基于GIOWA算子的安徽省全社会用电量组合预测研究

韦晨環娃 杨桂元 袁宏俊

（安徽财经大学数量经济研究所，安徽蚌埠 233030）

【摘 要】选取2000—2014年安徽省全社会用电量数据，采用多元线性回归、二次指数平滑以及多项式拟合3种预测方法，给出了单项预测值。引入广义诱导有序加权平均（GIOWA)算子，选取其3种特殊形式IOWA、IOWHA、IOWGA的算子，建立误差平方和最小组合预测祺型,并通过模型评价发现.这3种组合预测模型的误差均小于单项预测值，验证了模型的有效性。利用所建模型预测了未来五年的安徽省全社会用电总量，为电力规划和需求侧管理提供决策基础。

【关键词】安徽省；全社会用电量；组合预测；GIOWA算子；指数平滑；多元线性回归

【中图分类号】F224.0**【文献标识码】**A

【文章编号】1671-3079(2017)01-0057-07

电力需求分析与预测是电力系统规划和运行的基础工作，电力消费直接带动产业发展，拉动内需。近年来，国内电力需求一直呈上涨态势且幅度较大，预计今后一段时期内仍然会保持增长。全国电力需求总体来说是平衡的，但是局部地区也会出现电力供需紧张的情况。只有保持电力供需平衡，才不会影响到经济发展和居民生活问题。科学合理地预测全社会用电水平，将为电力规划和需求侧管理提供决策基础。全社会用电量预测结果是否准确，直接关系到电网安全以及供电的可靠性，影响电网经营企业的经营决策与经济效益。

安撤省目前正处于“转型、升级、保增长”的关键时期，电力供需变化直接影响安徽省的经济发展。由于电量的月度数据波动较大，要考虑到季节效应等问题，所以本文选取安徽省的全社会用电量的年度数据进行研究。本文采用多元线性回归预测，选取相关指标并做检验，解决了多重共线性以及异方差等问题并得到了比较满意的函数模型，同时，运用二次指数平滑预测以及多项式拟合预测，引入广义诱导有序加权平均算子，改变了传统组合预测固定权重的做法，附加诱导因子，建立误差平方和最小组合预测模型，提高预测精度，旨在为供电部门提供决策参考。

一、安徽省全社会用电量单项预测

（一）多元线性回归预测

根据相关文献以及资料整理，本文从《安徽省统计年鉴》中选取了2000－2014年六个自变量的数据，分别为人口（x_1/万人）、GDP（x_2/亿元）、能源消费总量（x_3/万吨）、居民消费水平（x_4/元）、CPI（x_5）以及城镇家庭人均可支配收入（x_6/元），因变量为安徽省全社会用电总量（y/亿千瓦时）。将数据导入Eviews3.1，[1]并做出各变量之间的变化趋势图。从趋势图里可以看出，自变量人口（x_1/万人）及CPI（x_5）与全社会用电总量y没有明显的线性关系，可剔除两个变量。对剩下的变量进行多元线性回归，得到的结果见表1。

表1　4个变量初步的回归结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−243.773 500	69.213 290	−3.522 062	0.0055
x_2	−0.017 458	0.026 559	−0.657 355	0.5258
x_3	0.056 014	0.017 622	3.178 691	0.0098
x_4	−0.006 895	0.045 431	−0.151 763	0.8824
x_6	0.065 173	0.013 707	4.754 715	0.0008
R-squared	0.998 463	Mean dependent var		839.820 000
Adjusted R-squared	0.997 848	S.D. dependent var		432.284 200
S.E. of regression	20.055 660	Akaike info criterion		9.096 102
Sum squared resid	4022.296 000	Schwarz criterion		9.332 118
Log likelihood	−63.220 760	F-statistic		1623.546 000
Durbin-Watson stat	2.438 582	Prob(F-statistic)		0.000 000

模型的计算结果表明，$R^2 = 0.998\ 463$，说明模型有很高的拟合优度，F检验也是高度显著的。但模型中一些变量的t统计量值较小，未通过检验。因此，需要对以上多元线性回归模型做适当的调整，按照统计检验程序，应该剔除变量而重新建立模型。下面考虑多重共线性的问题。

首先，做出x_2、x_3、x_4、x_6的相关系数矩阵。由相关系数矩阵可知，两两变量之间的相关系数较高，则证明存在多重共线性。因此，采用逐步回归法去检验和解决多重共线性问题。分别作出y与各自变量之间的一元回归，结果如表2所示。

表2　一元回归估计

变量	x_2	x_3	x_4	x_6
参数估计值	0.0697	0.1803	0.1229	0.0654
t统计量	30.9006	35.8721	31.9349	57.0106
R^2	0.9866	0.9899	0.9874	0.9961
$\bar{R}^2$	0.9855	0.9892	0.8640	0.9957

其中，加入x_6的方程$\bar{R}^2$最大，以x_6为基础，顺次加入其他变量逐步回归，结果如表3。经比较可以发现，加入x_2、x_4后参数的符号与预期相反，不符合经济意义，且t检验不显著；加入x_3后修正的$\bar{R}^2$有所提高，参数符号的经济意义合理，参数的t检验显著，所以保留x_3。再加入其他变量逐步回归。当加入x_2、x_4时，虽然$\bar{R}^2$有所增加，但其系数的符号与预期相反且参数的t检验不显著，这说明x_2、x_4引起了多重共线性，应予以剔除。将y与x_3、x_6进行拟合，得表4。

表 3　逐步回归结果

模型	x_6	x_2	x_3	x_4	$\bar{R}^2$
x_6、x_2	0.0910 (6.2427)	−0.0276 (−1.7656)			0.9968
x_6、x_3	0.0437 (6.9177)		0.0602 (3.4411)		0.9980
x_6、x_4	0.0841 (5.5484)			−0.0356 (−1.2416)	0.9965
x_6、x_3、x_2	0.0651 (4.9785)	−0.0210 (−1.8201)	0.0553 (3.4018)		0.9985
x_6、x_3、x_4	0.0617 (5.0067)		0.0595 (3.6347)	−0.0335 (−1.6588)	0.9984

表 4　最终回归结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−213.337 000	64.803 220	−3.292 075	0.0064
x_3	0.060 271	0.017 515	3.441 166	0.0049
x_6	0.043 772	0.006 327	6.917 745	0.0000
R-squared	0.997 995	Mean dependent var		839.820 000
Adjusted R-squared	0.997 661	S.D. dependent var		432.284 200
S.E. of regression	20.908 080	Akaike info criterion		9.095 005
Sum squared resid	5245.773 000	Schwarz criterion		9.236 615
Log likelihood	−65.212 540	F-statistic		2986.324 000
Durbin-Watson stat	1.828 768	Prob(F-statistic)		0.000 000

由表 4 可知，各参数 t 检验值均显著说明在其他解释变量不变的情况下，解释变量“能源消费总量（x_3/万吨）”与“家庭人均可支配收入（x_6/元）”分别对被解释变量“全社会用电量（y/亿千瓦时）”有显著影响。F 值很大，说明变量联合对 y 也有显著影响。DW 值为 1.8288，R 值为 0.9979，可见模型拟合效果较好。

对模型进行异方差检验可得辅助回归模型的 F 统计量值在显著水平 $\alpha = 0.05$ 的情况下，$\chi^2_{0.05}(2) = 5.99 > nR^2 = 4.3206$，模型不存在异方差性，得到最终的拟合函数如下：

$$\hat{Y} = -213.337 + 0.0603x_3 + 0.0438x_6 \tag{1}$$

这说明在其他因素不变的情况下，当能源消费总量 x_3 每增加 1 单位，家庭人均可支配收入 x_6 每增加 1 元时，安徽省全社会用电总量将分别增加 0.060 3 亿千瓦时、0.0438 亿千瓦时。由此可见，能源消费总量对全社会用电总量的影响比家庭人均可支配收入对全社会用电总量的影响要大。因此，对自变量 x_3 与 x_6 的 2000−2014 年数据进行预测得到未来五年数据代入函数，[2-4] 可得安徽省全社会用电总量 y 的未来五年预测值。

（二）指数平滑预测

指数平滑法作为一种时间序列的预测方法，具有预测准确、方法简单等优点，尤其对于那些不存在明显变化趋势的序列，预测效果良好。对于变化趋势明显的序列，采用二次指数平滑法对其预测，效果会更好。二次指数平滑法就是对一次指数平滑再进行一次平滑。

一次指数平滑预测方程形式：

$$\begin{aligned} \hat{Y}_{t+1} &= aY_t + (1-a)\hat{Y}_t \\ S_t^{(1)} &= aY_t + (1-a)S_{t-1}^{(1)} \end{aligned} \tag{2}$$

二次指数平滑预测方程的形式为：

$$\hat{Y}_{t+T} = a_t + b_t T$$

$$\begin{aligned} S_t^{(2)} &= \alpha S_{t-1}^{(1)} + (1-\alpha) S_{t-1}^{(2)} \\ a_t &= 2S_t^{(1)} - S_t^{(2)} \\ b_t &= \frac{\alpha}{1-\alpha}(S_t^{(1)} - S_t^{(2)}) \end{aligned} \qquad (3)$$

其中 Y_1，Y_2，…，Y_n 为一时间序列，α 为平滑系数，$S_t^{(1)}$、$S_{t+1}^{(1)}$ 为第 t、$t+1$ 期的一次平滑指数值，$S_t^{(2)}$、$S_{t+1}^{(2)}$ 为第 t、$t+1$ 期的二次平滑指数值，a_t 为线性回归方程的截距项，b_t 为线性回归方程的斜率，T 为以第 t 期为起始点的预测样本的期末值。

将安徽省全社会用电总量数据（单位：亿千瓦时）导入 Eviews，做出其趋势图如图 1。

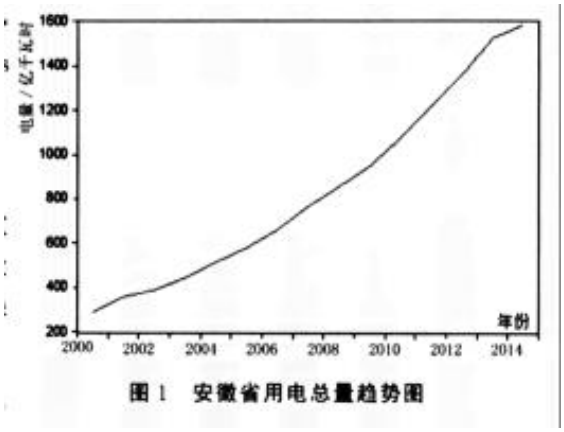

图 1　安徽省用电总量趋势图

根据图 1 可以看出安徽省全社会用电量序列增长趋势明显，采用二次指数平滑法预测社会用电量，通过 Eviews 操作可以得到未来五年预测数据。

（三）多项式拟合预测

已知离散点上的数据集 $[(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)]$，求得一解析函数 $y = f(x)$，使 $f(x)$ 在原离散点 x_i 上尽可能接近给定 y_i 的值，这一过程叫曲线拟合。最常用的曲线拟合是最小二乘法曲线拟合，拟合结果可使误差的平方和最小，即找出使下式最小的 $f(x)$：

$$\min \sum_{i=1}^{n} (f(x_i) - y_i)^2 \qquad (4)$$

根据图 1 的用电趋势图，选取三次多项式拟合，在 2000－2014 年，x 值分别取 1～15 之间的整数，以此类推，可得三次多项式函数如下：

$$y = -0.0751x^3 + 5.9448x^2 + 17.059x + 284.05 \tag{5}$$

公式（5）的可决系数为 0.9979，可见函数拟合效果较好。

下面对 x 分别取 16～20 之间的整数，可得 2015－2019 年的五年预测数据。将以上三种方法预测所得出的 2010－2019 年的数据列出来，如表 5 所示，其中 2010－2014 年数据用于下述组合预测模型的评价与检验，2015－2019 年数据用于未来五年的安徽省全社会用电总量的组合预测问题。

表 5　安徽省全社会用电总量单项预测值与实际值对比　　亿千瓦时

年份	实际值	多元线性回归预测	指数平滑预测	多项式拟合预测
2010	1077.92	1024.25	1044.90	1091.06
2011	1221.19	1172.53	1184.67	1215.04
2012	1361.10	1398.18	1346.16	1345.49
2013	1528.07	1546.62	1495.43	1481.98
2014	1585.18	1635.68	1677.48	1624.05
2015		1704.87	1698.16	1771.25
2016		1809.78	1803.49	1923.13
2017		1914.69	1908.81	2079.24
2018		2019.60	2014.14	2239.13
2019		2124.51	2119.47	2402.35

二、安徽省全社会用电量组合预测模型

（一）相关概念

设 $<v_1, a_1>, <v_2, a_2>, \cdots, <v_m, a_m>$ 为 m 个二维数组，令

$$f_w(<v_1, a_1>, <v_2, a_2>, \cdots, <v_m, a_m>) = \left(\sum_{i=1}^{m}(w_i a_{v-index(i)})^{\lambda}\right)^{\frac{1}{\lambda}} \tag{6}$$

则称函数 f_w 是由 $v_1, v_2, \cdots, v_m$ 所产生的 m 维广义诱导有序加权平均算子，简记为 GIOWA 算子。其中，v_i 称为 a_i 的诱导值，$v-index(i)$ 表示 $v_1, v_2, \cdots, v_m$ 中按从大到小顺序排列的第 i 个大的数的下标。[5] $W = (w_1, w_2, \cdots, w_m)^T$ 是加权向量，满足 $\sum_{i=1}^{m} w_i = 1$，$w_i \in [0, 1]$，$i = 1, 2, \cdots, m$。

对于广义诱导有序加权平均算子，可以对其中的 λ 取不同值来构造常见的几种算子：

1. 当 $\lambda = 1$ 时，为诱导有序加权算术平均（IOWA）算子，其形式为：

$$f_w(<v_1, a_1>, <v_2, a_2>, \cdots, <v_m, a_m>) = \sum_{i=1}^{m} w_i a_{v-index(i)} \tag{7}$$

2. 当 $\lambda \to 0$ 时，为诱导有序加权几何平均（IOWGA）算子，其形式为：

$$f_w(<v_1, a_1>, <v_2, a_2>, \cdots, <v_m, a_m>) = \prod_{i=1}^{m} a_{v-index(i)}^{w_i} \tag{8}$$

3. 当 $\lambda = -1$ 时，为诱导有序加权调和平均（IOWHA）算子，其形式为：

$$f_w(<v_1, a_1>, <v_2, a_2>, \cdots, <v_m, a_m>) = 1/\sum_{i=1}^{m}(w_i / a_{v-index(i)}) \tag{9}$$

（二）模型的建立

某一指标观察值序列为 $x_t(t=1,2,\cdots,N)$，假设有 m 个单项预测方法对该序列进行预测；第 i 种单项预测方法第 t 时刻的预测值记为 x_{it}（$t=1,2,\cdots,N$；$i=1,2,\cdots,m$）；$w_1,w_2,\cdots,w_m$ 为 m 种单项预测方法在组合预测中的加权系数，$\sum_{i=1}^{m}w_i=1$，（$w_i\in[0,1]$；$i=1,2,\cdots,m$）。

本文把预测精度 $v_{it}\in[0,1]$ 作为预测值的诱导值，预测精度用下式表示：

$$v_{it}=\begin{cases}1-|(x_t-x_{it})/x_t| & |(x_t-x_{it})/x_t|<1\\ 0 & |(x_t-x_{it})/x_t|\geqslant 1\end{cases} \tag{10}$$

其中 v_{it} 表示第 i 种方法在 t 时刻的预测精度，即为 x_{it} 的诱导值（其中 $t=1,2,\cdots,N$；$i=1,2,\cdots,m$）。这 m 种单项预测方法在 t 时刻的预测精度与其对应的预测值就组成 m 个二维向量 $<v_{1t},x_{1t}>$，$<v_{2t},x_{2t}>$，$\cdots$，$<v_{mt},x_{mt}>$，设 $W=(w_1,w_2,w_3,\cdots,w_m)^T$ 为各种单项预测方法在组合预测中的加权向量，按照从大到小的顺序对 m 种单项预测方法在 t 时刻的预测精度序列 $v_{1t},v_{2t},\cdots,v_{mt}$ 进行排列，$v-index(it)$ 表示第 i 种方法的预测精度的下标。为了求解方便，将对数误差平方作为优化准则。对于 $\lambda=1$、$\lambda\to 0$、$\lambda=-1$ 的 3 种不同情况，[6] 分别求得误差为：

$$e_{v-index(it)}=\begin{cases}x_t-x_{v-index(it)}\lambda=1\\ \ln x_t-\ln x_{v-index(it)}\lambda\to 0\\ \dfrac{1}{x_t}-\dfrac{1}{x_{v-index(it)}}\lambda=-1\end{cases} \tag{11}$$

再分别将公式（7）～（9）代入，可得 3 种组合预测方法的误差为：

$$e_t=\sum_{i=1}^{m}w_ie_{v-index(it)} \tag{12}$$

则前 N 期的预测误差平方和 S 可以表示为：

$$S=\sum_{t=1}^{N}(\sum_{i=1}^{m}w_ie_{v-index(it)})^2=\sum_{i=1}^{m}\sum_{j=1}^{m}w_iw_j(\sum_{t=1}^{N}e_{v-index(it)}e_{v-index(jt)}) \tag{13}$$

在误差平方和最小准则下，基于 GIOWA 算子组合预测模型如下形式：

$$\min S(W)=\sum_{i=1}^{m}\sum_{j=1}^{m}w_iw_j(\sum_{t=1}^{N}e_{v-index(it)}e_{v-index(jt)})$$

$$s.\ t.\begin{cases}\sum_{i=1}^{m}w_i=1\\ w_i\geqslant 0,\ i=1,2,\cdots,m\end{cases} \tag{14}$$

利用 lingo11 解（14）得到 3 种组合预测模型下的最优权向量为：

$$\begin{aligned}W_{\text{IOWA}}&=(0.8437,0.1563,0)\\ W_{\text{IOWGA}}&=(0.8311,0.1689,0)\\ W_{\text{IOWHA}}&=(0.8262,0.1576,0.0162)\end{aligned} \tag{15}$$

根据上述理论以及公式（15)求得的权重可作出 2010—2014 年的组合预测值对比，见表 6。

表 6　安徽省全社会用电量实际值与 3 种组合预测值对比　　亿千瓦时

年份	实际值	$\lambda=1$（IOWA）		$\lambda\to 0$（IOWGA）		$\lambda=-1$（IOWHA）	
		预测值	相对误差/%	预测值	相对误差/%	预测值	相对误差/%
2010	1077.92	1083.8495	0.55	1083.2646	0.55	1082.6998	0.55
2011	1221.19	1210.2908	0.89	1209.9060	0.89	1209.5576	0.89
2012	1361.10	1346.0559	1.11	1346.0474	1.11	1346.8870	1.11
2013	1528.07	1538.6249	0.69	1537.9763	0.69	1537.5012	0.69
2014	1585.18	1625.8687	2.57	1626.0160	2.57	1626.7438	2.57

由表 6 可得 3 种组合预测的预测值误差均较小，为了证明模型的有效性，下面选取 2010—2014 年的数据进行模型评价。

（三）模型评价

为了反映上文提出的基于 GIOWA 算子的误差平方和最小的组合预测模型对于安徽省全社会用电量预测的有效性，按照预测效果评价原则，选择以下常规的误差指标作为衡 M 精度的评价指标体系。

1. 误差平方和

$$SSE = \sum_{t=1}^{n} (x_t - \hat{x}_t)^2 \tag{16}$$

2. 均方误差

$$MSE = \frac{1}{n}\sqrt{\sum_{t=1}^{n} (x_t - \hat{x}_t)^2} \tag{17}$$

3. 平均绝对误差

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |x_t - \hat{x}_t| \tag{18}$$

4. 平均绝对百分比误差

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{x_t - \hat{x}_t}{x_t}\right| \tag{19}$$

5. 均方百分比误差

$$MSPE = \frac{1}{n}\sqrt{\sum_{t=1}^{n} \left(\frac{x_t - \hat{x}_t}{x_t}\right)^2} \tag{20}$$

表 7　预测指标评价体系

预测效果评价指标体系	SSE	MSE	MAE	MAPE	MSPE
二次指数平滑预测	12 231.9186	22.1196	41.8836	0.0302	0.0152
多项式拟合预测	9517.3091	19.5113	41.6918	0.0322	0.0155
多元线性回归预测	4089.2720	12.7895	23.9724	0.0167	0.0085
$\lambda = 1$(IOWA)	2147.2516	9.2677	16.6233	0.0116	0.0061
$\lambda \to 0$(IOWGA)	2148.1843	9.2697	16.4847	0.0115	0.0061
$\lambda = -1$(IOWHA)	2176.6697	9.3309	16.3240	0.0114	0.0062

从表 7 可以看出，本文提出的基于 GIOWA 算子的 3 种组合预测模型的各种误差指标值均明显低于三种单项预测的结果，从而表明本文提出的组合预测方法优于单项预测方法，证明了模型的有效性。下面将模型应用于安徽省全社会用电量未来五年的预测。

（四）模型应用

本文基于广义诱导有序加权平均算子的概念，提出了三种组合预测模型，并通过线性规划模型求出了组合预测的最优权系数 $w^* = (w_1^*, w_2^*, w_3^*)$，根据预测的连贯性原则，可以用其来进行未来五年的安徽省全社会用电量的组合预测，公式为：

$$f_w(<v_{1t}, x_{1t}>, <v_{2t}, x_{2t}>, <v_{3t}, x_{3t}>) = \sum_{i=1}^{3} w_i^* x_{v-order(i)} \quad (21)$$

若要对未来 k 年进行预测[4]，第 i 种预测方法的预测精度用最近 k 期拟合平均精度 $\frac{1}{k}\sum_{t=N-k+1}^{N} v_t$ 反映，本文中 k 为 5。根据以上公式以及前文中所求的最优权重，可得组合预测值如表 8 所示。

表 8　安徽省全社会用电总量预测　　亿千瓦时

组合预测模型	2015	2016	2017	2018	2019
$\lambda = 1$ (IOWA)	1759.8326	1905.4218	2053.5319	2204.8303	2358.9368
$\lambda \to 0$ (IOWGA)	1758.9064	1903.9854	2051.4467	2202.0485	2355.4162
$\lambda = -1$ (IOWHA)	1758.6426	1903.3093	2050.5173	2200.8466	2353.9250

三、结语

本文以误差平方和为 0 标函数，引入 GK)WA 算子改变了传统组合预测权重固定的做法，附加诱导因子将不同时点的预测对象重新赋予权重，较好地适应了经济发展变化的状态。取不同的参数值，得到的基于 IOWA、IOWGA、KOWHA 三种算子的组合预测模型，并通过预测效果评价指标体系，发现组合预测模型的预测精度远远高于其他三种单向预测方法。实例说明组合模型克服了单项预测和传统组合预测的缺点，大大提高预测精度。

从对安徽省未来五年的全社会用电量进行预测可知，该省未来时期的电力需求仍会呈上涨趋势。分产业看，除了第二产业用电平稳增长，第一产业以及第三产业用电增长都较快。主要是由于畜牧业、农用排灌以及交通运输业尤其是电气化铁路、电信和信息传输业用电高速增长。但对比之前几年，增长趋势减缓。“十二五”期间安傲省全社会用电量平均增速为 8.8%，受外部经济增速趋缓因素影响，未来的用电增速越来越慢。但用电笛求增速下降，电力供应紧张却没有缓解•这是因为目前电源在建规模严重不足，“十三五”期间电力供需矛盾将依旧突出，如何平衡电量需求是当前需要解决的一个重要问题。

为了有效预防电力短缺的现象，需要从年度数据出发，建立长期规划，从年度到月度再到日度规划，合理地安排电量供给，改善该省用电高峰的电量短缺情况。未来时期，安徽省在电力工业发展方面除了要保持适度增长外，还必须切实转变电力增长方式，注重质 M 和效率，实现能源与经济的和谐发展。

参考文献:

[1]张晓峒.计 M 经济学软件 Eviews 使用指南[M].天津：南歼大学出版社.2004:35—48.

[2]李昉•罗汉武•基于多元线性问归押论的河南宵全用电《预测[J].电网技术.2008(S1):124—126.

[3]王文圣，丁晶，赵五龙，等.基于偏最小二乘间归的年社会用电量预测究[J].中国电机工程学报，2003(10):17—21.

[4]基于 ARIMA 模智及回归分析的 K 域用电测方法研究[D]•南京：南京理工大学•2013.

[5]陈明.陈华友•基于 IOWGA 算 f 的最优绀合预测檨智及应用[J].统计与决策.2012(3):88—91.

[6]钟梅.杨桂元，袁宏俊•基于 GIOWA 算子的中国 FDI 预测研究[J].佳木斯大学学报，自然科学版 2015(6):956—960.

[7]陈华友，刘春林•基于 IOWGAW 算子的组合预测方法[J].预测.2003(6）:61_65.

[8]陈华友，盛昭瀚—类基于 IOWGAW 子的组合預测新方法[J].管理工程学报.2005(4>:36—39.