The Roy Bose simultaneous confidence interval approach to multivariate multitemporal pairwise comparisons within and between objects

Máirtín. P. Mac Siúrtáin

School of Biology and Environmental Science
Centre for Agriculture and Food
University College Dublin
Belfield, Dublin 4. Ireland.

Telephone: +353-1-716-7757
Mobile: +353-86-217-3519
Fax: +353-1-716-1104
Email: mms@ucd.ie

1. Introduction

A number of fundamental questions crop up all the time in multivariate multi-temporal investigations some of which include:

Do objects exhibit within and between object signature differences over time?
Which wavelengths are associated with each significant signature differences?
When exactly do significant signature differences occur?
What confidence can the user have in one or all signature differences?

The objective is to illustrate the Roy Bose simultaneous confidence interval approach to multivariate multitemporal pairwise comparisons within and between objects using results from a forest seedling remote sensing experiment.

A forest seedling remote sensing glasshouse experiment was designed to objectively quantify, model and monitor differences in the multitemporal multivariate mean spectral reflectance profiles within and between three conifer species, or objects, which were repeatedly measured on 17 successive Julian dates. The tree seedlings were all uniform one year old at the start of the experiment and fed an optimum liquid nutrient regime of 1:200 Solufeed™.

2. Experimental Design

The levels of the two fixed effects factors considered were species and time. The three species, or objects, were: Sitka spruce (SS), Japanese larch (JL) and lodgepole pine (LP). Sitka spruce and Lodgepole pine are evergreen while Japanese larch is deciduous. The 17 repeated measurements were recorded on 17 Julian dates from day 115 to 320. Four locations within the glasshouse were used as blocks, and the three seedlings types were randomly assigned to three Hiko trays within each block. The experimental unit was a Hiko tray, which held 40 seedlings in a 5x8 grid of cells each of which were filled with a uniform mixture of growing compost and perlite.

The experimental design was a randomized complete block design with three fixed species treatments and four blocks. The 12 experimental units were repeatedly measured on 17 successive dates. The design may also be considered as a two-way fixed effects layout with equal replication and or as split-plot-in-time design with whole plots being the three species and subplots being the 17 repeated measurements. The three levels of species are independent while the 17 levels of repeated measurements are clearly dependent and serially correlated. The five ordered response variables were percent spectral reflectance at wavelengths 480, 560, 640, 720 and 800nm. These defined an ordered five dimensional spectral reflectance profile, or vector, which was repeatedly measured on each experimental on each Julian day. With no missing observations the experimental database was fully balanced.

The experimental indices for this experiment were:

i	= 1, 2 , 3 = r	= number of species or objects: Sitka spruce (SS), Japanese larch (JL), lodgepole pine (LP)
j	= 1,2, ,..., 17	= the number of repeated measurements or Julian dates Julian day 115, 121, …, 302, 320
k	= 1,2,3,4 = n	= the number of blocks or replications: blocks 1-4
h	= 1,2,3,4,5= p	= the number of response variables or components; reflectance: blue (B) 480 nm, green (G) 560 nm red (G) 640 nm, near infra-red (N) 720 nm, near infra-red (M) 800 nm.

The experimental indices determine the number of p-dimensional observations and the degrees of freedom associated with each source of variation.

3. The Method of Analysis

The balanced spectral reflectance profile data were analysed as a split-plot-in-time model using multivariate analysis of variance (MANOVA). The primary multivariate hypothesis of interest was the composite hypothesis for the interaction of species and Julian dates. As expected, the interaction between species and dates was highly significant (Mac Siúrtáin, 2000:304). The interaction hypothesis tested the equality of 51 mean spectral reflectance profiles associated with the 3 species by 17 Julian dates treatment combinations.

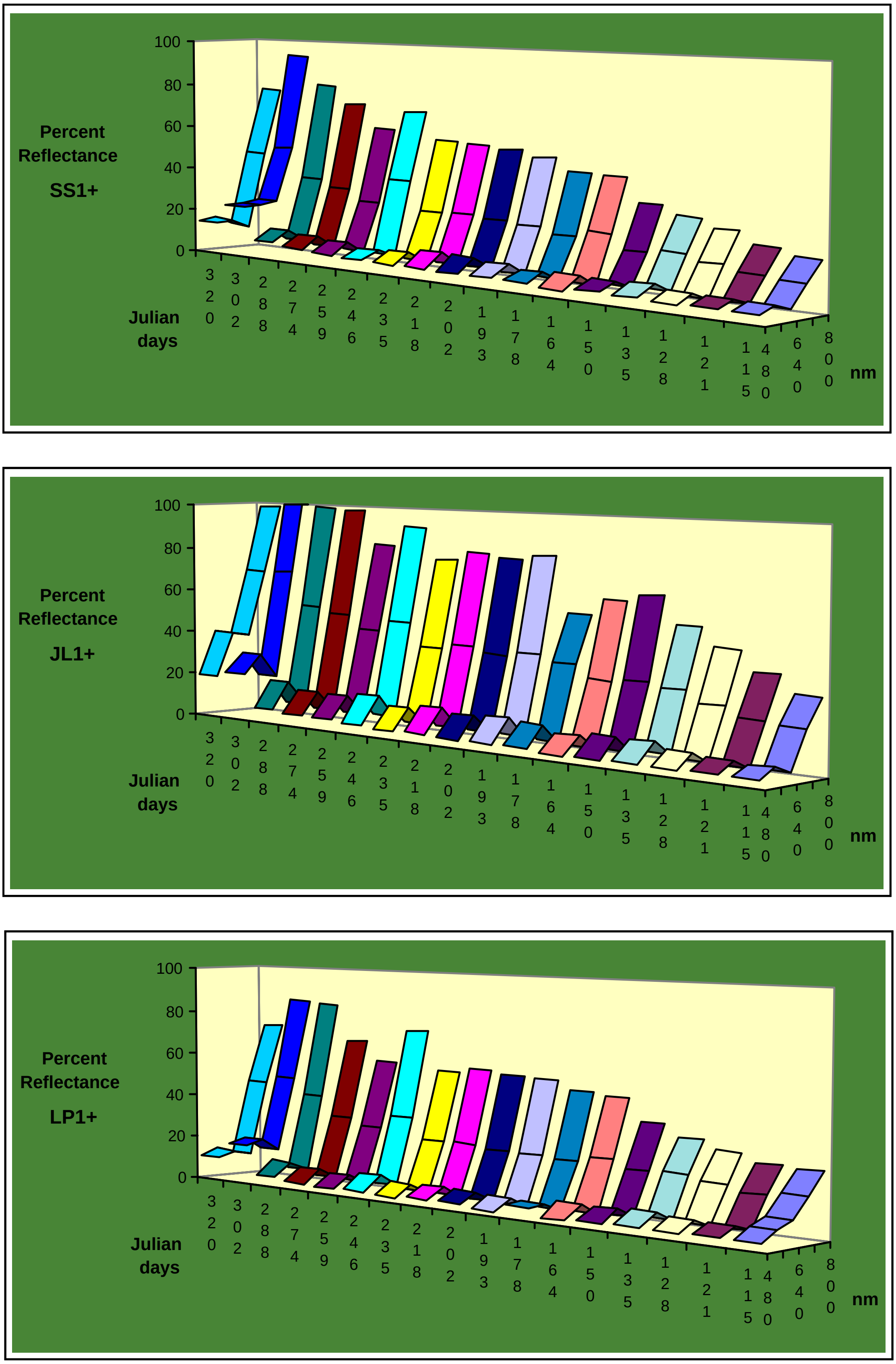
The multitemporal multivariate spectral reflectance profiles for Sitka spruce, Japanese larch and Lodgepole pine are presented (Figure 1). Percent reflectance (%) is on the vertical axis, Julian days is on the horizontal axis and wavelength (nm) is on the z axis. The five dimensional multivariate and multiemportal variation on 17 Julian days for each of the three species is very clearly illustrated (Figure 1).

The MANOVA provides a concise and objective partition of the pxp total, **T**, sums of squares and cross products matrix associated with each identifiable source of variation. The main effect hypotheses of equality of the three species, and the 17 Julian dates mean spectral reflectance profiles have no meaning as the former is confounded with time and the latter confounded with species.

The split-plot-in-time MANOVA provides two error sums of squares and cross products matrices, $\mathbf{E_1}$ and $\mathbf{E_2}$. The $\mathbf{E_1}$ error matrix related to the whole unit treatments or species, while the $\mathbf{E_2}$ error matrix related to the sub unit treatments or time intervals. $\mathbf{E_1}$ and $\mathbf{E_2}$. were used in the construction of all independent i.e. between species Roy-Bose simultaneous confidence intervals (RBSCIs). In contrast $\mathbf{E_2}$. alone was used in the construction of all dependent, i.e. within species RBSCIs (Morrison 1990: 232; Mac Siúrtáin, 2000:280-4).

Two new multivariate concepts of an ordered difference, Θ , and an ordered difference component vector, Δ , allowed the results of every pairwise hypothesis and its composite of five rudimentary component hypotheses to be reported clearly, objectively and with simultaneous confidence.

Figure 1 Multitemporal multivariate spectral reflectance profiles for Sitka spruce, Japanese larch and Lodgepole pine



4. Roy-Bose simultaneous confidence intervals

In the presence of significant species by time interaction Roy's critical test statistic, $\theta_{\alpha,s,m,n}$, alone leads directly to the construction of Roy-Bose simultaneous confidence intervals (RBSCSs) (Roy and Bose, 1953). By inversion RBSCIs lead to the testing of all unique dependent, and independent interaction pairwise comparisons and the associated component hypotheses, for the difference between all five components of the mean spectral reflectance profiles with simultaneous confidence of at least 95 percent. Each pairwise hypothesis is a composite of five rudimentary component hypotheses, each associated with one of the five response variables or wavelengths sampled. All unique within and between species pairwise comparisons for this experiment are presented (Table 1).

Table 1 All unique within and between species pairwise comparisons

Species		Sitka spruce SS	Japanese larch JL	Lodgepole pine LP
	Julian dates	1, 2 , ..., 16, 17	1, 2 , ..., 16, 17	1, 2 , ..., 16, 17
Sitka spruce SS	1 2 : 16 17	Within species pairwise comparisons 136		
Japanese larch JL	1 2 : 16 17	Between species pairwise comparisons 289	Within species pairwise comparisons 136	
Lodgepole pine LP	1 2 : 16 17	Between species pairwise comparisons 289	Between species pairwise comparisons 289	Within species pairwise comparisons 136

For dependent, i.e. within species, pairwise comparisons, there were 136 unique comparisons within each of the three species or 408 dependent comparisons for all species. For independent, i.e. between species, pairwise comparisons there were 289 unique comparisons between each pair of species or 867 independent comparisons. When combined the total number of dependent and independent pairwise comparisons was 1,275. Given that each comparisons involves five wavelengths this resulted in a total of 5100 component comparisons. .

RBSCIs provide simultaneous tests for the h = 1, 2,.. 5 = p component hypotheses of equality of the p response variables or components. If the individual h^{th} 95% confidence interval for the difference includes zero, the h^{th} component null hypothesis cannot be rejected. If the h^{th} 95% simultaneous confidence interval for the difference excludes zero, the h^{th} component null hypothesis may be rejected simultaneously with at least 95% confidence.

5. Concept of an ordered difference, Θ

When the number of response variables is p, the response 1xp vector consists p components; [X_1, X_2, X_3,..., X_p]. Furthermore, when the response variables are commensurable and follow a logical ordered sequence, the response vector consists of p commensurable ordered components.

When considering any pairwise comparison between two p dimensional mean vectors, the composite hypothesis of no difference between two mean vectors, $Ho: \mu_{ij} - \mu_{i'j'} = 0$, can be regarded as the intersection of p rudimentary component hypotheses, as follows:

$$Ho: \mu_{ij} - \mu_{i'j'} = 0 \qquad \rightarrow Ho: \begin{bmatrix} \mu_{ij.1} - \mu_{i'j'.1} \\ \mu_{ij.2} - \mu_{i'j'.2} \\ \mu_{ij.3} - \mu_{i'j'.3} \\ \vdots \\ \mu_{ij.p} - \mu_{i'j'.p} \end{bmatrix} = 0 \qquad \rightarrow \begin{matrix} Ho_{ij.1}: \mu_{ij.1} - \mu_{i'j'.1} = 0 \\ Ho_{ij.2}: \mu_{ij.2} - \mu_{i'j'.2} = 0 \\ Ho_{ij.3}: \mu_{ij.3} - \mu_{i'j'.3} = 0 \\ \vdots \\ Ho_{ij.p}: \mu_{ij.p} - \mu_{i'j'.p} = 0 \end{matrix} .$$

By inversion of the $100(1-\alpha)\%$ RBSCIs, each of the p rudimentary component hypotheses may be tested for inclusion or exclusion of zero. Each of the p rudimentary component hypotheses may or may not be rejected. If, for any pairwise comparison, none of the p rudimentary component hypotheses are rejected this is termed a zero order difference. If any one of the p hypotheses is rejected this is termed a 1^{st} order difference. If any two of the p hypotheses are rejected this is termed a 2^{nd} order difference. In like manner, if all p hypotheses are rejected this is termed a p^{th} order difference. Rejection of at least one of the rudimentary component hypotheses will result in rejection of the composite pairwise comparison. This led to the definition of an ordered difference, Θ, (Mac Siúrtáin, 2000:128) as:

> An ordered difference, Θ, is the number of component null hypotheses rejected at a specified and simultaneous level of confidence, $1-\alpha$, for any pairwise composite hypothesis of no difference between two p dimensional mean spectral reflectance profiles and $0 \le \Theta \le p$.

If Θ=0 for any pairwise comparison, it means that there was no significant difference between any of the p response variables. If Θ = 1, 2, 3, 4, 5 or p, it means that there were 1, 2, 3, 4, 5 or p simultaneously significant differences within the p component RBSCIs. Recall that every pairwise comparison of p components must assume an ordered difference, Θ, between 0 and p, i.e., $0 \le \Theta \le p$.

6. Concept of an ordered difference component vector, Δ

The main limitation of an ordered difference, Θ, is that it does not identify unambiguously which of the p response variables are associated with each ordered difference, Θ. The question arises, for each pairwise comparison, which, if any, of the p response variables, or wavelengths, are actually associated with each ordered difference? This question is answered by identifying unambiguously all of the possible ordered difference components between mean vectors, whose dependent 95% Roy-Bose simultaneous confidence interval includes zero. This led to the definition of an ordered difference component vector, Δ, (Mac Siúrtáin, 2000:130) as:

> An ordered difference component vector, Δ, is a 1xp vector, each ordered element of which may assume a value of 0, when associated with a non

significant difference or X_1, X_2, X_3, X_4, or X_p, when associated with one or more simultaneously significant components differences.

In the present context, the ordered difference component vector, Δ, is a 1xp row vector, each ordered element of which may assume a value of 0, B, G, R, N or M associated with no difference, blue 480nm, green 560nm, red 640nm, near infra-red 720nm or near infra-red 800nm simultaneous components differences respectively. The p elements of each ordered component may assume values from [0, 0, 0, 0, 0] indicating no difference in all of the wavelengths, when $\Theta = 0$, to [B, G, R, N, M] indicating simultaneous and significant differences in all five wavelengths, when $\Theta = 5$. Therefore [0, 0, 0, 0, 0]≤ Δ ≤, [B, G, R, N, M]. The [0, 0, 0, 0, 0] ordered difference component vector, Δ, clearly relates to a zero ordered difference, $\Theta = 0$, while the [B, G, R, N, M] ordered difference component vector, Δ, clearly relates to a fifth ordered difference, $\Theta = 5$.

The [0, 0, 0, 0, 0] ordered difference component vector, Δ, was also represented by a single [0] for ease of presentation and interpretation. For reasons of brevity, the square brackets have been deleted in the presentation of all ordered difference component vectors, Δ.

7. Matrix of ordered difference component vectors, Δ

A graphical means of presenting the ordered differences, Θ, and ordered difference component vectors, Δ, for all pairwise comparisons was required and this led to the definition of matrices of ordered Θ and Δ. Matrices of ordered Θ and Δ have been computed and examples are presented below. The matrices of ordered Θ and Δ, indicate the exact wavelengths associated within all ordered differences, Θ, for each and every pairwise comparison in the experiment. All of the entries within the matrices of ordered Δ will range from 0000 to BGRNM or $0 \le \Delta \le \text{BGRNM}$. The great benefit of these ordered matrices is that one can immediately identify exactly which of the p component wavelengths were associated with each and every simultaneously significant ordered difference, Θ, within and between each of the three tree species.

For example the cxc, i.e. 17x17, matrices of ordered differences, Θ, (Table 2) and associated ordered difference component vectors, Δ, within Sitka spruce (Table 3) are presented. These matrices provide the simultaneously significant results of all dependent pairwise comparisons for the difference between mean spectral reflectance profiles within Sitka spruce. The first row and column identify the exact Julian days involved in each pairwise comparison. Clearly it is illogical to make pairwise comparisons within species on the same Julian day or to repeat comparisons which have already been made. This explains why the diagonal and all lower-diagonal elements of the matrix are blank.

The nonzero elements in Table 3 identify the exact wavelengths within all order difference component vectors, Δ, associated with all order differences, Θ, with simultaneous confidence of at least 95%. It becomes evident that the order differences, Θ, associated with Julian day 302 are involve the changes in the blue, green and red wavelengths, which clearly suggests a change in reflectance within Sitka spruce on that day. In, contrast, the first order differences associated with Julian day 320 are primarily associated with significanct differences in blue reflectance, which suggests that the epicuticular wax associated with Sitka spruce foliage may be associated with simultaneously significance changes in the mean spectral reflectance profiles within the species.

Table 2 Matrix of dependent ordered differences, Θ, within Sitka spruce

Jday	115	121	128	134	150	164	178	193	204	218	235	246	259	274	288	302	320
115		0	0	0	0	0	0	0	0	0	0	0	0	1	1	4	2
121			0	0	0	0	0	0	0	0	0	0	0	0	1	4	2
128				0	0	0	0	0	0	0	0	0	0	0	1	4	1
134					0	0	0	0	0	0	0	0	0	0	0	4	1
150			**SS1+ vs. SS1+**			0	0	0	0	0	0	0	0	0	0	4	1
164							0	0	0	0	0	0	0	0	0	3	1
178								0	0	0	0	0	0	0	0	3	1
193									0	0	0	0	0	0	0	3	1
204										0	0	0	0	0	0	3	1
218											0	0	0	0	0	3	1
235												0	0	0	0	3	1
246													0	0	0	3	1
259														0	0	3	0
274															0	1	0
288																1	0
302																	1
320																	

"0" = zero order difference

Table 3 Matrix of dependent ordered difference component vectors, Δ, within Sitka spruce

Jday	115	121	128	134	150	164	178	193	204	218	235	246	259	274	288	302	320
115		0	0	0	0	0	0	0	0	0	0	0	0	- - - -M	- - - -M	BGR -M	B - - -M
121			0	0	0	0	0	0	0	0	0	0	0	0	- - - -M	BGR -M	B - - -M
128				0	0	0	0	0	0	0	0	0	0	0	- - - -M	BGR -M	B - - - -
134					0	0	0	0	0	0	0	0	0	0	0	BGR -M	B - - - -
150			**SS1+ vs. SS1+**			0	0	0	0	0	0	0	0	0	0	BGR -M	B - - - -
164							0	0	0	0	0	0	0	0	0	BGR - -	B - - - -
178								0	0	0	0	0	0	0	0	BGR - -	B - - - -
193									0	0	0	0	0	0	0	BGR - -	B - - - -
204										0	0	0	0	0	0	BGR - -	B - - - -
218											0	0	0	0	0	BGR - -	B - - - -
235												0	0	0	0	BGR - -	B - - - -
246													0	0	0	BGR - -	B - - - -
259														0	0	BGR - -	0
274															0	B - - - -	0
288																B - - - -	0
302																	B - - - -
320																	

"0" = zero order difference

8. Discussion

It is Roy's greatest root test statistic, $\theta_{\alpha,s,m,n}$, which is based on the union-intersection principle, that leads directly to the construction of RBSCIs for the pairwise difference within and between mean spectral reflectance profiles. By inversion of RBSCIs, it was possible to test all dependent and independent pairwise and component hypotheses, and all with simultaneous confidence of at least, 95 percent. Finally, by application of the new multivariate concepts of an ordered difference, Θ, and an ordered difference component vector, Δ, the component specific results of all dependent and independent pairwise and component hypotheses became apparent when the multitemporal multivariate results are presented as matrices.

The combination of the split-plot-in-time MANOVA and construction of RBSCIs for all 6,275 dependent and independent pairwise and component hypotheses, did objectively detect and pin point the exact response variables or wavelengths associated with all significant component differences, between pairs of mean spectral reflectance profiles within and between one year old Sitka spruce, Japanese larch and lodgepole pine seedlings and all with simultaneous confidence of at least 95 percent.

9. Conclusion

MANOVA and Roy-Bose simultaneous confidence intervals when combined with distribution matrices of ordered differences, Θ, and ordered difference component vectors, Δ, define the he Roy Bose simultaneous confidence interval approach to multivariate multitemporal pairwise comparisons within and between objects. While the objects used in this experiment were species of trees, the methodology may be equally applicable to any class of objects which exhibit multitemporal and multivariate changes in p-dimentsional response profiles or vectors with time. The methodology is most easily applied to balanced data.

10. References

Mac Siúrtáin, M. P. 2000. Multivariate Analysis of Spectral Reflectance Profiles in Forest Seedling Remote Sensing Experiments. PhD Thesis. National University of Ireland. 447p.

Morrison, D. F. 1990. Multivariate Statistical Methods. Third Edition. McGraw-Hill International Editions. Statistics Series. New York. 495p.

Roy, S. N. and R. C. Bose. 1953. Simultaneous Confidence Interval Estimation. Annals of Mathematical Statistics. 24, 513-36.