



北京大学

# 数据科学导引

## 期末大作业实验报告

题目： 风速预测实验

姓 名：	李子尧	严祚宇
学 号：	1500017776	1500017844
院 系：	元培学院	元培学院
专 业：	数据科学	数据科学

二〇一八 年 一 月

## 摘 要

风速预测是风力发电中的关键技术。现有的风速预测模型分为物理和统计及机器学习模型两种，后者中主要采用神经网络、时间序列分析、信号分析（包括小波变换、卡尔曼滤波等手段）等方式进行建模。统计模型相比物理模型更为简单，但是长期预测误差往往过大。

本次实验数据来自江苏南通某风场 2016 年 4 至 7 月风速实测与预测数据，数据量大，时空跨度不够广，质量一般。实验分为长期预测和短期预测两部分，长期预测采用 CNN+LSTM 与 CNN+MLP 两种神经网络框架，短期预测采用 ARIMA 与 ARWT 两种基于时间序列分析和信号分析的模型。

实验证实，LSTM 网络设计表现不如简单的 MLP 框架；短期预测上，ARIMA 模型更适合超短期（一分钟内）预测，ARWT 模型更适合中短期（半小时至一小时）预测；通过模拟真实情况的滚动预测实验，模型在风速变化较为稳定的情况下效果极佳。

**关键词：**风速预测 / 神经网络 / 时间序列 / 小波变换

## 1. 实验概览

随着化石能源的日益枯竭，环境问题例如全球变暖等的日益严重，发展可再生能源因此成为了一种必然。而风能作为一种污染、可再生能源，得到世界各国的高度重视，风力是目前最成熟的、最具有商业开发前景的可再生能源之一。但风力发电同时也存在着很多技术难题，包括风速的弱规律性、波动性、间歇性，以及风电能量密度低等特点，这些都对风力发电带来的极大的困难与挑战。其中，对于风速的准确预测，对于风电的开发是至关重要的。如果风电功率极不稳定，将会对风机、电机甚至整个电网产生致命威胁。

本次实验目标为风速预测。根据学术惯例，风速问题往往划分为短期预测和长期预测两类。短期预测往往是通过预测点前一段时间内的风速测量数据对于风速进行超前一步预测，进而更好的调节风机负载和方向。因此，短期预测往往需要预测风速变化的短时波动性，进而更好的预测精确时间点上的风速；长期预测则与天气预报相似，往往是通过大气、物理的模型对于收集到的气象信息加以分析和预报。与短期预报相比，长期风速预报不需要精确刻画风速变化的短时效应，而是预测长期风速变化趋势。

## 2. 相关研究

尽管预测难度大，风速变化依旧有迹可循。目前，该领域学术成果丰富，预测的主要模型可分为物理方法与统计及机器学习方法两类。物理方法主要考虑风速产生的气象背景，运用气象领域知识对风速变化模拟计算，这里略去；统计方法比之物理方法需要的资源更少，往往更加简单。2004 年，[1]利用时间序列的 ARMA 模型对美国风电功率进行预测；2010 年，[2]提出 ITSM，一种基于 ARIMA 的改进时间序列分析进行预测，其核心是利用小波变换分解风速后，使用时间序列模型与神经网络分别对子序列进行预测；2011 年，[3]组合 ARMA、VAR 等模型进行预测；同年[4]应用粒子群优化算法于神经网络进行预测；[5]比较了三种经典神经网络(FFBP、RBF、ADALINE)，认为风速预测精度受输入、学习策略和模型结构影响极大，没有最优方法；2012 年[6]结合小波变换与神经网络、2014 年[7]结合贝叶斯学习（与高斯过程结合）与神经网络进行预测；[8]利用神经网络和卡尔曼滤波混合建模进行风速短时间高精度预测，[9]基于卡尔曼滤波理论实现风速预测，通过求解三分钟内极大风速样本建立卡尔曼滤波测量方程以及状态方程，依靠卡尔曼滤波递推方程组来实现风速高精度向前多步预测等。

## 3. 数据预处理

本次大作业风速数据来自江苏南通某风场，使用的数据主要由两个部分组成，包括从 2016 年 4 月 1 日至 7 月 31 日，每隔不定时间，在六台风机的位置分别测得的风速历史数据，以及相应日期内每隔 15 分钟在六台风机的预测风速、气压与温度，这些预测的时间相距预测时间至少 24 小时。此外，历史数据

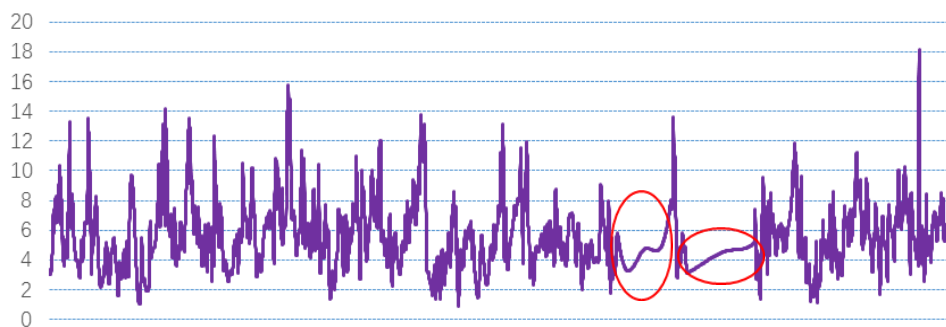


图 3.1 2016 年 4 月至 7 月 六号机风速情况

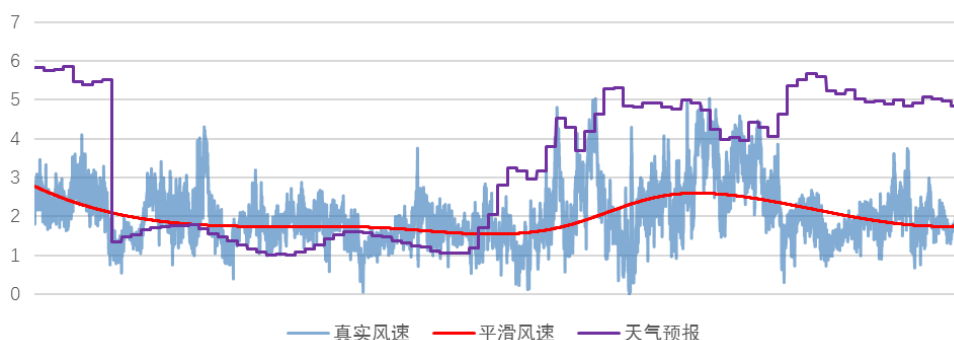


图 3.2 2016 年 7 月 12 日 一号机风速情况

与天气预报的数据均有不同程度的缺失。缺失情况上看，一至四号机较为完好，五、六号机缺失更多。整体而言，数据量大但时间跨度不够广，难拟合季节性模型；6 台风机基本无空间跨度，无法拟合空间模型。数据质量一般，6 月一段时间出现长时间缺失，数据完整性遭到破坏。

如前所述，本实验依照学术惯例，将问题分为长期风速预测与短期风速预测两个问题分别进行建模。在长期和短期这两个不同问题的框架下，考虑到二者之间的差异性，我们对于数据的预处理差异很大。

在长期预测下，风速的短期波动往往不予考虑。因此，我们首先拟合了真实数据的三次自然样条函数，并以  $15\text{min}^{-1}$  的频率进行采样，作为训练和评价的真实值；此外，由于数据中存在或长期或短期的缺失情况，考虑到缺失比例不大 ( $< 5\%$ )，因此我们对缺失值采用填充处理，填充的内容为当天该时刻的天气预报。以缺失比例最高的六号机为例，图 3.1 是采用上述缺失值填充策略得到的平滑结果。

可以看出，尽管利用风速预报数据进行缺失值填补的地方（图 3.1 中红色圈内的部分）由于采样点相对稀疏，相较其他地方更加平滑，但一方面由于缺失数量少，一方面天气预报与真实值大致水平相当、趋势一致，用其填补是一种很好的估计。

在短期预测下，这样的设计便不可行。短期、超短期预报要求对风速数据的短时波动也进行精确预报，因此在处理时必须考虑到风速的短时变化，平滑化的手段会丢失短期内的绝大部分信息，使任何的短期、超短期模型退化为均值估计。

然而，由于获得的数据并非均匀采样的，无法利用时间序列或信号分析手段，我们按照  $15\text{s}^{-1}$  的频率，对于数据进行 NN 线性插值采样，即用最靠近每个采样点的两个数据点的加权平均作为采样点风速的最精确估计，作为训练和评

价的标准。以 7 月 12 日一号机的风速情况为例，数据的处理结果如图 3.2。

可以看到，一方面三次样条插值的结果可以很好的反映风速的变化趋势，另一方面，已有的基于物理的风速预报虽然可以很好的近似风速的大致大小，其与真实值之间仍有一定差距，需要建立一个更好的模型进行预测。

## 4. 模型

在长期预测下，我们提出了基于 CNN-LSTM 和基于 CNN-MLP 的两种框架；在短期方面，我们利用时间序列分析和小波变换的方式进行预测。

### 4.1. 长期模型

由于天气预报对于风速真值的预测还有一定差距，我们自然想到，如果可以从历史数据中抽取出一一定的模式进行预报，那么对于风速的预测可以得到一定程度的修正。基于此，我们提出了输入为历史数据和天气预报的神经网络模型。对于历史数据，我们提取每个预测点前七天，每天该时刻以及该时刻前后一小时内的风速数据作为特征输入；对于天气预报数据，我们提取该时刻预报数据的风速、温度和压强作为第二组输入，采用图中模型进行训练。

考虑到相邻的日期、相邻的时间的风速数据更有融合的必要性的，我们对于矩阵形状的历史数据输入采用了单通道 CNN 模型提取特征，与用多层感知机提取的天气预报特征合并，作为大的 LSTM 的输入，得到了如图 4.1 的 CNN+LSTM 模型。然而，出乎意料的是，这个模型的效果欠佳。

通过残差分析我们认为，LSTM 的模型过于复杂，采用了简单 RNN 的框架，结果仍不如人意。通过进一步分析，我们认为，模型对于 RNN 类型的框架中的隐层信息并不能充分利用；再由于跨度长、没有真实数据进行进一步校正，长期预测的过程中隐层信息往往不能很好

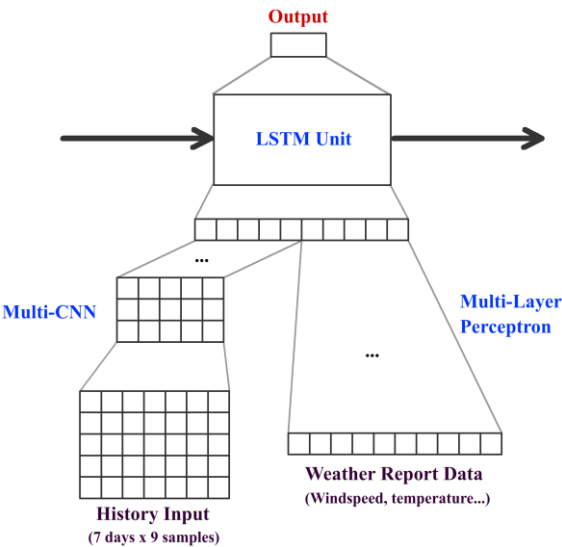


图 4.1 CNN+LSTM 框架

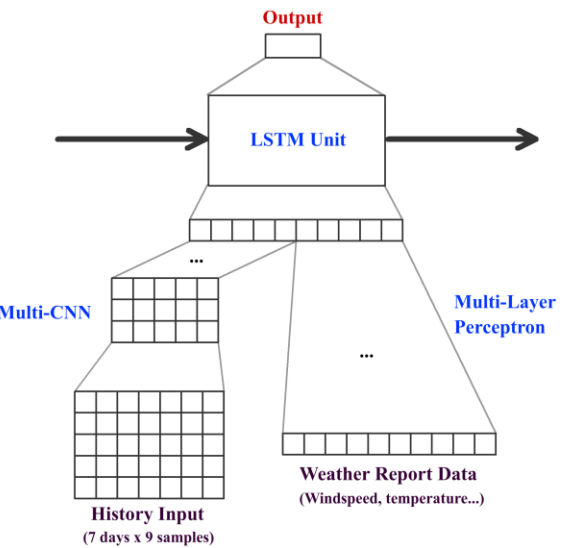


图 4.2 CNN+MLP 框架

反应真实值，于是我们抛弃了 RNN 框架，采用了直接利用 CNN 和多层感知机进行任意时刻风速预测的框架，如图 4.2。

### 4. 2. 短期模型

短期预测方面，我们主要采用了基于单纯时间序列分析和基于小波变换及时间序列（ARWT, AR(MA) and Wavelets Transformation）的两种方法。时间序列上，我们采用了经典的 ARIMA 模型。据检验，风速序列并不具有明显的日周期性；进一步，通过 Dickey-Fuller 检验以及 EACF 和 AIC 定阶[10]，我们确定 ARIMA(1,1,1)模型为最佳选择，用基于卡尔曼滤波的极大似然估计拟合模型；小波变换方面，我们采用了 Daubechies-4 基底进行二阶变换，并对于得到的小波分解后的三个系数分别进行 ARMA 模型，预测出一段时间后的小波系数并进行序列重构，得到预测的风速序列。

通过时间序列或小波变换的方式进行短期预测，我们采用了滚动预测的框架：在每个预测节点上，采集前若干时刻的风速数据作为训练样本，拟合出这一时刻的时间序列模型，并用这一时间序列模型进行向前一步或多步预测。

### 5. 实验结果

长期模型中，我们将 2016 年 4 月 1 日至 7 月 10 日的数据作为训练集，7 月 11 日至 31 日的数据作为测试集，并利用天气预报对于真实值估计的 MSE 作为评价的 Baseline。Baseline 结果的 MSE 为 3.50，MAE 结果为 1.49。

如前所述，我们反复调节 RNN 框架的参数，但是实验结果都不理想；将 RNN 框架简化为 CNN+MLP 的框架后，预测的 MSE 降至了 2.99。考虑到预测只用到了天气预报的数据和一部分历史信息，结果达到这一水平，可以论证我们的框架非常好的提取了特征。此外，我们注意到，这一模型对于六台机器预测的 MSE 分别为：

表 5.1 不同机器预测 MSE

机器	一	二	三	四	五	六
MSE	2.54	2.18	3.32	4.12	2.55	4.55

从上表可以看出，四、六号机数据中的离群点过多；而无论天气预报还是历史数据都无法预测出某个风速脉冲（这在四、六号机的数据中很常见，如图），再加之 MSE 受离群点影响极大，导致这两台机器的预测 MSE 很高。表中数据证实，我们的模型在正常的数据上，效果发挥更好。

短期模型中，由于数据过于庞大，我们采用了抽样出的 6\*100 个时间点作为预测节点，采集这些节点前 300 个风速数据作为训练样本，并进行 300 步向前预测（即预测这一时间节点后 15s~75min 的风速，每 15s 做一次预测），并分别依 MSE 和 MAE 进行模型评价。我们得到了以下有趣的结果：

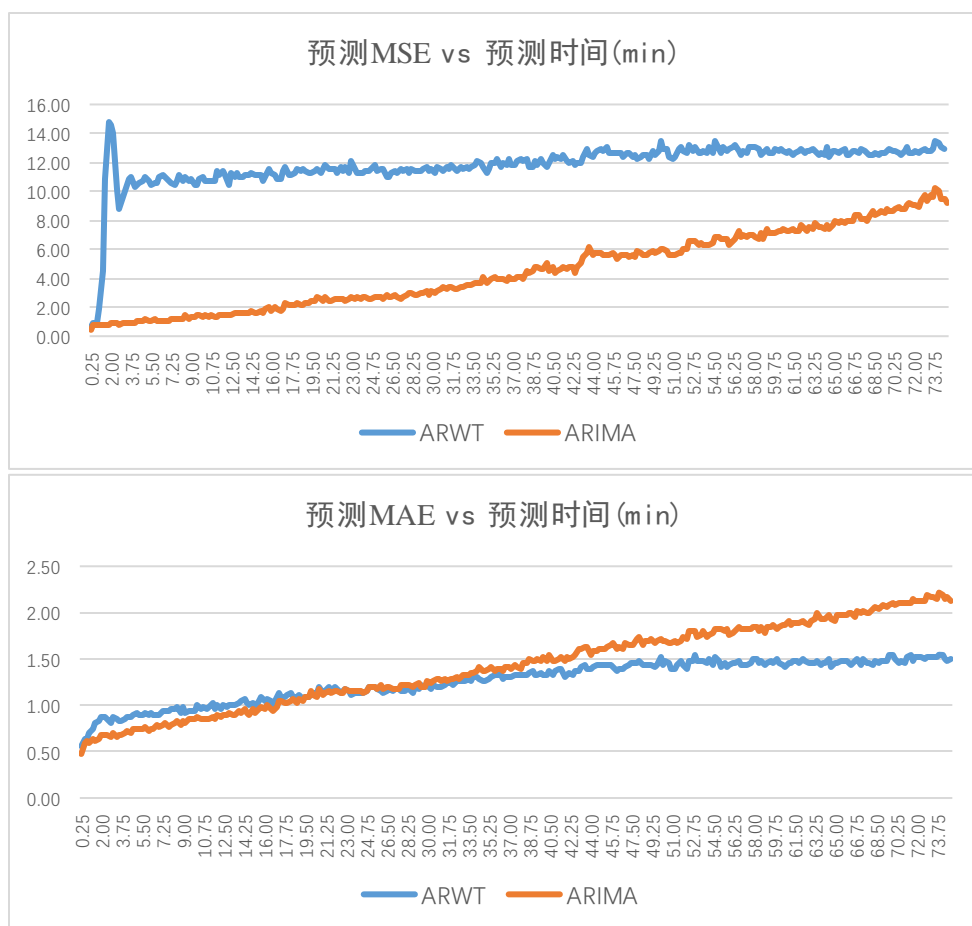


图 5.1 ARWT 与 ARIMA 预测误差

单看 MSE，ARWT 模型似乎很糟糕，在预测阶数很小的时候就有了很高的 MSE；但是通过 MAE，我们看到，当需要一个中短期的风速预测时，ARWT 模型得到的预测在更稳健的评测标准下更接近真值。我们进一步分析了两个模型预测结果的特点，见图 5.2。

通过图 5.2 我们可以看出，ARIMA 模型和基于 ARMA 模型的小波变换都存在“均值回归”的特点，对于序列变化，往往在短期内以一个线性趋势作为预测，并且之后以这个线性趋势为保持。因此，基于时间序列的模型不适合长期预测，但是在短期内可以往往把握时间序列变化的趋势性。

考虑到数据中噪声和离群点对于模型误差的严重影响，尤其是时间序列回归的保守性，我们又进行了一次滚动实验，通过没有明显离群点的 2016 年 4 月 1 日的 6 台机器的测量风速进行了  $6 \times 100$  次滚动预测实验，得到了良好的结果：一步向前预测数据误差 MSE 低于 0.1；8 步向前预测（2 分钟）误差 MSE 小于 0.2；60 步向前预测（15 分钟）误差 MSE 小于 1.0。

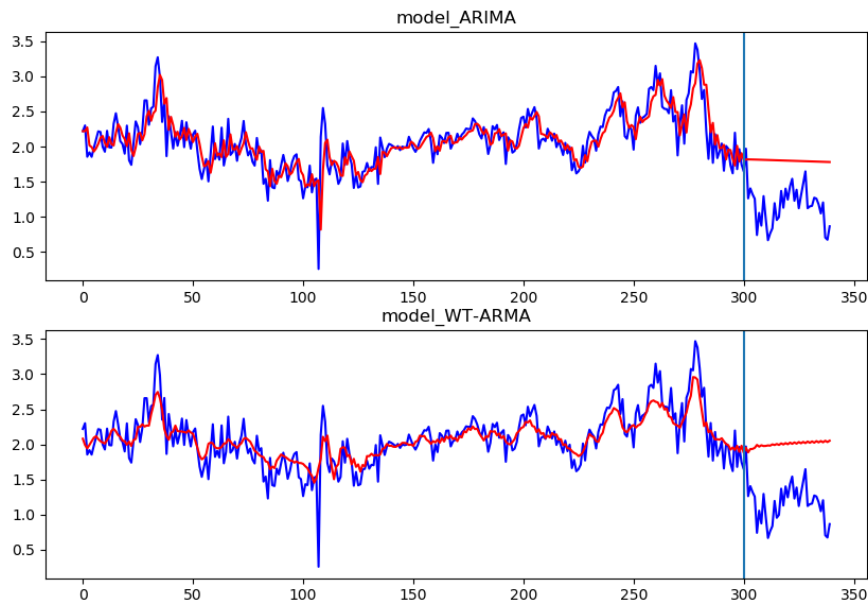


图 5.2 ARIMA 与 ARWT 预测结果

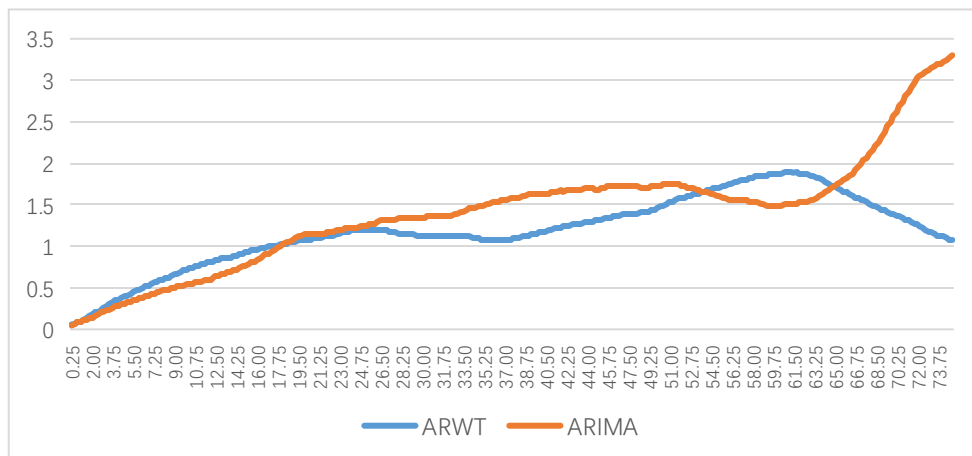


图 5.3 ARIMA 与 ARWT 滚动预测 MSE

## 6. 总结

在本次风速预测实验中，我们将风速预测问题按照学术惯例划分为短期和长期两类。短期模型往往需要捕捉风速的波动性；长期模型需要把握风速变化的趋势。我们基于数据限制，提出了长期预测的 CNN+MLP 和 CNN+RNN 两种框架，前者的表现优异；在短期模型中，考虑到天气预报的时效性，我们并未使用天气预报，采用了 ARIMA 和 ARWT 两种模型。实验证明前者更适合短期预测，后者更适合中短期预测。在滚动预测的实验中，二者的表现都良好；在抽样定点的实验中，受到较多的离群值影响，MSE 上二者表现稍逊，ARWT 尤其受影响，但 MAE 表现仍然稳健。在今后的探索中，我们希望可以再收集一些新的解释变量，从不同的侧面对风速进行预测。



## 参考文献

- [1] Milligan M, Schwartz M, Wan YH. Statistical wind power forecasting for U.S. Wind farms. 2004
- [2] Liu H, Tian H-Q, Chen C, Li Y-f. A hybrid statistical method to predict wind speed and wind power. *Renew Energy*, 2010, 35(8):1857-1861
- [3] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy*, 2011, 88(4):1405-1414
- [4] Bouzgou H, Benoudjit N. Multiple architecture system for wind speed prediction. *Applied Energy*, 2011, 88(7):2463-2471
- [5] Barbounis TG, Theocharis JB, Alexiadis MC, etc. Long-term wind speed and power forecasting using local recurrent neural network models. *IEEE Transactions on Energy Conversion*, 2006, 21(1):273-284
- [6] Tong JL, Zhao ZB, Zhang WY. A new strategy for wind speed forecasting based on autoregression and wavelet transform. *Proceedings of the 2nd international conference on remote sensing, environment and transportation engineering (RSETE)*, 2012.
- [7] A Tascikaraoglu, O Erdinc, M Uzunoglu, A Karakas, An adaptive load dispatching and forecasting strategy for a virtual power plant including renewable energy conversion units. *Applied Energy*, 2014, 119(15):445
- [8] Li SH. Wind Power Prediction Using Recurrent Multilayer Perceptron Neural Networks. *IEEE Conference Preceedings of Power Engineering Society General Meeting*, Toronto; Institute of Electrical and Electronics Engineers Inc., 2003:2325-2330
- [9] Kobayashi N, Shinanura M. Study of a Strong Wind Warning System. *Jr East Technical Review*, 2002:61-65
- [10] R. Tsay, *Analysis of Financial Time Series*, 2012