

# Automated Fluence Map Optimization based on Fuzzy Inference Systems

Joana Dias and Humberto Rocha

*Inesc-Coimbra, University of Coimbra, Portugal*

5 Tiago Ventura

*Medical Physics Department, IPOCFG,EPE, Coimbra, Portugal and Physics Department, University of Aveiro, Portugal*

10 Brígida Ferreira

*School of Allied Health Technologies Polytechnic Institute of Porto, Porto, Portugal*

Maria do Carmo Lopes

*Medical Physics Department, IPOCFG,EPE, Coimbra, Portugal*

**Purpose:** The planning of an intensity modulated radiation therapy treatment requires the optimization of the fluence intensities. The fluence map optimization (FMO) is many times based on a non-linear continuous programming problem, being necessary for the planner to define *a priori* weights and/or lower bounds that are iteratively changed within a trial-and-error procedure until an acceptable plan is reached. In this work we describe a new approach for FMO that releases the human planner from trial-and-error procedures, being a truly automated optimization process.

**Methods:** The FMO is represented by a voxel-based convex penalty continuous non-linear model. This model makes use of both weights and lower/upper bounds to guide the optimization process towards interesting solutions that are able to satisfy all the constraints defined for the treatment. All the model's parameters are iteratively changed by resorting to a fuzzy inference system. This system analyses how far the current solution is from a desirable solution, changing in a completely automated way both weights and lower/upper bounds. The fuzzy inference system is based on fuzzy reasoning that enables the use of common-sense rules within an iterative optimization process. The method is built in two stages: in a first stage an admissible solution is calculated, guaranteeing that all the treatment planning objectives are being satisfied. In this first stage, the algorithm tries to improve as most as possible the irradiation of the planning target volumes. In a second stage, the algorithm tries to improve organ sparing, without jeopardizing tumour coverage.

**Results:** The proposed methodology was applied to ten head-and-neck cancer cases already treated in the Portuguese Oncology Institute of Coimbra (IPOCFG) and signalized as complex cases. IMRT treatment was considered, with 7, 9 and 11 equidistant beam angles. It was possible to obtain admissible solutions for all the patients considered and with no human planner intervention. The results obtained were compared with the optimized solution using a similar optimization model but with human planner intervention. For the vast majority of cases, it was possible to improve organ sparing and at the same time to assure better tumour coverage.

**Conclusions:** Embedding a fuzzy inference system into FMO allows human planner reasoning to be used in the guidance of the optimization process towards interesting regions in a truly automated way. The proposed methodology is capable of calculating high quality plans within reasonable computational times and can be an important contribution towards fully automated radiation therapy treatment planning.

*Keywords:* Radiotherapy treatment planning, Fluence Map Optimization, Automated planning, Fuzzy Logic, IMRT

## 1.INTRODUCTION

Radiotherapy is the most technologically demanding cancer treatment approach, requiring a complex treatment planning process. In clinical practice, most of the times, the planning is done by resorting to a lengthy trial-and-error procedure assisted by a dedicated software (Treatment Planning System). In inverse treatment planning optimization, the planner will have to define several different parameters, like weights and bounds. These parameters are iteratively changed by the planner, using his own reasoning and experience, until an acceptable plan is reached. This process can be a very lengthy process, especially for difficult cases where proper tumour coverage and organ sparing prove difficult to obtain.

In this paper we will focus our attention on fluence map optimization (FMO) for IMRT treatment planning, where the beam angles are considered determined and fixed *a priori*.

There is a vast literature associated with FMO. FMO has been mainly tackled by constrained optimization models such that an objective function is optimized while meeting dose requirements<sup>1-6</sup>. The objective function usually considers a weighted sum of deviations from prescribed doses (underdose for PTVs and overdose for OARs). The constraints to consider will depend on the organs' functionality. It is possible to find examples of linear models<sup>7</sup>, mixed integer linear models<sup>8</sup>, nonlinear models<sup>9,10</sup>, and multiobjective models<sup>11</sup>. Most models known from the literature require some level of tuning and trial-and-error procedures, making treatment planning a lengthy and sometimes cumbersome process. To invest in automated treatment planning, taking advantage of current computational resources and optimization algorithmic developments, is an increasingly accepted path of research<sup>12</sup>. There have been very interesting steps in this direction<sup>13-17</sup>.

In this paper we propose a completely different methodological approach for FMO. A non-linear unconstrained continuous programming model will be used, and will be iteratively solved by having the model's parameters changed in an automated way using a fuzzy reasoning inference system. This methodology releases the human planner from trial-and-error procedures, being a completely automated approach for FMO. The physician will define the objectives that have to be achieved for a treatment plan to be considered admissible. These objectives can, in fact, be interpreted as constraints that are defined for the different structures: percentage of the PTV volume that receives at least a percentage of the prescribed dose, maximum or average doses that cannot be exceeded, maximum percentage of the structure's volume that receives more than a given dose, for example. In this methodological approach, the planner can define priorities that are associated with PTVs and organs at risk (OAR). For PTVs, priorities are used whenever it is not possible to comply with all the defined treatment objectives (meaning that some corresponding constraints are

being violated). Regarding OARs, priorities are used when the algorithm is trying to spare further these structures. A priority is simply an integer value belonging to a given interval.

85 The width of this interval can be defined taking into account the number of structures considered in the optimization and the degree of differentiation among structures' priorities that is intended. In the present case, we have considered an interval between 0 and 10.

The algorithm will then generate a treatment plan trying to comply with all the constraints defined, and without human intervention. To the best of our knowledge, it is the first time  
90 that fuzzy inference is applied for radiation therapy treatment planning.

## 2.METHODS

There are many ways of performing IMRT FMO. We have chosen to use a voxel-based convex penalty non-linear model where each voxel is penalized considering the square difference of the amount of dose received by the voxel and a given upper and/or lower bound.

95 This formulation yields a programming problem with only linear nonnegativity constraints on the fluence values. Considering that beam angles have already been fixed, let  $V$  represent the number of voxels,  $N$  the number of beamlets and  $D$  the dose matrix, such that  $D_{ij}$  represents the contribution of beamlet  $j$  to the total dose deposited in voxel  $i$ . We can thus calculate the

total dose received by voxel  $i$  as  $\sum_{j=1}^N D_{ij} w_j$  with  $w_j$  representing the weight of beamlet  $j$ . Let

100  $U_i$  be the upper bound associated with voxel  $i$ ,  $L_i$  the lower bound associated with voxel  $i$ ,  $\underline{\lambda}_i$  and  $\bar{\lambda}_i$  the penalty weights of underdose and overdose of voxel  $i$ , respectively. The FMO model can be defined as follows, where  $(\bullet)_+ = \max\{0, \bullet\}$ :

$$f(w) = \text{Min}_w \sum_{i=1}^V \left[ \underline{\lambda}_i \left( L_i - \sum_{j=1}^N D_{ij} w_j \right)_+^2 + \bar{\lambda}_i \left( \sum_{j=1}^N D_{ij} w_j - U_i \right)_+^2 \right] \quad (1)$$

$$\text{s.t. } w_j \geq 0, j = 1, \dots, N \quad (2)$$

Although this formulation allows unique weights and unique upper/lower bounds for each voxel, in the current approach weights are assigned by structure only so that every voxel in a given structure has the weight assigned to that structure. Considering a given structure  $S$  we can thus define  $\bar{\lambda}_S$ ,  $\underline{\lambda}_S$ ,  $U_S$  and  $L_S$  such that  $\bar{\lambda}_i = \bar{\lambda}_S, \forall i \in S$ ,  $\underline{\lambda}_i = \underline{\lambda}_S, \forall i \in S$ ,  $U_i = U_S, \forall i \in S$  and  $L_i = L_S, \forall i \in S$ . This nonlinear formulation implies that a very small amount of deviation from the established bounds may be accepted, but larger deviations are decreasingly tolerated. It should be stressed that the considered objective function does not have any meaningful clinical interpretation. It should be interpreted as a technical tool that will allow the guidance of the optimization procedure towards regions where admissible treatment plans can be found.

## 2. A. Initialization of the parameters

The upper and lower bounds ( $\underline{\lambda}_S$  and  $\bar{\lambda}_S$ ) are initialized considering the desired prescription defined by the physician. To illustrate this point, consider a PTV structure for which two objectives have to be attained: the physician has defined 70 Gy as the prescribed dose ( $D_p$ ) and, at the same time, he does not allow any PTV voxel to receive more than 107% of  $D_p$ . This PTV will have both upper and lower bounds associated with it. The lower bound will be first initialized as being equal to the prescribed dose in the PTV (70 Gy). The upper bound will be initialized as the maximum dose allowed (74.9 Gy). Regarding OARs, only upper bounds will be defined. These upper bounds will be equal to the maximum or mean dose allowed for the structure. Considering the spinal cord, for instance, if the physician defines 45 Gy as the maximum admissible dose in any spinal cord voxel, then this will be the upper bound considered. For parotids, if a mean dose of 26 Gy is considered admissible, then this will be the upper bound considered for each parotids' voxels.

Weights are considered equal to 1 for every voxel. This means that each voxel will have  
130 exactly the same weight regarding the optimization process, and no rescaling is made  
considering, for instance, the number of voxels in each structure. An exception to this rule is  
considered whenever one structure is delineated inside another one. This is very common in  
the PTVs, where a PTV structure can be totally located inside another PTV. In this case, the  
inner PTV structures should have a greater weight, because in reality each voxel in the inner  
135 PTV will belong to more than one structure, possibly with conflicting objectives. This is  
illustrated in FIG. 1, where a CT slice for one patient is depicted, and it is possible to see two  
PTVs with a prescribed dose of 70 Gy inside a PTV with a prescribed dose of 59.4 Gy.  
Voxels included in PTV<sub>70</sub> belong to two PTV structures at the same time, and they will be  
subject to conflicting constraints. This means that the voxels of PTV<sub>70</sub> should have a greater  
140 weight in the optimization procedure when they are considered as belonging to PTV<sub>70</sub> than  
when they are considered as belonging to PTV<sub>59.4</sub>.

The proposed methodology is intended to be a completely automated methodology for FMO,  
so we would not like to have the planner initializing or adjusting these weights by  
trial-and-error. As a matter of fact, the initialization of the weights considering values  
145 different than 1 will only have as consequence a decrease in the total computational time (the  
weights will be automatically updated by the algorithm, and giving a better starting point will  
only contribute to a decreased number of weights' updates), and not in the quality of the final  
treatment plan calculated. We have chosen to consider a very simple way of automatically  
initializing the weights, considering only the relation between the volume of the inner PTVs  
150 and the volume of the outer PTV: the smallest the volume, the greater the weight.

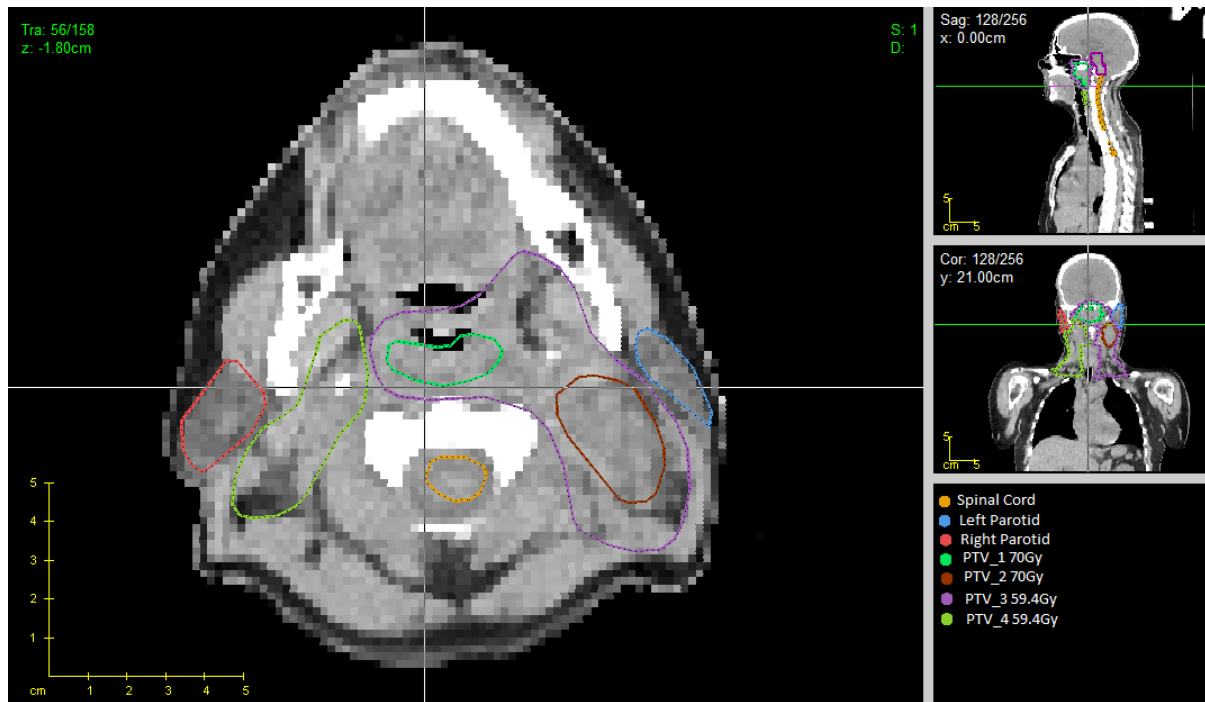


FIG. 1. Contoured structures in one CT slice for one representative patient. In this CT slice it is possible to identify 4 PTV structures, with two of them (PTV\_1 and PTV\_2) being located inside PTV\_3.

155

If the inner PTV volume is less than 5% of the outer PTV volume, then the weight of all inner PTV voxels will be equal to 50. If the volume is greater than 5% but less than or equal to 10%, then the weight will be equal to 10. In all other cases, it will be equal to 5. The choices of these thresholds and of the initial weights' values can seem as being arbitrary, and they really are arbitrary to some extent. Many other thresholds or initial values could be used, provided that they are based in the same reasoning.

160

## 2. B. First Phase: PTV Coverage

In the first phase of the automated methodology, the algorithm will try to achieve the desired PTV coverage, guaranteeing at the same time proper OAR sparing (it will try to ensure that all the treatment constraints are satisfied). This is done in an iterative way, first by trying to change upper and lower bounds only. If this is not sufficient, the structures' weights are also changed.

165

Lower and upper bounds are changed following very simple common-sense rules: if a given treatment constraint is being violated, then change the upper or lower bound that can contribute to the satisfaction of that constraint. As an example consider that the dose received by a given structure  $S$  (spinal cord, for instance) in the current treatment plan is greater than the acceptable value defined by the prescription (the maximum dose is greater than 45 Gy). Then we would like to increase the contribution for the objective function (1) of all voxels of the spinal cord, making the optimization procedure search for solutions such that the dose in these voxels is lower than in the current solution. This is achieved by decreasing the respective upper bound  $U_s$ , contributing to an increase of the term  $\left(\sum_{j=1}^N D_{ij}w_j - U_i\right)_+^2$  in the objective value, for all voxels  $i \in S$  that are receiving a dose greater than  $U_s$ . The optimization procedure will try to decrease  $\sum_{j=1}^N D_{ij}w_j, \forall i \in S$ , contributing to the satisfaction of the violated constraint. On the other hand, if the dose received by  $S$  is lower than the corresponding prescribed dose (in case of a PTV, for instance), then the current lower bound  $L_s$  should be increased, guaranteeing an increase in  $\left(L_i - \sum_{j=1}^N D_{ij}w_j\right)_+^2, \forall i \in S$ . This will force the optimization procedure to increase the dose received by  $S$ . The question that remains is *how much should these bounds be adjusted?* A common-sense rule will tell us that the greater the deviation of the current solution from an acceptable solution (the greater the violation of the corresponding constraint), the greater the adjustment. We can then consider that:

- If, for a given structure, the deviation between the received dose and the prescribed/accepted dose is *low* (the corresponding treatment constraint is being only slightly violated), then the upper/lower bound should be only *slightly* decreased/increased.



- 190 • If, for a given structure, the deviation between the received dose and the prescribed/accepted dose is *large*, then the upper/lower bound should undergo a *large* decrease/increase.
- If, for a given structure, the deviation between the received dose and the prescribed/accepted dose is *medium*, then the upper/lower bound should suffer a *medium* decrease/increase.
- 195

It can be very difficult to define in a crisp and clear way what we mean by *large*, *medium* or *low*. We understand these concepts and these common-sense rules, but how can we translate this into a computational algorithm? It is possible by resorting to fuzzy logic and fuzzy reasoning.

### 200 **2.B.1. Fuzzy Inference Mechanism**

Fuzzy logic allows us to mathematically represent concepts that we can understand but that may be difficult to define in a precise way<sup>17</sup>. It is based on the concept of *fuzzy sets*: sets that do not have clear boundaries so that a given element can belong to a set with only a partial degree of membership. What does *low* mean? What elements can belong to the set *low*?

205 Instead of considering a crisp definition, we will consider a membership function that will represent the *low* concept. This means that a given deviation can be considered *low* with a given degree of membership (a number between 0 and 1). The same thing happens with the *medium* and *large* concepts. Actually, a given deviation can belong simultaneously to more than one of these sets, with different degrees of membership.

210 Membership functions can be represented by many different ways, the only requisite being that they must vary between 0 and 1. In this work we have decided to work with triangular and trapezoidal membership functions. FIG. 2 and FIG. 3 illustrate the membership functions considered for determining the change in the bounds. The input is given by the percentage of deviation between the prescribed/accepted dose and the actual dose received (the input will

215 belong to [0%,100%], and this is simply a measure of how much the constraint is being

violated). The output is given by the percentage of change in the corresponding bound. A deviation of 9% in the current dose delivery to a given structure, for instance, will be interpreted as belonging simultaneously to the set *medium deviation* and *large deviation* with different membership functions. In FIG. 2 and FIG. 3, although the  $x$ -axis considers values from 0% to 100%, we only represent a subset of this interval, to illustrate in a better way the behaviour of the membership functions.

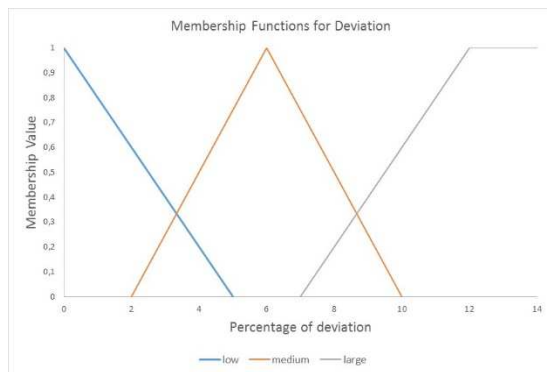


FIG. 2. Input (Percentage of Deviation) Membership Functions (deviations belonging to [12%,100%] are considered *large* with membership equal to 1)

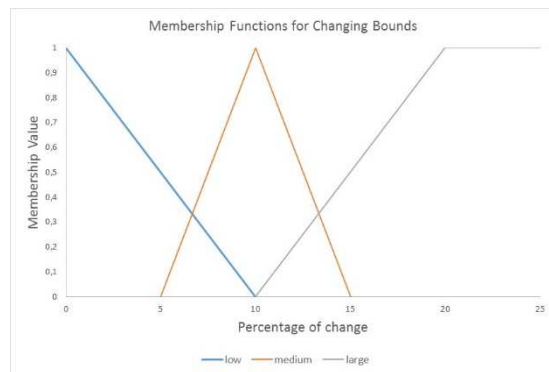


FIG. 3. Output (Percentage of Change in the Bound) Membership functions (changes belonging to [20%,100%] are considered *large* with membership equal to 1).

Input and output membership functions are connected through a set of fuzzy rules that constitute a fuzzy inference mechanism. These fuzzy rules are no more than simple if-then rules. “If deviation is *low* then the change in the corresponding bound is *low*” is an example of a fuzzy rule. Notice that the input to a fuzzy rule is a degree of membership (a value between 0 and 1, in this case the degree of membership of the current deviation to the set *low*), whereas the output of a fuzzy rule is a whole fuzzy set (the output fuzzy set *low*). The reasoning in evaluating a fuzzy rule is that if the antecedent is true to some degree of membership function, then the consequent will also be true to the same degree (*implication method*). In the present case, we have three fuzzy rules that should be simultaneously evaluated: the one already cited and two similar ones but that consider *medium* deviations implying *medium* changes and *large* deviations implying *large* changes to bounds. In a fuzzy inference system, all rules are simultaneously evaluated. After evaluating all the fuzzy rules,

235 we end up with a set of truncated output fuzzy sets, being then necessary to calculate a single  
 crisp value from these sets. This procedure, usually known as *defuzzification*, will first  
 aggregate all output fuzzy sets into a single one, and then will produce a crisp value. This can  
 be done by several different ways, being the most common one the centroid calculation that  
 returns the centre of the area under the curve. The whole process is illustrated in FIG. 4,  
 240 considering as example a deviation of 9%.

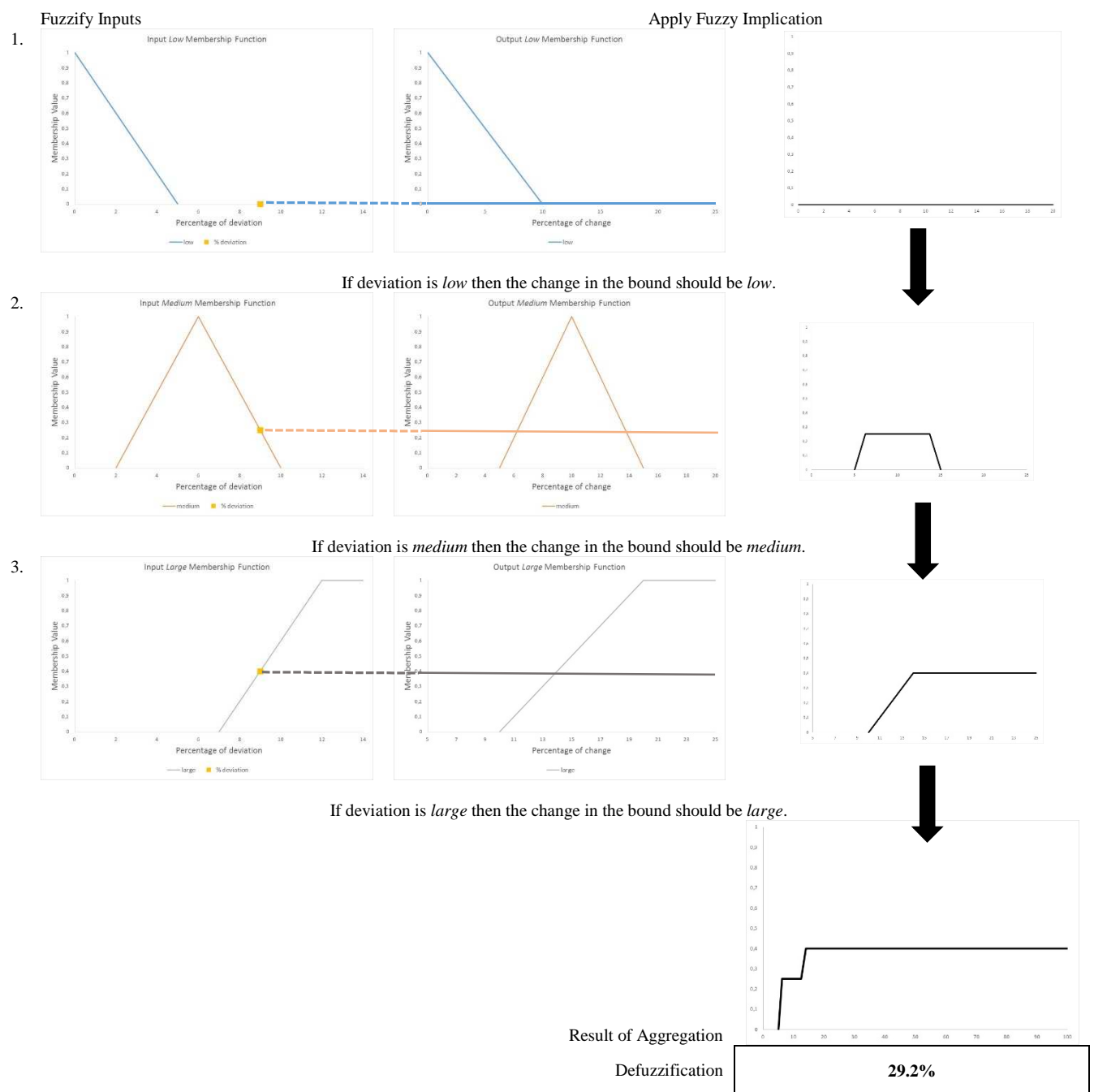


FIG. 4. Diagram representing the whole fuzzy inference system for changing the bounds

This value is evaluated for each of the three fuzzy rules: its membership function considering the input fuzzy set of each rule is calculated. Based on this membership value, the corresponding output fuzzy set is created by truncating the maximum membership value. The aggregation procedure aggregates all the fuzzy output sets, and based on this aggregated set a crisp value is calculated. This is the value that will determine the percentage of change in the corresponding bound.

### ***2.B.2. Automated mechanism for bounds update***

In each iteration of the algorithm, and for each and every structure considered, the algorithm will calculate the deviation between the current solution and the constraints defined for that structure. If there are deviations (meaning that some of the constraints for that structure are  
245 being violated), then the respective bounds will be updated resorting to the fuzzy inference mechanism described in the previous section. As an example, consider a given PTV structure with a prescribed dose of 70 Gy. The physician has defined as acceptable a solution such that 95% of the volume receives at least 95% of the prescribed dose ( $D_{95\%} \geq 95\% D_p$ ). Furthermore, the maximum dose allowed ( $D_{\max}$ ) for PTV voxels should be no greater than  
250 107% of the prescribed dose. Imagine that in the current solution  $D_{95\%} = 65.9$  Gy and at the same time  $D_{\max} = 78$  Gy. Regarding the  $D_{95\%}$  metric, the algorithm will calculate the current deviation in percentage (approximately 1%). Using the fuzzy inference mechanism, and the three fuzzy rules considered, it will determine a percentage of change (approximately 3.3%) and will increase the current structure's lower bound  $L_s$  using this percentage. At the same  
255 time, the algorithm will also calculate the deviation regarding the maximum dose achieved and the maximum dose accepted. In this case the deviation is approximately 4%. Applying the same fuzzy inference mechanism, the corresponding upper bound  $U_s$  will be updated (decreased) in 8.3%. As soon as the upper and lower bounds are updated for every structure considered in the optimization, the FMO is solved again, generating a new current solution.

260 The procedure is repeated until an admissible solution is found, or the upper/lower bounds reach defined thresholds. In the present work we consider changing the weight  $\bar{\lambda}_s$  whenever its upper bound  $U_s$  reaches a value lower than 15% of its initial value. The weight  $\underline{\lambda}_s$  is updated whenever  $L_s$  is greater than twice its original value. It is also possible to consider different thresholds leading to more or less frequent changes of weights. If the threshold  
265 associated with  $U_s$  is increased and the threshold associated with  $L_s$  is decreased, weights will be updated more often. This can either result in a decrease in the computational time or the algorithm having problems to converge to a solution. Actually, most of the times, only slight changes in the weights are needed to reach admissible solutions, so it is important to give enough time for the algorithm to adjust bounds  $U_s$  and  $L_s$  before changing the weights.

### 270 ***2.B.3.Automated mechanism for weights update***

If it is not possible to reach an admissible solution by manipulating lower and upper bounds only, the algorithm will automatically change the weights associated with structures that are violating the defined constraints. The weights are also changed according to a fuzzy inference system, similar to the one already described. The fuzzy rules used are also similar: if the  
275 deviation between the current solution and the admissible one is large/medium/small then increase the corresponding weight by a large/medium/small amount.

To prevent the algorithm from not converging, the maximum allowed change is increasing the weight by adding at most 10 to the current weight. The input fuzzy membership functions are the same depicted in FIG. 2. The output membership functions are different than the ones  
280 shown before and are depicted in FIG. 5. Whenever the weights are changed, the lower and upper bounds take their initial values, and the algorithm progresses by changing these bounds again, until an admissible solution is found.

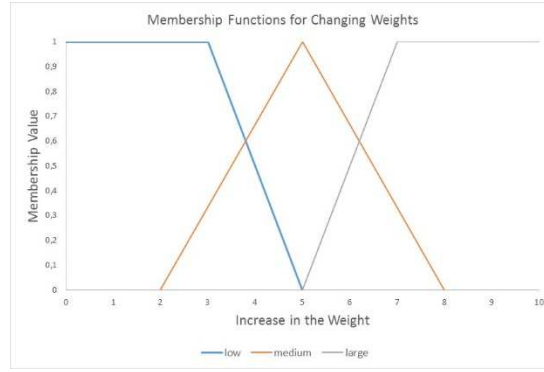


FIG. 5. Output Membership functions for updating weights

### 2.B.4.Achieving PTV Coverage

By changing the lower/upper bounds and also updating the weights associated with each structure, the algorithm will guide the search process and will, eventually, find a solution that satisfies all the constraints defined by the physician. If indeed an admissible solution is found, then, at this stage, the algorithm will try to improve even more PTV coverage, by being more demanding than the physician initially was. At the present moment, this is done by changing  $D_{V\%}$  to  $D_{(V+1)\%}$  : if the planner has defined that an admissible solution should have  $D_{95\%}$  greater than or equal to a given value  $A$ , then the algorithm will now try to guarantee that  $D_{96\%}$  is greater than or equal to  $A$ . The improvement of PTV coverage can also be done in other ways, depending on the type of constraints that were defined. The procedure is repeated until the algorithm is no longer capable of improving PTV coverage.

If, however, the algorithm is not capable of finding an admissible solution considering the initial prescribed constraints, it will even so try to calculate a solution. To this end, a slack associated with each PTV is considered, representing an acceptable deviation from the defined constraints (the constraint is slightly relaxed). Slacks represent a percentage such that if the current deviation is within that percentage, it is disregarded. As an example, imagine that the physician has defined the constraint  $D_{95\%} \geq 67 \text{ Gy}$ . Then this constraint will be considered satisfied if  $D_{95\%} \geq (1 - \text{slack}) \times 67 \text{ Gy}$ . The slacks associated with PTVs are

automatically determined by the algorithm, resorting again to a fuzzy inference mechanism. The physician can assign priorities to each PTV (the same for all PTVs or different priorities to different PTVs). These priorities are translated by the planner into a number belonging to  $[0, 10]$ , where 0 is no priority and 10 is the greatest priority of all (as already noticed, other intervals could be considered if a thinner differentiation between structures' priorities is desired). The greater the priority, the smaller the slack (meaning that only slight deviations from the desired constraints will be accepted). Three fuzzy rules are also considered: if the priority is high/medium/low then the slack is small/medium/high. FIG. 6 and FIG. 7 illustrate the membership functions considered.

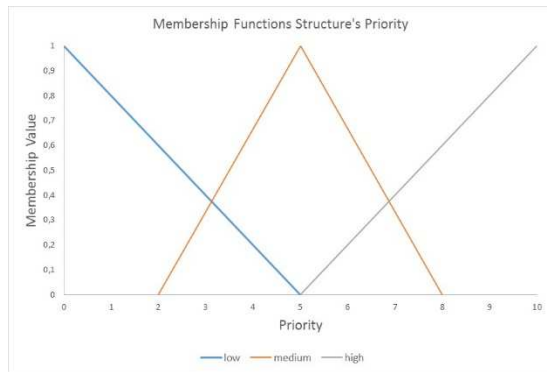


FIG. 6. Input Membership Functions considering the Structure's priority

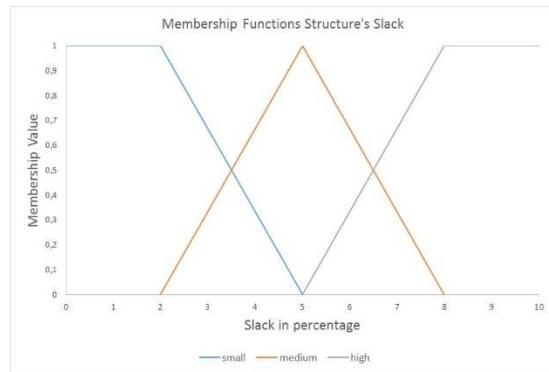


FIG. 7. Output Membership Function considering the slack associated with the structure's constraints.

The first phase of the algorithm is illustrated in FIG. 8.

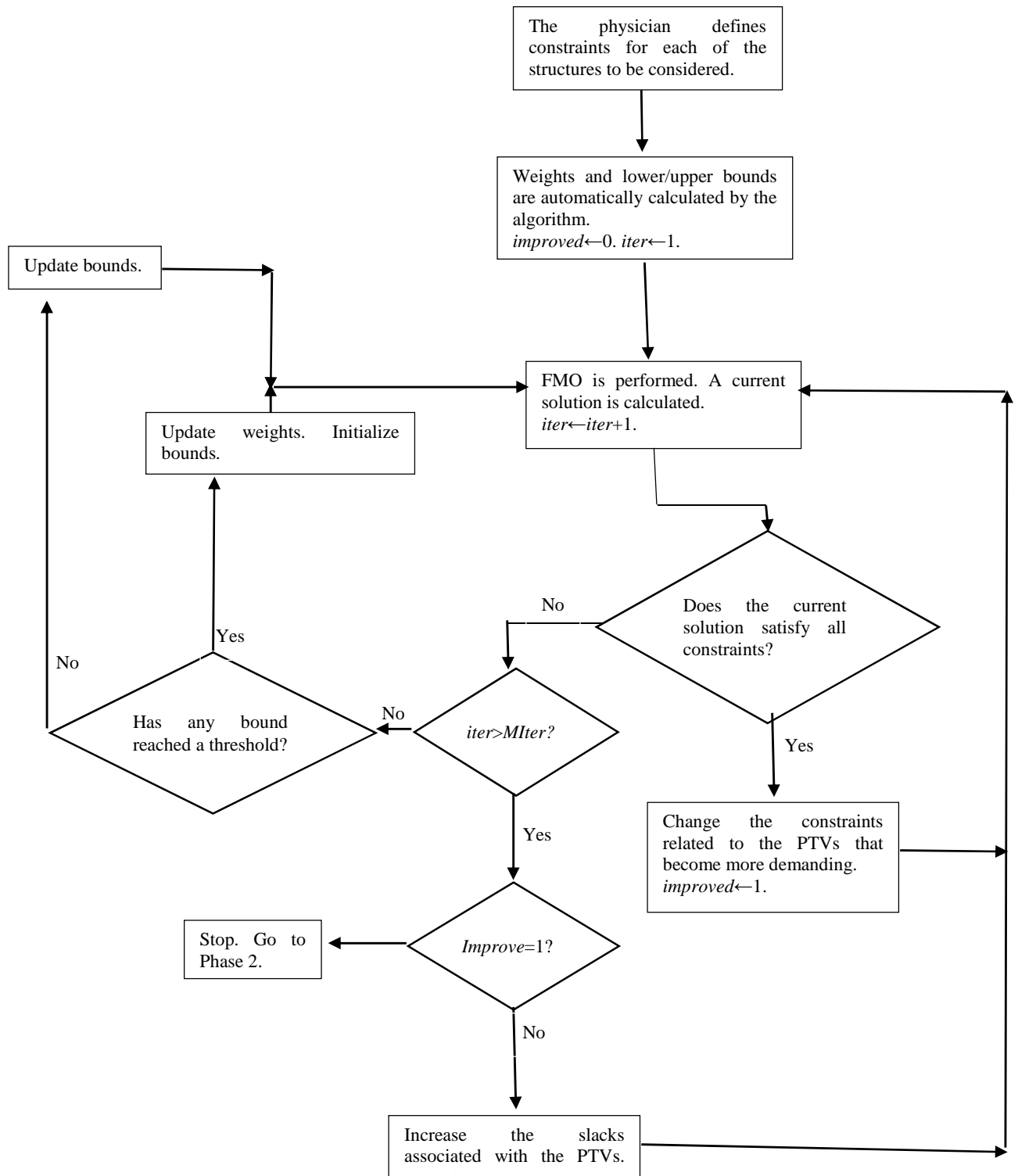


FIG. 8: Flowchart representing the algorithm's first phase.

## 315 2. C. Second Phase: OAR sparing

After the first phase, the algorithm has delivered a given solution satisfying as much as possible the physician defined treatment constraints. In the second phase, the algorithm will now try to improve OAR sparing, maintaining the achieved PTV coverage. The OAR sparing



is done by becoming more demanding in relation to the initial accepted doses. As an  
320 example, imagine that the physician has defined as a constraint stating that the spinal cord  
maximum dose should be no greater than 45 Gy. Then, the algorithm will be more  
demanding and will consider a constraint stating that the maximum dose acceptable in the  
spinal cord has to be less than or equal to  $\rho < 45\text{Gy}$ . The magnitude of this change ( $45 - \rho$ )  
will be automatically determined considering the difference between what is being achieved  
325 and what is considered admissible: if the constraint is being satisfied by a great amount then  
we can be more demanding and consider lower values for  $\rho$ , but if the constraint is being  
satisfied almost as an equality then we have to be less demanding and consider  $\rho$  closer to  
45. At the same time, slacks are defined for all the OARs that we are trying to spare in a  
better way, to give the algorithm flexibility to find better solutions. The physician can define  
330 different priorities for the OARs, in a manner similar to the one described for the PTVs. The  
greater the priority, the more demanding the algorithm will be for that structure, so the  
smaller the corresponding slack will be. Once again, all changes are automatically performed  
by resorting to fuzzy inference systems using the same type of fuzzy rules already described.  
The procedure is repeated until it is not possible to improve the current solution in a  
335 predetermined number of iterations or a global maximum number of iterations is reached.

### 3.RESULTS

The described methodology was applied to ten selected head-and-neck clinical examples  
where proper target coverage and organ sparing, in particular parotid sparing (parotid glands  
340 are in close proximity to or even overlapping with the PTV), proved to be difficult to obtain.  
In this study, the OARs used for treatment optimization were limited to the spinal cord, the  
brainstem and the parotid glands. For the head-and-neck cases in study there are two or more  
PTVs with different prescribed doses. We will in general refer to them as PTV1 (prescription

of 70 Gy) and PTV2 (prescription of 59.4 Gy). For most patients, at least one PTV1 is inside  
 345 PTV2. Table 1 presents the defined constraints for each of the considered structures, as well  
 as the priorities defined for each structure to be used if it is not possible to achieve the desired  
 PTV coverage, or in the second phase of the algorithm for OAR sparing.

**Table 1.** Prescribed doses for all the structures considered for FMO

Structure	Type of constraint		Limit	Priority
Spinal cord	Maximum dose	lower than	45 Gy	10
Brainstem	Maximum dose	lower than	54 Gy	10
Left parotid	Mean dose	Lower than	26 Gy	5
Right parotid	Mean dose	Lower than	26 Gy	5
PTV1	$D_{95\%}$	Greater than	66.5 Gy	10
PTV1	Maximum dose	Lower than	74.9 Gy	
PTV2	$D_{95\%}$	Greater than	56.4 Gy	10
PTV2	$V_{107\%}$	Lower than	Percentage of PTV1 volume inside PTV2 plus a 10% margin	
Body	Maximum dose	Lower than	80 Gy	2

350 Our tests were performed on an Intel Core i7 CPU 2.8 GHz computer with 4GB RAM and  
 Windows 7. We used CERR 3.2.2 version<sup>18</sup> and MATLAB 7.4.0 (R2007a). The dose was  
 computed using CERR's pencil beam algorithm (QIB), with corrections for heterogeneities.  
 For each of the ten head-and-neck cases, the sample rate used for Body was 32 and for the  
 remaining structures was 4 (meaning that each set of 32 Body voxels was considered as one  
 355 voxel in the optimization procedure and, for all other structures, one out of 4 voxels was used  
 in the optimization procedure). To address the convex non-linear formulation of the FMO  
 problem we used a trust-region-reflective algorithm (*fmincon*) of MATLAB 7.4.0 (R2007a)  
 Optimization Toolbox. The termination criteria for the algorithm are as follows: the  
 algorithm will stop after 75 iterations without being able to find an admissible solution, or if  
 360 300 iterations in total are reached.

The algorithm was applied considering 7, 9 and 11 coplanar equidistant beam angles. For all  
 patients it was possible to find an admissible solution, satisfying all of the defined constraints,  
 within reasonable computational times (Table 2). In the vast majority of cases, it was also  
 possible to improve both PTV coverage and OAR sparing, compared with the first admissible

365 solution found. In Table 2 the computational times are depicted. The times corresponding to the first admissible solution show how long it takes for the algorithm to calculate the first plan complying with all the treatment constraints. After this first solution, the algorithm continues, first trying to improve PTV coverage (phase 1) and then trying to improve OAR sparing (phase 2). The *improved solution* corresponds to the last solution calculated during  
370 phase 2.

**Table 2.** Computational times for the Fuzzy Inference based FMO

Number of angles	Time (minutes)	Patients									
		1	2	3	4	5	6	7	8	9	10
7	First admissible solution	8	15	9	8	18	36	6	10	4	45
	Improved solution	-	89	75	15	34	102	49	112	27	153
9	First admissible solution	12	12	9	5	15	14	5	9	5	56
	Improved solution	120	89	124	17	183	122	69	17	204	128
11	First admissible solution	47	19	22	10	31	27	13	95	10	27
	Improved solution	-	223	333	45	87	62	108	439	68	535

The treatment plans calculated (*Fuzzy FMO*) are also compared with the treatment plans  
375 calculated by using FMO without automated parameter tuning (*simple FMO*). In this case, both weights and bounds are defined by the planner. The bounds are kept constant and given by the defined constraints. Weights were optimized by a manual trial-and-error procedure.

The automated treatment plan was able to comply with all the defined treatment constraints, considering both PTV coverage and OAR sparing. Comparative results regarding PTV1 and  
380 PTV2 coverage are presented in FIG. 9A and FIG. 9B, considering 7, 9 and 11 equidistant beams with the simple and the fuzzy FMO. For all cases where PTV1 and/or PTV2 correspond to more than one structure, a weighted average was considered (considering the number of voxels in each structure). The horizontal line represents 95% of the prescribed dose. The treatment constraint determines that  $D_{95\%} \geq D_p$  so an admissible solution will  
385 correspond to values above this horizontal line. As can be seen, only simple FMO solutions present values under this admissibility threshold.

FIG. 9C and FIG. 9D present comparative results regarding spinal cord and brainstem. The horizontal line represents the maximum dose allowed in any voxel of these OARs, so we

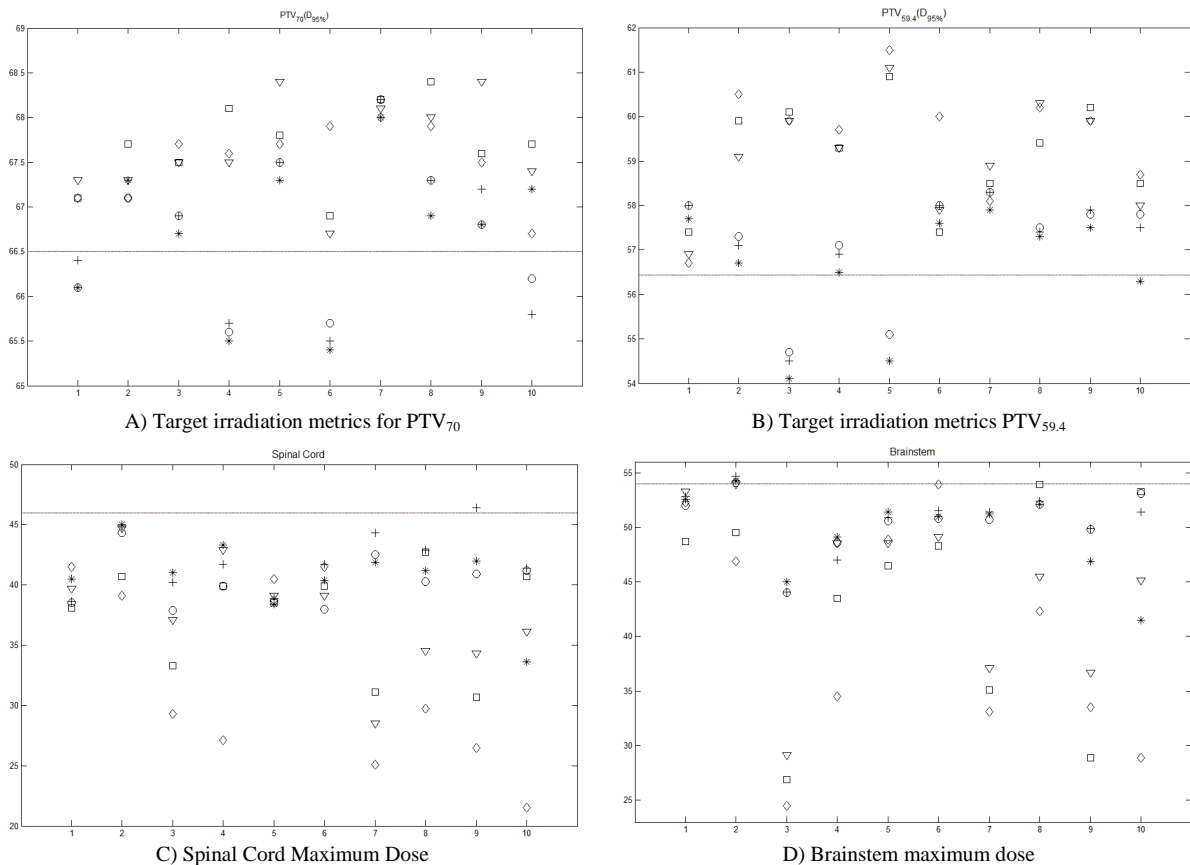
would like to have all values under this line. Almost all solutions comply with these

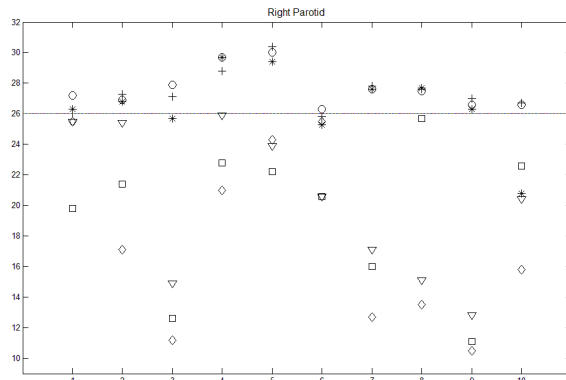
390 constraints, that are violated by few simple FMO solutions only. FIG. 9E and FIG. 9F depict similar results for the parotids. In this case the horizontal line represents the maximum acceptable value for the mean dose. It is possible to observe that the only values above the horizontal line correspond to simple FMO solutions. Solutions calculated by the proposed

395 fuzzy methodology satisfy all the constraints defined for OARs, and they are able of considerably improving organ sparing. More detailed dosimetry metrics information regarding all the 6 different solutions for each patient is given as supplemental material.

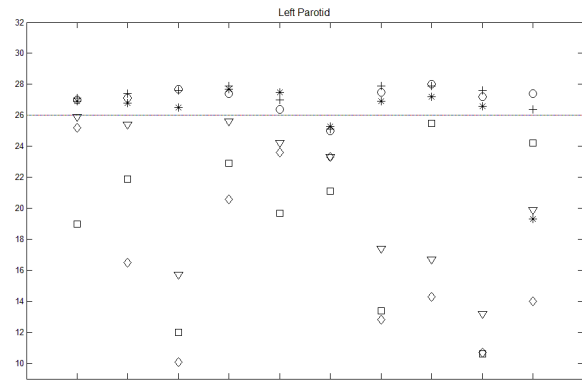
Comparative DVHs are also shown, considering two cases and 11 beams solutions (FIG. 10 and FIG. 11). These cases were chosen as representatives of the worst and best average solution improvement achieved by the fuzzy FMO methodology proposed compared to the

400 simple FMO solutions. DVHs for the 11 beam fuzzy FMO solution for all 10 cases are presented as supplemental material.





E) Right Parotid Mean Dose



F) Left Parotid Mean Dose

FIG. 9 Comparative Results Legend:

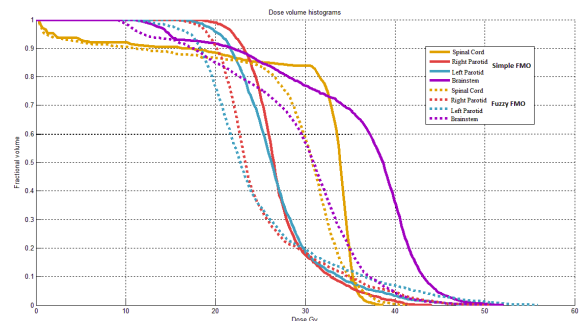
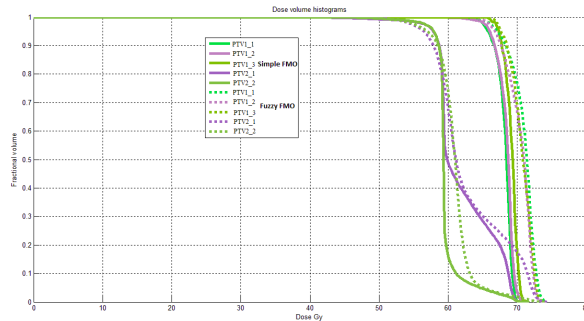
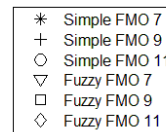


FIG. 10 DVHs for the 11 beam solution for Patient 1

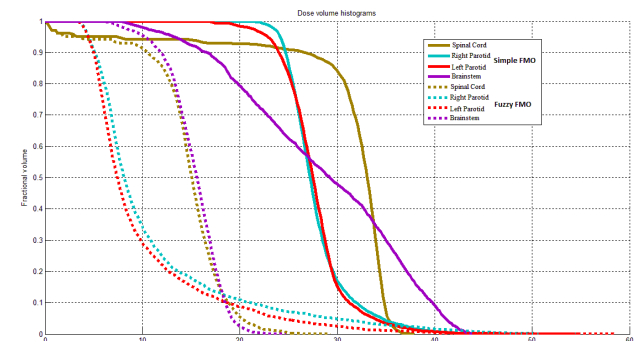
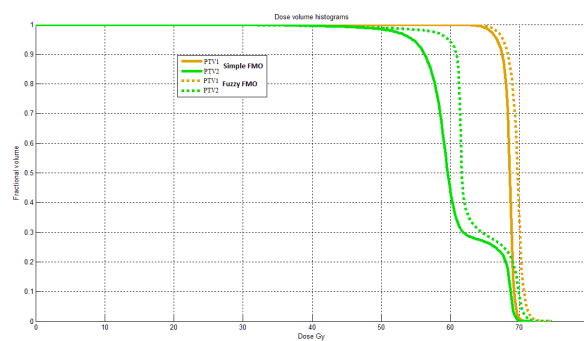


FIG. 11 DVHs for the 11 beam solution for Patient 3

## 4.DISCUSSION

In this paper we present a novel approach for FMO that releases the human planner from any parameter tuning or trial and error procedures. The proof of concept of this new approach was done resorting to ten head-and-neck cancer cases. The approach relies on a fuzzy inference

410 system that mimics the human planner reasoning within an iterative optimization procedure.  
Simple rules of the form *if the OAR is not spared enough then increase the importance of this*  
*OAR* are used to automatically update the FMO model's parameters.

It was possible to observe that the proposed automated FMO approach was able to generate  
admissible treatment plans, according to the defined constraints. The first admissible solution  
415 is generated within a short computational time window. The algorithm then tries to improve  
this calculated admissible solution, first guaranteeing a better PTV coverage, and then sparing  
as much as possible the OARs. The achieved results were compared with the equidistant  
solutions obtained by solving FMO with all parameters defined by the planner in a  
trial-and-error procedure. The Fuzzy FMO solutions present better dosimetry metrics than the  
420 simple FMO solution. This comparison can be considered unfair, since it is not possible to  
demonstrate that a real effort was made to tune the model's parameters for the simple FMO  
solution as much as possible. Actually, it is not trivial to compare trial-and-error and  
automated procedures, due to the intrinsic differences that exist in both approaches.

The proposed methodology presents as main advantage the fact that no human interaction is  
425 needed during treatment plan optimization. Prior to the beginning of the optimization, the  
planner only needs to define constraints based on the medical prescription and the priorities  
for each structure. The algorithm will automatically update both upper/lower bounds as well  
as weights (whenever necessary), guiding the search process into interesting regions where it  
is possible to find high quality solutions corresponding to acceptable treatment plans.

430 The computational results described considered FMO for step-and-shoot IMRT treatment  
planning. However, it will be possible to apply this same methodology for FMO applied to  
other treatment techniques, like VMAT. Future work will consider developments for VMAT,  
as well as the consideration of increased number of structures in the optimization process.  
Future work will also consider the use of similar fuzzy inference methodologies in IMRT  
435 beam angle optimization, especially for noncoplanar treatments.

## ACKNOWLEDGMENTS

This work was supported by FEDER. COMPETE. iCIS (CENTRO-07-ST24-FEDER-002003). Portuguese Foundation for Science and Technology under project grants  
440 UID/MULTI/00308/2013 and PTDC/EIA-CCO/121450/2010.

<sup>1</sup> Shepard, D.M., Ferris, M.C., Olivera, G.H., Mackie, T.R.: Optimizing the Delivery of Radiation Therapy to Cancer Patients. *SIAM Review*. 41, 721-744 (1999)

<sup>2</sup> Holder, A.: Designing Radiotherapy Plans with Elastic Constraints and Interior Point Methods. *Health Care Management Science*. 6, 5-16 (2003)

445 <sup>3</sup> Lim, G.J., Choi, J., Mohan, R.: Iterative solution methods for beam angle and fluence map optimization in intensity modulated radiation therapy planning. *OR Spektrum*. 30, 289-309 (2008)

<sup>4</sup> Lodwick, W.A., McCourt, S., Newman, F., Humphries, S.: Optimization Methods for Radiation Therapy Plans. In: C. Börgers, F. Natterer (eds) *Computational Radiology and Imaging: Therapy and Diagnostics*. pp. 229-249 (1999)

450 <sup>5</sup> Ólafsson, A., Wright, S.J.: Linear Programming Formulations and Algorithms for Radiotherapy Treatment Planning. *Optimization Methods and Software*. 21, 201-231 (2006)

<sup>6</sup> Preciado-Walters, F., Langer, M.P., Rardin, R.L., Thai, V.: Column generation for IMRT cancer therapy optimization with implementable segments. *Annals of Operations Research*. 148, 65-79 (2006)

455 <sup>7</sup> Romeijn, H.E., Ahuja, R.K., Dempsey, J.F., Kumar, A., Li, J.G.: A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning. *Physics in Medicine and Biology*. 48, 3521-3542 (2003)

<sup>8</sup> Lee, E.K., Fox, T., Crocker, I.: Integer Programming Applied to Intensity-Modulated Radiation Therapy Treatment Planning. *Annals of Operations Research*. 119, 165-181 (2003)

<sup>9</sup> Cheong, K., Suh, T., Romeijn, H., Li, J., Dempsey, J.: Fast Nonlinear Optimization with Simple Bounds for IMRT Planning. *Medical Physics*. 32, 1975-1975 (2005)

460 <sup>10</sup> Aleman, D.M., Kumar, A., Ahuja, R.K., Romeijn, H.E., Dempsey, J.F.: Neighborhood search approaches to beam orientation optimization in intensity modulated radiation therapy treatment planning. *Journal of Global Optimization*. 42, 587-607 (2008)

<sup>11</sup> Craft, D.L., Halabi, T.F., Shih, H.A., Bortfeld, T.R.: Approximating convex Pareto surfaces in multiobjective radiotherapy planning. *Medical Physics*. 33, 3399-3407 (2006)

465 <sup>12</sup> Sharpe, M.B., Moore, K.L., Orton, C.G.: Within the next ten years treatment planning will become fully automated without the need for human intervention. *Medical Physics*. 41, 120601 (2014)

<sup>13</sup> Breedveld, S., Storchi, P.R.M., Voet, P.W.J., Heijmen, B.J.M.: iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. *Medical Physics*. 39, 951-963 (2012)

470 <sup>14</sup> Rossi, L., Breedveld, S., Heijmen, B.J.M., Voet, P.W.J., Lanconelli, N., Aluwini, S.: On the beam direction search space in computerized non-coplanar beam angle optimization for IMRT prostate SBRT. *Physics in Medicine and Biology*. 57, 5441 (2012)

<sup>15</sup> Breedveld, S., Storchi, P.R.M., Keijzer, M., Heemink, A.W., Heijmen, B.J.M.: A novel approach to multi-criteria inverse planning for IMRT. *Physics in Medicine and Biology*. 52, 6339 (2007)

475 <sup>16</sup> Voet, P.W.J., Dirkx, M.L.P., Breedveld, S., Fransen, D., Levendag, P.C., Heijmen, B.J.M.: Toward fully automated multicriterial plan generation: a prospective clinical study. *International Journal of Radiation Oncology, Biology, Physics*. 85, 866-872 (2013)

<sup>17</sup> Zadeh, Lotfi A.: Fuzzy sets. *Information and control* 8, 338-353 (1965)

480 <sup>18</sup> Deasy, J.O., Blanco, A.I., Clark, V.H.: CERR: A computational environment for radiotherapy research. *Medical Physics*. 30, 979-985 (2003)