

pled with some input parameter settings, are we adequately exploring our program space?" While we have not answered that question here, we believe COSMOS provides a foundation for how to answer it. Past work has focused on describing problem landscapes, we accept that they are vast and possibly unknowable. We instead focus on determining how much of that space a given system with a given set of parameters can know. We also realize that performance metrics will vary depending upon the desired outcome. Fundamentally, we are not seeking one metric to rule them all. Instead we recognize the plurality of not only solutions, but systems.

7. ACKNOWLEDGMENTS

The authors would also like to thank Alexandre Passos for his statistical expertise and the CI Lab members at Hampshire College and Theory Lab members at the University of Massachusetts for their feedback. This work is supported in part by the National Science Foundation under NSF grants #CNS-0619337 and #1017817. Any opinions, findings, conclusions or recommendations expressed here are the author and do not necessarily reflect those of the sponsors.

8. REFERENCES

- [1] P. J. Angeline. An investigation into the sensitivity of genetic programming to the frequency of leaf selection during subtree crossover. In *Proceedings of the First Annual Conference on Genetic Programming*, GECCO '96, pages 21–29, Cambridge, MA, USA, 1996. MIT Press.
- [2] D. F. Barrero, B. Castaño, M. D. R-Moreno, and D. Camacho. Statistical distribution of generation-to-success in gp: application to model accumulated success probability. In *Proceedings of the 14th European conference on Genetic programming*, EuroGP'11, pages 154–165, Berlin, Heidelberg, 2011. Springer-Verlag.
- [3] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Press, 2nd edition, June 2001.
- [4] D. Costelloe and C. Ryan. On improving generalisation in genetic programming. In *Proceedings of the 12th European Conference on Genetic Programming*, EuroGP '09, pages 61–72, Berlin, Heidelberg, 2009. Springer-Verlag.
- [5] J. M. Daida, D. S. Ampy, M. Ratanasavetavadhana, H. Li, and O. A. Chaudhri. Challenges with verification, repeatability, and meaningful comparison in genetic programming: Gibson's magic. In *Proceedings of the Genetic and Evolutionary Computation Conference*, volume 2 of *GECCO '99*, pages 1851–1858, Orlando, FL, USA, 1999. Morgan Kaufmann.
- [6] J. M. Daida, R. R. Bertram, S. A. Stanhope, J. C. Khoo, S. A. Chaudhary, O. A. Chaudhri, and J. A. I. Polito. What makes a problem gp-hard? analysis of a tunably difficult problem in genetic programming. *Genetic Programming and Evolvable Machines*, 2(2):165–191, June 2001.
- [7] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2nd edition, Nov. 2001.
- [8] B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.
- [9] J. R. Koza. *Genetic programming: On the programming of computers by means of natural selection*. Complex adaptive systems. MIT Press, 1993.
- [10] S. Luke. Ecj : An evolutionary computation research system in java.
- [11] S. Luke. When short runs beat long runs. In L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H. Garzon, and E. Burke, editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, pages 74–80, San Francisco, California, USA, 7-11 July 2001. Morgan Kaufmann.
- [12] S. Luke and L. Panait. Is the perfect the enemy of the good? In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '02, pages 820–828, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.
- [13] J. Niehaus and W. Banzhaf. More on computational effort statistics for genetic programming. In *Proceedings of the 6th European conference on Genetic programming*, EuroGP'03, pages 164–172, Berlin, Heidelberg, 2003. Springer-Verlag.
- [14] N. Paterson and M. Livesey. Performance comparison in genetic programming. In *Late Breaking Papers at the 2000 Genetic and Evolutionary Computation Conference*, pages 253–260, Las Vegas, Nevada, USA, 2000.
- [15] N. P. Troutman, B. E. Eskridge, and D. F. Hougen. Is "best-so-far" a good algorithmic performance metric? In *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, GECCO '08, pages 1147–1148, New York, NY, USA, 2008. ACM.
- [16] M. Walker, H. Edwards, and C. Messom. Confidence intervals for computational effort comparisons. In *Genetic Programming. Proceedings of the 10th European Conference, EuroGP 2007*, 2007.
- [17] M. Walker, H. Edwards, and C. Messom. The reliability of confidence intervals for computational effort comparisons. In *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, GECCO '07, pages 1716–1723, New York, NY, USA, 2007. ACM.
- [18] M. Walker, H. Edwards, and C. Messom. Success effort and other statistics for performance comparisons in genetic programming. In *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*, pages 4631–4638, September 2007.