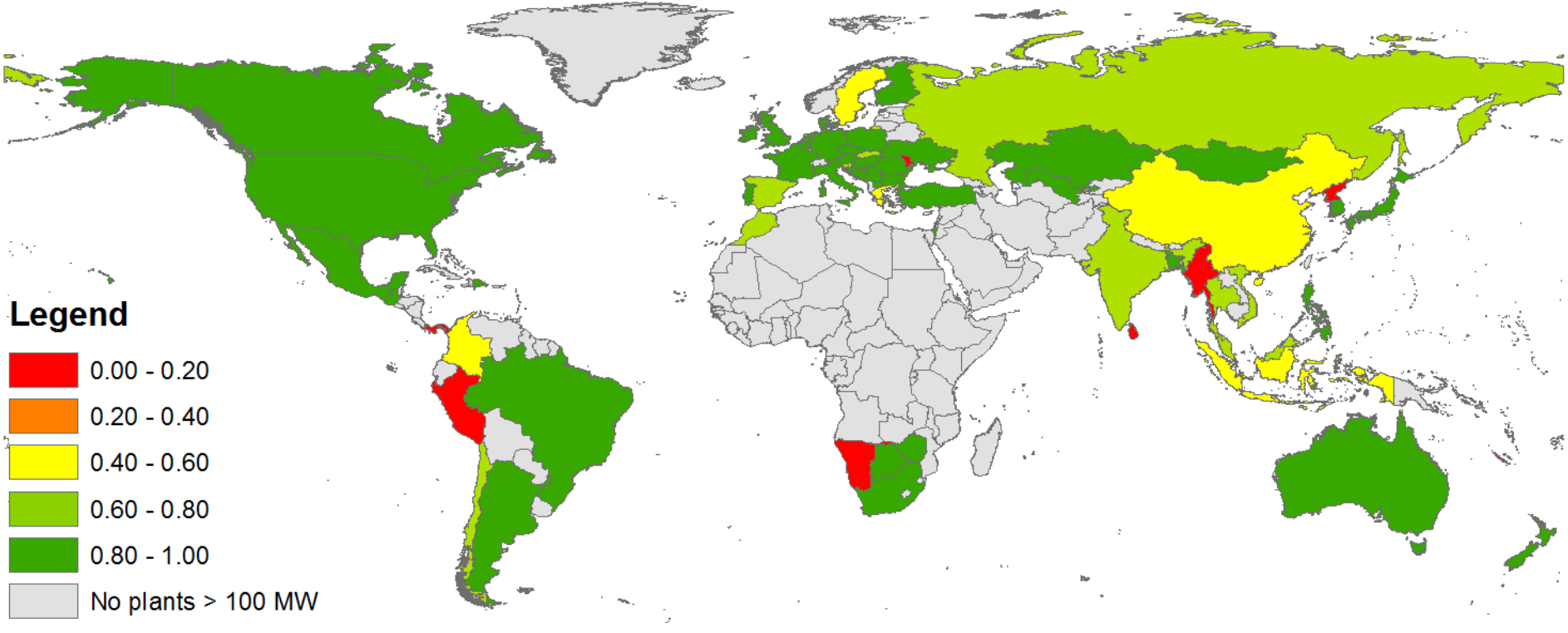


R^2 = 0.49 (validation set)

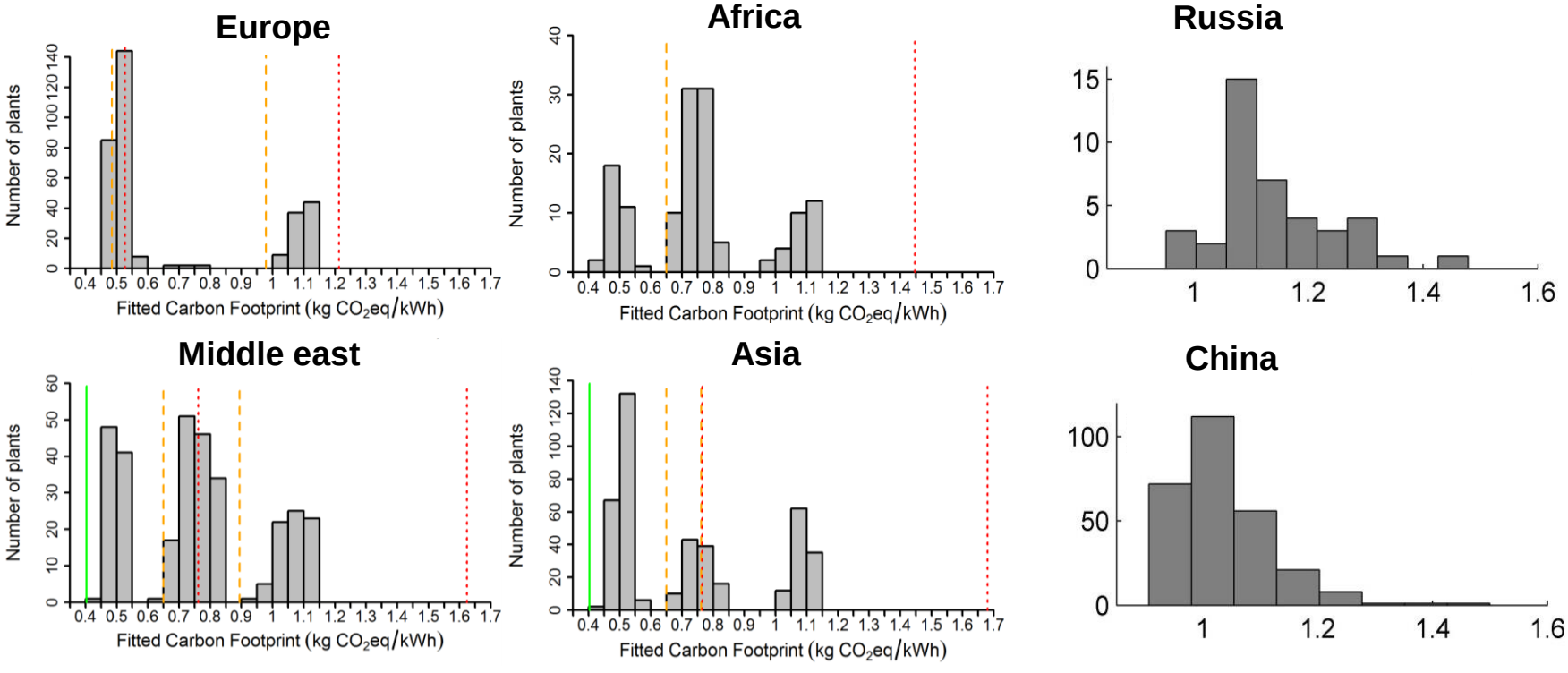
Steam pressure and lignite/non-lignite most important predictors

Model cannot reliably predict very high carbon footprints (>1.5 kg CO_2/kWh)

Direct and modelled data availability; coal plants



Application: world-wide distributions of carbon footprints



Carbon footprint (kg CO_2eq/kWh) gas and oil fuelled power plants

Carbon footprint (kg CO_2/kWh) coal fuelled power plants

Problem setting

LCA: environmental impact of a process/product

Requires considerable amount of data

Common problem: **lack of data in inventory and impact assessment**

Typically make assumptions to address the issue

Questions

- What can we do to fill these data gaps?

- How much data do we really need in impact assessment?

How much data do we really need?

Do we need data on 1500+ interventions and 100+ indicators for 15 different impact categories?

Impact assessment method	Number of indicators	Category
CML2001	49	Midpoint indicators
EDIP 2003	21	
Impact 2002	14	
ReCiPe	31	
TRACI	9	
EcoIndicator99	3	Endpoint indicators
Ecological scarcity2013	1	
EPS2000	1	
Impact 2002	3	
ReCiPe	3	
Fossil energy footprint	1	Resource footprints
Water footprint	1	
Material footprint	1	
Land footprint	1	

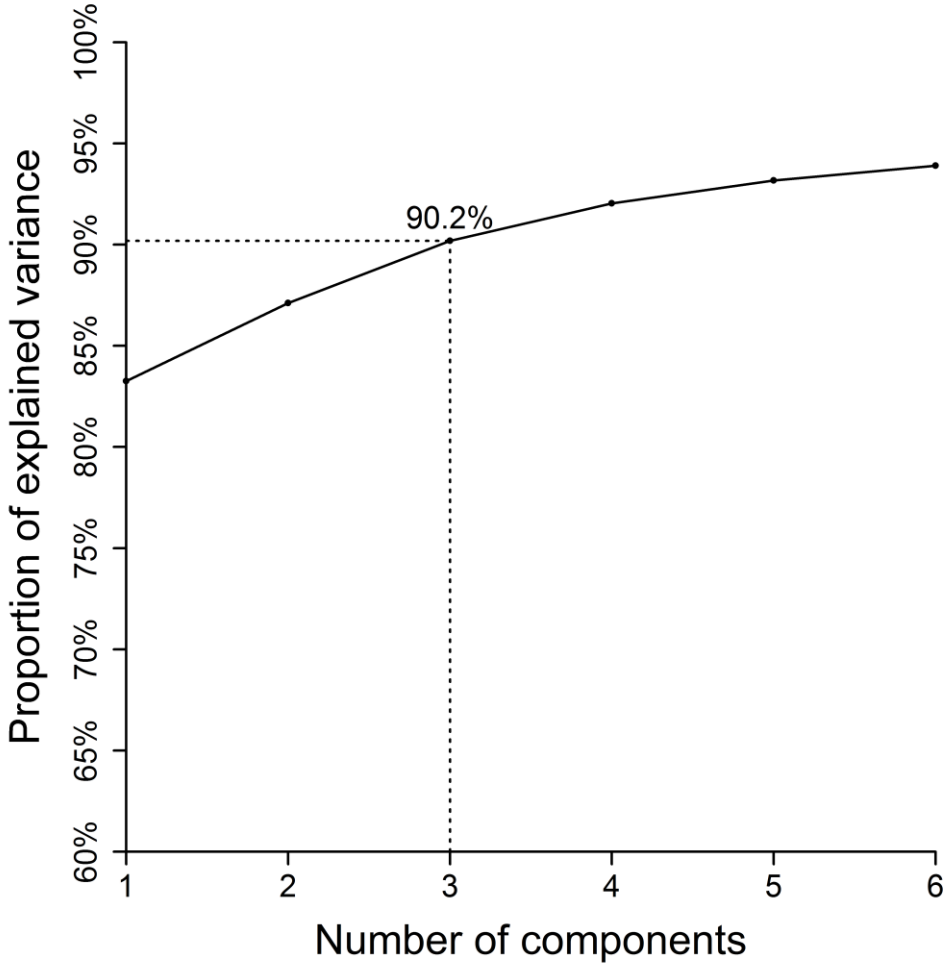
How much data do we really need?

PCA: Principal Component Analysis

- Based on the rank correlations between indicators
- Variance rewritten as a set of linear transformations of the original data
- The first component covers the maximum amount of variance, consecutive components cover less

First couple components cover the majority of the variance → potential for reduction

How much data do we really need?



To explain >90%:

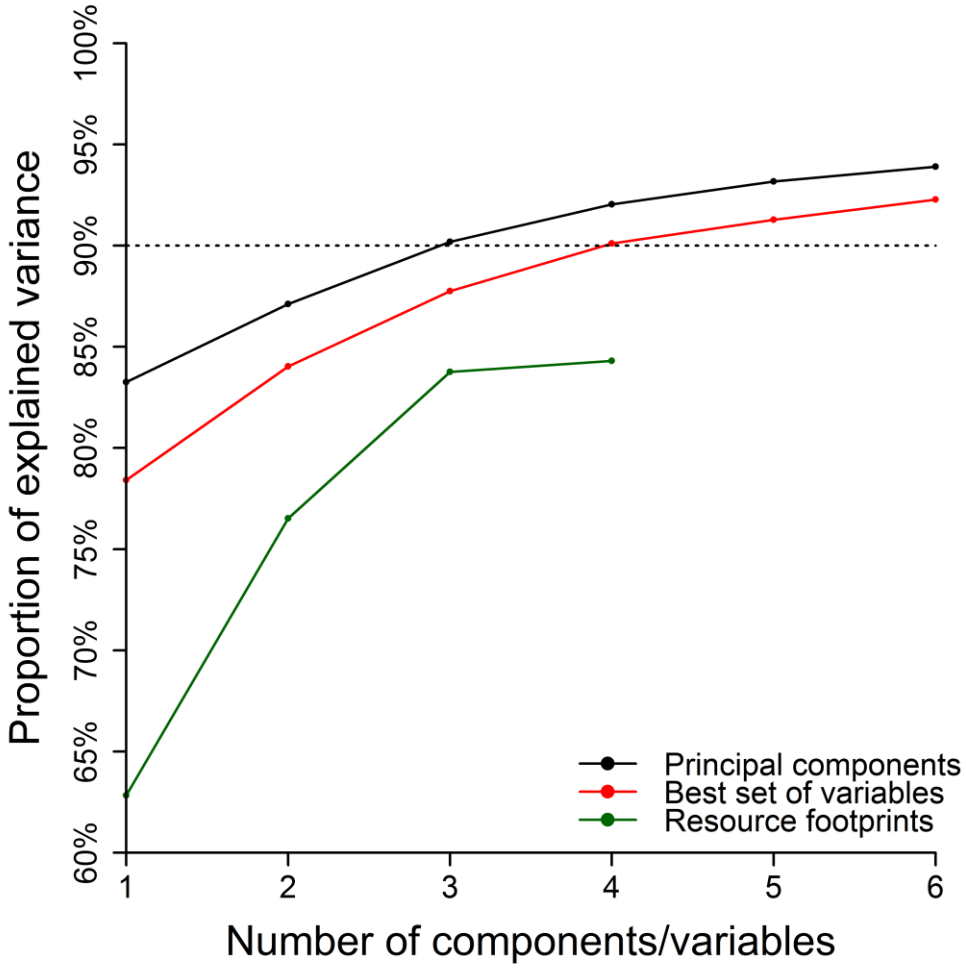
3 principal components

But: all original indicators are needed to calculate these....

So: How many original indicators?

→ use regression on PCA scores

How much data do we really need?



To explain >90%:
3 principal components

OR

Best 4 Indicators:

- Climate change
- Ozone depletion
- Land footprint
- Marine ecotoxicity

3 Resource footprints:
Energy, Land, Material
83.8%

Discussion

Regression approaches are a strong tool to fill data gaps

- Other application possibilities, e.g. farm specific or waste water treatment facility specific emissions
- Model uncertainty is introduced through the regression model

We can reduce the number of impacts, but:

- This is a purely statistical tool, however RACER (Relevant, Accepted, Credible, Easy and Robust) criteria are also important

Thank you for your attention

Zoran Steinmann

z.steinmann@science.ru.nl

References:

Steinmann, Zoran JN, et al. 2014. How to address data gaps in life cycle inventories: A case study on estimating CO2 emissions from coal-fired electricity plants on a global scale. *Environmental science & technology,* 48(9): 5282-5289.

Hauck, M. et al. 2017. Estimating the Greenhouse Gas Balance of Individual Gas-Fired and Oil-Fired Electricity Plants on a Global Scale. *Journal of industrial ecology, 21(1):127-135*

Steinmann, Zoran JN, et al. 2016. How many environmental impact indicators are needed in the evaluation of product life cycles? *Environmental science & technology,* 50(7): 3913-3919.