Pan-European Estimation of animal stocking densities at 1x1 km grid

Wolfgang Britz

Rural, Water & Ecosystem Resource Unit
IES, JRC/ISPRA

Context of the study

- Develop and validate downscaling algorithms for agricultural drivers of environmental impacts from administrative regions:
  - To allow for spatial analysis of nutrient balances, greenhouse emissions, bio-diversity, landscape ..
  - To allow for ex-ante spatial CAP impact analysis based on results of agricultural sector models
- Specifically here:
  - Estimate animal stocking densities,
  - In order to estimate manure output and economic performance indicators at a high spatial resolution

Overview on algorithm

- Three step procedure:
  1. Estimation of functional relationship between animal stocking densities and land use shares, yields and economic indicators for crops, based on statistical data for administrative regions
  2. Forecasting at 1x1 km grid, based on neighbouring cells and regions
  3. Correction of forecasts to ensure consistency with results at administrative regions, taking quality of forecast into account

Data for the Estimation Step

- Data source is the Farm Structure Survey 1999 (agricultural census in December, reporting land use and herds)
- Available from Eurostat at Pan-European level for NUTS II/III regions (ca. 500 regions for EU25)
- Land use and herd sizes corrected to match CAPRI data base (3-year average, herd sizes based on slaughtering statistics)

Variables in the Estimation

- Stocking densities estimated:
  - Animal activities in CAPRI (dairy cows, suckler cows, bulls, heifers for fattening, heifers for raising, calves for fattening, calves for raising, sows, pigs for fattening, sheep & goat for milk, sheep & goat for fattening, laying hens, poultry for fattening)
  - Aggregates of all animals, cattle & sheep, pigs & poultry, using livestock units as weights
- Explanatory variables:
  - Land use shares of agriculture, forestry; shares on agricultural land of cereals, Grandes Cultures, fallow land, fodder maize, grass land, all fodder crops, permanent crops
  - Yields of cereals, fodder maize
  - Total revenues and gross value added at producer prices per ha for cereals and Grandes Cultures
- All explanatory variables are possibly included in linear and quadratic form

Estimation procedure

- Member States are estimated individually when enough FSS regions were available, otherwise grouped
- Two step procedure:

1. Backwards elimination of insignificant variables based on OLS:
   - Remove step-wise least insignificant variables, as long:
     - The adjusted R squared is still increasing
     - The number of regressors exceeds 1/3 of the observations
     - There are variables with Prob=0 > 5%
2. Final non-linear tobit estimation, using the remaining variables

OLS/tobit regression function

Error term

Observed stocking density

Explanatory variables

$$e_{a,r} = Y_{a,r} - B_{a,const,r\in RM} - B_{a,r\in MS} - \sum_{i} B_{a,i,r\in RM} X_{i,r}^{\exp_{a,i}}$$

Constant term

Dummy for Member State

Regression coefficients

Exponential transformation
0.75 < exp < 1.25

a	animal activities
i	regressors
r	FSS regions
MS	Member states
RM	Block of member states

The tobit Estimator

- tobit estimator:
  - Respects the truncated nature of the variables to estimate, which are by definition non-negative
  - Thus prevents estimation bias resulting from the truncated nature of the variable
  - Maximum Likelihood Estimator, here assumption of normally distributed error term is maintained as in OLS

Tobit Maximum Likelihood estimator

Log of Likehood

squared Errors

Estimated variance of error terms

$$\max_{\hat{\sigma},B,e,\exp} \log(L) = \sum_{a,r \neg Y_{a,r}>0} -\frac{1}{2} e_{a,r}{}^{2} / \hat{\sigma}^{2} - \log(\hat{\sigma})$$

$$+ \sum_{a,r \neg Y_{a,r}=0} \log\left[1 - errorf\left(e_{a,r} / \hat{\sigma}\right)\right]$$

Log Prob. Density of error terms for non-zero obs.

Log Prob. Density of error terms for zero obs.

Estimation results (R2)

	Germany, Belgium & Netherlands	Greece	Spain and Portugal	France	Italy	UK and Irland	Norway, Finland, Denmark, Austria	New Member States, Bulgaria & Romania	Mean	Minimum
Cattle and sheep	0.98	0.92	0.92	0.92	0.92	0.97	0.73	0.99	0.92	0.73
Livestock units	0.97	0.91	0.82	0.91	0.89	0.95	0.72	0.98	0.89	0.72
Female calf for raising	0.95	0.93	0.91	0.81	0.92	0.94	0.69	0.99	0.89	0.69
Female calf for fattening	0.84	0.80	0.92	0.85	0.90	1.00	0.89	0.94	0.89	0.80
Dairy cows, high yield	0.95	0.89	0.91	0.84	0.88	0.97	0.66	0.99	0.89	0.66
Heifers raising	0.99	0.59	0.86	0.94	0.92	0.97	0.75	0.99	0.88	0.59
Male calf for fattening	0.94	0.87	0.93	0.85	0.51	0.92	0.95	0.93	0.86	0.51
Dairy cows, low yield	0.95	0.60	0.90	0.84	0.81	0.97	0.73	0.98	0.85	0.60
Sheep & goat for milk	0.91	0.82	0.94	0.51	0.78	0.87	0.94	0.99	0.85	0.51
Bulls, high final weight	0.94	0.87	0.79	0.82	0.72	0.94	0.67	0.96	0.84	0.67
Pigs and poultry	0.95	0.71	0.57	0.78	0.73	0.91	0.96	0.97	0.82	0.57
Other sheep and goat	0.94	0.68	0.93	0.46	0.69	0.87	0.95	0.99	0.81	0.46
Male calf for raising	0.89	0.89	0.85	0.75	0.54	0.96	0.55	0.99	0.80	0.54
Suckler cows	0.91	0.84	0.85	0.68	0.39	0.85	0.86	0.98	0.80	0.39
Heifers for fattening, low final weight	0.92	0.83	0.48	0.78	0.90	0.97	0.72	0.75	0.79	0.48
Heifers for fattening, high final weight	0.92	0.83	0.48	0.78	0.90	0.97	0.72	0.75	0.79	0.48
Sows	0.96	0.38	0.62	0.75	0.71	0.91	0.96	0.98	0.78	0.38
Bulls, low final weight	0.94	0.64	0.69	0.82	0.59	0.86	0.66	0.96	0.77	0.59
Pig for fattening	0.95	0.35	0.59	0.80	0.72	0.84	0.95	0.96	0.77	0.35
Laying hens	0.95	0.77	0.83	0.63	0.27	0.77	0.74	0.99	0.74	0.27
Poultry for fattening	0.96	0.54	0.53	0.70	0.20	0.87	0.61	0.99	0.68	0.20
Mean	0.94	0.75	0.78	0.77	0.71	0.92	0.78	0.95	0.82	
Minimum	0.84	0.35	0.48	0.46	0.20	0.77	0.55	0.75		

Animal Stocking Densities at 1x1 km grid

Results: Discussion

- For the majority of the cases stable results:
  - 55% of the regressions have a R2 > 0.85
  - 80% have a R2 > 0.70
- Some curious out layers in Greece (pigs), Italy (poultry, suckler cows!) and France (sheep & goats)
- Remember: all regression coefficients are significantly different from zero at last at the 95% level

Forecasting at 1x1 km grid

- Problem: Regression based on larger Farm Structure Survey (FSS) regions, but application for HSMUs
(samples of 1x1 km grid cells with identical parameters)
  - Spread of explanatory variables in HSMUs far greater then for FSS regions
  - Resulting in rather high forecast errors for certain HSMUs
- Remedy:
  - Forecast for single HSMU takes forecasts/data for surrounding HSMUs and FSS regions into account

Forecast and its standard error for single HSMUs

Forecast

Parameters from FSS estimation

Explanatory Variables for the HSMU

$$Y^{**}_{a,HSMU} = B_{a,const,HSMU'\in RM} - B_{a,HSMU'\in MS} - \sum_i B_{a,i,HSMU'\in RM} X_{i,HSMU'}^{\exp_{a,i}}$$

Forecast error

Estimated standard deviation of error term

“Distance” of X for the SHSMU from mean of sample

$$STDERR_{a,HSMU} = \hat{\sigma}\sqrt{1 + X_{HSMU}'(X_{HSMU\in RM}'X_{HSMU\in RM})^{-1}X_{HSMU}}$$

$$\hat{\sigma}^2 = \frac{SSE_{HSMU\in RM}}{n_{HSMU\in RM} - (k_{HSMU\in RM} + 1)}$$

HSMU forecasts based on surrounding regional units

Distance and size based weights for neighbouring HSMUs and regions

$$w_{HSMU,HSMU'} = \frac{\sqrt{size_{HSMU'}\left[km^2\right]}}{dist_{HSMU,HSMU'}\left[km\right]} \wedge dist_{HSMU,HSMU'} < 50km$$

$$w_{HSMU,FSS} = \frac{\sqrt{size_{FSS'}\left[km^2\right]}}{dist_{HSMU,FSS'}\left[km\right]} \wedge dist_{HSMU,FSS'} < 100km$$

Sum of weights

$$sumw_{HSMU=} \sum_{HSMU'} w_{HSMU,HSMU'} + \sum_{FSS} w_{HSMU,FSS}$$

Final forecast

Surrounding HSMU forecasts

Surrounding FSS observations

$$Y^{*}_{a,HSMU} = \frac{1}{sumw_{HSMU}}\left[\sum_{HSMU'} w_{HSMU,HSMU'} Y^{**}_{a,HSMU} + \sum_{FSS} w_{HSMU,FSS} Y_{a,FSS}\right]$$

Final Std.Err.

$$STDERR^{*}_{a,HSMU} = \frac{1}{sumw}\sqrt{\sum_{HSMU'}\left(w_{HSMU,HSMU'} STDERR_{a,HSMU'}\right)^2 + \sum_{FSS}\left(w_{HSMU,FSS}\hat{\sigma}\right)^2}$$

Animal Stocking Densities at 1x1 km grid

Consistency step

- Application of Highest Posterior Density estimator:
  - Maximize joint probability density of forecasts N(Y*,STDDEV*) for different animal types and LU categories for the HSMUs
  - So that
    - Herd sizes at FSS level are exhausted
    - LU categories add up from individual herds
    - Security bounds are not exceeded per HSMU
- Motivation:
  - Ensure mutual compatibility of downscaled results and statistical data
  - Exploit relations between data (adding up of livestock units from different animal categories)
  - Allow for plausible bounds on downscaled results

HDP estimator

Log of joint prob. Density
assuming normally distributed uncorrelated error terms

Deviation from forecasts weighted with std. Error

Weights for deviation: UAA size and LU units

$$\max_{Y} \log(L) = \frac{1}{\sum_{HSMU,a} UAAR_{HSMU} LU_a} \sum_{HSMU} \left(-\frac{1}{2} \left[\frac{Y_{a,HSMU} - Y^{*}_{a,HSMU}}{STDERR^{*}_{a,HSMU}} \right]^2 UAAR_{HSMU} LU_a \right)$$

Exhaustion of regional herds

$$s.t. \quad Y_{FSS,a} = \sum_{HSMU \subset FSS} Y_{HSMU,a} ha_{HSMU}$$

Adding up for animal categories

$$Y_{LU,HSMU} = \sum_{a} Y_{HSMU,a} LU_a$$

$$s.t. \quad Y_{CS,HSMU} = \sum_{a \subset CS} Y_{HSMU,a} LU_a$$

$$Y_{PP,HSMU} = \sum_{a \subset PP} Y_{HSMU,a} LU_a$$

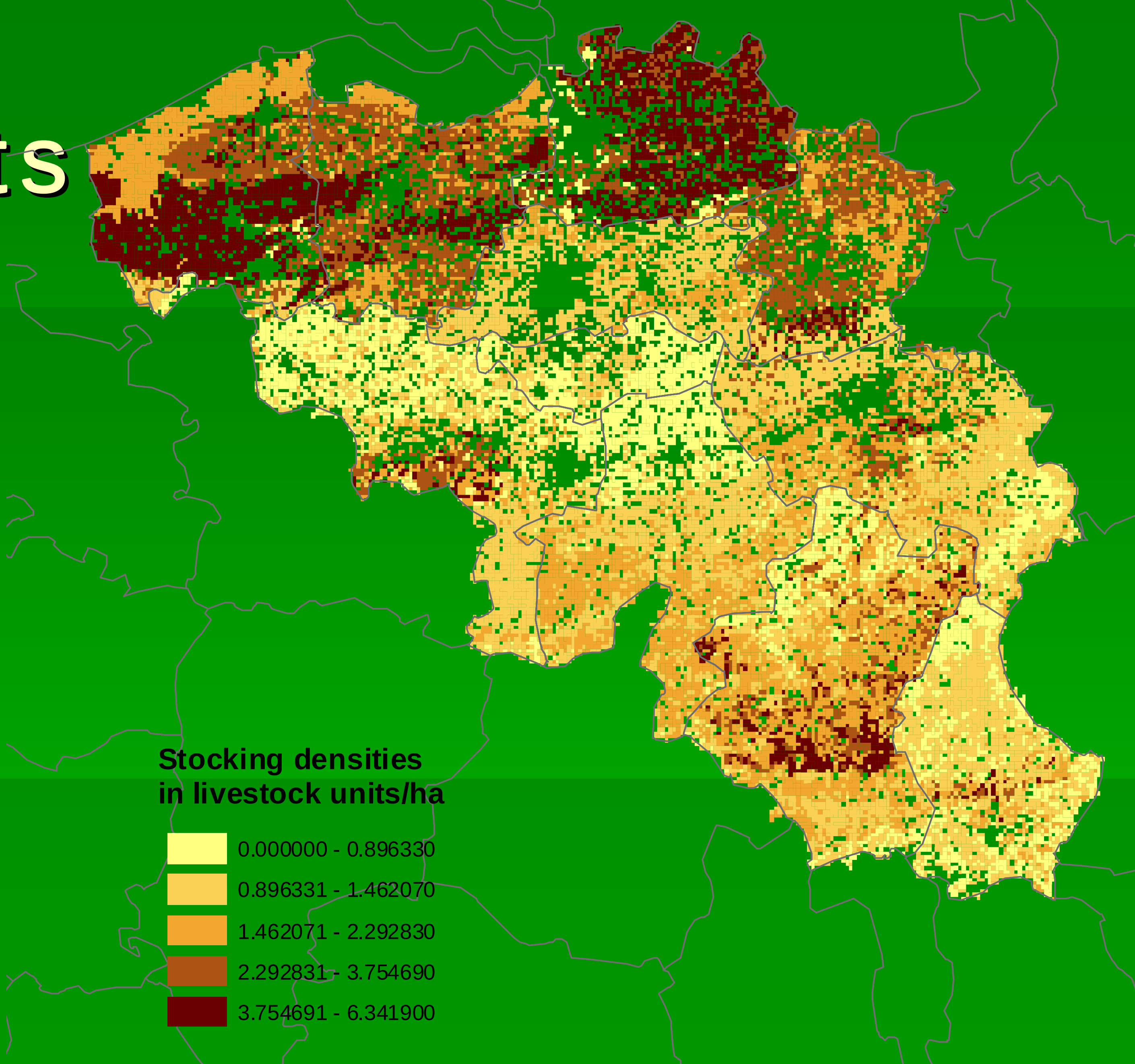
Plausibility bounds

$$s.t. \quad lowb_{FSS,a} \le Y_{HSMU,a} \le uppb_{FSS,a}$$

Results

Stocking densities in Belgium, and regional borders of the FSS regions

Green cells do not comprise agriculture



Animal Stocking Densities at 1x1 km grid

Result analysis

- Generally, the algorithm leads to a modest land use/soil type/elevation/slope – the main drivers for differences in cropping shares and yields inside administrative regions - based variance of stocking densities inside the FSS regions
- Despite the weighting with neighbouring forecasts and data for surrounding regions, in some cases FSS delineation remains visible:
  - High or low observed stocking densities for certain FSS regions are not well presented by the estimates for the HSMUs
  - Possible reasons: omitted variables (as e.g. distance to harbour, location of slaughter houses)
  - However: FSS regions may reflect natural boundaries as well …

Summary I

- Stocking densities are an important driver for environmental impacts of agriculture and are thus needed for spatial analysis
- Given current data limitation, a statistically based down-scaling algorithm for animal stocking densities from NUTS II/II to samples of 1x1 km grid cells was implemented
- Estimation revealed generally reasonable stable relation between crop production characteristics and stocking densities

Summary II

- Application of regression parameters to 1x1 km grid cells seems dangerous:
  - Explicit calculation of standard error of forecast for each HSMU to estimate, reflected in consistency step
  - Final forecast for each HSMU is distance and size weighted combination of surroundings HSMUs/regions
  - Final estimation step ensures mutual consistency with given statistical data for adminstrative regions

Outlook

- Along with cropping shares and yields, stocking densities are used to estimate organic and inorganic fertilizer application rates per HSMUs
- Fertilizer application rate estimation will take results from response surface estimation of bio-physical nutrient fate models into account
- Evaluation of alternative sources for the stocking density estimation (FADN, NUTS V statistics)