An indicator of solar radiation model performance based on a fuzzy expert system

Gianni Bellocchi*, Marco Acutis, Gianni Fila, and Marcello Donatelli

G. Bellocchi, G. Fila, and M. Donatelli, Research Institute for Industrial Crops, via di Corticella 133, 40128 Bologna, Italy; and M. Acutis, Department of Crop Science, via Celoria 2, 20133 Milan, Italy. *Corresponding author (g.bellocchi@isci.it).

ACKNOWLEDGEMENTS

The structure of the indicator was basically inspired by the fuzzy-based indicator of pesticide environment impact by van der Werf and Zimmer (1998), adapted with novel ideas and concepts to conform to the purpose of the paper.

The meteorological data sets used in this study are part of a large data base provided from several partners and used in previous studies. The list of people and institutions who contributed by providing data is too large to be reported here, and it can be found in the Acknowledgements section of the manual of the software RadEst3.00.

Research under the auspices of the Italian Ministry of Agricultural and Forestry Policies, project SIPEAA (http://www.sipeaa.it), paper no. 3.

An indicator of solar radiation model performance based on a fuzzy expert system

ABSTRACT

When evaluating models, various indices or test statistics are computed, quantifying the magnitude of model residuals, the correlation between estimates and measurements, patterns of residuals over external variables, etc. Such indices are variously related to each other, thus making model comparison difficult. Problems of this type emerge when testing solar radiation models. This paper proposes a fuzzy expert system to calculate a modular indicator, "I_{rad}", which reflects an expert perception about the quality of solar radiation models performance. Three modules were formulated reflecting, respectively, the magnitude of residuals ("Accuracy"), the correlation estimates and measurements ("Correlation"), and the presence/absence of patterns in the residuals against independent variables ("Pattern"). The modules "Accuracy" and "Pattern" resulted from the aggregation of three (relative root mean square error, modeling efficiency, t-Student probability) and two (pattern index versus day of the year and pattern index versus minimum air temperature) indices respectively, while the module "Correlation" was identified by a single index (Pearson's correlation coefficient). For each index two functions describing membership to the fuzzy subsets Favorable (F) and Unfavorable (U) have been defined. The expert system calculates the modules according to both the degree of membership of the indices to the subsets F and U, and a set of decision rules. Then the modules are aggregated into the indicator I_{rad}. Sensitivity analysis is presented, along with module and I_{rad} scores for some application cases.

Abbreviations: RMSE, root mean square error; RRMSE, relative root mean square error; EF, efficiency; r, correlation coefficient; PI, pattern index; I_{rad}, indicator of solar radiation model performance; BC, model Bristow-Campbell; CD, model Campbell-Donatelli; DB, model Donatelli-Bellocchi.

In the process of testing model performance, several indices and test statistics are commonly used (e.g., Smith et al., 1997; Martorana and Bellocchi, 1999; Metselaar, 1999; Yang et al., 2000). Some of them quantify the departure of the model response from experimental measurements (model residuals), while others focus on the correlation between model estimates and measurements. Other indices have been proposed recently to assess systematic behaviors of model residuals against model inputs or other independent variables (Donatelli et al., 2000, 2002a). The interpretation of these statistics is in itself essentially descriptive, and primarily based on scientific background, rather than on statistical significance (Willmott, 1982). Some users may appraise the degree of reliability of model outputs simply from the model's graphical display, while others may require more in-depth evaluation. Indeed, there is some evidence to suggest that the manner in which a study is performed (i.e., the extent to which the simulation study meets a user's expectations) is more important in forming a user's quality perception, than the results of a statistical evaluation (Robinson, 1998).

Each statistic allows only a partial insight into the model performance. Therefore, giving a solid quality judgment about model results one requires to simultaneously consider several statistics. Balancing different aspects involved in model testing, such as departure of estimates from measurements, modeling efficiency (i.e., better fit than the average of measurements), correlation measures, presence/absence of systematic behavior in the residuals, etc. is often complicated by the fact that different statistics may provide contrasting results. Hence, combining several statistics into one aggregated measure is desirable in order to have a comprehensive assessment of the model response. This would be helpful either when judging one model response, when evaluating its performance in a variety of conditions, or when one is requested to choose the best model out of a list of candidates.

Prior to aggregating statistics, the user must describe the constraints needed to evaluate the model (Bardossy et al., 1985). Generally, this is an internal process that is not typically well defined or thoroughly documented. Indeed, it varies from individual to individual according to a number of factors, which in fact influence and characterize each individual (Dubois and Prade, 1980). Personal preferences and intentions may strongly influence a user's judgment. To capture the knowledge and preferences of the user, weights may be used to establish the relative importance of complementary statistics. The weights are determined by the user and may change for different users. These subjective aspects cannot be completely removed from the evaluation process, but they may be proficiently captured and described in mathematical terms (Jones and Barnes, 2000). This could be achieved by aggregating indices by summation, multiplication, or a combination of both. These approaches pose mathematical and conceptual problems (Keeney and Raiffa, 1993), since evaluation statistics differ in their nature, dimensions and range of possible values.

Considering the inadequacy of such methods, a different approach to evaluate model performance can rely in setting up a fuzzy expert system (Hall and Kandel, 1991) using decision rules (fuzzification). This technique is robust on uncertain and imprecise data such as subjective judgments, and allows the aggregation of dissimilar measures in a consistent and reproducible way (Bouchon-Meunier, 1993).

Daily solar radiation is an important meteorological measure for various analyses in many applied sciences, including agricultural engineering, crop physiology, ecology, hydrology, meteorology, physics, and soil science (Brutsaert, 1982; Iqbal, 1983; Tracy et al., 1983). In spite of that, solar radiation is measured infrequently, and the access to reliable radiation data is limited (e.g., Thornton and Running, 1999). As a consequence, the need for accurate estimates of solar radiation exists.

A model for estimating global radiation based on daily temperature extremes was proposed by Bristow and Campbell (1984) using data from three stations in the northwestern U.S. Improved models were successively developed by Donatelli and Marletto (1994), Donatelli and Campbell (1998), Bechini et al. (2000) and, more recently, by Donatelli and Bellocchi (2000). Evaluating and comparing the models Bristow-Campbell (BC), Campbell-Donatelli

(CD) and Donatelli-Bellocchi (DB) against data sets from a wide range of latitudes was a complex process (Donatelli and Bellocchi, 2000) because of the number of statistical indices used. To overcome such a problem, in this paper we describe an expert system to calculate an indicator of model performance, "I_{rad}", which reflects an expert perception of the performance of radiation models.

METHODS

The structure of the indicator

It is assumed here that a comprehensive assessment about radiation model performance should consider: i) the ability of the model to produce small residuals, ii) the extent to which the model estimates are correlated with measurements, iii) the ability of the model to produce residuals uniformly distributed over the range of two relevant independent variables (day of the year, minimum air temperature).

According to such premises, we defined three indicator modules, named "Accuracy", "Correlation", and "Pattern". The value of each module depends on one or more indices (Table 1) and a set of decision rules. For each module, a dimensionless value between 0 (best model response) and 1 (worst model response) is calculated. The procedure, based on the multi-valued fuzzy set theory introduced by Zadeh (1965), follows the so-called Sugeno or Takagi-Sugeno-Kang method of fuzzy inference (Sugeno, 1985). This approach is computationally efficient and well-suited for mathematical analysis. It has been applied to a wide variety of problems, such as the design of an indicator for assessing environmental impact of pesticides (van der Werf and Zimmer, 1998), and the development of novel approaches to support decisions regarding sustainable development (Cornelissen et al., 2001).

Three membership classes (or subsets) were basically defined for all indices given in Table 1, according to an expert judgment: Favorable (F), Unfavorable (U), and partial (or fuzzy) membership. Several indices were aggregated into modules, and the modules in the final indicator, using fuzzy-based logic rules. The procedure is explained in detail in Appendix A.

The module "Accuracy"

The composition of the module "Accuracy" was based essentially on the suggestions of Yang et al. (2000). These authors found that a sound conclusion on model accuracy may be drawn using an index of the amount of residuals, (e.g., RMSE: root mean square error) (Fox, 1981), a measure of modeling efficiency (e.g., EF: efficiency) (Loague and Green, 1991), and a two-tailed paired t-test:

$$\mathrm{RMSE} = \left[\frac{\sum_{i=1}^{n} (D_i)^2}{n} \right]^{0.5} \qquad (1)$$

$$\mathrm{EF} = 1 - \frac{\sum_{i=1}^{n} (D_i)^2}{\sum_{i=1}^{n} (M_i - \overline{M})^2} \qquad (2)$$

$$t = \frac{\overline{D}}{s_D} \qquad (3)$$

where D_i is the difference E_i-M_i, E_i is the i-th estimated value, M_i is the i-th measured value, n is the number of pairs E_i/M_i, $\overline{M}$ is the average value of all measured values, $\overline{D}$ is the average of all differences D_i, and s_D is the standard error of the differences D_i. The computed t is compared against the critical t with 2·(n-1) degrees of freedom.

In place of RMSE, we considered it more appropriate to use a relative measure, the RRMSE (relative root mean square error), where:

$$\mathrm{RRMSE} = 100 \cdot \frac{\mathrm{RMSE}}{\overline{M}} \qquad (4)$$

The index RRMSE may vary from 0 to positive infinity. The smaller RRMSE, the better the model performance. RRMSE is a dimensionless index, allowing comparison among different model responses, regardless of units and range of values. Problems could arise because RRMSE tends to become unstable when $\overline{M}$ is close to zero, and is undefined when $\overline{M}$=0. However, this problem does not occur when computations are made on daily solar radiation data. The limit to the fuzzy subset F for this index was set equal to 20 (RMSE≤20 is F), while the limit to the subset U was established equal to 40 (RMSE≥40 is U). Both limits come from our expert judgment, working on multi-year weather data sets from about 200 locations (Donatelli and Bellocchi, 2000).

The index EF is very informative because it allows the immediate identification of inefficient models. EF is upper-bounded by 1, and it can assume negative values (lower-bounded at negative infinity). Negative values of EF indicate that the average value of all measured values is a better predictor than the model used. When estimating daily solar radiation, the limit for the subset U, EF=0.40 and the one for the subset F, EF=0.90, were chosen (EF≤0.40 is U, EF≥0.90 is F). Again, both limits are derived from our extensive experience, covering radiation estimates over a wide range of latitudes (Donatelli and Bellocchi, 2000).

In testing numerical estimates against measured data, the paired t-statistic is used to test the null hypothesis "average residual equal to zero" at a given probability level (statistical significance). If significant, the t-statistic shows that all the differences between estimated and measured values can not be attributed to experimental error. If not significant, the t-test can not prove that the estimated and measured values are identical, but it does indicate that there is no statistically significant reason for rejecting the hypothesis that the two outputs represent the same response. When the paired t-statistic is used to test the difference between estimated and measured data, the data set of D_i should be normally distributed with a null hypothesis "average D_i equal to zero". In our research all D_i data sets used were characterized by a roughly normal distribution. Generally, low t-values indicate a satisfactory response; however, under certain conditions they may not adequately characterize significant departure of simulation estimates from measured data. These conditions occur with high values of s_D. In such cases, low t values may be obtained with large departures of estimates from measurements and, consequently, the t-test may be unsuitable for evaluation purposes. Although cases of this type can theoretically occur in the evaluation process, they were not experienced with the data sets used in the present work. The paired t-test is handled here giving its significance level, i.e. P(t). Since P(t) represents the probability of t under random differences between pairs, then the best value for P(t) is 1, with 0 being the worst value. The selection of a particular statistical significance is problematic. Popular values are 0.01, 0.05, 0.10. The probability level of 0.05 is almost universally accepted as the threshold to reject a null hypothesis, therefore P(t)=0.05 was set as limit for U. Some scientists have used a higher level of statistical significance, e.g. 0.10 (Kedzie, 1997; Kleijnen et al., 1998; CTAP, 2000), because they were concerned about the lack of power in their test. Therefore, we set P(t)=0.10 as the limit for F, attributing a response in between P(t)=0.05 and P(t)=0.10 to a transition interval.

The value of the module "Accuracy" was calculated from the input indices according to eight decision rules, as summarized in Table 2. The expert reasoning runs as follows: if all indices are F, the value of the module is 0 (identity of estimates and measurements); if all indices are U, the value of the module is 1. In setting up the decision rules for the other combinations we had to decide on the relative importance of each index. In our experience, RRMSE and EF assume a relevant importance in the evaluation process, thus a fairly large weight (0.80) was attributed to the rule when both RRMSE and EF are U. The condition "$\overline{D}=0$" is a sound requisite but scarcely informative by itself, and may hide model outputs drifting towards large biases. Thus, the weight is low (0.20) when P(t) only is U.

The module "Correlation"

The value of the module "Correlation" depends on a single basic index, that is the correlation coefficient r (Addiscott and Whitmore, 1987), derived from the Pearson's simple linear correlation coefficient:

$$r = \frac{\sum_{i=1}^{n}(E_i - \overline{E}) \cdot (M_i - \overline{M})}{\left[\sum_{i=1}^{n}(E_i - \overline{E})^2 \cdot \sum_{i=1}^{n}(M_i - \overline{M})^2\right]^{0.5}} \qquad (5)$$

where $\overline{E}$ is the average of estimates. The coefficient r may vary from –1 (full negative correlation) to 1 (full positive correlation). The closer r is to 1, the better the model. Besides the indices based on differences, the coefficient of correlation r between estimates and measurements is commonly computed. The use of this index is questioned (e.g., Willmott, 1982), since its value is not related to the accuracy of estimate. However, the index r is a universal measure with multiple interpretations. For instance, Cahan (1987) looks at r as a measure of identity between standardized values. Moreover, the value of r may help recognise the fluctuation of the estimates among the n measurements (Kobayashi and Salam, 2000). For these reasons, the index r is generally still regarded as a useful measure of model performance.

The membership limits attributed here to the correlation coefficient come from the general categorization made by Hinkle et al. (1994), who took correlation coefficients equal to or greater than 0.90 as very high correlations (limit for F: r=0.90), and coefficients equal to or lower than 0.70 as moderate and little correlations (limit for U: r=0.70). Such limits do conform to our expert judgment.

It must be pointed out that statistical significance for correlation coefficients does not always imply practical significance. The limits attributed here are mere descriptors for the practical interpretation of correlation coefficients, and do not take into account statistical significance. The latter depends on the number of data points and can be verified by a t-test, provided that both estimated and measured series do conform to the assumptions required for the appropriate application of the test.

Given that there is only one index in the module, the computation of “Correlation” is simplified to two decision rules: if r is F then 0, if r is U then 1.

The module “Pattern”

The module “Pattern” accounts for two relevant independent variables in radiation models, the day of the year (DOY) and the daily minimum air temperature (Tmin). For the computation of pattern indices, the range of such independent variables is divided into four sub-ranges (quartiles), thus producing four groups of residuals. Pattern indices (PI) are based on the pairwise differences between average residuals of each quartile (Donatelli et al., 2000):

$$PI = \max_{l,m=1,\ldots,4;l\neq m}\left|\frac{1}{q_l}\cdot\sum_{i_l=1}^{q_l}R_{i_l} - \frac{1}{q_m}\cdot\sum_{i_m=1}^{q_m}R_{i_m}\right| \qquad (6)$$

where R is the model residual, l and m indicate two groups being compared, q_l and q_m represent the number of residuals in the groups, i_l and i_m identify each value of residuals in the groups. PI values have the same units as the variable under study (in this case MJ m^{-2} d^{-1}).

The pattern indices are targeted at pointing out macro-patterns in the residuals (Donatelli et al., 2002a). The presence of patterns usually means that the residuals contain structure that is not accounted for in the model. When applied to different types of plots of the residuals, pattern indices may provide meaningful information on the adequacy of different aspects of the model, such as lack of inputs, poor parameterization, etc.; therefore they should integrate difference-based and correlation-based indices when evaluating model performance.

We refer here to pattern indices computed against DOY and Tmin because daily radiation model residuals often show non-random distribution of residuals over the range of such variables, and some of the models commonly used for estimating radiation include parameters specifically devoted to account for patterns of this type.

The limits attributed to both pattern indices reflect the authors’ experience. Values of PI are considered F when equal to or lower than 1.0 MJ m^{-2} d^{-1}, and U when equal to or higher than 2.5 MJ m^{-2} d^{-1}. Examples of pattern indices computed with different models at tropical (Patos de Minas, Brazil) and temperate (Würzburg, Germany) sites in selected years are shown in Figure 1 and Figure 2, respectively.

The value of the module “Pattern” depends on the input indices according to decision rules summarized in Table 3. The same weight was attributed to PI_{doy} and PI_{Tmin}. If all indices are F, the value of the module is 0 (i.e., no pattern); if all indices are U, the expert weight is 1; if one index is F and the other is U, the weight is 0.5.

Aggregation of the modules

The three modules described above can be used to compare different radiation models. The modules might be ranked, for instance, by means of a multi-criteria analysis technique, using the modules as evaluation criteria. An alternative approach is to aggregate the three modules (second-level aggregation) in some way into an overall indicator (I_{rad}), reflecting a global judgment about model performance, again on a 0 to 1 scale. This can be done by summation, multiplication or a combination of both, according to aggregation schemes. We propose an aggregation of the three modules, which uses decision rules, as described above for the aggregation of indices into the modules.

The value of the indicator I_{rad} depends on the modules “Accuracy”, “Correlation”, and “Pattern” according to a set of eight decision rules (Table 4). The definition of the limits of the transition interval is the same for the three modules: we assigned complete membership to F if the value of the module is 0 (i.e., identity of estimates and measurements, unit correlation, no pattern of residuals versus independent variables) and complete membership to U if the value of the module is 1. In setting up the other decision rules we had to establish the relative importance of each module. As a general rule in model evaluation more emphasis is given to the amount of residuals, whereas the correlation of estimates versus measurements carries less weight. Because of their recent development, pattern indices are rarely used in model evaluation. Based on our experience, decreasing importance was given to the modules “Accuracy”, “Pattern”, and “Correlation” respectively. Thus, for instance, if “Accuracy” is U, then the weight is 0.55. If both “Accuracy” and “Pattern” are U, then the expert conclusion is 0.85. If both “Accuracy” and “Correlation” are U, then the conclusion is 0.70.

The relative incidence of each index on the indicator can be deduced by combining the weights of the indices into their own module with the ones of the modules into the indicator (Table 5).

RESULTS

Sensitivity analysis to input indices

In order to illustrate the functioning of the system, we present first the sensitivity showed by the indicator I_{rad} to variation of input values. Each input variable was varied over its transition interval, while the others were kept fixed either at the extremes of the transition interval, i.e. at U (Figure 3, top) and F (Figure 3, bottom), or at the median value (Figure 3, middle). The sensitivity analysis reflects the functioning of the system and provides some indication about the importance of each input index on the value of I_{rad}. However, one should be aware that the effect of the variation of a particular index over its transition interval on the value of I_{rad} is dependent on the value of the other indices. Therefore, results presented here should not be considered other than illustrations of the functioning of the system.

The extent to which indices affect the value of I_{rad} can be deduced from the decision rules involved in the process of aggregating indices in the modules, and the modules in the indicator. For instance, the input P(t) is very influential when all other inputs are U (Figure 3, top), while its effect is much smaller when the other inputs are F (Figure 3, bottom). This is the result of the mode of aggregation we adopted, which gives a lot of weight (0.8) to the rule in the module "Accuracy" when P(t) is F and both RRMSE and EF are U.

RRMSE tends to be more influential than the pattern indices when all other indices are U (Figure 3, top) as a consequence of the different weight attributed to these indices when they are F and the rest is U in the respective module (0.6 and 0.5 respectively). The opposite occurs when all other indices are kept F (Figure 3, bottom).

The influence of r when other inputs are U is somewhat large (Figure 3, top), due to the noticeable weight (0.85) attributed to the rule when the module "Correlation" is F and both other modules are U. Correlation coefficient and pattern indices exert the same incidence on the indicator (Table 5), thus their curves are somewhat complementary (Figure 3, top, middle, bottom).

Example applications

Another illustration of the functioning of the system is given by the computation of either the basic indices or the three modules and the I_{rad} indicator over yearly data sets of radiation data. Selected locations and years with dissimilar response in the various evaluation indices were used for the computations (Table 6).

Estimates of daily radiation were made using three models based on daily temperature extremes: BC (Bristow and Campbell, 1984), CD (Donatelli and Campbell, 1998), DB (Donatelli and Bellocchi, 2000). The models are described in Appendix B. Parameter values were determined at each location (except the parameter c of BC that was kept equal to 2, see Appendix B) using a multi-year calibration data set, through iterative steps aimed at minimizing both RRMSE and pattern indices. This procedure is consistent with the hypothesis underlying model formulations, for which model parameters are specifically devoted to reduce the magnitude of residuals (b in all models) or the presence of patterns (Tnc in CD, c1 and c2 in DB). The calibration procedure is described in detail in the documentation accompanying the software RadEst3.00 (Donatelli and Bellocchi, 2001; Donatelli et al., 2002b). Both software and documentation are freely downloadable via http from: http://www.isci.it/tools.

Calibrated model parameters for the locations and years selected are given in Table 7. The results are reported in Table 8 (evaluation indices) and Table 9 (modules and indicator). The outputs of Table 9 can be used to rank different radiation models with respect to the value of one or more modules or of the I_{rad} indicator.

The model DB gave the lowest value of the indicator (I_{rad}=0.0044 at Würburg), while BC gave the worst one (I_{rad}=0.5518 at Patos de Minas). The model DB gave the best response at six locations (Longreach, Los Baños, Matsumoto, Perugia, Sadore, Würzburg), CD was the best at three locations (Port Elizabeth, Patos de Minas, Smolensk), while BC was superior in one case (Poza Rica). In the sites where DB gave the best performance, the result was essentially due to the major ability of this model to deal with patterns (Table 8). The response associated to the model CD was more complex with reference to the contribution of each module to the indicator. In the estimates made at Patos de Minas we see that CD provided the best performance (I_{rad}=0.4425) due to the relative power of this model to keep small residuals ("Accuracy"=0.2451), although this was achieved at the cost of a partial presence of patterns ("Pattern"=0.9253). Conversely, CD was the best model at Port Elizabeth (I_{rad}=0.0315) because it allowed the best pattern control ("Pattern"=0.0533). At Smolensk CD was successful (I_{rad}=0.0887) because it gave the best correlation between estimates and measurements ("Correlation"=0.0050). At Poza Rica the model BC was the best according to all modules.

The results obtained here are again the consequence of the choices we made regarding the selection of input variables, the definition of their transition intervals, and the values given to the conclusions of the decision rules. These are illustrative results only and are not meant as conclusive findings on the ability of each model at specific sites. The investigation should be extended over multi-year data sets before providing final results. With large, multi-year data sets, a statistically-based investigation could be performed. In particular, a comprehensive assessment on the model performance would include the quantification of the variability associated to I_{rad}, the significance of the computed I_{rad} scores, and the correlation between I_{rad} scores and the respective modules. This would allow statistical separation of I_{rad} scores. Another option would involve cluster analysis to discriminate about model performance over a large amount of locations (or subsets of them), using both the I_{rad} values and the ones of the modules "Accuracy", "Correlation", and "Pattern".

DISCUSSION

In the design of a system to assess radiation model performance two major questions have to be answered: a) which input index should be taken into account, and b) how should the basic indices be aggregated. The method presented in this paper proposes an answer to both questions, but its relevance lies primarily in the answer it provides to the second question. The approach contains two key elements: the use of a fuzzy set and the use of decision rules. The use of a fuzzy set provides a well-designed solution to the problem of deciding the cut-off values for input indices: e.g., the limit between F response, U response, and transition response. The use of decision rules provides a "rational" aggregation of input indices in the related module: e.g., RRMSE, EF and P(t) in the module "Accuracy". The combination of these two concepts (limits in the response, mode of aggregation) in sets of fuzzy rules is attractive because, although the combinations of values of input indices are infinite, a single set of fuzzy rules connects them all.

In this application, the system is based on a compromise between operational suitability (the evaluation of radiation model performance) and flexibility (hierarchization of objectives and aggregation of preferences). It requires extended corroboration, considering that the objective of an expert system is the simulation of a human expert. The expert system is corroborated if it displays, under a variety of conditions, the same responses that a human expert would display. Experts are therefore invited to comment on the set-up and results of this system. If there is disagreement between expert perception of radiation model performance and the input of the system, the cause of this divergence will be examined in view of: a) choice of input indices; b) choice of the limits of the transition interval; c) formulation of the decision rules; and d) formulation of the mode of aggregation of the modules. All of these points may be modified according to expert consensus, after an extensive testing of the methodology. This process may be relatively time-demanding to perform on large data sets. For this purpose, the software IRENE (Fila et al., 2001, 2002), which supplies provisions for fuzzy-based aggregation, may be of help (free download from the site: http://www.isci.it/tools). IRENE allows the creation of re-usable modules and indicators (ASCII files), thus serving as a convenient means to support collaborative work among large, distributed network of scientists involved in creating and aggregating fuzzy-based components.

CONCLUSIONS

We propose a fuzzy expert system reflecting our expert judgment of the quality of radiation model performance. Providing usable values of basic evaluation indices, the method allowed to build an aggregated indicator with a modular structure. The system takes into account three types of input variables: magnitude of model residuals, correlation between estimated values and measured data, and patterns in the behavior of residuals versus independent variables. The resulting output can be used as a support to rank or choose among alternative models under a variety of conditions.

Although a wide expert examination is still required for a general consensus about weights and limits applied to indices and aggregated modules, the fuzzy sets suggested here may represent a pragmatic approach towards a satisfactory solution.

This approach to model evaluation is useful for a number of reasons:

1. it allows users to express mathematically individual or collective values and preferences (uncertainty factors);
2. it highlights the degree of model failure/goodness associated with each information source (i.e., model accuracy, correlation, patterns in the residuals);
3. it elucidates the degree of reliability of response associated with each alternative model;
4. it facilitates structuring of various components of the evaluation process;
5. it reduces several sources and levels of information into a single value;
6. it allows examination of operational equivalence between several indices and modules.

The modular structure also presents advantages. In the first place, users have access to both a synthetic indicator reflecting overall judgment and to each of the modules. This means a transparency of each step, and a control opportunity exists for anybody concerned by the process itself. Secondly, the mode of aggregation of modules can be changed and new modules can be eventually added. The multi-value nature of the issue we are dealing with is explicitly stated, the rules are easy to read, and the numerical scores used for their conclusion are easy to tune to match expert opinions.

The method illustrated here is flexible and can be extended to aggregating more indices of different types. The same method could be conveniently applied to state variables other than solar radiation, e.g. evapotranspiration. Further evolution may be the implementation of this system in exhaustive optimization algorithms and its use as a cost function in model parameters estimation.

APPENDIX A: Fuzzy expert systems

Zadeh (1965) proposed the use of fuzzy set theory to describe relationships that are best characterized by compliance to a collection of attributes. In setting up the set of decision rules in model evaluation the attributes are the basic evaluation indices, and the user must decide on the relative importance of each one. At the same time, the limit values beyond which the index is "certainly" favorable or unfavorable must be given by the user. With this procedure three membership classes are created for the index values: Favorable (F), Unfavorable (U), partial (or fuzzy) membership.

Fuzzy set theory addresses this type of problem by allowing one to define the "degree of membership" of an element in a set by means of membership functions (transition functions), that can take any value from the interval [0, 1]. The value 0 represents complete non-membership, the value 1 represents complete membership, and values in between are used to represent partial membership. For classical or "crisp" sets , the membership function only takes two values: 0 (non-membership) and 1 (membership).

The hierarchical structure of this technique is used to aggregate indices into first-level fuzzy indicators (modules) and, next, into a second-level fuzzy indicator. Each objective in the attribute hierarchy is given a weight. This process of aggregation may continue, hypothetically, until a final-level fuzzy indicator is achieved. The indicator developed here is a second-level indicator. For simplicity, in the example below only first-level aggregation is developed.

The aggregation process is accomplished by combining weighted fuzzy values. According to this approach we can characterize the shape of the membership function of each input index by the two limits of the "transition interval". Fuzzy membership functions may have different shapes, depending on someone's experience or even preference. We used membership functions that are S-shaped in the transition interval, since they provide smoother variations of the values of the inputs than functions that are linear in the transition interval. If x is the value of the index, α and γ lower and upper bound respectively, the S-function is flat at a value of 0 and 1 for $x \leq \alpha$ and for $x \geq \gamma$ respectively. Between α and γ the S-function is a quadratic function of x (Liao, 2002):

$$S(x;\alpha;\gamma)=\begin{cases} 0 & x\leq\alpha \\ 2\cdot\left(\dfrac{x-\alpha}{\gamma-\alpha}\right)^2 & \alpha\leq x\leq\beta \\ 1-2\cdot\left(\dfrac{x-\gamma}{\gamma-\alpha}\right)^2 & \beta\leq x\leq\alpha \\ 1 & x\geq\gamma \end{cases} \tag{7}$$

where $\beta=(\alpha+\gamma)/2$. Two adjacent fuzzy terms with S-shaped membership functions have 0.5 overlap at the midpoint between the two extremes. Eq. 7 can only represent the left-hand side. For the right-hand side, the complement of Eq. 7 is needed.

For each module we formulated a set of decision rules attributing values between 0 and 1 to an output variable according to the membership of its input indices to the fuzzy subsets F and U.

The linguistic description of these components is accomplished in the form of fuzzy rules with a relatively simple syntax. In fact, fuzzy rules inference involves computation of fuzzy rules which are mostly *if* … *then* … statements. When two indices are aggregated the principle of the method makes use of the conjunctive operator AND, as formalized by four rules:

if (x_1 is A_{11}) AND (x_2 is A_{12}) *then* (y_1 is B_1) [rule 1]

if (x_1 is A_{21}) AND (x_2 is A_{22}) *then* (y_2 is B_2) [rule 2]

if (x_1 is A_{31}) AND (x_2 is A_{32}) *then* (y_3 is B_3) [rule 3]

if (x_1 is A_{41}) AND (x_2 is A_{42}) *then* (y_4 is B_4) [rule 4]

where x_j (j=1, 2) is an input index, A_{ij} is a fuzzy subset, y_i is an output variable, B_i (i=1, 2, 3, 4) is a number. "x_j is A_{ij}" is called a premise of the i-th rule; "y_i is B_i" is called conclusion (or expert weight) of the i-th rule. In this case the decision rules consist of two premises (*if…* AND *if…*) linked by and followed by a conclusion (*then…*).

The process continues identifying the degree of truth in the premise portion of each rule and, then, aggregating the truth of linked conditions. Let $'x_1$ and $'x_2$ be the values taken by x_1 and x_2, and $A_{ij}('x_j)$ the membership value of $'x_j$ to the fuzzy set A_{ij} (given by the membership function that defines A_{ij}). According to Sugeno's inference method, when the premises are linked by a conclusion, the truth value of a decision rule is defined as the smallest of the truth values of its premises. Therefore, a fuzzy subset is assigned to each output variable for each rule using "min" aggregator, where min means "minimum value of". One can define w_1, w_2, w_3 and w_4 the truth values of the rules, as follows:

$w_1=\min(A_{11}('x_1), A_{12}('x_2))$ [truth value of the rule 1]

$w_2=\min(A_{21}('x_1), A_{22}('x_2))$ [truth value of the rule 2]

$w_3=\min(A_{31}('x_1), A_{32}('x_2))$ [truth value of the rule 3]

$w_4=\min(A_{41}('x_1), A_{42}('x_2))$ [truth value of the rule 4]

The first rule infers $w_1 \cdot B_1$, the second one $w_2 \cdot B_2$, and so on. The fuzzy sets that represent the outputs of each rule are combined by summation into a single fuzzy solution set ($'y_0$):

$$'y_0=w_1 \cdot B_1+w_2 \cdot B_2+w_3 \cdot B_3+w_4 \cdot B_4$$

A solution of this type is sometimes known as "singleton" output membership function, and it can be thought of as a pre-defuzzified fuzzy set.

The global output $'y$ is inferred by:

$$'y=(w_1 \cdot B_1+w_2 \cdot B_2+w_3 \cdot B_3+w_4 \cdot B_4)/(w_1+w_2+w_3+w_4)$$

This last operation (the weighted average of a few data points) is a common method (centroid calculation) adopted to reduce the final fuzzy set to a crisp value (defuzzification) in the Sugeno-type systems.

The computation of the aggregated indicator is primarily influenced by the truth value of each rule (w_i), for which the expert weight (B_i) is a multiplication factor. As a result of this insight, it should be clear that weights on rules are not simply measures of the "relative importance" of each rule. They are essentially measures of the importance of the increase from the worst to the best level of performance on one rule. The terms B_i are subjective elements introduced in the process of aggregation, which the global output may be sensitive to. This means that the quality of outputs is influenced by the weights of the individual rules. This crucial issue should be investigated within the context of interest, which the truth values depend on.

To illustrate Sugeno's inference method, a numerical example will be given for the composition of the module "Pattern", obtained by aggregating two pattern indices, PI_{doy} and PI_{Tmin} (see details in the paragraph "The structure of the indicator"). We gave three classes of model response with respect to the presence of patterns in the residuals against one independent variable (Figure 4a). We gave a response classified as "pattern" (PI_{doy}, $PI_{Tmin} \geq 2.5$ MJ m^{-2} d^{-1}) a membership value of 1 for the fuzzy subset U and a membership value of 0 for the fuzzy subset F. Model response classified as "no pattern" (PI_{doy}, $PI_{Tmin} \leq 1.0$ MJ m^{-2} d^{-1}) is given a membership value of 0 for the fuzzy subset U and a membership value of 1 for the fuzzy subset F (Figure 4b). The class of "borderline values" (1.0 MJ m^{-2} $d^{-1} <$ PI_{doy}, $PI_{Tmin} < 2.5$ MJ m^{-2} d^{-1}) falls within a "transition interval" in which the membership value for U increases from 0 (at PI_{doy}, $PI_{Tmin} = 1.0$ MJ m^{-2} d^{-1}) to 1 (at PI_{doy}, $PI_{Tmin} = 2.5$ MJ m^{-2} d^{-1}), and the membership value for F decreases from 1 to 0 (thus the functions characterizing F and U are complementary). In our example the reasoning for the four rules is:

if (PI_{doy} is F) AND *if* (PI_{Tmin} is F) *then* (B_1 is 0.0) [rule 1]

if (PI_{doy} is F) AND *if* (PI_{Tmin} is U) *then* (B_2 is 0.5) [rule 2]

if (PI_{doy} is U) AND *if* (PI_{Tmin} is F) *then* (B_3 is 0.5) [rule 3]

if (PI_{doy} is U) AND *if* (PI_{Tmin} is U) *then* (B_4 is 1.0) [rule 4]

You can notice the same weight is attributed here to each basic index.

Let's assume that PI_{doy} was computed equal to 2.00 MJ m^{-2} $d^{-1}$and PI_{Tmin} equal to 1.70 MJ m^{-2} d^{-1}. For both indices membership to fuzzy subsets F and U has to be defined. The membership functions of Figure 5 allow the calculation of the truth value of the premises, i.e. the degree of membership to the fuzzy subset concerned for PI_{doy} (Figure 5, top) and PI_{Tmin} (Figure 5, bottom). Results are shown in Table 10. The value of “Pattern” is calculated as the sum of the conclusions of the decision rules, weighted by the sum of their truth values:

$$\text{“Pattern”} = \frac{0.0 \cdot 0.2222 + 0.5 \cdot 0.2222 + 0.5 \cdot 0.5644 + 1.0 \cdot 0.4356}{0.2222 + 0.2222 + 0.5644 + 0.4356} = 0.5739$$

APPENDIX B: Daily solar radiation models

The general approach followed by the models used in this study for estimating solar radiation at ground level (Rad, MJ m^{-2} d^{-1}) is of the form:

$$Rad = tt_i \cdot R_a \qquad (8)$$

where tt_i is the atmospheric transmissivity and R_a is the extraterrestrial radiation (that is the calculated solar radiation outside the earth's atmosphere).

Based purely on solar geometry and the solar constant, an ordinary routine for estimating daily R_a at given latitudes (e.g., Swift, 1976; Bristow and Campbell, 1984; Campbell and Norman, 1998) was used.

As regards tt_i, three models to estimate it are here used:

- Model BC:

$$tt_i = \tau \cdot \left[1 - \exp\left(\frac{-b \cdot \Delta T_i^c}{\Delta T_m}\right)\right] \qquad (9)$$

- Model CD:

$$tt_i = \tau \cdot \{1 - \exp - [b \cdot f(Tavg) \cdot \Delta T_i^2 \cdot f(Tmin)]\} \qquad (10)$$

- Model DB:

$$tt_i = \tau \cdot \left[1 + c1 \cdot \sin\left(i \cdot \frac{\pi}{180} \cdot c2\right)\right] \cdot \left[1 - \exp\left(\frac{-b \cdot \Delta T_i^2}{\Delta T_w}\right)\right] \qquad (11)$$

τ	=	clear sky atmospheric transmissivity
i	=	day of the year
ΔT	=	$Tmax_i-(Tmin_i+Tmin_{i+1})/2$
ΔT_m	=	monthly mean ΔT (fixed mean)
ΔT_w	=	weekly mean ΔT (mobile mean)
f(Tavg)	=	$0.017 \cdot \exp[\exp(-0.053 \cdot Tavg_i)]$

$Tavg_i = (Tmax_i + Tmin_i)/2$

$f(Tmin) = \exp(Tmin_i/Tnc)$

where all temperatures are in °C.

The parameters b (temperature range coefficient) and c display the physics involved in the relationship, and determine the rate of increase of the exponential function as ΔT increases. The parameter c is usually kept equal to 2 (Ndlovu, 1994; Donatelli, 1995; Weiss et al., 2001). The parameters Tnc (summer night temperature factor), c1 and c2 (general seasonality factors) are empirical parameters accounting for seasonal effects.

REFERENCES

Addiscott, T.M., and A.P. Whitmore. 1987. Computer simulation of changes in soil mineral nitrogen and crop nitrogen during autumn, winter and spring. J. Agric. Sci. (Cambr.) 109:141-157.

Bardossy, A., I. Bogardi, and L. Duckstein. 1985. Composite programming as an extension of compromise programming. p. 375-408. *In* P. Serfini (ed.) Mathematics of multiple objective optimisation. Springer-Verlag, Wien, Austria.

Bechini, L., G. Ducco, M. Donatelli, and A. Stein. 2000. Modelling, interpolation and stochastic simulation in space and time of global solar radiation. Agric., Ecosyst. Environ. 81:29-42.

Bouchon-Meunier, B. 1993. La logique floue. Presses Universitaires de France, Paris, France.

Bristow, K.L. and G.S. Campbell. 1984. On the relationship between incoming solar radiation and daily maximum and minimum temperature. Agric. For. Meteorol. 31:159-166.

Brutsaert, W.H. 1982. Evaporation into the atmosphere: theory, history and applications. D. Rediel, Boston, USA.

Cahan, S. 1987. Stability and change in human characteristics. Wiley, New York.

Campbell, G.S., and J.M. Norman. 1998. An introduction to environmental biophysics, II edition. Springer, New York, USA.

Cornelissen, A.M.G., J. van den Berg, W.J. Koops, M. Grossman, and H.M.J. Udo. 2001. Assessment of the contribution of sustainability indicators to sustainable development: a novel approach using fuzzy set theory. Agric., Ecosyst. Environ. 86:173-185.

CTAP, 2000. Air quality trends in Illinois. p. 91-153. *In* CTAP (ed.) The changing Illinois environment: critical trends, Vol. 2 Air resources. Illinois Natural Resource Information, Springfield, USA.

Donatelli, M. 1995. Sistemi nella gestione integrata delle colture. PANDA project, General Series, Paper 1. ISA, Modena, Italy.

Donatelli, M., M. Acutis, and G. Bellocchi. 2000. Two statistical indices to quantify patterns of errors produced by models. p. 186. *In* Proc. International Crop Science Conference, 3rd, Hamburg, Germany. 17-22 August 2000. European Society for Agronomy, Hamburg, Germany.

Donatelli, M., M. Acutis, G. Bellocchi, and G. Fila. [2002a.] New indices to quantify patterns of residuals produced by model estimates. Agron. J. (submitted)

Donatelli, M., and G. Bellocchi. 2000. New methods to estimate global solar radiation. p. 186. *In* Proc. International Crop Science Conference, 3rd, Hamburg, Germany. 17-22 August 2000. European Society for Agronomy, Hamburg, Germany.

Donatelli, M., and G. Bellocchi. 2001. Estimate of daily global solar radiation: new developments in the software RadEst3.00. p. 213-214. *In* Proc. International Symposium Modelling Cropping Systems, 2nd, Florence, Italy. 16-18 July 2001. CNR - Institute for Biometeorology, Florence, Italy.

Donatelli, M., G. Bellocchi, and F. Fontana. [2002b.] RadEst3.00: Software to estimate daily radiation data from commonly available meteorological variables. Eur. J. Agron. (in press).

Donatelli, M., and G.S. Campbell. 1998. A simple model to estimate global solar radiation. p. 133-134. *In* Proc. European Society for Agronomy Congress, 5th, Vol. II, Nitra, Slovak Republic. 28 June-2 July 1998. The Slovak Agricultural University, Nitra, Slovak Republic.

Donatelli, M., and V. Marletto. 1994. Estimating surface solar radiation by means of air temperature. p. 352-353. *In* Proc. European Society for Agronomy Congress, 3rd, Abano-Padova, Italy. 18-22 September 1994. The Padova University, Padova, Italy.

Dubois, D., and H. Prade. 1980. Fuzzy sets and systems: theory and applications. Academic Press, New York, USA.

Fila, G., G. Bellocchi, M. Acutis, and M. Donatelli. 2001. IRENE: a software to test model performance. p. 215-216. *In* Proc. International Symposium Modelling Cropping Systems, 2nd, Florence, Italy. 16-18 July 2001. CNR - Institute for Biometeorology, Florence, Italy.

Fila, G., G. Bellocchi, M. Acutis, and M. Donatelli. [2002.] IRENE: a software to evaluate model performance. Eur. J. Agron. (in press).

Fox, D.G. 1981. Judging air quality model performance: a summary of the AMS workshop on dispersion models performance. Bull. Am. Meteorol. Soc. 62:599-609.

Hall, L.O., and A. Kandel. 1991. The evolution from expert systems to fuzzy expert systems. p. 3-21. *In* A. Kandel (ed.) Fuzzy expert systems theory. CRC Press, Boca Raton, USA.

Hinkle, D., W. Wiersma, and S. Jurs. 1994. Applied statistics for the behavioral sciences, III edition. Houghton Mifflin Company, Boston, USA.

Iqbal, M. 1983. An introduction to solar radiation. Academic Press, New York, USA.

Jones, D., and E.M. Barnes. 2000. Fuzzy composite programming to combine remote sensing and crop models for decision support in precision crop management. Agric. Syst. 65:137-158.

Kedzie, C.R. 1997. Quantitative analyses: the empty corner. *In* C.R. Kedzie (ed.) Communication and democracy: coincident revolutions and the emergent dictator's dilemma. [Online] Available at: http://www.rand.org/publications/RGSD/RGSD127/sec4.html (verified 11 Apr 2002). RGSD-127, RAND, Santa Monica, USA.

Keeney, R.L., and H. Raiffa. 1993. Decisions with multiple objectives: preferences and value tradeoffs. Cambridge University Press, New York, USA.

Kleijnen, J.P.C., Bettonvil, B., and W. van Groenendaal. 1998. Validation of trace-driven simulation models: a novel regression test. Management Science 44:812-819.

Kobayashi, K., and M.U. Salam. 2000. Comparing simulated and measured values using mean squared deviation and its components. Agron. J. 92:345-352.

Liao, T.W. 2002. A fuzzy C-medians variant for the generation of fuzzy term sets. Int. J. Intelligent Systems 17:21-43.

Loague, K., and R.E. Green. 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. J. Contam. Hydrol. 7:51-73.

Martorana, F., and G. Bellocchi. 1999. A review of methodologies to evaluate agroecosystem simulation models. Ital. J. Agron. 3:19-39.

Metselaar, K. 1999. Auditing predictive models: a case study in crop growth. Thesis, Wageningen Agricultural University, Wageningen, The Netherlands.

Ndlovu, L.S. 1994. Generating weather data for crop simulation model. Ph.D. diss. Biol. Syst. Eng. Dep., Washington State Univ., Pullman, WA.

Robinson, S. 1998. Service quality in the management of simulation projects. PhD Thesis, Lancaster University, United Kingdom.

Smith, P., J.U. Smith, D.S. Powlson, W.B. McGill, J.R.M. Arah, O.G. Chertov et al. 1997. A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. Geoderma 81:153-225.

Sugeno, M. 1985. An introductory survey of fuzzy control. Information Sciences 36:59-83.

Swift, L.W. 1976. Algorithm for solar radiation estimation on mountain slopes. Wat. Resour. Res. 12:108-112.

Thornton, P.E., and S.W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. Agr. For. Meteorol. 93:211-228.

Tracy, C.R., Hammond, K.A., Lechleitner, R.A., Smith II, W.J., Thompson, D.B., Whicker, A.D. et al. 1983. Estimating clear-day solar radiation: an evaluation of three models. J. Therm. Biol. 8:247-251.

Van der Werf, H.M.G., and C. Zimmer. 1998. An indicator of pesticide environmental impact based on a fuzzy expert system. Chemosphere 36:2225-2249.

Weiss, A., Hays, C.J., Hu, Q., and W.E. Easterling. 2001. Incorporating bias error in calculating solar irradiance: implications for crop yield simulations. Agron. J. 93:1321-1326.

Willmott C.J., 1982. Some comments on the evaluation of model performance. Bull. Am. Meteorol. Soc. 63:1309-1313.

Yang, J., D.J. Greenwood, D.L. Rowell, G.A. Wadsworth, and I.G. Burns. 2000. Statistical methods for evaluating a crop nitrogen simulation model, N_ABLE. Agric. Syst. 64:37-53.

Zadeh, L.A. 1965. Fuzzy sets. Information and Control 8:338-353.

Fig. 1 - Examples of pattern indices (PI) versus two independent variables (DOY: day of the year; Tmin: minimum air temperature) computed on residuals generated at Patos de Minas (Brazil) in the year 1997 by different radiation models (BC: Bristow-Campbell; CD: Campbell-Donatelli; DB: Donatelli-Bellocchi).

Fig. 2 - Examples of pattern indices (PI) versus two independent variables (DOY: day of the year; Tmin: minimum air temperature) computed on residuals generated at Würzburg (Germany) in the year 1998 by different radiation models (BC: Bristow-Campbell; CD: Campbell-Donatelli; DB: Donatelli-Bellocchi).

Fig. 3 - Sensitivity analysis of the indicator I_{rad} to variation of all input indices. Each input index is varied over its transition interval from 0.0 (completely favorable) to 1.0 (completely unfavorable) while the other input indices are kept at Unfavorable (top graph), at the median value of their transition interval (middle graph) or at Favorable (bottom graph). The traces of the pattern indices (PI_{doy} and PI_{Tmin}) are superimposed.

Fig. 4 - Graphical presentation of crisp (a) and fuzzy (b) sets for the pattern index (PI).

Fig. 5 - Membership to the fuzzy sets Favorable and Unfavorable for an hypothetical model response in terms of pattern indices: PI_{doy}=2.00, PI_{Tmin}=1.70.

TABLES

Table 1. The indicator modules "Accuracy", "Correlation" and "Pattern", and their inputs. For details, see the text.

Inputs	Indicator modules		
	"Accuracy"	"Correlation"	"Pattern"
RRMSE	x		
P(t)	x		
EF	x		
r		x	
PI_{doy}			x
PI_{Tmin}			x

Table 2. Summary of decision rules describing the effect of the input indices RRMSE, EF and P(t) on the module "Accuracy". F = Favorable, U = Unfavorable. For details, see the text.

RRMSE	EF	P(t)	Expert weight
Membership class			
F	F	F	0.00
F	F	U	0.20
F	U	F	0.40
F	U	U	0.60
U	F	F	0.40
U	F	U	0.60
U	U	F	0.80
U	U	U	1.00

Table 3. Summary of decision rules describing the effect of the input indices PI_{doy} and PI_{Tmin} on the module “Pattern”. F = Favorable, U = Unfavorable. For details, see the text.

PI_{doy}	PI_{Tmin}	Expert weight
Membership class		
F	F	0.00
F	U	0.50
U	F	0.50
U	U	1.00

Table 4. Summary of decision rules describing the effect of the three modules on the value of the indicator (I_{rad}).

"Accuracy"	"Correlation"	"Pattern"	Expert weight
Membership class			
F	F	F	0.00
F	F	U	0.30
F	U	F	0.15
F	U	U	0.45
U	F	F	0.55
U	F	U	0.85
U	U	F	0.70
U	U	U	1.00

Table 5. Relative incidence of each basic evaluation index on the value of the indicator (I_{rad}).

Index	**Relative incidence on I_{rad}**
RRMSE	0.4 x 0.55 = 0.22
EF	0.4 x 0.55 = 0.22
P(t)	0.2 x 0.55 = 0.11
r	1.0 x 0.15 = 0.15
PIdoy	0.5 x 0.30 = 0.15
PITmin	0.5 x 0.30 = 0.15

Table 6. Description of the ten locations used in this study: latitude, longitude, elevation, clear sky transmissivity, years, yearly rainfall, yearly average maximum air temperature (Tmax), yearly average minimum air temperature (Tmin), yearly average global solar radiation.

Location	Country	Latitude (deg)	Longitude (deg)	Elevation (m)	τ^*	Year	Rainfall** (mm)	Tmax** (°C)	Tmin** (°C)	Radiation** (MJ m^{-2} d^{-1})
Longreach	Australia	-23.43	144.27	191	0.77	1993	335	32.0	17.4	20.3
Los Baños	Philippines	14.17	121.25	300	0.67	1989	2092	30.2	23.4	15.4
Matsumoto	Japan	36.24	137.97	610	0.74	1994	671	19.0	7.3	15.0
Patos de Minas	Brazil	-18.60	-44.88	833	0.79	1997	1339	28.3	16.6	20.1
Perugina	Italy	43.08	12.50	351	0.77	1995	700	19.8	6.3	14.5
Port Elizabeth	South Africa	-33.98	25.62	59	0.73	1997	539	22.2	13.1	16.3
Poza Rica	Mexico	20.53	-97.45	60	0.73	1988	1340	29.3	19.3	15.9
Sadore	Niger	13.10	2.35	235	0.75	1995	486	36.7	22.4	22.3
Smolensk	Russia	54.80	31.88	175	0.73	1985	877	7.5	-0.3	9.9
Würzburg	Germany	49.77	9.97	268	0.74	1998	599	14.2	6.4	10.7

* Clear sky atmospheric transmissivity: the ratio between the value of solar radiation at earth surface during clear sky days and the value outside the earth atmosphere. See also APPENDIX B.

** Annual averages of rainfall, maximum (Tmax) and minimum (Tmin) air temperature, and global solar radiation.

Table 7. Model parameter values determined at each location. APPENDIX B for solar radiation models Bristow-Campbell (BC), Campbell-Donaatelli (CD) and Donatelli-Bellocchi (DB).

Model	Parameters	Locations									
		Longreach	Los Baños	Matsumoto	Patos de Minas	Perugia	Port Elizabeth	Poza Rica	Sadore	Smolensk	Würzburg
BC	b	0.121	0.184	0.093	0.120	0.103	0.198	0.108	0.144	0.110	0.117
	c	2	2	2	2	2	2	2	2	2	2
CD	b	0.318	0.477	0.27	0.38	0.225	0.477	0.353	0.469	0.367	0.367
	Tnc	106	28.8	59.5	106	106	22	88	106	69	106
DB	b	0.094	0.171	0.099	0.129	0.094	0.197	0.095	0.145	0.112	0.113
	c1	0.186	0.051	0.076	-0.122	-0.008	-0.012	0.043	0.012	0.066	0.048
	c2	0.490	1.192	1.605	1.520	1.070	1.079	0.439	0.958	1.036	1.171

Table 8. Response of three radiation models (BC, CD, DB) on ten yearly data sets. The quality of performance was evaluated in terms of the basic indices RRMSE, EF, P(t), r, PI_{doy} and PI_{Tmin}.

Location	Model	RRMSE	EF	P(t)	r	PI_{doy}	PI_{Tmin}
Longreach	BC	13.93	0.72	0.92	0.88	2.75	2.33
	CD	13.32	0.74	0.91	0.88	1.84	1.55
	DB	13.70	0.73	0.88	0.87	0.71	0.31
Los Baños	BC	19.90	0.61	0.83	0.81	1.59	1.65
	CD	19.28	0.63	0.87	0.82	2.21	1.02
	DB	19.14	0.64	0.94	0.80	0.56	0.74
Matsumoto	BC	20.67	0.79	0.90	0.78	1.74	2.23
	CD	20.25	0.80	0.89	0.99	2.39	1.26
	DB	20.83	0.79	0.89	0.92	0.88	1.56
Patos de Minas	BC	16.74	0.56	0.89	0.79	2.92	5.15
	CD	15.61	0.62	0.79	0.82	2.09	3.07
	DB	17.20	0.54	0.88	0.74	0.90	2.78
Perugia	BC	21.60	0.86	0.78	0.93	1.21	1.63
	CD	20.97	0.87	0.94	0.93	1.86	1.30
	DB	21.82	0.86	0.90	0.93	1.25	1.57
Port Elizabeth	BC	23.45	0.72	0.86	0.86	0.95	1.58
	CD	23.19	0.73	0.84	0.86	1.05	1.04
	DB	22.66	0.74	0.83	0.87	1.14	1.49
Poza Rica	BC	22.40	0.69	0.93	0.83	2.17	1.69
	CD	23.31	0.66	0.78	0.83	2.48	2.64
	DB	23.56	0.66	0.83	0.82	2.13	2.32
Sadore	BC	13.42	0.23	0.96	0.57	1.01	0.74
	CD	14.21	0.14	0.84	0.52	2.24	0.97
	DB	13.03	0.28	0.99	0.53	0.65	0.51
Smolensk	BC	39.44	0.78	0.96	0.88	2.16	0.64
	CD	38.20	0.80	0.98	0.89	1.44	1.14
	DB	41.02	0.77	0.98	0.88	0.98	0.31
Würzburg	BC	25.79	0.88	0.97	0.94	1.29	0.99
	CD	24.55	0.89	0.77	0.95	1.27	1.62
	DB	25.22	0.88	0.99	0.88	0.45	0.45

Table 9. Response of three radiation models (BC, CD, DB) on ten yearly data sets. The quality of performance was evaluated in terms of the indicator modules “Accuracy”, “Correlation” and “Pattern”, and the indicator I_{rad}.

Location	Model	“Accuracy”	“Correlation”	“Pattern”	I_{rad}
Longreach	BC	0.1037	0.0200	0.9872	0.3125
	CD	0.0819	0.0200	0.4615	0.1441
	DB	0.0925	0.0450	0.0000	0.0128
Los Baños	BC	0.2589	0.4050	0.4027	0.3241
	CD	0.2307	0.3200	0.4628	0.3168
	DB	0.2157	0.5000	0.0000	0.1613
Matsumoto	BC	0.0412	0.0000	0.6868	0.2442
	CD	0.0324	0.0050	0.5241	0.1664
	DB	0.0426	0.0050	0.1394	0.0167
Patos de Minas	BC	0.3181	0.5950	1.0000	0.5518
	CD	0.2451	0.3200	0.9253	0.4425
	DB	0.3373	0.9200	0.5000	0.4704
Perugia	BC	0.0200	0.0000	0.2181	0.0296
	CD	0.0084	0.0000	0.3775	0.0856
	DB	0.0215	0.0000	0.2050	0.0265
Port Elizabeth	BC	0.1565	0.0800	0.1495	0.0945
	CD	0.1394	0.0800	0.0032	0.0315
	DB	0.1162	0.0450	0.1283	0.0533
Poza Rica	BC	0.1661	0.2450	0.6367	0.3167
	CD	0.2254	0.2450	0.9998	0.4250
	DB	0.2311	0.3200	0.9016	0.4302
Sadore	BC	0.4000	1.0000	0.0001	0.3261
	CD	0.4000	1.0000	0.4700	0.5040
	DB	0.4000	1.0000	0.0000	0.3260
Smolensk	BC	0.4453	0.0200	0.4486	0.3773
	CD	0.4247	0.0050	0.1084	0.2154
	DB	0.4541	0.0200	0.0000	0.2271
Würzburg	BC	0.0704	0.0000	0.0374	0.0086
	CD	0.0423	0.0000	0.2373	0.0385
	DB	0.0579	0.0200	0.0000	0.0044

Table 10. Summary of decision rules describing the effect of the input indices PI_{doy} and PI_{Tmin} on the module "Pattern". Truth values of premises (w_i) and conclusions (B_i) for PI_{doy}=2.00 and PI_{Tmin}=1.70 are shown. F = Favourable, U = Unfavourable. For details, see the text.

PI_{doy}		PI_{Tmin}		Expert weight (B_i)	Truth value (w_i)	$w_i \cdot B_i$
Membership class	Membership value	Membership class	Membership value			
F	0.2222	F	0.5644	0.00	0.2222	0.0000
F	0.2222	U	0.4356	0.50	0.2222	0.1111
U	0.7778	F	0.5644	0.50	0.5644	0.2822
U	0.7778	U	0.4356	1.00	0.4356	0.4356
Fuzzy solution set ($'y_0$)						0.8289
Sum of truth values					1.4444	
Global output ($'y$)					0.5739	

FIGURES

Figure 1

Patos de Minas, 1997

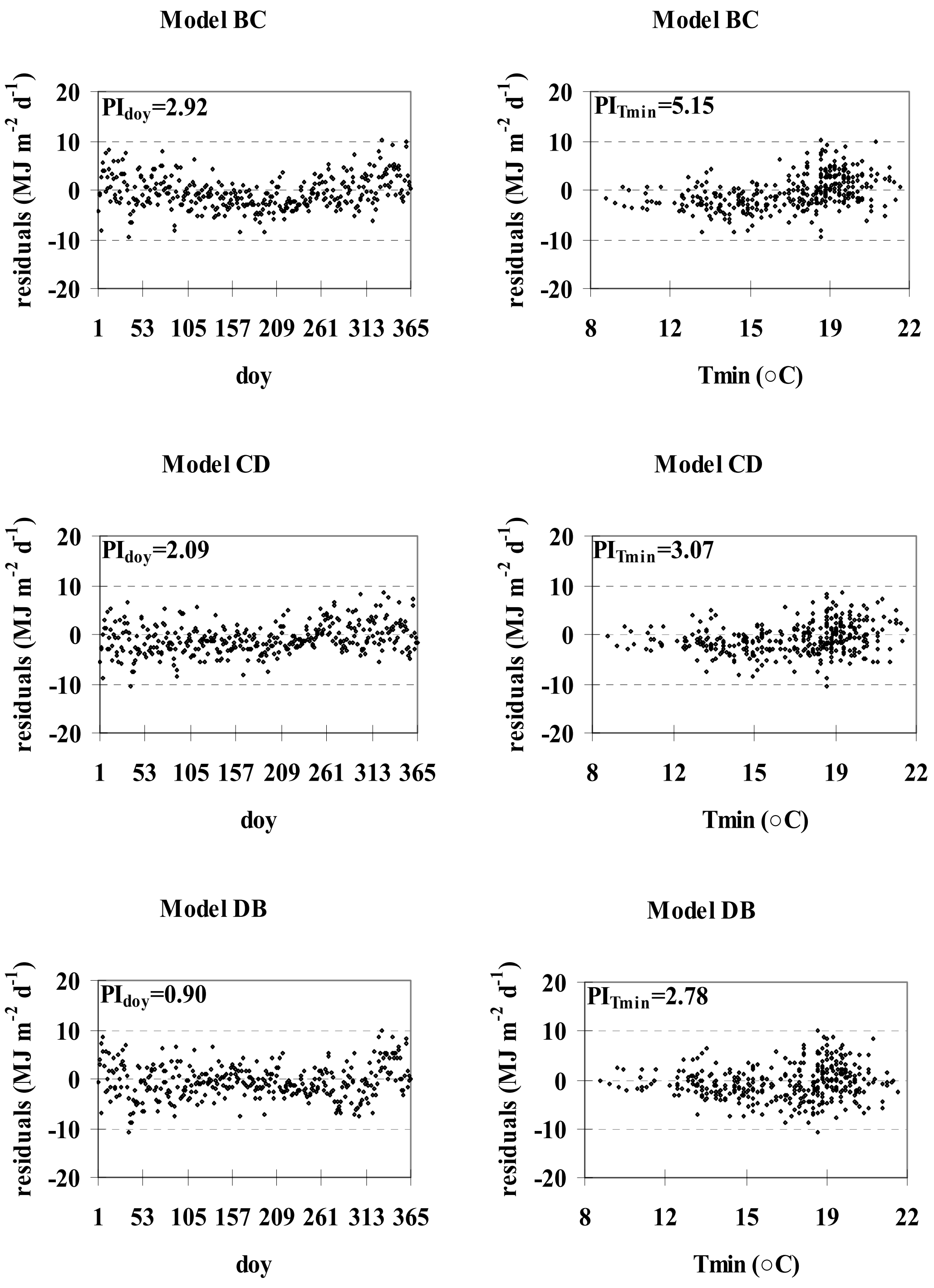


Figure 2

Würzburg, 1998

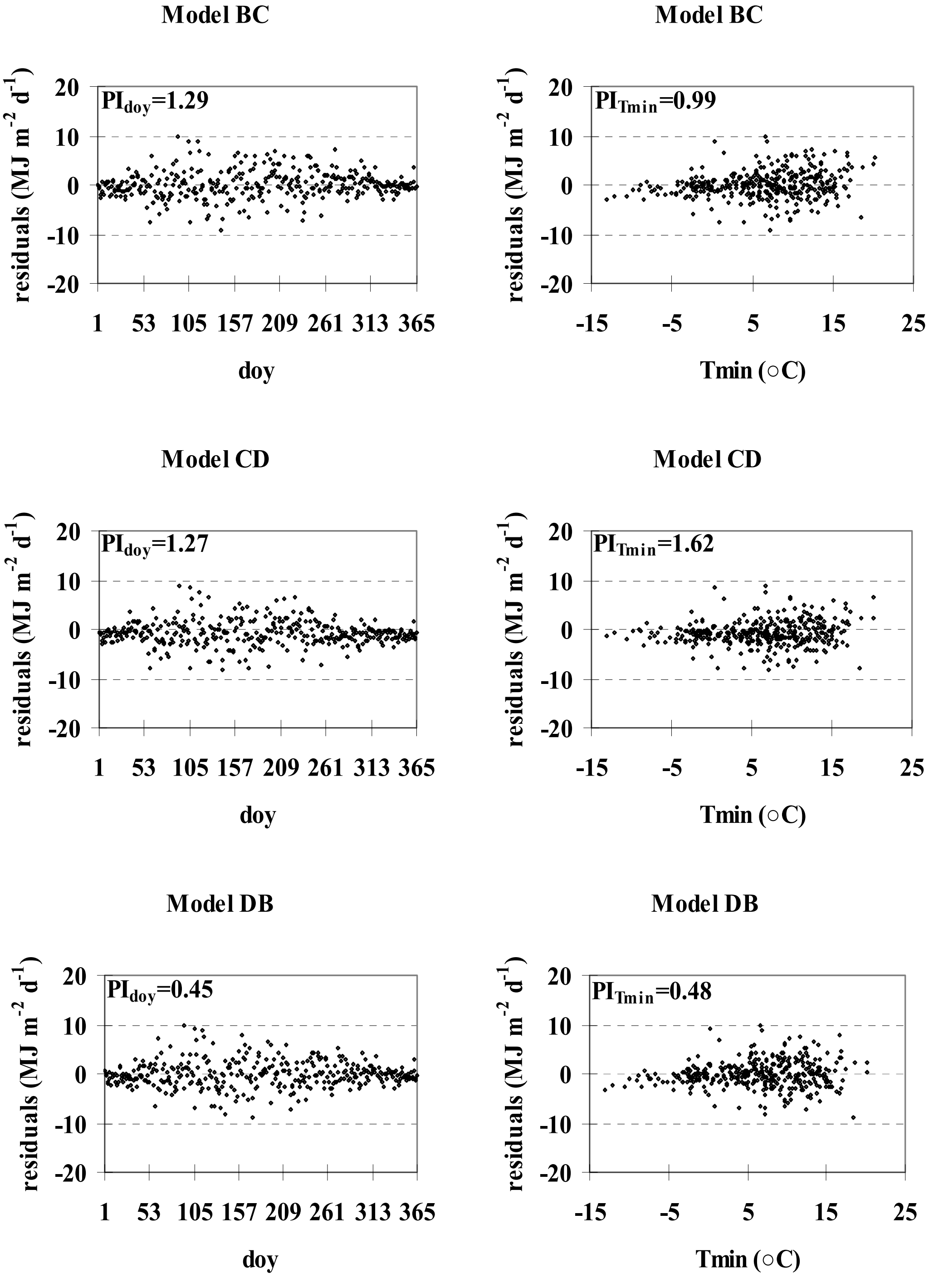
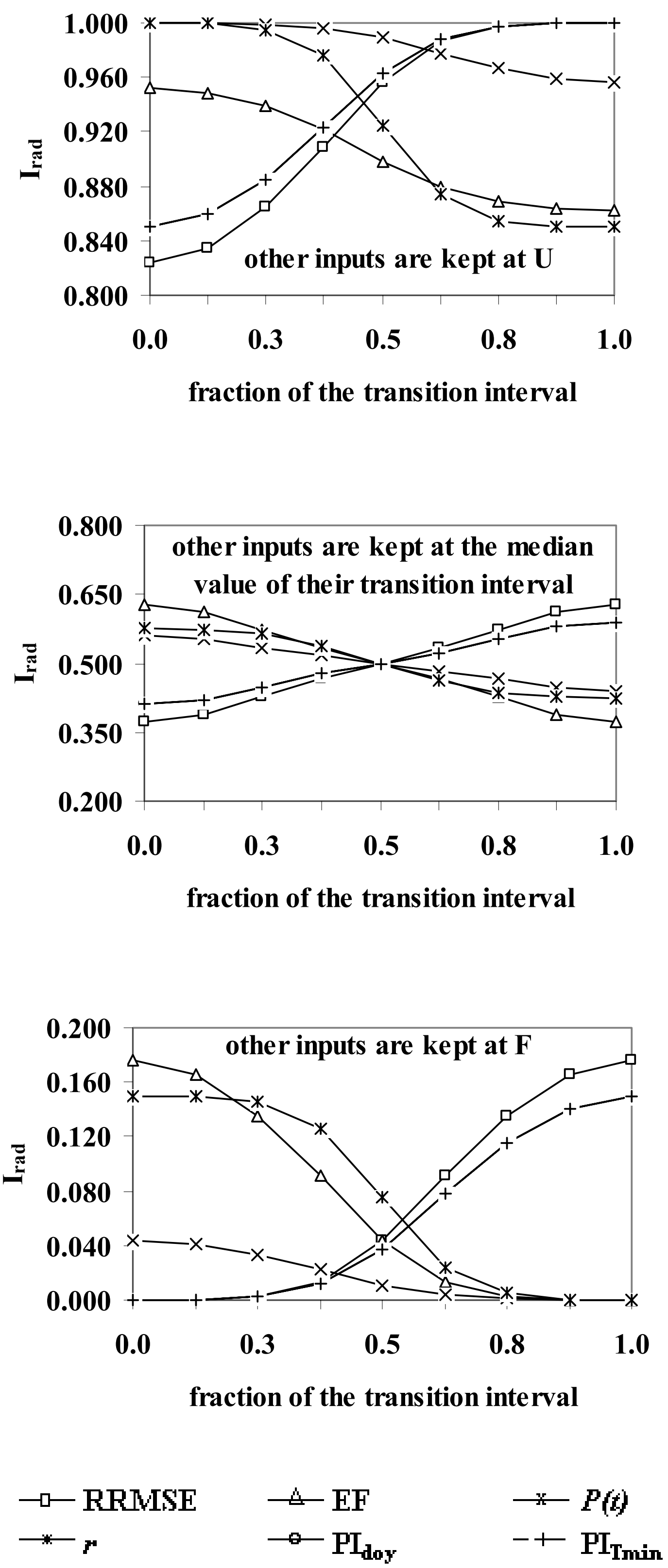
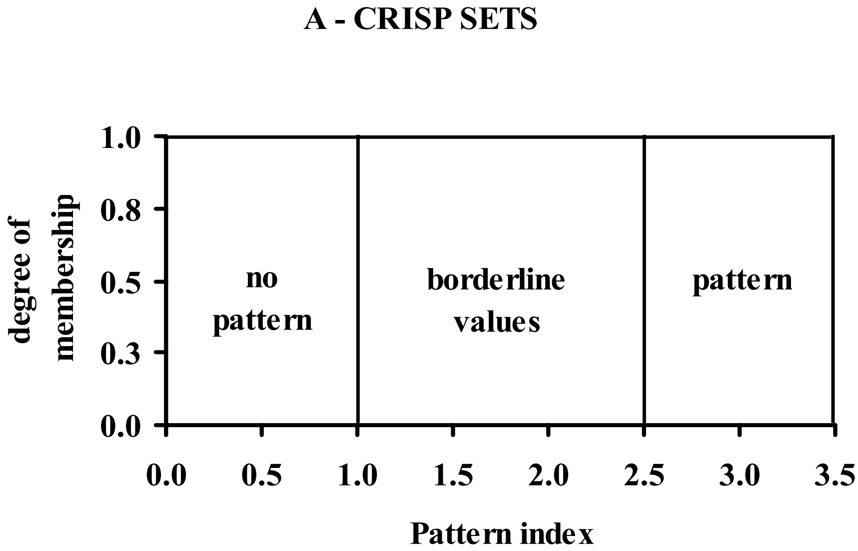


Figure 3



1.000
0.960
0.920
0.880
0.840
0.800
I_rad
other inputs are kept at U
0.0
0.3
0.5
0.8
1.0
fraction of the transition interval
0.800
0.650
0.500
0.350
0.200
I_rad
other inputs are kept at the median value of their transition interval
0.0
0.3
0.5
0.8
1.0
fraction of the transition interval
0.200
0.160
0.120
0.080
0.040
0.000
I_rad
other inputs are kept at F
0.0
0.3
0.5
0.8
1.0
fraction of the transition interval
RRMSE
EF
P(t)
r
PI_doy
PI_Tmin

Figure 4


A - CRISP SETS
degree of membership
1.0
0.8
0.5
0.3
0.0
no pattern
borderline values
pattern
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Pattern index

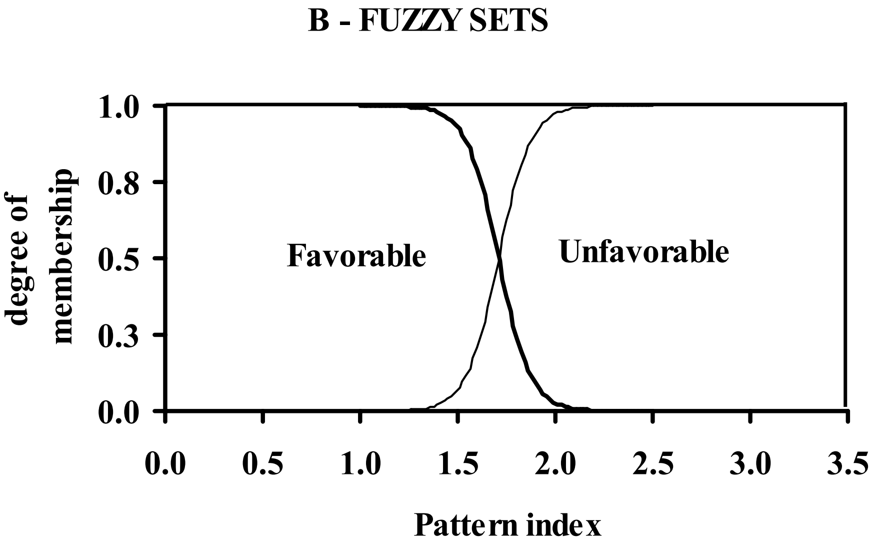
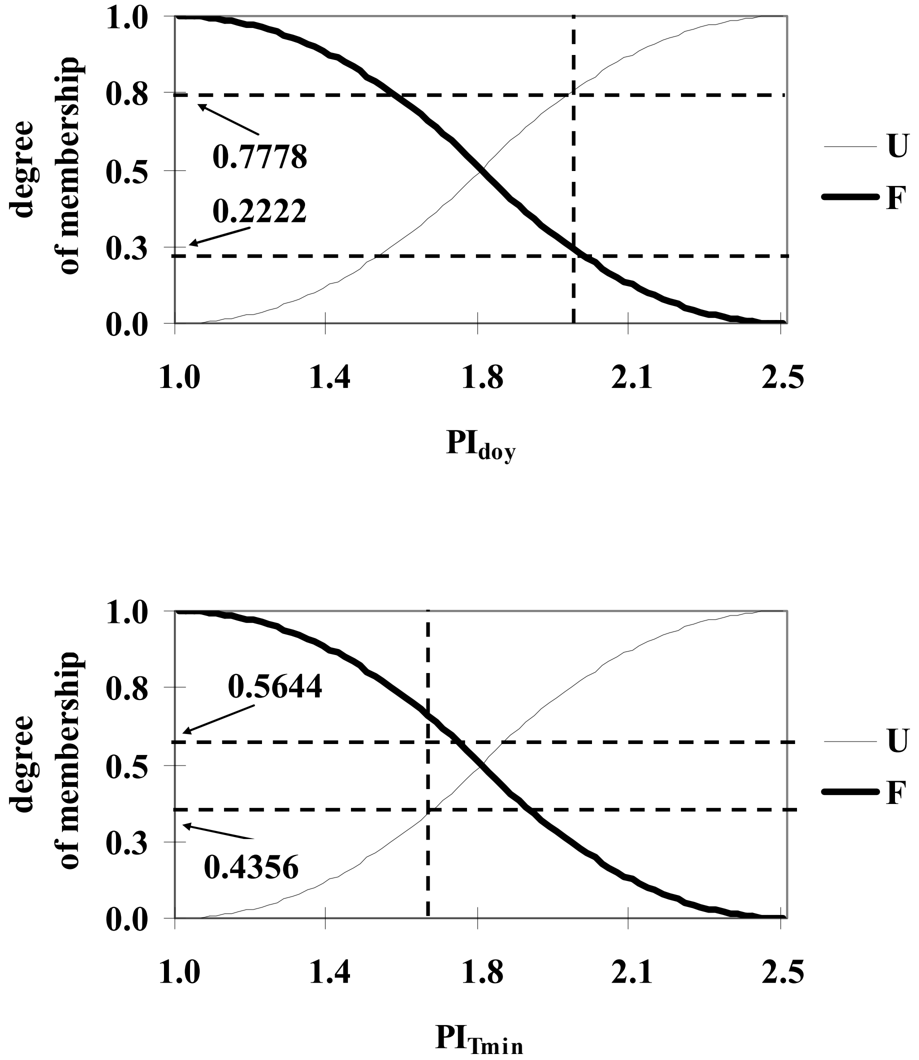

B - FUZZY SETS
degree of membership
1.0
0.8
0.5
0.3
0.0
Favorable
Unfavorable
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Pattern index

Figure 5


1.0
0.8
0.5
0.3
0.0
degree
of membership
0.7778
0.2222
1.0
1.4
1.8
2.1
2.5
PIdoy
U
F
1.0
0.8
0.5
0.3
0.0
degree
of membership
0.5644
0.4356
1.0
1.4
1.8
2.1
2.5
PITmin
U
F