

Prediction of seed coat proportion in lupins by near infrared reflectance spectroscopy

Preliminary results

Daniel Alomar

Facultad de Ciencias Agrarias, Universidad Austral de Chile

Mario Mera

Instituto de Investigaciones Agropecuarias, INIA-Carillanca

Scenario

- The world demand for plant protein and oil increases every year
- The Chilean salmon, poultry and swine industries require increasing amounts of protein
- Lupins are among the cropping options to produce protein in southern Chile

Additional domestication

- Lupins have a high percent of seed weight in the seed coat
- The seed coat has a high fraction of structural carbohydrates and low protein and oil contents, which lowers the nutritional value of the grain
- Lowering seed coat proportion is a priority among traits that require additional domestication

Seed coat proportion

- Genetic variation for seed coat proportion (SCP) in lupin species does exist
- Heritability values for SCP are strong ($H_{BS} = 0.95$ in *L. angustifolius*)
- The prospect to reduce this “primitive” characteristic by breeding is promising

Screening methods

- The standard method for assessing SCP involves manual separation of seed fractions, drying, weighing, which is cumbersome, time-consuming, and it destroys the seeds
- An alternative, rapid, reliable, non-destructive screening method is needed

NIRS

- Near infrared spectroscopy has been successfully used in a number of applications, including chemical and physical properties of grains
- NIRS can be applied on intact samples and it may be a fast and accurate technique, provided that:
 - a broad and representative set of calibration samples is used
 - the reference method is dependable
 - proper regression equations are developed

Material

- 70 lupin samples
 - 55 *L. angustifolius* (cultivated and wild; white, mottled, dark seeds)
 - 8 *L. albus* (all cultivated, white and speckled seeds)
 - 7 *L. luteus* (all cultivated, white and mottled seeds)
- A larger number of materials within each lupin species is being analysed

Reference method for SCP

- Dry seeds were allowed to absorb distilled water for ~2 h
- 20 random seeds per sample were processed
- The hull was removed manually, retaining the embryo with the cotyledons
- The two fractions were dried at 65°C for 12 h and weighed promptly
- SCP was reported as a percentage relative to the weight of the whole seeds

Coefficients of variation for SCP obtained with the reference method

- When estimated from field replications, with a sample of:
 - 10 seeds \rightarrow CV = 4.9%
 - 20 seeds \rightarrow CV = 4.5%
 - 40 seeds \rightarrow CV = 2.2%
- The error involved in SCP measurements is relatively small

Equipment

- NIRSystems 6500 scanning monochromator (Silver Springs, MD)
- PC with WINISI II software (Infrasoft International, Port Matilda, PA)



quartz ring cells

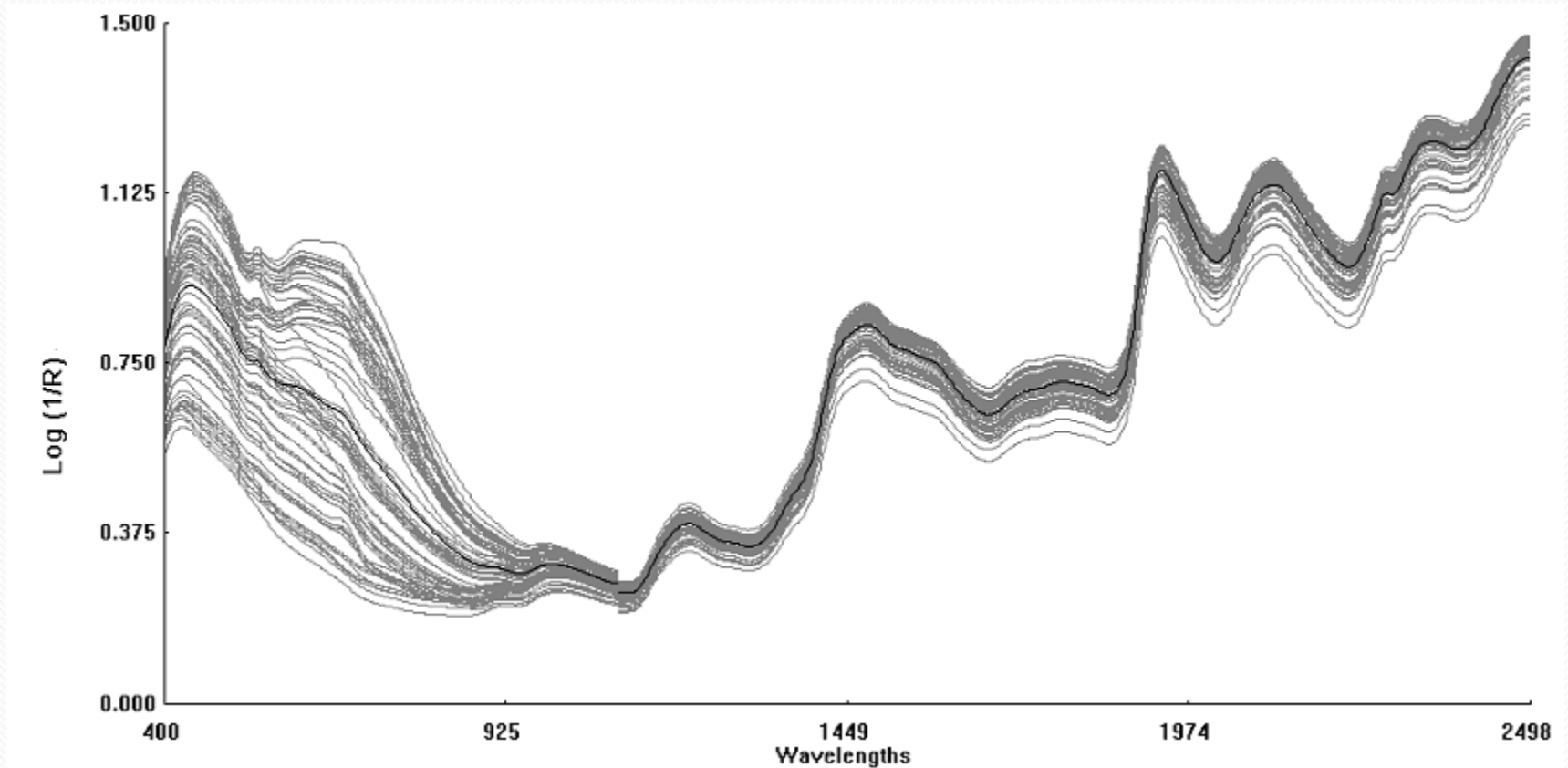


Spinning module

NIRS scanning

- Whole seed samples were scanned
- Spectra were collected in reflectance mode, from 400 to 2500 nm
- Optical data were transformed to microabsorbance units, the reciprocal \log_{10} reflectance ($\log 1/R$), and stored in suitable files

Spectra of 70 samples of lupin grains ($\log 1/R$)



NIR: region of the electromagnetic spectrum from ~ 800 to 2500 nm

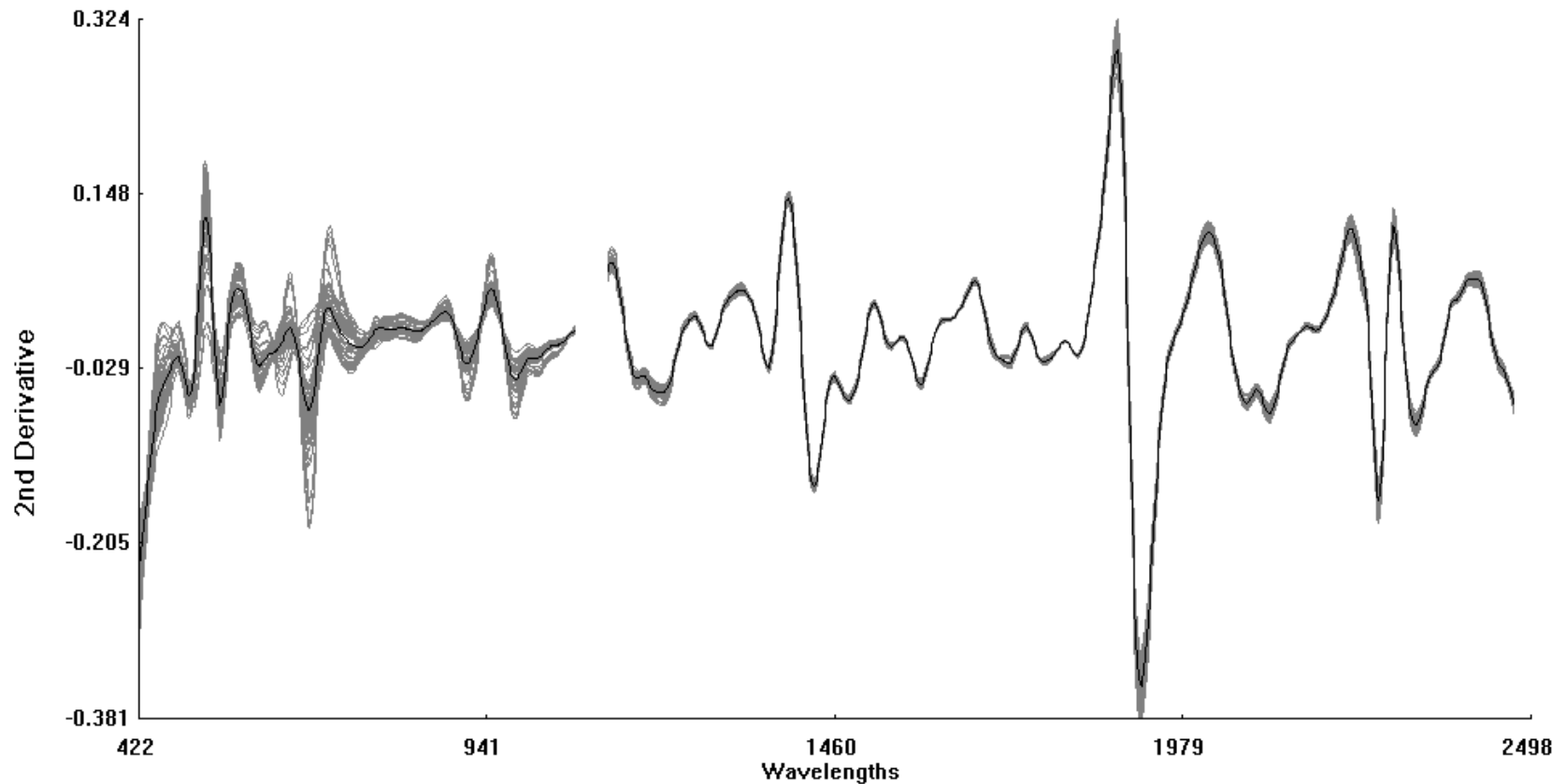
Treatment of spectral data

- Performed by WINISI II software
- Different mathematical treatments of the spectral data (derivative number, subtraction gap, smoothing interval) were tested to extract relevant information, with or without applying options for scatter correction - the “standard normal variate” and “detrend”

Spectra of 70 samples of lupin grains after treatment

Math: 2,8,8; SNV plus Detrend

Final treatment: a second derivative with a subtraction interval of 8 data points,
with a first smooth in a segment of 8 data points



NIR calibration

- Spectral data are confronted with data from the reference method
- Prediction equations were obtained by linear regression choosing the modified partial least squares method
- Reference set moments
 - Mean SCP: 24.3%
 - Standard deviation: 2.97
 - Coefficient of variation: 12.2%
 - Range: 15.97 to 27.42%

Cross-validation

- The prediction equations were checked
- The set of calibration samples was divided into four groups, then one group was excluded and considered as a set for prediction by an equation adjusted with the other three groups of samples
- The process was repeated cyclically with all groups, statistical terms of certainty and uncertainty were computed at each step, and averaged once all the groups of samples had been predicted

Best equation selected

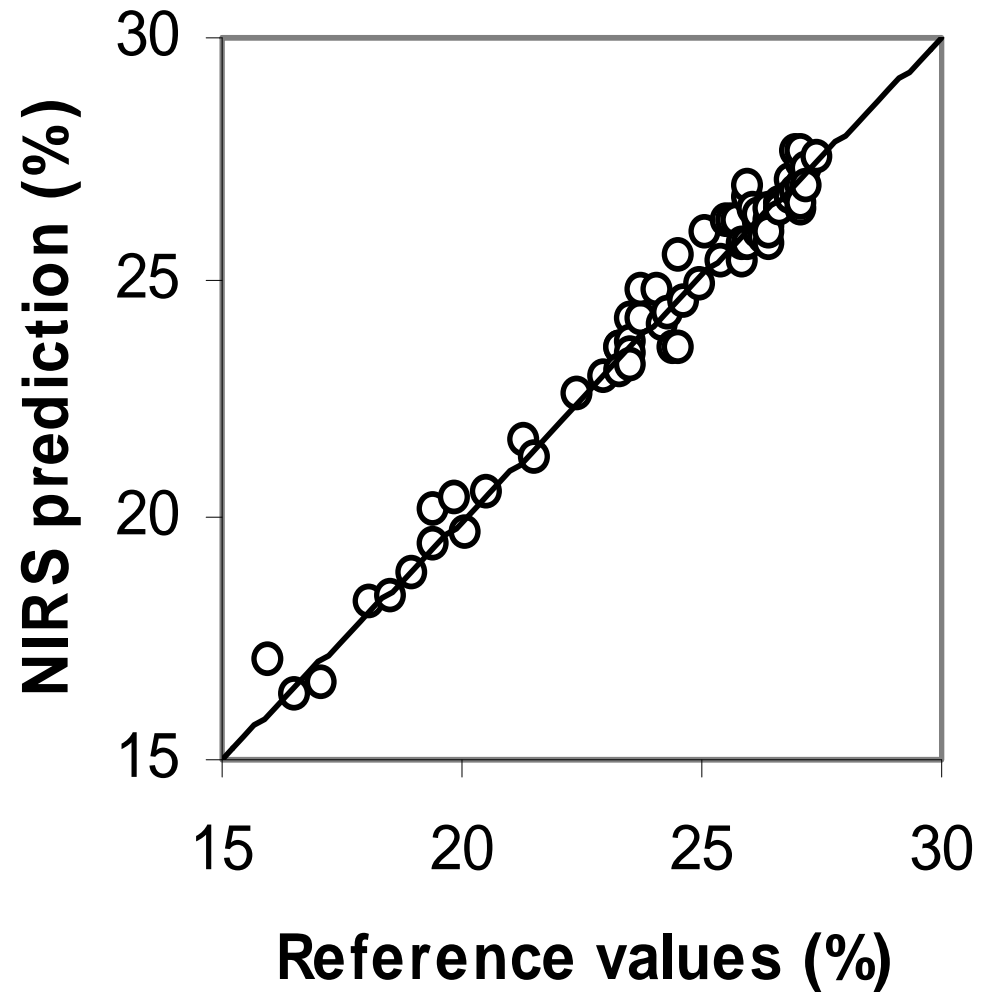
- The best equation was selected according to the highest coefficient of determination (R^2CV) and the lowest standard error of the cross validation ($SECV$)
- The equation selected included 8 terms and showed an **R^2CV of 0.95** and a **$SECV$ of 0.73**

Calibration output

Sample No	Pos	Rank	Actual	Pred	Resid	'T'
angustifolius 306	51	1	26,19	26,20	-0,01	-0,02
angustifolius 407	58	2	26,17	26,19	-0,02	-0,05
angustifolius 301	69	3	26,48	26,51	-0,03	-0,05
angustifolius 409	62	4	27,25	27,28	-0,03	-0,06
angustifolius 304	52	5	27,42	27,39	0,03	0,06
angustifolius 207	44	6	23,52	23,49	0,03	0,06
angustifolius 410	64	7	24,64	24,60	0,04	0,07
angustifolius 203	39	8	26,63	26,59	0,04	0,08
angustifolius 212	37	9	18,52	18,47	0,05	0,09
Taper	10	10	24,30	24,25	0,05	0,09
angustifolius 412	59	11	26,27	26,32	-0,05	-0,10
angustifolius 107	20	12	22,98	22,92	0,06	0,12
angustifolius 113	22	13	19,48	19,41	0,07	0,13
albus AL	3	14	20,53	20,46	0,07	0,14
Wonga	9	15	25,45	25,33	0,12	0,24
angustifolius 111	24	16	27,13	27,26	-0,13	-0,27
albus AS	1	17	17,70	17,56	0,14	0,28
angustifolius 303	53	18	21,35	21,49	-0,14	-0,28
angustifolius 101	29	19	22,47	22,62	-0,15	-0,29
angustifolius 406	63	20	26,86	27,05	-0,19	-0,38
angustifolius 307	50	21	26,89	26,70	0,19	0,38
angustifolius 411	56	22	23,52	23,74	-0,22	-0,44
Rumbo-B	5	23	18,97	18,74	0,23	0,47
angustifolius 405	67	24	23,32	23,56	-0,24	-0,48
angustifolius 308	45	25	25,90	25,65	0,25	0,50
angustifolius 401	61	26	24,25	23,99	0,26	0,52
angustifolius 201	41	27	26,02	25,76	0,26	0,53
Local-822	6	28	17,06	16,79	0,27	0,54
angustifolius 305	55	29	23,78	24,06	-0,28	-0,56
angustifolius 404	60	30	27,10	26,81	0,29	0,57
angustifolius 206	43	31	26,22	25,93	0,29	0,58
Italiano-912	8	32	16,53	16,23	0,30	0,60
angustifolius 109	23	33	26,10	26,42	-0,32	-0,65

Sample No	Pos	Rank	Actual	Pred	Resid	'T'
Lidar	12	34	26,32	26,00	0,32	0,65
Juno	18	35	25,04	24,71	0,33	0,65
angustifolius 108	21	36	26,45	26,10	0,35	0,71
Mielero-osc	16	37	25,70	26,06	-0,36	-0,72
angustifolius 211	33	38	23,36	23,00	0,36	0,73
ang 202 muy poco	38	39	26,69	26,33	0,36	0,73
angustifolius 403	65	40	27,27	26,90	0,37	0,73
angustifolius 311	46	41	25,75	26,18	-0,43	-0,86
angustifolius 210	35	42	26,98	26,55	0,43	0,86
Rose	14	43	21,54	21,11	0,43	0,87
angustifolius 105	28	44	27,11	26,64	0,47	0,93
Typtop-B	4	45	18,08	18,57	-0,49	-0,98
angustifolius 302	54	46	23,57	23,08	0,49	0,98
Merrit	15	47	23,35	23,84	-0,49	-0,99
angustifolius 413	57	48	26,41	25,91	0,50	0,99
angustifolius 205	40	49	24,11	24,61	-0,50	-0,99
angustifolius 209	32	50	27,16	27,70	-0,54	-1,08
angustifolius 112	25	51	26,03	26,57	-0,54	-1,08
angustifolius 309	70	52	19,87	20,45	-0,58	-1,16
angustifolius 310	49	53	26,98	27,57	-0,59	-1,18
angustifolius 213	34	54	23,50	24,12	-0,62	-1,23
angustifolius 204	42	55	27,07	26,42	0,65	1,29
angustifolius 312	48	56	25,84	25,16	0,68	1,36
angustifolius 408	68	57	19,40	20,09	-0,69	-1,38
angustifolius 110	19	58	25,52	26,24	-0,72	-1,43
albus AR	2	59	20,12	19,38	0,74	1,47
angustifolius 313	47	60	26,49	25,74	0,75	1,49
angustifolius 402	66	61	26,48	25,71	0,77	1,54
angustifolius 104	27	62	26,04	26,85	-0,81	-1,62
angustifolius 103	30	63	23,81	24,64	-0,83	-1,65
angustifolius 102	26	64	24,40	23,56	0,84	1,67
angustifolius 106	31	65	25,10	25,95	-0,85	-1,69
Mister	11	66	24,61	25,50	-0,89	-1,78
Legat	13	67	24,54	23,57	0,97	1,92
Marroqui-1230	7	68	15,97	17,05	-1,08	-2,16

Prediction of Lupin SCP



Discussion

- Calibration is considered adequate for screening purposes if the ratio of SECV to the standard deviation of the reference data is > 2.5 or 3.0
- In this case, $SD/SECV = 2.97/0.73 = 4.1$
- Therefore, NIRS has a high potential to screen lupin seeds for SCP



GOBIERNO DE CHILE
INIA



UNIVERSIDAD AUSTRAL DE CHILE
FACULTAD DE CIENCIAS AGRARIAS

Funding for this research was granted by CONICYT Chile
through FONDECYT Project 1070232

Thank you



ANOVA - 13 angustifolius genotypes

SCP – reference method

Response cub sem (%)

Whole Model

genotipo

rep

Summary of Fit

RSquare	0,946226
RSquare Adj	0,92382
Root Mean Square Error	0,626841
Mean of Response	25,11112
Observations (or Sum Wgts)	52

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	15	248,90659	16,5938	42,2309
Error	36	14,14548	0,3929	Prob > F
C. Total	51	263,05206		<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
genotipo	12	12	247,18497	52,4235	<.0001
rep	3	3	1,72162	1,4605	0,2416

LSMeans Differences Tukey HSD

Alpha= 0,050 Q= 3,54271

Level		Least Sq Mean
208-2-2-2	A	27,204360
P 22757 GS144	A	26,769493
Gunguru x Naturalizado	A	26,724127
P 28361 Wonga	A	26,653983
P 22785 GS173	A	26,445070
209-1-1-4-2	A	26,435904
P 26974 Gunguru	A	26,285061
P 26596 GRC5052A	A B	26,004126
P 28492 Chile1	B C	24,446708
P 26496 GRC5018A	C D	24,136376
P 26971 Danja	C D	23,309821
P 28130 MAR6825A	D	22,714119
THM 11257	E	19,315412

Levels not connected by same letter are significantly different

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	24,817205	0,17385447	24,8172
2	25,208373	0,17385447	25,2084
3	25,116580	0,17385447	25,1166
4	25,302322	0,17385447	25,3023

ANOVA - 13 angustifolius genotypes SCP – NIRS predictions

Response cub sem (%) NIRS predicted

Whole Model

Summary of Fit

RSquare	0,960744
RSquare Adj	0,94392
Root Mean Square Error	0,517191
Mean of Response	25,06824
Observations (or Sum Wgts)	51

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	15	229,12451	15,2750	57,1056
Error	35	9,36203	0,2675	Prob > F
C. Total	50	238,48654		<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
genotipo	12	12	228,40069	71,1564	<.0001
rep	3	3	0,49891	0,6217	0,6057

genotipo

LSMeans Differences Tukey HSD

Alpha= 0,050 Q= 3,54944

Level	Least Sq Mean
208-2-2-2	A 27,290000
Gunguru x Naturalizado	A B 27,002500
P 22785 GS173	A B 26,525000
P 22757 GS144	A B 26,495000
209-1-1-4-2	A B 26,420000
P 26974 Gunguru	A B 26,285000
P 28361 Wonga	A B C 25,964653
P 26596 GRC5052A	B C 25,840000
P 28492 Chile1	C D 24,630000
P 26496 GRC5018A	D E 24,055000
P 26971 Danja	E 23,140000
P 28130 MAR6825A	E 22,835000
THM 11257	F 19,605000

Levels not connected by same letter are significantly different

rep

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	25,013846	0,14344298	25,0138
2	24,987585	0,15120218	24,9142
3	25,090769	0,14344298	25,0908
4	25,242308	0,14344298	25,2423

Reference method for SCP

Data	20-seed sample	40-seed sample	10-seed sample	35-seed sample
Experiment	angustifolius	angustifolius	angustifolius	angustifolius
Sites	1	4	4	1
Years	1 (2005)	1 (2000)	2 (2000-01)	1 (2002)
Entries	13	14	14	960
Reps	4	3	4	1
N	52	168	448	960
Mean SCP (%)	25.26	22.66	23.47	23.02
SD	2.52	1.40	1.92	1.38
SE	0.35	0.11	0.09	0.04
CV-SCP (%)	4.5	2.2	4.9	6.0
CV-MSW (%)	6.1	10.0	10.2	12.8

MSW = mean seed weight

Outliers

- Two passes of elimination of outlier samples were applied, i.e., samples with high residuals of predicted versus reference values (critical T value: 2.5)
- The equation was finally constructed with data from 68 samples

Calibration

- The “standard normal variate” option scales each spectrum so that the standard deviation is 1.0 to help reduce particle size effects
- The “detrend” option removes the linear and quadratic curvature of each spectrum