Application

Name: Kari Norman

Student Data

First name	Middle name	Last name
Kari		**Norman**

Previous names (ex: maiden name)

E-mail

kari@normanfamily.org

Citizenship

Country of Birth: **United States**

U.S. Citizen: **Yes**

If no,

- *Country of Citizenship:*
- *Permanent Resident Alien:* **----**

If yes,

- *PRA Number:*
- *Port of Entry:*

References

List at least three persons familiar with your academic preparation and your technical abilities.
Please have these individuals mail the reference forms directly to Krell Institute.

	Title	First name	Last name	Institution	E-mail	Status
1.	Dr.	Ethan	White	University of Florida	ethan@weecology.org	Submitted
2.	Dr.	Lise	Aubry	Utah State University	lise.aubry@usu.edu	Submitted
3.	Dr.	Carl	Boettiger	University of California Berkeley	cboettig@berkeley.edu	Submitted
4.						

Name: Kari Norman

Academic Status

Current Academic Status: **First Year Doctoral Student**

Have you completed any academic credit towards your computational science/engineering doctoral degree? **Yes**

If yes, how many terms have you completed? (exclude summer) **1 Semester**

Official transcripts from every listed institution are a required component of the application including your Fall 2016 transcript, if applicable. Please see the instructions for more information on where to send the transcripts.

Doctoral Institution (Institution where you plan on completing your computational science and engineering doctorate or first choice doctoral university):

Institution	*Start Date*	*Expected End Date*	*Department*	*Academic Discipline*	*GPA*	*Degree*
University of California, Berkeley	08/2016	05/2021	Environmental Science, Policy, and Management	Ecology	4.0	PhD

Department Chair at Doctoral Institution:

First Name	*Last Name*	*Email*
George	Roderick	roderick@berkeley.edu

Other Doctoral Institution Choices (Answer only if not currently at doctoral institution)

			Department Chair Information	
Institution	*Department*	*Academic Discipline*	*Name*	*Email*

Higher Educational History (All university/colleges attended and degrees obtained with the exception of the doctoral degree listed above):

Institution	***Start Date***	***End Date Expected or Actual***	***Department***	***Academic Discipline***	***Degree***	***GPA***
Utah State University	08/2012	05/2016	Wildland Resources	Ecology	Bachelors	3.87
Utah State University	08/2012	05/2016	Mathematics and Statistics	Statistics	Bachelors	3.87
					None	
					None	
					None	

Name: Kari Norman

Graduate Advisor

The graduate advisor is the person from the preferred institution **who views and approves the Program of Study.**

First Name	*Last Name*
Carl	**Boettiger**

Institution	*Title (Dr., Ms., Professor, ...)*
University of California, Berkeley	**Dr.**

E-mail

cboettig@berkeley.edu

Address 1

ESPM Department, University of California

Address 2

130 Mulford Hall #3114

City	*State*	*Zip Code*
Berkeley	**CA**	**94720-3114**

Telephone	*Fax*

Name: Kari Norman

Program of Study

Listed are the courses in science and engineering, applied mathematics, and computer science that you agreed to take on your proposed Program of Study.

University: University of California, Berkeley

Course number	Course Title	Credit hours	Term and Year	Grade	Academic Level
Science/Engineering					
ESPM 298	Community Ecology	3S	Fall 2017		G
ESPM277	Advanced Topics in Conservation Biology	3S	Fall 2016	A	G
Mathematics and Statistics					
MATH222A	Partial Differential Equations	4S	Spring 2018		G
STAT238	Bayesian Statistics	3S	Fall 2017		G
Computer Science					
COMPSCI267	Applications of Parallel Computers	3S	Spring 2018		G
CS286A	Introduction to Database Systems	4S	Spring 2017		G

I have read this program of study and affirm that, in my opinion, it satisfies the fellowship program requirements. This POS has been approved by my advisor, **Carl Boettiger**, and I understand that, if offered a fellowship, my advisor and I are required to sign this page and send it to the Krell Institute.

Student's signature ____________________________ Date _______________

Graduate Advisor: **Carl Boettiger**
Graduate Advisor's Institute: **University of California, Berkeley**
Graduate Advisor signature ____________________________ Date _______________

Krell Institute (Office use only) __________________________________

Krell Institute, Attn: DOE CSGF Coordinator
1609 Golden Aspen Drive, Suite 101, Ames, IA 50010
Phone: 515-956-3696, Fax: 515-956-3699, csgf@krellinst.org

Name: Kari Norman

Course Description

ESPM 298: Community Ecology

Review of major concepts of community ecology, including in depth discussion of the historical development of the discipline, current issues and methodologies, and practical applications in areas such as natural resource management and biological conservation. Competition models, niche theory, Neutral theory, and community assembly will all be addressed.

ESPM277: Advanced Topics in Conservation Biology

A graduate level seminar covering advanced topics in conservation of biodiversity, focused on designing protected area networks. We will first lay the groundwork for the course by exploring the fundamental papers in ecology and conservation biology that led to systematic conservation planning. Then, we will study various issues at the current frontiers of the discipline, such as incorporating threats, costs, evolutionary processes, and ecosystem services into reserve network design. The class will encourage student engagement through discussions, group projects, peer instruction and peer review of essays.

MATH222A: Partial Differential Equations

The theory of boundary value and initial value problems for partial differential equations, with emphasis on nonlinear equations. Laplace's equation, heat equation, wave equation, nonlinear first-order equations, conservation laws, Hamilton-Jacobi equations, Fourier transform, Sobolev spaces.

STAT238: Bayesian Statistics

Bayesian methods and concepts: conditional probability, one-parameter and multiparameter models, prior distributions, hierarchical and multi-level models, predictive checking and sensitivity analysis, model selection, linear and generalized linear models, multiple testing and high-dimensional data, mixtures, non-parametric methods. Case studies of applied modeling. In-depth computational implementation using Markov chain Monte Carlo and other techniques. Basic theory for Bayesian methods and decision theory.

COMPSCI267: Applications of Parallel Computers

Models for parallel programming. Fundamental algorithms for linear algebra, sorting, FFT, etc. Survey of parallel machines and machine structures. Exiting parallel programming languages, vectorizing compilers, environments, libraries and toolboxes. Data partitioning techniques. Techniques for synchronization and load balancing. Detailed study and algorithm/program development of medium sized applications.

CS286A: Introduction to Database Systems

Access methods and file systems to facilitate data access. Hierarchical, network, relational, and object-oriented data models. Query languages for models. Embedding query languages in programming languages. Database services including protection, integrity control, and alternative views of data. High-level interfaces including application generators, browsers, and report writers. Introduction to transaction processing.

Name: Kari Norman

Other Planned Courses

Listed are the other courses you plan to take that you believe are particularly pertinent to your proposed or current research in the areas of Mathematics, Science and Engineering, and Computer Science.

Course number	Course Title	Credit hours	Term and Year	Grade	Academic Level
Science/Engineering					
ESPM 201A	Research Approaches in Environmental Science, Policy, and Management	3S	Spring 2017		G
ESPM174	Design and Analysis of Ecological Research	4S	Spring 2017		B
Computer Science					
CS289A	Introduction to Machine Learning	4S	Fall 2019		G

Name: Kari Norman

Course Description

ESPM 201A: Research Approaches in Environmental Science, Policy, and Management

Research projects and approaches in environmental science, policy, and management. An introduction to the diverse ways environmental problems are researched, comparing the approaches and methods of various disciplines represented among faculty and students.

ESPM174: Design and Analysis of Ecological Research

Surveys major designs and analyses for biological field and laborabory studies. Topics include data distributions; regression; analysis of variance; fixed and random effects; blocking, split plots, and repeated measures; maximum likelihood; Generalized Linear Models; basic computer programming. Relies on math to interpret and manipulate equations supported by computer simulations. Examples include population, ecosystem, behavioral, and evolutionary ecology.

CS289A: Introduction to Machine Learning

This course provides an introduction to theoretical foundations, algorithms, and methodologies for machine learning, emphasizing the role of probability and optimization and exploring a variety of real-world applications.

Name: Kari Norman

Completed Courses

Please list up to six courses you have completed that are particularly pertinent to your proposed or current research in the areas of Mathematics, Science and Engineering, and Computer Science. Please do not list entry level science/engineering or mathematics courses like Calculus I.

Course number	Course Title	Credit hours	Term and Year	Grade	Academic Level
BIOL5322	Programming and Database Management for Biologists	3S	Fall 2014	A	B
MATH2280	Ordinary Differential Equations	3S	Spring 2015	A	U
MATH5720	Introduction to Mathematical Statistics	3S	Spring 2016	B+	B
STAT5600	Applied Multivariate Statistics	3S	Spring 2015	A	B
WILD4600	Methods and Software for Spatial Conservation Planning	3S	Spring 2014	A	B
WILD6401	Population Ecology: State Variables, Vital Rates, Dynamics of Structured Populations	5S	Fall 2014	A	G

Name: Kari Norman

Course Description

BIOL5322: Programming and Database Management for Biologists

Fundamentals of the computer programming and database management necessary for conducting biological research. Data structure, analysis, and graphics creation will be explored using the programming language Python. Database creation, queries, and automated report creation will also be addressed. Students will apply these tools to biological questions resulting in a final project using independently obtained data.

MATH2280: Ordinary Differential Equations

Analytic solution techniques for ordinary differential equations. Initial value and boundary value problems and applications. Higher-order scalar equations, first-order linear systems, and Laplace transforms.

MATH5720: Introduction to Mathematical Statistics

Basic theory of point and interval estimation and hypothesis testing. Topics include: sufficiency and completeness; method-of-moments, best unbiased, maximum likelihood, Bayes', and empirical Bayes' estimators; Neyman-Pearson lemma; and likelihood ratio tests.

STAT5600: Applied Multivariate Statistics

Introduction to multivariate statistical procedures for data analysis. Topics include MANOVA, principal component analysis, factor analysis, clustering, and classification.

WILD4600: Methods and Software for Spatial Conservation Planning

At the end of the course, students will better understand the basic rational and methodology behind spatial conservation planning methods. Having worked on numerical examples, students have an insight into how these methods really operate on data. Hands-on topics include scoring, minimum set and maximum utility planning, irreplaceability, threats and dynamic landscapes, optimization, multi-action planning, etc. Students will also be able to set up and run basic analyses with the Zonation spatial prioritization software. The latter half of the course will be about the Zonation framework and software. Lecture material includes how decision rules and conservation value are defined in the spatial conservation planning context and the basics of solving reserve selection problems.

WILD6401: Population Ecology: State Variables, Vital Rates, Dynamics of Structured Populations

Theory and application of population ecology. Topics include monitoring and estimation of population abundance using capture-mark-recapture and distance sampling methods and examination of metapopulation dynamics through occupancy models. The importance of variation in life history strategies and the impact of that variation on reproduction, survival, and dispersal are also discussed. Perturbation analyses and demographic models are used to study life history evolution.

Name: Kari Norman

Research Statements

This information is vital to the overall evaluation of your application.

Field of Interest and the Role of Computational Science

a. In terms a general audience would understand, describe an important, outstanding scientific or engineering challenge in your field of interest where computational science can play an important role. (1/3)
b. Describe the particular science or engineering problem that you would like to pursue in your research. What would be the impact on the field and/or on science, engineering and/or society in general if this challenge could be successfully addressed? (2/3)

Efficiently allocating limited conservation resources is a pressing environmental policy issue of our generation. Doing so requires identifying regions of high importance for biodiversity both now and in the future as species shift in response to global change. Current biodiversity-based efforts to prioritize regions for conservation focus almost exclusively on biodiversity maps derived from geographic range map data for one or more taxa. These data are available for many taxa at global scales, but are poorly suited to conservation work due to their static nature and poor resolution(1). It is therefore imperative to provide an alternative method for quantifying and predicting biodiversity patterns, by using dynamic data that are more closely linked to conditions in local communities.

Due to extensive government and citizen efforts, survey data are increasingly available to fill this need. In contrast to range maps, surveys provide direct observations of the number of species and the number of individuals in the field. Comparisons of range map and survey data show large disparities in biodiversity estimates(2), and my previous work shows that these disparities are magnified in large-scale conservation prioritizations. Specifically, locations of hotspots determined from both data types have as little as 15% overlap in location, emphasizing the need to leverage survey data to inform conservation analyses.

In order to use survey data for comprehensive conservation analyses, we need a means to fill gaps in data availability across spatial and taxonomic extent. This presents a multi-faceted computational challenge as advanced modeling methods are necessary to address gaps, and large volumes of data are required to make those methods tractable. While multiple methods exist for fill spatial gaps, none yet address taxonomic gaps. My PhD work examines this question by developing an approach to predict assemblages across taxa using Joint Species Distribution Models (JSDM). It will be the first model that predicts assemblages for multiple taxa, an essential methodological task for applications in conservation and management.

1) Hurlbert, A.H. & Jetz, W. PNAS (2007)
2) Hurlbert, A.H. & White, E.P. Ecol Lett (2005)

Use of Computational Science in your research

In the research problem described in question 1, part b:

a. How will you use high performance computing with modeling and simulation and/or large data analysis? (1/3)
b. How will you use mathematics? (1/3)
c. How will you demonstrate the success of your approach? (1/3)

I will develop a JSDM that fills taxonomic data gaps by using one taxon to inform the prediction of another through a latent variable modeling approach, specifically addressing communities of birds and butterflies. Though multiple statistical approaches to JSDMs exist(1), latent variable model's ability to predict multiple response variables (species) with a reduced number of parameters make it especially suited to the task of predicting multi-taxa communities (2). Latent variable models use unobserved ("latent") predictors to account for unmeasured environmental or biotic variables. Taxa inhabiting similar scales and ecological spaces, such as birds and butterflies, are likely to respond to similar latent variables. This approach therefore provides a framework to use data rich birds to inform the prediction of data poor butterflies.

The first computing challenge will be compiling data from multiple large-scale survey efforts at different temporal and spatial scales. Temporal and spatial scale differences will be addressed by collapsing within time steps and grid cells across the landscape. Each cell will have data on species occurrence for birds and butterflies and corresponding climate data for that location. Cells will be split into testing and training data sets.

A suite of latent variable models will be fit for birds, with the best model selected by cross-validation. The latent variables from the chosen model will then be used in a model for butterfly prediction, built based on butterfly training cells. The butterfly model's predictive ability will then be assessed by cross-validation.

One of the limitations of latent variable models is that potential poor prediction of butterflies by the bird latent variables could be attributed to a number of ecological scenarios: birds and butterflies could be 1) responding to different environmental variables, 2) responding to the same variables but in different ways, 3) not responding strongly to environmental variables. The model will therefore also be validated against simulated bird and butterfly assemblages for each of those three scenarios to understand prediction shortfalls.

1. Harris, D.J. Methods Ecol (2015)
2. Warton, D.I. et al Trends Ecol Evol (2015)

Program of Study

Describe how the courses listed in your planned program of study would help prepare you to address the challenges you have described in questions 1 and 2. Discuss your rationale for choosing these courses.

My program of study builds upon my strong ecological foundation and extensive quantitative background with expansion into theory and tools. Both of these aspects of my education are essential as I seek to address questions at the confluence of cutting-edge computational approaches, ecological understanding, and societal need. The selected science courses represent a comprehensive review of the context for my project, Conservation Biology, paired with an in-depth study of the theory of community assembly. These courses work in tandem to ensure the models I build, and the ecological conclusions drawn from them, will be grounded in sound ecological theory, with application of those models to conservation questions as a top priority.

The computer science courses address topics fundamental to both my project and working in the field of computational ecology broadly. Introduction to Database Systems is essential for compiling and relating large-scale data sources in an efficient manner. Well-constructed databases greatly reduce the time sink of data management, and facilitate sharing of data sources between investigators in the academic and conservation communities. Similarly, Applications of Parallel Computers provides the skills necessary for reducing computational time through parallelization. Ecological simulations, like those in my project, are especially suited to parallelization due to inherently large parameter spaces.

In Mathematics, the selected courses ensure I have the theoretical underpinning necessary to address computational ecological questions effectively. Partial Differential Equations provide the basis of thought for spatial problems in ecology, and the development of models to improve the tools of conservation is an inherently spatial tasks. This course will allow me to go beyond conceptual understanding of past approaches to spatial questions, and on to development of novel methodology. Bayesian statistics also fills a vital gap in my knowledge by providing statistical tools for estimation of parameters with limited data. Its framework allows for prior ecological knowledge to inform model development, further bridging the divide between computational approaches and ecological theory.

Name: Kari Norman

Programming Languages and programming models

List (four at most) the programming languages and programming models with which you have experience. Provide a sentence that describes how you use them.

Python
Used for data formatting and mapping of biodiversity of different taxa across North America.

R
Statistical analysis including population demography assessments, machine learning techniques, mixed effects and non-parametric regression models, and data visualization.

SQL
Creation and use of a Postgresql spatial database for assessing biodiversity metrics based on spatial range maps.

Name: Kari Norman

List of Publications

Papers

Talks

K. Norman and L. Aubry, "Demographic Consequences of Climate Change in a Hibernator, the Uinta Ground Squirrel", Utah State Student Research Symposium, Utah State University, April 2015.

K. Norman and S. Null, "Modeling Streamflow in the Wasatch Mountain Region with Climate Change", iUtah Cohort Session, Salt Lake City, August 2013.

Posters

K. Norman and E. White, "Biodiversity Prioritization: A comparison of data types", Gordon Research Conference: Unifying Ecology Across Scales, University of New England, July 2016.

K.Norman and L. Aubry, "Demographic Consequences of Climate Change in the Uinta Ground Squirrel", Research on Capitol Hill, Salt Lake City, January 2016.

K. Norman and E. White, "Biodiversity Prioritization: A comparison of data types", Ecological Society of America, Baltimore, August 2015.

K. Norman and E. White, "Biodiversity Prioritization: A comparison of data types", National Conference of Undergraduate Research, Spokane, April 2015.

K. Norman and E. White, "Biodiversity Prioritization: A comparison of data types", Utah Conference of Undergraduate Research, Southern Utah University, February 2015.

K. Norman and S. Null, "Predicting Streamflow for Utah's Wasatch Mountains with Climate Change", iUtah Symposium, Salt Lake City, July 2013.

Name: Kari Norman

Laboratory and Research Experience/Other Employment

Begin with current or most recent employment. Please include employer, dates employment started and ended, position, and nature of work.

US Forest Service, Summer 2016, Botany Field Technician
Utah State University - Dr. Ethan White, 8/2014-5/2016, Undergraduate Research Fellow, Analyzed biodiversity patterns based on different data types.
Utah State University - Dr. Lise Aubry, 8/2014-5/2016, Undergraduate Research Fellow, Field surveys of Uinta ground squirrel populations and assessment of demographic structure using R.
University of Goettingen, Summer 2014, DAAD Rise Research Intern, Developed model to predict forest biomass from LiDAR data using nonparametric models.
iUtah EPSCoR Project, 5/2013-12/2013, Research Fellow, Developed model of streamflow for major watersheds in the Wasatch region.
Utah State University - Dr. Dan MacNulty, 8/2012-5/2013, Undergraduate Research Fellow, Analysis of wolf-elk interactions in predatory situations.

Academic Awards and Honors - Include undergraduate and graduate honors (if applicable).

University of California, Berkeley:
NSF NRT Fellowship, Data Science for the 21st Century 2016

Utah State University:
University and Departmental Honors, 2016
Dean's List, 2012-2106
Outstanding Statistics Undergraduate 2016
Undergraduate Research and Creative Opportunities Grant awardee 2015
Jeb Stuart Scholarship 2014-2015
iUtah iFellowship 2013
Presidential Scholar 2012-2015
Quinney Scholar 2012-2015
Undergraduate Research Fellow 2012-2015
Regents Scholar 2012-2014

Extracurricular Activities - Include technical societies and service organizations.

Organizational Committee, Expanding Your Horizons Network, 2016 - present
Vice President, USU Chapter, Society for Range Management, 2015-2016
Wildland Dept. Rep., Natural Resources Student Council, 2013-2016
Founding Member, Ecology Club, 2014

Name: Kari Norman

Additional Comments

Name: Kari Norman

DOE CSGF and Other Fellowships

The information that you provide will allow us to target our advertising more effectively. This information is confidential and is not used in review of the fellowship application.

1. Have you applied to other fellowship programs?

- ☐ DOE NNSA SSGF
- ☑ NSF
- ☐ DOD
- ☐ University-sponsored   Names of fellowships:
- ☐ Other   Names of fellowships:

2. How did you find out about the program?

- ☐ DOE CSGF poster
- ☐ DEIXIS, DOE CSGF annual publication
- ☐ Attended DOE CSGF talk
- ☐ Advertisement   Name the source:
- Word of mouth from
  - ☑ faculty
  - ☐ student
  - ☐ administrator
- ☐ Laboratory staff
- ☐ Institutional announcement
- ☐ Conference or meeting   Name:
- ☐ World Wide Web   List URL:
- ☐ Other   Explain:

Applicant Demographics

Applicant data is important in assessing the effectiveness of our efforts to solicit applications from a diverse population. Providing the information on this form is voluntary; omission of information will not affect any decision about your application. We appreciate your cooperation.

Race:	**Caucasian**
Gender:	**Female**
Physical/mental disability:	**No**