













ground	
Laboratory-based vis-NIR measurement	Proximally sensed vis-NIR measurement
Sample collection, Transportation, Drying, Grinding, and Sieving	In situ measurement Rapid !
Controlled indoor condition	Environmental effects (such as ambient light, soil moisture content, structure, temperature, dust, contamination by stones and excessive residues)
	🐺 Ma





























Wold et al. (1983)Suykens et al. (2002)PLSR aims to link the centred response varable vector, y, to the matrix of centred predictors, X, through k latent variables (factors) by:LS-SVM is an optimized algorithm based on standard SVM. It is an is a kernel-based learning methods use an implicit mapping of the input data in a high dimensional feature space, a special type of hyperplane defined by a kernel function, in which a regression model is built. The LS-SVM uses nonlinear regression function: $\mathbf{y} = \mathbf{b}_0 + \mathbf{x}_i \hat{\mathbf{b}}_i$ Where \mathbf{b}_0 is the intercept and $\hat{\mathbf{b}}_i$ regression vectors.Where $\mathbf{K}(\mathbf{x}, \mathbf{x}_i)$ is defined by the kernel function kernel (RBF), which is the typical general-purpose	PLSR	LS-SVM
PLSR aims to link the centred response transmitted algorithm based on standard SVM. predictors, X , through <i>k</i> latent variables (factors) by: $\begin{aligned} \mathbf{X} = \mathbf{t}_1 \mathbf{p}_1' + + \mathbf{t}_k \mathbf{p}_k' + \mathbf{E}_k \\ \mathbf{y} = \mathbf{t}_1 q_1 + + \mathbf{t}_k q_k + \mathbf{f}_k \\ \mathbf{y} = \mathbf{t}_1 q_1 + + \mathbf{t}_k q_k + \mathbf{f}_k \\ \mathbf{y} = \mathbf{t}_1 \mathbf{p}_1' + + \mathbf{t}_k \mathbf{p}_k' + \mathbf{E}_k \\ \mathbf{y} = \mathbf{t}_1 \mathbf{q}_1 + + \mathbf{t}_k \mathbf{p}_k' + \mathbf{E}_k \\ \mathbf{y} = \mathbf{t}_1 \mathbf{q}_1 + + \mathbf{t}_k \mathbf{p}_k' + \mathbf{E}_k \\ \mathbf{y} = \mathbf{t}_1 \mathbf{q}_1 + + \mathbf{t}_k \mathbf{p}_k + \mathbf{f}_k \\ \mathbf{y} = \mathbf{t}_1 \mathbf{q}_1 + + \mathbf{t}_k \mathbf{p}_k' + \mathbf{E}_k \\ \mathbf{y} = \mathbf{t}_1 \mathbf{q}_1 + + \mathbf{t}_k \mathbf{p}_k' + \mathbf{E}_k \\ \mathbf{y} = \mathbf{t}_1 \mathbf{q}_1 + + \mathbf{t}_k \mathbf{p}_k' + \mathbf{E}_k \\ \mathbf{y} = \mathbf{t}_1 \mathbf{q}_1 + + \mathbf{t}_k \mathbf{p}_k' + \mathbf{E}_k \\ \mathbf{y} = \mathbf{t}_1 \mathbf{q}_1 + + \mathbf{t}_k \mathbf{p}_k' + \mathbf{E}_k \\ \mathbf{y} = \mathbf{t}_1 \mathbf{q}_1 + + \mathbf{t}_k \mathbf{q}_k + \mathbf{q}_k \\ there transmitters are transmissional feature space, a special type of the input data in a high dimensional feature space, a special type of the input data in a high dimensional feature space, a special type of the input data in a high dimensional feature space, a special type of the input data in a high dimensional feature space, a special type of the input data in a high dimensional feature space, a special type of the input data in a high dimensional feature space, a special type of the input data in a high dimensional feature space, a special type of the input data in a high dimensional feature space, a special type of the input data in a high dimensional feature space. a special type of the input data in a high dimensional feature space a special type of the input data in a high dimensional feature space. a special type of the input data in a high dimensional feature space a special type of the input data in a high dimensional feature space a special data basis function kernel (RBF), which is the typical general-purpose the special space of the input data in a high dimensional feature special special space of the special space of the input data in a special special spe$	Wold et al. (1983)	Suykens et al. (2002)
regression vectors. function. We used radial basis function kernel (RBF), which is the typical general-purpose	PLSR aims to link the centred response varable vector, y , to the matrix of centred predictors, X , through <i>k</i> latent variables (factors) by: $\mathbf{X} = \mathbf{t}_1 \mathbf{p}'_1 + \dots + \mathbf{t}_k \mathbf{p}'_k + \mathbf{E}_k$ $\mathbf{y} = \mathbf{t}_1 q_1 + \dots + \mathbf{t}_k q_k + \mathbf{f}_k$ After the model parameters are estimated, they can be combined into the final prediction model as $\hat{\mathbf{y}} = \boldsymbol{b}_0 + \mathbf{x}_i \hat{\mathbf{b}}_i$ Where \boldsymbol{b}_n is the intercept and $\hat{\mathbf{b}}_i$ are	LS-SVM is an optimized algorithm based on standard SVM. It is an is a kernel-based learning algorithm. The kernel-based learning methods use an implicit mapping of the input data in a high dimensional feature space, a special type of hyperplane defined by a kernel function, in which a regression model is built. The LS-SVM uses nonlinear regression function: $y(x) = \sum_{i=1}^{n} a_i K(x, x_i) + b_0$ Where $K(x, x_i)$ is defined by the kernel
kernel: $K(x,x_i) = e^{-(x-x_i ^2)/2\sigma^2}$	regression vectors.	function. We used radial basis function kernel (RBF), which is the typical general-purpose kernel: $K(x,x_i) = e^{-(\ x-x_i\ ^2)/2\sigma^2}$

ART 1: Non-linea	ar data mining algori	thm				
			Assessment	t of predic	ction accuracies :	
Unit Mean	SD Medium Max	Min	(Chang et	al., 2001)	
			RPD >2.0 :	good,		
OC g/kg 16.93	8.52 14.57 36.29	4.12	RPD of 1.4-	-2.0 : mod	lerate	
			NPD \1.4.1	пассерга	ible.	
Tab	le 3-3 prediction acc	curacy of	f soil organi	ic carbo	n (g/kg)	
A 1	Measurement	calibrat	ion (n=78)	valid	ation (n=26)	÷
Algorithms	Conditions	R ²	RMSE	R ²	RMSE RPD	
DI CD	Laboratory-based	0.865	3.121	0.809	3.955 2.22	
FLOR	Proximally sensed	0.726	4.448	0.744	4.598 1.91	٦
LS-SVM	Proximally sensed	0.999	0.180	0.799	4.230 2.08	
(Ji et al., 2014)						
					and the second	















-		dataset	n	mean	St.dev	medium	max	min	Skew	
	00	all	104	16.93	8.52	14.57	36.29	4.12	412.2	
	UC .	calibration	70	17.82	9.07	14.96	36.29	4.13	394.14	
	g/kg	validation	34	15.09	7.05	14.23	33.62	4.12	273.04	
Table 4-4.	. PLSR	prediction accu	racies o	of SOC co	ntent usin Validati	g EPO-, DS-a	and PDS-	transfer R	rred field s	oectra
Table 4-4. Method	. PLSR Cali	prediction accu bration Datase (N=70)	racies o	of SOC co t*	ntent usin Validati (N	g EPO-, DS-a on Dataset I=34)	and PDS-	transfer R /log ₁₀	MSE (OC %)	pectra RPI
Table 4-4. Method	Cali	prediction accu bration Datase (N=70) Lab70	racies o	of SOC con t*	ntent usin Validati (N Li	ng EPO-, DS-a on Dataset I=34) ab34	and PDS- R ² 0.86	transfer R /log ₁₀ 0	MSE 0 (OC %)	RPI
Table 4-4. Method Original	Cali	prediction accu bration Datase (N=70) Lab70 Field70	racies o	of SOC con t* /	ntent usin Validati (N La Fie	g EPO-, DS-a on Dataset I=34) ab34 eld34	R ² 0.86	transfer R /log ₁₀ 0	MSE 1 (OC %) 0.099	pectra RPI 2.31 1.91
Table 4-4. Method Original EPO	Cali	prediction accu bration Datase (N=70) Lab70 Field70 ield 70_EPO	racies o	t* / / 50	ntent usin Validati (N L: Fie Field	g EPO-, DS-4 on Dataset I=34) ab34 eld34 34_EPO	R ² 0.86 0.78 0.80	transfer R /log ₁₀ 0 0	MSE 0 (OC %) 0.099 0.119 0.105	Pectra RPI 2.31 1.91 2.18



















		Sp	ectral pretreatment			
	≽ kenn 420-242	25 nn	n ≽log(1/Refl	ectance)	
	Saviziky-Gol		monthing (2 ± 11) Mean	center		
	> 1st derivatival	ays. ™t#a	NOUTING (2111) SINCU	Conc.		
	➤1 [™] derivauve ₃	ど1天	* MIG *			
	L	eave	-one-out cross validation			
	N	Num	her of PLSR factors=15			
T L	In 6 2 Wali datian atati				SOC) with meaning	
Tab	le 5-3 Validation statis	tics f	or predictions of soil organic o	arbon (SOC) with spectra	
Tab	ble 5-3 Validation statis from t	stics f he Cl	for predictions of soil organic on the soil spectral database (C	carbon (CSSD).	SOC) with spectra	
Tab	ole 5-3 Validation statis from t	tics f	for predictions of soil organic of hinese soil spectral database (C	carbon (CSSD).	SOC) with spectra RMSE	
Tab Methods	ole 5-3 Validation statis from t	tics f he Ch t*	for predictions of soil organic of hinese soil spectral database (Control of the soil of t	carbon (CSSD). R ²	SOC) with spectra RMSE / log ₁₀ (OC g/kg)	RPD
Tab Methods	ole 5-3 Validation statis from t Calibration dataset	tics f he Ch t*	or predictions of soil organic c hinese soil spectral database (C Validation dataset Lab 104	carbon (CSSD). R ² 0.813	SOC) with spectra RMSE / log ₁₀ (OC g/kg) 0.114	RPD 2.11
Tab Methods Original	ele 5-3 Validation statis from t Calibration dataset CSSD	stics f he Ch t*	or predictions of soil organic contract of soil organic contract of soil spectral database (Contract of the source	carbon (CSSD). R ² 0.813	SOC) with spectra RMSE / log ₁₀ (OC g/kg) 0.114	RPD 2.11
Tab Methods Original	cssb	tics f he Cl t*	or predictions of soil organic c hinese soil spectral database (C Validation dataset Lab 104 Field 104	carbon (CSSD). R ² 0.813 0.139	SOC) with spectra RMSE / log ₁₀ (OC g/kg) 0.114 1.042	RPD 2.11 0.23
Tab Methods Original	ele 5-3 Validation statis from t Calibration dataset CSSD EPO transferred	t*	ior predictions of soil organic of hinese soil spectral database (O Validation dataset Lab 104 Field 104	carbon (CSSD). R ² 0.813 0.139	SOC) with spectra RMSE / log ₁₀ (OC g/kg) 0.114 1.042	RPE 2.11 0.23
Tab Methods Original EPO	calibration dataset Calibration dataset CSSD EPO transferred	t* t* f 50	or predictions of soil organic c hinese soil spectral database (C Validation dataset Lab 104 Field 104 EPO transferred Field 104	carbon (CSSD). R ² 0.813 0.139 0.571	SOC) with spectra RMSE / log ₁₀ (OC g/kg) 0.114 1.042 0.161	RPE 2.11 0.23 1.50
Tab Methods Original EPO	CSSD CSSD CSSD	t* t* t* 50	ior predictions of soil organic c hinese soil spectral database (C Validation dataset Lab 104 Field 104 EPO transferred Field 104	carbon (<u>CSSD).</u> R² 0.813 0.139 0.571	SOC) with spectra RMSE / log ₁₀ (OC g/kg) 0.114 1.042 0.161	RPE 2.11 0.23 1.50
Tab Methods Original EPO DS	Calibration statis from t Calibration dataset CSSD EPO transferred CSSD CSSD	stics f he Cl t* / / 50 50	ior predictions of soil organic c hinese soil spectral database (C Validation dataset Lab 104 Field 104 EPO transferred Field 104 DS transferred Field 104	carbon (<u>CSSD).</u> R² 0.813 0.139 0.571 0.800	SOC) with spectra RMSE / log ₁₀ (OC g/kg) 0.114 1.042 0.161 0.117	RPI 2.11 0.23 1.50 2.06

1.7	field104 RPD=0.23			predict soil	carbo	on rapidly	
(1.4 1.1 1.1 0.8 1.1 0.8	lab104 RPD=2.11	¢	a'8	atment ≻log ₁₀ (-11) ≻Mean ≻SNV	(1/Reflet center	ectance) r	
9 0.5 0.2 -0.1 -0.4 0.	5 0.7 0.9 1.1	* * * * * * * * * * * * * * *	8 1.5 1.7	s validation actors=15 of soil organic of trail database ((carbon ((SOC) with spectra	
	Observed log. (O	lak	g-1)				
Methods	Calibration dataset	t*	Validati	on dataset	\mathbb{R}^2	RMSE / log ₁₀ (OC g/kg)	RPI
Methods	Calibration dataset	t*	Validati Lal	on dataset b 104	R ²	RMSE / log ₁₀ (OC g/kg) 0.114	RPI 2.11
Methods Original	Calibration dataset	t*	Validati Lal Fiel	on dataset b 104 id 104	R ² 0.813 0.139	RMSE / log ₁₀ (OC g/kg) 0.114 1.042	RPI 2.11 0.23
Methods Original EPO	Calibration dataset	t* / / 50	Validati Lal Fiel EPO transfe	on dataset b 104 ld 104 rred Field 104	R ² 0.813 0.139 0.571	RMSE / log ₁₀ (OC g/kg) 0.114 1.042 0.161	RPI 2.11 0.23 1.50
Methods Original EPO DS	Calibration dataset CSSD EPO transferred CSSD CSSD	t* / / 50 50	Validati Lal Fiel EPO transfe DS transfer	on dataset b 104 id 104 rred Field 104 rred Field 104	R ² 0.813 0.139 0.571 0.800	RMSE / log ₁₀ (OC g/kg) 0.114 1.042 0.161 0.117	RPI 2.11 0.23 1.50 2.00









ible 4. Compa direct s	rison a	mong the performa	nces of ex	ternal para	ameter o	rthogonalizatio	n (EPO),
	R	equirements	EPO	DS	PDS	Spiking	
If	transfer	samples are needed	Y	Y	Y	Y	
		Ji et al.		50	20		
		(2015b)	-	(82, 42)	(82, 42)	-	
		Viscarra Rossel et al.				74	
Nu	mber	(2009)	-	-	-	(892, 39)	
	of	Minasny et al.	60				
tra	insfer	(2011)	(271, 20)	-	-	-	
sai	nples	Ge et al.	177				
ne	eded	(2014)	(2017, 58)	-	-	-	
		Ji et al.	50	50		50	
		(2015a)	(1581, 54)	(1581, 54)	-	(1581, 54)	
	Spectr	al pre-treatment	N	Ν	Y or N	N	
	Re	-calibration	Y	N	Y or N	Y	



H		
	Table of contents	
	Background oil sands, tailings, remediation	
	Infrared Spectroscopy Reflectance NIK vs ALK-MIK	
	Objective	
	Methodology Samples, instrumentation, spectra pre-treatment	
	Results Soliclassification, PLS calibration, prediction of unknowns	
	Conclusion	























































