Interrelationship between evaluation metrics to assess agro-ecological models

Mattia Sanna[1], Marco Acutis[1], Gianni Bellocchi[2]

[1]DISAA - University of Milan, Milan (Italy)

[2]Grassland Ecosystem Research Unit – INRA, Clermont-Ferrand (France)

Overview

Topic: interrelationship between different evaluation metrics

Techniques: correlation analysis

Aim: identification of the most useful set of indicators for assessing simulation performance in agro-ecological modelling

Group	Indices
Bias	Mean Bias Error
	Squared Bias
	Fractional Bias
	Coefficient of Residual Mass
	Percent Bias
Accuracy	Mean Squared Error
	Root Mean Squared Error
	Root Mean Squared Relative Error
	Root Mean Squared Variation
	General Standard Deviation
	Normalized Mean Squared Error
Efficiency	Modelling Efficiency
	Modelling Percent Efficiency
	Modified Modelling Efficiency
Persistence Model Efficiency	Persistence Model Efficiency
Correlation/Regression	Pearson's Correlation Coefficient
	Spearman's Correlation Coefficient
	Coefficient of Determination
	Index of Agreement
Median based	Robust Modelling Efficiency
	Median Absolute Error
	Relative Median Absolute Error
Pattern	Pattern Index

Materials and Methods / Data

Variable	Observations	Models	Data sets	Combinations
AGB (Above Ground rice Biomass)	5 – 12	3	10	30
LAI (Leaf Area Index)	5 – 9	1	7	7
HARH (Hourly Air Relative Humidity)	5880	13	9	117
RAD (daily solar RADiation)	701 – 1826	3	10	30

On each combination, 21 performance indicators were calculated.

The interrelationship between performance indicators was assessed by computing the Spearman's rank correlation coefficient.

Materials and Methods / Statistical Analysis

Four Correlations Matrices

Correlations matrices between performance indicators were analysed to assess whether some sort of stable pattern could be detected

*The "degree of stability" of any single correlation can be expressed in terms of p-value: a correlation is defined as "**stable**", if its associated p-value is greater than 0.05.*

method for k independent parameters (Weaver et al., 2013) for testing the null hypothesis that k population correlations are the same

Sig.	MBE	SB	FB	CRM	MSE	RMSE	RMSRE	RMSV	GSD	NMSE	EF	EF1	PME	r	r_s	r^2	d	REF	MdAE	RMdAE	PI
MBE		0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.352	0.258	0.000	0.000	0.000	0.000
SB	0.000		0.000	0.000	0.088	0.088	0.126	0.138	0.061	0.141	0.006	0.002	0.006	0.000	0.000	0.000	0.000	0.000	0.012	0.007	0.049
FB	0.000	0.000		1.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.285	0.228	0.000	0.000	0.000	0.000
CRM	0.000	0.000	1.000		0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.285	0.228	0.000	0.000	0.000	0.000
MSE	0.000	0.088	0.000	0.000		1.000	0.001	0.000	0.000	0.000	0.002	0.004	0.068	0.016	0.000	0.000	0.001	0.078	0.000	0.000	0.030
RMSE	0.000	0.088	0.000	0.000	1.000		0.001	0.000	0.000	0.000	0.002	0.004	0.068	0.016	0.000	0.000	0.001	0.078	0.000	0.000	0.030
RMSRE	0.000	0.126	0.000	0.000	0.001	0.001		0.003	0.000	0.000	0.153	0.066	0.029	0.076	0.038	0.074	0.173	0.082	0.013	0.000	0.016
RMSV	0.007	0.138	0.008	0.008	0.000	0.000	0.003		0.016	0.000	0.000	0.000	0.000	0.186	0.007	0.102	0.364	0.000	0.000	0.000	0.141
GSD	0.000	0.061	0.000	0.000	0.000	0.000	0.000	0.016		0.302	0.000	0.000	0.504	0.009	0.000	0.003	0.017	0.014	0.000	0.000	0.371
NMSE	0.000	0.141	0.000	0.000	0.000	0.000	0.000	0.000	0.302		0.000	0.000	0.202	0.009	0.000	0.003	0.024	0.003	0.000	0.000	0.017
EF	0.000	0.006	0.000	0.000	0.002	0.002	0.153	0.000	0.000	0.000		0.000	0.136	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EF1	0.000	0.002	0.000	0.000	0.004	0.004	0.066	0.000	0.000	0.000	0.000		0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001
PME	0.000	0.006	0.000	0.000	0.068	0.068	0.029	0.000	0.504	0.202	0.136	0.001		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
r	0.000	0.000	0.000	0.000	0.016	0.016	0.076	0.186	0.009	0.009	0.001	0.001	0.000		0.000	0.314	0.348	0.011	0.037	0.000	0.000
r_s	0.001	0.000	0.001	0.002	0.000	0.000	0.038	0.007	0.000	0.000	0.000	0.000	0.000	0.000		0.094	0.080	0.000	0.002	0.000	0.000
CD	0.352	0.000	0.285	0.285	0.000	0.000	0.074	0.102	0.003	0.003	0.000	0.000	0.000	0.314	0.094		0.000	0.000	0.000	0.001	0.023
d	0.258	0.000	0.228	0.228	0.001	0.001	0.173	0.364	0.017	0.024	0.000	0.000	0.000	0.348	0.080	0.000		0.000	0.000	0.003	0.114
REF	0.000	0.000	0.000	0.000	0.078	0.078	0.082	0.000	0.014	0.003	0.000	0.000	0.000	0.011	0.000	0.000	0.000		0.006	0.000	0.016
MdAE	0.000	0.012	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.037	0.002	0.000	0.000	0.006		0.000	0.162
RMdAE	0.000	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.000		0.577
PI	0.000	0.049	0.000	0.000	0.030	0.030	0.016	0.141	0.371	0.017	0.000	0.001	0.001	0.000	0.000	0.023	0.114	0.016	0.162	0.577	

✓measure of the strength of the relationship between two performance indicators

✓useful to quantify their shared information

The stronger the relationship, the greater the amount of common information provided by the two indices.

Identification of all pairs of indicators, showing the same correlation in all the four variables

If a **stable** correlation between a pair of indicators is detected, the correlation coefficients, from a given combination of variables, can be considered as equivalent and consequently their **MEAN** can be calculated, without significant loss of information.

Conclusions

✓Each performance indicator is strongly related to at least one element of the **Optimal Subset**.

✓In each column it is always possible to identify a mean value of correlation coefficients greater than 0.5 in absolute value, with the only exceptions of **r^2** and **r**.

Modelling evaluation can be performed more effectively, avoiding the simultaneous use of strongly related indices!

This approach makes possible to select the most useful set of indicators: (1) including those showing a strong relationship with many other indices, as they somehow can condense a lot of information; (2) combining them such that the condensed information provided by each of them do not overlap, but mutually integrate.

Results / Performance Indicators Selection

1) **d** ← strong correlation with indices belonging to Bias group

2) **SB** ← strong correlation with indices belonging to Accuracy group

3) **RMSRE** ← strong correlation with indices belonging to Efficiency and Correlation/Regression groups

4) **PI** ← strong correlation with indices belonging to Median based group

5) **PME** ← no stable correlation with any of the previous indices

6) **r_s** ← correlation with **d** is stable, but the mean value of their correlation coefficients is very close to zero (-0.040)

Optimal Subset

Mean	MBE	FB	CRM	MSE	RMSE	RMSV	GSD	NMSE	EF	EF1	r	r^2	REF	MdAE	RMdAE
SB				0.664	0.664	0.355	0.506	0.603							
RMSRE									-0.535	-0.564	-0.481	-0.203	-0.667		
PME				-0.796	-0.796		-0.837	-0.858	0.936						
r_s												0.075			
d	-0.514	-0.518	0.518			0.250					0.105				
PI						0.694	0.634							0.572	0.611

Mean values of correlation coefficients calculated on AGB, LAI, HARH and RAD data

Only stable correlations are reported.

UNIVERSITY OF MILAN
DEPARTMENT OF AGRICULTURAL AND ENVIRONMENTAL SCIENCES
PRODUCTION, LANDSCAPE, AGROENERGY

INRA SCIENCE & IMPACT