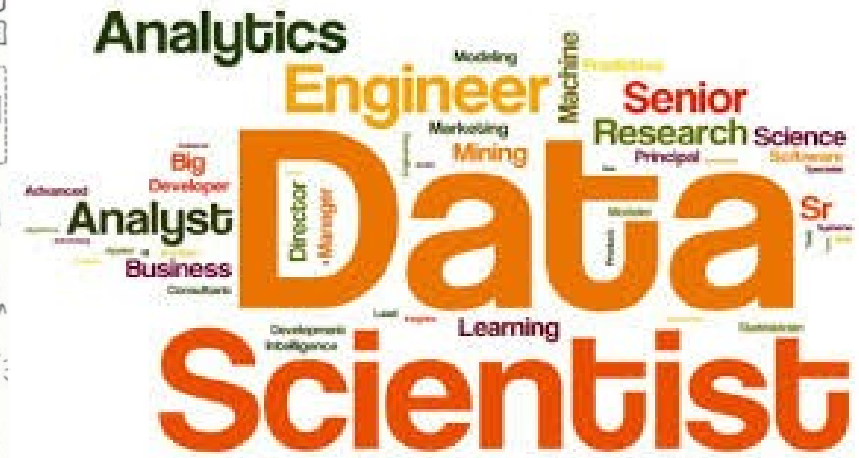
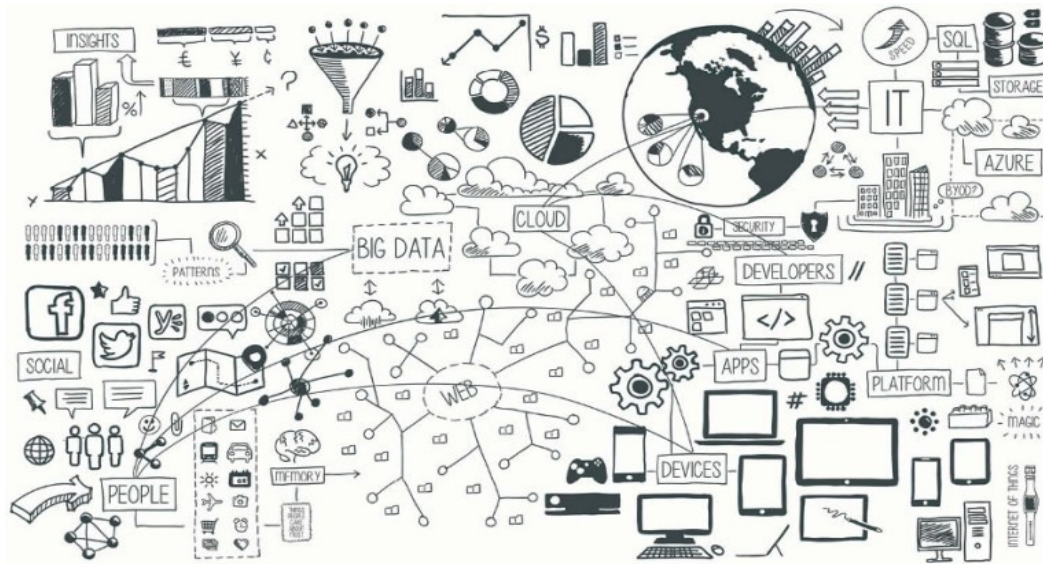


# Course Introduction



# Joaquín Bedia

**Grupo de Meteorología**  
**Univ. de Cantabria – CSIC**  
**MACC / IFCA**



**M1980 – Data Laboratory: Environment & Meteorology (16:00-18:00)**  
**Virtual classroom: <https://meet.jit.si/M1980>**

<b>03/25</b>	<b>X</b>	<b>Course introduction. Overview of climate4R</b>	<b>TL</b>	<b>JB</b>
03/26	J	Climatic System & Models (DM & ML in Climate Science)	T	SH
03/27	V	Data Repositories: ESGF & MARS	TL	SH
03/30	L	Data Repositories: ESGF & MARS	TL	SH
03/31	M	Lab: Climate4R – Example 1	L	JB
04/01	X	Lab: Climate4R – Example 2	L	JB
04/02	J	Downscaling: Data Mining in Climate	T	SH
04/03	V	Lab: downscaleR	L	JB
04/06	L	Evaluation and Validation	T	SH
04/07	M	Lab: Evaluation and Validation	L	JB
04/08	X	Impacts	L	JB & SH
04/13	L	Impacts	L	JB & SH

**SH** - Sixto Herrera | **JB** - Joaquín Bedia



A joint group of Univ. of Cantabria (**UC**) and Spanish National Research Council (**CSIC**) formed by 17 (6 staff) researchers based on Santander (Spain).



**Joaquín Bedia**



**Daniel San-Martín**



Our main research activities are:

**José M. Gutiérrez**  
Statistical  
downscaling



**Head**

**Miguel A. Rodríguez**  
Nonlinear  
physics



**Staff – Senior members**

**Antonio Cofiño**  
Computing  
data management



**Jesús Fernández**  
Numerical  
modeling



**Sixto Herrera**  
Data  
Mining



**M. Dolores Frías**  
Climate  
Services



### 3.1 RESULTADOS DE APRENDIZAJE

- Conocer los portales, bases de datos, repositorios, y el software y herramientas más relevantes para abordar un caso de uso en un área de conocimiento.
- Saber modelar problemas en cada área de conocimiento a un marco abstracto de Data Science e identificar qué puntos críticos pueden impactar el lograr los objetivos.

### 4. OBJETIVOS

Los datalab propuestos en el Máster de Data Science tienen como objetivo que el estudiante pueda conocer de la mano de expertos en las distintas áreas de conocimiento (física, medicina, genética, medioambiente, biodiversidad, economía, redes sociales, etc.) las técnicas y conjuntos de datos más relevantes en el entorno Open Science. En particular, el presente datalab se centrará en el problema de la regionalización estadística en predicción/proyección climática con técnicas de minería de datos.

### 3.1 RESULTADOS DE APRENDIZAJE

- Conocer los portales, bases de datos, repositorios, y el software y herramientas más relevantes para abordar un caso de uso en un área crítica
- Saber modelar casos de uso críticos pueden identificarlos

### 4. OBJETIVOS

Los data lab propiamente dichos (es decir, expertos en las disciplinas de ciencias sociales, etc.) las data lab se centra en la minería de datos.

### 5. MODALIDADES ORGANIZATIVAS Y MÉTODOS DOCENTES

ACTIVIDADES	HORAS DE LA ASIGNATURA
ACTIVIDADES PRESENCIALES	
HORAS DE CLASE (A)	
- Teoría (TE)	8
- Prácticas en Aula (PA)	15
- Prácticas de Laboratorio (PL)	
- Horas Clínicas (CL)	
Subtotal horas de clase	23
ACTIVIDADES DE SEGUIMIENTO (B)	
- Tutorías (TU)	7,5
- Evaluación (EV)	2
Subtotal actividades de seguimiento	9,5
<b>Total actividades presenciales (A+B)</b>	<b>32,5</b>
ACTIVIDADES NO PRESENCIALES	
Trabajo en grupo (TG)	7,5
Trabajo autónomo (TA)	35
Tutorías No Presenciales (TU-NP)	
Evaluación No Presencial (EV-NP)	
<b>Total actividades no presenciales</b>	<b>42,5</b>
<b>HORAS TOTALES</b>	<b>75</b>



### 3.1 RESULTADOS

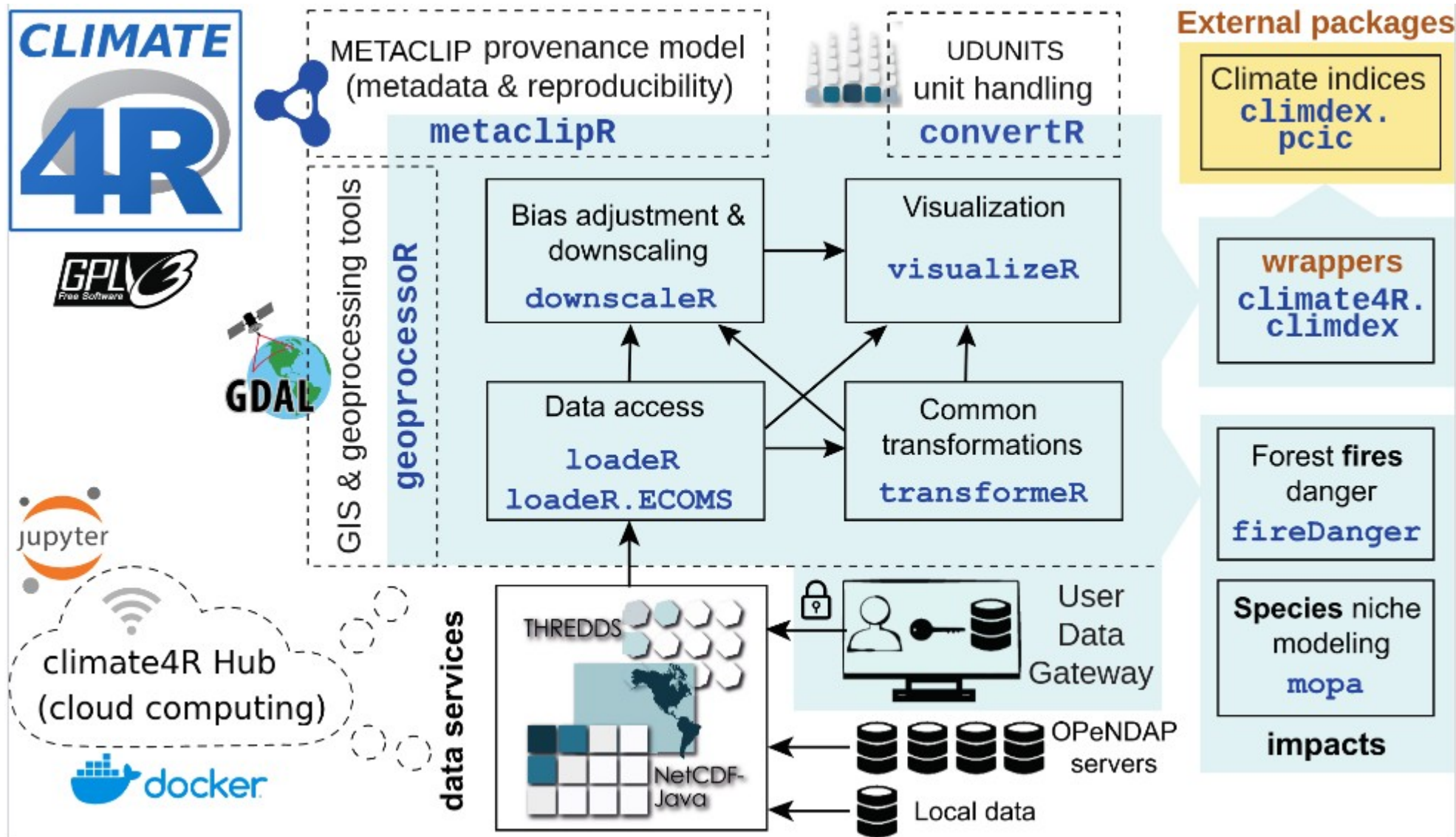
--Conocer los portales de uso en un área crítica  
--Saber modelar procesos críticos pueden i

### 4. OBJETIVOS

Los data lab propiamente expertos en las disciplinas sociales, etc.) las data lab se centra en minería de datos.

### 7. MÉTODOS DE LA EVALUACIÓN

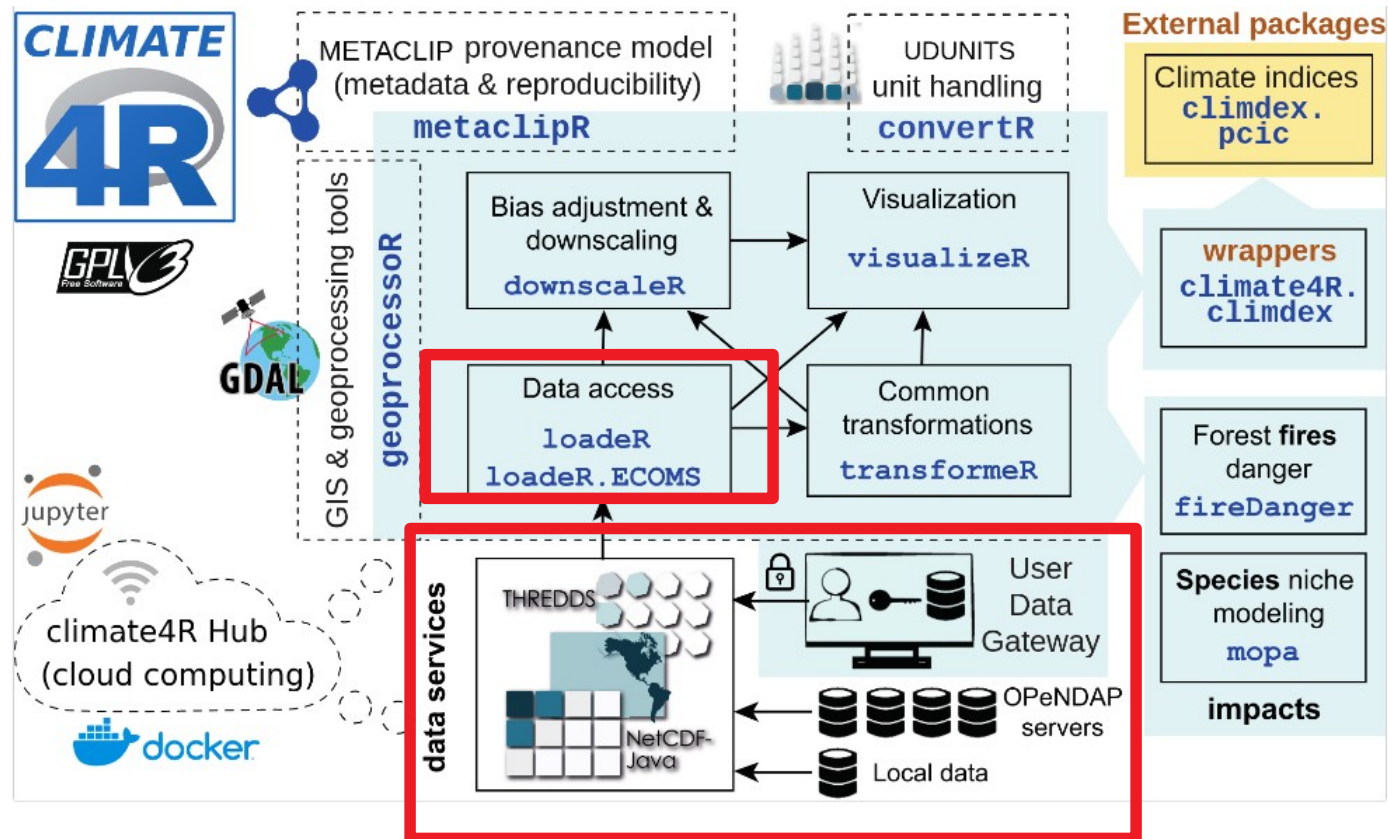
Descripción	Tipología	Eval. Final	Recuper.	%
Valoración de informes y trabajos escritos	Actividad de evaluación con soporte virtual	Sí	Sí	60,00
Calif. mínima	3,00			
Duración				
Fecha realización	Según el calendario			
Condiciones recuperación				
Observaciones	Evaluación de los trabajos de grupo e individuales entregados por el alumno.			
Valoración de exposiciones orales	Trabajo	Sí	Sí	40,00
Calif. mínima	0,00			
Duración				
Fecha realización	Durante el periodo de impartición de la asignatura.			
Condiciones recuperación				
Observaciones				
Realización de prácticas	Evaluación en laboratorio	No	No	0,00
Calif. mínima	0,00			
Duración				
Fecha realización	en las sesiones prácticas			
Condiciones recuperación				
Observaciones				
TOTAL				100,00
Observaciones				
Si la nota final del alumno fuese menor que 5 sobre 10, entonces la recuperación consistirá en la realización de cada una de las tareas en las que hubiera obtenido una calificación menor que 5 sobre 10. El procedimiento de evaluación de una actividad recuperable será equivalente al de la actividad original.				
Observaciones para alumnos a tiempo parcial				
Será necesaria la asistencia, por lo que se aplicará la misma evaluación que a los estudiantes a tiempo completo, teniendo en cuenta su disponibilidad de tiempo en cuanto a los plazos de entrega de los trabajos.				



## Tools for big data access and subsetting

### User Data Gateway: Santander Meteorology Group Database Repository

- Cofiño *et al.* 2018 (<http://dx.doi.org/10.1016/j.cliser.2017.07.001>)
- Iturbide *et al.* 2019 (<https://doi.org/10.1016/j.envsoft.2018.09.009>)

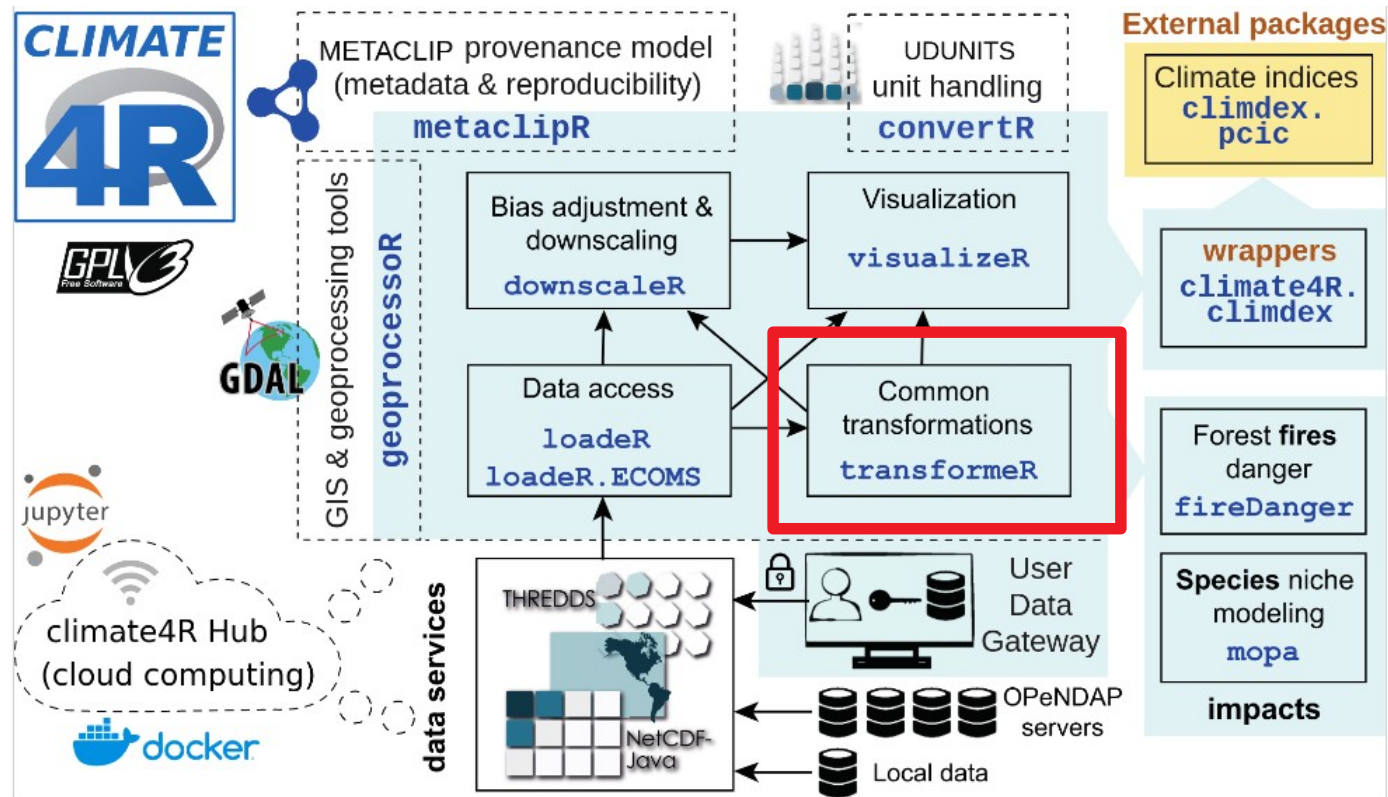




## Tools for climate data transformation.

- Common operations: subsetting, aggregation...
- Data Mining: clustering algorithms, principal component analysis...
- Others

See e.g., Iturbide *et al.* 2019 (<https://doi.org/10.1016/j.envsoft.2018.09.009>)

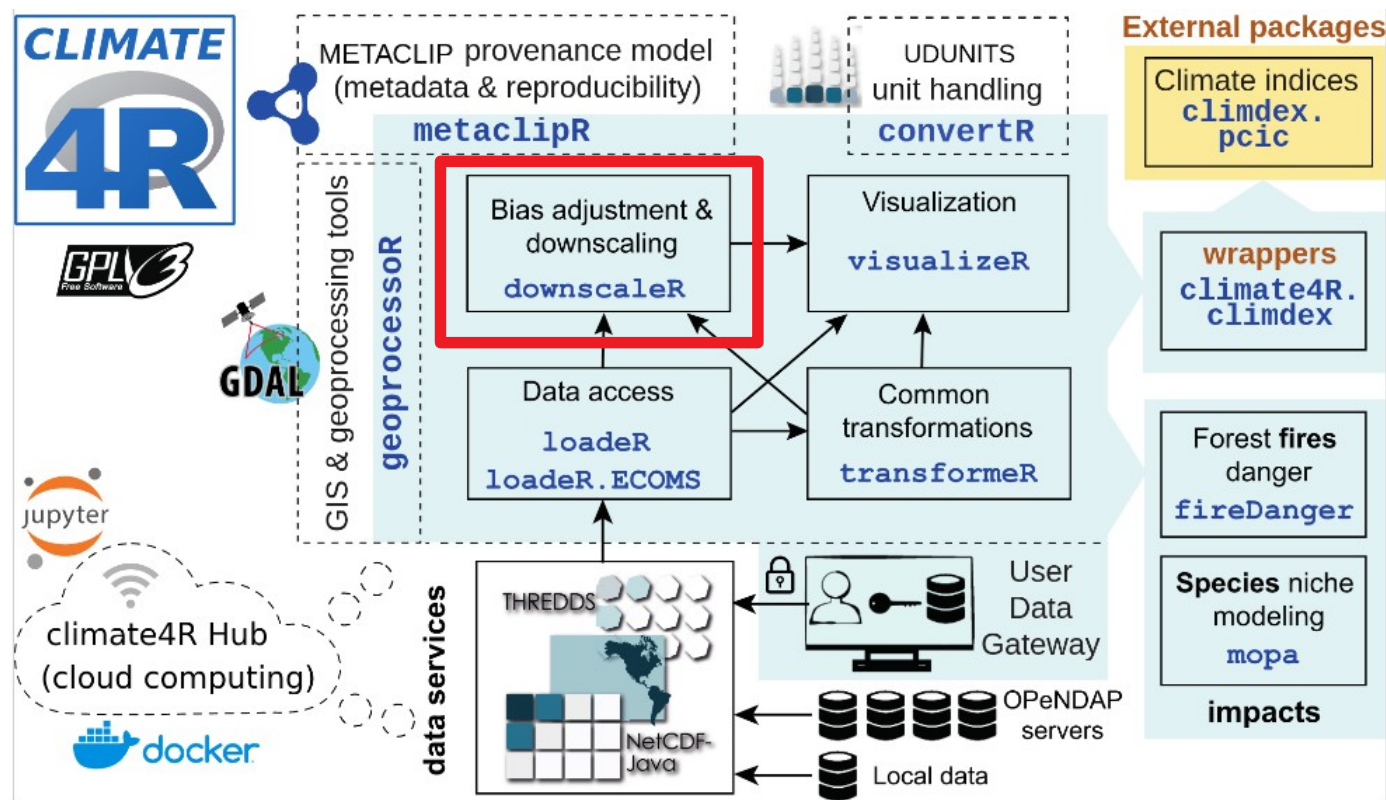


## Tools for **statistical downscaling** and bias correction of climate data

- Machine learning methods for statistical modelling and learning relationships between broad-scale and local-scale climate features
- Statistical treatment of climate model simulations (bias adjustment)

Bias adjustment: Iturbide *et al.* 2019 (<https://doi.org/10.1016/j.envsoft.2018.09.009>)

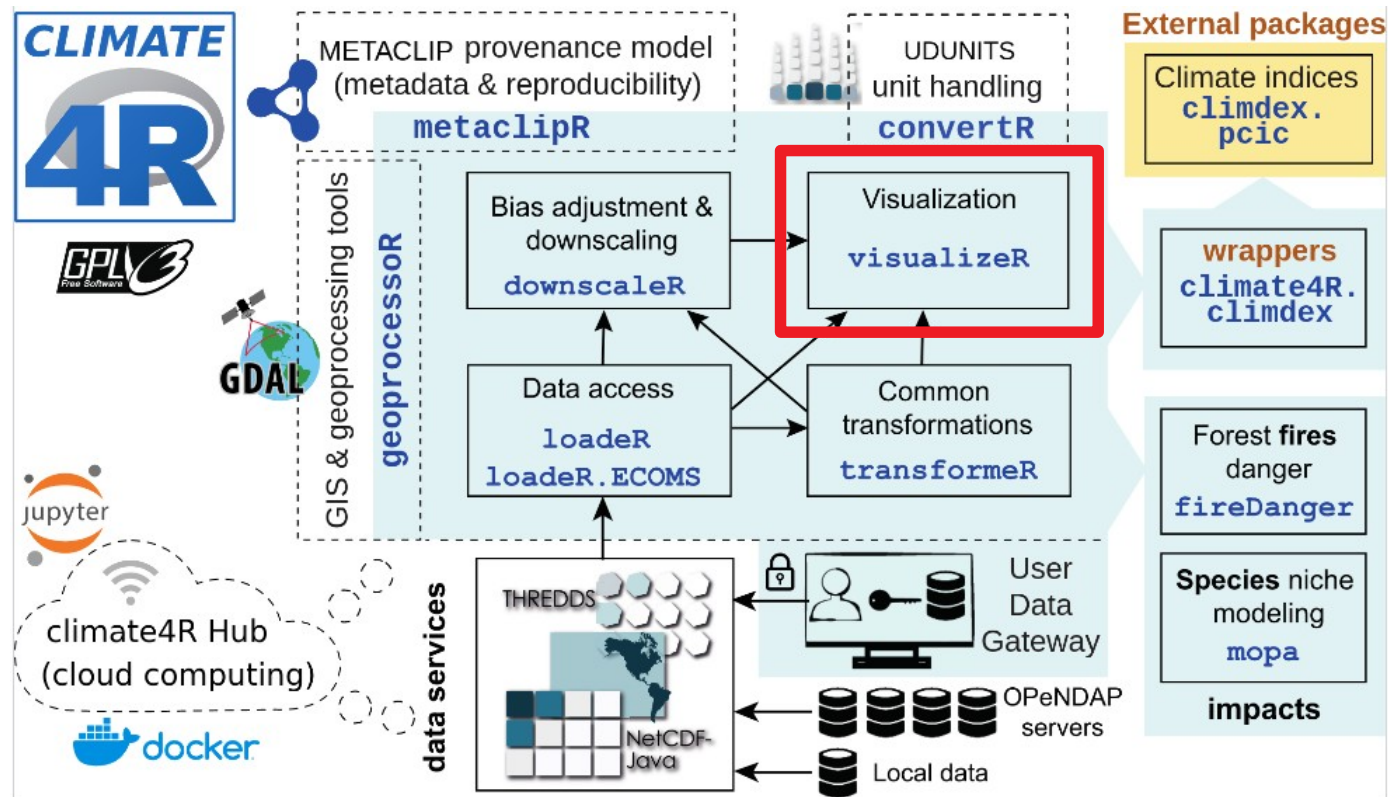
Statistical downscaling: Bedia *et al.* 2020 (<https://www.geosci-model-dev-discuss.net/gmd-2019-224/>)



## Climate data **visualization**

- General visualization capabilities (maps, time series, specific plots...)
- Communication of uncertainty in ensemble prediction through special plots

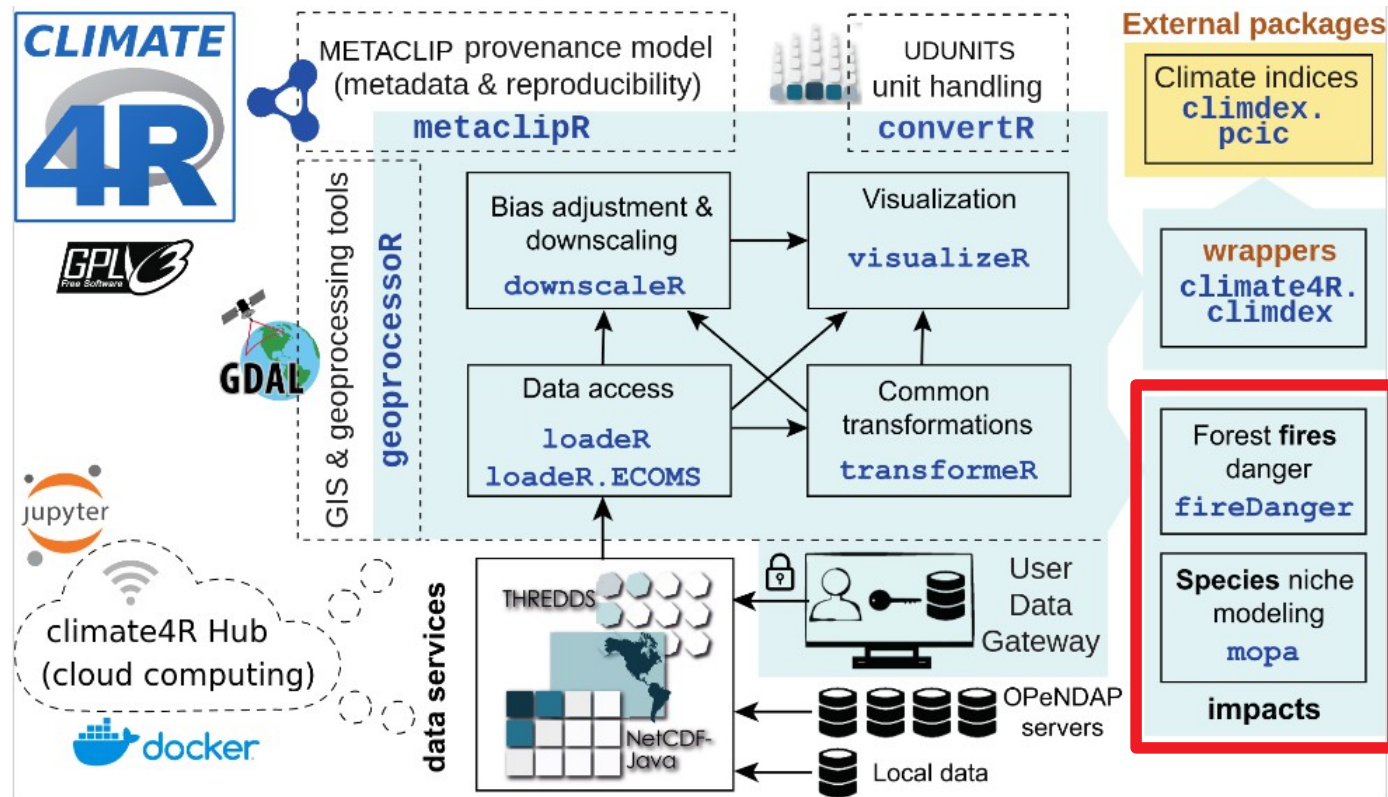
Frías *et al.* 2018 (<https://doi.org/10.1016/j.envsoft.2017.09.008>)





## Specific applications of climate4R on **sectorial impact studies**

- Forest Fire Danger prediction on seasonal time scales. Bedia *et al.* 2018  
<http://dx.doi.org/10.1016/j.cliser.2017.04.001>
- Species Distribution Modelling. Iturbide *et al.* 2018  
<https://journal.r-project.org/archive/2018/RJ-2018-019/index.html>





## Metadata encoding and data provenance tracking

- Based on the METACLIP semantic metadata framework  
<https://doi.org/10.1016/j.envsoft.2019.07.005>
- See also: <http://www.metaclip.org/>

