Comparing a band selection algorithm with PCA feature extraction from hyperspectral reflectance for N monitoring of processing tomato

by

Vito Aurelio Cerasola[a], Mirko Piani[a], Stefano Bona[b], Jochem Verrelst[c], Antonio Persichetti[d], Francesco Orsini[a], Giuseppina Pennisi[a], Luigi Manfrini[a], Giorgio Gianquinto[a]

[a] Department of Agricultural and Food Sciences (DISTAL), Alma Mater Studiorum–University of Bologna, Viale Fanin 44, Bologna 40127, Italy

[b] Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università, 16, Legnaro, PD 35020, Italy

[c] Image Processing Laboratory (IPL), Universitat de València, València, Spain

[d] Archetipo s.r.l., 35129 Padova, Italy

SUPPLEMENTARY MATERIAL

Table S1 Regression performance of the last 10 iterations in the GPR-BAT approach and identification of the optimal bands of each iteration. R^2 and *NRMSE* values are averaged across the four validation folds and reported ± the standard deviation

# Bands	R^2	*NRMSE* (%)	Spectral bands
AGB			
10	0.807±0.158	14.31±5.46	705, 776, 778, 780, 869, 871, 908, 911, 942, 1001
9	0.807±0.158	14.31±5.46	705, 776, 778, 780, 869, 908, 911, 942, 1001
8	0.807±0.158	14.32±5.47	705, 776, 778, 780, 869, 911, 942, 1001
7	0.807±0.158	14.33±5.48	705, 778, 780, 869, 911, 942, 1001
6	0.806±0.159	14.34±5.49	705, 778, 780, 869, 942, 1001
5	0.808±0.156	14.32±5.43	705, 778, 869, 942, 1001
4	0.864±0.061	12.24±3.29	778, 869, 942, 1001
3	**0.868±0.056**	**12.10±3.16**	**869, 942, 1001**
2	0.854±0.049	12.80±2.44	869, 1001
1	0.795±0.034	15.16±1.40	869
N uptake			
10	0.883±0.040	9.99±1.42	734, 747, 778, 780, 782, 809, 811, 818, 822, 869
9	0.883±0.040	9.98±1.43	734, 747, 778, 780, 782, 811, 818, 822, 869
8	0.882±0.039	10.01±1.38	734, 747, 778, 780, 811, 818, 822, 869
7	0.884±0.039	9.94±1.44	734, 747, 778, 780, 811, 818, 869
6	0.885±0.037	9.93±1.37	734, 747, 778, 780, 818, 869
5	0.877±0.065	9.65±2.73	747, 778, 780, 818, 869
4	0.882±0.061	9.47±2.65	747, 778, 818, 869
3	0.891±0.053	9.36±2.74	747, 778, 818
2	**0.885±0.060**	**9.52±3.02**	**747, 778**
1	0.801±0.068	12.92±2.56	778
NNI			
10	0.594±0.118	18.39±1.90	402, 404, 472, 483, 486, 488, 709, 711, 966, 968
9	0.592±0.119	18.40±1.91	402, 404, 472, 486, 488, 709, 711, 966, 968
8	0.593±0.118	18.38±1.89	402, 404, 472, 488, 709, 711, 966, 968
7	0.593±0.117	18.38±1.90	402, 404, 472, 709, 711, 966, 968
6	0.593±0.116	18.41±1.90	402, 404, 472, 709, 711, 966
5	0.594±0.115	18.38±1.88	402, 472, 709, 711, 966
4	0.596±0.118	18.39±1.93	402, 472, 711, 966
3	**0.592±0.123**	**18.45±1.91**	**402, 711, 966**
2	0.342±0.140	22.00±1.87	711, 966
1	0.328±0.161	23.04±2.01	966

Table S2 Regression performance of AGB and N uptake of the last 10 iterations in the GPR-BAT approach with the removal of soil (n=75) and identification of the optimal bands of each iteration. R^2 and *NRMSE* values are averaged across the four validation folds and reported ± the standard deviation

# Bands	R^2	*NRMSE* (%)	Spectral bands
AGB			
10	0.833±0.113	13.31±3.49	581, 583, 585, 587, 778, 869, 871, 873, 955, 1001
9	0.833±0.111	13.29±3.49	583, 585, 587, 778, 869, 871, 873, 955, 1001
8	0.833±0.111	13.30±3.49	583, 587, 778, 869, 871, 873, 955, 1001
7	0.833±0.112	13.30±3.52	583, 587, 778, 871, 873, 955, 1001
6	0.811±0.157	13.97±4.84	583, 587, 778, 873, 955, 1001
5	0.811±0.157	13.98±4.84	587, 778, 873, 955, 1001
4	0.898±0.031	10.47±2.38	778, 873, 955, 1001
3	**0.898±0.031**	**10.47±2.38**	**873, 955, 1001**
2	0.677±0.064	18.82±2.92	955, 1001
1	0.558±0.173	22.43±2.41	955
N uptake			
10	0.847±0.032	12.81±0.57	731, 740, 749, 751, 778, 780, 782, 784, 811, 818
9	0.845±0.037	12.86±0.83	731, 740, 751, 778, 780, 782, 784, 811, 818
8	0.844±0.039	12.90±0.91	731, 740, 751, 778, 780, 782, 811, 818
7	0.869±0.059	11.18±2.62	740, 751, 778, 780, 782, 811, 818
6	0.868±0.059	11.26±2.54	740, 751, 778, 780, 782, 818
5	0.870±0.060	11.13±2.68	740, 751, 780, 782, 818
4	0.889±0.054	10.34±2.96	740, 780, 782, 818
3	**0.889±0.054**	**10.34±2.96**	**740, 782, 818**
2	0.858±0.046	12.34±2.07	740, 782
1	0.778±0.065	15.49±1.05	782

Table S3 Spearman correlation coefficient (ρ) between the GPR standard deviation (σ) and the absolute prediction error (|measured − estimated|) of the out-of-fold predictions obtained from the 4-fold cross-validation, for each model and target variable (p-values in brackets; n = 83 for AGB and N uptake, n = 75 for NNI). A positive correlation indicates that observations with higher model-based uncertainty also showed larger prediction errors; *** indicate significant for $p<0.001$ and ns indicates not significant

Model	AGB	N uptake	NNI
GPR-Full bands	0.460 (***)	0.456 (***)	-0.129 (ns)
GPR-PCA	0.427 (***)	0.463 (***)	-0.059 (ns)
GPR-BAT	-0.076 (ns)	-0.076 (ns)	-0.140 (ns)

Table S4 Bands selected with GPR-BAT models after the truncation of the first and last 10 bands of the spectrum (421-979 nm). R^2 and *NRMSE* (%) are averaged across the four validation folds and reported ± the standard deviation

Crop trait	Spectral bands	R^2	*NRMSE* (%)
AGB	585, 618, 778, 975	0.876±0.056	11.48±2.40
N uptake	729, 782, 818, 869, 891	0.877±0.045	10.35±1.94
NNI	428, 444, 711, 846	0.663±0.050	16.10±1.57