REPRODUCIBLE PKG ENVS FOR MODELING WORKFLOWS

Pascal Sauer

Potsdam Institute for Climate Impact Research

STORYTIME

- HPC, ~50 users using same pkg library
- ~30 pkgs developed in-house, ~10 updates per day, most people want all updates

STORYTIME

- HPC, ~50 users using same pkg library
- ~30 pkgs developed in-house, ~10 updates per day, most people want all updates

1. 👨‍🎓 finishing up paper, run results today differ from yesterday

STORYTIME

- HPC, ~50 users using same pkg library
- ~30 pkgs developed in-house, ~10 updates per day, most people want all updates

1. 👩‍🎓 finishing up paper, run results today differ from yesterday
2. 👩‍💻 testing new feature in a central tool

STORYTIME

- HPC, ~50 users using same pkg library
- ~30 pkgs developed in-house, ~10 updates per day, most people want all updates

1. 👩🏻‍🎓 finishing up paper, run results today differ from yesterday
2. 👨🏼‍💻 testing new feature in a central tool
3. 👷🏾‍♂️ reproduce run from a month ago

WHAT IS A PKG ENV?

- installed pkgs
- specific to a project (folder)

WHAT IS A PKG ENV?

- installed pkgs
- specific to a project (folder)
- isolated from system-wide installed pkgs

WHAT IS A PKG ENV?

- installed pkgs
- specific to a project (folder)
- isolated from system-wide installed pkgs
- tools to
  - install/update/remove pkgs
  - document pkg env
  - restore pkg env

renv: Project Environments

A dependency management toolkit for R. Using 'renv', you can create and manage project-local R libraries, save the state of these libraries to a 'lockfile', and later restore your library as required. Together, these tools can help make your projects more isolated, portable, and reproducible.

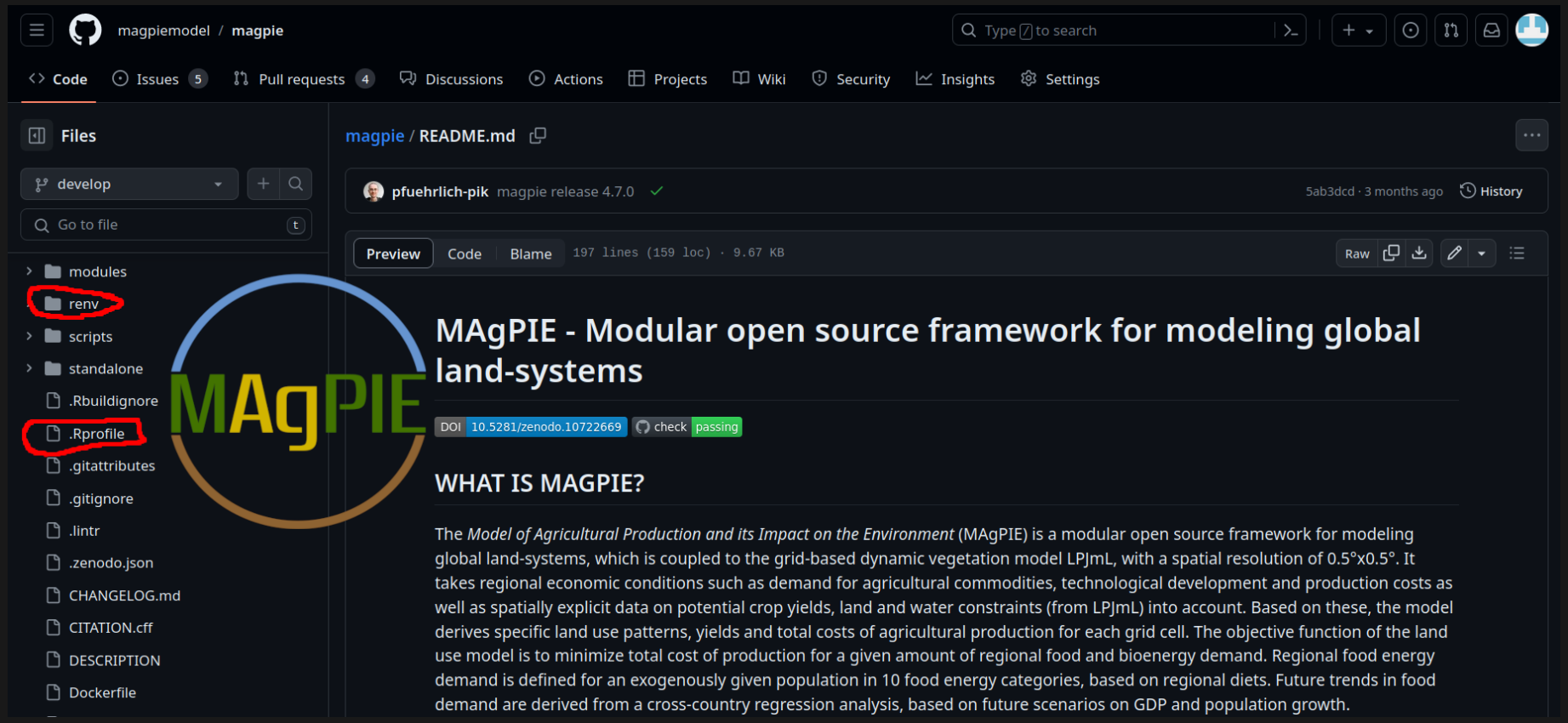
Version:	1.0.5
Imports:	utils
Suggests:	BiocManager, cli, covr, cpp11, devtools, gitcreds, jsonlite, knitr, miniUI, packrat, pak, R6, remotes, reticulate, rmarkdown, rstudioapi, shiny, testthat, uuid, waldo, yaml, webfakes
Published:	2024-02-29
Author:	Kevin Ushey [aut, cre], Hadley Wickham [aut], Posit Software, PBC [cph, fnd]
Maintainer:	Kevin Ushey <kevin at rstudio.com>
BugReports:	https://github.com/rstudio/renv/issues
License:	MIT + file LICENSE
URL:	https://rstudio.github.io/renv/, https://github.com/rstudio/renv
NeedsCompilation:	no
Materials:	README NEWS
In views:	ReproducibleResearch
CRAN checks:	renv results

RENV FILES & FOLDERS

```
renv/library/
.Rprofile
renv.lock
```

RENV.LOCK


```
    "magclass": {
      "Package": "magclass",
      "Version": "6.13.2",
      "Source": "Repository",
      "Repository": "https://rse.pik-potsdam.de/r/packages",
      "RemoteUrl": "https://github.com/pik-piam/magclass",
      "RemoteRef": "HEAD",
      "RemoteSha": "3c385d684d5cc3e7b90b8",
      "Hash": "dbc9e33e04a823f7219",
      "Requirements": [
        "abind",
        "data.table"
      ]
    },
```


magpiemodel / magpie
Type / to search
Code
Issues 5
Pull requests 4
Discussions
Actions
Projects
Wiki
Security
Insights
Settings
Files
develop
Go to file
modules
renv
scripts
standalone
.Rbuildignore
.Rprofile
.gitattributes
.gitignore
.lintr
.zenodo.json
CHANGELOG.md
CITATION.cff
DESCRIPTION
Dockerfile
magpie / README.md
pfuehrlich-pik magpie release 4.7.0
5ab3dcd · 3 months ago
History
Preview
Code
Blame
197 lines (159 loc) · 9.67 KB
Raw
MAgPIE - Modular open source framework for modeling global land-systems
DOI 10.5281/zenodo.10722669
check passing
WHAT IS MAGPIE?
The Model of Agricultural Production and its Impact on the Environment (MAgPIE) is a modular open source framework for modeling global land-systems, which is coupled to the grid-based dynamic vegetation model LPJmL, with a spatial resolution of 0.5°x0.5°. It takes regional economic conditions such as demand for agricultural commodities, technological development and production costs as well as spatially explicit data on potential crop yields, land and water constraints (from LPJmL) into account. Based on these, the model derives specific land use patterns, yields and total costs of agricultural production for each grid cell. The objective function of the land use model is to minimize total cost of production for a given amount of regional food and bioenergy demand. Regional food energy demand is defined for an exogenously given population in 10 food energy categories, based on regional diets. Future trends in food demand are derived from a cross-country regression analysis, based on future scenarios on GDP and population growth.

MAGPIE PKG ENVS

- clone MAgPIE from GitHub
- one time only: fill empty pkg env

MAGPIE PKG ENVS

- clone MAgPIE from GitHub
- one time only: fill empty pkg env
- on each start:
  - check for updates
  - create new pkg env specific to that run

DIFFERENCES TO REGULAR SETUP

- daily updates vs. stable pkg env
- for each run one stable pkg env
- quick pkg env creation needed

TECHNICAL DIFFERENCES

- lockfile not commited
- lockfile archive

TECHNICAL DIFFERENCES

- lockfile not commited
- lockfile archive
- main pkg env + one stable pkg env per run

TECHNICAL DIFFERENCES

- lockfile not commited
- lockfile archive
- main pkg env + one stable pkg env per run
- global symlink-based pkg cache
  - ✅ little disk space required (8.3GB for 3400 pkgs)
  - ✅ instant installation of cached packages

TECHNICAL DIFFERENCES

- lockfile not commited
- lockfile archive
- main pkg env + one stable pkg env per run
- global symlink-based pkg cache
  - ✅ little disk space required (8.3GB for 3400 pkgs)
  - ✅ instant installation of cached packages
  - ⛔ cache corrupt = all pkg envs break
  - ⛔ dev version with same version number in cache
  - ⛔ file permission issues

CONCLUSION

- ⛔ need to learn pkg env workflows
- ⛔ many different pkg envs

CONCLUSION

- ⛔ need to learn pkg env workflows
- ⛔ many different pkg envs

1. ✅ stable pkg env 👩‍🎓

CONCLUSION

- ⛔ need to learn pkg env workflows
- ⛔ many different pkg envs

1. ✅ stable pkg env 
2. ✅ isolated testing pkg env 

CONCLUSION

- ⛔ need to learn pkg env workflows
- ⛔ many different pkg envs

1. ✅ stable pkg env 👩‍🎓
2. ✅ isolated testing pkg env 👨‍💻
3. ✅ greatly increased reproducibility 👷🏾

CONCLUSION

- ⛔ need to learn pkg env workflows
- ⛔ many different pkg envs

1. ✅ stable pkg env 👩‍🎓
2. ✅ isolated testing pkg env 👨‍💻
3. ✅ greatly increased reproducibility 👷🏾
    - ⚠️ e.g. system libs not covered

FINAL SLIDE

pascal.sauer@pik-potsdam.de

github.com/pascal-sauer

slides on GitHub

renv homepage

slides made with reveal.js