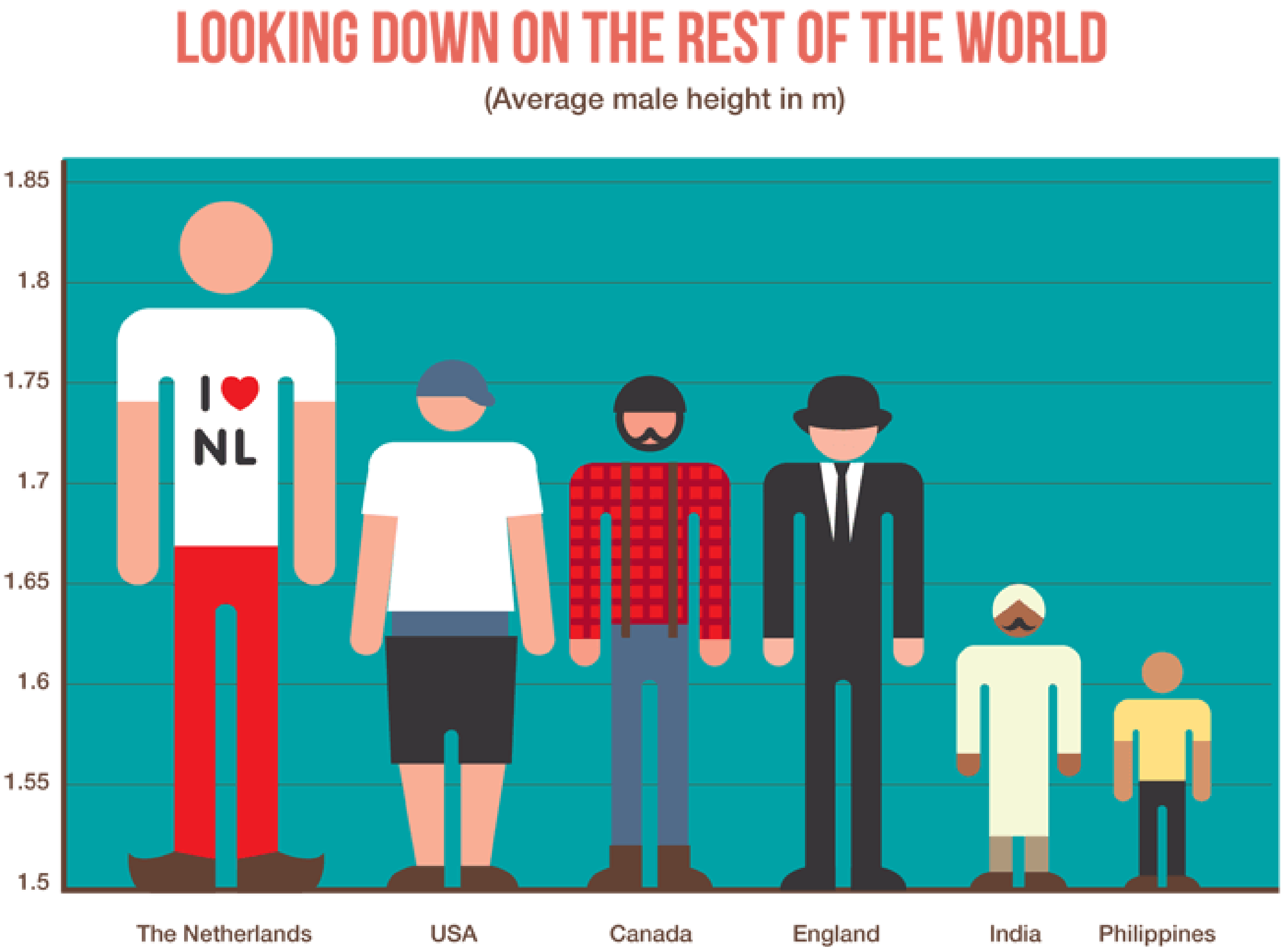
ISRIC
World Soil Information

Towards a sampling design for monitoring global soil organic carbon stocks

Gerard Heuvelink, Niels Batjes, Dick Brus, Jaap de Gruijter, Alex McBratney and Uta Stockmann

Wageningen Soil Conference, 28 August 2017

The Dutch are the tallest people in the world



Average height Dutch men: 182.5 cm

Required sample size in case of probability sampling

The Netherlands

- Population size 17,000,000
- Sample size needed to obtain standard error ≤ 0.5 cm: 200

China

- Population size 1,300,000,000
- Sample size needed to obtain standard error ≤ 0.5 cm: 200

In case of large populations, the standard error does not depend on population size

estimate of population mean

true population mean

population standard deviation

$$se = sd(\mu - \hat{\mu}) = \frac{\sigma}{\sqrt{n}}$$

sample size

Same holds for estimation of the global soil organic carbon stock

- Required sample size hardly depends on size of area: whether it is for a region, country or the entire Earth: sample size will be about the same
- We can compute it (in case of simple random sampling) from:

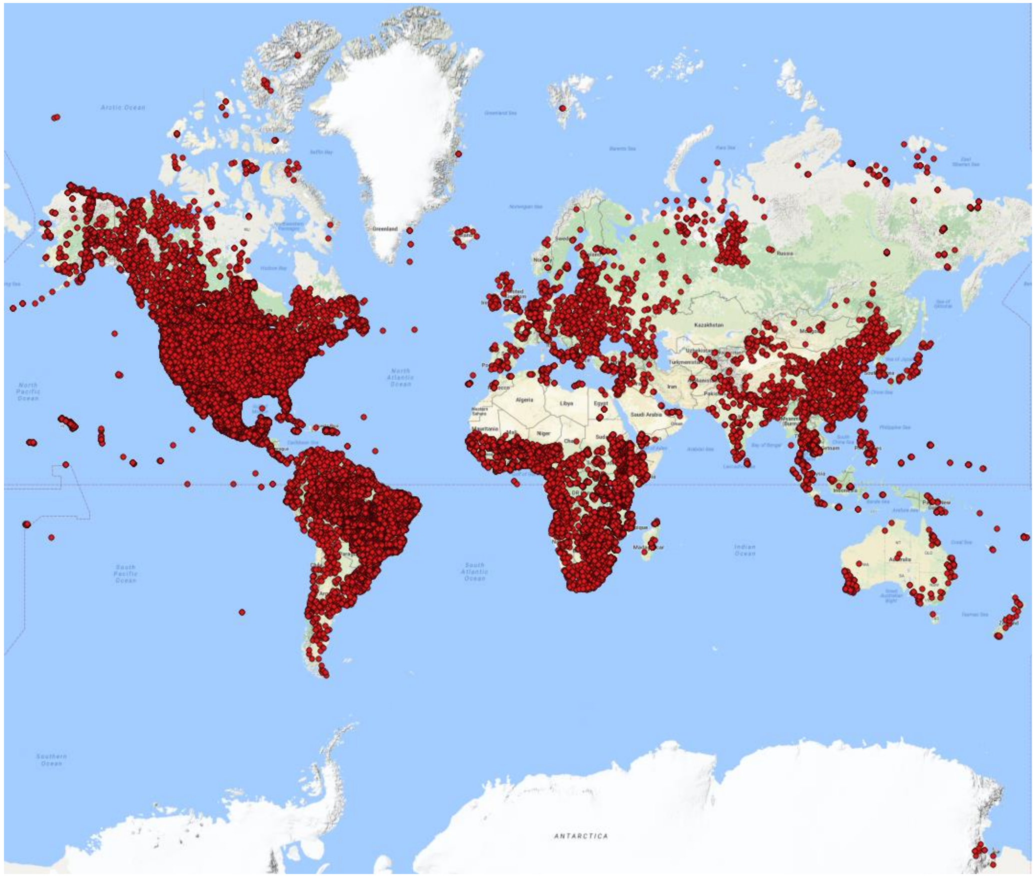
$$n = \left(\frac{\sigma}{se}\right)^2$$

- All that we need is the population standard deviation and the required standard error

Deriving the population standard deviation

- Target variable: soil organic carbon (SOC) stock 0-30 cm (ton/ha)
- ISRIC WoSIS soil database provides SOC stock observations for about 42,000 locations across the world (but note: no uniform spread)
- We used this dataset to compute the standard deviation, this gives:

$$\sigma = 5.5\ ton/ha$$



Deriving the standard error

- 4pour1000 initiative aims to compensate for annual fossil carbon CO_2 emission by sequestering 3.5 Gt C per year (4‰ of the total global stock)
- This means an increase of 17.5 Gt C every five years
- If we let the standard error be 10% of the intended increase over five years, then estimation accuracy will be sufficiently high to evaluate if the 4pour1000 objective is achieved
- We therefore used $se = 1.75\ Gt\ C$, this is about $0.12\ ton/ha$

A modest sample size will do the job

$$n = \left(\frac{\sigma}{se}\right)^2 = \left(\frac{5.5}{0.12}\right)^2 = 2100$$

- This is based on simple random sampling, while the sample size required can be much reduced using more elaborate sampling designs (see later)
- The 2100 refers to estimation of the stock at any one point in time. To estimate SOC stock change over time, it should be multiplied with $\sqrt{2}$, assuming independent sampling at the two points in time (but again, see later for a more efficient alternative)

But is all this really needed? After all, we have already so many SOC stock estimates

SOC stock (Gt C)		Source
0-1 m	0-2m	
1400	-	Post et al.(1982)
1460-1550	2380-2460	Batjes (1996)
1580	-	Kasting (1998)
1500	2460	Robert (2001)
1420	-	Hiederer and Köchy (2012)
1460	1920*	Wei et al. (2013; * to 2.3m)
1410 ± 150	2060 ± 220	Batjes (2016)

Carbon Managem

EODERMA

ELSEV

Globa

Uta Stockmann [a,*], José Padarian [a], Alex McBratney [a], Budiman Minasny [a]
Delphine de Brogn
Damien J. Field [a]

Differences much greater than 1.75 Gt C

What makes design-based estimation so attractive?

- It yields an accurate estimate for modest sample size
- It makes no assumptions, both the estimate and the associated accuracy measure are completely model-free
- Current model-based approaches seem not sufficiently accurate to assess change over time
- Efficiency of design-based estimation can be improved by using more elaborate sampling designs:
  - Stratified random sampling (for example using a soil type map for stratification), calculations that we did for France and New South Wales showed a 50% efficiency gain using the *Ospats* stratification method
  - Model-assisted sampling can benefit from models while still keeping the advantages of design-based sampling
  - For estimation of SOC stock change a static design (revisit locations) increases efficiency dramatically, but note that such approach is sensitive to manipulation

There are disavantages too

- Probability sampling is key, so problems may occur:
  - Parts of the world may be inaccessible because governments refuse access or there are too high risks (e.g. military conflict areas)
  - Parts of the world may be very poorly accessible because they are very remote or have harsh conditions (arctic, deserts, tropical rain forest)
- All observations should be collected and analysed in the same way using a standardised procedure
  - Many countries will want to use their own field protocol and laboratories, it will take considerable effort to get them to agree on a uniform, standardised approach
- Practical implementation will no doubt have imperfections, causing the outcomes to become vulnerable to criticism
- We only get estimates for the population as a whole, i.e. summary measures instead of maps

In summary

- Accurate estimation of the global soil organic carbon stock and changes therein over time using design-based statistics ('probability sampling') requires a remarkably small number of observations
- Estimates and associated accuracy measures are unbiased and completely model-free
- We believe this would make a valuable addition to the many modelling efforts that are already done, at relatively small cost
- Practical implementation will be a huge logistical challenge that requires commitment from many organisations
- The first challenge is to convince researchers (you!) and policy makers of the added value of such project: it can only work if there is broad support

Thank you