



Abstract

The use of R as part of a large-scale information management and decision system

Joris DE WOLF, Koen BRUYNSEELS, Rindert PEERBOLTE and Willem BROEKAERT CropDesign, Belgium

UseR Conference Vienna 14-17 June 2006

Why did *CropDesign* choose **R** as the statistical environment for its *high-throughput plant evaluation* platform *TraitMill*?

We give our experience since 2001 with R as part of an industrial-scale data management and decision support system.

UseR Conference Vienna 14-17 June 2006



TraitMill: Greenhouse



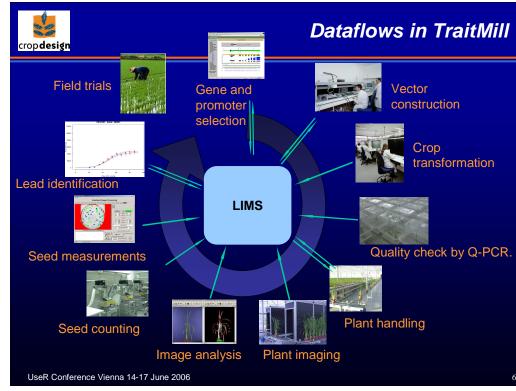
crop **design**

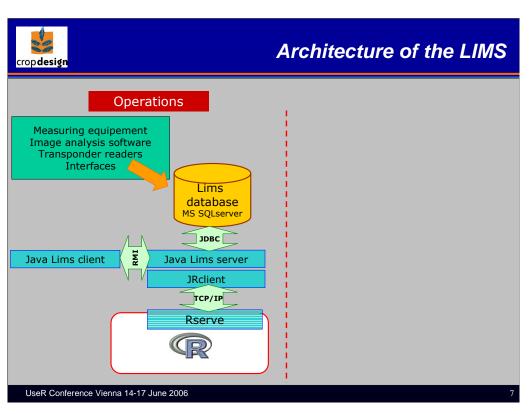
TraitMill: Plant imaging

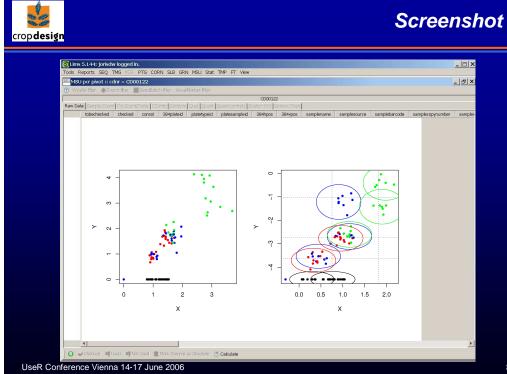


UseR Conference Vienna 14-17 June 2006



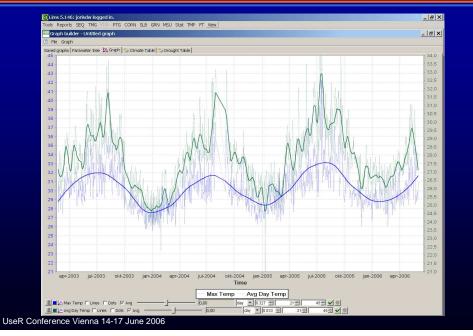


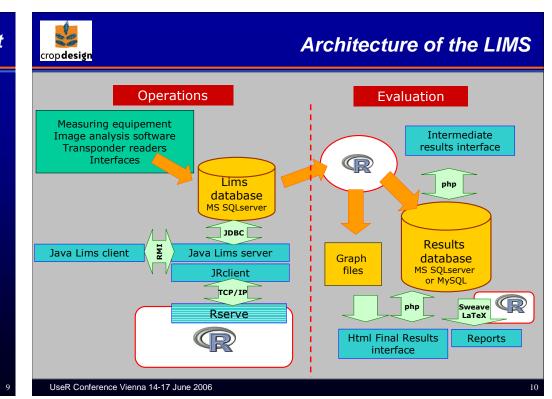






Screenshot





Screenshot crop **design** 😂 :: Cropdesign STAT :: - Mozilla Firefox ▼ 0 co 🔍 違 🕶 🍃 - 🥵 🔞 👔 🏫 🔝 http://stat.cropdesign.local/anova.php?IDexperiment=71798:sstatus=R8 allR allP client1 client2 STAT 👸 TUTOS 🥷 R Home > Single Experiment (Yield) > E0.000.007.576 (🏿 📾 🖽) 360607 - 050S0.000.397.196 A H R 🔟 🕠 OS4417-029A 17/18 75/95 360625 - 05050.000.397.214 A H R 🚳 🗔 🐷 0S4417-036A 18/18 75/95 360689 - 050S0.000.397.278 A H R 🗃 🗔 🐷 76/96 360722 - 050S0.000.397.311 A H R 🖾 🛅 • 0S4417-030A 76/96 360730 - 050S0.000.397.319 A H R 🗃 🗔 🐷 360739 - 050S0.000.397.328 A H R 🗃 🗔 🕟 🕟 OS4417-047A firstpan Height GNbfFlow GravityY AreaEmer AreaCycl RootDiam RootThick RootThin nrtotalseed CD15228 E0 000 007 576 - CD15228 AreaMay TR null Δ %Δ p-value 054417-037A 38735 38420 315 1 0.8956 eoo 48101 43073 5027 12 0.0385 43314 45955 -2642 -6 0.2729 40000 S4417-047A 49391 47902 1489 3 0.5356 Overall 44453 43744 709 2 0.4701 E0.000.007.576 - CD15228 EmerVigo UseR Conference Vienna 14-17 June 2006



Methods used

- General linear models
- Mixed linear models
- Generalised linear models
- Survival analysis
- Clustering
- Time series
- Graphical vizualisation

- base
- stats(4)
- graphics
- MASS
- nlme
- splines
- survival
- mvtnorm
- ellipse
- gtools
- cluster
- sweave
- ...

UseR Conference Vienna 14-17 June 2006



Reads Passwer 1 is Red 2

Set as Set of Se

Decision support



Main challenges

- Database connection
- Mimimize/optimize interactivity (400 dataset/week)
- Data preparation
- Data consistency checking (missing & odd values)
- Robustness of the methods
- Error capture
- Fall-back systems
- Standardize output
- Graphical vizualisation

UseR Conference Vienna 14-17 June 2006

1



Arguments for choosing R



Argument against choosing R

- Easy to integrate into the architecture of databases and interfaces
- © Various options of using the R engine

Decision Schedule for calling Validated Leads

- State-of-the-art statistical techniques
- © High quality and versatile graphical facilities
- Programming language
- Free, no problems with licenses
- Available for Linux

No interface for MS SQL server

UseR Conference Vienna 14-17 June 2006

UseR Conference Vienna 14-17 June 2006



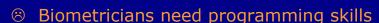
Positive experience with R



Negative experiences with R

- © Relatively easy programming language
- © Algorithmes and output can be easily stripped down and adapted to the needs
- © Quick assistance through a helpful user community
- Second Stable Stable

UseR Conference Vienna 14-17 June 2006



- Solution in the image of the standardized in the image of the standardized in the image of th
- Frequent updates not always compatible with old scripts
- Difficult to find help on old versions

UseR Conference Vienna 14-17 June 2006



Conclusion

R is a very valuable resource for this sort of application.

After 5 year of experience, we still think we made the right choice.