

# SHUD 模拟系统训练营学习笔记

---

## SHUD 模拟系统训练营学习笔记

### 前言

#### 1. SHUD 系统构成与基本原理回顾

- 1.1. 水文模型的分类与发展简史
  - 1.1.1. 总集式模型 (*Lumped Models*)
  - 1.1.2. 分布式模型 (*Distributed Models*)
  - 1.1.3. 物理模型 (*Physically-Based Models*)
  - 1.1.4. 概念模型 (*Conceptual Models*)
- 1.2. SHUD 模拟系统构成: SHUD, rSHUD, AutoSHUD
  - 1.2.1. SHUD (Simulator for Hydrologic Unstructured Domains)
  - 1.2.2. rSHUD
  - 1.2.3. AutoSHUD
- 1.3. SHUD 模型的发展历程 (与 PIHM 的渊源及区别)
  - 1.3.1. PIHM (Penn State Integrated Hydrologic Model)
  - 1.3.2. SHUD 与 PIHM 的关键区别
- 1.4. SHUD 模型的核心特点与优势
- 1.5. SHUD 系统的应用领域与价值
- 1.6. 引用规范与知识产权说明

#### 2. 系统环境与软件安装 (实操指南)

- 2.1. 目标平台选择 (Linux/Windows/Mac)
- 2.2. HPC 环境基础 (Slurm 简介, 远程连接工具 MobaXterm/Terminal 使用)
  - 2.2.1. 远程连接 (SSH)
  - 2.2.2. Slurm 任务管理系统
  - 2.2.3. 文件管理
- 2.3. 实操: 安装 SHUD 依赖库 (SUNDIALS/CVODE)
  - 2.3.1. 理解依赖库的作用
  - 2.3.2. 安装方法
  - 2.3.3. 常见问题
- 2.4. 实操: 编译 SHUD 模型可执行文件
  - 2.4.1. 理解 `Makefile` 文件的作用
  - 2.4.2. 详细步骤
  - 2.4.3. 检查编译结果
  - 2.4.4. 常见问题
- 2.5. 实操: 安装 rSHUD R 包
  - 2.5.1. 详细步骤
  - 2.5.2. 检查安装
  - 2.5.3. 常见问题
- 2.6. AutoSHUD 脚本环境配置与依赖
  - 2.6.1. 脚本结构
  - 2.6.2. 依赖

#### 3. SHUD 模型所需原始数据

- 3.1. 核心数据列表与要求
- 3.2. 数据实例

#### 4. 全球开放地理数据源介绍及获取

- 4.1. 常见全球数据集介绍
- 4.2. GHDC 平台的数据获取方法 (实操指南)
  - 4.2.1. 理解 GHDC 的功能
  - 4.2.2. 平台访问

- 4.2.3. 数据获取流程概述
- 4.2.4. 详细步骤：上传流域边界
- 4.2.5. 详细步骤：配置数据需求
- 4.2.6. 详细步骤：提交任务与下载数据
- 4.2.7. 理解 GHDC 生成的数据结构
- 4.2.8. GHDC 平台的局限性

## 5. SHUD 模型输入文件详解 (空间与水力学参数)

- 5.1. SHUD 模型输入文件结构 (复习与整合)
- 5.2. 空间数据文件 (.sp.\*) 详解
  - 5.2.1. .sp.mesh 文件
  - 5.2.2. .sp.att 文件
  - 5.2.3. .sp.riv 文件
  - 5.2.4. .sp.rivseg 文件
- 5.3. 水力学参数文件 (.para.\*) 详解
  - 5.3.1. .para.soil 文件
  - 5.3.2. .para.geol 文件
  - 5.3.3. .para.lc 文件

## 6. SHUD 模型输入文件详解 (配置文件与时间序列)

- 6.1. 模型配置文件 (.cfg.\*) 详解
  - 6.1.1. .cfg.para 文件
  - 6.1.2. .cfg.calib 文件
  - 6.1.3. .cfg.ic 文件
- 6.2. 时间序列数据文件 (.tsd.\*) 详解
  - 6.2.1. .tsd.forc 文件与关联的 .csv 文件
  - 6.2.2. .tsd.lai 文件
  - 6.2.3. .tsd.rl 文件
  - 6.2.4. .tsd.mf 文件
  - 6.2.5. .tsd.obs 文件

## 7. SHUD 模型输出文件详解

- 7.1. 输出文件命名规范与标识符解析
  - 7.1.1. 命名格式
  - 7.1.2. 输出文件列表与变量含义
  - 7.1.3. .cfg.para 文件中的输出频率设置回顾
  - 7.1.4. 理解不同变量类型的单位
- 7.2. 文本格式输出文件 (.csv) 详解
- 7.3. 二进制输出文件 (.dat) 详解

## 8. SHUD 模型构建与前处理 (AutoSHUD 实践)

- 8.1. SHUD 模型构建与前处理流程概述
  - 8.1.1. 流程步骤
  - 8.1.2. AutoSHUD 在流程中的作用
- 8.2. AutoSHUD 脚本结构与核心配置文件
  - 8.2.1. AutoSHUD 脚本结构
  - 8.2.2. 核心配置文件: .autoshoot.txt 文件详解
    - 8.2.2.1. 结构
    - 8.2.2.2. 关键参数详解
    - 8.2.2.3. 理解参数优先级
  - 8.2.3. AutoSHUD 工作流程与主要脚本功能 (实操指南)
    - 8.2.3.1. GetReady.R
    - 8.2.3.2. Step1\_RawDataProcessing.R
    - 8.2.3.3. Step2\_DataSubset.R
    - 8.2.3.4. Step3\_BuildModel.R

#### 8.2.4. 实操流程总结

### 9. SHUD 模型输入文件生成

- 9.1. SHUD 模型输入文件结构 (复习与整合)
- 9.2. 利用 AutoSHUD/rSHUD 构建模型输入文件 (实操指南)
  - 9.2.1. 理解 Step3 的作用 (或等效 rSHUD 函数的功能)
  - 9.2.2. 依赖关系
  - 9.2.3. 实操：运行 AutoSHUD Step3 脚本
  - 9.2.4. 检查生成的输入文件集
  - 9.2.5. 理解生成的输入文件内容
  - 9.2.6. 常见问题
  - 9.2.7. 利用生成的 GIS 文件进行验证

### 10. SHUD 模型模拟运行与初步结果分析

- 10.1. SHUD 模型运行准备
  - 10.1.1. 组织输入文件
  - 10.1.2. 实操：配置模型运行参数 (`.cfg para`)
  - 10.1.3. 实操：准备初始条件 (`.cfg ic`)
  - 10.1.4. 将文件组织到运行目录
- 10.2. SHUD 模型模拟运行 (实操指南)
  - 10.2.1. 理解不同运行环境的运行方式
  - 10.2.2. 详细步骤：在本地终端运行 SHUD
  - 10.2.3. 详细步骤：在 HPC 队列提交 SHUD 运行任务
  - 10.2.4. 常见问题
- 10.3. 模型预热 (*Spin-up*) (实操指南)
  - 10.3.1. 理解预热的目的
  - 10.3.2. 预热期的长度设置
  - 10.3.3. 预热方法
  - 10.3.4. 实操：进行预热模拟
  - 10.3.5. 实操：利用预热结果更新 `.cfg ic` 文件
- 10.4. SHUD 模型输出结果读取与初步分析 (实操指南)
  - 10.4.1. SHUD 模型输出文件组织
  - 10.4.2. 实操：使用 rSHUD 读取输出文件
  - 10.4.3. 理解读取后的数据结构
  - 10.4.4. 实操：初步结果检查与可视化

### 11. SHUD 模型参数率定

- 11.1. 参数率定理论与实践
  - 11.1.1. 参数率定的概念与必要性
  - 11.1.2. 参数的“代表性”与率定
  - 11.1.3. 率定的目标
  - 11.1.4. 率定的挑战
- 11.2. 常用的拟合优度指标 (复习与重点强调)
  - 11.2.1. 重要指标及其特点
  - 11.2.2. 指标选择的考量
- 11.3. 参数优化算法介绍 (以 CMA-ES 为例)
  - 11.3.1. 进化算法基本原理
  - 11.3.2. CMA-ES (*Covariance Matrix Adaptation - Evolution Strategy*) 核心思想
- 11.4. 参数率定配置 (实操指南)
  - 11.4.1. 率定文件组织
  - 11.4.2. 率定主程序 (`mine.R`) 关键配置
  - 11.4.3. 实操：配置率定算法参数文件 (`.cfg.cmaes.txt`)
  - 11.4.4. 实操：配置可率定参数范围文件 (`.calib.range.txt`)
  - 11.4.5. 准备观测数据
- 11.5. 提交与监控参数率定任务 (实操指南)

- 11.5.1. 理解率定任务的执行方式
- 11.5.2. 编写率定任务 Slurm 脚本 (`.sh` 文件)
- 11.5.3. 实操: 使用 `sbatch` 命令提交率定任务
- 11.5.4. 实操: 监控率定过程与日志解读
- 11.5.5. 常见问题
- 11.6. 参数率定结果解读
  - 11.6.1. 率定结果文件组织
  - 11.6.2. 实操: 读取率定结果文件
  - 11.6.3. 分析率定过程与收敛性
  - 11.6.4. 分析最终最优参数集
- 11.7. 基于最优参数重新运行模型
  - 11.7.1. 实操: 将最优参数写入模型输入文件
  - 11.7.2. 实操: 运行模型进行校准期和验证期模拟
  - 11.7.3. 常见问题

## 12. 综合案例实战: Arno River Basin 全流程建模

- 12.1. 流域概况与研究目标
  - 12.1.1. Arno River Basin 地理位置与水文气象特征
  - 12.1.2. 本案例建模与分析目标
- 12.2. 数据收集与预处理 (实操)
  - 12.2.1. 确定并获取所需数据源
  - 12.2.2. 利用 rSHUD/AutoSHUD 进行数据裁剪、投影转换和格式化
  - 12.2.3. 数据质量检查与初步分析
- 12.3. SHUD 模型构建 (实操)
  - 12.3.1. 配置 `.autoshoot.txt` 文件 (确认与微调)
  - 12.3.2. 运行 AutoSHUD Step3 脚本生成 SHUD 输入文件
  - 12.3.3. 检查生成的输入文件和模型网格
- 12.4. SHUD 模型模拟运行与预热 (实操)
  - 12.4.1. 配置 `.cfg.para` 和 `.cfg.ic` 文件
  - 12.4.2. 运行预热模拟并更新初始条件
  - 12.4.3. 运行正式模拟 (校准/验证期)
- 12.5. 参数率定 (实操)
  - 12.5.1. 准备观测数据并配置率定文件
  - 12.5.2. 提交并监控 HPC 上的率定任务
  - 12.5.3. 分析率定结果, 确定最优参数集
- 12.6. 结果可视化与分析 (实操)
  - 12.6.1. 基于最优参数进行校准期和验证期模拟
  - 12.6.2. 绘制模拟与观测径流对比图
  - 12.6.3. 计算并展示拟合优度指标
  - 12.6.4. 可视化关键水文变量的空间分布
  - 12.6.5. 分析水量平衡
- 12.7. 案例总结与讨论
  - 12.7.1. 模型在 Arno River Basin 的表现评估
  - 12.7.2. 建模过程中的挑战与经验
  - 12.7.3. 模型结果的意义与潜在应用

## 13. SHUD 模型应用案例回顾 (其他)

- 13.1. 理想化案例 (模型基础验证)
  - 13.1.1. V 形流域 (V-Catchment)
  - 13.1.2. Vauclin 水槽箱试验
- 13.2. 实际流域案例 (模型适用性)
  - 13.2.1. 美国加州 Cache 河流域
  - 13.2.2. 中国流域案例
- 13.3. 特定水文过程模拟

- 13.3.1. 洪水演进与预报
- 13.3.2. 城市内涝
- 13.3.3. 地下水动态与基流分割
- 13.3.4. 冰冻圈过程
- 13.4. 与外部系统或模型的集成
- 13.5. 理解模型选择需要服务于研究目的

## 14. SHUD 进阶讨论、源码与附录

- 14.1. SHUD 源码结构 (C/C++ 实现)
  - 14.1.1. 源码获取
  - 14.1.2. 主要模块
  - 14.1.3. 关键模块代码解析 (示例 - 概念性)
- 14. 并行实现
- 14.2. rSHUD R 包源码结构
  - 14.2.1. 源码获取
  - 14.2.2. 主要模块
  - 14.2.3. Rcpp 接口
- 14.3. 如何贡献代码与参与开发 (可选)
  - 14.3.1. 参与方式
  - 14.3.2. GitHub 协作流程 (通用流程)
- 14.4. SHUD 进阶讨论与展望
  - 14.4.1. 模型局限性与未来改进方向
  - 14.4.2. 高性能计算的进一步应用
  - 14.4.3. 与其他模型的耦合前景
  - 14.4.4. 数据与模型的协同发展
  - 14.4.5. 智能算法在水文模型中的应用
- 14.5. 附录
  - 14.5.1. 常见错误代码及排查指南
  - 14.5.2. 关键文件对照表
  - 14.5.3. 常用 R 函数速查表 (rSHUD 及数据处理)
  - 14.5.4. Slurm 命令速查表

# 前言

这份笔记是根据“数值方法在水文模型中的应用”训练营的录播内容整理而成，旨在系统化地呈现 SHUD 模拟系统的理论、方法和实际应用。通过这份笔记，您将学习如何构建、运行、率定和分析 SHUD 模型，并了解其在水文研究和应用中的价值。

“云气西行，云云然冬夏不辍；水泉东流，日夜不休；上不竭，下不满；小为大，重为轻；圜道也。”

——吕不韦《吕氏春秋》，公元前239年(?)

“上善若水，水善利万物而不争”

——老子《道德经》，公元前四世纪

科学技术是人工的演化。理论物理学家的理论验证需要等待宇宙现象的发生。科学家穷其一生也无法验证某些理论——因为自然现象的发生概率非常低。所以，人工实验就可以大幅度提高这种现象的发生概率。

——陈平，2019

## 1. SHUD 系统构成与基本原理回顾

本章将介绍 SHUD 模拟系统的整体构成、基本原理以及其在水文模型发展中的位置。理解这些基础概念是后续学习和应用 SHUD 模型的关键。

## 1.1. 水文模型的分类与发展简史

**水文模型 (Hydrological Model)** 是对真实水文系统（如流域）的简化数学表达，用于理解水文过程、预测水文现象（如洪水、干旱）以及评估人类活动和气候变化对水资源的影响。

水文模型可以根据其对流域内部过程描述的精细程度和空间处理方式，分为总集式模型和分布式模型；根据其理论基础，分为物理模型和概念模型。

### 1.1.1. 总集式模型 (*Lumped Models*)

- **概念:** 将整个流域视为一个单一的、均质的单元，不考虑流域内部水文过程和参数的空间变异性。输入（如降雨）和输出（如径流）都以流域平均值或总量的形式表示。
- **示例:** 经典的单位线模型 (Unit Hydrograph)、蓄满产流模型的部分简化形式 (如新安江模型中的一些简化模块，尽管新安江模型也包含分布式概念)。
- **核心假设:** 流域对降雨等输入的响应在空间上是均匀的。
- **优点:**
  - **结构简单:** 模型数学表达相对简单。
  - **参数较少:** 通常只有少量需要率定的参数。
  - **计算效率高:** 计算量小，运行速度快。
  - **数据需求低:** 对空间数据的精度和密度要求不高。
- **缺点:**
  - **无法描述空间异质性:** 不能反映流域内部不同位置水文过程的差异。
  - **物理意义模糊:** 参数通常缺乏明确的物理意义，依赖率定。
  - **适用性受限:** 不适用于需要精细化空间管理的任务，或研究空间变异性对水文过程影响的场景。
  - **可移植性差:** 率定得到的参数在不同流域或不同时期可能不适用。

### 1.1.2. 分布式模型 (*Distributed Models*)

- **概念:** 将流域划分为许多小的计算单元（如栅格单元、不规则三角网单元、子流域、水文响应单元 HRU 等），并对每个单元内的水文过程进行模拟，同时考虑单元间的物质和能量交换。
- **示例:** SWAT (Soil and Water Assessment Tool), VIC (Variable Infiltration Capacity), **SHUD (Simulator for Hydrologic Unstructured Domains)**, MIKE SHE, TopModel (半分布式)。
- **核心假设:** 考虑流域内部的空间异质性，每个单元都有其独立的物理参数和状态变量。
- **优点:**
  - **描述空间异质性:** 能够模拟流域内部水文过程的空间分布和动态变化。
  - **物理意义更强 (通常):** 许多分布式模型基于物理过程，参数具有一定的物理意义。
  - **适用性广:** 适用于精细化水资源管理、土地利用变化影响评估、非点源污染模拟、无资料流域模拟 (基于物理参数推求) 等。

- **支持多过程耦合:** 便于与地表过程、生态过程等其他模型耦合。
- **缺点:**
  - **结构复杂:** 模型包含大量计算单元和复杂的单元间交互逻辑。
  - **参数众多:** 需要为每个单元或每类单元定义参数, 参数获取和率定难度大。
  - **计算量大:** 通常需要较高的计算资源和较长的运行时间。
  - **数据需求高:** 对高精度的空间输入数据 (如 DEM, 土壤图, 土地利用图) 和气象驱动数据有较高要求。
  - **尺度效应:** 参数和过程在不同空间尺度上的有效性可能存在问题。

### 1.1.3. 物理模型 (*Physically-Based Models*)

- **概念:** 基于描述水文过程的物理定律 (如质量守恒方程、能量守恒方程、动量守恒方程、达西定律 (Darcy's Law)、圣维南方程 (Saint-Venant Equations)、曼宁公式 (Manning's Equation) 等) 构建的模型。模型参数通常具有明确的物理意义, 原则上可以通过实地测量或实验室分析获得。
- **示例:** SHUD, MIKE SHE, HydroGeoSphere, ParFlow。
- **核心假设:** 流域水文过程可以用已知的物理定律来描述。
- **优点:**
  - **理论基础扎实:** 参数具有明确的物理意义, 有助于理解水文过程的内在机制。
  - **可移植性强 (理论上):** 如果物理参数能够准确获取, 模型在不同流域或不同时期应具有较好的适用性。
  - **适用于无资料区域:** 在缺乏长期观测数据的区域, 可以通过测量或估算物理参数来进行模拟。
  - **支持情景分析:** 可以用于预测土地利用变化、气候变化等对水文过程的影响。
- **缺点:**
  - **数据获取困难:** 详细的、空间分布的物理参数 (如土壤水力特性、含水层参数) 获取成本高、难度大。
  - **数值求解复杂:** 控制方程通常是非线性偏微分方程, 需要复杂的数值求解方法 (如有限差分法、有限元法、有限体积法), 对计算资源要求高。
  - **模型简化不可避免:** 即使是物理模型, 也需要对真实世界的复杂性进行一定程度的简化和概化。
  - **尺度效应问题:** 实验室测量的点尺度参数如何应用于模型单元的块尺度, 存在不确定性。

### 1.1.4. 概念模型 (*Conceptual Models*)

- **概念:** 基于对流域水文过程概念性理解构建的模型, 通常将流域简化为一系列相互连接的储水库 (reservoirs) 或功能模块, 通过经验或半经验的数学关系描述水在这些储库间的转移和转化。
- **示例:** 新安江模型、Sacramento 模型 (SAC-SMA)、HBV 模型。
- **核心假设:** 流域对降雨的响应可以被一系列简化的产流、汇流和蓄泄过程所代表。
- **优点:**
  - **结构相对简单:** 数学表达通常比物理模型简单。
  - **参数较少:** 需要率定的参数数量相对较少。
  - **计算效率较高:** 计算量较小, 运行速度快。
  - **在有充足观测资料的流域表现良好:** 通过参数率定可以很好地拟合观测径流。

- 缺点:

- **参数缺乏物理意义:** 参数通常是集总的、经验性的, 难以直接测量, 严重依赖率定。
- **可移植性差:** 率定得到的参数在不同流域或气候条件下可能不适用。
- **难以描述内部机制:** 无法清晰揭示流域内部的物理过程和空间分布。
- **不适用于无资料区域:** 缺乏观测数据率定, 模型无法有效应用。

**SHUD 模型** 的定位是 **分布式、物理性水文模型**。它试图通过非结构化三角形网格上求解基于物理的控制方程, 来精细模拟流域的水文过程。

**图表建议 1.1:** 插入水文模型分类示意图 (展示总集式/分布式、物理/概念的四象限或层级关系, 并标注 SHUD 模型的位置)

## 1.2. SHUD 模拟系统构成: SHUD, rSHUD, AutoSHUD

SHUD 模拟系统由三个主要部分组成, 它们协同工作, 支持从数据准备到模型运行和结果分析的全流程:

### 1.2.1. SHUD (Simulator for Hydrologic Unstructured Domains)

- **定位:** 系统的核心计算引擎。

- **功能:**

- 采用 C/C++ 语言编写, 追求高性能计算。
- 基于**有限体积法 (Finite Volume Method, FVM)** 在非结构化三角形网格上求解水文过程的常微分方程组。
- 模拟包括冠层截留、积雪累积与消融、蒸散发、地表产流与汇流、土壤水运动 (入渗、非饱和流、饱和流)、地下水流动、河道水流以及各水文单元之间的耦合交换过程。
- 常微分方程组表示为:
$$\frac{d\mathbf{y}}{dt} = \mathbf{f}(t, \mathbf{y})$$
其中  $\mathbf{y}$  是系统的状态向量 (如各单元的水位、含水量),  $t$  是时间。

- **特点:**

- **物理性强:** 模型中的主要过程基于物理定律。
- **全耦合:** 地表水、土壤水、地下水、河道水等过程在每个时间步长内同时求解, 相互影响。
- **数值方法:** 使用 FVM 进行空间离散, 使用 SUNDIALS/CVODE 求解器进行时间积分, 保证数值解的稳定性和精度。
- **并行计算:** 支持 OpenMP 实现共享内存并行计算, 提高大流域模拟效率。

### 1.2.2. rSHUD

- **定位:** 一个基于 R 语言开发的地理信息系统 (GIS) 和水文分析工具箱, 作为 SHUD 模型的官方配套工具。

- **功能:**

- **数据前处理:** 流域边界和河网的提取与简化、DEM 处理、土壤和土地利用数据的处理与属性赋予、气象驱动数据的格式转换与空间插值、模型输入文件的生成。
- **数据后处理:** 模型输出结果的读取、分析、统计。



- **参数率定辅助:** 提供参数敏感性分析、参数优化算法接口 (如调用 CMA-ES) 的辅助函数。
- **可视化:** 绘制径流过程线、空间分布图、水量平衡图等。
- **全球开放数据接口:** 提供获取常用全球地理数据集的接口。
- **特点:** 利用 R 语言强大的数据处理、统计分析和可视化能力, 为 SHUD 用户提供便捷的数据处理和分析环境。通过 `Rcpp` 包, 部分计算密集型任务调用 C++ 代码以提高效率。

### 1.2.3. AutoSHUD

- **定位:** 一套基于 rSHUD 和其他 R 脚本构建的自动化建模流程。
- **功能:** 将 SHUD 建模的多个步骤 (数据准备、预处理、模型构建、初步运行、参数文件生成等) 封装成可配置的自动化脚本。用户通过修改一个核心配置文件 (`.autoshoot.txt`), 即可快速完成一个流域的 SHUD 模型搭建。
- **特点:** 旨在提高建模效率和可重复性, 降低用户进行 SHUD 建模的技术门槛。特别适用于对大量流域进行快速建模或进行情景分析。

图表建议 1.2: 插入 SHUD 模拟系统构成示意图 (展示 SHUD, rSHUD, AutoSHUD 三者及其在整个建模流程中的角色和数据流向, 箭头表示数据或控制流)

## 1.3. SHUD 模型的发展历程 (与 PIHM 的渊源及区别)

SHUD 模型是在 PIHM 模型的基础上发展而来的, 继承了其核心思想并进行了显著改进。

### 1.3.1. PIHM (Penn State Integrated Hydrologic Model)

- **起源:** 由 Christopher Duffy 教授团队在美国宾夕法尼亚州立大学开发, 其理论基础是 Duffy 于 1996 年提出的“两态积分平衡”(two-state integral-balance) 模型 [Duffy1996a]。
- **核心思想:** 将土壤水分运动简化为未饱和带和饱和带两个状态, 并使用基于有限体积法的积分平衡方程来描述水量交换。
- **关键版本与工具:**
  - **PIHM v1.0** ([Qu2004; Qu2007]): 由曲轶众博士开发, 是 PIHM 发展中的重要里程碑, 加入了蒸散发和河道计算。
  - **PIHM v2.0** ([Kumar2009]): 增强了陆面过程和土地利用对水文过程的影响。
  - **PIHMGis** ([Bhatt2012; Bhatt2014]): 由 Gopal Bhatt 开发, 是一个基于 QGIS 的图形化界面工具, 用于 PIHM 的数据准备、模型运行和结果分析, 极大地推动了 PIHM 的应用。
  - **耦合模块:** 基于 PIHM 发展了多个耦合模型, 如 Flux-PIHM (与 NOAH 陆面过程模型耦合 [Shi2014; Shi2015]), LE-PIHM (地貌演变模型 [Zhang2016]), RT-PIHM (化学反应运输模型 [Bao2017]) 等。
  - **MM-PIHM (Multi-Module PIHM):** 由石宇宁博士领导, 旨在整合各种 PIHM 相关模型。

### 1.3.2. SHUD 与 PIHM 的关键区别

虽然 SHUD 继承了 PIHM 的“两态积分平衡”和 FVM 思想, 但进行了全面的重新设计和代码重写, 两者在具体实现和功能上已不兼容。(参考 [04]--2.1 中的 01:33:44 及其后论述)

- **代码基础:** SHUD 采用全新的 C++ 面向对象编程, 旨在提高代码的模块化、可维护性和计算效率, 并避免 PIHM 旧代码中可能存在的内存泄漏等问题。
- **流域拓扑与河道处理:**
  - PIHM: 河道与两个三角形坡面相邻, 这种设计在处理复杂河网 (特别是平原区绵延曲折的河道) 时, 容易导致生成大量细小、不规则的三角形单元, 增加计算负担, 并可能引发数值不稳定 (如积水点 `sink` 问题)。用户通常需要大量手动修改河网以适应模型。
  - SHUD: 河道单元覆盖在三角形单元之上, 河道与坡面 (三角形单元) 的水量交换基于河道水位与地表水/地下水水头的水力梯度计算。这种**交叉拓扑关系 (intersecting topology)**显著简化了网格生成, 提高了计算效率和稳定性, 减少了对河网简化的依赖。
- **水文过程公式与算法:** SHUD 对部分水文过程 (如入渗、地下水补给、地表-河道交互) 的数学表达和数值算法进行了调整和优化, 以期获得更合理或高效的模拟。
- **数据结构与文件格式:** SHUD 采用了不同于 PIHM 的内部数据结构和算法, 并设计了更易读、更规范的输入输出文件格式, 统一了时间序列数据操作。
- **数值求解器与技术层面:**
  - SHUD 支持更新版本的 SUNDIALS/CVODE (如 CVODE 5.0+)。
  - 改进了 OpenMP 并行计算的实现。
  - 增加了模型输入数据和参数的自动检查机制。
  - 加入了模型调试选项, 方便监控计算过程中的异常值和内存操作。
  - 支持指定步长输出模型状态, 作为后续模型运行的初始条件。

**图表建议 1.3:** 插入 PIHM 和 SHUD 模型发展和分支图 (树状图或时间线, 清晰展示 PIHM 各主要版本、代表性耦合模型、PIHMGis, 以及 SHUD 如何从 PIHM 发展而来并形成独立分支, 可参考 [SHUD\_Book-master/CN/01-overview.Rmd] 中的 Figure 1)

## 1.4. SHUD 模型的核心特点与优势

- **基于物理过程:** 模型的核心方程基于公认的物理定律, 参数具有物理意义, 这使得模型不仅可以用于有资料流域的模拟, 也为无资料流域的参数推估和模拟提供了理论基础。
- **分布式与空间异质性表达:** 通过非结构化三角网格离散流域, 能够精细刻画地形、土壤、土地利用等空间异质性及其对水文过程的影响。每个网格单元都有其独立的邻居关系, 模型通过计算相邻单元间的水力梯度来确定水流方向和通量 ([04]--2.1 中的 01:53:09)。
- **全耦合模拟:** 模型在每个时间步内, 同时求解地表水、土壤水 (未饱和带与饱和带)、地下水和河道水之间的水量交换, 确保了各水文过程的相互作用和水量平衡的内部一致性 ([04]--2.1 中的 01:00:35, 01:08:50)。
- **非结构化网格的灵活性:**
  - **适应复杂地形:** 三角形网格能够很好地拟合不规则的流域边界和复杂的地形特征。
  - **局部加密:** 允许在重点关注区域 (如河道附近、城市区域) 使用更高密度的网格, 而在其他区域使用较粗的网格, 从而在保证精度的前提下优化计算效率 ([04]--2.1 中的 01:54:50, 02:07:56)。
- **可调的时空分辨率:**
  - **时间分辨率:** 模型内部时间步长由 CVODE 求解器根据数值稳定性自动调整。用户可以设置输出结果的时间频率, 从分钟级到日级甚至更长 ([04]--2.1 中的 02:07:07 建议模型运行时间步长小于 10 分钟)。

- **空间分辨率:** 网格单元的大小可以根据研究需求和数据可用性进行调整, 从几米到几公里不等 ([04]--2.1 中的 02:07:07 建议网格面积小于 10 平方公里, 但也可根据具体情况调整)。
- **高效稳定的数值求解:** 采用先进的常微分方程 (ODE) 求解器 SUNDIALS/CVODE [Hindmarsh2005], 该求解器由劳伦斯利弗莫尔国家实验室开发和维护, 能够有效处理水文模型中常见的大型、刚性方程组。
- **开源与社区协作:** SHUD 及其配套工具 rSHUD 和 AutoSHUD 均为开源项目, 源代码托管在 GitHub 上, 鼓励用户使用、修改、贡献代码和参与社区讨论, 共同推动模型发展。
  - 非商业用途通常是免费的。
  - 商业使用或涉及特定专利技术时, 需遵循相应的授权协议 ([04]--2.1 中的 02:49:59)。

## 1.5. SHUD 系统的应用领域与价值

SHUD 模型系统凭借其物理基础和分布式特性, 已在多个水文及相关领域展现出应用潜力:

- **水资源评价与管理:** 模拟和预测流域尺度及特定区域的水量平衡组分 (降水、蒸散发、径流、地下水补给与排泄、土壤水储量、积雪水当量、湖库蓄水量等) 的时空动态, 为水资源规划、调度和可持续利用提供科学依据。
- **洪水预报与风险评估:**
  - **山洪预警:** 模拟小流域在短历时强降雨作用下的快速产汇流过程和洪水演进, 为山洪灾害的临近预报和人员疏散提供支持 ([04]--2.1 中的 01:48:25 龙槽沟案例)。
  - **区域洪水模拟:** 模拟大中型流域的洪水过程, 包括洪峰流量、到达时间、淹没范围和深度, 服务于防洪减灾决策 ([04]--2.1 中的 01:52:58 休斯顿案例, 02:07:02 全国水情预报系统设想)。
  - **城市内涝模拟:** 结合高精度城市地形、下垫面不透水率和排水系统信息, 模拟城市暴雨内涝的形成、发展和消退过程。
- **生态水文过程研究:**
  - **植被-水分关系:** 模拟不同植被类型下的蒸散发过程、土壤湿度动态以及植被对水循环的影响。
  - **湿地水文:** 模拟湿地的水文连通性、水位波动和水交换过程。
- **环境水文与水质模拟 (通过耦合):** SHUD 作为水文驱动模块, 可以与污染物迁移转化模型、水化学模型等耦合, 模拟非点源污染物的产生与输移、营养盐循环、水体富营养化等过程。
- **地貌演变 (通过耦合):** 与泥沙输运和地貌演变模型耦合, 研究水流侵蚀、泥沙沉积以及长期水文作用下的地貌塑造过程 (如 LE-PIHM 案例)。
- **气候变化与人类活动影响评估:**
  - 在不同气候变化情景 (如未来降雨、温度变化) 或土地利用变化情景下, 模拟流域水文响应的变化, 评估其对水资源量、洪水风险等的影响。
  - 评估水利工程 (如水库修建与调度、灌溉引水、跨流域调水) 对下游水文情势的影响。
- **高寒区水文学:** 模拟高寒地区的积雪累积与消融过程、冻土的冻融循环及其对产汇流的影响 ([04]--2.1 中的 04:48 布哈河案例)。
- **无资料流域水文模拟:** 基于物理参数和全球/区域数据集, 对缺乏长期观测资料的流域进行水文模拟和水资源估算。

**图表建议 1.4:** 插入 SHUD 模型应用领域示例图 (可使用九宫格或概念图, 每个部分用一个小图标或简图示意一个应用领域, 如洪水淹没、地下水等值线、积雪覆盖、城市管网等)

# 1.6. 引用规范与知识产权说明

在使用 SHUD 模型系统进行科研工作并发表成果时，尊重知识产权并正确引用相关文献至关重要。

- 核心参考文献：
  - SHUD 模型主要文献：  
Shu, L., Ullrich, P. A., & Duffy, C. J. (2020). Simulator for Hydrologic Unstructured Domains (SHUD v1.0): numerical modeling of watershed hydrology with the finite volume method. *Geoscientific Model Development*, 13(6), 2743-2762. [ @SHU2020\_GMD ]
  - PIHM 基础理论与方法 (SHUD 继承):  
Qu, Y., & Duffy, C. J. (2007). A semidiscrete finite volume formulation for multiprocess watershed simulation. *Water Resources Research*, 43(8). [ @Qu2007 ]  
Duffy, C. J. (1996). A two-state integral-balance model for soil moisture and groundwater dynamics in complex terrain. *Water Resources Research*, 32(8), 2421-2434. [ @Duffy1996a ]
  - 其他相关 PIHM 家族文献 (根据研究内容选择性引用): 如涉及特定耦合模块或前处理工具，应引用相应的原始文献。
- 开源协议与使用权限：
  - SHUD 模型、rSHUD 工具包和 AutoSHUD 脚本通常遵循开源软件许可协议 (例如 GPL, MIT 等，具体请查阅各项目 GitHub 仓库中的 LICENSE 文件)。
  - 非商业用途: 开源协议通常允许用户免费地使用、复制、修改和分发软件，用于学术研究和非盈利性活动。
  - 商业用途: 如果计划将 SHUD 系统或其衍生产品用于商业目的，务必仔细阅读许可协议中关于商业使用的条款。某些情况下，可能需要联系开发者获取商业授权或遵守特定的商业许可条款 ([04]--2.1 中的 02:49:59, 02:59:59 提到了商业使用和专利问题)。
  - 专利技术: 训练营中提到，SHUD 系统的某些特定算法或自动化平台部署方案可能已申请专利 ([04]--2.1 中的 02:59:59, [06]--2.3 中 02:10:51)。如果您的应用涉及这些专利技术，即使源码是开源的，也可能需要获得专利授权。
  - 引用要求: 即使是开源软件，在使用并发表基于其成果时，也应按照学术惯例正确引用相关文献。
- 数据版权: 在使用全球或区域开放地理数据时，同样需要注意数据来源的版权和引用要求。多数开放数据集会提供其推荐的引用方式。

## 小结:

第一章节介绍了水文模型的基本分类和 SHUD 模拟系统的构成、发展历程、核心特点及应用价值。重点强调了 SHUD 作为一个分布式、物理性、全耦合水文模型的定位。同时，也明确了在使用 SHUD 系统时应遵循的引用规范和知识产权注意事项。

## 思考题:

1. 结合您自己的研究方向或感兴趣的水文问题，思考 SHUD 模型的哪些特点最能满足您的需求？为什么？
2. 在实际应用中，开源模型的“开放性”和商业软件的“成熟性与支持”各有何优劣？您如何权衡？
3. 理解 SHUD 与 PIHM 的主要区别，对于评估 SHUD 模型的适用性和预期其在特定问题上的表现有何帮助？

## 2. 系统环境与软件安装 (实操指南)

---

本章将详细介绍搭建和配置 SHUD 模拟系统运行环境所需的步骤，包括目标平台的选择、HPC 环境的基础知识、SHUD 核心依赖库的安装、SHUD 模型可执行文件的编译，以及 rSHUD 和 AutoSHUD 工具的安装与配置。本章内容包含大量实操步骤，请务必在您的计算环境中逐步尝试。

### 2.1. 目标平台选择 (Linux/Windows/Mac)

---

SHUD 模型及其配套工具可以在多种操作系统平台上运行，但不同平台的支持程度和推荐使用场景有所不同。

- **Linux:**
  - **推荐度:** 最高。
  - **原因:** Linux 系统提供了稳定、高效的编译和运行环境，特别适合科学计算和高性能计算。大多数 HPC 集群采用 Linux 操作系统。SHUD 模型的开发和测试也主要在 Linux 环境下进行。
  - **优势:** 依赖库 (如 SUNDIALS/CVODE, GDAL, NetCDF) 的安装和管理相对方便，系统资源利用率高，命令行工具丰富。
- **Mac OS X:**
  - **推荐度:** 较高。
  - **原因:** Mac OS X 基于 Unix，与 Linux 在命令行环境和部分系统库上具有相似性，编译和运行 SHUD 通常也比较顺利。
  - **优势:** 拥有良好的图形用户界面和开发工具。
  - **注意事项:** 部分系统库的安装可能需要使用 Homebrew 等包管理器。
- **Windows:**
  - **推荐度:** 中等。
  - **原因:**
    - **SHUD 模型核心:** 直接在 Windows 上编译和运行 SHUD 的 C/C++ 核心可能较为复杂，因为其依赖的 SUNDIALS/CVODE 库通常更易于在类 Unix 环境下编译。可以使用 WSL (Windows Subsystem for Linux) 来创建一个 Linux 环境在 Windows 上运行 SHUD，或者使用预编译的 Windows 版本 (如果提供)。
    - **rSHUD 和 AutoSHUD:** 由于是基于 R 语言的，可以在 Windows 上的 R 环境中正常运行。
  - **优势:** 用户界面友好，GIS 软件 (如 ArcGIS, QGIS) 支持良好。
  - **建议:** 如果主要在 Windows 环境工作，并需要运行 SHUD 核心模型，强烈建议安装 WSL 2。

**训练营环境:** 本次训练营提供的计算环境是基于 **Linux** 的 **HPC 集群**。

### 2.2. HPC 环境基础 (Slurm 简介, 远程连接工具 MobaXterm/Terminal 使用)

---

对于需要处理大规模数据、运行复杂模型或进行大量参数率定任务的情况，高性能计算 (High-Performance Computing, HPC) 集群是必不可少的。

#### 2.2.1. 远程连接 (SSH)

- **SSH (Secure Shell):** 一种网络协议，用于在不安全的网络上安全地执行远程命令和管理服务器。
- **连接工具:**
  - **Windows:**
    - **MobaXterm:** 功能强大的终端模拟器，集成了 SSH 客户端、X11 服务器、SFTP 文件传输等功能，界面友好，推荐使用。([05]--2.2 中 0:00:25 开始演示了 MobaXterm 的使用)
    - **PuTTY:** 轻量级的 SSH 和 Telnet 客户端。
  - **Linux/Mac OS X:**
    - 系统自带 **Terminal** (终端) 应用，内置 **ssh** 命令。
- **连接信息:** 通常需要服务器的 IP 地址 (或域名)、端口号 (Port)、您的用户名和密码。
- **实操：使用 SSH 连接到 HPC 服务器**
  1. 获取服务器的连接信息 (通常由管理员提供)。
  2. 打开您的 SSH 连接工具。
  3. 对于 **MobaXterm**:
    - 点击 "Session" -> "SSH"。
    - 在 "Remote host" 中输入服务器 IP 地址或域名。
    - 勾选 "Specify username" 并输入您的用户名。
    - 在 "Port" 中输入指定的端口号 (SHUD 训练营服务器端口通常为 32099)。
    - 点击 "OK"。
  4. 对于 **Terminal (Linux/Mac):**
    - 在命令行中输入以下命令 (替换尖括号中的内容):

```
ssh -p <port_number> <your_username>@<server_ip_or_hostname>
```

例如: `ssh -p 32099 shuser01@ghdc.ac.cn`

5. 首次连接时，可能会提示服务器的真实性未知，输入 **yes** 继续。
6. 输入您的密码 (输入时通常不显示字符)。
7. 成功登录后，您将看到服务器的命令行提示符。

图表建议 2.1: 插入 MobaXterm SSH 连接设置界面截图。

图表建议 2.2: 插入 Linux/Mac Terminal SSH 连接命令示例。

## 2.2.2. Slurm 任务管理系统

HPC 集群通常使用作业调度系统来管理和分配计算资源，Slurm (Simple Linux Utility for Resource Management) 是其中一种广泛使用的系统。

- **核心概念:**
  - **节点 (Node):** 集群中的一台物理计算机。
  - **核心 (Core / CPU):** 节点上的计算处理单元。



- **任务 (Task):** 一个独立的计算进程。
- **作业 (Job):** 用户提交给 Slurm 系统的一个或多个计算任务的集合。
- **队列/分区 (Queue / Partition):** Slurm 将计算节点分组管理，不同的队列可能有不同的资源限制 (如运行时长、可用核心数)。
- **作业脚本 (Job Script):** 一个包含 Slurm 指令和要执行命令的 Shell 脚本。
- **常用 Slurm 命令:** ([07]-3 中 00:35:16 开始涉及 Slurm 脚本和命令)
  - `sbatch <job_script.sh>`: 提交作业脚本到队列中等待执行。
  - `squeue [-u <username>]`: 查看当前队列中的作业状态 (您的作业或所有作业)。常见的状态有 `PD` (Pending, 等待资源), `R` (Running, 正在运行), `CG` (Completing, 即将完成), `CD` (Completed, 已完成), `F` (Failed, 失败)。
  - `scancel <job_id>`: 取消指定 ID 的作业。
  - `sinfo`: 查看集群节点和队列的状态。
  - `sacct -j <job_id>`: 查看已完成作业的详细信息。
- **实操: 查看当前任务队列**

1. 登录到 HPC 服务器后，在 Terminal 中输入：

```
squeue -u your_username # 将 your_username 替换为您的用户名
```

2. 这将列出您当前正在运行或等待运行的作业。

**图表建议 2.3:** 插入 Slurm 作业提交流程示意图 (用户编写脚本 -> sbatch -> 队列 -> 计算节点 -> 完成/失败)。

**图表建议 2.4:** 插入 `squeue` 命令输出示例及关键列解释 (JOBID, PARTITION, NAME, USER, ST, TIME, NODES, NODELIST)。

## 2.2.3. 文件管理

在远程服务器上进行文件和目录操作是日常工作的基础。

- **常用 Linux 命令:**
  - `pwd`: 显示当前工作目录的完整路径 (*Print Working Directory*)。
  - `ls [-lha]`: 列出当前目录下的文件和子目录 (*List*)。
    - `-l`: 以长格式显示详细信息。
    - `-h`: 以人类可读的格式显示文件大小 (如 K, M, G)。
    - `-a`: 显示所有文件，包括隐藏文件 (以 `.` 开头的文件)。
  - `cd <directory_path>`: 切换当前工作目录 (*Change Directory*)。
    - `cd ..`: 切换到上一级目录。
    - `cd ~` 或 `cd`: 切换到用户主目录。
  - `mkdir <directory_name>`: 创建新目录 (*Make Directory*)。
  - `cp <source_file> <destination_file_or_directory>`: 拷贝文件或目录 (*Copy*)。

- 拷贝目录时通常使用 `-r` (递归) 选项: `cp -r <source_directory> <destination_directory>`。
- `mv <source> <destination>`: 移动文件或目录, 也可用于重命名文件或目录 (Move)。
- `rm <file_name>`: 删除文件 (Remove)。此命令非常危险, 删除后通常无法恢复, 请谨慎使用!
  - 删除目录时使用 `-r` (递归) 选项: `rm -r <directory_name>`。
- `cat <file_name>`: 查看文本文件内容。
- `less <file_name>`: 分页查看文本文件内容 (按 `q` 退出)。
- `nano <file_name>` 或 `vim <file_name>`: 使用命令行文本编辑器编辑文件。
- 文件路径:
  - 绝对路径: 从根目录 (`/`) 开始的完整路径, 例如 `/home/shuser01/my_project`。
  - 相对路径: 相对于当前工作目录的路径, 例如 `../data` (上一级目录下的 `data` 文件夹)。
- 用户主目录、Scratch 空间、共享数据目录: ([05]--2.2 中 0:04:01 提到了文件目录组织)
  - 用户主目录 (`~` 或 `$HOME`): 通常用于存放个人配置文件、小型脚本等, 空间有限, 不宜存放大量计算数据或运行大规模计算任务。
  - Scratch 空间 (例如 `/scratch/<username>`): 通常是为用户提供的临时大容量存储空间, 读写速度较快, 推荐用于存放计算过程中的输入输出数据和运行大规模计算任务。Scratch 空间的数据可能不被备份, 且可能存在定期清理策略。
  - 共享数据目录 (例如 `/volume/data`): 通常由管理员维护, 存放公共数据集 (如全球 DEM, 土壤数据等) 或项目共享数据。
- 实操: HPC 文件操作练习
  1. 登录 HPC 后, 使用 `pwd` 查看您当前所在的目录 (通常是用户主目录)。
  2. 使用 `mkdir my_shud_workspace` 在主目录下创建一个新的工作目录。
  3. 使用 `cd my_shud_workspace` 进入该目录。
  4. 使用 `ls -lha` 查看当前空目录的内容。
  5. 尝试从老师提供的课程资料目录 (例如 `~/class2024/day2/`) 拷贝一个文件到您的工作目录。

## 2.3. 实操: 安装 SHUD 依赖库 (SUNDIALS/CVODE)

SHUD 模型依赖 SUNDIALS (SUite of Nonlinear and Differential/ALgebraic equation Solvers) 库中的 CVODE (C-language Variable-coefficient Ordinary Differential Equation solver) 求解器。

### 2.3.1. 理解依赖库的作用

CVODE 是一款先进的常微分方程 (ODE) 求解器, 能够高效稳定地求解 SHUD 模型中描述水文过程的复杂、刚性 (stiff) 的 ODE 系统。SHUD 将水文过程离散化后形成的方程组交由 CVODE 进行时间积分求解。

### 2.3.2. 安装方法

在训练营环境中, 通常会提供预编译好的库或者安装脚本。

- 方法 1: 使用提供的安装脚本 (推荐, 简化了编译配置过程)



## 1. 获取 SHUD 源码:

- 如果尚未获取, 请从指定位置 (例如课程共享目录或 GitHub) 将 SHUD 源码包复制或克隆到您的 HPC 工作目录 (例如, 用户主目录下的 `SHUD_source` 文件夹)。

```
# 假设 SHUD 源码在 /path/to/shared/SHUD_source.tar.gz
cd ~ # 进入用户主目录
mkdir SHUD_modelling
cd SHUD_modelling
cp /path/to/shared/SHUD_source.tar.gz .
tar -xvzf SHUD_source.tar.gz
cd SHUD_source # 假设解压后文件夹名为 SHUD_source
```

## 2. 执行 `configure.sh` 脚本: SHUD 源码中通常包含一个名为 `configure` 或 `configure.sh` (或类似名称, 如 `Script/installSundials.sh`) 的脚本, 用于自动下载、配置和安装 CVODE。

- 给予脚本执行权限:

```
chmod +x configure.sh
```

- 运行脚本:

```
./configure.sh
```

此脚本通常会将 SUNDIALS/CVODE 安装在用户主目录下的 `sundials/` 文件夹中。([06]--2.3 中 0:34:08 演示了脚本安装, [SHUD\_Book-master/CN/02-Install.Rmd] 也提供了脚本安装步骤)。

- 安装过程可能需要一段时间, 因为它需要从网络下载源码并进行编译。

### • 方法 2: 手动下载并编译 SUNDIALS/CVODE (如果脚本不可用或需要自定义安装)

- 请参考 [SHUD\_Book-master/CN/02-Install.Rmd] 中的详细手动安装步骤。关键步骤包括:

1. 从 SUNDIALS 官方网站下载指定版本的源码包 (例如 CVODE 5.x)。
2. 解压源码包。
3. 创建 `builddir` 目录并进入。
4. 使用 `ccmake ..` (或 `cmake ..` 配合参数) 配置编译选项。
  - 确保 `BUILD_CVODE = ON`。
  - 设置 `CMAKE_INSTALL_PREFIX` 到您的目标安装路径 (例如 `~/sundials`)。
  - 如果需要 OpenMP 支持, 启用相关选项。
5. 执行 `make` 和 `make install`。

## 2.3.3. 常见问题

- **缺少编译器或 CMake:** 错误信息会提示找不到 `gcc`, `g++` 或 `cmake`。需要联系管理员安装这些编译工具, 或者在用户环境下尝试安装 (如果权限允许)。
- **网络问题:** `configure.sh` 脚本下载 CVODE 源码失败。检查网络连接或尝试手动下载源码包。

- **权限问题:** 如果 `CMAKE_INSTALL_PREFIX` 指向系统目录 (如 `/usr/local`), `make install` 步骤可能因权限不足而失败。务必将其指向用户可写目录 (如 `~/sundials`)。
- **依赖库版本不匹配:** 如果系统中已存在其他版本的 SUNDIALS, 可能会产生冲突。确保 SHUD 编译时链接到正确安装的 CVODE 版本。

## 2.4. 实操：编译 SHUD 模型可执行文件

在成功安装 SUNDIALS/CVODE 之后, 就可以编译 SHUD 模型本身的源代码了。

### 2.4.1. 理解 `Makefile` 文件的作用

`Makefile` 是一个文本文件, 它定义了一系列规则来告诉 `make` 工具如何编译和链接 SHUD 的 C/C++ 源代码, 生成最终的可执行文件。

- **主要内容:**
  - **编译器指定:** 使用哪个 C++ 编译器 (如 `g++`)。
  - **编译选项:** 如优化级别 (`-O2`, `-O3`), 调试信息 (`-g`), C++ 标准 (`-std=c++17`) 等。
  - **头文件路径 (`-I`):** 告诉编译器在哪里查找所需的头文件 (如 SUNDIALS 的头文件)。
  - **库文件路径 (`-L`):** 告诉链接器在哪里查找所需的库文件 (如 SUNDIALS 的库文件)。
  - **链接库 (`-l`):** 指定需要链接的具体库文件 (如 `-lsundials_cvode`, `-lm`)。
  - **源文件列表:** 需要编译的 `.cpp` 文件。
  - **目标文件:** 最终生成的可执行文件名 (如 `shud` 或 `shud_omp`)。

### 2.4.2. 详细步骤

#### 1. 进入 SHUD 源码目录:

```
cd ~/SHUD_modelling/SHUD_source # 根据您的实际路径
```

#### 2. 编辑 `Makefile` 文件: 使用文本编辑器 (如 `nano Makefile` 或 `vim Makefile`) 打开。

- **关键修改:**
  - `SUNDIALS_DIR`: 这是最常需要修改的参数。确保它指向您之前安装 SUNDIALS/CVODE 的根目录。例如, 如果安装在 `~/sundials`, 则应设置为:

```
SUNDIALS_DIR = $(HOME)/sundials
```

注意 `$(HOME)` 代表用户主目录, 或者直接写绝对路径 `/home/your_username/sundials`。

- `OPENMP_FLAGS` 和 `OPENMP_LIBS`: 如果您希望编译支持 OpenMP 并行的版本 (`shud_omp`), 并且您的系统和 CVODE 都已支持 OpenMP, 请确保这些编译和链接标志是正确的。如果不需要并行或不确定, 可以暂时忽略。

- 保存并关闭 `Makefile`。

#### 3. 执行编译命令:

- 清理旧的编译产物 (推荐):

```
make clean
```

- 编译标准的 SHUD 可执行文件:

```
make shud
```

- 或者, 如果您的 `Makefile` 支持并配置了 OpenMP, 可以编译并行版本:

```
make shud_omp
```

### 2.4.3. 检查编译结果

- 编译过程会在终端显示许多信息, 留意是否有 `Error` 或 `Warning`。`Warning` 通常可以忽略, 但 `Error` 意味着编译失败。
- 如果编译成功, 将在当前目录 (SHUD 源码根目录) 下生成一个名为 `shud` (或 `shud_omp`) 的可执行文件。  
([06]--2.3 中 0:50:47 演示了编译成功后, `ls` 命令可以看到 `shud` 文件)
- 可以通过 `ls -l shud` 查看文件属性, 确认其具有可执行权限 (`x`)。

### 2.4.4. 常见问题

- **SUNDIALS\_DIR 路径错误:** 这是最常见的问题。`Makefile` 中找不到正确的 SUNDIALS 头文件或库文件, 会导致编译错误。务必仔细检查该路径是否与您实际安装 SUNDIALS 的路径完全一致。
- **编译器错误:**
  - 提示找不到编译器 (`g++: command not found`): 需要安装 C++ 编译器。
  - 代码语法错误: 如果 SHUD 源码本身存在问题或与您的编译器版本不兼容, 可能会出现 C++ 语法错误。
- **链接错误:**
  - 提示找不到 `-lsundials_cvode` 或其他库: 链接器无法找到所需的库文件。检查 `SUNDIALS_DIR` 下的 `lib` 目录是否存在相应的库文件, 并检查 `Makefile` 中的库文件路径和链接库名是否正确。
  - [06]--2.3 中 02:00:01 提到的 `sundials/sundials_lapack.h: No such file or directory` 错误, 就是典型的头文件找不到问题, 通常与 `SUNDIALS_DIR` 设置或 CVODE 安装不完整有关。
- **OpenMP 相关错误:** 如果编译 `shud_omp` 时出错, 可能是系统未安装 OpenMP 开发库, 或者 `Makefile` 中的 OpenMP 编译/链接标志不适用于您的系统。

## 2.5. 实操：安装 rSHUD R 包

rSHUD 是 SHUD 模型的 R 语言配套工具, 用于数据预处理、后处理和参数率定。

### 2.5.1. 详细步骤

#### 1. 确保 R 和 RStudio 已安装:

- 在 HPC 环境中, R 通常作为模块提供。您可能需要加载 R 模块:

```
module load R # 具体命令可能因 HPC 环境而异
```

- 您可以通过在远程服务器的 RStudio Server 界面进行操作，或者在本地安装 R 和 RStudio。

## 2. 安装 `devtools` 包: `devtools` 包用于从 GitHub 等源安装 R 包。

- 打开 R (或 RStudio 控制台), 输入:

```
if (!requireNamespace("devtools", quietly = TRUE)) {  
  install.packages("devtools")  
}  
library(devtools)
```

## 3. 安装 rSHUD 包:

```
devtools::install_github('SHUD-System/rSHUD')
```

这条命令会从 SHUD-System 的 GitHub 仓库下载 rSHUD 的最新源码并进行安装。安装过程中, 它会自动尝试安装 rSHUD 所依赖的其他 R 包。

## 4. 安装特定依赖包 (如果需要):

- rSHUD 依赖的一些包 (如 `RTriangle`) 可能不在 CRAN (官方 R 包仓库) 上, 或者需要特定版本, 需要单独从 GitHub 安装。
- 根据 `install_github('SHUD-System/rSHUD')` 执行过程中的错误或警告提示, 安装缺失的依赖。例如, 安装 `RTriangle`:

```
devtools::install_github('shulele/RTriangle/pkg')
```

- 关键依赖:** `sp`, `raster`, `rgdal`, `rgeos`, `ncdf4`, `xts`, `ggplot2`, `hydroGOF`, `Rcpp` 等。确保这些包都已成功安装。([06]--2.3 中 02:01:52 列出了一些依赖)

## 2.5.2. 检查安装

- 在 R 控制台中, 尝试加载 rSHUD 包:

```
library(rSHUD)
```

- 如果没有错误提示, 表示 rSHUD 包已成功安装并加载。

## 2.5.3. 常见问题

- `devtools` 安装失败:** 可能是网络问题或 R 版本过低。
- GitHub 下载失败:** 网络连接问题或 GitHub 仓库地址错误。
- 依赖包安装失败:**
  - 系统库缺失:** 某些 R 包 (如 `rgdal`, `ncdf4`, `sf`) 依赖于系统级别的地理空间库 (如 GDAL, PROJ, NetCDF, GEOS, UDUNITS2)。如果这些系统库未安装, 对应的 R 包会安装失败。错误信息通常会指明缺少的系统库。此时需要联系 HPC 管理员安装这些系统库, 或者在您有权限的系统上自行安装。

- **编译器问题:** 部分依赖包可能需要从源码编译, 如果系统中 C/C++ 编译器配置不当或版本不兼容, 可能导致安装失败。
- **R 包版本冲突:** 较少见, 但可能发生。

## 2.6. AutoSHUD 脚本环境配置与依赖

AutoSHUD 本身是一套 R 脚本, 其运行环境主要依赖于正确安装的 R 和 rSHUD 包。

### 2.6.1. 脚本结构

- **获取 AutoSHUD 脚本:** 从课程资料或 GitHub 仓库获取 AutoSHUD 的脚本文件夹。
- **主要文件与目录:**
  - `GetReady.R`: 初始化脚本。
  - `Step1_RawDataProcessing.R`, `Step2_DataSubset.R`, `Step3_BuildModel.R`: 主要建模步骤脚本。
  - `Rfunction/`: 包含辅助函数。
  - `SubScript/`: 包含特定数据处理的子脚本。
  - `.autoshoot.txt`: 核心配置文件 (在后续章节详细介绍)。

### 2.6.2. 依赖

- **R 环境和 rSHUD 包:** 必须已正确安装。
- **GDAL 命令行工具 (可选但推荐):** AutoSHUD 的某些 GIS 操作 (如投影转换、栅格裁剪) 可能会调用系统中的 GDAL 命令行工具 (如 `gdalwarp`, `gdal_translate`) 以提高效率。确保这些工具已安装并且在系统的 `PATH` 环境变量中。在 HPC 环境中, GDAL 通常作为模块提供, 可能需要加载 GDAL 模块 (`module load gdal`)。

#### 小结:

第二章节详细介绍了 SHUD 模型运行和数据处理所需的基础软件环境的搭建过程, 包括 HPC 环境的初步使用、SHUD 核心依赖库 SUNDIALS/CVODE 的安装、SHUD 模型可执行文件的编译, 以及 rSHUD R 包和 AutoSHUD 脚本环境的配置。完成这些步骤是顺利进行后续 SHUD 建模与分析的前提。在实操过程中, 务必仔细阅读错误信息, 并根据提示解决依赖问题。

#### 思考题:

1. 在编译 SHUD 或安装 R 包时, 遇到“找不到头文件 (.h file not found)”或“找不到库文件 (cannot find -lxxx)”的错误, 通常是什么原因造成的? 如何排查这类问题?
2. 为什么在 HPC 环境中, 软件的安装和管理通常采用模块 (module) 的方式? 这样做有什么好处?
3. 如果您的研究区域有非常高精度的 DEM 数据 (例如 1 米分辨率), 直接用于 SHUD 建模是否合适? 可能会遇到哪些问题?

## 3. SHUD 模型所需原始数据

在构建 SHUD 模型之前, 首要任务是收集和准备描述研究区域特征的各种原始数据。这些数据是模型输入的基础, 其质量和适用性直接影响模型的模拟精度和可靠性。

### 3.1. 核心数据列表与要求

运行 SHUD 模型需要多种类型的空间数据和时间序列数据。下表总结了这些数据及其基本要求：

类型	数据名	必要性	要求 (格式、步长、数量)	说明与关键点
空间数据	流域边界 (WBD)	高	矢量 (如 Shapefile)	必须是连续、封闭、唯一的单部分或多部分多边形，明确定义模型计算的地理范围。拓扑正确性至关重要。
空间数据	河流网络 (RIV)	高	矢量 (如 Shapefile)	河流线要素，必须具有明确的方向性 (通常从上游指向下游)，且在模型处理范围内应无向下分叉 (即每个河段只能有一个直接的下游河段，除非是流域出口)。定义了地表水的主要汇流路径。
空间数据	高程 (DEM)	高	栅格 (如 GeoTIFF)	数字高程模型，是计算坡度、坡向、汇流累积量、提取河网等地形参数的基础。其分辨率和精度直接影响模型对地形的表达。单位通常为米 (m)。
空间数据	土地利用分类 (LC)	中	矢量/栅格 (如 Shapefile, GeoTIFF)	描述地表覆盖类型 (如森林、草地、农田、城市、水体等)。不同土地利用类型对应不同的水文参数 (如曼宁糙率、蒸散发特性、不透水率)。模型通常使用分类编码。
空间数据	土壤分类 (SOIL)	中	矢量/栅格 (如 Shapefile, GeoTIFF)	描述表层土壤的类型。不同土壤类型对应不同的土壤物理和水力学参数 (如孔隙度、饱和导水率、土壤水特征曲线参数)。模型通常使用分类编码。
空间数据	地质分类 (GEOL)	中	矢量/栅格 (如 Shapefile, GeoTIFF)	描述深层地质或含水层介质的类型。不同地质类型对应不同的饱和带水力学参数 (如饱和导水率、给水度)。模型通常使用分类编码。
空间数据	气象站点位置 (FORC_Site)	中	矢量 (如 Shapefile) 或栅格 (格点中心坐标)	当使用站点或格点气象数据时，需要提供其空间位置信息，用于将气象驱动数据插值或分配到模型的计算单元。
参数	土地利用参数	中	表格 (\$n_{lc}\$ 组参数)	与土地利用分类编码对应的水力学参数 (如曼宁糙率 <code>ROUGH</code> ，根系深度 <code>RZD</code> ，不透水面积比 <code>IMPAF</code> ) 和能量平衡参数 (如反照率 <code>ALBEDO</code> ，最大叶面积指数 <code>LAIMAX</code> )。
参数	土壤参数	中	表格 (\$n_s\$ 组参数)	与土壤分类编码对应的土壤物理性质 (如沙粒、粉粒、粘粒百分比，有机质含量 <code>OM</code> ，容重 <code>BD</code> )，这些基础性质用于通过 <b>PTF (Pedotransfer Functions)</b> 推导水力学参数。



参数	地质层参数	中	表格 (\$n_{\{geol\}}\$ 组参数)	与地质分类编码对应的饱和带水力学参数 (如水平/垂直饱和导水率 <code>KSatH / KSatV</code> , 孔隙度 <code>ThetaS</code> )。
时间序列	气象驱动 ( <code>FORC_Data</code> )	高	小时 ~ 日 (如 <code>CSV</code> 或 <code>NetCDF</code> )	模型的外部强迫。包括: 降雨 ( <b>P</b> ), 气温 ( <b>T</b> ), 相对湿度 ( <b>RH</b> ), 风速 ( <b>WS</b> ), 太阳净辐射 ( <b>Rn_Net</b> ) (或短波/长波辐射分量), 气压 ( <b>Press</b> )。数据的时间分辨率和质量至关重要。
时间序列	叶面积指数 (LAI)	中	小时 ~ 月 (如 <code>CSV</code> )	描述植被冠层密度的季节性变化, 影响冠层截留和植被蒸腾。每种土地利用类型通常对应一条 LAI 时间序列。
时间序列	融雪因子 ( <code>MF</code> )	低	小时 ~ 月 (如 <code>CSV</code> )	用于基于度日模型的积雪融化计算 (在高寒区或有显著积雪过程的流域模拟时需要)。
时间序列	观测数据 ( <code>OBS_Data</code> )	高	小时 ~ 月 (如 <code>CSV</code> )	用于模型参数校准和性能验证。最常用的是流域出口的径流 ( <b>Discharge</b> ) 数据。也可包括地下水水位、实际蒸散发、土壤湿度等其他观测数据。观测数据的质量、连续性和长度对率定效果至关重要。

- 必要性说明:
  - 高: 对于模型的基本运行和产出有意义的结果是必需的。
  - 中: 对于提高模型的物理真实性、扩展模型功能或特定过程的准确模拟是重要的。省略这些数据模型仍可运行, 但结果的可靠性和适用性会降低。
  - 低: 主要用于特定的高级模块或非常精细的研究目标。

图表建议 3.1: 插入一个更详细的 SHUD 模型所需主要输入数据类型分类图, 可以包含数据格式、典型单位和数据来源示例的列。

### 3.2. 数据实例

以下是一些数据实例的可视化效果, 帮助理解各种数据的形态。

图表建议 3.2: 插入流域边界 (WBD) 和河流网络 (RIV) 叠加在 DEM 上的图。

- 描述: 清晰展示研究区域的地理范围和主要水系。例如, 使用红色线表示流域边界, 蓝色线表示河流网络, 背景为彩色渲染的 DEM。

图表建议 3.3: 插入土地利用分类图 (LC)。

- 描述: 以不同颜色表示流域内的主要土地利用类型, 并附带图例说明每种颜色对应的土地利用类别 (如森林、农田、城市等)。

图表建议 3.4: 插入土壤分类图 (SOIL) 或关键土壤属性图。

- **描述:** 类似土地利用图，展示土壤类型的空间分布。或者，展示某一关键土壤属性 (如砂粒含量百分比 SANDPPT) 的空间栅格图。

**图表建议 3.5:** 插入气象驱动数据站点分布图或格点覆盖图 (FORC\_Site/Grid)。

- **描述:** 如果使用站点数据，在流域图上标记气象站点的点位。如果使用格点数据 (如 CMFD, NLDAS)，展示格网覆盖在流域上的情况。

**图表建议 3.6:** 插入气象驱动时间序列数据示例图 (FORC\_Data)。

- **描述:** 选取一个代表性站点或格点，绘制其降雨 (柱状图) 和气温 (线图) 随时间变化的过程线，展示数据的季节性和波动性。

**图表建议 3.7:** 插入叶面积指数 (LAI) 时间序列示例图。

- **描述:** 选取几种典型的土地利用类型 (如森林、农田)，绘制其年内 LAI 的变化曲线，展示植被的生长和衰落周期。

**图表建议 3.8:** 插入观测径流过程线图 (OBS\_Data)。

- **描述:** 绘制流域出口观测径流随时间变化的过程线，展示洪峰、枯水期等水文特征。

## 小结:

第三章详细列出了 SHUD 模型运行所需的各类原始数据，并强调了它们的重要性、基本要求和数据格式。理解这些数据需求是成功构建和运行 SHUD 模型的第一步。在实际操作中，数据的质量 (精度、完整性、一致性) 和适用性 (时空分辨率、覆盖范围) 对模型结果有决定性影响。

## 思考题:

1. 在您的研究区域，获取哪些类型的原始数据可能最为困难？您计划如何应对数据缺失或质量不高的问题？
2. DEM 的分辨率如何影响流域边界和河网的提取精度，进而影响 SHUD 模型的模拟结果？
3. 对于缺乏详细土壤调查资料的地区，如何利用全球土壤数据库 (如 HWSD 或 SoilGrids) 来估算 SHUD 模型所需的土壤参数？这种方法存在哪些不确定性？

# 4. 全球开放地理数据源介绍及获取

在许多研究中，尤其是针对数据稀缺地区或进行大尺度模拟时，直接获取所有高精度的本地实测数据可能非常困难。幸运的是，目前有许多全球或区域尺度的开放地理数据集可以作为 SHUD 模型输入数据的重要来源或补充。本章将介绍一些常用的开放数据源，并重点讲解如何通过全球水文数据云 (GHDC) 平台获取这些数据。

## 4.1. 常见全球数据集介绍

以下是一些在水文建模中常用的全球或区域性开放地理数据集：

- **数字高程模型 (DEM - Digital Elevation Model):**
  - **SRTM (Shuttle Radar Topography Mission):**
    - **分辨率:** 全球大部分地区约 90 米，美国地区约 30 米。
    - **特点:** 应用广泛，数据质量相对较好，但可能存在数据空洞 (voids)，尤其是在高山和陡峭地区。



- **ASTER GDEM (Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model):**
  - **分辨率:** 全球约 30 米。
  - **特点:** 分辨率较高, 但数据噪声和伪影相对 SRTM 可能更多。
- **Merit DEM:**
  - **分辨率:** 全球约 90 米。
  - **特点:** 通过移除 SRTM 和 AW3D30 (ALOS World 3D - 30m) 中的树高、建筑物等误差, 并填补数据空洞, 提供了更高精度的全球高程数据。
- **WorldDEM™ (商业数据):** 提供全球 12 米分辨率的高精度 DEM, 但需付费。
- **土壤数据:**
  - **HWSD (Harmonized World Soil Database):**
    - **分辨率:** 全球约 1 公里。
    - **特点:** 整合了多个区域性土壤数据库, 提供土壤类型、质地 (沙粉粘含量)、有机碳、容重等属性。在中国区域, 其数据源部分来自于南京土壤所的 1:100 万土壤图。
  - **ISRIC SoilGrids:**
    - **分辨率:** 全球 250 米或 1 公里。
    - **特点:** 利用机器学习和全球土壤观测数据生成的栅格数据集, 提供多种土壤物理和化学属性 (如沙粉粘含量、有机碳、容重、pH、阳离子交换量、田间持水量、凋萎点等) 在不同深度的预测值。
  - **SSURGO/STATSGO (美国):** 由美国农业部 (USDA) 提供的美国本土高精度土壤调查数据。
- **土地利用/土地覆盖 (LULC - Land Use/Land Cover):**
  - **MODIS Land Cover Type Product (MCD12Q1):**
    - **分辨率:** 全球约 500 米。
    - **特点:** 提供多种分类系统 (如 IGBP, UMD, LAI/fPAR) 的年度土地覆盖图。
  - **USGS Global Land Cover Characterization (GLCC):**
    - **分辨率:** 全球约 1 公里。
    - **特点:** 基于 AVHRR 遥感数据, 提供多种分类系统。SHUD 教程中常提及的 USGS GLC 数据通常指此数据集或其衍生产品。
  - **GlobeLand30:**
    - **分辨率:** 全球 30 米。
    - **特点:** 由中国研制, 提供较高分辨率的全球土地覆盖数据。
  - **NLCD (National Land Cover Database - 美国):** 美国本土高精度土地覆盖数据。
- **河流网络与流域边界:**
  - **HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales):**
    - **特点:** 基于 SRTM DEM 生成的全球水文地理数据集, 包含河网、流域边界、汇流方向、汇流累积量等。有多种分辨率版本。

- **Merit Hydro (Merit Hydrography):**
  - **特点:** 基于 Merit DEM 和其他数据源, 提供了更高精度的全球河网、流域边界和河道宽度等水文地理数据。林佩蓉老师是该数据集的主要贡献者之一 [ @Yamazaki2019; @Lin2019 ]。( [09]--5 中 0:23:09 开始演示了 Merit Hydro 数据)
- **气象驱动数据 (再分析资料或卫星产品):**
  - **GLDAS (Global Land Data Assimilation System):**
    - **分辨率:** 全球 0.25 度或 1 度。
    - **时间分辨率:** 通常为 3 小时或月。
    - **特点:** 整合了多种卫星观测和地面观测数据, 通过陆面模型进行数据同化, 提供降水、气温、湿度、风速、辐射等变量。
  - **FLDAS (Famine Early Warning Systems Network Land Data Assimilation System):**
    - **分辨率:** 主要覆盖非洲, 0.1 度。
    - **时间分辨率:** 日或月。
    - **特点:** 类似 GLDAS, 但针对非洲地区优化。
  - **NLDAS (North American Land Data Assimilation System):**
    - **分辨率:** 美国本土 0.125 度。
    - **时间分辨率:** 小时。
    - **特点:** 针对北美地区的高分辨率陆面数据同化系统。
  - **ERA5 (ECMWF Reanalysis v5):**
    - **分辨率:** 全球约 30 公里 (0.25 度)。
    - **时间分辨率:** 小时。
    - **特点:** 欧洲中期天气预报中心 (ECMWF) 的最新一代全球气候再分析资料, 数据质量较高。
  - **CMFD (China Meteorological Forcing Dataset):**
    - **分辨率:** 中国区域 0.1 度。
    - **时间分辨率:** 3 小时。
    - **特点:** 融合了地面观测、卫星遥感和再分析资料, 专门针对中国区域的气象驱动数据集。( [04]--2.1 中的 02:13:24 讨论了 CMFD 数据的问题)

#### 选择数据源的考量:

- **研究区域:** 某些数据集只覆盖特定区域 (如 NLDAS, CMFD)。
- **空间分辨率:** 根据研究尺度和模型网格大小选择合适的分辨率。
- **时间分辨率与覆盖时段:** 确保气象数据的时间分辨率和覆盖时段满足模拟需求。
- **数据质量与不确定性:** 不同数据集的精度和可靠性不同, 了解其局限性。
- **数据格式与可获取性:** 考虑数据下载的便利性和后续处理的兼容性。
- **数据版权与引用要求:** 使用数据时务必遵守其版权规定并正确引用。

**图表建议 4.1:** 插入常用全球开放地理数据集汇总表 (包含数据集名称、主要变量、空间分辨率、时间分辨率/覆盖范围、主要特点和获取途径/官方链接)。

## 4.2. GHDC 平台的数据获取方法 (实操指南)

全球水文数据云 (Global Hydrologic Data Center, GHDC) 平台 ([ghdc.ac.cn](http://ghdc.ac.cn)) 为 SHUD 模型用户提供了一个便捷的数据获取和初步建模的入口。

### 4.2.1. 理解 GHDC 的功能

GHDC 平台集成了多种上述提到的全球基础地理数据 (如 ASTER GDEM, HWSO, USGS GLC) 和再分析气象驱动数据 (如 CMFD, NLDAS, GLDAS)。用户只需提供研究区域的流域边界文件, 平台即可自动完成以下任务:

- **数据裁剪:** 根据用户提供的边界裁剪全球数据集。
- **数据重投影 (如果需要):** 将不同来源的数据统一到合适的投影坐标系。
- **数据格式化:** 将数据转换为 SHUD 模型可直接使用的格式。
- **初步模型构建 (可选):** 根据用户指定的基本参数 (如最小网格数、含水层厚度), 自动生成一套初步的 SHUD 输入文件。

### 4.2.2. 平台访问

通过 Web 浏览器访问平台网址: `ghdc.ac.cn`。([05]--2.2 中 0:07:34 开始演示平台访问)

### 4.2.3. 数据获取流程概述

1. **上传流域边界:** 提供 `.zip` 格式的流域边界 `Shapefile`。
2. **配置数据需求:** 选择所需的数据集、时间范围和初步建模参数。
3. **提交任务:** 填写邮箱并提交。
4. **邮件确认:** 点击确认邮件中的链接激活任务。
5. **等待处理:** 后台服务器将处理您的请求。
6. **下载结果:** 收到包含下载链接的邮件, 下载处理好的数据包。

### 4.2.4. 详细步骤: 上传流域边界

1. **准备流域边界文件:**
  - 确保您的流域边界是一个有效的 `Shapefile` 文件。
  - 该 `Shapefile` 应包含 `.shp`, `.shx`, `.dbf`, `.prj` 等必需文件。
  - 将这些文件打包成一个 `.zip` 文件。
2. **上传文件:**
  - 在 GHDC 平台主页, 找到“上传流域边界”或类似功能的按钮/区域。
  - 点击上传, 选择您准备好的 `.zip` 文件。
  - 上传成功后, 平台通常会在地图上显示您上传的流域边界, 请务必检查其位置和形状是否正确。([05]--2.2 中 0:09:11 演示了上传过程和地图显示)

## 4.2.5. 详细步骤：配置数据需求

在上传边界文件后，您需要配置所需的数据和初步建模参数。([05]--2.2 中 0:14:13 开始演示参数配置界面)

### 1. 项目名称 (Project Name):

- 为您的项目指定一个唯一的英文名称 (不能包含中文、空格或特殊字符)。平台将使用此名称为生成的文件和文件夹命名。

### 2. 是否仅获取空间数据 (Only Spatial Data?):

- 如果选择 **Yes**，平台将只处理和提供空间数据 (DEM, 土壤, 土地利用等)，不进行 SHUD 模型文件构建。
- 如果选择 **No** (默认)，平台将在处理空间数据的基础上，进一步构建初步的 SHUD 模型输入文件。

### 3. 开始年份 (Start Year) 和结束年份 (End Year):

- 指定您需要的气象驱动数据的时间范围。

### 4. DEM 数据源 (DEM Source):

- 选择 DEM 数据来源，例如 **ASTER GDEM (30m)** 或 **Merit DEM (90m)**。

### 5. 最小网格单元数 (Number of Cells (Min)):

- 这是平台进行自动建模时，对生成的三角形网格数量的一个约束。

### 6. 含水层厚度 (Aquifer Depth (m)):

- 平台自动建模时使用的默认含水层厚度。

### 7. 气象驱动数据源 (Forcing Data Source):

- 选择所需的气象驱动数据，例如 **CMFD** (中国区域)，**NLDAS** (北美)，**GLDAS** (全球)。请注意不同数据源的空间覆盖范围和时间覆盖范围。

### 8. 邮箱地址 (Email):

- 输入您的有效邮箱地址，用于接收任务确认和结果下载链接。

## 4.2.6. 详细步骤：提交任务与下载数据

### 1. 提交: 完成所有配置后，点击“提交”或类似按钮。

### 2. 邮件确认:

- 平台会向您填写的邮箱发送一封确认邮件。
- 打开邮件，点击邮件中的确认链接，以激活您的数据处理任务。([05]--2.2 中 0:16:20 演示了确认邮件和激活链接)

### 3. 等待处理:

- 您的任务将在 GHDC 后台服务器上排队等待处理。处理时间取决于任务的复杂程度 (流域大小、数据时间长度、选择的数据集) 以及服务器当前的负载。
- 平台通常有任务队列机制，可能需要等待一段时间。([05]--2.2 中 0:17:34 提到了后台处理)

### 4. 下载结果:

- 任务处理完成后，您会收到第二封邮件，其中包含一个指向您数据的下载链接。

- 点击链接，或者使用命令行工具 (如 `wget`) 下载生成的 `.zip` 数据包。([05]--2.2 中 0:21:37 演示了 `wget` 下载命令)

```
wget -c -r <download_link_from_email> # -c 支持断点续传，-r 递归下载文件夹
```

## 4.2.7. 理解 GHDC 生成的数据结构

下载的数据包解压后，通常包含以下主要目录和文件：

- **ETV/ (Essential Terrestrial Variables):**
  - 存放处理后的基础地理空间数据，例如：
    - 裁剪后的 DEM (`DEM.tif`)。
    - 处理后的土壤数据 (`SOIL.tif`, `SOIL_ATT.csv`)。
    - 处理后的土地利用数据 (`LC.tif`, `LC_ATT.csv`)。
    - 根据 DEM 生成的河网 (`STM_DEM.shp`) 和流域边界 (`WBD_DEM.shp`)。
    - 气象格点中心位置文件 (`Meteorology.shp`)。
- **MODEL/Input/ :**
  - 存放 GHDC 自动生成的 SHUD 模型输入文件，以项目名开头，例如：
    - `[ProjectName].sp.mesh.dat`
    - `[ProjectName].sp.att.dat`
    - `[ProjectName].sp.riv.dat`
    - `[ProjectName].cfg.para.txt`
    - `[ProjectName].tsd.forc.txt`
    - 等其他 `.para.*` 和 `.cfg.ic` 文件。
- **GIS/ :**
  - 存放与模型域相关的、可以直接在 GIS 软件中打开的 Shapefile 文件，例如：
    - `domain.shp`: 模型三角形网格。
    - `river.shp`: 模型使用的河网。
    - `seg.shp`: 三角形与河网相交生成的河段线段。
- **TSD/ (或 Forcing/ ):**
  - 存放格式化后的气象驱动时间序列数据文件 (通常是多个 `.csv` 文件，每个文件对应一个气象格点或站点)。`.tsd.forc.txt` 文件中会列出这些 `.csv` 文件的相对路径。
- **fig/ (可选):**
  - 可能包含一些自动生成的预览图件，如 DEM 图、土地利用图等。

**图表建议 4.2:** 插入 GHDC 平台数据获取与建模流程图 (与图表建议 6.2 类似，但聚焦于 GHDC 的特定操作步骤和输出)。

**图表建议 4.3:** 插入 GHDC 生成的数据包典型目录结构图。

# 4.2.8. GHDC 平台的局限性

虽然 GHDC 平台非常便捷，但也存在一些局限性，用户在使用时需要了解：

- **数据源的局限性：**
  - 平台集成的数据源种类和版本是有限的。
  - 气象驱动数据的时间覆盖范围和空间分辨率可能无法满足所有研究需求。
- **自动建模的简化：**
  - 平台自动生成的模型配置 (如网格参数、河道参数、含水层参数) 是基于一些通用假设和默认值，可能不完全适合特定流域的精细化模拟。
  - 无法进行复杂的模型配置，如非均质含水层厚度、详细的河道断面参数、局部网格加密等。
- **数据质量的依赖：** 生成结果的质量高度依赖于所选全球数据集本身的质量。
- **需要后续处理：** GHDC 生成的数据和模型通常需要用户进行进一步的检查、修改和优化 (例如，使用 AutoSHUD 进行更精细的网格调整和参数匹配，或手动修改输入文件)。

尽管存在这些局限性，GHDC 仍然是一个非常有价值的工具，尤其适用于：

- 快速获取研究区域的基础数据集。
- 对 SHUD 模型进行初步的、快速的建模尝试。
- 教学和培训目的。
- 为后续更精细的建模工作提供一个良好的起点。

---

## 小结：

第四章介绍了在水文建模中常用的全球开放地理数据源，并详细阐述了如何通过 GHDC 平台获取这些数据并进行初步的模型构建。理解不同数据源的特点和 GHDC 平台的使用方法，可以帮助您高效地为 SHUD 模型准备输入数据。同时，也需要认识到 GHDC 平台的局限性，以便在后续工作中进行必要的补充和优化。

## 思考题：

1. 如果您的研究区域缺乏高质量的实测气象数据，您会选择哪个全球再分析气象数据集？选择的依据是什么？
2. GHDC 平台自动生成的 SHUD 模型输入文件中，哪些参数您认为最有可能需要根据您的具体研究区域进行调整？为什么？
3. 在使用 Merit Hydro 数据集提取河网时，如何根据流域的实际情况 (如河网密度、最小汇流面积阈值) 来选择合适的河网级别或进行处理？

---

# 5. SHUD 模型输入文件详解 (空间与水力学参数)

在通过 GHDC 平台或 AutoSHUD 工具初步准备好数据后，理解 SHUD 模型核心输入文件的具体结构和含义至关重要。本章将详细解析定义模型计算域几何特征和物理属性的空间数据文件 (.sp.\*) 和水力学参数文件 (.para.\*)。

## 5.1. SHUD 模型输入文件结构 (复习与整合)

在深入细节之前，我们再次回顾一下 SHUD 输入文件的总体组织：

- 文件分类与命名:

- 所有输入文件均以项目名 (例如 `my_project`) 开头。
- 通过文件标识符区分文件类型和内容, 如 `my_project.sp.mesh.dat`。
- 主要类别包括:
  - `.sp.*`: 空间数据 (网格、河网、属性)
  - `.para.*`: 水力学参数 (土壤、地质、土地利用)
  - `.cfg.*`: 模型配置 (运行参数、校准参数、初始条件)
  - `.tsd.*`: 时间序列数据 (气象驱动、LAI 等)

**图表建议 5.1:** 插入 SHUD 模型输入文件体系结构图 (展示各类输入文件及其相互关联, 例如 `.sp.att` 中的索引如何指向 `.para.soil` 中的具体参数行)

## 5.2. 空间数据文件 (`.sp.*`) 详解

这些文件定义了模型的几何结构和空间属性的离散化表达。它们通常由建模工具 (如 rSHUD 或 AutoSHUD) 自动生成。所有坐标均基于模型所采用的投影坐标系。

### 5.2.1. `.sp.mesh` 文件

此文件是 SHUD 模型的核心空间数据文件, 定义了非结构化三角形网格的拓扑关系和节点几何信息。  
([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.2.1 节有详细表格, [06]--2.3 中 01:49:58 开始解读此文件)

- **结构:** 包含两个主要的表格块。

#### 1. 表格块 1: 三角形单元信息 (Cell Information)

- 文件头:

```
[Number of Cells N_cell] [Number of Columns (typically 8)]
```

- 数据表 (每行代表一个三角形单元):

列名	含义	示例值	单位	备注
ID	单元 (三角形) 的唯一标识符	1	-	从 1 到 $N_{\{cell\}}$
Node1	构成该三角形的第一个节点的 ID	278	-	对应下方“节点信息”表中的节点 ID。节点顺序通常按逆时针或顺时针排列。
Node2	构成该三角形的第二个节点的 ID	340	-	
Node3	构成该三角形的第三个节点的 ID	342	-	
Nabr1	与 Node1-Node2 边共享的邻居单元的 ID	86	-	如果为边界边，则为负值 (如 -1, -2, -3, -4 分别代表不同类型的边界)。
Nabr2	与 Node2-Node3 边共享的邻居单元的 ID	84	-	
Nabr3	与 Node3-Node1 边共享的邻居单元的 ID	93	-	
Zmax	该三角形单元中心的平均地表高程	77.68	m	通常由 DEM 插值得到。

## 2. 表格块 2: 节点信息 (Node Information)

- 文件头:

```
[Number of Nodes N_node] [Number of Columns (typically 5)]
```

- 数据表 (每行代表一个节点):

列名	含义	示例值	单位	备注
ID	节点的唯一标识符	1	-	从 1 到 $N_{\{node\}}$
X	节点的 X 坐标	-2148535	m	在模型使用的投影坐标系下。
Y	节点的 Y 坐标	2026400	m	在模型使用的投影坐标系下。
AqDepth	该节点处的含水层厚度 (从地表到不透水基岩)	10	m	SHUD 假设每个节点有其对应的含水层厚度。
Elevation	节点的 Z 坐标 (地表高程)	79.6	m	通常直接从 DEM 获取。

- 重要性:



- 定义了计算域的离散化方式，是所有水文过程计算的基础。
- 拓扑关系 (`Nabr1`, `Nabr2`, `Nabr3`) 决定了单元间的水平水流路径。
- 节点高程和含水层厚度定义了每个单元的垂向结构。

- 常见问题:

- **拓扑错误:** 邻居关系不正确 (例如，一个边指向了错误的邻居，或者本应是边界却指向了内部单元)，会导致水流计算错误。
- **恶劣三角形:** 存在角度过小 (锐角) 或过大 (钝角) 的三角形，可能导致数值计算不稳定。
- **节点信息不一致:** 节点坐标与高程数据源不匹配。
- **含水层厚度异常:** `AqDepth` 为零或负值，或与实际地质情况严重不符。

**图表建议 5.2:** 插入 `.sp.mesh` 文件中两个表格块的示例截图 (突出表头和前几行数据)。

**图表建议 5.3:** 插入一个简单的三角形网格示例图，标注单元 ID、节点 ID、邻居关系，并展示部分节点坐标和高程。

## 5.2.2. `.sp.att` 文件

此文件为每个三角形单元赋予一组属性索引，这些索引将单元与具体的水力学参数 (土壤、地质、土地利用) 和气象驱动数据关联起来。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.2.2 节有详细表格，[06]--2.3 中 02:05:05 开始解读此文件)

- **结构:** 一个表格块。

- **文件头:**

```
[Number of Cells N_cell] [Number of Columns (typically 7-9)]
```

- **数据表 (每行代表一个三角形单元):**

列名	含义	示例值	备注
ID	单元(三角形)的唯一标识符	1	与 <code>.sp.mesh</code> 中单元 ID 一一对应。
SOIL	土壤类型索引	48	指向 <code>.para.soil</code> 文件中定义的土壤类型行号 (从 1 开始)。
GEOL	地质类型索引	4	指向 <code>.para.geol</code> 文件中定义的地质类型行号。
LC	土地利用类型索引	3	指向 <code>.para.lc</code> 文件中定义的土地利用类型行号, 也关联到 <code>.tsd.lai</code> 等。
FORC	气象驱动站点/格点索引	1	指向 <code>.tsd.forc</code> 文件中列出的气象数据文件 (从 1 开始)。
MF	融雪因子时间序列索引 (可选)	0	指向 <code>.tsd.mf</code> 文件中定义的融雪因子时间序列列号 (如果使用)。0 表示不使用。
BC	边界条件索引 (可选)	0	用于指定特定单元的边界条件 (如定流量、定水头)。0 表示无特殊边界条件。
SS	源汇项索引 (可选)	0	用于指定特定单元的源汇项 (如抽水、灌溉)。0 表示无特殊源汇项。
LAKE	湖泊单元标识与索引 (可选)	0	如果单元属于湖泊, 则为湖泊 ID (从 1 开始)。0 或负值表示非湖泊单元。

- 重要性:
  - 实现了流域空间异质性的参数化表达, 是分布式模型的核心。
  - 通过索引将每个计算单元链接到其特定的物理属性和外部驱动。
- 常见问题:
  - 索引越界: `SOIL`, `GEOL`, `LC`, `FORC` 等索引值超出了对应参数文件或数据列表的实际范围。
  - 属性分配不合理: 空间叠加分析错误导致单元被赋予了不正确的土壤或土地利用类型。

图表建议 5.4: 插入 `.sp.att` 文件示例截图 (突出表头和前几行数据, 并用箭头示意各索引列指向的对应参数文件)。

图表建议 5.5: 插入流域土地利用类型空间分布图 (使用 `.sp.att` 中的 `LC` 索引和 `.para.lc` 中的土地利用名称, 对网格单元进行着色)。

### 5.2.3. `.sp.riv` 文件

此文件定义了流域内的河流网络, 包括河段的连接关系和基本几何属性。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.2.3 节有详细表格, [06]--2.3 中 02:23:40 开始解读此文件)

- 结构: 通常包含三个表格块 (但在 SHUD 实际应用中, 后两个表格块通常合并到 `.para.lc` 或通过 `.cfg.calib` 间接定义, 这里主要关注第一个表格块)。

1. 表格块 1: 河段基本信息 (River Segment Information)

■ 文件头:

```
[Number of River Segments N_riv] [Number of Columns (typically 5-6)]
```

■ 数据表 (每行代表一个河段):

列名	含义	示例值	单位	备注
ID	河段的唯一标识符	1	-	从 1 到 $N_{riv}$ 。
DOWN	该河段直接下游的河段 ID	2	-	如果为流域出口，则为负值 (如 -1, -2, -3)。正确的下游连接是构建河网拓扑的关键。
Type	河流参数类型索引	1	-	关联到河流参数表，定义河道形状、糙率等。在 SHUD 中，这些参数常通过 <code>.para.lc</code> (针对河道单元) 或 <code>.cfg.calib</code> 统一或分类设置。
Slope	河床底平均坡度	0.0018	m/m	影响河道水流速度。
Length	河段长度	271.66	m	河段的空间长度。
BC	边界条件索引 (可选)	0	-	如果此河段有特殊边界条件 (如闸门控制、定流量输入)。

- **表格块 2 & 3 (河道形状和水力学参数):** 在 SHUD 中，这些参数的详细定义和赋值逻辑更为灵活，通常在 `.para.lc` (将河道视为一种特殊的土地利用类型，赋予其曼宁糙率等) 或通过 `.cfg.calib` 文件进行全局或分类别的校准。`.sp.riv` 主要定义拓扑和基本几何。

● 重要性:

- 定义了流域地表水汇流的主要路径。
- 河段的坡度和长度影响水流的传播时间和速度。

● 常见问题:

- **下游 (DOWN) 连接错误:** 导致河网拓扑不正确，如形成孤立河段、循环流或错误的出口。
- **坡度异常:** 出现负坡度或非常平缓的坡度 (接近零)，可能由 DEM 数据质量或河网提取算法引起，影响水流计算。
- **河段长度过长或过短:** 过长的河段可能导致水流演算失真，过短的河段可能增加不必要的计算量。

图表建议 5.6: 插入 `.sp.riv` 文件第一个表格块的示例截图。  
图表建议 5.7: 插入流域河网图，可以根据河段级别 (Strahler 序) 或坡度用不同颜色或线宽表示。

5.2.4. `.sp.rivseg` 文件

此文件详细记录了每个河段是如何被三角形单元的边界切割成更小的线段的，即河段与三角形单元的精确空间相交关系。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.2.4 节有详细表格)

- 结构: 一个表格块。
  - 文件头:

```
[Number of River Segments (fine scale) N_seg] [Number of Columns (typically 4)]
```

- 数据表 (每行代表一个被三角形切割出的河道线段):

列名	含义	示例值	单位	备注
ID	线段的唯一标识符	1	-	从 1 到 $N_{\text{seg}}$ 。
iRiv	该线段所属的原始河段 ID	1	-	对应 <code>.sp.riv</code> 文件中的河段 ID。
iEle	该线段所在的三角形单元 ID	527	-	对应 <code>.sp.mesh</code> 文件中单元的 ID。这定义了河道与哪个单元交换水量。
Length	该线段的长度	63.98	m	一个原始河段的 <code>Length</code> 等于其所有 <code>rivseg</code> 线段的 <code>Length</code> 之和。

- 重要性:
  - 精确定义了河道与陆地计算单元之间的水力联系界面。
  - 模型计算河道与三角形单元之间的地表水和地下水交换时，会依据此文件确定交换发生的具体单元和交换长度。
- 常见问题: 通常由建模工具自动生成，较少出现人工编辑错误。但如果上游的 `.sp.mesh` 或 `.sp.riv` 文件存在拓扑问题，可能会导致此文件生成错误。

图表建议 5.8: 插入 `.sp.rivseg` 文件示例截图。

图表建议 5.9: 插入一个局部放大的图，清晰显示一条或几条河段是如何被三角形网格的边切割成多个 `rivseg` 线段的，并标注 `iRiv` 和 `iEle` 的对应关系。

### 5.3. 水力学参数文件 (`.para.*`) 详解

这些文件存储了与不同土壤类型、地质类型和土地利用类型相关的物理参数。这些参数直接输入到模型的控制方程中，决定了水的运动和储存特性。

#### 5.3.1. `.para.soil` 文件

定义了模型中不同土壤类型在未饱和带和表层的水力学特性。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.3.1 节, [06]--2.3 中 02:53:24 左右开始解读此文件)

- 结构: 一个表格。
  - 文件头: 通常指明参数名称，但模型实际读取时可能依赖列的顺序。
  - 数据表 (每行代表一种土壤类型):

列名 (示例)	含义	典型单位	备注
INDEX	土壤类型索引	-	从 1 开始, 与 <code>.sp.att</code> 中的 <code>SOIL</code> 列对应。
KsatV	饱和状态下垂直方向的水力传导度	m/day	也被称为渗透系数, 是影响入渗速率的关键参数。
ThetaS	饱和含水量 (等效于孔隙度 $\phi$ )	m <sup>3</sup> /m <sup>3</sup>	土壤完全饱和时, 水分所占的体积百分比。决定了土壤的最大储水能力。
ThetaR	残留含水量	m <sup>3</sup> /m <sup>3</sup>	当土壤基质势非常低 (非常干燥) 时, 仍能保持在土壤孔隙中的水分体积百分比。
InfD	有效入渗深度 (或类似影响入渗的参数)	m	SHUD 模型中可能包含一个与 Green-Ampt 模型中湿润锋深度概念相关的参数, 用于简化入渗计算。
Alpha	van Genuchten 方程参数 $\alpha$	1/m	van Genuchten 模型是描述土壤水特征曲线 (SWCC) 的常用经验公式。 $\alpha$ 与进气吸力有关。
Beta (或 n)	van Genuchten 方程参数 $\beta$ (或 $n$ )	-	影响 SWCC 的形状, $n = \beta / (\beta - 1)$ 。
AreaF_mac_v	垂直大孔隙面积比例	- (0 到 1)	土壤中垂直大孔隙所占的面积分数, 影响优先流。
KsatV_mac	大孔隙饱和垂直水力传导度	m/day	大孔隙的导水能力, 通常远大于土壤基质的导水能力。

- **重要性:** 这些参数直接控制降雨入渗、水分在未饱和土壤中的垂直运动、土壤的持水能力以及向饱和地下水的补给。
- **参数获取:**
  - **实测:** 实验室或野外测量 (如双环入渗仪、压力膜仪、土壤柱实验)。成本高, 通常只针对少量代表性样点。
  - **PTF (Pedotransfer Functions):** 基于易于获取的土壤基本物理性质 (如沙粒、粉粒、粘粒百分比, 有机质含量, 容重) 通过经验统计关系推求水力学参数。这是目前分布式水文建模中获取空间分布土壤参数的主要方法。常用的 PTF 包括 Rosetta [Schaap2001] 等。([07]--3 中 01:13:49 提到了 PTF 的应用)
- **常见问题:**
  - 参数值超出物理合理范围 (例如孔隙度大于 1 或小于 0)。
  - PTF 选择不当或输入到 PTF 的基础土壤属性数据质量不高, 导致推导出的水力学参数不准确。
  - 空间上不同土壤类型间的参数差异未能体现真实情况。

图表建议 5.10: 插入 `.para.soil` 文件示例截图 (突出表头和关键参数列)。

图表建议 5.11: 插入土壤水特征曲线 (SWCC) 和非饱和导水率曲线示意图, 并标注 van Genuchten 参数 ( $\alpha$ ,  $\beta$ / $n$ ,  $\theta_s$ ,  $\theta_r$ ) 对曲线形状的影响。

5.3.2. `.para.geol` 文件

定义了模型中不同地质单元 (通常指饱和含水层) 的水力学特性。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.3.2 节, [06]--2.3 中 02:53:24 左右开始解读此文件)

- 结构: 与 `.para.soil` 类似, 一个表格。
  - 数据表 (每行代表一种地质类型):

列名 (示例)	含义	典型单位	备注
<code>INDEX</code>	地质类型索引	-	从 1 开始, 与 <code>.sp.att</code> 中的 <code>GEOL</code> 列对应。
<code>KsatH</code>	饱和状态下水平方向的水力传导度	m/day	控制地下水在饱和含水层中的水平流动速率, 是地下水流动模型中的核心参数。
<code>KsatV</code>	饱和状态下垂直方向的水力传导度	m/day	影响饱和带与上下层 (如未饱和带、下伏弱透水层) 之间的垂向水量交换。
<code>ThetaS</code>	饱和含水量 (孔隙度 $\phi_s$ )	m <sup>3</sup> /m <sup>3</sup>	饱和含水层的总孔隙度。
<code>Sy</code>	给水度 (Specific Yield)	m <sup>3</sup> /m <sup>3</sup>	(可选, 有时用 $\phi_e = \theta_s - \theta_r$ 有效孔隙度代替) 描述含水层在重力作用下能够释放的水量比例。
<code>Ss</code>	储水率 (Specific Storage)	1/m	(可选, 主要用于承压含水层) 描述单位体积含水层在单位水头下降时因水和介质骨架压缩释放的水量。
<code>AreaF_mac_H</code>	水平大孔隙/裂隙面积比例	- (0 到 1)	描述饱和含水层中水平方向优先流通道的占比。
<code>KsatH_mac</code>	大孔隙/裂隙饱和和水平水力传导度	m/day	优先流通道的导水能力。
<code>Dmac</code>	大孔隙/裂隙影响深度 (或类似参数)	m	描述优先流通道在含水层中的分布特征。([04]--2.1 中 03:01:02 提到了饱和带大孔隙深度参数 <code>GEOL_DMAC</code> )

- 重要性: 这些参数决定了地下水的储存、运动和排泄特性, 对流域的基流过程、干旱期径流、地下水与地表水的交互有决定性影响。
- 参数获取:
  - 水文地质调查与抽水试验: 获取含水层参数最直接的方法, 但成本高, 覆盖范围有限。
  - 文献查阅与经验值: 根据区域地质图和类似地区的研究成果获取参数的典型取值范围。
  - 地球物理勘探: 如电阻率法、地震反射法等, 可辅助推断含水层结构和某些属性。

- 常见问题:
  - 缺乏详细的地质资料导致参数概化严重。
  - 水平和垂直导水率的各向异性未得到合理考虑。
  - 大孔隙/裂隙参数难以确定，通常作为率定参数。

图表建议 5.12: 插入 `.para.geol` 文件示例截图。

图表建议 5.13: 插入典型含水层结构示意图，标注水平/垂直导水率、孔隙度、给水度等参数。

### 5.3.3. `.para.lc` 文件

定义了不同土地利用/土地覆盖类型对水文过程 (主要是地表过程和蒸散发) 的影响参数。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.3.3 节, [06]--2.3 中 03:03:07 开始解读此文件)

- 结构: 一个表格。
  - 数据表 (每行代表一种土地利用类型):

列名 (示例)	含义	典型单位	备注
INDEX	土地利用类型索引	-	从 1 开始, 与 <code>.sp.att</code> 中的 <code>LC</code> 列对应。
LAIMAX	最大叶面积指数 (LAI)	m <sup>2</sup> /m <sup>2</sup>	植被生长最茂盛时的 LAI 值。
RMIN	最小冠层阻抗 (Minimum Canopy Resistance)	s/m	植被气孔完全张开时的最小阻力, 影响潜在蒸腾。
ALBEDO	地表反照率	- (0 到 1)	地表反射太阳短波辐射的比例, 影响净辐射和能量平衡。
VEGFRAC	最大植被覆盖度	- (0 到 1)	植被冠层投影面积占地面积的最大比例。
ROUGH	地表曼宁糙率系数	s/m <sup>(1/3)</sup>	描述地表对水流的阻力, 影响地表径流流速。
RZD (或 RootDepth)	根系深度	m	植被根系主要分布的深度, 影响植被从土壤中吸水的范围, 从而影响实际蒸腾。
IMPAF	不透水面积比例	- (0 到 1)	地表不透水部分 (如屋顶、道路) 所占的面积分数, 影响产流和入渗。
ISMAX	最大冠层截留量	m (或 mm)	植被冠层所能截留的最大降雨量。超过此量的降雨会成为穿透雨或茎流。
RoughLength	空气动力学粗糙度长度 (可选)	m	用于更复杂的蒸散发计算模型 (如 Penman-Monteith 的某些形式), 影响湍流交换。

- 重要性:
  - 土地利用参数直接影响降雨的再分配 (截留、穿透雨)、地表径流的产生和汇流速度、以及蒸散发过程 (能量分配、水分来源)。
  - 在人类活动影响显著的流域 (如城市化、农业开发), 这些参数尤为重要。
- 参数获取:



- **文献查阅:** 不同植被类型和地表覆盖的典型参数值已有大量研究。
- **遥感反演:** 如 LAI, 植被覆盖度, 反照率等可以从遥感数据 (如 MODIS 产品) 反演得到。
- **野外观测与实验:** 如测量不同下垫面的曼宁糙率。
- **常见问题:**
  - 土地利用分类系统与参数表不匹配。
  - 参数值未能反映特定区域的植被或地表真实状况。
  - 对于混合像元或变化剧烈的土地利用, 参数的代表性存在问题。

图表建议 5.14: 插入 `.para.lc` 文件示例截图。

图表建议 5.15: 插入不同土地利用类型 (如森林、草地、农田、城市) 的关键参数对比图 (例如, 用条形图对比它们的曼宁糙率、根系深度、不透水面积比和最大冠层截留量)。

### 小结:

第五章节详细解析了 SHUD 模型的核心输入文件: 空间数据文件 (`.sp.*`) 和水力学参数文件 (`.para.*`)。理解这些文件的结构、各列数据的物理含义以及参数的来源和不确定性, 对于正确构建 SHUD 模型、进行合理的参数设置、以及后续模型率定和结果分析至关重要。这些文件共同构成了对研究流域物理特征的数字化描述。

### 思考题:

1. 在 `.sp.mesh` 文件中, 三角形单元的邻居信息是如何确定的? 如果一个单元位于流域边界, 其邻居信息如何表示?
2. PTF (Pedotransfer Functions) 在推导土壤水力学参数 (`.para.soil`) 时扮演了什么角色? 使用 PTF 的主要优势和潜在的误差来源是什么?
3. 对于城市化区域, `.para.lc` 文件中的哪些参数对模拟产汇流过程最为敏感? 为什么?

## 6. SHUD 模型输入文件详解 (配置文件与时间序列)

在前一章介绍了定义模型空间结构和物理属性的文件后, 本章将聚焦于控制模型运行行为的配置文件 (`.cfg.*`) 和驱动模型运行的时间序列数据文件 (`.tsd.*`)。这些文件共同构成了模型运行的“指令集”和“外部驱动力”。

### 6.1. 模型配置文件 (`.cfg.*`) 详解

配置文件用于设定模型运行的控制参数、参数校准的范围以及模型的初始状态。

#### 6.1.1. `.cfg.para` 文件

此文件是 SHUD 模型最核心的运行控制文件, 包含了模拟时间、求解器参数、输出控制以及部分物理过程模块的开关等。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.3.1 节有详细表格, [06]--2.3 中 01:39:34 左右开始解读此文件)

- **结构:** 纯文本文件, 每行定义一个参数及其值。
- **关键参数与含义:**

		典		
--	--	---	--	--



参数名	含义	型 值/ 范 围	单位	备注
<code>VERBOSE</code>	是否输出详细的屏幕运行信息	0 或 1	-	1 表示输出更多调试信息，0 表示简洁输出。
<code>INIT_MODE</code>	初始条件模式	0 到 3	-	0 =Relief (基于地形高程猜测初始地下水位), 1 =Dry (干初始条件), 2 =Default guess (内部默认值), 3 =Warm start (从 <code>.cfg.ic</code> 文件读取)。 通常预热后使用 3。
<code>ASCII_OUTPUT</code>	是否输出 <code>.csv</code> 格式的结果文件	0 或 1	-	1 表示输出。
<code>BINARY_OUTPUT</code>	是否输出 <code>.dat</code> 格式的结果文件	0 或 1	-	1 表示输出 (推荐)。
<code>SPINUPDAY</code>	模型预热天数 (或数据点数, 取决于时间单位)	非 负 整 数	day/step	在计算拟合优度或进行长期分析时, 将忽略这段时间的模拟结果。
<code>SCR_INTV</code>	屏幕输出间隔	非 负 整 数	min	模型运行时在终端打印进度的频率 (分钟)。
<code>ABSTOL</code>	CVODE 求解器绝对容差	例 如 1e- 4	-	影响求解精度和计算时间。
<code>RELTOL</code>	CVODE 求解器相对容差	例 如 1e- 4	-	影响求解精度和计算时间。
<code>INIT_SOLVER_STEP</code>	CVODE 求解器初始时间步长	例 如 1	min	求解器开始积分时的初始时间步长。
<code>MAX_SOLVER_STEP</code>	CVODE 求解器最大允许时间步长	例 如 10 或 60	min	限制求解器的最大时间步长, 防止步长过大导致数值不稳定或错过关键过程。
<code>ET_STEP</code>	蒸散发计算步长	例 如 60	min	蒸散发模块的计算频率。
<code>START</code>	模型模拟开始时间 (相对于气象驱动数据起始点的天数或步)	非 负 整	day/step	定义模拟的起始点。

	数)	数		
END	模型模拟结束时间 (相对于气象驱动数据起始点的天数或步数)	非负整数	day/step	定义模拟的结束点。模拟总时长为 END - START。
FROZEN_SOIL	是否启用冰冻圈模块 (控制冻土参数化方案)	0 或 1	-	1 表示启用。([06]--2.3 中 02:42:39 提到此参数)
DEBUG_MODE	是否启用调试模式 (输出更多中间变量和检查信息)	0 或 1	-	通常仅在模型开发或详细排错时使用。
Unsat_Mode	未饱和带水流计算模式 (可选, 高级参数)	整数	-	控制未饱和带水流的计算方法或简化程度。
Surf_Mode	地表水流计算模式 (可选, 高级参数)	整数	-	控制地表径流演算方法 (如运动波 vs 扩散波)。([08]--4 中 01:49:29 提到相关内容)

- **输出频率设置:**
  - 文件后半部分包含一系列以 dt\_ 开头的参数, 用于控制各个输出变量的输出频率 (单位为分钟)。([07]--3 中 01:39:34 演示了修改输出频率)
  - 例如:
    - dt\_ye\_gw = 1440 表示每天输出一次地下水水位 (.eleygw)。
    - dt\_Qr\_down = 60 表示每小时输出一次河道向下游流量 (.rivqdown)。
    - 设置为 0 表示不输出该变量。
- **重要性:** .cfg.para 文件直接控制模型的运行行为和计算细节, 其设置对模拟结果的精度、稳定性和计算效率有显著影响。
- **常见问题:**
  - 模拟时间范围 (START, END) 超出气象驱动数据范围。
  - 求解器参数 (ABSTOL, RELTOL, MAX\_SOLVER\_STEP) 设置不当, 导致计算不稳定或耗时过长。
  - 输出频率设置过高, 导致产生大量不必要的输出文件, 占用磁盘空间。

**图表建议 6.1:** 插入 .cfg.para 文件关键参数示例截图 (突出显示 START, END, MAX\_SOLVER\_STEP, 以及一个 dt\_\* 输出频率参数)。

### 6.1.2. .cfg.calib 文件

此文件定义了 SHUD 模型中所有可能被校准的参数及其在模型内部的基准乘子或加数。在实际进行参数率定时, 优化算法会读取此文件作为参考, 并根据 .calib.range.txt 文件中指定的范围和开关来调整这些参数的实际应用值。([07]--3 中 01:18:16 开始详细解读此文件)

- **结构:** 纯文本文件, 每行定义一个可校准参数及其默认的调整系数。
- **内容 (每行):**

[Parameter\_Name\_in\_SHUD] [Multiplier\_or\_Adder\_Value]

- `[Parameter_Name_in_SHUD]`: SHUD 模型内部识别的参数名称, 例如 `GEOL_KSATH` (地质层水平饱和导水率), `SOIL_KINF` (土壤表层渗透系数), `LC_ROUGH` (土地利用曼宁糙率), `AQ_DEPTH+` (含水层厚度增量), `TS_PRCP` (降雨乘子)。
- `[Multiplier_or_Adder_Value]`:
  - 对于大多数物理参数 (如导水率、糙率、孔隙度等), 此值是一个乘性系数。模型内部的实际参数值 = `.para.*` 文件中的基础参数值 × 此乘子。默认值为 `1`, 表示不调整基础参数。
  - 对于某些参数 (如以 `+` 结尾的参数, 例如 `AQ_DEPTH+`, `TS_SFCTMP+`, 或一些初始条件调整参数), 此值是一个加性系数。模型内部的实际参数值 = `.para.*` 文件中的基础参数值 + 此加数。默认值为 `0`, 表示不调整基础参数。

- 作用:

- 提供一个集中的位置来管理所有可能参与校准的参数及其默认调整方式。
- 在参数率定过程中, 优化算法会读取此文件中的参数列表, 并修改这些乘子或加数的值, 从而间接调整模型中实际使用的物理参数。
- 当不进行参数率定时, 模型会直接使用此文件中定义的乘子/加数 (通常为 `1` 或 `0`) 来应用 `.para.*` 文件中的基础参数。

- 参数分类举例 (根据参数名后缀):

- 乘性调整: `GEOL_KSATH`, `SOIL_KINF`, `LC_ROUGH`, `TS_PRCP`, `ET_ETP` (针对时间序列或ET计算的调整系数)。
- 加性调整: `AQ_DEPTH+`, `TS_SFCTMP+`, `IC_GW+`, `IC_RIV+` (通常用于调整绝对值或初始条件)。

- 常见问题:

- 参数名称与 SHUD 模型内部定义不一致。
- 乘子/加数的默认值设置不当, 导致初始模拟结果偏差过大。

图表建议 6.2: 插入 `.cfg.calib` 文件部分示例截图 (包含乘性调整参数和加性调整参数的例子)。

### 6.1.3. `.cfg.ic` 文件

此文件定义了模型模拟开始时刻所有计算单元的状态变量初始值。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.3.3 节有详细表格, [07]--3 中 01:31:02 开始解读此文件)

- 结构: 通常包含两个表格块。

1. 表格块 1: 三角形单元初始条件 (Cell Initial Conditions)

- 文件头:

```
[Number of Cells N_cell] [Number of Columns (typically 6)]
```

- 数据表 (每行代表一个三角形单元):

列名	含义	单位	备注
ID	单元 (三角形) ID	-	与 <code>.sp.mesh</code> 中单元 ID 一一对应。
Canopy	冠层截留储量 (水深)	m	
Snow	雪水当量储量 (水深)	m	
Surface	地表积水储量 (水深)	m	
Unsat	未饱和带储水量 (水深)	m	等效水深，需要结合孔隙度换算为实际含水量。
GW	饱和带地下水水位 (高程)	m	相对于某个基准面的水位高程，或相对于不透水基岩的饱和含水层厚度。

2. 表格块 2: 河段初始条件 (River Initial Conditions)

■ 文件头:

```
[Number of River Segments N_riv] [Number of Columns (typically 2)]
```

■ 数据表 (每行代表一个河段):

列名	含义	单位	备注
ID	河段 ID	-	与 <code>.sp.riv</code> 中河段 ID 一一对应。
Stage	河道水位 (或水深)	m	相对于河床底的水位深度，或相对于某个基准面的水位高程。

- 作用: 为模型的动态模拟提供起始状态。不合理的初始条件会引入“预热期效应”，影响模拟初期的结果。
- 获取方式:
  - 默认生成: AutoSHUD 或 GHDC 可能会根据地形或默认值生成一个初始的 `.cfg.ic` 文件。
  - 模型预热 (Spin-up): 推荐方式。运行模型一段较长时间 (预热期)，将预热结束时刻的模型状态输出保存为新的 `.cfg.ic` 文件，作为后续正式模拟或率定的初始条件。
- 常见问题:
  - 初始值与实际情况严重不符，导致模型运行不稳定或预热时间过长。
  - 文件格式错误或单元/河段 ID 与 `.sp.mesh` / `.sp.riv` 不匹配。

图表建议 6.3: 插入 `.cfg.ic` 文件两个表格块的示例截图。

6.2. 时间序列数据文件 (`.tsd.*`) 详解

这些文件提供了驱动模型运行的随时间变化的外部输入，主要是气象数据，也包括植被的季节性参数等。

### 6.2.1. .tsd.forc 文件与关联的 .csv 文件

.tsd.forc 文件是一个列表文件，它指向实际存储气象驱动数据的一个或多个 .csv 文件。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.4.1 节有详细表格，[06]--2.3 中 02:03:36 开始解读此文件)

- **.tsd.forc 文件结构:**
  - **第一行:** [Number of Forcing Sites/Grids N\_forc] | [Start Date (YYYYMMDD)]
    - N\_forc: 气象数据文件 (通常是 .csv 文件) 的数量。
    - Start Date: 这些气象数据文件的时间序列起始日期。
  - **第二行:** [Path to Forcing Data Files]
    - 指定实际 .csv 气象数据文件所在的目录路径 (可以是绝对路径或相对于 .tsd.forc 文件的相对路径)。
  - **后续 \$N\_{forc}\$ 行:** 每行列出一个 .csv 气象数据文件的文件名。
    - 这些文件名将与 .sp.att 文件中 FORC 列的索引值 (从 1 开始) 对应。
- **关联的 .csv 气象驱动数据文件结构 (每个文件对应一个站点或格点):**
  - **文件头 (可选, 但 SHUD 通常不读取):** 可能包含列名。
  - **数据表 (每行代表一个时间步长):**

列名 (示例)	含义	典型单位	备注
Day (或 Time)	时间 (通常是模拟开始以来的天数)	day	
PRCP	降雨量	m/day (或 mm/day)	注意单位! SHUD 内部通常期望 m/day。
TEMP	平均气温	°C	
RH	平均相对湿度	- (0 到 1)	注意范围! 不是百分比。
Wind	平均风速	m/s (或 m/day)	模型计算蒸散发时需要。
Rn (或 SRAD, LRAD)	太阳净辐射 (或短波/长波分量)	W/m² (或 MJ/m²/day)	模型计算蒸散发和能量平衡时需要。单位和变量名需与模型内部期望一致。
PRESS	平均气压	Pa (或 hPa, kPa)	影响空气密度, 进而影响蒸散发。

- **重要性:** 气象驱动数据是模型运行的主要外部输入，其质量和时空分辨率直接决定了模拟结果的准确性。
- **常见问题:**
  - .tsd.forc 中列出的 .csv 文件路径或文件名错误，导致模型找不到数据。

- `.csv` 文件中的数据单位、时间步长或日期范围与模型设置不符。
- 数据缺失或存在异常值。
- 气象站点/格点空间分布不均或代表性不足。

图表建议 6.4: 插入 `.tsd.forc` 文件示例截图。

图表建议 6.5: 插入一个关联的 `.csv` 气象驱动数据文件示例截图。

## 6.2.2. `.tsd.lai` 文件

定义不同土地利用类型的叶面积指数 (LAI) 随时间的变化。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.4.2 节)

- 结构: 一个表格。
  - 文件头: `[Number of Time Steps N_time] | [Number of Land Cover Types N_lc] | [Start Date (YYYYMMDD)]`
  - 数据表:
    - 第一列: 时间 (通常是模拟开始以来的天数)。
    - 后续  $N_{lc}$  列: 每一列对应一种土地利用类型 (与 `.para.lc` 中的 `INDEX` 顺序对应) 的 LAI 值 ( $m^2/m^2$ )。
- 作用: 为模型提供动态的植被参数, 用于更精确地计算冠层截留和植被蒸腾。

## 6.2.3. `.tsd.rl` 文件

定义不同土地利用类型的空气动力学粗糙度长度 ( $z_0$ ) 随时间的变化 (可选, 高级)。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.4.3 节)

- 结构: 类似 `.tsd.lai` 文件, 但数值为粗糙度长度 (m)。
- 作用: 用于更复杂的蒸散发计算 (如 Penman-Monteith 方程的某些形式)。如果 `.para.lc` 中已包含固定的粗糙度长度, 则此文件可能不是必需的。

## 6.2.4. `.tsd.mf` 文件

定义融雪因子随时间的变化 (在高寒区或有显著积雪过程时使用)。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.4.4 节)

- 结构: 一个表格。
  - 文件头: `[Number of Time Steps N_time] | [Number of Melt Factor Series N_mf] | [Start Date (YYYYMMDD)]`
  - 数据表:
    - 第一列: 时间。
    - 后续  $N_{mf}$  列: 每一列对应一种融雪因子时间序列。 `.sp.att` 文件中的 `MF` 列将单元与这些序列关联。
- 作用: 用于基于度日模型的积雪融化计算, 融雪因子通常随季节变化。

## 6.2.5. `.tsd.obs` 文件

包含用于模型校准和验证的观测数据时间序列。([SHUD\_Book-master/CN/03-Inputfiles.Rmd] 3.4.5 节)

- **结构:** 一个表格。

- **文件头:** [Number of Time Steps N\_time] | [Number of Observation Series N\_obs] | [Start Date (YYYYMMDD)]

- **数据表:**

- 第一列: 时间。
    - 后续  $N_{\text{obs}}$  列: 每一列对应一个观测站点或观测变量的时间序列值 (如径流量  $\text{m}^3/\text{s}$  或  $\text{m}^3/\text{day}$ )。

- **作用:** 在参数率定过程中, 模型模拟结果将与此文件中的观测数据进行比较。

**图表建议 6.6:** 插入 `.tsd.lai` (或 `.tsd.mf`, `.tsd.obs`) 文件示例截图, 展示其时间序列表格结构。

## 小结:

第六章节详细介绍了 SHUD 模型的配置文件 (`.cfg.*`) 和时间序列数据文件 (`.tsd.*`)。 `.cfg para` 控制模型的核心运行逻辑, `.cfg.calib` 和 `.cfg.ic` 分别用于参数校准的基准和模型的初始状态。 `.tsd.*` 文件则提供了模型运行所需的外部时间变化驱动, 其中气象驱动数据 (`.tsd.forc` 及其关联的 `.csv`) 是最为关键的。正确理解和配置这些文件是成功运行和应用 SHUD 模型的基础。

## 思考题:

1. `.cfg para` 文件中的 `MAX_SOLVER_STEP` 参数对模型计算的稳定性和效率有何影响? 在不同类型的模拟 (如长期连续模拟 vs. 短期洪水事件模拟) 中应如何合理设置?
2. 为什么 SHUD 模型采用将参数的“基础值” (在 `.para.*` 文件中) 与“调整系数” (在 `.cfg.calib` 文件中) 分开的方式进行参数管理? 这种方式在参数校准和情景分析中有何优势?
3. 在准备 `.tsd.forc` 相关的 `.csv` 气象驱动数据时, 如何处理多个气象站点或格点数据, 以得到能够代表整个流域或各个子区域的气象输入? 泰森多边形法有哪些优缺点?

# 7. SHUD 模型输出文件详解

在 SHUD 模型成功运行后, 会生成一系列输出文件, 记录了模拟过程中各个水文变量的时空动态。理解这些输出文件的命名规范、格式和内容, 是进行模型结果分析、可视化和验证的关键步骤。

## 7.1. 输出文件命名规范与标识符解析

为了方便用户识别和管理大量的模拟结果文件, SHUD 模型对输出文件采用了规范的命名格式。

### 7.1.1. 命名格式

[Project Name].[Identifier].[Format]

或者, 更具体地, 标识符部分可以展开为:

[Project Name].[Model Unit][Variable Type][Variable Name].[Format]



- **[Project Name]**: 项目名称, 与输入文件和运行命令中指定的项目名一致 (例如 `ccw`, `test`, `arno_basin`)。
- **[Identifier]**: 文件内容标识符, 由以下三部分组合而成, 清晰地指明了文件包含的数据内容:
  - **[Model Unit]**: 文件中数据对应的模型计算单元类型。
    - `ele`: 代表三角形单元 (*Element*)。
    - `riv`: 代表河段 (*River*)。
    - `lak`: 代表湖泊 (*Lake*) (如果湖泊模块启用)。
  - **[Variable Type]**: 文件中数值所代表的变量类型。
    - `y`: 代表状态变量 (*State variable*)。通常表示在某个时刻, 模型单元内储存的水量或水位。其单位一般是长度 (例如, 水深 `m`)。
    - `v`: 代表单位通量 (*Specific flux*) 或速率。表示单位面积 (对于三角形单元) 或单位长度 (有时用于河道侧向交换) 上的水流量或物质通量。其单位通常是  $L^3 L^{-2} T^{-1}$  (即  $L T^{-1}$ , 如  $\text{\$/m/day}$ ) 或  $L^3 L^{-1} T^{-1}$  (即  $L^2 T^{-1}$ , 如  $\text{\$/m}^2/\text{day}$ )。例如, 降雨强度、蒸散发速率、入渗速率等。
    - `q`: 代表体积流量 (*Volumetric flux*)。表示在单位时间内流过某个界面或某个模型单元的总水量体积。其单位通常是  $L^3 T^{-1}$  (如  $\text{\$/m}^3/\text{day}$  或  $\text{\$/m}^3/\text{s}$ )。例如, 河道流量、三角形单元间的侧向地表径流量等。
  - **[Variable Name]**: 具体的变量名称缩写 (例如 `ic` 代表冠层截留, `snow` 代表积雪, `surf` 代表地表水, `gw` 代表地下水, `stage` 代表水位, `down` 代表向下游的流量等)。
- **[Format]**: 文件的存储格式。
  - `csv`: 文本格式 (*Comma Separated Values*)。易于人工阅读和使用常见的表格软件 (如 Excel, WPS 表格) 打开。
  - `dat`: 二进制格式。文件通常更小, 读写效率更高, 是 SHUD 模型推荐的默认输出格式, 特别适用于大规模数据和长时间序列的模拟结果。

## 7.1.2. 输出文件列表与变量含义

下表列出了 SHUD 模型在启用相应输出选项后可能生成的主要输出文件及其详细含义。请注意, 实际生成哪些文件取决于您在 `.cfg para` 文件中对各变量输出频率的设置。单位中  $L$  代表长度 (通常为米 `m`),  $T$  代表时间 (通常为天 `day` 或秒 `s`, 取决于您的模型时间单位和输出设置)。

标识符 ( [Identifier] )	模型单元 ( [Model Unit] )	变量类型 ( [Variable Type] )	变量名 ( [Variable Name] )	含义	典型单位 (\$L=m, \$T\$=day)
eleyic	ele	y	ic	三角形单元, 冠层截留储量 (水深)	m
eleysnow	ele	y	snow	三角形单元, 雪水当量储量 (水深)	m
eleysurf	ele	y	surf	三角形单元, 地表积水储量 (水深)	m
eleyunsat	ele	y	unsat	三角形单元, 未饱和带储水量 (等效水深)	m
eleygw	ele	y	gw	三角形单元, 饱和带地下水水位 (相对于基岩或某一基准面的高程)	m
elevetp	ele	v	etp	三角形单元, 潜在蒸散发速率	\$m/day\$ (或 \$m^3/(m^2 \cdot day)\$)
eleveta	ele	v	eta	三角形单元, 实际蒸散发速率	\$m/day\$
elevetic	ele	v	etic	三角形单元, 冠层截留蒸发速率	\$m/day\$
elevettr	ele	v	ettr	三角形单元, 植被蒸腾速率	\$m/day\$
elevetev	ele	v	etev	三角形单元, 土壤直接蒸发速率	\$m/day\$
elevprcp	ele	v	prcp	三角形单元, 降雨强度/速率	\$m/day\$
elevnetprcp	ele	v	netprcp	三角形单元, 净降雨速率 (降雨 - 冠层截留)	\$m/day\$
elevinfil	ele	v	infil	三角形单元, 入渗速率	\$m/day\$
elevexfil	ele	v	exfil	三角形单元, 出渗速率 (从地下水到地表)	\$m/day\$
elevrech	ele	v	rech	三角形单元, 地下水补给速率 (从未饱和带到饱和带)	\$m/day\$
elegsurf	ele	q	surf	三角形单元, 流向邻居单元的地表径流总量	\$m^3/day\$
elegsub	ele	q	sub	三角形单元, 流向邻居单元的地下水侧向流总量	\$m^3/day\$
rivystage	riv	y	stage	河段, 河道水位 (相对于河床底或某一基准面)	m
rivqup	riv	q	up	河段, 从上游河段流入该河段的总流量	\$m^3/day\$
rivqdown	riv	q	down	河段, 从该河段流向下游河段的总流量	\$m^3/day\$
rivqsurf	riv	q	surf	河段, 与相邻三角形单元之间的地表水交换流量	\$m^3/day\$
rivqsub	riv	q	sub	河段, 与相邻三角形单元之间的地下水交换流量	\$m^3/day\$
.lakystage etc.	lak	y / q / v	...	湖泊相关变量 (水位、入湖/出湖流量、湖面蒸发/降雨等)	...

- 流量方向约定:** 对于河段与单元的交流流量 ( `.rivqsurf`, `.rivqsub` ), 通常正值表示从单元流入河道, 负值表示从河道流向单元。但具体约定需参考模型文档或代码实现。

**图表建议 7.1:** 插入一个 SHUD 模型输出文件命名规范的详细解释图 (类似图表建议 4.1, 但更侧重于各组成部分的含义)。

**图表建议 7.2:** 插入一个更详细的 SHUD 核心输出变量分类表 (类似图表建议 4.2, 但包含更多变量并强调其物理意义和单位)。

### 7.1.3. `.cfg para` 文件中的输出频率设置回顾

正如第六章所述, 您可以在 `.cfg para` 文件中通过以 `dt_` 开头的参数来精确控制每个输出变量的生成频率。例如, `dt_Qr_down = 1440` 表示模型将每 1440 分钟 (即每天) 输出一组所有河段向下游的流量数据到 `[ProjectName].rivqdown.[Format]` 文件中。将频率设置为 0 则表示不输出该变量, 这对于节省磁盘空间和减少不必要的计算后处理非常有用, 尤其是在参数率定阶段, 可能只关心少数关键变量。

### 7.1.4. 理解不同变量类型的单位

在分析 SHUD 输出结果时, 准确理解各变量的单位至关重要。

- **状态变量 ( $y$ ):**
  - 其值通常表示的是等效水深或水位高程, 单位为长度 ( $m$ )。
  - 例如, `.eleysurf` (地表积水储量) 表示在三角形单元上平均积水深度为多少米。要计算该单元上的总积水体积, 需要将其乘以该单元的面积。
  - `.eleygw` (地下水水位) 表示的是相对于某个基准面 (如海平面或模型定义的最低点) 的地下水自由水面的高程, 或者在非承压含水层中, 也可以是相对于不透水基岩的饱和含水层厚度。
  - `.eleyunsat` (未饱和带储水量) 也是以等效水深表示, 即如果将未饱和带中的所有水分提取出来并均匀铺在单元表面, 其水深为多少。要换算成实际的体积含水量 ( $m^3/m^3$ ), 需要除以未饱和带的厚度。
- **单位通量 ( $v$ ):**
  - 其值表示单位面积上的流量速率, 单位是  $L T^{-1}$  (例如  $m/day$ )。
  - 例如, `.elevprcp` (降雨通量) 的值为  $0.01 m/day$  表示该单元接收到的平均降雨强度为每天 10 毫米。
  - 要计算该单元接收到的总降雨体积, 需要将其乘以该单元的面积和对应的时间步长。
- **体积流量 ( $q$ ):**
  - 其值表示单位时间内的总体积流量, 单位是  $L^3 T^{-1}$  (例如  $m^3/day$  或  $m^3/s$ )。
  - 例如, `.rivqdown` (河道向下游流量) 表示在每个输出时间步长内, 从该河段流向下流的总水量。
  - `.elegsurf` 表示从一个三角形单元的各条边流向其邻居单元的地表径流总和 (通常是净流出量)。

**单位转换:** 在进行数据对比或分析时, 经常需要进行单位转换。例如, 将  $m^3/s$  的观测径流转换为模型输出常用的  $m^3/day$ 。rSHUD 提供了一些辅助函数进行单位转换, 但理解原始单位和转换逻辑非常重要。

## 7.2. 文本格式输出文件 (`.csv`) 详解

当在 `.cfg para` 中设置 `ASCII_OUTPUT = 1` 时, 模型会生成 `.csv` 格式的输出文件。

- **结构:**

```
[N_columns_data], [StartTime_YYYYMMDD.HHMMSS]
[Time_step_1_value (e.g., days)], [Cell1_Value], [Cell2_Value], ..., [CellN_Value]
[Time_step_2_value (e.g., days)], [Cell1_Value], [Cell2_Value], ..., [CellN_Value]
...
[Time_step_m_value (e.g., days)], [Cell1_Value], [Cell2_Value], ..., [CellN_Value]
```

- **第一行 (文件头):**

- 第一个数字 `N_columns_data` 表示数据列的数量 (不包括第一列的时间列)。例如, 如果输出的是所有三角形单元的地表水储量, 则 `N_columns_data` 等于总的三角形单元数  $N_{\text{cell}}$ 。
- 第二个字符串 `StartTime_YYYYMMDD.HHMMSS` 表示模拟的开始日期和时间。

- **后续数据行 (每行代表一个输出时间步长):**

- **第一列:** 当前时间步长相对于模拟开始时间的的时间值。其单位由 `.cfg para` 中设置的输出时间步长单位决定 (例如, 如果以天为单位输出, 则这里是第几天)。
- **后续 `N_columns_data` 列:** 对应各个模型单元 (三角形单元或河段, 按其 `ID` 顺序排列) 在该时间步长的变量值。

- **优缺点:**

- **优点:** 文件内容可直接用文本编辑器或电子表格软件 (如 Excel, LibreOffice Calc) 打开和查看, 便于快速检查和初步分析。
- **缺点:** 对于大量的计算单元或长时间的模拟, `.csv` 文件会变得非常巨大, 占用大量磁盘空间, 并且读写效率较低。

**图表建议 7.3:** 插入一个 `.csv` 输出文件示例截图 (例如 `.rivqdown.csv`), 清晰展示文件头和前几行数据, 并用标注解释各部分的含义。

## 7.3. 二进制输出文件 (`.dat`) 详解

当在 `.cfg para` 中设置 `BINARY_OUTPUT = 1` (默认推荐) 时, 模型会生成 `.dat` 格式的二进制输出文件。

- **结构:**

- 文件内部以二进制形式连续存储数值, 没有文本分隔符。
- **文件头:**
  - 第一个数值 (通常为 `double`, 8字节): 数据列的数量 `N_columns_data` (与 `.csv` 文件中的含义相同)。
  - 第二个数值 (通常为 `double`, 8字节): 模拟开始时间, 通常以某种数值形式表示 (例如从某个纪元开始的总秒数或天数, 具体编码需查阅模型实现或 `rSHUD` 读取逻辑)。
- **后续数据块 (每个时间步长一个数据块):**
  - 每个数据块首先是该时间步长的时间值 (通常为 `double`)。
  - 然后是该时间步长下, 所有模型单元的变量值 (按 `ID` 顺序排列, 每个值为 `double`)。
- 所有数据紧密排列, 没有额外的格式化字符。

- **读取:**

- 不能直接用文本编辑器查看。

- 需要使用专门的程序或脚本读取。**rSHUD** 包中的 `readout()` 函数能够自动识别并正确解析 SHUD 生成的 `.dat` 文件。
- 其他编程语言 (如 Python, Matlab, C++) 也可以编写脚本来读取这种格式, 但需要知道确切的数据类型 (通常是 `double`) 和文件组织结构。

- 优缺点:

- 优点:

- **文件体积小:** 相比 `.csv` 文件, `.dat` 文件大大减小了存储空间占用。
    - **读写效率高:** 二进制读写速度远快于文本文件, 对于大规模数据处理和分析非常重要。

- 缺点:

- **不易直接查看:** 无法直接用文本编辑器打开检查内容。
    - **依赖特定读取工具:** 需要 **rSHUD** 或自定义脚本才能读取。

**图表建议 7.4:** 插入 `.dat` 文件结构的概念示意图 (用方框表示二进制数据块, 标注文件头信息和后续数据流的组织方式, 强调其与 `.csv` 的区别在于没有分隔符和文本编码)。

## 小结:

第七章详细介绍了 SHUD 模型输出文件的命名规则、文件结构以及主要变量的含义。理解输出文件的组织方式和内容是进行模型结果分析、可视化、参数校准和模型验证的基础。选择合适的输出格式 (`.csv` 或 `.dat`) 和输出频率对管理数据和提高分析效率非常重要。

## 思考题:

1. 在 SHUD 模型的 `.cfg para` 文件中, 如何根据您的研究目标 (例如, 关注洪峰 vs. 关注长期水量平衡) 来合理设置不同变量的输出频率 (`dt_*` 参数)?
2. `.eleygw` (地下水水位) 和 `.eleyunsat` (未饱和带储量) 都是以长度 `m` 为单位, 它们分别代表什么物理意义? 在进行水量平衡分析时, 如何将它们转换为实际的水量体积?
3. 如果您需要用 Python 而不是 R 来读取 SHUD 的 `.dat` 输出文件, 您认为需要了解哪些关于文件格式的关键信息才能正确解析数据?

# 8. SHUD 模型构建与前处理 (AutoSHUD 实践)

在前面的章节中, 我们已经了解了 SHUD 模型所需的基础数据、数据源以及核心输入文件的结构。本章将重点介绍如何利用 **AutoSHUD** 这一自动化工具集, 高效地完成从原始数据到 SHUD 模型输入文件的构建和前处理过程。AutoSHUD 旨在简化繁琐的数据处理步骤, 提高建模效率和可重复性。

## 8.1. SHUD 模型构建与前处理流程概述

从获取原始的地理空间数据和气象时间序列数据, 到最终生成 SHUD 模型可以直接运行的输入文件, 通常需要经历以下几个关键步骤:

### 8.1.1. 流程步骤

1. 数据准备与收集 (Data Acquisition):

- 收集研究流域的 DEM、河网、流域边界、土壤类型图、土地利用图等空间数据。
- 收集气象驱动数据 (降雨、气温、湿度、风速、辐射、气压等) 和其他必要的时间序列数据 (如观测径流、LAI 等)。
- 确保数据的格式、投影、时间范围和分辨率等满足基本要求。

## 2. 数据预处理 (Data Pre-processing):

- 空间数据处理:
  - 对 DEM 进行填洼、裁剪等操作。
  - 对河网数据进行拓扑检查、简化、方向修正等。
  - 对流域边界进行检查与修正。
  - 将所有空间数据统一到相同的投影坐标系。
  - 对土壤图和土地利用图进行栅格化或矢量化, 并与 DEM 对齐。
- 时间序列数据处理:
  - 对气象驱动数据进行质量控制、插补缺测值、单位转换。
  - 如果使用站点数据, 可能需要进行空间插值 (如泰森多边形法、反距离权重法) 以获得面平均气象输入。

## 3. 流域离散化与网格生成 (Domain Discretization & Mesh Generation):

- 基于处理后的 DEM 和河网数据, 利用 Delaunay 三角剖分等算法生成非结构化三角形网格。
- 控制网格生成的参数, 如最小单元数量 (NCELL)、最大单元面积 (MAX\_AREA)、三角形最小内角 (MIN\_ANGLE) 等, 以保证网格质量。
- 将河网嵌入到三角形网格中, 确定河段与三角形单元的相交关系。

## 4. 属性赋予与参数化 (Attribute Assignment & Parameterization):

- 将预处理好的土壤类型、土地利用类型等空间属性数据叠加到生成的三角形网格上, 为每个单元赋予相应的属性索引。
- 根据属性索引, 从预定义的参数查找表 (.para.soil, .para.geol, .para.lc) 中为每个单元或每类单元匹配相应的水力学参数。

## 5. SHUD 输入文件生成 (Input File Generation):

- 将所有处理好的空间几何信息、拓扑关系、单元属性、水力学参数、模型运行配置、初始条件以及时间序列驱动数据, 按照 SHUD 模型规定的格式写入相应的输入文件 (.sp.\*, .cfg.\*, .para.\*, .tsd.\*).

### 8.1.2. AutoSHUD 在流程中的作用

AutoSHUD 是一套基于 R 语言和 rSHUD 包开发的自动化脚本集合。它的核心目标是将上述复杂且耗时的建模前处理流程自动化。用户通过配置一个核心的文本文件 (.autoshoot.txt), 指定输入数据源、建模参数和输出路径, AutoSHUD 即可自动执行从数据下载、预处理、网格生成到 SHUD 输入文件生成的绝大部分步骤。

- 主要优势:
  - 提高效率: 大大减少手动操作的时间。
  - 增强可重复性: 确保了建模流程的标准化和结果的可复现。



- **降低技术门槛:** 即使对 GIS 操作和数据处理不太熟悉的用户, 也能通过配置参数快速搭建模型。
- **支持多种数据源:** 内置了对多种全球开放数据集 (如 CMFD, NLDAS, HWSO, GLC) 的处理逻辑。

**图表建议 8.1:** 插入一个详细的 SHUD 模型构建与前处理流程图, 清晰展示从原始数据到最终 SHUD 输入文件的各个步骤, 并突出 AutoSHUD 在其中扮演的角色。

**图表建议 8.2:** 插入 AutoSHUD 脚本内部主要步骤 (Step1, Step2, Step3 等) 之间的数据流和逻辑关系图 (类似于 [06]--2.3 中 01:06:21 的示意图, 但更侧重于数据流动)。

## 8.2. AutoSHUD 脚本结构与核心配置文件

理解 AutoSHUD 的脚本组织方式和核心配置文件的含义是有效使用该工具的前提。

### 8.2.1. AutoSHUD 脚本结构

AutoSHUD 通常以一个包含多个 R 脚本和辅助文件夹的包的形式提供。

#### • 主目录:

- `GetReady.R`: 初始化脚本, 负责加载必要的库、读取配置文件、设置全局变量。
- `Step1_RawDataProcessing.R`: 第一步处理脚本, 主要负责原始空间数据的读取、检查和初步处理。
- `Step2_DataSubset.R`: 第二步处理脚本, 负责裁剪和处理全球/区域数据集, 并将属性与流域关联。
- `Step3_BuildModel.R`: 第三步处理脚本, 负责流域离散化、网格生成、参数赋予和 SHUD 输入文件生成。
- `Step4_SHUD.R` (可选): 用于直接调用 SHUD 模型可执行文件进行模拟 (通常在参数率定脚本中更常用)。
- `Step5_ResultVisualization.R` (可选): 用于对 SHUD 输出结果进行初步可视化。
- `.autoshoot.txt` (或类似名称的配置文件): **核心配置文件**, 用户通过修改此文件来控制整个建模流程。

#### • 子目录:

- `Rfunction/`: 存放大量可重用的辅助 R 函数, 被主步骤脚本调用。这些函数通常封装了具体的数据转换、GIS 操作、统计计算等功能。
- `SubScript/`: 存放更细粒度的子步骤脚本, 通常对应特定数据源的处理逻辑 (例如, 处理 CMFD 数据的脚本, 处理 HWSO 土壤数据的脚本等)。`Step2_DataSubset.R` 会根据配置调用这些子脚本。
- `Example/`: 包含一个或多个示例项目的完整数据和配置文件, 供用户参考和测试。
- `Table/`: 存放一些静态的查找表, 如土地利用分类与参数的对应表、颜色表等。

### 8.2.2. 核心配置文件: `.autoshoot.txt` 文件详解

`.autoshoot.txt` 文件是用户与 AutoSHUD 交互的主要接口。通过修改此文件中的参数, 可以定制化建模流程以适应不同的研究区域和需求。

#### 8.2.2.1. 结构

- 纯文本文件, 每行定义一个参数。
- 参数名和参数值之间用空格或制表符分隔。
- 以 `#` 开头的行为注释行, 会被程序忽略。



8.2.2.2. 关键参数详解

(此处内容与第六部分 6.2.2.2 节重复，为保持笔记的自包含性，这里再次列出并强调其在 AutoSHUD 流程中的作用)

参数名	含义	示例值 (根据 [06]--2.3 01:51:37 左右的配置)	备注
ProjectName	项目名称	Futunxi	模型的唯一标识。
StartYear, EndYear	模拟时间范围 (年)	1980, 1997 (或 1980 根据后续修改)	定义气象驱动数据的年份范围。
DIR_OUT	建模结果输出根目录	../FutunxiDeployCMFD	AutoSHUD 生成的 SHUD 输入文件等将保存到此目录下的 ./deploy/ 子文件夹中。.. 表示上一级目录。
WBD	流域边界文件路径	../FutunxiBuild/data/WBD.shp	指向用户提供的流域边界 Shapefile。
RIV	河流网络文件路径	../FutunxiBuild/data/MeritRiver.shp	指向用户提供的河流网络 Shapefile。
DEM	DEM 文件路径	../FutunxiBuild/DEM.tif	指向用户提供的 DEM 栅格文件。
CRS	投影信息文件路径 (可选)	(示例中未显式指定, 可留空或注释掉)	如果提供, 所有输出将采用此投影; 否则自动确定。
FORCING_DATA	气象驱动数据源选择	0.5	0.5 代表使用服务器上存储的 CMFD 再分析资料。
FORCING_PATH	气象驱动数据源路径	/volume/data/forc/CMFD/	当 FORCING_DATA 选择服务器数据时, 此为数据存放的根目录。
SOIL_DATA	土壤数据源选择	0.1	0.1 代表使用服务器上存储的 HWSD 全球土壤数据。
LC_DATA	土地利用数据源选择	0.1	0.1 代表使用服务器上存储的 USGS GLC 数据。
NCELL	最小网格单元数量约束	1000	
MAX_AREA	最大网格单元面积约束	10 (单位 km²)	
MIN_ANGLE	三角形最小角度约束	30 (单位 度)	
TOL_WBD	流域边界简化容差	500 (单位 m)	
TOL_RIV	河流分段最大长度	2000 (单位 m)	
RIV_WIDTH, RIV_DEPTH	一级河流默认宽度和深度	3, 4 (单位 m)	
AQ_DEPTH	含水层默认厚度	6 (单位 m)	
QUICK_MODE	快速模式开关	0	
FROZEN_MODULE	冰冻圈模块开关	0	
START_TIME, END_TIME	模型运行时间范围 (可选)	(示例中未启用, 可留空或注释掉)	

8.2.2.3. 理解参数优先级

- 本地数据优先: 如果用户在 .autoshoot.txt 中为 WBD, RIV, DEM 等参数提供了具体的本地文件路径, AutoSHUD 将优先使用这些本地文件。

- **数据源选择:** 如果未提供本地文件路径, AutoSHUD 将根据 `FORCING_DATA`, `SOIL_DATA`, `LC_DATA` 等参数指定的代码 (如 `0.5` 代表 CMFD) 和对应的 `FORCING_PATH` 等路径, 从服务器或指定位置获取和处理全球/区域数据集。
- **投影:** 如果指定了 `CRS` 文件, 则以此为准; 否则自动生成。

**图表建议 8.3:** 插入一份完整的 `.autoshoot.txt` 文件示例, 并对每一行或关键参数块进行详细注释, 解释其含义、可选值和对建模的影响。

**图表建议 8.4:** 插入一个决策流程图, 展示 AutoSHUD 如何根据 `.autoshoot.txt` 中的参数 (特别是数据源选择和本地文件路径) 决定数据获取和处理的逻辑。

### 8.2.3. AutoSHUD 工作流程与主要脚本功能 (实操指南)

AutoSHUD 的建模过程通常分解为几个按顺序执行的 *R* 脚本。

#### 8.2.3.1. GetReady.R

- **功能:**
  - **加载必要的 R 包:** 如 `rSHUD`, `sp`, `raster`, `rgdal` 等。
  - **读取 `.autoshoot.txt` 配置文件:** 将配置文件中的参数解析并加载到 *R* 的工作环境中, 作为全局变量供后续脚本使用。
  - **设置工作路径和输出目录:** 根据配置创建必要的输出文件夹结构。
- **实操:**
  1. [在 *RStudio* 中打开 AutoSHUD 项目, 或者在 Terminal 中导航到 AutoSHUD 脚本所在的根目录。]
  2. [执行 `GetReady.R` 脚本。]

```
# 在 RStudio 控制台中执行
source('GetReady.R')
# 或者在 Terminal 中执行 (如果 GetReady.R 被设计为可独立运行的脚本)
# Rscript GetReady.R
```

- **输出与检查:** 脚本执行后, 检查 *R* 工作空间中是否已定义了配置文件中的变量 (例如, 在 *RStudio* 的 Environment 窗格中查看)。检查指定的输出目录是否已创建。

#### 8.2.3.2. Step1\_RawDataProcessng.R

- **功能:**
  - **读取用户指定的原始空间数据:** 包括流域边界 (`WBD`)、河流网络 (`RIV`) 和 DEM。
  - **数据有效性初步检查:** 例如, 检查 `Shapefile` 是否完整, DEM 是否可读。
  - **投影转换与统一:** 如果用户指定的原始数据投影与目标投影 (由 `CRS` 参数或自动确定) 不同, 则进行重投影。
  - **数据裁剪 (初步):** 可能根据流域边界对 DEM 进行初步裁剪。
- **实操:**
  1. [确保 `GetReady.R` 已成功执行, 并且 `.autoshoot.txt` 中正确配置了 `WBD`, `RIV`, `DEM` 的路径。]
  2. [执行 `Step1_RawDataProcessng.R` 脚本。]

```
source('Step1_RawDataProcessng.R')
```

- 输出与检查:

- 在配置的输出目录 (`DIR_OUT` 下的 `DataPRE/GCS` 和 `DataPRE/PCS`) 中会生成处理后的空间数据文件 (通常是 `GeoTIFF` 和 `Shapefile` 格式)。
- 在 `fig/` 目录下可能会生成一些预览图件, 如原始 DEM 图、处理后的流域边界和河网图等。
- **关键检查点:** 投影是否正确统一, 裁剪范围是否合理。

### 8.2.3.3. Step2\_DataSubset.R

- 功能:

- **处理全球/区域数据集:** 根据 `.autoshoot.txt` 中的 `SOIL_DATA`, `LC_DATA`, `FORCING_DATA` 等配置, 调用相应的子脚本 (`SubScript/` 目录下) 来获取、裁剪、重采样和格式化土壤、土地利用和气象驱动数据。
  - 例如, 如果配置使用 HWSD 土壤数据, 则会调用处理 HWSD 数据的子脚本, 根据流域范围提取土壤类型, 并可能将其转换为栅格格式。
  - 如果配置使用 CMFD 气象数据 ([06]--2.3 中 01:55:33 开始演示 CMDF 数据处理), 则会从服务器路径读取相应的 NetCDF 文件, 提取指定时间范围和空间范围的数据, 并将其转换为 SHUD 可用的 `.csv` 格式。
- **属性空间化:** 将处理好的土壤、土地利用等属性信息与流域范围或初步网格进行空间关联。
- **时间序列格式化:** 将气象驱动数据整理成 SHUD 所需的时间序列格式。

- 实操:

1. [确保 `Step1_RawDataProcessng.R` 已成功执行。]
2. [确保 `.autoshoot.txt` 中正确配置了所需数据源及其路径。]
3. [执行 `Step2_DataSubset.R` 脚本。]

```
source('Step2_DataSubset.R')
```

- 输出与检查:

- 在 `DataPRE/` 目录下会生成处理后的土壤、土地利用等栅格或矢量文件。
- 在 `Forcing/` 目录下会生成格式化后的气象驱动数据文件 (通常是多个 `.csv` 文件, 每个对应一个站点/格点, 以及一个指向这些文件的列表文件, 如 `meteo.txt` 或类似名称)。
- 在 `fig/` 目录下可能会生成土壤、土地利用分布图, 以及气象数据的时间序列图等。
- **关键检查点:** 土壤和土地利用分类是否正确覆盖流域, 气象数据的时间范围和数值是否合理。

### 8.2.3.4. Step3\_BuidModel.R

- 功能:

- 流域离散化与网格生成:

- 利用 `rSHUD::triangulate` 等函数, 基于预处理好的 DEM、河网以及 `.autoshoot.txt` 中指定的网格控制参数 (`NCELL`, `MAX_AREA`, `MIN_ANGLE`, `TOL_WBD`, `TOL_RIV`) 生成非结构化三角形网格。
- 确定河段与三角形单元的相交关系。
- 属性与参数赋予:
  - 将 Step2 生成的土壤、土地利用属性空间数据与生成的三角形网格叠加, 为每个单元赋予属性索引。
  - 根据属性索引, 从内部或外部的参数查找表 (`Table/` 目录或 `rSHUD` 内置) 中查找并分配水力学参数。
- SHUD 输入文件生成:
  - 将所有几何信息、拓扑关系、属性索引、水力学参数、模型运行配置 (基于默认值和 `.autoshoot.txt` 的部分设置)、默认初始条件以及格式化后的时间序列数据列表, 按照 SHUD 要求的格式写入到最终的输入文件集 (`.sp.*`, `.cfg.*`, `.para.*`, `.tsd.forc` 等)。
- 实操:
  1. [确保 `Step1_RawDataProcessng.R` 和 `Step2_DataSubset.R` 已成功执行。]
  2. [执行 `Step3_BuidModel.R` 脚本。]

```
source('Step3_BuidModel.R')
```

- 输出与检查:
  - 在 `DIR_OUT` 下的 `deploy/<project_name>/MODEL/Input/` 目录中会生成完整的 SHUD 输入文件集。
  - 在 `DIR_OUT` 下的 `deploy/<project_name>/MODEL/gis/` 目录中会生成可用于 GIS 软件查看的模型网格 (`domain.shp`)、河网 (`river.shp`) 和河段分割 (`seg.shp`) 文件。
  - 在 `fig/` 目录下可能会生成最终模型网格图、单元面积分布直方图等。
  - 关键检查点:
    - 检查生成的输入文件是否完整。
    - 使用 GIS 软件打开 `domain.shp` 和 `river.shp`, 检查网格质量、河网与网格的匹配情况。
    - 抽查部分输入文件的内容, 如 `.sp.mesh` 中的邻居关系, `.sp.att` 中的属性索引是否在合理范围。

## 8.2.4. 实操流程总结

通常, AutoSHUD 的建模流程是按顺序执行上述脚本:

1. 配置: 修改 `.autoshoot.txt` 文件。
2. 初始化: 运行 `GetReady.R`。
3. 原始空间数据处理: 运行 `Step1_RawDataProcessng.R`。
4. 属性数据与气象数据处理: 运行 `Step2_DataSubset.R`。
5. 模型构建与输入文件生成: 运行 `Step3_BuidModel.R`。

每一步完成后，都应检查其输出和日志，确保该步骤成功执行且结果合理，再进行下一步。

### 小结:

第八章节详细介绍了 AutoSHUD 自动化建模工具的脚本结构、核心配置文件 `.autoshoot.txt` 的各项参数，以及通过依次执行 `GetReady.R` 和 `Step1` 到 `Step3` 脚本来完成从原始数据到 SHUD 模型输入文件的整个前处理和构建过程。理解 AutoSHUD 的工作流程和配置方法，将极大地提升 SHUD 建模的效率和规范性。

### 思考题:

1. 在 `.autoshoot.txt` 中，气象驱动数据源参数 `FORCING_DATA` 和 `FORCING_PATH` 是如何协同工作的？如果我想使用我自己准备的、非标准格式的本地气象站点数据，AutoSHUD 是否能直接处理？如果不能，我应该如何修改或扩展 AutoSHUD 的功能？
2. AutoSHUD 生成的 SHUD 输入文件中，水力学参数 (`.para.soil`, `.para.geol`, `.para.lc`) 的值是如何确定的？它们是基于全球默认值，还是会根据用户提供的土壤和土地利用数据进行更精细的匹配？
3. 如果 `Step3_BuidModel.R` 脚本在生成三角形网格时耗时过长或失败，可能是什么原因？应如何调整 `.autoshoot.txt` 中的网格控制参数 (`NCELL`, `MAX_AREA`, `MIN_ANGLE` 等) 来优化网格生成过程？

## 9. SHUD 模型输入文件生成

在第六部分 (对应原计划的第五章) 中，我们已经详细介绍了 AutoSHUD 如何通过 `Step3_BuidModel.R` 脚本，将预处理好的数据整合成 SHUD 模型可读取的输入文件。为了避免内容上的大量重复，并保持框架的逻辑递进，本章将重点复习和整合输入文件生成的关键环节，并强调在此过程中用户需要关注的检查和验证步骤。

我们假设已经通过 AutoSHUD 的 `Step1_RawDataProcessng.R` 和 `Step2_DataSubset.R` (或等效的手动/rSHUD 处理) 完成了以下工作：

- 原始空间数据 (DEM, 河网, 流域边界) 已被读取、裁剪、投影转换并保存。
- 土壤、土地利用等空间属性数据已根据流域范围提取并栅格化/矢量化。
- 气象驱动数据已根据配置下载、提取并初步格式化。

现在，我们将聚焦于如何将这些中间成果转化为 SHUD 模型的最终输入文件集。

### 9.1. SHUD 模型输入文件结构 (复习与整合)

再次强调，SHUD 模型运行依赖于一套结构化、格式化的输入文件。理解这些文件的作用和相互关系是关键。

- 核心文件类别回顾：
  - `.sp.*`: 定义空间几何与拓扑 (网格单元、河段、单元属性)。
  - `.para.*`: 定义物理参数 (土壤、地质、土地利用的水力学和物理特性)。
  - `.cfg.*`: 控制模型运行行为 (时间、求解器、输出、校准基准、初始条件)。
  - `.tsd.*`: 提供时间序列驱动 (气象、LAI 等)。

### 9.2. 利用 AutoSHUD/rSHUD 构建模型输入文件 (实操指南)

AutoSHUD 中的 `Step3_BuidModel.R` 脚本是实现这一目标的核心。如果您不使用 AutoSHUD 的完整流程，也可以参考其内部逻辑，使用 rSHUD 包中的相应函数手动或半自动地生成这些文件。

### 9.2.1. 理解 Step3 的作用 (或等效 rSHUD 函数的功能)

`Step3_BuidModel.R` (或 rSHUD 中的核心建模函数，如 `rSHUD::ModelBuild`) 的主要任务包括：

1. **读取预处理数据:** 加载由 Step1 和 Step2 (或等效过程) 生成的 DEM、河网、流域边界、处理后的土壤图、土地利用图等。
2. **非结构化网格生成:**
  - 基于 DEM 和河网，利用 Delaunay 三角剖分算法 (`rSHUD::triangulate`) 生成三角形不规则网络 (TIN)。
  - 应用 `.autoshoot.txt` 中定义的网格控制参数 (`NCELL`, `MAX_AREA`, `MIN_ANGLE`, `TOL_WBD`, `TOL_RIV`) 来优化网格质量和密度。
  - 输出 `.sp.mesh` 文件，包含单元拓扑和节点几何信息。
3. **属性赋予:**
  - 将处理后的土壤、土地利用栅格数据与生成的三角形网格进行空间叠加。
  - 为每个三角形单元确定其主要的土壤类型和土地利用类型。
  - 将确定的类型编码 (索引) 写入 `.sp.att` 文件。
  - 同时，根据泰森多边形法或最近邻法，为每个单元分配一个气象驱动站点/格点的索引，也写入 `.sp.att`。
4. **河流网络处理:**
  - 将原始河网与三角形网格进行空间相交分析。
  - 生成 `.sp.riv` 文件，包含河段的连接关系和基本几何属性。
  - 生成 `.sp.rivseg` 文件，详细记录每个河段被三角形单元切割成的线段及其与单元的对应关系。
5. **参数文件生成:**
  - 根据 `.sp.att` 文件中每个单元的土壤、地质、土地利用类型索引。
  - 从内部参数查找表或用户提供的参数表 (例如 `Table/` 目录下的 `.csv` 文件) 中，查找并写入 `.para.soil`, `.para.geol`, `.para.lc` 文件。
6. **配置文件生成:**
  - 生成默认的 `.cfg.para` (运行参数), `.cfg.calib` (校准参数基准), 和 `.cfg.ic` (初始条件) 文件。这些文件的部分内容会参考 `.autoshoot.txt` 中的设置。
7. **时间序列数据索引文件生成:**
  - 生成 `.tsd.forc` 文件，列出所有处理好的气象驱动数据文件 (`.csv` 格式) 及其路径。
  - 类似地，如果配置了 LAI、融雪因子等，也会生成对应的 `.tsd.lai`, `.tsd.mf` 等索引文件。
8. **GIS 输出文件生成:**
  - 为了方便用户在 GIS 软件中查看和验证模型结构，通常会输出模型网格 (`domain.shp`)、使用的河网 (`river.shp`) 和河段分割 (`seg.shp`) 的 Shapefile 文件到 `gis/` 目录。



## 9.2.2. 依赖关系

`Step3_BuidModel.R` 的成功执行高度依赖于前面步骤生成的中间数据文件的完整性和正确性。例如：

- 准确的、经过投影转换和裁剪的 DEM、河网、流域边界。
- 已栅格化并与 DEM 对齐的土壤和土地利用分类图。
- `.autoshoot.txt` 中正确的网格控制参数和数据路径配置。

## 9.2.3. 实操：运行 AutoSHUD Step3 脚本

1. **确保前序步骤已完成:** 确认 `GetReady.R`, `Step1_RawDataProcessing.R`, 和 `Step2_DataSubset.R` 已成功运行, 并且相关的中间数据已生成在预期的目录中。
2. **检查 `.autoshoot.txt` 配置:** 再次确认 `.autoshoot.txt` 中的网格参数 (`NCELL`, `MAX_AREA`, `MIN_ANGLE` 等) 和相关路径设置符合您的需求。
3. **执行脚本:**

```
# 在 RStudio 控制台或 Terminal 中, 确保当前工作目录是 AutoSHUD 项目的根目录
source('Step3_BuidModel.R')
# 或者 Rscript Step3_BuidModel.R
```

4. **监控输出:** 观察 R 控制台或终端的输出信息。`Step3` 通常会打印较多的日志, 包括网格生成的统计信息 (单元数、面积分布、角度分布)、属性赋予的概况等。([06]--2.3 中 01:55:33 及其后演示了 Step3 的运行和输出信息, 例如输出了网格数量、面积、分辨率、河段数量和长度等统计)

## 9.2.4. 检查生成的输入文件集

脚本成功运行后, 在 `.autoshoot.txt` 中 `DIR_OUT` 参数指定的输出目录下的

`deploy/<ProjectName>/MODEL/Input/` 文件夹中, 应该能找到一套完整的 SHUD 输入文件。

- **对照文件列表:** 参考本笔记第三、四、五章节中描述的 SHUD 输入文件列表, 检查是否所有必需的文件都已生成。
- **检查文件大小:** 对于一些包含大量数据的文件 (如 `.sp.mesh`, `.sp.att`), 检查其文件大小是否在一个合理的范围内 (过小可能表示数据缺失或处理错误)。

## 9.2.5. 理解生成的输入文件内容

回顾第三、四、五章节关于各个输入文件结构和内容的详细描述, 打开部分关键文件进行抽查。

- `.sp.mesh`: 检查单元数量是否与日志输出一致, 节点坐标范围是否与流域范围匹配。
- `.sp.att`: 检查 `SOIL`, `GEOL`, `LC`, `FORC` 等索引值是否在合理的范围内 (例如, 如果只有 3 种土壤类型, `SOIL` 索引不应大于 3)。
- `.para.soil`, `.para.geol`, `.para.lc`: 检查参数值是否具有物理意义, 是否与预期的类型特征相符。
- `.cfg.para`: 检查模拟时间、输出频率等是否按配置生成。
- `.tsd.forc`: 检查列出的气象数据文件名和路径是否正确, 起始日期是否正确。



**图表建议 9.1:** 插入一个流程图, 展示 `Step3_BuidModel.R` (或核心建模函数) 如何利用预处理数据 (DEM, 河网, 土壤图, 土地利用图, 气象数据列表) 和配置参数 (`.autoshoot.txt`) 来生成 SHUD 的各类输入文件 (`.sp.*`, `.para.*`, `.cfg.*`, `.tsd.*`)。

## 9.2.6. 常见问题

- **脚本中断或报错:**
  - **原因:** 前序步骤数据缺失或格式错误; `.autoshoot.txt` 配置参数不合理 (如网格参数导致无法生成有效网格); rSHUD 包或其依赖包函数出错。
  - **排查:** 仔细阅读 R 控制台的错误信息和警告信息, 定位问题发生在哪一步或哪个函数调用。检查相关输入数据和配置文件。
- **生成的网格质量差:**
  - **现象:** 出现大量狭长三角形、角度过小的三角形, 或者网格密度分布不均。
  - **原因:** `MIN_ANGLE` 设置过小; `MAX_AREA` 与 `NCELL` 设置不协调; `TOL_WBD` 或 `TOL_RIV` 设置不当导致边界或河网过于复杂或简化过度。
  - **解决:** 调整 `.autoshoot.txt` 中的网格控制参数, 重新运行 `Step3`。
- **属性赋予错误:**
  - **现象:** `.sp.att` 中某些单元的属性索引为 NA 或异常值。
  - **原因:** 预处理的土壤/土地利用栅格数据未能完全覆盖生成的网格范围, 或者空间叠加分析出错。
  - **解决:** 检查 Step2 输出的属性栅格数据, 确保其覆盖范围和分辨率。
- **参数文件内容异常:**
  - **现象:** `.para.*` 文件中的参数值全为 NA 或与预期严重不符。
  - **原因:** 参数查找表 (`Table/` 目录中的 `.csv`) 内容错误或与属性索引不匹配。
  - **解决:** 检查参数查找表的正确性。

## 9.2.7. 利用生成的 GIS 文件进行验证

AutoSHUD 通常会在 `deploy/<ProjectName>/MODEL/gis/` 目录下生成可供 GIS 软件查看的 Shapefile 文件, 如 `domain.shp` (三角形网格), `river.shp` (模型使用的河网), `seg.shp` (河段与单元的相交线段)。

- **实操: 在 QGIS 或 ArcGIS 中打开这些 Shapefile。**
- **检查内容:**
  - **domain.shp:**
    - **网格形状和密度:** 是否符合预期? 是否存在明显的恶劣三角形?
    - **与流域边界和河网的叠合:** 网格是否良好地覆盖了流域, 并与河网对齐?
    - **属性可视化:** 尝试将其与 `.sp.att` 文件关联 (例如通过单元 ID), 并根据土壤类型、土地利用类型或分配的气象站点对着色, 检查空间分布的合理性。
  - **river.shp:**
    - **河网连通性:** 河网是否连续? 是否存在断头河或孤立河段? 下游连接是否正确?
  - **seg.shp:**

- 与 `domain.shp` 和 `river.shp` 的叠合: 检查河段分割是否正确反映了河网与三角形网格的相交情况。

图表建议 9.2: 插入一张在 QGIS (或类似 GIS 软件) 中叠加显示 `domain.shp` (已按某种属性如土地利用类型着色)、`river.shp` 和原始流域边界的截图, 用于展示建模结果的空间正确性。

## 小结:

第九章复习并整合了 SHUD 模型输入文件生成的关键环节, 重点强调了在 AutoSHUD (或 rSHUD) 生成这些文件后, 进行详细检查和验证的重要性。通过检查生成文件的完整性、内容合理性以及利用 GIS 工具进行空间可视化, 可以有效发现和修正建模前处理过程中的潜在问题, 为后续的模型模拟和参数率定打下坚实的基础。

## 思考题:

1. 如果 `.sp.mesh` 文件中存在大量角度非常小的三角形 (例如小于 10 度), 这会对 SHUD 模型的数值计算产生什么影响? 应如何调整 AutoSHUD 的参数来改善这种情况?
2. 在检查 `.sp.att` 文件时, 如果发现某个三角形单元的 `SOIL` 索引超出了 `.para.soil` 文件中定义的土壤类型总数, 这意味着什么? 如何追溯这个错误的来源?
3. `.tsd.forc` 文件中列出的气象驱动 `.csv` 文件名通常是基于格点坐标或站点编号生成的。如果这些文件名与 `.sp.att` 文件中 `FORC` 列的索引不匹配, 模型运行时会出现什么问题? 如何修正这种不匹配?

# 10. SHUD 模型模拟运行与初步结果分析

在成功构建并生成了 SHUD 模型的全套输入文件之后, 下一步就是实际运行模型进行模拟, 并对输出结果进行初步的读取和分析。本章将详细介绍模型运行前的准备工作、不同环境下的运行方法、模型预热的重要性与操作, 以及如何使用 rSHUD 工具对模拟结果进行初步的检查和可视化。

## 10.1. SHUD 模型运行准备

在启动 SHUD 模型可执行文件之前, 需要确保所有输入文件已正确组织, 并且模型的运行参数和初始条件已根据模拟需求进行了合理配置。

### 10.1.1. 组织输入文件

1. **创建运行目录:** 建议为每次模拟或每个情景创建一个独立的运行目录, 以保持文件组织的清晰。
2. **复制输入文件:**
  - 将通过 AutoSHUD (通常在 `deploy/<ProjectName>/MODEL/Input/` 目录下) 或手动方式生成的完整 SHUD 输入文件集 (`.sp.*`, `.cfg.*`, `.para.*`, `.tsd.forc` 等) 复制到您创建的运行目录中。
  - 确保 `.tsd.forc` 文件中引用的所有气象驱动数据 `.csv` 文件也已复制到 `.tsd.forc` 文件第一行所指定的相对或绝对路径下 (通常是运行目录下的 `forcing/` 或类似子目录)。
3. **放置 SHUD 可执行文件:**
  - 将您编译好的 SHUD 可执行文件 (`shud` 或 `shud_omp`) 复制到运行目录中。
  - 或者, 如果 SHUD 可执行文件所在的路径已添加到系统的 `PATH` 环境变量中, 则无需复制, 可以直接在任何位置调用。

**图表建议 10.1:** 插入一个 SHUD 模型典型运行目录结构示意图 (展示运行目录下包含 `input/` 子目录存放各类输入文件, 以及 `shud` 可执行文件, 并可能有一个 `forcing/` 子目录存放气象 `.csv` 文件)。

### 10.1.2. 实操：配置模型运行参数 (`.cfg para`)

`.cfg para` 文件是控制 SHUD 模型运行行为的核心。在运行前务必仔细检查和配置。

- 打开文件: 使用文本编辑器打开运行目录下的 `[ProjectName].cfg para.txt` 文件。
- 关键参数回顾与配置:
  - **START** 和 **END**: ([07]--3 中 01:38:02 演示了修改这两个参数)
    - 定义模型模拟的开始和结束时间。单位通常是相对于气象驱动数据起始日期的天数。
    - 例如, 如果气象数据从 2001-01-01 开始, 要模拟 2001 年全年, 则 `START = 0` (或 `1`, 取决于约定), `END = 364` (或 `365`)。
    - 务必确保 **START** 和 **END** 定义的时段在您的气象驱动数据覆盖范围内。
  - 输出频率 (`dt_*` 参数): ([07]--3 中 01:39:34 演示了修改输出频率)
    - 单位是分钟。
    - 对于初步运行或参数率定, 通常只关心少数关键输出, 如流域出口径流 (`dt_Qr_down = 1440` 表示每天输出)。其他不必要的输出可以设置为 `0` 以节省磁盘空间和运行时间。
  - 求解器参数:
    - `MAX_SOLVER_STEP`: 最大允许时间步长 (分钟)。例如, 对于日输出或小时输出的模拟, 可以设置为 `10`, `30`, 或 `60`。较小的值通常更稳定, 但计算耗时。
    - `ABSTOL`, `RELTOL`: 绝对和相对容差。例如 `1e-4` 或 `1e-5`。较小的值精度更高, 但计算耗时。
  - 输出格式:
    - `BINARY_OUTPUT = 1` (推荐, 输出 `.dat` 二进制文件)。
    - `ASCII_OUTPUT = 0` (除非需要人工查看小型结果, 否则不推荐)。
  - 其他参数: 如 `SPINUPDAY`, `INIT_MODE` 等, 在预热阶段会重点配置。对于首次运行, 可以暂时保持 AutoSHUD 生成的默认值。
- 保存文件。

### 10.1.3. 实操：准备初始条件 (`.cfg ic`)

`.cfg ic` 文件为模型提供了所有状态变量的初始值。

- 来源:
  - AutoSHUD 或 GHDC 通常会根据一些简单规则 (如高程、默认含水层比例) 生成一个初始的 `.cfg ic` 文件。
  - 最佳实践是通过模型预热 (Spin-up) 生成。
- 对于首次运行或测试:
  - 可以使用 AutoSHUD 生成的默认 `.cfg ic` 文件。
  - **注意:** 如果模型在初始阶段运行非常缓慢或出现数值不稳定, 很可能是因为这个默认的初始条件与流域的实际准稳定状态相差太远。

- **文件结构回顾:** 包含三角形单元和河段的初始状态变量值 (如水深、水位)。([07]--3 中 01:31:02 详细解读了其结构)

### 10.1.4. 将文件组织到运行目录

确保所有必需的、已配置好的输入文件都在 SHUD 可执行文件能够访问的路径下。通常是将 AutoSHUD 生成的 `deploy/<ProjectName>/MODEL/Input/` 目录下的所有内容, 以及 `deploy/<ProjectName>/Forcing/` 目录下的所有 `.csv` 文件, 复制到您新建的运行目录中对应的 `input/` 和 `forcing/` 子目录。

## 10.2. SHUD 模型模拟运行 (实操指南)

### 10.2.1. 理解不同运行环境的运行方式

- **本地终端 (Linux/Mac Terminal, Windows MobaXterm/PowerShell with WSL):**
  - 直接在命令行执行 SHUD 程序。
  - 优点: 操作直观, 便于快速测试和调试小型案例。
  - 缺点: 长时间运行会占用当前终端; 大规模计算受本地机器资源限制; 不适合并行率定。
- **HPC 队列 (Slurm):**
  - 通过提交作业脚本到计算节点执行。
  - 优点: 可以利用 HPC 集群的强大计算资源, 适合长时间、大规模模拟和并行任务。
  - 缺点: 需要编写作业脚本, 熟悉 Slurm 命令。

### 10.2.2. 详细步骤: 在本地终端运行 SHUD

1. **打开终端:** 启动 Terminal (Linux/Mac) 或 MobaXterm (Windows)。
2. **进入运行目录:** 使用 `cd` 命令切换到包含 SHUD 输入文件和可执行文件的运行目录。

```
cd /path/to/your/shud_run_directory
```

3. **执行 SHUD 命令:** ([05]--2.2 中 00:59:52 演示了运行命令)

```
./shud <project_name>
```

- `<project_name>` 必须与您的输入文件名中的项目名部分完全一致。
- 例如, 如果您的输入文件是 `mybasin.sp.mesh.dat` 等, 则命令为 `./shud mybasin`。
- 如果运行的是 OpenMP 并行版本 (`shud_omp`), 可以指定线程数:

```
./shud_omp <project_name> -n <number_of_threads>
```

例如, 使用 4 个线程: `./shud_omp mybasin -n 4`。

4. **监控终端输出:**

- 模型开始运行后, 会在终端打印运行日志, 包括读取文件信息、模拟进度 (当前模拟到的天数、百分比、每个时间步的迭代次数等)、警告和错误信息。([05]--2.2 中 01:00:00 开始解读运行日志)

- 观察日志可以判断模型是否正常运行。

5. 中断运行: 如果需要提前终止模拟, 在终端中按下 `Ctrl + C`。

### 10.2.3. 详细步骤: 在 HPC 队列提交 SHUD 运行任务

#### 1. 准备 Slurm 作业脚本 ( `.sh` 文件):

- 创建一个文本文件 (例如 `run_shud_slurm.sh`), 内容如下:

```
#!/bin/bash
#SBATCH --job-name=shud_sim_projectX # 任务名称, 可自定义
#SBATCH --output=shud_sim_%j.out    # 标准输出日志文件名 (%j 会被替换为作业ID)
#SBATCH --error=shud_sim_%j.err      # 标准错误日志文件名
#SBATCH --partition=your_partition   # 指定作业提交的队列/分区 (需根据实际HPC配置填写)
#SBATCH --nodes=1                    # 通常 SHUD 单次模拟在单个节点上运行
#SBATCH --ntasks-per-node=1          # 每个节点运行一个 SHUD 任务
#SBATCH --cpus-per-task=40           # 申请的 CPU 核心数 (如果运行 shud_omp, 可以设置为节点最大核心数)
#SBATCH --mem=16G                    # 申请的内存量 (例如 16GB, 根据模型规模调整)
#SBATCH --time=02:00:00              # 任务最大运行时长 (例如 2 小时)

# 加载可能需要的模块 (根据 HPC 环境配置, 例如 R, GCC, SUNDIALS, GDAL 等)
# module load R/4.2.0
# module load gcc/9.3.0
# module load sundials/5.8.0

# 定义工作目录和项目名
WORK_DIR="/path/to/your/shud_run_directory_on_hpc" # 必须是 HPC 上的绝对路径
PROJECT_NAME="my_project_on_hpc"

# 进入工作目录
cd ${WORK_DIR}

# 执行 SHUD 命令
# 如果是串行版本
./shud ${PROJECT_NAME}
# 如果是 OpenMP 并行版本, 确保 SLURM_CPUS_PER_TASK 与 -n 参数匹配
# export OMP_NUM_THREADS=${SLURM_CPUS_PER_TASK}
# ./shud_omp ${PROJECT_NAME} -n ${SLURM_CPUS_PER_TASK}

echo "SHUD simulation finished."
```

- 注意: 脚本中的路径 (`WORK_DIR`) 必须是 HPC 文件系统上的绝对路径。队列名称 (`your_partition`)、申请的核心数、内存和运行时长需要根据您的 HPC 环境和模拟需求进行调整。

2. 上传作业脚本和输入文件到 HPC: 使用 `scp` 命令或 MobaXterm 的 SFTP 功能, 将作业脚本和完整的 SHUD 运行目录上传到 HPC。

#### 3. 登录 HPC 并提交作业:

- 通过 SSH 登录 HPC。
- 进入存放作业脚本的目录。

- 使用 `sbatch` 命令提交作业：

```
sbatch run_shud_slurm.sh
```

系统会返回一个作业 ID (`job_id`)。

#### 4. 监控任务状态:

```
squeue -u your_username # 查看您的作业状态  
scontrol show job <job_id> # 查看特定作业的详细信息
```

- 5. **查看日志:** 任务完成后 (或失败后), 在作业脚本中指定的输出和错误日志文件 (`shud_sim_<job_id>.out`, `shud_sim_<job_id>.err`) 中查看运行信息。

## 10.2.4. 常见问题

- **文件找不到 (File not found 或类似错误):**
  - 检查 `.cfg para` 中指定的输入文件路径是否正确 (相对于运行目录)。
  - 检查项目名称是否与输入文件名中的项目名部分完全一致。
  - 确保所有必需的输入文件都已复制到运行目录中。
- **参数读取错误 (Error reading parameter 或类似错误):**
  - 检查 `.cfg para`, `.cfg calib`, `.para.*` 等配置文件和参数文件的格式是否正确, 数值是否在合理范围内。
  - 注意文本文件的编码 (应为 UTF-8) 和换行符 (Linux 通常为 LF)。
- **内存不足 (通常在 HPC 日志中提示 oom-kill 或内存分配失败):**
  - 对于大型流域或高分辨率网格, 模型可能需要大量内存。
  - 在 HPC 作业脚本中增加申请的内存量 (`--mem`)。
  - 优化模型配置, 如减少输出变量或降低输出频率。
- **运行超时 (HPC 作业因超出 walltime 而被终止):**
  - 在 HPC 作业脚本中增加申请的运行时长 (`--time`)。
  - 优化模型以提高计算效率。
- **数值不稳定 (模型在模拟过程中提前终止, 日志中可能出现大量 CVOICE 错误或迭代失败信息):**
  - **初始条件不合理:** 这是常见原因, 需要进行模型预热。
  - **模型参数不合理:** 某些参数组合可能导致物理过程计算出现极端情况。
  - **网格质量差:** 存在恶劣三角形。
  - **求解器参数设置不当:** `MAX_SOLVER_STEP` 过大, 或容差 (`ABSTOL`, `RELTOL`) 设置不合适。
  - **气象驱动数据异常:** 驱动数据中存在极端值或不合理数据。

**图表建议 10.2:** 插入一个典型的 SHUD 运行日志片段截图 (来自本地终端或 HPC 的 `.out` 文件), 并高亮显示关键信息, 如模拟天数、百分比、每步迭代次数、警告等。



## 10.3. 模型预热 (Spin-up) (实操指南)

模型预热是确保模拟结果不受初始条件任意猜测影响的关键步骤。

### 10.3.1. 理解预热的目的

- 水文系统具有“记忆性”，特别是地下水和土壤水储量。如果初始条件与流域的长期平均状态（或准稳定状态）相差很远，模型需要一段时间来调整内部状态，直到其响应主要由外部驱动（如气象）决定，而不是由初始猜测值主导。这段调整时间就是预热期。
- 目标:** 使模型达到一个动态平衡或准稳定状态，从而可以更准确地模拟后续时段的真实水文响应，并为参数率定提供更可靠的基准。

### 10.3.2. 预热期的长度设置

- 影响因素:** 流域大小、气候特征、地质条件、土壤类型、主要产流机制。
- 一般原则:**
  - 响应慢的系统 (如大型平原流域、深厚含水层、地下水主导型流域) 需要**更长**的预热期，可能需要数年甚至数十年。
  - 响应快的系统 (如小型山区流域、地表流主导型流域) 预热期可以**相对较短**，可能几个月到一两年。
- 判断标准 (经验性):**
  - 观察流域总储水量 (特别是地下水储量) 的时间序列，当其年际变化趋于稳定，不再有明显的单向增加或减少趋势时，可认为预热基本完成。
  - 观察流域出口基流部分的稳定性。
- 实践建议:** 通常从几年开始尝试，如果储量变化仍不明显，则逐步增加预热时长。

### 10.3.3. 预热方法

- 选择一段较早的、具有代表性的气象驱动数据作为预热期驱动。
- 使用默认或猜测的初始条件 (`.cfg.ic`) 开始模拟。
- 运行模型覆盖整个预热期。
- 将预热期结束时刻的模型状态 (所有单元的水量、水位等) 保存下来，作为新的初始条件。

### 10.3.4. 实操：进行预热模拟

- 配置 `.cfg para` 文件:**
  - `START = 0` (或其他预热起始点)。
  - `END = <预热期总天数>`。
  - `INIT_MODE = 0` 或 `1` 或 `2` (使用默认或猜测的初始条件)。
  - 确保输出频率设置合理，至少在预热结束时能够输出模型状态 (例如，可以将 `.update` 文件的输出频率设置为预热期长度对应的分钟数，或者确保在 `END` 时刻有状态输出)。
- 运行 SHUD 模型** (如 10.2 节所述)。
- 监控预热过程:** 观察关键状态变量 (如流域平均地下水储量) 的变化趋势。



### 10.3.5. 实操：利用预热结果更新 `.cfg.ic` 文件

SHUD 模型在运行时，会按照 `.cfg para` 中 `dt_ye_*` 等参数指定的频率，或者在特定条件下 (如参数校准的每次模拟结束时)，输出包含当前所有状态变量的文件。这些文件通常命名为 `[ProjectName].update_[TotalMinutes].[Format]`，其中 `[TotalMinutes]` 是从模拟开始到当前时刻的总分钟数。

#### 1. 找到预热结束时刻的状态文件：

- 确定预热模拟结束时对应的总分钟数。例如，如果预热 3 年 (假设每年 365 天)，则总天数为  $3 \times 365 = 1095$  天。如果模型时间单位是天，输出频率是按天，那么在第 1095 天结束时会有输出。如果模型内部时间步长或输出是以分钟为单位，则需换算。
- 在模型输出目录中找到对应的 `.update_[TotalMinutes].dat` (或 `.csv`) 文件。

#### 2. 替换旧的 `.cfg.ic` 文件：

- 将找到的 `.update` 文件复制到您的 SHUD 输入目录 (`input/` 或类似)。
- 将其重命名为与项目匹配的初始条件文件名，例如 `[ProjectName].cfg.ic.dat` (或 `.txt` 如果是 ASCII)。
- 覆盖原有的 `.cfg.ic` 文件。

```
# 示例：假设预热结束在模拟的第 1095 天，输出频率为每天 (1440 分钟)
# 结束时的总分钟数为 1095 * 1440 = 1576800
# 假设输出目录为 ./output/，输入目录为 ./input/，项目名为 mybasin
cp ./output/mybasin.update_1576800.dat ./input/mybasin.cfg.ic.dat
```

#### 3. 修改 `.cfg para` 中的 `INIT_MODE`：

- 在后续的正式模拟或参数率定中，将 `.cfg para` 文件中的 `INIT_MODE` 设置为 3 (Warm start)，这样模型就会从更新后的 `.cfg.ic` 文件读取初始条件。

**图表建议 10.3:** 插入一个流域总储水量或平均地下水储量随预热时间变化的曲线图，展示储量如何从初始猜测值逐渐趋于稳定的过程。

## 10.4. SHUD 模型输出结果读取与初步分析 (实操指南)

模型运行成功并生成输出文件后，可以使用 rSHUD 工具包在 R 环境中读取和分析这些结果。

### 10.4.1. SHUD 模型输出文件组织

回顾第七章节，输出文件通常保存在 `.cfg para` 中指定的输出目录下 (`output/` 子目录)，并遵循特定的命名规范。

### 10.4.2. 实操：使用 rSHUD 读取输出文件

- 启动 R 或 RStudio。
- 加载 rSHUD 包：

```
library(rSHUD)
```

### 3. 设置 SHUD 项目环境变量:

- 这一步告诉 rSHUD 函数在哪里找到您的模型输入和输出文件。

```
# 替换为您的实际项目名、输入输出路径和 SHUD 版本
rshud_wd <- "/path/to/your/shud_run_directory" # SHUD 运行的根目录
project_name <- "my_project_on_hpc"

# 使用 rSHUD 的全局配置函数 (如果可用)
# 或者手动构建路径 (如下面的 readout 示例)
# shud.env(projectname = project_name, inpath = file.path(rshud_wd, "input"), outpath
= file.path(rshud_wd, "output"))
```

注: 较新版本的 rSHUD 可能推荐在调用 `readout` 时直接指定完整路径, 而不是依赖全局环境变量。请参考您使用的 rSHUD 版本文档。

### 4. 使用 `readout()` 函数读取输出文件: ([05]--2.2 中 01:25:26 开始演示 `readout` 的使用)

- `readout()` 函数是读取 SHUD 输出文件的核心。它需要指定要读取的变量的关键词和模型单元类型。

```
# 示例: 读取流域出口 (假设出口河段 ID 为 outlet_id) 的向下游流量
# 假设输出文件在 ./output/my_project_on_hpc.out/ 目录下
# 并且项目名为 my_project_on_hpc

# 构建输出文件路径
fn_rivqdown <- file.path(rshud_wd, "output", paste0(project_name, ".out"),
paste0(project_name, ".rivqdown.dat"))

# 读取整个 rivqdown 文件 (所有河段)
rivqdown_all_reaches <- readout(file = fn_rivqdown, keyword = "qdown", model_unit
= "riv")

# 如果只需要特定河段 (例如, 假设出口河段是第 N 个河段, 其 ID 在 .sp.riv 中可以查到)
# outlet_reach_index <- N
# outlet_discharge <- rivqdown_all_reaches[, outlet_reach_index]

# 示例: 读取所有三角形单元的平均地表积水深度
fn_eleysurf <- file.path(rshud_wd, "output", paste0(project_name, ".out"),
paste0(project_name, ".eleysurf.dat"))
eleysurf_all_cells <- readout(file = fn_eleysurf, keyword = "surf", model_unit =
"ele")
```

- 关键词 (`keyword`) 和模型单元 (`model_unit`): 需与第七章输出文件列表中的变量名和模型单元对应。
- 文件路径 (`file`): 推荐直接提供文件的完整路径, 以避免依赖全局工作目录设置。

## 10.4.3. 理解读取后的数据结构

- `readout()` 函数返回的数据通常是:
  - 时间序列对象 (如 `xts`): 如果加载了 `xts` 包, 并且数据包含时间信息, `readout` 会尝试将其转换为 `xts` 对象。 `xts` 对象非常适合进行时间序列分析和绘图。

- **数据框 (data.frame) 或矩阵 (matrix):** 如果不转换为 `xts`, 或者数据本身不适合时间序列, 则可能返回这些结构。
- **数据维度:**
  - 对于空间分布的变量 (如 `.eleysurf`, `.rivqdown`), 返回的数据通常是一个**时间 x 空间单元**的矩阵或数据框, 行代表时间步长, 列代表不同的三角形单元或河段。

## 10.4.4. 实操：初步结果检查与可视化

### 1. 绘制流域出口进流过程线:

```
# 假设 outlet_discharge 是包含出口径流的 xts 对象
# plot(outlet_discharge, main = "Simulated Outlet Discharge", ylab = "Discharge
# (m3/day)")
# 如果 rivqdown_all_reaches 是所有河段的数据, 并且你知道出口河段的列索引 outlet_idx
plot(rivqdown_all_reaches[, outlet_idx], main = "Simulated Outlet Discharge", ylab =
"Discharge (m3/day)")
```

### 2. 与观测数据对比 (如果可用):

- 读取观测径流数据 (也应为 `xts` 对象或类似结构)。
- 确保模拟和观测数据在时间上对齐 (公共时间段, 相同时间步长或进行聚合)。
- 在同一张图上绘制模拟和观测过程线。

```
# 假设 obs_discharge 是观测径流的 xts 对象
# 对齐时间 (示例, 可能需要更复杂的处理)
# common_period <- intersect(index(outlet_discharge), index(obs_discharge))
# sim <- outlet_discharge[common_period]
# obs <- obs_discharge[common_period]
#
# plot(obs, main = "Simulated vs. Observed Discharge", ylab = "Discharge
# (m3/day)", col = "blue")
# lines(sim, col = "red")
# legend("topright", legend = c("Observed", "Simulated"), col = c("blue", "red"),
# lty = 1)
```

**图表建议 10.4:** 插入一张模拟与观测 (如果提供了) 流域出口径流过程线对比图。

**图表建议 10.5:** 插入一张模拟与观测流量频率曲线 (FDC) 对比图。

### 3. 可视化空间输出变量:

- 例如, 绘制某个时刻或平均时段的地下水水位空间分布。
- 这通常需要将 `readout` 读取的矩阵数据与模型网格的 Shapefile (`domain.shp`) 关联起来 (通过单元 ID), 然后在 GIS 软件 (如 QGIS) 中或使用 R 的空间绘图包 (如 `ggplot2` + `sf`) 进行可视化。

```
# 示例概念：计算平均地下水水位
# fn_eleygw <- file.path(rshud_wd, "output", paste0(project_name, ".out"),
# paste0(project_name, ".eleygw.dat"))
# eleygw_data <- readout(file = fn_eleygw, keyword = "gw", model_unit = "ele")
# mean_gw_levels <- colMeans(eleygw_data, na.rm = TRUE) # 计算每个单元的平均地下水水位

# 后续需要将 mean_gw_levels 与网格的 Shapefile 属性表合并，并用 GIS 软件或 R 绘图
```

**图表建议 10.6:** 插入一张模拟地下水水位或实际蒸散发 (ETa) 在某个代表性时刻或平均时段的空间分布图。

#### 4. 计算并检查水量平衡要素:

- **主要组分:** 降雨 (P), 蒸散发 (ET), 径流 (Q), 储量变化 ( $\Delta S$ )。
- **平衡方程 (概念):**  $P - ET - Q = \Delta S$
- **计算:**
  - 总降雨量: 对 `.elevprcp` (降雨速率) 进行空间积分 (乘以单元面积再求和) 和时间积分。
  - 总蒸散发量: 对 `.eleveta` (实际蒸散发速率) 进行空间积分和时间积分。
  - 总径流量: 对流域出口的 `.rivqdown` (向下游流量) 进行时间积分。
  - 总储量变化: 计算模拟期初和期末的总储水量 (冠层水、雪水、地表水、未饱和带水、饱和带地下水) 之差。
- **检查:** 验证水量平衡方程是否近似成立。由于模型简化和数值误差，完全精确的平衡可能难以达到，但应在一个可接受的误差范围内。

**图表建议 10.7:** 插入一张流域水量平衡各组分 (P, ET, Q,  $\Delta S$ ) 的累积量随时间变化的曲线图，或年均水量平衡柱状图。

#### 小结:

第十章详细介绍了 SHUD 模型模拟运行的准备工作、在不同环境下的运行方法、模型预热的重要性与操作步骤，以及如何使用 rSHUD 对输出结果进行初步的读取、可视化和分析。这些步骤构成了从模型构建到初步结果评估的完整流程。通过初步分析，可以对模型的行为有一个基本了解，并为后续的参数率定和详细研究奠定基础。

#### 思考题:

1. 在进行模型预热时，如何判断预热期是否足够长？除了观察地下水储量，还有哪些指标可以用来评估模型的准稳定状态？
2. 如果模型运行日志中出现大量的 CVODE 求解器警告 (例如 `CVODE_TOO_MUCH_WORK` 或 `CVODE_ERR_FAILURE`)，可能是什么原因？如何调整 `.cfg.para` 中的求解器参数来尝试解决？
3. 在初步分析输出结果时，如果发现模拟的洪峰时间与观测相比有明显提前或滞后，这可能暗示了模型中哪些参数或过程描述存在问题？

## 11. SHUD 模型参数率定

在前面的章节中，我们已经学习了如何构建 SHUD 模型并运行初步的模拟。然而，由于模型参数的不确定性和模型的简化，直接使用默认或基于经验的参数往往难以准确地再现观测到的水文过程。因此，**参数率定 (Parameter Calibration)** 成为了模型应用中至关重要的一步。本章将详细介绍参数率定的理论基础、常用方法、配置过程以及结果解读。

## 11.1. 参数率定理论与实践

### 11.1.1. 参数率定的概念与必要性

- **概念:** 参数率定是一个通过系统地调整模型参数，使得模型模拟结果与实际观测数据 (如流域出口径流) 之间的差异最小化的过程。
- **必要性:**
  - **参数不确定性:** 即使是物理模型，其参数的获取也存在不确定性。实测参数通常是点尺度的，难以代表模型计算单元的块尺度有效参数；通过 PTF 等经验方法推导的参数本身也带有误差。
  - **模型结构不确定性:** 任何模型都是对真实物理过程的简化，这种简化必然导致模型结构与现实存在差异。
  - **输入数据不确定性:** 降雨、气温等气象驱动数据，以及 DEM、土壤、土地利用等空间数据都存在测量和插值误差。
  - 参数率定在一定程度上可以补偿这些不确定性，使得模型能够更好地拟合观测数据。

### 11.1.2. 参数的“代表性”与率定

- 对于物理模型，我们期望参数具有明确的物理意义。但在率定过程中，为了使模拟结果更好地匹配观测，有时得到的“最优”参数值可能会偏离其实际物理范围。
- 此时，率定得到的参数更多地被视为一种**有效参数 (Effective Parameter)** 或 **代表性参数 (Representative Parameter)**，它代表了在当前模型结构和数据条件下，能够最好地再现流域整体响应的参数组合。
- 理解这一点有助于我们客观看待率定结果，并避免对参数的物理意义进行过度解读。

### 11.1.3. 率定的目标

- **主要目标:** 找到一组参数组合，使得模型在**校准期 (Calibration Period)** 内，模拟结果与观测数据之间的某种（或多种）拟合优度指标达到最优。
- **次要目标 (同样重要):** 确保模型在独立的**验证期 (Validation Period)** 内，仍能保持较好的模拟性能，以检验模型的泛化能力和参数的稳定性。

### 11.1.4. 率定的挑战

- **异参同效 (Equifinality):** 在复杂的非线性模型中，可能存在多组完全不同的参数组合，它们都能产生相似的、同样“好”的模拟结果。这使得确定唯一的“真实”参数集变得困难。
- **参数敏感性 (Parameter Sensitivity):** 模型输出对不同参数的敏感程度不同。某些参数的微小变动可能引起输出的显著变化，而另一些参数则影响甚微。识别敏感参数是高效率定的前提。
- **多指标冲突:** 不同的拟合优度指标可能关注模拟结果的不同方面 (如洪峰大小、总量平衡、低水过程等)。优化一个指标可能会导致另一个指标变差。
- **计算量大:** 分布式水文模型通常参数众多，且单次模拟耗时较长。参数率定过程需要运行模型成百上千次，对计算资源要求很高，通常需要在 HPC 环境下进行。

- **观测数据质量:** 观测数据的误差和不确定性会直接影响率定结果的可靠性。

**图表建议 11.1:** 插入参数率定过程示意图 (流程图, 展示从选择参数、设定范围、运行模型、计算拟合优度、调整参数到收敛的迭代过程)。

**图表建议 11.2:** 插入异参同效示意图 (例如, 展示两条不同的径流过程线, 它们由不同的参数集生成, 但都与观测数据拟合得较好)。

## 11.2. 常用的拟合优度指标 (复习与重点强调)

拟合优度指标 (Goodness-of-Fit, GoF) 是量化模型模拟结果与观测数据之间吻合程度的数学工具。在 SHUD 训练营中, 常用的指标包括:

### 11.2.1. 重要指标及其特点

(以下内容与第九部分 2.1 节有重叠, 此处为复习和在率定上下文中的强调)

- **纳什效率系数 (Nash-Sutcliffe Efficiency, NSE):**
  - **公式:**  $NSE = 1 - \frac{\sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^N (Q_{obs,i} - \overline{Q_{obs}})^2}$
  - **解读:** 越接近 1 越好。\$NSE > 0.5\$ 通常被认为是可接受的。对高流量和峰值敏感。
- **克林瓦鲁德效率系数 (Kling-Gupta Efficiency, KGE):**
  - **公式:**  $KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}$ , 其中 \$r\$ 是相关系数, \$\beta = \overline{Q\_{sim}} / \overline{Q\_{obs}}\$ 是偏倚比, \$\gamma = (CV\_{sim} / CV\_{obs})\$ 是变异系数比。
  - **解读:** 越接近 1 越好。综合考虑了相关性、偏倚和变异性。
- **确定系数 (\$R^2\$):**
  - **公式:**  $R^2 = \left( \frac{\sum_{i=1}^N (Q_{obs,i} - \overline{Q_{obs}})(Q_{sim,i} - \overline{Q_{sim}})}{\sqrt{\sum_{i=1}^N (Q_{obs,i} - \overline{Q_{obs}})^2} \sqrt{\sum_{i=1}^N (Q_{sim,i} - \overline{Q_{sim}})^2}} \right)^2$
  - **解读:** 0 到 1 之间, 越接近 1 表示线性相关性越好。不反映量级偏差。
- **误差百分比 (Percentage Bias, PBias):**
  - **公式:**  $PBias = \frac{\sum_{i=1}^N (Q_{sim,i} - Q_{obs,i})}{\sum_{i=1}^N Q_{obs,i}} \times 100\%$
  - **解读:** 越接近 0% 越好。正值表示高估, 负值表示低估。反映总体水量平衡。
- **均方根误差 (Root Mean Square Error, RMSE):**
  - **公式:**  $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})^2}$
  - **解读:** \$\ge 0\$, 越接近 0 越好。单位与模拟变量一致。

### 11.2.2. 指标选择的考量

- **单一指标 vs. 多指标:** 仅依赖单一指标可能导致对模型性能的片面评估。例如, 高 NSE 可能伴随着较大的 PBias。
- **研究目标:** 如果研究侧重于洪水预报, NSE 或峰值误差可能更重要; 如果侧重于水资源总量评估, PBias 可能更关键。
- **数据特性:** 对于存在大量低流量的干旱期数据, NSE 可能表现不佳, 此时 KGE 或对数 NSE 可能更合适。



- **多目标优化:** 在参数率定中, 可以尝试同时优化多个指标, 或者将多个指标加权组合成一个综合目标函数。  
([07]--3 中 02:01:20 演示了将 NSE 和 PBias 的对数进行加权组合)

图表建议 11.3: 插入一个表格, 对比不同拟合优度指标的优点、缺点和主要关注点。

## 11.3. 参数优化算法介绍 (以 CMA-ES 为例)

手动调整大量参数以达到最优拟合是非常低效的。因此, 通常采用自动参数优化算法。

### 11.3.1. 进化算法基本原理

进化算法 (Evolutionary Algorithms, EAs) 是一类模仿生物进化过程的全局优化算法。

- **基本流程:**
  1. **初始化种群 (Population):** 随机生成一组参数组合 (个体/染色体)。
  2. **适应度评估 (Fitness Evaluation):** 对种群中的每个个体 (参数组合), 运行 SHUD 模型, 并根据模拟结果与观测数据计算其适应度值 (通常是拟合优度指标)。
  3. **选择 (Selection):** 根据适应度值, 选择表现较好的个体进入下一代或作为父代。
  4. **遗传操作 (Genetic Operators):**
    - **交叉 (Crossover / Recombination):** 组合两个或多个父代个体的参数, 产生新的子代个体。
    - **变异 (Mutation):** 对子代个体的某些参数进行小范围的随机扰动。
  5. **形成新种群:** 用生成的子代替换部分或全部旧种群。
  6. **终止条件判断:** 如果达到最大迭代代数或适应度满足要求, 则停止; 否则返回步骤 2。

### 11.3.2. CMA-ES (Covariance Matrix Adaptation - Evolution Strategy) 核心思想

CMA-ES 是一种高效的、非梯度的进化策略算法, 特别适用于处理复杂、非线性、非凸的优化问题。([04]--2.1 中的 0:57:59 开始介绍 CMA-ES)

- **核心机制:**
  - **协方差矩阵自适应 (CMA):** CMA-ES 的关键在于它能够学习参数之间的相关性, 并自适应地调整参数采样的协方差矩阵。这意味着算法可以学习到参数空间中“有希望”的搜索方向和形状。
  - **步长控制 (Step-size Control):** 算法还自适应地调整全局的搜索步长。
  - **去随机化:** 通过累积演化路径来减少随机性对搜索方向的影响。
- **优点:**
  - 对初始参数不敏感。
  - 能够处理参数间复杂的相互依赖关系。
  - 在许多具有挑战性的优化问题上表现出色, 收敛速度相对较快。
  - 通过其采样机制, 有助于跳出局部最优解, 探索更广阔参数空间 ([04]--2.1 中 0:58:33 强调其避免局部最优的能力)。



图表建议 11.4: 插入进化算法基本流程图。

图表建议 11.5: 插入 CMA-ES 搜索过程的概念示意图 (展示多维参数空间中, 采样椭球如何随迭代自适应地旋转和缩放, 逼近最优解)。

## 11.4. 参数率定配置 (实操指南)

在 SHUD 系统中, 使用 CMA-ES 进行参数率定主要通过 rSHUD 中的脚本 (`mine.R` 或类似脚本) 调用, 并依赖于以下配置文件:

### 11.4.1. 率定文件组织

- 在您的项目工作目录下, 创建一个专门用于参数率定的子目录 (例如 `calib_run_01`)。
- 将以下文件组织到该目录或其子目录中:
  - **SHUD 模型输入文件:** 从 AutoSHUD 生成的 `deploy/<ProjectName>/MODEL/Input/` 目录完整复制过来, 并放在一个子目录 (例如 `input/`) 中。确保所有 `.sp.*`, `.cfg.ic`, `.para.*`, `.tsd.forc` 以及关联的 `.csv` 气象文件都在。
  - **SHUD 可执行文件:** `shud` 或 `shud_omp`。
  - **率定主程序脚本:** 例如 `mine.R`。
  - **CMA-ES 配置文件:** `.cfg.cmaes.txt`。
  - **参数范围文件:** `.calib.range.txt`。
  - **观测数据文件:** 例如 `observed_flow.rds` 或 `.tsd.obs.txt`。
  - **辅助 R 函数脚本:** 例如 `functions_calib.R`, `objective_function.R`。
  - **Slurm 作业脚本:** 用于在 HPC 上提交率定任务 (例如 `submit_calib.sh`)。

图表建议 11.6: 插入一个典型的参数率定工作目录结构图。

### 11.4.2. 率定主程序 (`mine.R`) 关键配置

`mine.R` (或您自定义的类似脚本) 是驱动整个率定流程的入口。

- 加载必要的 R 包:

```
library(rSHUD)
library(hydroGOF) # 用于计算拟合优度
library(xts)      # 用于处理时间序列数据
library(ggplot2)  # 用于绘图
library(doParallel) # 并行计算
library(doMC)     # 并行计算 (Linux/Mac)
# 可能还有其他自定义函数源文件
source("calib/functions_calib.R")
source("calib/objective_function.R")
```

- 设置工作目录和项目信息:



```
# stopfitness = cfg.cmaes$stopfitness
# # ... 其他 CMA-ES 控制参数 ...
# ))
# 这是一个示意性调用, 实际 rSHUD 中可能有封装好的函数如 rSHUD::shud_calib_cmaes()
```

- 停止并行集群:

```
parallel::stopCluster(cl)
```

- 保存结果。

### 11.4.3. 实操：配置率定算法参数文件 (.cfg.cmaes.txt)

此文件直接控制 CMA-ES 算法的行为。

- 结构: 纯文本, 每行 参数名 值。
- 关键参数: ([07]--3 中 01:38:02 开始演示了部分参数修改)
  - `lambda`: 种群大小 (每代个体数)。应小于等于您在 Slurm 脚本中申请的 CPU 核心数。例如, `lambda 120`。
  - `stopfitness`: 适应度停止阈值。例如, 如果目标是最大化 NSE (最优为 1), 且希望在 NSE 达到 0.75 时停止, 那么此值可以设为 `1 - 0.75 = 0.25` (因为 CMA-ES 通常最小化目标函数)。具体如何转换为算法的停止条件, 需查阅 `mine.R` 或 `rSHUD` 中 CMA-ES 接口的实现。
  - `maxgen`: 最大迭代代数。例如, `maxgen 20`。
  - `sigma`: 初始全局步长/标准差的缩放因子。例如, `sigma 0.5`。
  - `updateic`: 是否在每次迭代更新子代模型的初始条件 (通常为 `0`, 预热独立进行)。
  - `walltime`: 单个 SHUD 模型模拟允许的最大运行时长 (秒)。例如, `walltime 7200` (2小时)。
  - `nspingup`: 用于计算 GOF 时, 跳过的初始数据点数量 (预热期)。例如, 如果日输出, 预热 365 天, 则 `nspingup 365`。
  - `minlength`: 用于计算 GOF 的最小有效数据对数量。例如, `minlength 300`。

### 11.4.4. 实操：配置可率定参数范围文件 (.calib.range.txt)

此文件指定哪些 SHUD 参数参与率定, 以及它们的取值范围和采样尺度。([07]--3 中 01:20:18 开始详细解读此文件内容和格式)

- 结构: 纯文本, 每行定义一个参数的率定规则。
- 列定义:
  1. 参数名: 与 `.cfg.calib` 文件中的参数名完全一致。
  2. On/Off (0 或 1): `1` 表示该参数参与率定; `0` 表示不参与 (使用 `.cfg.calib` 中的默认值)。
  3. Log Scale (0 或 1): `1` 表示在对数空间搜索该参数; `0` 表示在线性空间搜索。
  4. Min Value: 参数的最小值 (在线性或对数空间, 取决于上一列)。
  5. Max Value: 参数的最大值。

- 示例行:

```
GEOL_KSATH 1 1 0.001 10 # 地质层水平导水率, 对数尺度, 范围 0.001*基准值 到 10*基准值
SOIL_KINF 1 1 0.01 100 # 土壤渗透系数, 对数尺度
LC_ROUGH 1 0 0.5 2.0 # 土地利用曼宁糙率, 线性尺度, 范围 0.5*基准值 到 2.0*基准值
AQ_DEPTH+ 0 0 -2 5 # 含水层厚度调整量 (不参与率定)
```

- 设置原则:

- **选择敏感参数:** 优先选择对模型输出 (如径流) 影响显著的参数进行率定。
- **合理范围:** 参数范围应基于其物理意义、文献经验值以及对研究区域的先验知识来设定, 避免过宽或过窄。
- **Log Scale 选择:** 对于变化范围跨越多个数量级的参数 (如导水率), 使用对数尺度通常更有效。

## 11.4.5. 准备观测数据

- **数据格式:** 确保观测数据 (如径流) 已处理成与模型输出频率一致的时间序列 (通常是 `.rds` 格式的 `xts` 对象, 或 `.tsd.obs.txt` 格式)。
- **单位一致:** 确保观测数据的单位与模型输出单位一致, 或在目标函数中进行转换。
- **时间对齐:** 确保观测数据的时间戳与模型模拟的时间戳能够正确对齐, 以便进行点对点的比较。

## 11.5. 提交与监控参数率定任务 (实操指南)

参数率定任务通常计算量较大, 推荐在 HPC 集群上通过 Slurm 等作业调度系统提交和运行。

### 11.5.1. 理解率定任务的执行方式

- **主从结构:** 率定主程序 (`mine.R`) 作为主进程, 在 HPC 的一个计算节点上运行。
- **并行模拟:** `mine.R` 内部会根据 CMA-ES 算法的配置 (如种群大小 `lambda`), 生成多组参数试验。对于每一组参数, 它会启动一个独立的 SHUD 模型模拟子任务。
- **资源调度:** 这些 SHUD 子任务可以利用 `doParallel` 或 `doMC` 等 R 包在申请到的 CPU 核心上并行执行。
- **迭代优化:** 每个子任务模拟完成后, `mine.R` 收集其输出, 计算拟合优度, 然后 CMA-ES 算法根据所有个体的表现生成下一代的参数种群, 重复此过程。

### 11.5.2. 编写率定任务 Slurm 脚本 (`.sh` 文件)

您需要一个 Slurm 作业脚本来向 HPC 系统申请资源并启动您的率定主程序 (`mine.R`)。([07]--3 中 01:46:46 详细讲解了 Slurm 脚本的编写)

- **脚本示例 (`submit_calibration.sh`):**

```
#!/binbash
#SBATCH --job-name=shud_calib_arno # 任务名称 (可自定义, 例如包含您的用户名和项目名)
#SBATCH --output=slurm_calib_%j.out # 标准输出日志文件 (%j 会被作业ID替换)
#SBATCH --error=slurm_calib_%j.err # 标准错误日志文件
#SBATCH --partition=your_queue # 指定作业提交的队列/分区 (需根据实际HPC配置填写)
#SBATCH --nodes=1 # 率定主程序通常在单个节点上运行
```

```

#SBATCH --ntasks-per-node=1          # 每个节点运行一个率定主任务
#SBATCH --cpus-per-task=40           # **核心数，应等于或大于 .cfg.cmaes 中的 lambda 值**
#SBATCH --mem=32G                    # 申请的内存量（例如 32GB，根据模型规模和 lambda 调整）
#SBATCH --time=24:00:00              # 任务最大运行时长（例如 24 小时）

# 设置工作目录（替换为您的实际率定工作目录在HPC上的绝对路径）
WORK_DIR="/scratch/your_username/arno_basin/calibration_run_01"
cd ${WORK_DIR}

# 加载 R 模块（如果需要，根据 HPC 环境配置）
# module load R/4.2.0
# module load gcc/9.3.0 # 可能需要 R 包编译
# module load gdal/3.x.x proj/x.x # 如果 rSHUD 或其依赖包需要

echo "Starting SHUD calibration job: ${SLURM_JOB_NAME} (ID: ${SLURM_JOB_ID})"
echo "Working directory: $(pwd)"
echo "Number of CPUs allocated: ${SLURM_CPUS_PER_TASK}"

# 执行率定主程序 mine.R，并将 R 的输出重定向到日志文件
# 确保 mine.R 脚本具有可执行权限，或者使用 Rscript 调用
Rscript mine.R > mine_R_output.log 2>&1

echo "SHUD calibration job finished."

```

#### • 关键 Slurm 参数:

- `--job-name`: 方便您在队列中识别任务。
- `--cpus-per-task`: **非常重要!** 该值应等于或大于您在 `.cfg.cmaes.txt` 中设置的 `lambda` (种群大小)，以确保每个参数组合的模拟都能有核心可用。如果 `lambda` 为 120，而您只申请了 40 个核心，那么同一代中的 120 个模拟任务会分批在这些核心上运行。
- `--mem`: 总内存需求。如果 `lambda` 较大，且每个 SHUD 模拟占用一定内存，需要估算总内存。
- `--time`: 估算整个率定过程 (所有代) 完成所需的时间。

- **Rscript 调用:** 使用 `Rscript mine.R > mine_R_output.log 2>&1` 可以将 `mine.R` 脚本中通过 `print()` 或 `cat()` 输出的信息以及任何错误信息都重定向到 `mine_R_output.log` 文件中，便于后续查看。

### 11.5.3. 实操：使用 `sbatch` 命令提交率定任务

1. **上传文件:** 确保您的率定工作目录 (包含所有 SHUD 输入、可执行文件、`mine.R`, `.cfg.cmaes.txt`, `.calib.range.txt`, 观测数据文件以及 `submit_calibration.sh` 脚本) 已完整上传到 HPC 的 Scratch 空间。
2. **登录 HPC。**
3. **进入工作目录:** `cd /scratch/your_username/arno_basin/calibration_run_01`
4. **提交作业:**

```
sbatch submit_calibration.sh
```

如果提交成功，系统会返回一个作业 ID (e.g., Submitted batch job 12345)。记下这个 ID。

## 11.5.4. 实操：监控率定过程与日志解读

### 1. 查看作业状态:

```
squeue -u your_username  
# 或者查看特定作业ID的状态  
# squeue -j <job_id>
```

- **PD** (Pending): 任务正在排队等待资源。
- **R** (Running): 任务正在运行。
- **CG** (Completing): 任务即将完成。

### 2. 查看 Slurm 输出日志:

- 在任务运行期间或完成后，查看 `slurm_calib_<job_id>.out` 和 `slurm_calib_<job_id>.err` 文件。
- `.out` 文件通常包含 Slurm 的一些运行信息和您脚本中 `echo` 的内容。
- `.err` 文件包含 Slurm 运行脚本时遇到的系统级错误。

### 3. 查看 `mine.R` 输出日志 (`mine_R_output.log`):

- 这是最重要的日志文件，记录了率定算法的详细过程。([07]--3 中 01:50:40 开始查看此日志)
- 内容解读:
  - **初始化信息:** 加载包、读取配置等。
  - **各代信息 (Generation X):**
    - 通常会打印当前是第几代。
    - **参数采样:** 可能会打印出当前代生成的参数组合 (取决于 `mine.R` 的详细程度)。
    - **SHUD 模拟启动信息:** 对于每个参数组合，调用 SHUD 模型运行时，可能会打印启动命令或提示。
    - **拟合优度计算:** 每个 SHUD 子任务模拟完成后，会计算其拟合优度。日志中可能会打印每个个体的 GOF 值。
    - **本代最优/平均 GOF:** 总结当前这一代所有个体的 GOF 表现。
    - **CMA-ES 内部状态:** 可能打印算法的步长、协方差矩阵等信息。
  - **错误与警告:** 如果某个 SHUD 子任务运行失败 (例如，模型崩溃、输出文件缺失、模拟时长超过 `walltime`)，`mine.R` 通常会捕获这个错误，并为该参数组合赋予一个极差的拟合优度值 (如 `-9999`)。日志中应记录这些失败信息。([07]--3 中 01:52:32 演示了因文件缺失导致返回 `-9999`)
  - **收敛信息:** 当达到最大代数或 `stopfitness` 条件满足时，算法会停止，并报告最终结果。

## 11.5.5. 常见问题

### • Slurm 作业无法启动 (长时间 **PD** 或直接 **F**):

- 资源申请不合理 (核心数、内存、时长超过队列限制)。
- 脚本路径错误或权限问题。

- HPC 系统问题。
- **mine.R 脚本执行错误** (查看 `mine_R_output.log` 和 `slurm_calib_<job_id>.err`):
  - R 包未正确加载。
  - 文件路径错误 (找不到输入文件、观测数据或配置文件)。
  - `ObjectiveFunction` 内部错误 (例如, SHUD 子任务调用失败, 拟合优度计算出错)。
- **大量 SHUD 子任务失败:**
  - **参数范围不合理:** 生成的参数组合导致 SHUD 模型运行不稳定或崩溃。检查 `.calib.range.txt`。
  - **SHUD 模型本身的问题:** 确保 SHUD 模型在单个参数组合下能够稳定运行。
  - **初始条件问题:** 如果未进行充分预热, 不稳定的初始条件可能导致许多模拟失败。
  - **计算资源不足:** 单个 SHUD 模拟所需的内存或时间超出了隐性限制。

## 11.6. 参数率定结果解读

率定任务完成后, 需要分析 `CMAES_OUT/` 目录下的结果文件。

### 11.6.1. 率定结果文件组织

(回顾第十部分 11.4.1 节关于此目录的介绍)

- `CMAES_OUT/`
  - `gov.csv`: 各代最优 (通常是前5个) 个体的拟合优度值。
  - `gof_gen<X>.csv`: 第 X 代所有个体的拟合优度值。
  - `parameters_gen<X>.csv`: 第 X 代所有个体的参数值。
  - `best_solution.csv` (或类似名称): 整个率定过程中的最优参数集及其对应的拟合优度。
  - `fig/`: 可能包含各代参数分布、GOF演化等图件。
  - `configure_calib/`: 可能包含各代生成的 `.cfg.calib` 文件副本。

### 11.6.2. 实操：读取率定结果文件

使用 R 脚本读取 `CMAES_OUT/` 目录下的关键 `.csv` 文件。

```
# 假设工作目录已设置为包含 CMAES_OUT 的目录
# 或者提供完整路径
gof_summary <- read.csv("CMAES_OUT/gov.csv")
best_solution_params <- read.csv("CMAES_OUT/best_solution.csv") # 文件名可能不同

# 查看数据结构
# print(head(gof_summary))
# print(best_solution_params)
```

### 11.6.3. 分析率定过程与收敛性

- 绘制拟合优度随迭代代数的变化:



- 从 `gov.csv` 或 `gof_gen<X>.csv` 文件中提取每代的最优（或平均）拟合优度值。
- 绘制拟合优度指标 (如 NSE, KGE, -RMSE, -PBias\_abs) 随迭代代数 (Generation) 变化的曲线。
- **观察:** 曲线是否逐渐趋于平稳, 表明算法已收敛或接近收敛。如果曲线仍在显著改善, 可能需要增加 `maxgen`。如果曲线波动剧烈或过早停滞, 可能算法陷入局部最优或参数范围不当。([07]--3 中 02:00:24 开始查看并讨论 `gov.csv` 的内容和收敛情况)

- **绘制关键参数值随迭代代数的变化:**

- 从 `parameters_gen<X>.csv` 或 `best_solution.csv` (如果它记录了各代最优) 中提取关键参数的值。
- 绘制这些参数值随迭代代数变化的曲线。
- **观察:** 参数是否收敛到某个特定值或在一个较小范围内波动。如果参数持续大幅度波动, 可能表明该参数不敏感或模型对该参数的识别能力较弱。

**图表建议 11.7:** 插入一张展示某关键拟合优度指标 (如 NSE) 随迭代代数变化的曲线图, 清晰显示收敛趋势。

**图表建议 11.8:** 插入一张或几张展示几个关键参数 (如某个导水率、糙率) 的值随迭代代数变化的曲线图。

## 11.6.4. 分析最终最优参数集

- **检查参数值的物理合理性:**

- 查看 `best_solution.csv` 中最优参数组合的值。
- 判断这些值是否在其物理意义的可接受范围内。例如, 孔隙度不应大于 1, 水力传导度不应为负数等。
- 如果最优参数值非常极端或不符合常识, 即使拟合优度很高, 也需要警惕, 可能模型结构存在问题或率定过程有偏差。

- **与先验知识对比:** 将最优参数值与文献报道的类似流域或土壤/植被类型的参数值进行比较。

## 11.7. 基于最优参数重新运行模型

在获得一组满意的最优参数后, 需要使用这组参数重新运行模型, 以生成用于校准期和验证期详细分析的模拟结果。

### 11.7.1. 实操：将最优参数写入模型输入文件

1. **获取最优参数:** 从 `CMAES_OUT/best_solution.csv` (或您选定的其他最优参数文件) 中提取最优参数值。
2. **更新 `.cfg.calib` 文件:**
  - 创建一个新的 SHUD 运行目录, 或在现有运行目录中操作。
  - 将 SHUD 原始输入文件 (包括一个包含默认乘子/加数的 `.cfg.calib` 文件) 复制到此目录。
  - **手动或通过脚本**将最优参数值更新到这个 `.cfg.calib` 文件中对应的参数行。确保参数名称匹配, 并且乘性/加性关系正确。

### 11.7.2. 实操：运行模型进行校准期和验证期模拟

1. **配置 `.cfg.para` 文件:**
  - 设置 `START` 和 `END` 以覆盖整个校准期和验证期。
  - 确保 `INIT_MODE = 3` (使用预热后的 `.cfg.ic` 文件)。

- 设置合理的输出频率，以便详细分析。
- 2. 确保使用更新后的 `.cfg.calib` 和预热后的 `.cfg.ic`。
- 3. 运行 SHUD 模型。
- 4. 结果分析:
  - 分别提取校准期和验证期的模拟结果与观测数据。
  - 计算并比较两个时期的拟合优度指标。
  - 绘制过程线、FDC 曲线、散点图等进行详细对比。([07]--3 中 02:08:20 开始展示最终率定结果的图件)
  - 评估模型在验证期的表现，判断参数的稳定性和模型的泛化能力。如果验证期结果显著差于校准期，可能存在过拟合问题。

### 11.7.3. 常见问题

- **最优参数导致模型运行失败:** 有时，优化算法找到的“数学最优”参数组合在物理上可能导致模型不稳定。需要检查参数值是否极端，或微调参数。
- **验证期效果不佳 (过拟合 Overfitting):** 模型在校准期表现很好，但在验证期表现明显下降。
  - **原因:** 参数过多、校准期数据代表性不足、模型过于复杂等。
  - **对策:** 减少率定参数数量 (选择最敏感的)、使用更长的或更多样化的校准期数据、简化模型结构 (如果可能)。

---

#### 小结:

第十一章详细介绍了 SHUD 模型参数率定的实践操作，包括如何通过 Slurm 提交和监控 HPC 上的率定任务，如何解读率定过程的日志和输出结果文件，以及如何分析率定过程的收敛性和最终得到的最优参数集。最后，强调了使用最优参数重新运行模型以进行校准期和验证期评估的重要性。成功的参数率定是模型能够可靠应用于实际研究和预测的关键。

#### 思考题:

1. 如果您的参数率定任务在 HPC 上因超出申请的运行时长 (`--time`) 而被终止，但从日志看算法尚未完全收敛，您应该如何操作以继续率定过程？
2. 在分析率定结果时，如果发现多个参数在迭代过程中并未收敛到稳定值，而是持续在较大范围内波动，这可能说明什么问题？
3. 除了拟合优度指标，您认为还有哪些标准可以用来判断一组率定参数是否“好”？

---

#### 小结:

第十章详细介绍了进行 SHUD 模型参数率定前的各项配置工作，包括组织率定文件、配置率定主程序 (`mine.R`) 的关键部分、设置 CMA-ES 算法参数 (`.cfg.cmaes.txt`)、定义可率定参数及其范围 (`.calib.range.txt`)，以及准备用于比较的观测数据。正确的配置是成功进行自动化参数率定的基础。

#### 思考题:

1. 在 `.cfg.cmaes.txt` 中，`lambda` (种群大小) 和 `maxgen` (最大代数) 这两个参数如何共同影响率定的总计算量和找到最优解的可能性？

2. 为参数设置率定范围时，如果范围设置过窄可能会有什么问题？如果范围设置过宽又可能会有什么问题？
3. 如果您的观测径流数据是日平均流量 ( $m^3/s$ )，而 SHUD 模型输出的河道流量 (`.rivqdown`) 是日总量 ( $m^3/day$ )，您应如何在目标函数中处理这种单位不一致？

## 12. 综合案例实战：Arno River Basin 全流程建模

本章将通过一个完整的实例——意大利阿尔诺河流域 (Arno River Basin) 的 SHUD 模型构建、运行、参数率定和结果分析，将前面章节学习到的理论知识和操作步骤串联起来。目标是提供一个从零开始、端到端的 SHUD 建模演练，使您能够独立完成类似的研究项目。

(免责声明: 本案例旨在教学和演示 SHUD 建模流程。所使用的数据源、参数选择和具体结果可能与真实的 Arno 河流域研究存在差异，不应直接用于实际决策。实际研究中，数据获取、处理和模型验证需要更严谨的步骤和本地化知识。)

### 12.1. 流域概况与研究目标

#### 12.1.1. Arno River Basin 地理位置与水文气象特征

- 地理位置:** 阿尔诺河是意大利中部托斯卡纳地区最主要的河流，发源于亚平宁山脉，流经佛罗伦萨和比萨等重要城市，最终注入利古里亚海。
- 流域面积:** 约 8,200 平方公里 (具体模拟区域可能根据数据可获得性和研究目标选取子流域)。
- 地形:** 上游为山区，地势陡峭；中下游逐渐过渡到丘陵和平原。
- 气候:** 典型的地中海气候，夏季炎热干燥，冬季温和湿润，降雨主要集中在秋冬季节。
- 水文特征:** 径流具有显著的季节性变化，秋冬季易发生洪水，夏季则可能出现枯水。流域内有多个水库和人类活动影响。

图表建议 12.1: 插入 Arno River Basin 地理位置图 (在意大利地图上标出流域范围)。

图表建议 12.2: 插入 Arno River Basin DEM 图，叠加主要河网和重要城市位置。

#### 12.1.2. 本案例建模与分析目标

- 目标1 (建模与运行):** 利用全球开放数据源，构建 Arno River Basin (或其代表性子流域) 的 SHUD 水文模型，并成功运行。
- 目标2 (参数率定):** 选取一个或多个历史洪水事件或一段连续时期，利用流域出口 (例如佛罗伦萨或比萨附近) 的实测径流数据，对 SHUD 模型的关键参数进行率定。
- 目标3 (结果分析与可视化):** 评估模型在校准期和验证期的表现，分析关键水文变量 (如径流、地下水、蒸散发) 的时空动态，并进行水量平衡分析。
- 目标4 (流程演示):** 完整演示从数据准备到结果输出的 SHUD 建模全流程。

### 12.2. 数据收集与预处理 (实操)

#### 12.2.1. 确定并获取所需数据源

数据类型	建议数据源 (示例)	分辨率/备注
流域边界 (WBD)	HydroSHEDS, Merit Hydro, 或本地提供的 Shapefile	根据研究尺度选择
河流网络 (RIV)	HydroSHEDS, Merit Hydro, 或本地提供的 Shapefile	与流域边界匹配
高程 (DEM)	SRTM 30m/90m, ASTER GDEM 30m, Merit DEM	30m 或 90m
土地利用 (LC)	MODIS Land Cover (MCD12Q1), GlobeLand30, CORINE (欧洲)	500m, 30m, 或更高 (CORINE)
土壤 (SOIL)	ISRIC SoilGrids, HWSD, ESDB (欧洲土壤数据库)	250m/1km
气象驱动	ERA5, GLDAS, 或欧洲区域性再分析资料 (如 E-OBS)	0.1° - 0.25°, 3小时/日
观测径流	GRDC (Global Runoff Data Centre), 本地水文机构数据	日平均流量 ( $m^3/s$ )

- **实操步骤 (概念性):**
  1. **定义研究区域:** 明确 Arno 河流域的具体模拟范围 (例如, 以某个水文站控制的上游流域为界)。
  2. **数据下载:** 从上述数据源或其他可靠途径下载所需数据。
    - 对于全球数据, 可能需要注册账户并使用专门的下载工具或脚本。
    - **GHDC 平台:** 如果平台支持该区域且数据源符合要求, 可优先使用 GHDC 获取初步数据。
  3. **数据整理:** 将下载的数据按类型存放到项目工作目录的不同子文件夹中。

### 12.2.2. 利用 rSHUD/AutoSHUD 进行数据裁剪、投影转换和格式化

- **实操步骤 (以 AutoSHUD 为例):**
  1. **准备 `.autoshoot.txt` 配置文件:**
    - 设置 `ProjectName` 为 `ArnoBasin` (或类似)。
    - 指定 `WBD`, `RIV`, `DEM` 的本地文件路径 (如果已下载)。
    - 选择合适的气象驱动 (`FORCING_DATA` 和 `FORCING_PATH`), 例如选择服务器上已有的 ERA5 或 GLDAS 数据, 并指定路径。
    - 选择土壤 (`SOIL_DATA`) 和土地利用 (`LC_DATA`) 数据源及路径。
    - 设置合适的网格参数 (`NCELL`, `MAX_AREA`, `MIN_ANGLE` 等)。
    - 设置模拟时间范围 (`StartYear`, `EndYear`), 确保覆盖可用的观测径流时段。
  2. **运行 AutoSHUD 脚本:**
    - 执行 `GetReady.R`。
    - 执行 `Step1_RawDataProcessng.R` (处理用户提供的 WBD, RIV, DEM)。
    - 执行 `Step2_DataSubset.R` (处理气象、土壤、土地利用数据)。
  3. **检查中间输出:** 查看 `DataPRE/` 和 `Forcing/` 目录下的输出文件, 确保数据已按流域范围正确处理。

### 12.2.3. 数据质量检查与初步分析

- 空间数据:
  - 在 QGIS 中加载处理后的 DEM、河网、流域边界、土壤图、土地利用图，检查其空间范围、投影是否一致，是否存在明显的错误或伪影。
  - 检查河网是否与 DEM 匹配，水流方向是否合理。
- 时间序列数据:
  - 绘制气象驱动数据 (如降雨、气温) 的时间序列图，检查其季节性、年际变化是否符合区域气候特征，是否存在异常值或大量缺测。
  - 绘制观测径流的时间序列图，识别洪水事件、枯水期等，并与降雨过程进行初步对比。

图表建议 12.3: 插入 Arno River Basin 处理后的 DEM、河网、土地利用和土壤类型空间分布图。

图表建议 12.4: 插入 Arno River Basin 代表性站点/格点的降雨和气温时间序列图。

图表建议 12.5: 插入 Arno River Basin 出口观测径流过程线图。

## 12.3. SHUD 模型构建 (实操)

### 12.3.1. 配置 `.autoshoot.txt` 文件 (确认与微调)

- 在完成数据预处理后，再次检查并确认 `.autoshoot.txt` 中的网格生成参数 (`NCELL`, `MAX_AREA`, `MIN_ANGLE`, `TOL_WBD`, `TOL_RIV`)、含水层厚度 (`AQ_DEPTH`)、一级河道参数 (`RIV_WIDTH`, `RIV_DEPTH`) 等设置是否合理。
- 基于流域面积和地形复杂程度，调整网格参数以获得合适的网格密度和质量。例如，对于 Arno 河流域 (约 8200 km<sup>2</sup>)，如果目标分辨率为 1-2 km，则 `NCELL` 可能在几千到一万之间。

### 12.3.2. 运行 AutoSHUD Step3 脚本生成 SHUD 输入文件

- 实操步骤:
  1. 在 R 环境中执行 `Step3_BuidModel.R` 脚本。
  2. 监控脚本运行日志，注意网格生成的统计信息和任何警告或错误。

### 12.3.3. 检查生成的输入文件和模型网格

- 导航到 `deploy/ArnoBasin/MODEL/Input/` 目录，检查 SHUD 输入文件是否完整生成。
- 在 QGIS 中加载 `gis/domain.shp` 和 `gis/river.shp`，检查最终生成的模型网格和河网的质量、拓扑关系以及与流域边界的匹配情况。

图表建议 12.6: 插入 Arno River Basin 最终生成的 SHUD 模型三角形网格图 (可叠加河网和流域边界)。

图表建议 12.7: 插入 Arno River Basin 三角形单元面积分布直方图和最小角度分布直方图 (评估网格质量)。

## 12.4. SHUD 模型模拟运行与预热 (实操)

### 12.4.1. 配置 `.cfg.para` 和 `.cfg.ic` 文件

1. 复制输入文件到运行目录: 将 `deploy/ArnoBasin/MODEL/Input/` 下的所有文件和 `deploy/ArnoBasin/Forcing/` 下的 `.csv` 文件复制到 HPC 上的一个新的运行目录 (例如 `/scratch/your_username/arno_basin/run_spinup`)。
2. 初始 `.cfg.ic`: 使用 AutoSHUD 生成的默认 `.cfg.ic` 文件。
3. 配置 `.cfg.para` 进行预热:
  - `START = 0`
  - `END = <预热期总天数>` (例如, 选择 5-10 年作为初步预热期,  $5 \times 365 = 1825$  天)。
  - `INIT_MODE = 0` 或 `2` (使用地形猜测或默认猜测的初始条件)。
  - 设置输出频率, 确保在预热结束时输出模型状态 (例如, `dt_ye_gw = 1440*1825` 只在最后一天输出, 或者设置较小的 `SCR_INTV` 并监控屏幕输出的储量变化)。通常, 预热阶段可以关闭大部分详细输出以节省空间。

## 12.4.2. 运行预热模拟并更新初始条件

1. 编写并提交 Slurm 作业脚本 运行预热模拟。
2. 监控模拟过程。
3. 预热模拟完成后, 找到输出目录中对应于预热结束时刻的 `.update_[TotalMinutes].dat` 文件。
4. 将其复制到 `input/` 目录并重命名为 `ArnoBasin.cfg.ic.dat`, 覆盖旧的初始条件文件。

## 12.4.3. 运行正式模拟 (校准/验证期)

1. 配置 `.cfg.para` 进行正式模拟:
  - `START = <预热期结束后的第一天>` (例如 `1825`, 如果预热了 1825 天)。
  - `END = <整个模拟期结束的天数>` (例如, 预热 5 年, 校准 5 年, 验证 5 年, 则 `END` 可能为  $15 \times 365$ )。
  - `INIT_MODE = 3` (使用更新后的 `.cfg.ic` 文件)。
  - 设置合理的输出频率, 以便后续分析 (例如, `dt_Qr_down = 1440` 输出日径流)。
2. 编写并提交 Slurm 作业脚本 运行正式模拟。

## 12.5. 参数率定 (实操)

假设我们选择一段包含典型洪水事件和枯水期的时段 (例如 2-3 年) 作为校准期。

### 12.5.1. 准备观测数据并配置率定文件

1. 观测径流数据:
  - 获取 Arno 河流域出口 (如佛罗伦萨站) 对应校准期的日平均径流数据。
  - 将其处理为 `xts` 对象并保存为 `.rds` 文件 (例如 `calib/arno_obs_flow_daily.rds`), 或符合 `.tsd.obs.txt` 格式的文本文件。确保单位为  $\text{m}^3/\text{s}$  (如果模型输出也是这个单位, 或者在目标函数中统一)。
2. 配置 `.cfg.cmaes.txt`:



- `lambda`: 根据可用核心数设置 (如 40 或 80)。
- `maxgen`: 设置最大迭代代数 (如 20-50)。
- `stopfitness`: 根据选择的目标函数设置 (如 NSE 目标为 1, 则设为  $1 - 0.75 = 0.25$ )。
- `nspingup`: 设置与校准期开始对应的预热数据点数 (如果校准期紧随预热期, 则为0; 如果单独运行校准期, 则需要包含校准期之前的预热数据点数)。
- `minlength`: 校准期有效数据长度 (例如, 一年日数据为 360)。
- `walltime`: 根据单次模拟校准期所需时间估算。

### 3. 配置 `.calib.range.txt`:

- 选择关键的敏感参数进行率定 (如土壤渗透系数、地质层导水率、曼宁糙率、部分土地利用参数如根系深度、截留参数等)。
- 为选定的参数设置合理的取值范围和采样尺度 (线性或对数)。

## 12.5.2. 提交并监控 HPC 上的率定任务

### 1. 修改 `mine.R` 脚本:

- 更新项目名、输入/输出路径、观测数据路径。
- 确保目标函数 (`ObjectiveFunction`) 正确读取模拟输出并与观测数据进行比较, 计算所选的拟合优度指标。

### 2. 编写并提交 Slurm 作业脚本 (`submit_calib_arno.sh`) 来运行 `mine.R`。

### 3. 监控任务状态和日志文件。

## 12.5.3. 分析率定结果, 确定最优参数集

- 查看 `CMAES_OUT/` 目录下的结果。
- 绘制拟合优度随迭代代数的变化曲线, 判断收敛情况。
- 查看 `best_solution.csv` (或类似文件) 获取最优参数组合。
- 检查最优参数的物理合理性。

图表建议 12.8: 插入 Arno River Basin 参数率定过程中某一拟合优度指标 (如 NSE) 随迭代代数的变化曲线。

图表建议 12.9: 插入 Arno River Basin 率定得到的最优参数与默认参数的对比表或图。

## 12.6. 结果可视化与分析 (实操)

### 12.6.1. 基于最优参数进行校准期和验证期模拟

1. 将最优参数更新到 `.cfg.calib` 文件中。
2. 配置 `.cfg.para` 分别运行覆盖校准期和验证期的模拟 (使用相同的预热后初始条件)。

### 12.6.2. 绘制模拟与观测径流对比图

- 使用 rSHUD 或其他绘图工具, 绘制校准期和验证期的模拟径流与观测径流的过程线、FDC 曲线和散点图。



图表建议 12.10: 插入 Arno River Basin 校准期模拟与观测径流过程线对比图。

图表建议 12.11: 插入 Arno River Basin 验证期模拟与观测径流过程线对比图。

图表建议 12.12: 插入 Arno River Basin 校准期和验证期的 FDC 曲线对比图。

图表建议 12.13: 插入 Arno River Basin 校准期和验证期的模拟与观测径流散点图 (含 1:1 线和  $R^2$ )。

### 12.6.3. 计算并展示拟合优度指标

- 分别计算校准期和验证期的 NSE, KGE, PBias, RMSE 等指标。

图表建议 12.14: 插入一个表格, 汇总 Arno River Basin 校准期和验证期的各项拟合优度指标值。

### 12.6.4. 可视化关键水文变量的空间分布

- 选择几个代表性时刻 (如洪峰期、枯水期) 或时段平均, 利用 rSHUD 和 GIS 工具可视化地下水水位、实际蒸散发、土壤湿度等的空间分布。

图表建议 12.15: 插入 Arno River Basin 在某一洪水事件期间的模拟地表水深或淹没范围图。

图表建议 12.16: 插入 Arno River Basin 在某一干旱期或平均状态下的模拟地下水埋深空间分布图。

### 12.6.5. 分析水量平衡

- 计算模拟期内 (如校准期或验证期) 的总降雨量、总实际蒸散发量、总径流量和总储量变化 ( $\Delta S = P - ET - Q$ )。
- 评估模型的整体水量平衡闭合程度。

图表建议 12.17: 插入 Arno River Basin 校准期/验证期的年均水量平衡柱状图或饼图。

## 12.7. 案例总结与讨论

### 12.7.1. 模型在 Arno River Basin 的表现评估

- 根据拟合优度指标和可视化结果, 总结模型在模拟 Arno 河流域水文过程方面的优势和不足。
- 讨论模型对洪水事件、枯水期、季节性变化的再现能力。

### 12.7.2. 建模过程中的挑战与经验

- 数据获取的难点 (例如, 高质量、长序列的本地气象和径流数据)。
- 参数本地化的挑战 (如何将全球参数适用于特定流域)。
- 模型率定中的不确定性和计算资源需求。

### 12.7.3. 模型结果的意义与潜在应用

- 模型结果对理解 Arno 河流域水文过程的贡献。
- 模型在当地水资源管理、洪水预警、气候变化影响评估等方面的潜在应用价值。

小结:

第十二章通过一个完整的 Arno River Basin 案例，演示了从数据收集、预处理、SHUD 模型构建、参数率定到结果分析和可视化的全流程。本案例旨在将前面章节的理论和操作技能整合应用，帮助您掌握独立完成一个 SHUD 建模项目的能力。请注意，这仅为一个教学示例，实际研究需要更深入的数据考证和模型验证。

思考题:

1. 在 Arno River Basin 案例中，如果缺乏本地实测的高精度 DEM，使用全球 DEM (如 SRTM 90m) 会对哪些后续建模步骤和最终结果产生主要影响？
2. 针对 Arno 河流域的地中海气候特征 (夏季干旱、秋冬季多雨)，在参数率定时应特别关注哪些水文过程或参数？
3. 如果率定结果显示模型对 Arno 河流域的夏季低流量模拟较差，可能的原因有哪些？如何进一步改进模型或率定策略？

## 13. SHUD 模型应用案例回顾 (其他)

除了第十二章详细介绍的 Arno River Basin 综合案例外，SHUD 模型及其前身 PIHM 已在全球多个不同特征的流域进行了广泛应用，验证了其模拟能力并揭示了特定区域的水文规律。本章将回顾一些具有代表性的应用案例，以进一步展示 SHUD 模型的适用性和在不同水文问题研究中的潜力。

### 13.1. 理想化案例 (模型基础验证)

理想化案例通常具有简化的几何形状和明确的边界条件，主要用于检验模型核心算法的正确性、数值解的稳定性以及对基本物理过程的描述能力。

#### 13.1.1. V 形流域 (V-Catchment)

- 参考文献: [ @Shen2010 ] (作为对比参考), [ @SHU2020\_GMD ] (SHUD 应用)
- 目标: 验证模型对理想化地形下地表径流、坡面汇流、河道径流以及质量守恒的模拟能力。
- 案例特点:
  - 几何形状: 由两个对称倾斜的矩形坡面汇入中央的一个倾斜矩形河道构成，形成一个 "V" 字形集水区。
  - 边界条件: 坡面和河床通常假设为不透水。
  - 驱动条件: 在整个流域上施加均匀、恒定强度的降雨，持续一段时间后停止。
- SHUD 应用与关键结果:
  - SHUD 模型能够很好地模拟从降雨开始到坡面产流、水流汇集至河道、最终从流域出口流出的完整过程。
  - 模拟的坡面总出流量和河道总出流量能够与总降雨量（乘以流域面积）在数值上保持良好的一致性，验证了模型的质量守恒特性。
  - 模拟的径流过程线（坡面出流和河道出流）的形态（起涨时间、洪峰流量、退水曲线）可以与解析解（如果存在）或其他成熟模型的数值解进行对比。
  - [ @SHU2020\_GMD ] 中展示了 SHUD 模拟结果与文献中其他模型结果的对比，显示出良好的一致性。

**图表建议 13.1:** 插入 V 形流域的几何形状和参数示意图 (标注坡面尺寸、坡度、曼宁糙率、河道尺寸、坡度、曼宁糙率以及降雨强度和历时)。

**图表建议 13.2:** 插入 SHUD 模拟 V 形流域的径流过程线图 (包含降雨过程、坡面总出流过程线、河道出口出流过程线, 并可与文献结果对比)。

### 13.1.2. Vauclin 水槽箱试验

- **参考文献:** [ @Vauclin1979 ] (原始实验), [ @SHU2020\_GMD ] (SHUD 应用)
- **目标:** 验证模型对非饱和带水分入渗、再分布、地下水补给以及饱和-非饱和耦合过程的模拟能力。
- **案例特点:**
  - 实验装置: 一个二维垂直剖面的沙箱, 一端有定水头边界, 底部和另一端为不透水边界, 顶部局部区域施加恒定强度的入渗 (模拟灌溉或降雨) 。
  - 观测数据: 在沙箱内部不同位置 and 不同时刻观测地下水位的变化。
- **SHUD 应用与关键结果:**
  - SHUD 模型将二维沙箱离散化为一系列垂向单元或特殊的二维网格。
  - 模型能够模拟入渗水流在非饱和带的运移、地下水位的抬升以及向定水头边界的侧向排泄过程。
  - 模拟的地下水水位在不同时间和空间位置上的变化与 [ @Vauclin1979 ] 的实验观测数据吻合良好, 验证了模型对饱和-非饱和流耦合描述的准确性。
  - 同时也揭示了模型简化 (如 SHUD 默认的单层非饱和带) 在微观尺度实验中可能存在的局限性, 例如对湿润锋的精细刻画。

**图表建议 13.3:** 插入 Vauclin 水槽箱实验装置示意图 (标注沙箱尺寸、边界条件、入渗区域、观测点位置)。

**图表建议 13.4:** 插入 SHUD 模拟 Vauclin 实验的地下水水位在不同时刻的空间分布图, 并与观测数据点进行对比。

## 13.2. 实际流域案例 (模型适用性)

将模型应用于具有复杂地形、多样的下垫面条件和真实气象驱动的实际流域, 是检验模型综合性能和实际应用价值的关键。

### 13.2.1. 美国加州 Cache 河流域

- **参考文献:** [ @SHU2020\_GMD ]
- **目标:** 在一个地形陡峭、具有明显地中海气候特征的真实中尺度流域, 评估 SHUD 模型的参数率定效果和水文过程模拟能力。
- **案例特点:**
  - 面积约 196 km<sup>2</sup>, 地形起伏大, 平均坡度 38%。
  - 地中海气候: 冬季湿润多雨, 夏季炎热干燥, 径流季节性变化显著。
  - 有 USGS 径流观测站点数据用于模型率定和验证。
- **SHUD 应用与关键结果:**
  - 利用全球数据集 (NLDAS-2 气象驱动, SRTM DEM, STATSGO 土壤, NLCD 土地利用) 构建模型。
  - 采用 CMA-ES 算法进行参数率定, 校准期和验证期的 NSE 等指标表现良好。

- 成功模拟了流域出口的日径流过程，包括洪峰和基流的动态变化。
- 展示了模型输出其他水文变量的能力，如地下水水位的空间分布和流域水量平衡组分的年际变化。

**图表建议 13.5:** 插入 Cache 河流域的地理位置图、DEM 图、土地利用图、土壤类型图以及 SHUD 模型网格划分图。

**图表建议 13.6:** 插入 Cache 河流域模拟与观测日径流过程线对比图 (包含校准期和验证期，并标注 NSE、KGE、PBias 等指标)。

**图表建议 13.7:** 插入 Cache 河流域年均地下水水位空间分布图或某个典型时刻的地下水埋深图。

## 13.2.2. 中国流域案例

SHUD 模型及其开发团队已在中国多个不同气候区和地理环境的流域开展了应用研究，积累了丰富的建模经验。

- **黑河流域 (上游):** ([04]--2.1 中的 03:41:54 开始提及)
  - **区域特点:** 中国西北干旱半干旱内陆河流域，高山冰雪融水和山区降水是主要水源。
  - **研究内容:** 通常关注冰雪融水对径流的贡献、水资源时空分布、生态需水等。SHUD 在此区域的应用与其他模型 (如 SWAT, VIC, TOPMODEL, WEAP, GBHM) 进行了对比，在日径流模拟方面取得了可比的结果。
  - **图表建议 13.8:** 插入黑河流域上游区域图及 SHUD 模拟结果与观测径流或其他模型结果的对比图。
- **布哈河流域/青海湖流域:** ([04]--2.1 中的 04:48:34, 01:20:20, 02:27:00 开始提及)
  - **区域特点:** 青藏高原高寒流域，青海湖是中国最大的内陆咸水湖，其水位变化和入湖径流是区域生态环境的重要指示。
  - **研究内容:** 模拟主要入湖河流 (如布哈河) 的径流过程，分析湖泊水量平衡 (降水、蒸发、入流、地下水交换)，预测湖泊水位变化。SHUD 的湖泊模块在此得到应用。
  - **挑战:** CMFD 等再分析气象数据在高原地区的降雨强度可能存在偏差，影响洪峰模拟；冻土过程对产流的影响需要合理参数化或物理表达。([04]--2.1 中 02:27:00 讨论了 CMFD 低估洪峰的问题及原因)
  - **图表建议 13.9:** 插入青海湖流域及主要入湖河流图，以及 SHUD 模拟的布哈河径流过程线或青海湖水位变化曲线与观测对比图。
- **深圳河流域:** ([04]--2.1 中的 01:35:21 开始提及，由刘海帆博士完成)
  - **区域特点:** 快速城市化区域，人类活动影响显著，地势相对平缓。
  - **研究内容:** 模拟城市化背景下的径流响应和地下水动态。刘海帆博士的工作展示了 SHUD 在这类受扰动流域模拟日径流和地下水位的潜力，与观测数据吻合较好。
  - **图表建议 13.10:** 插入深圳河流域土地利用图 (突出城市建成区)，以及模拟与观测的日径流和地下水水位对比图。
- **龙槽沟流域 (山洪案例):** ([04]--2.1 中的 01:48:25 开始提及)
  - **区域特点:** 四川盆地边缘山区小流域 (约 7 km<sup>2</sup>)，易发突发性山洪。
  - **研究内容:** 模拟特定暴雨事件引发的山洪过程，利用高分辨率雷达降雨数据作为驱动，检验模型在高时空分辨率下的洪水预警能力。结果显示模型能够在山洪发生前给出预警信号。
  - **价值:** 体现了分布式物理模型在小流域精细化洪水模拟和预警方面的潜力。
  - **图表建议 13.11:** 插入龙槽沟流域地形图和雷达降雨空间分布图，以及模拟的出口洪水过程线 (突出洪峰到达时间)。
- **大河流域/汾河流域 (华北):** ([04]--2.1 中的 01:46:57, 02:10:19 开始提及，与商业公司合作项目)

- **区域特点:** 华北平原及其上游山区的典型流域, 水资源短缺, 人类活动 (如水库调度、农业灌溉) 影响复杂。
- **研究内容:** 模拟特定洪水事件 (如 2012 年 7 月、2023 年 7 月洪水), 服务于防洪减灾业务需求。
- **挑战:** 如何准确获取和表达人类活动对水文过程的影响是建模的关键。参数率定需要针对特定事件或时期进行。
- **图表建议 13.12:** 插入大清河或汾河流域图, 以及某次洪水事件的模拟与观测径流过程线对比图。
- **其他商业测试案例 (如福建石壁水库):** ([04]--2.1 中的 01:43:37)
  - **目标:** 服务于水库来水预报等业务需求。
  - **特点:** 通常具有较好的本地实测气象和水文数据, 模型模拟精度要求高。

## 13.3. 特定水文过程模拟

SHUD 模型不仅能模拟流域出口的总径流, 还能对内部的特定水文过程进行详细刻画。

### 13.3.1. 洪水演进与预报

- SHUD 能够模拟洪水波在河道网络中的传播和演进过程, 包括洪峰流量、峰现时间的变化。
- 当河道水位超过堤岸高程时, 模型可以模拟洪水漫溢到相邻三角形单元, 形成地表淹没。

**图表建议 13.13:** 插入洪水演进动画的关键帧截图序列, 展示洪水波从上游向下游传播以及可能的淹没范围扩展过程。

### 13.3.2. 城市内涝

- 结合高分辨率城市 DEM、详细的土地利用分类 (特别是不同等级的不透水面) 和可能的简化的城市排水系统参数, SHUD 可以用于模拟城市区域在强降雨事件下的地表积水和内涝范围。
- [`@SHU2020_GMD`] 中提及了休斯顿 Harvey 飓风期间的城市内涝模拟, 并将模拟淹没范围与卫星影像和保险理赔数据进行了对比, 取得了较好的一致性。([04]--2.1 中的 01:52:58)

**图表建议 13.14:** 插入城市内涝模拟淹没范围图, 并可与遥感影像或其他验证数据进行叠加对比。

### 13.3.3. 地下水动态与基流分割

- SHUD 模拟了每个三角形单元的饱和地下水水位动态。
- 模型可以输出河段与相邻单元之间的地下水交换流量 (`.rivqsub`), 从而可以分析河道是“得水河段”还是“失水河段”。
- 通过分析河道总出流中来自地表径流 (`.rivqsurf`) 和地下水补给 (`.rivqsub`) 的贡献, 可以实现对径流的组分分割, 区分快流 (地表径流) 和慢流 (基流)。([04]--2.1 中 01:08:50 提到了全耦合可以做基流分割, 01:40:29 展示了布哈河地下水与河道交换的空间分布)

**图表建议 13.15:** 插入流域出口总径流过程线, 并用不同颜色区域表示其中地表径流贡献和基流贡献随时间的变化。

**图表建议 13.16:** 插入一张流域河网图, 用不同颜色表示河段是“得水河段”(地下水补给河流)还是“失水河段”(河水渗漏补给地下水)。

### 13.3.4. 冰冻圈过程



- **积雪:** SHUD 包含基于度日因子法的积雪累积和消融模块。模型可以输出每个单元的雪水当量 (`.eleysnow`)。
- **冻土:** 模型中包含一个参数化的冻土模块, 通过计算冻结指数来影响土壤的渗透性和导水性, 从而影响产流过程。([04]--2.1 中的 04:48:34 介绍了冻土参数化方案, 并展示了其对径流模拟的影响)

图表建议 13.17: 插入模拟的流域积雪雪水当量在某个冬季时刻的空间分布图。

图表建议 13.18: 插入有/无冻土模块影响下, 高寒流域径流过程线的对比图, 展示冻土对冬季基流和春季融雪径流的影响。

## 13.4. 与外部系统或模型的集成

- **全球水文数据云 (GHDC):** SHUD 模型与 GHDC 平台紧密集成, 用户可以通过 GHDC 快速获取全球范围的流域基础数据并自动构建初步的 SHUD 模型 ([04]--2.1 中的 01:57:45 详细介绍了 GHDC 和自动化部署平台)。
- **多模型耦合:** SHUD 的物理基础和模块化设计使其具有与其他领域模型耦合的潜力, 如水质模型、生态模型、农业模型、地貌演变模型等。PIHM 家族的多个耦合模型 (Flux-PIHM, LE-PIHM, RT-PIHM) 为此提供了范例。
- **决策支持系统:** SHUD 的模拟结果可以作为输入, 为更上层的决策支持系统 (如洪水风险管理系统、水资源优化调度系统) 提供水文情景。([04]--2.1 中 01:55:14 提到了美国军方利用 PIHM 进行水文-农业-经济-社会耦合, 以评估区域稳定性)

## 13.5. 理解模型选择需要服务于研究目的

- **“所有模型都是错的, 但有些是有用的” (George Box):** 这是水文建模领域的一句名言。没有一个模型能够完美复制真实世界的所有复杂过程。
- **适用性而非普适性:** 选择模型时, 应首先明确研究目标和要解决的关键科学问题。
  - 如果只关心流域出口的总水量或简单径流过程, 概念模型或集总模型可能已足够, 且数据需求和计算成本较低。
  - 如果需要理解流域内部水文过程的空间分布、物理机制、或进行土地利用/气候变化影响的精细评估, 分布式物理模型 (如 SHUD) 更具优势。
  - 如果研究重点是精细的地下水三维流动, 可能需要专业的地下水模型 (如 MODFLOW), SHUD 可以作为其地表水文边界条件。([08]--4 中的 01:57:20 讨论了与 MODFLOW 耦合的可能性)
- **数据可获得性与计算资源:** 模型的复杂性应与可用数据的质量和计算资源相匹配。

### 小结:

第十三章节通过回顾 SHUD 模型在多种理想化和实际流域中的应用案例, 展示了其广泛的适用性和在解决不同水文问题方面的能力。从基础的汇流过程验证, 到复杂的城市内涝、冰冻圈水文以及与外部系统的集成, SHUD 都表现出作为一个强大科研和应用工具的潜力。同时也强调了, 选择合适的模型始终应以具体的研究目标和实际条件为出发点。

### 思考题:

1. 在回顾的中国流域案例中, 您认为哪些案例对您自己的研究最具启发性? 为什么?
2. SHUD 模型在模拟山洪和城市内涝这类快速响应事件时, 高时空分辨率的输入数据 (如雷达降雨、高精度 DEM) 为何如此重要?
3. 基于您对 SHUD 模型特点的理解, 设想一个可以利用 SHUD 进行的创新性研究课题 (可以是您自己研究方向的拓展, 或其他感兴趣的水文问题)。

# 14. SHUD 进阶讨论、源码与附录

本章将引导您进入 SHUD 模拟系统的更深层次，包括其核心 C/C++ 源码和 rSHUD R 包的结构概览，如何参与开源贡献，以及对模型未来发展的一些讨论。附录部分则提供了一些实用速查信息。

## 14.1. SHUD 源码结构 (C/C++ 实现)

SHUD 模型的核心计算引擎是用 C/C++ 编写的，以实现高效的数值计算和内存管理。对源码结构有基本了解，有助于深入理解模型内部机制，甚至进行二次开发和定制。

### 14.1.1. 源码获取

- SHUD 模型的官方源码托管在 GitHub 上: <https://github.com/SHUD-System/SHUD>
- 您可以使用 `git clone` 命令将源码仓库克隆到本地。

### 14.1.2. 主要模块

SHUD 的源码通常组织在 `src/` 目录下，并根据功能划分为若干子目录和模块。典型的结构可能包括: ([SHUD 源码 Mapping] 提供了文件与模块的对应关系)

- main.cpp:**
  - 模型的主入口程序。
  - 负责解析命令行参数 (如项目名称、配置文件路径、输出路径等)。
  - 初始化模型控制对象，启动模拟流程。
- classes/ (或类似结构):**
  - 存放定义模型核心对象的类文件 (`.h` 头文件和 `.cpp` 实现文件)。
  - `Element`: 代表一个三角形计算单元，封装其属性、状态变量和相关的计算逻辑。
  - `River`: 代表一个河段计算单元。
  - `Lake`: 代表一个湖泊/水库计算单元 (如果湖泊模块启用)。
  - `Model_Control` (或 `ModelConfigure`): 负责整个模型的初始化、时间循环控制、调用物理过程计算、与求解器交互等。
  - `IO`: 负责模型输入文件的读取和输出文件的写入。
  - `CommandIn`: 解析和处理命令行输入。
  - 其他辅助类，如 `Node` (节点), `TimeSeriesData`, `Parameters` 等。
- Equations/ (或类似结构):**
  - 存放实现具体水文物理过程控制方程和通量计算的函数。
  - 例如，计算地表径流 (基于曼宁公式或圣维南方程简化形式)、入渗 (基于 Green-Ampt, Richards 或其他参数化方法)、蒸散发 (基于 Penman-Monteith 或其他经验公式)、地下水流动 (基于达西定律) 等的函数。
  - `cvode_config.cpp`: 通常是 CVODE 求解器的接口适配层，定义了 CVODE 所需的右端项函数  $f(t,y)$ ，该函数会调用 `Equations/` 和 `ModelData/` 中的函数来计算状态变量的变化率。



- **ModelData/ (或类似结构):**
  - 负责管理模型运行时所需的数据结构，如存储所有单元和河段的状态变量、参数、属性等。
  - 包含模型数据的初始化 (`MD_initialize.cpp`)、从输入文件读取数据到内部数据结构 (`MD_readin.cpp`)、在每个时间步更新状态和通量 (`MD_Update.cpp`, `MD_ElementFlux.cpp`, `MD_RiverFlux.cpp` 等)、以及与 CVODE 求解器交互准备数据。
- **并行实现 (OpenMP):**
  - SHUD 支持使用 OpenMP 进行共享内存并行计算。
  - 并行化通常应用于计算量最大的循环部分，例如遍历所有三角形单元进行状态更新或通量计算。相关的 OpenMP 指令 (`#pragma omp parallel for` 等) 会嵌入在这些循环中。
  - 例如，在 `ModelData/MD_f_omp.cpp` 中可能包含并行计算模型状态变化率的函数。

### 14.1.3. 关键模块代码解析 (示例 - 概念性)

深入理解源码需要具备 C++ 编程基础和数值方法知识。以下仅为概念性示例：

- **模型时间循环 (在 `Model_Control` 或 `main.cpp` 中):**

```
// 伪代码
time = startTime;
while (time < endTime) {
    // 调用 CVODE 求解器进行一个时间步的积分
    CVODE_Solve(cvode_mem, next_time_step, y_vector, &time, CV_NORMAL);
    // y_vector 包含了所有单元的状态变量

    // 更新模型内部状态 (从 y_vector 提取)
    UpdateModelState(y_vector);

    // 计算和输出当前时刻的结果 (如果达到输出频率)
    if (ShouldOutput(time)) {
        WriteOutput(time);
    }
    next_time_step = CalculateNextTimeStep(); // 确定下一个输出/积分时间点
}
```

- **CVODE 右端项函数  $f(t,y)$  (在 `Equations/cvode_config.cpp` 或 `ModelData/MD_f.cpp` 中):**

```
// 伪代码 - CVODE 调用此函数计算 dy/dt
int shud_rhs_function(realtype t, N_Vector y, N_Vector ydot, void *user_data) {
    ModelData *data = (ModelData*) user_data; // 获取模型数据指针

    // 将 N_Vector y 中的状态变量解包到模型内部数据结构中
    UnpackStateVariables(y, data);

    // 遍历所有三角形单元，计算其状态变化率
    for (int i = 0; i < num_elements; ++i) {
        CalculateElementFluxesAndRates(i, data, ydot_element_i);
        // ydot_element_i 是该单元状态变量的变化率
    }
}
```

```

// 遍历所有河段单元，计算其状态变化率
for (int j = 0; j < num_rivers; ++j) {
    CalculateRiverFluxesAndRates(j, data, ydot_river_j);
}

// ... (湖泊等其他单元)

// 将计算得到的所有变化率打包到 N_Vector ydot 中
PackStateVariableRates(ydot, data);
return 0;
}

```

## 1.4. 并行实现

SHUD 的 OpenMP 并行主要通过通过在计算密集型的循环前添加 `#pragma omp parallel for` 等指令来实现。例如，在计算所有三角形单元的蒸散发、入渗或地表径流时，这些单元的计算通常可以并行进行。

**图表建议 14.1:** 插入 SHUD C/C++ 源码主要模块结构图 (展示 `src` 目录下的主要模块及其大致的相互调用关系或数据流向)。

**图表建议 14.2:** 插入 SHUD 模型内部核心计算流程与 CVODE 求解器交互的简化示意图 (概念图，强调数据如何在模型主控、物理方程模块、数据管理模块和 CVODE 之间传递)。

## 14.2. rSHUD R 包源码结构

rSHUD 作为 SHUD 的重要配套工具，其 R 源码也遵循一定的组织结构，方便用户理解和扩展。

### 14.2.1. 源码获取

- rSHUD 的官方源码托管在 GitHub 上: <https://github.com/SHUD-System/rSHUD>
- 您可以通过 `devtools::install_github('SHUD-System/rSHUD')` 安装，也可以克隆仓库查看源码。

### 14.2.2. 主要模块

rSHUD 的源码主要位于其包内的 `R/` 目录下，每个 `.R` 文件通常包含一组功能相关的函数。

- **数据输入/输出函数:**
  - `readinput.R` (或类似名称): 包含读取 SHUD 模型各类输入文件的函数。
  - `writeInput.R` (或类似名称): 包含生成和写入 SHUD 模型各类输入文件的函数。
  - `readout.R`: 核心的输出文件读取函数，支持 `.csv` 和 `.dat` 格式。
- **GIS 处理函数:**
  - `Func_GIS.R`: 可能包含通用的 GIS 操作函数。
  - `GIS_*.R` (如 `GIS_DEMProcess.R`, `GIS_RiverNetwork.R`): 针对特定 GIS 数据类型 (DEM, 河网等) 的处理函数。
- **网格与流域处理函数:**
  - `MeshDomain.R`: 与模型域、网格生成和属性赋予相关的函数。

- `triangulate.R`: 调用 `RTriangle` 包进行 Delaunay 三角剖分的接口函数。
- **参数处理与校准辅助函数:**
  - `Func_PTF.R`: 实现或调用 Pedotransfer Functions (PTF) 以从基础土壤属性推导水力学参数。
  - `ModelBC.R`: 处理模型边界条件相关的函数。
  - 与参数校准相关的辅助函数 (如读取校准配置文件、准备校准运行等)。
- **时间序列处理函数:**
  - `Hydro_obs.R`: 处理水文观测数据 (如径流)。
  - `readTSD.R`: 读取 SHUD 的 `.tsd.*` 时间序列数据文件。
  - 与时间序列聚合、插值、分析相关的函数。
- **可视化函数:**
  - `plot.R`: 提供基础的绘图功能。
  - `plotMap.R`: 专门用于绘制空间分布图 (如模型网格、属性分布)。
  - 可能包含调用 `ggplot2` 等高级绘图包的函数。

### 14.2.3. Rcpp 接口

为了提高某些计算密集型操作 (特别是几何计算、大规模数据循环处理) 的效率, rSHUD 使用 `Rcpp` 包将部分功能用 C++ 实现, 并通过 Rcpp 接口在 R 中调用。

- **C++ 源码:** C++ 实现的函数通常位于 rSHUD 包的 `src/` 目录下 (例如 `src/triTopology.cpp`, `src/polygonArea.cpp`)。
- **Rcpp 接口文件:** `src/RcppExports.cpp` 文件由 Rcpp 自动生成, 它建立了 R 函数与 C++ 函数之间的桥梁。
- **在 R 中的调用:** 用户在 R 中调用的是封装好的 R 函数, 这些 R 函数内部通过 `.Call` 或类似机制调用底层的 C++ 实现。

图表建议 14.3: 插入 rSHUD R 包源码主要模块结构图 (展示 `R/` 目录下的主要 R 脚本文件分类及其功能)。

图表建议 14.4: 插入 rSHUD 中 R 调用 C++ (通过 Rcpp) 的简化示意图 (概念图, 展示 R 代码如何通过 Rcpp 接口调用底层的 C++ 函数, 以提高计算效率)。

## 14.3. 如何贡献代码与参与开发 (可选)

SHUD 是一个开源项目, 欢迎社区用户为其发展做出贡献。

### 14.3.1. 参与方式

- **报告 Bug:** 如果在使用 SHUD 或 rSHUD 过程中发现错误, 可以在对应的 GitHub 仓库中提交 Issue。请详细描述错误发生的环境、步骤和错误信息。
- **提出功能建议 (Feature Request):** 如果您认为模型或工具可以增加某些有用的功能, 也可以通过 GitHub Issues 提出建议。
- **贡献代码:**
  - 修复已知的 Bug。
  - 实现新的功能模块或改进现有算法。

- 优化代码性能。
- **改进文档与教程:** 帮助完善用户手册、API 文档、示例教程等。
- **在社区中提供技术支持:** 在论坛、邮件列表或 GitHub Issues 中帮助其他用户解决问题。

### 14.3.2. GitHub 协作流程 (通用流程)

1. **Fork 仓库:** 将官方的 SHUD 或 rSHUD GitHub 仓库 Fork 到您自己的 GitHub 账户下。
2. **Clone 仓库:** 将您 Fork 后的仓库 Clone 到您的本地计算机或 HPC 环境。

```
git clone https://github.com/YOUR_USERNAME/SHUD.git
# 或者 rSHUD
```

3. **创建分支 (Branch):** 在进行任何修改之前, 从主分支 (通常是 `main` 或 `master`) 创建一个新的特性分支或修复分支。

```
git checkout -b my_new_feature
```

4. **进行修改与开发:** 在新的分支上修改代码、添加功能、修复 Bug。
5. **测试:** 确保您的修改没有引入新的问题, 并通过了必要的测试。
6. **提交 (Commit):** 将您的修改提交到本地分支。

```
git add .
git commit -m "Description of my changes"
```

7. **推送 (Push):** 将您的本地分支推送到您在 GitHub 上的 Forked 仓库。

```
git push origin my_new_feature
```

8. **创建拉取请求 (Pull Request, PR):** 在 GitHub 网站上, 从您的 Forked 仓库的 `my_new_feature` 分支向官方仓库的主分支发起一个 Pull Request。在 PR 中详细描述您的修改内容、目的和测试情况。
9. **代码审查与合并:** 项目维护者会审查您的 PR, 可能会提出修改意见。经过讨论和修改后, 如果代码被接受, 维护者会将其合并到官方仓库中。

## 14.4. SHUD 进阶讨论与展望

### 14.4.1. 模型局限性与未来改进方向

- **土壤水分运动:**
  - **多层土壤模型:** 当前 SHUD 主要采用简化的两层 (非饱和-饱和) 或单层非饱和带模型。未来可以考虑引入更精细的多层土壤模型, 以更好地模拟垂向水分和溶质运移。
  - **非饱和带水平流动:** 目前 SHUD 通常忽略非饱和带的水平流动, 这在某些特定地形和土壤条件下可能引入误差。选择性地加入或改进非饱和带水平流的表达是一个方向。
- **冰冻圈过程:**

- **冻土物理过程:** 当前冻土模块多为参数化方案 ([04]--2.1 中的 04:48:34, 01:30:08)。未来应引入更基于物理的冻融过程模型, 考虑土壤热量传输和相变。
- **冰川模块:** 整合更完善的冰川动态和冰雪融水模块。
- **河道过程:**
  - **复杂河道形态:** 对于辫状河、地下河、受潮汐影响的河口等复杂河道, 现有的一维河道模型可能不足, 需要更专业的河流水动力模块。
  - **河床演变:** 与泥沙输运模型耦合, 模拟河床冲淤。
- **植被与蒸散发:**
  - **动态植被:** 考虑植被生长、竞争和演替的动态过程, 而非静态的土地利用类型。
  - **更精细的蒸散发模型:** 耦合更复杂的双源或多源蒸散发模型。
- **数值方法与计算效率:**
  - **求解器优化:** 探索更适合特定水文问题的 ODE/PDE 求解策略。
  - **并行计算:** 进一步优化 OpenMP 效率, 探索 MPI 等跨节点并行技术以支持更大规模模拟。
  - **自适应网格:** 开发能够根据模拟过程动态调整网格密度的技术。
- **参数不确定性与数据同化:** 结合更先进的参数优化算法和数据同化技术 (如集合卡尔曼滤波 EnKF), 以量化和减小模型预测的不确定性。

## 14.4.2. 高性能计算的进一步应用

- **大规模流域模拟:** 利用 HPC 进行超大流域 (如整个长江、黄河流域) 的高分辨率、长时序模拟。
- **集合预报:** 运行大量不同参数、不同初始条件或不同气象驱动的模式, 进行概率预报和不确定性评估。
- **实时洪水预报系统:** 将 SHUD 模型部署在实时数据流和 HPC 环境中, 实现业务化的洪水预报。

## 14.4.3. 与其他模型的耦合前景

SHUD 作为一个物理过程清晰的水文模型, 为与其他领域模型的耦合提供了良好基础。

- **生态模型:** 耦合植被动态模型、生物地球化学循环模型, 研究水文过程对生态系统的影响及反馈。
- **农业模型:** 耦合作物生长模型, 评估水资源对农业生产的影响, 优化灌溉管理。
- **地貌模型:** 耦合泥沙输运和地貌演变模型, 研究长期水文地貌过程。
- **大气模型:** 实现陆气双向耦合, 更准确地模拟局地气候和水循环。
- **社会经济模型:** 将水文模拟结果作为输入, 评估水资源变化对社会经济系统的影响, 支持水资源管理决策。  
([04]--2.1 中 01:55:14 提到了美军的项目)

## 14.4.4. 数据与模型的协同发展

- **高精度 ETV 数据:** 遥感技术和地面观测网络的发展为模型提供了越来越丰富和高精度的数据源, 这反过来也对模型提出了更高的要求。
- **数据同化:** 将实时或近实时的观测数据 (如卫星土壤湿度、积雪、地下水储量变化等) 融入模型, 提高模拟精度和预测能力。

- **模型作为数据生成的工具:** 利用物理模型在数据稀缺地区生成高质量的“虚拟观测数据”，服务于其他研究或模型。

### 14.4.5. 智能算法在水文模型中的应用

- **参数率定:** 利用机器学习算法 (如遗传算法、粒子群优化、贝叶斯优化) 改进参数优化效率和全局搜索能力。
- **模型代理/替代模型 (Surrogate Models):** 训练机器学习模型来模拟复杂物理模型的输入输出关系，以大幅减少计算时间，用于敏感性分析或快速情景模拟。
- **过程识别与数据挖掘:** 利用数据挖掘技术从大量模拟输出和观测数据中发现新的水文规律或改进模型结构。
- **混合建模:** 将物理模型与数据驱动模型相结合，发挥各自优势。

## 14.5. 附录

### 14.5.1. 常见错误代码及排查指南

(本节内容将在后续详细填充，根据训练营中遇到的实际问题和常见错误进行总结)

- **示例:**
  - **SHUD 模型运行错误:**
    - `Error: Input file [filename] not found.` -> 检查文件路径和名称。
    - `CVODE Error: CV_TOO_MUCH_WORK` -> 尝试减小 `MAX_SOLVER_STEP` 或调整容差。
    - `Error: Negative depth/storage in element/river [ID]` -> 模型发散，检查参数、初始条件或网格。
  - **AutoSHUD/rSHUD 脚本错误:**
    - `Error: package 'xxx' could not be loaded` -> 安装或加载缺失的 R 包。
    - `Error in file(con, "r") : cannot open the connection` -> 文件路径错误或文件不存在。

### 14.5.2. 关键文件对照表

(本节内容将在后续详细填充，提供 SHUD 输入/输出文件标识符与物理变量、计算单元的详细对照表)

### 14.5.3. 常用 R 函数速查表 (rSHUD 及数据处理)

(本节内容将在后续详细填充，列出 rSHUD 包中常用函数及其功能，以及进行数据处理时常用的基础 R 函数)

- **示例:**
  - `rSHUD::readout()`: 读取 SHUD 输出文件。
  - `rSHUD::triangulate()`: 生成 Delaunay 三角网格。
  - `rSHUD::shud.env()`: 设置 SHUD 项目环境变量。
  - `xts::apply.daily()`, `xts::apply.monthly()`: 时间序列聚合。
  - `raster::raster()`, `raster::crop()`, `raster::projectRaster()`: 栅格数据处理。
  - `sf::st_read()`, `sf::st_transform()`, `sf::st_intersection()`: 矢量数据处理。

## 14.5.4. Slurm 命令速查表

(本节内容将在后续详细填充，列出 Slurm 常用命令及其选项，方便用户在 HPC 上管理作业)

- 示例:

- `sbatch <script.sh>`: 提交作业。
  - `squeue -u <username>`: 查看用户作业。
  - `scancel <job_id>`: 取消作业。
  - `sinfo -p <partition>`: 查看队列信息。
- 

### 小结:

第十四章节 (原计划第十三部分) 探讨了 SHUD 模型的源码结构、参与开源贡献的方式，并对模型未来的发展方向进行了展望。附录部分将提供实用的速查信息。通过本章的学习，希望您能对 SHUD 系统有更深入的理解，并为可能的进阶使用或参与开发做好准备。

### 思考题:

1. 如果您发现 SHUD 模型在模拟您研究区域的某一特定水文过程 (例如，融雪径流的峰现时间) 时存在系统性偏差，您会考虑从哪些方面 (如输入数据、模型参数、模型结构本身) 入手进行改进？
  2. 开源模型的发展依赖社区的贡献。结合您的专业背景和兴趣，您认为可以从哪些方面为 SHUD 社区做出贡献？
  3. 在未来的水文模型发展中，您认为物理模型与数据驱动模型 (如机器学习) 应该是一种怎样的关系？是相互替代还是融合发展？
-