Proposal for Outlier and Trend plots to be used in the new EU-GHG inventory report

Adrian Leip, 09/10/2014

# Outlier plots

Outlier plots are very useful because

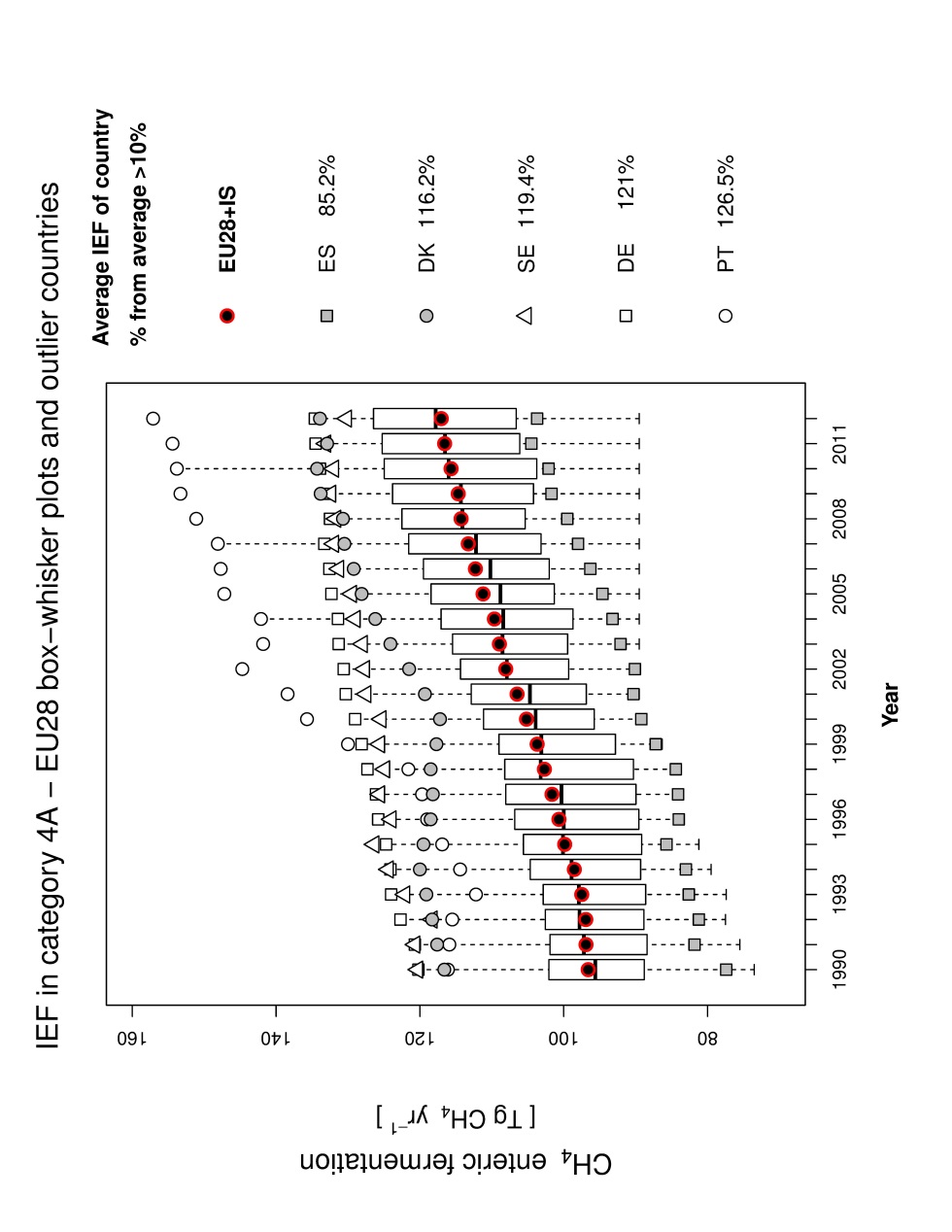
* They allow the sector experts to identify issues that need to be discussed with the countries
* They could be made useful also in the NIR by guiding for which countries additional information is useful to understand the IEF for EU28+IS

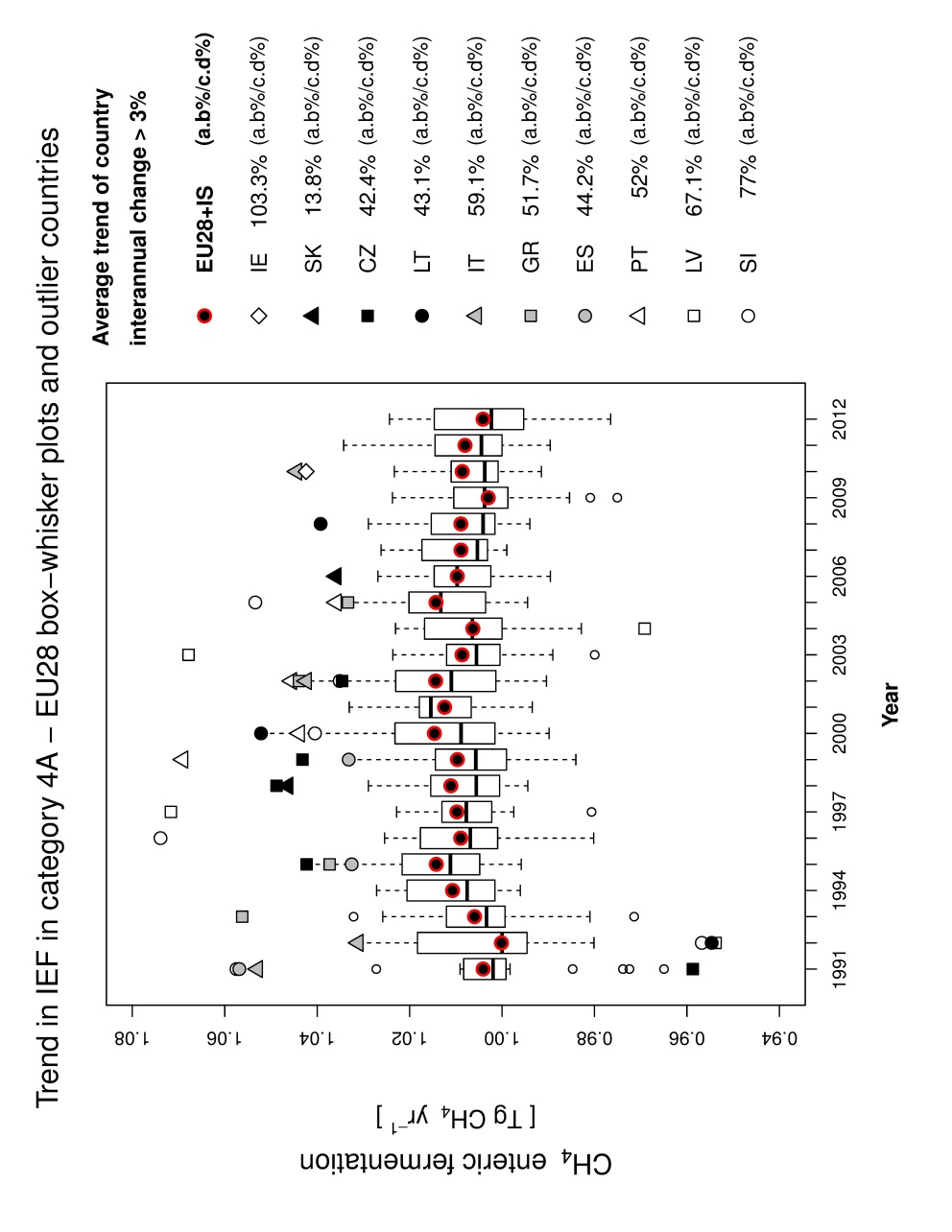
The plots should show information relevant for EU28+IS and limit country-specific information only to those countries where “outliers” occur.

The **suggestion** is to use box-whisker plots that show the mean, lower/upper quartiles and a ‘whisker’ which defines the outliers as being beyond a certain factor of the inter-quartile range.

* The countries to be plotted are chosen on a criterion, having an IEF that diverts from the EU28+IS average by more than 10%, but no more than 10 countries would be selected. However more meaningful and “sophisticated” criteria could be used (e.g. using their ‘weight’ for EU28+IS IEF, on the basis of more refined statistical evaluations, etc.)
* The full time series of the countries is then plotted, also for those years when their IEF is close to the EU28+IS average
* The figure could be complemented with a table giving quantitative data
* Note that the current figure plots EU28+IS average IEF without accounting for the different weights of the countries, this needs still to be implemented.
* Another figure could be designed focusing on the interannual changes of the IEF
* Additional ‘flags’ could be introduced (e.g. ring around some points) that indicate if an outlier has been justified (and explained in the text or a table) or if the reason is still unclear, or to highlight certain characteristics.

The examples below show only the principle of the figures suggested. Details would need to be discussed, such as selection criteria, information to present beside the data (a/b%...).





# Trend plots

This proposal is based on the discussions on 19/06/2014, as also reflected in the minutes, in particular:

* Plots similar to the EEA trend plots or the trend plots used in the agriculture chapter already might be a good display item to be used in the sectoral chapters
* In view of the increased numbers of MS, and also the decision to focus the EU-IR on EU-wide relevant data, those plots should contain only the most important countries
* Additional information on the contribution of AD and IEF (in case of trend in emissions), as already done previously, is desirable

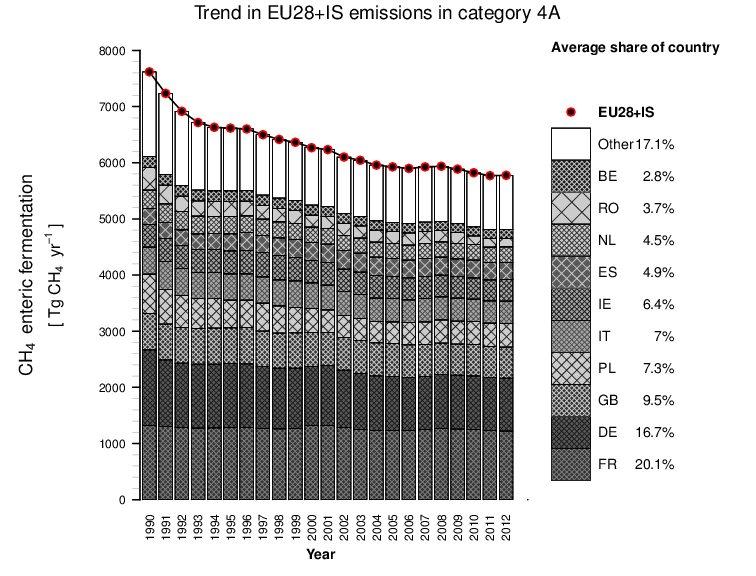
The EEA plots currently are based directly on the xml-files submitted by the countries. They are downloaded, filtered and relevant data are written into text files. Loops over various graphical scripts then produce the plots that are distributed. So far, the scripts use the gnuplot-language. Here I propose to use the R-statistical graphical package for two reasons (i) according to my experience, R graphical packages are more flexible (and powerful) then gnuplot; (ii) the tight integration with statistical packages makes it very attractive for combination with other evaluations, such as additional checks (e.g. those made for agriculture), or even outlier/trend analyses (e.g. for the Theil-Sen estimator used in the outlier tool, a package ‘openair’ for the analysis of air pollutants exists already). R runs on Linux and Windows.

The figures below have been constructed with a few additional ideas:

1. Symbols for countries should be identical in all plots throughout the IR. That means that a careful selection of each symbol must be done so that in all combinations the countries are clearly recognizable. As given in Table 1 I used four dimensions: a combination of background and pattern colors, pattern density and pattern angle. As Table 1 shows this is completely flexible. So far I used only grey colors (to make it readable on black/white printers and for color-blind persons) and diagonal patterns.
2. For each plot all countries are ranked and the most important x countries are plotted, the other lumped into a last category “Others”. Here a simple criterion of the average share of contribution to total CH4 emissions over the whole time period is used and the upper 10 countries are plotted. It is proposed that a common criterion is developed as ‘default’ procedure for all plots, even though for specific plots another evaluation of the countries can be done. Also the ‘default’ number of countries should be the same for all chapters, even though also a flexible solution (all countries with a share > 5% or so…) could be used.
3. The two figures below show the same numbers, once at total emissions (and shares) and once for annual changes. In the latter, a dummy is plotted for the contribution of AD/IEF in case of emissions. In case a plot is done e.g. for indirect N2O emissions, other solutions such as contribution from N-input/Leaching fraction/EF are possible.

The layout of the plots is obviously preliminary and would need to be discussed/optimized.

The current proposal is applicable to any ‘cumulative’ data (ie activity data, emissions) as a start. In case this finds a general agreement, work on plots for non-cumulative data (ie. IEFs, parameters) will follow.



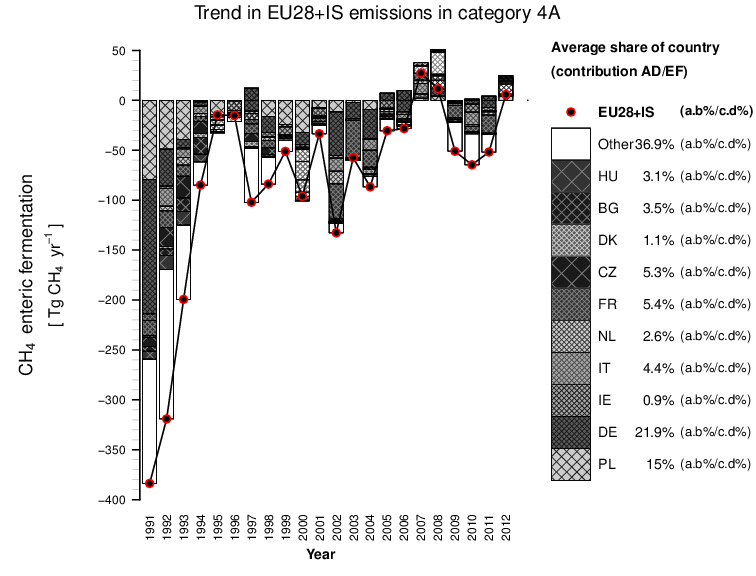


Table 1. Attributes for country-plots

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | cnt color | dens | coll | ang1 | ang2 |
| AT | grey10 | 20 | grey50 | 45 | -45 |
| BE | grey10 | 20 | grey80 | 45 | -45 |
| BG | grey10 | 10 | grey50 | 45 | -45 |
| CY | grey10 | 10 | grey80 | 45 | -45 |
| CZ | grey10 | 5 | grey50 | 45 | -45 |
| DE | grey35 | 20 | grey10 | 45 | -45 |
| DK | grey35 | 20 | grey80 | 45 | -45 |
| EE | grey35 | 10 | grey10 | 45 | -45 |
| ES | grey35 | 10 | grey80 | 45 | -45 |
| FI | grey35 | 5 | grey10 | 45 | -45 |
| FR | grey20 | 20 | grey50 | 45 | -45 |
| GB | grey20 | 20 | grey80 | 45 | -45 |
| GR | grey20 | 10 | grey50 | 45 | -45 |
| HR | grey20 | 10 | grey80 | 45 | -45 |
| HU | grey20 | 5 | grey50 | 45 | -45 |
| IE | grey65 | 20 | grey10 | 45 | -45 |
| IT | grey65 | 20 | grey30 | 45 | -45 |
| LT | grey65 | 10 | grey10 | 45 | -45 |
| LU | grey65 | 10 | grey30 | 45 | -45 |
| LV | grey65 | 5 | grey10 | 45 | -45 |
| MT | grey80 | 20 | grey10 | 45 | -45 |
| NL | grey80 | 20 | grey20 | 45 | -45 |
| PL | grey80 | 10 | grey10 | 45 | -45 |
| PT | grey80 | 10 | grey20 | 45 | -45 |
| RO | grey80 | 5 | grey10 | 45 | -45 |
| SE | grey90 | 20 | grey10 | 45 | -45 |
| SI | grey90 | 20 | grey20 | 45 | -45 |
| SK | grey90 | 10 | grey10 | 45 | -45 |
| Other | white | 0 | grey20 | 45 | -45 |