# Report on the South-East Asia Poverty Hotspots – Operational Appendix

The intense analytical work was essentially aimed at

**a.** preparing three databases at the administrative level for the RoI (the six SE-Asia countries plus Yunnan);

**b.** analysing them via multivariate techniques (Principal Components Analysis + unsupervised Clustering).

Many of the variables included in those datasets were derived from digital maps, which had to be suitably checked, recoded and resized in order to adapt them to the purpose. The operation required the writing of some software from scratch, as well as the adaptation of existing software.

As requested by the TORs, the available digital images of the RoI were reviewed, evaluating their utility and applicability for the research.

All the images considered, briefly commented below, have generally been converted to WINDISP format and resampled to the resolution of 2.5’ (0.0416666667 deg), also using some of them at the resolution of 1km when convenient.

A shapefile merging 125 administrative units for the Yunnan with the 207 administrative units of the other countries has been created, and generally used for all the operations carried out at the administrative level. From it a mask was derived, that allowed the exclusion of the external areas from the operations carried out at pixel level.

**Not all the available images proved useful or reliable**, as it will be said in detail. Should a digital data set be created, a selection would be necessary.

## 1. The 1-km SPOT/VGT images

The VGT1 sensor mounted on the SPOT4 platform became operational in April 1998, issuing images at 1km (0.0089285714 deg), both daily or dekadal syntheses (S10). The images are available on the VITO site (free.vgt.vito.be) as continental windows, in HDF format and zip-compressed. Unfortunately, as our RoI was split between two such windows, it was necessary to download the whole set of NDVI S10 images for SE-Asia and Asian Islands, from April 1, 1998 to November 3, 2005.

CROP_VGT was used to extract from each of them, in IDA format, the area included in the RoI. The two subareas were then joined, obtaining a whole series of 276 dekadal images spanning from Apr1, 1998 to Nov3, 2005. Each image, in geographic projection, had 3249 lines of 3081 samples, covering the RoI between the UL corner (92E, 29.5N) and the LR corner (119.5E, 0.5N).

The 276 dekadal images thus obtained were processed: the long series of each pixel, consisting of 276 values, was made more regular by interpolating through maxima and simulating the missing (cloudy) values.

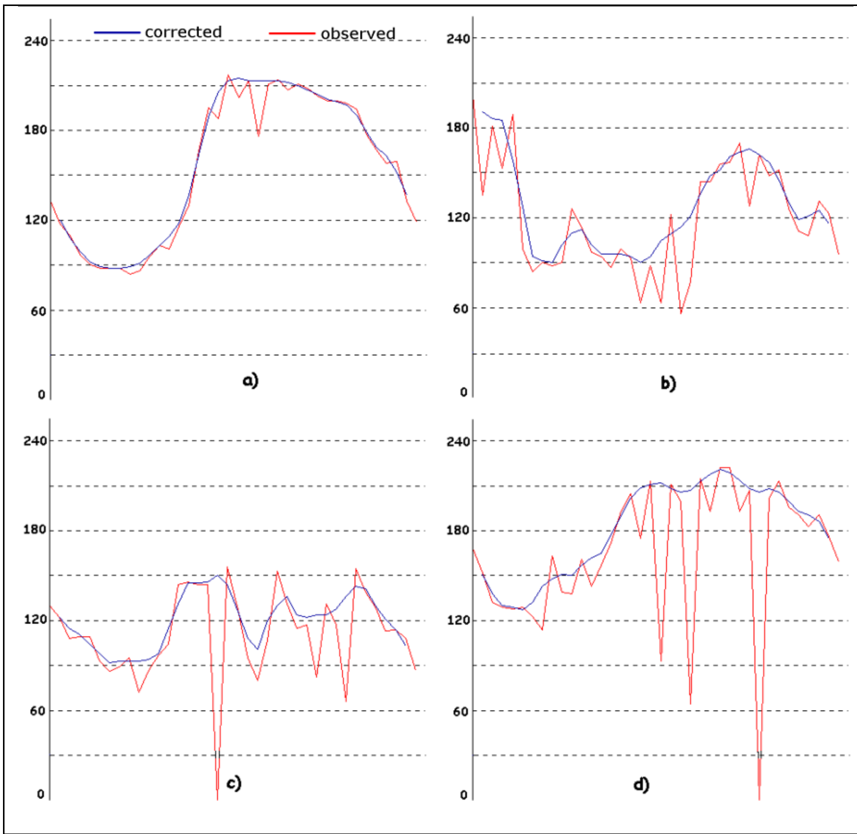


**Fig. 1 – Example of NDVI observed and corrected pixel profiles.**

Here only a few dekads are shown, but the correction was actually carried out, for each pixel, on the whole series of its 276 values from Apr1, 1998 to Nov3, 2005.

From the images output by this procedure the **36 dekadal historical averages** (AVE) and **standard deviations** (STDEV) were eventually computed, by dekad and pixel.

**Click here** to display an example of a VGT historical average image (September, second dekad), masked and in shades of grey. The image size is actually 3249 x 3081 pixels.

## 2. Computation of the series of images representing the relative interannual variability.

The relative inter-annual variability, assumed as an indicator of risk of cropping failure, was computed by pixel and dekad, obtaining a series of 36 images. The images used as input were

- the dekadal series of the NDVI values (historical averages 1998-2005), pixel by pixel;
- the dekadal series of the standard deviations of the NDVI value over years 1998-2005;

both computed processing the complete series Apr98-Nov2005, after reconstructing clouded pixels.

For each pixel and dekad, the relative inter-annual variability is defined as

$$i.a.\mathrm{var}\,iability = \frac{stdev(rv)}{AVE(rv)} = \frac{slope * \Delta bc2}{slope * bc1 + \mathrm{int}\,ercept} \qquad (1)$$

where *AVE(rv)* is the real value of the historical average NDVI, *stdev(rv)* is the real value of the inter-annual standard deviation (expressed in NDVI units), bc1 is the NDVI byte count value, bc2 is the byte count in the stdev image. For VGT NDVI images, slope = 0.004 and intercept = -0.1.

Byte count 25 is critical for NDVI VGT images, as it corresponds to NDVI = 0, which causes the output value to diverge. Byte counts close to 25 also give some problems, as they correspond to very low NDVI values. This results in excessively high values of the inter-annual variability.

To cope with this, the formula (1) was modified as follows:

$$i.a.\,variability = \frac{stdev(rv)}{AVE(rv)} = \frac{slope * \Delta bc2}{slope * bc1 + intercept} * attenuationfactor(AVE)$$

where *attenuation factor(ave) = 1 - exp(-643.78*ave*ave)*

The output values were coded onto [1..250] and saved as images IDA of type 200, using for all images common values of slope and intercept determined on purpose. Working at pixel level is more critical than working at district level, as there is no averaging advantage.

It makes no sense to consider NDVI values less than, say, 0.050. These are scarcely meaningful for production, but are presumably associated with high values of variability for mere mathematical reasons. This has a strong impact on the scale used for the output images (slope and intercept), compressing the really significant values into a smaller sub-interval, with loss of details.

A set of ratio images have been produced for the 36 dekad, and actually this is what occurs. The ratio value is very low for almost all pixels, compressed by the effect of the few pixels with a high ratio, that eat up almost all the range in [1..250]. Therefore, pixels with historical average NDVI values lower than 0.050 have been excluded, setting conventionally the output value to 252 (253 is used for clouds, 254 for land outside the Region of Interest, 255 for water mask) and coding in [1..250] only pixels with NDVI values greater or equal than 0.050.

The disadvantage is that for some pixels the series of the inter-annual variability by dekad will present some invalid values 252. The fact can be overlooked if what is actually used is some map of average i.a. variability over the year, or something similar. Instead, it becomes a problem for multivariate analyses carried out at pixel level.

**Eventually, we have three series of 36 dekadal images** that respectively provide, by pixel and dekad, the historical averages, the inter-annual standard deviations, the relative inter-annual variability.

The first series was the input for an **eco-climatic classification** (see point 4 below).

From each series a synthetic image was also computed, by averaging up over dekads, to be used as a relevant indicator both at administrative and pixel level. These synthetic images represent

1) the average Vegetation level over a standard year;

2) the standard deviation of the vegetation level over the year (the intra-annual variation, useful to sort out forests from cultivated land), not to be confused with

3) the inter-annual variation, that is an indicator of risk.

## 3. The Elevation Data

The **SRTM** (Shuttle Radar Topographic Mission) data for the whole RoI were downloaded from the site http://srtm.csi.cgiar.org. The Elevation images, 16 bits per pixel and with a resolution of 90 m, were already spatially interpolated in order to eliminate data gaps still present in the NASA version. Click here for a sample image of Yunnan.

The 5° x 5° degrees tiles downloaded were in ASCII format, 6000 x 6000 pixels each. As 23 such tiles were necessary to fully cover the RoI, the memory burden caused by their simultaneous use was quite heavy. However, as the DEM data had to be used as the background to help interpret the eco-climatic classification of the Region, whose resolution was 1 km, the 90m resolution for the DEM meant a waste of resources; all DEM data were then resized to 360m, saved in BIL format with their accompanying text info, and used as such. The size of each resized tile, 16 bits per pixel, is 4.5MB.

## 4. An eco-climatic classification

Some classifications with tentatively different numbers of classes were carried out on the series of 36 historical averaged NDVI VGT images (see point 1 above), assumed to represent the vegetation evolution in a standard year. The purpose was to single out forested and cultivated areas, identifying different types of vegetation cycles over the year, in particular cropped area with a double- and even a triple-peaked cycle. The analytical sequence included a PCA + an unsupervised clustering. The area outside the RoI was masked, while water pixels within the RoI were considered valid and assigned. Only 49 cloudy pixels on the Yunnan mountains could not be reconstructed: the classifications can therefore be considered practically exhaustive.

Routines included in Addapix for Windows were used to construct the classifications, to visualize the average profiles of the classes and to inspect the vegetative profiles of individual pixels in specific areas. Two classifications, with 20 and 40 classes respectively, appeared particularly interesting: the former was sufficiently synthetic to allow an easy description, the latter more burdensome, but detailed enough as to put in clear evidence double-peaked areas.

The classes were re-ordered according to their average vegetation level, the last one always gathering water pixels. For each classification carried out, the program also creates an appropriate palette to visualize it. The underlying DEM data is an invaluable support to understand the features and the spatial distribution of the classes. See this PPT presentation as an example.

## 5. An excerpt of the GLC2000 classification

The Global Land Cover classification, produced from the set of S10 VGT images for year 2000 by a team of international partners, co-ordinated by the GVM Unit of the EC Joint Research Center in Ispra, was also used as reference for the singling out of cropped areas.

The data was downloaded from www-gvm.jrc.it/glc2000. Also in this case it was necessary to assemble an image of our RoI from two different GLC2000 classified images, one for China, the other for SE Asia, and coping with the fact that the legendas of the two images were different.

## 6. Disaster hazards

Raster images mapping the distribution of the hazard for the occurring of some types of disasters were downloaded from ftp.ciesin.columbia.edu/pub/hotspots. The disasters possibly affecting our RoI were **cyclones, drought, flood, landslide and earthquakes**. The images were published by Dilley and others, on the basis of the historical distribution of events. For example, the distribution of cyclones hazard took into account more than 1,600 storm tracks for the years 1980-2000 for the Atlantic, Pacific, and Indian Oceans.

The images have the resolution of 2.5’ (0.041666667 deg) and cover all the planet from 85°N to 58°S. The window for our RoI was cropped from them, and duly masked.

In all single-hazard images pixels with null hazard had value 0. All pixels with a significant hazard were split in ten equal-frequency groups, and the value assigned to a pixel corresponded to the decile to which it belonged. This enabled a quantitative treatment of data.

The images were used, in different occasions, both at the original 2.5’ resolution, or resized at 1 km.

Also a multihazard image was available on the site, but the way it was coded made it almost impossible to retrieve the combination of hazards affecting each pixel. In single-hazard images, due to the nature of the phenomenon mapped, values came in large spatial blocks. It was quite straightforward to perform a classification assuming as input the five single-hazard image and processing data quantitatively. **Click here** to display the resulting map, showing the spatial distribution of combined disaster hazards.

## 7. Gridded Population and population density of the world images (CIESIN-Columbia)

Raster images downloaded from the site http://sedac.ciesin.columbia.edu/gpw/global.jsp give the population per pixel (only for the northern emisphere) every five years from 1990 to 2015 (images 2005-2015 are estimated).

The resolution is 2.5’ (pixel size = 0.041666667 deg) and pixel values are coded using 32 bits. The extraction of the RoI, both maintaining the original 2.5’ resolution or resampled to 5’ in order to allow use with other layers at 5’, was done with ENVI for all available years.

In the cropped images (698 lines by 662 samples, common size of all images of our RoI at the resolution of 2.5’) pixel values, expressing the pixel’s population in units, ranged from 0 to 1,500,000 ca. Out of the urban areas, specially in forested parts, the population value was only some hundreds or even some tens of units, and the recoding to one byte per pixel for use with WINDISP was not possible: everything out of towns would have been flattened.

The images were therefore left 32-bit coded, extracting directly from them the values relative to the various administrative areas. For some particular use, some software was written to split each of them into three suitable 1-byte images; not worth to describe here how and why.

## 8. Construction of the grid image mapping the 2005-15 average population growth rate

Two raster images (at 2.5’ and 5’) were computed to map the population annual growth rate, directly from the two CIESIN 32-bit population maps for 2005 and 2015. Software written on purpose was used for this operation.

Results were coded 8-bit in IDA format, computing slope and intercept from the real values encountered so as to get byte counts in the range [1..250]. Obviously, this simply means **to re-discover** the rate used to project the population to 2015, rate that CIESIN does not declare explicitly.

From the image it appears evident that the CIESIN projection was indeed a very rough operation: growth rate values are almost uniform all over Cambodia, VietNam and Laos; they are spatially well detailed and differentiated only for Thailand and Malaysia. For Myanmar and Yunnan the rate appears to vary by administrative units (which are few and very large in Myanmar, quite small in Yunnan).

**Click here** to display an image where each single rate value has been represented with a different color, in order to visualize even tiny differences.

**The reliability of the gridded population distribution, and of the growth rate computed from it, is therefore questionable.**

Values for pixels with null initial or final population could not be computed, and they had to be declared **invalid** in the resulting image. This puts an obstacle to the use of this image (as it occurs for some others...) as an input for multivariate analyses carried out at pixel level.

## 9. The global AEZ/IIASA-FAO raster images

The plates processed and used were downloaded from the CD attached to the volume *Global Agro-ecological Assessment for Agriculture in the 21$^{st}$ Century – Methodology and Results*, by Fischer, van Velthuizen, Shah and Nachtergaele.

The Plates **listed here** were checked, and the window corresponding to the RoI was cropped from each of them. It was then decided to use only some of them to extract statistics at the administrative level.

The original images used categorical codes, though each code had a quantitatuive meaning in the original plate legenda. In order to handle them quantitatively, they were suitably recoded as **shown here**, and all resampled to 2.5’.

The information offered by the plates is not complete. Large areas, e.g., most of Yunnan, are often void of data. This hinders the use of these images as an input for multivariate procedures.

## 10. The minNPP images

The **minNPP** images were received from FAO at the 5' (0.0833333333° deg) resolution. They have been reprojected to 2.5’ with some corrections, obtaining images exactly matching the others relative to the RoI.

From the minNPP images for 2005 and 2015 (byte counts coded in [0..253]), the following two images were derived:

- the absolute variation NPPAbsVar05-15.img (values coded in [1..250])
- the relative variation NPPRelVar05-15.img (values coded in [1..250]).

## 11. What is still pending

- A description of the eco-climatic classification: more synthetic (20 classes) or more detailed (40 classes)? How organized? Is it necessary?
- A description of a pixel-based classification that was carried out on the three synthetic images mentioned under point 2, derived from VGT historical averages and standard deviations. Is it worth?
- Maybe a further attempt to carry out a multivariate analysis on the set of images described in this document, and **summarized here**, at the 2.5’ scale. Some analyses have been tried already, with no satisfactory result so far, due to the poor quality of most inputs and to the areas void of data existing in them.