We thank the reviewer for the many useful comments. They have been incorporated into the revised draft, and we detail here how we have done so.

1- In section 2.1, the authors states that entropy is due to Harmancioglu (1981). This is incorrect. Concept of entropy evolves with Carnot and Boltzman. The entropy in information sense (which is used by the authors) is due to Shannon. This entire section should be carefully re-written and placed in the proper context of authors who have actually developed the fundamental notion. This is corrected as follows:

Shannon (1948) exploited the information theory to quantify the information loss in the transmission of a given message. The study was carried out in a communication channel and Shannon focused in physical and statistical constraints that limit the message transmission. Moreover, the measure does not addresses, in this way, the meaning of the message. In general, information entropy due to Shannon (1948) is a quota of uncertainty combined with a random variable. In the theory of communication and transmission of information Shannon (1948) had introduced a statistical concept of entropy. Therefore, entropy is the amount of information which are measured in bits and contained per average instance of a character in a stream of characters. Following Shannon (1948), Harmancioglu (1981) stated that entropy is a quantitative compute of the information content of a series of data since reduction of uncertainty, by making observations, equals the same amount of gain in information.

2-

Furthermore, for benefit of other researchers on Ocean Sciences, it could make sense to identify and refer to work of other researchers who have been dealing with entropy and chaotic signals in oceanography. This particularly concerns work of Sannasiraj, such as Enhancing tidal prediction accuracy in a deterministic model using chaos theory, Advances in water resources 27 (7), 761-77, 2004.

This corrected as follows:

The implementation of Entropy in dynamical model of tidal has been addressed in the study of Sannasiraj et al., (2004). In this_study Entropy and chaotic signals are used to enhance tidal prediction accuracy beside to the genetic algorithm which was used for the optimization of the local tidal model parameters.

Therefore, Marghany (2001) and Marghany and van Genderen, (2014) implemented entropy to determine the degree of uncertainty of random oil spill footprint discrimination in SAR satellite data. In a definition adopted from information theory, Cloude and Pottier, (1996), entropy is the numerical expression of random objects footprint boundaries in SAR images. In using this concept, oil spill footprint can be measured indirectly based on the degree of the reduction of multiplicative speckle noises and uncertainty of look-alike effects. The main hypothesis is the oil spill footprint boundaries have larger entropy compared to surrounding environment. Hence, in order to quantitatively assess the cumulative effect of uncertainty in oil spill footprint, entropy can be used as a metric for population diversity of oil spill footprint boundaries which are stored at each intersection of the column j and row i of the various slick areas. At the rear of Amorocho and Espildora, (1973) and Harmancioglu (1981); Magrghany and van Genderen (2014), the uncertainty (C) associated with the oil spill pixel value of x_i for a random variable X is then written as

3-

> Similarly, multi-objective optimisation cannot be associated with Coello et al. (2002). There are quite a few other authors whose work on multi-objective optimisation precedes Coello. For example,
>
> Ching-Lai Hwang; Abu Syed Md Masud (1979). Multiple objective decision making, methods and applications: a state-of-the-art survey. Springer-Verlag. ISBN 978-0-387-09111-2.
>
> Kaisa Miettinen (1999). Nonlinear Multiobjective Optimization. Springer. ISBN 978-0-7923-8278-2.
>
> In context of evolutionary algorithms, the main reference would be:
>
> Kalyanmoy Deb (2001), Multi-Objective Optimization Using Evolutionary Algorithms, Wiley
>
> This is very important, since in this section the authors also seem to be claiming that Marghany (2014b) and Gunawan (2004) have introduced main optimisation concepts in multi-objective sense, which is obviously not the case. I would suggest that author carefully reviews this section and recognise the original authors.

This is corrected by adding the suggested references in text as follows:

Comprehending Hwang and Masud (1979); Miettinen (1999);Deb (2001); Coello et al., (2002), the multi-objective optimization (MOP) has already been successfully adopted to solve uncertainty of object detection in SAR images as shown in Marghany (2014a) and (2014b) studies. In general, MOP consists of *n* decision variable parameters, *k* objective functions and *m* constraints (Deb 2001). Multi-objective Optimization (Deb 2001) aims at conducting optimization for a range of functions as follows:

Following Deb (2001), Marghany (2014b), used entropy based MOEA E-MOEA for the optimization of oil spill detection from SAR data. In this regard, the entropy of oil spill footprint boundaries must be coded into a Genetic Algorithm syntax form i.e. the chromosome form. In this problem, the chromosome consists of a number of genes where every gene corresponds to a coefficient in the n^{th}-order surface fitting polynomial

4-

On page 1273, authors explains that he is combining all objectives into single objective. The author describes that weights are chosen randomly and objectives

added. This is sub-optimal procedure. In turn, the most widely accepted multi-objective algorithm in evolutionary computing is so-called non-dominated sorting Genetic Algorithm. See for example, A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II K Deb, S Agrawal, A Pratap (2000) - http://repository.ias.ac.in/83498/

This is state-of-the-art algorithm with numerous open source freely available implementations that would as a result provide full Pareto front. I would suggest author to explore performance of NSGA instead of relying on de-facto single objective description of the problem.

This is corrected by adding new section as follows:

2.2.3. Non-dominated Sorting Genetic Algorithm NSGA-II

This section presents a brief description of NSGA-II relevant to this study. NSGA-II is the second version of the famous "*Non-dominated Sorting Genetic Algorithm*" based on the work of Prof. Kalyanmoy Deb for solving *non-convex* and *non-smooth* single and multi-objective optimization problems. Its main features are: (i) A sorting non-dominated procedure where all the individual are sorted according to the level of non-domination; (ii) It implements elitism which stores all non-dominated solutions, and hence enhancing convergence properties; (iii) It adapts a suitable automatic mechanics based on the crowding distance in order to guarantee diversity and spread of solutions; and (iv) Constraints are implemented using a modified definition of dominance without the use of penalty functions.

Perhaps, there is not exist one best solution in the case of multiple objectives. Therefore, there exists a set of solutions which are superior to rest of solutions in the search space when all objectives are considered but are inferior to other solutions in the space in one or more objectives. These solutions are known as Pareto-optimal solutions or nondominated solutions.

The efficiency of NSGA lies in the way multiple objectives are reduced to dummy fitness function using nondominated sorting procedures. Consequently, NSGA can solve practically any number of objectives. In this regard, this algorithm can handle both minimization and maximization problems.

In order to sort a population of size N for $E(\beta_1),...............,E(\beta_N)$ according to the level of non-domination, each solution *m* must be compared with every other solution in the population to find if it is dominated. This requires comparisons $O(E(\beta_m))_N$ for each solution, where is *m* is the number of different pixels belong to oil spill, look-alikes, and sea roughness, and low wind zones.

The initialized population N of $E(\beta_1),...............,E(\beta_N)$ is sorted based on the level of non-domination. Let *S* is each solution which must be compared to other every solution to determine the level of domination. In this regard, the fast sort algorithm was given by Deb et al., (2000) can be explored in oil spill automatic detection as follows:

for each individual $E(\beta_1)$ in main population *P* do the following

Initialize $S_{E(\beta_1)} = \Phi$. This set Φ would include all the individuals of $E(\beta_n)_N$ which is being dominated by $E(\beta_1)$.

Initialize $n_{E(\beta_1)} = 0$. This would be the number of individuals that dominate $E(\beta_1)$ i.e. no individuals dominate $E(\beta_1)$ then $E(\beta_1)$ belongs to the first front; set rank for individual $E(\beta_1)$ to one i.e. $E(\beta_1)_{rank} = 1$.

for each individual *m* in *P*

if $E(\beta_1)$ dominated *m* then
. add *m* to the set Φ i.e. $\Phi = \Phi \bigcup \{m\}$

*else if *m* dominates $E(\beta_1)$ then
. increment for domination counter for $E(\beta_1)$ i.e. $n_{E(\beta_1)} = n_{E(\beta_1)} + 1$
Let the first front set F_1 and then update by adding $E(\beta_1)$ to front 1 i.e. $F_1 = F_1 \bigcup \{E(\beta_1)\}$

Initialize the front counter to one. *i*=1
Then $F_i \neq \Phi$
Let $Q \neq \Phi$. The set for sorting the individuals for $(i+1)^{th}$ front
for each individual $E(\beta_1)$ in front F_i

For every individuals *m* in $S_{E(\beta_1)}$ ($S_{E(\beta_1)}$ is the set of individuals dominated by $E(\beta_1)$)

. $n_{E(\beta_1)} = n_{E(\beta_1)} - 1$, decrement the domination count for individual *m*.

. if $n_{E(\beta_1)}$=0 then none of the individuals in the subsequent fronts would dominate *m*. Hence set $E(\beta_1)_{rank} = i+1$. Update the set Q with individual *m* i.e. $Q = Q \bigcup m$.

-increment the front by one.

-Now the set Q is the next front and hence $F_i = Q$.

2.2.3.1. Crowding Distance

Following Deb et al., (2000), the moment the non-dominated sort is achieved the crowding distance is designated. All the individuals in the population are assigned as crowding distance value since the individuals are selected based on rank and crowding distance. Crowding distance is assigned front wise and comparing the crowding distance between two individuals in different front is meaning less. The crowing distance is estimated as follows:

- For each front F_i, the number of individuals is represented by *N*.
  - Reset the distance d_j to be zero for all the individuals of $E(\beta_1)$ i.e. $F_i(d_j) = 0$,where j corresponds to j^{th} individual of $E(\beta_j)$ in front F_i.
  - For every objective function f

    *Sort the $E(\beta_j)$ in front F_i based on objective f i.e. $E(\beta_j)$=sort (F_i, f).

    *Assign infinite distance to boundary values for each individual $E(\beta_j)$ in F_i i.e.

$$E(\beta_{d_1}) = \infty \text{ and } E(\beta_{d_n}) = \infty$$

*for K= 2 to *(n-1)*

. $E(\beta_{d_K}) = E(\beta_{d_K}) + \dfrac{E(\beta)(K+1).q - E(\beta)(K-1).q}{f_q^{\max} - f_q^{\min}}$

. $E(\beta_{d_K}).q$ is the value of q^{th} objective function of the K^{th} individual in $E(\beta_{d_K})$

The main concept behind the crowing distance is estimating the Euclidian distance between each individual $E(\beta_j)$ in a front F_i which is based on their q objectives in the q dimensional hyper space. The individuals $E(\beta_j)$ in the boundary are always selected since they have infinite distance assignment.

Selection. Once the individuals $E(\beta_j)$ are sorted based on non-domination and with crowding distance $E(\beta_{d_K})$ assigned, the selection is carried out using a crowded comparison operator $\prec_n$ which is based on

(1) non-domination rank $E(\beta_1)_{rank}$ i.e. individuals $E(\beta_j)$ in front F_i will have their rank as $E(\beta_1)_{rank}$ =i.

(2) crowding distance $E(\beta_{d_K})$

- $E(\beta_1) \prec_n m$
  - $E(\beta_1)_{rank} < m_{rank}$
  - or if $E(\beta_1)$ and m belong to the same front F_i then $F_i(E(\beta_{d_K}) > F_i(d_m)$ i.e. the crowing distance should be more.

The individuals $E(\beta_1)$ are chosen by exercising a binary contest selection with crowed comparison-operator $\prec_n$. Following Deb (2001),the point with lower rank of $E(\beta_1)_{rank} < m_{rank}$ is preferred between two solutions. Else the point that is included in region with less number of $E(\beta_1)$ points is selected. Therefore, the diversity with non-dominated solutions is presented by using the crowding comparison procedure which is used in the tournament selection and during the population reduction phase. Since solutions compete with their crowding distance.

2.2.3.2. Recombination and Selection.

The offspring population is merged with the current generation population and variety is completed to set the individuals of the next generation. Elitism is confirmed, subsequently all best individuals are included in the population. In this context, population is now sorted based on non-domination. Subsequently, the new generation is filled by each front till the population size surpasses the existing population size. For instance, the population exceeds *N* when adding all the individuals in front F_i then the individuals in front F_i is chosen based on their crowding distance in the descending order until the population size is *N*.

In results section we add figure 6b and figure 7b for NGSA-II.

Fig. 6. shows the output result of E-MMGA and NSGA-II. Clearly, E-MMGA is able to produce four different segmentation boundaries. However, NSGA-II can produce sharper segmentation boundaries than E-MMGA. In NSGA-II algorithm, oil spill footprint discriminated and identified by sharp vector that separates it from surrounding features i.e., sea surface, look-alikes and land boundaries (Fig.6b). Besides, Fig. 7a shows that the thick oil spill footprint has highest E-MMGA value of 2 than medium and light oil spill. Nevertheless, NSGA-II is able to produce different clusters of oil spill footprint thickness as compared to E-MMGA with highest value of NSGA-II of 2.5. This indicates that NSGA-II can identify clearly the level oil spill footprint spreading accurately than E-MMGA.

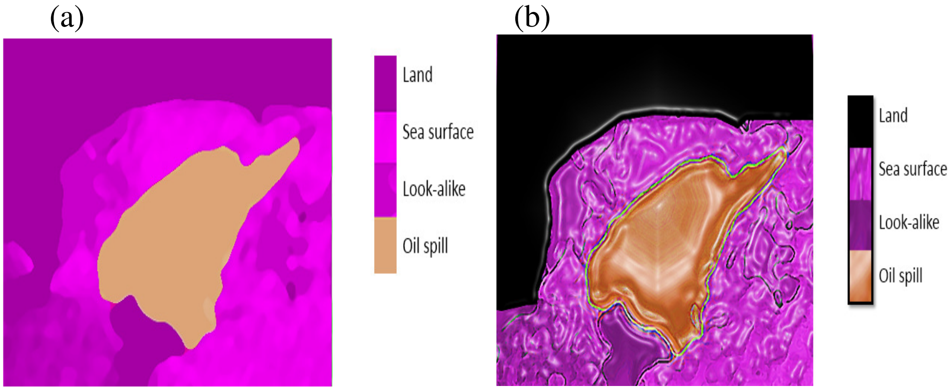

Figure 6. Optimization solutions for oil spill discrimination in COSMO-SkyMed using (a) E-MMGA and (b) NSGA-II.

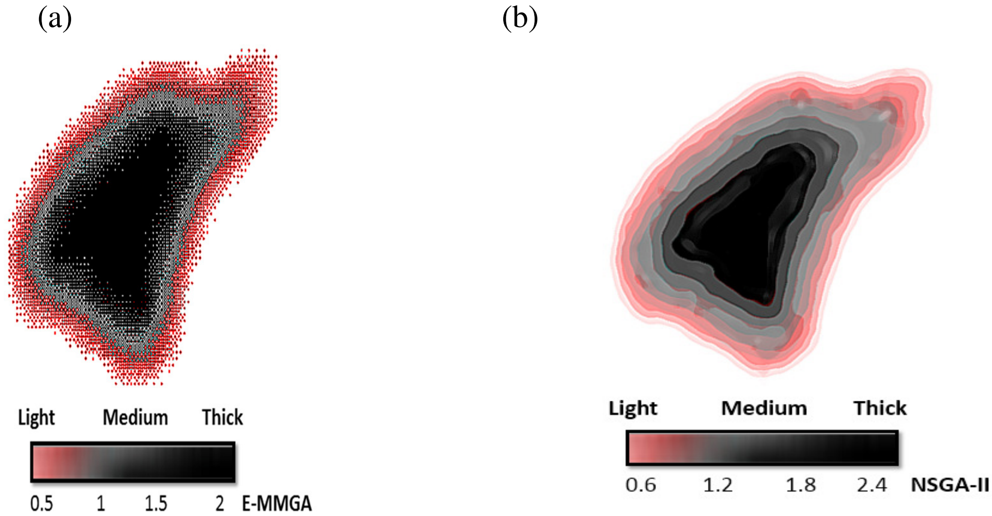

Figure 7. Oil spill footprint Categories by (a) E-MMGA and (b) NSGA-II.

Figure 8 illustrates the nondominated solution of different algorithms. From Figures 8 , it is clear that the solution of NSGA-II (Figure 8b) is much better than, Entropy, and E-MMGA. Further, Entropy solution is far from real Pareto front while, the solution of E-MMGA is gathered around the center of the Pareto front. Under this circumstance, E-MMGA tends to concentrate in one part of the Pareto front. On the other hand, NSGA-II maintained high degrees of diversity of their solutions during the searching of best optimal solution for either oil spill footprint detection or oil spill spreading level in COSMO-SkyMed satellite data. In this regard, the NSGA-II is able to better distribute its population along the obtained front than Entropy and E-MMGA.

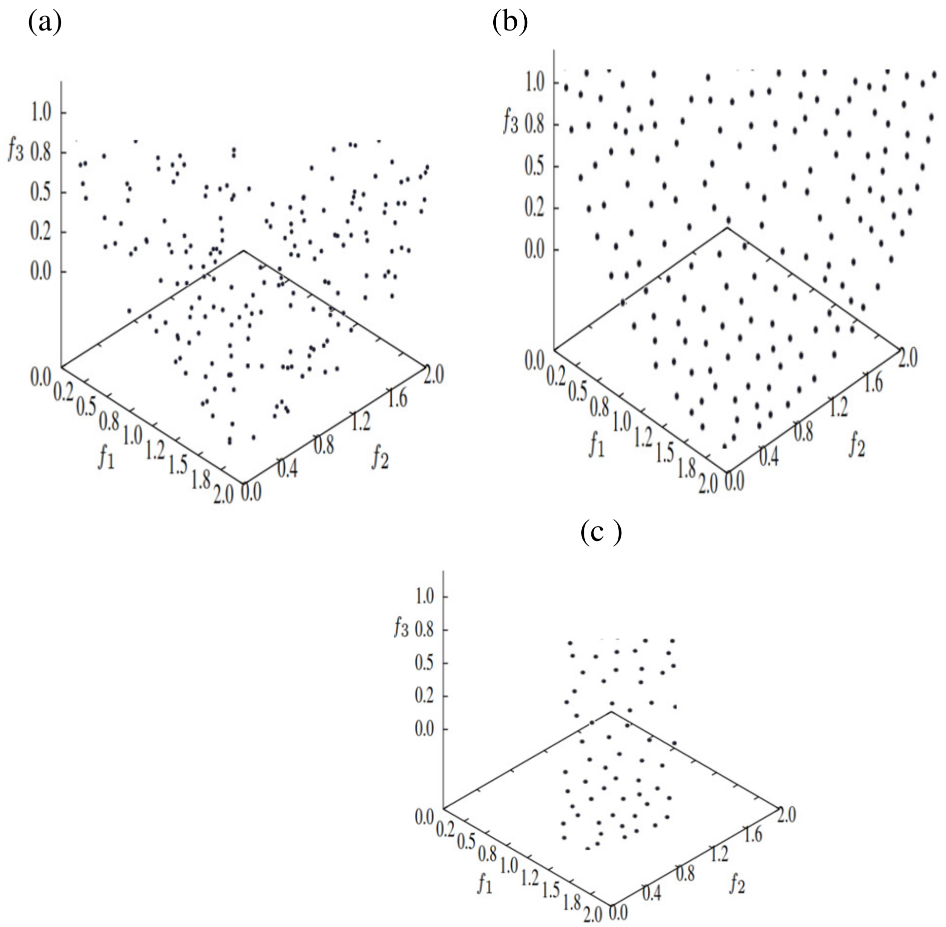

Figure 8. Final Nondominated solutions by using (a) E-MMGA, (b) NSGA-II and (c) Entropy

This is mainly because each multi-objective function in NSGA-II tends to bias its population towards the extreme edges of the Pareto frontier. This is confirms the work was done by Deb et al., (2001). Compared to Entropy algorithm and E-MMGA, NSGA-II is able to identify the look-alike footprint boundaries and discriminate accurately between, oil spill and look-alike, and surrounding sea surface with standard error of 0.04 and fastest computing time of 65 sec (Table 2). NSGA-II can accurately identify the sharpest morphological boundary of oil spill and assigned by different segmentation layer in COSMO-SkyMed satellite data as compared to Entropy algorithm and E-MMGA. In fact, NSGA-II provides a set of compromised solutions called Pareto optimal solution since no single solution can optimize each of the objectives separately. The decision maker is provided with the set of Pareto optimal solutions in order to choose solution based on the decision maker's criteria. This sort of NSGA-II solution technique is called nondominanted since decision is taken after searching is finished. This confirms the work done by Deb (2000) and Deb et al., (2001) In this context, the Pareto-optimization approach does not require any a priori preference decisions between the conflicting of oil spill, look-alike, land, and surrounding sea footprint boundaries. Further, Pareto-optimal points have form Pareto-front as shown in Fig. 8 in the multi-objectives function of the COSMO-SkyMed data space. Finally, NGSA-II has advantages on Entropy and E-MMGA because (i) NGSA-II explicit diversity preservation mechanism;(ii) overall complexity of NSGA-II is at most $O(MN^2)$ and;(iii) elitism does not allow an already found Pareto optimal solution to be deleted. This agreed with Deb et al., (2001).

Table 2. Accuracy performance of different algorithms

Algorithms	**Iterations**	**Time (sec)**	**Standard error**
Entropy	200	240	1.2
E-MMGA	700	140	0.89
NSGA-II	1200	65	0.04

References

Ching-Lai Hwang; Abu Syed Md Masud (1979). Multiple objective decision making, methods and applications: a state-of-the-art survey. Springer-Verlag. ISBN 978-0387-09111-2.

Deb,K. (2001), Multi-Objective Optimization Using Evolutionary Algorithms, Wiley.

Deb, K.; Agrawal, S.; Pratap, A.; Meyarivan, T. A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II; In Parallel Problem Solving from Nature-PPSN VI, Proceeding of 6th International Conference, Paris, France, 18–20 September 2000; Schoenauer, M., Deb, K., Rudolph, G., Yao, X., Lutton, E., Merelo, J.J., Schwefel, H.P., Eds.; Springer: Berlin, Germany, 2000; Volume 1917, pp. 849–858.

Miettinen,K. (1999). Nonlinear Multiobjective Optimization. Springer. ISBN 978-07923-8278-2.

Sannasiraj, S. A., Zhang, H., Babovic, V., & Chan, E. S. (2004). Enhancing tidal prediction accuracy in a deterministic model using chaos theory. *Advances in water resources*, *27*(7), 761-772.

Shannon, C. (1948) A mathematical theory of communication. Bell System Tech. J. 27, 379-423. Available online: http://cm.belllabs.com/cm/ms/what/shannonday/shannon1948.pdf (accessed on 4 December 2015).