Uncertainty Analysis and Statistical Validation of Spatial Environmental Models

PE&RC Course 9-13 December 2024

Gerard Heuvelink and Sytze de Bruin

WAGENINGEN UNIVERSITY
WAGENINGEN UR

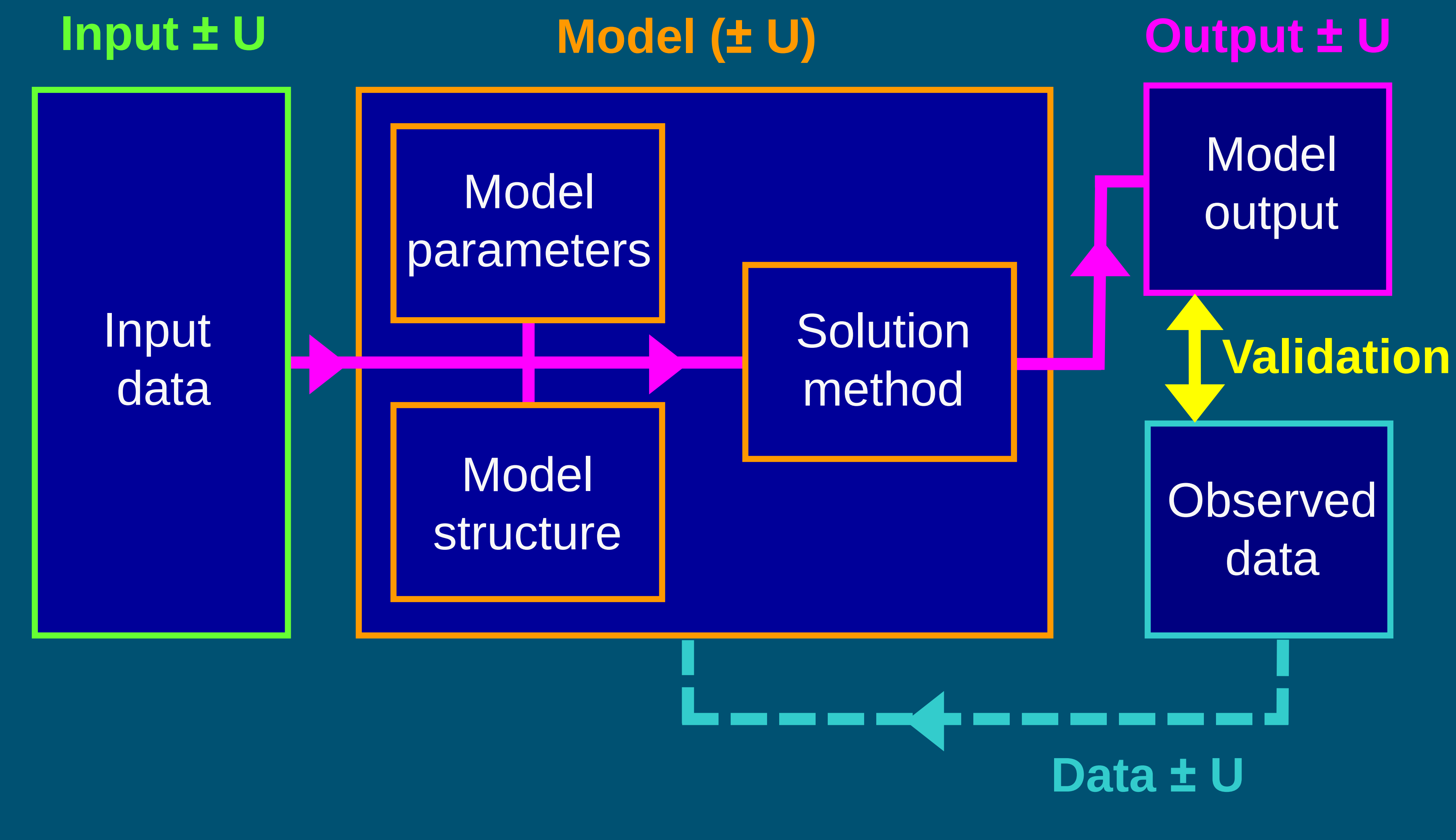
This week's programme

Post-graduate course

Uncertainty Analysis and Statistical Validation of Spatial Environmental Models (9-13 December 2024)

<table>
<tr><th>Day</th><th>9:00 – 12:00</th><th>12:00 – 13:30</th><th colspan="3">13:30 – 15:45</th><th>15:45 – 16:15</th><th>16:15 – 17:00</th></tr>
<tr><td>Mon. 9 Dec.</td><td>Course overview, lecture and exercises: probabilistic modelling of uncertainty, Taylor series method</td><td>Lunch break</td><td colspan="3">Computer practical Taylor series method: vegetation indices</td><td>Feedback computer practical</td><td>Continue practical, own data or advice</td></tr>
<tr><td>Tue. 10 Dec.</td><td>Lecture and exercises Monte Carlo method, including uncertainty source contributions and scenarios</td><td>Lunch break</td><td colspan="3">Computer practical Monte Carlo method: vegetation indices</td><td>Feedback computer practical</td><td>Continue practical, own data or advice</td></tr>
<tr><td>Wed. 11 Dec.</td><td>Lecture and exercises uncertainty in model parameters and model structure, including Bayesian calibration</td><td>Lunch break</td><td colspan="3">Computer practical Bayesian calibration and propagation of model parameter uncertainty</td><td>Feedback computer practical</td><td>Continue practical, own data or advice</td></tr>
<tr><td>Thu. 12 Dec.</td><td>Lecture and exercises statistical validation and cross-validation of spatial model outputs (maps), including sampling for validation, spatial cross-validation and reliability plots</td><td>Lunch break</td><td colspan="3">Computer practical statistical validation and cross-validation of spatial model outputs</td><td>Feedback computer practical</td><td>Continue practical, own data or advice</td></tr>
<tr><td>Fri. 13 Dec.</td><td>Lecture and computer practical uncertainty of spatial averages and totals</td><td>Lunch break</td><td>Finish computer practical and feedback</td><td>Course evaluation</td><td>Uncertainty game</td><td></td><td></td></tr>
</table>

Uncertainty propagation and model validation overview



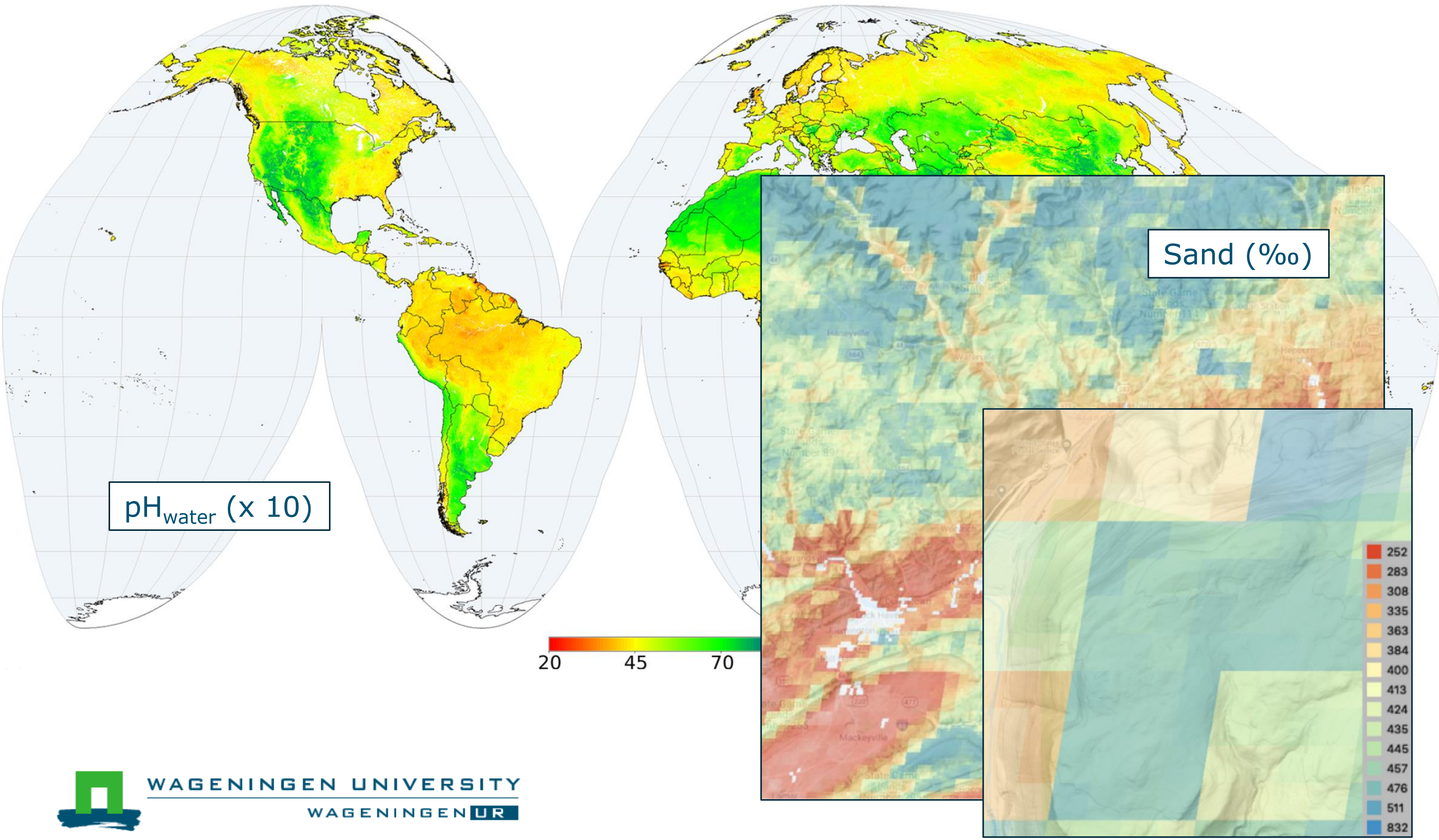
Change of support

- In the first three days of this course we analysed how uncertainty propagates through spatial models
- We obtained a **probability distribution** of the model output at each location in the area of interest
- But what is a location? Is that a point or a grid cell?
- And what if we want to know the uncertainty of **spatial averages or totals**?
- These are very relevant questions because many users are not interested in values at ‘points’ but in values for ‘blocks’ (e.g. CO_2 emission, rainfall, crop yield, aboveground biomass)
- So we need to account for a **change of support**. We will see that this can have a major effect on the uncertainty

The scale triplet

- Many studies refer to 'scale' but it is not always clear what this means
- In cartography, scale is the ratio of the distance on a (paper) map and the distance in reality (e.g. scale 1:10,000), but that no longer works in a digital world
- Nowadays scale is a triplet of three components:
    1. **Extent** (size of study area)
    2. **Resolution** (size of grid cells of a raster map)
    3. **Support** (area or volume over which observations or predictions are made)
- Of these three, support is perhaps the most important, but it is often ignored
- Many confuse resolution and support, but these are not the same!

SoilGrids predicts at 250 m x 250 m resolution, but what is the support?



WAGENINGEN UNIVERSITY
WAGENINGEN UR

Support is particularly important when we quantify uncertainty

Uncertainty averages out, so we expect smaller uncertainty at larger supports

$$uncertainty\ of\ spatial\ average \neq average\ of\ spatial\ uncertainty$$

We propagated the uncertainty, but what was the support of the following model outputs?

- Spectral vegetation index Lorraine study area
- Topsoil CN ratio
- Rainfall-Runoff model (spatial and/or temporal support?)
- USLE erosion

Without a change of support, the output usually has the same support as that of the inputs, and so also the propagated uncertainty refers to that support

This also means that inputs must be brought to a common support before running a model

From Leopold et al. (2006), see also literature folder

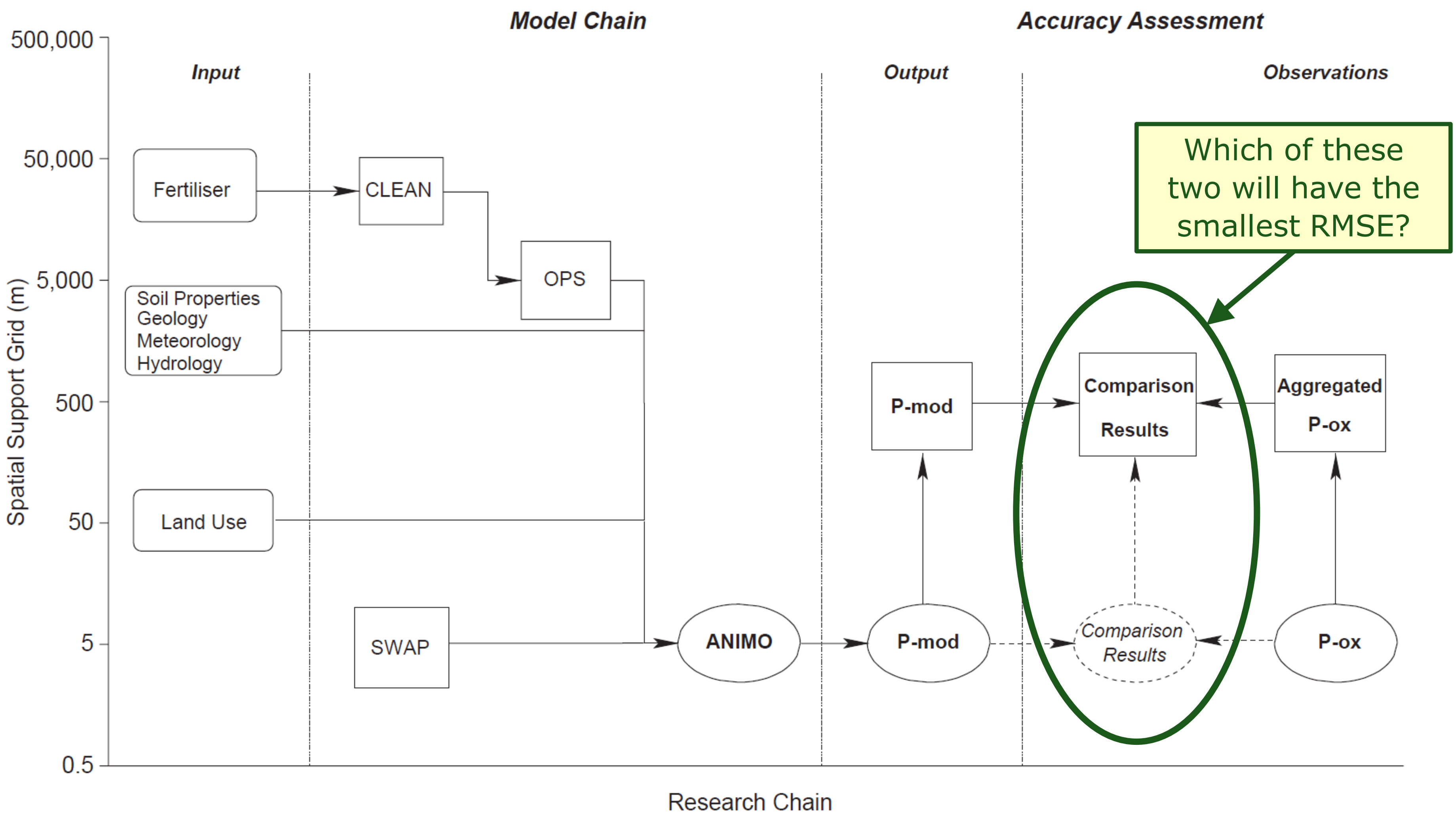


Fig. 1. Building blocks of STONE operate at different spatial supports. Comparison of STONE output (P-mod) and observations (P-ox) can be done on the block support (route with solid arrows) and point support (route with dashed arrows).

Averaging reduces uncertainty, but how much depends on the degree of (spatial) correlation

Check this out yourself with today's Exercise

Exercise

- How many dice must we roll to be 95% certain that their mean is between 3 and 4? Between 3.4 and 3.6? Between 3.49 and 3.51?

- Ozone (air quality) is measured every 40 m along a 1 km transect (25 measurements in total). Due to spatial variability each single measurement is quite uncertain: assume a normal distribution with mean of 70 ppb, standard deviation 30 ppb
- Suppose human health at stake if Ozone > 100 ppb. What is the probability that this occurs at a given point along the transect?
- What is the probability that the average Ozone along the whole transect (i.e., the average of 25 measurements) is above the threshold? Compute this for various degrees of correlation between the 25 concentrations (assuming all pairs have the same correlation).
- Use R script "Exercise Friday.R" on TEAMS
- What should we get if the correlation is 1?

To combine uncertainty propagation with change of support, we must account for spatial correlation of input errors

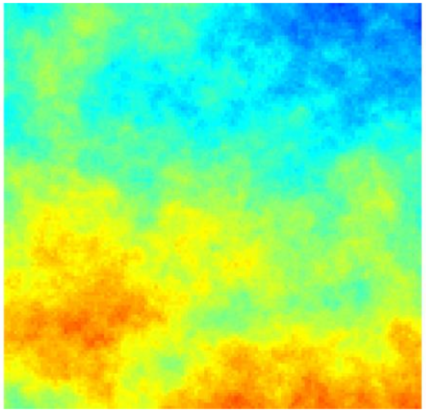
1. DEFINITION of a (statistical) uncertainty model for spatial objects and attributes

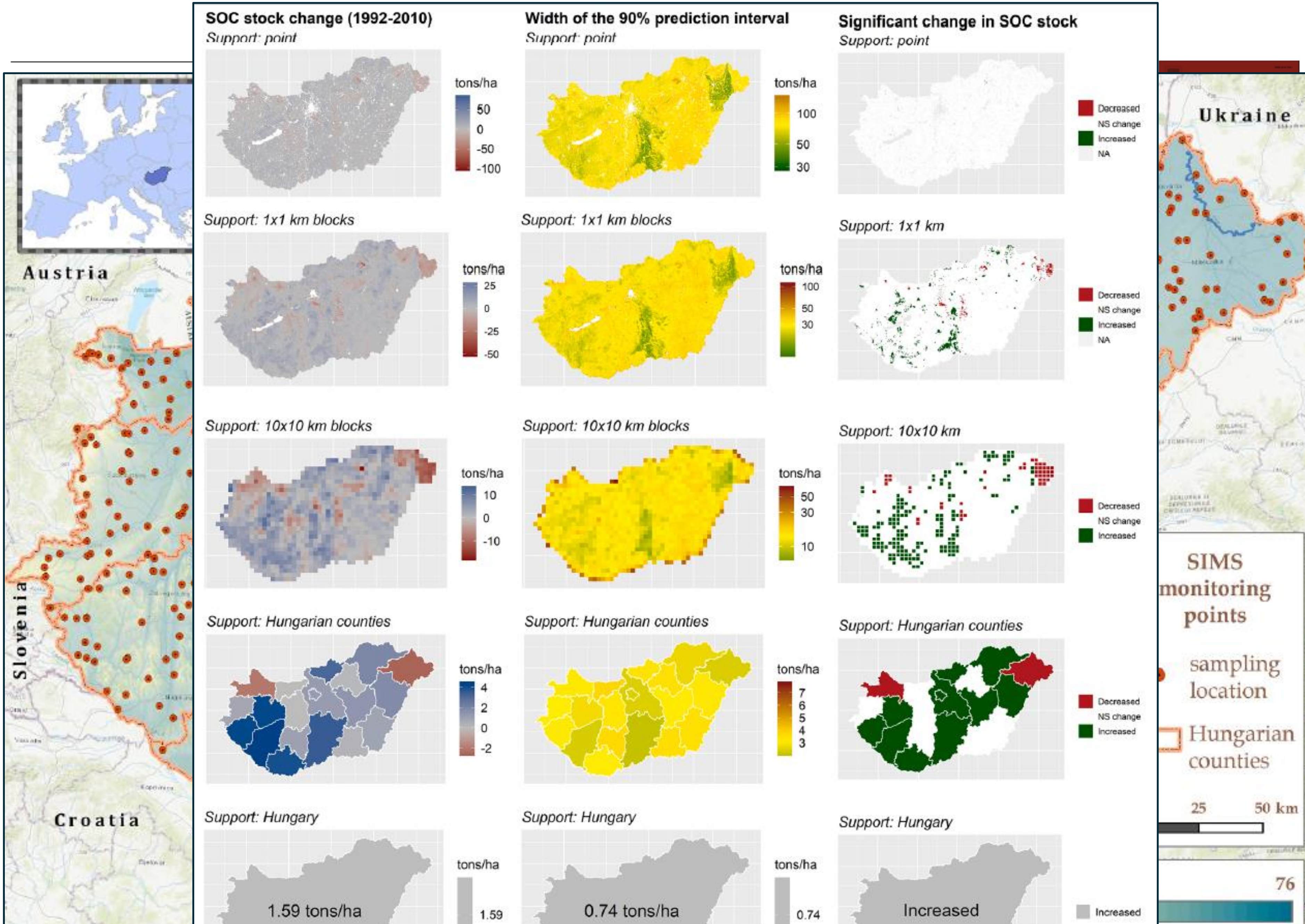
2. IDENTIFICATION of the uncertainty model (estimate its parameters)

3. Perform the actual UNCERTAINTY PROPAGATION ANALYSIS

➔ Steps 1 and 2 must account for spatial correlation

➔ This can be done with geostatistics (next year's course ☺)





patial aggregation
ant for carbon ac-
hat observed SOC

Propagation of ‘point’ support uncertainty to ‘block’ support model output is not that difficult:

1. Quantify **point support input uncertainty** with a probability distribution as done before
2. Generate **possible realities** of these inputs by random sampling from the distribution
3. **Propagate the uncertainty** to the point support model output using the Monte Carlo method
4. **Spatially aggregate** the model output for each individual realisation
5. Compute **summary statistics** of the aggregated model outputs, use quantiles or the standard deviation to quantify uncertainty

Today’s computer practical

- Study area in Cameroon
- QUEFTS model simulates maize yield from multiple inputs, among others soil chemical properties
- Four soil chemical properties are uncertain, maps of possible realities at point support have already been generated
- Propagate soil property uncertainty to maize yield at point support, next average to districts and to the entire study area
- How much does uncertainty decrease with increase of spatial support?