

水文模型参数敏感性快速定量评估的 RSMSobol方法

孔凡哲¹, 宋晓猛¹, 占车生², 叶爱中³

(1. 中国矿业大学 资源与地球科学学院, 江苏 徐州 221008;

2. 中国科学院地理科学与资源研究所 陆地水循环及地表过程重点实验室, 北京 100101;

3. 北京师范大学全球变化与地球系统科学研究院, 北京 100875)

摘要:水文模型参数敏感性分析是模型不确定性量化研究的重要环节,其可以有效识别关键参数,减少模型率定的不确定性,提高模型优化效率。然而如何快速有效地定量评估参数敏感性已成为当前大尺度分布式水文模型优化的瓶颈。针对传统的全局定量敏感性分析方法在多参数复杂水文模型的不足,本文采用基于统计学习理论的支持向量机(SVM)建立非参数响应曲面(称为代理模型),再结合基于方差的Sobol方法,建立了基于响应曲面方法的Sobol定量全局敏感性分析方法(RSMSobol方法),实现复杂模型系统参数敏感性的快速定量评估。本文选用淮河流域的日尺度分布式时变增益水文模型进行实例研究,采用水量平衡系数(WB), Nash-Sutcliffe效率系数(NS)和相关系数(RC)三个目标函数综合评价模拟效果。研究结果显示RSMSobol方法在实现定量全局敏感性分析的同时降低了模型运行时耗,提高了模型评估效率,且与传统定量方法Sobol方法具有同样的评估效果。该方法的有效应用为大型复杂水文动力模拟系统的参数定量敏感性评价提供了参考,为模型参数进一步优化提供了可靠依据。

关键词:代理模型;响应曲面方法;敏感性分析;支持向量机;淮河流域

水文模型作为研究流域水文过程及其演化规律的重要工具^[1],采用数学公式和物理方程对高度复杂的水文过程进行概念化和抽象化。因此水文过程模拟必然存在极大不确定性^[2],其来源主要有4个方面^[3-4]:模型输入数据、用于率定的模型输出数据、模型参数以及模型结构。为了得到可靠的模拟结果,参数识别、模型率定和不确定性量化成为模型构建的重要环节^[5]。在给定数据的条件下,参数识别与率定的关键就是要在模型复杂性与预测不确定性之间找到较好的平衡^[6]。对于一个参数众多的水文模型,率定过程则变得非常复杂,然而对模型结果起决定作用的仅仅是少数的重要参数。为此开展水文模型参数敏感性分析也成为模型率定优化的前提和必要条件,其可以有效评价模型参数对模型输出的贡献率和影响程度^[7-8]。

敏感性分析主要为了实现以下几个内容:1)确定哪些参数是重要参数,对模型输出贡献较大;2)确定不同参数组合对模型模拟效果的影响,以验证模型参数之间的相互作用;3)确定不敏感参数,减少参数率定过程中的相关计算量及其不确定性。通常敏感性分析包括局部敏感性分析和全局敏感性分析,局部敏感性分析只检验单个因子对模型的影响程度,由于

收稿日期: 2011-05-02; 修订日期: 2011-06-08

基金项目: 国家水体污染控制与治理重大专项(2009ZX07210-006); 国家重点基础研究发展计划(973计划)项目(2010CB428403); 中国矿业大学基本科研业务费大学生创新项目 [Foundation: National Grand Science and Technology Special Project of Water Pollution Control and Improvement, No.2009ZX07210-006; National Key Basic Research Program of China (973 Program), No.2010CB428403; China University of Mining and Technology Student Innovation Project of Fundamental Research Funds for Central Universities]

作者简介: 孔凡哲(1964-),男,江苏徐州人,博士,教授,主要从事流域水文模拟与流域产汇流过程研究。
E-mail: kongfz3@126.com

通讯作者: 占车生(1975-),男,湖北黄冈人,博士,主要从事遥感水文学和大尺度水文模型模拟研究。
E-mail: zhancs2006@gmail.com

其简单快捷,可操作性强,也被广泛应用,但是由于未考虑因子之间相互影响等因素,致使结果有一定的局限性;而全局敏感性分析同时考虑多个因子对模型输出的影响,分析各因子之间的相互作用对模型输出的影响,为全面认识各因子的敏感性程度提供参考,比较适合多参数的水文模型研究。目前,许多全局敏感性分析方法被广泛应用于水文模型参数的敏感性评价中^[9],如多元回归法, LH-OAT方法^[10], Morris-OAT方法^[5]和傅里叶幅度灵敏度检验法(Fourier Amplitude Sensitivity Test, FAST)^[11]为主的定性方法和以基于方差分解理论的Sobol方法^[7, 12]和扩展傅里叶幅度灵敏度检验法(Extend FAST)^[9, 13]为主的定量分析方法。然而对于这些作为“黑箱子”模型的传统分析方法通过直接计算参数的方差计算参数的敏感度,并不能充分利用样本的所有信息。尽管这些定量方法是可靠的和稳健的,但是由于其需要大量的计算,数以千次的模型评价都导致其应用的困难^[14],特别是对于多参数的复杂水文模型。随着计算机模拟仿真技术的不断提高与发展,代理模型技术被逐渐应用于复杂模型的不确定性分析^[15-16]、敏感性分析^[14-15, 17-18]以及参数优化^[16, 19]等,使快速实现模型不确定性量化成为可能。当前常用的代理模型技术包括响应曲面方法(Response Surface Methodology, RSM)^[17]和状态相关参数法(State dependent parameter, SDP)^[14],其中响应曲面方法是基于统计理论的数学模型,如多项式回归模型、人工神经网络^[17]和支持向量机^[17]等。虽说上述方法已经被成功应用于诸多工程设计问题中,如Sathyanarayanamurthy和Chinnam结合三种代理模型(Kriging模型、径向基神经网络模型和支持向量机模型)和基于方差的定量敏感性分析方法(Sobol方法和FAST方法)分析复杂的工程模型设计的敏感性问题^[18], Ratto等^[14]提出的SDP方法和Sobol方法结合也取得了较好的应用效果。但上述方法目前还很少应用于水文模型中。为此,本文尝试利用代理模型技术评价分布式水文模型的参数敏感性,快速实现多参数复杂模型的定量全局敏感性分析。利用基于统计学习理论的支持向量机(support vector machine, SVM)构建非参数响应曲面作为分布式水文模型的代理模型,结合Sobol方法和SVM的响应曲面法形成一个新的全局敏感性定量评估方法—RSMSobol方法,并将其方法应用于淮河蚌埠闸以上流域的分布式时变增益水文模型中,结果Sobol方法的结果进行比较,从而可以验证方法的有效性。

1 研究方法

1.1 Sobol方法

Sobol方法最早是由Sobol于1993年提出的^[20],该方法的核心将模型分解为单个参数及参数之间相互组合的函数^[7]。假设模型 $Y = F(X) = F(x_1, x_2, \dots, x_k)$, x_i 服从 $[0, 1]$ 均匀分布,则可分解为:

$$F(x_1, x_2, \dots, x_k) = \sum_{i=1}^k F_i(x_i) + \sum_{i=1}^k \sum_{j>i}^k F_{ij}(x_i, x_j) + \dots + F_{1, 2, \dots, k}(x_1, x_2, \dots, x_k) \quad (1)$$

式中: $F_i(x_i)$ 表示包含 x_i 的函数项, $F_{ij}(x_i, x_j)$ 表示同时包含 x_i 和 x_j 的函数项,依此类推,则总的方差可以分解成:

$$V = \sum_{i=1}^k V_i + \sum_{i=1}^k \sum_{j>i}^k V_{ij} + \dots + V_{1, 2, \dots, k} \quad (2)$$

式中: V 为总方差, V_i 为 x_i 的单因子方差, V_{ij} 为 x_i 和 x_j 的双因子方差, $V_{1, 2, \dots, k}$ 为所有因子作用的方差,对上式进行归一化,设:

$$S_{i_1, \dots, i_n} = \frac{V_{i_1, \dots, i_n}}{V} \quad (3)$$

则第 i 个参数的一阶、二阶和总敏感度 S_i , S_{ij} 和 S_{Ti} 分别为:

$$S_i = \frac{V_i}{V} \quad (4)$$

$$S_{ij} = \frac{V_{ij}}{V} \quad (5)$$

$$S_{T_i} = \sum S_{(i)} \quad (6)$$

式中, $S_{(i)}$ 为所有包含第 i 个参数的敏感度。

1.2 响应曲面方法

本研究结合美国劳伦斯·利弗莫尔国家实验室开发的 PSUADE (Problem Solving environmental for Uncertainty Analysis and Design Exploration), 利用其提供的多种响应曲面方法, 将一个大型复杂动力系统模型的复杂输入输出关系概化成一个代理模型, 即用此近似表达这种高度复杂高维非线性关系, 从而提高模型的计算速度和降低模型的计算量和复杂度, 同时在一定程度上保证模型计算精度^[3]。运用响应曲面模型的关键就是合理的样本数据 (对应的输入输出数据和空间填充数据) 和合理的拟合方法, 对于拟合方法经过发展已经形成了包括参数型 (如线性回归, Legendre 多项式和非线性回归函数等) 和非参数型 (如多元自适应回归样条方法, 人工神经网络, 高斯过程, 支持向量机等) 两大类。本文采用基于统计学习理论的支持向量机进行响应曲面构建, 其方法在工业参数优化中得到部分应用, 并取得较合理的效果^[21-22]。

支持向量机 (Support Vector Machines, SVM) 是一种基于结构风险最小化原则的函数拟合方法, 对于回归估计, 函数表达式可表示为:

$$Y = F(X) = a^T X + b \quad (7)$$

式中, a 是权重向量, b 是转移向量。

定义以下优化问题:

$$\min \frac{1}{2} |a|^2 \quad (8)$$

$$\text{s. t. } \begin{cases} Y_i - a^T X - b \leq \varepsilon \\ a^T X + b - Y_i \leq \varepsilon \end{cases}$$

当约束调价不满足时, 则引入松弛变量 Ξ_i , Ξ_i^* 和惩罚参数 C , 对超出 ε -不敏感区的样本数据进行惩罚, 此时最优化问题变为:

$$\min \frac{1}{2} |a|^2 + C \sum_{i=1}^l (\Xi_i + \Xi_i^*) \quad (9)$$

$$\text{s. t. } \begin{cases} Y_i - a^T X - b \leq \varepsilon + \Xi_i \\ a^T X + b - Y_i \leq \varepsilon + \Xi_i^* \\ \Xi_i, \Xi_i^* \leq 0 \end{cases}$$

基于 Lagrangian 函数条件, 线性关系可以转为如下优化问题:

$$\max - \frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) X_i^T X_j - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l Y_i (\alpha_i + \alpha_i^*) \quad (10)$$

$$\text{s. t. } \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i - \alpha_i^* \in [0, C] \end{cases}$$

式中 α_i^* , α_i 是 Lagrange 因子, 非线性关系在给定的核函数 $K(X_i, X_j)$ 变为:

$$\max - \frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(X_i, X_j) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l Y_i (\alpha_i + \alpha_i^*) \quad (11)$$

对于非线性回归, 函数表达式为: $Y = F(X) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(X_i, X_j) + b \quad (12)$

生成 SVM 响应曲面的步骤为: 1) 选择核函数 $K(X, X')$; 2) 生成训练样本, 模拟; 3) 用训练样本计算核矩阵; 4) 解决以上的优化问题得到 α_i, α_i^* ; 5) 计算权重向量 $a = \sum_{i=1}^l (\alpha_i^* - \alpha_i) X_i$;

6) 利用Karush-Kuhn-Tucker (KKT) 条件计算偏置向量 b ; 7) 最终得到函数表达式 (12)。

1.3 RSMSobol方法

针对模型计算的复杂性,本文采取结合响应曲面的Sobol方法,完成主效应(一阶敏感度)、交互作用(二阶敏感度)和总敏感度计算,具体方法为:

(1) 主效应分析 主效应分析是基于McKay提出的方差分析方法^[23-24],其本质是条件期望方差的统计分析。根据数理统计方法有方差可分解为:

$$V(Y) = V(E(Y|X_k)) + E(V(Y|X_k)) \quad (13)$$

式中: $V(Y)$ 和 $E(Y)$ 为输出变量 Y 的方差和均值, X_k 是第 k 个输入,右边第一项 $V(E(Y|X_k))$ 是输出变量条件期望的方差(条件是 X_k),可以记作 $VCE(X_k)$,表示 X_k 影响的情况下 Y 的变化,第二项 $E(V(Y|X_k))$ 是误差项或剩余项,表示不考虑 X_k 影响的 Y 的变异程度。

本研究中采用重复式的拉丁超立方设计样本进行计算。在样本中,当 X_k 取 x_{kj} ($j = 1, 2, \dots, s$,其中 s 是水平数)时,条件期望为:

$$\bar{Y}_j = \frac{1}{r} \sum_{i=1}^r Y_{ij} \quad (14)$$

$$\text{方差为:} \quad V(Y) = \frac{1}{sr} \sum_{j=1}^s \sum_{i=1}^r (Y_{ij} - \bar{Y}_j)^2 \quad (15)$$

式中: r 为重复式抽样次数, Y_{ij} 表示第 i 次抽样,第 j 个水平下的输出变量值, $\bar{Y}$ 为所有样本输出的均值, $\bar{Y}_j$ 为第 j 个水平条件下 r 次重复式抽样输出变量的均值。

则在 $X_k = x_{kj}$ 条件下的条件方差 $V(Y|X_k)$ 及其期望 $E(V(Y|X_k))$ 分别为:

$$V(Y|X_k = x_{kj}) = \frac{1}{r} \sum_{i=1}^r (Y_{ij} - \bar{Y}_j)^2 \quad (16)$$

$$E[V(Y|X_k = x_{kj})] = \frac{1}{sr} \sum_{j=1}^s \sum_{i=1}^r (Y_{ij} - \bar{Y}_j)^2 \quad (17)$$

另根据文献[23]中:

$$\begin{aligned} E\left(\sum_{j=1}^s (\bar{Y}_j - \bar{Y})^2\right) &\approx sV(\bar{Y}_j) \\ &= s\{V[E(\bar{Y}_j|X_k = x_{kj})] + E[V(\bar{Y}_j|X_k = x_{kj})]\} \\ &= s\left\{V[E(Y|X_k = x_{kj})] + \frac{1}{r}E[V(Y|X_k = x_{kj})]\right\} \end{aligned} \quad (18)$$

$$\text{可推导出:} \quad VCE(X_k) = \frac{1}{s} \sum_{j=1}^s (\bar{Y}_j - \bar{Y})^2 - \frac{1}{sr^2} \sum_{j=1}^s \sum_{i=1}^r (Y_{ij} - \bar{Y}_j)^2 \quad (19)$$

$$\text{对于分析结果的相关比}\eta^2\text{可以定义为:} \quad \eta^2 = VCE(X_k)/V(Y) \quad (20)$$

(2) 交互作用分析 针对不相关输入因素的双因素交互作用分析是在主效应分析的基础上扩展开来^[20],假定满足如下公式:

$$V(Y) = V(E(Y|X_i, X_k)) + E(V(Y|X_i, X_k)) \quad (21)$$

式中: X_i 和 X_k 分别是两个输入因素,右边第一项 $V(E(Y|X_i, X_k))$ 是双因素相互作用的期望方差,也可以记作 $VCE(X_i, X_k)$,第二项 $E(V(Y|X_i, X_k))$ 为残差项。

$$\text{则双因素相关比}\eta^2(X_i, X_k)\text{可由公式计算:} \quad \eta^2(X_i, X_k) = VCE(X_i, X_k)/V(Y) \quad (22)$$

高的相关比表示 X_i 和 X_k 两个输入因素对输出变异的共同作用很明显,贡献很显著。那么基于交互项的方差 $V(X_i, X_k)$ 可定义为:

$$V(X_i, X_k) = V(E(Y|X_i, X_k)) - V(E(Y|X_i)) - V(E(Y|X_k)) \quad (23)$$

本研究利用重复的正交设计样本计算 $V(X_i, X_k)$,其中 X_i 和 X_k 分别取值为 X_{ij} 和 X_{ik} ($j = 1,$

2, ..., S, l = 1, 2, ..., S, S 为水平数)。基于此样本的均值和方差为:

$$\bar{Y} = \frac{1}{S^2 R} \sum_{j=1}^S \sum_{l=1}^S \sum_{r=1}^R Y^{(r)}(X_i = X_{ij}, X_k = X_{kl}) \quad (24)$$

$$V(Y) = \frac{1}{SR} \sum_{j=1}^S \sum_{l=1}^S \sum_{r=1}^R [Y^{(r)}(X_i = X_{ij}, X_k = X_{kl}) - \bar{Y}]^2 \quad (25)$$

式中: $Y^{(r)}$ 表示第 r 组重复样本, 则条件期望为:

$$\bar{Y}(X_i = X_{ij}, X_k = X_{kl}) = \frac{1}{R} \sum_{r=1}^R Y^{(r)}(X_i = X_{ij}, X_k = X_{kl}) \quad (26)$$

和条件期望方差为:

$$VCE(X_i, X_k) = \frac{1}{S^2} \sum_{j=1}^S \sum_{l=1}^S (\bar{Y}_{lj} - \bar{Y})^2 - \frac{1}{S^2 R^2} \sum_{j=1}^S \sum_{l=1}^S \sum_{r=1}^R (Y_{(lj, r)} - \bar{Y}_{lj})^2 \quad (27)$$

最后可得到 $V(X_i, X_k)$ 的计算公式为:

$$V(X_i, X_k) = VCE(X_i, X_k) - VCE(X_i) - VCE(X_k) \quad (28)$$

式中: $VCE(X_i)$ 和 $VCE(X_k)$ 分别表示 X_i 和 X_k 的条件期望方差, 可通过主效应分析计算得到。

(3) 总敏感度分析 如果输入是相关的, 则以上所说的方差分解方法就不在有效, 本研究采用另外一种方法, 总敏感度算法, 定义为:

$$S_{Ti} = V_i + \sum_{j \neq i} V_{ij} + \sum_{i \neq j, i \neq k} V_{ijk} + \dots + V_{12\dots M} \quad (29)$$

式中: M 是输入因素个数, V_{ij} 是输入因素 i 和 j 同时作用条件下的方差, 依此类推。

则总敏感度可定义为: $S_{Ti} = E(V(Y|X_{\sim i}))/V(Y)$ (30)

式中: 下标 $\sim i$ 表示除了第 i 个因素的其他所有因素。

2 实例研究

本文将该方法应用到淮河流域的分布式时变增益水文模型中, 重点讨论模型参数的敏感性。分布式时变增益水文模型 DTVGM 是建立在 GIS/DEM 基础上, 通过 GIS/RS 提取陆地表面单元坡度、流向、水流路径、河流网络、流域边界和土地覆被等信息, 将单元 TVGM 扩展到由 DEM 划分的流域单元网格上进行非线性产流计算, 并利用 DEM 提取出的汇流网络进行网格汇流演算, 从而得到流域出口断面的流量过程。模型基本原理可参见文献[25-28]。

淮河流域地处长江流域和黄河流域之间, 位于东经 112°~121°、北纬 31°~36°, 流域面积 27 万 km²。淮河流域由淮河和沂沭泗两大水系组成, 废黄河以南为淮河水系, 以北为沂沭泗水系。淮河水系集水面积 19 万 km² (图 1), 约占流域总面积的 70%。流域年平均降雨量约 900 mm, 其中 70%~80% 的降水集中在夏季, 因而径流分布也极不均匀。淮河的年平均流量为 853 m³/s, 汛期洪峰流量可达 11000 m³/s 以上, 而枯季流量几乎为零。本文选择淮河蚌埠闸以上流域 1965-2008 年的日降雨量、蒸发量、水文站流量等资料构建淮河流域日尺度分布式时变增益模型, 其中分为两个不同时期 (1965-2000 和 2001-2008) 进行模型的适应性以及 RSMSobol 方法的稳健性验证。土地利用数据采用 1995-1996 年的全国土地利用数据库中的淮河流域土地利用数据 (1:100 万), 数据来源为中国科学院资源环境科学数据中心。DEM 数据来自美国地质调查局 (USGS) 的最新 SRTM DEM 数据, 分辨率为 3'×3'。基于淮河流域实测的二级河网对 DEM 进行了校正, 采用校正后的 DEM 提取流域信息 (坡度、流向、河网、子流域等)。采用 100 km 的阈值, 将蚌埠闸以上淮河流域划分为 441 个子流域, 水文气象数据通过地学统计插值模型拓展到 441 个子流域上。

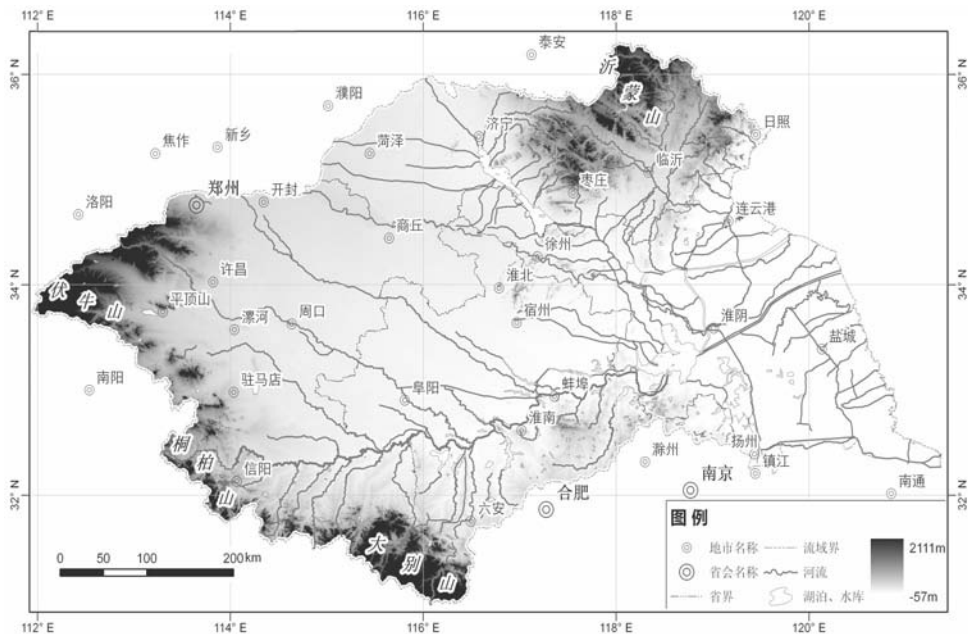


图1 淮河流域位置图

Fig. 1 Location of the Huaihe River Basin in China

由于模型参数较多,实现全部参数的定量化敏感性分析较难,加之有些参数本身对模型影响较小,因此为了有效降低模型参数的不确定性和计算时耗,在前期对模型研究的基础上,本文选择8个较重要的参数进行全局敏感性分析。由于研究区各子流域本身空间变异性差异不大,可对不同的子流域部分模型参数采用相同的取值,构建成文件数据库供模型调用,特别对于本文重点讨论的模型参数,为了在计算过程中减少不必要的损耗和模型运行时间以及连续高效地运行模型,本研究中采用统一值。对于参数的取值范围,在综合考虑模型运行过程中的实际影响和流域具体情况外,基于最不利原则采用相对较大的参数范围,为此本文确定的8个参数的取值范围见表1,且假定所有参数均服从均匀分布。

本文采用正交拉丁超立方 OALH 采样方法^[29-30],样本个数为1024组,并采用3个不同目标函数值,即水量平衡系数 WB, Nash-Sutcliffe 效率系数 NS 和相关系数 RC,具体目标函数为:

$$WB = \sum_{i=1}^n Q_{s,i} / \sum_{i=1}^n Q_{o,i} \tag{31}$$

$$NS = 1 - \frac{\sum_{i=1}^n (Q_{o,i} - Q_{s,i})^2}{\sum_{i=1}^n (Q_{o,i} - \bar{Q}_o)^2} \tag{32}$$

$$RC = \frac{\sum_{i=1}^n (Q_{o,i} - \bar{Q}_o) (Q_{s,i} - \bar{Q}_s)}{\sqrt{\sum_{i=1}^n (Q_{o,i} - \bar{Q}_o)^2 \sum_{i=1}^n (Q_{s,i} - \bar{Q}_s)^2}} \tag{33}$$

式中: Q_s 和 Q_o 分别是模拟流量和观测流量, $\bar{Q}$ 是相应流量的均值, n 为时间序列长度。

本文选择 Gauss 径向基函数 $K(X_i, X_j) = \exp(-\gamma \|X_i - X_j\|^2)$ 为 SVM 的核函数,其中 PSUADE

表1 分布式时变增益水文模型参数取值范围

Tab. 1 Parameters and ranges of distributed time variant gain model

参数	物理意义	参数范围
g1	时变增益因子	[0.01, 1.0]
g2	时变增益因子	[0.01, 5.0]
Kr	壤中流出流系数	[0.01, 1.0]
Kaw	蒸发权重系数	[0.01, 1.0]
RoughRss	曼宁公式糙率系数	[0.001, 0.1]
Wmi	最小土壤湿度	[0.01, 0.40]
WM	上层饱和土壤湿度	[0.40, 1.0]
ThickU	上层土壤厚度	[400, 800]

的 SVM 工具默认的初始条件设置为 $\gamma = 1$ (取值范围 10^{-6} - 10^6), $e = 10^{-6}$ (取值范围 10^{-6} -1)。经多次调试比较,调整上述 2 个参数值,发现当 $\gamma = 0.1, e = 10^{-5}$ 时响应输出目标值范围达到最大(表 2)。根据

表 2 SVM 模型参数调试结果比较

Tab. 2 Results of parameter calibration in the SVM

方案	ϵ	γ	WB	NS	RC
初始条件	10^6	1	[0.54478, 0.66809]	[0.25697, 0.52652]	[0.72351, 0.91236]
方案1	10^5	0.1	[0.37985, 0.78177]	[0.21774, 0.72460]	[0.70515, 0.99207]
方案2	10^4	0.1	[0.38011, 0.78152]	[0.21784, 0.72438]	[0.70514, 0.99198]
方案3	10^5	0.01	[0.37820, 0.72937]	[0.26195, 0.38837]	[0.64938, 0.90225]
方案4	10^5	10	[0.66334, 0.66914]	[0.33245, 0.37240]	[0.77142, 0.79817]
方案5	10^5	1	[0.54479, 0.66809]	[0.25697, 0.52651]	[0.72351, 0.91235]

1024 组样本及其目标函数值,构建基于 SVM 的非参数响应曲面,并进行响应曲面交互误差验证。为了能够直观反映响应曲面关系,本文选择双参数 (g_1 和 g_2) 来构建不同目标函数 (WB, NS 和 RC) 的三维关系 (图 2)。从交互验证的误差分布来看,插值误差符合响应曲面构建基本要求^[29-30]。由此可以得出基于 SVM 的响应曲面作为原始模型的代理模型可以很好的反映样本数据的特征信息,较好的实现模型仿真。

代理模型验证结果合理后,则通过 SVM 响应曲面生成 10 万组样本对模型参数进行全局敏感性评价 (表 3, 表 4 和表 5)。从表 3 的主效应和表 4 的总敏感度分析结果可以看出模型参数时变增益因子 g_1 和 g_2 是重要的参数,另外,所有目标函数中各参数敏感度之和均小于 1,说明参数之间存在一定的相互作用。从表 5 中可以看出 g_1 和 g_2 与其他参数的二阶敏感度也较大。且对于不同目标函数,其参数敏感性存在一定差异,产流参数 g_1 和 g_2 对水量平衡系数 WB 表现较为重要,而汇流参数 RoughRss 对 Nash-Sutcliffe 效率系数 NS 和相关系数 RC 的响应影响明显。同时对于不同时期的数据样本通过 RSMSobol 方法进行对比试验,同样得出相似的结果 (表 3, 表 4),说明方法具有很好的稳健性。

为了验证结果的可靠性,本文采用 Sobol 方法对参数敏感性进行定量分析比较 (表 3 和

表 3 不同目标函数条件的一阶敏感度比较

Tab. 3 First order sensitivity indices for different objective functions

参数	WB			NS			RC		
	RSMSobol ^a	RSMSobol ^b	Sobol	RSMSobol ^a	RSMSobol ^b	Sobol	RSMSobol ^a	RSMSobol ^b	Sobol
g_1	0.179	0.230	0.178	0.023	0.032	0.014	0.158	0.177	0.177
g_2	0.482	0.309	0.361	0.246	0.186	0.182	0.048	0.017	0.052
Kr	0.007	0.002	0.005	0 ^c	0 ^c	0 ^c	0.003	0.004	0.001
Wni	0.013	0.039	0.014	0.002	0.001	0.001	0.014	0.005	0.023
WM	0.038	0.081	0.048	0.005	0.002	0.002	0.028	0.051	0.042
Kav	0.063	0.030	0.039	0.007	0.002	0.001	0.021	0.029	0.04
RoughRss	0.005	0.007	0.003	0.105	0.172	0.062	0.399	0.384	0.404
ThickU	0.013	0.018	0.014	0.001	0.001	0 ^d	0.009	0.012	0.011

注: a-数据序列为 1965-2000 年, b-数据序列为 2001-2008 年, c-数值数量级为 10^{-5} , d-数值数量级为 10^{-4} , 近似取值为 0

表 4 不同目标函数条件的总敏感度比较

Tab. 4 Total sensitivity indices for different objective functions

参数	WB			NS			RC		
	RSMSobol ^a	RSMSobol ^b	Sobol	RSMSobol ^a	RSMSobol ^b	Sobol	RSMSobol ^a	RSMSobol ^b	Sobol
g_1	0.368	0.423	0.373	0.543	0.544	0.513	0.330	0.305	0.305
g_2	0.698	0.578	0.687	0.839	0.717	0.818	0.155	0.077	0.149
Kr	0.008	0.016	0.010	0.002	0.013	0 ^c	0.020	0.031	0.012
Wni	0.043	0.066	0.041	0.022	0.026	0.023	0.049	0.014	0.043
WM	0.044	0.097	0.059	0.040	0.068	0.054	0.105	0.128	0.121
Kav	0.088	0.053	0.084	0.037	0.015	0.091	0.054	0.059	0.049
RoughRss	0.005	0.012	0.004	0.412	0.571	0.497	0.590	0.627	0.591
ThickU	0.013	0.022	0.019	0.023	0.017	0.009	0.041	0.072	0.05

注: a-数据序列为 1965-2000 年, b-数据序列为 2001-2008 年, c-数值数量级为 10^{-4} , 近似取值为 0。

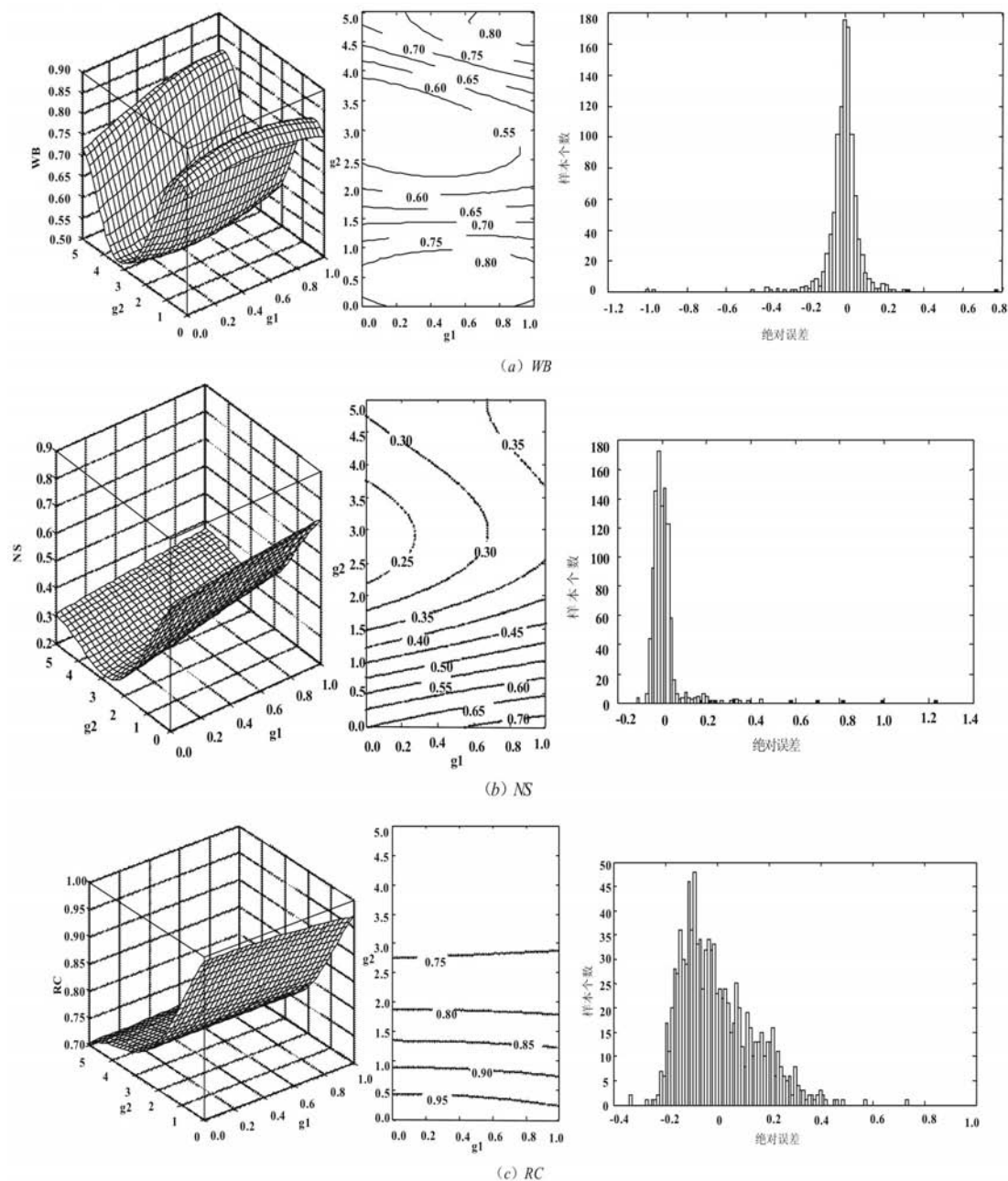


图2 基于不同目标函数的SVM响应曲面及其交互验证误差分布

Fig. 2 SVM-based response surface and its cross validation for different objective functions

表4),同样得出以上结果。从模型物理意义上说,模型参数 g_1 和 g_2 作为产流因子,对模型水量平衡WB影响较为显著,而对于日尺度模型,汇流参数中的糙率系数RoughRss对模型汇流过程及模拟流量过程影响显著,从而影响Nash-Sutcliffe效率系数NS和实测与模拟流量的相关系数RC,对于土壤饱和含水量WM和蒸散发折算系数Kaw而言,其对模型输出的影响也挺重要,属于相对敏感参数。

从结果上来看,两种分析方法的结果保持高度的一致性,但是对于Sobol方法进行计算过程中,需要大量的样本数据和时耗,如在Windows XP系统,Intel Core 2 T5800 CPU,主频

2.00 GHz, 内存 2.00 GB 条件下, 运行模型 10000 次, 耗时约 2300 分钟, Sobol 方法虽能够给出定量化的结果, 但是评价过程中巨大的计算量对于复杂模型应用起来较为困难, 且计算过程中将所有样本作为独立个体看待, 而无法充分考虑样本特征及其输入输出的关系^[14]。

本文所用的 RSMSobol 方法在给出定量关系的同时大大降低计算时耗, 其仅需在 Windows XP 系统下运行初始模型 1024 次耗时约 210 分钟, 在 1024 次的基础上构建响应曲面从而提供 10 万组样本, 在 Centos 5.5 系统, Intel Pentium IV CPU, 主频 1.80 GHz, 内存 256 MB 条件下, 通过 PSUADE 平台进行一阶敏感度、二阶敏感度和总敏感度分别耗时约 5 分钟、40 分钟和 1 分钟, 可见其大大提高了模型计算效率。同时构建的响应曲面模型可较好地反映输入参数与模型评价函数之间的响应关系, 提供连续平滑的样本数据, 为敏感性分析提供足够的样本信息。

3 结论

本文对当前多参数复杂水文模型的敏感性问题进行讨论, 针对传统经典分析方法 (如 GLUE, Sobol 和 Extend FAST 等) 在复杂模型参数敏感性分析的应用不足, 为了实现复杂模型的快速定量敏感性评估, 引入一种代理模型方法, 通过支持向量机构建非参数响应曲面仿真复杂模型系统, 结合基于方差的 Sobol 定量评价方法, 建立了新的定量全局敏感性分析方法--RSMSobol 方法。在响应曲面代理模型的基础上深入探讨了淮河流域日尺度分布式时变增益水文模型参数的敏感性。通过比较 Sobol 方法和 RSMSobol 方法, 发现采用代理模型技术对复杂模型参数敏感性进行快速定量评估具有较好的可行性和实用性, 为进一步的模型率定与参数优化提供了坚实基础。同时, RSMSobol 方法也为其他大型复杂水动力学模拟系统提供了新的不确定性量化分析思路, 实现快速有效地定量评估复杂模型系统参数敏感性, 缓解目前大型复杂模型系统的不确定性量化低效高时耗问题。但本文的研究仍有一些不足, 对于响应曲面模型的优化, 即如何实现构建最优的响应曲面关系成为本研究需要进一步解决的问题。

致谢: 作者特别感谢美国 Lawrence Livermore National Laboratory 的 Charles Tong 教授和 Yunwei Sun 教授提供 PSUADE 的帮助和相关指导, 以及中国科学院地理科学与资源研究所张永勇博士提供部分数据 (2001-2008 年蚌埠站日径流量)。

表 5 不同目标函数条件下的二阶敏感度
Tab. 5 Two-way interaction sensitivity index based on different objective functions

WB	g2	Kr	Wmi	WM	Kaw	RoughRss	ThickU
g1	0.374	0.027	0.020	0.021	0.026	0.008	0.003
g2		0.179	0.152	0.099	0.094	0.067	0.024
Kr			0.001	0 ^a	0.023	0.001	0 ^a
Wmi				0 ^a	0.003	0.001	0.003
WM					0.020	0.002	0.005
Kaw						0.050	0.007
RoughRss							0.003
NS	g2	Kr	Wmi	WM	Kaw	RoughRss	ThickU
g1	0.369	0.005	0.007	0.003	0.004	0.015	0 ^a
g2		0.097	0.119	0.021	0.018	0.188	0.147
Kr			0.001	0.001	0 ^a	0.018	0 ^a
Wmi				0 ^a	0 ^a	0.013	0.002
WM					0.002	0.030	0.001
Kaw						0.014	0 ^a
RoughRss							0.071
RC	g2	Kr	Wmi	WM	Kaw	RoughRss	ThickU
g1	0.037	0.065	0.088	0.109	0.042	0.062	0.076
g2		0.006	0.006	0.001	0.013	0.051	0.010
Kr			0 ^a	0.006	0.004	0.075	0.003
Wmi				0.001	0.002	0.051	0.002
WM					0.003	0.037	0 ^a
Kaw						0.177	0.004
RoughRss							0.010

注: a-数值数量级为 10⁻⁴, 为保留小数后三位近似取 0。

参考文献(References)

- [1] Song Xiaomeng, Kong Fanzhe. Application of Xinanjiang model coupling with artificial neural networks. *Bulletin of Soil and Water Conservation*, 2010, 30(6): 135-138, 144. [宋晓猛, 孔凡哲. 新安江模型和人工神经网络的耦合应用. *水土保持通报*, 2010, 30(6): 135-138, 144.]
- [2] Li Dan, Zhang Xiang, Zhang Yang et al. Application of interval analysis of sensitivity of parameters of hydrologic model. *Advances in Science and Technology of Water Resources*, 2011, 31(1): 29-32, 41. [李丹, 张翔, 张杨等. 水文模型参数敏感性的区间分析. *水利水电科技进展*, 2011, 31(1): 29-32, 41.]
- [3] Engeland K, Xu C Y, Gottschalk L. Assessing uncertainties in a conceptual water balance model using Bayesian methodology. *Hydrological Science Journal*, 2005, 50(1): 45-63.
- [4] Song Xiaomeng, Zhan Chesheng, Kong Fanzhe et al. A review on uncertainty analysis of large-scale hydrological cycle modeling system. *Acta Geographica Sinica*, 2011, 66(3): 396-406. [宋晓猛, 占车生, 孔凡哲等. 大尺度水循环模拟系统不确定性研究进展. *地理学报*, 2011, 66(3): 396-406.]
- [5] Campolongo F, Cariboni J, Saltelli A. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 2007, 22: 1509-1518.
- [6] Wang Gansheng, Xia Jun, Chen Junfeng. A multi-parameter sensitivity and uncertainty analysis method to evaluate relative importance of parameters and model performance. *Geographical Research*, 2010, 29(2): 263-270. [王纲胜, 夏军, 陈军峰. 模型多参数灵敏度与不确定性分析. *地理研究*, 2010, 29(2): 263-270.]
- [7] Ren Qiwei, Chen Yangbo, Zhou Haolan et al. Global sensitivity analysis of TOPMODEL parameters based on Sobol method. *Yangtze River*, 2010, 41(19): 91-94, 107. [任启伟, 陈洋波, 周浩澜等. 基于Sobol法的TOPMODEL模型全局敏感性分析. *人民长江*, 2010, 41(19): 91-94, 107.]
- [8] Saltelli A, Tarantola S, Campolongo F et al. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons, Ltd., 2004.
- [9] Ren Qiwei, Chen Yangbo, Shu Xiaojuan. Global sensitivity analysis of Xinanjiang model parameter based on Extend FAST method. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 2010, 49(3): 127-134. [任启伟, 陈洋波, 舒晓娟. 基于Extend FAST方法的新安江模型参数全局敏感性分析. *中山大学学报: 自然科学版*, 2010, 49(3): 127-134.]
- [10] van Griensven A, Meixner T, Grunwald S et al. A global sensitivity analysis tool for the parameters of multi-variable catchment model. *Journal of Hydrology*, 2006, 324(1-4): 10-23.
- [11] Fu Xiang, Chu Xuefeng, Tan Guangming. Sensitivity analysis for an infiltration-runoff model with parameter uncertainty. *Journal of Hydrologic Engineering*, 2010, 15(9): 243-251.
- [12] Yang T, Reed P, van Werkhoven K et al. Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. *Water Resources Research*, 2007, 45, W06415, doi: 10.1029/2006WR005813.
- [13] Xu C, Gertner G. Extending a global sensitivity analysis technique to models with correlated parameters. *Computational Statistics & Data Analysis*, 2007, 51(12): 5579-5590.
- [14] Ratto M, Pagano A, Young P. State dependent parameter metamodeling and sensitivity analysis. *Comput. Phys. Comm.*, 2007, 177: 863-876.
- [15] Bonin O. Sensitivity analysis and uncertainty analysis for vector geographical applications//Caetano M, Painho M. 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 16 November 2005, 319-328.
- [16] Jin R, Du X, Chen W. The use of metamodeling techniques for optimization under uncertainty. *Structural and Multidisciplinary Optimization*, 2003, 25(2): 99-116.
- [17] Ascoug II J C. Key criteria and selection of sensitivity analysis methods applied to natural resource models. *International Congress on Modeling and Simulation Proceedings*, Salt Lake City, UT, November 6-11, 2005, 2463-2469.
- [18] Sathyanarayanamurthy H, Chinnam R B. Metamodels for variable importance decomposition with applications to probabilistic engineering design. *Computers & Industrial Engineering*, 2009, 57: 996-1007.
- [19] Lin Y S, Wang J, Wang X M et al. Optimization of butanol production from corn straw hydrolysate by clostridium acetobutylicum using response surface method. *Chinese Science Bulletin*, 2011, 56(14): 1422-1428.
- [20] Sobol I. Sensitivity analysis for non-linear mathematical models. *Mathematical modeling & Computational Experiment*, 1993, 1: 407-414.
- [21] He Zhen. A study on the small sample response surface methodology based on SVM. *Industrial Engineering Journal*, 2006, 9(5): 6-10, 27. [何桢. 基于支持向量机的小样本响应曲面法研究. *工业工程*, 2006, 9(5): 6-10, 27.]
- [22] Cui Qingan. SVM-based nonparametric dual response surface methodology. *Journal of Tianjin University*, 2006, 39(8): 1008-1014. [崔庆安. 一种基于支持向量机的非参数双响应曲面法. *天津大学学报*, 2006, 39(8): 1008-1014.]
- [23] McKay M D. Evaluating prediction uncertainty. Los Alamos National Laboratory Technical Report NUREG/CR-6311,

- LA-12915-MS, 1995.
- [24] Tong C. Self-validated variance-based methods for sensitivity analysis of model outputs. *Reliability Engineering and System Safety*, 2010, 95(3): 301-309.
- [25] Xia Jun, Wang Gangsheng, Lv Aifeng A research on distributed time variant gain modeling. *Acta Geographica Sinica*, 2003, 58(5): 789-796. [夏军, 王纲胜, 吕爱锋. 分布式时变增益流域水循环模拟. *地理学报*, 2003, 58(5): 789-796.]
- [26] Xia Jun, Wang Gangsheng, Tan Ge et al. Development of distributed time-variant gain model for nonlinear hydrological systems. *Science in China: Series D*, 2005, 48(6): 713-723.
- [27] Xia Jun, Ye Aizhong, Wang Gangsheng. A distributed time-variant gain model applied to Yellow River (I): Model theories and structures. *Engineering Journal of Wuhan University*, 2005, 38(6): 10-15. [夏军, 叶爱中, 王纲胜. 黄河流域时变增益分布式水文模型(I): 模型的原理与结构. *武汉大学学报: 工学版*, 2005, 38(6): 10-15.]
- [28] Xia Jun, Ye Aizhong, Qiao Yunfeng et al. An application research on distributed time-variant gain hydrological model in Wuding River of Yellow River. *Journal of Basic Science and Engineering*, 2007, 15(4): 457-465. [夏军, 叶爱中, 乔云峰等. 黄河无定河流域分布式时变增益水文模型的应用研究. *应用基础与工程科学学报*, 2007, 15(4): 457-465.]
- [29] Tong C, Graziani F. A practical global sensitivity analysis methodology for multi-physics applications. *Computational Methods in Transport. Lecture Notes in Computational Science and Engineer*, 2008, 62: 277-299.
- [30] Tong C. Quantifying uncertainties of a soil-foundation structure-interaction system under seismic excitation. *Lawrence Livermore National Laboratory Technical Report UCRL-TR-402883*, CA, 2008.

An Efficient Quantitative Sensitivity Analysis Approach for Hydrological Model Parameters Using RSMSobol Method

KONG Fanzhe¹, SONG Xiaomeng¹, ZHAN Chesheng², YE Aizhong³

(1. School of Resource and Earth Science, China University of Mining & Technology, Xuzhou 221008, Jiangsu, China;

2. Key Laboratory of Water Cycle & Related Land Surface Processes,

Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;

3. State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China)

Abstract: Sensitivity analysis of hydrological models is a key step for model uncertainty quantification. It can identify the dominant parameters, reduce the model calibration uncertainty, and enhance the model optimization efficiency. However, how to effectively validate a model and identify the dominant parameters for a large-scale complex distributed hydrological model is a bottle-neck to achieve the parameters optimization. There are some shortcomings for classical approaches, e.g. time-consuming and high computation cost, to quantitatively assess the sensitivity of the multi-parameters complex hydrological model. For this reason, a new approach was applied in this paper, in which the support vector machine was used to construct the response surface (a surrogate model) at first. Then it integrated the SVM-based response surface with the Sobol method, i.e. the RSMSobol method, to achieve the quantification assessment of sensitivity for complex models. Taking the distributed time-variant gain model in the Huaihe River Basin as a case study, we selected three objective functions (i.e. water balance coefficient WB, Nash-Sutcliffe efficiency coefficient NS, and correlation coefficient RC) to assess the model as the output responses for sensitivity analysis. The results show that the RSMSobol method can not only achieve the quantification of the sensitivity, and also reduce the computational cost, with good accuracy compared to the classical approaches.

Key words: meta-modeling approach; response surface methodology; sensitivity analysis; support vector machines; Huaihe River Basin