**基于多源气象数据的风电功率误差传播分析研究方案**

**1. 数据预处理与探索性分析**

**1.1 数据质量检查**

* **缺失值处理**：分析各变量缺失模式，制定插补策略
* **异常值检测**：识别和处理极端气象条件和功率异常
* **时间对齐验证**：确保所有数据源的时间戳一致性

**1.2 基础统计分析**

* **变量分布特征**：各气象变量的概率分布分析
* **相关性矩阵**：变量间的线性和非线性相关关系
* **垂直梯度分析**：不同高度间风速、风向的关系
* **季节性分析**：识别气象变量的季节性和日变化模式

**1.3 预测误差特征分析**

误差定义：

- ECMWF误差 = ec\_variable - obs\_variable

- GFS误差 = gfs\_variable - obs\_variable

分析内容：

- 各变量预测误差的统计特性（均值、方差、偏度、峰度）

- 误差的时间相关性和空间相关性

- 不同天气条件下的误差分布差异

**2. 功率模型构建（从简单到复杂）**

**2.1 阶段一：简单功率曲线模型**

**单变量模型**：

P = f(U)

选择代表性高度（如70m）的风速作为输入

**模型选择**：

* 多项式拟合
* 分段线性拟合
* Sigmoid函数拟合

**评估指标**：

* RMSE, MAE, R²
* 不同风速区间的拟合精度

**2.2 阶段二：多变量线性模型**

**模型形式**：

P = β₀ + β₁×U + β₂×θ + β₃×T + β₄×ρ + ε

其中：

- U: 选定高度的风速

- θ: 对应风向

- T: 温度

- ρ: 空气密度

**特征工程**：

* 风向的正弦余弦变换
* 风速的多项式项
* 变量标准化处理

**2.3 阶段三：多变量非线性模型**

**多项式扩展**：

P = β₀ + Σβᵢxᵢ + Σβᵢⱼxᵢxⱼ + Σβᵢxᵢ² + ...

包含交互项和高次项

**机器学习模型**：

* 随机森林回归
* 神经网络（MLP）
* 支持向量回归

**模型对比**：

* 交叉验证评估
* 过拟合检测
* 可解释性 vs 精度权衡

**3. 误差传播分析框架**

**3.1 数学理论基础**

**线性化误差传播**：

Var(P) ≈ Σᵢ (∂f/∂xᵢ)² × Var(xᵢ) + 2ΣᵢΣⱼ (∂f/∂xᵢ)(∂f/∂xⱼ) × Cov(xᵢ,xⱼ)

**敏感性系数计算**：

* 解析梯度（线性模型）
* 数值梯度（非线性模型）
* 局部敏感性 vs 全局敏感性

**3.2 误差源分解**

**观测误差传播**：

以观测数据为基准，分析各变量观测精度对功率估算的影响

**预测误差传播**：

分析 ECMWF 和 GFS 预测误差对功率估算的影响

对比两种预测源的误差特性

**高度选择影响**：

对比不同高度观测数据的敏感性差异

分析最优观测高度选择

**4. 敏感性分析深入研究**

**4.1 条件敏感性分析**

**分风速区间**：

* 低风速区（0-5 m/s）
* 中风速区（5-12 m/s）
* 高风速区（>12 m/s）

**分季节分析**：

* 春夏秋冬四季的敏感性差异
* 大气稳定度对敏感性的影响

**分天气类型**：

* 稳定天气 vs 不稳定天气
* 不同风向扇区的敏感性

**4.2 多维敏感性分析**

**Sobol敏感性指数**：

* 一阶敏感性：单变量直接影响
* 二阶敏感性：变量间交互影响
* 总敏感性：包含所有高阶影响

**相对重要性排序**：

量化各误差源的相对贡献：

1. 哪个变量的观测误差影响最大？

2. 哪个高度的观测最重要？

3. ECMWF vs GFS 哪个预测源更可靠？

**5. 实用化分析与建议**

**5.1 观测精度改进建议**

**定量化收益分析**：

- 风速测量精度提升0.1 m/s → 功率误差减少X%

- 风向测量精度提升1° → 功率误差减少Y%

- 温度测量精度提升0.5°C → 功率误差减少Z%

**成本效益评估**：

* 不同传感器升级的优先级排序
* 投入产出比分析

**5.2 预测数据优化建议**

**数据源选择指导**：

- 不同气象条件下 ECMWF vs GFS 的选择建议

- 多模式集成的权重优化

- 偏差订正的重点方向

**6. 研究创新点总结**

**6.1 方法论创新**

* **系统性框架**：完整的多源、多高度误差传播分析
* **垂直梯度敏感性**：首次系统分析不同观测高度的影响
* **条件化敏感性**：不同气象条件下的敏感性变化规律

**6.2 实用价值**

* **工程指导**：具体的观测精度改进建议
* **预测优化**：数值预报产品的使用指导
* **成本优化**：传感器升级的优先级排序

**6.3 数据优势**

* **多源对比**：观测、ECMWF、GFS三源完整对比
* **时空完整**：548天连续数据，四个观测高度
* **高质量验证**：真实功率数据验证

**7. 实施计划**

**Phase 1 (1-2周)：数据探索与预处理**

**Phase 2 (2-3周)：功率模型构建与对比**

**Phase 3 (3-4周)：误差传播分析实现**

**Phase 4 (2-3周)：敏感性分析与结果解释**

**Phase 5 (1-2周)：实用化建议与论文撰写**

差异最小的是哪儿种条件呢？

差异最大的是哪儿种条件呢？

哪儿种条件下对power的映射最明显？

已经知道的：  
1）10m 与power最为相关  
2）estimate power 通过shap解释两种机器学习模型，给出的特征贡献较为相似，凸显了10m风速重要。

3）比较特殊的月份1月，