

# 1. 决策树模型

---

## 1.1 西班牙数据集

---

西班牙数据包含的特征属性: year, month, day, wind\_power, wind\_direction, wind\_speed

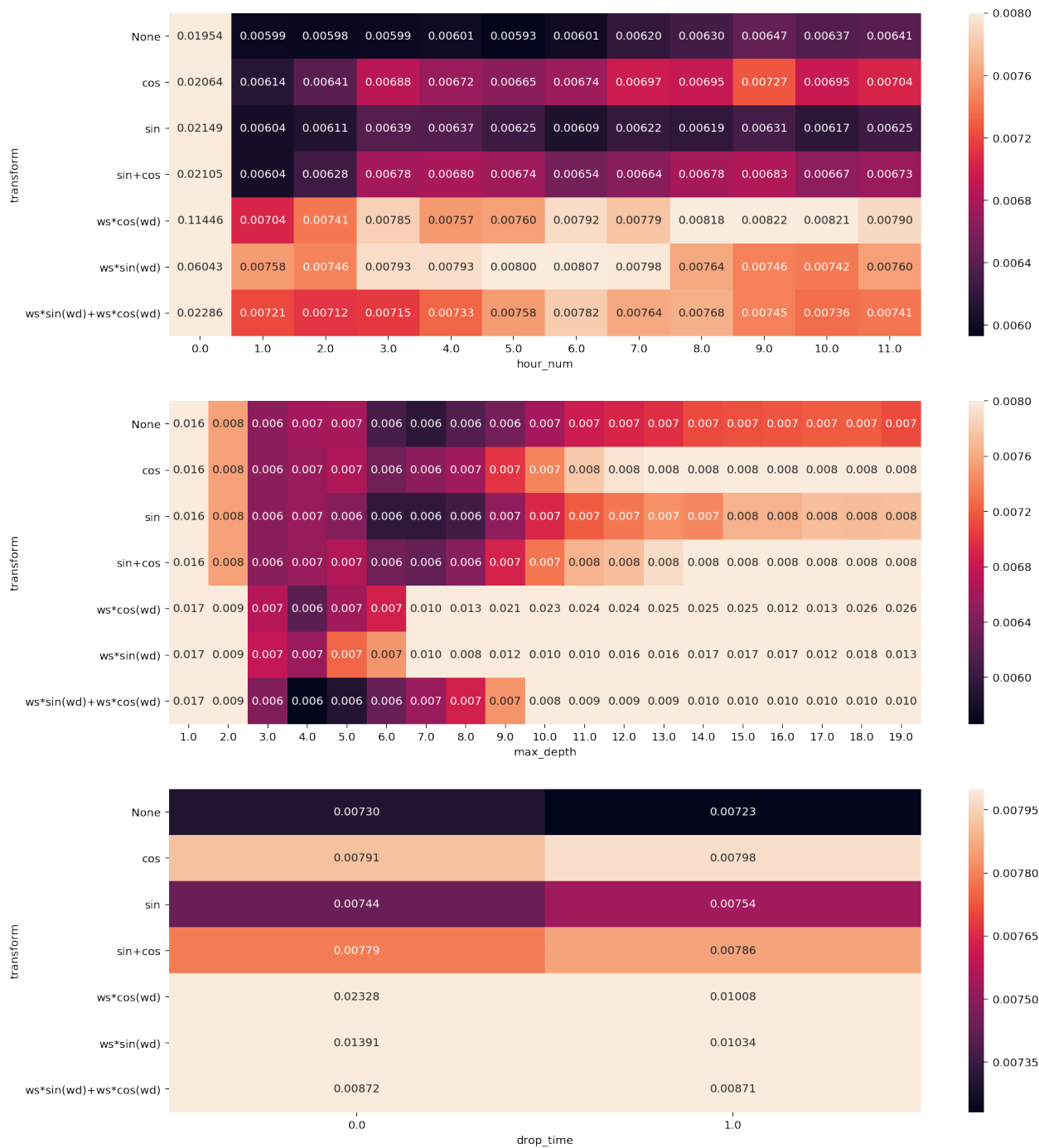
训练集: [6426, 10426] len: 4000 (小时)

测试集: [14389, 15389] len: 1000 (小时)

输入说明: (所有数据已归一化)

1. **hour\_num**: t-L 中的L
2. **transform**: 对输入的 wind\_direction, wind\_speed 进行转换  
{ None: 无转换 ws wd  
'sin': wd sin(wd)  
'cos': wd cos(wd)  
'sin+cos': wd sin(wd) cos(wd)  
'ws\*sin(wd)': wd\*sin(wd)  
'ws\*cos(wd)': wd\*cos(wd)  
'ws\*sin(wd)+ws\*cos(wd)': wd\*sin(wd) wd\*cos(wd)}
3. **drop\_time**: 是否删除时间特征 ['Year', 'Month', 'Day', 'Hour']

对不同的输入形式, 不同的决策树最大深度 **max\_depth** 进行测试可得:



可以看到:

1. L 为 0 时效果较差, 为1或2时效果较好
2. 数据转换在

- 1 **None**: 无转换 ws wd
- 2 **'sin'**: wd sin(wd)
- 3 **'cos'**: wd cos(wd)
- 4 **'sin+cos'**: wd sin(wd) cos(wd)

上的表现较佳, 其中 None 和 'sin' 的表现优于 'cos' 和 'sin+cos', 而在其他四种组合方式上表现较差。

3. 是否添加时间属性 (drop\_time) 对结果没有较大影响

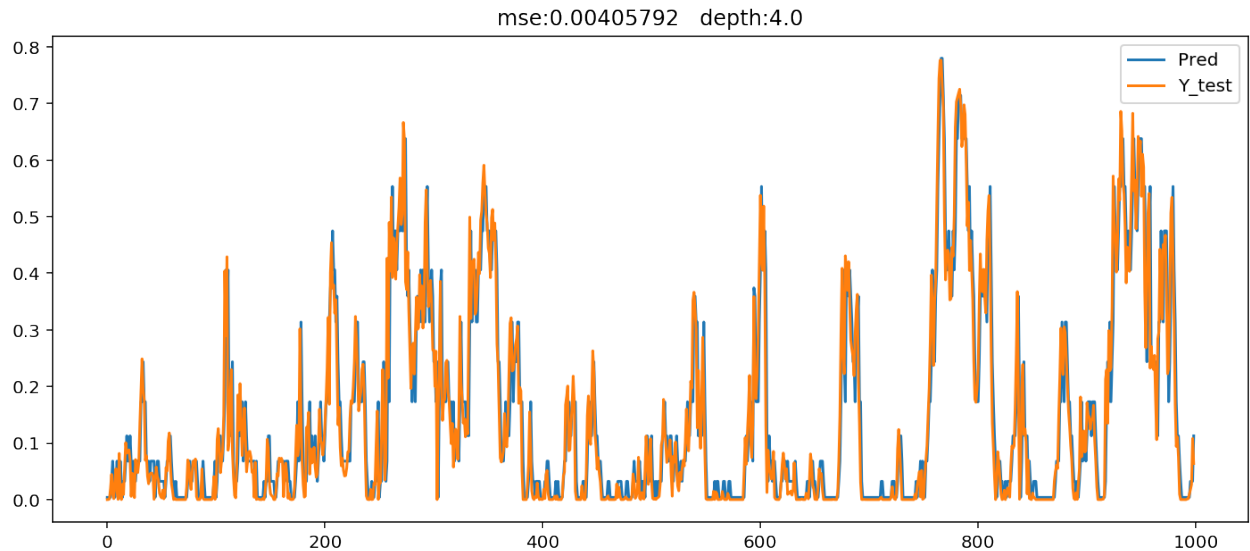
可以得到最优的参数组合为:

```

1 {'drop_time': 1.0,
2   'hour_num': 1.0,
3   'max_depth': 4.0,
4   'transform': 'ws*cos(wd)'}

```

对应最优 mse 为 0.00405792，预测图如下所示：



## 1.2 美国数据集

美国数据包含的特征属性：year, month, day, minute, wind\_power, wind\_direction, wind\_speed, air\_temperature, surface\_air\_pressure, density

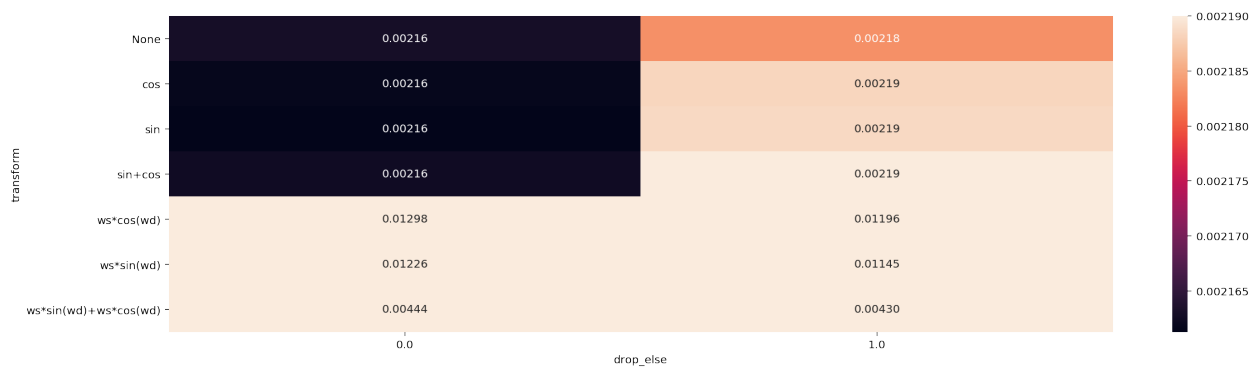
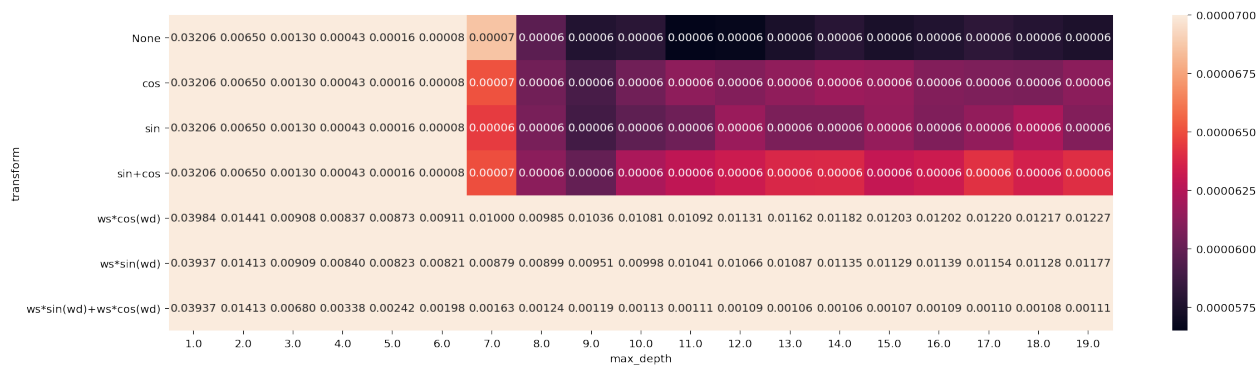
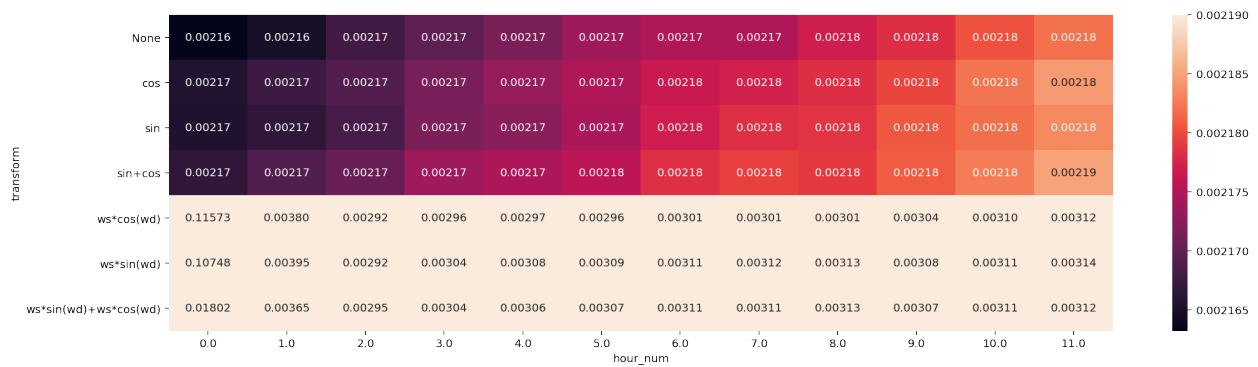
训练集：[6426, 10426] len: 4000（分钟）

测试集：[12000, 13000] len: 1000（分钟）

输入说明：（所有数据已归一化）

1. **hour\_num**: t-L 中的L
2. **transform**: 对输入的 wind\_direction, wind\_speed 进行转换  
 { None: 无转换 ws wd  
 'sin': wd sin(wd)  
 'cos': wd cos(wd)  
 'sin+cos': wd sin(wd) cos(wd)  
 'ws\*sin(wd)': wd\*sin(wd)  
 'ws\*cos(wd)': wd\*cos(wd)  
 'ws\*sin(wd)+ws\*cos(wd)': wd\*sin(wd) wd\*cos(wd)
3. **drop\_time**: 是否删除时间特征 ['Year', 'Month', 'Day', 'Hour', 'Minute']
4. **drop\_else**: 是否删除其他特征 ['air\_temperature', 'surface\_air\_pressure', 'density']

对不同的输入形式，不同的决策树最大深度 **max\_depth** 进行测试可得：



可以看到:

1. L 为 0 时效果最好, 随着 L 的增长效果逐渐变差
2. 数据转换在

1. **None**: 无转换 ws wd
2. **'sin'**: wd sin(wd)
3. **'cos'**: wd cos(wd)
4. **'sin+cos'**: wd sin(wd) cos(wd)

上的表现较佳, 其中 None 的表现优于 'sin', 'cos' 和 'sin+cos', 而在其他四种组合方式上表现较差。

3. 不添加时间属性 (drop\_time=True) 时效果较佳
4. 不添加其他属性 (drop\_else=True) 时效果较佳

可以得到最优的参数组合为：

```
1 {'drop_else': 0.0,  
2  'drop_time': 1.0,  
3  'hour_num': 0.0,  
4  'max_depth': 10.0,  
5  'transform': 'sin+cos'}
```

对应最优 mse 为  $2.5709867 \times 10^{-5}$ ，预测图如下所示：

