

气象数据订正系统

第一章 气象数据订正系统概述	2
1.1 系统简介	2
1.2 开发背景	2
1.3 设计原则	2
1.4 运行环境要求	2
第二章 气象数据订正系统功能介绍	4
2.1 路径设置模块	4
2.2 数据导入模块	5
2.3 站点展示模块	6
2.4 格点数据展示模块	6
2.5 数据预处理模块	7
2.6 数据建模模块	8
2.7 数据订正模块	9
第三章 系统特点与创新点	11
第四章 系统总结	12

第一章 气象数据订正系统概述

1.1 系统简介

气象数据订正系统是一套集数据采集、处理、建模、订正与可视化于一体的专业气象数据智能处理平台。系统基于现代气象学理论与人工智能技术，利用多源数据融合、异常检测、机器学习订正等方法，对原始气象观测与数值模式数据进行质量控制与偏差修正，从而显著提升数据精度与可靠性。

系统通过模块化设计实现从“数据导入—预处理—建模—订正—输出”的全流程自动化处理，能够满足气象业务、科研分析、新能源预测、农业气象服务等多领域的应用需求。平台采用前后端分离架构，前端以React框架实现高交互界面，后端基于Node.js与Python计算引擎构建，支持高并发与海量数据处理，兼顾可视化与高性能计算。

本系统以“智能化、模块化、可扩展”为核心设计理念，致力于构建一个兼具科研深度与工程实用性的气象数据处理工具，帮助用户高效完成数据质量提升与订正分析工作。

1.2 开发背景

随着气象观测与模式模拟技术的不断发展，气象数据规模呈指数级增长。然而，不同来源、不同分辨率、不同时间尺度的数据之间往往存在偏差。观测误差、仪器漂移、地形修正不足、模式系统误差等问题都会导致数据的系统性偏差，进而影响气象预测、气候评估及相关行业模型（如风电、光伏功率预测）的精度。

传统的订正方法多采用线性回归或偏差统计校正模型，无法有效捕捉复杂非线性关系。而机器学习方法在多变量回归、特征选择及非线性建模方面表现突出，已逐渐成为气象数据质量提升的重要技术途径。

本系统正是基于这一技术背景研发，通过机器学习和数据挖掘方法，对气象数据进行系统化、智能化的订正与可视化处理，填补了传统订正方法精度低、自动化程度不足的空缺。

1.3 设计原则

气象数据订正系统在设计过程中遵循以下核心原则：

高效性：采用并行计算和异步任务管理机制，确保大规模气象数据的快速处理。

准确性：通过机器学习与统计建模相结合的方式，大幅提升订正结果的精度。

稳定性：设置完善的异常捕获、日志记录与数据备份机制，保障系统长期稳定运行。

易用性：前端界面直观友好，提供图形化路径选择、进度显示与结果导出功能，降低用户使用门槛。

可扩展性：系统采用模块化与接口化设计，支持新算法与新数据源的快速集成。

1.4 运行环境要求

软件环境：

操作系统：Windows 7/10/11，麒麟系统等主流操作系统

运行环境：React框架，Node.js

用户硬件配置

组件	最低配置	推荐配置
CPU	Intel Core i3 或同等	Intel Core i7 / i9, 主频 \geq 2GHz
内存	8GB	32GB及以上
显卡	集成显卡	独立显卡
显示器	分辨率 \geq 1920×1080	高清宽屏
其他设备	鼠标、键盘	鼠标、键盘

第二章 气象数据订正系统功能介绍

本章将详细介绍气象数据订正系统的整体功能框架与核心业务流程。系统采用模块化设计思想，将复杂的气象数据处理流程分解为一系列逻辑清晰、功能独立的子模块。这些模块通过数据流相互衔接，共同构成一个从原始数据输入到订正结果输出的完整闭环处理链路。

系统整体工作流程如下（如图1所示）：用户首先通过路径设置模块配置数据源，继而使用数据导入模块将数据载入系统数据库。随后，数据可经由两条路径进行处理：一是通过站点展示模块或格点数据展示模块进行初步的可视化探查；二是进入核心处理流程，依次经过数据预处理模块的清洗与规整、数据建模模块的模型训练与评估，最终由数据订正模块应用训练好的模型完成数据的智能订正与输出。所有模块均围绕SQLite数据库进行数据交互，确保了数据的一致性与流程的连贯性。该架构确保了系统兼具灵活性与健壮性，用户既可以按顺序执行完整流程，也可以根据需求单独使用特定模块，满足了从数据探索、模型研发到业务化订正的不同应用场景需求。

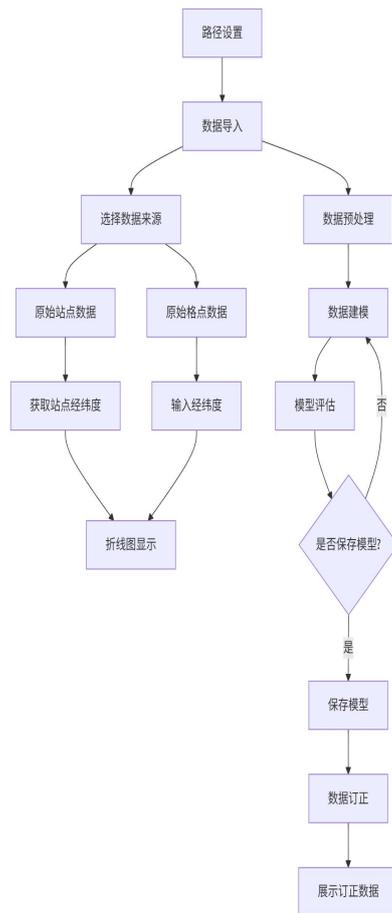


图1 系统框架图

2.1 路径设置模块

功能概述

路径设置模块是气象数据订正系统的“入口配置中枢”，其核心功能是为后续所有数据处理流程提供准确、合法的数据源定位。用户通过本模块可分别指定原始站点观测数据与格点预报数据的存储根目录，系统将据此建立数据访问通道，并将路径信息持久化保存至本地配置文件，确保后续模块在启动或运行时能自动加载并调用正确的数据源路径。

该模块采用直观的图形化界面设计，极大降低了用户的操作门槛，即使非专业技术人员也能快速完成基础环境配置。同时，模块内置智能校验机制，有效避免因路径错误导致的后续流程中断，提升了系统的健壮性与用户体验。

界面布局与交互设计 (如图2示)

如图所示，模块界面结构清晰，分为两个独立的功能区域，分别对应不同类型数据的路径设置：

原始数据文件夹设置

标题明确标识为“原始数据文件夹设置”，用于配置站点观测数据的存放路径。

输入框内已预填示例路径：`data/station/湖北省82个国家气象站-观测数据`，提示用户应填写包含具体观测站点数据文件的父级目录。

用户可通过点击输入框右侧的“浏览”按钮（图中未显示，但为标准交互），调出系统文件选择对话框，以图形化方式选择目标文件夹，避免手动输入路径可能引发的格式或拼写错误。

输入框旁设有红色星号“*”，表明此字段为必填项，系统在保存前会强制校验其有效性。

格点数据文件夹设置

标题为“格点数据文件夹设置”，用于配置格点预报产品的存放路径。

示例路径为 `C:/CARAS`，表明系统支持任意磁盘分区下的绝对路径，适用于不同部署环境。

同样配备“浏览”功能（隐含交互）及必填标识“*”，确保用户必须指定有效的格点数据源。

全局操作按钮

在界面右下角设有醒目的蓝色“保存设置”按钮。用户完成路径填写后，点击该按钮即可触发系统执行以下操作：

对两个输入框中的路径进行合法性校验

若校验通过，则将路径信息写入系统配置文件（如 `config.ini` 或 `settings.json`）；

若校验失败，系统将弹出友好提示框，明确指出错误原因（例如：“指定的原始数据文件夹不存在，请重新选择”），引导用户修正。

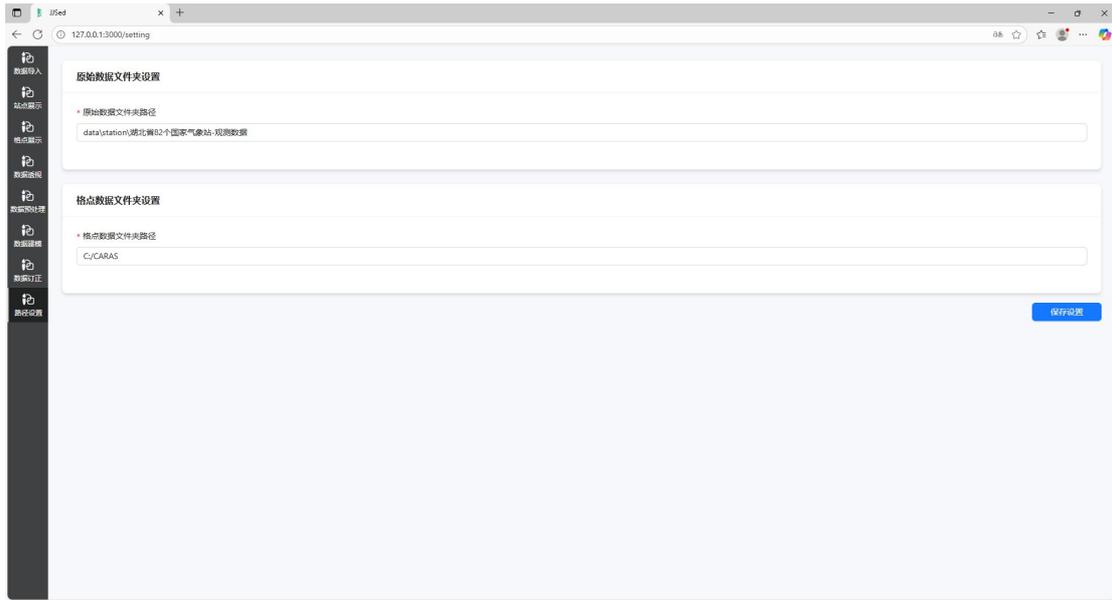


图2 路径设置图

2.2 数据导入模块

功能描述：

数据导入模块是气象数据订正系统的“数据中枢”，承担着将原始观测或预报文件从本地磁盘批量读取、解析并写入系统核心数据库（SQLite）的关键任务。它是连接外部数据源与内部处理流程的桥梁，为后续的数据可视化、预处理、建模与订正提供结构化、标准化的数据基础。

本模块支持对站点观测数据（如CSV格式）进行自动扫描、智能识别与高效入库，其设计充分考虑了海量文件处理的性能需求与用户操作的便捷性。通过直观的列表展示、状态跟踪与一键启动机制，用户可清晰掌握数据导入进度，及时发现并处理异常文件，确保数据流的完整性与可靠性。

输出结果：

导入后的数据存储于SQLite数据库中，并自动生成数据索引表。

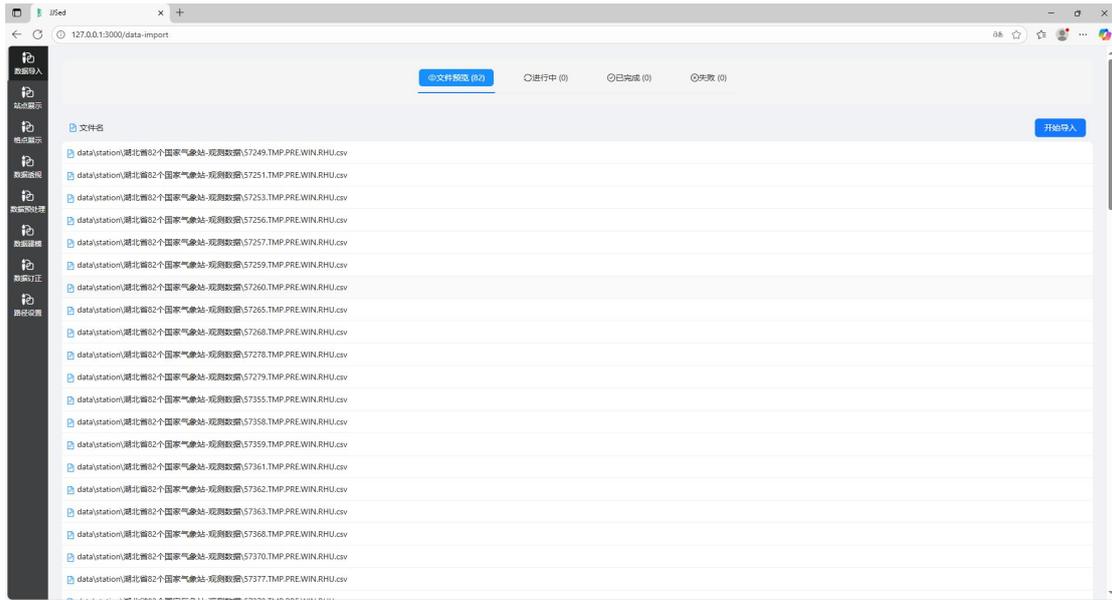


图3 数据导入图

2.3 站点展示模块

功能概述

站点数据展示模块是气象数据订正系统中的“可视化探查引擎”，旨在为用户提供直观、交互性强的单站时序数据分析能力。用户可通过该模块，基于站点经纬度或名称快速定位目标观测站，并在自定义时间范围内，动态绘制并对比温度、湿度、风速、降水等关键气象要素的变化趋势。该模块不仅服务于数据质量初判与异常识别，更支持科研人员对特定站点气候特征进行深度挖掘，是连接原始数据与智能建模的重要桥梁。

核心功能与交互设计 (如图4示)

界面顶部提供多维度筛选控件：用户可选择数据来源（如“原始观测数据”或“格点预报数据”）、指定具体站点（如“竹溪”）、输入经/纬度范围辅助定位、选择气象要素（如“温度”）及设定日期区间（如“2010-01-01 至 2010-01-31”）。点击“查询”按钮后，系统即时渲染下方折线图，清晰呈现选定要素在时间轴上的波动形态。

图表区域支持高级交互功能：用户可通过鼠标拖拽缩放时间轴，聚焦于特定时段；悬停于曲线上任意点，将弹出浮动标签，精确显示该时刻的数值（如“2010-01-22 08:00 温度 5.5°C”），极大提升数据解读效率。此外，模块还提供“导出数据”功能，可将当前图表所用原始数据以CSV格式下载，便于离线分析或报告撰写。

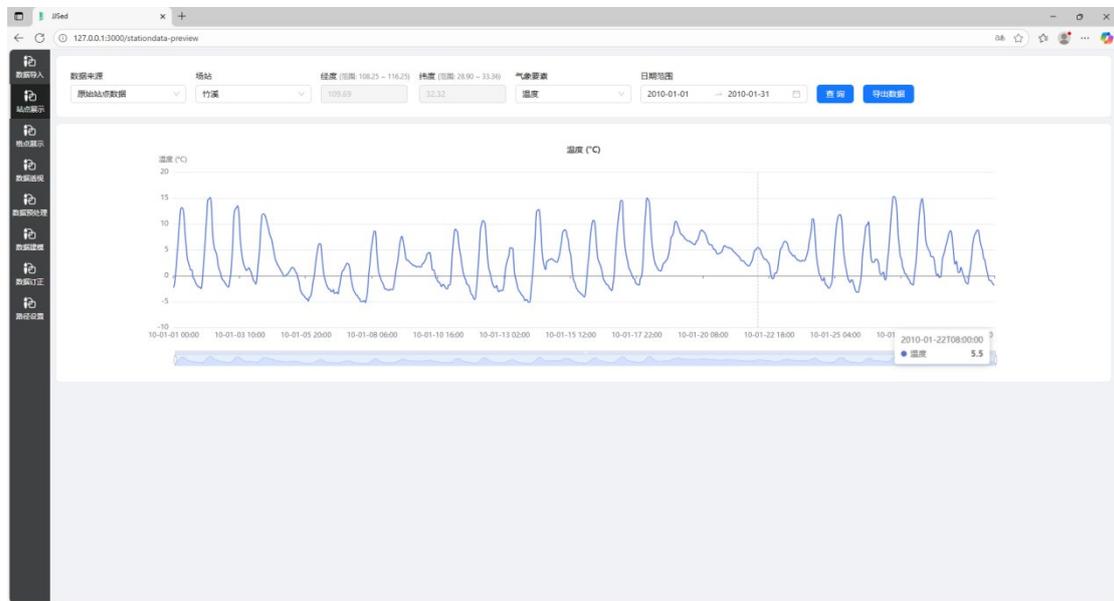


图4 站点展示图

2.4 格点数据展示模块

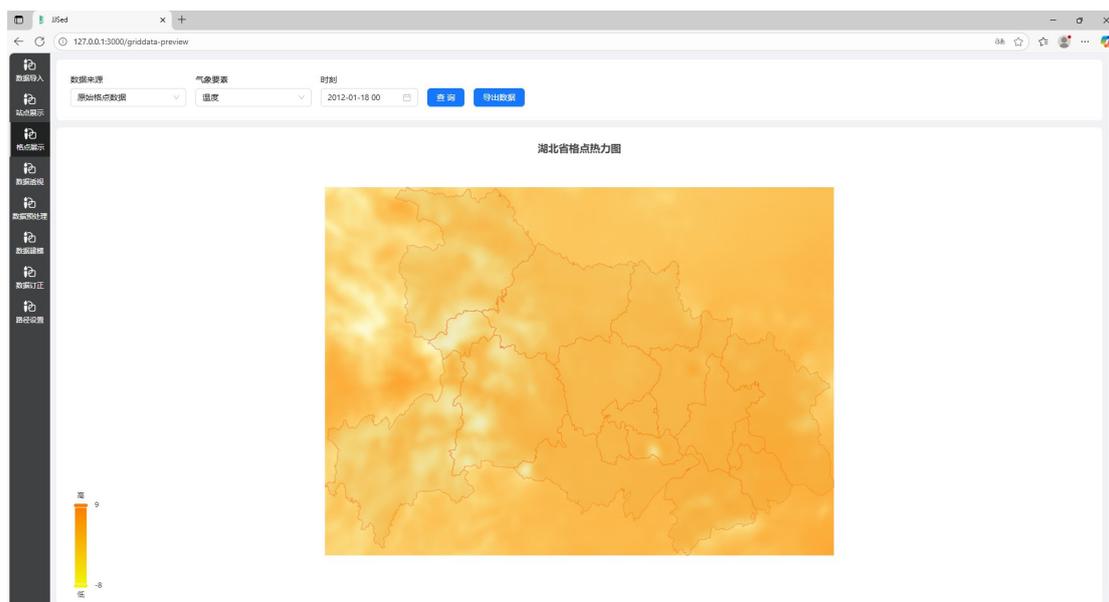
功能概述

格点数据展示模块是系统面向区域尺度气象分析的核心可视化工具，专为呈现高分辨率、空间连续的格点化气象场而设计。该模块允许用户通过交互式时间轴选择任意预报或分析时次，系统随即动态渲染覆盖全省（如图所示“湖北省”）的热力图，直观展现温度、降水、风场等关键气象要素的空间分布格局。其目标在于帮助用户快速把握天气系统的宏观结构、演变趋势与区域差异，为模式产品评估、灾害预警及订正模型输入提供强有力的空间视角支持。

界面布局与交互设计（如图5示）

界面顶部控制区简洁高效：用户可切换“数据来源”（如“原始格点数据”），选择待展示的“气象要素”（如“温度”），并通过“时次”选择器精确指定某一时刻（如“2012-01-18 00:00”）。点击“查询”按钮后，主视图区即刻生成一幅色彩渐变的热力图，以地理轮廓为边界，清晰勾勒出湖北省内各区域的温度梯度分布。左侧配色条（Colorbar）标注了数值范围（如“低 -8 至 高 9”），使颜色与物理量值一一对应，便于快速解读。

模块核心在于“时间轴滑动”功能——用户可通过拖动时间控件或点击前后箭头，逐时浏览气象场的时空演变过程，形成类似“动画”的动态效果，极大增强了对天气系统移动、发展和消散的认知深度。同时，“导出数据”按钮支持将当前时刻的格点矩阵数据以标准格式（如NetCDF或GeoJSON）下载，供进一步科研或业务应用。



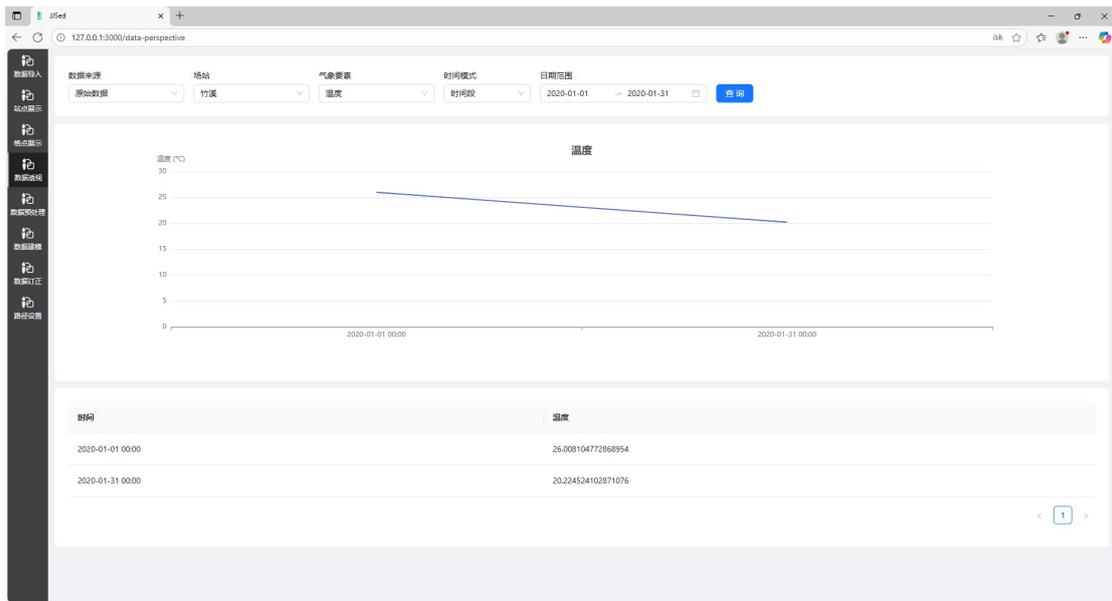


图5 数据展示图

2.5 数据预处理模块

功能概述

数据预处理模块是气象数据订正系统中的“数据净化与特征引擎”，承担着将原始、杂乱、不规整的观测或格点数据转化为高质量、标准化、可建模输入的关键任务。该模块通过一套系统化、可配置化的清洗与转换流程，对导入的气象数据进行深度质量控制与结构优化，确保后续建模环节所使用的数据具备完整性、一致性与统计可比性，从而大幅提升模型训练的稳定性和预测精度。

核心功能与交互设计 (如图6示)

界面顶部提供灵活的参数配置区：用户可通过下拉菜单选择待处理的“气象变量”（如“温度”、“相对湿度”、“过去1小时降水量”、“2分钟平均风速”等），并设定“开始年份”、“结束年份”及“任务并发数”，以控制处理范围与资源分配。点击醒目的蓝色“开始预处理”按钮后，系统即启动后台批处理引擎，对指定时间段内的所有数据执行多步骤清洗操作。

下方任务管理区实时展示处理进度：包含“进行中”、“已完成的任务”、“失败的任务”三个标签页，便于用户监控不同批次的执行状态。在“进行中”页面，系统会列出当前正在处理的“任务ID”、“种类”、“数据时间”及“状态”，支持用户随时暂停、重试或查看日志。当无数据时，界面显示“暂无数据”图标，提示用户需先完成数据导入或选择有效变量。

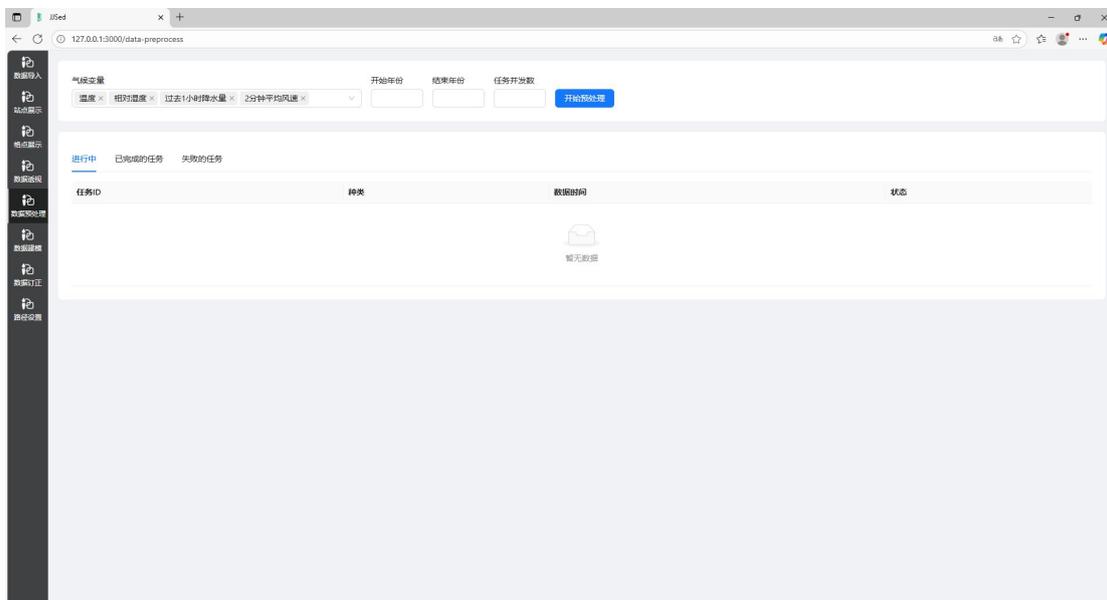


图6数据预处理图

2.6 数据建模模块

功能概述

数据建模模块是气象数据订正系统的“智能大脑”，承担着从清洗后的历史数据中学习气象要素内在规律、构建高精度订正模型的核心使命。该模块集成了多种主流机器学习与深度学习算法，支持用户根据数据特性与业务需求灵活选择模型架构，并通过可视化界面配置关键超参数，实现从数据到模型的端到端自动化训练。系统内置科学的数据划分策略与多维度评估体系，确保所训练模型不仅在训练集上表现优异，更具备良好的泛化能力与业务实用性。

核心功能与交互设计 (如图8示)

界面顶部为模型配置区，提供高度可定制化的参数设置：

算法选择：下拉菜单支持切换“线性回归”、“随机森林”、“XGBoost”、“LSTM”等主流模型，当前图示中选中的是“LightGBM”。

变量配置：用户可指定目标变量（如“温度”）及输入特征变量（如“相对湿度”、“风速”、“过去1小时降水量”等），支持多变量联合建模。

划分方法：提供“按年份划分”或“随机划分”两种策略，便于用户按时间序列特性或统计独立性进行训练/验证集分离。

季节控制：可限定模型仅针对特定季节（如“全年”、“夏季”、“冬季”）进行训练，提升模型在特定气候背景下的适应性。

时间范围：设定“起始年份”、“结束年份”、“测试年份”，明确训练数据与独立测试数据的时间边界，避免数据泄露。

超参数调优：开放关键参数调节，如“学习率”、“树数量”、“叶子数”、“采样比例”、“特征采样”、“L2正则化”、“早停轮数”等，满足高级用户对模型性能的精细调控。

点击“提交任务”按钮后，系统将自动启动后台训练引擎，并在下方“进行中”标签页实时显示任务列表，包含“任务ID”、“算法”、“变量”、“划分方法”、“季节”、“起止年份”、“测试数据”、“进度”及“操作”列，支持用户随时查看、暂停或下载结果。

模型训练与评估 (如图8示)

当任务完成时，系统会弹出详细评估窗口，直观展示模型性能：

训练过程图：绘制Loss曲线，对比训练损失与验证损失随迭代次数的变化趋势，帮助判断模型是否收敛、过拟合或欠拟合。

评估指标表：提供多维度量化评价，包括：

相关系数 (CC)：衡量预测值与真实值的线性相关强度；

均方根误差 (RMSE)：反映预测偏差的整体幅度；

平均绝对误差 (MAE)：体现预测值偏离真实值的平均距离；

平均相对误差 (MRE)：以百分比形式表达误差规模；

决定系数 (R^2)：评估模型解释数据变异的能力；

Pearson相关系数：补充衡量变量间线性关联度。

所有指标均精确到小数点后六位，确保科研级严谨性。用户可通过“查看报告”或“保存”按钮导出完整训练日志与评估结果，便于归档、复现或用于后续模型集成。

应用价值与技术亮点

本模块不仅提供“开箱即用”的默认配置，更赋予用户充分的自由度去探索不同算法与参数组合，是科研人员进行模型对比实验、业务人员优化订正效果的理想平台。其内置的自动化交叉验证、早停机制与多指标评估体系，有效规避了人为经验依赖，保障了模型开发的科学性与可重复性，为最终生成稳定、可靠、高精度的气象数据订正产品奠定了坚实基础

基础

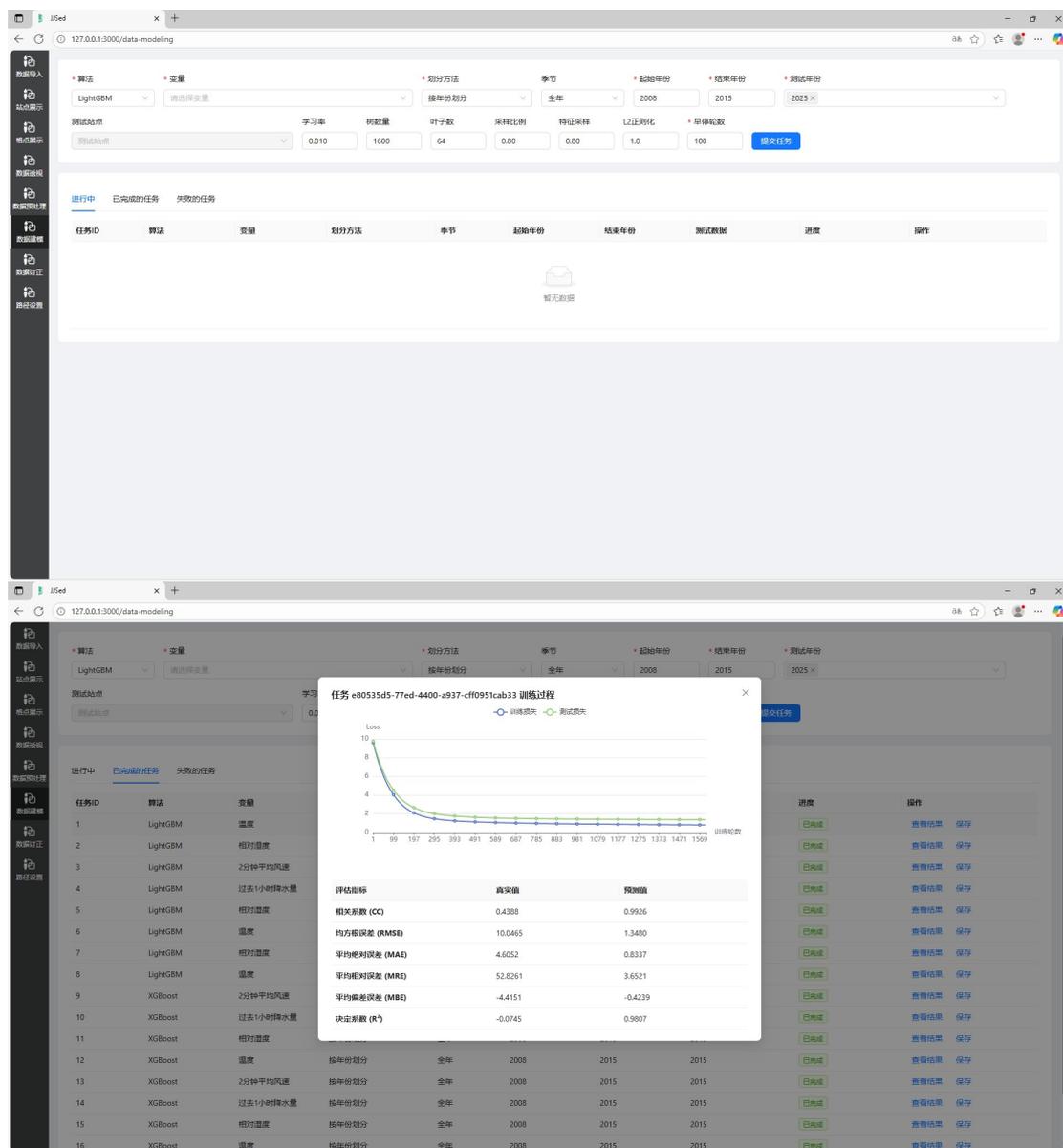


图7 数据建模图

2.7 数据订正模块

功能概述

数据订正模块是气象数据订正系统的“最终执行单元”，负责将经过训练与验证的最优模型应用于待订正的原始格点或站点数据，生成高精度、偏差校正后的气象产品。作为整个处理链路的出口环节，该模块不仅实现了从“模型”到“业务成果”的转化，更通过灵活的任务调度与结果管理机制，支持批量、定时或按需的订正服务，满足科研分析与业务化运行的双重需求。

本模块紧密依赖于2.6节中训练完成的模型文件与参数配置，确保订正过程与建模阶段在数据分布、特征空间和时间窗口上完全一致，从而保障订正结果的科学性与可复现性。

核心功能与交互设计（基于图8）

界面顶部为任务配置区，提供清晰的操作入口：

模型选择：下拉菜单列出所有已成功训练并保存的模型（如“温度_XGBoost_2010-2020”），用户可一键选用目标模型。

数据源指定：支持选择“原始格点数据”或“站点观测数据”作为待订正对象，系统自动匹配模型输入维度。

时间范围设定：用户可指定需订正的起止日期（如“2021-01-01 至 2021-12-31”），适用于历史数据回算或实时滚动订正场景。

输出配置：可勾选“保存订正结果至数据库”、“导出为CSV/NetCDF格式”、“生成偏差对比图”等选项，满足不同下游应用需求。

点击“开始订正”按钮后，系统启动后台推理引擎，自动加载模型权重、读取待处理数据、执行特征对齐与预测计算，并将订正结果写入SQLite数据库的corrected_data表中。下方任务面板实时显示“进行中”、“已完成”、“失败”三类任务状态，每项任务均标注模型名称、数据类型、时间范围及进度条，支持用户随时中断或重试。

技术实现与质量保障

模型一致性校验：系统在订正前自动检查待处理数据的特征列、时间频率、单位制是否与训练阶段一致，若存在不匹配（如缺失关键输入变量），将终止任务并提示错误，防止“错配推理”。

批量并行处理：针对大范围格点数据，系统采用分块并行推理策略，利用多核CPU加速计算，显著缩短订正耗时。

结果可追溯性：每条订正记录均关联原始数据ID、所用模型ID、执行时间戳及版本号，确保结果可审计、可回溯。

偏差可视化对比：订正完成后，系统可自动生成“原始 vs 订正”双曲线图（站点）或双热力图（格点），直观展示偏差修正效果，辅助用户判断订正合理性。

应用价值与输出成果

本模块输出的订正数据可直接用于：
气象业务中的高精度实况分析与短临预报；
新能源（风电、光伏）功率预测模型的输入驱动场；
农业气象灾害风险评估与灌溉决策支持；
气候变化研究中的长期序列均一化处理。

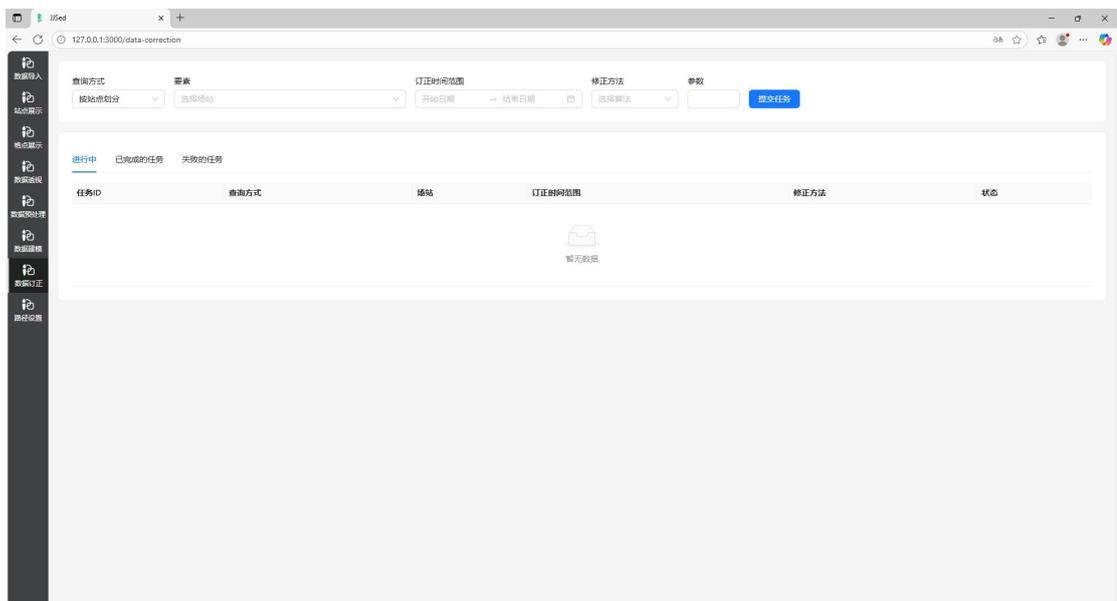


图8 数据订正图

第三章 系统特点与创新点

本系统在设计理念、技术实现与应用价值上均具备显著优势，其核心特点与创新点可归纳为以下四个方面：

1 全流程一体化设计

系统突破了传统气象数据处理工具功能单一、流程割裂的局限，创造性构建了从数据接入、质量探查、智能预处理、机器学习建模到业务化订正应用的端到端解决方案。该设计将原本分散于多个软件、需要人工干预衔接的环节无缝集成，实现了“一键式”数据质量提升，极大提升了业务与科研工作的效率，并降低了因人为操作导致错误的风险。

2 模型驱动的智能订正

本系统的核心创新在于将先进的机器学习算法深度嵌入气象数据订正流程。相较于依赖固定公式或简单统计回归的传统订正方法，系统所采用的随机森林、XGBoost、LSTM等模型具备强大的非线性拟合与时序特征学习能力，能够更精准地捕捉观测数据与模式数据之间复杂的误差关系，从而实现对系统性偏差的更有效修正，显著提升订正结果的客观性与准确性。

3 交互式可视化分析与决策支持

系统并非“黑箱”操作，而是提供了丰富、直观的可视化交互界面。用户不仅可以通过站点折线图、格点热力图多维度探查数据特征与问题，还能在模型训练阶段实时评估不同参数下的性能指标，为模型优化提供直观依据。这种“可视化分析-模型调优-结果验证”的闭环，将专家的领域知识与数据驱动模型相结合，构成了强大的人机协同决策支持系统。

4 高可扩展的模块化架构

系统采用前沿的模块化与前后端分离架构设计，使得核心数据处理、算法模型与用户界面高度解耦。这种设计赋予了系统卓越的可扩展性与维护性。未来，新增数据源的支持、新订正算法的集成或新可视化图表的开发，均可作为独立模块便捷地嵌入现有框架，确保了系统能够持续演进，快速响应气象科学与业务需求的未来发展。

第四章 系统总结

气象数据订正系统作为面向现代气象业务与科研需求的专业化智能平台，成功实现了从原始数据到高质量订正产品的全流程闭环处理。系统以“智能化、模块化、可扩展”为核心设计理念，深度融合气象学理论、数据科学方法与工程化软件架构，构建了一个兼具科学严谨性与操作便捷性的综合解决方案。

在功能层面，系统完整覆盖了数据接入、质量探查、清洗预处理、机器学习建模、智能订正与结果输出六大关键环节。通过路径设置、数据导入、站点/格点可视化等模块，用户可高效完成数据准备与初步诊断；借助灵活可配的数据预处理引擎，系统能自动识别并修复缺失值、异常值，完成标准化与特征衍生；在建模阶段，集成线性回归、随机森林、XGBoost、LSTM等多种先进算法，支持多维度超参数调优与科学评估，确保模型性能最优；最终，数据订正模块将训练好的模型无缝应用于业务数据，输出高精度、低偏差的订正结果，并支持多种标准格式导出，便于下游系统集成。

在技术层面，系统采用前后端分离架构，前端基于 React 实现高交互性用户界面，后端融合 Node.js 服务调度与 Python 高性能计算引擎，兼顾响应速度与计算能力。所有模块围绕 SQLite 数据库协同工作，保障数据一致性与流程连贯性。同时，系统内置任务队列、异常捕获、日志追踪与结果可追溯机制，确保长时间、大规模运行的稳定性与可靠性。

更为重要的是，本系统突破了传统订正方法依赖经验公式、难以适应复杂非线性误差的局限，通过数据驱动的机器学习范式，显著提升了订正精度与泛化能力。其交互式可视化设计不仅增强了用户对数据与模型的理解，更促进了“人机协同”的科学决策模式，使气象数据质量控制从“经验主导”迈向“智能驱动”。

气象数据订正系统不仅在技术架构上具备先进性，在实际应用中也展现出强大的实用性与推广价值。它可广泛服务于气象预报、气候研究、新能源（风电/光伏）功率预测、农业气象、水文模拟及防灾减灾等多个关键领域，为提升我国气象数据产品质量、支撑精细化气象服务提供坚实的技术底座，具有显著的社会效益与广阔的应用前景。