

1. esn+决策树模型

1.1 西班牙数据集（时间间隔一小时）

1.1.1 对输入的形式以及决策树最大深度进行搜索

1.1.2 对 esn 参数进行搜索

1.2 美国数据集（时间间隔五分钟）

1.2.1 对输入的形式以及决策树最大深度进行搜索

1.2.2 对 esn 参数进行搜索

1. esn+决策树模型

1.1 西班牙数据集（时间间隔一小时）

西班牙数据包含的特征属性：year, month, day, wind_power, wind_direction, wind_speed

1.1.1 对输入的形式以及决策树最大深度进行搜索

训练集：[6426, 10426] len: 4000（时间间隔1小时）

测试集：[14389, 15389] len: 1000（时间间隔1小时）

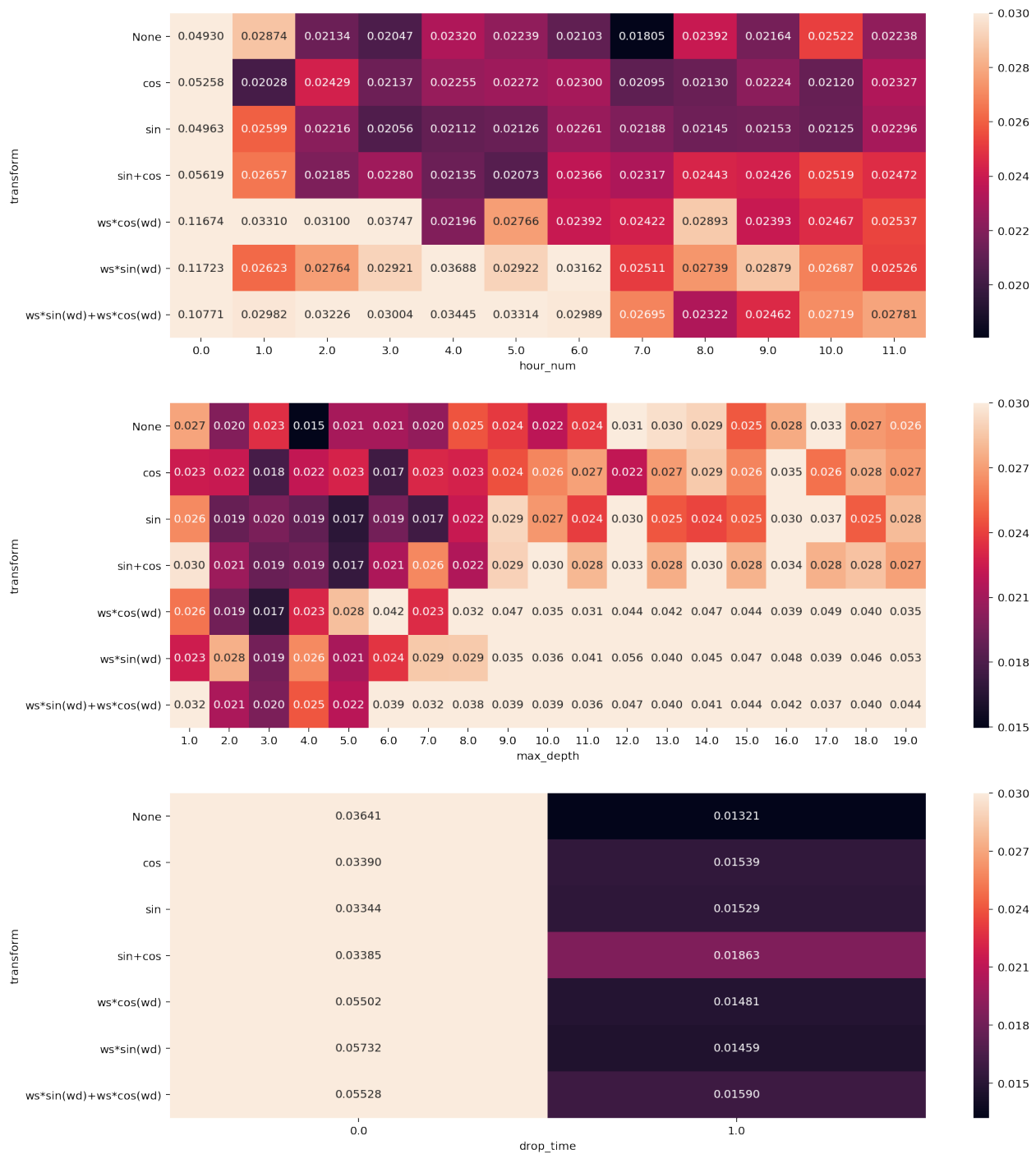
输入说明：（所有数据已归一化，时间间隔1小时）

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']

搜索范围：

```
1 param_grid1 = {'transform': [None, 'sin', 'cos', 'sin+cos', 'ws*sin(wd)',  
2                               'ws*cos(wd)', 'ws*sin(wd)+ws*cos(wd)],  
3                   'hour_num': np.arange(0,12),  
4                   'drop_time': [True, False],  
5                   'max_depth': np.arange(1,20)}
```

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：



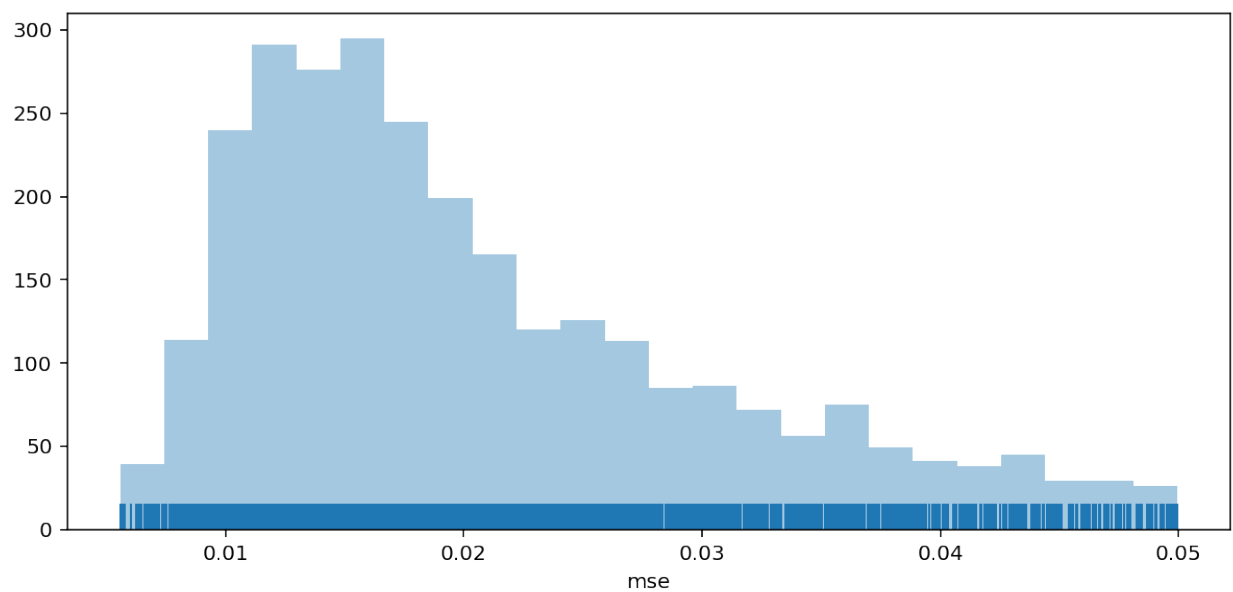
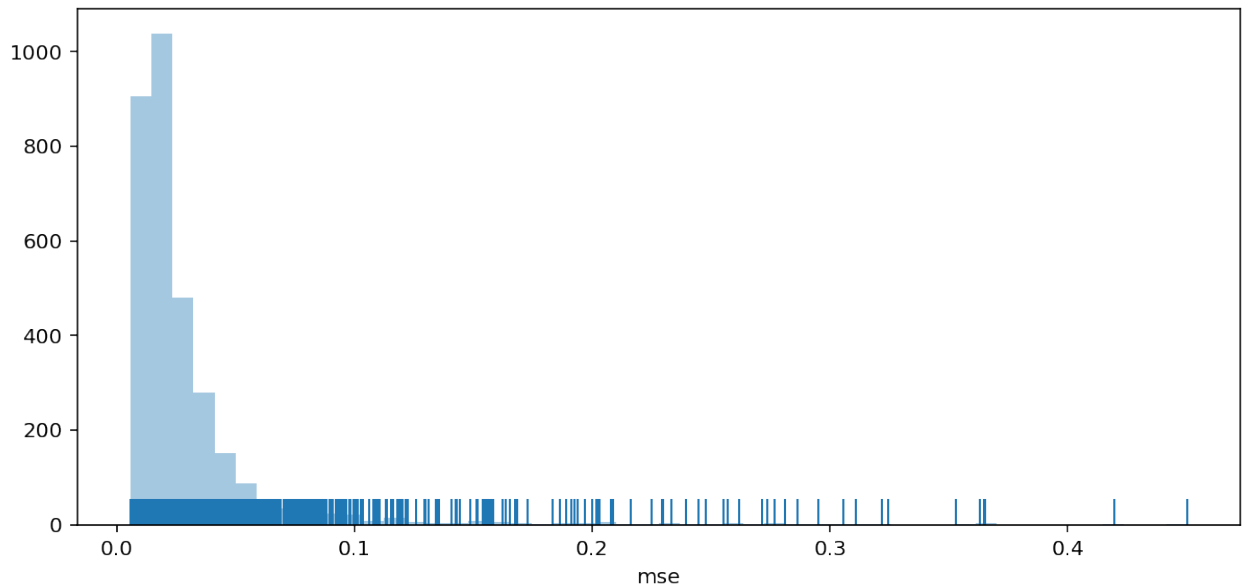
可以看到:

1. L 为 0 小时时效果较差，为3~10时效果较好
2. 数据转换在【同决策树模型】

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 、'sin' 和 'cos' 的表现优于 '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. 所有测试 mse 分布如下:

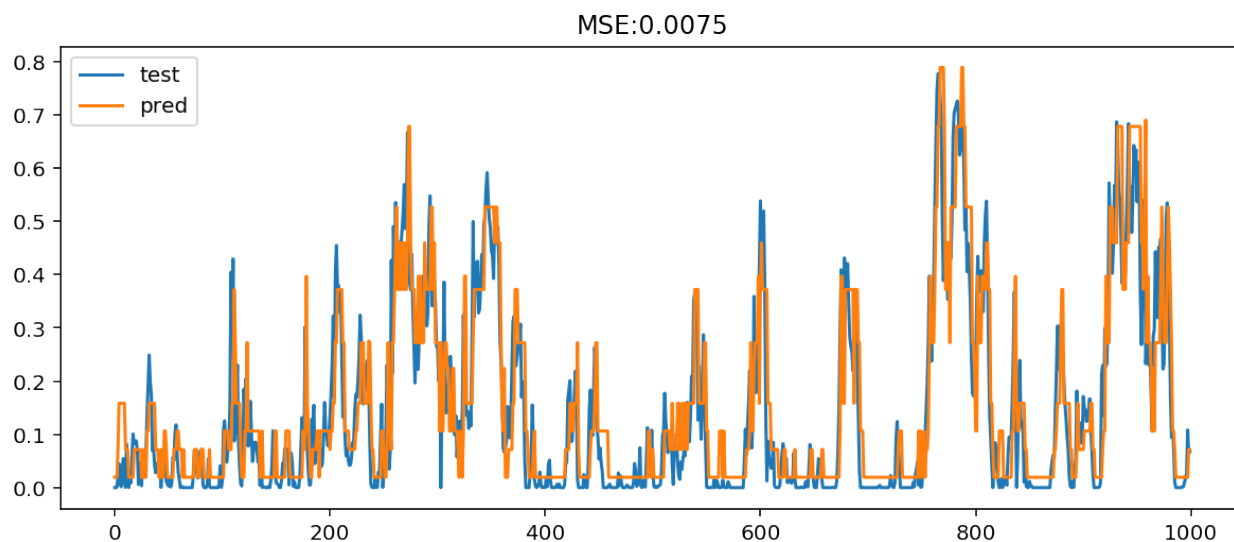


其中最优的参数组合为：

```
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 0.0055729145，预测图如下所示：



1.1.2 对 esn 参数进行搜索

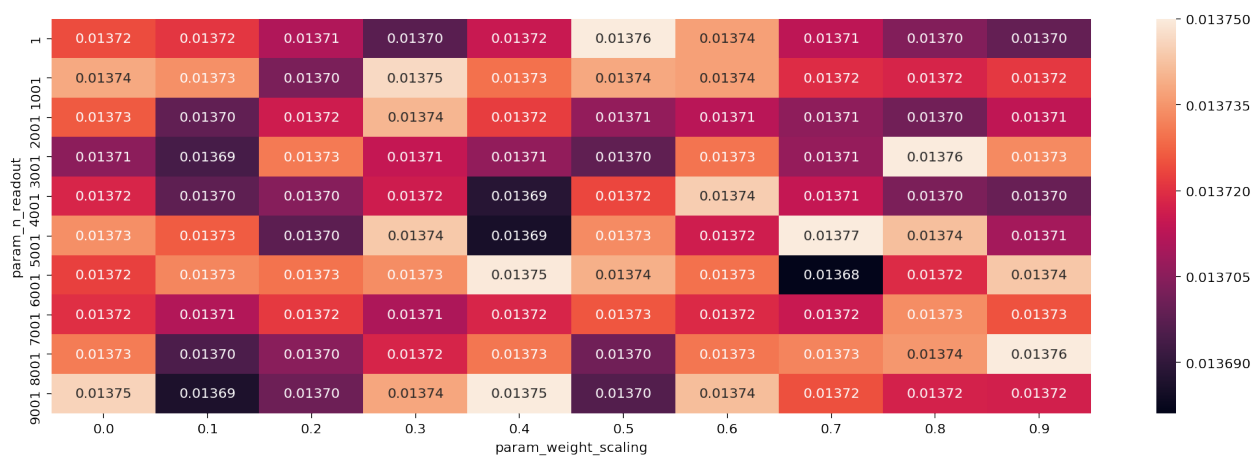
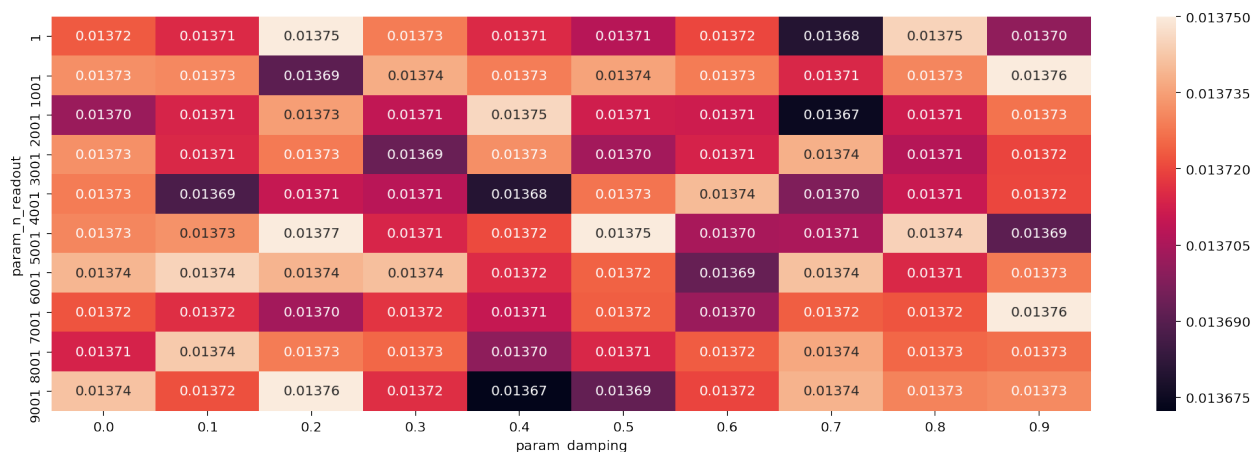
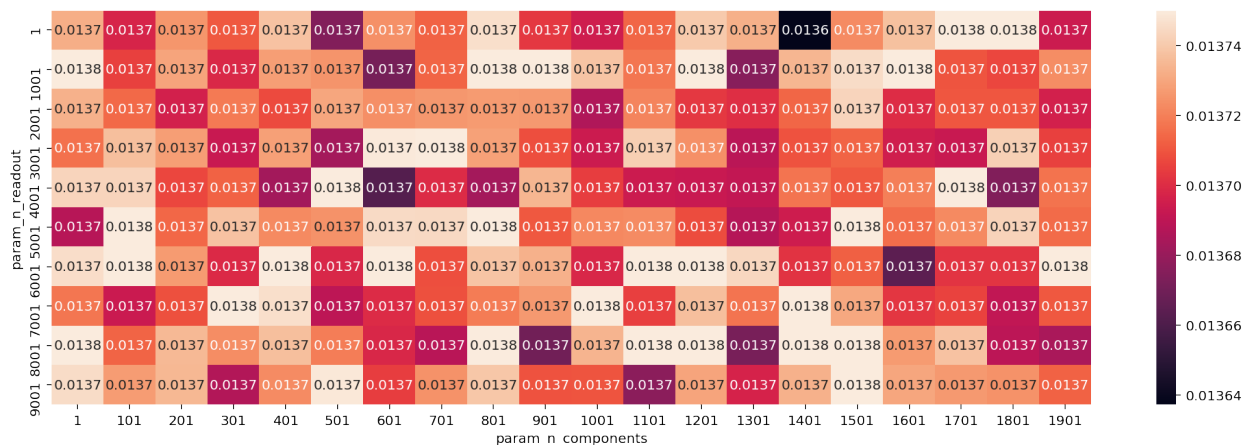
接下来在 1.1.1 的最优参数基础上，对 esn 参数进行网络搜索，搜索范围如下所示：

```
1 esn_param = {  
2     'n_readout': np.arange(1,10000,1000),  
3     'n_components': np.arange(1,2000,100),  
4     'damping': np.arange(0,1,0.1),  
5     'weight_scaling': np.arange(0,1,0.1)}
```

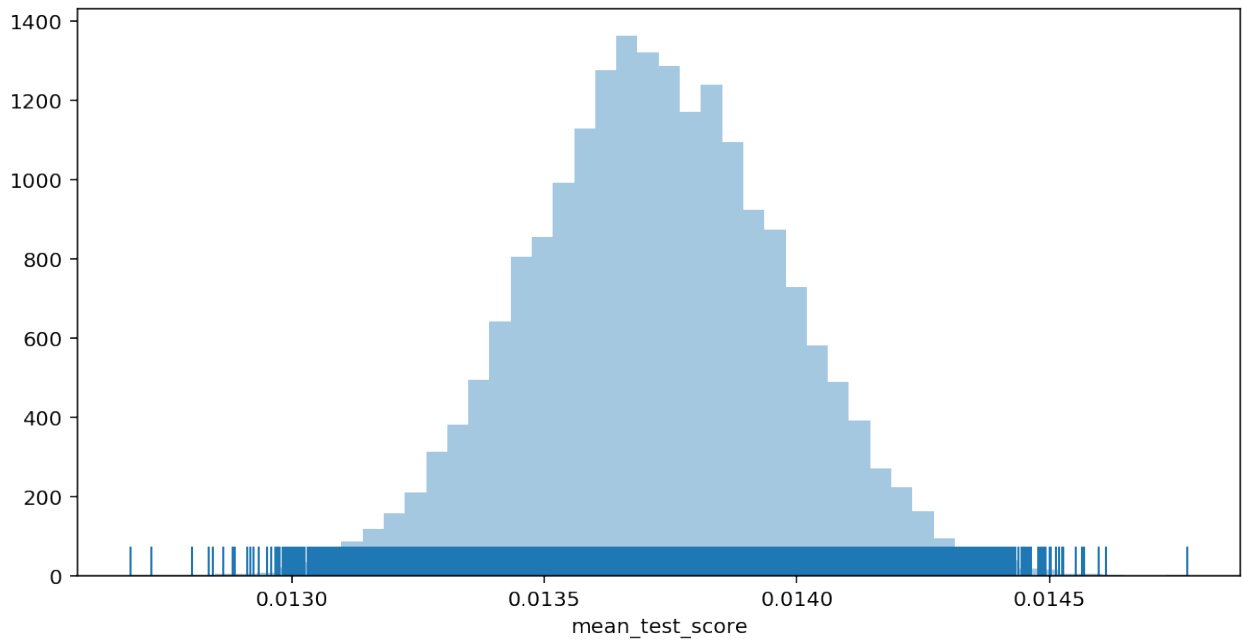
训练范围 [6426, 9427]，采用三折交叉验证，训练集大小 2000，验证集大小 1000

测试范围 [14389, 15390]，测试集大小 1000

结果如下：



所有的测试 mse 分布如下:

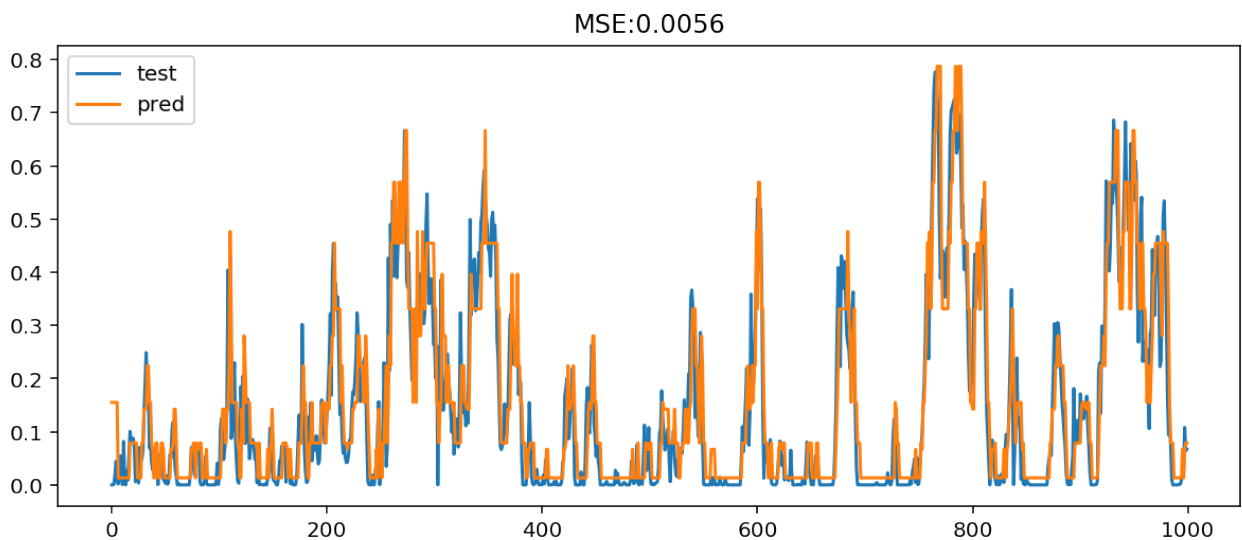


其中 esn 最优参数为：

```
1 {'damping': 0.3,  
2  'n_components': 1801,  
3  'n_readout': 4001,  
4  'weight_scaling': 0.4}
```

最优 mse 为：0.0126793948

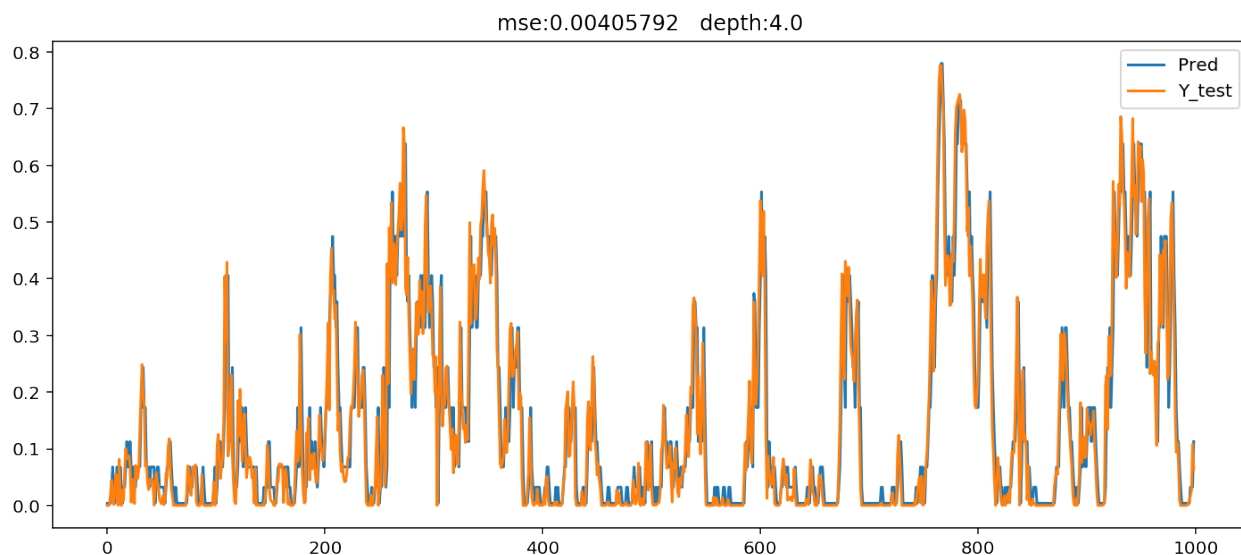
采用该参数后十次测试的平均 mse 为 0.006066137，其中一个预测图如下所示：



可以看到：

1. 对比优化前的 mse 0.0076，经过网络搜索优化 esn 参数下的测试 mse 为 0.0056，有了一定提升。
2. 与单纯决策树模型（mse 0.0041）对比，加入 esn 储蓄计算后的决策树模型并没有得到提升。

单纯决策树模型预测图如下：



1.2 美国数据集（时间间隔五分钟）

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

1.2.1 对输入的形式以及决策树最大深度进行搜索

训练集：[3001,7001] len: 4000（时间间隔5分钟）

测试集：[2000,3000] len: 1000（时间间隔5分钟）

输入说明：（所有数据已归一化，时间间隔5分钟）

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']

网络搜索范围：

```
1 param_grid2 = {'transform': [None, 'sin', 'cos', 'sin+cos', 'ws*sin(wd)',  
2                               'ws*cos(wd)', 'ws*sin(wd)+ws*cos(wd)],  
3                 'hour_num': np.arange(0,12),  
4                 'drop_time': [True, False],  
5                 'max_depth': np.arange(1,20)}
```

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：



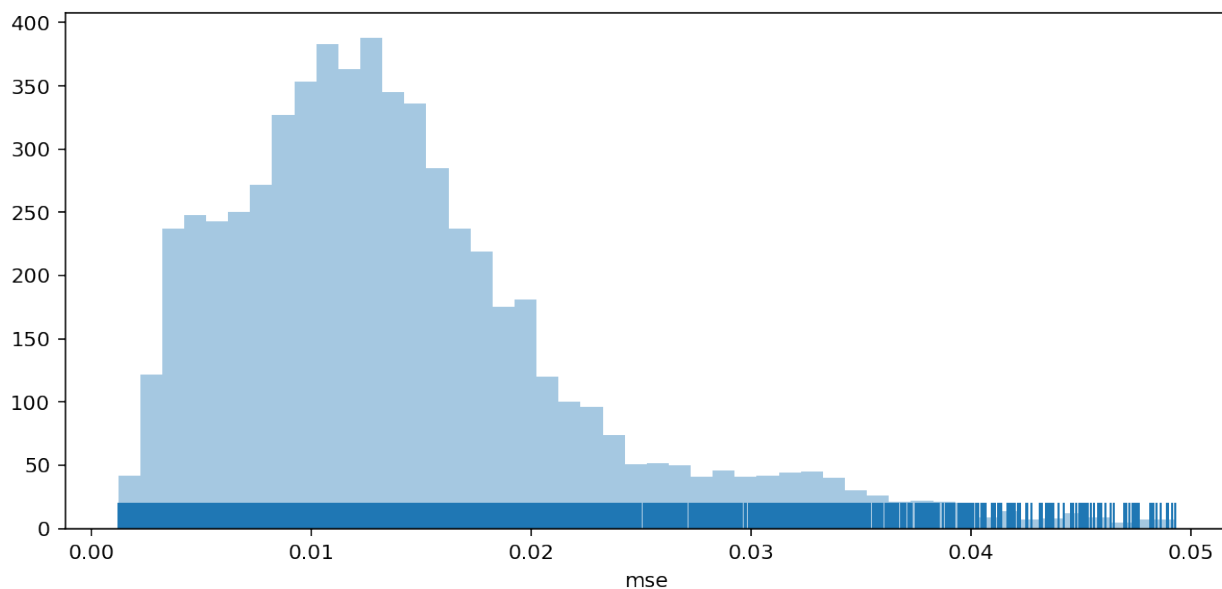
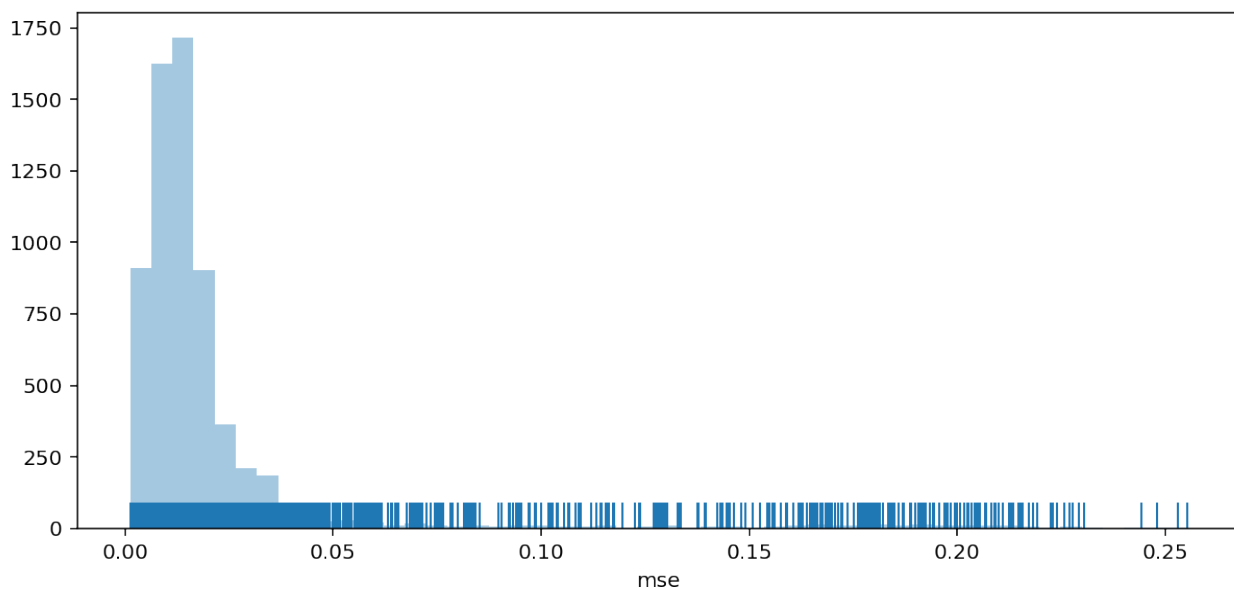
可以看到:

1. L 为 3 时效果最好
2. 数据转换在

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

上的表现较佳，其中 None 的表现优于 'sin' 和 'cos'，而在其他五种组合方式上表现较差。

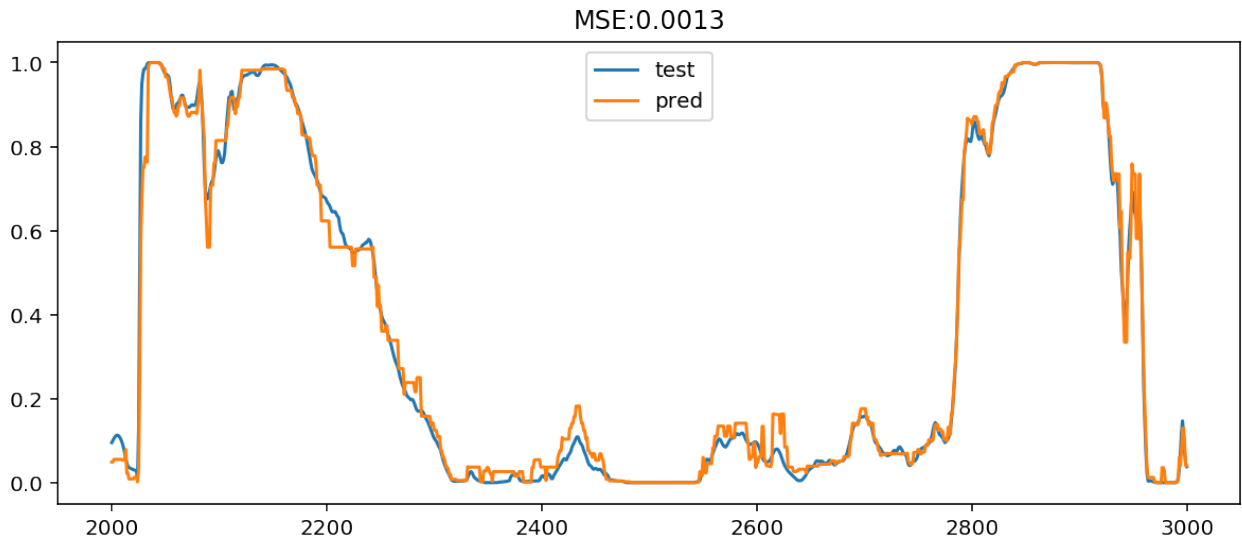
3. 不添加时间属性 (drop_time=True) 时效果较佳
4. 不添加其他属性 (drop_else=True) 时效果较佳
5. 所有测试 mse 分布如下:



可以得到最优的参数组合为:

```
1  {'drop_else': 1.0,  
2   'drop_time': 1.0,  
3   'hour_num': 0.0,  
4   'max_depth': 17.0,  
5   'transform': 'None'}
```

对应最优 mse 为 0.00120134, 采用该最优参数的预测图如下所示:



1.2.2 对 esn 参数进行搜索

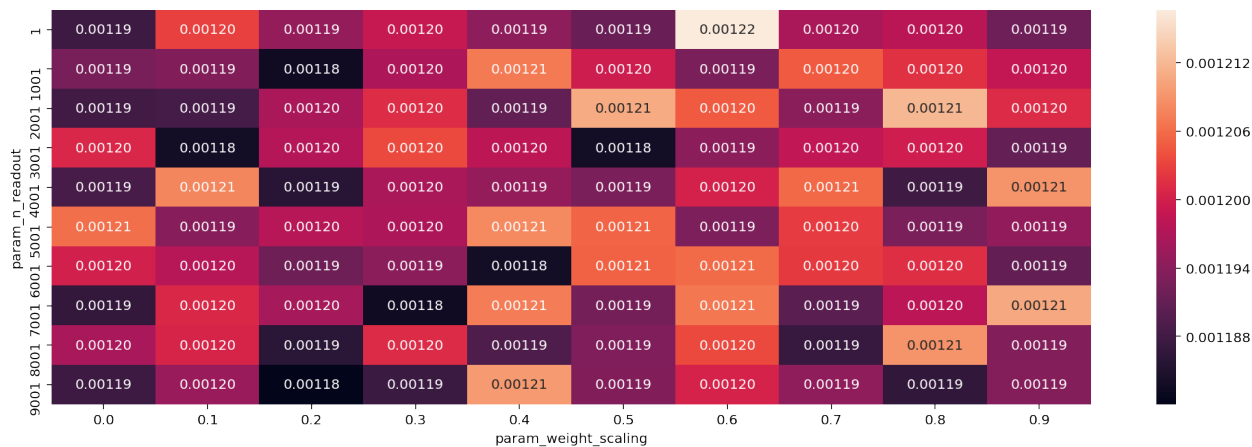
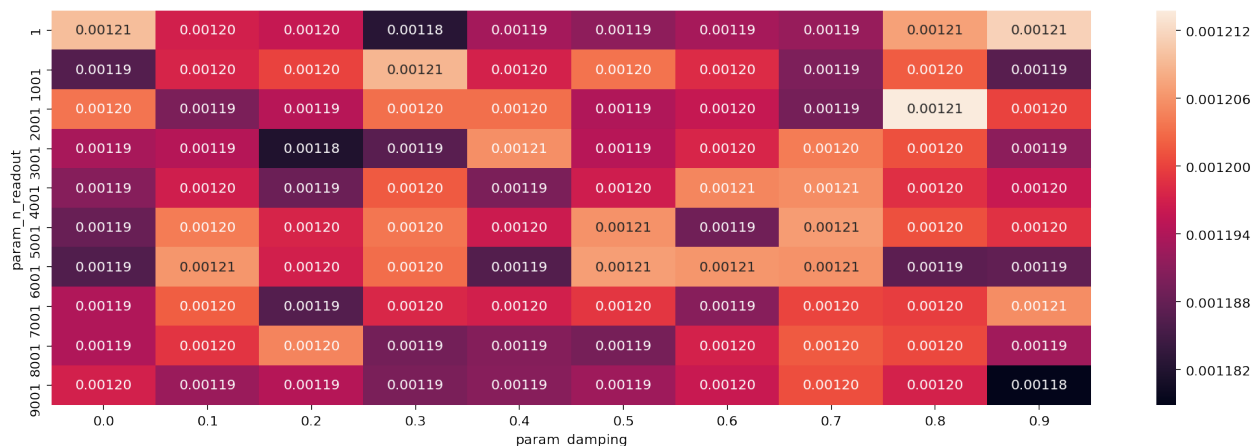
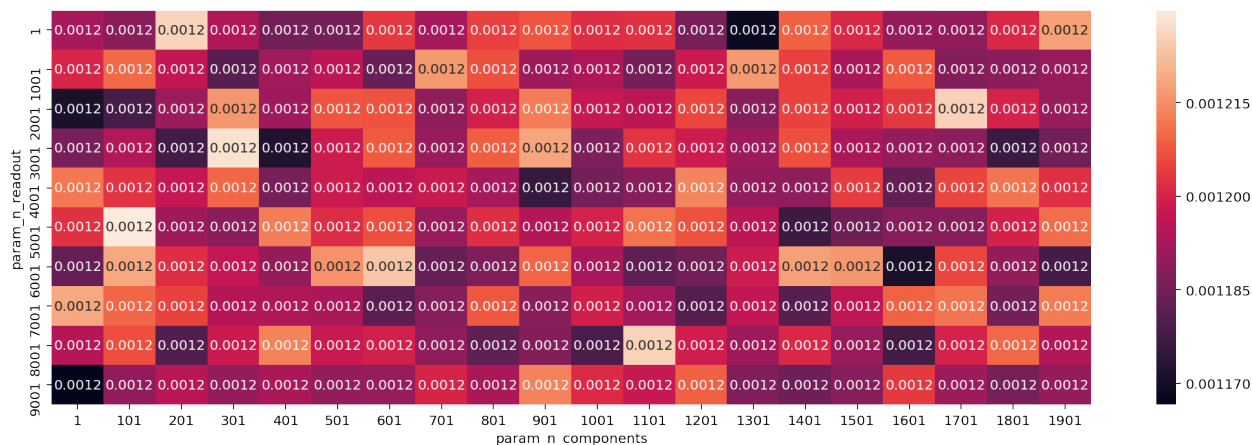
接下来在 1.2.1.1 的最优参数基础上，对 esn 参数进行网络搜索，搜索范围如下所示：

```
1 esn_param = {  
2     'n_readout': np.arange(1,10000,1000),  
3     'n_components': np.arange(1,2000,100),  
4     'damping': np.arange(0,1,0.1),  
5     'weight_scaling': np.arange(0,1,0.1)}
```

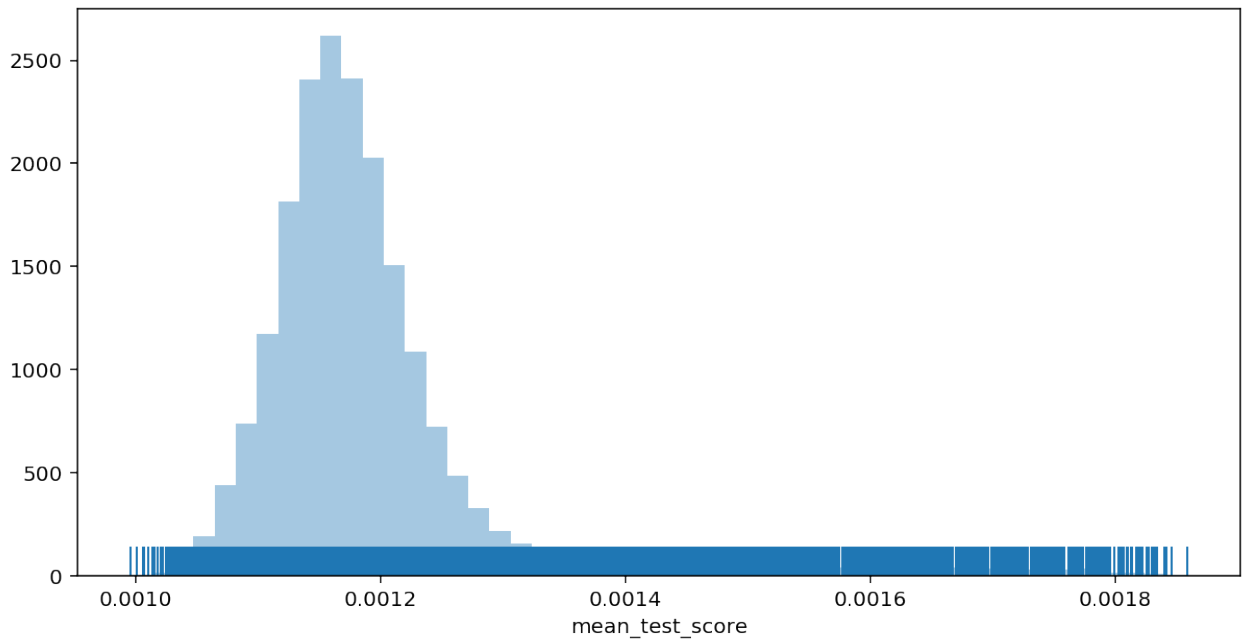
训练范围 [3001, 7002]，采用三折交叉验证，训练集大小 2000，验证集大小 1000

测试范围 [2000, 3001]，测试集大小 1000

结果如下：



所有的测试 mse 分布如下:

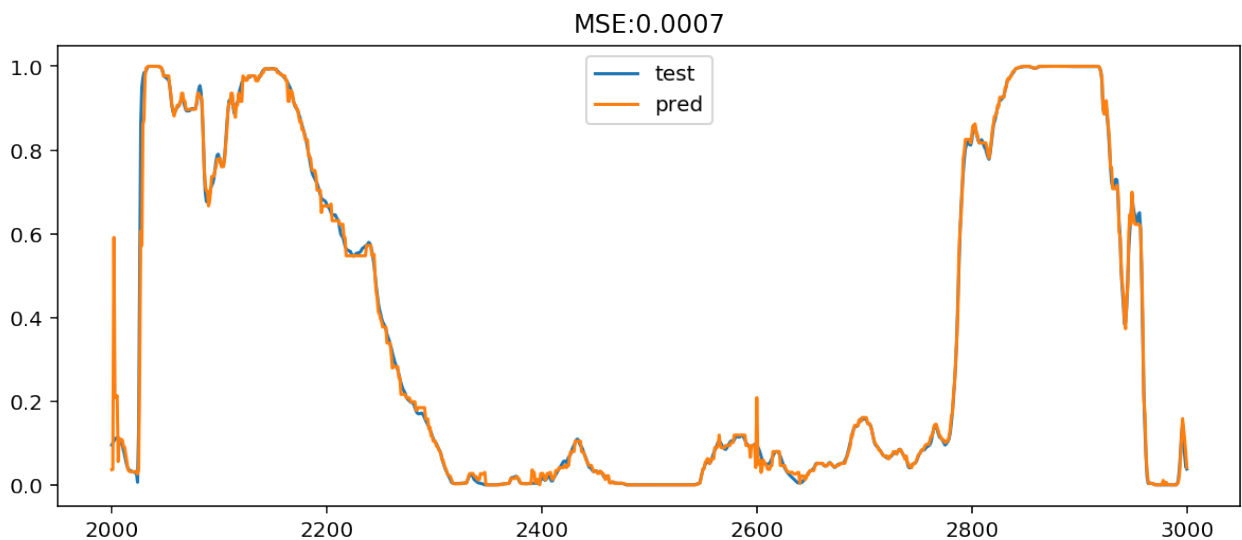


其中 esn 最优参数为：

```
1 {'damping': 0.9,  
2  'n_components': 1401,  
3  'n_readout': 3001,  
4  'weight_scaling': 0.8}
```

最优 mse 为：0.000995083

采用该参数后十次测试的平均 mse 为 0.0007898476，其中一个预测图如下所示：



可以看到：

1. 对比优化前的 mse 0.0013，经过网络搜索优化 esn 参数下的测试 mse 为 0.0007，有了一定提升。
2. 与单纯决策树模型 (mse 2.5709867e-05) 对比，加入 esn 储蓄计算后的决策树模型并没有得到提升。

单纯决策树模型预测图如下：

mse:1.346e-05 depth:10.0

