Hello Serena!

I have prepared some forest data for you. Try and see if it works, if it doesn’t, then we will at least have learned something about the data.

We decided to scrap the distinction between deciduous and coniferous trees and go just by total volume, age and coverage.

The forest in Sweden has been measured during three years: 2000, 2005, and 2010. Since your data varies in time, and the forest does too (it grows, gets cut, etc), I have tried to match each fishing occasion to the corresponding forest data that is likely to be the best fit. Most of your data is post 1995, which is perfect. I assigned the data as below:

* All electrofishing occasions before and including 2000 will be assigned 2000-data.
* All electrofishing occasions from 2001 to 2005 will be assigned 2005-data.
* All electrofishing occasions 2006 and onwards will be assigned 2010-data.

Forest data was assigned to a specific site by creating a circular buffer of 700m radius around the point, corresponding to an area of ca 154 hectares. The reason for choosing an area as huge as 150 hectares is to decrease the error. The document supplied with the data shows a figure with the error magnitude in relation to the studied area. Errors seem pretty high below 50-100 hectares and flatten out at areas larger than that, see the figure at the end of this document.

Within each circle the following info was extracted:

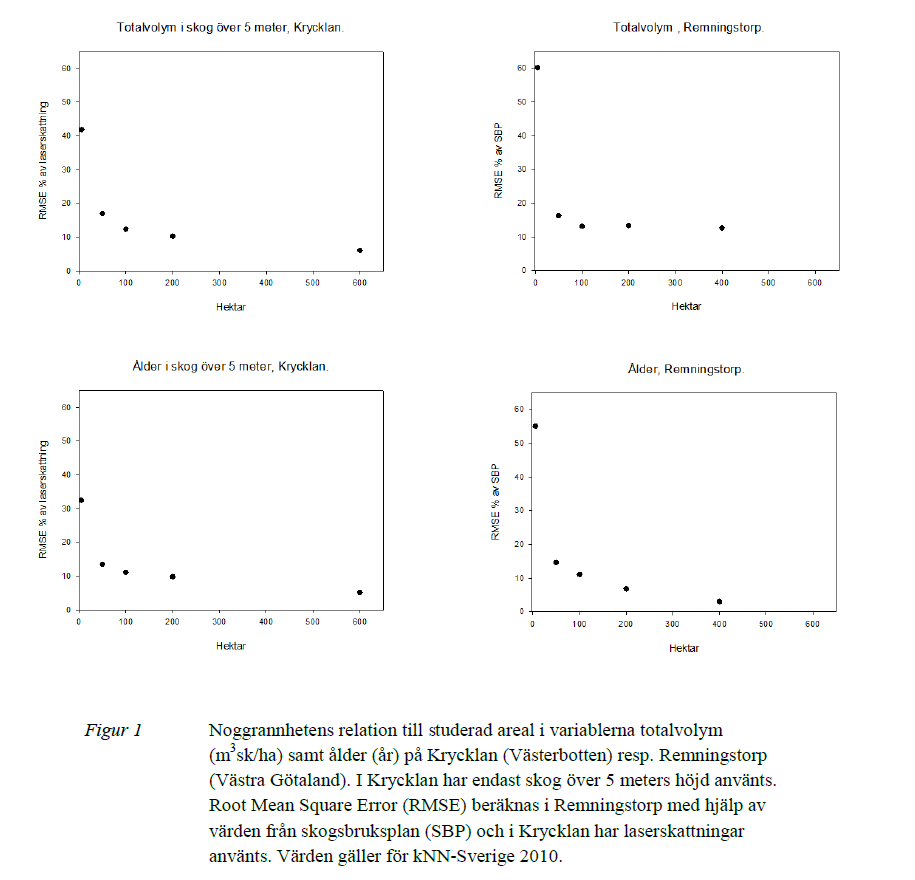
* Coverage (how much of the circle that consisted of forest in hectares).
* Mean age of the forest within the circle
* Total volume of forest within the circle (m3)

We noticed that there were some empty patches in the data layers (in less than 5% of the instances). I e-mailed the person responsible for the database and asked why. I was told that in rare instances clouds can be in the way of the satellites which cause these data gaps. In those cases, I used the mean value from the other two years to fill in the gap. You can identify these cases (and exclude them if you want) by the variable “clouds=1”.

Some sites don’t have any data at all, maybe they are outside the data range for the forest data (in fjällen maybe); these are just specified as “NA”.

If you have any questions or if something feels off – just ask.

Mvh Calle och Erik



Hej Calle (Cc Erik),

Great job! I am really looking forward to including the new variables in the models!

You made it quite fast, I am impressed :0

A few questions:

1) You have reported the mean coverage in hectares in a area of ca 154 hectares. What was the original resolution of the data? I mean, how large are the pixels that you have used to estimate the average?

2) Was the pixel size the same for data related to age and volume?

3) Is the total volume of forest a function of coverage? Or to be clearer: of course the two are correlated, but was the volume estimated independently from the coverage or is derived matematically from the coverage?

Btw,  thanks for the bonus variable "Clouds", very thoughtful.

Ha det canonbra!

Serena

Hej Serena,

Thanks!

1.    The original resolution of the data is 25x25 meters, which is quite high (but the creators of the data stress that the study area should be a quite large when using this data). The pixel size means that a circular buffer zone around an electrofishing site could contain approximately 2400 pixels of information. The coverage is equal to the number of pixels with information within the buffer, instead of number of pixels, it is converted to hectares in your file (if I remember correctly). In agricultural areas where there may be very little forest, only a few pixels carry information of forest age, and so the coverage is low.

2.    Pixel size was 25x25m for all layers included

3.    The forest volume is provided in standardized units of m3 of forest per hectare for each 25x25m pixel (a little mind boggling at first). A standardized volume of forest per pixel alone will not say much if we don’t know the coverage! So the mean density (m3 of forest per hectare) within the buffer is multiplied with the coverage (number of hectares) within the buffer to receive the total volume of forest within the vicinity of a site.

I pasted a picture of how the data looks together with the sites and buffers so you’ll get an idea. 

The layer in black and white represents forest, the darker it is, the older it is, the lighter, the younger. You see the patches of white, these are most definitely clear-cut areas. The dark patches are older patches of forest. The pink dots are your sites, and the alpha’d buffers around them are the 154 hectare “sample areas”. The green color represents areas where the forest layer has No Data, and are sites with, for example, agricultural areas, lakes, swamps, cities, etc. The number of filled pixels within the buffer is the coverage.

Hope that makes sense!

Cheers, Calle

Thanks Calle for the detailed explanation! It makes perfectly sense.

I imagine that the pixels of forest are derived from satellite images. But what about the density? Is that also calculated from satellite images? and the age? Do you know if there is any ground-truth been done? I guess so, but I just can't help having questions ;)

Cheers!

Serena

Good!

No problem, it’s good to have questions towards huge chunks of data like this.

The forest data is a combination of satellite data and field inventory data (called Riksskogstaxeringen in Swedish). In the field inventory 11000 samples are taken scattered evenly nation-wide, whereof half falls within production forest and I would guess the rest into other kinds of forest. It says the collective total sampled area is 130 ha, which is less than the buffer I chose around one of your sites. Yikes!

I imagine the forest data is collected with satellites and calibrated and controlled with the field data.

Mvh Calle