Supplementary materials captions

S1 Mahalanobis distance of 270 samples in invasive PLS model

S2 Mahalanobis distance of 360 samples in non-invasive PLS model

S3 Raw NIR spectra of six medicinal herbs in powder

S4 Performance parameters of gallic acid invasive PLS correction models with different pretreatment methods

S5 Performance parameters of amygdalin invasive PLS correction models with different pretreatment methods

S6 Performance parameters of albiflorin invasive PLS correction models with different pretreatment methods

S7 Performance parameters of paeoniflorin invasive PLS correction models with different pretreatment methods

S8 Performance parameters of ferulic acid invasive PLS correction models with different pretreatment methods

S9 Performance parameters of naringin invasive PLS correction models with different pretreatment methods

S10 Performance parameters of neohesperidin invasive PLS correction models with different pretreatment methods

S11 Performance parameters of gallic acid non-invasive PLS correction models with different pretreatment methods

S12 Performance parameters of amygdalin non-invasive PLS correction models with different pretreatment methods

S13 Performance parameters of albiflorin non-invasive PLS correction models with different pretreatment methods

S14 Performance parameters of paeoniflorin non-invasive PLS correction models with different pretreatment methods

S15 Performance parameters of ferulic acid non-invasive PLS correction models with different pretreatment methods

S16 Performance parameters of naringin non-invasive PLS correction models with different pretreatment methods

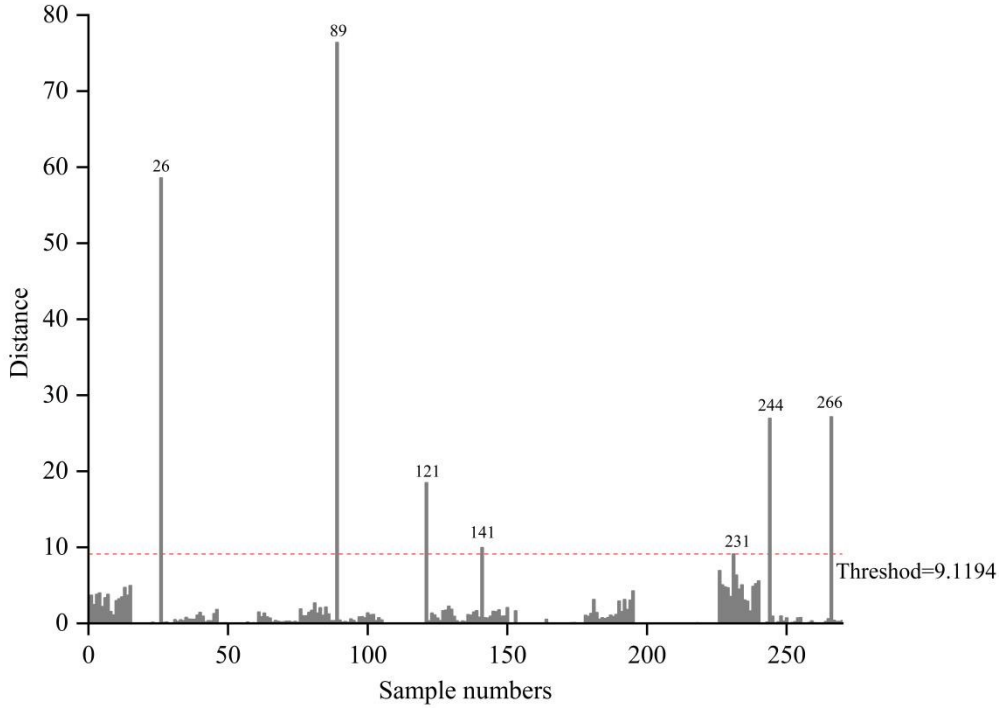
S17 Performance parameters of neohesperidin non-invasive PLS correction models with different pretreatment methods

S18 Performance parameters of each component content invasive model of powder treated by different variable screening methods

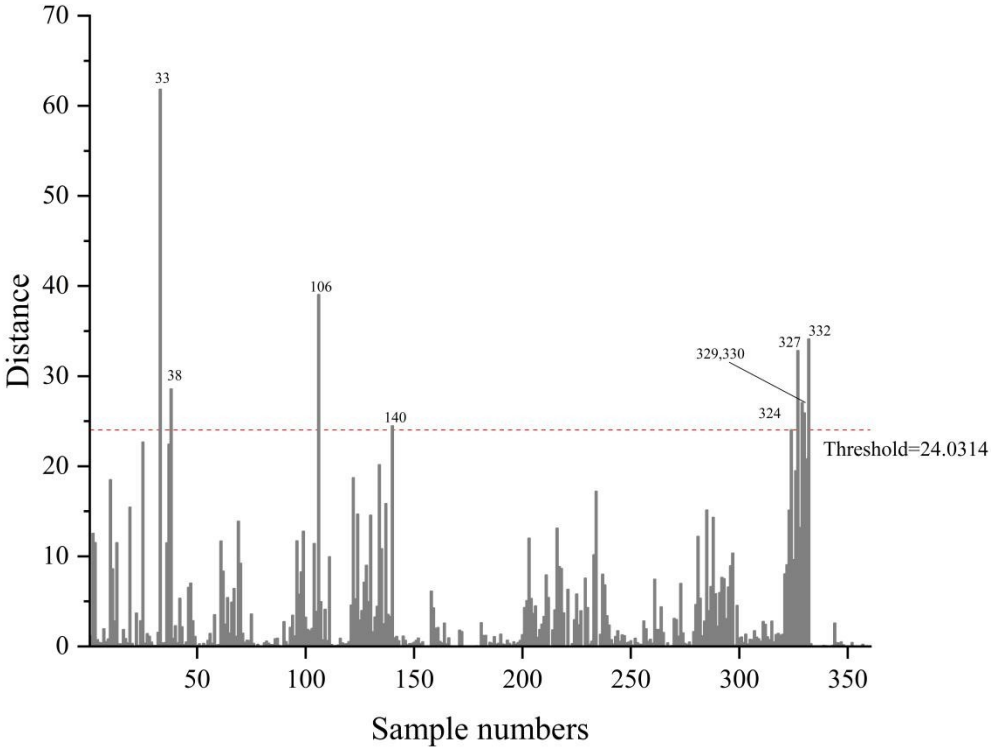
S19 Performance parameters of each component content non-invasive model of powder treated by different variable screening methods

S20 Comprehensive powder properties of 21 batches Xuefu Zhuyu Capsule raw powder samples

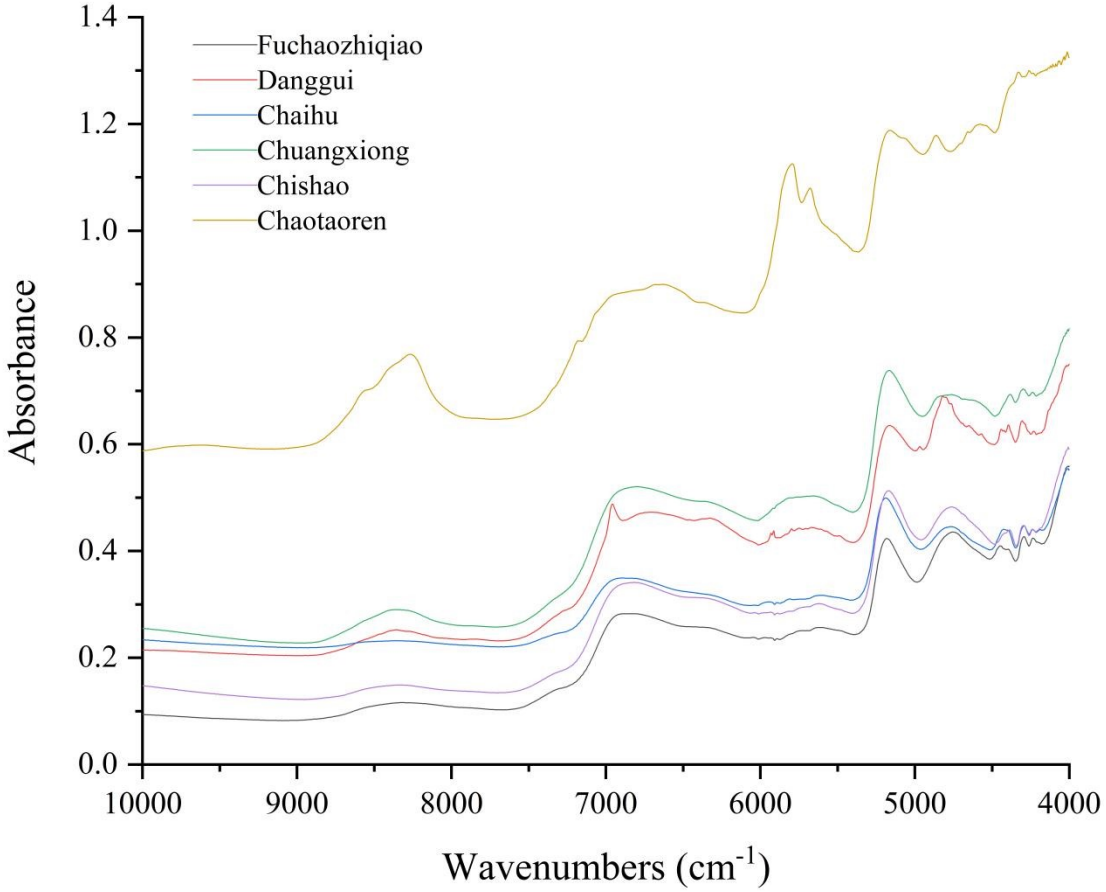
S1 Mahalanobis distance of 270 samples in invasive PLS model



S2 Mahalanobis distance of 360 samples in non-invasive PLS model



S3 Raw NIR spectra of six medicinal herbs in powder



S4 Performance parameters of gallic acid invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	7	0.4095	0.3159	0.8662	0.2748	0.8851
MSC	7	0.3587	0.2884	0.8919	0.2588	0.9038
SNV	8	0.3562	0.2812	0.8976	0.2516	0.9074
1^{st} derivative	5	0.3502	0.2672	0.9087	0.2363	0.9183
2^{nd} derivative	7	0.3187	0.1886	0.9568	0.2282	0.9195
Savitzky-Golay	7	0.4145	0.3109	0.8710	0.2745	0.8883
SNV+1^{st} derivative	4	0.3463	0.2833	0.8959	0.2759	0.8475
SNV+Savitzky-Golay	8	0.3575	0.2857	0.8939	0.2516	0.9072
1^{st} derivative+Savitzky-Golay	5	0.3541	0.2729	0.9043	0.2349	0.9164

S5 Performance parameters of amygdalin invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	8	4.4529	3.2564	0.8202	3.1193	0.8059
MSC	7	4.1020	3.2485	0.8213	2.6967	0.8661
SNV	7	4.0668	3.2561	0.8203	2.7732	0.8653
1^{st} derivative	5	3.9855	3.0156	0.8514	2.7352	0.8670
2^{nd} derivative	7	3.5962	2.4974	0.9040	2.6413	0.8670
Savitzky-Golay	8	4.5442	3.2691	0.8183	3.1296	0.8048
SNV+1^{st} derivative	5	3.9026	3.0872	0.8427	2.6890	0.8684
SNV+Savitzky-Golay	7	4.0721	3.2724	0.8180	2.7711	0.8653
1^{st} derivative+Savitzky-Golay	4	3.9765	3.2504	0.8211	2.9574	0.8325

S6 Performance parameters of albiflorin invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	5	6.6174	5.0136	0.6575	4.1410	0.7603
MSC	3	6.8103	5.1264	0.6210	4.2844	0.7374
SNV	11	6.1499	4.2865	0.7918	4.2003	0.8036
1^{st} derivative	4	6.2683	4.6795	0.7316	3.9472	0.7966
2^{nd} derivative	9	5.5294	3.0156	0.9103	5.0820	0.7420
Savitzky-Golay	5	6.6638	5.0142	0.6572	4.1542	0.7588
SNV+1^{st} derivative	3	6.6401	4.9723	0.6685	4.0511	0.7684
SNV+Savitzky-Golay	11	6.3352	4.3762	0.7796	4.1763	0.8042
1^{st} derivative+Savitzky-Golay	4	6.2594	4.7014	0.7276	3.9102	0.7972

S7 Performance parameters of paeoniflorin invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	8	5.9451	4.3502	0.7552	4.0255	0.7324
MSC	6	5.7351	4.4488	0.7381	4.4305	0.7028
SNV	6	5.8338	4.4613	0.7359	4.3712	0.7864
1^{st} derivative	8	5.9451	4.3503	0.7552	4.0257	0.7324
2^{nd} derivative	7	5.0939	3.4357	0.8669	3.8502	0.7688
Savitzky-Golay	8	5.8901	4.3862	0.7491	4.0033	0.7366
SNV+1^{st} derivative	9	5.5777	3.573	0.8538	4.4812	0.7105
SNV+Savitzky-Golay	8	5.9451	4.3501	0.7552	4.0257	0.7324
1^{st} derivative+Savitzky-Golay	4	5.7950	4.4990	0.7287	4.0321	0.7120

S8 Performance parameters of ferulic acid invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	8	0.2678	0.2119	0.9006	0.2168	0.9204
MSC	7	0.2662	0.2179	0.8937	0.2129	0.9146
SNV	10	0.2530	0.1743	0.9530	0.1729	0.9463
1^{st} derivative	5	0.2423	0.1945	0.9175	0.2059	0.9277
2^{nd} derivative	7	0.2192	0.1727	0.9362	0.1550	0.9556
Savitzky-Golay	7	0.2654	0.2156	0.8962	0.2167	0.9194
SNV+1^{st} derivative	9	0.2265	0.1469	0.9547	0.1497	0.9578
SNV+Savitzky-Golay	8	0.2649	0.2172	0.8945	0.2056	0.9182
1^{st} derivative+Savitzky-Golay	5	0.2439	0.1969	0.9152	0.2091	0.9260

S9 Performance parameters of naringin invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	5	14.7361	10.0632	0.4813	7.2193	0.5895
MSC	4	15.1622	10.0411	0.4669	7.1090	0.6125
SNV	4	15.0638	10.0504	0.4716	7.1571	0.6247
1^{st} derivative	5	13.9444	9.6623	0.6377	8.4810	0.6271
2^{nd} derivative	6	12.8589	8.3501	0.7779	10.5821	0.5383
Savitzky-Golay	5	14.7862	10.0532	0.4767	7.28622	0.5785
SNV+1^{st} derivative	10	13.7695	8.5963	0.7587	9.7602	0.5788
SNV+Savitzky-Golay	4	15.1200	7.2195	0.5859	10.0410	0.4666
1^{st} derivative+Savitzky-Golay	5	14.1082	9.7501	0.6621	8.5922	0.6181

S10 Performance parameters of neohesperidin invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	5	12.1778	9.0312	0.5945	7.6532	0.6313
MSC	4	12.7188	9.1282	0.5628	7.3611	0.6538
SNV	8	12.5295	8.7780	0.6471	7.6305	0.6826
1^{st} derivative	5	11.8489	8.4261	0.6999	7.1424	0.7099
2^{nd} derivative	6	10.6217	6.9775	0.8253	7.3634	0.7125
Savitzky-Golay	5	12.2697	9.0520	0.5890	7.8235	0.6170
SNV+1^{st} derivative	3	11.9589	8.9153	0.6228	7.0662	0.6984
SNV+Savitzky-Golay	7	12.5530	8.8835	0.6294	7.6160	0.6734
1^{st} derivative+Savitzky-Golay	4	11.9219	8.7701	0.6502	7.5784	0.6884

S11 Performance parameters of gallic acid non-invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	9	0.3293	0.2434	0.6996	0.2409	0.6222
MSC	11	0.3066	0.2278	0.7566	0.2424	0.6542
SNV	11	0.3098	0.2270	0.7590	0.2470	0.6514
1^{st} derivative	8	0.2759	0.1973	0.8343	0.2420	0.6432
2^{nd} derivative	6	0.2742	0.1750	0.8759	0.2589	0.6373
Savitzky-Golay	6	0.3371	0.2554	0.6369	0.2537	0.5628
SNV+1^{st} derivative	8	0.2854	0.2070	0.8129	0.2309	0.6814
SNV+Savitzky-Golay	11	0.3108	0.2314	0.7451	0.2461	0.6437
1^{st} derivative+Savitzky-Golay	8	0.2874	0.2087	0.8088	0.2417	0.6496

S12 Performance parameters of amygdalin non-invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	10	4.2871	3.3042	0.8152	3.5334	0.7582
MSC	6	4.4519	3.4514	0.7927	3.4142	0.7680
SNV	6	4.4598	3.4411	0.7944	3.5434	0.7728
1^{st} derivative	5	3.8951	3.1683	0.8339	3.6809	0.7707
2^{nd} derivative	6	3.7048	2.4427	0.9095	3.3903	0.7848
Savitzky-Golay	10	4.3417	3.3224	0.8127	3.4938	0.7623
SNV+1^{st} derivative	5	3.9246	3.2021	0.8294	3.4982	0.7908
SNV+Savitzky-Golay	9	4.3101	3.3602	0.8070	3.4431	0.7857
1^{st} derivative+Savitzky-Golay	5	3.9075	3.2063	0.8289	3.5542	0.7702

S13 Performance parameters of albiflorin non-invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	12	5.8491	4.2982	0.7708	4.2201	0.7716
MSC	4	6.1541	4.8071	0.6707	4.0331	0.7216
SNV	4	6.1369	4.8082	0.6706	4.2241	0.7230
1^{st} derivative	2	5.6815	4.6807	0.7016	4.3513	0.7123
2^{nd} derivative	5	5.5368	3.6444	0.8507	4.5332	0.7394
Savitzky-Golay	11	5.9695	4.4272	0.7503	4.1201	0.7681
SNV+1^{st} derivative	2	5.6496	4.6438	0.7095	4.2484	0.7393
SNV+Savitzky-Golay	4	6.1488	4.8140	0.6690	4.2385	0.7197
1^{st} derivative+Savitzky-Golay	2	5.7151	4.6974	0.6979	4.3382	0.7043

S14 Performance parameters of paeoniflorin non-invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	13	4.3472	3.3919	0.8346	3.4584	0.7873
MSC	8	4.5996	3.5386	0.8157	3.4523	0.7576
SNV	8	4.6206	3.5418	0.8154	3.6124	0.7597
1^{st} derivative	3	4.4429	3.7529	0.7847	3.4354	0.7350
2^{nd} derivative	7	4.1387	2.6204	0.9093	3.7251	0.7441
Savitzky-Golay	13	4.5841	3.4446	0.8280	3.4045	0.7896
SNV+1^{st} derivative	5	4.3395	3.4290	0.8299	3.3715	0.7919
SNV+Savitzky-Golay	8	4.5781	3.5541	0.8136	3.6115	0.7590
1^{st} derivative+Savitzky-Golay	4	4.4553	3.6667	0.7978	3.3335	0.7492

S15 Performance parameters of ferulic acid non-invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	9	0.5624	0.4467	0.8369	0.3338	0.8675
MSC	8	0.5933	0.4396	0.8432	0.3322	0.8603
SNV	8	0.5929	0.4409	0.8421	0.3448	0.8581
1^{st} derivative	4	0.5494	0.4345	0.8476	0.3368	0.8650
2^{nd} derivative	7	0.5195	0.3125	0.9280	0.3702	0.8528
Savitzky-Golay	9	0.5725	0.4507	0.8332	0.3340	0.8664
SNV+1^{st} derivative	8	0.5302	0.3680	0.8967	0.2866	0.9087
SNV+Savitzky-Golay	11	0.5757	0.4185	0.8608	0.3049	0.8960
1^{st} derivative+Savitzky-Golay	5	0.5451	0.4287	0.8525	0.3159	0.8845

S16 Performance parameters of naringin non-invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	5	12.9336	8.5091	0.4706	5.8231	0.6473
MSC	4	13.5241	8.4923	0.4558	5.9542	0.6645
SNV	4	13.5898	8.5094	0.4703	6.0762	0.6683
1^{st} derivative	3	12.7166	5.8203	0.4882	6.1345	0.6427
2^{nd} derivative	5	12.1468	8.1644	0.6413	6.3020	0.7028
Savitzky-Golay	5	12.9539	8.5032	0.4646	5.8422	0.6379
SNV+1^{st} derivative	4	13.2367	8.5130	0.5197	6.1023	0.7174
SNV+Savitzky-Golay	4	13.6618	8.5031	0.4646	6.1254	0.6631
1^{st} derivative+Savitzky-Golay	3	12.7459	8.5184	0.4847	6.1381	0.6391

S17 Performance parameters of neohesperidin non-invasive PLS correction models with different pretreatment methods

Preprocessing methods	PCs	RMSECV (%)	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Raw spectra	5	10.5811	8.0502	0.6117	8.7673	0.5424
MSC	4	11.3661	8.2151	0.5534	7.9135	0.5202
SNV	4	11.4170	8.2144	0.5539	8.1133	0.5214
1^{st} derivative	4	10.4441	7.7601	0.6708	8.6960	0.5825
2^{nd} derivative	6	9.76517	5.9853	0.8442	7.8041	0.6653
Savitzky-Golay	5	10.5942	8.0495	0.6120	8.7754	0.5416
SNV+1^{st} derivative	9	9.99413	6.7184	0.7904	8.6172	0.6123
SNV+Savitzky-Golay	4	11.4132	8.2138	0.5539	8.1277	0.5205
1^{st} derivative+Savitzky-Golay	4	10.4463	7.8233	0.6599	8.8575	0.5680

S18 Performance parameters of each component content invasive model of powder treated by different variable screening methods

Components	Methods	Numbers of variables	PCs	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Gallic acid	Full	1543	7	0.2389	0.9071	0.2282	0.9195
	CARS	18	3	0.2710	0.9083	0.2389	0.9071
	UVE	736	7	0.2027	0.9483	0.2243	0.9272
Amygdalin	Full	1543	7	2.4970	0.9040	2.6412	0.8671
	CARS	39	5	3.2315	0.8193	3.1092	0.8391
	UVE	622	7	2.5413	0.8996	2.7025	0.8591
Albiflorin	Full	1543	5	3.0156	0.9103	5.0822	0.7453
	CARS	20	5	4.3257	0.7692	3.9656	0.8413
	UVE	527	7	3.7884	0.8487	4.1481	0.7666
Paeoniflorin	Full	1543	8	3.4353	0.8669	3.8501	0.7717
	CARS	13	5	3.2221	0.8484	2.6923	0.9055
	UVE	548	5	3.4693	0.8089	2.6754	0.8711
Ferulic acid	Full	1543	8	0.1727	0.9362	0.1554	0.9556
	CARS	18	7	0.1523	0.9517	0.1054	0.9791
	UVE	729	6	0.1561	0.9483	0.1156	0.9733
Naringin	Full	1543	5	8.3514	0.7779	10.5806	0.5383
	CARS	20	2	10.0712	0.4557	7.4396	0.6379
	UVE	359	4	9.8494	0.5467	8.6933	0.6362
Neohesperidin	Full	1543	5	6.9774	0.8253	7.3633	0.7125
	CARS	12	3	8.2722	0.6884	6.0615	0.7618
	UVE	366	4	7.6891	0.7421	6.0594	0.8206

S19 Performance parameters of each component content non-invasive model of powder treated by different variable screening methods

Components	Methods	Numbers of variables	PCs	RMSEC (%)	Rc^2	RMSEP (%)	Rp^2
Gallic acid	Full	1543	5	0.1918	0.8476	0.2650	0.6607
	CARS	27	3	0.2173	0.7871	0.2518	0.6532
	UVE	483	5	0.1918	0.8455	0.2568	0.6462
Amygdalin	Full	1543	6	2.4427	0.9095	3.3903	0.7848
	CARS	36	5	3.1274	0.8392	3.4975	0.7742
	UVE	423	4	2.9540	0.8600	3.5293	0.7617
Albiflorin	Full	1543	5	3.6444	0.8507	4.5332	0.7394
	CARS	9	4	4.3921	0.7560	4.2532	0.7526
	UVE	345	2	4.4022	0.7543	4.1971	0.7634
Paeoniflorin	Full	1543	7	2.6204	0.9093	3.7251	0.7441
	CARS	21	3	3.5164	0.8187	3.5550	0.7460
	UVE	358	5	3.1285	0.8640	3.6232	0.7538
Ferulic acid	Full	1543	7	0.3125	0.9280	0.3702	0.8528
	CARS	66	2	0.4990	0.7823	0.4637	0.7553
	UVE	583	5	0.3728	0.8936	0.3330	0.8820
Naringin	Full	1543	5	8.1644	0.6413	6.3020	0.7028
	CARS	9	3	8.5132	0.5192	6.7605	0.6136
	UVE	298	5	8.1290	0.6480	6.9603	0.6282
Neohesperidin	Full	1543	6	5.9853	0.8442	7.8041	0.6653
	CARS	15	5	7.8262	0.6594	9.2044	0.4923
	UVE	313	4	7.1543	0.7494	7.6321	0.6727

S20 Comprehensive properties of 21 batches Xuefu Zhuyu Capsule raw powder samples

Powder properties	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
D_{10} (μm)	34.58	23.03	24.64	46.40	25.56	26.80	46.41	28.24	26.03	41.73	26.04	26.96
D_{50} (μm)	146.0	99.60	86.49	151.6	109.3	93.46	152.8	112.2	91.75	153.9	112.8	92.34
D_{60} (μm)	174.3	136.3	104.1	181.9	148.8	111.8	184.4	146.2	109.2	187.3	156.9	110.6
D_{90} (μm)	306.8	386.9	176.3	313.1	444.1	190.3	320.3	338.4	184.9	343.6	379.8	191.5
RA (°)	71.05	64.83	60.12	61.54	63.10	59.44	64.15	57.43	62.85	62.77	59.65	60.87
CA (°)	60.16	56.67	47.04	52.52	53.94	56.44	51.61	49.32	52.75	56.85	58.20	56.14
DA (°)	10.89	8.16	13.08	9.02	9.16	3.00	12.54	8.11	10.10	5.92	1.45	4.73
FPA (°)	54.88	52.10	61.02	56.00	53.58	59.41	53.31	58.05	53.24	56.68	57.55	58.62
AD (g/cm^3)	0.36	0.31	0.28	0.35	0.31	0.29	0.35	0.33	0.30	0.35	0.34	0.26
TD (g/cm^3)	0.55	0.58	0.55	0.57	0.60	0.56	0.57	0.57	0.56	0.56	0.59	0.55
Compressibility (%)	35	47	49	39	48	48	39	42	46	38	42	53
Dispersion (%)	2	2	1	3	2	3	1	1	2	2	1	2
Uniformity	5.04	5.92	4.22	3.92	5.82	4.17	3.97	5.18	4.2	4.49	6.03	4.1
FI	47**	43**	41.5**	48**	45**	44.5**	46**	46**	45.5**	49**	45**	44.5**
JFI	34*	34.25*	43*	40**	38.5*	29.75*	41**	39*	39*	32.25*	27*	29.75*

****good

***ordinary

**poor

*very poor

*continued table

Powder properties	P13	P14	P15	P16	P17	P18	P19	P20	P21
D_{10} (μm)	41.11	26.40	27.20	41.44	26.00	25.23	67.03	26.88	28.66
D_{50} (μm)	153.4	113.2	89.27	142.2	96.41	87.63	193.5	98.96	95.61
D_{60} (μm)	188.1	161.4	106.4	173.3	127.7	106.3	227.4	129.7	114.1
D_{90} (μm)	344.2	433.1	187.1	313.8	359.8	211.8	357.4	350.7	195.0
RA (°)	62.29	63.37	59.36	57.41	60.79	60.46	56.69	60.21	64.00
CA (°)	59.26	61.26	57.14	55.81	57.77	52.37	52.51	58.82	58.29
DA (°)	3.03	2.11	2.22	1.60	3.02	8.09	4.18	1.39	5.71
FPA (°)	54.03	57.4	56.77	55.7	58.18	55.99	52.34	55.61	55.94
AD (g/cm^3)	0.35	0.29	0.27	0.33	0.3	0.27	0.36	0.3	0.26
TD (g/cm^3)	0.59	0.58	0.55	0.57	0.59	0.56	0.57	0.56	0.52
Compressibility (%)	41	50	51	42	49	52	37	46	50
Dispersion (%)	3	2	-1	1	1	2	2	3	2
Uniformity	4.58	6.11	4.03	4.18	4.91	4.21	3.39	4.83	3.98
FI	47.5**	43**	44.5**	47.5**	44.5**	45.5**	51**	45.5**	46**
JFI	27*	26*	26.5*	32*	29.5*	37*	35*	27*	27*

****good

***ordinary

**poor

*very poor