

Paul E. Anderson, Ph.D.

Curriculum Vitae

Associate Professor
Director, M.S. and B.S. Data Science & Analytics
Department of Computer Science
University of Charleston (College of Charleston)
66 George St., Charleston, SC 29424

andersonpe2@cofc.edu
(843) 953-8151
<http://anderson-data-science.com>

EDUCATION

Ph.D., Computer Science and Engineering, Wright State University (WSU) (GPA 4.0)	2010
Advisor: Michael L. Raymer	
Dissertation: Algorithmic techniques employed in the quantification and characterization of nuclear magnetic resonance spectroscopic data	
M.S., Computer Science, Wright State University (GPA 4.0)	2006
Advisor: Michael L. Raymer	
Thesis: A computational framework for analyzing chemical modification and limited proteolysis experimental data used for high confidence protein structure prediction	
B.S., Computer Engineering, Wright State University, Summa Cum Laude (GPA 3.85)	2004

ACADEMIC EXPERIENCE

M.S. Data Science and Analytics	Director	2018+
College of Charleston	Associate Professor	2017+
B.S. Data Science Program	Director	2012+
College of Charleston	Assistant Professor	2011 - 2017
Graduate Program in Marine Biology, College of Charleston	Adjunct Faculty Member	2015+
Consortium of Universities (CRFP)	Postdoctoral Fellow	2010 – 2011
Wright State University (WSU)	Adjunct Professor	2007 – 2011
Consortium of Universities (CRFP)	Graduate Research Fellow	2009 – 2010
Wright State University	Graduate Research Assistant	2004 – 2009

OTHER PROFESSIONAL EXPERIENCE

Air Force Research Laboratory	Bioinformatician & Computational Scientist	2009-2011
Forensic Bioinformatics	Bioinformatics Software Engineer	2008
LexisNexis	Capacity and Performance Engineer	2001 – 2004
Sterling Commerce	Software Engineering Intern	1999 – 2000

FELLOWSHIPS, SCHOLARSHIPS, AND AWARDS

Nominated for Distinguished Research Award	2017
Nominated for William V. Moore Teacher Scholar Award	2016
C. Richard Crosby Distinguished Teaching Co-Chair	2015+
Nominated for Distinguished Research Award	2014
Nominated for Faculty of the Year, School of Science and Mathematics, College of Charleston, ExCEL Awards	
Nominated for Faculty of the Year, School of Science and Mathematics, College of Charleston, ExCEL Awards	2013
Nominated for Faculty of the Year, School of Science and Mathematics, College of Charleston, ExCEL Awards	2012
Nominated for President Leo I. Higdon Outstanding Leadership Award, College of Charleston, ExCEL Awards	2012
Consortium of Universities Postdoctoral Fellow	2010 – 2011
Highly competitive national postdoctoral fellowship. Formed as a partnership between the Consortium of Universities and several Department of Defense agencies: http://www.consortium-research-fellows.org/	
Wright State University President's Award for Excellence (Outstanding Collaborative Units)	2010
Consortium of Universities Graduate Research Fellow	2009
Best student presentation at Ohio Collaborative Conference on Bioinformatics (2 nd place)	2007
Dayton Area Graduate Studies Research Fellowship (DAGSI)	2006 – 2009
Highly competitive and merit-based research fellowship. Full three-year award covering tuition and stipend.	
Nominee for Excellence in Teaching Award	2005 – 2009
Graduate Student Excellence Award	2006

HONORS

Phi Kappa Phi Honor Society	2002+
President of Tau Beta Pi, Ohio Mu Chapter	2002
Golden Key International Honor Society	2001+
National Engineering Honor Society (Tau Beta Pi)	2001+

RESEARCH INTERESTS

Data science, data mining, machine learning, bioinformatics, biomedical informatics, computational science, biostatistics, e-science, cloud computing, evolutionary computation, pattern recognition, biological modeling, metabolomics, proteomics, transcriptomics, genomics, and computer science education.

* Indicates an undergraduate student

+ Indicates a graduate student

BRIEF BIO

Dr. Paul Anderson graduated in 2004 from Wright State University with a B.S. degree in Computer Engineering. He received his Master of Computer Science in 2006 and his Ph.D. in Computer Science & Engineering in June 2010. After graduation, Dr. Anderson was awarded a Consortium of Universities Research Fellowship to study as a Bioinformatics and Computational Research Scientist for the Air Force Research Laboratory (AFRL). Dr. Anderson has published 30+ peer-reviewed articles in the fields of data science, data mining, machine learning, computational intelligence, metabolomics, genomics, e-Science, bioinformatics, cloud computing, biomedical informatics, cancer informatics, and computer science & engineering education.

At present, Dr. Anderson is an associate professor in the Computer Science Department at the University of Charleston in South Carolina historical known as the College of Charleston. He is the director of the B.S. in Data Science (first such program in the country) and the M.S. in Data Science and Analytics (first such program in South Carolina). Dr. Anderson directs the Data Science Research Group at the College of Charleston where his team specializes in developing and applying data mining, machine learning, and artificial intelligence to the fields of bioinformatics, genomics, biomedical informatics, and metabolomics. Complementary to this applied work, Dr. Anderson's lab is interested in the foundations of data science with a current focus on explainable ranking algorithms and the rankability of data. He currently has multidisciplinary projects underway in genomics, metabolomics, toxicology, marine biology, cancer informatics, and medical and marine genomics. He is the director of the Institute for Graph Analytics and Research in Data Science (IGARDS) - a collaborative interdisciplinary institute aimed at broadening foundational data science research (<http://igards.github.io>). Dr. Anderson served as the primary investigator for the Omics NSF Research Experience for Undergraduates at the College of Charleston (<http://omics.cofc.edu>).

COLLABORATORS

Dr. Amy Langville
Department of Mathematics, College of Charleston

Dr. Annalisa Calini
Department of Mathematics, College of Charleston

Dr. Frank Marí and Dr. Benjamin Neely
National Institute of Standards and Technology

Dr. Jihad Obeid
Center for Biomedical Information at the Medical
University of South Carolina

Dr. Dennis Watson, Dr. Robert Wilson, Dr. Starr
Hazard, and Dr. Patricia Watson
Center for Medical Genomics at the Medical University
of South Carolina

Dr. Andrew Shedlock
Department of Biology, College of Charleston

Dr. Robert Chapman
Hollings Marine Laboratory

Dr. Fran VanDolah and Lab
National Oceanic and Atmospheric Administration

Dr. Michael Peterson
University of Hawaii at Hilo

Dr. Nicholas Reo and Magnetic Resonance Lab
Wright State University

Dr. Nicholas DelRaso and Dr. Deirdre Mahle
Air Force Research Laboratory

Grier Allen (CEO)
BoomTown ROI

Edward Murray (CTO)
Ceterus

* Indicates an undergraduate student

+ Indicates a graduate student

TEACHING

DISC/DATA 101	Introduction to Discovery Informatics/Data Science
DATA 210	Data Set Organization and Management*
DATA 495	Data Science Capstone
CSCI 220	Computer Programming I
CSCI 220L	Computer Programming I Lab
CSCI 221	Computer Programming II
CSCI 250	Computer Organization
CSCI 334	Data Mining
CSCI 470	Principles of Artificial Intelligence
CSCI 320	Programming Languages
CSCI 462	Software Engineering Practicum
CSIS 604	Distributed Computing
CSIS 638	Advanced Topics in Database Systems
CITA 295	Computer in the Arts Seminar
CITA 495	Computer in the Arts Capstone
BIOL 453L	Genomics
HONS 380	Bioinformatics

TEACHING – INDIVIDUAL ENROLLMENT COURSES

The number in parentheses represents the number of times the individual enrollment course has been taught.

CSCI 399 (29)	Tutorial/independent study
CSCI 499 (11)	Honors year-long bachelor's essay
DISC 495 (5)	Data Science Capstone (when taught as individual enrollment)
CSIS 698 (6)	Graduate Thesis
CSCI 299 (2)	Tutorial/independent study
CSCI 397 (2)	Independent study
CSIS 699 (3)	Graduate independent study
DISC 397 (1)	Independent study
CSIS 691 (1)	Graduate independent study

* Indicates an undergraduate student

+ Indicates a graduate student

JOURNAL AND CONFERENCE ARTICLES IN SUBMISSION OR REVISION

Langville, A., Chartier, T., and **Anderson, P.** “The Rankability of Data” in revision at SIAM Journal on Mathematics of Data Science (SIMODS).

Bearden, D., Yost, J., Casu, F., Gaylord, G., Leffler, J., Denson, M., **Anderson P.**, Barrows F., Sandifer, P.. “Investigation of Graded Levels of Soybean Meal Diets for Red Drum, *Sciaenops ocellatus*, using Quantitative PCR Derived Biomarkers” in revision at Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics.

PEER-REVIEWED JOURNAL AND CONFERENCE ARTICLES

Newman, T.* and **Anderson, P.**, Alignment-Based Topic Extraction Using Word Embedding. In *Workshop on Interactive Adaptive Learning*, 2018 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (pp. 60-73).

Blair, W.⁺, Davis, L.*, & **Anderson, P. E.** (2017). Streamlining the genomics processing pipeline via named pipes and persistent Spark datasets. In *Bioinformatics and Bioengineering (BIBE)*, 2017 IEEE 17th International Conference on, pp. 35-38. IEEE, 2017.

Turner, C. A.⁺, Jacobs, A. D.*, Marques, C. K. *, Oates, J. C., Kamen, D. L., **Anderson, P. E.**, & Obeid, J. S. (2017). Word2Vec inversion and traditional text classifiers for phenotyping lupus. *BMC Medical Informatics and Decision Making*, 17(1), 126.

Neely, M. G., Morey, J. S., Anderson, P., Balmer, B. C., Ylitalo, G. M., Zolman, E. S., ... others. (2017). Skin Transcriptomes of common bottlenose dolphins (*Tursiops truncatus*) from the northern Gulf of Mexico and southeastern US Atlantic coasts. *Marine Genomics*.

Morey, J. S., Neely, M. G., Lunardi, D., **Anderson, P. E.**, Schwacke, L. H., Campbell, M., & Van Dolah, F. M. (2016). RNA-Seq analysis of seasonal and individual variation in blood transcriptomes of healthy managed bottlenose dolphins. *BMC Genomics*, 17(1), 720. <http://doi.org/10.1186/s12864-016-3020-8>

Kopel, S.*, Fluette, K., Glen, G., & **Anderson, P. E.** (2016). Fastfood Elastic Net: Combining Variable Selection with Kernel Expansion Approximations. In *International Workshop on Machine Learning, Optimization and Big Data* (pp. 427–432).

Labuzzetta, C. J. *, Antonio, M. L. *, Watson, P. M., Wilson, R. C., Laboissonniere, L. A., Trimarchi, J. M., ... **Anderson, P. E.** (2016). Complementary feature selection from alternative splicing events and gene expression for phenotype prediction. *Bioinformatics*, 32(17), i421--i429.

Parthasarathy, G.⁺, Olmsted, A., & Anderson, P. (2016). Natural language processing pipeline for temporal information extraction and classification from free text eligibility criteria. In *Information Society (i-Society), 2016 International Conference on* (pp. 120–121).

Paul, M. R. *, Levitt, N. P. *, Moore, D. E. *, Watson, P. M., Wilson, R. C., Denlinger, C. E., Watson, D.K., and **Anderson, P. E.** (2016). Multivariate models from RNA-Seq SNVs yield candidate molecular targets for biomarker discovery: SNV-DA. *BMC Genomics*, 17(1), 1–19. doi:10.1186/s12864-016-2542-4

DelRaso, N. J., Harville, D. L., Chamberlain, M. L., **Anderson, P. E.**, Sibomana, I., Raymer, M. L., and Reo, N.V. (2016). Urinary Metabolite Profiles May be Predictive of Cognitive Performance under Conditions of Acute Sleep Deprivation. *Current Metabolomics*, 4(1), 63-77.

Sogin, E. M.⁺, Putnam, H. M., **Anderson, P. E.**, & Gates, R. D. (2016). Metabolomic signatures of increases in temperature and ocean acidification from the reef-building coral, *Pocillopora damicornis*. *Metabolomics*, 12(4), 1-12.

* Indicates an undergraduate student

⁺ Indicates a graduate student

-
- Moore, D. E.*, Fluette, K. A. *, Milne, H. J. *, Shedlock, A. M., & **Anderson, P. E.** (2015). ccKOPLS: Confounder-correcting kernel-based orthogonal projections to latent structures. In *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on* (pp. 897–903). doi:10.1109/BIBM.2015.7359803
- Anderson, P. E.**, Nash, T. *, & McCauley, R. (2015). Facilitating Programming Success in Data Science Courses Through Gamified Scaffolding and Learn2Mine. In *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education* (pp. 99–104).
- Johnson, J. G., Paul, M. R.*, Kniffin, C. D., **Anderson, P. E.**, Burnett, L. E., & Burnett, K. G. (2015). High CO₂ alters the hypoxia response of the Pacific whiteleg shrimp (*Litopenaeus vannamei*) transcriptome including known and novel hemocyanin isoforms. *Physiological genomics*, 47(11), 548-558.
- Sullivan, C. V, Chapman, R. W., Reading, B. J., & **Anderson, P. E.** (2015). Transcriptomics of mRNA and egg quality in farmed fish: Some recent developments and future directions. *General and Comparative Endocrinology*, 221, 23–30.
- Sogin EM⁺, **Anderson P**, Williams P, Chen CS, Gates RD. Application of ¹H-NMR metabolomic profiling for reef-building corals. *PloS one* 9.10 (2014): e111274.
- Anderson, P.**, Turner, C.,* Dierksheide, J.* and McCauley, R. An Extensible Online Environment for Teaching Data Science Concepts through Gamification, in *Proceedings of the 44th Annual Frontiers in Education Conference (FIE'14)*, October 22-25, 2014, Madrid, Spain, pp 1336 - 1343.
- Scheiber M.N., Watson P.M., Rumboldt T., Stanley, C.*, Wilson, R.C., Findlay V.J., **Anderson P.E.**, and Watson D.K.. (2014). FLII expression is correlated with breast cancer cellular growth, migration, and invasion and altered gene expression. *Neoplasia*, 16(10), 801–813. PMID: 25379017; PMC4212256
- Anderson, P.E.**, Paul, M.R.*, McCaffrey, V.A.*, Wilson, R., Hazard, E.S., Denlinger, C.E., Watson, P.M., Watson, D.K. (2014) Predictive Modeling of Lung Cancer Recurrence using Alternative Splicing Events versus Differential Expression Data. In *Proceedings of IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp 1 - 8.
- Anderson, P.**, Bowring, J., McCauley, R., Pothering, G., Starr, C. (2014) An Undergraduate Degree in Data Science: Curriculum and a Decade of Implementation Experience. in *Proceedings of 45th SIGCSE*, pp. 145 – 150, 2014.
- Turner, C.*, Dierksheide, J.*, & **Anderson, P.** (2014) Learn2Mine: Data Science Practice and Education through Gameful Experiences. *International Journal of E-Education, E-Business, E-Management and E-Learning*, 4(3), 243–248. doi:10.7763/IJEEEE.2014.V4.338
- P. Anderson**, E. Pharr⁺, and M. Peterson (2013) Providing Cloud-based Metabolite Annotations for NMR Spectroscopic Data with Semantic Web Integration, in *Proceedings of the ISCA 5th International Conference on Bioinformatics and Computational Biology (BiCOB)*, pp. 237–244.
- C. D. Ruark, C. E. Hack, P. J. Robinson, **P. E. Anderson**, and J. M. Gearhart (2013) Quantitative structure--activity relationships for organophosphates binding to acetylcholinesterase, *Archives of toxicology*, vol. 87, no. 2, pp. 281–289.
- E. Pharr⁺ and **P. Anderson** (2012) gLIMS: Google Cloud-based Laboratory Information Management System for NMR Metabolomics, in *Proceedings of 3rd Annual ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB)*, pp. 595–597.
- P. Anderson** and M. Peterson (2012) Empirical Metabolite Identification via GA Feature Selection and Bayes Classification, in *Proceedings of 3rd Annual ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB)*, pp. 591–593.

* Indicates an undergraduate student

+ Indicates a graduate student

-
- A. Ranabahu⁺, **P. Anderson**, and A. Sheth (2011) The Cloud Agnostic e-Science Analysis Platform, IEEE Internet Computing, *IEEE Internet Computing*, vol. 15, no. 6, pp. 85–89.
- Manjunatha, A.⁺, **Anderson, P.**, Ranabahu, A.⁺, and Sheth, A. (2011). Identifying and Implementing the Underlying Operators for Nuclear Magnetic Resonance based Metabolomics Data Analysis. *In Proceedings of 3rd International Conference on Bioinformatics and Computational Biology (BICoB)*, New Orleans, LA.
- Gunaratna, K., **Anderson, P.**, Ranabahu, A., and Sheth, A (2010). A Study in Hadoop Streaming with Matlab for NMR data processing. *In Proceedings of 2nd IEEE International Conference on Cloud Computing (Cloudcom)*, Indianapolis, IN.
- Mahle, D. A., **Anderson, P. E.**, DelRaso, N. J., Raymer, M. L., Neuforth, A. E., and Reo, N. V. (2010). A Generalized Model for Metabolomic Analyses: Application to Dose and Time Dependent Toxicity. *Metabolomics*, 7(2), 206-216.
- Anderson, P. E.**, Sahoo, S. S., Manjunatha, A., Ranabahu, A. H., DelRaso, N. J., Reo, N. V., Sheth, A. P., and Raymer, M. L. (2010). Cloud-based Map-Reduce Architecture for Nuclear Magnetic Resonance based Metabolomics. *Proceedings of the 7th Microsoft Research eScience Workshop*.
- Anderson, P. E.**, Mahle, D. A., Raymer, M. L., Reo, N. V., DelRaso, N. J., and Doom, T. E. (2011). Dynamic Adaptive Binning: Dynamic quantification of nuclear magnetic resonance spectroscopic data. *Metabolomics*, 7(2), 179-190.
- Anderson P. E.**, M. L. Raymer , Kelly, B. J., Reo, N. V., DelRaso, N. J., and Doom, T. E. (2009), Characterization of ¹H NMR spectroscopic data and the generation of synthetic validation sets, *Bioinformatics*, 25, 2992 – 3000.
- Anderson, P.E.**, Reo, N. V., DelRaso, N. J., Doom, T. E., and Raymer, M. L. (2008) Gaussian binning: A new kernel-based method for processing NMR spectroscopic data for metabolomics, *Metabolomics*, 4:3, 261 – 272.
- Kelly, B. J., **Anderson, P. E.**, Reo, N. V., DelRaso, N. J., Doom, T. E., and Raymer, M. L. (2007). A proposed statistical protocol for the analysis of metabolomic toxicological data derived from NMR spectroscopy. *In Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2007)*, volume II, pages 1414 – 1418, Boston, Massachusetts.
- Anderson, P. E.**, Raiford, D. W., Sweeney, D. J., Doom, T. E., and Raymer, M. L. (2005). Stochastic model of protease-ligand reactions. *In Proceedings of the 5th IEEE Symposium on Bioinformatics and Bioengineering (BIBE 2005)*, pages 306–310, Minneapolis, Minnesota.

* Indicates an undergraduate student

⁺ Indicates a graduate student

KEYNOTE AND CONTRIBUTED PRESENTATIONS

Panel Speaker, Technological advancements that are transforming the healthcare industry, the challenges and future directions, BIOINFORMATICS (BIOSTEC), the 11th International Joint Conference on Biomedical Engineering Systems and Technologies

NSF Data Science Education Webinar (2017)

NSF Workshop on Data Science Education (October 2015)

Keynote Speaker, Lewis and Clark Big Data Workshop (May 2015)

Plenary Talk, "Data Science and Big Data – Can it change science and medicine?" South Carolina Junior Academy of Science (2014)

University of Hawaii, Manoa, "Invited Speaker on Integrative Omics," University of Hawaii, Manoa. (May 2014).

CONFERENCE & WORKSHOP PRESENTATIONS, POSTERS, AND ABSTRACTS (SUBSET)

Newman, T.* and **Anderson, P.** (2018), Alignment-Based Topic Extraction Using Word Embedding. *Workshop on Interactive Adaptive Learning at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD)*.

Blair, W.⁺, Davis, L.*, & **Anderson, P. E.** (2017). Streamlining the genomics processing pipeline via named pipes and persistent Spark datasets, *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*.

Kopels, S.*, Fluette K.*, Glen, G.*, and **Anderson, P.** (2016) Fastfood Elastic Net: Combining Variable Selection with Kernel Expansion Approximation. 2nd International Workshop on Machine Learning (MOD 2016).

Labuzzetta, C.J.*, Antonio, M. L. *, Watson, P.M., Wilson, R. C., Trimarchi, J.M., Watson, D.K., and **Anderson, P.E.** (2016) Complementary Feature Selection from Alternative Splicing Events and Gene Expression for Phenotype Prediction. *In Computational Biology (ECCB), 2016 Annual European Conference on*.

Labuzzetta, C.*, Antonio, M.*, and **Anderson, P.E.** (2016) MVPq: A novel filtering method for isoform-based features from alternative splicing events enhances phenotype prediction accuracy via machine learning techniques in RNA-Seq experiments. 20th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2016).

Antonio, M.*, Labuzzetta, C.*, and **Anderson, P.E.** (2016) *METRO: A pipeline for comparative meta-analysis and visualization of gene and pathway expression in RNA-Seq experiments*. 20th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2016).

Emilia M. Sogin, Hollie M. Putnam, Craig Nelson, **Paul Anderson** and Ruth D. Gates, Coral Associated Microbial Partners Influence Hologiont Metabolite Profiles. *International Conference for Reef Studies (ICRS 2016)*.

Turner, C.A.* , **Anderson, P.E.**, Oates, J.C., Kamen, D.L., Obeid, J.S. Improving Lupus Phenotyping Using Natural Language Processing and Machine Learning. *2015 Summit on Translational Bioinformatics. March 2015*. Information Available at: <http://www.amia.org/jointsummits2015/tbi-call-participation>

Johnson, J. G., Paul, M.*, Kniffin, C. D., **Anderson, P. E.**, Burnett, L. E., & Burnett, K. G. (2015). Deep sequencing of the hepatopancreas transcriptome reveals new isoforms of hemocyanin and their regulation in response to low O₂/high CO₂ in the Pacific whiteleg shrimp, *Litopenaeus vannamei*. In *INTEGRATIVE AND COMPARATIVE BIOLOGY* (Vol. 55, pp. E90–E90).

Moore, David E.*; Fluette, Kellan A.*; Milne, Heather J.; Shedlock, Andrew M.; **Anderson, Paul E.**,

* Indicates an undergraduate student

⁺ Indicates a graduate student

"ccKOPLS: Confounder-correcting kernel-based orthogonal projections to latent structures," *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*.

- Anderson, P. E.**, Nash, T.*, & McCauley, R. (2015, June). Facilitating Programming Success in Data Science Courses through Gamified Scaffolding and Learn2Mine. *2015 ACM Conference on Innovation and Technology in Computer Science Education*.
- Jacobs, A.* and **Anderson P.E.** (2015). Galaxy Classification Using Deep Belief Networks. *2015 School of Science and Mathematics Research Symposium*. Award of Merit.
- Moore, D.* and **Anderson P.E.** (2015). ccOPLS: Confounder Correcting Orthogonal Projections to Latent Structures. *2015 School of Science and Mathematics Research Symposium*. Award of Merit.
- Fluette, K.* and **Anderson P.E.** (2015). Extending scalable deep neural networks to multi-way classification. *2015 School of Science and Mathematics Research Symposium*. Award of Merit.
- Hurlburt D. and **Anderson P.E.** (2015). Comparison of Effectiveness of Convolutional Neural Networks. *2015 School of Science and Mathematics Research Symposium*.
- Anderson, P.**, Bowring, J., McCauley, R., Pothering, G., Starr, C., (2014). "An Undergraduate Degree in Data Science: Curriculum and a Decade of Implementation Experience." *Presentation and Proceedings of 45th SIGCSE, Atlanta, GA*.
- Turner, C.*, Dierksheide, J.*, & **Anderson, P.**, (2014). "Learn2Mine: Data Science Practice and Education through Gameful Experiences." *International Conference on E-Education, E-Business, E-Management and E-Learning, Toronto, Canada*.
- Emilia M. Sogin, Hollie Putnam, **Paul Anderson**, and Ruth Gates (2014). Variation in Coral Metabolite Production After Exposure to Global Climate Change Stressors is Species Specific. Presentation and abstract: *Ocean Sciences Meeting, Honolulu, HI*.
- Sogin E.M., **Anderson P.**, Horgen D., Williams P. and Gates R.D. (2014) Differential Expression of Betaines Links to Performance Limits in Corals. Presentation and Abstract: *Western Society of Naturalists, Oxnard/Ventura, CA*.
- Anderson, P.** and Peterson, M. (2013). NMR-based Metabolite Identification by Iteratively Updating Evolved Bayesian Networks. Poster and abstract: *21st Annual International Conference on Intelligent Systems for Molecular Biology (ISMB), Berlin, Germany*.
- P. Anderson**, E. Pharr⁺, and M. Peterson, "Providing Cloud-based Metabolite Annotations for NMR Spectroscopic Data with Semantic Web Integration," Presentation and Paper: *ISCA 5th International Conference on Bioinformatics and Computational Biology (BICOB-2013), 2013, Honolulu, HI*.
- Turner, C.*, and **Anderson. P.E.** (2012). Learn2Mine: An Open-Source Cloud-based Informatics Platform for Integrated Teaching and Data Exploration. Poster and Short Paper: *8th IEEE International Conference on eScience, Chicago, IL*.
- Pharr, E.⁺ and **Anderson, P.E.** (2012). gLIMS: Google Cloud-based Laboratory Information Management System for NMR Metabolomics. Poster and Paper: *3rd Annual ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB), Orlando, FL*.
- Anderson, P.** and Peterson, M. (2012). Empirical Metabolite Identification via GA Feature Selection and Bayes Classification. Poster and Paper: *3rd Annual ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB), Orlando, FL*.
- Morgan, J. C.*, Chapman, R. W., and **Anderson, P.E.** (2012). A Next Generation Sequence Processing and Analysis Platform with Integrated Cloud-storage and High Performance Computing Resources. Poster and abstract: *3rd Annual ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-*

* Indicates an undergraduate student

⁺ Indicates a graduate student

BCB), Orlando, FL.

- Turner, C.* and **Anderson, P. E.** (2012). An open cloud-based informatics platform for integrated teaching and data exploration. *Poster and abstract: National Conference on Undergraduate Research*. Ogden, UT.
- Hack, C. E., Fleming, E. J., **Anderson, P. E.**, Meghan K. Makley, Peter J. Robinson, Mark W. Lisanby, and Gearhart, J. M. (2011). Mechanistically-based Computational Approaches to Predicting Host Immune Response to Biological Warfare Agents. Presentation and abstract: *Aerobiology in Biodefense IV*. Richmond VA.
- Fleming, E. J., **Anderson, P. E.**, Meghan K. Makley, Peter J. Robinson, Gearhart, J. M., and Hack, C. E. (2011). Mechanistically-based Computational Model of the Host Immune Response to Inhaled Type A and Type B *Francisella tularensis* Strains. Poster and abstract: *Aerobiology in Biodefense IV*. Richmond VA.
- Hack, C. E., Fleming, E. J., **Anderson, P. E.**, and Gearhart, J. M. (2011). Mechanistically-based computational model of the host immune response to biological warfare agents: Application to tularemia. Poster and abstract: *48th Annual Society of Toxicology*. Washington, DC.
- Hack, C. E., Fleming, E. J., **Anderson, P. E.**, Makley, M. K., Robinson, P. J., and Gearhart, J. M. (2011). Mechanistically-based computational model of the host immune response to biological warfare agents: Application to tularemia. Poster and abstract: *ASM Biodefense and Emerging Diseases Research Meeting*. Washington, DC.
- Fleming, E. J., Makley, M. K., Robinson, P. J., Lisanby, M. W., **Anderson, P. E.**, Gearhart, J. M., and Hack, C. E. (2010). Mechanistically-Based Computational Model of the Host Immune Response to Biological Warfare Agents: Application to Tularemia. Presentation and abstract: *Chemical and Biological Defense Science and Technology Conference*. Orland, FL.
- Gunaratna, K., **Anderson, P.**, Ranabahu, A., and Sheth, A (2010). A Study in Hadoop Streaming with Matlab for NMR data processing. Presentation, abstract, and paper: *2nd Annual IEEE International Conference on Cloud Computing (Cloudcom)*, Indianapolis, IN.
- Manjunatha, A., Ranabahu, A., **Anderson, P.** and Sheth, A. (2010). A study of generating customized data intensive scientific workflows with domain specific language. Poster, abstract, and paper: *2nd Annual IEEE International Conference on Cloud Computing (Cloudcom)*, Indianapolis, IN.
- Anderson, P. E.**, Sahoo, S. S., Manjunatha, A., Ranabahu, A. H., DelRaso, N. J., Reo, N. V., Sheth, A. P., and Raymer, M. L. (2010). Cloud-based Map-Reduce Architecture for Nuclear Magnetic Resonance based Metabolomics. Presentation, abstract, and paper: *7th Microsoft Research eScience Workshop*.
- Anderson, P. E.**, Ruark, C. D., Hack, C. E., Robinson, P. J., and Gearhart, J. M. (2010). A comparison of QSAR Multivariate Statistical Regression Techniques to Predict Chemical Warfare Agent Biologically-Based Dose-Response Model Parameters. Poster and abstract: *Chemical and Biological Defense Science and Technology Conference*. Orland, FL.
- Manjunatha, A., **Anderson, P.**, Sahoo, S., Ranabahu, A., Raymer, M., and Sheth, A. (2010). Cloud Based Scientific Workflow for NMR Data Analysis. Poster and abstract: *18th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB)*, Boston, MA.
- Manjunatha, A., **Anderson, P.**, Sahoo, S., Ranabahu, A., Raymer, M., and Sheth, A. (2010). Semantically Annotated RESTful Services for Large-scale Metabolomics Data Analysis. Poster and abstract: *Ohio Collaborative Conference on Bioinformatics (OCCBIO)*, Columbus, OH.
- Mahle, D. A., **Anderson, P. E.**, Gearhart, J. M., and Reo, N. V. (2010). Effect of Low Level Exposure to Diisopropylfluorophosphate (DFP) on Regional Brain Metabolism in F344 Rats. Poster and abstract: *49th annual Society of Toxicology Conference (SOT)*, Salt Lake City, Utah.

* Indicates an undergraduate student

+ Indicates a graduate student

-
- Anderson, P. E.**, Maynard, C. M., Hodson, N. R., Kelly, B. J., Reo, N. V., DelRaso, N. J., Doom, T. E., and Raymer, M. L. (2009). A web-based framework for the distribution of bioinformatics techniques: Orthogonal projection on latent structures and principal component analysis implemented as RESTful web services. Poster and abstract: *OCCBio Collaborative Conference on Bioinformatics (OCCBIO)*, Cleveland, OH.
- Anderson, P. E.**, Reo, N. V., DelRaso, N. J., Doom, T. E., and Raymer, M. L. (2008). Gaussian binning: A new kernel-based method for processing NMR spectroscopic data for metabolomics. Presentation and abstract: *11th annual Force Health Protection Conference (FHP)*, Albuquerque, New Mexico.
- Reo, N. V., Neuforth, A. E., Couch, W., Raymer, M. L., **Anderson, P. E.**, Mahle, D. A., and DelRaso, N. J. (2008). A Time and Dose Response Metabonomics Study of D-Serine Toxicity in Rats. Poster and abstract: *47th annual Society of Toxicology Conference (SOT)*, Seattle, Washington.
- Kelly, B. J., **Anderson, P. E.**, Reo, N. V., DelRaso, N. J., Doom, T. E., and Raymer, M. L. (2007). Comparison of Statistical Techniques for the Analysis of Metabolic Toxicological Data Derived from NMR Spectroscopy. Presentation, abstract, and paper: *Ohio Collaborative Conference on Bioinformatics (OCCBIO)*, Miami, OH.
- Anderson, P. E.**, Mitchell, L., Alter, G. M., and Raymer, M. L. (2006). Protein structure probing through limited proteolysis. Poster and abstract: *Ohio Collaborative Conference on Bioinformatics (OCCBio)*, Athens, Ohio.
- Anderson, P. E.**, Raiford, D. W., Sweeney, D. J., Doom, T. E., and Raymer, M. L. (2005). Stochastic model of protease-ligand reactions. Poster, abstract, and paper: *5th IEEE Symposium on Bioinformatics and Bioengineering (BIBE 2005)*, Minneapolis, Minnesota.
- Presented at the Ohio Graduate Student Symposium on Computer and Information Science & Engineering (2005). Topic was *Protein Structure Prediction assisted by limited proteolysis*.

GRADUATE RESEARCH PROJECTS/THESES ADVISED

- Characterization of coral metabolome under different pH and temperature regimes, via 1H NMR. Julie Loewenstein, Master's Thesis, 2017 – 2019 (expected)
- Streamlining the genomics processing pipeline via persistent Spark datasets*. Walter Blair, Master's Thesis, 2016-2018.
- Big Data Genomics: A Spark Distributed Computing Framework*. Tom Nash⁺, Master's Thesis, 2016-2017.
- Characterization of growth anomalies in *Porites compressa* using elemental analysis and metabolomics*. Erik Andersson, Master's Thesis, 2016-2017.
- Natural Language Processing Pipeline for Temporal Information Extraction and Classification from Free Text Eligibility Criteria*. Gayathri Parthasarathy, Master's Project 2016.
- Biomedical Informatics Phenotyping using Machine Learning and Natural Language Processing*. Clayton Turner⁺, Master's Thesis, 2016.
- gLIMS: Google Cloud-based Laboratory Information Management System for NMR Metabolomics*. Edward Pharr⁺, Master's Thesis, 2012 - 2014.

UNDERGRADUATE RESEARCH AND HONORS PROJECTS ADVISED

- Biogeography and Computational Analysis of Zeta proteobacteria*. Katherine Duchinski*, 2018
- Natural Language Processing and Word Representation for Classification*. Sarah Wiegrefe, 2017
- Optimization Methods for SVMs*. Kaya Tollas*, 2016.
- Algorithms in Bioinformatics*. Luca Carvalho De Oliveira* and Leonardo De Melo Joao*, 2016.

* Indicates an undergraduate student

+ Indicates a graduate student

- Computational Transcriptions Software Development*. Katherine Duchinski*, 2016.
- Deep Neural Networks and Kernel Approximation Methods*. Sonia Kopels*, 2016.
- Natural Language Processing and Word Embedding*. Matheus Rodrigues Freitag Sr*, 2016.
- Creating a Feature Selection Pipeline Using Apache Spark*. Cassios Marques*, 2015.
- Data Science in Econometrics*, Pasha Sadeghian*, 2015.
- Advanced Algorithms for Data Science Research*, Alexander Jacobs*, 2015
- Data Science Research: Support Vector Machines*, David Moore*, 2015
- Exchange Rate Prediction based on interest rates*, Daniel Hurlburt*, Honors Thesis, 2015
- Deep Learning Kernel Methods*, Kellan Fluette*, 2015
- Predictive Forecasting of Operating Room Efficiency*, Sarah Wiegrefe*, 2015.
- Algorithms in Data Science: Convolution Deep Belief Networks*, Daniel Hurlburt*, 2015
- Metabolomics Cyberinfrastructure Development*, Kellan Fluette*, 2014
- Learn2Mine - An Integral Learning Environment*, Clayton Turner*, Honors Thesis, 2014
- Honors Thesis - Bioinformatics Algorithm Development for Genomics-enabled Research in the Life Sciences*.
Victoria McCaffrey*, Honors Thesis, 2014.
- Bioinformatics Infrastructure Development for Cancer Genomics*. Matt Paul*, 2013 - 2014
- Prediction and Annotation of Genomic Repeat Dynamics using Hidden Markov Models*. Kelsey Yetsko*, Fall
2013 - 2015.
- Exploring Expression Variability in Lung Cancer Patients through Deep-Learning Dimensionality Reduction*.
Victoria McCaffrey* and Matt Paul*, Summer 2013-2014.
- ETS Transcription Factor Biomarker Identification*. Connor Stanley*, Summer 2013 – Summer 2014.
- Next Generation Bioinformatics Algorithms and Cyberinfrastructure*. Victoria McCaffrey*, Matt Paul*, Connor
Stanley*, and Jeremy Morgan*.
- Bifrost: A Life Sciences Data Connector for Cloud Services*. Jeremy Morgan*, 2013 - 2014.
- Midgard: A Next Generation Sequence Processing and Analysis Platform with Integrated Cloud-storage and
High Performance Computing Resources*. Jeremy Morgan*, Summer 2012 – Summer 2014.
- Predicting Protein Secondary Structure*. Jeremy Morgan* Independent Study, Spring 2012.
- Learn2Mine: An open cloud-based informatics platform for integrated teaching and data exploration*. Clayton
Turner*, Spring 2012 - 2014.
- Cloud Computing: Making Clouds More Reachable – A Domain Specific Language Approach for Complex
Cloud Centric Applications*, Ashwin K. Manjunatha⁺, M.S. Thesis, 2011.
- Cloud-computing map-reduce implementations of NMR analysis tools*, Brandon Egbert*, Honors Project, 2010.
- A metabolite NMR identification algorithm and toolkit*, Zachary Markwell*, 2010.
- Comparison of NMR spectral alignment techniques*, Katie Timmerman*, Honors Project, 2010.
- Spectral Viewer: A java based NMR spectral quantification and analysis tool*, Christopher J. Maynard*, 2009.
- A web-based framework for the distribution of bioinformatics techniques*, Christopher J. Maynard* and Nathan R.
Hodson*, 2009.

* Indicates an undergraduate student

+ Indicates a graduate student

GRANTS

- Anderson P.** (PI), “Lung Cancer Relapse Prediction using Alternative Splicing Biomarkers”, SC Bioinformatics Pilot Project (BIPP) Program, amount \$10,000, 2017
- Anderson P.** (PI), “Towards FutureCast”, Ceterus, amount \$15,000, June 2017 – August 2017
- Anderson P.** (PI), “Deep Learning Image Classification”, BoomTown ROI, amount \$61,165, June 2016 – August 2016
- Anderson P.** (PI), “Red Drum Sequencing and Bioinformatics”, SC Department of Natural Resources/SC Sea Grant Consortium, amount \$5,000, January 2015 –May 2015
- Anderson P.** (PI), “Computational Genomics Support”, National Oceanic and Atmospheric Administration, amount \$16,000, August 2013 – September 2016
- Anderson P.** (PI), “Electronic Health Medical Recording Phenotyping and Machine Learning”, Medical University of South Carolina, amount \$25,778, July 2016 – June 2017.
- Duchinski, K. and **Anderson P.**, “Pipeline for Analysis and Visualization of Gene and Pathway Expression in Breast Cancer”, College of Charleston SURF Program, amount \$6,500, June 2016 – August 2016.
- Duchinski, K. and **Anderson P.**, “Bioinformatics Development and Digital Transcriptomics for FLI1 in Human Breast Cancer Cells”, College of Charleston MAYS Program, amount \$6,000, January 2016 – June 2016.
- Anderson, Paul E** (PI), "Electronic Health Medical Record Phenotyping using Machine Learning and NLP," Sponsored by MUSC, State governmental agency, awarded \$25,778.00. (July 2015 - June 2016).
- Anderson, Paul E** (PI), "Computational Genomics Support, 2015-2020," Federal governmental agency, awarded \$2,826.00. (June 2015 - June 2016).
- Anderson P.** (PI) and Shedlock A. (Co-PI), “REU Site: Next-Generation Bioinformatics for Genomics-enabled Research in the Life Sciences”, NSF Research Experience for Undergraduates, Division of Biological Infrastructure, amount \$334,662, Awarded.
- Yetsko K., Shedlock A., and **Anderson P.**, “Prediction and Annotation of Genomic Repeat Dynamics using Hidden Markov Models”, College of Charleston MAYS Program, amount \$2,000, October 2013 – June 2014.
- Anderson P.** (PI), Shedlock, A. (Co-PI), Watson, D., “Building Next-Generation Bioinformatics Cyberinfrastructure for Genomics-enabled Research and Education in the Charleston Scientific Community”, NSF Support SC EPSCoR GEAR: Cybinfrastructure Program, amount \$100,000, June 2013 – Jun 2014.
- Anderson P.** (PI), “Learn2Mine: An integrated learning environment to introduce undergraduate students to data mining, computer science, and data science”, SC EPSCoR GEAR: Research Experience, amount \$17,000, June 2013 – Jun 2014.
- Anderson, P.** and Turner C.¹, “Learn2Mine: An integrated learning and data mining system for scientists and students,” Summer Undergraduate Research with Faculty, College of Charleston. Funding: \$6,000.
- Anderson, P.** (PI), “Developing an Analysis Pipeline for Next Generation Sequencing,” South Carolina Sea Grant Consortium, Seed Grant. Funding: \$10,000.

* Indicates an undergraduate student

+ Indicates a graduate student

COLLEGE OF CHARLESTON SERVICE ACTIVITIES

Member, Graduate School Assessment Committee (2018)

Director of Master's in Data Science and Analytics Program (2018 – present)

Director of Discovery Informatics/Data Science Program (2012 – present)

Senator, College of Charleston, Computer Science, (2014 – 2016).

College of Charleston Research and Development Committee (2015)

Collaboration Agreement with University of San Francisco MS in Analytics Program

Directed the Department of Computer Science and Association for Computing Machinery Club High school Programming Competition (2013 and 2014)

Directed Data Science Program Annual Competition (2015 – present)

Association for Computing Machinery Student Club Advisor (2011 – 2013)

School of Sciences & Mathematics Scholarship Committee (2013 and 2014)

School of Science and Mathematics High Performance Computing Task Force (2012)

School of Sciences & Mathematics Scholarship Committee (2012)

Undergraduate Research and Creative Activities Review Committee (2011 – present)

PROFESSIONAL SERVICE ACTIVITIES

Reviewer, Comparative Biochemistry and Physiology - Part D: Genomics and Proteomic

Conference Chair, BIOINFORMATICS (BIOSTEC) 2018

Program Committee, CIBCB (2016 – Present)

Editorial Board, Advances in Bioinformatics Journal (2016 – present)

Reviewer 2015+ AMIA Translational Bioinformatics Conference

Reviewer 2016 IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2016)

Data Science Undergraduate Degree Panel Presentation at SIGCSE 2014.

Program Committee for International Conference on Bioinformatics Models, Methods, and Algorithms (2013 – present)

Chair of Session, 5th International Conference on Bioinformatics and Computational Biology (BiCoB)

Reviewer, IEEE Conference on Computational Intelligence and Computational Biology (2015 – present)

Reviewer, Chemometrics and Intelligent Laboratory Systems (2015 – present)

Reviewer, Bioinformatics (2009 – present)

Reviewer, Journal of Supercomputing (2013 – present)

Reviewer, Metabolomics (2008 – present)

Book Review, *Introduction to Technical Problem Solving with MATLAB* by J Sticklen and M. Eskil (2010)

* Indicates an undergraduate student

+ Indicates a graduate student

Reviewer, IEEE Transactions on Evolutionary Computation (2009 – 2011)

Reviewer, 6th IEEE Symposium on Bioinformatics and Bioengineering (2006)

PROFESSIONAL SOCIETIES

Association for Computing Machinery (ACM)

Institute of Electrical and Electronics Engineers (IEEE) Computer Society

SELECT PUBLICALLY AVAILABLE SOFTWARE

For a complete list see <http://anderson-lab.github.io> and <https://github.com/Anderson-Lab>.

Complimentary Domain Prioritization. <https://github.com/Anderson-Lab/ComplementaryDomainPrioritization>

Single Nucleotide Variant Discriminant Analysis (SNV-DA). <https://github.com/Anderson-Lab/SNV-DA>

Confounding Correcting Predictive Analytics (CCPredict). <https://github.com/Anderson-Lab/CCPredict>

Orthogonal Project onto Latent Structures (OPLS). <https://github.com/Anderson-Lab/OPLS>

C2G2: A Next Generation Sequence Processing and Analysis Platform with Integrated Cloud-storage and High Performance Computing Resources.

Learn2Mine: A cloud-based integrated machine learning, data mining, and learning experience.

<http://learn2mine.appspot.com/>

NMR-based Metabolomics Analysis Toolbox. <https://github.com/BirgPublicizer/Metabolomics-Analysis-Toolbox>

Forensic DNA analysis java-based application (GenoStat) that generates DNA profile match statistics (RMP, CPI and things like sibling match probabilities) as well performing mixture resolution (separating mixtures into their contributor components). <http://www.bioforensics.com/genostat/index.html>

Developed protonation and hydrogen bonding, structure alignment kernel, and structure datasets for the biomolecular structure software package called SPADETM. <http://www.spadeweb.org/>

THESIS AND POSTDOC ADVISORS

Professor Michael L. Raymer, Ph.D.

Department of Computer Science and Engineering

Wright State University, Dayton, Ohio 45435

(937) 775-5110

michael.raymer@wright.edu

Professor Travis E.W. Doom, Ph.D.

Department of Computer Science and Engineering

Wright State University, Dayton, Ohio 45435

(937) 775-5105

travis.doom@wright.edu

Nicholas J. DelRaso, Ph.D.

Air Force Research Laboratory

Wright-Patterson AFB, OH 45433

(937) 904-9572

Nicholas.DelRaso@wpafb.af.mil

* Indicates an undergraduate student

+ Indicates a graduate student