15th IFAC Symposium on System Identification
Saint-Malo, France, July 6-8, 2009 Pages 1016-1021

System Identification in Production Ecology: from theory to agroforestry practice

Karel J. Keesman*, Anil Graves, Wopke van der Werf***, Paul Burgess****

** Wageningen University, Systems & Control Group, P.O. Box 17, 6700 AA Wageningen, The Netherlands (Tel: +31 317 483780; e-mail: karel.keesman@wur.nl)*
*** Cranfield University, Silsoe, MK45 4DT, Bedfordshire, United Kingdom (e-mail: a.graves/p.burgess@cranfield.ac.uk)*
**** Wageningen University, Group Crop & Weed Ecology, P.O. Box 43, 6700 AA, Wageningen,, The Netherlands (e-mail:wopke.vanderwerf@wur.nl)*

Abstract: This paper introduces a system identification approach to agricultural ecosystems. In particular, the identification of an agroforestry system, combining trees with crops, is subject of study. Typically, for these systems $N < p$, where N is the number of data points and p the number of parameters in a (process-based) model. In this paper, we follow a constrained optimization approach, in which the constraints are found from literature or are given by experts. Given the limited *a priori* systems knowledge and very limited data sets, after decomposition of the parameter estimation problem and after model adaptation, we were able to produce an acceptable fit to validation data from a real-world agroforestry experiment.

1. INTRODUCTION

Silvo-arable agroforestry (AF) comprises widely-spaced trees intercropped with arable crops. Recent findings indicate that modern silvo-arable production systems are very efficient in terms of resource use, and could introduce an innovative agricultural production system that will be both environment-friendly and economically profitable. Growing high quality trees in association with arable crops in European fields may improve the sustainability of farming systems, diversify farmers' incomes, provide new products to the wood industry, and create novel landscapes of high value. The key question is then where in Europe to implement such a system and under which conditions is such a system profitable.

For an economic evaluation of an AF system that could be affected by seasonal decisions, the development of a bio-physical model linked to an economic model is indispensible. To allow the analysis of a full tree rotation, the bio-physical model should be of limited complexity. Hence, for analyses of full tree rotations a simple daily time-step biophysical model, called Yield-SAFE, has been developed (see van der Werf *et al.*, 2007; Keesman *et al.*, 2007).

For AF systems, mathematical modelling has become a major tool to increase the understanding of the underlying crop/tree growth mechanisms under light, water and nutrient competition, and to predict (long-term) yields of tree and crop for economical analyses (Graves *et al*., 2007). In addition to this, mathematical models of AF systems have been used to predict and assess the environmental effects of agroforestry at the landscape scale (Palma *et al.*, 2007a,b). However, the mathematical model can only be applied successfully if it is a proper description, in terms of model structure and model parameters, of the underlying process. Hence, theoretical modelling is most often not enough. There is also a need for identification of the model from experimental data. In this paper, identification basically comprises: calibration, validation and model adaption, which often results in an identification loop (see *e.g*. Ljung, 1987).

However, in solving the European AF allocation problem the identification exercise is strongly hampered by the limited availability of data. More specifically, we have only information from yield tables for trees, yield databases for crops and two experimental AF sites with only 12 years of data. Hence, in short the problem is: given the prior knowledge on AF systems, as represented by Yield-SAFE and expert's experiences, and given limited data sets, how to estimate the unknown model parameters and how in case of deficiencies to adapt the model structure.

The objective of the paper is to present a methodology for the identification of agricultural ecosystems from limited experimental data. As stated before, our focus is on agroforestry and the AF model Yield-SAFE (van der Werf *et al*., 2007) has been used as a starting point.

2. BACKGROUND

2.1 Summary of model

Yield-SAFE describes tree and crop growth in arable, forestry, and silvo-arable systems according to light and water availability. The model consists of seven state equations expressing the temporal dynamics of: (1) tree biomass; (2) tree leaf area; (3) number of shoots per tree; (4) crop biomass; (5) crop leaf area index; (6) heat sum, and (7) soil water content. The main outputs of the model are the growth dynamics and final yields of trees and crops. Daily inputs are temperature, radiation and precipitation. Planting densities, initial biomasses of tree and crop species, and soil parameters must be specified. Yield-SAFE contains 21

parameters, *i.e.* 6 tree parameters (per species), 9 crop parameters (per species) and 6 soil parameters (per location).

2.2 Experimental data

Potential yield, as a result of potential growth, is determined foremost by temperature (which drives developmental and phenological processes) and radiation (which drives photosynthesis) but is unaffected by water and nutrients as these are assumed to be non-limiting under the potential growth assumption (van Ittersum and Rabbinge, 1997). Potential yields for holm oak, stone pine and poplar in the Mediterranean region, wild cherry, walnut and poplar in the Atlantic region were obtained from experts and literature. These potential yields present just one single measurement.

Data for potential production of durum wheat, grain maize and sunflower were obtained from simulation results using the crop growth model STICS (Brisson *et al.*, 2003). Meteorological data from Montpellier for the period 1997-2002 were used. The same procedure, using STICS, has been followed to obtain potential yields for forage maize and winter wheat. For these two crop species, meteorological data from Wageningen for 1980-1984 were used.

Actual yields (i.e. locally attained yields; van Ittersum and Rabbinge, 1997) for each tree and crop species have been found from databases and from experts.

At experimental sites in the UK from 1992 onwards, the height of each tree in each arable treatment was measured after leaf fall. The diameters of the same trees were measured at breast height (1.3 m above the ground) each winter from 1994 onwards at the Leeds University site in Branham and from 1995 onwards at The Cranfield University site at Silsoe. Timber volume was estimated by first assuming the trunk is a perfect cylinder, with a volume calculated from height and diameter, and then multiplying this calculated volume by a form factor to account for taper of the trunk (Burgess *et al.*., 2004). The form factor was derived from poplar yield tables, given in Christie (1994). Each year, grain, bean or pea yield within each poplar-hybrid x arable-treatment plot was determined by harvesting with a plot combine. Corresponding measurements were also taken within the monocropped control area.

In addition to yield and meteorological data, plot management data has been collected as well. Typically, plot management includes choices with respect to the cropped area expressed as a proportion of the total silvo-arable area, crop sowing date (for each year in the tree cycle), tree stand density, proportion of trees thinned (time-dependent), proportion of tree biomass pruned (time-dependent) and proportion of tree shoots pruned (time-dependent). At last *a priori* parameter values and ecologically acceptable ranges have been specified.

2.3 Problem formulation

From the previous sections, we conclude that we have obtained an estimation problem with $N < p$, where N is the number of measurements and p the number of parameters. In the past, several solutions have been proposed, as *e.g.* the *minimum length* solution (see Menke, 1989; Lawson and Hanson, 1995), which for linear regression problem uses the right semi-inverse of the data matrix, application of regularization techniques (among which Tikhonov regularization is very popular, see *e.g.* Johanson, 1997) and constrained optimization (see *e.g.* Fletcher, 1980).

In the following, we will use a constrained optimization solution, where the constraints have been obtained from experts. The constraints on parameters are basically expressed in terms of individual parameter ranges.

3. METHODOLOGY

Generally, for the identification of dynamic systems the data set is split into a calibration set and a validation set to allow cross-validation. However, as mentioned before, in our case only a very limited set of data is available (see section 2.2). Hence, for this specific case in which we do not have any access to data from mature AF systems, we propose the following overall system identification loop, as presented in Figure 1 (see Appendix). Notice, in particular, the decompositions of the parameter estimation problem, which will be further illuminated in section 4.

In particular, the following procedure is implemented:

1. perform sensitivity analysis (SA) on both the tree and crop dynamics
2. estimate the most dominant parameters from monoculture data of potential growth of tree and crop
3. estimate additional parameters from monoculture data of actual growth of tree and crop
4. validate the models for tree and crop growth on monoculture data from different sites
5. validate the agroforestry model, which includes tree-crop interactions, using data from running AF system experiments.

First, prior to the calibration step, a sensitivity analysis is performed. This sensitivity analysis detects which parameters dominate the output behaviour of the system, in this case the crop and tree yields. Most often, only a few parameters appear to be dominant. Consequently, only the most dominant parameters are estimated from experimental data, while the others are just fixed at values from literature and given by experts (see also Ioslovich *et al.*, 2004). Hence, in our case the most dominant tree parameters have been estimated from potential yield data found in literature or found in databases. On the contrary, the most dominant crop parameters have been estimated from data generated by the widely accepted crop model STICS (Brisson *et al.*, 2003). In the next step, actual yield data, reflecting the effects of soil water and other management factors, have been used to modify the light and water related parameters for the tree and crop. In order to avoid unrealistic light parameters, as the light use efficiency, a management factor has been

introduced. This management factor is subsequently multiplied with the light use efficiency. Then, the models for tree and crop growth have been validated by data from different experimental sites with monocultures. Finally, the agroforestry model has been cross-validated by data from experimental sites in Leeds and Silsoe.

4. RESULTS

4.1 Sensitivity analysis

A first indication of which parameters dominate the yield predictions is obtained by applying a one-at-a-time sensitivity analysis. In this analysis we vary each parameter ±10% while fixing the other parameters at their nominal value, so that only two runs per parameter are needed. Let us denote the *j*th parameter by p_j with nominal value p_j^o. Given the 10% variation, the lower and upper values are defined as $\underline{p}_j := 0.9 p_j^o$ and $\overline{p}_j := 1.1 p_j^o$, respectively. If we define the corresponding yields as $\underline{y}_j := y(\underline{p}_j)$ and $\overline{y}_j := y(\overline{p}_j)$, the normalized sensitivity coefficient is given by

$$S_j = \frac{\overline{y}_j - \underline{y}_j}{\overline{p}_j - \underline{p}_j} \cdot \frac{p_j^o}{y(p_j^o)} \qquad (1)$$

On the basis of this measure, the sensitivity analysis revealed that for potential tree growth, that is neglecting the effects of water and management factors, the light use efficiency and the initial shoot number were most dominant. The most dominant parameters for potential crop growth are: light use efficiency, temperature sum to plant emergence, temperature sum at which partitioning to leaves starts to reduce, temperature sum at which partitioning to leaves is nihil and temperature sum till harvest.

4.2 Calibration using potential and actual yields

Tree volumes were calibrated by matching the predicted and potential volumes in the harvest year. Crop yields, on the other hand, were calibrated by matching the mean predicted and potential yields for the duration of the tree rotation.

The calibrated model was than used to predict tree and crop yields in "typical" arable, forestry, and silvo-arable systems for the landscape test sites in Spain, France, and the Netherlands using standard management practice. In this step, a single parameter, the transpiration coefficient, is adjusted such that output from the model over the duration of the tree component matched the actual or reference monoculture tree and crop yield for each network site.

4.3 Cross validation

Application of the parameter values in the simulation model raised a number of issues. In particular, it has been observed that predicted timber volumes for individual trees at low densities, especially at 50 trees ha^{-1} or less, are too high (i.e. > 150%) relative to those in forest stands. Thus, intercrop yields were thought to decline too rapidly. The procedure of adjusting the transpiration coefficient to calibrate predicted model output to actual crop and tree yields was found to result in an over-estimate of water-use, when actual crop or tree yields were low. The critical soil water tension value (pF_{crit}; *i.e.* log of the soil water potential in water cm) beyond which crop growth was predicted to decline, was wrongly assumed to be 2.3.

4.4 Model adaptation

Through observation of model responses to changes in the parameters, review of literature, and expert opinion, the following strategy was gradually developed for calibration of Yield-SAFE for actual yields.

With respect to the first observation, use of a lower light extinction coefficient for trees (k_t), although reducing final per tree volume and increasing intercrop yields at low densities, was found to produce incorrect timber increment profiles. In particular, the initial growth of the trees at low densities was well below that for trees at high densities, whereas evidence generally shows that volume increment of individual trees at low densities is at least as great as that at high densities.

The use of a phased light-extinction coefficient based on the leaf area (LA in m^2 $tree^{-1}$) of the tree, rather than a constant reduction, was implemented so that the coefficient was highest when tree leaf area was lowest, but lowest when tree leaf area was highest. The modified light-extinction coefficient is defined as

$$k_t = b + b\left(\frac{a}{LA + a}\right) \qquad (2)$$

which introduces two new empirical parameters *a* and *b*. Observation of the effect of adjusting these parameters on tree and crop yields and estimates of the light extinction coefficient suggested that the most suitable value for *a* was 10, and for *b* was 0.4.

The phased light-extinction coefficient ensured initial tree-volume increment at low tree densities was not superseded by initial tree-volume increment at high tree densities, whilst later tree-volume increment at low density was reduced, so that final tree volumes were lower and intercrop yields were higher. The effect of a phased light extinction coefficient compared with the constant light extinction coefficient, initially assumed, is shown for continuous wheat (Figure 2a) and cherry (50 trees ha^{-1}) (Figure 2b).

A more realistic value for the critical soil water potential (pF_{crit}) at which crop and tree growth was assumed to decline was given by using a value of 2.9 for crops and a value of 4.0 for trees which were assumed to be less sensitive to water stress than crops. Observations of results showed that a high critical soil water tension for crops increased crop yields (Figure 2c) and reduced tree growth (Figure 2d). A pF_{crit} of 2.9 for crops was chosen as this was the soil water tension for 50% water availability in a medium soil (Burgess *et al.*, 2004).

However, despite adjusting the structure of k_t and changing critical soil water potentials for trees and crops (pF_{crit}) to more realistic values, re-calibration of trees and crops, given fixed prior bounds on the transpiration coefficient, often led to model outputs that could not match actual tree and crop yields. Hence, at last, new bounds on transpiration coefficient values, depending on tree and crop species and location within the Atlantic and Mediterranean zones, and additional use of an harvest index and a "management factor" were introduced.

These three parameters (transpiration coefficient, harvest index and management factor) for trees and crops were adjusted within a priori defined bounds, so that the parameter values of the transpiration coefficient will stay within acceptable bounds. The bounds on the transpiration coefficient ensure that tree-crop competition for water was not excessively distorted.

The response of the model showed that increasing the water use requirements (represented by the parameter γ) of either the tree or the crop reduced both crop yields (Figure 2e) and tree growth (Figure 2f). The converse was also true. As the vapour pressure deficit in Mediterranean zones is generally greater than in Atlantic zones, water-use requirements were defined to be higher in the Mediterranean zone than in the Atlantic zone (van Keulen, pers. comm. 2004). The harvest index was adjusted within a pre-defined range for each crop according to the location and yield of the site. Existing literature suggests that harvest index varies between crops and for the same crop, tends to be low when yields are low and high when crop yields were high. The management factor was generally only used when adjusting the water-use requirements and the harvest index to the bounds could not be used to match predicted and actual yields.

After implementation of the model adjustment with respect to light extinction coefficient, critical pF values, transpiration coefficient, harvest index and management factor, a set of parameter values has been obtained.

The final step is now to validate the adjusted model structure with its parameter set on the basis of data from the experimental AF sites in Silsoe and Leeds. The calibrated model was then run to calculate growth trajectories and yields (under water limitation) for crops and trees within a silvo-arable system over a 30 year tree rotation. The predicted relative crop yields for the first twelve years (Figure 3) generally matched the experimental results. This match between data and simulation results in the agroforestry situation provides further evidence for the validity for the modelling concept and calibration philosophy. Recall that so far the model was not fitted to any data from the agroforestry stand, but only to data from pure stands of crops or trees. Thus, the rather good fit of the model to the yields in an actual agroforestry experiment provides evidence that it correctly captures the essence of the crop-tree interactions.

REFERENCES

Brisson, N., C. Gary, E. Justes, R. Roche, B. Mary, D. Ripoche, D. Zimmer, J. Sierra, P. Bertuzzi, P. Burger, F. Bussière, Y.M. Cabidoche, P. Cellier, P. Debaeke, J.P. Gaudillère, C. Hénault, F. Maraux, B. Seguin, H. Sinoquet (2003). An overview of the crop model STICS. *European Journal of Agronomy* **18**, 309-332.

Burgess, P.J., Graves, A.R., Metselaar, K., Stappers, R., Keesman, K., Palma, J, Mayus, M., van der Werf, W. (2004). Description of the Plot-SAFE Version 0.3. Unpublished document, 15 September 2004. Silsoe, Bedfordshire: Cranfield University, 52 pp. Available http://montpellier.inra.fr/safe/. (May 5 2005).

Christie, J.M. (1994) Provisional yield tables for poplar in Britain. Forestry Commission Technical Paper 6. Forestry Commission, Edinburgh, 36 pp.

Graves, A.R., P.J. Burgess, J.H.N. Palma, F. Herzog, G. Moreno, M. Bertomeu, C. Dupraz, F. Liagre, K. Keesman, W. van der Werf , A. Koeffeman de Nooy, J.P. van den Briel (2007). Development and application of bio-economic modelling to compare silvoarable, arable, and forestry systems in three European countries. *Ecological Engineering* **29(4)**, 434-449.

Fletcher, R. (1980). *Practical Methods of Optimization*. Vol 1: Unconstrained Optimization. John Wiley & Sons.

Ioslovich, I., P-O Gutman and I. Seginer (2004). Dominant parameter selection in the marginally identifiable case. *Math. Comp. Simul.*, **65(1-2)**, 127-136.

Johansen, T.A. (1997) On Tikhonov regularization, bias and variance in nonlinear system identification. Automatica **33**, 441-446.

Keesman, K.J., W. van der Werf and H. van Keulen (2007). Mathematical Production Ecology: Analysis of a Silvo-arable Agro-forestry System. *Math. Biosci*. **209(2)**, 608-623.

Lawson, C.L. and R.J. Hanson (1995) *Solving Least Squares Problems*. SIAM, pp.337.

Ljung, L. (1987). *System Identification: Theory for the User,* 2nd ed. Prentice Hall, New Jersey, 609 pp.

Menke, W. (1989) Geophysical Data Analysis: Discrete Inverse Theory. Academic Press, pp 289.

Palma J H N, Graves A R, Bunce R G H, Burgess P J, de Filippi R, Keesman K J, van Keulen H, Liagre F, Mayus M and Moreno G. (2007a) Modeling environmental benefits of silvoarable agroforestry in Europe. *Agriculture, Ecosystems & Environment* **119**, 320-334.

Palma, J.H.N., A.R. Graves, P.J. Burgess, K.J. Keesman, H. van Keulen, M. Mayus, Y. Reisner and F. Herzog (2007b) Methodological approach for the assessment of environmental effects of agroforestry at the landscape scale. *Ecol. Eng*., **29(4)**, 450-462.

Van Ittersum, M.K. and R. Rabbinge (1997). Concepts in production ecology for analysis and quantification of agricultural input-output combinations. *Field Crops Research* **52**, 197-208.

van der Werf, W., K.J. Keesman, P. Burgess, A. Graves, D. Pilbeam, L.D. Incoll, K. Metselaar, M. Mayus, R. Stappers, H. van Keulen, J. Palma & C. Dupraz (2007). Yield-SAFE: a parameter-sparse process-based dynamic model for predicting resource capture, growth and production in agroforestry systems. *Ecological Engineering* **29(4)**, 419-433.

Appendix A. Figures

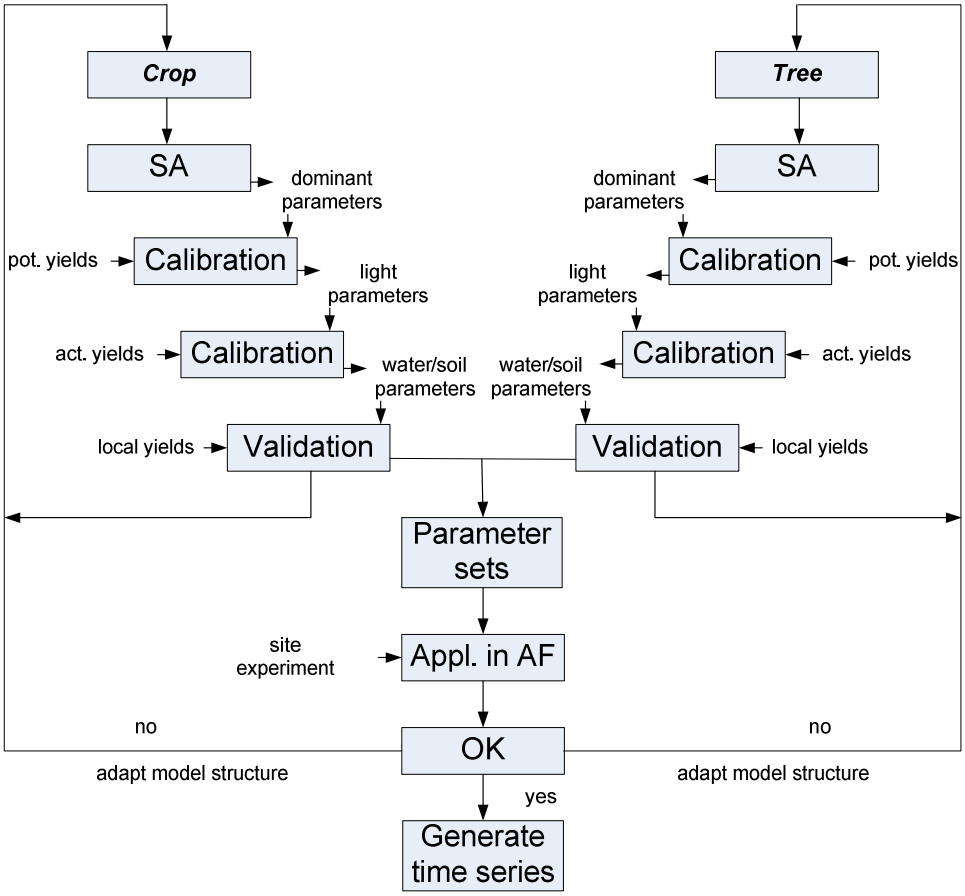
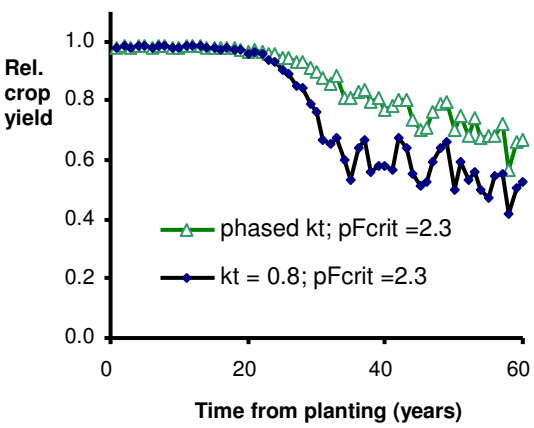
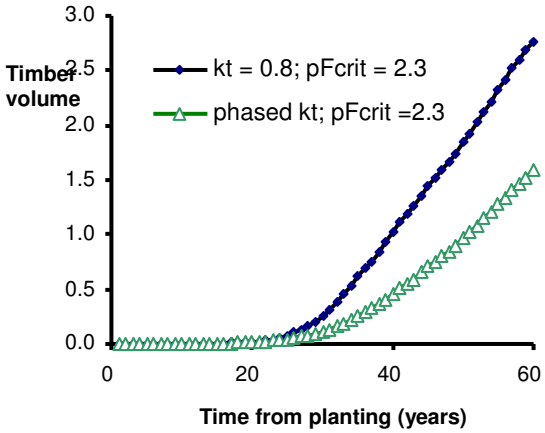


Fig. 1 System identification loop for AF systems containing: sensitivity analysis (SA), calibration, (cross) validation and revision steps, with a restart of the cycle if validation using experts' knowledge and/or (agroforestry) site experiments yields results that are not satisfactory.

a) Effect of phased k_t on relative crop yield



b) Effect of phased k_t on tree yield



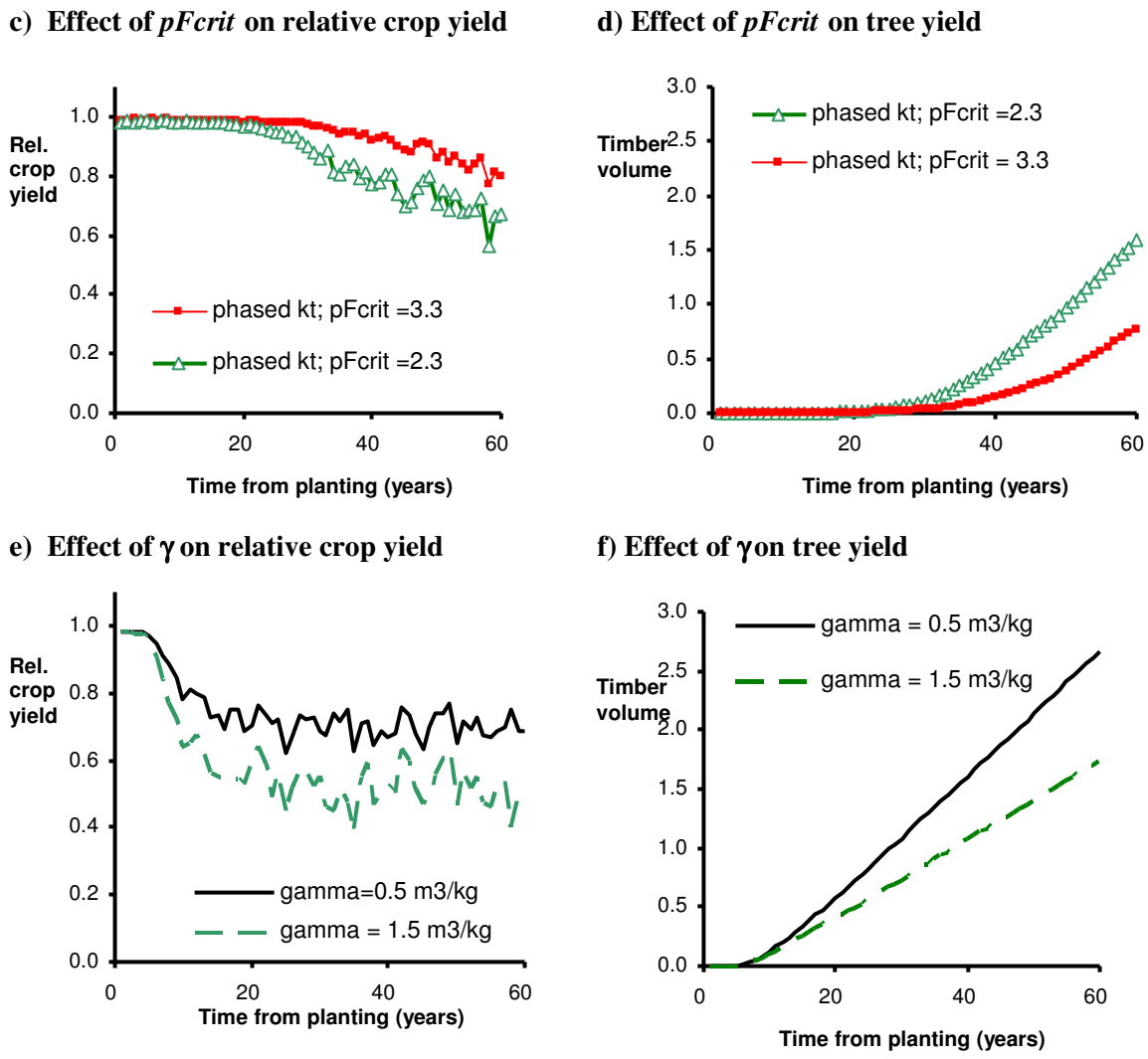


Fig. 2 Comparison of a) relative wheat yields (-) and b) tree yields (m^3 $tree^{-1}$) at a constant k_t (0.8) and phased k_t (Eqn. 2) for wild cherry and a crop pF critical value of 2.3, c) relative crop yields and d) tree yields for two different pF critical values, e) relative crop yields and f) tree yields for two different values of γ.

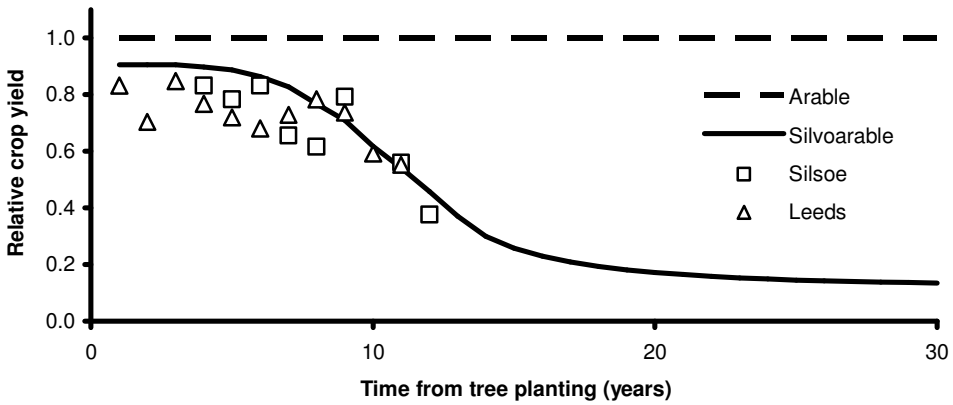


Fig 3 Validation of Yield-SAFE: model prediction of relative yield of continuous winter wheat, compared with monoculture wheat yield, in a poplar agroforestry stand (156 trees ha^{-1}), compared to observed relative crop yields in Silsoe and Leeds agroforestry experiments, 1992-2004 (open symbols).



https://dspace.org/

System identification in production ecology: from theory to agroforestry practice

Keesman, K. J.; Graves, Anil; van der Werf, Wopke; Burgess, Paul J.

2009-07-08T00:00:00Z
http://dspace.lib.cranfield.ac.uk/handle/1826/5455

Downloaded from a DSpace repository. https://dspace.org/