

HPCA-AI 博士组会 2019-11-12

资料（数据）同化

(一) 资料同化能做什么？

气象观测数据过少（，无法建立足够精确的初值数据供模式使用。）

利用现有信息或观测资料（包括模式信息、观测信息及其他先验信息），来定义一个最大可能精确（符合实际情况）、与模式协调的大气状态（即获得输入模式的初值）。

最核心的问题：同时降低初值与「模式」和「实际情况」的误差。

(二) 效果

资料同化是决定了预报效果的根本要素。

初值效果对预报的影响非常大。

欧洲中心的资料（数据）同化系统最好，生成的初值非常好。

(三) 怎么做

目标函数的极小化问题，通过同化让初值与模式（模拟的情况）和观测数据（实际的情况）的轨迹最近。

核心问题：

- 最小方差，最小二乘(从函数角度进行优化)
- 最大似然 (从概率角度进行优化)
- 贝叶斯理论 (从概率角度进行优化)

B: 背景误差协方差矩阵 (高维度的稠密矩阵，计算量非常大) –
模式模拟性能的误差(模式与真实情况的误差)

R: 观测误差协方差矩阵 (对角矩阵，稀疏矩阵)

H: 算子误差

解决 B 的问题最重要，用以下方法去估计 B，这实际上是数据降维问题（通过数据降维来降低计算量）：

a) HL 方法

b) MMC 方法

时间轴、多级互补，使用流函数、似函数假设替代原数据，将稠密矩阵转化为稀疏矩阵。

c) 集合预报 (Ensemble Forecast) 方法

计算量的产生与「局对化」有很强的关系。
每个网格点都要做「局对化」

选择 20 组初值，并行地输入模式进行计算；并行地作为下一个阶段模式的初值。每次需要获得预报值的时候，取平均值。
随着时间的演变，离散度会越来越小，每隔一段时间后要冷启动。

* 对 B 的处理，决定了分析场的好坏。

目前的资料同化理论，已经无法适应高分辨率模式了。

「高分辨模式与实际情况的误差」增大，「原有的同化资料与高分辨模式的误差」也会增大，两个误差会叠加，进一步放大「资料同化 (理论)」与「模式」的不匹配。

(四) 已有思路:

1) 将卷积替换 BOX 切分，看能否解决边界问题

2) 历史数据选择集合版本

从历史库中找到有足够代表性样本，作为集合预报的样本

相似形

气候气象大数据: 过去 50 年的数据。

如果能找到，效果会很好。但是很难找到拥有足够的样本。

3) Inflation 误差

自适应的添加误差与阶段。

阶段性累加误差，延长离散度。

4) 完成一个标准化的测试系统