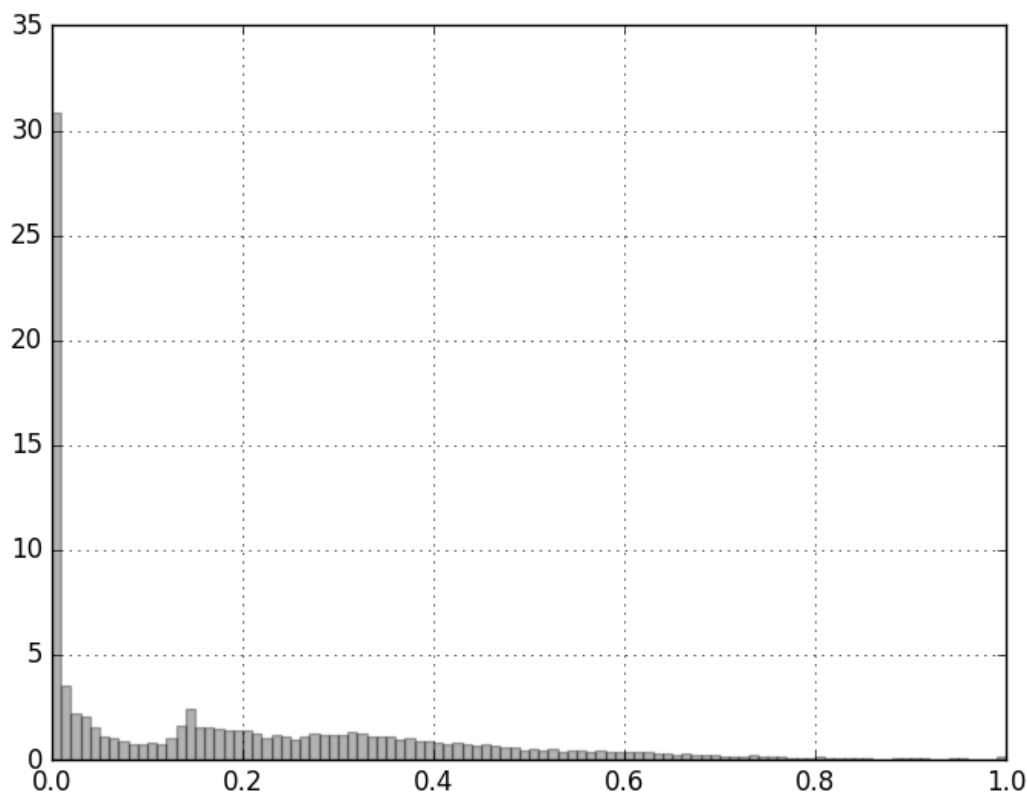


VAR和AR模型及ARMA模型对比分析结果

1.AR模型与VAR模型

对超短期功率进行直方图分析，画出概率密度曲线；



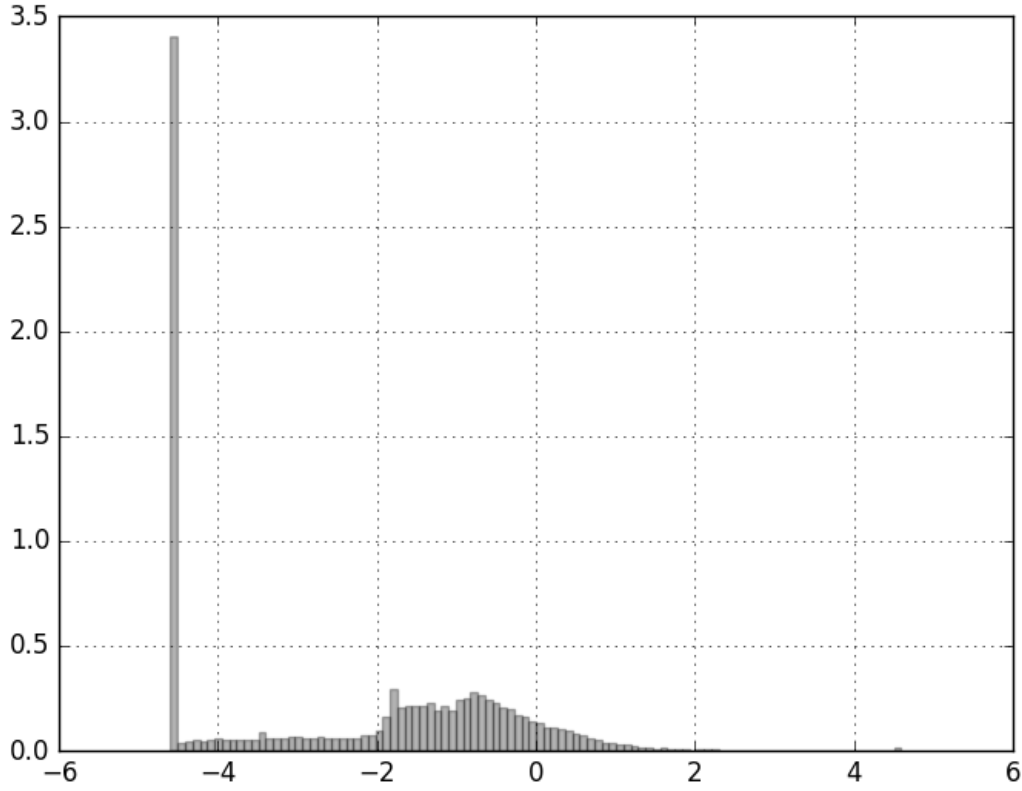
《Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions》指出，风机功率是对数正态分布，为了适应线性条件下进行预测，需要在高斯分布的情况下建模，因此提出了广义对数变换，将功率变换为正态分布，变换公式为：

$$y_t = \gamma(x_t; \nu) = \ln \left(\frac{x_t^\nu}{1 - x_t^\nu} \right), \quad \nu > 0, \quad x_t \in (0, 1) \quad (1)$$

while the inverse transformation, referred to as inverse generalised logit (IGL), is defined as

$$x_t = \gamma^{-1}(y_t; \nu) = \left\{ 1 + \frac{1}{\exp(y_t)} \right\}^{-1/\nu}, \quad \nu > 0, \quad y_t \in \mathbb{R} \quad (2)$$

变换后y的概率密度图；



考虑到随机过程的非线性和功率的双边界性质。（0分布过于集中，需要仔细分析原因）在这里采用广义对数 - 正态分布和边界处概率密度的离散连续混合形式。积累分布函数可以写为：

$$Y_{t+k} \sim \omega_{t+k}^0 \delta_{\gamma(\epsilon; \nu)} + \mathcal{N}(\mu_{t+k}, \sigma_{t+k}^2) \mathbf{1}_{\bar{\mathcal{D}}_y} + \omega_{t+k}^1 \delta_{\gamma(1-\epsilon; \nu)} \quad (6)$$

where $\bar{\mathcal{D}}_y$ is the complement of \mathcal{D}_y and $\mathbf{1}_{\bar{\mathcal{D}}_y}$ the indicator function for this open interval, i.e.

$$\mathbf{1}_{\bar{\mathcal{D}}_y} = \begin{cases} 1, & y \in \bar{\mathcal{D}}_y \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The weight $(1 - \omega_{t+k}^0 - \omega_{t+k}^1)$ does not appear anymore in (6). This is because the predictive density for Y_{t+k} takes the form of a censored Normal instead. This permits to straightforwardly derive the expressions for the weights ω_{t+k}^0 and ω_{t+k}^1 ,

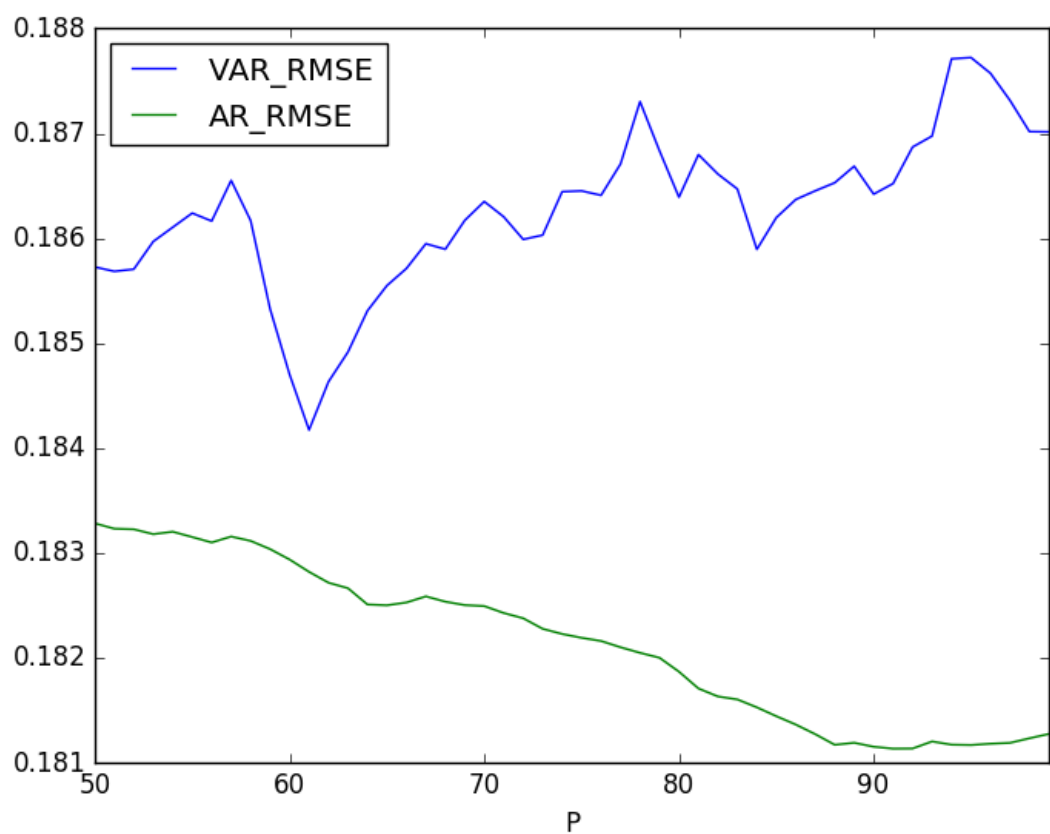
$$\omega_{t+k}^0 = \Phi\left(\frac{\gamma(\epsilon; \nu) - \mu_{t+k}}{\sigma_{t+k}}\right), \quad \omega_{t+k}^1 = 1 - \Phi\left(\frac{\gamma(1-\epsilon; \nu) - \mu_{t+k}}{\sigma_{t+k}}\right) \quad (8)$$

因此可以用AutoRegression模型进行预测。

实验中采用某电厂2015/08/23-2016/08/01日的15分钟数据作为训练数据。

采用2016/08/01-2016/11/01日的15分钟数据作为实验数据。

对比AR和VAR模型的均方根误差

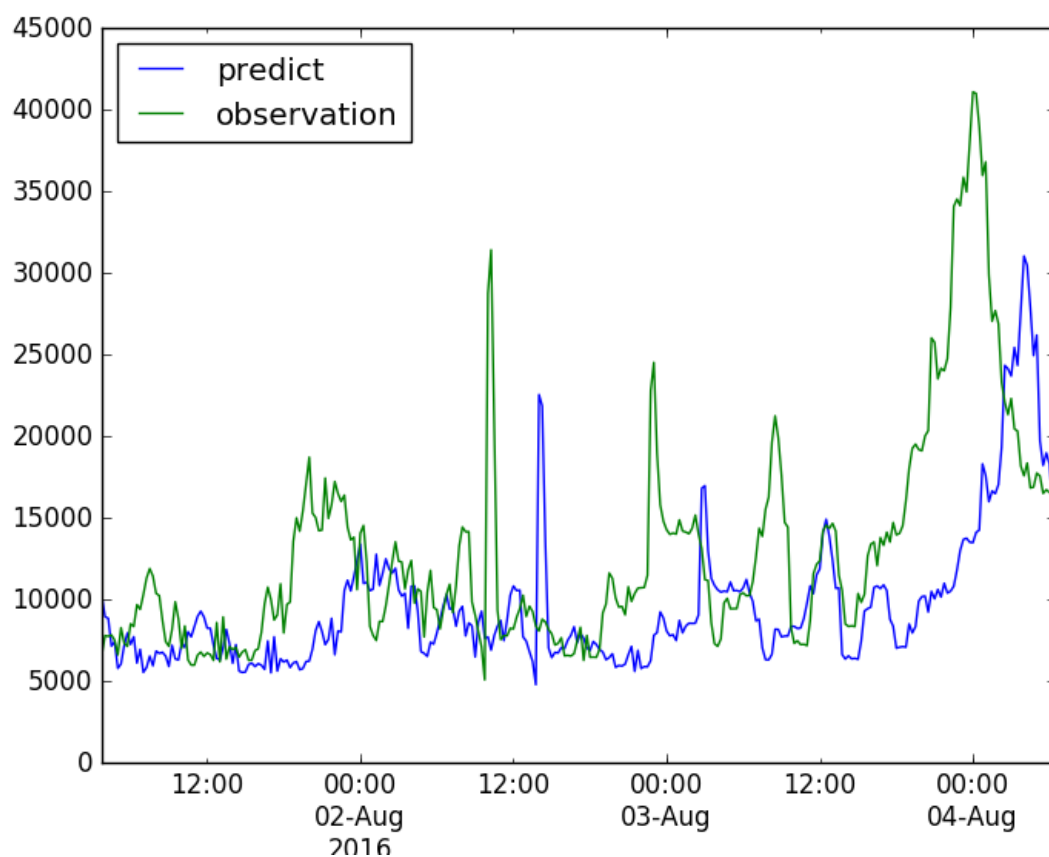


VAR 随着阶数P的增加超过一个值之后均方根误差并没有显著下降，并且没有超过AR模型。

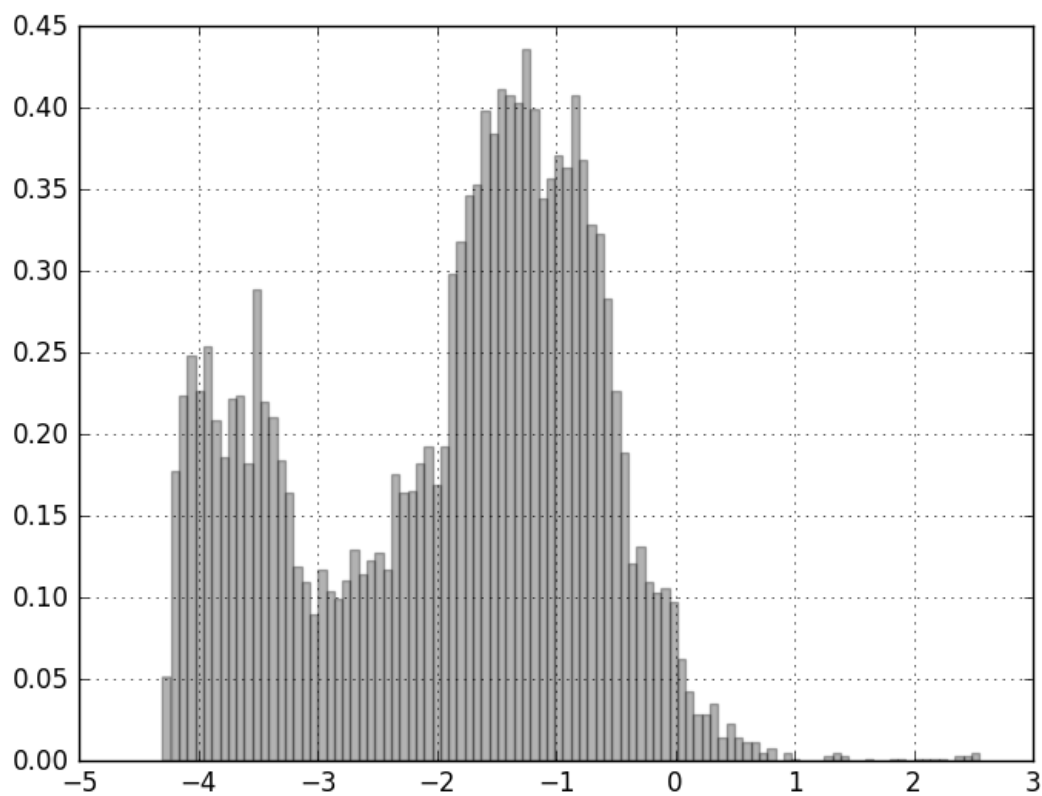
AR模型随着阶数P的均方根误差增长有明显的下降，考虑到模型算法复杂度和效率的结合选取参数p=90，分析其各个月份的均方根误差。

month	RMSE
8	0.174618598
9	0.141792849
10	0.179298075
point_mean	0.181156346435

预测值与观测值对比图：



预测值的分布：



2.AR模型与ARMA

AR模型选取参数 $p=2$, ARMA模型选取参数 $p=2, q=2$

arma RMSE

month	RMSE
8	0.183920
9	0.141944
10	0.182765
point_mean	0.185415109176

ar RMSE

month	RMSE
8	0.184014
9	0.141972
10	0.182742
point_mean	0.185444757458

实验中ARMA和AR模型效果相差不大。

3.后续计划，对于实验出现的一些结果，需要更深入分析其原因，比如在边界值过于集中，预测结果中差异较大的点产生的原因。

未来考虑的优化方向，后续加入动态遗忘因子，利用多模型复合预测，提高模型对于突变天气的适应性，增加模型算法复杂度，考虑非线性模型。