The important role of versioning your code

Versioning as an essential part of a reproducible workflow

Stephan Siemen
Development Section, Forecast Department
Why versioning code is so essential for reproducible workflows

• Track changes
  – Makes them visible for colleagues
  – Lets the team see what has changed to understand their impact

• Revert if necessary, e.g. in case of regressions
  – Records also what has not worked → don’t only commit what worked

• Makes it easy to share code and keep up-to-date with developments
  – Powerful when linked with Continuous Integration/Deployment (CI/CD)

• But can you version more than just your code?
  – Yes, and you should → data & configurations
Git – the right choice for versioning

• Git is well designed and very powerful
  – Lots to learn, but in the end it you survive with a small set of commands
  – Big community and ecosystem of tools
  – Very good documentation

• Not always easy
  – Have to be careful to not leave branches open for too long
  – Easy to start again from a fresh clone
  – Initial struggle to version larger binary files
  – Helps to have a GUI to keep track of history and branches
  – And there is always Google … ;(-)
At ECMWF we use internally Bitbucket
Gitflow workflow
GitHub – bringing versioning to the next level

• Version control → social media
  – Collaboration platform
  – Linked into many other systems/services

• Introducing pull requests
  – Extra layer to encourage code review and testing

• Trigger actions & services on commits & PRs
The GitHub workflow

GitHub

Organization Repository

Fork

Pull Request

Personal GitHub Repository

Clone

Local Machine

Upstream

Local Files

Master Branch

Merge

Feature Branch
Supercharging your workflows with GitHub

• CI/CD
  – connecting services like TravisCI, CircleCI, ...

• Enforce code reviews in pull requests

• Code contribution agreements
Embracing Open Development for ECMWF software

- ECMWF now available on GitHub
  - Allows for easier code contributions
First success – Windows port of software

• Ported key libraries to Windows
  – Need help from community
  – Test suites are running from GitHub
  – Work was done by company Old Reliable Tech

• Package software on conda & pip
  – Easy reach for Python community
Takeaway messages …

• Versioning is essential for reproducible workflows
  – Not just to do reverts ;-)  

• Versioning can be the start of robust development workflows
  – CI/CD, pull requests, code reviews, …  

• Community can now contribute to ECMWF software through a common workflow