

1
2 10.1071/SR13077_AC
3 © CSIRO 2013

4 **Supplementary Material: *Soil Research*, 2013, 51(7&8), 577–595.**

5

6

7 **Predicting contents of carbon and its component fractions in Australian soils from**
8 **diffuse reflectance mid-infrared spectra**

9

10 *J. A. Baldock^{A,B}, B. Hawke^A, J. Sanderman^A, and L. M. Macdonald^A*

11

12 ^ACSIRO Land and Water/Sustainable Agriculture Flagship, PMB 2 Glen Osmond, SA 5064,
13 Australia.

14 ^BCorresponding author. Email: jeff.baldock@csiro.au

15 Table 1. Summary statistics for the distributions of OC, IC, TC and TN contents obtained for soil
 16 samples collected from the various regions and included in the generation of project specific
 17 MIR/PLSR algorithms.

	Variable	Count	Average	Standard deviation	Median	Minimum	Maximum	Skewness	Kurtosis
NSW_DECCW	OC (mg C/g)	1752	16.5	13.5	12.1	2.2	121.0	2.4	8.7
	IC (mg C/g)				no data available				
	TC (mg C/g)				no data available				
	TN (mg N/g)				no data available				
NSW_MCMA	OC (mg C/g)	948	11.0	8.8	7.3	0.8	48.7	1.6	2.6
	IC (mg C/g)				no samples contained carbonate				
	TC (mg C/g)				no samples contained carbonate				
	TN (mg N/g)				940	0.9			
NSW_Orgill	OC (mg C/g)	284	18.4	17.6	12.0	2.5	101.0	2.0	4.5
	IC (mg C/g)				insufficient samples with carbonate data				
	TC (mg C/g)				insufficient samples with carbonate data				
	TN (mg N/g)				284	1.6			
NSW	OC (mg C/g)	3548	10.8	7.4	8.5	0.4	53.4	1.4	2.7
	IC (mg C/g)	416	1.8	2.7	1.2	0.0	22.2	4.5	26.3
	TC (mg C/g)	3547	11.0	7.5	8.7	0.4	53.4	1.4	2.6
	TN (mg N/g)	3151	0.9	0.6	0.7	0.1	4.3	1.5	3.0
Qld_cropping	OC (mg C/g)	1723	13.2	5.4	12.3	1.3	32.0	0.3	-0.9
	IC (mg C/g)	377	2.5	2.6	1.7	0.0	22.3	3.2	16.7
	TC (mg C/g)	1723	13.7	5.2	13.0	1.3	34.3	0.2	-0.8
	TN (mg N/g)	1388	1.0	0.5	0.9	0.1	3.1	0.6	0.1
Qld_rangeland	OC (mg C/g)	664	6.3	4.4	4.6	1.3	34.3	2.5	7.7
	IC (mg C/g)	294	4.3	8.2	1.5	0.0	56.4	3.7	15.0
	TC (mg C/g)	664	8.2	7.2	5.6	1.3	60.8	3.0	12.5
	TN (mg N/g)	344	0.5	0.3	0.4	0.2	2.2	2.4	6.8
SA	OC (mg C/g)	963	9.7	5.2	8.6	1.3	37.9	1.4	3.0
	IC (mg C/g)	272	11.8	14.1	6.2	0.0	68.7	1.7	2.6
	TC (mg C/g)	963	13.1	10.0	10.6	1.3	76.0	2.5	8.6
	TN (mg N/g)	963	0.9	0.4	0.8	0.2	3.0	1.1	1.6
Tas	OC (mg C/g)	1640	28.1	18.4	23.3	2.3	120.3	1.2	1.4
	IC (mg C/g)				insufficient samples with carbonate data				
	TC (mg C/g)				insufficient samples with carbonate data				
	TN (mg N/g)				1582	2.2			
Vic_Gippsland	OC (mg C/g)	738	20.1	16.0	15.8	1.2	132.6	1.9	5.5
	IC (mg C/g)				no data available				
	TC (mg C/g)				no data available				
	TN (mg N/g)				no data available				
Vic	OC (mg C/g)	3278	13.5	15.1	7.7	0.2	143.5	2.5	7.9
	IC (mg C/g)	1033	6.1	8.2	3.2	0.0	55.4	3.0	10.3
	TC (mg C/g)	3278	15.5	15.1	10.4	0.2	143.5	2.2	6.8
	TN (mg N/g)	2188	1.3	1.5	0.8	0.0	18.4	2.7	13.3
WA	OC (mg C/g)	4205	12.0	14.2	6.7	0.0	147.6	2.6	9.3
	IC (mg C/g)				insufficient samples with carbonate data				
	TC (mg C/g)				insufficient samples with carbonate data				
	TN (mg N/g)				insufficient samples with total N data				

18

19

20 Table 2. Calibration and validation model parameters and statistics derived for predicting OC, POC,
 21 HOC and ROC contents of the 312 fractionated soils using the following modelling approaches: 1)
 22 random cross validation, 2) full cross validation, 3) test set validation using samples selected from a
 23 PCA analysis of spectra, and 4) test set validation using samples selected from a PLSR analysis. For
 24 the random cross validation approach 20 segments containing 15 or 16 samples were used. The full
 25 cross validation used a leave one out approach incorporating all 312 samples. For the two test set
 26 validation approaches the optimum 250 samples, as identified using a Kennard-Stone algorithm,
 27 were used for calibration and the remaining 62 samples were used for validation.
 28

Analyte	Method	Model	Slope	Intercept	r2	RMSE
sqrt SOC	Random Cross Validation	Calibration	0.930	0.271	0.930	0.448
	Full Cross Validation	Calibration	0.930	0.272	0.930	0.448
	Test Set Validation (PCA)	Calibration	0.929	0.276	0.929	0.453
	Test Set Validation (PLS)	Calibration	0.926	0.288	0.926	0.475
	Random Cross Validation	Validation	0.926	0.287	0.926	0.465
	Full Cross Validation	Validation	0.926	0.286	0.924	0.469
	Test Set Validation (PCA)	Validation	0.932	0.337	0.925	0.465
	Test Set Validation (PLS)	Validation	0.964	0.169	0.942	0.353
sqrt POC	Random Cross Validation	Calibration	0.867	0.223	0.867	0.398
	Full Cross Validation	Calibration	0.867	0.223	0.867	0.398
	Test Set Validation (PCA)	Calibration	0.853	0.242	0.853	0.424
	Test Set Validation (PLS)	Calibration	0.853	0.243	0.853	0.432
	Random Cross Validation	Validation	0.851	0.246	0.844	0.432
	Full Cross Validation	Validation	0.851	0.248	0.844	0.432
	Test Set Validation (PCA)	Validation	0.837	0.379	0.829	0.417
	Test Set Validation (PLS)	Validation	0.824	0.382	0.805	0.408
sqrt HUM	Random Cross Validation	Calibration	0.894	0.300	0.894	0.380
	Full Cross Validation	Calibration	0.894	0.300	0.894	0.380
	Test Set Validation (PCA)	Calibration	0.887	0.324	0.887	0.401
	Test Set Validation (PLS)	Calibration	0.897	0.290	0.898	0.368
	Random Cross Validation	Validation	0.889	0.313	0.884	0.399
	Full Cross Validation	Validation	0.889	0.312	0.885	0.397
	Test Set Validation (PCA)	Validation	0.889	0.271	0.921	0.301
	Test Set Validation (PLS)	Validation	0.851	0.472	0.905	0.377
sqrt ROC	Random Cross Validation	Calibration	0.859	0.274	0.859	0.314
	Full Cross Validation	Calibration	0.859	0.274	0.859	0.314
	Test Set Validation (PCA)	Calibration	0.856	0.280	0.856	0.328
	Test Set Validation (PLS)	Calibration	0.852	0.287	0.852	0.326
	Random Cross Validation	Validation	0.855	0.280	0.852	0.324
	Full Cross Validation	Validation	0.852	0.290	0.848	0.327
	Test Set Validation (PCA)	Validation	0.881	0.236	0.823	0.297
	Test Set Validation (PLS)	Validation	0.896	0.237	0.853	0.304

29

30

31 Table 3. Summary statistics for the state project based MIR/PLSR models constructed for sqrt_TC.

State Group		PLSR Factors	No. samples	Slope	Offset	r	R ²	RMSE	SE	Bias	s	RPD
NSW_DECCW	Calibration Validation											
No data available												
NSW_MCMA	Calibration Validation											
No samples contained carbonate												
NSW_Orgill ¹	Calibration Validation											
Insufficient samples with carbonate data												
NSW	Calibration	6	2295	0.866	0.423	0.931	0.87	0.391	0.391	-0.000	1.07	2.7
	Validation		1252	0.878	0.381	0.928	0.86	0.391	0.391	0.003	1.05	2.7
Qld_cropping	Calibration	7	1153	0.908	0.334	0.953	0.91	0.223	0.223	0.000	0.74	3.3
	Validation		570	0.887	0.399	0.954	0.91	0.215	0.215	-0.011	0.72	3.3
Qld_rangeland	Calibration	7	442	0.935	0.174	0.967	0.94	0.253	0.253	0.000	1.00	3.9
	Validation		222	0.947	0.117	0.967	0.94	0.25	0.25	-0.024	0.99	3.9
SA	Calibration	6	649	0.937	0.214	0.968	0.94	0.294	0.294	0.000	1.17	4.0
	Validation		314	0.968	0.127	0.962	0.92	0.314	0.316	0.017	1.13	3.6
Tas	Calibration Validation											
Insufficient samples with carbonate data												
Vic_Gippsland	Calibration Validation											
No data available												
Vic	Calibration	9	2189	0.952	0.170	0.976	0.95	0.369	0.369	0.000	1.69	4.6
	Validation		1089	0.948	0.183	0.977	0.95	0.352	0.353	-0.002	1.64	4.6
WA	Calibration Validation											
Insufficient samples with carbonate data												

32 ¹ due to the smaller number of samples, a random cross validation approach was used rather than test set validation thus the number of samples included
33 in the calibration and validation sets were equivalent.

34

35 Table 4. Summary statistics for the state project based MIR/PLSR models constructed for sqrt_IC.

State Group		PLSR Factors	No. samples	Slope	Offset	r	R ²	RMSE	SE	Bias	s	RPD
NSW_DECCW	Calibration Validation	No data available										
NSW_MCMA	Calibration Validation	No samples contained carbonate										
NSW_Orgill	Calibration Validation	Insufficient samples with carbonate data										
NSW	Calibration	8	267	0.846	0.168	0.920	0.85	0.309	0.310	0.000	0.79	2.5
	Validation		149	0.911	0.135	0.934	0.87	0.293	0.301	0.039	0.81	2.7
Qld_cropping	Calibration	8	250	0.879	0.172	0.937	0.88	0.228	0.229	-0.000	0.67	2.9
	Validation		124	0.823	0.267	0.944	0.89	0.247	0.247	0.016	0.73	2.9
Qld_rangeland	Calibration	4	197	0.947	0.087	0.973	0.95	0.273	0.273	-0.000	1.19	4.3
	Validation		97	0.945	0.131	0.976	0.95	0.286	0.295	0.038	1.33	4.5
SA	Calibration	4	188	0.973	0.077	0.986	0.97	0.300	0.386	0.000	1.83	6.1
	Validation		84	0.970	0.147	0.975	0.95	0.399	0.420	0.057	1.80	4.4
Tas	Calibration Validation	Insufficient samples with carbonate data										
Vic_Gippsland	Calibration Validation	No data available										
Vic	Calibration	5	673	0.929	0.151	0.964	0.93	0.343	0.344	-0.000	1.29	3.7
	Validation		360	0.946	0.125	0.959	0.92	0.363	0.364	0.010	1.27	3.5
WA	Calibration Validation	Insufficient samples with carbonate data										

36

37

38 Table 5. Summary statistics for the state project based MIR/PLSR models constructed for sqrt_TN.

State Group		PLSR Factors	No. samples	Slope	Offset	r	R ²	RMSE	SE	Bias	s	RPD
NSW_DECCW	Calibration											
	Validation											
No data available												
NSW_MCMA	Calibration	8	626	0.950	0.045	0.975	0.95	0.073	0.073	0.000	0.33	4.5
	Validation		324	0.895	0.094	0.943	0.89	0.110	0.11	-0.001	0.33	3.0
NSW_Orgill ¹	Calibration	4	284	0.876	0.144	0.936	0.88	0.178	0.179	0.000	0.51	2.8
	Validation		284	0.873	0.149	0.931	0.87	0.185	0.185	0.001	0.51	2.7
NSW	Calibration	10	2042	0.868	0.120	0.932	0.87	0.106	0.106	0.000	0.29	2.8
	Validation		1109	0.871	0.113	0.932	0.87	0.106	0.106	-0.003	0.29	2.8
Qld_cropping	Calibration	10	931	0.898	0.100	0.947	0.90	0.073	0.073	0.000	0.23	3.1
	Validation		457	0.892	0.101	0.948	0.90	0.073	0.073	-0.005	0.23	3.1
Qld_rangeland ¹	Calibration	6	344	0.876	0.088	0.936	0.88	0.066	0.066	0.000	0.19	2.9
	Validation		344	0.872	0.09	0.931	0.87	0.069	0.069	-0.000	0.19	2.7
SA	Calibration	11	649	0.913	0.081	0.955	0.91	0.067	0.067	0.000	0.23	3.4
	Validation		314	0.877	0.113	0.938	0.88	0.074	0.074	-0.001	0.22	2.9
Tas	Calibration	7	1046	0.858	0.202	0.926	0.86	0.168	0.168	0.000	0.45	2.7
	Validation		536	0.853	0.205	0.916	0.84	0.171	0.171	-0.006	0.43	2.5
Vic_Gippsland	Calibration											
	Validation											
No data available												
Vic	Calibration	9	1465	0.941	0.059	0.97	0.94	0.129	0.129	0.000	0.55	4.2
	Validation		723	0.888	0.112	0.937	0.89	0.186	0.186	-0.004	0.53	2.9
WA	Calibration											
	Validation											
Insufficient samples with total N data												

39 ¹ due to the smaller number of samples, a random cross validation approach was used rather than test set validation thus the number of samples included
 40 in the calibration and validation sets were equivalent.
 41
 42