Evaluation Model of Agricultural Production Risk and the Application

农业生产风险评估模型及应用

Prof. Qiao Zhang

Agricultural Information Institute (AII),

Chinese Academy of agricultural sciences (CAAS)

For DelPHE 4th Workshop

June 22,2009

Contents
目 录

1. Agricultural yield risk: concept, classification and feature

- Concept(概念)

Crop yield risk was defined as the degree that farmer’s actual yield differ from the expected yield for uncertainty and bounded rationality.

农业生产风险是指由于不确定性和有限理性，农业生产的实际产量偏离其预期产量的程度。

- Note (说明)

Most crop yield risk was the consequence of adverse Weather. However, crop yield risk is not the same as natural hazard risk, it also depend on other factors such as variety, investment and farm management.

农业生产风险主要是由不利的气候引起的，但是它并不等同于自然灾害风险，其风险水平还与作物品种、田间投资和管理等因素密切相关。

1. Agricultural yield risk: concept, classification and feature

- Classification(类别)

Agricultural yield risk was refer to “in-between” risk that is neither highly independent nor highly correlated (Skees and Barnett, 1999).

and it was also considered as the mixture of two kind of risk: non-catastrophic risk and catastrophic risk, also known as normal risk and extreme risk (Ker and Coble, 2003).

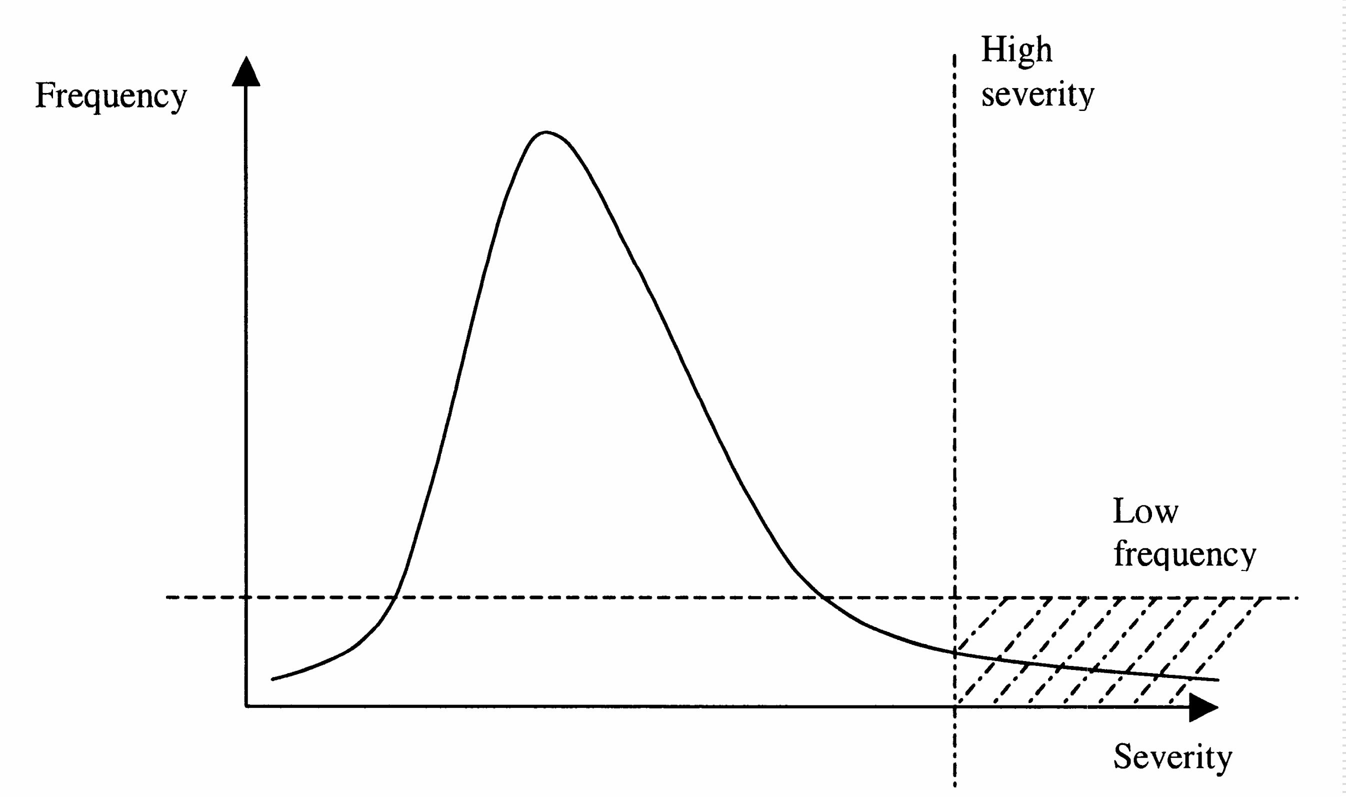
农业生产风险是一种介于完全独立和高度相关之间的一种“中间”风险，它可以看作是两种风险（常规风险和极端风险）的组合。即农业生产风险可以分为常规风险（非巨灾风险）和极端风险（巨灾风险）两类。

1. Agricultural yield risk: concept, classification and feature

Feature(特征)

Normal risk r
their expecta
Extreme risk

- Normal
  常规风险,
- Extrem
  极端风险



2. Agricultural yield risk assessment

- What is yield risk assessment?(风险评估的本质)

Assessing the agricultural yield risk involves determining the probability of loss and the expected level of loss when losses occur. More formally, one is generally interested in a measure of the PDF underlying the events that trigger loss. Thus, the concept of modeling yield risk for the purpose of assessing yield risk is fully analogous to modeling the probability distribution for the crop yield in question (Goodwin and Mahul, 2004).

农业生产风险评估包括估计单产损失的大小及损失发生的概率，一般来讲，人们更关心后者。在农业生产风险评估过程中，对生产风险的定量化估算就是对相应农作物单产概率分布进行拟和的过程**(Goodwin and Mahul, 2004)**。

2. Agricultural yield risk assessment

- For normal risk(常规风险评估)

● Although the normal risk occur frequently, the consequence result from it are modest, and most values of some indicators that can express the yield risk level, such as yield loss mostly are located at center region. thereby, traditional (classic) statistic methods were reliable to modeling the normal yield risk.

● The basis idea is to use historic yield data to derive the probability distribution of normal yield risk

虽然农业生产中的常规风险发生频率很高，但是每次造成的损失都不大，反映常规生产风险的单产波动概率分布集中于均值附近。因此，利用传统的或经典的概率统计方法就可以对常规生产风险进行有效评估。

对常规生产风险进行评估的基本思想是采用概率统计学的方法，利用单产历史数据来估测农作物常规生产风险的概率分布。

2. Agricultural yield risk assessment

□ For extreme risk(极端风险评估)

- Traditional statistics mostly focus on the laws governing average, while catastrophic events could lead to severe consequences and they fall in the tails of specific distributions. So, it will have misleading or biased if modeling extreme risk using traditional statistic approaches.
- Extreme value theory could provide a promising solution to modeling extreme risk since it is primarily concerned with the quantification of the stochastic behavior of a process at usually the largest, the smallest, or the events over a threshold.

传统的统计方法大都关注于均值，即基本统计都是建立在中间数据的基础之上，而极端风险（巨灾风险）产生的破坏后果特别严重，集中于分布的尾端，因此，如果用传统的统计方法来拟和和评估极端风险就会产生很大的偏差和误差。

由于极值理论专注于小概率尾端事件或高于某阀值之上的事件，它克服了用传统统计方法拟合极端风险存在的各种不足，提供了对极端风险进行拟合、评估的稳健方法。

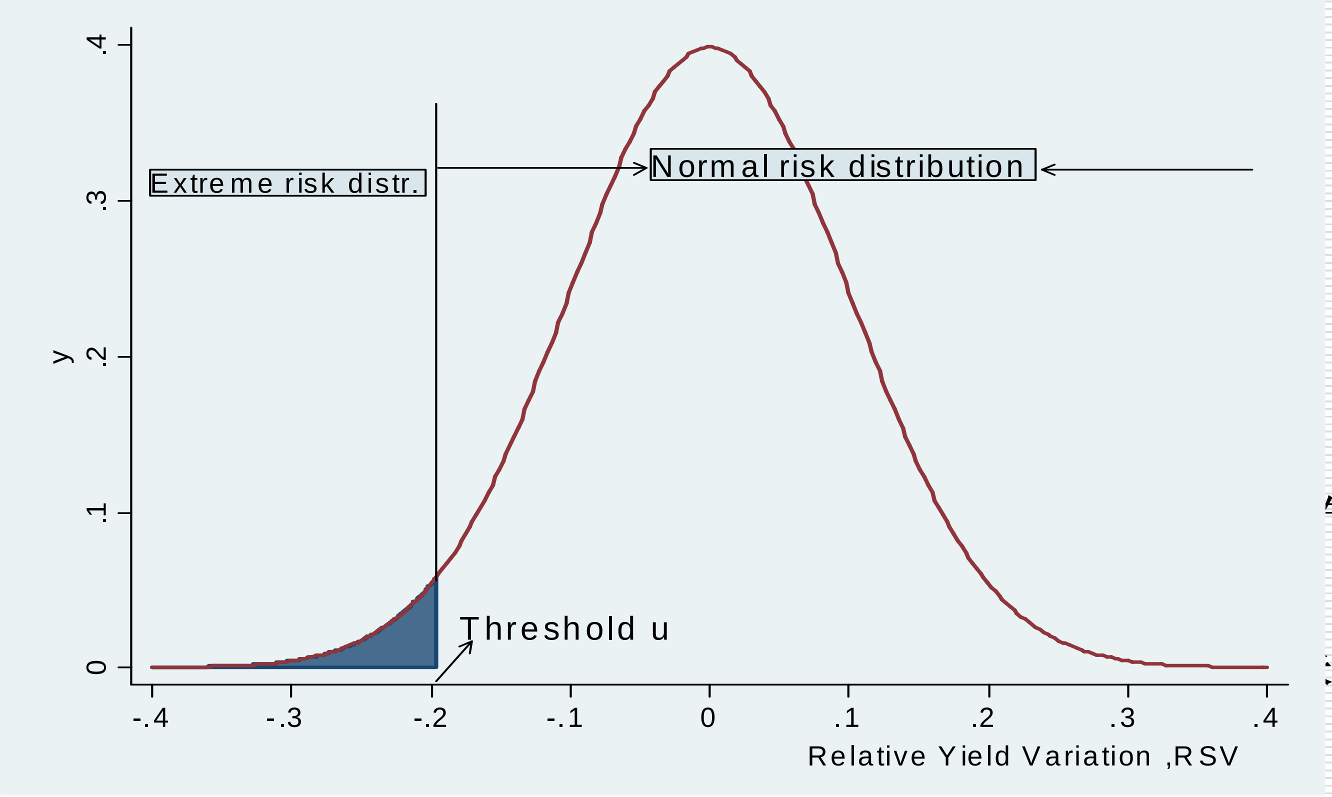
2. Agricultural yield risk assessment

- For overall yield risk(整体生产风险评估)

● As agricu
risk, the over
unknown dist
reflecting nor

● Threshol
the normal y
at the left tai
distribution w

如同农业生产
也是常规风险分
阀值**u**是将两
u在分布的左尾
值服从极端风险
年）一遇事件的



3. Approaches of Modeling yield risk

- Modeling normal yield risk(常规风险量化估计的方法)

Basically, th
be fall into two
对常规生产风险

Under this approach, a specifically prior distribution is selected firstly, and then it's parameters are estimated based on the observed data using MLE or GMM approach
所谓参数方法，就是首先假定其服从一个参数分布，然后应用极大似然估计利用样本数据估计出该分布的参数值。

Parametric method
参数方法

Advantage: work well when the underlying distribution is correct assigned. also have better Performance when sample is limited. 优点：适合于小样本情况，当已知潜在分布时效果良好
Weakness: have a prior assumption that one know the correct distribution which is not true in reality. may result in an imprecise prediction and misleading inference 缺点：存在先验性假设，有主观性，易产生误差

nonparametric method
非参数方法

3. Approaches of Modeling yield risk

- Modeling normal yield risk(常规风险拟和的方法)

Basically, the approaches used to modeling normal yield risk can be fall into two broad areas, parametric & nonparametric methods

对常规生产风险进行拟和的方法大体可以分为参数和非参数方法两类

Parametric method
参数方法

nonparametric method
非参数方法

Advantage: free of functional forms and distribution assumptions, impervious to specification errors and can result in more accurate and robust models. 优点：不用事先假定分布形态，分布形式自由，估计结果准确。

Weakness: reliance on a large sample and has the potential difficulty in measuring rare events.
缺点：需要大样本数据，不适用于小样本事件分布的估计

It is an alternative approach to modeling crop yield distributions, and the simplest one is the histogram and the commonly used one is the empirical distribution
非参数方法不用事先假定样本序列服从某种分布，直接用样本信息拟和作物分布。直方图是最简单的非参数方法，最常用的非参数方法是经验分布

3. Approaches of Modeling yield risk

Modeling extreme yield risk(极端风险量化估计的方法)

As the study of Hao(2005), extreme value model have obviously advantage in modeling and assessing extreme risk then traditional parametric method.

已有研究证实，相比传统方法，极值模型在拟和、评估极端风险方面具有明显的优势

Extreme value theory (EVT) dates back from the late 1920s to the early 1940s and have been extensively applied in many subjects during last several decades. From 1990s, the applications of EVT in modeling financial extremes have become more and more popular since early 1990s.

极值理论起源于**20**世纪**20**年代后期至**40**年代早期，近几十年来已在多个学科中得到发展应用。**90**年代以来，极值理论越来越多的应用于金融领域极端风险的建模

Generally, there are two principal kinds of approaches in modeling extreme values: the block maxima model (BMM) and the peak-over-threshold (POT) model.

总体而言，极值理论可以分为两个大类：组最大模型（**BMM**）和超越阀值模型（**POT**）

3. Approaches of Modeling yield risk

Block Maxima Model (BMM模型)

- The BMM approach focuses on the statistical behavior of the largest or smallest value in a sequence of independent random variables.

  BMM 模型是指在一系列独立随机事件中只关注最大值或最小值的一种极值统计方法

- Assume M_n be the maximum of the process over n independent random variables with a common distribution function F,

$$M_n = \max\{x_1, x_2, \cdots, x_n\}$$

$$p(M_n \le z) = p(x_1 \le z, x_2 \le z, \cdots, x_n \le z) = \{F(z)\}^n$$

3. Approaches of Modeling yield risk

Block Maxima Model (BMM模型)

- Normalize M_n to get a non-degenerated limiting distribution

$$\lim_{n\to\infty} p\left(\frac{M_n-\mu}{\sigma}\le z\right)=G(z)=\begin{cases}\exp\{-[1+\xi(\frac{x-\mu}{\sigma})]^{-1/\xi}\} & when\ \xi\neq 0\\ \exp\{-\exp[-(\frac{x-\mu}{\sigma})]\} & when\ \xi=0\end{cases}$$

Where G() belong the generalized extreme value (GEV) family, μ, σ *and* ξ are the location, scale and shape parameters respectively. G=Gumbel when ξ=0, weibull when ξ<0, and Frechet when ξ>0

G()是一个广义极值分布，μ，σ and ξ 分别是该分布的位置、尺度和形状参数。当ξ趋近于0时，G（）的具体分布形式为Gumbel分布， 小于0时为Weibull分布，大于0时为Frechet分布。

3. Approaches of Modeling yield risk

Peak over Threshold Model (POT模型)

- Since BMM only care about the largest or smallest events, it is an inefficient approach if other data on the tail are available and of interest and too narrow to be applied to a wide range of problems. POT can compensate such shortcomings and be used to model all large (small) observations that exceed (fall below) a high (low) threshold.

由于**BMM**模型只关注于最大值或最小值，当我们想要利用其它尾部信息时，该方法就无能为力了。另外，只关注最大值或最小值，就使本就样本量较小的尾部极值数据进一步减少。**POT**模型克服了**BMM**模型存在的这些缺点，可以用来对超过或低于某一阀值的样本数据进行分析、模拟和建模。

3. Approaches of Modeling yield risk

Peak over Threshold Model (POT模型)

- Assume u is the threshold, then the stochastic behavior of these events whose values are greater than threshold (u) can be represented by following conditional probability function. 假定**u**为阀值，则样本值在阀值**u**之上的事件为尾部事件，其分布可由如下条件概率函数表示，

$$P(X > u + y \mid X > u) = \frac{1 - F(u+y)}{1 - F(u)}, \qquad y > 0$$

- Pickands(1975), Balkema and Haan(1974) have shown that if the block maxima have an approximate distribution of GEV, then threshold excesses have a corresponding Generalized Pareto Distribution (GPD). The distribution function of (x-u) conditional on x>u can be approximated by, 已有研究证实，如果**BMM**模型服从渐近的广义极值分布的话，则超越阀值的样本也服从一个相对应的广义皮尔洛分布（**GPD**），阀值之上样本在**x>u**条件下的分布函数为，

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)^{-1/\xi}$$

$$where\ \sigma_u = \sigma + \xi(u - \mu)$$

3. Approaches of Modeling yield risk

Peak over Threshold Model (POT模型)

• In the POT model, the determination of the threshold u is crucial. Too low a threshold is likely to violate the asymptotic basis of the model and lead to a bias; too high a threshold will generate few observations left to estimate the parameters of the tail distribution function and cause high variance.

在**POT**模型中，阀值**u**的确定是利用**POT**模型取得良好效果的关键。阀值太低会使模型的渐近性得不到满足，产生偏差；而过高的阀值又会使可利用的样本容量过少，导致很高的误差。

• Threshold u can be determined by mean residual life plot (Coles,2001), goodness-of-fit test (Gumble,1958) and bootstrap method (de Haan, 1989). Among them, mean residual life plot is the most popular approach.

确定阀值**u**的方法有平均残差生命图、拟和优度检验和自举法，其中又以第一种最为常用。

• An ideal mean excess plot should be approximately a straight line in u with slope $\xi/(1-\xi)$. Empirically, the locus of point $\{(u, \frac{1}{n_i}\sum_{i=1}^{n_u}(x_{(i)}-u)) : u < x_{\max}\}$ is termed of the mean excess plot.

一个理想的平均残差图应该是一个截距为**u**，斜率为ξ **/ (1-**ξ**)**的直线。各个阀值水平下的平均超越值（**x-u, x>u)**形成的轨迹构成了平均残差图。

3. Approaches of Modeling yield risk

□ Parameters estimation of GEV&GPD (两种极值模型的参数估计）

• Maximum likelihood procedures (MLE) can be utilized to estimate the GEV parameters and the GPD parameters given the threshold u.
对于两种极值模型来说，极大似然估计法都可以用来估计两模型中参数的值

• The log-likelihood for GEV distribution is,
GEV 模型的对数似然函数形式为：

$$\ln L(\mu,\sigma,\xi,z) = -n\log(\sigma) - (1+1/\xi)\sum_{i=1}^{n}\log[1+\xi(\frac{z_i-\mu}{\sigma})] - \sum_{i=1}^{n}[1+\xi(\frac{z_i-\mu}{\sigma})]^{-1/\xi}$$

• The log-likelihood for GPD distribution is,
GPD 模型的对数似然函数形式为：

$$\ln L(\sigma,\xi) = -k\log\sigma - (1+1/\xi)\sum_{i=1}^{k}\log(1+\xi y_i/\sigma)$$

4. Case & Application

- Normal yield risk evaluation (常规生产风险评估)

  Illustrate by an example: Corn yield risk analysis and evaluation in China （中国玉米生产风险分析和评估）

- Extreme yield risk evaluation (极端生产风险评估)

  Illustrate by another example: Extreme Rainfall risk analysis in Jilin province （吉林省极端降雨风险分析评估）

Case: Corn yield risk evaluation in China

案例一，中国玉米生产风险分析

- **Background** (背景)

Corn is one of the main crop in China ,and the Corn Belt in China is from Northeast to Southwest.

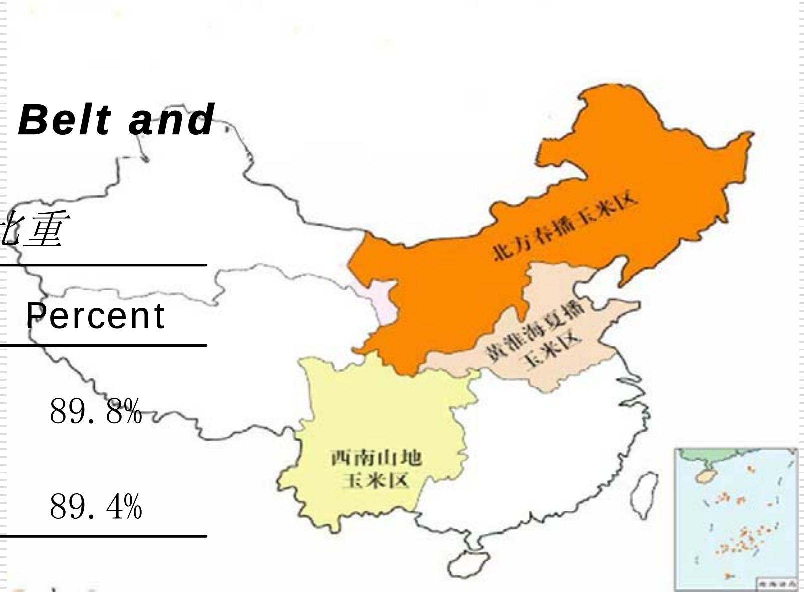
玉米是中国主要粮食作物之一，我国玉米生产分布大致呈现一个从东北到西南的斜长带。

The acreage and production in Corn Belt and whole country

主产区玉米播种面积和产量占全国的比重

	Corn Belt	Nation	Percent
Acreage（10^4Ha）	2213.43	2464.0779	89.8%
Output（10^4t）	10826.44	12114.725	89.4%

Average of 2000-2005



Background (背景)

We had selected eight province as the sample of Chinese Crop Belt, there are *Heilongjiang, Jilin, Liaoning, Shandong, Henan , Sichuan, Guizhou and Guangxi*)

我们选取了八个省份作为我国玉米主产区的样本代表，分别为：黑龙江、吉林、辽宁、山东、河南、四川、贵州和广西

The percent of Sample acreage &production to Corn Belt
八省份玉米播种面积和产量占主产区的比重

	HLJ	JL	LN	HN	SD	SC	GX	GZ	Sum
Acreage	9.5%	11.8%	7.0%	10.6%	11.3%	5.4%	2.5%	3.2%	61.4%
Output	8.5%	14.0%	8.2%	10.1%	13.8%	4.9%	1.6%	3.1%	64.1%

Average of 2000-2005

Analysis approach(分析方法)

Step 1: Collect yield data and got the relative stochastic variation (RSV)

第一步：收集单产数据，计算相对随机波动（RSV）

a: Modeling the yield trend use LMA

利用直线滑动平均法拟和作物单产趋势

b: Detrend（去除趋势）

剔除作物单产的趋势

Analysis approach (分析方法)

Step 2: Select candidate parameterization of yield distribution for different regions

第二步：采用参数方法，为不同地区选定合适的单产分布形式

a: Compare the skewness and kurtosis of samples with the moment ratio diagrams

将各样本的偏度、峰度值与距比率图进行比较

b: AD test

AD检验

Analysis approach (分析方法)

Step 3: Using MLE to estimate the parameters of yield distributions ,then got the PDF & CDF for different regions

第三步：利用极大似然估计法估计各单产分布的参数，得到各样本的概率密度函数和累积概率密度函数

Analysis results (分析结果)

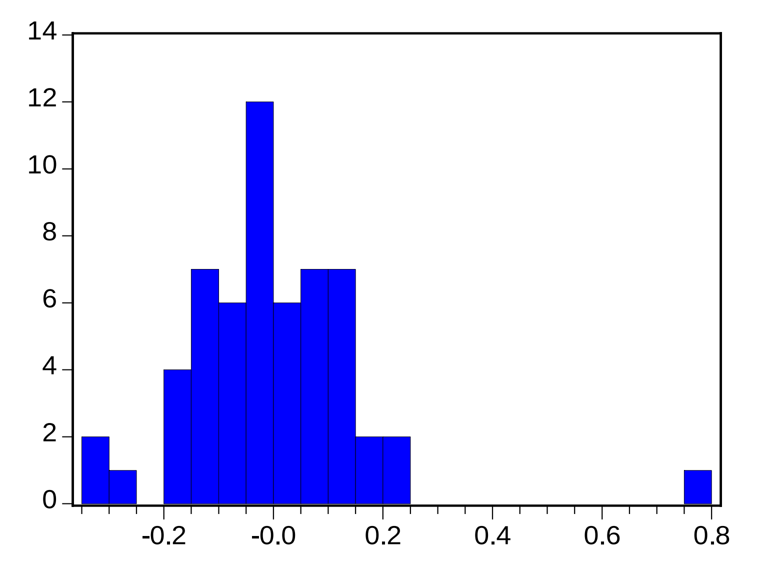
The histogram, PDF and yield distribution of these sample is:

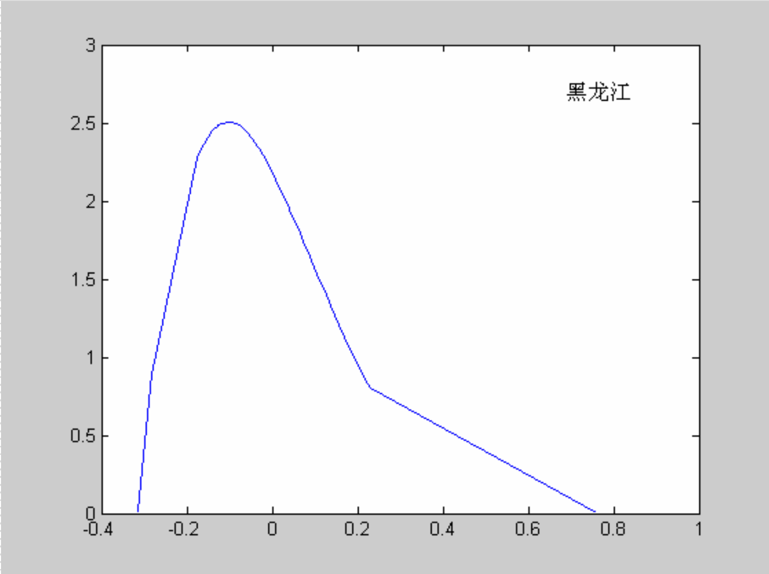
各样本省的单产分布直方图、概率密度函数和单产分布模拟图如下所示，

➢ Heilongjiang (Weibull distribution)

黑龙江（服从Weibull分布）

$$f(x) = 4.2873 \times \left(\frac{x}{1.7575}\right)^{6.5350} \times \exp\left[-\left(\frac{x}{1.7575}\right)^{7.5350}\right]$$

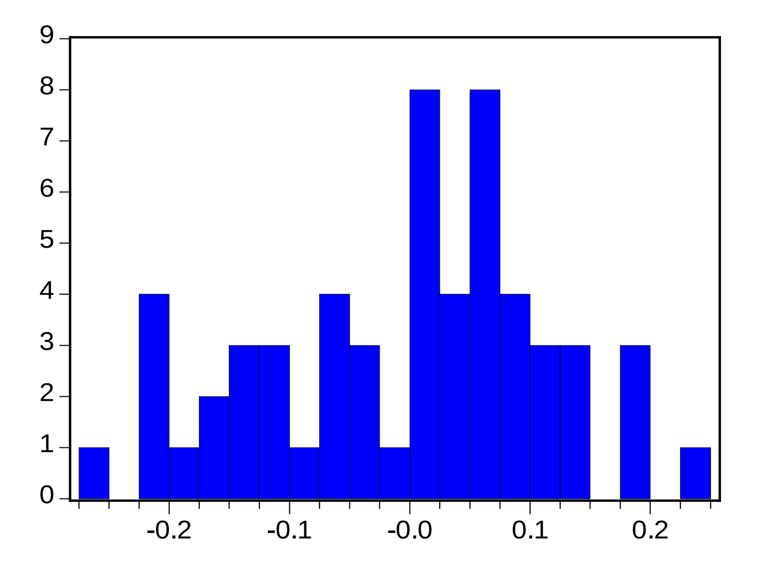


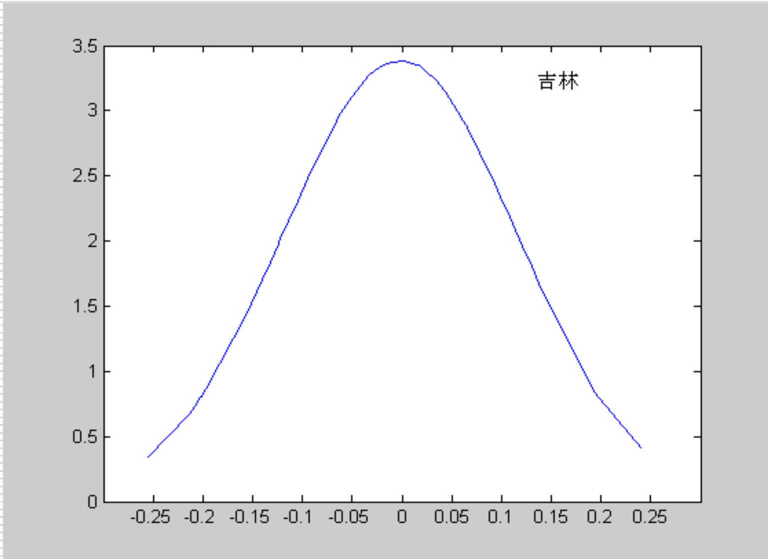


➢ Jilin (Normal distribution)

吉林 （服从正态分布）

$$f(x)=\frac{1}{0.2952}\exp[-\frac{(x+0.002014)^2}{0.0277}]$$

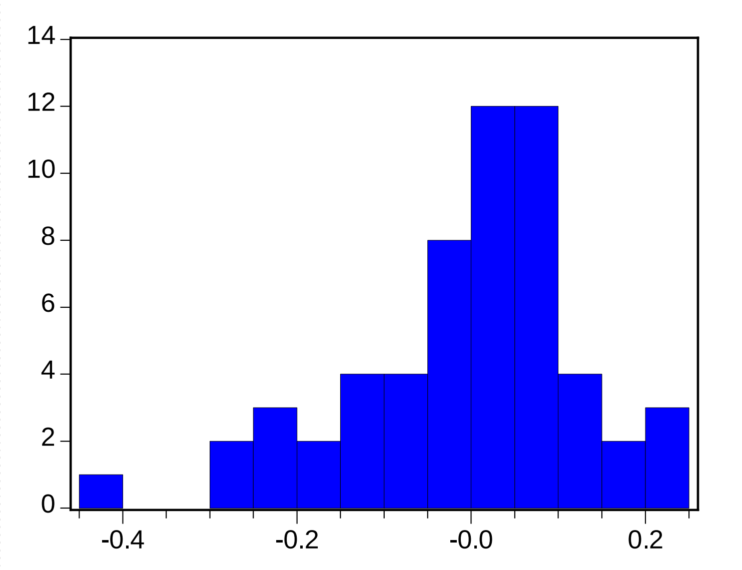


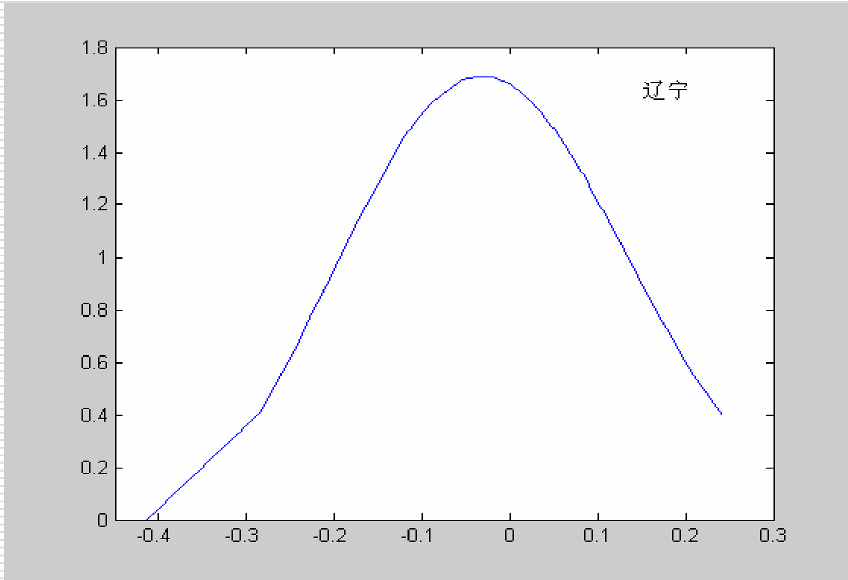


➢ Liaoning (Weibull distribution)

辽宁（服从Weibull分布）

$$f(x) = 1.0836 \times \left(\frac{x}{2.89}\right)^{2.1316} \times \exp\left[-\left(\frac{x}{2.89}\right)^{3.1316}\right]$$

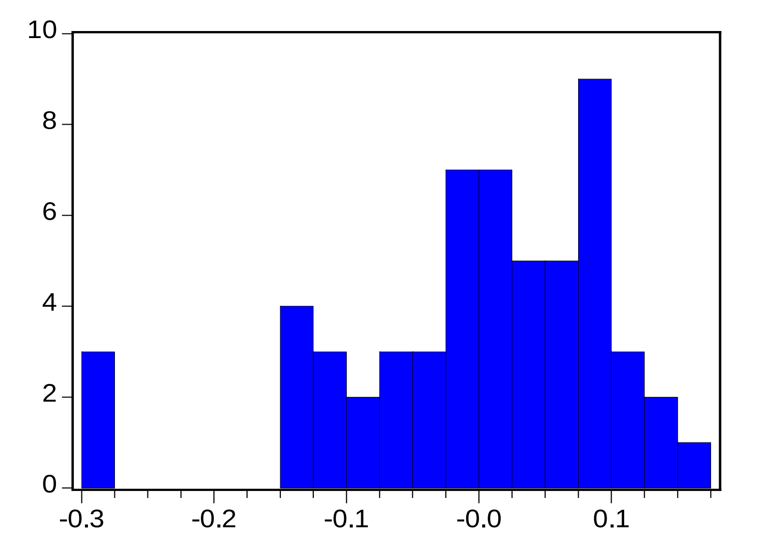


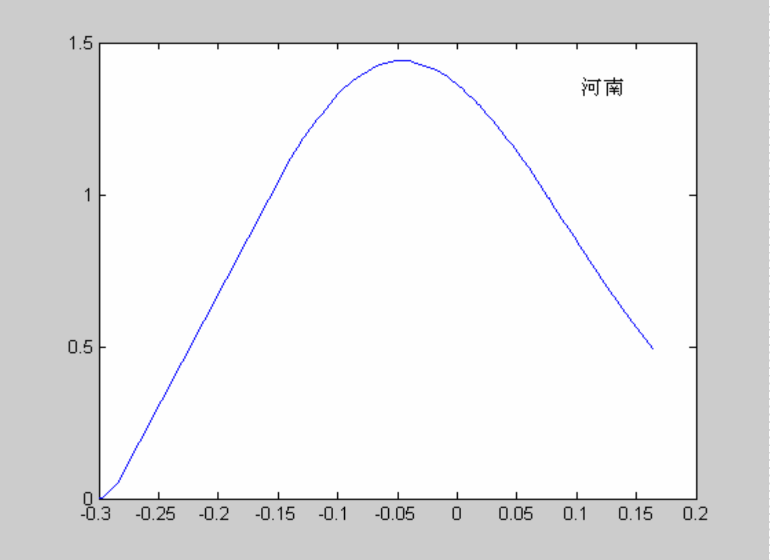


➢ Henan （Weibull distribution)

河南（服从Weibull分布）

$$f(x)=1.0455\times(\frac{x}{2.4114})^{1.5212}\times\exp[-(\frac{x}{2.4114})^{2.5212}]$$

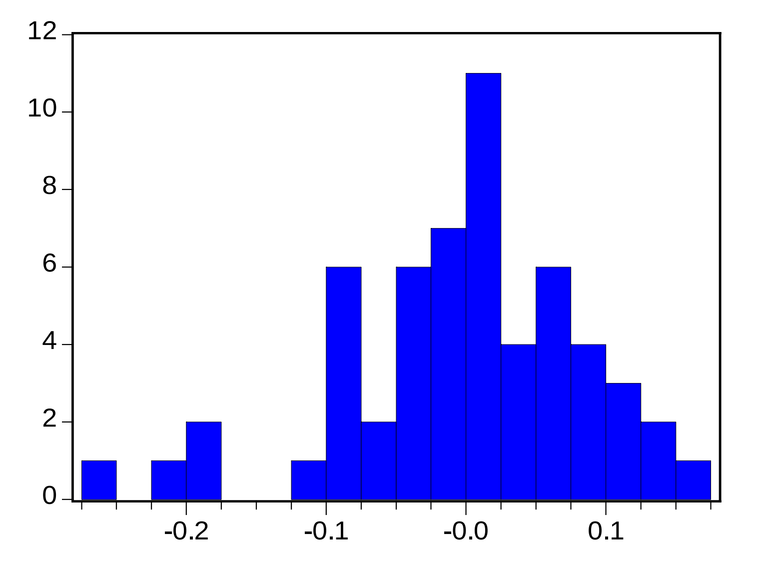


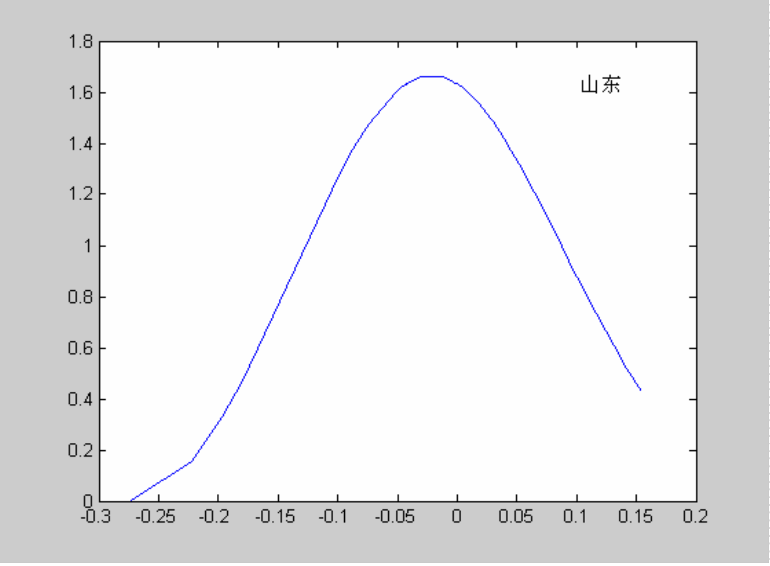


➢ Shandong (Weibull distribution)

山东（服从Weibull分布）

$$f(x) = 1.0428 \times \left(\frac{x}{2.8814}\right)^{2.0047} \times \exp\left[-\left(\frac{x}{2.8814}\right)^{3.0047}\right]$$

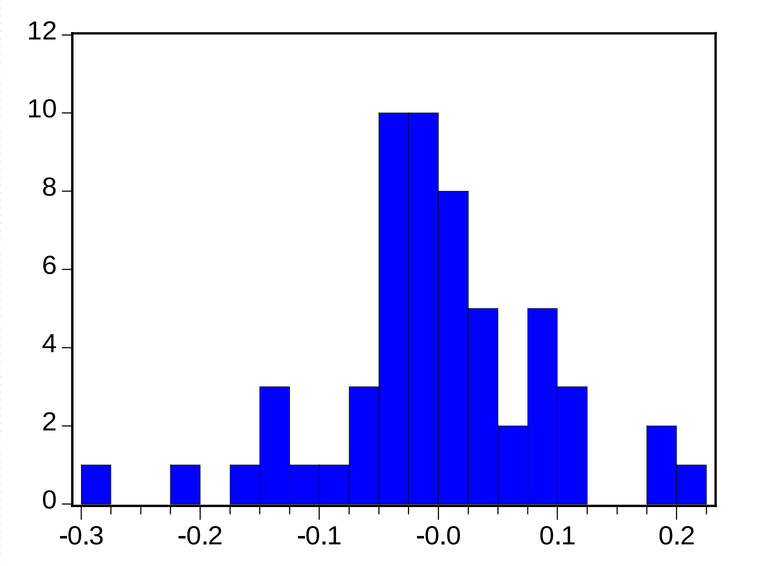


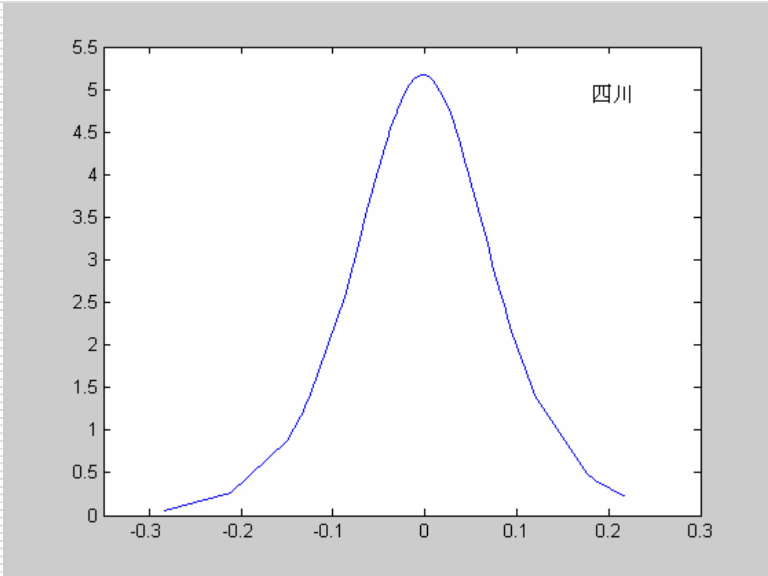


➢ Sichuan (Logistic distribution)

四川（服从Logistic分布）

$$f(x) = 0.048335 \times e^{-\frac{x+0.002068}{0.048335}} \times (1+e^{-\frac{x+0.002068}{0.048335}})^{-2}$$

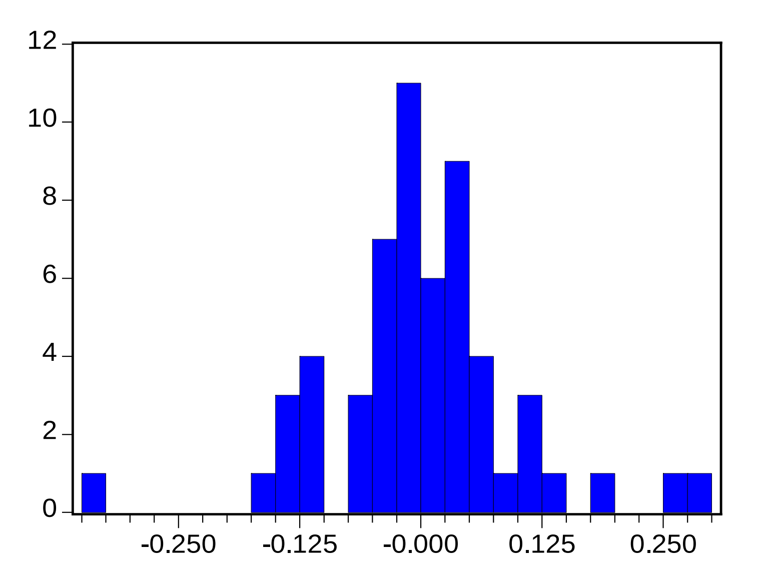


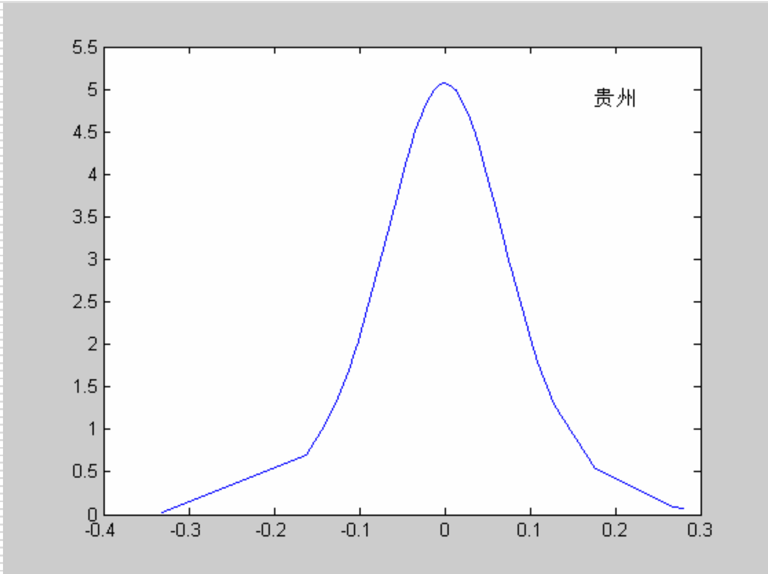


➢ Guizhou (Logistic distribution)

贵州（服从Logistic分布）

$$f(x)=\frac{1}{0.049286}\times e^{-\frac{x-0.000298}{0.049286}}\times(1+e^{-\frac{x-0.000298}{0.049286}})^{-2}$$

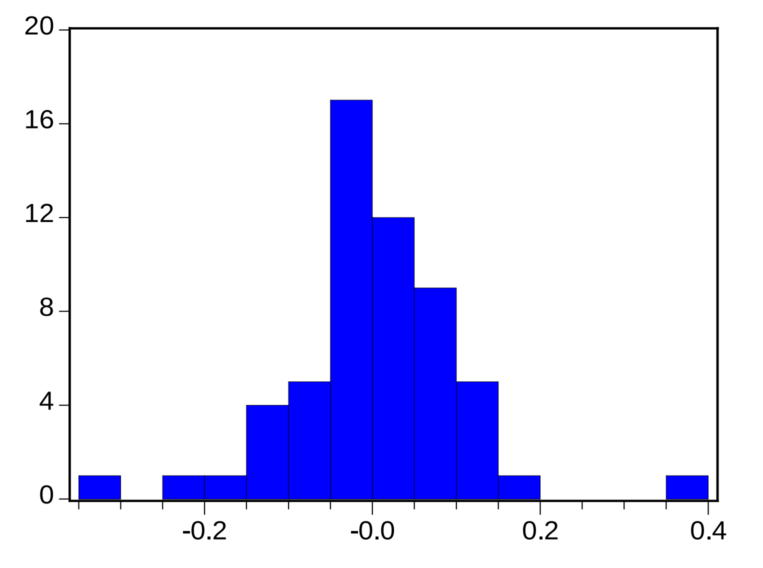


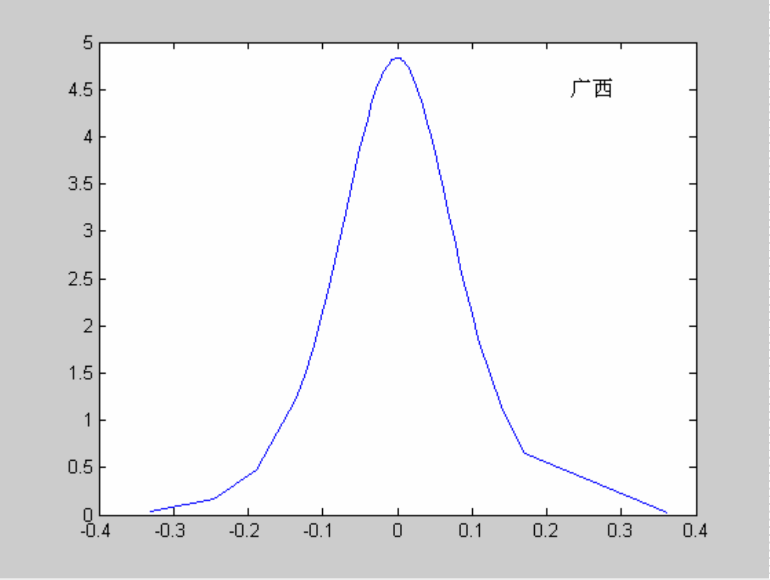


➢ Guangxi (Logistic distribution)

广西（服从Logistic分布）

$$f(x)=\frac{1}{0.051701}\times e^{-\frac{x+0.000622}{0.051701}}\times(1+e^{\frac{x+0.000622}{0.051701}})^{-2}$$





Case: Extreme rainfall risk analysis in Jilin province

案例二，吉林省极端降雨风险评估

In fact, the yield history data should be used to modeling the extreme yield risk as mentioned above. But, there are few yield observations that excess a given threshold. So, we just here utilized the maximum rainfall data per day of Jilin province to show how to use POT model.

事实上，按照上面的介绍，我们应该利用阀值之上的单产历史数据来对极端生产风险进行评估建模。然而，单产历史数据的样本量本身就不是太多，超越某一阀值的极端单产数据更少。因此，本报告中利用吉林省**1956-2007**年玉米生长季节日最大降雨量数据对吉林省的洪涝灾害风险进行评估，仅以此为例说明极值理论方法**POT**模型的应用。

Summary of the maximum rainfall level pre day from the twelve weather station in Jilin province

表 6　吉林省所选 12 站点至建站来年日降水最大 251 个值的特征总结

	N	Average	median	standard deviation	skewness	kurtosis	Maximum	Minimum
通榆县	251	354.99	303	166.63	2.349	7.263	1310	209
乾安县	251	355.13	304	162.12	1.794	3.371	1045	200
前郭	251	346.90	304	145.65	2.251	6.310	1062	215
抚余	251	399.87	331	224.07	3.841	19.154	1887	238
长岭县	251	381.31	321	187.96	3.097	15.266	1715	218
春市郊	251	438.67	363	204.50	2.006	4.105	1304	262
永吉和市郊	251	411.85	360	160.16	2.103	5.477	1193	271
蛟河	251	409.66	370	154.78	2.358	7.331	1289	267
桦甸	251	424.33	381	142.74	1.861	4.252	1156	290
双辽	251	397.91	335	197.26	2.687	9.143	1524	237
四平城郊	251	461.74	407	188.64	2.642	9.640	1571	294
敦化市	251	372.27	326	160.89	2.873	11.301	1387	239
Average		396.22	342.08	174.62	2.49	8.55	1370.25	245.00
Minimum		346.90	303	142.74	1.79	3.371	1045	200
Maximum		461.74	407	224.07	3.84	19.154	1887	294

资料来源：国家气象局　中国气象科学数据共享服务网

Source: Han(2008)

The determination of threshold

The mean residual life plot

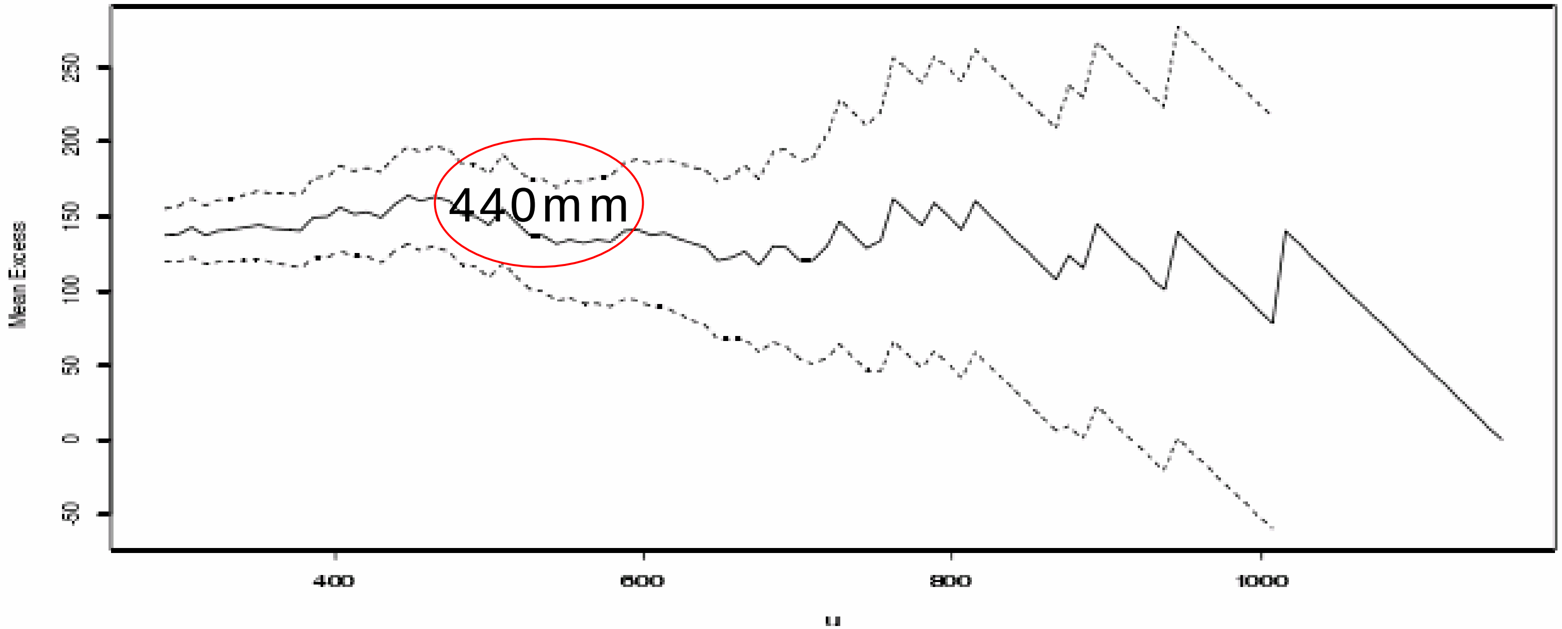


图 3 桦甸县的超额平均降水检验图。

Once the threshold u was determined, we can estimate the parameters of GPD using MLE approach. The parameters and distribution function of GPD is,

在确定阀值**u**以后，我们就可以利用**MLE**方法估计出广义皮尔洛分布的参数值，得到其分布函数为，

$$\sigma_u = 174.62 \;; \qquad \varepsilon = -0.1173$$

$$H(x) = 1 - \left(1 - 0.1173\frac{x}{174.62}\right)^{\frac{1}{0.1173}}$$

Model Check/模型检验

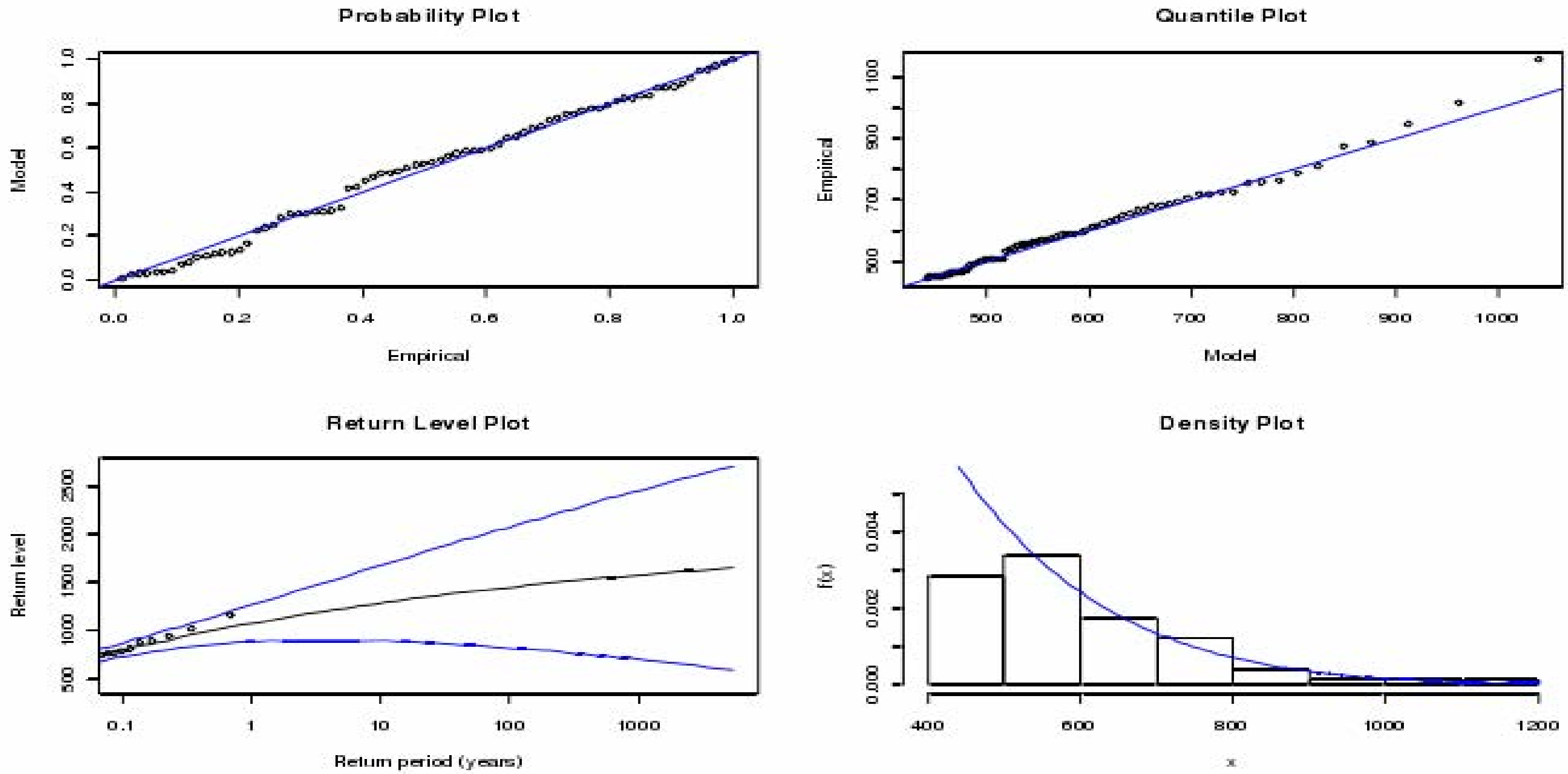


图 5　桦甸县极端降水分布图（包括 PP 图、QQ 图、Return Level 图、概率密度图）

Source：Han(2008)

Extreme Rainfall risk of Jilin Province

表 7 POT 法估计吉林省地区极端降水情况

	N	threshold	Numb exce TH	ex R	scale par MLE	scale par Std. Err	shape par MLE	shape par Std. Err	$E(U_0)$ (mm)	100-y prec
通榆县	251	620	15	21.8	300.95	105.5	-0.318	0.25	228.3	1482.70
乾安县	251	620	20	29.1	233.34	79.9	-0.470	0.28	158.73	1104.66
前郭	251	530	23	33.4	228.47	75.16	-0.267	0.264	180.32	1286.46
抚余	251	440	62	90.2	148.21	31.48	0.372	0.176	236.00	11869
长岭县	251	750	7	10.1	429.11	339.08	-0.203	0.73	356.6	3043.2
春市郊	251	700	29	42.2	287.22	91.68	-0.358	0.27	211.5	1462.11
永吉市郊	251	720	13	18.9	251.58	147.90	-0.381	0.536	181.75	1343.22
蛟河	251	660	15	21.8	301.05	102.24	-0.384	0.248	217.52	1403.16
桦甸	251	440	77	112.9	174.62	26.76	-0.117	0.103	156.32	1430.37
双辽	251	660	17	24.7	446.75	152.79	-0.429	0.271	106.92	1663.03
四平城郊	251	760	12	17.4	402.38	159.92	-0.370	0.343	293.43	1778.42
敦化市	251	590	19	27.6	274.05	85.46	-0.173	0.220	190.96	1768.47

Note:

N..TH=number of the observations that excess the threshold 超过阀值的个数

Ex R=Excess rate per year 年降水量超越门阀值的比率

E(U)=mean of the excess value against threshold 降雨量超过阀值的平均水平

5. Discussion & future plan
讨论及下一步工作打算

We need to investigate the relation between extreme weather events and the yield loss, then we can evaluate the crop overall yield risk using traditional statistic and EVT approach

下一步，我们需要找出作物单产水平、极端损失与极端气象事件，如降雨量之间的明确定量关系。这样，我们就能够通过对气象变量的分析得出作物极端生产风险水平的评估。


Thanks