Hybrid Modelling of Land-Atmosphere Fluxes: Estimating Evapotranspiration using Combined Physics-Based and Data-Driven Machine Learning

Reda ElGhawi[1,2,3], Alexander J. Winkler[1], Basil Kraft[1,2], Christian Reimers[1], Marco Körner[3], and Markus Reichstein[1]

[1] Max Planck Institute for Biogeochemistry
Biogeochemical Integration
Jena, Germany

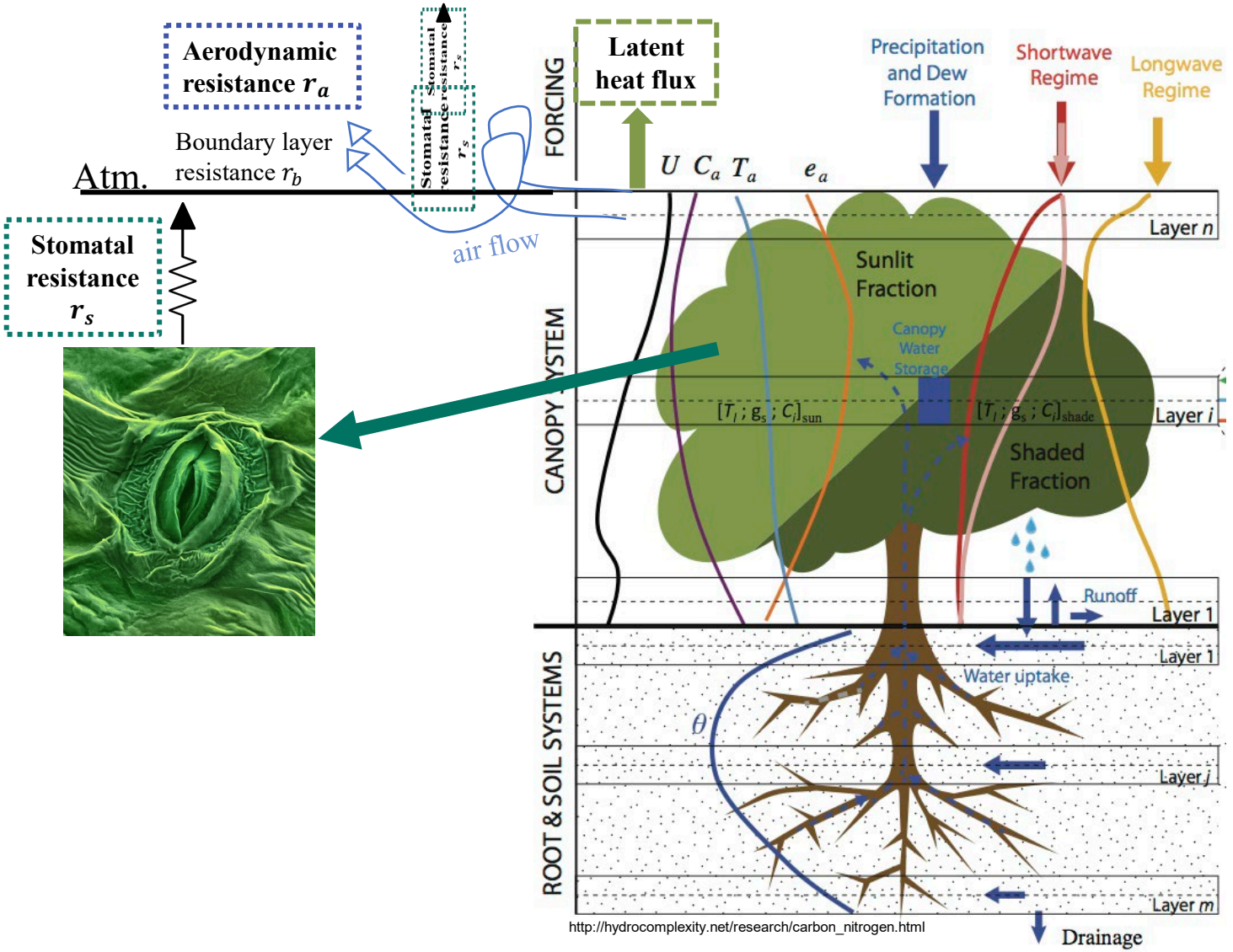
[2] International Max Planck Research School
for Global Biogeochemical Cycles
Jena, Germany

www.bgc-jena.mpg.de/bgi
relghawi@bgc-jena.mpg.de

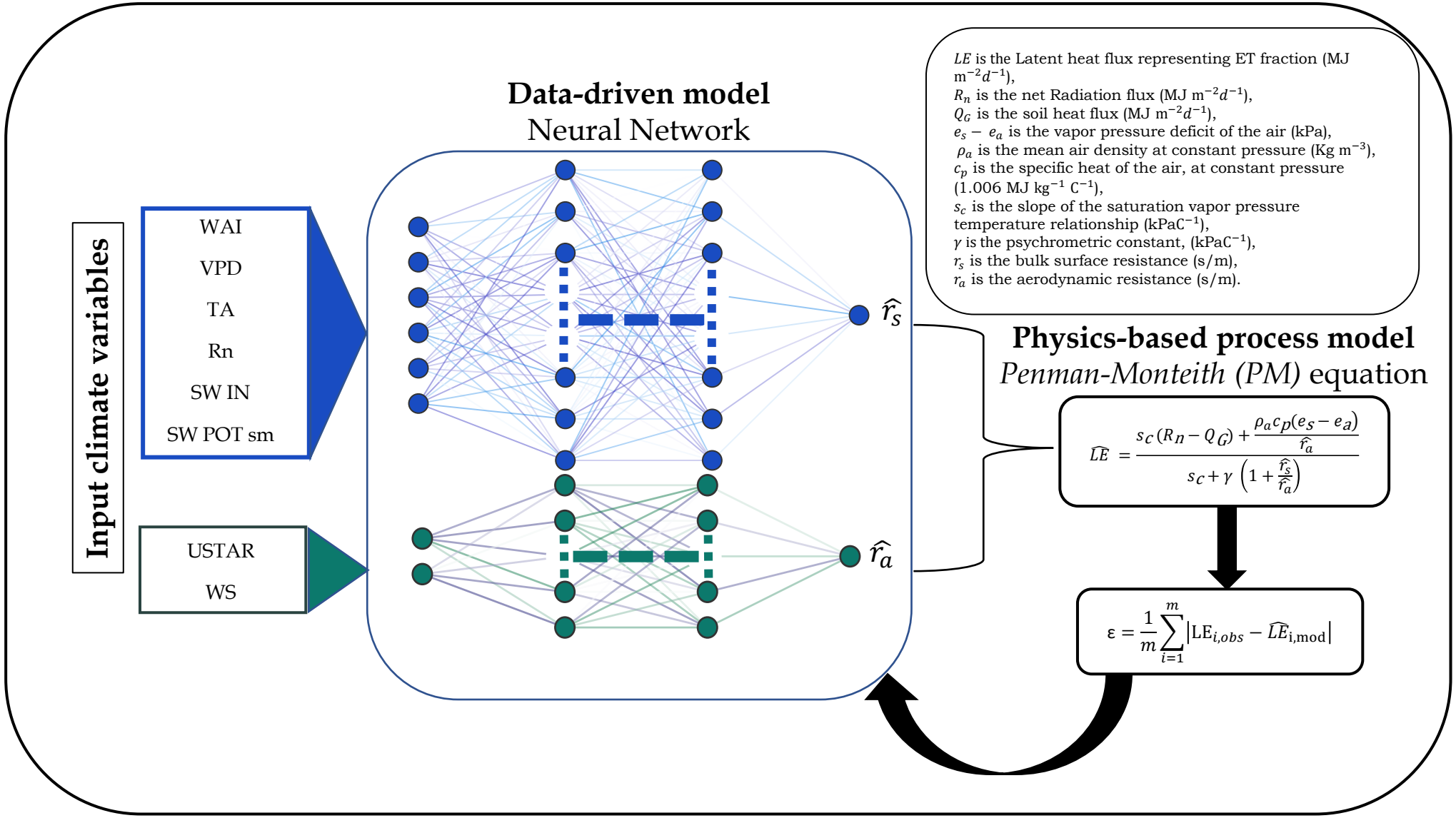
[3] Technical University of Munich
TUM School of Engineering and Design
Department of Aerospace and Geodesy
Chair of Remote Sensing Technology
Munich, Germany

www.asg.ed.tum.de/lmf
marco.koerner@tum.de

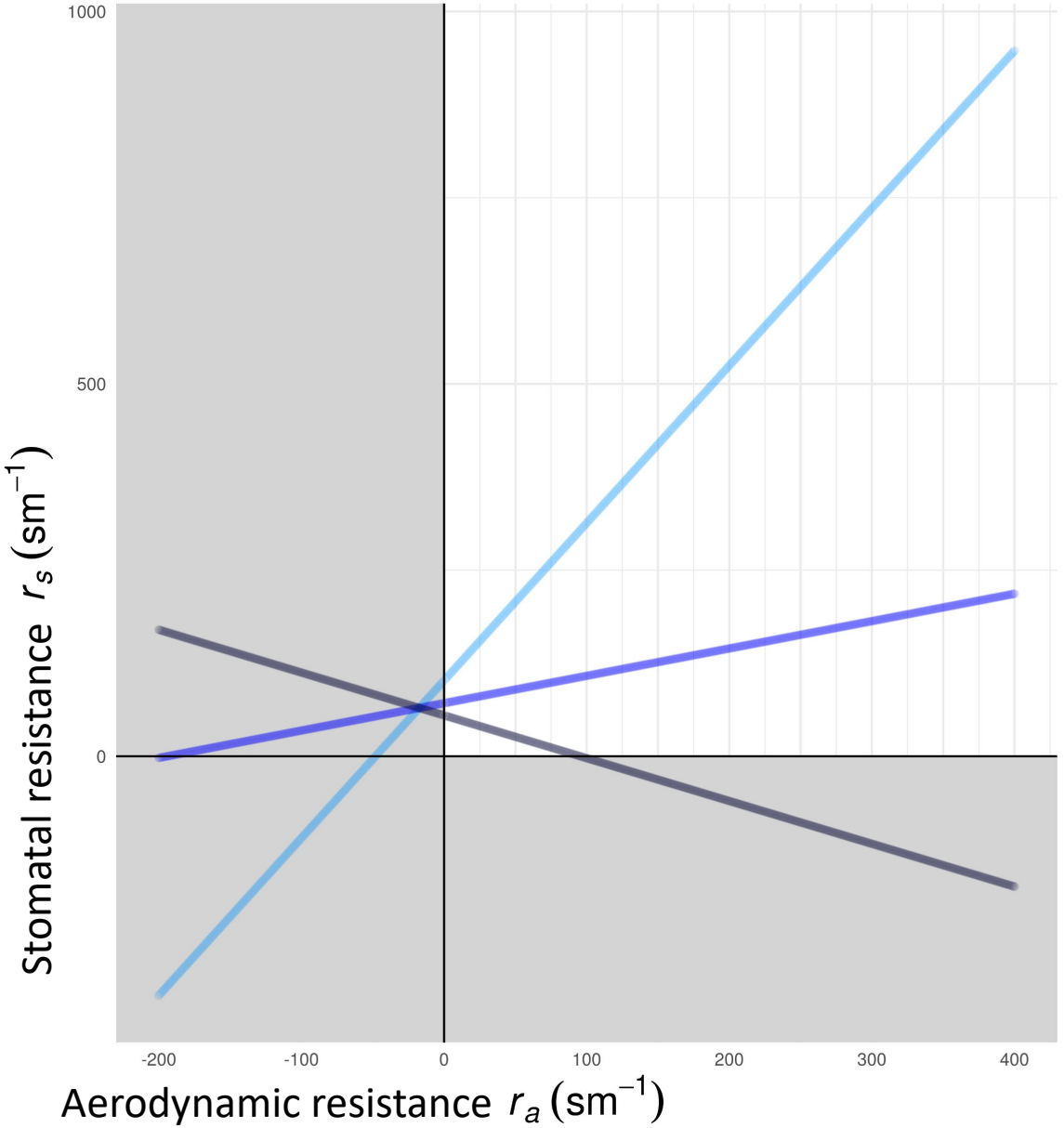
Evapotranspiration as a Key Flux: Looking into *Stomatal* and *Aerodynamic Resistance*



Hybrid Model



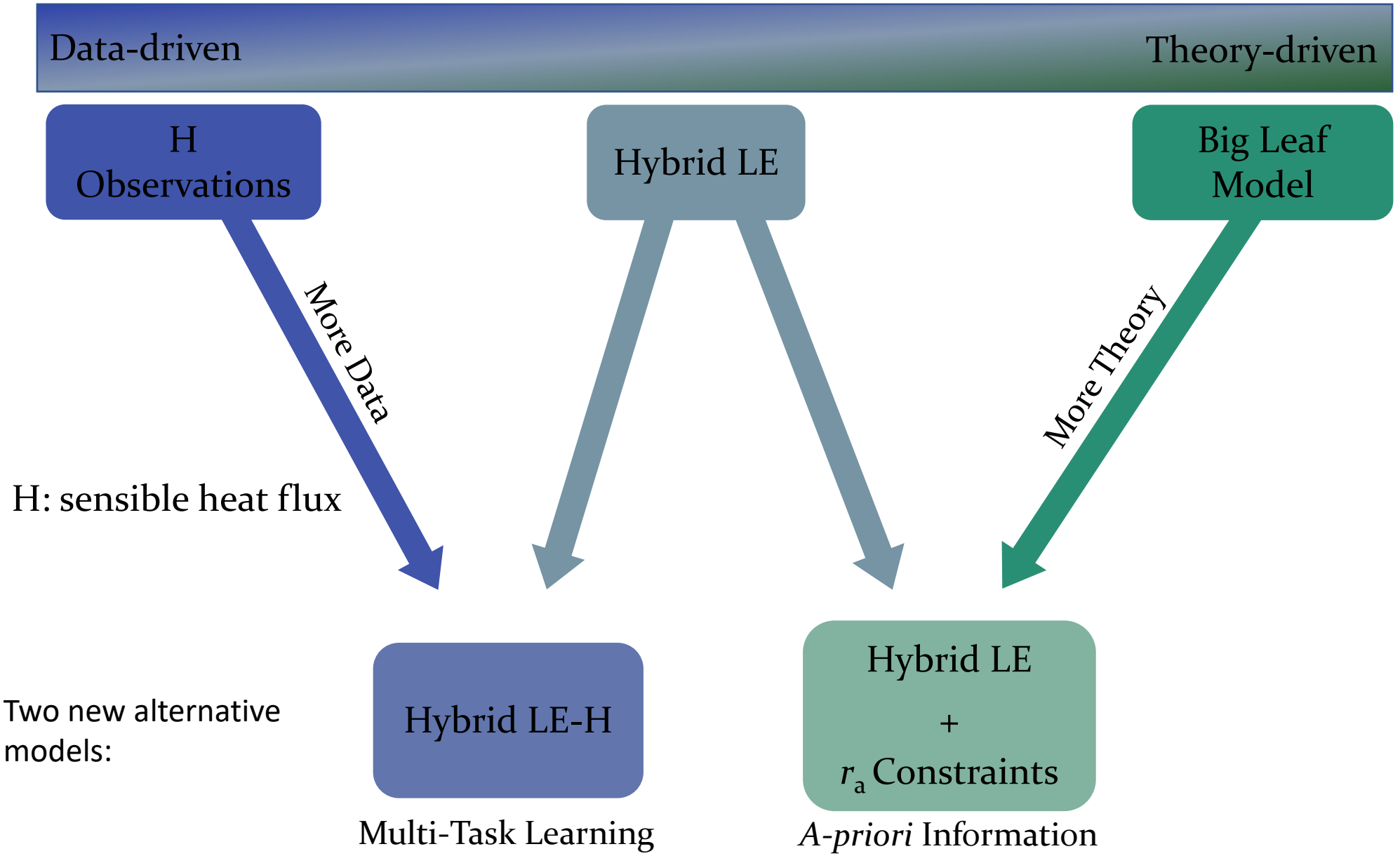
Equifinality in the physics-based component of unconstrained hybrid models



Latent heat flux values (in Wm^{-2}):

- LE=234
- LE=334
- LE=434

$$\widehat{LE} = \frac{s_C(R_n - Q_G) + \frac{\rho_a c_p (e_S - e_a)}{\hat{r}_a}}{s_C + \gamma\left(1 + \frac{\hat{r}_s}{\hat{r}_a}\right)}$$


Data-driven
Theory-driven
H Observations
Hybrid LE
Big Leaf Model
More Data
More Theory
H: sensible heat flux
Two new alternative models:
Hybrid LE-H
Multi-Task Learning
Hybrid LE + r_a Constraints
A-priori Information

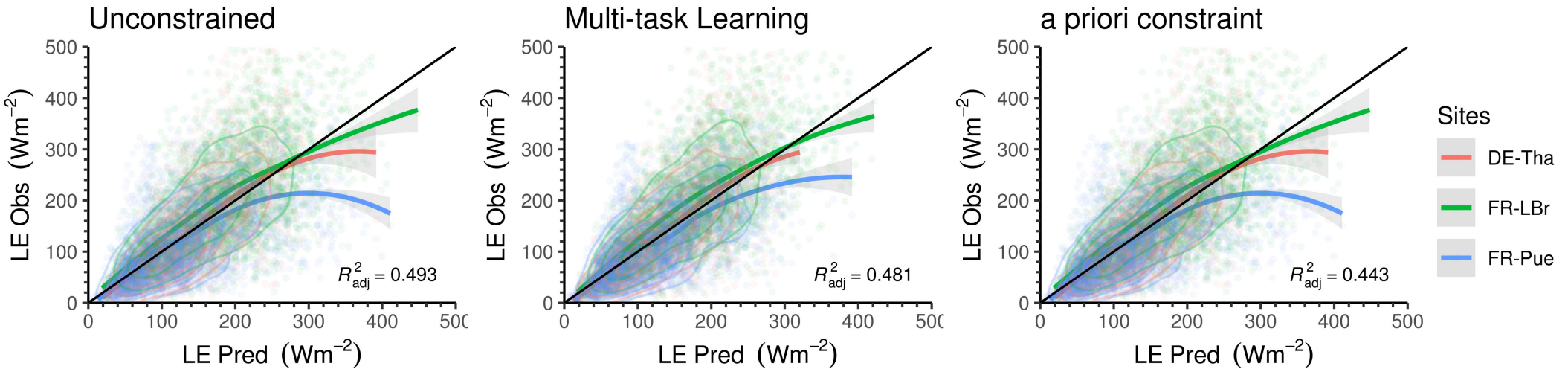
The FLUXNET 2015 dataset

Site ID	IGBP	Elevation (m)	Mean Annual Temperature (°C)	Mean Annual Precipitation (mm)	Data Availability
DE-Tha	ENF	385	8.2	843	19 years (1996 - 2014)
FR-Pue	EBF	270	13.5	883	15 years (2000 - 2014)
FR-LBr	ENF	61	13.6	900	12 years (1996 - 2008)
CH-Cha	GRA	393	9.5	1136	10 YEARS (2005 - 2014)
DE-Gri	GRA	385	7.8	901	11 YEARS (2004 - 2014)
US-Var	GRA	129	15.8	559	15 YEARS (2000 - 2014)

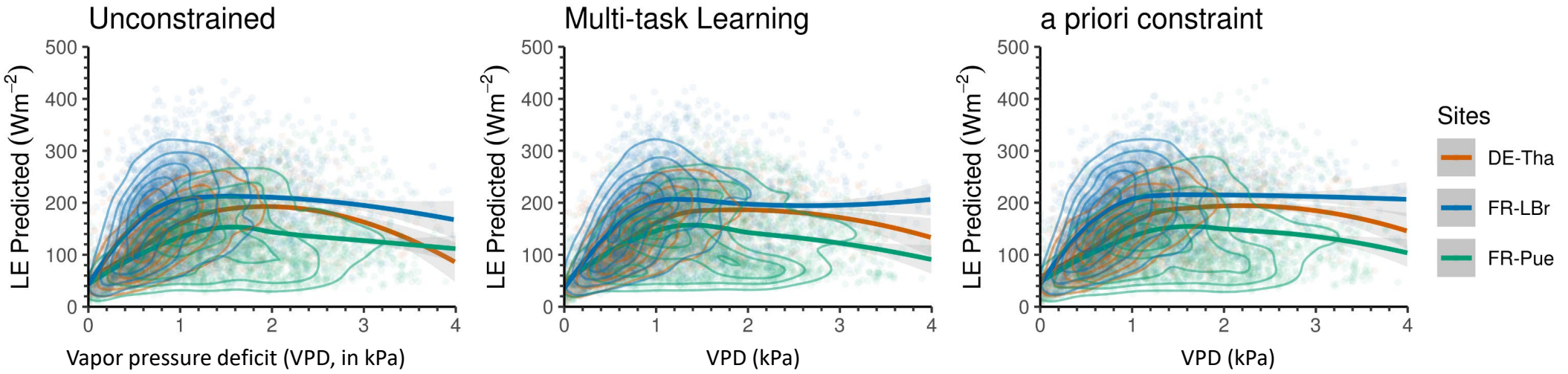
•ENF (Evergreen Needleleaf Forests: Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 meters. Almost all trees remain green all year. Canopy is never without green foliage).
•EBF (Evergreen Broadleaf Forests: Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 meters. Almost all trees and shrubs remain green year-round. Canopy is never without green foliage).
•GRA (Grasslands: Lands with herbaceous types of cover. Tree and shrub cover is less than 10%. Permanent wetlands lands with a permanent mixture of water and herbaceous or woody vegetation. The vegetation can be present in either salt, brackish, or fresh water.)

MAX PLANCK INSTITUTE
FOR BIOGEOCHEMISTRY

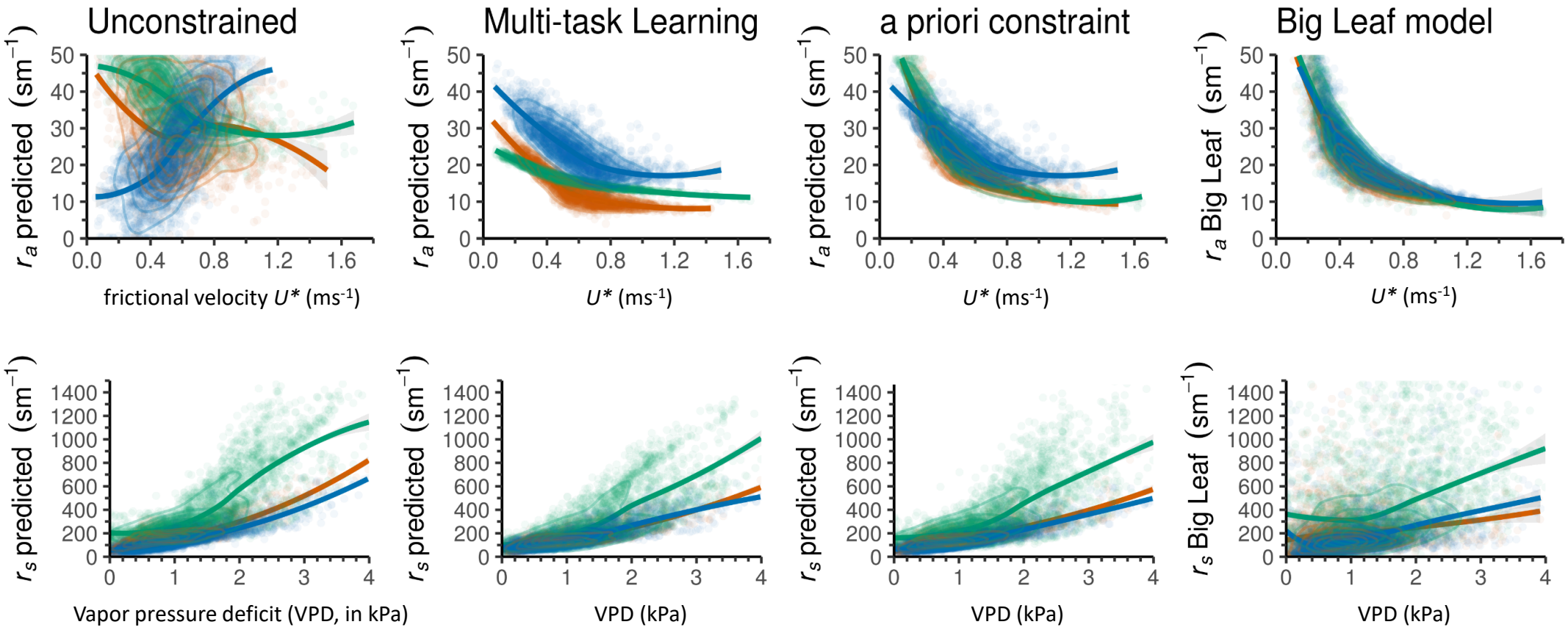
Evaluating *predictions* against *observations*



Evaluating *latent heat flux* predictions against relevant *climate variables*



Evaluating *predictions* against relevant *climate variables*



Predicted latent variable evaluated from unconstrained model against constrained model		R^2 metric value averaged over sites
r_s latent variable	a priori constraint	R^2 =0.901
	Multi-Task Learning	R^2 =0.826
r_a latent variable	a priori constraint	R^2 =0.055
	Multi-Task Learning	R^2 =0.261

Conclusion

- Hybrid modelling allows
  - the retrieval of *latent variables*
  - that are *physically interpretable*

  in comparison to both *purely data-driven* and *physics-based approaches*, and
- *Equifinality* can be circumvented by inducing more theory or data
- Data-driven modelling ***needs*** to be boosted by domain expertise!