

第一章 绪论

课程性质和任务

气象统计方法是大气科学专业的一门专业主干课，它是在概率统计、线性代数等课程基础上，学习利用统计方法对气象特征及成因进行诊断或对气象现象进行预报的方法。通过本课程的教学，使学生能够掌握对天气、气候现象进行诊断及预报的常用统计学方法的基本原理和技术，包括挑选与统计检验预报因子、建立预报工具、统计与分析气象资料中的统计规律，并且培养学生运用所学的方法与技能从事科学研究的能力，为将来从事实际气象业务预报及科学研究工作打下基础。

1.1 统计诊断方法

1.1.1 定义与重要性

统计学方法的重要性

上世纪 50-60 年代的论文，图像以手绘为主，受限于计算机能力限制，资料无法充分利用，统计方法应用不多。自 20 世纪 70 年代来，统计学方法在包括大气科学在内的各个领域得到广泛应用。大气运动有**确定性的规律**，也具有**随机性**，但对于**月、季以上时间尺度的天气气候特征和变化规律**，其统计学方法仍然是主要研究方法之一。天气预报中的具体要素预报仍然以统计学方法为主。

需要注意：统计预报是数据驱动的预报方法，数值预报是基于动力学方法的预报。

统计学 研究关于变率、**不确定性**及面对不确定性时如何决策的学科。评估和量化不确定性，也包括针对不确定性做推断和预报，是统计学的重要组成部分。在大气科学中，包括描述性统计和推断性统计。目前的预报都是概率预报。

诊断分析 包括**统计诊断与动力诊断**，统计诊断是 20 世纪 70 年代发展的一门统计学分支，对统计建模和统计推断过程进行诊断。实际应用中，两者应当相辅相成，只有统计诊断无法令人信服。

三板斧：统计诊断→动力分析→数值模式，是优秀论文的一般范式。

- ① 它以收集到的**大量数据**为基础，建立经验公式或推断模型，并对其合理性给出判断。
- ② 通过一些统计量来检查**数据、模型、推断方法**中可能存在的病患问题，提出治疗解决的方法。
- ③ 为克服模型与客观实际之间可能存在的差异，需要一种**诊断方法**，判断实际数据和模型之间是否存在较大偏差，并采取相应对策。

数据的至关重要性

以台风资料为例，对于统计强台风个数趋势这一相同课题的论文，使用来自日本和美国的资料，得到的结论截然相反，且每个人都给出了合理的物理解释。其中至少有一个是错误的，这体现了研究中数据的至关重要性，也启示我们阅读文献时需要具有批判性思维。

1.1.2 方法的内容

气候诊断 统计诊断分析方法是一种直观的认知事物的手段。

研究内容

- ① **要素本身**：应用统计方法了解某要素的时空分布、变化规律、气候异常等，即**充分分析研究对象**。
- ② **要素之间**：探索要素与其他因素之间的联系，研究造成该相关性的**成因**及其背后的**物理机制**。
- ③ **要素比较**：**模式验证**，对数值模拟结果与实际变化状况之间差异进行诊断。

台风案例

全球每年台风生成个数稳定在 80~90 个，无论全球气候如何变化。西北太平洋占据三分之一，这

是要素本身。随后，分析哪些因素会影响台风，比如水汽、海温、季风槽（高正涡度）等，这些因素的改变导致了台风的变化。最后，进行敏感性数值模式实验以验证结论。

诊断步骤

对天气、气候现象进行统计诊断分析，一般分为四步：

- ① **收集资料**：从研究的实际问题出发，确定诊断对象，收集资料。资料应该**准确性、精确性、均一性、代表性和比较性**。（台风定强资料的主观推导产品具有很大的不确定性，40 年来来技巧无明显提升）
- ② **资料预处理**：根据研究问题的具体需要做预处理，如距平化、标准化等。
- ③ **选择诊断方法**：根据研究目的和研究对象，选择合适方法（距离分析等）。
- ④ **科学综合和诊断**：对结果应该进行**显著性检验**，将统计计算结果和气象知识结合起来，进行科学综合和细致分析，了解其可信程度。**没有物理意义的结果是失真的，无分析价值。**

1.2 气候统计预测

概念

利用统计模型对气候系统的**未来变化状态**作出估计。统计模型是在利用大量过去气候资料对气候系统内部或与其它变量之间关系的变化规律及特征分析基础上建立的。统计模型是基于过去的资料，却用于预测未来，这种外推的可靠性需要特别注意。

基本假设

- ① 气候变化的**成因与物理机制**至少**在预测期间与观测一致**。统计规律在时间上的一致性
- ② **气候系统在预测期间保持稳定**。气候变化中存在突变
南大西洋这一没有台风出现的地方出现台风，当然有大量论文探讨机制，但也不能忽略观测技术提升带来的数据敏感度变化。

基本要素

- ① **预测对象**：例如**某区域（局地、全球）、某时期（季节、月）**的旱涝趋势、冷暖趋势。
注意：研究年变化至少需要区分为冬夏两个季节，否则全年平均会淹没信息。
- ② **预测依据**：通常为从某些**统计上显著相关**的预报因子群提取的有效信息。
- ③ **预测技术**：根据数据性质、预测对象和预测因子特点，选择合适的**统计预测模型**。（线性/非线性）
- ④ **预测结果**：对未来气候变化状态时间、空间、数量、性质等方面的预测。

预测模型的选择

需要研究因变量和自变量的关系时，不应该直接计算线性相关系数并剔除相关系数小的因素，最好先作图分析，人工评估是否存在二次项等非线性关系。

预测技术

- 定性预测**：主观性强，综合分析过去、现在、未来各种因素之间的可能相互影响情况，寻找气候要素发展规律，推断出气候要素未来的发展趋势。（例如 SST 越高，台风越多）
- 定量预测**：技术更为复杂，需要基于统计数学模型或**数值模式**。（例如 SST 增加 1℃，台风增加 7%）
方法有时间序列模型、多元回归模型、动态系统模型、变量场预测方法、神经网络等。

五大特性

气候变量满足五大特性（五大特性不一定同时满足）：

- ① **持续性**：气候变量的历史变化中升降趋势的持久程度。
- ② **相似性**：气候变化在某一时期与另一时期变化形式相似。
- ③ **周期性**：气候变化趋势经过一定时间间隔后重复出现，ENSO 循环、波动震荡、太阳辐射等。
- ④ **最大最小可能性**：变化数量，变量历史变化的数量在一定时间内有适当范围，给出历史变化频率。
- ⑤ **转折点**：气候变量变化中某一时期明显的特征，发生质的突变（均值、方差、斜率等）。

基本步骤

收集资料→建立统计模型→**统计检验**（任何统计结论都需要统计检验）→预测结论

统计优点

- ① 可以从历史资料中概括出天气演变的统计性质，找出现象之间的联系。
- ② 与动力学结合，从大量的统计规律中总结发现必然的自然规律。
- ③ 可以利用少数台站资料，做出大范围预报。

统计缺点

- ① 方法无法预报**历史资料中没有出现过的天气**。
- ② 预报效果的好坏与**变量和资料的数量、质量**有关（大数定律大于 30 年）。
- ③ 预报量与因子之间的**统计相关性关系**，由于气候变迁等原因也会发生改变。