phasized in coevolution. For a fitness function emphasizing non-overlapping time spans, coevolution emphasized longer time spans.

In Figure 7, examining overall spread of the maximum catch difference in each network, we can see that for the overlap-favored fitness function the coevolutionary GA actually identified the lowest average differences throughout the networks. Coupled with the fact that it produced the highest modularity networks (Figure 3), we can conclude that coevolutionary GA traded off large differences for higher overlap in years. Paired with shorter time spans considered (recall Figure 6), this trade off would naturally not produce the most interesting networks for fisheries experts. For the non-overlap favored network, in terms of overall spread of the data, PAMDGA consistently produced large maximum average differences within its network when noting reach of whisker and outliers (but its results in this respect are not statistically different than GA around the median and are generally comparable to GA). However, for non-overlapping time spans, PAMDGA dominates the other algorithms when considering both longer time span lengths (Figure 6) and large average differences (Figure 7) together. By maximizing both the time span length and catch difference metrics, PAMDGA produces high modularity networks (Figure 3) that were chosen as most useful by a fisheries expert for the geographical area under study.

6. EXPERT EXAMINATION OF VISUALIZATIONS

A fisheries expert, the fourth author, examined the best networks located by all algorithms (GA, SA, Coevolutionary GA, and PAMDGA). He was asked to rate each network edge of the best networks. Network edges were visualized using GTdiff as two temporal bins and one corresponding difference graph. The first two grids are the temporal bins in GTdiff, and display average catch in kilograms in each spatial grid element. The two temporal bins are ordered sequentially based on the last year of the time spans; if the last year is the same, the temporal bins are ordered by the first year. The colour scale spans from light yellow (lowest average catch) to brown (largest average catch). The third grid is the difference graph, where the difference in average catch between the two time spans is displayed as a positive (green) or negative (red) change. No change in catch is represented by white, and the degree of saturation of green and red is used to represent positive and negative differences, respectively. The visualization tool was used to change the resolution of the final networks to 10 x 10. The resolution does not change the content of the networks and allowed for viewing of trends by the expert and presentation in publication rather than examination on the high resolution display used in the study.

During the time period examined (1980 to 2005), there were known anomalies (large differences in catch over time): Fisheries scientists reported that the population levels of cod dropped suddenly in the early 1990s, which prompted a moratorium on cod fisheries starting in 1992. In addition, other smaller changes known to fisheries scientists occurred during particular years. The three options for the rating of each difference graph by the expert were: No (meaning no difference relevant to fisheries scientists appeared), Relevant (a difference relevant to fisheries scientists appeared), or Salient (a special case of Relevant indicating that an important biological shift was identified). The ratings of each of the difference graphs (individual edges) in the final best network of each algorithm out of 50 trials are summarized in Table 1 as the number of responses for each rating and total differences rated.

Our expert found that for almost all results for every algorithm using the overlap favored fitness function, differences were not emphasized appropriately. In particular, there were many instances of small changes in catch between two time spans with a very large degree of overlap in the final networks. Despite this, the expert indicated that PAMDGA provided the largest number of relevant differences, with both coevolutionary algorithms significantly outperforming the standard GA and SA algorithms.

The expert found that for non-overlap favored fitness, SA had the largest proportion of relevant differences. However, there were only 10 differences in the best SA network. The expert also indicated that none of the differences located in the SA were of particular interest; they simply presented known overall trends with no particular anomalous changes identified. In contrast, PAMDGA presented the highest proportion of salient anomalies in the data. PAMDGA gleaned from the dataset a remarkably higher percentage of salient differences than any other algorithm (3/16, or 18.75%). We present all three difference graphs for PAMDGA that represented salient differences for the expert in Figure 8.

The top difference graph in Figure 8 clearly shows the decline of cod stocks from the decade leading up to the moratorium of 1992 to 1993 on northern cod compared to a sample later year in the future. In the middle graph of Figure 8, the darker red across the region indicates that cod stocks dropped considerably after the years 1983 to 1991, which were prior to the moratorium on cod (1992 to 1993). This difference graph nicely contrasts those previous years of abundance with post moratorium levels of 1994 to 1997. In the bottom difference graph of Figure 8, years mostly covering the time period of the moratorium (1991 to 1994) are compared to years following the moratorium by several years (2000 to 2004). The interesting aspect of these data for the expert was the red in the difference graph depicting the decline in cod stocks post-moratorium in the northeast of the grid with the exception of a place southwest of Newfoundland (light green portion of grid). This area actually corresponds to known areas where levels of cod during the moratorium were lower than the current levels.

Table 1: Ranking of Difference Graphs

Overlap Favored				
	No	Relevant	Salient	Differences
GA	5	2	0	7
SA	6	1	0	7
CoEvGA	6	11	0	17
PAMDGA	2	5	0	7
Overlap Not Favored				
	No	Relevant	Salient	Differences
GA	10	6	1	17
SA	3	7	0	10
CoEvGA	6	6	1	13
PAMDGA	6	7	3	16

7. CONCLUSIONS

This work described the application of coevolutionary GA algorithms to the simultaneous discovery of problem solution (interesting network containing anomalous large changes in catch data) and underlying community structure. A large real-world network based on geospatial fisheries catch data over a 25 year span in Atlantic Canada was used as the basis for performance comparisons. The fitness function for the analysis of the networks used modularity (the Q metric) with a community membership function that either did or did not favor overlapping of communities to emphasize different relationships between time periods. Four search algorithms were compared: GA, SA, standard coevolutionary GA, and PAM DGA (refined coevolutionary GA). The best networks found by all algorithms were examined using a prototype visualization tool designed for fisheries scientists.

Results indicated that the coevolutionary algorithm produced superior Q-based fitness results for overlap-favored fitness. However, the fisheries expert indicated that non-overlap favoring fitness provided more interesting final networks. With respect to the non-overlap favored fitness function, PAMDGA provided high fitness networks that combined both large differences and extended time spans in its chosen networks to a greater degree than the other algorithms. Upon examination of the final networks by the expert, he found that PAMDGA located the largest number of interesting known trends in the catch data.

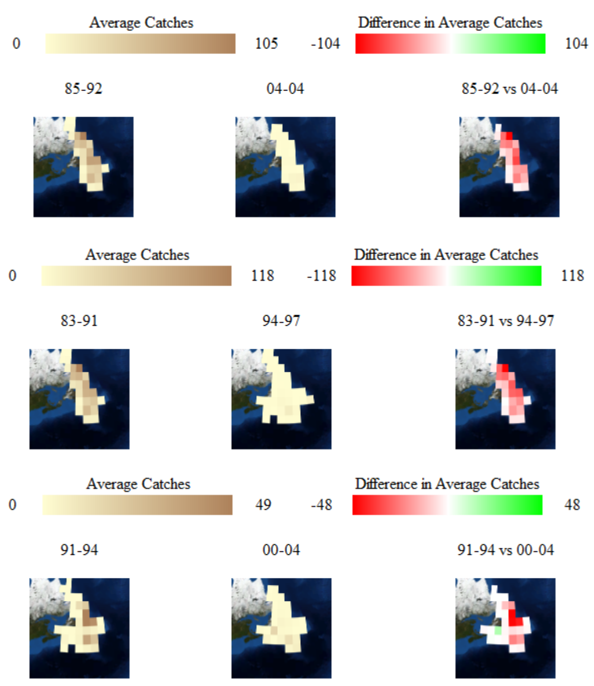

Figure 8: Salient difference graphs selected by expert from the highest Q, no overlap favored network produced by PAMDGA. Catches are shown in thousands of kg.

8. ACKNOWLEDGMENTS

The authors wish to thank Fisheries and Oceans Canada (DFO) for making available the data used in the case study. This work was supported by a Strategic Projects Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) held by O.H., R.D., and W.B.

9. REFERENCES

[1] A. Gog, D. Dumitrescu, and B. Hirsbrunner. Community detection in complex networks using collaborative evolutionary algorithms. In *ECAL'07: Proceedings of the 9th European Conference on Advances in Artificial Life*, pages 886–894, Berlin, Heidelberg, 2007. Springer-Verlag.

[2] S. Gregory. Fuzzy overlapping communities in networks. Technical Report CSTR-10-008, University of Bristol, October 2010.

[3] O. Hoeber, G. Wilson, S. Harding, R. Enguehard, and R. Devillers. Exploring geo-temporal differences using gtdiff. In *2011 IEEE Pacific Visualization Symposium (PacificVis)*, pages 139 –146, New York, USA, 2011. IEEE Press.

[4] X. Liu, D. Li, S. Wang, and Z. Tao. Effective algorithm for detecting community structure in complex networks based on ga and clustering. In *ICCS '07: Proceedings of the 7th International Conference on Computational Science, Part II*, pages 657–664, Berlin, Heidelberg, 2007. Springer-Verlag.

[5] M. E. J. Newman. Analysis of weighted networks. *Phys. Rev. E*, 70(5):056131, Nov 2004.

[6] V. Nicosia, G. Mangioni, V. Carchiolo, and M. Malgeri. Extending the definition of modularity to directed graphs with overlapping communities. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03024, 2009.

[7] M. Tasgin and H. Bingol. Community detection in complex networks using genetic algorithm. In *ECCS '06: Proc. of the European Conference on Complex Systems*, Apr. 2006.

[8] G. Wilson, S. Harding, O. Hoeber, R. Devillers, and W. Banzhaf. Detecting anomalies in spatiotemporal data using genetic algorithms with fuzzy community membership. In *ISDA 2010: 10th International Conference on Intelligent Systems Design and Applications*, pages 97–102, Chennai, India, 2010. Research Publishing Services.

[9] G. Wilson and M. Heywood. Introducing probabilistic adaptive mapping developmental genetic programming with redundant mappings. *Genetic Programming and Evolvable Machines*, 8:187–220, June 2007.