

# 基于再抽样和析因分析的生态水文系统不确定性分析技术研究

水文系统作为复合生态系统的关键组成之一，其呈现出异常复杂的多功能性、多目标性和动态性。生态水文系统的这种多层次、多组分和高互动的特征在珠三角地区尤为显著。一方面，珠三角地区河网密布交错，自然生态系统和流域水文系统之间存在着多重、复杂的互动关系。受人类活动和气候变化的共同影响，近 10 年来珠三角地区的生态水文关系显著恶化，如地表植被退化、土壤酸化等生态问题，直接影响了水文循环的产汇流过程，致使地表径流量增大、水土流失严重，进而演变成洪水、泥石流等自然灾害。水文情势的突变进一步加剧生态系统的退化，加速生态系统服务功能和生物多样性的丧失。水文和生态系统之间相互作用，紧密关联。另一方面，随着珠三角地区城市化进程的加快，各类开发建设项目也日益增多，开发类型和规模前所未有，由此引发的城市水土流失问题也愈发严重。水土流失不仅破坏城市地表生态景观、打乱地表水系、污染饮用水源，而且破坏城市生态环境、生活环境和投资环境，进而制约社会经济的可持续发展。因此，解析珠三角地区生态水文系统的复杂互动关系，辨识系统在模拟过程中呈现出的异常复杂的不确定性，是研究珠三角复合生态系统的重要组成部分。

基于此，为了定量揭示珠三角地区复合生态水文系统的多层复杂性，我们耦合了再抽样和方差分析方法，开发了再抽样析因不确定性定量分析技术。将该技术用于珠三角生态水文系统的研究，以表征在模拟过程中出现的不确定性、追踪水文模型响应的不确定性来源，并量化每一个不确定性来源及其交互作用的影响。研究结果有利于揭示珠三角地区水文生态系统的复杂性、不确定性和动态性。

本研究运用了三参数简化模型和四参数日径流水文模型两个案例阐述所开发方法的有效性，并将其应用于珠三角增江流域的日径流模拟，以揭示流域生态水文的复杂性。

传统的基于 ANOVA 的敏感性分析方法概括为：

$$SST = \sum_{i=1}^k SS_i + \sum_{i=1}^k \sum_{j>1}^k SS_{ij} + \dots + SS_{1, 2, \dots, k}$$

其中  $SS_i$  代表单因子效应， $SS_{ij}$  至  $SS_{1, 2, \dots, k}$  代表  $k$  个因子之间的交互作用，

SST 代表所有因子和交互的方差和。

单二次抽样 ANOVA 方法指只对模型中一个参数进行二次抽样, 其表达式如下所示:

$$SST^j = \sum_{t_1=1}^{T_1} \sum_{t_2=1}^{T_2} \cdots \sum_{h=1}^2 \cdots \sum_{t_k=1}^{T_k} (Y^{t_1, t_2 \cdots g(h, j) \cdots t_k} - Y^{o, o \cdots g(o, j) \cdots o})^2$$

当  $i = n$  时,

$$SS_i^j = T_1 T_2 \cdots T_{n-1} T_{n+1} \cdots T_k \sum_{h=1}^2 (Y^{t_1, t_2 \cdots g(h, j) \cdots t_k} - Y^{o, o \cdots g(o, j) \cdots o})^2$$

当  $i \neq n$ ,

$$SS_i^j = 2 \times T_1 T_2 \cdots T_{i-1} T_{i+1} \cdots T_{n-1} T_{n+1} \cdots T_k \sum_{t_i=1}^{T_i} (Y^{o, o \cdots t_i \cdots g(o, j) \cdots o} - Y^{o, o \cdots g(o, j) \cdots o})^2$$

多二次抽样方法指多个模型参数同时被二次抽样, 可以表达为:

$$SST^j = \sum_{t_1=1}^{T_1} \sum_{t_2=1}^{T_2} \cdots \sum_{h_p=1}^2 \cdots \sum_{h_q=1}^2 \cdots \sum_{t_k=1}^{T_k} (Y^{t_1, t_2 \cdots g(h_p, j_p) \cdots g(h_q, j_q) \cdots t_k} - Y^{o, o \cdots g(o, j_p) \cdots g(o, j_q) \cdots o})^2$$

当  $i = p, \cdots, q$  时,

$$SS_i^j = T_1 \times T_2 \times \cdots \times T_k \sum_{h_p=1}^2 \cdots \sum_{h_q=1}^2 (Y^{t_1, t_2 \cdots g(h_p, j_p) \cdots g(h_q, j_q) \cdots t_k} - Y^{o, o \cdots g(o, j_p) \cdots g(o, j_q) \cdots o})^2$$

当  $i \neq p, \dots, q$ ,

$$SS_i^j = 2 \times \cdots \times 2 \times T_1 \times T_2 \cdots T_{i-1} \times T_{i+1} \cdots T_k \sum_{t_i=1}^{T_i} (Y^{o, o \cdots t_i \cdots g(o, j_p) \cdots g(o, j_q) \cdots o} - Y^{o, o \cdots g(o, j_p) \cdots g(o, j_q) \cdots o})^2$$

全二次抽样方法指模型中所有参数都被二次抽样, 数学表达式如下:

$$SST^j = \sum_{t_1=1}^2 \sum_{t_2=1}^2 \cdots \sum_{t_k=1}^2 (Y^{g(h_1, j_1), g(h_2, j_2) \cdots g(h_k, j_k)} - Y^{g(o, j_1), g(o, j_2) \cdots g(o, j_k)})^2$$

$$SS_i^j = \sum_{h_1=1}^2 \sum_{h_2=1}^2 \cdots \sum_{h_k=1}^2 (Y^{g(h_1, j_1), g(h_2, j_2) \cdots g(h_k, j_k)} - Y^{g(o, j_1), g(o, j_2) \cdots g(o, j_k)})^2$$

以一个简单的三参数模型为例, 该模型表达式为:

$$F_3(x_1, x_2, x_3) = x_1 \times x_3 + x_1 \times \sin\left(\frac{\pi}{2} \times x_2\right) + x_2 \times e^{|x_3|} + x_1 \times x_2 \times x_3$$

其中， $x_1$ ， $x_2$ ， $x_3$ 是独立变量，且在 $[0, 1]$ 之间均匀分布。

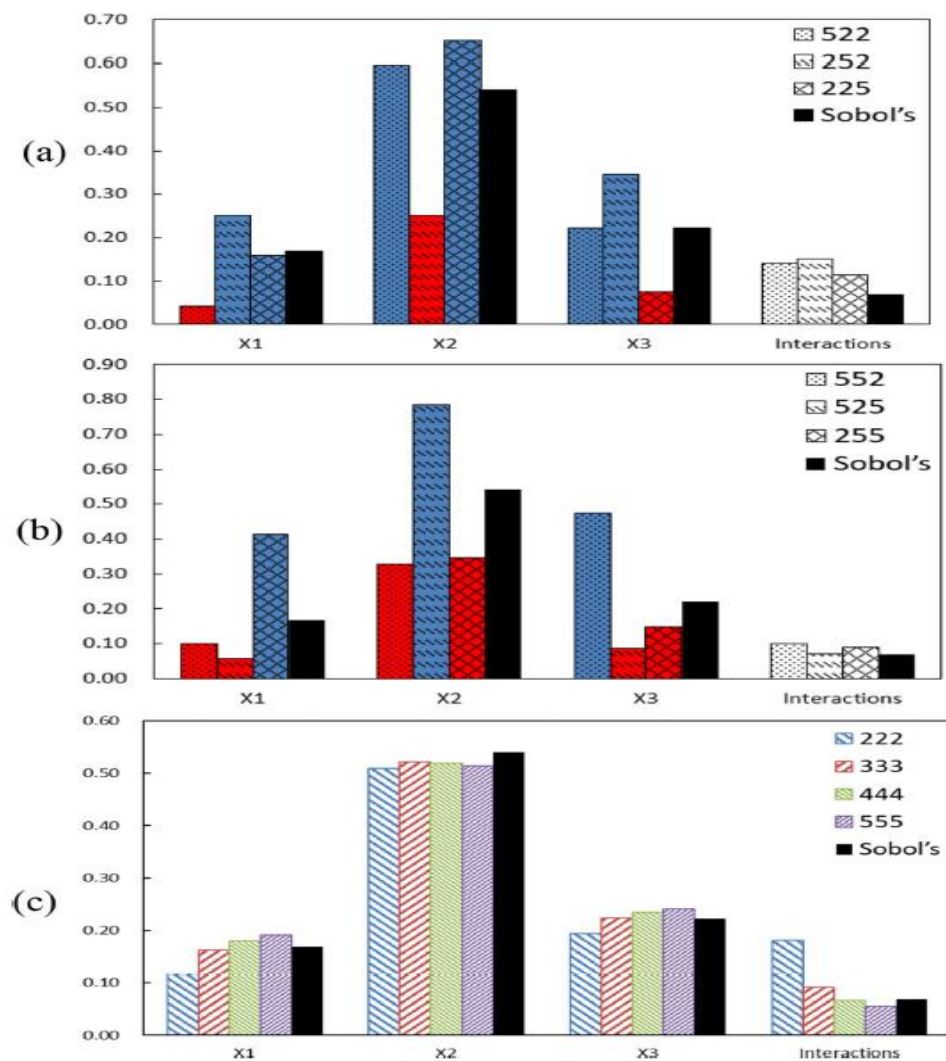


图 1 三个参数及其交互的敏感性：(a) 单二次抽样、(b) 多二次抽样、  
(c) 全二次抽样

本研究采用单二次抽样、多二次抽样和全二次抽样方法对三个参数及其交互作用的敏感性进行定量分析，敏感性结果如图 1 所示。结果表明，应用三种方法得到的参数敏感性具有显著差异。以 Sobol's 敏感性结果为基准，可以看出，参数的单因子敏感性会因为二次抽样而显著降低。这说明单二次抽样方法和多二次抽样方法会显著低估参数的敏感性。此外，对于全二次抽样方法而言，三个参数的个体敏感性和相互作用敏感性随参数水平的变化而变化。随着参数水平从 222 增加到 555， $x_1$  和  $x_3$  的个体敏感性分别从 11.7% 和 19.4% 逐渐增加到 19.1% 和

24.1%。同时，交互参数灵敏度从 18.1%逐渐下降到 5.5%。 $x_2$  的个体灵敏度保持相对稳定，从 50.9%到 52.2%不等。单参数和参数间交互作用的敏感性会受到二次抽样参数水平的影响。增加抽样参数的水平数会略微增加该参数的敏感性。

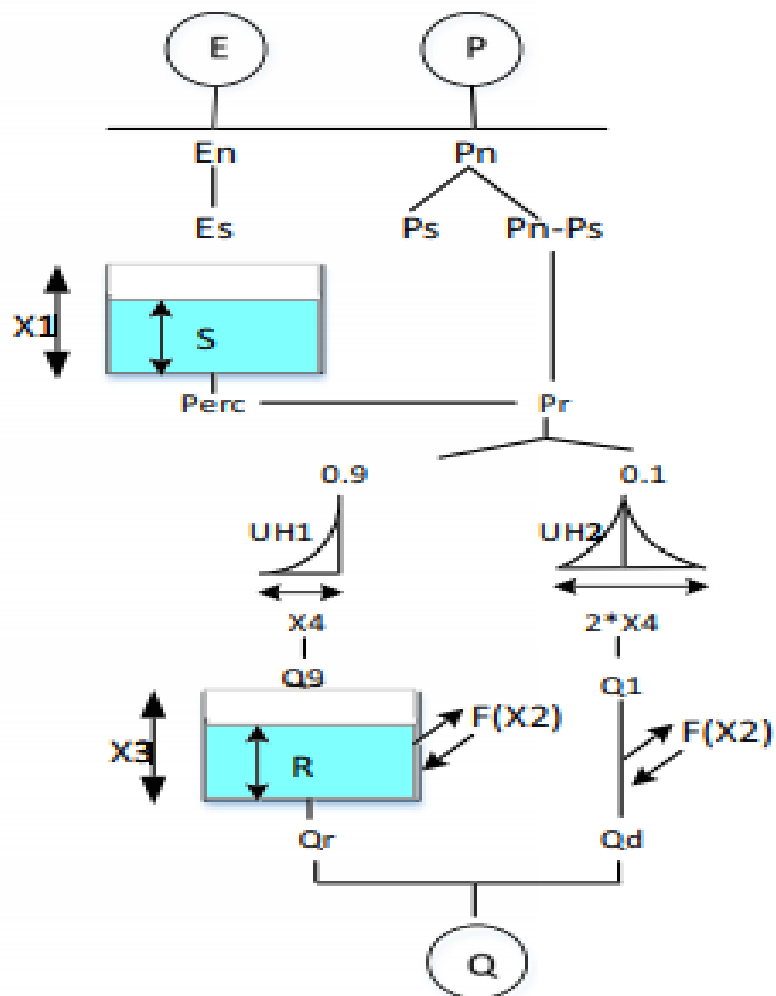


图 2 GR4J 模型结构图。

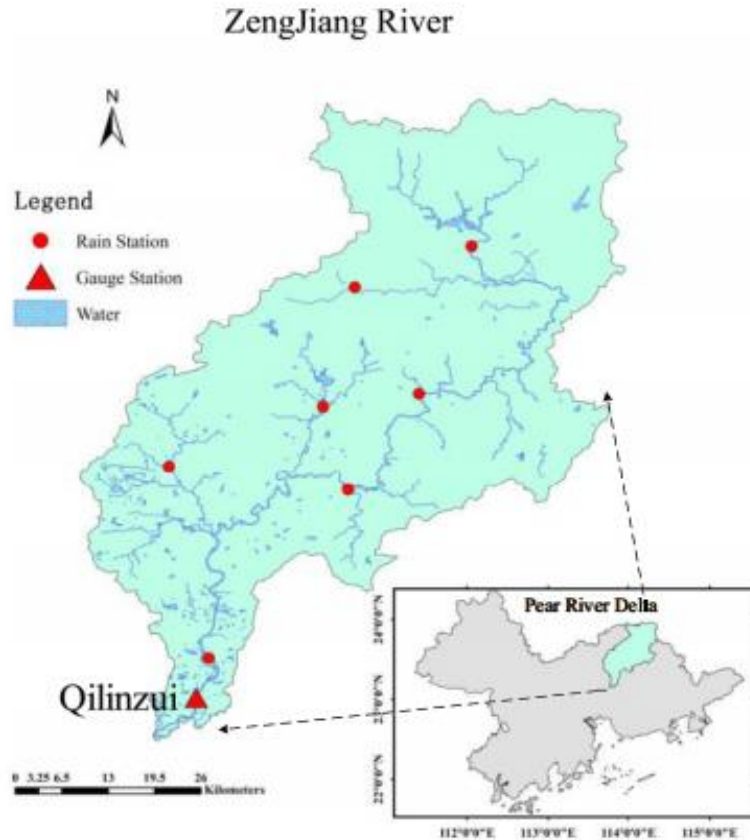


图 3 增江流域示意图。

为进一步研究上述三种方法在水文模拟中的可用性,这三种方法被应用于概念式水文模型 GR4J 中(详见图 2),以模拟增江流域(详见图 3)的径流量。GR4J 是降雨径流模型之一,在流域水文建模方面具有很强的基础性和有效性。该模型采用产流水库、汇流水库这两个线性水库进行产汇流计算,对湿润地区进行洪水预报和水资源规划,具有简便、准确等特点。该模型只包含 4 个参数,分别是:  $x_1$  产流水库容量(mm)、  $x_2$  地下水交换系数(mm)、  $x_3$  汇流水库容量(mm)、  $x_4$  单位线汇流时间(day)。本文选用 GR4J 模型进行水文模拟和预报研究。其次,将该模型应用于增江,其为东江支流,是珠三角地区重要的流域之一。本研究所使用的数据(日蒸发量、日降水量和日径流)均来自于麒麟咀水文站,数据采集的时间范围为 2009–2015 年。麒麟咀水文站以上流域总面积为 2866 km<sup>2</sup>, 占增江流域(3160 km<sup>2</sup>) 的 91%。流域年平均气温和降水量分别为 21.6 °C 和 2188 mm。

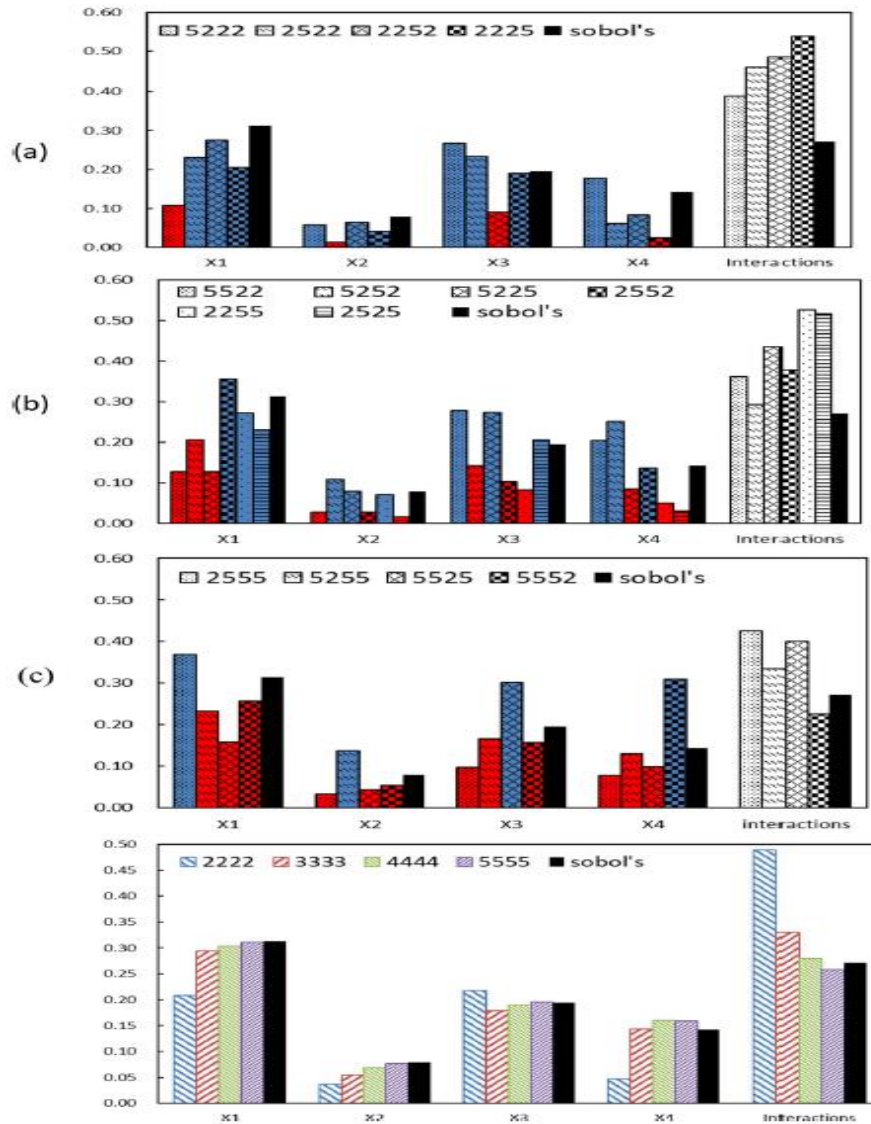


图 4 GR4J 模型参数及其交互在不同二次抽样方法下的敏感性。

模型参数及其交互效应的敏感性如图 4 所示。结果表明，增江流域的径流量对  $x_1$  的变化最为敏感。与案例研究 1 结果类似，二次抽样会在一定程度上降低参数的敏感性，且各参数及其交互作用的敏感性排序也受到抽样方法的影响。此外，在全采样方差分析方法中，所有参数都在其变化范围内以不同级别进行子采样。在这项研究中，将测试四种情景，每个参数分别具有 2、3、4 或 5 个级别。如图 4 所示，4 个参数的个体敏感性和相互作用敏感性随参数水平的变化而变化。对于全二次抽样方法而言，随着参数水平的增加（从 2 水平至 5 水平），相应参数的敏感性也逐渐增加。参数  $x_1$ ， $x_2$  和  $x_4$  的敏感性分别从 20.1、3.7 和 4.7% 增

长到 31.0、7.6 和 15.8%。与此同时，参数  $x_3$  的敏感性从 21.7%减少到 17.8%，而交互效应的敏感性从 48.9%减少到 25.9%。结果表明，参数水平将影响全采样方差分析方法中的个体和交互敏感性。具体来说，最敏感的参数和交互作用的灵敏度一般会降低，而其他参数的灵敏度随着参数级别的增加而增加。

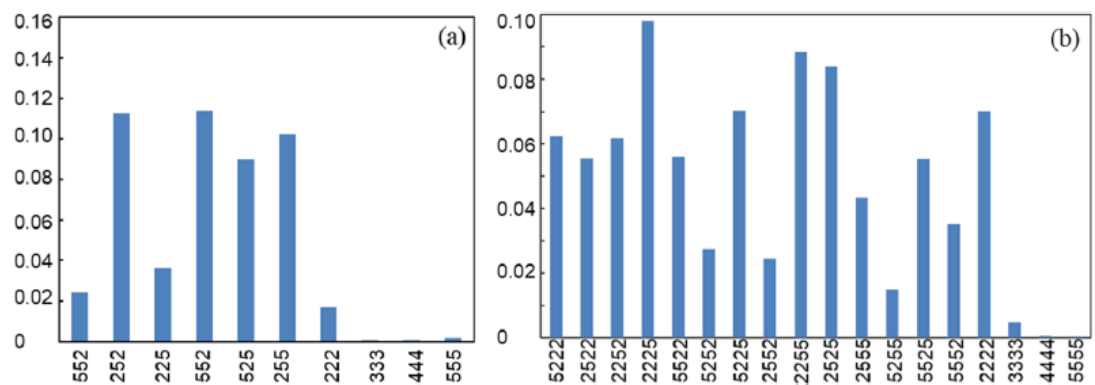


图 5 基于 Sobol's 法与二次抽样法的水文模型参数敏感性的对比分析。

本研究还对比了上述三种二次抽样 ANOVA 算法与 Sobol's 算法在计算模型参数敏感性方面的差异（如图 5 所示）。结果表明，与单二次抽样 ANOVA 算法和多二次抽样 ANOVA 算法相比较，全二次抽样 ANOVA 算法可以产出更为可靠的结果。此外，为得到较为可靠的敏感性分析结果，建议每个参数至少设置 3 至 4 个水平。许多研究表明 Sobol's 算法计算负担较重。相比而言，二次抽样 ANOVA 算法是一种更高效的算法。例如，对于案例 1 来说，Sobol's 算法需要运行 10000 次以得到稳定的结果，而二次抽样 ANOVA 算法只需要 256 次。尽管相较于 Sobol's 算法，二次抽样 ANOVA 算法并不能提供更好的结果，但是它极大地降低了运算要求，且可用于非数值化分析。



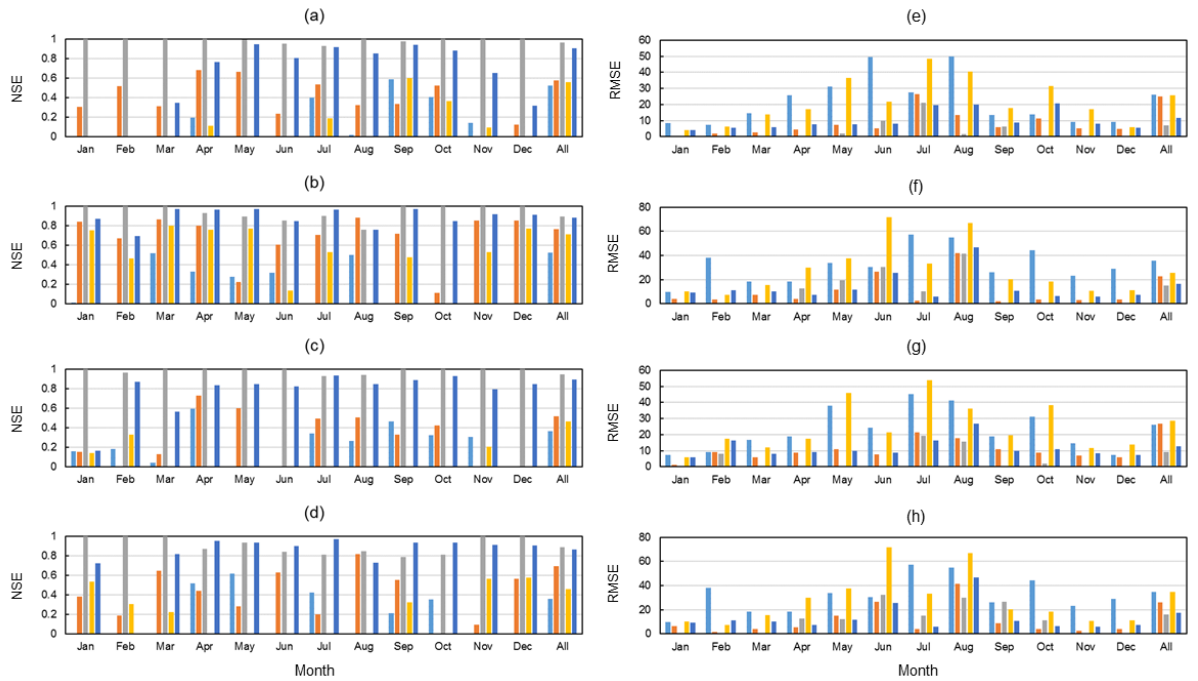


图 6 离散主单调推断(DPMI)和线性回归(LM)校正月、年水文模拟的验证精度。

此外，水文模拟受到各种偏差和不确定性的挑战(例如，水文模型差异、偏差修正方案、预测因子选择、径流大小和时间尺度等)。为了弥补现有对系统偏差和不确定性分析研究的不足，本研究全面分析了偏差修正水文模拟中气象条件与径流在多重不确定性下的复杂响应。这将通过开发基于离散主单调推理(DPMI)的综合偏差修正水文模拟不确定性量化(DiBHUQ)方法来实现。DiBHUQ 的有效性将通过北江支流中洲河流域的偏差修正水文模拟的案例研究得到验证(如图 6 所示)。

在本研究中，应用 DiBHUQ 的方法对水文模拟和偏差修正系统进行了全面的偏差和不确定性分析。这些偏差不确定性被分为两组，即，方法不确定性和系统不确定性。前者包括 2 个水文模型，5 个偏差修正方案，5 个预测因子组合，11 个性能指标；后者包括 6 个分位数区间和 2 个时间尺度。首先，本研究探讨了 DPMI 在偏差修正中的可行性。其次，通过该方法揭示了一系列关于这些偏差不确定性的发现。这些研究成果对中洲河的流域的水文模拟、流域管理以及预防洪涝灾害具有重要意义。一些成果如下：

(a) DPMI 与常用的偏差修正方法(如线性回归)在多各方面偏差修正的准确



性进行了系统的比较。DPMI 在偏差修正方面性能较多元线性回归提升在 4%到 16% 之间。DPMI 在雨季有一定的不足，但在低流量模拟修正方面有很大的优势。对 DPMI 建模过程的分析表明，在时滞的修正组合中，气象因素对偏差修正的贡献要大于模拟量。其原因是由于随着时滞预测因子数量的增加，模拟流场的影响减小，温度成为修正径流模拟偏差的主导因素。

(b) 由于水文过程与模型概念化过程的不完全匹配，水文模型的不确定性对水文模拟影响较大。因此，采用多种水文模型进行相关研究是必要的。同时，偏差修正方案(如样本选择和排序)也是水文模拟不确定度的主要来源之一。根据模拟径流随机抽样排序数据的方案和随机抽样原始数据的方案(即 SOR)提供最佳的偏差修正精度。在所有偏差修正方案和所有预测因子组合中，偏差修正的精度在旱季和雨季之间表现出一致的季节间差异。预测因子的选择虽然比偏差修正方案和水文模型的影响小，但其影响不容忽视。通过对不同预测因子条件下的建模精度的比较，表明流域气象条件与径流的主要滞后时间均在 2 天左右。

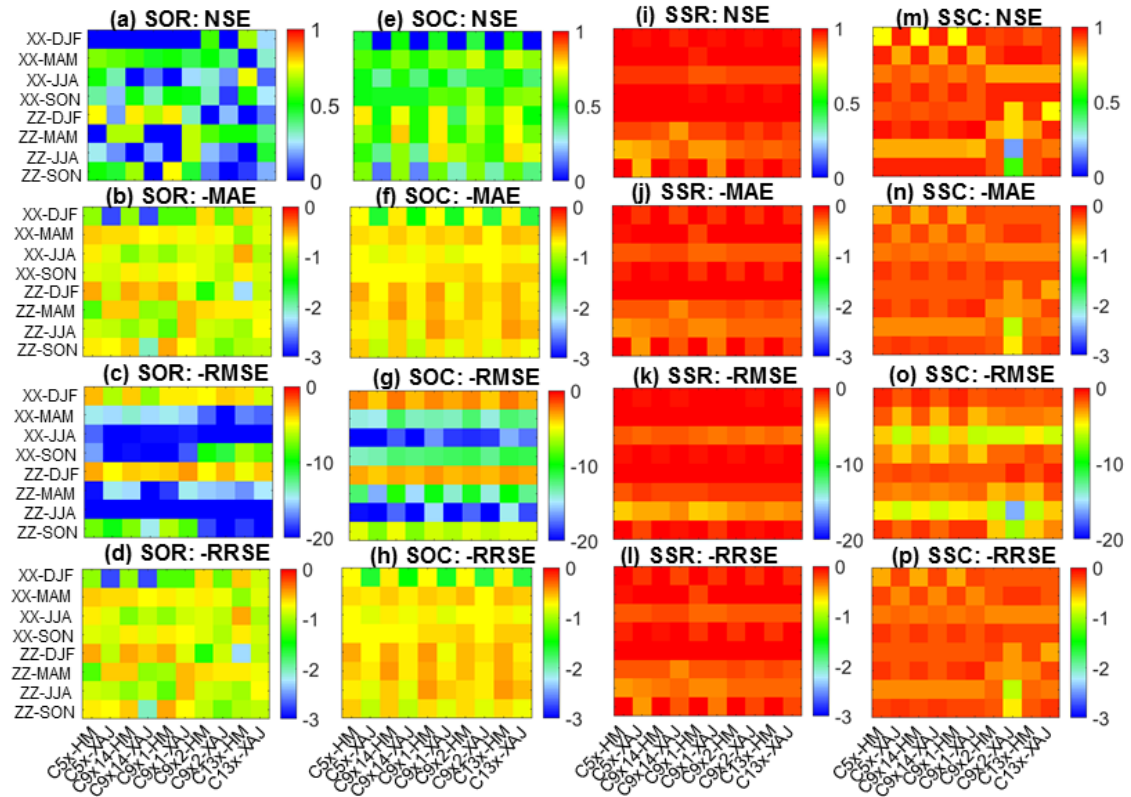


图 7 模拟性能的季节间变化(x 轴:预测因子选择与水文模型的组合;y 轴:流域和季节的组合)。



图 8 水文模型性能指标的分布和相关性。

图 8 展示了水文模型中性能指标的分布和相关性，结果表明，变异系数更大的指标具有更显著的区分度，所以本研究选择的指标是 Nash-Sutcliffe 效率系数（NSE）、均方根误差（RMSE）和平均绝对误差（MAE）。此外，大多数指标在正相关或负相关方面具有显著性。变异系数表明，这 9 个指标的离散化程度较低，而 MAE 和 RMSE 的离散化程度较高。

表 1 基于偏差修正的不同水文模型的预测准确度。

	SUC	DiPMI-SOC	DiPMI-SOR	LM-SOC	LM-SOR
Pearson	0.83	0.76	0.98	0.74	0.95
Kendall	0.61	0.63	0.99	0.54	0.82
Spearman	0.79	0.81	0.99	0.72	0.95
MAE	12.41	11.75	0.58	14.33	5.93
RMSE	26.19	25.14	6.99	25.88	11.76
NRMSE_1	0.074	0.046	0.012	0.046	0.021
NRMSE_2	0.28	0.26	0.02	0.27	0.05
RAE	1.07	1.10	1.13	1.10	1.09
NASH	0.52	0.57	0.96	0.55	0.90
RRSE	0.69	0.64	0.18	0.66	0.30
IOA	0.89	0.85	0.99	0.84	0.97

表 1 展示了基于 Hymod 模拟和 DiPMI 偏差修正的各种验证准确度，包括未修正对照组和线性回归（LM）修正对照组。在不同的流域、水文模型、时间尺度、偏差修正方案和性能指标的条件下，DiPMI 偏差修正和 LM 偏差修正的结果表明，与 LM 相比，DiPMI 作为一种偏差修正方法具有显著的优势。研究发现 DiPMI 在所有修正条件下性能均优于传统方法。同时，在消除月尺度水文模拟偏差方面，非一致性分布方案（SSC）优于其它方案。此外，两流域气象条件的水文响应具有显著的一日滞后性。整体而言，DiPMI 具有更好的适应性和稳定性。在输入条件不佳的情况下，DiPMI 仍然可以有效地完成偏差修正。

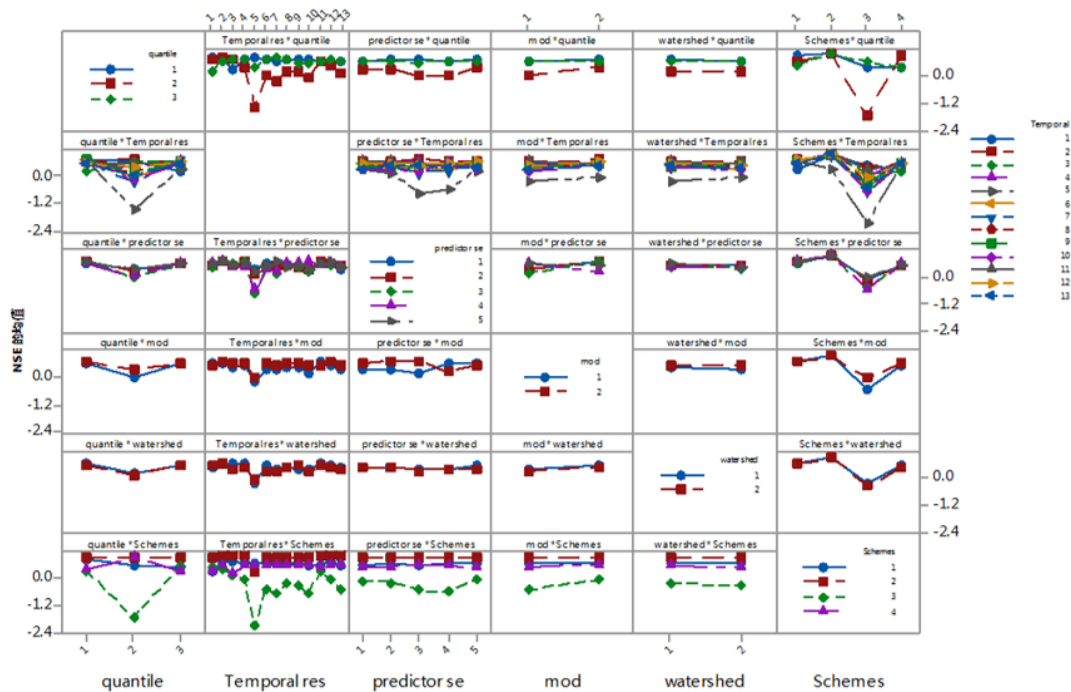


图 9 全因子交互效应图。图中由左到右因子依次为：分位数区间、时间尺度、预测因子组合、水文模型、流域和修正方案。

此外，本研究还可以在各个方面进行改进。例如，数据不确定性、参数不确定性、空间尺度不确定性和气候模型的不确定性，这些不确定性可能对水文模拟产生重要影响。同时，对预测因子相互作用的分析可能有助于提高模拟效率，如在偏差修正中进行系统的因子分析，能够全面揭示各种不确定性之间的相互作用。因此，本研究开展了系统的因子分析以揭示因子间内在联系。图 9 展示了 DiPMI 模型的全因子交互效应图。由图可知，六个因子间都存在或强或弱的交互作用，其中，模型结构和其它因子（如模型参数、偏差修正方法）的交互作用对 NSE 的影响极小。结果表明 DiPMI 模型不确定性贡献度由大到小排序为：偏差修正方案、水文模型、流域和自变量选择。综上所述，DiPMI 模型可通过聚类过程帮助我们了解流域的气象与水文交互过程。

表 2 气候预测因子和径流之间的相关性

站点	气候预测因子								
	降雨	气压	风速	平均温度	水汽压	相对湿度	日最低温	日高温	相对湿度
西江梧州	0.52	-0.75	-0.27	0.71	0.77	0.41	0.72	0.69	0.08
北江石角	0.42	-0.56	-0.19	0.50	0.55	0.47	0.51	0.46	0.01
北江横石	0.79	-0.53	-0.18	0.42	0.52	0.59	0.47	0.36	-0.05
东江博罗	0.71	-0.61	0.04	0.54	0.61	0.55	0.58	0.48	-0.08

（注：灰色背景框代表相关性没有通过 95%置信度的显著性检验。）

为了识别影响珠三角地区径流的最主要气象因子，本研究分析了 9 种气象因子和径流之间 95%置信度下的相关性，如表 2 所示。由表 2 可以看出，气压和风速和径流之间存在负的相关性，其余 7 种气象因子和径流之间显示正的相关性。与此同时，气象因子和径流的关系存在地域差异。例如，西江梧州站的径流和降雨之间的相关性是 0.52，而北江横石站的径流和降雨之间的相关性是 0.79。水汽压对径流的影响在西江比较大（相关性 0.77），对北江径流的影响相对比较小（相关性 0.52）。相关性的分析是为了筛选出最主要的气候预测因子，为建立统计模型的提供理论依据。

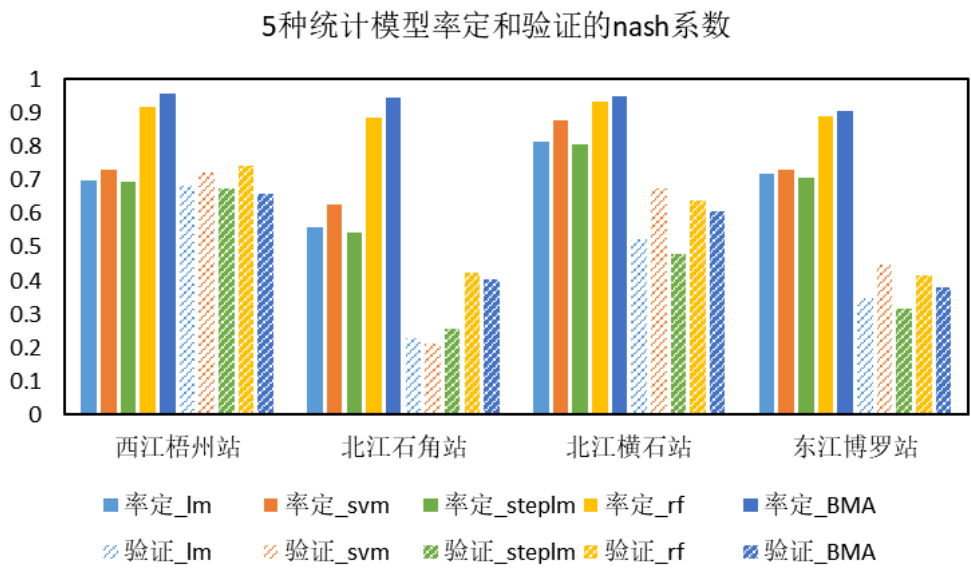


图 10 五种统计模型在 4 个站点的率定和验证效果

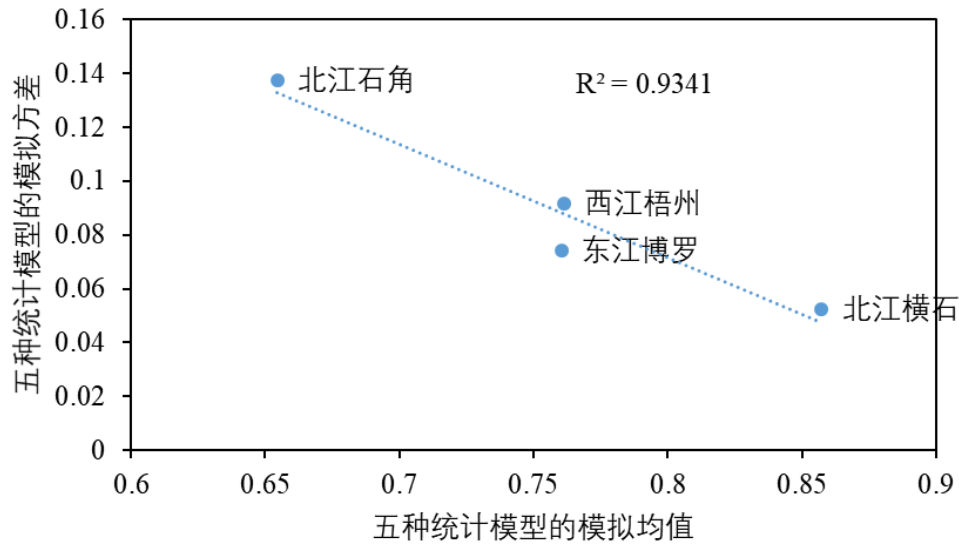


图 11 五种统计模型在 4 个站点模拟效果的均值和方差。

其次，本研究采用五种统计模型建立珠三角地区径流预测模型。这五种统计模型分别为线性回归模型(lm)，支持向量机模型(svm)，逐步回归模型(step1m)，随机森林模型(rf)以及基于贝叶斯模型平均的集合模型(BMA)。本研究用 25 年历史数据(1955 - 1979)做模型率定，5 年历史数据(1980 -1984)做验证。图 10 展示了五种统计模型在 4 个站点的率定和验证效果(以 nash 系数为例)。从图 10 可以看出，5 种统计模型在不同站点模拟效果不同，其中随机森林模型和贝叶斯集合模型在率定期表现最好。然而在验证期，这两种模型并不一定是最优的模型。例如在北江横石站和东江博罗站，支持向量机模型在率定期有最优的模拟效果(nash 最高)。但整体而言，率定期的模拟效果优于验证期的模拟效果，这说明气候因子和径流的关系在验证期和率定期发生了变化。相比于西江，北江和东江的降水径流关系年际变化较大、且更易受模型不确定性的影响。另外，分析五种模拟模型在 4 个站点模拟效果的均值和方差，如图 11 所示。可以看出，统计模型之间的差异和模拟效果之间存在显著的线性关系( $R^2=0.93$ )。这说明，模型准确度由站点气候因子-径流关系决定，受模型结构差异的影响较弱。不同站点，气候因子对径流的解释度不同。



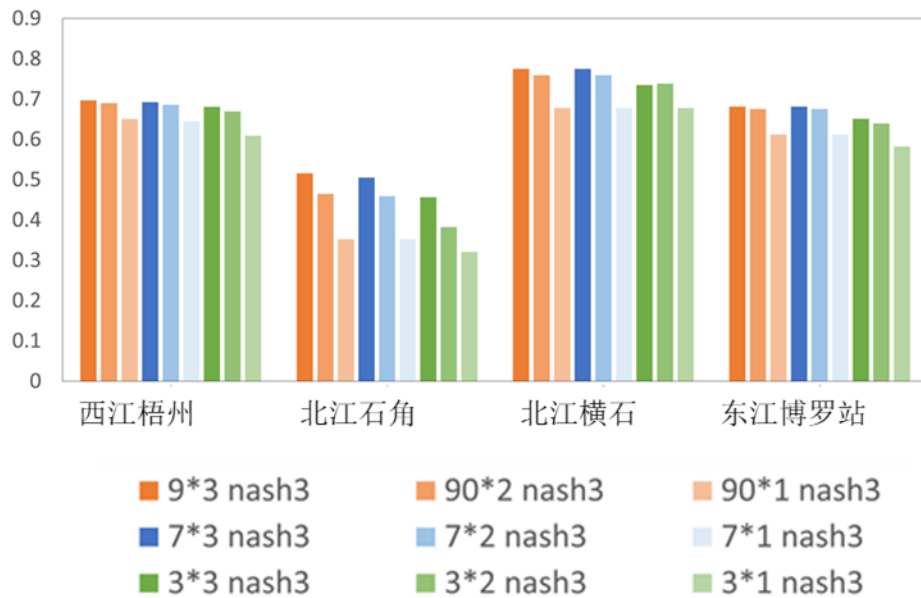


图 12 不同预测因子组合下集合预报模型在 4 个站点模拟效果

此外，为了定量分析当月气候因子，前期气候因子，以及下垫面和人类活动的影响，本研究选择不同气候因子种类，并考虑不同时长的前期气候，针对 9 种组合情景进行模拟分析，结果如图 12 所示。

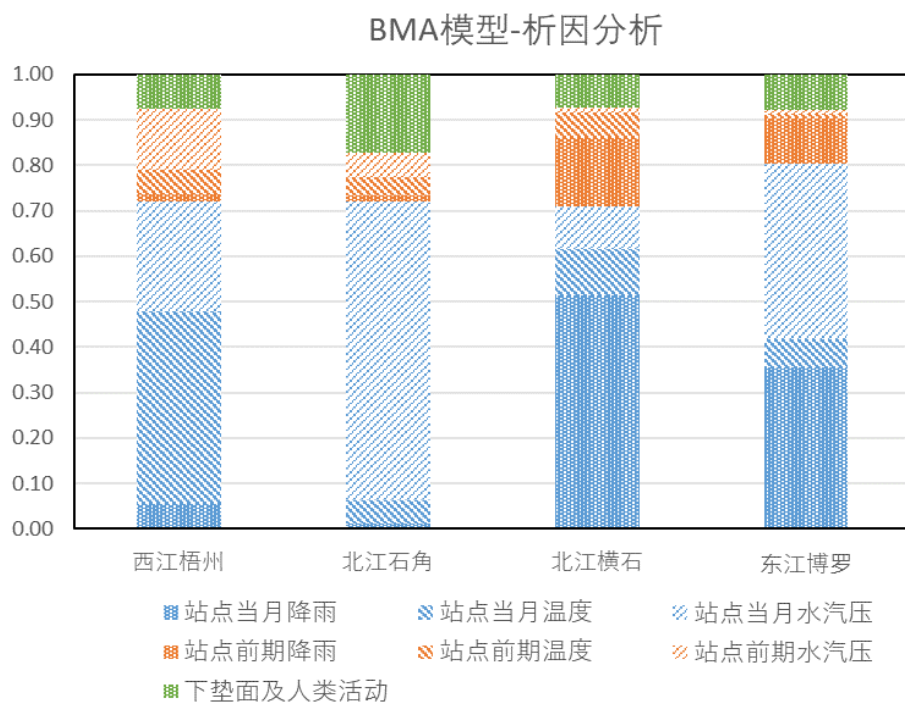


图 13 基于 BMA 多模型集合的多水平-析因分析。



---

最后，进一步利用析因分析技术，对图 12 模拟效果进行解析，结果如图 13 所示。由图可以看出，降雨对径流变化的贡献随地域变化。如北江横石站，本地当月降雨对径流变化的贡献是 51%，而在西江梧州站，本地当月降雨对径流的贡献仅为 5%。对北江石角站而言，当地的水汽压对径流变化的影响最大。整体而言，在珠三角地区，站点当月气候因子对径流变化起主要贡献（71%-80%），其次是前期气候条件（11%-22%），最后是下垫面变化和人类活动的影响（7%-17%）。

综上所述，解析珠三角地区生态水文系统的复杂互动关系，辨识系统在模拟过程中呈现出的异常复杂的不确定性，是研究珠三角复合生态系统的重要组成部分。因此，为了定量揭示珠三角地区复合生态水文系统的多层复杂性，我们耦合了再抽样和方差分析方法，开发了再抽样析因不确定性定量分析技术。将该技术用于珠三角生态水文系统的研究，以表征在模拟过程中出现的不确定性、追踪水文模型响应的不确定性来源，并量化每一个不确定性来源及其交互作用的影响。本研究有利于对复合生态系统进行结构化、多层式的系统辨识，剖析其繁杂的内在系统特征和互动机理。同时，研究结果可为生态风险预报预警及生态安全保障提供科学可靠的决策支持。