**Interview guide**

1. Who are you?
2. What is your current role, and how long have you worked there?
3. What are your major job responsibilities right now?
4. Tell me a bit about your training. What was your path to this position?
5. Walk me through a typical day in your work life.

II. How did you get interested in machine learning?

1. What kind of machine learning or data science applications are you interested in using?
2. What (or who) got you interested in this technique?
3. If the project goes well, what would be the ideal outcome?
4. What do you think is the advantage of using machine learning in this way?
5. Where are you at in the project right now?

III. What kind of tools are you using?

1. What kind of computer are you using?
   1. Personal? Lab machine?
   2. Cluster maintained by University? Cloud?
2. What kind of software are you using/going to use?
3. Did you consider any other tools? How did you pick the one you chose?
4. Tell me what it was like learning to use the software.

IV. Data management

1. Do you have an estimate for how big your data is?
2. What kind of data is it? Is it a table? Something else?
3. Where does the data “live”?
4. Where does your data come from?
5. Is the data in a form that you could analyze immediately? Or will it need to be transformed or cleaned?
6. Does your data have any special considerations for management, like privacy?
   1. If so, has this created any challenges for you?

V. Training

1. Where did you learn how to code and/or manage data?
2. Have you ever taken a class? What was it like?
   1. What did it cover?
   2. What do you wish it had covered?
3. Have you ever participated in a workshop, like Software Carpentry? What was it like?
4. Have you ever attended any MIDAS events? Does your lab have an affiliation?
5. Do you use any online resources or communities? Tell me about it.
   1. Do you feel welcome asking questions? Sharing your ideas?
6. What do you wish you could find training for?
7. What has been the hardest part of learning these skills?

VI. Team collaboration

1. Who else is part of the project?
2. What are the roles of everyone in the project?
3. How do you share code?
4. How do you share data?
5. How do you share results of analyses?
6. Walk me through the last time you needed to share some code, data, or a result with someone else in the lab.
7. What are the hardest parts of working with a team on this project?

VII. Publication beyond the lab

1. When your work is complete, how will you share your findings with the broader community?
2. Who is your audience? Do you have a professional society or journal in mind?
3. How common are machine learning methods in this area?
4. How do you expect your community/reviewers to respond?
5. Do you plan to make any of your code, data, or models available to the community?
   1. How?
   2. Have you ever tried this before?
   3. Do you anticipate any challenges?
6. Are there any benefits you anticipate for sharing (or not) the artifacts of this project?

VIII. Survey of tools used

* Version control software (Git, Mercurial, etc.)
* GitHub or another Git repository hosting service
* Cloud computing
* Computing cluster
* Virtual environment
* Virtual machine
* Command line