

Table I. Comparison between the bivariate power distribution  $P(l, s)$  of Kulldorff's circular scan statistic and Tango and Takahashi's flexible scan statistic for the hot-spot cluster  $A = \{14, 15, 20\}$ . The nominal  $\alpha$ -level was set to 0.05 and 1000 trials were carried out [11].

Length $l$	Kulldorff					Tango and Takahashi					
	0	Include $s$ hot-spot regions			Total	Length $l$	0	Include $s$ hot-spot regions			Total
1	0	0			0	1	0	0			0
2	1	0	0		1	2	0	0	0		0
3	0	0	0	738	738	3	0	0	0	142	142
4	0	0	0	134	134	4	0	0	0	116	116
5	0	0	0	39	39	5	0	0	0	137	137
6	0	0	0	12	12	6	0	0	0	149	149
7	0	0	0	9	9	7	0	0	0	165	165
8	0	0	0	1	1	8	0	0	0	131	131
9	0	0	2	3	5	9	0	0	0	84	84
10	0	0	0	2	2	10	0	0	0	27	27
11	0	0	0	4	4	11	0	0	0	11	11
12	0	0	0	12	12	12	0	0	0	2	2
13	0	0	0	14	14	13	0	0	0	0	0
14	0	0	0	3	3	14	0	0	0	0	0
15	0	0	0	6	6	15	0	0	0	0	0
Total	1	0	2	977	980	Total	0	0	0	964	964
	Usual power = 0.980					Usual power = 0.964					

useful information for evaluating the CDT performance. An observation of the bivariate power distribution suggests that it can be summarized into a single index that takes a value in the interval  $[0, 1]$ .

We now define the extended power of the CDTs based on the bivariate distribution  $P(l, s)$  by introducing penalties for the FPs and FNs as

$$I(w^-, w^+) = \sum_{l \geq 1} \sum_{s \geq 0} W(l, s; w^-, w^+) P(l, s) \tag{2}$$

where  $W(l, s; w^-, w^+)$  is a weight function such that

$$W(l, s; w^-, w^+) = \begin{cases} \sqrt{(1 - \min\{w^-(s^* - s), 1\}) (1 - \min\{w^+(l - s), 1\})} & (s \leq l; 0 \leq s \leq s^*, 1 \leq l) \\ 0 & (\text{otherwise}) \end{cases} \tag{3}$$

and  $w^-$  and  $w^+$  are the predefined penalties for the FNs and FPs (per region), respectively, with the following constraint:

$$0 \leq w^+ \leq w^- \leq 1 \tag{4}$$

Table II. Comparison between the bivariate power distribution  $P(l, s)$  of Kulldorff's circular scan statistic and Tango and Takahashi's flexible scan statistic for the non-circular hot-spot cluster  $C = \{14, 15, 26, 27\}$ . The nominal  $\alpha$ -level was set to 0.05 and 1000 trials were carried out [11].

Length $l$	Kulldorff						Tango and Takahashi						
	Include $s$ hot-spot regions					Total	Include $s$ hot-spot regions					Total	
	0	1	2	3	4		Length $l$	0	1	2	3		4
1	1	0				1	1	0	0				0
2	0	0	351			351	2	0	0	0			0
3	2	0	4	0		6	3	0	0	0	0		0
4	0	0	3	0	0	3	4	0	0	0	0	138	138
5	2	0	2	0	0	4	5	0	0	0	3	147	150
6	1	0	0	0	0	1	6	1	0	0	2	200	203
7	0	0	0	81	0	81	7	0	1	0	4	147	152
8	0	0	10	18	38	66	8	0	0	2	9	107	118
9	0	0	2	0	26	28	9	0	0	0	10	71	81
10	0	0	0	29	3	32	10	1	0	2	5	28	36
11	0	0	1	13	1	15	11	0	0	0	0	10	10
12	0	0	2	4	60	66	12	0	0	0	0	2	2
13	0	0	0	5	62	67	13	0	0	0	0	0	0
14	0	0	0	10	27	37	14	0	0	0	0	0	0
15	0	0	0	6	37	43	15	0	0	0	0	0	0
Total	6	0	375	166	254	801	Total	2	1	4	33	850	890
	Usual power = 0.801						Usual power = 0.890						

The inequality between  $w^-$  and  $w^+$  is set because it is natural to consider that an FN is more important for a CDT than an FP. It should be noted that  $W(l, s; w^-, w^+)$  implies the geometric mean of '1—penalty for FNs' and '1—penalty for FPs,' hence we can easily obtain  $0 \leq I \leq 1$ . Next we consider the following three special cases:

1.  $w^- = w^+ = 0$ , i.e.  $I(0, 0) = \sum_{l \geq 1} \sum_{s \geq 0} P(l, s)$  equals the usual power.
2.  $w^- = w^+ = 1$ , i.e.  $I(1, 1) = P(s^*, s^*)$  denotes the power to detect the true cluster accurately.
3.  $w^- = 1$ ,  $w^+ = 0$ , i.e.  $I(1, 0) = \sum_{l \geq 1} P(l, s^*)$  denotes the power for which the MLC includes all the regions within the true cluster.

The weight function  $W(l, s; w^-, w^+)$  and the estimated extended power  $I(w^-, w^+)$  for these three special cases are shown in Table III for Kulldorff's and Tango and Takahashi's scan statistics using the simulated results of Table I (circular cluster model A with  $s^* = 3$ ). All three results show that Kulldorff's scan statistic is better than that of Tango and Takahashi's for the circular cluster A. However, the difference varies with the values of  $w^-$  and  $w^+$ .

A natural procedure of determining these two penalties will be derived as follows. First, we consider the penalty for the FNs given by

$$\min\{w^-(s^* - s), 1\}$$

Table III. Weight function  $W(l, s; w^-, w^+)$  and estimated extended power  $I(w^-, w^+)$  for Kulldorff's circular and Tango and Takahashi's flexible scan statistic tests for the circular hot-spot cluster A.

Length $l$	Include $s$ hot-spot regions			Length $l$	Include $s$ hot-spot regions				
	0	1	2		3	0	1	2	3
		$W(l, s; 0, 0)$				$W(l, s; 1, 1)$			
1	1	1			1	0	0		
2	1	1	1		2	0	0	0	
3	1	1	1	1	3	0	0	0	1
4	1	1	1	1	4	0	0	0	0
5	1	1	1	1	5	0	0	0	0
6	1	1	1	1	6	0	0	0	0
7	1	1	1	1	7	0	0	0	0
8	1	1	1	1	8	0	0	0	0
9	1	1	1	1	9	0	0	0	0
10	1	1	1	1	10	0	0	0	0
11	1	1	1	1	11	0	0	0	0
12	1	1	1	1	12	0	0	0	0
13	1	1	1	1	13	0	0	0	0
14	1	1	1	1	14	0	0	0	0
15	1	1	1	1	15	0	0	0	0
Kulldorff's $I(0, 0) = 0.980$				Kulldorff's $I(1, 1) = 0.738$					
Tango and Takahashi's $I(0, 0) = 0.964$				Tango and Takahashi's $I(1, 1) = 0.142$					
		$W(l, s; 1, 0)$				$W(l, s; 1/3, 1/3)$			
1	0	0			1	0	0.577		
2	0	0	0		2	0	0.471	0.816	
3	0	0	0	1	3	0	0.333	0.667	1
4	0	0	0	1	4	0	0	0.471	0.816
5	0	0	0	1	5	0	0	0	0.577
6	0	0	0	1	6	0	0	0	0
7	0	0	0	1	7	0	0	0	0
8	0	0	0	1	8	0	0	0	0
9	0	0	0	1	9	0	0	0	0
10	0	0	0	1	10	0	0	0	0
11	0	0	0	1	11	0	0	0	0
12	0	0	0	1	12	0	0	0	0
13	0	0	0	1	13	0	0	0	0
14	0	0	0	1	14	0	0	0	0
15	0	0	0	1	15	0	0	0	0
Kulldorff's $I(1, 0) = 0.977$				Kulldorff's $I(1/3, 1/3) = 0.870$					
Tango and Takahashi's $I(1, 0) = 0.964$				Tango and Takahashi's $I(1/3, 1/3) = 0.316$					

where  $s = 1, \dots, s^*$ . If  $s = s^*$  the FNs are eliminated and the penalty is set to zero. If  $s = 0$ , the penalty for the FNs attains the maximum value of 1, i.e.

$$w^-(s^* - 0) = 1 \quad \text{or} \quad w^- = 1/s^* \tag{5}$$

On the other hand, the penalty for the FPs is given by

$$\min\{w^+(l - s), 1\}$$

where  $l = s, \dots, m$ . If  $l = s$ , the FPs are eliminated and the penalty is set to zero. We assume that the *maximum allowable number* of regions detected as the FPs is  $l_0 (\geq s^*)$ . Then, when  $l = s + l_0$ , the penalty for the FPs attains the maximum value of 1, i.e.

$$w^+ l_0 = 1 \quad \text{or} \quad w^+ = 1/l_0 \quad (6)$$

Generally,  $l_0$  should be determined according to the ratio of the relative weight of the FP to that of the FN, i.e.  $w^+/w^-$ , which might depend on the cost of misclassifications.

For the previous examples on the circular cluster model A, we consider the following two cases for  $l_0 = s^*$  and  $l_0 = 2s^*$ :

1.  $l_0 = s^* = 3$ :  $I(1/3, 1/3) = 0.870$  (Kulldorff) and 0.316 (Tango and Takahashi). The weight function  $W(l, s; 1/3, 1/3)$  for this case is shown in Table III.
2.  $l_0 = 2s^* = 6$ :  $I(1/3, 1/6) = 0.906$  (Kulldorff) and 0.614 (Tango and Takahashi).

For the non-circular cluster model C shown in Table II, we estimated the following values:

1.  $I(0, 0) = 0.801$  (Kulldorff) and 0.890 (Tango and Takahashi);
2.  $I(1, 1) = 0.000$  (Kulldorff) and 0.138 (Tango and Takahashi);
3.  $I(1, 0) = 0.254$  (Kulldorff) and 0.850 (Tango and Takahashi);
4.  $I(1/4, 1/4) = 0.253$  (Kulldorff) and 0.483 (Tango and Takahashi);
5.  $I(1/4, 1/8) = 0.371$  (Kulldorff) and 0.719 (Tango and Takahashi).

These results indicate that Tango and Takahashi's scan statistic can detect the non-circular clusters such as C more accurately as compared with that of Kulldorff's.

#### 4. PROFILE OF THE EXTENDED POWER

In several cases, it is difficult to set the value of  $l_0$  in advance. For dealing with such situations, we consider a new measure  $I(w^-, w^+)$  to be a function of the ratio  $r = w^+/w^-$  ( $0 \leq r \leq 1$ ) with  $w^- = 1/s^*$ . Then, we define the *profile of the extended power* as

$$Q(r|s^*) = I(1/s^*, r/s^*), \quad (0 \leq r \leq 1) \quad (7)$$

which represents the extended power continuously for all the values of  $r$  where  $l_0 = s^*/r$ . Figure 2 shows the plots of the profile  $Q(r|s^*)$  against  $r$  ( $0 \leq r \leq 1$ ) for the two types of scan statistics applied to the circular cluster model A (Figure 2(a)) and the non-circular cluster model C (Figure 2(b)). The profile  $Q(r|s^*)$ , where  $r = 1$  implies the extended power when  $w^- = w^+$ , i.e.  $l_0 = s^*$ . On the other hand,  $r = 0.5$  implies  $l_0 = 2s^*$ , and  $r = 0$  implies that  $l_0 = \infty$  and the penalties for the FPs is 0. In Figure 2(a), Kulldorff's scan statistic is shown to be uniformly more powerful than that of Tango and Takahashi. The difference in the values of  $I$  is very small when  $r$  is small, but it increases considerably with  $r$ . On the other hand, Figure 2(b) shows that Tango and Takahashi's scan statistic is uniformly more powerful than that of Kulldorff, and the difference in the values of  $I$  appears almost constant irrespective of the values of  $r$ . In certain cases, depending on the shape of the true cluster, there will be situations that two values of  $I$  cross each other at some point  $r'$ . Therefore, the profile of the extended power is expected to play an important role in the simultaneous evaluation and comparison of several CDTs.

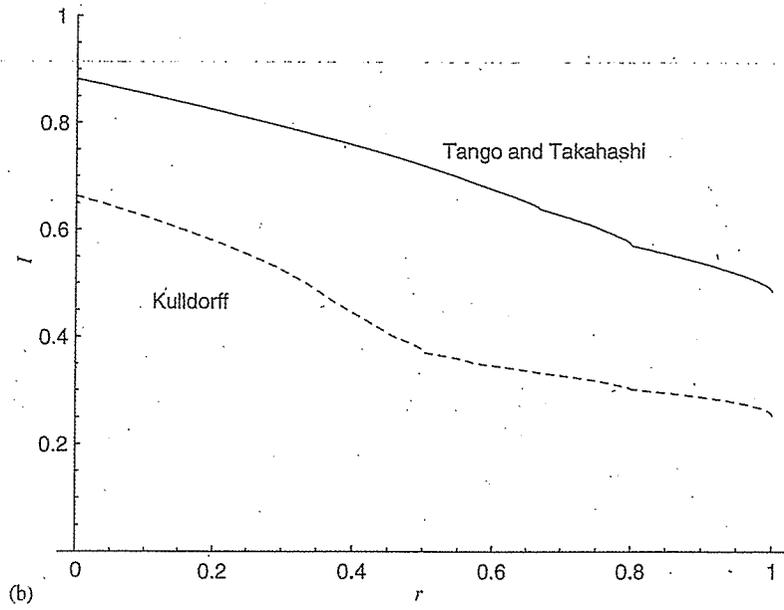
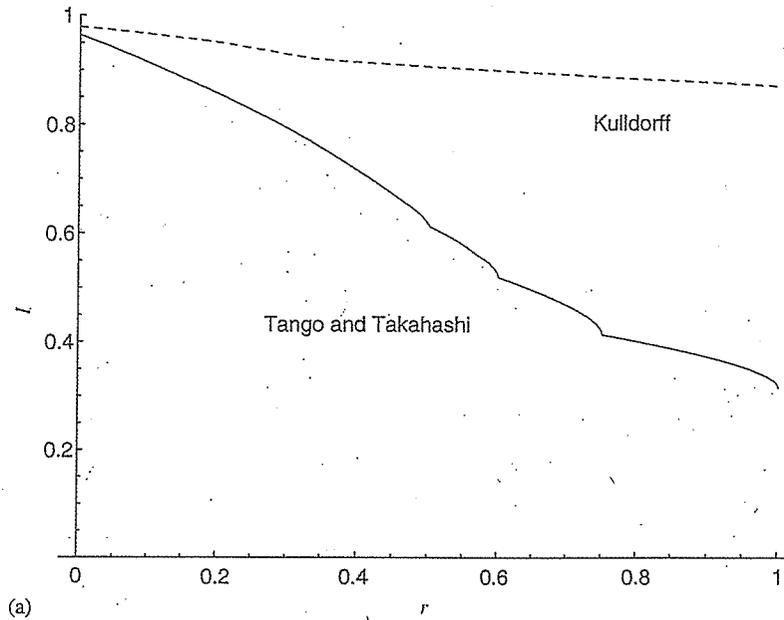


Figure 2. Profile of the extended power  $Q(r|s^*) = I(1/s^*, r/s^*)$  for Kulldorff's circular and Tango and Takahashi's flexible scan statistics: (a)  $A = \{14, 15, 20\}$  ( $s^* = 3$ ); and (b)  $C = \{14, 15, 26, 27\}$  ( $s^* = 4$ ).

## 5. GUIDANCE FOR THE APPLICATION

In this section, we will give some guidance for the application of the proposed procedures in order to investigate the performance of some CDT. At first, we consider a situation where the number of cluster is assumed to be one.

*Step 0:* As an entire study region, use a real region such as Tokyo Metropolitan area or area of upstate New York where real regional populations and number of cases of some disease under study are available.

*Step 1:* Assume a true (practically feasible) cluster model where the location, the length  $s^*$  and its relative risk should be determined. Set  $P(l, s) = 0$ .

*Step 2:* Under the above alternative hypothesis, generate a disease map based upon a random sample of  $n$  total number of cases. If possible,  $n$  should be set equal to the actual total number of cases of disease under study.

*Step 3:* Apply the CDT to the random sample. If we could identify the significant MLC with  $l$  and  $s$ , then set  $P(l, s) \leftarrow P(l, s) + 1$ .

*Step 4:* Repeat steps 2–3 on a large number of random replications (e.g. 1000 times) and estimate the bivariate power distribution (1).

*Step 5:* If we can easily set the value of  $l_0$ , then calculate the extended power (2) where  $w^- = 1/s^*$  and  $w^+ = 1/l_0$ . Otherwise, plot the profile of the extended power (7).

Of course, we can repeat the above procedure for a single different cluster model.

Next, we consider a situation when multiple non-overlapping clusters, say  $k$  kinds, of different size and shape exist in the study area. In this case, the proposed extended power and its profile can be applied in many ways. One simple method will be to estimate a single integrated bivariate power distribution where multiple true clusters are combined into an integrated disconnected cluster area wherein the length of cluster area is defined as the total number of lengths of individual cluster, i.e.  $s^* = s_1^* + s_2^* + \dots + s_k^*$ . To perform this simple method for multiple clusters, steps 1 and 3 of the procedure above for a single cluster model should be modified as follows:

*Step 1:* Assume a cluster model consisting of  $k$  non-overlapping true clusters where the locations, the length  $s_1^*, s_2^*, \dots, s_k^*$  and their relative risks  $\theta_1, \theta_2, \dots, \theta_k$  should be determined. Set  $P(l, s) = 0$ .

*Step 3:* Apply the CDT to the random sample. If we could identify  $m$  significant clusters, i.e. the significant MLC with length  $l_1$ , the significant secondary cluster with length  $l_2, \dots$ , and the  $m$ th cluster with length  $l_m$ , then set  $l \leftarrow l_1 + l_2 + \dots + l_m$  and count  $s$  true regions included in the identified  $l$  regions. Then, set  $P(l, s) \leftarrow P(l, s) + 1$ .

## 6. DISCUSSION

Though the usual power has been widely used for evaluating the CDT, the usual power is not always appropriate since the purpose of CDT is to both reject the null hypothesis and identify the cluster areas accurately. In order to evaluate the CDT, the key problem is two types of misclassifications. One is FN, and the other is FP. Then, Tango and Takahashi [11] proposed a bivariate power distribution classified according to the number of regions detected as the MLC and the number of true hot-spot regions included in the MLC. The bivariate power distribution shows us very useful information for evaluating the CDT performance. However,

an observation of the bivariate power distribution suggests that it can be summarized into a single index.

In this paper, we proposed a new measure termed extended power, which takes a value in the interval  $[0, 1]$ , to evaluate the performance of the CDTs. This extended power is defined as the weighted sum of the bivariate power distribution  $P(l, s)$  wherein the weight is given by the geometric mean of (1—penalty for the FNs) and (1—penalty for the FPs). This measure includes the usual power as a special case.

The extended power could be calculated by using different values of  $w^-$  and  $w^+$ , which means that the researcher could set different values under different situation. Any values satisfying constraint (4) could be used depending on the situations. This expandability might be desirable for some researchers. But for the researchers who are not familiar with this, it may be unhelpful. Then, we proposed the quantities  $w^- = 1/s^*$  and  $w^+ = 1/l_0$  as the values of the penalties for FN and FP, respectively, where  $s^*$  is the length of the assumed true cluster, and  $l_0$  is the maximum allowable number of regions detected as the FPs. Theoretically,  $l_0$  should be determined according to the ratio of the relative weight of the FP to that of the FN. In general, however, we think that  $l_0 = s^* \sim 3s^*$  are reasonable choices.

If we can estimate the average cost per region of the FNs and FPs, we can consider the cost comparison proposed by Tango and Takahashi [11]. Let  $C^-$  and  $C^+$  denote the average costs of the FNs and FPs per region, and the random variables  $L$  and  $S$  denote the length  $l$  of a detected cluster and the number of true regions  $s$  included in the detected cluster, respectively. Then,  $s^* - S$  represents the number of FNs, and  $L - S$  represents the number of FPs. The expected total cost  $C$  is then given by

$$C = C^- \{E(s^* - S) + r_c E(L - S)\}$$

where  $r_c = C^+/C^-$  is the ratio of the average cost of FNs to the average cost of FPs. In this case, we also assume  $r_c \leq 1$ , similar to the case of the ratio  $r$  of the two penalties. Therefore, it might be possible to set  $r = r_c$  for the proposed extended power.

However, in many situations, it might be difficult to set the value of  $l_0$  in advance. Then, we proposed the profile of the extended power  $Q(r|s^*)$ , which can conveniently be used to evaluate and compare different CDTs without fixing the value of the ratio  $r = w^+/w^-$  with  $w^- = 1/s^*$ . The profile could be plotted automatically based upon the bivariate power distribution  $P(l, s)$ . This profile is expected to play an important role in the simultaneous evaluation and comparison of several CDTs. The performance of CDT could be illustrated by the plots of the profile clearly. Furthermore, if we can assume some prior distribution of  $r$ , we can also consider the expected value of extended power as

$$E[Q(r|s^*)] = \int_0^1 r p(r) dr \quad (8)$$

where  $p(r)$  is a probability density function of  $r$ . For example, when  $r$  is distributed according to the uniform distribution  $U(0, 1)$ , an area under the profile in Figure 2 implies the expected power, and using this index, we can compare the performance of Kulldorff's scan statistic and that of Tango and Takahashi.

In the Monte Carlo simulation performed in this study, the number of clusters was assumed to be one. However, the proposed extended power of the CDTs and its profile can be applied for multiple clusters of different shapes. In Section 5, we described a simple method in

which we apply the extended power and its profile to an integrated cluster area as an overall performance measure. However, this simple method ignores the variability of the FP and the FN for individual true cluster and so, the method which can take such variability into account might be better depending on the situation. We would like to leave such a method in our future work.

Finally, it should be noted that the proposed extended power and its profile are solely based upon a Monte Carlo simulation study, not on some theoretical basis due primarily to the fact that it is quite difficult to derive theoretical bivariate power distribution for alternative cluster models.

## REFERENCES

1. Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel JF, Bertolini R (eds). *Disease Mapping and Risk Assessment for Public Health*. Wiley: New York, 1999.
2. Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data*. Wiley: New York, 2004.
3. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Statistics in Medicine* 1995; 14:799–810.
4. Kulldorff M. A spatial scan statistic. *Communications in Statistics* 1997; 26:1481–1496.
5. Kulldorff M. *SaTScan v4.0: Software for the Spatial and Space-time Scan Statistics*. Information Management Services Inc., <http://www.satscan.org/>, 2004.
6. Besag J, Newell J. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society, Series A* 1991; 154:143–155.
7. Turnbull B, Iwano E, Burnett W, Howe H, Clark L. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *American Journal of Epidemiology* 1990; 132(suppl.):136–143.
8. Tango T. A test for spatial disease clustering adjusted for multiple testing. *Statistics in Medicine* 2000; 19:191–204.
9. Duczmal L, Assunção R. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics and Data Analysis* 2004; 45:269–286.
10. Patil GP, Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics* 2004; 11:183–197.
11. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* 2005; 4:11.
12. Moran PAP. Notes on continuous stochastic phenomena. *Biometrika* 1950; 37:17–23.
13. Whitmore AS, Friend N, Brown BW, Holly EA. A test to detect clusters of disease. *Biometrika* 1987; 74:631–635.
14. Oden N. Adjusting Moran's I for population density. *Statistics in Medicine* 1995; 14:17–26.
15. Tango T. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Statistics in Medicine* 1995; 14:2323–2334.
16. Rogerson PA. The detection of clusters using a spatial version of the chi-square goodness-of-fit statistic. *Geographical Analysis* 1999; 31(1):130–147.
17. Bonetti M, Pagano M. The interpoint distance distribution as a descriptor of point patterns, with an application to spatial disease clustering. *Statistics in Medicine* 2005; 24:753–773.
18. Glaz J, Naus J, Wallenstein S. *Scan Statistics*. Springer: Berlin, 2001.
19. Kulldorff M, Tango T, Park PJ. Power comparisons for disease clustering tests. *Computational Statistics and Data Analysis* 2003; 42:665–684.
20. Takahashi K, Yokoyama T, Tango T. *FlexScan: Software for the Flexible Spatial Scan Statistic*. National Institute of Public Health: Japan, 2005.

# 需要密度に対する供給効果を最大化するAEDの最適配置地点

片岡 裕介・浅見 泰司・浅利 靖・郡山 一明

Optimal allocation of AED's by maximization of supply effect on demand density

Yusuke KATAOKA, Yasushi ASAMI, Yasushi ASARI and Kazuaki KOHRIYAMA

GIS－理論と応用

Theory and Applications of GIS

Vol.14 No.2/2006

抜 刷

67

GIS/A 地理情報システム学会

# 需要密度に対する供給効果を最大化する AED の最適配置地点

片岡裕介・浅見泰司・浅利 靖・郡山一明

Optimal allocation of AED's by maximization of supply effect on demand density

Yusuke KATAOKA, Yasushi ASAMI, Yasushi ASARI and Kazuaki KOHRIYAMA

**Abstract:** A probability density function of demand for automated external defibrillators (AED's), which is based on the distribution of demand points by kernel density estimation method, is estimated from the data on occurrences of cardiopulmonary arrests. The probability that someone can be saved by AED's is formulated using probability of survival to hospital discharge. "Supply effect" is expressed as demand density multiplied by this probability of saving, and the locations of AED's are optimized by maximizing supply effect in the entire region. The results show that optimal locations tend to cover the center of the city and that the supply effect is approximately proportional to the number of AED's.

**Keywords:** AED (automated external defibrillator), 密度分布 (density distribution), 最大被覆 (maximal covering), 空間パターン (spatial pattern), 生存退院率 (probability of survival to hospital discharge)

## 1. はじめに

AED (自動体外式除細動器) は、突発的な心停止状態への救急措置に有効とされる医療機器である。取り扱いが簡便で、既に一般市民による使用が認められている状況のなか、全国の自治体には設置に向けた早急な対処が求められている。現状の AED の設置場所としてその多くは駅、空港やイベント会場などの大規模な集客施設、もしくは医療機関が主となる傾向がみられる。人命に直接的に関わる事態である AED の設置に際しては、地域の社会的そして経済的な事情も加味されなければならない。そこで殊に強調されるべきは、より多くの装置を用意する努力もさることながら、装置の配置地点に関する慎重な検討であろう。

AED 設置に関する最近の研究をみると、Malcom et al. (2004) は、地域の人口密度と心停止が発生した場所との関連についての調査を行った。その結果として、人口密度が高い場所であるほど、病院外で心停止が発生する割合が減少する一方で、自宅外で発生する割合については増加する傾向があったことが報じられている。また、Crocco et al. (2004) では、集客施設で必要とされる、AED の設置数に関する数理モデルの提案がなされた。AED の効果が注目されつつある現在の社会的状況を鑑みても、その設置地点に関する検討が今後求められていくであろうし、この要求に応える最も直接的で有効な方法は、理論的に得られた設置地点が示されることであろう。ならびに GIS の分野としても、危機情報の管理および解析等の面から、より一層の貢献が望まれるものと考ええる。

そこで本稿では、AED の最適配置地点に関する考察をおこなうにあたり、主に以下の点を目的とし

片岡：〒 277-8568 千葉県柏市柏の葉 5-1-5  
東京大学空間情報科学研究センター  
Center for Spatial Information Science, University of Tokyo  
E-mail ukataoka@csis.u-tokyo.ac.jp

ている。まず、心停止発生地点（需要点）の空間パターンにもとづいた地域全体の装置の需要量を表す密度分布に対して、装置との位置関係で決定される救命確率をもって各配置地点（供給点）周辺で重み付けした「供給効果」を最大化する配置問題を提案する。さらに、実際の心停止発生地点をデータとして用いた場合の最適配置地点を示すとともに、地域の救命確率および「供給効果」の地理的分布についても表すことで、本稿で提案される方法の有効性を実証する。

## 2. 需要量の空間パターンと供給効果の定式化

本節では、まず地域における需要量の空間パターンを把握する。さらに、装置の設置に伴って装置周辺に生じると考えられる救命確率を定義し、需要量に救命確率を加味した「供給効果」を定式化する。

### 2.1. 需要点にもとづく密度分布の推定

通常、需要量を把握する際には、対象となるものの特性を考慮し、それに応じて適切な方法が用いられる。ここでは、実際に装置が必要とされた地点、つまり過去に心停止が発生した地点をAEDの「需要点（各点で均質）」とし、需要点の分布パターンにより地域の需要量が決定されるものと考えた。とはいえ、心停止が発生した地点のみに需要があるとするのは、同一地点で再発する可能性を考えても妥当性を欠く。よって、過去に数多く発生した地域を潜在的に需要が高い地域であると見なすこととする。従って各需要点周辺の需要量は、需要点で最大であり、周囲に拡がるにつれて減衰する確率密度分布で表現することができよう。ただし、ごく短期間のうちでは、心停止の発生が偶然であることを免れない。そのため需要点数については、需要を代表するのに十分な数である必要がある<sup>1)</sup>。

需要量の地域全体の確率密度分布についてはカーネル法を用いて推定する (Silverman, 1986)。カーネル法とは、点分布を形成する各点の中心に、カーネルと呼ばれる密度分布の山を置き、それら全てのカーネルの合計をもって全体の確率密度関数を求める密度推定法である (図1参照)。心停止発生地点は、

自宅、医療・福祉施設等といった一般に日常生活の多くを過ごす場所での発生、および外出先での発生からなると考えられる。つまり、地域の人口分布のみならず他の要因をも加味した、AEDの潜在的な需要量を表す密度分布がカーネル法で推定される。

なお、本稿ではカーネル関数のパラメータとなるバンド幅選択に際して、Sheather and Jones (1991)の方程式解によるプラグイン法を採用する<sup>2)</sup>。

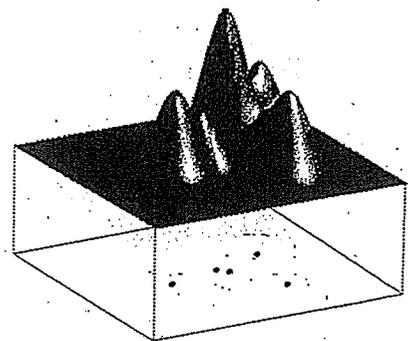


図1 点分布と密度分布

カーネル法によると、平面上の点 $(x, y)$ における確率密度関数 $\hat{f}(x, y)$ は、点分布の各点 $i$ の座標 $(x_i, y_i)$ を用いて以下のように表せる (Bowman and Azzellini, 1997)。

$$\hat{f}(x, y) = \frac{1}{n} \sum_i K\left(\frac{x-x_i}{h_x}\right) K\left(\frac{y-y_i}{h_y}\right) \quad (1)$$

$(K(\cdot))$ :カーネル関数,  $n$ :点の総数,  
 $h_x, h_y$ :各座標値に関するバンド幅

また式(1)で、カーネル関数として正規分布を用いて、各需要点の座標が $(x_{1,i}, y_{1,i})$ とすると、点分布にもとづく地域全体の確率密度関数は、以下の式(2)のようにできる。

$$\hat{f}(x, y) = \frac{1}{2\pi h_x h_y n_1} \sum_{i=1}^{n_1} \exp\left(-\frac{(x-x_{1,i})^2}{2h_x^2} - \frac{(y-y_{1,i})^2}{2h_y^2}\right) \quad (2)$$

$(n_1$ :需要点数)

式(2)は、需要点から推定された確率密度関数であると同時に、需要量の地理的分布を表す空間パターンとして捉えられる。

## 2.2. 供給効果の定式化

次に、設置される装置周辺の需要量に対して、装置による効果の程度を表す「供給効果」について考える。何よりも装置が有効に使用されるためには、心停止発生時点から限られた時間内に処置が施される必要がある。その際、各施設立地点から一定距離内にある領域に含まれる需要量を最大化する配置問題として知られる、最大被覆問題 (Maximal Covering Location Problem) があげられる。最大被覆問題は、Church and ReVelle (1974) により提案された後、救急施設などを対象とした離散平面における配置問題として発展するなかで、様々な解法がこれまでに提案されてきた (Brotcorne et al., 2003)。ただ先述のとおり、地域の需要量は点分布であるよりも、各地点の潜在的な需要を表す確率密度分布で表現される方が適切であろう。さらに AED の場合では、心停止に始まり処置に至るまでの時間に、救命確率が大きく依存するとされるため (American Heart Association, 2000)、被覆される領域内においても場所によっては効果の程度は著しく異なると考えられる。同様に Alsalloum et al. (2006) は、救急車の配置モデルにおいて、通常用いられる一様に被覆される領域を、定めた制限時間内に被覆される確率に置き換えた方法を提案しているが、AED の場合においても類似した状況にあると考えてよい。

そこで、装置から距離  $d$  の地点における救命確率は、経過時間に依存する救命確率を、距離的な問題に置き換えた以下の関数型で近似できるとする<sup>3)</sup>。

$$g(d) = \exp(-ad^2) \quad (3)$$

( $a$ : パラメータ)

本稿では、心停止が発生した現場に居合わせた「救助者」が装置を取りに行くことを前提としている。救助者の置かれた状況を考えると、設置場所がわかっている装置については、装置に向かって取りに行くことが可能である。ところが、当然ながら救助者は装置の全設置点を把握しているわけではない。事前知識も不確実であることに加え、このように一刻を争う事態では行動が直感に大きく左右され、状況として通常の施設選択の場合とは異なるともい

える。つまり、救助者は必ずしも現場から最も近くにある装置を常に把握しているわけではなく、ある地点において装置が心停止にもたらす効果は、心停止発生地点から最も近くにある装置との関係性のみに影響するとは限らない<sup>4)</sup>。すると装置の「認知度」の拡がり、装置からの距離に伴う救命確率の減衰に代替されているとしたときに、近くに設置されている装置が多いほど、設置に伴う効果は上昇するものとも考えられる。以上より、ある地点における AED 設置によって与えられる影響の程度は、各装置の救命確率の和として説明できるものと仮定する。

よって式(3)を用いると、地点  $(x, y)$  での救命確率の和は、各装置の設置地点の座標  $(x_{2,j}, y_{2,j})$  を用いて以下のようにできる。

$$g(x, y) = \sum_{j=1}^{n_2} \exp\left(-a\left((x-x_{2,j})^2 + (y-y_{2,j})^2\right)\right) \quad (4)$$

( $n_2$ : 装置の設置数)

以上をふまえ、装置の効果が期待される装置周辺において、救命確率を需要量に加味することで得られる、需要と装置の効果に依存する量を本稿では「供給効果」と定義する。なお、理解を容易にするために一次元で考えると、需要量と「供給効果」の関係は以下の図2のようになる。

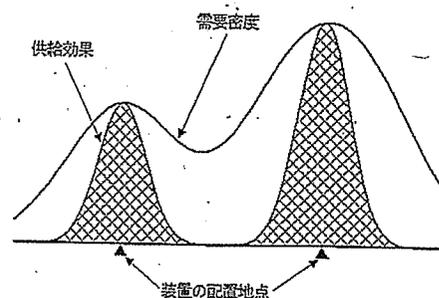


図2 需要密度に対する供給効果

これより、ある地点における需要に対して救命可能な量となる「供給効果」は、その地点の需要量に救命確率の和を乗じたもので表されることから、「供給効果」の地域全体の総和 (図2の網掛け部分) は、以下の式(5)のように定式化される。

$$\begin{aligned}
L_1(x, y) &= \iint \hat{f}(x, y) \cdot g(x, y) dx dy \\
&= \iint \left\{ \frac{1}{2\pi h_x h_y n_i} \sum_{i=1}^n \exp\left(-\frac{(x-x_{1,i})^2}{2h_x^2} - \frac{(y-y_{1,i})^2}{2h_y^2}\right) \right. \\
&\quad \left. \times \sum_{j=1}^{n_2} \exp\left(-a((x-x_{2,j})^2 + (y-y_{2,j})^2)\right) \right\} dx dy \\
&= \frac{1}{4n_1 \sqrt{1+2ah_x^2} \sqrt{1+2ah_y^2}} \\
&\quad \times \sum_{i=1}^n \sum_{j=1}^{n_2} \exp\left(-\frac{a(x_{1,i}-x_{2,j})^2}{1+2ah_x^2} - \frac{a(y_{1,i}-y_{2,j})^2}{1+2ah_y^2}\right) \\
&\quad \times \operatorname{erf}\left(\frac{x-x_{1,i}+2ah_x^2(x-x_{2,j})}{h_x \sqrt{2+4ah_x^2}}\right) \\
&\quad \times \operatorname{erf}\left(\frac{y-y_{1,i}+2ah_y^2(y-y_{2,j})}{h_y \sqrt{2+4ah_y^2}}\right)
\end{aligned} \tag{5}$$

式(5)において、各項における積分定数については省略した。また、 $\operatorname{erf}(\cdot)$ は誤差関数をあらわし、以下のように定義される。

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \tag{6}$$

### 3. 供給効果を最大化する配置問題

本節では、2節で定式化された「供給効果」を最大化する、AEDの最適配置問題を提案する。

まず、式(5)で積分区間を無限領域としてよいことから、以下の式(7)を得る。

$$\begin{aligned}
L_2(x_{2,1}, x_{2,2}, \dots, x_{2,n_2}, y_{2,1}, y_{2,2}, \dots, y_{2,n_2}) \\
&= \frac{1}{n_1 \sqrt{1+2ah_x^2} \sqrt{1+2ah_y^2}} \\
&\quad \times \sum_{i=1}^n \sum_{j=1}^{n_2} \exp\left(-\frac{a(x_{1,i}-x_{2,j})^2}{1+2ah_x^2} - \frac{a(y_{1,i}-y_{2,j})^2}{1+2ah_y^2}\right)
\end{aligned} \tag{7}$$

式(7)で定式化された「供給効果」を、このように設置場所の候補地が全くない状況で、単純に最大化しようとするならば、需要量の多い領域を幾度も被覆してしまう可能性がある。それは、同一地点に装置を複数台設置するという結果を得ることを意味する。ところが、ある装置のごく近辺で、複数の心停止患者がほぼ同時に発生する状況は想定しづらく、いわば「容量の制限がない」と条件付けられるAEDの問題に限っては、同一地点に解が重複して

得られる状態は好ましくない。そのため、ここでは装置間の距離についての制約を設けることにより、救命確率の和が1を超える地点が存在しないものとする<sup>5)</sup>。つまり、装置の効果が期待できる各装置の設置地点からの限界の距離を $d_{\min}$ とおき、ある装置から $d_{\min}$ の距離内には他の装置が設置されることがないものとする。よって、各装置間の距離を $d_{s,t}$ で表し、制約条件として式(9)とおくことで、最適化されるとき各配置地点の座標値は、式(7)が最大化されるとき解集合として与えられる。つまり、本稿で定式化された配置問題は、以下の多変数最大化問題に帰着する。

$$\text{Maximize } L_2(x_{2,1}, x_{2,2}, \dots, x_{2,n_2}, y_{2,1}, y_{2,2}, \dots, y_{2,n_2}) \tag{8}$$

subject to

$$d_{s,t} = \sqrt{(x_{2,s} - x_{2,t})^2 + (y_{2,s} - y_{2,t})^2} \geq d_{\min} \quad \forall s, t \tag{9}$$

### 4. 適用事例

前節で提案されたAEDの配置問題の適用事例を扱う。需要点として弘前市の心停止発生地点のデータを用いた場合の、AEDの最適配置地点を示す。

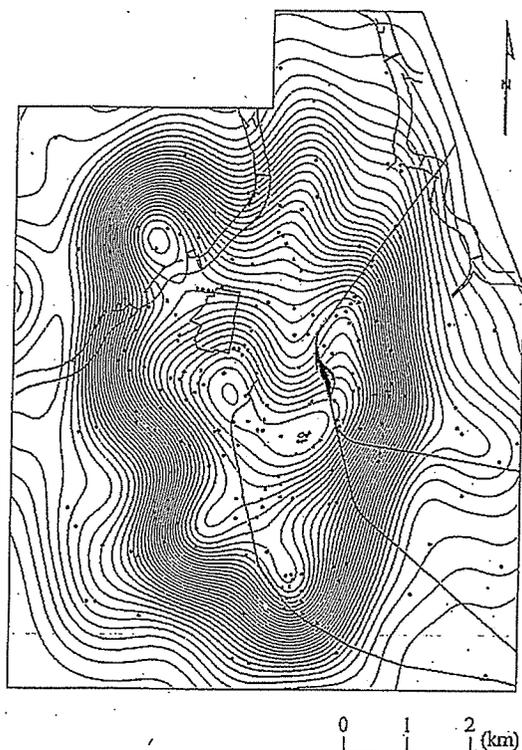
#### 4.1. 心停止発生地点とその密度分布

需要点となる心停止発生地点としては、弘前地区消防事務組合が過去3年間に扱った内因性心肺停止で、原因疾患が心疾患と推定された411例を分析に用いた。

始めに、当データ内の属性として含まれている住所表記から位置情報を取得した<sup>6)</sup>。各需要点の位置と、式(2)より得られる各地点の確率密度を表す等値線(間隔は $1.0 \times 10^{-9}$ )を図3に示す。なお、カーネル関数のバンド幅選択は、2.1.で述べたように、Sheather and Jones (1991)によるプラグイン法を採用した。

図3の需要点にもとづく密度分布について見てみると、弘前駅や弘前城の南側に隣接する市役所から少し離れた地域、および土手町と呼ばれるメインストリート付近を中心として、心停止発生密度が非常に高いとされる地域が拡がっていることが確認された。また、市街地部周縁である地域帯で密度分布が

大きく変化しており、概して人口分布の変化に対応しているものと理解される。



(図中の表記については以下の通り。●：需要点、および需要量密度を示す等値線。背景地図として、河川、鉄道、弘前城を表記。)

図3 心停止発生地点とその密度分布

#### 4.2. 処置までの時間にもとづく救命確率式

本稿では、式(3)で定式化された装置からの距離にもとづく救命確率式を求めるために、処置までの時間に依存する「生存退院率」を用いることとする。「生存退院率」とは処置による蘇生の成功率を意味し、倒れてから除細動の開始までに要する時間が1分経過するごとに約10%低下するとされることから<sup>7)</sup>、AEDのような問題においては国内外を問わず重要な知見とされている。これにもとづき、10分経過した時点で救命確率がほぼ0になるように式(3)のパラメータを求めた ( $a=4.32 \times 10^{-5}$ )。なお、実際のパラメータの決定においては、本稿で設定された仮定にもとづき、分速80mで装置に向かっ

て直進するとしたうえで、往復にかかる時間を考え5分経過した400mの時点で救命確率が  $1.0 \times 10^{-3}$  となるとした<sup>8)</sup>。

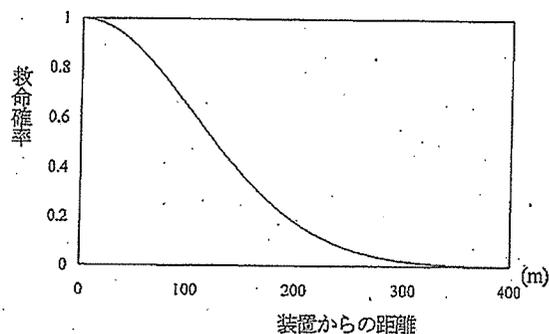


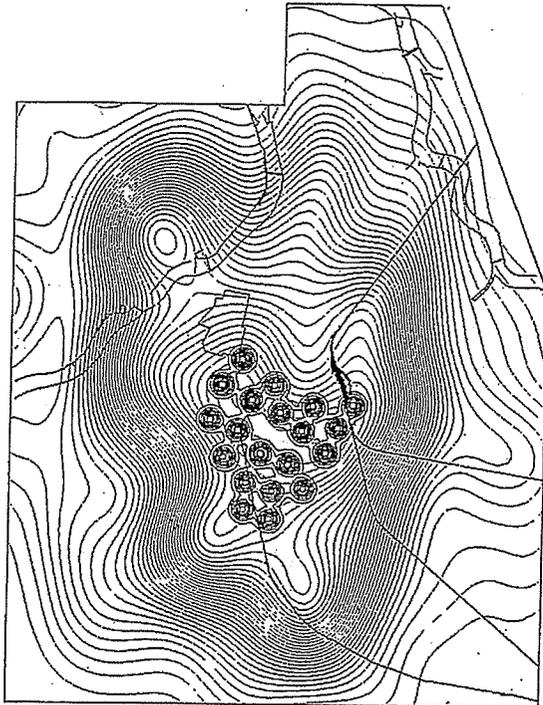
図4 各装置からの救命確率

#### 4.3. AEDの最適配置地点

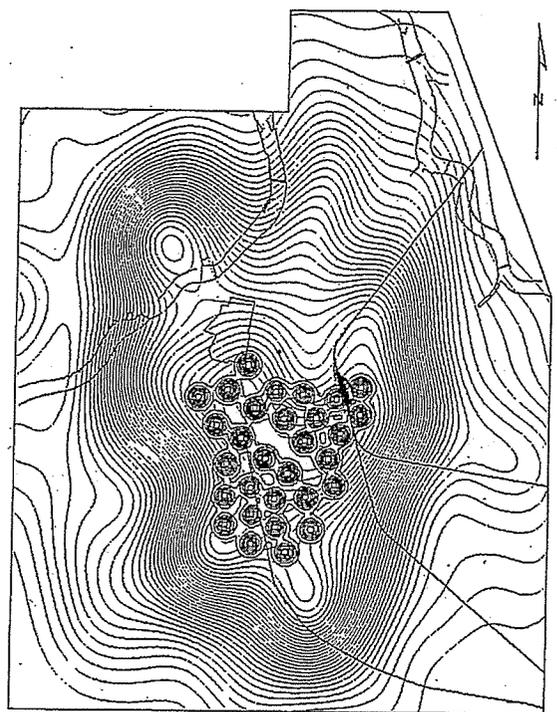
次に式(8)を用いて、AEDの設置数を1点から30点までの間で所与とし最適地を求める。前項での考察をふまえて、式(9)において  $d_{\min}=400\text{m}$  とし、式(8)を多変数の制約条件付き最大化問題として数値計算をおこなった結果<sup>9)</sup>、分析の対象とした設置数の全ての場合で各最適地が対象地域内に得られた。そのなかで、設置数が「20点」、「30点」のときの各最適地、および任意の地点における「供給効果」の密度分布を示す等値線(間隔は  $5.0 \times 10^{-8}$ )を図5に、また、装置による救命確率の密度分布を示す等値線(間隔は  $2.0 \times 10^{-1}$ )を図6に示す。参考のために、図3の需要点の密度分布についても各図に併せて示した。

得られた最適配置地点の分布の特徴として、4.1.で装置の需要密度が高いとされる地域において、設置数が増加するに伴い、最適地が密度分布の「山」の頂点付近からその周辺に向かって、満遍なくカバーするように得られるという傾向が表れている。

また、図6の救命確率の密度分布は、需要分布に依らない各地点での装置の効果を示しており、各等値線が救命確率で表された装置までの一定の到達距離の分布を意味している。従って、一刻を争う状況下では、視覚化された判断基準として特に有効な手がかりとなることも期待できるであろう。

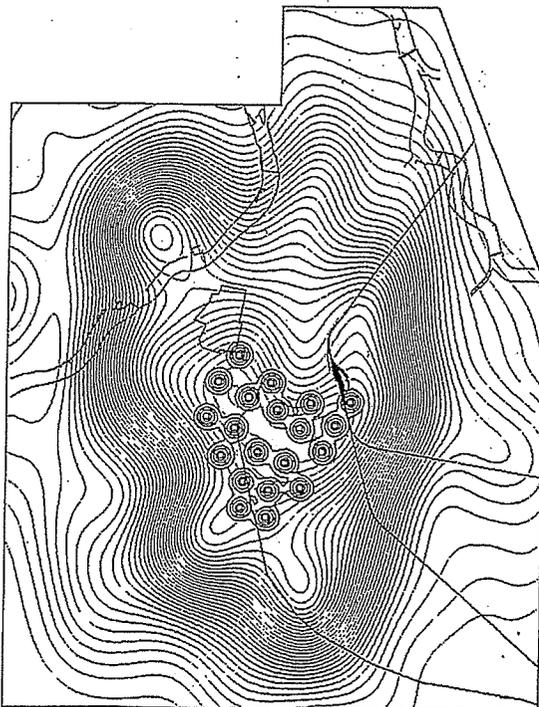


20 点配置

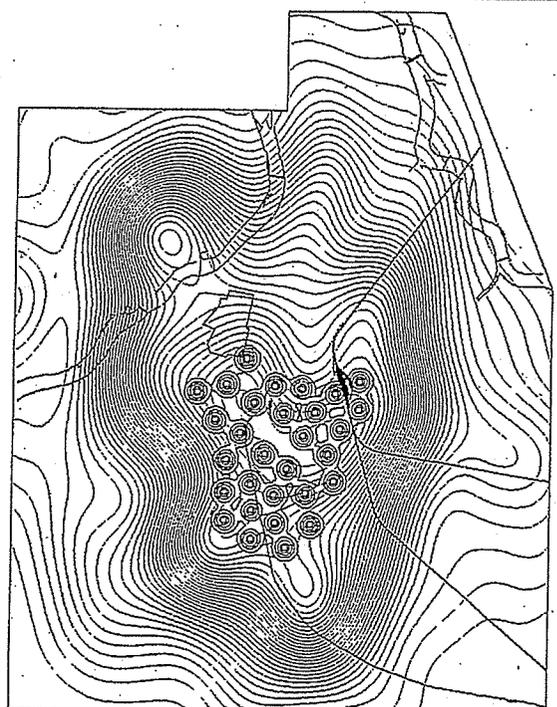


30 点配置

図5 AEDの最適配置地点と供給効果の密度分布



20 点配置



30 点配置

図6 AEDの最適配置地点と救命確率の密度分布

0 1 2 (km)

(図5,6中の表記については以下の通り。□: AEDの最適配置地点, および各々の密度(濃線), 需要量密度(淡線)を示す等値線。背景地図として, 河川, 鉄道, 弘前城を表記。)

なお、図5と図6における密度分布が非常によく似た様相を呈している理由は、最適地が得られる範囲内で需要量に大きな変化がないことによる。

#### 4.4. 設置数に伴う効果

図7に、各設置数と式(8)の目的関数値との関係を示す。例えば、「供給効果」を表す目的関数値が0.01となるときでは、最適地に設置された装置によって地域全体の需要量の1%がカバーされることを示している。

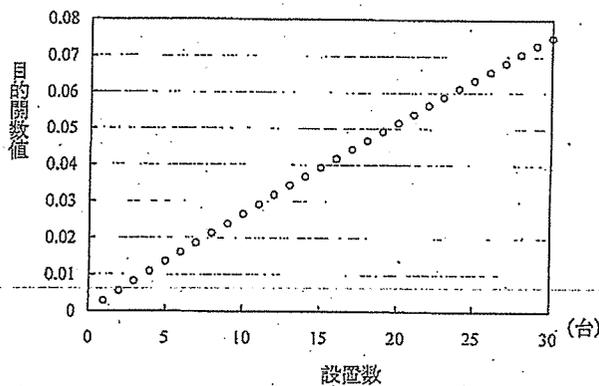


図7 設置数と目的関数値

図7を見る限りでは、設置数の増加に従って「供給効果」はほぼ線型に増加することから、少なくとも設置数が30点までにおいては、需要がほぼ同等に高い地域に装置が割り当てられていると考えられる。「供給効果」の変化の割合は、結局のところ需要量の分布状況に大きく依存するものであるが、一般的には、設置数が少ない初期の段階では急激に増加し、その後緩やかに増加を続けることが予測できる。言い換えれば、設置数が増加するにつれ、需要密度の高い地域から需要密度の低い地域へと、装置で新たにカバーする領域が移行して行くといえる。

さらにここで得た結果は、より多くのAEDを地域全体として有することで、地域医療に貢献できることを意味するとともに、供給効果の上昇の程度が鈍らないこの範囲の設置数においては、AEDの配置の検討が特に重要とされるべきであることを示唆している。

従来の限られた施設以外にも設置が検討される状

況にあるなか、適切な配置地点についての議論は一層重要性を増すことが予想される。現実には、すべての最適地にAEDを配置することは非常に困難ではあるが、問題の重要性からも現状の配置状況の適切性は常に問われるものである。そのためには、何よりも理論的な最適配置地点、およびその効果については厳密に示されている必要がある。

#### 5. おわりに

本稿では、AEDの配置地点に対する検討が必要であるとの認識にたち、実際の心停止発生地点からなる需要点より推定された確率密度分布に、生存退院率にもとづく救命確率を加味したAEDの最適配置問題を提案した。なかでも、地域全体の需要量に対する「供給効果」を定式化するとともに、配置地点の座標値に関する多変数最適化問題としてその解を数値的に得た。AEDの最適配置問題では、装置がカバーできる範囲が限られているとともに、装置までの距離が非常に重大な要因である。本稿では、対象地域内で密度分布が比較的高い場所が広範囲に広がっていたために、装置の最適配置地点は特定の地域に集中して得られる傾向が顕著であった。しかし、心停止発生地点に影響を及ぼすとされる、高齢者の人口分布や、人的そして物的な集積度に大きな偏りがある地域では、地域全体としての需要量の地理的な変化が激しくなる。その際には、求まる最適地点もある程度分散するといった、今回とは異なる結果を得ることも予想される。要は、AEDの最適配置地点は直感的に導き出すことが困難であるだけに、十分な検討が望まれる問題であると考えてよいであろう。

また、AEDの場合に、設置に伴う用地取得や周辺環境への配慮などの問題は、通常の施設立地の場合ほど重大ではないと考えられる。装置に関する維持管理などの問題はあっても、明らかに立地不可能であるような条件を除けば、この点で大きく性質が異なることが理解される。つまり、現段階において候補地として装置の設置地点を事前に限定するのではなく、連続的な平面上に得ようとする本稿の試みは、現状に対しても有意義で実効性の高い問題と

位置付けて良いと言える。

最後に、今後の課題について述べる。AEDの設置が希望される施設として、「駅」、「学校」、「公共施設」などの、現段階で積極的に設置が進んでいる施設を挙げる声が一般には大きいものの、「コンビニエンスストア」あるいは「交番」への設置を希望する意見も少なくない<sup>10)</sup>。「コンビニエンスストア」や「交番」は、地域住民に比較的認知されている施設であり、例えば深夜のような時間を問わない緊急時への対応の面でも期待できる。また、市街地のみならず居住地域も含めて幅広く分布していることから、有効なAEDの配置地点として、需要量との関係性を含めた検討をおこなうことも重要であると考えられる。これらをふまえて今後は、既存施設へのAEDの設置を視野に入れた配置問題についても考察していきたい。

#### 謝辞

本研究をおこなうにあたり、国立保健医療科学院技術評価部丹後俊郎部長、東京大学空間情報科学研究センター丸山祐造助教授および東京大学大学院工学系研究科都市工学専攻貞広幸雄助教授より貴重なコメントを頂いた。また、弘前地区消防事務組合には救急車による搬送データの提供をはじめ、多大なご協力を頂いた。

本研究は、厚生労働科学研究費補助金（健康科学総合研究事業）「地理及び社会状況を加味した地域分析方法の開発に関する研究」（代表：浅見泰司）の支援を受けた。ならびに、東京大学空間情報科学研究センターの研究用空間データ利用を伴う共同研究（研究番号75）として、以下のデータを利用した。（株）ゼンリン提供：ZmapTownII 青森県弘前市記して謝意を表す。

#### 補注

1) 本稿の適用事例で扱っている心停止発生数に関しては、季節などの時期による発生数の変動の可能性や、設置するAEDの数との関係を吟味し、過去3年分の411例を本稿で分析を行うのに十分なサンプル数とした。実際の密度分布を

見る限りでも、市街地のなかで人口が多い居住地域ほど密度が高くなる傾向を示すことなどから、需要量をよく代表しているものと判断した。

- 2) Sheather and Jones のプラグイン法では、式(1)の $\hat{f}$ の推定値が、理論的な考察によるバンド幅からではなく、データから推定されるという点に特徴がある。Simonoff (1996) では、一般に最良なバンド幅選択法が発見されていないことを踏まえつつも、プラグイン法が「最先端」の方法であるとともに、多方面で有効な選択法であると紹介されている。
- 3) 処置に至るまでの時間の経過と後出の生存退院率との関係に加えて、道路状況などの地理的な条件を考慮した結果、距離の増大に伴う効果の減衰を示す関数式として適当であると判断した。
- 4) 例えば、本稿で想定されるAEDの場合とは異なり、救急車の出動が前提となるような事例では、発生地点から最も近くにある立地点が選択されるのが妥当であろう。
- 5) 計算上、被覆される領域外でも1を超える地点は存在するが、本稿でも述べるように影響を無視できるほど微少であることから、装置から一定距離以上の地点での効果は無視できることが前提とされている。
- 6) 位置情報の取得に際しては、東京大学空間情報科学研究センターが提供するアドレスマッチングサービスを用いた。
- 7) American Heart Association (2000) によれば、心停止発生から1分経過するにつれ、生存退院率が約7%~10%で減少するとされている。これに基づいて本稿では、10分経過した時点で生存退院率がほぼ0になるとした。
- 8) AEDの効果が期待できる範囲がごく限られていることに加え、対象地域では道路密度の粗密が分析に大きな影響を与えないと考えられることから、本稿では直線距離を採用した。
- 9) 最適化計算では、MATLABのOptimization Toolboxを使用した。

- 10) AED (自動体外式除細動器) 普及を推進する会  
による「AED の設置場所に関するアンケート」  
より < [http://www.aedjapan.com/vote/qnaire2.  
cgi](http://www.aedjapan.com/vote/qnaire2.cgi) >.

#### 参考文献

- American Heart Association (2000) Guidelines 2000 for cardiopulmonary resuscitation and emergency cardiovascular care: International consensus on science, *Circulation*.
- Alsalloum, O.I. and Rand, G.K. (2006) Extensions to emergency vehicle location models. *Computers & Operations Research*, 33, 2725-2743.
- Bowman, A. and Azzelini, A. (1997) *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press.
- Brotcorne, L., Gilbert Laporte, G. and Semet, F. (2003) Ambulance location and relocation models. *European Journal of Operational Research*, 147, 451-463.
- Church R.L. and ReVelle C.S. (1974) The maximal covering location problem. *Papers of the Regional Science Association*, 32, 101-18.
- Crocco, T.J., Sayre, M.R., Liu, T., Davis, S.M., Cannon, C. and Potluri, J. (2004) Mathematical determination of external defibrillators needed at mass gatherings. *Prehospital Emergency Care*, 8, 292-297.
- Malcom, G.E., Thompson, T.M. and Coule, P.L. (2004) The location and incidence of out-of-hospital cardiac Arrest in Georgia. *Prehospital Emergency Care*, 8, 10-14.
- Sheather, S.J. and Jones, M.C. (1991) A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society series B*, 53, 683-690.
- Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- Simonoff, J.S. (1996) *Smoothing Methods in Statistics*. New York: Springer-Verlag.

# 空間ドキュメント管理システムの設計と実装

白石 陽   有川 正俊   相良 毅   浅見 泰司

2007

『DEWS』B7-10

# 空間ドキュメント管理システムの設計と実装

白石 陽<sup>†</sup> 有川 正俊<sup>†</sup> 相良 毅<sup>‡</sup> 浅見 泰司<sup>†</sup>

<sup>†</sup> 東京大学 空間情報科学研究センター 〒277-8568 千葉県柏市柏の葉 5-1-5

<sup>‡</sup> 東京大学 生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

E-mail: <sup>†</sup> {siraisi, arikawa, asami}@csis.u-tokyo.ac.jp, <sup>‡</sup> sagara@fkl.iis.u-tokyo.ac.jp

あらまし 本稿では、住所情報を含むデジタルドキュメントを、直感的なインタフェースで扱うことのできるシステムとして、空間ドキュメント管理システム (SDMS: Spatial Document Management System) を提案する。SDMS は、ドラッグ&ドロップという簡単な操作で、即座に、ドキュメント中の住所情報を、点として地図上に表示することができる。SDMS では、非定型の一般的なデジタルドキュメントから、住所情報を抽出し、アドレスマッチング技術を利用して POI (Point of Interest) を生成し、地図上に表示する。SDMS を用いることにより、多くの一般的なユーザが、簡単に、デジタルドキュメントを空間的に整理し、地図化することが可能となる。

キーワード 空間ドキュメント, 位置情報, GIS, ジオコーディング, ジオパース, Point of Interest (POI)

## 1. はじめに

近年の位置情報関連技術の進歩は目覚ましい。位置測位技術の発達、GPS 搭載携帯電話の普及、位置情報サービスの高度化、地図 API の公開、Web コンテンツへの緯度経度情報の埋め込みなど、位置情報を扱う様々な枠組みが整ってきている[1,2,3,4]。緯度経度という絶対的な位置情報は、こうしたサービスを支える基盤となる情報であるが、多くの人にとっては、ただの数字の列にしか見えないため、必ずしも人間にやさしい表現とは言えない。それに対して、住所や地名などの位置情報は、そのままでは地図に表示することはできない間接的な表現であるが、人間の理解しやすいものである。日常的に利用される自然言語による表現は、コミュニケーションにおいても効果的であり、案内文書、報告書、店舗一覧、施設一覧、光化学スモッグ発生情報など多くのデジタルドキュメント上で利用されている。今後、緯度経度のような“直接位置参照情報”が目に見えない形で埋め込まれたドキュメントが増えていくことは容易に想像できるが、依然として、住所や地名などの“間接位置参照情報”を含む様々なドキュメントが、多くの場面で利用されていくことは間違いない。本研究では、こうした住所や地名などの位置情報を含むデジタルドキュメントを「空間ドキュメント」と呼ぶ。インターネット上の Web ページ、メールの添付ファイル、自分のコンピュータの中のファイルなど様々な空間ドキュメントが存在する。そのファイル形式も、HTML, Word, Excel, TEXT など多様である。

本稿では、一般ユーザが日常的に利用している、このような多種多様な空間ドキュメントを、容易に管理、検索、地図化できるツールとして、空間ドキュメント管理システム (Spatial Document Management System,

以下 SDMS と呼ぶ) を提案する。SDMS は、ドラッグ&ドロップという単純な操作で、住所や地名などを自動的に抽出し、緯度経度を算出して、Point of Interest (POI) として地図上に表示することができる。SDMS を用いることによって、インターネット上やパソコン内に存在する様々なドキュメントを空間的に整理し、把握することができる。また、地域行政、保健医療、公衆衛生、危機管理などの現場において、こうした場所情報を含むドキュメントは、貴重な情報源であるにも関わらず、地図化する簡易な手段がなかったために、死蔵されている場合も多く、空間ドキュメントを扱うことのできる SDMS に対する潜在的なニーズは高いと考えられる。

以下、本稿では、2 章で関連研究を挙げ、提案システムの位置付けについて述べた後、3 章で提案システムの説明を行う。4 章では、動作例を説明した後、考察と今後の課題を述べる。最後に、5 章でまとめを行う。

## 2. 関連研究

### 2.1. 地理情報システムとの比較

住所、地名、郵便番号などの間接位置参照情報を緯度経度などの直接位置参照情報に変換することをジオコーディング (geo-coding) と言う。住所を緯度経度に変換する処理は、特に、アドレスマッチングと呼ばれる。それに対して、ドキュメント中から住所や地名などの地理情報を抽出することをジオパース (geo-parse) 処理と呼ぶ。

空間情報を管理できるシステムとして地理情報システム (GIS: Geographic Information System)[5]がある。GIS は、様々な空間情報の管理、解析、統合、可視化を行うことのできる便利なツールである。ジオコーデ