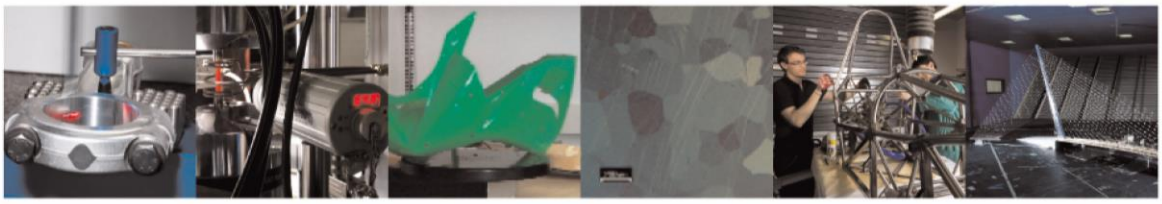
POLITECNICO MILANO 1863

DIPARTIMENTO DI MECCANICA

mecc



The patient assignment problem in home health care: using a data-driven method to estimate the travel times of care givers

Yalcindag, Semih; Matta, Andrea; Şahin, Evren; Shanthikumar, J. George

This is a post-peer-review, pre-copyedit version of an article published in FLEXIBLE SERVICES AND MANUFACTURING JOURNAL. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10696-015-9222-6

DIPARTIMENTO DI MECCANICA ■ POLITECNICO DI MILANO
via G. La Masa, 1 ■ 20156 Milano ■ EMAIL (PEC): pecmecc@cert.polimi.it
http://www.mecc.polimi.it
Rev. 0

Noname manuscript No.
(will be inserted by the editor)

The Patient Assignment Problem in Home Health Care: Using A Data-Driven Method to Estimate the Travel Times of Care Givers

Semih Yalçındağ · Andrea Matta · Evren Şahin · J. George Shanthikumar

Received: date / Accepted: date

Abstract Home Health Care is one of the recent service systems where human resource planning has a great importance. The assignment of patients to care givers is a relevant issue that the Home Health Care service provider must address before generating the daily routes. The assignment decision is typically made without knowing the visiting sequence, which creates some uncertainties and disparities regarding the effective workload of care givers. However, taking into account travel times in the care giver workload while solving the assignment problem is not straightforward, because travel times can also be affected by clinical conditions of patients and their homes.

Providing good travel time estimates that would be used in the assignment decision is the specific topic this paper focuses on. In particular, we propose a data-driven method to estimate the travel times of care givers in the assignment problem when their routes are not available yet. The method, based on the Kernel Regression technique, uses the travel times observed from previous periods to estimate the time necessary for visiting a set of patients

S. Yalçındağ
Dipartimento di Meccanica, Politecnico di Milano, Via La Masa 1, Milano, Italy
Department of Management Sciences, HEC Montréal, 3000 Chemin de la Cote-Sainte-Catherine, Montréal, Canada
E-mail: semih.yalcindag@polimi.it, semih.yalcindag@hec.ca

A. Matta
School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai, China
E-mail: matta@sjtu.edu.cn

E. Şahin
Laboratoire Génie Industriel, CentraleSupelc, Grande Voie des Vignes, Chatenay-Malabry Cedex, France
E-mail: evren.sahin@centralesupelec.fr

J. G. Shanthikumar
Krannert School of Management, Purdue University, 403 W. State Street West Lafayette, Indiana, USA
E-mail: shanthikumar@purdue.edu

located in specific geographical locations. The main advantage offered by this technique is the empirical modelling of the travel routes generated by care givers. Numerical results based on realistic problem instances indicate that the proposed estimation method performs better than the Average Value and k-Nearest Neighbor search methods and can be successfully used in a two–stage approach that first assigns patients to care givers and then defines their routes.

Keywords Home health care · Resource assignment · Kernel regression

1 Introduction

With the ever increasing cost of operations and various constraints coming from customers or service operators, the service industry is faced with the tough challenge of offering better service quality while keeping costs as low as possible. This issue is even more important for *mobile services* [10, 19] that involve the traveling of service operators among customer sites and eventually, the realization of on-site activities. Indeed, home delivery, equipment (appliance, elevator, etc.) installation and repair services [18, 32] are typical examples of such services that include the transportation of goods and personnel (competencies) spending some time at customers' places. For those services, issues regarding the planning process need to be supported by innovative decision making models. The planning process of interest developed in this paper is deciding which operator is in charge of which customers, i.e. the assignment problem.

Home Health Care (HHC) is an example of such mobile services that has known a fast recent growth in the health care sector, representing an alternative to the conventional hospitalization in developed countries [36, 41]. Indeed, HHC providers deliver medical, paramedical and social services to patients in their homes. Among decisions related to HHC resources planning are issues such as the dimensioning of materials and equipments, the determination of the required number of care givers, the partitioning of the territory served by the HHC provider into districts, the resource allocation to the districts, the assignment of care givers to patient visits and the specification of routing plans for each care giver [6].

The heterogeneity of HHC services and territories where they operate generated different operations management approaches; we refer to [36] for a description of several planning levels and processes adopted in current practice. Hence, in practice, once the patient is admitted to the HHC service, according to his/her therapeutic project, the resource assignment problem determines which care givers (operators) will provide care for which patients. Daily routes are then defined for each care giver, which specify the sequence in which patients are visited. Within this context, integrating routing considerations while assigning patients to care givers (without explicitly solving the assignment and routing problems at the same time) is a challenging research question.

Main factors to consider in the assignment problem are the visiting times, the travelling times and the professional skills required to deliver the service to a specific patient. Nevertheless, other characteristics than the geographical locations of patients would have an impact on the assignment of patients. Examples of such characteristics can stem from features related to patient care requirements (i.e., their care profile) or the geographical aspects of the territory where the HHC provider operates. For instance, the visit of a patient requiring a blood test would most probably be done early in the morning, although it could be optimal to visit him/her at the end of the day if only geographical coordinates are considered. Thus, this would have an impact on the assignment decision that has to be held. Other features such as the information regarding the availability, for a given day, of patient family members that help care givers is another motivating feature that the HHC provider has to take into account while deciding for the assignments.

In summary, these characteristics, that are referred in this paper as *patient attributes*, have an impact on decision making at assignment level because they may force care givers to realize longer routes, thus increasing their effective assigned workload. Therefore, considering at assignment level the care giver travel times as the result of a travel time minimization problem (that considers only geographical locations) would likely lead to non optimal decisions. In practice, the HHC planner would assign a patient to a different care giver than the one to whom (s)he would be assigned when only the geographical criterion based on Euclidean traveling times are used. Such features can be captured by the available historical data that would give information regarding the choices made in previous routes accomplished by a given care giver. A data–driven approach would then enable to estimate (future) travel times based on care giver's specific past behaviors. These estimations could be used to obtain higher quality assignments. The development of a data–driven travel time estimation method for supporting decision making at assignment level is the objective of this paper.

More specifically, we propose Kernel Regression as a travel time estimation method for capturing the features that mostly impacted the choices of care giver routes in the previous periods. In order to demonstrate the effectiveness of the method, the solution of the assignment problem is used to generate the route of each care giver in a *two–stage approach* as in many HHC organizations. Then, the performance of the two–stage approach using the Kernel Regression method is compared to the one generated by a *simultaneous approach* in which assignment and routing problems are jointly solved. The comparison is based on a simplified setting of a HHC planning problem, in which some sources of complexities such as time windows, synchronization constraints and care givers' capacities are not taken into consideration. This simplified experimental setting enables to focus on the accuracy and effectiveness of the Kernel Regression method rather than on the comprehensiveness of constraints that may be considered in HHC assignment and routing problems.
As patient attributes, only the geographical location is considered in the numerical experiments. The main reason is to test the validity of the approach in

the most unfavorable situation for the Kernel Regression method when only travel time minimization matters and the simultaneous approach is theoretically the most appropriate one.

The results, based on realistic problem instances, show that the proposed estimation method performs better than existing ones (i.e., the Average Value (AV) and k-Nearest Neighbor (kNN) search methods) and can be successfully adopted in the assignment problem. Another merit of the proposed method is that it enables to solve the assignment problem by integrating features related to the routing problem. This paves the way to develop approximate two–stage approach methods for solving large scale instances of HHC assignment and routing problems.

The paper is organized as follows. Section 2 provides a literature review on the HHC planning problem. Section 3 describes the proposed travel time estimation method. Section 4 presents the models of the planning approaches used in the comparison whereas section 5 provides the numerical results. Finally, section 6 presents concluding remarks and some future research directions.

2 Literature review

Existing research contributions on HHC can be split into two parts. The first category of papers concerns works that focus on the assignment problem as part of a two–stage planning process (assignment and routing stages), whereas the second category of papers represents works related to the simultaneous approach. Since the present paper focusses on the assignment problem, this section gives more details on the first category of papers.

Within this first category of papers, Hertz and Lahrichi [25] propose two different mixed integer programming models for assigning care givers to patients. The objective of both models is to balance nurses' workloads by minimizing a weighted sum of the visit load (based on the weight of each visit), the case load (due to the number of patients assigned) and the travel load (related to the distances traveled) while respecting constraints related to maximum acceptable loads and continuity of care. The travel load is calculated on the basis of the average distance of the patient location from the district where care giver works. Since the estimate does not consider sequencing, it should be accurate for small districts. Borsani *et al.*[12] propose assignment and scheduling models where the output of the assignment model is incorporated as the input to the scheduling model. In this work, the assignment process is held to ensure workload balance among care givers while respecting continuity of care, qualification requirements and geographical coherence constraints. Travel times are constant and independent from the sequence.

An extended modeling framework related to the assignment problem of the HHC services was developed by Lanzarone *et al.* [30], where authors provide different assignment models to balance care givers' workloads by considering several peculiarities of HHC services, such as care givers' skills, the geographical locations of patients and care givers, and the stochastic patient requests.

Travel times are modeled as in [12]. The same problem with stochastic demand is then tackled in the works of Lanzarone and Matta [29,31] who propose simple policies to assign patients to care givers instead of mathematical programming. Carello and Lanzarone [14] develop a cardinality–constrained robust assignment model where their aim is to exploit the potentialities of a mathematical programming formulation and to evaluate the capability of such model in reducing the costs related to nurses' overtimes. Also in this case, the travel time for reaching homes is the same for all patients and care givers. Lastly, Koeleman *et al.* [28] represent the HHC system as a Markov chain and they develop admittance policies for patients with the use of a trunk reservation heuristic to control the system by considering a general visiting time containing a travel load that does not consider routes.

We note that a common characteristic of assignment related papers are the need of balancing the workload among care givers. Indeed, this problem is extremely important to have equal working conditions in the same organization.

The second category of papers focuses on the simultaneous assignment and routing problems. The recent work of Hulshof *et al.* [27] proposes a taxonomic review on planning–related decisions in health care services, including HHC. Papers presented in this review consider the exact distances among patients' homes to calculate the care giver workloads, but this makes the problem significantly more complex. Since this paper is focused on the assignment problem, we only present the list the of these references [2,4,5,7–9,13,15,16,20,21,26, 33–35,39,42,46].

Furthermore, we note that in these works either heuristic methods (e.g., Genetic Algorithm, Tabu Search, etc.) are adopted, or small instance sets are used to solve the developed models [39,40]. Large scale problems have not been solved with exact methods yet.

In conclusion, the current literature on HHC services does not consider patient attributes to estimate travel times when the assignment problem is solved. Contrary to existing works, in this paper, we propose a data driven approach to consider such attributes. More specifically, factors related to patients geographical locations such as traffic conditions to reach them, the accessibility of patients homes, etc. are integrated into a Kernel Regression method to estimate care givers (effective) travel times while solving the assignment model.

3 Data-driven travel time estimation method

The travel time of a care giver in the assignment problem may be affected by several patient related features (attributes) such as the care profile of patients (i.e., pathology, type and intensity of care), the temporal constraints (i.e., availability of the family member that is present for help) and the geographical locations of patients. In this work, we focus solely on features that are related to the geographical locations of patients, but the approach proposed for estimating travel time is generic enough to consider other types of patient attributes (with the ones available in the considered historical data).

We use historical information to integrate routing related considerations in the assignment problem by estimating care givers' past travel times based on patients' geographical locations. Indeed, there are several factors related to patients' geographical locations that have an impact on the travel time of care givers. Examples of such factors are related to daily traffic conditions (i.e., dense or calm), personal preferences of care givers and/or difficulties related to the access to patients' homes.

To illustrate this, consider a real example of a care giver tour which has to visit 7 different patients (identified as A-F) in a particular day. The care giver considers several geographical and physical aspects while planning his/her route. For example, due to high traffic density, (S)he chooses to visit patient A at the end of the working day. Similarly, due to the absence of an elevator, (S)he chooses to visit patient D at the beginning of the working day since (S)he feels more energetic. Thus, according to such personal preferences, (s)he executes a route that is not optimal from a total travel time minimization perspective. For the given case, if (S)he wanted to obtain her route as the optimal one, according to the travel distance (time) minimization, (S)he would need to travel 16.2 km with the following sequence Center-A-B-C-D-E-F-Center. However, since (S)he considers some other features, the observed executed tour length turns out to be 20 km with the sequence of Center-D-C-E-F-B-A-Center. Thus, it can be concluded that in practice planners may not take into account only the criteria on travel distances but also other features while planning visits.

3.1 Selection of the technique

We propose a non-parametric method to estimate travel times from real data observations because of their distribution-free property and the asymptotic convergence of some estimators. Specifically, Kernel Regression (KR) is used to estimate the travel time functions.

KR is a non-parametric regression technique that does not require a predetermined form, as the predictor is built with the information derived from existing data [47]. KR exploits correlations existing among observations by assuming a radial basis function explaining the data. In our context, since HHC patients have unique characteristics depending on their features (i.e., geographical location, care profile, etc.), KR seems to be proper to estimate the travel time to visit a set of patients. Indeed, such data-driven approach is important for HHC services since historical observations would enable to capture what really happenes in the system in terms of executed planning decisions. For instance, for some reason, if a patient has been visited in the first order of the visiting sequence for a certain period of time then it is likely to observe similar behaviors for the following periods as well. Thus, the KR technique would enable to integrate this situation when estimating travel times by assigning a certain weight to that patient for that specific sequence based on the information coming from historical observations. In particular, as far as the management of new patients, KR approach is quite flexible in managing any type of patient independent of his/her status (new or not) in the system. Even if the patient is new in the system, he/she can be located in a place which is close to one of the previously cared patients (available in the historical data) via weights that are assigned by KR. Hence, travel times can be estimated in a more realistic way via the use of KR.

To our knowledge, although such data–driven approaches and the KR technique have been used for problems such as inventory control, call center staffing and dynamic assortment optimization [45], they have not been applied to the HHC setting yet.

There are some advantages to use this technique in HHC services. First, this method uses past data to infer the travel time related to a set of patients having specific attributes. Since the method needs several samples to build its estimators, HHC service fits quite well because it is a periodic and repetitive service type. Thus, a particular patient can be observed several times in the past data and the Kernel estimator gains significance by time. Another advantage is related to districting, which is a priori step involved in the HHC planning problem before the assignment is tackled. The districting process consists of partitioning a territory into smaller areas [6] and such consideration is beneficial for KR to be able to perform with high accuracy even using a lower number of historical information.

The use of Kernel or other regression techniques for larger areas would require a more important volume of historical data not available in practice. Hence, the proposed method is an efficient way to estimate travel times for small regions without requiring large volume of historical information.

3.2 Proposed method

This section describes the method developed for estimating care giver travel times.

Given the set of patients $\mathcal{P} = \{1, \ldots, P\}$ a care giver must visit in a route, we want to estimate the route travel time using past observations. Let $\mathbf{x} = \{x_i, i = 2(p-1) + l, p \in \mathcal{P}; l = 1, 2\}$ be a $2P$-dimension vector containing the geographical locations of the patients in the set $\mathcal{P}$.

Let $\mathcal{S}_P = \{1, \ldots, S\}$ represent the set containing the historical data used for building estimates. $\mathcal{S}_P$ is thus the set of all the routes (or sequences) of length P, i.e. all routes that required P patients to be visited. Let also $\mathbf{y^0} = \{y^0_s, s \in \mathcal{S}_P\}$ denote the travel time of the care giver related to each of the S routes, whereas $\mathbf{x^0} = \{x^0_{si}, i = 2(p-1) + l, s \in \mathcal{S}_P; p \in \mathcal{P}; l = 1, 2\}$ is an $S \times 2P$ array denoting the geographical locations of all the patients visited in the history.

The KR technique estimates the expectation of the outcome Y conditional on the random variable variable X, $\mathbb{E}(Y|X)$. The main reason for using KR is the scarcity of restrictions on the functional relationship between X and the outcome Y. This relationship can be formulated as follows:

$$Y = \tau(X) + \varepsilon$$

where τ is an unknown function, and ε is the error term, which is independent and identically distributed with mean 0 and variance $\sigma^2(X)$.

We consider the case of Multivariate Kernel Regression method because our response variable Y depends on a vector of exogenous variables $\mathbf{X}$. Thus, we aim to estimate the following conditional expectation:

$$\mathbb{E}(Y|\mathbf{X}) = \mathbb{E}(Y|x_1, ..., x_{2P}) = \tau(\mathbf{x}),$$

where $\mathbf{x}$ is the point to be evaluated, i.e. patients in the set $\mathcal{P}$.

To estimate the unknown function, we use the Nadaraya-Watson estimator [47]:

$$\widehat{\tau}(\mathbf{x}) = \frac{\sum_{s \in \mathcal{S}_P} K(\frac{\mathbf{x_s^0} - \mathbf{x}}{\mathbf{h}}) y_s^0}{\sum_{s \in \mathcal{S}_P} K(\frac{\mathbf{x_s^0} - \mathbf{x}}{\mathbf{h}})}, \tag{1}$$

where $K(.)$ is a $2P$ dimensional kernel function, $\mathbf{h}$ is the *bandwidth* vector and $\mathbf{x_s^0} = (x_{s1}^0, \ldots, x_{s,2P}^0)$ contains the geographical locations for each of the P patients in the route s. In the Nadaraya-Watson approach, the function τ is estimated with a locally weighted average by using the Kernel as a weighting function. The selection of the bandwidth value is relevant, as it affects the predictor's smoothness. Several methods are available in the literature to select a value for $\mathbf{h}$ and we use the optimal bandwidth technique suggested by Bowman and Azzalini [11].

The Kernel function, $K(\mathbf{z_s})$, is chosen as the widely applied Gaussian Kernel,

$$K(\mathbf{z_s}) = \prod_{i \in \mathcal{I}} \frac{1}{\sqrt{2\pi}} e^{-z_{si}^2}, \tag{2}$$

where $z_{si} = (\frac{x_{si}^0 - x_i}{h_i})$ and $\mathcal{I} = \{1, \ldots, 2P\}$.

To obtain more accurate estimates using the KR technique, we considered the uniformization of the input data to be used in the KR function. The same patient can occupy different positions (i.e., the rank in the input matrix) in different routes. Thus, if we do not uniformize the input data, the KR function might not recognize that these (multiple) patients refer to the same single patient and may spend unnecessary time providing a better estimate by building a structure across all dimensions (as the double of the number of patients). To avoid this computation, we use a simple ordering technique for the input data that rearranges the patients order according to their geographical locations. This technique sorts the patients by a function of their location coordinates: $x_{i1} + a x_{i2}$ ($i \in \mathcal{P}$ and a is a positive integer).

3.3 Implementation procedure

Historical data values $(\mathbf{x^0}, \mathbf{y^0})$ are used to build the estimator $\hat{\tau}(\mathbf{x})$. Hence, given a set of patients to visit in a day represented by $\mathbf{x}$, we can use the expression in (1) to estimate the time that the care giver needs to reach patients' homes.

Algorithm 1

procedure TRAVEL TIME ESTIMATION
Step 1 Initialization:
  $P \leftarrow$ *number of patients*
  $S \leftarrow$ *number of routes with* P *patients*
  $\mathbf{h} \leftarrow \mathbf{h_0}, a \leftarrow a_0$
Step 2 Uniformization of historical data:
  for $s = 1, \ldots, S$ **do**
    for $p = 1, \ldots, P$ **do**
      <Calculate $OV_{sp} = x^0_{s,2(p-1)+1} + ax^0_{s,2(p-1)+2}$>
    <Sort patients in descending order of OV_{sp}>
Step 3 Uniformization of actual patients:
  for $p = 1, \ldots, P$ **do**
    <Calculate $OV_p = x_{2(p-1)+1} + ax_{2(p-1)+2}$>
  <Sort patients in descending order of OV_p>
Step 4 Calculation of weights:
  for $s = 1, \ldots, S$ **do**
    <Calculate $w_s = K(\frac{\mathbf{x_s^0} - \mathbf{x}}{\mathbf{h}})$ using equation (2)>
  <Calculate $\hat{\tau} = \frac{\sum_{s=1}^{S} w_s y_s^0}{\sum_{s=1}^{S} w_s}$>

Algorithm 1 represents the implementation procedure of the KR technique. Initially, some parameters are configured: P is set equal to the number of patients in the route under consideration, the coefficient a and the bandwidth $\mathbf{h}$ values are calculated after analysis of historical data or are estimated based on experience. The next step (step 2) is to rearrange the order of patients in each historical route by applying uniformization (step 2). The same is done on the patients of the route under consideration (step 3). Step 4 establishes the weight w_s for each sequence and use them in combination with the observed travel times y_s^0 for calculating the estimate $\hat{\tau}$.

The procedure can be customized by using only the historical sequences related to a specific care giver. Notice that $\hat{\tau}$ depends on the number of patients in the routes, thus an estimator has to be built for each route length observed, e.g. routes involving 5 patients, 6 patients and so on.

Assessing the performance of the proposed method is not easy mainly because of two reasons. The first reason is related to the effectiveness of the method. Indeed, it is more important to quantify at which degree the method contributes to produce better assignments, rather than assessing its accuracy in estimating the travel times related to some sets of patients. The second reason is related to the problem of judging the quality of an assignment. To do this, we build optimal routings after having solved the assignment problem. In this way, we can finally evaluate if the proposed method used at assignment level can bring benefits in terms of effective (i.e., after routing) workload balancing and total travel times. This assessment is the objective in the next two sections. Section 4 describes the models used in the assessment of the effectiveness of the proposed method, whereas section 5 reports the numerical results.

4 Planning approaches

In this section, we describe the models used to assess the effectiveness of the proposed KR method. Hence, the Kernel estimator is used in a two–stage approach, which sequentially solves the assignment and routing problems. Results of the two–stage approach are compared with those obtained from a simultaneous approach, which is theoretically the most accurate one to solve assignment and routing (see also Figure 1). The purpose is to demonstrate that the use of the proposed method leads to results as good as the simultaneous approach. We voluntarily consider a simple HHC setting that operates with the assumptions described below. These assumptions are valid for both the two–stage and simultaneous approaches.

4.1 Problem definition

This section describes the specific assignment and routing problems by defining the assumptions and the objective function used in the two–stage and simultaneous approaches.

Patients in the set $\mathcal{A} = \{1, \ldots, N\}$ have to be assigned to care givers and a route plan for each care giver has to be designed. Each patient $i \in \mathcal{A}$ is assumed to have a deterministic demand λ_i (expressed as a service time), which denotes the total amount of care volume (e.g., expressed in minutes) the patient requires on a single day. At the HHC service provider we worked with, this parameter is often considered a standard value and, without loss of generality, has the same value for homogeneous patients: Note that the last assumption is not strictly necessary for the models we develop in this paper. Each visit requires only one care giver and does not have precise time windows to be respected. Synchronous visits are not considered. Finally, each patient is visited once a day.

We consider a single category of care givers in the set $\mathcal{B} = \{1, \ldots, |\mathcal{B}|\}$ and assume that all care givers have identical skills. In practice, HHC operators are often divided into several districts (i.e., groups) based on their main skills and the geographical areas they serve. Since we are interested in assignment and routing decisions, we assume that the districts have been defined earlier.

Each care giver $k \in \mathcal{B}$, is assumed to have a capacity a_k corresponding to the maximum amount of time that (s)he works according to his (her) contract.

We also assume that, beyond this capacity, care givers can handle an excess load, i.e. overtime is allowed for care givers, although possible overtime costs are not explicitly integrated into the models. A patient can be assigned to only one care giver in the set of existing care givers.

HHC providers using the two–stage approach often solve the assignment problem over a larger horizon than the routing problem, e.g. assignments are done for a whole week while routes are defined for each day. If a simultaneous approach is followed, the problem is solved over smaller horizons, often on a single day. While comparing the two approaches, in order to have identical planning periods, we consider a planning period of one day for both approaches.

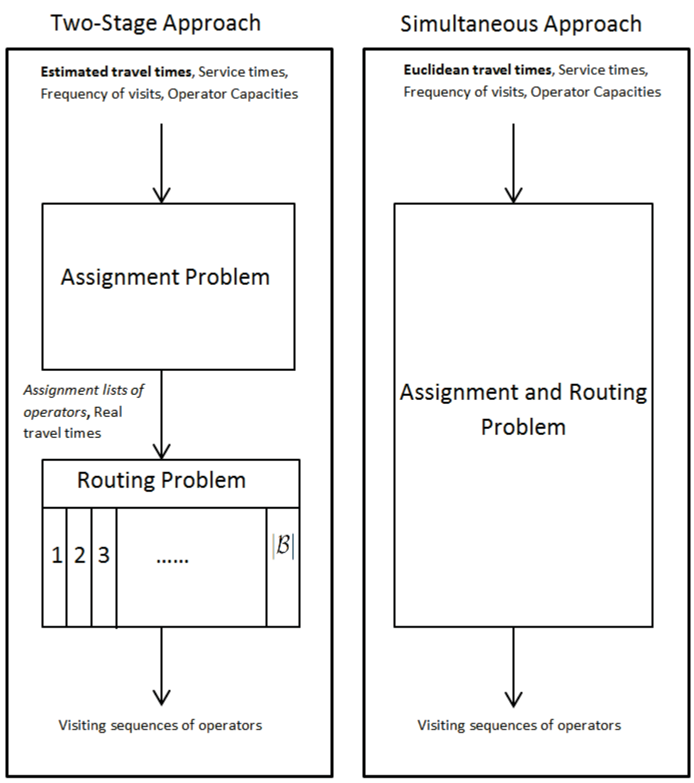


Fig. 1 Two-Stage Approach vs. Simultaneous Approach

The objective is to balance the utilization rates of each care giver (defined as the ratio between the actual workload of the care giver and his (her) capacity) and to minimize the total traveling times of care givers. Workload balancing is important to ensure that the workload dispersion across all care givers is as equal as possible. Minimizing the total travel time is also important for employing care givers more efficiently (i.e., to serve all patients and/or to perform other coordination activities in the common HHC center when care givers do not perform visits) and to reduce the corresponding traveling cost.

4.2 Two–stage approach

The assignment problem is formulated as follows:

$$\min \quad Z = h + \gamma \sum_{k \in \mathcal{B}} y_k \tag{3}$$

$$\text{s.t.} \sum_{k \in \mathcal{B}} x_{ik} = 1 \quad \forall i \in \mathcal{A} \tag{4}$$

$$y_k = \hat{\tau_k} \quad \forall k \in \mathcal{B} \tag{5}$$

$$w_k = \sum_{i \in \mathcal{A}} \lambda_i x_{ik} + y_k \quad \forall k \in \mathcal{B} \tag{6}$$

$$h \geq \frac{w_k}{a_k} \quad \forall k \in \mathcal{B} \tag{7}$$

$$x_{ik} \in \{0,1\} \quad \forall i \in \mathcal{A}, \forall k \in \mathcal{B} \tag{8}$$

$$w_k, y_k \geq 0 \quad \forall k \in \mathcal{B} \tag{9}$$

Decision variable x_{ik} equals 1 if patient i is assigned to care giver k and 0 otherwise. w_k is a continuous variable used to calculate the workload of care giver k. y_k denotes the travel time of care giver k, and γ is a penalty parameter used to balance the tradeoff between the total travel time and workload balancing. The auxiliary variable h is used to estimate the maximum utilization rate among all care givers.

Equation (4) implies that any patient must be assigned to only one care giver. Equation (5) assigns to y_k the estimation of the travel time. This estimation is done with Algorithm 1. It is worthwhile to point out that $\hat{\tau_k}$ is a function of the set of patients assigned to care giver k, thus in general $\hat{\tau_k} = f(x_{1k}, \ldots, x_{Nk})$. Equation (6) defines the workload of each care giver k. Inequality (7) expresses the maximum utilization rate h, which is minimized in the objective function (3) together with the penalized sum of travel times.

At the routing level, a TSP model is used to create each care giver's route for a single day. Hence, with patient lists obtained from the assignment step, $|\mathcal{B}|$ independent TSP models are solved, and each care giver's visiting sequence is determined. In other words, the assignment model's output is incorporated into the routing problem, and the routes of $|\mathcal{B}|$ care givers are obtained by solving a set of $|\mathcal{B}|$ TSP models. We use a classical TSP model [38] with travel time minimization objective where deterministic travel times are obtained as Euclidean distances.

4.3 Simultaneous approach

As described before, the simultaneous approach is used to make the assignment and routing decisions at the same time. It should be noted that, the routing part of the HHC problem is actually a VRP with HHC related constraints, and we do not intend to have a significant contribution to the VRP literature. Rather, we use this problem as a benchmark to be able to analyze the performance of the two–stage approach using the proposed Kernel estimator. The only adaptation that is done here is the objective function. Hence, we formulate the simultaneous approach as a VRP with the objective function that balances the trade-off between workload balancing and total travel time minimization.

Note that in the basic VRP, capacity restriction of care givers are imposed as hard constraint. However, in this work we used a variant of VRP model in which the capacity constraints are not considered. Such a model is named as mTSP [3]. For simplicity, we name this model as VRP throughout this work and it is presented as follows:

$$\min \quad Z = h + \gamma \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{A}} t_{ij} \sum_{k\in\mathcal{B}} x_{ijk} \tag{10}$$

$$\text{s.t.} \sum_{i\in\mathcal{A}} \sum_{k\in\mathcal{B}} x_{ijk} = 1 \quad \forall j \in \mathcal{A} \tag{11}$$

$$\sum_{i\in\mathcal{A}} x_{ipk} = \sum_{j\in\mathcal{A}} x_{pjk} \quad \forall p \in \mathcal{A}, \forall k \in \mathcal{B} \tag{12}$$

$$\sum_{j\in\mathcal{A}} x_{1jk} = 1 \quad \forall k \in \mathcal{B} \tag{13}$$

$$u_i - u_j + 1 \leq N + 1(1 - \sum_{k\in\mathcal{B}} x_{ijk}) \quad \forall i, j \in \mathcal{A}, i \neq j \neq 1 \tag{14}$$

$$w_k = \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{A}} x_{ijk} t_{ij} + \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{A}} x_{ijk} \lambda_j \quad \forall k \in \mathcal{B} \tag{15}$$

$$h \geq \frac{w_k}{a_k} \quad \forall k \in \mathcal{B} \tag{16}$$

$$x_{ijk} \in \{0, 1\} \quad \forall i, j \in \mathcal{A}, \forall k \in \mathcal{B} \tag{17}$$

$$w_k \geq 0 \quad \forall k \in \mathcal{B} \tag{18}$$

Different than the previously presented formulation, here the decision variable x_{ik} is modified as x_{ijk} in order to identify the visiting sequences of patients and takes the value 1 if care giver k visits the patient j immediately after patient i and 0 otherwise. In particular, the continuous decision variable u_i is used to indicate the sequence in which patient i is visited ($i \neq 1$) and t_{ij} is used as the traveling time between patient i and j (i.e., calculated as the Euclidean distances).

Equation (11) states that each patient should be visited exactly once. Equation (12) is the flow conservation constraints, which ensures that once a care giver visits the patient, then he must also depart from this patient. Equation (13) ensures that each care giver serves exactly once and Equation (14) is the sub-tour elimination constraint. The total workload of each care giver k is defined by the Equation (15). Inequality (16) calculates the maximum utilization rate h, which is minimized in the objective function (10) together with the penalized total travel time of care givers.

4.4 Solution approach used

In this work, we use a solution method based on Genetic Algorithm (GA) to solve models presented in sections 4.2 and 4.3 especially with large instances. GAs are coded in Matlab R2014b and run on a 2.2GHz processor.
Although it is possible to use CPLEX for both the two–stage and simultaneous approaches, due to the computational complexity encountered in the simultaneous approach especially for large instances, GA is preferred to be able to obtain results in a reasonable time.
More specifically, there are two main reasons why we choose to implement GA for the two–stage approach with KR. The first one is that the KR function used in the assignment problem is not unique, but there is one function for each care giver and for each possible group size of assignable patients. Implementing this in a GA method is much easier rather than in mathematical programming . The second reason is the non-linear property of the Kernel function.

Description of the GA used: GA is initialized with the randomly generated population (chromosome) matrix. Then, for each chromosome, the fitness values (the objective function value) are obtained. A predetermined number of chromosomes are then randomly selected, and two individual with minimum fitness values are selected. With the chosen chromosomes, basic GA operations are performed (mutation and crossover) to populate the next generation (i.e., children). The procedure is repeated for several iterations until the exit condition (maximum number of iterations) is satisfied .

How to call the travel time function: With respect to the KR estimation, because the regression function is fitted to calculate directly the travel time of a care giver, the incorporation of this estimation in the assignment problem is complex. As mentioned before, GA is adopted to be able to cope with such complexities. Indeed, it is not difficult to embed the KR function into this heuristic approach since, at each iteration of the GA, the assignment list of patients is known for each care giver. Thus, direct computation of travel times for each care giver can be completed using the generated KR function (each care giver has its own set of KR functions, one for each number of assignable patients). Then, the algorithm can proceed for the next step where the fitness value is obtained.

More details related to the GAs (including some accuracy analysis) are given in Appendix A.1, A.2 and A.3 respectively.

5 Numerical Results

In this section, we present results of the comparison between the two–stage approach using KR estimator and the simultaneous approach. First, we present the details of the available data and the generation of instances. Then, we analyze small problem instances and extend the analysis to medium as well as large size problems. Furthermore, the two–stage approach is executed with two other estimation methods existing in the literature to numerically quantify the added value of using the Kernel estimator instead of others. Lastly, a sensitivity analysis is carried out.

5.1 Available data and instance generation

Two groups of data sets are used to conduct the experiments in this work. The first one is real data that is provided by an Italian HHC provider whereas the second one is the benchmark data that is used in the VRP literature [23, 17].

GA is used to be able to solve the realistic instances (i.e., 150 patients and 15 care givers). Since such heuristic approach does not guarantee optimality and it is also not possible to solve such instances with the CPLEX solver, smaller instances (up to 25 patients and 8 care givers) are also generated to compare the two-stage approach with the simultaneous approach under the light of optimality. For this case, benchmark data files that are available from the VRP literature are used to generate other instances to analyze the performance with different problem settings (i.e., a higher number of cities and larger distances between them). More details on these data sets are provided in the following subsections.

5.1.1 Data available from real case

Instances used in the experiments pertaining Group A are obtained from the analysis of a real case from which we collected the type and number of patients that have received service in the analyzed period (six days), the number of visits and the days of execution per each patient, the standard service time per visit (45 minutes) and the city in which each patient is located. From this data, it is observed that the area served by the HHC provider is composed of closely located 7 main cities (due to priori districting process) where each patient is living in (see Table 1 for distances between cities). Although the patient's detailed address is not available, we know that patients are closely located to each other within the cities.

5.1.2 Generation of real instances

Since data available from the real case does not include the address and the visiting sequences of patients, we have randomly generated patient locations and have optimally calculated the travel time for reaching them.

Table 1 Distances among cities for real data (in km)

	1	**2**	**3**	**4**	**5**	**6**	**7**
1	0	3.3	5.5	5.3	8.0	7.5	5.5
2		0	4.6	5.8	9.8	9.1	7.3
3			0	2.3	6.6	5.9	4.7
4				0	4.4	3.6	2.5
5					0	0.7	2.6
6						0	2.0
7							0

Patients, and the cities where they live, are generated according to the probabilities reported on Table 2. Then, the exact location of each patient is randomly sampled in the selected cities (using the cities' geographical coordinates) from a normal distribution with mean and standard deviation equal to 1 and 0.5 km, respectively (see Figure 2).
The minimum care giver travel time is calculated by solving a travel salesman problem (TSP). This travel time is considered in the experiment as the *historical value* on which the Kernel estimator will be built.

The procedure is repeated as the number of samples we want to generate. At the end of the generation process, the historical data set contains the locations $\mathbf{x^0}$ and the travel times $\mathbf{y^0}$ for each *observed* sample. Algorithm 2 explains in detail the data generation procedure.

Algorithm 2

procedure GENERATION OF HISTORICAL DATA
  $S \leftarrow$ *number of samples*
  $P \leftarrow$ *number of patients in a route*
  for $s = 1, \ldots, S$ **do**
    for $p = 1, \ldots, P$ **do**
      <Randomly sample the city from discrete distribution on Table 2>
      for $l = 1, 2$ **do**
        <Randomly sample z_l from distribution $N(1, 0.5)$>
        <Calculate $x^0_{s,2(p-1)+l} = x_{city,l} + z_l$>
    <Calculate y^0_s solving a TSP problem>

5.1.3 Data available from the VRP literature and instance generation

The data sets that are extracted from the VRP literature[1] and considered in this work are as follows [23,17]: E16-3m, E16-5m, E21-4m, E23-5s and E26-8m where the first two digits indicate the number of customers (patients) and the digit after "-" is used to identify the number of care givers. For simplicity, we refer to these instances as B.1-B.5 in the upcoming sections.

[1] *http://www.or.deis.unibo.it/research_pages/ORinstances/VRPLIB/Simmetric_CVRP.zip*

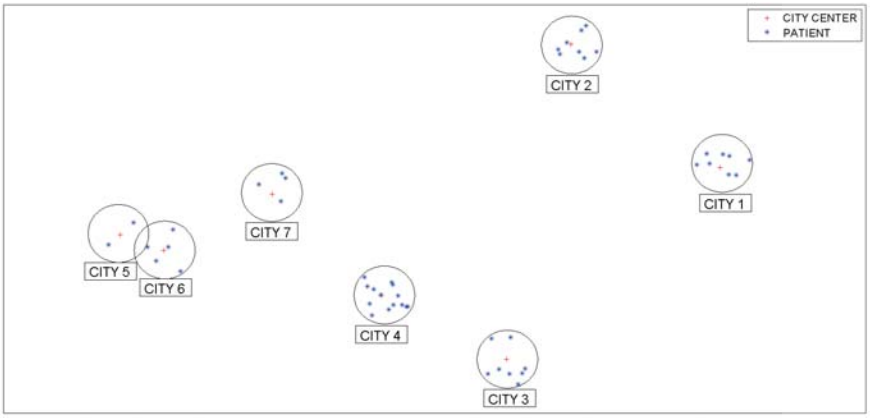


Fig. 2 Map of cities and patient locations in instance 1. Circles represent the area where, according to distribution $N(1, 0.5)$, 95% of the patients are randomly generated

Table 2 Location and probabilities of cities for real data

City	$x_{city,1}$	$x_{city,2}$	Probability
1	49.21	42.32	0.1544
2	51.63	40.01	0.1611
3	48.16	36.92	0.1544
4	46.14	38.02	0.2819
5	41.90	39.05	0.0604
6	42.60	38.79	0.1007
7	44.36	39.75	0.0872

These instances include the number of customers (patients), the number of care givers, customer locations and demand information. Note that, since a variant of VRP is adopted in our research (i.e., mTSP), and a trade-off objective is used instead of a travel time minimization criterion, solutions provided in the literature with these data sets are not comparable with the ones that are presented in this numerical analysis.

Different than the real case data, here the number of available points in the corresponding data set (i.e., 16 for E16-3m or 26 for E26-8m) are considered as the number of main cities (instead of having 7 cities as in section 5.1.2). In addition to that, the distances between these cities are also larger than the real data (see Table 1) where the maximum distance difference between these cities are approximately 21.68 km for E16-3m and E16-5m, 62.94 km for E021-4m, 85.63 km for E26-8m, and 145.17 km for E23-5s.

Historical data is generated with a similar process as it is done for the real case (see section 5.1.2). Similar to the real case, it is also assumed that the locations of patients are close to each other and to the city center they belong to. The only difference is that patients, and cities where they live, are generated with equal probabilities instead of the ones shown in Table 2. The rest of the generation process is applied as is it is explained in section 5.1.2.

5.2 Test on small size instances

In this part, results are presented with the use of small instances with 15 to 25 patients (plus the HHC center) and 3 to 8 care givers. Since the instances are small, CPLEX Solver is able to provide optimal solutions for the simultaneous approach which represents the benchmark solution for analyzing the two-stage approach with KR method. Thus, in Table 3 optimal solutions of the simultaneous approach (denoted as Simultaneous(CPLEX)) are compared with the ones of the two-stage approach that are obtained with the use of the proposed GA. In addition to these solutions, results of the simultaneous approach that are obtained with GA (Simultaneous(GA)) are also presented to be able to analyze the performance of the GA. More performance analysis on the GAs is presented in Appendix A.3.

The dimensions (i.e., number of patients and care givers) of Group B instances are already explained in Section 5.1.3. This information is used to generate Group A (i.e., Instances A.1-A.5) instances of Table 3 where dimensions of each instance belonging to this group is the same as the one of Group B according to the number following the letter (i.e., Instance A.3 has the same dimensions of Instance B.3 etc).

In Table 3, the total travel time obtained in the two–stage approach (calculated over all care givers) with the KR method is shown by T(KR), and the workload balance value between the maximally and minimally utilized care givers is shown by B(KR). Similarly, the total travel time in the simultaneous approach is denoted by T(VRP), and the balancing value is denoted by B(VRP). Because models are solved to balance the trade–off between care giver workload balancing and care givers' total travel times, the corresponding value is denoted as $Z_{(.)}$, which equals $h(.) + \gamma T(.)$ where $h(.)$ corresponds to the maximum care giver utilization level.

The presented T(KR) values are obtained by solving several (as the number of care givers) independent TSP models, with the outputs obtained from the assignment stage and summing the result of each TSP model across all care givers. T(VRP) values are directly calculated from the corresponding simultaneous model as the sum of each care giver's route time.

For the cases where we use GA, all results (i.e., T(.) and B(.) values) are obtained as the average values from replications of the algorithms, and corresponding $Z_{(.)}$ values are calculated based on a 95% confidence interval. To keep the computational effort of the experiments low, we executed 10 independent replications. In obtaining the results, the size of the GA population is set equal to 100, the number of the GA iterations is set to 1000, the probability of crossover (p_c) is set to 0.9 and the probability of mutation (p_m) is set equal to 0.1. Furthermore, the Kernel method estimates travel times using a history with length $S = 100$. Since the historical data used in the experiments consider only travel time considerations, all care givers have the same behavior. Thus, the same Kernel function is used independently from the care giver. All these settings are used for the entire numerical analysis.

Table 3 Results with small instances and $\gamma = 1/500$

Instance	**Two-Stage KR**			**Simultaneous(GA)**			**Simultaneous(CPLEX)**		
	T(KR)	B(KR)	Z_{KR}	T(VRP)	B(VRP)	Z_{VRP}	T(VRP)	B(VRP)	Z_{VRP}
A.1	37.59	0.021	0.581 ± 0.0000	37.59	0.021	0.581 ± 0.0000	37.59	0.021	0.581
A.2	68.71	0.022	0.742 ± 0.0007	68.57	0.023	0.742 ± 0.0001	68.55	0.022	0.742
A.3	55.44	0.013	0.647 ± 0.0011	54.84	0.013	0.646 ± 0.0006	54.64	0.013	0.646
A.4	68.23	0.016	0.573 ± 0.0041	63.62	0.018	0.563 ± 0.0000	63.62	0.018	0.563
A.5	99.43	0.019	0.535 ± 0.0016	98.61	0.019	0.533 ± 0.0003	98.32	0.019	0.533
B.1	284.92	0.067	1.308± 0.0298	261.14	0.057	1.250 ± 0.0050	260.22	0.061	1.249
B.2	332.43	0.053	1.435± 0.0235	322.22	0.055	1.402 ± 0.0058	320.85	0.059	1.399
B.3	383.41	0.042	1.503± 0.0465	354.43	0.045	1.426 ± 0.0107	352.27	0.049	1.420
B.4	678.20	0.177	2.068± 0.0534	626.62	0.157	1.932 ± 0.0096	613.55	0.160	1.906
B.5	605.28	0.122	1.981± 0.0394	582.58	0.134	1.917 ± 0.0172	573.63	0.149	1.896

As Table 3 shows, results obtained by the two-stage method are close to the optimal results of the simultaneous approach (see Simultaneous(CPLEX)). The differences are approximately 2% (average of 5 instances) and 7.6% for the Group A and Group B instances respectively when total travel time values ($T(.)$) are considered. Thus, the two-stage method provides better quality solutions for the real case (Group A) where a more limited number of closely located cities is considered. However, a difference of 7.6% is not very high and can be reduced by increasing the length of the history (S) as it will be shown in the sensitivity analysis section.

Another observation concerns the performance of the GA. As it can be seen, results of the simultaneous approach that are obtained with the GA are almost equal to the optimal ones. For both groups of instances, the differences for the total travel time are less than or equal to 1% on average.

In the following sections, results are provided with the medium and large instances of real data (Group A) to analyze the behavior of the two-stage approach with KR method. Due to complexity issues, GA is used to perform these experiments.

5.3 Test on medium and large size instances

The first part of results presented in Table 4 are obtained with 5 different medium size instances (i.e., Instances A.6-A.10) with 56 patients and 7 care givers. Whereas in the second part, the results are reported with five different large size instances (i.e., Instances A.11-A.15) that are composed of 150 patients and 15 care givers. For both cases the first instances are taken from real data (i.e. A.6 and A.11) whereas 8 additional instances are generated.

As Table 4 indicates, results obtained by the two–stage method with medium size instances are close to the results of the simultaneous approach. The differences are approximately 6.9% (i.e., average of 5 instances) for the total travel time. These differences are acceptable but still need to be improved by increasing the length of the historical information.

We also executed these experiments without uniformizing data as in Algorithm 1. Hence, the gap without using data uniformization was 23.05% for the total travel time, this proves the need of data rearrangement before the predictor is calculated.

Table 4 Results with medium and large instances and $\gamma = 1/1000$

Instance	Two-Stage KR			Simultaneous (GA)		
	T(KR)	B(KR)	Z_{KR}	T(VRP)	B(VRP)	Z_{VRP}
A.6	70.86	0.020	0.851 ± 0.0031	66.05	0.022	0.845 ± 0.0014
A.7	81.50	0.017	0.865 ± 0.0058	77.83	0.010	0.857 ± 0.0026
A.8	79.07	0.018	0.862 ± 0.0044	74.52	0.014	0.854 ± 0.0016
A.9	73.13	0.024	0.857 ± 0.0052	66.17	0.021	0.845 ± 0.0016
A.10	72.19	0.022	0.853 ± 0.0035	68.20	0.025	0.848 ± 0.0018
A.11	214.52	0.028	1.192 ± 0.0090	181.57	0.029	1.157 ± 0.0064
A.12	206.95	0.026	1.184 ± 0.0098	165.34	0.027	1.138 ± 0.0065
A.13	207.51	0.026	1.183 ± 0.0105	166.90	0.029	1.140 ± 0.0085
A.14	207.71	0.022	1.183 ± 0.0105	162.87	0.026	1.134 ± 0.0049
A.15	209.35	0.025	1.185 ± 0.0075	171.35	0.028	1.144 ± 0.0079

For larger instances, although the difference between the objective function values of the two–stage approach and the simultaneous approach are comparable (approximately 3.75%), the difference between the total travel times are quite larger (approximately 23.47%). In section 5.5, this will be improved by increasing the number of historical data.

5.4 Comparison with other estimators

To be able to analyze the performance of the proposed KR estimator, we compare the results of two–stage method with KR technique with two other methods. The first method is the simple Average Value (AV) approach and the other one is another regression method based on k-Nearest Neighbor(kNN) search.

In the AV approach, the estimation of the travel time required to visit a particular patient is calculated as the average travel time to reach him (her) from all other patients [48].

In the kNN method only the local neighborhood is used to obtain the prediction for each care giver based on past observations. This estimator is presented with the same notation provided for KR in section 3.2 as follows:

$$\hat{\tau}_k(\mathbf{x}) = \frac{1}{m} \sum_{\forall \mathbf{s}:\mathbf{x_s^0} \in \mathcal{M}_m(\mathbf{x})} y_s^0 \tag{19}$$

where $\mathcal{M}_m(\mathbf{x}) \subset \mathcal{M}$ is the neighborhood of $\mathbf{x}$ defined as the m closest points $\mathbf{x_s^0}$ in the history set. Closest points are identified according to the Euclidean distances identified in $\mathbb{R}^{2P}$(i.e. $\mathbf{d}(\mathbf{x}, \mathbf{x_s^0}) = \|\mathbf{x} - \mathbf{x_s^0}\|$) with the following order statistics $0 \leq d_1 \leq d_2 \leq \ldots \leq d_m \leq \ldots \leq d_S$; the m closest ones are defined as the nearest neighborhood of $\mathbf{x}$, where the parameter m is selected with the method presented in the work of Gyorfi [24].

These two other travel time estimation methods are used in expression (5) of the assignment model formulated in section 4.1. Results are presented in Table 5 with the medium and large instance sets. Similar to previous tables, care givers' total travel time obtained in the two–stage approach with the AV and kNN methods are shown by T(AV) and T(kNN), respectively, and the workload balance value between the maximally and minimally utilized care givers is shown by B(AV) and B(kNN), respectively. These results have to be compared with those in Tables 4.

Table 5 Results with with medium and large instances, $\gamma = 1/1000$ with different travel time estimators

Instance	**Two-Stage AV**			**Two-Stage kNN**		
	T(AV)	B(AV)	Z_{AV}	T(kNN)	B(kNN)	Z_{kNN}
A.6	143.88	0.017	0.943	73.89	0.027	0.857 ± 0.0078
A.7	157.67	0.006	0.958	87.40	0.013	0.870 ± 0.0186
A.8	158.24	0.003	0.956	82.83	0.018	0.867 ± 0.0089
A.9	118.97	0.030	0.920	82.91	0.022	0.867 ± 0.0164
A.10	120.23	0.029	0.920	80.04	0.021	0.863 ± 0.0071
A.11	335.73	0.012	1.324	215.56	0.034	1.199 ± 0.0320
A.12	321.61	0.015	1.310	216.16	0.030	1.198 ± 0.0251
A.13	329.28	0.011	1.317	214.10	0.029	1.193 ± 0.0267
A.14	326.21	0.011	1.313	208.41	0.025	1.184 ± 0.0165
A.15	321.90	0.012	1.309	221.02	0.026	1.199 ± 0.0155

As it can be seen from the table, the KR method outperforms both the AV and kNN methods for medium and large instances. Although, results of KR and kNN methods seem closer for large instances, the solution times for the kNN are very long (1-3 hours) with respect to the KR method (from 2-4 minutes). In particular, it can also be observed that results with data–driven approaches (i.e KR and kNN) are performing much more better than the AV approach.

Note that the results of the the two–stage approach with the AV technique are obtained with the ILOG Cplex 12.3 solver, whereas the results corresponding to the kNN technique are obtained with GA as we do for the KR method.

5.5 Sensitivity analysis

Results obtained from increasing the number of historical data (for $S = 100$ and $S = 1000$) for all the instances are presented in Table 6. As the number of historical data increases, the performance of the two–stage approach with KR technique significantly increases. For instance, with the small instances the differences are decreased to 0.81% for the Group A and 2.70% for the Group B instances. In particular, it is also observed that the differences with medium and large instances are decreased to 2.89% and 3.08% respectively.

When the solutions of these tables are compared with the solutions of the simultaneous approach, we can conclude that with enough number of historical

Table 6 Results with with different sample size of history (from 100 to 1000)

Instance	γ	S	Two-Stage KR T(KR)	B(KR)	Z_{KR}
A.1	1/500	100	37.59	0.021	0.581 ± 0.0000
		1000	37.59	0.021	0.581 ± 0.0000
A.2		100	68.71	0.022	0.742 ± 0.0007
		1000	68.61	0.022	0.742 ± 0.0001
A.3		100	55.44	0.013	0.647 ± 0.0011
		1000	55.16	0.013	0.647 ± 0.0011
A.4		100	68.23	0.016	0.573 ± 0.0041
		1000	65.16	0.019	0.566 ± 0.0030
A.5		100	99.43	0.019	0.535 ± 0.0016
		1000	98.90	0.019	0.534 ± 0.0008
B.1		100	284.92	0.067	1.308 ± 0.0298
		1000	270.39	0.049	1.266 ± 0.0290
B.2		100	332.43	0.053	1.435 ± 0.0235
		1000	320.75	0.067	1.401 ± 0.0041
B.3		100	383.41	0.042	1.503 ± 0.0465
		1000	366.55	0.031	1.456 ± 0.0073
B.4		100	678.20	0.177	2.068 ± 0.0534
		1000	630.34	0.153	1.939 ± 0.0128
B.5		100	605.28	0.122	1.981 ± 0.0394
		1000	589.59	0.119	1.931 ± 0.0380
A.6	1/1000	100	70.86	0.020	0.851 ± 0.0031
		1000	68.16	0.023	0.848 ± 0.0035
A.7		100	81.50	0.017	0.865 ± 0.0058
		1000	79.35	0.015	0.860 ± 0.0061
A.8		100	79.07	0.018	0.862 ± 0.0044
		1000	76.79	0.017	0.859 ± 0.0041
A.9		100	73.13	0.024	0.857 ± 0.0052
		1000	68.70	0.025	0.852 ± 0.0040
A.10		100	72.19	0.022	0.853 ± 0.0035
		1000	69.84	0.027	0.852 ± 0.0043
A.11		100	214.52	0.028	1.192 ± 0.0090
		1000	176.15	0.026	1.149 ± 0.0043
A.12		100	206.95	0.026	1.184 ± 0.0098
		1000	166.46	0.028	1.140 ± 0.0094
A.13		100	207.51	0.026	1.183 ± 0.0105
		1000	172.44	0.029	1.146 ± 0.0075
A.14		100	207.71	0.022	1.183 ± 0.0105
		1000	173.01	0.026	1.147 ± 0.0065
A.15		100	209.35	0.025	1.185 ± 0.0075
		1000	175.07	0.028	1.148 ± 0.0102

points, the two-stage approach with KR method presents similar solutions in comparison to the simultaneous approach even for large instances. Note also that $S = 1000$ is not unrealistic, since a care giver usually works 250 days per year.

Results obtained from increasing the number of patients are presented (while keeping the same number of care givers 7) in Table 7. As it can be seen, the performance of the KR function starts to decrease in terms of total travel time. This is due to the increase of dimensionality of the Kernel function's argument. For example, in the instances with $N = 105$, from 8 to 15 patients are assigned to care givers and the estimator loses in accuracy.

Table 7 Results with different number of patients per care giver based on instance A.6

N	Two-Stage AV			Two-Stage KR			Simultaneous (GA)		
	T(AV)	B(AV)	Z_{AV}	T(KR)	B(KR)	Z_{KR}	T(VRP)	B(VRP)	Z_{VRP}
56	143.88	0.017	0.943	68.16	0.023	0.848 ± 0.0035	66.05	0.022	0.845 ± 0.0014
70	154.61	0.008	1.143	74.98	0.021	1.046 ± 0.0048	69.49	0.023	1.037 ± 0.0023
105	155.02	0.018	1.611	107.29	0.016	1.553 ± 0.0088	87.27	0.021	1.529 ± 0.0060

Lastly, for simplicity, only two different penalty values are used while obtaining the provided results of this section, $\gamma = 1/500$ for small and $\gamma = 1/1000$ for medium and large instances. To be able to show that the two–stage approach using KR estimation is also consistent with other penalty values, we plot Figure 3 using data from the first medium instance (see Instance A.6). With this figure, we show the trade-off between the workload balancing and total travel times of care givers for decreasing values of the penalty term; results refer to the assignment phase of the two-stage process. As expected, it appears that when we decrease the penalty value, the effect of the total travel time decreases while better workload balancing is ensured. Thus, the KR predictor seems to correctly guide the assignment problem in different situations.

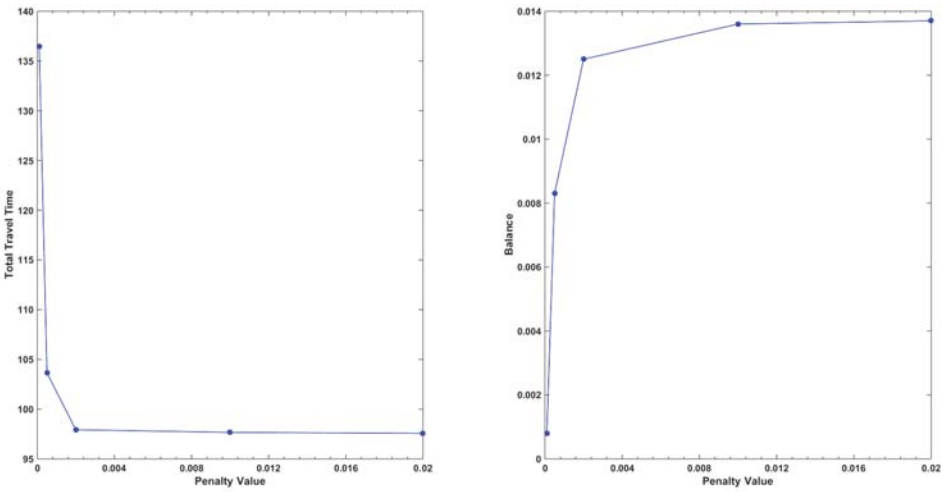


Fig. 3 The trade-off between the workload balancing and the total travel time minimization for the assignment phase of the two-stage model

6 Conclusion

In this work, we have proposed a Kernel regression method to estimate travel times in the HHC assignment problem using the historical routing information of care givers. We have analyzed the performance of the proposed estimator and showed the improvements achieved in comparison with the AV and data-driven kNN regression techniques. Then, we have compared the performance of the two–stage and simultaneous approaches. We have found that the two–stage approach with KR method provides results similar to those of the simultaneous approach.

The use of the KR technique is promising for HHC organizations where the number of patients and care givers can be significant and the assignment and routing problems have different time scales, e.g. care givers' assignment lists are gathered weekly, and routes are defined daily.

Furthermore, because the solutions of the two–stage approach are comparable with the simultaneous approach, this process seems to be a promising tool for approximately solving the HHC VRP. However, this approach should be tested on further experiments and more realistic planning settings including time windows constraints, synchronization of visits and limited care givers' availability. Indeed, this work is limited to a basic application of the HHC services where we do not consider care giver capacities or different qualifications. Thus, one important extension will be to consider multiple planning horizons with different care giver skills and capacities to be able to analyze this approach in more realistic situations.

Another relevant extension of this work would be the consideration of other HHC features since the actual paper focuses on the geographical locations of patients that actually might not be the only criteria for defining care givers' visits. Indeed, other features such as patients' care profiles (i.e., corresponding pathology), special service requests (i.e., requests for clinical tests) and temporal constraints (i.e., requests for visits at specific times) would impact the real care giver behaviors and therefore the associated travel times.

Lastly, this work can also be extended by considering differently structured historical data that is not gathered by solving multiple TSPs but instead obtained via different methods (i.e., using nearest neighborhood search etc.) or directly from real world data.

References

1. Akjiratikarl C., Yenradee P., Drake P.R.: PSO-based algorithm for home care worker scheduling in the UK, Computers and Industrial engineering, **53**, 559–583, (2007).
2. Begur S.V., Miller D.M., Weaver J.R.: An integrated spatial Decision Support System for scheduling and routing home health care nurses, Interfaces, **27**, 35–48, (1997).
3. Bektas T.: The multiple traveling salesman problem: an overview of formulations and solution procedures, Omega, **34**, 209–219, (2006).
4. Ben Bachouch R., Fakhfakh M., Guinet A., Hajri-Gabouj S.: Planification de la tournee des infirmiers dans une structure de soins a domicile, Conference Francophone en Gestion et Ingenierie des Systemes Hospitaliers (GISEH), Switzerland, (2008).

5. Ben Bachouch R., Fakhfakh M., Guinet A., Hajri-Gabouj S.: A model for scheduling drug deliveries in a French homecare, International Conference on Industrial Engineering and Systems Management (IESM), Montreal, Canada, (2009).
6. Benzarti E., Sahin E., Dallery Y.: Operations management applied to home care services: Analysis of the districting problem. Decision Support Systems **55**, 587–598, (2013).
7. Bertels S., Fahle T.: A hybrid setup for a hybrid scenario: combining heuristics for the home health care problem, Computers and Operations Research, **33**, 2866–2890, (2006).
8. Bredstörm D., Rönnqvist M.: A branch and price algorithm for the combined vehicle routing and scheduling problem with synchronization constraints. Technical report, Department of Finance and Management Science, Norwegian School of Economics and Business Administration, (2007).
9. Bredstörm D. and Rönnqvist M.: Combined vehicle routing and scheduling with temporal precedence and synchronization constraints, European Journal of Operational Research, 191(1), 19–31, (2008).
10. Brucker P., Qu R., Burke E.: Personnel scheduling: Models and complexity, European Journal of Operational Research, **210(3)**, 467–473, (2011).
11. Bowman A. W. and Azzalini A.: Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations. Oxford: Oxford University Press., (1997).
12. Borsani V., Matta A., Beschi G., Sommaruga, F.: A home care scheduling model for human resources, in Proceedings of Int. Conf. on Service Systems and Service Management, France, 449–454, (2006).
13. Cappanera, P., Scutellà, M.G.: Joint assignment, scheduling and routing models to home care optimization: a pattern based approach. Technical Report, TR-13-05, Dipartimento di Informatica, Universita di Pisa, (2013).
14. Carello, G., Lanzarone, E.: A cardinality-constrained robust model for the assignment problem in Home Care services, European Journal of Operational Research, **2**, 748–762, (2014).
15. Chahed S., Marcon E., Sahin E., Feillet D., Dallery Y.: Exploring new operational research opportunities within the Home Care context: the chemotherapy at home, Health Care Management Science, **12**, 179–191, (2009).
16. Cheng E., Rich J.L.: A home care routing and scheduling problem. Technical Report TR98-04, Department of Computational And Applied Mathematics, Rice University, (1998).
17. Christofides, N., Mingozzi, A., and Toth, P., Exact algorithms for the vehicle routing problem, based on spanning tree and shortest path relaxations, Mathematical Programming, 20, 255–282, (1981).
18. Cordeau J.F., Laporte G., Pasin F., Ropke S.: Scheduling technicians and tasks in a telecommunications company, Journal of Scheduling, **13(4)**, 393–409, (2010).
19. De Causmaecker P., Demeester P., Vanden Berghe G., Verbeke B.: Analysis of real-world personnel scheduling problems, in: Proceedings of the 5th International Conference on the Practice and Theory of Automated Timetabling, Pittsburgh, PA, USA, 183–198, (2005).
20. Elbenani, B., Ferland, J.A., Gascon, V.: Mathematical Programming Approach for Routing Home Care Nurses, Proceedings of the 2008 IEEE, (2008)
21. Eveborn P., Flisberg P., Ronnqvist M.: LAPS CAREan operational system for staff planning of home care, European Journal of Operational Research, **171**, 962–976, (2006).
22. Goldberg D.E.: Genetic algorithms in search optimization and machine learning, Addison Wesley Publishing Company, USA, (1989).
23. Gillett, B., Miller L, A Heuristic for the Vehicle Dispatching Problem, Operations Research, 22, 340–349, (1974).
24. Gyorfi L.: The rate of convergence of k-nn regression estimates and classification rules, IEEE Transactionosn Information Theory, **27(3)**, 362–364, (1981).
25. Hertz A., Lahrichi N.: A patient assignment algorithm for home care services, Journal of the Operational Research Society, **60**, 481–495 (2009).
26. Hiermann G., Prandtstetter M., Rendl A., Puchinger J., Raidl G.R., Metaheuristics for solving a multimodal home-healthcare scheduling problem. Central European Journal of Operations Research, **23(1)**, 89–113, (2015).

27. Hulshof, P.J.H., Kortbeek, N., Boucherie, R.J., Hans, E.W., Bakker, P.J.M.: Taxonomic classification of planning decisions in health care: a review of the state of the art in OR/MS. Health Systems, **1**, 129–175, (2012).
28. Koeleman P.M. , Bhulai S., Van Meersbergen M.: Optimal patient and personnel scheduling policies for care-at-home service facilities, European Journal of Operational Research, **219**, 557–563,
29. Lanzarone E., Matta A.: A cost assignment policy for home care patients, Flexible Service and Manufacturing Journal, **24(4)**, 465–495, (2012).
30. Lanzarone E., Matta A., Sahin E.: Operations Management Applied to Home Care Services: The Problem of Assigning Human Resources to Patients," Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions, **42**, 1346–1363, (2012).
31. Lanzarone, E., Matta, A. : Robust nurse-to-patient assignment in home care services to minimize overtimes under continuity of care. Operations Research for Health Care, In Press.
32. Lim A., Rodrigues B., Song L.: Manpower allocation with time windows, Journal of the Operational Research Society, **55**, 1178–1186, (2004).
33. Liu R., Xie X., Augusto V., Heuristic algorithms for a vehicle routing problem with simultaneous delivery and pickup and time windows in home health care, European Journal of Operational Research, **230(3)**, 475–486, (2013).
34. Liu R., Xie X., Garaix T, Hybridization of tabu search with feasible and infeasible local searches for periodic home health care logistics, Omega, **47**, 17–32, (2014).
35. Mankowska D. S., Meisel F., Bierwirth C., The home health care routing and scheduling problem with interdependent services, Health Care Manag. Sci., **17(1)**, 15–30, (2014)
36. Matta A., Chahed S., Sahin E., Dallery Y.: Modelling home care organisations from an operations management perspective, Flexible Services and Manufacturing Journal, (2014).
37. Miller B.L, Goldberg D.E.: Genetic algorithms, tournament selection, and the effects of noise, Journal of Complex Systems, **9**, 193–212, (1995).
38. Miller C.E., Tucker A.W. and Zemlin R.A.: Integer programming formulation for traveling salesman problems, Journal of the ACM, **3**, 326–329, (1960).
39. Nickel S., Schröder M., Steeg J.: Mid-term and short-term planning support for home health care services, European Journal of Operational Research, **60**, 574–587, (2012).
40. Rasmussen M. S., Justesen T., Dohn A., Larsen J., Home care crew scheduling problem: Preference-based visit clustering and temporal dependencies, European Journal of Operational Research, **219(3)**, 598–610, (2012).
41. Sahin E., Matta A.: A contribution to operations management related issues and models for home care structures, Internatinal Journal of Logistics Reseach and Application, Available online (2015).
42. Thomsen K.: Optimization on home care, Thesis Report, Informatics and Mathematical Modeling, Technical University of Denmark, (2006).
43. Trautsamwieser A., Gronalt M., Hirsch P.: Securing home health care in times of natural disasters, OR Spectrum, **3**, 787–813, (2011).
44. Vidal, T., Crainic, T. G., Gendreau, M., Prins, C.: A unified solution framework for multi–attribute vehicle routing problems. Operations Research, **234(3)**, 658–673, (2014).
45. Vivek R.: Data Driven Learning Approaches in Operations Management Problems, Ph.D. Thesis, UC Berkeley: Industrial Engineering and Operations Research, (2012).
46. Kergosien Y., Lente Ch., Billaut J.-C., Home health care problem: An extended multiple Traveling Salesman Problem, Multidisciplinary International Conference on Scheduling: Theory and Applications (MISTA), Dublin, Ireland, (2009).
47. Wand M.P., Jones M.C.: Kernel Smoothing. London: Chapman and Hall, (1995).
48. Yalcindag S., Matta A., Sahin E.: Operator assignment and routing problems in home health care services, in: Proceedings of CASE 2012, Seoul, Korea, 325–330, (2012).

A Details for the Genetic Algorithms

GA is an adaptive search procedure applied to a set of solutions that uses the properties from population genetics (i.e., crossover and mutation) to guide the search. At each iteration, GA discards some solutions (poor ones) and generate new ones based on superior members of the current set of solutions. Evaluation of the solutions (e.g., poor or good) is based on a problem specific function that is named as *fitness function*. The general representation of the GA is presented in the Algorithm 3 below and problem specific components are explained in the following parts according to the adopted GAs.

Algorithm 3 Genetic Algorithm

1. Generate an initial population
2. Evaluate the chromosomes with the fitness function
3. Perform selection operation with tournament system
4. Perform crossover and mutation and check feasibility
5. Repeat step 2,3 and 4 until the stopping criterion is met

A.1 GA for the Assignment Problem with KR Approach

Each solution in GA is represented as a chromosome with two parts as the size of number of patients and care givers ($N+|\mathcal{B}|$). The first part of each chromosome contains the information for the care giver–patient match which is represented by a permutation of integers from 1 to N. On the other hand, the second part is used to show the number of patients that each care giver needs to provide the service for and consists of $|\mathcal{B}|$ non-negative integers. For example, Figure 4 illustrates the matching between 3 care givers and 9 patients. According to this figure, the first care giver is responsible to visit 4 locations (patients) of the chromosome patient 1, patient 4, patient 5 and patient 9, Then, the second one would visit the next 2 locations (i.e., patient 3 and patient 7) and the the last one would visit the 3 remaining locations (i.e., patient 2, patient 6 and patient 8).

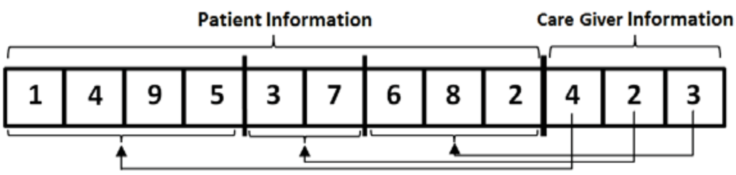


Fig. 4 Chromosome representation for the assignment problem with KR approach

The fitness function is the objective function of the assignment model given in the Equation (3), which is trying to balance the trade–off between care giver workload balancing and their total travel times. Since the fitness function is calculated for the assignment problem, the order of the patients in the chromosome that are matched with the care givers is not necessarily important. This is because travel time values are estimated with the use of the KR approach (or any other estimation method) by only using the patient IDs.

Selection process involves in choosing the chromosomes that would serve as parents for the next population generation. The tournament system [37] is used in which q chromosomes are randomly selected from the population. Then, two chromosomes with the minimum fitness function values are selected among these q individuals to be used as the two parents. This process is performed several times to populate the next generation.

Once the patents are selected a crossover operation is performed with a certain probability (p_c). First the chromosome is splitted into two parts according to the patient and care giver information. Then, with the part related to the patient information, the order crossover operation is performed to populate two children chromosomes (offsprings). Within this procedure, two cut points are randomly chosen from parent and parts between these cut points are mapped into two offsprings chromosomes. From the second cut point in one parent, the remaining genes are filled in the order that they appear in the other parent. After the order crossover operation, second part of the parent chromosomes is swaped and copied into the offsprings as well (see Figure 5 for an example of the crossover operation).

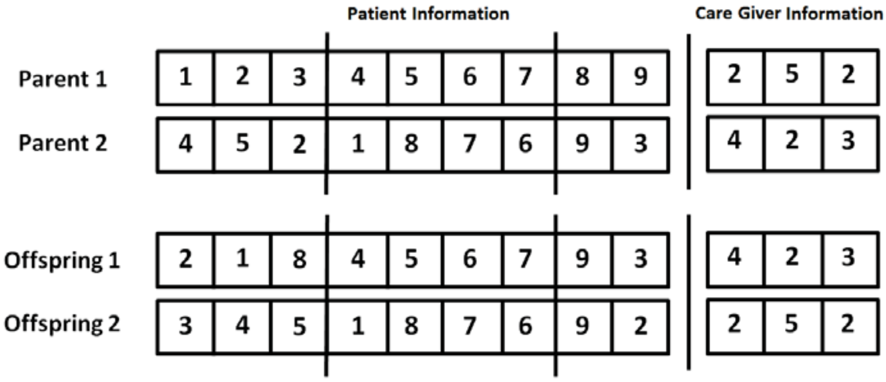


Fig. 5 Crossover Operation

Once the crossover operation is finalized, the mutation operation is held for the offspring chromosomes with a probability of p_m. The mutated chromosomes are obtained by randomly choosing two points between 1 and N and simply changing their places from the first part of where the patient information are stored. No mutation operation is performed for the second part of the chromosome.

Since the population matrix is generated according to the constraint where each patient can only be assigned to single care giver, feasibility is always ensured throughout the whole procedure.

It is also important to note that elitist selection process is also considered where the best chromosome in a generation is carried over the next one without any change.

A.2 GA for the TSP

Since TSP only deals with the visiting sequences of a single care giver, in the GA the chromosome represents the visiting sequence of the corresponding care giver. Thus, the chromosome represented in Figure 4 can also be used for this algorithm by only considering the first part which corresponds to the patient information.

The fitness function is the objective function of the TSP model which is trying to minimize the total travel time of the care giver. It is important to note that, for the fitness calculation, the order of patients must be considered as the they appear in any chromosome.

The population selection, crossover, mutation and elitism operations are the same as the previously described GA (see section A.1).

Since the population matrix is generated according to the constraint where each patient can be visited only once, feasibility is always ensured through out the whole procedure.

A.3 GA for the VRP

The chromosome representation of the VRP is the same as the one provided for the first GA (see section A.1).

The fitness function is the objective function of the VRP given in the Equation (10), which is trying to balance the trade-off between care giver workload balancing and their total travel times. As in the previous algorithm, the order of the patients must be considered as they are in any chromosome.

The population selection, crossover, mutation, elitism and feasibility operations are the same as the previously described GA (see section A.1).

We provide details about the performance of the implemented GA for the VRP. Remind that the VRP model that is used in this paper is the modified variant of the mTSP problem where instead of only minimizing the total travel time of care givers, we try to balance the trade-off between care giver workload balancing and their total travel times. To be able to analyze the performance of the GA, we only provide results based on total travel time minimization which is the basic model that is present in the literature. To do so, in addition to the h and γ terms of the objective function (Equation (10)) of the corresponding model, we also eliminate the constraints for the workload balancing (Equations (15) and (16)).

Table 8 shows the objective function values minimized by the implemented GA and the method used as benchmark [44]. All the results are obtained with same group of instances used in section 5. The first group (Instance B.1-B.5) corresponds to the first set of small problem instances with 15 to 25 patients (and a health care center) and 3 to 8 care givers. These ones are obtained from the benchmark instances that are used in the VRP literature ([23,17]). The second set of small problem instances (Instance A.1-A.5) are as the same size of Group S.1 and generated from real data as described in section 5. The third and last group of instances are also generated from real data and the difference lays on the dimensions. Hence, the third group corresponds to the medium sized instances (Instances A.6-A.10) where each instance have 56 patients and 7 care givers (and a health care center). The last group (Instances A.11-A.15) corresponds to the larger problem instances that have 150 patients and 15 care givers (and a health care center). Here the solutions are obtained with the same procedure as provided in the numerical result section.

Table 8 Accuracy analysis of the GA for VRP

Instance	CPLEX	UHGS	GA	Average error %
B.1	250.04	-	250.64 ± 0.68	0.24
B.2	285.27	-	288.27 ± 1.48	0.75
B.3	294.43	-	296.36 ± 3.14	0.66
B.4	601.67	-	601.96 ± 0.51	0.05
B.5	476.04	-	480.29 ± 3.57	0.80
A.1	24.54	-	24.54 ± 0.00	0.00
A.2	51.81	-	51.81 ± 0.00	0.00
A.3	36.12	-	36.43 ± 0.38	0.84
A.4	36.85	-	37.02 ± 0.19	0.48
A.5	70.93	-	71.86 ± 0.46	1.32
A.6	-	37.03	37.19 ± 0.11	0.56
A.7	-	44.05	44.26 ± 0.21	0.48
A.8	-	37.41	37.59 ± 0.29	0.49
A.9	-	36.92	37.00 ± 0.13	0.20
A.10	-	38.42	38.58 ± 0.07	0.44
A.11	-	53.84	54.77 ± 0.20	1.73
A.12	-	54.36	54.61 ± 0.20	0.47
A.13	-	54.11	54.54 ± 0.35	0.78
A.14	-	54.30	54.50 ± 0.20	0.37
A.15	-	55.14	55.44 ± 0.36	0.55

The solutions provided by the GA for the instances A.6 to A.15 are compared with the Unified Hybrid Generic Search method (UHGS) presented in the paper of Vidal *et al.* [44]. On the other hand, solution of the GA for both group of small sized instance (Instances B.1-B.5 and A.1-A.5) are directly compared with the ones that are executed by the ILOG CPLEX solver. It is observed that the maximum average error for all the groups of instances is less than 1.73%.