Combining Soft Computing and Statistical Methods to Improve Data Analysis Solutions

IC0702

Start date: 01/04/2008

End date: 31/03/2012

Year: 2/3

Frank Klawonn

Vice Chair

University of Applied Sciences Braunschweig/Wolfenbüttel

Germany

COST is supported by
the EU RTD Framework Programme

ESF provides the
COST Office through an EC contract

Scientific context and objectives (1/2)

- Soft Computing, as an engineering science, and Statistics, as a classical branch of mathematics, emphasize different aspects of data analysis
- **Soft Computing**
  focuses on quickly obtaining working solutions that meet the needs arising in applications.
- **Statistics**
  focuses on establishing objective conclusions by rigorously analyzing all possible situations.
- **Objective of the Action**
  Improving the currently rather limited dialogue and interaction between these areas can lead to new and improved data analysis methods.

Scientific context and objectives (2/2)

Soft Computing Methods and Models

- often lack sound mathematical foundations,
- rarely make their underlying assumptions explicit, thus impeding a reliable transfer to new applications.
- are seldom checked rigorously and monitored w.r.t. performance and robustness,
- can rarely be generalized or easily transferred

• Statistical Methods and Models

- tend to focus on models, the mathematical properties of which are easy to analyze,
- constrain the set of eligible models, thus perhaps ruling out the most suitable or promising ones,
- can be difficult to understand or apply for a non-mathematician.

Working groups

1. Working Group A
   Model Selection and Validation
   - Statistical validation and monitoring of Soft Computing models
   - Model selection and validation for neural networks

2. Working Group B
   Bio-inspired Metaheuristics
   - Metaheuristics (evolutionary algorithms in particular) as Estimators
   - Estimation of distribution algorithms

3. Working Group C
   Statistics with Imperfect and Incomplete Data
   - Statistics with fuzzy data (estimation and regression)
   - Psychological versus statistical complexity

Results vs. Objectives

- Gather researchers from soft computing and statistics in working groups, stimulate exchange and discussion.
  (successfully achieved: organized several exchange events)
- Execute short term scientific missions, preferably between a statistics host and a soft computing visitor or vice versa.
  (successfully achieved: exceeded the projected number of STSMs)
- Organize training schools to disseminate statistical and soft computing knowledge and successful joint applications.
  (successfully achieved: summer course at IFSA/EUSFLAT 2009)
- Apply for new research projects in which statisticians and soft computing researchers work togetherand develop new data analysis methods.
  (still achieved yet: progress to this goal is slower than hoped for, but some progress is finally being made in the working groups)

Significant highlights in Science or Networking (1/2)

- Summer Course "**Soft Computing and Statistics**"
  July 17-19, 2009
  University of Lisbon, Portugal
  http://www.cost-ic0702.org/summercourse/
- Combined with Tutorials of IFSA/EUSFLAT 2010
  http://ifsa2009.ist.utl.pt/
  (one day out of a total of 3 days)
- Topics: Clustering, Feature Selection, Fuzzy Modelling, Fuzzy and Artificial Neural Networks, Robust and Fuzzy Statistics
- 11 high-level lecturers from 7 countries (including USA), > 30 participants (early stage researchers) from 10 countries on first two days, > 70 participants on last day (tutorials).

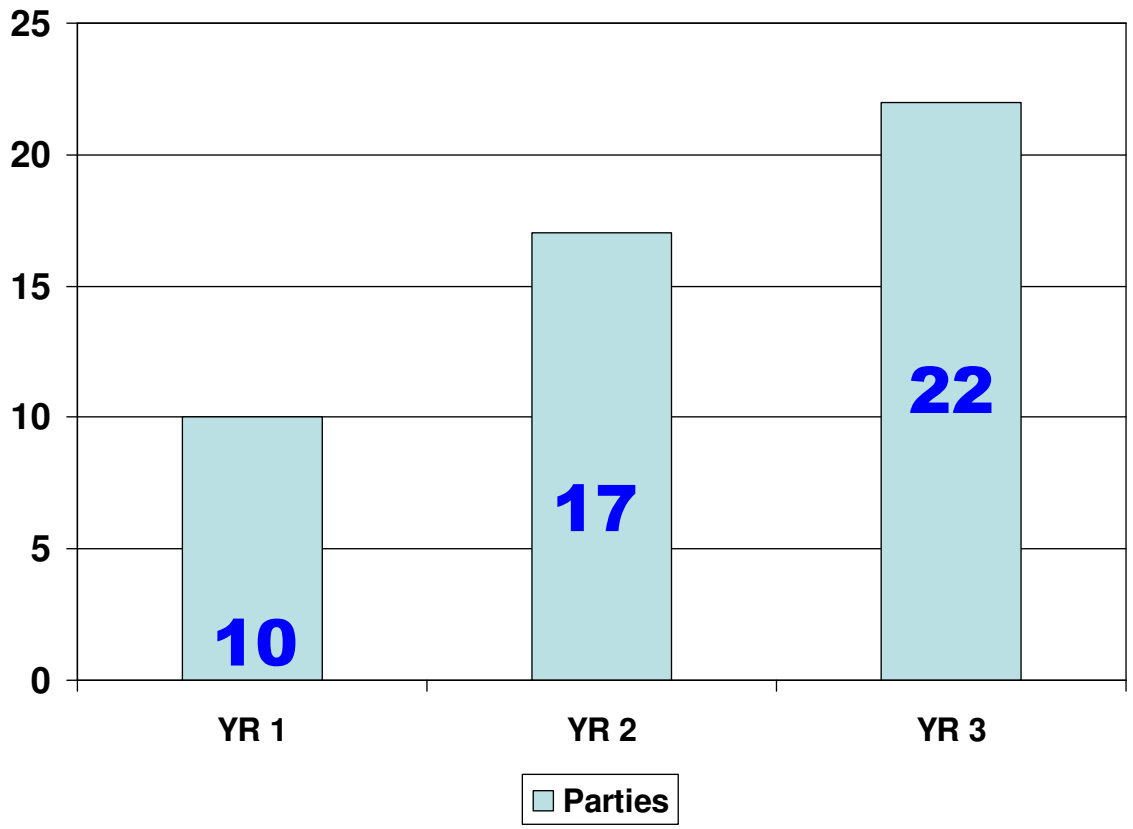
Significant highlights in Science or Networking (2/2)

- 18 short-term scientific mission (3 more than planned).
- Most of the STSMs lasted 3-4 weeks, some even longer.
- Topics covered applications of statistical and soft computing techniques in diverse areas, especially
  - statistics with fuzzy and incomplete data
  - temporal, spatial, and graph data
  - decision support, model evaluation, optimization
  - mining document collections, interest rates and option pricing etc.
- For already completed STSMs draft research papers are available and will be submitted in the course of this year.
- Noteworthy STSM: “A comprehensive PLS environment to Structural Equation Modeling“, Giorgio Russolillo
  (from Italy to France, 3.5 months, two draft publications)

Challenges

- We planned to achieve or at least get closer to the goal of compiling and submitting a research proposal in the area of soft computing and statistics.
- Progress towards this goal has been slower than hoped for, because the exchange about the research areas of the working group members took longer than expected.
- It is difficult to agree on a sufficiently narrow range of topics.
- Discussion groups have been formed in working group C, which work towards a project applications in subareas. The challenge is to actually reach a project proposal.
- Second core challenge: maintain and improve the success of the short-term scientific missions: at least as many STSMs next year as we had this year.

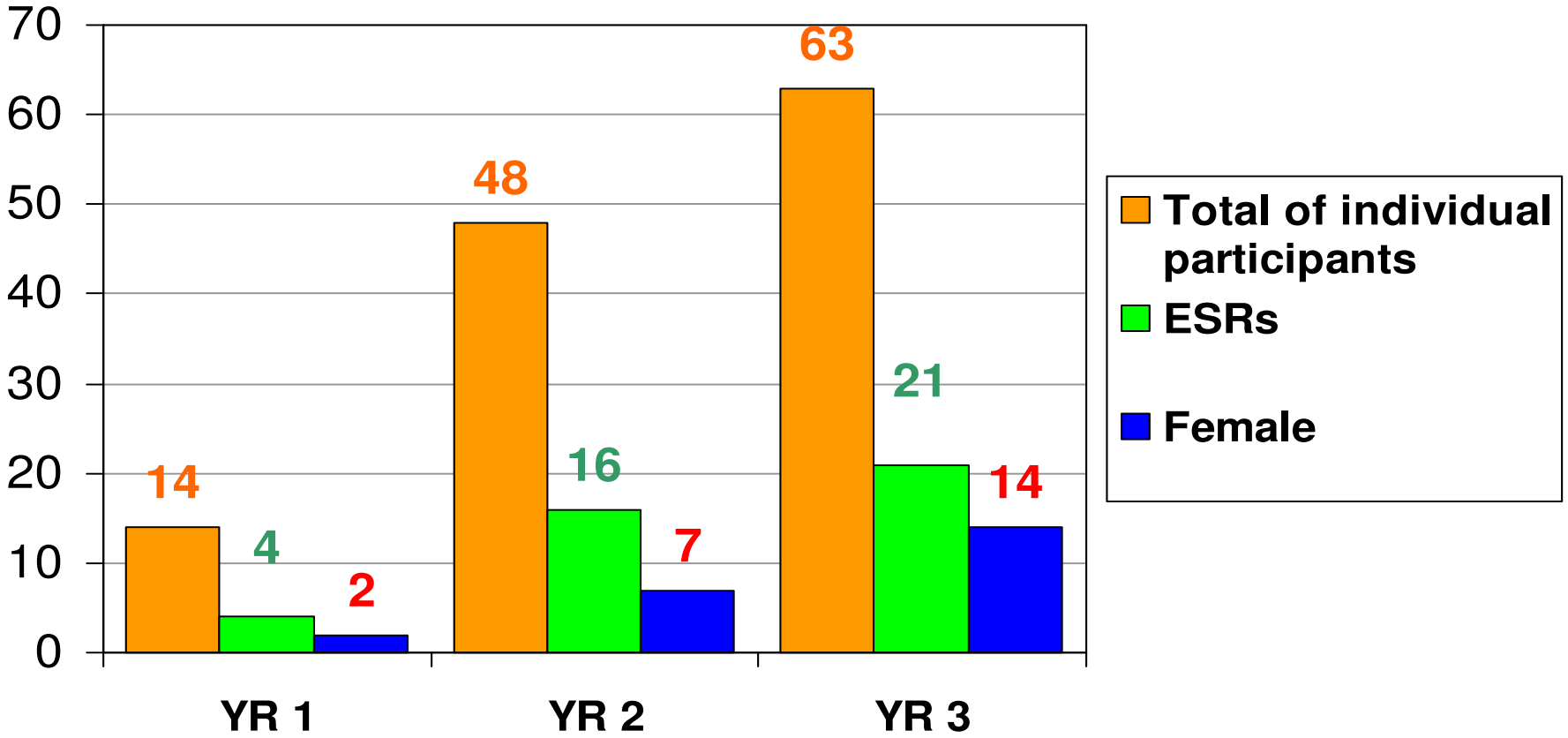
Action Parties



Grant Holder:

- European Centre for Soft Computing
- Raul del Coso
- Spain

Action participants



Use of COST instruments

	YR 1	YR 2	YR 3	YR 4
No. of MC / WG meetings	2 (3)	2	2	
No. of STSMs	9	18	15	
No. of workshops / conferences	1	1	1	
No. of joint publications	12	35	(60)	
No. of training schools	1	1	1	