Supplementary material

Table S1. Statistical analysis of training sets of the MLR for the grape TSS content, when the entire datasets were used. The analyses when the datasets from variable selection were used, are given in the supplementary material.

Parameter	**Dataset 1** harvest 1 2016		**Dataset 2** harvest 2 2016		**Dataset 3** both harvests 2016		**Dataset 4** both harvests 2016 and 2017	
	train set	test set	train set	test set	train set	test set	train set	test set
N	41	11	32	9	74	19	132	33
Mean	17.35	18.08	18.62	18.93	17.82	18.82	17.67	17.01
Median	17.63	18.60	18.86	18.74	17.90	18.86	17.84	17.46
Min	13.44	15.18	14.07	16.12	13.44	14.84	10.18	11.05
Max	20.93	20.71	22.42	21.58	22.42	21.58	22.42	20.34
Range	7.49	5.53	8.35	5.46	8.98	6.74	12.24	8.88
Standard deviation	1.68	1.83	1.97	1.56	1.88	1.76	2.17	0.27
Coefficient of variation	0.10	0.10	0.10	0.08	0.10	0.09	0.12	0.13

Table S2. Performance of Partial Least Squares (PLS) models for TSS prediction from the entire dataset of spectra.

Parameter	**Dataset 1 harvest 1 2016**		**Dataset 2 harvest 2 2016**		**Dataset 3 both harvests 2016**		**Dataset 4 both harvests 2016 and 2017**	
	c	**cv**	**c**	**cv**	**c**	**cv**	**c**	**cv**
Coefficient of determination (R^2)	0.922	0.751	0.866	0.189	0.967	0.740	0.927	0.727
Mean	17.51	17.49	18.69	18.74	18.03	18.04	17.54	17.57
Standard deviation (SD)	1.671	1.617	1.767	2.181	1.872	1.854	2.126	2.056
Mean squared error (MSE)	0.238	0.756	0.482	2.924	1.121	0.944	0.355	1.331
Root mean squared error (RMSE)	0.488	0.869	0.694	1.710	0.348	0.972	0.596	1.154
Standard error of Calibration* (SEC)	0.234	0.226	0.279	0.345	0.195	0.193	0.166	0.161
Residual Prediction Deviation (for calibration) (RPD)	3.570	2.003	2.735	1.110	5.473	1.960	3.705	1.914

c = for calculated values; cv = for cross-validation values

* Calibration of the model is its construction from the X_train set of spectra

Table S3. Performance of Partial Least Squares (PLS) models for TSS prediction from the dataset opt_Xc.

Parameter	opt_Xc for harvest 1 of 2016		opt_Xc for harvest 2 of 2016		opt_Xc for both harvests of 2016		opt_Xc for both harvests of 2016 and 2017	
	c	cv	c	cv	c	cv	c	cv
R^2	0.998	0.919	1.000	0.871	0.981	0.899	0.957	0.884
Mean	17.51	17.51	18.69	18.75	18.03	18.04	17.54	17.53
SD	1.739	1.642	1.898	1.729	1.887	1.835	2.160	2.108
MSE	0.006	0.245	0.000	0.466	0.067	0.366	0.212	0.563
RMSE	0.078	0.495	0.020	0.683	0.260	0.605	0.460	0.751
SEC	0.244	0.230	0.300	0.273	0.197	0.191	0.169	0.165
RPD	22.391	3.514	93.702	2.779	7.334	3.146	4.797	2.942

c = for calculated values; cv = for cross-validation values

Table S4. Performance of Partial Least Squares (PLS) models for TSS prediction from the dataset absxcoefs.

Parameter	absxcoefs for harvest 1 of 2016		absxcoefs for harvest 2 of 2016		absxcoefs for both harvests of 2016		absxcoefs for both harvests of 2016 and 2017	
	c	cv	c	cv	c	cv	c	cv
R^2	0.998	0.919	1.000	0.871	0.981	0.899	0.957	0.884
Mean	17.51	17.51	18.69	18.75	18.03	18.04	17.54	17.53
SD	1.739	1.642	1.898	1.729	1.887	1.835	2.160	2.108
MSE	0.006	0.245	0.000	0.466	0.067	0.366	0.212	0.563
RMSE	0.078	0.495	0.020	0.683	0.260	0.605	0.460	0.751
SEC	0.244	0.230	0.300	0.273	0.197	0.191	0.169	0.165
RPD	22.391	3.514	93.702	2.779	7.334	3.146	4.797	2.942

c = for calculated values; cv = for cross-validation values

Table S5. Performance of MLR models for TSS prediction from the entire dataset of spectra.

Parameter	Dataset 1 harvest 1 2016		Dataset 2 harvest 2 2016		Dataset 3 both harvests 2016		Dataset 4 both harvests 2016 and 2017	
	c	p	c	p	c	p	c	p
R^2	1.000	0.722	1.000	-3.314	1.000	0.543	1.000	0.746
Mean	17.35	18.09	18.62	20.70	17.82	19.04	17.67	17.01
SD	1.682	1.676	1.977	3.246	1.885	2.287	2.17	2.070
MSE	0.000	0.932	0.000	10.500	0.000	1.194	0.000	1.146
RMSE	0.000	0.965	0.000	102.46	0.000	1.092	0.000	1.070
SEC	0.266	0.530	0.355	1.147	0.220	0.539	0.189	0.366
RPD	$> 10^{13}$	1.870	$> 10^{13}$	0.481	$> 10^{13}$	1.480	$> 10^{13}$	1.986

c = for calculated values; p = for predicted values

Table S6. Performance of MLR models for TSS prediction from the opt_Xc dataset of spectra.

Parameter	Dataset 1 harvest 1 2016		Dataset 2 harvest 2 2016		Dataset 3 both harvests 2016		Dataset 4 both harvests 2016 and 2017	
	c	p	c	p	c	p	c	p
R^2	1.000	0.911	1.000	0.503	1.000	0.570	0.999	-5.693
Mean	17.35	18.08	18.62	19.56	17.82	19.03	17.64	15.86
SD	1.682	1.853	1.977	1.682	1.885	2.282	2.170	6.799
MSE	0.000	0.297	0.000	1.209	0.000	1.341	0.001	34.728
RMSE	0.000	0.544	0.000	1.099	0.000	1.158	0.031	5.893
SEC	0.266	0.586	0.355	0.594	0.220	0.537	0.189	1.202
RPD	$> 10^{13}$	3.360	$> 10^{13}$	1.418	$> 10^{13}$	1.526	83.438	0.386

c = for calculated values; p = for predicted values

Table S7. Performance of MLR models for TSS prediction from the absxcoefs dataset of spectra.

Parameter	Dataset 1 harvest 1 2016		Dataset 2 harvest 2 2016		Dataset 3 both harvests 2016		Dataset 4 both harvests 2016 and 2017	
	c	p	c	p	c	p	c	p
R^2	1.000	0.968	1.000	0.569	1.000	0.687	0.999	-5.693
Mean	17.35	18.04	18.62	19.50	17.82	19.02	17.67	15.86
SD	1.682	1.854	1.977	1.777	1.885	2.168	2.170	6.799
MSE	1.640	0.105	0.000	1.048	0.000	0.977	0.001	34.728
RMSE	1.280	0.324	0.000	1.023	0.000	0.988	0.031	5.893
SEC	0.266	0.586	0.355	0.628	0.220	0.511	0.189	1.202
RPD	$> 10^{13}$	5.653	$> 10^{13}$	1.523	$> 10^{13}$	1.788	83.438	0.386

c = for calculated values; p = for predicted values

Table S8. Selected wavenumbers (cm^{-1}) most correlated to the TSS content and in agreement with published literature on the vibration wavenumbers of sugar O-H and C-H bonds.

Most correlated to sugars (from bibliography)	Among all 110 selected wavenumbers	Among top 30 most correlated wavenumbers
5848-6452	6101.89, 5831.89, 5893.61, 6410.45, 6449.02, 6449.02, 6225.31, 6402.74, 6140.46, 6132.74, 6132.74, 6202.17	5415.33, 6418.17, 6233.03 6109.6, 6101.89
10162.6	10236.7, 10074.7, 10221.2	10236.7, 10074.7
10416.67	10498.9, 10398.7, 10676.4	10676.4, 10498.9
10989	11031.2, 11015.8, 11023.5, 11000.4, 10853.8, 10799.8,10761.2	11031.2, 11015.8, 11154.7
9881.42-9784.74	9688.96	