

# 数据同化

$u_k$ :  $k$ 时刻下模拟出的场

$v_k$ :  $k$ 时刻下观测值

$u^{DA}$ : 某时间窗口下同化后的场

$M$ : VOCs模拟

$H$ : 场到观测点的映射关系

$$u_{k+1} = Mu_k$$

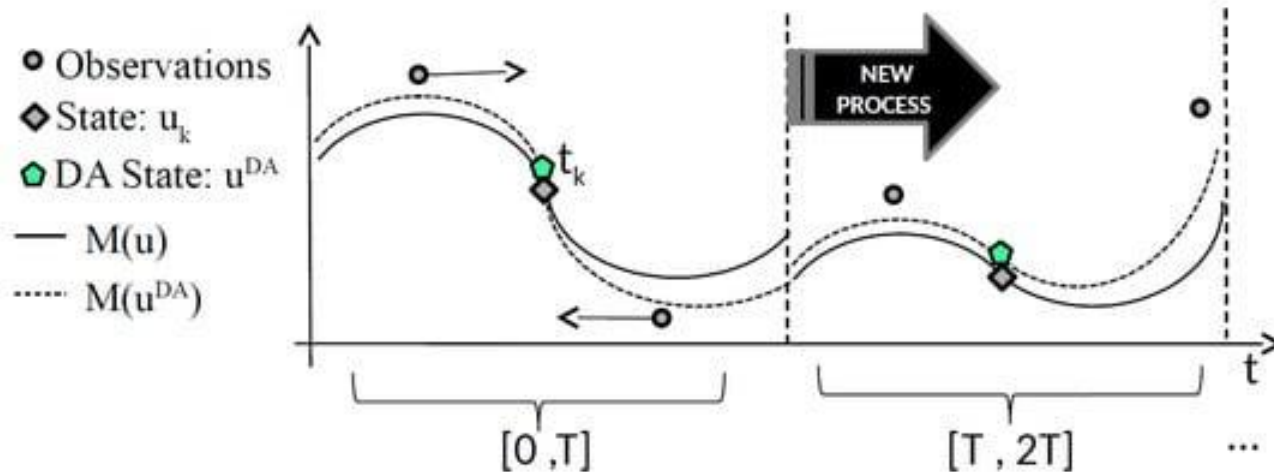
$$v = Hu + \text{err}$$

$$u_k^{DA} = \underset{u}{\operatorname{argmin}} \left\{ \|u - u_k\|^2 B_k^{-1} + \|v_k - Hu\|^2 R_k^{-1} \right\}$$

VOCs模拟模型误差协方差矩阵

观测误差协方差矩阵

数据同化：在观测值和模拟值中间找一个中间状态，综合考虑模拟的结果和观测结果，作为下一个时间步的初始输入，以获得下一个时间步更准确预测结果。



$B_k$  与  $R_k$  的计算:

$$R_k = \sigma^2 I$$

$$B_k = V_k V_k^T$$

$$V_k = \{V_{jk}\}_{j=1,2,\dots,nk}$$

$$V_{jk} = u_j - \bar{u}$$

$$\bar{u} = \operatorname{mean}\{u_j\}_{j=1,2,\dots,nk}$$

# 数据同化

$u_k$ :  $k$ 时刻下模拟出的场

$v_k$ :  $k$ 时刻下观测值

$u^{DA}$ : 某时间窗口下同化后的场

$M$ : VOCs模拟

$H$ : 场到观测点的映射关系

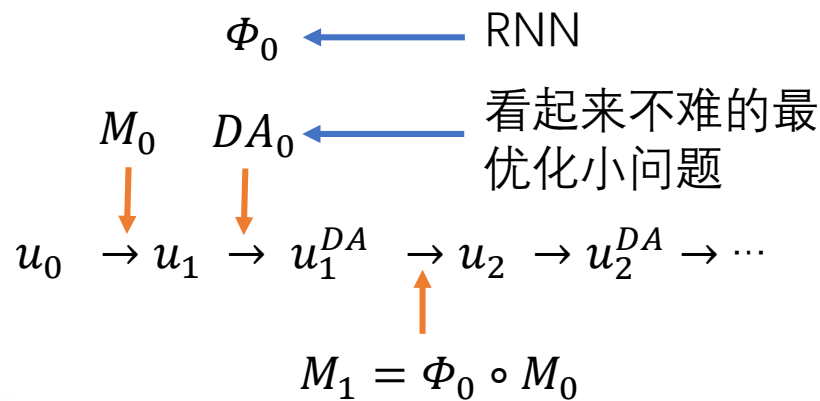
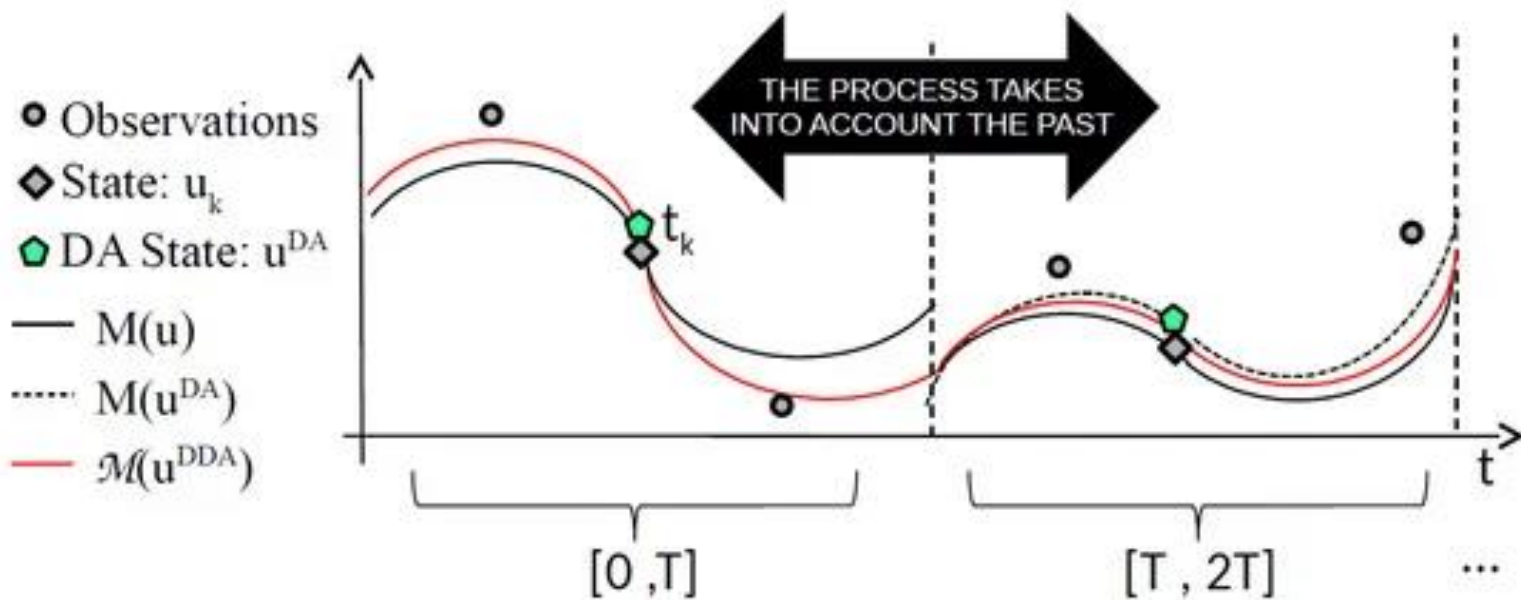
$$u_{k+1} = Mu_k$$

$$v = Hu + \text{err}$$

$$u_k^{DA} = \text{argmin}_u \{ \|u - u_k\|^2 B_k^{-1} + \|v_k - Hu\|^2 R_k^{-1} \}$$

$\Phi$ : 数据同化

$$M_{k+1} = \Phi_k \circ M_k$$

$$u_k^{DA} = \Phi_k u_k$$


# 数据同化：模拟方法M

## 使用LBM进行VOCs模拟的数据同化

解决了之前提出的几个问题:

- ✓ 加快LBM模拟：支持模拟内部切换外部状态信息，优化掉不必要的文件I/O。
- ✓ 降低输入数据规模：完成污染物场3D->2D映射程序开发

仍然还存在一些问题:

- 数据规模还是太大：B很难计算与存储，同化过程维数高
- $u \rightarrow v$ 之间映射关系非线性/不连续，同化过程本身不易求解

## 使用机器学习进行VOCs模拟的数据同化

- ✓ VOCs模拟快
- ✓ 数据规模小，同化涉及的矩阵容易计算和存储，例如：同化过程中的矩阵B可以轻松放得下
- ✓ 可以进一步使用深度学习+同化的思路，持续矫正模拟过程（不仅仅优化初始场）

## 短期计划

- 完成同化过程的最优化问题求解器开发
- 初步搭建神经网络 $\Phi$ 的训练模型，利用同化求解器产生模型训练数据
- 思考与调研：后续如何部署该系统？实现随时间推移，新的观测数据进来，自动训练 $\Phi_k$ 并更新 $M_{k+1}$ （考虑NLP领域相关问题与解决方案）