

Article

Support Vector Regression Approach to Predict the Design Space for the Extraction Process of *Pueraria lobata*

Yaqi Wang^{1,2}, Yuanzhen Yang², Jiaojiao Jiao¹, Zhenfeng Wu² and Ming Yang^{1,2,*}

¹ College of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 610072, China; wangyaqi_3@163.com (Y.W.); jqiao6@163.com (J.J.)

² Key Laboratory of Modern Preparation of Traditional Chinese Medicine, Ministry of Education, Jiangxi University of Traditional Chinese Medicine, Nanchang 330004, China; yangyuanzhen666@163.com (Y.Y.); zfwu527@163.com (Z.W.)

* Correspondence: tcmmyang@163.com; Tel.: +86-791-8711-8658

Received: 2 August 2018; Accepted: 18 September 2018; Published: 20 September 2018



Abstract: A support vector regression (SVR) method was introduced to improve the robustness and predictability of the design space in the implementation of quality by design (QbD), taking the extraction process of *Pueraria lobata* as a case study. In this paper, extraction time, number of extraction cycles, and liquid–solid ratio were identified as critical process parameters (CPPs), and the yield of puerarin, total isoflavonoids, and extracta sicca were the critical quality attributes (CQAs). Models between CQAs and CPPs were constructed using both a conventional quadratic polynomial model (QPM) and the SVR algorithm. The results of the two models indicated that the SVR model had better performance, with a higher R^2 and lower root-mean-square error (RMSE) and mean absolute deviation (MAD) than those of the QPM. Furthermore, the design space was predicted using a grid search technique. The operational range was extraction time, 24–51 min; number of extraction cycles, 3; and liquid–solid ratio, 14–18 mL/g. This study is the first reported work optimizing the design space of the extraction process of *P. lobata* based on an SVR model. SVR modeling, with its better prediction accuracy and generalization ability, could be a reliable tool for predicting the design space and shows great potential for the quality control of QbD.

Keywords: *Pueraria lobata*; SVR; QPM; extraction process; QbD; design space

1. Introduction

Radix puerariae (RP) is the root of *Pueraria lobata*, known as ‘Gegen’ in Chinese. RP was one of the earliest herbal resources used for food and medicine: it was firstly documented in “Shen Nong’s Herbal Classic” and classified as middle grade for the prevention and treatment of fever, diabetes, diarrhea, and cardiovascular and cerebrovascular diseases. Furthermore, modern studies have shown that RP extract exhibits potential bioactivity for the treatment of several immune disorders, such as atopic dermatitis [1], osteoporosis [2], and Alzheimer’s disease [3]. Isoflavonoids are believed to be the major active components in RP. Puerarin is the major and most important component in RP with extensive pharmacological activities such as hepatoprotection [4], anti-atherogenic effects [5], and anti-cancer effects [6]. The content of puerarin ($\geq 2.4\%$) is regarded as the quality indicator of RP according to the Pharmacopoeia of the People’s Republic of China.

Extraction is a key operation process in the manufacturing of health food, dietary supplements, and medicine, especially for traditional Chinese medicine (TCM). The quality by design (QbD) concept has been applied in the pharmaceutical field as a well-established tool for both formulation and manufacturing process development. According to ICH guideline Q8, QbD is implemented with

several steps, including risk analysis, diagnosis of potential critical quality attributes (CQAs) and process parameters (CPPs), construction of mathematical models, optimization of the design space, selection of a control strategy, and continual improvement in the product lifecycle [7,8]. Among these steps, design space, as a reliable operation range, has gained increasing attention in the food and drug industries. According to the definition of design space, the quality is guaranteed at any combination of independent variables (process parameters) within the space. Consequently, the design space is regarded as a zone of robustness, as no significant fluctuations should be observed within the space.

Design space is established on the basis of a sound understanding of the effect of the interaction of CQAs and CPPs on the quality of the product. Therefore, statistical and multivariate analysis models are essential in the implementation of QbD. The predictive ability of the mathematical model shows the robustness and predictability of the design space. Furthermore, a well-fitted model helps us not only to gain a clear understanding of the connection and the intrinsic regular pattern between CPPs and CQAs, but also to gain regulatory flexibility. Thus, constructing a reliable model is first and foremost. The robustness of the model directly influences the batch-to-batch consistency of the quality of products.

Response surface methodology (RSM), due to its data visualization and handling ability, has become the most widely used method to express multidimensional relationships between CQAs and CPPs. The quadratic polynomial model (QPM) is the most common algorithm for response surface optimization. Design of experiment (DoE) analysis techniques such as analysis of variance and fitted regression models are used frequently. A multivariate knowledge space may be delineated to find regions of risk or optimal performance, which are often graphically illustrated by figures and known as response surfaces. The literature has shown that the QPM has good performance when used for relatively simple and linear cases [9,10]. The QPM also has some well-known limitations that may affect the prediction accuracy [11]. The multidimensional relationships observed in the pharmaceutical area are often complex and nonlinear. The predictions of models based on the linear regression algorithm exhibit poor estimation [12]. Therefore, the QPM may not be the most applicable algorithm for accurate prediction of the design space of QbD.

Support vector regression (SVR) is a promising kernel-based machine learning algorithm developed by Vapnik and Cortes [13]. The SVR approach can optimize complex nonlinear problems by using an exclusive objective function that minimizes the structural risk of the model. The introduction of the kernel function allows nonlinear problems to be linearly solved in a higher dimension compared with its original dimensional feature space. Thus, SVR has a global optimum and exhibits excellent prediction accuracy. Considering its remarkable generalization performance, SVR has attracted particular attention and been extensively used in applications including atmospheric science prediction [14], drug design [15], credit rating analysis [16], protein structure and function prediction [17], and metabolomics [18]. In most of these cases, the performance of the SVR model is better than that of traditional machine learning approaches. To our knowledge, few studies of the design space model have been developed based on the SVR model.

The aim of this present study was to explore the practicability of using an SVR model for predicting a design space. For the RP extraction process, there are several analytical methods based on the QPM for response surface optimization [19]. A single algorithm may not be credible enough for model development. To the best of our knowledge, there is no study in the literature comparing the QPM and the SVR algorithm for the extraction process of RP. Thus, in this paper, the RP extraction process was optimized as a case study. The extraction time, number of extraction cycles, and liquid–solid ratio were identified as CPPs, and the yield of puerarin, total isoflavonoids, and extracta sicca were the CQAs. Models between CQAs and CPPs were constructed using both the QPM and the SVR algorithm based on the Box–Behnken design. The performance of the two models was analyzed and compared. Then, the design space was calculated and optimized using a grid search technique. This is the first study on optimization of the design space of the extraction process of RP using the SVR algorithm.

2. Results

2.1. Box–Behnken Design

Box–Behnken design is one of the most commonly used DoEs for RSM. Three influential factors (independent variables) for the extraction process of RP were investigated, including extraction time (X_1 , min), extraction cycles (X_2 , cycles), and liquid–solid ratio (X_3 , mL/g). Seventeen experimental runs were arranged using Box–Behnken design. Their experimental results were considered as dependent variables. In this work, Y_1 , Y_2 , and Y_3 represented the yields of puerarin (%), extracta sicca (%), and total isoflavonoids (%), respectively. The entire dataset obtained using Box–Behnken design was considered as the training set and adopted to establish the QPM and SVR fitted models. To further evaluate the performance of the two fitted models, cross-validation and external validation approaches were adopted. Four sets of external validation values—random combinations of independent variables along with experimental responses—were used as the test set to evaluate the quality of the fitted models. The experimental datasets are shown in Table 1.

Table 1. Results of training and test set.

No.	Factors			Response Variables		
	X_1 (min)	X_2 (cycles)	X_3 (mL/g)	Y_1 (%)	Y_2 (%)	Y_3 (%)
Training set						
1	35	3	5	3.93	28.64	16.00
2	35	3	15	5.22	34.11	43.86
3	10	3	10	3.63	27.53	28.95
4	35	2	10	4.03	28.06	16.40
5	60	3	10	4.74	34.44	26.82
6	35	2	10	3.90	27.97	20.60
7	35	1	5	1.22	13.52	39.20
8	10	2	15	3.24	25.18	27.57
9	60	2	5	3.24	26.67	13.37
10	10	1	10	1.76	14.82	7.73
11	35	2	10	3.93	28.25	15.41
12	60	1	10	2.40	23.47	12.66
13	35	2	10	4.03	30.40	23.03
14	60	2	15	4.28	33.46	36.94
15	35	2	10	3.77	28.90	21.69
16	10	2	5	2.32	19.67	8.46
17	35	1	15	2.56	20.74	20.70
Test set						
1	25	2	10	3.50	25.92	22.70
2	30	2	8	3.45	25.37	17.62
3	15	2	15	3.46	26.35	34.41
4	20	2	15	3.45	27.78	35.40

2.2. QPM Analysis

A five-fold cross-validation method was used for training and evaluation of both the QPM and the SVR model. The regression coefficient and parameters of QPM and SVR are listed in Table 2. All the fitted R^2 (training set) values of QPM were more than 0.96, which shows that more than 96% of the variation could be explained by QPM. For the yields of puerarin (Y_1) and extracta sicca (Y_2), the predicted R^2 (test set) values were higher than 0.90, which shows that 90% of the variation could be predicted by QPM. The cross-validation R^2 values were above 0.80. The closer the cross-validation R^2 is to 1, the better the generalization capability of the statistical model. The regression coefficients and p -values of the constructed QPM equation are presented in Table 3. For the yields of puerarin (Y_1) and extracta sicca (Y_2), the effects of extraction time (X_1), extraction cycles (X_2), and liquid–solid ratio (X_3) were all significant model terms. For the yield of total isoflavonoids (Y_3), there was no significant

model term, whereas the fitted R^2 (training set) of the QPM was more than 0.96. This phenomenon may be caused by the combined effect of each model term playing a greater role than any single model term.

Table 2. Statistical parameters of the quadratic polynomial model (QPM) and the support vector regression (SVR) model.

		QPM			SVR		
		R^2	RMSE	MAD	R^2	RMSE	MAD
Y_1	Training set	0.985	0.127	0.111	0.983	0.132	0.077
	Test set	0.903	0.191	0.164	0.918	0.175	0.133
	Cross-validation	0.802	0.457	0.366	0.846	0.403	0.329
Y_2	Training set	0.988	0.641	0.514	0.982	0.789	0.596
	Test set	0.944	0.946	0.797	0.975	0.636	0.559
	Cross-validation	0.908	1.795	1.429	0.954	1.272	1.031
Y_3	Training set	0.964	1.906	1.56	0.961	2.005	1.646
	Test set	0.706	4.12	4.02	0.765	3.683	3.606
	Cross-validation	0.724	5.311	4.567	0.821	4.281	3.834

Table 3. Coefficients of the constructed QPM equation.

		QPM									
		C_0	X_1^2	X_2^2	X_3^2	$X_1 X_2$	$X_1 X_3$	$X_2 X_3$	X_1	X_2	X_3
Y_1	Regression coefficient	0.751	-1.524	-1.674	-1.124	2.157	3.859	2.237	0.470	0.120	-0.050
	<i>p</i> -value	0.014 *	0.006 *	0.003 *	0.023 *	0.275	0.771	0.903	0.003 *	0.001 *	0.003 *
Y_2	Regression coefficient	9.371	-3.317	-11.287	-6.567	11.257	26.075	13.050	-1.740	1.280	-1.750
	<i>p</i> -value	0.001 *	0.132	0.001 *	0.012 *	0.413	0.542	0.41	0.003 *	0.001 *	0.001 *
Y_3	Regression coefficient	1.402	0.158	-1.702	8.478	5.412	17.347	5.582	-7.060	4.460	11.080
	<i>p</i> -value	0.698	0.979	0.777	0.186	0.273	0.477	0.104	0.491	0.053	0.478

* Significant at the 0.05 level.

2.3. SVR Analysis

2.3.1. Parameter Optimization for SVR

Due to the good general performance and the small number of parameters to be adjusted, the Gaussian radial basis function (RBF) was employed as the kernel function in this study. Thus, the SVR performance depends on the combination of three parameters: the capacity parameter C , the kernel function parameter ε , and the kernel parameter σ .

In the SVR formulation, the capacity parameter C represents the tradeoff between the margin maximization and the training error minimization. If the C value is too high or too low, the algorithm may over- or under-fit the training set. The kernel function parameter ε signifies the width of the ε -insensitive zone used to fit the training data. Moreover, the corresponding kernel parameter σ strongly influences the number of support vectors. Thus, the combination of the three parameters controls the accuracy and generalization performance of the regression estimate.

To choose the parameters of the model, this paper adopted the method of cross-validation based on a grid search avoiding blindness and randomness. The key is to find out which combination of the three parameters gives the highest prediction accuracy. The procedure for optimizing the parameters of SVR is as follows (schematic shown in Figure 1):

Step 1: Start SVR using an RBF kernel and find the R^2 , $RMSE$, and MAD for the run with the test set.

Step 2: Repeat the procedure by varying ε from 2^{-15} to 2^{-1} (typically 15 values) and find the R^2 , $RMSE$, and MAD for each run (15 runs).

Step 3: Repeat the procedure by varying C (capacity control) from 2^6 to 2 (typical six values: 2^6 , 2^5 , 2^4 , 2^3 , 2^2 , 2) and find the R^2 , $RMSE$, and MAD for each run (15×6 runs).

Step 4: Repeat the procedure by varying σ from 2^{-15} to 2^{-1} (typically 15 values) and find the R^2 , $RMSE$, and MAD for each run ($15 \times 6 \times 15$ runs).

Step 5: Steps 1–4 are repeated for the development of each model of Y_1 , Y_2 , and Y_3 .

The optimized combination of the three parameters for the SVR model is shown in Table 4.

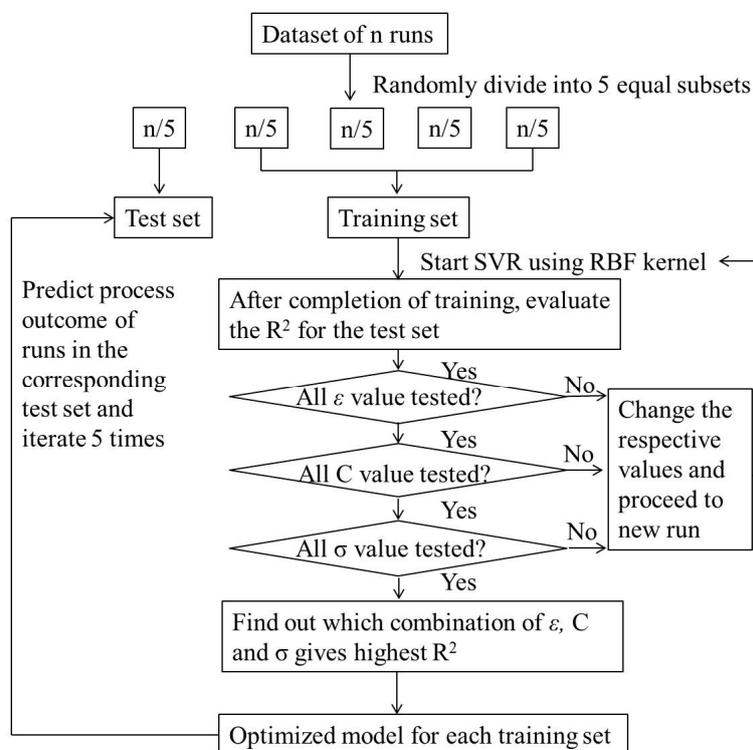


Figure 1. Schematic for SVR algorithm implementation.

Table 4. Optimized parameters for the SVR model.

	SVR		
	σ	C	ϵ
Y_1	2^{-4}	2^2	2^{-6}
Y_2	2^{-5}	2^4	2^{-3}
Y_3	2^{-4}	2^5	2^{-3}

2.3.2. Evaluation of Models

The performance of the two models was mainly evaluated using two validation techniques: cross-validation and test set. The parameters of the QPM and the SVR model are listed in Table 2. Although the R^2 values of the training set were similar between QPM and SVR, for the test set and cross-validation group, the R^2 values of the SVR model were higher than those of the QPM. Furthermore, the $RMSE$ and MAD values of SVR for the test set and cross-validation were lower than those of QPM. The high R^2 and low $RMSE$ and MAD indicate the good prediction and generalization performance of the SVR model.

Figure 2 depicts a comparison between predicted and experimental values for the training and test sets with QPM and SVR. Obviously, the values predicted by SVR are more closely matched with the experimental runs than are those by QPM. Hence, SVR is superior to QPM for predicting the yield value of the extraction process of RP.

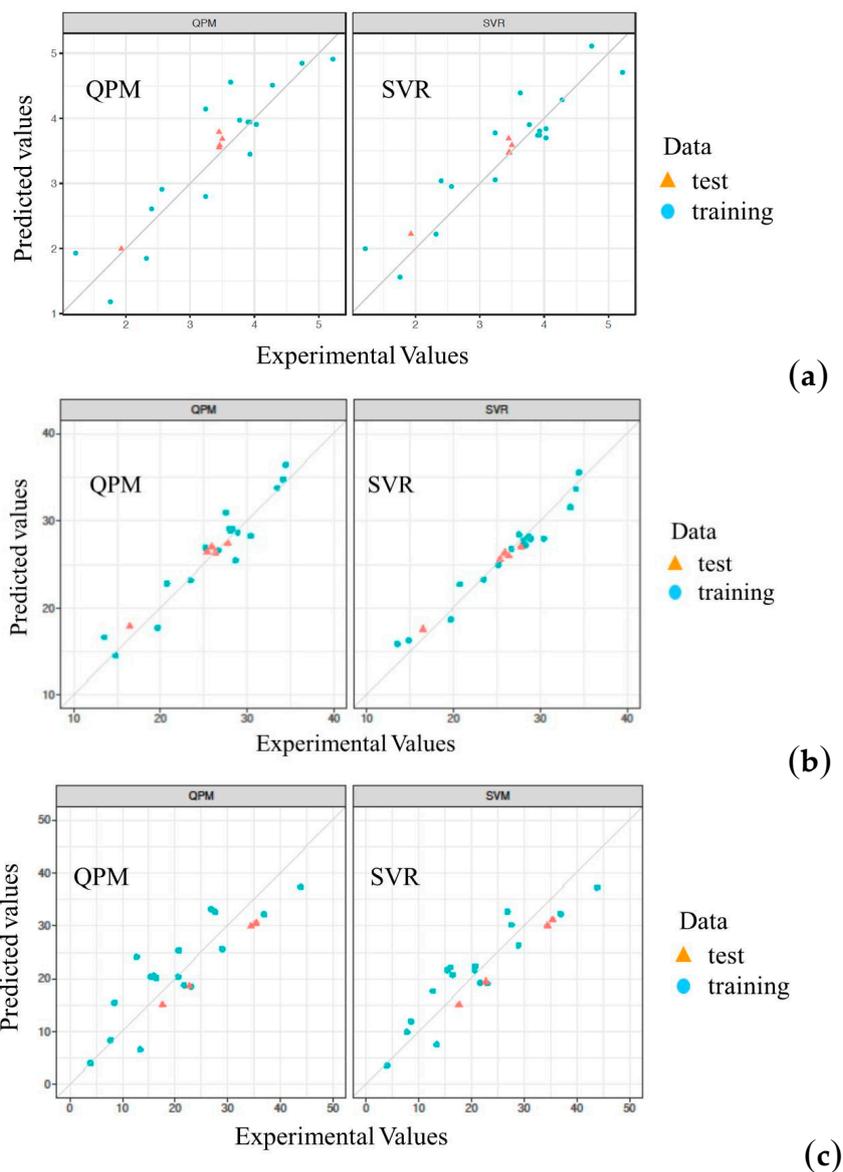


Figure 2. Comparison of predicted and experimental values for QPM and SVR. (a) Y_1 ; (b) Y_2 ; (c) Y_3 .

2.4. Design Space

In this study, the aim of the optimization was to obtain the design space (operating range) which could provide the maximum yield value of active components and minimum of extracta sicca. Figure 3 shows the design space of the optimized yield value (D) based on the SVR model. It was observed that the yield value of D increased with the extraction time (X_1) up to 20 min, and then declined for X_1 greater than 55 min. This phenomenon may be caused by the increased extraction time accelerating the mass transfer. When the extraction time continuously increases, the extracta sicca grows faster than the active components' dissolution rate. When the extraction time reaches 55 min, the D value begins to decline. However, the number of extraction cycles (X_2) and liquid–solid ratio (X_3) both had a positive effect on the extraction efficiency. Increased X_2 and X_3 enhanced the extraction efficiency greatly, which shows that the concentration gradient of the solvent plays a critical role in the RP extraction process. However, a continuously increasing number of extraction cycles and liquid–solid ratio may result in a heavy load for the subsequent concentration and drying process.

For a maximum yield value of active components and minimum extracta sicca, a D value of 0.9 was set as the lower limit. The design space was calculated using a grid search method.

The four-dimensional design space is shown in Figure 3d. The design space is an irregular polygon, and the operational range for the extraction process is X_1 , 24–51 min; X_2 , 3 cycles; and X_3 , 14–18 mL/g.

The predicted operating range was validated using an external validation approach which consisted of a combination of CPPs that never occurred in the calibration set. Verification results are shown in Table 5. All the optimized D values were higher than 0.9, in perfect agreement with the predicted values. The good correlation between these results indicates that the design space was reliable in predicting the operating range.

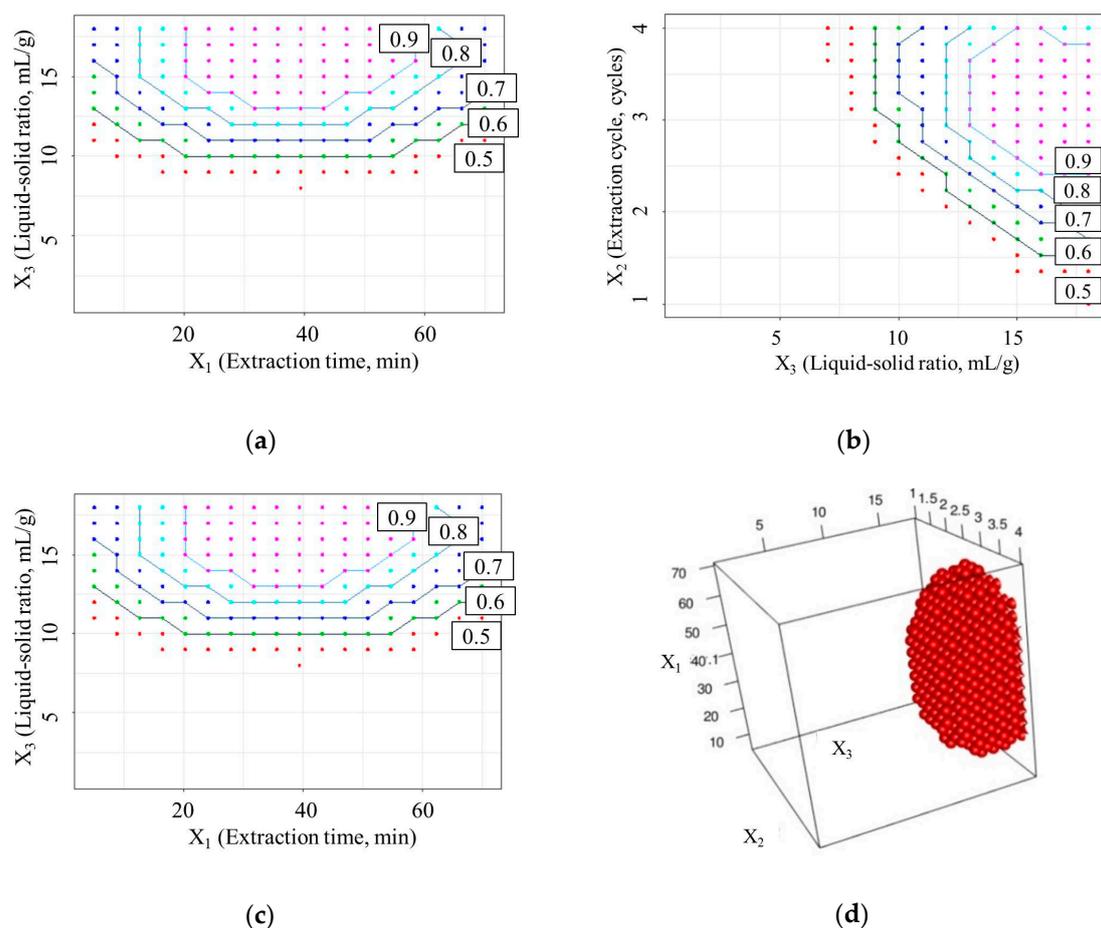


Figure 3. Design space and contour plots: (a) extraction cycle (X_2) fixed at 3 cycles; (b) extraction time (X_1) fixed at 35 min; (c) liquid-to-solid ratio (X_3) fixed at 9 mL/g; (d) four-dimensional operation space ($D > 0.9$).

Table 5. Responses for verification experiments.

No.	Factors			Predicted D Value	Experimental D Value
	X_1 (min)	X_2 (cycles)	X_3 (mL/g)		
1	35	3	14	0.98	0.99
2	40	3	15	1.03	1.01

3. Materials and Methods

3.1. Materials

RP was collected from Pingwu, Sichuan Province (China). Reference samples of puerarin and rutin were purchased from the National Institute for the Control of Pharmaceuticals and Biological Products (Beijing, China). Acetonitrile (chromatographic grade) and methanol (chromatographic

grade) were obtained from Merck (Darmstadt, Germany). Deionized water was produced using a Milli-Q academic water purification system (Millipore, Bedford, MA, USA). Sodium nitrite (analytical grade), aluminum nitrate (analytical grade), and sodium hydroxide (analytical grade) were purchased from Sigma (Saint Louis, MO, USA).

3.2. Apparatus

An Agilent 1200 HPLC equipped with a variable-wavelength ultraviolet detector (Agilent Technologies, Santa Clara, CA, USA) was used for HPLC analysis. A Phenomenex reversed-phase Gemini C₁₈ column (250 × 4.6 mm, 5 μm) and a Phenomenex C₁₈ guard column (Phenomenex, Torrance, CA, USA) were used for chromatographic analysis. A Shimadzu UV-2550 UV-vis Spectrophotometer (Shimadzu, Kyoto, Japan) was used for UV analysis.

3.3. Procedures

According to the Chinese pharmacopoeia, heat reflux extraction was used to extract active components from RP using water as the solvent. Before extraction, RP samples were oven-dried at 60 °C for 12 h. Each sample was cut into half-inch cubes. A 100 g sample was placed in a round-bottom flask and soaked with 5–15 mL/g (500–1500 mL) of water for 30 min. Then, heat reflux extraction was performed for 1–3 cycles of 10–60 min each according to the experimental design. After extraction, the mixture was filtered with gauze. The supernatant was subjected to HPLC, UV analysis, and weight measurement of extracta sicca. Dried extracta sicca were obtained by firstly evaporating in a 60 °C water bath and then drying in a 105 °C oven for 6 h. The yields of puerarin (Y_1), extracta sicca (Y_2), and total isoflavonoids (Y_3) were calculated using the equation $Y = (\text{mass of analyte}/\text{mass of plant material}) \times 100\%$. Each experiment was conducted in triplicate and the average yield value was used for statistical analysis.

3.4. HPLC Analysis

The yield of puerarin (Y_1) extracted from RP was analyzed using an HPLC approach. The detection wavelength was 254 nm. The gradient elution of the mobile phase contained (A) acetonitrile and (B) water with 0.1% formic acid. Gradient procedures were as follows: 0–25 min, 11% A; 25–30 min, 11–25% A; 30–40 min, 25–40% A. The flow rate was 1.0 mL/min and the injection volume was 10 μL. Column temperature was maintained at 30 °C. The linear range of puerarin was 10–1000 μg/mL, $Y = 28.981X + 112.29$ ($R^2 = 0.9997$). The limits of detection (LOD) and quantification (LOQ) were 0.6 and 2 μg/mL, respectively.

3.5. UV Analysis

The total quantity of isoflavonoids (Y_3) extracted from RP was measured using an aluminum nitrate colorimetric method described by Saeed [20]. Briefly, in a 10 mL test tube, 1 mL of the extract, 0.4 mL 5% sodium nitrite, and 0.4 mL 10% aluminum nitrate were mixed. After 6 min, 4 mL 4% sodium hydroxide and 4.2 mL 75% ethanol were added, mixed well, and left to stand for 15 min. Then, the absorbance of the solution was measured against a prepared reagent blank at 510 nm using a UV-vis spectrophotometer. The standard curve for total isoflavonoids was found using a rutin standard solution (0–80 mg/L) under the same procedure described above. The content of total isoflavonoids was expressed as milligrams of rutin equivalent per gram of dried sample. All the samples were analyzed in duplicate.

3.6. Establishment of Models

3.6.1. QPM

A second-order response function was applied to establish a mathematical model that relates the response measured to the independent variables

$$Y = b_0 + \sum_{i=1} b_i X_i + \sum_{i=1} b_{ii} X_i^2 + \sum_{i,j=1} b_{ij} X_{ij} \quad (1)$$

where Y is the response, b_0 is a constant, b_i is the linear coefficient, b_{ii} is the quadratic coefficient, b_{ij} is the two-factor interaction coefficient, and X_i and X_j are the process parameters.

The quality of the fitted models was expressed using the coefficient of determination (R^2), $RMSE$, and MAD .

3.6.2. SVR

By introducing the kernel function, the original input was mapped into the feature space. The ultimate mathematical form of the kernel-based SVR is shown in Equation (2)

$$f(x) = \sum_{i=1}^N (\alpha_i^* - \alpha_i) K(x_i, x) + b \quad (2)$$

$$b = \begin{cases} y_i - f(x_i)_{b=0} - \varepsilon & \text{For } \alpha_i \in (0, C) \\ y_i - f(x_i)_{b=0} + \varepsilon & \text{For } \alpha_i^* \in (0, C) \end{cases} \quad (3)$$

where α_i and α_i^* are the optimized Lagrange multipliers, $K(x_i, x)$ denotes the kernel function describing the dot product in the feature space, b is the bias parameter, and x_i and y_i denote the i th support vector and the corresponding target output, respectively. The coefficients α and α^* have an intuitive interpretation as forces pushing and pulling the regression estimate $f(x)$ towards the measurements y .

The kernel function is defined in terms of the dot product of the mapping function as given by

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (4)$$

where $\Phi(x)$ is the high-dimensional feature space being nonlinearly mapped to from the input space x . There exist several choices for the kernel function K , including linear, polynomial, splines, and radial basis functions. With respect to the support vector regression, the function which is broadly employed is the Gaussian RBF.

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (5)$$

3.7. Optimization

A function was used to optimize the three CQAs simultaneously: $D = Y_1 + Y_3 - Y_2$. D is the optimized yield value of Y_1 , Y_2 , and Y_3 . All the CQA data were normalized using the equation $Y' = (Y - Y_{\min}) / (Y_{\max} - Y_{\min})$. In our study, the goal of optimization was to maximize yield value of active components and minimize the extracta sicca. Extraction time X_1 (0–70 min), number of extraction cycles X_2 (1–4 cycles), and liquid–solid ratio X_3 (0–18 mL/g) were chosen as the recognition space. A grid search method was used to seek the design space.

The software RStudio (v1.1.456, RStudio, Boston, MA, USA) and R v3.5 were used to construct the QPM and SVR model, and to perform the grid search method.

4. Conclusions

In this study, the SVR algorithm was used to improve model performance for finding the design space of the RP extraction process. The results indicate that the SVR model performs better, with higher R^2 and lower $RMSE$ and MAD than those of QPM. Furthermore, verification results were in perfect agreement with the predicted values, which indicated that the SVR model was reliable in predicting the operating range for the extraction process of RP. This study is the first reported work optimizing the design space of the RP extraction process based on an SVR model. SVR modeling has better prediction

accuracy and generalization ability, and shows great potential for the quality control of QbD as a reliable tool for predicting the design space.

Author Contributions: Y.W. and M.Y. conceived and designed the experiments; J.J. and Y.Y. performed the experiments; Z.W. gave some advice for improving the paper; Y.W. analyzed the data and wrote the paper.

Funding: This research was funded by the National Key Research and Development Program of China (No. 2017YFC1702904) and National Standardization Program of Traditional Chinese Medicine (No. ZYBZH-C-SC-54).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, J.H.; Jeon, Y.D.; Lee, Y.M.; Kim, D.K. The suppressive effect of puerarin on atopic dermatitis-like skin lesions through regulation of inflammatory mediators in vitro and in vivo. *Biochem. Biophys. Res. Commun.* **2018**, *498*, 707–714. [CrossRef] [PubMed]
2. Tanaka, T.; Tang, H.J.; Yu, F.N.; Michihara, S.; Uzawa, Y.; Zaima, N.; Moriyama, T.; Kawamura, Y. Kudzu (*Pueraria lobata*) vine ethanol extracts improve ovariectomy-induced bone loss in female mice. *J. Agric. Food Chem.* **2011**, *59*, 13230–13237. [CrossRef] [PubMed]
3. Koirala, P.; Seong, S.H.; Jung, H.A.; Choi, J.S. Comparative molecular docking studies of lupeol and lupenone isolated from *Pueraria lobata* that inhibits BACE1: Probable remedies for Alzheimer's disease. *Asian Pac. J. Trop. Med.* **2017**, *10*, 1117–1122. [CrossRef] [PubMed]
4. Li, R.; Xu, L.Y.; Liang, T.; Li, Y.W.; Zhang, S.J.; Duan, X.Q. Puerarin mediates hepatoprotection against CCl₄-induced hepatic fibrosis rats through attenuation of inflammation response and amelioration of metabolic function. *Food Chem. Toxicol.* **2013**, *52*, 69–75. [CrossRef] [PubMed]
5. Cheung, D.W.S.; Koon, C.M.; Ng, C.F.; Leung, P.C.; Fung, K.P.; Poon, S.K.S.; Lau, C.B.S. The roots of *Salvia miltiorrhiza* (Danshen) and *Pueraria lobata* (Gegen) inhibit atherogenic events: A study of the combination effects of the 2-herb formula. *J. Ethnopharmacol.* **2012**, *143*, 859–866. [CrossRef] [PubMed]
6. Hu, Y.F.; Li, X.; Lin, L.; Liang, S.T.; Yan, J. Puerarin inhibits non-small cell lung cancer cell growth via the induction of apoptosis. *Oncol. Rep.* **2018**, *39*, 1731–1738. [CrossRef] [PubMed]
7. U.S. Department of Health and Human Services Food and Drug Administration. Guidance for Industry PAT-A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. Available online: <https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm070305.pdf> (accessed on 1 September 2004).
8. ICH. ICH Harmonised Tripartite Guideline: Pharmaceutical Development Q8 (R2). Available online: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf (accessed on 1 August 2009).
9. Zhai, C.H.; Xuan, J.B.; Fan, H.L.; Zhao, T.F.; Jiang, J.L. The application of SVR model in the improvement of QbD: A case study of the extraction of podophyllotoxin. *Drug Dev. Ind. Pharm.* **2018**, *44*, 1506–1511. [CrossRef] [PubMed]
10. Wang, Y.Q.; Wu, Z.F.; Ke, G.; Yang, M. An effective vacuum assisted extraction method for the optimization of labdane diterpenoids from *Andrographis paniculata* by response surface methodology. *Molecules* **2015**, *20*, 430–445. [CrossRef] [PubMed]
11. Liu, R.; Li, X.; Zhang, W.; Zhou, H.H. Comparison of nine statistical model based warfarin pharmacogenetic dosing algorithms using the racially diverse international warfarin pharmacogenetic consortium cohort database. *PLoS ONE* **2015**, *10*, e0135784. [CrossRef] [PubMed]
12. Norioka, T.; Hayashi, Y.; Onuki, Y.; Andou, H.; Tsunashima, D.; Yamashita, K.; Takayama, K. A novel approach to establishing the design space for the oral formulation manufacturing process. *Chem. Pharm. Bull.* **2013**, *61*, 39–49. [CrossRef] [PubMed]
13. Liang, Y.Z.; Xu, Q.S.; Li, H.D.; Cao, D.S. *Support Vector Machines and Their Application in Chemistry and Biotechnology*; CRC Press: New York, NY, USA, 2011; pp. 15–16.
14. Jang, H.S.; Bae, K.Y.; Park, H.S.; Sung, D.K. Solar power prediction based on satellite images and support vector machine. *IEEE Trans. Sustain. Energy* **2016**, *7*, 1255–1263. [CrossRef]
15. Yosipof, A.; Guedes, R.C.; Garcia-Sosa, A.T. Data mining and machine learning models for predicting drug likeness and their disease or organ category. *Front. Chem.* **2018**, *6*, e162. [CrossRef] [PubMed]

16. Zhong, H.M.; Miao, C.Y.; Shen, Z.Q.; Feng, Y.H. Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings. *Neurocomputing* **2014**, *128*, 285–295. [[CrossRef](#)]
17. Cai, C.Z.; Xiao, H.G.; Yuan, Q.F.; Liu, X.H.; Wen, Y.F. Function prediction for DNA-/RNA-binding proteins, GPCRs, and drug ADME-associated proteins by SVM. *Protein Pept. Lett.* **2008**, *15*, 463–468. [[CrossRef](#)] [[PubMed](#)]
18. Eliasson, M.; Rannar, S.; Trygg, J. From data processing to multivariate validation-essential steps in extracting interpretable information from metabolomics data. *Curr. Pharm. Biotechnol.* **2011**, *12*, 996–1004. [[CrossRef](#)] [[PubMed](#)]
19. Wong, K.H.; Li, G.Q.; Li, K.M.; Razmovski-Naumovski, V.; Chan, K. Optimisation of Pueraria isoflavonoids by response surface methodology using ultrasonic-assisted extraction. *Food Chem.* **2017**, *231*, 231–237. [[CrossRef](#)] [[PubMed](#)]
20. Saeed, N.; Khan, M.R.; Shabbir, M. Antioxidant activity, total phenolic and total flavonoid contents of whole plant extracts *Torilis leptophylla* L. *BMC Complement. Altern. Med.* **2012**, *12*, 221. [[CrossRef](#)] [[PubMed](#)]

Sample Availability: RP samples and the reference compounds are available from the authors.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).