American Meteorological Society 104th Annual Meeting – January 30, 2024

U.S. NAVAL RESEARCH LABORATORY

Herbie

A python package to retrieve numerical weather prediction data

Brian K. Blaylock
Naval Research Laboratory, Monterey, California, USA

Data Sources

NOMADS

ECMWF Open-data

Azure

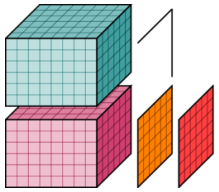
aws

Google Cloud Platform

Insert long list of NWP models here

Thank you NODD program!

Data Tools



xarray

cfgrib

pandas

matplotlib

MetPy

polars

Thank you developers!

Data Sources

NOMADS

ECMWF Open-data

Azure

aws

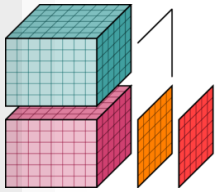
Google Cloud Platform

Insert long list of NWP models here

Access Gap

*How do I **get** the data?*

Data Tools



xarray

cfgrib

pandas

matplotlib

MetPy

polars

*Just because something is **public** doesn't mean it's easily **accessible***

Data Sources

Data Tools

Access Gap

*How do I **get** the data?*

URLs are not intended for human consumption

```
FILE = "https://blah.blah.s3.blah.com/blah/blah.grib2"
tool.open_dataset(FILE)
```

*Just because something is **public** doesn't mean it's easily **accessible***

Data Sources

Data Tools


```python
from herbie import Herbie
Herbie("2024-01-01", model='hrrr')
```

Goal: Help scientist spend less time discovering/wrangling data.

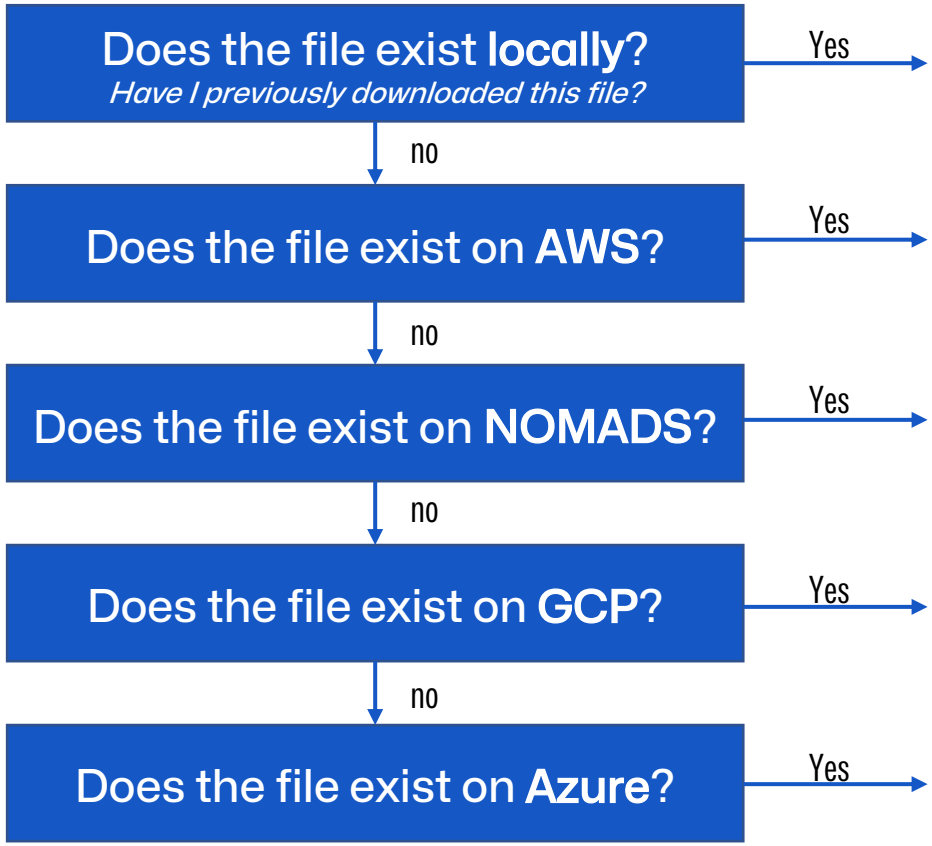
Data Sources

Data Tools

Goal: Help scientist spend less time discovering/wrangling data.

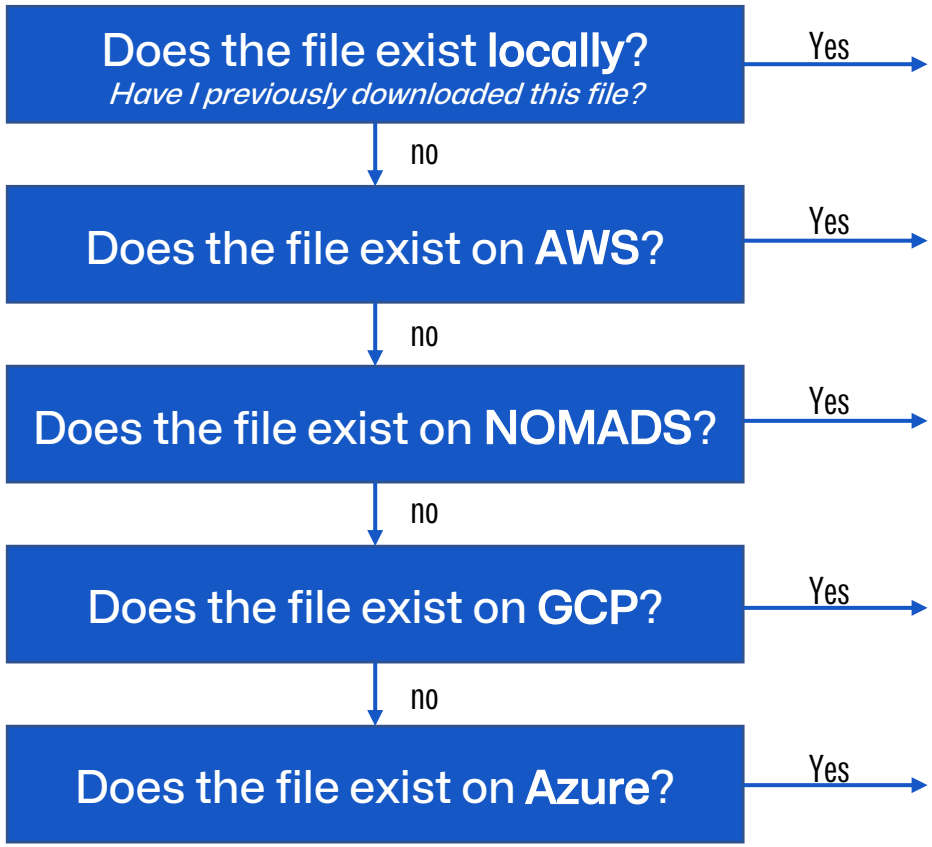
How Herbie Works

```
from herbie import Herbie
H = Herbie("2024-01-28 12:00", model='hrrr', fxx=5)
```



How Herbie Works

```python
from herbie import Herbie
H = Herbie("2024-01-28 12:00", model='hrrr', fxx=5)
```



Herbie Object

- **.grib** — URL or path to GRIB2 file
- **.idx** — URL to GRIB2 index file
- **...** — *(other attributes)*
- **.inventory()** — DataFrame of file contents
- **.download()** — Download the file to local disk
- **.xarray()** — Open the GRIB2 file with **xarray+cfgrib**

How Herbie Works

Herbie can discover these models...

- HRRR
- RAP
- ECMWF
- GFS
- GEFS
- HAFS
- HRDPS
- NAM
- NAVGEM
- NBM
- RTMA
- URMA
- RRFS*
- (potentially more)

and looks for data in these places...

- NOMADS
- Amazon Web Services
- Google Cloud Platform
- Microsoft Azure
- University of Utah Pando
- ECMWF Open data
- Meteorological Service of Canada
- (potentially more)

Herbie can be extended to discover other NWP datasets if

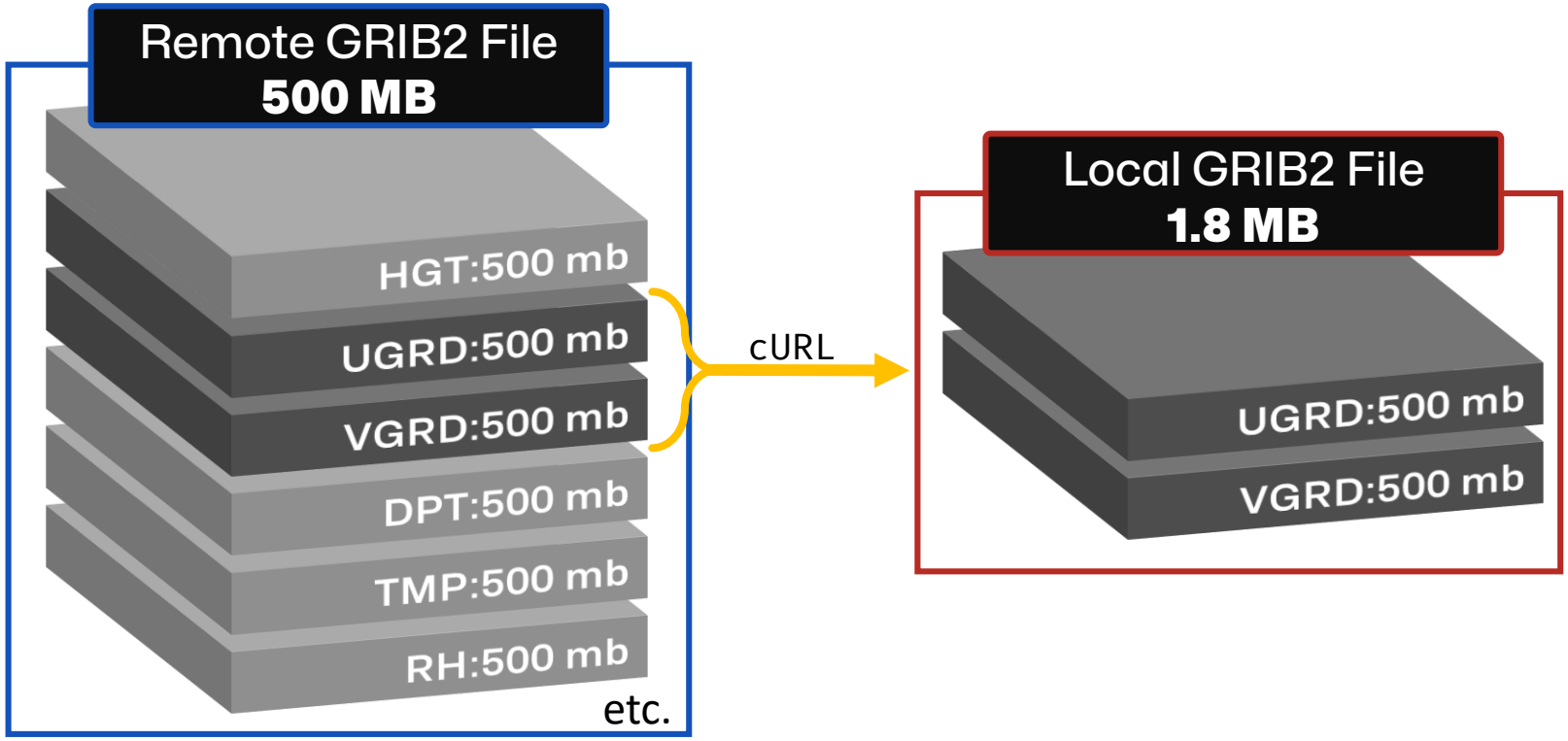
1. Data is in GRIB2 format
2. Can be accessed via https
3. (ideally) has a wgrib2-style index file.

How Herbie Works

```
H = Herbie("2024-01-01 12:00", model="gfs")
H.download("[U|V]GRD:500 mb")
```

https://noaa-gfs-bdp-pds.s3.amazonaws.com/gfs.20240101/12/atmos/gfs.t12z.pgrb2.0p25.f000

Herbie found this



Only download the data you need; Subset file by individual GRIB message

How Herbie Works

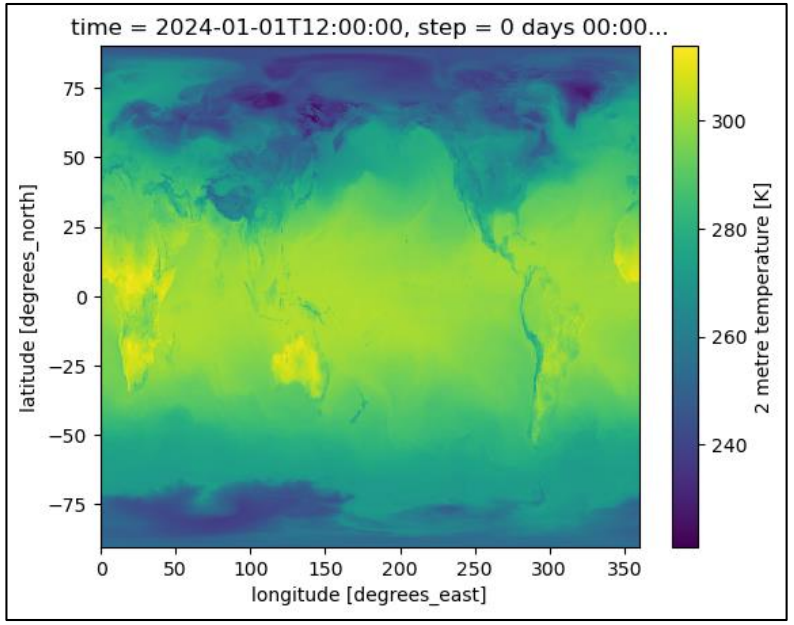
```
H = Herbie("2024-01-01 12:00", model="gefs", member="mean")
ds = H.xarray("TMP:2 m")
```

https://noaa-gefs-pds.s3.amazonaws.com/gefs.20240101/12/atmos/pgrb2ap5/geavg.t12z.pgrb2a.0p50.f000

Herbie found this

xarray.Dataset

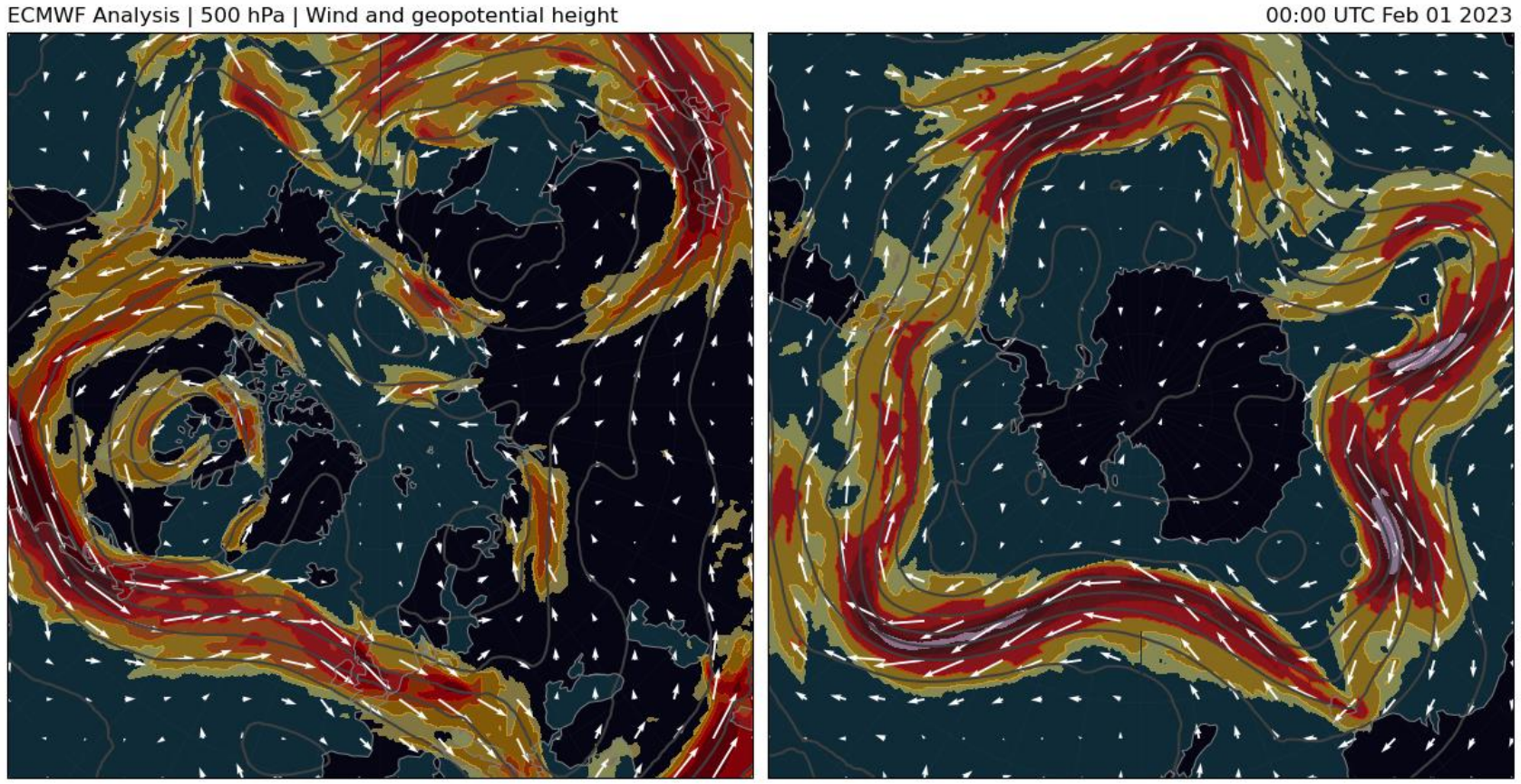
► Dimensions:	(**latitude**: 361, **longitude**: 720)		
▼ Coordinates:			
time	()	datetime64[ns]	2024-01-01T12:00:00
step	()	timedelta64[ns]	00:00:00
heightAboveGro...	()	float64	2.0
latitude	(latitude)	float64	90.0 89.5 89.0 ... -89.5 -90.0
longitude	(longitude)	float64	0.0 0.5 1.0 ... 358.5 359.0 359.5
valid_time	()	datetime64[ns]	2024-01-01T12:00:00
▼ Data variables:			
t2m	(latitude, longitude)	float32	245.1 245.1 245.1 ... 254.2 254.2
gribfile_projection	()	object	None
► Indexes: (2)			
► Attributes: (12)			



Only download the data you need; Subset file by individual GRIB message

Example: ECMWF 500 hPa wind analysis

```
for date in DATES:
    H = Herbie(date, model="ecmwf", fxx=0)
    ds = H.xarray("[gh|u|v]:500 mb")
    make_wind_map(ds)
```



No publically-accessible HRRR data after 48 hours.

HRRR unofficially archived at the University of Utah for field programs.

University of Utah HRRR archive made public

Official HRRR archive (and many NWP datasets) available through ***NOAA Open Data Dissemination Program*** (NODD)

HRRR available in Zarr format on AWS

2016 **2018** **2020** **2022** **2024**

Simple local-server data access scripts

HRRR

HRRR_archive_download

HRRR-B

RAP GFS NBM

Herbie

ECMWF GEFS RTMA URMA GDAS HRDPS HAFS

I should write some scripts so I don't have to repeat myself.

I'll put my code on GitHub. **I have no idea what I'm doing**, but it seems like a good idea.

Whoa, people found my code. I should clean it and make it user friendly.

There, I packaged my code and made it easier to use.

This code can be extended to access data from other sources and models, not just HRRR.

I'm getting a lot of emails about Herbie. ***What have I done!***

I feel like I'm starting to understand GitHub.

GitHub is an awesome tool.

Time to clean up Herbie's code again

Summary

is a Python package for accessing NWP data.

```
conda install -c conda-forge herbie-data
```

Herbie's popularity demonstrates

The need for good data discovery and access tools

Data providers: consider how you can improve your tools/docs to help users understand, get, and use your data.

If you still need an invitation, here is it...

Participate in the open source community

Your participation in open source makes it even better.

- Ask questions on Stack Overflow, GitHub Issues/Discussions, Discord, etc.
- Practice using GitHub with a "hobby project" (your grad school code)

What you will gain...

1. **Better software**
2. **Collaborative code development is a *highly desirable job skill*.**
3. **Friends**

*Even if you can't share code publicly (e.g. DoD, private company), you can still foster an "open source" culture with your colleagues.

Future of Herbie

Herbie continues to be a hobby project.
Its code is far from perfect, and that's OK. I'm learning a lot.

I'll keep "driving" Herbie until it becomes obsolete, or maybe Herbie will become the best NWP data access tool with your help.

Give Herbie a Star

https://github.com/blaylockbk/Herbie

Documentation

https://herbie.readthedocs.io

Install

```
conda install -c conda-forge herbie-data
```