Australian Journal of Crop Science AJCS

Aust J Crop Sci. 19(01):69-75 (2025) | ISSN:1835-2707
Supplementary 1 & 2

The impact of climactic and spectral variables on soybean productivity: An approach to spatial panel data

Edilza Martins Silva[1*], Alex Paludo[1], Willyan Ronaldo Becker[1], Priscila Pigatto Gasparin[2], Luciana Pagliosa Carvalho Guedes[1], Jerry Adriani Johann[1]

Supplementary 1.

Table 1. Global spatial correlation of the residuals from the Simple Regression Model.

Harvest Year - Cross section	Global Moran Index
2010/2011	0.201*
2011/2012	0.151*
2012/2013	0.179*
2013/2014	0.206*
2014/2015	0.315*
2015/2016	0.279*
2016/2017	0.297*
2017/2018	0.301*
2018/2019	0.219*
2019/2020	0.227*

*(p<0.05).

Table 2 Diagnosis to choose the panel data spatial model.

Lagrange Multiplier Test	Statistics	p-value
LM lag	5,169*	2.2×10^{-16}
LM error	5,067*	2.2×10^{-16}
LM lag (robust)	181*	2.2×10^{-16}
LM error (robust)	78*	2.2×10^{-16}

*(p<0.05).

Table 3. Results of the Pooled, Fixed-effects with no dependence, SAR with fixed effects, and SEM without fixed effects models for the Soybean productivity dependent variable.

Variables	Pooled	Fixed effects without dependence	SAR Fixed effects	SEM Fixed effects
Rain_1b_MVDD_CD	-0.006^{NS}	0.113*	0.088*	0.097*
Rain_SD_MVDD_1a	-0.067*	-0.133*	-0.096*	-0.107*
EVI_1b_MVDD_1a	0.176*	0.174*	0.139*	0.138*
EVI_SD_MVDD_2a	0.328*	0.260*	0.198*	0.204*
T_{me}_SD_2b_MVDD	0.046*	-0.208*	0.182*	0.224*
T_{min}_SD_2b_MVDD	-0.286*	-0.219*	-0.222*	-0.264*
T_{max}_1b_MVDD_1a	-0.050*	0.0001^{NS}	-0.003^{NS}	-0.004^{NS}
SR_MVDD_CD	0.245*	-0.030*	-0.037*	0.051*
ETp_MVDD_2a_CD	-0.224*	0.001^{NS}	0.014*	0.024*
Constant	0.46*	-	-	-
Number of observations	16,990	16,990	16,990	16,990
R^2	0.198	0.109	-	-
Adjusted R^2	0.198	0.009	-	-
Spatial Lag ($\hat{\rho}$)	-	-	0.58*	
Spatial Error ($\hat{\lambda}$)	-	-	-	0.58*
AIC	-3,229	-10,271	-12,253	-10,232

*(p<0.05), NS: Not Significant.

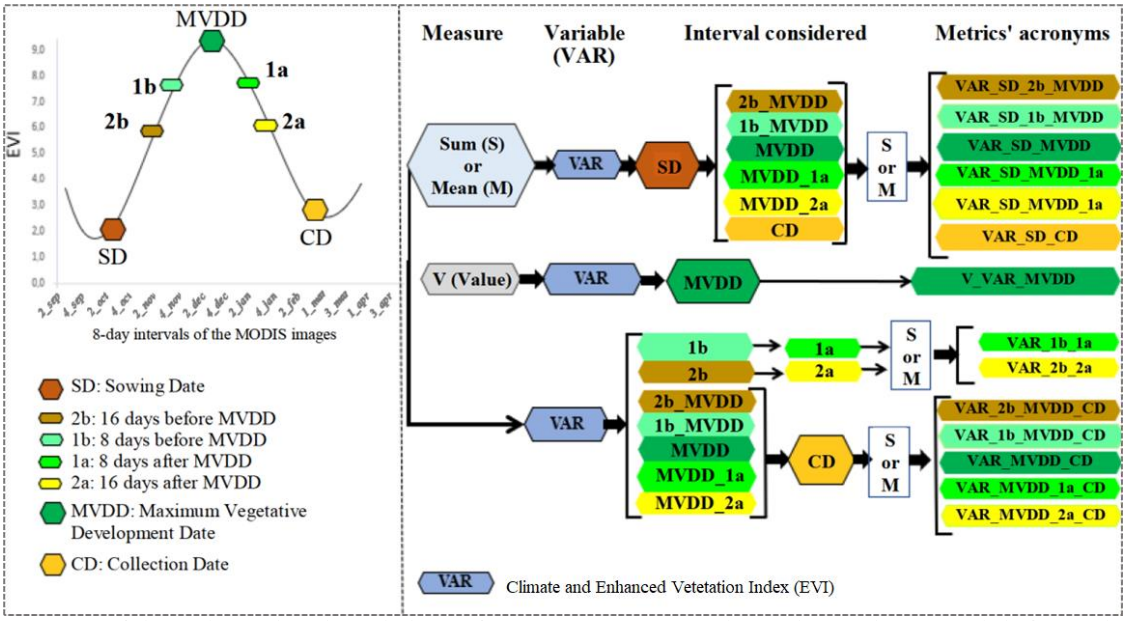

Fig 1. Representation of the soybean phenological phases, from SD, MVDD, CD and 8- and/or 16-day intervals before and after MVDD, and representation of the Acronyms associated with the intervals for the creation of AMVs.

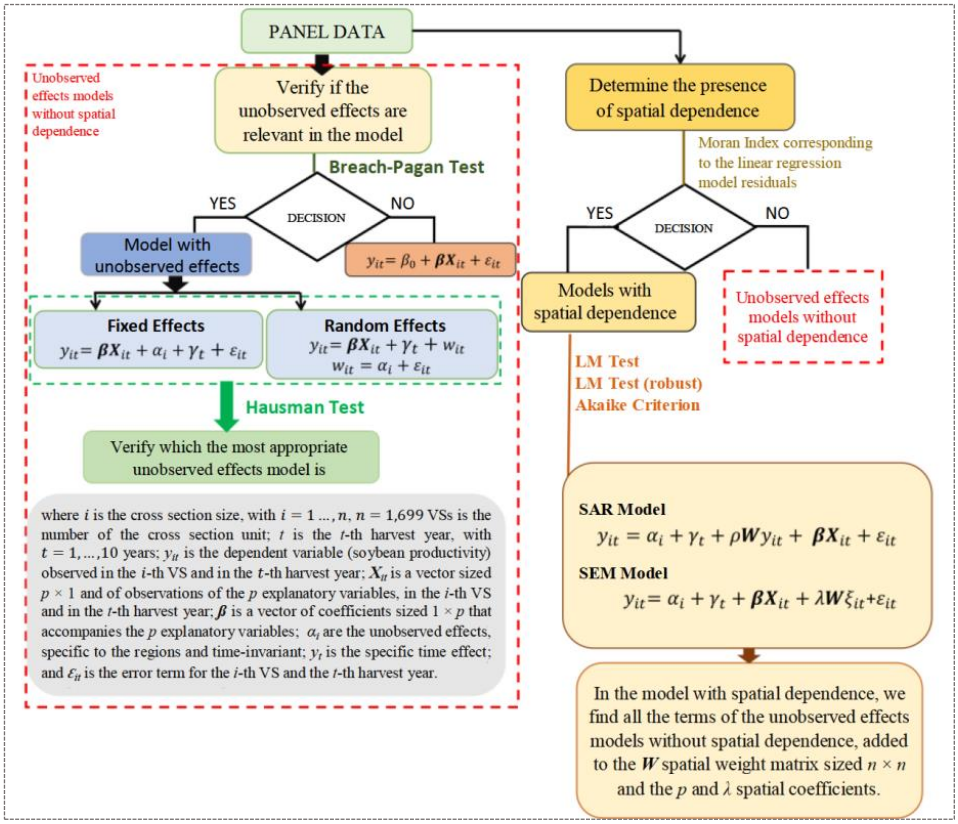

Fig 2. Representation of cross-section, time series and panel data databases (research data). Source: The authors (2022).

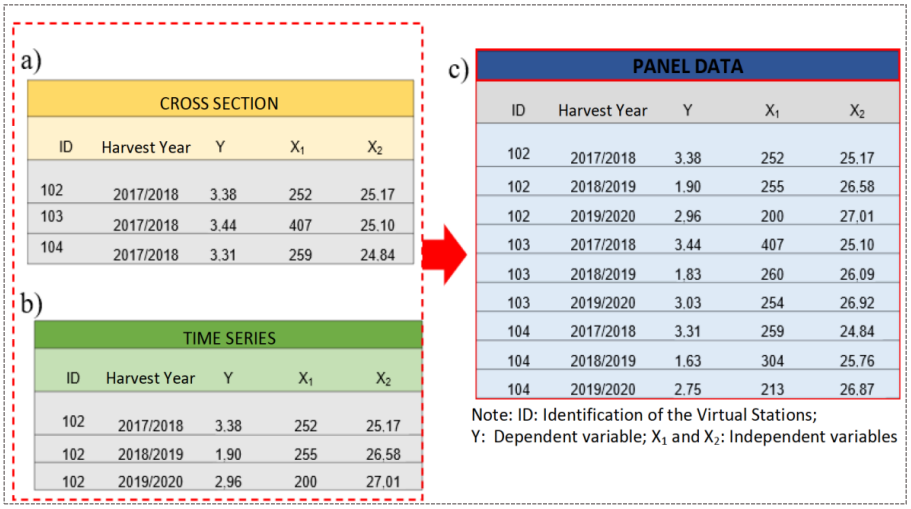

Fig 3. Process corresponding to selection of the model.

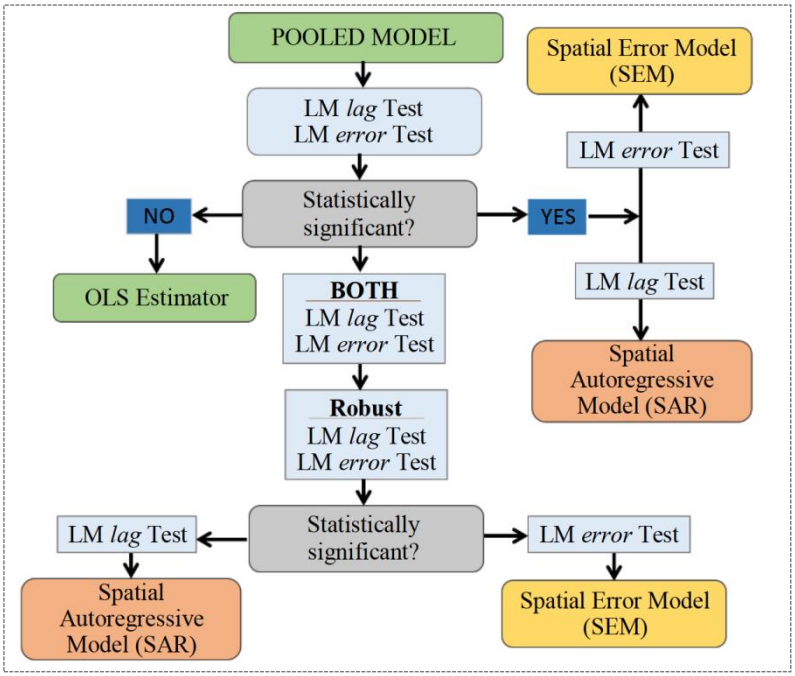

Fig 4. Process to choose the spatial model, based on Anselin et al. (2008) and Ortiz et al. (2022).

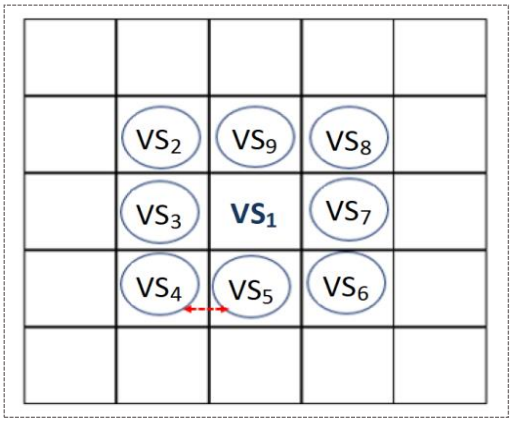

Fig 5. Overflow, VS_1 and $i = 1, \cdots, 9$ are examples of virtual stations: Local Indirect Effect (blue circle) and *feedback* effect (red arrow).

The Variance Inflation Factor (VIF) measure was used for each explanatory variable equation (1).

$$\mathrm{VIF}_u = \frac{1}{1-R_u^2} \quad (1)$$

where R_u^2 is the partial determination coefficient of explanatory variable X_u in relation to the other variables, with $u = 1, \ldots, p$, where p is the number of explanatory variables. VIF values above 10 indicate multicollinearity problems between explanatory variables (Marques et al., 2022).

After verifying the existence of multicollinearity, variables among the 98 AMV were selected using the Non-Negative Garrote (NNG) method, as it presents the best results in papers related to the selection of variables in a panel data set (Vrigazova, 2017).

The NNG estimator is a scaled version of the Ordinary Least Squares (OLS) estimate equation (2) (Yuan and Lin, 2007):

$$arg\ min \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{Z}_d\|^2 + n\,\eta \sum_{u=1}^{p} d_u \quad (2)$$

where $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)^T$, $\boldsymbol{Z}_d = \boldsymbol{Z}_{pn} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{LS}$, in which $\widehat{\boldsymbol{\beta}}^{LS}$ is the vector of the estimated parameters of the p explanatory variables using the Ordinary Least Squares (OLS) method, $\boldsymbol{X}$ is a design matrix sized $p \times n$, where n is the number of observations (1,699 VSs in this study), $\eta > 0$ is the fit parameter obtained through cross-validation, and $d_u = d_u(\eta)\widehat{\boldsymbol{\beta}}^{LS} > 0$ is a shrinkage factor function, with $u = 1, \ldots, p$,. Furthermore, a non-negative garrotte can be illustrated under orthogonal designs, where $\mathbf{XX'} = \boldsymbol{I}_n$. In this case, Equation (2) has an explicit form defined by Equation (3).

$$d_u(\eta) = \left(1 - \frac{\eta}{\widehat{\boldsymbol{\beta}}^{LS^2}}\right) \quad (3)$$

For coefficients whose OLS estimates have values tending to infinity, the shrinkage factor will be close to 1. However, for redundant coefficients, the shrinking factor was 0. Thus, the NNG estimate of the regression coefficient is defined as $d_u\ \widehat{\boldsymbol{\beta}}_u^{LS} = \widehat{\boldsymbol{\beta}}_j^{NG}, u = 1, \ldots, p$ (Yuan and Lin, 2007). To apply the selection method, the data were divided into two subsets: training database (80%) and test database (20%).

Spatial analysis Global Moran Index (I) Equation (4)

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} z_i z_j\ w_{ij}}{S_0 \sum_{i=1}^{n} z_i^2} \quad (4)$$

where $n = 1{,}699$ is the number of VSs in this study; $z_i = (y_i - \bar{y})$ and $z_j = (y_j - \bar{y})$ are the observed values of the productivity variable (y) in the i-th and j-th VSs, respectively, with $i \neq j$ with $i, j = 1, \ldots, n$ centered on the mean ($\bar{y}$) of the variable under study; w_{ij} are the elements of $\mathbf{W}$, which is a symmetric spatial weighting matrix, sized $n \times n$, where its elements represent the proximity between regions i and j regions. If the observation presented a common border, then $w_{ij} = 1$; otherwise, the w_{ij} value is 0. Finally, we have that S_0 is the sum of the w_{ij} elements.