

基于卷积神经网络的中国夏季日最高气温延伸期预报模型——版本 V01

实施人：谢洁宏^{1,2}，徐邦琪¹

1. 南京信息工程大学大气科学学院 2. 广东省揭阳市气象局

第一实施人联系方式：jiehongx@foxmail.com

完成时间：2021-06

- 预报模型实施路线：**挑选与预报量（中国夏季 10~90 d 日最高气温）具有物理联系的预报因子，利用非传统带通滤波方法（Hsu et al, 2014）提取可实时获取的要素低频场，基于超前滞后相关探索延伸期（提前 5~30 d）可预报性来源，并基于高相关区域选定关键区预报因子作为特征输入，利用卷积神经网络（Convolutional Neural Network, CNN）挖掘学习前兆信号与后期气温演变的映射关系，构建经验预测模型——CNN 模型（CNNM）。实施路线详见图 1.1。

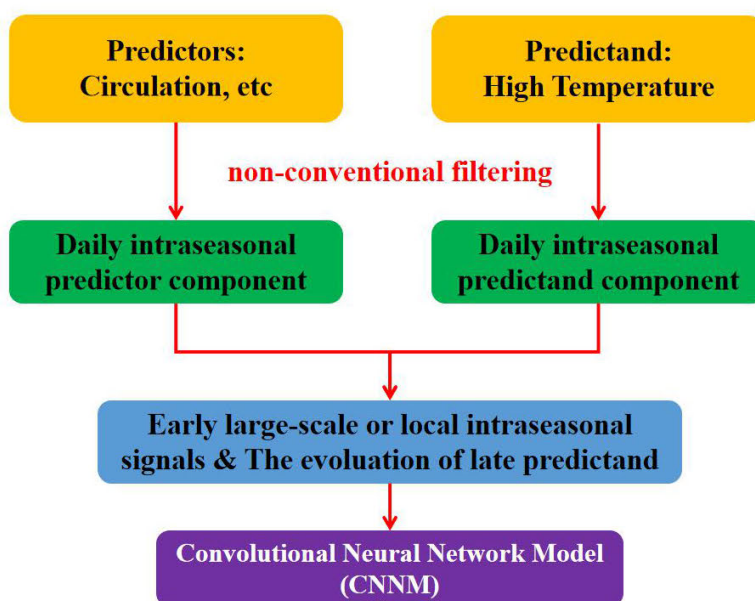


图 1.1 基于 CNN 的中国夏季日最高气温延伸期预报模型实施路线

- 预报对象：**北半球夏季（5-10 月）中国日最高气温 10~90 d 低频分量，由预测其前 6 个主

模态时间系数重构所得，主模态基准期为 1961~2019 年夏季，前 6 模态累计方差贡献约 75%。30~90 d、10~30 d 与 10~90 d 日最高气温前 6 主模态与方差贡献如图 1.2 所示。经对比(图略)，直接预报 10~90 d 分量技巧较优于 10~30 d 与 30~90 d 分量的组合预报(10~30 d + 30~90 d)，因此实时预报中直接对 10~90 d 分量场进行预报。

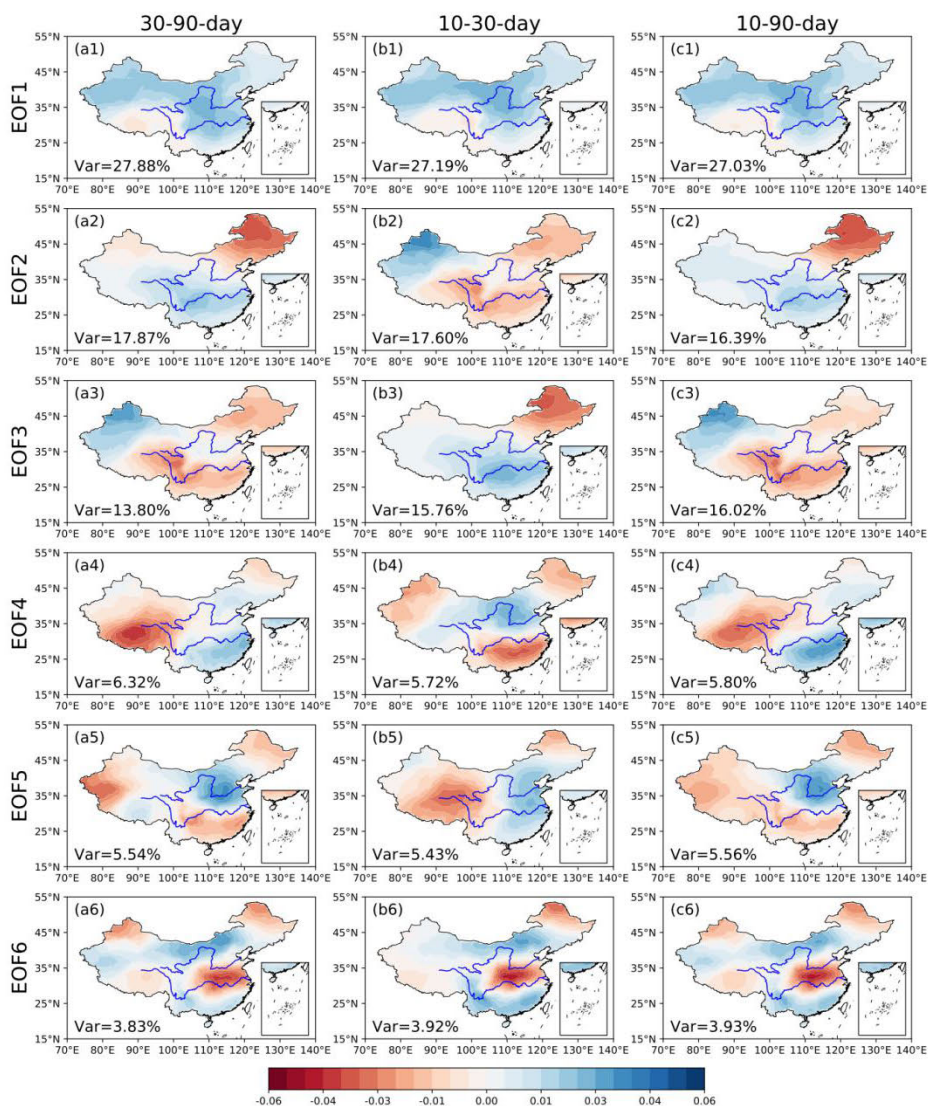


图 1.2 基准期中国夏季日最高气温 (a1-6) 30~90 d、(b1-6) 10~30 d 与 (c1-6) 10~90 d 分量主模态空间分布及对应方差贡献 (Var)

3. 预报因子：海平面气压 (SLP)、925 hPa 温度场 (T925)、850 hPa 高度场 (H850)、500 hPa 垂直速度 (W500)、700 hPa 比湿 (RH700)，经超前滞后相关选取预报因子关

键区域，并测试对应输入特征的模型预测表现，最终选取空间范围为-10~80°N, 40~180°E。

4. **预报思路：**利用前 6 候低频预报因子预测后 6 候气温 10~90 d 分量主模态时间系数，预报因子与预报量之间的映射关系由 CNN 挖掘并学习。

5. **CNNM 结构：**由卷积层、最大值池化层与全连接层组件构成，各组件间包含构建非线性映射关系的激活函数层。卷积操作是 CNN 最基础但最核心的操作，起特征提取作用，可从输入特征（前 6 候预报因子）中提取更深层次的隐性特征；池化是一种降采样操作，可降低特征图维度（分辨率），而最大值池化通过取最大值的方式，保留了指定区域的最显著的信号并进行去噪；全连接层在整个 CNN 中起“分类器”作用，而激活函数层用于构建输入与输出间的非线性映射关系；卷积层、池化层和激活函数层等操作是将原始数据映射到隐层特征空间，全连接层起到将学到的“分布式特征表示”映射到样本标记空间的作用。本研究所研发 CNNM 具体结构如图 1.3 所示。

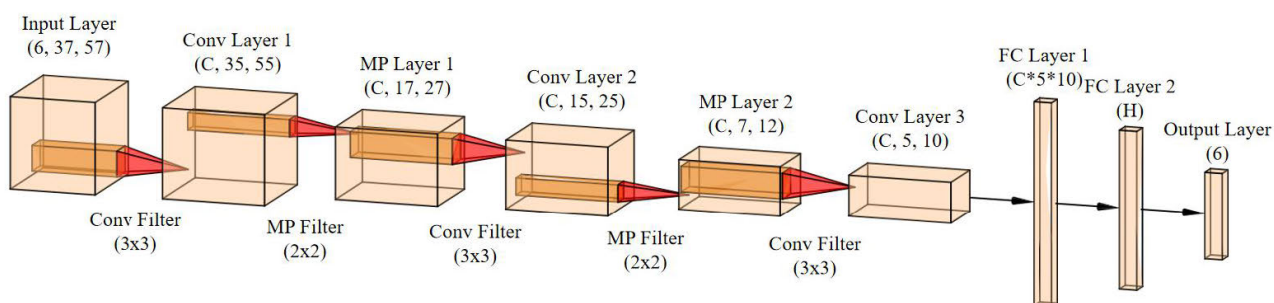


图 1.3 所构建 CNNM 结构，包含 3 层卷积层（Conv）、2 层最大值池化层（MP）与 2 层全连接层（FC）。各组件间使用激活函数构建输入与输出间的非线性映射关系。输入输出张量和滤波器（卷积核、池化核）对应维度已于各组件旁进行标识，维度描述中“C”与“H”分别代表卷积核数与全连接隐藏层数

6. **建模思路：**每种预报因子分别构建 4 类不同配置的网络，区别在于卷积层卷积核数（C）与全连接层中的隐藏层数（H）。具体配置分别为 C81H81（即 81 个卷积核、81 个隐藏

层，下同）、C81H64、C64H81、C64H64，每类配置构建 8 个随机种子（对应 8 个不同的网络初始权重，可类比数值模式集合成员具有不同初始扰动），最终预测结果为多预报因子、多配置、多种子的等权重集合。

7. 气候态、异常与低频分量的计算——使用非传统带通滤波

- 7.1. 以 1961~2019 年作为气候态基准期，将前 2 谐波分量作为逐日年循环。
- 7.2. 原始场减去对应年循环得到异常，异常减去其前 45 d 的平均（day -45 ~ day 0）以去除 90 d 以上甚低频扰动，保留 90 d 以下季节内异常。
- 7.3. 将经上述步骤获取异常进行前 5 d 平均（day -5 ~ day 0）以滤除 10 d 以下天气尺度高频扰动，进而获取较纯净的 10~90 d 低频分量。

8. 高温热浪监测指标：

- 8.1. 对逐日日最高气温进行 5 d 滑动平均，以滑动前 35°C 以上高温日数的百分比确定滑动后的对应的高温阈值。
- 8.2. 滑动后区域内某日最高气温 $\geq 35^\circ\text{C}$ 的格点数占当日全区百分比 P_0 ($P_0=20\%$)，同时区域平均最高气温大于等于本区域 80 百分位值判定为区域高温日。
- 8.3. 若前一日满足条件 8.2，当日满足条件 8.2 且与前一日格点重合度大于等于 50%，则把当日判定为区域高温持续日。
- 8.4. 若区域高温日之后紧接着 2 d 或 2 d 以上高温持续日，即连续高温日达到 3 d 及以上，则判断为一次区域持续性高温过程。
- 8.5. 按上述定义的持续高温过程，标记为_0；若两个持续过程间隔 1 d，视作同一个持续过程，标记为_1；若两个持续过程间隔 2 d，视作同一个持续过程，标记为_2。

9. **监测区域：**华南 (Huanan, 21~26°N, 105~120°E)、长江 (ChangJiang, 26~32°N, 105~122°E)、黄淮 (Huanghuai, 32~37°N, 105~122°E)、华北 (HuaBei, 37~43°N, 105~122°E)

参考文献:

Hsu et al., 2014, A spatial-temporal projection model for 10-30 day rainfall forecast in South China. *Clim Dyn.*