Chapter 7
Capture, Capacities, and Thresholds

7.1 Introduction

There can be a number of different problem settings under which covering models can be defined and applied. Many of the models that are discussed in this chapter were originally inspired by issues associated with retail and competition. Although one at first blush may think of retail siting and coverage models as having little in common, except for something obvious like a pizza chain attempting to locate facilities so that it can deliver pizzas everywhere in a city within 30 min, there are surprisingly many retail elements that can be defined and addressed using coverage constructs.

Christaller (1933) attempted to address the question of why retail centers were arranged the way they were across a region. He reasoned that such centers needed a sustainable level of customers, called the threshold. Without a retail establishment attracting a threshold amount of customers, a business was not viable. Further, if a retail center had a large number of customers, much larger than the threshold, it was an invitation for other firms to enter that market area and establish their own facilities. Each of these elements can also be modeled using covering concepts, albeit with added components.

One of the first models to explicitly address a retail problem using a covering framework was the list selection problem in direct mail advertising (Dwyer and Evans 1981). Such a problem is characterized by a retail company wanting to rent various magazine subscriber lists in order to launch a mail-order catalogue, often for an upcoming holiday or event. The key issue is to determine which lists to rent in order to cover/reach as many of the potential customers as possible. The mathematical program developed by Dwyer and Evans (1981) has the exact same structure as the maximal covering location problem introduced and formulated in Chap. 2.

Another retail-based covering model is the maximum capture or sphere of influence location problem (ReVelle 1986). This model was inspired by the early work of Hotelling (1929) where two vendors compete for customers. The model

R. L. Church, A. Murray, *Location Covering Models*, Advances in Spatial Science,
https://doi.org/10.1007/978-3-319-99846-6_7

developed by ReVelle (1986) was based upon taking the perspective of one of the competitors in an attempt to capture as many of the other's customers while locating several retail facilities. Although this model is quite simple, it encompasses a key element in retail location, that of serving a region better than a competitor.

Finally, many settings suggest that facilities can handle only so many customers at the same time without being congested or losing service quality. That is, there is a limit or capacity to reasonably handle customers. In many applications, it is assumed that each facility can handle all of the customers it serves, but in other contexts this is simply not true. Consequently, the use of a capacity on each facility may be appropriate. This too, can be addressed in a covering model.

Collectively, the above examples have hopefully demonstrated that there are many possible applications for covering models in retail and service location. In the remainder of this chapter, we provide details on models that are designed to capture customers or an audience, models that maintain viability through the use of thresholds and capacity conditions, and models that carve up a region into franchise areas. In the next section, we discuss the problem of capturing customers in retail service provision.

7.2 Maximum Capture

Hotelling (1929) considered two competitors, each selling water from their own artesian wells along a linear market. In his highly constrained model of a game between the two competitors, he demonstrated that when each of the competitors maximized their own profits, an equilibrium would eventually be reached at which the linear market between the two water vendors would be equally divided with each customer visiting their closest market.[1] ReVelle (1986) envisioned a similar setting, but defined it on a network of nodes and arcs instead of a linear market. It was assumed that some of the nodes were places of potential retail location and other nodes were points of demand. Consider the situation that there are already one or more firms that offer the same product/service with the same price across this network. Presumably, the competing facilities have already divided the market with customers patronizing their closest facility.[2] With respect to this network of

[1]Hotelling (1929) suggested that if the two vendors could change location, they would eventually reach a price equilibrium while locating close together at the center of the linear market. Here they would attract an equal share of the customers. However, d'Aspremont et al. (1979) has since shown that a price equilibrium does not hold when the vendors are sufficiently close together.

[2]As in ReVelle (1986), we assume here that customers see no difference in price or offerings between competitors, so they patronize their closest facility. This may not be the case when making multi-purpose trips, like a person stopping off at a facility on their way to work, or on their way home.

demands, existing competing firms, and customer patronage patterns, ReVelle (1986) posed the following problem:

> If a retail operation plans to invest in one of more new facilities, where should the firm locate those facilities in order to maximize new customer gains?

When any new facility opens, the patronage patterns change. ReVelle (1986) suggested that, even if one lacks a reliable model of shopping behavior, a firm would still need some metric to measure the impact of locating any new facilities. He suggested the "maximum capture or maximum sphere of influence metric". Essentially, a demand node was considered to be captured from a competitor if a new facility was located closer to that demand than any existing competitor's facility. That is, a site which captures a large number of customers from existing retailers would be viewed as providing those customers with a closer, more accessible, facility. ReVelle (1986) noted that if the firm already has some facilities, customers at their existing facilities should not be counted as being captured if any of their newer facilities are closer to these existing customers. That is, the number being captured are customers from other firms because they are now provided with a closer facility. It should be noted that this problem is related to the Condorcet location problem, in which the objective is to locate a number of facilities in such a manner that a majority of people are better off (Hansen and Thisse 1981). Given this criteria, a solution to the maximum capture problem when 50% or more of the people (in terms of closeness) are captured meets the Condorcet property.

Consider the following notation:

I = set of demand nodes that are available for capture
J = set of eligible facility sites
i = index of demand nodes
j = index of facility sites
J_O = set of currently occupied facility sites
J_N = set of sites not currently occupied, but are eligible to have a new facility
d_{ij} = shortest distance or travel time between demand node i and facility site j
S_i = distance from node i to its closest existing facility

$$N_i = \{ j \in J_N \mid d_{ij} < S_i \}$$

$$\tilde{N}_i = \{ j \in J_N \mid d_{ij} = S_i \}$$

p = number of facilities to be located
a_i = demand representing potential customers at node i

$$x_j = \begin{cases} 1, & \text{if a facility is located at node } j \\ 0, & \text{otherwise} \end{cases}$$

$$y_i = \begin{cases} 1, & \text{if demand } i \text{ is now served by a new facility that is closer than } S_i \\ 0, & \text{otherwise} \end{cases}$$

$$z_i = \begin{cases} 1, & \text{if demand } i \text{ is now served by a new facility exactly at the distance } S_i \\ 0, & \text{otherwise} \end{cases}$$

The coverage sets N_i and $\tilde{N}_i$ are worth a little more discussion given their significance here. N_i is the set of available facility sites that are strictly closer to demand i than its current closest facility. Differing from this, $\tilde{N}_i$ is the set of sites j (occupied or unoccupied) that are equally distant to the closest current facility to demand i. Building on this discussion, two situations arise for capturing new customers in a region. Siting a new facility represents a case where it is now closer to demand than any existing facility of a competitor. The variables y_i track this situation. The second case occurs when a new facility is equally close to demand as an existing facility of a competitor. The variables z_i track this situation. We can now define the maximum capture location problem (MAXCAP) as:

$$\text{MAXCAP}: \quad \textit{Maximize} \sum_{i \in I} a_i y_i + \sum_{i \in I} 0.5 \; a_i z_i \tag{7.1}$$

Subject to:

$$\sum_{j \in N_i} x_j \geq y_i \quad \forall i \in I \tag{7.2}$$

$$\sum_{j \in \tilde{N}_i} x_j \geq z_i \quad \forall i \in I \tag{7.3}$$

$$y_i + z_i \leq 1 \quad \forall i \in I \tag{7.4}$$

$$\sum_{j \in J_N} x_j = p \tag{7.5}$$

$$x_j \in \{0, 1\} \quad \forall j \in J_N \tag{7.6}$$

$$y_i \in \{0, 1\} \quad \forall i \in I \tag{7.7}$$

$$\mathrm{z_i} \in \{0, 1\} \quad \forall i \in I \tag{7.8}$$

MAXCAP is structured as originally formulated by ReVelle (1986). The objective, (7.1), seeks to maximize the total capture of new customers. The objective counts two types of customers as being captured. The first corresponds to capturing all customers that are now closer to one of the newly located facilities than any existing facilities. This is tracked in the first term of the objective. Whenever a demand node i is counted as closer than any existing facility, then the variable $y_i = 1$ and its population, a_i, is included in the amount being captured. The second type of capture involves those demands that are now equidistant to an existing facility and a newly sited facility. For those demands that now find their new closest facility at the same distance as an existing facility, it is assumed that the existing facility and new facility will share those customers equally. Thus, when a demand is equidistant between their closest new facility and their closest existing facility, the amount considered captured is one-half of the demand. This reduced level of capture is accounted for in the second term of the objective, where $z_i = 1$ means a new facility has been placed at the exact same distance from demand i as an existing facility. Constraints (7.2) allow variable y_i to equal one in value whenever one of more new facilities have been located among the sites j that are closer to demand i (that is, sites in the set N_i). Constraints (7.3) allow variable z_i to be one in value whenever a facility is located at the same distance to node i as an existing competitor's facility (that is, sites in the set $\tilde{N}_i$). To prevent any double counting, constraint (7.4) restricts capture to be counted in at most one way, complete capture or shared capture. Obviously, if a demand is provided both types of capture, the highest valued capture (i.e., complete) will be selected. Constraints (7.5) indicate that p new facilities are to be sited. Constraints (7.6), (7.7) and (7.8) specify binary restrictions on decision variables.

This formulation of MAXCAP does have two exceptions that require a more nuanced structure. Anytime there is a new facility that is closer than an existing facility, the above model considers that demand as being completely captured, regardless of the locations of other new facilities. This fits the definition of capture offered by ReVelle (1986). The above model also handles the case where each demand has a unique closest existing facility and where at most one new facility is located that is equidistant to that customer. In this case, the demand would be considered to be equally shared between the new and existing facilities, unless, of course, a different new facility is located even closer to capture all demand.

If a demand has two existing facilities the same distance away and the closest new facility is also located at that same distance, then the market needs to be shared three ways as the new facility would only capture a third of the demand not half. This is a nuance that can easily be added into the model objective. Following this logic, in the unlikely event that a demand has three existing closest facilities, and a new facility is located at that exact same distance, then the amount captured would be a fourth of the demand. This too, can easily be added to the objective by modifying captured fractions as appropriate to the second term of (7.1).

The problem becomes a bit more complex, however, when two new facilities are located equidistant from a demand that is the same distance away from its closest existing facility. The above model would consider the capture to be a half, but really the two new facilities and the existing facility would represent a split of demand into thirds, yielding a capture of two-thirds and not a half. Such a circumstance requires further model modification. Consider new variables:

$$w_i = \begin{cases} 1, & \text{if demand } i \text{ has a second new facility that is exactly at the distance } S_i \\ 0, & \text{otherwise} \end{cases}$$

These new variables will be one only when two new facilities have been located equidistant to demand i and where that distance equals S_i, the closest distance of an existing facility. We can modify the objective as follows:

$$Maximize \sum_{i \in I} a_i y_i + \sum_{i \in I} 0.5 \ \ a_i z_i + \sum_{i \in I} 0.166 \ \ a_i w_i \tag{7.9}$$

The basic idea is that a second, new, equidistant facility will add to the half already captured, raising the total to two-thirds, or an additional one-sixth (0.166). We also need to modify constraint (7.3) in the following manner:

$$\sum_{j \in \tilde{N}_i} x_j \geq z_i + w_{\mathrm{i}} \quad \forall i \in I \tag{7.10}$$

as well as introduce a new constraint:

$$y_i + w_i \leq 1 \quad \forall i \in I \tag{7.11}$$

These additions would also require constraints for enforcing the binary restrictions on new variables, w_i. The idea is that if two new facilities are located equidistant to a demand that is at that same distance to an existing facility, then the associated variable w_i would be equal to one. Constraints (7.10) will allow both w_i and z_i to be one in value when two equal distant facilities to demand i are located at the same distance as an existing facility serving this demand. If that is the case, the objective will count two-thirds of the demand as having been captured. If one new facility is located that is the same distance to demand i as the closest existing facility, then this constraint will allow only one of the two variables, w_i or z_i, to be one in value. In that case, the objective will force z_i to be one as it represents an added share of a half, and not a sixth. Finally, constraints (7.11) prevent w_i from being one in value if some other new facility is located even closer, where the demand is fully captured (i.e., $y_i = 1$).[3]

Chapter 6 introduced several models that were based upon expanding the definition of coverage. Rather than using a simple metric, say service within a maximal standard S, to define if coverage has been provided or not, Chap. 6 presented the GMCLP (general maximal covering location problem) that involved several service standards and valued the coverage provided differently for each standard. Chapter 6 also demonstrated how these models could be formulated as equivalent p-median problems, just as was done for the LSCP and MCLP in Chap. 2. ReVelle (1986) made the same connection for MAXCAP, providing a mechanism for using a

[3]This nuance is formulated differently than in ReVelle's original maximum capture paper, as the constraints proposed by ReVelle (1986) actually force demand nodes where there could be multiple shares of capture to have at least one new facility located at that distance, clearly an unintended feature.

transformed distance matrix in the p-median in order to structure MAXCAP.[4] The above model has been expanded to include the feature that some facilities may already exist and the extended form of MAXCAP optimizes capture while some of the existing facilities may be relocated, some may be kept, and others may be added (ReVelle and Serra 1991). Serra et al. (1999) and Drezner et al. (2002) suggest that each facility needs to have a minimum level of customers to be viable, a subject that is discussed later in this chapter. Other extensions have included a hierarchical set of facilities (Serra et al. 1992), uncertainty of demand (Serra et al. 1996), chance constraints on capture (Colomé et al. 2003), a weighted network (Eiselt and Laporte 1989), and a form of the problem involving shrinking a facility chain and ceding as few customers as possible in the process (ReVelle et al. 2007).

Given that we have formulated a model that involves a firm entering a market with an established competitor, a logical question might be: what if there were two firms, a leading firm (A) and a follower firm (B), where the leader firm moves into the market first and the follower firm makes site choices after the leader firm has made theirs? This is exactly the problem posed by Serra and ReVelle (1994). Although they recognized that this problem was a Stackelberg game, their model was that of the leader alone, which is a version of MAXCAP given above. They recognized shortcomings in their model, however, and proposed two heuristics that simulated the course of two firms making decisions and where firm A would site in response to the choices of firm B, providing a mechanism to examine resulting market share. The idea is that after firm A makes site choices of where to locate, Serra and ReVelle (1994) would then apply MAXCAP to see what firm B will do in response. Then, given firm B's selection, they would test possible moves by firm A one at a time associated with each facility in order to assess the likely response of firm B. The idea is to keep those moves that firm A makes where the move represents an improvement in the market share for firm A over the previous locational choice as evaluated after firm B has made their choices. The interest, of course, is to uncover the best strategy for firm A. This type of heuristic (moving facilities one at a time to seek out better locations) has proven relatively robust, but has not been tested in any complete way involving a game between two competitors. It is important to recognize that competitive games, like the above description involving two firms, is a complex problem that cannot be formulated as a single level optimization problem. Rather, it has been approached as a bi-level problem. Further discussion is beyond the scope of this chapter, and remains a promising area for future research, although several bi-level models are described in Chap. 9 for a different type of covering problem.

[4] Note that this equivalent form of the p-median problem cannot handle the case when more than one new facility is placed at the same distance of a competitors involving a given customer.

7.3 Capturing/Intercepting Flow

Most of this book involves cases where demand for service coverage is represented by a set of discrete points or encompassed by a continuous bounded area. However, these two situations do not necessarily address all types of demand, like that for a quick service restaurant, a gas station or a commercial vehicle enforcement facility (truck scale). Much of the potential demand for a restaurant may be reflected by discrete points, but there is another type of demand or target audience that cannot be represented by points or areas. This type of demand has been called “flow” or “traffic” based. Retail location experts are always interested in the amount of traffic, or flow, that streets experience when making site selection choices. Even though many firms might make location decisions based upon the population within a primary shopping zone of 5–7 miles, as an example, firms are also interested in those sites that experience high traffic volumes. That is, the quality of a potential facility site can be a function of both nearby demand and volume of traffic that happens to pass by. Traffic volume is not only of interest for retail site location but is also of value to transportation planners.

Traffic along a road segment over a day or a week can be can viewed as the sum of a number of trips being made that traverse that segment. In order to forecast those areas that will experience traffic congestion, and the need for road investment, better traffic flow control, etc., transportation planners estimate the travel demand that will occur between two locations. It is not uncommon for these locations to be transportation analysis zones, an official reporting unit by the US Census. The interest is in the traffic that originates in one area that is destined for another area. Given a network of nodes and arcs representing major streets/roads that connect two locations, daily traffic volumes indicate some level of interaction between an origin and a destination. Such interaction follows particular paths through the network. It is precisely this interaction that may be of value in facility siting.

Consider the following notation:

$i =$ index of network nodes representing origins (entire set denoted I)

$k =$ index of network nodes representing destinations (entire set denoted K)

$$Q = \{(i,k)|i \in I, k \in K\}$$

$q =$ index of origin-destination pairs

$t_q =$ volume of traffic that occurs between origin-destination pair q

$j =$ index of potential facility location (entire set denoted J)

$p =$ number of facilities to be sited

$N_q =$ set of locations j that can intercept or capture flow of origin-destination pair q

$$\delta_{qj} = \begin{cases} 1, & \text{if } j \in N_q \\ 0, & \text{otherwise} \end{cases}$$

The set Q represents all origin-destination pairs, and through analysis we can determine which potential facility sites j are capable of capturing/intercepting a given origin-destination q pair. We call the set of sites that can capture origin-destination pair q set N_q. This information can then be used to specify an indicator coefficient, δ_{qj}, summarizing whether or not a potential facility j captures/intercepts the flow of origin-destination q pair.

Let's say that we wish to locate a quick service restaurant specializing in hamburgers and we want a site that has the largest number of people traveling passed it during the week. To do this we would have to identify the site j that has the largest traffic volume. This can be calculated as follows for each potential site j:

$$\sum_{q \in Q} \delta_{qj} t_q \tag{7.12}$$

The best site j would be the one with the highest traffic flow, determined by evaluating (7.12) in each case. But, if we wish to locate two or more hamburger joints having the highest number of people that pass by at least one of the restaurants each day, then the problem becomes a bit more complicated. For example, picking the site that experiences the second highest volume of traffic, using equation (7.12), may well identify a site with considerable flow in common with the highest volume site. In this case, the second joint will actually cannibalize sales from the first chosen site. As the goal is to find a site with the most total traffic flow, we want to keep cannibalization to a minimum. Consider the following additional notation:

$$x_j = \begin{cases} 1, & \text{if a facility located at site } j \\ 0, & \text{otherwise} \end{cases}$$

$$y_q = \begin{cases} 1, & \text{if a flow volume generated by origin-destinationn pair } q \text{ is captured} \\ 0, & \text{otherwise} \end{cases}$$

Hodgson (1990) structured a model to identify the best locations for capturing origin-destination flow along the lines described above. This problem is called the Flow Capture Location Model (FCLM) and formulated as follows:

$$\text{FCLM} : \mathit{Maximize} \sum_{q \in Q} t_q y_q \tag{7.13}$$

Subject to:

$$\sum_{j \in N_q} x_j \geq y_q \quad \forall q \tag{7.14}$$

$$\sum_{j \in J} x_j = p \tag{7.15}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.16}$$

$$y_q \in \{0, 1\} \quad q \in Q \tag{7.17}$$

The variable y_q is used to identify whether a given flow volume for origin-destination pair q passes by one or more of the selected facility sites. Even if a volume of flow generated by a given origin-destination pair passes by more than one of the selected facility locations, flow capture is counted only once. This means that the objective (7.13) represents the total traffic volume captured one or more times, so there will be no double counting of any flows. Thus, once captured there is no incentive to locate any facilities to intercept that same origin-destination flow again. Constraints (7.14) specify that flow q can be captured when at least one facility is located at a site capable of capture. When this occurs $y_q$will equal one in value. Otherwise the constraint will force y_q to be zero in value, indicating that flow q has not been captured. Constraint (7.15) specifies that exactly p facilities will be located. Constraints (7.16) and (7.17) impose binary restrictions on decision variables.

As discussed in previous chapters, only the location siting variables x_j technically need to be formally specified as binary when a general integer programming solver is used as the y_q variables will take only zero–one values as long as these variables are bounded to be no larger than 1 in value. One should also recognize the similarity of this model to that of the maximal covering location problem detailed in Chap. 2. In fact, it is exactly the same except the form has been defined so that it represents the problem of capture/intercepting flows instead of covering demand.

The FCLM can be used for a number of different applications, including the selection of billboards in an advertising campaign and the location of truck weighing stations as well as siting (or expanding or contracting) a chain of retail establishments. In each of these cases, the problem is to capture or intercept the maximum amount of flow. Interesting forms of this problem that have been developed and includes the capture of people who might evade a system of intercepting facilities like inspection stations (Marković et al. 2015), paths (Gutiérrez-Jarpa et al. 2010), alternative fuel stations (Hong and Kuby 2016), and a generalized form that addresses many possible nuances (Zeng et al. 2010).

7.4 Capacities

When solving for a configuration of facilities to cover an area (represented by points or other spatial objects), the workload of individual facilities is often overlooked. In many cases, this is a realistic approach. Consider, for example, a system of sirens designed to warn residents of potential danger (e.g., severe storms, wildfires, tornados, etc.). When locating such a system to cover an area, the main issue is to ensure spatial coverage so that all can hear a siren. Covering everything spatially as efficient as possible is the primary concern. The number of people served by an individual siren is not limited. The same can be true for other coverage based service systems, like sensors, cameras, beacons, etc. But, there are systems where the major concern is to seek high levels of coverage, where service at individual facilities can be congested; that is, there may be some limit on how many people can be served by an individual facility. This means that facilities may experience capacity limitations, and those capacities should not be exceeded at each located facility. To ensure that capacities will not be exceeded, explicit tracking of demand service by each facility is necessary. Further, ensuring that the total demand assigned to a given facility does not exceed its capacity is also required. The fact that we need to track or assign individual demands to facilities means that any adopted approach must rest on an allocation process. Accordingly, a covering location-allocation approach is necessary for addressing capacity issues, in contrast to many of the models in this book that account for coverage in a manner that is not explicit in terms demand service assignment. For example, the LSCP involves finding the smallest number of facilities needed to cover every demand, but there is no accounting for which facility is providing service, just that all demand is covered. In fact, some demand will be covered by many facilities while others will be uniquely covered by only one facility. Although we know that every demand is covered, it is not possible to know the amount of demand that is to be served at a facility without adding allocation variables. This is also true for the MCLP. To illustrate the complications that can arise, consider Fig. 7.1 showing a solution derived using the MCLP ($p = 13$). The facilities in this case serve households, and summarized in Fig. 7.1 is the total demand allocated to each sited facility. Of the 9116 households served by this solution, the workload for each facility varies considerably. This ranges from a low of 245 to a high of 1152. There is much variability from the average of 701, and as a result significant inequities across facilities may be experienced along with the potential to exceed facility capabilities (e.g., workload of 1152) as well as

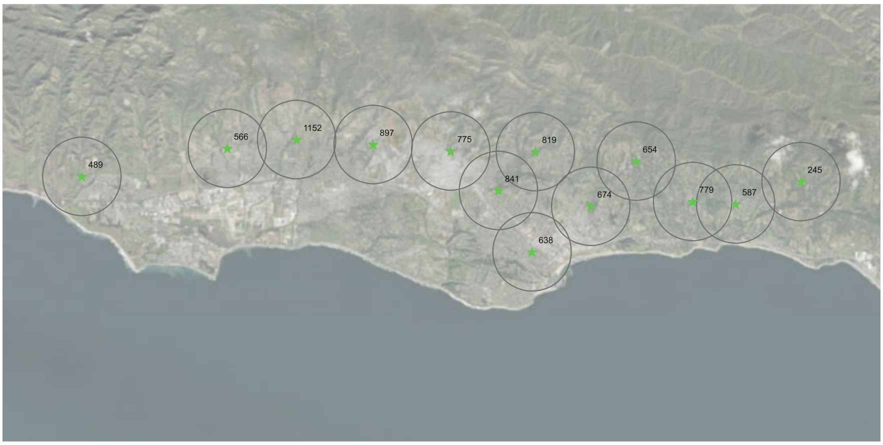

Fig. 7.1 Facility sites identified using MCLP with closest assignment workloads indicated

unreasonably low demand for service (e.g., workload of 245) in some cases. Variability can be addressed through the imposition of capacities for each facility. The early efforts to address capacity issues in coverage modeling include the works of Chung et al. (1983), Smogy and Church (1985), Chung (1986), Current and Storbeck (1988), and Pirkul and Schilling (1989, 1991).

When introducing capacities into a coverage problem, an allocation process must be conceived. The way in which we might allocate demand is dependent upon our assumptions about the service being located as well as the amount of information concerning demand behavior. There are three classical ways to define how demand might be allocated amongst facilities: closest assignment (user optimal), equal assignment probability (based upon the principle of insufficient reason), and system convenience (system optimal). These three demand allocation/assignment approaches hinge on the definition of a system's operational characteristics. Let's say that a system attempts to provide service to a demand by responding to a call for service, like "I want a pizza in 30 minutes". If a system of pizza restaurants covers potential demand across a city within 30 min, and there are overlapping facility service areas, then a central call center can coordinate the response for service, allocating the order to a particular pizza restaurant for preparation, baking and delivery. Assuming that several outlets can satisfy the 30 min standard, then the call center can make the choice as to which of those outlets will do the task, likely picking the facility that has idle capacity. That is, the system makes the choice and not the individual customer with the intent of making the system operate as efficiently as possible. This is a system optimal assignment. If users, on the other hand, make the choice to call a particular facility for their pizza, or decide to travel to the facility to order and eat their pizza, then each individual user makes their choice based upon what maximizes their utility best, not the system. This represents a user optimal approach. A user optimal approach requires some type of choice formula, or utility function, which can be used to allocate demand to individual facilities. Finally, what if we do not know what each individual will do or how they will make their choice regarding facility assignment coverage. If we have no prior knowledge about allocation, then any choice set of alternatives for a customer will be probabilistic. That is, for three facilities covering a given individual demand, as an example, then the individual is assumed to patronize each facility one third of the time. This approach is based upon the work of Bernouli and Laplace (Dupont 1977/78) and is called the principle of insufficient reason. Bernouli, and later Laplace, suggested that if there was no reason to think that one alternative/event was preferred over another, then each choice or event will occur equally likely, hence the equal probabilities of choice in assigning demand to individual facilities in their coverage set. The remainder of this section discusses coverage models for each of these three approaches.

7.4.1 System Optimal Perspective

We begin with the system optimal approach, where the system decides which facility will serve a given demand and where all demands must be covered. Most of the

literature has adopted this form of allocation when adding capacity restrictions to covering facilities. Consider the following notation:
i = index of demand points/areas/objects (entire set denoted I)
j = index of potential facility sites (entire set denoted J)
S = desired maximal service standard (travel distance or time)
d_{ij} = shortest distance or travel time between demand i and potential facility j

$$N_i = \{j \mid d_{ij} \leq S\}$$

$$\Psi_j = \{i \mid d_{ij} \leq S\}$$

p = number of facilities to be located
a_i = amount of demand at i

$$x_j = \begin{cases} 1, & \text{if a facility is located at } j \\ 0, & \text{otherwise} \end{cases}$$

C_j = capacity of potential facility j
z_{ij} = fraction of demand i that is assigned to facility j

The introduction of the z_{ij} variables now enable allocation and tracking of service. Accordingly, a capacitated version of the LSCP is possible, as done in Current and Storbeck (1988). Here we formulate the system-optimal perspective of the capacitated location set covering problem—system optimal (CLSCP-SO) as:

$$\text{CLSCP-SO} : \textit{Minimize} \sum_{j \in J} x_j \tag{7.18}$$

Subject to:

$$\sum_{j \in N_i} z_{ij} = 1 \quad \forall i \in I \tag{7.19}$$

$$\sum_{i \in I} a_i z_{ij} \leq C_j x_j \quad \forall j \in J \tag{7.20}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.21}$$

$$z_{ij} \geq 0 \quad \forall i \in I, j \in N_i \tag{7.22}$$

The CLSCP-SO involves locating just enough facilities and associated capacity such that all demand is served within the capacity limits of each facility, given the coverage capabilities of each facility. Demand allocation is done using variables z_{ij}, indicating the fraction of demand i served by facility j. The objective, (7.18), is equivalent to the original LSCP detailed in Chap. 2. Constraints (7.19) require that the sum of the fractional assignments of demand i add up to one, which means that 100% of that demand will be allocated within the coverage standard (to one or perhaps several facilities). Of course, all assignments must be within a maximal service standard of demand i as well. Constraints (7.20) require that the assigned demand to each facility j cannot exceed its established capacity, C_j. Observe that if a site j is not chosen for a facility, then $x_j = 0$ and the effective capacity on the right hand side of constraint (7.20) will equal zero (the product of zero times C_j). This will restrict the sum of assignments to that site to zero, effectively ensuring that allocations are made only to those sites that are selected for facilities. This model assigns demand at the convenience of the system, such that allocations all fit within the capacity, even though demand may be served by facilities that are not their closest or their preferred choice. Constraints (7.21) impose binary integer restrictions on facility siting variables. Constraints (7.22) indicate non-negativity conditions on assignment variables, which means that demand can be split and portions may be assigned to different facilities.

It should come as little surprise that the CLSCP-SO can be cast in a form that maximizes coverage. The system optimal form of the CMCLP (capacitated maximal covering location problem—system optimal, CMCLP-SO) was first detailed in Chung et al. (1983) and Smogy and Church (1985). The formulation is as follows:

$$\text{CMCLP-SO}: \textit{Maximize} \sum_{i \in I} \sum_{j \in N_i} a_i z_{ij} \tag{7.23}$$

Subject to:

$$\sum_{j \in J} z_{ij} = 1 \quad \forall i \in I \tag{7.24}$$

$$\sum_{i \in I} a_i z_{ij} \leq C_j x_j \quad \forall j \in J \tag{7.25}$$

$$\sum_{j \in J} x_j = p \tag{7.26}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.27}$$

$$z_{ij} \geq 0 \quad \forall i \in I, j \in N_i \tag{7.28}$$

The objective, (7.23), is to maximize total demand covered within the desired maximum service standard. Constraints (7.24) ensure that each demand fully assigns for service. Note that allocation of demand to a facility may be beyond the stipulated service standard. The rationale is that demand beyond the standard may still seek out service, and as a result needs to be served. Constraints (7.25) impose capacities on facilities serving demand. Constraint (7.26) stipulates that p facilities are to be sited. Finally, binary restrictions are indicated for facility variables, constraints (7.27), and non-negativity is required on allocation variables, constraints (7.28).

An interesting distinction in the CMCLP-SO formulation is that there is no service standard limit in the assignment of demand. Demand may well be served beyond the maximum desired service standard, but the goal is in fact to maximize demand served within the standard. This goal is reflected in the objective, (7.23).

One may well be tempted to restrict service capacity to only those within the service standard. Such a constraint form would be as follows:

$$\sum_{i \in \Psi_j} a_i z_{ij} \leq C_j x_j \quad \forall j \in J \tag{7.29}$$

Unfortunately, the use of constraints (7.29) would ignore actual service demand within the system for those beyond the service standard, yet still obtaining service from the system. Such individuals remain likely and able to be served, just not within the desired standard.

The CMCLP-SO here is a restricted form of that developed by Smogy and Church (1985), but is equivalent to that of Chung et al. (1983) (see also Chung 1986). In Chung (1986) there is a suggestion to add the following Balinski type constraints:

$$z_{ij} \leq x_j \quad \forall i \in I, j \in N_i \tag{7.30}$$

The addition of constraints (7.30) in the CMCLP-SO helps to encourage integer solutions in linear programming based techniques, such as branch and bound. This would be considered a *tighter* LP form for this problem. The literature on the use of Balinski constraints in location-allocation models is quite extensive. Such constraints can improve the performance of general purpose software in solving many types of location-allocation models, especially those with similar structure to that of the p-median problem (mentioned in Chap. 2).

The extended model formulated in Smogy and Church (1985) was designed to allocate health personnel, but also ensure service was limited to a manageable number of patients. The approach made it possible to locate more than one facility (heath professional) at a given site through the use of an x_j variable that was allowed to be any positive integer in value.

7.4.2 User Optimal Perspective

When operating a set of facilities, like retail or many public services, people will decide exactly which facility they will visit. That is, rather than the system dispatching service to the customer/demand, like pizza delivery and or EMS response, the customer goes to the facility. In the former case, the system can decide which facility will respond to serve each demand, but in the latter case the user makes the decision as to which facility they will attend. In many cases that will boil down to the demand going to their closest facility. This is often observed for facilities like public libraries, post offices, etc. If we assume a closest assignment paradigm, we need to ensure that the system has enough capacity at each facility to serve all demand while assuming each demand will go to their closest facility. This is an allocation process called closest assignment.

For any facility j, all sites $j^{'}$ that are as close of closer to demand i can be identified. Formally, we define:

$$\Omega_{ij} = \left\{ j' \middle| d_{ij'} \leq d_{ij} \right\}$$

This then is the set of potential facility sites as close or closer to demand i, and includes facility j. Using this notation, we can define the capacitated LSCP with closest assignment (CLSCP-CA) as follows:

$$\text{CLSCP-CA} : \textit{Minimize} \sum_{j \in J} x_j \tag{7.31}$$

Subject to:

$$\sum_{j \in N_i} z_{ij} = 1 \quad \forall i \in I \tag{7.32}$$

$$\sum_{i \in I} a_i z_{ij} \leq C_j x_j \quad \forall j \in J \tag{7.33}$$

$$z_{ij} \leq x_j \quad \forall i \in I, j \in N_i \tag{7.34}$$

$$\sum_{j' \in \Omega_{ij}} z_{ij'} \geq x_j \quad \forall i \in I, j \in N_i \tag{7.35}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.36}$$

$$z_{ij} \geq 0 \quad \forall i \in I, j \in N_i \tag{7.37}$$

The CLSCP-CA differs from the CLSCP-SO because of the use of closest assignment constraints (7.34) and (7.35). These constraints require that demand be allocated/served by its closest sited facility. Constraints (7.35) ensure that each demand wholly assigns to their closest facility, or in the case where a demand has two or more equidistant closest facilities, the sum of that demand's assignment across the set of equally close facilities must be one. This may not handle all of the issues of closest assignment raised in Gerrard and Church (1996) as they demonstrate how demand can be forced to equally share their assignment across a set of equally close facilities, in addition to forcing closest assignment.

As before, we can also formulate a closest assignment form of the user optimality perspective in addressing capacitated maximal coverage. Here the problem involves maximizing coverage (that is, service within some desired standard of service) while locating p facilities. An expectation is that demand would go to their closest facility, even when they are not covered. That is, just because a demand is not covered does not mean that the demand would not go to their closest facility when they needed service. Because of this, sufficient capacity must exist within the system to serve all demand, however some demand are served within the maximum coverage standard and some are not.

The closest assignment form of the capacitated MCLP (capacitated maximal covering location problem with closest assignment, CMCLP-CA) is as follows:

$$\text{CMCLP-CA}: \quad \textit{Maximize} \sum_{i \in I} \sum_{j \in N_i} a_i z_{ij} \tag{7.38}$$

Subject to:

$$\sum_{j \in J} z_{ij} = 1 \quad \forall i \in I \tag{7.39}$$

$$\sum_{i \in I} a_i z_{ij} \leq C_j x_j \quad \forall j \in J \tag{7.40}$$

$$\sum_{j \in J} x_j = p \tag{7.41}$$

$$z_{ij} \leq x_j \quad \forall i \in I, j \in J \tag{7.42}$$

$$\sum_{j' \in \Omega_{ij}} z_{ij'} \geq x_j \quad \forall i \in I, j \in J \tag{7.43}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.44}$$

$$z_{ij} \geq 0 \quad \forall i \in I, j \in N_i \tag{7.45}$$

The objective, (7.38), is to maximize total demand covered within the desired maximum service standard. Constraints (7.39) require demand to be completely allocated to facilities. As was the case for the CMCLP-SO, allocation of demand to a facility may be beyond the stipulated service standard. Constraints (7.40) impose capacities on facilities serving demand. Constraint (7.41) stipulate that p facilities are to be sited. The major distinction of the CMCLP-CA is the inclusion of closest assignment constraints (7.42) and (7.43). Constraints (7.43) in particular require that each demand assign to their closest facility, or facilities in situations where they may be equidistant. Again, one caveat is that particular cases may arise where (7.42) and (7.43) insufficient (see Gerrard and Church 1996). Binary restrictions are indicated for facility variable, constraints (7.27), and non-negativity is required on allocation variables, constraints (7.28).

Little research has actually focused on user optimal situations in coverage modeling. A notable exception is the work of Gerrard (1996), with the CLSCP-CA and CMCLP-CA inspired by his research. It should also be pointed out that there are other forms of user assignment that can be used as well when crafting a user optimal location covering problem. Potential examples include the incorporation of assignments through the gravity model and utility based multinomial logit models. Such models, for all practical purposes, have yet to be developed.

7.4.3 Equal Fraction Perspective

The third perspective noted for allocation of demand is that of equal probability when there is insufficient information on demand preferences. The is referred to here as an equal assignment fraction approach because without additional information, one must assume that probabilistically there is an equal chance among potential choices. The idea is that if there are several facilities that can serve a given demand (within some market range), and if we lack information regarding preferences to make allocations to these facilities, then we are left with a situation where assignments are equally probable. That is, if a demand location is served within the range of three facilities, then assuming that one-third will go to one of the facilities, a third to the next and a final third to the last of the three facilities is the only reasonable conclusion. As mentioned above, the fractions of assignment will be equal based upon the principle of insufficient reason first raised by Bernoulli in the 1600s. It has also been called the principle of indifference and the equal distribution of ignorance (Dembski and Marks 2009). The model that we present below is inspired by the work of Balakrishnan and Storbeck (1991), although their work will be discussed later in the chapter.

The nature of capacitated location models, in general, requires a set of allocation variables, in addition to variables that represent the selection of specific sites for facilities. These two components represent the very heart of facility location-allocation modeling. The models in the previous two sections used allocation variables, z_{ij}, representing each possible assignment of a demand i to a potential facility j. Somewhat unique in the equal assignment case is that specific allocations of demand to specific facilities are not necessary. For instance, say demand i is served within the service standard range S by three different facilities. If we assume that demand i will be equally split between the three "covering" facilities, then we can determine and calculate the portions of demand i assigning to each individual facility by just knowing how many times demand i is covered. Thus, consider the following type of decision variable:

$$y_i^k = \begin{cases} 1, & \text{if demand } i \text{ is served (or covered) by exactly } k \text{ facilities} \\ 0, & \text{otherwise} \end{cases}$$

In the case of location set covering, we know each demand would be served at least once, but it is possible to be served by all sited facilities. Thus, the number of options is the set K, where $K = \{1, 2, 3, \ldots, k^{\max}\}$. Given this, the contribution of demand i for a given facility j providing service coverage is:

$$\sum_{k \in K} \left(\frac{1}{k}\right) a_i y_i^k \tag{7.46}$$

where the fraction of $1/k$ represents the portion of demand a_i that will be assigned to facility j given that there are exactly k facilities providing service to demand i. From this basis, we can now formulate an equal assignment form of the capacitated LSCP (capacitated location set covering problem with equal assignment, CLSCP-EA) as:

$$\text{CLSCP-EA}: \quad \textit{Minimize} \ \sum_{j \in J} x_j \tag{7.47}$$

Subject to:

$$\sum_{j \in N_i} x_j - \sum_{k \in K} k \ y_i^k = 0 \quad \forall i \in I \tag{7.48}$$

$$\sum_{k \in K} y_i^k = 1 \quad \forall i \in I \tag{7.49}$$

$$\sum_{i \in \Psi_j} \sum_{k \in K} \left(\frac{1}{k} \right) a_i y_i^k \le C_j x_j \quad \forall j \in J \tag{7.50}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.51}$$

$$y_i^k \in \{0, 1\} \quad \forall i \in I, k \in K \tag{7.52}$$

Objective (7.47) is the same as the classic form of the LSCP, where the number of facilities providing service is minimized. Constraints (7.48) and (7.49) together define feasible values of y_i^k. Constraints (7.49) indicate that exactly one y_i^k variable will be equal to 1 for each demand i. Given this, constraints (7.48) account for the number of facilities that cover demand i exactly k times. With the definition of the set K involving only positive integers greater than zero, constraints (7.48) do not allow a given demand to be left uncovered. That is, at the bare minimum, demand i must be covered at least once. Constraints (7.50) ensure that the assigned demand to each facility does not exceed its capacity. Constraints (7.51) and (7.52) impose binary integer requirements.

An extension to address the option of equal assignment for the capacitated form of the MCLP (capacitated maximal covering location problem with equal assignment, CMCLP-EA) is as follows:

$$\text{CMCLP-EA}: \quad \textit{Minimize} \ \sum_{i \in I} a_i y_i^0 \tag{7.53}$$

Subject to:

$$\sum_{j \in N_i} x_j - \sum_{k=1}^{p} k \ y_i^k + y_i^0 = 0 \quad \forall i \in I \tag{7.54}$$

$$\sum_{k=0}^{p} y_i^k = 1 \quad \forall i \in I \tag{7.55}$$

$$\sum_{j \in J} x_j = p \tag{7.56}$$

$$\sum_{i \in I} \sum_{k=1}^{p} \left(\frac{1}{k}\right) a_i y_i^k \leq C_j x_j \quad \forall j \in J \tag{7.57}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.58}$$

$$y_i^k \in \{0, 1\} \quad \forall i \in I, k \in \{0, 1, 2, \ldots, p\} \tag{7.59}$$

The CMCLP-EA differs in subtle but important ways than the CLSCP-EA, but also the other CMCLP formulations. First, objective (7.47) now minimizes the total amount of demand not covered or served. This is equivalent to maximizing demand covered, a distinction highlighted in Chap. 2 for the MCLP. Second, constraints (7.54) have been extended to account for the case where no coverage is provided within the desired service standard. In particular, either one or more facilities are located so that a given demand i is covered or the associated y_i^0 variable is forced to equal 1. This option is also included in constraints (7.55). Third, there is constraint (7.56) specifying that exactly p facilities are to be sited. Constraints (7.57) establish capacity limits on sited facilities. In this case, however, only demand that is covered is allocated to the located facilities. This is in contrast to the assignment of all demand in the cases of CMCLP-SO and CMCLP-CA. Extension to account for the allocation of all demand in terms of capacity considerations is possible, but will require the use of assignment variables. We leave such a variant of the CMCLP-EA as a topic for future research. Binary integer requirements are imposed in constraints (7.58) and (7.59) for all decision variables.

7.5 Thresholds

The previous section was devoted to ensuring that located facilities would not be overwhelmed with demand when there are limits to what each facility can handle. We addressed three different methods of allocating demand to facilities and formulated coverage models, extensions of the LSCP and MCLP. Collectively, these capacitated models provide ways for helping to identify facility siting configurations that address a major issue in planning: functional capacity limits. But there is another issue that is important in planning a system, dealing with minimum service demand thresholds at each facility. As discussed in Chap. 1, Christaller (1933) developed a theory of central places, involving spatial patterns of service/retail centers across a region. He reasoned that the market area of each central place was limited to some distance that he called the "range", a distance beyond which people were unwilling to travel for shopping and services. The demand that was enclosed within the range distance represented the highest potential for a market center, or even an individual facility. He further suggested that each center needed some minimum amount of demand or customers in order to be profitable and stay in business. He called this the “threshold”. Simply put, if the demand within the range of a site is not as large as the threshold for a given type of facility or service, then that site is not viable for that type of service. For all practical purposes, many activities require some threshold of service demand in order to be viable. This includes a restaurant, a dry cleaning business, a grocery store, etc. Location models, beyond those developed in central place theory, often ignore the need to address threshold issues. That does not mean that some approaches have not included constraints on facility thresholds. For example, the classic warehouse location model of Geoffrion and Bride (1978) includes capacity constraints as well as threshold constraints.

Perhaps the first covering model developed using threshold constraints was the regional solid waste planning model of Church (1980). This model was developed for the Tennessee Valley Authority (TVA), a federal agency. They wanted to carve up the region into solid waste planning areas or districts. Each district was to have a designated center, a place where solid waste recycling would be economically viable. The model that TVA employed divided the region of 201 counties into districts, by assigning counties to those locations selected as district centers. The location model sited centers and assigned counties to these centers based on optimizing three different objectives: minimize the weighted distances of possible waste haulage (to each designated center from the other counties in each district), maximize the amount of waste that could be economically transported to centers, and minimize the number of centers that would have an expected volume of waste that was less than 1000 tons per week. The first objective represents that of the *p*-median problem, the second objective reflects a maximal covering goal, and the third objective amounted to a desired threshold. Specifically, the TVA wanted as many viable recycling centers as possible, recognizing that a base level of activity (tonnage per week) was important. The threshold objective was a means for seeking this out within a multi-objective model. Solutions could be generated on the Pareto frontier based on all three objectives. The basic idea was twofold: one, covering models can often be paired with other location objectives resulting in a multi-objective location model; and two, it is possible to either require that all facilities in a solution meet a minimum threshold of business or service demand or the number of facilities can be minimized that do not have an expected demand that equals or exceeds a minimum threshold. For the remainder of this section, we will assume that thresholds should be enforced everywhere, rather than minimizing the number of facilities that do not meet a minimum threshold.

Three different approaches for demand assignment were explored in the previous section. They included system optimal, user optimal, and equal assignment. The former two approaches utilized capacity constraints (7.20) and (7.33) for the LSCP, and similarly imposed in the MCLP using constraints (7.25) and (7.40). Given this, it is easy to structure constraints that enforce minimum thresholds as:

$$\sum_{i \in I} a_i z_{ij} \geq T_j x_j \quad \forall j \in J \tag{7.60}$$

where T_j represents the minimum required threshold of service demand at facility j. One could readily add constraint (7.60) to the CLSCP-SO, CMCLP-SO, CLSCP-CA or CMCLP-CA models in order to address threshold requirements.

Capacity constraints (7.50) and (7.57) for the equal fraction approaches also suggest a straightforward structure for imposing thresholds:

$$\sum_{i \in I} \sum_{k=1}^{p} \left(\frac{1}{k}\right) a_i y_i^k \geq T_j x_j \quad \forall j \in J \tag{7.61}$$

Accordingly, the CLSCP-EA and CMCLP-EA could be extended by adding constraints (7.61) in order to address threshold requirements.

A variant of the CMCLP-EA involves substituting threshold constraints (7.61) for capacity constraints (7.57), yielding a form of the McTHRESH model of Balakrishnan and Storbeck (1991) developed for market location. Other covering based models that have involved threshold constraints include the works of Carreras and Serra (1999), Hong and Kuby (2016) and Drezner et al. (2002). Another interesting approach involving covering models and thresholds is that of Storbeck (1988, 1990) where theoretical central place patterns were analyzed. Storbeck (1988, 1990) was able to demonstrate that the central place patterns of Christaller (1933) could be generated using a protected threshold covering model applied to a uniform triangular lattice of demand and facility points.

When developing covering models for specific applications, both capacity and threshold constraints may be necessary. The issue at hand is that when such conditions are added to a covering problem, additional allocation variables or coverage service level variables are likely needed in order to track the exact demand loading on each sited facility. This often requires additional constraints as well. Altogether, the models are generally more difficult to solve using general purpose mixed integer programming software, possibly to the extent that exact solution is not possible. One of the reasons for this is that capacity and threshold constraints are a form of knapsack constraints. Knapsack-type constraints usually encourage fractional values among the allocation variables in a linear programming relaxation. This usually results in much larger branch and bound trees in the search for an optimal solution, and concomitantly longer solution times. As a result, development of specialized solution approaches has been necessary (see for example Carnes and Shmoys 2008).

7.6 Franchise Territory Design

Problems that deal with locating retail and service facilities, like a franchise operation (e.g., 7-Eleven, Carl's Jr., Dairy Queen, Subway, etc.), share a lot in common with Christaller's central place theory. In fact, several covering models have been developed to analyze central place patterns as well as locate franchise facilities. Although Christaller (1933) was principally interested in the location of villages, towns, and cities, where each of them had a variety of services and types of retail establishments, the same issues apply whether it is an individual facility being located or the emergence of a town or city.[5] We concentrate here on the issues of locating a set of facilities across an area, and in particular franchisee facilities. It is important to remember the concepts of threshold and range as they apply to a given type of business. The range represents the furthest extent to which a facility might capture or attract demand/customers and threshold represents the minimum amount of demand (or number of customers) such that an individual facility is viable/profitable. One can think of the types of allocation (system optimal, user optimal, and equal assignment) as approaches to estimate the amount of demand that will be served by a facility. This will need to be as much or larger than the threshold for that type of business. In fact, we can structure models for multiple retail facility location using the constraints and variables already used. For example, the CMCLP-CA could be modified such that the capacity constraints, (7.40), are removed and threshold constraints (7.61) appended, giving us a retail chain facility location model. This would be an alternative to the McTHRESH model discussed above.

There is, however, an interesting problem that arises when retail establishments are franchisee owned and franchisor licensed. Wherever they are located, the franchisor would like to ensure that each facility maintains a threshold of demand. What is particularly interesting and important is that there are two perspectives, one of the franchisor who wants to locate and license as many viable franchisee facilities as possible, and the other perspective of the franchisee who hopes for a large market area and customer base with as few other facilities (or none) of the same franchisor in the region as possible. That is, the franchisor wants a lot of facilities in order to saturate a market area and the franchisee wants to see as few as possible. This problem was first raised by Current and Storbeck (1994), which they termed FVF (franchisor vs. franchisee).

[5]To be complete, Christaller (1933) suggests a hierarchy of centers and types of goods. For instance a village may have a grocery store, a hardware store, and a gas station, whereas a town will offer all of those goods as well as many other facilities, like clothing stores, automobile dealerships, and a hospital. Cities have an even larger set of services, including all of those offered in villages and towns, but in addition even "higher-ordered" goods and services. These include such businesses as medical specialists and high-end jewelry stores. It is important to note that Church and Bell (1990) have demonstrated that co-location of competitors can exist in stable market central place configurations, so that, depending on the nature of the business, competing firms will locate in the same area.

The FVF problem recognizes the tradeoff between the perspective of the franchisor wanting many facilities and the franchisee wanting as few as possible. Consider the following additional or modified notation:

$$\delta_{ij} = \begin{cases} 1, & \text{if demand } i \text{ is within the range of potential outlet location } j \\ 0, & \text{otherwise} \end{cases}$$

$$\theta_{ij} = \begin{cases} 1, & \text{if demand } i \text{ is within the threshold of potential outlet location } j \\ 0, & \text{otherwise} \end{cases}$$

$$x_j = \begin{cases} 1, & \text{if facility location } j \text{ is chosen for a franchisee outlet} \\ 0, & \text{otherwise} \end{cases}$$

$$r_i = \begin{cases} 1, & \text{if demand } i \text{ is not within the range of a facility} \\ 0, & \text{otherwise} \end{cases}$$

$$t_i = \begin{cases} 1, & \text{if demand } i \text{ is not within the threshold of a facility} \\ 0, & \text{otherwise} \end{cases}$$

e_i = number of additional facilities for which demand i is within the range
w_1, w_2, w_3 = importance weights for three different objective terms.

The Franchisor versus Franchisee (FVF) model can be structured as:

$$\text{FVF}: \quad \textit{Minimize} \quad -w_1 \sum_{j \in J} x_j + w_2 \sum_{i \in I} a_i t_i + w_3 \sum_{i \in I} a_i e_i \tag{7.62}$$

Subject to:

$$\sum_{j \in J} \delta_{ij} x_j - e_i + r_i = 1 \quad \forall i \in I \tag{7.63}$$

$$\sum_{j \in J} \theta_{ij} x_j + t_i = 1 \quad \forall i \in I \tag{7.64}$$

$$x_j \in \{0, 1\} \quad \forall j \in J \tag{7.65}$$

$$t_i \in \{0, 1\} \quad \forall i \in I \tag{7.66}$$

$$r_i \in \{0, 1\} \quad \forall i \in I \tag{7.67}$$

The objective, (7.62), of the FVF contains three terms, each weighted by an importance weight reflecting the relationship of one to the others. The first term is the franchisor's main goal of maximizing the number of outlets being located. The second term attempts to minimize the number of people who are not in the threshold of a facility. This, too, is an objective of the franchisor, who would like nothing better than having all demand served within the threshold of a facility. The third term of the objective minimizes the total demand that experiences multiple facilities outside the threshold distance but inside the range distance. This represents the objective of the franchisee, who wishes to keep competition within the market range to be at a minimum and potentially increase profits beyond the threshold of viability. Constraints (7.63) and (7.64) make use of coverage matrices, δ_{ij} and θ_{ij}, indicating demand within the range and threshold standards. They are derived in advance and represent spatial information and relationships structured in the model. Constraints (7.63) effectively track how many facilities are in the range of a demand. Similarly, constraints (7.64) account for whether a demand is part of the threshold demand of a facility. Notice that constraint (7.64) also ensures that demand i will not be within the threshold of two or more facilities, thus making the facility viable in terms of a minimum threshold of customers. But, if all customers are within a threshold of a facility, then the market will be saturated with viable facilities (the goal of a franchisor). One might consider the first two objectives as being complementary, with potential to omit the second objective. But, Current and Storbeck (1994) argue that the second term can be used to tease out alternate optima when trading off the first and third objectives, conceivably an important issue.

7.7 Summary and Concluding Comments

Many models have been applied in facility location by retail chains, and they are often considered proprietary. Although books such as *Location, Location, Location: how to select the best site for your business* by Salvaneschi (1996) provide many details on the issues of site choice, the overall approach that is suggested could be classified as a guided seat-of-the pants process, once potential trade or market demand has been determined within 10 or 15 min of travel time form a potential location. Guides such as this book and others often neglect to state exactly how to address the multiple facility location problem, where sales at one facility might be cannibalized by another. Such guides also often lack good methods to estimate facility revenue, given a potential customer base within certain travel times to a site.

Large companies today may have a GIS department where volumes of data on customer purchases can be analyzed within a spatial context. The type of data that is collected by a retail chain includes a wealth of customer information, even knowing who comes to a facility and uses a coupon cut from a newspaper or printed from an online website. This data can and is being used in sophisticated ways to develop better patronage models that earlier works such as those in Salvaneschi (1996) did not anticipate. These companies are using models to select multiple facility locations simultaneously, as compared to the one at a time approach discussed by Salvaneschi (1996). Hodgson et al. (1996) apply the FCLP in such an extended manner, utilizing a road network and associated origin-destination travel in Edmonton, Canada. Spaulding and Cromley (2007) detail application of MAXCAP integrated into a GIS framework.

Estimating how many customers will be attracted to a new facility as compared to existing competitor facilities has been a problem that many researchers have attempted to tackle with varying degrees of success. Reilly (1929) was one of the first to delineate market areas using a gravity model. Market boundaries were assumed to be those points at which half of all customers travel to one competitor and half travel to the other competitor. The gravity model, and other models like a multinomial logit model, can take into account the differences in facility size as well as travel costs and other amenities. The sophistication of modeling customer demand has advanced considerably in the last 10 years in the era of big data. Customer demand models are now based upon first dividing the population into different lifestyle profiles, e.g., young married couples. Each type of facility attracts different proportions of people based upon their profiles, e.g., affluent-retired. Thus, the capture of demand can be modeled with a more refined approach (Benati 1999). But, one can still view this as a customer capture problem, and the model structures reviewed here can include more refined estimates of what is being captured. It is also possible to extend the capture process from complete capture and exact sharing of customers with portioning schemes a company experiences with their competitors.

References

Balakrishnan PV, Storbeck JE (1991) McTHRESH: modeling maximum coverage with threshold constraints. Environ Plan B: Plan Des 18(4):459–472

Benati S (1999) The maximum capture problem with heterogeneous customers. Comput Oper Res 26(14):1351–1367

Carnes T, Shmoys D (2008) Primal-dual schema for capacitated covering problems. In: International conference on integer programming and combinatorial optimization. Springer, Berlin, pp 288–302

Carreras M, Serra D (1999) On optimal location with threshold requirements. Socio-Econ Plan Sci 33(2):91–103

Christaller W (1933) Central places in Southern Germany. The pioneer work in theoretical economic geography (trans: Carlisle Baskinin 1966). Prentice-Hall, Englewood Cliffs

Chung CH (1986) Recent applications of the maximal covering location planning (MCLP) model. J Oper Res Soc 37:735–746

Chung CH, Schilling DA, Carbone R (1983) The capacitated maximal covering problem: a heuristic. In: Proceedings of 14th annual Pittsburgh conference on modeling and simulation, pp 1423–1428

Church RL (1980) Developing solid waste planning regions for the Tennessee Valley Authority. In: Proceedings of the 11th annual Pittsburgh conference on modelling and simulation, vol 11, pp 1611–1618

Church RL, Bell TL (1990) Unpacking central place geometry I: single level theoretical k systems. Geogr Anal 22(2):95–115

Colomé R, Lourenço HR, Serra D (2003) A new chance-constrained maximum capture location problem. Ann Oper Res 122(1):121–139

Current JR, Storbeck JE (1988) Capacitated covering models. Environ Plan B: Plan Des 15 (2):153–163

Current JR, Storbeck JE (1994) A multiobjective approach to design franchise outlet networks. J Oper Res Soc 45(1):71–81

d'Aspremont C, Gabszewicz JJ, Thisse JF (1979) On Hotelling's "Stability in competition". Econ: J Econ Soc 1145–1150

Dembski WA, Marks RJ (2009) Bernoulli's principle of insufficient reason and conservation of information in computer search. In: IEEE international conference on systems, man and cybernetics, 2009. SMC 2009, IEEE, pp 2647–2652

Drezner T, Drezner Z, Shiode S (2002) A threshold-satisfying competitive location model. J Reg Sci 42(2):287–299

Dupont P (1977/78) Laplace and the Indifference Principle in the *Essai philosophique des probabilités*. Rend Sem Mat Univ Politec Torino 36:125–137

Dwyer FR, Evans JR (1981) A branch and bound algorithm for the list selection problem in direct mail advertising. Manage Sci 27(6):658–667

Eiselt HA, Laporte G (1989) The maximum capture problem in a weighted network. J Reg Sci 29 (3):433–439

Geoffrion A, Bride RM (1978) Lagrangean relaxation applied to capacitated facility location problems. AIIE Trans 10(1):40–47

Gerrard RA (1996) The location of service facilities using models sensitive to response distance, facility workload, and demand allocation. PhD Dissertation, University of California, Santa Barbara

Gerrard RA, Church RL (1996) Closest assignment constraints and location models: properties and structure. Locat Sci 4(4):251–270

Gutiérrez-Jarpa G, Donoso M, Obreque C, Marianov V (2010) Minimum cost path location for maximum traffic capture. *Computers & Industrial Engineering* 58(2):332–341

Hansen P, Thisse JF (1981) Outcomes of voting and planning: Condorcet, Weber and Rawls locations. J Public Econ 16(1):1–15

Hodgson MJ (1990) A flow-capturing location-allocation model. Geogr Anal 22(3):270–279

Hodgson MJ, Rosing KE, Leontien A, Storrier G (1996) Applying the flow-capturing location-allocation model to an authentic network: Edmonton, Canada. Eur J Oper Res 90(3):427–443

Hong S, Kuby M (2016) A threshold covering flow-based location model to build a critical mass of alternative-fuel stations. J Transp Geogr 56:128–137

Hotelling H (1929) Stability in competition. Econ J 39(1929):41–57

Marković N, Ryzhov I, Schonfeld P (2015) Evasive flow capture: optimal location of weigh-in-motion systems, tollbooths, and security checkpoints. Networks 65(1):22–42

Pirkul H, Schilling D (1989) The capacitated maximal covering location problem with backup service. Ann Oper Res 18(1):141–154

Pirkul H, Schilling DA (1991) The maximal covering location problem with capacities on total workload. Manage Sci 37(2):233–248

Reilly WJ (1929) Methods for study of retail relationships. Research Monograph, 4, Bureau of Business Research, The University of Texas, Austin

ReVelle C (1986) The maximum capture or "sphere of influence" location problem: hotelling revisited on a network. J Reg Sci 26(2):343–358

ReVelle C, Serra D (1991) The maximum capture problem including relocation. INFOR 29 (2):130–138

ReVelle C, Murray AT, Serra D (2007) Location models for ceding market share and shrinking services. Omega 35(5):533–540

Salvaneschi L (1996) Location, location, location: how to select the best site for your business. Oasis Press/PSI Research, Grants Pass

Serra D, ReVelle C (1994) Market capture by two competitors: the preemptive location problem. J Reg Sci 34(4):549–561

Serra D, Marianov V, ReVelle C (1992) The maximum-capture hierarchical location problem. Eur J Oper Res 62(3):363–371

Serra D, Ratick S, ReVelle C (1996) The maximum capture problem with uncertainty. Environ Plan B: Plan Des 23(1):49–59

Serra D, ReVelle C, Rosing K (1999) Surviving in a competitive spatial market: the threshold capture model. J Reg Sci 39(4):637–650

Smogy C, Church RL (1985) Balancing access and service cover. Paper presented at the North American Meetings of the Regional Science Association, Philadelphia, PA
Spaulding BD, Cromley RG (2007) Integrating the maximum capture problem into a GIS framework. J Geogr Syst 9(3):267–288
Storbeck JE (1988) The spatial structuring of central places. Geogr Anal 20(2):93–110
Storbeck JE (1990) Classical central places as protected thresholds. Geogr Anal 22(1):4–21
Zeng W, Castillo I, Hodgson MJ (2010) A generalized model for locating facilities on a network with flow-based demand. Netw Spat Econ 10(4):579–611