New Methodology for Statistical Imputation in the FAOSTAT Production Domain

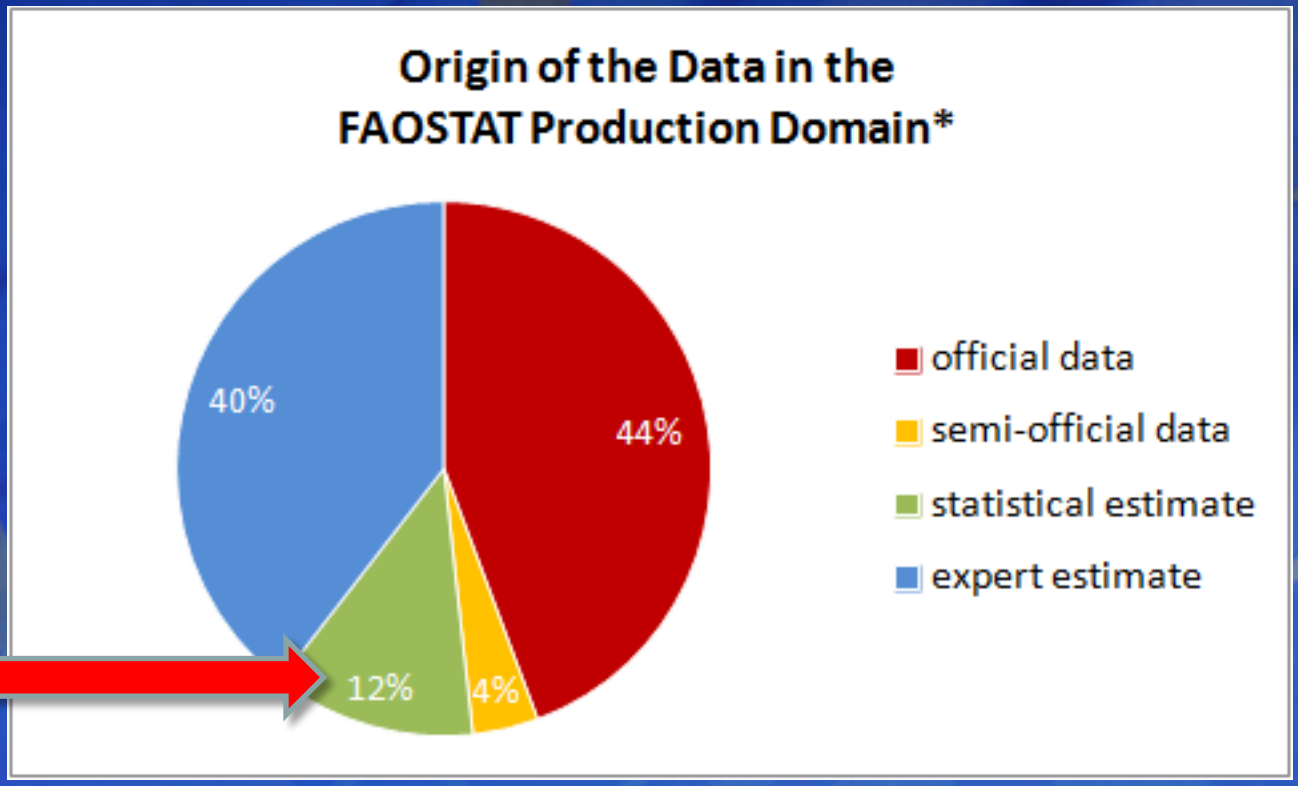
Onno Hoffmeister, Statistician, ESS

Rome, 11 April 2013

Contents

- Context
- Description of the Methodology
- Evaluation of the Results

Scope of Statistical Imputations



Focus of this presentation

*) Data on production and input factors of primary agricultural products in FAOSTAT from 2000 to 2010.

Expert Estimates

Based on the best available knowledge of the Country Statistician, taking into account:

- background knowledge of the conditions in the country (studies and reports, country visits, country experience, …)
- fit with other elements of the SUA
- technical conversion factors in other years
- time trend of the data series

Potential Gains from Statistical Imputation

- economization of work load
- objectivity
- accuracy

Concretization of the Aim

Common objectives of imputations (Kalton and Kasprzyk 1982):

- preserving the distribution
- deriving reliable sample estimators
- “assigning values at the micro level and thereby allowing analyses to be conducted as if the data set were complete”

Main objective for FAOSTAT

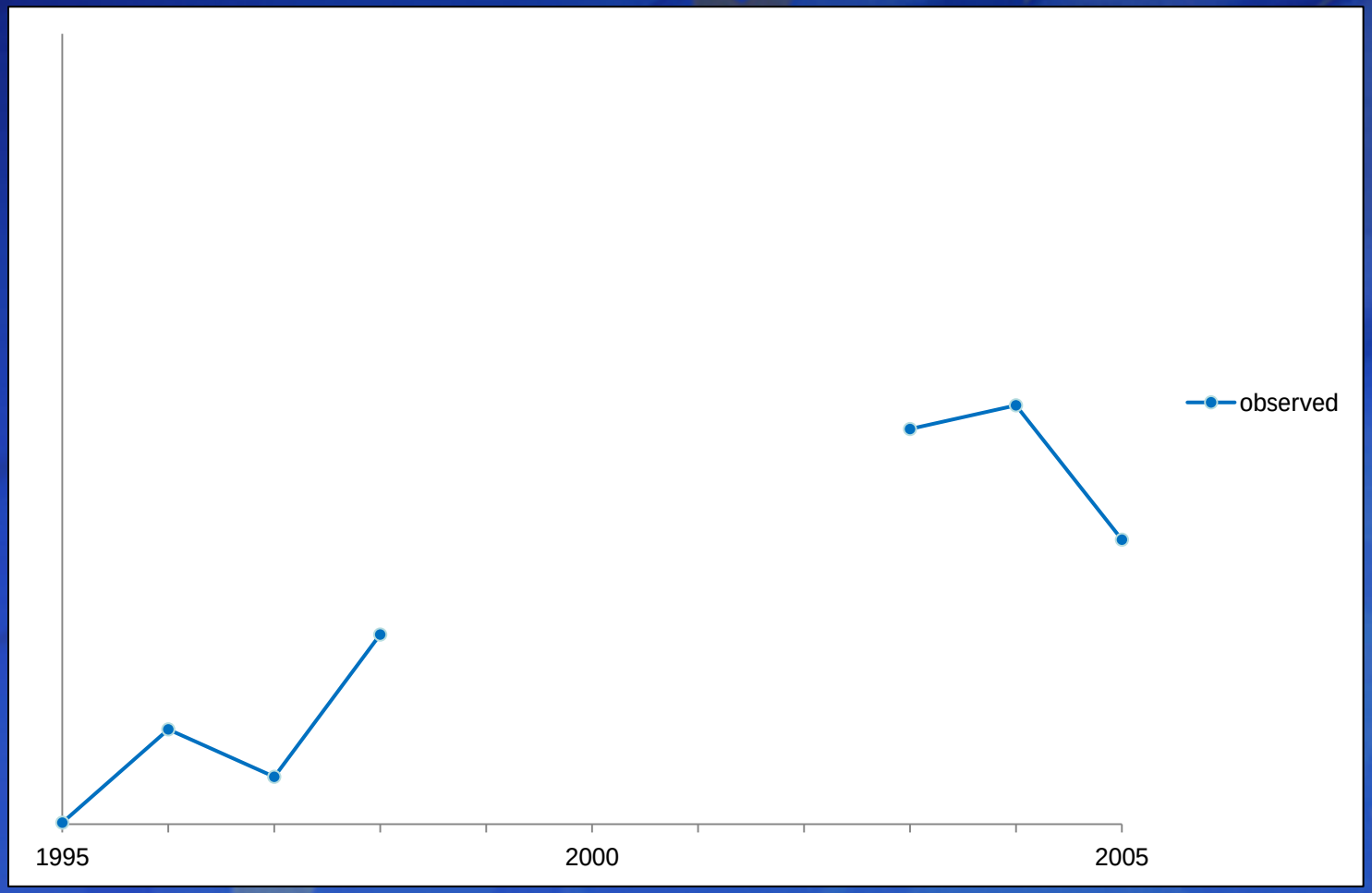
Applied Methodology

Growth rate is benchmarked on a related aggregate.

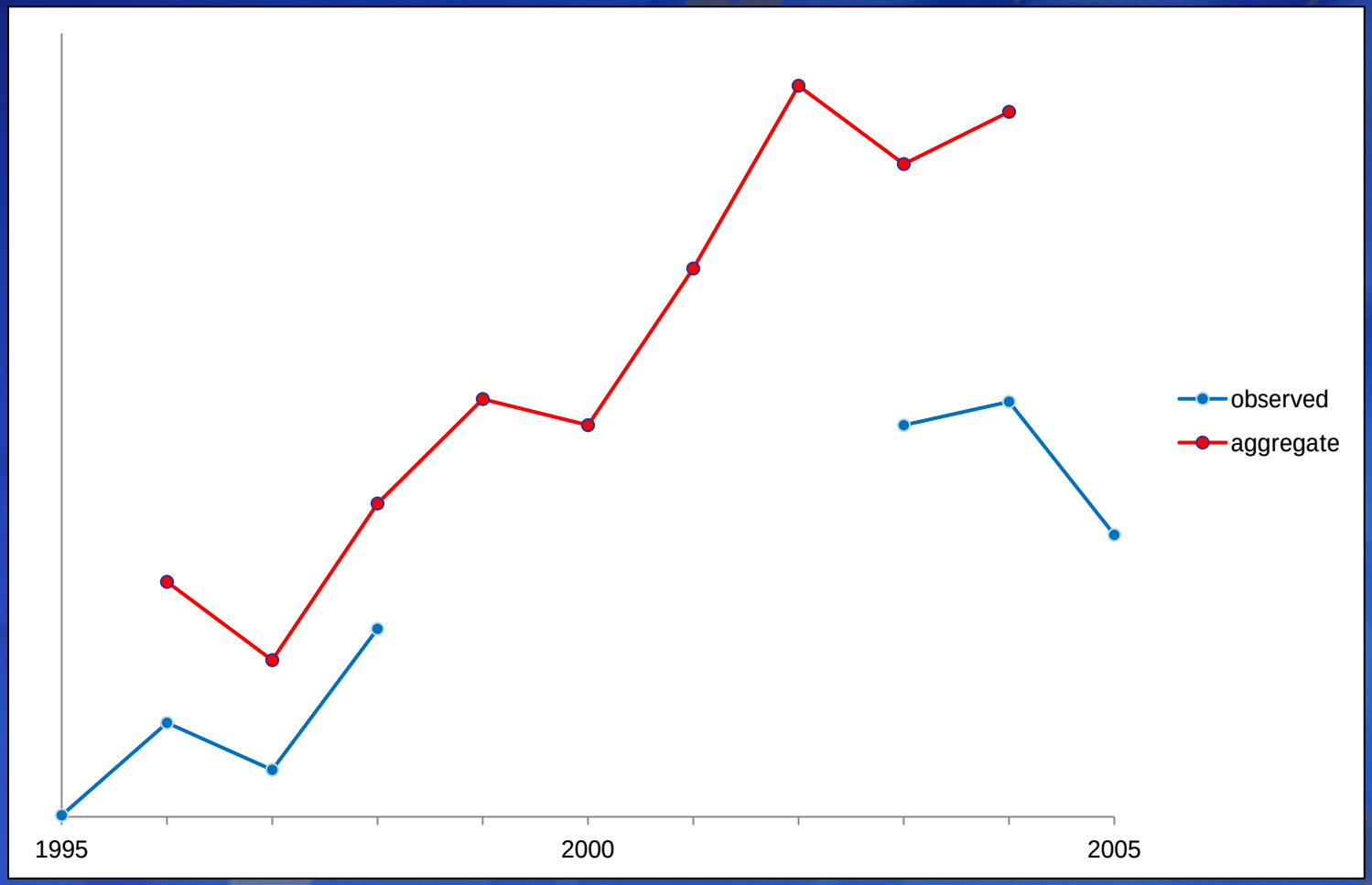
$$\bar{r} = \left(\left(\frac{\sum_{c \neq c_0} y_{c,t-l+1}}{\sum_{c \neq c_0} y_{c,t-l}} \right) \left(\frac{\sum_{c \neq c_0} y_{c,t-l+2}}{\sum_{c \neq c_0} y_{c,t-l+1}} \right) \cdots \left(\frac{\sum_{c \neq c_0} y_{c,t}}{\sum_{c \neq c_0} y_{c,t-1}} \right) \right)^{1/l} - 1$$

$$\hat{y}_{c_0,t} = y_{c_0,t-l} \left(1 + \bar{r}\right)^l$$

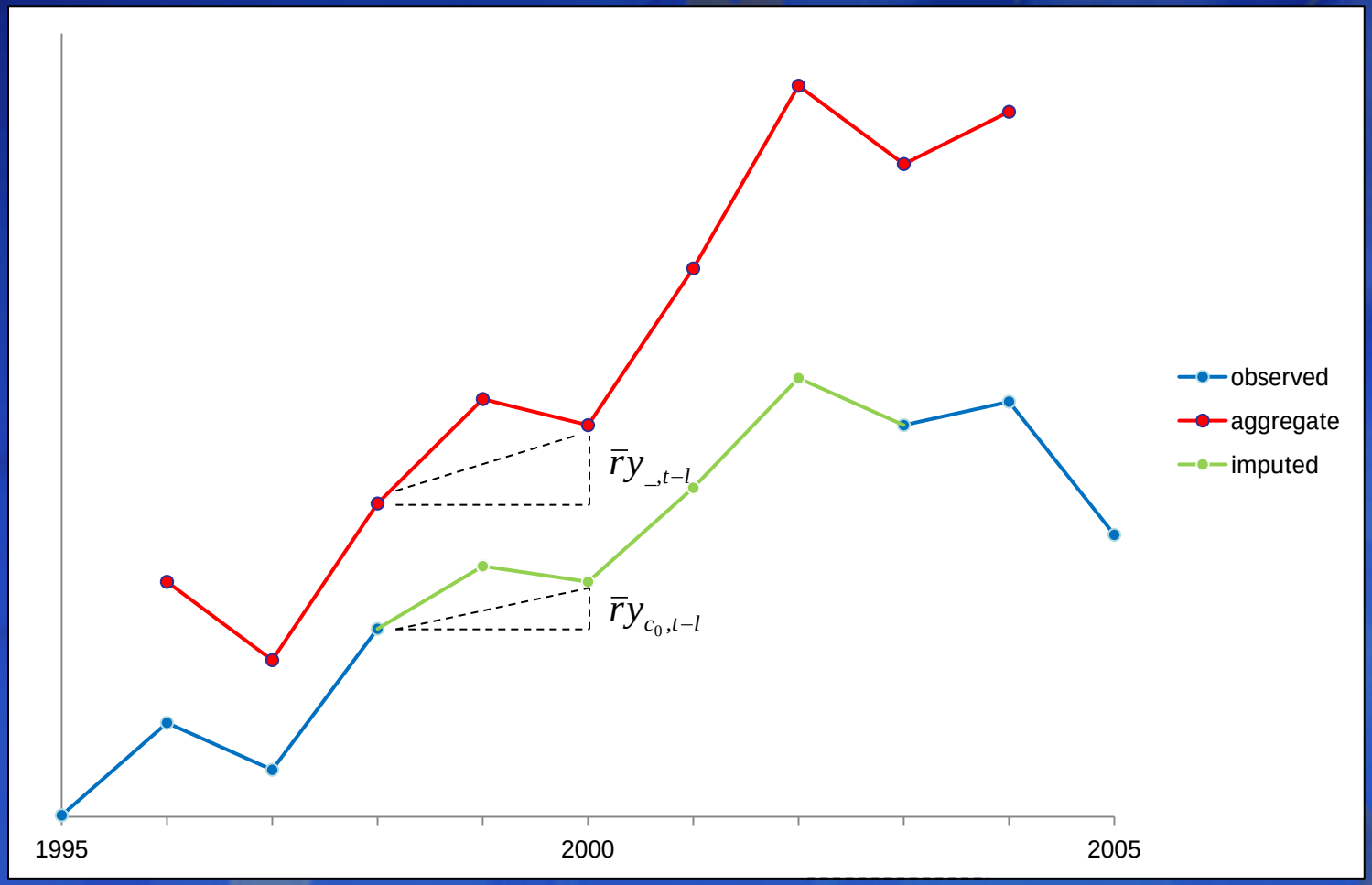
Applied Methodology



Applied Methodology



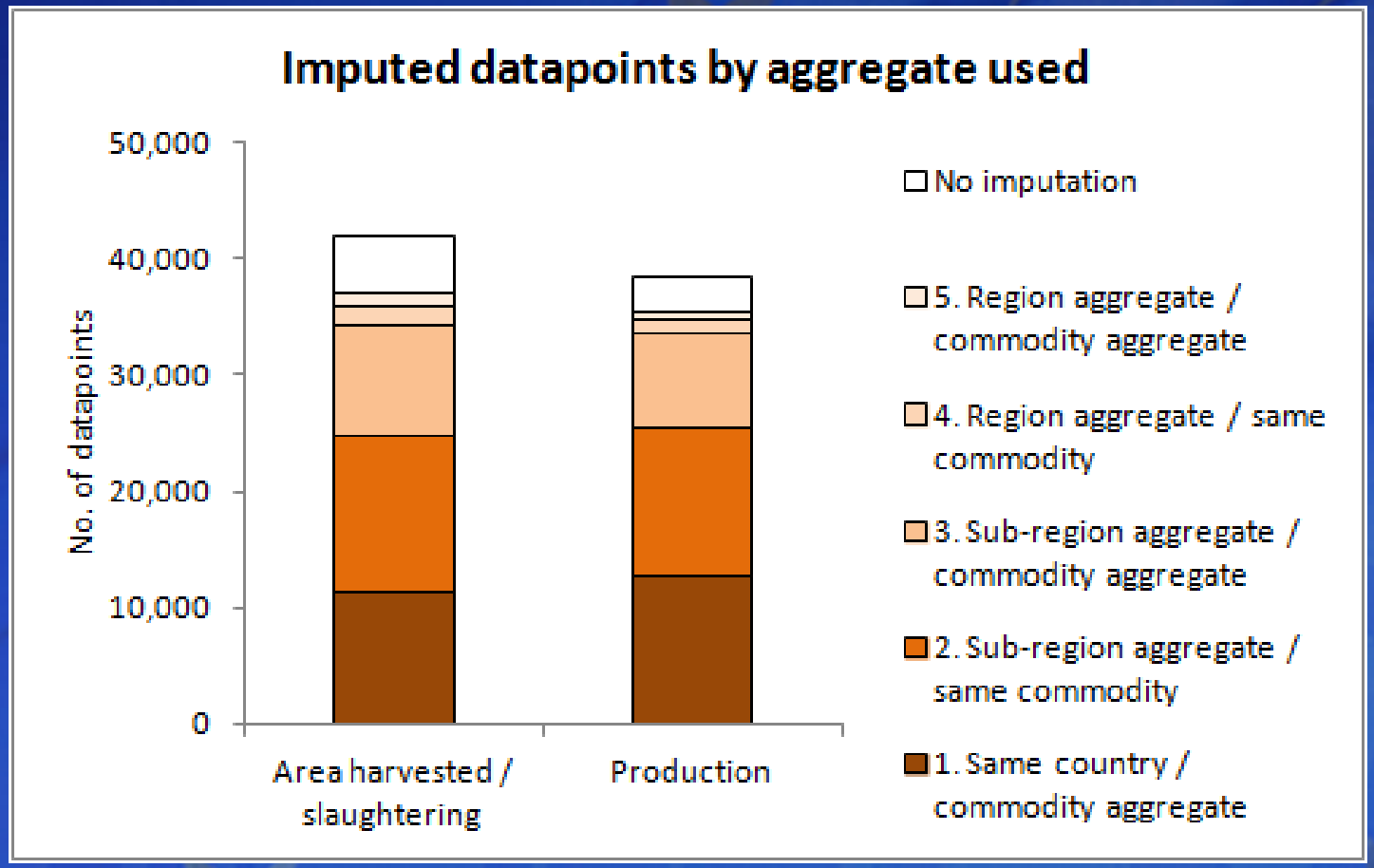
Applied Methodology



Selection of the Aggregate to Benchmark on

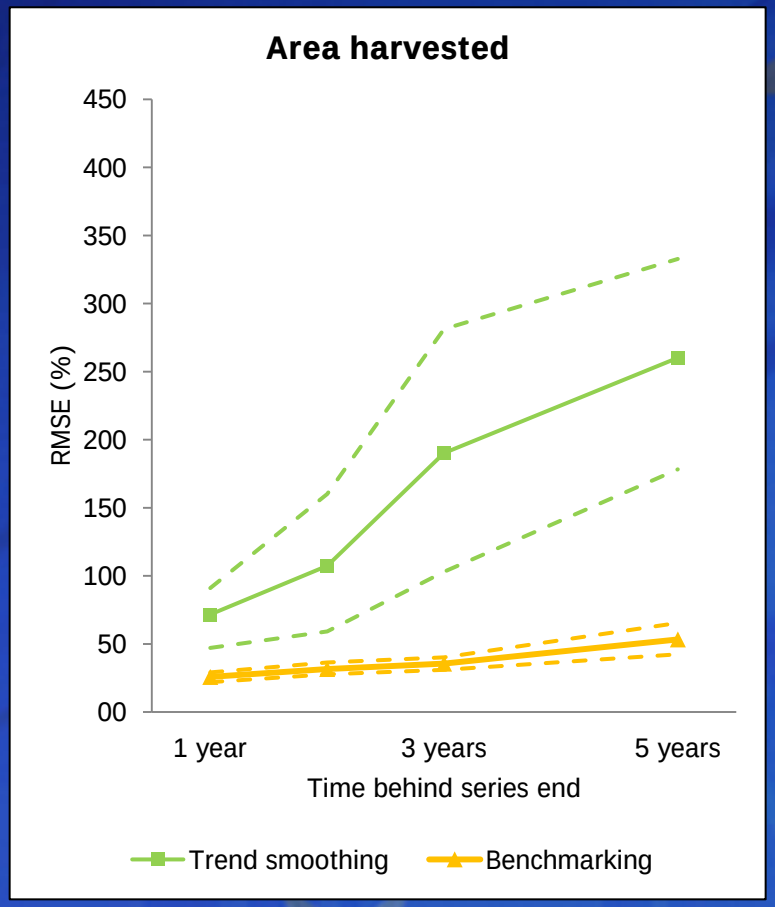
Priority	Aggregate
1.	Same country / same commodity
2.	Sub-region aggregate / same commodity
3.	Sub-region aggregate / commodity aggregate
4.	Region aggregate / same commodity
5.	Region aggregate / commodity aggregate

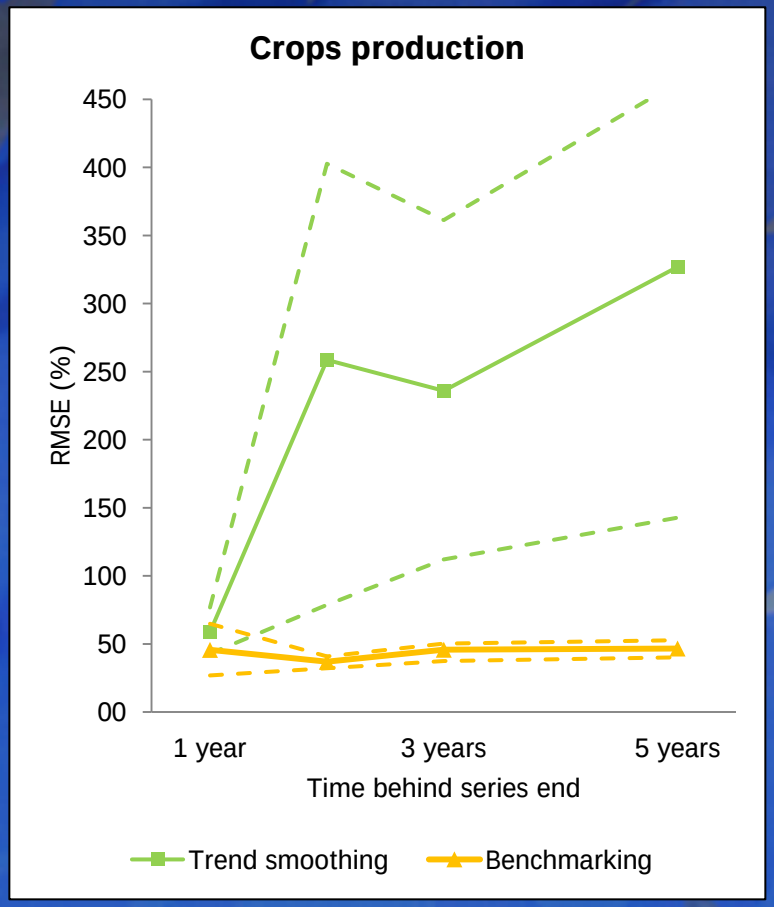
Results - Coverage



Scope: FAOSTAT data in the production domain, 1990-2009.

Results - Accuracy





Work proceeds …

- Review of imputed values during data compilation
- Discussion of the approach at conferences and workshops
- Further refinement
  - fit with other SUA elements
  - application in time-series gaps
  - broadening of the scope
- Development and testing of other approaches

Thank you very much for your attention!